Deciphering the function of the human *Dihydrofolate Reductase 2* gene

Paola Drago, MSc

Supervisor: Prof Anne Parle McDermott

School of Biotechnology
Dublin City University

A thesis submitted for the degree
of Doctor of Philosophy

August 2022

Declaration

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:  ID No.: 17213388 Date: 8 August 2022

A mia nonna

Gentilezza, e forza.

Nei tuoi occhi, Vivo.

«Un paese ci vuole, non fosse che per il gusto di andarsene via. Un paese vuol dire non essere soli, sapere che nella gente, nelle piante, nella terra c'è qualcosa di tuo, che anche quando non ci sei resta ad aspettarti».

CESARE PAVESE, La luna e i falò

All the lights couldn't put out the dark
Running through my heart
[. . .]
Shine
Step into the light
Shine
So bright sometimes
Shine
I'm not ever going back

HARRY STYLES, Lights Up

Il cielo d'Irlanda è un oceano di nuvole e luce [. . .]
Il cielo d'Irlanda ha i tuoi occhi se guardi lassù
Ti annega di verde e ti copre di blu [. . .]
Il cielo d'Irlanda a volte fa il mondo in bianco e nero
Ma dopo un momento i colori li fa brillare più del vero [. . .]
Il cielo d'Irlanda è Dio che suona la fisarmonica
Si apre e si chiude col ritmo della musica [. . .]
Dovunque tu stia ballando con zingari o re
Il cielo d'Irlanda si muove con te
Il cielo d'Irlanda è dentro di te

FIORELLA MANNOIA, Il cielo d'Irlanda

Acknowledgements

I welcome this opportunity to thank Ireland, Irish and non-Irish people I met throughout the years living in this magical land. You all made me feel at home.

I would like to start by thanking my supervisor, Professor Anne Parle-McDermott, for being a continuous source of encouragement and optimism throughout. Thanks, Anne, for choosing me and introducing me to the Irish academic world. I feel you created the conditions for a healthy workplace, built on solid interpersonal connections. Thanks for seeing the human side of us ahead of our work. A special thanks for encouraging me to get out of my comfort zone and helping me acquire confidence in skills that were never my forte - like presenting -. Your feedback has always been precious.

This PhD journey has been long and very hard at times. I couldn't have done it without my beautiful colleagues, now friends. It is hard to explain with words the gratitude I feel for meeting you. Thank you for making me feel always welcomed and part of a fantastic group. Despite the cultural differences and age gap, I never felt foreign. The sense of belonging I experienced with you is something I will always cherish.

Niamh, we shared the same project, desk room, and lab. These things made us work in very close contact. But our bond goes way beyond that. Thanks for always being there when I needed somebody to listen to me. During my most challenging times, you have been a real support, a real friend. I felt heard and seen by you: that means the world to me! I admire your kindness and firmness, your being wise and silly (sometimes!). You are a wonderful human being, and I am proud to call you my friend.

Molly, you brought so much joy, energy, and positive vibrations to our group. Your enthusiasm has really brightened up some of my darkest days. I loved our end-of-the-day conversations when we lost track of time, our car sing-along and nights out (or in!). I learned from you that having a healthier work/life balance was essential for my well-being. Thanks for always supporting me and genuinely celebrating my successes.

Conor, thanks for believing in me and my capacities when it was easier to give way to discouragement. Your constant reminding me that I could do it gave me the strength to keep going. Thanks for the John Mayer lip-sync performances across the PC screens, the "Careless Whisper" duets, the stretching routines and the hilarious "would-you-rather...?" sessions. You have totally contributed to making the last two years unforgettable.

Darren and Arabelle, even though we haven't seen each other on a daily basis, your presence in my journey has been equally important. Arabelle, you have always been so nice to me. Thanks for your patience in listening to my endless rants. Darren, your humour and chill attitude inspired me to take things less seriously. Thanks for your encouraging words (and unexpected shots!). And finally, to Dearbhla, Helena, Ciara and Resti, a big thank you for making the office such a joyful and vibrant environment.

I am also profoundly grateful to all the School of Biotechnology technical staff. Monica, Deirdre, Graham, Kasia, Allison, Carmen, Janice, David, Teresa, and Patricia, thanks for your invaluable help over the years. You have always been thoughtful and generous, and kind. Also, a special shout-out to the school secretary Mary Rafter. Thanks for your joyful and kind words and priceless assistance. Finally, to all lecturers from DCU, School of Biotech goes a sincere thanks for always showing interest in my research and myself as a student and as a person. A few questions on the way, a smile, a nod, a comment, they all seem tiny and meaningless acts, but they add up to build a sense of belonging to a community. So, Thanks, School of Biotech, for making me feel a part of you!

Dr Úna Prendergast: thanks for making those long hours in front of a microscope such a pleasant time. Your vivacity, friendliness, and precious help meant a lot to me in the final stressful months of my PhD.

A special thank goes to Dr Fiona Gallagher for her guidance on how to improve my Academic English. It was an honour and pleasure to have been able to know you. Your passion, enthusiasm and humanity are an inspiration to me.

Now moving out from DCU, my friend Bea is the first person I want to thank. You were the first friend I made in Ireland and will always have a special spot in my heart. We shared a lot of trips, hiking, laughs, serious and deep conversations. I was always welcomed with understanding and compassion, curiosity and generosity. Even if there are 3000 Kms between us now, our connection is solid. Thanks for being the beautiful person and friend that you are.

To my other dear friends Aleksandra, Karolina and Luana, a huge thanks for making my time in Ennis a beautiful memory. Spending time with you filled my heart with joy and peace. Thanks for your goodness, kindness and hospitality.

My deepest gratitude is to my family. To my grandma, my heartfelt appreciation for supporting my education since primary school. Without your financial aid and moral support, I wouldn't be where I am now. You have been a saving light during my darkest times and a beam of hope to get through them. To my father, thanks for always believing in me and my abilities; thanks for being proud of me. To my mother, I always looked at your generosity and kindness as an example to be followed; thanks for always being there when I needed you, even when I wasn't able to say it. To my sister, you are the joy of my life. Thanks for being my number one supporter; thanks for constantly reminding me how much you love and are proud of me. Thanks for always making me laugh. Your presence in my life is the most precious gift, and I am so honoured to be your sister.

I owe a debt of gratitude to Gabri, partner for twelve years and future husband. I must thank you for your infinite patience and understanding. Your presence and continuous support over the years, especially during my PhD journey, have been more important than ever. Thanks for holding space for me, thanks for always being gentle and compassionate. You are the place where I feel safe and can be my truest self: you are home.

Table of Contents

DECLARATION II

ACKNOWLEDGEMENTS V

TABLE OF CONTENTS VIII

ABBREVIATIONS XIV

LIST OF TABLES XVII

LIST OF FIGURES XVIII

ABSTRACT XXI

CHAPTER 1 INTRODUCTION 1

1.1 OVERVIEW 2

1.2 FOLATE COFACTORS: STRUCTURE AND BIOLOGICAL RELEVANCE 2

1.3 ONE-CARBON METABOLISM 5

1.3.1 COMPARTMENTALISATION OF ONE-CARBON METABOLISM 6

1.3.2 ONE-CARBON METABOLISM IN MITOCHONDRIA 9

1.3.3 ONE-CARBON METABOLISM IN CYTOPLASM 11

1.3.3.1 Folate-dependent de novo purine biosynthesis 12

1.3.3.2 De novo thymidylate synthesis 13

1.3.3.3 Homocysteine remethylation pathway 17

1.4 DIHYDROFOLATE REDUCTASE (DHFR) 18

1.4.1 DHFR STRUCTURE AND KINETICS 20

1.4.2 DHFR MECHANISMS OF REGULATION 22

1.4.3 HUMAN DIHYDROFOLATE REDUCTASE GENE FAMILY 26

1.5 DIHYDROFOLATE REDUCTASE 2 (DHFR2) 26

1.6 GENE-EDITING TECHNIQUES 28

1.6.1 CRISPR/CAS: FROM MICROBIAL IMMUNITY TO GENOME EDITING TECHNOLOGY 29

1.6.2 CRISPR/CAS DELIVERY SYSTEMS 32

1.6.3 DNA REPAIR MECHANISMS 33

1.7 AIMS AND OBJECTIVES 35

CHAPTER 2 MATERIALS AND METHODS 39

2.1 MATERIALS 40

2.1.1 REAGENTS 40

2.1.2 SOLUTIONS 43

2.1.3 MAMMALIAN CELL LINES 43

2.1.4 PROKARYOTE CELL LINES 43

2.2 METHODS 44

2.2.1 MICROBIAL CULTURE 44

2.2.2 PLASMID MINI PREP 44

2.2.3 PLASMID MAXI PREP 44

2.2.4 ENDOTOXIN REMOVAL 45

2.2.5 Restriction Enzyme Digestion 45
2.2.5.1 Double Digestion 45
2.2.5.2 Single Digestion 45
2.2.6 DNA Cloning Ligation 46
2.2.6.1 Plasmid Modification by Annealed Oligo Cloning 46
2.2.6.2 In-Fusion Cloning 46
2.2.6.3 Golden Gate Cloning 46
2.2.7 Bacteria transformation 46
2.2.7.1 OneShot TOP10 competent cells 46
2.2.7.2 XL10 Gold Ultracompetent cells 47
2.2.7.3 Stellar competent cells 47
2.2.8 Gel electrophoresis 47
2.2.9 Purification of gel extracted PCR products and vectors 48
2.2.10 Quantification of DNA and RNA concentration 48
2.2.11 Sanger Sequencing 48
2.2.12 Polymerase Chain Reaction (PCR) 49
2.2.13 Mammalian Cell Culture Methods 49
2.2.13.1 Cell Maintenance 49
2.2.13.2 Cell Count 49
2.2.13.3 Recovery of cells from long term storage 50
2.2.13.4 Single-cell Cloning 50
2.2.13.5 Dose-response curve for puromycin selection 50
2.2.14 Mycoplasma test 50
2.2.15 Transfection 51
2.2.16 Cell fixation and staining 51
2.2.17 DNA extraction from mammalian cells 52
2.2.18 RNA extraction from mammalian cells 52
2.2.19 RNA integrity check 53
2.2.20 Reverse-Transcription of RNA to cDNA (and genomic DNA removal) 53
2.2.21 Genomic contamination assay 54
2.2.22 Quantitative Reverse Transcription PCR (RT-qPCR) 54
2.2.23 Protein extraction 55
2.2.24 Bradford protein assay (protein quantification) 55
2.2.25 DHFR enzymatic assay 56
2.2.26 Hybridisation Chain Reaction IHC + RNA-FISH 56
2.2.26.1 Fixation of HepG2 cells on a chambered slide 57
2.2.26.2 Protein Detection (immunochemistry) 57
2.2.26.3 RNA detection (RNA-FISH) 57
2.2.26.4 Amplification (Hybridisation Chain Reaction) 57
2.2.27 Bioinformatic Resources 58
2.2.27.1 Primer BLAST 58
2.2.27.2 CRISPR gRNA design tools 58
2.2.27.3 Geneious Prime 2019.2 and 2022.1.1 58
2.2.27.4 SnapGene Viewer 58
2.2.27.5 R Studio 58

2.2.27.6 Image J 59

CHAPTER 3 AN RNP-BASED CRISPR/CAS METHOD TO GENERATE A DHFR2 KNOCKOUT IN HEPG2 CELLS 60

3.1 INTRODUCTION 61
3.1.1 DHFR2 KNOCKOUT OVERVIEW 61
3.2 CONSTRUCTION OF THE RIBONUCLEOPROTEIN COMPLEX 63
3.2.1 CRRNAS DESIGN 63
3.2.2 RIBONUCLEOPROTEINS IN VITRO ASSEMBLY 64
3.3 TRANSFECTION OF CRISPR/CAS9 RIBONUCLEOPROTEIN (RNP) 66
3.3.1 PRELIMINARY STEPS 66
3.3.2 REVERSE TRANSFECTION 66
3.3.3 SINGLE-CELL CLONING 67
3.4 MULTI-POPULATION SCREENING FOR IDENTIFICATION OF POSITIVE TRANSFORMANTS 67
3.4.1 DIRECT PCR OF GENOMIC DNA FROM MAMMALIAN CELLS 67
3.4.2 PCR CONFIRMATION OF THE DHFR2 ORF DELETION 68
3.4.2.1 Methods 68
3.4.2.2 Results 68
3.4.3 RT-PCR CONFIRMATION OF THE DHFR2 ORF DELETION 69
3.4.3.1 Methods 69
3.4.3.2 Results 70
3.5 DISCUSSION 75

CHAPTER 4 AN MMEJ-ASSISTED CRISPR/CAS APPROACH TO GENERATE A DHFR KNOCK-DOWN IN HEPG2 CELLS 78

4.1 INTRODUCTION 79
4.1.1 OVERVIEW 79
4.1.2 THE PITCH SYSTEM 80
4.1.3 APPLICATION OF THE PITCH STRATEGY FOR DHFR GENE EDITING 81
4.2 SELECTION OF SUITABLE GRNAS AND GENERATION OF RECOMBINANT CLONES (METHODS) 83
4.2.1 INVESTIGATION AND SELECTION OF DHFR SUITABLE GRNAS 83
4.2.2 ADAPTATION OF PITCH VECTORS TO DHFR GENE EDITING EXPERIMENT (OVERVIEW) 84
4.2.3 CONSTRUCTION OF THE ALL-IN-ONE VECTOR – INTERMEDIATE PLASMID 84
4.2.4 CONSTRUCTION OF THE ALL-IN-ONE VECTOR – GOLDEN GATE ASSEMBLY 86
4.2.5 CONSTRUCTION OF THE μΩ VECTOR (MICROHOMOLOGY VECTOR) 88
4.3 CONFIRMATION OF THE CORRECT CONSTRUCTION OF THE DHFR PITCH VECTORS (RESULTS) 90
4.3.1 INSERTION OF THE DHFR-SPECIFIC GRNA IN THE INTERMEDIATE PLASMID 90
4.3.2 SEQUENTIAL INSERTION OF THE DHFR GRNA AND PITCH GRNA IN THE ALL-IN-ONE VECTOR 90
4.3.3 SUBSTITUTION OF THE FBL 5' AND 3' MICROHOMOLOGIES WITH DHFR-SPECIFIC SEQUENCES 91
4.4 DHFR GENE EDITING VIA PITCH: TRANSFECTION AND SELECTION OF MONOCLONAL CELL LINES (METHODS AND RESULTS) 93
4.4.1 DOSE-RESPONSE CURVE FOR PUROMYCIN SELECTION OF HEPG2 CELLS 93
4.4.2 TRANSFECTION OF ALL-IN-ONE AND μΩ VECTORS IN HEPG2 CELLS 94

4.4.3 PUROMYCIN SELECTION 95
4.4.4 FLUORESCENCE OBSERVATION 96
4.4.5 EGFP-2A-PURO^R CASSETTE IN FRAME INSERTION CONFIRMED VIA PCR 96
4.4.5.1 Results 97
4.4.6 SINGLE-CELL CLONING 97
4.5 MULTI-POPULATION SCREENING FOR IDENTIFICATION OF POSITIVE TRANSFORMANTS (METHODS) 98
4.5.1 GENOMIC SCREENING VIA PCR 98
4.5.2 DHFR RNA EXPRESSION CONFIRMATION VIA RT-PCR 99
4.5.3 DHFR ENZYME PRESENCE CONFIRMATION VIA WESTERN BLOT 99
4.5.4 DHFR ENZYMATIC ACTIVITY ASSESSMENT 100
4.6 MULTI-POPULATION SCREENING FOR IDENTIFICATION OF POSITIVE TRANSFORMANTS (RESULTS) 100
4.6.1 DNA SCREENING POINTS OUT FREQUENT INDELS AROUND THE DHFR CLEAVAGE SITE 100
4.6.2 DHFR TRANSCRIPTION IS NOT COMPROMISED 101
4.6.3 MONOCLONAL LINE A5 DISPLAYS A REDUCED EXPRESSION OF DHFR PROTEIN 101
4.6.4 MONOCLONAL LINE A5 PRESENTS 90 % LESS REDUCTASE ACTIVITY 101
4.7 DISCUSSION 114

CHAPTER 5 FUNCTIONAL ANALYSIS OF DHFR2 KNOCKOUT IN HEPG2 CELL LINE. EFFECTS OF LOSS-OF-FUNCTION ON DHFR AND ONE-CARBON METABOLISM 119

5.1 OVERVIEW 120
5.1.1 RATIONALE FOR CELL PROLIFERATION ANALYSIS 120
5.1.2 TARGETED TRANSCRIPTOME ANALYSIS: CHOICE OF GENES 121
5.1.3 GLOBAL TRANSCRIPTOMIC AND PROTEOMIC ANALYSIS: DHFR2 LOF BROADER EFFECTS 122
5.1.4 ASSESSMENT OF THE OCM STATE UPON DHFR2 LOSS: INVESTIGATION OF THE FOLATE PROFILE 123
5.1.5 RESTORATION OF THE ORIGINAL PHENOTYPE: COMPLEMENTATION TEST 123
5.1.6 EVALUATION OF OPPOSING EFFECTS IN DHFR2 OVEREXPRESSION MODEL 123
5.2 METHODS 124
5.2.1 GENERATION OF CELL GROWTH CURVES 124
5.2.2 INVESTIGATION OF THE OCM-ASSOCIATED GENES AND DHFR2-NEIGHBOURING GENES RNA EXPRESSION VIA RT-QPCR 125
5.2.3 MRNA AND LNCRNA GLOBAL EXPRESSION PROFILING VIA MICROARRAY 125
5.2.4 QUANTIFICATION OF OCM INTERMEDIATES VIA MASS SPECTROMETRY 126
5.2.5 COMPLEMENTATION TEST OF THE DHFR2 KNOCKOUT LINE: PLASMID PREPARATION AND SUCCEEDING CELL LINE ANALYSIS 127
5.3 RESULTS 127
5.3.1 THE LOSS OF DHFR2 DOES NOT SIGNIFICANTLY IMPACT CELL PROLIFERATION 127
5.3.2 THE LOSS OF DHFR2 DOWN-REGULATES THE GENES INVOLVED IN THE DE NOVO THYMIDYLATE SYNTHESIS 128
5.3.3 OCM DOWNSTREAM PATHWAYS ARE DOWN-REGULATED IN THE DHFR2 KNOCKOUT 129
5.3.4 DHFR2 LOSS INDUCES A GLOBAL DECREASE IN FOLATE METABOLITES 131
5.3.5 NORMAL DHFR TRANSCRIPT LEVELS ARE RESTORED UPON COMPLEMENTATION 132
5.3.5.1 Transfection of DHFR2 recombinant clone and relative analysis 132
5.3.5.2 Investigation of transcriptome profile in relation to the DHFR2 KO altered gene expression 133

5.3.6 DHFR2 OVEREXPRESSION LEADS TO INCREASED DHFR TRANSCRIPTIONAL LEVELS 133
5.4 DISCUSSION 153

CHAPTER 6 FUNCTIONAL ANALYSIS OF THE DHFR KNOCK-DOWN IN HEPG2 CELLS. EFFECTS OF DHFR IMPAIRMENT ON DHFR2 AND ONE-CARBON METABOLISM 158

6.1 OVERVIEW 159
6.2 METHODS 160
6.3 RESULTS 160
6.3.1 THE REDUCED DHFR ACTIVITY DOES NOT IMPACT THE HEPG2 CELL GROWTH 160
6.3.2 DHFR IMBALANCES SIGNIFICANTLY IMPACT SHMT2, ALDH1L1 AND DHFR2 EXPRESSION 160
6.3.3 DHFR KNOCK-DOWN PRODUCES REARRANGEMENTS OF THE CELLULAR METABOLIC PATHWAYS 161
6.3.4 DHFR IMPAIRMENT INDUCES A GLOBAL DECREASE IN FOLATE METABOLITES 163
6.3.5 TRANSFECTION OF DHFR ORF DOES NOT REVERT THE TRANSCRIPTIONAL PROFILE 164
6.4 DISCUSSION 183

CHAPTER 7 DHFR2 COLOCALISES WITH DHFR IN HUMAN EMBRYONIC HEART 189

7.1 INTRODUCTION 190
7.1.1 RNA MODIFICATIONS ARE CRUCIAL TO RNA FUNCTION 191
7.1.2 A-TO-I EDITING CHANGES THE RNA FATE 192
7.1.3 A-TO-I MODIFICATIONS IN THE 3'UTR OF RNAS MEDIATE PARASPECKLE RETENTION 193
7.1.4 SECONDARY STRUCTURE IMPACTS RNA STABILITY AND FUNCTION 195
7.1.5 DHFR2 INTERACTIONS MAY BE INDICATORS OF FUNCTION 196
7.2 MATERIALS AND METHODS 197
7.2.1 HYBRIDISATION CHAIN REACTION (HCR) 197
7.2.2 HCR EXPERIMENTAL DESIGN AND EXECUTION 197
7.3 RESULTS 201
7.3.1 DHFR2 201 AND 202 FOLD INTO UNIQUE SECONDARY STRUCTURES 201
7.3.2 HCR PROBES SPECIFICALLY BIND TO THEIR TARGETS 201
7.3.3 DHFR2 201 COLOCALISES WITH DHFR RNA 202
7.3.4 SUCCESSFUL MULTIPLEXED DETECTION OF FOUR PROTEIN/RNA PROBES 202
7.3.5 DHFR AND DHFR2 COLOCALISE IN HUMAN EMBRYONIC HEART 203
7.4 DISCUSSION 217

CHAPTER 8 GENERAL DISCUSSION AND CONCLUSIONS 222

8.1 GENERAL DISCUSSION 223
8.2 CONCLUSION AND FUTURE WORK 233

APPENDICES 236

A. PX330A-1X2, CRISPR/CAS9 VECTOR BACKBONE. 237
B. PX330S-2-PITCH, CRISPR/CAS9 PLASMID TARGETING THE EXTREMITIES OF THE EGFP-2A-PUROR CASSETTE. 238

C. pCRIS-PITCHv2-FBL, PITCH DONOR VECTOR FOR EGFP-2A-PUROR INSERTION INTO HUMAN FBL LOCUS. 239

D. pX330A-FBL/PITCH, CRISPR/CAS9 PLASMID TARGETING FBL. 240

E. PCR PRIMERS USED TO CONSTRUCT CRISPR/CAS RECOMBINANT CLONES. 241

F. PCR PRIMERS USED TO CONTROL THE INSERTION OF THE EGFP-2A-PURO CASSETTE. 242

G. PCR PRIMERS USED TO CHECK FOR GENOMIC DNA CONTAMINATION OF RNA SAMPLES (MTHFD1 R653Q ASSAY). 243

H. SANGER SEQUENCING OF DHFR GRNA. 244

I. SANGER SEQUENCING OF THE ALL-IN-ONE VECTOR. 245

J. SANGER SEQUENCING OF THE 5'END OF THE Ωμ VECTOR. 247

K. SANGER SEQUENCING OF THE 3'END OF THE Ωμ VECTOR. 248

L. SANGER SEQUENCING OF DHFR CLEAVAGE AREA FROM THE POLYCLONAL LINE AND ALIGNMENT WITH DHFR WILD TYPE. 249

M. SANGER SEQUENCING OF DHFR CLEAVAGE AREA FROM THE A1 MONOCLONAL LINE AND ALIGNMENT WITH DHFR WILD TYPE. 250

N. SANGER SEQUENCING OF DHFR CLEAVAGE AREA FROM THE A2 MONOCLONAL LINE AND ALIGNMENT WITH DHFR WILD TYPE. 251

O. SANGER SEQUENCING OF DHFR CLEAVAGE AREA FROM THE A4 MONOCLONAL LINE AND ALIGNMENT WITH DHFR WILD TYPE. 252

P. SANGER SEQUENCING OF DHFR CLEAVAGE AREA FROM THE A5 MONOCLONAL LINE AND ALIGNMENT WITH DHFR WILD TYPE. 253

Q. SANGER SEQUENCING OF GENE CASSETTE INSERTION AT ITS 3' END ON DHFR FROM THE POLYCLONAL LINE AND ALIGNMENT WITH EXPECTED AMPLICON. 254

R. SANGER SEQUENCING OF GENE CASSETTE INSERTION AT ITS 3' END ON DHFR FROM THE A1 MONOCLONAL LINE AND ALIGNMENT WITH EXPECTED AMPLICON. 255

S. R STUDIO ANALYSIS ON THE EFFECTS OF HT/NEAA ON HEPG2 CELL GROWTH. 256

T. DIRECT PCR OF GENOMIC DNA FROM MAMMALIAN CELLS. 263

U. PRIMERS USED FOR THE DHFR2 DELETION DETECTION ASSAY (AND DIRECT PCR). 264

V. RT-PCR PRIMERS USED FOR THE 201-202 ASSAY (DETECTION OF DHFR2 201 & 202 ISOFORMS). 264

W. RT-QPCR PRIMERS USED TO TEST THE DHFR2 KNOCKOUT AND DHFR KNOCK-DOWN CELL LINES. 265

X. PCMV6-AC-GFP. MAMMALIAN EXPRESSION VECTOR (ORIGENE). 267

Y. PCMV6-AC-DHFR2-GFP (VARIANT 1). MAMMALIAN EXPRESSION VECTOR (ORIGENE). 268

Z. PCMV6-AC-DHFR-GFP. MAMMALIAN EXPRESSION VECTOR (ORIGENE). 269

AA. COEFFICIENTS OF INTENSITY CORRELATION OF COLOCALIZING OBJECTS IN A DUAL-COLOUR IMAGE. 270

BB. SCREENING OF SUITABLE PAM SITES ACROSS DHFR GENE (ZOOM ON EXON1-INTRON1). 271

REFERENCES 272

Abbreviations

1C	One carbon
ADAR	Adenosine deaminase acting on RNA
AF	Alexa Fluor
AICAR	Aminoimidazolecarboxamide ribonucleotide transformylase
Ala	Alanine
Amp	Ampicillin
Arg	Arginine
Bp	Base pairs
cDNA	Complementary DNA
CHO	Chinese hamster ovary
CRISPR/Cas	Clustered Regularly Interspaced Short Palindromic Repeats / CRISPR-associated protein
Cys	Cysteine
DHF	Dihydrofolate
DHFR	Dihydrofolate Reductase
DHFR2	Dihydrofolate Reductase 2
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
DPBS (or PBS)	Dulbecco's phosphate-buffered saline
dsRNA	Double-strand RNA
dTMP	Deoxythymidine monophosphate (aka thymidylate)
dTMP-SC	dTMP synthesis complex
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphate
dUMP	Deoxyuridine monophosphate
dUTP	Deoxyuridine triphosphate
E. coli	Escherichia coli
EMSA	Electrophoretic Mobility Shift Assay
FAD	Flavin adenine dinucleotide
FBS	Foetal bovine serum
FFPE	Formalin-fixed paraffin-embedded
fMet	Formyl-Methionine
FPGS	Folylpolyglutamate Synthase
FR	Folate Receptor
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GAR	Glycinamide ribonucleotide transformylase
Gln	Glutamine
Glu	Glutamate
H. sapiens	Homo sapiens
HBSS	Hank's Balanced Salt Solution
HCR	Hybridization chain reaction
HT	Hypoxanthine and thymidine
IHC	Immunohistochemistry
IMS	Industrial methylated spirits

IPTG	Isopropyl ß-D-1-thiogalactopyranoside
IR*Alu*s	Inverted Repeat *Alu*s
Kb	Kilo base pairs
KDa	Kilo Dalton
LB	Lysogeny Broth or Luria-Bertani medium
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
Leu	Leucine
lncRNA	Long non-coding RNA
MFT	Mitochondrial Folate Transporter
$MgCl_2$	Magnesium Chloride
miRNA	micro-RNA
mRNA	Messenger RNA
MTHFD 1/2	Methylenetetrahydrofolate dehydrogenase 1/2
MTHFD 1/2 L	Methylenetetrahydrofolate dehydrogenase 1/2 Like
MTHFR	Methylenetetrahydrofolate reductase
MTHFS	Methenyltetrahydrofolate synthetase
MTR	Methionine synthase
MTX	Methotrexate
NADH	Reduced Nicotinamide Adenine Dinucleotide
NADPH	Reduced Nicotinamide Adenine Dinucleotide Phosphate
NEAA	Non-essential amino acids
NEAT1	Nuclear Paraspeckle Assembly Transcript 1
NTDs	Neural Tube Defects
OCM	One-Carbon Metabolism
Oligo dT	Oligo deoxythymine
ORF	Open Reading Frame
PCFT	Proton-Coupled Folate Transporter
PCR	Polymerase Chain Reaction
PIC	Pre-Initiation Complex
PRM LC-MS/MS	Parallel reaction monitoring Liquid chromatography-tandem mass spectrometry
PRPP	Phosphoribosyl Pyrophosphate
PSPC1	Paraspeckle Component 1
Rb	Retinoblastoma
RFC	Reduced Folate Carrier
RNA	Ribonucleic acid
RNA-FISH	RNA fluorescence in situ hybridization
RNP	Ribonucleoprotein
Rpm	Revolutions per minute
RPS3	Ribosomal Protein S3
RT-PCR	Reverse Transcription Polymerase Chain Reaction
RT-qPCR	Quantitative Reverse Transcription Polymerase Chain Reaction
S.O.C.	Super Optimal broth with Catabolite repression
SAM	S-Adenosyl Methionine
Ser	Serine
SHMT 1/2	Serine Hydroxymethyltransferase
SSC	Saline-Sodium Citrate

ssDNA	Single-strand DNA
Str	Streptomycin
SUMO	Small Ubiquitin-like Modifier
TBE	Tris/Borate/EDTA
THF	Tetrahydrofolate
TK	Thymidine Kinase
tRNA	Transfer RNA
TYMS	Thymidylate Synthetase
UMFA	Unmetabolised Folic Acid
UPLC-MS/MS	Ultra-high performance liquid chromatography-tandem mass spectrometry
UTR	Untranslated Region
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

List of Tables

TABLE 3.1 CRRNAS SELECTED TO TARGET DHFR2 GENE. 63

TABLE 4.1 STATUS OF TRANSFECTED CELLS AFTER SIX DAYS OF PUROMYCIN ADMINISTRATION AT DIFFERENT FREQUENCY AND CONCENTRATION. 104

TABLE 7.1 DIFFERENT PROBE COMBINATIONS WITH RELATIVE FLUOROPHORE AND AMPLIFIER POOL. 200

List of Figures

FIGURE 1.1 CHEMICAL STRUCTURE OF FOLIC ACID AND TETRAHYDROFOLATE. 4

FIGURE 1.2 PRINCIPAL FOLATE INTERMEDIATES. 5

FIGURE 1.3 SCHEMATIC OF ONE-CARBON METABOLISM. 8

FIGURE 1.4 MITOCHONDRIAL ONE-CARBON METABOLISM. 10

FIGURE 1.5 CYTOPLASMIC OCM, PURINE DE NOVO SYNTHESIS. 13

FIGURE 1.6 CYTOPLASMIC OCM, THYMIDYLATE DE NOVO SYNTHESIS. 15

FIGURE 1.7 THYMIDYLATE DE NOVO SYNTHESIS IN NUCLEUS. 16

FIGURE 1.8 CYTOPLASMIC OCM, REMETHYLATION OF HOMOCYSTEINE AND METHIONINE CYCLE. 18

FIGURE 1.9 DHFR SECONDARY AND TERTIARY STRUCTURE. 20

FIGURE 1.10 DHFR CATALYTIC REACTION. 21

FIGURE 1.11 DHFR LNCRNA-MEDIATED TRANSCRIPTIONAL REGULATION. 23

FIGURE 1.12 DHFR AUTO-REGULATION MODEL. 25

FIGURE 1.13 CLUSTAL PAIRWISE ALIGNMENT OF DHFR AND DHFR2 PREDICTED SEQUENCE. 27

FIGURE 1.14 CRISPR/CAS-MEDIATED DNA INTERFERENCE IN ADAPTIVE IMMUNITY IN PROKARYOTES. 30

FIGURE 1.15 ENGINEERED CAS9 ENZYME ACTIVATED BY SYNTHETIC SINGLE-GUIDE RNA (SGRNA). 31

FIGURE 1.16 SCHEMATIC OF THREE DNA REPAIR MECHANISMS. 35

FIGURE 1.17 COMPARISON BETWEEN THE GENE-EDITING METHODS. 36

FIGURE 3.1 FORMATION OF THE CRRNA:TRACRRNA DUPLEX, AND ITS ASSOCIATION WITH CAS9. 64

FIGURE 3.2 SCHEMATIC REPRESENTATION OF THE DHFR2 ORF DELETION SITES AND PCR DETECTION. 71

FIGURE 3.3 DHFR2 DELETION DETECTION PCR ASSAY ON POLYCLONAL LINE. 72

FIGURE 3.4 DHFR2 DELETION DETECTION PCR ASSAY ON MONOCLONAL LINES. 73

FIGURE 3.5 RT-PCR DHFR2 201-202 ASSAY ON THE FIVE DHFR2 KO MONOCLONAL LINES. 74

FIGURE 4.1 DIAGRAM OF THE CASSETTE INSERTION ACCORDING TO THE PITCH SYSTEM. 81

FIGURE 4.2 PITCH STRATEGY APPLIED TO THE DHFR KNOCKOUT. 82

FIGURE 4.3 REPRESENTATION OF GUIDE RNA (GRNA) STRUCTURE AND FUNCTION. 83

FIGURE 4.4 DOUBLE BBSI RESTRICTION SITE AT U6 PROMOTER 3' END. 85

FIGURE 4.5 SECTION OF THE INTERMEDIATE PLASMID (PREVIOUSLY PX330A-1X2), INDICATING THE LOCATION OF THE LAC OPERON AND BSA I RESTRICTION SITES. 86

FIGURE 4.6 CONSTRUCTION OF ALL-IN-ONE VECTOR BY GOLDEN GATE ASSEMBLY. 87

FIGURE 4.7 CONSTRUCTION OF THE μΩ VECTOR VIA PCR AND IN-FUSION LIGATION. 89

FIGURE 4.8 DHFR KNOCKOUT MODEL OF TRANSFECTION AND SELECTION IN HEPG2 CELL LINE. 102

FIGURE 4.9 PUROMYCIN SELECTION OF HEPG2 CELLS POST-TRANSFECTION. 103

FIGURE 4.10 PCR AMPLIFICATION OF GENE CASSETTE EXTREMITIES IN CONTINUITY WITH DHFR. 105

FIGURE 4.11 CASSETTE EXTREMITIES AMPLIFICATION. 106

FIGURE 4.12 SINGLE-CELL CLONING TECHNIQUE. 107

FIGURE 4.13 PCR ASSAYS USED TO IDENTIFY THE STATUS OF DHFR KNOCKOUT. 108

FIGURE 4.14 PCR AMPLIFICATION OF THE 3' END ASSAY AND DHFR FULL-LENGTH ASSAY ON DHFR KO MONO- AND POLYCLONAL LINES. 109

FIGURE 4.15 SEQUENCING DATA OF DHFR CLEAVAGE AREA FROM THE POLYCLONAL LINE AND ALIGNMENT WITH DHFR WILD-TYPE. TYPE LINE USING THE DHFR FULL-LENGTH ASSAY. 110

FIGURE 4.16 RT-PCR AMPLIFICATION OF DHFR MAIN ISOFORM ASSAY. 111

FIGURE 4.17 WESTERN BLOT IMAGING. 112

FIGURE 4.18 DHFR ENZYMATIC ASSAY. 113

FIGURE 5.1 DHFR2 AND NSUN3 SHARED PROMOTER. 122

FIGURE 5.2 PERMUTATION TEST FOR PAIRWISE COMPARISONS. 135

FIGURE 5.3 GROWTH CURVES OF HEPG2 DHFR2 KNOCKOUT LINE COMPARED TO HEPG2 WILD-TYPE. 136

FIGURE 5.4 ENDOGENOUS LEVELS OF DHFR2, DHFR, TYMS AND SHMT1 RNAS IN DHFR2 KO AND HEPG2 LINES. 137

FIGURE 5.5 SCATTER PLOT OF DHFR2 KO LNCRNA AND MRNAS AFTER FILTERING. 138

FIGURE 5.6 VOLCANO PLOT OF DHFR2 KO LNCRNA AND MRNAS. 139

FIGURE 5.7 PATHWAY ANALYSIS OF THE DIFFERENTIALLY EXPRESSED (DE) RNAS IN DHFR2 KO. 140

FIGURE 5.8 GO ANALYSIS OF THE DOWN-REGULATED RNAS IN DHFR2 KO (PIE CHARTS). 141

FIGURE 5.9 GO ANALYSIS OF THE DOWN-REGULATED RNAS IN DHFR2 KO (BAR CHART). 142

FIGURE 5.10 GO ANALYSIS OF THE UP-REGULATED RNAS IN DHFR2 KO (PIE CHARTS). 143

FIGURE 5.11 GO ANALYSIS OF THE UP-REGULATED RNAS IN DHFR2 KO (BAR CHART). 144

FIGURE 5.12 FOLATE METABOLITES PROFILE IN DHFR2 KO AND HEPG2 WILD-TYPE LINES. 145

FIGURE 5.13 CONCENTRATION OF FOLATE INTERMEDIATES AND THEIR RELATIVE POLYGLUTAMATED FORMS IN DHFR2 KO AND HEPG2 WILD-TYPE LINES. 146

FIGURE 5.14 RELATIVE PROPORTIONS (EXPRESSED AS % OF TOTAL FOLATE) OF FOLATE METABOLITES IN DHFR2 KO AND HEPG2 WILD-TYPE LINES. 147

FIGURE 5.15 DHFR2-GFP EXPRESSION IN DHFR2 KNOCKOUT LINE (FLUORESCENCE). 148

FIGURE 5.16 DHFR2-GFP RNA EXPRESSION IN DHFR2 KNOCKOUT LINE. 149

FIGURE 5.17 DHFR2-GFP PROTEIN EXPRESSION IN DHFR2 KNOCKOUT LINE. 150

FIGURE 5.18 GENE EXPRESSION AFTER TRANSFECTION WITH PCMV6-AC-DHFR2 IN DHFR2 KNOCKOUT CELLS. 151

FIGURE 5.19 DHFR GENE EXPRESSION AFTER TRANSFECTION WITH PCMV6-AC-DHFR2 IN HEPG2 CELLS. 152

FIGURE 6.1 PERMUTATION TEST FOR PAIRWISE COMPARISONS. 166

FIGURE 6.2 GROWTH CURVES OF HEPG2 DHFR KNOCK-DOWN LINE COMPARED TO HEPG2 WILD-TYPE. 167

FIGURE 6.3 ENDOGENOUS LEVELS OF DHFR, DHFR2, SHMT2 AND ALDH1L1 RNAS IN DHFR KD AND HEPG2 LINES. 168

FIGURE 6.4 SCATTER PLOT OF DHFR KD LNCRNA AND MRNAS AFTER FILTERING. 169

FIGURE 6.5 VOLCANO PLOT OF DHFR KD LNCRNA AND MRNAS. 170

FIGURE 6.6 PATHWAY ANALYSIS OF THE DIFFERENTIALLY EXPRESSED (DE) RNAS IN DHFR KD. 171

FIGURE 6.7 GO ANALYSIS OF THE DOWN-REGULATED RNAS IN DHFR KD (PIE CHARTS). 172

FIGURE 6.8 GO ANALYSIS OF THE DOWN-REGULATED RNAS IN DHFR KD (BAR CHART). 173

FIGURE 6.9 GO ANALYSIS OF THE UP-REGULATED RNAS IN DHFR KD (PIE CHART). 174

FIGURE 6.10 GO ANALYSIS OF THE UP-REGULATED RNAS IN DHFR KD (BAR CHART). 175
FIGURE 6.11 FOLATE METABOLITES PROFILE IN DHFR KD AND HEPG2 WILD-TYPE LINES. 176
FIGURE 6.12 CONCENTRATION OF FOLATE INTERMEDIATES AND THEIR RELATIVE POLY-GLUTAMATED FORMS IN DHFR KD (AT DIFFERENT TIMES AND CONDITIONS) AND HEPG2 WILD-TYPE LINES. 177
FIGURE 6.13 RELATIVE PROPORTIONS (EXPRESSED AS % OF TOTAL FOLATE) OF FOLATE METABOLITES IN DHFR KD (AT DIFFERENT TIMES AND CONDITIONS) AND HEPG2 WILD-TYPE LINES. 178
FIGURE 6.14 DHFR-GFP EXPRESSION IN DHFR KNOCK-DOWN LINE (FLUORESCENCE). 179
FIGURE 6.15 DHFR-GFP RNA EXPRESSION IN DHFR KNOCK-DOWN LINE. 180
FIGURE 6.16 DHFR-GFP PROTEIN EXPRESSION IN DHFR KNOCKOUT LINE. 181
FIGURE 6.17 GENE EXPRESSION AFTER TRANSFECTION WITH PCMV6-AC-DHFR IN DHFR KNOCK-DOWN LINE. 182
FIGURE 6.18 DHFR/DHFR2 FEEDBACK LOOP REGULATION. 188
FIGURE 7.1 SCHEMATIC OF DHFR2 201 AND 202 ISOFORMS. 191
FIGURE 7.2 PARASPECKLE RETENTION MECHANISM. 194
FIGURE 7.3 HCR IHC + RNA-FISH MECHANISM OF ACTION. 199
FIGURE 7.4 DHFR2 201 & 202 SECONDARY STRUCTURE PREDICTION. 204
FIGURE 7.5 CONFOCAL IMAGING OF PROTEINS AND RNAS DETECTED VIA HCR IHC AND HCR RNA-FISH FOR SINGLE PROBES IN HEPG2 CELLS ON A SLIDE. 205
FIGURE 7.6 CONFOCAL IMAGING OF RPS3 AND DHFR2 201 PROTEIN. 206
FIGURE 7.7 CONFOCAL IMAGING OF RPS3 AND DHFR2 202 PROTEIN. 207
FIGURE 7.8 CONFOCAL IMAGING OF PSPC1 AND DHFR2 201 PROTEIN. 208
FIGURE 7.9 CONFOCAL IMAGING OF PSPC1 AND DHFR2 202 PROTEIN. 209
FIGURE 7.10 CONFOCAL IMAGING OF DHFR AND DHFR2 201 PROTEIN. 210
FIGURE 7.11 CONFOCAL IMAGING OF DHFR AND DHFR2 202 PROTEIN. 211
FIGURE 7.12 CONFOCAL IMAGING OF DHFR ^{RNA} AND DHFR2 201 RNA. 212
FIGURE 7.13 CONFOCAL IMAGING OF DHFR ^{RNA} AND DHFR2 202 RNA. 213
FIGURE 7.14 CONFOCAL IMAGING OF RPS3, PSPC1, DHFR2 201 AND DHFR2 202 PROTEIN. 214
FIGURE 7.15 CONFOCAL IMAGING OF DHFR, DHFR ^{RNA}, DHFR2 201 AND DHFR2 202 PROTEIN. 215
FIGURE 7.16 CONFOCAL IMAGING OF DHFR, DHFR2 201 AND DHFR2 202 PROTEIN. 216
FIGURE 8.1 DHFR ENZYMATIC ASSAY. 226
FIGURE 8.2 FOLATE METABOLITES PROFILE IN DHFR KD, DHFR2 KO AND HEPG2 WILD-TYPE LINES. 227
FIGURE 8.3 DHFR KNOCK-DOWN MODEL: EFFECTS OF DHFR IMPAIRMENT ON DHFR AND DHFR2. 228
FIGURE 8.4 DHFR2 KNOCKOUT MODEL: EFFECTS OF DHFR2 LOSS ON DHFR. 229
FIGURE 8.5 DHFR EXPRESSION LEVELS IN RESPONSE TO DHFR2. 230
FIGURE 8.6 DHFR/DHFR2 FEEDBACK LOOP REGULATION. 231

Abstract

Paola Drago

Deciphering the function of the human Dihydrofolate Reductase 2 gene

DHFR2 is part of the large Dihydrofolate reductase gene family, in which DHFR is the mainly characterised member. For a long time, DHFR was considered the only gene of the family able to make a functional protein, with the remaining members catalogued as pseudogenes. DHFR has been thoroughly studied, considering its crucial role in facilitating the entry of folic acid in the One-Carbon Metabolism pathway, and its implications in neural tube defects and cancer. In the past decade, DHFRP4 was renamed DHFR2 and classified as a protein-coding gene since it harbours an intact open reading frame that can result in a protein, in vitro. However, no information was available in the literature about the endogenous DHFR2 protein, its enzymatic activity, or cellular localisation.

This thesis aimed to decipher the role of native DHFR2 and its overall function in relation to DHFR and the entire One-Carbon Metabolism. HepG2 knock out cell lines, lacking either DHFR or DHFR2, were established and comprehensively investigated. In particular, the effects of DHFR2 loss were examined in relation to DHFR and other OCM-related genes. The analysis of RNA expression patterns, performed via RTqPCR and RNA microarray, indicated that the lack of DHFR2 directly impacted OCM-related genes, with particular attention on DHFR, whose RNA levels showed a drastic decrease in abundance. The loss of DHFR2 has a broader impact on the One-Carbon Metabolism, as demonstrated by the Folate Metabolite Profiling study. It revealed that the concentrations of folate intermediates halved in the DHFR2 KO line compared to the wild type. Pathway Analysis, performed upon RNA microarray data, indicated that DNA replication and repair pathways and pyrimidine metabolism were mainly downregulated after DHFR2 knock out. This can be explained as a direct consequence of DHFR down-regulation and OCM reduced activity. In the light of the overall results, we propose that DHFR2 acts as a long non-coding RNA, primarily involved in regulating DHFR expression and, consequently, One-Carbon Metabolism.

Chapter 1
Introduction

1.1 Overview

One Carbon Metabolism (OCM) is an intricate metabolic network which represents the hub for many fundamental cell pathways, such as purines and thymidylate synthesis, methylation, and metabolism of some amino acids. As the name implies, OCM facilitates the one-carbon groups to be processed and activated by chemical modification, starting with folate-derived donors (Fox & Stover, 2008). Natural or synthetic forms of folate can enter OCM solely through the reduction mediated by the enzyme Dihydrofolate Reductase. Ergo, it is considered of extraordinary importance for the maintenance of OCM and, in turn, of normal cell functions (Bailey & Caudill, 2012). The recent discovery of a second active Dihydrofolate Reductase (DHFR2), which is expressed and possibly localised to the mitochondria (Anderson et al., 2011; McEntee et al., 2011), opens up a whole new set of questions and opportunities to better understand the role of DHFR2 and its influence in OCM.

1.2 Folate cofactors: structure and biological relevance

The generic term folate indicates a group of water-soluble compounds that present a similar chemical structure to folic acid or pteroylmonoglutamate (Lucock, 2000). Folic acid and its derivatives are all included in the class of vitamin B_9 (Naderi & House, 2018). The discovery of folate is attributed to Lucy Wills in 1931 (Wills, 1931). Ten years after, Mitchell, Snell and Williams isolated the molecule for the first time from spinach leaves (Mitchell et al., 1941, 1944). Since then, many discoveries have been made on the role of folate in metabolic pathways and its involvement in health and disease (Bailey et al., 2015; Bo et al., 2020; Molloy & Scott, 2001; Stover, 2004, 2012).

Folate is an essential vitamin, necessary for normal cell functioning. Suboptimal folate intake can lead to various pathologies, i.e., some forms of anaemia, cardiovascular diseases and cancer (Blom & Smulders, 2011; Scaglione & Panzavolta, 2014). Therefore, optimal folate status is protective for such pathologies and has been shown to prevent neural tube defects during embryogenesis, a period of persistent cell division (Ebara, 2017; Ohrvik & Witthoft, 2011; Viswanathan et al., 2017).

Neural tube defects (NTDs) are a major cause of neonatal deaths and lifelong severe pathologies (Greene & Copp, 2014). Periconceptional folate supplementation was proven to be a protective factor (Czeizel & Dudás, 1992; 'Prevention of Neural Tube Defects', 1991), but due to the unplanned nature of most pregnancies, this tool was not sufficient to reduce the burden of NTDs (Fischer et al., 2017). Considering that Western diets are often linked to malnutrition with suboptimal levels of folate (García-Montero et al., 2021), many countries have introduced mandatory food fortification in a broader attempt to reduce the impact of congenital birth defects and other pathologies (Crider et al., 2011; Mills, 2017).

As mammals cannot synthesise the pteroylglutamic acid, which is the core of the vitamin, they rely on exogenous sources, such as food or supplements (Iyer & Tomar, 2009). The folate core consists of three moieties – pterin, para-aminobenzoic acid and L-glutamic acid – as shown in Figure 1.1. The natural forms of folate present a reduced pteridine ring and additional molecules of glutamic acid linked through gamma-peptide bonds (Finglas & Wright, 2002), making them polyglutamated. Food folates are naturally present in foods such as leafy vegetables, pulses, citrus fruits, and liver (Nazki et al., 2014; Ohrvik & Witthoft, 2011). The synthetic form of folate occurs in a monoglutamyl state and is the most oxidised form of folate, making it very suitable for food fortification and supplementation (Iyer & Tomar, 2009; Naderi & House, 2018; *Office of Dietary Supplements - Folate*, n.d.)

The polyglutamyl forms of folate are required to be hydrolysed into their monoglutamic forms prior to be absorbed in the intestines, whilst folic acid (a term referring to the synthetic variant only) can be absorbed immediately as it is not polyglutamated (Cochrane et al., 2020). Although folic acid is chemically stable, it cannot be actively used as a coenzyme. In fact, it needs to be reduced into tetrahydrofolate, the bioactive form of the vitamin (Fig. 1.1) (Shane, 2008)

Once the monoglutamated serum folates are transported into the cells, five to eight glutamate residues are added through a γ-linked peptide bond by the enzyme Folylpoly-gamma-glutamate synthetase (FPGS) (Hoffbrand et al., 1977; Moran, 1999). The aim is to increase their affinity for enzymes (Fowler, 2001) and prevent their immediate export

from the cell (Schirch & Strong, 1989). Therefore, cellular folate is predominantly polyglutamated.

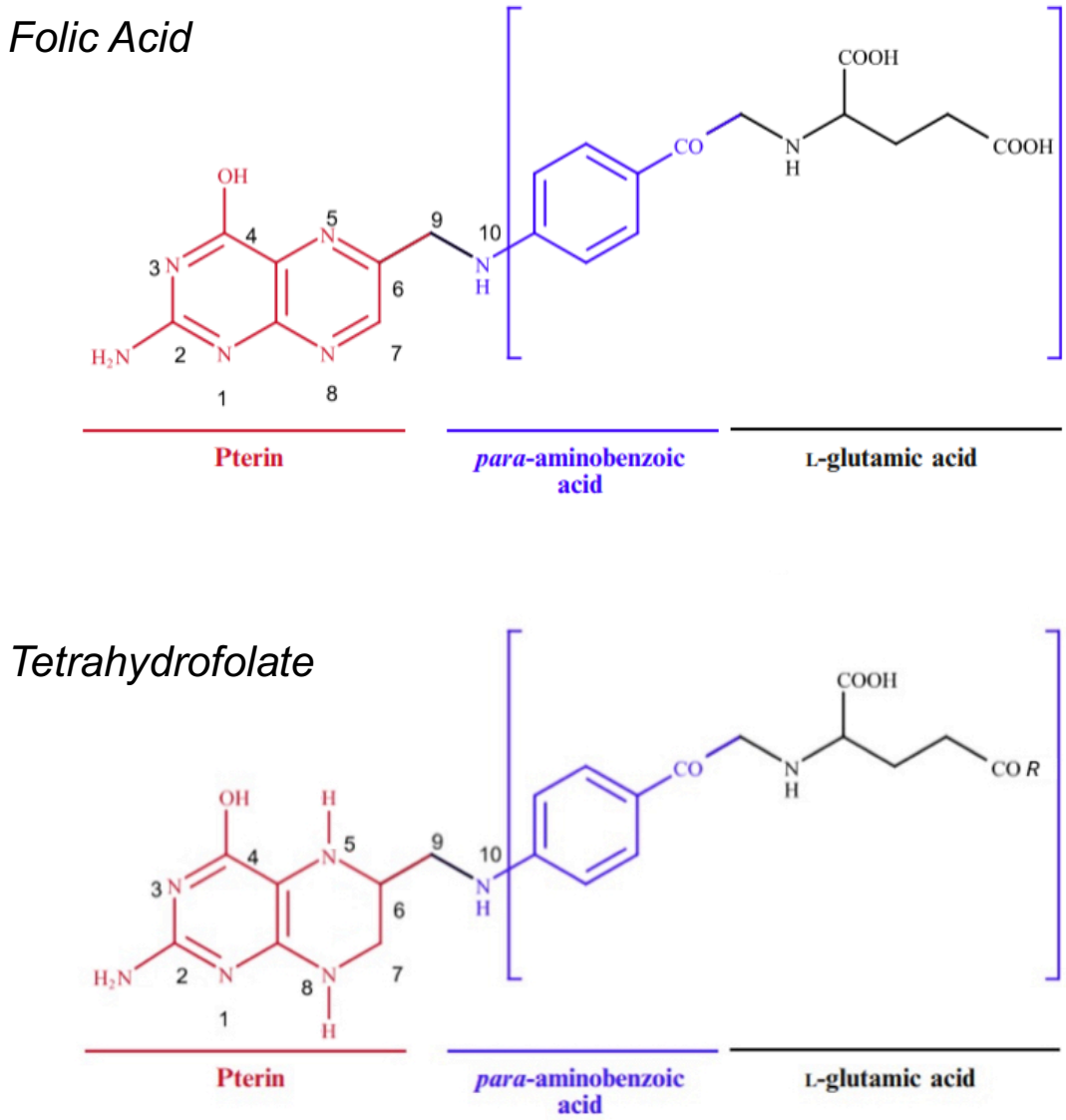


Figure 1.1 Chemical structure of Folic Acid and Tetrahydrofolate. The three moieties composing the molecules are indicated in different colours: pterin (red), para-aminobenzoic acid (purple), L-glutamic acid (black). N5 and N8 of the pterin ring are reduced in tetrahydrofolate (THF). THF is typically polyglutamated as indicated by the terminal -R. Image adapted from Naderi & House, 2018 using Paint Pad Lite v. 6.0.3.

Folate is a particularly important molecule in cell metabolism due to its ability to accept and donate one-carbon (1C) groups (Bailey & Gregory, 1999). These activated groups can go through a series of oxidation and reduction reactions, varying their structure into

methanol, formaldehyde or formate (Shane, 2008; Stover & Field, 2011). Therefore, the folate molecule acts both as a carrier and coenzyme. The 1C groups, in diverse redox states, can eventually be donated to different substrates, feeding important cellular processes (Ducker & Rabinowitz, 2017). Figure 1.2 illustrates the main folate intermediates. The ensemble of reactions entailing the conversion of folate into its several intermediates is known as One-Carbon Metabolism.

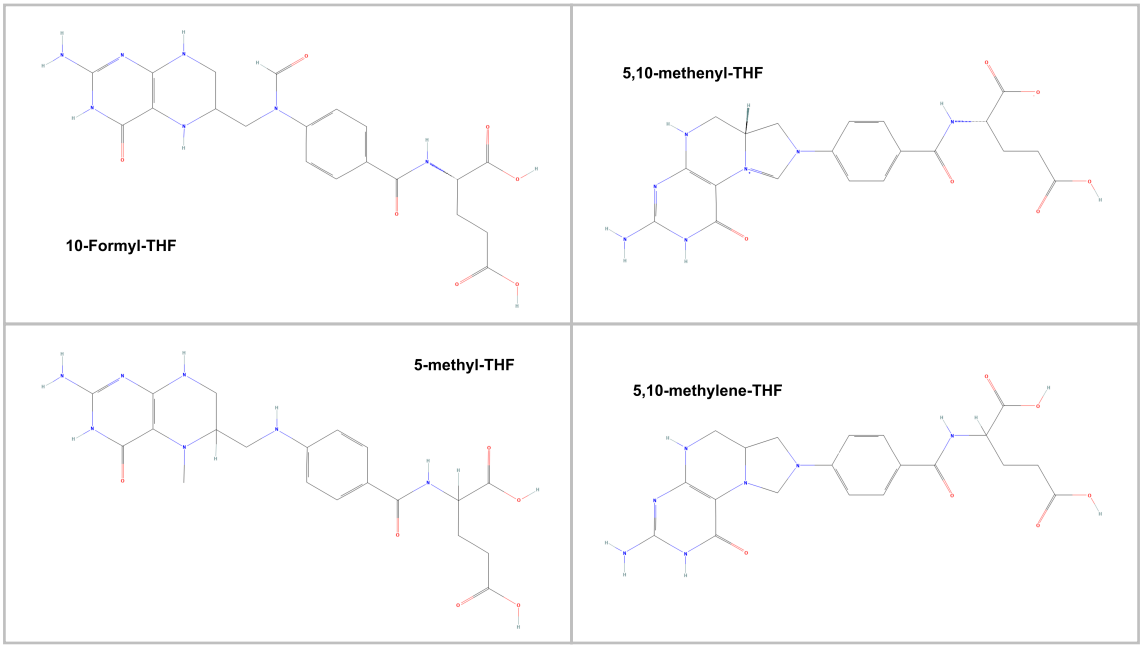


Figure 1.2 Principal folate intermediates. Formyl-THF, the most oxidised coenzyme, carries a formyl group at N10. In Methenyl- and Methylene-THF, the carbon group bridges between N5 and N10. Methyl-THF is the most reduced form with a methyl group at N5. Images taken from PubChem (NCBI).

1.3 One-Carbon Metabolism

One-carbon metabolism is an intricate network of metabolic reactions, which constitutes the hub for essential cell pathways such as purines and thymidylate synthesis, methylation processes, and amino acids homeostasis (Fox & Stover, 2008). OCM and annexed pathways work in a balanced manner responding to the cell's metabolic needs; therefore, they are tightly regulated (Clare et al., 2019; Stover, 2004). Imbalances within OCM are linked to several pathologies; cancer, neural tube defects, and cardiovascular diseases are among the most studied in this regard (Stover, 2012).

One-carbon metabolism is highly partitioned in Eukaryotes and takes place in three different cellular compartments: cytoplasm, mitochondria and nucleus (Fox & Stover, 2008; Scotti et al., 2013; Tibbetts & Appling, 2010). Many enzymes come into play within OCM, aiding in the chemical rearrangement of one-carbon groups. Isozymes, which catalyse the same reactions, are often found in two distinct cellular compartments, i.e., cytoplasm and mitochondria (Anderson & Stover, 2009; Krupenko & Krupenko, 2018). Other folate enzymes, instead, can translocate to a separate organelle to respond to specific requests; for example, the migration of the thymidylate synthesis complex (dTMP-SC) to the nucleus during cell division (Chon et al., 2017).

Tetrahydrofolate (THF) is able to carry one-carbon units at different oxidation states upon interconversion mediated by the OCM enzymes (Ducker & Rabinowitz, 2017). The different forms of folate are specifically required for particular pathways. For example, 10-formyl-THF is utilised for synthesising the purine ring in the cytoplasm (Baggott & Tamura, 2015) and for formylating methionyl-tRNA in the mitochondria (Minton et al., 2018); 5,10-methylene-THF mediates the conversion of dUMP to dTMP, in the nucleus and mitochondria (Anderson et al., 2011; Woeller et al., 2007); and the one-carbon moiety of 5-methyl-THF serves to re-methylate methionine from homocysteine in the cytoplasm (Crider et al., 2012). As the concentration of folate-binding proteins exceeds that of their substrates, each biosynthetic pathway is in constant competition for the limiting pool of folates (Herbig et al., 2002). In addition to the compartmentalisation (Appling, 1991) (discussed in detail in the next section), metabolic channelling was proven to be an efficient way to maintain continuity between groups of reactions and avoid the diffusion of the substrates. Folate channelling is also crucial for preventing oxidative degradation of the labile folate coenzymes; it was demonstrated that folate intermediates are protected from random oxidative events when bound to enzymes (Stover & Field, 2011; Swarbrick et al., 2008).

1.3.1 Compartmentalisation of One-Carbon Metabolism

In the 1950s, mitochondria were identified as the site of oxidation of the one-carbon donors. With the advent of genetic and biochemical approaches, an increasing number of enzymes involved in the one-carbon metabolism were identified (Tibbetts & Appling,

2010). The discovery of these metabolic components and their functions led to the proposal of an OCM model where the cytoplasmic and mitochondrial pathways run parallel and in opposite directions, creating a cycle (Appling, 1991).

The compartmentalisation of OCM pathways can be accomplished via the physical separation of the enzymes, which localise to specific organelles due to their genetic and structural differences (Fig. 1.3). The composition of the folate pools differs between the cytoplasm and mitochondria (Horne et al., 1989; Lin et al., 1993; Shane, 2001). Such distinct separation is reached via polyglutamylation. In fact, polyglutamated folates are not subject to transportation across the mitochondrial membranes. Thus, FPGS activity traps folates in their respective compartments (Lawrence et al., 2014). Furthermore, the formation of metabolons - multi-enzyme complexes - helps maintain folate pools in specific compartments, supporting the orderly flux of the activated carbons (Nijhout et al., 2004; Scotti et al., 2013). Other than attaching glutamic tails to folates, FPGS is responsible for channelling folates to the metabolon compartments contrasting metabolites diffusion and random oxidation (Stark et al., 2021).

The compartmentalisation, which occurs on a larger scale by separating the pathways in different cell compartments (as stated above), can be also achieved on a smaller scale by the formation of multi-enzyme complexes. As a result, spatial separation of the metabolic reactions is achieved. Moreover, metabolons that assemble solely in response to metabolic needs, create an additional temporal partition within the OCM (Lan et al., 2018; Scotti et al., 2013). In fact, the cell cycle plays an essential role in regulating the retrieval of specific folate enzymes in a given cellular compartment, as demonstrated by the translocation of the dTMP-SC to the nucleus upon SUMOylation (Anderson et al., 2007; Woeller et al., 2007), or the MTHFS-dependent delivery of 10-formyl-THF to the purinosome to enhance purine de novo synthesis (Field et al., 2011).

Finally, mitochondria have a particularly relevant role within OCM as providers of one-carbon units (in the form of formate). About 75 % of one-carbon groups involved in the cytoplasmic folate reactions come from mitochondria (Pike et al., 2010) via serine catabolism. Formate production is dependent on respiration, as well as the balance of the NAD+/NADH ratio (with MTHFD2L/1L producing NADH) (Garcia-Martinez & Appling, 1993). Mitochondrial oxidation uses NADH cofactors, thus leading to the maintenance

of a differential electrochemical potential of NADH/NADPH that finally helps with shuttling formate to the cytoplasm (Ducker & Rabinowitz, 2017).

The architecture of OCM follows the rule of compartmentalisation for maintaining high efficiency, implementing oxidation of one-carbon groups in the mitochondria with reduction of the same species in the cytoplasm (Ducker et al., 2016). Despite the compartmental design of OCM appearing to be highly efficient, the evolutionary advantages of this dual setting have yet to be fully understood. A recent study suggests that this composition enables major flexibility due to the possibility to uncouple the two pathways (Zheng et al., 2018).

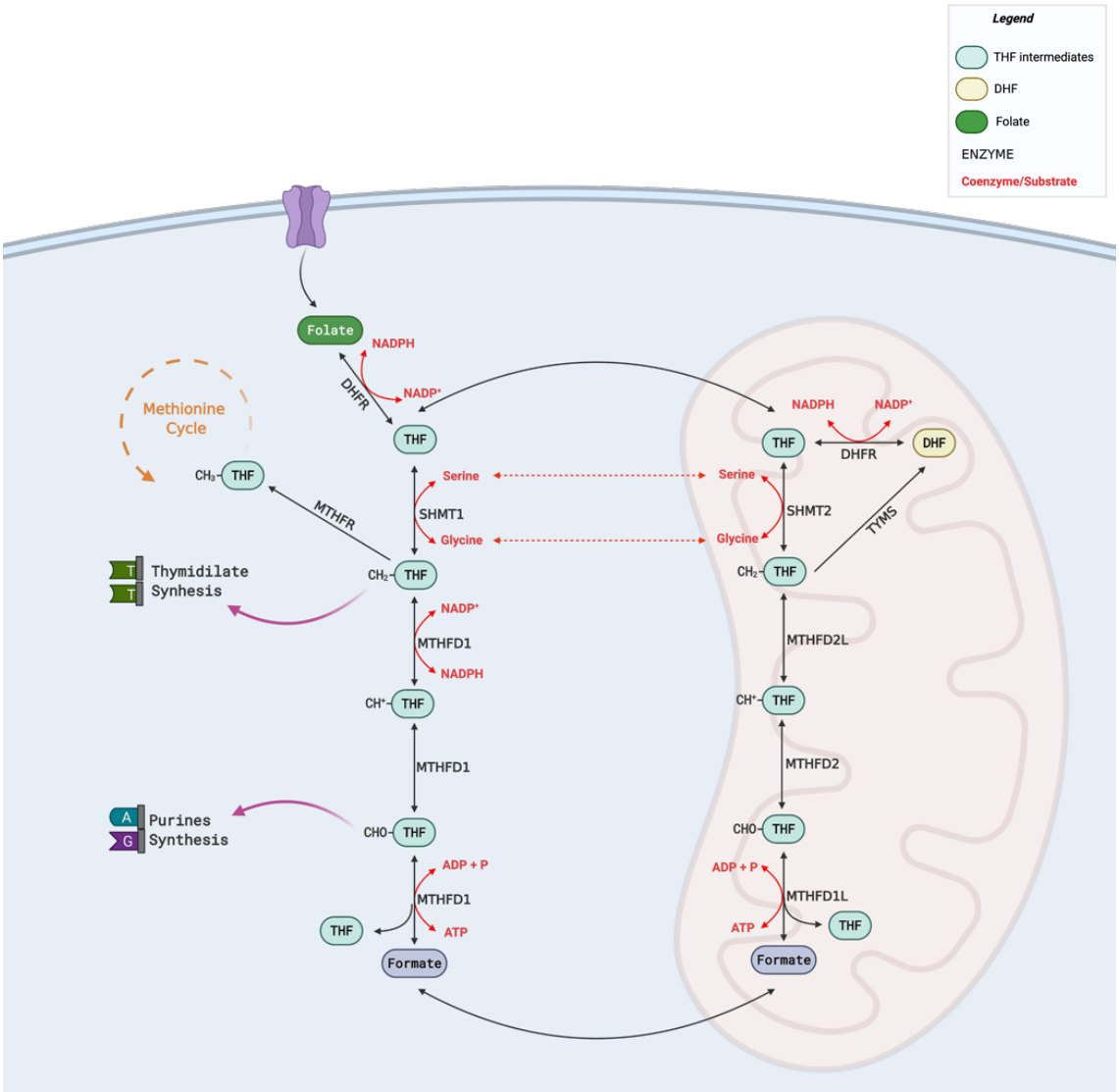


Figure 1.3 Schematic of One-Carbon Metabolism. Mitochondrial and cytoplasmic pathways are specular. Most of the reactions can work in both directions but the general sense of reaction proceeds toward oxidation in the mitochondria, and toward reduction in the cytoplasm. Image created with BioRender.com; adapted from Bryant et al., 2018.

1.3.2 One-Carbon Metabolism in Mitochondria

Mitochondria contain roughly 40 % of the total cell folate (Lin et al., 1993; Shin et al., 1976). Upon import mediated by MFT (mitochondrial folate transporter) (Lawrence et al., 2011; Titus & Moran, 2000) and polyglutamation by FPGS (Lawrence et al., 2014), the folate intermediates are activated via the addition of 1C groups derived from the catabolism of amino acids such as serine, glycine, sarcosine and choline (Pike et al., 2010; Tibbetts & Appling, 2010). The 1C flux proceeds toward oxidation to finally produce formate, which will serve as the primary substrate for the cytoplasmic OCM (Stover & Field, 2011). The mitochondrial OCM is mainly supported by serine catabolism, which in turn produces glycine, an amino acid involved in other cellular functions such as heme biosynthesis, transsulfuration pathway, and protein synthesis (Brosnan & Brosnan, 2020; Meiser et al., 2016; Pietzke et al., 2020; Razak et al., 2017). As demonstrated by several studies (Ducker et al., 2016; Taylor & Hanna, 1982; Zheng et al., 2018), disruption of mitochondrial OCM leads to glycine auxotrophy, indicating that glycine production is a crucial part of the mitochondrial OCM. The mitochondrial enzymes can be variably expressed in different tissues and developmental stages; also, their chemical equilibrium can change in accordance with metabolic requests favouring different outcomes. For example, in embryonic development, the mitochondrial flux favours serine catabolism and formate production to supply 1C unit to the cytoplasmic purine biosynthesis. In contrast, in adult tissues, the OCM equilibrium in mitochondria may favour serine production (Christensen & MacKenzie, 2006).

SHMT2 (Serine Hydroxymethyltransferase 2) converts serine to glycine, attaching the 1C moiety to THF, thus forming methylene-THF. Subsequently, the bifunctional enzyme MTHFD2/L (Methylenetetrahydrofolate dehydrogenase 2/L) oxidises methylene-THF to formyl-THF, passing through the intermediate methenyl-THF. The formyl-THF is the substrate of MTHFD1L, which frees THF by making formate, which in turn is released to the cytoplasm (Christensen & MacKenzie, 2006; Fox & Stover, 2008; Lan et al., 2018; Tibbetts & Appling, 2010). A schematic of the mitochondrial OCM reactions is illustrated in Figure 1.4.

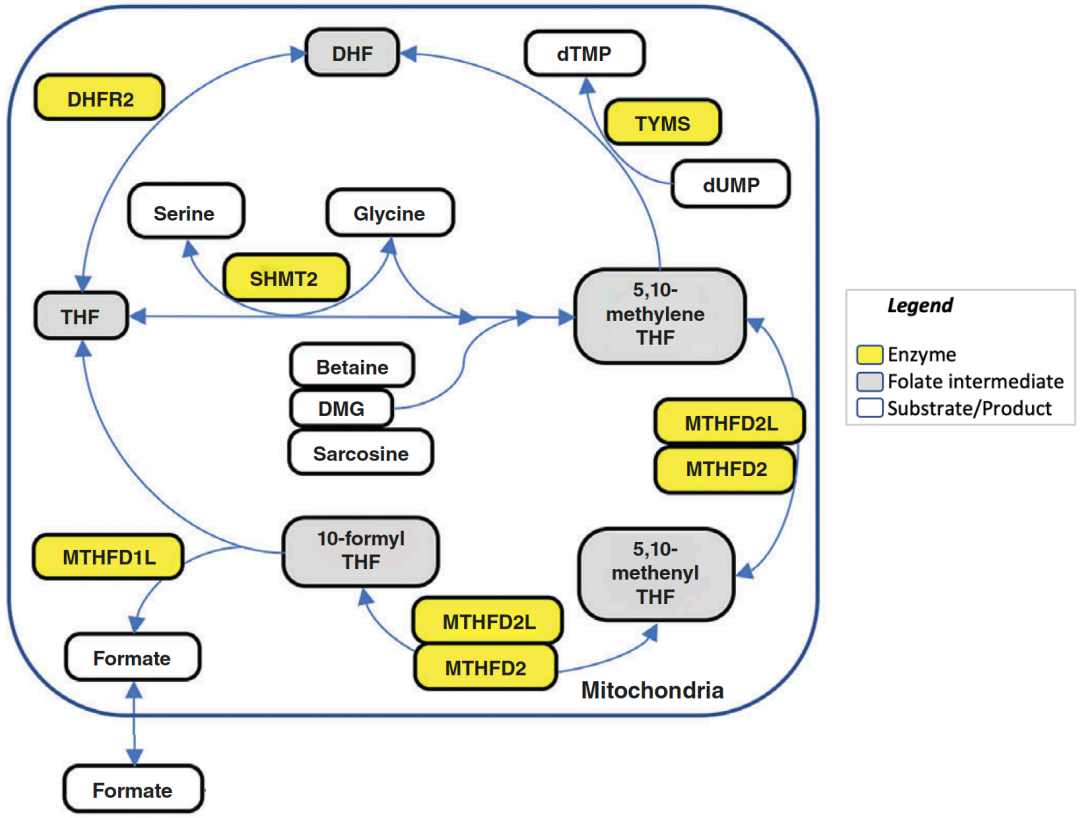


Figure 1.4 Mitochondrial One-Carbon Metabolism. The donation of 1C group to THF, from SHMT2, is followed by the oxidation of methylene-THF to formyl-THF via MTHFD2/L. Finally, MTHFD1L converts formyl-THF to formate, which is transported to the cytoplasm. Alternatively, methylene-THF can be used in the Thymidylate de novo synthesis pathway, with TYMS converting dUMP to dTMP and releasing DHF, which is recycled back to THF by DHFR2. Image adapted from Lan et al., 2018.

In addition to the production of formate and glycine, the mitochondrial OCM was shown to accommodate *de novo* thymidylate synthesis. The enzyme TYMS (Thymidylate synthase) unidirectionally transfers the 1C group from methylene-THF to dUMP, forming dTMP, which is directly used in mitochondrial DNA replication and repair (Alonzo et al., 2018; Anderson et al., 2011). TYMS is also the only enzyme able to change the chemical structure of the principal carrier THF; in fact, the donation of 1C to dUMP is accompanied by the loss of two electrons, with the formation of DHF (Ducker & Rabinowitz, 2017). The discovery of a mitochondrial Dihydrofolate reductase isozyme, DHFR2, able to reconvert DHF to THF, indicates the presence of a complete mitochondrial dTMP de novo synthesis pathway composed of SHMT2, TYMS and DHFR2 (Anderson et al., 2011).

Finally, the production of fMet-tRNA^{Met} is folate-dependent. The mitochondrial enzyme MTFMT (methionyl-tRNA formyltransferase) utilises the OCM product formyl-THF to create the modified tRNA used to initiate translation in the mitochondria (Dickerman et al., 1967). The supply of specifically activated 1C groups to mitochondrial tRNA links OCM to mitochondrial well-being, as demonstrated by Minton, (2018). Impairment of the mitochondrial branch of OCM led to decreased mitochondrial translation and proteins with defective N-terminal ends (Minton et al., 2018). The correlation between mitochondrial OCM and translation may subsequently lead to folate-mediated modulation of the OXPHOS pathway, as all mitochondria-coded proteins belong to the respiratory chain (Chial & Craig, 2008). Although this hypothesis has yet to be explored, the interdependence of mitochondrial OCM and oxidative phosphorylation has been largely demonstrated (Bao et al., 2016; Fiddler et al., 2021, 2021, 2022; Xiu & Field, 2020).

1.3.3 One-Carbon Metabolism in Cytoplasm

The main form of circulating folate is 5-methyl-THF monoglutamated. Cellular uptake relies on several types of transporters, such as RFC (reduced folate carrier), FR (folate receptor) and PCFT (proton-coupled folate transporter). Each transporter is differentially distributed among tissues. Additionally, they present a distinct affinity for folate and operate at a specific pH (Shulpekova et al., 2021; Zhao et al., 2011). Once transported in the intracellular space, the folate molecules are quickly loaded with a variable number of glutamate residues (generally two to nine) by the FPGS enzyme (Lawrence et al., 2014; Shane, 1989; Wagner, 2001). Folate polyglutamates are, in turn, less prone to bind to transporters and more driven toward the OCM enzymes (D. W. Kim et al., 1996). Polyglutamylation represents an evolutionary adaptation as demonstrated by the higher number of glutamate residues in eukaryotes compared to bacteria (Leung et al., 2013; Lu et al., 2007).

