Unravelling the role of the human Dihydrofolate reductase 2 gene

Niamh Bookey, BSc

Supervisor: Prof. Anne Parle-McDermott

School of Biotechnology, Dublin City University

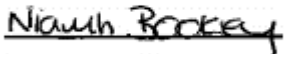
A thesis submitted for the degree

of Doctor of Philosophy



August 22

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy 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)

ID No: 17213243

Date: 04/08/22

“Our greatest glory is not in never falling, but in rising every time we fall”

- Oliver Goldsmith

“There is no limit to what we, as women, can accomplish”

- Michelle Obama

“Even though this life, this love is brief. I’ve got some people who carry me”

- Dermot Kennedy

Acknowledgements

First of all, I would like to thank my supervisor Prof. Anne Parle-McDermott. From our first encounter during my undergraduate degree, to the pursuit of this PhD you have always provided me with great encouragement, support and self-belief. I am so unbelievable grateful for your guidance throughout the last 9 years. Thank you for believing in me and giving me the opportunity to complete this PhD.

I want to take the opportunity to also express my appreciation towards all my colleagues that I have met on my PhD journey. A huge thank you to our collaborators, Prof. Nicholas Greene and Dr. Kit-Yi Leung located in University College London. The significant results revealed in this thesis would not have been uncovered without your invaluable support. A special thank you to all the lecturers and technical staff in DCU that helped me throughout my PhD. I am extremely grateful for all the hard work you put in during the pandemic so that I could continue my research. I would also like to thank Dr. Michael Henry from the proteomics facility in DCU for the many hours he spent with me brainstorming potential experiments, teaching me about mass spectrometry and explaining experimental results. Moreover, I want to thank my examiners Dr. Ciaran Fagan and Prof. Lorraine Brennan for the interesting discussions we had during my viva and for making it such an enjoyable experience.

A PhD is often considered a lonely experience, this was not the case for me as I was lucky enough to have shared the journey with some of my closest friends. Paola, Molly, Darren and Conor, I am so thankful that we were put together at the start of each of our journeys. You are all fantastic scientist and I feel privileged to have worked alongside each of you. I cannot wait to see what you all achieve next in your careers. I am even more thankful for the friendship we have developed over the last 5 years. You have been there for me during the challenging parts and there to celebrate all my achievements; and for that I am very grateful.

I had the privilege of meeting some amazing people on my Dyslexia journey; from teachers to fellow Dyslexics. I would like to thank each and every one of you for your guidance and support throughout the last 2 decades. There is no way this PhD thesis would have been possible without you.

To all my friends outside academia, I want to take the opportunity to express my gratitude towards you. Brid, Katie, Keeshi, Lauren and Sadhbh you have all provided much needed distractions from the demanding PhD life; from the hundreds of walks during the pandemic, the countless amounts of concerts to travelling the world together. I am eternally grateful for your encouragement and support throughout the last 5 years.

The exact same can be said for my extremely supportive family. My grandparents, aunties, uncles and cousins, your curiosity about my research was always welcomed; even if you didn't fully understand it.

A very special thank you has to go to Emma and Sean. We may have killed each other as kids but I am so thankful for our relationship now. You have provided a safe space during this stressful PhD and for that I am so appreciative. I am even more thankful for your encouragement throughout the last 5 years; it motivated me on days I felt like giving up. A shout out to Mike and Penny for allowing me to stay in their house for the last year.

Finally, Daddy and Mammy, words can't describe how thankful I am for your love, support and guidance over the last 3 decades. I am eternally grateful for all the time you dedicated to me and money you spent on me during my Dyslexia journey. None of this would have been possible if you hadn't got me diagnosed or sat with me reading nonsense words from Toe by toe. We have all come a long way from reading the Tara and Ben books. From the bottom of my heart, thank you for the lifetime of support.

This work was funded by Science Foundation Ireland (SFI) in collaboration with the Biotechnology and Biological Sciences Research Council (BBSRC): 17/BBSRC/3416.

Table of contents

Acknowledgments....iii
Table of contents....vi
Abbreviations....x
Units....xix
List of figures....xxi
List of tables....xxiv
Abstract....xxvi

CHAPTER 1. INTRODUCTION....1

1.1. OVERVIEW....2
1.2. VITAMIN B_9....2
1.2.1. History of folate and folic acid....3
1.2.2. Folate and folic acid chemical structures....3
1.2.3. Absorption of folate and folic acid....5
1.3. ONE-CARBON METABOLISM....6
1.3.1. One-carbon metabolism in the mitochondria....7
1.3.2. One-carbon metabolism in the cytoplasm....8
1.3.3. One-carbon metabolism in the nucleus....9
1.4. ONE-CARBON METABOLISM IN HEALTH AND DISEASE....11
1.4.1. One-carbon metabolism and neural tube defects....11
1.4.2. One-carbon metabolism and other diseases....12
1.5. DIHYDROFOLATE REDUCTASE GENE FAMILY....14
1.5.1. Dihydrofolate reductase gene....15
1.5.2. Dihydrofolate reductase 2 gene....21
1.6. AIMS AND OBJECTIVES....26

CHAPTER 2. MATERIALS AND METHODS....28

2.1. MATERIALS....29
2.1.1. Reagents....29
2.1.2. Solutions....31
2.1.3. Cell lines....33
2.1.4. Tissue samples....34
2.1.5. RNA samples....34
2.2. METHODS....35
2.2.1. Generation of HepG2 DHFR knockdown and DHFR2 knockout cell lines....35
2.2.2. General cell culture methods....35
2.2.3. DNA extraction from cultured cells....36
2.2.4. RNA extraction from cultured cells....36
2.2.5. DNA/RNA quantification....36
2.2.6. RNA integrity check....36
2.2.7. Agarose gel electrophoresis....37
2.2.8. DNase treatment and reverse transcription (RT) of RNA samples....37
2.2.9. Bioinformatic tools used for primer design....38
2.2.10. Genomic contamination check on newly synthesised cDNA....39
2.2.11. Investigative polymerase chain reactions (PCR)....39
2.2.12. PCR product clean-up for Sanger Sequencing....41
2.2.13. Blunt-end cloning of PCR product into a cloning vector for Sanger Sequencing....42

2.2.14. Cloning vector transformed into competent *E. coli* cells 42
2.2.15. Isolation and purification of vector for Sanger Sequencing 42
2.2.16. Sanger Sequencing of purified PCR products and purified plasmids 43
2.2.17. Bioinformatic tools used for analysing Sanger Sequencing results 43
2.2.18. Quantitative polymerase chain reaction (qPCR) 44
2.2.19. Protein extraction from cell lines and tissues 45
2.2.20. Sample preparation for liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis 46
2.2.21. LC-MS/MS analysis 47
2.2.22. Western Blot analysis 50

CHAPTER 3. INVESTIGATING THE TRANSLATION STATUS OF THE HUMAN *DHFR2* GENE THROUGH ITS ASSOCIATION WITH THE RIBOSOME 52

3.1. INTRODUCTION 53
3.1.1. Overview 53
3.1.2. The role of the ribosome during translation 53
3.1.3. Ribosome profiling as a method to study the translatability of an RNA molecule 54
3.1.4. Polysome profiling as a method to study the translatability of an RNA molecule 57
3.1.5. Sucrose cushion ultracentrifugation RT-qPCR as a method to study the translatability of an RNA molecule 57
3.1.6. Aims and objectives 59
3.2. METHOD 59
3.2.1. Examining the relationship between the DHFR2 RNA and the ribosome through bioinformatic analysis of Ribo-seq datasets 59
3.2.2. Investigation of the DHFR2_202 transcript in four cell lines by RT-PCR analysis 60
3.2.3. Examining the relationship between the ribosome and the DHFR2_201/202 transcripts by performing the sucrose cushion ultracentrifugation RT-qPCR experiment 61
3.3. RESULTS 65
3.3.1. Ribo-seq data reveals the DHFR2 RNA associates with the ribosome 65
3.3.2. Transcript DHFR2_202 confirmed by RT-PCR for sucrose cushion experiment 70
3.3.3. Ribosome fractionation by sucrose cushion ultracentrifugation shows the DHFR2 transcripts differentially associate with the ribosome 75
3.4. DISCUSSION 80

CHAPTER 4. INVESTIGATING THE TRANSLATABILITY OF THE HUMAN DHFR2 ORF THROUGH THE GENERATION OF THE RECOMBINANT DHFR2 PROTEIN 85

4.1. INTRODUCTION 86
4.1.1. Overview 86
4.1.2. Generation of recombinant protein and its function 87
4.1.3. Aims and Objectives 89
4.2. METHOD 90
4.2.1. Transfection of pCMV6-AC-DHFR2 vectors into HepG2 DHFR2 knockout cell line 90
4.2.2. Generation of humanised DHFR2 mouse model 90
4.2.3. Protein extraction for Western Blot and LC-MS/MS analysis 91
4.2.4. Western Blot analysis of HepG2 DHFR2 knockout cell line transfected with pCMV6-AC vectors 92
4.2.5. LC-MS/MS sample preparation 92
4.2.6. Searching for recombinant DHFR2 specific peptides using data-dependent acquisition (DDA) LC-MS/MS analysis 92

4.2.7. Searching for recombinant DHFR2 specific peptides using parallel reaction monitoring (PRM) LC-MS/MS analysis 93
4.3. RESULTS 93
4.3.1. Recombinant DHFR2 protein was synthesised in HepG2 DHFR2 knockout cells 93
4.3.2. Recombinant DHFR2 protein was not expressed in humanised DHFR2 mouse model 98
4.4. DISCUSSION 101

CHAPTER 5. INTERROGATING THE ENDOGENOUS HUMAN DHFR2 PROTEIN 104

5.1. INTRODUCTION 105
5.1.1. Overview 105
5.1.2. The proteome and how it is studied 105
5.1.3. MS as a technique for identifying the composition of a sample 106
5.1.4. Shotgun LC-MS method for identifying the composition of a sample 107
5.1.5. LC-MS/MS method for identifying the composition of a sample 108
5.1.6. Aims and objectives 111
5.2. METHODS 112
5.2.1. DDA and PRM LC-MS/MS analysis on adult cell lines/tissue, HepG2 DHFR knockdown cells and proteasome inhibitor-treated HepG2 cells 112
5.2.2. MRM LC-MS/MS analysis on human embryonic and foetal tissues 114
5.3. RESULTS 118
5.3.1. The DHFR2 protein was not detected in any of the adult cell lines or tissues examined 118
5.3.2. The DHFR2 protein is not synthesised to carry out the reductase activity when the expression of the DHFR enzyme is reduced 122
5.3.3. The DHFR2 protein is not subject to ubiquitination 123
5.3.4. The DHFR2 protein was detected in human embryonic and foetal heart tissue 125
5.4. DISCUSSION 131

CHAPTER 6. ARE THE DHFR2 TRANSCRIPTS SUBJECT TO TRANSLATIONAL INHIBITION VIA PARASPECKLE NUCLEAR RETENTION? 136

6.1. INTRODUCTION 137
6.1.1. Overview 137
6.1.2. The nucleus and its subdomains 137
6.1.3. The paraspeckle 138
6.1.4. DHFR2's IRAlu elements and possible A to I RNA editing sites 141
6.1.5. Aims and Objectives 141
6.2. METHODS 142
6.2.1. Analysing transcriptome data to determine if the IRAlu elements in the DHFR2 RNA are subject to A to I editing 142
6.2.2. Determining the 3'UTR DHFR2 genomic sequence in the HepG2 cell line to use as the reference sequence for the investigative RT-PCR 142
6.2.3. Investigative RT-PCR to determine if the DHFR2 transcripts are subject to A to I editing in the HepG2 cell line 143
6.3. RESULTS 144
6.3.1. The REDIportal V2.0 database revealed multiple A to I edits in the DHFR2 IRAlu elements 144
6.3.2. RT-PCR on the DHFR2 IRAlu elements revealed possible A to I RNA edited sites in HepG2 cells 146
6.4. DISCUSSION 151

CHAPTER 7. MODULATING THE EXPRESSION OF THE HUMAN *DHFR2* GENE TO DETERMINE ANY ADDITIONAL BIOLOGICAL FUNCTIONS 154

7.1. INTRODUCTION 155
7.1.1. Overview 155
7.1.2. Modulating the expression of a specific gene for functional analysis 155
7.1.3. Differential expression analysis on knockout and overexpression models 159
7.1.4. Aims and Objectives 160
7.2. METHODS 160
7.2.1. Cell culture preparation for differential protein expression analysis 160
7.2.2. Protein extraction and LC-MS/MS sample preparation 161
7.2.3. LC-MS/MS analysis 161
7.2.4. Differential protein expression analysis using Progenesis QI 162
7.3. RESULTS 163
7.3.1. Knocking out the dihydrofolate reductase genes impacted the expression of hundreds of other proteins in HepG2 cells 163
7.3.2. Overexpressing the *DHFR2* gene impacted the expression of several hundred other proteins in HepG2 cells 170
7.4. DISCUSSION 176

CHAPTER 8. CONCLUSION 181

8.1. GENERAL 182
8.2. CHAPTER OBSERVATIONS 183
8.3. LIMITATIONS ASSOCIATED WITH THIS PHD THESIS 189
8.4. CONTRIBUTIONS AND CONCLUDING REMARKS 189
8.5. FUTURE WORK 190
8.5.1. Recommended experiments for assessing the translational capabilities of the DHFR2 transcripts 190
8.5.2. Recommended experiments for the recombinant DHFR2 protein 191
8.5.3. Recommended experiments for investigating the endogenous DHFR2 protein 191
8.5.4. Recommended experiments for investigating the translational regulation of the DHFR2_201 and DHFR2_202 transcripts 192
8.5.5. Recommended experiments to determine which DHFR molecule the DHFR2 RNA binds to and regulates 192

BIBLIOGRAPHY 193

APPENDICES 229

Abbreviations

10-CHO-THF	10-formyl-THF
101B	Human induced pluripotent stem (iPS)-derived hepatocytes cell line
17158-Coriell	B-Lymphocyte cell line
5-CH_3-THF	5-methyl-THF
5,10-CH^+-THF	5,10-methenyl-THF
5,10-CH_2-THF	5,10-methylene-THF
A to I	Adenosine to inosine
ACADVL	Acyl-CoA dehydrogenase very long chain
ACOX1	Acyl-CoA oxidase 1
ACTB	Beta-actin
ADAR	Adenosine deaminases
ADP	Adenosine diphosphate
AGC	Automatic gain control
AHCY	S-adenosylhomocysteine hydrolase
ALDH1L1	Aldehyde dehydrogenase family 1 member L1
ALDH1L2	Aldehyde dehydrogenase 1 family member L2
ANOVA	An analysis of variance
ASNS	Asparagine synthetase (glutamine-hydrolyzing)
ATP	Adenosine triphosphate
BBSRC	Biotechnology and Biological Sciences Research Council
BH2	Dihydrobiopterin
BH4	Tetrahydrobiopterin
BHMT	Betaine--homocysteine S-methyltransferase
BSA	Bovine serum albumin

cDNA	Complementary DNA
CDS	Coding sequence
CEMIP2	Cell migration inducing hyaluronidase 2
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonat
CHO	Chinese hamster ovary cells
CHX	Cycloheximide
CID	Collision-induced dissociation
CMV	Cytomegalovirus
CO_2	Carbon dioxide
CORO1B	Coronin 1B
COX4I1	Cytochrome C oxidase subunit 4I1
CPS1	Carbamoyl-phosphate synthase 1
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
CSS	Cushion sucrose solution
CVD	Cardiovascular disease
DBHS	Drosophila behaviour human splicing
DCU	Dublin City University
DDA	Data-dependent acquisition
DHF	Dihydrofolate
DHFR	Dihydrofolate reductase
DHFR2	Dihydrofolate reductase 2
DHFRP1	Dihydrofolate reductase pseudogene 1
DHFRP2	Dihydrofolate reductase pseudogene 2
DHFRP3	Dihydrofolate reductase pseudogene 3

DHFRP5	Dihydrofolate reductase pseudogene 5
DHFRP6	Dihydrofolate reductase pseudogene 6
DMEM	Dulbecco's modified eagle medium
DMGDH	Dimethylglycine dehydrogenase
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNA-PKcs	DNA-dependent protein kinase catalytic subunit
dNTP	Deoxynucleotide triphosphates
DSB	Double-stranded break
dsRNA	Double-stranded RNA
dTMP	Deoxythymidine monophosphate
DTT	Dithiothreitol
dUMP	Deoxyuridine monophosphate
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EHHADH	Enoyl-CoA hydratase and 3-hydroxyacyl CoA dehydrogenase
ELISA	Enzyme-linked immunosorbent assay
eNOS	Endothelial nitric oxide synthase
ESI	Electrospray ionisation
ETD	Electron transfer dissociation
FASP	Filter aided sample preparation
FBS	Fetal bovine serum
FDR	False discovery rate
FISH	Fluorescence in situ hybridization
FPGS	Foly-polyglutamate synthase

FR	Folate receptors
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GART	Phosphoribosylglycinamide formyltransferase, phosphoribosylglycinamide synthetase, phosphoribosylaminoimidazole synthetase
GCPII	Glutamate carboxypeptidase II
gDNA	Genomic DNA
GFP	Green fluorescent protein
gRNA	Guide RNA
HCD	High energy collision dissociation
HCl	Hydrochloric acid
HCR	Hybridisation chain reaction
HDR	Homology-directed repair
HEK293	Human embryonic kidney 293 cell line
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
HepG2	Hepatocellular carcinoma cell line
hg38	Genome reference consortium human reference 38
HPLC	High-performance liquid chromatography
HRP	Horseradish peroxidase
HSPD1	Heat shock protein family D (Hsp60) member 1
HT	Sodium hypoxanthine and thymidine
HuH7	Human hepatocellular carcinoma cell line
IMR32	Human neuroblastoma cell line
IPA	Ingenuity Pathway Analysis
iPS	Undifferentiated human induced pluripotent stem cells
IRAlu	Inverted repeated Alu

IRES	Internal ribosome entry site
KCl	Potassium chloride
LB	Luria broth
LC-MS	Liquid chromatography-mass spectrometry
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
LMNB1	Lamin B1
lncRNA	Long noncoding RNA
MALDI	Matrix-assisted laser desorption/ionisation
MAT	Methionine adenosyltransferase
MetRNA	Methionyl-tRNA
MFT	Mitochondrial folate transporter
MG132	Carbobenzoxy-Leu-Leu-leucinal
$MgCl_2$	Magnesium chloride
miRNA	MicroRNA
MMEJ	Microhomology-mediated end joining
MRM	Multiple reaction monitoring
mRNA	Messenger RNA
MS	Mass spectrometry
MS	Methionine synthase
MTFMT	Mitochondrial methionyl-tRNA formyltransferase
MTHFD1	Methylenetetrahydrofolate dehydrogenase, cyclohydrolase and formyltetrahydrofolate synthetase 1
MTHFD1L	Methylenetetrahydrofolate dehydrogenase, cyclohydrolase and formyltetrahydrofolate synthetase 1L
MTHFD2	Methylenetetrahydrofolate dehydrogenase, cyclohydrolase and formyltetrahydrofolate synthetase 2

MTHFD2L	Methylenetetrahydrofolate dehydrogenase, cyclohydrolase and formyltetrahydrofolate synthetase 2L
MTHFR	Methylenetetrahydrofolate reductase
MTR	Methionine synthase
MTX	Methotrexate
$NADP^+$	Nicotinamide adenine dinucleotide phosphate
NADPH	Nicotinamide adenine dinucleotide phosphate, reduced form
NEAT1	Nuclear enriched abundant transcript 1
NEP	Human neuroectodermal progenitor stem cells
NHEJ	Nonhomologous end-joining
NICB	National Institute for Cellular Biology
NO	Nitric oxide
NRT	No reverse transcriptase
NTC	No-template control
NTDs	Neural tube defects
OCM	One-carbon metabolism
ORF	Open reading frame
P/S	Penicillin-streptomycin
$P54^{NRB}$/NONO	54 kDa nuclear RNA- and DNA-binding protein/Non-POU-domain-containing octamer binding protein
PAM	Protospacer adjacent motif
PBS	Dulbecco's phosphate-buffered saline
PCFT	Proton-coupled folate transporter
PCR	Polymerase chain reaction
PITCh	Precise integration into target chromosomes

POLG	DNA polymerase subunit gamma
POLR2A	Polymerase II, RNA, subunit A
PPI	Protein-protein interaction
PRM	Parallel reaction monitoring
PSF/SFPQ	Proline-glutamine rich splicing factor
PSPC1	Paraspeckle component 1
PVDF	Polyvinylidene difluoride
qPCR	Quantitative PCR
Rb	Retinoblastoma
RCN1	Reticulocalbin 1
RFC	Reduced folate carrier
Ribo-seq	Ribosome profiling
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RPFdb v2.0 beta	Ribosome-protected mRNA fragments database version 2 beta
RPFs	Ribosome protected mRNA fragments
rRNA	Ribosomal RNA
RT	Reverse transcription
RT-PCR	Reverse transcription PCR
RT-qPCR	Reverse transcription quantitative PCR
S.O.C.	Super optimal broth
SAH	S-adenosylhomocysteine
SAM	S-adenosylmethionine
SARDH	Sarcosine dehydrogenase
SDS	Sodium dodecyl sulfate

SEDI	Surface-enhanced laser desorption/ ionisation
SFI	Science Foundation Ireland
SHMT1	Serine hydroxymethyltransferase 1
SHMT2	Serine hydroxymethyltransferase 2
SHMT2α	Serine hydroxymethyltransferase 2α
SINEs	Short interspersed nuclear elements
SNP	Single-nucleotide polymorphism
SRM	Selected reaction monitoring
ssRNA	Single stranded RNA
SUMO	Small ubiquitin-like modifier
SUPT6H	SPT6 homolog, histone chaperone and transcription elongation factor
TALEs	Transcription-activator-like effectors
TBE	Tris-borate-EDTA
TBST	Tris buffered saline with tween
TDA	Targeted data acquisition
tGFP	Turbo green fluorescent protein
THF	Tetrahydrofolate
TOM20	Translocase of outer mitochondrial membrane 20
Tracr	Trans-activating CRISPR
tRNA	Transfer RNA
TUBA1A	Tubulin alpha 1A
TYMS	Thymidylate synthetase
UCL	University College London
UTR	Untranslated region
UV	Ultraviolet

VWA8	Von Willebrand factor A domain containing 8
XRCC4	X-ray repair cross complementing 4
ZFNs	Zinc fingers nucleases

Units

%	Percentage
°C	Degrees Celsius
Å	Angstrom
bp	Base pair
cm	Centimetre
Da	Daltons
g	Gram
Hour	Hours
kb	Kilobase
kDa	Kilodaltons
kV	Kilovolts
L	Litre
M	Molar
m/z	Mass-to-charge ratio
mg	Milligram
Min	Minutes
mL	Millilitre
mm	Millimetres
mM	Millimolar
ms	Millisecond
ng	Nanogram
nL	Nanolitre
nm	Nanometres
nM	Nanomolar

p-value	Probability value
pg	Picogram
ppm	Parts per million
RPKM	Reads per kilobase per million mapped reads
rpm	Rotations per minute
Sec	Seconds
U	Units
V	Volts
v/v	Volume per volume
w/v	Weight per volume
x g	Times gravity
µg	Microgram
µg/d	Microgram per day
µL	Microlitre
µm	micrometre
µM	Micromolar

List of figures

Figure 1.1 Chemical Structure of folic acid and folate. Both molecules are composed of three large sub-components: pteridine ring, para-amino-benzoic acid and glutamate....4

Figure 1.2 Absorption of folate and folic acid from the small intestine....6

Figure 1.3 Compartmentalisation of folate one-carbon metabolism pathway: cytoplasm, nucleus and mitochondria....10

Figure 1.4 The chromosomal location of the genes that make up the dihydrofolate reductase gene family....14

Figure 1.5 Mechanism of action by the DHFR enzyme....15

Figure 1.6 Crystal Structure of the DHFR Protein from Swiss Model....16

Figure 1.7 The model for DHFR translational up-regulation....18

Figure 1.8 Schematic diagram of the NO pathway that occurs in endothelial cells....21

Figure 1.9 The 6 DHFR2 RNA isoforms annotated by Ensembl....22

Figure 1.10 Sequence alignment comparing the amino acid sequence of the human DHFR protein versus the predicted DHFR2 protein, using Snapgene....24

Figure 3.1 Schematic showing the role of the ribosome during translation of an mRNA molecule into a protein....54

Figure 3.2 Ribosome profiling schematic....56

Figure 3.3 A schematic diagram of a sucrose cushion ultracentrifugation coupled with RT-qPCR....58

Figure 3.4 Ribo-seq results for the *GAPDH* and *POLG* genes, across multiple cells and tissues....66

Figure 3.5 Ribo-seq results for the *MTHFD1L, TYMS, DHFR* and *DHFR2* genes, across multiple cells and tissues....68

Figure 3.6 Ribo-seq results for the 5'UTR, CDS and 3'UTR regions of the *DHFR2* gene, across multiple cells and tissues....69

Figure 3.7 RNA integrity test on various RNA samples....71

Figure 3.8 Genomic contamination check on the newly synthesised cDNA from the Coriell, HepG2, HEK293 and undifferentiated human iPS samples....72

Figure 3.9 Illustration of where the DHFR2-202 PCR assay primers bind to DHFR2_202 RNA isoform....73

Figure 3.10 DHFR2_202 Exon 1 assay PCR product (10 µL) visualised on a 1% (w/v) agarose gel stained with ethidium bromide....74

Figure 3.11 DHFR2_202 Exon 1-2 assay PCR product (10 µL) visualised on a 1% (w/v) agarose gel stained with ethidium bromide....75

Figure 3.12 RNA integrity check on whole lysate control, supernatant and pellet sucrose cushion ultracentrifugation RNA samples....76

Figure 3.13 Genomic contamination check on the newly synthesised cDNA from the whole lysate control, supernatant and pellet sucrose cushion ultracentrifugation samples....77

Figure 3.14 Sucrose cushion ultracentrifugation RT-qPCR results on *GAPDH, NEAT1, DHFR* and *DHFR2* genes....79

Figure 4.1 Recombinant protein expression using a mammalian expression system 87

Figure 4.2 Generation of humanised DHFR2 mouse model using genome editing CRISPR-Cas9 gene-editing technology 91

Figure 4.3 Western Blot analysis of recombinant tGFP and DHFR2-tGFP using an anti-tGFP antibody 94

Figure 4.4 Western Blot analysis of recombinant tGFP-DDK-Myc and DHFR2-DDK-Myc using an anti-DDK antibody 95

Figure 4.5 Three-dimensional distribution plots from the PRM LC-MS/MS analysis on the HepG2 DHFR2 knockout cells transfected with the pCMV6-AC-DHFR2-tGFP and pCMV6-DHFR2-DDK-Myc vectors 98

Figure 5.1 Schematic of Shotgun LC-MS 108

Figure 5.2 Schematic of the TDA LC-MS/MS technique 110

Figure 5.3 Sequence alignment comparing the amino acid sequence of the human DHFR protein and the predicted DHFR2 protein showing the possible tryptic peptides unique to the DHFR2 protein 121

Figure 5.4 Example of the three-dimensional distribution plots generated from the PRM LC-MS/MS analysis 122

Figure 5.5 Western Blot analysis of DMSO treated control sample and MG132 treated sample using an anti-ubiquitin antibody. 124

Figure 5.6 Chromatogram results of the DHFR2 peptide LLPEYPGVLSDVQEGK in (A) whole protein lysate containing the recombinant DHFR2 protein, (B) Carnegie stage 21-22 heart tissue (14755), (C) post-conceptual week 9-10 heart tissue (14326) and (D) post-conceptual week 15-17 heart tissue (14559). 130

Figure 6.1 The structure of a paraspeckle 139

Figure 6.2 Schematic of 'normal' mRNA exportation out of the nucleus vs. RNA nuclear retention via paraspeckles 140

Figure 6.3 The location of the three Alu elements (AluSp_Alu, AluJb_Alu and FLAM_A_Alu) within the 3'UTR of the DHFR2_201 and DHFR2_202 transcripts according to REDIportal V2.0 145

Figure 6.4 Exon 2 of the DHFR2_201 and DHFR2_202 transcripts showing the location of the ORF, three Alu elements in the 3'UTR and the eight A to I RNA edits listed in Table 6.1.. 146

Figure 6.5 A to I gDNA reference PCR assay PCR product (7 µL) visualised on a 1% (w/v) agarose gel stained with SYBR Safe 147

Figure 6.6 RNA integrity test on the extracted HepG2 RNA. RNA (500 ng) was resolved on a 1% (w/v) agarose gel stained with SYBR Safe 148

Figure 6.7 Genomic contamination check on the newly synthesised cDNA from HepG2 RNA 148

Figure 6.8 A to I cDNA PCR assays PCR product (7 µL) visualised on a 1% (w/v) agarose gel stained with SYBR Safe 149

Figure 6.9 Sections of the Sanger Sequencing chromatograms showing the sites where the adenosine (A) nucleotide has been A to I edited 151

Figure 7.1 Schematic representation of the CRISPR/Cas9 and non-homologous end joining system to disrupt a gene within the genome of cultured cells 157

Figure 7.2 Schematic diagram showing the different mechanisms cells use to generate endogenous protein and recombinant protein 158

Figure 7.3 Compound abundance graphs showing the normalised abundance of (A) the ALDH1L1, (B) the DHFR, (C) the SHMT1 and (D) TYMS proteins within the four replicates for the HepG2 wildtype and the HepG2 DHFR2 knockout cell lines 165

Figure 7.4 Compound abundance graphs showing the normalised abundance of (A) the ALDH1L1, (B) the DHFR, (C) the SARDH and (D) the SHMT1 proteins within the four replicates for the HepG2 wildtype and the HepG2 DHFR knockdown cell lines 168

Figure 7.5 Compound abundance graphs showing the normalised abundance of (A) the recombinant DHFR2, (B) the DHFR and (C) the SARDH proteins within the four replicates for the HepG2 cells transfected with the pCMV6-AC-tGFP vector and HepG2 transfected with the pCMV6-AC-DHFR2-tGFP vector 172

Figure 8.1 Proposed DHFR and DHFR2 regulatory mechanism 188

List of tables

Table 2.1 Breakdown of each of the human embryonic/foetal tissues 34

Table 2.2 Primers used for genomic contamination check 39

Table 2.3 Primers used for DHFR2_202 RT-PCR assays 40

Table 2.4 Primers used for A to I DHFR2 assays 41

Table 2.5 Primers and probes used in RT-qPCR assay on ribosome bound and free HepG2 fraction 45

Table 2.6 The amino acid sequence of the DHFR and DHFR2 tryptic peptide 49

Table 2.7 DHFR2 peptide information used for the PRM LC-MS/MS workflow 50

Table 2.8 Antibody concentrations used for Western Blot analysis 51

Table 4.1 Dihydrofolate reductase peptides detected in the DDA LC-MS/MS analysis from HepG2 DHFR2 knockout cells transfected with DHFR2-tGFP Vector and DHFR2-DDK-Myc Vector 96

Table 4.2 Dihydrofolate reductase peptides detected in the PRM LC-MS/MS analysis from HepG2 DHFR2 knockout cells transfected with DHFR2-tGFP vector and DHFR2-DDK-Myc vector 97

Table 4.3 Dihydrofolate reductase peptides detected in the DDA LC-MS/MS analysis from wildtype and humanised DHFR2 embryo mouse tissue 100

Table 4.4 Dihydrofolate reductase peptides detected in the DDA LC-MS/MS analysis from wildtype postnatal mouse and humanised DHFR2 postnatal mouse tissue 100

Table 5.1 DHFR2 specific peptides included in the Skyline MRM DHFR2 LC-MS/MS workflow 116

Table 5.2 Peptides included in the Skyline MRM DHFR2 LC-MS/MS workflow 118

Table 5.3 Dihydrofolate reductase peptides detected in the DDA LC-MS/MS analysis on the adult cell lines and tissues 120

Table 5.4 Dihydrofolate reductase peptides detected in the DDA LC-MS/MS analysis on DMSO and MG132 treated cells 125

Table 5.5 DHFR2 specific peptides detected in whole protein lysate containing the recombinant DHFR2 protein 126

Table 5.6 Peptides identified in the whole protein lysate containing the recombinant DHFR2 protein 127

Table 5.7 Peptides identified in the Carnegie stage 21-22 (embryo) tissues 128

Table 5.8 Peptides identified in the post-conceptual week 9-10 (foetus) tissue 128

Table 5.9 Peptides identified in the post-conceptual week 15-17 (foetus) tissue 129

Table 6.1 High-frequency A to I RNA edits identified in the AluJb element by REDIportal V2.0 145

Table 6.2 The A to I cDNA PCR assay plasmids Sanger Sequencing results 150

Table 7.1 Impact of knocking out the human *DHFR2* gene in the HepG2 cells on the expression of OCM proteins 164

Table 7.2 Impact of down-regulating the human *DHFR* gene in the HepG2 cells on the expression of OCM proteins 167

Table 7.3 Impact of overexpressing the human *DHFR2* gene in the HepG2 cells on the expression of OCM proteins 171

Table 7.4 Proteins found to have consistent effects in the HepG2 knockout and overexpression cell line 174

Table 7.5 Proteins found to have opposing effects in the HepG2 knockout and overexpression cell line 175

Abstract

Unravelling the role of the human Dihydrofolate reductase 2 gene

Niamh Bookey

The dihydrofolate reductase gene family consists of two functional genes. The *DHFR* gene is the more researched gene of this family. It encodes an enzyme that reduces dihydrofolate to tetrahydrofolate in a cell. This reduction is an important step in the one-carbon metabolism pathway, which is responsible for purine/thymidylate biosynthesis and homocysteine re-methylation. The remaining functional gene is known as the *DHFR2* gene. There was very little knowledge about this retrogene prior to this project. Given the importance of the *DHFR* gene, it was vital the function of the *DHFR2* gene was elucidated. Therefore this thesis aimed to characterise it by focusing on the expression and behaviours of its RNA and protein. Firstly, the translatability of the DHFR2 transcripts was assessed by performing a sucrose cushion ultracentrifugation RT-qPCR experiment. The two main DHFR2 transcripts showed differential translational capabilities, with one of them potentially encoding a protein. Secondly, the expression of the endogenous DHFR2 protein was investigated using LC-MS/MS. There was no evidence of the native DHFR2 protein in the human cells/tissues examined, apart from heart tissue from early development. This result suggests that the *DHFR2* gene encodes a protein to supply dihydrofolate reductase activity to rapidly growing embryos and foetuses. Thirdly, mechanisms of translational regulation were considered for the DHFR2 transcripts as the native protein had not been detected in all the cells/tissues examined. There was evidence that some of the DHFR2 transcripts were being translationally inhibited via paraspeckle nuclear retention. Finally, an additional role for the *DHFR2* gene was investigated as the DHFR2 RNA was found to be highly abundant in multiple cells/tissues despite the protein being undetectable in the same samples. Mass spectrometry-based differential analysis on a DHFR2 knockout and overexpression model identified that the DHFR2 RNA has a role in regulating the expression of the *DHFR* gene. To conclude, the results gathered in this thesis revealed the biological roles of the human *DHFR2* gene. These significant findings will pave the way for future research on this functional gene which may have a major impact on human health and disease.

Chapter 1. Introduction

1.1. Overview

One-carbon metabolism (OCM) is an interconnected network of biochemical pathways, including those required for de novo purine biosynthesis, de novo thymidylate biosynthesis, homocysteine re-methylation and methionyl-tRNA synthesis (Fox & Stover, 2008). This complex pathway has been extensively studied since the discovery of folic acid in the 1940s, due to its importance in human health and disease. Folate deficiency and polymorphisms within some of the OCM genes have been linked with a broad range of health issues and clinical diseases. These include neural tube defects (NTDs), cardiovascular disease (CVD), stroke, megaloblastic anaemia, cancer and metabolic disorders (Bailey et al., 2015).

The human dihydrofolate reductase gene family is one of the key gene families within the OCM pathway. It consists of two functional genes and five pseudogenes that are dispersed across the genome (Anagnou et al., 1984, 1988; Anderson et al., 2011; Funanage et al., 1984; Maurer et al., 1984; McEntee et al., 2011; Shimada et al., 1984). The dihydrofolate reductase (*DHFR*) gene is the most researched gene in this family. It encodes an enzyme that carries out the dihydrofolate reductase activity in the cytoplasm, nucleus and mitochondria of a cell (Anderson et al., 2011, 2012; Appling, 1991; Woeller et al., 2007). This enzyme reduces dihydrofolate (DHF) to its biologically active form tetrahydrofolate (THF), using nicotinamide adenine dinucleotide phosphate (NADPH) as an electron donor. This reduction is an extremely important step in OCM; thus being essential for normal cell function (Fox & Stover, 2008). Therefore, the *DHFR* gene has become an extremely valuable drug target for the treatment of multiple types of cancer and some inflammatory diseases (Bailey, 1995; Cronstein, 1996). The remaining functional gene is a retrogene and is known as the dihydrofolate reductase 2 (*DHFR2*) gene (McEntee et al., 2011). To date, very little is known about this gene. However, given the importance of the *DHFR* gene to human health, it is vital the overall function of the *DHFR2* gene is elucidated. This project aims to characterise this gene.

This first chapter begins with a general overview of folate, and the OCM pathway, followed by its role in human health and disease. This is followed by an overview of the *DHFR* gene family with particular emphasis on the *DHFR2* gene. This chapter concludes with the aims and objectives.

1.2. Vitamin B_9

Vitamin B_9 is an essential water-soluble vitamin that provides one-carbons for OCM, a complex network of interconnected folate-dependent metabolic pathways that is responsible for nucleic acid biosynthesis and methylation processes (McNulty et al., 2020). Therefore, this vitamin is

essential for normal cell function. A deficiency in vitamin B_9 results in suboptimal folate levels in the body which can lead to the development of many clinical diseases, including anaemia, CVD, cancer and NTDs. Optimal folate status is effective at preventing or treating some of these pathologies. Therefore, this vitamin is important for maintaining an individual's health and is crucial for the early stages of embryonic development (Ami et al., 2016; Daly et al., 1995; Kim et al., 2015; Scaglione & Panzavolta, 2014; Scott et al., 1994; Warzyszynska & Kim, 2014).

1.2.1. History of folate and folic acid

Folate is the naturally occurring version of vitamin B_9 (Naderi & House, 2018). It was initially discovered by Lucy Wills in 1931 when investigating the nutritional deficiencies that caused anaemia in pregnant women (Wills, 1931). Over a decade later the term folate was conceived when the molecule was isolated from a spinach leaf (*folium* meaning leaf in Latin) and used as a growth factor for *Streptococcus faecalis* (Mitchell et al., 1941). Ever since this discovery, folate has been extensively studied in relation to human health and disease (Rosenberg, 2012).

Folic acid, also called pteroylglutamic acid, is the synthetic form of vitamin B_9. It was first synthesised in 1943 by Robert Angier from Lederle Laboratories in the United States (Angier et al., 1945; Rosenberg, 2012). In the subsequent years following this breakthrough, folic acid was employed to treat megaloblastic anaemia and many other clinical diseases (Rosenberg, 2012).

1.2.2. Folate and folic acid chemical structures

Folate and folic acid have very similar chemical structures. They are composed of three distinct sub-components: pteridine ring, para-amino-benzoic acid and glutamate (Figure 1.1). The pteridine ring is bound to para-amino-benzoic acid via a methylene bridge. The para-amino-benzoic acid is subsequently linked to a glutamate unit by a peptide linkage (Zempleni et al., 2013). A key difference between folate and folic acid is the degree to which the pteridine ring is oxidised. Folic acid is fully oxidized and does not contain a carbon substitution. Whereas, folate is the term usually used to refer to the reduced forms and may contain a carbon moiety at N5 and/or N10 atoms (Chan et al., 2013; Tam et al., 2012). A second key difference between folate and folic acid is the number of glutamates in their tails. Folic acid always exists as a monoglutamate, meaning it only has 1 glutamate. Whereas, folate exists as a polyglutamate, meaning there are several glutamates on its tail (Melse-Boonstra et al., 2002). The longer glutamate tail on the folate molecules makes it unstable and much more susceptible to damage by heat, oxidation and ultraviolet light (Bassett &

Sammán, 2010). The shorter glutamate tail on the folic acid molecule not only makes it more stable but also allows for more efficient transportation across cell membranes and absorption into the blood. For this reason, folic acid has a higher bioavailability in the human body than folate (Cieślik & Cieślik, 2018).

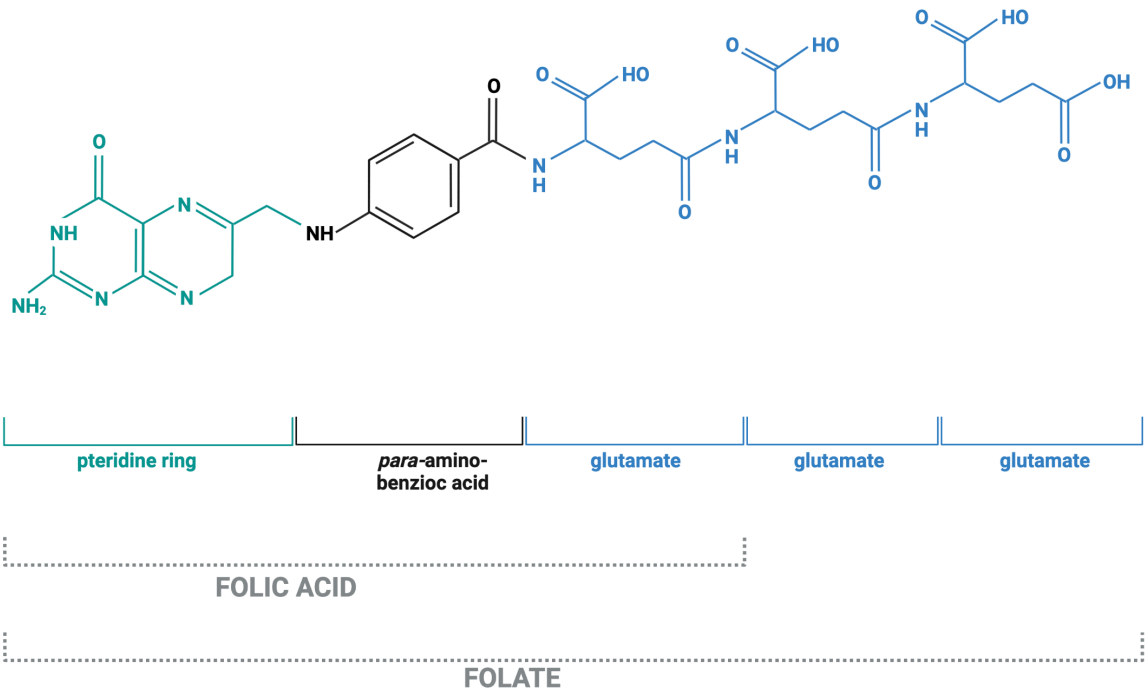


Figure 1.1 Chemical Structure of folic acid and folate. *Both molecules are composed of three large sub-components: pteridine ring, para-amino-benzoic acid and glutamate. One key difference between folate and folic acid is the number of glutamates in their tails. Folic acid always exists as a monoglutamate, meaning it only has 1 glutamate. Whereas, folate exists as a polyglutamate, meaning there is more than 1 glutamate on its tail. Created with BioRender.com, Adapted from Zempleni et al., 2013.*

1.2.3. Absorption of folate and folic acid

Mammalian cells are capable of synthesising the three sub-components of vitamin B_9 but are unable to combine the pteridine ring with the para-amino-benzoic acid (Stanger, 2002). Therefore, this essential vitamin must be obtained through one's diet. The World Health Organisation advises that the Recommended Daily Intake for a healthy adult is 400 µg/d and 600 µg/d for women who are planning a pregnancy or during pregnancy (Food and Nutrition Board, 1998; Molloy et al., 2017; World Health Organization, 2005, 2015). This recommendation can be achieved by consuming foods that contain naturally occurring folates such as green leafy vegetables, citrus fruits, beans or cereals. Or by ingesting the synthetic form of vitamin B_9, which comes in the form of folic acid supplements or folic acid fortified foods (Greenberg et al., 2011).

Once vitamin B_9 is consumed it travels down the digestive tract until it reaches the intestines where it is absorbed into the bloodstream (Figure 1.2) (Karthika & Pati, 2019). For this process to occur vitamin B_9 has to exist as its monoglutamate form. Therefore, folic acid can be immediately absorbed into the proximal small intestine. However, as natural folate exists as a polyglutamate form it has to be shortened to only have one glutamate unit. This reaction is carried out by a folate conjugase known as glutamate carboxypeptidase II (GCPII), which is found in the apical brush border of the small intestine (Naderi & House, 2018; Visentin et al., 2014). The monoglutamate molecules are subsequently transported into the enterocyte of the small intestine via the high-affinity proton-coupled folate transporter (PCFT) (Said, 2011). In this compartment, the monoglutamate molecules are converted to 5-methyl-THF (5-CH_3-THF). This is a multi-step process that requires the DHFR and methylenetetrahydrofolate reductase (MTHFR) enzymes. The 5-CH_3-THF molecules are delivered to the hepatic portal vein which transports them to the liver (Blom & Smulders, 2011). The liver ultimately distributes the 5-CH_3-THF molecules to all the cells in the body via the systemic bloodstream. The 5-CH_3-THF molecule enters a cell through either the family of folate receptors (FR) or the reduced folate carrier (RFC) (Visentin et al., 2014). It is then converted into the polyglutamate form via the foly-polyglutamate synthase (FPGS) enzyme. The addition of the glutamate units to the 5-CH_3-THF molecule prevents it from being subject to transporter-mediated efflux. This version of folate is the main form of intracellular folate (Naderi & House, 2018; Nazki et al., 2014). The DHFR enzyme reduces the polyglutamate folate to THF. The coenzyme THF enters the OCM cycle to support several metabolic reactions (Fox & Stover, 2008).

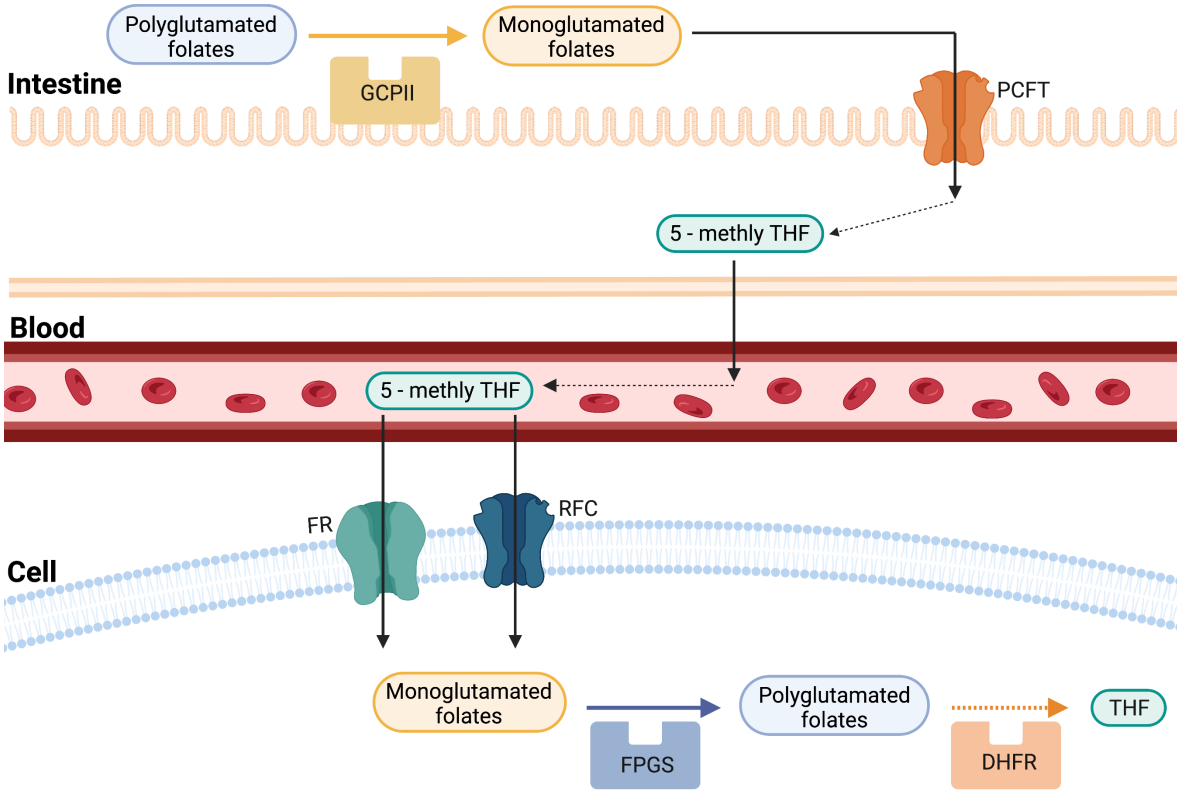


Figure 1.2 Absorption of folate and folic acid from the small intestine. *Folate and folic acid arrive in the intestinal tract. Folic acid is absorbed immediately whereas folate has to be hydrolysed to a monoglutamate form via the GCPII enzyme. The monoglutamate molecules are transported into the small intestine through the PCFT and converted to 5-methyl-THF with the help of the DHFR and MTHFR enzymes. The 5-methyl-THF molecule enters the bloodstream where it travels around the body. It enters a cell through the FR or the RFC, where It is converted into the polyglutamate form via the FPGS enzyme. The DHFR enzyme reduces the polyglutamate folate to THF which enters the OCM cycle to support several metabolic reactions. Created with BioRender.com, Adapted from Blom & Smulders, 2011.*

1.3. One-carbon metabolism

One-carbon metabolism is a complex network of interconnected folate-dependent metabolic pathways (Stover, 2009). Folate supplies activated one-carbon units to support each of the metabolic reactions in these pathways. The coenzyme THF is the most versatile derivative of folate. It is able to carry activated one-carbon units at multiple different oxidation states (Tibbetts & Appling, 2010). THF can exist as 10-formyl-THF (10-CHO-THF), 5-methyl-THF (5-CH_3-THF), 5,10-methenyl-THF (5,10-CH^+-THF), 5,10-methylene-THF (5,10-CH_2-THF). All these different forms of THF with a bound carbon are referred to as the one-carbon pool (Tibbetts & Appling, 2010; Tjong et al., 2022). The one-carbon pool supplies one-carbon units for de novo purine biosynthesis, de novo

thymidylate biosynthesis, homocysteine re-methylation and methionyl-tRNA (MetRNA) synthesis (Fox & Stover, 2008).

In eukaryotes, OCM reactions occur in three separate compartments of the cell i.e., the cytoplasm, mitochondria and nucleus (Figure 1.3). Although OCM is compartmentalised, each organelle is connected through the unidirectional movement of one-carbon donors (formate, serine and glycine) across intracellular membranes (Tibbetts & Appling, 2010). For example, the majority of one-carbon units that are required for cytoplasmic reactions are derived from mitochondrial formate (Tibbetts and Appling, 2010).

1.3.1. One-carbon metabolism in the mitochondria

OCM in the mitochondria is responsible for generating one-carbon units in the form of folate for cytoplasmic OCM, producing glycine, synthesising formylated MetRNA for protein synthesis and de novo thymidylate biosynthesis (Figure 1.3) (Stover, 2009).

The one-carbon flux in the mitochondria proceeds in the oxidation direction. THF is transported into the mitochondria via the mitochondrial folate transporter (MFT). However, the amount of THF transported into the mitochondria is not sufficient to supply one carbons to all the metabolic reactions in this organelle (Tibbetts & Appling, 2010). Therefore, one-carbon donors (serine, glycine and formate) are transported from the cytoplasm to fuel mitochondrial OCM. THF is converted to 10-CHO-THF with several folate intermediates in between; THF and serine are converted to 5,10-CH_2-THF and glycine by the serine hydroxymethyltransferase 2 (SHMT2) enzyme, 5,10-CH_2-THF is oxidised to 5,10-CH^+-THF via the methylenetetrahydrofolate dehydrogenase, cyclohydrolase and formyltetrahydrofolate synthetase 2L (MTHFD2L) enzyme and 5,10-CH^+-THF is catalysed to 10-CHO-THF by the MTHFD2 enzyme (Fox & Stover, 2008). The resulting 10-CHO-THF can be hydrolysed to formate by the MTHFD1L enzyme to be exported to the cytoplasm. Alternatively, 10-CHO-THF can be employed to synthesise MetRNA for mitochondrial protein synthesis (Stover, 2009).

De novo thymidylate biosynthesis has also been identified in the mammalian mitochondria. The 5,10-CH_2-THF folate coenzyme that was converted from THF by the SHTM2 enzyme for previously mentioned metabolic reactions can be alternatively utilised for de novo thymidylate biosynthesis. 5,10-CH_2-THF and deoxyuridine monophosphate (dUMP) are converted to deoxythymidine monophosphate (dTMP) and DHF by the thymidylate synthetase (TYMS) enzyme. dTMP is an essential precursor for DNA biosynthesis and has been shown to prevent uracil accumulation in mitochondrial DNA. DHF is converted back to THF by the DHFR enzyme in an NADPH-dependent

manner. THF can be utilised for any of the mitochondrial OCM reactions described in this section (Anderson et al., 2011).

1.3.2. One-carbon metabolism in the cytoplasm

OCM can be split into three different pathways in the cytoplasm: de novo purine biosynthesis pathway, de novo thymidylate biosynthesis pathway and homocysteine re-methylation pathway. These three pathways are interlocked and compete with each other for one-carbon units (Figure 1.3) (Fox & Stover, 2008).

The one-carbon flux in the mitochondria proceeds towards reduction in the cytoplasm (as opposed to the mitochondria). Mitochondrial formate enters the cytoplasm where it is converted to 10-CHO-THF in an ATP-dependent reaction catalysed by the MTHFD1 enzyme. 10-CHO-THF serves as a donor of C2 and C8 carbons for purine biosynthesis. 10-CHO-THF-dependent purine biosynthesis only happens if the purine nucleotide salvage pathway does not generate an adequate amount of purine nucleotides (Stover, 2009). The purines generated from both the de novo and salvage pathway incorporate into DNA and RNA molecules. Furthermore, purine metabolites provide the cell with energy and cofactors to boost cellular proliferation and cell survival (Alur, 1999).

The activated one-carbon units belonging to 10-CHO-THF can also be employed to initiate the de novo thymidylate biosynthesis pathway. The MTHFD1 enzyme reduces 10-CHO-THF to 5,10-CH_2-THF in the presence of NADPH. 5,10-CH_2-THF and dUMP are subsequently converted to dTMP and DHF by the TYMS enzyme. dTMP is an essential precursor for DNA biosynthesis and therefore, vital for normal cell function. DHF is converted to THF by the DHFR enzyme in an NADPH-dependent manner. The SHMT 1/2α enzymes catalyse the conversion of THF back to 5,10-CH_2-THF. 5,10-CH_2-THF can be utilised for subsequent cycles of de novo thymidylate synthesis (Scotti et al., 2013; Stover, 2009; Tibbetts & Appling, 2010).

Alternatively, 5,10-CH_2-THF can be further reduced to 5-CH_3-THF by the MTHFR enzyme using NADPH as an electron donor. 5-CH_3-THF can subsequently supply one-carbon units for the remethylation of homocysteine to methionine. This reaction is carried out by the B_{12}-dependent methionine synthase (MTR) enzyme (Obeid, 2013; Stover, 2009). Methionine is the universal amino acid employed to initiate the translation of mRNA into a protein. Methionine can also be converted into S-adenosylmethionine (SAM) by the enzyme methionine adenosyltransferase (MAT). SAM acts as a high energy methyl donor in the cell and therefore is involved in the regulation of multiple cellular processes (Ouyang et al., 2020; Scotti et al., 2013). Following the donation of the methyl

group, S-adenosylhomocysteine (SAH) is generated. SAH is hydrolysed to adenosine and homocysteine by S-adenosylhomocysteine hydrolase (AHCY). The successful generation of homocysteine closes the methionine cycle (Froese et al., 2019).

1.3.3. One-carbon metabolism in the nucleus

In the nucleus, folates are employed to facilitate de novo thymidylate biosynthesis (Stover & Field, 2011). The nuclear de novo thymidylate biosynthesis pathway is similar to the cytoplasmic version of this pathway (Figure 1.3) (Tibbetts & Appling, 2010). Cytoplasmic THF enters the nucleus where it is converted to 5,10-CH_2-THF by the MTHFD1 enzyme (Field et al., 2014, 2016). The TYMS enzyme subsequently employs 5,10-CH_2-THF and dUMP to generate dTMP, yielding DHF as a secondary product (Stover, 2009). The dTMP nucleotides are employed to maintain DNA replication and aid with DNA repair in the nucleus. The DHFR enzyme reduces DHF to THF in the presence of NADPH and the SHMT1/2α enzymes convert THF back to 5,10-CH_2-THF using serine as the one-carbon donor. 5,10-CH_2-THF can be used to restart the de novo thymidylate biosynthesis cycle (Stover & Field, 2011).

The enzymes required for nuclear de novo thymidylate biosynthesis do not naturally exist in this compartment. Therefore, during the S phase of the cell cycle, the DHFR, TYMS, SHMT1/2α and MTHFD1 enzymes located in the cytoplasm are subjected to posttranslational modifications (Anderson et al., 2012; Fox et al., 2009; Spizzichino et al., 2021). This modification is mediated by the ubiquitin-conjugating enzyme 9 with the small ubiquitin-like modifier (SUMO) (Anderson et al., 2007; Woeller et al., 2007). The modifications made to the enzymes act as a signal for nuclear importation. Once all the enzymes are imported into the nucleus, DHFR, TYMS and SHMT1/2α form a dTMP synthesis complex. The SHMT1/2α section of this complex is anchored into the nuclear lamina at sites of replication (Murthy & Reddy, 2006). The multienzyme complex and MTHFD1 enzymes are now ready to facilitate de novo thymidylate biosynthesis for nuclear DNA replication and repair (Anderson et al., 2012).

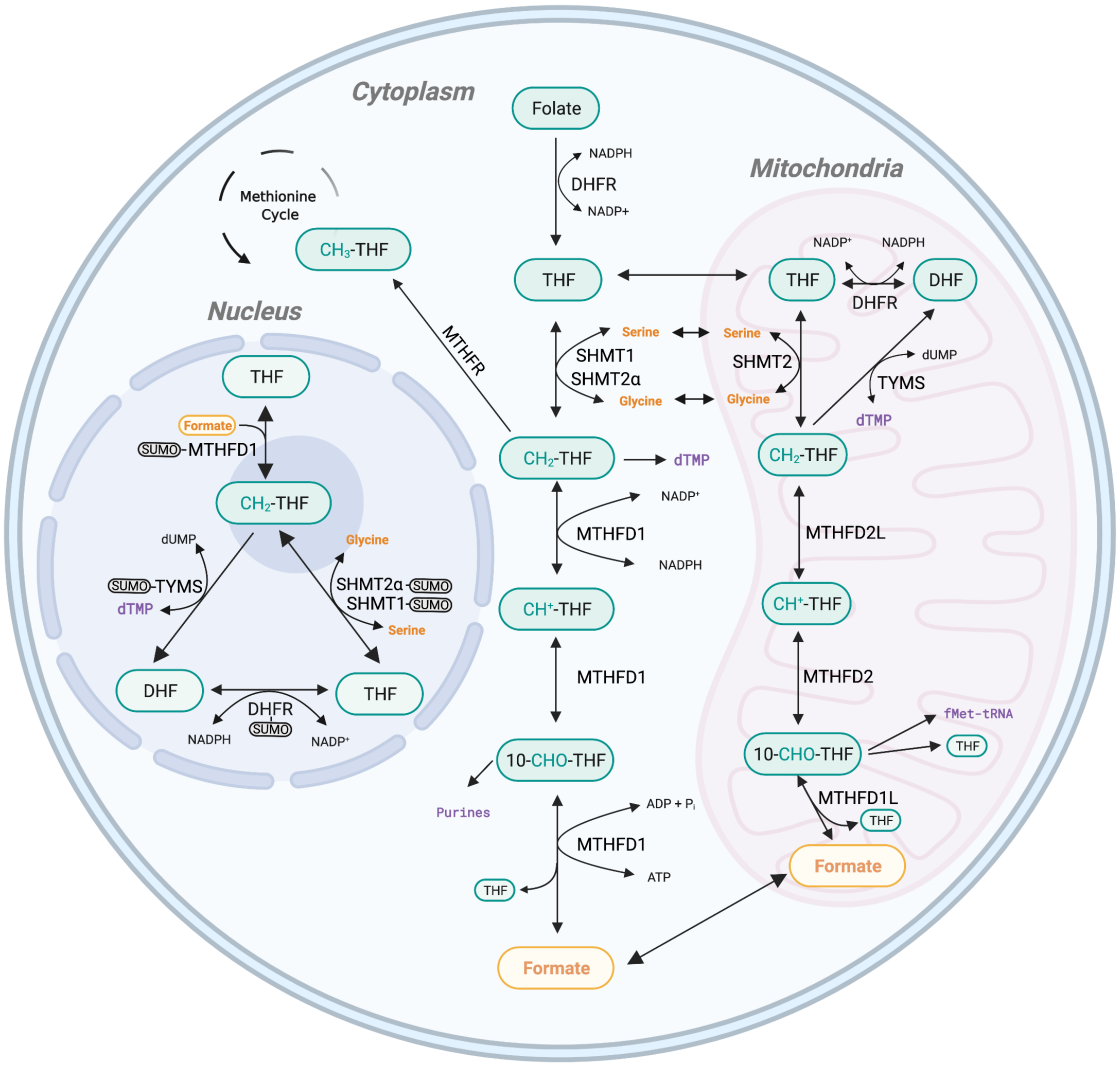


Figure 1.3 Compartmentalisation of folate one-carbon metabolism pathway: cytoplasm, nucleus and mitochondria. *The OCM pathway in the cytoplasm is responsible for de novo purine biosynthesis, de novo thymidylate biosynthesis and homocysteine re-methylation. The OCM pathway in the mitochondria is responsible for generating one-carbon units in the form of folate for cytoplasmic OCM, producing glycine, synthesising formylated MetRNA for protein synthesis and de novo thymidylate biosynthesis. The OCM pathway in the nucleus is responsible for de novo thymidylate biosynthesis. The metabolic reactions that generate these end products (purple) require one-carbon units. THF and some of its derivatives gain one-carbon units from one-carbon donors (orange) such as serine, glycine and formate. Therefore, the activated one-carbon units (green) bound to THF carry and supply the one-carbon units to the metabolic reactions that occur in the OCM pathway. OCM enzymes (black) are required throughout the whole pathway to catalyse all these reactions. Created with BioRender.com, Adapted from Tibbetts & Appling, 2010.*

1.4. One-carbon metabolism in health and disease

Impairments within the OCM pathway can have serious consequences on an individual's health and the risk of developing a disease. Folate deficiency has been linked with a broad range of health issues and clinical diseases, these include NTDs, stroke, megaloblastic anaemia and cancer (Bailey et al., 2015).

1.4.1. One-carbon metabolism and neural tube defects

The need for one-carbon units is highest during embryonic development as the cells are rapidly proliferating. Therefore, suboptimal folate levels in the mother can result in NTDs (Beaudin & Stover, 2009). NTDs are a group of birth defects that arise from the failure of neurulation during the very early stages of human embryogenesis (Imbard et al., 2013). The two most common NTDs are spina bifida and anencephaly. The association between NTDs and folate deficiency was initially proposed by Richard Smithells in 1976 (Smithells et al., 1976). Subsequent research showed that supplementing pregnant women with a high dose of folic acid reduced the risk of new-borns having NTDs, by up to 70% (Ami et al., 2016). This is the reason why the world health organisation recommends a higher dosage of 600 µg/d for women who are planning a pregnancy or women that are pregnant (Food and Nutrition Board, 1998; Molloy et al., 2017; World Health Organization, 2005, 2015).

Furthermore, polymorphisms in certain OCM genes have been linked to an increased risk of NTDs. This includes the *MTHFD1, MTHFD1L, MTHFR, DHFR* and *DHFR2* genes (Greene & Copp, 2014; Johnson et al., 2004; Pangilinan et al., 2021; Parle-McDermott et al., 2007; Stover, 2011).

The 677 C>T single-nucleotide polymorphism (SNP) in the *MTHFR* gene is one of the most widely studied polymorphisms in relation to NTDs (Kirke et al., 2004; Shields et al., 1999; van der Put et al., 1995; Whitehead et al., 1995). Both the homozygous (TT) and heterozygous (CT) mutated genotypes have been found to have a significant impact on the MTHFR enzyme activity. The enzyme activity was only 50% of that of the CC genotype. The reduction in the MTHFR enzyme activity subsequently results in suboptimal blood folate concentrations and elevated homocysteine levels (Frosst et al., 1995). These conditions in early pregnancy can result in the development of NTDs and sometimes cause embryo lethality (Kirke et al., 2004; Shields et al., 1999; van der Put et al., 1995; Whitehead et al., 1995). The combined CT and TT genotypes account for over a quarter of NTDs in Ireland (Kirke et al., 2004).

The 19 bp deletion polymorphism in the promoter of the *DHFR* gene has also been extensively studied over the years due to its association with NTDs. Johnson et al. 2004 showed that the DHFR 19 bp deletion in mothers is associated with an increased risk of having a child with spina bifida (Johnson et al., 2004). Whereas, Parle-McDermott et al. 2007 observed that Irish mothers with the deletion have a lower risk of having offspring with NTDs (Parle-McDermott et al., 2007). Therefore, the scientific community remains undecided about the link between the DHFR 19 bp deletion and NTDs.

A relatively new genetic association study of an Irish cohort found an association between 3 SNPs within the *DHFR2* gene with an increased risk of developing NTDs. The rs17855824, rs10454213 and rs7645522 SNPs are located in different areas of the *DHFR2* gene and are in linkage disequilibrium with each other. This result did not replicate in the UK cohort but does not rule out a functional role in NTDs (Pangilinan et al., 2021)

Even though there is strong evidence that suggests that certain polymorphisms in OCM genes result in NTDs, the situation is not that simple. NTDs seem to be caused by a combination of multiple factors, including numerous genetic and environmental factors (Greene & Copp, 2014).

1.4.2. One-carbon metabolism and other diseases

1.4.2.1. One-carbon metabolism and links to stroke

Stroke is the second leading cause of death worldwide, accounting for over six million deaths in 2019 (World Health Organization, 2020). Suboptimal folate levels and elevated homocysteine levels are strongly associated with the probability of having a stroke (Burgess et al., 2020; Furie & Kelly, 2006). Unlike NTDs, reducing the elevated homocysteine levels via folic acid supplementation only had a modest effect on reducing the risk of a stroke. Therefore, scientists remain sceptical about folic acid supplementation as the only preventative measure for reducing the risk of a stroke (Hsu et al., 2018; Kaye et al., 2020; Lonn et al., 2006; Tian et al., 2017).

Polymorphisms within certain OCM genes have been found to have a greater association with the incidence of a stroke. A genome-wide association study found a variant within the 5' untranslated region (UTR) of the Aldehyde Dehydrogenase Family 1 Member L1 (*ALDH1L1)* gene was linked to an increased risk of ischemic stroke. This is the most common type of stroke, which occurs when the arteries to the brain become narrowed or blocked, causing severely reduced blood flow (Williams et al., 2014).

1.4.2.2. One-carbon metabolism and links to megaloblastic anaemia

Megaloblastic anaemia is a type of anaemia caused by the impairment of DNA synthesis in red blood cells. This is mainly caused by severe vitamin B_{12} and/or folate deficiency (Castellanos-Sinco et al., 2015). Extremely low levels of vitamin B_{12} and folate result in thymidylate deficiency. This leads to uracil being incorporated into DNA during replication in place of thymidine. This cellular event impairs the maturation rate of the nucleus of red blood cells. However, the cytoplasmic maturity remains unaffected. For this reason, the bone marrow produces unusually large, structurally abnormal, immature red blood cells known as megaloblasts. The majority of megaloblasts die before reaching maturity via p53/p21 mediated cellular apoptosis (Aslinia et al., 2006; Chakraborty & Stover, 2020; Koury et al., 2000). Megaloblastic anaemia is a very treatable disease through vitamin B_{12} and folic acid supplementation (Castellanos-Sinco et al., 2015).

1.4.2.3. One-carbon metabolism and links to cancer

There has long been an association between folate levels and the risk of developing certain types of cancers. As mentioned in Section 1.4.2.2, folate deficiency promotes massive uracil misincorporation during DNA replication. The cell tries to repair this mistake by removing the uracil base from the DNA strand. This produces deletions, double-strand breaks and chromosomal breaks in the DNA. The damage to the DNA caused by this repair mechanism may contribute to the increased risk of cancer. Low folate status has been linked to colorectum, oesophagus, stomach, pancreas, lungs, breast, cervix, ovary and breast cancers (Kim, 1999). The most significant association is found with colorectal cancer (Bailey et al., 2003; Giovannucci, 2002; Kim, 1999, 2003). However it is not that simple, folate status alone is not enough to increase the risk of developing cancer. It is often dependent on the severity and timing of folate deficiency and polymorphisms within the genes of the OCM pathway (Kim, 2003). Studies have shown MTHFR polymorphisms in the presence of low plasma folate levels have been linked to the colorectal cancer incidence (Kim et al., 2015).

Even though folic acid supplementation may help to reduce the chances of developing cancer in normal tissues, there are also drawbacks associated with folic acid supplementation. A high folate diet plus folic acid supplementation results in high levels of folate circulating in the blood. Increased levels of folate in the body have been found to promote the progression of established precancerous and cancer cells (Ebbing et al., 2009; Kim, 2004, 2006, 2008; Pieroth et al., 2018).

Therefore, the recommended daily intake of folate and folic acid must be followed to avoid extremely low and high levels in the body (Pieroth et al., 2018).

1.5. Dihydrofolate reductase gene family

The dihydrofolate reductase gene family is one of the key gene families within the OCM pathway. It consists of seven highly homologous genes that are dispersed across the human genome (Figure 1.4). This includes the functional *DHFR* and *DHFR2* genes located on chromosomes 5 and 3 respectively. Along with five intronless pseudogenes known as DHFRP1, DHFRP2, DHFRP3, DHFRP5 and DHFRP6 which are found on chromosomes 2, 6 and 18. (Anagnou et al., 1984, 1988; Funanage et al., 1984; Maurer et al., 1984; Shimada et al., 1984).

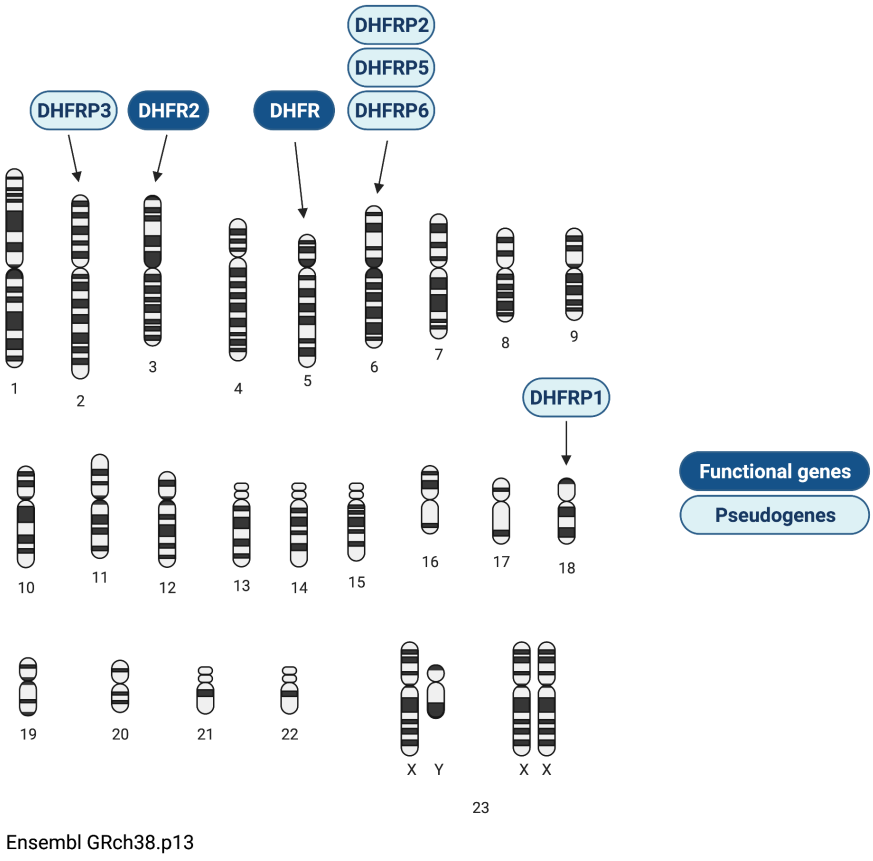


Figure 1.4 The chromosomal location of the genes that make up the dihydrofolate reductase gene family. *The functional DHFR gene and DHFR2 gene are found on chromosomes 5 and 3 respectively. The pseudogenes DHFRP1, DHFRP2, DHFRP3, DHFRP5 and DHFRP6 on chromosomes 2, 6 and 18. Created with BioRender.com.*

The *DHFR* gene was originally discovered in 1984 by Chen et al. and has subsequently become the most researched gene of this gene family (Chen et al., 1984). Nearly three decades later the second dihydrofolate reductase gene was identified by McEntee et al., 2011 and Anderson et al., 2011. The remaining five genes within the dihydrofolate reductase gene are thought to have accumulated numerous mutations over the years that have rendered them non-functional (Thomas, 2001).

1.5.1. Dihydrofolate reductase gene

The human *DHFR* gene is located on chromosome 5q11.2 and extends for approximately 30 kb downstream (Askari & Krajinovic, 2010). Ensembl has predicted that the human *DHFR* gene encodes seven RNA isoforms identified as DHFR_201 through DHFR_207. The DHFR_201 transcript is considered the main DHFR processed mRNA molecule. It is approximately 3900 bp in length and is broken down into six exons. This DHFR transcript contains a 561 bp open reading frame (ORF) which encodes a 187 amino acid long protein, known as the DHFR enzyme (Anagnou et al., 1984; Chen et al., 1984; Ensembl - DHFR, 2019; Maurer et al., 1984). This enzyme carries out the dihydrofolate reductase activity in the cytoplasm, nucleus and mitochondria of a cell (Anderson et al., 2011, 2012; Appling, 1991; Hughes, unpublished; Woeller et al., 2007). This consists of catalysing the reduction of dietary folic acid down to DHF and subsequently converting DHF to its biologically active form THF in an NADPH-dependent manner (Figure 1.5) (Stover, 2009). This reduction is extremely important as THF supplies one-carbon units for several of the metabolic reactions within the OCM pathway. Therefore, is essential for sustaining DNA biosynthesis, amino acid synthesis, cell proliferation and cellular methylation (Schnell et al., 2004).

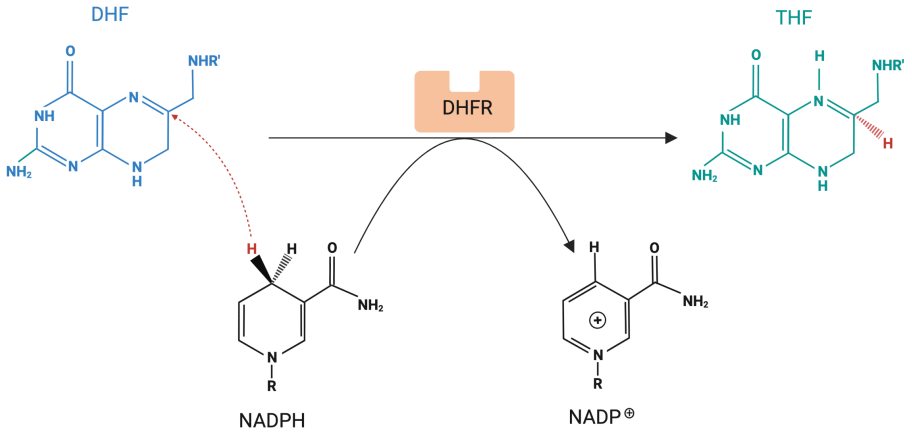


Figure 1.5 Mechanism of action by the DHFR enzyme. *The DHFR enzyme catalyses the reduction of 7,8-dihydrofolate to 5,6,7,8-tetrahydrofolate, via NADPH oxidation. Created with BioRender.com, Adapted from Stover, 2009.*

1.5.1.1. DHFR enzyme structure

The human DHFR enzyme is a relatively small protein with a molecular weight of approximately 21 kDa. It is composed of multiple secondary structural elements including an eight-stranded β sheet backbone (βA - βH) connected to four α-helices (αB, αC, αE, αF) by extended loops (Oefner et al., 1988). The eight-stranded β sheet can be broken down into seven parallel strands and a C-terminal antiparallel strand. A combination of all the secondary structural elements contributes to the final tertiary structure of the DHFR enzyme (Figure 1.6). The active site of the DHFR enzyme is the location where the substrate DHF and cofactor NADPH bind. The active site of the DHFR enzyme is situated between two subdomains, the adenosine-binding subdomain and the loop subdomain (Figure 1.6) (Matthews et al., 1977). The adenosine-binding subdomain consists of the βB, βC, βD and βE strands; the αC and αE helices; and the C-D loop. The loop subdomain is made up of the βA, βF, βG, and βH strands; αB and αF helices; and the Met20, F-G, and G-H loops. The size of the active site and thus the capability of substrates/cofactors to bind to the DHFR enzyme is controlled by the movements of these two subdomains (Blakley, 1995; Mauldin et al., 2012).

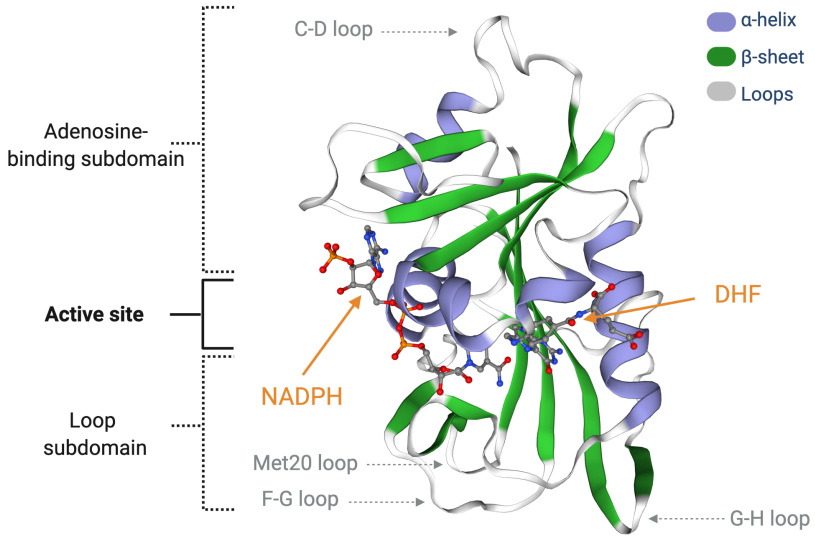


Figure 1.6 Crystal Structure of the DHFR Protein from Swiss Model. *The α-helixes are shown in purple, β-sheets in green and the loops in grey in this ribbon diagram. DHFR protein is broken down into two subdomains: the adenosine-binding subdomain and the loop subdomain. The active site of DHFR is located between these two subdomains. The image illustrates the binding site for the NADPH cofactor and DHF substrate, these ligands are shown as ball and stick molecules. Created with swissmodel.expasy.org.*

1.5.1.2. DHFR regulation

As the DHFR enzyme has such a vital role in maintaining normal cellular function it is expressed in every cell in the body at all stages of development. However, the level at which the DHFR protein is expressed varies depending on multiple factors. These factors tightly regulate the expression of the *DHFR* gene at multiple levels: transcriptional, translational and post-translational (Abali et al., 2008).

The degree to which the *DHFR* gene is being transcribed depends on what phase of the cell cycle the cell is in. The transcriptional rate of the *DHFR* gene increases near the end of the G1 phase and peaks during the S phase. Thus, the transcriptional rate of the *DHFR* gene is at its highest during DNA replication (Azizkhan et al., 1993). This transcriptional regulation is controlled by the transcription factors Sp1 and E2F. These ubiquitous transcription factors are able to modulate the transcription of the *DHFR* gene as it contains a TATA-less promoter. The Sp1 factor is constantly bound to the promoter region of the *DHFR* gene inducing transcription. Whereas, E2F binds to the promoter to repress the transcription of DHFR in the G0 and early G1 phases (Abali et al., 2008; Jensen et al., 1997). The function of the E2F transcription factor can be altered with the help of a regulatory protein complex known as pocket proteins. The pocket protein complex is made up of three regulatory proteins: retinoblastoma (Rb) and its two homologs, p107 and p130. Therefore, E2F can initiate the transcription of DHFR with the help of the pocket proteins and Sp1, in conjunction with the entrance into S-phase (Abali et al., 2008; Jensen et al., 1997; Slansky & Farnham, 1996).

The transcription of the *DHFR* gene is also controlled by a non-coding DHFR transcript. This non-coding DHFR RNA isoform is generated from an alternative promoter located approximately 400 bp upstream of the major promoter. The non-coding DHFR RNA molecule binds to DHFR's main promoter and forms a stable complex. This results in the pre-initiation structure dissociating from the major promoter, thus preventing transcription of the *DHFR* gene from starting (Abali et al., 2008; Martianov et al., 2007).

The translation of the *DHFR* gene is governed by an auto-regulatory mechanism. This is where the intact DHFR protein binds to an 82 bp section within the coding region of the DHFR mRNA to suppress its translation (Ercikan et al., 1993; Tai et al., 2004). This mechanism has become part of a bigger model explaining how the translation process of the *DHFR* gene is regulated (Figure 1.7). The DHFR protein exists as two conformers that are in equilibrium. The first is the DHFR protein bound to NADPH. The addition of an inhibitor (e.g. Methotrexate (MTX)) or a substrate (e.g. DHF) moves the DHFR enzyme:NADPH complex towards the catalytic form, ready to facilitate the reduction of

DHF to THF. The second conformer is the DHFR protein bound to its own mRNA to inhibit translation. The addition of an inhibitor (e.g. MTX) or a substrate (e.g. DHF) changes the conformation of the DHFR enzyme:mRNA complex. Resulting in the release of the DHFR mRNA leading to the resumption of its translation (Abali et al., 2008; Ercikan-Abali et al., 1997).

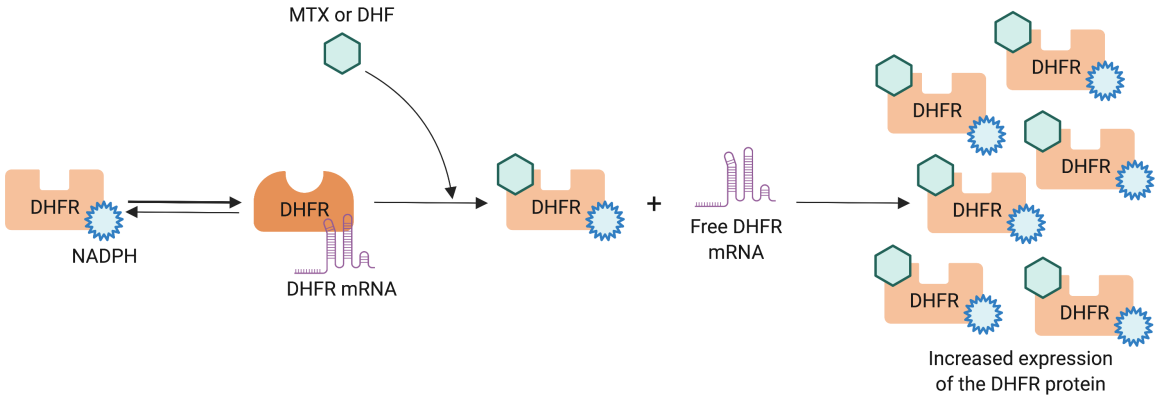


Figure 1.7 The model for DHFR translational up-regulation. *The DHFR protein exists as two conformers: DHFR protein bound to NADPH and DHFR protein bound to its own mRNA. If MTX or DHF binds to the DHFR enzyme:NADPH complex the equilibrium moves towards the catalytic form. Whereas, if MTX or DHF binds to the DHFR enzyme:mRNA complex it results in a conformational change which releases the DHFR mRNA. This change leads to an increased expression of the DHFR protein. Created with BioRender.com, Adapted from Abali et al., 2008.*

Finally, at a post-translational level, the DHFR protein is prone to sumoylation and phosphorylation. The DHFR coding sequence contains a SUMO motif at residues 178-181. The putative sumoylation site within this motif is residue 179 which is a lysine amino acid. The sumoylation of this motif aids the translocation of the DHFR protein into the nucleus to form part of the de novo thymidylate synthesis pathway, described in more detail in Section 1.3.3 (Anderson et al., 2007). The S145 and S168 amino acid residues within the DHFR protein are also subject to phosphorylation. The cellular role of this modification remains unanswered (Skierka et al., 2019). However, several scientists have suggested that the phosphorylation of these residues could regulate the expression of the DHFR protein (Klein et al., 2016; Maguire et al., 2008; Skierka et al., 2019; St-Denis et al., 2015).

1.5.1.3. DHFR as a therapeutic target

The vital role of the DHFR enzyme in cell proliferation has made it a well-established drug target for the treatment of several clinical diseases, including rheumatoid arthritis, head cancer, neck cancer, breast cancer, lung cancer, acute lymphoblastic leukaemia and non-Hodgkin's lymphoma (Cronstein, 1996; Raimondi et al., 2019). MTX is one of the most effective antifolate drugs that can be administered for these clinical diseases (Tian & Cronstein, 2007). This drug tightly binds to the DHFR enzyme and inhibits its reductase activity. It does this by specifically targeting the DHFR translational auto-regulatory mechanism described in Section 1.5.1.2 and Figure 1.7 (Skacel et al., 2005). The suppression of DHFR by folate antagonists (MTX) results in a decreased availability of THF. For this reason, the cells do not have one-carbon units to facilitate thymidylate and purines biosynthesis. Therefore, there are not sufficient nucleotides to support DNA synthesis and repair (Cronstein, 1996; Kremer, 2004). Thus, MTX is an ideal drug for treating diseases where uncontrollable cell growth is a key pathology (Hsieh et al., 2009). However, there is one major problem associated with prescribing this antifolate drug to certain patients. There is a possibility of patients developing MTX drug resistance. It can be acquired via amplification of the *DHFR* gene or through mutations to the *DHFR* gene that is specific for MTX binding. This results in a decreased affinity for any antifolate drug, rendering them ineffective (Roper et al., 2004; Skacel et al., 2005). Therefore, carefully monitoring the effectiveness of this drug is required.

1.5.1.4. The effect of polymorphisms within the *DHFR* gene

Several polymorphisms within the *DHFR* gene have been associated with disease susceptibility and response to drug treatment (Askari and Krajinovic, 2010).

The 19 bp deletion located within the DHFR promoter region has been extensively studied over the years due to its association with developing certain clinical diseases. One of which is its link to NTDs, as described in Section 1.4.1. Johnson et al. 2004 and Parle-McDermott et al. 2007 described conflicting views in relation to the 19 bp deletion in mothers and its link to offspring with NTDs. Therefore, the association between the DHFR 19 bp deletion and NTDs remains undecided in the scientific community. However, the DHFR 19 bp deletion allele has been found to have a protective role in adult acute lymphoblastic leukaemia patients (Gemmati et al., 2009).

As mentioned in Section 1.5.1.3, variations within the 3'UTR of the *DHFR* gene have been found to affect MTX efficiency. One of the first SNPs to be discovered in the DHFR 3'UTR was the 829 C>T SNP (rs1677669). The 829 C>T substitution was shown to induce DHFR expression which results in

a reduced MTX-related cytotoxicity (Goto et al., 2001). Two subsequent polymorphisms, 35289 A>T and 1171 A>T were also identified to have a similar effect (Askari & Krajinovic, 2010).

1.5.1.5. The additional role of the DHFR enzyme

The DHFR enzyme has been found to have an additional role in the Nitric Oxide (NO) pathway (Figure 1.8). This pathway generates the compound nitric oxide, which is involved in regulating cerebral blood flow and cellular oxygenation (Chen et al., 2008a). Nitric oxide is synthesised in pulmonary vascular endothelial cells by the endothelial nitric oxide synthase (eNOS) enzyme. Tetrahydrobiopterin (BH4) is an essential cofactor for the eNOS enzyme. The enzymatic coupling of eNOS and BH4 generates NO molecules which maintain normal endothelial function (Figure 1.8). NO deficiency through decreased availability of BH4 causes endothelial dysfunction. Problems within the endothelium have been linked to health issues and clinical diseases, including hypertension, hypercholesterolemia, diabetes mellitus and heart failure (Klinger & Kadowitz, 2017; Luiking et al., 2010; Tousoulis et al., 2012). Oxidative stress is one mechanism that can decrease the availability of BH4; this mechanism converts BH4 to dihydrobiopterin (BH2). BH2 can subsequently bind to the eNOS enzyme which generates superoxide instead of NO. The binding of BH2 to the eNOS enzyme decreases the availability of free eNOS enzymes that would be used to generate NO with the remainder of the BH4 cofactors, ultimately resulting in endothelial dysfunction (Figure 1.8) (Bendall et al., 2014; Dumitrescu et al., 2007). The DHFR enzyme has been found to recycle BH2 back to the BH4 cofactor. This reaction increases the bioavailability of BH4 and freely available eNOS enzymes, allowing them to couple together to generate sufficient NO to maintain normal endothelial function (Figure 1.8) (Crabtree et al., 2009, 2011).

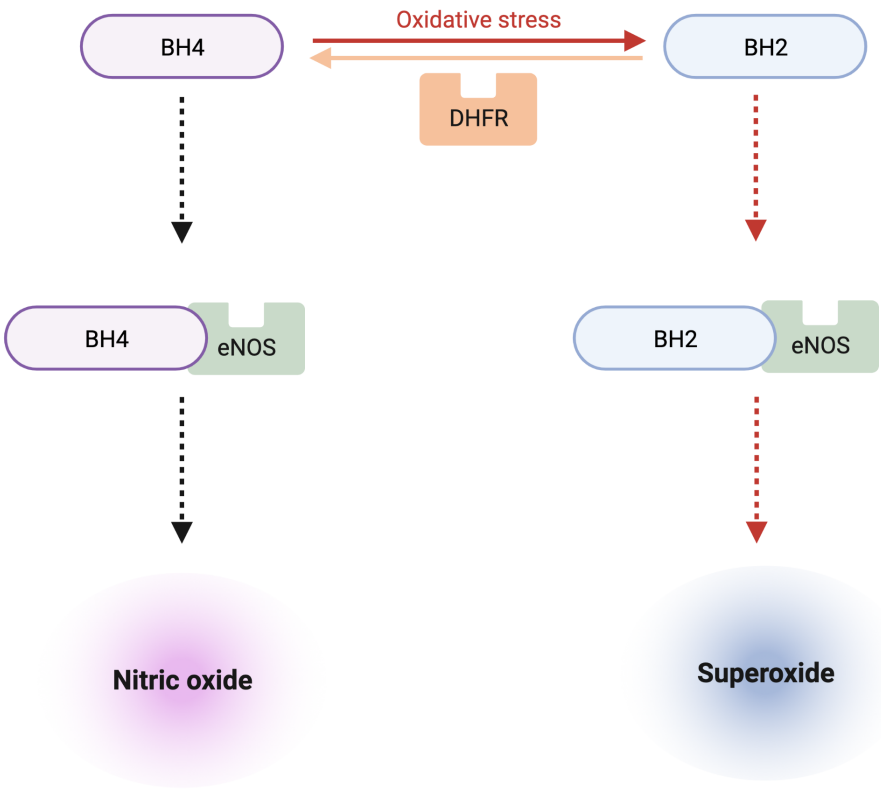


Figure 1.8 Schematic diagram of the NO pathway that occurs in endothelial cells. *The eNOS enzyme binds with the BH4 cofactor to generate NO molecules, which maintain normal endothelial function. Oxidative stress converts BH4 to BH2 which generates superoxide when bound to the eNOS enzyme. This reaction reduces the BH4 pools and freely available eNOS enzymes to generate NO, resulting in endothelial dysfunction. The DHFR enzyme has been found to convert BH2 back to the BH4 cofactor. This reaction increases the bioavailability of BH4 and freely available eNOS enzymes, allowing them to couple together to generate sufficient NO to maintain normal endothelial function. Created with BioRender.com, Adapted from Crabtree et al., 2009, 2011.*

1.5.2. Dihydrofolate reductase 2 gene

The second dihydrofolate reductase gene, *DHFR2*, spans a 6 kb region located on chromosome 3q11.2 of the human genome. It originated from the reverse transcription of mature mRNAs belonging to the *DHFR* gene being inserted back into the genome, i.e., *DHFR2* is a retrogene (Fablet et al., 2009; McEntee et al., 2011). Hughes et al. (2015) reported that the *DHFR2* gene was not a feature in the genome of all mammals, such as rodents. Therefore, the *DHFR2* retrogene appears to be unique to primates.

Ensembl has predicted that the *DHFR2* gene encodes six RNA isoforms identified as DHFR2_201 through DHFR2_206 (Figure 1.9). DHFR2_201, DHFR2_202 and DHFR2_206 contain a 561 nucleotide long ORF that is predicted to synthesise a 187 amino acid long protein, known as the DHFR2 protein. DHFR2_205 encodes a truncated version of the DHFR2 protein. DHFR2_203 is capable of synthesising the DHFR2 protein but contains a nonsense mutation resulting in it being subject to nonsense-mediated decay. And finally, DHFR2_204 does not have the DHFR2 ORF and is predicted to be a processed transcript (Ensembl, 2019).

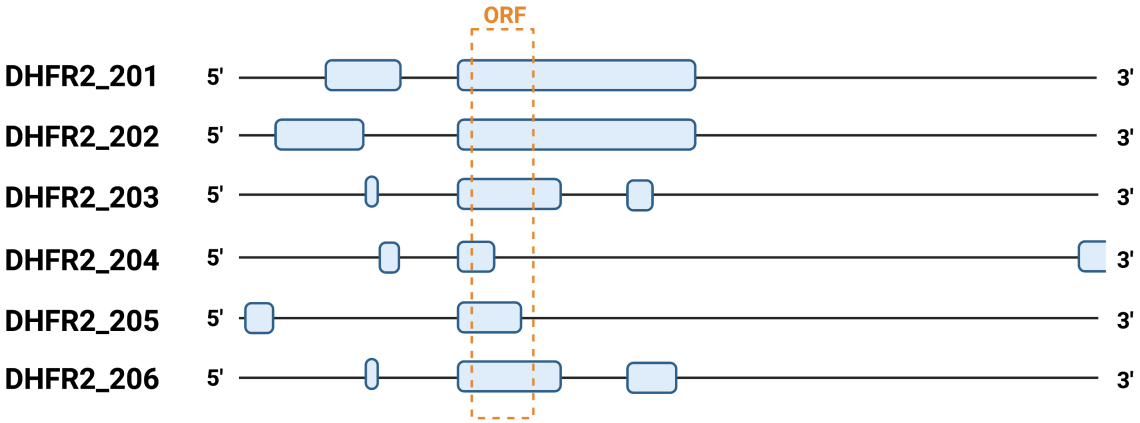


Figure 1.9 The 6 DHFR2 RNA isoforms annotated by Ensembl. *Exons are shown as blue boxes, introns as black lines and the DHFR2 ORF is marked with the orange box. DHFR2_201, DHFR2_202 and DHFR2_206 transcripts are predicted to generate the 187 amino acid DHFR2 protein as they contain the entire ORF. DHFR2_203 is predicted to generate the DHFR2 protein but is subject to nonsense-mediated decay. DHFR2_205 is predicted to synthesise a truncated version of the DHFR2 protein, containing 114 amino acids. Finally, the DHFR2_204 mRNA isoform is predicted to not generate a protein and is marked as a processed transcript. Created with BioRender.com, Adapted from ensembl.org/index.html.*

The DHFR2_201 transcript is considered one of the most plausible mRNA isoforms to generate the functional DHFR2 protein. Therefore, all the research carried out prior to this project only concentrated on the DHFR2_201 mRNA isoform. A region of the DHFR2_201 transcript has been experimentally confirmed in various human cells (Anderson et al., 2011; McEntee et al., 2011). Furthermore, public databases have ample evidence of the DHFR2_201 mRNA isoform in multiple tissues, summarised at: https://www.proteinatlas.org/ENSG00000178700-DHFR2/summary/rna. However, despite the DHFR2_201 transcript being easily detected across all cell and tissue types examined there was no evidence of the DHFR2 protein. This included experimental research

performed in the APMcD laboratory and external research projects available on public databases, available at: https://www.proteomicsdb.org/proteomicsdb/#overview. When comparing the amino acid sequence for the DHFR protein and predicted DHFR2 protein it reveals that the two are 92% homologous (Figure 1.10). This makes distinguishing between the two very challenging. Therefore, this could be one of the possible reasons why the DHFR2 protein was not identified.

The similarity between the amino acid sequence of the DHFR protein and the predicted DHFR2 protein suggests that the DHFR2 protein could function as a dihydrofolate reductase enzyme. Model predictions were previously carried out in the APMcD laboratory to determine if the motifs required for dihydrofolate reductase catalytic activity were conserved in the DHFR2 protein (Figure 1.10) This modelling was performed using the online tool known as SWISS-MODEL (https://swissmodel.expasy.org/). The analysis revealed that the amino acids required for DHF binding are fully conserved in the DHFR2 amino acid sequence. This result suggests that the DHFR2 protein has the capabilities to carry out dihydrofolate reductase activity. However, the NADPH binding site does not appear to be fully intact in the DHFR2 amino acid sequence. Three out of the 38 amino acid residues responsible for NADPH binding appear to be different in the DHFR2 sequence. The most significant being the Trp25 residue found within the DHFR sequence being an Arg in the DHFR2 sequence. The additional NH_3 moiety found on the Arg residue more than likely lodges in the active site for DHF and NADPH. This may result in DHFR2 having a lower affinity for DHF and/or NADPH. Thus, the DHFR2 protein could provide dihydrofolate reductase activity in the cell but it may need to employ a smaller cofactor, such as NADH (Hughes, unpublished).



Figure 1.10 Sequence alignment comparing the amino acid sequence of the human DHFR protein versus the predicted DHFR2 protein, using Snapgene*. This software estimated that the two sequences share 92% homology, with only 15 amino acid differences. The four motifs required for DHFR catalytic activity are shown in the yellow (NADPH binding sites) or green (folate binding site) box. The folate binding site is conserved in the predicted DHFR2 protein. However, three of the amino acid residues within the NADPH binding site were not conserved. The most significant amino acid change is the Trp25 residue in DHFR being an Arg in DHFR2, shown underlined in orange. The amino acids important for DHFR mRNA binding are conserved in the DHFR2 sequence, underlined in blue. The proposed DHFR sumoylation motif (purple box) and phosphorylation sites (underlined in red) are also conserved in the DHFR2 protein. Created with snapgene.com, Adapted from McEntee et al., 2011.*

Due to the fact that the endogenous DHFR2 protein was not detectable, the APMcD and Stover lab groups switched to using the recombinant version of the protein to uncover some more information about this potential protein (Anderson et al., 2011; McEntee et al., 2011). Anderson et al., 2011 discovered the mitochondria contains the de novo thymidylate synthesis pathway to support mitochondrial DNA replication. McEntee et al., 2011 demonstrated that the DHFR2-GFP recombinant protein localised to the mitochondria in unsynchronised human embryonic kidney 293 (HEK293) cells. Therefore, dihydrofolate reductase activity was subsequently revealed in this organelle as it is one of the enzymes that drive de novo thymidylate synthesis (Anderson et al., 2011). Due to no prior evidence of the DHFR enzyme in the mitochondria and the fact that the recombinant protein was found to localise to this compartment, it was concluded that the function

of the *DHFR2* gene was to encode the mitochondrial reductase enzyme (Anderson et al., 2011; McEntee et al., 2011; Wang et al., 1967). After, the publication of these results the APMcD lab continued to investigate the mitochondrial reductase enzyme. Liquid chromatography-tandem mass spectrometry (LC-MS/Ms) analysis on mitochondrial fractions from HEK293 cells failed to detect the endogenous DHFR2 protein. However, the endogenous DHFR protein was identified in the mitochondrial fractions. This information revealed that endogenous human DHFR localises to the mitochondria to carry out the reductase activity (Hughes, unpublished). Therefore, the function of the *DHFR2* gene was once again unknown. This project aimed to identify why primates have evolved to have a second functional dihydrofolate reductase gene.

Despite the DHFR2_201 RNA being highly abundant in all cells/tissues (Anderson et al., 2011; McEntee et al., 2011), there is no evidence of the native DHFR2 protein. This preliminary research suggests that the DHFR2 protein could be highly regulated at the translational level, cell/tissue specific or only present under particular cellular conditions. A genetic association study of an Irish cohort found an association between 3 SNPs within the *DHFR2* gene and the risk of developing NTDs, described in Section 1.4.1 (Pangilinan et al., 2021). These findings form the basis for the hypothesis of this project, that the *DHFR2* gene encodes a reductase enzyme that is only expressed and functional during early development to aid OCM.

1.6. Aims and Objectives

This project intends to fully characterise the human *DHFR2* gene, with particular attention on identifying its overall cellular function and its role in the OCM pathway. The vision is to examine its expression and behaviours at a DNA, RNA and protein level using a number of approaches. This PhD thesis will mainly focus on investigating DHFR2 RNA and protein, to address the following questions:

1. Do the DHFR2_201 and DHFR2_202 transcripts have the capabilities to be translated into the DHFR2 protein?
2. Can the recombinant DHFR2 protein be synthesised using a mammalian-expression system? Or is there something within the DHFR2 ORF repressing the expression?
3. Does the *DHFR2* gene encode a functional protein? Is the native DHFR2 protein expressed in specific cell/tissue types (e.g. human embryonic tissue)? Is the DHFR2 protein translated under particular cellular conditions (e.g. in the absence of the DHFR protein)?
4. Are the DHFR2_201 and DHFR2_202 RNA transcripts subject to translational regulation via RNA editing?
5. Does modulating the expression of the *DHFR2* gene affect other proteins within a cell? Does altering the expression of the *DHFR2* gene impact other proteins within the OCM pathway?

The above research questions will be met by achieving the following specific Aims and Objectives.

Aim 1: Interrogation of the translatability of DHFR2 RNA by assessing ribosome binding.

- Objective 1.1: Bioinformatically analyse ribosome profiling data available on RPFdb v2.0 beta that is specific to the *DHFR2* gene.
- Objective 1.2: Sucrose density gradient centrifugation RT-qPCR analysis will be performed to assess the association of the two main DHFR2 transcripts (DHFR2_201 and DHFR2_202) with the ribosome.

Aim 2: Determine if the recombinant human DHFR2 protein can be synthesised in mammalian cells.

- Objective 2.1: Generate recombinant DHFR2 protein using a mammalian expression system. This technique is robust but not fully representative of a native cellular environment.

- Objective 2.2: Synthesise recombinant DHFR2 protein in a humanised DHFR2 mouse model. This cellular environment is more representative of the natural one.

Aim 3: Investigate the presence of the endogenous DHFR2 protein in various cell/tissue types and particular cellular conditions.

- Objective 3.1: Data-dependent acquisition (DDA) and parallel reaction monitoring (PRM) LC-MS/MS on a range of adult cells and tissues to determine if the endogenous DHFR2 protein is present.
- Objective 3.2: DDA and PRM LC-MS/MS on several cellular conditions to determine if the endogenous DHFR2 protein is present when the DHFR enzyme is absent or if the native DHFR2 protein is subject to degradation.
- Objective 3.3: Multiple reaction monitoring (MRM) LC-MS/MS on a range of human embryonic and foetal tissues to determine if the endogenous DHFR2 protein is present.

Aim 4: Determine if the DHFR2 transcripts are subject to translational regulation via RNA editing.

- Objective 4.1: RT-qPCR analysis will be performed to determine if the DHFR2 transcripts are prone to RNA editing. This could lead to them being retained in the nucleus and ultimately lead to translational repression.

Aim 5: Assess the impact of modulating the expression of the *DHFR2* gene on overall protein expression to determine if the *DHFR2* gene has an important cellular function.

- Objective 5.1: Mass spectrometry-based differential protein expression analysis will be carried out on a HepG2 DHFR2 knockout cell line compared to the HepG2 control cell line.
- Objective 5.2: Mass spectrometry-based differential protein expression analysis will be carried out on a HepG2 DHFR2 overexpression cell line compared to the HepG2 control cell line.

Chapter 2. Materials and Methods

2.1. Materials

2.1.1. Reagents

Agilent: XL10-Gold Ultracompetent Cells (Cat. no. 200314).

Beckman: Tube, Thickwall, Polycarbonate, 1 mL, 11 x 34 mm (Cat. no. 343778).

Bioline: 5x Polymerase Chain Reaction (PCR) Reaction Buffer (Cat. no. MB-109I), Bioscript 200 U/µL (Cat. no. BIO-27036), Oligo dT Primers (Cat. no. BIO-38029), Random Hexamer Primers (Cat. no. BIO-38028).

Bio-Rad Laboratories: Quick Start Bradford 1x Dye Reagent 1 L (Cat. no. 5000205), Quick Start Bradford Protein Assay Kit 1 (Cat. no. 5000201).

Biotage: Isolute C18 Cartridges (Cat. no. 220-0010-A).

Fisher Scientific: Ethanol (Cat. no. 10048291), Macherey-Nagel NucleoSpin Gel and PCR Clean-up Kit (Cat. no.740609.10), Methanol (Cat. no. M/4056/17).

Laborgeraete Beranek GmbH: Adapter for thick-walled tubes 1 mL, 11x34 mm, 'long version' (Cat. no. 11015).

Merck-Millipore: Microcon-10 kDa Centrifugal Filter Unit with Ultracel-10 Membrane (Cat. no. MRCPRT010).

New England Biolabs: 100 bp DNA Ladder (Cat. no. N3231L), Gel Loading Dye Purple 6x (Cat. no. B7024S), Q5 High-Fidelity DNA Polymerase (Cat. no. M0491S), Quick-load 1 kb DNA Ladder (Cat. no. N0468S), ssRNA Ladder (Cat. no. N0362S).

OriGene: Anti-DDK Antibody (Cat. no. TA50011-100), Anti-turboGFP Antibody (Cat. no. TA150071), pCMV6-AC-DHFR2-DDK-Myc Expression Vector (Cat. no. RC208452), pCMV6-AC-DHFR2-tGFP Expression Vector (Cat. no. RG232027), pCMV6-AC-tGFP Expression Vector (Cat. no. PS100010).

Precellys: Tissue Homogenizing CKMix (Cat. no. P000918-LYSK0-A).

Promega: ProteaseMAX Surfactant Trypsin Enhancer (Cat. no. V2071), Trypsin Gold Mass Spectrometry Grade (Cat. no. V5280).

Qiagen: DNeasy Blood & Tissue Kit (Cat. no. 69506), RNeasy Mini Kit (Cat. no. 74104).

Roche: Ampicillin (Cat. no. 91084621), FastStart Essential DNA Probes Master (Cat. no. 06402682001), LightCycler 480 Multiwell Plate 96, white (Cat. no. 04729692001).

Sigma-Aldrich: 10X PCR Buffer without Magnesium Chloride ($MgCl_2$) (Cat. no. P2317), Acetic Acid (Cat. no. A6283), Agarose (Cat. no. A9539), Ammonium Bicarbonate (Cat. no. 09830), Boric Acid (Cat. no. B6768), 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonat (CHAPS) (Cat. no. C9426), Cycloheximide (CHX) (Cat. no. C1988), Dimethyl Sulfoxide Hybri-max Sterile (DMSO) (Cat. no. D2650), DNase I Amplification Grade (Cat. no. AMPD1), dNTPs Set 100 mM (Cat. no. DNTP100), Dithiothreitol (DTT) (Cat. no. 10197777001), Dulbecco's Phosphate Buffered Saline (PBS) (Cat. no. D8537), Ethidium Bromide Solution (Cat. no. E1510), Ethylenediaminetetraacetic Acid (EDTA) (Cat. no. E9884), Formic Acid (Cat. no. F0507), Glycerol (Cat. no. G5516), Glycine (Cat. no. G8898), Iodoacetamide (Cat. no. I1149), Luria Broth (LB) (Cat. no. L3522), LB Broth with Agar (Cat. no. L2897), MG132 (Cat. no. 474790), $MgCl_2$ (Cat. no M8266), $MgCl_2$ solution for PCR (Cat. no. M8266), N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES) (Cat. no. H4034), Nuclease Mix (100X) (Cat. no. GE80-6501-42), Potassium Chloride (KCl) (Cat. no. P9541), Ponceau S (Cat. no. P3504), Ribonuclease Inhibitor Human 20,000 U/mL (Cat no. R2520), Sodium Chloride (Cat. no. S3014), Sodium Dodecyl Sulfate (SDS) (Cat. no. L3771), Sodium Hydroxide (Cat. no. 221465), Sorbitol (Cat. no. S1876), Sucrose (Cat. no. 84097), Taq DNA Polymerase 5 U/µL (Cat. no. D4545), Thiourea (Cat. no. T8656), Trifluoroacetic acid (Cat. no. 91707), Tris-HCl (Cat. no. T5941), Trizma Base (Cat. no. T1503), Tween-20 (Cat. no. P1379), Ubiquitin Monoclonal Antibody (Cat. no. 14-6078-82), Universal ProbeLibrary Probe #46 (Cat. no. 04688066001), Universal ProbeLibrary Probe #57 (Cat. no. 04688546001), Universal ProbeLibrary Probe #60 (Cat. no. 04688589001), Universal ProbeLibrary Probe #72 (Cat. no. 04688953001), Universal ProbeLibrary Probe #75 (Cat. no. 04688988001), Universal ProbeLibrary Probe #89 (Cat. no. 04689143001), Urea (Cat. no. U5378), Water Molecular Biology (Cat. no. W4502).

Thermo Fisher Scientific: 1 Step Transfer Buffer (Cat. no. 84731), Acetonitrile (Cat. no. 51101), C18 Spin Columns (Cat. no. 89873), Chloroform – Isoamyl Alcohol Mixture (Cat. no. 25666), CloneJET PCR Cloning Kit (Cat. no. K123240), ExoSAP-IT Express PCR Product Cleanup Reagent (Cat. no. 75001), Gel Loading Buffer II (Denaturing PAGE) (Cat. no. AM8546G), GeneJET Plasmid Miniprep Kit (Cat. no. K0502), Gibco Dulbecco's Modified Eagle Medium (DMEM), High Glucose, GlutaMAX Supplement (Cat. no. 61965026), Gibco DMEM, High Glucose, Pyruvate (Cat. no. 41966029), Gibco Essential 8 Flex Medium Kit (Cat. no. A2858501), Gibco Fetal Bovine Serum (FBS) (Cat. no. 10270106), Gibco GlutaMAX Supplement (Cat. no. 35050061), Gibco RPMI 1640 Medium, Gibco Trypsin-EDTA (0.25%) Phenol Red 100 mL (x5) (Cat. no. 25200056), GlutaMAX Supplement (Cat. no. 61870010), Glass Beads diam. 0.5 mm, for Mini-Beadbeater (Cat. no. Z250465), Goat Anti-Rabbit IgG (H+L) Cross-adsorbed Secondary, Horseradish Peroxidase (HRP) (Cat. no. G-21234), Halt Protease and Phosphatase Inhibitor Cocktail (100X) (Cat. no. 78440), Lipofectamine 3000

Transfection Reagent (Cat. no. L3000008), Novex 4-20% Tris-Glycine Mini Gels, Wedgewell Format, 12-well (Cat. no. XP04202BOX), NuPAGE LDS Sample Buffer (4X) (Cat. no. NP0007), NuPAGE Sample Reducing Agent (10X) (Cat. no. NP0004), MEM Non-Essential Amino Acids (100X) (Cat. no. 11140035), PageRuler Plus Prestained Protein Ladder, 10 to 250 kDa (Cat no. 26619), Penicillin-Streptomycin (10 000 U/mL) (Cat. no. 15140122), PureLink RNA Mini Kit (Cat. no. 12183025), PVDF Transfer Membrane, 0.45 µm, 26.5 cm x 3.75 m (Cat. no. 88518), Rat Anti-Mouse IgG (H+L) Secondary Antibody, HRP (Cat. no. 04-6020), Sodium Hypoxanthine and Thymidine (HT) Supplement (100X) (Cat. no. 11067030), SuperScript III First-Strand Synthesis SuperMix (Cat. no. 18080400), Super Optimal Broth (S.O.C.) Medium (Cat. no. 15544034), SuperSignal West Pico PLUS Chemiluminescent Substrate (Cat. no. 34580), SYBR Safe DNA Gel Stain (Cat. no. S33102), TRIzol Reagent (Cat. no. 15596026), TURBO DNase (2 U/µL) 1000 U (Cat. no. AM2238), Western Blotting Filter Paper, 7 cm x 8.4 cm (Cat. no. 84783).

Waters: Polypropylene Screw Neck Vial (Cat. no. 186002640).

2.1.2. Solutions

Activation solution: 50% (v/v) acetonitrile.

Alkylation solution: 8 M urea, 0.05 M iodoacetamide in tris buffer 1 (pH 8.5).

Ammonium bicarbonate solution: 50 mM ammonium bicarbonate.

Ampicillin: dissolved in molecular grade water to 100 µg/mL.

Antibody solution: 1% (w/v) non-fat milk powder in 0.05% (v/v) (Tris Buffered Saline with Tween) TBST buffer (pH 7.4).

Blocking solution: 5% (w/v) non-fat milk powder in 0.05% (v/v) TBST buffer (pH 7.4).

Cushion sucrose solution (CSS): 300 mM sorbitol, 20 mM HEPES pH 7.5, 1 mM EDTA, 5 mM $MgCl_2$, 10 mM KCl, 10% (v/v) glycerol, 100 µg/mL CHX, Halt Protease and Phosphatase Inhibitor Cocktail (1x).

DHFR2 synthetic peptide: lyophilised peptide reconstituted in resuspension solution to a final concentration of 0.115 ng/µL.

DTT solution: 0.5 M DTT in 50 mM ammonium bicarbonate.

Elution solution: 70% (v/v) acetonitrile.

Equilibrium solution: 2% (v/v) acetonitrile, 0.1% (v/v) trifluoracetic acid.

LB Agar: 35 g of LB agar per litre of distilled water.

LB Broth: 20 g of LB broth per litre of distilled water.

MG132 stock solution: 1 mg MG132 in 1 mL DMSO.

Ponceau stain: 0.1% (w/v) Ponceau S, 5% (v/v) acetic acid.

ProteaseMAX surfactant solution (1%) (w/v): ProteaseMAX Surfactant powder in 50 mM ammonium bicarbonate.

Protein lysis buffer (DCU): 7 M urea, 2 M thiourea, 4% (w/v) CHAPS, 30 mM trizma base, 1x Protease and Phosphatase Inhibitor, 1x nuclease mix, pH 8.5.

Protein lysis buffer (UCL): 100 mM ammonium bicarbonate, 100 mM DTT, 1% (w/v) sodium deoxycholate, pH 8.

Resuspension solution: 2% (v/v) acetonitrile, 0.1% (v/v) formic acid.

Rinse solution A: 0.5 M sodium chloride.

Rinse solution B: 2% (v/v) acetonitrile, 0.1% (v/v) trifluoracetic acid.

Sequence grade trypsin solution (0.4 µg/µL): Trypsin Gold powder in 50 mM acetic acid.

Sequence grade trypsin solution (1 µg/µL): Trypsin Gold powder in 50 mM acetic acid.

Solvent A: 2% (v/v) acetonitrile and 0.1% (v/v) formic acid

Solvent A1: 0.1% (v/v) formic acid.

Solvent B: 80% (v/v) acetonitrile, 0.08% (v/v) formic acid.

Solvent B1: acetonitrile, 0.1% (v/v) formic acid.

Sucrose solution: 60% (w/v) sucrose, 20 mM HEPES pH 7.5, 1 mM EDTA, 5 mM $MgCl_2$, 10 mM KCl, 10% (v/v) glycerol, 100 µg/mL CHX.

Tris-borate-EDTA (TBE) buffer (10X): 89 mM trizma base, 90 mM boric acid, 20 mM EDTA, pH 8.3.

TBST buffer (0.05%): 0.15 M sodium chloride, 0.01 M trizma base, 0.05% (v/v) Tween-20, pH 7.4.

Trifluoroacetic acid solution: 0.1% (v/v) trifluoroacetic acid.

Tris buffer 1: 36.5 mM trizma base, 72 mM Tris-HCl, pH 8.5.

Tris buffer 2: 80.6 mM trizma base, 30 mM Tris-HCl, pH 7.9.

Tris glycine buffer (10X): 0.25 M trizma base, 1.9 M glycine, 1% (w/v) SDS, pH 8.3.

Urea solution: 8 M urea.

Wash solution 1: 8 M urea in tris buffer 1 (pH 8.5).

Wash solution 2: 8 M urea in tris buffer 2 (pH 7.9).

2.1.3. Cell lines

Human hepatocellular carcinoma (HepG2) cells: Sigma (Cat. no. 85011430), cultured by Paola Drago (DCU).

HepG2 DHFR knockdown cells: *DHFR* gene was knocked down in HepG2 cells (Sigma, Cat. no. 85011430) using CRISPR-Cas9, generated and cultured by Paola Drago (DCU).

HepG2 DHFR2 knockout cells: *DHFR2* gene was knocked out from HepG2 cells (Sigma, Cat. no. 85011430) using CRISPR-Cas9, generated and cultured by Paola Drago (DCU).

Human hepatocellular carcinoma (HuH7) cells: cultured by Dr Sandra Castro (UCL).

Human-induced pluripotent stem (iPS)-derived hepatocytes cells: differentiated from undifferentiated human iPS cells by Dr Sandra Castro (UCL).

Human neuroblastoma (IMR32) cells: cultured by Dr Sandra Castro (UCL).

Human neuroectodermal progenitor (NEP) stem cells: differentiated from undifferentiated human iPS cells by Denise Burtenshaw (DCU).

Undifferentiated human iPS cells: HipSci (Cat. no. 77650065), cultured by Denise Burtenshaw (DCU).

2.1.4. Tissue samples

Human embryonic/foetal tissue: Human developmental Biology Resource, see Table 2.1 for more information on the individual tissues.

Table 2.1 Breakdown of each of the human embryonic/foetal tissues

Stage	Tissue	Human embryo number	Aliquot number
Carnegie stage 21-22	Kidney	12613	5
	Heart	14755	5
	Lung	11364	10
	Liver	12092	15
	Brain	12210	3
Post-conceptual week 9-10	Kidney	10951	5
	Heart	14326	9
	Lung	12412	10
	Liver	12542	8
	Brain	12543	23
Post-conceptual week 15 -17	Kidney	14104	8
	Heart	14559	9
	Lung	13209	57
	Liver	13454	22
	Brain	13008	12

Human ovary tissue: BioIVT (Cat. no. 1283091F).

Human placenta, umbilical cord and membrane tissue: sourced from Dr Michelle Murphy (URV).

Human testis tissue: BioIVT (Cat. no. 114500A1).

Mouse embryos tissue – E11.5 (wildtype and humanised DHFR2 model with C57BL/6J background): generated by Dr Zoe Crane-Smith (UCL).

Postnatal mouse tissue – P20 (wildtype and humanised DHFR2 model with C57BL/6J background): generated by Dr Zoe Crane-Smith (UCL).

2.1.5. RNA samples

B-Lymphocyte (17158-Coriell) cells: Coriell Institute (Cat. no. NA17158), RNA was previously extracted using the Ultraspec-II and stored in a - 80 °C freezer.

HepG2 cells: Sigma (Cat. no. 85011430), RNA extracted using the Qiagen RNA Extraction Kit (See Section 2.2.4) and stored in a - 80 °C freezer.

Human iPS cells: HipSci (Cat. No. 77650065), RNA extracted using the Qiagen RNA Extraction Kit (See Section 2.2.4) and stored in a - 80 °C freezer.

Immortalized human embryonic kidney cells (HEK293): ATCC (Cat. no. CRL-1573), RNA previously extracted using the Qiagen RNeasy Kit and stored in a - 80 °C freezer.

2.2. Methods

2.2.1. Generation of HepG2 DHFR knockdown and DHFR2 knockout cell lines

A DHFR knockdown and a DHFR2 knockout cell line were generated by Paola Drago (DCU) using two different CRISPR/Cas9 strategies.

The DHFR knockdown was created in the HepG2 cell line using a CRISPR gene knock-in technique, termed PITCh (Precise Integration into Target Chromosomes) (Nakade et al., 2014). The PITCh system was employed to create a double-strand break in the coding sequence belonging to the *DHFR* gene and an enhanced GFP-2A-puromycin cassette was inserted. The addition of this cassette ultimately disrupted the *DHFR* gene. The mutated population was selected by treating the cells with puromycin and screened by PCR genotyping.

The DHFR2 knockout was created in the HepG2 cell line using the double-cut nonhomologous end-joining (NHEJ) strategy (Bauer et al., 2015). The double-cut NHEJ system was employed to create two double-strand breaks at the start and end of the DHFR2 coding sequence. The resulting extremities were re-joined by NHEJ. This resulted in the deletion of the entire ORF of the *DHFR2* gene. A monoclonal DHFR2 knockout cell line was obtained by single-cell cloning and screened by PCR.

2.2.2. General cell culture methods

HepG2 wildtype, DHFR knockdown and DHFR2 knockout cells were cultured in DMEM, high glucose (4,500 mg/L of D-Glucose), GlutaMAX, supplemented with 10% (v/v) FBS.

The human iPS-derived hepatocytes cell line is human iPS cells that were differentiated into hepatocyte-like cells for 21 days. These cells were grown and differentiated according to Rashidi et al., (2018).

HuH7 cells were grown in DMEM, high glucose (4,500 mg/L of D-Glucose), pyruvate (110 mg/L of Sodium Pyruvate), supplemented with 10% (v/v) FBS, 1% (v/v) Glutamax and 1% (w/v) Penicillin-streptomycin (P/S) media.

The IMR32 cell line was grown in RPMI-1640 media supplemented with 10% (v/v) FBS, 1% (w/v) P/S and 2.5% (v/v) HEPES.

The undifferentiated human iPS cells and human NEP stem cells were cultured in 8 Flex Media and 1% (w/v) P/S solution. The NEP stem cells were differentiated from the undifferentiated human iPS cells according to Di Luca et al., (2021).

All cells were grown in a sterile 5% carbon dioxide (CO_2) incubator at 37 °C until approximately 80% confluency was reached. After cells were washed with PBS they were harvested using Trypsin-EDTA (0.25% (v/v)). The cells were pelleted and stored in a - 80 °C freezer until required.

2.2.3. DNA extraction from cultured cells

A cell pellet from Section 2.2.2 was washed with 1 mL of PBS buffer (1x). DNA was extracted using the Qiagen DNeasy Blood and Tissue Kit, with no deviations from the manual.

2.2.4. RNA extraction from cultured cells

A cell pellet from Section 2.2.2 was washed with 1 mL of PBS buffer (1x). The RNA was extracted using the Qiagen RNA Extraction Kit, with no deviations from the manual.

2.2.5. DNA/RNA quantification

The quality and concentration of DNA/RNA samples were measured using a Nanodrop (One/One Microvolume UV-Vis Spectrophotometer by Thermo Fisher Scientific), at an absorbance of 260 nm.

2.2.6. RNA integrity check

The integrity of the RNA was analysed by mixing 500 ng of RNA in a total volume of 5 µL with an equal volume of Gel Loading Buffer II (Denaturing PAGE). The samples were loaded on a 1% (w/v) agarose gel as per Section 2.2.7.

2.2.7. Agarose gel electrophoresis

DNA/RNA products were run on a 1% (w/v) agarose gel unless stated otherwise. This was prepared by dissolving 1 g of agarose powder in a 100 mL TBE (1x) buffer. The gel stain (1 µL of 10 mg/mL ethidium bromide or 1x SYBR Safe Stain) was added once the mixture cooled to 55 °C. The mixture was poured into the gel mould. Once solidified the gel was covered with TBE (1x) buffer. DNA samples were mixed with purple DNA gel loading dye while RNA samples were mixed with RNA denaturing loading dye, with volumes stated in each section. DNA/RNA ladder (5 µL) was added to the first well of the gel and the DNA/RNA samples were added to subsequent wells. The gel was left to run for 1-2 hour at 100 volts. The gel was visualised and imaged under UV light (DNR Mini-Bis Pro Bio-Imaging System or Azure biosystems Azure 200 Gel Imaging Workstation).

2.2.8. DNase treatment and reverse transcription (RT) of RNA samples

2.2.8.1. DNase treatment and reverse transcription using Sigma and Bioline reagents

RNA samples were DNase treated using the DNase I amplification grade kit from Sigma. The RNA samples (1 µg) were diluted to a volume of 4 µL using water and mixed with 0.5 µL of reaction buffer and 0.5 µL of DNase I (1 U/µL). The reactions were incubated at room temperature for 15 min. The reactions were mixed with 0.5 µL of Stop Solution and incubated at 70 °C for 10 min using a Thermal Cycler (Applied Biosystems GeneAmp PCR System 9700).

The DNase treated RNA samples were reverse transcribed by adding 2 µL of 50 ng/µL Random Hexamers Primer Mix (Bioline) and 4 µL of 50 µM Oligo dT (Bioline) to 1 µg of RNA. The reactions were incubated at 70 °C for 5 min using a Thermal Cycler. The samples were left to cool on ice for 1 min before adding 4 µL 5x reaction buffer (Bioline), 40 U Ribonuclease Inhibitor Human (Sigma), 200 U Bioscript (Moloney Murine Leukaemia Virus Reverse Transcriptase) (Bioline), 1 µL of 10 mM dNTPs (Sigma) and 0.5 µL water. They were incubated at 25 °C for 10 min, 42 °C for 60 min and 70 °C for 15 min. The complementary DNA (cDNA) was stored at in a - 20 °C freezer until required.

2.2.8.2. DNase treatment and reverse transcription using Thermo Fisher Scientific kits

RNA samples were DNase treated using the Invitrogen TURBO DNase Kit from Thermo Fisher Scientific. A total of 1 µg of RNA was diluted to a volume of 5 µL of water. TURBO DNase buffer (1x) and TURBO DNase (2 U) were added to the RNA sample and incubated at 37 °C for 30 min. The DNase activity was stopped by adding 15 mM EDTA and incubated at 75 °C for 10 min.

The DNase treated RNA samples were reverse transcribed using the Invitrogen SuperScript III First-Strand Synthesis SuperMix. Annealing buffer (1 µL), primer (1 µL of 50 µM oligo(dT) or 50 ng/µL random hexamers) and RNase/DNase-free water (0.2 µL) was added to 1 µg of RNA. They were incubated at 65 °C for 5 min and immediately placed on ice for 1 min. SuperScript III (Moloney Murine Leukaemia Virus Reverse Transcriptase)/RNaseOUT (RNase Inhibitor) Enzyme Mix (2 µL) and 2X First-Strand Reaction Mix (10 µL) were added to the reaction. The reactions containing the oligo (dT) primers were incubated at 50 °C for 50 min and 80 °C for 5 min. The reactions with the random hexamers primers were incubated at 25 °C for 10 min, 50 °C for 50 min and 80 °C for 5 min. The cDNA was stored in a - 20 °C freezer until required.

If the cDNA is being employed for quantitative PCR (qPCR) a no reverse transcriptase control (NRT) is required to check the RNA for residual genomic contamination. This sample is prepared in a similar manner to the cDNA samples, apart from the reverse transcriptase (SuperScript III/RNaseOUT Enzyme Mix) was left out of the reverse transcription step.

2.2.9. Bioinformatic tools used for primer design

Ensembl (www.ensembl.org/index.html) is an online genome browser that annotates genes, computes multiple alignments, and predicts regulatory functions. Primers were designed based on Ensembl's annotation for the gene of interest.

NCBI Primer Basic Local Alignment Search Tool (Blast) (www.ncbi.nlm.nih.gov/tools/primer-blast/) was used to determine the chances of the primers binding and amplifying other areas of the genome/transcriptome. It was also used to check both primers had similar GC% and melting temperatures.

Sigma's Oligo Evaluator tool (http://www.oligoevaluator.com) / **IDT's OligoAnalyzer** (https://eu.idtdna.com/calc/analyzer) was employed to access the GC%, melting temperature, self-complementarity and secondary structure of the primers.

NEB's Tm Calculator (https://tmcalculator.neb.com/#!/main) was used to estimate an appropriate annealing temperature for the primers when using Q5 High-Fidelity DNA polymerase.

2.2.10. Genomic contamination check on newly synthesised cDNA

A PCR assay that targets an intron flanking region of the human methylenetetrahydrofolate dehydrogenase, cyclohydrolase and formyltetrahydrofolate synthetase 1 (*MTHFD1)* gene was used to check for genomic DNA (gDNA) contamination in the newly synthesised cDNA. Primers were sourced from Sigma and sequences are shown in Table 2.2. The PCR reactions were set up as follows: 2.5 µL of 10 X PCR buffer without $MgCl_2$, 1.5 µL of 25 mM $MgCl_2$, 0.5 µL of 10 mM dNTPs, 1.25 µL of 10 µM for each primer, 1 U of Taq DNA Polymerase, and 1 µL of cDNA in a final volume of 25 µL. A no-template control (NTC) was also set up omitting template DNA from the reaction. The reactions were left at the following conditions in a Thermal Cycler (Applied Biosystems GeneAmp 9700 PCR System): 95 °C for 3 min for the pre-incubation stage; 94 °C for 30 sec, 58° C for 1 min and 72 °C for 1 min for the amplification stage for 35 cycles and 72 °C for 10 min for the final extension stage. PCR products (10 µL) were mixed with 3 µL purple DNA gel loading dye. A 100 bp DNA ladder and the PCR products mixed with loading dye were run on a 1% (w/v) agarose gel, as per Section 2.2.7. A band at 330 bp revealed gDNA whereas a band at 232 bp represented cDNA.

Table 2.2 Primers used for genomic contamination check

Assay	Primer	Name	Sequence
MTHFD1 R653Q	Forward	FP MTHFD1 GenCont	5' CACTCCAGTGTTTGTCCATG 3'
	Reverse	RP MTHFD1 GenCont	5' GCATCTTGAGAGCCCTGAC 3'

2.2.11. Investigative polymerase chain reactions (PCR)

2.2.11.1. DHFR2_202 PCR assays

2.2.11.1.1. DHFR2_202 exon 1 assay

Primers were purchased from Sigma Aldrich, with primer sequences found in Table 2.3. The PCR reactions were set up as follows: 5 µL of 10X PCR buffer without $MgCl_2$, 2 µL of 25mM $MgCl_2$, 1.25 µL of 10 mM dNTPs, 1 µL of 10 µM forward primer, 2 µL of 10 µM reverse primer, 2.5 U of Taq DNA Polymerase, and 3 µL of cDNA (17158-Coriell, HepG2, HEK293 and undifferentiated iPS) in a final volume of 50 µL. A positive PCR control was prepared similarly, except 1 pg of the DHFR2_202 synthetic gBlock was used as template DNA (sourced from IDT, see Appendix A for the sequence). A NTC was also set up omitting template DNA from the reaction. The reactions were left at the following conditions in a Thermal Cycler (Applied Biosystems GeneAmp 9700 PCR System): 94 °C

for 5 min for the pre-incubation stage; 94 °C for 30 sec, 65 °C for 2 min and 72 °C for 3 min for the amplification stage for 35 cycles and 72 °C for 7 min for the final extension stage. A 100 bp DNA ladder and the PCR products (10 µL) mixed with purple DNA gel loading dye (3 µL) were run on a 1% (w/v) agarose gel, according to Section 2.2.7.

2.2.11.1.2. DHFR2_202 exon 1 -2 assay

Primers were purchased from Sigma Aldrich, with primer sequences found in Table 2.3. Exon 1-2 assay was set up as follows: 5 µL of 10X PCR buffer without $MgCl_2$, 2 µL of 25 mM $MgCl_2$, 0.75 µL of 10 mM dNTPs, 1 µL of 10 µM forward primer, 2 µL of 10 µM reverse primer, 2.5 U of Taq DNA Polymerase, and 3 µL of cDNA (17158-Coriell, HepG2, HEK293 and undifferentiated iPS) in a final volume of 50 µL. A positive (1 pg DHFR2_202 synthetic gBlock) and NTC were also set up for this assay. The reaction settings were the same as the settings in Section 2.2.11.1.1 apart from the annealing temperature being changed to 61 °C. A 100 bp DNA ladder and the PCR products (10 µL) mixed with purple DNA gel loading dye (3 µL) were run on a 1% (w/v) agarose gel, according to Section 2.2.7.

Table 2.3 Primers used for DHFR2_202 RT-PCR assays

Assay	Primer	Name	Sequence
DHFR2_202 exon 1	Forward	FP Exon 1	5' ATCCCAGCACCCTCCCTCGCCA 3'
	Reverse	RP Exon 2	5' CTCCGCGTCCTGGGAAGTATC 3'
DHFR2_202 exon 1-2	Forward	FP Exon 1-2	5' CATCCACTTTCATTCTCACGTTTC 3'
	Reverse	RP Exon 1-2	5' GAAGAAGTTGTGGTCATTCTCTG 3'

2.2.11.2. A to I DHFR2 PCR assays

2.2.11.2.1. A to I DHFR2 gDNA reference PCR assay

Primers were designed using the tools described in Section 2.2.9 and purchased from IDT, with primer sequences found in Table 2.4. PCR reactions were set up as follows: 10 µL of Q5 reaction buffer, 1 µL of 10 mM dNTPs, 2.5 µL of 10 µM forward primer, 2.5 µL of 10 µM reverse primer, 1 U of Q5 High-Fidelity DNA Polymerase and 10 ng of gDNA (HepG2) in a final volume of 50 µL. A NTC was also set up omitting template DNA from the reaction. The reactions were left at the following conditions in a Thermal Cycler (Applied Biosystems GeneAmp 9700 PCR System): 98 °C for 30 sec

for the pre-incubation stage; 98 °C for 10 sec, 66 °C for 30 sec and 72 °C for 30 sec for the amplification stage for 30 cycles and 72 °C for 5 min for the final extension stage. A 1 kb DNA ladder and the PCR products (7 µL) mixed with purple DNA gel loading dye (3 µL) were run on a 1% (w/v) agarose gel, according to Section 2.2.7.

2.2.11.2.2. A to I DHFR2 cDNA PCR assays

Primers were designed using the tools described in Section 2.2.9 and purchased from IDT, with primer sequences found in Table 2.4. PCR reactions were set up as follows: 10 µL of Q5 reaction buffer, 1 µL of 10 mM dNTPs, 2.5 µL of 10 µM forward primer, 2.5 µL of 10 µM reverse primer, 1 U of Q5 High-Fidelity DNA Polymerase, and 2 µL of cDNA (HepG2) in a final volume of 50 µL. A NTC was also set up omitting template DNA from the reaction. The reactions were left at the following conditions in a Thermal Cycler: 98 °C for 30 sec for the pre-incubation stage; 98 °C for 10 sec, 66 °C for 30 sec and 72 °C for 30 sec for the amplification stage for 30 cycles and 72 °C for 5 min for the final extension stage. A 100 bp DNA ladder and the PCR products (7 µL) mixed with purple DNA gel loading dye (3 µL) were run on a 1% (w/v) agarose gel, according to Section 2.2.7.

Table 2.4 Primers used for A to I DHFR2 assays

Assay	Primer	Name	Sequence
A to I DHFR2 gDNA Assay	Forward	FP DHFR2 gDNA AtoI	5’ TAACCCCCACTCCTTGGTT 3’
	Reverse	RP DHFR2 gDNA AtoI	5’ GCTAAGCAGTCAGTCACCAC 3’
A to I DHFR2 cDNA Assay	Forward	FP DHFR2 cDNA AtoI	5’ GTTTGACTTTGGTTTTGGGC 3’
	Reverse	RP DHFR2 cDNA AtoI	5’ CAAGAAAGGGTACATGTAACCC 3’

2.2.12. PCR product clean-up for Sanger Sequencing

2.2.12.1. ExoSAP Clean-up

PCR reactions with one PCR product were purified as follows: 14 µL ExoSAP-IT Express PCR Product Clean-up was added to 35 µL of PCR product. The mixture was incubated in a Thermal Cycler (Applied Biosystems GeneAmp 9700 PCR System) at 37 °C for 4 min followed by 80 °C for 1 min.

2.2.12.2. Gel Extraction and Purification

PCR products with multiple products were purified using a gel extraction purification kit. Four 50 µL PCR reactions with the same template DNA were pooled together and mixed with 40 µL purple DNA gel loading dye. A 100 bp DNA ladder was loaded into the first well and the PCR product mixed with loading dye was subsequently loaded into a large well of a 2% (w/v) agarose gel and electrophoresed as per Section 2.2.7. The PCR band(s) of interest were extracted from the gel using a scalpel and subsequently purified using the Macherey-Nagel NucleoSpin Gel and PCR Clean-up Kit, according to the manufacturer's protocol.

2.2.13. Blunt-end cloning of PCR product into a cloning vector for Sanger Sequencing

PCR products were ligated into a pJET1.2/blunt cloning vector using the CloneJET PCR Cloning Kit from Thermo Fisher Scientific. The sticky ends were removed from the PCR product by setting up the following reaction: 10 µL of reaction buffer 1, 1 µL of unpurified PCR product, 1 µL of Blunting Enzyme and made up to 18 µL with water. The mixture was briefly vortexed and incubated at 70 °C for 5 min. The PCR product was ligated into the cloning vector by adding 50 ng of the pJET1.2/blunt cloning vector and 5 U of T4 DNA Ligase to the mixture. The mixture was briefly vortexed and incubated at 22 °C for 5 min.

2.2.14. Cloning vector transformed into competent *E. coli* cells

The newly formed plasmids from Section 2.2.13 were transformed into XL10-Gold Ultra-competent cells by initially adding 2 µL of the ligation mixture into 50 µL of the cells and placing on ice for 30 min. The cells were incubated at 42 °C for 30 sec and then immediately put on ice for 2 min. S.O.C. media (250 µL) was added to the cells and incubated at 37 °C for 1 hour in a shaking incubator at 225 rpm. Three different volumes (50 µL, 100 µL and 150 µL) of the transformed cells were spread onto separate LB agar plates containing 100 µg/mL ampicillin. The plates were inverted and incubated overnight at 37 °C.

2.2.15. Isolation and purification of vector for Sanger Sequencing

Single colonies were inoculated in 3 mL of LB broth containing 100 µg/mL ampicillin and incubated overnight at 37 °C while shaking at 230 rpm. The vector was recovered using the GeneJET Plasmid

Miniprep Kit from Thermo Fisher Scientific (all solutions/buffers provided in the kit). The overnight culture was centrifuged at 2,500 x *g* for 4 min. The cell pellet was resuspended in 250 µL of the resuspension solution (inverted 6 times), 250 µL of the lysis solution (inverted 6 times) and 350 µL of the neutralization solution (inverted 6 times). The solution was centrifuged at 12,000 x *g f*or 5 min. The supernatant was transferred into the spin column and centrifuged at 12,000 x *g f*or 1 min, the flow-through was discarded. Wash solution (500 µL) was added to the column, centrifuged at 12,000 x *g* for 1 min and the supernatant was discarded. The wash step was repeated. The column was centrifuged at 12,000 x *g* for 1 min to dry the membrane and then transferred into a recovery tube. Elution buffer (30 µL) was placed on the centre of the column and left at room temperature for 2 min. To recover the plasmid DNA from the spin column it was centrifuged at 12,000 *x g* for 2 min. The DNA was stored in a - 20 °C freezer until required.

2.2.16. Sanger Sequencing of purified PCR products and purified plasmids

Purified PCR products/plasmids and primers were prepared for Sanger Sequencing according to the company guidelines. They were sequenced by Source Bioscience (Riverstown 5 Complex, Riverstown, Industrial Estate, Tramore, Co. Waterford) or Eurofins Genomics Europe Sequencing Gmbh (European Genome and Diagnostics Centre, Jakob-Stadler-Platz 7, 78467 Konstanz, Germany). The results were analysed using the tools described in Section 2.2.17.

2.2.17. Bioinformatic tools used for analysing Sanger Sequencing results

Geneious Prime (www.geneious.com/academic/) **/ Snapgene** (https://www.snapgene.com/): are tools used to visualise biological data, they were employed to analyse the DNA chromatogram from Sanger Sequencing of the PCR products and cloning vectors.

Clustal Omega (www.ebi.ac.uk/Tools/msa/clustalo/) is a multiple sequence alignment program that was used to compare the Sanger Sequencing results back to the reference sequence obtained from Ensembl.

BLAST (https://blast.ncbi.nlm.nih.gov) finds areas that are comparable between the input sequence and sequences already on the databases; it subsequently measures the statistical significance of the alignment. This tool was used to identify the origin of the Sanger Sequences if it did not perfectly match back to the reference sequence.

2.2.18. Quantitative polymerase chain reaction (qPCR)

2.2.18.1. qPCR analysis on ribosome bound and free cDNA HepG2 fractions

Probe-based qPCR assays were designed for the GAPDH (ribosome-bound RNA control), NEAT1 (ribosome free RNA control), DHFR (dihydrofolate reductase gene family control), DHFR2_All, DHFR2_201 and DHFR2_202 transcripts using Roche's ProbeFinder software (https://lifescience.roche .com/global_en/brands/universal-probe- library.html). The NICB Primer Blast tool was also employed to check the primers did not bind and amplify other areas of the genome, as per Section 2.2.9. Primer and probe designs are in Table 2.5. Both products were sourced from Sigma.

The NRT control samples for the whole lysate control, supernatant and pellet were initially examined to determine did the RNA contain any residual genomic contamination. The NRT controls were used as the template DNA in all the qPCR assays that will be used in the investigation. They were set up in a final reaction volume of 20 µL, made up of 10 µL FastStart Essential DNA Probes Master, 0.4 µL of 10 µM of probe (apart from DHFR2_All assay (0.6 µL of 10 µM)), 0.6 µL of 10 µM of each primer (apart from GAPDH assay (1 µL of 10 µM) and DHFR2_All assay (1.2 µL of 10 µM)), 1 µL of template DNA (NRT control – equivalent of 50 ng cDNA) and molecular grade water. The qPCR reactions were prepared in triplicate. The reactions were loaded onto the white version of the LightCycler 480 Multiwell Plate 96. The qPCR runs were carried out on the LightCycler 96 system (Roche) under the following conditions: 95 °C for 10 min for the pre-cycling step; followed by 45 cycles of 95 °C for 30 sec, 60 °C for 30 sec and 72 °C for 30 sec followed by fluorescence acquisition (FAM ƛex = 470 nm, ƛem =514 nm) and finally at 40 °C for 10 sec for the cooling step. No amplification on the Light Cycler indicates no residual genomic contamination within that RNA sample.

Once the quality control checks were completed, the investigative RT-qPCR assays were performed. A NTC, whole lysate control, supernatant and pellet reaction was set up for the GAPDH, NEAT1, DHFR, DHFR2_All, DHFR2_201 and DHFR2_202 assays according to the previous paragraph. The template DNA was molecular grade water for the NTC and 50 ng of cDNA for the supernatant sample, pellet sample and whole lysate control. The Light Cycler conditions were the same as the previous paragraph. Relative abundance of the target in the pellet and supernatant reactions were calculated using the LightCycler 96 qPCR System Instrument Software, Roche. The expression ratios were calculated relative to the target in the whole lysate control sample. The relative ratios of the target in the pellet and supernatant were plotted on a bar chart using Microsoft excel. The ratio errors calculated by the LightCycler 96 qPCR System Instrument Software (Roche) were added to

the bar charts as errors bars. The fold difference of the target in the pellet and supernatant was calculated by dividing the larger relative ratio by the smaller relative ratio.

Table 2.5 Primers and probes used in RT-qPCR assay on ribosome bound and free HepG2 fraction

Assay	Primer	Name	Sequence	Probe #*
GAPDH	Forward	FP GAPDH	5' CTCTGCTCCTCCTGTTCGAC 3'	60
	Reverse	RP GAPDH	5' ACGACCAAATCCGTTGACTC 3'	
NEAT1	Forward	FP NEAT1	5' AGTGAATGTGCACCCTTGG 3'	46
	Reverse	RP NEAT1	5' AACAAACCACGGTCCATGA 3'	
DHFR	Forward	FP DHFR	5' GAACTCGTGACCGCAAGC 3'	75
	Reverse	RP DHFR	5' GGCTCAAGCCGGTAATCC 3'	
DHFR2_All	Forward	FP DHFR2_All	5' AATTTCGCGGCATTCTTG 3'	72
	Reverse	RP DHFR2_All	5' GGTTAACACCTCCGAACTTGC 3'	
DHFR2_201	Forward	FP DHFR2_201	5' CGGACCTTAGAAAGTCACACATC 3'	89
	Reverse	RP DHFR2_201	5' GCGAAATTCCCTTCTTCAAAT 3'	
DHFR2_202	Forward	FP DHFR2_202	5' CGTCCAGAAGCGTCTCATTC 3'	57
	Reverse	RP DHFR2_202	5' AAGCTCTCAGCGGGACAAT 3'	

2.2.19. Protein extraction from cell lines and tissues

2.2.19.1. Reducing frozen tissue samples into a fine powder for protein extraction

Frozen tissue was initially chopped into smaller pieces using a pre-chilled and sterilised PVC pipe cutter. The tissue was wrapped in prechilled aluminium foil and placed on dry ice. A sterile grinder (Krups F203 Coffee and Spice grinder) was pre-chilled by grinding several pieces of dry ice in it. The smaller tissue chunks were placed into the chilled grinder and ground down to a fine powder. The frozen powder was quickly transferred into a pre-cooled 1.5 mL tube using a chilled spatula, approximately to the 50 µL mark. The powdered tissue was stored in a - 80 °C freezer until required.

2.2.19.2. Protein extraction from ground tissue and cultured cells

The cells were sheared by homogenising 50 µL of powdered tissue or $5x10^6$ cells in 500 µL of protein lysis buffer (DCU). They were incubated at room temperature for 1 hour on a rotator. The lysate was centrifuged at 18,300 x *g* for 15 min at 4 °C and the supernatant was retained. A Bradford assay

was employed to calculate the protein concentration of each sample. The BSA (bovine serum albumin) standard curve and protein samples were prepared in triplicate according to the manufacturer's instructions and loaded onto a flat-bottom 96-well plate (Greiner Bio-One). The absorbance was read on the Tecan Infinite 200 PRO plate reader at a wavelength of 595 nm. The concentration of each protein sample was quantified by subtracting the background noise from the absorbance and substituting it into the equation of the line generated from the BSA standard curve. The protein lysate was diluted to 1 µg/µL in freshly prepared ammonium bicarbonate solution and 100 µL aliquots were prepared. The samples were stored in a - 80 °C freezer until required.

2.2.20. Sample preparation for liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis

2.2.20.1. Filter aided sample preparation (FASP) and protein digestion

Protein lysate (100 µg) from Section 2.2.19.2 was initially reduced by adding 1 µL of DTT solution and incubated at 56 °C for 20 min. After leaving the samples to cool for 10 min, 200 µL of urea solution was added and the sample was transferred into the filter device of the Microcon-10 kDa Centrifugal Filter Unit. The unit was centrifuged at 14,000 *x g* for 40 min at 20 °C and the flow-through was discarded. The membrane-bound proteins were washed by adding 200 µL of Wash solution 1 to the device and centrifuged at 14,000 *x g* for 40 min at 20 °C. The flow-through was discarded. The sample was alkylated by adding 100 µL of wash alkylating solution to the filter and vortexed at 600 rpm for 1 min. The unit was incubated in a dark room for 15 min at room temperature and subsequently centrifuged at 14,000 *x g* for 40 min at 20 °C. The flow-through was discarded. The sample was washed by adding 100 µL of Wash solution 2 to the device and centrifuged at 14,000 *x g* for 40 min at 20 °C. The flow-through was discarded. This wash step was repeated. The filter device was transferred into a new filtrate tube. The membrane-bound protein was enzymatically digested by adding 100 µL of fresh ammonium bicarbonate solution, 2 µL Sequence Grade Trypsin solution (1 µg/µL) and 1 µL ProteaseMAX Surfactant solution to the filter unit. The lids were parafilmed closed and incubated overnight at 37 °C at 300 rpm in a shaking incubator. The unit was centrifuged at 14,000 *x g* for 40 min at 20 °C to obtain the tryptic peptides. To elute any remaining peptides, 50 µL of rinse solution A was added to the membrane and centrifuged at 14,000 *x g* for 20 min at 20 °C. The filter unit was discarded. The digestion was halted by adding 2% (v/v) acetonitrile and 0.1% (v/v) trifluoracetic to the tryptic peptide solution. The samples were stored in a - 80 °C freezer until required.

2.2.20.2. Protein purification of enzymatically digested peptides

The enzymatically digested peptides from Section 2.2.20.1 were purified for LC-MS/MS analysis using the C18 Spin Column system (Thermo Fisher Scientific). Activation solution (200 µL) was added to the spin column and centrifuged at 1,500 *x g* for 1 min at 20 °C. The flow-through was discarded and this step was repeated. Equilibrium solution (200 µL) was added to the column and centrifuged at 1,500 *x g* for 1 min at 20 °C. The flow-through was discarded and this step was repeated. The enzymatically digested peptides from Section 2.2.20.1 were loaded onto the spin column and centrifuged at 1,500 x *g* for 1 min at 20 °C. The flow-through was retrieved and this step was repeated twice. After discarding the flow-through, the resin-bound peptides were washed by adding 200 µL of rinse solution B to the column and centrifuged at 1,500 x *g* for 1 min at 20 °C. The flow-through was discarded and this step was repeated. Elution solution (20 µL) was added to the spin column and centrifuged at 1,500 x *g* for 1 min at 20 °C to obtain the purified tryptic peptide. This elution step was repeated. The spin-column was discarded. The samples were dried using a SpeedVac vacuum concentrator (Savant DNA110) under the following conditions: mode/vent of vacuum-high vapor, vacuum level of 13 metric millibar, speed of 17 x *g*, drying rate of medium (ambient temperature) and time range of 2-5 hour until the solution is completely evaporated. The dried samples were stored in a - 80 °C freezer until required.

2.2.21. LC-MS/MS analysis

LC-MS/MS analysis was carried out in the Proteomics Facility in the National Institute for Cellular Biology (NICB) in DCU, if not stated otherwise. LC-MS/MS analysis consisted of reverse-phased capillary high-pressure liquid chromatography on the UltiMate 3000 RSLCnano System (Thermo Scientific) in line with an Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Scientific).

The lyophilised samples from Section 2.2.20.2 were resuspended in 15 µL of resuspension solution and then sonicated using a sonicating waterbath (Precision ultrasonic cleaning, DP201–00, Ultrawave) under the following conditions: power of 280 watts, frequency of 30 kilohertz, ambient temperature and time of 5 min. The peptide concentration was quantified using the Nanodrop One spectrophotometer (Labtech instruments), at an absorbance of 280 nm. The samples were diluted down to a working solution of 1 µg/µL using resuspension solution for data-dependent acquisition (DDA) and parallel reaction monitoring (PRM) LC-MS/MS analysis.

2.2.21.1. DDA LC-MS/MS analysis

One microlitre of the working solution of the digested protein samples from Section 2.2.21 was loaded onto a trapping column cartridge (PepMap100, C18, 300 µm x 5 mm) (Thermo Scientific) using 2% (v/v) acetonitrile, 0.1% (v/v) trifluoracetic acid at a flow rate of 25 µL/min for 3 min. The peptides were then resolved on AcclaimPepMap100 75 µm x 50 cm, nanoViper C18, 3 µm, 100 Å (Thermo Fisher Scientific) where a binary gradient was employed; solvent A and 2–25% solvent B for 100 min followed by solvent A and 25–50% solvent B for a further 20 min. The column flow rate was 300 nL/min and the column oven temperature was 35 °C. The Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Scientific) was set to data-dependent acquisition using a top speed approach (cycle time of 3 sec) with a voltage of 2.0 kV and a capillary temperature of 320 °C. Full scans within the 380-1500 *m/z* range were carried out in the Orbitrap mass analyser using a resolution of 120,000 (at *m/z* 200), automatic gain control (AGC) target value of $4x10^5$ and a maximum ion injection time of 50 ms. The top-speed acquisition algorithm was used to identify the number of selected precursor ions for fragmentation. The Quadrupole selected the precursor ions for fragmentation with an isolation width of 1.6 Da. The peptides with a charge state between 2 + to 6 + were examined and a dynamic exclusion was applied after 60 sec. Precursor ions were fragmented using higher energy collision-induced dissociation with a normalised collision energy of 28%. The linear ion trap measured the resulting MS/MS ions under the following scan conditions: AGC target value of $2x10^4$ and maximum fill time of 35 ms. An MS/MS output file was generated, uploaded onto the ProteomeDiscoverer 2.2 software (Thermo Fisher Scientific) and compared with a reference data set from UniProtKB (taxonomy: *Homo sapiens* 26,592 sequences and *Mus musculus* 17,556 sequences) to identify the peptides in the sample. The following search parameters were used for peptide identification: peptide mass tolerance of 20 ppm, MS/MS mass tolerance of 0.6 Da, two missed cleavages were allowed, carbamidomethylation of cysteine was defined as a fixed modification, and methionine oxidation defined as a variable modification. Only highly confident peptide identifications with a false discovery rate (FDR) ≤ 0.01 (identified using a SEQUEST HT workflow coupled with Percolator validation in Proteome Discoverer 2.2) were isolated. Finally, Proteome Discoverer 2.2 generated a report with the peptides identified in the sample.

2.2.21.2. PRM LC-MS/MS analysis

PRM LC-MS/MS analysis was employed to target DHFR2 specific peptides in the cell and tissue samples. Table 2.6 shows the amino acid sequence of the selected DHFR2 peptide and the equivalent DHFR tryptic peptide. There is one amino acid difference between the DHFR2 and DHFR version of this peptide, the DHFR peptide has a proline residue (115 Da) whereas the DHFR2 peptide has a leucine amino acid (131 Da). The mass difference between these amino acids give the DHFR2 and DHFR peptides a significant *m/z* difference, making it relatively straightforward for the mass spectrometer to differentiate between them. An AQUA Basic Light synthetic version of the DHFR2 peptide was purchased from Thermo Fisher Scientific to characterise this peptide and generate a workflow to target the endogenous/recombinant version. The DHFR2 synthetic peptide (0.115 ng) was injected into the LC-MS/MS system and a full MS scan was performed, as per Section 2.2.21.1. The results (Table 2.7) from this scan were used to generate a targeted workflow on the MS instrument. The protein sample would be separated for 60 min in the UltiMate 3000 RSLCnano System. The Quadrupole would select a number of precursor ions for fragmentation with an isolation width of 1.6 Da. If a precursor ion matched the *m/z*, charge state and retention time window from the DHFR2 target list it would be selected for fragmentation using higher energy collision-induced dissociation with a normalised collision energy of 35%. The Orbitrap mass analyser would measure the resulting MS/MS at a resolution of 30,000 (at *m/z* 200) under the following scan conditions: AGC target value of $5x10^4$ and maximum fill time of 300 ms. To confirm this approach could identify the DHFR2 peptide within a complex sample, 2.3 pg of the DHFR2 synthetic peptide was spiked into 1 µg of enzymatically digested protein sample and analysed using the PRM workflow. This two-step optimisation was carried out for each batch of samples to account for the instrument's variability. One microlitre of the working solution of the digested protein samples from Section 2.2.21 was subsequently analysed for the endogenous/recombinant DHFR2 peptide using the optimised PRM workflow.

Table 2.6 The amino acid sequence of the DHFR and DHFR2 tryptic peptide

Name	Peptide Sequence
DHFR Specific Tryptic Peptide	EAMNH**P**GHLK
DHFR2 Specific Tryptic Peptide	EAMNH**L**GHLK

Table 2.7 DHFR2 peptide information used for the PRM LC-MS/MS workflow

Peptide sequence	Modification	Charge	*m/z*	Retention time start (min)	Retention time stop (min)
EAmNHLGHLK	Oxidation	2	583.2934	15	17
EAmNHLGHLK	Oxidation	3	386.1984	15	17
EAMNHLGHLK	-	3	383.8668	18	20
EAMNHLGHLK	-	2	575.2968	18	20

2.2.22. Western Blot analysis

2.2.22.1. Separation of protein on 4-20% Tris-Glycine gel

Protein samples were diluted in NuPAGE LDS Sample buffer (1x) and NuPAGE Sample Reducing Agent (1x), concentrations stated in specific sections. The mixture was heated to 95 °C for 10 min and left to cool before resolving on a Novex 4-20% Tris-Glycine Mini Gels in 1x Tris-Glycine buffer. A 10 µL volume of PageRuler Plus Prestained protein ladder was added to the first well and the samples were added to subsequent wells. The proteins were separated by running the gels at 175 volts until the dye reached the end of the gel.

2.2.22.2. Transfer of proteins onto PVDF membrane and Ponceau staining

The gel from Section 2.2.22.1 was removed from its cassette and soaked in a 1-Step Transfer buffer for 30 min. To activate the PVDF transfer membrane it was incubated in methanol for 30 sec. The PVDF transfer membrane and four pieces of Western Blotting filter paper were equilibrated by soaking them in a 1-Step Transfer buffer for 5 min. The proteins were transferred from the gel onto the PVDF membrane using Thermo Fisher Scientific Pierce Power Blotter. It was assembled on the anode as follows: 2 Western Blotting filter papers, PVDF transfer membrane, 4-20% tris-glycine gel and 2 Western Blotting filter papers. The blotter roller was used to eliminate any trapped bubbles. The cathode was attached to the anode and inserted into the cassette unit. The protein was left to transfer to the PVDF membrane for 10 min. To confirm even protein loading and successful electrotransfer the membrane was soaked in Ponceau stain for 10 min. The membrane was washed with deionised water until protein bands were visible and imaged. The membrane was soaked in 0.1 M sodium hydroxide for 30 sec to remove any residual Ponceau stain.

2.2.22.3. Immunoblotting

The PVDF membrane from Section 2.2.22.2 was blocked in 50 mL of blocking solution for 2 hour at room temperature. The membrane was subsequently incubated in primary antibody diluted in antibody solution (See Table 2.8 for concentrations used) for 2 hour at room temperature. The membrane was washed three times for 5 min each using TBST buffer (0.05% (v/v)). The membrane was then incubated in secondary antibody diluted in antibody solution (See Table 2.8 for concentrations used) overnight at 4 °C. The next day the membrane was washed three times for 5 min each using TBST buffer (0.05% (v/v)).

Table 2.8 Antibody concentrations used for Western Blot analysis

Primary Antibody	Dilution	Secondary Antibody	Dilution
Anti-tGFP antibody	1:2,000	Anti-Rabbit Antibody	1:5,000
Anti-DDK antibody	1:2,000	Anti-Mouse Antibody	1:1,000
Anti-Ubiquitin antibody	1:250	Anti-Mouse Antibody	1:1,000

2.2.22.4. Protein detection using chemiluminescence

The bound target proteins were detected using SuperSignal West Pico PLUS Chemiluminescent Substrate. Equal volumes of Reagent A and Reagent B were incubated with the membrane from Section 2.2.22.3 for 5 min at room temperature. The proteins were visualised and imaged using Syngene's GeneGnome Bio Imaging System.

Chapter 3. Investigating the translation status of the human *DHFR2* gene through its association with the ribosome

3.1. Introduction

3.1.1. Overview

The central dogma of molecular biology states that DNA is transcribed into RNA and subsequently translated into a functional protein (Crick, 1970). However, there are some instances where this principle is not followed, such as retroviruses, prions and noncoding RNAs (Koonin, 2012).

The *DHFR2* gene is transcribed into several RNA isoforms (Ensembl, 2019; McEntee et al., 2011) in human cell lines and tissues (Anderson et al., 2011; McEntee et al., 2011; Protein Atlas, 2021). However, there is no prior evidence that the DHFR2 RNA is translated into an endogenous protein. Therefore, one of the aims of this project was to examine the translational capabilities of the two main DHFR2 transcripts, DHFR2_201 and DHFR2_202 (Figure 1.9). The best approach for investigating the translational status of an RNA molecule is to measure its association with the ribosome. The ribosome plays a vital role in the initiation and elongation stages of translation. For this reason, if an RNA molecule has a high association with the ribosome it should be translated into a protein. In contrast, if an RNA molecule has little to no association with the ribosome it is unlikely that it generates a protein (Dalla Venezia et al., 2019; King & Gerber, 2016). Therefore, this chapter turns its attention to investigating the relationship between the DHFR2 transcripts and the ribosome. Two experimental methodologies were employed to reach this aim. The first approach extracted and assessed ribosome profiling data for the DHFR2 RNA from an online directory known as ribosome protected RNA fragments database version 2.0 beta (RPFdb v2.0 beta). The main disadvantage of using online databases is they consider all RNA isoforms of a gene as one and the detail for the individual transcripts are lost. Therefore, a sucrose cushion ultracentrifugation step coupled with reverse transcriptase - quantitative polymerase chain reaction (RT-qPCR) was experimentally performed to target individual DHFR2 transcripts, DHFR2_201 and DHFR2_202.

3.1.2. The role of the ribosome during translation

The process of translating a messenger RNA (mRNA) molecule into a protein is intricate and tightly regulated within each cell. The ribosome is a supramolecular ribonucleoprotein complex that plays a critical role in this mechanism. A ribosome binds to an mRNA molecule (~20-30 nucleotides) at the ribosome binding site and travels along the mRNA molecule decoding the sequence to synthesise a polypeptide (Dalla Venezia et al., 2019). This unique polypeptide later folds into an active 3D protein in the endoplasmic reticulum (Figure 3.1) (Sharma, 2013). Multiple ribosomes can bind to an mRNA molecule and initiate translation at one time, this structure is otherwise known

as a polysome. Therefore at any given time in a cell, there is a pool of mRNA molecules bound to the ribosome that are being actively translated. At the same time, there will be a group of free mRNA molecules that are not associated with the ribosome and are not being translated into a protein (Pringle et al., 2019). Numerous methodologies have been developed to investigate the translational status of a particular mRNA molecule utilising this information (King & Gerber, 2016).

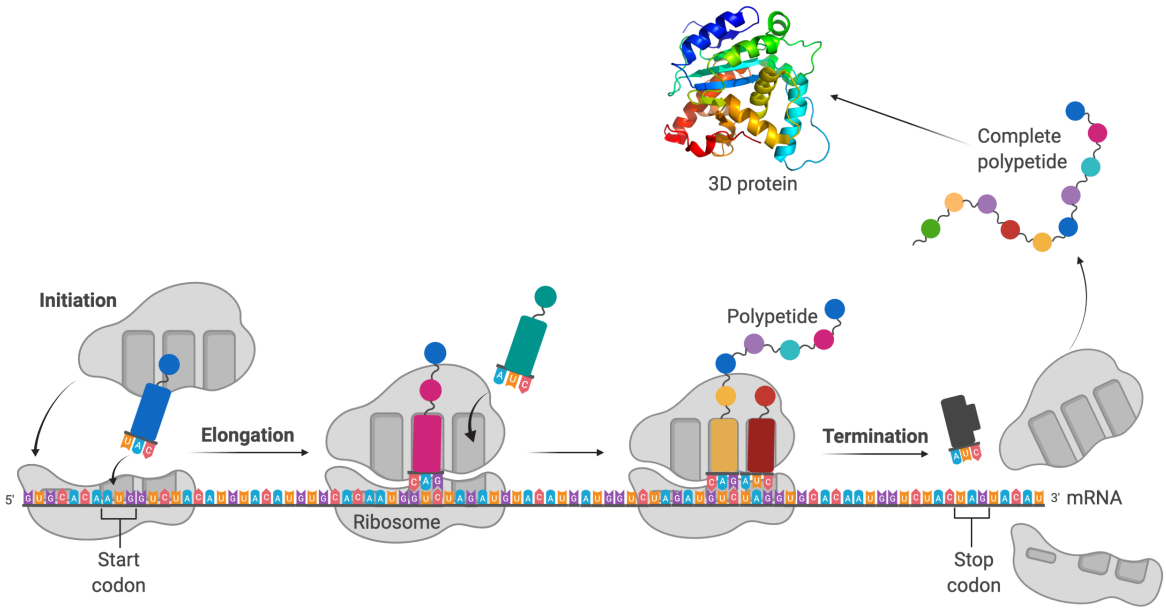


Figure 3.1 Schematic showing the role of the ribosome during translation of an mRNA molecule into a protein. *In eukaryotes, translation is divided into three stages: initiation, elongation and termination. The ribosome binds to the ribosome binding site of mature mRNA and starts to move along the molecule. When the ribosome reaches a start codon it initiates translation. The ribosome travels along the mRNA molecule decoding the sequence to synthesise a polypeptide. Once the ribosome reaches a stop codon it disassociates from the mRNA molecule and the polypeptide is released. This unique polypeptide later folds into an active 3D protein in the endoplasmic reticulum. Created with BioRender.com, Adapted from Dalla Venezia et al., 2019.*

3.1.3. Ribosome profiling as a method to study the translatability of an RNA molecule

Ribosome profiling, also known as Ribo-seq, is by far the most powerful tool used to assess the global translation in a cell at a given point (Ingolia, 2014). This technique provides a list of mRNA molecules that have been found to bind to the ribosome, as well as at what rate this binding has

occurred. Furthermore, it identifies the position on the mRNA molecule where the ribosome was situated. Therefore, this technique can be employed to identify potential novel proteins, short peptides and alternative protein isoforms (Brar & Weissman, 2015).

Ribo-seq is a complex method that comprises of multiple experimental techniques and extensive computational analysis (King & Gerber, 2016). The detailed protocol must be followed to ensure accurate measurements are recorded, see Figure 3.2. Eukaryotic cells in culture are incubated in cycloheximide (CHX) or flash-frozen to halt the actively translating ribosomes (Ingolia et al., 2012). The extracted cell lysate is subject to ribonuclease treatment, typically RNase I, which digests RNA that is not protected by the ribosome. The mixture is loaded onto a sucrose gradient and centrifuged at a high speed. This separates the molecules in the mixture into different layers based on their density. Ribosome protected mRNA fragments (RPFs) will have a much greater density than the free RNA leftover from the ribonuclease treatment. Therefore, RPFs and the remaining free RNA will be in different fractions. Each layer is carefully isolated and the RNA is extracted from the fraction known to have the RPFs (King & Gerber, 2016). This RNA is sequenced using next-generation sequencing technology. The data generated is processed to create a high-quality dataset. This includes eliminating reads that are not of a specific length and reads that match back to ribosomal RNA and transfer RNA molecules. The remaining reads are aligned to the reference genome and the reads per kilobase per million mapped reads (RPKM) are calculated for each gene (Wang et al., 2019), using the following formula (Mortazavi et al., 2008; Wagner et al., 2012):

$$\text{RPKM of a gene} = \frac{\text{Number of reads mapped to a gene} \times 10^{3*} \times 10^{6**}}{\text{Total number of mapped reads from library} \times \text{gene length (bp)}}$$

*10^3 normalises for gene length
**10^6 normalises for sequencing depth factor

Therefore, for Ribo-seq analysis, the RPKM value is the normalised unit that measures the level at which an RNA molecule binds to the ribosome (Wagner et al., 2012).

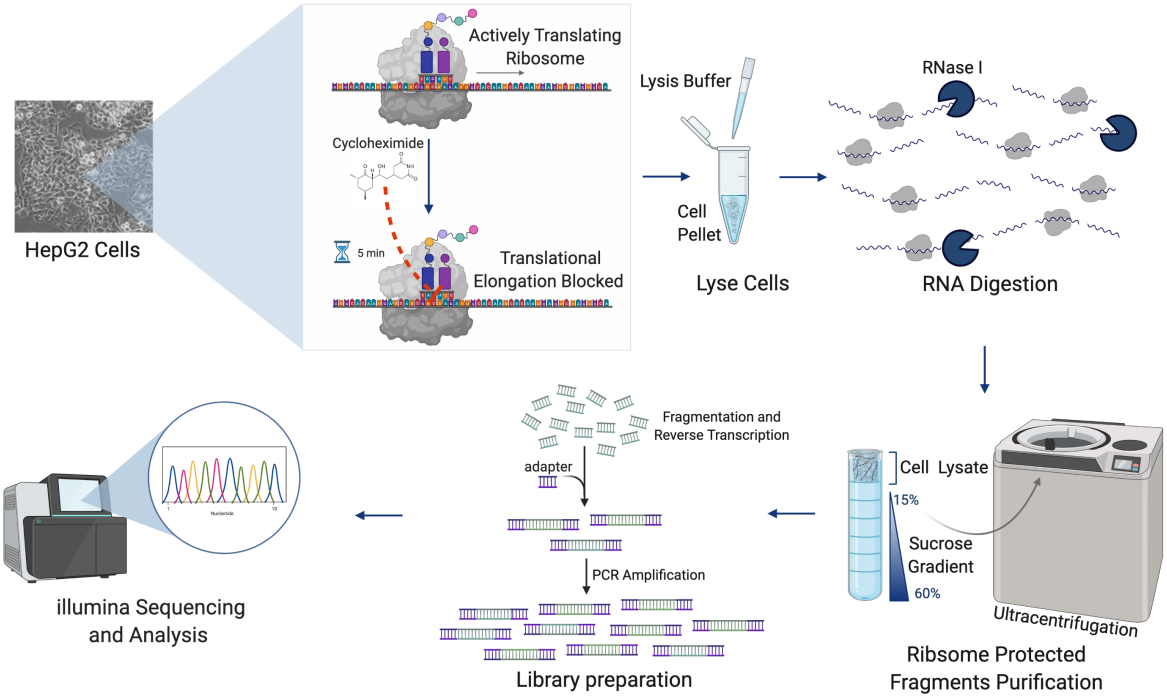


Figure 3.2 Ribosome profiling schematic*. Actively growing cells are incubated in cycloheximide or flash-frozen to halt translation. The cells are harvested and cell lysate is extracted. The lysate is treated with RNase I to digest the RNA that is not protected by the ribosome. The remaining sample is gently loaded onto a sucrose gradient and centrifuged at a high speed. This results in the molecules being separated depending on their density. Each fraction is carefully isolated from the tube and the RNA is extracted from the fraction known to have the ribosome protected mRNA fragment. The RNA is subject to Illumina sequencing. The sequencing results are analysed to determine if the mRNA of interest binds to the ribosome and if so, at what rate. Created with BioRender.com, Adapted from Ingolia et al., 2012.*

There are numerous online resources that were built for depositing and visualising published Ribo-seq data sets. The most established databases are GWIPS-viz, RPFdb and TranslatomeDB (Xie et al., 2016). The latest version of RPFdb, RPFdb v2.0 beta is the largest and most versatile directory for Ribo-seq data. Furthermore, it has an optimised pipeline for processing and normalising the raw data from different Ribo-seq studies (Wang et al., 2019). Therefore, Ribo-seq data from RPFdb v2.0 beta was utilised to assess the translational status of the DHFR2 RNA.

3.1.4. Polysome profiling as a method to study the translatability of an RNA molecule

Polysome profiling is the traditional gold standard method for analysing the translation of mRNA in a cell at a given time. This technique also identifies the rate of translation of an mRNA molecule by the degree of its association with the ribosomes. However, unlike ribosome profiling, it does not identify the position on the mRNA molecule where the ribosome was situated (Chassé et al., 2017). Actively translating mRNA will have a greater number of ribosomes attached which makes them denser. Whereas poorly translating or non-translating mRNA will have very little or no ribosomes associated making these molecules lighter in weight (Zarai et al., 2016). Polysome analysis utilises this information to separate out cellular mRNAs based on their densities (Panda et al., 2017).

Polysome profiling is also a complex method that requires highly trained scientists and specialised equipment (King & Gerber, 2016). Again, cycloheximide is used to treat actively growing cells, which halts translation and stops the translocation of the ribosomes from the mRNA molecules (Pringle et al., 2019). The extracted cell lysate is loaded onto a linear sucrose gradient and centrifuged at a high speed. This step separates the cellular mRNA into layers based on their density or in other words the number of ribosomes associated. The top layers from the sucrose gradient will contain the non-translating/poorly translating mRNA whereas the bottom layers will consist of the actively translating mRNA (Poria & Ray, 2017). Each layer is carefully separated into fractions using a flow cell in line with a spectrophotometer (Mehta et al., 2012). A polysome profile is generated during the isolation process to confirm optimal separation. The RNA is isolated from each fraction and RT-qPCR is used to determine which fraction the target mRNA is most abundant in. Ultimately identifying the rate of translation of that gene in that particular cell or tissue (Panda et al., 2017).

3.1.5. Sucrose cushion ultracentrifugation RT-qPCR as a method to study the translatability of an RNA molecule

The polysome profiling technique requires a specific set of skills and equipment to be performed, which may not be available to less specialised laboratories (King & Gerber, 2016). A simpler technique known as a sucrose cushion ultracentrifugation RT-qPCR can be employed instead (Rivera et al., 2015). This method uses the same principle as the polysome profiling method but does not reveal the rate at which the mRNA molecule is translated. It can only predict if an mRNA molecule is actively translating or non-translating/poorly translating (Mehta et al., 2012). The protocol for this technique is similar to the polysome profiling method, apart from the gradient element of the sucrose separation step (Figure 3.3). The cycloheximide treated cell lysate is loaded onto a sucrose

cushion and centrifuged at a high speed. This results in a supernatant that has the ribosome-free RNA and a pellet that contains the ribosome-bound mRNA (Mehta et al., 2012). This technique was employed to determine the translational status of the two main DHFR2 transcripts, DHFR2_201 and DHFR2_202.

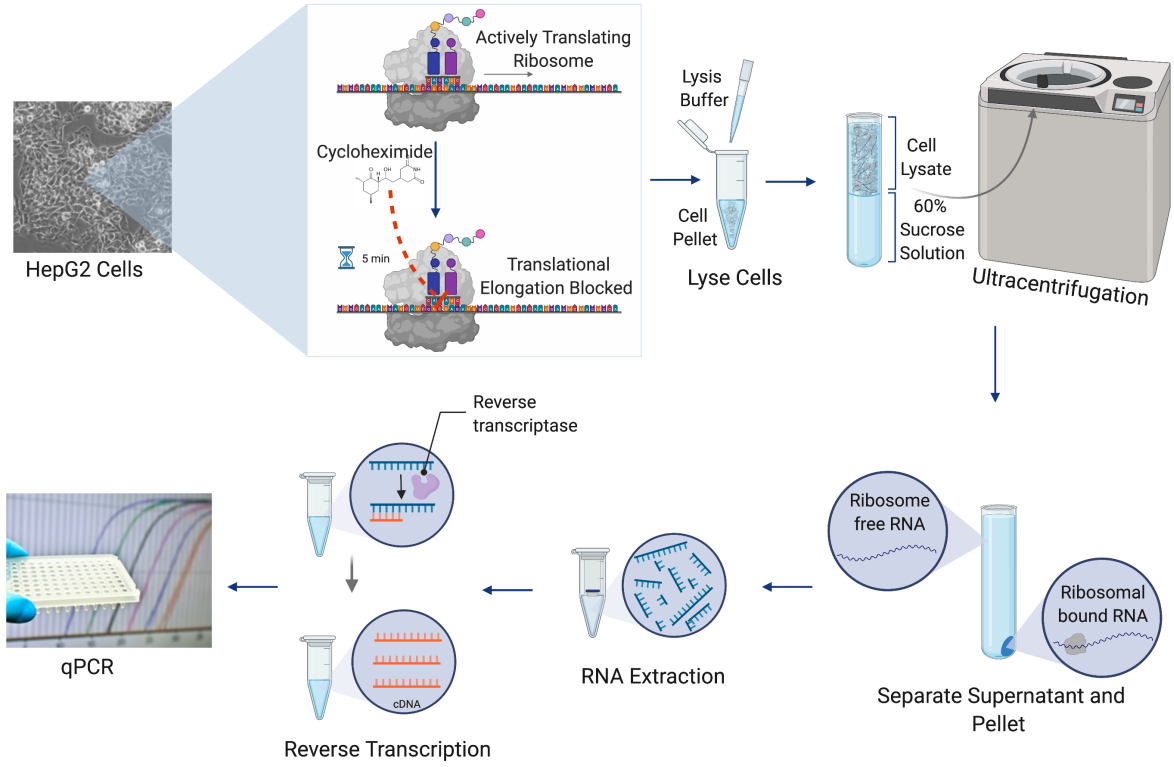


Figure 3.3 A schematic diagram of a sucrose cushion ultracentrifugation coupled with RT-qPCR. *Actively growing cells are treated with cycloheximide to halt the ribosome and ultimately inhibit translation elongation. The cells are harvested and cell lysate is extracted. The lysate is gently loaded onto a 60% (w/v) sucrose solution and centrifuged at a high speed. This results in a supernatant that has ribosome-free RNA and a pellet that contains ribosome-bound mRNA. The supernatant and pellet are isolated and RNA is extracted from each fraction. RT-qPCR is used to determine which fraction contains a higher abundance of the mRNA of interest. Created with BioRender.com, Adapted from Mehta et al., 2012; Rivera et al., 2015.*

3.1.6. Aims and objectives

The main aim of this chapter was to investigate the translational capabilities of the DHFR2 RNA, more specifically the DHFR2_201 and DHFR2_202 transcripts. To examine this, the relationship between the DHFR2 RNA and the ribosome was explored. To achieve this aim, the following objectives were proposed:

1. Analyse Ribo-seq data available on RPFdb v2.0 beta that is specific to the *DHFR2* gene.
2. Determine if Ensembl predicted the DHFR2_202 transcript correctly by targeting it in several cell types by reverse transcriptase-polymerase chain reaction (RT-PCR). This objective is required to carry out the sucrose cushion ultracentrifugation RT-qPCR experiment.
3. Perform a sucrose cushion ultracentrifugation RT-qPCR experiment on HepG2 cells and assess the data that targets the DHFR2_201 and DHFR2_202 transcripts.

3.2. Method

3.2.1. Examining the relationship between the DHFR2 RNA and the ribosome through bioinformatic analysis of Ribo-seq datasets

Ribosome profiling data from hundreds of different studies are available on the RPFdb v2.0 beta database (http://sysbio.gzzoc.com/rpfdb/). These studies were utilised to assess the relationship between the DHFR2 RNA and the ribosome. A criterion was generated for choosing samples for this experiment: human origin, wild-type cell lines with no treatments, normal tissue and/or cancerous tissue. Taking this criterion into account, 23 samples were chosen to investigate out of the thousands available on the directory. They included the following human cell lines: A549 (adenocarcinomic alveolar basal epithelial), CN34 (pleural fluid from breast carcinoma), GM19204 (B-Lymphocyte), GM19238 (B-Lymphocyte), HCT116 (colorectal carcinoma), HEK293 (embryonic kidney), HEK293T (embryonic kidney), HeLa (cervical carcinoma), ESC (embryonic stem cells), MCF10A (mammary epithelial), MDA (mammary epithelial carcinoma), MM1S (myeloma), NCCIT (pluripotent embryonal carcinoma), NPC (neural progenitor), PC3 (prostate carcinoma), SHSY5Y (neuroblastoma), SUM159 (mesenchymal triple-negative breast cancer), THP1 (acute monocytic leukaemia) and fibroblasts. Plus the following human tissues: brain normal, brain cancerous, kidney normal and kidney cancerous. See Appendix B for more information. The Ribo-seq data belonging to these 23 samples were downloaded from the RPFdb v2.0 beta platform as a CSV file. Within each

file, there was a list of genes that were found to be protected by the ribosome in that sample. For each of the genes identified in the sample, it also had the normalised RPKM value (described in Section 3.1.3) for the entire RNA molecule, 5' untranslated region (UTR), coding sequence (CDS) and 3'UTR.

The RPKM values for the glyceraldehyde 3-phosphate dehydrogenase (*GAPDH),* DNA polymerase subunit gamma (*POLG*)*,* methylenetetrahydrofolate dehydrogenase, cyclohydrolase and formyltetrahydrofolate synthetase 1L *(MTHFD1L),* thymidylate synthetase *(TYMS), DHFR* and *DHFR2* genes were extracted from the 23 samples using RStudio (https://www.rstudio.com/). The RPKM values for the *GAPDH* and *POLG* genes were examined first. These genes were used as ubiquitously expressed controls, with POLG specific to the mitochondria. Any sample that had outlier RPKM values for the *GAPDH* and *POLG* gene were excluded from subsequent analyses. The RPKM values for the *MTHFD1, TYMS* and *DHFR* genes were subsequently examined. These genes were used as one-carbon metabolism (OCM) controls. More specifically, the *TYMS* gene was the de novo thymidylate biosynthesis pathway control, the M*THFD1L* gene was used as the OCM mitochondrial control and the *DHFR* gene was the dihydrofolate reductase gene family control. Once the analysis of the control genes was complete, the RPKM values for the *DHFR2* gene were analysed to determine the relationship between the DHFR2 RNA and the ribosome. Microsoft Excel was used to generate bar charts illustrating the Ribo-seq results for the six genes.

3.2.2. Investigation of the DHFR2_202 transcript in four cell lines by RT-PCR analysis

Two RT-PCR assays were designed to investigate the presence of the DHFR2_202 transcript in numerous cell lines; this analysis was required to perform the sucrose cushion ultracentrifugation RT-qPCR experiment.

3.2.2.1. Cell culture, RNA extraction, RNA integrity, complementary DNA (cDNA) synthesis and genomic contamination check

HepG2 cells and undifferentiated human induced pluripotent stem (iPS) cells were grown according to Chapter 2, Section 2.2.2. RNA was extracted from these cells as per Chapter 2, Section 2.2.4. RNA that was previously extracted from 17158-Coriell (B-Lymphocyte) and HEK293 cells were obtained from the laboratories - 80 °C freezers, see Chapter 2, Section 2.1.5 for additional information on extraction methods. The concentration of each RNA was quantified according to Chapter 2, Section

2.2.5. The integrity of all the RNA samples was assessed as per Chapter 2, Section 2.2.6. The RNA samples were subsequently DNase treated and reverse transcribed using the Sigma and Bioline reagents outlined in Chapter 2, Section 2.2.8.1. The newly synthesised cDNA was checked for residual genomic contamination as per Chapter 2, Section 2.2.10.

3.2.2.2. DHFR2_202 PCR assays on newly synthesised cDNA

The first PCR assay was designed to target exon 1 of the DHFR2_202 transcript. PCR reactions were prepared according to Chapter 2, Section 2.2.11.1.1 using the 17158-Coriell, HepG2, HEK293 and human undifferentiated iPS cells cDNA synthesised in Section 3.2.2.1. The reactions were incubated in the Thermal Cycler and visualised on an agarose gel as per Chapter 2, Section 2.2.11.1.1.

The second PCR assay was designed to target exon 1 and exon 2 of the DHFR2_202 transcript. PCR reactions were prepared according to Chapter 2, Section 2.2.11.1.2 using the 17158-Coriell, HepG2, HEK293 and undifferentiated human iPS cells cDNA synthesised in Section 3.2.2.1. The reactions were incubated in the Thermal Cycler and visualised on an agarose gel as per Chapter 2, Section 2.2.11.1.2.

3.2.2.3. Sanger Sequencing of DHFR2_202 PCR products

The PCR products from Section 3.2.2.2 were prepared for Sanger Sequencing as follows: PCR reactions with one PCR product were purified using the ExoSAP Clean-up reagent, as per Chapter 2, Section 2.2.12.1; whereas PCR products with multiple products were purified using the Macherey-Nagel NucleoSpin Gel and PCR Clean-up Kit, detailed in Chapter 2, Section 2.2.12.2. The purified PCR products were sent for Sanger Sequencing as per Chapter 2, Section 2.2.16. The results were analysed using the tools mentioned in Chapter 2, Section 2.2.17.

3.2.3. Examining the relationship between the ribosome and the DHFR2_201/202 transcripts by performing the sucrose cushion ultracentrifugation RT-qPCR experiment

3.2.3.1. Ribosome inhibitor treatment on actively growing HepG2 cells

HepG2 cells were grown to 80% confluency in three T175 flasks, according to Chapter 2, Section 2.2.2. The cells were washed with 10 mL of warm phosphate-buffered saline (PBS). Active translation was halted by adding 28 mL serum-free DMEM containing cycloheximide (CHX) at a final

concentration of 100 µg/mL to each flask. The flasks were placed in the 5% carbon dioxide incubator at 37 °C for 5 min. The cells were washed twice with ice-cold PBS containing CHX (100 µg/mL). The PBS from the last wash step was left in the flask and the cells were harvested by gently scraping them into it using a cell scraper. The contents of two flasks were merged for the sample and the remaining flask was for the whole lysate control. They were pelleted by centrifuging at 2,000 x *g* for 5 min at 4 °C.

3.2.3.2. Lysis of CHX treated HepG2 cells

The sample and whole lysate control cell pellets from Section 3.2.3.1 were washed in 1 mL of PBS containing CHX (100 µg/mL) and pelleted again by centrifuging at 2,000 x *g* for 5 min at 4 °C. They were resuspended in 250 µL of cushion sucrose solution (CSS) buffer and glass beads were added. The cells were sheared by vortexing (VORTEX 3, IKA Shakers) the samples at 2,500 rpm for 20 sec and returning to the ice for 20 sec for eight cycles. Once, the cells were fully lysed, an additional 300 µL of CSS buffer was added. To remove large cellular debris the lysate was centrifuged at 2,000 x *g* for 5 min at 4 °C and the supernatant was retained. To eliminate small cellular debris the lysate was centrifuged at 10,000 x *g* for 15 min at 4 °C and the supernatant was saved.

3.2.3.3. Separation of ribosome bound/free RNA using sucrose cushion ultracentrifugation

The supernatant belonging to the sample (500 µL) from Section 3.2.3.2 was gently loaded onto 400 µL of 60% (w/v) sucrose solution in a thick wall polycarbonate tube. The sample was centrifuged at 55,000 x *g* for 2.5 hour at 4 °C in a Beckman Coulter Optima XL Series Ultracentrifuge with the Type 70.1 Ti Rotor using adaptors supplied by Laborgeraete Beranek GmbH. The supernatant (ribosome free RNA) and pellet (ribosome-bound RNA) were carefully separated and retained. The pellet was resuspended in 500 µL of CSS buffer. The whole lysate control was stored at 4 °C during this period.

3.2.3.4. RNA isolation and purification of ribosome bound and free RNA HepG2 fractions

Total RNA was extracted from the pellet sample, supernatant sample and whole lysate control sample using the TRIzol method. An equal volume of ice-cold TRIzol reagent was mixed with the pellet sample, supernatant sample and whole lysate control. After the samples were incubated at room temperature for 5 min, 0.2 mL of chloroform per 1 mL of TRIzol reagent was added. The

samples were vigorously vortexed for 15 sec and left at room temperature for 3 min. The samples were centrifuged at 12,000 x *g* for 15 min at 4 °C which causes phase separation. The upper aqueous phase which contains the RNA was carefully isolated.

To obtain ultra-pure RNA the pellet sample, supernatant sample and whole lysate control were purified using the PureLink RNA Mini Kit. An equal volume of 70% (v/v) ethanol was mixed with each RNA sample. It was transferred into a spin column and centrifuged at 12,000 x *g* for 15 sec. The supernatant was discarded and 700 µL of wash buffer I was added to the column. It was centrifuged at 12,000 x *g* for 15 sec, the supernatant removed, and the spin column placed in a new collection tube. Wash buffer II (500 µL) was added to the column, centrifuged at 12,000 x *g* for 15 sec and the supernatant was discarded. The second wash step was repeated. The spin column was centrifuged at 12,000 x *g* for 15 sec to dry the membrane and then transferred into a recovery tube. RNase-free water (40 µL) was placed in the centre of the column and left at room temperature for 1 min. To recover the RNA from the spin column it was centrifuged at 12,000 x *g* for 2 min. The RNA was stored in a - 80 °C freezer until required.

3.2.3.5. RNA integrity check and reverse transcription of ribosome bound and free RNA HepG2 fractions

The concentration and purity of the pellet sample, supernatant sample and whole lysate control RNA were determined using the Nanodrop as per Chapter 2, Section 2.2.5. The integrity of the RNA was analysed according to Chapter 2, Section 2.2.6. The pellet sample, supernatant sample and whole lysate control RNA was DNase treated and reverse transcribed as per Chapter 2, Section 2.2.8.2, only using random hexamer primers. A no reverse transcriptase control (NRT) was also prepared for the pellet sample, supernatant sample and whole lysate control RNA according to Chapter 2, Section 2.2.8.2; these samples were generated to check for residual genomic DNA. The NRT and cDNA samples were initially checked for genomic contamination using the basic method according to Chapter 2, Section 2.2.10.

3.2.3.6. qPCR genomic contamination check on NRT control cDNA

Due to qPCR being a lot more sensitive than cold PCR a more stringent genomic contamination check was subsequently carried out to approve the cDNA samples for qPCR analysis. The best way to check cDNA for genomic contamination is to run the NRT controls as the template DNA in all the

qPCR assays that will be used in the investigation. Any amplification on the Light Cycler indicates residual genomic contamination. The pellet sample, supernatant sample and whole lysate control NRT samples generated in Section 3.2.3.5 were used as template DNA in the six qPCR assays (GAPDH, nuclear enriched abundant transcript 1 (NEAT1), DHFR, DHFR2_All, DHFR2_201, DHFR_202) that will be employed in Section 3.2.3.7. These reactions were set up and run according to Chapter 2, Section 2.2.18.1.

3.2.3.7. qPCR on ribosome bound and free cDNA

Once all quality control checks were performed, the investigative qPCR on the ribosome bound and free cDNA was carried out. A non-template control (NTC), whole lysate control, supernatant and pellet reaction was set up for the six qPCR (GAPDH, NEAT1, DHFR, DHFR2_All, DHFR2_201, DHFR2_202) assays according to Chapter 2, Section 2.2.18.1. The GAPDH assay was used as the ribosome-bound RNA control. The NEAT1 assay was employed as the ribosome-free RNA control. And the DHFR assay was used as the dihydrofolate reductase gene family control. The Light Cycler conditions and data analysis for this experiment is described in Chapter 2, Section 2.2.18.1.

3.3. Results

3.3.1. Ribo-seq data reveals the DHFR2 RNA associates with the ribosome

There are numerous online resources that were built for depositing and visualising published Ribo-seq data, including the RPFdb v2.0 beta database (Wang et al., 2019; Xie et al., 2016). The ribosome profiling data that is available on the RPFdb v2.0 beta platform was employed to assess the relationship between the ribosome and the DHFR2 RNA. To do this the Ribo-seq data of 23 different cell lines and tissues were isolated according to Section 3.2.1. See Appendix B for more information on the samples. The RPKM values for the *GAPDH, POLG, MTHFD1L, TYMS, DHFR* and *DHFR2* genes were extracted from the 23 cell/tissue samples and analysed.

Even though the RPFdb v2.0 beta platform normalises all the RPKM values from different samples, it was essential that genes that are ubiquitously expressed were included as quality controls. Ubiquitously expressed genes should have similar RPKM values throughout all cell and tissue types examined. Therefore, any samples that have an RPKM value that deviates from the other samples should be eliminated from subsequent analysis. This quality control step prevents inaccurate data samples from being included in the investigative analysis. The *GAPDH* and *POLG* genes were used as the ubiquitously expressed controls (Figure 3.4). The *GAPDH* gene was chosen as the overall internal control as its expression is high and consistent in many different cells and tissues (Jo et al., 2019). The HeLa cell line and kidney normal tissue RPKM values were significantly lower than the other samples in the *GAPDH* gene (Figure 3.4 (A)). Therefore, these samples were removed from subsequent analysis. The *POLG* gene was employed as the mitochondrial internal control. A mitochondrial internal control was also included in this initial quality control step in case the *DHFR2* gene encodes a mitochondrial reductase enzyme as McEntee et al., 2011 and Anderson et al., 2011 suggested. The SUM159 cell line RPKM value was significantly higher than the other samples in the *POLG* gene (Figure 3.4 (B)). Therefore, the SUM159 cell line was removed from subsequent analysis. In the remaining 20 samples, the binding of the GAPDH and POLG mRNA to the ribosome was approximately the same order of magnitude in all cells and tissues. Thus, the 20 human cell and tissue samples were employed to perform the remaining analysis.

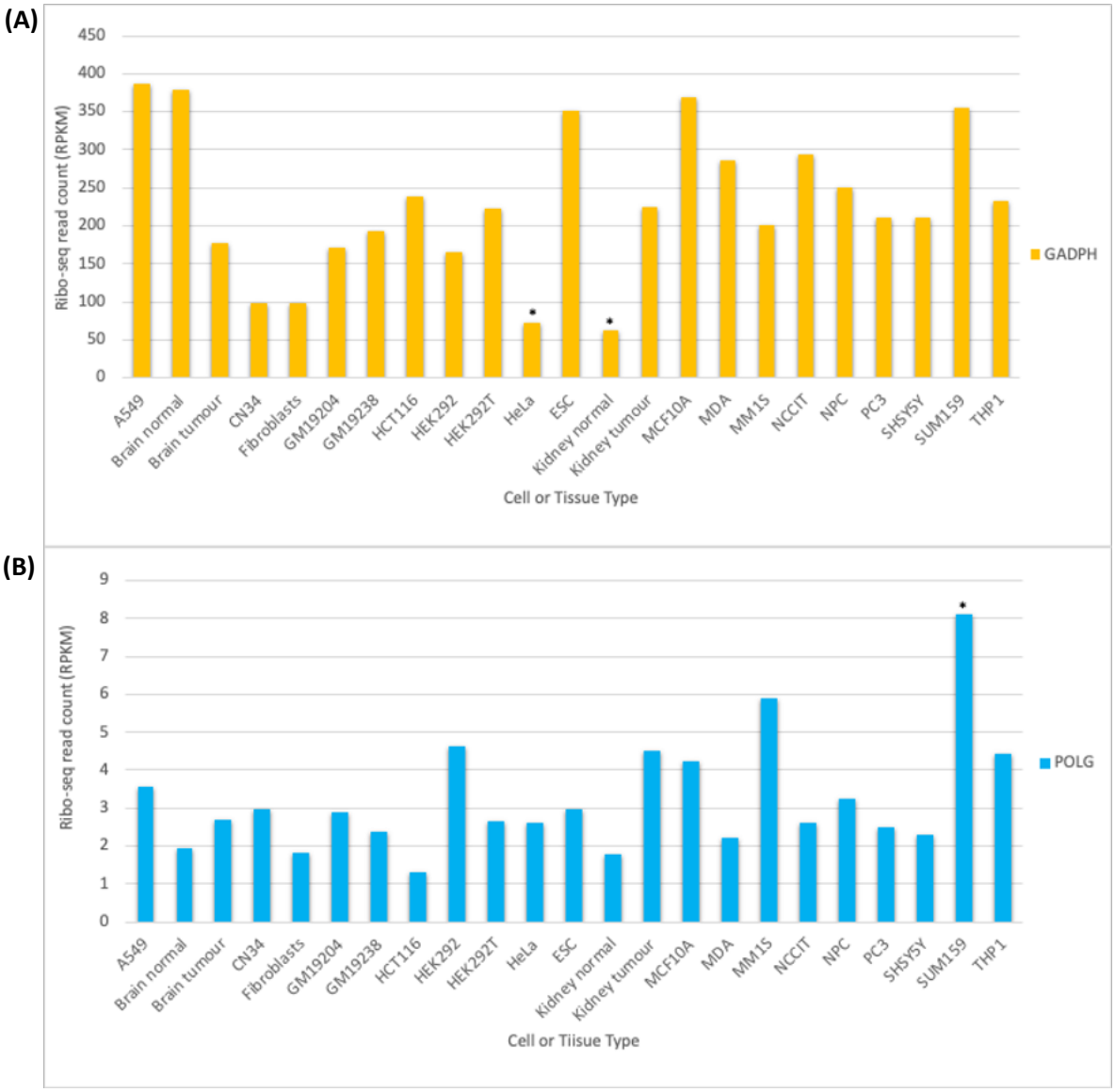


Figure 3.4 Ribo-seq results for the GAPDH and POLG genes, across multiple cells and tissues. *This data was obtained from the RPFdb v2.0 beta database. Graph A represents the RPKM results for the GADPH gene, the ubiquitously expressed control. Graph B represents the RPKM results for the POLG gene, the ubiquitously expressed mitochondrial control. The three outliers (*) in both graphs were removed from subsequent analysis. In the remaining samples, the binding of the GAPDH and POLG mRNA to the ribosome was approximately the same order of magnitude in all cells and tissues. Created with Microsoft excel.*

The RPKM values for the *MTHFD1, TYMS* and *DHFR* genes were subsequently examined in the remaining 20 samples (Figure 3.5). These genes were used as one-carbon metabolism controls. The *MTHFD1, TYMS* and *DHFR* genes encode essential enzymes within the OCM pathway. Therefore they should have a significant association with the ribosome. The M*THFD1L* gene was used as the OCM mitochondrial control. An OCM mitochondrial control was included in this analysis in case the DHFR2 protein localises to the mitochondria as McEntee et al., 2011 suggested. As predicted, the ribosome profiling data revealed an association between the MTHFD1L mRNA and the ribosome in the majority of cells and tissues (Figure 3.5 (A)). The *TYMS* gene was included as the de novo thymidylate biosynthesis pathway control. A de novo thymidylate biosynthesis pathway control was included in case the *DHFR2* gene encodes a reductase enzyme that plays a role in de novo thymidylate biosynthesis. There was an even greater association seen between the TYMS mRNA and the ribosome in the majority of samples examined (Figure 3.5 (A)). The *DHFR* gene was included in this analysis as it is part of the same gene family as the *DHFR2* gene. As expected, the ribosome profiling data revealed that the DHFR mRNA binds to the ribosome in most of the cell and tissue samples (Figure 3.5 (B)).

After carrying out all the previous quality control steps the RPKM values for the *DHFR2* gene were assessed. The DHFR2 RNA was found to bind to the ribosome in all 20 samples examined (Figure 3.5 (B)). Each sample showed a different degree of association, with the most significant association observed in the prostate cancer cells (PC3), brain tumour tissue and neural progenitor cells (NPC).

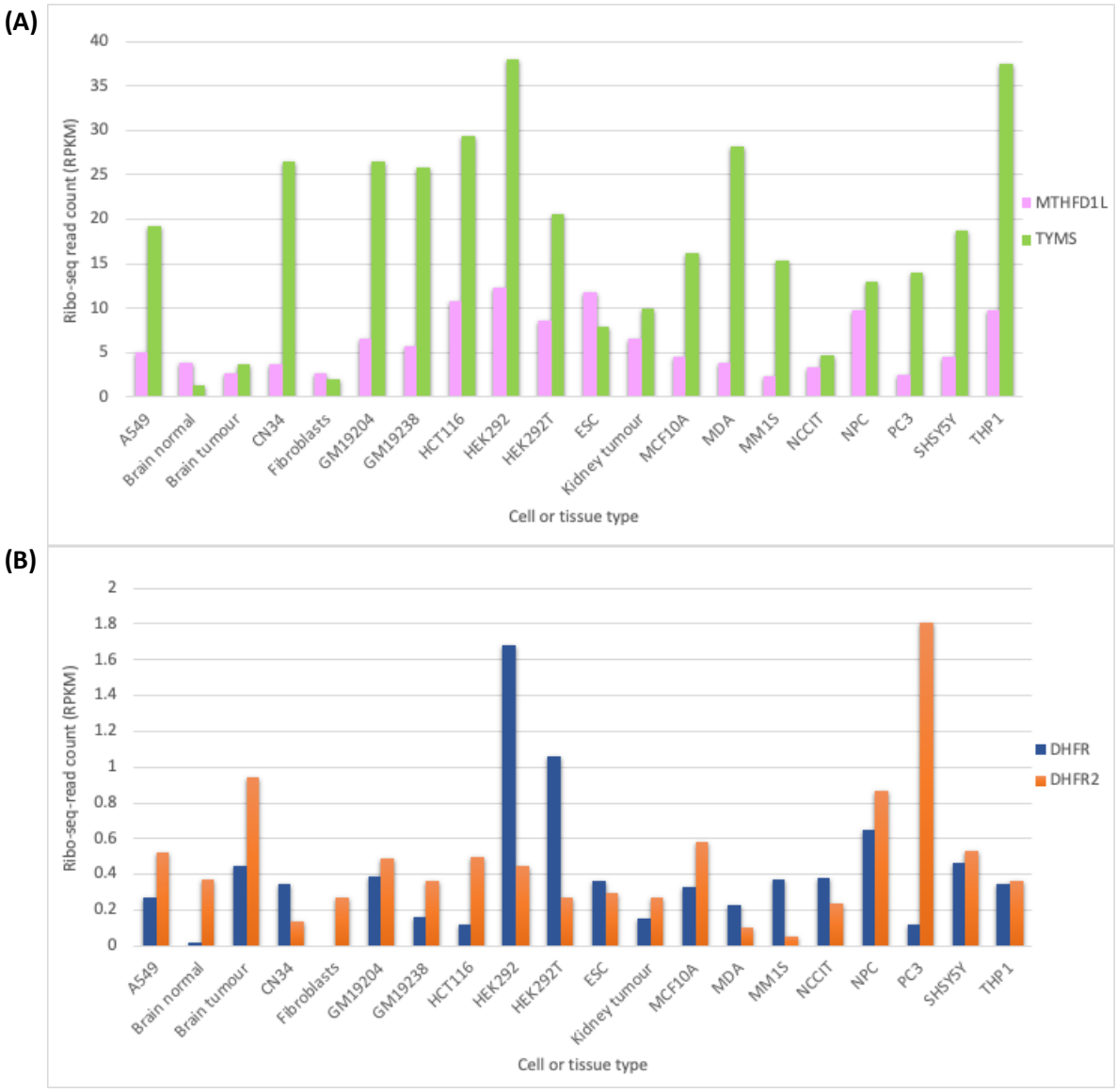


Figure 3.5 Ribo-seq results for the MTHFD1L, TYMS, DHFR and DHFR2 genes, across multiple cells and tissues. *This data was obtained from the RPFdb v2.0 beta database. Graph A illustrates the RPKM values for the MTHFDL1 and TYMS genes, the OCM pathway controls. This data shows that the ribosome associates with the MTHFD1L mRNA and even more so with the TYMS mRNA. Graph B compares the RPKM values for the two functional genes of the dihydrofolate reductase family, the DHFR and DHFR2 gene. This data shows that the ribosome associates with both the DHFR and DHFR2 mRNA. Created with Microsoft excel.*

The Ribo-seq data available for the *DHFR2* gene was further interrogated to determine which regions of the gene had a greater association with the ribosome. The RPKM values for the overall gene were broken down into three regions; the 5'UTR, the CDS and the 3'UTR (Figure 3.6). This analysis showed that the ribosome binds to the 5'UTR and more so to the coding sequence of the DHFR2 RNA. There was little to no association observed between the ribosome and the 3'UTR of the DHFR2 RNA in any of the samples. This result is consistent with an mRNA molecule that is translated into a protein.

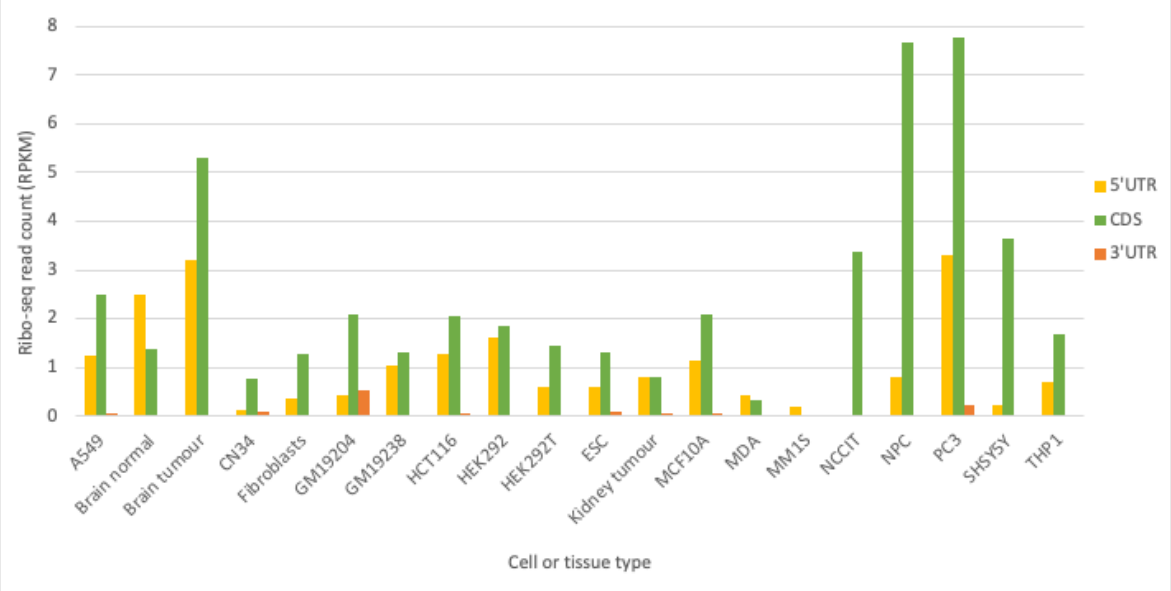


Figure 3.6 Ribo-seq results for the 5'UTR, CDS and 3'UTR regions of the DHFR2 gene, across multiple cells and tissues. *This data was obtained from the RPFdb v2.0 beta database. This graph compares the RPKM values of the different regions of the DHFR2 gene. Ultimately showing that the ribosome binds to the 5'UTR and even more so to the CDS of the DHFR2 RNA. There was little to no association between the ribosome and the 3'UTR of the DHFR2 RNA in any of the cell and tissue samples. Created with Microsoft excel.*

3.3.2. Transcript DHFR2_202 confirmed by RT-PCR for sucrose cushion experiment

Ensembl has annotated six different DHFR2 RNA isoforms, DHFR2_201 through to DHFR2_206. The predicted functions are described in Chapter 1, Section 1.5.2 and the predicted sequences are illustrated in Figure 1.9. Ensembl has also predicted that the DHFR2_201 and DHFR2_202 transcripts are the most plausible mRNA isoforms to generate a functional DHFR2 protein. Therefore, it was imperative that the translational capabilities of both DHFR2 transcripts be investigated. The RPFdb v2.0 beta database considers all RNA isoforms of a gene as one and the detail for the individual transcripts are lost. Therefore, a sucrose cushion ultracentrifugation step coupled with RT-qPCR was employed to target individual DHFR2 transcripts. Previous studies have confirmed a section of the DHFR2_201 transcript by RT-qPCR in multiple cell lines (McEntee et al., 2011). However, there was no experimental evidence of the DHFR2_202 RNA isoform. Therefore, two RT-PCR assays were designed to investigate the DHFR2_202 transcript in numerous cell lines before performing the translatability study on the DHFR2_201 and DHFR2_202 transcripts.

3.3.2.1. Quality control checks on RNA and newly synthesised cDNA for the DHFR2_202 PCR analysis

RNA is a very unstable molecule that is prone to degradation, which can impact downstream applications (Uzman, 2001). Therefore, the integrity of the RNA must be examined prior to use. Intact RNA will appear as two distinct bands on an agarose gel, which are the 28S and 18S rRNA structures. Partially degraded RNA samples will have the two bands with the addition of a low molecular weight smear. Completely degraded RNA will show up as a quite low molecular weight smear (Sarhan et al., 2014). Therefore, the integrity of all RNA samples that were going to be synthesised into cDNA for DHFR2_202 PCR analysis were examined by running them on an agarose gel (Section 3.2.2.1). Two distinct bands were visualised for the 17158-Coriell (B-Lymphocyte), HEK293, undifferentiated human iPS and HepG2 RNA samples (Figure 3.7). This result meant that the RNA was completely intact and ready to be reverse transcribed into cDNA.

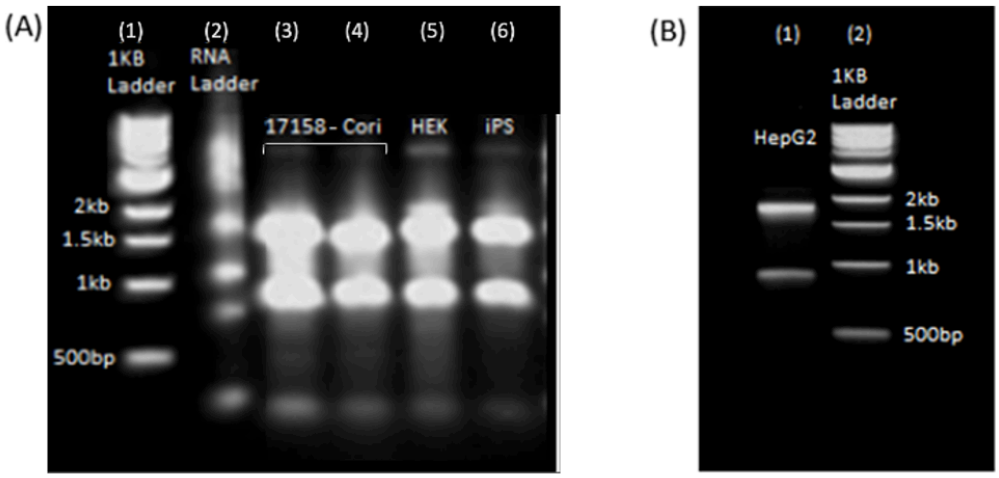


Figure 3.7 RNA integrity test on various RNA samples. *RNA (500ng) was resolved on a 1% (w/v) agarose gel stained with ethidium bromide. Intact RNA results in two distinct bands, representing 28S and 18S. Whereas completely degraded RNA appears as a quite low molecular weight smear, with no distinct bands. The figure shows the results of this test: (A) (1) 1 kb DNA Ladder, (2) RNA Ladder, (3&4) 17158-Coriell (B-Lymphocyte) RNA, (5) HEK293 RNA and (6) undifferentiated human iPS RNA, (B) (1) HepG2 RNA and (2) 1 kb DNA Ladder. The 28S and 18S bands were visible in all samples, therefore all the samples passed the integrity check.*

Newly synthesised cDNA should be checked for residual genomic contamination before proceeding to investigative PCR assays. This check was carried out by performing an in-house intron spanning MTHFD1 PCR assay on each cDNA sample (Section 3.2.2.1). Amplification of cDNA results in a PCR product of 232 bp whereas genomic DNA (gDNA) generates a 330 bp product. Figure 3.8 shows there were no bands present in the NTC. Furthermore, a defined band was present at 232 bp in the 17158-Coriell, HepG2, HEK293 and undifferentiated human iPS cDNA samples. Therefore, these samples passed the genomic contamination check and were ready for downstream application.