One branch of the one-carbon metabolism occurs in the cytoplasm, functioning as a fundamental metabolic network for purines and thymidylate synthesis and remethylation of homocysteine (Fox & Stover, 2008). The virtual journey of 1C groups across the cytosolic OCM starts with the efflux of mitochondrial formate, which

constitutes the most oxidised 1C group. Following the reductive flux typical of the cytoplasmic pathway, three of the main folate intermediates are encountered: formyl-THF, methylene-THF and methyl-THF, respectively involved in purine synthesis, thymidylate synthesis and remethylation of homocysteine. However, these three pathways compete for the limiting pool of folate intermediates. Particularly, methylene-THF is contended between dTMP synthesis and methylation cycle (Stover & Field, 2011).

1.3.3.1 Folate-dependent de novo purine biosynthesis

Purine de novo biosynthesis is a 10-step pathway, activated exclusively when the salvage pathway is insufficient to meet the nucleotide cellular demand (Hartman & Buchanan, 1959; Yamaoka et al., 1997). The purine de novo pathway is folate-dependent, as formyl-THF serves as the entry point by supplying the C2 and C8 carbons of the purine ring, mainly via the action of GAR (glycinamide ribonucleotide transformylase) and AICAR (aminoimidazolecarboxamide ribonucleotide transformylase) (Baggott & Tamura, 2015; Greasley et al., 2001; Manieri et al., 2007; Vergis et al., 2001).

The salvage pathway is the preferential mode to obtain purines in a single step, where hypoxanthine (HT) is linked to phosphoribosyl pyrophosphate (PRPP), generating inosine-5-monophosphate, a purine precursor (Torres, 2013). When the salvage pathway is active, a feedback mechanism ensures the inhibition of de novo purine biosynthesis (Henderson & Khoo, 1965; Yamaoka et al., 2001).

An additional requirement that limits the use of the de novo pathway is the formation of the purinosome (Pedley & Benkovic, 2017). The six enzymes involved in the nucleotide biosynthesis assemble in a metabolon and rapidly dissociate when exposed to purines (An et al., 2008). Although the enzymes are present in the cytoplasm, they are not functional unless arranged in a multienzyme complex. This mechanism enables the temporal compartmentalisation of a metabolic pathway within the same cellular compartment (Lan et al., 2018).

Mitochondrial formate is transferred to THF by MTHFD1, thus producing formyl-THF (Fig. 1.5). It was hypothesised that MTHFD1 could deliver formyl-THF to the purinosome, even though no evidence has been found yet. Contrarily, MTHFS (methenyltetrahydrofolate synthetase) was shown to colocalise with the purinosome

(Field et al., 2011). Beyond its main function of converting 5-formyl-THF (an alternative form of formyl-THF) to methenyl-THF, MTHFS can tightly bind 10-formyl-THF, which acts as an inhibitor (Field et al., 2006). While the 5-formyl-THF does not serve as a coenzyme but as storage of formyl-THF (Stover & Schirch, 1993), the association with 10-formyl-THF ensures that the formyl reserve (5-formyl-THF) is mobilised only in the event of 10-formyl-THF depletion, that is during de novo purine biosynthesis. This regulatory mechanism enables the delivery of 10-formyl-THF to the purinosome only when necessary (Field et al., 2009).

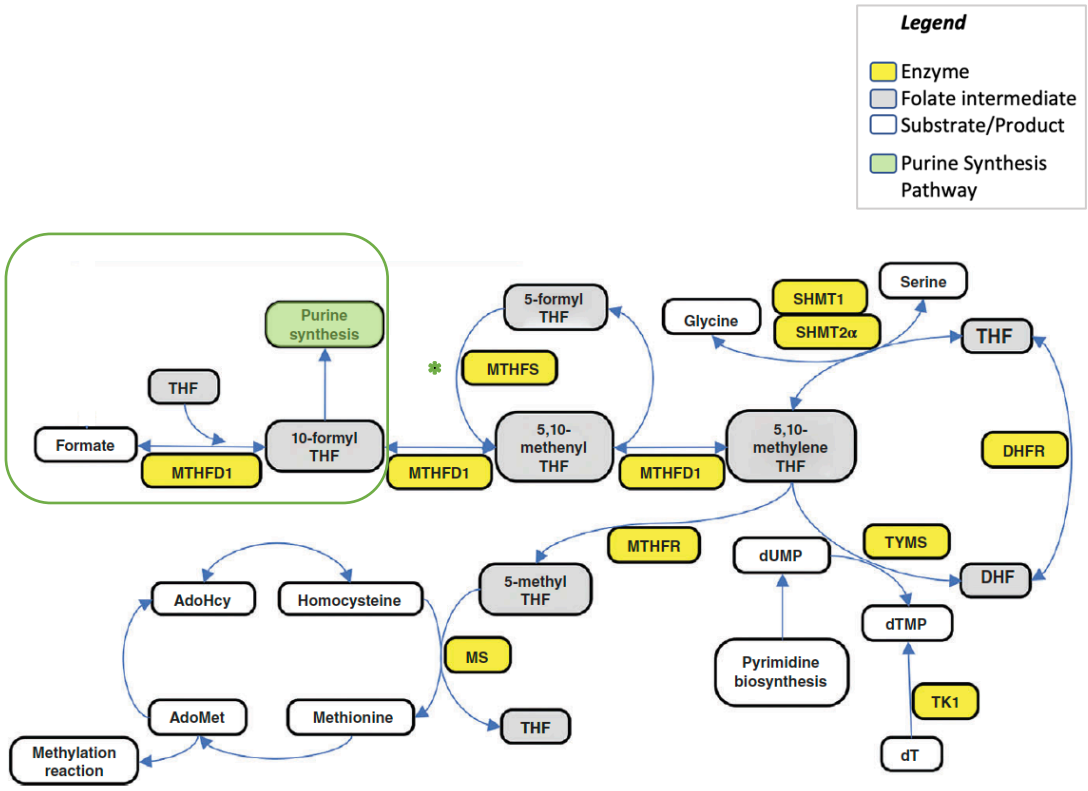


Figure 1.5 Cytoplasmic OCM, Purine de novo synthesis. Mitochondrial-derived formate is transferred to a THF molecule by MTHFD1 producing Formyl-THF, which may be used in purine de novo synthesis. Image adapted from Lan et al., 2018.

1.3.3.2 De novo thymidylate synthesis

MTHFD1 is a trifunctional enzyme. In addition to the C-terminal domain, which can synthesise formyl-THF (discussed in the previous section), it possesses cyclohydrolase and dehydrogenase domains at the N-terminal. The cyclohydrolase activity is required

in the formation of methenyl-THF (from formyl-THF), which is further transformed into methylene-THF by dehydrogenation (Hum et al., 1988). Alternatively to MTHFD1, SHMT1 can produce the same intermediate by transferring 1C groups from cytosolic serine to THF, with the production of methylene-THF and glycine (Fox & Stover, 2008). This reaction is analogous to that performed by the mitochondrial isozyme SHMT2, although SHMT1 cannot compensate for the lack of SHMT2 (Chasin et al., 1974). Methylene-THF is the key substrate for the interconversion of dUMP (uridylate monophosphate) to dTMP (thymidylate monophosphate) by TYMS (thymidylate synthase). TYMS is a unique enzyme due to its ability to transfer the activated carbon alongside two electrons resulting in the oxidation of the substituted THF to DHF. The enzyme DHFR reconstitutes THF by reducing DHF in an NADPH-dependent reaction (Fox & Stover, 2008). Figure 1.6 illustrates the pathway.

Isotope tracer studies have shown the preferential incorporation of SHMT1-derived methylene-THF in thymidylate molecules (Herbig et al., 2002), therefore concluding that the de novo thymidylate synthesis is mainly sustained by the cooperating activity of SHMT1, TYMS and DHFR. Thymidylate biosynthesis also relies on a salvage pathway, where Thymidine Kinases (TKs) add a 5'-phosphate to thymidine restoring the dTMP pools. However, the salvage pathway is not sufficient to address the cell dTMP requirements, especially in preparation for cell division (Field et al., 2016). It has been demonstrated that pyrimidine biosynthesis is upregulated ahead of S-phase and reverted to standard expression immediately after (Ash et al., 1995; Ayusawa et al., 1986). Impairments in the thymidylate pools cause the misincorporation of deoxyuridine nucleotides in the DNA due to the incapability of DNA polymerase to discriminate between dUTP and dTTP. This causes chromosome instability and eventual breakage with consequences for cancer development and neuronal damage (Blount et al., 1997; Hazra et al., 2010; Hori et al., 1984).

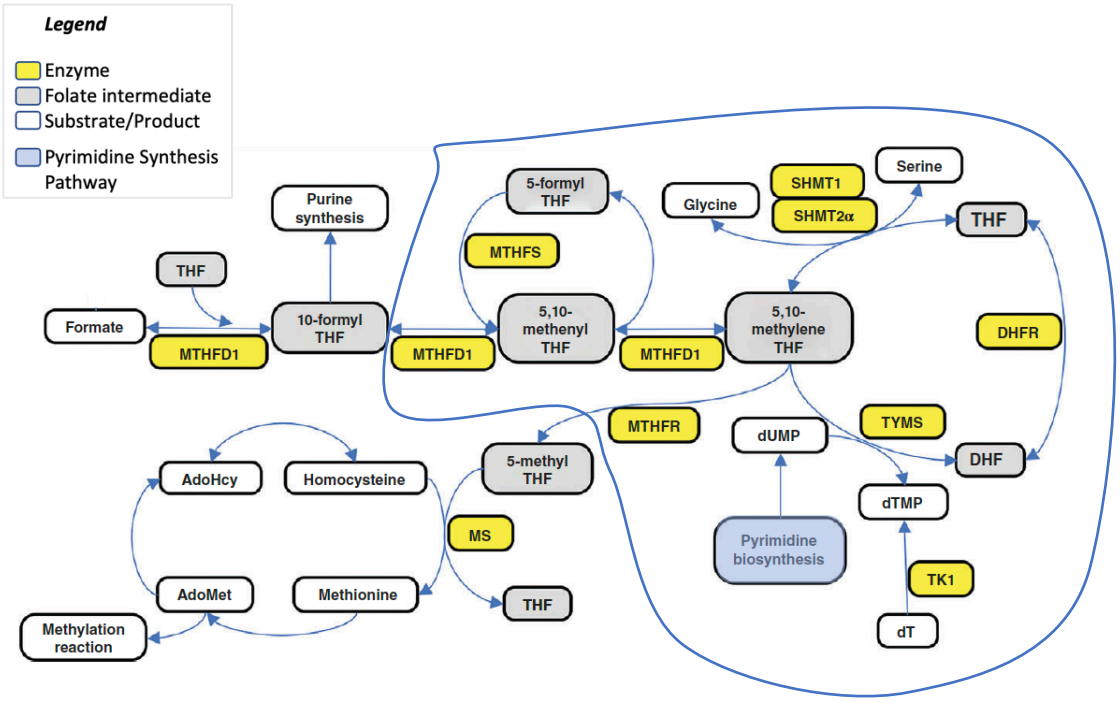


Figure 1.6 Cytoplasmic OCM, Thymidylate de novo synthesis. Methylene-THF, derived from either MTHFD1 or SHMT1/2a is used as a carbon donor in the dTMP synthesis reaction catalysed by TYMS. The discarded DHF is recycled into THF by DHFR. Image adapted from Lan et al., 2018.

1.3.3.2.1 Nuclear de novo thymidylate synthesis

During the S-phase of the cell cycle, SHMT1, TYMS and DHFR undergo nuclear translocation upon SUMOylation (Anderson et al., 2007; Woeller et al., 2007). The Small Ubiquitin-like Modifier is covalently attached to each enzyme, inducing their transportation to the nuclear compartment, where they assemble into a complex at the replication sites (Anderson et al., 2012). SHMT1 was shown to localise in the cytoplasm during the G1-phase and also in the nuclear periphery and nucleus in the S, G2, and M phases of the cell cycle. Evidence suggests that SUMOylation occurs at the nuclear pore leading to nuclear import (Woeller et al., 2007). Residual serine hydroxymethyltransferase activity was observed in SHMT1$^{-/-}$ cells due to the expression of a secondary cytoplasmic/nuclear enzyme, SHMT2α, encoded from *SHMT2* via alternative splicing (Anderson & Stover, 2009). Similar mechanisms are activated for nuclear import of TYMS and DHFR, which were found to be modified by SUMO1 in vitro (Anderson et al., 2007, 2012) but also possess SUMO2 potential target sites (Hendriks et

al., 2015; Tammsalu et al., 2014). MTHFD1 was also found translocating to the nucleus, maximising its presence during phase S. The nuclear MTHFD1 increment responds to fluctuations in folate levels. In the event of folate depletion, the nuclear MTHFD1 translocation happens at the expense of the cytoplasmic remethylation pathway due to the reduction of cytoplasmic methylene-THF pools (Field et al., 2014, 2015; Kamynina et al., 2017). Nuclear OCM is shown in Figure 1.7.

Once translocated, SHMT1, TYMS, and DHFR assemble in a dTMP-synthesis complex (dTMP-SC) which anchors the nuclear lamina through SHMT1 (Anderson et al., 2012). Evidence suggests that the dTMP-SC assembles with DNA polymerase and other enzymes of the DNA replication (Stover & Field, 2011). Together, they form a mega-complex called Replitase (Murthy & Reddy, 2006). The active participation of dTMP-SC at the replication fork was hypothesised to directly provide thymidylate units and lower the risk of uracil misincorporation (Stover & Field, 2011).

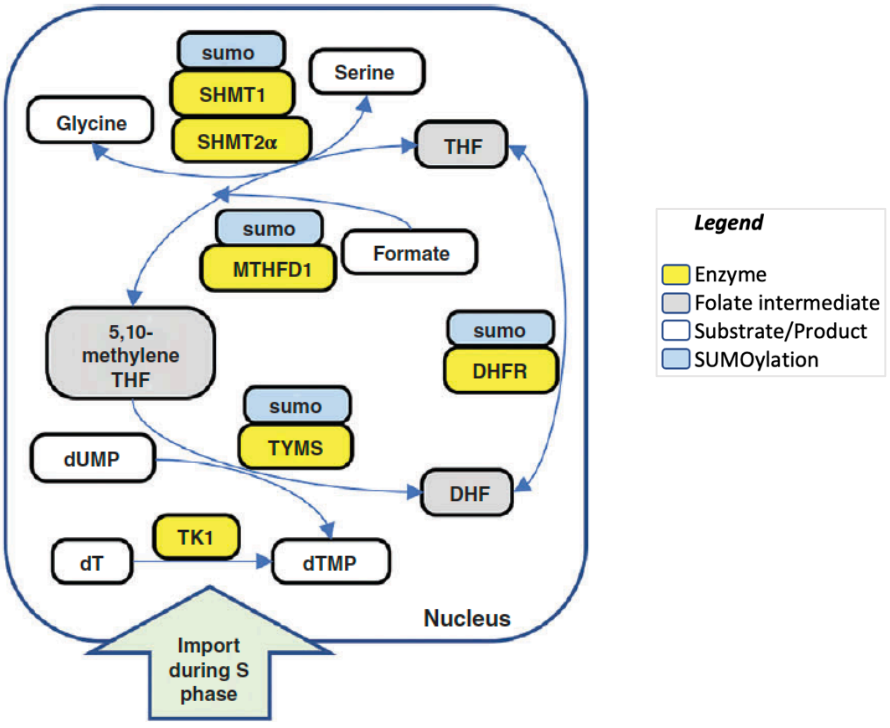


Figure 1.7 Thymidylate de novo synthesis in nucleus. The main components of the dTMP-SC (TYMS, DHFR and SHMT1/2a) and MTHFDL1 translocate to the nucleus before DNA replication. The nuclear import is dependent on SUMOylation of the enzymes. Image adapted from Lan et al., 2018.

1.3.3.3 Homocysteine remethylation pathway

MTHFR (5,10-MethyleneTHF reductase) catalyses the reduction of methylene-THF to methyl-THF, competing with TYMS for the aldehyde substrate (Selhub, 1999). The MTHFR reaction is irreversible; thus methyl-THF becomes an obligate substrate for the remethylation of homocysteine (Appling, 1991; Sumner & Matthews, 1992; Trimmer et al., 2001). Methionine synthase (MTR) catalyses said reaction with the contribution of vitamin B12. Methionine, upon conversion to SAM (S-adenosyl-methionine), serves as main donor of methyl groups to all macromolecules (DNA, RNA, lipid, protein) (Fig. 1.8) (Mudd et al., 2007; Sanderson et al., 2019). An accumulation of methyl-THF can cause purine and thymidylate synthesis impairments. Such an event occurs in the case of severe B12 deficiencies (Palmer et al., 2017; Wang et al., 2020). SAM regulates MTHFR in a negative feedback mechanism (Jencks & Mathews, 1987). Disruptions of the methionine cycle lower SAM levels with consequent methyl-THF accumulation.

MTHFR is expressed in all tissue types, with a peak in testis where high levels of DNA methylation are required (Gaughan et al., 2000). Deficiencies of the MTHFR activity are generally associated with a common polymorphism, 677 C>T, which causes the substitution of Alanine 222 with Valine in the enzyme's catalytic site. This mutation leads to an unstable binding with the cofactor FAD (Pejchal et al., 2006). The MTHFR 677C>T mutation is associated with mild hyperhomocysteinemia (Jacques et al., 1996) and DNA hypomethylation (Tsang et al., 2015) and with a higher risk of cardiovascular diseases (Cortese & Motti, 2001) and neural tube defects (Tabatabaei et al., 2022). It was recently demonstrated that riboflavin (FAD) is able to modulate the MTHFR 677C>T phenotype in homozygous adults (MTHFR 677 TT), with subsequential effects on DNA methylation patterns (Amenyah et al., 2020). Specifically, alterations of DNA methylation of hypertension-related genes in response to riboflavin supplementation provide additional information on the direct link between OCM and epigenetic phenomena, such as DNA methylation (Amenyah et al., 2021). In addition, the Stover group proposed a stochastic model where MTHFR 677C>T polymorphism is causative for changed SHMT activity and consequent impaired de novo thymidylate synthesis (Misselbeck et al., 2017).

Through the activity of MTHFR, the OCM is directly linked with the methylation cycle and relative cellular processes, such as the epigenetic regulation of gene expression (Froese et al., 2019; Raghubeer & Matsha, 2021; Van Winkle & Ryznar, 2019). OCM disturbances, either caused by enzymatic impairment (Yadav et al., 2021) or nutritional state (Crider et al., 2012; Kruman & Fowler, 2014) are linked to altered epigenetic patterns and disease (Clare et al., 2019; Ducker & Rabinowitz, 2017; Liew & Gupta, 2015).

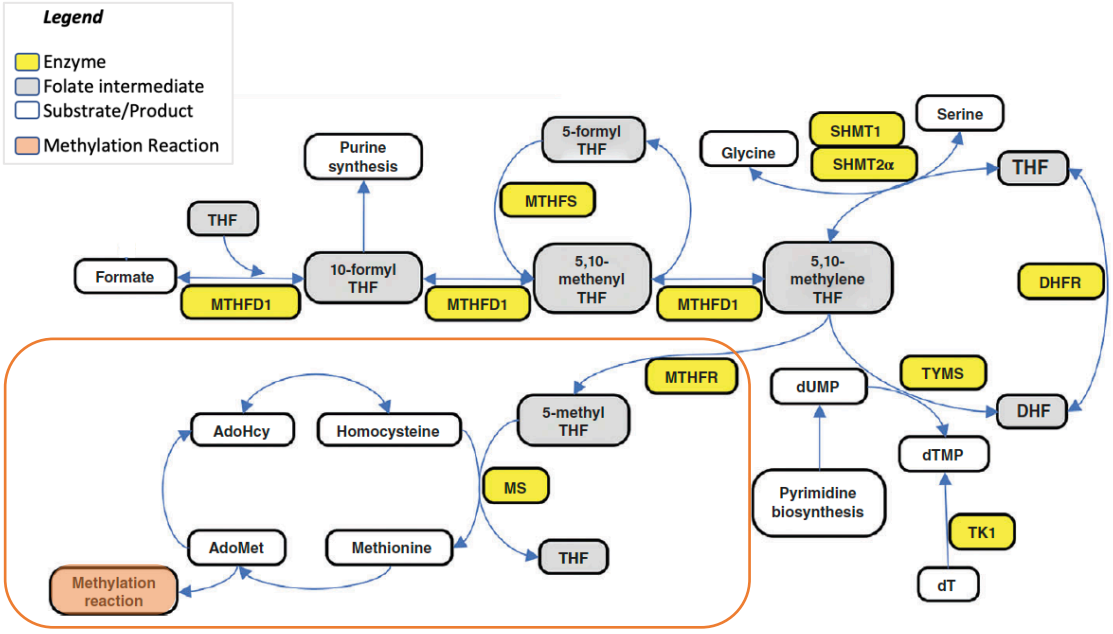


Figure 1.8 Cytoplasmic OCM, Remethylation of homocysteine and Methionine cycle. The thymidylate synthesis pathway and methylation cycle compete for the substrate methylene-THF. When MTHFR use it, it irreversibly forms Methyl-THF, which serves to remethylate homocysteine and produce methionine, thus feeding the methylation cycle. Image adapted from Lan et al., 2018.

1.4 Dihydrofolate Reductase (DHFR)

Dietary folate from non-fortified foods is polyglutamated and chemically reduced, while folic acid (from fortified foods and supplements) comes in a fully oxidised monoglutamated form (Lucock, 2000; Scaglione & Panzavolta, 2014; Shane, 2001). Although the natural form of folate has a reasonable degree of intestinal absorbance, the synthetic molecule reaches a major absorption capacity (Bailey et al., 1984, 1988;

Caudill, 2010; Gregory et al., 1991; Sauberlich et al., 1987). The limiting factor for the uptake of natural folate can be attributable to the process of de-conjugation (Melse-Boonstra et al., 2002). At the luminal brush of the small intestine, the enzyme gamma-glutamyl hydrolase eliminates the polyglutamine tail of folates to enable their transportation into the enterocytes (Shane, 1989). Once absorbed, the natural forms of folate are readily transformed into methyl-THF, which constitutes the main form of circulating folate. On the other side, even though monoglutamated folic acid can readily cross the enterocytes membrane, the cellular capacity to convert it to THF is not comparable to that expressed for the natural folate (Menezo et al., 2022). Therefore, the majority of folic acid arrives intact to the portal vein. The liver is the principal organ for the reduction of folic acid to THF through the action of the dihydrofolate reductase enzyme (DHFR) (Patanwala et al., 2014). However, it is important to note that the hepatic DHFR activity is slow and variable among human samples (Bailey & Ayling, 2009). This phenomenon could explain the presence of unmetabolised folic acid in plasma and urine. Although folic acid supplementation/fortification was demonstrated to be protective against cancer and useful for the prevention of neural tube defects (Czeizel & Dudás, 1992; 'Prevention of Neural Tube Defects', 1991), its activation is limited by liver saturation and can lead to high circulating levels of unmetabolised folic acid (UMFA) (Bailey & Ayling, 2009). Public and scientific debate has focused on the potential side effects of UMFA, yet rigorous investigation is still required to assess the causal relation between UMFA and adverse health outcomes (Field & Stover, 2018; Maruvada et al., 2020).

DHFR plays a crucial role in the bioavailability of folic acid. It is converted into THF in a double reaction that forms DHF as an intermediate product. DHFR exerts this function primarily in the liver (Wright et al., 2007). However, DHFR is a ubiquitous enzyme, found in all tissue types and cell compartments (Human Protein Atlas https://www.proteinatlas.org/ENSG00000228716-DHFR/tissue) (Sjöstedt et al., 2020). Its primary role is to convert DHF - discarded upon dTMP de novo synthesis – to THF, thus recycling the main 1C group carrier of the OCM. As discussed in the previous sections, THF carries 1C moieties at different oxidation states, thus providing several activated groups for anabolic pathways such as purine and thymidylate synthesis and

methylation reactions (Morellato et al., 2021). Due to its crucial role in normal cell physiology, inactivation of DHFR results in depletion of folate pools, impaired DNA synthesis and eventually cell death (Hsieh et al., 2013). DHFR function is essential in cell metabolism; therefore, it is maintained throughout evolution, as confirmed by the presence of Dihydrofolate Reductase enzymes from bacteria to humans. Also, due to its relevance in cell proliferation, DHFR has been largely characterised and examined as a pharmacological target (Abali et al., 2008).

1.4.1 DHFR structure and kinetics

The human DHFR enzyme consists of 187 amino acids organised into seven parallel β-strands and a C-terminal antiparallel strand. Five α-helices interlayer against the β-sheet. In addition, a polyproline-like helix, eight tight turns, and several loops contribute to the final structure of the enzyme (Blakley, 2006; Carugo & Argos, 1997; Davies et al., 1990; Oefner et al., 1988), as shown in Figure 1.9.

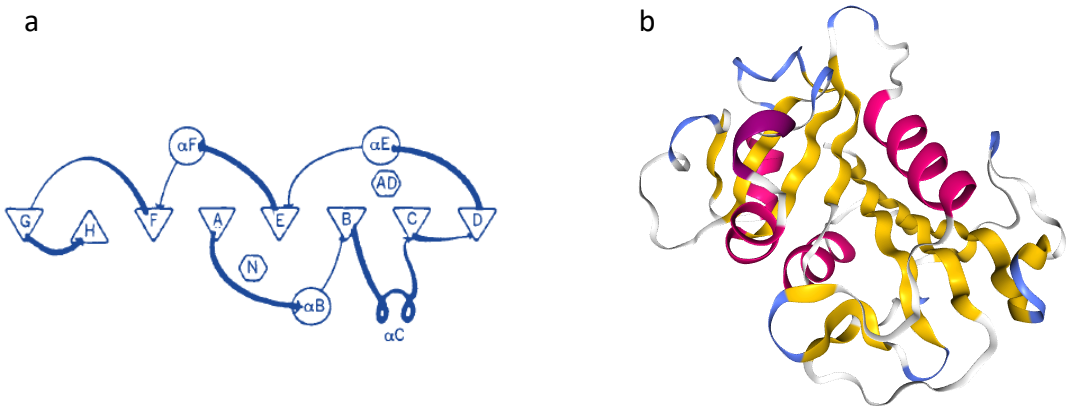


Figure 1.9 DHFR secondary and tertiary structure. a) Secondary structure diagram. β-strands are indicated by triangles, and α-helices by circles. α-C helix is indicated as a two-turn helix: its axis is perpendicular to the β-sheet, contrarily to all other helices, with axes parallel to the central β-sheet. The N and AD hexagons show the contact areas of the nicotinamide and adenine mononucleotide moieties of NADPH, respectively. Image adapted from Blakeley, 1995. b) Tertiary structure diagram. β-strands in yellow, α-helices in pink, β-turns in blue and coils in white. Image taken from PBD.

The reductase reaction catalysed by DHFR is shown in Figure 1.10 (Fischer et al., 2010) and can be summarised as follows:

7,8-Dihydrofolate + NADPH + H^+ ⇄ 5,6,7,8-tetrahydrofolate + $NADP^+$

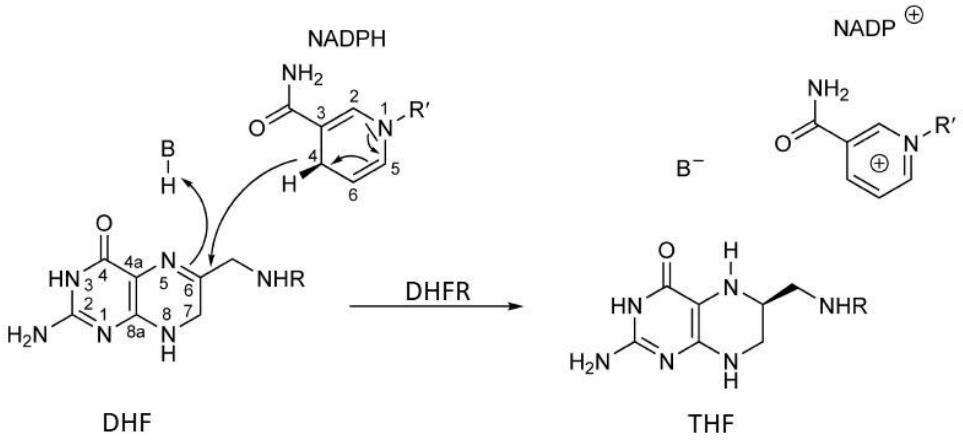


Figure 1.10 DHFR catalytic reaction. DHF is reduced to THF via NADPH oxidation. Adapted from Fischer et al., 2010

The kinetics of the reaction can be divided into five steps, comprising the formation of several intermediate states. The holoenzyme E:NADPH sees DHFR binding the coenzyme NADPH; upon interaction with DHF, DHFR forms the so-called Michaelis complex, E:NADPH:DHF, followed by three product complexes E:NADP+:THF, E:THF and E:NADPH:THF, which is formed before THF can be released. In fact, after NADPH oxidation, THF cannot be discharged until a new NADPH molecule refills the active site. Molecular motions determine the interconversion of the enzyme into different catalytic intermediates, which are the basis of the reaction kinetics (Abali et al., 2008). These motions are controlled by residues of the active site and remote residues, which contribute to local and global effects. It has been proven that these amino acids are evolutionarily conserved among all phylogenetic groups, confirming the influence of the structure on enzymatic dynamics, kinetics and catalysis (Adesina et al., 2021; Francis et al., 2013; Francis & Kohen, 2014).

DHFR binding and catalytic sites of *E. coli* and *H. sapiens* are sufficiently diverged (proteins share 26 % homology) with different catalytic efficiency. Comparative analysis of DHFR sequence from 233 species identified three mutation sites involved in the divergence of DHFR functions throughout evolution. The change of these sites on E.coli (with those of *H.sapiens*) was enough for the bacterial enzyme to acquire the human DHFR catalytic properties. This experiment suggests that the conformational motions

are highly conserved and common between pro- and eukaryotes, despite sequence divergence (Liu et al., 2013).

1.4.2 DHFR mechanisms of regulation

DHFR expression is cell cycle-dependent, with a maximal transcriptional rate at the turn of G1 and S phase, when a broad pool of precursors is required for DNA duplication. DHFR contains a TATA-less promoter regulated by E2F and Sp1 factors when binding to their respective target sequences, 5'-TTTCGCGCCAAA-3' and 5'-GGGGCGGGGC-3' (Abali et al., 2008; Slansky & Farnham, 1996). Evidence suggests that Sp1 is constantly bound to the DHFR promoter while E2F is a transcriptional regulator that can activate or repress transcription according to the interaction status (Good et al., 1996; Schilling & Farnham, 1994).

Another group of regulatory factors was found to regulate DHFR transcription. The Rb (retinoblastoma) factors Rb/p105, p107 and Rb2/p130 (collectively called 'pocket proteins') are responsible for the cell cycle checkpoint at G1. Rb factors generally repress gene transcription by binding E2F (Giacinti & Giordano, 2006). The Rb proteins change their phosphorylation status according to the cell cycle. Hyperphosphorylation of the pocket proteins induces inactivation and dissociation from E2F, which switches from repressive to activating form. As a result, *DHFR* expression increases tenfold at the entrance of the S-phase (Abali et al., 2008). *In vitro* studies showed that DHFR E2F sites act primarily as repressive elements in G_0 and G_1, while Sp1 can induce transcription upon serum supplementation in quiescent cells. The S-phase activation of Sp1 in DHFR can be mediated by the cell-cycle responsive Rb proteins (Good et al., 1996; Jensen et al., 1997).

The CpG islands along DHFR are unmethylated. The entirety of the gene is associated with nucleosomes except for the promoter region, which is instead covered with a multi-enzymatic complex (Shimada et al., 1986). The DHFR promoter is actively involved in *DHFR* epigenetic regulation. The Rb proteins p107 and p130 can recruit the histone methyltransferase SUV39H1 at the E2F site and silence DHFR during the G_0 and early G_1 phases of the cell cycle. In particular, Lysine 9 of Histone 3 changed from a

hypermethylated state (G_0) to a hyperacetylated state during the G_1/S transition, with a consequent transcriptional activation (Nicolas et al., 2003).

Promoters are considered modulatory DNA structures, which, by assembling with cis-regulatory elements, ensure flexibility in gene expression. The presence of multiple promoters was shown to be frequent in multicellular organisms. The use of alternative promoters can influence transcription itself and the turnover and translation efficiency of the mRNA (Ayoubi & Van De Yen, 1996). *DHFR* harbours a secondary promoter situated ~400 bp upstream of the main transcription start site. It produces an alternative transcript (long non-coding RNA) able to inhibit transcription from the major promoter. Blume et al. (2003) demonstrated that DHFR lncRNA could sequester Sp1 and Sp3 in vitro, thus altering the formation of the pre-initiation complex (PIC). PIC is made of general transcription factors bound to the core promoter; its formation is necessary for RNA Polymerase II to start the transcription (Petrenko et al., 2019).

Further analysis showed the lncRNA stably annealed with the major promoter in a DNA:RNA triplex. This association resulted in the direct interaction of the lncRNA with the transcription factor IIB and the subsequent disassembly of the preinitiation complex from the main promoter (Fig. 1.11). This type of transcriptional repression was demonstrated in quiescent cells (Martianov et al., 2007).

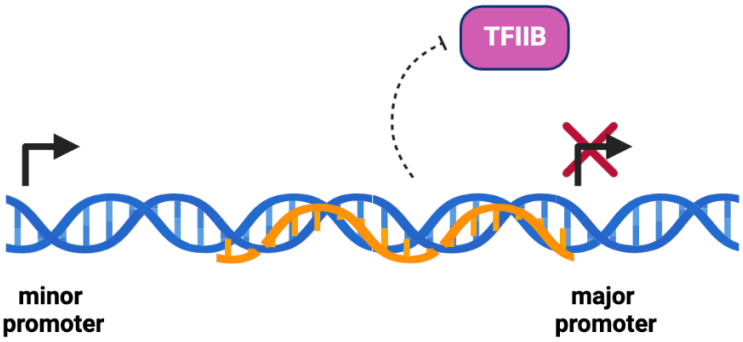


Figure 1.11 DHFR lncRNA-mediated transcriptional regulation. Transcription from the DHFR minor promoter produces a long non-coding RNA, which stably assemble with DNA. The formation of a DNA:RNA triplex leads to the dissociation of the pre-initiation complex at the major promoter due to TFIIB sequestration. Image adapted from Angrand et al., 2015.

DHFR regulation is also implemented by an auto-regulatory translational mechanism, by which DHFR binds its own mRNA, inhibiting translation. Methotrexate (MTX) is a common chemotherapeutic drug which acts as a folate antagonist (Rajagopalan et al., 2002; Visentin et al., 2012). Upon MTX treatment, cells respond with an initial increase in DHFR protein levels (Bertino et al., 1962, 1963) despite the unaltered level of RNA (Cowan et al., 1986). This translational induction by MTX was further investigated, leading to the discovery of a translational suppression mechanism acted by DHFR on its own RNA. An area of ~100 bp, containing two stem-loops, was identified on DHFR mRNA as a contact point for DHFR (Ercikan et al., 1993). Site-directed mutagenesis studies identified amino acid residues involved in RNA recognition. Cys6 residue is crucial, as its substitution with either Ala or Ser completely disrupted DHFR's ability to bind its cognate mRNA (Tai et al., 2002). In addition, amino acids Glu30, Leu22, and Ser118 (all belonging to the NADPH binding site) were involved in the MTX-mediated upregulation mechanism. A model of autoregulation was proposed, where DHFR exists in an equilibrium of two conformers binding either NADPH or DHFR mRNA. When DHF or MTX associates with DHFR:NADPH, the equilibrium moves towards the catalytic form; otherwise, when the substrate binds to DHFR:mRNA, the mRNA is released and translated with a consecutive increase in the protein level (Fig. 1.12) (Abali et al., 2008).

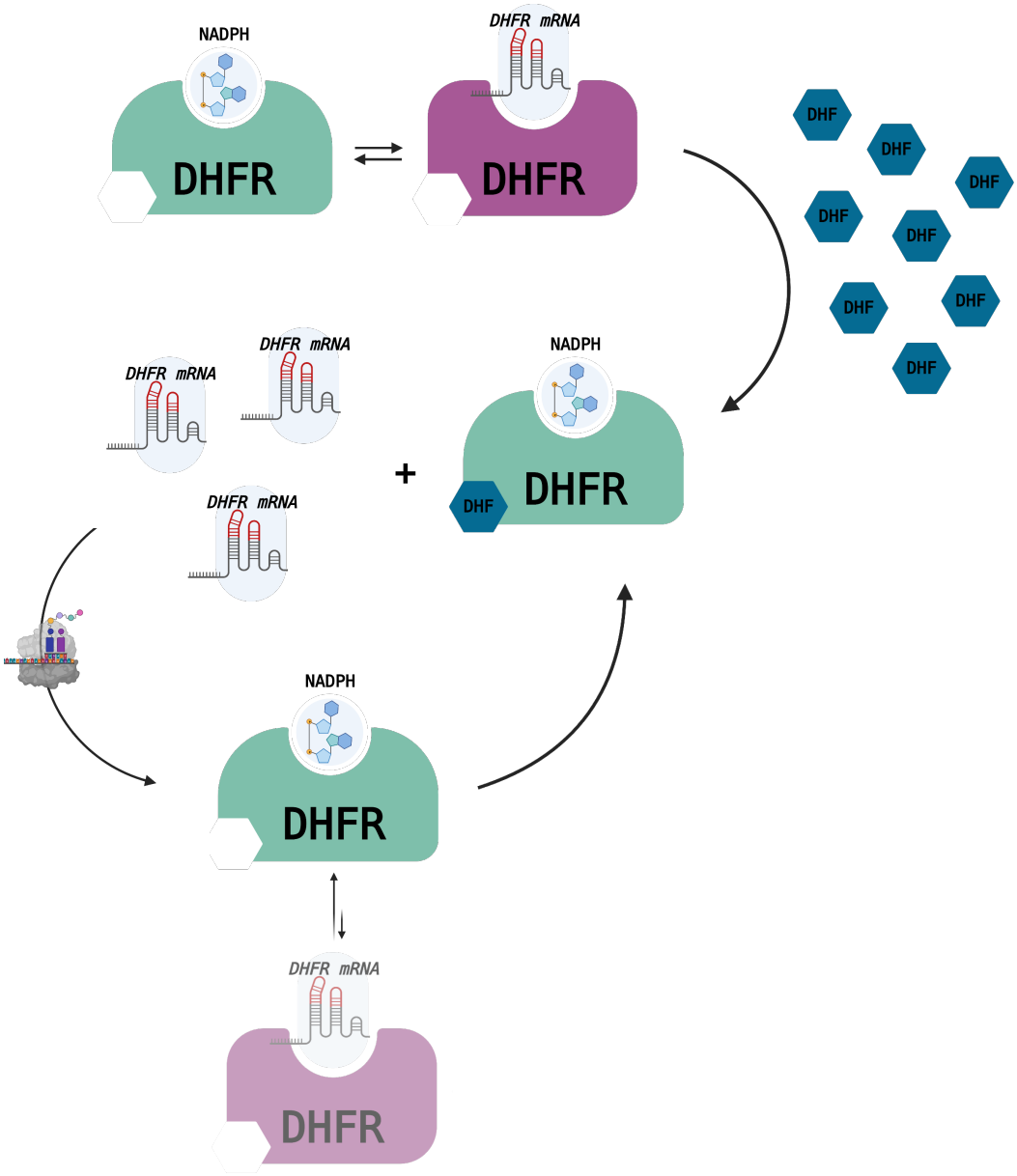


Figure 1.12 DHFR auto-regulation model. DHFR exists in two conformers, in association with NADPH or mRNA. DHF binding induces the catalysis in the E:NADPH:DHF form; otherwise, DHF leads to mRNA dissociation and relative translation. As a result, the DHFR protein levels rise. Similarly, MTX induces a DHFR upregulation without a correspondent increase in THF (product of the DHFR reaction). Image created with BioRender.com; inspired by Abali et al., 2008.

1.4.3 Human Dihydrofolate Reductase gene family

The *DHFR* gene is located on chromosome 5q11.2-q13.3. It is a 30 kb long gene consisting of six exons with five alternating introns. The primary processed mRNA is 3800 bp long, but two additional transcripts can be produced as a result of the use of different polyadenylation signals on exon 6. The 564 bp-long open reading frame encodes for a 21 KDa protein (187 aa) (Chen et al., 1984; Maurer et al., 1984).

Several pseudogenes have been identified as part of the DHFR gene family. It is a widely held view that they originate from the genomic reintegration of DHFR mRNA since they are intronless and interspersed into the genome. DHFRP1 and DHFRP2 are located at chromosomes 18 and 6, respectively. They possess poly-A tail fragments, which confirm their origin as mRNA (Anagnou et al., 1984). DHFRP3 harbours just half of the coding sequence of the original DHFR gene, and it is located at chromosome 2 (Shimada et al., 1984). Finally, DHFR2 (previously DHFRP4; DHFRL1), located at chromosome 3, was recently re-annotated as a functional retrogene, thanks to the work of the Parle-McDermott and Stover's groups (Anderson et al., 2011; McEntee et al., 2011).

1.5 Dihydrofolate Reductase 2 (DHFR2)

The previously reported DHFR2 pseudogene rose to a new life after McEntee and Anderson proposed a novel function for the *DHFR2* gene, coding a mitochondrial reductase enzyme (Anderson et al., 2011; McEntee et al., 2011).

DHFR2 mRNA was identified in several cell lines, confirming that DHFR2 transcription is active in many tissue types, including differentiating cell lines (*H.sapiens* DHFR2 ensg00000178700 Expression Atlas) (Papatheodorou et al., 2020). Comparison between the amino acid sequence of DHFR and DHFR2 showed 92 % of identity with high conservation of four functional motifs. Residues involved in the creation and maintenance of the active sites are mostly conserved, except for three amino acids, including Arg25 (Fig. 1.13) (McEntee et al., 2011). A Trp occupies the same position in DHFR. This amino acid difference is particularly relevant since the additional NH3 moiety of Arg lodges in the proximity of the DHF and NADPH active sites. In addition, molecular

modelling predicted the decrease of the net electrostatic forces exerted by Arg32 and Gln35 on DHF (Gao et al., 2013). This data might explain the 30 % reduction in the catalytic activity of the recombinant DHFR2 compared to DHFR. In fact, despite having a similar Km value for NADPH, DHFR2 presents a 10-fold higher Km value for DHF (McEntee et al., 2011).

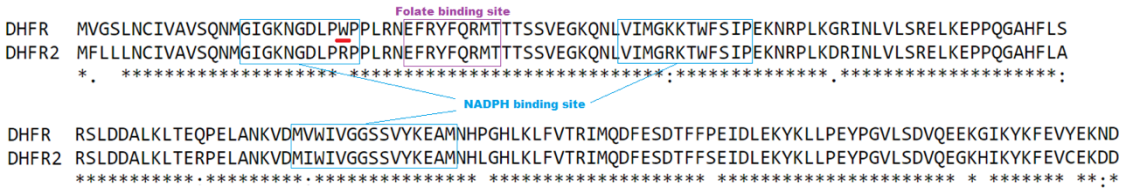


Figure 1.13 Clustal pairwise alignment of DHFR and DHFR2 predicted sequence. The folate (DHF/THF) binding site amino acids are enclosed in a purple box. The NADPH binding site is tridimentionally composed by the amino acids found in the blue boxes. Arg25 substitution is underlined in red.

The recombinant DHFR2-GFP protein was found to localise in mitochondria of unsynchronised HEK293 (McEntee et al., 2011). Anderson validated and extended this data by proposing DHFR2 as a member of the newly discovered mitochondrial *de novo* dTMP synthesis pathway. They demonstrated that intact mitochondria extracted from

CHO and HepG2 cells could synthesise ^{3}H-dTMP starting from unlabelled dUMP, NADPH and [2,3-^{3}H]-L-serine. Upon silencing of DHFR2 RNA, depletion of reductase activity was observed in mitochondria. These results, together with confirmation of a DHFR2 mitochondrial protein via immunoblotting, led to the conclusion that DHFR2 is a mitochondrial isozyme involved in OCM (Anderson et al., 2011)

It is important to note that neither of the above studies investigated the presence of DHFR in mitochondria. This hypothesis was explored by the Parle-McDermott lab, which confirmed the presence of DHFR, as opposed to DHFR2, in HEK293 mitochondria (Bookey et al., unpublished), in a similar fashion to rodent DHFR (Hughes et al., 2015). These novel results may suggest that the reductase activity identified in mitochondria was due to DHFR, not DHFR2. To date, however, there has been little conclusive evidence on the dihydrofolate reductase role within the mitochondrial *de novo*

Thymidylate synthesis, and the known data seem to be contradictory. Therefore, further work is required to gain a better understanding of the functional role of DHFR2.

1.6 Gene-editing techniques

Genome editing and gene editing are two terms commonly referring to a group of techniques for genome engineering. This term refers to the process of introducing site-specific alterations to genomic DNA. From a single nucleotide to entire genes, this relatively recent group of biotechnologies concerns the insertion, deletion and modification of specific areas of the genetic code (Haimovich et al., 2015). To be more precise, the terms gene and genome editing carry slightly different meanings, even though they are often used interchangeably. Gene editing is referred to modifications occurring in a gene. In contrast, genome editing focuses on modifying non-gene areas (to insert new genes, for example) or regulatory regions to modulate the expression of downstream genes (Robb, 2019).

Genome editing techniques are becoming increasingly accurate, efficient and straightforward to use. These characteristics, along with a decrease in processing time and costs, make genome-editing the most versatile tool for creating new disease models (Gupta & Musunuru, 2014), besides generating transgenic animals and performing gene functional analysis (Khalil, 2020).

All genome editing techniques share the use of programmable nucleases, which work in a targeted manner. Four different types of nucleases have been discovered and adapted for editing purposes: Meganucleases, Zinc-finger, TALEN (transcription activator-like effector nucleases), and CRISPR/Cas (clustered regularly interspaced short palindromic repeats / CRISPR-associated protein). They all introduce a double-strand break (DSB) at the recognised genomic site. This is consequently tracked and adjusted by one of the endogenous repair systems of the cell: homology-directed repair (HDR), non-homologous end joining (NHEJ), or other minor systems (Fernández et al., 2017; Khalil, 2020).

Genome editing techniques based on meganucleases, zinc-fingers and TALEN require the engineering of new nucleases for any new target site, as the enzyme specificity for the DNA target resides in their structural motives. Conversely, the CRISPR/Cas enzymes are ribonucleoproteins that acquire their specificity by coupling with a small RNA. Therefore, the protein side of Cas remains unchanged, while different RNAs can be easily designed and assembled to work with it. This feature makes CRISPR/Cas exceptionally specific and easy to use (Jiang et al., 2013; Khalil, 2020; Lino et al., 2018).

The most recent among the genome editing techniques, CRISPR/Cas, has already taken over the other methodologies due to its broad applicability: from agriculture and environment to medicine, spreading from basic research to clinical treatments (Hough & Ajetunmobi, 2017). On the molecular side, CRISPR/Cas can not only produce indels (insertions/deletions) but engineered versions of the Cas enzyme allow more subtle applications, like regulation of gene expression and monitoring of chromatin movements (Rodríguez-Rodríguez et al., 2019).

1.6.1 CRISPR/Cas: from microbial immunity to genome editing technology

CRISPR/Cas editing system represents the most recent significant revolution in molecular biology, as demonstrated by awarding the 2020 chemistry Nobel Prize to Emmanuelle Charpentier and Jennifer Doudna (Ledford & Callaway, 2020). The discovery of a bacterial adaptive immune system that can 'memorise' pieces of genetic material from viruses and later use it to target and destroy the same invading organism was rapidly exploited to develop a simple and accurate genome-editing technique (Lander, 2016).

The first CRISPR locus was accidentally discovered in 1987 (Ishino et al., 1987). Most bacterial and archaeal species possess such a locus, made of short repeats interrupted by unique foreign DNA pieces. Twenty years after that initial discovery, Barrangou and team (2007) demonstrated that the spacer sequences represent an adaptive immune system, active upon re-infection. Short fragments of invading bacteriophages are promptly inserted into the CRISPR locus and, at a later time, used as a defensive mechanism (Barrangou et al., 2007; Wright et al., 2016).

The CRISPR system operates in three phases: adaptation, expression and interference. During adaptation, fragments of foreign DNA are incorporated into the CRISPR array as novel “spacers”, interspaced with the repeats. While the spacers are variable by nature, the repeats are copies of the same sequence. The expression phase sees the *cas* genes transcribed, along with the CRISPR repeat-spacer array, with the following formation of a long pre-crRNA. This precursor RNA is furtherly processed to give mature crRNAs, which, in the interference phase, assemble with the Cas proteins targeting and cleaving exogenous DNA (Barrangou, 2015; Jiang & Doudna, 2017; Rath et al., 2015) (Fig. 1.14).

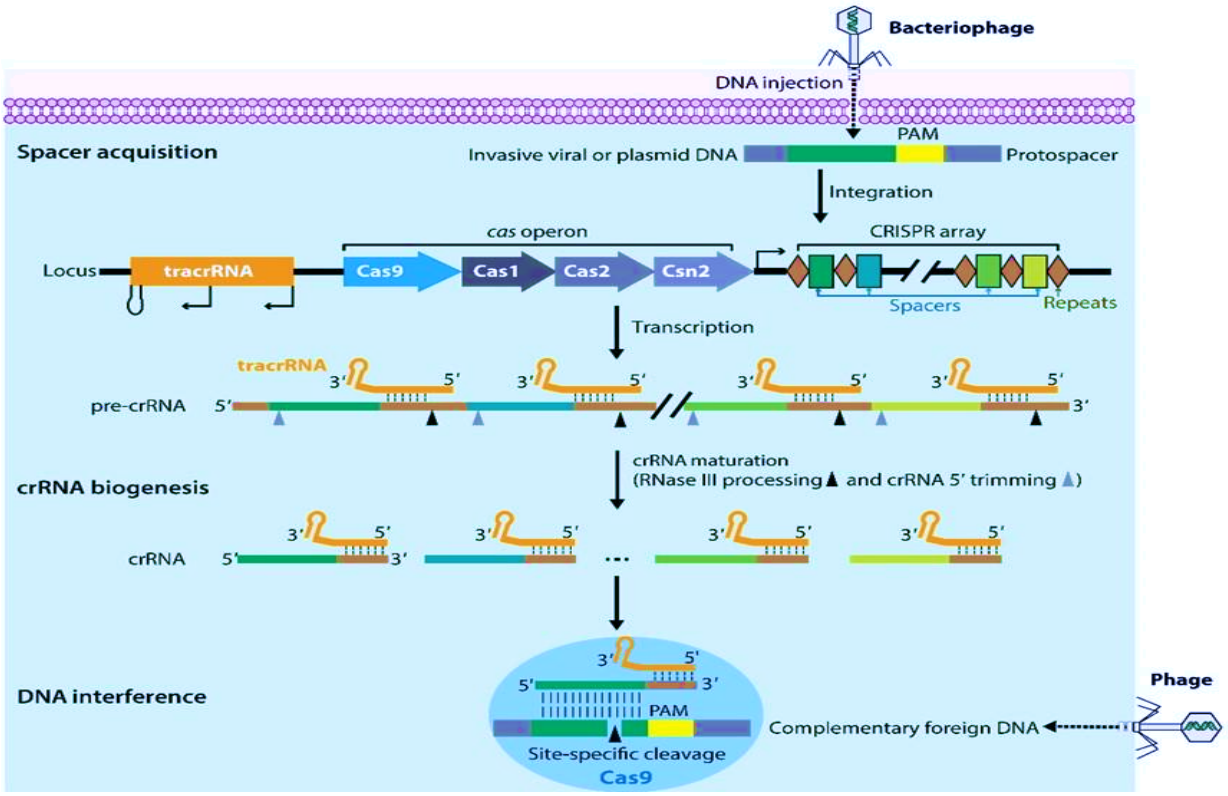


Figure 1.14 CRISPR/Cas-mediated DNA interference in adaptive immunity in Prokaryotes. When foreign DNA is introduced into a Prokaryotic cell, a portion of it is integrated into the CRISPR array as a spacer. The spacers (coloured boxes) are interspaces by repetitive sequences, the repeats (brown diamonds). Preceding the CRISPR array is the Cas operon, carrying all Cas enzymes’ genes and the tracrRNA, encoding a non-coding RNA complementary to the repeats. When the CRISPR locus is expressed, a long pre-crRNA originates from the CRISPR array, which associates with the tracrRNAs. The pre-crRNA goes through maturation thanks to the action of RNase III and other nucleases (from the cas operon), leading to the formation of a single functional RNA duplex (crRNA:tracrRNA). When associated with the tracrRNA, the crRNA binds to the Cas9 enzyme and, via complementarity, recognises the foreign DNA. The Cas9 endonuclease can lead to the cut of foreign DNA, thus actively interfering with a viral infection. Image adapted from Jiang & Doudna, 2017.

The Prokaryotic CRISPR locus harbours a tracrRNA gene, in addition to the *cas* operon and the CRISPR array. In fact, the crRNA (CRISPR RNA), which carries the 'invaders' specific sequence, must form a duplex with the tracrRNA (*trans*-activating crRNA) to bind and activate the Cas enzyme. The Cas enzyme is a ribonucleoprotein endonuclease, which introduces DSB in a targeted manner, thanks to the association with its RNA components crRNA/tracrRNA (Jiang & Doudna, 2017; Mohamadi et al., 2020; Sternberg & Doudna, 2015).

In 2012, the Doudna group showed that a chimeric molecule, called sgRNA (single guide RNA) or simply gRNA, made by the fusion of crRNA and tracrRNA, was able to assemble with and direct Cas to the target site and cut it (Fig. 1.15) (Jinek et al., 2012). Along with the complementarity to the target site, which is mediated by crRNA, Cas has to recognise a PAM sequence (Protospace Adjacent Motif) that can be located at the 5' or 3' of the target site and be composed of a different set of nucleotides depending on the species. For example, *S.pyogenes* Cas9 (one of the most commonly used) identifies the PAM sequence "NGG" at the 3' end (Leenay & Beisel, 2017). The Cas endonuclease uses PAM to discern between self and non-self, avoiding destroying the CRISPR locus itself (Jinek et al., 2012).

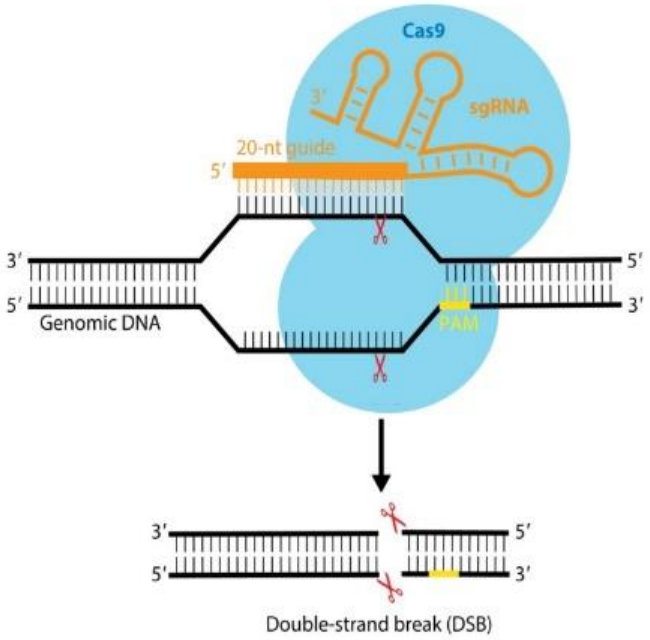


Figure 1.15 Engineered Cas9 enzyme activated by synthetic single-guide RNA (sgRNA). The Cas9 endonuclease is guided by the 20-nt sequence of the sgRNA, designed explicitly by researchers to target genomic DNA. Image adapted from Jiang & Doudna, 2017.

The extraordinary potentiality of the CRISPR system was immediately exploited. Doudna and Zhang groups were the first to engineer *S. pyogenes* and *S. thermophilus* Cas9 and test different combinations of sgRNAs or crRNAs/tracrRNAs on human and mouse cells (Cong et al., 2013; Jinek et al., 2013). Since 2013, an exponential number of papers contemplating the use of a CRISPR/Cas technology have been published (Adli, 2018), thus demonstrating an ongoing revolution in the field of genome editing. CRISPR/Cas indeed represents the most straightforward and cheap approach among all the genome editing techniques (Kouranova et al., 2016).

1.6.2 CRISPR/Cas delivery systems

The system's primary key is the design of a target-specific RNA that guides the Cas9 enzyme to virtually any sequence of any organism. These two elements -Cas9 and gRNA- can be delivered into cells in several ways. The heterologous Cas9 expression, along with the gRNA expression, can be obtained by lentivirus or plasmid transfection, direct DNA/RNA injection, or in vitro assembly of the entire ribonucleoprotein complex (Cas9 + gRNA). Different strategies can be equally accepted and effective according to the application and organism (Kouranova et al., 2016; Sternberg & Doudna, 2015).

Plasmids have been largely used as Cas9 and gRNA carriers, primarily because of their reasonable cost and versatility. A plasmid could contain both the Cas9 gene and the gRNA. There is also a possibility to engineer plasmids to include several gRNA to carry out a multiplex genome editing (Sakuma et al., 2014). Plasmids resist cellular degradation for longer, consenting to a more thorough editing process but, at the same time, increasing the risk of introducing off-target modifications. Also, a definite possibility of plasmid integration into the genome exists and could lead to unwanted genomic rearrangements (Kouranova et al., 2016).

A method that is starting to gain popularity due to its numerous advantages is Cas9 RNP (ribonucleoprotein). The Cas9 RNP consists of a purified Cas9 enzyme that can be easily assembled with any desired gRNA in vitro. Both elements -enzyme and RNA- can be purchased from specialised companies. Contrary to plasmid transfection, the adoption of RNP eliminates the risk of genomic integration and reduces expression variability (*The Problems of Using Plasmids for CRISPR Genome Editing*, n.d.). A plasmid may be

expressed or not, or differentially expressed depending on cell type and cell status (Al-Dosari & Gao, 2009).

Furthermore, as demonstrated by Kim and Kouranova, off-target modifications are less prone to occur when pre-assembled RNPs are delivered (Kim et al., 2014; Kouranova et al., 2016). To conclude, RNPs are very efficient because they don't rely on the cell expression system; instead, they are ready to detect the target site and cut it. They are also degraded relatively quickly, reducing off-target activity (Mout et al., 2017; Vakulskas et al., 2018).

1.6.3 DNA repair mechanisms

As per CRISPR/Cas function, introducing a targeted double-strand break in the genome would not be sufficient to produce a permanent genomic modification. It is, however, the trigger that induces the engagement of the cell's repair mechanisms. Therefore, endogenous repair systems become an active and essential part of genome-editing technology (Cubbon et al., 2018).

Eukaryotic cells have established a series of protective mechanisms to preserve their integrity and minimise the disruptive effect of genetic material damage. DNA is naturally exposed to many types of endogenous and exogenous stresses. Ionising radiation, ROS (reactive oxygen species), and chemotherapeutic drugs can all cause the break of one or both strands of DNA. Even some programmed cell events, like meiotic recombination, can cause fracture of the nucleotide chain. These events may compromise DNA stability and eventually lead to chromosomal aberrations, deletions, and apoptosis unless repair mechanisms are implemented. All Eukaryotes present two different main tools to repair DSB: homology-directed repair (HDR) and non-homologous end joining (NHEJ) (Cooper, 2000; Featherstone & Jackson, 1999).

The homology-directed repair pathway is a recombinational process that generally uses a second copy of the cleaved chromosome, or a piece of homologous DNA, as a template (Karran, 2000). Homologous recombination is not based on a single linear mechanism. Still, the enzymatic machinery assembling at the breaking point can vary depending on the type of DSB, the cell cycle phase, and even the cell type (Haber, 2000). This repair

mechanism is very complex and variable among organisms, but the core steps remain unchanged. Initially, the DSB is recognised and extended until forming two ssDNA ends. When sequence complementarity is found, the strands face a mutual exchange with the subsequent formation of a joint molecule –single or double Holliday junction– that will finally be resolved into two separated repaired chromosomes (Kowalczykowski, 2015).

A recombination event cannot occur outside the S/G2 phase of the cell cycle, as DNA replication is not happening. Therefore, the cell relies on the non-homologous end joining (NHEJ) (M. Liu et al., 2019; Yang et al., 2020). NHEJ evolved to adapt to various DNA DSBs; hence, it is characterised by flexibility and iteration. Despite having a separate evolutionary origin, both Pro- and Eukaryotes show the same enzymatic machinery, made of DNA-binding proteins, nuclease, polymerase, and ligase. The objective of this repair machine is to directly join the DNA extremities, avoiding large deletions or chromosome loss at the expense of introducing errors. The terminal ends of the damaged DNA undergo a series of digestion/polymerisation steps independently until the ligase complex recognises that the strands are similar enough to be annealed and ligated (Lieber, 2010; Pannunzio et al., 2018).

Depending on the nature of the breakpoint and the repair circumstances, HDR and NHEJ could either compete or interact. These mechanisms were considered the sole cell repair pathways for a long time. Recently, a third mechanism has been identified and denominated Microhomology-Mediated End Joining (MMEJ). Evidence suggests that MMEJ is a physiological repair process in normal cells and, despite being an error-prone system, is highly effective in protecting genome integrity. The critical feature of MMEJ is the ability to recognise and anneal short homologous sequences (5-25 bp) flanking the DSB. This mechanism facilitates the joining of the extremities, although leading to deletions (Kawahara et al., 2016; McVey & Lee, 2008; Seol et al., 2018). Although some aspects of the MMEJ remain unclear, the repair process can be summarised as follows. Initially, the ends resulting from the DSB are identified and partially resected until small homologies are met on both strands. These overlapping regions are now the core of the repairing site, leading to the trimming of the 3’ heterologous flap on each strand. Finally, a polymerase fills the remaining gaps, and a ligase restores the double-strand structure (Wang & Xu, 2017). A diagram of the three systems is shown in Figure 1.16.

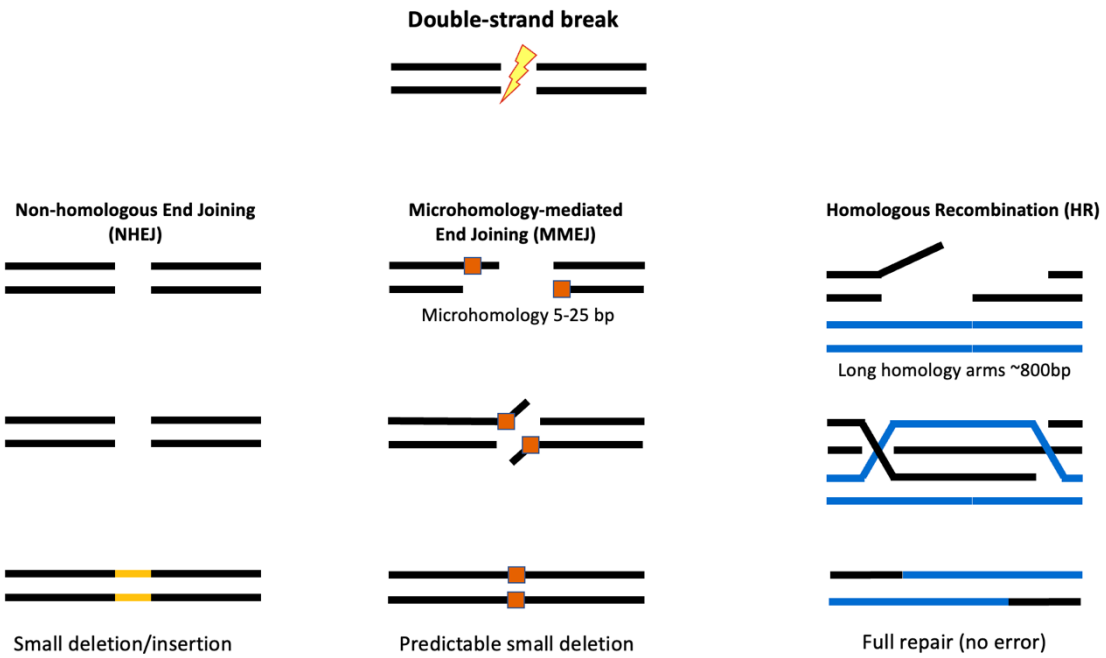


Figure 1.16 Schematic of three DNA repair mechanisms: Non-homologous End Joining (NHEJ) (on the left), Microhomology-mediated End Joining (MMEJ) (in the middle), Homologous recombination (HR) (on the right). Image adapted from Kawahara et al., 2016.

1.7 Aims and Objectives

Considering the limited scientific literature available on DHFR2, our research questions stemmed from the hypothesis that the endogenous DHFR2 protein might not be made or be functional (despite the possibility of producing a recombinant enzyme), as explained in Section 1.5 (last paragraph). In addition, evidence of interactions occurring between *DHFR2* RNA and DHFR protein (McEntee et al., 2011) opened the possibility of a regulatory function for *DHFR2* in form of RNA. Because of the fragmented knowledge on *DHFR2*, our study proposes to thoroughly investigate the function of endogenous *DHFR2* (if any is found), either in form of protein or RNA and its implication in One-Carbon Metabolism and Cell Growth.

The primary limitation in the design of the experiments was due to the striking similarity between DHFR and DHFR2 proteins. This feature makes standard immunostaining techniques inadequate for the unequivocal identification of DHFR2 in environments

where DHFR is equally expressed. Furthermore, the methods employed in this research, part of which are described in this thesis, respond to the necessity to overcome this practical limitation.

Although the amino acid homology between DHFR and DHFR2 amounts to 92%, the two genes differ significantly, in terms of length, number of exons and overall structure. These unique features have been exploited through the use of differential gene-editing strategies, thus implementing the specificity of the methodology (Fig. 1.17). The final goal was to create a loss-of-function model (DHFR2 knockout) in order to explore DHFR2 function; alongside that, a DHFR knockout line would provide us with a DHFR-free environment to detect and localise DHFR2 unequivocally, other than exploring DHFR2 compensatory functions (in relation to DHFR).

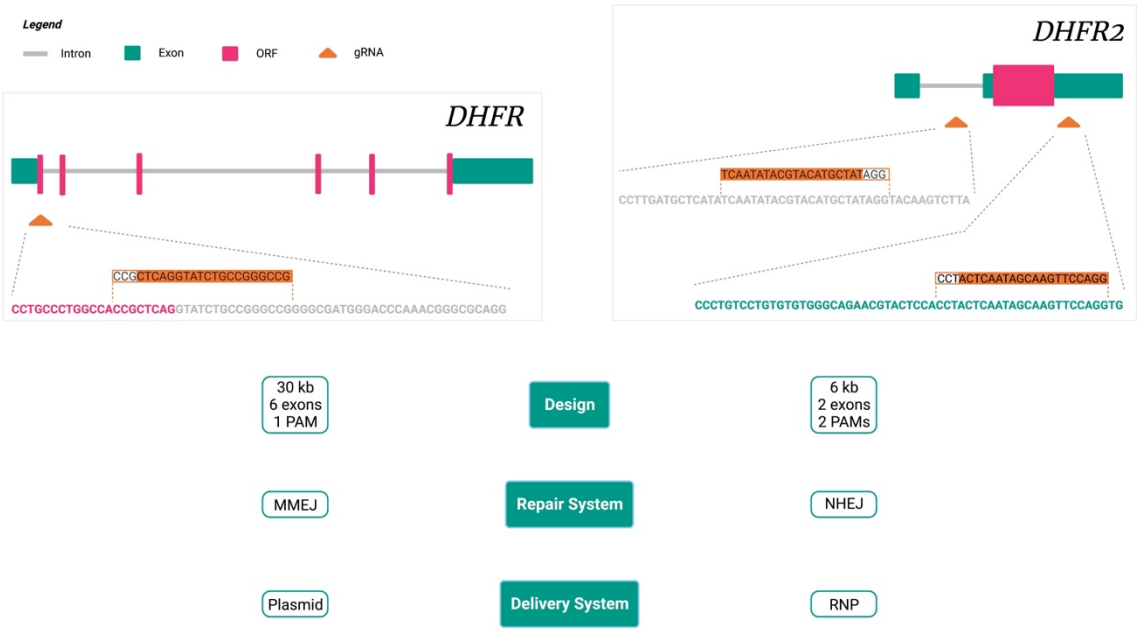


Figure 1.17 Comparison between the gene-editing methods. The image indicates DHFR and DHFR2 genes' structural features with relative gRNAs. A comparison of genes' characteristics, and deployed repair and delivery systems are included.

The aims of this thesis can be summarised as follows:

- To explore whether DHFR2 function is relevant to cell growth.
- To verify the role of DHFR2 in One-Carbon Metabolism, with emphasis on the thymidylate de novo synthesis.
- To examine the relation between DHFR and DHFR2 in terms of direct interaction and functional interdependence (if any).
- To understand the molecular mechanisms that regulate DHFR2.
- To investigate potential differences in localisation, interaction and function of DHFR2 RNA isoforms.

The following objectives will help in reaching the goals of the investigation.

- Establishment of a DHFR knockout cell line (HepG2) using CRISPR/Cas9.
    - Assessment of cell proliferation via cell count, plotting and statistical analysis.
    - Evaluation of the DHFR2 (and additional OCM genes) transcriptional level via RTqPCR.
    - Global transcriptomics screening via microarray for mRNAs and lncRNAs.
    - Profiling of folate metabolites in OCM via UPLC-MS/MS.

- Establishment of a DHFR2 knockout cell line (HepG2) using CRISPR/Cas9.
    - Assessment of cell proliferation via cell count, plotting and statistical analysis.
    - Evaluation of the DHFR (and additional OCM genes) transcriptional level via RTqPCR.
    - Global transcriptomics screening via microarray for mRNAs and lncRNAs.
    - Assessment of DHFR:DHFR2 interdependence via DHFR enzymatic activity test.
    - Profiling of folate metabolites in OCM via UPLC-MS/MS.

- Investigation of localisation and interactions of DHFR2 isoforms in HepG2 cell line and embryonic tissues via Hybridisation Chain Reaction methods.
    - Association with DHFR protein and RNA.
    - Association with ribosomes.
    - Association with nuclear components (paraspeckles).

The following chapters present the work in chronological order, starting with the establishment of the gene-edited lines, DHFR2 knockout (Chapter 3) and DHFR knock-down (Chapter 4). Both gene-edited lines have been thoroughly screened (cell proliferation, transcriptome analysis, folate profiling, enzymatic assay). The results of the analyses are described in Chapter 5 for DHFR2 knockout and Chapter 6 for DHFR knock-down. Finally, Chapter 7 illustrates the investigation of DHFR2 RNA sequence and structural features and association with DHFR, ribosomes and paraspeckles.

Chapter 2
Materials and Methods

2.1 Materials

2.1.1 Reagents

AbCam: Dihydrofolate Reductase Assay Kit (Colorimetric) (Cat. no. ab239705), Mouse monoclonal [1L4] to PSPC1 - C-terminal (Cat. no. ab214012), Rabbit monoclonal [EP1332Y] to alpha Tubulin (Cat. no. ab52866), Rabbit monoclonal [EPR7808] to RPS3 – Ribosome (Cat. no. ab128995).

Addgene: plasmid pCRIS-PITChv2-FBL (Cat. no. 63672), plasmid pX330A-1x2 (Cat. no. 58766), plasmid pX330A-FBL/PITCh (Cat. no. 63671), plasmid pX330S-2-PITCh (Cat. no. 63670).

Agilent: MycoSensor PCR Assay Kit (Cat. no. 302109).

Ambion: RNaseZAP (Cat. no. AM9780).

Bioline: 5x Reaction Buffer (Cat. no. MB-109I), BioScript (Cat. no. BIO-27036), Isolate II Plasmid Mini Kit (Cat. no. BIO-52056), MyTaq HS DNA Pol (Cat. no. BIO-21111), Oligo dT (Cat. no. BIO-38029), Random Hexamer Primers (Cat. no. BIO-38028).

BioRad: Quick Start Bovine Serum Albumin Standard (Cat. no. 5000206), Quick Start Bradford 1x Dye Reagent (Cat. no. 5000205).

Gibco: Dulbecco's phosphate-buffered saline (DPBS) (Cat. no. 14190094), Fetal Bovine Serum qualified (FBS)(Cat. no. 10270106), HT Supplement 100 X (Cat. no. 11067030), MEM Non-Essential Amino Acids 100 X (Cat. no. 11140035), Opti-MEM (Cat. no. 11058021), Trypan Blue Solution 0.4% (Cat. no. 15250061), Trypsin-EDTA 0.25% (Cat. no. 25200056), Cell Culture Freezing Medium (Cat. no. 12648010), DMEM high glucose GlutaMAX Supplement (Cat. no. 61965026), Geneticin Selective Antibiotic (G418 Sulfate) (50 mg/mL) (Cat. no. 10131035), HBSS calcium magnesium no phenol red 500 mL (Cat. no. 14025050), Kanamycin Sulfate (Cat. no. 15160047), Poly-D-Lysine (Cat. no. A3890401).

Greiner Bio-One: 24-well plate (Cat. no. 662160), 6-well plates (Cat. no.657160), T25 flasks (Cat. no. 690175CI), T75 flasks (Cat. no. 658175CI).

Ibidi: μ-Slide 8 Well Glass Bottom (Cat. no. 80827)

IDT: Alt-R CRISPR-Cas9 crRNA DHFR2 3' (Cat. no. 74905117), Alt-R CRISPR-Cas9 crRNA DHFR2 5' (Cat. no. 74905116), Alt-R CRISPR-Cas9 tracrRNA (Cat. no. 1072532), Alt-R S.p. HiFi Cas9 Nuclease (Cat. no. 1081060), gBlocks, Nuclease-Free Duplex Buffer (Cat. no. 11-05-01-12).

Invitrogen: Lipofectamine RNAiMAX (Cat. no. 13778100), PureLink RNA Mini Kit (Cat. no. 12183018A), SuperScript III First-Strand Synthesis SuperMix (Cat. no. 18080400), ezDNase Enzyme (Cat. no. 11766051), PureLink DNase Set 50 preps (Cat. no. 12185010), SYBR Safe DNA Gel Stain (Cat. no. S33102), Lipofectamine 3000 Transfection Reagent 0.75ml (Cat. no. L3000008), Image-iT Fixative Solution (4% formaldehyde methanol-free) (Cat. no. R37814), Image-IT LIVE Intracellular Membrane and Nuclear Labeling Kit (Cat. no. I34407), SlowFade Diamond Antifade Mountant (Cat. no. S36963), UltraPure SSC, 20X (Cat. no. 15557044).

Mirus: MiraCLEAN Endotoxin Removal Kit (Cat. no. MIR 5910), Puromycin Dihydrochloride (Cat. no. MIR5940), TransIT-X2 (Cat. no. MIR6003).

Molecular Instruments: GAPDH probe, DHFR2_201 probe, DHFR2_202 probe, DHFR probe, Amplifier B2, Amplifier B3, Amplifier B5, Probe Hybridization Buffer, Probe Wash Buffer, Donkey Anti-Mouse Secondary Antibody, Donkey Anti-Rabbit Secondary Antibody, Antibody Buffer.

NanoEntek: AccuChip kit (Cat. no. AD4K-200).