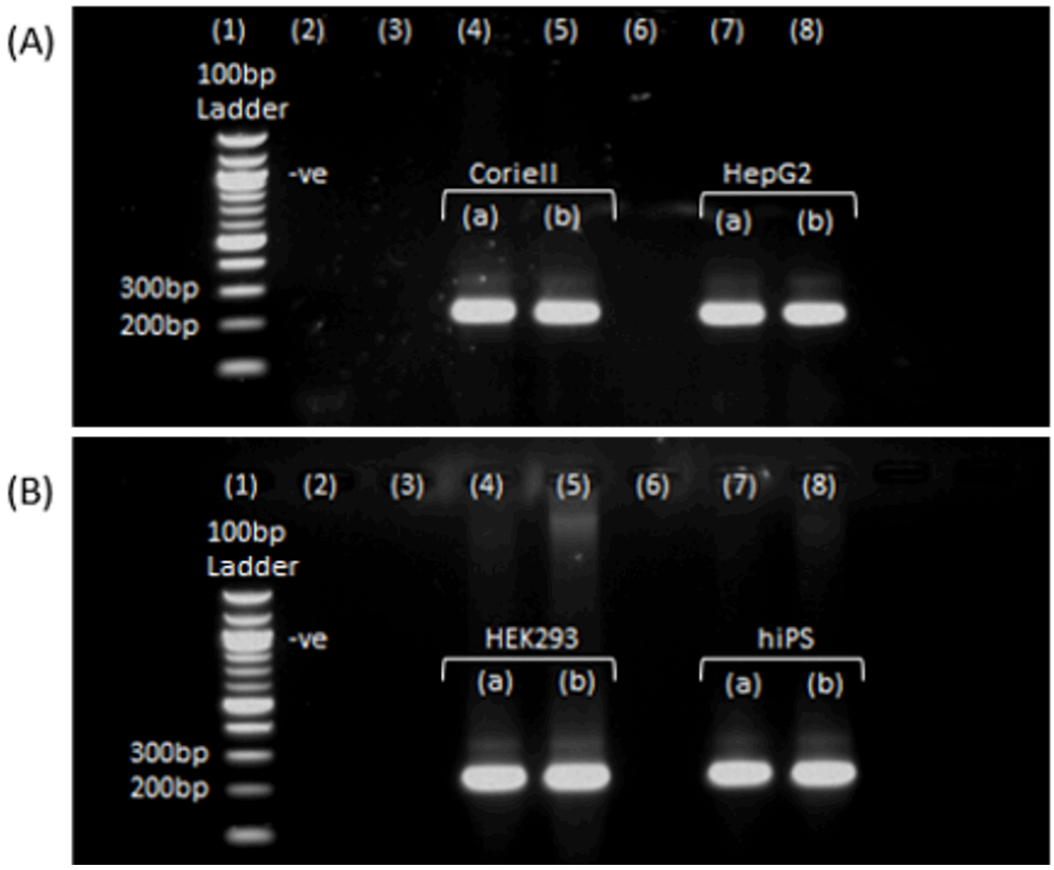


Figure 3.8 Genomic contamination check on the newly synthesised cDNA from the Coriell, HepG2, HEK293 and undifferentiated human iPS samples. *The PCR products (10 µL) were resolved on a 1% (w/v) agarose gel stained with ethidium bromide. A band at 330 bp revealed gDNA whereas a band at 232 bp represented cDNA. The figure shows the results of the genomic contamination test: (A) (1) 100 bp DNA Ladder, (2) NTC, (3) empty, (4&5) 17158-Coriell cDNA PCR reactions, (6) empty, (7&8) HepG2 cDNA PCR reactions. (B) (1) 100 bp DNA Ladder, (2) NTC, (3) empty, (4&5) HEK293 cDNA PCR reactions, (6) empty, (7&8) undifferentiated human iPS cDNA PCR reactions. This figure shows that there were bands visible at 232 bp in each sample, meaning that all the samples passed the genomic contamination check.*

3.3.2.2. Confirmation of Ensembl's annotated DHFR2_202 isoform

Two PCR assays were designed to target the predicted DHFR2_202 RNA isoform. The first assay (Exon 1 assay) was designed to confirm the presence/absence of exon 1 and the second (Exon 1-2 assay) to investigate if exon 1 and exon 2 were part of the same RNA isoform (Figure 3.9).

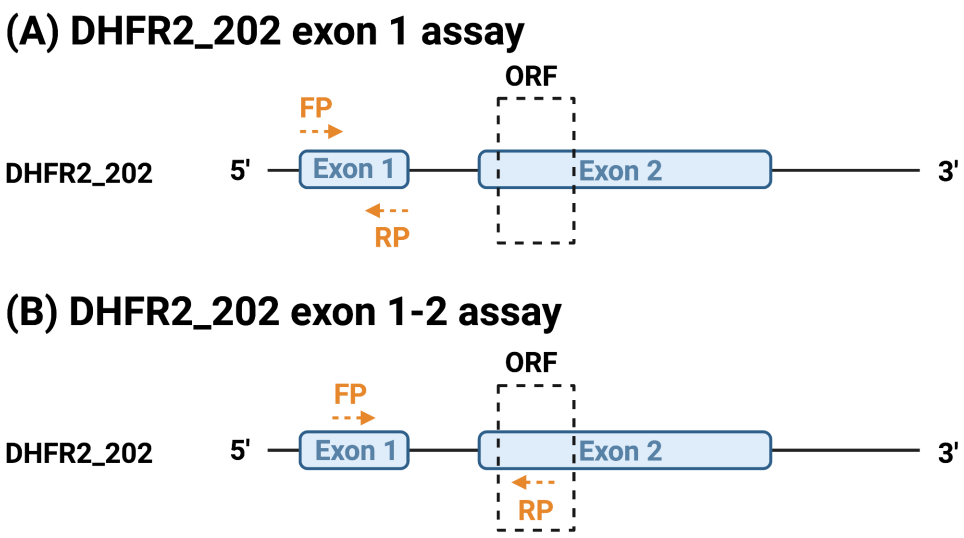


Figure 3.9 Illustration of where the DHFR2-202 PCR assay primers bind to DHFR2_202 RNA isoform. *(A) Exon 1 assay was designed to target Exon 1, which would result in a 335 bp PCR product. (B) Exon 1-2 assay to investigate if Exon 1 and Exon 2 were part of the same isoform, which would result in a 383 bp PCR product. Created with BioRender.com, Adapted from ensembl.org/index.html.*

The Exon 1 assay was designed to amplify the entire exon 1 region, which would result in a 335 bp PCR product. PCR results were visualised in an agarose gel (Figure 3.10), according to Section 3.2.2.2. There were no bands detected in the NTC. A band was visualised at 335 bp in the positive control, 17158-Coriell, HepG2, HEK293 and undifferentiated human iPS cDNA samples. The intensity of the band differed between each cell type, with the brightest seen in both HEK293 and HepG2 cells. Sanger Sequencing of the PCR product confirmed the presence of Exon 1 of the DHFR2_202 transcript (Appendix C).

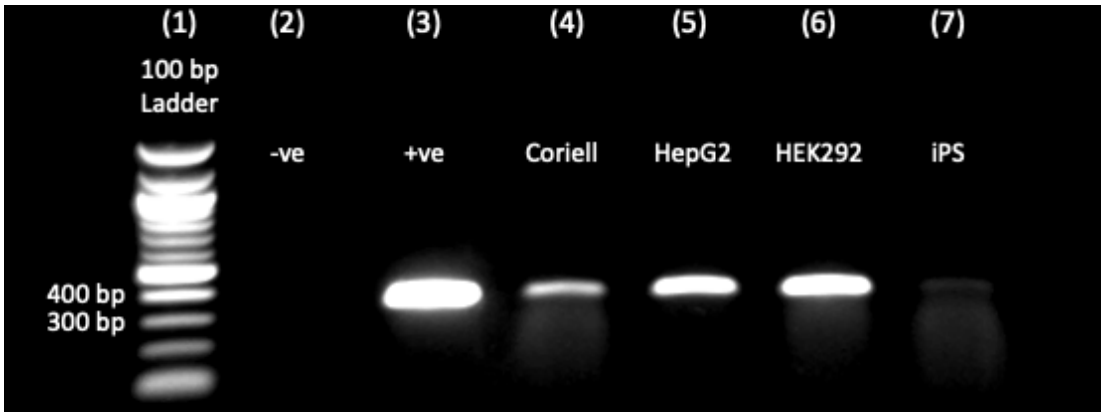


Figure 3.10 DHFR2_202 Exon 1 assay PCR product (10 µL) visualised on a 1% (w/v) agarose gel stained with ethidium bromide. *This assay was designed to determine the presence of the first exon belonging to the DHFR2_202 transcript that was predicted by Ensembl. This would result in a PCR product of 335 bp. The figure shows the results of this PCR assay: (1) 100 bp DNA Ladder, (2) NTC, (3) positive control, (4) 17158-Coriell cDNA PCR reaction, (5) HepG2 cDNA PCR reaction, (6) HEK293 cDNA PCR reaction and (7) undifferentiated human iPS cDNA PCR reaction. This gel showed no bands in NTC. There were bands present at 335 bp in the positive control and all cDNA PCR reactions.*

The Exon 1-2 assay was designed to flank intron 1 of the DHFR2 RNA, which would result in a 383 bp PCR product. The PCR results are shown in (Figure 3.11), according to Section 3.2.2.2. There were no bands present in the NTC. A band was visualised at 383 bp in the positive control, 17158-Coriell, HepG2, HEK293 and undifferentiated human iPS samples. Each sample varied in intensity, with the HEK293 and HepG2 showing the brightest band. Sanger Sequencing of the 383 bp PCR product confirmed exon 1 and exon 2 were part of the same mRNA isoform (Appendix D). An additional band with a higher molecular weight was visualised in the 17158-Coriell and HepG2 samples near the 500 bp mark. Sanger Sequencing of this PCR product revealed it aligned back to the DHFR2_201 transcript (Appendix E).

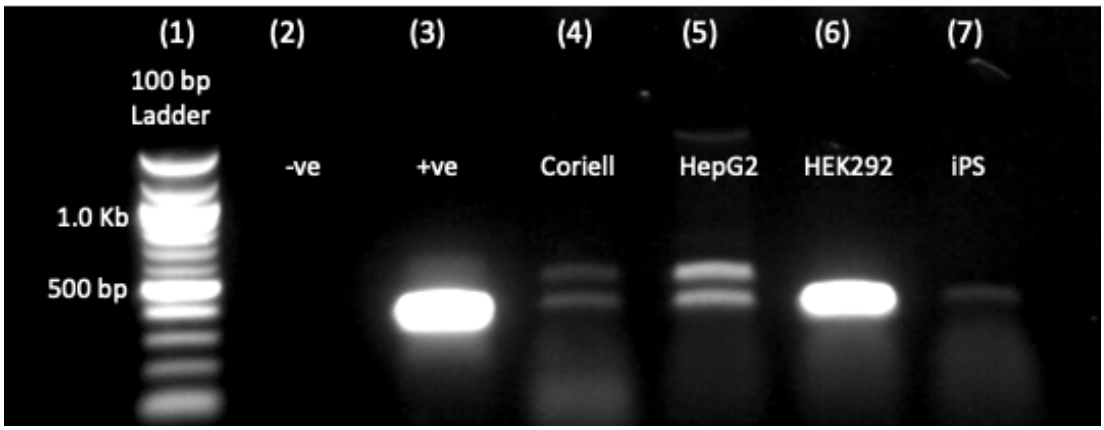


Figure 3.11 DHFR2_202 Exon 1-2 assay PCR product (10 µL) visualised on a 1% (w/v) agarose gel stained with ethidium bromide. *This assay was designed to confirm the presence of the first and second exon belonging to the DHFR2_202 transcript, which would give a PCR product of 383 bp. The figure shows the results of this PCR assay: (1) 100 bp DNA Ladder, (2) NTC, (3) positive control, (4) 17158-Coriell cDNA PCR product, (5) HepG2 cDNA PCR product, (6) HEK293 cDNA PCR product and (7) undifferentiated human iPS cDNA PCR product. This gel showed no bands in NTC. There were bands present at 383 bp in the positive control and all cDNA PCR reactions. A second, larger band was visualised in the 17158-Coriell and HepG2 samples near the 500 bp mark.*

3.3.3. Ribosome fractionation by sucrose cushion ultracentrifugation shows the DHFR2 transcripts differentially associate with the ribosome

A sucrose cushion ultracentrifugation RT-qPCR technique (as described in Section 3.1.5) was the second approach employed to investigate the translational status of the *DHFR2* gene. This technique was chosen over ribosome and polysome profiling analysis as the necessary equipment was not available to carry them out. Whole lysate control, supernatant and pellet sucrose cushion ultracentrifugation RNA samples were isolated and reverse transcribed into cDNA, according to Section 3.2.3.1 - 3.2.3.5. Six qPCR assays (GAPDH, NEAT1, DHFR, DHFR2_All, DHFR2_201 and DHFR2_202) were carried out on the whole lysate control, supernatant sample and pellet sample cDNA (Section 3.2.3.7). The results from the qPCR revealed which fraction the target mRNA was more abundant in, the ribosome-bound or ribosome-free fraction.

3.3.3.1. Quality control checks on the RNA and newly synthesised cDNA from the sucrose cushion ultracentrifugation step prior to qPCR analysis

As mentioned in Section 3.3.2.1, it is imperative that the integrity of the RNA is examined prior to downstream applications. The whole lysate control, supernatant and pellet sucrose cushion ultracentrifugation RNA samples were run on an agarose gel to assess the integrity, according to Section 3.2.3.5. Figure 3.12 shows there were two distinct bands for each of the samples. These bands represented the 28S and 18S rRNA structures. This result meant that the RNA was completely intact and ready to be reverse transcribed into cDNA.

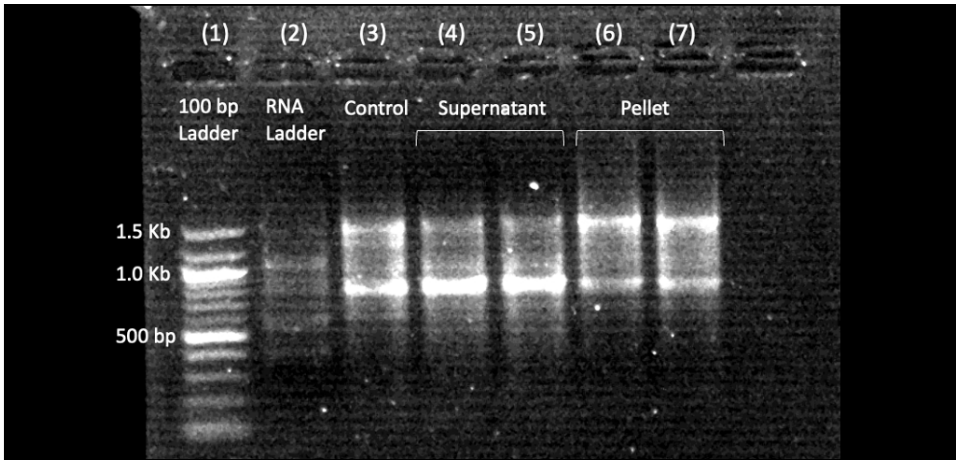


Figure 3.12 RNA integrity check on whole lysate control, supernatant and pellet sucrose cushion ultracentrifugation RNA samples. *The RNA (500 ng) was resolved on a 1% (w/v) agarose gel stained with SYBR safe. The figure shows the results of this test: (1) 100 bp DNA Ladder, (2) RNA Ladder, (3) whole cell lysate RNA (4&5) supernatant sample RNA and (6&7) pellet sample RNA. The 28S and 18S bands were visible in all samples, therefore all samples passed the integrity check.*

As mentioned in Section 3.3.2.1, it is also essential that newly synthesised cDNA is checked for residual genomic contamination prior to performing investigative PCR assays. The whole lysate control, supernatant and pellet cDNA samples were tested for genomic contamination using the intron spanning MTHFD1 PCR assay described in Section 3.2.3.5. Figure 3.13 shows there were no bands present in the NTC and the NRT control. Furthermore, the whole lysate control, supernatant and pellet sucrose cushion ultracentrifugation cDNA samples only had the cDNA band at 232 bp. Therefore, all cDNA samples passed this genomic contamination check.

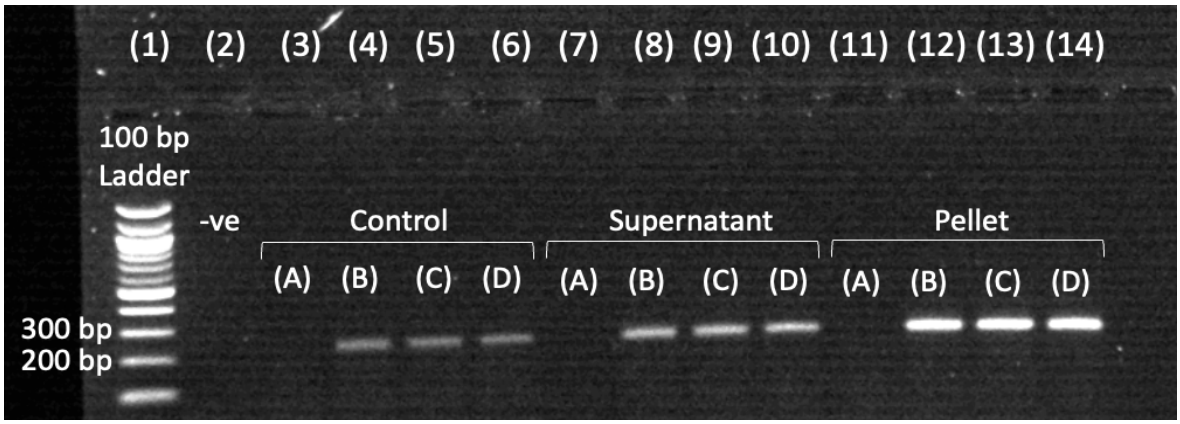


Figure 3.13 Genomic contamination check on the newly synthesised cDNA from the whole lysate control, supernatant and pellet sucrose cushion ultracentrifugation samples. *The PCR products (10 µL) were resolved on a 1% (w/v) agarose gel stained with SYBR safe. A band at 330 bp reveals gDNA whereas a band at 232 bp represents cDNA. (1) 100 bp DNA Ladder, (2) NTC, (3-6) whole lysate control cDNA PCR reactions, (7-10) supernatant cDNA PCR reactions and (11-14) pellet cDNA PCR reactions. Sample A: NRT control and sample B-D: cDNA reaction. This gel shows there were no bands in NTC and NRT controls. There were bands present at 232 bp in all cDNA PCR reactions, meaning that all the samples passed the genomic contamination check.*

Due to qPCR being a lot more sensitive than cold PCR a more stringent genomic contamination check was subsequently carried out to approve the cDNA samples for qPCR analysis. The best way to check cDNA for genomic contamination is to make an NRT control and run it as the template DNA in all the qPCR assays that will be used. Any amplification visualised on the Light Cycler represents residual genomic contamination (Lion, 2001). Therefore, qPCR reactions were set up according to Section 3.2.3.6 for the six assays (GAPDH, NEAT1, DHFR, DHFR2_All, DHFR2_201, DHFR2_202) that will be employed in the sucrose cushion ultracentrifugation qPCR experiment (Section 3.3.3.2). The NRT cDNA made from the whole lysate control, supernatant and pellet RNA was used as the template DNA for these assays. There was no amplification detected on the Light Cycler for any of these six assays. Therefore, the cDNA samples from each fraction also passed this genomic contamination check and were ready for downstream application.

3.3.3.2. RT-qPCR on ribosome-bound and ribosome-free mRNA fractions shows the DHFR2 RNA isoforms differentially associate with the ribosome

Six RT-qPCR assays were performed on the quality-controlled cDNA from the whole lysate control, supernatant and pellet sucrose cushion ultracentrifugation RNA samples, according to Section 3.2.3.7. These assays targeted GAPDH, NEAT1, DHFR and DHFR2 (DHFR2_All, DHFR2_201 and DHFR2_202) transcripts. The relative expression of each target RNA was calculated for the pellet (ribosome-bound RNA) and supernatant (ribosome-free mRNA) fractions (Figure 3.14); the expression ratios were relative to the target in the whole lysate control sample (Chapter 2, Section 2.2.18.1).

The assay that amplified the GAPDH transcripts was used as the ribosome-bound control. As expected, the RT-qPCR analysis showed the GAPDH mRNA was significantly more abundant in the pellet which was the ribosome-bound fraction. The RNA expression of GAPDH was 7.47 higher in the ribosome-associated fraction (Figure 3.14 (A)). The assay that targeted the NEAT1 RNA isoforms was used as the ribosome-free control. As predicted, the RT-qPCR analysis revealed the NEAT1 RNA was slightly more abundant in the supernatant which was the ribosome-free fraction. The RNA expression of the *NEAT1* gene was 1.43 fold higher in the ribosome-free fraction compared to the ribosome-bound fraction (Figure 3.14 (B)). The assay that amplified the DHFR transcripts was used as the dihydrofolate reductase gene family control. The DHFR mRNA was present at a higher concentration in the ribosome-free fraction when compared to the ribosome-bound sample, with a fold difference of 4.28 (Figure 3.14 (C)).

Once the quality control assays were performed the investigative DHFR2 RT-qPCR assays could be carried out. An assay that targeted all the DHFR2 RNA isoforms was performed to compare the results to the Ribo-seq data. The RT-qPCR analysis showed that the DHFR2 RNA was more abundant in the ribosome-associated fraction. The DHFR2 RNA expression was 6.87 fold higher in the ribosome-associated fraction when compared to the ribosome-free sample (Figure 3.14 (D)). Finally, two assays were designed to target the two main DHFR2 RNA isoforms, DHFR2_201 and DHFR2_202. The RT-qPCR analysis revealed the DHFR2_201 transcript was significantly more abundant in the ribosome-bound fraction, with a fold difference of 6.04 (Figure 3.14 (E)). Whereas, the DHFR2_202 RNA isoform was slightly more abundant in the ribosome-free fraction. The DHFR2_202 transcript expression was 1.82 fold greater in the ribosome-free fraction (Figure 3.14 (F)).

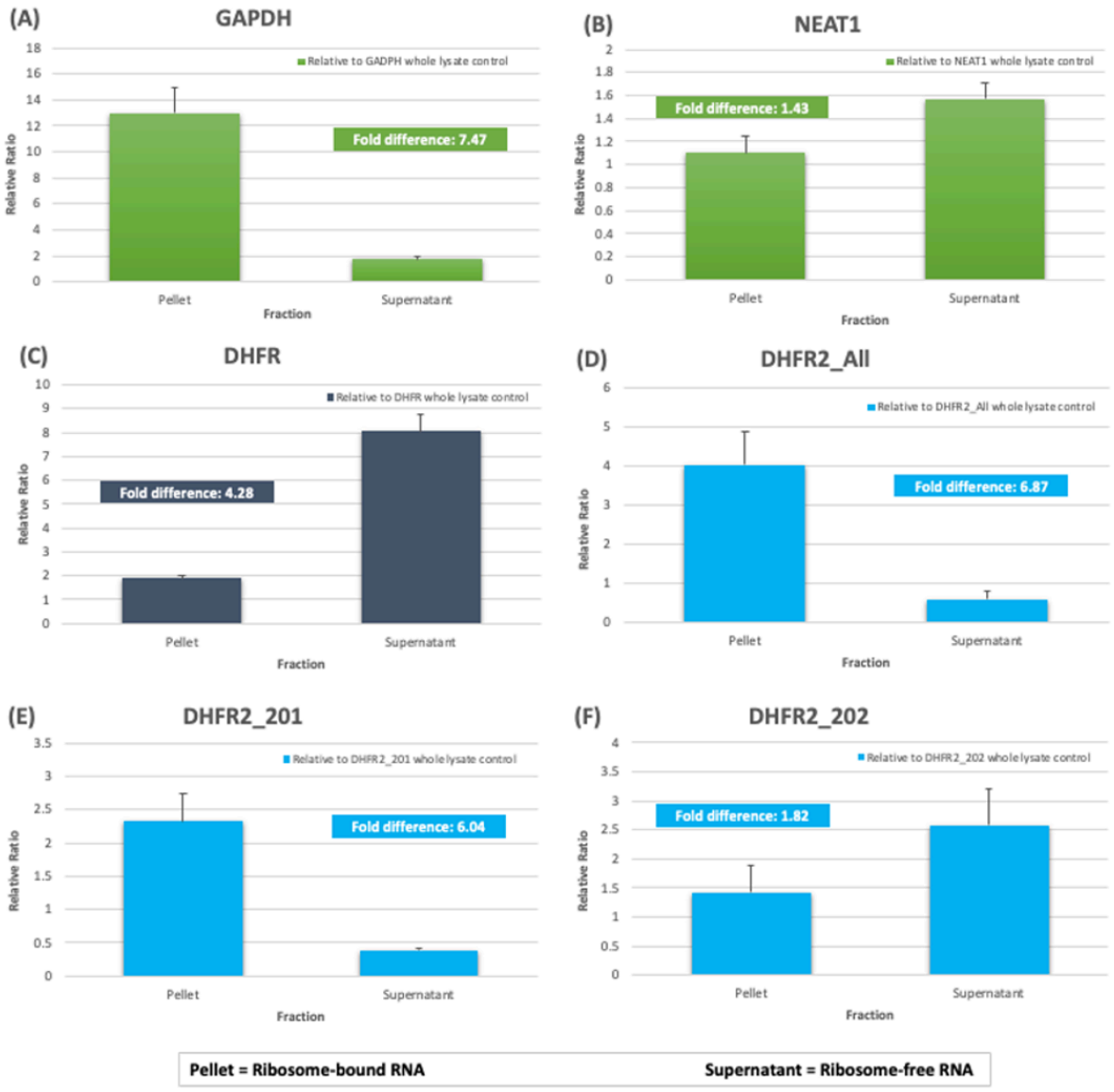


Figure 3.14 Sucrose cushion ultracentrifugation RT-qPCR results on GAPDH, NEAT1, DHFR and DHFR2 genes. *Six RT-qPCR assays targeting these genes were performed on the whole lysate control, supernatant sample and pellet sample. The relative expression of the target RNA was calculated for the pellet (ribosome-bound RNA) and supernatant (ribosome-free RNA) fractions using the LightCycler 96 qPCR System Instrument Software, Roche. The expression ratios were relative to the target in the whole lysate control sample. The error bars represent the ratio error which was calculated by the LightCycler 96 qPCR System Instrument Software, Roche. (A) The GAPDH assay was the ribosome-bound control, the RNA was a lot more abundant in the pellet fraction. (B) The NEAT1 assay was the ribosome-free control, there was a higher concentration of its RNA in the supernatant fraction. (C) The DHFR assay was the dihydrofolate reductase gene family control, the RNA was present at a higher concentration in the ribosome-free fraction. (D) The DHFR2_All assay showed that the DHFR2 RNA was more abundant in the ribosome-associated fraction. (E) The DHFR2_201 assay revealed the DHFR2_201 transcript was significantly more abundant in the pellet fraction. (F) The DHFR2_202 assay showed the DHFR2_202 RNA isoform was slightly more abundant in the supernatant fraction. Created with Microsoft excel.*

3.4. Discussion

The overall aim of this chapter was to investigate the translational status of the human *DHFR2* gene. A secondary aim was to confirm specific isoforms of DHFR2 RNA as they relate to translation. This was examined by studying the relationship between the DHFR2 RNA and the ribosome. Two different approaches were employed to investigate this; ribosome profiling and sucrose cushion ultracentrifugation RT-qPCR.

Ribosome profiling is the most powerful tool used to assess the global translation in a cell at a given point (Ingolia, 2014). This technique calculates the rate of binding between a target mRNA molecule and the ribosome. There are multiple online resources available to visualise published Ribo-seq data, including RPFdb v2.0 beta (Wang et al., 2019). The data from 23 studies were downloaded from this database and the RPKM values for six genes were examined (Section 3.2.1 for methods, Section 3.3.1 for results). The *GAPDH* gene was used as the ubiquitously expressed control and the *POLG* gene was employed as the ubiquitously expressed mitochondrial control. Both of these genes encode a functional protein that should be expressed at a consistent level across different cells and tissues. As expected, there was a consistent level of association seen between these genes and the ribosome (Figure 3.4). However, the RPKM values of three samples appeared to be outliers in either the *GAPDH* or *POLG* gene. Therefore, these samples were excluded from subsequent analyses. The *MTHFD1L* gene was used as the OCM mitochondrial control. The *MTHFD1L* gene encodes an enzyme that catalyses the production of formate from 10-formyl-THF in the mitochondria (Prasannan et al., 2003). Formate is one of the main suppliers of one-carbon groups to the OCM pathway to support purine synthesis, thymidylate biosynthesis and the provision of methyl groups (Brosnan & Brosnan, 2016). Therefore, MTHFD1L is one of the key enzymes of the folate cycle and thus its mRNA should be rapidly translated into a protein. As predicted, there was a high association between the MTHFD1L mRNA and the ribosome in the majority of cells and tissues (Figure 3.5 (A)). The *TYMS* gene was used as the de novo thymidylate biosynthesis pathway control. The *TYMS* gene encodes an enzyme that is responsible for the reductive methylation of deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP). Therefore, the TYMS enzyme is essential for DNA and RNA synthesis (Fox & Stover, 2008; Stover, 2009). For this reason, the TYMS mRNA should have a strong translation rate. As expected, the ribosome was found to bind to TYMS mRNA at an extremely high level in the majority of cell and tissue samples (Figure 3.5 (A)). The translation rate of the *DHFR* gene was investigated as it is part of the same gene family as the *DHFR2* gene. As previously mentioned, the DHFR enzyme generates THF, which acts as a one-carbon donor to support vital anabolic reactions in the OCM pathway (Schnell et al., 2004). Therefore, the DHFR

mRNA should be actively translated into a protein in most cells and tissues. As predicted, the ribosome profiling revealed that the DHFR mRNA associated with the ribosome in most of the cell and tissue samples (Figure 3.5 (B)). Finally, the Ribo-seq data was interrogated to determine if the *DHFR2* gene has the capabilities to be translated into a functional protein. The ribosome profiling data revealed that the DHFR2 RNA was bound to the ribosome in the majority of cell lines and tissues (Figure 3.5 (B)). Therefore, this result suggests that the DHFR2 RNA has the potential to be translated into a protein.

A huge advantage of performing Ribo-seq analysis is that it identifies the position on the mRNA molecule where the ribosome binds. The ribosome densities are usually higher in the 5'UTR and even more so in the coding regions of a transcript (Ingolia, 2010). Therefore, this technique can identify novel proteins, short peptides and alternative protein isoforms (Brar & Weissman, 2015). RPFdb v2.0 beta does not provide the exact locations of the ribosome binding sites along an mRNA molecule. It does however categorise the overall RPKM value of a gene into three regions; 3'UTR, CDS and 5'UTR. As expected, there was a higher association seen in the 5'UTR and even more so in the CDS regions of the *DHFR2* gene. Moreover, there was little to no association observed between the ribosome and the DHFR2 3'UTR (Figure 3.6). This data revealed that the predicted DHFR2 ORF is the most likely region within this gene to potentially encode a functional protein. Furthermore, this result eliminates an ORF located within the 3'UTR of the *DHFR2* gene that could generate an alternative protein.

Even though the RPFdb v2.0 beta revealed a large amount of information about the translational status of the DHFR2 RNA, there are some limitations associated with it. The main disadvantage of using online databases is they consider all DHFR2 RNA isoforms as one and the detail for the individual transcripts are lost. For this reason, there is uncertainty whether all or the majority of mRNA isoforms bind to the ribosome. Therefore, this analysis would not be sufficient to conclude the translational capability of the *DHFR2* gene. This issue could be overcome by experimentally performing ribosome profiling or other ribosome purification techniques that are designed to target individual DHFR2 transcripts. However, prior to performing the ribosome purification experiment, the DHFR2 transcripts needed to be identified and experimentally confirmed.

Ensembl has annotated six different DHFR2 RNA isoforms, DHFR2_201 through to DHFR2_206. The predicted functions are described in Chapter 1, Section 1.5.2 and the predicted sequences are illustrated in Figure 1.9. This software tool has predicted that the DHFR2_201 and DHFR2_202 transcripts are the most plausible mRNA isoforms to be functional. The research carried out prior to this project was focused on the DHFR2_201 transcript. This is due to it containing the entire

DHFR2 ORF and also being highly abundant across a panel of cells and tissues (Anderson et al., 2011; McEntee et al., 2011). However, bioinformatic analysis of published PacBio sequencing data revealed that the DHFR2_202 transcript was also abundant in a vast number of cell and tissue samples (Tyminski, unpublished). Furthermore, this mRNA isoform also had the whole DHFR2 ORF. The only obvious difference between the DHFR2_201 and DHFR2_202 transcripts is the exon 1 nucleotide sequence (Appendix F). Therefore, it was imperative that the translational capabilities of both DHFR2 transcripts be investigated. Prior to carrying out the sucrose cushion ultracentrifugation RT-qPCR experiment the presence of the DHFR2_202 transcript needed to be confirmed. Two RT-PCR assays specific to the DHFR2_202 transcript were performed on four cell lines to investigate this. The RT-PCR analysis confirmed the presence of exon 1 in all cell lines tested (Figure 3.10). The results also showed that exon 1 and 2 reside on the same mRNA isoform (Figure 3.11). Therefore, these results confirmed the presence of the DHFR2_202 transcript. Thus the sucrose cushion ultracentrifugation RT-qPCR experiment could now be performed on the DHFR2_201 and DHFR2_202 transcripts to see do they have similar translational capabilities.

A sucrose cushion ultracentrifugation RT-qPCR technique was subsequently employed to investigate the translational status of the main DHFR2 transcripts, DHFR2_201 and DHFR2_202 (Section 3.2.3 for methods, Section 3.3.3.2 for results). This enrichment technique is an adaptation of the polysome profiling method; the gold standard approach for analysing the translation of mRNA. The sucrose cushion ultracentrifugation initially separates mRNA molecules into two fractions, actively translating and non-translating/poorly translating. RT-qPCR analysis on these fractions determines which group the target RNA is more abundant in. Ultimately predicting whether the RNA is being actively translated by the ribosome or not.

Quality control checks were carried out on the sucrose cushion fractions prior to performing the investigative RT-qPCR assays. GAPDH is a protein-coding gene that is essential for the existence of a cell. (Tarze et al., 2007). Therefore, the *GAPDH* gene should have an extremely high translation rate. For this reason, an assay that amplified the GAPDH transcripts was used as the ribosome-bound control. As predicted, the RT-qPCR analysis showed the GAPDH mRNA was significantly more abundant in the ribosome-bound fraction (Figure 3.14 (A)). NEAT1 is a long non-coding RNA that is found to maintain the architectural integrity of nuclear bodies known as paraspeckles (Ahmed et al., 2018). For this reason, the NEAT1 RNA should not have a significant relationship with the ribosome. Therefore, an assay that targeted the NEAT1 RNA isoforms was used as the ribosome-free control. As expected, the RT-qPCR analysis revealed the NEAT1 RNA was slightly more abundant in the ribosome-free fraction (Figure 3.14 (B)). Therefore, the ribosome-free and ribosome-bound fractions passed the quality control checks.

The two functional genes within the dihydrofolate reductase gene family were the next targets of the sucrose cushion ultracentrifugation RT-qPCR analysis. The original characterised *DHFR* gene was used as the dihydrofolate reductase gene family control. Again, this essential gene codes for an enzyme that supplies one-carbon units to the OCM pathway (Schnell et al., 2004). Therefore, the DHFR RNA should associate with the ribosome resulting in a high translation rate. Unexpectedly, the DHFR mRNA was present at a higher concentration in the ribosome-free fraction when compared to the ribosome-bound sample (Figure 3.14 (C)). This result does not imply that the DHFR mRNA is not being translated it just suggests that there is a higher abundance of free DHFR RNA present. The autoregulatory translation mechanism of the *DHFR* gene may explain this surprising result, this mechanism is described in Chapter 1, Section 1.5.1.2 and Figure 1.7. The DHFR protein has the ability to inhibit the translation of the DHFR mRNA (Abali et al., 2008; Ercikan et al., 1993). Therefore, there may be a large amount of free mRNA present in a cell to carry out this vital function. The translation capabilities of the *DHFR2* retrogene were subsequently investigated. An RT-qPCR assay that targeted all transcripts was initially performed. The RT-qPCR analysis showed that the DHFR2 mRNA was more abundant in the ribosome-associated fraction (Figure 3.14 (D)). This result suggests that the DHFR2 transcripts are binding to the ribosome and have the ability to be translated into a protein. This outcome is consistent with the results from the ribosome profiling on the DHFR2 RNA. However, this result does not reveal if it is all or the majority of DHFR2 RNA isoforms that bind to the ribosome. Therefore, the translational status of the two main DHFR2 transcripts were interrogated further. Two assays that target the DHFR2_201 and DHFR2_202 transcripts were executed. The RT-qPCR analysis revealed the DHFR2_201 transcript was significantly more abundant in the ribosome-bound fraction (Figure 3.14 (E)). This result suggests that the DHFR2_201 transcript has the capabilities to be synthesised into a dihydrofolate reductase protein. Therefore, proteomic analysis will need to be performed to determine if the predicted DHFR2 protein is actually being generated. The RT-qPCR analysis on the DHFR2_202 RNA isoform showed it was present at a higher concentration in the ribosome-free fraction (Figure 3.14 (F)). This result suggests that the DHFR2_202 transcript does not have the capabilities to be synthesised into a protein. Taking into account this result and the fact that the DHFR2_202 transcript is highly expressed across multiple cell lines (Section 3.3.2), this transcript may have a meaningful biological role at the RNA level. The DHFR2_202 transcript could autoregulate the translation of its own protein, the DHFR protein or other proteins within the OCM pathway. The differential association between the two main DHFR2 transcripts and the ribosome is an interesting finding. This result prompts the question “What are the differences between these two transcripts that could cause these opposing effects?”. The only obvious difference in the DHFR2_201 and DHFR2_202

transcripts is the exon 1 nucleotide sequence (Appendix F). This suggests that there is something within the DHFR2_201 exon 1 sequence that initiates translation and/or something within the DHFR2_202 exon 1 sequence that blocks translation. This could be anything from a regulatory element within one of the transcripts to the secondary structure of the RNA. Further analysis would need to be performed on both transcripts to investigate all possible hypotheses.

Chapter 4. Investigating the translatability of the human DHFR2 ORF through the generation of the recombinant DHFR2 protein

4.1. Introduction

4.1.1. Overview

Earlier work carried out in the APMcD laboratory uncovered huge difficulties with expressing high levels of recombinant human DHFR2 protein in bacteria cells. Eventually, small quantities of soluble DHFR2 protein were obtained from extremely large volumes of media (Hughes, unpublished). However, *E. coli* expression systems do not contain the same chaperone proteins as eukaryotic cells. Therefore, they are not equipped to fold eukaryotic recombinant proteins properly resulting in them sometimes not being biologically functional (Sørensen & Mortensen, 2005). Thus there is no guarantee that the enzyme made from an *E. coli* expression system is biologically active. Therefore, the DHFR2 protein generated in *E. coli* cells could only be used to a certain extent and left some research questions unanswered. For example, "does the recombinant DHFR2 protein harbour reductase activity?". For this reason, the APMcD laboratory moved on to trying to express recombinant human DHFR2 protein in HEK293 (human embryonic kidney) and CHO (Chinese hamster ovary) cells. However, the mammalian expression systems employed failed to generate the recombinant DHFR2 protein. Therefore, characterising the DHFR2 protein came to a halt (Hughes, unpublished). As methods have evolved since performing these experiments it was time to retry to synthesise the recombinant DHFR2 protein in mammalian cells.

Therefore, this chapter aimed to generate the recombinant human DHFR2 protein using two separate methods. The first approach utilised a traditional mammalian expression system with strong transcriptional and translational signals in an attempt to generate the recombinant protein. This consisted of transfecting a DHFR2 expression vector into the HepG2 DHFR2 knockout (KO) cell line. This technique is a robust method for generating recombinant protein, however, it is not fully representative of DHFR2's natural cellular environment (Oza et al., 2015). Secondly, a humanised DHFR2 mouse model that was initially generated for complementation was exploited to determine if it could generate the recombinant DHFR2 protein. This mouse model was generated by knocking out the mouse Dhfr open reading frame (ORF) and inserting the human DHFR2 coding sequence in its place. Therefore this system used the strong transcriptional and translational signals belonging to the mouse *Dhfr* gene to try to generate the human DHFR2 protein. Thus, this in vivo model may provide a more representative cellular environment to allow the human DHFR2 protein to be translated (Zhu et al., 2019). The proteome of the HepG2 DHFR2 knockout cells transfected with a DHFR2 expression vector and the DHFR2 mouse model was subsequently interrogated by liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis to establish if the recombinant DHFR2 protein was synthesised. LC-MS/MS was the preferred method to identify the DHFR2 protein

throughout this project as it can easily distinguish between the highly homologous DHFR and predicted DHFR2 protein.

4.1.2. Generation of recombinant protein and its function

A recombinant protein is a manipulated version of a native protein that is generated by a recombinant DNA technology (Sørensen & Mortensen, 2005). This is usually achieved by isolating the coding sequence belonging to the gene of interest and inserting it into an expression vector using a molecular cloning method. The newly formed vector (referred to as a recombinant clone) is introduced into the expression system and the cells machinery synthesises the recombinant protein at a high rate (Bertero et al., 2017) (Figure 4.1). Recombinant protein expression has become one of the most important techniques in biochemistry and has provided valuable breakthroughs in biomedical biotechnology (Pham, 2018). This is due to it being relatively simple, quick and inexpensive to produce large quantities of protein (Sørensen & Mortensen, 2005). Furthermore, it can be used to characterise a novel protein, from revealing its structure to identifying its overall function (Khan, 2013).

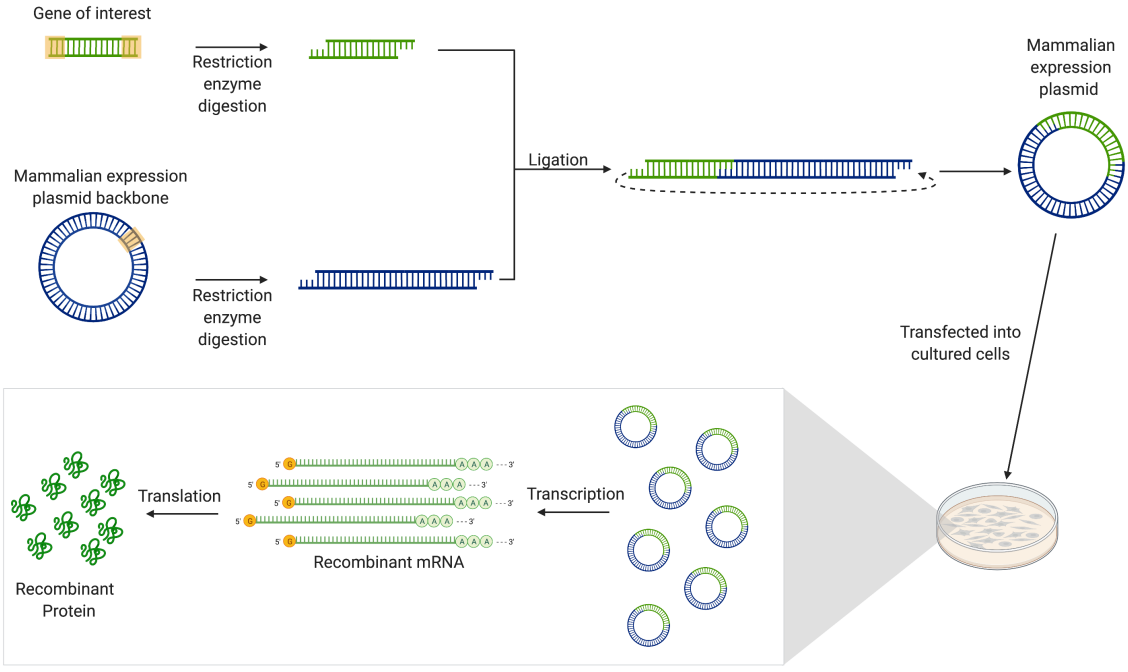


Figure 4.1 Recombinant protein expression using a mammalian expression system. *The coding sequence belonging to the gene of interest is isolated. The gene of interest fragment and the expression plasmid backbone is digested to generate sticky ends. The two fragments are ligated together to create the mammalian expression plasmid. The newly formed vector (referred to as the recombinant clone) is introduced into the expression system and the cells machinery synthesises the recombinant protein at a high rate. Created with BioRender.com, Adapted from Bertero et al., 2017.*

4.1.2.1. Summary of the different expression systems employed to generate recombinant protein

Synthesising recombinant proteins requires the use of an expression system. There are various types of hosts to choose from, including insect, bacterial, plant, yeast, algal or mammalian. Each expression system has its own respective advantages and disadvantages (Wingfield, 2015). Therefore, these need to be considered when trying to choose the correct host for the protein of interest (Khan, 2013).

4.1.2.1.1.Bacterial expression systems to synthesise recombinant protein

Bacterial expression systems are one of the most widely used systems to synthesise recombinant proteins. This is due to the low cost, short turnaround time and the high yield associated with them (Terpe, 2006). *E. coli* is the bacterial host of choice as it has an extremely quick doubling time of 20 minutes (Rosano & Ceccarelli, 2014). However, this strategy has limitations when trying to produce eukaryotic proteins. Bacteria cells do not have the same chaperone proteins as eukaryotic cells. Therefore, they are not equipped to fold the recombinant proteins properly. This causes the formation of highly aggregated proteins known as inclusion bodies. Resulting in the recombinant protein not being biologically functional. There are ways to solubilise inclusion bodies (e.g. incubate in a high concentration of denaturants or chaotropes) and refold the proteins (e.g. remove the solubilisation agent by dialysis). However, it is an extremely tricky task to perform and rarely ends successful (Sørensen & Mortensen, 2005). A second limitation arising from using bacteria cells as the expression system is that they do not have the machinery necessary to carry out post-translational modifications on eukaryotic proteins. Furthermore, there is no technique available to overcome this issue. Therefore, if the protein of interest depends on a post-translational modification to become biologically active this expression system is not a suitable option (Bhatwa et al., 2021).

4.1.2.1.2.Mammalian expression systems to synthesise recombinant protein

Mammalian expression systems are the platform of choice for producing and studying eukaryotic proteins. This is the case for performing scientific research and manufacturing biopharmaceuticals (Tripathi & Shrivastava, 2019). This is due to the eukaryotic hosts having a similar cellular environment to the protein's original cell type. In other words, eukaryotic cells have the proper chaperones to fold the recombinant proteins correctly and to perform relevant post-translational

modifications. Therefore, mammalian expression systems have the highest probability of producing a functional active eukaryotic protein (Wingfield, 2015). Multiple different mammalian cell lines have been employed over the years for recombinant protein expression. HEK293 and CHO cell lines are the most widely used to date (Khan, 2013). Even though, this method has huge advantages there are also some limitations that need to be considered when choosing a mammalian expression system. It is a complex technique that requires highly trained scientists and specialised equipment. Furthermore, mammalian cells are extremely delicate and have an exceptionally slow doubling time of 14 to 48 hours. For these reasons, using mammalian expression systems to synthesise recombinant protein is time-consuming and highly expensive (Tripathi & Shrivastava, 2019).

Taking into account both the advantages and disadvantages of using mammalian expression systems to synthesise recombinant protein, it was decided that the benefits outweigh the limitations. Therefore, a mammalian expression system was employed in this chapter to determine if the recombinant DHFR2 protein could be generated.

4.1.3. Aims and Objectives

The main aim of this chapter was to determine if a recombinant human DHFR2 protein could be synthesised in mammalian cells. To achieve this aim, the following objectives were proposed:

1. Investigate if the recombinant DHFR2 protein is synthesised when HepG2 DHFR2 knockout cells are transfected with a DHFR2 expression vector.
2. Determine if the recombinant DHFR2 protein is expressed when the human DHFR2 ORF is knocked into the mouse genome.

4.2. Method

4.2.1. Transfection of pCMV6-AC-DHFR2 vectors into HepG2 DHFR2 knockout cell line

HepG2 DHFR2 knockout cells were generated according to Chapter 2, Section 2.2.1 and grown to 70% confluency as per Chapter 2, Section 2.2.2. Three T25 flasks were seeded at a cell density of $2x10^6$ and 5 mL DMEM, supplemented with 10% (v/v) FBS. The flasks were incubated for 24 hour in a 5% carbon dioxide incubator at 37 °C. The pCMV6-AC-tGFP (control), pCMV6-AC-DHFR2-tGFP (Appendix G) and pCMV6-AC-DHFR2-DDK-Myc (Appendix H) expression vectors were transfected into individual T25 flasks using the Lipofectamine 3000 Reagent, according to the manufacturer's protocol. The transfected cells were incubated for 48 hour in a 5% carbon dioxide incubator at 37 °C. The cells were harvested and pelleted as per Chapter 2, Section 2.2.2. This work was performed by Paola Drago (DCU).

4.2.2. Generation of humanised DHFR2 mouse model

Dual CRISPR guide RNAs (gRNA) and a repair template were designed to remove the mouse Dhfr coding sequence from the mouse genome and replace it with the human DHFR2 ORF (Figure 4.2). This was performed by injecting the CRISPR gRNAs and a repair template into Embryonic stem (ES) cells derived from 129 mouse substrains. The resulting colonies were screened and positive colonies were used to generate chimeras from pregnant C57BL/6J females. Positive chimeras (presenting the 129 agouti coat colour) were then crossed to C57BL/6J mice to assess for germline transmission of the humanised allele. Subsequent positive F1 mice were then crossed to C57BL/6J mice for at least 5 generations to create stable humanised mouse lines on a C57BL/6J background. These mouse models were generated by Dr Zoe Crane-Smith (UCL).

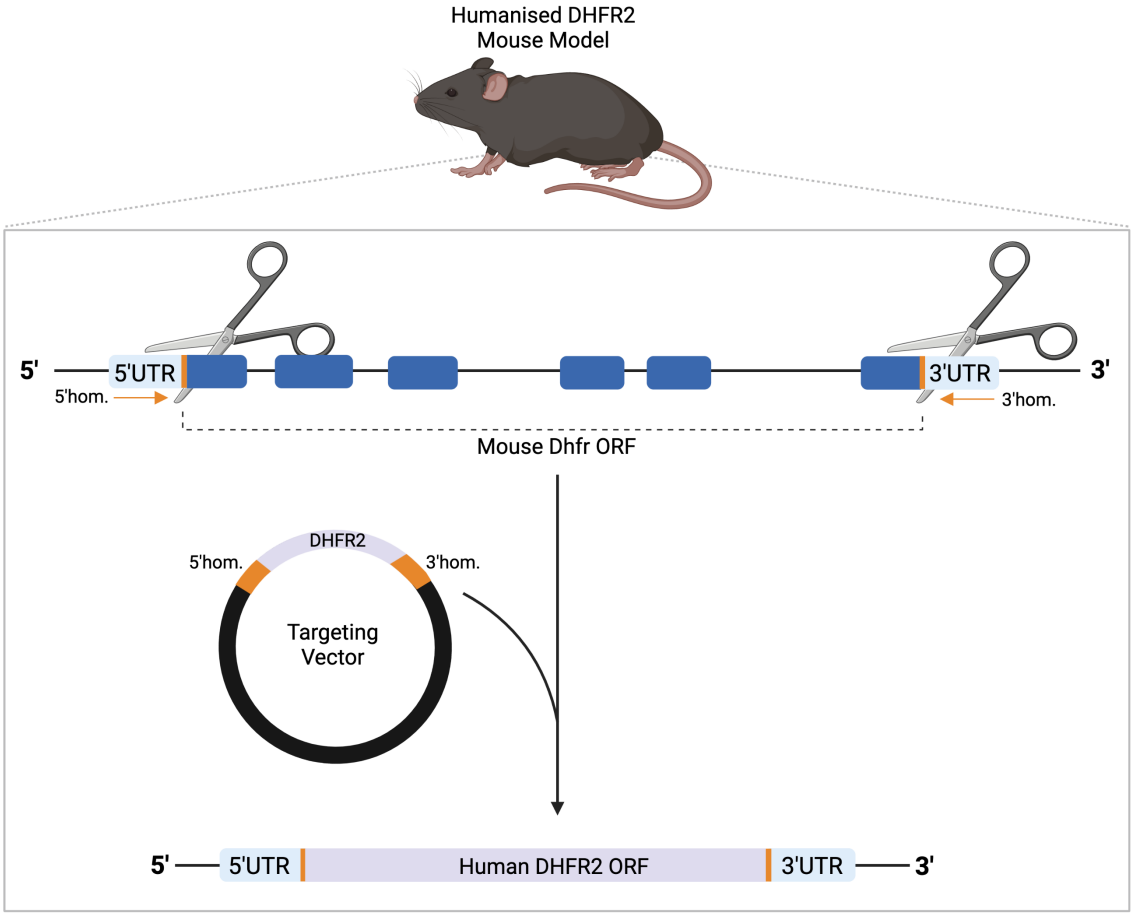


Figure 4.2 Generation of humanised DHFR2 mouse model using genome editing CRISPR-Cas9 gene-editing technology. *The coding sequence belonging to the mouse Dhfr gene was knocked out using the Cas9 endonuclease. The human DHFR2 open reading frame was inserted in its place using homology-directed repair. Created with BioRender.com.*

4.2.3. Protein extraction for Western Blot and LC-MS/MS analysis

Tissue from wildtype (C57BL/6J background) and humanised DHFR2 mice (Section 4.2.2) was isolated in UCL and sent to DCU. These samples included whole E11.5 day embryos and tissues from P20 day juvenile mice (brain, kidney, heart, liver, lung, spleen and stomach). The frozen tissue was reduced to a fine powder according to Chapter 2, Section 2.2.19.1. Whole protein lysates were extracted from these ground tissue samples and the HepG2 DHFR2 knockout cell line transfected with the pCMV6-AC-tGFP, pCMV6-AC-DHFR2-tGFP and pCMV6-AC-DHFR2-DDK-Myc vectors (Section 4.2.1), as described in Chapter 2, Section 2.2.19.2.

4.2.4. Western Blot analysis of HepG2 DHFR2 knockout cell line transfected with pCMV6-AC vectors

Western Blot analysis was carried out on the protein lysate from HepG2 DHFR2 knockout cell line transfected with the pCMV6-AC-tGFP, pCMV6-AC-DHFR2-tGFP and pCMV6-AC-DHFR2-DDK-Myc vectors, as per Chapter 2, Section 2.2.22. A total of 50 µg of protein was denatured and ran on a 4–20% Tris-glycine gel. The membrane that contained the recombinant tGFP proteins was initially incubated in a polyclonal antibody toward tGFP and subsequently probed with a monoclonal anti-rabbit HRP conjugated secondary antibody, concentrations stated in Chapter 2, Table 2.8. The membrane that contained the recombinant DDK-Myc proteins was initially incubated in a monoclonal antibody toward DDK and subsequently probed with a monoclonal rat anti-mouse HRP conjugated secondary antibody, concentrations stated in Chapter 2, Table 2.8.

4.2.5. LC-MS/MS sample preparation

Protein lysate from the HepG2 DHFR2 knockout cell line transfected with pCMV6-AC-tGFP, pCMV6-AC-DHFR2-tGFP and pCMV6-AC-DHFR2-DDK-Myc vectors and the mouse models (E11.5 day mouse embryo and P20 day juvenile mouse (brain, kidney, heart, liver, lung, spleen and stomach) from wildtype and humanised DHFR2 mice) were prepared for LC-MS/MS analysis as per Chapter 2, Section 2.2.20.1 and Section 2.2.20.2.

4.2.6. Searching for recombinant DHFR2 specific peptides using data-dependent acquisition (DDA) LC-MS/MS analysis

All the trypsinised whole protein lysate samples from Section 4.2.5 were resuspended in 15 µL of resuspension solution and then sonicated using a sonicating waterbath (Precision ultrasonic cleaning, DP201–00, Ultrawave) under the following conditions: power of 280 watts, frequency of 30 kilohertz, ambient temperature and time of 5 min. The peptide concentration was quantified using the Nanodrop One spectrophotometer (Labtech instruments), at an absorbance of 280 nm. The samples were diluted down to a working solution of 1 µg/µL using resuspension solution for DDA LC-MS/MS analysis (Chapter 2, Section 2.2.21.1) to search for a recombinant DHFR2 specific peptide.

4.2.7. Searching for recombinant DHFR2 specific peptides using parallel reaction monitoring (PRM) LC-MS/MS analysis

The 1 µg/µL working stock of the trypsinised whole protein lysate samples from Section 4.2.6 were also subject to PRM DHFR2 LC-MS/MS analysis (Chapter 2, Section 2.2.21.2) to search for a recombinant DHFR2 specific peptide.

4.3. Results

4.3.1. Recombinant DHFR2 protein was synthesised in HepG2 DHFR2 knockout cell line

A mammalian cell-based expression system was employed to try to generate the recombinant DHFR2 protein. Prior attempts at using mammalian cell lines to generate the recombinant protein were unsuccessful, this included HEK293 and CHO cell lines. For this reason, a different cell line was used, the HepG2 DHFR2 knockout cell line (Section 4.2.1). Furthermore, two different sized C-terminal fusion tags were employed, a turboGFP (tGFP) and a DDK-Myc tag.

The pCMV6-AC-DHFR2-tGFP (Appendix G) and pCMV6-DHFR2-DDK-Myc (Appendix H) mammalian expression vectors were transfected into the HepG2 DHFR2 knockout cell line as per Section 4.2.1. The resulting protein lysates were assessed for the recombinant DHFR2 protein. Western Blotting using antibodies against the DHFR2 protein could not be employed as anti-DHFR2 antibodies are prone to cross-react with the DHFR protein; resulting in false-positive results. However, Immunoblotting using antibodies against the specific fusion tag could be employed to detect the protein complex. Western Blot analysis was performed on the protein lysate from the HepG2 DHFR2 knockout cell line transfected with the pCMV6-AC-tGFP and pCMV6-AC-DHFR2-tGFP expression vector using an anti-tGFP antibody, according to Section 4.2.4 (Figure 4.3). The HepG2 DHFR2 knockout cell line transfected with the pCMV6-AC-tGFP control vector should generate the tGFP protein (26 kDa). An intense band of this size was visualised in this sample. An additional band around 18 kDa was also detected in this sample. This is a common occurrence with tGFP and is a result of the protein having different folding states (Geertsma et al., 2008). The HepG2 DHFR2 knockout cell line transfected with the pCMV6-AC-DHFR2-tGFP vector should generate the DHFR2-tGFP protein (47 kDa). The Immunoblot in Figure 4.3 showed a distinct band at this size, confirming the presence of the recombinant DHFR2-tGFP protein.

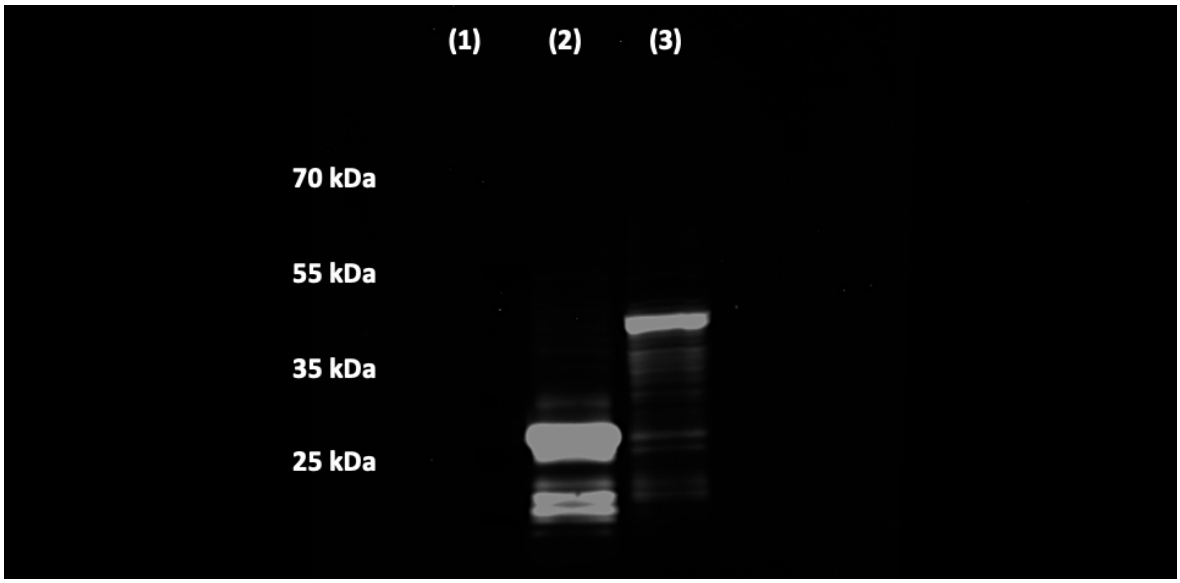


Figure 4.3 Western Blot analysis of recombinant tGFP and DHFR2-tGFP using an anti-tGFP antibody. *The pCMV6-AC-tGFP and pCMV6-AC-DHFR2-tGFP mammalian expression vectors were transfected into the HepG2 DHFR2 knockout cell line. The resulting proteins (50 µg) were separated on a 4-20% tris-glycine gel and visualised on a chemiluminescence Western Blot, probed with an anti-tGFP primary antibody and anti-mouse secondary antibody. (1) Protein ladder, (2) HepG2 DHFR2 knockout cell line plus pCMV6-AC-tGFP control vector, and (3) HepG2 DHFR2 knockout cell line plus pCMV6-AC-DHFR2-tGFP vector. The recombinant tGFP protein generated from the pCMV6-AC-tGFP expression vector should result in a band at 26 kDa; an intense band of this size was visualised in Lane 2. An additional band around 18 kDa was also detected in this sample. The recombinant DHFR2-tGFP protein generated from the pCMV6-AC-DHFR2-tGFP expression vector should result in a band at 47 kDa; an intense band of this size was visualised in Lane 3.*

Western Blot analysis was also carried out on the protein lysate from the HepG2 DHFR2 knockout cell line transfected with the pCMV6-AC-DHFR2-DDK-Myc expression vector using an anti-DDK antibody, as per Section 4.2.4 (Figure 4.4). The recombinant DHFR2-DDK-Myc protein should result in a band at 27 kDa. Figure 4.4 showed an extremely faint band at that size, confirming the presence of the recombinant DHFR2-DDK-Myc protein.

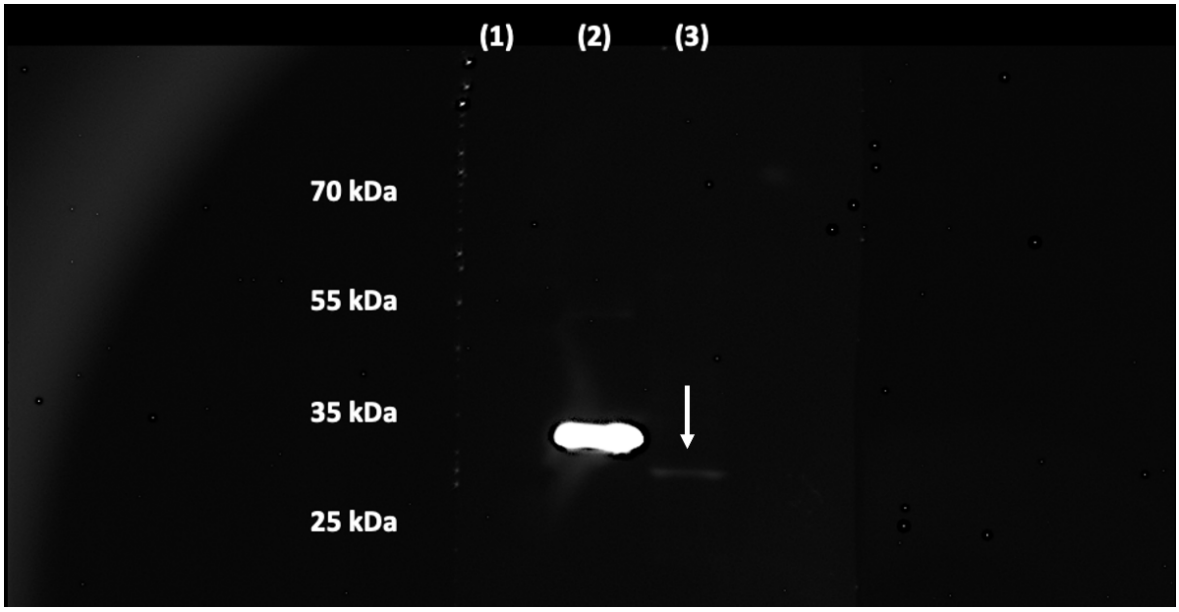


Figure 4.4 Western Blot analysis of recombinant tGFP-DDK-Myc and DHFR2-DDK-Myc using an anti-DDK antibody. *The pCMV6-DHFR2-DDK-Myc mammalian expression vector was transfected into the HepG2 DHFR2 knockout cell line. The resulting proteins (50 µg) were separated on a 4-20% tris-glycine gel and visualised on a chemiluminescence Western Blot, probed with an anti-DDK primary antibody and anti-mouse secondary antibody. (1) Protein ladder, (2) Purified recombinant tGFP-DDK-Myc protein, and (3) HepG2 DHFR2 knockout cell line plus pCMV6-DHFR2-DDK-Myc vector. The purified tGFP-DDK-Myc protein should be approximately 32 kDa; a band of this size was detected in Lane 2. The recombinant DHFR2-DDK-Myc protein generated from the pCMV6-DHFR2-DDK-Myc expression vectors should result in a band at 27 kDa. A very faint band of this size was visualised in Lane 3, pointed out by the white arrow.*

LC-MS/MS was the preferred method to identify the DHFR2 protein in this project; this is due to it being able to easily distinguish between the highly homologous DHFR and predicted DHFR2 protein. To be consistent throughout this thesis, data-dependent acquisition LC-MS/MS analysis was carried out on the protein lysate from the HepG2 DHFR2 knockout cell line transfected with both the DHFR2 expression vectors to confirm the results obtained in Figure 4.3 and Figure 4.4. The DDA LC-MS/MS methodology is discussed in greater detail in Chapter 5. DDA LC-MS/MS analysis was initially performed on protein lysate from the HepG2 DHFR2 knockout cell line. This analysis was carried out to confirm the absence of the endogenous DHFR2 protein in the HepG2 DHFR2 knockout cell line. The DDA LC-MS/MS analysis on this sample did not detect any DHFR2 specific peptides. It did reveal one peptide unique to the DHFR protein (Table 4.1). This result confirmed the presence of

only the DHFR protein in the HepG2 DHFR2 knockout cell line. DDA LC-MS/MS analysis was then carried out on the protein lysate from the HepG2 DHFR2 knockout cell line transfected with the pCMV6-AC-DHFR2-tGFP and the pCMV6-AC-DHFR2-DDK-Myc expression vectors (Table 4.1). The instrument identified peptides shared by the human DHFR and DHFR2 proteins in both samples. The mass spectrometer also detected two peptides unique to the DHFR2 protein in the HepG2 DHFR2 knockout cell line transfected with the pCMV6-AC-DHFR2-tGFP vector (Table 4.1). This result confirmed the presence of the recombinant DHFR2-tGFP in this sample. The DDA LC-MS/MS analysis on the protein lysate from the HepG2 DHFR2 knockout cell line transfected with the pCMV6-AC-DHFR2-DDK-Myc vector did not recognise any peptides that were unique to the DHFR2 or the DHFR protein (Table 4.1); therefore, it could not be concluded which protein the MTTTSSVEGK peptide belonged to. A more sensitive targeted LC-MS/MS method was employed to interrogate this sample for the recombinant DHFR2-DDK-Myc protein.

Table 4.1 Dihydrofolate reductase peptides detected in the DDA LC-MS/MS analysis from HepG2 DHFR2 knockout cells transfected with DHFR2-tGFP Vector and DHFR2-DDK-Myc Vector

Peptide Sequence	Modifications	Protein Name	DHFR2 KO	DHFR2 KO & DHFR2-tGFP Vector	DHFR2 KO & DHFR2 -DDK-Myc Vector
MTTTSSVEGK	-	DHFR/DHFR2	x	✓	✓
INLVLSR	-	DHFR/DHFR2	x	✓	x
LTEQPELANK	-	DHFR	✓	x	x
EAMNHLGHLK	-	DHFR2	x	✓	x
LTERPELANK	-	DHFR2	x	✓	x

The targeted PRM DHFR2 workflow described in Chapter 2, Section 2.2.21.2 was employed to interrogate protein lysate from the HepG2 DHFR2 knockout cell line transfected with the pCMV6-AC-DHFR2-DDK-Myc vector for the recombinant DHFR2-DDK-Myc protein. The PRM LC-MS/MS methodology is discussed in much greater detail in Chapter 5. This PRM workflow was adapted to target the two DHFR2 specific peptides identified in the HepG2 DHFR2 knockout cell line transfected with the pCMV6-AC-DHFR2-tGFP plasmid (Table 4.1); this included the EAMNHLGHLK and LTERPELANK DHFR2 specific peptides. The characterisation of these peptides using a synthetic version of the peptide (described in Chapter 2, Section 2.2.21.2) was not required as the DDA LC-MS/MS analysis had revealed the *m/z*, charge state and retention time for these peptides. This information was used to adapt the PRM DHFR2 workflow to identify only these two peptides at their specific *m/z*, charge state and retention time window. The PRM DHFR2 LC-MS/MS workflow

was tested using the trypsinised cell lysate from the HepG2 DHFR2 knockout cell line transfected with the pCMV6-AC-DHFR2-tGFP vector. There were three MS/MS responses triggered when analysing this sample, confirming the presence of two DHFR2 specific peptides and one post-translationally modified version (Table 4.2 and Figure 4.5). Again, verifying the presence of the recombinant DHFR2-tGFP protein. Furthermore, confirming the PRM DHFR2 LC-MS/MS workflow was successful at identifying the two DHFR2 specific peptides. Subsequently, protein lysate from the HepG2 DHFR2 knockout cell line was then assessed using the PRM DHFR2 LC-MS/MS workflow to again confirm the absence of the endogenous DHFR2 protein in this cell line. The PRM LC-MS/MS analysis on this sample did not detect any DHFR2 specific peptides (Table 4.2). Therefore, the result from both LC-MS/MS experiments showed no evidence of the endogenous DHFR2 protein in the HepG2 DHFR2 knockout cell line. Finally, the protein lysate from the HepG2 DHFR2 knockout cell line transfected with the pCMV6-AC-DHFR2-DDK-Myc vector was assessed. There was one MS/MS response triggered when analysing this sample, meaning one of the DHFR2 specific peptides were detected (Table 4.2 and Figure 4.5). Therefore, this targeted experiment confirmed the presence of the recombinant DHFR2-DDK-Myc in this sample.

Table 4.2 Dihydrofolate reductase peptides detected in the PRM LC-MS/MS analysis from HepG2 DHFR2 knockout cells transfected with DHFR2-tGFP vector and DHFR2-DDK-Myc vector

Peptide Sequence	Modifications	Protein Name	DHFR2 KO	DHFR2 KO & DHFR2-tGFP Vector	DHFR2 KO & DHFR2-DDK-Myc Vector
EAMNHLGHLK	-	DHFR2	x	✓	x
EAmNHLGHLK	Oxidation (M)	DHFR2	x	✓	x
LTERPELANK	-	DHFR2	x	✓	✓

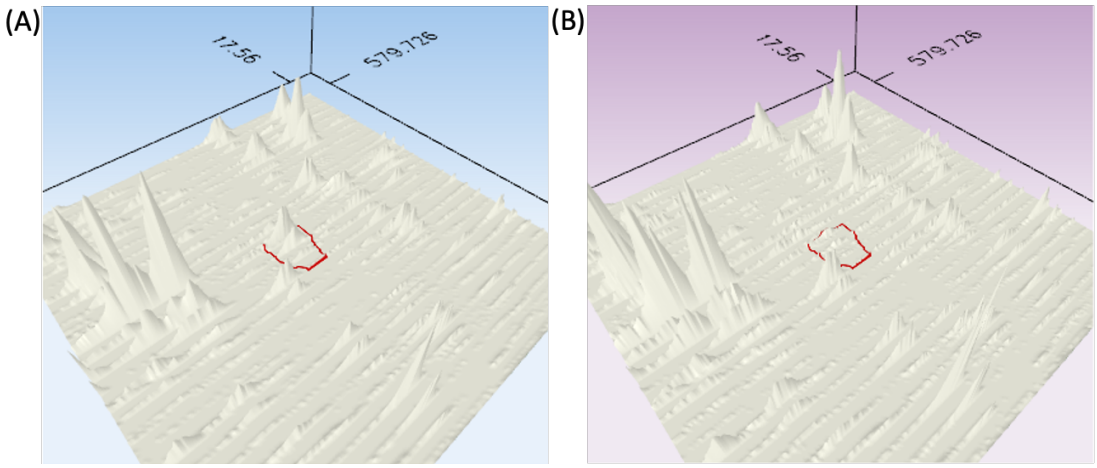


Figure 4.5 Three-dimensional distribution plots from the PRM LC-MS/MS analysis on the HepG2 DHFR2 knockout cells transfected with the pCMV6-AC-DHFR2-tGFP and pCMV6-DHFR2-DDK-Myc vectors. *The pCMV6-AC-DHFR2-tGFP and pCMV6-DHFR2-DDK-Myc mammalian expression vectors were transfected into the HepG2 DHFR2 knockout cell line. The resulting proteins were analysed using the PRM LC-MS/MS method. If any of the unique DHFR2 peptides were detected in the sample an MS/MS response would occur and the peaks would be recorded. (A) LC-MS/MS Spectrum results from the HepG2 DHFR2 knockout cells transfected with the pCMV6-AC-DHFR2-tGFP expression vector. An MS/MS response occurred and two unique recombinant DHFR2 peptides were detected. (B) LC-MS/MS Spectrum results from the HepG2 DHFR2 knockout cells transfected with the pCMV6-DHFR2-DDK-Myc expression vector. An MS/MS response occurred and one unique recombinant DHFR2 peptide was identified.*

4.3.2. Recombinant DHFR2 protein was not expressed in humanised DHFR2 mouse model

Phylogenetic analysis revealed that the RNA-mediated DHFR duplication event that gave rise to the *DHFR2* gene in humans occurred across the mammalian tree. However, there was no evidence of an intact reductase ORF or RNA expression in mice. Therefore, it is likely that the *DHFR2* gene in mice is a processed pseudogene (Hughes et al., 2015). The Nervous System Development Research Group at UCL generated a humanised DHFR2 mouse model where they knocked out the mouse Dhfr ORF and inserted the human DHFR2 coding sequence in its place (Section 4.2.2 and Figure 4.2). This mouse model was generated to determine whether the expression of the *DHFR2* gene could complement the mouse *Dhfr* gene. This mouse model was exploited to also determine whether the recombinant human DHFR2 protein could be synthesised using the strong transcriptional and translational signals from the mouse *Dhfr* gene.

The homozygote humanised DHFR2 mouse models were embryonically lethal (Greene, personal communication). For this reason, all subsequent research was performed on the heterozygote mice models i.e., mice containing one copy of the mouse *Dhfr* and one copy of the human *DHFR2* allele.

Antibody-based methods could not be employed to assess the humanised DHFR2 mouse tissue for the recombinant human DHFR2 protein. This is due to the similarity between the amino acid sequence of the human DHFR2 and mouse Dhfr protein. Resulting in commercially available antibodies for either protein cross-reacting. Therefore, the LC-MS/MS methods described in Section 4.3.1 and Chapter 5 were employed to analyse the proteome of a number of tissues belonging to the humanised DHFR2 mice and wildtype mice, Section 4.2.6 - 4.2.7. These samples included whole E11.5 day embryos and tissues from P20 day juvenile mice (brain, kidney, heart, liver, lung, spleen and stomach).

First off, DDA and PRM LC-MS/MS analysis was carried out on the wild-type mouse tissue (Table 4.3 and Table 4.4). This analysis did not detect any peptide unique to the DHFR2 protein, confirming that wildtype mice do not synthesise the native version of the protein. DDA LC-MS/MS analysis was then carried out on the humanised DHFR2 mouse tissue (Table 4.3 and Table 4.4). The instrument identified peptides shared by the mouse Dhfr and human DHFR2 proteins. These peptides were found in the whole embryo, kidney, liver and spleen tissue. However, the instrument did not recognise any peptides that were unique to the DHFR2 protein in any of the tissues. Moreover, the mass spectrometer did identify a number of peptides with a sequence exclusive to the mouse Dhfr protein. Confirming the presence of the Dhfr protein in the embryo, kidney, liver, spleen and stomach tissue. There were no dihydrofolate reductase peptides identified in the brain, heart and lung tissue of the postnatal mouse. Even though other OCM proteins were detected including Mthfd1, Mthfd1l and Shmt2. The PRM DHFR2 workflow described in Section 4.3.1 was used to interrogate the humanised DHFR2 mouse tissue further for the recombinant DHFR2 specific peptides. The PRM LC-MS/MS analysis did not detect the recombinant DHFR2 peptides in the humanised DHFR2 mouse tissues. Therefore, both LC-MS/MS methods did not detect the recombinant DHFR2 protein in tissues belonging to the humanised DHFR2 mouse model.

Table 4.3 Dihydrofolate reductase peptides detected in the DDA LC-MS/MS analysis from wildtype and humanised DHFR2 embryo mouse tissue

Peptide Sequence	Modifications	Protein Name	Wildtype embryo mouse tissue	Humanised DHFR2 embryo mouse tissue
MTTTSSVEGK	-	Dhfr/DHFR2	✓	✓
QNLVIMGR	-	Dhfr/DHFR2	✓	✓
TWFSIPEK	-	Dhfr/DHFR2	✓	X
VDMVWIVGGSSVYQEAMNQPGHLR	-	Dhfr	✓	X
LLPEYPGVLSEVQEEK	-	Dhfr	✓	✓
LIEQPELASK	-	Dhfr	✓	✓
FEVYEK	-	Dhfr	✓	X
YKFEVYEK	-	Dhfr	✓	X
SLDDALR	-	Dhfr	✓	X

Table 4.4 Dihydrofolate reductase peptides detected in the DDA LC-MS/MS analysis from wildtype postnatal mouse and humanised DHFR2 postnatal mouse tissue

Peptide Sequence	Modifications	Protein Name	Kidney		Liver		Spleen		Stomach	
			WT	H	WT	H	WT	H	WT	H
MTTTSSVEGK	-	Dhfr/DHFR2	X	X	✓	X	✓	✓	X	X
QNLVIMGR	-	Dhfr/DHFR2	✓	✓	✓	✓	✓	✓	✓	X
TWFSIPEK	-	Dhfr/DHFR2	X	✓	X	X	✓	X	X	X
NGDLPWPPLR	-	Dhfr	✓	✓	X	X	X	X	X	X
LLPEYPGVLSEVQEEK	-	Dhfr	✓	✓	✓	✓	✓	✓	✓	✓
IMQEFESDTFFPEIDLGK	-	Dhfr	✓	X	✓	X	X	X	X	X
VDMVWIVGGSSVYQEAMNQPGHLR	-	Dhfr	✓	X	X	X	✓	X	X	X
LIEQPELASK	-	Dhfr	✓	✓	✓	✓	X	X	✓	X
YKFEVYEK	-	Dhfr	✓	X	X	X	X	X	X	X
MVRPLNCIVAVSQNMGIGK	-	Dhfr	X	X	✓	X	X	X	X	X
INIVLSR	-	Dhfr	X	X	✓	X	X	X	X	X

WT= wildtype postnatal mouse tissue, H= humanised DHFR2 postnatal mouse tissue

4.4. Discussion

A human cell-based expression system was employed in an attempt to generate recombinant DHFR2 protein. Prior attempts at using human cell lines were unsuccessful, this included HEK293 and CHO cell lines. Therefore, a cell line with a different tissue origin was selected, the HepG2 (liver hepatocellular carcinoma) cell line. Furthermore, the DHFR2 knockout version of this cell line was employed to eliminate the expression of the *DHFR2* gene that might have interfered with the end result. A pCMV6-AC mammalian expression vector was chosen as it has extremely strong signals to initiate transcription and translation. The cytomegalovirus (CMV) enhancer and promoter is one of the strongest DNA element used to generate recombinant mRNA in mammalian cells (Kim et al., 2004). A combination of a 5'-terminal 7-methylguanosine cap, a 3'-terminal poly(A) tail and a Kozak consensus sequence ensures the recombinant mRNA is efficiently translated into recombinant protein. Therefore, by employing robust transcriptional and translational signals there should be high levels of recombinant protein being expressed (Kim et al., 2004). The addition of a fusion tag to the protein allows for easy identification and purification of the recombinant protein (Riggs, 2013). Two different sized C-terminal fusion tags were tested to eliminate them as the possible expression issue, tGFP (26 kDa) and DDK-Myc (6 kDa) tag.

The resulting proteins from the HepG2 DHFR2 knockout cells transfected with the pCMV6-AC-DHFR2-tGFP and pCMV6-DHFR2-DDK-Myc expression vectors were assessed for the recombinant DHFR2 protein (Section 4.3.1). Western Blot analysis using antibodies against the fusion tags confirmed the presence of the recombinant DHFR2-tGFP (Figure 4.3) and DHFR2-DDK-Myc proteins (Figure 4.4). It appears from these Immunoblots that the DHFR2-tGFP protein was a lot more abundant than the DHFR2-DDK-Myc protein. DDA and PRM LC-MS/MS analysis was employed to confirm the results from the Western Blots (Section 4.3.1). DDA LC-MS/MS analysis identified the recombinant DHFR2 protein in the HepG2 DHFR2 knockout cells transfected with the pCMV6-AC-DHFR2-tGFP. Similar analysis on the pCMV6-AC-DHFR2-DDK-Myc clone did not detect the recombinant DHFR2 protein (Table 4.1). Therefore, the PRM LC-MS/MS method was employed to interrogate the pCMV6-AC-DHFR2-DDK-Myc clone further for the DHFR2 protein. This technique successfully detected the recombinant DHFR2 protein in the HepG2 DHFR2 knockout cells transfected with the pCMV6-AC-DHFR2-DDK-Myc vector (Table 4.2). This result showed that the PRM LC-MS/MS approach was much more sensitive than the DDA LC-MS/MS method. Furthermore, this set of experiments revealed that the recombinant DHFR2 protein can be synthesised when there are extremely strong artificial signals to initiate transcription and translation. The generation of the recombinant human DHFR2 protein in a eukaryotic expression system is a huge breakthrough

for this project. The recombinant protein can be used as a control for subsequent analysis on the endogenous DHFR2 protein. Furthermore, it will aid in characterising the DHFR2 protein and ultimately help determine the function of the human *DHFR2* gene.

The use of expression systems to synthesise recombinant protein is one of the most important techniques in biochemistry and biotechnology (Pham, 2018). This is due to it being relatively simple, quick and inexpensive to produce large quantities of protein (Sørensen & Mortensen, 2005). Furthermore, it can be used to characterise a novel protein, from revealing its structure to identifying its overall function (Khan, 2013). However, the environment in which the recombinant protein is generated is not fully representative of the natural cellular environment. Therefore, the generation of the recombinant DHFR2 protein using strong artificial signals does not reveal much information about the translatability of the DHFR2's ORF. Thus a second approach was employed to determine whether the recombinant DHFR2 protein could be synthesised in a similar cellular environment to its native one. A humanised DHFR2 mouse model that was initially generated for complementation was exploited to answer this research question. The humanised DHFR2 mouse model was generated by knocking out the mouse Dhfr ORF and inserting the human DHFR2 coding sequence in its place. Therefore this system used the strong transcriptional and translational signals belonging to the mouse *Dhfr* gene to try to generate the human DHFR2 protein.

In the process of trying to generate the homozygote DHFR2 mouse models, it was revealed that all the embryos were lethal (Greene, personal communication). This result suggests that the human DHFR2 protein cannot substitute for the mouse Dhfr protein. Therefore, with little to no dihydrofolate reductase activity in the mice, the OCM pathway would be disrupted. This results in a severe deficiency in the cellular pools of thymidylate and purines. This hinders nucleic acid synthesis which is required for the rapidly proliferating stages of embryogenesis, resulting in the embryo not developing properly and eventually dying. The heterozygote DHFR2 embryos were viable and their phenotypes were similar to that of the wildtype mice. All subsequent research was performed on the heterozygote mice models for these reasons.

Even though the human DHFR2 protein could not compensate for the loss of the mouse Dhfr enzyme it did not mean the DHFR2 protein was not being expressed. Therefore, the next logical step was to determine if the heterozygote DHFR2 mice were synthesising the recombinant DHFR2 protein. DDA and PRM LC-MS/MS analysis was performed on protein lysate from numerous tissues belonging to the heterozygote DHFR2 embryos and juvenile mice (Section 4.3.2). This analysis found no proof of the recombinant DHFR2 protein in any of the tissue samples examined (Table 4.3 and Table 4.4). This result implied that the strong transcriptional and translational signals belonging to

the mouse *Dhfr* gene cannot synthesise the recombinant DHFR2 protein. There may be a regulator element or signal located within the DHFR2 ORF repressing the expression of the DHFR2 protein in cellular environments that is similar to its native one.

Taking into account all the results from this chapter, it suggests that there may be a regulator element or signal within the DHFR2 coding sequence blocking the expression of the DHFR2 protein. Furthermore, the translational signals belonging to the mouse *Dhfr* gene were not strong enough to reverse this repression to synthesise the recombinant human DHFR2 protein. As the cellular environment for the humanised DHFR2 mouse model is similar to the human DHFR2's native environment, this result may also suggest that the synthesis of the endogenous human DHFR2 protein is also repressed. Further analysis will need to be carried out on the *DHFR2* gene to determine if this is the case for the endogenous DHFR2 protein. Nevertheless, the extremely strong artificial translational signals in the mammalian expression system appeared to be sufficient to synthesise the recombinant human DHFR2 protein. The generation of the recombinant human DHFR2 protein in a eukaryotic expression system is a huge step forward for this project. The characterisation experiments that were previously halted from failure to generate recombinant human DHFR2 protein in a eukaryotic expression system can resume. These experiments have the potential to determine the cellular function of the human *DHFR2* gene.

Chapter 5. Interrogating the endogenous human DHFR2 protein

5.1. Introduction

5.1.1. Overview

The APMcD laboratory previously carried out liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis on human embryonic kidney 293 (HEK293) and hepatocellular carcinoma (HepG2) cells. This analysis did not detect peptides unique to the endogenous human DHFR2 protein (Hughes, unpublished). Moreover, Chapter 4 hypothesised that the expression of the native DHFR2 protein may be translationally repressed. This information suggests that the *DHFR2* gene may not encode the predicted dihydrofolate reductase protein. However, Chapter 3 confirmed that at least one of the DHFR2 transcripts binds to the ribosome and therefore has the capabilities to be translated into a protein. Therefore, the next logical step for this project was to perform high definition, focused analysis to determine if the predicted endogenous DHFR2 protein is expressed. Two different LC-MS/MS approaches were employed to search for the native DHFR2 protein. Data-dependent acquisition (DDA) LC-MS/MS analysis was initially carried out to examine the entire proteome of cell lines/tissues and to look for DHFR2 specific peptides. A targeted data acquisition (TDA) LC-MS/MS technique was also pursued to only target peptides unique to the DHFR2 protein. The difficulties associated with detecting the endogenous DHFR2 protein suggested that the DHFR2 protein could be cell/tissue-specific, only present under particular cellular conditions and/or a low abundance protein (Hughes, unpublished). Thus, the proteome of eight cell lines, six tissues and two different cellular contexts were examined for the DHFR2 protein using the two LC-MS/MS approaches. The majority of cell lines and tissue types that were chosen were similar to embryonic/foetal tissue. This was due to a genetic association study that found an association between 3 single nucleotide polymorphisms (SNPs) within the *DHFR2* gene and the risk of developing neural tube defects (NTDs) (Pangilinan et al., 2021). The findings from this study suggest that the *DHFR2* gene may play an important role during the development stages. (Pangilinan et al., 2021).

5.1.2. The proteome and how it is studied

In 1996 Marc Wilkins defined the proteome as the overall set of proteins expressed within a cell, tissue or organism at a given time (Wilkins et al., 1996). This blanket term includes all protein isoforms and all post-translationally modified versions of a protein. For this reason, the proteome is significantly larger than the genome (Garrels, 2001). In addition, the proteome within a cell is

constantly changing in response to both internal and external stimuli. Therefore, the proteome is extremely complex and a challenge to study (Tyers & Mann, 2003).

Proteomics is defined as the large-scale study of the proteome. This term refers to a wide range of techniques that aims to perform protein identification, protein quantification, protein function, protein localization, protein modification, protein structure or protein-protein interaction studies (Cho, 2007). The conventional techniques used to carry out proteomic analysis are categorised into antibody-based methods (e.g. ELISA and Western Blotting), gel-based methods (e.g. two-dimensional gel electrophoresis) and chromatography-based methods (e.g. ion-exchange chromatography). These traditional methods are extremely time-consuming and labour-intensive for the quantity of data gathered. In addition, the specificity and sensitivity of these techniques are not adequate to carry out certain types of research (Aslam et al., 2017). Thus, high throughput methods have been developed to overcome these issues. Over the last decade, protein microarray technology has become one of the most powerful tools to study the proteome (Chandramouli & Qian, 2009). However, this technique is limited by the availability of good quality antibodies against the protein of interest (Stoevesandt et al., 2014; Sutandy et al., 2013). Therefore, mass spectrometry (MS) has become the most widely-used method for analysing the proteome as it does not depend on commercially available antibodies (Pusch et al., 2003)

5.1.3. MS as a technique for identifying the composition of a sample

MS is an analytical technique used to identify and characterise molecules present in a given sample (Pusch et al., 2003). The mass spectrometer does this by ionising the components within a sample and identifying them based on their mass-to-charge ratio (*m/z*) (Nakorchevsky & Yates, 2012). MS is not sufficient when analysing an extremely complex sample. This is due to the possibility of several molecules within a sample having a comparable mass imprint and fragmentation pattern, resulting in a false-positive result (Aslam et al., 2017; Busch, 2003). To overcome this limitation a separation step known as high-performance liquid chromatography (HPLC) can be performed prior to MS analysis. This whole technique is commonly known as liquid chromatography-mass spectrometry (LC-MS). LC-MS provides superior specificity and sensitivity compared to just MS (Pitt, 2009). Therefore, LC-MS is the traditional method used to separate, identify and characterise molecules present in a given sample (Grebe & Singh, 2011).

There are three main LC-MS approaches used to assess the proteome: bottom-up which examines enzymatically digested peptides; top-down which assesses whole proteins; and middle-down which

studies partially digested proteins (Zhang et al., 2013). Bottom-up LC-MS is the most frequently used approach out of all the strategies for protein identification and characterisation (Amunugama et al., 2013).

5.1.4. Shotgun LC-MS method for identifying the composition of a sample

The term "shotgun LC-MS" refers to bottom-up LC-MS analysis performed on an extremely complex sample (Yates, 1998). This method is summarised in Figure 5.1. Whole protein lysate is extracted from a cell line/tissue and digested into peptides using a protease, such as trypsin (Giansanti et al., 2016). The resulting peptides are injected into the HPLC instrument which separates them based on their polarity or size (Mant et al., 2007). The peptide mixture is subsequently transported to the mass spectrometer. This instrument has three different components: an ion source, a mass analyser and a detector system. The peptides are initially converted into ions in the ion source by matrix-assisted laser desorption/ionisation (MALDI), surface-enhanced laser desorption/ ionisation (SEDI), or electrospray ionisation (ESI). The charged ions are drawn into a mass analyser and separated based on their *m/z*. There are two different types of mass analysers: scanning and ion-beam (e.g. Quadrupole and Time of flight) and trapping (e.g. Ion trap, Fourier transform ion cyclotron resonance and Orbitrap) (Zhang et al., 2010). The detector system records the experimental spectra. Established databases (e.g. Mascot, SEQUEST or Andromeda) compare the experimental spectrum with theoretical ones to identify peptides within the sample (Wang & Wilson, 2013).

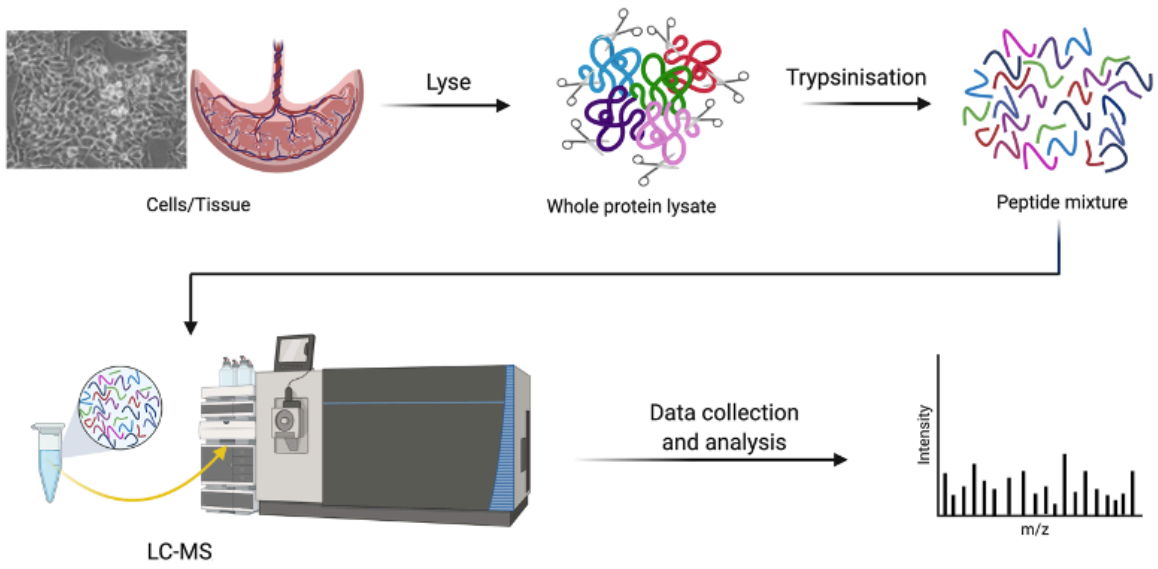


Figure 5.1 Schematic of Shotgun LC-MS. Protein lysate is isolated from cells/tissue and subjected to trypsinisation. *The resulting peptides are separated using an HPLC instrument and subsequently analysed using a mass spectrometer. The data is collected and analysed using robust proteomic software. Created with BioRender.com, Adapted from Nättinen et al., 2021.*

5.1.5. LC-MS/MS method for identifying the composition of a sample

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) is a similar technique to LC-MS; the only difference is that LC-MS/MS performs at least two mass analysis steps on an individual sample, with a dissociation step performed in between (Büyükköroğlu et al., 2018). The addition of a second mass analyser improves the specificity and sensitivity. Therefore, LC-MS/MS has become the gold standard method when examining extremely complex protein samples (Han et al., 2008).

A typical protocol for LC-MS/MS consists of the following steps. The peptide mixture is separated by HPLC and converted into ions by an ion source. The ions are drawn into the first mass analyser which carries out a full scan on them and determines which ions proceed to the fragmentation step (Adaway et al., 2015; Smith, 2013). There are multiple approaches that can be employed in this step to decide which ions proceed for fragmentation, including DDA (discussed in Section 5.1.5.1) and TDA (discussed in Section 5.1.5.2) LC-MS/MS. The ions selected by the first mass analyser are subjected to fragmentation by collision-induced dissociation (CID), high energy collision dissociation (HCD) or electron transfer dissociation (ETD) (Shen et al., 2012). The fragmented ions are drawn into the second mass analyser where they are separated and analysed to identify the proteins within the sample (Adaway et al., 2015; Smith, 2013).

5.1.5.1. DDA LC-MS/MS

The DDA LC-MS/MS approach is a relatively simple, quick and inexpensive technique used to examine peptides in a sample. It consists of the first mass analyser carrying out a full scan on the ions and only allowing ions within a certain range of precursor intensity and/or charge state to proceed to the fragmentation step (Rauniyar et al., 2017). The preselected fragmented ions enter the second mass analyser where they are separated and identified (Adaway et al., 2015; Smith, 2013).

Given that the DDA LC-MS/MS approach is relatively straightforward it was one of the techniques that was employed to examine a panel of cell lines and tissues for the endogenous DHFR2 protein.

5.1.5.2. TDA LC-MS/MS

The TDA LC-MS/MS method is a more complex approach compared to the DDA LC-MS/MS method. It consists of the first mass analyser carrying out a full scan of the ions and only allowing ions with specific characteristics to proceed to the fragmentation step. The preselected fragmented ions are drawn into the second mass analyser where they are separated and assessed (Adaway et al., 2015; Smith, 2013).

Given that the TDA LC-MS/MS method only retains ions with specific characteristics it is vital that the peptide information inputted into the workflow is correct. Therefore, TDA LC-MS/MS analysis requires several optimisation and validation steps (Heywood et al., 2016). An example of a full TDA LC-MS/MS method is summarised in Figure 5.2. This TDA LC-MS/MS technique has three main steps. Firstly, the peptide(s) of interest are characterised by running the trypsinised recombinant or synthetic version of the peptide through the mass spectrometer in discovery mode. The *m/z*, charge state and retention time recorded for the peptide(s) are used to program the mass spectrometer to identify only these peptide(s) at their specific characteristics. Secondly, the workflow is tested by analysing an enzymatically digested whole protein lysate containing the trypsinised recombinant protein or the synthetic peptide. The recombinant or synthetic version of the peptide must be detected in the sample to move on to the last step. Finally, the targeted workflow is employed to identify the endogenous version of the peptide(s) in trypsinised whole protein lysate. An MS/MS response will occur if the peptide(s) are identified in the sample, whereas there will be no MS/MS response if the peptide(s) are not detected (Carapito et al., 2015; Quandt et al., 2014; Vandenbrouck et al., 2016).

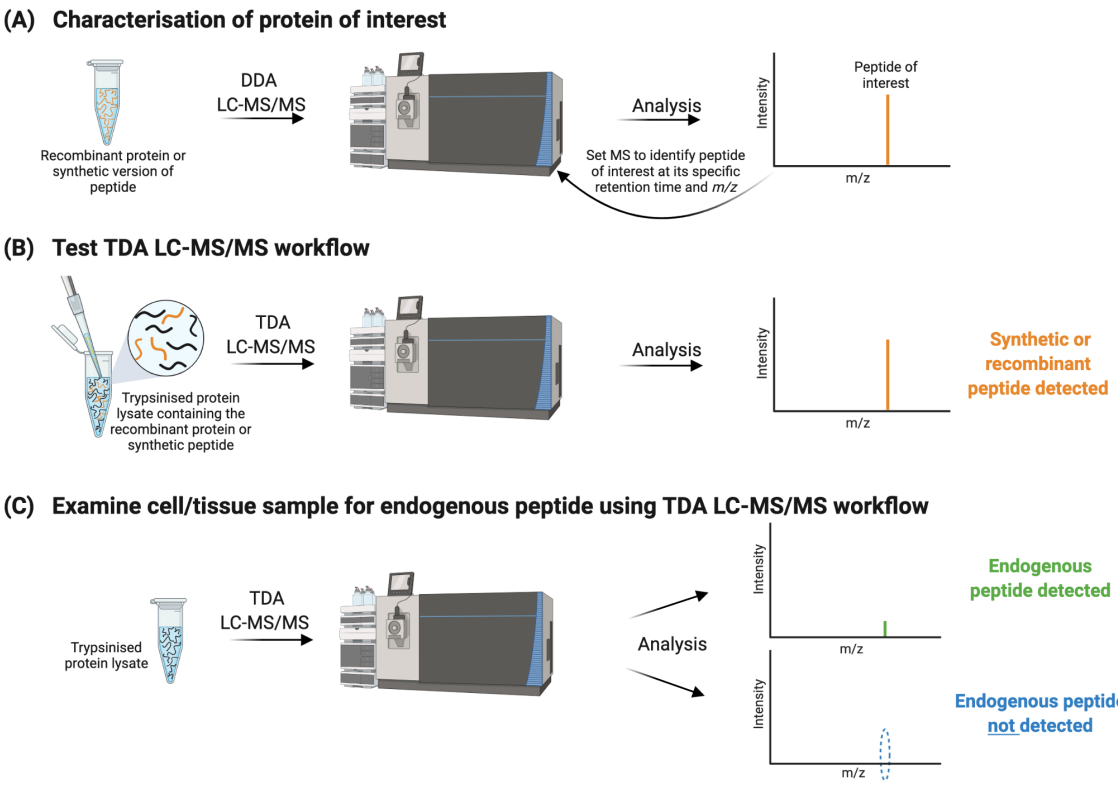


Figure 5.2 Schematic of the TDA LC-MS/MS technique. *(A) The peptide of interest is characterised by running the recombinant or synthetic version of the peptide through the mass spectrometer in discovery mode. The m/z and retention time of the peptide of interest is used to program the mass spectrometer to only identify peptides with these specific characteristics. (B) The workflow is tested by running a trypsinised whole protein lysate sample containing the recombinant protein or the synthetic peptide through it. If the mass spectrometer detects the recombinant or synthetic peptide it has passed the validation process. (C) The final step of this method is to search for the endogenous peptide in an enzymatically digested whole protein lysate sample using the TDA LC-MS/MS program. An MS/MS response will occur if the peptide is identified in that sample. Created with BioRender.com, Adapted from Kolker et al., 2006.*

TDA LC-MS/MS analysis has unparallel sensitivity and is a great option when investigating low abundance proteins (Vidova & Spacil, 2017). Therefore, TDA LC-MS/MS is the second technique that was employed to examine a panel of cell lines and tissues for the endogenous DHFR2 protein.

5.1.5.3. Different Targeted Data Acquisition LC-MS/MS methods

There are multiple methods within TDA LC-MS/MS including parallel reaction monitoring (PRM), selected reaction monitoring (SRM) and multiple reaction monitoring (MRM) (MacLean et al., 2010). All three TDA LC-MS/MS methodologies are highly sensitive, specific and reproducible (Ronsein et al., 2015). The only difference in these methodologies is how the second mass analyser identifies the peptides. The PRM technique identifies a peptide by detecting all the fragmented ions from that peptide. Whereas, the SRM and MRM methods detect a peptide based on one-two preselected fragmented ions belonging to that peptide. Due to this the SRM and MRM methods have slightly more sensitivity and specificity when compared to the PRM technique (Ronsein et al., 2015).

5.1.6. Aims and objectives

The main aim of this chapter was to interrogate cell lines and tissues for the presence of the endogenous human DHFR2 protein. To achieve this aim, the following objectives were proposed:

1. Investigate the presence of the endogenous DHFR2 protein in a range of adult cells and tissues using DDA and PRM LC-MS/MS analysis. This analysis may reveal if the endogenous human DHFR2 protein is cell/tissue specific.
2. Determine if the endogenous DHFR2 protein is only detected under particular cellular contexts or conditions. Perform DDA and PRM LC-MS/MS analysis on:
    a. The HepG2 DHFR knockdown cell line. This analysis will reveal if the native DHFR2 protein is synthesised to carry out the reductase activity when the expression of the DHFR enzyme is reduced.
    b. Cells treated with a proteasome inhibitor known as MG132. This experiment will determine if the endogenous DHFR2 protein is subject to proteasome-mediated degradation.
3. Investigate the presence of the endogenous DHFR2 protein in a range of human embryonic and foetal tissues using MRM LC-MS/MS analysis. This analysis has the potential to uncover if the endogenous human DHFR2 protein is expressed during development.

5.2. Methods

5.2.1. DDA and PRM LC-MS/MS analysis on adult cell lines/tissue, HepG2 DHFR knockdown cells and proteasome inhibitor-treated HepG2 cells

These experiments were carried out in the Proteomics Facility in the National Institute for Cellular Biology (NICB) in DCU, Ireland.

5.2.1.1. Cell culture and protein extraction for the adult cell/tissue specific experiment

Human placental, umbilical cord, membrane, ovary and testis tissue was reduced to a fine powder according to Chapter 2, Section 2.2.19.1. The following human cell lines were cultured as per Chapter 2, Section 2.2.2: HepG2, undifferentiated induced pluripotent stem cells (iPS), a hepatocyte-derived carcinoma cell line (HuH7), iPS-derived hepatocytes (101B), a neuroblastoma cell line (IMR32) and neuroectodermal progenitor stem cells (NEP). Whole protein lysates were extracted from these ground tissue samples and cell lines as described in Chapter 2, Section 2.2.19.2.

5.2.1.2. Cell culture and protein extraction for the DHFR complementation experiment

A HepG2 DHFR knockdown cell line was generated using the CRISPR-Cas9 system according to Chapter 2, Section 2.2.1. The HepG2 DHFR knockdown cell line was cultured according to Chapter 2, Section 2.2.2. Whole protein lysates were extracted from the HepG2 DHFR knockdown cell line as described in Chapter 2, Section 2.2.19.2.

5.2.1.3. Cell culture, proteasome inhibitor treatment and protein extraction for the DHFR2 ubiquitination experiment

5.2.1.3.1.Carbobenzoxy-Leu-Leu-leucinal (MG132) incubation time curve on HepG2 cells

A time curve was carried out to determine the optimal incubation period to treat HepG2 cells with the proteasome inhibitor MG132. Six T25 flasks were grown to approximately 80% confluency according to Chapter 2, Section 2.2.2. The dimethyl sulfoxide (DMSO) control flask was incubated in DMEM containing 10% (v/v) FBS and 0.005% (v/v) DMSO in a 5% CO_2 incubator at 37 °C for 8 hour. The remaining five flasks were treated with DMEM containing 10% (v/v) FBS and 10 µM

MG132 solution. Each of these flasks were incubated in a 5% CO_2 incubator at 37 °C for a different period: 1 hour, 2 hour, 4 hour, 8 hour and 24 hour. The cells were harvested and pelleted according to Chapter 2, Section 2.2.2 after the stated incubation times. The cells were lysed, protein extracted and quantified as per Chapter 2, Section 2.2.19.2. Western Blot analysis was carried out on these samples to determine which incubation time for MG132 treatment accumulated the most ubiquitin proteins, as per Chapter 2, Section 2.2.22. A total of 75 µg of protein was denatured and ran on a 4–20% Tris-glycine gel. The membrane was initially incubated in a monoclonal antibody toward Ubiquitin and subsequently probed with a monoclonal rat anti-mouse HRP conjugated secondary antibody, concentrations stated in Chapter 2, Table 2.8.

5.2.1.3.2.Treatment with proteasome inhibitor MG132 on HepG2 cells for LC-MS/MS analysis

HepG2 cells were grown to approximately 80% confluency in two T75 flasks according to Chapter 2, Section 2.2.2. One flask was treated with DMSO as a control and the other was treated with the MG132 inhibitor as per Section 5.2.1.3.1, with 8 hour being the optimal incubation time. The cells were harvested, lysed, proteins isolated and concentration calculated as per Section 5.2.1.3.1. Western Blot analysis was employed to confirm a build-up of ubiquitinated proteins in the MG132 treated sample compared to the DMSO control, as per Section 5.2.1.3.1.

5.2.1.4. LC-MS/MS sample preparation for the adult cell/tissue specific experiment, DHFR complementation experiment and DHFR2 ubiquitination experiment

Protein lysate from Sections 5.2.1.1, 5.2.1.2 and 5.2.1.3.2 were prepared for LC-MS/MS analysis as per Chapter 2, Section 2.2.20.1 and Section 2.2.20.2.

5.2.1.5. DDA and PRM LC-MS/MS analysis for the adult cell/tissue specific experiment, DHFR complementation experiment and DHFR2 ubiquitination experiment

All the trypsinised whole protein lysate samples from Section 5.2.1.4 were resuspended in 15 µL of resuspension solution and then sonicated using a sonicating waterbath (Precision ultrasonic cleaning, DP201–00, Ultrawave) under the following conditions: power of 280 watts, frequency of 30 kilohertz, ambient temperature and time of 5 min. The peptide concentration was quantified using the Nanodrop One spectrophotometer (Labtech instruments), at an absorbance of 280 nm. The samples were diluted down to a working solution of 1 µg/µL using resuspension solution for

DDA LC-MS/MS analysis (Chapter 2, Section 2.2.21.1) and PRM DHFR2 LC-MS/MS analysis (Chapter 2, Section 2.2.21.2) to search for a unique DHFR2 peptide. The PRM LC-MS/MS technique was the TDA method of choice as it was the system that was set up in the Proteomics facility in DCU.

5.2.2. MRM LC-MS/MS analysis on human embryonic and foetal tissues

This experiment was carried out in the Nervous System Development and Translational Mass Spectrometry research laboratories at University College London's (UCL) Institute of Child Health, London, United Kingdom.

5.2.2.1. Protein extraction from human embryonic and foetal tissues

Kidney, heart, lung, liver and brain tissue were obtained for three different stages of human development by the Nervous System Development research group UCL, UK. The three stages were the Carnegie stage 21-22 (embryo), post-conceptual week 9-10 (foetus) and post-conceptual week 15-17 (foetus). See Chapter 2, Section 2.1.4 for more information about the origin of these tissues. Approximately 100 µg of each tissue was isolated and placed into a tissue homogeniser tube (Precellys). The lysis buffer (200 µL) was added to the tissue homogeniser tube. The samples were lysed using a Minilys personal homogeniser from Bertin Technologies. It was turned on to the highest speed for 10 sec for 3 runs. The samples were centrifuged at 16,000 x *g* at room temperature for 15 min. The supernatant containing the protein was placed into a clean tube and incubated at 85 °C for 20 min to denature the proteins. Once the samples had cooled down, 50 µL of 600 nM iodoacetamide was added and they were incubated in a dark place for 20 min.

5.2.2.2. Trypsinisation, purification and lyophilisation of human embryonic and foetal protein

A total of 5 µL of Trypsin Gold Mass Spectrometry Grade (0.4 µg/µL) was added to each of the protein samples from Section 5.2.2.1. The samples were incubated overnight at 37 °C in a waterbath to trypsinise the proteins. The next morning 5 µL of 6% (v/v) of trifluoroacetic acid was added to each sample and placed on a rotator for 10 min at room temperature to stop the digestion. They were subsequently centrifuged at 16,000 x *g* at room temperature for 10 min. The supernatant containing the trypsinised proteins were placed into a clean tube.

A solid-phase extraction cleanup was performed on the enzymatically digested peptides. The Isolute C18 cartridges were placed on a retort stand and 1 mL of 60% (v/v) acetonitrile, 0.1% (v/v) trifluoroacetic acid was added to solvate them. A plunger was used to push the first drop through the column and the remainder of the solution was left to gravity to pass through. The columns were equilibrated by adding 1 mL of 0.1% (v/v) trifluoroacetic acid to them. The plunger was used to slowly push all the buffer through the column. This equilibrium step was repeated. The trypsinised peptides samples were loaded onto the column and the solution was left to gravity to pass through. The bound peptides were washed by adding 1 mL of 0.1% (v/v) trifluoroacetic acid to the column. The wash buffer was left to gravity to pass through the column. The peptides were eluted from the column by adding 500 µL of 60% (v/v) acetonitrile, 0.1% (v/v) trifluoroacetic acid to the column and left under gravity. The eluent was collected in a fresh 1.5 mL tube. The samples were dried using a Eppendorf's Concentrator plus under the following conditions: mode/vent of vacuum-high vapor, vacuum level of 20 metric millibar, speed of 20 x *g*, drying rate of medium (ambient temperature) and time of 10 hour.

5.2.2.3. Generation of Skyline MRM DHFR2 LC-MS/MS workflow

Skyline is an open-source software tool that has been developed at the University of Washington by Brendan MacLean to accelerate the targeted LC-MS/MS field. Skyline allows scientists to build TDA, DDA and data-independent acquisition (DIA) workflows. It has the ability to generate PRM, SRM and MRM targeted data acquisition LC-MS/MS programs (MacLean et al., 2010). The MRM LC-MS/MS technique was chosen to search for the DHFR2 specific peptides in the human embryonic/foetal tissue as it was the system that was set up in the Translational Mass Spectrometry research laboratories at UCL.

Prior to examining the human embryonic/foetal tissue for the DHFR2 protein the MRM LC-MS/MS workflow had to be built and validated on Skyline. Protein sequences belonging to the human DHFR2 and DHFR protein were obtained from UniProt (https://www.uniprot.org/) and pasted directly into the Skyline software. Skyline has the ability to predict peptides generated from an enzymatic digest; the peptide's characteristics (retention time, charge state and *m/z* ratio) and the ideal fragmented ions to target. This information was obtained for the DHFR2 and DHFR proteins. Peptides that were common to both proteins were eliminated from the workflow. The peptide information belonging to the DHFR protein was subsequently removed. Therefore, the MRM DHFR2 LC-MS/MS workflow only contained trypsinised peptides that were unique to the DHFR2 protein, this consisted of eight peptides (Table 5.1).

Table 5.1 DHFR2 specific peptides included in the Skyline MRM DHFR2 LC-MS/MS workflow

Protein name	Peptide sequence
DHFR2	NGDLPRPPLR
	QNLVIMGR
	EPPQGAHFLAR
	LTERPELANK
	VDMIWIVGGSSVYK
	EAMNHLGHLK
	IMQDFESDTFFSEIDLEK
	LLPEYPGVLSDVQEGK

5.2.2.4. Validation of Skyline MRM DHFR2 LC-MS/MS workflow using recombinant DHFR2 protein

The next logical step was to validate the MRM DHFR2 LC-MS/MS workflow by running an enzymatically digested whole protein lysate containing the trypsinised DHFR2 recombinant protein. The cell lysate from Chapter 4, Section 4.3.1 was prepped in DCU according to Chapter 2, Section 2.2.20.1 and Section 2.2.20.2. The lyophilised peptides were sent to the Translational Mass Spectrometry research laboratory at UCL. The peptides were resuspended in 100 µL of 5% (v/v) acetonitrile, 0.1% (v/v) trifluoroacetic acid and placed on a shaker at room temperature for 10 min at 1,000 rpm. The peptide mixture was centrifuged at 16,000 x *g* at room temperature for 10 min and the supernatant was transferred into a polypropylene screw neck vial for MRM LC-MS/MS analysis. The MRM LC-MS/MS analysis for this experiment was performed on the ACQUITY Ultra Performance LC system (Waters) coupled to the Xevo TQ-S Tandem Quadrupole Mass Spectrometer (Waters). The peptide mixture was loaded onto an ACQUITY UPLC Peptide CSH C18 Column (Waters) where a binary gradient was employed: 95% Solvent A1, 5% Solvent B1 for 1 min; 80% Solvent A1, 20% Solvent B1 for 4.5 min; 70% Solvent A1, 30% Solvent B1 for 9.8 min; 100% Solvent B1 for 10.5 min; 100% Solvent B1 for 11.99 min; 95% Solvent A1, 5% Solvent B1 for 12 min; 95% Solvent A1, 5% Solvent B1 for 15 min and holding at 95% Solvent A1, 5% Solvent B1. The separation used a flow rate of 0.15 mL/min and a column temperature of 50 °C. The peptides were transported to the mass spectrometer where they were ionised using electrospray ionisation under the following parameters: spray voltage 2.8 kV, cone voltage 25 V, source offset voltage 30 V, desolvation temperature 400 °C, desolvation gas flow 600 L/hour, cone gas flow 150 L/hour and nebuliser gas flow 7 bar. The first mass analyser (Quadrupole) allowed the ionised peptides that were preselected for the 8 DHFR2 unique peptides through for fragmentation. The ionised peptides

were fragmented using a collision gas flow of 0.25 mL/min. The second mass analyser (Quadrupole) measured the resulting ions. The Skyline software was employed to determine if any ions identified by the second mass analyser matched back to the predetermined ions assigned to the 8 unique DHFR2 peptides.

5.2.2.5. MRM LC-MS/MS analysis on the human embryonic and foetal tissues

The lyophilised samples from Section 5.2.2.2 were resuspended in 50 µL of 5% (v/v) acetonitrile, 0.1% (v/v) trifluoroacetic acid. The samples were placed on a shaker at room temperature for 10 min at 1,000 rpm. They were subsequently centrifuged at 16,000 x *g* at room temperature for 10 min and the supernatant was transferred into polypropylene screw neck vials.

The Skyline MRM DHFR2 LC-MS/MS workflow generated in Section 5.2.2.3 and validated in Section 5.2.2.4 was adapted for the analysis of the human embryonic and foetal tissues. The DHFR2 specific peptides that were not identified in the whole protein lysate containing the recombinant DHFR2 protein (Section 5.2.2.4) were removed from the workflow. Internal cytoplasmic, nucleic and mitochondrial control peptides were added to the workflow to confirm sample preparation was successful. See Table 5.2 for the list of peptides included in the workflow. The human embryonic and foetal peptide mixtures in the polypropylene screw neck vials were assessed using the adapted Skyline MRM DHFR2 LC-MS/MS workflow. The HPLC and mass spectrometer instrument and settings were the same as in Section 5.2.2.4.

Table 5.2 Peptides included in the Skyline MRM DHFR2 LC-MS/MS workflow

Protein name	Peptide sequence	Precursor ion charge	Precursor ion *m/z*	Fragment ion
DHFR2	NGDLPRPPLR	3	378.8841	y9 - 510.8011++ y6 - 368.2348++
	QNLVIMGR	2	465.7631	y5 – 575.334+ y4 – 476.2650+
	VDMIWIVGGSSVYK	2	777.4053	y10 – 1095.583+ y8 – 796.4199+
	LLPEYPGVLSDVQEGK	2	872.4618	y14 – 759.3777++ y12 – 646.3301++
ACTB (Cytoplasm)	AGFAGDDAPR	2	488.7278	y7 – 701.3213+ y6 – 630.2842+
	DLTDYLMK	2	499.7468	y6 – 770.3753+ y5 – 669.3276+
TUBA1A (Cytoplasm)	DVNAAIATIK	2	508.2928	y8 – 801.4829+ y6 – 616.4028+
LMNB1 (Nucleus)	LVEVDSGR	2	437.7351	y6 – 662.3104+ y5 – 533.2678+
	AGGPTTPLSPTR	2	577.8118	y6 – 670.3883+ y9 – 485.2718++
HSPD1 (Mitochondria)	GIIDPTK	2	372.2185	y5 – 573.3243+ y3 – 345.2132+
COX4I1 (Mitochondria)	VNPIQGLASK	2	513.8007	y6 – 603.3461+ y8 – 407.2451++
TOM20 (Mitochondria)	LPTISQR	2	407.7427	y5 – 604.3413+ y6 – 351.2007++

5.3. Results

5.3.1. The DHFR2 protein was not detected in any of the adult cell lines or tissues examined

As mentioned in Section 5.1.1, it was hypothesised that the DHFR2 protein could be cell/tissue specific. In addition, a genetic association study of an Irish cohort found an association between 3 SNPs within the *DHFR2* gene and the risk of developing NTDs (Pangilinan et al., 2021). Even though this result did not replicate in a British cohort, these findings still suggest that the DHFR2 protein may only be expressed and functional during human development. The regulations in Ireland do not allow research on human embryonic/foetal tissue. Therefore, adult human cell lines and tissues that are similar to embryonic/foetal tissue were employed to search for the native DHFR2 protein.

This included the HepG2 cells, undifferentiated iPS cells, HuH7 cells, 101B cells, IMR32 cells, NEP cells, placental tissue, umbilical cord tissue, membrane tissue, ovaries tissue and testis tissue. The data-dependent acquisition LC-MS/MS method was initially employed to examine the entire proteome of these cell lines/tissues and to look for peptides that were unique to the DHFR2 protein, as per Section 5.2.1.1, Section 5.2.1.4 and Section 5.2.1.5. For each cell and tissue type examined, the instrument detected 20,000 - 35,000 peptides from 2,300 - 5,000 different proteins. This is the expected amount of proteins identified from a whole lysate sample using the Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific), meaning the sample preparation was successful. The DDA LC-MS/MS analysis detected peptides shared by the DHFR and DHFR2 proteins in some of the cell lines and tissues, including the HepG2, undifferentiated iPS, HuH7 and NEP samples (Table 5.3). However, the instrument did not recognise any peptides that were unique to the DHFR2 protein in any of the cell lines or tissues. Moreover, the mass spectrometer did identify a number of peptides with a sequence exclusive to the DHFR protein in the HepG2, undifferentiated iPS, HuH7, IMR32, NEP and testis samples (Table 5.3). This result confirmed the presence of the DHFR enzyme in these cell lines and tissues. Furthermore, this result suggested that the peptides identified in these samples that were shared by both proteins belonged to the DHFR protein. Confirming the non-detection of the endogenous DHFR2 protein in these samples. There were no DHFR and DHFR2 peptides identified in the 101B cells, placental, umbilical cord, membrane and ovary tissue samples. This result means that the endogenous DHFR and DHFR2 protein was not detected in these samples; even though other OCM proteins were detected including MTHFD1, MTHFD1L, MTHFD2, SHMT1 and SHMT2. Despite the use of one of the most advanced mass spectrometers available, this result suggested that the DDA LC-MS/MS method was not sensitive enough for the analysis that was required. This is a well-known limitation associated with DDA LC-MS/MS analysis on a complex sample (Wang & Wilson, 2013). Therefore, a TDA LC-MS/MS method would be the optimal approach to overcome this challenge.

Table 5.3 Dihydrofolate reductase peptides detected in the DDA LC-MS/MS analysis on the adult cell lines and tissues

Peptide Sequence	Modifications	Protein Name	HepG2	iPS	HuH7	IMR32	NEP	Testis
MTTTSSVEGK	-	DHFR / DHFR2	✓	✓	✓	x	✓	X
INLVLSR	-	DHFR / DHFR2	✓	x	x	x	x	X
NGDLPWPPLR	-	DHFR	x	x	x	✓	x	x
LTEQPELANK	-	DHFR	x	✓	✓	x	x	✓
VDMVWIVGGSSVYK	-	DHFR	x	✓	x	✓	x	x
EAmNHPGHLK	Oxidation (M)	DHFR	✓	x	x	x	x	x
EAMNHPGHLK	-	DHFR	x	✓	x	x	x	x
IMQDFESDTFFPEIDLEK	-	DHFR	✓	x	x	x	x	x
LLPEYPGVLSDVQEEK	-	DHFR	✓	✓	x	✓	x	X
YKFEVYEK	-	DHFR	x	✓	x	x	✓	x

There was a possibility that the DHFR2 protein was present at such low abundance it was rendered undetectable by the mass spectrometer when carrying out the DDA LC-MS/MS analysis on the adult cell lines and tissues. Therefore, a TDA LC-MS/MS method known as PRM was employed to interrogate these samples further for the DHFR2 protein, as per Section 5.2.1.5. This approach eliminates the background noise from the other peptides in a sample and only searches for a preselected peptide(s) belonging to the protein of interest. The DHFR2 protein has eight tryptic peptides that are unique to this protein that could have been targeted using this approach, shown in Figure 5.3. The EAMNHLGHLK tryptic peptide was chosen to be the target for the PRM LC-MS/MS analysis by process of elimination. The four DHFR2 tryptic peptides (QNLVIMGR, EPPQGAHFLAR, LTERPELANK and VDMIWIVGGSSVYK) that have one amino acid difference which is similar to the amino acid on the DHFR peptide (illustrated as : in Figure 5.3) were eliminated to prevent false-positive results (Shen et al., 2009). The two DHFR2 tryptic peptides (IMQDFESDTFFSEIDLEK and LLPEYPGVLSDVQEGK) that correspond with the DHFR peptides that are phosphorylated were also ruled out in case they were post-translationally modified (Anderson et al., 2007; Skierka et al., 2019). That left the NGDLPRPPLR and EAMNHLGHLK unique DHFR2 tryptic peptides. Either peptide could have been targeted in the PRM LC-MS/MS method as they had a significant *m/z* difference when compared to their DHFR counterpart, making it relatively straightforward for the mass spectrometer to differentiate between them. The EAMNHLGHLK peptide was chosen as the recombinant version of this tryptic peptide was identified by the Orbitrap Fusion Tribrid mass

spectrometer (Thermo Scientific) using both the DDA and PRM LC-MS/MS methods, Chapter 4, Section 4.3.1.

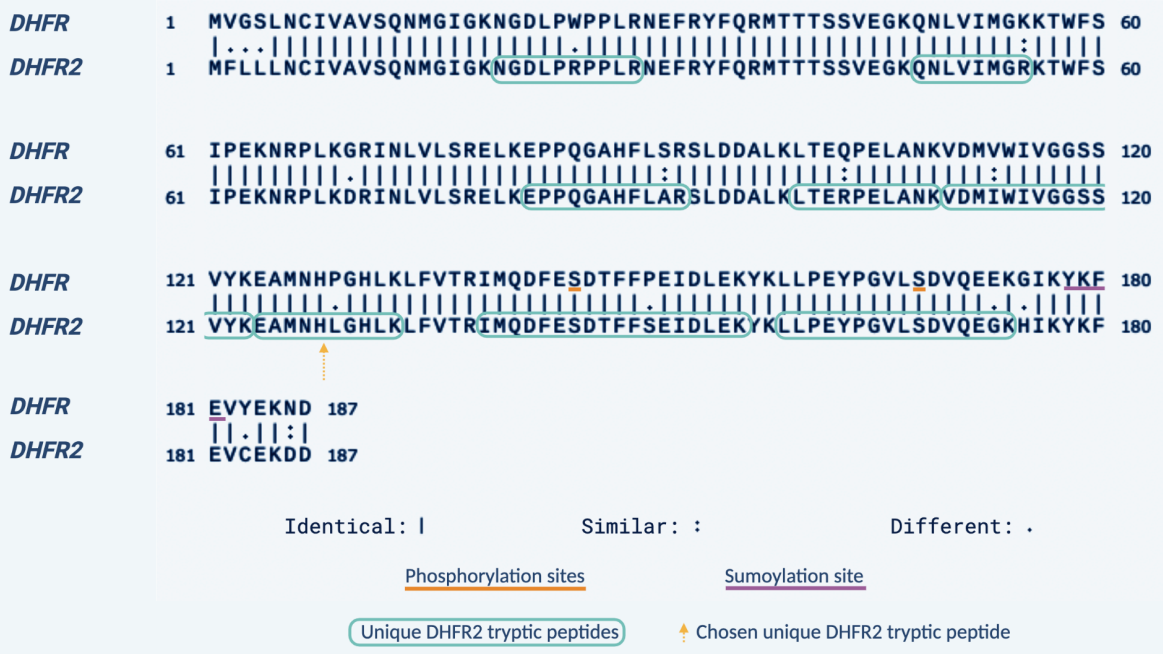


Figure 5.3 Sequence alignment comparing the amino acid sequence of the human DHFR protein and the predicted DHFR2 protein showing the possible tryptic peptides unique to the DHFR2 protein (green box). *The figure shows eight possible tryptic peptides unique to the DHFR2 protein that could have been employed for the PRM LC-MS/MS method. However, six were eliminated due to the amino acid difference being too similar (:) and possible phosphorylation sites (underlined in orange). The chosen DHFR2 tryptic peptide is shown by the yellow arrow. Created with BioRender.com, Adapted from McEntee et al., 2011.*

A synthetic version of the selected DHFR2 peptide was run through the mass spectrometer in discovery mode to determine the *m/z*, charge state and retention time of the non-modified and any post-translationally modified versions of the peptide. The *m/z* and retention time for both charged states (+2 and +3) were identified for the non-modified and a post-translationally modified version of the synthetic peptide i.e., oxidation of the methionine amino acid. This data was used to generate a workflow on the mass spectrometer to only identify peptides at these specific characteristics. The workflow was tested by running an enzymatically digested whole protein lysate spiked with the synthetic peptide through the mass spectrometer. Two MS/MS responses were recorded in the mass spectrometer. This means both versions of the synthetic peptide were detected, see an example of the results in Figure 5.4 (A). Both of these optimisation steps were performed prior to each analysis to account for the instrument's variability. Finally, all the

trypsinised whole protein lysate samples from the adult cell lines and tissues were examined using the optimised and validated PRM LC-MS/MS method. There was no MS/MS response triggered when analysing the HepG2 cells, undifferentiated iPS cells, HuH7 cells, 101B cells, IMR32 cells, NEP cells, placental tissue, umbilical cord tissue, membrane tissue, ovaries tissue and testis tissue. See an example of the results in Figure 5.4 (B). Confirming the non-detection of the endogenous version of the DHFR2 peptide in any sample examined.

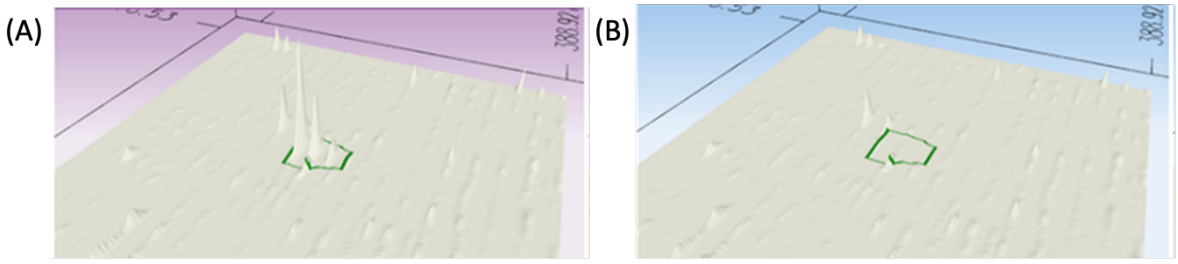


Figure 5.4 Example of the three-dimensional distribution plots generated from the PRM LC-MS/MS analysis. *(A) The isotopic pattern shown for the undifferentiated human iPS sample spiked with the synthetic version of the DHFR2 specific peptide. An MS/MS response occurred, identifying a normal and posttranslationally modified version of the synthetic peptide. (B) The isotopic pattern for the undifferentiated human iPS sample. No MS/MS response occurred, meaning no version of the endogenous DHFR2 peptide was detected in this sample. Created on Progenesis QI.*

5.3.2. The DHFR2 protein is not synthesised to carry out the reductase activity when the expression of the DHFR enzyme is reduced

The role of the DHFR2 protein remained unanswered as the LC-MS/MS analysis on the panel of adult cell lines and tissues failed to detect a DHFR2 specific peptide. Therefore, the next hypothesis was investigated i.e., the DHFR2 protein is only expressed and functional under particular cellular contexts or conditions. More specifically, to determine if the DHFR2 protein is synthesised to provide reductase activity when the DHFR protein expression is eliminated or significantly reduced. A HepG2 DHFR knockdown cell line was generated using the CRISPR-Cas9 system (Chapter 2, Section 2.2.1). The resulting proteome of this knockdown model was assessed for the DHFR2 protein using the LC-MS/MS approaches described in Chapter 2, Section 2.2.21 and employed in Section 5.3.1.

The DDA LC-MS/MS analysis on the HepG2 DHFR knockdown cell line identified approximately 22,000 peptides from 3,700 proteins. This is the expected amount of proteins identified from a

whole cell lysate sample using an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific), meaning the sample preparation was successful. As the *DHFR* gene was knocked down and the expression of the DHFR2 protein was yet to be detected, other proteins belonging to the OCM pathway were used to further assess the quality of the raw data. Peptides belonging to the MTHFD1, MTHFD1L, MTHFD2, SHMT1 and SHMT2 proteins were identified by the mass spectrometer. Once the raw data passed the quality control steps the data set was then examined for the dihydrofolate reductase proteins. The mass spectrometer did not identify any peptides belonging to the DHFR protein. This result meant the CRISPR-Cas9 knockdown successfully eliminated or drastically reduced the expression of the DHFR protein. The DDA LC-MS/MS analysis on the HepG2 DHFR knockdown also did not detect any peptides that align back to the DHFR2 protein. This result meant that the endogenous DHFR2 protein was not detected in the HepG2 DHFR knockdown cell line using the DDA LC-MS/MS method. Therefore, the PRM DHFR2 LC-MS/MS method described in Chapter 2, Section 2.2.21.2 and Section 5.3.1 was employed to interrogate the sample further. The PRM LC-MS/MS analysis did not detect the native DHFR2 peptide in protein lysate from the HepG2 DHFR knockdown cell line. Confirming the non-detection of the DHFR2 protein in the HepG2 DHFR knockdown model.

5.3.3. The DHFR2 protein is not subject to ubiquitination

Given that the DHFR2 protein was not identified in any of the cell and tissue samples or the HepG2 DHFR knockdown cell line there was a possibility it was subject to protein degradation. One of the main methods for intracellular protein degradation is via the ubiquitin-proteasome pathway (Ciechanover, 1994). Therefore, HepG2 cells were treated with DMSO (control sample) or a proteasome inhibitor known as MG132; this peptide aldehyde halts ubiquitination and results in an accumulation of ubiquitinated proteins within a cell. A build-up of ubiquitinated proteins was confirmed in the cells treated with MG132 by Western Blot using an antibody against ubiquitin (Figure 5.5). The resulting proteome from the DMSO and the MG132 treated cells were assessed for the DHFR2 protein using the LC-MS/MS approaches described in Chapter 2, Section 2.2.21 and employed in Section 5.3.1. The DDA LC-MS/MS analysis on the HepG2 cells treated with DMSO and MG132 identified approximately 30,000 - 32,000 peptides from 4,800 - 4,900 proteins. As already described, this is the expected amount of proteins identified from a whole cell lysate sample using an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific), meaning the sample preparation was successful. Furthermore, peptides belonging to proteins that have been found to be ubiquitinated (e.g. apolipoprotein B (Yeung et al., 1996)) were identified at a higher abundance in

the HepG2 cells treated with MG132 compared to the control cells treated with DMSO. This result also confirmed the MG132 inhibitor successfully blocked the ubiquitination of proteins in the HepG2 cells. Once the raw data passed the quality control steps the data set was then examined for the dihydrofolate reductase proteins. DDA LC-MS/MS analysis detected one peptide common to the DHFR and DHFR2 protein in the DMSO and the MG132 treated cells, see Table 5.4. However, the instrument did not recognise any peptides that were unique to the DHFR2 protein in either cell lysate. The mass spectrometer did identify a number of peptides unique to the DHFR protein in the DMSO and the MG132 treated cells, see Table 5.4. This result confirmed the presence of the DHFR enzyme in both samples. Furthermore, this result suggested that the peptides identified in these samples that were shared by both proteins belonged to the DHFR protein. Confirming the non-detection of the endogenous DHFR2 protein in the DMSO treated control sample and the MG132 treated sample. The PRM LC-MS/MS method described in Chapter 2, Section 2.2.21.2 and Section 5.3.1 was subsequently employed to interrogate the samples further. The PRM LC-MS/MS analysis did not identify the native DHFR2 peptide in either of the samples. This result meant that the endogenous DHFR2 protein was not detected in the DMSO and the MG132 treated cells using the PRM LC-MS/MS method.

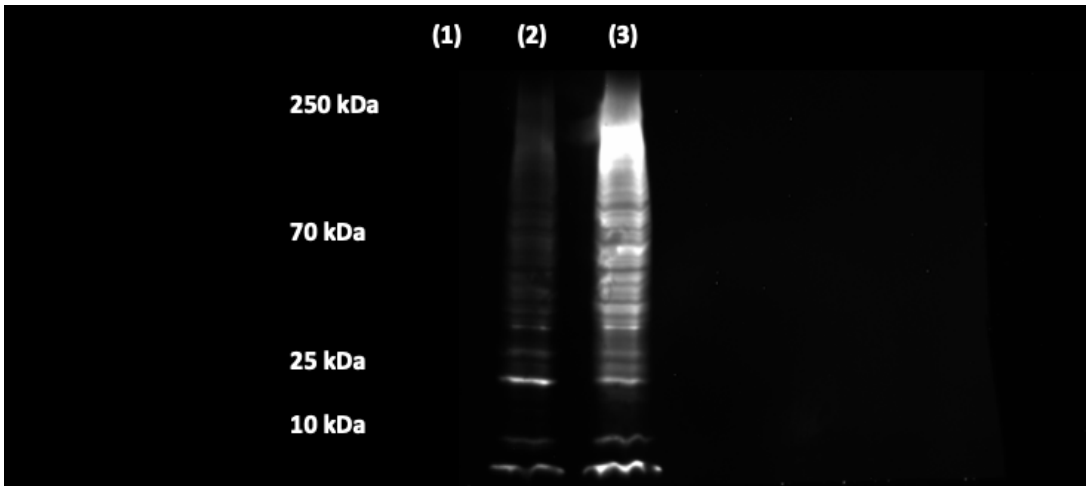


Figure 5.5 Western Blot analysis of DMSO treated control sample and MG132 treated sample using an anti-ubiquitin antibody. *HepG2 cells were treated with DMSO or MG132 proteasome inhibitor. The resulting proteins (75 µg) were separated on a 4-20% Tris-glycine gel and visualised using chemiluminescence Western Blot, probed with an anti-ubiquitin primary antibody and anti-mouse secondary antibody. (1) Protein ladder, (2) DMSO treated HepG2 protein lysate and (3) MG132 treated HepG2 protein lysate. The sample treated with DMSO should have a low amount of ubiquitinated proteins. The faint bands in lane 2 show low levels of ubiquitinated proteins. The sample treated with MG132 should have a build-up of ubiquitinated proteins. The intense smear in Lane 3 reveals high levels of ubiquitinated proteins.*

Table 5.4 Dihydrofolate reductase peptides detected in the DDA LC-MS/MS analysis on DMSO and MG132 treated cells

Peptide Sequence	Modifications	Protein Name	DMSO Sample - Control	MG132 Sample
MTTTSSVEGK	-	DHFR/DHFR2	✓	✓
QNLVIMGK	-	DHFR	✓	✓
EPPQGAHFLSR	-	DHFR	✓	✓
LTEQPELANK	-	DHFR	✓	✓
VDMVWIVGGSSVYK	-	DHFR	✓	✓
IMQDFESDTFFPEIDLEK	-	DHFR	✓	x
LLPEYPGVLSDVQEEK	-	DHFR	✓	✓

5.3.4. The DHFR2 protein was detected in human embryonic and foetal heart tissue

Given that the endogenous DHFR2 protein had not been detected in any of the cell lines and tissues or in any of the cellular conditions created, the role of this gene remained unanswered. The focus returned to the original hypothesis that the DHFR2 protein may play a key role in the OCM pathway during early development. Therefore, to determine if this hypothesis was in fact true, human embryonic/foetal tissue needed to be examined for the native DHFR2 protein. The regulations in Ireland do not allow research on human embryonic/foetal tissue. Therefore, this experiment was performed in the United Kingdom where human embryonic/foetal research is permitted.

Prior to examining the human embryonic/foetal tissue for the endogenous DHFR2 protein, an MRM LC-MS/MS workflow had to be generated on Skyline to target the peptides unique to the DHFR2 protein. This workflow was initially set to search for the eight DHFR2 specific peptides (Table 5.1), as per Section 5.2.2.3. The MRM DHFR2 LC-MS/MS workflow was tested by running a trypsinised whole protein lysate containing the recombinant DHFR2 protein, according to Section 5.2.2.4. The MRM LC-MS/MS analysis on this sample identified four out of the eight recombinant DHFR2 specific peptides, see Table 5.5. Thus, the MRM DHFR2 LC-MS/MS workflow was successful at detecting the recombinant DHFR2 protein and therefore deemed capable of identifying the endogenous DHFR2 protein if present in a whole protein lysate sample.

Table 5.5 DHFR2 specific peptides detected in whole protein lysate containing the recombinant DHFR2 protein

Protein name	Peptide sequence	Whole protein lysate with recombinant DHFR2 protein
DHFR2	NGDLPRPPLR	✓
	QNLVIMGR	✓
	EPPQGAHFLAR	x
	LTERPELANK	x
	VDMIWIVGGSSVYK	✓
	EAMNHLGHLK	x
	IMQDFESDTFFSEIDLEK	x
	LLPEYPGVLSDVQEGK	✓

Once the MRM DHFR2 LC-MS/MS workflow had been validated the next step was to examine the human embryonic/foetal tissue for the native DHFR2 protein. Kidney, heart, lung, liver and brain tissue were obtained from three different stages of human development. The three stages were the Carnegie stage 21-22 (embryo), post-conceptual week 9-10 (foetus) and post-conceptual week 15-17 (foetus). See Chapter 2, Section 2.1.4 for more information about the origin of these tissues. The tissue samples were prepared for LC-MS/MS according to Section 5.2.2.1 - 5.2.2.2. The MRM DHFR2 LC-MS/MS workflow was adapted prior to analysing the human embryonic and foetal tissues; the DHFR2 specific peptides that were not identified in the recombinant DHFR2 clone lysate (Table 5.5) were removed from the workflow and internal cytoplasmic (ACTB and TUBA1A), nucleic (LMNB1) and mitochondrial (HSPD1, COX4I1 AND TOM20) control peptides were added to the workflow to confirm sample preparation was successful. See Table 5.2 for the list of peptides included in the workflow. The whole protein lysate containing the recombinant DHFR2 protein and the trypsinised human embryonic/foetal tissues lysates were subsequently analysed using the adapted MRM DHFR2 LC-MS/MS workflow. The whole protein lysate containing the recombinant DHFR2 protein was employed as a control for this run. The MRM DHFR2 LC-MS/MS analysis identified the four DHFR2 specific peptides in this sample (Table 5.6) meaning the workflow and instrumentation were functioning as expected.

Table 5.6 Peptides identified in the whole protein lysate containing the recombinant DHFR2 protein

Protein name	Peptide sequence	Whole protein lysate with recombinant DHFR2 protein
DHFR2	NGDLPRPPLR	✓
	QNLVIMGR	✓
	VDMIWIVGGSSVYK	✓
	LLPEYPGVLSDVQEGK	✓
ACTB (Cytoplasm)	AGFAGDDAPR	✓
	DLTDYLMK	✓
TUBA1A (Cytoplasm)	DVNAAIATIK	✓
LMNB1 (Nucleus)	LVEVDSGR	✓
	AGGPTTPLSPTR	✓
HSPD1 (Mitochondria)	GIIDPTK	✓
COX4I1 (Mitochondria)	VNPIQGLASK	✓
TOM20 (Mitochondria)	LPTISQR	✓

The human embryonic/foetal tissues were subsequently assessed for the internal cytoplasmic (ACTB and TUBA1A), nucleic (LMNB1) and mitochondrial (HSPD1, COX4I1 AND TOM20) control peptides (Section 5.2.2.5). All of the tissues had at least one of the cytoplasmic, nucleic and mitochondrial control peptides present (Table 5.7, Table 5.8 and Table 5.9) meaning the sample preparation was successful at solubilising the membranes and releasing all the proteins.

Given that the instrumentation and the trypsinised samples had passed the quality control steps the human embryonic/foetal tissues were finally examined for the four peptides that were unique to the DHFR2 protein. The NGDLPRPPLR, QNLVIMGR and VDMIWIVGGSSVYK peptides were not detected in any of the human embryonic/foetal tissues examined (Table 5.7, Table 5.8 and Table 5.9). The LLPEYPGVLSDVQEGK DHFR2 peptide was also not identified in the kidney, lung, liver and brain tissue from the three developmental stages examined. There was however evidence that the LLPEYPGVLSDVQEGK DHFR2 specific peptide was present in the heart tissue from the Carnegie stage 21-22 (Table 5.7), post-conceptual week 9-10 (Table 5.8) and post-conceptual week 15-17 samples (Table 5.9). Figure 5.6 shows the chromatograms generated for the LLPEYPGVLSDVQEGK peptide in the control DHFR2 recombinant clone lysate (A), heart tissue lysate from Carnegie stage 21-22 (B), heart tissue lysate from post-conceptual week 9-10 (C) and heart tissue lysate post-conceptual week 15-17 (D). The chromatograms for each of these samples clearly shows the heavy-ion (y14 – 759.3777++) and the light-ion (y12 – 646.3301++) belonging to the LLPEYPGVLSDVQEGK peptide were identified by the Quadrupole. Furthermore, the heavy and light ions were detected at the same retention time; this result implies that the two ions detected were part of the same

peptide. In addition, the retention time for these ions was within a window of error (0.1) of the predicted retention time for the LLPEYPGVLSDVQEGK peptide (shown as explicit in the graphs in Figure 5.6). Therefore, this result suggested that the LLPEYPGVLSDVQEGK peptide was detected in the heart tissue lysate from Carnegie stage 21-22, post-conceptual week 9-10 and post-conceptual week 15-17. The detection of this peptide in this tissue suggested that the endogenous DHFR2 protein is expressed in heart tissue during embryo and foetal development.

Table 5.7 Peptides identified in the Carnegie stage 21-22 (embryo) tissues

Protein name	Peptide sequence	Kidney (12613)	Heart (14755)	Lung (11364)	Liver (12092)	Brain (12210)
DHFR2	NGDLPRPPLR	x	x	x	x	x
	QNLVIMGR	x	x	x	x	x
	VDMIWIVGGSSVYK	x	x	x	x	x
	LLPEYPGVLSDVQEGK	x	✓	x	x	x
ACTB (Cytoplasm)	AGFAGDDAPR	✓	✓	✓	✓	✓
	DLTDYLMK	✓	✓	✓	x	✓
TUBA1A (Cytoplasm)	DVNAAIATIK	✓	✓	✓	✓	✓
LMNB1 (Nucleus)	LVEVDSGR	✓	✓	✓	✓	✓
	AGGPTTPLSPTR	✓	✓	✓	✓	✓
HSPD1 (Mitochondria)	GIIDPTK	✓	✓	✓	✓	✓
COX4I1 (Mitochondria)	VNPIQGLASK	✓	✓	✓	✓	✓
TOM20 (Mitochondria)	LPTISQR	✓	✓	✓	✓	✓

Table 5.8 Peptides identified in the post-conceptual week 9-10 (foetus) tissue

Protein name	Peptide sequence	Kidney (10951)	Heart (14326)	Lung (12412)	Liver (12542)	Brain (12542)
DHFR2	NGDLPRPPLR	x	x	x	x	x
	QNLVIMGR	x	x	x	x	x
	VDMIWIVGGSSVYK	x	x	x	x	x
	LLPEYPGVLSDVQEGK	x	✓	x	x	x
ACTB (Cytoplasm)	AGFAGDDAPR	✓	✓	✓	✓	✓
	DLTDYLMK	✓	✓	✓	x	✓
TUBA1A (Cytoplasm)	DVNAAIATIK	✓	✓	✓	✓	✓
LMNB1 (Nucleus)	LVEVDSGR	✓	✓	✓	✓	✓
	AGGPTTPLSPTR	✓	✓	✓	✓	✓
HSPD1 (Mitochondria)	GIIDPTK	✓	✓	✓	✓	✓
COX4I1 (Mitochondria)	VNPIQGLASK	✓	✓	✓	✓	✓
TOM20 (Mitochondria)	LPTISQR	x	✓	x	✓	x

Table 5.9 Peptides identified in the post-conceptual week 15-17 (foetus) tissue

Protein name	Peptide sequence	Kidney (14104)	Heart (14559)	Lung (13209)	Liver (13454)	Brain (13008)
DHFR2	NGDLPRPPLR	x	x	x	x	x
	QNLVIMGR	x	x	x	x	x
	VDMIWIVGGSSVYK	x	x	x	x	x
	LLPEYPGVLSDVQEGK	x	✓	x	x	x
ACTB (Cytoplasm)	AGFAGDDAPR	✓	✓	✓	✓	✓
	DLTDYLMK	✓	✓	✓	x	✓
TUBA1A (Cytoplasm)	DVNAAIATIK	✓	✓	✓	✓	✓
LMNB1 (Nucleus)	LVEVDSGR	✓	✓	✓	✓	✓
	AGGPTTPLSPTR	✓	✓	✓	✓	✓
HSPD1 (Mitochondria)	GIIDPTK	✓	✓	✓	✓	✓
COX4I1 (Mitochondria)	VNPIQGLASK	✓	✓	✓	✓	✓
TOM20 (Mitochondria)	LPTISQR	✓	✓	x	✓	✓

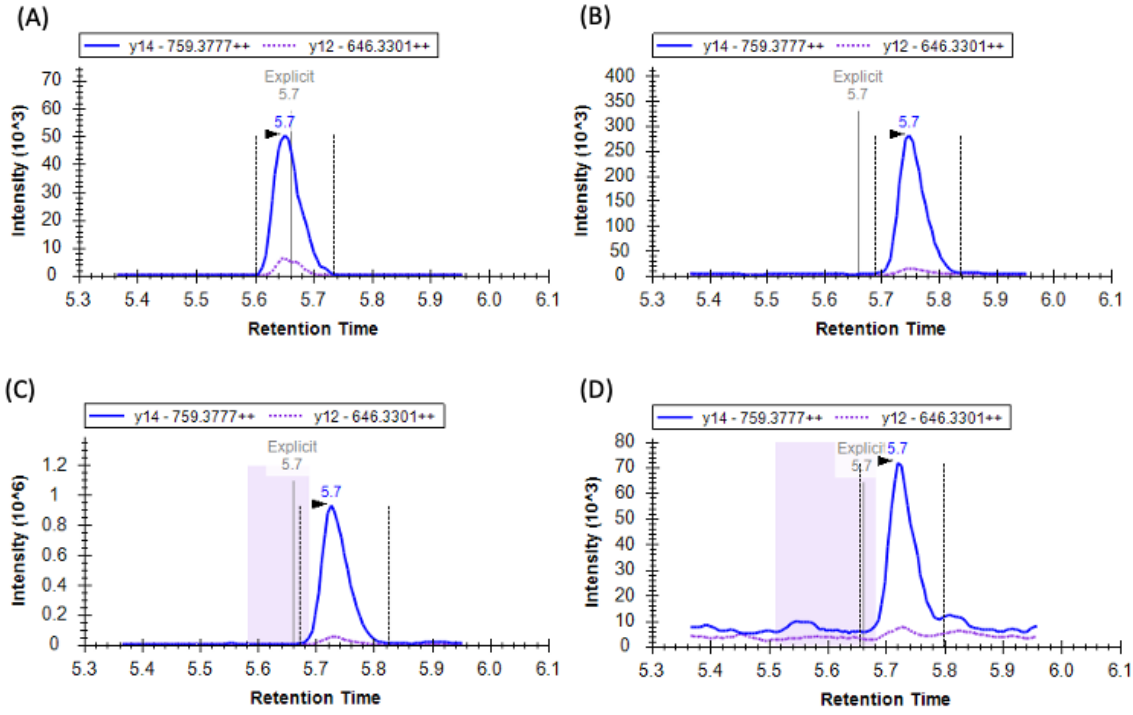


Figure 5.6 Chromatogram results of the DHFR2 peptide LLPEYPGVLSDVQEGK in (A) whole protein lysate containing the recombinant DHFR2 protein, (B) Carnegie stage 21-22 heart tissue (14755), (C) post-conceptual week 9-10 heart tissue (14326) and (D) post-conceptual week 15-17 heart tissue (14559). *The blue line represents the heavy ion (y14 – 759.3777++) and the dotted purple line represents the light ion (y12 – 646.3301++). Given that the heavy and light ions were detected at the same retention time and that their retention time was within a window of error (0.1) of the predicted 5.7 retention time. This suggests that the LLPEYPGVLSDVQEGK peptide was detected in the heart tissue lysate from Carnegie stage 21-22, post-conceptual week 9-10 and post-conceptual week 15-17. Created on Skyline.*

5.4. Discussion

Prior research showed that, despite a relative abundance of mRNA isoforms arising from the *DHFR2* gene (McEntee et al., 2011), there was no evidence of an endogenous DHFR2 protein (Hughes, unpublished). This suggested that the DHFR2 protein expression could be cell/tissue specific, only present under particular cellular conditions and/or a low abundance protein. Therefore, the overall aim of this chapter was to interrogate a range of cell lines and tissues for the presence of the native human DHFR2 protein. The similarity between the DHFR and DHFR2 amino acid sequences made choosing an approach very challenging. Antibody-based methods such as Western Blotting, ELISA and protein microarray could not be employed to carry out this analysis due to commercially available anti-DHFR2 antibodies being prone to cross-react with the DHFR protein. The APMcD lab assessed this by testing four anti-DHFR2 antibodies from different commercial companies and found all bound to the DHFR protein (Bookey, unpublished). Therefore, LC-MS/MS analysis was the best option as it separates and identifies enzymatically digested peptides according to their specific *m/z.* Several of the DHFR and DHFR2 tryptic peptides have an amino acid difference resulting in a contrasting *m/z* (Figure 5.3); this makes it relatively straightforward for the mass spectrometer to distinguish between the DHFR and DHFR2 proteins.

The first set of LC-MS/MS experiments were carried out to investigate whether the DHFR2 protein was expressed in a specific adult cell line or tissue. The initial hypothesis was that the translation of the DHFR2 protein was blocked until it was required to support dihydrofolate reductase activity in a cell. This mechanism may occur during times of high proliferation, such as during embryogenesis or cancer growth. A genetic association study of an Irish cohort found a significant association between three SNPs within the *DHFR2* gene that are in linkage disequilibrium with each other, with an increased risk of NTDs. However, the association was not replicated in a UK cohort (Pangilinan et al., 2021). These findings supported the hypothesis that the *DHFR2* gene may play an important role during embryogenesis. Given that the regulations in Ireland do not allow research to be carried out on human embryonic/foetal tissue, adult cell lines and tissues similar to embryonic/foetal tissue were examined for the native DHFR2 protein. These cell lines included a HepG2, undifferentiated iPS, HuH7, iPSC-derived hepatocytes, IMR32 and NEP. In addition to these cell lines, human placental, umbilical cord, membrane, ovaries and testis tissue were also examined. The DDA LC-MS/MS analysis did not show any proof that the endogenous DHFR2 protein was expressed in any of the samples at the specific time of protein extraction (Section 5.3.1 and Table 5.3). Various proteins associated with the OCM pathway were identified in each cell line and tissue protein lysate, with the DHFR enzyme detected in the majority of samples (Table 5.3). Similar results were

observed when the PRM LC-MS/MS strategy was employed to determine if DHFR2 was a rare protein and rendered undetectable in the DDA LC-MS/MS analysis due to the background noise. The native DHFR2 protein was absent from all cell and tissue lysates investigated using the EAMNHLGHLK peptide as the identifier (Section 5.3.1). Both methods also considered but did not detect any post-translational versions of the protein (oxidation on methionine or carbamidomethylation on cysteine). These posttranslational modifications were considered as they are easily determined and adjusted for by the mass spectrometer (Parker et al., 2010). The DHFR protein has been found to be prone to other posttranslational modifications including phosphorylation and sumoylation (Anderson et al., 2007; Skierka et al., 2019); this could also be the case for the DHFR2 protein. These types of modifications are more challenging to account for by the mass spectrometer and require specific analysis (Parker et al., 2010). Therefore, the DHFR2 peptides that correspond with the DHFR peptides that contain these modifications were not considered as the target for the PRM LC-MS/MS strategy in DCU. To conclude, this analysis found no evidence that the *DHFR2* gene encodes a functional protein in the adult cell lines and tissues examined. This suggested that the predicted DHFR2 protein does not supply dihydrofolate reductase activity to these adult cell lines and tissues.

The next set of LC-MS/MS experiments were performed to examine if the DHFR2 protein is only expressed and functional under a particular cellular condition. The DHFR protein is considered to be one of the most important enzymes in maintaining the OCM and, subsequently, for supporting normal cell function. Decreased availability of the DHFR enzyme has a detrimental effect on overall cell health. It causes a deficiency in the cellular pools of thymidylate and purines, which hinders nucleic acid synthesis and inevitably results in cell death (Abali et al., 2008). Therefore, there was a possibility the genome encodes an alternative enzyme that can compensate for DHFR loss. There are numerous situations where two or more enzymes within a gene family can carry out a similar function. For example, the serine hydroxylmethyltransferase (SHMT) enzyme is an essential enzyme of the OCM pathway that exists in three forms: SHMT1/SHMT2α versions localise to the cytoplasm and SHMT2 in the mitochondria (Giardina et al., 2018). For this reason, the next hypothesis was that the DHFR2 protein was synthesised to carry out the reductase activity in a cell when the DHFR protein expression was eliminated or significantly reduced. DDA LC-MS/MS analysis on the HepG2 DHFR knockdown protein lysate did not identify any peptides relating to proteins of the dihydrofolate reductase gene family (Section 5.3.2). In other words, it revealed the CRISPR-Cas9 knockdown was successful at eliminating or drastically reducing the expression of the DHFR protein. Along with that, there was no proof that the DHFR2 protein was expressed in this cell line. The PRM LC-MS/MS strategy revealed similar results to the DDA LC-MS/MS analysis, in that the mass

spectrometer did not identify the DHFR2 protein. Therefore, this investigation found no evidence of the endogenous DHFR2 protein in the HepG2 DHFR knockdown cell line, indicating the DHFR2 protein is not synthesised to carry out the reductase activity in a cell when the expression of the DHFR enzyme is diminished or lost.

Given that there was no evidence of the DHFR2 protein from the previous experiments there was a possibility that the DHFR2 protein was subject to degradation. The ubiquitin-proteasome pathway is the primary method for intracellular protein degradation (Ciechanover, 1994). This pathway is used by eukaryotes to regulate protein quality. Damaged, unfolded or redundant proteins typically have intrinsic degradation signals, which results in them being covalently tagged with multiple ubiquitin proteins. The ubiquitin-tagged proteins are recognised by the 26S proteasome, which degrades them via proteolysis (Hershko & Ciechanover, 1992; Shang & Taylor, 2011). Therefore, HepG2 cells were treated with a proteasome inhibitor known as MG132 to examine if the native DHFR2 protein was being ubiquitinated. MG132 is a peptide aldehyde that inhibits the proteolytic activity of the 26S proteasome complex, resulting in a build-up of ubiquitinated proteins (Lee, 1998). The DDA LC-MS/MS analysis on the HepG2 cells treated with MG132 did not show any proof that the endogenous DHFR2 protein was expressed (Section 5.3.3). Various proteins associated with the OCM pathway were identified in this protein lysate, including the DHFR protein (Table 5.4). Similar results were observed when the PRM LC-MS/MS strategy was carried out on the protein lysate from the HepG2 cells treated with MG132 (Section 5.3.3). Therefore, this analysis found no evidence that the DHFR2 protein was susceptible to degradation via the ubiquitin-proteasome pathway. There are other mechanisms that perform cellular protein degradation including via the lysosomal pathway. This pathway mainly targets membrane-bound or organelle-associated proteins (Shang & Taylor, 2011). Therefore, this mode of protein degradation was not investigated for the DHFR2 protein as it more than likely does not fit into either of these categories.

Given that the endogenous DHFR2 protein had not been detected in any of the cell lines and tissues or in any of the cellular conditions created, the role of this gene remained unanswered. The focus returned to the original hypothesis that the *DHFR2* gene may play an important role during embryogenesis. It was concluded that the adult cell lines and tissues that were similar to the human developmental tissue were not fully representative of the cellular environment of human embryos and foetuses. Therefore, to determine if this hypothesis was in fact true, human embryonic/foetal tissue needed to be examined for the native DHFR2 protein. The regulations in Ireland do not allow research on human embryonic/foetal tissue. Therefore, this experiment was performed in the United Kingdom where human embryonic/foetal research is permitted. Kidney, heart, lung, liver and brain tissue were obtained from three different stages of human development: the Carnegie

stage 21-22, post-conceptual week 9-10 and post-conceptual week 15-17. MRM LC-MS/MS analysis that targeted DHFR2 peptides was carried out on this tissue in the Translational Mass Spectrometry research laboratories in UCL (Section 5.2.2). MRM LC-MS/MS analysis did not detect the DHFR2 specific peptides in the majority of the tissue samples examined (Section 5.3.4). One of the unique DHFR2 peptides was however detected in all three stages of the heart tissue (Section 5.3.4). The detection of only one peptide from a protein is common when the protein of interest is present at an extremely low abundance (Canha-Gouveia et al., 2019; Graves & Haystead, 2002). The detection of this peptide in this tissue implied that the endogenous DHFR2 protein is expressed in heart tissue during embryo and foetal development (Section 5.3.4 and Table 5.7 - Table 5.9). This result suggested that the DHFR2 protein could supply dihydrofolate reductase activity in the heart of rapidly growing embryos and foetuses.

This chapter has shown that the DHFR2 protein is an extremely challenging protein to study, even when the gold standard method for identifying proteins was utilised. Despite interrogating a panel of adult cells/tissues there was no evidence that the native DHFR2 protein was synthesised at a detectable abundance level in any of the cell lines and tissues examined. It was shown that the DHFR2 protein was not synthesised to carry out the reductase activity in a cell when the DHFR protein expression is diminished or lost. This chapter's results also demonstrated that the endogenous DHFR2 protein is unlikely to be ubiquitinated and rapidly degraded. However, an extremely significant finding was found when interrogating the human embryo and foetal tissue. This analysis found that the endogenous DHFR2 protein was expressed in heart tissue during embryo and foetal development. This suggested that the DHFR2 protein could supply dihydrofolate reductase activity in the heart during early development. This is a significant finding for this project however, functional analysis needs to be performed on the native DHFR2 protein found within this heart tissue to determine if it harbours dihydrofolate reductase activity or if it has a different cellular role; for example in the endothelial nitric oxide pathway like the DHFR enzyme (Chapter 1, Section 1.5.1.5 and Figure 1.8) (Crabtree et al., 2009, 2011). It is also worth noting that adult heart tissue has not been examined for the native DHFR2 protein. Therefore, TDA LC-MS/MS analysis should be performed on adult heart tissue to determine if the DHFR2 protein is a heart-specific protein or a developmental heart-specific protein.

There are some limitations associated with this chapter as a whole. The adult cell/tissue specific, DHFR complementation and ubiquitination experiments were carried out in DCU, Ireland. Whereas the human embryo/foetal tissue experiment was performed at UCL, London. Performing these experiments in different facilities meant the sample preparation, instrumentation, TDA LC-MS/MS method and software were different. Both mass spectrometers and TDA LC-MS/MS methodologies

were highly sensitive and specific. This is verified by the detection of the low abundance recombinant DHFR2 protein in both DCU and UCL. Therefore, both facilities should have been able to detect the endogenous DHFR2 protein if expressed at a low abundance. However, the MRM LC-MS/MS method employed in UCL would have superior sensitivity (Tang et al., 2022). Therefore, if the native DHFR2 protein was present at an extremely low abundance in the adult cell/tissue specific, DHFR complementation and ubiquitination samples the PRM LC-MS/MS method employed in DCU may have missed it. Therefore, the adult cell/tissue specific, DHFR complementation and ubiquitination samples should be reanalysed using the preparation, instrumentation, MRM LC-MS/MS method and software employed in UCL, London to confirm the results obtained in DCU.

Chapter 6. Are the DHFR2 transcripts subject to translational inhibition via paraspeckle nuclear retention?

6.1. Introduction

6.1.1. Overview

Out of all the cell lines and tissues examined for the endogenous DHFR2 protein in Chapter 5, the protein was only detected in human embryonic and foetal heart tissue. This result suggests that the cellular role of the DHFR2 protein is tissue specific. It also implies that the translation of the DHFR2 protein is tightly regulated or completely blocked in certain cells and tissues. Furthermore, there is ample evidence that different DHFR2 RNA isoforms are expressed in all cell lines and tissues examined to date (McEntee et al., 2011, Chapter 3). Both of these results suggest that the regulation of the DHFR2 protein most likely occurs before the DHFR2 RNA is translated into a protein. Structural features and regulatory sequences within an RNA molecule are the main mechanisms that facilitate translational inhibition (Gebauer & Hentze, 2004). However, alternative mechanisms have been identified over the last decade that can control the expression of a gene, this includes paraspeckle nuclear retention. Paraspeckles are located in the interchromatin space of the nucleus and bind to RNA molecules that have undergone a high level of adenosine (A) to inosine (I) editing. The binding of these edited RNA molecules to the paraspeckle prevents them from being exported into the cytoplasm, ultimately inhibiting translation of the RNA molecule (Bond & Fox, 2009). Paraspeckle nuclear retention could be the reason why the DHFR2 protein is undetectable in certain cell and tissue types, despite high levels of the DHFR2 transcripts being detected. Therefore, the aim of this chapter was to determine could the DHFR2 transcripts be translationally regulated by paraspeckle nuclear retention.

6.1.2. The nucleus and its subdomains

The nucleus of a eukaryotic cell is made up of multiple different structural compartments that make it extremely robust. One of the main regions in the nucleus is the interchromatin space. The interchromatin compartment consists of multiple channels in and around the individual chromosomal territories (Matera, 1999). The interchromatin space contains substructures such as Cajal bodies, promyelocytic leukaemia bodies, nuclear speckles and paraspeckles. These subnuclear bodies are thought to regulate multiple cellular mechanisms such as ribosome biogenesis, transcription and RNA splicing (Anantharaman et al., 2016; Fox & Lamond, 2010).

6.1.3. The paraspeckle

Paraspeckles are one of the most recent subnuclear bodies to be discovered (Andersen et al., 2002; Fox et al., 2002). Fox et al., (2002) named them paraspeckles as they observed them in the interchromatin space near to, but not associated with the nuclear speckles. Paraspeckles are small in size (of 0.5–1 μm in diameter) and have an irregular, oblong shape. Each nucleus contains between 5 – 20 foci that are unevenly distributed across the interchromatin compartment (Fox et al., 2002). Paraspeckles have been identified in the nucleus of all mammalian cell lines and tissues examined (Andersen et al., 2002; Clemson et al., 2009; Fox et al., 2002; Prasanth et al., 2005; Sunwoo et al., 2009); however, human embryonic stem cells are the exception (Chen & Carmichael, 2009).

6.1.3.1. The formation of the paraspeckle

Paraspeckles are made up of the NEAT1 long noncoding RNA (lncRNA) and *Drosophila* Behaviour Human Splicing (DBHS) family of proteins. The DBHS proteins consist of the 54 kDa nuclear RNA- and DNA-binding protein/Non-POU-domain-containing octamer binding protein ($P54^{NRB}$/NONO), paraspeckle component 1 (PSPC1) and proline-glutamine rich splicing factor (PSF/SFPQ) proteins (Knott et al., 2016).

Paraspeckle formation occurs near the site of NEAT1 lncRNA transcription to prevent the RNA from being immediately exported out of the nucleus (Sasaki et al., 2009). Once both NEAT1 transcripts are transcribed by RNA polymerase II, the longer NEAT1 transcript known as NEAT1_202 forms a complex with the mammalian DBHS family of proteins (Fox et al., 2005; Myojin et al., 2004). Multiple NEAT1 RNA–DBHS protein complexes bind together to form a structural scaffold that is the backbone of the paraspeckle (Figure 6.1) (Fox & Lamond, 2010). The shorter NEAT1 transcript known as NEAT1_201 has been found to bind to the peripheral area of the paraspeckle and is thought to have a regulatory role (Naganuma & Hirose, 2013).

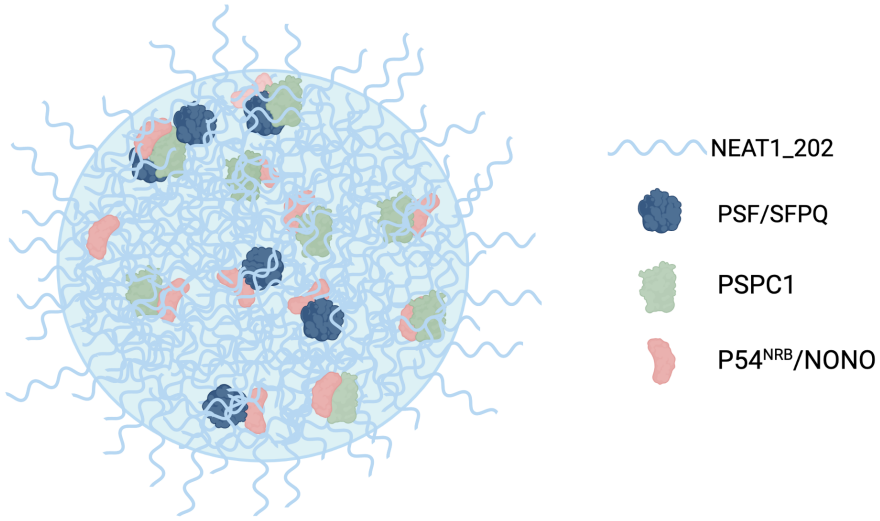


Figure 6.1 The structure of a paraspeckle. Once the NEAT1_202 lncRNA is transcribed it forms a complex with the DBHS family of proteins (P54^{NRB}/NONO, PSPC1 and PSF/SFPQ proteins). *Several NEAT1 RNA–DBHS protein complexes bind together to form a paraspeckle. Created with BioRender.com, Adapted from Fox & Lamond, 2010.*

6.1.3.2. The function of the paraspeckle

Paraspeckles are involved in regulating the expression of specific genes in differentiated cells. They have been linked to regulating specific cellular processes including stress responses, differentiation and circadian rhythm maintenance. These responses have been associated with the progression of cancer, neurodegenerative disease and viral infection (Pisani & Baron, 2019; Wang & Chen, 2020).

6.1.3.3. Mechanism of paraspeckle-mediated gene expression

Paraspeckles regulate the expression of genes by targeting the RNA intermediate. The RNA molecules have to contain specific elements that are subject to RNA editing to be recognised by the paraspeckle. More specifically, the RNA molecules must have inverted repeated Alu (IRAlu) elements in their 3' untranslated region (UTR). Interactions between these elements result in them hybridising together and forming a double-stranded RNA (dsRNA) molecule (Chen & Carmichael, 2012). The IRAlu elements within the dsRNA molecule are subsequently prone to adenosine (A) to inosine (I) editing. A group of enzymes known as adenosine deaminases (ADAR) catalyses the hydrolytic deamination of C6 of the adenosine residue, which converts it to inosine (Bass & Weintraub, 1988; Chen & Carmichael, 2009; Wagner et al., 1989). Edited RNA molecules subsequently bind to the paraspeckle. More precisely, the inosine residues in the edited RNA

molecules have an extremely high affinity for the P54^{NRB}/NONO paraspeckle protein (Zhang & Carmichael, 2001). The binding of this RNA molecule to the paraspeckle prevents it from being exported to the cytoplasm for translation (Figure 6.2). Therefore, this type of nuclear retention regulates the expression of specific proteins (Bond & Fox, 2009; Chen & Carmichael, 2009). This mechanism can retain RNA molecules indefinitely or can release the RNA molecules into the cytoplasm in response to specific stimuli. For example, signals released in response to cellular stress can mediate the release of the RNA molecules from the paraspeckle (Prasanth et al., 2005).

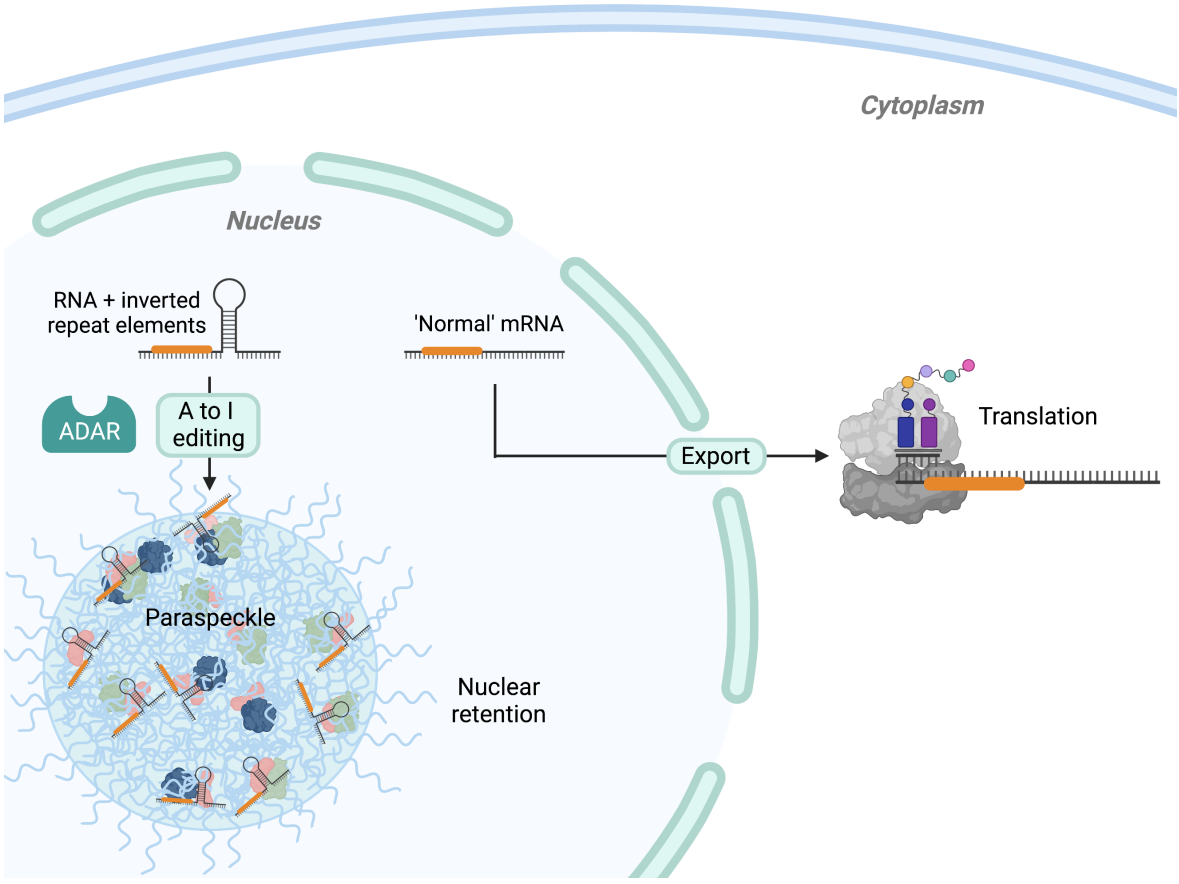


Figure 6.2 Schematic of 'normal' mRNA exportation out of the nucleus vs. RNA nuclear retention via paraspeckles. *All RNA molecules are transcribed in the nucleus from DNA. 'Normal' mRNA molecules are exported out of the nucleus into the cytoplasm where they bind to ribosomes to initiate translation. RNA molecules that contain inverted repeat elements in their 3'UTR can be subjected to A to I editing by the ADAR enzymes. This form of editing on an RNA molecule results in it binding to a paraspeckle which retains it in the nucleus. Created with BioRender.com, Adapted from Fox & Lamond, 2010.*

6.1.4. DHFR2's IRAlu elements and possible A to I RNA editing sites

Chen et al., (2008b) examined human mRNA and expressed sequence tags sequences from the UCSC genome browser and identified 333 genes within the human genome that had IRAlu elements in their 3'UTR. Subsequent analysis of complementary DNA (cDNA) sequences from Genbank showed the majority of the 333 genes were prone to A to I RNA editing. This result suggested that these genes could be retained in the nucleus by the paraspeckles (Chen et al., 2008b).

The *DHFR2* gene was one of the 333 genes identified to have IRAlu elements in its 3'UTR. However, Chen et al., (2008b) did not provide a list of which of the 333 genes that were found to be susceptible to A to I RNA editing. Even though there was evidence of the *DHFR2* gene containing IRAlu elements in its 3'UTR there was no proof that they were prone to A to I editing. Therefore, it was essential this was investigated to determine if the DHFR2 transcripts could be retained by the paraspeckle.

6.1.5. Aims and Objectives

The main aim of this chapter was to determine if the DHFR2 transcripts could be bound to paraspeckles and retained in the nucleus as a means to control its translation. To achieve this aim, the following objectives were proposed:

1. Confirm the DHFR2 transcripts contain IRAlu elements in their 3'UTR by bioinformatic analysis.
2. Investigate if any of the residues within DHFR2's IRAlu elements are subject to A to I editing by:
    a. Bioinformatically analysing deep transcriptome data available on an A to I RNA editing database known as REDIportal V2.0.
    b. RT-PCR amplifying DHFR2's IRAlu elements, cloning the PCR product into a vector, Sanger Sequencing the individual colonies and comparing the sequencing results to the reference genomic DNA (gDNA).

6.2. Methods

6.2.1. Analysing transcriptome data to determine if the IRAlu elements in the DHFR2 RNA are subject to A to I editing

REDIportal V2.0 (http://srv00.recas.ba.infn.it/atlas/) is the largest and most comprehensive collection of RNA editing events in humans. REDIportal V2.0 identified these editing sites by interrogating hundreds of deep transcriptome datasets (Picardi et al., 2017).

REDIportal V2.0 was first employed to confirm the DHFR2 transcripts contained IRAlu elements in its 3'UTR, as Chen et al., (2008b) stated. The JBrowse tab was selected, followed by hg38 and the chromosome location of the *DHFR2* gene was entered (chr3:94057901..94062998 (5.1 kb)). The short interspersed nuclear elements (SINEs) tab in the search results confirmed that the DHFR2_201 and DHFR2_202 transcripts contained IRAlu elements, known individually as AluJb_Alu and FLAM_A_Alu.

REDIportal V2.0 was then used to determine if any of the residues in DHFR2's IRAlu elements were A to I edited. The search option was selected, followed by search positions. The following parameters were selected: organism name – *Homo sapien*, genome version - hg38, gene name - *DHFR2*, location - Alu, genic region - 3'UTR. REDIportal V2.0 provided a list of all the A to I edits identified in the IRAlu elements in the DHFR2 transcripts. The A to I edits that were found in over a thousand samples were examined.

6.2.2. Determining the 3'UTR DHFR2 genomic sequence in the HepG2 cell line to use as the reference sequence for the investigative RT-PCR

6.2.2.1. Cell culture, DNA extraction and quantification

HepG2 cells were grown and pelleted according to Chapter 2, Section 2.2.2. Genomic DNA was extracted from the cell pellet as per Chapter 2, Section 2.2.3. The concentration of the DNA samples was determined according to Chapter 2, Section 2.2.5.

6.2.2.2. A to I genomic DNA (gDNA) PCR assay, PCR product clean-up and Sanger Sequencing

A PCR assay was designed to amplify the 3'UTR region of the DHFR2 transcripts that contained the Alu elements. This assay was employed to determine the genomic version of this region to use as the reference sequence for the investigative RT-PCR. PCR reactions were prepared according to Chapter 2, Section 2.2.11.2.1 using the HepG2 DNA extracted in Section 6.2.2.1. The reactions were incubated in the Thermal cycler and visualised on an agarose gel as per Chapter 2, Section 2.2.11.2.1. The PCR products were purified using the ExoSAP Clean-up reagent, as per Chapter 2, Section 2.2.12.1. The purified PCR products were sent for Sanger Sequencing as per Chapter 2, Section 2.2.16. The results were analysed using the tools mentioned in Chapter 2, Section 2.2.17. The resulting sequence was used as the reference sequence for the investigative RT-PCR.

6.2.3. Investigative RT-PCR to determine if the DHFR2 transcripts are subject to A to I editing in the HepG2 cell line

6.2.3.1. Cell culture, RNA extraction, RNA quantification and RNA integrity check

HepG2 cells were grown and pelleted according to Chapter 2, Section 2.2.2. RNA was extracted from these cell pellets as per Chapter 2, Section 2.2.4. The concentration of each RNA was quantified according to Chapter 2, Section 2.2.5. The integrity of all the RNA samples was assessed as per Chapter 2, Section 2.2.6.

6.2.3.2. DNase treatment and reverse transcription of RNA samples

The RNA samples from Section 6.2.3.1 were DNase treated and reverse transcribed using the ThermoFisher Scientific kit outlined in Chapter 2, Section 2.2.8.2. Random hexamer primers were used for the HepG2 RNA.

6.2.3.3. Genomic contamination check on the newly synthesised cDNA

The newly synthesised cDNA was checked for residual genomic contamination as per Chapter 2, Section 2.2.10.

6.2.3.4. A to I complementary DNA (cDNA) PCR assay

A PCR assay was designed to amplify the region in the 3'UTR of the DHFR2 transcripts that contained the AluJb_Alu element. PCR reactions were prepared according to Chapter 2, Section 2.2.11.2.2 using the newly synthesised HepG2 cDNA from Section 6.2.3.2. The reactions were incubated in the Thermal cycler and visualised on an agarose gel as per Chapter 2, Section 2.2.11.2.2.

6.2.3.5. Cloning of PCR product into a vector, purification of vector and Sanger Sequencing

The PCR products were inserted into the pJET1.2/blunt cloning vector according to Chapter 2, Section 2.2.13. The new vectors were transformed into competent *E. coli* cells and single colonies were isolated as per Chapter 2, Section 2.2.14. Fifteen single colonies were inoculated into a culture and left to grow overnight as per Chapter 2, Section 2.2.15. The vectors were isolated and purified as per Chapter 2, Section 2.2.15. The purified vectors concentration was determined as per Chapter 2, Section 2.2.5. The recombinant clones were sent for Sanger Sequencing as per Chapter 2, Section 2.2.16. The results were aligned back to the reference sequence (Section 6.2.2.2) using the tools mentioned in Chapter 2, Section 2.2.17 to determine if any of the clones were subject to A to I RNA editing.

6.3. Results

6.3.1. The REDIportal V2.0 database revealed multiple A to I edits in the DHFR2 IRAlu elements

REDIportal V2.0 database contains a catalogue of A to I RNA edits that have been identified in the human transcriptome. This data has been sourced from deep transcriptome sequencing of 9642 human samples from 549 individuals (Mansi et al., 2021).

REDIportal V2.0 was first employed to confirm the DHFR2 transcripts contained IRAlu elements in their 3'UTR, as Chen et al., (2008b) stated (Section 6.2.1). This database revealed that the DHFR2_201 and DHFR2_202 transcripts contained three Alu elements in their 3'UTR. These three Alu elements are known as AluSp_Alu, AluJb_Alu and FLAM_A_Alu (Appendix I and Figure 6.3). Of these three Alu elements, REDIportal V2.0 predicted AluJb_Alu and FLAM_A_Alu were IRAlu elements.

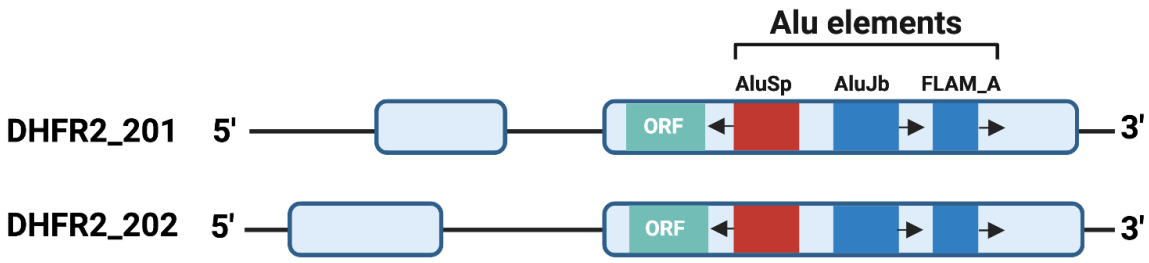


Figure 6.3 The location of the three Alu elements (AluSp_Alu, AluJb_Alu and FLAM_A_Alu) within the 3'UTR of the DHFR2_201 and DHFR2_202 transcripts according to REDIportal V2.0. *The AluJb_Alu and FLAM_A_Alu elements are on the negative strand, which is the same as the DHFR2 gene. Whereas the AluSp_Alu element is located on the positive strand. Created with BioRender.com, Adapted from REDIportal V2.0 (http://srv00.recas.ba.infn.it/atlas/).*

The REDIportal V2.0 database was subsequently employed to determine if the IRAlu elements located in the 3'UTR of the DHFR2 transcripts were prone to A to I RNA editing (Section 6.2.1). There were 53 A to I edits recorded in the AluJb element. Eight out of the 53 A to I RNA edits were present in over 1,000 samples (

Table 6.1 and Figure 6.4). Brain, heart and skin tissue was found to have the highest frequency of these eight A to I RNA edits. It is also worth noting that the A to I edits were quite frequent in liver, muscle, blood, adipose and lung tissue. The FLAM_A_Alu element did not have as many A to I RNA edits as the AluJb element. Only 12 A to I edits were found within this Alu element, all of which were present at a low frequency.

Table 6.1 High-frequency A to I RNA edits identified in the AluJb element by REDIportal V2.0

A to I RNA modifications number	A to I RNA Editing Position (hg38)	Alu element	Number of Samples identified in
#1	94059249	AluJb	1580
#2	94059352	AluJb	4611
#3	94059358	AluJb	3817
#4	94059360	AluJb	5908
#5	94059362	AluJb	1443
#6	94059370	AluJb	6070
#7	94059387	AluJb	4491
#8	94059495	AluJb	1489

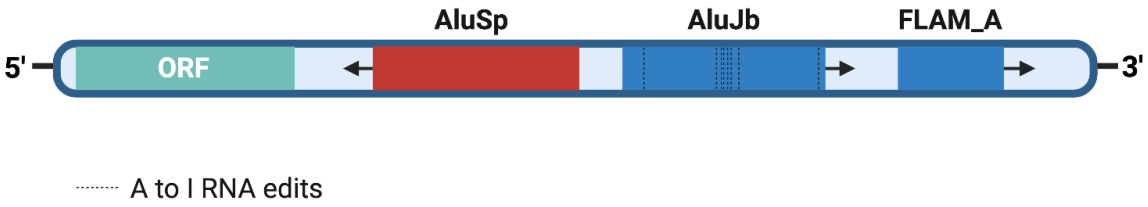


Figure 6.4 Exon 2 of the DHFR2_201 and DHFR2_202 transcripts showing the location of the ORF, three Alu elements in the 3'UTR and the eight A to I RNA edits (dashed line) listed in Table 6.1. *Created with BioRender.com, Adapted from REDIportal V2.0 (http://srv00.recas.ba.infn.it/atlas/).*

6.3.2. RT-PCR on the DHFR2 IRAlu elements revealed possible A to I RNA edited sites in HepG2 cells

The next logical step was to examine the transcriptome of a human cell line to determine if any of the A to I RNA edits identified in Section 6.3.1 were present. HepG2 is a human liver hepatocellular carcinoma cell line and it was picked for this analysis. There were two reasons why this cell line was chosen. The first being the DHFR2 protein had not been detected in this cell line (Chapter 5, Section 5.3.1). The second reason was that the analysis in Section 6.3.1 found that liver tissue had a high frequency of these edits.

6.3.2.1. The DHFR2 3'UTR genomic sequence in HepG2 cells revealed by PCR and Sanger Sequencing to use as the reference sequence for the investigative RT-PCR

The A to I gDNA PCR assay was designed to amplify the region within the *DHFR2* gene that was predicted to contain the three Alu elements (AluSp_Alu, AluJb_Alu and FLAM_A_Alu). This assay was performed on genomic DNA to reveal the reference sequence for the investigative RT-PCR. PCR results were visualised on an agarose gel (Figure 6.5), according to Section 6.2.2.2. There were no bands detected in the non-template control (NTC). As expected a band was visualised at 1670 bp in the two HepG2 gDNA samples. Sanger Sequencing (Appendix J) of this PCR product revealed the genomic sequence of these Alu elements.

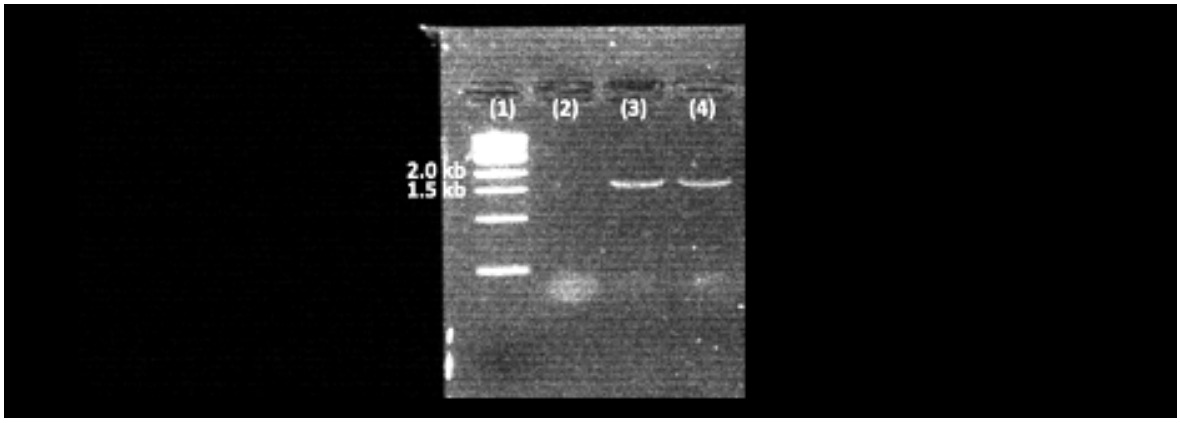


Figure 6.5 A to I gDNA reference PCR assay PCR product (7 µL) visualised on a 1% (w/v) agarose gel stained with SYBR Safe. *This assay was designed to amplify the 3'UTR of the DHFR2 gene to use as a reference sequence for the investigative A to I cDNA assays. This assay should result in a band at 1670 bp. This figure shows the results obtained for this PCR assay: (1) 1 kb DNA Ladder, (2) NTC and (3&4) HepG2 gDNA PCR reaction. This gel showed no bands in NTC. There were bands present at 1670 bp in the gDNA PCR reactions.*

6.3.2.2. A small number of A to I RNA edits were identified in the DHFR2 IRAlu elements in HepG2 cells

Prior to performing the investigative RT-PCR assay, there were several steps and quality control checks that needed to be carried out. The first check included assessing the integrity of the RNA prior to DNase treatment and reverse transcription. HepG2 RNA was run on an agarose gel to assess its integrity, according to Section 6.2.3.1. Figure 6.6 shows there were two distinct bands for the HepG2 RNA which represented the 28S and 18S rRNA structures. This result meant that the RNA was completely intact and was ready to be DNase treated and reverse transcribed into cDNA. Following reverse transcription, the newly synthesised cDNA had to be checked for residual genomic contamination prior to performing the investigative RT-PCR assay. The HepG2 cDNA was tested for genomic contamination using the intron spanning MTHFD1 PCR assay described in Chapter 2, Section 2.2.10. Figure 6.7 shows there were no bands present in the NTC. Furthermore, HepG2 cDNA samples were found to only have the cDNA band at 232 bp. Therefore, all cDNA samples passed this genomic contamination check and were ready to be used in the investigative RT-PCR assay.

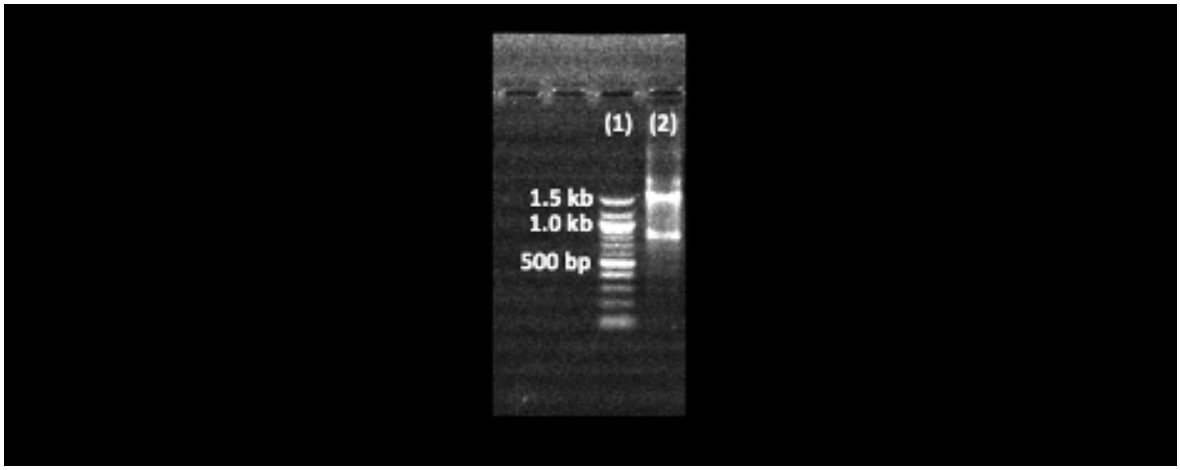


Figure 6.6 RNA integrity test on the extracted HepG2 RNA. RNA (500 ng) was resolved on a 1% (w/v) agarose gel stained with SYBR Safe. *Intact RNA results in two distinct bands, representing 28S and 18S. Whereas completely degraded RNA appears as a quite low molecular weight smear, with no distinct bands. The figure shows the results obtained for this test: (A) (1) 100 bp DNA Ladder and (2) HepG2 RNA. The 28S and 18S bands were visible in the HepG2 RNA, therefore it passed the integrity check.*

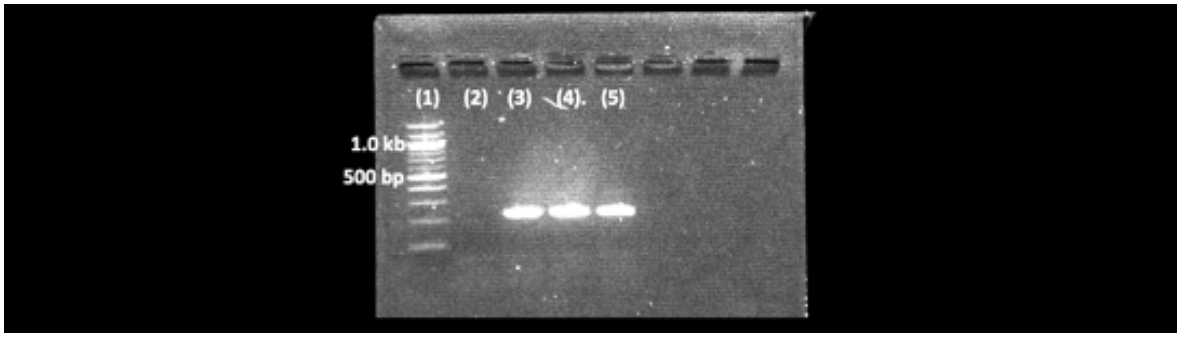


Figure 6.7 Genomic contamination check on the newly synthesised cDNA from HepG2 RNA. *The PCR products (10 µL) were resolved on a 1% (w/v) agarose gel stained with SYBR Safe. A band at 330 bp revealed gDNA whereas a band at 232 bp represented cDNA. The figure shows the results obtained for the genomic contamination test: (A) (1) 100 bp DNA Ladder, (2) NTC and (3-5) HepG2 cDNA PCR reactions. This figure revealed that there were bands visible at 232 bp in each sample, meaning all the samples passed the genomic contamination check.*

The next step was to perform the A to I cDNA PCR assay on the quality-controlled HepG2 cDNA. The A to I cDNA PCR assay was designed to amplify one of the IRAlu elements within the DHFR2_201 and DHFR2_202 transcripts. The two Alu elements were not targeted as it would result in an extremely long PCR product making it difficult to clone into an expression vector for Sanger Sequencing. The AluJb element was chosen to target over the FLAM_A_Alu element as the analysis in Section 6.3.1 found it had significantly more A to I RNA edits and the edits were present at a

much higher frequency. The PCR results for the A to I cDNA PCR assay were visualised on an agarose gel (Figure 6.8), according to Section 6.2.3.4. There were no bands detected in the NTC. As expected a band was visualised at 500 bp in the HepG2 cDNA sample.

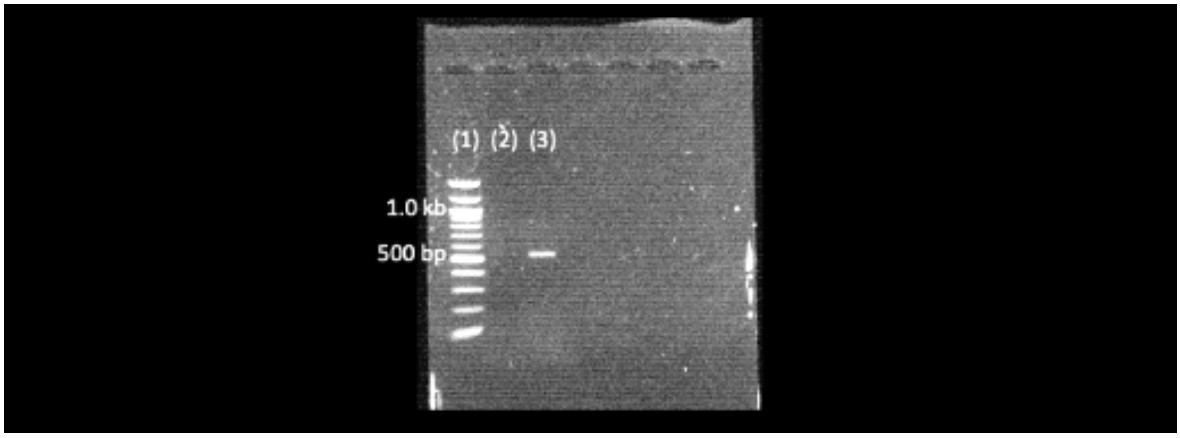


Figure 6.8 A to I cDNA PCR assays PCR product (7 µL) visualised on a 1% (w/v) agarose gel stained with SYBR Safe. *This assay was designed to amplify DHFR2's* AluJb element *to determine if any residues were prone to A to I RNA editing. This assay should result in a band at 500 bp. The figure shows the results obtained for this PCR assay: (1) 100 bp DNA Ladder, (2) NTC and (3) HepG2 cDNA PCR reaction. This gel showed no bands in NTC. There were bands present at 500 bp in the cDNA PCR reactions.*

The PCR products from Figure 6.8 could not be Sanger Sequenced at this point; this is due to inconsistent A to I editing across the same transcripts by the ADAR enzyme complex (Yablonovitch et al., 2017). This would result in the lower abundance A to I edits being missed and/or extremely messy Sanger Sequencing reads (Sakurai et al., 2014). Therefore, the PCR products were cloned into a vector according to Section 6.2.3.5 to isolate single PCR products and ultimately assess individual DHFR2 RNA molecules for A to I edits. Fifteen clones were isolated and subject to Sanger Sequencing. The sequencing results were compared to the gDNA reference sequence generated in Section 6.3.2.1 (Appendix K-Y). If an adenosine residue within an RNA molecule has been edited to an inosine nucleotide it will appear as a guanosine (G) residue in the cDNA sequence (Sakurai et al., 2014). The Sanger Sequencing results for each clone are summarised in Table 6.2. Thirteen out of the fifteen clones examined aligned perfectly back to the genomic reference sequence (Appendix K, M, N, O, P, R, S, T, U, V, W, X and Y). This result meant that none of the RNA residues within these clones was subject to any type of RNA editing, including A to I editing. The A to I HepG2 clone number two had one nucleotide that did not align with the genomic reference sequence. The expected adenosine nucleotide at position 94059362 (hg38) was a guanosine (G) residue (Table 6.2,

Appendix L and Figure 6.9 (A)). Clone number seven also had two nucleotides that did not align with the genomic reference sequence. The expected adenosine nucleotide at positions 94059249 and 94059360 (hg38) were guanosine (G) residues (Table 6.2, Appendix Q and Figure 6.9 (B-C)). The replacement of the adenosine nucleotide with the guanosine residue in the cDNA sequence is indicative of an A to I RNA editing site. Therefore, the guanosine (G) residues in clones two and seven that were expected to be adenosine nucleotides were sites of A to I RNA editing.

Table 6.2 The A to I cDNA PCR assay plasmids Sanger Sequencing results

A to I HepG2 clone number	A to I RNA edits within AluJb Alu element	A to I RNA modifications number	A to I RNA Editing Position (hg38)	See Appendix
#1	x	-	-	K
#2	✓	#5	94059362	L
#3	x	-	-	M
#4	x	-	-	N
#5	x	-	-	O
#6	x	-	-	P
#7	✓	#1 #4	94059249 94059360	Q
#8	x	-	-	R
#9	x	-	-	S
#10	x	-	-	T
#11	x	-	-	U
#12	x	-	-	V
#13	x	-	-	W
#14	x	-	-	X
#15	x	-	-	Y

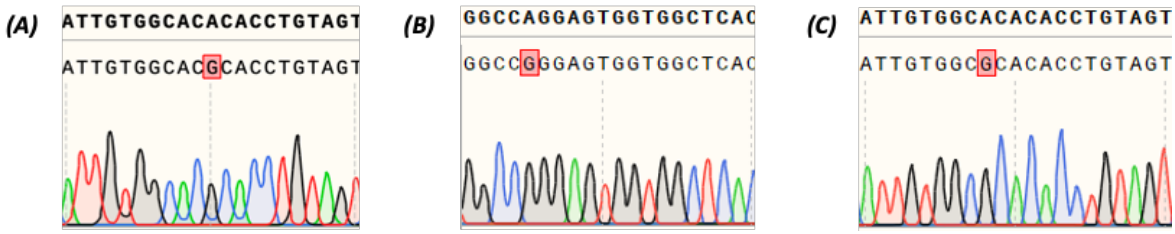


Figure 6.9 Sections of the Sanger Sequencing chromatograms showing the sites where the adenosine (A) nucleotide has been A to I edited. The inosine nucleotide appears as guanosine (G) in the cDNA sequence. *(A) A to I HepG2 clone number two showing the expected adenosine nucleotide at position 94059362 (hg38) was guanosine (red box). (B) Clone number seven showing the expected adenosine nucleotide at position 94059249 (hg38) was guanosine (red box). (C) Clone number seven showing the expected adenosine nucleotide at position 94059360 (hg38) was guanosine (red box). Performed on Snapgene.*

6.4. Discussion

Post-transcriptional modifications of RNA molecules are essential for many biological processes and ultimately for normal cellular function. A to I editing is the most common type of RNA editing that occurs in a cell (Ouyang et al., 2018). A to I RNA editing is responsible for multiple cellular mechanisms including gene expression regulation (Burns et al., 1997; Higuchi et al., 1993; Levanon et al., 2005), alternative splicing (Rueter et al., 1999), nonsense-mediated RNA decay (Agranat et al., 2008), RNA interference (Bass, 2006) and miRNA-mediated translational repression (Borchert et al., 2009). The role of A to I RNA editing in RNA nuclear retention (Ma et al., 2021) was the one of interest for this chapter. A to I edits within IRAlu elements in the 3'UTR of an RNA molecule results in it binding to the paraspeckle. Therefore, the exportation of the RNA molecule into the cytoplasm is blocked and translation is inhibited (Prasanth et al., 2005). This type of regulation could be one of the reasons why the DHFR2 protein is undetectable in certain cells and tissues despite high levels of the DHFR2 transcripts being recorded. Therefore, in this chapter, the DHFR2 transcripts were examined to determine if this could be the case.

For RNA molecules to be retained by the paraspeckle it first must contain IRAlu elements in its 3'UTR. These elements ultimately bind together to form a dsRNA molecule which is the optimal structure for paraspeckle binding (Bass & Weintraub, 1988). Therefore, the first step was to determine if any of the DHFR2 transcripts contained IRAlu elements in their 3'UTR. The REDIportal V2.0 database predicted that only the DHFR2_201 and DHFR2_202 transcripts contained Alu elements in their 3'UTR. Furthermore, two (AluJb_Alu and FLAM_A_Alu) out of three of these

elements were IRAlu elements (Section 6.3.1 and Figure 6.3). Interactions between the AluJb_Alu and FLAM_A_Alu elements may result in them hybridising together and forming dsRNA molecules, which is the ideal structure for paraspeckle binding (Bass & Weintraub, 1988; Chen & Carmichael, 2012). The remaining DHFR2 RNA isoforms did not contain the AluJb_Alu and FLAM_A_Alu elements in their 3'UTR; therefore, they were eliminated from further analysis.

For dsRNA molecules to be retained by the paraspeckle the IRAlu elements within the 3'UTR must also be subject to A to I editing. This is due to the $P54^{NRB}$/NONO paraspeckle protein having a much higher affinity for the inosine residue (Zhang & Carmichael, 2001). Two different methods were subsequently employed to determine if the IRAlu elements in the 3'UTR of the DHFR2_201 and DHFR2_202 transcripts were prone to A to I RNA editing.

The first approach utilised an RNA editing database known as REDIportal V2.0 to examine DHFR2's IRAlu elements (Section 6.2.1). This database contains a catalogue of A to I RNA edits that have been identified in the human transcriptome through deep transcriptome sequencing of 9642 human samples (Mansi et al., 2021). REDIportal V2.0 identified several A to I edits in the AluJb_Alu and FLAM_A_Alu IRAlu elements. The AluJb_Alu element contained over four times more A to I RNA edits than the FLAM_A_Alu element. Moreover, 15% of the A to I edits identified in the AluJb_Alu element by REDIportal V2.0 were present at a high frequency (≥ 1000 samples) (Table 6.1). These RNA modifications were most commonly found in the brain, heart, skin, liver, muscle, blood, adipose and lung tissue. Whereas all the A to I edits identified in the FLAM_A_Alu element were present at a lower frequency (≤ 1000 samples). Therefore, this result showed evidence that the IRAlu elements within some of the DHFR2_201 and/or DHFR2_202 transcripts were A to I RNA edited.

A conventional experimental method was subsequently employed to determine if the IRAlu elements in the 3'UTR of the DHFR2_201 and DHFR2_202 transcripts were subject to A to I RNA editing (Sakurai et al., 2014). This consisted of comparing cDNA sequences of the AluJb_Alu element with the corresponding genomic reference sequence using RT-PCR, cloning and Sanger Sequencing (Section 6.2.2 and 6.2.3). The two Alu elements were not targeted in the PCR assay as it would result in an extremely long PCR product making it difficult to clone into an expression vector for Sanger Sequencing. The AluJb_Alu element was chosen to target over the FLAM_A_Alu element as the analysis in Section 6.3.1 found it had significantly more A to I RNA edits and the edits were present at a much higher frequency. The experiment was performed on HepG2 cells which is a human liver hepatocellular carcinoma cell line. There were two reasons why this cell line was chosen, the first being the DHFR2 protein had not been detected in this cell line (Chapter 5, Section 5.3.1). The

second reason being the analysis carried out in Section 6.3.1 found that there was a high frequency of A to I RNA edits in the AluJb_Alu element in the liver tissue samples. The Sanger Sequencing of the A to I cDNA recombinant clones identified two clones to have A to I RNA edits. Several measures were taken to prevent false-positive results from occurring including designing a relatively small PCR product (500 bp), the use of a high-fidelity DNA polymerase during the RT-PCR, recommended number of PCR cycles (30 cycles) and Sanger Sequencing of both strands of the PCR product (Hestand et al., 2016; Potapov & Ong, 2017; Victoria Wang et al., 2012). Given that these measures were executed it is unlikely that the A to I edits identified on two of the clones were PCR or Sanger Sequencing errors. Thus, the result from this experiment shows evidence that the IRAlu elements within some of the DHFR2_201 and/or DHFR2_202 transcripts were A to I RNA edited.

To conclude, this chapter confirmed that the DHFR2_201 and DHFR2_202 transcripts contain IRAlu elements in their 3'UTR. It also showed that some of the transcripts were prone to A to I RNA editing. These edited transcripts may have bound to the $P54^{NRB}$/NONO paraspeckle protein which resulted in them being retained in the nucleus (Zhang & Carmichael, 2001). Thus, decreasing the amount of readily available DHFR2_201 and DHFR2_202 transcripts in the cytoplasm to be translated into a protein, ultimately reducing the chances of the DHFR2 protein being expressed (Pisani & Baron, 2019; Zhang & Carmichael, 2001; Zinshteyn & Nishikura, 2009). As a result of these findings, it is concluded that paraspeckle nuclear retention could be one of the mechanisms that is inhibiting the translation of the DHFR2 transcripts in certain cell and tissue types. However, it is not clear the level at which the IRAlu elements need to be A to I edited for the RNA molecule to be retained by the paraspeckle. Therefore, further experiments need to be performed to determine if some of the DHFR2_201 and/or DHFR2_202 RNA isoforms are targeted by the paraspeckle. For example, performing a co-localisation experiment on the DHFR2_201/DHFR2_202 RNA and the paraspeckle protein to determine if they are situated in the same location in the cell. There are multiple fluorescent-based techniques that can be employed to recognise specific nucleic acids and/or proteins including fluorescence in situ hybridization (FISH) or hybridisation chain reaction (HCR) (Evanko, 2004; Meyer et al., 2017).

Moreover, this chapter showed that the majority of the DHFR2 transcripts were not subjected to A to I editing. This result suggests that these transcripts were not bound to the $P54^{NRB}$/NONO paraspeckle protein and retained in the nucleus. Thus they were free to be exported into the cytoplasm for translation. Given this information and the fact that the DHFR2 protein was not detected in the HepG2 cells or other cell lines/tissues (Chapter 5, Section 5.3.1), there must be other regulatory mechanisms inhibiting the translation of these unedited DHFR2 RNA molecules. Further analysis needs to be carried out to determine these additional mechanisms.