New England BioLabs: 100bp DNA Ladder (Cat. no. N3231L), 1kb DNA Ladder (Cat. no. N3232L), Bbs I – HF (Cat. no. R3539S), Bgl II (Cat. no. R0144S), Bsa I-HF (Cat. no. R3535), CutSmart Buffer (Cat. no. B7204S), Gel Loading Dye, Purple 6X (Cat. no. B7024S), Mlu I (Cat. no. R0198S), Nde I (Cat. no. R0111S), NEBuffer 3:1 (Cat. no. B7203S), Q5 High-Fidelity DNA Pol (Cat. no. M0491S), Quick Ligation Kit (Cat. no. M2200S), T4 DNA Ligase Reaction Buffer (Cat No. B0202S), Xho I (Cat. no. R0146S), ssRNA Ladder (Cat. no. N0362S), RNA Loading Dye (2X) (Cat. no. B0363A).

Origene: pCMV6-AC-DHFRL1 (GFP-tagged) - transcript1 (10ug) (Cat. no. RG232027), pCMV6-AC-DHFRL1 (Myc-DDK-tagged) - transcript variant 2 (10ug) (Cat. no. RC208452), DHFR Mouse Monoclonal Antibody (Cat. no. TA500543S).

Promega: Wizard SV Gel-PCR Clean-Up (Cat. no. A9281).

Qiagen: DNeasy Blood & Tissue Kit (Cat. no. 69504), Plasmid maxi kit (Cat. no. 12162), Proteinase K (Cat. No. 19131), RNeasy Mini Kit (Cat. no. 74104).

Roche: Ampicillin (Cat. No. 91084621), Faststart Essential Dna Probes Master (Cat. no. 06402682001), Universal Probe Library Set, Human (Cat. no. 04683633001), LightCycler 480 plates 96well (Cat. no. 04729692001).

Sarstedt: 100mm Dish (Cat. no. 664160), 6-well plate (Cat. No. 833920), 96-well plate (Cat. no. 831835300).

Sigma: 10X PCR buffer without Magnesium Chloride (Cat. No. P2317), Agarose (Cat. no. A9539), Cell Freezing Media W/ DMSO (Cat. No. S-002-D), DNase I kit (Cat. No. AMPD1), dNTPs Set 100 mM (Cat. No. DNTP100), Dulbecco's Modified Eagle's Medium DMEM hi-glucose (Cat. No. D5796), Ethidium Bromide (Cat. no. E1510), LB broth (Cat. No. L3522), LB broth with agar (Cat. No. L2897), Magnesium Chloride (Cat. No. M8787), Orange G (Cat. no. O3756), Primers, Ribonuclease inhibitor human (Cat. No. R2520), Streptomycin (Cat. no. S6501), Taq DNA Polymerase 5U/µl (Cat. No.D4545), Water molecular biology reagent (Cat. No. W4502-1L).

Takara: In-Fusion HD Cloning Plus (Cat. No. 638909), NucleoSpin Gel and PCR Clean-up (Cat. No. 740609), S.O.C. Medium (Cat. No. ST0215).

Thermo Fisher Scientific: 12-well plate (Cat. no. 150628), 5-bromo-4-chloro-3-indolyl ß-D-galactopyranoside (X-Gal) (Cat. no. 10113253), ExoSAP-IT Express PCR Product Cleanup Reagent (Cat. No. 75001), Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Cat. no. R0392), GeneJET Plasmid Miniprep Kit (Cat. no. K0503), M-PER Mammalian Protein Extraction Reagent (Cat. no. 78503), Halt Protease Inhibitor Cocktail EDTA-Free (100X) (Cat. no. 87785), Ethanol 99%, absolute, extra pure (Cat. no. 10048291), Pierce 20X PBS Tween20 Buffer (Cat. no. 28352), TWEEN 20 (50% Solution) (Cat. no. 003005).

VWR: SuperFrost Microscope Slides (Cat. no. 631-0909), Cover glass 22 x 26 mm (Cat. no. 631-0131), Cover glass 15 mm ∅ (Cat. no. 631-1579).

2.1.2 Solutions

Ampicillin: dissolved in molecular grade water. Stock solution 100mg/ml, working solution 100 µg/ml

Streptomycin: dissolved in molecular grade water. Stock solution 50mg/ml, working solution 50 µg/ml

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

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

TBE 10X: 48.44gTris HCl, 12.37g Boric Acid, 1.5g EDTA, 500ml H_2O

Agarose gel: unless differently specified, all gels are 1% agarose in TBE 1X

IPTG: resuspended in molecular grade water to a final concentration of 100 mM

X-Gal: resuspended in DMSO 99.5 % to a final concentration of 50 mM

2.1.3 Mammalian Cell Lines

Hep G2 Cell Line human (hepatocyte carcinoma): ATCC (Cat.no. 85011430)

2.1.4 Prokaryote Cell Lines

OneShot TOP 10 (*Escherichia coli*): Invitrogen (Cat.no.C4040)

XL10-Gold Ultracompetent Cells (*Escherichia coli*): Stratagene (Cat.no.200314)

Stellar Competent Cells (*Escherichia coli* HST08): Takara (Cat.no.ST0213)

2.2 Methods

2.2.1 Microbial Culture

Bacteria were cultured on the surface of LB agar plates supplemented with the specific antibiotic (100 µg/ml ampicillin; 50 µg/ml streptomycin; 1 X kanamycin) and grown overnight in a static incubator at 37°C. When requested from the adopted procedure, single colonies were isolated using a disposable loop and inoculated into liquid LB broth added with the antibiotic. The colonies were allowed to grow overnight in a shaking incubator at 225 rpm, 37°C.

2.2.2 Plasmid Mini Prep

A single colony was inoculated and cultured overnight in 3 ml LB broth containing either 100 µg/ml ampicillin or 50 µg/ml streptomycin or 1 X kanamycin (according to the resistance gene harboured), then centrifuged at 1500 x g for 3 minutes. After removing the supernatant, the pelleted bacteria were lysed and the plasmids were extracted using either the Isolate II Plasmid Mini Kit (Bioline) or the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) as per manufacturer's protocol.

2.2.3 Plasmid Maxi Prep

Overnight bacterial cultures, obtained inoculating a single colony into 100 ml of media (LB broth supplemented with either 100 µg/ml Ampicillin or 1 X Kanamycin) were harvested and pelleted by centrifugation at 6000 x g for 15 minuntes at 4°C. The DNA was isolated using Qiagen Plasmid Maxi Kit. The pellet was resuspended in buffer P1, P2 and P3 before being filtered through the specific column (QIAGEN-tip). The plasmid DNA, trapped in the column membrane, was eluted in buffer QF and then precipitated using room-temperature isopropanol. After a centrifugation step at 15,000 x g for 30 minutes at 4°C, the DNA pellet was washed with 70 % ethanol, centrifuged and finally dissolved in 50 µl molecular biology grade water.

2.2.4 Endotoxin Removal

Plasmids isolated through a maxi prep were further purified using MiraCLEAN Endotoxin Removal Kit in a bid to remove bacteria outer membrane residues, as per manufacturer's protocol. The plasmid DNA samples were diluted to a 1 µg/µl concentration and mixed with 0.1 volumes of MiraCLEAN Buffer. After a 5-minute incubation on ice, 0.03 volumes of EndoGO Extraction Reagent were added to the sample, followed by an additional 5-minute incubation on ice (intermittent vortexing was applied to the sample). The DNA samples were finally heated at 50°C, chilled and centrifuged for one minute at 14,000 x g. A distinct separation was formed between the clear upper aqueous phase containing the clean DNA and the lower pink phase containing the endotoxin, allowing the removal of the former.

2.2.5 Restriction Enzyme Digestion

2.2.5.1 Double Digestion

Restriction enzyme digestions of plasmid DNA was performed to confirm identity. After plasmid DNA was extracted, a double digestion was performed as a confirmation step. The following restriction enzymes were used and combined in sets of two for obtaining a distinctive pattern of bands resolved in an agarose gel (electrophoresis): *Xho* I / *Bgl II* for pX330A-1x2, *Mlu I* / *Bgl II* for pX330S-2-PITCh, *Nde I* / *Bsa I – HF* for pCRIS-PITChv2-FBL and *Mlu I* / *Bgl II* for pX330A-FBL/PITCh. The reaction mix was as follows: ddH_2O, 10X NEBuffer 3:1/CutSmart, 1µg Plasmid DNA, 10U each Restriction Enzyme. The reaction mix was incubated at 37°C for 1 hour.

2.2.5.2 Single Digestion

Single digestions required only one restriction enzyme and the same reaction mix described for double digestion. As the fragments deriving from single digestions were used further for ligation reactions, the complete cut of the restriction sites was assured by incubating the reactions overnight (instead of 1 hour, as per manufacturer's protocol).

2.2.6 DNA Cloning Ligation

2.2.6.1 Plasmid Modification by Annealed Oligo Cloning

The insert harbouring the DHFR gRNA sequence was ligated into the linearised pX330A-1x2 using the Quick Ligation kit (NEB). A molar ratio of 1:3 vector to insert was adopted and the following formula was used to calculate the correct amount of insert: [(ng of vector x size of insert) / size of vector] x (insert:vector ratio). The ligation mix was set as follows: ddH_2O, 2X Quick Ligase Buffer, Linearised Vector, Insert, 1 µl Quick Ligase. The mix was incubated at 25°C for 5 minutes, chilled on ice and stored at -20°C until used for bacteria transformation.

2.2.6.2 In-Fusion Cloning

The donor vector backbone and the gene cassette were ligated together using an In-Fusion reaction. Vector (50 ng) and Insert (100 ng) were mixed thoroughly with 5X In-Fusion PreMix and ddH_2O and incubated for 15 minutes at 50°C, chilled on ice and stored at -20°C until ready for bacteria transformation.

2.2.6.3 Golden Gate Cloning

The Golden Gate assembly was accomplished using a single digestion/ligation reaction, performed as follows: pX330A-1x2-DHFR and pX330S-2-PITCh (1:2 ratio), 10X T4 DNA Ligation Buffer, 10U *Bsa I* – HF, 1 µl QuickLigase, ddH_2O. The reaction was carried out within a GeneAmp PCR System 9700 (AB) thermal cycler at the following conditions: 37°C, 5 minutes; 16°C, 10 minutes for 25 cycles. An additional reaction was further digested with *Bsa I* – HF at 37°C for 30 minutes, followed by heat inactivation at 80°C for 5 minutes. Both digestion/ligation mixes were used to transform bacteria.

2.2.7 Bacteria transformation

2.2.7.1 OneShot TOP10 competent cells

A vial of OneShot TOP10 was let thaw on ice and immediately transformed with 10 ng plasmid DNA. After a 30-minute incubation on ice, a heat shock was performed by moving the cells to a 42°C water bath for 30 seconds and placed back on ice for 2

minutes. Cells were supplemented with 250 µl LB broth and incubated for 1 hour in a shaking incubator at 37°C, 225 rpm. Two different volumes (20 µl and 100 µl) of transformed bacteria were then plated (LB agar + Amp/Str) and let grow overnight in a static incubator at 37°C. Single colonies were picked to be inoculated in LB broth + Amp/Str and grown overnight.

2.2.7.2 XL10 Gold Ultracompetent cells

XL10 Gold (50 µl) was thawed on ice and supplemented with 2 µl Golden Gate Assembly reaction mix (pX330A-1x2-DHFR + pX330S-2-PITCh), then incubated on ice for 30 minutes. After a 30-second heat shock at 42°C, the cells were placed back on ice for 2 minutes, added with 300 µl S.O.C. medium (from In-Fusion kit) and placed into a shaking incubator for 1 hour at 37°C. Two different volumes of cells (20 µl and 100 µl) were streaked onto IPTG/X-Gal Ampicillin plates. These plates were prepared by spreading 40 µl IPTG 100 mM on the surface of 100 µg/ml Ampicillin plates and let dry for 30 minutes in the proximity of a Bunsen burner. Then, 120 µl X-Gal 50 mM were added the same way. The plates were stored in the fridge wrapped in foil to avoid X-Gal degradation.

2.2.7.3 Stellar competent cells

Stellar Competent Cells (50 µl) were mixed gently after thawing and added with 2.5 µl In-Fusion reaction mixture. A 30-minute incubation on ice was followed by heat shock at 42°C for exactly 45 seconds and ice incubation for 2 minutes. The cells were supplemented with 450 µl S.O.C. and shaken at 225 rpm, 37°C for 1 hour. To assure the growth of colonies, all cells were plated in three different volumes (5 µl, 100 µl, all the remaining).

2.2.8 Gel electrophoresis

Unless indicated otherwise, all agarose gels were prepared by mixing 1% agarose (w/v) in either 50 ml or 100 ml TBE 1X. The solution was brought to a boil using a common microwave, then let cool on the bench until reaching a temperature of approximately 55°C. Ethidium Bromide (stock solution 10 mg/ml) was added to a final concentration of 0.1 µl/ml and poured into the mould. According to the expected product size, the PCR reaction mix was supplemented with either Orange G (5X) or Purple Gel Loading Dye

(6X) before being loaded into wells. Gels were run at 90-100 V for 45-60 minutes and visualised through DNR Mini-Bis Pro Bio-Imaging System. Conversely, gels described in Chapters 5 and 6 were stained with SYBR Safe (Invitrogen) and visualised through Azure 200 Gel Imaging Workstation (Cambridge Bioscience).

2.2.9 Purification of gel extracted PCR products and vectors

Pictures taken by DNR Mini-Bis Pro Bio-Imaging System allowed the visualization of the desired DNA fragment and facilitated their excision using a scalpel. The gel slice was collected into a 1.5 ml tube and mixed with Membrane Binding Solution. The tube was incubated at 65°C for 10 minutes and together vortexed every few minutes to help the gel dissolve thoroughly. The DNA was purified by centrifugation using the Wizard SV Gel and PCR Clean-Up System as per manufacturer's protocol.

Contrarily, NucleoSpin Gel and PCR Clean-up was solely used to purify vector and insert DNA fragments to construct the Donor vector. After loading and running a 1% agarose gel with the two amplified DNA pieces (25 µl), the relative bands were excised out and dissolved in Buffer NTI (200 µl buffer per 100 mg gel). The sample was incubated at 50°C for 10 minutes and vortexed every 2-3 minutes until the slice was completely dissolved. The DNA was eluted from the agarose using the NucleoSpin columns and buffers as per manufacturer's protocol. To increase the yield of DNA, two consecutive elution steps (20 µl and 15 µl) were performed with elution buffer pre-warmed at 70°C.

2.2.10 Quantification of DNA and RNA concentration

The concentration of all DNA and RNA samples were measured using either a NanoDrop 1000 or NanoDrop One Microvolume UV-Vis Spectrophotometer (Thermo Fisher Scientific).

2.2.11 Sanger Sequencing

All PCR products were sequenced by either Source BioScience (Tramore, Ireland) or Eurofins Genomics (Ebersberg, Germany).

2.2.12 Polymerase Chain Reaction (PCR)

All reactions were performed in the Applied Biosystem GeneAmp PCR system 9700 unless otherwise stated.

2.2.13 Mammalian Cell Culture Methods

2.2.13.1 Cell Maintenance

HepG2 cells (hepatocyte carcinoma) were cultured in Dulbecco's Modified Eagle's Medium (DMEM) hi-glucose, supplemented with 10% Fetal Bovine Serum at 37°C, in a 5% CO_2 incubator. Cells were generally grown in 100 mm dishes containing 12 ml medium. At 80% confluence, cells were harvested using 3 ml Trypsin-EDTA 0.25% after one/two washing steps with DPBS 1X to eliminate the medium residues. When different vessels were used, the medium and trypsin volumes were changed accordingly. Cells were left detaching for 5 minutes in the incubator (37°C, 5% CO_2), then the trypsin was deactivated by adding the same volume of medium followed by centrifugation at 150 x g for 3 minutes.

2.2.13.2 Cell Count

Pelleted cells were generally resuspended in 5.5 ml complete media. Cell suspension (0.5 ml) was transferred into an Eppendorf, to which 0.1 ml Trypan Blue was added. The mixture was intermittently vortexed for 10 seconds and incubated at room temperature for 5 minutes. An aliquot of cells (20 µl) was pipetted on one of the haemocytometer's chambers and a coverslip was placed over the specimen. Cells were counted in the four corner squares and the average number was calculated. To determine the cell concentration, the average number was multiplied by the dilution factor (1.2) and by the area under the coverslip (10^4).

Specifically to the Cell Growth Curve (Chapters 5 and 6), cells were counted using the ADAM Automated Cell Counter. Pelleted cells were resuspended in an appropriate volume of growth media to dilute to a final concentration of $5x10^4$ cells/ml to $4x10^6$ cells/ml. To count total cells and viable cells, two different suspensions were prepared by adding 50µl of Accustain solution T and solution N to 50µl of cells. After brief

vortexing, 12µl of each suspension were loaded in the appropriate channel (T or N) of the Accuchip 4x slide.

2.2.13.3 Recovery of cells from long term storage

Cells stored in liquid nitrogen were resuscitated by rapidly transferring into the biological safety cabinet, where the cryovial lid was slightly twisted and tightened back (using a tissue soaked in 70% IMS) to remove nitrogen residues and immediately moved to a 37°C water bath. When a few ice crystals were still present (~ 2 minutes), the vial was moved back to the BSC and the whole content was transferred to a universal tube containing complete media. The tube was gently swung and centrifuged (150 x g, 3 minutes). Cells were resuspended and counted before being seeded.

2.2.13.4 Single-cell Cloning

Cells were seeded into a 96-well plate at a 1 x 10^4 cells/ml concentration using regular media added with conditioned media in a 1:1 ratio Each well was filled with a final volume of 200 µl media. Cell populations from a single cell were selected and further passaged into a 12-well plate, then to a 6-well plate and finally to a 100mm dish upon reaching confluency per each used vessel. Cells were incubated and grown in a 5 % CO_2 incubator at 37 °C. Conditioned media was prepared from media in which cells had been grown for 24 hours, centrifuged at 10 x g for 10 minutes and filtered on a 0.2 µm filter.

2.2.13.5 Dose-response curve for puromycin selection

Cells were seeded into a 24-well plate at a 3 x 10^5 cells/ml concentration, containing 0.5 ml each. Cells were grown into a 5 % CO_2 incubator at 37 °C for two days when they reached a 60-70 % confluence. Media was replaced with fresh DMEM + 10 % FBS and puromycin at increasing concentrations: 0, 0.25, 0.5, 1, 1.5, 2, 4, 8 µg/ml final concentration, in duplicate. Puromycin-added media was replaced every two days. Cells were observed for six days. Optimal killing concentration was inferred by observation.

2.2.14 Mycoplasma test

Cell cultures were tested for *Mycoplasma* infection using the MycoSensor PCR assay kit (Agilent). The supernatant from cell cultures was used in the PCR reaction after being

boiled at 95°C and treated with the StrataClean resin. Alongside the test samples, negative and positive controls were also tested. The primer mix would generate a 315 bp band in the positive control and in the samples if infected with any of the most common Mycoplasma species. An internal control (producing a 500bp band) was added to all tubes to confirm the absence of PCR inhibitors. Each 50 μl reaction included: 5 μl 10X PCR Buffer, 0.5 μl Taq DNA Polymerase (5U/μl), 6 μl $MgCl_2$ 25mM, 1 μl dNTP/dUTP mix, 4 μl Internal control template, 2 μl Mycoplasma primer set, 5 μl *Mycoplasma orale* positive control template/Resin-treated supernatant, H_2O. The PCR amplification was performed in a thermal cycler (GeneAmp PCR System 9700 AB), 35 cycles at the following temperatures: 94°C for 30 seconds, 55°C for 1 minute, 72°C for 1 minute.

2.2.15 Transfection

Cell transfection was employed to artificially introduce plasmid DNA into cells. At the moment of transfection, cells were between passage 5 and 25 post-thaw. They were passaged every three days to avoid them entering senescence. To ensure an efficient transfection, cells had to be actively growing; thus they were passaged the day before transfection. When counted, cells had to be over 90% viable. HepG2 cells were generally seeded at 1.5 x 10^5 cells/well in a 24-well plate, a concentration that allowed them to be 70-90% confluent on the following day. The seeding concentration was scaled up accordingly when larger vessels were used. To ensure the dissociation of clumps, cells were resuspended with the help of a p200 tip inserted on top of a 5ml/10ml serological pipette and pipetted up and down 5 times. Having single cells is essential to improve transfection efficiency. On the day of transfection, the lipid-DNA complexes were assembled in Opti-MEM. Lipofectamine 3000 Transfection Reagent (Invitrogen) was used (as per manufacturer's protocol) along with 500 ng plasmid DNA per well. The mixture was added to the cells, which were incubated (37°C, 5% CO_2) for 48 h before being analysed.

2.2.16 Cell fixation and staining

A 22 x 26 mm cover glass was placed onto a well of a 6-well plate (alternatively, 15 mm ∅ cover glass for 24-well plate) for cells to be cultured on its surface. At 70-80%

confluence, cells could be fixed and stained for fluorescence microscopy. After media removal and PBS wash, cells were fixed with 1 ml 4% formaldehyde (Invitrogen Image-iT Fixative Solution) for 15 minutes at 37°C. Cells were then washed three times with HBSS (Gibco) and covered with 1 ml staining solution (Invitrogen Image-IT LIVE Intracellular Membrane and Nuclear Labeling Kit) for 10 minutes at room temperature. The staining solution was removed, and the cells were washed twice in HBSS. The cover glass was carefully removed from the plate and mounted, facing down, into a slide with the aid of 30 µl mounting media (Invitrogen SlowFade Diamond Antifade Mountant). The edges were sealed with clear nail polish. The slides were stored in the dark until imaging.

2.2.17 DNA extraction from mammalian cells

The Qiagen DNeasy Blood & Tissue kit was used to purify total DNA from mammalian cells. Cultured cells were harvested at sub-confluence status (maximum 5 x 10^6 cells), pelleted and resuspended in 200 µl 1 X PBS. After adding 20 µl proteinase K (20 mg/ml) and 200 µl lysis buffer (buffer AL), cells were vortexed and incubated at 56°C for 10 minutes. Ethanol 100 % (200 µl) was added before moving the whole lysate into the appropriate mini columns. The following centrifugation step allowed the DNA to bind to the membrane and eliminate the rest of the cell components in the flow-through. Two consecutive washing steps led to removing any residues from the membrane, consenting to the elution of pure DNA in molecular grade water (100 µl)

2.2.18 RNA extraction from mammalian cells

Working surfaces and pieces of equipment were thoroughly wiped with RnaseZap prior to the start of the extraction. Upon trypsinisation, cells were pelleted by centrifugation (300 x g, 5 minutes), snap-frozen (liquid nitrogen) and stored at -80°C for a short time.

Total RNA was extracted using the Qiagen RNeasy Mini Kit (Chapters 3 and 4). The frozen pellet was resuspended in 600 µl RLT buffer by vortexing. The homogenisation of the lysate was achieved by passing the sample through the QIAshredder spin column using centrifugation. The sample was then washed with ethanol 70% and pipetted into the RNeasy spin column. A centrifugation step allowed the RNA to bind to the membrane

while the flow-through was discarded. Before the total RNA could be eluted in 40 µl RNase-free water, three washing steps were performed.

The PureLink RNA Mini Kit was also used (Chapters 5 and 6). The frozen pellet was resuspended in an appropriate volume of Lysis Buffer added with a reducing agent (1% 2-mercaptoethanol or 40 mM DTT). The lysate was then syringed 5-10 times through an 18-21 gauge needle to obtain a homogenised suspension. The sample was added with one volume of 70% ethanol before being transferred to the Spin Cartridge and centrifuged. After three washing steps, the purified RNA was eluted in an appropriate volume of RNase-free water.

Traces of genomic DNA were eliminated using DNase treatment performed either during (on-column PureLink DNase Treatment) or post RNA purification (DNase I; ezDNase).

2.2.19 RNA integrity check

The integrity of all RNA samples (both freshly extracted and frozen) was assessed prior to cDNA synthesis. An aliquot of RNA was diluted to 50 ng/µl. NEB RNA Loading Dye (2X) was added to 5µl of RNA and heated at 70°C for 2 minutes, then cooled on ice for at least 1 minute. The sample was loaded into a 1% agarose gel and electrophoresed at 100 V for 1 hour. An intact total RNA sample would resolve as two bands (corresponding to the largest ribosomal RNAs: 18S and 28S), with the 28S band showing a 2 x higher intensity than the 18S band. Degraded RNA resolves as a low molecular weight smear.

2.2.20 Reverse-Transcription of RNA to cDNA (and genomic DNA removal)

A *DNase I* (alternatively, ezDNase) treatment was performed before the reverse-transcription, aiming to eliminate any DNA residues left from the RNA extraction. RNA (2 µg) was diluted in a total volume of 8 µl. *DNase I* (1 unit/µL) and DNase buffer -1 µl each- were added to the sample and incubated for 15 minutes at room temperature. The DNase activity was stopped by adding 1 µl DNase Stop Solution. The mixture was incubated at 70°C for 10 minutes and cooled on ice. The DNase-treated RNA had to be split in two to proceed with the reverse-transcription, thus working with 1µg RNA per tube. Each sample was added with 2µl random hexamers (50 ng/µL) + 1 µl oligo dT (270

ng/µL), incubated at 70°C for 5 minutes and cooled on ice for at least 1 minute. Ribonuclease inhibitor (1 µl, concentrated 30,000-50,000 units/ml) was added to the mixture, along with 1 µl dNTPs 10 mM, 1 µl BioScript, 4 µl 5X Reaction Buffer, 1.5 µl H_2O. The reaction was incubated in a thermal cycler (GeneAmp PCR System 9700 AB) under the following conditions: 25°C for 10 minutes, 42°C for 60 minutes, 70°C for 15 minutes. The resulting cDNA was stored at -20°C.

In addition, cDNA used in the experiments described in Chapters 5 and 6 was synthesised through the SuperScript III First-Strand Synthesis SuperMix. Total RNA (1 µg) was combined with 1 µl Oligo-dT (50 µM) and 1 µl annealing buffer and incubated at 65°C for 5 minutes. After adding 2X First-Strand Reaction Mix and SuperScript III/RNaseOUT Enzyme Mix (2 µl), the mixture was incubated at 50°C for 50 minutes. The reaction was inactivated by increasing the temperature to 85°C for 5 minutes and then stored at 20°C.

2.2.21 Genomic contamination assay

A genomic contamination assay was performed on every cDNA sample to ensure the DNase treatment was effective. This PCR assay (MTHFD1 R653Q) (Appendix G) leads to the amplification of a 232bp band from the cDNA component and a 330bp band from genomic DNA, if any is present in the sample. The reaction mix is as follows: 10X PCR buffer, 0.2 µl Taq DNA Polymerase (5U/ µl), 3 µl $MgCl_2$ (25mM), 4 µl dNTPs (2.5mM), 2.5 µl MTHFD1 Fwd (10µM), 2.5 µl MTHFD1 Rev (10µM), 1 µl cDNA, H_2O to a final volume of 50 µl. The PCR mixture was put in a thermal cycler (GeneAmp PCR System 9700 AB)at the following conditions: initial denaturation at 95°C for 3 minutes; 35 cycles at 94°C for 30 seconds, 58°C for 1 minute, 72°C for 1 minute; final elongation at 72°C for 10 minutes.

2.2.22 Quantitative Reverse Transcription PCR (RT-qPCR)

All assays were designed using the Roche Universal Probe Library Design Tool (discontinued), which indicated the probes to use (Roche Universal Probe Library Set, Human) alongside the primers (manufactured by Sigma) and the master mix (Faststart Essential Dna Probes Master). A comprehensive list of the assays, with relative primer/probe sequence and concentration, can be found in the Appendix W. All assays were run in triplicate on the LightCycler 96 (Roche) under the following conditions: pre-

incubation at 95°C for 10 minutes, 45 cycles of amplification divided into 30 seconds at 95°C, 30 seconds at 60°C, 30 seconds at 72°C followed by a cooling step at 40°C for 10 seconds. Amplification data were visualised and analysed through the Lightcycler 96 software.

2.2.23 Protein extraction

Cells were cultured in T125 intending to obtain enough proteins. Cells were collected by trypsinisation when 80% confluent and snap-frozen (stored at -80°C for a short time). On the day of protein extraction, the pellet was washed with PBS to eliminate any traces of phenol red (from media and trypsin). The volume of lysis buffer (Thermo Fisher M-PER Mammalian Protein Extraction Reagent) was calculated based on the estimated cell pellet weight, considering at least 1 ml M-PER Reagent for each 100 mg of wet cell pellet. The lysis buffer was supplemented with a protease inhibitor (Thermo Fisher Halt Protease Inhibitor Cocktail EDTA-Free, 100 X) and added to the cell pellet. After spending 10 minutes in the shaking incubator, the resuspended/lysed cells were centrifuged at 14000 x g for 15 minutes. The supernatant, which contained free proteins, was transferred to a clean tube and stored at -80°C for further analysis. An aliquot (25μl) of protein was used for estimating the sample concentration by Bradford assay.

2.2.24 Bradford protein assay (protein quantification)

Quick Start Bovine Serum Albumin (BioRad) was used to make a standard curve, in triplicate. The unknown samples were measured as neat and as 1:5 and 1:10 dilutions, each in triplicate. Both standards and samples were added to a 96-well plate with a volume of 4 μl. Quick Start Bradford 1x Dye Reagent was added using a multichannel pipette (200 μl per well), making sure to mix the content of the wells avoiding the production of bubbles. The plate was incubated for 5-10 minutes at room temperature before being read at the spectrophotometer (Tecan I-Control Infinite 200) at 595 nm. The average of the absorbance values of the triplicates was calculated, and the blank was subtracted from it. The standard curve was plotted, and the relative equation was generated. The absorbance values of the unknown samples were substituted into the standard equation to deduce the sample concentration (mg/ml).

2.2.25 DHFR enzymatic assay

DHFR enzymatic activity was assessed using a colourimetric assay (Abcam Dihydrofolate Reductase Assay Kit), which exploits the natural ability of DHFR to reduce NADPH. NADPH consumption is read as a decrease in absorbance at 340 nm. The test was performed on a 96-well plate in the Tecan I-Control Infinite 200 spectrophotometer. The samples consisted of purified total proteins extracted from HepG2 cell lines (section 2.2.23). An NADPH standard curve was prepared and analysed in endpoint mode. Each tested sample consisted of total protein, NADPH and DHF. An identical sample minus DHF was run alongside this, serving as background control. Each plate also contained a positive control made of purified DHFR (+ DHF, + NADPH), a negative control lacking the protein component (DHF and NADPH included in the reaction), and a Blank made solely of the buffer. All samples were read in kinetic mode for 60 minutes. The Blank absorbance values were subtracted from those of the samples, and a ΔOD was calculated between t0 and t60 for both test samples and background samples. The ΔOD values were applied to the standard curve equation to find the number of moles of NADPH used. The following equation was used to calculate DHFR activity:

$$\text{DHFR mU/mg} = (\text{n mol sample} - \text{n mol background} / \Delta t) * \text{mg protein in well}$$

where 1 U is the amount of DHFR that oxidises 1 μmol NADPH per minute at pH 7.5 at room temperature.

2.2.26 Hybridisation Chain Reaction IHC + RNA-FISH

The HCR (Hybridisation Chain Reaction) technology was applied to an Immunohistochemistry and RNA-FISH (Fluorescence *in situ* Hybridisation) multiplexed localisation experiment. The protocol consisted of three stages plus the initial preparation of fixed cells on a chambered slide.

The description of the following steps refers to mammalian cells on a chambered slide. The protocol for tissue sections on a slide was slightly different, but the steps of the procedure remain the same. For detailed protocols, depending on the sample type, please see the Molecular Instrument HCR IHC + HCR RNA-FISH web page (https://www.molecularinstruments.com/hcr-ihc-hcr-rnafish-protocols).

The volume of all reagents used equals 300 μl per chamber unless otherwise stated.

2.2.26.1 Fixation of HepG2 cells on a chambered slide

A chambered slide (Ibidi μ-Slide 8 Well Glass Bottom) was coated with 0.01% poly-D-Lysine (Gibco) and incubated for 30 minutes at room temperature. After removing the coating solution and washing twice with molecular grade water, cells were plated and grown until the desired confluence. Cells were fixed (as described in section 2.2.16) and permeabilised overnight with 70% ice-cold ethanol at -20°C.

2.2.26.2 Protein Detection (immunochemistry)

The chambers were washed off the ethanol with PBS (2 x 5 minutes) and added with Antibody Buffer. The slide was incubated at room temperature for 1 h with gentle agitation. The Antibody Buffer was replaced with Primary Antibody solution and incubated overnight at 4°C with gentle agitation. The excess antibodies were washed with PBS-Tween (3 x 5 minutes). Initiator-labelled secondary antibody solution (1 μg/ml) was added and incubated for 1 h at room temperature with gentle agitation. The excess antibodies were washed with PBS-Tween (3 x 5 minutes).

2.2.26.3 RNA detection (RNA-FISH)

The samples were post-fixed and washed with SSC. A pre-hybridisation step was done by adding Probe Hybridisation Buffer and incubating for 30 minutes at 37°C. The Probe Hybridisation Buffer was replaced with Probe Solution (16 nM) and incubated overnight at 37°C. The excess probes were washed with Probe Wash Buffer (4 x 5 minutes), and then washed again with SSC-Tween.

2.2.26.4 Amplification (Hybridisation Chain Reaction)

The Amplification Buffer was added and the sample was incubated for 30 minutes at room temperature. Then, it was replaced by the hairpin solution (60 nM), which contained all snap-cooled hairpins h1 and h2. The slide was incubated overnight in a dark place at room temperature. The excess hairpins were washed with SSC (5 x 5 minutes). The slide was mounted by adding 150 μl mounting media per chamber and stored at 4°C protected from light before imaging.

2.2.27 Bioinformatic Resources

2.2.27.1 Primer BLAST

Primer-BLAST (Basic Local Alignment Search Tool) is a publicly available bioinformatics tool that helps to design and/or check the specificity of a pair of primers to be used in a Polymerase Chain Reaction (https://www.ncbi.nlm.nih.gov/tools/primer-blast/).

2.2.27.2 CRISPR gRNA design tools

Several web tools have been consulted to analyse the most suitable gRNAs to use in DHFR and DHFR2 knock-out strategies.

- DESKGEN Guide Picker (https://www.deskgen.com/guide-picker/#/)
- PITCh designer 2.0 (http://www.mls.sci.hiroshima-u.ac.jp/smg/PITChdesigner/index.html)
- IDT Alt-R Custom Cas9 crRNA Design Tool (https://eu.idtdna.com/site/order/designtool/index/CRISPR_CUSTOM)

2.2.27.3 Geneious Prime 2019.2 and 2022.1.1

Sequence analysis software used to visualise and compare Sanger sequenced products with their expected sequence. (https://www.geneious.com).

2.2.27.4 SnapGene Viewer

Molecular Biology software used to visualise plasmid sequences and annotate specific features such as restriction enzymes sites and primers locations.

2.2.27.5 R Studio

Statistical analysis of cell growth was performed using R Studio. Particularly, the growth of several Hep G2 populations (wild-type, DHFR knock-out, DHFR2 knock-out; ±HT/NEAA) was monitored throughout a period of ten days, and a comparison was made using several statistical tools, i.e. t-test, ANOVA test, linear regression, permutation test. All statistical tests were accompanied by graphic visualisation so to understand better growth trends (for details, see Appendix S). (RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/).

2.2.27.6 Image J

Java-based programme used for image analysis. The plugin JACoP was used for colocalisation studies.

Chapter 3
An RNP-based CRISPR/Cas method to generate a DHFR2 knockout in HepG2 cells

3.1 Introduction

3.1.1 DHFR2 knockout overview

In loss-of-function studies, the disruption of a gene, and consecutively of its protein, is conducive to changes in the molecular phenotype, with alteration of specific cell functions. To fully understand the biological role of *DHFR2*, a knockout of the gene was planned. The disruption of DHFR2 was anticipated to potentially impact DHFR function, as its interaction with DHFR was previously demonstrated (McEntee et al., 2011). The discovery that DHFR2 RNA is able to bind the DHFR enzyme led to the hypothesis of a new player in the autoregulatory pattern of DHFR, in which the protein binds its own RNA and therefore modulates its translational levels (Abali et al., 2008; Ercikan-Abali et al., 1997; Tai et al., 2004). As a result of DHFR translational modulation, we expected consequent imbalances within the One-Carbon Metabolism, with potential effects on cell proliferation and death.

DHFR2 is a retrogene derived from reverse transcription of a DHFR RNA back into the genome. The entire ORF resides within the same exon. That peculiarity, along with its relatively short length, makes DHFR2 the perfect target for a rapid and easy engineering method. The goal to produce a knockout is reached by a double-cut strategy, in which two DSBs, mediated by CRISPR/Cas, allow the complete deletion of the Open Reading Frame.

The paper "Generation of Genomic Deletions in Mammalian Cell Lines via CRISPR/Cas9" by Bauer et al. 2015 describes a simple CRISPR/Cas approach to achieve genomic deletions in mammalian cells. The idea behind this strategy is very straightforward. If two double-strand breaks are introduced at the edges of a gene, this will be removed entirely, and the resulting extremities will be re-joined together by NHEJ. This technique, however simple, was able to produce the deletion of the entire *Pim1* gene, 8 kb long, in MEL cells (Bauer et al., 2015). This strategy was adapted onto DHFR2, aiming to delete the ORF as a whole, bringing about the effective knockout of the gene.

The similarity rate between DHFR2 and DHFR was a significant concern when it came to choosing the gRNAs targeting the DHFR2 ORF flanking regions. An accurate design,

coupled with a Cas9 RNP, increased the method's specificity. We opted for the use of a HiFi Cas9 (IDT Alt-R S.p. Cas9) complexed with a tracrRNA (the constant part of a gRNA) and a crRNA (the target-specific part of a gRNA). While the tracrRNA has an invariable sequence - that adapts into the relative Cas9 activating it -the crRNA has to be specifically designed to detect the target sequence and guide Cas9 over it.

Note to the Reader: This chapter discusses a collection of methods used to establish a knockout line. Therefore, the following sections will not be characterised by the standard partition in Methods and Results. The majority of the following sections will describe methods; occasionally, subsections of Methods and Results are integrated.

3.2 Construction of the Ribonucleoprotein complex

3.2.1 crRNAs design

The Custom Alt-R CRISPR-Cas9 guide RNA tool (*CRISPR-Cas9 Guide RNA Design Checker | IDT*, n.d., p. 9) was used to design two crRNAs that could allow a unique cut at both ends of DHFR2 ORF. All possible target sites were visualised with the relative on- and off-target activity. The crRNAs shown in Table 3.1 were selected as presenting the best combination of highest on-target and off-target activity.

Table 3.1 crRNAs selected to target DHFR2 gene.

Name crRNA	*Strand*	*Location*	*Sequence*	*PAM*	*On-Target*	*Off-Target*
5' DHFR2	-	Intron 1	TCAATATACGTACATGCTAT	AGG	74	57
3' DHFR2	+	Exon 2	CCTGGAACTTGCTATTGAGT	AGG	73	54

The two sequences are strategically located at either edge of the ORF, which starts at the beginning of exon 2. The sequence block below shows a section of DHFR2: in red exon 1 and 2, in black intron 1, highlighted in yellow the ORF, bold and underlined the two target sequences (crRNAs).

CTCATTCAGATAGAGGTACCCAGCCGGGCCCCAGGTCTCACCTGGGTCAGCATTGTCCCGCTGAGAGCTT
TCAAACAACAGGTGTACGCGAATCAGTATCAAAAAGCCTCCAGGGAACTGAAAAAAACAGTCTGACTTCC
GGGATTCGCGGGCATCCACTTTCATTCTCACGTTTCCTCTTCAATTTCTCGCGAGATACTGATTTGCTCG
CGCTATAGAATTCTTGTCGAAGAGGCTTAGCTCCCGCGGGAGATTTCGTGCGCAGTTACTTCCGGTAGCT
GGTAAAGGCTGATACTTCCCAGGACGCGGAGGTAACGGGCCAGGGCCAAAGCGACTTTCGCTACTTGGAT
TGGTCGGCGTAGCTTTGGGCGGCCGGACCTTAGAAAGTCACACATCTGCGCGCCTGTGCGGCCCCTGCTT
CTGCGGATGCTGAGGTGGGAGGATCGCTTGGGCAGAAGTTTGAGGTTGCAGTGACCCGTGATCGCGCCAC
TGCACTGCAGCGTGGGCGACAGAGCGAGGCTTTGTCTCAAAAATAAATAAAAAGTCGCATATTGTGCTGA
ACAGTGTGTATATGCATGGGGCTCAATAAAGGAATGAGTTTTGGGGTTATGTTGTATGATGACTGATTTA
AAATATTCTCAAAGCGAAGCTAGTTACTGACTTAATCTACCAAATCTAGTGACCCCTGTTAACCTAGACA
TTCAGTACCCAAGGATAAAAGGGTGGAATATAATGAAAAAGAAATTAAGTCATGCTTTTCCCCAGACCTG
GCCCATGGGGATAATCTCTCCTTTTAATGGAGAGTGTTAAACTGAAGTAGGCATGTAAAGCGGTTACAAC
AGTGATTAGGAAACAGTAAATGCCCAAATGTTAGCTGCTATTATTTTCACTATTGTGATCATTGTGATTT
TTGTCGCGTTGATACCAAATGCGTGAAGACCAAAAGGACCCTCCACTCTAGCCTCCCCAATTCACCTTTA
CAAGGGAGCAATGTTTAATGGTTCTCCAAGGAAGATGTTTATTACACCTATTTTGACACGAATAGAAGTT
TATTTGCTCTTTTGGTGTCATGCTTCAGTAACTATCATTTGTAGCAGCAATCAGCAATTAGAAAAAACCG
GTTTGATATGGTCCCTGATAGTTAATAATCATCATTTTGAGTTGTTCCTTGATGCTCATA***TCAATATACG***
TACATGCTATAGGTACAAGTCTTATATTCATGTAAAAACATGTCTTTCCTTCTAATCTGATTTGAAGCTT
AATATATTTATTTATTTAAACAAGTTTCCAAGTCACCTAAAAAATGCATCTTATTAGAATTTAGATTTGG
TACAAAAGACTATGATAATCAGGAATGAAATGATGCCATCTCTTGGATGCCGTCTGTTTAATGTTTGTGG
AAAAGTCGGTCTAACCTTTCAGCCTCTGTATTTTTTATAATTTTATAAATTTATATTTTATAAATATTTT
ATAAATTTATACTTTATAAATATATTTTATATAATTTATTTTTCTATGTTCATTCCAGAATGTTGCAATT
AATTAATACAGTGGTATTGTATTTGCAATGTAGGCACGTAAAAAAATTTGAAGAAGGGAATTTCGCGGCA
TTCTTGGCCTGGCTTCCTGGCGTAGCCAGCAAGTTCGGAGGTGTTAACCGCTGCTGTCATGTTTCTTTTG
CTAAACTGCATCGTCGCTGTGTCCCAAAACATGGGCATCGGCAAGAACGGGGACCTGCCCAGGCCGCCGC

TCAGGAATGAATTCAGGTATTTCCAGAGAATGACCACAACTTCTTCAGTAGAGGGTAAACAGAATCTGGT
GATTATGGGTAGGAAGACCTGGTTCTCCATTCCTGAGAAGAATCGACCTTTAAAGGATAGAATTAATTTA
GTTCTCAGCAGAGAACTCAAGGAACCTCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGATGCCT
TAAAACTTACTGAACGACCAGAATTAGCAAATAAAGTAGACATGATTTGGATAGTTGGTGGCAGTTCTGT
TTATAAGGAAGCCATGAATCACCTAGGCCATCTTAAACTATTTGTGACAAGGATCATGCAGGACTTTGAA
AGTGACACGTTTTTTTCAGAAATTGACTTGGAGAAATATAAACTTCTGCCTGAATACCCAGGTGTTCTCT
CTGATGTCCAGGAGGGGAAACACATCAAGTACAAATTTGAAGTATGTGAGAAGGATGATTAATATGAAGG
TGTTTTCTGGTTTAAGTTGTTCCCCCTCCCTCTGAGAAAAGTATGCATTTTTACATTAGAAAAGGGACTT
TTGTTGACTTCAGATCTATGGATAATTATTTCTAAGCAACGTGTTTTTATTCCTCACTAATCTTGGCTAT
ATCAGATACCATTTATGAAACATTCTTGCTATAACTGTCTCTCCAAGACCCCGACTGAGTCCCCAGCACC
TGCTACAGTGAGCTGCCATTCCACACCCATCACATGTGGCACTCTTGCCACTCCTTGACATTGTCAGGCT
TTTCTAATGTTGGTAGTATTTATTAAAGATGAAGATGCACATACCCTTCAGCTGAGCAGTTTCACTAGTA
GGAAATACCAAAAGCTTCGTACATGTATATCCAGAGGTTTGTAGACAAATGTTGCAGCCTTTTTTGTAAC
AGTGAAAAACTGAAAACAACCTGGAAGTCCAGTGATGGGAAAATGAATATATTTCTGTCTTAGATTGGGG
AACCCAAAGCAGATTCCAAGACTGAAATTTAAGTGAAAGCAGTTTATTTGCTAGGTCATACCAGAAGTCA
TCAATCGAAGTATGGAGAAATGGAACTGAGAAGGTAAAAACCAGTTCAAAATCAGTGAGGAGGTTCTCAA
TGGTAACGAGCTCCATACTGCTGAGATACAGGGAAACGGAGGGGAGAAAGCTAGAGTATTTAACCCCCAC
TCCTTGGTTGTCAGCTCCCTGTCCTGTGTGTGGGCAGAACGTACTCCACCT***ACTCAATAGCAAGTTCCAG***
GTGTTTGCTGAAAGAAGCTGCTGTAATGTACGGGAACGGTGAATGCCAAACACTTAAAGCAATTCCATGT
TTAAGTATGTAAACTCTTCATACTTTTTTTTTTTTTCTTTTTAAGACAGAATTTCACTCTTGTTGCCCA
GGCGGAGTGCAATGGAACAACCTTGGCTCACGGCAACTTCAGCCTTCCCAGGTTCAAGCAATTCTCCTGC
CTCAGGCTCCTGGGTGGCTGGGACTACAGGCGCGTGCCACCACGCCAGGCTAATTTTGTATTTTTGGTGG

3.2.2 Ribonucleoproteins in vitro assembly

Two distinct RNPs were assembled, one per cleavage site. An RNP comprises a crRNA, a tracrRNA and a HiFi Cas 9 (Fig. 3.1). The crRNA is the only variable piece of the complex, thus being specifically customised. The remaining components -tracrRNA and HiFi Cas 9- are not variable; therefore, they were ordered directly from IDT. The IDT user guide "Cationic lipid delivery of CRISPR ribonucleoprotein complexes into mammalian cells" describes the protocol to assemble the ribonucleoproteins and transfect the cells.

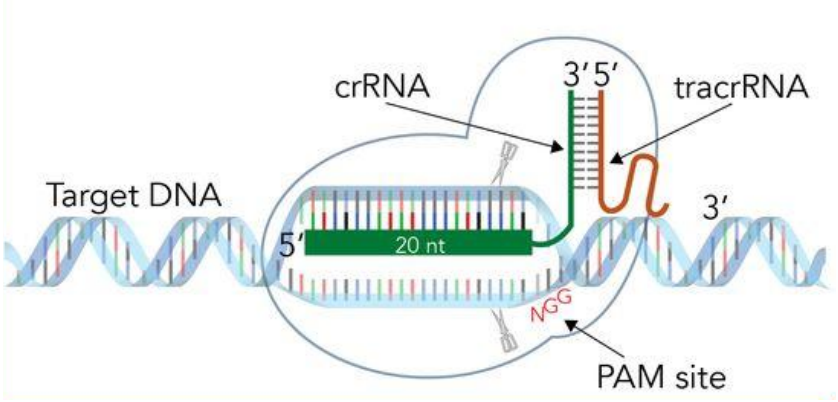


Figure 3.1 Formation of the crRNA:tracrRNA duplex, and its association with Cas9. The ribonucleoprotein, so formed, can recognise the target site through the target-specific portion of crRNA. Upon PAM site recognition by Cas9, the cleavage activity is triggered. Image adapted from IDT website.

The lyophilised RNAs were resuspended to a 100 µM stock concentration. The crRNA and tracrRNA were assembled first to obtain an RNA:RNA duplex. As this strategy requires the simultaneous cleavage of two DNA spots, we prepared two separate duplexes; one made of 5' DHFR2 crRNA + tracrRNA, the other one having 3' DHFR2 crRNA + tracrRNA. The duplex complex was diluted in Nuclease-free Duplex buffer at a final concentration of 1 µM. Two duplex mixes were prepared and heated at 95 °C for 5 minutes, then cooled down at room temperature. Mix composition as follows:

5' Duplex	**3' Duplex**
2 µl 100 µM 5' DHFR2 crRNA	2 µl 100 µM 3' DHFR2 crRNA
2 µl 100 µM tracrRNA	2 µl 100 µM tracrRNA
196 µl Nuclease-free Duplex buffer	196 µl Nuclease-free Duplex buffer

The HiFi Cas 9 was provided at a 61 µM concentration. To obtain a 1 µM working concentration, the enzyme was resuspended in Opti-MEM in a final volume of 400 µl. The final step sees the joining of the enzyme with the crRNA:tracrRNA duplex. The relative amount of the reagents was defined in relation to the size of the well. The protocol is optimised for transfection on 96-well plates but, to save time, we performed the transfection on 100mm dishes, sure of the efficiency of the method (previous trial experiments were performed to assess the effectiveness of the whole knockout experiment). The surface of a 100 mm dish is approximately 128-fold bigger than a single well of a 96-well plate. Considering that, all reagents volumes were adapted accordingly (see below). The RNP mixes were incubated for 5 minutes at RT.

5' DHFR2 RNP	**3' DHFR2 RNP**
192 µl 1µM 5' Duplex	192 µl 1µM 3' Duplex
192 µl 1µM HiFi Cas 9	192 µl 1µM HiFi Cas 9
2816 µl OptiMEM	2816 µl OptiMEM

3.3 Transfection of CRISPR/Cas9 ribonucleoprotein (RNP)

3.3.1 Preliminary steps

HepG2 cells were used to obtain a DHFR2 knockout. Cells were recovered from long term storage (Chapter 2, Section 2.2.13.3) and grown for at least four passages before being tested for Mycoplasma infections (Chapter 2, Section 2.2.14). Cells were grown in DMEM hi-glucose media supplemented with 10 % FBS and maintained in a 5 % CO_2 incubator at 37 °C.

3.3.2 Reverse transfection

Reverse transfection is a way to transfect genetic material into cells by seeding the cells on top of the transfection reagent. Reverse transfection requires a shorter time than the more traditional forward transfection, in which cells need to be seeded a day before transfection.

Transfecting an entire 100 mm dish required large volumes of reagents. Notably, 3.2 ml of RNPs was necessary for the transfection. The whole volume was reached by adding 1.6 ml of 5' DHFR2 RNP and 1.6 ml of 3' DHFR2 RNP. These reagents were also added with 153.6 µl Lipofectamine RNAiMax transfection agent and 3046.4 µl OptiMEM. The mixture was incubated at RT for 20 minutes. The entire volume (6.4 ml) was spread onto a dish, topped by 12.8 ml HepG2 cells resuspended at a concentration of 400,000 cells/ml in complete media (DMEM + 10 % FBS). Cells were incubated for 48 hours in a 5 % CO_2 incubator at 37 °C. The same amount of cells was seeded and grown in parallel as a negative control. Media was replaced every second day, allowing the cells to fully grow and recover from the toxic action of the lipofectamine. They reached full confluence in 10 days. Cell count was performed as described in Materials and Methods (Chapter 2, Section 2.2.13.2). An aliquot of 5 x 10^6 cells was collected to perform DNA extraction. Also, a 10 ml cell suspension (1 x 10^4 cells/ml) was prepared for the single-cell serial dilution.

3.3.3 Single-cell cloning

Single-cell serial dilution was carried out following the Corning protocol "Cell cloning by serial dilution in 96 well plates" (Ryan, n.d.) (Chapter 2, Section 2.2.13.4). All cell populations were screened for DHFR2 deletion by PCR assay.

3.4 Multi-population screening for identification of positive transformants

The DHFR2 knock out strategy described in this chapter is quite immediate in its application. Still, contrarily to most transfection methodologies, this one doesn't exploit any selection marker to make the screening of the knock out cell lines more facile. Therefore, the screening process is entirely based on the PCR detection of the ORF deletion. DNA extraction cannot be performed on a low concentration of cells, which brings to the necessity to grow cells into large populations fully. A preliminary screening method was attempted to avoid growing cells that did not receive the intended knockout. It required no DNA extraction and could be applied to small volumes of cells. After confirming the intended DHFR2 deletion, the positive lines could be kept growing and further tested for genetic rearrangement.

3.4.1 Direct PCR of genomic DNA from mammalian cells

The direct PCR methodology was inspired by colony PCR, a method usually performed on bacteria. Similarly, a little aliquot of growing cells was treated and used directly in a PCR reaction as a source of DNA. The proof of concept was obtained working on 96 well plate transfected cells (HepG2 DHFR2 Knock out). Cells were detached by trypsinisation and collected by centrifugation at 15,000 x g for 1 minute. The supernatant was discarded, and cells were either resuspended in 50 µl of molecular grade water or Tris buffer (10 mM, pH 8.5). The cell suspension was boiled at 99 °C for 10 minutes, followed by an ice shock for 1 minute and centrifugation at 15,000 x g for 1 minute. The resulting supernatant, containing cellular DNA, was directly used in PCR. With water or Tris, both preparations were tested using either Taq DNA Pol or Q5 DNA Pol in a 25 µl reaction.

The details of the PCR can be found in Appendix T. The preliminary data showed that the DHFR2 deletion band was present in all samples, albeit requiring optimisation. Because of time constraints, the method was not fully employed to screen the populations, but it proved to be an excellent approach to simplify the screening of multiple populations.

3.4.2 PCR confirmation of the DHFR2 ORF deletion

The polyclonal line (post-transfection) and fourteen cell populations, derived from single-cell cloning, were kept growing until ready for DNA extraction (DNeasy Blood and Tissue kit, Qiagen). A PCR assay, called DHFR2 Deletion Detection assay (Appendix U), was used to assess the occurrence of the designed deletion over the DHFR2 gene. In the event of a double cut, an entire portion of the gene will be lost, leading to a 1797-bp deletion. The primers strategically bind to both deletion flanking regions, allowing the amplification of the deletion area. This assay was designed to be a diagnostic test, showing a 2250-bp band if the knockout did not occur and a 453-bp band in case of a successful deletion (Fig. 3.2).

3.4.2.1 Methods

The column-extracted DNA was used to perform the DHFR2 deletion detection assay as follows: 10X Taq buffer, 2.5 µl dNTPs 2.5 mM, 2 µl $MgCl_2$, 0.5 µl of each primer DHFR2 Fb and DHFR2 Re (both 10 µM), 1 µl DNA (100 ng/µl), 0.25 µl Taq DNA Pol (5U/µl) and H_2O to a final volume of 25 µl. The PCR was run at the following conditions: pre-incubation at 95 °C for 5 minutes, amplification for 35 cycles at 94 °C for 1 minute, 59.6 °C for 1 minute, 72 °C for 2.5 minutes, and final elongation at 72 °C for 10 minutes. PCR products were electrophoresed on a 1 % agarose gel.

3.4.2.2 Results

Different volumes of DNA extracted from the polyclonal cell line were tested after transfection, showing the method's high efficiency. The transfection was highly successful, as demonstrated by the band at 453 bp indicative of deletion. The polyclonal cell line also showed the amplification of the full-length DHFR2 ORF (non-deleted alleles) (Fig. 3.3). This test defined the detection limit for the negative transfectants to 1:10

dilution of genomic DNA (corresponding to 10 ng DNA per reaction). The PCR reaction naturally favours the amplification of a shorter amplicon (deletion), hence the importance of assessing the minimum DNA concentration at which the longer band (non-deletion) can also be amplified.

The same PCR assay was performed on all cell populations maturated upon serial dilution. Fourteen cell lines deriving from 1 and 2 cells were maintained. All monoclonal lines were screened first, showing all possible outcomes (Fig. 3.4).

Lines A1, A3 and B1 were heterozygous, revealing amplification of both deletion and non-deletion options. A4 was the only line not presenting the deletion band; therefore, DHFR2 ORF was not excised. Finally, A2 showed amplification of the 453bp band alone, indicating the DHFR2 knockout on both alleles.

3.4.3 RT-PCR confirmation of the DHFR2 ORF deletion

The monoclonal cell lines were also tested for the presence of DHFR2 RNA. This assay was utilised as a confirmation step to prove the molecular phenotype of all monoclonal cell lines. DHFR2 presents two main isoforms, which differ just in their exon 1. Isoform DHFR2-201 and DHFR-202 share a small portion of their exon 1, with isoform 201 more extended at its exon 1 3' end. The 201-202 assay is designed to amplify both isoforms as having a forward primer binding to their common sequence on exon 1 and the reverse primer binding at the start of exon 2. The expected size of the amplicons is 308 bp and 185 bp for DHFR2-201 and DHFR2-202, respectively. The primers used are highlighted and described in Appendix V.

3.4.3.1 Methods

RNA was extracted from the A1, A2, A3, A4 and B1 cell lines using the Qiagen RNeasy Mini Kit as described in section 2.2.16 (Chapter 2). The RNA integrity was assessed before proceeding with reverse transcription. A genomic contamination assay was also performed on the freshly made cDNA to validate the removal of any genomic carry-over (Chapter 2, Sections 2.2.19/20/21).

Niamh Bookey optimised the 201-202 assay. The reaction mix was composed of the following: 10X Taq buffer, 0.6 µl dNTPs 10 mM, 1 µl $MgCl_2$ 25 mM, 0.6 µl FPDHFRL1Var1&2 10 µM, 0.8 µl RPDHFRL1Var1&2 10 µM, 1.5 µl cDNA, 0.25 µl Taq DNA Pol (5U/µl), H_2O to a final volume of 25 µl. The samples were PCR amplified as follows: pre-incubation at 95 °C for 3 minutes, amplification for 35 cycles at 94 °C for 30 seconds, 53 °C for 1 minute, 72 °C for 1 minute, and final elongation at 72 °C for 10 minutes. PCR products are separated through gel electrophoresis (Chapter 2, Section 2.2.8).

3.4.3.2 Results

Figure 3.5 serves as additional evidence of the molecular phenotypes of DHFR2 knock out monoclonal lines. Line A4, which didn't show any sign of ORF deletion in the PCR, revealed two intense bands corresponding to the two main DHFR2 isoforms, confirming that the gene has not received an ORF deletion. A1, A3 and B1 lines, classified as heterozygous, displayed very faint bands. Although the RT-PCR is not a quantitative method, it can be speculated that the presence of just one functioning allele diminishes the total amount of DHFR2 RNAs. Finally, the A2 line does not show amplification, validating it as a homozygous recessive line for DHFR2. According to the PCR and RT-PCR results, A2 was the chosen monoclonal line to carry out the functional experiments described in Chapter 5.

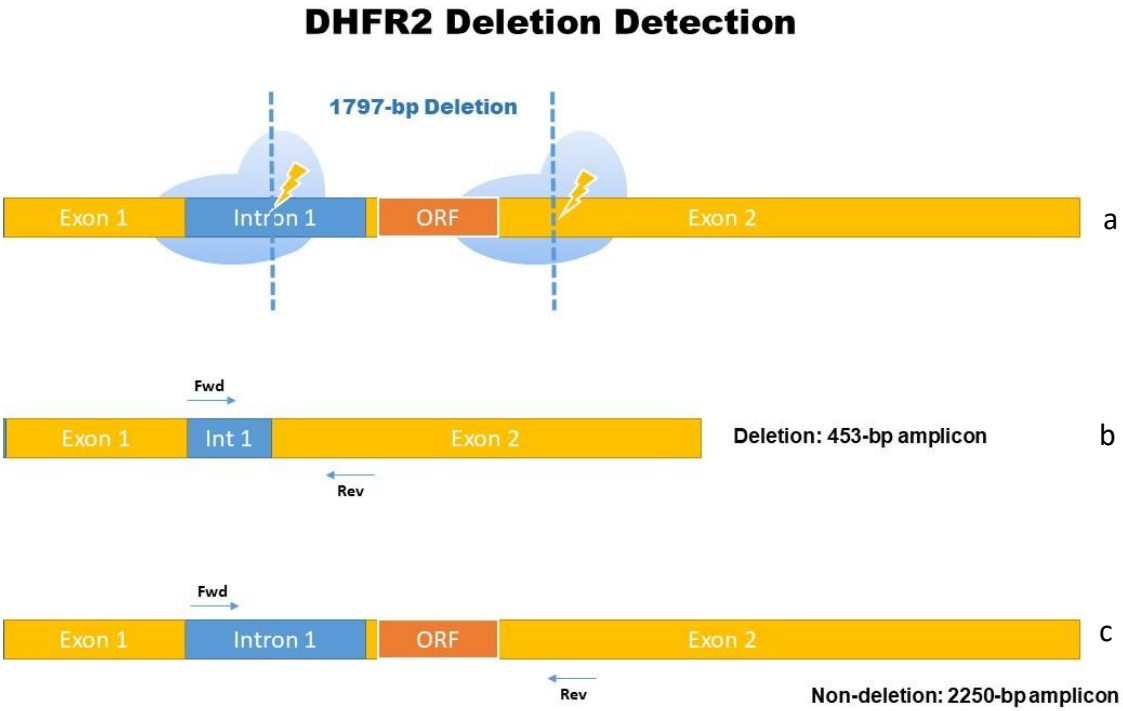


Figure 3.2 Schematic representation of the DHFR2 ORF deletion sites and PCR detection. a) The cleavage sites are located within intron 1 and exon 2. The successful cutting of both sites results in the deletion of 1797 bp inclusive of the entire DHFR2 ORF. b & c) Using primers sitting just outside the deletion sites, two different outcomes are possible: successful deletion resulting in the amplification of a 453 bp fragment (b); unsuccessful deletion resulting in the amplification of a 2250 bp fragment (c). Image created with BioRender.com.

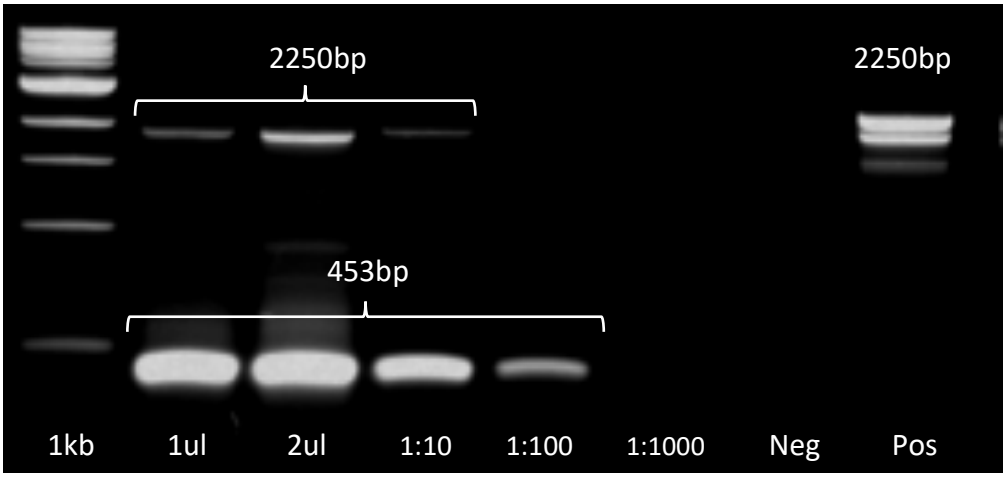


Figure 3.3 DHFR2 deletion detection PCR assay on polyclonal line. gDNA serial dilution test to assess the detection limit for both bands. Resolution on a 1 % (w/v) agarose gel. 1kb ladder on the left. Genomic DNA extracted from wild-type HepG2 line was used as positive control. Negative control contains no DNA. Different final concentrations of genomic DNA from the DHFR2 KO polyclonal line were tested. The bands at 453 bp are indicative of successful DHFR2 ORF deletion. The bands at 2250 bp indicate unsuccessful DHFR2 ORF deletion.

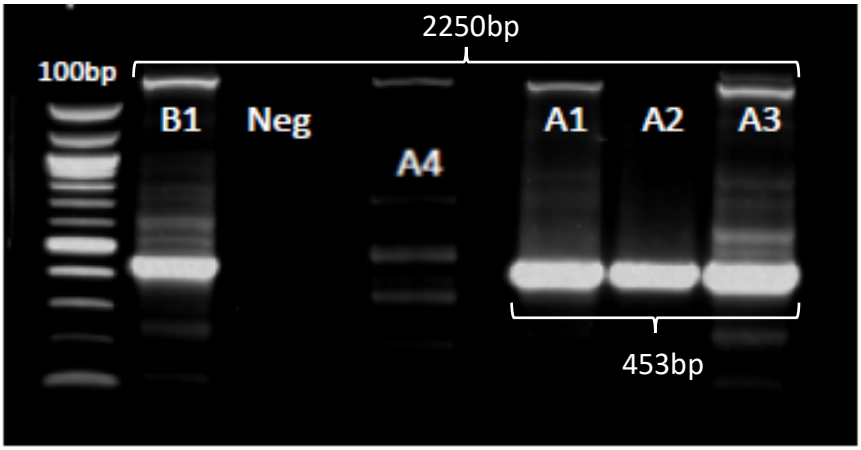


Figure 3.4 DHFR2 deletion detection PCR assay on monoclonal lines. Resolution on a 1 % (w/v) agarose gel. 100 bp ladder on the left. Negative control contains no DNA. Genomic DNA extracted from five monoclonal lines was tested. Lines B1, A1 and A3 are heterozygous for DHFR2 deletion, showing bands at both 2250 bp and 453 bp, with the latter indicating the successful deletion of the DHFR2 ORF. Line A4 shows the 2250 bp band corresponding to unsuccessful knockout, with no additional band indicating ORF deletion. Line A2 only presents the 453 bp band, indicating a successful deletion on both alleles.

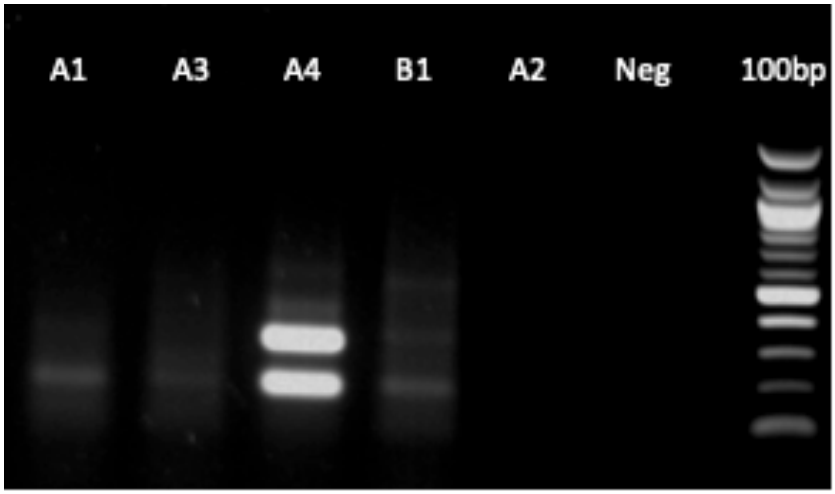


Figure 3.5 RT-PCR DHFR2 201-202 assay on the five DHFR2 KO monoclonal lines. Resolution on a 1 % (w/v) agarose gel. 100 bp ladder on the left. Negative control contains no DNA. RNA was extracted from the five DHFR2 KO monoclonal lines, reverse-transcribed into cDNA and tested using the 201-202 assay. Line A4 displays two bright bands corresponding to DHFR2 isoforms 201 and 202. The two bands are also present in the B1 line at a lower intensity. Lines A1 and A3 display only the 185bp band, corresponding to isoform 202. Line A2 shows no amplification, thus confirming the presence of a bi-allelic DHFR2 ORF deletion.

3.5 Discussion

The knock out strategy discussed in this chapter is based on the 2015 paper "Generation of Genomic Deletions in Mammalian Cell Lines via CRISPR/Cas9" (Bauer et al., 2015). The paper describes the feasibility and facility of producing large deletions using a CRISPR/Cas genome engineering method by inducing two DSBs at the extremities of the gene targeted for deletion. Bauer and colleagues assembled two plasmids harbouring the specific gRNA sequences and Cas9 gene. Cells were transformed by electroporation, sorted into 96 well plates by FACS and finally screened by PCR. The results are auspicious, showing how even substantial deletions can be accomplished by this method (Bauer et al., 2015).

We combined the rationale behind this study with a more efficient delivery element, such as the ribonucleoproteins (Burger et al., 2016; Gagnon et al., 2014; Hoshijima et al., 2019; Vakulskas et al., 2018). Introducing not one but two cuts to delete a piece of DNA is ingenious, but the use of plasmids as gRNA/Cas9 carriers could decrease the overall performance of the method (Chandrasekaran et al., 2018; S. Kim et al., 2014; Liang et al., 2015; Xu et al., 2021), as explained in Section 3.1.3. Using adequately designed RNPs, potential restrictions of plasmid delivery were overcome, and the overall strategy was brought to much higher efficiency.

The double deletion approach is a very effective and straightforward method (Bauer et al., 2015; Schmieder et al., 2018; Treuren & Vishwanatha, 2018), which, however, has to consider the structure and sequence of the target gene. DHFR2 retrogene layout, with the entire ORF harboured within one exon (human DHFR2 gene, *ENSG00000178700*) (Cunningham et al., 2022) and two accessible PAM sites, conveniently located at both sides of the ORF itself, made the application of this knockout method accessible. Furthermore, the repair of the 1.8 kb resulting deletion can be left to the NHEJ system, which does not require a DNA template to seal the gap (Gallagher & Haber, 2018; Pannunzio et al., 2018; Symington & Gautier, 2011). This implies that the overall design is rapid and straightforward since choosing the PAM sites (and relative gRNAs) is the only variable that requires planning.

The simplicity of the method, however, came with a cost. The lack of a selection marker, such as an antibiotic resistance gene or a fluorescent marker, made the screening process the longest part of the procedure. Many single cells had to be grown into relatively large monoclonal populations before they could be tested for the intended mutation. In fact, as reported by Foster et al. (2018), the use of a marker-less strategy becomes suitable when the targeted gene is associated with an easily identifiable phenotype (Foster et al., 2018). Even though our work did not include the use of organisms, the molecular phenotype (DNA, RNA) could be easily screened. This, in conjunction with the high editing frequency, enabled the selection of a monoclonal line without the use of a canonical selection marker (Agudelo et al., 2017; R. M. Martin et al., 2019). Direct PCR of genomic DNA showed the potential to ease the screening process by identifying potentially positive transformants in advance. With this method, negative transformants could be spotted in the early stage of the process and eliminated. Direct PCR could broaden the chance of focusing only on the positive samples (heterozygous or homozygous), thus redirecting the resources (time and effort) onto the worthwhile lines.

An additional limitation of the screening process was the use of the single-cell cloning method. This way of manually seeding cells at very low concentrations in the hope of getting enough wells filled with just one cell is not the most efficient (Morley et al., 1983; Staszewski, 1990). On a total of 96 wells, an average of 5 wells contained a single cell, meaning the necessity to repeat the procedure multiple times in case of a negative outcome, both in terms of the number of monoclonal lines and associated positive transformants. The low probability of selecting single cells, associated with the human error rate -due to the difficulty of observing an individual cell at the microscope- (Soitu et al., 2020) invites preferring an alternative method, such as FACS (Gross et al., 2015; Herzenberg et al., 2002; S. Liu, 2020).

The success of the ORF deletion was evaluated via PCR, with an assay able to discern between the intact and deleted form of the *DHFR2* gene. This PCR assay also gave insight about mono- and biallelic knockouts, thus enabling the identification of homozygous and heterozygous lines. Despite the assay working quite well, in the condition of heterozygosity, it existed a higher chance of reduced amplification of the longer

amplicon (i.e. failed ORF deletion). The reason for this phenomenon is the competition between a short and a long template when competing for amplification in the same polymerase reaction, with a shorter amplicon favoured versus a larger one (Kleter et al., 1998; Suzuki & Giovannoni, 1996). An additional PCR assay could be integrated into the screening process to avoid an uneven representation of the two alleles. The supplementary assay should amplify internally to the deletion area so that only the alleles with an intact ORF would generate an amplicon. Performing these two assays in parallel would enhance the possibility of a correct screening of the monoclonal lines.

In summary, the double-cut deletion strategy associated with an RNP delivery system proves to be a simple and time-effective method. Indeed, some adjustments could be made to improve the screening process but, overall, this genome-editing technique offers an extremely valid option to quickly obtain a permanently edited cell line.

To conclude, the DHFR2 knockout line represents a valuable tool for exploring DHFR2 function. The functional analyses deployed to screen the DHFR2 knockout line will be described in Chapter 5, while the following Chapter (4) will illustrate the CRISPR/Cas strategy used to establish a DHFR knock-down line.

Chapter 4
An MMEJ-assisted CRISPR/Cas approach to generate a DHFR knock-down in HepG2 cells

4.1 Introduction

4.1.1 Overview

DHFR is a critical component of One-Carbon Metabolism. It is the only entry point for dietary folate to the metabolic cycle through its conversion into THF (Bailey & Ayling, 2009). Over the past decade, a second active Dihydrofolate Reductase-like gene (*DHFR2*) was discovered, and a recombinant protein was constructed (McEntee et al., 2011). So far, the current literature does not report any clear identification of native DHFR2. Investigating DHFR2 properties, translatability and function remain a significant area of interest. That is especially important in the light of a possible DHFR/DHFR2 co-regulation and/or One-Carbon Metabolism modulation.

The *DHFR2* open reading frame produces a protein with 92 % homology to DHFR. The remarkable amino acid similarity between DHFR and DHFR2 brings along a series of technical challenges that have hindered a complete characterisation of the DHFR2 endogenous protein. The difficulty of explicitly targeting the protein with standard techniques led us to attempt to generate a DHFR KO cell line to approach this problem. The permanent elimination of the DHFR gene and protein provides the ideal condition to investigate DHFR2 presence, activity, location and relation with OCM comprehensively. Another crucial aspect that can be explored using the knockout line is the DHFR2 capability to compensate for the lack of DHFR, as commonly featured by other OCM enzymes (Anderson & Stover, 2009; Krupenko et al., 2015; Nilsson et al., 2019). In general, an established DHFR knockout cell line represents a powerful tool in advancing the research around the yet unanswered questions about DHFR2 function and relation with DHFR and OCM at large.

The chosen genome-editing technique for seeking a DHFR knockout relied on CRISPR/Cas. The editing strategy, called PITCh, was designed to achieve the target gene's knockout by inserting a gene cassette. The cassette, carrying an antibiotic resistance gene and a fluorescent protein gene, facilitated the selection of the correctly-edited cells. This method was chosen over several others because of the intelligent design, easy selection and favourable location of the CRISPR/Cas target sites on DHFR.

4.1.2 The PITCh system

Genome engineering by programmable nucleases facilitates the production of many different genomic rearrangements, such as knockouts and knock-ins. Depending on the type of intended mutation and adopted strategy, a preferential DNA repair system must be exploited to reach the desired outcome.

A gene knockout is a genetic modification in which a gene is made dysfunctional. Knockouts can be obtained through different techniques. Historically, HDR was mainly used, with the first applications dating back to the eighties (Capecchi, 2005). It employed a DNA segment containing the desired modification, typically in combination with a selection marker (Brown, 2002). The construct also had to carry a homology section (6-14 kb) to facilitate the recombination and thus the substitution of the target gene. However, this strategy appeared inefficient, with favourable recombinational events occurring at a meagre rate (Hall et al., 2009).