Chapter 7. Modulating the expression of the human *DHFR2* gene to determine any additional biological functions

7.1. Introduction

7.1.1. Overview

The results from the previous chapters have given a good indication of the expression and translation patterns of the DHFR2 RNA and protein. Whilst this information suggests that the DHFR2 protein may have an important cellular role in embryonic and foetal heart tissue it is yet to be proven that the protein is functional. As working with human embryonic tissue is not possible in the Republic of Ireland this part of the project could not be explored further at DCU. To keep the project moving forward all the results gathered for the *DHFR2* gene were assessed. One result became of huge interest; the fact that the DHFR2 RNA was highly abundant in multiple cell lines and tissues (McEntee et al., 2011) despite the DHFR2 protein not being detected in the same samples (Chapter 5). This result suggests that the *DHFR2* gene may have a biological role at the RNA level, possibly as a regulatory RNA. Therefore, the focus of the project shifted to determining does the *DHFR2* gene have a second biological role.

Altering the expression of a specific gene is one of the most powerful mechanisms for determining its biological role (Prelich, 2012). Therefore, in this chapter, the expression of the *DHFR2* gene was modulated in a cell line where the DHFR2 protein was not identified i.e., the HepG2 cell line (Chapter 5, Section 5.3.1). Differential expression analysis on the DHFR2 modulated cell lines compared to their wildtype counterparts would reveal any links between *DHFR2* and other genes within the human genome. To do this a HepG2 DHFR2 knockout model and a HepG2 DHFR2 overexpression model was generated. The protein expression within these models was subsequently compared to the HepG2 control sample using liquid chromatography-tandem mass spectrometry (LC-MS/MS) to identify proteins that displayed differential expression.

7.1.2. Modulating the expression of a specific gene for functional analysis

Modulating the expression of a specific gene is a powerful mechanism for determining its biological role and physiological significance. The expression of a gene can either be down or up-regulated (Prelich, 2012).

7.1.2.1. The technique for down-regulating the expression of a target gene

Gene knockout is a technique used to completely and permanently silence a specific gene. The process of knocking out a gene consists of two major steps. Firstly, a double-stranded break (DSB) is generated in the genome of the target gene by a nuclease. There are several nucleases on the market to do this including zinc fingers nucleases (ZFNs), transcription-activator-like effectors (TALEs) and clustered regulatory interspaced short palindromic repeats (CRISPR) coupled with the CRISPR-associated protein 9 (Cas9) (Laufer & Singh, 2015). Secondly, the gap in the genome is repaired via non-homologous end-joining (NHEJ), microhomology-mediated end joining (MMEJ) or homology-directed repair (HDR) (Li et al., 2020). NHEJ directly ligates the two DNA extremities which often causes indels at the site of repair resulting in frameshifts within the gene (Lieber, 2010). MMEJ uses short homologies at each side of the DSB to mediate the joining of the two extremities, this often results in small deletions within the gene (Sfeir & Symington, 2015). HDR introduces a DNA sequence via a donor template at the DSB site while repairing the gap in the genome (Nambiar et al., 2019). Ultimately these three repair mechanisms can permanently disrupt the gene of interest (Li et al., 2020).

7.1.2.1.1.CRISPR/Cas9 and NHEJ systems

The CRISPR/Cas9 system is the superior method for generating DSB within a target gene (Brunner et al., 2019). It has been adapted from a naturally occurring genome editing system found in bacteria to defend against viruses (Jinek et al., 2012). The CRISPR/Cas9 system has two components: a single guide RNA (sgRNA) and the Cas9 nuclease (Figure 7.1). The sgRNA is composed of the CRISPR RNA (crRNA) and the trans-activating CRISPR (tracr) RNA. The crRNA is a 20 nucleotide sequence that is complementary to the target region. The tracr RNA is a section of RNA that acts as a binding scaffold for the Cas9 nuclease. Once the Cas9 nuclease binds to the tracr RNA its enzymatic activity is activated, thus causing the crRNA to guide the Cas9/sgRNA ribonucleoprotein complex to the target site where it binds to the target sequence (Jinek et al., 2013). The Cas9 enzyme requires a protospacer adjacent motif (PAM) sequence at the 5’ or 3’ of the target site to perform its function (Leenay & Beisel, 2017). Once everything is in place the Cas9 enzyme creates a DSB in the target sequence. Two sgRNA molecules can be designed to generate two DSB in the target DNA sequence, which will eliminate a section of the gene (Cong et al., 2013; Jinek et al., 2013).

A repair system is subsequently activated to fix the gap in the genome. The NHEJ system is a very efficient repair mechanism that is most active in the cell (Rodgers & McVey, 2016). The Ku complex recognises the broken ends of the DNA and recruits DNA-dependent protein kinase catalytic subunit (DNA-PKcs) in complex with the endonuclease Artemis (Figure 7.1). The phosphorylation of DNA-PKcs activates Artemis which inserts or removes nucleotides at the extremities of the DSB. Ligase IV, X-Ray Repair Cross Complementing 4 (XRCC4) and Cernunnos bind together to form a complex that ligates the broken ends together, thus repairing the double-stranded break (Chang et al., 2017). However, any nucleotide that was removed by the nuclease will be missing from the genome. Therefore, this repair mechanism is ideal if the intent is to knock out a gene (Davis & Chen, 2013).

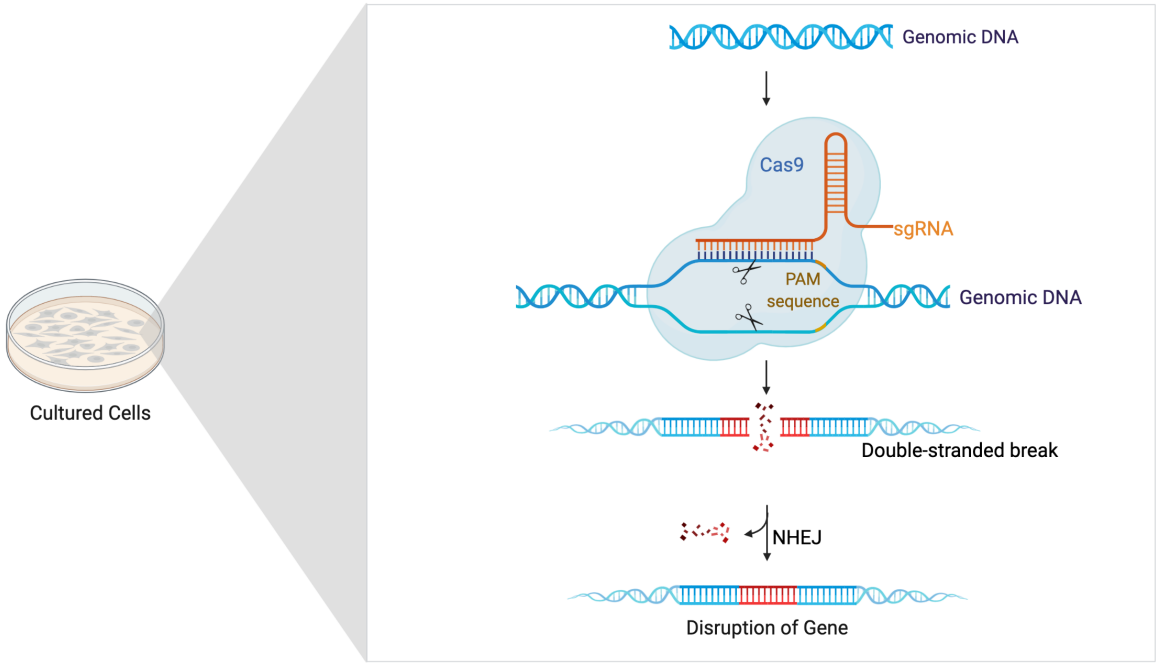


Figure 7.1 Schematic representation of the CRISPR/Cas9 and non-homologous end joining system to disrupt a gene within the genome of cultured cells. *The sgRNA molecule binds and activates the Cas9 enzyme. The sgRNA part of the complex binds to the target gene and directs the Cas9 enzyme to create a double-strand break. For this double-strand break to occur Cas9 enzyme needs a PAM sequence located at the 5' or 3' of the target. Two separate cuts can occur to remove a section of DNA from the genome. There are multiple repair mechanisms that can be employed to eliminate the gap in the genome, including NHEJ. NHEJ directly joins the two DNA extremities through a series of digestion/polymerisation steps. This process often causes indels at the site of repair resulting in a frameshift. Therefore, the break in the genome is repaired but the gene is disrupted. Created with BioRender.com, Adapted from Li et al., 2020.*

For the reasons stated above, the CRISPR/Cas9 and NHEJ systems were employed to knockout the *DHFR2* gene from the genome of the human HepG2 cell line.

7.1.2.2. The technique for overexpressing a specific gene

Gene overexpression is a technique used to dramatically increase the expression levels of a gene's mRNA and protein in a cell or organism (As discussed in Chapter 4). This mechanism can also be exploited to determine the biological role of a target gene (Prelich, 2012).

Within the cell, the endogenous mRNA and protein belonging to the gene of interest are still synthesised at a normal level (Figure 7.2). However, a recombinant expression system is also utilised in the cell to express extremely high levels of the gene's mRNA and protein. This is achieved by transfecting the cells with an expression vector that has been genetically engineered to contain the coding sequence of the gene of interest (Capriotti et al., 2020). The expression vector contains all the appropriate signals for transcription and translation to occur in the cell, including an enhancer, promoter, terminator and a polyadenylation tail motif (Makrides, 2003). Once the expression vector is in the cell it exploits the cellular machinery to transcribe the gene into recombinant mRNA and subsequently translate it into the recombinant protein (Figure 7.2).

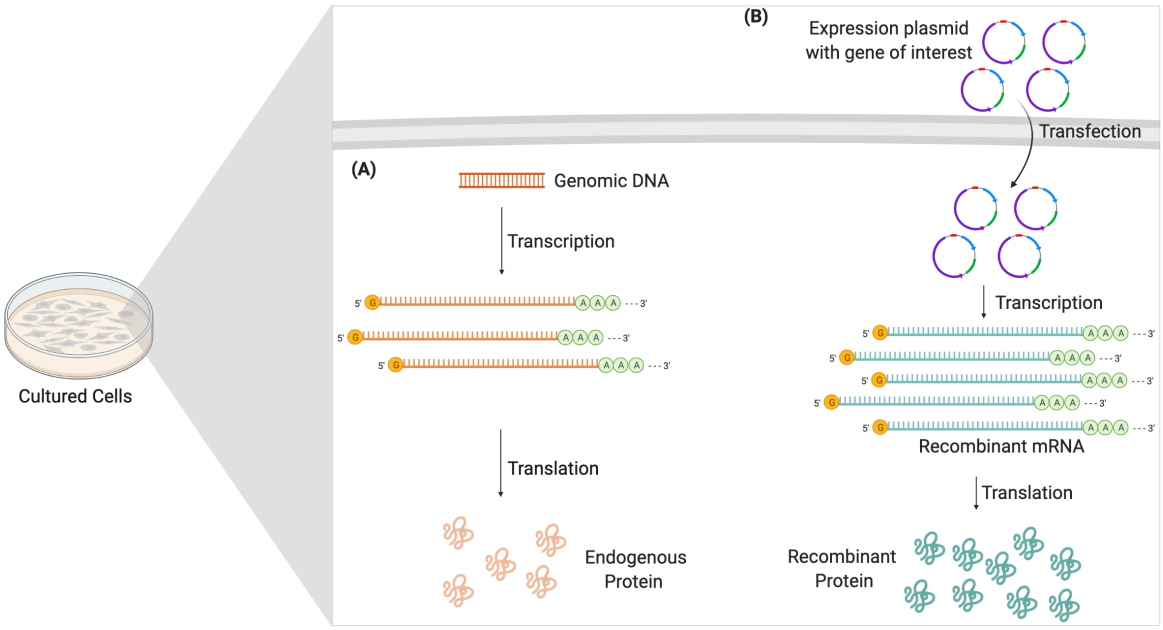


Figure 7.2 Schematic diagram showing the different mechanisms cells use to generate endogenous protein and recombinant protein. *Pathway (A) illustrates the generation of endogenous proteins in cultured cells. A gene is transcribed into an RNA intermediate which is subsequently translated into a native protein. Pathway (B) shows the mechanism of how cells generate recombinant proteins. An expression vector containing the gene of interest is transfected into the cells. The cellular machinery transcribes the gene in the plasmid into recombinant mRNA. The mRNA molecules are then translated into a recombinant protein. Typically recombinant protein expression generates more protein compared to endogenous protein expression. Created with BioRender.com, Adapted from Capriotti et al., 2020.*

This type of gene overexpression system was employed to up-regulate the *DHFR2* gene in the human HepG2 cell line.

7.1.3. Differential expression analysis on knockout and overexpression models

Differential expression analysis is a technique used to identify statistically significant differences in expression levels between two or more experimental groups. This is the type of analysis employed to assess the impact of knocking out or overexpressing a specific gene. It has the potential to reveal the gene's cellular role (McDermaid et al., 2019).

Differential expression analysis can be carried out at the RNA level (e.g. RNA-seq or RT-qPCR) and the protein level (e.g. LC-MS/MS) (Kim et al., 2021; Shi et al., 2022). All techniques require a control sample per experimental condition and each sample to be prepared in triplicate. Once each individual sample has been analysed by RNA-seq, RT-qPCR or LC-MS/MS the statistical analysis can be performed. There are several data analysis programmes that can be employed to carry out this analysis including Python and R. There are also software tools that have been specifically developed for differential expression analysis. For example, the Progenesis QI, MaxQuant and Proteios Software have all been designed to assess label-free LC-MS/MS data. Progenesis QI has been found to be the most consistent tool for this type of differential expression analysis (Välikangas et al., 2018).

7.1.3.1. Progenesis QI data analysis programme

Progenesis QI normalises the data and calculates the mean relative abundance for each cellular condition. An analysis of variance (ANOVA) statistical test is subsequently performed on the mean values calculated for each cellular condition compared to its control sample. The ANOVA test returns a probability value (p-value) which reveals if the difference in expression is significant. A p-value less than 0.05 is considered statistically significant. If the expression of a protein has been found to be statistically significant Progenesis QI calculates the fold change between the cellular condition and the control sample (Jaiswal et al., 2017). The Progenesis QI software subsequently generates a report with the proteins that have been found to be significantly differentially expressed.

The Progenesis QI tool was used to determine which proteins were impacted when the *DHFR2* gene was knocked out and overexpressed in HepG2 cells.

7.1.4. Aims and Objectives

In this chapter, the impact of altering the expression of the human *DHFR2* gene on the HepG2 proteome was assessed. To achieve this aim, the following objectives were proposed:

1. Mass spectrometry-based differential protein expression analysis on a HepG2 DHFR2 knockout cell line compared to the HepG2 wildtype.
2. Mass spectrometry-based differential protein expression analysis on a HepG2 DHFR2 overexpression cell line compared to the HepG2 wildtype transfected with an empty vector.

7.2. Methods

7.2.1. Cell culture preparation for differential protein expression analysis

7.2.1.1. Cell culture of HepG2 wildtype, HepG2 DHFR2 knockout and HepG2 DHFR knockout cell lines

The HepG2 DHFR2 knockout cell line was generated as per Chapter 2, Section 2.2.1. The HepG2 DHFR knockout cell line was generated as a control according to Chapter 2, Section 2.2.1. The HepG2 wildtype, HepG2 DHFR2 knockout and HepG2 DHFR knockout cell lines were grown to 90% confluency in a T75 flask according to Chapter 2, Section 2.2.2. The confluent T75 flask containing the HepG2 wildtype cell line was split into four T25 flasks and grown for 48 hour according to Chapter 2, Section 2.2.2. This step was also carried out for the confluent T75 flask containing the HepG2 DHFR2 and HepG2 DHFR2 knockout cell line. Each flask was harvested and pelleted as per Chapter 2, Section 2.2.2. This work was performed by Paola Drago (DCU).

7.2.1.2. Cell culture of HepG2 wildtype transfected with pCMV6-AC-tGFP (empty vector) and HepG2 wildtype transfected with pCMV6-AC-DHFR2-tGFP

Eight T25 flasks containing the HepG2 wildtype cell line were prepared, at a cell density of $2x10^6$. DMEM (5 mL), supplemented with 10% (v/v) Fetal Bovine Serum (FBS) was added to each flask. All the flasks were incubated for 24 hour in a 5% carbon dioxide incubator at 37 °C. The pCMV6-AC-tGFP expression vector was transfected into four of the T25 flasks using the Lipofectamine 3000 Reagent, according to the manufacturer's protocol. The pCMV6-AC-DHFR2-tGFP vector was

transfected in the remaining four flasks as per the previous step. The transfected cells were incubated for 48 hour in a 5% carbon dioxide incubator at 37 °C. The cells were harvested and pelleted as per Chapter 2, Section 2.2.2. This work was performed by Paola Drago (DCU).

7.2.2. Protein extraction and LC-MS/MS sample preparation

Whole protein lysates were extracted from each cell pellet from Section 7.2.1.1 and 7.2.1.2 according to Chapter 2, Section 2.2.19.2. The protein lysate from each sample was prepared for LC-MS/MS analysis as per Chapter 2, Section 2.2.20.1 and Section 2.2.20.2.

7.2.3. LC-MS/MS analysis

LC-MS/MS analysis was carried out in the Proteomics Facility in the National Institute for Cellular Biology (NICB) in DCU. LC-MS/MS analysis consisted of reverse-phased capillary high-pressure liquid chromatography on the UltiMate 3000 RSLCnano System (Thermo Scientific) in line with an Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Scientific).

The lyophilised samples from Section 7.2.2 were resuspended in 15 µL of resuspension solution and then sonicated using a sonicating waterbath (Precision ultrasonic cleaning, DP201–00, Ultrawave) under the following conditions: power of 280 watts, frequency of 30 kilohertz, ambient temperature and time of 5 min. The peptide concentration was quantified using the Nanodrop One spectrophotometer (Labtech instruments), at an absorbance of 280 nm. The samples were diluted down to a working solution of 1 µg/µL using resuspension solution for LC-MS/MS analysis. One microlitre of the working solution was loaded onto a trapping column cartridge (PepMap100, C18, 300 µm x 5 mm) (Thermo Scientific) using 2% (v/v) acetonitrile, 0.1% (v/v) trifluoracetic acid at a flow rate of 25 µL/min for 3 min. The peptides were then resolved on AcclaimPepMap100 75 µm x 50 cm, nanoViper C18, 3 µm, 100 Å (Thermo Fisher Scientific) where a binary gradient was employed; solvent A and 2–25% solvent B for 100 min followed by solvent A and 25–50% solvent B for a further 20 min. The column flow rate was 300 nL/min and the column oven temperature was 35 °C. The Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Scientific) was set to data-dependent acquisition using a top speed approach (cycle time of 3 sec) with a voltage of 2.0 kV and a capillary temperature of 320 °C. Full scans within the 380-1500 *m/z* range were carried out in the Orbitrap mass analyser using a resolution of 120,000 (at *m/z* 200), automatic gain control (AGC) target value of $4x10^5$ and a maximum ion injection time of 50 ms. The top-speed acquisition algorithm was used to identify the number of selected precursor ions for fragmentation. The

Quadrupole selected the number of precursor ions for fragmentation with an isolation width of 1.6 Da. The peptides with a charge state between 2 + to 6 + were examined and a dynamic exclusion was applied after 60 sec. Precursor ions were fragmented using higher energy collision-induced dissociation with a normalised collision energy of 28%. The linear ion trap measured the resulting MS/MS ions under the following scan conditions: AGC target value of $2x10^4$ and maximum fill time of 35 ms.

7.2.4. Differential protein expression analysis using Progenesis QI

Each sample within each of the investigative cellular conditions were initially processed using the Progenesis QI for Proteomics software (version 3.1; Waters). A reference LC-MS file was automatically selected by this software. The retention times of all of the other samples were aligned to this reference run and the peak intensities were normalised. The normalised data was subsequently filtered based on (i) peptide features with ANOVA < 0.05 between experimental conditions, (ii) mass peaks (features) with charge states from +2,+3,+4 and (iii) greater than one isotope per peptide. MS/MS output files were generated, exported to ProteomeDiscoverer 2.2 software (Thermo Fisher Scientific) and compared with a reference data set from UniProtKB (taxonomy: *Homo sapiens* 26,592 sequences) to identify the peptides in the samples. The following search parameters were used for peptide identification: peptide mass tolerance of 20 ppm, MS/MS mass tolerance of 0.6 Da, two missed cleavages were allowed, carbamidomethylation of cysteine was defined as a fixed modification, and methionine oxidation defined as a variable modification. Furthermore, a false discovery rate (FDR) of ≤ 0.01 was selected to only identify highly confident peptides (identified using a SEQUEST HT workflow coupled with Percolator validation in Proteome Discoverer 2.2). Proteome Discoverer 2.2 generated reports with the peptides identified in the samples. These output files were imported back into the Progenesis QI software for further analysis. Protein identifications were reviewed, and only those which passed the following criteria were considered differentially expressed between experimental groups with high confidence and statistical significance: an ANOVA p-value of ≤ 0.05 between experimental groups; proteins with a minimum of 2 unique peptides contributing to the identification and a ≥ 1.5 fold change in relative abundance between the two experimental conditions. The proteins that were found to be significantly differentially expressed between the control and the investigative cellular condition were initially assessed for the OCM proteins. The proteins were subsequently imported into the QIAGEN Ingenuity Pathway Analysis (IPA) (https://digitalinsights.qiagen.com/products-

overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-ipa/) software and STRING (https://string-db.org/) database to perform network and protein function analysis.

7.3. Results

In a bid to determine if the human *DHFR2* gene had an additional role as a regulatory RNA a DHFR2 knockout and a DHFR2 overexpression cell line was generated (Section 7.2.1.1 and 7.2.1.2). These models were created in a cell line where the DHFR2 protein was not identified, i.e., the HepG2 cell line. The protein expression within these models were compared to a HepG2 control (wildtype for the knockout and wildtype transfected with pCMV6-AC-tGFP vector for the overexpression) using LC-MS/MS. Statistical analysis revealed proteins that displayed differential expression in the DHFR2 knockout and a DHFR2 overexpression cell line (Section 7.2.4).

7.3.1. Knocking out the dihydrofolate reductase genes impacted the expression of hundreds of other proteins in HepG2 cells

7.3.1.1. Eliminating the *DHFR2* gene from the HepG2 genome altered the expression of hundreds of other proteins

The *DHFR2* gene was knocked out of the genome of HepG2 cells by Paola Drago. The knockout of the *DHFR2* gene in the HepG2 cells was confirmed by PCR and reverse transcription PCR (RT-PCR). Protein-based methods were not used to confirm the successful knockout of the *DHFR2* gene as there was no evidence of the protein in wildtype HepG2 cells (Chapter 5, Section 5.3.1). The HepG2 DHFR2 knockout cell line was subsequently assessed to determine the impact of silencing the *DHFR2* gene on overall protein expression patterns. Mass spectrometry-based differential expression analysis compared protein expression between HepG2 wildtype cells and HepG2 DHFR2 knockout cells (Section 7.2.4). The differential expression analysis identified 493 proteins to be either significantly up-regulated or down-regulated in the HepG2 DHFR2 knockout cells (Appendix Z). Of these 493 proteins, the expression of 272 were down-regulated and the remaining 221 were up-regulated. The data was subsequently interrogated for the proteins within the OCM pathway (Table 7.1 and Figure 7.3). It revealed that the expression of four of the OCM proteins were altered when the *DHFR2* gene was eliminated from HepG2 cells. The expression of the aldehyde dehydrogenase 1 family member L1 (ALDH1L) (22.499 fold), DHFR (2.404 fold), serine hydroxymethyltransferase 1 (SHMT1) (1.543 fold) and thymidylate synthetase (TYMS) (5.249 fold)

protein was found to be significantly reduced in the HepG2 DHFR2 knockout cells (Table 7.1 and Figure 7.3). The DHFR, SHMT1 and TYMS enzymes are all involved in the de novo thymidylate synthesis pathway in OCM (Anderson et al., 2012). Therefore, the *DHFR2* gene may have a role in regulating this pathway through its interaction with one or all of these proteins.

Table 7.1 Impact of knocking out the human DHFR2 gene in the HepG2 cells on the expression of OCM proteins

Accession Number	Protein Name	Description	HepG2 WT vs. HepG2 DHFR2 knockout	
			ANOVA (p)	Fold
O75891	ALDH1L1	Aldehyde dehydrogenase family 1 member L1	0.00000432	↓ 22.499
Q3SY69	ALDH1L2	Aldehyde dehydrogenase family 1 member L2	-	-
Q93088	BHMT	Betaine--homocysteine S-methyltransferase	-	-
P00374	DHFR	Dihydrofolate reductase	0.00014019	↓ 2.404
Q86XF0	DHFR2	Dihydrofolate reductase 2	-	-
Q9UI17	DMGDH	Dimethylglycine dehydrogenase, mitochondrial	-	-
Q96DP5	MTFMT	Methionyl-tRNA formyltransferase, mitochondrial	-	-
Q99707	MS	Methionine synthase	-	-
P11586	MTHFD1	Methylenetetrahydrofolate dehydrogenase, cyclohydrolase and formyltetrahydrofolate synthetase 1	-	-
Q6UB35	MTHFD1L	Methylenetetrahydrofolate dehydrogenase, cyclohydrolase and formyltetrahydrofolate synthetase 1L	-	-
P13995	MTHFD2	Methylenetetrahydrofolate dehydrogenase, cyclohydrolase and formyltetrahydrofolate synthetase 2	-	-
Q9H903	MTHFD2L	Methylenetetrahydrofolate dehydrogenase, cyclohydrolase and formyltetrahydrofolate synthetase 2L	-	-
P42898	MTHFR	Methylenetetrahydrofolate reductase	-	-
Q9UL12	SARDH	Sarcosine dehydrogenase, mitochondrial	-	-
P34896	SHMT1	Serine hydroxymethyltransferase 1, cytosolic	0.00544504	↓ 1.534
P34897	SHMT2	Serine hydroxymethyltransferase 2, mitochondrial	-	-
P04818	TYMS	Thymidylate synthase	0.00093476	↓ 5.249

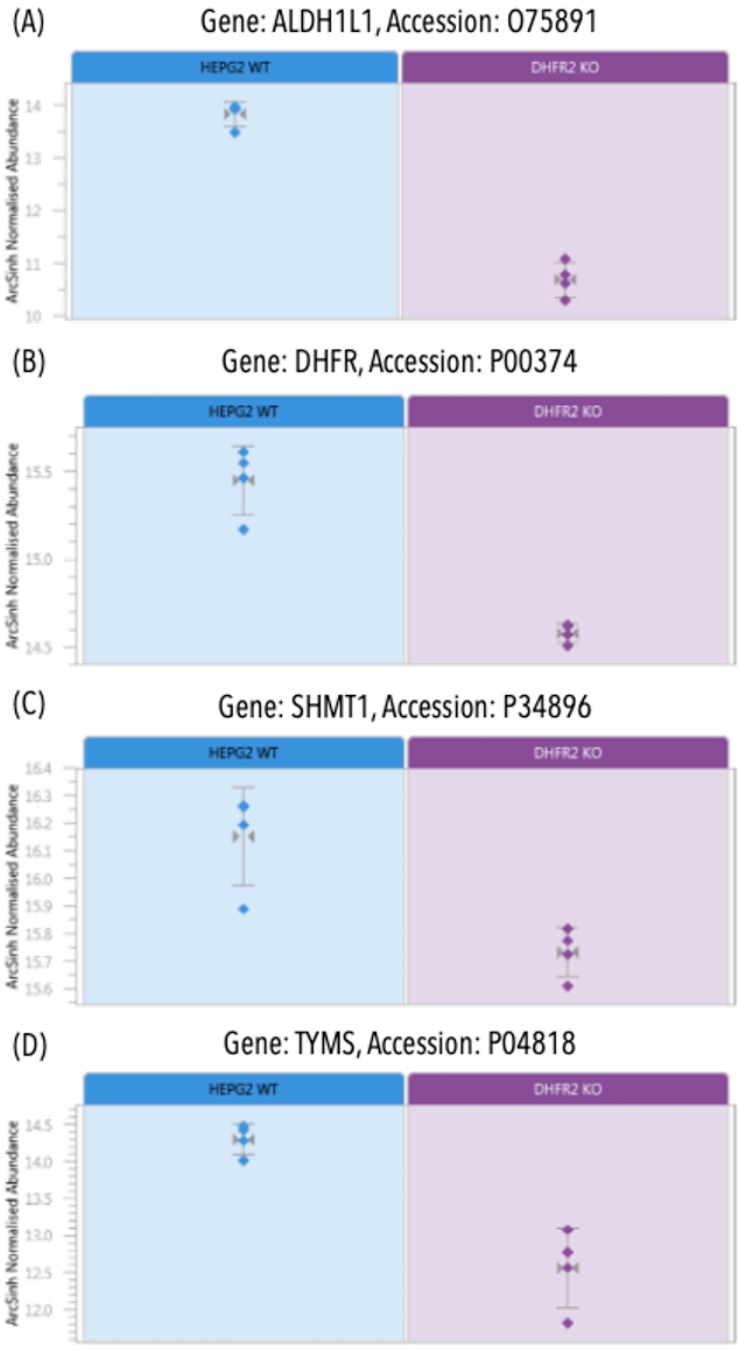


Figure 7.3 Compound abundance graphs showing the normalised abundance of (A) the ALDH1L1, (B) the DHFR, (C) the SHMT1 and (D) TYMS proteins within the four replicates for the HepG2 wildtype (blue) and the HepG2 DHFR2 knockout (purple) cell lines. *The expression of the ALDH1L1, DHFR, SHMT1 and TYMS proteins were all found to be significantly reduced in the HepG2 DHFR2 knockout cell lines compared to the HepG2 wildtype. See Table 7.1 for the p-values and fold change for these genes. Created on Progenesis QI.*

The Qiagen IPA software and STRING database were employed to identify the biological networks and pathways that were impacted in the HepG2 DHFR2 knockout cells. The STRING database calculated the protein-protein interaction (PPI) enrichment p-value to be < 1.0e–16 for the differential expressed proteins in the HepG2 DHFR2 knockout model. A low value like this indicated that this set of proteins was not random but in fact, there was a strong interaction among them (Hu et al., 2021). The network and protein function analysis carried out on Qiagen IPA revealed that the majority of these proteins were linked with the following networks: (i) small molecule biochemistry, lipid metabolism, cell morphology (Appendix AA); (ii) hereditary disorder, neurological disease, organismal injury and abnormalities (Appendix AB); (iii) cellular assembly and organisation, developmental disorder, hereditary disorder (Appendix AC). Appendix AA – AC shows diagrams of these networks in relation to the DHFR2 knockout cell model. The STRING database assessed the differential expressed proteins in the HepG2 DHFR2 knockout model more thoroughly and identified specific pathways that were impacted. The de novo thymidylate synthesis pathway and the response to methotrexate were found to be the most significantly affected processes in this dataset. Other networks that STRING identified to be impacted were the response to hydrostatic pressure, glucuronate catabolic process to xylulose 5-phosphate, de novo pyrimidine nucleobase biosynthetic pathway, fatty acid beta-oxidation using acyl-coa oxidase, cardiac muscle cell adhesion, mitochondrial ribosome assembly, alditol catabolic process, negative regulation of fibrinolysis and the tetrahydrofolate metabolic process.

7.3.1.2. Reducing the expression of the *DHFR* gene in HepG2 cells impacted hundreds of other proteins

A CRISPR/Cas9 system was employed to knockout the parental homologue of DHFR2 i.e., the *DHFR* gene in HepG2 cells, to use as a control for the HepG2 DHFR2 knockout model. The knockout of the *DHFR* gene in the HepG2 cells was assessed by PCR, RT-PCR, Western Blot, LC-MS/MS and enzymatic activity analysis. This analysis showed that the *DHFR* gene was not completely silenced however it did reveal that the expression of the protein was reduced by 1.626 fold (Table 7.2). This result meant that the *DHFR* gene was not completely silenced but that it was down-regulated. Therefore, this model will be referred to as the HepG2 DHFR knockdown cell line going forward. The HepG2 DHFR knockdown cell line was subsequently assessed to determine the impact of reducing the expression of the *DHFR* gene on overall protein expression. Mass spectrometry-based differential expression analysis compared protein expression between HepG2 wildtype cells and HepG2 DHFR knockdown cells (Section 7.2.4). The differential expression analysis identified 723 proteins to be

either significantly up-regulated or down-regulated in the HepG2 DHFR knockdown cells (Appendix AD). Of these 723 proteins, the expression of 389 were down-regulated and the remaining 334 were up-regulated. As expected, the expression of the DHFR protein was found to be reduced in the HepG2 DHFR knockdown model (Table 7.2 and Figure 7.4). The data was subsequently interrogated for the remaining proteins within the OCM pathway (Table 7.2 and Figure 7.4). It revealed that the expression of three other OCM proteins were altered when the *DHFR* gene was down-regulated in HepG2 cells. The expression of the ALDH1L (2.304 fold), sarcosine dehydrogenase (SARDH) (1.782 fold) and SHMT1 (1.613 fold) protein were found to be significantly reduced in the HepG2 DHFR knockdown cells (Table 7.2 and Figure 7.4). This result was not surprising given the essential role of the *DHFR* gene in OCM.

Table 7.2 Impact of down-regulating the human DHFR gene in the HepG2 cells on the expression of OCM proteins

Accession Number	Protein Name	Description	HepG2 WT vs. HepG2 DHFR knockdown	
			ANOVA (p)	Fold
O75891	ALDH1L1	Aldehyde dehydrogenase family 1 member L1	0.00023981	↓ 2.304
Q3SY69	ALDH1L2	Aldehyde dehydrogenase family 1 member L2	-	-
Q93088	BHMT	Betaine--homocysteine S-methyltransferase	-	-
P00374	DHFR	Dihydrofolate reductase	0.01845623	↓ 1.626
Q86XF0	DHFR2	Dihydrofolate reductase 2	-	-
Q9UI17	DMGDH	Dimethylglycine dehydrogenase, mitochondrial	-	-
Q96DP5	MTFMT	Methionyl-tRNA formyltransferase, mitochondrial	-	-
Q99707	MS	Methionine synthase	-	-
P11586	MTHFD1	Methylenetetrahydrofolate dehydrogenase, cyclohydrolase and formyltetrahydrofolate synthetase 1	-	-
Q6UB35	MTHFD1L	Methylenetetrahydrofolate dehydrogenase, cyclohydrolase and formyltetrahydrofolate synthetase 1L	-	-
P13995	MTHFD2	Methylenetetrahydrofolate dehydrogenase, cyclohydrolase and formyltetrahydrofolate synthetase 2	-	-
Q9H903	MTHFD2L	Methylenetetrahydrofolate dehydrogenase, cyclohydrolase and formyltetrahydrofolate synthetase 2L	-	-
P42898	MTHFR	Methylenetetrahydrofolate reductase	-	-
Q9UL12	SARDH	Sarcosine dehydrogenase, mitochondrial	0.00274048	↓ 1.782
P34896	SHMT1	Serine hydroxymethyltransferase 1, cytosolic	0.00511007	↓ 1.613
P34897	SHMT2	Serine hydroxymethyltransferase 2, mitochondrial	-	-
P04818	TYMS	Thymidylate synthase	-	-

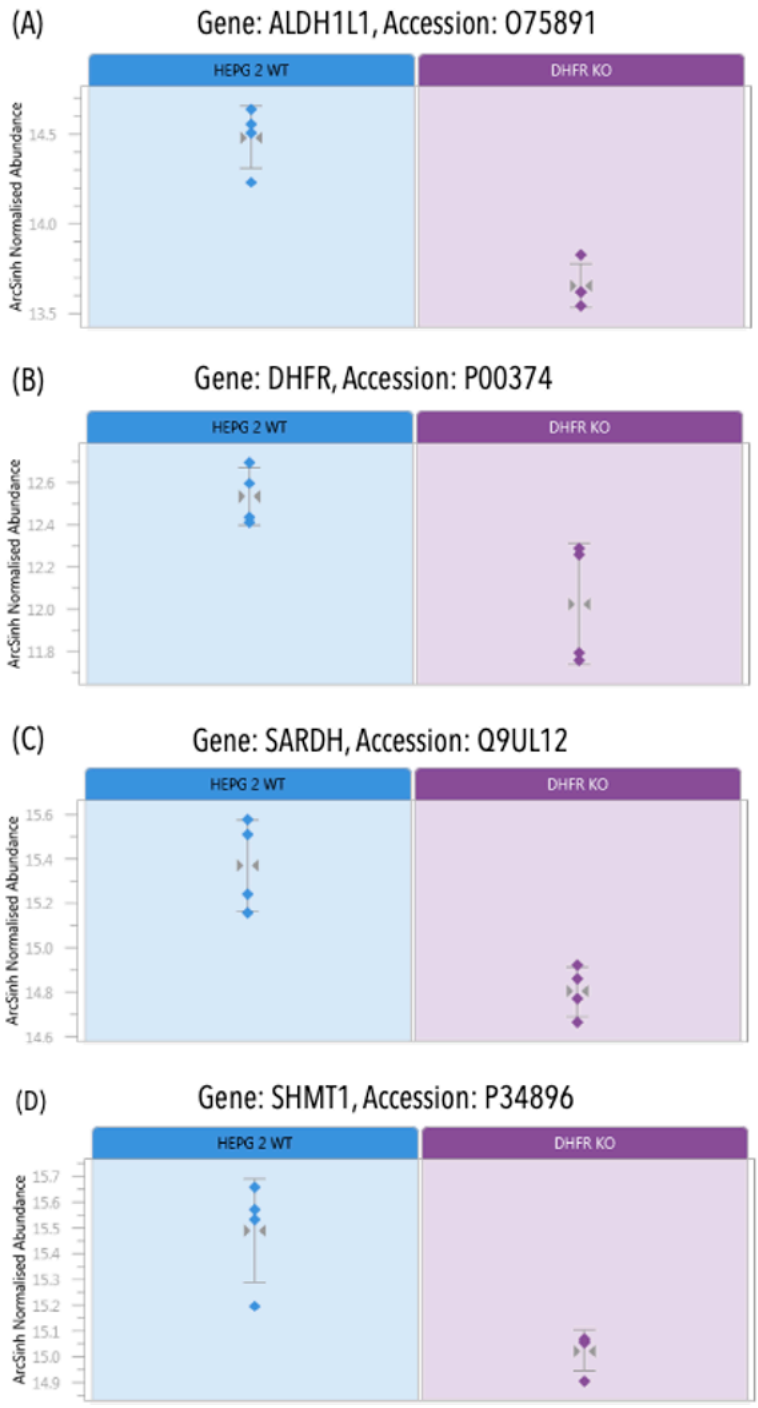


Figure 7.4 Compound abundance graphs showing the normalised abundance of (A) the ALDH1L1, (B) the DHFR, (C) the SARDH and (D) the SHMT1 proteins within the four replicates for the HepG2 wildtype (blue) and the HepG2 DHFR knockdown (purple) cell lines. *The expression of the ALDH1L1, DHFR, SARDH and SHMT1 proteins were all found to be significantly reduced in the HepG2 DHFR knockdown cell lines compared to the HepG2 wildtype. See Table 7.2 for the p-value and fold change for these genes. Created on Progenesis QI.*

The Qiagen IPA software and STRING database were also employed to identify the biological networks and pathways that were impacted in the HepG2 DHFR knockdown model. STRING calculated the PPI enrichment p-value to be < 1.0e–16 for the differential expressed proteins in the HepG2 DHFR knockdown model, the same value that was seen for the HepG2 DHFR2 knockout model. The network and protein function analysis carried out on Qiagen IPA revealed that the majority of these proteins were linked with the following networks: (i) cancer, cell death and survival, organismal injury and abnormalities (Appendix AE); (ii) RNA post-transcriptional modification, hereditary disorder, neurological disease (Appendix AF). Appendix AE and AF shows diagrams of these networks in relation to the DHFR knockdown cell model. The STRING database interrogated the differential expressed proteins in the HepG2 DHFR knockdown model further and identified the following pathways to be the most impacted: tRNA-guanine transglycosylation, protein import into mitochondrial intermembrane space, endonucleolytic cleavage to generate mature 5-end of small subunit ribosomal ribonucleic acid, L-ascorbic acid metabolism and SRP-dependent cotranslational protein targeting to the membrane.

7.3.1.3. Comparing the differential expressed proteins in the HepG2 DHFR2 knockout model to the HepG2 DHFR knockdown model

The differentially expressed proteins in the HepG2 DHFR2 knockout model and the HepG2 DHFR knockdown model were compared to identify proteins that were common to both. There were 243 proteins shared between these two datasets (Appendix AG). There were 27 proteins identified that had an opposing response to knocking out the *DHFR2* gene and knocking down the *DHFR* gene. Whereas, 216 proteins were found to have a consistent response to silencing either of the dihydrofolate reductase genes. Out of these 216 proteins, 119 were down-regulated and 97 were up-regulated. The 243 proteins that were common to the HepG2 DHFR2 knockout model and the HepG2 DHFR knockdown model were subsequently interrogated for the OCM proteins. The expression of the ALDH1L, DHFR and SHMT1 proteins were significantly reduced in both datasets (Table 7.1 and Table 7.2). Qiagen IPA and STRING were employed to determine if the 243 proteins that were common to both models were a significant dataset and if so which networks were impacted the most. STRING calculated the PPI enrichment p-value to be < 1.0e–16 for this dataset, meaning this set of proteins was not random but in fact, there was a strong interaction among them. The network and protein function analysis carried out on Qiagen IPA revealed that the majority of these proteins were linked with (i) gene expression, protein synthesis, cancer; (ii)

cellular movement, cell-to-cell signalling and interaction, inflammatory response; (iii) cancer, organismal injury and abnormalities, cell death and survival.

7.3.2. Overexpressing the *DHFR2* gene impacted the expression of several hundred other proteins in HepG2 cells

The *DHFR2* gene was overexpressed in HepG2 cells by Paola Drago using the technique described in Section 7.1.2.2. The overexpression of the DHFR2-tGFP RNA in the HepG2 cells was determined by RT-PCR and RT-qPCR. The expression of the recombinant DHFR2-tGFP protein was confirmed by Western Blot and LC-MS/MS analysis. The HepG2 DHFR2 overexpression cell line was subsequently assessed to determine the impact of up-regulating the *DHFR2* gene on overall protein expression. Mass spectrometry-based differential expression analysis compared protein expression between HepG2 cells transfected with the pCMV6-AC-tGFP (empty) vector (control) and HepG2 cells transfected with the pCMV6-AC-DHFR2-tGFP vector (Section 7.2.4). The differential expression analysis identified 273 proteins to be either significantly up-regulated or down-regulated in the HepG2 DHFR2 overexpression cells (Appendix AH). Of these 273 proteins, the expression of 144 were down-regulated and the remaining 129 were up-regulated. As expected, the recombinant DHFR2 protein was one of the 129 proteins that were found to be up-regulated, with a fold change of 15.05 (Table 7.3 and Figure 7.5). The data was subsequently interrogated for the remaining proteins within the OCM pathway (Table 7.3 and Figure 7.5). It revealed that the expression of two other OCM proteins was altered when the *DHFR2* gene was overexpressed in HepG2 cells. The expression of the DHFR (4.42 fold) and SARDH (2.48 fold) protein was found to be significantly up-regulated in the HepG2 DHFR2 overexpression cells (Table 7.3 and Figure 7.5)).

Table 7.3 Impact of overexpressing the human DHFR2 gene in the HepG2 cells on the expression of OCM proteins

Accession Number	Protein Name	Description	HepG2 WT + pCMV6-AC-tGFP vs. HepG2 WT + pCMV6-AC-DHFR2-tGFP	
			ANOVA (p)	Fold
O75891	ALDH1L1	Aldehyde dehydrogenase family 1 member L1	-	-
Q3SY69	ALDH1L2	Aldehyde dehydrogenase family 1 member L2	-	-
Q93088	BHMT	Betaine--homocysteine S-methyltransferase	-	-
P00374	DHFR	Dihydrofolate reductase	0.00002083	↑ 4.42
Q86XF0	DHFR2	Dihydrofolate reductase 2	0.00000015	↑ 15.05
Q9UI17	DMGDH	Dimethylglycine dehydrogenase, mitochondrial	-	-
Q96DP5	MTFMT	Methionyl-tRNA formyltransferase, mitochondrial	-	-
Q99707	MS	Methionine synthase	-	-
P11586	MTHFD1	Methylenetetrahydrofolate dehydrogenase, cyclohydrolase and formyltetrahydrofolate synthetase 1	-	-
Q6UB35	MTHFD1L	Methylenetetrahydrofolate dehydrogenase, cyclohydrolase and formyltetrahydrofolate synthetase 1L	-	-
P13995	MTHFD2	Methylenetetrahydrofolate dehydrogenase, cyclohydrolase and formyltetrahydrofolate synthetase 2	-	-
Q9H903	MTHFD2L	Methylenetetrahydrofolate dehydrogenase, cyclohydrolase and formyltetrahydrofolate synthetase 2L	-	-
P42898	MTHFR	Methylenetetrahydrofolate reductase	-	-
Q9UL12	SARDH	Sarcosine dehydrogenase, mitochondrial	0.013338	↑ 2.48
P34896	SHMT1	Serine hydroxymethyltransferase 1, cytosolic	-	-
P34897	SHMT2	Serine hydroxymethyltransferase 2, mitochondrial	-	-
P04818	TYMS	Thymidylate synthase	-	-

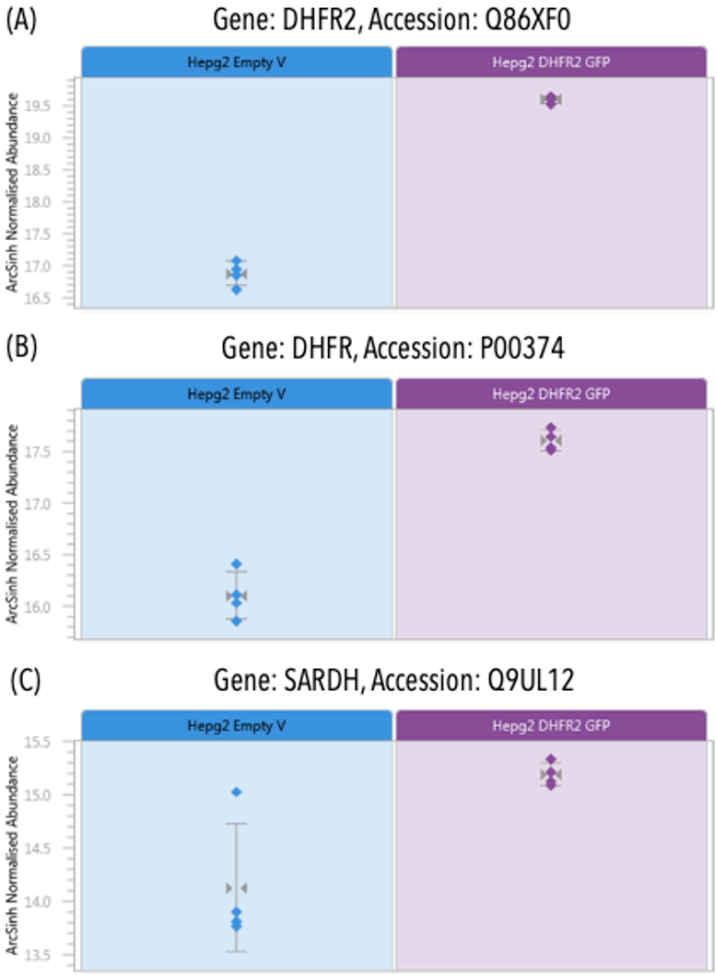


Figure 7.5 Compound abundance graphs showing the normalised abundance of (A) the recombinant DHFR2, (B) the DHFR and (C) the SARDH proteins within the four replicates for the HepG2 cells transfected with the pCMV6-AC-tGFP vector (blue) and HepG2 cells transfected with the pCMV6-AC-DHFR2-tGFP vector (purple). *The expression of the recombinant DHFR2, native DHFR and endogenous SARDH proteins were all found to be significantly up-regulated in the HepG2 DHFR2 overexpression cells. See Table 7.3 for the p-value and fold change for these genes. Created on Progenesis QI.*

The Qiagen IPA software and STRING database were also employed to identify the biological networks and pathways that were impacted in the HepG2 DHFR2 overexpression model. STRING calculated the PPI enrichment p-value to be < 1.0e–16 for the differential expressed proteins in the HepG2 DHFR2 overexpression model, meaning this set of proteins was not random but in fact, there was a strong interaction among them. The network and protein function analysis carried out on Qiagen IPA revealed that the majority of these proteins were linked with the following networks: (i) hereditary disorder, metabolic disease, neurological disease (Appendix AI); (ii) cancer, endocrine system disorders, organismal injury and abnormalities (Appendix AJ). Appendix AI and AJ shows diagrams of these networks in relation to the DHFR2 overexpression cell model. The STRING database interrogated the differential expressed proteins in the HepG2 DHFR2 overexpression model further and identified the following pathways to be the most impacted: viral translational termination-reinitiation, tetrahydrofolate biosynthetic process, IRES-dependent viral translational initiation, formation of the cytoplasmic translation initiation complex, cellular response to aldehyde, cytoplasmic translational initiation and glutamine metabolic process.

7.3.2.1. Comparing the differential expressed proteins in the HepG2 DHFR2 knockout model to the HepG2 DHFR2 overexpression model

The differentially expressed proteins in the HepG2 DHFR2 knockout model and the HepG2 DHFR2 overexpression model were compared to identify proteins that were common to both. There were 38 proteins shared between these two datasets (Appendix AK). There were 25 proteins found to have a consistent response to down-regulating and up-regulating the *DHFR2* gene (Table 7.4). This included the ABCC1, ACACA, AFDN, APOE, ARID3A, ASPH, ATXN2L, CARS1, CKB, DGAT1, H2BC12, H4C1, HK2, HNF4A, HSD17B10, MGAT2, NDUFA12, RIF1, RPL10A, RPL14, RPS10, RSL1D1, SUN2, TOP2A and YARS1 proteins. The majority of these proteins have been linked with either cellular metabolism (ABCC1, ACACA, APOE, CKB, DGAT1, HK2, HSD17B10, NDUFA12, RPL10A, RPL14 and RPS10) or ribosome biogenesis/translation (CARS1, RPL10A, RPL14, RPS10, RSL1D1 and YARS1). There were 13 proteins identified that had an opposing response to knocking out and overexpressing the *DHFR2* gene (Table 7.5). This included the ACADVL, ACOX1, ASNS, CEMIP2, ORO1B, CPS1, DHFR, EHHADH, GART, POLR2A, RCN1, SUPT6H and VWA8 proteins. Several of these proteins are involved in the carboxylic acid catabolic process (ACADVL, ACOX1, CEMIP2 and EHHADH) or the carboxylic acid biosynthetic process (ASNS, CPS1, DHFR and GART).

Table 7.4 Proteins found to have consistent effects in the HepG2 knockout and overexpression cell line

Accession Number	Protein Name	Description	HepG2 WT vs. HepG2 DHFR2 knockout		HepG2 WT + pCMV6-AC-tGFP vs. HepG2 WT + pCMV6-AC-DHFR2-tGFP	
			ANOVA (p)	Fold	ANOVA (p)	Fold
P33527	ABCC1	Multidrug resistance-associated protein 1	0.00064144	↑ 1.76	0.01911579	↑ 1.67
Q13085	ACACA	Acetyl-CoA carboxylase 1	0.00134604	↑ 1.57	0.00934038	↑ 1.66
P55196	AFDN	Afadin	0.00216526	↑ 1.58	0.02486876	↑ 1.51
P02649	APOE	Apolipoprotein E	0.00093689	↑ 1.57	0.01183213	↑ 2.65
Q99856	ARID3A	AT-rich interactive domain-containing protein 3A	0.00234298	↑ 1.50	0.02074146	↑ 3.11
Q12797	ASPH	Aspartyl/asparaginyl beta-hydroxylase	0.00920819	↑ 1.83	0.00190771	↑ 2.09
Q8WWM7	ATXN2L	Ataxin-2-like protein	0.00091048	↑ 1.56	0.02571023	↑ 1.86
P49589	CARS1	Cysteine--tRNA ligase, cytoplasmic	0.00671332	↑ 1.50	0.01982138	↑ 2.98
P12277	CKB	Creatine kinase B-type	0.00054993	↑ 1.77	0.00482668	↑ 5.83
O75907	DGAT1	Diacylglycerol O-acyltransferase 1	0.00017846	↓ 1.79	0.02259145	↓ 3.38
O60814	H2BC12	Histone H2B type 1-K	0.00008900	↓ 2.11	0.00743107	↓ 3.92
P62805	H4C1	Histone H4	0.00610037	↓ 1.83	0.00207687	↓ 1.88
P52789	HK2	Hexokinase-2	0.00000311	↑ 1.70	0.04111291	↑ 2.76
P41235	HNF4A	Hepatocyte nuclear factor 4-alpha	0.00008025	↓ 1.89	0.01894997	↓ 2.67
Q99714	HSD17B10	3-hydroxyacyl-CoA dehydrogenase type-2	0.00318818	↓ 1.51	0.02911252	↓ 969.52
Q10469	MGAT2	Alpha-1,6-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase	0.00085378	↓ 1.76	0.01537076	↓ 27.60
Q9UI09	NDUFA12	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 12	0.01726195	↓ 1.58	0.00378634	↓ 193.11
Q5UIP0	RIF1	Telomere-associated protein RIF1	0.00464843	↓ 1.60	0.01685801	↓ 1.56
P62906	RPL10A	60S ribosomal protein L10a	0.00608445	↑ 1.60	0.01078478	↑ 2.00
P50914	RPL14	60S ribosomal protein L14	0.04489312	↑ 1.86	0.00276071	↑ 1.60
P46783	RPS10	40S ribosomal protein S10	0.00056176	↓ 1.51	0.00063810	↓ 2.03
O76021	RSL1D1	Ribosomal L1 domain-containing protein 1	0.02762234	↑ 1.62	0.00089569	↑ 2.96
Q9UH99	SUN2	SUN domain-containing protein 2	0.00449768	↓ 1.58	0.03533514	↓ 1.74
P11388	TOP2A	DNA topoisomerase 2-alpha	0.00038509	↓ 2.07	0.03991862	↓ 1.85
P54577	YARS1	Tyrosine--tRNA ligase, cytoplasmic	0.00017517	↑ 1.59	0.02400512	↑ 4.23

Table 7.5 Proteins found to have opposing effects in the HepG2 knockout and overexpression cell line

Accession Number	Protein Name	Description	HepG2 WT vs. HepG2 DHFR2 knockout		HepG2 WT + pCMV6-AC-tGFP vs. HepG2 WT + pCMV6-AC-DHFR2-tGFP	
			ANOVA (p)	Fold	ANOVA (p)	Fold
P49748	ACADVL	Very long-chain specific acyl-CoA dehydrogenase, mitochondrial	0.00049556	↓ 1.57	0.00025789	↑ 1.79
Q15067	ACOX1	Peroxisomal acyl-coenzyme A oxidase 1	0.00339544	↓ 1.65	0.00998623	↑ 1.67
P08243	ASNS	Asparagine synthetase [glutamine-hydrolyzing]	0.00005449	↑ 4.33	0.01352700	↓ 36.44
Q9UHN6	CEMIP2	Cell surface hyaluronidase	0.00001322	↑ 2.20	0.02315722	↓ 15.84
Q9BR76	CORO1B	Coronin-1B	0.00058599	↑ 1.59	0.00442998	↓ 12.98
P31327	CPS1	Carbamoyl-phosphate synthase [ammonia], mitochondrial	0.00265963	↓ 1.64	0.01453463	↑ 2.37
P00374	DHFR	Dihydrofolate reductase	0.00014019	↓ 2.40	0.00002083	↑ 4.42
Q08426	EHHADH	Peroxisomal bifunctional enzyme	0.00090430	↓ 1.55	0.02255708	↑ 1.89
P22102	GART	Trifunctional purine biosynthetic protein adenosine-3	0.00006607	↑ 1.53	0.02243793	↓ 1.50
P24928	POLR2A	DNA-directed RNA polymerase II subunit RPB1	0.00280944	↓ 1.55	0.01141808	↑ 2.40
Q15293	RCN1	Coatomer subunit delta	0.00005254	↑ 1.64	0.01515474	↓ 1.93
Q7KZ85	SUPT6H	Transcription elongation factor SPT6	0.00016004	↓ 1.60	0.02635284	↑ 4.17
A3KMH1	VWA8	von Willebrand factor A domain-containing protein 8	0.00118105	↓ 1.63	0.01654083	↑ 3.42

The dataset that contained the 38 proteins (consistent or opposing response to down-regulating and up-regulating the *DHFR2* gene) was interrogated for the proteins within the OCM pathway. The DHFR protein was the only OCM protein identified in this analysis. When the *DHFR2* gene was knocked out the expression of the DHFR protein was reduced by 2.404 fold; whereas when the *DHFR2* gene was overexpressed the expression of the DHFR protein was increased by 4.42 fold. Therefore this consistent result suggested that the expression of the *DHFR2* gene impacted the expression of the *DHFR* gene.

Network and protein function analysis was not performed on the 38 proteins that were common to both datasets, 25 proteins found to have a consistent response to down-regulating and up-regulating the *DHFR2* gene or the 13 proteins identified that had an opposing response to knocking out and overexpressing the *DHFR2* gene. This is due to the number of proteins in these datasets being far too low to get a statistically significant result.

7.4. Discussion

Modulating the expression of a specific gene is a powerful mechanism for determining its biological role and physiological significance. It can reveal the processes, networks and pathways that the target gene is involved in. It can even identify specific genes that are linked to the gene of interest (Prelich, 2012). This technique was employed to determine if the human *DHFR2* gene had an additional cellular role as a regulatory RNA. The HepG2 (liver hepatocellular carcinoma) cell line was chosen to examine the effect of altering the expression of the *DHFR2* gene. There were two reasons why this cell line was chosen. The first being the DHFR2 protein had not been detected in this cell line (Chapter 5, Section 5.3.1). The second was that the HepG2 cell line is extremely robust making it straightforward to manipulate (Harjumäki et al., 2019).

There are several ways in which the expression of a gene can be modulated, one of which is reducing its expression (Section 7.1.2.1) (Prelich, 2012). A gene knockout is the method of choice as it completely and permanently silences the gene at the DNA level (Boettcher & McManus, 2015). This methodology was employed in HepG2 cells to eliminate the expression of the *DHFR2* gene. The knockout of the *DHFR2* gene in the HepG2 cells was assessed by PCR and RT-PCR. This analysis confirmed a homozygous knockout of the *DHFR2* gene in the HepG2 cell line (Drago, unpublished). Protein-based methods were not utilised to confirm the successful knockout of the *DHFR2* gene as there was no evidence of the protein in wildtype HepG2 cells (Chapter 5, Section 5.3.1). The gene knockout technology was also employed to generate a HepG2 DHFR knockout cell line to use as a control in the subsequent analysis. This was assessed by PCR, RT-PCR, Western Blot, LC-MS/MS and enzymatic activity analysis. The Western Blot and LC-MS/MS analysis showed that the *DHFR* gene was not completely silenced however it did reveal that the expression of the protein was significantly reduced. Furthermore, enzymatic activity analysis showed that the dihydrofolate reductase activity was reduced by 88.74% in this model when compared to the HepG2 wildtype. This result meant that the *DHFR* gene was not completely silenced but that it was significantly down-regulated (Drago, unpublished).

Differential expression analysis is a technique used to identify statistically significant differences in expression levels between two or more experimental groups (McDermaid et al., 2019). This type of analysis was employed to assess the impact of knocking out the *DHFR2* and knocking down the *DHFR* gene on the proteome of HepG2 cells. Differential expression analysis was performed using the mass spectrometer comparing the HepG2 DHFR2 knockout and HepG2 DHFR knockdown cell lines to the HepG2 wildtype cell line. The down-regulation of both dihydrofolate reductase genes had a huge impact on the overall expression of the proteome in HepG2 cells. Over 700 proteins

were found to be differentially expressed when the *DHFR* gene was down-regulated (Section 7.3.1.2). This result was not surprising given its essential role in supplying one-carbon units to several metabolic reactions within the OCM pathway. Furthermore, the DHFR enzyme is essential for sustaining DNA biosynthesis, amino acid synthesis, cell proliferation and cellular methylation (Schnell et al., 2004). Almost 500 proteins were found to be differentially expressed when the *DHFR2* gene was silenced (Section 7.3.1.1). This large dataset suggested that the *DHFR2* gene, like the *DHFR* gene, also has an important cellular role.

When knocking down the *DHFR* gene there was an expectation that it would impact proteins within the OCM that work alongside the DHFR enzyme. For example, proteins that are involved in the de novo thymidylate synthesis pathway, such as the SHMT and TYMS enzymes (Anderson et al., 2012). As predicted, the expression of the SHMT1 enzyme was found to be reduced in the HepG2 DHFR knockdown (Table 7.2 and Figure 7.4). This result suggested that the de novo thymidylate synthesis pathway was impaired when the *DHFR* gene was down-regulated. Given that the de novo thymidylate synthesis pathway is one of several pathways within the interconnected OCM network; it was not surprising that the expression of other OCM proteins were also altered. The expression of ALDH1L and SARDH enzymes was reduced in the DHFR knockdown model (Table 7.2 and Figure 7.4). The ALDH1L enzyme converts 10-formyl-THF (10-CHO-THF), $NADP^+$ and water to THF, NADPH and carbon dioxide in the cytoplasm. The ALDH1L enzyme catalyses this reaction to eliminate excess 10-formyl-THF in a cell (Horita & Krupenko, 2017; Krupenko, 2009). If the de novo thymidylate synthesis pathway is impaired it will result in a decrease in the number of activated THF molecules, including 10-formyl-THF (Tjong et al., 2022; Zarou et al., 2021). The reduction of 10-formyl-THF in the cell could lead to lower demand for the ALDH1L enzyme, causing a decrease in its expression. This could be the reason why the expression of the ALDH1L enzyme decreased when the *DHFR* gene was knocked down. The SARDH enzyme converts THF and sarcosine to 5,10-methylene-THF (5,10-CH_2-THF) and glycine in the mitochondria. Glycine can contribute one-carbon units to the OCM if there is insufficient formate in the cell. Apart from being part of the same complex pathway, there is no known link between the SARDH and DHFR enzyme (Bar-joseph et al., 2012; Mudd et al., 1980; Porter et al., 1985; Simile et al., 2018). Therefore, there was no explanation for why the expression of the SARDH enzyme was impacted when the *DHFR* gene was down-regulated.

As the function of the *DHFR2* gene had not been fully identified it was unclear which proteins within the OCM pathway would be affected by knocking it out. The expression of the three main enzymes (DHFR, SHMT1 and TYMS) involved in the de novo thymidylate syntheses pathway were found to be reduced when the *DHFR2* gene was silenced (Table 7.1 and Figure 7.3). Furthermore, the network and protein function analysis carried out on STRING identified the de novo thymidylate

synthesis pathway to be one of the most affected pathways in the DHFR2 knockout model (Section 7.3.1.1). This result suggested that the de novo thymidylate synthesis pathway was impaired when the *DHFR2* gene was knocked out. Therefore, the *DHFR2* gene may regulate the de novo thymidylate synthesis pathway through its interaction with the *DHFR, SHMT1* and/or *TYMS* genes. Considering that the de novo thymidylate synthesis pathway is one of a number of pathways within the interconnected OCM network, it was not surprising that the expression of other OCM proteins were altered. The expression of ALDH1L protein was drastically reduced in the DHFR2 knockout model (Table 7.1 and Figure 7.3). The reason why the expression of this enzyme was impacted in this model could be the reason stated in the previous paragraph.

Comparing the DHFR knockdown dataset to the DHFR2 knockout dataset revealed a clear correlation between them. First off, there were 243 proteins shared between the two cell models (Appendix AG), which is a significant sized dataset. Furthermore, the STRING database calculated an extremely low PPI enrichment p-value for this dataset, meaning these set of proteins were not random but in fact, there was a strong interaction among them (Section 7.3.1.3). Moreover, when the data was interrogated for the OCM proteins the expression of similar proteins appeared to be altered in both models, the ALDH1L, DHFR and SHMT1 proteins (Section 7.3.1.3). This correlation between the DHFR knockdown and DHFR2 knockout cell lines suggested that the cellular function of the two dihydrofolate reductase genes are interlinked. Therefore, the *DHFR2* gene may regulate the de novo thymidylate synthesis pathway through interactions with the *DHFR* gene.

A second method that can be utilised to determine the biological function of a gene is by increasing the expression of that gene (Section 7.1.2.2) (Prelich, 2012). Overexpression technology was employed to also investigate the cellular role of the *DHFR2* gene. The *DHFR2* gene was up-regulated by transfecting the pCMV6-AC-DHFR2-tGFP expression vector into HepG2 cells and allowing the cells to grow for two days (Section 7.2.1.2). The overexpression of the recombinant DHFR2-tGFP RNA in the HepG2 cells was confirmed by RT-PCR and RT-qPCR. The expression of the recombinant DHFR2-tGFP protein was determined by Western Blot and LC-MS/MS analysis (Drago, unpublished). A pCMV6-AC-tGFP (empty) expression vector was also prepared as a control sample for this experiment (Section 7.2.1.2). The expression of the recombinant tGFP protein in this control sample was confirmed by Western Blot analysis (Drago, unpublished). Differential expression analysis was employed to assess the impact of overexpressing the *DHFR2* gene on the proteome of HepG2 cells. Differential expression analysis was performed using the mass spectrometer comparing the HepG2 DHFR2 overexpression cell line to the HepG2 control sample (Section 7.2.4). Over 200 proteins were found to be differentially expressed when the *DHFR2* gene was up-regulated (Section 7.3.2). This substantial dataset suggested that the *DHFR2* gene has an important cellular role. When the data

was interrogated for the OCM proteins the expression of several OCM proteins appeared to be altered. As expected, the recombinant DHFR2 protein was found to be up-regulated in the HepG2 DHFR2 overexpression model (Table 7.3 and Figure 7.5). In addition, the expression of two other OCM proteins was increased, the DHFR and SARDH protein (Table 7.3 and Figure 7.5). Apart from being part of the same complex pathway, there is no known link between the SARDH enzyme and the *DHFR2* gene. Therefore, there was no explanation for why the expression of the SARDH enzyme was impacted when the *DHFR2* gene was overexpressed. The *DHFR* gene however is one of interest as it is part of the same gene family as the *DHFR2* gene. Moreover, the DHFR protein was found to be the second most impacted protein in the HepG2 overexpression model, after the recombinant DHFR2 protein. This result proposed that the expression of the *DHFR2* gene impacted the expression of the DHFR enzyme. Thus, this significant finding ultimately suggested that the *DHFR2* gene has a role in regulating the expression of the *DHFR* gene.

There were a small number of proteins found within both the DHFR2 knockout dataset and the DHFR2 overexpression dataset. When this dataset was interrogated for the OCM proteins the DHFR protein was the only OCM protein identified. When the *DHFR2* gene was knocked out the expression of the DHFR protein was reduced; whereas when the *DHFR2* gene was overexpressed the expression of the DHFR protein was increased. Therefore this consistent result clearly shows that the expression of the *DHFR2* gene impacted the expression of the *DHFR* gene. Ultimately showing that the *DHFR2* gene regulates the expression of the *DHFR* gene.

Taking the results from this chapter into account a hypothesised regulatory model for the *DHFR2* gene was generated; (i) the *DHFR2* gene directly regulates the expression of the *DHFR* gene, (ii) this regulation impacts the efficiency of the de novo thymidylate synthesis pathway and (iii) the effect on the de novo thymidylate synthesis pathway ultimately alters the expression of the remaining proteins within the pathway (SHMT1 and TYMS). This model could explain why the expression of the three enzymes involved in the de novo thymidylate synthesis pathway (DHFR, SHMT1 and TYMS) were impacted in the DHFR2 knockout cell line; whereas only the expression of the DHFR protein was affected in the DHFR2 overexpression cell line. The DHFR2 knockout model had several weeks to impact the expression of other proteins within the proteome of HepG2 cells. Thus, the regulation within the DHFR2 knockout model had time to impact the DHFR protein and the other de novo thymidylate synthesis proteins. Whereas, the DHFR2 overexpression model only had two days to impact the expression of other proteins within the proteome of HepG2 cells. Therefore, the regulation within the DHFR2 overexpression model only had time to affect the expression of the DHFR protein.

To conclude, the experiments in this chapter clearly showed that the *DHFR2* gene has an additional role in regulating the expression of the *DHFR* gene. This statement was formed based on the impact observed on the DHFR protein when the *DHFR2* gene was silenced and overexpressed (Table 7.1, Table 7.3, Figure 7.3 and Figure 7.5). This conclusion is also supported by the fact that the DHFR2 knockout and overexpression models had opposing effects on the expression of the DHFR protein. As the endogenous DHFR2 protein was not detected in the HepG2 wildtype cell line (Chapter 5, Section 5.3.1) or the HepG2 DHFR2 knockout cell line (Chapter 4, Section 4.3.1) this result suggests that it is the DHFR2 RNA that carries out this regulatory role. The most plausible option is that the DHFR2 RNA transcripts regulate the expression of the *DHFR* gene through interactions with its protein. This prediction is based on an electrophoretic mobility shift experiment that showed that the DHFR2 RNA has the ability to bind to the DHFR protein (McEntee et al., 2011). However, further experiments need to be performed to conclude how the DHFR2 RNA regulates the expression of the *DHFR* gene. A co-localisation experiment on the DHFR2 RNA and DHFR RNA/protein would determine which molecule the DHFR2 RNA binds to and regulates. There are multiple fluorescent-based techniques that can be employed to recognise specific nucleic acids and/or proteins including fluorescence in situ hybridization (FISH) or hybridisation chain reaction (HCR) (Evanko, 2004; Meyer et al., 2017).

Chapter 8. Conclusion

8.1. General

The first and only experimental investigations into the human dihydrofolate reductase 2 (*DHFR*2) retrogene were published in 2011 (Anderson et al., 2011; McEntee et al., 2011), thus limiting the amount of data and knowledge in relation to the function of this gene. However, *DHFR2* is part of the same gene family as the well-known, and extensively characterised, dihydrofolate reductase (*DHFR*) gene. The *DHFR* gene encodes an enzyme that carries out the dihydrofolate reductase activity in the cytoplasm, nucleus and mitochondria of a cell (Anderson et al., 2011, 2012; Appling, 1991; Hughes, unpublished; Woeller et al., 2007). Dihydrofolate reductase activity is essential for one-carbon metabolism (OCM), which is responsible for purine/thymidylate biosynthesis and homocysteine re-methylation. These products are required to maintain normal cell function (Fox & Stover, 2008). For this reason, the *DHFR* gene has become an extremely valuable drug target for the treatment of multiple types of cancer and some inflammatory diseases (Bailey, 1995; Cronstein, 1996). Given the importance of the *DHFR* gene to human health, it was vital that the role of the second functional dihydrofolate reductase gene was elucidated.

The overall DHFR2 project aimed to fully characterise the human *DHFR2* gene, with particular attention on identifying its overall cellular function and its role in the OCM pathway. The plan was to determine the function of the *DHFR2* gene by examining its expression and behaviours at a DNA, RNA and protein level. This PhD thesis mainly focused on characterising specific DHFR2 RNA isoforms and the predicted DHFR2 protein. Firstly, the translatability of the DHFR2 transcripts was assessed by (i) performing a sucrose cushion ultracentrifugation coupled with RT-qPCR and (ii) attempting to generate the recombinant DHFR2 protein in a humanised DHFR2 mouse model. Secondly, the expression of the endogenous DHFR2 protein was investigated in a panel of human cell lines/tissues and under two cellular conditions, using data-dependent acquisition (DDA) and targeted-dependent acquisition (TDA) liquid chromatography-tandem mass spectrometry (LC-MS/MS). Thirdly, translational regulation by paraspeckle nuclear retention was considered for the DHFR2 transcripts. Finally, the effect of modulating the *DHFR2* gene on the proteome was assessed by carrying out mass spectrometry-based differential analysis on the DHFR2 knockout and overexpression models. These experiments provided evidence that the *DHFR2* gene encodes a protein that is specifically synthesised in the heart during early development; possibly to supply dihydrofolate reductase activity to a rapidly growing embryos and foetuses. Furthermore, the findings showed that the *DHFR2* gene has an additional role at the RNA level regulating the expression of its parent homologue, the *DHFR* gene, and in turn the other de novo thymidylate synthesis genes.

8.2. Chapter observations

Prior to this project, the *DHFR2* gene was found to be transcribed into several RNA isoforms (Ensembl, 2019; McEntee et al., 2011) in multiple human cell lines and tissues (Anderson et al., 2011; McEntee et al., 2011; Protein Atlas, 2021); however, there was no evidence that the DHFR2 RNA was translated into the predicted endogenous protein. Therefore, in Chapter 3 the translational capabilities of the DHFR2 transcripts were examined. The best approach for investigating the translational status of an RNA molecule is to measure its association with the ribosome (Dermit et al., 2017); thus, two different methodologies were employed to assess the relationship between the DHFR2 transcripts and the ribosome. The first approach exploited ribosome profiling (Ribo-seq) data that had been generated for other studies and published on the online directory known as ribosome protected mRNA fragments database version 2.0 beta (RPFdb v2.0 beta)(Wang et al., 2019; Xie et al., 2016). Interrogating the Ribo-seq of several human cell lines and tissues on this database revealed that the DHFR2 transcripts associate with the ribosome; this result suggested that the DHFR2 RNA has the potential to be translated into a protein (Chapter 3, Section 3.3.1). Moreover, this experiment also showed that the predicted DHFR2 open reading frame (ORF) was the most likely region within this gene to potentially encode a functional protein (Chapter 3, Section 3.3.1). These findings implied that the DHFR2 RNA has the capabilities to be translated into a dihydrofolate reductase enzyme. Even though a significant result was obtained from the RPFdb v2.0 beta database, this directory considered all the DHFR2 RNA isoforms as one; thus the detail for the individual transcripts was lost. For this reason, a sucrose cushion ultracentrifugation RT-qPCR technique was employed to investigate the translational status of individual DHFR2 transcripts. The DHFR2_201 and DHFR2_202 RNA isoforms were chosen for investigation as they were found by Ensembl to be the most plausible mRNA isoforms to generate a functional DHFR2 protein. The RT-qPCR results on the sucrose cushion ultracentrifugation fractions revealed that the DHFR2_201 transcript binds to the ribosome; suggesting it has the capabilities to be synthesised into a protein (Chapter 3, Section 3.3.3.2). Therefore, proteomic analysis needed to be performed to determine if the predicted DHFR2 protein is expressed and functional. The RT-qPCR results on the sucrose cushion ultracentrifugation fractions also revealed that the DHFR2_202 RNA isoform does not associate with the ribosome; suggesting this transcript does not encode a protein (Chapter 3, Section 3.3.3.2). Taking into account this result and the fact that the DHFR2_202 transcript is highly expressed across multiple cell lines, this transcript may have a meaningful biological role at the RNA level. Therefore, further analysis should be performed to determine if this is the case.

The translatability of the human *DHFR2* gene was also the focus of Chapter 4. The aim of this chapter was to determine could the DHFR2 ORF generate the predicted human DHFR2 protein in mammalian cells. To answer this research question, two different approaches were employed to try to synthesise the recombinant DHFR2 protein from the DHFR2 ORF. A traditional mammalian cell-based expression system that was transfected with a DHFR2 expression vector was the first approach undertaken. Western Blotting and LC-MS/MS analysis on this recombinant clone revealed that the recombinant DHFR2 protein had been successfully synthesised (Chapter 4, Section 4.3.1). The generation of the recombinant human DHFR2 protein in a eukaryotic expression system was a huge breakthrough for this project. It meant that the DHFR2 protein could be synthesised in mammalian cells when there were strong artificial signals initiating transcription and translation. Furthermore, the recombinant DHFR2 protein could help characterise the endogenous version of the protein and could also be used as a control for subsequent analysis. Even though this result was significant it did not reveal much information about the translatability of the DHFR2's ORF; this is due to the cellular environment not being fully representative of DHFR2's natural environment. Thus, a second approach was employed to determine could the recombinant DHFR2 protein be synthesised from DHFR2's ORF in a cellular environment that was similar to its native one. A humanised DHFR2 mouse model that was initially generated for complementation was exploited to answer this research question. The coding sequence belonging to the *DHFR2* gene had been inserted into the mouse genome in place of the mouse Dhfr ORF. LC-MS/MS analysis on the humanised DHFR2 mouse model revealed that the recombinant DHFR2 protein had not been synthesised (Chapter 4, Section 4.3.2). This result implied that the transcriptional and translational signals belonging to the mouse *Dhfr* gene were not sufficient to synthesise the recombinant DHFR2 protein. As the cellular environment for the humanised DHFR2 mouse model is similar to DHFR2's native environment, this result also suggested that the synthesis of the endogenous human DHFR2 protein may also be repressed.

Chapter 4 hypothesised that the expression of the native DHFR2 protein may be translationally inhibited. This information suggested that the *DHFR2* gene may not encode the predicted dihydrofolate reductase protein. However, Chapter 3 confirmed that at least one of the DHFR2 transcripts binds to the ribosome and therefore has the capabilities to be translated into a protein. Therefore, the next logical step for this project was to perform high definition, focused analysis to determine if the predicted endogenous DHFR2 protein was expressed. Difficulties associated with detecting the endogenous DHFR2 protein prior to this project suggested that the DHFR2 protein may only be cell/tissue specific, present under particular cellular conditions or a low abundance protein (Hughes, unpublished). Therefore, the overall aim of Chapter 5 was to interrogate a range

of cell lines and tissues for the presence of the native human DHFR2 protein. A genetic association study of an Irish cohort found an association between 3 SNPs within the *DHFR2* gene and the risk of developing NTDs. Even though this result did not replicate in a British cohort (Pangilinan et al., 2021), these findings still suggested that the DHFR2 protein may only be expressed and functional during human development. The regulations in Ireland do not allow research on human embryonic/foetal tissue. Therefore, adult human cell lines and tissues that are similar to embryonic/foetal tissue were employed to search for the native DHFR2 protein. This included HepG2 cells, undifferentiated iPS cells, HuH7 cells, 101B cells, IMR32 cells, NEP cells, placental tissue, umbilical cord tissue, membrane tissue, ovaries tissue and testis tissue. DDA LC-MS/MS and a TDA LC-MS/MS approach known as parallel reaction monitoring (PRM) found no evidence that the *DHFR2* gene encodes a functional protein in these adult cell lines and tissues (Chapter 5, Section 5.3.1). This suggested that the predicted DHFR2 protein does not supply dihydrofolate reductase activity to these adult cell lines and tissues. The next experiment examined if the DHFR2 protein was only expressed and functional under particular cellular conditions. More specifically, if the DHFR2 protein was synthesised to provide reductase activity when the DHFR protein expression was eliminated or significantly reduced. DDA LC-MS/MS and PRM LC-MS/MS analysis on the HepG2 DHFR knockdown model found no proof of the endogenous DHFR2 protein (Chapter 5, Section 5.3.2); indicating the DHFR2 protein is not synthesised to carry out the reductase activity in a cell when the expression of the DHFR enzyme is diminished or lost. Given that there was no evidence of the DHFR2 protein from the previous experiments there was a possibility that the DHFR2 protein was subject to degradation via the ubiquitin-proteasome pathway. DDA LC-MS/MS and PRM LC-MS/MS analysis on cells treated with the MG132 proteasome inhibitor found no evidence of the native DHFR2 protein (Chapter 5, Section 5.3.3). This result suggested that the DHFR2 protein is not susceptible to degradation by the ubiquitin-proteasome pathway. Given that the endogenous DHFR2 protein had not been detected in any of the cell lines and tissues or in any of the cellular conditions created, the role of this gene remained unanswered. The focus returned to the original hypothesis that the *DHFR2* gene may play an important role during embryogenesis. It was concluded that the adult cell lines and tissues that were similar to the human developmental tissue were not fully representative of the cellular environment of human embryos and foetuses. Therefore, to determine if this hypothesis was in fact true, human embryonic/foetal tissue needed to be examined for the native DHFR2 protein. As the regulations in Ireland do not allow research to be carried out on human embryonic/foetal tissue, this experiment was performed in the United Kingdom where it is permitted. A TDA LC-MS/MS approach known as multiple reaction monitoring (MRM) was performed on kidney, heart, lung, liver and brain tissue from three different stages of

human development: the Carnegie stage 21-22, post-conceptual week 9-10 and post-conceptual week 15-17. The MRM LC-MS/MS analysis detected the endogenous DHFR2 protein in heart tissue from all three stages (Chapter 5, Section 5.3.4). This significant finding suggested that the DHFR2 protein was synthesised in the heart during early development; possibly to supply dihydrofolate reductase activity to rapidly growing embryos and foetuses.

Out of all the cell lines and tissues examined for the endogenous DHFR2 protein in Chapter 5, the protein was only detected in human embryonic and foetal heart tissue. This result suggested that the cellular role of the DHFR2 protein is tissue specific. It also implied that the translation of the DHFR2 protein is tightly regulated or completely blocked in certain cells and tissues. Furthermore, there is ample evidence that different DHFR2 RNA isoforms are expressed in all cell lines and tissues examined to date (McEntee et al., 2011, Chapter 3). Both of these results suggested that the regulation of the DHFR2 protein most likely occurs before the DHFR2 RNA is translated into a protein. Therefore, in Chapter 6 different mechanisms of translational regulation were considered for the DHFR2 transcripts. One mechanism of translational regulation that became of interest for the DHFR2 transcripts was paraspeckle nuclear retention. Paraspeckles bind to RNA molecules and prevent them from being exported into the cytoplasm; ultimately inhibiting the translation of that RNA molecule (Bond & Fox, 2009). For RNA molecules to be retained by the paraspeckle it first must contain IRAlu elements in its 3'UTR (Chen & Carmichael, 2012). Chen et al., (2008b) and the REDIportal V2.0 database confirmed that the DHFR2_201 and DHFR2_202 transcripts contained IRAlu elements in its 3'UTR, known individually as AluJb_Alu and FLAM_A_Alu (Chapter 6, Section 6.3.1). The IRAlu elements within the 3'UTR of an RNA molecule must also be subject to A to I editing to be retained by the paraspeckle (Bass & Weintraub, 1988; Chen & Carmichael, 2009; Wagner et al., 1989). Two different methods were employed to determine if the IRAlu elements in the 3'UTR of the DHFR2_201 and DHFR2_202 transcripts were prone to A to I RNA editing. The first approach exploited an RNA editing database known as REDIportal V2.0 (Picardi et al., 2017). This directory identified several A to I edits within DHFR2's AluJb_Alu and FLAM_A_Alu IRAlu elements (Chapter 6, Section 6.3.1). The second approach employed a more conventional experimental method; it consisted of comparing complementary DNA (cDNA) sequences of the AluJb_Alu element to the corresponding genomic reference sequence using RT-PCR, cloning and Sanger Sequencing (Sakurai et al., 2014). The RT-PCR analysis identified several A to I edits within DHFR2's AluJb_Alu element (Chapter 6, Section 6.3.2.2). Both approaches revealed that some of the DHFR2_201 and/or DHFR2_202 transcripts were prone to A to I RNA editing. These edited transcripts may have bound to the $P54^{NRB}$/NONO paraspeckle protein which resulted in them being retained in the nucleus (Zhang & Carmichael, 2001). The retention of these transcripts in the nucleus would decrease the

amount of readily available DHFR2_201 and DHFR2_202 transcripts in the cytoplasm, ultimately reducing the chances of the DHFR2 protein being translated (Pisani & Baron, 2019; Zhang & Carmichael, 2001; Zinshteyn & Nishikura, 2009). As a result of these findings, it was concluded that paraspeckle nuclear retention could be one of the mechanisms that is inhibiting the translation of the DHFR2 transcripts in certain cell and tissue types.

Chapter 5 revealed that the *DHFR2* gene may play an important cellular role in embryonic and foetal heart tissue; however it is yet to be proven that the protein is functional. As working with human embryonic tissue is not possible in Ireland this part of the project could not be explored further at DCU. Therefore, the focus of this project turned to determining did the *DHFR2* gene have any additional cellular roles. A previous observation became of huge interest, the DHFR2 RNA is highly abundant in multiple cell lines and tissues despite the DHFR2 protein being undetectable in the same samples (McEntee et al., 2011; Hughes, unpublished; Chapter 5). This observation suggested that the *DHFR2* gene may have a biological function at the RNA level. Thus, the aim of Chapter 7 was to determine did the *DHFR2* gene have a second cellular role as a regulatory RNA. Altering the expression of a specific gene is one of the most powerful mechanisms for determining its biological role and physiological significance. It has the ability to reveal the processes, networks and pathways that the target gene is involved in; furthermore, it can even identify specific genes that are linked to the gene of interest (Prelich, 2012). Therefore, this technique was employed to determine if the human *DHFR2* gene had an additional cellular role at the RNA level. The *DHFR2* gene was knocked out and overexpressed in a cell line where the DHFR2 protein had not been identified i.e., the HepG2 cell line (Chapter 5, Section 5.3.1). Mass spectrometry-based differential analysis on the two models revealed that the expression of hundreds of proteins were altered when the *DHFR2* gene was knocked out and up-regulated (Chapter 7, Section 7.3). The most significant impact recorded was with the DHFR enzyme; the expression of the DHFR protein was reduced when the *DHFR2* gene was knocked out, whereas the expression of the DHFR protein was a lot higher when the *DHFR2* gene was up-regulated (Chapter 7, Section 7.3). This result showed that the expression of the *DHFR2* gene impacted the levels of the DHFR enzyme. Thus, the DHFR2 RNA may directly regulate the expression of the *DHFR* gene via interactions with its RNA and/or protein. The most plausible option is that the DHFR2 RNA transcripts regulate the expression of the *DHFR* gene through interactions with its protein. This prediction was based on an electrophoretic mobility shift experiment that showed that the DHFR2 RNA has the ability to bind to the DHFR protein (McEntee et al., 2011). Interactions between the DHFR2 RNA and the DHFR protein may disrupt DHFR's established auto-regulation system (Figure 8.1). DHFR's auto-regulation mechanism is where the DHFR protein binds to DHFR mRNA to suppress its translation (Ercikan et al., 1993; Tai et al., 2004).

Therefore, if the DHFR2 RNA binds to the DHFR protein it could prevent the DHFR mRNA from interacting with the protein; subsequently allowing the DHFR RNA to be translated into a protein (Figure 8.1). The effect of knocking out *DHFR2* gene on the expression of the DHFR protein supports this hypothesis. Alternatively, if the DHFR2 RNA levels are diminished in the cell the DHFR protein would be free to bind to the DHFR mRNA ultimately inhibiting its translation. The impact of overexpressing the *DHFR2* gene on the expression of the DHFR protein also supports this statement.

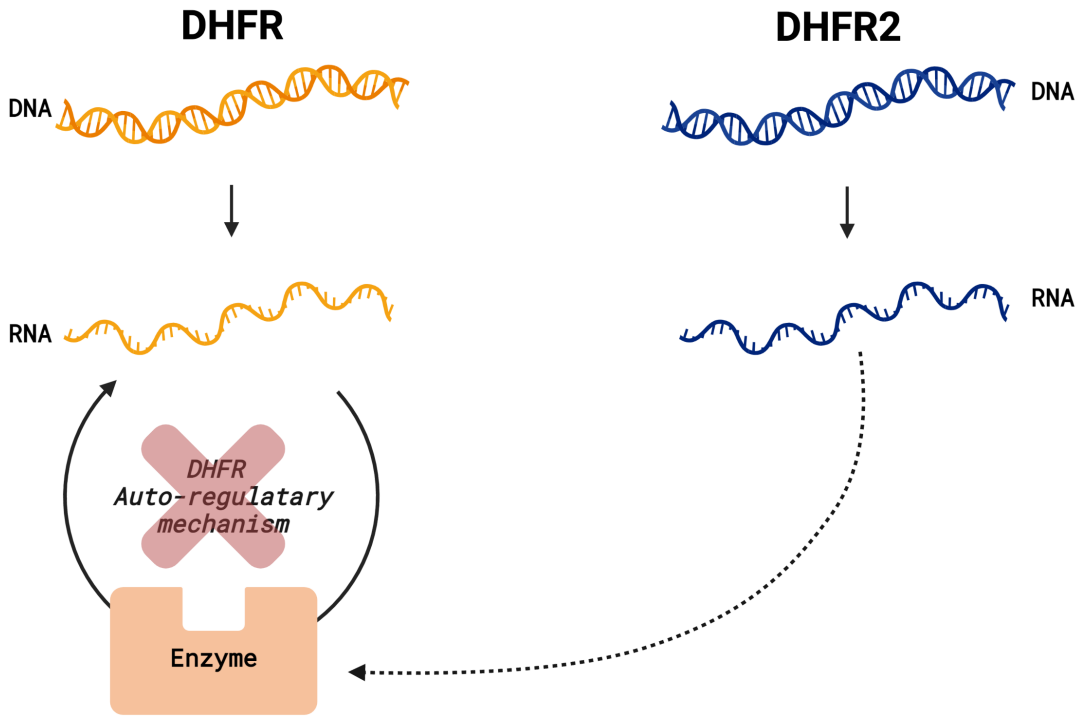


Figure 8.1 Proposed DHFR and DHFR2 regulatory mechanism. *The DHFR2 RNA directly regulates the expression of the DHFR gene via interactions with its protein. The interaction between the DHFR2 RNA and the DHFR protein may disrupt DHFR's auto-regulation mechanism. If the DHFR2 RNA binds to the DHFR protein it could prevent the DHFR mRNA from interacting with the protein; subsequently allowing the DHFR RNA to be translated into a protein. Alternatively, if the DHFR2 RNA levels are diminished in the cell the DHFR protein would be free to bind to the DHFR mRNA ultimately inhibiting its translation. Created with BioRender.com.*

The expression of several of the remaining OCM proteins were also affected when the *DHFR2* gene was silenced and up-regulated. The expression of the two other enzymes (SHMT1 and TYMS) involved in the de novo thymidylate syntheses pathway were impacted when the expression of the *DHFR2* gene was eliminated (Chapter 7, Section 7.3.1.1). This result suggested that the de novo thymidylate synthesis pathway was impacted when the expression of the *DHFR2* gene was

modulated for a significant time. Therefore, the *DHFR2* gene may indirectly regulate the de novo thymidylate synthesis pathway through its interaction with the *DHFR* gene.

8.3. Limitations associated with this PhD thesis

Chapter 5 focused on examining multiple cell lines/tissues and several cellular contexts for the endogenous DHFR2 protein. It is worth noting that there are some limitations associated with the experiments performed in this chapter. The adult cell/tissue specific, DHFR complementation and ubiquitination experiments were carried out in DCU, Ireland; with the native DHFR2 protein not detected in any of these samples (Chapter 5, Section 5.3.1-5.3.3). Whereas the human embryo/foetal tissue experiment was performed at UCL, London; where the endogenous DHFR2 protein was identified in the human embryonic and foetal heart tissue (Chapter 5, Section 5.3.4). Executing these experiments in different locations meant the sample preparation, instrumentation, TDA LC-MS/MS method and software utilised were different. Both facilities employed state of the art mass spectrometers and highly sensitive TDA LC-MS/MS methodologies. Therefore, both research centres should have the ability to detect the endogenous DHFR2 protein even if it is expressed at a very low abundance. However, the MRM LC-MS/MS method employed in UCL would have superior sensitivity (Tang et al., 2022). Therefore, if the native DHFR2 protein was present at an extremely low abundance in the adult cell/tissue specific, DHFR complementation and ubiquitination samples the PRM LC-MS/MS method employed in DCU may have missed it. Therefore, the adult cell/tissue specific, DHFR complementation and ubiquitination samples should be reanalysed using the preparation, instrumentation, MRM LC-MS/MS method and software employed in UCL, London to confirm the results obtained in DCU.

8.4. Contributions and concluding remarks

Knowledge and data in relation to the human *DHFR*2 retrogene was limited prior to this project. The findings in this PhD thesis have provided an insight into the expression and behaviours of the *DHFR2* gene. In the process of characterising the *DHFR2* gene, it was revealed that it encodes a protein that is expressed in the heart during early development; possibly to supply dihydrofolate reductase activity to rapidly growing embryos and foetuses. This result was a major breakthrough for this project and even more so for the OCM field. However, functional analysis needs to be performed on the endogenous heart protein to fully conclude its role in human embryos and foetuses. The research carried out in this thesis also showed that the *DHFR2* gene has an additional

role at the RNA level. The DHFR2 RNA has the ability to regulate the expression of the *DHFR* gene and in turn, the other de novo thymidylate synthesis genes. This was a huge result for this project as the *DHFR* gene plays such an important role in OCM which is required to maintain normal cell function (Fox & Stover, 2008). Furthermore, given that the *DHFR* gene is an extremely important drug target for the treatment of multiple types of clinical diseases (cancer and inflammatory diseases) (Bailey, 1995; Cronstein, 1996), the *DHFR2* gene now has the potential to also become a valuable antifolate drug target that could be targeted at the RNA level and used in tandem with current anti-DHFR enzyme drugs. To conclude, the research performed in this thesis has revealed the biological roles of the human *DHFR2* gene. These significant findings have paved the way for future research on this functional gene that may have a major impact on human health and disease.

8.5. Future work

Like any research topic, there is an overwhelming amount of experiments that could be carried out on the *DHFR2* gene. This section will describe the most beneficial experiments that could be performed.

8.5.1. Recommended experiments for assessing the translational capabilities of the DHFR2 transcripts

The differential association found between the two main DHFR2 transcripts (DHFR2_201 and DHFR2_202) and the ribosome was an interesting find given that they are quite homologous. The only obvious difference in the DHFR2_201 and DHFR2_202 transcripts is the exon 1 nucleotide sequence (Appendix F). The contrasting sequence at the start of these transcripts may result in them having different RNA secondary structures. The secondary structure of RNA has been found to determine the binding efficiency of the ribosome to RNA, which subsequently dictates the translational rate of the RNA molecule (de Smit & van Duin, 1990). Therefore, the next logical step would be to investigate the secondary structures of the DHFR2_201 and DHFR2_202 transcripts to determine whether this could be the reason why they have differential translational capabilities. There are several online RNA secondary structure software packages that could be employed to carry out this experiment including RNAstructure and ViennaRNA (Gruber et al., 2008; Reuter & Mathews, 2010).

Chapter 3 investigated the translational capabilities of only the DHFR2_201 and DHFR2_202 transcripts, as they were found by Ensembl to be the most plausible mRNA isoforms to generate a functional DHFR2 protein. However, Ensembl has predicted that the DHFR2_206 transcript also contains the 561 nucleotide long ORF that has the potential to synthesise the DHFR2 protein (Ensembl, 2019). Therefore, the sucrose cushion ultracentrifugation RT-qPCR technique employed in Chapter 3 could be employed to assess the translational capabilities of the DHFR2_206 transcript to determine whether it encodes the DHFR2 protein.

8.5.2. Recommended experiments for the recombinant DHFR2 protein

The generation of the recombinant human DHFR2 protein in the mammalian expression system could be used to characterise the DHFR2 protein. The colorimetric dihydrofolate reductase assay kit from Abcam could be employed to determine whether the recombinant human DHFR2 protein harbours reductase activity (i) to reduce dihydrofolate to tetrahydrofolate to aid one-carbon metabolism or (ii) to convert dihydrobiopterin to tetrahydrobiopterin to support nitric oxide synthesis (Crabtree et al., 2009, 2011; Fox & Stover, 2008; Kohen, 2015). The results from this experiment has the potential to reveal the function of the endogenous DHFR2 protein.

8.5.3. Recommended experiments for investigating the endogenous DHFR2 protein

There are several experiments that could be performed to expand the information known about the endogenous DHFR2 protein. Firstly, the protein samples from the adult cell/tissue specific, DHFR complementation and ubiquitination experiment could be reanalysed using the preparation, instrumentation, MRM LC-MS/MS method and software employed in UCL, London to confirm the results obtained in DCU (Section 8.3). Secondly, MRM LC-MS/MS analysis could be performed on adult heart tissue to determine if the DHFR2 protein is a heart-specific protein or a developmental heart-specific protein. Finally, functional analysis could be performed on the endogenous DHFR2 protein found in the heart tissue of human embryos and foetuses to determine its cellular role. An enzymatic activity experiment would be one of the most valuable functional experiments. The colorimetric dihydrofolate reductase assay described in Section 8.5.2 could be employed to assess the enzyme activity of the endogenous DHFR2 protein (Kohen, 2015).

8.5.4. Recommended experiments for investigating the translational regulation of the DHFR2_201 and DHFR2_202 transcripts

In Chapter 6 it was concluded that several of the DHFR2_201 and DHFR_202 transcripts may be retained in the nuclease by the paraspeckle. This was due to the IRAlu elements in the 3'UTR of these transcripts being A to I edited. However, it is not clear the level at which the IRAlu elements need to be A to I edited for the RNA molecule to be retained by the paraspeckle. Therefore, further experiments need to be performed to determine if some of the DHFR2_201 and/or DHFR2_202 RNA isoforms are actually targeted by the paraspeckle. A co-localisation experiment on the DHFR2_201/DHFR2_202 RNA and the paraspeckle protein would determine if they are situated in the same location in the cell. There are multiple fluorescent-based techniques that can be employed to recognise specific nucleic acids and/or proteins including fluorescence in situ hybridization (FISH) or hybridisation chain reaction (HCR) (Evanko, 2004; Meyer et al., 2017).

This chapter also showed that the majority of the DHFR2_201 and DHFR2_202 transcripts were not translationally regulated by the paraspeckle nuclear retention mechanism. Therefore, there must be other regulatory mechanisms inhibiting the translation of these RNA molecules. There are multiple different mechanisms of translational regulation (e.g. RNA secondary structure or cis-acting elements within the transcripts that are regulated by protein complexes or microRNAs) that could be investigated for the DHFR2_201 and DHFR2_202 RNA isoforms (Gebauer & Hentze, 2004; Kervestin & Amrani, 2004; Oliveto et al., 2017).

8.5.5. Recommended experiments to determine which DHFR molecule the DHFR2 RNA binds to and regulates

There are multiple experiments that could be carried out to determine which DHFR molecule the DHFR2 RNA binds to and regulates. A co-localisation experiment on the DHFR2 RNA and DHFR RNA/protein is one of the methodologies that could be employed to answer this research question. There are multiple fluorescent-based co-localisation techniques that could be employed to do this including FISH or HCR (Evanko, 2004; Meyer et al., 2017).

Bibliography

Abali, E. E., Skacel, N. E., Celikkaya, H., & Hsieh, Y. (2008). Chapter 9 Regulation of Human Dihydrofolate Reductase Activity and Expression. In *Vitamins & Hormones* (Vol. 79, pp. 267–292). Elsevier. https://doi.org/10.1016/S0083-6729(08)00409-3

Adaway, J. E., Keevil, B. G., & Owen, L. J. (2015). Liquid chromatography tandem mass spectrometry in the clinical laboratory. *Annals of Clinical Biochemistry*, *52*(Pt 1), 18–38. https://doi.org/10.1177/0004563214557678

Agranat, L., Raitskin, O., Sperling, J., & Sperling, R. (2008). The editing enzyme ADAR1 and the mRNA surveillance protein hUpf1 interact in the cell nucleus. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(13), 5028–5033. https://doi.org/10.1073/pnas.0710576105

Ahmed, A. S. I., Dong, K., Liu, J., Wen, T., Yu, L., Xu, F., Kang, X., Osman, I., Hu, G., Bunting, K. M., Crethers, D., Gao, H., Zhang, W., Liu, Y., Wen, K., Agarwal, G., Hirose, T., Nakagawa, S., Vazdarjanova, A., & Zhou, J. (2018). Long noncoding RNA NEAT1 (nuclear paraspeckle assembly transcript 1) is critical for phenotypic switching of vascular smooth muscle cells. *Proceedings of the National Academy of Sciences*, *115*(37), E8660–E8667. https://doi.org/10.1073/pnas.1803725115

Alur, M. D. (1999). Metabolic Pathways | Nitrogen Metabolism. In R. K. Robinson (Ed.), *Encyclopedia of Food Microbiology* (pp. 1288–1298). Elsevier. https://doi.org/10.1006/rwfm.1999.1005

Ami, N., Bernstein, M., Boucher, F., Rieder, M., Parker, L., Canadian Paediatric Society, D. T. and H. S. C., Canadian Paediatric, & Committee, D. T. and H. S. (2016). Folate and neural tube defects: The role of supplements and food fortification. *Paediatrics & Child Health*, *21*(3), 145–154. https://doi.org/10.1093/pch/21.3.145

Amunugama, R., Jones, R., Ford, M., & Allen, D. (2013). Bottom-Up Mass Spectrometry–Based Proteomics as an Investigative Analytical Tool for Discovery and Quantification of Proteins in Biological Samples. *Advances in Wound Care*, *2*(9), 549–557. https://doi.org/10.1089/wound.2012.0384

Anagnou, N. P., Antonarakis, S. E., O'Brien, S. J., Modi, W. S., & Nienhuis, A. W. (1988). Chromosomal localization and racial distribution of the polymorphic human dihydrofolate reductase pseudogene (DHFRPI). *American Journal of Human Genetics*, *42*(2), 345–352.

Anagnou, N. P., O'Brien, S. J., Shimada, T., Nash, W. G., Chen, M. J., & Nienhuis, A. W. (1984). Chromosomal organization of the human dihydrofolate reductase genes: Dispersion, selective amplification, and a novel form of polymorphism. *Proceedings of the National Academy of Sciences of the United States of America*, *81*(16), 5170–5174.

Anantharaman, A., Jadaliha, M., Tripathi, V., Nakagawa, S., Hirose, T., Jantsch, M. F., Prasanth, S. G., & Prasanth, K. V. (2016). Paraspeckles modulate the intranuclear distribution of paraspeckle-associated Ctn RNA. *Scientific Reports*, *6*(1), 34043. https://doi.org/10.1038/srep34043

Andersen, J. S., Lyon, C. E., Fox, A. H., Leung, A. K. L., Lam, Y. W., Steen, H., Mann, M., & Lamond, A. I. (2002). Directed proteomic analysis of the human nucleolus. *Current Biology: CB*, *12*(1), 1–11. https://doi.org/10.1016/s0960-9822(01)00650-9

Anderson, D. D., Quintero, C. M., & Stover, P. J. (2011). Identification of a de novo thymidylate biosynthesis pathway in mammalian mitochondria. *Proceedings of the National Academy of Sciences*, *108*(37), 15163–15168.

Anderson, D. D., Woeller, C. F., Chiang, E.-P., Shane, B., & Stover, P. J. (2012). Serine hydroxymethyltransferase anchors de novo thymidylate synthesis pathway to nuclear lamina for DNA synthesis. *The Journal of Biological Chemistry*, *287*(10), 7051–7062. https://doi.org/10.1074/jbc.M111.333120

Anderson, D. D., Woeller, C. F., & Stover, P. J. (2007). Small ubiquitin-like modifier-1 (SUMO-1) modification of thymidylate synthase and dihydrofolate reductase. *Clinical Chemical Laboratory Medicine*, *45*(12), 1760–1763. https://doi.org/10.1515/CCLM.2007.355

Angier, R. B., Boothe, J. H., Hutchings, B. L., Mowat, J. H., Semb, J., Stokstad, E. L., Subbarow, Y., Waller, C. W., Cosulich, D. B., Fahrenbach, M. J., Hultquist, M. E., Kuh, E., Northey, E. H., Seeger, D. R., Sickels, J. P., & Smith, J. M. (1945). Synthesis of a compound identical with the l. casei factor isolated from liver. *Science (New York, N.Y.)*, *102*(2644), 227–228. https://doi.org/10.1126/science.102.2644.227

Appling, D. R. (1991). Compartmentation of folate-mediated one-carbon metabolism in eukaryotes. *The FASEB Journal*, *5*(12), 2645–2651.

Askari, B. S., & Krajinovic, M. (2010). Dihydrofolate reductase gene variations in susceptibility to disease and treatment outcomes. *Current Genomics*, *11*(8), 578–583. https://doi.org/10.2174/138920210793360925

Aslam, B., Basit, M., Nisar, M. A., Khurshid, M., & Rasool, M. H. (2017). Proteomics: Technologies and Their Applications. *Journal of Chromatographic Science*, *55*(2), 182–196. https://doi.org/10.1093/chromsci/bmw167

Aslinia, F., Mazza, J. J., & Yale, S. H. (2006). Megaloblastic Anemia and Other Causes of Macrocytosis. *Clinical Medicine and Research*, *4*(3), 236–236.

Azizkhan, J. C., Jensen, D. E., Pierce, A. J., & Wade, M. (1993). Transcription from TATA-less promoters: Dihydrofolate reductase as a model. *Critical Reviews in Eukaryotic Gene Expression*, *3*(4), 229–254.

Bailey, L. B. (1995). *Folate in Health and Disease* (p. 22). Marcel Dekker, Inc., New York.

Bailey, L. B., Rampersaud, G. C., & Kauwell, G. P. A. (2003). Folic acid supplements and fortification affect the risk for neural tube defects, vascular disease and cancer: Evolving science. *The Journal of Nutrition*, *133*(6), 1961S-1968S. https://doi.org/10.1093/jn/133.6.1961S

Bailey, L. B., Stover, P. J., McNulty, H., Fenech, M. F., Gregory, J. F., Mills, J. L., Pfeiffer, C. M., Fazili, Z., Zhang, M., Ueland, P. M., Molloy, A. M., Caudill, M. A., Shane, B., Berry, R. J., Bailey, R. L., Hausman, D. B., Raghavan, R., Raiten, D. J., & Raiten, D. J. (2015). Biomarkers of Nutrition for Development-Folate Review. *The Journal of Nutrition*, *145*(7), 1636S-1680S. https://doi.org/10.3945/jn.114.206599

Bar-joseph, I., Pras, E., Reznik-Wolf, H., Marek-Yagel, D., Abu-Horvitz, A., Dushnitzky, M., Goldstein, N., Rienstein, S., Dekel, M., Pode-Shakked, B., Zlotnik, J., Benarrosh, A., Gillery, P., Hofliger, N., Auray-Blais, C., Garnotel, R., & Anikster, Y. (2012). Mutations in the sarcosine dehydrogenase gene in patients with sarcosinemia. *Human Genetics*, *131*(11), 1805–1810. https://doi.org/10.1007/s00439-012-1207-x

Bass, B. L. (2006). How Does RNA Editing Affect dsRNA-mediated Gene Silencing? *Cold Spring Harbor Symposia on Quantitative Biology*, *71*, 285–292. https://doi.org/10.1101/sqb.2006.71.037

Bass, B. L., & Weintraub, H. (1988). An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell*, *55*(6), 1089–1098. https://doi.org/10.1016/0092-8674(88)90253-x

Bassett, M. N., & Sammán, N. C. (2010). Folate content and retention in selected raw and processed foods. *Archivos Latinoamericanos De Nutricion*, *60*(3), 298–305.

Bauer, D. E., Canver, M. C., & Orkin, S. H. (2015). Generation of genomic deletions in mammalian cell lines via CRISPR/Cas9. *Journal of Visualized Experiments: JoVE*, *95*, e52118. https://doi.org/10.3791/52118

Beaudin, A. E., & Stover, P. J. (2009). Insights into metabolic mechanisms underlying folate-responsive neural tube defects: A minireview. *Birth Defects Research. Part A, Clinical and Molecular Teratology*, *85*(4), 274–284. https://doi.org/10.1002/bdra.20553

Bendall, J. K., Douglas, G., McNeill, E., Channon, K. M., & Crabtree, M. J. (2014). Tetrahydrobiopterin in Cardiovascular Health and Disease. *Antioxidants & Redox Signaling*, *20*(18), 3040–3077. https://doi.org/10.1089/ars.2013.5566

Bertero, A., Brown, S., & Vallier, L. (2017). Chapter 2—Methods of Cloning. In M. Jalali, F. Y. L. Saldanha, & M. Jalali (Eds.), *Basic Science Methods for Clinical Researchers* (pp. 19–39). Academic Press. https://doi.org/10.1016/B978-0-12-803077-6.00002-3

Bhatwa, A., Wang, W., Hassan, Y. I., Abraham, N., Li, X.-Z., & Zhou, T. (2021). Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in Escherichia coli and Strategies to Address Them for Industrial Applications. *Frontiers in Bioengineering and Biotechnology*, *9*. https://www.frontiersin.org/article/10.3389/fbioe.2021.630551

Blair, J. D., Hockemeyer, D., Doudna, J. A., Bateup, H. S., & Floor, S. N. (2017). Widespread Translational Remodeling during Human Neuronal Differentiation. *Cell Reports*, *21*(7), 2005–2016. https://doi.org/10.1016/j.celrep.2017.10.095

Blakley, R. L. (1995). Eukaryotic dihydrofolate reductase. *Advances in Enzymology and Related Areas of Molecular Biology*, *70*, 23–102. https://doi.org/10.1002/9780470123164.ch2

Blom, H. J., & Smulders, Y. (2011). Overview of homocysteine and folate metabolism. With special references to cardiovascular disease and neural tube defects. *Journal of Inherited Metabolic Disease*, *34*(1), 75–81. https://doi.org/10.1007/s10545-010-9177-4

Boettcher, M., & McManus, M. T. (2015). Choosing the Right Tool for the Job: RNAi, TALEN or CRISPR. *Molecular Cell*, *58*(4), 575–585. https://doi.org/10.1016/j.molcel.2015.04.028

Bond, C. S., & Fox, A. H. (2009). Paraspeckles: Nuclear bodies built on long noncoding RNA. *Journal of Cell Biology*, *186*(5), 637–644. https://doi.org/10.1083/jcb.200906113

Borchert, G. M., Gilmore, B. L., Spengler, R. M., Xing, Y., Lanier, W., Bhattacharya, D., & Davidson, B. L. (2009). Adenosine deamination in human transcripts generates novel microRNA binding sites. *Human Molecular Genetics*, *18*(24), 4801–4807. https://doi.org/10.1093/hmg/ddp443

Brar, G. A., & Weissman, J. S. (2015). Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nature Reviews Molecular Cell Biology*, *16*(11), 651–664. https://doi.org/10.1038/nrm4069

Brosnan, M. E., & Brosnan, J. T. (2016). Formate: The Neglected Member of One-Carbon Metabolism. *Annual Review of Nutrition*, *36*(1), 369–388. https://doi.org/10.1146/annurev-nutr-071715-050738

Brunner, E., Yagi, R., Debrunner, M., Beck-Schneider, D., Burger, A., Escher, E., Mosimann, C., Hausmann, G., & Basler, K. (2019). CRISPR-induced double-strand breaks trigger recombination between homologous chromosome arms. *Life Science Alliance*, *2*(3), e201800267. https://doi.org/10.26508/lsa.201800267

Burgess, K., Bennett, C., Mosnier, H., Kwatra, N., Bethel, F., & Jadavji, N. M. (2020). The Antioxidant Role of One-Carbon Metabolism on Stroke. *Antioxidants (Basel, Switzerland)*, *9*(11), E1141. https://doi.org/10.3390/antiox9111141

Burns, C. M., Chu, H., Rueter, S. M., Hutchinson, L. K., Canton, H., Sanders-Bush, E., & Emeson, R. B. (1997). Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature*, *387*(6630), 303–308. https://doi.org/10.1038/387303a0

Busch, K. L. (2003). Mass Spectrometry. In R. A. Meyers (Ed.), *Encyclopedia of Physical Science and Technology (Third Edition)* (pp. 145–158). Academic Press. https://doi.org/10.1016/B0-12-227410-5/00406-3

Büyükköroğlu, G., Dora, D. D., Özdemir, F., & Hızel, C. (2018). Chapter 15—Techniques for Protein Analysis. In D. Barh & V. Azevedo (Eds.), *Omics Technologies and Bio-Engineering* (pp. 317–351). Academic Press. https://doi.org/10.1016/B978-0-12-804659-3.00015-4

Calviello, L., Mukherjee, N., Wyler, E., Zauber, H., Hirsekorn, A., Selbach, M., Landthaler, M., Obermayer, B., & Ohler, U. (2016). Detecting actively translated open reading frames in ribosome profiling data. *Nature Methods*, *13*(2), 165–170. https://doi.org/10.1038/nmeth.3688

Canha-Gouveia, A., Paradela, A., Ramos-Fernández, A., Prieto-Sánchez, M. T., Sánchez-Ferrer, M. L., Corrales, F., & Coy, P. (2019). Which Low-Abundance Proteins are Present in the Human Milieu of Gamete/Embryo Maternal Interaction? *International Journal of Molecular Sciences*, *20*(21), 5305. https://doi.org/10.3390/ijms20215305

Capriotti, L., Baraldi, E., Mezzetti, B., Limera, C., & Sabbadini, S. (2020). Biotechnological Approaches: Gene Overexpression, Gene Silencing, and Genome Editing to Control Fungal and Oomycete Diseases in Grapevine. *International Journal of Molecular Sciences*, *21*(16), 5701. https://doi.org/10.3390/ijms21165701

Carapito, C., Lane, L., Benama, M., Opsomer, A., Mouton-Barbosa, E., Garrigues, L., Gonzalez De Peredo, A., Burel, A., Bruley, C., Gateau, A., Bouyssié, D., Jaquinod, M., Cianferani, S., Burlet-Schiltz, O., Van Dorsselaer, A., Garin, J., & Vandenbrouck, Y. (2015). Computational and mass-spectrometry-based workflow for the discovery and validation of missing human proteins: Application to chromosomes 2 and 14. *Journal of Proteome Research*, *14*(9), 3621–3634.

Castellanos-Sinco, H. B., Ramos-Peñafiel, C. O., Santoyo-Sánchez, A., Collazo-Jaloma, J., Martínez-Murillo, C., Montaño-Figueroa, E., & Sinco-Ángeles, A. (2015). Megaloblastic anaemia: Folic acid and vitamin B12 metabolism. *Revista Médica Del Hospital General De México*, *78*(3), 135–143. https://doi.org/10.1016/j.hgmx.2015.07.001

Chakraborty, J., & Stover, P. J. (2020). Deoxyuracil in DNA in Health and Disease. *Current Opinion in Clinical Nutrition and Metabolic Care*, *23*(4), 247–252. https://doi.org/10.1097/MCO.0000000000000660

Chan, Y.-M., Bailey, R., & O'Connor, D. L. (2013). Folate1. *Advances in Nutrition*, *4*(1), 123–125. https://doi.org/10.3945/an.112.003392

Chandramouli, K., & Qian, P.Y. (2009). Proteomics: Challenges, Techniques and Possibilities to Overcome Biological Sample Complexity. *Human Genomics and Proteomics: HGP*, 2009, 239204. https://doi.org/10.4061/2009/239204

Chang, H. H. Y., Pannunzio, N. R., Adachi, N., & Lieber, M. R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nature Reviews Molecular Cell Biology*, *18*(8), 495–506. https://doi.org/10.1038/nrm.2017.48

Chassé, H., Boulben, S., Costache, V., Cormier, P., & Morales, J. (2017). Analysis of translation using polysome profiling. *Nucleic Acids Research*, *45*(3), e15–e15. https://doi.org/10.1093/nar/gkw907

Chen, K., Pittman, R. N., & Popel, A. S. (2008a). Nitric Oxide in the Vasculature: Where Does It Come From and Where Does It Go? A Quantitative Perspective. *Antioxidants & Redox Signaling*, *10*(7), 1185–1198. https://doi.org/10.1089/ars.2007.1959

Chen, L.-L., & Carmichael, G. G. (2009). Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: Functional role of a nuclear noncoding RNA. *Molecular Cell*, *35*(4), 467–478. https://doi.org/10.1016/j.molcel.2009.06.027

Chen, L.-L., & Carmichael, G. G. (2012). Nuclear Editing of mRNA 3'-UTRs. *Current Topics in Microbiology and Immunology*, *353*, 111–121. https://doi.org/10.1007/82_2011_149

Chen, L.-L., DeCerbo, J. N., & Carmichael, G. G. (2008b). Alu element-mediated gene silencing. *The EMBO Journal*, *27*(12), 1694–1705. https://doi.org/10.1038/emboj.2008.94

Chen, M. J., Shimada, T., Moulton, A. D., Cline, A., Humphries, R. K., Maizel, J., & Nienhuis, A. W. (1984). The functional human dihydrofolate reductase gene. *The Journal of Biological Chemistry*, *259*(6), 3933–3943.

Cho, W. C. S. (2007). Proteomics Technologies and Challenges. *Genomics, Proteomics & Bioinformatics*, *5*(2), 77–85. https://doi.org/10.1016/S1672-0229(07)60018-7

Chun, S. Y., Rodriguez, C. M., Todd, P. K., & Mills, R. E. (2016). SPECtre: A spectral coherence--based classifier of actively translated transcripts from ribosome profiling sequence data. *BMC Bioinformatics*, *17*(1), 482. https://doi.org/10.1186/s12859-016-1355-4

Ciechanover, A. (1994). The ubiquitin-proteasome proteolytic pathway. *Cell*, *79*(1), 13–21.

Cieślik, E., & Cieślik, I. (2018). Occurrence and significance of folic acid. *Pteridines*, *29*(1), 187–195. https://doi.org/10.1515/pteridines-2018-0017

Clemson, C. M., Hutchinson, J. N., Sara, S. A., Ensminger, A. W., Fox, A. H., Chess, A., & Lawrence, J. B. (2009). An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Molecular Cell*, *33*(6), 717–726. https://doi.org/10.1016/j.molcel.2009.01.026

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science (New York, N.Y.)*, *339*(6121), 819–823. https://doi.org/10.1126/science.1231143

Crabtree, M. J., Hale, A. B., & Channon, K. M. (2011). Dihydrofolate reductase protects endothelial nitric oxide synthase from uncoupling in tetrahydrobiopterin deficiency. *Free Radical Biology & Medicine*, *50*(11), 1639–1646. https://doi.org/10.1016/j.freeradbiomed.2011.03.010

Crabtree, M. J., Tatham, A. L., Hale, A. B., Alp, N. J., & Channon, K. M. (2009). Critical Role for Tetrahydrobiopterin Recycling by Dihydrofolate Reductase in Regulation of Endothelial Nitric-oxide Synthase Coupling: relative importance of the de novo biopterin synthesis versus salvage pathways. *Journal of Biological Chemistry*, *284*(41), 28128–28136. https://doi.org/10.1074/jbc.M109.041483

Crappé, J., Ndah, E., Koch, A., Steyaert, S., Gawron, D., De Keulenaer, S., De Meester, E., De Meyer, T., Van Criekinge, W., Van Damme, P., & Menschaert, G. (2015). PROTEOFORMER: Deep proteome coverage through ribosome profiling and MS integration. *Nucleic Acids Research*, *43*(5), e29. https://doi.org/10.1093/nar/gku1283

Crick, F. (1970). Central Dogma of Molecular Biology. *Nature*, *227*(5258), 561–563. https://doi.org/10.1038/227561a0

Cronstein, B. N. (1996). Molecular therapeutics. Methotrexate and its mechanism of action. *Arthritis & Rheumatism*, *39*(12), 1951–1960. https://doi.org/10.1002/art.1780391203

Dalla Venezia, N., Vincent, A., Marcel, V., Catez, F., & Diaz, J.-J. (2019). Emerging Role of Eukaryote Ribosomes in Translational Control. *International Journal of Molecular Sciences*, *20*(5), 1226. https://doi.org/10.3390/ijms20051226

Daly, L. E., Kirke, P. N., Molloy, A., Weir, D. G., & Scott, J. M. (1995). Folate levels and neural tube defects. Implications for prevention. *J Amer Med Assn*, *274*(21), 1698–1702. https://doi.org/10.1001/jama.1995.03530210052030

Davis, A. J., & Chen, D. J. (2013). DNA double strand break repair via non-homologous end-joining. *Translational Cancer Research*, *2*(3), 130–143. https://doi.org/10.3978/j.issn.2218-676X.2013.04.02

de Smit, M. H., & van Duin, J. (1990). Secondary structure of the ribosome binding site determines translational efficiency: A quantitative analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 87(19), 7668–7672. https://doi.org/10.1073/pnas.87.19.7668

Dermit, M., Dodel, M., & Mardakheh, F. K. (2017). Methods for monitoring and measurement of protein translation in time and space. *Molecular Biosystems*, *13*(12), 2477–2488. https://doi.org/10.1039/c7mb00476a

Di Luca, M., Fitzpatrick, E., Burtenshaw, D., Liu, W., Helt, J. C., Hakimjavadi, R., Corcoran, E., Gusti, Y., Sheridan, D., Harman, S., Lally, C., Redmond, E. M., & Cahill, P. A. (2021). The calcium binding protein S100β marks hedgehog-responsive resident vascular stem cells within vascular lesions. *Npj Regenerative Medicine*, *6*(1), 1–15.

Dumitrescu, C., Biondi, R., Xia, Y., Cardounel, A. J., Druhan, L. J., Ambrosio, G., & Zweier, J. L. (2007). Myocardial ischemia results in tetrahydrobiopterin (BH4) oxidation with impaired endothelial function ameliorated by BH4. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(38), 15081–15086. https://doi.org/10.1073/pnas.0702986104

Ebbing, M., Bønaa, K. H., Nygård, O., Arnesen, E., Ueland, P. M., Nordrehaug, J. E., Rasmussen, K., Njølstad, I., Refsum, H., Nilsen, D. W., Tverdal, A., Meyer, K., & Vollset, S. E. (2009). Cancer incidence and mortality after treatment with folic acid and vitamin B12. *J Amer Med Assn*, *302*(19), 2119–2126. https://doi.org/10.1001/jama.2009.1622

Ensembl. (2019). *Gene: DHFR2 (ENSG00000178700)—Summary—Homo sapiens—Ensembl genome browser 96*. https://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000178700;r=3:94047836-94063389

Ensembl - DHFR. (2019). *Gene: DHFR (ENSG00000228716)—Summary—Homo sapiens—Ensembl genome browser 96*. https://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000228716;r=5:80626226-80654983;t=ENST00000504396

Ercikan, E., Banerjee, D., Waltham, M., Schnieders, B., Scotto, K. W., & Bertino, J. R. (1993). Translational regulation of the synthesis of dihydrofolate reductase. *Advances in Experimental Medicine and Biology*, *338*, 537–540. https://doi.org/10.1007/978-1-4615-2960-6_109

Ercikan-Abali, E. A., Banerjee, D., Waltham, M. C., Skacel, N., Scotto, K. W., & Bertino, J. R. (1997). Dihydrofolate reductase protein inhibits its own translation by binding to dihydrofolate reductase mRNA sequences within the coding region. *Biochemistry*, *36*(40), 12317–12322. https://doi.org/10.1021/bi971026e

Evanko, D. (2004). Hybridization chain reaction. *Nature Methods*, *1*(3), 186–186. https://doi.org/10.1038/nmeth1204-186a

Fablet, M., Bueno, M., Potrzebowski, L., & Kaessmann, H. (2009). Evolutionary Origin and Functions of Retrogene Introns. *Molecular Biology and Evolution*, *26*(9), 2147–2156. https://doi.org/10.1093/molbev/msp125

Field, M. S., Kamynina, E., Agunloye, O. C., Liebenthal, R. P., Lamarre, S. G., Brosnan, M. E., Brosnan, J. T., & Stover, P. J. (2014). Nuclear Enrichment of Folate Cofactors and Methylenetetrahydrofolate Dehydrogenase 1 (MTHFD1) Protect de Novo Thymidylate Biosynthesis during Folate Deficiency. *The Journal of Biological Chemistry*, *289*(43), 29642–29650. https://doi.org/10.1074/jbc.M114.599589

Field, M. S., Kamynina, E., & Stover, P. J. (2016). MTHFD1 Regulates Nuclear de novo Thymidylate Biosynthesis and Genome Stability. *Biochimie*, *126*, 27–30. https://doi.org/10.1016/j.biochi.2016.02.001

Food and Nutrition Board. (1998). Dietary Reference Intakes for Thiamin, Riboflavin, Niacin, Vitamin B6, Folate, Vitamin B12, Pantothenic Acid, Biotin, and Choline. National Academies Press (US). http://www.ncbi.nlm.nih.gov/books/NBK114310/

Fox, A. H., Bond, C. S., & Lamond, A. I. (2005). P54nrb forms a heterodimer with PSP1 that localizes to paraspeckles in an RNA-dependent manner. *Molecular Biology of the Cell*, *16*(11), 5304–5315. https://doi.org/10.1091/mbc.e05-06-0587

Fox, A. H., Lam, Y. W., Leung, A. K. L., Lyon, C. E., Andersen, J., Mann, M., & Lamond, A. I. (2002). Paraspeckles: A novel nuclear domain. *Current Biology: CB*, *12*(1), 13–25. https://doi.org/10.1016/s0960-9822(01)00632-7

Fox, A. H., & Lamond, A. I. (2010). Paraspeckles. *Cold Spring Harbor Perspectives in Biology*, *2*(7), a000687. https://doi.org/10.1101/cshperspect.a000687

Fox, J. T., Shin, W. K., Caudill, M. A., & Stover, P. J. (2009). A UV-responsive internal ribosome entry site enhances serine hydroxymethyltransferase 1 expression for DNA damage repair. *The Journal of Biological Chemistry*, *284*(45), 31097–31108. https://doi.org/10.1074/jbc.M109.015800

Fox, J. T., & Stover, P. J. (2008). Chapter 1 Folate-Mediated One-Carbon Metabolism. *Vitamins and Hormones*, *79*, 1–44.

Fritsch, C., Herrmann, A., Nothnagel, M., Szafranski, K., Huse, K., Schumann, F., Schreiber, S., Platzer, M., Krawczak, M., Hampe, J., & Brosch, M. (2012). Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. *Genome Research*, *22*(11), 2208–2218. https://doi.org/10.1101/gr.139568.112

Froese, D. S., Fowler, B., & Baumgartner, M. R. (2019). Vitamin B12, folate, and the methionine remethylation cycle—Biochemistry, pathways, and regulation. *Journal of Inherited Metabolic Disease*, *42*(4), 673–685. https://doi.org/10.1002/jimd.12009

Frosst, P., Blom, H. J., Milos, R., Goyette, P., Sheppard, C. A., Matthews, R. G., Boers, G. J. H., den Heijer, M., Kluijtmans, L. A. J., van den Heuve, L. P., & Rozen, R. (1995). A candidate genetic risk factor for vascular disease: A common mutation in methylenetetrahydrofolate reductase. *Nature Genetics*, *10*(1), 111–113. https://doi.org/10.1038/ng0595-111

Funanage, V. L., Myoda, T. T., Moses, P. A., & Cowell, H. R. (1984). Assignment of the human dihydrofolate reductase gene to the q11—Q22 region of chromosome 5. *Molecular and Cellular Biology*, *4*(10), 2010–2016.

Furie, K. L., & Kelly, P. J. (2006). Homocyst(e)ine and stroke. *Seminars in Neurology*, *26*(1), 24–32. https://doi.org/10.1055/s-2006-933306

Garrels, J. I. (2001). Proteome. In S. Brenner & J. H. Miller (Eds.), *Encyclopedia of Genetics* (pp. 1575–1578). Academic Press. https://doi.org/10.1006/rwgn.2001.1042

Gebauer, F., & Hentze, M. W. (2004). Molecular mechanisms of translational control. *Nature Reviews Molecular Cell Biology*, *5*(10), 827–835. https://doi.org/10.1038/nrm1488

Geertsma, E. R., Groeneveld, M., Slotboom, D.-J., & Poolman, B. (2008). Quality control of overexpressed membrane proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(15), 5722–5727. https://doi.org/10.1073/pnas.0802190105

Gemmati, D., De Mattei, M., Catozzi, L., Della Porta, M., Serino, M. L., Ambrosio, C., Cuneo, A., Friso, S., Krampera, M., Orioli, E., Zeri, G., & Ongaro, A. (2009). DHFR 19-bp insertion/deletion polymorphism and MTHFR C677T in adult acute lymphoblastic leukaemia: Is the risk reduction due to intracellular folate unbalancing? *American Journal of Hematology*, *84*(8), 526–529. https://doi.org/10.1002/ajh.21451

Ghaderi, M., Sabahi, F., Sadeghi-Zadeh, M., Khanlari, Z., Jamaati, A., Mousavi-Nasab, D., Majidi-Gharenaz, N., Ajorloo, M., & Fazeli, M. (2014). Construction of an eGFP Expression Plasmid under Control of T7 Promoter and IRES Sequence for Assay of T7 RNA Polymerase Activity in Mammalian Cell Lines. *Iranian Journal of Cancer Prevention*, *7*(3), 137–141.

Giansanti, P., Tsiatsiani, L., Low, T. Y., & Heck, A. J. R. (2016). Six alternative proteases for mass spectrometry–based proteomics beyond trypsin. *Nature Protocols*, *11*(5), 993–1006. https://doi.org/10.1038/nprot.2016.057

Giardina, G., Paone, A., Tramonti, A., Lucchi, R., Marani, M., Magnifico, M. C., Bouzidi, A., Pontecorvi, V., Guiducci, G., Zamparelli, C., Rinaldo, S., Paiardini, A., Contestabile, R., & Cutruzzolà, F. (2018). The catalytic activity of serine hydroxymethyltransferase is essential for de novo nuclear dTMP synthesis in lung cancer cells. *The FEBS Journal*, *285*(17), 3238–3253. https://doi.org/10.1111/febs.14610

Giovannucci, E. (2002). Epidemiologic studies of folate and colorectal neoplasia: A review. *The Journal of Nutrition*, *132*(8 Suppl), 2350S-2355S. https://doi.org/10.1093/jn/132.8.2350S

Gonzalez, C., Sims, J. S., Hornstein, N., Mela, A., Garcia, F., Lei, L., Gass, D. A., Amendolara, B., Bruce, J. N., Canoll, P., & Sims, P. A. (2014). Ribosome Profiling Reveals a Cell-Type-Specific Translational Landscape in Brain Tumors. *The Journal of Neuroscience*, *34*(33), 10924–10936. https://doi.org/10.1523/JNEUROSCI.0084-14.2014

Goodarzi, H., Nguyen, H. C. B., Zhang, S., Dill, B. D., Molina, H., & Tavazoie, S. F. (2016). Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression. *Cell*, *165*(6), 1416–1427. https://doi.org/10.1016/j.cell.2016.05.046

Goto, Y., Yue, L., Yokoi, A., Nishimura, R., Uehara, T., Koizumi, S., & Saikawa, Y. (2001). A novel single-nucleotide polymorphism in the 3'-untranslated region of the human dihydrofolate reductase gene with enhanced expression. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research*, *7*(7), 1952–1956.

Graves, P. R., & Haystead, T. A. J. (2002). Molecular biologist's guide to proteomics. *Microbiology and Molecular Biology Reviews : MMBR*, *66*(1), 39–63; table of contents. https://doi.org/10.1128/MMBR.66.1.39-63.2002

Grebe, S. K., & Singh, R. J. (2011). LC-MS/MS in the Clinical Laboratory – Where to From Here? *The Clinical Biochemist Reviews*, *32*(1), 5–31.

Greenberg, J. A., Bell, S. J., Guan, Y., & Yu, Y. (2011). Folic Acid Supplementation and Pregnancy: More Than Just Neural Tube Defect Prevention. *Reviews in Obstetrics and Gynecology*, *4*(2), 52–59.

Greene, N. D. E., & Copp, A. J. (2014). Neural Tube Defects. *Annual Review of Neuroscience*, *37*, 221–242. https://doi.org/10.1146/annurev-neuro-062012-170354

Grow, E. J., Flynn, R. A., Chavez, S. L., Bayless, N. L., Wossidlo, M., Wesche, D. J., Martin, L., Ware, C. B., Blish, C. A., Chang, H. Y., Pera, R. A. R., & Wysocka, J. (2015). Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature*, *522*(7555), 221–225. https://doi.org/10.1038/nature14308

Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., & Hofacker, I. L. (2008). The Vienna RNA Websuite. *Nucleic Acids Research*, 36, W70–W74. https://doi.org/10.1093/nar/gkn188

Han, X., Aslanian, A., & Yates, J. R. (2008). Mass Spectrometry for Proteomics. *Current Opinion in Chemical Biology*, 12(5), 483–490. https://doi.org/10.1016/j.cbpa.2008.07.024

Harjumäki, R., Nugroho, R. W. N., Zhang, X., Lou, Y.-R., Yliperttula, M., Valle-Delgado, J. J., & Österberg, M. (2019). Quantified forces between HepG2 hepatocarcinoma and WA07 pluripotent stem cells with natural biomaterials correlate with in vitro cell behavior. *Scientific Reports*, *9*(1), 7354. https://doi.org/10.1038/s41598-019-43669-7

Hershko, A., & Ciechanover, A. (1992). The ubiquitin system for protein degradation. *Annual Review of Biochemistry*, *61*, 761–807.

Hestand, M. S., Van Houdt, J., Cristofoli, F., & Vermeesch, J. R. (2016). Polymerase specific error rates and profiles identified by single molecule sequencing. *Mutation Research*, *784–785*, 39–45. https://doi.org/10.1016/j.mrfmmm.2016.01.003

Heywood, W. E., Baud, A., Bliss, E., Sirka, E., Schott, J. M., Zetterberg, H., Galimberti, D., Sebire, N. J., & Mills, K. (2016). A High Throughput, Multiplexed and Targeted Proteomic CSF Assay to Quantify Neurodegenerative Biomarkers and Apolipoprotein E Isoforms Status. *Journal of Visualized Experiments : JoVE*, *116*, 54541. https://doi.org/10.3791/54541

Higuchi, M., Single, F. N., Köhler, M., Sommer, B., Sprengel, R., & Seeburg, P. H. (1993). RNA editing of AMPA receptor subunit GluR-B: A base-paired intron-exon structure determines position and efficiency. *Cell*, *75*(7), 1361–1370. https://doi.org/10.1016/0092-8674(93)90622-w

Horita, D. A., & Krupenko, S. A. (2017). Modeling of interactions between functional domains of ALDH1L1. *Chemico-Biological Interactions*, *276*, 23–30. https://doi.org/10.1016/j.cbi.2017.04.011

Hsieh, Y.-C., Skacel, N. E., Bansal, N., Scotto, K. W., Banerjee, D., Bertino, J. R., & Abali, E. E. (2009). Species-Specific Differences in Translational Regulation of Dihydrofolate Reductase. *Molecular Pharmacology*, *76*(4), 723–733. https://doi.org/10.1124/mol.109.055772

Hsu, C.-Y., Chiu, S.-W., Hong, K.-S., Saver, J. L., Wu, Y.-L., Lee, J.-D., Lee, M., & Ovbiagele, B. (2018). Folic Acid in Stroke Prevention in Countries without Mandatory Folic Acid Food Fortification: A Meta-Analysis of Randomized Controlled Trials. *Journal of Stroke*, *20*(1), 99–109. https://doi.org/10.5853/jos.2017.01522

Hu, Y., Zheng, M., Zhang, D., Gou, R., Liu, O., Wang, S., & Lin, B. (2021). Identification of the prognostic value of a 2-gene signature of the WNT gene family in UCEC using bioinformatics and real-world data. *Cancer Cell International*, *21*(1), 516. https://doi.org/10.1186/s12935-021-02215-0

Hughes, L., Carton, R., Minguzzi, S., McEntee, G., Deinum, E. E., O'Connell, M. J., & Parle-McDermott, A. (2015). An active second dihydrofolate reductase enzyme is not a feature of rat and mouse, but they do have activity in their mitochondria. *FEBS Letters*, *589*(15), 1855–1862.

Imbard, A., Benoist, J.-F., & Blom, H. J. (2013). Neural tube defects, folic acid and methylation. *International Journal of Environmental Research and Public Health*, *10*(9), 4352–4389. https://doi.org/10.3390/ijerph10094352

Ingolia, N. T. (2010). Chapter 6—Genome-Wide Translational Profiling by Ribosome Footprinting. In *Methods in Enzymology* (Vol. 470, pp. 119–142). Academic Press. https://doi.org/10.1016/S0076-6879(10)70006-9

Ingolia, N. T. (2014). Ribosome profiling: New views of translation, from single codons to genome scale. *Nature Reviews Genetics*, *15*(3), 205–213. https://doi.org/10.1038/nrg3645

Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M., & Weissman, J. S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nature Protocols*, *7*(8), 1534–1550. https://doi.org/10.1038/nprot.2012.086

Jaiswal, R. K., Kumar, P., Sharma, A., Mishra, D. K., & Yadava, P. K. (2017). Proteomic identification of proteins differentially expressed following overexpression of hTERT (human telomerase reverse transcriptase) in cancer cells. *PloS One*, *12*(7), e0181027. https://doi.org/10.1371/journal.pone.0181027

Jensen, D. E., Black, A. R., Swick, A. G., & Azizkhan, J. C. (1997). Distinct roles for Sp1 and E2F sites in the growth/cell cycle regulation of the DHFR promoter. *Journal of Cellular Biochemistry*, *67*(1), 24–31.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*, *337*(6096), 816–821. https://doi.org/10.1126/science.1225829

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., & Doudna, J. (2013). RNA-programmed genome editing in human cells. *ELife*, *2*, e00471. https://doi.org/10.7554/eLife.00471

Jo, J., Choi, S., Oh, J., Lee, S.-G., Choi, S. Y., Kim, K. K., & Park, C. (2019). Conventionally used reference genes are not outstanding for normalization of gene expression in human cancer research. *BMC Bioinformatics*, *20*(10), 245. https://doi.org/10.1186/s12859-019-2809-2

Johnson, W., Stenroos, E., Spychala, J., Chatkupt, S., Ming, S. X., & Buyske, S. (2004). New 19 bp deletion polymorphism in intron-1 of dihydrofolate reductase (DHFR): A risk factor for spina bifida acting in mothers during pregnancy? *American Journal of Medical Genetics. Part A*. https://doi.org/10.1002/ajmg.a.20505

Karthika, K. V., & Pati, H. P. (2019). Megaloblastic Anemia and Dual Nutritional Deficiency Anemia. In R. Saxena & H. P. Pati (Eds.), *Hematopathology: Advances in Understanding* (pp. 49–57). Springer. https://doi.org/10.1007/978-981-13-7713-6_3

Kaye, A. D., Jeha, G. M., Pham, A. D., Fuller, M. C., Lerner, Z. I., Sibley, G. T., Cornett, E. M., Urits, I., Viswanath, O., & Kevil, C. G. (2020). Folic Acid Supplementation in Patients with Elevated Homocysteine Levels. *Advances in Therapy*, *37*(10), 4149–4164. https://doi.org/10.1007/s12325-020-01474-z

Khan, K. H. (2013). Gene Expression in Mammalian Cells and its Applications. *Advanced Pharmaceutical Bulletin*, *3*(2), 257–263. https://doi.org/10.5681/apb.2013.042

Kervestin, S., & Amrani, N. (2004). Translational regulation of gene expression. *Genome Biology*, 5(12), 359. https://doi.org/10.1186/gb-2004-5-12-359

Kim, J. A., Vetrivel, P., Kim, S. M., Ha, S. E., Kim, H. H., Bhosale, P. B., Heo, J. D., Lee, W. S., Senthil, K., & Kim, G. S. (2021). Quantitative Proteomics Analysis for the Identification of Differential Protein Expression in Calf Muscles between Young and Old SD Rats Using Mass Spectrometry. *ACS Omega*, *6*(11), 7422–7433. https://doi.org/10.1021/acsomega.0c05821

Kim, J. W., Jeon, Y. J., Jang, M. J., Kim, J. O., Chong, S. Y., Ko, K. H., Hwang, S. G., Oh, D., Oh, J., & Kim, N. K. (2015). Association between folate metabolism-related polymorphisms and colorectal cancer risk. *Molecular and Clinical Oncology*, *3*(3), 639–648. https://doi.org/10.3892/mco.2015.520

Kim, K.-J., Kim, H.-E., Lee, K.-H., Han, W., Yi, M.-J., Jeong, J., & Oh, B.-H. (2004). Two-promoter vector is highly efficient for overproduction of protein complexes. *Protein Science : A Publication of the Protein Society*, *13*(6), 1698–1703. https://doi.org/10.1110/ps.04644504

Kim, Y. (1999). Folate and carcinogenesis: Evidence, mechanisms, and implications. *The Journal of Nutritional Biochemistry*, *10*(2), 66–88. https://doi.org/10.1016/s0955-2863(98)00074-6

Kim, Y. (2003). Role of Folate in Colon Cancer Development and Progression. *The Journal of Nutrition*, *133*(11), 3731S-3739S. https://doi.org/10.1093/jn/133.11.3731S

Kim, Y. (2004). Will mandatory folic acid fortification prevent or promote cancer? *The American Journal of Clinical Nutrition*, *80*(5), 1123–1128. https://doi.org/10.1093/ajcn/80.5.1123

Kim, Y. (2006). Folate: A magic bullet or a double edged sword for colorectal cancer prevention? *Gut*, *55*(10), 1387–1389. https://doi.org/10.1136/gut.2006.095463

Kim, Y. (2008). Folic Acid Supplementation and Cancer Risk: Point. *Cancer Epidemiology and Prevention Biomarkers*, *17*(9), 2220–2225. https://doi.org/10.1158/1055-9965.EPI-07-2557

King, H. A., & Gerber, A. P. (2016). Translatome profiling: Methods for genome-scale analysis of mRNA translation. *Briefings in Functional Genomics*, *15*(1), 22–31. https://doi.org/10.1093/bfgp/elu045

Kirke, P. N., Mills, J. L., Molloy, A. M., Brody, L. C., O'Leary, V. B., Daly, L., Murray, S., Conley, M., Mayne, P. D., Smith, O., & Scott, J. M. (2004). Impact of the MTHFR C677T polymorphism on risk of neural tube defects: Case-control study. *BMJ (Clinical Research Ed.)*, *328*(7455), 1535–1536. https://doi.org/10.1136/bmj.38036.646030.EE

Klein, S., Meng, R., Montenarh, M., & Götz, C. (2016). The Phosphorylation of PDX-1 by Protein Kinase CK2 Is Crucial for Its Stability. *Pharmaceuticals (Basel, Switzerland)*, *10*(1), E2. https://doi.org/10.3390/ph10010002

Klinger, J. R., & Kadowitz, P. J. (2017). The Nitric Oxide Pathway in Pulmonary Vascular Disease. *The American Journal of Cardiology*, *120*(8, Supplement), S71–S79. https://doi.org/10.1016/j.amjcard.2017.06.012

Knott, G. J., Bond, C. S., & Fox, A. H. (2016). The DBHS proteins SFPQ, NONO and PSPC1: A multipurpose molecular scaffold. *Nucleic Acids Research*, *44*(9), 3989–4004. https://doi.org/10.1093/nar/gkw271

Kohen, A. (2015). Dihydrofolate reductase as a model for studies of enzyme dynamics and catalysis. *F1000Research*, *4*, F1000 Faculty Rev-1464. https://doi.org/10.12688/f1000research.6968.1

Kolker, E., Higdon, R., & Hogan, J. M. (2006). Protein identification and expression analysis using mass spectrometry. *Trends in Microbiology*, *14*(5), 229–235. https://doi.org/10.1016/j.tim.2006.03.005

Koonin, E. V. (2012). Does the central dogma still stand? *Biology Direct*, *7*(1), 27. https://doi.org/10.1186/1745-6150-7-27

Koury, M. J., Price, J. O., & Hicks, G. G. (2000). Apoptosis in megaloblastic anemia occurs during DNA synthesis by a p53-independent, nucleoside-reversible mechanism. *Blood*, *96*(9), 3249–3255. https://doi.org/10.1182/blood.V96.9.3249

Kremer, J. M. (2004). Toward a better understanding of methotrexate. *Arthritis & Rheumatism*, *50*(5), 1370–1382. https://doi.org/10.1002/art.20278

Krupenko, S. A. (2009). FDH: An aldehyde dehydrogenase fusion enzyme in folate metabolism. *Chemico-Biological Interactions*, *178*(1–3), 84–93. https://doi.org/10.1016/j.cbi.2008.09.007

Laufer, B. I., & Singh, S. M. (2015). Strategies for precision modulation of gene expression by epigenome editing: An overview. *Epigenetics & Chromatin*, *8*(1), 34. https://doi.org/10.1186/s13072-015-0023-7

Lee, D. H. (1998). Proteasome inhibitors: Valuable new tools for cell biologists. *Trends in Cell Biology*, *8*(10), 397–403.

Leenay, R. T., & Beisel, C. L. (2017). Deciphering, Communicating, and Engineering the CRISPR PAM. *Journal of Molecular Biology*, *429*(2), 177–191. https://doi.org/10.1016/j.jmb.2016.11.024

Levanon, E. Y., Hallegger, M., Kinar, Y., Shemesh, R., Djinovic-Carugo, K., Rechavi, G., Jantsch, M. F., & Eisenberg, E. (2005). Evolutionarily conserved human targets of adenosine to inosine RNA editing. *Nucleic Acids Research*, *33*(4), 1162–1168. https://doi.org/10.1093/nar/gki239

Li, H., Yang, Y., Hong, W., Huang, M., Wu, M., & Zhao, X. (2020). Applications of genome editing technology in the targeted therapy of human diseases: Mechanisms, advances and prospects. *Signal Transduction and Targeted Therapy*, *5*(1), 1–23. https://doi.org/10.1038/s41392-019-0089-y

Lieber, M. R. (2010). The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End Joining Pathway. *Annual Review of Biochemistry*, *79*, 181–211. https://doi.org/10.1146/annurev.biochem.052308.093131

Lion, T. (2001). Current recommendations for positive controls in RT-PCR assays. *Leukemia*, *15*(7), 1033–1037. https://doi.org/10.1038/sj.leu.2402133

Loayza-Puch, F., Rooijers, K., Buil, L. C. M., Zijlstra, J., F. Oude Vrielink, J., Lopes, R., Ugalde, A. P., van Breugel, P., Hofland, I., Wesseling, J., van Tellingen, O., Bex, A., & Agami, R. (2016). Tumour-specific proline vulnerability uncovered by differential ribosome codon reading. *Nature*, *530*(7591), 490–494. https://doi.org/10.1038/nature16982

Loayza-Puch, F., Rooijers, K., Zijlstra, J., Moumbeini, B., Zaal, E. A., Oude Vrielink, J. F., Lopes, R., Ugalde, A. P., Berkers, C. R., & Agami, R. (2017). TGFβ1-induced leucine limitation uncovered by differential ribosome codon reading. *EMBO Reports*, *18*(4), 549–557. https://doi.org/10.15252/embr.201744000

Lonn, E., Yusuf, S., Arnold, M. J., Sheridan, P., Pogue, J., Micks, M., McQueen, M. J., Probstfield, J., Fodor, G., Held, C., Genest, J., & Heart Outcomes Prevention Evaluation (HOPE) 2 Investigators. (2006). Homocysteine lowering with folic acid and B vitamins in vascular disease. *The New England Journal of Medicine*, *354*(15), 1567–1577. https://doi.org/10.1056/NEJMoa060900

Luiking, Y. C., Engelen, M. P. K. J., & Deutz, N. E. P. (2010). Regulation of nitric oxide production in health and disease. *Current Opinion in Clinical Nutrition and Metabolic Care*, *13*(1), 97–104. https://doi.org/10.1097/MCO.0b013e328332f99d

Ma, Y., Dammer, E. B., Felsky, D., Duong, D. M., Klein, H.-U., White, C. C., Zhou, M., Logsdon, B. A., McCabe, C., Xu, J., Wang, M., Wingo, T. S., Lah, J. J., Zhang, B., Schneider, J., Allen, M., Wang, X., Ertekin-Taner, N., Seyfried, N. T., De Jager, P. L. (2021). Atlas of RNA editing events affecting protein expression in aged and Alzheimer's disease human brain tissue. *Nature Communications*, *12*(1), 1–16. https://doi.org/10.1038/s41467-021-27204-9

MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., & MacCoss, M. J. (2010). Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*, *26*(7), 966–968. https://doi.org/10.1093/bioinformatics/btq054

Maguire, M., Nield, P. C., Devling, T., Jenkins, R. E., Park, B. K., Polański, R., Vlatković, N., & Boyd, M. T. (2008). MDM2 Regulates Dihydrofolate Reductase Activity through Monoubiquitination. *Cancer Research*, *68*(9), 3232–3242. https://doi.org/10.1158/0008-5472.CAN-07-5271

Makrides, S. C. (2003). Vectors for gene expression in mammalian cells. *New Comprehensive Biochemistry*, *38*, 9–26. https://doi.org/10.1016/S0167-7306(03)38002-0

Mansi, L., Tangaro, M. A., Lo Giudice, C., Flati, T., Kopel, E., Schaffer, A. A., Castrignanò, T., Chillemi, G., Pesole, G., & Picardi, E. (2021). REDIportal: Millions of novel A-to-I RNA editing events from thousands of RNAseq experiments. *Nucleic Acids Research*, *49*(D1), D1012–D1019. https://doi.org/10.1093/nar/gkaa916

Mant, C. T., Chen, Y., Yan, Z., Popa, T. V., Kovacs, J. M., Mills, J. B., Tripet, B. P., & Hodges, R. S. (2007). HPLC Analysis and Purification of Peptides. *Peptide Characterization and Application Protocols*, *386*, 3–55. https://doi.org/10.1007/978-1-59745-430-8_1

Martianov, I., Ramadass, A., Serra Barros, A., Chow, N., & Akoulitchev, A. (2007). Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature*, *445*(7128), 666–670. https://doi.org/10.1038/nature05519

Matera, A. G. (1999). Nuclear bodies: Multifaceted subdomains of the interchromatin space. *Trends in Cell Biology*, *9*(8), 302–309. https://doi.org/10.1016/S0962-8924(99)01606-2

Matthews, D. A., Alden, R. A., Bolin, J. T., Freer, S. T., Hamlin, R., Xuong, N., Kraut, J., Poe, M., Williams, M., & Hoogsteen, K. (1977). Dihydrofolate reductase: X-ray structure of the binary complex with methotrexate. *Science (New York, N.Y.)*, *197*(4302), 452–455. https://doi.org/10.1126/science.17920

Mauldin, R. V., Sapienza, P. J., Petit, C. M., & Lee, A. L. (2012). Structure and Dynamics of the G121V Dihydrofolate Reductase Mutant: Lessons from a Transition-State Inhibitor Complex. *PLoS ONE*, *7*(3), e33252–e33252. https://doi.org/10.1371/journal.pone.0033252

Maurer, B. J., Barker, P. E., & Masters, J. N. (1984). Human dihydrofolate reductase gene is located in chromosome 5 and is unlinked to the related pseudogenes. *ISOTOPENPRAXIS*, *20*(1), 1484–1488.

McDermaid, A., Monier, B., Zhao, J., Liu, B., & Ma, Q. (2019). Interpretation of differential gene expression results of RNA-seq data: Review and integration. *Briefings in Bioinformatics*, *20*(6), 2044–2054. https://doi.org/10.1093/bib/bby067

McEntee, G., Minguzzi, S., O'Brien, K., Ben Larbi, N., Loscher, C., O'Fágáin, C., & Parle-McDermott, A. (2011). The former annotated human pseudogene dihydrofolate reductase-like 1 (DHFRL1) is expressed and functional. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(37), 15157–15162.

McNulty, H., Strain, J. J., Hughes, C. F., Pentieva, K., & Ward, M. (2020). Evidence of a Role for One-Carbon Metabolism in Blood Pressure: Can B Vitamin Intervention Address the Genetic Risk of Hypertension Owing to a Common Folate Polymorphism? *Current Developments in Nutrition*, *4*(1), nzz102. https://doi.org/10.1093/cdn/nzz102

Mehta, P., Woo, P., Venkataraman, K., & Karzai, A. W. (2012). Ribosome purification approaches for studying interactions of regulatory proteins and RNAs with the ribosome. *Methods in Molecular Biology (Clifton, N.J.)*, *905*, 273–289. https://doi.org/10.1007/978-1-61779-949-5_18

Melse-Boonstra, A., de Bree, A., Verhoef, P., Bjørke-Monsen, A. L., & Verschuren, W. M. M. (2002). Dietary monoglutamate and polyglutamate folate are associated with plasma folate concentrations in Dutch men and women aged 20-65 years. *The Journal of Nutrition*, *132*(6), 1307–1312. https://doi.org/10.1093/jn/132.6.1307

Meyer, C., Garzia, A., & Tuschl, T. (2017). Simultaneous detection of the subcellular localization of RNAs and proteins in cultured cells by combined multicolor RNA-FISH and IF. *Methods (San Diego, Calif.)*, *118–119*, 101–110. https://doi.org/10.1016/j.ymeth.2016.09.010

Miettinen, T. P., & Björklund, M. (2015). Mevalonate Pathway Regulates Cell Size Homeostasis and Proteostasis through Autophagy. *Cell Reports*, *13*(11), 2610–2620. https://doi.org/10.1016/j.celrep.2015.11.045

Mitchell, H. K., Snell, E. E., & Williams, R. J. (1941). THE CONCENTRATION OF "FOLIC ACID." *Journal of the American Chemical Society*, *63*(8), 2284–2284. https://doi.org/10.1021/ja01853a512

Molloy, A. M., Pangilinan, F., & Brody, L. C. (2017). Genetic Risk Factors for Folate-Responsive Neural Tube Defects. *Annual Review of Nutrition*, *37*, 269–291. https://doi.org/10.1146/annurev-nutr-071714-034235

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, *5*(7), 621–628. https://doi.org/10.1038/nmeth.1226

Mudd, S. H., Ebert, M. H., & Scriver, C. R. (1980). Labile methyl group balances in the human: The role of sarcosine. *Metabolism - Clinical and Experimental*, *29*(8), 707–720. https://doi.org/10.1016/0026-0495(80)90192-4

Murthy, S., & Reddy, G. P.-V. (2006). Replitase: Complete machinery for DNA synthesis. *Journal of Cellular Physiology*, *209*(3), 711–717. https://doi.org/10.1002/jcp.20842

Myojin, R., Kuwahara, S., Yasaki, T., Matsunaga, T., Sakurai, T., Kimura, M., Uesugi, S., & Kurihara, Y. (2004). Expression and functional significance of mouse paraspeckle protein 1 on spermatogenesis. *Biology of Reproduction*, *71*(3), 926–932. https://doi.org/10.1095/biolreprod.104.028159

Naderi, N., & House, J. D. (2018). Recent Developments in Folate Nutrition. In *Advances in Food and Nutrition Research* (Vol. 83, pp. 195–213). Elsevier. https://doi.org/10.1016/bs.afnr.2017.12.006

Naganuma, T., & Hirose, T. (2013). Paraspeckle formation during the biogenesis of long non-coding RNAs. *RNA Biology*, *10*(3), 456–461. https://doi.org/10.4161/rna.23547

Nakade, S., Tsubota, T., Sakane, Y., Kume, S., Sakamoto, N., Obara, M., Daimon, T., Sezutsu, H., Yamamoto, T., Sakuma, T., & Suzuki, K. T. (2014). Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nature Communications*, *5*(1), 5560. https://doi.org/10.1038/ncomms6560

Nakorchevsky, A., & Yates, J. R. (2012). 1.17 Mass Spectrometry. In E. H. Egelman (Ed.), *Comprehensive Biophysics* (pp. 341–375). Elsevier. https://doi.org/10.1016/B978-0-12-374920-8.00121-1

Nambiar, T. S., Billon, P., Diedenhofen, G., Hayward, S. B., Taglialatela, A., Cai, K., Huang, J.-W., Leuzzi, G., Cuella-Martin, R., Palacios, A., Gupta, A., Egli, D., & Ciccia, A. (2019). Stimulation of CRISPR-mediated homology-directed repair by an engineered RAD18 variant. *Nature Communications*, *10*(1), 3395. https://doi.org/10.1038/s41467-019-11105-z

Nättinen, J., Aapola, U., Nukareddy, P., & Uusitalo, H. (2021). Looking deeper into ocular surface health: An introduction to clinical tear proteomics analysis. *Acta Ophthalmologica*. https://doi.org/10.1111/aos.15059

Nazki, F. H., Sameer, A. S., & Ganaie, B. A. (2014). Folate: Metabolism, genes, polymorphisms and the associated diseases. *Gene*, *533*(1), 11–20. https://doi.org/10.1016/j.gene.2013.09.063

Obeid, R. (2013). The Metabolic Burden of Methyl Donor Deficiency with Focus on the Betaine Homocysteine Methyltransferase Pathway. *Nutrients*, *5*(9), 3481–3495. https://doi.org/10.3390/nu5093481

Oefner, C., Arcy, A. D. ', & Winkler, F. K. (1988). *Crystal structure of human dihydrofolate reductase complexed with folate* (Eur. J. Biochem, Vol. 174, pp. 377–385). https://febs.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1432-1033.1988.tb14108.x

Oliveto, S., Mancino, M., Manfrini, N., & Biffo, S. (2017). Role of microRNAs in translation regulation and cancer. *World Journal of Biological Chemistry*, 8(1), 45–56. https://doi.org/10.4331/wjbc.v8.i1.45

Ouyang, Y., Wu, Q., Li, J., Sun, S., & Sun, S. (2020). S-adenosylmethionine: A metabolite critical to the regulation of autophagy. *Cell Proliferation*, *53*(11), e12891. https://doi.org/10.1111/cpr.12891

Ouyang, Z., Ren, C., Liu, F., An, G., Bo, X., & Shu, W. (2018). The landscape of the A-to-I RNA editome from 462 human genomes. *Scientific Reports*, *8*(1), 12069. https://doi.org/10.1038/s41598-018-30583-7

Oza, J. P., Aerni, H. R., Pirman, N. L., Barber, K. W., ter Haar, C. M., Rogulina, S., Amrofell, M. B., Isaacs, F. J., Rinehart, J., & Jewett, M. C. (2015). Robust production of recombinant phosphoproteins using cell-free protein synthesis. *Nature Communications*, *6*(1), 8168. https://doi.org/10.1038/ncomms9168

Panda, A. C., Martindale, J. L., & Gorospe, M. (2017). Polysome Fractionation to Analyze mRNA Distribution Profiles. *Bio-Protocol*, *7*(3), e2126. https://doi.org/10.21769/BioProtoc.2126

Pangilinan, F., Finlay, E. K., Molloy, A. M., Abaan, H. O., Shane, B., Mills, J. L., Brody, L. C., & Parle-McDermott, A. (2021). A dihydrofolate reductase 2 (DHFR2) variant is associated with risk of neural tube defects in an Irish cohort but not in a United Kingdom cohort. *American Journal of Medical Genetics, Part A*, *185*(4), 1307–1311.

Parker, C. E., Mocanu, V., Mocanu, M., Dicheva, N., & Warren, M. R. (2010). *Mass Spectrometry for Post-Translational Modifications*. CRC Press/Taylor & Francis. http://www.ncbi.nlm.nih.gov/pubmed/21882444

Parle-McDermott, A., Pangilinan, F., Mills, J. L., Kirke, P. N., Gibney, E. R., Troendle, J., O'Leary, V. B., Molloy, A. M., Conley, M., Scott, J. M., & Brody, L. C. (2007). The 19-bp deletion polymorphism in intron-1 of dihydrofolate reductase (DHFR) may decrease rather than increase risk for spina bifida in the Irish population. *American Journal of Medical Genetics Part A*, *143A*(11), 1174–1180. https://doi.org/10.1002/ajmg.a.31725

Pham, P. V. (2018). Chapter 19 - Medical Biotechnology: Techniques and Applications. In D. Barh & V. Azevedo (Eds.), *Omics Technologies and Bio-Engineering* (pp. 449–469). Academic Press. https://doi.org/10.1016/B978-0-12-804659-3.00019-1

Picardi, E., D'Erchia, A. M., Lo Giudice, C., & Pesole, G. (2017). REDIportal: A comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Research*, *45*(Database issue), D750–D757. https://doi.org/10.1093/nar/gkw767

Pieroth, R., Paver, S., Day, S., & Lammersfeld, C. (2018). Folate and Its Impact on Cancer Risk. *Current Nutrition Reports*, *7*(3), 70–84. https://doi.org/10.1007/s13668-018-0237-y

Pisani, G., & Baron, B. (2019). Nuclear paraspeckles function in mediating gene regulatory and apoptotic pathways. *Non-Coding RNA Research*, *4*(4), 128–134. https://doi.org/10.1016/j.ncrna.2019.11.002

Pitt, J. (2009). Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *The Clinical Biochemist. Reviews*, *30*(1). https://pubmed.ncbi.nlm.nih.gov/19224008/

Poria, D. K., & Ray, P. S. (2017). Polysome Analysis. *Bio-Protocol*, *7*(6). https://doi.org/10.21769/BioProtoc.2192

Porter, D. H., Cook, R. J., & Wagner, C. (1985). Enzymatic properties of dimethylglycine dehydrogenase and sarcosine dehydrogenase from rat liver. *Archives of Biochemistry and Biophysics*, *243*(2), 396–407. https://doi.org/10.1016/0003-9861(85)90516-8

Potapov, V., & Ong, J. L. (2017). Examining Sources of Error in PCR by Single-Molecule Sequencing. *PLoS ONE*, *12*(1), e0169774. https://doi.org/10.1371/journal.pone.0169774

Prasannan, P., Pike, S., Peng, K., Shane, B., & Appling, D. R. (2003). Human mitochondrial c1-tetrahydrofolate synthase: gene structure, tissue distribution of the mRNA, and immunolocalization in CHO cells. *The Journal of Biological Chemistry*, *278*(44), 43178–43187. https://doi.org/10.1074/jbc.M304319200

Prasanth, K. V., Prasanth, S. G., Xuan, Z., Hearn, S., Freier, S. M., Bennett, C. F., Zhang, M. Q., & Spector, D. L. (2005). Regulating gene expression through RNA nuclear retention. *Cell*, *123*(2), 249–263. https://doi.org/10.1016/j.cell.2005.08.033

Prelich, G. (2012). Gene Overexpression: Uses, Mechanisms, and Interpretation. *Genetics*, *190*(3), 841–854. https://doi.org/10.1534/genetics.111.136911

Pringle, E. S., McCormick, C., & Cheng, Z. (2019). Polysome Profiling Analysis of mRNA and Associated Proteins Engaged in Translation. *Current Protocols in Molecular Biology*, *125*(1), e79–e79. https://doi.org/10.1002/cpmb.79

Protein Atlas. (2021). *DHFR2 protein expression summary—The Human Protein Atlas*. https://www.proteinatlas.org/ENSG00000178700-DHFR2/summary/rna)

Pusch, W., Flocco, M. T., Leung, S.-M., Thiele, H., & Kostrzewa, M. (2003). Mass spectrometry-based clinical proteomics. *Pharmacogenomics*, *4*(4), 463–476. https://doi.org/10.1517/phgs.4.4.463.22753

Quandt, A., Espona, L., Balasko, A., Weisser, H., Brusniak, M. Y., Kunszt, P., Aebersold, R., & Malmström, L. (2014). Using synthetic peptides to benchmark peptide identification software and search parameters for MS/MS data analysis. *EuPA Open Proteomics*, *5*, 21–31.

Raimondi, M. V., Randazzo, O., La Franca, M., Barone, G., Vignoni, E., Rossi, D., & Collina, S. (2019). DHFR Inhibitors: Reading the Past for Discovering Novel Anticancer Agents. *Molecules*, *24*(6), 1140. https://doi.org/10.3390/molecules24061140

Raj, A., Wang, S. H., Shim, H., Harpak, A., Li, Y. I., Engelmann, B., Stephens, M., Gilad, Y., & Pritchard, J. K. (2016). Thousands of novel translated open reading frames in humans inferred by ribosome footprint profiling. *ELife*, *5*, e13328. https://doi.org/10.7554/eLife.13328

Rashidi, H., Luu, N. T., Alwahsh, S. M., Ginai, M., Alhaque, S., Dong, H., Tomaz, R. A., Vernay, B., Vigneswara, V., Hallett, J. M., Chandrashekran, A., Dhawan, A., Vallier, L., Bradley, M., Callanan, A., Forbes, S. J., Newsome, P. N., & Hay, D. C. (2018). 3D human liver tissue from pluripotent stem cells displays stable phenotype in vitro and supports compromised liver function in vivo. *Archives of Toxicology*, *92*(10), 3117–3129.

Rauniyar, N., Peng, G., Lam, T. T., Zhao, H., Mor, G., & Williams, K. R. (2017). Data-Independent Acquisition and Parallel Reaction Monitoring Mass Spectrometry Identification of Serum Biomarkers for Ovarian Cancer. *Biomarker Insights*, *12*, 1177271917710948. https://doi.org/10.1177/1177271917710948

Reuter, J. S., & Mathews, D. H. (2010). RNAstructure: Software for RNA secondary structure prediction and analysis. BMC Bioinformatics, 11(1), 129. https://doi.org/10.1186/1471-2105-11-129

Riggs, P. (2013). Fusion Protein. In S. Maloy & K. Hughes (Eds.), *Brenner's Encyclopedia of Genetics (Second Edition)* (pp. 134–135). Academic Press. https://doi.org/10.1016/B978-0-12-374984-0.00565-9

Rivera, M. C., Maguire, B., & Lake, J. A. (2015). Isolation of ribosomes and polysomes. *Cold Spring Harbor Protocols*, *2015*(3), 293–299. https://doi.org/10.1101/pdb.prot081331

Rodgers, K., & McVey, M. (2016). Error-prone repair of DNA double-strand breaks. *Journal of Cellular Physiology*, *231*(1), 15–24. https://doi.org/10.1002/jcp.25053

Ronsein, G. E., Pamir, N., von Haller, P. D., Kim, D. S., Oda, M. N., Jarvik, G. P., Vaisar, T., & Heinecke, J. W. (2015). Parallel reaction monitoring (PRM) and selected reaction monitoring (SRM) exhibit comparable linearity, dynamic range and precision for targeted quantitative HDL proteomics. *Journal of Proteomics*, *0*, 388–399. https://doi.org/10.1016/j.jprot.2014.10.017

Roper, C., Pearce, R., Nair, S., Sharp, B., Nosten, F., & Anderson, T. (2004). Intercontinental Spread of Pyrimethamine-Resistant Malaria. *Science*, *305*(5687), 1124–1124. https://doi.org/10.1126/science.1098876

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

Rosenberg, I. H. (2012). A History of the Isolation and Identification of Folic Acid (Folate). *Annals of Nutrition and Metabolism*, *61*(3), 231–235. https://doi.org/10.1159/000343112

Rueter, S. M., Dawson, T. R., & Emeson, R. B. (1999). Regulation of alternative splicing by RNA editing. *Nature*, *399*(6731), 75–80. https://doi.org/10.1038/19992

Said, H. M. (2011). Intestinal absorption of water-soluble vitamins in health and disease. *The Biochemical Journal*, *437*(3), 357–372. https://doi.org/10.1042/BJ20110326

Sakurai, M., Ueda, H., Yano, T., Okada, S., Terajima, H., Mitsuyama, T., Toyoda, A., Fujiyama, A., Kawabata, H., & Suzuki, T. (2014). A biochemical landscape of A-to-I RNA editing in the human brain transcriptome. *Genome Research*, *24*(3), 522–534. https://doi.org/10.1101/gr.162537.113

Sarhan, M. A. A., Shati, A. A., & Elsaid, F. G. (2014). Biochemical and molecular studies on the possible influence of the Brassica oleracea and Beta vulgaris extracts to mitigate the effect of food preservatives and food chemical colorants on albino rats. *Saudi Journal of Biological Sciences*, *21*(4), 342–354. https://doi.org/10.1016/J.SJBS.2013.11.004

Sasaki, Y. T. F., Ideue, T., Sano, M., Mituyama, T., & Hirose, T. (2009). MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(8), 2525–2530. https://doi.org/10.1073/pnas.0807899106

Scaglione, F., & Panzavolta, G. (2014). Folate, folic acid and 5-methyltetrahydrofolate are not the same thing. *Xenobiotica; the Fate of Foreign Compounds in Biological Systems*, *44*(5), 480–488. https://doi.org/10.3109/00498254.2013.845705

Schnell, J. R., Dyson, H. J., & Wright, P. E. (2004). Structure, Dynamics, and Catalytic Function of Dihydrofolate Reductase. *Annual Review of Biophysics and Biomolecular Structure*, *33*(1), 119–140. https://doi.org/10.1146/annurev.biophys.33.110502.133613

Scott, J., Kirke, P., Molloy, A., Daly, L., & Weir, D. (1994). The role of folate in the prevention of neural-tube defects. *The Proceedings of the Nutrition Society*, *53*(3), 631–636. https://doi.org/10.1079/pns19940072

Scotti, M., Stella, L., Shearer, E. J., & Stover, P. J. (2013). Modeling cellular compartmentation in one-carbon metabolism. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, *5*(3), 343–365. https://doi.org/10.1002/wsbm.1209

Sfeir, A., & Symington, L. S. (2015). Microhomology-mediated end joining: A back-up survival mechanism or dedicated pathway? *Trends in Biochemical Sciences*, *40*(11), 701–714. https://doi.org/10.1016/j.tibs.2015.08.006

Shang, F., & Taylor, A. (2011). Ubiquitin-proteasome pathway and cellular responses to oxidative stress. *Free Radical Biology & Medicine*, *51*(1), 5–16. https://doi.org/10.1016/j.freeradbiomed.2011.03.031

Sharma, R. (2013). Biochemical Mechanisms of Fatty Liver and Bioactive Foods: Fatty Liver, Diagnosis, Nutrition Therapy. *Bioactive Food as Dietary Interventions for Liver and Gastrointestinal Disease*, 623–655. https://doi.org/10.1016/B978-0-12-397154-8.00181-0

Shen, C., Sheng, Q., Dai, J., Li, Y., Zeng, R., & Tang, H. (2009). On the estimation of false positives in peptide identifications using decoy search strategy. *Proteomics*, *9*(1), 194–204. https://doi.org/10.1002/pmic.200800330

Shen, Y., Tolić, N., Purvine, S. O., & Smith, R. D. (2012). Improving Collision Induced Dissociation (CID), High Energy Collision Dissociation (HCD), and Electron Transfer Dissociation (ETD) Fourier Transform MS/MS Degradome–Peptidome Identifications Using High Accuracy Mass Information. *Journal of Proteome Research*, *11*(2), 668–677. https://doi.org/10.1021/pr200597j

Shi, Y., Lee, J.-H., Kang, H., & Jiang, H. (2022). A Two-Part Mixed Model for Differential Expression Analysis in Single-Cell High-Throughput Gene Expression Data. *Genes*, *13*(2), 377. https://doi.org/10.3390/genes13020377

Shields, D. C., Kirke, P. N., Mills, J. L., Ramsbottom, D., Molloy, A. M., Burke, H., Weir, D. G., Scott, J. M., & Whitehead, A. S. (1999). The "thermolabile" variant of methylenetetrahydrofolate reductase and neural tube defects: An evaluation of genetic risk and the relative importance of the genotypes of the embryo and the mother. *American Journal of Human Genetics*, *64*(4), 1045–1055.

Shimada, T., Chen, M. J., & Nienhuis, A. W. (1984). A human dihydrofolate reductase intronless pseudogene with an Alu repetitive sequence: Multiple DNA insertions at a single chromosomal site. *Gene*, *31*(1–3), 1–8.

Sidrauski, C., McGeachy, A. M., Ingolia, N. T., & Walter, P. (2015). The small molecule ISRIB reverses the effects of eIF2α phosphorylation on translation and stress granule assembly. *ELife*, *4*, e05033. https://doi.org/10.7554/eLife.05033

Simile, M. M., Latte, G., Feo, C. F., Feo, F., Calvisi, D. F., & Pascale, R. M. (2018). Alterations of methionine metabolism in hepatocarcinogenesis: The emergent role of glycine N-methyltransferase in liver injury. *Annals of Gastroenterology*, *31*(5), 552–560. https://doi.org/10.20524/aog.2018.0288

Skacel, N., Menon, L. G., Mishra, P. J., Peters, R., Banerjee, D., Bertino, J. R., & Abali, E. E. (2005). Identification of Amino Acids Required for the Functional Up-regulation of Human Dihydrofolate Reductase Protein in Response to Antifolate Treatment. *Journal of Biological Chemistry*, *280*(24), 22721–22731. https://doi.org/10.1074/jbc.M500277200

Skierka, K., Wilamowski, P., Wielechowska, M., Cysewski, D., Senkara, E., Wińska, P., Bretner, M., & Cieśla, J. (2019). Human dihydrofolate reductase is a substrate of protein kinase CK2α. *Biochemical and Biophysical Research Communications*, *513*(2), 368–373. https://doi.org/10.1016/j.bbrc.2019.03.186

Slansky, J. E., & Farnham, P. J. (1996). Introduction to the E2F family: Protein structure and gene regulation. *Current Topics in Microbiology and Immunology*, *208*, 1–30. https://doi.org/10.1007/978-3-642-79910-5_1

Smith, R. W. (2013). Mass Spectrometry. In J. A. Siegel, P. J. Saukko, & M. M. Houck (Eds.), *Encyclopedia of Forensic Sciences (Second Edition)* (pp. 603–608). Academic Press. https://doi.org/10.1016/B978-0-12-382165-2.00250-6

Smithells, R. W., Sheppard, S., & Schorah, C. J. (1976). Vitamin dificiencies and neural tube defects. *Archives of Disease in Childhood*, *51*(12), 944–950. https://doi.org/10.1136/adc.51.12.944

Sørensen, H. P., & Mortensen, K. K. (2005). Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microbial Cell Factories*, *4*(1), 1. https://doi.org/10.1186/1475-2859-4-1

Spizzichino, S., Boi, D., Boumis, G., Lucchi, R., Liberati, F. R., Capelli, D., Montanari, R., Pochetti, G., Piacentini, R., Parisi, G., Paone, A., Rinaldo, S., Contestabile, R., Tramonti, A., Paiardini, A., Giardina, G., & Cutruzzolà, F. (2021). Cytosolic localization and in vitro assembly of human de novo thymidylate synthesis complex. *The FEBS Journal*, *n/a*(n/a). https://doi.org/10.1111/febs.16248

Stanger, O. (2002). Physiology of folic acid in health and disease. *Current Drug Metabolism*, *3*(2), 211–223. https://doi.org/10.2174/1389200024605163

St-Denis, N., Gabriel, M., Turowec, J. P., Gloor, G. B., Li, S. S.-C., Gingras, A.-C., & Litchfield, D. W. (2015). Systematic investigation of hierarchical phosphorylation by protein kinase CK2. *Journal of Proteomics*, *118*, 49–62. https://doi.org/10.1016/j.jprot.2014.10.020

Stoevesandt, O., Taussig, M. J., & He, M. (2014). Protein microarrays: High-throughput tools for proteomics. *Expert Review of Proteomics*, *6*(2), 145–157. https://doi.org/10.1586/epr.09.2

Stover, P. J. (2009). One-Carbon Metabolism–Genome Interactions in Folate-Associated Pathologies. *The Journal of Nutrition*, *139*(12), 2402–2405. https://doi.org/10.3945/jn.109.113670

Stover, P. J. (2011). Polymorphisms in 1-carbon metabolism, epigenetics and folate-related pathologies. *Journal of Nutrigenetics and Nutrigenomics*, *4*(5), 293–305. https://doi.org/10.1159/000334586

Stover, P. J., & Field, M. S. (2011). Trafficking of Intracellular Folates. *Advances in Nutrition*, *2*(4), 325–331. https://doi.org/10.3945/an.111.000596

Sunwoo, H., Dinger, M. E., Wilusz, J. E., Amaral, P. P., Mattick, J. S., & Spector, D. L. (2009). MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Research*, *19*(3), 347–359. https://doi.org/10.1101/gr.087775.108

Sutandy, F. X. R., Qian, J., Chen, C., & Zhu, H. (2013). Overview of Protein Microarrays. *Current Protocols in Protein Science*, *72*(1), 2711–27116. https://doi.org/10.1002/0471140864.ps2701s72

Tai, N., Schmitz, J. C., Chen, T., & Chu, E. (2004). Characterization of a cis-acting regulatory element in the protein-coding region of human dihydrofolate reductase mRNA. *The Biochemical Journal*, *378*(Pt 3), 999–1006. https://doi.org/10.1042/BJ20031396

Tam, C., O'Connor, D., & Koren, G. (2012). Circulating Unmetabolized Folic Acid: Relationship to Folate Status and Effect of Supplementation. *Obstetrics and Gynecology International*, *2012*, e485179. https://doi.org/10.1155/2012/485179

Tang, L., Swezey, R. R., Green, C. E., & Mirsalis, J. C. (2022). Enhancement of sensitivity and quantification quality in the LC-MS/MS measurement of large biomolecules with sum of MRM (SMRM). *Analytical and Bioanalytical Chemistry*, *414*(5), 1933–1947. https://doi.org/10.1007/s00216-021-03829-z

Tarze, A., Deniaud, A., Le Bras, M., Maillier, E., Molle, D., Larochette, N., Zamzami, N., Jan, G., Kroemer, G., & Brenner, C. (2007). GAPDH, a novel regulator of the pro-apoptotic mitochondrial membrane permeabilization. *Oncogene*, *26*(18), 2606–2620. https://doi.org/10.1038/sj.onc.1210074

Terpe, K. (2006). Overview of bacterial expression systems for heterologous protein production: From molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology*, *72*(2), 211–222. https://doi.org/10.1007/s00253-006-0465-8

Thomas, E. (2001). Pseudogene. *Encyclopedia of Genetics*, 1579–1580. https://doi.org/10.1006/RWGN.2001.1048

Tian, H., & Cronstein, B. (2007). Understanding the mechanisms of action of methotrexate: Implications for the treatment of rheumatoid arthritis. *Bulletin of the NYU Hospital for Joint Diseases*, *65*(3), 168–173.

Tian, T., Yang, K.-Q., Cui, J.-G., Zhou, L.-L., & Zhou, X.-L. (2017). Folic Acid Supplementation for Stroke Prevention in Patients With Cardiovascular Disease. *The American Journal of the Medical Sciences*, *354*(4), 379–387. https://doi.org/10.1016/j.amjms.2017.05.020

Tibbetts, A. S., & Appling, D. R. (2010). Compartmentalization of Mammalian Folate-Mediated One-Carbon Metabolism. *Annual Review of Nutrition*, *30*(1), 57–81. https://doi.org/10.1146/annurev.nutr.012809.104810

Tirosh, O., Cohen, Y., Shitrit, A., Shani, O., Le-Trilling, V. T. K., Trilling, M., Friedlander, G., Tanenbaum, M., & Stern-Ginossar, N. (2015). The Transcription and Translation Landscapes during Human Cytomegalovirus Infection Reveal Novel Host-Pathogen Interactions. *PLoS Pathogens*, *11*(11), e1005288. https://doi.org/10.1371/journal.ppat.1005288

Tjong, E., Dimri, M., & Mohiuddin, S. S. (2022). Biochemistry, Tetrahydrofolate. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK539712/

Tousoulis, D., Kampoli, A.-M., Tentolouris, C., Papageorgiou, N., & Stefanadis, C. (2012). The role of nitric oxide on endothelial function. *Current Vascular Pharmacology*, *10*(1), 4–18. https://doi.org/10.2174/157016112798829760

Tripathi, N. K., & Shrivastava, A. (2019). Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. *Frontiers in Bioengineering and Biotechnology*, *7*, 420. https://doi.org/10.3389/fbioe.2019.00420

Tyers, M., & Mann, M. (2003). From genomics to proteomics. *Nature*, *422*(6928), 193–197. https://doi.org/10.1038/nature01510

Uzman, A. (2001). Molecular Cell Biology (4th edition): Harvey Lodish, Arnold Berk, S. Lawrence Zipursky, Paul Matsudaira, David Baltimore and James Darnell; Freeman & Co., New York, NY, 2000, 1084 pp., list price $102.25, ISBN 0-7167-3136-3. https://doi.org/10.1016/S1470-8175(01)00023-6

Välikangas, T., Suomi, T., & Elo, L. L. (2018). A comprehensive evaluation of popular proteomics software workflows for label-free proteome quantification and imputation. *Briefings in Bioinformatics*, *19*(6), 1344–1355. https://doi.org/10.1093/bib/bbx054

Van der Put, N. M., Steegers-Theunissen, R. P., Frosst, P., Trijbels, F. J., Eskes, T. K., van den Heuvel, L. P., Mariman, E. C., den Heyer, M., Rozen, R., & Blom, H. J. (1995). Mutated methylenetetrahydrofolate reductase as a risk factor for spina bifida. *Lancet (London, England)*, *346*(8982), 1070–1071. https://doi.org/10.1016/s0140-6736(95)91743-8

Vandenbrouck, Y., Lane, L., Carapito, C., Duek, P., Rondel, K., Bruley, C., Macron, C., Gonzalez De Peredo, A., Couté, Y., Chaoui, K., Com, E., Gateau, A., Hesse, A. M., Marcellin, M., Méar, L., Mouton-Barbosa, E., Robin, T., Burlet-Schiltz, O., Cianferani, S., Pineau, C. (2016). Looking for Missing Proteins in the Proteome of Human Spermatozoa: An Update. *Journal of Proteome Research*, *15*(11), 3998–4019.

Victoria Wang, X., Blades, N., Ding, J., Sultana, R., & Parmigiani, G. (2012). Estimation of sequencing error rates in short reads. *BMC Bioinformatics*, *13*(1), 185. https://doi.org/10.1186/1471-2105-13-185

Vidova, V., & Spacil, Z. (2017). A review on mass spectrometry-based quantitative proteomics: Targeted and data independent acquisition. *Analytica Chimica Acta*, *964*, 7–23. https://doi.org/10.1016/j.aca.2017.01.059

Visentin, M., Diop-Bove, N., Zhao, R., & Goldman, I. D. (2014). The Intestinal Absorption of Folates. *Annual Review of Physiology*, *76*, 251–274. https://doi.org/10.1146/annurev-physiol-020911-153251

Wagner, G. P., Kin, K., & Lynch, V. J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in Biosciences*, *131*(4), 281–285. https://doi.org/10.1007/s12064-012-0162-3

Wagner, R. W., Smith, J. E., Cooperman, B. S., & Nishikura, K. (1989). A double-stranded RNA unwinding activity introduces structural alterations by means of adenosine to inosine conversions in mammalian cells and Xenopus eggs. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(8), 2647–2651. https://doi.org/10.1073/pnas.86.8.2647

Wang, F. K., Koch, J., & Stokstad, E. L. (1967). Folate coenzyme pattern, folate linked enzymes and methionine biosynthesis in rat liver mitochondria. *Biochemische Zeitschrift*, *346*(5), 458–466.

Wang, H., Yang, L., Wang, Y., Chen, L., Li, H., & Xie, Z. (2019). RPFdb v2.0: An updated database for genome-wide information of translated mRNA generated from ribosome profiling. *Nucleic Acids Research*, *47*(D1), D230–D234. https://doi.org/10.1093/nar/gky978

Wang, P., & Wilson, S. R. (2013). Mass spectrometry-based protein identification by integrating de novo sequencing with database searching. *BMC Bioinformatics*, *14*(Suppl 2), S24–S24.

Wang, Y., & Chen, L.-L. (2020). Organization and function of paraspeckles. *Essays in Biochemistry*, *64*(6), 875–882. https://doi.org/10.1042/EBC20200010

Warzyszynska, J. E., & Kim, Y.-I. J. (2014). Folate in Human Health and Disease. In *ELS*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470015902.a0002268.pub2

Whitehead, A. S., Gallagher, P., Mills, J. L., Kirke, P. N., Burke, H., Molloy, A. M., Weir, D. G., Shields, D. C., & Scott, J. M. (1995). A genetic defect in 5,10 methylenetetrahydrofolate reductase in neural tube defects. *QJM: Monthly Journal of the Association of Physicians*, *88*(11), 763–766.

Wiita, A. P., Ziv, E., Wiita, P. J., Urisman, A., Julien, O., Burlingame, A. L., Weissman, J. S., & Wells, J. A. (2013). Global cellular response to chemotherapy-induced apoptosis. *ELife*, *2*, e01236. https://doi.org/10.7554/eLife.01236

Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J.-C., Yan, J. X., Gooley, A. A., Hughes, G., Humphery-Smith, I., Williams, K. L., & Hochstrasser, D. F. (1996). From Proteins to Proteomes: Large Scale Protein Identification by Two-Dimensional Electrophoresis and Arnino Acid Analysis. *Bio/Technology*, *14*(1), 61–65. https://doi.org/10.1038/nbt0196-61

Williams, S. R., Yang, Q., Chen, F., Liu, X., Keene, K. L., Jacques, P., Chen, W.-M., Weinstein, G., Hsu, F.-C., Beiser, A., Wang, L., Bookman, E., Doheny, K. F., Wolf, P. A., Zilka, M., Selhub, J., Nelson, S., Gogarten, S. M., Worrall, B. B., Study, the F. H. (2014). Genome-Wide Meta-Analysis of Homocysteine and Methionine Metabolism Identifies Five One Carbon Metabolism Loci and a Novel Association of ALDH1L1 with Ischemic Stroke. *PLoS Genetics*, *10*(3), e1004214–e1004214. https://doi.org/10.1371/JOURNAL.PGEN.1004214

Wills, L. (1931). TREATMENT OF "PERNICIOUS ANAEMIA OF PREGNANCY" AND "TROPICAL ANAEMIA." *British Medical Journal*, *1*(3676), 1059–1064.

Wingfield, P. T. (2015). Overview of the Purification of Recombinant Proteins. *Current Protocols in Protein Science, 80*, 6.1.1-6.1.35. https://doi.org/10.1002/0471140864.ps0601s80

Woeller, C. F., Anderson, D. D., Szebenyi, D. M. E., & Stover, P. J. (2007). Evidence for Small Ubiquitin-like Modifier-dependent Nuclear Import of the Thymidylate Biosynthesis Pathway. *Journal of Biological Chemistry*, *282*(24), 17623–17631. https://doi.org/10.1074/jbc.M702526200

World Health Organization. (2005). *Vitamin and mineral requirements in human nutrition*. World Health Organization. https://apps.who.int/iris/handle/10665/42716

World Health Organization. (2015). Guideline: Optimal serum and red blood cell folate concentrations in women of reproductive age for prevention of neural tube defects. World Health Organization. https://apps.who.int/iris/handle/10665/161988

World Health Organization. (2020). *The top 10 causes of death*. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

Xie, S.-Q., Nie, P., Wang, Y., Wang, H., Li, H., Yang, Z., Liu, Y., Ren, J., & Xie, Z. (2016). RPFdb: A database for genome wide information of translated mRNA generated from ribosome profiling. *Nucleic Acids Research*, *44*(Database issue), D254–D258. https://doi.org/10.1093/nar/gkv972

Yablonovitch, A. L., Deng, P., Jacobson, D., & Li, J. B. (2017). The evolution and adaptation of A-to-I RNA editing. *PLOS Genetics*, *13*(11), e1007064. https://doi.org/10.1371/journal.pgen.1007064

Yates, J. R. (1998). Mass spectrometry and the age of the proteome. *Journal of Mass Spectrometry: JMS*, *33*(1), 1–19. https://doi.org/10.1002/(SICI)1096-9888(199801)33:1<1::AID-JMS624>3.0.CO;2-9

Yeung, S. J., Chen, S. H., & Chan, L. (1996). Ubiquitin-proteasome pathway mediates intracellular degradation of apolipoprotein B. *Biochemistry*, *35*(43), 13843–13848. https://doi.org/10.1021/bi9618777

Zarai, Y., Margaliot, M., & Tuller, T. (2016). On the Ribosomal Density that Maximizes Protein Translation Rate. *PLOS ONE*, *11*(11), e0166481. https://doi.org/10.1371/journal.pone.0166481

Zarou, M. M., Vazquez, A., & Vignir Helgason, G. (2021). Folate metabolism: A re-emerging therapeutic target in haematological cancers. *Leukemia*, *35*(6), 1539–1551. https://doi.org/10.1038/s41375-021-01189-2

Zempleni, J., Suttie, J. W., III, J. F. G., & Stover, P. J. (2013). *Handbook of Vitamins*. CRC Press.

Zhang, G., Annan, R. S., Carr, S. A., & Neubert, T. A. (2010). Overview of Peptide and Protein Analysis by Mass Spectrometry. *Current Protocols in Protein Science*, *62*(1). https://doi.org/10.1002/0471140864.ps1601s62

Zhang, Y., Fonslow, B. R., Shan, B., Baek, M.-C., & Yates, J. R. (2013). Protein Analysis by Shotgun/Bottom-up Proteomics. *Chemical Reviews*, *113*(4), 2343–2394. https://doi.org/10.1021/cr3003533

Zhang, Z., & Carmichael, G. G. (2001). The fate of dsRNA in the nucleus: A p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. *Cell*, *106*(4), 465–475. https://doi.org/10.1016/s0092-8674(01)00466-4

Zhu, F., Nair, R. R., Fisher, E. M. C., & Cunningham, T. J. (2019). Humanising the mouse genome piece by piece. *Nature Communications*, *10*(1), 1845. https://doi.org/10.1038/s41467-019-09716-7

Zinshteyn, B., & Nishikura, K. (2009). Adenosine-to-inosine RNA editing. *Wiley Interdisciplinary Reviews. Systems Biology and Medicine*, *1*(2), 202–209. https://doi.org/10.1002/wsbm.10

Appendices

Appendix A *DHFR2_202 gBlock Sequence. The DHFR2_202 gBlock sequence sourced from IDT. This sequence is identical to a section of Ensembl's version of the DHFR2-202 transcript (Transcript ID: ENST00000394221.2); exon 1 (grey) and exon 2 (blue).*

5'ATCCCAGCACCCTCCCTCGCCACCTCGGCCGCGTCCAGAAGCGTCTCATTCAGATAGAGGTACCCAGCC
GGGCCCCAGGTCTCACCTGGGTCAGCATTGTCCCGCTGAGAGCTTTCAAACAACAGGTGTACGCGAATC
AGTATCAAAAAGCCTCCAGGGAACTGAAAAAAACAGTCTGACTTCCGGGATTCGCGGGCATCCACTTTC
ATTCTCACGTTTCCTCTTCAATTTCTCGCGAGATACTGATTTGCTCGCGCTATAGAATTCTTGTCGAAGAGG
CTTAGCTCCCGCGGGAGATTTCGTGCGCAGTTACTTCCGGTAGCTGGTAAAGGCTGATACTTCCCAGGAC
GCGGAGGCACGTAAAAAAATTTGAAGAAGGGAATTTCGCGGCATTCTTGGCCTGGCTTCCTGGCGTAG
CCAGCAAGTTCGGAGGTGTTAACCGCTGCTGTCATGTTTCTTTTGCTAAACTGCATCGTCGCTGTGTCCCA
AAACATGGGCATCGGCAAGAACGGGGACCTGCCCAGGCCGCCGCTCAGGAATGAATTCAGGTATTTCCA
GAGAATGACCACAACTTCTTCAGTAGA3'

Appendix B *Information on 23 samples chosen for Ribo-seq analysis.*

Study Number	Sample Number	Cell or Tissue Type	Cell or Tissue Source	Reference
GSE82232	SRX1818903	A549	Adenocarcinomic alveolar basal epithelial	(Gonzalez et al., 2014)
GSE51424	SRX690808	Brain normal	Brain normal	(Gonzalez et al., 2014)
GSE51424	SRX690809	Brain tumour	Brain tumour	(Gonzalez et al., 2014)
GSE77292	SRX1549421	CN34	Pleural fluid from breast carcinoma	(Goodarzi et al., 2016)
GSE69906	SRX1059593	Fibroblasts	Fibroblasts	(Tirosh et al., 2015)
GSE75290	SRX1447296	GM19204	B-Lymphocyte	(Raj et al., 2016)
GSE75290	SRX1447297	GM19238	B-Lymphocyte	(Raj et al., 2016)
GSE58207	SRX569214	HCT116	Colorectal carcinoma	(Crappé et al., 2015)
GSE73136	SRX1254413	HEK293	Embryonic kidney	(Calviello et al., 2016)
GSE65778	SRX870804	HEK293T	Embryonic kidney	(Sidrauski et al., 2015)
ERA390450	ERX635825	HeLa	Cervical carcinoma	(Miettinen & Björklund, 2015)
GSE100007	SRX2916238	ESC	Embryonic stem sells	(Blair et al., 2017)
GSE59820	SRX663295	Kidney normal	Kidney normal	(Loayza-Puch et al., 2016)
GSE59820	SRX663298	Kidney tumour	Kidney tumour	(Loayza-Puch et al., 2016)
GSE69922	SRX1059908	MCF10A	Mammary epithelial	(Loayza-Puch et al., 2017)
GSE77317	SRX1550636	MDA	Mammary epithelial carcinoma	(Goodarzi et al., 2016)
GSE48785	SRX320144	MM1S	Myeloma	(Wiita et al., 2013)
GSE63570	SRX765942	NCCIT	Pluripotent embryonal Carcinoma	(Grow et al., 2015)
GSE100007	SRX2916242	NPC	Neural progenitor cells	(Blair et al., 2017)
GSE74510	SRX1402243	PC3	Prostate carcinoma	(Loayza-Puch et al., 2016)
GSE75947	SRX1478977	SHSY5Y	Neuroblastoma	(Chun et al., 2016)
GSE69921	SRX1059905	SUM159	Mesenchymal triple-negative breast cancer	(Loayza-Puch et al., 2017)
GSE39561	SRX170358	THP1	Acute monocytic leukaemia	(Fritsch et al., 2012)

Appendix C *DHFR2_202 Sanger Sequencing results for the Exon 1 assay analysed using Snapgene. The electropherogram data from the DHFR2_202 Sanger Sequencing results for Exon 1 assay, (A) DHFR2_202 Exon 1 forward primer and (B) DHFR2_202 Exon 1 reverse primer. The electropherograms have been clipped at the beginning to remove any unresolved peaks, a common result from DNA sequencing. The sequencing results align back to Exon 1 of the DHFR2_202 transcript that was predicted by Ensembl.*

(A)



(B)



Appendix D *DHFR2_202 Sanger Sequencing results for the Exon 1-2 assay analysed using Snapgene. The electropherogram data from the DHFR2_202 Sanger Sequencing results for Exon 1-2 assay, (A) DHFR2_202 Exon 1-2 forward primer and (B) DHFR2_202 Exon 1-2 reverse primer. The electropherograms have been clipped at the beginning to remove any unresolved peaks, a common result from DNA sequencing. The sequencing results align back to Exon 1-2 of the DHFR2_202 transcript that was predicted by Ensembl.*

(A)

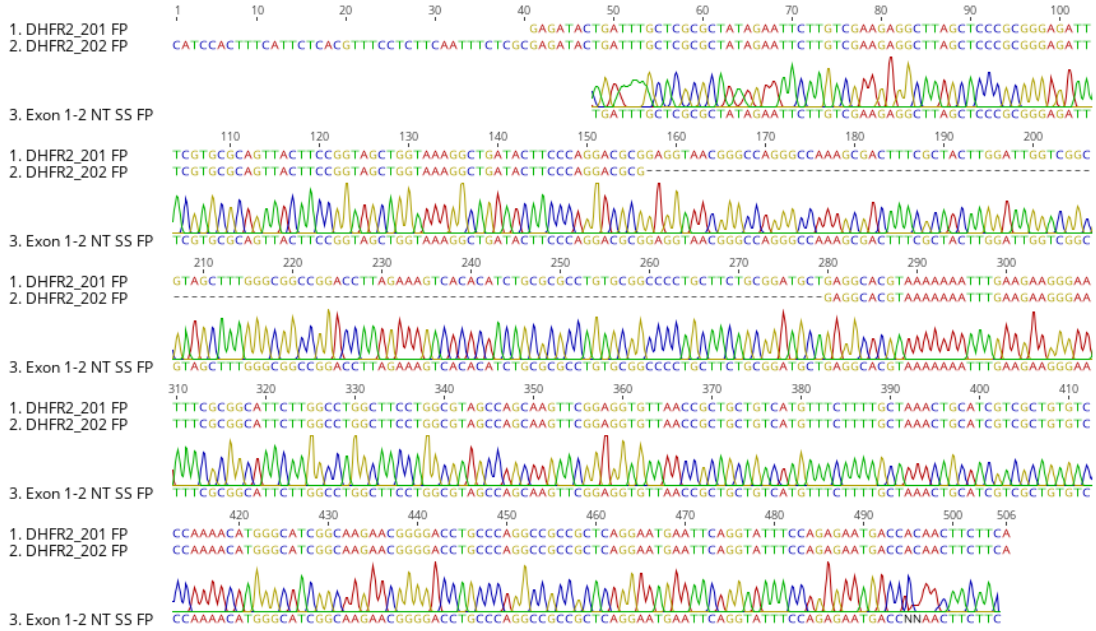


(B)

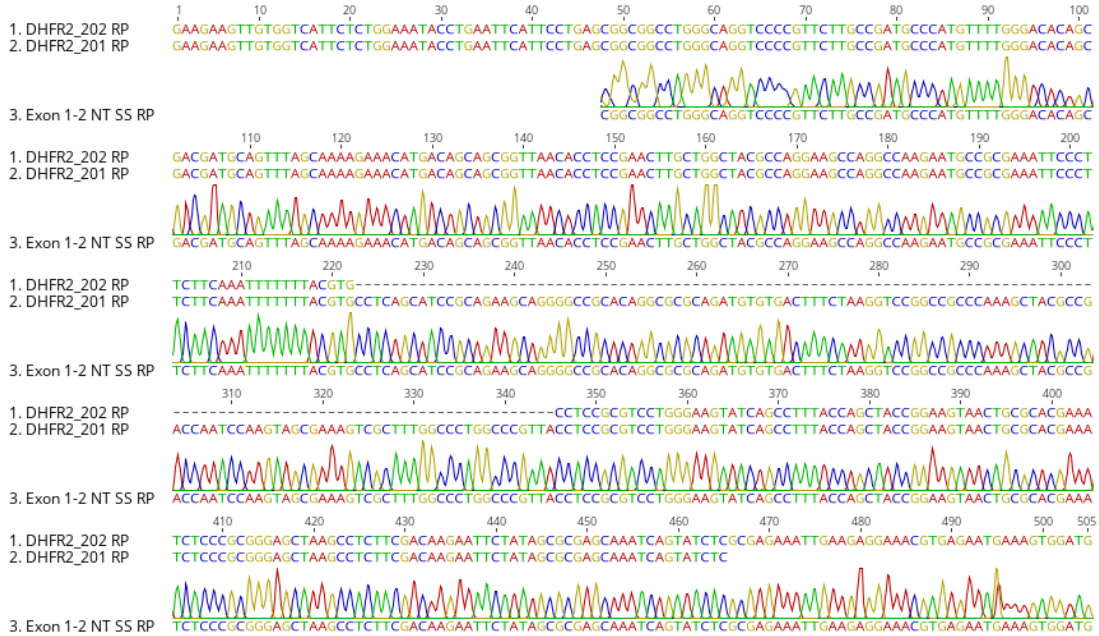


Appendix E *DHFR2_202 Sanger Sequencing results for the Exon 1-2 assay additional band. The electropherogram data from the DHFR2_202 Sanger Sequencing results for Exon 1-2 assay additional band, (A) DHFR2_202 Exon 1-2 forward primer and (B) DHFR2_202 Exon 1-2 reverse primer. The electropherograms have been clipped at the beginning to remove any unresolved peaks, a common result from DNA sequencing. The sequencing results align back to Exon 1-2 of the DHFR2_201 transcript.*

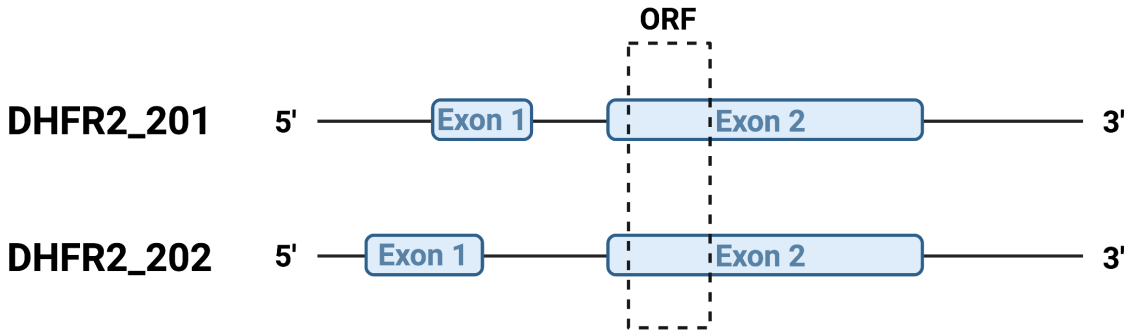
(A)



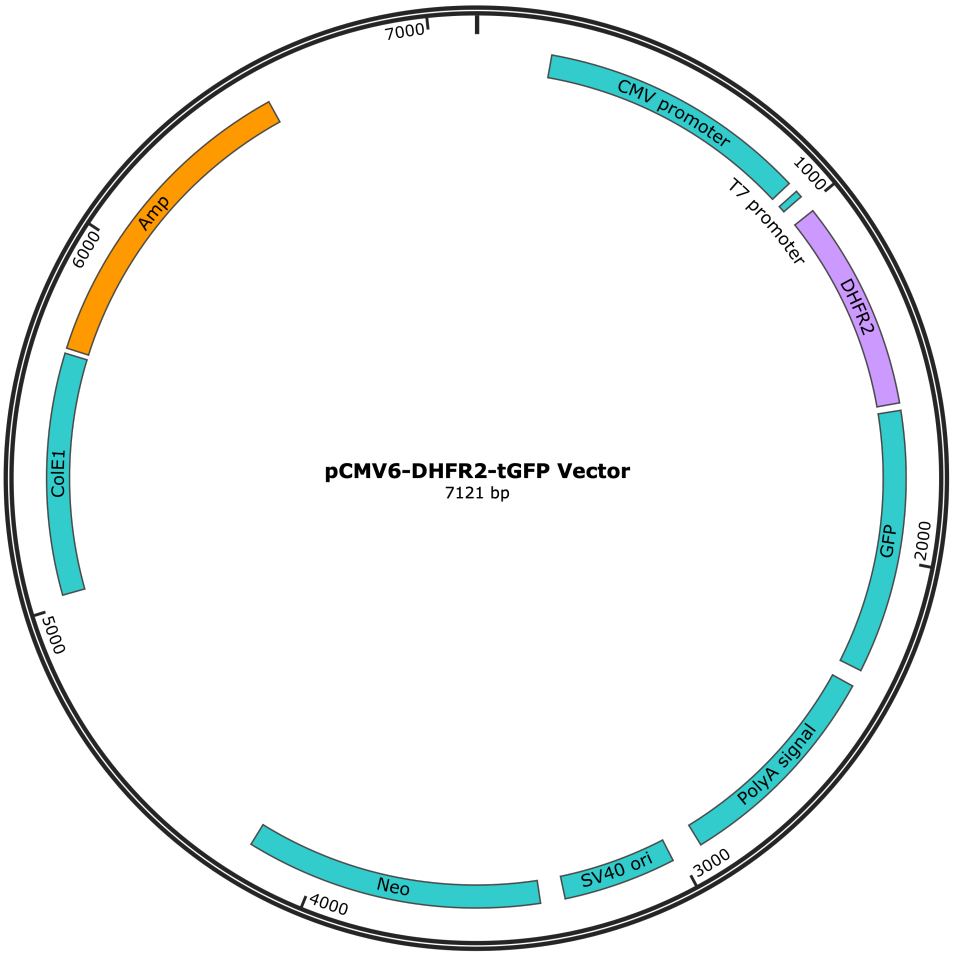
(B)



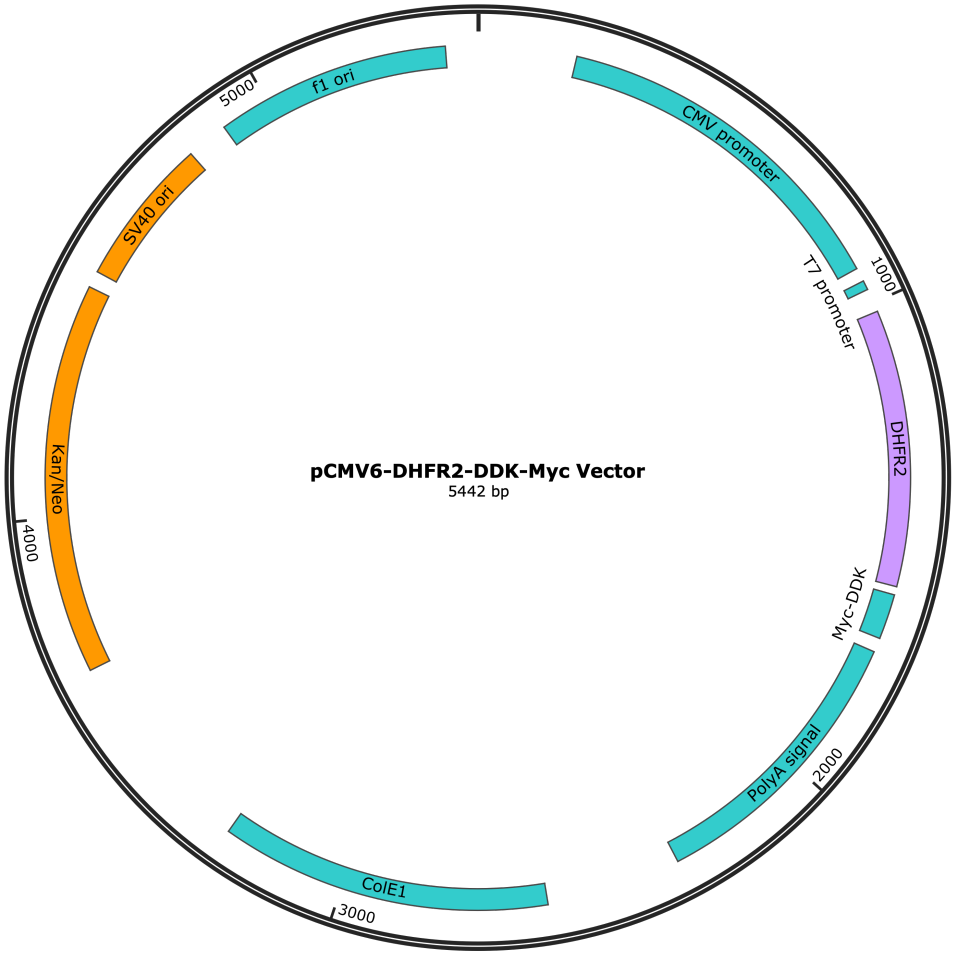
Appendix F *Illustration of the DHFR2_201 and DHFR2_202 mRNA isoforms as annotated by Ensembl. Created with BioRender.com, Adapted from ensembl.org/index.html.*



Appendix G *Vector Map of the pCMV6-AC-DHFR2-tGFP expression vector.*



Appendix H *Vector Map of the pCMV6-AC-DHFR2-DDK-Myc expression vector.*



Appendix I *Sequences for the Alu elements located in the 3'UTR of the DHFR2 transcripts, obtained from REDIportal V2.0 (http://srv00.recas.ba.infn.it/atlas/).*

- **AluSp_Alu Sequence (+ strand): chr3:94059782..94060076 (295 bp) – hg38**

  GCCAAGTGCAGTGGCTGATGCCTGTAATCCCAGCACCCCTGAAGGCCGAGGCAGGAGGATCACC
  TGAGGTCAGGAGTTCGAGACTAGCCTGACCAACATGGAATGTCCCCACCAAAAATACAAAATTAG
  CCTGGCGTGGTGGCACGCGCCTGTAGTCCCAGCCACCCAGGAGCCTGAGGCAGGAGAATTGCTT
  GAACCTGGGAAGGCTGAAGTTGCCGTGAGCCAAGGTTGTTCCATTGCACTCCGCCTGGGCAACA
  AGAGTGAAATTCTGTCTTAAAAAGAAAAAAAAAAAAA

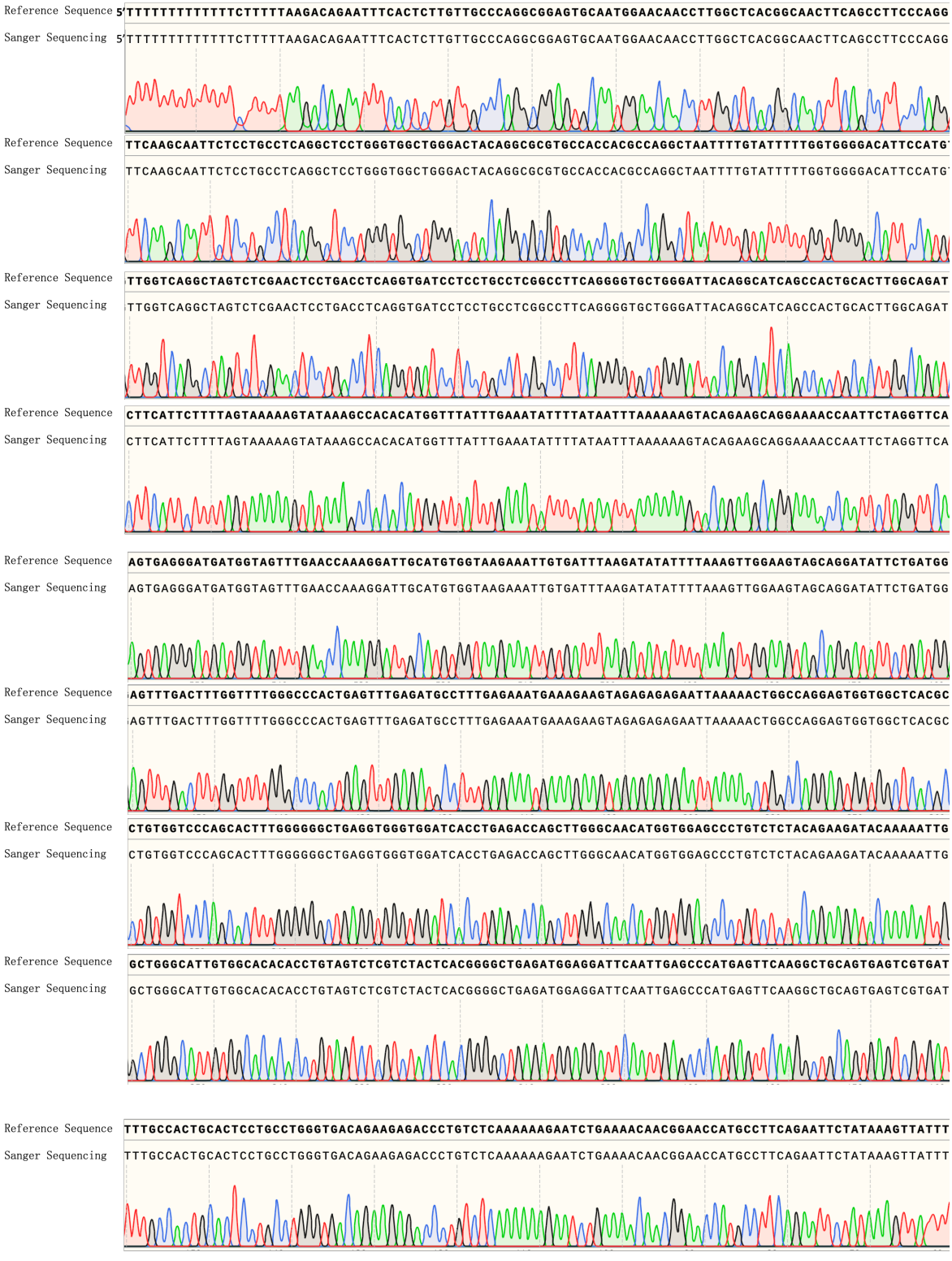
- **AluJb_Alu Sequence (- strand): chr3:94059224..94059499 (276 bp) – hg38**