Knockout strategies based on the NHEJ repair system seem to be more successful. In the occurrence of a DSB, the NHEJ mechanism is the principal pathway to be activated throughout the entire cell cycle. Being essentially an emergency repair system, the main objective of NHEJ is to re-join the loose extremities at the cost of introducing frameshift mutations (Chang et al., 2017; Gaj et al., 2013).

A knockout strategy based on NHEJ repair is highly effective but screening the lines would be time-consuming. Contrarily, in the homologous recombination-based method, the possibility of a knockout via a selection marker's introduction represents an advantage in the screening step but producing a DNA construct with long homology arms is rather laborious.

In 2014, a Japanese group (Nakade et al., 2014) developed a novel knockout system called PITCh, Precision Integration into Target Chromosome, which exploits the Microhomology-mediated End Joining repair pathway. This method was presented as a win-win strategy to maintain a certain level of precision with a facilitated design, as it tried to merge the pros of HDR and NHEJ-based methods into a brand new way. As explained in section 3.1.4 (Chapter 3), MMEJ uses short homologies to repair DSBs, with a certain deletion level. This feature would be perfect to exploit when a selection marker

is used to disrupt a native gene. That was accomplished by the Suzuki & Yamamoto group, who published *"MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems"* in 2016 (Sakuma et al., 2016). They described a protocol to insert an EGFP-2A-Puro cassette into a gene locus in HEK293T cells, therefore obtaining the knock out of a gene and the simultaneous insertion of a double selective marker. The PITCh system uses programmable nucleases to produce DSBs both in the target chromosome and the donor vector; with the latter presenting short homologous arms that are used by MMEJ to repair the scission in the genomic DNA, leading to a precise insertion of the donor vector into the chromosome (Fig. 4.1) (Sakuma et al., 2016). For this procedure, two plasmids have been made: a donor vector containing the gene cassette flanked by the desired microhomologies, and a nuclease vector, containing the Cas9 gene and two gRNAs (one for the target gene and one for the gene cassette).

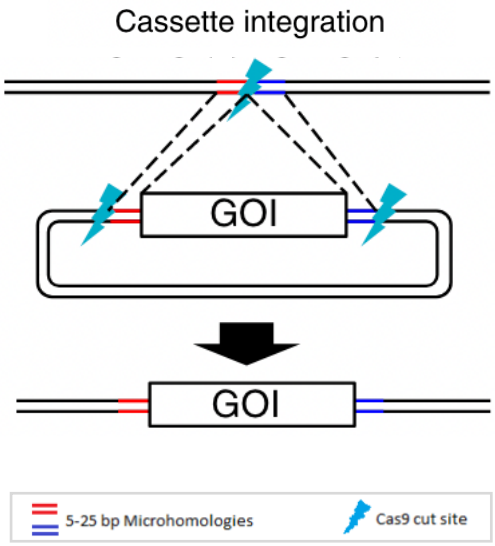


Figure 4.1 Diagram of the cassette insertion according to the PITCh system. CRISPR/Cas-mediated cuts are introduced both in the intended genomic site and at the edges of the gene cassette. Microhomologies, specifically designed at either side of the cassette, are exploited by MMEJ to repair the DSB and simultaneously integrate the cassette. Image adapted from Sakuma et al., 2016.

4.1.3 Application of the PITCh strategy for DHFR gene editing

The PITCh strategy was adopted to produce a DHFR knockout in Hep G2 cells. This gene editing experiment aimed to disrupt DHFR function by introducing a gene cassette to facilitate the selection and screening of the mutated population. The Sakuma protocol (Sakuma et al., 2016) was adapted to allow the coupling of the PITCh design with the CRISPR/Cas technology.

The four plasmids described in the protocol were publicly available from Addgene. They were used as backbones to construct donor and nuclease vectors for the DHFR knockout experiment. The donor vector will henceforth be referred to as "μΩ vector". It contained

an EGFP-2A-Puro cassette flanked by 5' and 3' microhomologies for the DHFR locus and target sites for PITCh-gRNA. The nuclease vector, called from now on "all-in-one vector", contained the Cas9 gene, preceded by a PITCh-gRNA (targeting the edges of the gene cassette) and a second gRNA –DHFR-gRNA– (targeting DHFR).

Upon plasmid transfection in HepG2 cells, the all-in-one vector was expressed to give Cas9 enzyme assembling with either PITCh-gRNA or DHFR-gRNA, thus introducing cuts in both the *DHFR* gene and the gene cassette extremities. Once excised from the plasmid, the MMEJ system used the EGFP-2A-Puro cassette to repair the DSB on DHFR, thanks to the microhomologies flanking the cassette itself (Fig. 4.2).

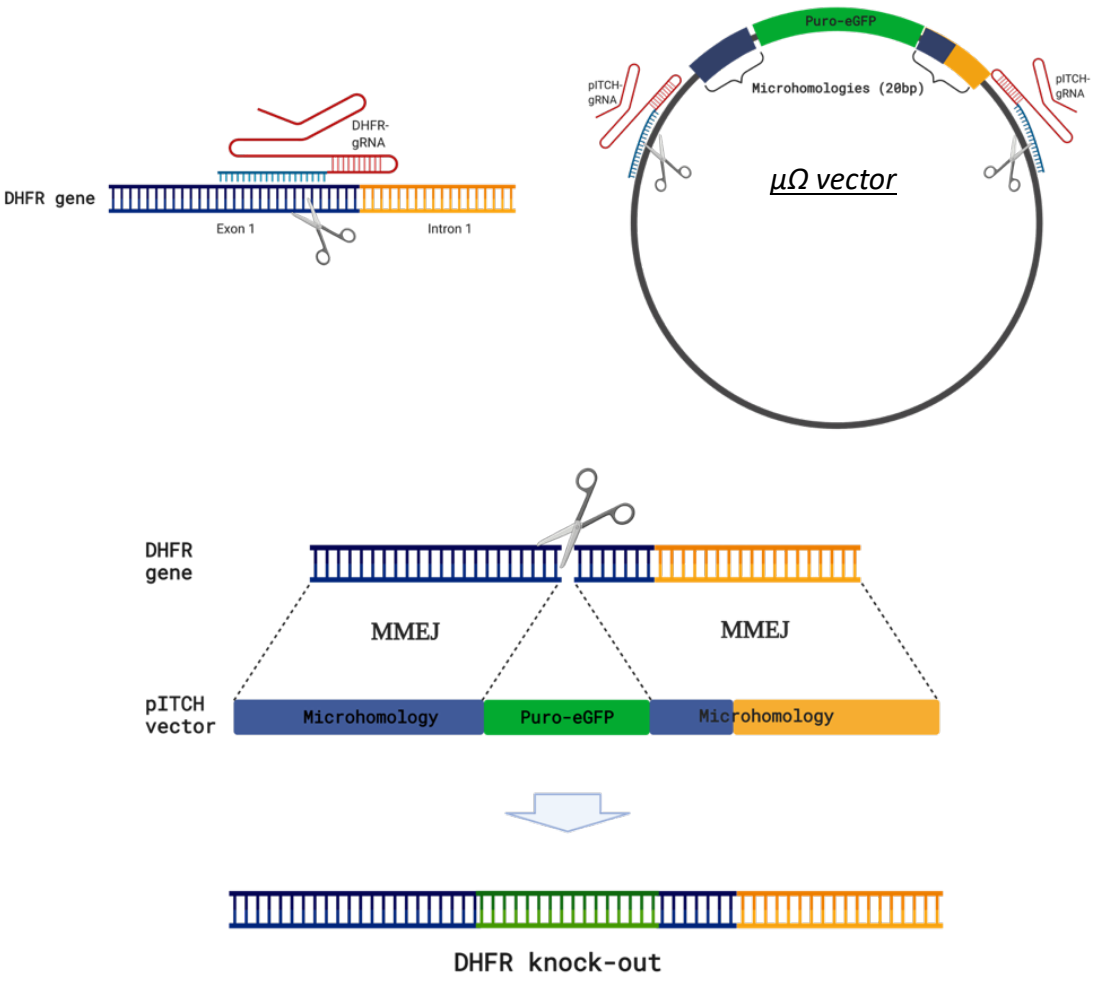


Figure 4.2 PITCh strategy applied to the DHFR knockout. DHFR-gRNA mediates the cut of DHFR at the end of exon 1 (top left). A PITCh-gRNA mediates the excision of the cassette out of the PITCh Vector. The microhomologies flanking the cassette are exploited by MMEJ to repair the DNA break. The insertion of the gene cassette into DHFR results also in the gene knockout. Image created with BioRender.com

4.2 Selection of suitable gRNAs and generation of recombinant clones (Methods)

4.2.1 Investigation and selection of DHFR suitable gRNAs

The *DHFR* gene sequence was thoroughly screened for suitable PAM sites, using two different online tools: Deskgen Guide Picker (*DESKGEN Cloud | Guide Picker*, n.d.) and PITCh designer 2.0 (*PITCh KnockIn | PITCh Designer 2.0*, n.d.). Our initial analysis was conducted considering the usage of SpCas9 (from *S.pyogenes*). It is important to notice that the PAM site is not included in the gRNA sequence, despite being essential to the recognition of the target site by the Cas9 enzyme. After a PAM sequence was identified in a site of interest (i.e. within the ORF), a gRNA could be designed to complement the target site (cleavage site). Each possible gRNA was assessed for on and off-target activity of the relative Cas9. The selected gRNA was 5' CGGCCCGGCAGATACCTGAG CGG 3' with the terminal CGG as PAM. The corresponding target site is located at the 3' end of *DHFR* exon 1. This specific gRNA was preferred as presenting the highest values of on- and off-target activity, respectively, 76 and 90, on a scale of 1 to 100 (Appendix BB include a full list of PAM sites). Note that the selected 20-nt gRNA sequence refers to the variable part of the gRNA, carrying specificity for the cleavage site. The full gRNA contains an invariable part that binds to the endonuclease (Fig. 4.3).

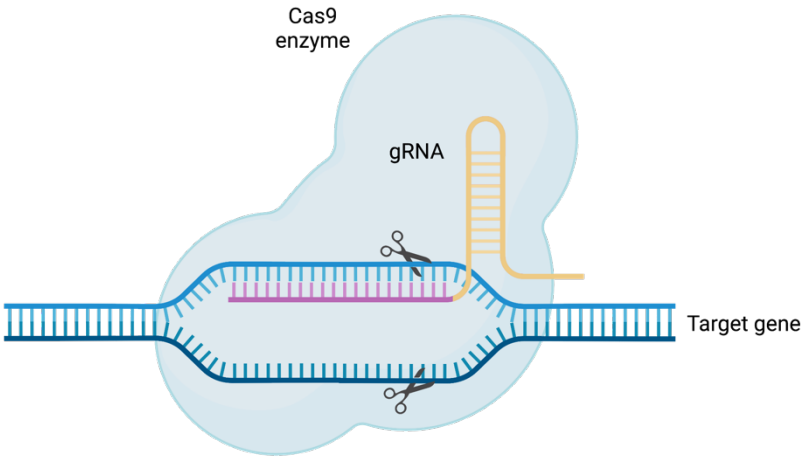


Figure 4.3 Representation of guide RNA (gRNA) structure and function. A gRNA is composed of a constant part (in yellow) which recognises the Cas9 enzyme, and a variable part (in pink), which anneals with the target DNA. This adaptable part is designed by the researcher. Image created with BioRender.com.

4.2.2 Adaptation of PITCh vectors to DHFR gene editing experiment (overview)

The All-in-one and μΩ vectors were obtained by molecular cloning techniques starting from four Addgene plasmids deposited from the Takashi Yamamoto lab: pX330A-1x2, pX330S-2-PITCh, pCRIS-PITChv2-FBL and pX330A-FBL/PITCh (Appendices A, B, C, D, respectively). The plasmids arrived as bacterial stabs. Bacteria were recovered by streaking and subsequent inoculation. Plasmids were extracted using a commercial kit and double digested to confirm their identity (Chapter 2, Section 2.2.5.1).

The All-in-One vector's construction necessitated two sequential molecular cloning procedures: the construction of a DHFR-specific gRNA and the assembly of DHFR-gRNA and PITCh-gRNA in a single plasmid. In the first step, the pX330A-1x2 was modified by inserting the variable part of the gRNA specific to *DHFR* into the vector. The gRNA 'gene' is composed of a constant part –already present in the pX330A-1x2 backbone– and a variable part specific to the target site (*DHFR*, in our case). In the second step, the Golden Gate Assembly method was required to excise the entire PITCh-gRNA out of pX330S-2-PITCh and ligate it downstream to the previously made DHFR-gRNA plasmid (intermediate plasmid). Therefore, the All-in-One vector, obtained by these two cloning steps, finally contained DHFR-gRNA, PITCh-gRNA and the Cas9 gene. Each one of these three genetic elements possesses its own promoter.

The development of the μΩ vector required the application of a single cloning technique. The microhomology sequences flanking the EGFP-2A-Puro cassette in pCRIS-PITChv2-FBL had to be replaced with DHFR-specific microhomologies. The details of the three methods are illustrated in the following sections.

4.2.3 Construction of the All-in-One vector – Intermediate Plasmid

The pX330A-1x2 plasmid harboured the Cas9 gene, a gRNA scaffold and an Ampicillin resistance gene (Appendix A). It was designed to enable the insertion of the targeting part of the gRNA, just upstream of the scaffold, but it was also structured to allow the cloning of a second complete gRNA downstream of the first one. In my design, the upstream gRNA was DHFR-specific, followed by a PITCh-gRNA designed to target the EGFP-2A-Puro cassette on the μΩ vector. The Ran protocol (Ran et al., 2013) was

followed to insert the DHFR-specific gRNA between the U6 promoter and the gRNA scaffold (Fig. 4.4). Note that an additional G was added at the 5' end of the insert because the U6 RNA Pol III promoter prefers a G as the first base of its transcripts. U6 contains two adjacent *Bbs I* restriction sites used to ligate the DHFR-gRNA when presenting with the correct overhangs. Two oligos containing the DHFR-gRNA sequence, one forward and one reverse, were designed with additional *Bbs I* overhangs and 5'-phosphorylated ends ("Top DHFR Insert" and "Bottom DHFR Insert")(Appendix E).

To prepare the insert, the two primers were resuspended at a final concentration of 100 µM. A volume of 1 µl of each primer was added to 1 µl of T4 DNA Ligase buffer 10X to a final volume of 10 µl. Using a thermal cycler, the reaction mixture was heated at 95 °C for 5 minutes and then cooled down to 25 °C, 5 °C per minute. In addition, pX330A-1x2 was digested with *Bbs I*, then run on a 1 % agarose gel to identify and separate the linearised from the undigested plasmid (Chapter 2, Sections 2.2.5.2 and 2.2.8). The band corresponding to the linearised plasmid was excised and purified using Wizard SV Gel and PCR Clean-Up System (Chapter 2, Section 2.2.9). The linearised plasmid and the DHFR-gRNA insert were then ligated using the QuickLigase kit and transformed into OneShot TOP10 cells (Chapter 2, Sections 2.2.6.1 and 2.2.7.1). Plasmids were isolated and Sanger sequenced by Source Bioscience (Chapter 2, Sections 2.2.2 and 2.2.11). "U6Fwd_Control DHFR insert" primer (Appendix E) was used to confirm the presence of the DHFR-gRNA in frame with promoter and gRNA scaffold.

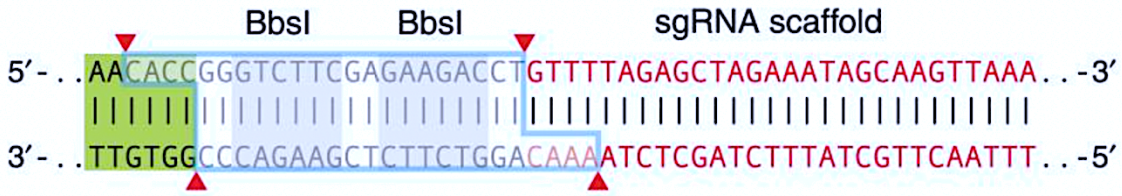


Figure 4.4 Double BbsI restriction site at U6 promoter 3' end. When both restriction sites are cleaved off, the desired gRNA, harbouring the correct overhangs, can be ligated in. The resulting construct will have in frame U6 promoter, gRNA target sequence and gRNA scaffold. Image adapted from Ran et al., 2013.

4.2.4 Construction of the All-in-One vector – Golden Gate Assembly

This second step aimed to obtain the All-in-One vector with the DHFR-gRNA, PITCh-gRNA and Cas9 in series. The second cycle of digestion/ligation was applied to both the intermediate plasmid and pX330S-2-PITCh. The PITCh-gRNA (443bp), contained in the pX330S-2-PITCh plasmid (Appendix B), could be easily excised by an enzymatic cut catalysed by *Bsa I – HF.* The same restriction enzyme was used to create a gap (two cuts) downstream of the DHFR-gRNA on the intermediate plasmid to accommodate the PITCh-gRNA (Chapter 2, Section 2.2.5.2). These cleavage sites are located between LacZα and its promoter, thus leading to their excision (459bp, positions 8503 and 8962) (Fig. 4.5). Both plasmids were digested and ligated in the same reaction mix, which was later used to transform XL10 Gold cells (Fig. 4.6) (Chapter 2, Sections 2.2.6.3 and 2.2.7.2). These *E.coli* cells allowed the Blue/White selection when grown with IPTG and X-Gal. Blue colonies were indicative of the original undigested plasmid, as the LacZα gene was still active and able to metabolise X-Gal, which eventually oxidises into a blue molecule. Contrarily, the white colonies could either contain the re-ligated DHFR-gRNA plasmid lacking the LacZα gene or the DHFR-gRNA plasmid containing the PITCh-gRNA. Transformants with just pX330S-2-PITCh were negatively selected by ampicillin, as they are streptomycin-resistant. To define which white colonies had the correct insert, twelve CFUs were screened by colony PCR. Each selected colony was also used to prepare a replica plate. MyTaq HS DNA Pol (Bioline) was used as polymerase suitable for colony PCR. The reaction mix was as follows: 1 µl MyTaq HS DNA Pol (5U/µl), 10 µl 5X MyTaq reaction buffer, 2 µl CRISPR-step2-Fwd (10 mM), 2 µl CRISPR-step2-Rev (10 mM), H_2O to a final volume of 50 µl. The PCR was run at the following conditions: pre-

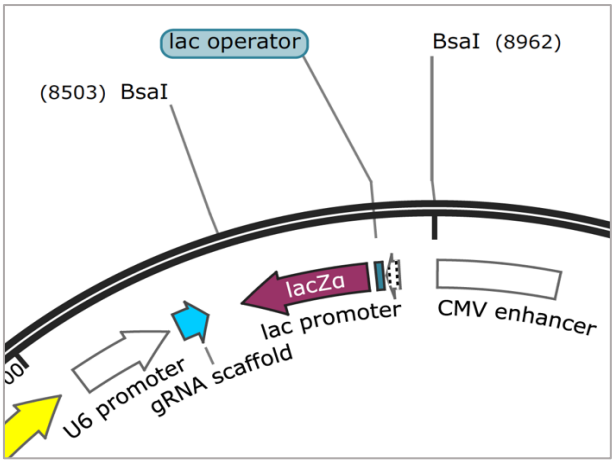


Figure 4.5 Section of the intermediate plasmid (previously pX330A-1x2), indicating the location of the Lac operon and Bsa I restriction sites. The Bsa I restriction sites are used to eliminate the Lac operon (made of lacZα gene, lac operator and promoter), and insert an additional gRNA. The extra gRNA will be located downstream the constitutive gRNA (U6 promoter and gRNA scaffold). Image created with SnapGene Viewer.

incubation at 95 °C for 1 minute, amplification for 28 cycles at 95 °C for 15 seconds, 67 °C for 15 seconds, 72 °C for 50 seconds, and final elongation at 72 °C for 50 seconds. The PCR products were run on a 1% agarose gel (Chapter 2, Section 2.2.8), confirming that all amplicons had the intended product, which resolved in a 900 bp band. For further confirmation, four of the twelve tested colonies were prepped for plasmid isolation and sent to Sanger sequence (Chapter 2, Sections 2.2.2 and 2.2.11), adopting the same primers of the PCR reaction (Appendix E).

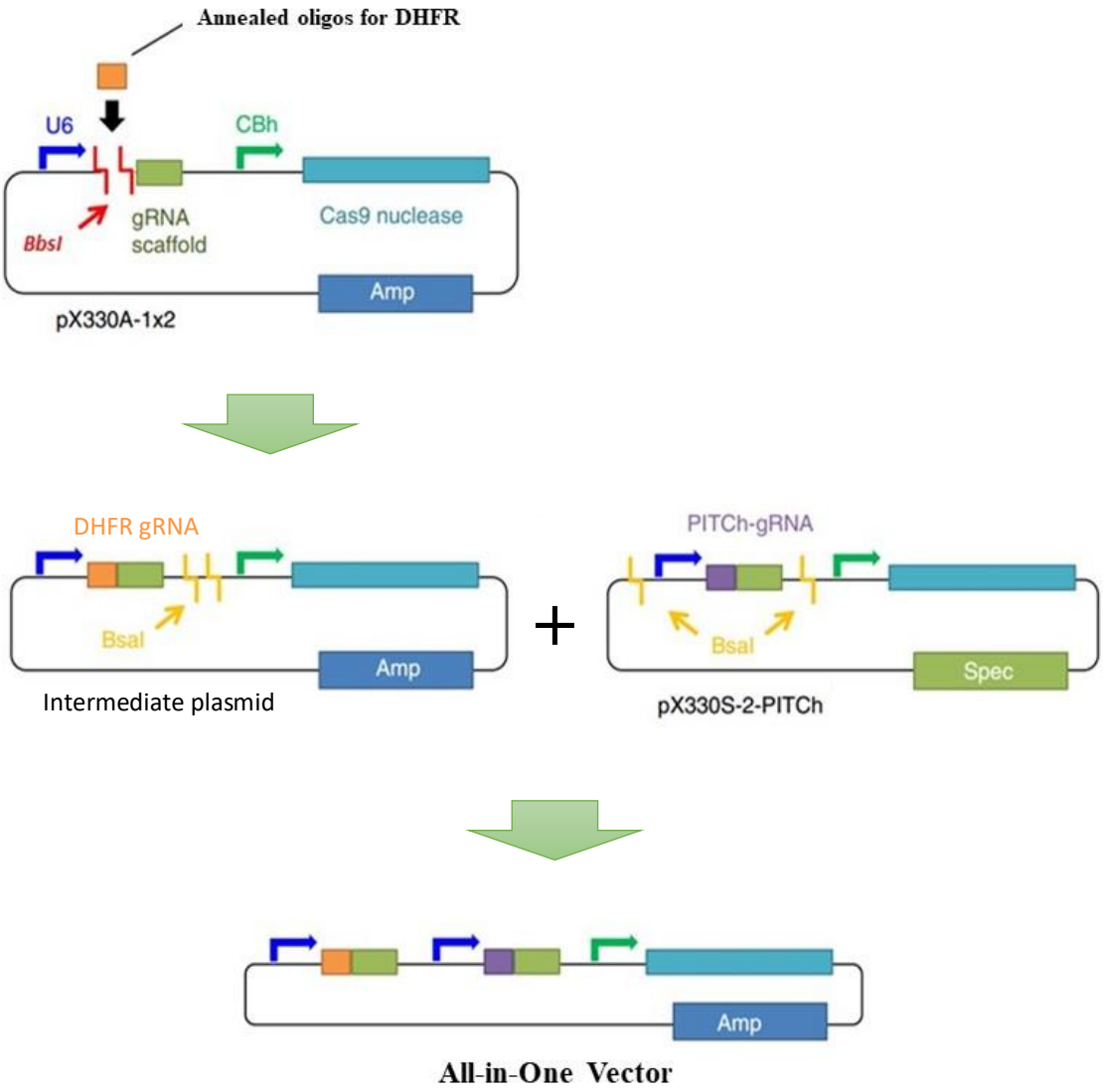


Figure 4.6 Construction of All-in-One vector by Golden Gate Assembly. After pX330A-1x2 was modified by inserting the DHFR gRNA (thus creating the intermediate plasmid), the PITCh gRNA was excised from the pX330S-2-PITCh vector and ligated into the intermediate plasmid. The resulting All-in-One Vector carried the DHFR gRNA, the PITCh gRNA and the Cas9 gene. Image adapted from Sakuma et al., 2016.

4.2.5 Construction of the μΩ vector (microhomology vector)

The pCRIS-PITChv2-FBL plasmid (Appendix C) harboured the EGFP-2A-Puro cassette flanked by FBL (fibrillarin) - specific microhomologies and PITCh-gRNA target sites. The FBL-specific microhomologies were substituted with DHFR-specific microhomologies, which correspond to the 5' and 3' extremities forming on *DHFR* after Cas9 cleaving action (Fig. 4.7). These targeted microhomologies facilitated the repair of DHFR double-strand break via MMEJ. To replace the current microhomologies on pCRIS-PITChv2-FBL, two separate PCR reactions were set: one to amplify the vector backbone and the other to amplify the cassette and, simultaneously, introduce the new DHFR homologous sequences.

The vector was amplified starting from two generic primers, 5#-rev DonorVector and 3#-fwd DonorVector (Appendix E). Instead, the second reaction required gene-specific primers that contained the desired DHFR microhomologies. These primers, 5Fwd DonorVect Mo and 3rev DonorVect Mo (Appendix E) were very long and prone to form secondary structures, making it necessary to use high annealing temperature and a robust polymerase (Q5 High-Fidelity DNA Pol). Both reactions, described below, presented the same reagents, except for the primer pairs. The PCR mix was composed of 1.25 µl pCRIS-PITChv2-FBL (1 ng/µl), 5X Q5 reaction buffer, 5X Q5 high GC Enhancer, 0.5 µl dNTPs (10 mM), 0.375 µl Q5 Pol (2000 units/ml), 2.5 µl of each primer 10 mM (5#-rev DonorVector, 3#-fwd DonorVector; 5Fwd DonorVect Mo, 3rev DonorVect Mo), and H_2O to a final volume of 25 µl. The cycling conditions were the following: pre-incubation at 94 °C for 2 minutes, amplification for 36 cycles at 98 °C for 10 seconds, 72 °C for 90 seconds, and final extension at 72 °C for 4 minutes.

The expected bands of ~1400 bp for the cassette and ~4600 bp for the vector backbone were confirmed by gel electrophoresis (Chapter 2, Section 2.2.8). The gel slices corresponding to the desired DNA fragments were excised with the help of a scalpel and purified using the NucleoSpin Gel and PCR Clean-up system (Chapter 2, Section 2.2.9). The vector backbone and the gene cassette, flanked now by DHFR microhomologies, were ligated together to restore the whole plasmid. An In-Fusion ligation was set, followed by transformation into Stellar competent cells (Chapter 2, Sections 2.2.6.2 and

2.2.7.3),. Four transformants were selected for plasmid isolation and sequencing (Chapter 2, Sections 2.2.2 and 2.2.11). Primers Contr DonorUpstr and Contr DonorDownst (Appendix E) were used to amplify the gene cassette's 5’ and 3’ end.

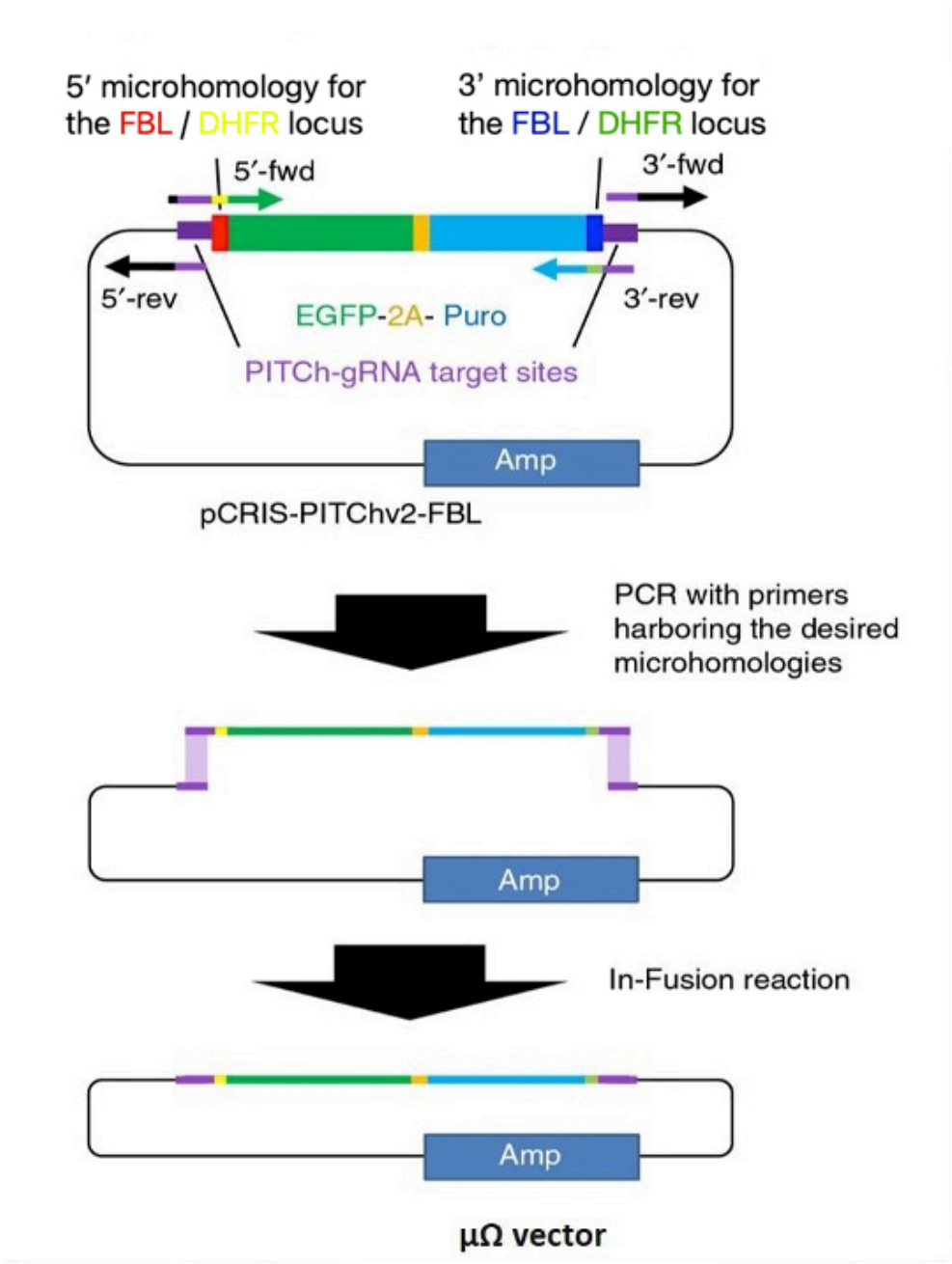


Figure 4.7 Construction of the μΩ vector via PCR and In-Fusion ligation. The pCRIS-PITChv2-FBL plasmid backbone was amplified with primers starting from the PITCh gRNA target sites outwards. On the other hand, the primers amplifying the gene cassette carried DHFR specific microhomologies, which substituted the FBL ones. The very long primers annealed with the PITCh gRNA target sites (in purple) and with a portion of the cassette (EGFP, in green, for the 5’ primer and Puromycin resistance gene, in light blue, for the 3’ primer). The central part of each primer contained the new DHFR microhomologies (in yellow, for the 5’ primer and in blue, for the 3’ primer). Once the two DNA segment were amplified, they were ligated together with an In-Fusion reaction. Image adapted from Sakuma et al., 2016.

4.3 Confirmation of the correct construction of the DHFR PITCh vectors (Results)

4.3.1 Insertion of the DHFR-specific gRNA in the intermediate plasmid

The annealing of the two 5' phosphorylated oligos "Top DHFR insert" and "Bottom DHFR insert" (Appendix E) resulted in the following insert:

```
5    C A C C g C G G C C C G G C A G A T A C C T G A G          3

3              C G C C G G G C C G T C T A T G G A C T C C A A A A   5
```

The correct insertion of the DHFR insert into pX330A-1x2 was confirmed by Sanger sequencing (Appendix H), thus generating the intermediate plasmid.

NNNNNNNNNNNCGANANNNGCTGTTAGAGAGATAATTGGAATTAATTTGACTGTAAACA
CAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCA
GTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTT
CGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGCGGCCCGGCAGATAC
CTGAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGA
AAAAGTGGCACCGAGTCGGTGC

Legend: U6promoter 21nt DHFR-gRNA gRNA scaffold

4.3.2 Sequential insertion of the DHFR gRNA and PITCh gRNA in the All-in-One vector

The sequencing results showed the correct insertion of the PITCh-gRNA into the intermediate plasmid, effectively creating the All-in-One vector. The sequence also showed the presence of the DHFR-gRNA, giving additional proof that the intermediate DHFR-gRNA plasmid was correctly built (Appendix I).

NNNNNNNNNNNTGATTCTTCNTATTTGCATATACGATACAAGGCTGTTAGAGAGATAAT
TGGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGT
AATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATG
CTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACG
AAACACCGCGGCCCGGCAGATACCTGAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATA
AGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTGTTTTAGA
GCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTTTTAGCGCGTGCGCCAATTCTGC
AGACAAATGGCTCTAGAGGCATGTGAGGGCCTATTTCCCATGATTCCTTCATATTTGCA
TATACGATACAAGGCTGTTAGAGAGATAATTGGAATTAATTTGACTGTAAACACAAAGA
TATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTA
AAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTT
CTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGCATCGTACGCGTACGTGTTGT
TTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG
GCACCGAGTCGGTGCTTTTTTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAG
TCCGTTTTTAGCGCGTGCGCCAATTCTGCAGACAAATGGCTCTAGAGGTACCCGTTACA
TAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTC
AATAGTAA

Legend: U6promoter DHFR-gRNA PITCh-gRNA gRNA scaffold

4.3.3 Substitution of the FBL 5’ and 3’ microhomologies with DHFR-specific sequences

The gene cassette could not be entirely amplified due to its length (1.4 kb). Still, it was possible to amplify the two extremities separately, ensuring that DHFR microhomologies immediately flanked the cassette with PITCh target sites outwards (Appendices J and K). The replacement of the 5’ microhomology sees the PITCh-gRNA target sites immediately followed by the 5’ DHFR Microhomology. A few base pairs downstream, the EGFP starting codon is shown. Sequencing results are as follows:

NNNNNNNNNNNNNNCGTTNANNTTGATTATTGACTAGTTATTAATAGTAATCAATTACG
GGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATA**GCATCGTACGCGTAC**
GTGTTTGG*GACCTGCCCTGGCCACCGCTCC*CCGGATCCATGGTGAGCAAGGGCGAGGAG
CTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAA
GTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGT
TCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACC
TACGGCGTGCAGTGCTTCAGCCGTTACCCCGACCACATGAAGCAGCACGACTTC

Legend: 5Fwd DonorVect Mo primer comprising

PITCh-gRNA target *5'DHFR microhomology* EGFP starting codon

Similarly, but in the opposite direction, the replacement of the 3' microhomology is composed of the end of the Puromycin resistance gene, the 3' DHFR microhomology and the PITCh target site. Sequencing results are as follows:

NNNNNNNNNNNNNNNGCGGCCGANCGCGCCGGGGTGCCCGCCTTCCTGGAGACCTCCG
CGCCCCGCAACCTCCCCTTCTACGAGCGGCTCGGCTTCACCGTCACCGCCGACGTCGAG
GTGCCCGAAGGACCGCGCACCTGGTGCATGACCCGCAAGCCCGGTGCCTGA*AGGTATCT*
*GCCGGGCCGGGG***CCAAACACGTACGCGT**ACGATGCTCTAGAATGCTGATGGGCTAGCAA
AATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCAT

Legend: 3rev DonorVect Mo (complementary sequence) comprising

End of Puro^{R} *3'DHFR microhomology* **PITCh gRNA target site**

4.4 DHFR gene editing via PITCh: transfection and selection of monoclonal cell lines (Methods and Results)

Creating a novel cell line model required thorough planning, spanning from the choice of the cell line to the gene-editing strategy, from the transfection system to the maintenance of the new line.

The cell line chosen to generate a DHFR knockout was HepG2. HepG2 are human hepatocellular carcinoma cells; they present an epithelial morphology and are suitable for transfection (*Hep G2 [HEPG2] ATCC® HB-8065™*, n.d.). HepG2 cells are non-tumorigenic, with a 48-hour doubling time. Its robustness, stability and easy handling make HepG2 one of the most widely used cell lines for *in vitro* experiments (Donato et al., 2015). Despite all these qualities, it has to be considered that the novel cell line will be deficient in DHFR. Previous studies, including McEntee, 2011 (McEntee et al., 2011), demonstrated that DHFR lacking cells could not survive unless supplemented with Hypoxanthine and Thymidine (HT) (Urlaub et al., 1983; Urlaub & Chasin, 1980). Because of that, HepG2 cells subjected to DHFR knockout were permanently supplemented with HT. The impact of HT on cell growth has been assessed before gene-editing, with the result that no significant variation in number was observed (Appendix S).

A diagram of the transfection and selection process is shown below, in Figure 4.8. Two populations were kept as a negative control, one of whom supplemented with HT. An FBL knockout was performed as a positive control, using the original plasmids purchased from Addgene (pCRIS-PITChv2-FBL and pX330A-FBL/PITCh). For the DHFR knockout, the All-in-One and μΩ vectors were transfected into HepG2 cells. Positive transformants were selected by puromycin. The resulting polyclonal population was screened via PCR and fluorescence observation. These same methods were employed to screen the selected monoclonal lines obtained from single-cell cloning.

4.4.1 Dose-response curve for puromycin selection of HepG2 cells

HepG2 cells were recovered from liquid nitrogen storage and grown for at least four passages. Before being transfected, cells were tested for Mycoplasma infection (Agilent

MycoSensor PCR assay kit). Unless otherwise stated, cells were grown into Sarstedt 100mm Dishes.

A puromycin kill curve was performed to assess the optimal puromycin concentration for drug selection (Chapter 2, Section 2.2.13.5). Cells treated with concentrations of puromycin below 1 μg/ml had no major impact on their growth. Puromycin was effective to kill the cells when above 1 μg/ml. Therefore, it was concluded that 1, 1.5, and 2 μg/ml are, respectively, low, optimal and high puromycin concentrations to use on HepG2 for drug selection.

4.4.2 Transfection of All-in-One and μΩ vectors in HepG2 cells

When exogenous DNA is artificially inserted into mammalian cells, it necessitates a high purity level. Hence, all plasmids used for transfection were isolated using Qiagen Plasmid Maxi Kit and cleaned from residual bacterial debris using Mirus MiraCLEAN Endotoxin Removal Kit (Chapter 2, Sections 2.2.3/4).

Two plasmids (All-in-One and μΩ vector, for DHFR knockout; pCRIS-PITChv2-FBL and pX330A-FBL/PITCh for FBL knockout -positive control-) were co-transfected (Chapter 2, Section 2.2.15) with the help of Mirus TransIT-X2, a low-toxicity, high-efficiency transfection reagent. Cells were passaged the day before transfection to be 60-70 % confluent the following day. A 6-well plate was prepared alongside to be used as negative control (untransfected). A 4:1 TransIT-X2:DNA ratio was used (4μl transfection reagent per μg plasmid) and 12 μg of total plasmid DNA, divided into 8 μg All-in-One vector and 4 μg μΩ vector (2:1 ratio). The identical amounts for the positive control were 8 μg pX330A-FBL/PITCh and 4 μg pCRIS-PITChv2-FBL. Transfection was performed in triplicates to allow each dish to be treated with a different concentration of puromycin: 1, 1.5 and 2 μg/ml. The transfection mix (quantities refer to a 100 mm dish) was prepared following the manufacturer's protocol. Warm OptiMEM (1.2 ml) was added to 12 μg plasmid DNA and 48 μl TransIT-X2 and incubated at RT for 30 minutes. Each dish was supplemented with the transfection mix, added dropwise, and incubated for 72 hours in a 5 % CO_2 incubator at 37 °C. Observation of the transfected dishes, 24 hours post-transfection, revealed an unexpected growth compared with the negative control. This increase in the cell population impacted drug selection.

4.4.3 Puromycin selection

Compared to the untransfected control, cells transfected with TransIT-X2 showed a rapid and unforeseen increase in proliferation. Puromycin effectiveness is directly linked to cell abundance; thus, it was supposed to not act effectively as a selection tool if cells were overnumbered. For this reason, cells were passaged 1:2, 48 hours after transfection. In contrast, cells transfected with FBL plasmids (positive control) were left growing.

Puromycin selection was commenced 72 hours after transfection, using different antibiotic concentrations. The positive control dishes were treated with 1, 1.5 and 2 µg/ml puromycin. Supplemented with puromycin, the media was replaced every two days for six days. Equal treatment was reserved for the 6-well plate negative control. As of DHFR transfection, the resulting six dishes (1:2 passaging) were treated as follows: two supplemented with HT and NEAA, and treated with puromycin 1 µg/ml; two supplemented with HT and NEAA, and treated with puromycin 2 µg/ml; two not supplemented with HT and NEAA, and treated with puromycin 1.5 µg/ml. Equal treatment was reserved for the 6-well plate negative control. As illustrated in Figure 4.9, three plates, one per treatment type, were supplied with fresh media and antibiotics every day; the other three had it replaced every second day, both groups for six days.

After 6 days of puromycin administration, cells from positive control were not affected by the antibiotic treatment, confirming that fully confluent cells were resistant to puromycin, even at a high dose (2 µg/ml). Contrarily, cells from DHFR transfection, which were split in half before the treatment, were responsive to differential puromycin dosage. At the end of the 6-days treatment, cell populations were widely different in number, as shown in Table 4.1

Cells treated with puromycin 2 µg/ml (both administration times) and 1.5 µg/ml (daily administration) were maintained and left recovering for four weeks, adding complete media regularly. The addition of HT/NEAA was continued only on cells initially treated with 2 µg/ml puromycin. Cells not supplemented with HT/NEAA had a granular morphology, less healthy than those receiving supplements.

4.4.4 Fluorescence Observation

Cells that survived puromycin treatment were expected to have the gene cassette integrated in frame. Besides giving resistance to puromycin, the gene cassette provided the presence of eGFP. Fluorescence observation on living cells was attempted using a Nikon Ti-E epifluorescence microscope. Only a few green signals could be observed in correspondence to large clumps of cells. Unfortunately, pictures of fluorescence could not be taken due to technical reasons.

4.4.5 EGFP-2A-Puro^R cassette in frame insertion confirmed via PCR

PCR confirmed the correct insertion of the EGFP-2A-Puro^R cassette. Total DNA was extracted from DHFR knock out polyclonal lines using the Qiagen DNeasy Blood & Tissue Kit and eluted in 200 µl H_2O (Chapter 2, Section 2.2.17). To avoid competitive amplification within the same assay, two different sets of primers were designed and used to separately amplify the 5' and 3' ends of the gene cassette embedded in the *DHFR* gene in frame with the gene promoter (Fig. 4.10).

To amplify the cassette 5' end, two primers called "DHFR KO Fwd" and "Invert cassette Rev" were used (in blue, in the above sequence block). The amplicon has an expected length of 151 bp. At the other extremity, the amplification of the 3' end was obtained by the use of "DHFR KO Rev" and "Invert cassette Fwd", with a resulting amplicon of 463 bp (Appendix F).

Both reaction mixes were set as follows: 0.5 µl Q5 Pol (2000U/ml), 5X Q5 buffer, 0.5 µl dNTPs 10mM, 1.25 µl Fwd primer 10 µM (either DHFR KO Fwd or Invert cassette Fwd), 1.25 µl Rev primer 10 µM (either Invert cassette Rev or DHFR KO Rev), 1 µl DNA (200 ng/µl) and H_2O to a final volume of 25 µl. The PCR was run at the following conditions: pre-incubation at 98 °C for 30 seconds, amplification for 35 cycles at 98 °C for 10 seconds, 64 °C (5'end assay) or 65 °C (3' assay) for 30 seconds, 72 °C for 40 seconds, and final elongation at 72 °C for 2 minutes.

4.4.5.1 Results

The 5’ and 3’ end assays performed on DNA extracted from cells initially treated with puromycin 2 µg/ml daily show the expected bands at 151 and 463 bp (Fig. 4.11). This confirms the successful integration of the EGFP-2A-Puro cassette on DHFR.

The assays were repeated on the other polyclonal population treated with puromycin 2 µg/ml alternate day, and they also amplified the expected bands at 151 bp and 463bp. Contrarily, the population treated with puromycin 1.5 µg/ml daily didn’t resolve any band at the 3’end, indicating a partial cassette insertion (not shown).

4.4.6 Single-cell cloning

Cells previously treated with 2 µg/ml puromycin daily were used to obtain monoclonal lines via single-cell cloning. The protocol for cell cloning by serial dilution in 96 well plates provided by Corning (Ryan, n.d.) was adopted (Chapter 2, Section 2.2.13.4). The top left well (A1) was filled with 200 µl of cell suspension, and 1:2 dilutions were performed from A1 downwards until all wells in the column had 100 µl of cells. The same type of dilution was performed to dilute cells from the first column to the last, using an 8-channel pipettor (Figure 4.12). It was possible to identify wells containing just a single cell after 5-6 days. Three weeks were required to allow single cells to grow enough to passage into a 24-well plate. A week after, cells were moved to a 12-well plate and finally (two weeks later) to a 100mm dish. Four monoclonal lines have been selected and further screened.

4.5 Multi-population screening for identification of positive transformants (Methods)

The four monoclonal lines, together with the polyclonal line, were screened to assess which of them harboured a biallelic knock out of the DHFR gene. Genomic DNA from all five populations was tested via PCR. In addition, RT-PCR was performed on cDNA reverse-transcribed from RNA extracted from all the samples. Finally, the presence and functionality of the DHFR enzyme were determined using Western blot and DHFR enzymatic activity assay, respectively.

4.5.1 Genomic screening via PCR

The assays used to screen the polyclonal line for the gene cassette insertion (Section 4.4.5) were also utilised to screen the monoclonal lines. The primers used to amplify the cassette at the 5' and 3' ends were designed to be exchangeable; that is, primer 'DHFR KO Fwd' and 'DHFR KO Rev' could be coupled together, effectively detecting any wild-type allele or a shorter insertion. The three PCR assays allowed me to screen for homozygous/heterozygous populations. A schematic of the three assays is shown in Figure 4.13.

Total DNA was extracted from DHFR knock out monoclonal lines using the Qiagen DNeasy Blood & Tissue Kit and eluted in 200 µl H_2O (Chapter 2, Section 2.2.17).

All three reaction mixes were set as follows: 0.5 µl Q5 Pol (2000U/ml), 5X Q5 buffer, 0.5 µl dNTPs 10mM, 1.25 µl Fwd primer 10 µM (either DHFR KO Fwd or Invert cassette Fwd), 1.25 µl Rev primer 10 µM (either Invert cassette Rev or DHFR KO Rev), 1 µl DNA (200 ng/µl) and H_2O to a final volume of 25 µl. The PCR was run at the following conditions: pre-incubation at 98 °C for 30 seconds, amplification for 35 cycles at 98 °C for 10 seconds, 64 °C (5'end assay) or 65 °C (3' end assay and DHFR full-length assay) for 30 seconds, 72 °C for 40 seconds, and final elongation at 72 °C for 2 minutes. The PCR products were run on a 1 % agarose gel (90V, 1 h) (Chapter 2, Section 2.2.8). Note that the initial concentration of positive control was unknown; therefore, the band's intensity is not correlated to that of the other samples.

All PCR products from the 'DHFR full-length' assay were cleaned using ExoSAP-IT and sent to sequence (Chapter 2, Section 2.2.11).

4.5.2 DHFR RNA expression confirmation via RT-PCR

The screening of the four monoclonal lines and the polyclonal line was continued by analysing the DHFR RNA expression levels. As described in Section 2.2.18 (Chapter 2), total RNA was extracted using the PureLink RNA Mini Kit followed by ezDNase treatment. After checking the RNA integrity (Chapter2, Section 2.2.19), 5 μg RNA from each population was reverse-transcribed into cDNA (Chapter 2, Section 2.2.20) and examined for genomic contamination (Chapter2, Section 2.2.21) (Appendix G for primers sequence).

The cDNA was tested on the assay 'DHFR mains', designed to detect all main DHFR RNA isoforms. The reaction mix was set as follows: 10X PCR buffer, 0.2 μl Taq DNA Polymerase (5U/μl), 2 μl dNTPs 2.5 mM, 1 μl Fwd DHFR mains, 1 μl Rev DHFR mains, 1.5 μl $MgCl_2$ (25mM), 62.5 ng cDNA and H_2O to a final volume of 25 μl. The RT-PCR was run at the following conditions: pre-incubation at 94 °C for 2 minutes, amplification for 35 cycles at 94 °C for 30 seconds, 57.5 °C for 30seconds, 72 °C for 30 seconds, and final elongation at 72 °C for 5 minutes. The PCR products were then run onto a 2 % agarose gel (90 V, 90 minutes) (Chapter 2, Section 2.2.8).

4.5.3 DHFR enzyme presence confirmation via Western Blot

Further confirmation of *DHFR* expression was obtained verifying if a protein could be made. To address this, cells from each monoclonal, polyclonal, and wild-type line were pelleted and lysed to isolate total protein (Chapter 2, Section 2.2.23). Proteins were quantified using a Bradford assay (Chapter 2, Section 2.2.24), and 25 μg of protein per sample was loaded into a 4-20 % Tris-Glycine gel. Following the separation of the bands via electrophoresis, the proteins were transferred into a PVDF membrane ready for Western Blot. Niamh Bookey carried out the entire procedure; thus, this thesis will not report the detailed methods.

4.5.4 DHFR enzymatic activity assessment

All cell lines were tested for DHFR enzymatic activity, in the case of compromised enzyme function. Total proteins were extracted from each cell line (including monoclonal, polyclonal and wild-type) and quantified (Chapter 2, Sections 2.2.23/24). A colorimetric assay was used to measure DHFR reductase activity based on NADPH consumption (Chapter 2, Section 2.2.25).

4.6 Multi-population screening for identification of positive transformants (Results)

4.6.1 DNA screening points out frequent indels around the DHFR cleavage site

The 5'end assay did not give a clear amplification of any of the tested samples (not shown). The 3'end assay confirmed the cassette insertion at that extremity for the polyclonal line (as seen in the previous section) and the monoclonal line A1. None of the other monoclonal lines presented amplification of the cassette as per the planned design. Three prominent bands were amplified from monoclonal line A4 instead, indicating possible rearrangements at the insertion site (Fig. 4.14, left). The DHFR full-length assay would give a 236 bp band if no large rearrangements were made. All monoclonal and polyclonal lines amplified a band at around 250 bp. Even though it is impossible to assess minor length differences via gel electrophoresis, all the amplified bands seem to correspond to the band resolved for the positive control (Fig. 4.14, right).

However, the possibility of point mutations introduced after the Cas9 cleavage remains. That would be the case of NHEJ activation in the place of MMEJ. This hypothesis was verified via sequencing. Although the sequencing results present a certain number of unidentifiable bases, the overall picture shows various small indels and point mutations, which, despite not impacting on the length of the amplicon, may produce severe mutations (i.e. missense, frameshift, etc.) reflecting on the protein structure and function. An example is reported below, in Figure 4.15. For the sequencing data of all lines, please see Appendices L to R.

4.6.2 DHFR transcription is not compromised

The assay 'DHFR mains' produces an amplicon of 78 bp, detecting all DHFR isoforms. All monoclonal lines displayed the said band, similarly to the positive control (Fig. 4.16). Therefore, DHFR continued to be transcribed. However, this method cannot indicate the expected presence of point mutations passed from the DNA to the RNA. In addition, transcripts arising from the allele receiving the gene cassette might be hard to amplify with RT-PCR due to the transcript's length or possible rearrangements in the area where the primers sit.

4.6.3 Monoclonal line A5 displays a reduced expression of DHFR protein

The Western Blot analysis revealed that the DHFR protein is present in all the samples (Fig. 4.17). Nonetheless, the monoclonal line A5 showed a meagre amount of DHFR protein compared to the other samples. This may indicate regulatory pathways acting on DHFR expression, effectively reducing its translatability. Alternatively, the CRISPR/Cas activity on DHFR could have introduced a non-sense mutation making the protein more prone to degradation. To note that a white-light picture of the membrane, inclusive of a protein ladder was taken to estimate the molecular weight of the bands (not shown).

4.6.4 Monoclonal line A5 presents 90 % less reductase activity

DHFR activity is expressed in mU/mg, where 1 U is the amount of DHFR that oxidises 1 μmol NADPH per minute at pH 7.5 at room temperature. The analysis of the lines showed that the A5 monoclonal line presents the least DHFR activity, showing a 90 % decrease in reductase activity compared to the wild-type (Fig. 4.18). These results, concurring with the Western Blot data, led to the conclusion that A5 is the most suitable line to perform further experiments, despite being a knock-down cell line.

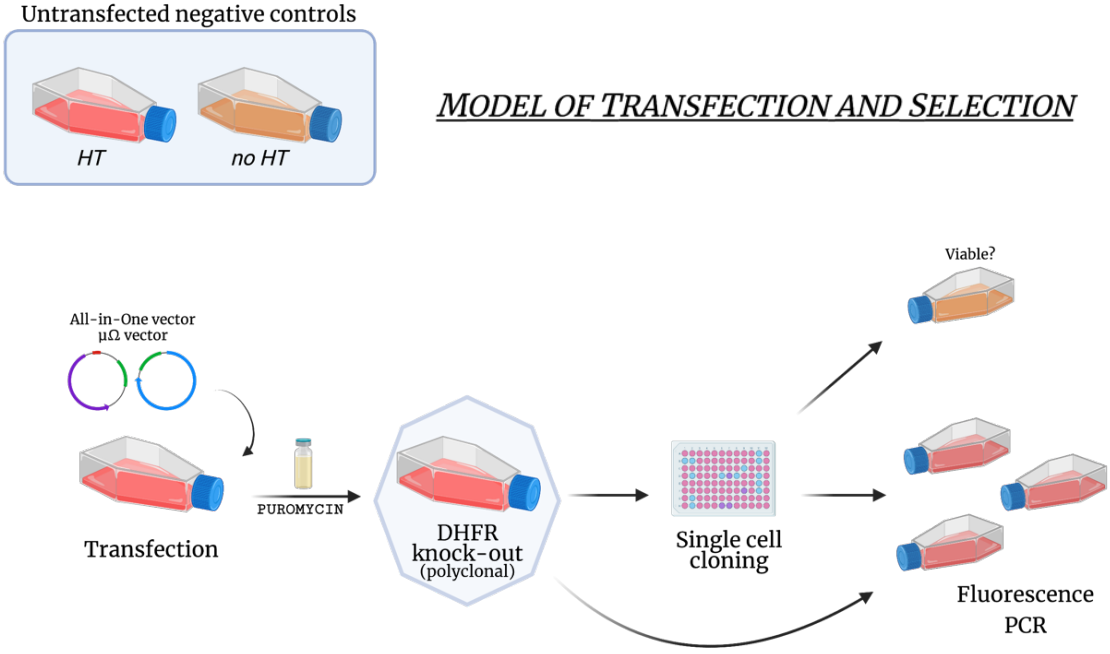


Figure 4.8 DHFR Knockout model of transfection and selection in HepG2 cell line. HepG2 cells were transfected with All-in-One and µΩ vectors in Mirus TransIT-X2 transfection reagent, in triplicate. Two untransfected negative controls were seeded at the same concentration, and treated alternatively with or without HT supplements. After 72 hours from transfection, positive transformants were selected by puromycin. The resulting polyclonal lines were tested for EGFP expression via fluorescence observation. DNA samples were also extracted and amplified via PCR in search of the integrated gene cassette. Monoclonal lines were obtained from the polyclonal one, through a single-cell cloning methods. Similarly to the polyclonal line screening, all monoclonal lines were tested via fluorescence observation and PCR. Image created with BioRender.com

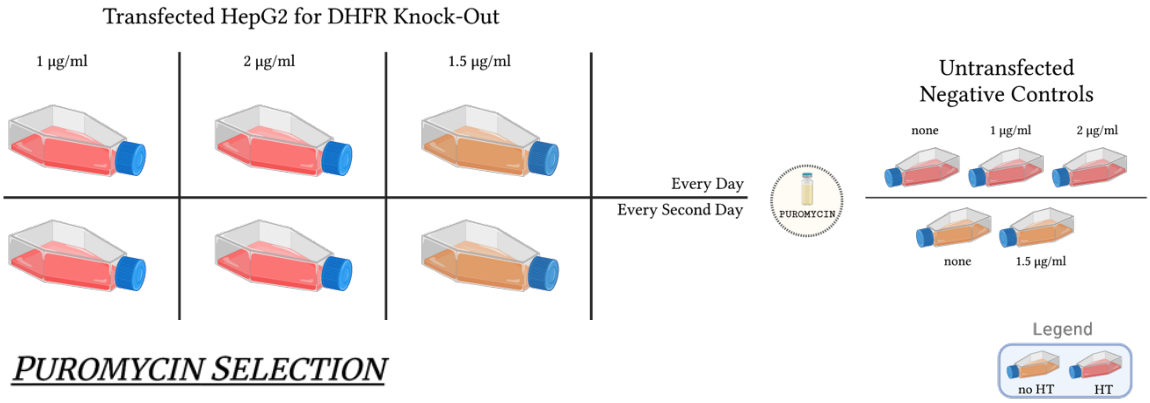


Figure 4.9 Puromycin selection of HepG2 cells post-transfection. Because of an unexpected growth post-transfection, each of the three cell populations was split in two. Four dishes were supplemented with HT, two were maintained in complete media only. The supplemented populations were treated with either 1 or 2 mg/ml puromycin, while the non-supplemented cells were treated with 1.5 mg/ml puromycin. To further differentiate the antibiotic exposure and cell response, three dishes, one per treatment type, had the media changed every day; while the other three dishes, one per treatment type, were supplied with fresh media every second day. All cell populations underwent puromycin treatment for a total of six days. In parallel, untransfected cell populations were treated with HT and puromycin at the same conditions as the transfected cells. In addition, two untransfected cell populations, with and without HT, were maintained as negative controls. Image created with BioRender.com

Table 4.1 Status of transfected cells after six days of puromycin administration at different frequency and concentration.

		Puromycin Administration	
		Daily	Alternate day
Puromycin Concentration	1 µg/ml	80 % confluent with large portion of detached cells	100 % confluent and moderately healthy morphology
	1.5 µg/ml	Over 90 % dead cells	65 % confluent but apoptotic morphology; large portion of detached cells
	2 µg/ml	Over 95 % dead cells. A few cells remain attached with healthy morphology	40 % confluent but apoptotic morphology

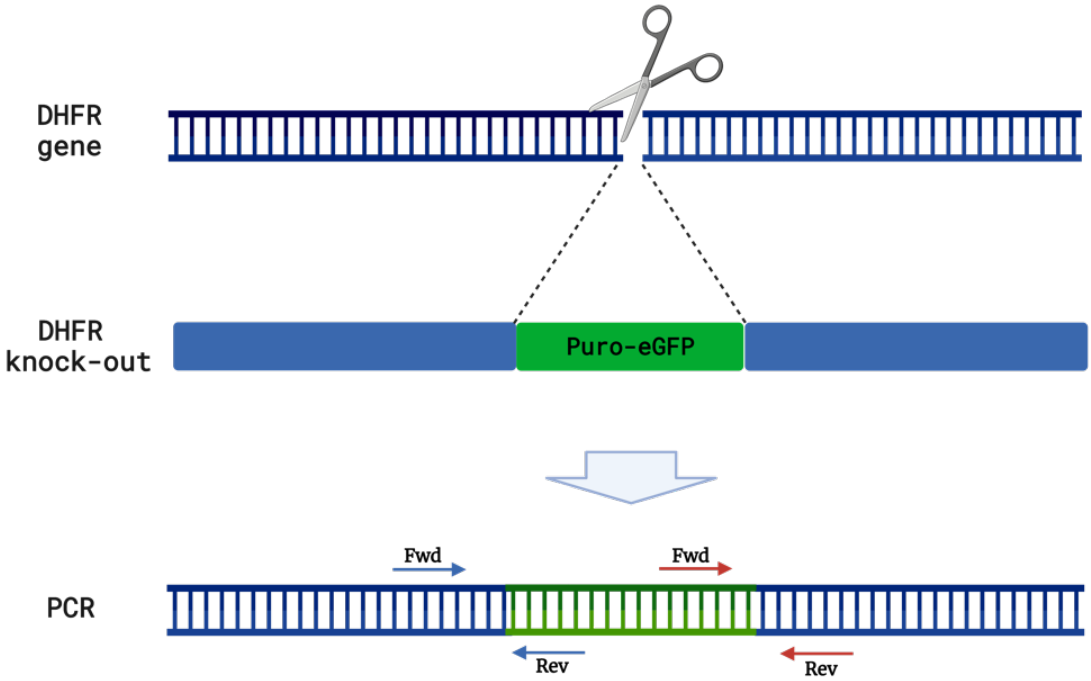


Figure 4.10 PCR amplification of gene cassette extremities in continuity with DHFR. On the occasion of correct gene cassette insertion, two distinct PCR assays could detect the merging sites at both sides of the cassette and DHFR. To amplify the 5' end of the integrated cassette, two primers, in blue, were used: the forward primer sitting on DHFR and the reverse primer sitting on the cassette. Vice versa, the 3'end primers, indicated in red, were designed with the forward on the cassette and the reverse on DHFR. Image created with BioRender.com.

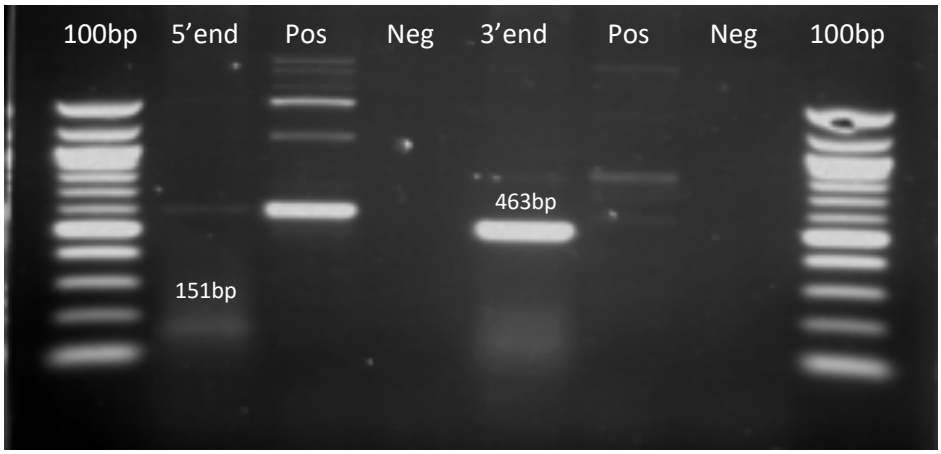


Figure 4.11 Cassette extremities amplification. Resolution on a 1 % (w/v) agarose gel. Positive control contains DNA from HepG2 untransfected cells. Negative control has no added DNA. 100bp ladder has been loaded on both sides. The amplified bands, relative to the 5' and 3' end of the gene cassette inserted into DHFR gene, correspond to the expected length, 151bp and 463, respectively.

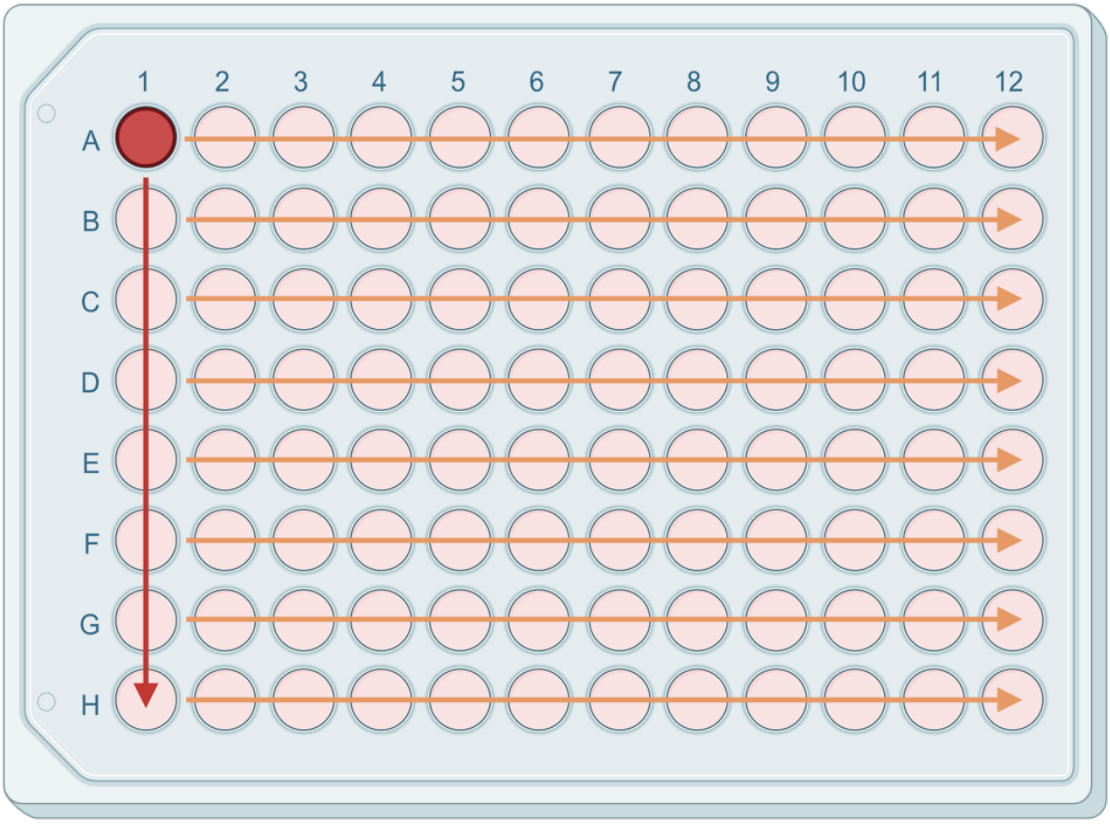


Figure 4.12 Single-cell cloning technique. Well A1 was filled with 200μl of cell suspension. All other wells were filled with 100μl complete media with supplements. Starting from A1, 100μl of cell suspension was moved downwards, mixing well at each passage. The final 100μl of diluted cells were wasted. Line 1 was then topped up with 100μl conditioned media per well. Using the same rationale, cells were diluted horizontally, from lines 1 to 12. At the end of the procedure, each well contained 200μl of resuspended cells. Image adapted from Ryan, n.d. and created with BioRender.com.

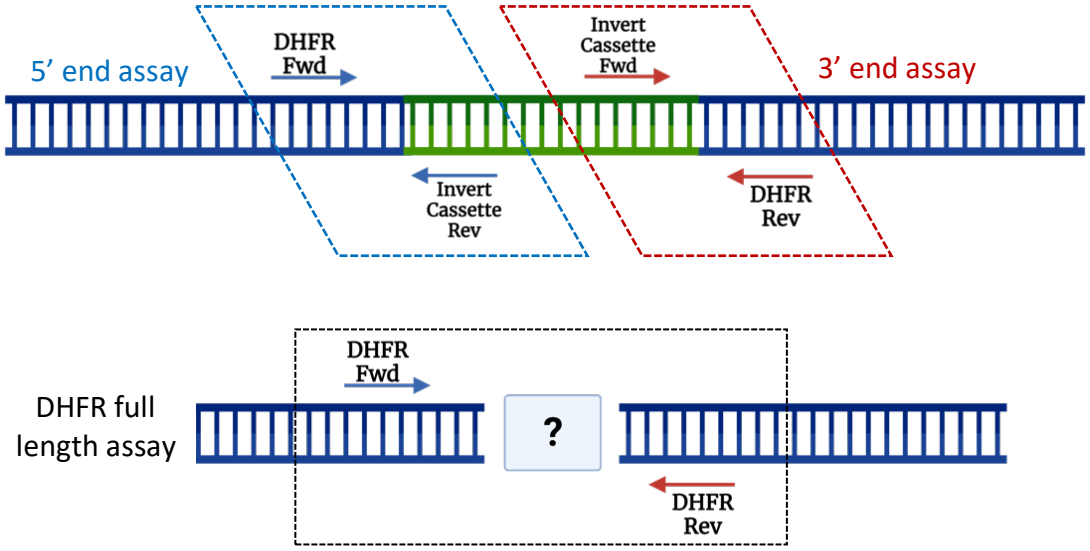


Figure 4.13 PCR assays used to identify the status of DHFR knockout. At the top, the 5’ and 3’ end assays (also described in Fig. 4.10) were used to detect the cassette insertion sites on DHFR (in blue and red, respectively). With the same annealing temperature, the forward primer from the 5’end assay (DHFR Fwd, in blue) and the reverse primer from the 3’ end assay (DHFR Rev, in red) could be coupled in a third PCR assay. This assay, named DHFR full-length assay (bottom), helped detect the native DHFR sequence in case of failed knockout; therefore, it was essential to identify DHFR +/+ or +/- lines. In addition, the DHFR full-length assay detected any minor rearrangements that occurred in that area upon CRISPR/Cas9 activity. Large rearrangements, such as the cassette insertion, were not likely detected because of the length of the amplicon (> 1.5 kb). Image created with BioRender.com.

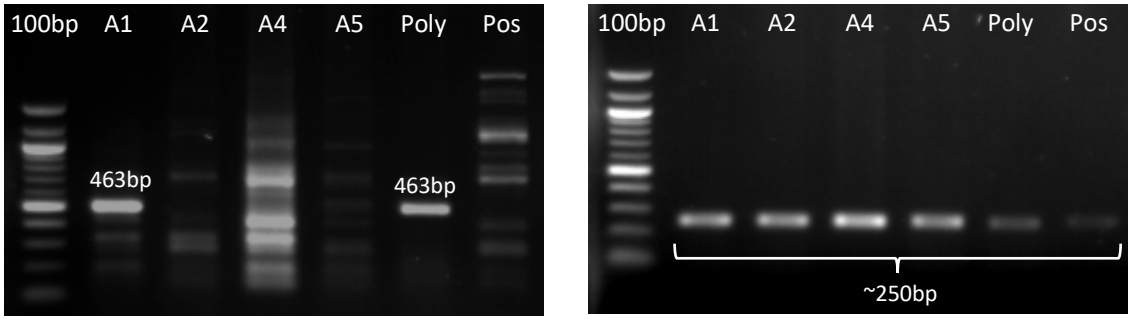


Figure 4.14 PCR amplification of the 3' end assay and DHFR full-length assay on DHFR KO mono- and polyclonal lines. Resolution on a 1 % (w/v) agarose gel. 100bp ladder on the left side of both gels. Positive control contains DNA from HepG2 untransfected cells. The picture on the left side shows the bands relative to the 3' end assay, amplified in all monoclonal and polyclonal lines. Only the polyclonal and monoclonal A1 lines display the expected band at 463bp, indicating the insertion of the gene cassette at the 3' extremity. The DHFR full-length assay (picture on the right) showed amplification in all the samples. The ~250bp band was expected in the wild-type DHFR, thus indicating the failed cassette insertion in at least one allele.

DHFR wildtype

GTCGCTGTGTCCCAGAACATGGGCATCGGCAAGAACGGGGACCTGCCCTGGCCACCGCTCAGGTATCT
GCCGGGCCGGGGCGATGGGACCCAAACGGGCGCAGGCTGCCCACGGTCGGGGTGGCCGACTCCCGGCG
AGAGGATGGGGCCAGACTTGCGGTCTGCGCTGGCAGGAAGGGTGGGCCCGACTGGATTCCCCTTTTCT
GCTGCGCGGGAGGCCCAGTTGCTGATTTCTGC

DHFR – Polyclonal line (from DHFR full length assay)

CCCCNNGANGGGNNNGCCTGGC_ACCGCTCAGGTTCTGCCGGGCCGGGGCGATGGGACCCAAACGGGCNC
AGGNNNNNNNNNGNCGGGGTACCTGGGCGGGACGCGCCAGGCCNNNTNNNNGNGAGAGGATGGGGCN
AGACTTGCGGTCTGCGCAGGCAGNAAGGGTGGGCCCGACTGGATTCCCCTTTTCTGCTGCGCGGGAGGCCCA
GTTGCTGATTTCTACTN

Score	Expect	Identities	Gaps	Strand
206 bits(228)	3e-58	168/207(81%)	21/207(10%)	Plus/Plus

```
DHFR wt  47  CCTGGCCACCGCTCAGGTATCTGCCGGGCCGGGGCGATGGGACCCAAACGGGCGCAGGCT  106
Query    17  CCTGGC-ACCGCTCAGGT-TCTGCCGGGCCGGGGCGATGGGACCCAAACGGGCNCAGGNN  74
                 D              D
DHFR wt  107 GCCCACGGTCGGGGT-----GGCCGA---CTCCCGGC-----------GAGAGGATGGGG  147
Query    75  NNNNNNNGNCGGGGTACCTGGGCGGGACGCGCCAGGCCNNNTNNNNGNGAGAGGATGGGG  134
                          *   *  *   I
DHFR wt  148 CCAGACTTGCGGTCTGCGCTGGCAGGAAGGGTGGGCCCGACTGGATTCCCCTTTTCTGCT  207
Query    135 CNAGACTTGCGGTCTGCGCAGGCAGNAAGGGTGGGCCCGACTGGATTCCCCTTTTCTGCT  194
                          *
DHFR wt  208 GCGCGGGAGGCCCAGTTGCTGATTTCT  234
Query    195 GCGCGGGAGGCCCAGTTGCTGATTTCT  221
```

Figure 4.15 Sequencing data of DHFR cleavage area from the polyclonal line and alignment with DHFR wild-type. The top sequence block indicates the amplicon deriving from the amplification of a HepG2 wild-type line using the DHFR full-length assay. The primers sequence is shown in blue and red. Highlighted in green and purple, the extremities deriving after CRISPR/Cas9 cut. The second sequence block corresponds to the sequencing results from the polyclonal line. As evidenced by the alignment, the region surrounding the CRISPR/Cas9 cleavage site shows several mutations, such as deletions (indicated by the red D), insertions (indicated by the green I) and substitutions (indicated by the orange asterisk). Moreover, additional areas didn't resolve into a clear sequencing (indicated by N).

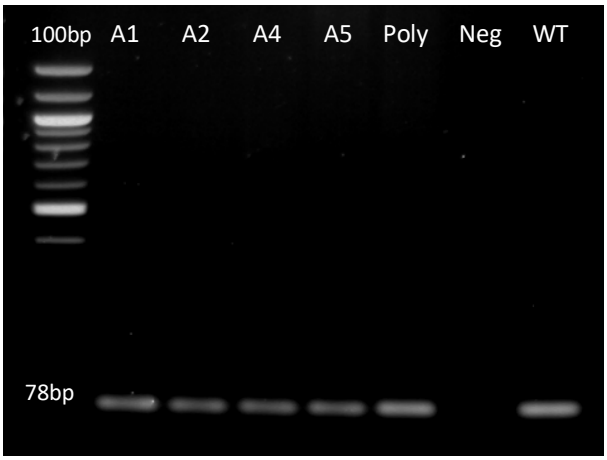


Figure 4.16 RT-PCR amplification of DHFR main isoform assay. Resolution on a 1 % (w/v) agarose gel. 100bp ladder on the left side. Negative control contains no DNA. All samples show the 78bp band, indicative of standard DHFR RNA expression.

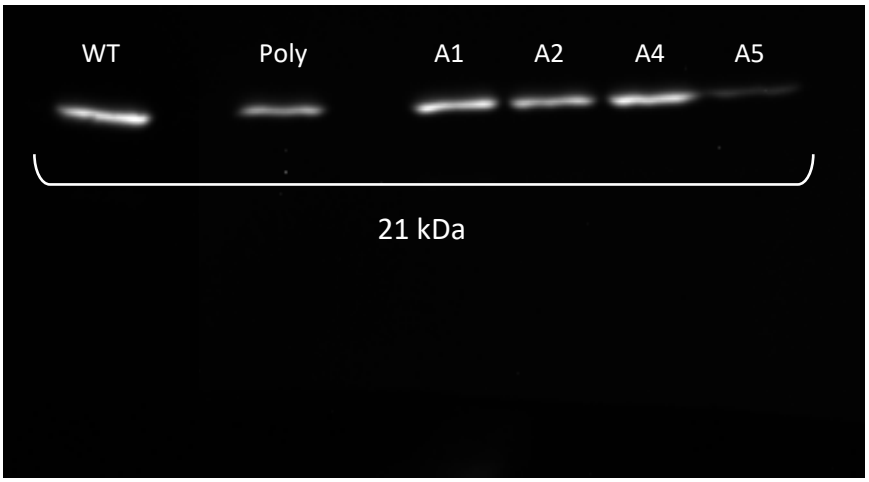


Figure 4.17 Western Blot imaging. Antibody-bound proteins were detected using SuperSignal West Pico PLUS Chemiluminescent Substrate and imaged using Syngene's GeneGnome Bio Imaging System. All samples show the presence of DHFR protein. However, A5 monoclonal line displays the lowest intensity among the lines.

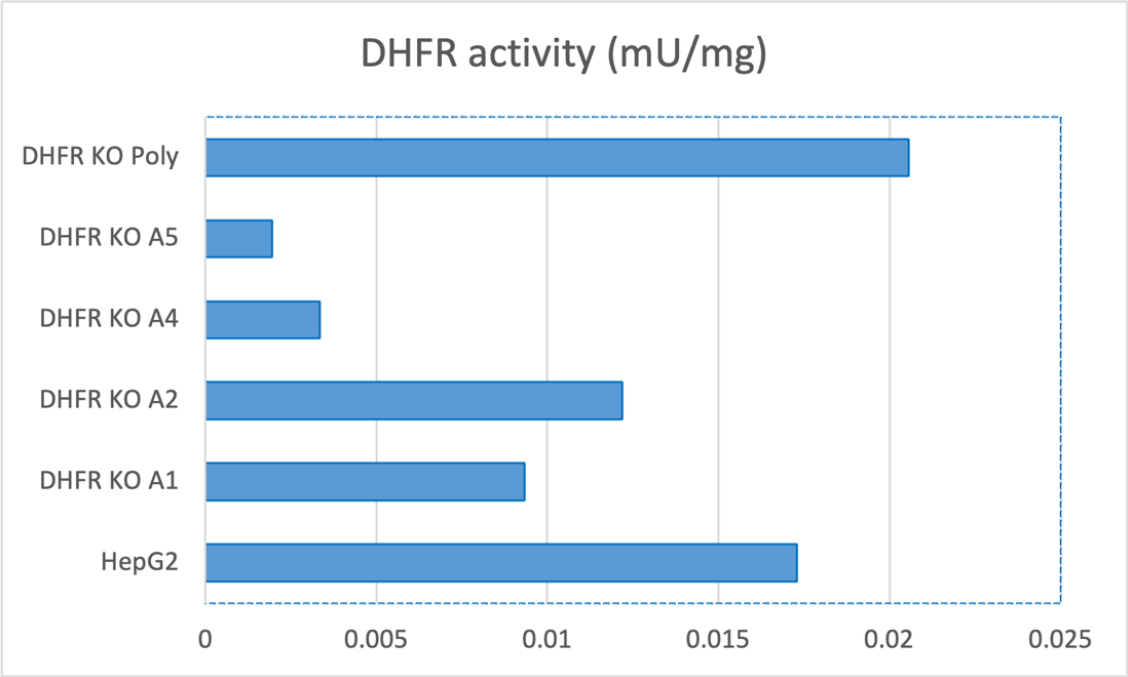


Figure 4.18 DHFR enzymatic assay. The DHFR activity to reduce NADPH was measured in wild-type, mono- and polyclonal lines. The graph shows increased activity in the polyclonal line compared to the wild-type. Contrarily, all monoclonal lines show a significant reduction in activity, with a peak of -90 % presented by the A5 line.

4.7 Discussion

This Chapter discussed the process of creating a novel HepG2 cell line harbouring a DHFR knockout. The lack of DHFR would provide the perfect molecular environment to investigate the presence and function of DHFR2 without risking misdetection due to the remarkable amino acid similarity between DHFR and DHFR2. CRISPR/Cas was the chosen gene-editing method to introduce a DSB at the level of the DHFR gene. The CRISPR/Cas machinery elements were introduced through a plasmid. An additional plasmid, carrying a gene cassette with DHFR-homologous arms, was also transfected. The resulting polyclonal line was screened for the effective cassette insertion, alongside the four monoclonal lines deriving from it. The aim was to select one line presenting a biallelic cassette insertion, thus including a complete DHFR disruption.

The PITCh strategy exploits homologous micro-arms to facilitate the insertion of a gene cassette in a specific genomic site, i.e. DHFR. In this case, the EGFP-2A-Puro^{R} cassette represented the tool for simultaneous gene disruption and transformants selection. The antibiotic resistance gene allowed the survival of the correctly edited cells upon puromycin treatment; the EGFP gene was a fast and direct tool to confirm the occurred insertion (Nakade et al., 2014; Sakuma et al., 2014, 2016). The cassette was designed to be introduced downstream of the DHFR exon 1 in frame and under control of the DHFR promoter. Therefore, the antibiotic treatment would have negatively selected cells with a non-specific cassette insertion. To further confirm the correct insertion site, two PCR assays were designed to amplify the 5' and 3' insertion spots into DHFR.

The screening and selection of the ideal DHFR KO cell line have seen a multilevel analysis, in which DNA, RNA, protein and enzyme activity were examined in search of DHFR impairment.

The investigation of the DHFR cleavage site in the polyclonal line clearly showed the insertion of the gene cassette. Unfortunately, the PCR amplification of the 5'/3' cassette ends in the monoclonal lines was not equally clear, with failed amplification of the 5' insertion extremity. The 3' end PCR showed a specific band in the A1 line only. In addition, all the mono- and polyclonal lines proved to have maintained a "normal" allele, with no cassette insertion. However, the sequencing of the relative amplicons revealed

that the DHFR gene has been subject to random mutations, as shown in other studies (Bennett et al., 2020; Kosicki et al., 2022). We can conclude that all monoclonal lines are heterozygous, with at least one allele presenting minor rearrangements (indels/point mutations). The other allele seems to have a partial cassette insertion in A1. For the remaining lines, the condition of the cassette insertion remains unclear. Rearrangements on the site remain possible (dependent on the simultaneous activation of NHEJ and MMEJ) (M. Liu et al., 2019; Miyaoka et al., 2016) but undetectable due to the sequence-specific nature of PCR primers. Nonetheless, these data confirm the efficacy of the CRISPR/Cas technology.

To explain why the cassette appeared to be present in the polyclonal line but then lost in the monoclonal lines, we can consider three factors: number of screened lines, functional adaptation, and target gene function.

Upon single-cell cloning, only four monoclonal cell lines could be detected. It is plausible that, despite the promising results shown by the polyclonal line, each of the monoclonal lines presented an entirely different profile, with none of them amplifying both cassette extremities. Screening a larger set of monoclonal lines could have resulted in the expected outcome. The limitation deriving from the manual isolation of single cells could be overcome by employing sorting machines like FACS (Gross et al., 2015; Herzenberg et al., 2002).

DNA from the polyclonal line was extracted and tested immediately after transfection and over time. Only the 3'end of the gene cassette consistently amplified, contrary to the 5' end band seen in the post-transfection sample only. This data suggest that the polyclonal population changed over time, losing those cells carrying the entire cassette, or portions of it, confirming that the loss of transgenes is a common phenomenon (Kong et al., 2009; Nakanishi et al., 2002). The loss of the transgene could be due to the lethality associated with the importance of the DHFR gene or to recombination events (Bravo et al., 2020). This hypothesis is also corroborated by the annotation of cells' changed morphology. The aspect of the cells, either poly- or monoclonal, improved gradually, going from very granular and compact to less grainy and elongated. Also, the cells' ability to firmly adhere seemed compromised initially and moved towards normal behaviour over time. By the time the single cells were seeded and grown into full

monoclonal populations, those cells harbouring a complete DHFR KO could not have survived, despite the addition of supplements. Therefore, the loss of severely impaired cells can explain why the surviving monoclonal lines had minor rearrangements, still compatible with life. Overall, the cells have demonstrated an incredible plasticity, making cell adaptation a key for survival (Fodale et al., 2011; McGarry & Olson, 2016).

The small number of monoclonal lines and the change and adaptation of the cell population over time can all be related to the nature of the target gene, DHFR. From the literature, we know that DHFR -/- lines are not viable unless supplemented with HT (Urlaub et al., 1983; Urlaub & Chasin, 1980). More recent studies employing CRISPR/Cas technology coupled with bioinformatic analysis (Blomen et al., 2015; Hart et al., 2015; T. Wang et al., 2015) indicate DHFR as one of the ~3000 essential human genes, whose loss of one or both alleles are not tolerated, leading to lethality (Bartha et al., 2018). However, the Online GEne Essentiality database (OGEE v.3) shows variable essentiality for DHFR depending on tissue-type and deployed experimental method, indicating that gene essentiality is not a static but context-dependent feature (Gurumayum et al., 2021). I speculate that cells with a total DHFR KO did not survive despite the supplementation, and the number of monoclonal lines was too small to find a positive bi-allelic knockout.

The lines were tested for DHFR RNA expression to assess if any of the DNA mutations had affected the nature of the DHFR transcripts. From the RT-PCR data, the transcript's length did not change compared to the wild-type. That, however, does not reflect the nature of the RNA sequence, which could carry small mutations. Sequencing the amplicons would have facilitated the analysis, but it was not performed. However, if any partial insertion of the cassette was present on the other allele, the PCR could not have amplified it due to the specific design of primers.

Despite the unclear picture of the genetic rearrangements and the maintained RNA expression of DHFR, the Western Blot analysis and the DHFR enzymatic activity assay showed a differential abundance and functionality of the enzyme. The monoclonal line A5 appeared to express very little enzyme, with consequent low enzymatic activity compared to the other KO lines and wild-type. Therefore, for the aim of this research, the choice of A5 became obvious. A DHFR knock-down line still proves to be a very

powerful tool for investigating the function of a gene, particularly when essential (Fei et al., 2016; S. Sun et al., 2007, 2011).

To sum up, a homozygous DHFR -/- line was not obtained. This could have been related primarily to the nature of the gene. In fact, the complete elimination of DHFR could be related to non-viable cells, as demonstrated by Li et al. (2019), where murine DHFR biallelic knockouts were lethal, whilst the correspondent heterozygous mutants developed hypertension and abdominal aortic aneurysm (Li et al., 2019).

Limitations of the method can also be listed starting from the cleavage site. Despite being the best site along the *DHFR* gene, the cleavage happened at the very end of exon 1, overlapping with the splicing site. This may have facilitated the elimination of the cassette (if partial insertion occurred just at the 3'end, with consequential disruption on the intronic sequence only) during RNA processing, thus limiting the efficacy of the overall method. An alternative design could have considered different PAM sites, employing Cas enzymes from microorganisms other than *S. pyogenes* (most commonly used) (Nidhi et al., 2021) or engineered versions with improved specificity and efficiency (Ledford, 2020; M.-S. Liu et al., 2020; Tsuchida et al., 2022). Finally, relying on MMEJ as a repair system might not be the safest solution. The knockout of a gene prone to induce cell death when functionally impaired, like DHFR, should rely on a more robust repair mechanism, less susceptible to sequence mistakes. Using a gene cassette with very long homology arms could have produced better insertion rates (Rozov et al., 2019; J.-P. Zhang et al., 2017) yet a higher death rate.

The PITCh strategy is a smart strategy, presented as an improvement of the classical knockout methods (Aida et al., 2016; Ezaki et al., 2022; Kawabe et al., 2018; Sakuma et al., 2016). The use of microhomologies simplified the construction of the insert. In addition, the nature of CRISPR/Cas consented to the transfection of a single plasmid carrying the Cas gene and as many gRNA as needed, allowing the cleavage of multiple sites (i.e. DHFR and the gene cassette). Cutting the gene cassette out of the plasmid (linearised DNA) increased the chances of recombination, balancing the reduced efficacy of short homology arms.

Nevertheless, creating a monoclonal line was proven to be more difficult than expected. This is primarily due to the nature of the target gene and some technical limitations. The production and screening of monoclonal lines required the most time and energy. Replacing single-cell cloning with FACS sorting would have enormously increased the number of monoclonal lines to screen for cassette insertion.

In conclusion, the crucial role of DHFR in cell metabolism and viability does not make it an ideal target for a knockout, thus turning the task into a very arduous process. Beyond the nature of the target gene, the method can be considered pretty solid if also accompanied by technical improvement. So much for DHFR knockout specifically, the use of a different Cas enzyme -and associated PAM sites- could have opened to more suitable cutting sites. Finally, extending the homologous arms in length could have had a more significant impact on the likelihood of accurate cassette insertion. The associated analysis of the DHFR knock-down line will be thoroughly discussed in Chapter 6, in line with the investigation of the DHFR2 knockout line, reported in Chapter 5.

Chapter 5 Functional analysis of DHFR2 knockout in HepG2 cell line. Effects of loss-of-function on DHFR and One-Carbon Metabolism

5.1 Overview

Loss-of-function (LOF) approaches are a powerful tool for understanding the genetic mechanisms behind cell functions and behaviour. The loss of a specific cellular function, either partial or complete, can be obtained by targeting DNA, RNA or protein. The disruption of the normal cell physiology produces phenotypic effects that need to be interpreted. The chosen method to achieve a loss-of-function, the nature of the outcome (total or partial loss), and the complexity of the manipulated gene/genes must be considered when it comes to the analysis of the novel phenotype. Furthermore, robust and broad analyses should be carried out to produce a comprehensive understanding of the gene function and avoid inaccurate conclusions (Hartenian & Doench, 2015; Housden et al., 2017; Mullenders & Bernards, 2009; Teng et al., 2013).

The creation and validation of the DHFR2 knockout cell line were followed by extensive analysis (in order to characterise the line) ranging from cell viability to screening of the transcriptomic and proteomic profiles. Some analyses were chosen based on the current literature and the function assigned to DHFR2, its homologue DHFR and the One-Carbon Metabolism pathway.

The aims of this Chapter are to explore whether DHFR2 function is relevant to cell growth; to verify the role of DHFR2 within OCM; to assess if functional interdependence exists between DHFR and DHFR2. These aims were achieved via the use of the following methods: Cell growth analysis, RTqPCR, RNA microarray, Folate Profiling, Enzymatic Assay, Complementation and Overexpression.

5.1.1 Rationale for cell proliferation analysis

DHFR2 was annotated as a mitochondrial reductase, possibly reducing DHF to THF, similarly to its cytoplasmic counterpart DHFR (Anderson et al., 2011; McEntee et al., 2011). As DHFR is an essential enzyme for life, and its disruption is not compatible with viability (Huennekens, 1994; Raimondi et al., 2019; Schweitzer et al., 1990; Urlaub & Chasin, 1980), we intended to test if a similar effect could be attributed to DHFR2 as well. Therefore, a proliferation assessment was carried out, ensuring also to determine

a possible rescue effect on behalf of HT (as occurs with DHFR-deprived cell lines, surviving only in presence of HT) (Urlaub et al., 1983).

5.1.2 Targeted transcriptome analysis: choice of genes

Continuing with the investigation of the effects of the DHFR2 absence, we tried to evaluate the role of DHFR2 on the OCM. Assuming DHFR2 is a constituent part of the mitochondrial OCM (Anderson et al., 2011), the loss of its function would create unbalances in this very complex pathway. In addition, McEntee, 2011 demonstrated the capacity of the DHFR2 RNA to bind to the DHFR enzyme, proposing a regulatory mechanism (McEntee et al., 2011). In order to investigate how the lack of DHFR2 impacted DHFR expression along with that of other OCM components, a limited set of genes was chosen for the investigation of the transcriptional patterns. The selection was finalised with the help of the STRING database (version 11.0), where only predicted interactions were reported for DHFR2.

By thoroughly examining the DHFR2 gene structure, we noticed that DHFR2 shared the promoter with the neighbouring gene NSUN3, coding for a mitochondrial RNA methyltransferase (EPD, Eukaryotic Promoter Database – new version 006) (Fig. 5.1). Since the late eighties, many studies have reported the presence of divergent promoters throughout the species, from Bacteria to humans (Beck & Warren, 1988; Lepoivre et al., 2013; Seila et al., 2009; Trinklein et al., 2004). The analysis of the transcripts arising from divergent promoters revealed several different combinations, among which coding/non-coding bidirectional transcription, with an increasingly high number of studies reporting functionality associated with these, once dismissed as 'junk', ncRNAs (non-coding RNAs). Some of the functions of these ncRNAs are associated with the reorganisation of local chromatin, regulation of neighbouring genes or influence on distal genes (Wei et al., 2011). The gene pair could be co-expressed and co-regulated, sharing function and tissue specificity or be unidirectionally expressed and functionally unrelated (Chen et al., 2007; Jiménez-Badillo et al., 2017; Voth et al., 2009). Given the above, we considered the hypothesis of DHFR2 acting as a lncRNA (long non-coding RNA), potentially regulating flanking genes or distant genes (e.g., OCM-related genes).

Hence, DHFR2 flanking genes (NSUN3 and ARL13B) were added to the list of OCM-related genes to screen for transcriptional changes.

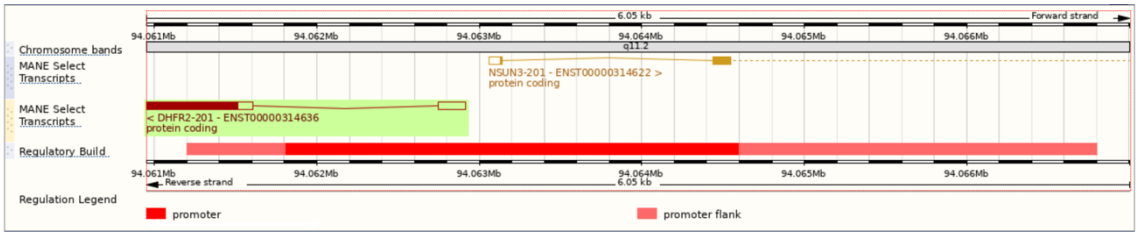


Figure 5.1 DHFR2 and NSUN3 shared promoter. Image created using the Ensembl platform (release 107).

5.1.3 Global transcriptomic and proteomic analysis: DHFR2 LOF broader effects

Investigating potential new targets of the DHFR2 activity - beyond OCM genes and DHFR2 neighbouring genes- was considered the next logical step. To achieve a broader inspection of the molecular phenotype of the DHFR2 knockout line, we implemented the use of omics technologies, specifically microarray (lncRNAs and mRNAs) and mass spectrometry-based proteomics (this method and relative results will not be discussed in this thesis). The microarray technology allows the quantification of a pre-defined set of RNA via hybridisation with anti-sense probes. This may look like a limitation, especially when compared with RNA sequencing, which can capture the entirety of the RNAs in a sample (Lowe et al., 2017). However, despite RNA sequencing presenting higher sensitivity and the capacity to detect novel transcripts, the amount of data produced would be particularly large, thus requiring massive data processing (expert personnel and high power computing facilities). Meanwhile, microarrays offer well-defined hybridisation protocols and analysis pipelines, as well as a relatively low cost (S. A. M. Martin et al., 2016). Also, bearing in mind that the investigation of the DHFR2 knockout can be considered a small study, the microarray method was deemed the most suitable choice.

5.1.4 Assessment of the OCM state upon DHFR2 loss: investigation of the Folate Profile

In addition to the RT-qPCR investigation of the OCM-associated genes (discussed at the beginning of this section), a further examination of the One-Carbon Metabolism was planned, with emphasis on the metabolites. As described in several studies, folate metabolite profiling is an effective method to provide evidence of variations in the One-Carbon Metabolism (Burren et al., 2006; Leung et al., 2013, 2017; Pai et al., 2015). The OCM is a complex network whose primary action is to provide one-carbon units to interlinked metabolic pathways. As a result, subtle alterations in OCM functionality can result in more critical outcomes (Stover, 2009). For example, 10-formyl-THF acts as a precursor for purines biosynthesis, while 5-methyl-THF is employed as a methyl donor in the methylation cycle. Imbalances of specific folate intermediates (in relation to total folate) may suggest disturbances of certain steps of the OCM and help predict the impact on the cellular function (i.e., DNA synthesis, methylation, etc.) (Leung et al., 2013). The analysis of the OCM metabolites was therefore used to estimate if specific reactions were compromised by the loss of DHFR2.

5.1.5 Restoration of the original phenotype: complementation test

Once all the characterisation experiments were carried out, a genetic complementation experiment was attempted, testing if a DHFR2 expressing transgene was able to rescue the DHFR2$^{-/-}$ phenotype. The simplicity and robustness of the method make the genetic complementation test a fundamental genetic tool for gene assignment since mutations occurring in different genes can result in similar phenotypes (Yook, 2005). The transcriptional levels of the previously tested genes were assessed, along with the global proteomics profile of the complemented line. Unfortunately, due to time constraints, I was not able to repeat the entire set of experiments carried out for the DHFR2 knockout characterisation on the complemented line.

5.1.6 Evaluation of opposing effects in DHFR2 overexpression model

Finally, DHFR2 overexpression was performed on the HepG2 cell line to fully understand the role of DHFR2. Overexpression is a powerful tool that allows exploring how an

increase in gene products (either RNA or protein) can lead to dysregulation of a different set of pathways compared to those impacted in a knockout model. In fact, gene overexpression may be as disruptive as gene loss (Prelich, 2012).

To conclude, although extensive examination of the DHFR2 knockout has been accomplished, more experiments could be performed in the future. The creation and maintenance of a permanent DHFR2 knockout line indeed represent a beneficial long-term tool for advancing research in the field.

5.2 Methods

5.2.1 Generation of Cell Growth Curves

The growth of the DHFR2 knockout cell line was tested with and without supplementation (HT/NEAA). A wild-type HepG2 cell line was used as a control and grown either in the presence or absence of supplementation. Each experiment was run in triplicate. Cell viability and number were monitored every three days for a total of ten days using the ADAM automated cell counter (Chapter 2, Section 2.2.13.2). Total and viable cells were counted, with only the latter considered in the growth analysis.

The comparison of the cell growth of the two cell populations was assessed via 'compareGrowthCurves', an algorithm included in the 'statmod' package for statistical modelling in RStudio, created by Gordon Smyth (Baldwin et al., 2007; Elso et al., 2004). This algorithm enables one to run pairwise comparisons between two or more groups of growth curves through a permutation test. A data frame is returned upon data submission containing the observed statistics (Stat), an estimated p-value and an adjusted p-value (for multiple testing).

The data from the cell count was plotted into graphs to facilitate the comprehension of the growth trends of the tested cell lines and relative comparison. The data visualisation package 'ggplot2' was employed to produce all the graphs in RStudio.

5.2.2 Investigation of the OCM-associated genes and DHFR2-neighbouring genes RNA expression via RT-qPCR

RNA was extracted from the DHFR2 knockout and HepG2 wild-type lines, and copied into cDNA, according to the methods described in Chapter 2, Sections 2.2.18/19/20/21. The RT-qPCR experiments were performed as stated in Chapter 2, Section 2.2.22. The comprehensive list of primers and probes used for the assays can be found in Appendix W.

Four different endogenous control genes (*GAPDH, GUS, RPS13* and *TBP*) were tested to identify the more stable reference gene shared by both cell populations. *GUS* and *TBP* were both chosen to normalise the expression levels of the tested assays and used in accordance with each assay's Cq values. The relative fold gene expression of samples was calculated using the 2–ΔΔCt method (LightCycler 96 software - Roche). Statistically significant differences in the relative expression ratios were assessed using a one-way ANOVA test (Microsoft Excel), with a p-value threshold of 0.05 or below.

5.2.3 mRNA and lncRNA global expression profiling via microarray

RNA was isolated from the HepG2 wild-type and the DHFR2 knockout cell lines in triplicate according to Section 2.2.18 (Chapter 2). The purity and concentration of total RNA were determined (Chapter 2, Section 2.2.10). Only the samples presenting A260/280 and A260/230 above 2 and concentrations above 1 µg/µl (40 µl total volume) were considered eligible for microarray analysis. The samples were shipped to Arraystar Inc. (MD, USA) to perform the mRNA/lncRNA expression profiling. Upon sample receipt, Arraystar performed total RNA quality control once more and subsequently proceeded with cRNA synthesis and labelling (passing through a cDNA step). The labelled cRNA was then hybridised onto the array (Human LncRNA Array v5.0, Arraystar), washed and scanned (Agilent Scanner G2505C). Raw data were extracted (Agilent Feature Extraction software v. 11.0.1.1) and finally analysed (normalisation and processing by GeneSpring GX v12.1 software package, Agilent). A comprehensive report including a list of differentially expressed mRNAs/lncRNAs, Scatter Plots, Pathway analysis and Gene ontology analysis was returned.

5.2.4 Quantification of OCM intermediates via mass spectrometry

Folate metabolite profiling was performed via ultraperformance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS) by Dr Kit-Yi Leung at the Institute of Child Health, UCL, London. Cells were grown in our lab until confluency, harvested by trypsinisation and counted (Chapter2, Sections 2.2.13.2). Each sample consisted of around $2x10^7$ pelleted cells that were snap-frozen and stored at -80 °C until ready for shipment in dry ice. Cell pellets were resuspended in a buffer containing 20mM ammonia acetate, 0.1 % ascorbic acid, 0.1 % citric acid and 100mM DTT at pH 7. Cell suspensions were sonicated using a hand-held sonicator at 40 % amplitude, on ice, for 10 seconds. Total proteins were extracted by precipitation with 2 volumes of acetonitrile. The samples were mixed for 2 minutes and centrifuged at 12,000 x g, 4 °C, for 15 minutes. Supernatants, containing the proteins, were lyophilised and stored at -80 °C prior to mass spectrometry (Leung et al., 2013).

Samples were re-hydrated with 30 µl ultrapure water, centrifuged (at 12'000 x g, 4 °C, for 5 minutes) and transferred into glass vials for UPLC-MS/MS. Folate metabolites were resolved by reversed-phase chromatography. The system used Acquity UPLC BEH C18 columns (50mm x 2.1mm; 1.7µm bead size, Waters Corporation, UK) and two buffers as solvents (Buffer A: 5 % methanol, 95 % Milli-Q water and 5mM dimethylhexylamine at pH 8.0; Buffer B: 100 % methanol). The column was equilibrated using Buffers A+B on a 95:5 ratio, and then 20 µl of sample was injected. Subsequently, the buffers were injected as follows: Buffers A+B (95:5) for 1 minute, Buffer B 5-60 % gradient for 9 minutes, Buffer B for 1 minute. Afterwards, the columns were re-equilibrated for 4 minutes. Folate metabolites were eluted at a flow rate of 200 nl/min and run through a XEVO-TQS mass spectrometer (Waters Corporation, UK) operating in negative-ion mode (Capillary 2.5kV, Source temperature 150°C, Desolvation temperature 600°C, Cone gas flow rate 150 L/h and Desolvation gas flow rate 1200 L/h). Folate intermediates were measured by Multiple Reaction Monitoring (MRM), using optimised cone voltage and collision energy for precursor and product ions (Leung et al., 2013).

MassLynx software (Waters) was used to extract the peak areas of the individual folate metabolites. The total peak area of each folate intermediate was then indicated as a percentage of the total folate, in each sample (Leung et al., 2013).

5.2.5 Complementation test of the DHFR2 knockout line: plasmid preparation and succeeding cell line analysis

To validate if the phenotypic effects discovered in the DHFR2 knockout line could be reverted, a complementation test was performed in order to rescue the DHFR2-deficient cell line. The following plasmids were used for the complementation test: pCMV6-AC-DHFRL1 (GFP-tagged) (transcript 1) (also referred to as pCMV6-AC-DHFR2) and the empty vector pCMV6-AC-GFP (Appendices X and Y). All plasmids were transformed into OneShot TOP10 cells and plated (Chapter 2, Section 2.2.7.1). Single colonies were inoculated (Chapter 2, Section 2.2.1) and grown in volumes as large as to allow the plasmids isolation by Maxi prep (Chapter 2, Section 2.2.3). An aliquot of plasmid was Sanger sequenced by Eurofins, and after confirmation of the correct sequence, the plasmids were used for complementation. DHFR2 knockout cells were transfected as indicated in Chapter 2, Section 2.2.15. Cells were grown for 9 days after transfection, during which aliquots of cells were collected to be tested for tGFP expression via Fluorescence observation (after cell fixation and staining, Chapter2, Section 2.2.16), RT-PCR and RT-qPCR (Chapter 2, Sections 2.2.18/22 and 2.2.12), Western Blot (Chapter 2, Sections 2.2.23-24 and Chapter 4, Section 4.5.3) and Proteomics Analysis (performed by Niamh Bookey; this thesis will not report the details of this specific analysis). The complemented line was finally tested for OCM gene expression (RT-qPCR) to examine the possibility of reversion to the wild-type expression profile.

5.3 Results

5.3.1 The loss of DHFR2 does not significantly impact cell proliferation

The assessment of growth differences in the DHFR2 knockout cell population was firstly performed to evaluate if the loss of DHFR2 impacted cell viability and proliferation. In the hypothesis of DHFR2 being involved in the Thymidylate de novo pathway, the impact of HT/NEAA (Hypoxanthine Thymidylate / Non-essential amino acids) addition to cell growth was also investigated (Urlaub & Chasin, 1980). The cell growth trends of the DHFR2 knockout, with and without supplementation, were compared to those of the

HepG2 wild-type, under the same conditions. The comparison among the six possible pairwise combinations of DHFR2 knockout (DHFR2KO) and HepG2 (WT) treated with or without HT/NEAA supplementation (Suppl) were investigated, returning p-values above 0.5 (Fig. 5.2). The adjusted p-values for all the comparisons were equal to 1, indicating no statistical significance in the differences between the growth curves of DHFR2 knockout and HepG2 wild-type cell lines, regardless of the supplementation status.

The growth trends are illustrated in Figure 5.3. The four samples show an almost identical growth pattern, with a sudden diversification on day 10. This unexpected twist seems particularly impressive in the HepG2 wild-type line supplemented with HT/NEAA, while the other three samples seem to maintain a more similar course.

5.3.2 The loss of DHFR2 down-regulates the genes involved in the de novo thymidylate synthesis

The DHFR2 knockout cell line offers a unique environment to investigate DHFR2 function and how the lack of it impacts the cell's transcriptional levels. The focus was mainly placed on DHFR and other OCM-associated genes. As the number of genes involved in the One-Carbon Metabolism is considerably large, the choice fell onto a limited number of them, particularly those showing up- or down-regulation in the MS-based proteomics analysis of the same line (conducted by Niamh Bookey). The following six OCM-related genes were investigated using relative quantification by RT-qPCR: *DHFR, SHMT1, SHMT2, TYMS, ALDH1L1 and SARDH*. In addition, gene expression of *DHFR2* flanking genes *NSUN3* and *ARL13B* were also included in the investigation of gene expression changes via RT-qPCR. Expression levels were detected in all the tested genes, with significant down-regulation effects (relative to HepG2 wild type) in most samples (*SHMT2* and *ALDH1L1* expression levels did not show any significant difference). As expected, no DHFR2 expression was detected in the DHFR2 knockout sample. Regarding the investigated OCM-related genes, the expression levels of *DHFR, SHMT1* and *TYMS* showed a dramatic reduction in the DHFR2 knockout compared to HepG2 wild-type, as shown in Figure 5.4. These findings suggest that DHFR2 is involved in the regulation of the *de novo* thymidylate synthesis complex, formed by DHFR, SHMT1 and TYMS.

5.3.3 OCM downstream pathways are down-regulated in the DHFR2 knockout

RT-qPCR is the golden standard for expression analysis, and it is the best solution when the targets to investigate are known. In the case of a broader screening for transcriptome analysis, especially when investigating a novel cell model, global methods such as microarrays become the most suitable solution (Camarillo et al., 2011). The choice to use the Arraystar Human LncRNA Arrays V5 sprang from the possibility of systematically profiling mRNAs and lncRNAs, simultaneously. An extensive number of probes, divided between about 20,000 for mRNA and 40,000 for lncRNAs, are included in the array. The Arraystar proprietary database extensively collects RNA sequences from all major public databases and knowledge-based mining of scientific publications and their 'lncRNA discovery pipeline' searching through over 47 Tb of RNA-Seq data. Finally, the array probes (60 nt in length) detect individual transcripts by identifying specific exons or splice junctions.

Raw lncRNA and mRNA data were normalised, and low-intensity signals were filtered. After filtering, the lncRNA and mRNA array data quality was assessed and shown through scatter plots. This first analysis of the DHFR2 knockout transcriptome showed the up-regulation of 1154 lncRNAs and 671 mRNAs and a down-regulation of 790 lncRNAs and 712 mRNAs (Fig. 5.5). The correlation between the DHFR2 KO transcriptome profile and that of the control (HepG2) appears to be strong, as indicated by Pearson's correlation coefficient.

Differentially expressed lncRNAs and mRNAs between the two samples were filtered based on the Fold Change filtering, set at ≥ 2.0 and p-values of ≤ 0.05. The elaborated data, shown in the form of Volcano plots, demonstrated that 1155 lncRNAs and 1048 mRNAs were differentially regulated in the DHFR2 knockout line (Fig. 5.6).

To assign biological significance to the list of genes found to be differentially expressed in the knockout model, two different analyses have been performed: Pathway and Gene Ontology analyses. Pathway analysis is a functional analysis that maps genes to KEGG pathways, using p-values below 0.05 to indicate the statistical significance of the pathway in function of the studied condition (Reimand et al., 2019). Pathways are considered more significant the lower the p-value is. The biological pathways that are

enriched among the DE (differentially expressed) genes were visualised through an enrichment score for both down- and up-regulation gene sets. The down-regulated pathways seem to be related to DNA replication and repair, and cell cycle regulation. On the other side, the up-regulated pathways show a more heterogeneous pattern with several signalling pathways involved (Fig. 5.7).

Similarly to Pathway Analysis, Gene Ontology (GO) Enrichment analysis is commonly used to interpret high-throughput data and form valid hypotheses about the observed changes in the investigated samples (Tomczak et al., 2018). Gene Ontology is a ubiquitously used vocabulary encompassing three independent domains: biological process, molecular function and cellular component. Through this systematic approach, every gene and gene product can be universally recognised and described (Ashburner et al., 2000).

For each of the three GO categories, a pie chart was used to indicate the number of biological processes, molecular functions or cellular components affected by the up- or down-regulation of specific gene sets in the DHFR2 knockout model. Fisher's exact test was employed to evaluate the level of correspondence between the DE list and the GO annotation list. The resulting p-values are indicative of the significance of the GO Terms. The p-values were used to extrapolate an enrichment score plotted onto bar charts relative to up- and down-regulated gene sets.

Examination of the down-regulated gene sets showed that the major impact of loss of DHFR2 expression is suffered by the cell cycle and cell division, alongside DNA replication and DNA metabolic processes (Fig. 5.8a). This effect was also observed in the Pathway Analysis. According to these results, the cell components mainly influenced are the nucleus and chromosomes, and organelles lumen (Fig. 5.8b). Concerning the molecular functions, nucleotide and nucleoside binding are affected, together with enzymatic activities linked to DNA, such as helicase, ATPase and DNA-binding activities (Fig. 5.8c). The GO Enrichment score bars can be visualised in Figure 5.9.

Regarding the up-regulated gene sets, the biological processes more highly affected by the DHFR2 knockout are those of cell signalling and cell communication, comprising response to organic and oxygen-containing compounds, and secretion and export (Fig.

5.10a). Consequently, the vesicles and endomembrane systems are the cellular components that experience the major rearrangements (Fig. 5.10b). Finally and correspondingly, the most significant proportion of enriched genes falls in signalling receptor binding for the GO Molecular Function Classification (Fig. 5.10c). The Enrichment Score bar chart relative to the just described GO domains can be consulted in Figure 5.11.

5.3.4 DHFR2 loss induces a global decrease in Folate Metabolites

The status of OCM was tested in the DHFR2 knockout line to assess if imbalances in any of the folate intermediates were present. The main monoglutamated forms of folates were primarily evaluated. These include DHF, THF, CHO-THF (5-formyl-THF), 5-CH3-THF (5-methyl-THF), 5,10-CH2-THF (5,10-methylene-THF), 5,10-CH=THF (5,10-methenyl-THF). The same intermediates' polyglutamated forms (with 2 to 7 glutamates attached) were also investigated. Folate concentration was calculated as pmol/mg protein. In addition, the percentage of each form of folate in relation to total folate was calculated.

Comparing the concentration of all different types of folate between DHFR2 Knockout and HepG2 wild-type, it is evident that there is an overall reduction in folate concentration. However, they seem to maintain the same proportions. All forms of folate are present at halved concentration in the knockout line, except for formyl-THF, which decreases to around 20 % compared to the wild-type (Fig. 5.12).

A similar trend is observed in the polyglutamated forms of folate, with a sharp concentration drop in the DHFR2 knockout sample, as shown in Figure 5.13.

When the concentrations of the folate intermediates are correlated to total folate, it becomes clear that the proportions are pretty similar, with slightly higher percentages in THF and 5’-methyl-THF in the DHFR2 knockout. Contrarily, the rate of formyl-THF had lowered in the knockout line compared to the wild-type (Fig. 5.14).

5.3.5 Normal DHFR transcript levels are restored upon complementation

5.3.5.1 Transfection of DHFR2 recombinant clone and relative analysis

In order to assess if the new phenotype (i.e., reduction of the thymidylate de novo synthesis, impairment of the OCM) observed in the DHFR2 knockout line was due to the loss of DHFR2, a complementation test was the right tool to test this hypothesis by re-inserting the missing gene. Hence, the DHFR2 knockout cells were transfected with pCMV6-AC-DHFR2 plasmid (along with pCMV6-AC-GFP used as a control), which harboured the DHFR2 ORF. The plasmid expressed a fusion protein made of DHFR2 with a C-terminal turboGFP (tGFP). To determine the success of the complementation experiment, the presence of tGFP was extensively examined under different facets: fluorescence observation and tGFP RNA and protein detection.

The observation of fixed and stained cells (DAPI) expressing either pCMV6-AC-DHFR2 or pCMV6-AC-GFP at the fluorescence microscope indicated that the plasmids show the highest expression level on day 2 post-transfection. The fluorescence seems to reduce over time, with little green (tGFP) fluorescence observed on day 9. Fluorescence due to pCMV6-AC-DHFR2 expression was particularly low, with just a few sparse signals on day 2, which entirely disappeared by day 9 (Fig. 5.15).

The presence of the tGFP RNA was confirmed by RT-PCR (not shown) and quantified by RT-qPCR (relative quantification, normalised to *GAPDH*) (primers and probe listed in Appendix W). The trend observed through fluorescence observation appears to be consistent, as shown in Figure 5.16. The expression levels of pCMV6-AC-DHFR2 are relatively high on days 2 and 6, dropping on day 9. On the other hand, pCMV6-AC-GFP expression is impressively high on day 2 but significantly decreases the following days. Considering the expression level on day 2, it was expected to find an equal fluorescent tGFP signal, even though that is not the case. This might indicate a certain difficulty for DHFR2 RNA to be translated, as demonstrated by the Western Blot analysis. Figure 5.17 shows the antibody detection of tGFP on the DHFR2 KO sample complemented with the empty vector pCMV6-AC-GFP. Still, no fluorescent fusion protein detection was observed on the sample complemented with pCMV6-AC-DHFR2. However, the proteomics analysis of the complemented samples showed a small amount of the

recombinant DHFR2 protein (data not shown), indicating a low translation rate of recombinant DHFR2 despite an abundance of mRNA.

The ensemble of these results gives evidence of successful transfection and restored DHFR2 expression, with additional confirmation that the translation of DHFR2 protein is somehow hindered, in line with the current data showing no evidence of an endogenous DHFR2 protein in human adult tissues (proteomics data, not shown in this thesis).

5.3.5.2 Investigation of transcriptome profile in relation to the DHFR2 KO altered gene expression

The analysis of the DHFR2 knockout complemented line was continued with the RT-qPCR investigation of the OCM genes (and DHFR2 flanking genes) that were found to be up- or down-regulated in the knockout line in an attempt to establish whether the restored DHFR2 expression has an effect in reverting the observed transcriptional changes.

The complementation with pCMV6-AC-DHFR2 induced *DHFR, SHMT1* and *TYMS* RNA expression levels to increase (Fig. 5.18). It is interesting to notice that even the complementation with pCMV6-AC-GFP produced an up-regulation of said genes. However, in the case of *SHMT1* and *TYMS*, the expression increment is equivalent in the samples transfected with either pCMV6-AC-DHFR2 or pCMV6-AC-GFP, indicating that the more abundant expression of *SHMT1* and *TYMS* is likely due to the incorporation of a plasmid, and not to the specific effect of the DHFR2 ORF. On the contrary, DHFR exhibited a more significant rise in the sample receiving the pCMV6-AC-DHFR2 plasmid than the empty vector (pCMV6-AC-GFP). This may indicate that the regulatory target of DHFR2 is only DHFR, with SHMT1 and TYMS indirectly regulated upon DHFR down-regulation.

5.3.6 DHFR2 overexpression leads to increased DHFR transcriptional levels

In order to verify the hypothesis that *DHFR2* acts as a regulatory RNA on *DHFR* (Sections 5.3.2 and 5.3.5.2), a transient DHFR2-overexpressing HepG2 line was created. The transfection procedure and following analyses were performed as described in Section

5.2.5, using pCMV6-AC-DHFR2 and pCMV6-AC-GFP on the HepG2 wild-type line. Upon transfection, the cell line was screened for tGFP expression via RT-PCR, RT-qPCR and Western Blot (not shown), thus confirming the success of the procedure. cDNA, copied from RNA extracted on day 2 post-transfection, was interrogated on the DHFR expression levels. As illustrated in Figure 5.19, DHFR relative expression appeared to be significantly raised compared to both HepG2 wild-type and HepG2 transfected with the empty vector. These findings suggest and further confirm that *DHFR2* has a role in the regulation of *DHFR*.

	Group1	Group2	Stat	P.Value	adj.P.Value
1	DHFR2KO	DHFR2KO Suppl	-0.08846713	0.61	1
2	DHFR2KO	WT	-0.28416181	0.53	1
3	DHFR2KO	WT Suppl	-0.69069432	0.50	1
4	DHFR2KO Suppl	WT	-0.20218532	0.66	1
5	DHFR2KO Suppl	WT Suppl	-0.64859915	0.59	1
6	WT	WT Suppl	-0.53031881	0.82	1

Figure 5.2 Permutation test for pairwise comparisons. Test run between groups of growth curves using the algorithm 'compareGrowthCurves' in RStudio. DHFR2KO, HepG2 DHFR2 Knockout line; WT, HepG2 wild-type line; Suppl, Hypoxanthine + Thymidine (HT) / Non-essential Amino acids (NEAA) supplementation

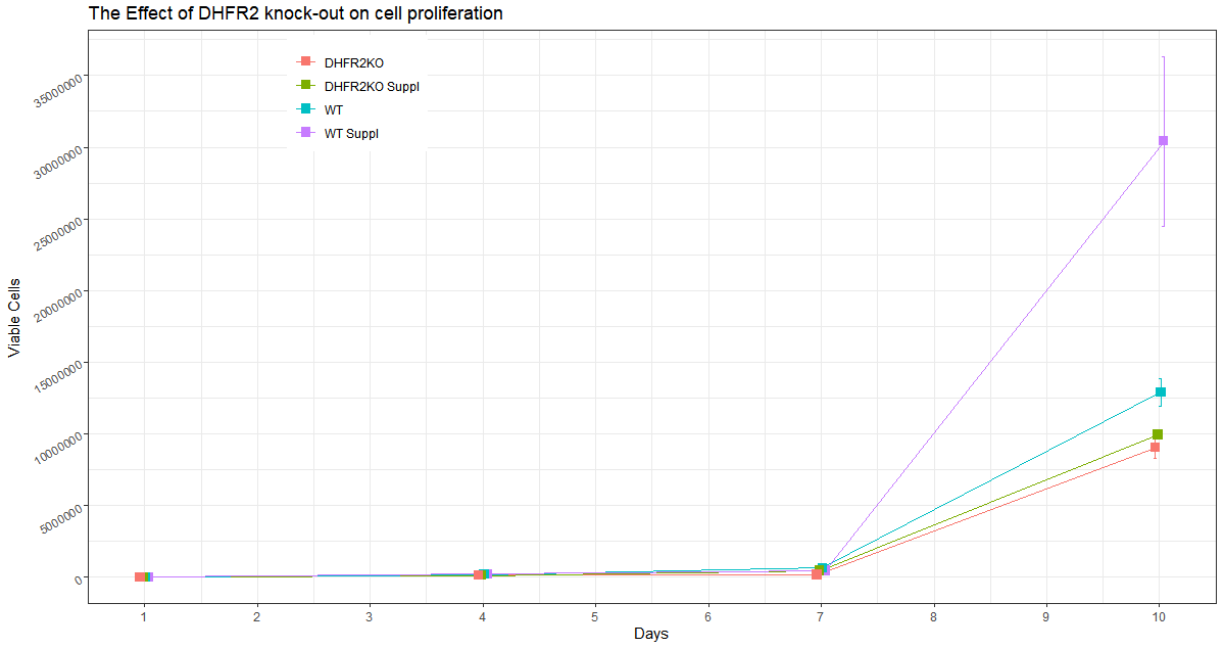


Figure 5.3 Growth curves of HepG2 DHFR2 knockout line compared to HepG2 wild-type Cellular growth curves determined by PI (Propidium Iodide) staining method combined with advanced image analysis by the automated fluorescence cell counter ADAM-MC. All measures were performed in triplicate. DHFR2KO, HepG2 DHFR2 Knockout line; WT, HepG2 wild-type line; Suppl, Hypoxanthine + Thymidine (HT) / Non-essential Amino acids (NEAA) supplementation.

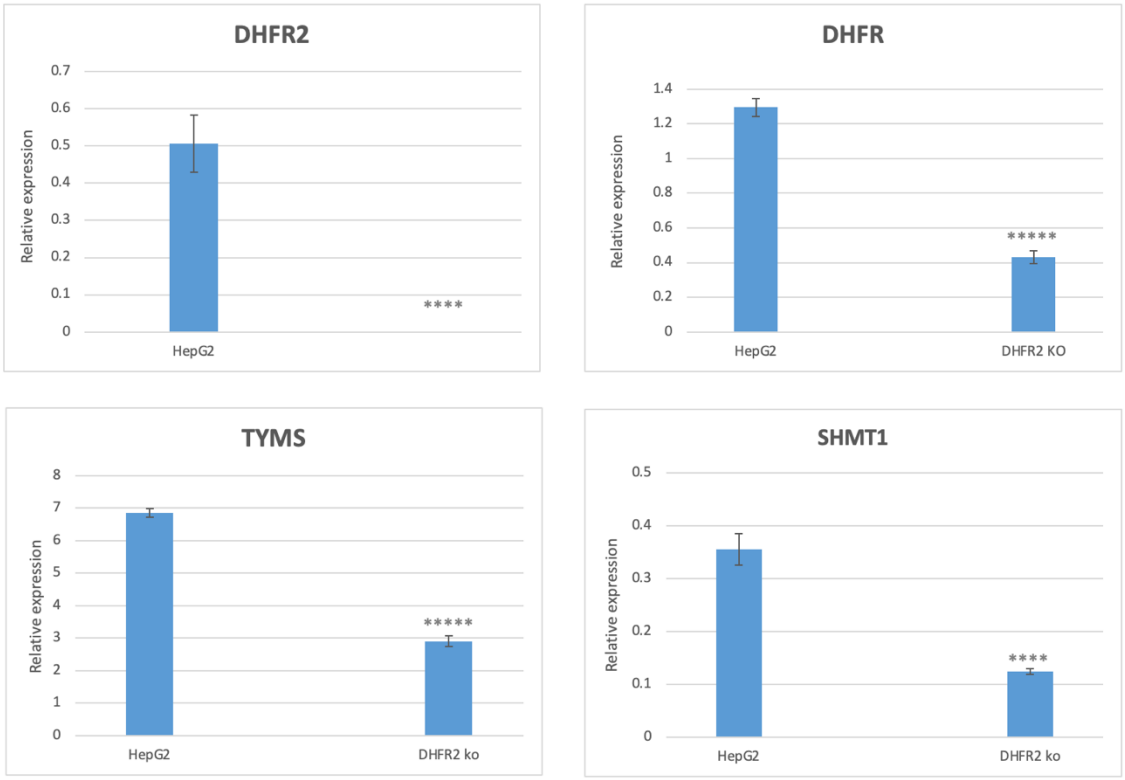


*Figure 5.4 Endogenous levels of DHFR2, DHFR, TYMS and SHMT1 RNAs in DHFR2 KO and HepG2 lines. Expression of DHFR2 RNA is absent in the DHFR2KO line, confirming the validity of the cell model. DHFR, TYMS and SHMT1 showed a drastic drop in expression compared to the wild-type line. Differences in relative expression ratios were compared using one-way ANOVA. P-value threshold was set at ≤ 0.05, yet the statistical significance of the expression differences was markedly higher. The asterisks designate the statistical significance expressed by the p-value as follows: *P$\leq$0.05, **P$\leq$0.01, ***P$\leq$0.001, ****P$\leq$0.0001, *****P$\leq$0.00001. Error bars shown in all graphs are SEM (n=3).*

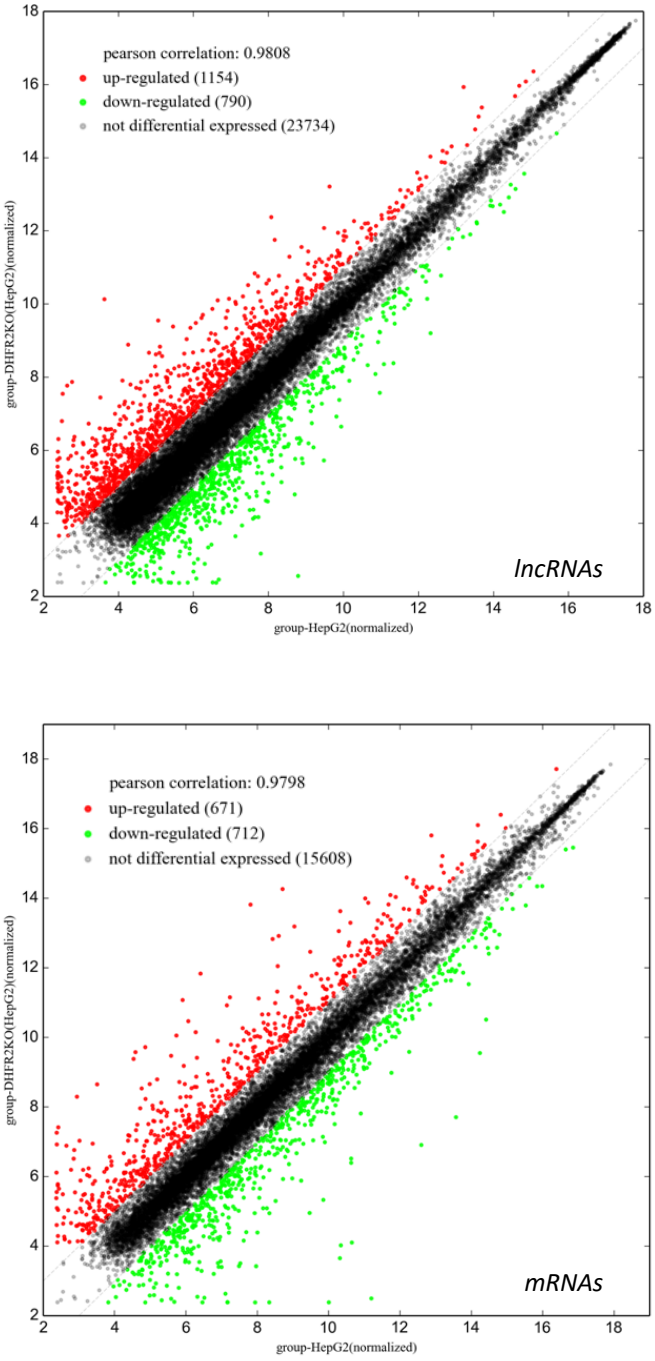


Figure 5.5 Scatter Plot of DHFR2 KO lncRNA and mRNAs after filtering. Raw signal intensities were normalised by quantile normalisation method, and low-intensity RNAs were eliminated. The filtered RNAs were plotted. The expression variation between DHFR2KO and HepG2 can be visualised via a Scatter plot. The values of the X and Y axes are the averaged normalised signal values of the groups (log2 scaled). The grey dash lines are Fold Change Lines (default fold change value is 2.0). The RNAs above and below the grey dash lines indicated more than 2.0 fold change.

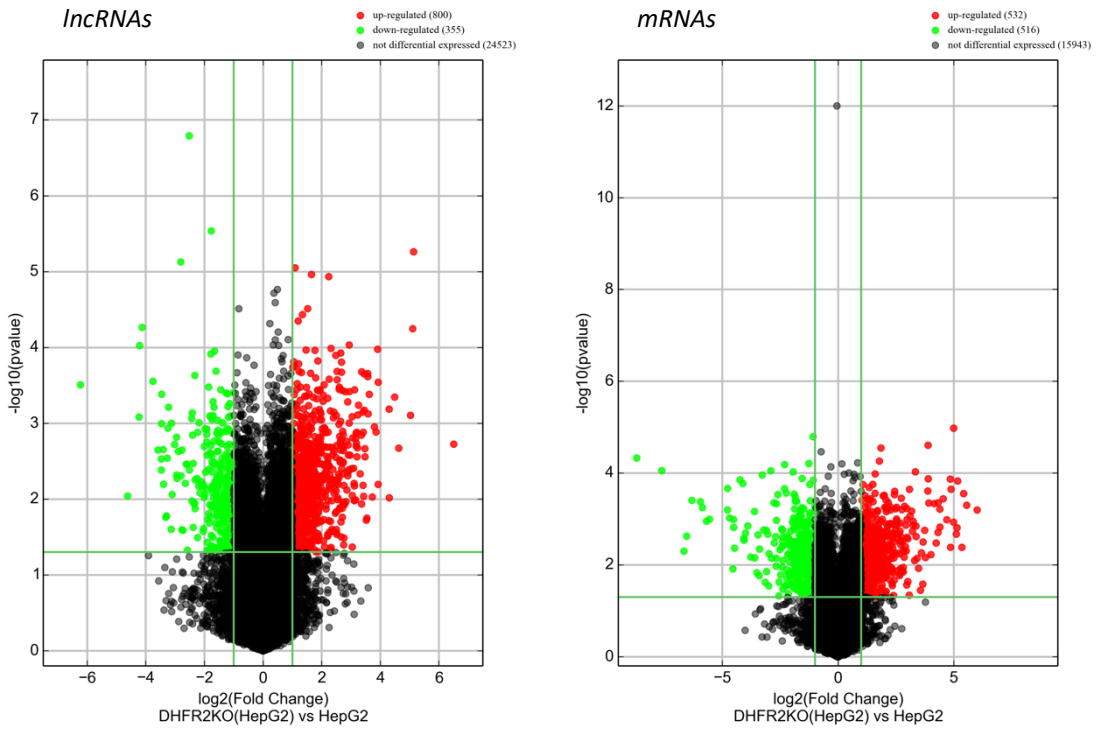


Figure 5.6 Volcano Plot of DHFR2 KO lncRNA and mRNAs. Differentially expressed lncRNAs and mRNAs with statistical significance were identified through Volcano filtering between DHFR2 KO and HepG2. The threshold is Fold change ≥ 2 and p-value ≤ 0.05. The Volcano plot displays large magnitude changes which are also statistically significant. On the y-axis is plotted the $-\log_{10}$ (p-value) and on x-axis the $\log_2$ fold change in RNAs expression between the two samples.

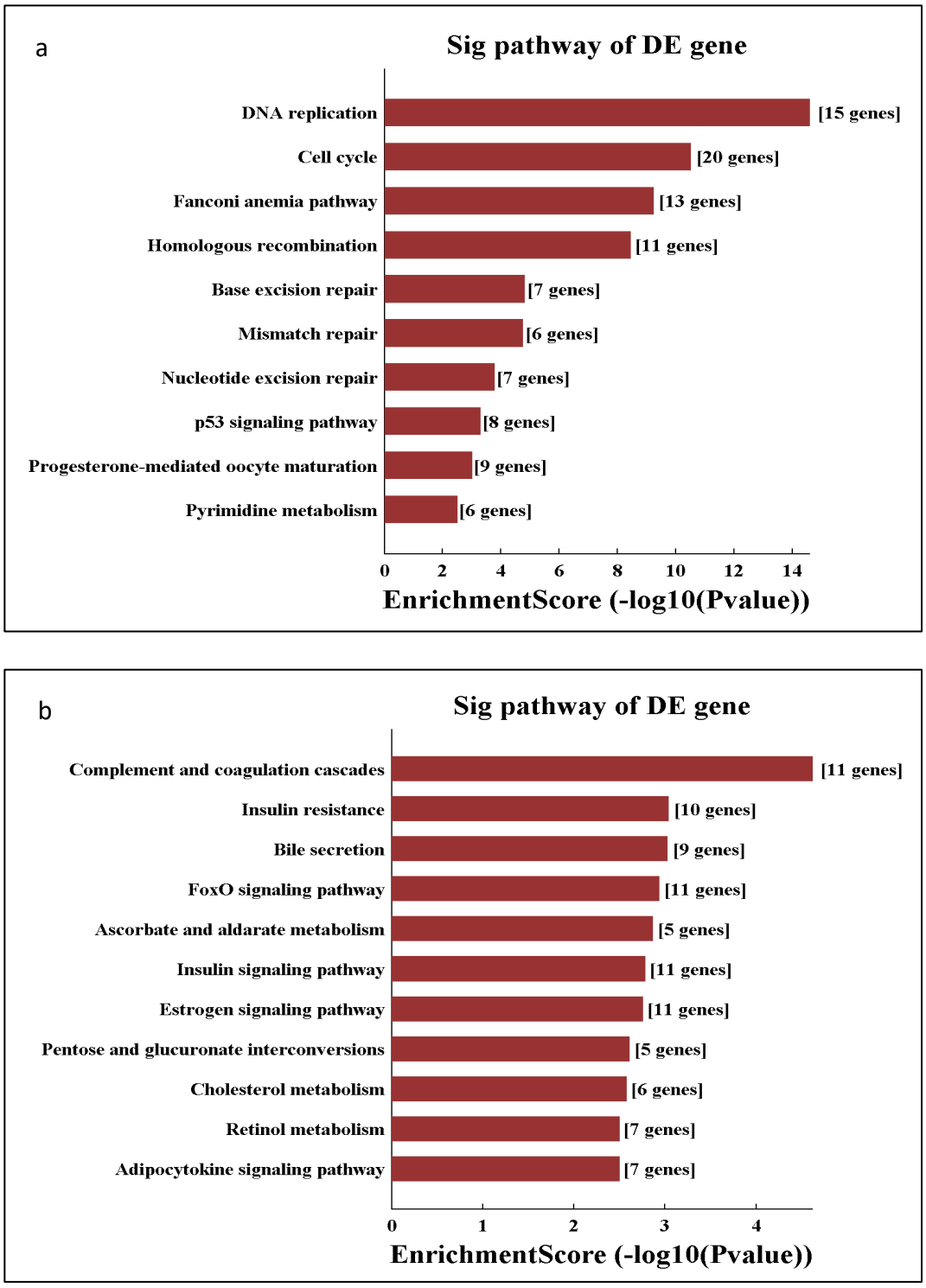


Figure 5.7 Pathway analysis of the differentially expressed (DE) RNAs in DHFR2 KO. a) Down-regulated pathways, b) Up-regulated pathways. Pathway analysis maps genes to KEGG pathways. This analysis allowed determining the biological pathways with significant enrichment of DE genes. The bar plot shows the top ten Enrichment score values of the significant enrichment pathway.

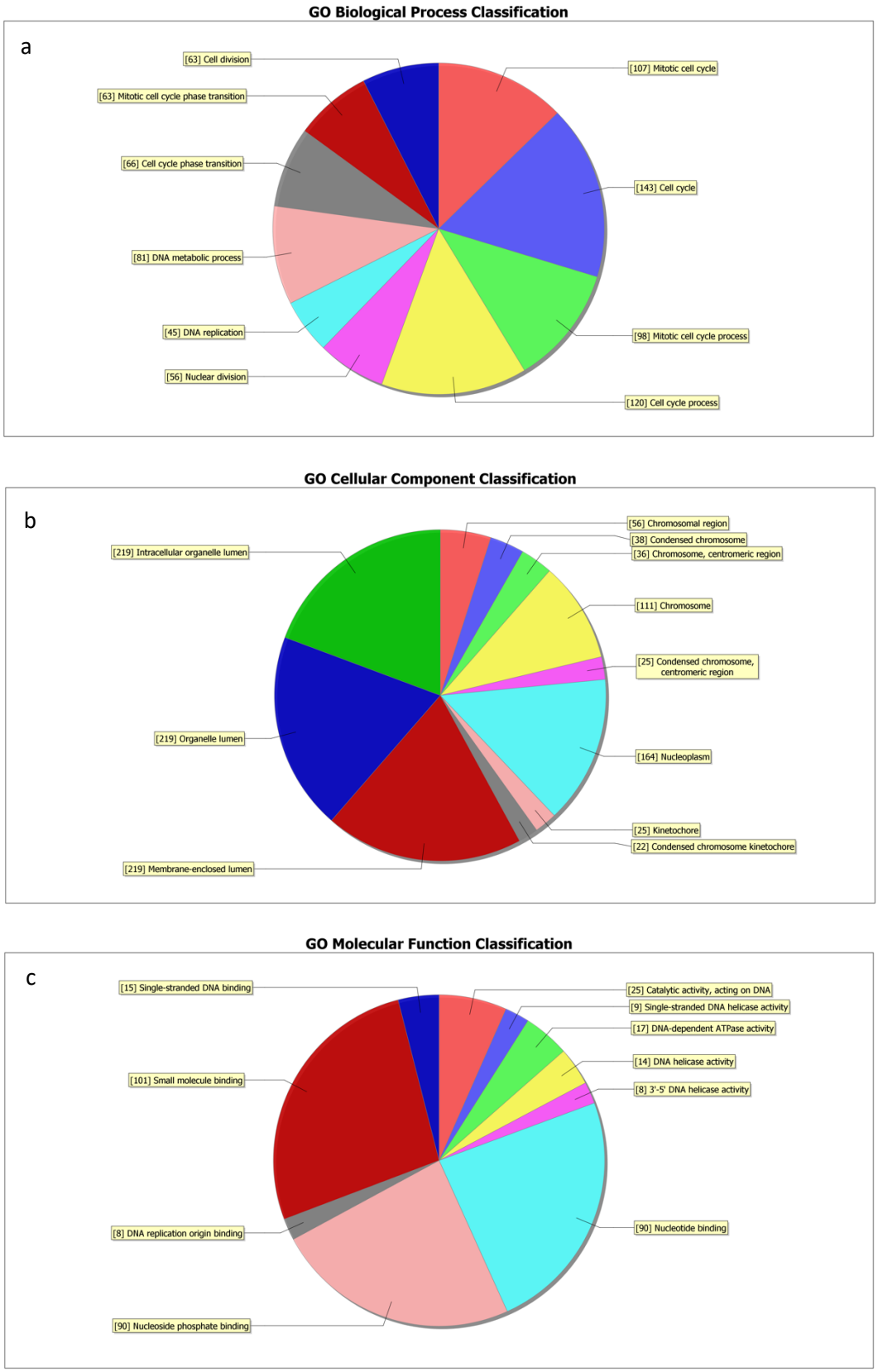


Figure 5.8 GO analysis of the down-regulated RNAs in DHFR2 KO (Pie Charts). GO (gene ontology) analysis associates DE genes with the three GO categories Biological Processes (a), Cellular Component (b) and Molecular Function (c). The pie charts show the top ten counts of the significant enrichment terms per each category.

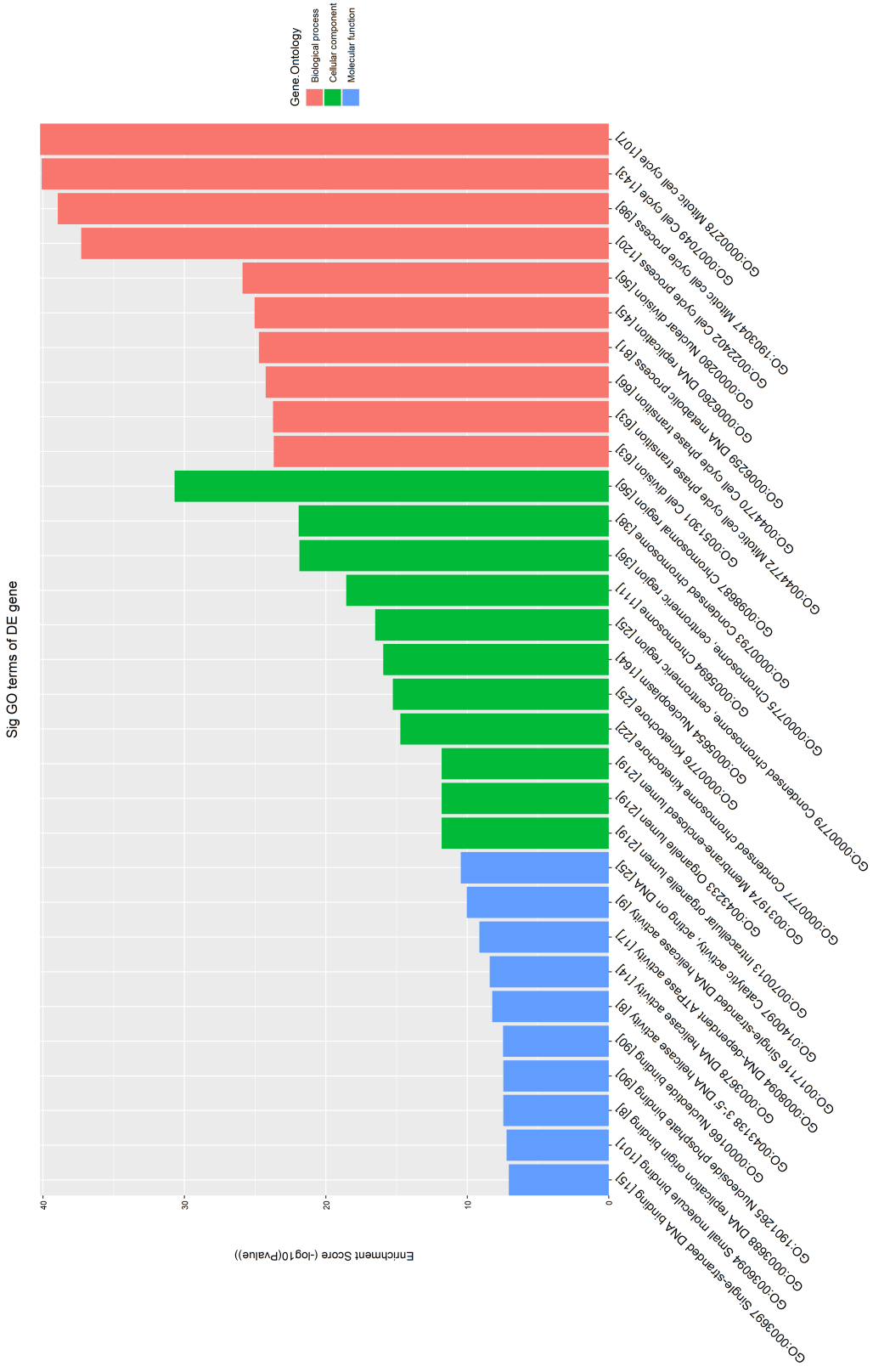


Figure 5.9 GO analysis of the down-regulated RNAs in DHFR2 KO (Bar Chart). The bar plot shows the top ten Enrichment Score values of the significant enrichment terms per each category: Biological Process in red, Cellular Component in green and Molecular Function in blue.

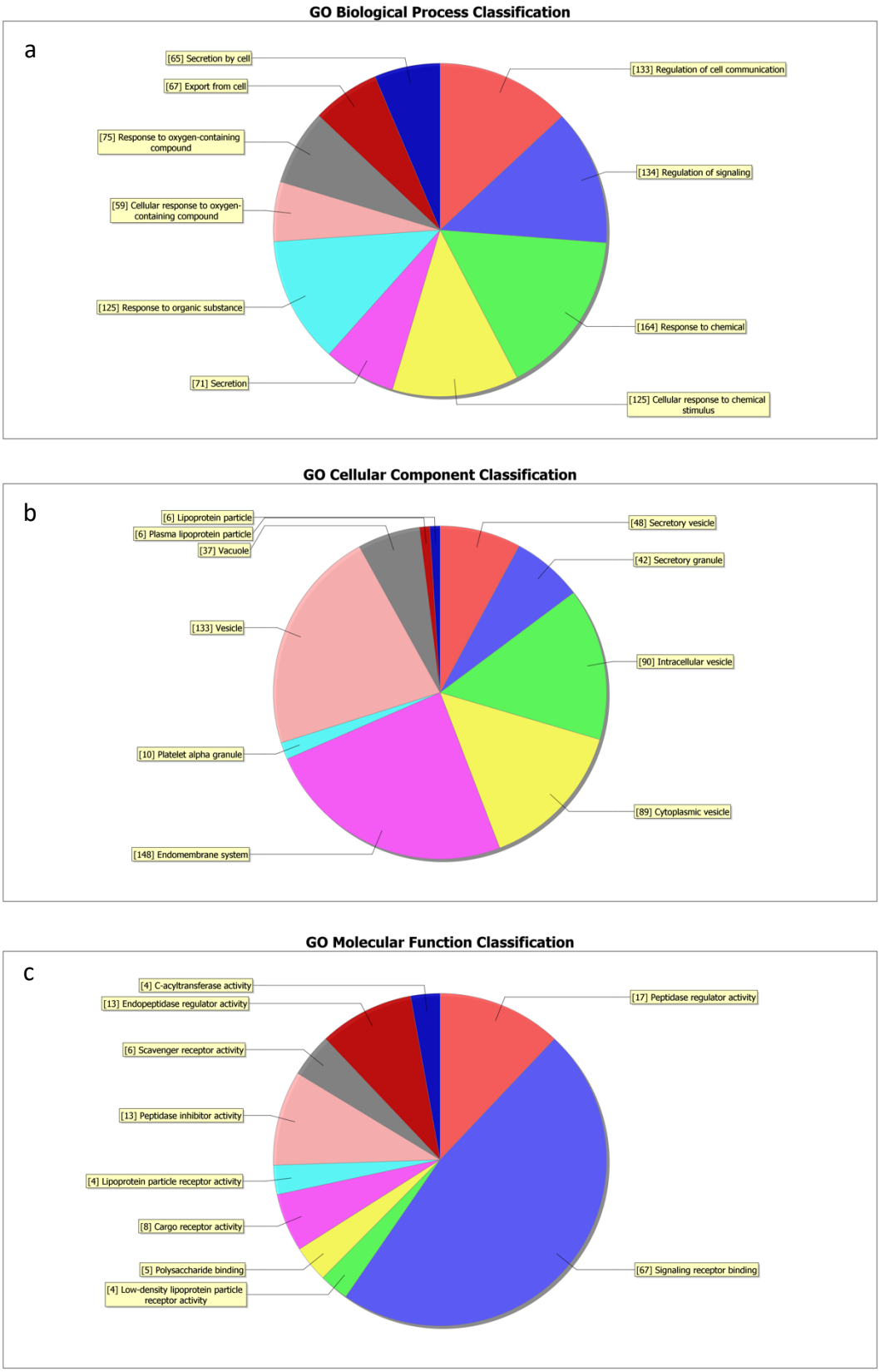


Figure 5.10 GO analysis of the up-regulated RNAs in DHFR2 KO (Pie Charts). GO (gene ontology) analysis associates DE genes with the three GO categories Biological Processes (a), Cellular Component (b) and Molecular Function (c). The pie charts show the top ten counts of the significant enrichment terms per each category.

Figure 5.11 GO analysis of the up-regulated RNAs in DHFR2 KO (Bar Chart). The bar plot shows the top ten Enrichment Score values of the significant enrichment terms per each category: Biological Process in red, Cellular Component in green and Molecular Function in blue.

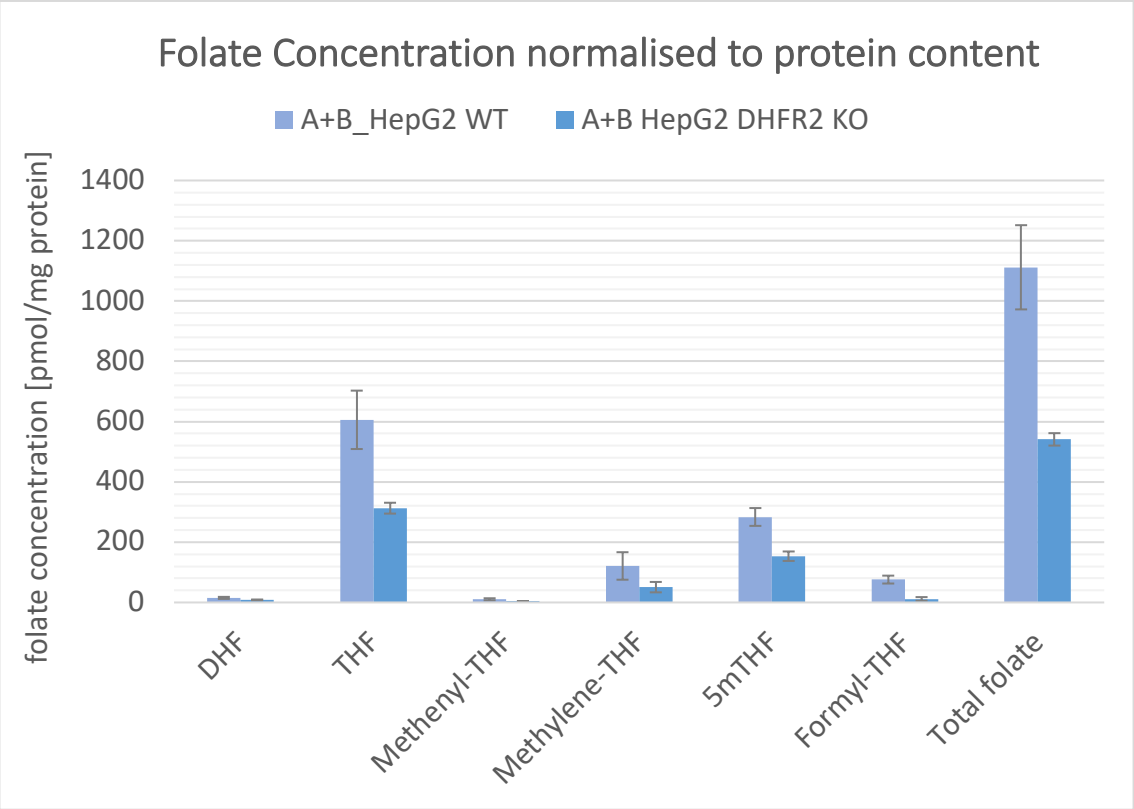


Figure 5.12 Folate metabolites profile in DHFR2 ko and HepG2 wild-type lines. The concentrations were normalised to protein content and expressed as pmol/mg protein. The measures were performed in duplicate (as indicated by A+B).