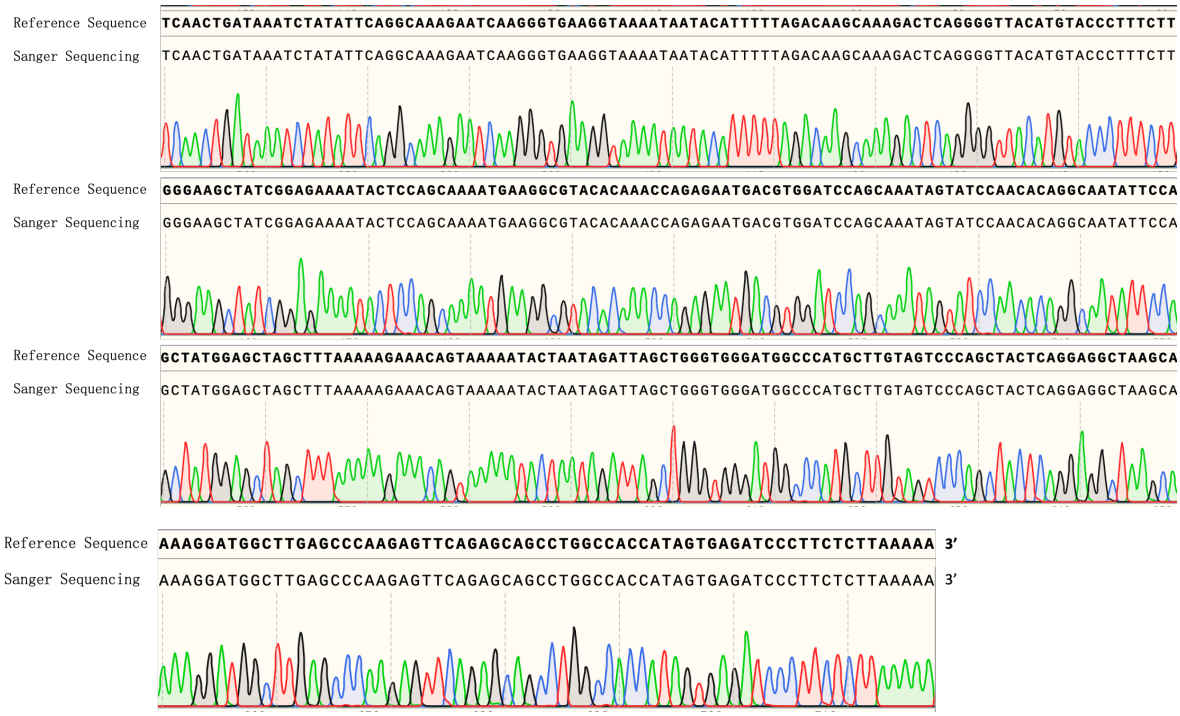
  GGCCAGGAGTGGTGGCTCACGCCTGTGGTCCCAGCACTTTGGGGGGCTGAGGTGGGTGGATCAC
  CTGAGACCAGCTTGGGCAACATGGTGGAGCCCTGTCTCTACAGAAGATACAAAAATTGGCTGGG
  CATTGTGGCACACACCTGTAGTCTCGTCTACTCACGGGGCTGAGATGGAGGATTCAATTGAGCCC
  ATGAGTTCAAGGCTGCAGTGAGTCGTGATTTTGCCACTGCACTCCTGCCTGGGTGACAGAAGAGA
  CCCTGTCTCAAAAAAGAA

- **FLAM_A_Alu Sequence (- strand): *chr3:94058810..94058929 (120 bp)* – hg38**

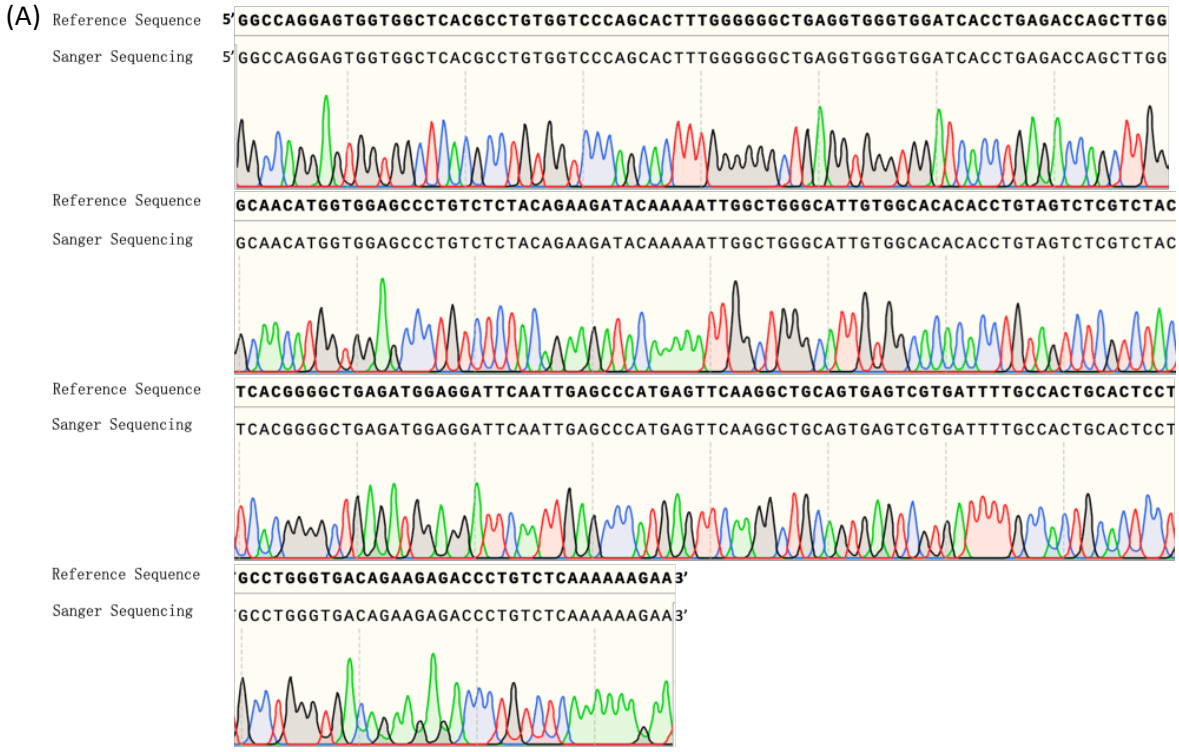
  GCTGGGTGGGATGGCCCATGCTTGTAGTCCCAGCTACTCAGGAGGCTAAGCAAAAGGATGGCTT
  GAGCCCAAGAGTTCAGAGCAGCCTGGCCACCATAGTGAGATCCCTTCTCTTAAAAA

Appendix J *An alignment of Sanger Sequencing results for the A to I gDNA HepG2 reference PCR product and Ensembl's annotation of the DHFR2's genomic sequence. The A to I gDNA HepG2 reference PCR product aligns perfectly back to Ensembl's annotation of the DHFR2's genomic sequence. Performed on Snapgene.*

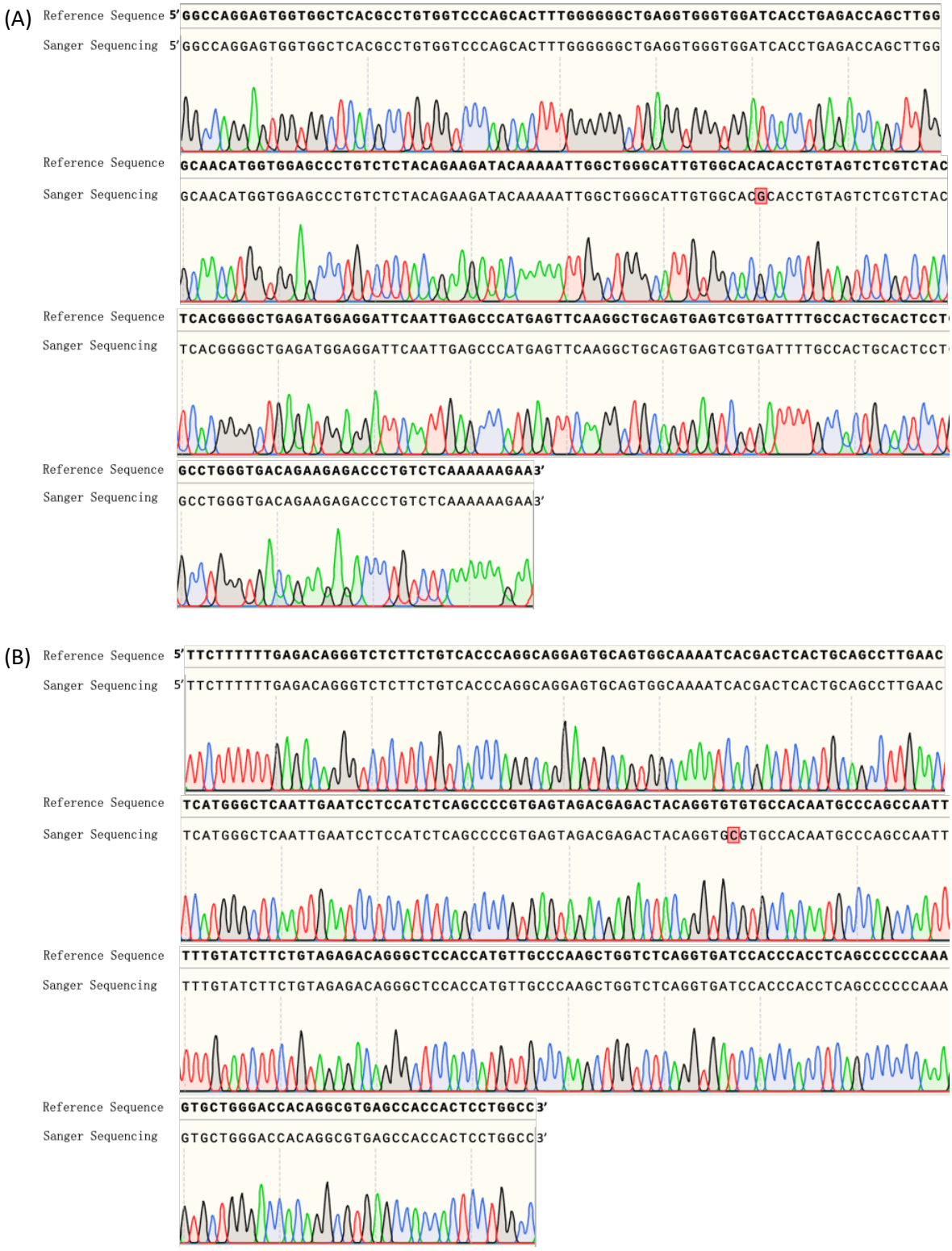



Reference Sequence
TCAACTGATAAATCTATATTCAGGCAAAGAATCAAGGGTGAAGGTAAAATAATACATTTTTAGACAAGCAAAGACTCAGGGGTTACATGTACCCTTTCTT
Sanger Sequencing
TCAACTGATAAATCTATATTCAGGCAAAGAATCAAGGGTGAAGGTAAAATAATACATTTTTAGACAAGCAAAGACTCAGGGGTTACATGTACCCTTTCTT
Reference Sequence
GGGAAGCTATCGGAGAAAATACTCCAGCAAAATGAAGGCGTACACAAACCAGAGAATGACGTGGATCCAGCAAATAGTATCCAACACAGGCAATATTCCA
Sanger Sequencing
GGGAAGCTATCGGAGAAAATACTCCAGCAAAATGAAGGCGTACACAAACCAGAGAATGACGTGGATCCAGCAAATAGTATCCAACACAGGCAATATTCCA
Reference Sequence
GCTATGGAGCTAGCTTTAAAAAGAAACAGTAAAAATACTAATAGATTAGCTGGGTGGGATGGCCCATGCTTGTAGTCCCAGCTACTCAGGAGGCTAAGCA
Sanger Sequencing
GCTATGGAGCTAGCTTTAAAAAGAAACAGTAAAAATACTAATAGATTAGCTGGGTGGGATGGCCCATGCTTGTAGTCCCAGCTACTCAGGAGGCTAAGCA
Reference Sequence
AAAGGATGGCTTGAGCCCAAGAGTTCAGAGCAGCCTGGCCACCATAGTGAGATCCCTTCTCTTAAAAA 3′
Sanger Sequencing
AAAGGATGGCTTGAGCCCAAGAGTTCAGAGCAGCCTGGCCACCATAGTGAGATCCCTTCTCTTAAAAA 3′

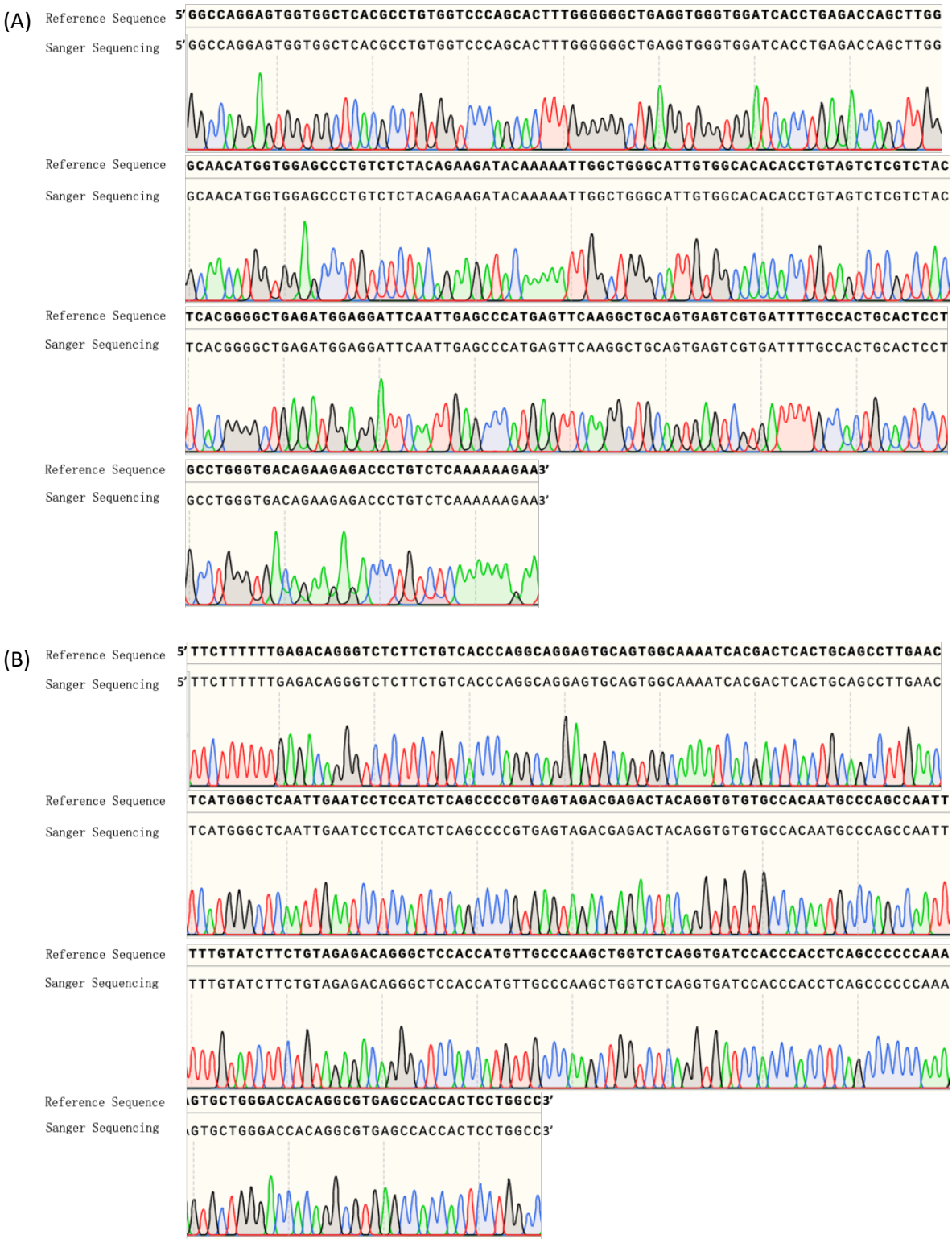
Appendix K *An alignment of Sanger Sequencing results for the A to I cDNA HepG2 PCR Product # 1 and the reference sequence (genomic sequence). (A) shows the reverse primer results and (B) illustrates the forward primer results. The A to I cDNA HepG2 PCR Product # 1 sequencing results aligned perfectly to the reference sequence, meaning there was no A to I edits in the AluJb_Alu element of this sample. Performed on Snapgene*.



Appendix L *An alignment of Sanger Sequencing results for the A to I cDNA HepG2 PCR Product # 2 and the reference sequence (genomic sequence). (A) shows the reverse primer results and (B) illustrates the forward primer results. The A to I cDNA HepG2 PCR Product # 2 sequencing results aligned to the reference sequence, apart from one nucleotide (red box). This nucleotide difference represents an A to I edit in the AluJb_Alu element of this sample. Performed on Snapgene.*



Appendix M *An alignment of Sanger Sequencing results for the A to I cDNA HepG2 PCR Product # 3 and the reference sequence (genomic sequence). (A) shows the reverse primer results and (B) illustrates the forward primer results. The A to I cDNA HepG2 PCR Product # 3 sequencing results aligned perfectly to the reference sequence, meaning there was no A to I edits in the AluJb_Alu element of this sample. Performed on Snapgene*.

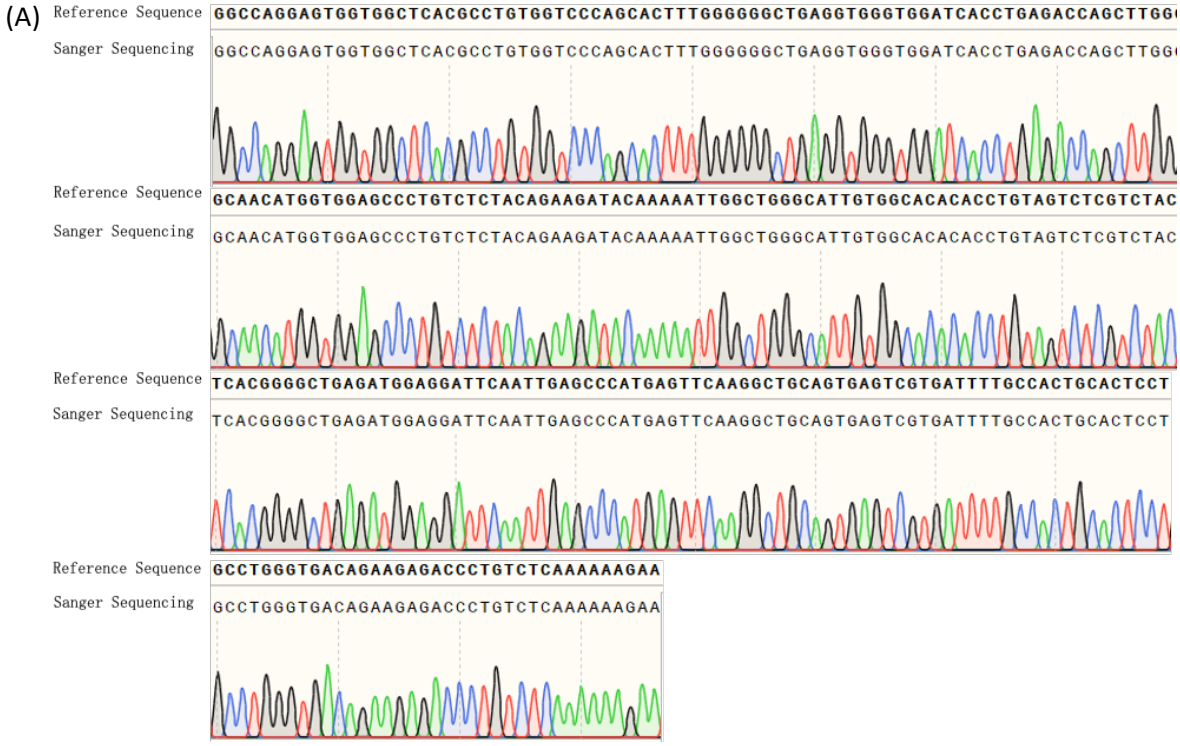


Appendix N *An alignment of Sanger Sequencing results for the A to I cDNA HepG2 PCR Product # 4 and the reference sequence (genomic sequence). (A) shows the reverse primer results and (B) illustrates the forward primer results. The A to I cDNA HepG2 PCR Product # 4 sequencing results aligned perfectly to the reference sequence, meaning there was no A to I edits in the AluJb_Alu element of this sample. Performed on Snapgene.*

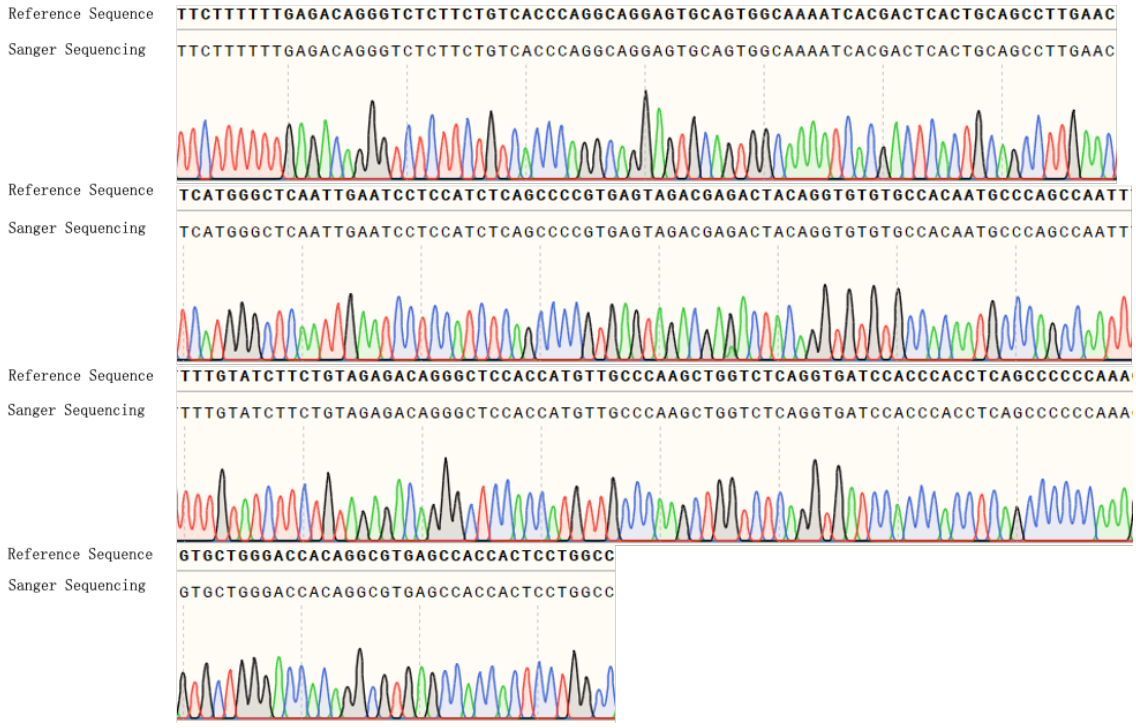


Appendix O *An alignment of Sanger Sequencing results for the A to I cDNA HepG2 PCR Product # 5 and the reference sequence (genomic sequence). (A) shows the reverse primer results and (B) illustrates the forward primer results. The A to I cDNA HepG2 PCR Product # 5 sequencing results aligned perfectly to the reference sequence, meaning there was no A to I edits in the AluJb_Alu element of this sample. Performed on Snapgene.*

(A)

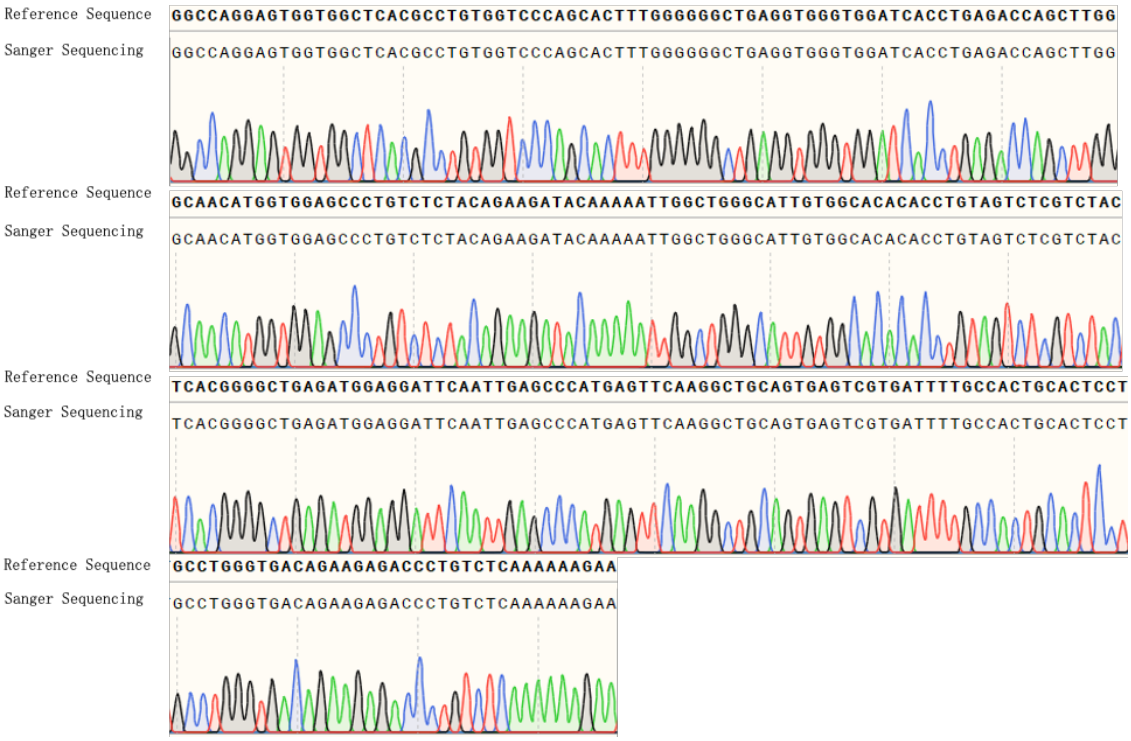


(B)

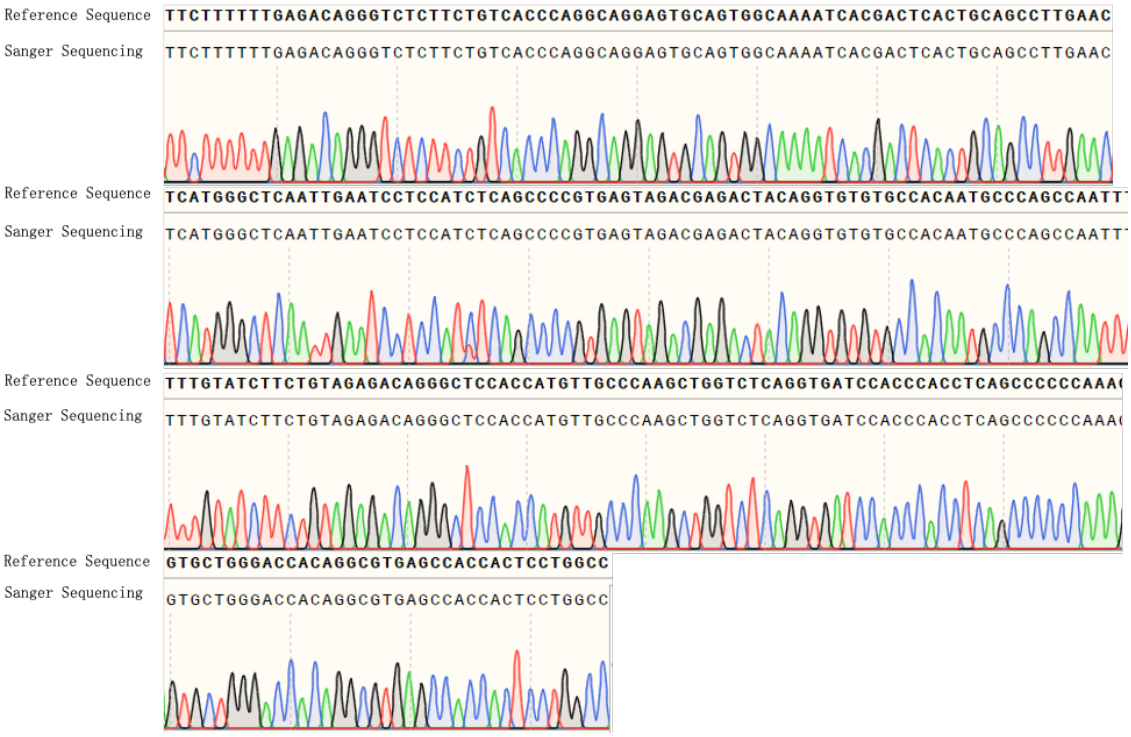


Appendix P *An alignment of Sanger Sequencing results for the A to I cDNA HepG2 PCR Product # 6 and the reference sequence (genomic sequence). (A) shows the reverse primer results and (B) illustrates the forward primer results. The A to I cDNA HepG2 PCR Product # 6 sequencing results aligned perfectly to the reference sequence, meaning there was no A to I edits in the AluJb_Alu element of this sample. Performed on Snapgene.*

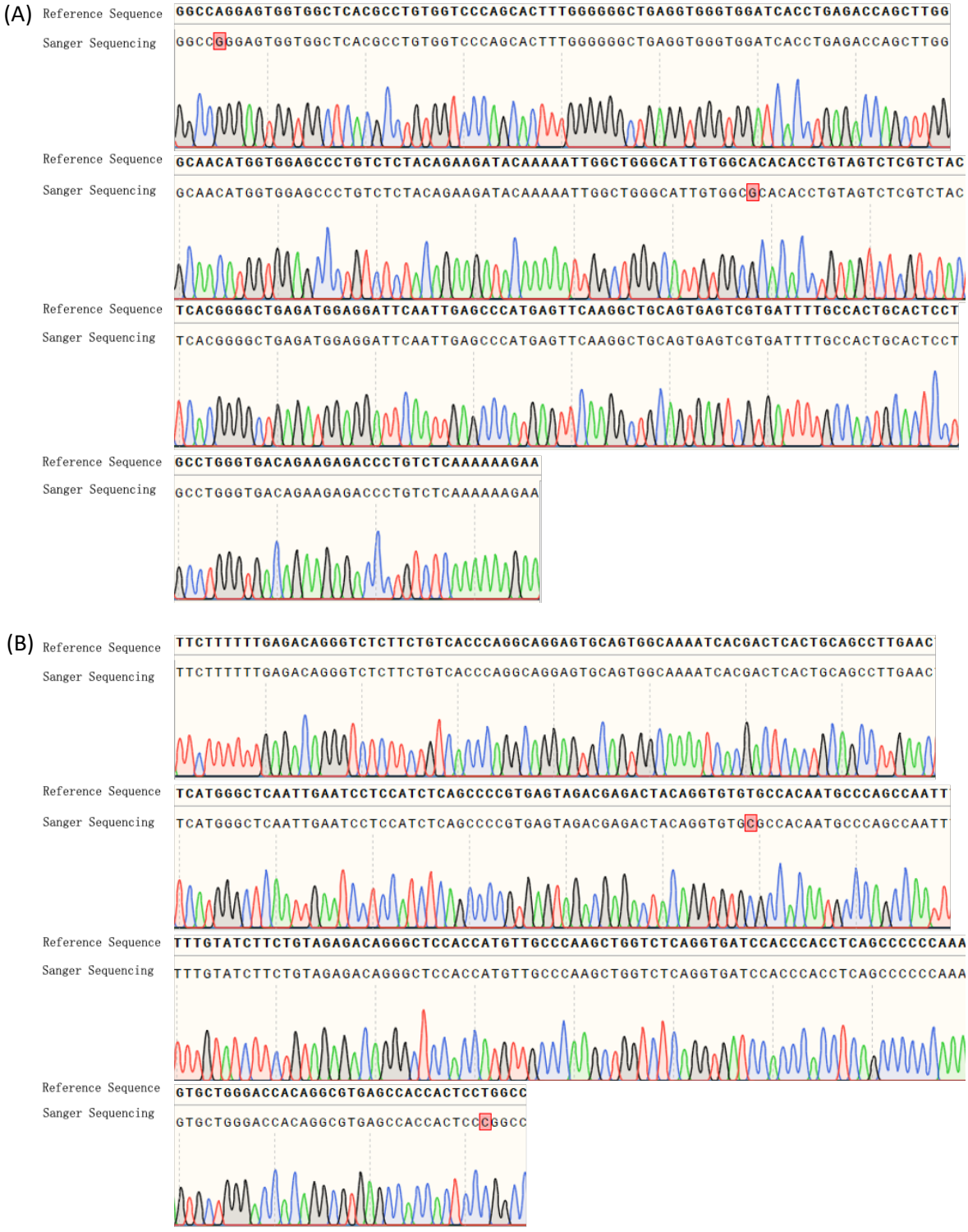
(A)



(B)

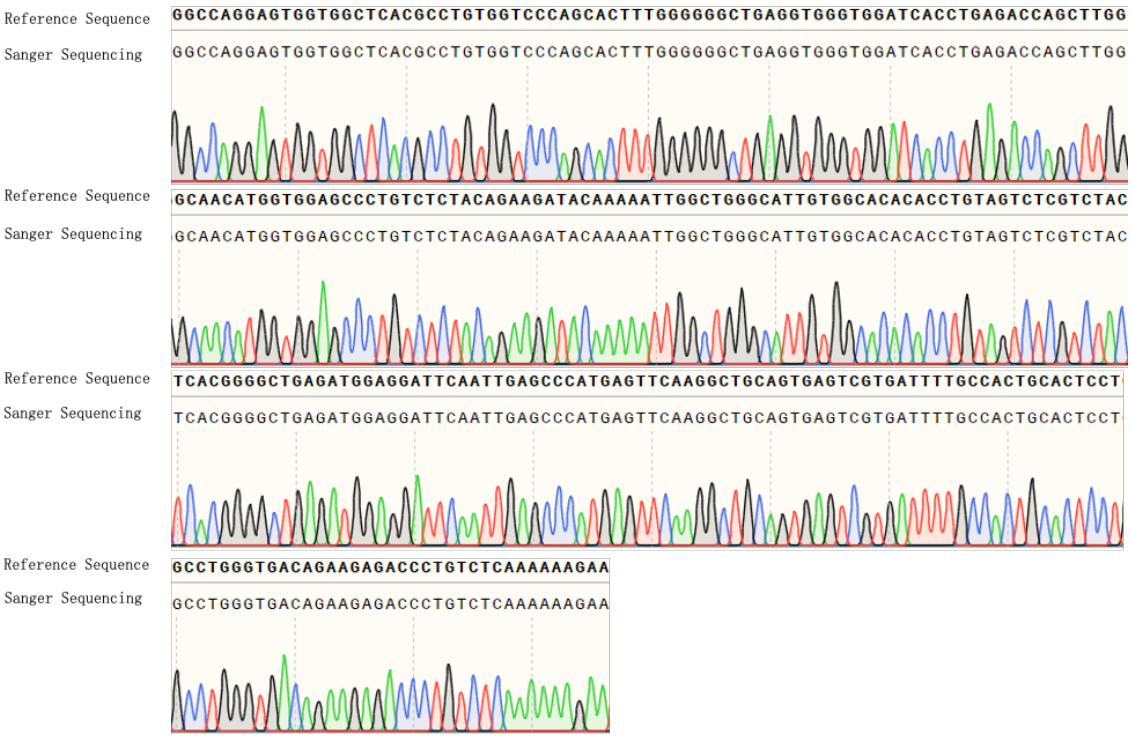


Appendix Q *An alignment of Sanger Sequencing results for the A to I cDNA HepG2 PCR Product # 7 and the reference sequence (genomic sequence). (A) shows the reverse primer results and (B) illustrates the forward primer results. The A to I cDNA HepG2 PCR Product # 7 sequencing results aligned to the reference sequence, apart from two nucleotides (red boxes). These nucleotide differences represent A to I edits in the AluJb_Alu element of this sample. Performed on Snapgene.*

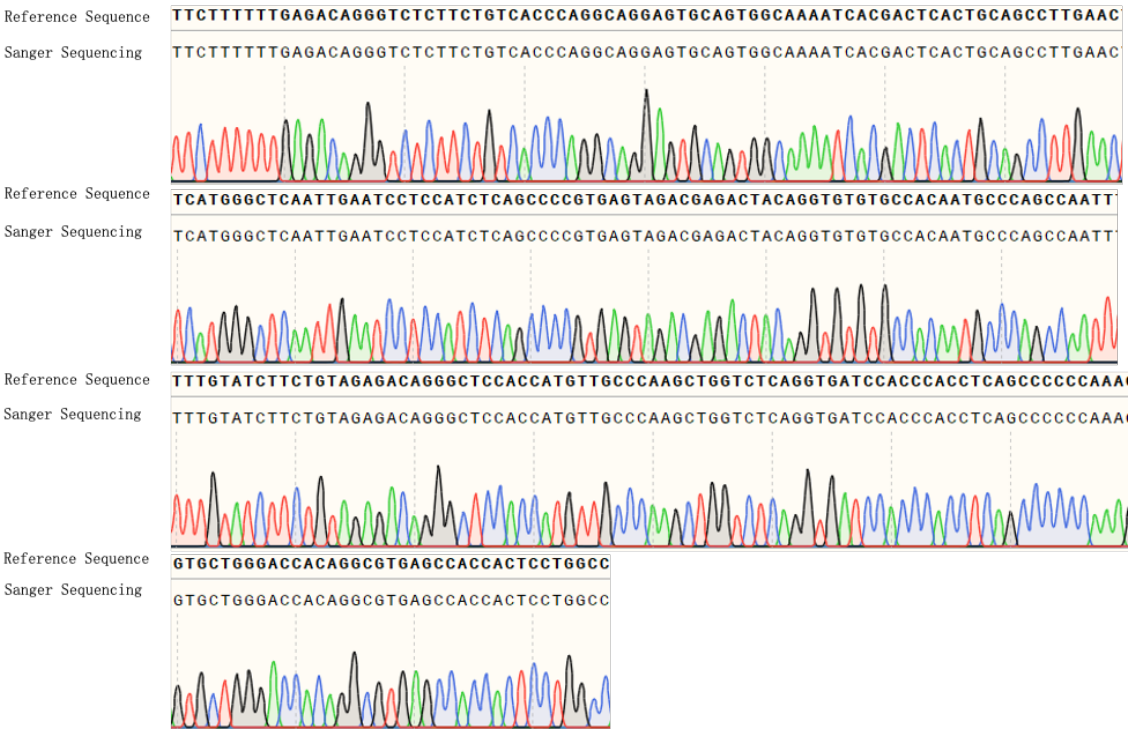


Appendix R *An alignment of Sanger Sequencing results for the A to I cDNA HepG2 PCR Product # 8 and the reference sequence (genomic sequence). (A) shows the reverse primer results and (B) illustrates the forward primer results. The A to I cDNA HepG2 PCR Product # 8 sequencing results aligned perfectly to the reference sequence, meaning there was no A to I edits in the AluJb_Alu element of this sample. Performed on Snapgene.*

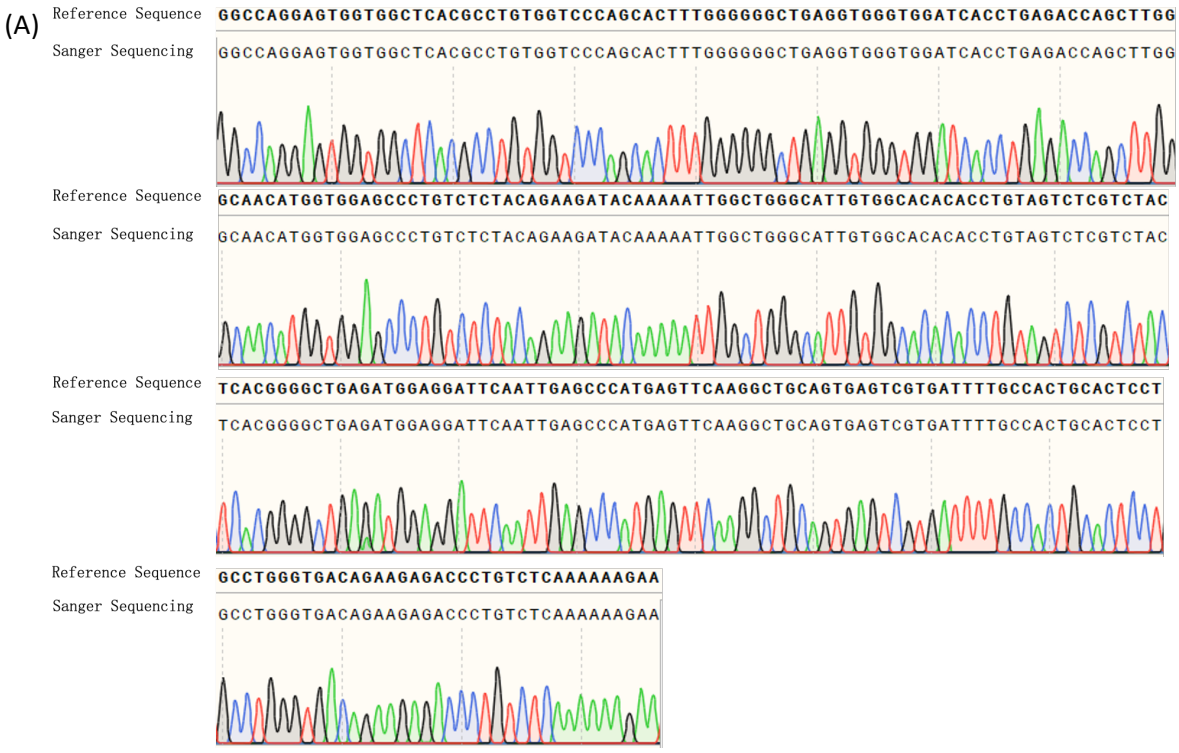
(A)

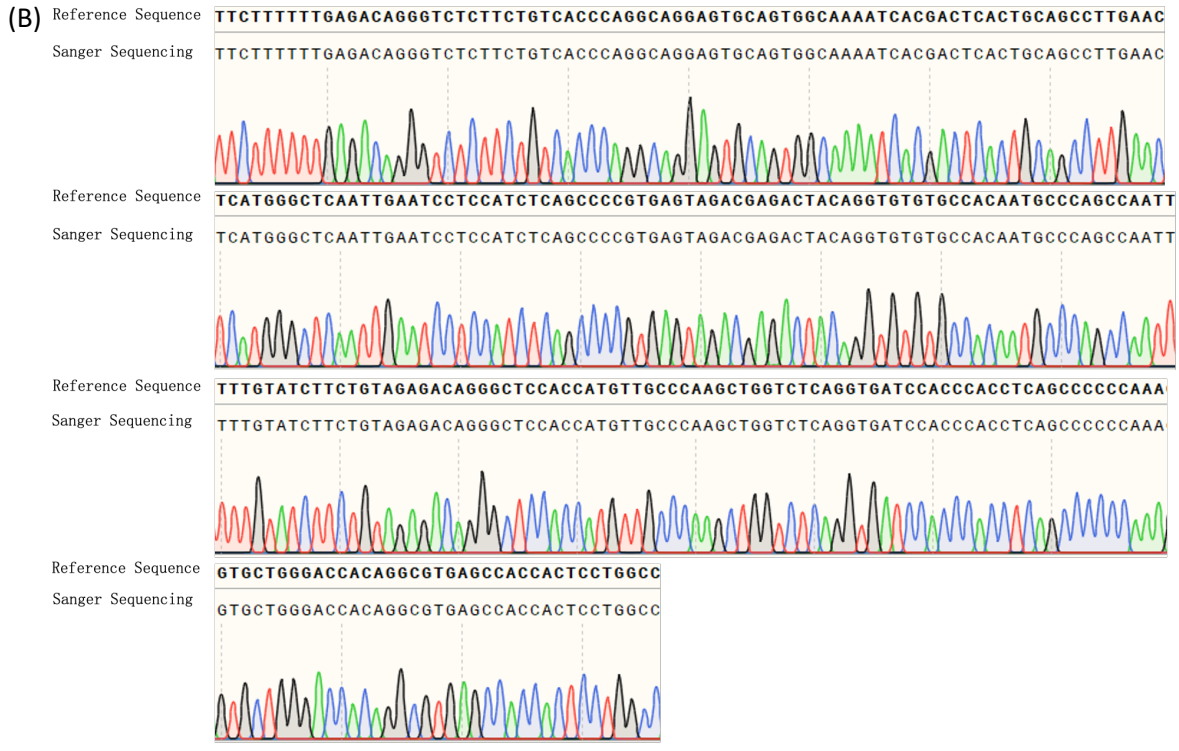


(B)

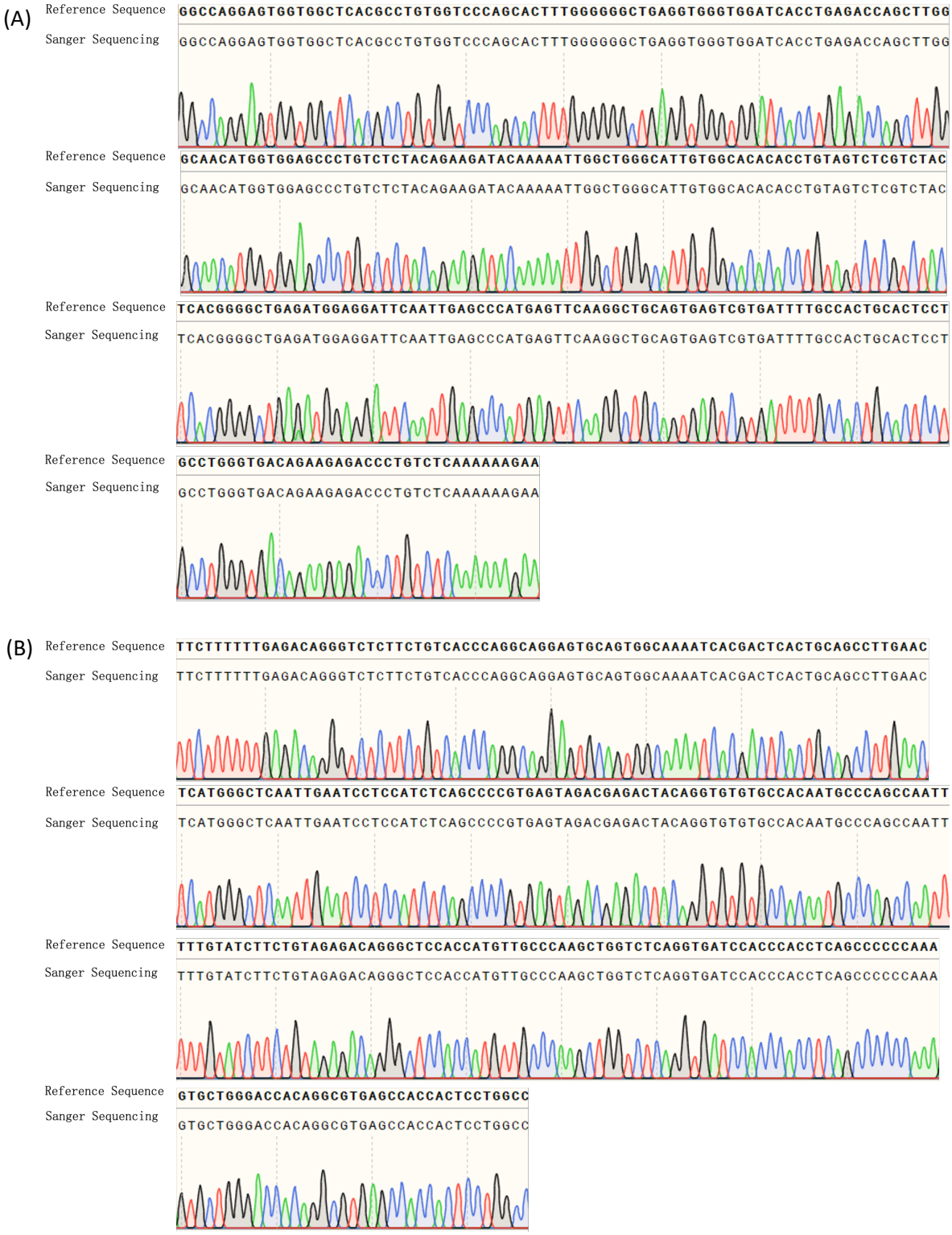


Appendix S *An alignment of Sanger Sequencing results for the A to I cDNA HepG2 PCR Product # 9 and the reference sequence (genomic sequence). (A) shows the reverse primer results and (B) illustrates the forward primer results. The A to I cDNA HepG2 PCR Product # 9 sequencing results aligned perfectly to the reference sequence, meaning there was no A to I edits in the AluJb_Alu element of this sample. Performed on Snapgene.*

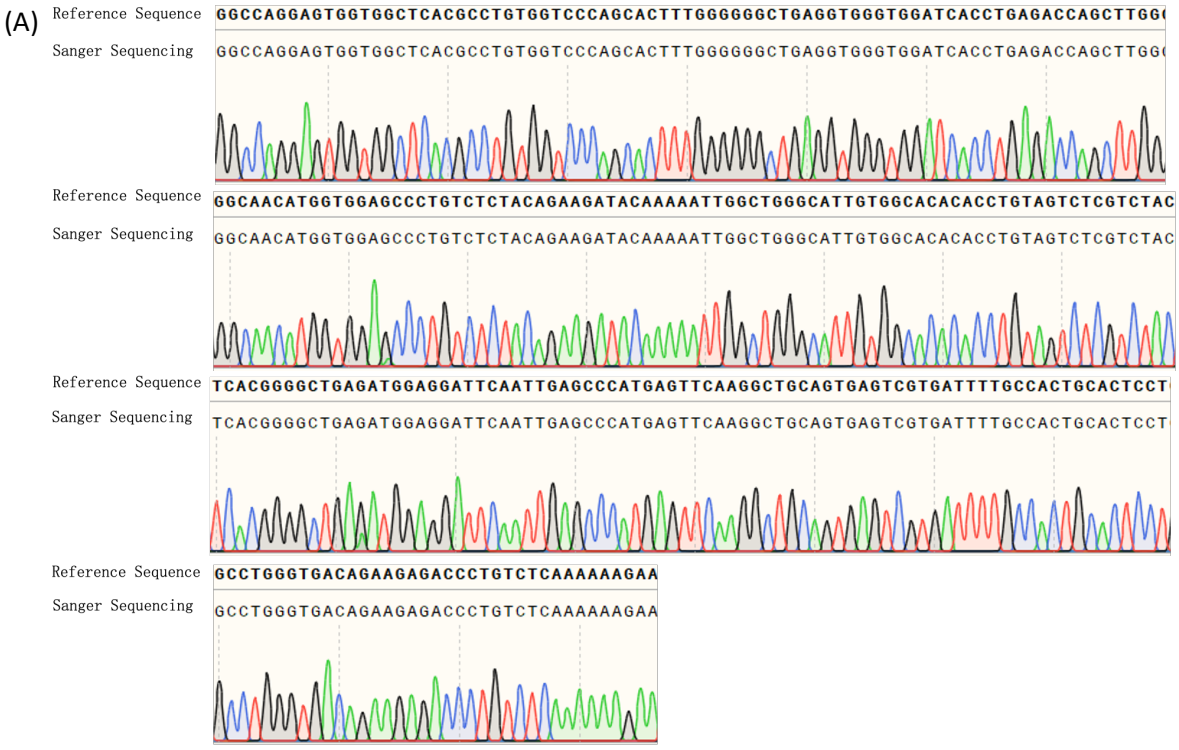


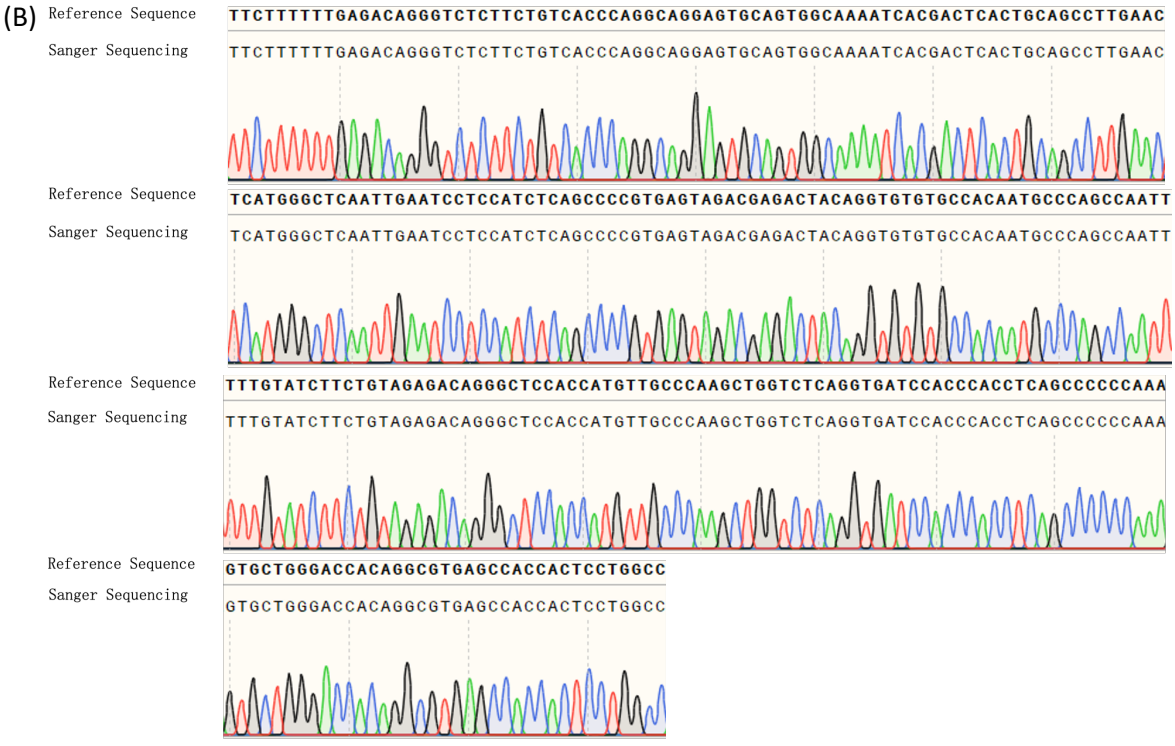


Appendix T *An alignment of Sanger Sequencing results for the A to I cDNA HepG2 PCR Product # 10 and the reference sequence (genomic sequence). (A) shows the reverse primer results and (B) illustrates the forward primer results. The A to I cDNA HepG2 PCR Product # 10 sequencing results aligned perfectly to the reference sequence, meaning there was no A to I edits in the AluJb_Alu element of this sample. Performed on Snapgene.*

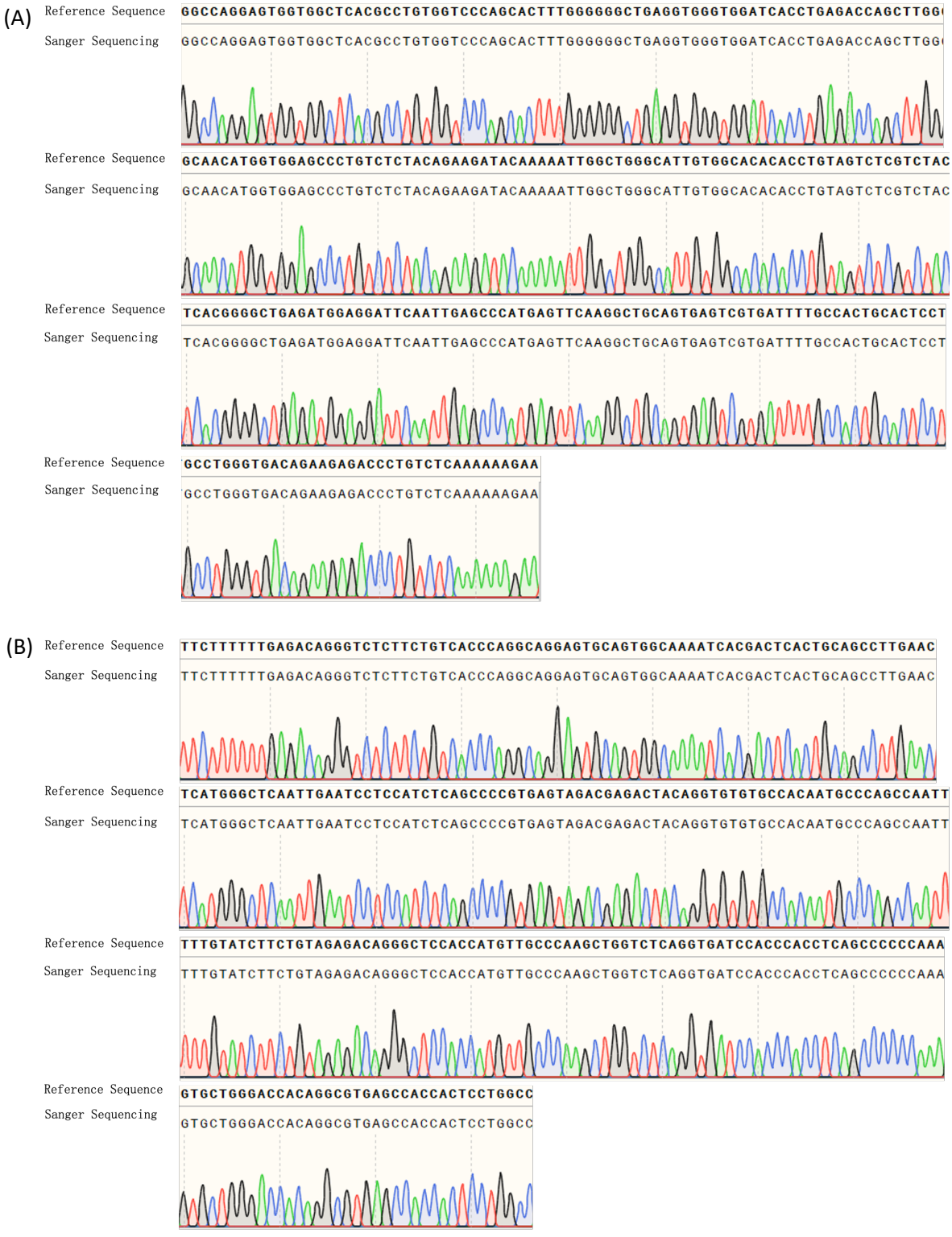


Appendix U *An alignment of Sanger Sequencing results for the A to I cDNA HepG2 PCR Product # 11 and the reference sequence (genomic sequence). (A) shows the reverse primer results and (B) illustrates the forward primer results. The A to I cDNA HepG2 PCR Product # 11 sequencing results aligned perfectly to the reference sequence, meaning there was no A to I edits in the AluJb_Alu element of this sample. Performed on Snapgene.*

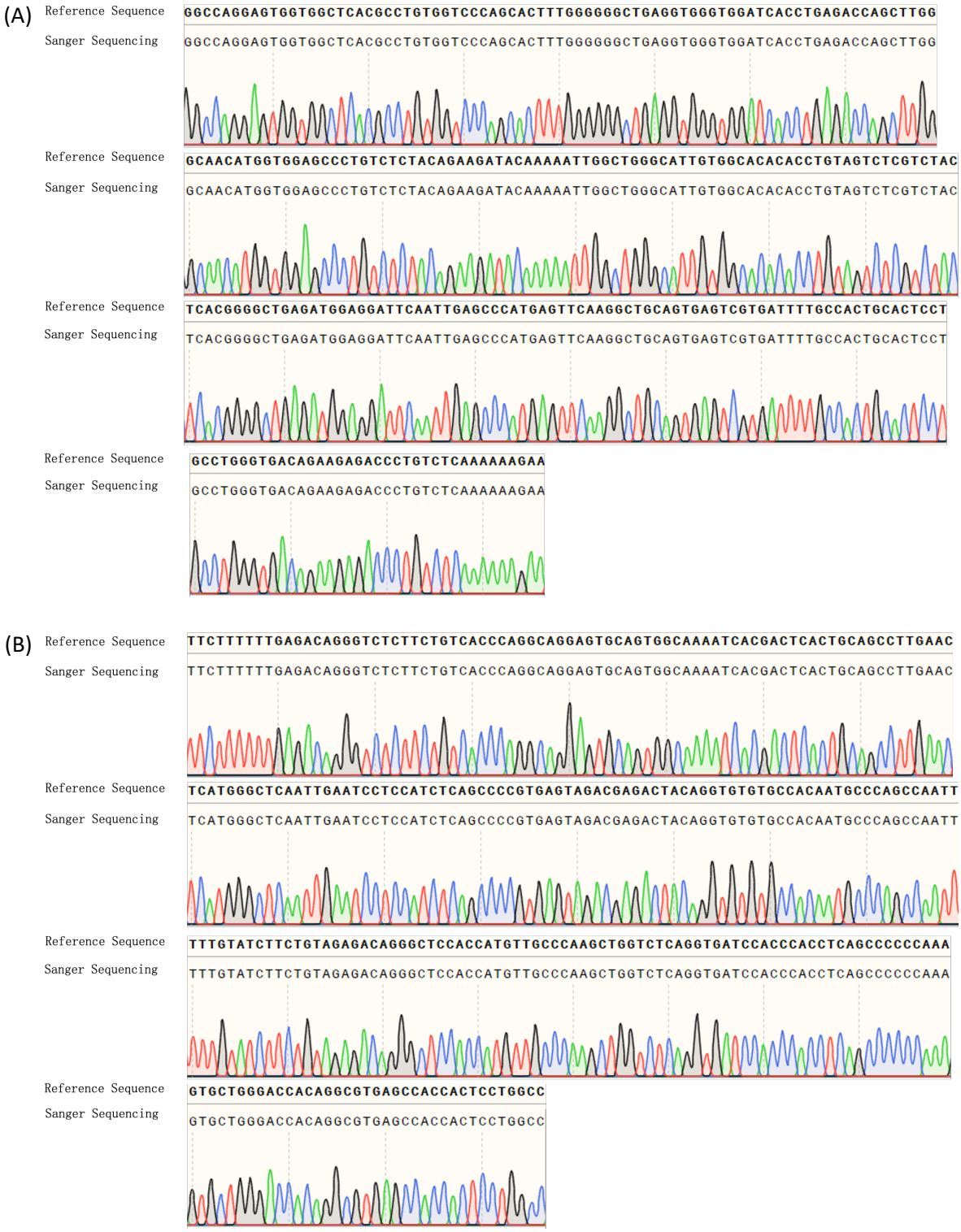




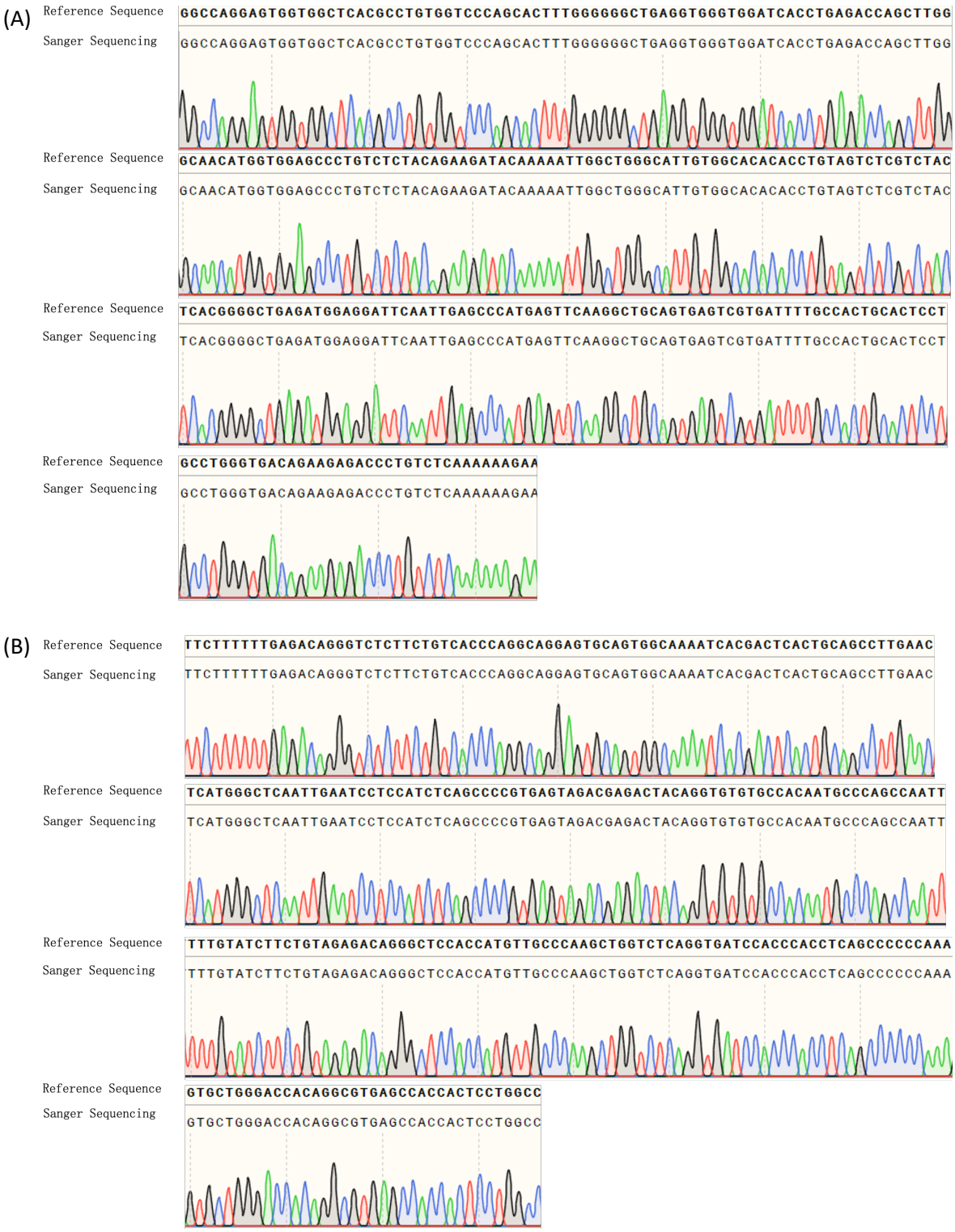
Appendix V *An alignment of Sanger Sequencing results for the A to I cDNA HepG2 PCR Product # 12 and the reference sequence (genomic sequence). (A) shows the reverse primer results and (B) illustrates the forward primer results. The A to I cDNA HepG2 PCR Product # 12 sequencing results aligned perfectly to the reference sequence, meaning there was no A to I edits in the AluJb_Alu element of this sample. Performed on Snapgene*.



Appendix W *An alignment of Sanger Sequencing results for the A to I cDNA HepG2 PCR Product # 13 and the reference sequence (genomic sequence). (A) shows the reverse primer results and (B) illustrates the forward primer results. The A to I cDNA HepG2 PCR Product # 13 sequencing results aligned perfectly to the reference sequence, meaning there was no A to I edits in the AluJb_Alu element of this sample. Performed on Snapgene*.

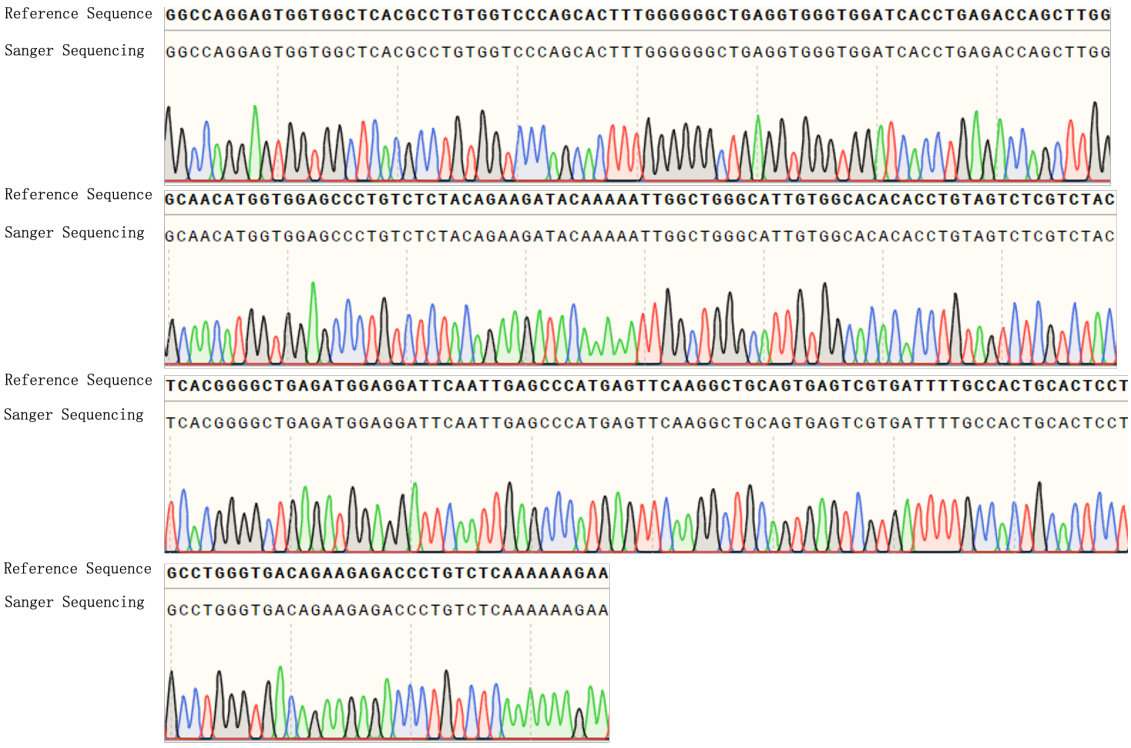


Appendix X *An alignment of Sanger Sequencing results for the A to I cDNA HepG2 PCR Product # 14 and the reference sequence (genomic sequence). (A) shows the reverse primer results and (B) illustrates the forward primer results. The A to I cDNA HepG2 PCR Product # 14 sequencing results aligned perfectly to the reference sequence, meaning there was no A to I edits in the AluJb_Alu element of this sample. Performed on Snapgene*.

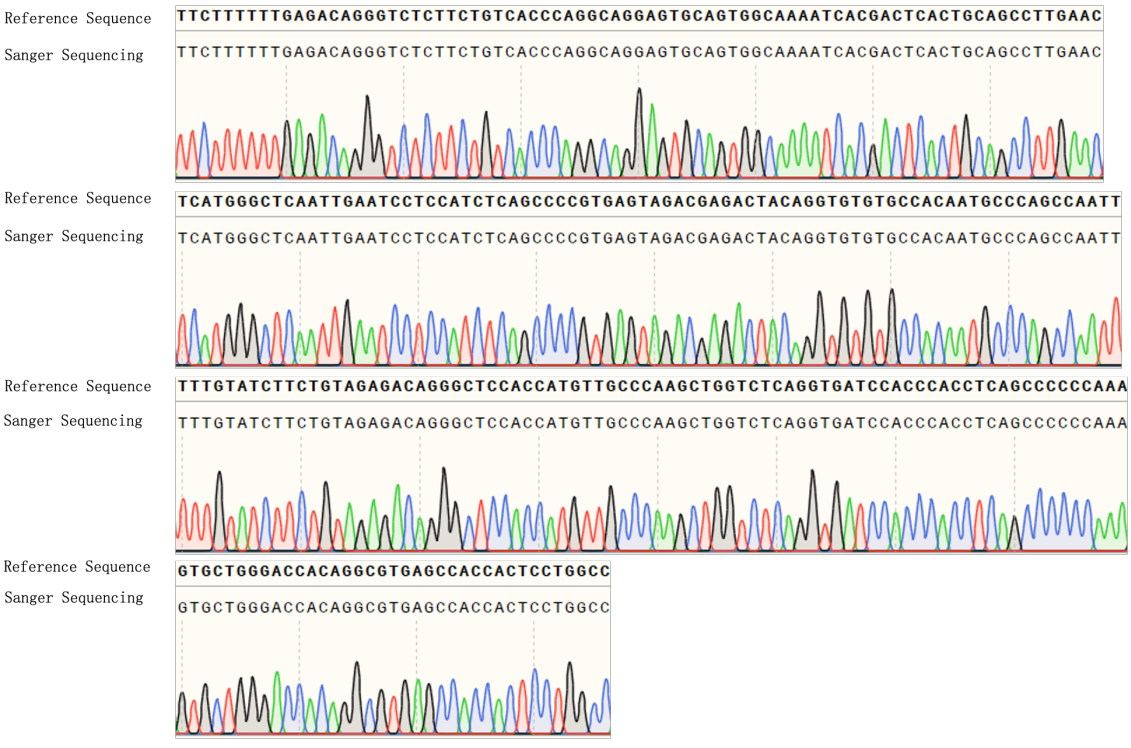


Appendix Y *An alignment of Sanger Sequencing results for the A to I cDNA HepG2 PCR Product # 15 and the reference sequence (genomic sequence). (A) shows the reverse primer results and (B) illustrates the forward primer results. The A to I cDNA HepG2 PCR Product # 15 sequencing results aligned perfectly to the reference sequence, meaning there was no A to I edits in the AluJb_Alu element of this sample. Performed on Snapgene.*

(A)



(B)



Appendix Z *Differentially expressed proteins identified in the HepG2 DHFR2 knockout cell line when compared to HepG2 wildtype.*

Accession	Gene Name	Anova (p)	Fold change
Q9NQ94	A1CF	2.01E-05	-1.9088
Q13685	AAMP	2.26E-05	-2.8605
P49588	AARS1	5.51E-04	1.5150
Q9UDR5	AASS	1.15E-03	-1.6670
Q9NRK6	ABCB10	2.01E-02	-1.9150
P33527	ABCC1	6.41E-04	1.7558
Q92887	ABCC2	2.35E-05	-3.6147
P42765	ACAA2	6.85E-04	-1.5579
Q13085	ACACA	1.35E-03	1.5733
P45954	ACADSB	1.46E-03	-1.6205
P49748	ACADVL	4.96E-04	-1.5703
Q86TX2	ACOT1	1.15E-03	-1.5734
O14734	ACOT8	6.55E-04	-1.8643
Q15067	ACOX1	3.40E-03	-1.6524
Q99424	ACOX2	2.91E-05	-3.3197
Q96CM8	ACSF2	5.84E-04	-1.7921
P33121	ACSL1	1.52E-04	-1.8178
O95573	ACSL3	5.54E-05	2.0052
Q9ULC5	ACSL5	8.35E-06	-3.0748
Q68CK6	ACSM2B	8.38E-06	-9.0781
Q9NR19	ACSS2	1.24E-03	1.6629
Q03154	ACY1	3.65E-03	1.5189
O14672	ADAM10	3.05E-02	-1.8383
Q9UEY8	ADD3	2.89E-04	-1.6825
P08319	ADH4	4.04E-05	-4.2364
P28332	ADH6	1.40E-04	-1.5095
P55196	AFDN	2.17E-03	1.5815
P52594	AGFG1	6.20E-04	1.5354
O00116	AGPS	5.47E-05	-2.8803
O00468	AGRN	4.07E-06	5.1859
P21549	AGXT	1.88E-03	-1.7474
P02765	AHSG	2.37E-03	1.7273
P00568	AK1	1.29E-04	1.7041
Q13740	ALCAM	1.74E-05	-4.8433
P00352	ALDH1A1	2.61E-05	-1.9908
O75891	ALDH1L1	4.32E-06	-22.4986
P51648	ALDH3A2	1.64E-03	-1.6075
Q02252	ALDH6A1	5.41E-05	-2.4347
P09972	ALDOC	1.30E-04	2.6260
Q9BT22	ALG1	5.29E-03	-1.5107
Q9H553	ALG2	2.50E-03	-1.6394
Q9Y673	ALG5	1.08E-03	-1.5899
Q9H6U8	ALG9	4.89E-03	-2.0572
Q9P2R3	ANKFY1	1.60E-03	-1.6674
P53680	AP2S1	2.62E-03	-1.6241
Q06481	APLP2	2.04E-03	1.6033
Q9HDC9	APMAP	6.67E-03	-1.8212
P02647	APOA1	5.09E-05	-1.9923
P02649	APOE	9.37E-04	1.5672
P02749	APOH	1.45E-04	-3.1342
Q9NP61	ARFGAP3	7.84E-05	2.8566
P78540	ARG2	1.65E-05	3.0729
Q8N392	ARHGAP18	5.51E-03	1.5828
Q92888	ARHGEF1	6.57E-03	2.0057
Q99856	ARID3A	2.34E-03	1.5015
Q6NXE6	ARMC6	6.87E-03	1.5685
P51690	ARSL	3.04E-04	-1.8252
P07307	ASGR2	9.55E-04	-1.5130
P08243	ASNS	5.45E-05	4.3257
Q12797	ASPH	9.21E-03	1.8333
Q7L266	ASRGL1	1.38E-04	2.1357
Q6PL18	ATAD2	1.68E-03	-1.6337
P54709	ATP1B3	6.70E-03	-1.6877
P23634	ATP2B4	3.74E-03	-1.9997
Q15904	ATP6AP1	1.87E-04	-2.0324
Q8WWM7	ATXN2L	9.10E-04	1.5575
Q9Y679	AUP1	4.80E-05	-1.6281
O14965	AURKA	2.05E-02	-1.6685
P46379	BAG6	1.04E-02	1.5773
Q9P287	BCCIP	6.50E-04	1.5373
P21953	BCKDHB	1.65E-03	-1.6661
Q86WA6	BPHL	7.27E-04	-1.7392
Q96K17	BTF3L4	8.74E-03	1.6814
A6NCS6	C2orf72	4.13E-03	-1.6609
P35218	CA5A	1.63E-04	-2.5198
P27708	CAD	4.40E-04	1.5914
P49589	CARS1	6.71E-03	1.5027
O14936	CASK	3.91E-03	-1.7159
P20810	CAST	1.82E-03	1.8587
P04040	CAT	1.07E-04	-1.6911
P0DN79	CBSL	4.96E-04	-1.8619
P29279	CCN2	1.55E-07	18.2271
Q9Y5K6	CD2AP	7.11E-03	1.6204
P04233	CD74	1.05E-04	-1.7459
P30260	CDC27	2.35E-03	-1.6156
P19022	CDH2	6.60E-06	-3.4642
Q96SZ6	CDK5RAP1	1.98E-03	-2.1173
Q9UHN6	CEMIP2	1.32E-05	2.1990
Q9HA82	CERS4	2.15E-02	-1.6640
P23141	CES1	3.47E-07	-2.9569
Q96JM3	CHAMP1	3.85E-03	1.5545
Q9BSY4	CHCHD5	9.74E-03	1.9078
Q8NET6	CHST13	6.62E-03	-1.7684
Q6FI81	CIAPIN1	3.67E-03	1.5743
Q07065	CKAP4	2.35E-05	2.2443
P12277	CKB	5.50E-04	1.7748
O00299	CLIC1	3.56E-05	1.5785
O75153	CLUH	8.46E-04	-1.5219
Q9Y2R0	COA3	2.46E-03	-1.5276
Q53SF7	COBLL1	5.82E-03	-1.7252
P39060	COL18A1	7.20E-04	-1.9679
Q8NI60	COQ8A	3.76E-04	-1.7994
Q9BR76	CORO1B	5.86E-04	1.5921
P00450	CP	5.56E-05	3.7506
P31327	CPS1	2.66E-03	-1.6431
Q9P2I0	CPSF2	2.10E-03	1.5120
P19784	CSNK2A2	5.38E-04	1.5592
P21291	CSRP1	8.51E-06	5.3035
P35221	CTNNA1	6.51E-06	-1.9074
O60716	CTNND1	4.88E-05	-1.5104
P78310	CXADR	5.03E-06	1.7940
P00167	CYB5A	9.91E-04	-1.6348
Q02318	CYP27A1	1.88E-04	-1.7803
Q16643	DBN1	1.05E-06	2.3803
P11182	DBT	4.35E-03	-1.8321
P32321	DCTD	4.56E-03	1.5080
Q7Z4W1	DCXR	1.08E-04	-1.6104
Q92466	DDB2	4.49E-04	1.8493
P20711	DDC	1.33E-03	-1.7927
P39656	DDOST	3.29E-02	-1.5230
Q9NUI1	DECR2	1.06E-03	-1.6423
O75907	DGAT1	1.78E-04	-1.7856
P00374	DHFR	1.40E-04	-2.4042
Q96LJ7	DHRS1	8.72E-03	-1.7398
Q13268	DHRS2	1.78E-04	-2.0130
Q7L2E3	DHX30	5.83E-04	-2.0785
O60610	DIAPH1	2.24E-03	-1.5116
P11532	DMD	8.28E-05	-1.7571
P50570	DNM2	2.13E-03	1.6914
Q9Y6K1	DNMT3A	1.27E-03	-1.5912
Q96N67	DOCK7	5.20E-03	1.5926
Q14126	DSG2	6.81E-06	1.7290
P15924	DSP	3.88E-06	1.7311
Q99956	DUSP9	5.24E-04	-1.5456
P30084	ECHS1	6.35E-04	-1.6295
P42126	ECI1	7.00E-04	-1.5141
O75521	ECI2	6.82E-03	-1.8205
P00533	EGFR	1.28E-05	1.8170
Q8NDI1	EHBP1	1.20E-03	2.0144
Q08426	EHHADH	9.04E-04	-1.5453
Q7L2H7	EIF3M	1.85E-03	-14.5293
O95163	ELP1	9.83E-04	-1.6344
Q9HC35	EML4	3.76E-05	1.8159
P22413	ENPP1	6.29E-05	2.3590
O43491	EPB41L2	1.01E-05	1.5852
P29317	EPHA2	2.38E-05	1.7733
P07099	EPHX1	2.13E-03	-1.6306
P62495	ETF1	7.83E-05	1.5664
P00734	F2	3.93E-04	1.7572
O95864	FADS2	4.28E-03	-1.7958
Q9NZB2	FAM120A	3.21E-04	-1.6616
Q9Y6X4	FAM169A	1.44E-04	1.6660
Q14320	FAM50A	1.46E-04	1.5762
Q9NYY8	FASTKD2	1.02E-03	-1.6545
Q7L8L6	FASTKD5	4.71E-03	-1.6673
P55899	FCGRT	1.98E-05	-2.6752
P22830	FECH	3.56E-03	-1.5193
Q13451	FKBP5	1.02E-04	1.6546
P49326	FMO5	6.32E-06	-1.9481
Q53EP0	FNDC3B	4.06E-03	1.5293
O95954	FTCD	1.47E-05	-1.8920
Q8TAE8	GADD45GIP1	4.84E-03	-1.7762
Q10472	GALNT1	1.06E-03	-1.6715
P41250	GARS1	1.59E-05	1.5278
P22102	GART	6.61E-05	1.5345
O43903	GAS2	2.54E-04	-1.7165
Q04446	GBE1	3.34E-03	1.9598
Q92947	GCDH	2.86E-04	-2.0881
P31150	GDI1	1.03E-02	1.7178
P19440	GGT1	3.10E-02	-1.5892
P32189	GK	2.97E-05	1.8535
P35754	GLRX	1.08E-04	1.8459
P00367	GLUD1	2.02E-04	-1.9879
O60547	GMDS	2.01E-05	1.9041
Q9Y223	GNE	2.58E-02	1.5553
O15228	GNPAT	1.62E-03	-3.4016
P15586	GNS	2.15E-03	1.6015
P17174	GOT1	4.51E-04	1.6326
P51654	GPC3	9.56E-04	1.7635
Q9UBQ7	GRHPR	2.07E-04	-1.5256
P28799	GRN	1.03E-04	2.3328
P08263	GSTA1	4.08E-05	-2.0287
Q9Y2Q3	GSTK1	4.70E-04	-1.8759
P06899	H2BC11	1.87E-04	-2.2350
O60814	H2BC12	8.90E-05	-2.1126
Q5TEC6	H3-2	6.44E-03	-1.8455
P62805	H4C1	6.10E-03	-1.8257
Q6Y1H2	HACD2	4.36E-03	-1.6301
Q16836	HADH	3.88E-03	-1.5196
P69905	HBA1	1.82E-05	7.9147
Q9ULT8	HECTD1	1.82E-04	-2.2946
Q9NRZ9	HELLS	1.14E-06	-8.1484
P49773	HINT1	2.46E-02	-1.5355
Q9BX68	HINT2	1.92E-03	-1.5499
P52789	HK2	3.11E-06	1.6973
Q01581	HMGCS1	3.41E-04	-1.5011
P41235	HNF4A	8.03E-05	-1.8886
Q99714	HSD17B10	3.19E-03	-1.5095
P37059	HSD17B2	1.18E-05	-3.1063
O95757	HSPA4L	1.47E-03	1.7694
P04792	HSPB1	5.43E-06	3.0386
P05362	ICAM1	6.18E-05	2.0143
P08833	IGFBP1	1.10E-06	28.6730
Q8N6C5	IGSF1	1.15E-03	-1.8554
A1L0T0	ILVBL	7.82E-04	-1.7825
Q9NV31	IMP3	8.17E-04	-2.0745
P29218	IMPA1	9.38E-04	2.1523

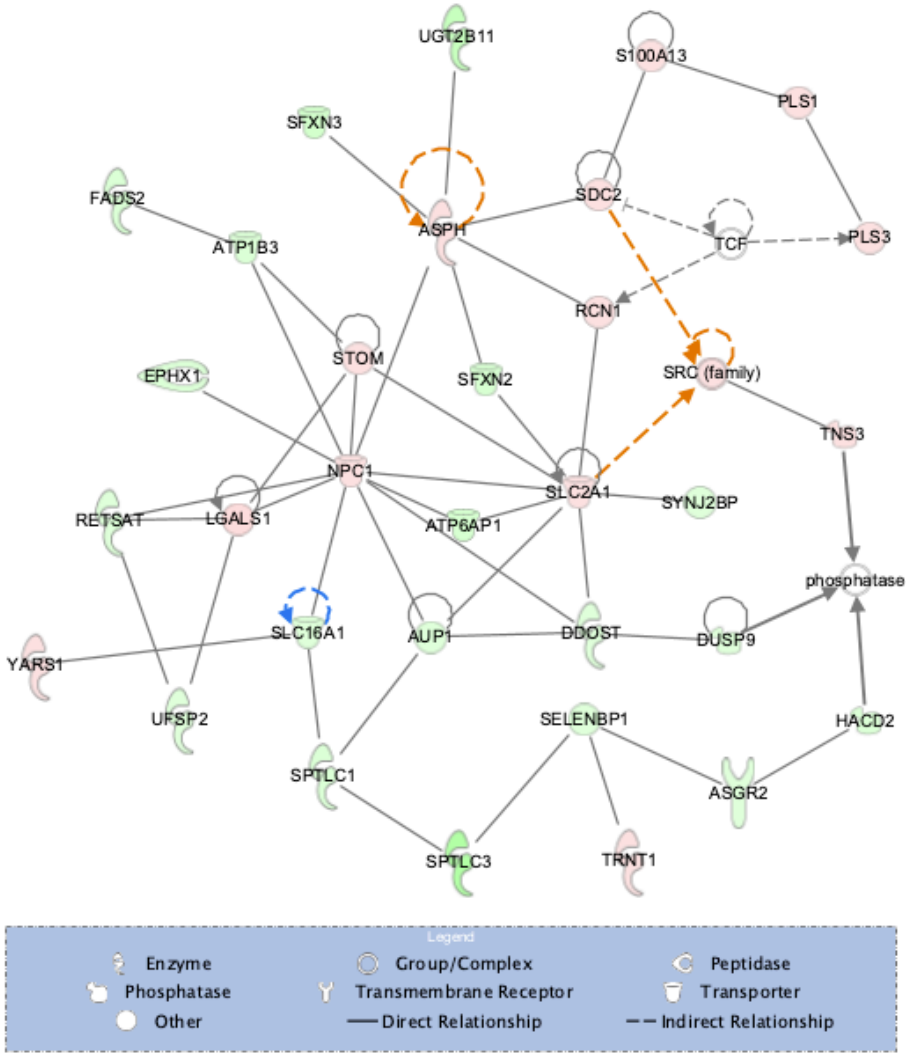
Accession	Gene Name	Anova (p)	Fold change
Q9H1B7	IRF2BPL	6.09E-03	1.8878
Q9NPH2	ISYNA1	2.82E-05	1.8619
P17301	ITGA2	2.17E-05	2.9602
P06756	ITGAV	5.22E-04	2.2294
P05556	ITGB1	2.39E-03	1.6535
P19823	ITIH2	1.27E-04	-1.6278
Q14571	ITPR2	3.94E-03	-1.6109
Q6NYC1	JMJD6	5.51E-03	1.9909
Q07866	KLC1	6.38E-03	1.6998
Q9NSK0	KLC4	4.15E-05	-1.6463
P05783	KRT18	8.15E-05	1.8391
P05787	KRT8	8.52E-06	1.8830
Q16719	KYNU	2.10E-03	1.6400
O00515	LAD1	4.49E-03	1.6899
O15230	LAMA5	8.09E-05	2.0597
P07942	LAMB1	4.32E-05	2.7058
P55268	LAMB2	1.68E-06	-8.5226
P11047	LAMC1	1.59E-03	1.5190
P01130	LDLR	5.28E-05	4.4600
P09382	LGALS1	1.45E-08	5.4910
P17931	LGALS3	1.70E-05	3.5152
Q9UHB6	LIMA1	8.40E-04	1.8058
Q6ZN17	LIN28B	1.69E-04	-1.9508
P49257	LMAN1	1.89E-03	-1.5279
P02545	LMNA	2.16E-04	1.5041
Q93052	LPP	6.74E-04	1.6201
P30533	LRPAP1	3.83E-04	-1.6388
Q32MZ4	LRRFIP1	1.28E-03	1.8133
P61626	LYZ	1.22E-05	2.2120
Q9Y5V3	MAGED1	1.03E-02	-1.5480
P21397	MAOA	8.49E-04	-2.1336
P27338	MAOB	4.05E-05	-2.5616
P27816	MAP4	5.38E-04	1.5996
O95819	MAP4K4	6.15E-06	1.5026
Q16539	MAPK14	1.77E-02	-1.7956
Q15555	MAPRE2	1.78E-03	1.5527
P56192	MARS1	3.64E-05	1.5100
O15232	MATN3	6.93E-05	131.6123
Q96RQ3	MCCC1	1.97E-03	-1.7496
Q9HCC0	MCCC2	5.04E-03	-1.6721
P49736	MCM2	2.45E-03	-1.9280
P33992	MCM5	3.74E-02	-1.5924
Q6UX53	METTL7B	1.12E-04	-1.8122
O95140	MFN2	6.00E-03	-1.7229
Q10469	MGAT2	8.54E-04	-1.7627
O95822	MLYCD	4.46E-03	-1.8221
Q96EY8	MMAB	7.74E-03	-1.6190
P22033	MMUT	5.69E-05	-2.1612
Q9UBU8	MORF4L1	1.07E-03	-1.7636
Q9HCE1	MOV10	5.12E-05	-1.5007
Q5T653	MRPL2	1.56E-02	2.0554
Q9BZE1	MRPL37	6.99E-03	-1.5777
Q92552	MRPS27	4.43E-03	-1.7467
P02795	MT2A	3.22E-09	88.7325
Q5VT66	MTARC1	7.36E-04	-1.7569
Q9Y6C9	MTCH2	2.09E-03	-1.6627
Q7Z6M4	MTERF4	3.34E-03	-1.5829
P55157	MTTP	2.36E-05	-1.7204
P53602	MVD	1.67E-04	1.5118
Q15746	MYLK	9.26E-03	-1.6305
O00159	MYO1C	2.51E-03	1.8135
O94832	MYO1D	3.20E-02	-1.6491
Q4G0N4	NADK2	3.73E-04	-1.6478
P43490	NAMPT	8.01E-04	1.9144
Q9NR45	NANS	1.78E-04	-1.8746
Q9HCD5	NCOA5	2.45E-03	-1.5401
Q9BTX1	NDC1	2.42E-02	2.5022
Q92597	NDRG1	3.65E-04	2.9635
Q9UI09	NDUFA12	1.73E-02	-1.5841
P56556	NDUFA6	4.26E-03	-1.5591
Q96TA1	NIBAN2	3.18E-04	2.1001
P14543	NID1	3.42E-05	2.0828
Q9GZT8	NIF3L1	1.43E-02	-1.5437
O75323	NIPSNAP2	2.64E-03	-1.5125
Q9UFN0	NIPSNAP3A	9.42E-04	-1.6087
Q9BSC4	NOL10	4.93E-03	-1.5256
O15118	NPC1	2.86E-03	1.6340

Accession	Gene Name	Anova (p)	Fold change
Q9H0P0	NT5C3A	6.89E-04	-1.8053
Q5TFE4	NT5DC1	6.76E-04	-1.9757
Q9H857	NT5DC2	2.79E-02	-1.5499
Q8TB37	NUBPL	4.46E-03	-1.5618
Q9BRJ7	NUDT16L1	9.37E-03	-1.6073
Q14980	NUMA1	4.93E-03	1.5127
Q8WUM0	NUP133	6.32E-03	-1.5337
Q12769	NUP160	5.42E-04	-1.8221
Q8TEM1	NUP210	3.64E-03	-1.8443
Q8NFH5	NUP35	4.45E-03	-1.5308
Q8NFH3	NUP43	7.44E-03	-1.6166
O60313	OPA1	2.87E-03	-1.6292
O15460	P4HA2	5.30E-06	2.1301
P00439	PAH	6.65E-07	-2.7741
Q13177	PAK2	2.13E-03	1.6088
P05165	PCCA	4.29E-03	-1.6354
Q8WW12	PCNP	6.09E-05	1.5390
Q9UHG3	PCYOX1	2.53E-03	-1.5518
Q15118	PDK1	9.09E-03	1.9658
Q6P996	PDXDC1	1.75E-02	-1.5314
O00764	PDXK	7.28E-04	1.5061
Q5T2W1	PDZK1	1.00E-08	-5.0143
Q9BY49	PECR	7.38E-04	-1.8790
Q9GZU2	PEG3	1.03E-04	2.3619
Q99623	PHB2	2.47E-03	-1.5211
Q8IXK0	PHC2	5.50E-04	1.5596
Q8TBX8	PIP4K2C	3.87E-04	1.7794
Q9P0Z9	PIPOX	5.98E-05	-2.0540
P30613	PKLR	1.58E-02	-1.5881
Q13393	PLD1	1.18E-06	2.5682
P53350	PLK1	1.38E-04	-1.7667
Q9Y342	PLLP	6.92E-06	1.8975
Q14651	PLS1	2.62E-04	1.7105
P13797	PLS3	2.49E-04	1.6157
Q9Y2S7	POLDIP2	2.70E-03	-1.5520
P24928	POLR2A	2.81E-03	-1.5456
P19387	POLR2C	2.39E-04	-1.5414
P62487	POLR2G	2.07E-04	-1.8387
Q15166	PON3	1.67E-04	-4.0843
Q06203	PPAT	9.80E-03	1.5728
Q9UNP9	PPIE	2.30E-02	-1.7026
Q9Y570	PPME1	6.02E-03	1.6536
O75170	PPP6R2	2.15E-05	1.8548
Q5H9R7	PPP6R3	6.24E-03	1.7383
P17252	PRKCA	6.07E-05	-1.7272
Q9UF12	PRODH2	2.43E-03	-1.6631
Q13523	PRPF4B	2.40E-03	-1.5165
P28062	PSMB8	8.38E-03	1.8620
Q9H7Z7	PTGES2	1.73E-03	1.6020
P15151	PVR	1.52E-04	1.5949
P62820	RAB1A	2.74E-04	-3.5001
Q92878	RAD50	1.07E-02	1.5280
Q8IY67	RAVER1	1.07E-02	-1.5347
P02753	RBP4	2.59E-05	-1.7152
Q15293	RCN1	5.25E-05	1.6439
P78509	RELN	4.66E-04	-1.8041
Q6NUM9	RETSAT	7.33E-04	-1.6214
Q15493	RGN	4.81E-03	-1.8728
Q5UIP0	RIF1	4.65E-03	-1.5987
Q8IUF8	RIOX2	1.08E-03	-1.5419
Q9NWS8	RMND1	3.95E-03	-1.8023
O00584	RNASET2	4.84E-04	1.7016
Q99496	RNF2	1.10E-02	1.8068
O75116	ROCK2	1.14E-03	1.6617
P27635	RPL10	6.05E-03	1.6038
P62906	RPL10A	6.08E-03	1.6004
P50914	RPL14	4.49E-02	1.8567
P18621	RPL17	2.69E-03	1.9529
Q02543	RPL18A	1.71E-02	2.3067
P46778	RPL21	2.14E-02	1.6330
P61254	RPL26	3.57E-03	2.2458
P46776	RPL27A	8.03E-03	1.6982
P39023	RPL3	2.75E-02	1.8091
P46783	RPS10	5.62E-04	-1.5091
P62280	RPS11	2.16E-03	1.7469
P62277	RPS13	6.78E-03	1.5234
P62841	RPS15	3.06E-02	-1.7171

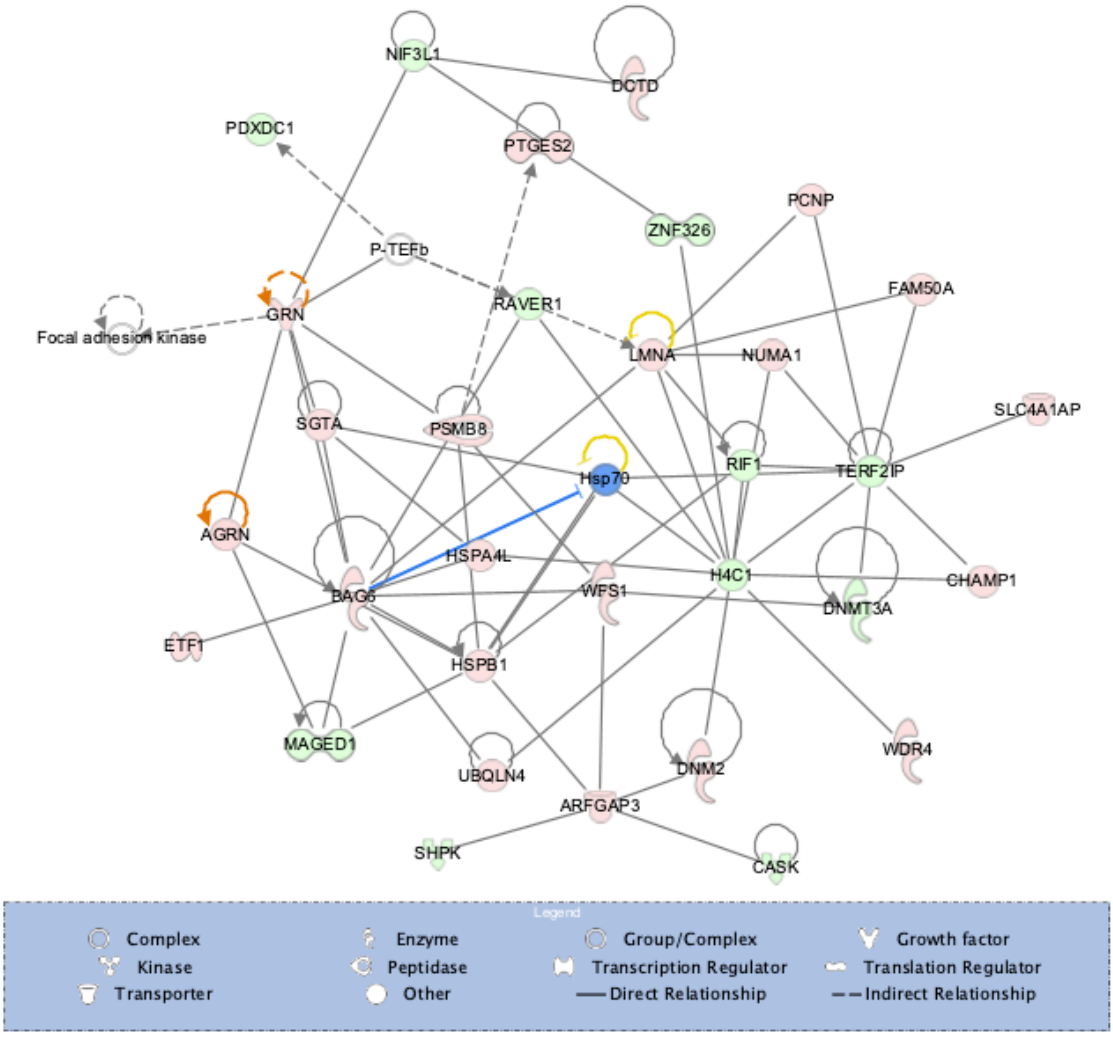
Accession	Gene Name	Anova (p)	Fold change
P62244	RPS15A	1.27E-04	-2.1952
P62269	RPS18	9.84E-04	1.5460
P62266	RPS23	1.63E-02	1.5321
P62847	RPS24	1.87E-03	1.6263
P46781	RPS9	1.06E-02	1.8225
P08865	RPSA	5.38E-04	-2.3238
P31350	RRM2	4.32E-04	-2.4552
Q14684	RRP1B	3.26E-03	-1.6095
O76021	RSL1D1	2.76E-02	1.6163
Q96T51	RUFY1	2.95E-04	-1.7178
Q99584	S100A13	3.03E-06	1.9011
P25815	S100P	7.82E-05	2.2022
Q9NR31	SAR1A	6.92E-03	-1.5720
O00767	SCD	3.31E-03	-1.7093
P22307	SCP2	6.57E-03	-1.5726
Q9HB40	SCPEP1	2.41E-05	1.9034
Q12765	SCRN1	3.26E-04	1.8818
P34741	SDC2	2.91E-03	3.5843
Q15437	SEC23B	1.18E-02	-1.9805
Q13228	SELENBP1	1.05E-02	-1.5544
P01009	SERPINA1	7.43E-04	-1.9278
P01011	SERPINA3	1.82E-06	5.4778
P05154	SERPINA5	3.19E-04	-1.9761
P50453	SERPINB9	1.99E-02	1.5421
P05121	SERPINE1	1.30E-06	23.5293
Q86TU7	SETD3	1.71E-04	-1.5558
P31947	SFN	2.38E-06	4.0214
Q96NB2	SFXN2	1.69E-03	-1.9242
Q9BWM7	SFXN3	4.49E-05	-2.0949
O43765	SGTA	7.51E-03	1.5273
Q9UJC5	SH3BGRL2	4.65E-04	-1.5292
P34896	SHMT1	5.45E-03	-1.5340
Q9UHJ6	SHPK	2.71E-02	-1.6081
Q9NXA8	SIRT5	1.19E-03	-1.5027
P53985	SLC16A1	2.15E-03	-1.6148
O15427	SLC16A3	3.94E-03	1.6482
P11166	SLC2A1	1.70E-05	3.0384
P11169	SLC2A3	1.21E-04	1.8701
Q9Y6M5	SLC30A1	6.92E-03	1.5872
Q96QD8	SLC38A2	7.01E-06	2.7610
Q8WUX1	SLC38A5	3.69E-04	-1.7148
Q6P5W5	SLC39A4	1.03E-05	-3.0342
Q9BWU0	SLC4A1AP	3.03E-05	1.8431
Q9UN76	SLC6A14	1.16E-03	4.2169
P28370	SMARCA1	1.63E-03	1.5955
B2RUZ4	SMIM1	3.30E-04	-1.8419
Q9NXE4	SMPD4	1.44E-03	-2.0494
Q9Y5X1	SNX9	1.44E-03	1.5678
O94875	SORBS2	4.45E-05	-1.8817
Q00796	SORD	1.44E-04	-1.7418
Q92673	SORL1	2.19E-04	2.6969
O60271	SPAG9	9.49E-03	1.7694
O15269	SPTLC1	1.22E-03	-1.6737
Q9NUV7	SPTLC3	1.22E-06	-3.7379
Q14534	SQLE	3.76E-03	-1.5085
Q9Y6N5	SQOR	6.79E-05	2.4082
P12931	SRC	1.49E-04	1.5500
Q13586	STIM1	3.99E-03	-1.7063
Q9P289	STK26	2.24E-03	1.6884
Q9UEW8	STK39	1.03E-04	1.8385
P27105	STOM	3.20E-05	2.3637
Q13033	STRN3	2.03E-04	1.5643
P53597	SUCLG1	5.48E-04	-1.5028
P50225	SULT1A1	2.14E-04	-2.2158
P0DMN0	SULT1A4	3.94E-04	-1.5124
Q06520	SULT2A1	5.05E-05	-1.7009
Q9UH99	SUN2	4.50E-03	-1.5841
Q7KZ85	SUPT6H	1.60E-04	-1.6021
P57105	SYNJ2BP	6.83E-03	-1.5326
Q01995	TAGLN	2.30E-05	6.3529
P26639	TARS1	6.97E-04	1.5974
O60343	TBC1D4	7.79E-03	1.5274
O75764	TCEA3	8.73E-05	-1.6164
Q13428	TCOF1	1.59E-03	1.5417
Q9NYB0	TERF2IP	1.42E-02	-1.8067
Q9UGI8	TES	6.28E-03	2.0405
Q12800	TFCP2	5.72E-04	-1.8352