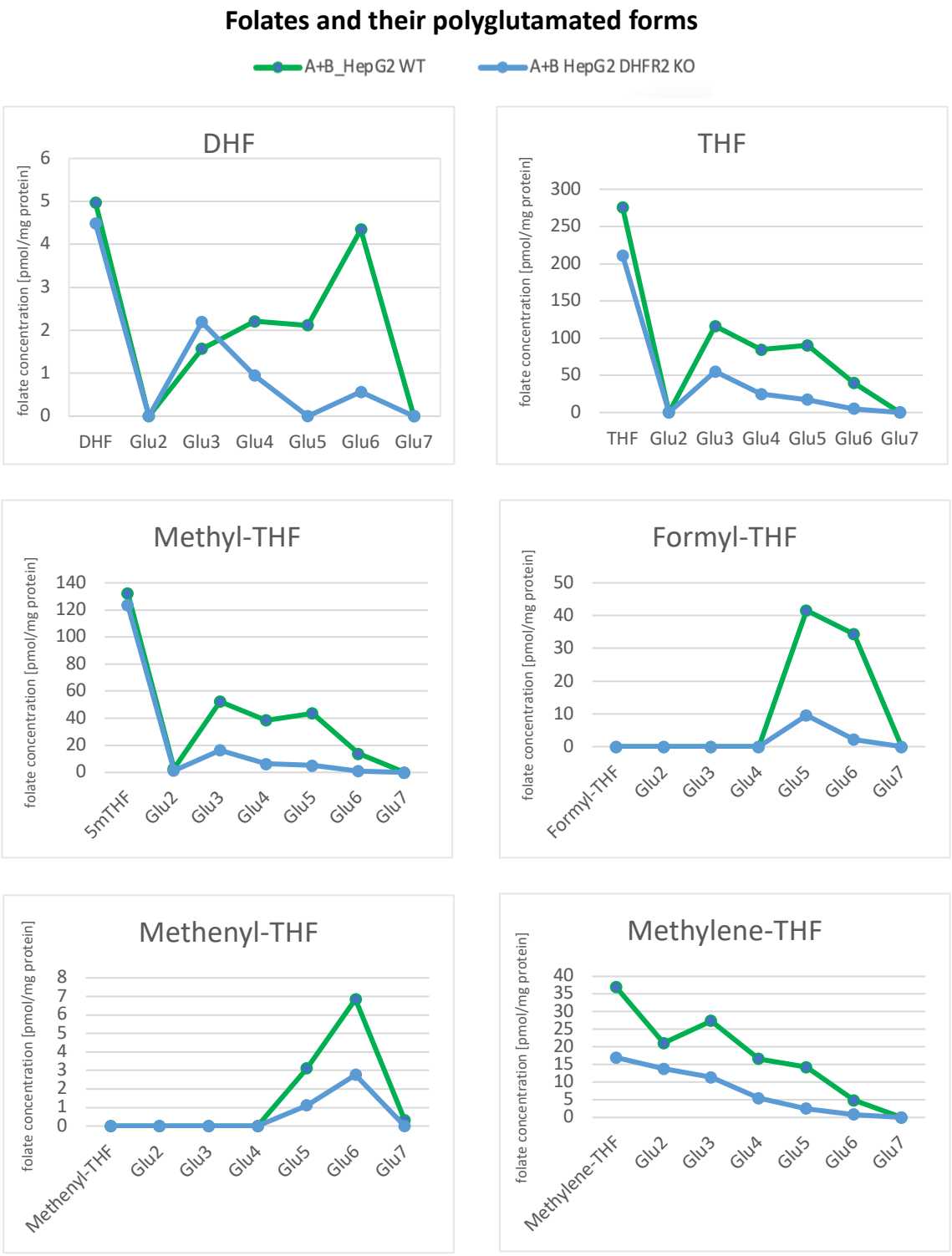


Figure 5.13 Concentration of Folate Intermediates and their relative polyglutamated forms in DHFR2 ko and HepG2 wild-type lines. Glu2-7 indicate the number of glutamate molecules attached. The concentrations were normalised to protein content and expressed as pmol/mg protein. The measures were performed in duplicate (A+B).

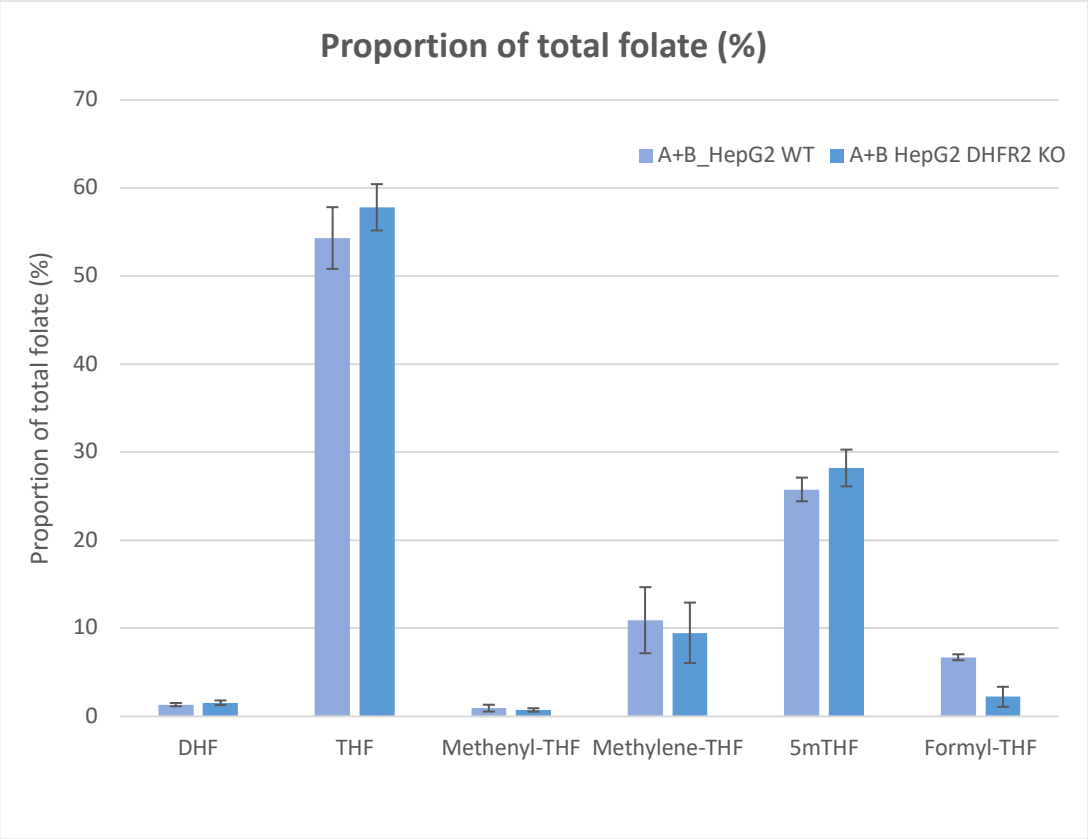


Figure 5.14 Relative proportions (expressed as % of total folate) of folate metabolites in DHFR2 KO and HepG2 wild-type lines. The percentages for each folate intermediate were calculated considering the sum of all glutamated forms (n1–7).

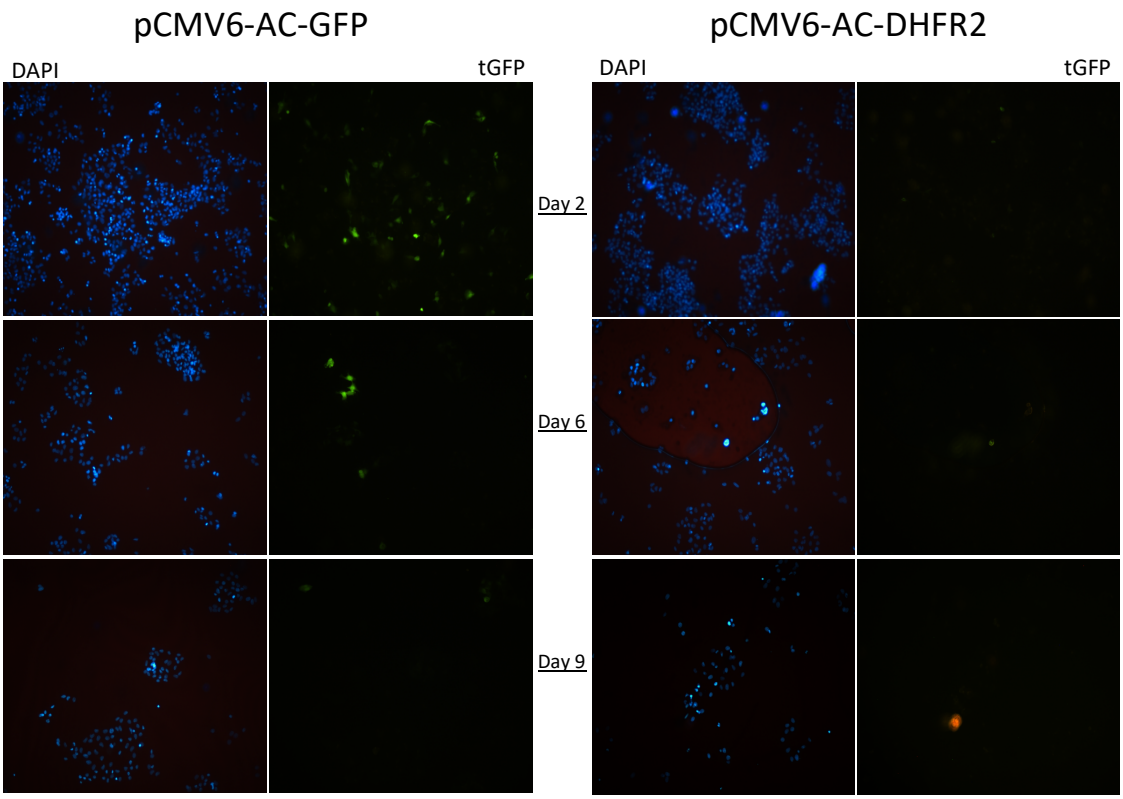


Figure 5.15 DHFR2-GFP expression in DHFR2 knockout line (Fluorescence). Cells transfected with pCMV6-AC-DHFR2(gfp) and pCMV6-AC-GFP (control) were fixed and counter-stained with DAPI at days 2, 6 and 9. Fluorescence was observed using a Leica DFC 500 and pictures taken via the Leica Application Suite software.

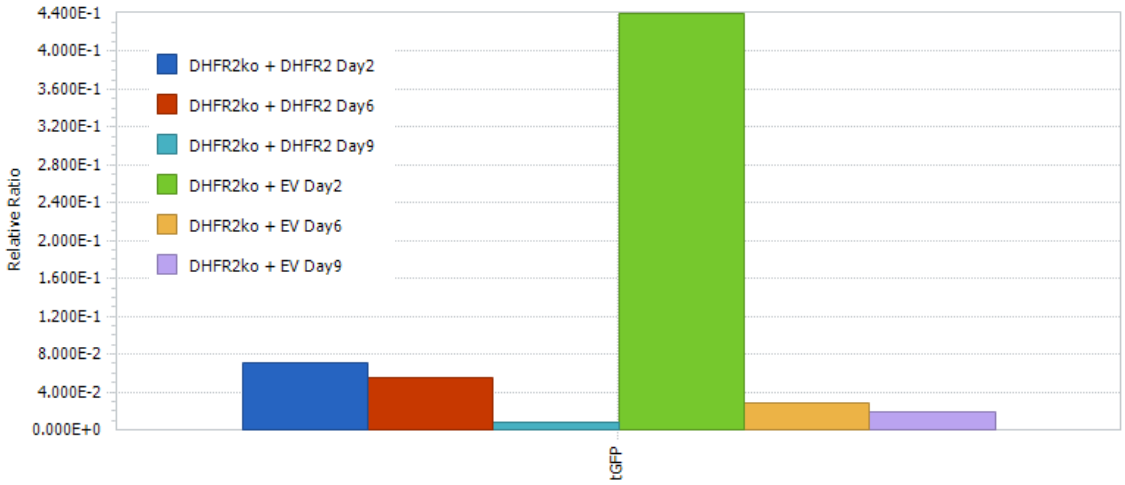


Figure 5.16 DHFR2-GFP RNA expression in DHFR2 knockout line. RNA samples were isolated 2, 6, and 9 days post-transfection (DHFR2-GFP and GFP only) and GFP relative expression was measured against GAPDH. DHFR2ko + DHFR2 indicates the DHFR2 knockout line transfected with pCMV6-AC-DHFR2(gfp). DHFR2ko + EV indicates the same line transfected with the empty vector pCMV6-AC-GFP (negative control).

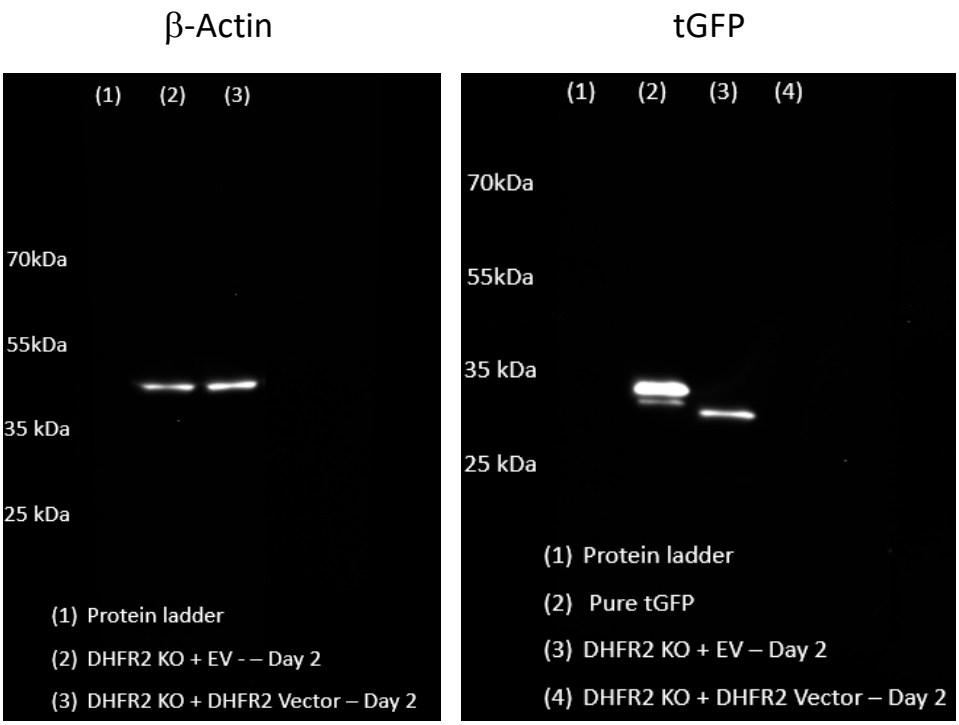


Figure 5.17 DHFR2-GFP protein expression in DHFR2 knockout line. Western blot analysis testing the detection of tGFP in DHFR2 knockout line transfected with either DHFR2 vector (DHFR2-GFP) or Empty Vector (GFP only). No band corresponding to DHFR2-GFP was identified. Pure tGFP was included as a positive control, and β-Actin as quality control. 25 μg protein loaded per sample. Origene tGFP antibody (TA150071), CST β-Actin antibody (8H10010).

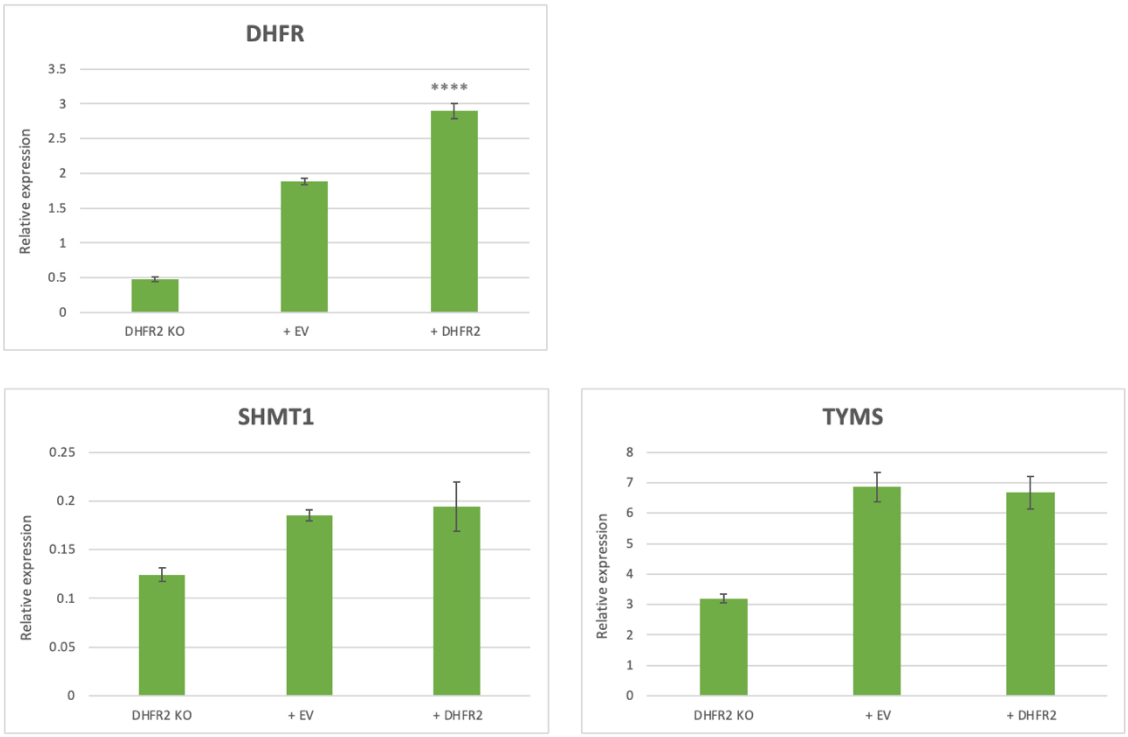


*Figure 5.18 Gene expression after transfection with pCMV6-AC-DHFR2 in DHFR2 knockout cells. DHFR relative expression is the only gene to present a significant increase upon DHFR2 transfection, comparing it to the +EV sample. No significant differences were observed for the other genes, comparing the +EV and +DHFR2 samples. Differences in relative expression ratios were compared using one-way ANOVA. The asterisks designate the statistical significance expressed by the p-value as follows: *P≤0.05, **P≤0.01, ***P≤0.001, ****P≤0.0001, *****P≤ 0.00001. Error bars shown in all graphs are SEM (n=3).*

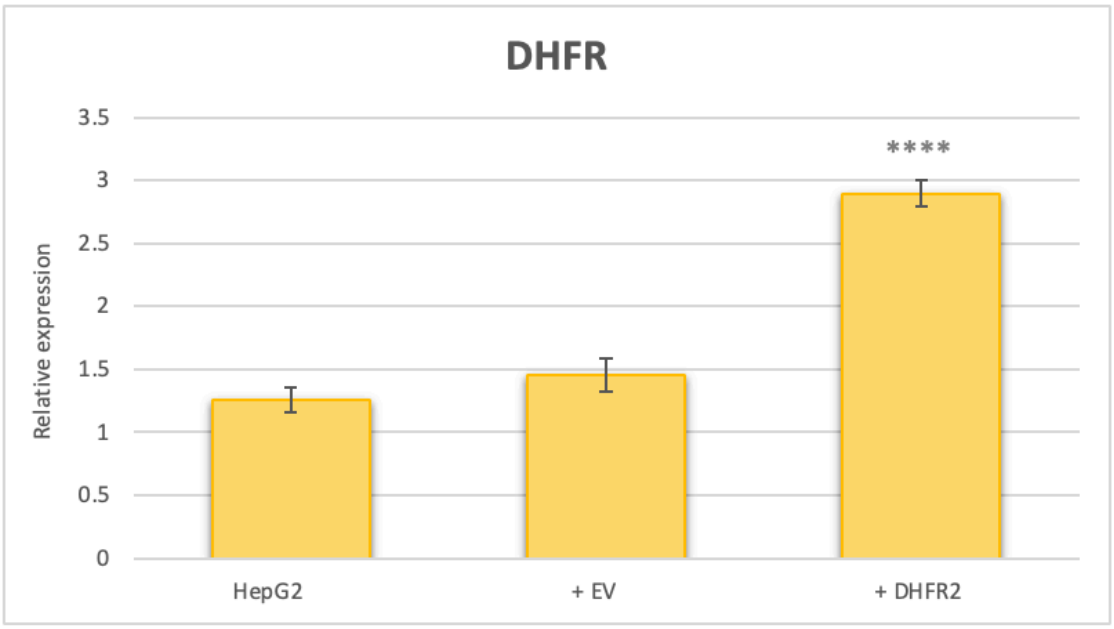


*Figure 5.19 DHFR gene expression after transfection with pCMV6-AC-DHFR2 in HepG2 cells. DHFR relative expression presents a significant increase compared to both HepG2 wild-type and HepG2 + EV (gfp-only), indicating a direct regulatory activity of DHFR2 on DHFR. Differences in relative expression ratios were compared using one-way ANOVA. The asterisks designate the statistical significance expressed by the p-value as follows: *P≤0.05, **P≤0.01, ***P≤0.001, ****P≤0.0001, *****P≤0.00001. Error bars shown in all graphs are SEM (n=3).*

5.4 Discussion

The characterisation of DHFR2 dates back to 2011, when two concurrent papers agreed that the previously annotated pseudogene DHFRL1 (later renamed DHFR2) was actually expressed and active as a dihydrofolate reductase enzyme, in the mitochondria. Both papers demonstrated that a recombinant version of DHFR2 protein was functional as a reductase enzyme and localised in the mitochondria (Anderson et al., 2011; McEntee et al., 2011).

The mitochondrial localisation of the DHFR2-GFP fusion protein in HeLa cells (Anderson et al., 2011) and HEK293 cells (McEntee et al., 2011) led to the conclusion that the immunoblot bands resolved in mitochondrial fractions from HepG2 cells were specific to DHFR2 (Anderson et al., 2011). However, as highlighted in both studies, DHFR and DHFR2 primary sequences are 92 % identical making it impossible for a polyclonal antibody to distinguish between the two isozymes. Therefore, the identification of a DHFR2 protein in HepG2 mitochondria could have been due to DHFR, even though the current literature does not report any mitochondrial localisation for this enzyme. Following the publication of the McEntee paper in 2011, the Parle-McDermott laboratory pursued further analysis of isolated mitochondria from HEK293 cells via mass-spectrometry to guarantee the specific characterisation of peptides that were unique to either of the two proteins. The results demonstrated that no DHFR2 protein was present, but DHFR was identified in the mitochondria for the first time (Bookey et al., in progress).

Considering that a recombinant DHFR2 protein could be made (albeit at a relatively low level) (proteomics data, not shown), but no native enzyme could be found, the experiments presented in this thesis were built to explore the function of the endogenous DHFR2 protein, if it could in fact be detected. Generation of the DHFR2 knockout cell line made it possible to assess DHFR2 function and the pathways connected to it, confirming that its functionality is exerted through its RNA form.

Our overall results suggested that *DHFR2* can regulate *DHFR* expression and indirectly modulate *SHMT1* and *TYMS*, making it an important player in regulating the dTMP synthesis complex for the de novo thymidylate synthesis. Results from the RT-qPCR

investigation resonate with the proteomics profile of the DHFR2 knockout line (results not shown), which finds a significant down-regulation in the RNA and protein levels of these three folate-dependent enzymes.

The de novo dTMP synthesis relies on the subsequential activity of the three OCM enzymes SHMT1, TYMS and DHFR. In particular, SHMT1 is able to convert serine to methylene-THF by transferring a 1C group from the amino acid to a molecule of THF. The methylene-THF is the substrate of TYMS, which passes the activated 1C to dUMP, thus creating dTMP. This reaction releases DHF, which is again converted into THF by DHFR, closing the cycle (Field et al., 2016b). Upon SUMOylation, the three enzymes translocate to the nucleus and assemble into a dTMP synthesis complex (dTMP-SC). The dTMP-SC has been found anchored to the nuclear lamina and in DNA replication sites, providing DNA precursors on-site (Anderson et al., 2007, 2012; Hendriks & Vertegaal, 2016; Murthy & Reddy, 2006; Woeller et al., 2007). This might be especially important in preventing uracil misincorporation (Anderson et al., 2012; MacFarlane et al., 2008, 2011; Paone et al., 2014) and subsequent DNA damage (Blount et al., 1997). The complex has been seen migrating in the nucleus during DNA replication and repair (Field et al., 2018), but it was also found in the cytoplasm (Spizzichino et al., 2020); finally, dTMP synthesis was observed in mitochondria, although using the mitochondrial isozyme SHMT2 (Anderson et al., 2011).

Therefore, we can associate the DHFR2 regulatory function to the nuclear/cytoplasmic de novo dTMP synthesis pathway since all three enzymes of the complex undergo significant down-regulation due to *DHFR2* removal. However, it is curious that the complementation test did not restore *SHMT1* and *TYMS* expression levels, yet it does for *DHFR*. This result may suggest that the main DHFR2 regulatory trait is exerted on DHFR, and only after, the *DHFR* regulation leads to altered *SHMT1* and *TYMS* levels. This apparently ambiguous result could be due to the timing of the analysis; in fact, the complemented line was tested 48 hours post-transfection, and this could not have been enough for the *DHFR* increased RNA levels to induce a subsequent increase in *TYMS* and *SHMT1*.

Although further evidence may be required to clarify the relation between DHFR2 and TYMS/SHMT1, the direct interdependence of DHFR from DHFR2 seems evident. This

evidence is also corroborated by the overexpression study, where adding a DHFR2-expressing vector to a HepG2 wild-type line caused the up-regulation of *DHFR*, at both RNA and protein levels. Overall, these results prove that *DHFR2* directly modulates *DHFR* expression in a proportional way.

A further confirmation that *DHFR2* is required to maintain optimal levels of the enzymes of the One-Carbon Metabolism arrives from the Pathway and Gene Ontology analysis obtained from the RNA microarray data. It was shown that cell division, DNA replication and repair pathways, and pyrimidine metabolism are significantly down-regulated. These cellular functions are tightly linked to the OCM network and its dependent pathways. OCM can be considered the central hub for several metabolic reactions, such as purine and thymidylate synthesis, amino acids homeostasis and methylation reactions (i.e. epigenetics) (Xiu & Field, 2020). In synthesis, the OCM's primary function is to integrate, activate and transfer 1C groups at different oxidation states, using THF as a carrier and enzymatic cofactor (Ducker & Rabinowitz, 2017). Our previous results (RT-qPCR) showed that the main affected pathway is the pyrimidine metabolism, which would be sufficient to compromise DNA replication, repair and subsequent cell division. It is worth noticing how despite the down-regulation of these pathways, the growth of cells does not seem to be significantly impaired. A possible explanation for this might be that we are dealing with a cancerous line - HepG2 – which might have found alternative ways (i.e., salvage pathways) to suppress apoptosis and keep growing despite the possible misincorporation of uracil (MacFarlane et al., 2011). In addition, the chosen statistical method for analysing the growth curves took into account growth trends in their entirety. In contrast, significant differences between DHFR2 knockout and wild-type were found on days 7 and 10 when a traditional ANOVA test was applied (data not shown).

Even though other components of OCM do not demonstrate significant expression imbalances - beyond *DHFR, SHMT1*, and *TYMS* - the global OCM equilibrium could be impaired due to the uneven distribution of folate intermediates (Stover & Field, 2011). The folate enzymes are present in cells at exceeding concentrations compared to their substrates, leading to substrate competition (Clifford et al., 1990; Horne et al., 1989; Kim et al., 1996). Therefore, regulating the metabolic flux among interconnected

pathways is vital and achieved via several mechanisms. The folate intermediates present variable-length polyglutamated tails to keep them inside the cells and improve their affinity for the relative enzymes (Suh et al., 2001). Furthermore, metabolic channelling prevents folate diffusion and improves the overall performance (Stover & Field, 2011). Due to the architecture and multi-layer regulation of OCM, impairment of an enzyme (or enzymatic complex) reflects back to the functionality of the entire network. A prominent example of this phenomenon is the *MTHFR* genetic variant 677 C→T. The reduced efficiency of MTHFR to oxidise methylene-THF to methyl-THF leads to an enlarged pool of methylene-THF, which, in turn, favours the de novo dTMP synthesis at the expense of homocysteine re-methylation (Misselbeck et al., 2017; Quinlivan et al., 2005).

Applying this knowledge to the folate metabolite profiling of the DHFR2 knockout line, it becomes clear that the overall OCM activity is diminished, with a general pattern of halved concentration for all folate intermediates, in either native or polyglutamated forms. This general reduction of folates can be attributable to the lowered DHFR activity. In particular, the conversion of unmetabolised folic acid (from FBS, in the case of cultured cells) to DHF represents the main supply of extracellular folate to the OCM. Hence, when this contribution collapses, the total amount of cellular folate drops.

In terms of the proportion of the single intermediates, we noticed a reduction in the levels of methylene-THF. We suggest that this is due to the lowered SHMT1 activity, which does not convert THF to methylene-THF at an efficient rate. In addition, the reduced TYMS activity leads to minor incorporation of 1C groups from methylene-THF to dUMP (to form dTMP). The impaired efficiency of the dTMP-SC shifts the equilibrium towards the methionine cycle, with a consequent increase in the percentage of methyl-THF. At the other end of the cytoplasmic OCM reactions, formyl-THF is produced starting from mitochondrial formate. We propose that the overall levels of formate decrease because of DHFR impairment and subsequent minor provision of THF. Therefore, the formyl-THF levels are drastically reduced in an attempt to feed the thymidylate de novo pathway (MTHFD1 equilibrium towards methylene-THF). Cells' attempt to support the metabolic pathways can be observed in the 22-fold decreased expression of ALDH1L1 (proteomics data), which converts formyl-THF to THF and CO2+H2O. This reaction,

which eliminates the activated 1C groups (-formyl) from the OCM is tremendously down-regulated in order to redirect formyl-THF towards the cytoplasmic pathways; thus, contributing to its moderate increase despite DHFR's reduced activity.

The ensemble of these results clearly demonstrates that the activity of *DHFR2* is essential in regulating the One-Carbon Metabolism. Although *DHFR2* regulatory effects seem to be directly played on *DHFR*, this single modulation has repercussions on the entire network, with extensive downstream consequences. Even though a regulatory function can be assigned to *DHFR2*, the form through which this regulation is accomplished has yet to be understood. As per our proteomics data (mass-spectrometry global and targeted approach for DHFR2 unique peptides), HepG2 cells (plus several other adult tissue types) do not possess a DHFR2 protein. These findings, along with the functional evaluation of the DHFR2 knockout phenotype, support the hypothesis that *DHFR2* acts as a long non-coding RNA. Moreover, the 2011 Anderson paper reported that *DHFR2* silencing via siRNA in HepG2 turned into abolished reductase activity in the mitochondria (Anderson et al., 2011). Although they logically associated this event with the silencing of the DHFR2 protein, we propose that the inactivation of the *DHFR2* RNA interfered with the endogenous levels of DHFR, with the consequent withdrawal of the mitochondrial DHFR pools. Nevertheless, further investigation is required to confirm this theory and better understand the mechanism of action of *DHFR2*.

After examining the DHFR2 knockout line, the DHFR knock-down line was investigated in an equivalent fashion. The details of the analyses and relative results can be found in the following chapter.

Chapter 6
Functional analysis of the DHFR knock-down in HepG2 cells. Effects of DHFR impairment on DHFR2 and One-Carbon Metabolism

6.1 Overview

Establishing the relation between genotype and phenotype is key in molecular biology. By using a cause-effect paradigm, gene knockout approaches have been central in loss-of-function studies, where the function of a gene could be easily identified. However, uncovering the function of essential genes via gene knockout was not an option due to the resulting lethality associated with the phenotype (Du, 2020). Knock-down models are also included in loss-of-function studies, even though the obtained effect is a partial suppression of the gene function (hypomorphic LOF mutations). The knock-down models were generally obtained by targeting RNAs or proteins (e.g., RNA interference and protein inhibitors), making the phenotypic effect temporary and reversible. Nevertheless, the genome can also be targeted, inducing a permanent perturbation, with long-term effects on the cell phenotype (Housden et al., 2017).

The creation of a DHFR knock-down model was not intentional. The design of the CRISPR/Cas strategy was, in fact, intended to produce a DHFR gene knockout. However, due to the vital function of DHFR, cells with severe disruption of the gene were lethal. On the other hand, cells with a non-efficient expression of DHFR, supplemented with non-essential amino acids and DNA precursors, were maintained in culture. The resulting DHFR knock-down model was an equally powerful tool to explore the influence of the DHFR impairment on DHFR2 and produce supporting evidence of their interconnection and further confirm the essential role of DHFR within One-Carbon Metabolism.

The DHFR knock-down cell line was tested in parallel with the DHFR2 knockout line; therefore, the same set of analyses was performed as described in Chapter 5. This chapter will report the results and discussion relative to the study of the DHFR knock-down line. Specifically, the main aim addressed in this Chapter is to explore the impact of DHFR down-regulation on Cell Growth, OCM and DHFR2.

6.2 Methods

The methods used to investigate the DHFR knock-down cell line are listed in the previous chapter (Chapter 5, Section 5.2).

6.3 Results

6.3.1 The reduced DHFR activity does not impact the HepG2 cell growth

Similarly to the analysis done on the DHFR2 knockout line, the DHFR knock-down line was investigated for cell proliferation (Chapters 2 and 5, Sections 2.2.13.2 and 5.3.1).

The permutation test – run via 'compareGrowthCurves' – of all pairwise combinations of the samples displayed p-values considerably above the significance limit, implying the absence of major differences between the DHFR knock-down line and the relative wild-type cell line (Fig. 6.1). Likewise, no substantial difference was observed between samples treated with additional HT/NEAA versus non-supplemented ones.

In addition, Figure 6.2 graphically illustrates the analogous growth trends for DHFR knock-down and HepG2 wild-type, both with or without HT/NEAA supplementation. The plot shows minor differences in proliferation, with a moderate increase in cell numbers for the supplemented wild-type on day 10.

6.3.2 DHFR imbalances significantly impact SHMT2, ALDH1L1 and DHFR2 expression

The monoclonal DHFR knock-down line was selected for its reduced DHFR presence and reductase activity. Despite the translational levels of the protein appearing to be decreased (Chapter 4, Section 4.5.3), the RNA could easily be detected via RT-PCR (Chapter 4, Section 4.5.2). An RT-qPCR method was employed to quantify the relative expression of the *DHFR* RNA in the DHFR knock-down cell model compared to the HepG2 wild-type. Unexpectedly, the DHFR RNA levels were significantly up-regulated, indicating a self-regulating mechanism attempting to compensate for the drop in the enzymatic levels (Fig. 6.3).

The same six OCM-related genes tested in the DHFR2 knockout line (Chapter 5, Section 5.3.2) were investigated in the DHFR knock-down model to assess if the increment in DHFR RNA and/or the reduction in the cognate enzyme had an influence on regulating the homonymous pathway. It was found that *SHMT2* and *ALDH1L1* presented a significant decrease in the RNA levels, while DHFR2 displays a moderate up-regulation, as illustrated in Figure 6.3. In addition, *SHMT1* (down-regulated) and *SARDH* (up-regulated) show a modest yet significant expression change (data not shown).

6.3.3 DHFR knock-down produces rearrangements of the cellular metabolic pathways

RNA samples from the DHFR knock-down cell line were tested for transcriptome profiling alongside the control line HepG2 wild-type. The global transcriptomic analysis was performed by Arraystar lnc., using their proprietary Human LncRNA Arrays V5, which covers over 60,000 transcripts between mRNAs and lncRNAs. Raw data were normalised and filtered for low-intensity signals. Figure 6.4 illustrates the correlation between HepG2 and DHFR knock-down RNA pools, with 699 up-regulated and 833 down-regulated lncRNAs. The mRNA change in expression is shown in Figure 6.4, indicating 373 up-regulated and 477 down-regulated mRNAs. In both correlations, Pearson's coefficient is above 0.98, indicating a strong positive correlation between the two RNA pools.

These data were subsequently filtered for Fold Change (≥ 2) and p-value (≤ 0.05) to obtain the list of differentially expressed genes. The data were displayed through Volcano plots (Fig. 6.5). The lncRNAs with a significantly different expression profile amount to 542 (307 up-regulated and 235 down-regulated), while the mRNAs are 458 (234 up-regulated and 224 down-regulated).

The differentially expressed (DE) genes were clustered into biological pathways, and each pathway was then evaluated for statistical significance via Pathway Analysis. The relevant pathways were assigned with an enrichment score derived from the p-value, indicative of the change that occurred in the pathway itself compared to the control.

The significant pathways of DE genes for down-regulation presented a relatively low enrichment score range, with highest values of 1.5-2 (enrichment score corresponds to $-\log_{10}$(p-value)). Signalling and cancer-associated pathways were the most affected. On the other hand, pathways affected by the over-expression in DHFR knock-down had a higher enrichment score overall. Drug metabolism pathways presented the biggest change, especially those involving cytochrome p450 and other cellular metabolic pathways. Refer to Figure 6.6 for the specific pathways and relative enrichment score.

Gene Ontology Analysis allowed for a more comprehensive investigation of the biological mechanisms influenced by a particular condition: the DHFR abnormal expression, with enzymatic impairment and RNA up-regulation. GO analysis permitted the study of the significant differential expression patterns within three domains: biological process, molecular function and cellular component.

The GO Biological Process classification indicated that the down-regulated genes in DHFR knock-down line are involved in cell signalling, communication, and adhesion. Consequently, the Cellular Components mostly affected are the membrane and the extracellular space. Finally, the category of Molecular Function sees cation-binding and receptor activity as the functions with the largest number of differentially expressed genes. Pie charts relative to the classification of the three GO domains are displayed in Figure 6.7. The enrichment score for the significant entries of the three domains expressed as a bar chart can be found in Figure 6.8.

Moving to the GO analysis of the up-regulated genes in the DHFR knock-down line, it becomes clear that many metabolic pathways are significantly influenced, such as those relative to small molecules, organic acids (carboxylic acids), and purine and uronic acid metabolic processes. The cellular component mostly affected is the endomembrane system, with molecular functions associated mainly with transferase and homodimerization activity. The pie and bar charts relative to the GO classifications and relative enrichment score are shown in Figures 6.9 and 6.10.

6.3.4 DHFR impairment induces a global decrease in Folate Metabolites

DHFR is a principal constituent of the One-Carbon Metabolism, thanks to its ability to reduce DHF to THF, the main carrier of 1C units. In the DHFR knock-down line, the levels of the DHFR protein are reduced, with a consequent reduction of enzymatic activity. To assess if and how this enzymatic impairment affects the OCM, a Folate Metabolite Profiling was conducted on the cell line at different growth stages and with/without HT/NEAA supplementation. Six types of folate intermediates, among which DHF, THF, CHO-THF (5-formyl-THF), 5-CH3-THF (5-methyl-THF), 5,10-CH2-THF (5,10-methylene-THF), 5,10-CH=THF (5,10-methenyl-THF), and their relative polyglutamated forms were investigated via UPLC-MS/MS methods. The concentration of each metabolite was calculated (in duplicate) and normalised to protein content. Also, the folate concentration was determined in relation to total folate content.

The folate concentrations declined in the DHFR knock-down line at all stages compared to the HepG2 wild-type (Fig. 6.11). The trends of folate concentration seem to maintain a certain proportion for all folate intermediates. The reduction of folate metabolites concentration was more prominent in the non-supplemented DHFR knock-down (yellow bar). Also, the supplemented cells seem to partially recover over time, with higher folate concentrations at passage 33 compared to passage 23 (orange and grey bars). These data are consistent with the microscope observation of cells improving their general aspect over time (cells showed an unhealthy appearance after the knock-down was established). Despite the cells showing a clear recovery tendency throughout time, which can be observed through the general status of the OCM, the sudden subtraction of the HT/NEAA supplementation has caused a rapid reversal of the trend. The total folate of the non-supplemented sample amounts to just ~24 % of that of the wild-type line. A similar trend can be observed in the polyglutamated folate forms, as shown in Figure 6.12.

The proportions of the folate intermediates in relation to total folate showed a discreet consistency, with a few exceptions. THF and Methenyl-THF display the highest percentages in the non-supplemented DHFR kd line; contrarily, those two intermediates show slightly lower rates in DHFR Knock-down P33 and wild-type. Regarding the Methylene-THF, DHFR Knock-down P33 and wild-type seem to have a more similar trend

with higher percentages than the other two samples again. Also, DHFR Knock-down P33 displays the highest percentage of 5'methyl-THF and the lowest percentage of formyl-THF among the four samples (Fig. 6.13).

6.3.5 Transfection of DHFR ORF does not revert the transcriptional profile

A complementation test was performed on the DHFR knock-down cell line to reverse the molecular phenotype and determine if the observed changes were uniquely due to the reduction of DHFR activity. The complementation was obtained by transfecting the pCMV6-AC-DHFR plasmid (Appendix Z) into the DHFR kd line alongside a transfection control (pCMV6-AC-GFP). The pCMV6-AC-DHFR plasmid expressed a DHFR fusion protein with a C-terminal tGFP, while pCMV6-AC-GFP (control) expressed tGFP only. The outcome of the complementation was assessed by examining the presence of green fluorescence and the expression of tGFP RNA and protein.

The observation of transfected cells at the fluorescence microscope indicated the presence of tGFP, proof of successful transfection. As also observed in the DHFR2 complementation experiment, the plasmids presented a peak of expression immediately after transfection (day 2), tending to decrease over time, despite the antibiotic selection (neomycin/G148) (Fig. 6.14).

The tGFP RNA expression was evaluated via RT-PCR (picture not shown) and RT-qPCR (Fig. 6.15), where a maximum relative expression could be observed on day 2 in both transfected lines. The RNA levels drastically dropped on days 6 and 9. For this reason, a Western Blot analysis was conducted on protein samples extracted on day 2 only (Fig. 6.16). The tGFP was identified in both samples, with the band corresponding to DHFR-tGFP fusion protein at a higher molecular weight. Both lanes presented an additional band (at a proportional distance) indicating the formation of dimers.

The investigation of the expression profiles of OCM-related genes (Section 6.3.2) was repeated in the complemented line. Relative expression of DHFR, DHFR2, SHMT1, SHMT2, ALDH1L1 and TYMS was assessed in the DHFR knock-down line, in parallel with the complemented line (+ pCMV6-AC-DHFR) and the control (pCMV6-AC-GFP).

As shown in Figure 6.17, DHFR relative expression skyrocketed in the complemented line compared to DHFR knock-down and DHFR knock-down plus the empty vector pCMV6-AC-GFP. Despite such a significant increase in the levels of DHFR RNA (ORF only + tGFP), none of the genes found to be up- or down-regulated in the DHFR knock-down line recovered their original transcriptional pattern. In particular, DHFR2 RNA was up-regulated equally in the complemented sample and the control, indicating that pCMV6-AC-DHFR does not produce any significant change in DHFR2 expression. In contrast, SHMT2 and ALDH1L1 presented a further decrease in expression levels in the complemented samples. These findings may suggest that the down-regulation of SHMT2 and ALDH1L1 in the DHFR knock-down line was due to the increased levels of DHFR RNA and by adding more DHFR RNA via plasmid transfection, the down-regulation of SHMT2 and ALDH1L1 was exacerbated, leading to a further drop in their RNA levels.

	Group1	Group2	Stat	P.Value	adj.P.Value
1	DHFRkd	DHFRkd Suppl	-0.12836932	0.75	1
2	DHFRkd	WT	-0.19700677	0.68	1
3	DHFRkd	WT Suppl	-0.42681146	0.59	1
4	DHFRkd Suppl	WT	-0.07068987	0.68	1
5	DHFRkd Suppl	WT Suppl	-0.32336379	0.62	1
6	WT	WT Suppl	-0.26264070	0.53	1

Figure 6.1 Permutation test for pairwise comparisons. Test run between groups of growth curves using the algorithm 'compareGrowthCurves' in RStudio. DHFRkd, HepG2 DHFR Knock-down line; WT, HepG2 wild-type line; Suppl, Hypoxanthine + Thymidine (HT) / Non-essential Amino acids (NEAA) supplementation.

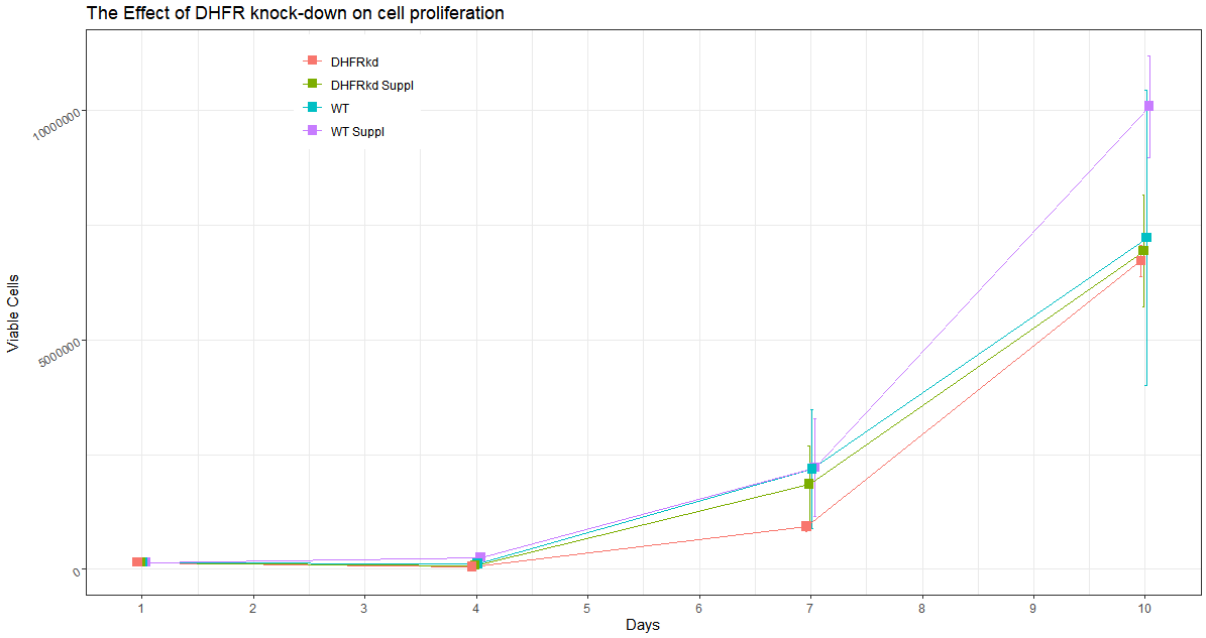


Figure 6.2 Growth curves of HepG2 DHFR knock-down line compared to HepG2 wild-type. Cellular growth curves determined by PI (Propidium Iodide) staining method combined with advanced image analysis by the automated fluorescence cell counter ADAM-MC. All measures were performed in triplicate. DHFRkd, HepG2 DHFR Knock-down line; WT, HepG2 wild-type line; Suppl, Hypoxanthine + Thymidine (HT) / Non-essential Amino acids (NEAA) supplementation.

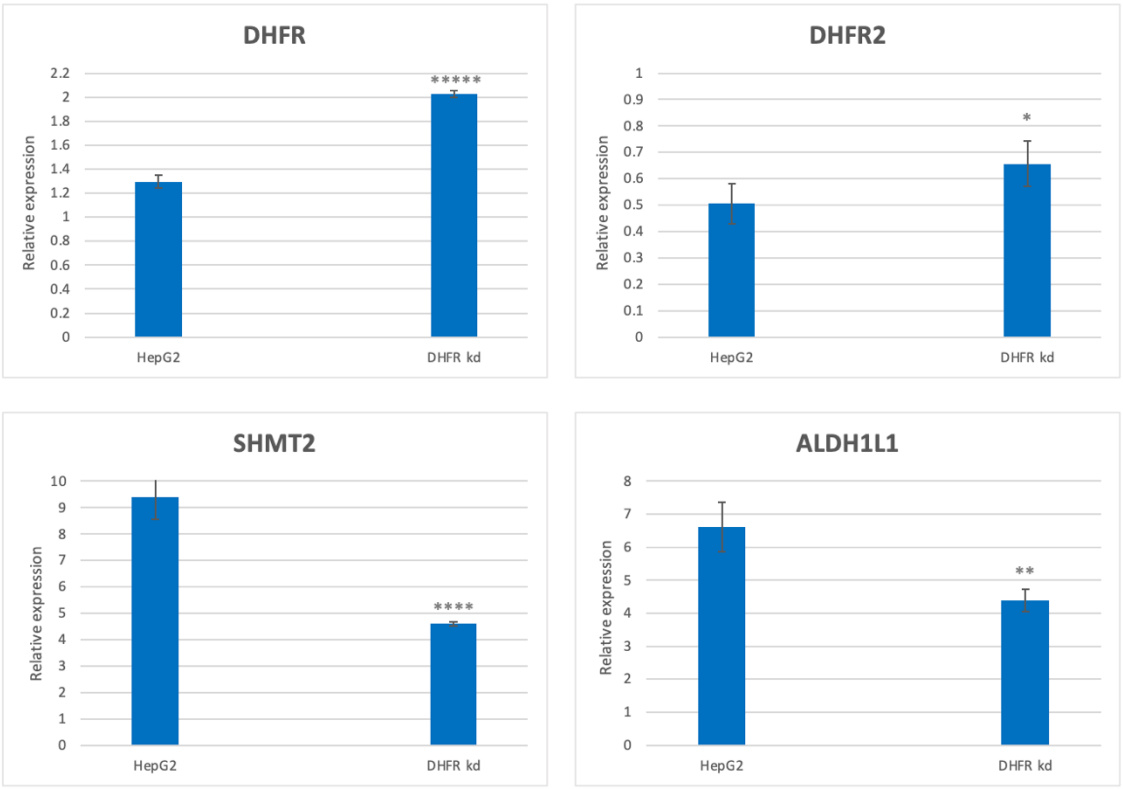


*Figure 6.3 Endogenous levels of DHFR, DHFR2, SHMT2 and ALDH1L1 RNAs in DHFR kd and HepG2 lines. Expression of DHFR RNA is unexpectedly increased in the DHFR kd model. Similarly, the levels of DHFR2 undergo a more moderate rise. Contrarily SHMT2 and ALDH1L1 are significantly down-regulated in the DHFR kd line compared to the HepG2 wild-type line. Differences in relative expression ratios were compared using one-way ANOVA. P-value threshold was set at ≤ 0.05, yet the statistical significance of the expression differences was markedly higher. The asterisks designate the statistical significance expressed by the p-value as follows: *P≤0.05, **P≤0.01, ***P≤0.001, ****P≤0.0001, *****P≤0.00001. Error bars shown in all graphs are SEM (n=3).*

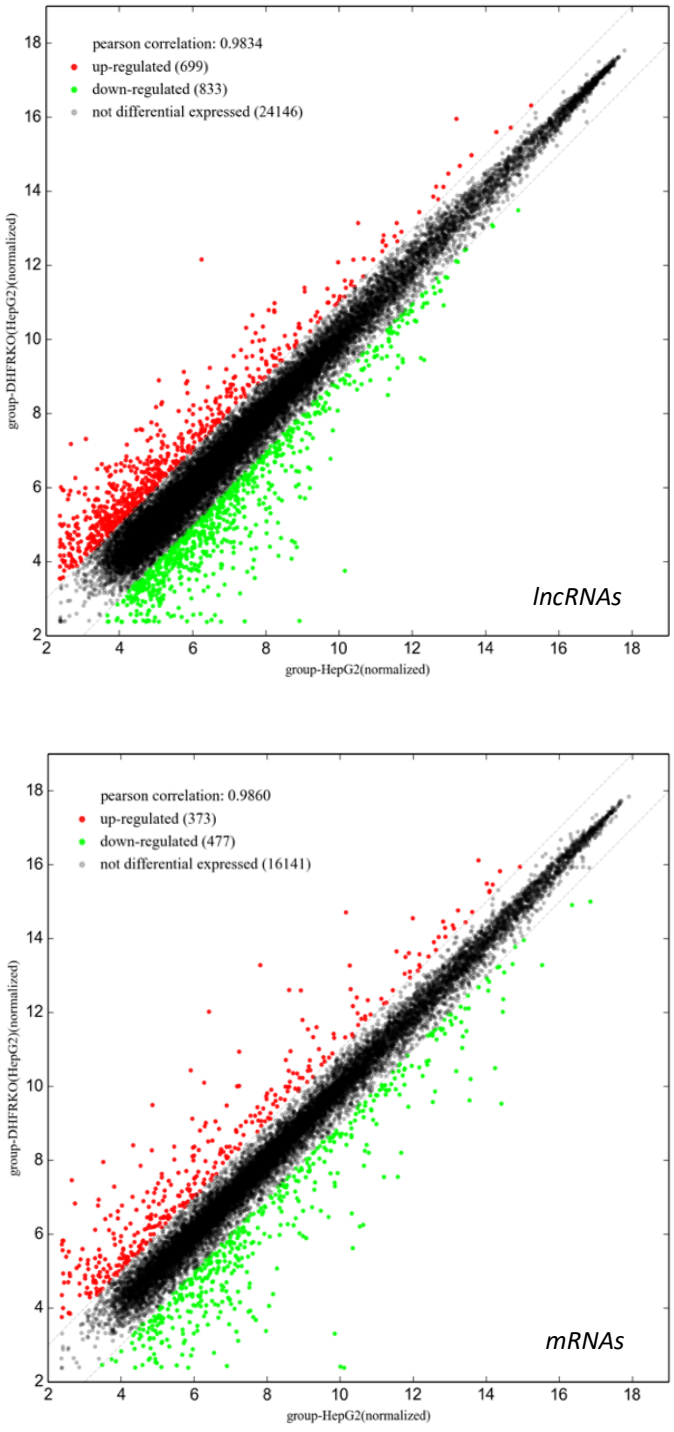


Figure 6.4 Scatter Plot of DHFR kd lncRNA and mRNAs after filtering. Raw signal intensities were normalized by quantile normalization method and low intensity RNAs were eliminated. The filtered RNAs were plotted. The expression variation between DHFR kd and HepG2 can be visualised via Scatter plot. The values of X and Y axes are the averaged normalized signal values of the groups (log2 scaled). The grey dash lines are Fold Change Lines (default fold change value is 2.0). The RNAs above and below the grey dash lines indicated more than 2.0 fold change.

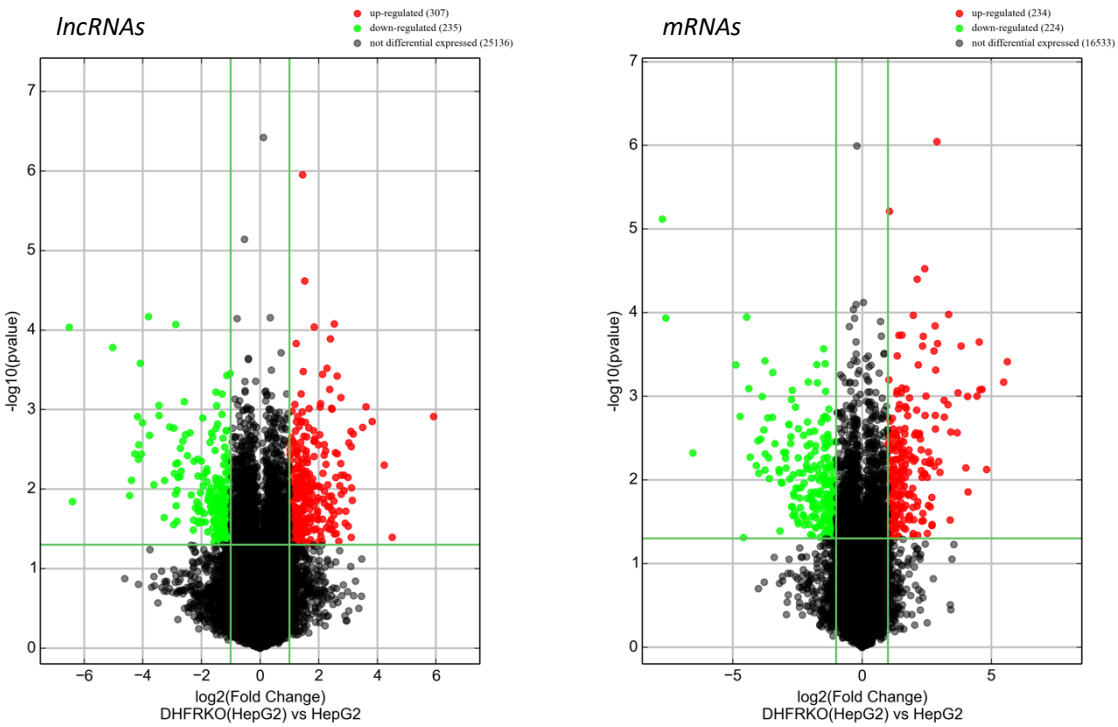


Figure 6.5 Volcano Plot of DHFR kd lncRNA and mRNAs. Differentially expressed lncRNAs and mRNAs with statistical significance were identified through Volcano filtering between DHFR kd and HepG2. The threshold is Fold change ≥ 2 and p-value ≤ 0.05. The Volcano plot displays large magnitude changes which are also statistically significant. On the y-axis is plotted the $-\log_{10}$ (p-value) and on x-axis the $\log_2$ fold change in RNAs expression between the two samples.

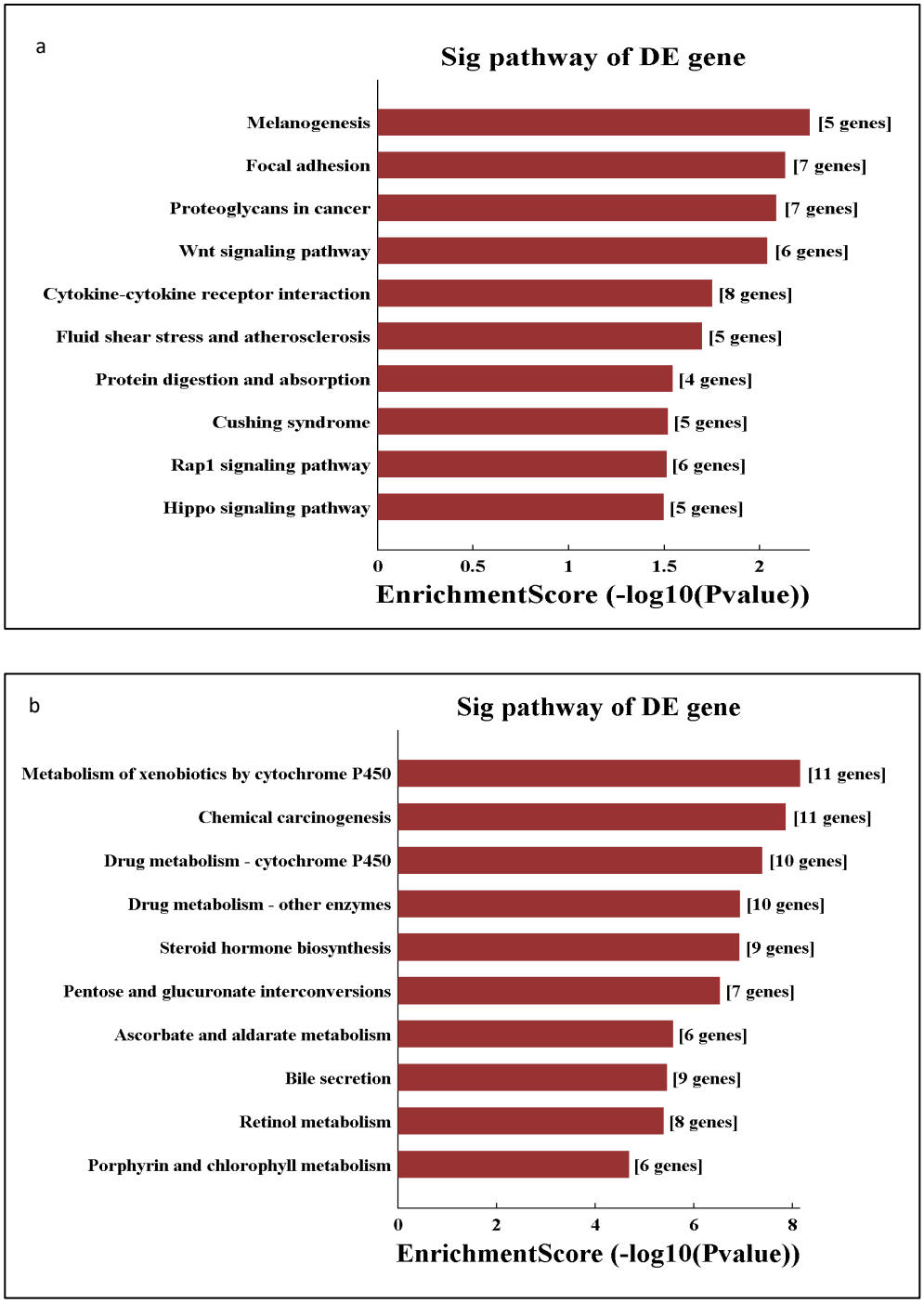


Figure 6.6 Pathway analysis of the differentially expressed (DE) RNAs in DHFR kd. a) Down-regulated pathways, b) Up-regulated pathways. Pathway analysis maps genes to KEGG pathways. This analysis allowed determining the biological pathways with significant enrichment of DE genes. The bar plot shows the top ten Enrichment score values of the significant enrichment pathway.

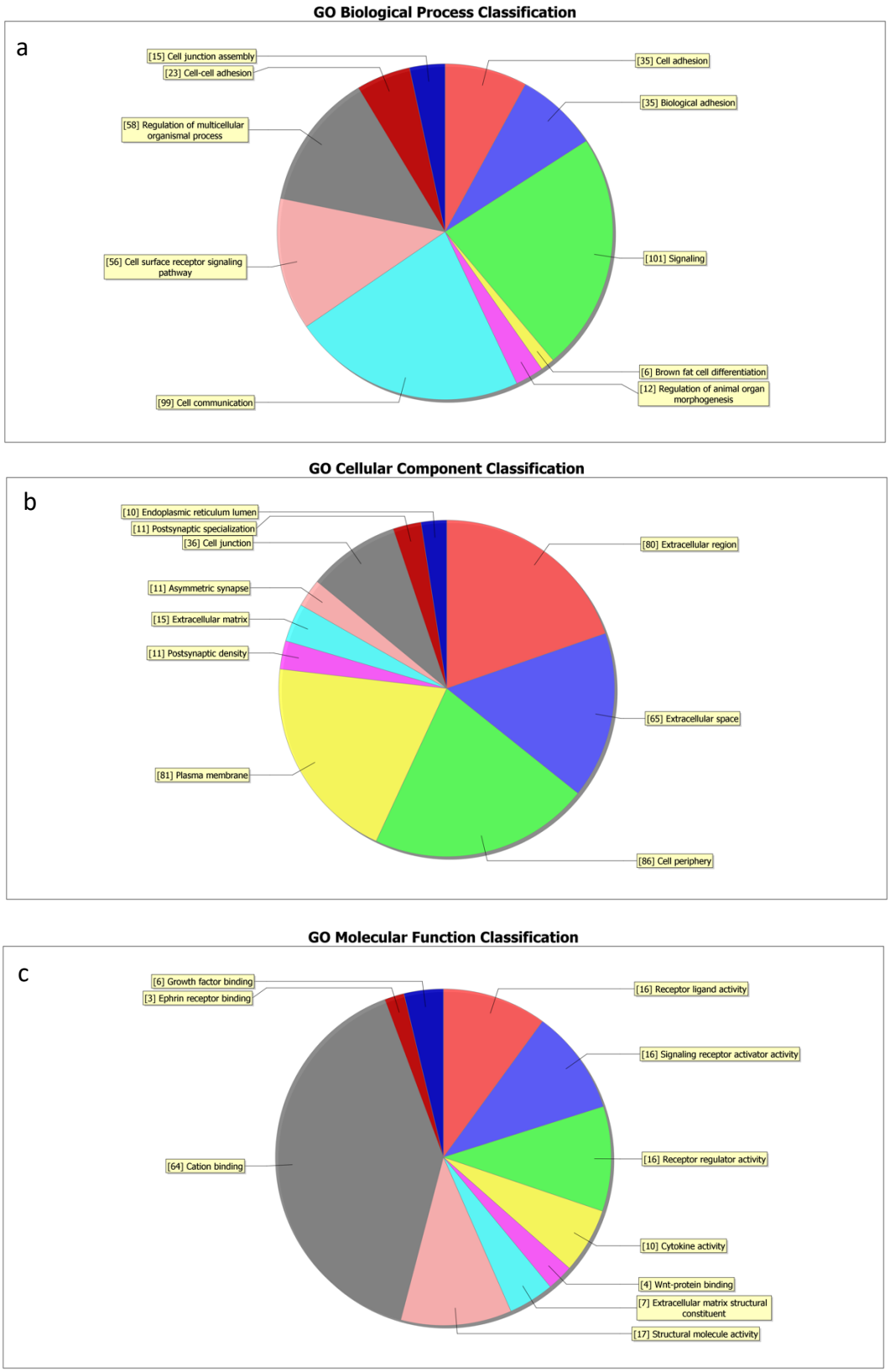


Figure 6.7 GO analysis of the down-regulated RNAs in DHFR kd (Pie Charts). GO (gene ontology) analysis associates DE genes with the three GO categories Biological Processes (a), Cellular Component (b) and Molecular Function (c). The pie charts show the top ten counts of the significant enrichment terms per each category.

Figure 6.8 GO analysis of the down-regulated RNAs in DHFR kd (Bar Chart). The bar plot shows the top ten Enrichment Score values of the significant enrichment terms per each category: Biological Process in red, Cellular Component in green and Molecular Function in blue.

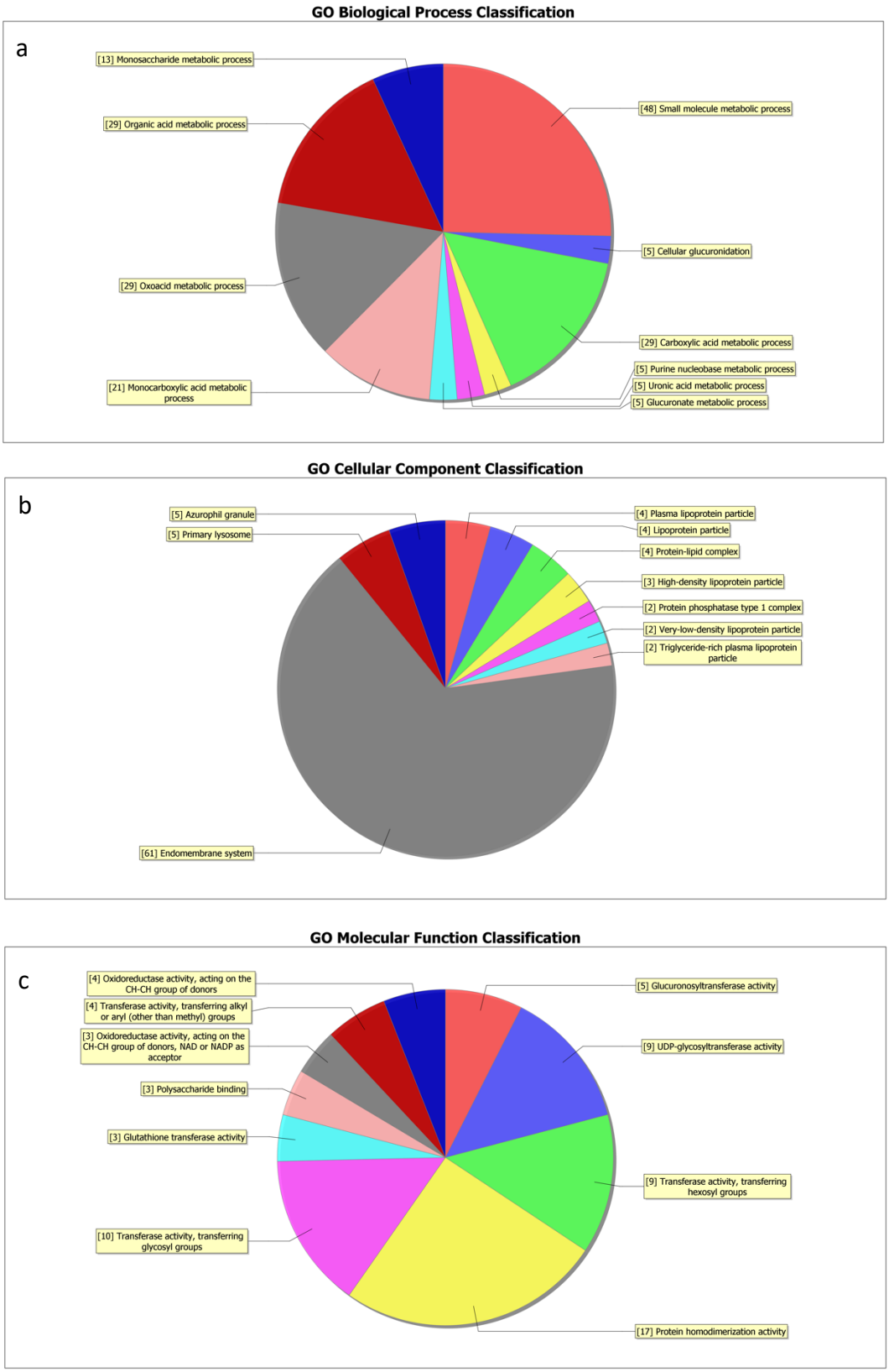


Figure 6.9 GO analysis of the up-regulated RNAs in DHFR kd (Pie Chart). GO (gene ontology) analysis associates DE genes with the three GO categories Biological Processes (a), Cellular Component (b) and Molecular Function (c). The pie charts show the top ten counts of the significant enrichment terms per each category.

Figure 6.10 GO analysis of the up-regulated RNAs in DHFR kd (Bar Chart). The bar plot shows the top ten Enrichment Score value of the significant enrichment terms per each category: Biological Process in red, Cellular Component in green and Molecular Function in blue.

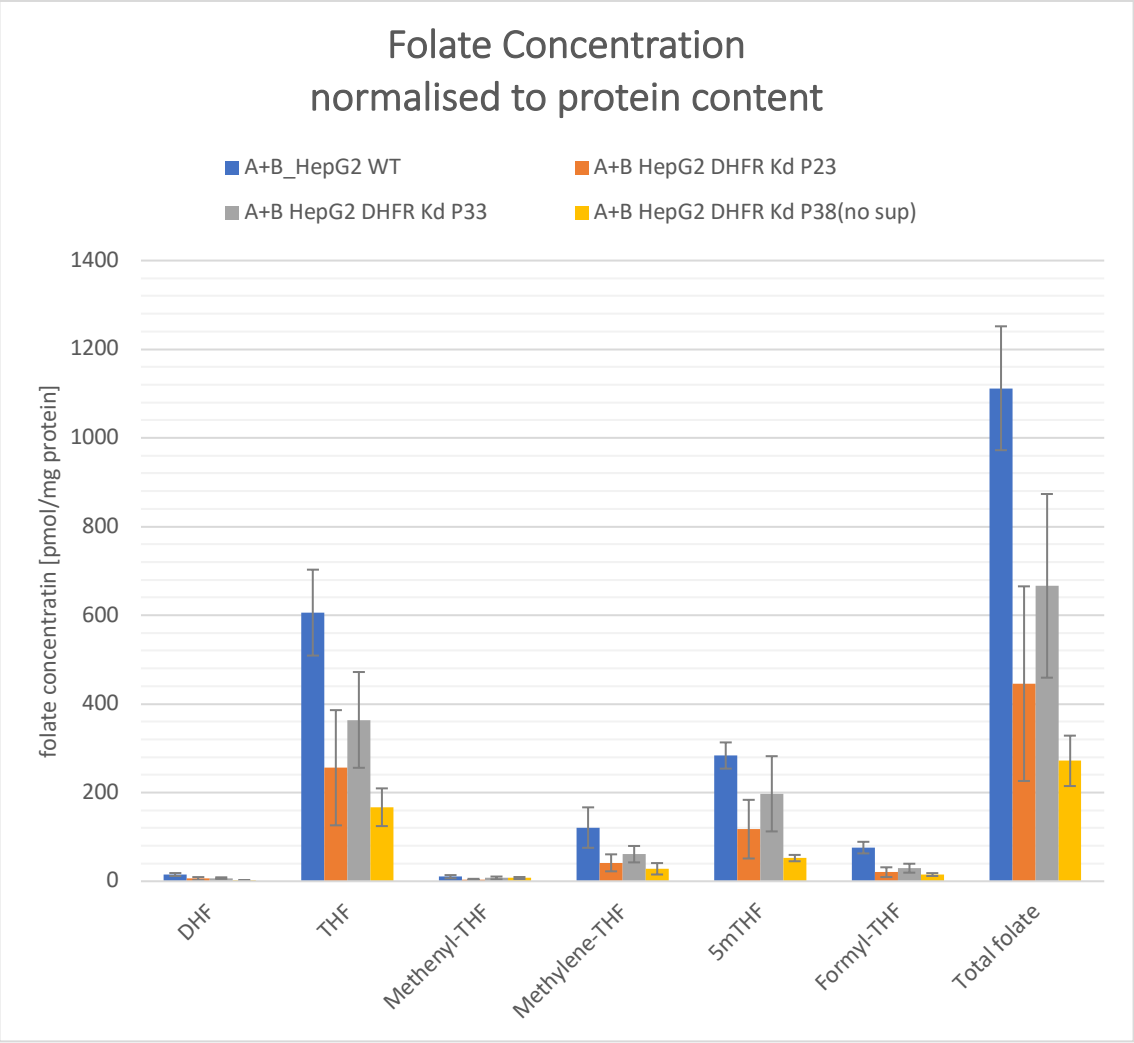


Figure 6.11 Folate metabolites profile in DHFR kd and HepG2 wild-type lines. The DHFR kd line was tested at different points in time (passages 23, 33, 38); HepG2 wild-type line and DHFR kd p38 were grown without supplementation, while DHFR kd p23 and p33 were treated with HT/NEAA supplements. The concentrations were normalised to protein content and expressed as pmol/mg protein. The measures were performed in duplicate.

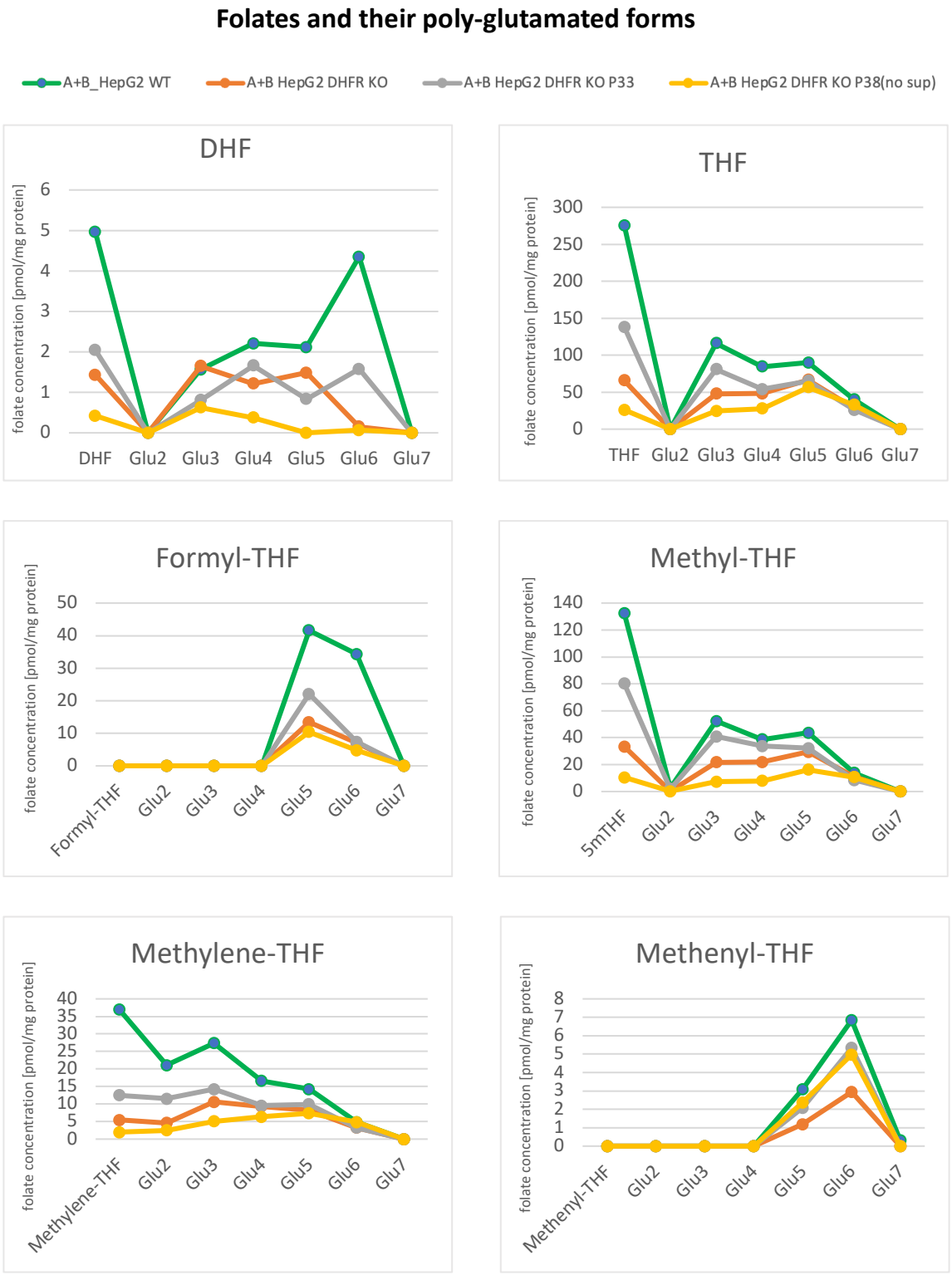


Figure 6.12 Concentration of Folate Intermediates and their relative poly-glutamated forms in DHFR kd (at different times and conditions) and HepG2 wild-type lines. Glu2-7 indicate the number of glutamate molecules attached. The concentrations were normalised to protein content and expressed as pmol/mg protein. The measures were performed in duplicate.

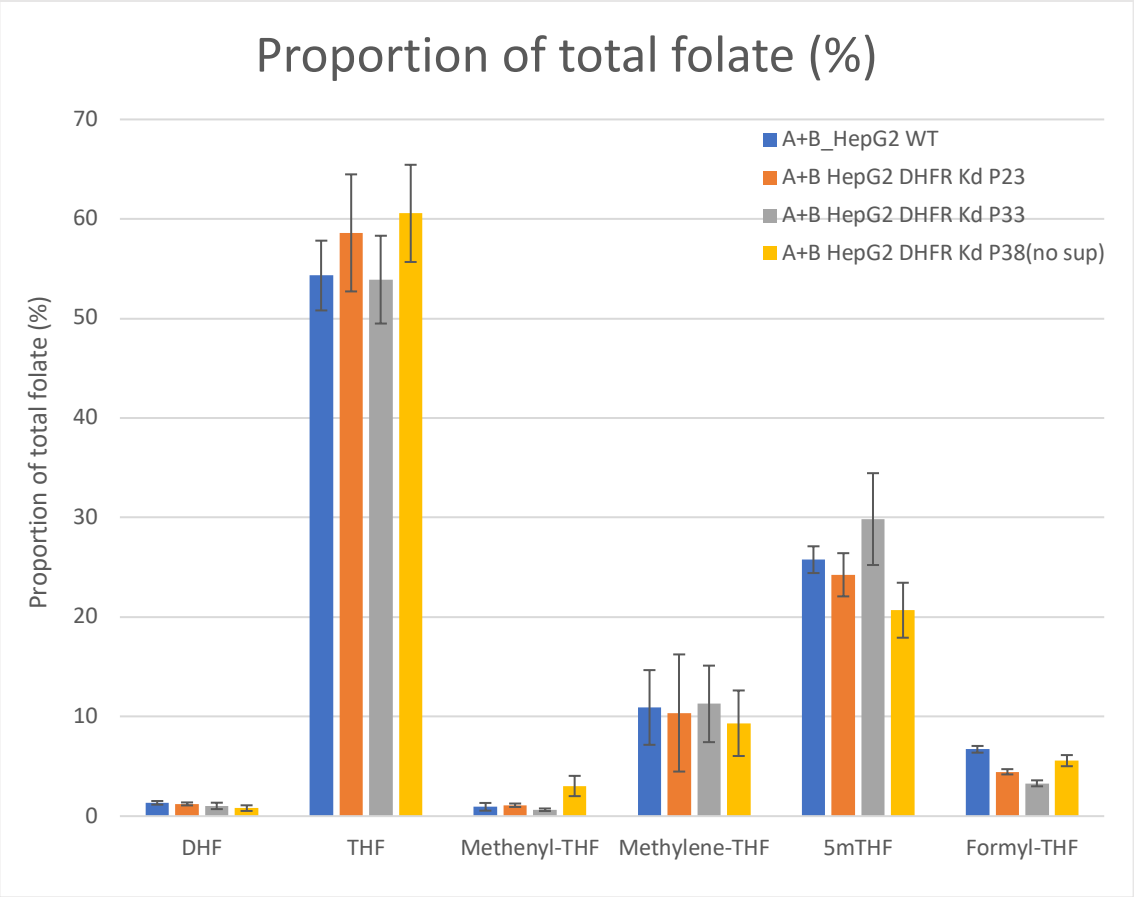


Figure 6.13 Relative proportions (expressed as % of total folate) of folate metabolites in DHFR kd (at different times and conditions) and HepG2 wild-type lines. The percentages for each folate intermediate were calculated considering the sum of all glutamated forms (n1–7).

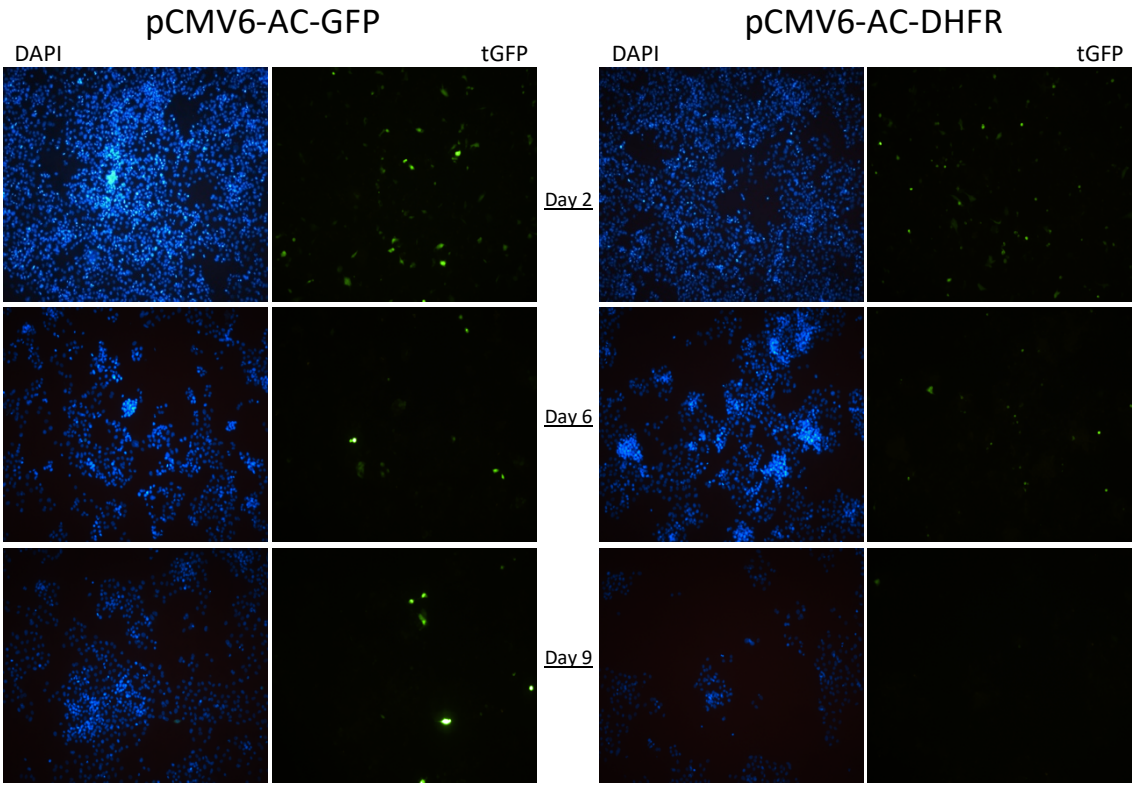


Figure 6.14 DHFR-GFP expression in DHFR knock-down line (Fluorescence). Cells transfected with pCMV6-AC-DHFR(gfp) and pCMV6-AC-GFP (control) were fixed and counter-stained with DAPI at days 2, 6 and 9. Fluorescence was observed using a Leica DFC 500 and pictures taken via the Leica Application Suite software.

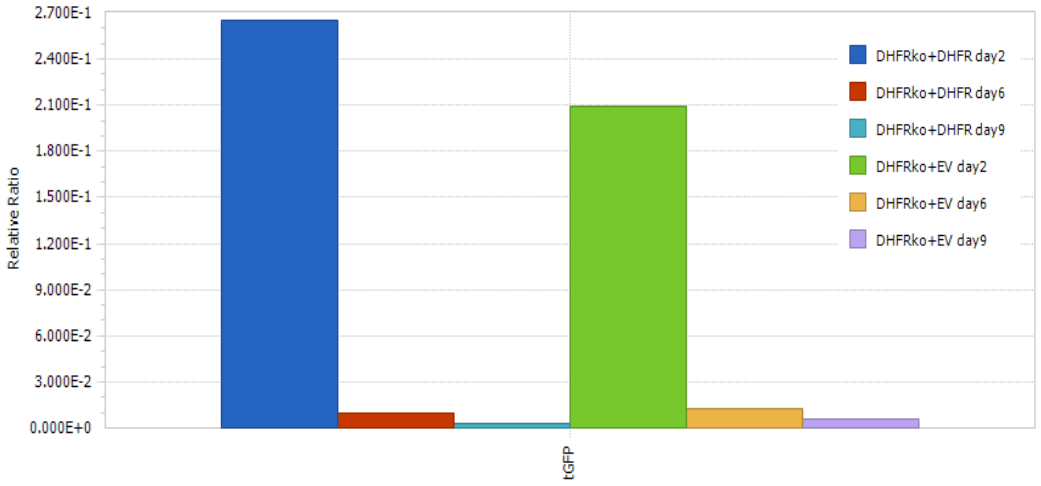


Figure 6.15 DHFR-GFP RNA expression in DHFR knock-down line. RNA samples were isolated 2, 6, and 9 days post-transfection (DHFR-GFP and GFP only) and GFP relative expression was measured against RPS13. DHFRko + DHFR indicates the DHFR knock-down line transfected with pCMV6-AC-DHFR(gfp). DHFRko + EV indicates the same line transfected with the empty vector pCMV6-AC-GFP (negative control).

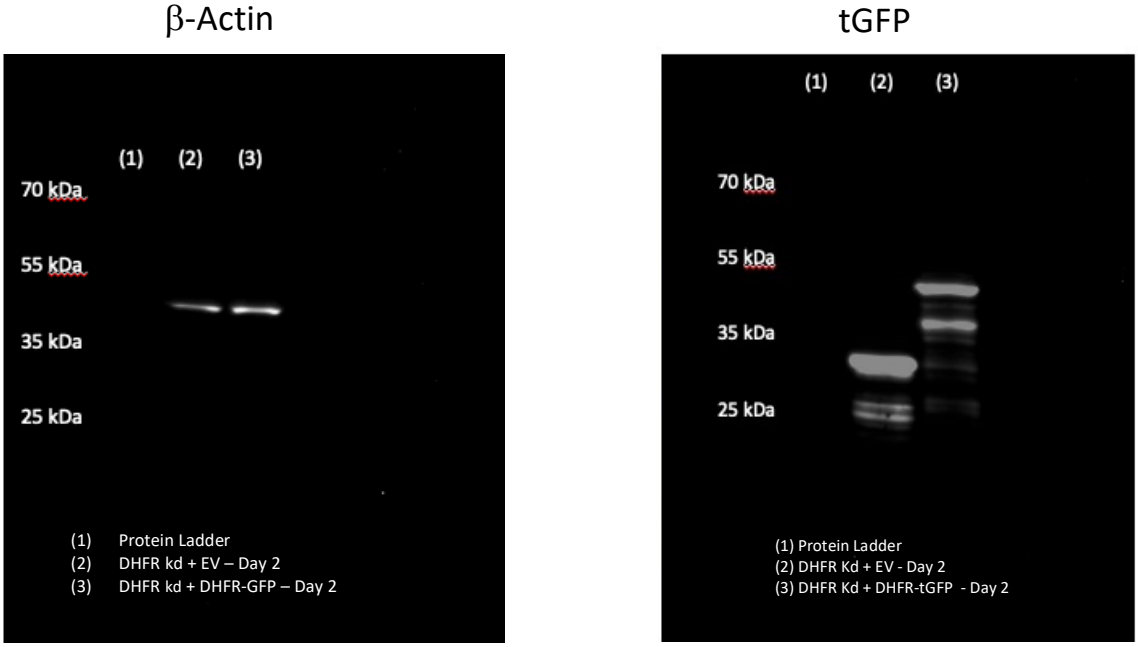


Figure 6.16 DHFR-GFP protein expression in DHFR knockout line. Western blot analysis testing the detection of tGFP in DHFR2 knockout line transfected with either DHFR2 vector (DHFR-gfp) or Empty Vector (gfp only). No band corresponding to DHFR2-GFP was identified. Pure tGFP was included as positive control, and β-Actin as quality control. 25 μg protein loaded per sample. Origene tGFP antibody (TA150071), CST β-Actin antibody (8H10010).

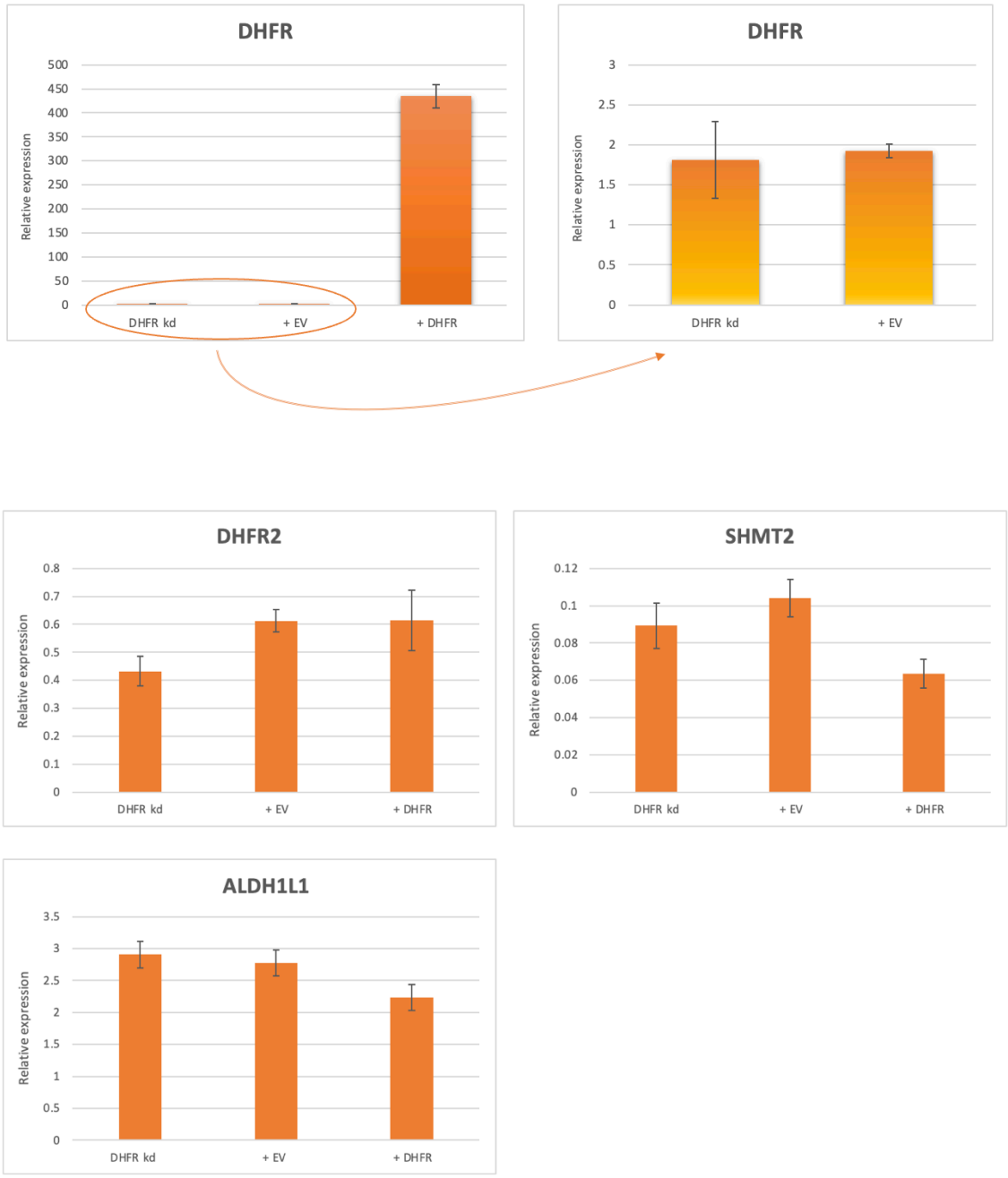


Figure 6.17 Gene expression after transfection with pCMV6-AC-DHFR in DHFR knock-down line. DHFR RNA levels skyrocketed compared to the DHFR knock-down line and the same transfected with empty vector (EV). The expression levels relative to DHFR2, SHMT2, ALDH1L1 are not restored upon DHFR expression, with similar increase between samples transfected with DHFR and GFP-only vectors. The mild rise in expression is, therefore, not attributable to the specific effect of DHFR. Differences in relative expression ratios were calculated using one-way ANOVA. P-value threshold was set at ≤ 0.05. None of the tested genes passed the threshold, comparing +EV and +DHFR samples. Error bars shown in all graphs are SEM (n=3).

6.4 Discussion

Studying the gene's function using loss-of-function approaches can be particularly arduous in the case of pleiotropic and essential genes, where even a partial suppression of expression can result in alteration of unrelated functions (Housden et al., 2017). The Chinese Hamster Ovary (CHO) DG44 cell line (Urlaub et al., 1983) has been the dominant mammalian cell model with a complete deficiency in DHFR activity. However, since the advent of CRISPR/Cas, a few more DHFR knockout models have been attempted (Lee et al., 2021; Li et al., 2019; Mensah et al., 2019).

The HepG2 DHFR knock-down line established in our lab presented a significantly reduced amount of DHFR protein, with a consequent reduction in the reductase activity. Several indels have been reported in the region between exon 1 and intron 1 (Chapter 4, Section 4.6.1). Although we cannot exclude the presence of significant mutations on exon 1 (sequencing reported two ambiguous bases at exon 1)(Appendix P), the majority of mutations are carried by intron 1, which is associated with translational rate and protein stability (Lee et al., 2021; Li et al., 2019; Mensah et al., 2019). Therefore, we proposed that our cell model has an unstable DHFR protein, thus more susceptible to degradation.

DHFR deficient cells would be the ideal system to investigate the presence and function of a hypothetical DHFR2 protein. Furthermore, in the case of an inefficient DHFR activity, cells could be prompted to up-regulate the alternative Dihydrofolate reductase (2) to sustain their metabolic requests, as occurs for SHMT2α (Anderson & Stover, 2009). Hence, the main goal of creating a DHFR-defective line was to use it as a unique system for investigating DHFR2 properties.

Cell viability was assessed before exploring the impact of DHFR knock-down on DHFR2 and other genes and pathways. To our surprise, the effect of the DHFR knock-down on cell growth seems to be marginal, either with or without HT/NEAA supplementation. Comparing this result with the well characterised CHO DG44 line, we knew that cultured cells lacking DHFR necessitated purine and pyrimidine precursors, such as hypoxanthine and thymidine, and glycine to thrive (Urlaub & Chasin, 1980). However, cells will keep growing even without supplements as they can acquire enough nutrients from the

culturing media (*CHO/DhFr- 94060607 | Sigma-Aldrich*, n.d.). Therefore, a (deoxy-) ribonucleosides-depleted media could have been employed to assess better the real impact of the DHFR impairment on cell proliferation.

The experimental evidence on the relative expression of OCM-related genes in the DHFR knock-down line showed the most significant decrease in expression for *ALDH1L1* and *SHMT2*. ALDH1L1 is a cytosolic enzyme that converts formyl-THF to THF and CO2. It is considered a general regulator of the OCM due to its ability to irreversibly remove 1C groups from the folate pool, reducing the overall OCM metabolic capacity (Krupenko et al., 2015). Because of its tumor suppressor activity, *ALDH1L1* was found to be down-regulated in cancer (Krupenko & Krupenko, 2018). The drastic reduction in DHFR activity represents a deficit for the highly proliferating cancerous cells (HepG2). The subsequent down-regulation of *ALDH1L1* can be explained as an attempt to maintain sustainable levels of folate intermediates by avoiding the elimination of formyl-THF, which, in turn, can keep feeding the purine de novo biosynthesis pathway.

The down-regulation of *SHMT2*, on the other hand, could corroborate the idea of a DHFR mitochondrial localisation and participation in the mitochondrial dTMP de novo synthesis, even though a concomitant drop in *TYMS* RNA levels was not reported. In addition, the SHMT2 RT-qPCR assay was designed to amplify an area comprised between exons 2 and 3. This results in the impossibility of distinguishing between *SHMT2* and *SHMT2*α, which localise in the mitochondria and cytoplasm/nucleus, respectively (Anderson & Stover, 2009). SHMT2 is not uniquely needed in the mitochondrial de novo thymidylate synthesis. Still, it also catalyses the first reaction (of four) to produce formate, which is used as a 1C donor in the cytoplasmic side of the One-Carbon Metabolism (Ducker & Rabinowitz, 2017). Serine catabolism initiated by SHMT2 is also crucial in downstream pathways dependent on the OCM, such as methionine formylation of mitochondrial initiator tRNAs. It has been demonstrated that the loss of *SHMT2* disturbs the translation of mitochondrially encoded proteins, with consequences for the mitochondrial respiration (Minton et al., 2018; Morscher et al., 2018). The link between mitochondrial OCM and respiration has been further demonstrated in mouse embryonic fibroblast cells, where the impact was less severe in a folate-sufficient diet. Furthermore, *SHMT2* heterozygosity and folate-deficient diet impaired dTMP synthesis,

as demonstrated by the mtDNA uracil misincorporation (Fiddler, Xiu, et al., 2021a). Therefore, *SHMT2* down-regulation can significantly impact mitochondrial fitness, overall OCM and cellular well-being.

Despite the significant *SHMT2* RNA down-regulation in DHFR knock-down, a parallel reduction in the protein levels was not observed (results not shown). A possible explanation could be the recently discovered SHMT1 regulation via *SHMT2* RNA binding. This regulatory system has a double effect, as fine-tuning the levels of mitochondrial SHMT2 while modulating SHMT1 activity. Moreover, the SHMT1 substrate/s and the RNA compete for the binding (Guiducci et al., 2019). In the light of this evidence, we can suggest that the down-regulation of SHMT1 (RNA and protein) in the DHFR knock-down line is causative of the more sustained *SHMT2* RNA drop. As a possible consequence of *SHMT2* down-regulation, an increase of *SARDH* (RNA) was observed. SARDH catalyses sarcosine oxidation to glycine, which, in turn, can be used as a 1C donor by the Glycine Cleavage System (GCS) with the final formation of methylene-THF (same product of the SHMT2 reaction) (Ducker & Rabinowitz, 2017; Kikuchi et al., 2008; Lionaki et al., 2022). It is plausible to suggest that *SHMT2* down-regulation is at least partially compensated by SARDH activity.

Finally, *DHFR2* RNA shows a moderate up-regulation. Considering the conclusions put forward in Chapter 5 about the regulatory function of *DHFR2* RNA on *DHFR,* the up-regulation of *DHFR2*, consecutive to the DHFR knock-down, corroborates the hypothesis of a co-regulation of the two genes. Hence, the low levels of DHFR protein could induce an increase in *DHFR2* RNA, which, in turn, produces a rise in *DHFR* RNA levels, as also demonstrated by the RT-qPCR results. The rise in *DHFR* RNA, not followed by the same increase in the protein levels, can additionally be explained by the auto-regulatory mechanism typical of DHFR. DHFR can bind its own RNA, which will be released when the equilibrium favours the DHF substrate. As a result, the free RNA could be translated, supporting the cellular need for DHFR enzymatic activity (Abali et al., 2008). Although, in the DHFR knock-down model, the instability of the DHFR protein or its inability to bind its own RNA, coupled with a possible up-regulation due to *DHFR2* modulation, could all contribute to the actual scenario where RNA and protein levels are not proportional, or rather, antithetical (RNA up-regulation and protein down-regulation).

Another essential aspect to consider is the DHFR knock-down's overall effect on cellular pathways and functions. The transcriptomic profile analysis emerged that signalling pathways (Wnt, Hippo, Rap1) tend to be down-regulated with consequences for cell receptors, cell communication, and adhesion. In particular, a significant connection between OCM and the Wnt signalling pathway was recently demonstrated (Albrecht et al., 2019). The Wnt signalling pathway is an intricated and conserved transduction pathway that coordinates paracrine and autocrine communication, modulation of protein catabolism, and control of endolysosomes (Acebron et al., 2014; Albrecht et al., 2018; Hartung et al., 2017; Nusse, 2005; Nusse & Clevers, 2017). The correct functioning of the Wnt pathway relies on SAM (S-adenosylmethionine); therefore, methionine depletion or methotrexate treatment inhibits it (Albrecht et al., 2018, 2019). Other signalling pathways are down-regulated in the DHFR knock-down line. Signalling pathways, in fact, heavily depend on methylation reactions, where SAM is the principal contributor. Together, OCM and methionine cycle allow the redistribution of 1C groups to control cell fate and homeostasis via energy and redox power modulation, genetic and epigenetic control, and signalling (Rosenzweig et al., 2018). We speculate that the drastic reduction in Dihydrofolate Reductase activity determines a re-arrangement of the OCM fluxes, with consequent effects on cell transduction pathways.

On the other hand, the up-regulated pathways encompass a diverse metabolic profile, more typical of a hepatic cell line. Steenbergen (2018) demonstrated that culturing hepatoma cells in human serum (HS) instead of FBS leads to metabolic reprogramming by restoring cellular functions proper to hepatic cells, such as an increase in glycogenesis and β-oxidation, increase in cytochrome p450 metabolic rates, restoration of secretory processes, in addition to a reverse of the cancer metabolic profile (reduction of Warburg effect) (Steenbergen et al., 2018). Although the DHFR knock-down line does not present a differentiation input, its metabolism seems to swap towards the re-establishment of lipid and carbohydrate metabolism and some liver-specific functions, like degradation of xenobiotics and bile secretion in a similar fashion to the HS-cultured cells. Some of the enriched pathways include steroid hormone biosynthesis, pentose and glucuronate interconversions, and ascorbate and aldarate metabolism. As our cells were kept growing in media + FBS, thus, under the effect of growth factors, the cancerous

phenotype was maintained. Still, significant metabolic re-arrangements were made, possibly to sustain the high energy requirement of the cells upon OCM impairment.

A pattern of metabolic adaptation was also inferred by the microscopic observation of DHFR knock-down cells over time. Shortly after the knock-down was established, the cells appeared stressed, with a granular morphology and a higher mortality rate (despite the proliferation rate being sustained). The longer the cells survived in culture, the healthier they looked. Because of this, the folate metabolite profiling was carried out on cells at different moments in time (passages 23 and 33). Moreover, after the cell line was fully stabilised (passage 38), we removed HT/NEAA supplementation and carried out this additional test. The data demonstrate a deficit in all folate intermediates in the three conditions, yet their relative range differs significantly. A severe drop in THF, methylene-THF, formyl-THF and total folate is evident in the early stage knock-down (p23), which appears to be partially rescued after ten passages. Despite the cells being put on a path of recovery, the elimination of HT/NEAA from the media provoked an even more drastic drop in the folates population. These results align with the findings of the implementation of alternative metabolic pathways discussed in the previous paragraph. It could also provide an additional explanation of why the growth curve analysis did not show significant differences.

In summary, the reduction in DHFR activity creates conspicuous impairment of the OCM and subsequent metabolic re-arrangements to ensure cell survival, with promising results in terms of proliferation. Despite the division rate remaining almost unchanged, the DHFR diminished activity closely affected some OCM genes in an attempt to maintain a sufficient functioning of the metabolic network. For example, *ALDH1L* down-regulation saved formyl-THF from being degraded, and *SHMT2* under-expression might have compromised mitochondrial OCM in favour of the cytoplasmic counterpart, ensuring genomic DNA replication.

Despite the application of a mass-spec targeted method, no DHFR2 specific peptides were identified, thus concluding that DHFR2 is not able to compensate for the lack/reduction of DHFR. On the other hand, surprisingly, *DHFR2* RNA levels underwent up-regulation. As the rise in RNA was not accompanied by a corresponding increase/display in protein, our hypothesis of *DHFR2* functioning as a lncRNA is further

supported by this data. In particular, we speculate that the decline in DHFR activity triggers the up-regulation of *DHFR2* lncRNA, which - being a direct modulator of DHFR – induces an increase in *DHFR* expression. This leads to a rise in *DHFR* RNA not followed by an equal surge in the corresponding enzyme (in our cell model). This mechanism would close the regulatory cycle, creating a feedback loop where DHFR is regulated by DHFR2, which in turn is modulated by DHFR (Fig. 6.18). However, the transfection of the DHFR ORF (pCMV6-AC-DHFR) did not lower DHFR2 expression, as expected. This could be due to the lack of regulatory parts of DHFR (e.g., introns or UTRs) that could mediate the sensing response from DHFR2. However, further work is required to confirm the details of the proposed DHFR/DHFR2 regulatory model and resolve the aspects that remained unclear.

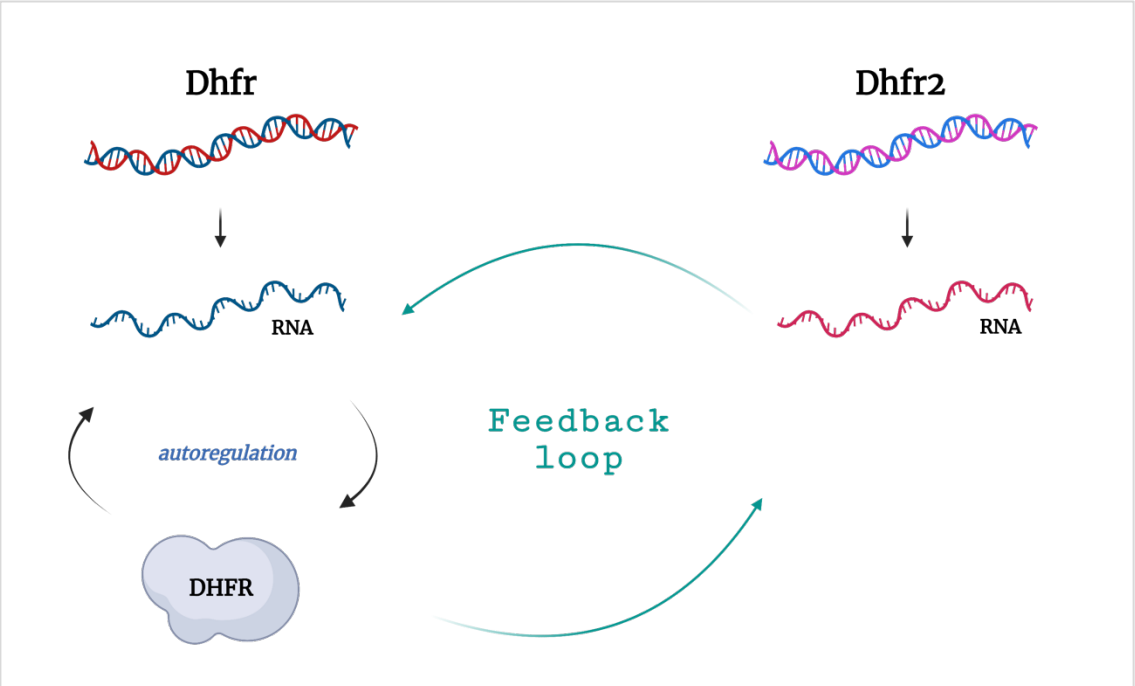


Figure 6.18 DHFR/DHFR2 feedback loop regulation. DHFR2 RNA regulates DHFR RNA (and protein, consecutively). Low levels of DHFR enzyme trigger an increase in DHFR2 which, in turn, leads to a DHFR RNA rise. DHFR enzyme level are intrinsically regulated by the protein ability to bind its own RNA in an auto-regulatory mechanism. DHFR2 RNA might be involved in the auto-regulatory mechanism or act on DHFR RNA/protein independently.

Chapter 7
DHFR2 colocalises with DHFR in human embryonic heart

7.1 Introduction

The analysis of the DHFR2 knockout and DHFR knock-down HepG2 cell lines (Chapters 5 and 6) led to the conclusion that DHFR2 actively regulates DHFR, possibly as a long non-coding RNA. This hypothesis comes from a joint screening of the transcriptional and translational patterns of the edited cell lines. The search for the DHFR2 endogenous protein did not give positive results, despite the use of a powerful mass spectrometry apparatus (Orbitrap Fusion Tribrid Mass Spectrometer for LC-MS/MS and PRM LC-MS/MS) and a targeted approach aimed at DHFR2 specific peptides. The investigation was also carried out on several adult tissue types and stem cells (HepG2, iPSC, IMR32, Huh7, 101B, NEP, placenta, umbilical cord, ovary and testis) to assess the possibility of a tissue-specific expression of the DHFR2 protein. This search was equally unsuccessful.

Several DHFR2 mRNA isoforms contain an intact ORF, which is able to produce a protein (at a low rate), as demonstrated by the proteomics analysis of recombinant DHFR2 in the complementation experiments (Bookey et al., unpublished). Despite the translational potentiality of DHFR2, the endogenous protein could not be identified. Together with the consistent RNA expression (observed in all tissue types) (Chapter 1, Section 1.5), this phenomenon led to the hypothesis that regulatory mechanisms are in place to inhibit DHFR2 RNA translation in order to facilitate its function as a regulatory RNA. Therefore, our research moved towards investigating the untranslated regions (5' and 3' UTRs) of the gene, emphasising secondary structure and post-transcriptional modifications.

Untranslated regions are known to modulate post-transcriptional regulation, including RNA stability, transport and subcellular localisation. Nucleotide patterns forming at UTRs can impact the association with Ribonucleoproteins or regulatory RNAs, leading to specific outcomes (Mignone et al., 2002). In addition, RNA modifications were shown to have a large impact on the formation of higher-order structures, making them an important component in the analysis of UTR-related functions (Leppek et al., 2018).

Our investigation was primarily directed at the two main DHFR2 transcripts, 201 and 202. GenBank (Benson et al., 2015) reports only two transcripts for DHFR2 (NM_176815.5; NM_001195643.2), while Ensembl (Cunningham et al., 2022) shows six

of them (DHFR2 ENSG00000178700). Experimental and bioinformatic data collected by our lab confirmed the existence of Ensembl transcripts, plus additional variants (not shown). Even though most DHFR2 transcripts could be classified as rare, due to their low abundance and tissue-specificity, isoforms 201 and 202 harboured a consistent expression in most of the PCR-screened tissues (HepG2, iPSC, Coriell, HEK 293, HT-29). DHFR2 201 and 202 are almost identical in sequence, with a 123bp additional fragment carried by transcript 201, as shown in Figure 7.1.

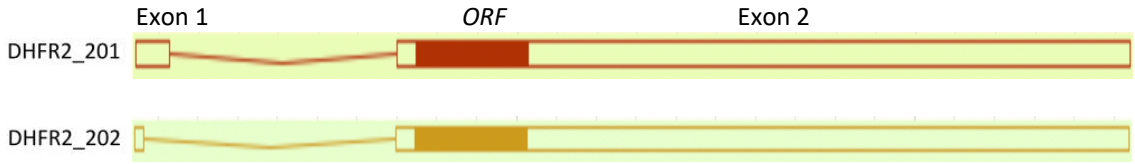


Figure 7.1 Schematic of DHFR2 201 and 202 isoforms. Exon 2, containing the open reading frame, is identical between the two transcripts. Intron 1 of DHFR2 201 is slightly longer than that of 202. Image taken from Ensembl.

7.1.1 RNA modifications are crucial to RNA function

Similarly to DNA, RNA molecules are subject to modifications that play a role in cellular processes. More than 160 types of RNA modification have been identified, opening up the emerging field of epitranscriptomics (Kadumuri & Janga, 2018; Kumar & Mohapatra, 2021). The regulation of RNA stability can be pursued via RNA modification, with consequent regulation of cellular processes such as transcription, splicing, RNA export, translation and degradation. Therefore, RNA modifications can be considered one of the keys to the regulation of gene expression (Boo & Kim, 2020; Nachtergaele & He, 2018).

RNA modifications can be divided into two main groups: chemical modifications and nucleotide substitution. The former encompasses modified bases like N6-methyladenosine, pseudouridine and 5-methylcytidine. The chemical modification of RNA bases may induce structural remodelling leading to altered accessibility for RNPs (ribonucleoproteins). In addition to the type of modification, its position has a role in the interaction with certain protein domains, which generally leads to stabilisation or degradation of the RNA (Roundtree et al., 2017; Zhao et al., 2017). On the other hand,

the nucleotide substitutions are predominantly mediated by deaminase enzymes which convert Cytidine to Uridine and Adenosine to Inosine (Gu et al., 2012). This type of modification is involved in proteomic diversity due to alternative splicing and recoding. It is also associated with RNA stability and degradation due to changes in the secondary structure and miRNAs accessibility. For this reason, nucleotide substitutions are particularly relevant in development, tissue-specificity, immunity and diseases like cancer (Christofi & Zaravinos, 2019; Eisenberg & Levanon, 2018).

7.1.2 A-to-I editing changes the RNA fate

The deamination of Adenosine to Inosine (A-to-I) is a common phenomenon occurring in all Metazoans (Porath et al., 2017; Yablonovitch et al., 2017), which sees Alu elements as the major target of this phenomenon (Athanasiadis et al., 2004). Primates genomes have seen a massive expansion of these transposable elements (DeCerbo & Carmichael, 2005; Deininger & Batzer, 2002; Lander et al., 2001), and as a natural consequence, their transcriptome goes through a vast number of A-to-I substitutions, precisely at the level of *Alu* elements (Bazak et al., 2014; Chen et al., 2008; Schaffer & Levanon, 2021).

The A-to-I editing is mediated by the ADAR (adenosine deaminase acting on RNA) protein family. These enzymes specifically recognise double-stranded RNA (dsRNA) structures, which occur in the presence of Inverted Repeat *Alu*s (IR*Alu*s) (Levanon et al., 2004; Savva et al., 2012). A-to-I editing has vital importance, as demonstrated in a study where patients with mutations in ADAR1 and consequently an impaired A-to-I editing system, presented an abnormal activation of the interferon signalling pathway, leading to Aicardi–Goutières syndrome (Rice et al., 2012).

The editing occurring in the coding region is named recoding-type editing, as it reshapes protein functions (i.e., truncated protein) (Khermesh et al., 2016). This phenomenon is relatively common for ion channels and neurotransmitter receptors (Hood & Emeson, 2012; Rosenthal & Seeburg, 2012). However, the total of mammalian genes subject to this type of modification accounts for ~80 genes (Nishikura, 2016). The majority of the editing sites are located in introns and UTR sequences. Despite the modifications occurring in non-coding regions, A-to-I editing has a major impact on cell biology (Bahn et al., 2012; Hundley & Bass, 2010). Some of the implications are the suppression of the

interferon response, gene silencing via heterochromatin formation, degradation by endonucleases, exonisation of intronic *Alu*s, and paraspeckle retention (Nishikura, 2016; Slotkin & Nishikura, 2013).

7.1.3 A-to-I modifications in the 3'UTR of RNAs mediate paraspeckle retention

The control of gene expression is built on the dynamic interactions between DNA, RNA, and proteins. Specific cell factors play important roles in this interplay through their ability to mediate protein-protein and protein-nucleic acid interactions. The DBHS (Drosophila behaviour/human splicing) protein family act as a molecular scaffold for many cell functions involving gene regulation. The human orthologues are PSF, PSPC1 and $p54^{nrb}$ (Knott et al., 2016). These three proteins were among the first to be identified as components of the paraspeckles (Nakagawa et al., 2018), which are mammalian-specific subnuclear bodies found in the interchromatin space, where they exert a role in control of gene expression (Bond & Fox, 2009). Paraspeckles assemble around the lncRNA NEAT1 (Yamazaki & Hirose, 2015). NEAT1 bears an architectural role; in fact, embryonic stem cells, which do not express NEAT1, lack paraspeckles (Chen & Carmichael, 2009).

In 2001, Zhang and Carmichael demonstrated that A-to-I hyper-edited dsRNAs were retained in the nucleus by $p54^{nrb}$, which holds a binding specificity to Inosine (Z. Zhang & Carmichael, 2001). Specifically, the IR*Alu*s contained in the RNA 3'UTR are subject to A-to-I modifications by ADAR via the formation of dsRNA structures. Upon editing, the Inosines are recognised and bound by $p54^{nrb}$ and retained in the nucleus (Nishikura, 2016). Figure 7.2 schematically illustrates the binding mechanism. The murine gene *mCAT2*, which encodes for an arginine transporter, possesses two mRNAs; one cytoplasmic transcript that provides for basal levels of the protein; the other, named CTN-RNA, is retained in the nucleus in association with $p54^{nrb}$. This alternative RNA presents the same ORF but an extended 3'UTR with embedded *Alu* elements that mediate the nuclear retention. When a stress response is activated, the 3'UTR gets cleaved, and the RNA is quickly transported to the cytoplasm implementing the production of the arginine transporter (Prasanth et al., 2005). A similar mechanism of gene expression regulation was demonstrated in the human gene *Nicn1,* with one

isoform retained in paraspeckles and the other exported in the cytoplasm (Chen et al., 2008).

Comprehensive bioinformatics analysis of the human cDNA and EST databases led to identifying 333 human genes with IR*Alu*s in their 3'UTR (Chen et al., 2008). DHFR2 was one of those genes, with three IR*Alu*s in its 3'UTR, shown to be susceptible to A-to-I editing. This discovery makes it plausible to assume that DHFR2 may be subject to paraspeckle retention, thus explaining the absence of DHFR2 protein regardless of the substantial transcriptional levels that are observed in most cells and tissues.

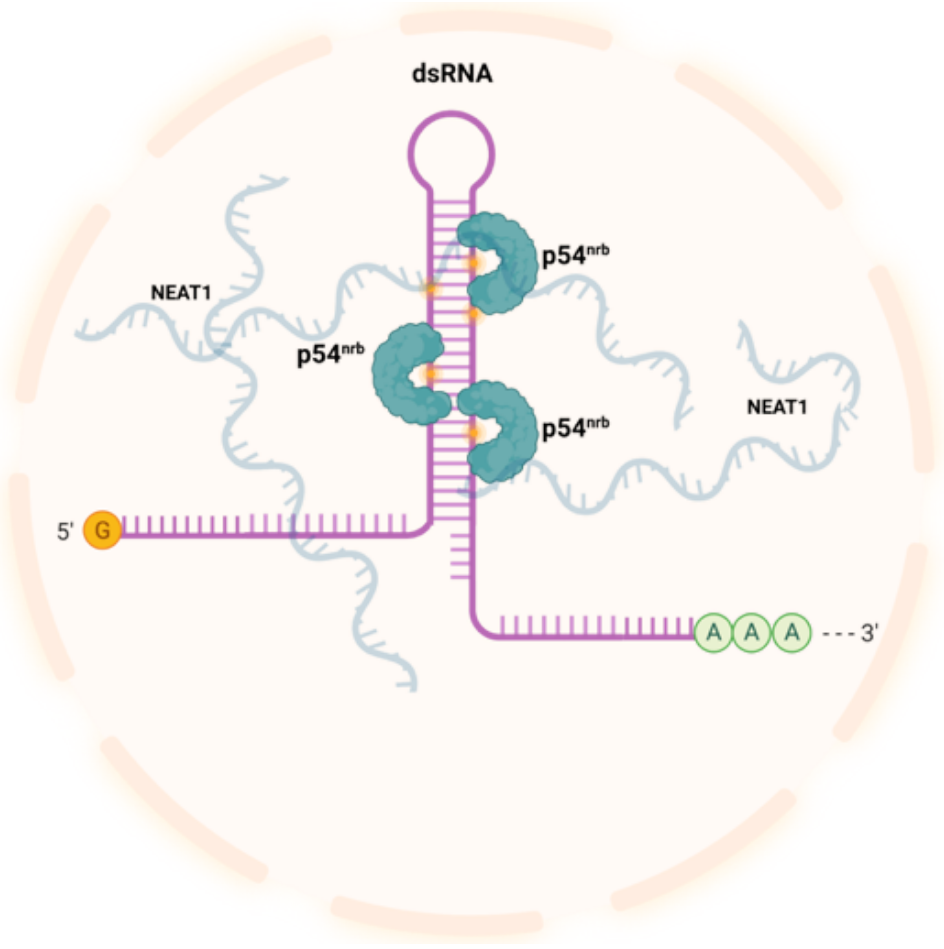


Figure 7.2 Paraspeckle retention mechanism. RNA harbouring Alu elements are prone to form stem-loop structures, targeted by ADAR enzymes leading to A-to-I editing. p54nrb leads to RNA retention by binding to the Inosines (shown as orange dots on the dsRNA). Image created with BioRender.com

7.1.4 Secondary structure impacts RNA stability and function

RNA is a versatile molecule, easily adaptable to many interactions, from ions and vitamins to more complex molecules such as protein and other RNAs. All RNAs can shift among different conformations, with a variable extent of stability (Chełkowska-Pauszek et al., 2021). RNA heterogeneity in shape and interaction can be achieved via the formation of secondary structures (Sharp, 2009; Wan et al., 2011). Intramolecular base pairing is the basic level of secondary structure formation. In conjunction with cofactors and the association of RNPs, the RNA can acquire a tertiary structure and, in turn, a function (catalysis, scaffolding, regulation) (Vandivier et al., 2016). RNA structure is a crucial feature of gene regulation, as RNA dynamic changes can alter the interaction with other cell components, thus impacting biological functions (Wang et al., 2021).

RNA structure forms during transcription and then undergoes modification while migrating from chromatin to cytoplasm. In this process, post-transcriptional modifications and RNPs binding represent the major impact on structural change and the associated function (Lorenz et al., 2016; Mauger et al., 2019; Wong & Pan, 2009). The primary structure of a nascent RNA is thought to influence the RNA polymerase activity with consequent modulation of the transcriptional rate. The more structured 5'UTRs tend to show a lower transcriptional rate. Moreover, RNA secondary structure at ribosome binding sites affects translational rate and capacity. Finally, a higher rate of secondary structure correlates with a shorter half-life (L. Sun et al., 2019).

Since the detection of the DHFR2 protein in many adult tissue types could not be accomplished, the ability of the DHFR2 RNA to bind the ribosomes was investigated. Bioinformatic Analysis of Riboseq Data confirmed that DHFR2 RNA was found to be associated with ribosomes in most human tissues. Experimental confirmation was attempted via Sucrose Cushion Ultracentrifugation on HepG2 cells (experiments performed by Niamh Bookey). The ribosome-bound and ribosome-free fractions were tested for the presence of both DHFR2 main transcripts, 201 and 202. It was found that 201 is predominantly bound to the ribosomes, while 202 is more evenly distributed between the fraction, with a predominance in the ribosome-free fraction.

The differential distribution of the DHFR2 isoforms may indicate possible different functions associated with each transcript. As DHFR2 201 and 202 only differ in a small portion of the 5'UTR, we speculate that they could fold into unique secondary structures, involved in differential functionality.

7.1.5 DHFR2 interactions may be indicators of function

Studying DHFR2 has always been puzzling due to the complexity and peculiarity of the gene. *DHFR2* is expressed, with substantial levels of RNA reported for most human tissues (data available from Protein Atlas) (Karlsson et al., 2021). Despite the remarkable RNA pool, its cognate protein could not be identified. Furthermore, of the two main transcripts - 201 & 202 – the former has been found in association with ribosomes. Both isoforms possess an intact ORF, able to produce a functional protein. Their 3'UTR harbours IR*Alu*s, prone to A-to-I modification and leading to possible paraspeckle retention. In addition, 201 & 202 differ exclusively in their 5'UTRs, which fold in distinctive secondary structures, possibly involved in transcript regulation. Although not in the form of protein, DHFR2 RNA regulates DHFR expression (Chapter 5), likely by interacting with the enzyme itself (McEntee et al., 2011).

Mechanisms of action, regulatory pattern and functionality may be deducted by the localisation and interactions of the DHFR2 transcripts. This final set of experiments aimed to assess the interactions between DHFR2 201/202 and the ribosomes, the paraspeckles, and DHFR. The experiments were carried out on cultured cells first and then repeated on embryonic tissue, addressing a potential divergent behaviour by DHFR2 in the early stages of development. In particular, if the retention in paraspeckles were a correct hypothesis, the absence of these nuclear bodies in embryonic cells could have shown a yet unobserved function of DHFR2.

7.2 Materials and Methods

7.2.1 Hybridisation Chain Reaction (HCR)

RNA-FISH and IHC methods are essential tools to investigate the spatial organisation of molecules in their native environment. The traditional methods use probes labelled with fluorophores or reporter enzymes (i.e., horseradish peroxidase (HRP)-labelled oligonucleotide probes). These methods present limitations, especially for multiplexing experiments (serial amplification of the targets, qualitative staining, diffusion of reporter molecules). On the other hand, the hybridisation chain reaction offers a superior alternative for multiplexed, high-resolution and quantitative IHC and RNA-FISH experiments. The method relies on fluorophore-labelled DNA probes, which self-assemble in an enzyme-free, isothermal polymerisation reaction. In the RNA-FISH, the specificity is acquired via metastable DNA hairpin pairs; once annealed, in couple, with the target RNA, they are recognised by the initiator probe. The IHC specificity depends on the antibodies, with the secondary antibody labelled with the initiator probe. For both RNA-FISH and IHC, the initiator triggers an amplification reaction in which two alternating hairpins (fluorophore-labelled) assemble into a fluorescent amplification polymer (Schwarzkopf et al., 2021). An illustration of the method can be found in Figure 7.3.

7.2.2 HCR experimental design and execution

The design of the experiments revolved mainly around the multiplexing of probes. The best combinations of fluorophores were chosen to not cross-react. Each fluorophore was associated with an HCR amplifier pool, which consisted of the fluorophore-labelled hairpins. It was essential to select the correct amplifiers, with the specific fluorophore targeting the intended molecule. Secondary antibodies, RNA probes and HCR amplifiers were ordered from Molecular Instruments Inc., along with proprietary buffers required for the procedure. Primary antibodies were ordered from separate companies (Chapter 2, Section 2.1.1). A schematic of the probes combinations and relative fluorophores is shown below in Table 7.1.

To address some of the open questions about DHFR2, colocalisation experiments were performed on HepG2 cultured cells and human embryonic tissues. The protocol can be found in Chapter 2, Section 2.2.26. The hybridised and fixed cells were visualised through a Leica TCS SP8 Confocal & STED Super-Resolution Microscope at Nano Research Facility (NRF), DCU. Images were acquired via the Microscope Software Platform LAS X Life Science (Leica) by Dr Úna Prendergast (NRF, DCU). Human embryonic material was obtained from the MRC-Wellcome Trust Human Developmental Biology Resource (HDBR) located at the Institute of Child Health, London. Sections of paraffin-embedded early-stage embryos were prepared by Chloe Santos (ICH, UCL). One to four sections were mounted per slide and de-paraffinized prior to in situ Hybridisation. The HCR IHC + RNA-FISH protocol for FFPE tissue sections was followed (link at Chapter 2, Section 2.2.26). The fluorescence was observed through a ZEISS Axio Observer Microscope (pictures acquired by Chloe Santos). The HCR experiments on human embryonic material were performed at the Institute of Child Health, UCL, London (Professor Nicholas Greene laboratory).

All images were analysed via the JACoP algorithm (Bolte & Cordelières, 2006) in ImageJ (Schneider et al., 2012). Manders' and Pearson's coefficients were calculated to evaluate the degree of colocalisation between two channels. The statistical significance of the resulting coefficients was assessed via the Costes' test.

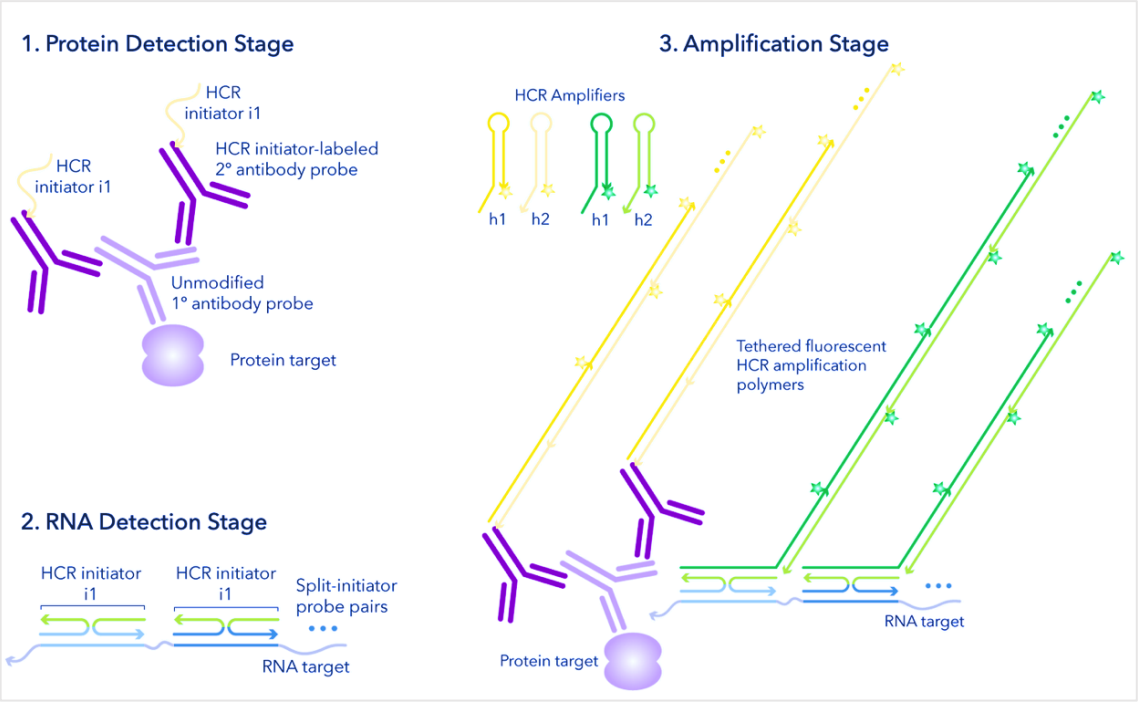


Figure 7.3 HCR IHC + RNA-FISH mechanism of action. The method is divided in three phases which are performed consecutively. 1.Protein detection: standard IHC method with primary antibody targeting the protein of interest; the secondary antibody is labelled with HCR initiator probe. 2. RNA detection: split-initiator probes detect the RNA target in pairs. 3. Amplification: specific initiators trigger self-amplification creating a fluorescent polymer per each targeted molecule. Image adapted from the [Molecular Instrument]() webpage.

Table 7.1 Different probe combinations with relative fluorophore and amplifier pool.

EXPERIMENT 1: DHFR2 RNA LOCALISATION	
1a: Nuclear Retention	**1b: Ribosomal Retention**
PSPC1 (paraspeckle-specific protein) (anti-mouse): Alexa Fluor 514 DHFR2_202 (RNA): Alexa Fluor 594	RPS3 (ribosome-specific protein) (anti-rabbit): Alexa Fluor 488 DHFR2_202 (RNA): Alexa Fluor 594
PSPC1 (paraspeckle-specific protein) (anti-mouse): Alexa Fluor 514 DHFR2_201 (RNA): Alexa Fluor 546	RPS3 (ribosome-specific protein) (anti-rabbit): Alexa Fluor 488 DHFR2_201 (RNA): Alexa Fluor 546
Multiplex all: PSPC1 + DHFR2_202 + DHFR2_201 + Ribosome	

EXPERIMENT 2: DHFR2 RNA INTERACTS WITH DHFR	
2a: Interaction with DHFR Protein	**2b: Interaction with DHFR RNA**
DHFR (protein) (anti-mouse): Alexa Fluor 488 DHFR2_202 (RNA): Alexa Fluor 594	DHFR_main (RNA): Alexa Fluor 514 DHFR2_202 (RNA): Alexa Fluor 594
DHFR (protein) (anti-mouse): Alexa Fluor 488 DHFR2_201 (RNA): Alexa Fluor 546	DHFR_main (RNA): Alexa Fluor 514 DHFR2_201 (RNA): Alexa Fluor 546
Multiplex all: DHFR + DHFR2_202 + DHFR2_201 + DHFR_main	

HCR Amplifier pools			
B1 (AF *488*)	**B2** (AF *546*)	**B3** (AF *594*)	**B5** (AF *514*)

7.3 Results

7.3.1 DHFR2 201 and 202 fold into unique secondary structures

A sequence comparison (Ensembl release 106) revealed that DHFR2 201 and 202 RNAs uniquely differ in their 5'UTR, with isoform 201 owning an additional 123bp. IPknot, software for predicting RNA secondary (Sato et al., 2011), unveiled a notable diversity in secondary structure, with a 202 5'UTR more compactedly structured (Fig. 7.4). The difference in the 5'UTR secondary structure could provide an explanation for the reduced ribosome association by isoform 202 (discussed in Section 7.1.4). Furthermore, certain stem-loops, hairpins and pseudoknots have been reported to pause or stall ribosome movement (Bao et al., 2020; Caliskan et al., 2017; Yan et al., 2015). This phenomenon might explain the binding of DHFR2 201 to the ribosomes, not followed by its translation into protein.

7.3.2 HCR probes specifically bind to their targets

The first step in the in-situ hybridisation experiment was to test each probe individually (and relative HCR amplifier), following the HCR RNA-FISH and HCR IHC protocols. The targeted proteins were PSPC1 (paraspeckles), RPS3 (ribosomes), and DHFR (enzyme), whilst the targeted RNAs were *GAPDH* (control), *DHFR, DHFR2 201, DHFR2 202.* As shown in Figure 7.5, all probes produced realistically specific signals as evidenced by their correct localisation (evident for protein targets) and relative abundance (noticeable in RNA targets). Despite the RNA probes are designed to be highly specific, an additional indication of their efficiency is deduced from the consistency between observed fluorescence and the expression level. Using data from the Consensus dataset for hepatic tissue (The Human Protein Atlas) (Sjöstedt et al., 2020), we assessed the proportionality between normalised RNA expression (nTPM) and fluorescence (by eye). The following RNA expression values, DHFR2 nTPM 6.7; DHFR nTPM 60.1; GAPDH nTPM 2557.9, are consistent with the observed fluorescence.

7.3.3 DHFR2 201 colocalises with DHFR RNA

The HCR RNA-FISH + IHC protocol was designed to allow multiplexed experiments in which both proteins and RNAs could be subjected to hybridisation simultaneously. We tested probes in pairs at first to check if the detection was possible. DHFR2 RNA isoforms 201 and 202 were tested for colocalisation with ribosomes (Fig. 7.6 and 7.7), paraspeckles (Fig. 7.8 and 7.9), the DHFR enzyme (Fig. 7.10 and 7.11), and cognate RNA (Fig. 7.12 and 7.13). The detection of protein targets appeared to be particularly abundant, despite the minimal working concentration for primary antibodies. The signals coming from the DHFR2 RNA targets were less ubiquitous, making it easier to detect possible interactions as per DHFR RNA and DHFR2 201, which show overlapping localisation (Fig.7.12). To ensure the colocalising signals are accurate, the images were exported in ImageJ, and the colocalising signals were analysed using the JACoP function in ImageJ. A comprehensive list of the statistical tests for each probe pair is supplied in Appendix AA. In addition to the probe pairs listed above, DHFR protein + DHFR RNA colocalisation was tested and served as a control due to the documented interaction between the two (image not shown) (Abali et al., 2008; Ercikan-Abali et al., 1997; McEntee et al., 2011). In fact, Pearson's coefficient amounted to 0.94, indicating a strong correlation. Contrarily, none of the investigated pairs showed Pearson's coefficient above 0.9, except for 'DHFR RNA + DHFR2 201' with a coefficient equal to 0.989. This data highly suggest that this interaction could be the key to the regulatory mechanism by which DHFR2 regulates DHFR.

7.3.4 Successful multiplexed detection of four protein/RNA probes

The HCR technology is suitable for multiplexing, with several probes used at once (Schwarzkopf et al., 2021). A four-probe experiment was attempted aiming to reduce variability among the samples and have all probes hybridise the same group of cells. Figures 7.14 and 7.15 illustrate the multiplexed detection of PSPC1, RPS3, DHF2 201 & 202, and DHFR (enzyme & RNA), DHFR2 201 & 202, respectively. The images show the same expression patterns compared to the previous experiment and no additional information. Thus, this can be interpreted as an indicator of the robustness of the method.

7.3.5 DHFR and DHFR2 colocalise in human embryonic heart

The HCR experiment was originally designed to compare the localisation of DHFR2 transcripts in both HepG2 and embryonic cells, particularly concerning paraspeckle retention. While we performed the above experiments in HepG2 cells, cultured embryonic cells could not be easily provided within a restricted time. Alternatively, paraffin-conserved embryos were investigated. Even though the two sets of experiments are not directly comparable, exploring DHFR2 behaviour in embryonic tissues was deemed a unique opportunity.

The initial hypothesis of DHFR2 being retained in the paraspeckles could not be strongly confirmed in HepG2 cells, even though PSPC1 and DHFR2 201 presented a Pearson's coefficient of 0.683, which indicates a certain degree of correlation. Since paraspeckles do not assemble in embryonic cells (Chen & Carmichael, 2009), the DHFR2 RNAs could exhibit a particular activity in undifferentiated cells. Two multiplexed experiments were performed on human embryos (whole mount). DHFR2 201 & 202 were tested in one experiment along with PSPC1 (paraspeckles); in another experiment, the two DHFR2 RNAs were multiplexed with DHFR RNA and protein. Unfortunately, a software issue related to the detection of fluorophore AF 514 did not allow the observation of PSPC1 and DHFR RNA. Therefore, the only available images are those relative to DHFR2 201, 202 and DHFR protein. As shown in Figure 7.16, the major expression of all three probes is concentrated at the heart level. Although this initial discovery needs to be further investigated, it provides precious data on DHFR2 tissue-specificity during embryonic development. Moreover, this observed differential expression may be likely linked to a specific function exerted by DHFR2 in embryonic development only.

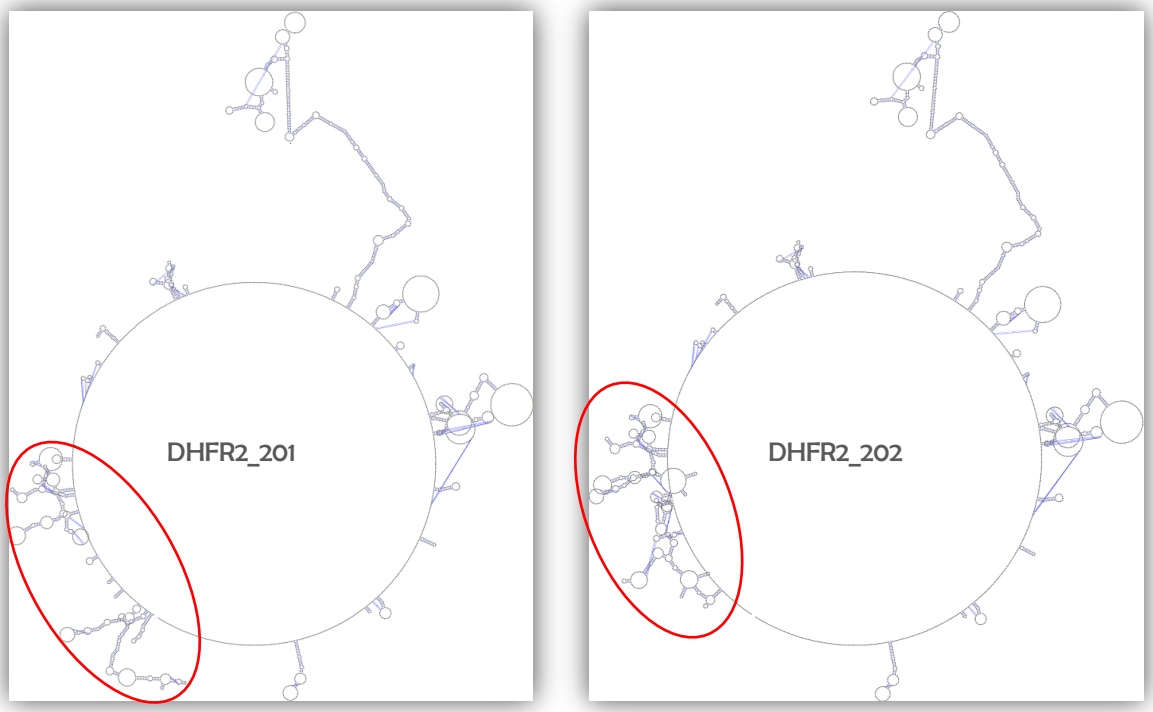


Figure 7.4 DHFR2 201 & 202 secondary structure prediction. The 5'UTRs are circled in red. Image created with IPKnot web server.

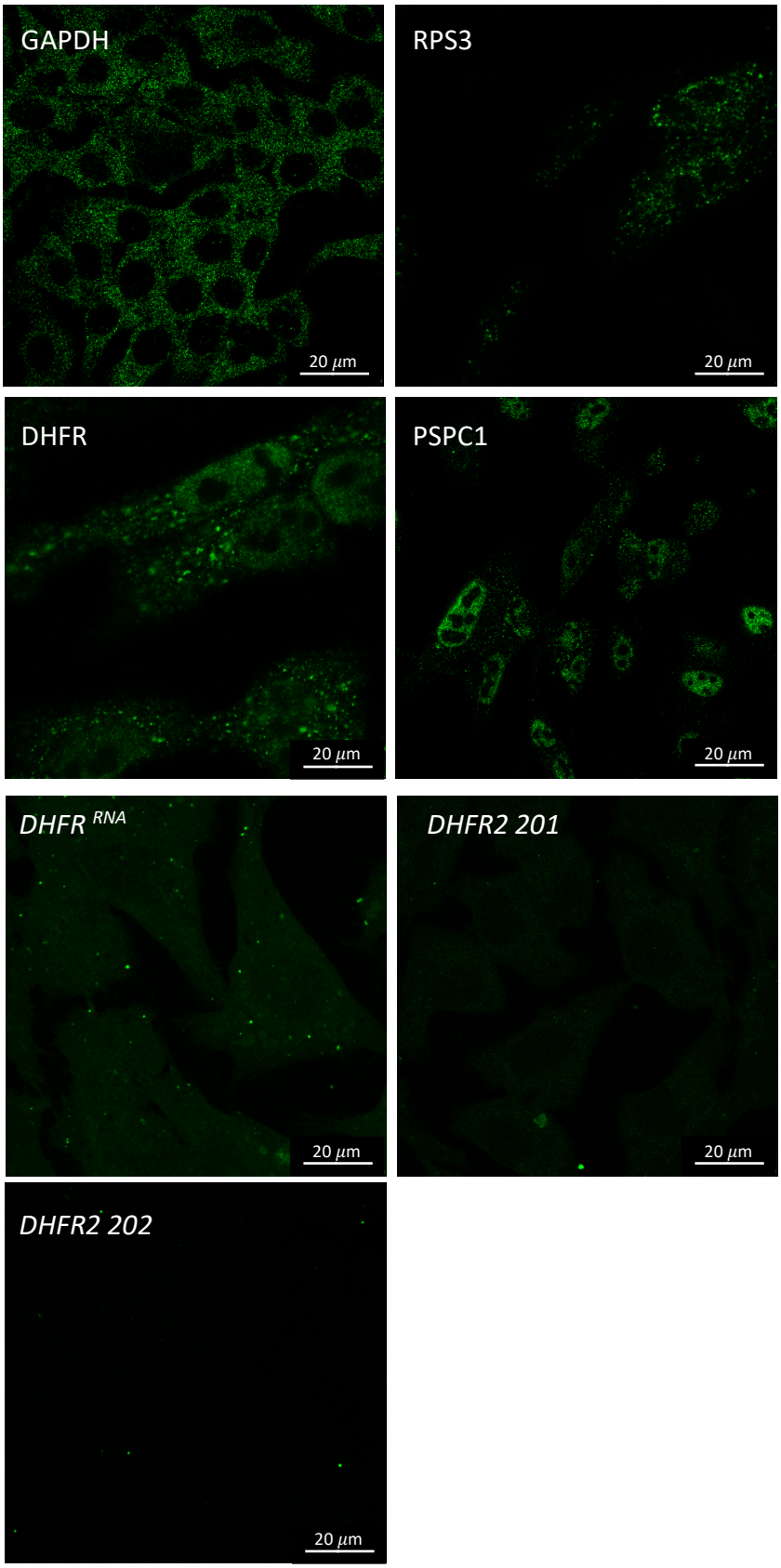


Figure 7.5 Confocal imaging of proteins and RNAs detected via HCR IHC and HCR RNA-FISH for single probes in HepG2 cells on a slide. Target proteins: RPS3 (AlexaFluor 488), PSPC1 (AlexaFluor 514), DHFR (AlexaFluor 488). Target RNAs: GAPDH (AlexaFluor 594), DHFR ^{RNA} (AlexaFluor 514), DHFR2 201 (AlexaFluor 546), DHFR2 202 (AlexaFluor 594). Magnification 100X.

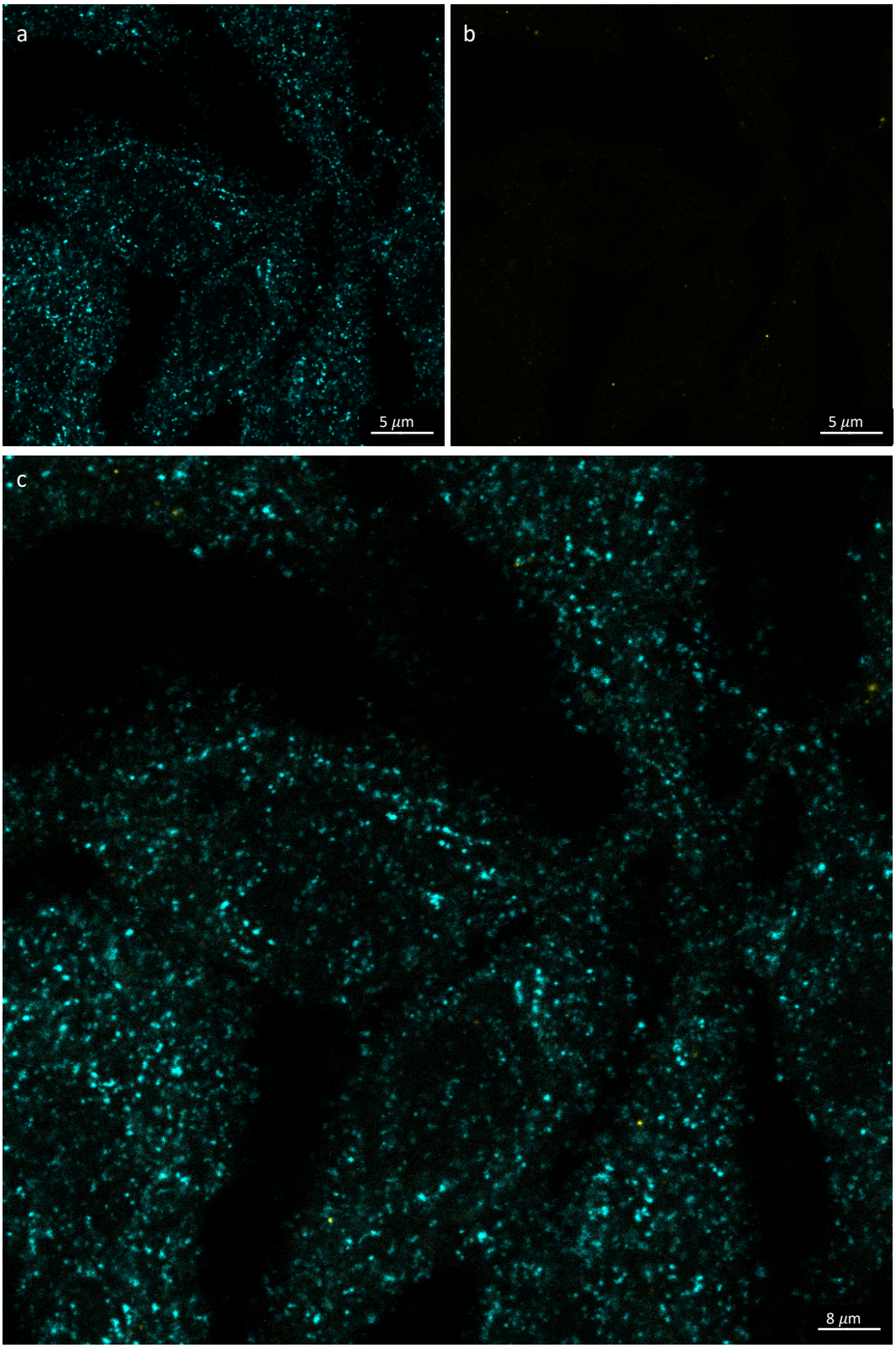


Figure 7.6 Confocal imaging of RPS3 and DHFR2 201 protein:RNA 2-plex detected via HCR IHC + RNA-FISH in HepG2 cells on a slide. 65×65 µm pixels. Targets: RPS3 (AlexaFluor 488) and DHFR2 201 (AlexaFluor 546). a. RSP3 (cyan); b. DHFR2 201 (yellow); c. Merge. Magnification 100X.