Accession	Gene Name	Anova (p)	Fold change
P10646	TFPI	9.65E-03	1.9870
P21980	TGM2	2.25E-03	1.6715
P07996	THBS1	2.75E-05	1.9345
Q9Y5L4	TIMM13	2.23E-04	-1.6788
O14925	TIMM23	3.41E-04	-1.7075
O60220	TIMM8A	1.39E-03	-1.5388
P35625	TIMP3	1.98E-03	-4.0646
Q07157	TJP1	9.64E-03	1.5171
Q3LXA3	TKFC	1.66E-04	-1.8237
Q9BXJ8	TMEM120A	1.76E-03	-1.9448
Q96HV5	TMEM41A	9.73E-04	-1.7897
Q9BTV4	TMEM43	2.22E-03	-1.5012
Q68CZ2	TNS3	3.21E-03	1.5631
P11387	TOP1	2.71E-02	1.5885
P11388	TOP2A	3.85E-04	-2.0674
Q53FA7	TP53I3	1.00E-03	1.7747
P14373	TRIM27	5.33E-04	1.7925
Q7Z2T5	TRMT1L	7.26E-04	1.5995
Q96Q11	TRNT1	1.71E-02	2.0470
Q2NL82	TSR1	6.99E-03	1.5195
Q13509	TUBB3	1.14E-06	2.4448
P23258	TUBG1	1.84E-04	-1.6294
Q12792	TWF1	7.96E-03	1.5063
Q9NNW7	TXNRD2	5.98E-03	-1.5567
P04818	TYMS	9.35E-04	-5.2486
P26368	U2AF2	2.05E-03	-1.7611
Q16222	UAP1	3.63E-03	-1.6406
P61077	UBE2D3	2.82E-02	1.6094
P62253	UBE2G1	1.36E-04	2.2413
P60604	UBE2G2	7.44E-03	-2.3886
Q9H832	UBE2Z	4.09E-03	1.8844
Q9BZL1	UBL5	3.71E-03	-1.5442
Q9NRR5	UBQLN4	1.07E-02	1.5609
Q9NUQ7	UFSP2	1.02E-02	-1.5111
Q6UWM9	UGT2A3	1.58E-03	-1.5902
O75310	UGT2B11	3.09E-05	-2.7393
Q96T88	UHRF1	2.14E-06	-5.5880
P11172	UMPS	4.38E-04	-1.5905
Q93008	USP9X	5.63E-03	2.0354
Q08AM6	VAC14	1.06E-03	1.8726
Q5ST30	VARS2	9.75E-04	-1.6255
Q6EMK4	VASN	1.83E-03	-1.5184
P50552	VASP	1.75E-04	1.6365
Q9UK41	VPS28	7.33E-03	2.0395
A3KMH1	VWA8	1.18E-03	-1.6350
P23381	WARS1	7.26E-03	1.7266
P57081	WDR4	2.66E-03	1.5739
O15213	WDR46	5.58E-03	-1.5402
O76024	WFS1	1.76E-05	2.1360
O75191	XYLB	1.52E-02	-1.5476
P54577	YARS1	1.75E-04	1.5897
Q04917	YWHAH	2.24E-04	1.5248
Q9ULU4	ZMYND8	1.35E-03	1.6984
Q5BKZ1	ZNF326	1.58E-03	-1.6047
Q15942	ZYX	1.97E-04	1.7422

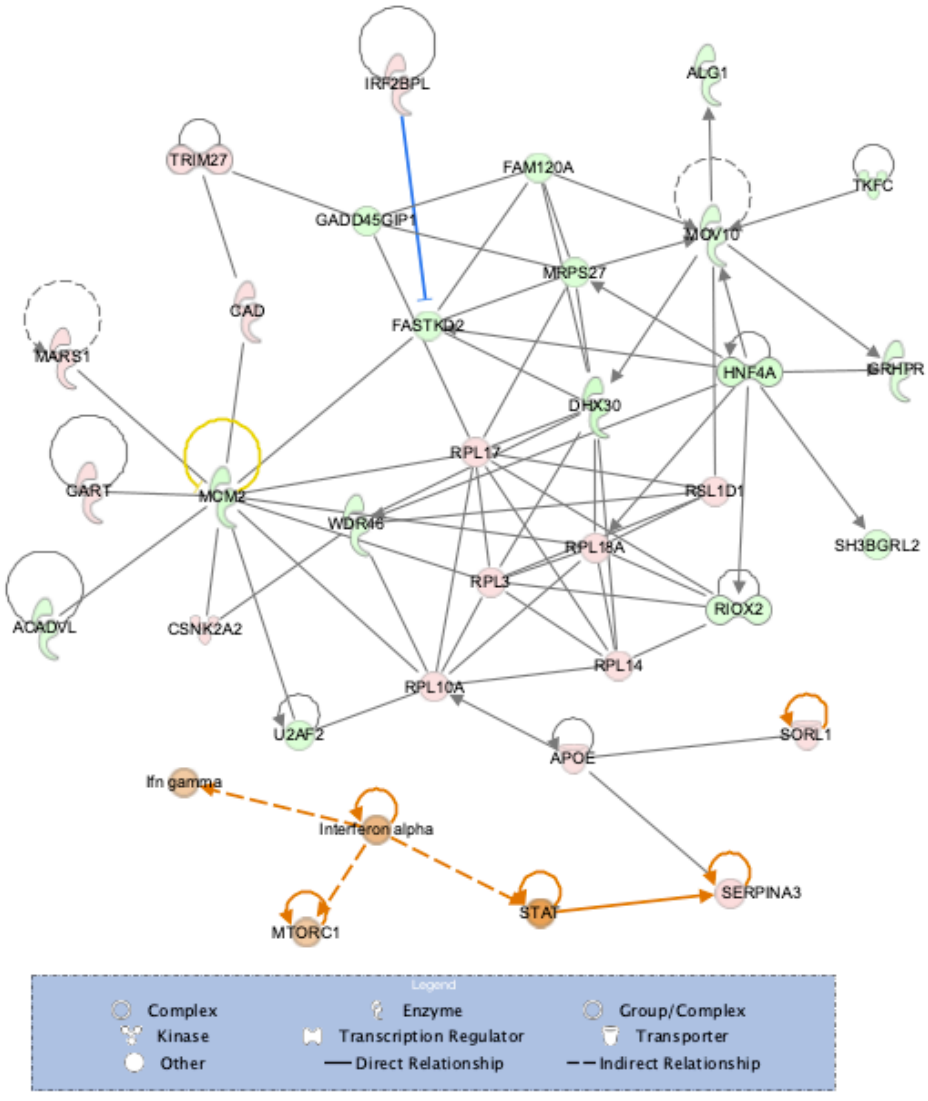
Appendix AA *Network Analysis of the differentially expressed proteins within the DHFR2 knockout cell line revealed that the small molecule biochemistry, lipid metabolism, cell morphology network was one of the most impacted. Drawn Using the IPA Software. Green represents those that are down-regulated, light green represents slightly down-regulated, light red represents slightly up-regulated, orange arrow leads to activation, blue arrow leads to inhibition and white represents those that were added from the Ingenuity Knowledge Base.*



Appendix AB *Network Analysis of the differentially expressed proteins within the DHFR2 knockout cell line revealed that the hereditary disorder, neurological disease, organismal injury and abnormalities network was one of the most impacted. Drawn Using the IPA Software. Green represents those that are down-regulated, light green represents slightly down-regulated, light red represents slightly up-regulated, blue represents predicted inhibition, orange arrow leads to activation, blue arrow leads to inhibition, yellow arrow represents the findings inconsistent with state of downstream molecule and white represents those that were added from the Ingenuity Knowledge Base.*



Appendix AC *Network Analysis of the differentially expressed proteins within the DHFR2 knockout cell line revealed that the cellular assembly and organisation, developmental disorder, hereditary disorder network was one of the most impacted. Drawn Using the IPA Software**.** Green represents those that are down-regulated, light green represents slightly down-regulated, light red represents slightly up-regulated, orange represents predicted activation, orange arrow leads to activation, blue arrow leads to inhibition and yellow arrow represents the findings inconsistent with state of downstream molecule.*



Appendix AD *Differentially expressed proteins identified in the HepG2 DHFR knockdown cell line when compared to HepG2 wildtype.*

Accession	Gene Name	Anova (p)	Fold change
Q9NRG9	AAAS	1.40E-02	-1.74
Q86V21	AACS	1.39E-02	1.69
Q9UDR5	AASS	2.82E-04	-1.84
P80404	ABAT	3.88E-04	-1.99
P08183	ABCB1	1.96E-04	1.95
P33527	ABCC1	2.56E-03	1.66
Q9UG63	ABCF2	1.10E-03	2.25
P42765	ACAA2	5.62E-04	-1.66
Q9H3P7	ACBD3	2.31E-06	1.67
P53396	ACLY	8.89E-05	1.53
Q86TX2	ACOT1	5.08E-04	-1.58
Q15067	ACOX1	7.81E-04	-1.90
Q99424	ACOX2	1.82E-04	-2.17
Q96CM8	ACSF2	9.69E-04	-1.76
P33121	ACSL1	2.28E-04	-1.92
O95573	ACSL3	2.15E-04	1.77
O60488	ACSL4	6.36E-05	1.60
Q9ULC5	ACSL5	4.66E-05	-2.09
Q08AH3	ACSM2A	5.39E-06	-5.51
Q68CK6	ACSM2B	7.16E-06	-5.65
Q9NR19	ACSS2	1.55E-05	2.26
O96019	ACTL6A	1.23E-02	-1.56
O43707	ACTN4	1.74E-04	-1.51
P08319	ADH4	3.12E-05	2.98
P28332	ADH6	3.70E-04	-1.52
Q16186	ADRM1	7.33E-04	1.56
O00116	AGPS	1.39E-05	3.80
O00468	AGRN	5.07E-07	-8.33
P21549	AGXT	1.21E-03	-1.67
O95831	AIFM1	4.55E-03	-1.69
P54819	AK2	1.42E-03	-1.52
P27144	AK4	2.37E-03	-1.77
P14550	AKR1A1	7.75E-04	-1.78
O60218	AKR1B10	1.65E-08	5.19
P13716	ALAD	2.86E-04	-1.82
P00352	ALDH1A1	4.53E-05	-1.54
P30837	ALDH1B1	4.87E-03	-1.64
O75891	ALDH1L1	2.40E-04	-2.30
P51648	ALDH3A2	5.56E-03	-1.57
P51649	ALDH5A1	2.70E-03	-1.58
Q02252	ALDH6A1	2.15E-04	-1.92
P09972	ALDOC	2.22E-03	-1.55
O94973	AP2A2	1.94E-03	-2.42
Q06481	APLP2	6.73E-04	1.78
P02652	APOA2	1.31E-04	1.61
P04114	APOB	2.65E-03	-1.62
P02655	APOC2	1.44E-03	1.82
P02749	APOH	1.17E-02	-2.26
O60306	AQR	4.58E-03	-1.73
P53365	ARFIP2	1.68E-02	1.95
P78540	ARG2	1.96E-06	3.27
Q92888	ARHGEF1	6.12E-04	1.61
P40616	ARL1	2.53E-03	2.00
P49407	ARRB1	9.47E-04	2.00
P15848	ARSB	1.51E-05	1.59
P51690	ARSL	9.11E-03	-1.97
P07306	ASGR1	2.04E-03	-1.59
P07307	ASGR2	1.58E-03	-1.79
P04424	ASL	2.58E-04	1.52
P08243	ASNS	6.09E-05	4.20
Q9BZE9	ASPSCR1	1.25E-03	-1.66
Q7L266	ASRGL1	5.23E-06	3.76
P25705	ATP5F1A	4.61E-03	-1.53
Q93050	ATP6V0A1	5.64E-05	1.77
P61421	ATP6V0D1	1.10E-02	1.57
P21283	ATP6V1C1	1.58E-03	1.63
O14965	AURKA	3.35E-05	3.03
P46379	BAG6	5.23E-05	1.81
Q9UQB8	BAIAP2	8.20E-05	2.14
Q16611	BAK1	1.27E-03	1.98
Q9UIG0	BAZ1B	1.03E-02	2.20
P50895	BCAM	1.95E-04	-1.98
P54687	BCAT1	4.12E-04	1.51
Q9P287	BCCIP	8.71E-03	1.58
P21953	BCKDHB	2.66E-03	-1.67
Q02338	BDH1	2.67E-03	1.57
P54132	BLM	2.41E-05	1.78
Q14692	BMS1	6.01E-04	1.56
Q86WA6	BPHL	1.90E-05	-1.82
Q15059	BRD3	2.37E-05	-1.62
Q9H0W9	C11orf54	5.68E-05	-1.76
A6NCS6	C2orf72	4.90E-03	-1.75
P01024	C3	2.46E-05	-1.76
Q5T6V5	C9orf64	3.73E-03	-1.54
P35218	CA5A	5.17E-05	-2.05
Q9BY67	CADM1	2.41E-04	-2.79
Q14012	CAMK1	2.12E-06	-1.64
O14936	CASK	1.77E-03	-2.22
P20810	CAST	2.99E-03	1.58
P04040	CAT	7.19E-05	-1.84
P16152	CBR1	1.34E-03	-2.63
Q8N4T8	CBR4	1.80E-04	-2.38
Q8IX12	CCAR1	1.03E-03	-1.59
P14635	CCNB1	4.27E-04	5.09
Q5ZPR3	CD276	2.67E-02	-1.76
P13987	CD59	2.16E-03	-1.56
P08962	CD63	1.21E-04	1.96
P04233	CD74	6.35E-05	-2.22
Q9UJX2	CDC23	2.67E-02	2.05
P19022	CDH2	1.70E-06	-4.04
Q00534	CDK6	1.21E-03	-2.11
O95674	CDS2	1.94E-03	-1.51
P49715	CEBPA	8.30E-03	1.59
Q92879	CELF1	9.09E-04	-1.57
Q9UHN6	CEMIP2	6.48E-05	1.69
Q9HA82	CERS4	1.20E-02	-1.84
Q9UEE9	CFDP1	3.85E-04	-2.23
Q8N4Q1	CHCHD4	2.51E-03	-2.15
Q8NE62	CHDH	2.31E-03	-1.84
Q9BWS9	CHID1	3.27E-05	-1.51
Q8NET6	CHST13	3.58E-05	-2.55
Q9Y3D0	CIAO2B	2.45E-03	-1.55
P12277	CKB	1.19E-05	2.69
O14967	CLGN	8.80E-04	-2.07
O00299	CLIC1	7.36E-05	1.65
Q9Y696	CLIC4	5.39E-03	-1.52
P09497	CLTB	1.20E-04	1.72
P10909	CLU	2.24E-05	-2.10
P09543	CNP	4.94E-05	-1.64
Q9NYJ1	COA4	2.18E-04	-2.27
Q96BR5	COA7	9.66E-04	-2.63
Q8WTW3	COG1	1.21E-04	1.70
P83436	COG7	8.01E-03	1.50
P39060	COL18A1	5.54E-04	-2.57
Q5HYK3	COQ5	3.22E-04	-2.95
Q8NI60	COQ8A	7.31E-05	-1.91
O75208	COQ9	2.16E-03	-2.36
Q9BR76	CORO1B	1.33E-03	1.58
Q14019	COTL1	4.14E-06	4.04
P24311	COX7B	5.77E-04	-1.81
O75131	CPNE3	1.08E-03	-1.54
Q9BRF8	CPPED1	2.05E-03	-1.63
P43155	CRAT	1.38E-03	-1.53
P50238	CRIP1	4.02E-05	-3.24
P35221	CTNNA1	7.71E-06	-1.68
Q8WYA6	CTNNBL1	6.94E-04	-2.10
O60716	CTNND1	2.22E-05	-1.76
Q69YN2	CWF19L1	1.50E-02	1.79
P00167	CYB5A	1.06E-03	-1.55
Q8TAV3	CYP2W1	1.20E-03	-1.66
P51398	DAP3	9.82E-03	-1.52
Q96EP5	DAZAP1	2.26E-02	-1.51
Q16643	DBN1	1.32E-06	1.80
P11182	DBT	9.15E-03	-2.39
Q7Z4W1	DCXR	1.13E-05	-1.69
Q92466	DDB2	3.30E-05	-2.36
Q9H0S4	DDX47	1.57E-02	-1.61
Q16698	DECR1	1.60E-02	-1.66
O75907	DGAT1	3.74E-05	-1.82
Q15392	DHCR24	6.32E-03	-1.87
P00374	DHFR	1.85E-02	-1.63
Q96LJ7	DHRS1	1.66E-03	-2.05
Q7L2E3	DHX30	1.00E-02	-2.18
Q92620	DHX38	4.82E-03	-1.51
O60610	DIAPH1	1.90E-02	-1.77
O60832	DKC1	1.52E-03	-1.52
Q9NXW2	DNAJB12	4.64E-04	1.76
Q13217	DNAJC3	4.91E-04	1.52
Q96N67	DOCK7	1.55E-02	1.88
Q14126	DSG2	8.29E-03	-1.52
P60981	DSTN	1.00E-04	1.78
P33316	DUT	5.99E-04	-1.69
Q99848	EBNA1BP2	3.69E-05	1.60
P30084	ECHS1	3.32E-04	-1.70
O75521	ECI2	1.73E-03	-2.61
Q5VYK3	ECPAS	2.04E-02	1.79
Q15029	EFTUD2	1.03E-04	-1.52
Q8NDI1	EHBP1	1.09E-03	2.49
Q9H223	EHD4	1.60E-03	1.76
Q08426	EHHADH	3.09E-03	-1.54
Q9BY44	EIF2A	1.26E-05	1.53
Q9UI10	EIF2B4	2.67E-03	1.56
O75822	EIF3J	1.67E-04	1.62
P38919	EIF4A3	1.28E-03	-1.54
P78344	EIF4G2	1.78E-04	1.52
Q8IZ81	ELMOD2	5.65E-04	1.66
O95163	ELP1	5.97E-04	-1.56
Q9HC35	EML4	1.63E-04	1.69
P22413	ENPP1	4.24E-05	2.85
O14638	ENPP3	1.50E-03	-1.54
O43491	EPB41L2	1.90E-06	1.52
P16422	EPCAM	7.10E-04	-2.13
P29317	EPHA2	5.72E-05	2.30
P07099	EPHX1	5.39E-03	-1.56
P34913	EPHX2	4.41E-04	-1.88
Q9H6S3	EPS8L2	3.27E-03	-1.59
O75616	ERAL1	2.44E-02	-1.52
Q9UKR5	ERG28	5.38E-03	1.50
Q7Z2K6	ERMP1	7.41E-04	1.73
Q01780	EXOSC10	2.35E-04	1.52
Q9Y624	F11R	4.84E-03	-1.90
P00734	F2	9.23E-06	2.32
Q9NZB2	FAM120A	1.39E-02	-1.68
Q96C01	FAM136A	1.08E-03	-1.62
Q96A26	FAM162A	2.40E-03	-2.43
Q9Y6X4	FAM169A	1.46E-03	1.57
Q14320	FAM50A	1.05E-04	1.85
Q9UK22	FBXO2	4.95E-05	-2.06
P55899	FCGRT	5.11E-04	-2.00
P37268	FDFT1	1.73E-03	1.74
P62942	FKBP1A	4.21E-03	-1.67
Q13045	FLII	2.72E-03	1.64
O75955	FLOT1	3.07E-03	1.53
P49326	FMO5	7.79E-08	-9.61
P02751	FN1	3.70E-04	-1.59
Q16658	FSCN1	6.60E-04	-1.53
O95954	FTCD	7.38E-05	-1.74

Accession	Gene Name	Anova (p)	Fold change
Q96I24	FUBP3	1.37E-03	-1.79
P51114	FXR1	5.62E-04	1.64
Q8TAE8	GADD45GIP1	4.98E-03	-1.64
P51570	GALK1	3.33E-04	-1.65
Q96C23	GALM	1.04E-04	-1.56
Q14353	GAMT	6.07E-04	-1.69
P41250	GARS1	1.59E-05	1.61
A0A0B4J2D	GATD3B	7.87E-04	-1.79
Q04446	GBE1	9.94E-03	1.55
Q92947	GCDH	1.63E-04	-2.18
P30047	GCHFR	6.99E-03	-1.69
P48506	GCLC	1.86E-02	1.67
P48507	GCLM	1.09E-03	2.00
Q7L5D6	GET4	4.26E-03	1.54
P32189	GK	1.17E-06	1.93
P23378	GLDC	4.19E-04	-1.77
Q9HC38	GLOD4	9.92E-04	-1.62
P35754	GLRX	5.30E-05	1.92
P00367	GLUD1	1.41E-04	-1.99
O60547	GMDS	1.29E-06	2.46
O75496	GMNN	1.98E-05	3.44
Q9Y5P6	GMPPB	3.01E-03	1.63
P29992	GNA11	1.79E-03	-2.36
P04899	GNAI2	2.59E-03	-1.50
Q5JWF2	GNAS	7.24E-03	-1.56
P62873	GNB1	6.20E-03	-1.67
Q9Y223	GNE	6.26E-04	2.33
Q13823	GNL2	4.37E-04	2.41
P15586	GNS	7.33E-07	2.07
Q08379	GOLGA2	7.21E-04	1.98
Q08378	GOLGA3	2.30E-04	2.65
Q13439	GOLGA4	6.20E-05	1.64
Q8TBA6	GOLGA5	4.13E-03	1.55
Q14789	GOLGB1	4.18E-03	1.52
O14653	GOSR2	2.56E-05	1.87
P17174	GOT1	2.02E-05	2.11
P51654	GPC3	1.73E-07	-234.55
Q14451	GRB7	3.78E-06	-3.53
Q9UBQ7	GRHPR	2.92E-04	-1.63
P28799	GRN	3.00E-03	-2.23
P08263	GSTA1	1.76E-03	1.58
Q9Y2Q3	GSTK1	1.16E-03	-1.65
P21266	GSTM3	1.78E-06	-4.12
P78417	GSTO1	6.45E-05	1.67
Q00403	GTF2B	9.08E-04	1.76
Q9BZE4	GTPBP4	6.79E-04	1.67
P16402	H1-3	3.37E-04	2.05
O60814	H2BC12	7.56E-05	-2.23
P62805	H4C1	2.00E-03	-1.99
Q9P035	HACD3	6.03E-03	-1.98
Q16836	HADH	6.66E-04	-1.76
P55084	HADHB	2.81E-03	-1.54
P12081	HARS1	9.36E-05	1.51
P69905	HBA1	2.57E-02	1.51
P53701	HCCS	3.62E-03	-1.94
Q9BSH5	HDHD3	1.84E-04	-2.97
Q9ULT8	HECTD1	3.16E-06	2.25
Q9NRZ9	HELLS	3.25E-03	1.70
P06865	HEXA	1.89E-04	1.71
P31937	HIBADH	1.51E-02	-1.59
P52789	HK2	1.87E-07	1.67
Q2TB90	HKDC1	4.31E-03	1.66
P04439	HLA-A	3.19E-02	-3.31
P17096	HMGA1	1.91E-04	1.81
P52926	HMGA2	3.53E-03	-2.19
Q01581	HMGCS1	1.25E-05	1.80
P50135	HNMT	1.64E-03	-2.41
P22626	HNRNPA2B1	3.25E-03	-1.60
O14979	HNRNPDL	1.28E-03	-1.58
P55795	HNRNPH2	3.68E-03	-1.71
P14866	HNRNPL	1.63E-03	-1.67
O43390	HNRNPR	3.59E-03	-1.55
Q86VS8	HOOK3	1.03E-02	1.57
P00738	HP	1.43E-03	1.61
P37235	HPCAL1	9.48E-04	1.78
P32754	HPD	4.14E-04	2.21
P00492	HPRT1	4.32E-04	1.60

Accession	Gene Name	Anova (p)	Fold change
Q99714	HSD17B10	9.50E-03	-1.54
P51659	HSD17B4	8.53E-03	-1.52
Q92506	HSD17B8	4.75E-05	-2.07
Q6YN16	HSDL2	1.09E-03	-1.73
P10809	HSPD1	4.68E-03	-1.54
Q7Z6Z7	HUWE1	1.39E-02	-1.54
P05362	ICAM1	2.18E-04	1.77
O43837	IDH3B	1.55E-02	-1.73
P01344	IGF2	1.72E-03	2.07
P11717	IGF2R	2.41E-05	1.66
Q8N6C5	IGSF1	1.45E-02	-1.66
A1L0T0	ILVBL	1.07E-03	-1.63
Q16891	IMMT	4.99E-03	-1.55
Q9NV31	IMP3	2.53E-04	-5.11
P46940	IQGAP1	4.69E-04	1.56
Q7Z5L9	IRF2BP2	4.01E-02	1.52
P05161	ISG15	9.28E-03	1.57
Q9NPH2	ISYNA1	6.83E-05	-2.23
P17301	ITGA2	5.82E-04	1.53
P23229	ITGA6	1.49E-03	-1.70
Q14571	ITPR2	3.80E-04	-1.80
Q6NYC1	JMJD6	1.64E-04	2.49
Q92945	KHSRP	8.09E-04	-1.66
Q9BW19	KIFC1	6.39E-06	2.74
Q07866	KLC1	1.19E-02	1.55
P52292	KPNA2	2.72E-06	2.07
P04264	KRT1	2.07E-02	-2.98
P13645	KRT10	6.81E-03	-48.67
P05783	KRT18	3.04E-05	1.83
P05787	KRT8	2.13E-05	1.80
P35527	KRT9	2.30E-02	-3.10
O15230	LAMA5	2.39E-06	-4.70
P07942	LAMB1	8.96E-04	-1.68
P55268	LAMB2	9.77E-07	-8.15
P11047	LAMC1	2.48E-05	-3.01
O43813	LANCL1	5.26E-05	-1.91
Q71RC2	LARP4	2.23E-04	4.12
Q15031	LARS2	1.04E-02	-2.28
P09382	LGALS1	6.77E-09	3.69
P17931	LGALS3	3.51E-11	21.38
Q08380	LGALS3BP	2.40E-02	-1.57
Q6ZN17	LIN28B	8.51E-05	-1.91
Q9BU23	LMF2	5.11E-03	-2.03
Q32MZ4	LRRFIP1	2.87E-03	1.64
Q86X29	LSR	3.10E-04	2.22
Q96GA3	LTV1	4.61E-05	3.61
P61626	LYZ	2.06E-07	2.42
Q9BQ69	MACROD1	1.04E-02	-3.06
Q9P0M6	MACROH2A2	2.86E-03	-1.72
P21397	MAOA	3.07E-05	-3.23
P27338	MAOB	3.95E-05	-2.66
Q02750	MAP2K1	4.73E-03	1.54
P27816	MAP4	5.04E-04	1.51
P29966	MARCKS	6.28E-05	-2.18
Q96GW9	MARS2	4.04E-03	1.59
P31153	MAT2A	9.77E-05	2.13
Q96RQ3	MCCC1	2.06E-03	-1.78
Q9HCC0	MCCC2	7.25E-03	-1.57
P49736	MCM2	1.21E-04	1.54
Q14566	MCM6	3.10E-04	1.63
P33993	MCM7	3.98E-04	1.66
P40925	MDH1	4.15E-03	-1.77
Q9H8H3	METTL7A	3.53E-04	-1.64
P55081	MFAP1	1.50E-02	-1.52
O95140	MFN2	4.48E-03	2.08
Q9UNW1	MINPP1	1.05E-04	-1.70
P46013	MKI67	2.52E-04	1.57
O95822	MLYCD	1.92E-03	-2.20
Q96EY8	MMAB	2.69E-03	-1.68
P22033	MMUT	1.77E-03	-1.58
Q9Y3A3	MOB4	4.25E-04	1.76
Q9HCE1	MOV10	2.20E-05	-1.59
O00566	MPHOSPH10	4.17E-03	-1.51
P52815	MRPL12	6.12E-03	1.51
Q9NX20	MRPL16	8.92E-03	-1.54
Q9NRX2	MRPL17	1.12E-02	-1.79
Q5T653	MRPL2	2.56E-02	1.97

Accession	Gene Name	Anova (p)	Fold change
Q9BQ48	MRPL34	3.09E-03	1.53
Q9BZE1	MRPL37	8.43E-03	-1.63
Q96DV4	MRPL38	5.74E-03	-1.66
Q9BRJ2	MRPL45	2.54E-02	-1.64
Q9Y676	MRPS18B	5.43E-03	-1.58
Q9Y3D9	MRPS23	3.04E-03	-1.57
P82663	MRPS25	4.80E-03	-1.53
Q92552	MRPS27	2.32E-03	-2.11
Q9NP92	MRPS30	7.39E-03	-1.85
P82675	MRPS5	8.29E-04	-2.88
Q9Y2R9	MRPS7	2.88E-02	-1.65
Q13330	MTA1	1.14E-02	1.53
Q5VT66	MTARC1	1.81E-03	-1.53
Q9Y6C9	MTCH2	1.38E-03	-1.59
P55157	MTTP	7.55E-06	-1.80
O75431	MTX2	1.31E-02	-1.91
P53602	MVD	4.59E-06	2.08
Q7Z406	MYH14	1.77E-05	-2.56
O00159	MYO1C	2.08E-03	1.81
Q9BXJ9	NAA15	7.35E-03	1.56
Q9UK23	NAGPA	1.87E-02	1.53
P43490	NAMPT	2.63E-03	1.56
Q9NR45	NANS	3.93E-04	-1.61
Q99747	NAPG	9.08E-04	1.57
P49321	NASP	1.55E-03	-1.61
Q8NCW5	NAXE	1.77E-02	-1.52
Q6PIU2	NCEH1	2.30E-04	1.88
Q9UI09	NDUFA12	4.41E-03	-1.77
O95182	NDUFA7	2.80E-03	-1.65
O14561	NDUFAB1	1.59E-02	-1.51
Q9P032	NDUFAF4	3.10E-03	1.58
Q9Y6M9	NDUFB9	1.46E-02	-1.63
O75380	NDUFS6	1.80E-03	-1.91
P49821	NDUFV1	1.04E-03	-1.99
P19404	NDUFV2	2.44E-02	-1.58
P14543	NID1	2.31E-06	-5.99
Q9GZT8	NIF3L1	2.86E-03	-2.14
Q9BYG3	NIFK	5.52E-05	1.93
O75323	NIPSNAP2	1.00E-04	-2.01
Q9ULX3	NOB1	4.41E-06	1.62
Q9H8H0	NOL11	9.48E-03	-1.84
P78316	NOP14	1.94E-04	1.61
Q9Y3C1	NOP16	1.24E-04	1.53
Q86U38	NOP9	6.44E-04	-2.16
Q8TAT6	NPLOC4	4.70E-04	1.64
P16083	NQO2	2.07E-02	1.52
O95478	NSA2	5.16E-05	8.24
Q9H857	NT5DC2	1.35E-02	-2.63
P21589	NT5E	1.15E-02	2.03
Q9BSD7	NTPCR	2.55E-03	-1.70
Q9H1E3	NUCKS1	1.99E-03	-2.09
Q9Y266	NUDC	3.37E-04	1.59
Q9BRJ7	NUDT16L1	7.32E-03	-1.71
Q14980	NUMA1	1.21E-02	-2.05
P57740	NUP107	7.40E-04	-1.51
Q8WUM0	NUP133	2.11E-03	-1.51
Q12769	NUP160	1.58E-03	-1.94
Q92621	NUP205	3.76E-03	-1.55
Q8TEM1	NUP210	5.41E-04	-2.24
Q9NX40	OCIAD1	3.89E-03	-1.72
O60313	OPA1	5.67E-03	-1.56
Q96FW1	OTUB1	8.87E-03	-2.52
O95747	OXSR1	1.41E-02	-1.55
O15460	P4HA2	1.54E-04	2.29
Q9UKS6	PACSIN3	6.38E-03	-1.56
P00439	PAH	4.14E-06	-1.97
O43252	PAPSS1	2.49E-03	-1.61
Q15365	PCBP1	3.23E-03	-1.86
P05165	PCCA	1.67E-03	-1.67
Q15004	PCLAF	1.43E-04	3.07
Q15154	PCM1	1.14E-03	1.63
P12004	PCNA	2.45E-03	1.56
Q9UHG3	PCYOX1	6.83E-04	-1.68
Q99447	PCYT2	2.36E-03	-1.64
Q14690	PDCD11	4.78E-03	1.57
Q53EL6	PDCD4	2.99E-07	-4.63
Q8WUM4	PDCD6IP	3.31E-03	-1.91

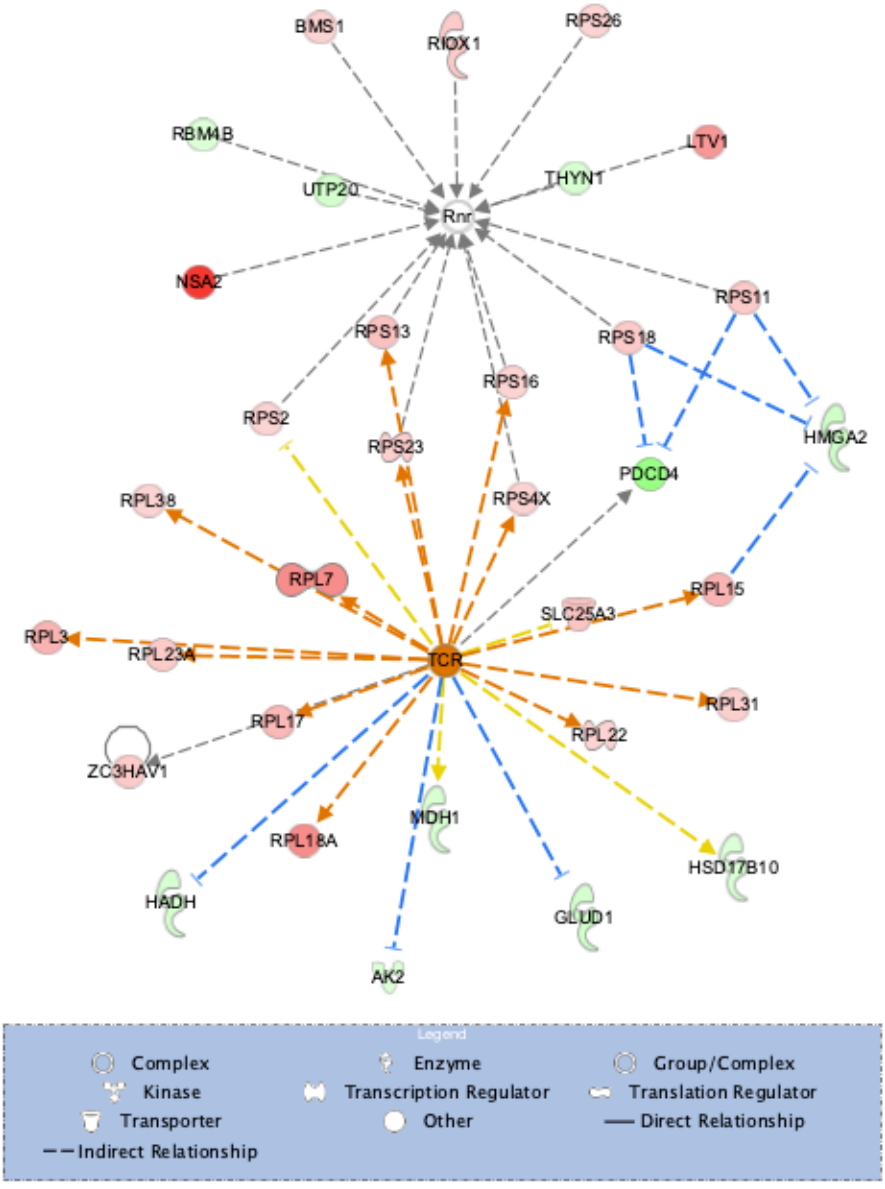
Accession	Gene Name	Anova (p)	Fold change
Q96HC4	PDLIM5	3.87E-06	2.95
Q8NCN5	PDPR	2.43E-03	-1.60
Q9NTI5	PDS5B	1.56E-04	-2.05
Q5T2W1	PDZK1	4.02E-06	-2.19
P30086	PEBP1	6.11E-04	-1.51
Q9UBV8	PEF1	2.11E-02	1.52
Q86TG7	PEG10	2.35E-05	-1.74
Q9BRX2	PELO	3.34E-03	1.76
O00264	PGRMC1	5.40E-03	-1.99
Q9NRX4	PHPT1	5.14E-04	-2.19
Q8TBX8	PIP4K2C	5.37E-05	1.81
Q9P0Z9	PIPOX	2.50E-04	-1.80
Q5JRX3	PITRM1	1.58E-04	2.03
Q13393	PLD1	1.36E-04	1.64
P53350	PLK1	2.45E-05	2.19
O60568	PLOD3	5.87E-04	-1.63
O43660	PLRG1	3.09E-03	-1.60
Q14651	PLS1	8.05E-04	1.59
O15031	PLXNB2	2.43E-03	-1.94
O15305	PMM2	2.26E-03	1.51
Q10713	PMPCA	4.36E-03	-1.55
Q9NRX1	PNO1	2.96E-03	1.63
P00491	PNP	2.33E-03	1.63
P28340	POLD1	4.63E-03	-1.68
O95602	POLR1A	1.14E-04	1.89
P24928	POLR2A	8.25E-05	-1.58
P30876	POLR2B	2.86E-03	1.60
P30405	PPIF	4.05E-04	-1.66
O15355	PPM1G	1.60E-02	-1.61
Q6IN85	PPP4R3A	4.82E-03	2.10
O75170	PPP6R2	6.98E-04	2.14
P50897	PPT1	2.28E-04	-1.57
P32119	PRDX2	1.07E-03	-1.85
P30048	PRDX3	3.24E-03	-1.96
Q13162	PRDX4	1.11E-02	-1.80
P10644	PRKAR1A	3.29E-05	1.81
P17252	PRKCA	1.25E-03	-1.53
Q05655	PRKCD	1.79E-04	-1.85
Q9UF12	PRODH2	1.63E-03	-1.66
P07225	PROS1	2.33E-04	4.44
Q13523	PRPF4B	5.17E-03	1.96
O75475	PSIP1	2.24E-06	-1.55
P25787	PSMA2	1.00E-03	1.53
P62195	PSMC5	1.80E-03	1.50
P55036	PSMD4	1.09E-04	1.59
Q16401	PSMD5	1.11E-04	-2.68
Q06323	PSME1	3.49E-04	-1.57
Q14997	PSME4	4.16E-03	1.56
O95758	PTBP3	1.51E-03	-1.54
O75127	PTCD1	7.55E-04	1.53
Q96EY7	PTCD3	7.87E-03	-1.93
Q9H7Z7	PTGES2	8.81E-04	1.51
Q13308	PTK7	1.61E-04	-2.02
P06454	PTMA	1.33E-02	2.12
Q9Y3E5	PTRH2	5.06E-03	-1.66
P15151	PVR	4.94E-04	1.91
Q13610	PWP1	7.22E-04	1.60
Q96C36	PYCR2	3.19E-03	-1.69
Q6ZRP7	QSOX2	2.22E-03	1.54
Q9BXR0	QTRT1	1.27E-04	-1.71
Q9H974	QTRT2	2.06E-05	-1.80
P61006	RAB8A	5.67E-05	-2.13
Q92696	RABGGTA	9.28E-03	1.56
P53611	RABGGTB	2.82E-03	1.70
P63000	RAC1	2.66E-04	1.53
Q5T160	RARS2	2.77E-03	-1.87
Q8IY67	RAVER1	5.07E-03	-1.98
Q09028	RBBP4	2.73E-04	-1.66
Q16576	RBBP7	1.69E-05	-1.96
O75884	RBBP9	7.10E-03	-1.73
Q9NWB1	RBFOX1	3.81E-04	-1.88
P49756	RBM25	6.45E-03	-1.57
Q5T8P6	RBM26	8.49E-03	1.52
Q9BQ04	RBM4B	5.71E-03	-1.59
Q9Y5S9	RBM8A	1.01E-02	-2.59
P38159	RBMX	1.61E-04	-1.60
Q14257	RCN2	1.12E-03	-1.65

Accession	Gene Name	Anova (p)	Fold change
Q00765	REEP5	5.21E-05	1.57
P78509	RELN	4.88E-06	-3.50
Q9Y3B8	REXO2	2.42E-04	1.75
P35251	RFC1	4.52E-03	-2.54
P40937	RFC5	3.10E-03	-1.68
Q15493	RGN	2.15E-03	-2.21
Q8IXI1	RHOT2	1.11E-02	-1.51
Q9H6W3	RIOX1	2.35E-02	1.80
Q9NWS8	RMND1	3.34E-03	-1.60
Q9Y508	RNF114	7.64E-05	1.65
Q5VTR2	RNF20	3.50E-03	1.90
O75116	ROCK2	7.51E-04	1.73
P27635	RPL10	1.21E-03	2.23
P62906	RPL10A	5.95E-03	1.62
P26373	RPL13	5.38E-03	2.86
P40429	RPL13A	2.49E-03	4.26
P50914	RPL14	1.49E-03	3.31
P61313	RPL15	8.91E-03	2.56
P18621	RPL17	1.06E-03	2.53
Q07020	RPL18	2.58E-03	3.56
Q02543	RPL18A	2.74E-03	3.87
P84098	RPL19	7.59E-03	3.22
P46778	RPL21	2.48E-03	2.12
P35268	RPL22	3.28E-04	1.50
P62750	RPL23A	8.38E-04	1.78
P83731	RPL24	1.41E-03	3.69
P61254	RPL26	1.03E-03	2.59
P61353	RPL27	5.61E-03	2.65
P46776	RPL27A	4.16E-04	2.63
P46779	RPL28	3.05E-03	3.03
P39023	RPL3	9.34E-03	2.56
P62899	RPL31	1.76E-04	1.63
P62910	RPL32	1.32E-03	1.87
P42766	RPL35	6.35E-03	4.01
P18077	RPL35A	7.66E-04	2.27
Q9Y3U8	RPL36	1.52E-03	2.18
Q969Q0	RPL36AL	9.88E-03	2.92
P61513	RPL37A	2.04E-03	1.97
P63173	RPL38	1.58E-03	1.54
P36578	RPL4	7.74E-03	3.38
Q02878	RPL6	5.76E-03	4.34
P18124	RPL7	5.30E-03	3.87
P62424	RPL7A	3.73E-03	4.75
P62917	RPL8	9.61E-03	4.14
P62280	RPS11	1.13E-03	1.88
P62277	RPS13	2.86E-04	2.06
P62249	RPS16	3.75E-04	1.55
P62269	RPS18	7.84E-05	1.83
P15880	RPS2	7.23E-04	1.58
P62266	RPS23	5.11E-03	1.72
P62847	RPS24	7.87E-04	1.64
P62854	RPS26	2.30E-03	1.53
P62979	RPS27A	2.76E-04	-2.05
P62701	RPS4X	1.42E-03	1.54
P22090	RPS4Y1	2.00E-03	1.59
P62753	RPS6	1.48E-03	2.25
P62241	RPS8	9.77E-03	2.53
P46781	RPS9	8.75E-04	2.81
P31350	RRM2	2.21E-05	3.56
Q14684	RRP1B	4.60E-04	1.60
Q9UHA3	RSL24D1	1.08E-03	1.76
A6ZKI3	RTL8C	1.79E-04	3.00
Q99584	S100A13	1.39E-04	2.29
P25815	S100P	1.67E-05	2.95
Q15424	SAFB	2.11E-04	-1.60
Q14151	SAFB2	4.57E-04	-1.50
Q9Y512	SAMM50	5.67E-03	-1.63
Q9UL12	SARDH	2.74E-03	-1.78
Q9NP81	SARS2	1.09E-02	-2.14
Q14108	SCARB2	1.32E-03	1.81
Q8WVM8	SCFD1	8.01E-05	1.54
P22307	SCP2	3.21E-03	-1.53
Q9HB40	SCPEP1	8.00E-05	2.25
O94855	SEC24D	3.41E-04	1.60
Q13228	SELENBP1	5.99E-04	-1.70
Q9NVA2	SEPTIN11	1.62E-03	1.59
Q9UHD8	SEPTIN9	1.41E-03	1.52

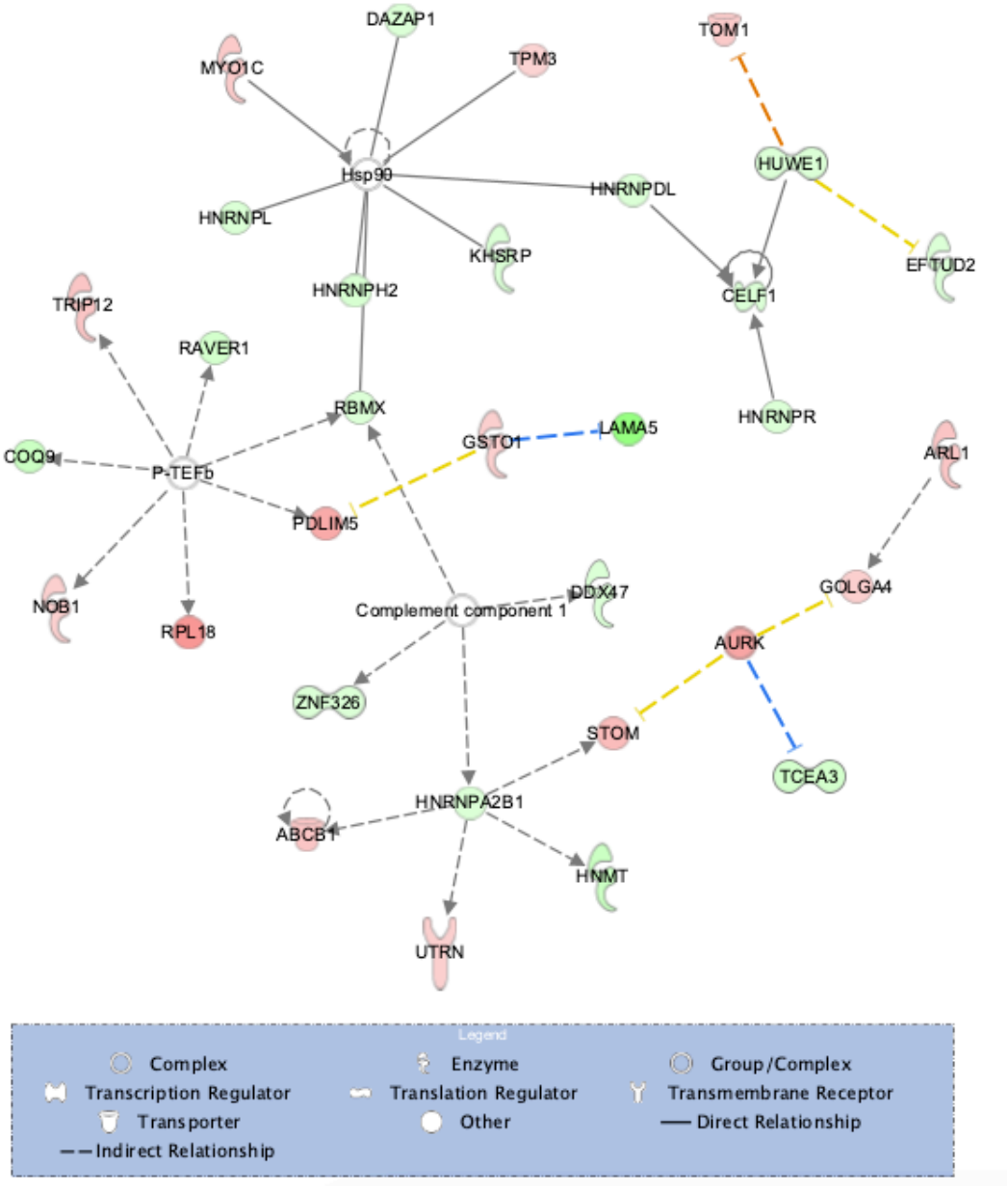
Accession	Gene Name	Anova (p)	Fold change
P01011	SERPINA3	2.54E-04	2.08
P35237	SERPINB6	1.30E-03	1.83
P36955	SERPINF1	5.57E-04	-1.58
Q15459	SF3A1	1.71E-03	-1.59
P31947	SFN	1.29E-08	6.47
Q9Y371	SH3GLB1	6.94E-04	3.14
P34896	SHMT1	5.11E-03	-1.61
A0MZ66	SHTN1	1.05E-03	2.49
Q9NXA8	SIRT5	1.10E-04	-1.60
P63208	SKP1	4.55E-05	-1.60
O15427	SLC16A3	2.69E-03	1.66
O75746	SLC25A12	6.38E-03	-2.32
Q9UJS0	SLC25A13	9.20E-03	-1.60
O43772	SLC25A20	5.73E-03	-1.86
Q00325	SLC25A3	3.93E-03	2.04
O14975	SLC27A2	1.41E-02	1.54
P11169	SLC2A3	3.47E-05	-2.59
Q8WUX1	SLC38A5	1.97E-03	-1.55
P30825	SLC7A1	1.39E-03	1.68
Q6GMV2	SMYD5	6.49E-03	-2.16
Q96DI7	SNRNP40	1.71E-04	-1.60
Q13425	SNTB2	6.17E-03	-2.11
P00441	SOD1	5.82E-04	-1.68
P04179	SOD2	3.58E-04	1.54
O94875	SORBS2	2.19E-03	-1.68
Q00796	SORD	1.84E-04	-1.85
Q9NUV7	SPTLC3	6.21E-04	-2.16
Q9Y6N5	SQOR	1.84E-04	2.51
Q13501	SQSTM1	4.64E-06	6.29
P08240	SRPRA	3.96E-04	1.58
Q9BXP5	SRRT	1.04E-03	-1.80
Q13247	SRSF6	3.66E-04	1.51
Q16629	SRSF7	2.14E-02	-2.02
Q13242	SRSF9	1.91E-02	1.52
Q13586	STIM1	4.89E-03	-2.66
Q9Y6E0	STK24	3.63E-03	1.54
Q9P289	STK26	5.02E-03	1.70
Q9UEW8	STK39	7.10E-04	1.62
P27105	STOM	1.22E-05	2.38
Q13033	STRN3	5.60E-05	2.09
Q13190	STX5	3.79E-04	1.70
Q8NBJ7	SUMF2	1.18E-05	1.93
Q9UH99	SUN2	6.35E-04	-2.16
Q7KZ85	SUPT6H	3.91E-03	-1.59
Q96A49	SYAP1	4.21E-02	1.53
O60506	SYNCRIP	7.01E-03	-4.10
Q9BSH4	TACO1	2.14E-02	-1.74
Q01995	TAGLN	8.44E-06	7.64
P26639	TARS1	9.60E-04	1.50
O14907	TAX1BP3	1.77E-03	1.79
Q9BTW9	TBCD	2.77E-03	-4.12
Q12788	TBL3	2.24E-02	-1.64
O75764	TCEA3	2.20E-06	-1.97
O14776	TCERG1	4.89E-04	-1.51
Q9UGI8	TES	8.23E-03	1.53
P02787	TF	1.05E-05	-1.53
Q12800	TFCP2	3.46E-03	-1.86
Q96FV9	THOC1	5.09E-04	-1.69
Q13769	THOC5	7.98E-03	-1.73
Q6I9Y2	THOC7	2.26E-04	-2.49
Q9P016	THYN1	7.63E-04	-1.72
P31483	TIA1	1.14E-03	-1.75
Q9BSF4	TIMM29	1.71E-02	-1.73
O60220	TIMM8A	1.24E-03	-1.52
Q9Y5J7	TIMM9	1.56E-02	-1.52
Q9NPL8	TIMMDC1	6.89E-03	1.94
Q07157	TJP1	6.81E-03	1.61
P04183	TK1	9.72E-04	1.89
Q13445	TMED1	1.24E-02	-1.55
Q9BXJ8	TMEM120A	2.35E-03	-1.79
Q9HC07	TMEM165	1.32E-04	1.99
Q9H330	TMEM245	8.17E-05	-1.88
Q9BTV4	TMEM43	5.87E-03	-1.74
Q8NBN3	TMEM87A	5.47E-04	1.57
P42167	TMPO	3.26E-05	-1.57
O60784	TOM1	2.90E-04	1.77
O96008	TOMM40	1.19E-02	-1.54

Accession	Gene Name	Anova (p)	Fold change
P11387	TOP1	1.37E-02	1.84
P06753	TPM3	1.10E-03	1.54
P12270	TPR	4.04E-04	-1.86
P13693	TPT1	6.90E-05	1.64
Q9ULW0	TPX2	9.18E-06	2.86
Q9Y296	TRAPPC4	2.60E-03	1.66
O43715	TRIAP1	1.73E-06	-2.28
P14373	TRIM27	4.84E-05	2.50
Q14669	TRIP12	5.30E-03	2.00
O95900	TRUB2	2.80E-03	1.95
P43897	TSFM	1.73E-03	-2.79
Q2NL82	TSR1	1.03E-04	1.84
Q5R3I4	TTC38	1.12E-04	-1.82
P07437	TUBB	2.03E-03	-2.02
Q13509	TUBB3	3.29E-04	1.71
Q12792	TWF1	1.96E-03	1.94
P40222	TXLNA	8.90E-05	1.76
O14530	TXNDC9	3.45E-04	2.25
Q16881	TXNRD1	7.38E-05	1.93
Q9NNW7	TXNRD2	2.74E-03	-1.54
P26368	U2AF2	6.62E-03	-1.75
P60604	UBE2G2	1.75E-04	-2.44
Q9C0C9	UBE2O	3.65E-03	1.92
Q9NPD8	UBE2T	1.73E-03	2.86
Q92575	UBXN4	2.09E-02	-1.93
O94888	UBXN7	1.50E-02	-1.91
P15374	UCHL3	8.59E-04	1.62
Q9Y5K5	UCHL5	5.55E-04	1.65
P36537	UGT2B10	1.04E-04	2.36
P16662	UGT2B7	8.06E-04	2.21
Q9BRT2	UQCC2	1.74E-03	-1.75
P47985	UQCRFS1	1.57E-02	-1.58
P51784	USP11	4.53E-03	2.65
P45974	USP5	2.13E-02	2.13
Q9BVJ6	UTP14A	5.87E-04	1.52
O75691	UTP20	8.55E-03	-1.99
P46939	UTRN	7.62E-06	1.64
Q9UK41	VPS28	4.82E-03	1.73
P04004	VTN	7.71E-04	-2.62
O76024	WFS1	2.28E-04	1.94
Q96S55	WRNIP1	2.13E-03	1.55
Q9NQH7	XPNPEP3	1.46E-04	-2.21
Q9HAV4	XPO5	6.71E-04	-1.73
Q9UIA9	XPO7	2.43E-03	-1.58
P54577	YARS1	6.51E-05	1.64
P07947	YES1	6.49E-03	-1.77
P25490	YY1	1.35E-04	-3.02
Q86VM9	ZC3H18	7.94E-04	-2.20
Q7Z2W4	ZC3HAV1	3.21E-03	1.75
Q9NUD5	ZCCHC3	3.68E-03	-1.99
O75844	ZMPSTE24	9.26E-04	1.59
Q9ULU4	ZMYND8	7.00E-03	-1.74
Q5BKZ1	ZNF326	6.18E-03	-1.51
Q15942	ZYX	8.71E-05	1.89

Appendix AE *Network Analysis of the differentially expressed proteins within the DHFR knockdown cell line revealed that the cancer, cell death and survival, organismal injury and abnormalities network was one of the most impacted. Drawn Using the IPA Software**.** Green represents those that are down-regulated, light green represents slightly down-regulated, red represents those that are up-regulated, light red represents slightly up-regulated, orange represents predicted activation, orange arrow leads to activation, blue arrow leads to inhibition, yellow arrow represents the findings inconsistent with state of downstream molecule and white represents those that were added from the Ingenuity Knowledge Base.*



Appendix AF *Network Analysis of the differentially expressed proteins within the DHFR knockdown cell line revealed that the RNA post-transcriptional modification, hereditary disorder, neurological disease network was one of the most impacted. Drawn Using the IPA Software. Green represents those that are down-regulated, light green represents slightly down-regulated, red represents those that are up-regulated, light red represents slightly up-regulated, orange arrow leads to activation, blue arrow leads to inhibition, yellow arrow represents the findings inconsistent with state of downstream molecule and white represents those that were added from the Ingenuity Knowledge Base.*



Appendix AG *Differentially expressed proteins shared by the HepG2 DHFR2 knockout and DHFR knockdown cell line when compared to HepG2 wildtype.*

		DHFR2 Knockout		DHFR Knockout	
Accession	Gene Name	Anova (p)	Fold change	Anova (p)	Fold change
Q9UDR5	AASS	1.15E-03	-1.67	2.82E-04	-1.84
P33527	ABCC1	6.41E-04	1.76	2.56E-03	1.66
P42765	ACAA2	6.85E-04	-1.56	5.62E-04	-1.66
Q86TX2	ACOT1	1.15E-03	-1.57	5.08E-04	-1.58
Q15067	ACOX1	3.40E-03	-1.65	7.81E-04	-1.90
Q99424	ACOX2	2.91E-05	-3.32	1.82E-04	-2.17
Q96CM8	ACSF2	5.84E-04	-1.79	9.69E-04	-1.76
P33121	ACSL1	1.52E-04	-1.82	2.28E-04	-1.92
O95573	ACSL3	5.54E-05	2.01	2.15E-04	1.77
Q9ULC5	ACSL5	8.35E-06	-3.07	4.66E-05	-2.09
Q68CK6	ACSM2B	8.38E-06	-9.08	7.16E-06	-5.65
Q9NR19	ACSS2	1.24E-03	1.66	1.55E-05	2.26
P08319	ADH4	4.04E-05	-4.24	3.12E-05	2.98
P28332	ADH6	1.40E-04	-1.51	3.70E-04	-1.52
O00116	AGPS	5.47E-05	-2.88	1.39E-05	3.80
O00468	AGRN	4.07E-06	5.19	5.07E-07	-8.33
P21549	AGXT	1.88E-03	-1.75	1.21E-03	-1.67
P00352	ALDH1A1	2.61E-05	-1.99	4.53E-05	-1.54
O75891	ALDH1L1	4.32E-06	-22.50	2.40E-04	-2.30
P51648	ALDH3A2	1.64E-03	-1.61	5.56E-03	-1.57
Q02252	ALDH6A1	5.41E-05	-2.43	2.15E-04	-1.92
P09972	ALDOC	1.30E-04	2.63	2.22E-03	-1.55
Q06481	APLP2	2.04E-03	1.60	6.73E-04	1.78
P02749	APOH	1.45E-04	-3.13	1.17E-02	-2.26
P78540	ARG2	1.65E-05	3.07	1.96E-06	3.27
Q92888	ARHGEF1	6.57E-03	2.01	6.12E-04	1.61
P51690	ARSL	3.04E-04	-1.83	9.11E-03	-1.97
P07307	ASGR2	9.55E-04	-1.51	1.58E-03	-1.79
P08243	ASNS	5.45E-05	4.33	6.09E-05	4.20
Q7L266	ASRGL1	1.38E-04	2.14	5.23E-06	3.76
O14965	AURKA	2.05E-02	-1.67	3.35E-05	3.03
P46379	BAG6	1.04E-02	1.58	5.23E-05	1.81
Q9P287	BCCIP	6.50E-04	1.54	8.71E-03	1.58
P21953	BCKDHB	1.65E-03	-1.67	2.66E-03	-1.67
Q86WA6	BPHL	7.27E-04	-1.74	1.90E-05	-1.82
A6NCS6	C2orf72	4.13E-03	-1.66	4.90E-03	-1.75
P35218	CA5A	1.63E-04	-2.52	5.17E-05	-2.05
O14936	CASK	3.91E-03	-1.72	1.77E-03	-2.22
P20810	CAST	1.82E-03	1.86	2.99E-03	1.58
P04040	CAT	1.07E-04	-1.69	7.19E-05	-1.84
P04233	CD74	1.05E-04	-1.75	6.35E-05	-2.22
P19022	CDH2	6.60E-06	-3.46	1.70E-06	-4.04
Q9UHN6	CEMIP2	1.32E-05	2.20	6.48E-05	1.69
Q9HA82	CERS4	2.15E-02	-1.66	1.20E-02	-1.84
Q8NET6	CHST13	6.62E-03	-1.77	3.58E-05	-2.55
P12277	CKB	5.50E-04	1.77	1.19E-05	2.69
O00299	CLIC1	3.56E-05	1.58	7.36E-05	1.65
P39060	COL18A1	7.20E-04	-1.97	5.54E-04	-2.57
Q8NI60	COQ8A	3.76E-04	-1.80	7.31E-05	-1.91
Q9BR76	CORO1B	5.86E-04	1.59	1.33E-03	1.58
P35221	CTNNA1	6.51E-06	-1.91	7.71E-06	-1.68
O60716	CTNND1	4.88E-05	-1.51	2.22E-05	-1.76
P00167	CYB5A	9.91E-04	-1.63	1.06E-03	-1.55
Q16643	DBN1	1.05E-06	2.38	1.32E-06	1.80
P11182	DBT	4.35E-03	-1.83	9.15E-03	-2.39
Q7Z4W1	DCXR	1.08E-04	-1.61	1.13E-05	-1.69
Q92466	DDB2	4.49E-04	1.85	3.30E-05	-2.36
O75907	DGAT1	1.78E-04	-1.79	3.74E-05	-1.82
P00374	DHFR	1.40E-04	-2.40	1.85E-02	-1.63
Q96LJ7	DHRS1	8.72E-03	-1.74	1.66E-03	-2.05
Q7L2E3	DHX30	5.83E-04	-2.08	1.00E-02	-2.18
O60610	DIAPH1	2.24E-03	-1.51	1.90E-02	-1.77
Q96N67	DOCK7	5.20E-03	1.59	1.55E-02	1.88
Q14126	DSG2	6.81E-06	1.73	8.29E-03	-1.52
P30084	ECHS1	6.35E-04	-1.63	3.32E-04	-1.70
O75521	ECI2	6.82E-03	-1.82	1.73E-03	-2.61
Q8NDI1	EHBP1	1.20E-03	2.01	1.09E-03	2.49
Q08426	EHHADH	9.04E-04	-1.55	3.09E-03	-1.54
O95163	ELP1	9.83E-04	-1.63	5.97E-04	-1.56

		DHFR2 Knockout		DHFR Knockout	
Accession	Gene Name	Anova (p)	Fold change	Anova (p)	Fold change
Q9HC35	EML4	3.76E-05	1.82	1.63E-04	1.69
P22413	ENPP1	6.29E-05	2.36	4.24E-05	2.85
O43491	EPB41L2	1.01E-05	1.59	1.90E-06	1.52
P29317	EPHA2	2.38E-05	1.77	5.72E-05	2.30
P07099	EPHX1	2.13E-03	-1.63	5.39E-03	-1.56
P00734	F2	3.93E-04	1.76	9.23E-06	2.32
Q9NZB2	FAM120A	3.21E-04	-1.66	1.39E-02	-1.68
Q9Y6X4	FAM169A	1.44E-04	1.67	1.46E-03	1.57
Q14320	FAM50A	1.46E-04	1.58	1.05E-04	1.85
P55899	FCGRT	1.98E-05	-2.68	5.11E-04	-2.00
P49326	FMO5	6.32E-06	-1.95	7.79E-08	-9.61
O95954	FTCD	1.47E-05	-1.89	7.38E-05	-1.74
Q8TAE8	GADD45GIP1	4.84E-03	-1.78	4.98E-03	-1.64
P41250	GARS1	1.59E-05	1.53	1.59E-05	1.61
Q04446	GBE1	3.34E-03	1.96	9.94E-03	1.55
Q92947	GCDH	2.86E-04	-2.09	1.63E-04	-2.18
P32189	GK	2.97E-05	1.85	1.17E-06	1.93
P35754	GLRX	1.08E-04	1.85	5.30E-05	1.92
P00367	GLUD1	2.02E-04	-1.99	1.41E-04	-1.99
O60547	GMDS	2.01E-05	1.90	1.29E-06	2.46
Q9Y223	GNE	2.58E-02	1.56	6.26E-04	2.33
P15586	GNS	2.15E-03	1.60	7.33E-07	2.07
P17174	GOT1	4.51E-04	1.63	2.02E-05	2.11
P51654	GPC3	9.56E-04	1.76	1.73E-07	-234.55
Q9UBQ7	GRHPR	2.07E-04	-1.53	2.92E-04	-1.63
P28799	GRN	1.03E-04	2.33	3.00E-03	-2.23
P08263	GSTA1	4.08E-05	-2.03	1.76E-03	1.58
Q9Y2Q3	GSTK1	4.70E-04	-1.88	1.16E-03	-1.65
O60814	H2BC12	8.90E-05	-2.11	7.56E-05	-2.23
P62805	H4C1	6.10E-03	-1.83	2.00E-03	-1.99
Q16836	HADH	3.88E-03	-1.52	6.66E-04	-1.76
P69905	HBA1	1.82E-05	7.91	2.57E-02	1.51
Q9ULT8	HECTD1	1.82E-04	-2.29	3.16E-06	2.25
Q9NRZ9	HELLS	1.14E-06	-8.15	3.25E-03	1.70
P52789	HK2	3.11E-06	1.70	1.87E-07	1.67
Q01581	HMGCS1	3.41E-04	-1.50	1.25E-05	1.80
Q99714	HSD17B10	3.19E-03	-1.51	9.50E-03	-1.54
P05362	ICAM1	6.18E-05	2.01	2.18E-04	1.77
Q8N6C5	IGSF1	1.15E-03	-1.86	1.45E-02	-1.66
A1L0T0	ILVBL	7.82E-04	-1.78	1.07E-03	-1.63
Q9NV31	IMP3	8.17E-04	-2.07	2.53E-04	-5.11
Q9NPH2	ISYNA1	2.82E-05	1.86	6.83E-05	-2.23
P17301	ITGA2	2.17E-05	2.96	5.82E-04	1.53
Q14571	ITPR2	3.94E-03	-1.61	3.80E-04	-1.80
Q6NYC1	JMJD6	5.51E-03	1.99	1.64E-04	2.49
Q07866	KLC1	6.38E-03	1.70	1.19E-02	1.55
P05783	KRT18	8.15E-05	1.84	3.04E-05	1.83
P05787	KRT8	8.52E-06	1.88	2.13E-05	1.80
O15230	LAMA5	8.09E-05	2.06	2.39E-06	-4.70
P07942	LAMB1	4.32E-05	2.71	8.96E-04	-1.68
P55268	LAMB2	1.68E-06	-8.52	9.77E-07	-8.15
P11047	LAMC1	1.59E-03	1.52	2.48E-05	-3.01
P09382	LGALS1	1.45E-08	5.49	6.77E-09	3.69
P17931	LGALS3	1.70E-05	3.52	3.51E-11	21.38
Q6ZN17	LIN28B	1.69E-04	-1.95	8.51E-05	-1.91
Q32MZ4	LRRFIP1	1.28E-03	1.81	2.87E-03	1.64
P61626	LYZ	1.22E-05	2.21	2.06E-07	2.42
P21397	MAOA	8.49E-04	-2.13	3.07E-05	-3.23
P27338	MAOB	4.05E-05	-2.56	3.95E-05	-2.66
P27816	MAP4	5.38E-04	1.60	5.04E-04	1.51
Q96RQ3	MCCC1	1.97E-03	-1.75	2.06E-03	-1.78
Q9HCC0	MCCC2	5.04E-03	-1.67	7.25E-03	-1.57
P49736	MCM2	2.45E-03	-1.93	1.21E-04	1.54
O95140	MFN2	6.00E-03	-1.72	4.48E-03	2.08
O95822	MLYCD	4.46E-03	-1.82	1.92E-03	-2.20
Q96EY8	MMAB	7.74E-03	-1.62	2.69E-03	-1.68
P22033	MMUT	5.69E-05	-2.16	1.77E-03	-1.58
Q9HCE1	MOV10	5.12E-05	-1.50	2.20E-05	-1.59
Q5T653	MRPL2	1.56E-02	2.06	2.56E-02	1.97

Accession	Gene Name	DHFR2 Knockout		DHFR Knockout	
		Anova (p)	Fold change	Anova (p)	Fold change
Q9BZE1	MRPL37	6.99E-03	-1.58	8.43E-03	-1.63
Q92552	MRPS27	4.43E-03	-1.75	2.32E-03	-2.11
Q5VT66	MTARC1	7.36E-04	-1.76	1.81E-03	-1.53
Q9Y6C9	MTCH2	2.09E-03	-1.66	1.38E-03	-1.59
P55157	MTTP	2.36E-05	-1.72	7.55E-06	-1.80
P53602	MVD	1.67E-04	1.51	4.59E-06	2.08
O00159	MYO1C	2.51E-03	1.81	2.08E-03	1.81
P43490	NAMPT	8.01E-04	1.91	2.63E-03	1.56
Q9NR45	NANS	1.78E-04	-1.87	3.93E-04	-1.61
Q9UI09	NDUFA12	1.73E-02	-1.58	4.41E-03	-1.77
P14543	NID1	3.42E-05	2.08	2.31E-06	-5.99
Q9GZT8	NIF3L1	1.43E-02	-1.54	2.86E-03	-2.14
O75323	NIPSNAP2	2.64E-03	-1.51	1.00E-04	-2.01
Q9H857	NT5DC2	2.79E-02	-1.55	1.35E-02	-2.63
Q9BRJ7	NUDT16L1	9.37E-03	-1.61	7.32E-03	-1.71
Q14980	NUMA1	4.93E-03	1.51	1.21E-02	-2.05
Q8WUM0	NUP133	6.32E-03	-1.53	2.11E-03	-1.51
Q12769	NUP160	5.42E-04	-1.82	1.58E-03	-1.94
Q8TEM1	NUP210	3.64E-03	-1.84	5.41E-04	-2.24
O60313	OPA1	2.87E-03	-1.63	5.67E-03	-1.56
O15460	P4HA2	5.30E-06	2.13	1.54E-04	2.29
P00439	PAH	6.65E-07	-2.77	4.14E-06	-1.97
P05165	PCCA	4.29E-03	-1.64	1.67E-03	-1.67
Q9UHG3	PCYOX1	2.53E-03	-1.55	6.83E-04	-1.68
Q5T2W1	PDZK1	1.00E-08	-5.01	4.02E-06	-2.19
Q8TBX8	PIP4K2C	3.87E-04	1.78	5.37E-05	1.81
Q9P0Z9	PIPOX	5.98E-05	-2.05	2.50E-04	-1.80
Q13393	PLD1	1.18E-06	2.57	1.36E-04	1.64
P53350	PLK1	1.38E-04	-1.77	2.45E-05	2.19
Q14651	PLS1	2.62E-04	1.71	8.05E-04	1.59
P24928	POLR2A	2.81E-03	-1.55	8.25E-05	-1.58
O75170	PPP6R2	2.15E-05	1.85	6.98E-04	2.14
P17252	PRKCA	6.07E-05	-1.73	1.25E-03	-1.53
Q9UF12	PRODH2	2.43E-03	-1.66	1.63E-03	-1.66
Q13523	PRPF4B	2.40E-03	-1.52	5.17E-03	1.96
Q9H7Z7	PTGES2	1.73E-03	1.60	8.81E-04	1.51
P15151	PVR	1.52E-04	1.59	4.94E-04	1.91
Q8IY67	RAVER1	1.07E-02	-1.53	5.07E-03	-1.98
P78509	RELN	4.66E-04	-1.80	4.88E-06	-3.50
Q15493	RGN	4.81E-03	-1.87	2.15E-03	-2.21
Q9NWS8	RMND1	3.95E-03	-1.80	3.34E-03	-1.60
O75116	ROCK2	1.14E-03	1.66	7.51E-04	1.73
P27635	RPL10	6.05E-03	1.60	1.21E-03	2.23
P62906	RPL10A	6.08E-03	1.60	5.95E-03	1.62
P50914	RPL14	4.49E-02	1.86	1.49E-03	3.31
P18621	RPL17	2.69E-03	1.95	1.06E-03	2.53
Q02543	RPL18A	1.71E-02	2.31	2.74E-03	3.87
P46778	RPL21	2.14E-02	1.63	2.48E-03	2.12
P61254	RPL26	3.57E-03	2.25	1.03E-03	2.59
P46776	RPL27A	8.03E-03	1.70	4.16E-04	2.63
P39023	RPL3	2.75E-02	1.81	9.34E-03	2.56
P62280	RPS11	2.16E-03	1.75	1.13E-03	1.88
P62277	RPS13	6.78E-03	1.52	2.86E-04	2.06
P62269	RPS18	9.84E-04	1.55	7.84E-05	1.83
P62266	RPS23	1.63E-02	1.53	5.11E-03	1.72
P62847	RPS24	1.87E-03	1.63	7.87E-04	1.64
P46781	RPS9	1.06E-02	1.82	8.75E-04	2.81
P31350	RRM2	4.32E-04	-2.46	2.21E-05	3.56
Q14684	RRP1B	3.26E-03	-1.61	4.60E-04	1.60
Q99584	S100A13	3.03E-06	1.90	1.39E-04	2.29
P25815	S100P	7.82E-05	2.20	1.67E-05	2.95
P22307	SCP2	6.57E-03	-1.57	3.21E-03	-1.53
Q9HB40	SCPEP1	2.41E-05	1.90	8.00E-05	2.25
Q13228	SELENBP1	1.05E-02	-1.55	5.99E-04	-1.70
P01011	SERPINA3	1.82E-06	5.48	2.54E-04	2.08
P31947	SFN	2.38E-06	4.02	1.29E-08	6.47
P34896	SHMT1	5.45E-03	-1.53	5.11E-03	-1.61
Q9NXA8	SIRT5	1.19E-03	-1.50	1.10E-04	-1.60
O15427	SLC16A3	3.94E-03	1.65	2.69E-03	1.66

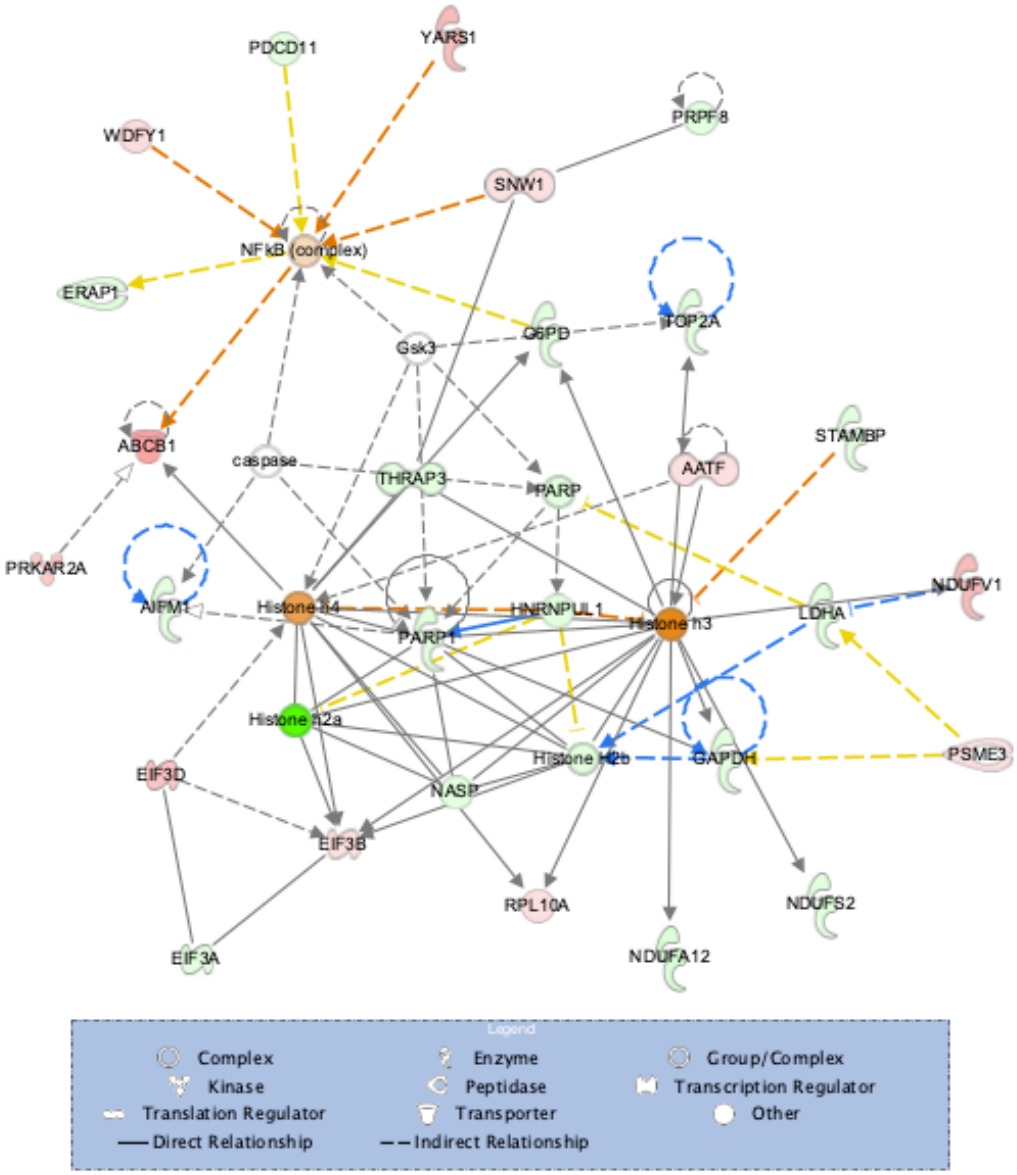
Accession	Gene Name	DHFR2 Knockout		DHFR Knockout	
		Anova (p)	Fold change	Anova (p)	Fold change
P11169	SLC2A3	1.21E-04	1.87	3.47E-05	-2.59
Q8WUX1	SLC38A5	3.69E-04	-1.71	1.97E-03	-1.55
O94875	SORBS2	4.45E-05	-1.88	2.19E-03	-1.68
Q00796	SORD	1.44E-04	-1.74	1.84E-04	-1.85
Q9NUV7	SPTLC3	1.22E-06	-3.74	6.21E-04	-2.16
Q9Y6N5	SQOR	6.79E-05	2.41	1.84E-04	2.51
Q13586	STIM1	3.99E-03	-1.71	4.89E-03	-2.66
Q9P289	STK26	2.24E-03	1.69	5.02E-03	1.70
Q9UEW8	STK39	1.03E-04	1.84	7.10E-04	1.62
P27105	STOM	3.20E-05	2.36	1.22E-05	2.38
Q13033	STRN3	2.03E-04	1.56	5.60E-05	2.09
Q9UH99	SUN2	4.50E-03	-1.58	6.35E-04	-2.16
Q7KZ85	SUPT6H	1.60E-04	-1.60	3.91E-03	-1.59
Q01995	TAGLN	2.30E-05	6.35	8.44E-06	7.64
P26639	TARS1	6.97E-04	1.60	9.60E-04	1.50
O75764	TCEA3	8.73E-05	-1.62	2.20E-06	-1.97
Q9UGI8	TES	6.28E-03	2.04	8.23E-03	1.53
Q12800	TFCP2	5.72E-04	-1.84	3.46E-03	-1.86
O60220	TIMM8A	1.39E-03	-1.54	1.24E-03	-1.52
Q07157	TJP1	9.64E-03	1.52	6.81E-03	1.61
Q9BXJ8	TMEM120A	1.76E-03	-1.94	2.35E-03	-1.79
Q9BTV4	TMEM43	2.22E-03	-1.50	5.87E-03	-1.74
P11387	TOP1	2.71E-02	1.59	1.37E-02	1.84
P14373	TRIM27	5.33E-04	1.79	4.84E-05	2.50
Q2NL82	TSR1	6.99E-03	1.52	1.03E-04	1.84
Q13509	TUBB3	1.14E-06	2.44	3.29E-04	1.71
Q12792	TWF1	7.96E-03	1.51	1.96E-03	1.94
Q9NNW7	TXNRD2	5.98E-03	-1.56	2.74E-03	-1.54
P26368	U2AF2	2.05E-03	-1.76	6.62E-03	-1.75
P60604	UBE2G2	7.44E-03	-2.39	1.75E-04	-2.44
Q9UK41	VPS28	7.33E-03	2.04	4.82E-03	1.73
O76024	WFS1	1.76E-05	2.14	2.28E-04	1.94
P54577	YARS1	1.75E-04	1.59	6.51E-05	1.64
Q9ULU4	ZMYND8	1.35E-03	1.70	7.00E-03	-1.74
Q5BKZ1	ZNF326	1.58E-03	-1.60	6.18E-03	-1.51
Q15942	ZYX	1.97E-04	1.74	8.71E-05	1.89

Appendix AH *Differentially expressed proteins identified in the HepG2 DHFR2 overexpression model when compared to HepG2 pCMV6-AC-tGFP (empty) vector model.*

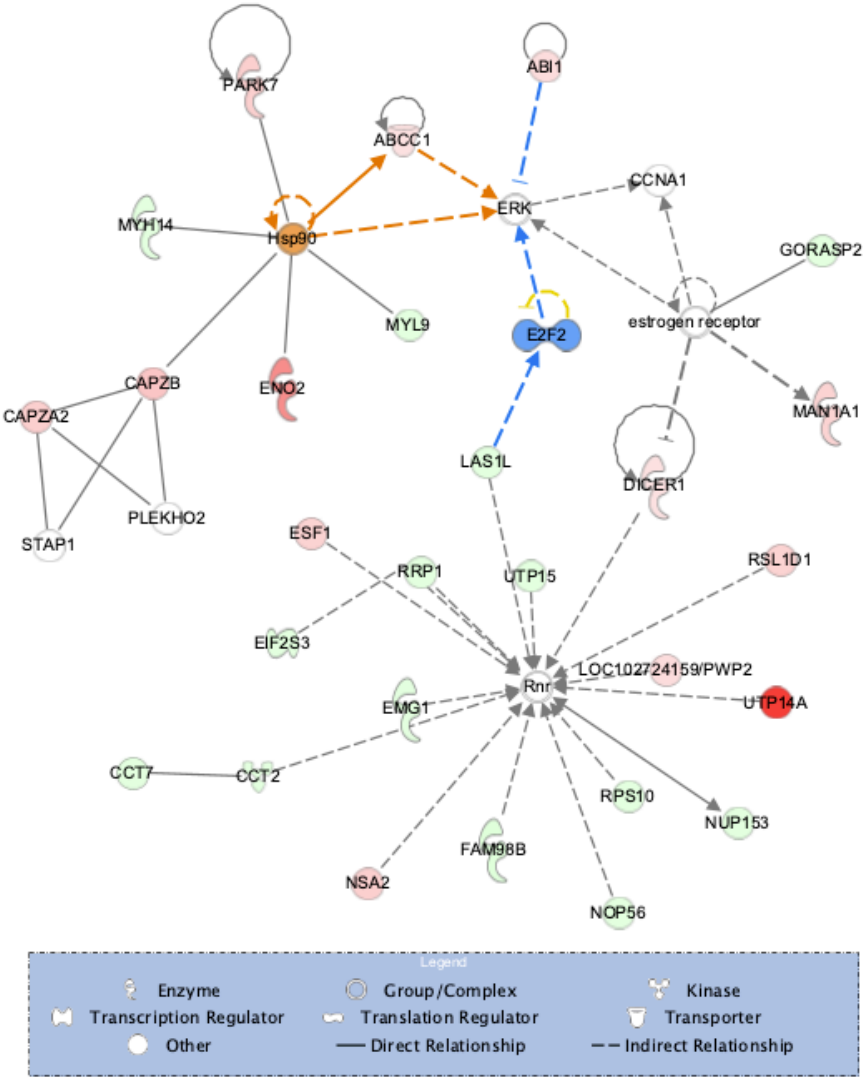
Accession	Gene Name	Anova (p)	Fold change
Q9NY61	AATF	9.31E-03	1.93
P08183	ABCB1	9.61E-03	5.98
P33527	ABCC1	1.91E-02	1.67
Q8N2K0	ABHD12	9.61E-03	3.46
Q96IU4	ABHD14B	1.24E-02	2.66
Q8IZP0	ABI1	2.57E-02	2.26
Q13085	ACACA	9.34E-03	1.66
Q9H845	ACAD9	1.87E-02	1.86
P49748	ACADVL	2.58E-04	1.79
P24752	ACAT1	1.46E-02	-1.50
Q9BWD1	ACAT2	2.47E-02	3.78
Q9UKV3	ACIN1	4.60E-02	-10.46
Q15067	ACOX1	9.99E-03	1.67
P24666	ACP1	1.35E-02	-2.55
O60488	ACSL4	1.28E-02	-1.88
P55196	AFDN	2.49E-02	1.51
P23526	AHCY	1.05E-02	-1.70
Q9BQI0	AIF1L	4.14E-02	-1.78
O95831	AIFM1	4.83E-02	-3.73
Q12904	AIMP1	1.66E-02	-26.71
P54886	ALDH18A1	1.74E-02	1.65
P30837	ALDH1B1	3.94E-02	1.53
P49419	ALDH7A1	2.93E-02	-2.40
O43747	AP1G1	1.56E-02	3.08
P02649	APOE	1.18E-02	2.65
P61204	ARF3	2.16E-02	-12.33
P18085	ARF4	4.08E-02	1.98
Q8N6T3	ARFGAP1	3.58E-02	7.36
Q99856	ARID3A	2.07E-02	3.11
P07306	ASGR1	6.83E-03	-1.85
P08243	ASNS	1.35E-02	-36.44
Q12797	ASPH	1.91E-03	2.09
Q9NVI7	ATAD3A	2.76E-02	-1.64
P25705	ATP5F1A	4.70E-02	-1.64
P06576	ATP5F1B	1.37E-02	3.19
P36542	ATP5F1C	3.18E-02	-7.67
P38606	ATP6V1A	6.90E-03	-31.93
Q8WWM7	ATXN2L	2.57E-02	1.86
Q9NX62	BPNT2	4.27E-02	-3.87
Q13895	BYSL	5.47E-03	-1.99
P0C0L4	C4A	1.63E-02	1.57
Q86VP6	CAND1	2.98E-02	-2.02
P27824	CANX	4.56E-02	2.02
Q01518	CAP1	2.45E-02	-1.55
P47755	CAPZA2	4.35E-02	3.26
P47756	CAPZB	2.40E-02	3.83
Q9Y2V2	CARHSP1	2.49E-02	1.70
P49589	CARS1	1.98E-02	2.98
P78371	CCT2	3.78E-02	-4.98
Q99832	CCT7	1.34E-02	-1.53
Q99459	CDC5L	1.08E-02	-1.66
Q9NYV4	CDK12	1.40E-02	1.73
Q03701	CEBPZ	3.12E-02	-9.38
Q9UHN6	CEMIP2	2.32E-02	-15.84
Q14839	CHD4	2.18E-02	3.05
P12277	CKB	4.83E-03	5.83
Q16740	CLPP	2.01E-02	-1.53
Q96KP4	CNDP2	1.37E-02	-1.70
Q9Y2B0	CNPY2	4.68E-02	-2.63
Q8NBJ5	COLGALT1	9.01E-03	-1.73
P53621	COPA	2.11E-03	1.66
P53618	COPB1	1.37E-03	3.89
Q9BR76	CORO1B	4.43E-03	-12.98
P09669	COX6C	2.96E-02	2.44
P31327	CPS1	1.45E-02	2.37
P46108	CRK	2.76E-03	5.77
P46109	CRKL	7.33E-03	-5.93
Q08257	CRYZ	1.50E-02	-12.20
P51398	DAP3	2.65E-04	2.30
Q92841	DDX17	3.16E-02	2.32
O00571	DDX3X	1.95E-02	-1.57
O75907	DGAT1	2.26E-02	-3.38
P00374	DHFR	2.08E-05	4.42
Q86XF0	DHFR2	1.53E-07	15.05
Q92620	DHX38	4.98E-03	-2.42
Q9UPY3	DICER1	4.83E-03	1.63
P10515	DLAT	7.32E-03	1.61
P31689	DNAJA1	2.58E-02	2.87
P23919	DTYMK	1.05E-02	-1.82
Q15075	EEA1	2.45E-03	-1.94
P13639	EEF2	6.40E-03	-1.59
Q9H4M9	EHD1	1.15E-02	-1.80
Q08426	EHHADH	2.26E-02	1.89
P41091	EIF2S3	3.97E-02	-5.10
Q14152	EIF3A	2.78E-02	-1.57
P55884	EIF3B	1.89E-02	1.56
B5ME19	EIF3CL	1.33E-03	-2.14
O15371	EIF3D	3.64E-02	3.74
Q13347	EIF3I	3.33E-02	2.84
P60842	EIF4A1	2.13E-02	1.80
Q15370	ELOB	2.26E-02	-78.31
Q92979	EMG1	3.04E-02	-41.34
P09104	ENO2	6.71E-05	7.35
P58107	EPPK1	9.59E-03	-1.52
Q9NZ08	ERAP1	1.35E-02	-68.45
O94905	ERLIN2	4.34E-03	-8.49
P10768	ESD	7.09E-04	1.58
Q9H501	ESF1	3.34E-02	2.96
Q52LJ0	FAM98B	2.89E-02	-157.76
Q9Y4F1	FARP1	1.32E-02	1.91
P22570	FDXR	1.98E-02	2.19
P02679	FGG	2.05E-02	7.48
P02751	FN1	4.33E-02	-2.52
Q9Y2H6	FNDC3A	4.23E-03	2.70
P11413	G6PD	2.43E-02	-1.57
Q10471	GALNT2	2.67E-02	-11.06
P04406	GAPDH	4.18E-02	-1.55
P22102	GART	2.24E-02	-1.50
Q96RP9	GFM1	8.94E-03	-9.86
Q06210	GFPT1	1.56E-02	-1.59
P23378	GLDC	1.65E-02	1.71
O94925	GLS	1.24E-02	2.31
P49915	GMPS	4.25E-02	-9.24
P62873	GNB1	2.30E-02	-301.12
Q14789	GOLGB1	3.78E-02	-26.32
O00461	GOLIM4	2.70E-02	-1.57
Q9H8Y8	GORASP2	2.14E-03	-2.14
P00505	GOT2	2.13E-02	-2.57
Q9HAV7	GRPEL1	3.80E-02	-1.94
P15170	GSPT1	1.29E-02	3.71
P48637	GSS	3.98E-02	2.46
P0C0S5	H2AZ1	2.97E-02	-2846.12
O60814	H2BC12	7.43E-03	-3.92
P62805	H4C1	2.08E-03	-1.88
P52789	HK2	4.11E-02	2.76
P09429	HMGB1	7.29E-03	-1.68
P41235	HNF4A	1.89E-02	-2.67
Q9BUJ2	HNRNPUL1	1.38E-02	-5.47
P37235	HPCAL1	2.01E-02	-24.07
Q99714	HSD17B10	2.91E-02	-969.52
P48723	HSPA13	1.02E-02	-2.67
Q9NSE4	IARS2	3.31E-02	-7.99
Q12906	ILF3	4.02E-02	-1.66
P46940	IQGAP1	1.89E-02	-1.72
Q96CN7	ISOC1	7.37E-03	4.77
P14923	JUP	2.07E-02	-3.20
Q8IZA0	KIAA0319L	2.26E-02	2.63
P33176	KIF5B	4.40E-02	-3.21
Q14974	KPNB1	1.32E-02	-1.88
P04264	KRT1	1.31E-02	3.01
P13645	KRT10	3.06E-02	4.81
P35908	KRT2	2.50E-02	3.41
Q86UP2	KTN1	3.84E-03	1.55
O43813	LANCL1	3.43E-02	-2.12
Q6PKG0	LARP1	2.08E-02	2.79
Q71RC2	LARP4	1.42E-02	2.14
Q9Y4W2	LAS1L	1.67E-02	-1.67
Q14847	LASP1	3.91E-02	-1.83
P00338	LDHA	1.99E-02	-2.52
Q08380	LGALS3BP	4.40E-02	6.54
P18858	LIG1	3.52E-02	2.43
Q03252	LMNB2	4.60E-03	-1.64
Q8N1G4	LRRC47	2.45E-02	-2.25
P33908	MAN1A1	1.60E-02	2.70
Q16706	MAN2A1	5.38E-04	1.61
Q15691	MAPRE1	2.29E-02	-1.67
P43243	MATR3	3.58E-02	3.50
Q10469	MGAT2	1.54E-02	-27.60
Q9BYD1	MRPL13	3.30E-02	3.31
Q9P015	MRPL15	2.99E-02	-2.44
Q9Y2Q9	MRPS28	1.25E-02	1.80
P52701	MSH6	2.95E-02	-3.08
Q96DH6	MSI2	4.08E-02	-19.74
O75431	MTX2	7.93E-03	-48.84
Q7Z406	MYH14	6.69E-03	-6.10
P24844	MYL9	2.24E-02	-2.89
P49321	NASP	3.91E-02	-2.36
Q9UI09	NDUFA12	3.79E-03	-193.11
O14561	NDUFAB1	1.97E-02	-10.27
O75306	NDUFS2	1.85E-02	-2.29
P49821	NDUFV1	1.37E-02	4.26
Q9BYT8	NLN	3.50E-02	2.98
Q13423	NNT	1.02E-02	3.63
Q5JPE7	NOMO2	3.90E-02	-1.75
P46087	NOP2	2.03E-02	3.51
O00567	NOP56	1.17E-02	-2.36
Q86U38	NOP9	6.05E-03	12.23
O95478	NSA2	2.23E-03	3.06
O95989	NUDT3	1.48E-02	2.85
Q8WV74	NUDT8	1.97E-03	1.76
P49790	NUP153	1.47E-02	-1.73
Q9NWU1	OXSM	2.66E-02	-47.89
P07237	P4HB	2.67E-02	2.64
Q99497	PARK7	2.15E-02	3.15
P09874	PARP1	3.84E-02	-2.86
Q14690	PDCD11	2.06E-02	-33.29
Q9H2J4	PDCL3	5.13E-03	-1.77
Q6L8Q7	PDE12	4.01E-02	-4.30
P52209	PGD	4.42E-02	-2.62
Q99959	PKP2	1.92E-02	-14.62
O60664	PLIN3	3.67E-02	-2.55
O43660	PLRG1	6.63E-03	-201.62
P24928	POLR2A	1.14E-02	2.40
P62937	PPIA	1.84E-02	3.96
P23284	PPIB	2.13E-02	-12.33
P30048	PRDX3	4.20E-02	-10.06
P13861	PRKAR2A	2.81E-02	3.38
Q6P2Q9	PRPF8	8.18E-03	-1.63
P35998	PSMC2	1.83E-02	-2.40
P61289	PSME3	7.96E-03	1.64
Q96EY7	PTCD3	3.04E-02	-3.65
Q15269	PWP2	1.51E-02	2.25
P46060	RANGAP1	4.49E-02	-2.20
P54136	RARS1	2.61E-02	2.97
Q15293	RCN1	1.52E-02	-1.93
Q8IXI1	RHOT2	3.86E-02	-2.34
Q5UIP0	RIF1	1.69E-02	-1.56
Q9H4A4	RNPEP	5.25E-03	-7.56
P62906	RPL10A	1.08E-02	2.00
P50914	RPL14	2.76E-03	1.60

Accession	Gene Name	Anova (p)	Fold change
Q07020	RPL18	3.91E-02	1.57
P62829	RPL23	4.50E-02	-5.13
Q02878	RPL6	5.68E-03	2.23
P62917	RPL8	1.30E-02	1.68
P04843	RPN1	8.25E-03	-10.27
Q96P16	RPRD1A	3.46E-02	1.86
P46783	RPS10	6.38E-04	-2.03
P25398	RPS12	3.49E-02	-2.15
P46782	RPS5	1.90E-02	2.52
P56182	RRP1	1.20E-02	-1.51
O76021	RSL1D1	8.96E-04	2.96
Q9NQC3	RTN4	4.74E-03	6.04
Q9UL12	SARDH	1.33E-02	2.48
Q9UPN6	SCAF8	7.36E-03	2.51
Q15436	SEC23A	3.67E-03	2.59
Q9UGP8	SEC63	3.04E-02	1.53
Q15637	SF1	7.19E-03	-2.77
O95470	SGPL1	3.65E-02	2.64
P53007	SLC25A1	2.46E-02	-2.46
O75746	SLC25A12	3.39E-02	2.61
Q15043	SLC39A14	3.84E-03	-6.98
P62314	SNRPD1	3.17E-02	2.16
Q13573	SNW1	8.38E-03	1.67
P04179	SOD2	3.03E-02	2.90
P18583	SON	5.79E-03	-1.52
P61011	SRP54	2.35E-04	1.54
P05455	SSB	3.28E-02	2.65
O95630	STAMBP	1.09E-02	-28.37
Q9UJZ1	STOML2	2.39E-02	3.67
Q9Y3F4	STRAP	1.23E-02	-2.09
Q9UH99	SUN2	3.53E-02	-1.74
Q7KZ85	SUPT6H	2.64E-02	4.17
O75410	TACC1	1.65E-02	4.34
Q9BSH4	TACO1	1.92E-02	2.19
Q13148	TARDBP	1.65E-02	2.94
Q9Y4P3	TBL2	2.57E-02	1.81
Q9Y2W1	THRAP3	8.21E-03	-5.15
O60830	TIMM17B	6.00E-03	-3.86
Q96SK2	TMEM209	9.18E-03	-1.62
Q9H330	TMEM245	7.25E-03	-1.50
Q6ZXV5	TMTC3	1.59E-02	-3.47
Q9NS69	TOMM22	1.44E-02	4.18
P11388	TOP2A	3.99E-02	-1.85
Q8WWH5	TRUB1	4.51E-02	-1.89
P07437	TUBB	3.17E-02	1.56
O95881	TXNDC12	2.58E-02	-10.20
O15042	U2SURP	2.81E-02	3.21
O94874	UFL1	2.78E-02	2.31
Q9NYU2	UGGT1	6.84E-03	2.96
Q16851	UGP2	2.06E-02	-1.73
Q53GS9	USP39	1.58E-02	-1.55
Q9BVJ6	UTP14A	3.53E-02	11.98
Q8TED0	UTP15	3.12E-02	-1.52
P46939	UTRN	2.66E-02	3.37
P26640	VARS1	1.11E-02	-34.19
P45880	VDAC2	1.87E-02	-23.87
Q96QK1	VPS35	1.96E-02	1.83
A3KMH1	VWA8	1.65E-02	3.42
Q8IWB7	WDFY1	3.06E-02	2.29
Q15061	WDR43	2.35E-02	3.07
P54577	YARS1	2.40E-02	4.23
P16989	YBX3	3.25E-02	-7.89
Q96KR1	ZFR	2.62E-02	-2.36

Appendix AI *Network Analysis of the differentially expressed proteins within the DHFR2 overexpression cell line revealed that the hereditary disorder, metabolic disease, neurological disease network was one of the most impacted. Drawn Using the IPA Software**.** Green represents those that are down-regulated, light green represents slightly down-regulated, red represents those that are up-regulated, light red represents slightly up-regulated, orange represents predicted activation, orange arrow leads to activation, blue arrow leads to inhibition, yellow arrow represents the findings inconsistent with state of downstream molecule and white represents those that were added from the Ingenuity Knowledge Base.*



Appendix AJ *Network Analysis of the differentially expressed proteins within the DHFR2 overexpression cell line revealed that the cancer, endocrine system disorders, organismal injury and abnormalities network was one of the most impacted. Drawn Using the IPA Software. Light green represents slightly down-regulated, red represents those that are up-regulated, light red represents slightly up-regulated, orange represents predicted activation, blue represents predicted inhibition, orange arrow leads to activation, blue arrow leads to inhibition, yellow arrow represents the findings inconsistent with state of downstream molecule and white represents those that were added from the Ingenuity Knowledge Base.*



Appendix AK *Differentially expressed proteins that were found in both the HepG2 DHFR2 knockout and DHFR2 overexpression cell lines.*

Accession	Gene	Anova (p)	Fold change	Anova (p)	Fold change
P33527	ABCC1	6.41E-04	1.76	1.91E-02	1.67
Q13085	ACACA	1.35E-03	1.57	9.34E-03	1.66
P49748	ACADVL	4.96E-04	-1.57	2.58E-04	1.79
Q15067	ACOX1	3.40E-03	-1.65	9.99E-03	1.67
P55196	AFDN	2.17E-03	1.58	2.49E-02	1.51
P02649	APOE	9.37E-04	1.57	1.18E-02	2.65
Q99856	ARID3A	2.34E-03	1.50	2.07E-02	3.11
P08243	ASNS	5.45E-05	4.33	1.35E-02	-36.44
Q12797	ASPH	9.21E-03	1.83	1.91E-03	2.09
Q8WWM7	ATXN2L	9.10E-04	1.56	2.57E-02	1.86
P49589	CARS1	6.71E-03	1.50	1.98E-02	2.98
Q9UHN6	CEMIP2	1.32E-05	2.20	2.32E-02	-15.84
P12277	CKB	5.50E-04	1.77	4.83E-03	5.83
Q9BR76	CORO1B	5.86E-04	1.59	4.43E-03	-12.98
P31327	CPS1	2.66E-03	-1.64	1.45E-02	2.37
O75907	DGAT1	1.78E-04	-1.79	2.26E-02	-3.38
P00374	DHFR	1.40E-04	-2.40	2.08E-05	4.42
Q08426	EHHADH	9.04E-04	-1.55	2.26E-02	1.89
P22102	GART	6.61E-05	1.53	2.24E-02	-1.50
O60814	H2BC12	8.90E-05	-2.11	7.43E-03	-3.92
P62805	H4C1	6.10E-03	-1.83	2.08E-03	-1.88
P52789	HK2	3.11E-06	1.70	4.11E-02	2.76
P41235	HNF4A	8.03E-05	-1.89	1.89E-02	-2.67
Q99714	HSD17B10	3.19E-03	-1.51	2.91E-02	-969.52
Q10469	MGAT2	8.54E-04	-1.76	1.54E-02	-27.60
Q9UI09	NDUFA12	1.73E-02	-1.58	3.79E-03	-193.11
P24928	POLR2A	2.81E-03	-1.55	1.14E-02	2.40
Q15293	RCN1	5.25E-05	1.64	1.52E-02	-1.93
Q5UIP0	RIF1	4.65E-03	-1.60	1.69E-02	-1.56
P62906	RPL10A	6.08E-03	1.60	1.08E-02	2.00
P50914	RPL14	4.49E-02	1.86	2.76E-03	1.60
P46783	RPS10	5.62E-04	-1.51	6.38E-04	-2.03
O76021	RSL1D1	2.76E-02	1.62	8.96E-04	2.96
Q9UH99	SUN2	4.50E-03	-1.58	3.53E-02	-1.74
Q7KZ85	SUPT6H	1.60E-04	-1.60	2.64E-02	4.17
P11388	TOP2A	3.85E-04	-2.07	3.99E-02	-1.85
A3KMH1	VWA8	1.18E-03	-1.63	1.65E-02	3.42
P54577	YARS1	1.75E-04	1.59	2.40E-02	4.23