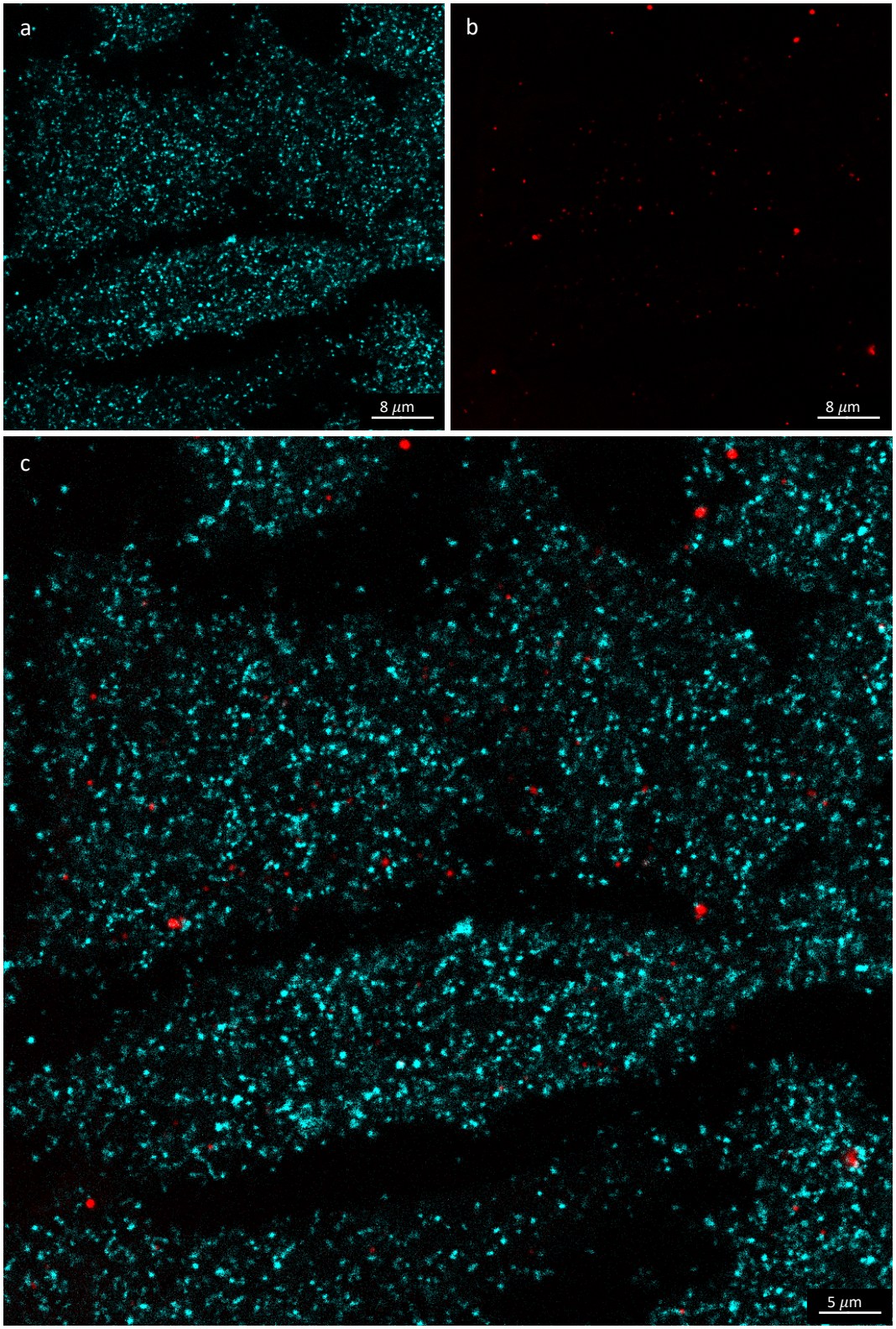


Figure 7.7 Confocal imaging of RPS3 and DHFR2 202 protein:RNA 2-plex detected via HCR IHC + RNA-FISH in HepG2 cells on a slide. 65x65 µm pixels. Targets: RPS3 (AlexaFluor 488) and DHFR2 202 (AlexaFluor 546). a. RSP3 (cyan); b. DHFR2 202 (red); c. Merge. Magnification 100X

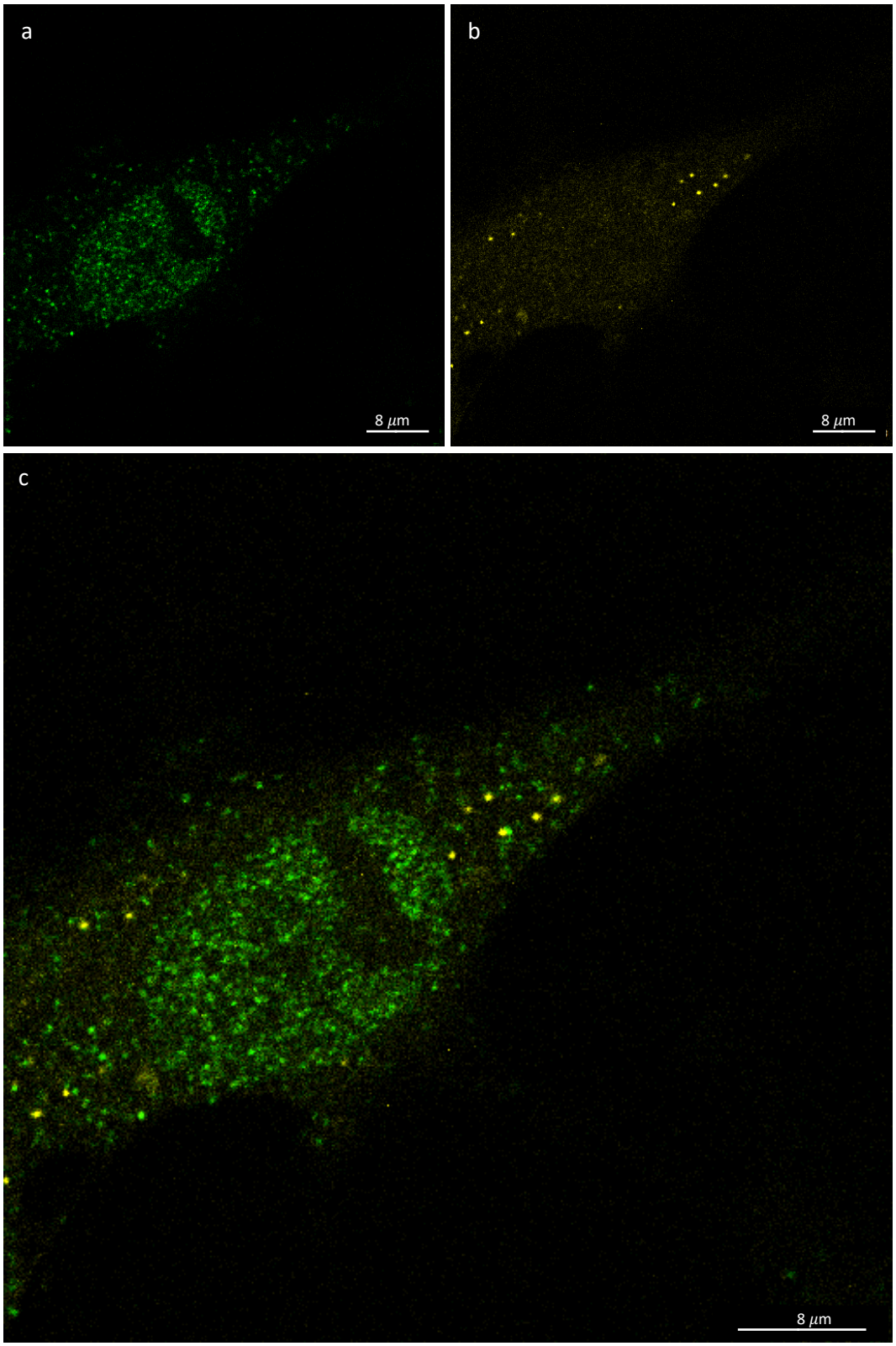


Figure 7.8 Confocal imaging of PSPC1 and DHFR2 201 protein:RNA 2-plex detected via HCR IHC + RNA-FISH in HepG2 cells on a slide. 52x52 µm pixels. Targets: PSPC1 (AlexaFluor 514) and DHFR2 201 (AlexaFluor 546). a. PSPC1 (green); b. DHFR2 201 (yellow); c. Merge. Magnification 100X.

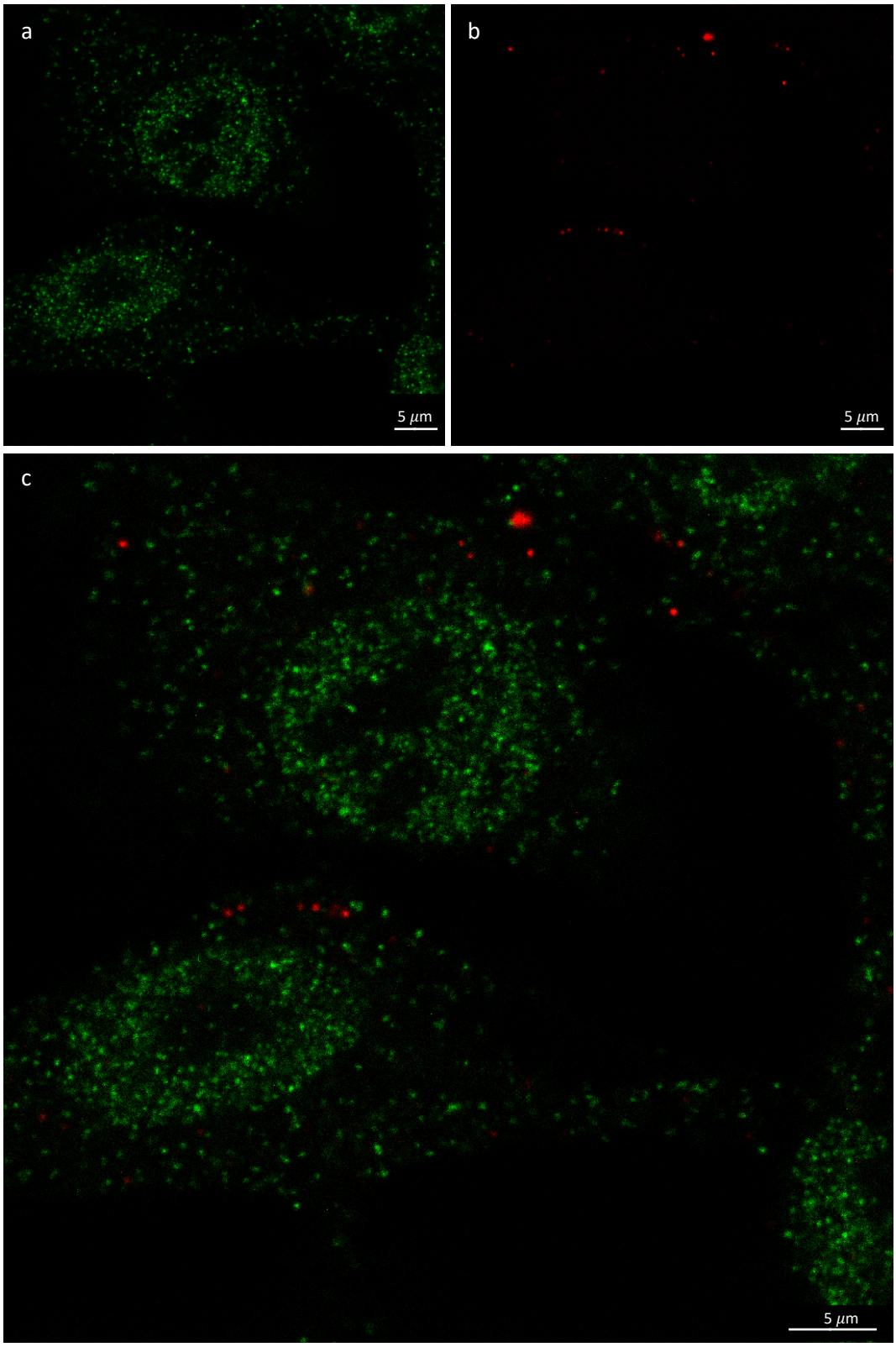


Figure 7.9 Confocal imaging of PSPC1 and DHFR2 202 protein:RNA 2-plex detected via HCR IHC + RNA-FISH in HepG2 cells on a slide. 52x52 µm pixels. Targets: PSPC1 (AlexaFluor 514) and DHFR2 202 (AlexaFluor 594). a. PSPC1 (green); b. DHFR2 202 (red); c. Merge. Magnification 100X.

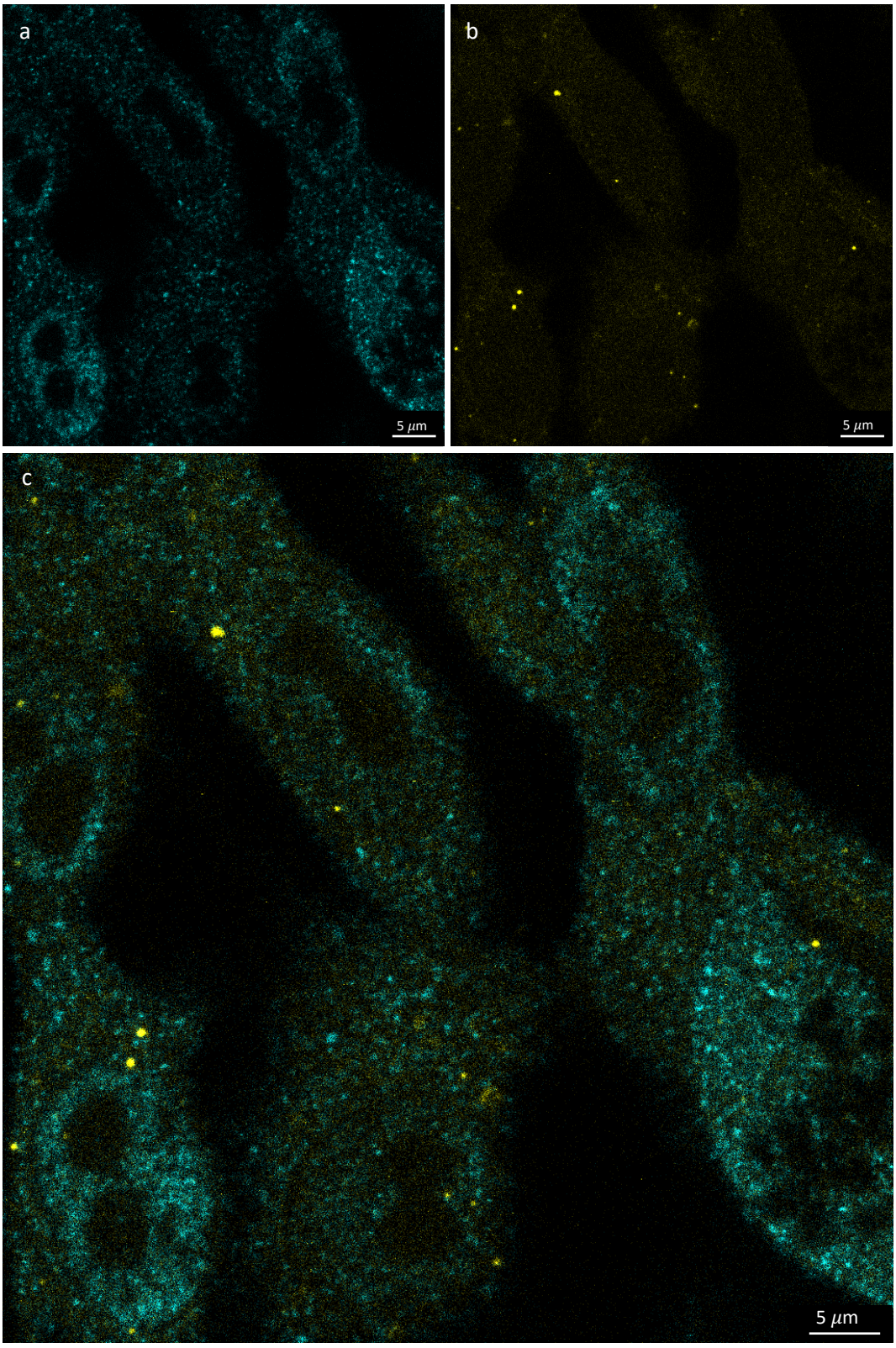


Figure 7.10 Confocal imaging of DHFR and DHFR2 201 protein:RNA 2-plex detected via HCR IHC + RNA-FISH in HepG2 cells on a slide. 52x52 µm pixels. Targets: DHFR (AlexaFluor 488) and DHFR2 201 (AlexaFluor 546). a. DHFR (cyan); b. DHFR2 201 (yellow); c. Merge. Magnification 100X.

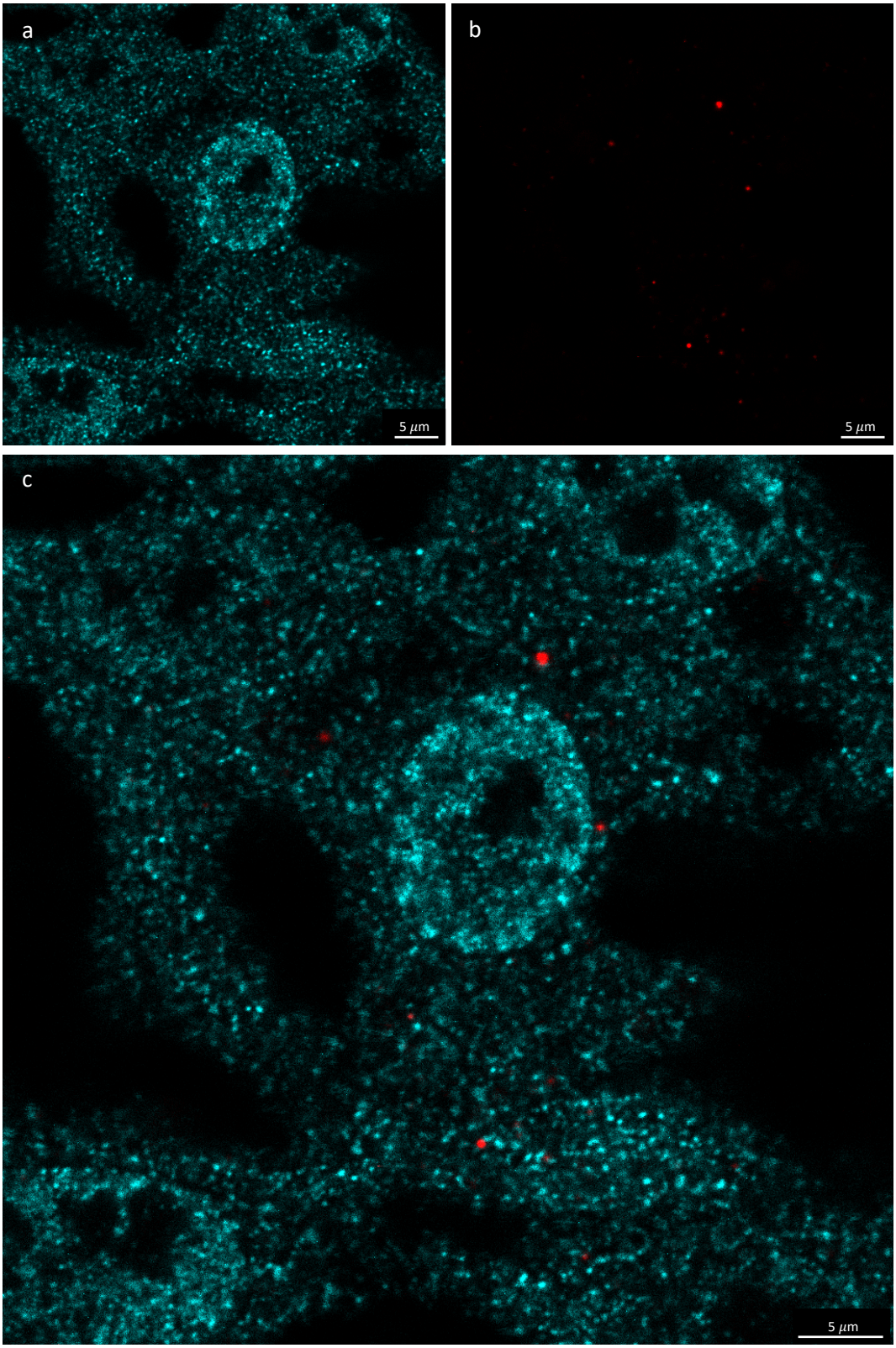


Figure 7.11 Confocal imaging of DHFR and DHFR2 202 protein:RNA 2-plex detected via HCR IHC + RNA-FISH in HepG2 cells on a slide. 52x52 µm pixels. Targets: DHFR (AlexaFluor 488) and DHFR2 202 (AlexaFluor 594). a. DHFR (cyan); b. DHFR2 202 (red); c. Merge. Magnification 100X.

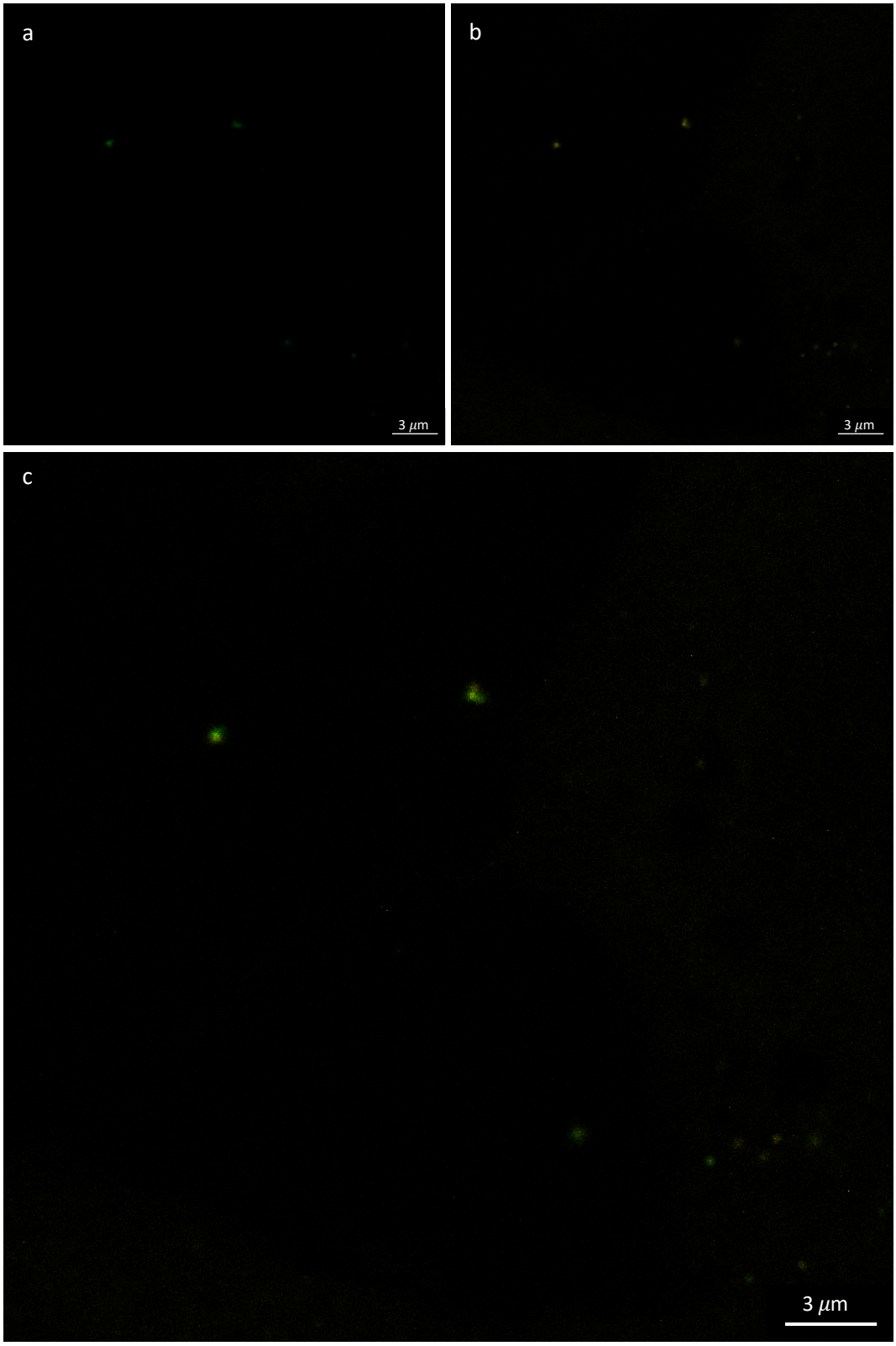


Figure 7.12 Confocal imaging of DHFR ^{RNA} and DHFR2 201 RNA:RNA 2-plex detected via HCR RNA-FISH in HepG2 cells on a slide. 33x33 µm pixels. Targets: DHFR ^{RNA} (AlexaFluor 514) and DHFR2 201 (AlexaFluor 546). a. DHFR ^{RNA} (green); b. DHFR2 201 (yellow); c. Merge. Magnification 100X.

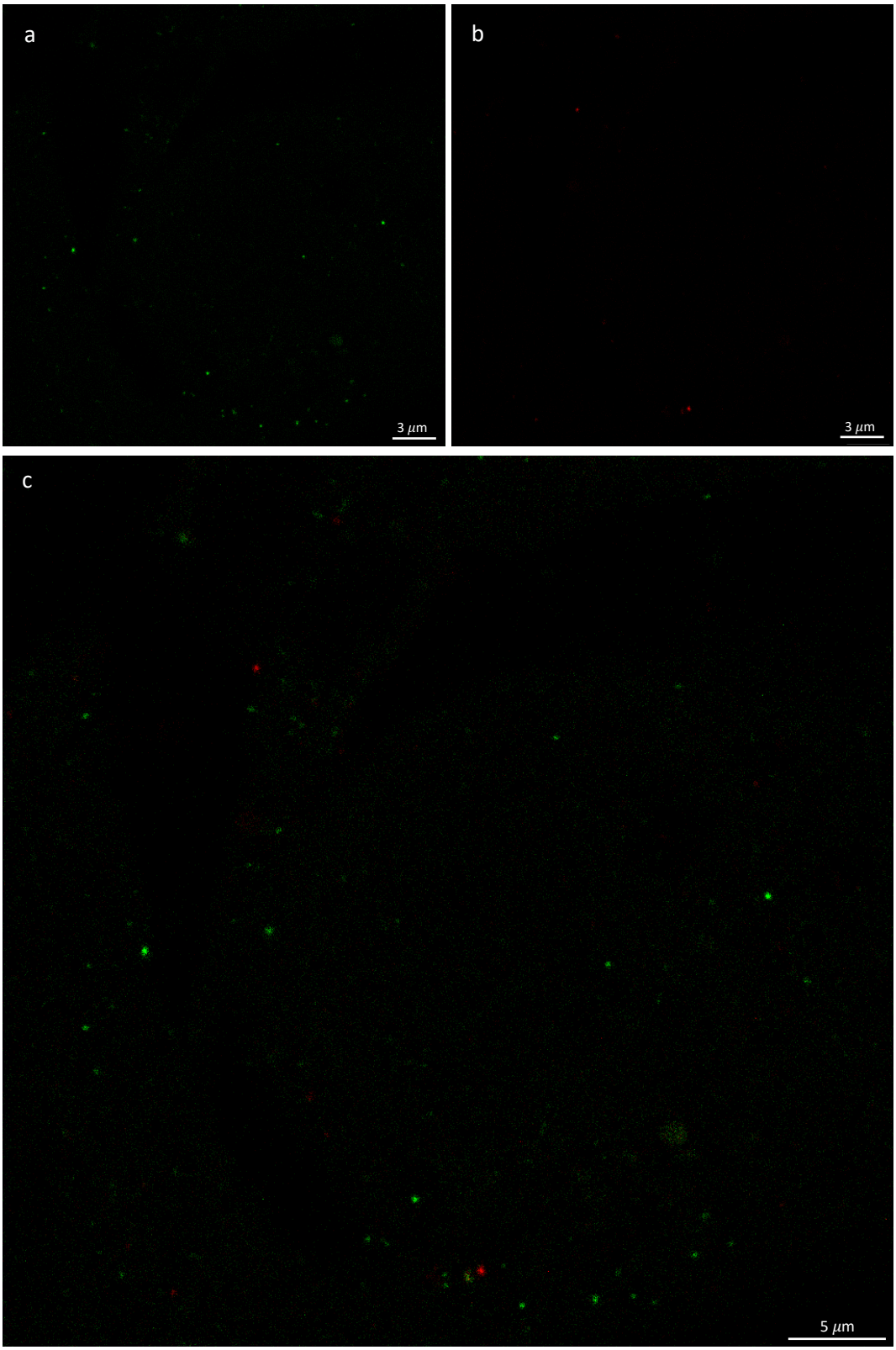


Figure 7.13 Confocal imaging of DHFR ^{RNA} and DHFR2 202 RNA:RNA 2-plex detected via HCR RNA-FISH in HepG2 cells on a slide. 52x52 µm pixels. Targets: DHFR ^{RNA} (AlexaFluor 514) and DHFR2 202 (AlexaFluor 594). a. DHFR ^{RNA} (green); b. DHFR2 202 (red); c. Merge. Magnification 100X.

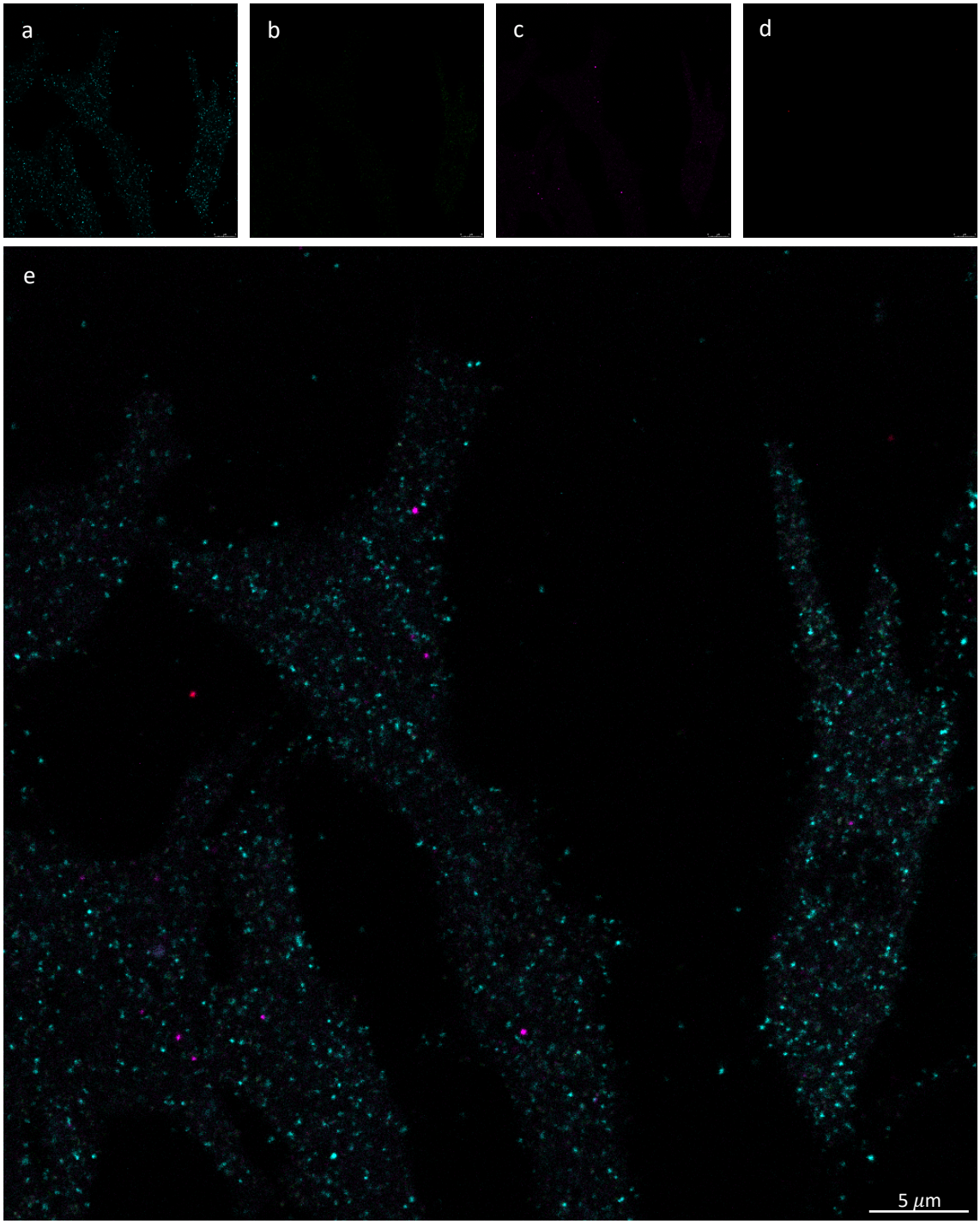


Figure 7.14 Confocal imaging of RPS3, PSPC1, DHFR2 201 and DHFR2 202 protein:RNA 4-plex detected via HCR IHC + RNA-FISH in HepG2 cells on a slide. 92x92 µm pixels. Targets: RPS3 (AlexaFluor 488), PSPC1 (AlexaFluor 514), DHFR2 201 (AlexaFluor 546), DHFR2 202 (AlexaFluor 594). a. RSP3 (cyan); b. PSPC1 (green); c. DHFR2 201 (magenta); d. DHFR2 202 (red); E. Merge. Magnification 100X.

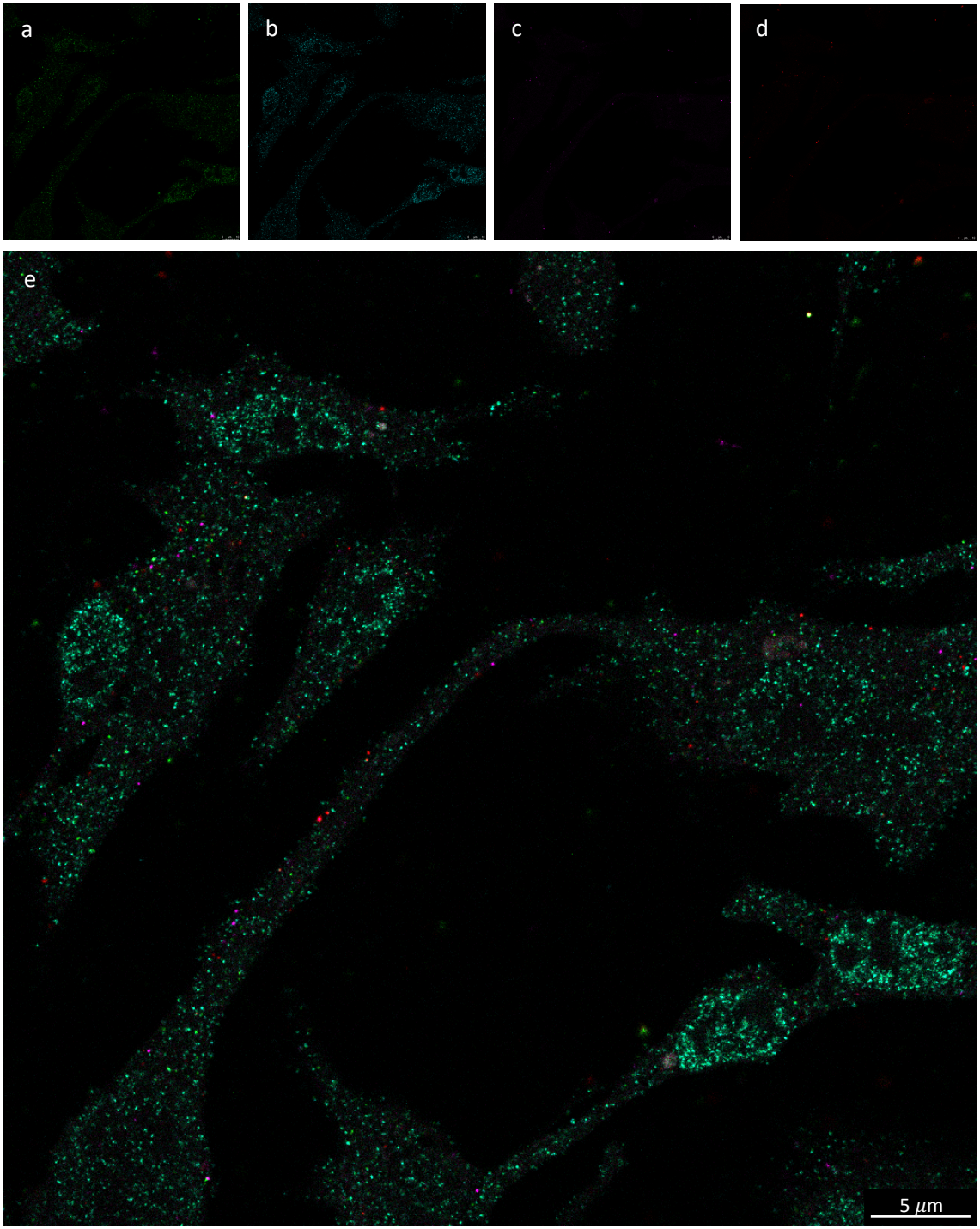


Figure 7.15 Confocal imaging of DHFR, DHFR ^{RNA}, DHFR2 201 and DHFR2 202 protein:RNA 4-plex detected via HCR IHC + RNA-FISH in HepG2 cells on a slide. 155x155 µm pixels. Targets: DHFR (AlexaFluor 488), DHFR ^{RNA} (AlexaFluor 514), DHFR2 201 (AlexaFluor 546), DHFR2 202 (AlexaFluor 594). a. DHFR ^{RNA} (green); b. DHFR protein (cyan); c. DHFR2 201 (magenta); d. DHFR2 202 (red); e. Merge. Magnification 100X.

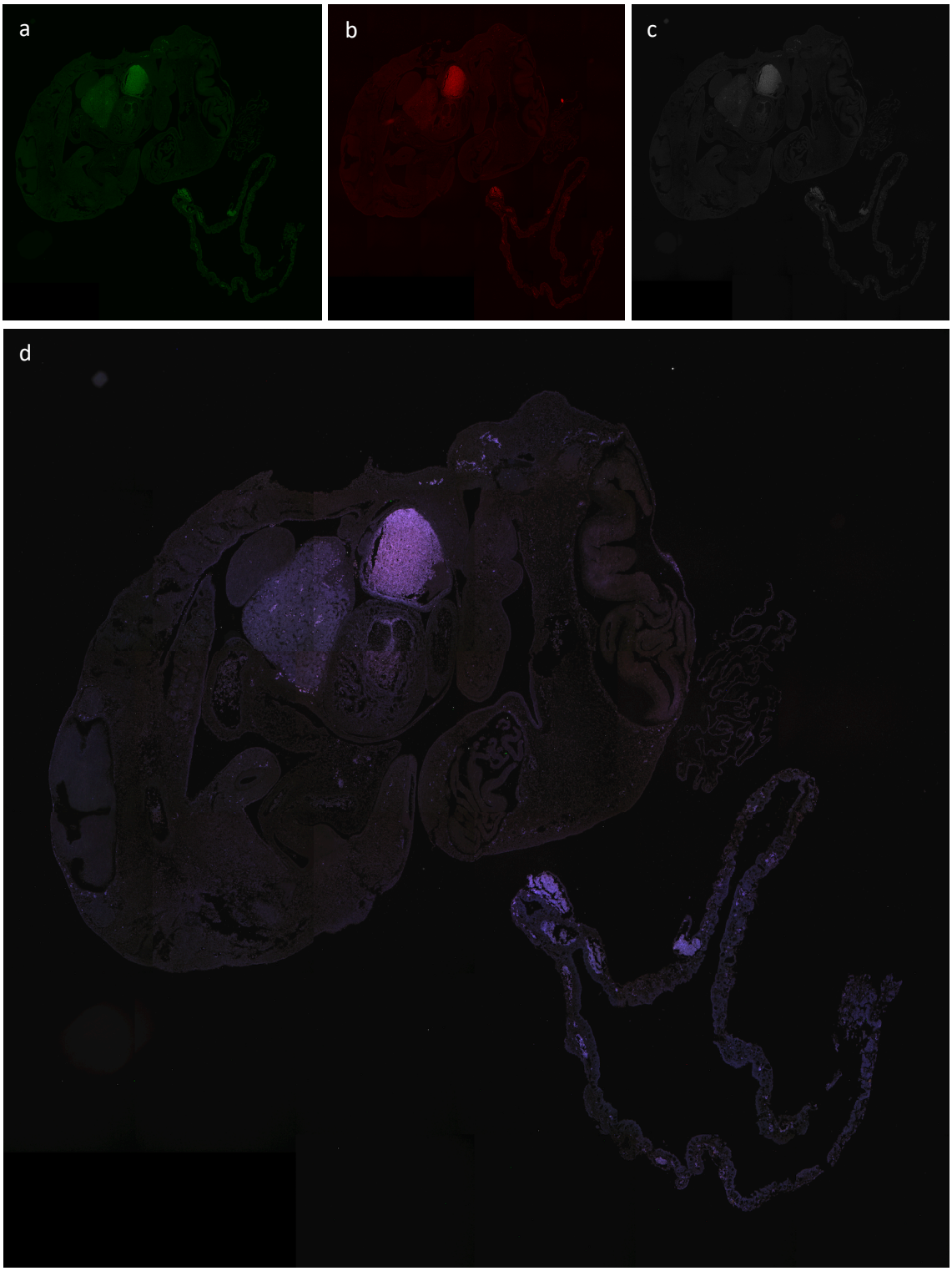


Figure 7.16 Confocal imaging of DHFR, DHFR2 201 and DHFR2 202 protein:RNA 3-plex detected via HCR IHC + RNA-FISH in embryo sections on a slide. Magnification: 10X. Targets: DHFR (AlexaFluor 488), DHFR2 201 (AlexaFluor 546), DHFR2 202 (AlexaFluor 594). a. DHFR protein (green); b. DHFR2 201 (red); c. DHFR2 202 (white); d. Merge.

7.4 Discussion

The present chapter of the thesis was designed to determine the localisation of DHFR2 main isoforms, 201 and 202 and their possible interaction with DHFR, the ribosomes and paraspeckles. After a careful evaluation of the more traditional methods to investigate RNA-protein interactions (EMSA, RNA pull-down, RPA, FISH + IHC), the novel Hybridisation Chain Reaction was deemed to be the best choice to assess many targets simultaneously and have a more general indication of the cellular fate of DHFR2, both in adult and embryonic cells. HCR is a revolutionary way of performing joint IHC and RNA-FISH, with a unified approach for multiplexing. This is facilitated by DNA hairpins and initiator molecules and allows the method to reach automatic background suppression with highly-specific signals. Due to the fluorescence amplification (iso-thermal reaction), the signal increases linearly with the amount of target, leading to a quantitative and high-resolution method (Choi et al., 2018; Schwarzkopf et al., 2021).

The differential association of DHFR2 201 and 202 with the ribosomes (Section 7.1.4) led to a deepened investigation of the differences between the two isoforms, with the consequent discovery of possible regulatory patterns associated with the 5' and 3' UTRs. The 5'UTRs of the two transcripts are different, with the DHFR2 202 upstream sequence more highly folded than that of 201 (Section 7.3.1). The divergent secondary structures could explain the differential behaviour in relation to the ribosomal association; in fact, highly structured RNAs tend to have a shorter half-life and could compromise translation by ribosome stalling (L. Sun et al., 2019). Although the 201 preferential interaction was already proven using a sucrose cushion method, a confirmation experiment was attempted with HCR. This was also used as quality control for the rest of the experiments.

On the other hand, both isoforms possess an identical 3'UTR with IR*Alu*s, making them prone to A-to-I modifications (experimentally demonstrated by Niamh Bookey) and subsequent paraspeckle retention. This hypothesis was the most intriguing, providing a logical explanation of why the DHFR2 protein could not be identified despite the significant RNA levels. Furthermore, the lack of paraspeckles in embryonic cells made the hypothesis even more appealing in the possibility of detecting the DHFR2 protein

uniquely in embryos. For this reason, the entire set of HCR experiments was repeated on human embryos. To note that due to the lack of specific legislation on the matter of stem cell research in Ireland, this part of the study relating to human embryonic material was entirely carried out in the United Kingdom, at the Institute of Child Health, University College London.

The knockout DHFR2 cell line analysis indicated a robust regulatory pattern of DHFR2 on DHFR (Chapter 5). Due to the missed identification of the DHFR2 protein in the DHFR knock-down cell line (and additional other adult and stem cell lines) (Chapter 6), the DHFR2 regulatory function was believed to be due on its RNA. The response elicited by DHFR2 RNA upon complementation (DHFR2 ORF plasmid) on DHFR indicated that the regulation occurred through the DHFR2 Open Reading Frame (possibly containing interaction sites). With the DHFR2 RNA + DHFR enzyme association demonstrated in previous studies (McEntee et al., 2011), our goal was to determine if the DHFR2 mRNA isoforms 201 and 202 could differentially bind DHFR protein. To make the investigation complete, DHFR RNA was also tested.

Although the HCR method proved to be very efficient, the results had to be carefully interpreted. A strikingly evident colocalisation could not be subjectively identified by observing the superimposed images of two probes. The nature of the tested proteins can partially explain this. Contrarily to structural proteins that localise orderly, enzymes like DHFR and ribosomal proteins can be found all over the cells. Also, PSPC1 was considered the golden standard for paraspeckle detection (Fox & Lamond, 2010), but it can also be found in other nuclear areas and, occasionally, in the cytoplasm (Lang & Jou, 2021; Wang et al., 2017). Therefore, HCR's high specificity and amplified resolution of the signals probably did not help identify the most significant clusters exclusively.

A solution came from the use of imaging software able to reduce human error and analyse colocalising signals with high confidence. Colocalisation can be intended as the ensemble of co-occurrence, the spatial juxtaposition of two signals, and correlation, which regards the overlapping of the signals plus their proportional distribution within cellular structures (Dunn et al., 2011). The combined examination of the Pearson's (r) and Mander's (M1, M2) coefficients underlined a clear association between the DHFR RNA and DHFR2 201 RNA (r=0.989; M1=0.998, M2=0.992). This unexpected but thrilling

result provides new information on the DHFR-DHFR2 interaction and regulatory mechanism. Even though the possibility of a DHFR2 201-DHFR interaction remains open (Pearson's coefficient of 0.617 indicates a certain degree of correlation), the DHFR regulation via RNA-RNA interaction enhances the support for the DHFR2 lncRNA hypothesis (Sebastian-delaCruz et al., 2021).

Additionally, a moderate correlation was found between DHFR2 201 RNA and PSPC1 protein (r=0.683; M1=0.998, M2=0.8), therefore maintaining the possibility of paraspeckle retention for this isoform. In general, cancerous cell lines could be considered a limiting model due to their unique genetic patterns and relative functioning, which may actually differ from that of "normal" cells (Carter & Shieh, 2010).

While DHFR2 201 RNA seems to be correlated to a few targets at different degrees, it is surprising that DHFR2 202 presents correlation coefficients close to 0 for most of the co-tested targets. This is an unforeseeable outcome considering the high sequence similarity between 201 and 202. This result indicated that the 5'UTR, possibly via secondary structure formation, is key to a distinct functionality for the two isoforms. It may also suggest a novel function for DHFR2 202, which we could not point out, or it may indicate no function. Undoubtedly, the possibility of technical issues linked to the experimental procedure remains plausible.

The HCR investigation of the DHFR2 RNAs and DHFR enzyme in embryos sees them all colocalising predominantly in heart tissue and the surrounding lungs and liver. The low magnification power did not allow us to investigate the probe localisation at the intracellular level, thus restricting the comparison with the HepG2 samples. Nevertheless, discovering a tissue-specific colocalisation of DHFR and DHFR2 RNAs in human embryos was very promising. This result gained even more importance in light of the unprecedented identification of the DHFR2 protein in embryonic heart tissue. Considering that the proteomics targeted analysis of adult tissues did not include cardiac tissues, the tissue specificity of DHFR2 protein remains to be explored. The paraspeckle retention hypothesis could still stand in place in the case of DHFR2 being a cardiac protein in embryogenesis only or throughout life.

Undoubtedly, further analysis must be carried out to elucidate the apparently multifaceted nature of DHFR2, which has proven to act as a regulatory RNA in adult (cancerous) cells and as a protein in developing embryos. Several ongoing experiments will help clarify the regulatory mechanisms involved in the DHFR2 functions in the form of RNA and protein. For example, RNA immunoprecipitation is a powerful technique to explore the physical association between a specific protein and RNAs (Gagliardi & Matarazzo, 2016). Applied to paraspeckle-specific proteins, it will allow us to understand if nuclear retention is a mechanism by which DHFR2 RNA is modulated. Furthermore, testing different cell types, including non-cancerous lines, would guarantee a thorough evaluation of the regulatory mechanism, which is variable per se, as the retention is directly proportional to the amount of A-to-I edited bases (Chen et al., 2008).

One limitation of the HCR study was the impossibility of using a nuclear stain, such as DAPI, as it would have interfered with the acquisition of the other signals (Leica TCS SP8 specific issue). This means that we could not establish whether the DHFR-DHFR2 RNA interaction took place in the cytoplasm or in the nucleus. Furthermore, in the future, the DHFR-DHFR2 RNA:RNA interaction could be explored via specific low-throughput methods such as SPR (surface plasmon resonance), co-sedimentation and single-molecule FRET. Note that an investigation of the RISE database (RNA Interactome from Sequencing Experiments) (Gong et al., 2018; Ju et al., 2018) did not show evidence of a DHFR-DHFR2 interaction.

An additional note concerns the specificity of antibodies. Even though both primary and secondary antibodies have been validated by the supplier companies, further proof of specificity could have been provided. Western Blots using purified target protein alongside a cell lysate (from the same cell line to be tested) could offer an indication of specificity for primary antibodies (Pillai-Kastoori et al., 2020). Furthermore, repeating the HCR experiment while omitting the primary antibody would provide a control test for unspecific binding of the secondary antibodies (Hewitt et al., 2014). These two methods are deemed to be the easiest and most rapid ways to confirm the validity of our antibodies, although other methods could be deployed to further test antibody efficiency and specificity (Brooks & Lindsey, 2018; Daneshtalab et al., 2010; Torlakovic et al., 2015).

An additional strategy to assess probe specificity (either antibody or RNA) is to use siRNA to silence the expression of the target molecule and determine if the tested probe will bind to unspecific targets (Olds & Li, 2016). However, the innovative design of the HCR RNA probes guarantees the highest specificity and suppression of background (Schwarzkopf et al., 2021); in other words, even in the occurrence of unspecific binding, the amplification reaction will not take place, thus enhancing the overall performance. In summary, considering the characteristics of the novel HCR methodology, the companies' antibodies validation and the consistency of the probes hybridisation across different experimental setups, we are confident about the accuracy of the reported results.

To conclude, this study did not respond fully to all the original questions on DHFR2; it even added some more. However, the data has given additional support to some of the hypotheses including the DHFR:DHFR2 interaction, the DHFR2 201/202 differential association with ribosomes, the possible retention of 201 into paraspeckles, and the embryonic heart-specific expression of DHFR2. Preliminary data has been gathered via the HCR study but requires confirmation and validation by more specific methods, i.e. paraspeckle immunoprecipitation, RNA:RNA interaction studies, etc. DHFR2 RNAs, and 201 particularly, seem to be regulated through their own structure and multiple interactions. Each part of the DHFR2 RNA sequence and structure own the key to its function, with essential modulator moieties found in the 5'UTR (secondary structure may explain 201 & 202 different behaviour), 3'UTR (A-to-I editing possibly leading to paraspeckle retention), and ORF (translation into a protein, and modulation of DHFR). Therefore, a more rigorous bioinformatic screening of the DHFR2 sequence searching for RNPs interaction signals and post-transcriptional modification sites coupled with a search for lncRNA mechanisms of action may help advance the understanding of this simple yet very complex gene.

Chapter 8
General discussion and conclusions

8.1 General discussion

The present study was designed to determine the function of the human *DHFR2* gene. Two previous studies had identified DHFR2 as a mitochondrial reductase whose function was analogous to the cytoplasmic/nuclear isozyme DHFR. Although the annotated DHFR2 function and organelle localisation fit the OCM design perfectly, the published results presented limitations. McEntee investigated only the recombinant form of DHFR2 (McEntee et al., 2011) and Anderson used standard immunostaining techniques which could not clearly distinguish between DHFR and DHFR2 (Anderson et al., 2011). The additional discovery of DHFR localising in the mitochondria of HEK293 cells (Parle-McDermott lab, unpublished) urged a further investigation to elucidate the nature of the native DHFR2 protein.

To circumvent the technical limitations due to the outstanding similarity between DHFR and DHFR2, the creation of a DHFR knockout line was thought to provide an ideal model for the detection and study of endogenous DHFR2. Such a cell line would have offered the opportunity to test the capacity of DHFR2 to compensate for the lack of the main enzyme DHFR as demonstrated previously in relation to another OCM enzyme, SHMT2a (Anderson & Stover, 2009). Unfortunately, cells with biallelic DHFR knockout were not viable; thus, we had to rely on a DHFR knock-down model instead. The selected mutated DHFR cell line had indels between the end of exon 1 and the beginning of intron 1, with the latter involved in protein stability (Noé et al., 2003). The line presented a residual presence of the DHFR enzyme and a 90 % decrease in reductase activity, as demonstrated by Western Blot, Proteomics analysis and DHFR enzymatic assay (Chapter 4).

A DHFR2 knockout line was also established to investigate DHFR2 properties within the cellular environment (Chapters 3 and 5). This loss-of-function model enabled the evaluation of DHFR2 importance for cell proliferation, its involvement in the OCM and interdependence with DHFR, other than assessing DHFR2 possible roles as a trans-acting factor. Therefore, creating two knockout cell lines aimed to study DHFR2 structure, localisation and compensatory abilities via the DHFR knock-down model and its overall function via the DHFR2 knockout line.

A mass spectrometry-based global and targeted analysis was carried out to assess whether the drastic DHFR decrease in the DHFR knock-down line led to a compensatory increase in the DHFR2 levels. No DHFR2 specific peptides were identified and this precluded any further investigation of the endogenous DHFR2 protein itself. However, the DHFR knock-down line was useful to assess the effects of reduced DHFR activity on OCM and *DHFR2* expression (discussed in Chapter 6 and summarized in the following sections).

On the OCM side, a few genes were majorly dysregulated. First of all, ALDH1L1, which eliminates the activated carbon of formyl-THF, was down-regulated in an attempt to maintain the folate pools. In fact, the DHFR decreased activity is responsible for lowering the overall pool of THF, the main carrier of activated 1C groups. SHMT2 was also down-regulated, possibly favouring cytoplasmic OCM over the mitochondrial counterpart. However, to not fully compromise the mitochondrial pathway, SARDH increase would explain the provision of an alternative source of carbon groups derived from the oxidation of sarcosine to glycine (Chapter 6, Sections 6.3.2 and 6.4).

Major metabolic rearrangements were identified via transcriptomics profiling. The partial loss of the DHFR activity seems to have pushed the cells towards a metabolic profile more typical of hepatic cells. Therefore, an increase in glycogenesis, β-oxidation and cytochrome p450 metabolic rates were observed, in addition to the restoration of secretory processes. Moreover, the Warburg effect (increased glucose uptake and fermentation typical of cancer cells) (Liberti & Locasale, 2016) appeared to be reduced, in line with the reversing of the cancer metabolic profile (Chapter 6, Sections 6.3.3 and 6.4).

A pattern of metabolic adaptation was also hypothesised by microscopic observation and confirmed by folate metabolite profiling. Cells tested immediately after establishing the knock-down cell line were particularly deficient in folate intermediates and looked unhealthy. Over time, cells appeared to recover, as demonstrated by the overall folate increase. The cells showed an extraordinary power of adaptation, leaning on diversified metabolic pathways and exploiting any source available, such as the NEAA/HT supplements. Therefore, the elimination of the additional nutrients from the media led

to a drastic setback in the availability of folate intermediates, as demonstrated by the Folate metabolite Profiling (Chapter 6, Sections 6.3.4 and 6.4).

Despite the DHFR2 protein was not detected in several adult tissue types, including our DHFR knock-down model, the loss-of-function study (Chapter 5) has revealed surprising implications for DHFR2. In the first place, the lack of DHFR2 induced a drastic down-regulation of DHFR, TYMS and SHMT1, major components of the thymidylate *de novo* synthesis in the cytoplasm and nucleus (Chapter 5, Section 5.3.2). However, the complementation test exposed a restoration of normal expression levels solely for DHFR (Chapter 5, Section 5.3.5); thus, hypothesising that the regulatory response observed on the dTMP-SC was directed to DHFR and secondarily to TYMS and SHMT1 (via DHFR) (Chapter 5, Section 5.4). The overexpression study further confirmed the interdependence between DHFR and DHFR2, as an up-regulation of DHFR2 RNA caused a proportional rise in DHFR RNA (and protein) (Chapter 5, Sections 5.3.6 and 5.4).

The effects of DHFR2 loss on thymidylate synthesis were also observed in the transcriptomic profiling, where DNA replication and repair pathways, plus pyrimidine metabolism were primarily down-regulated (Chapter 5, Section 5.3.3). This trend is easily explained by the overall folate pools, which are reduced by half (Chapter 5, Section 5.3.4). The reduction of the DHFR levels can explain this drastic effect. In relation to the proportions of folates, the reduced activity of SHMT1 causes the decrease of methylene-THF and the final impairment of dUMP to dTMP conversion, both for reduction of the substrate (methylene-THF) and relative enzyme (TYMS). This behaviour was consistent with the impairment of the dTMP *de novo* pathway (Chapter 5, Section 5.4).

Both the cell models that were generated in this thesis endured a DHFR impairment, direct in the case of the DHFR knock-down and indirect for the DHFR2 knockout. However, it was surprising to observe how the broader effects were different despite the common alteration of the OCM. This could be explained by the residual DHFR activity, which hovered around 50 % in the DHFR2 knockout compared to ~10 % in the DHFR knock-down, as shown in Figure 8.1. As expected, the irreversible genetic modification of *DHFR* caused a drastic activity reduction with global consequences for cell metabolism. However, the distinctive decline in DHFR activity was unforeseen for the DHFR2 knockout. These results reflect the remarkable impact of the DHFR2 gene on

the regulation of DHFR. The regulatory response exerted by DHFR2 can be recognised in one of the crucial experiments performed by Anderson. They used a DHFR2-specific siRNA to silence it. The result was the complete abolition of reductase activity from mitochondria (Anderson et al., 2011). Even though the authors explained this outcome by the loss of DHFR2 enzymatic activity in the mitochondria (our lab subsequently showed that it is DHFR in the mitochondria and not DHFR2), we suggest that DHFR2 silencing caused DHFR to withdraw from the mitochondrial compartment, restricting the remaining pool of enzymes in the cytoplasm.

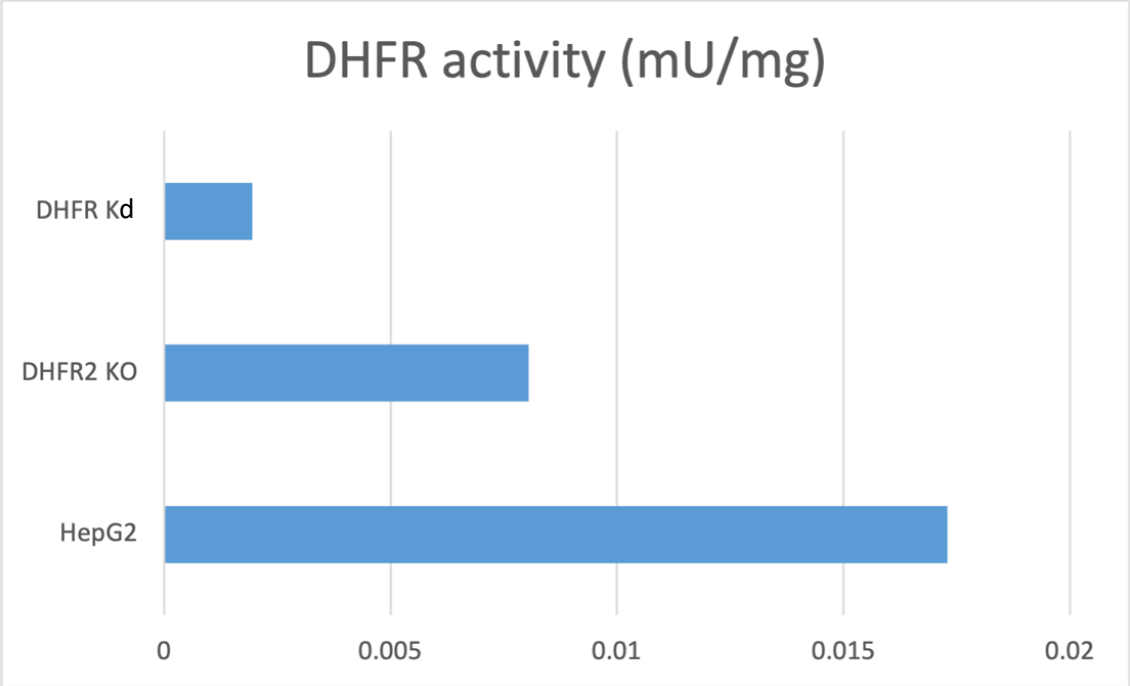


Figure 8.1 DHFR enzymatic assay. The DHFR activity to reduce NADPH was measured in wild-type HepG2, DHFR knock-down and DHFR2 knockout lines. The graph shows a significant reduction in activity of both edited lines, with DHFR2 KO and DHFR Kd retaining 46% and 11 % reductase activity, respectively.

The examination and comparison of the folate metabolite profiling between the two cell models showed a maintained pattern of dysregulated folates, as illustrated in Figure 8.2. A drastic drop in all forms of folate is observed in both lines. In particular, the folate levels of the DHFR2 knockout place midway between early-stage (*in orange*) and post-recovery DHFR knock-down (*in grey*) but remain higher than unsupplemented DHFR knock-down (*in yellow*). The fact that DHFR is the main altered enzyme in both cell models could resonate with the similar oscillations of folate intermediates. However, when considering the unsupplemented DHFR knock-down, all folate metabolites are around half concentration compared to DHFR2 knockout. This evidence suggests that DHFR2 is an excellent modulator of DHFR and overall OCM activity, but its loss is not equivalent to that of DHFR itself.

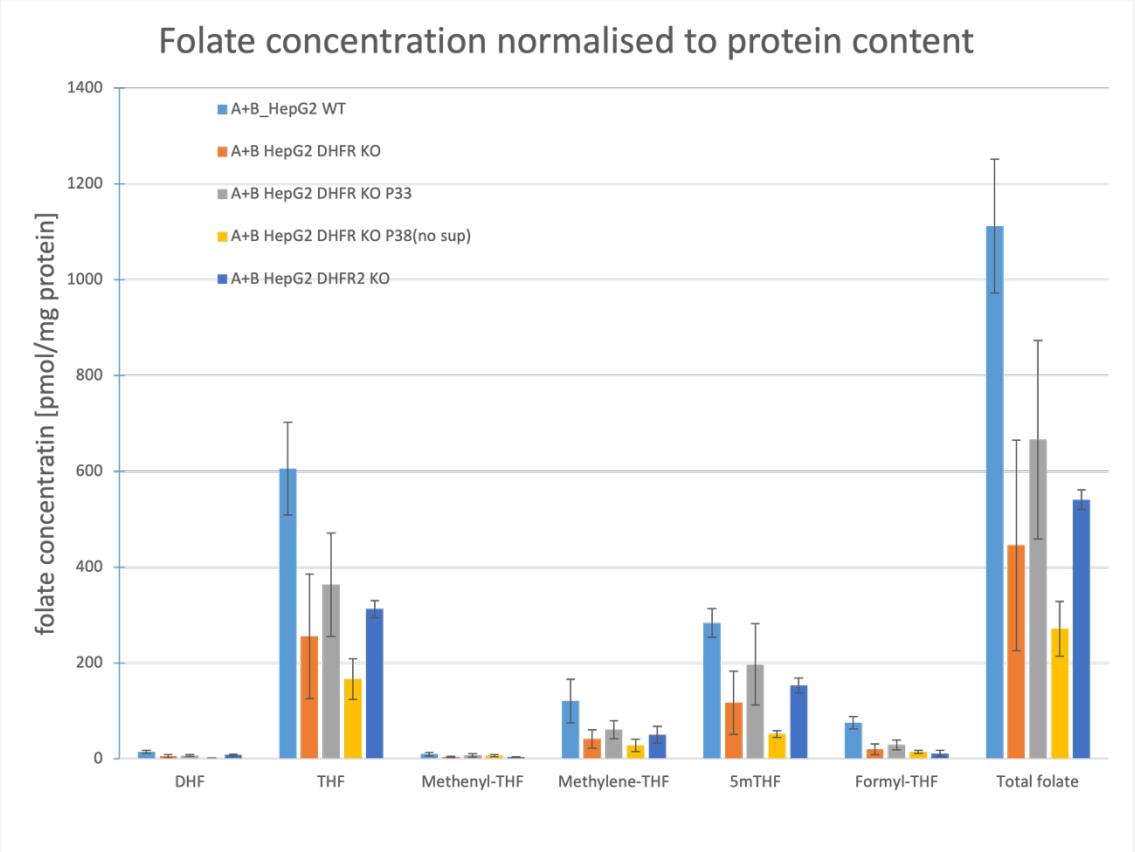


Figure 8.2 Folate metabolites profile in DHFR kd, DHFR2 KO and HepG2 wild-type lines. The DHFR kd line was tested at different points in time (passages 23, 33, 38); HepG2 wild-type, DHFR2 KO and DHFR kd p38 were grown without supplementation, while DHFR kd p23 and p33 were treated with HT/NEAA supplements. The concentrations were normalised to protein content and expressed as pmol/mg protein. The measures were performed in duplicate.

Further investigation of the interdependence between DHFR and DHFR2 involved analysing the RT-qPCR and proteomics data jointly. Figures 8.3 and 8.4 summarise the effects of the DHFR impairment and DHFR2 loss on DHFR2 and DHFR, respectively. The DHFR knock-down model possesses an altered DHFR enzyme with 90 % less reductase activity. While a reduced cellular concentration of the enzyme (proteomics data) was expected, it was surprising to observe a 1.6x increase in the relative RNA expression. This phenomenon was explained via the autoregulatory mechanism, typical of DHFR (Abali et al., 2008). Therefore, we hypothesised that because of DHFR instability and poor functionality (in the knockdown model), a feedback mechanism was signalling to make more protein, leading to increased RNA levels. Note that DHFR2 RNA levels also undergo a moderate increase due to potential involvement in the regulatory feedback (Fig. 8.3). On the other side, the DHFR2 knockout completely eliminates the DHFR2 RNA expression. As a result, DHFR levels (RNA and protein) considerably dropped, indicating a direct dependence on DHFR2 (Fig. 8.4).

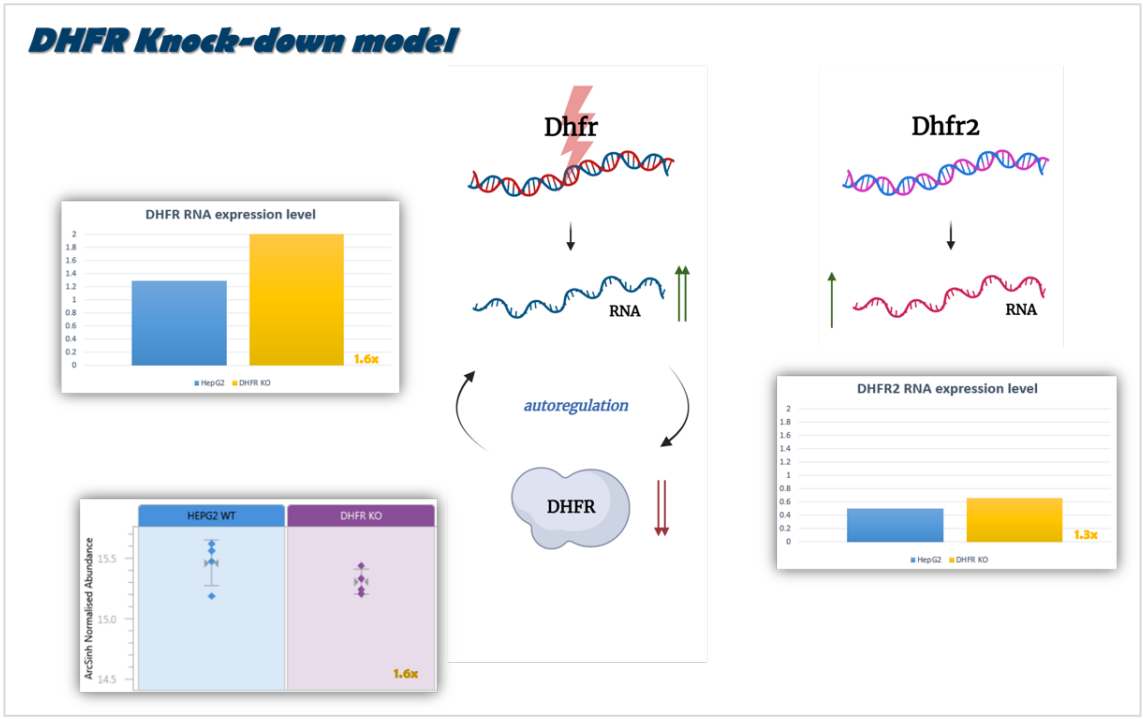


Figure 8.3 DHFR knock-down model: effects of DHFR impairment on DHFR and DHFR2. DHFR and DHFR2 relative expression and protein abundance were measured in comparison with HepG2 wild-type.

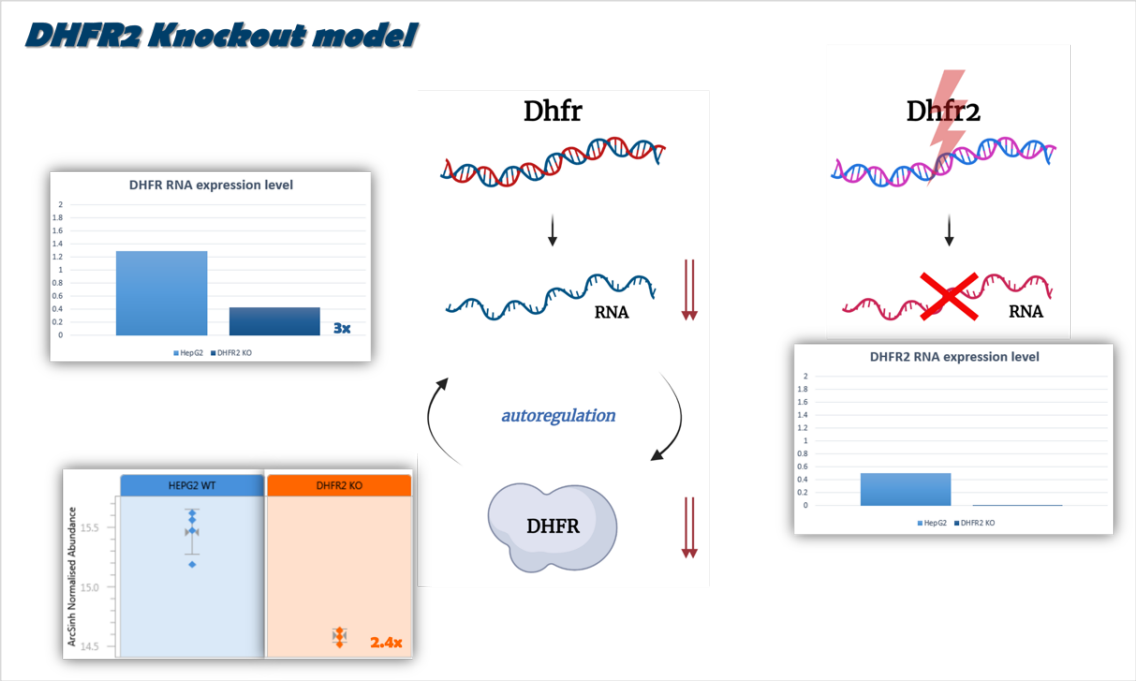


Figure 8.4 DHFR2 knockout model: effects of DHFR2 loss on DHFR. DHFR and DHFR2 relative expression and protein abundance were measured in comparison with HepG2 wild-type.

To comprehensively elucidate the link between DHFR and DHFR2, the complementation test and overexpression experiment provided additional data supporting this regulatory mechanism (Fig. 8.5). If DHFR2 loss created down-regulation of DHFR, the complementation of the DHFR2 KO line with a DHFR2 plasmid restored the DHFR levels (RNA and protein). Equally, the overexpression of DHFR2 via plasmid transfection in a wild-type HepG2 line induced an increase of both DHFR RNA and protein.

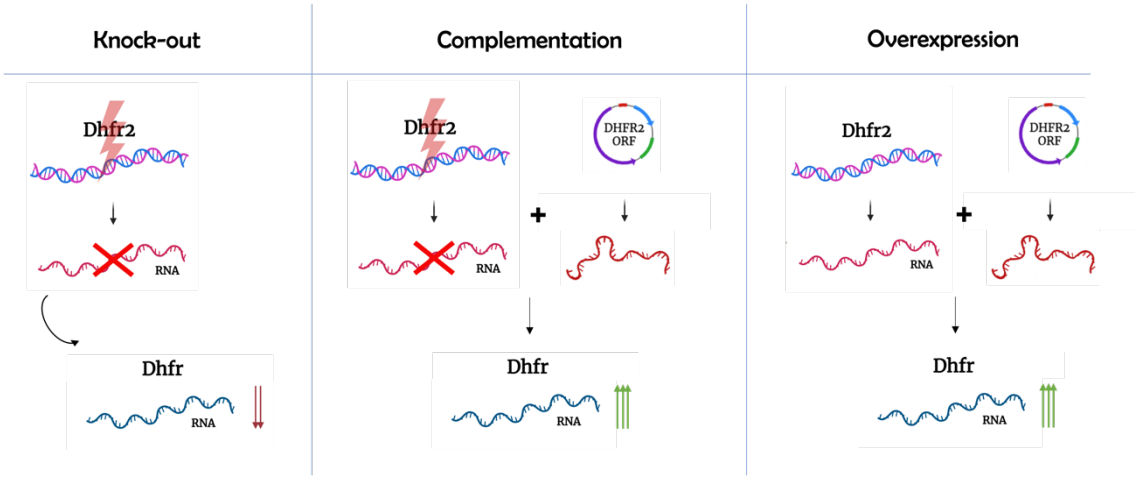


Figure 8.5 DHFR expression levels in response to DHFR2. Comparison among DHFR2 knockout, DHFR2 KO complemented line and overexpression model. In the absence of DHFR2, DHFR levels drop, but they are restored when a DHFR2 plasmid is expressed. Similarly, a HepG2 line (expressing normal levels of DHFR2 RNA) experiences an increase in DHFR levels following the DHFR2 plasmid transfection.

In the light of the collective findings, we envisaged a regulatory model, as shown in Figure 8.6. According to this model, DHFR2 RNA directly regulates DHFR RNA in a proportional way, with similar effects on the enzyme itself. Therefore, low levels of DHFR2 RNA correspond to decreased levels of DHFR; similarly, high concentrations of DHFR2 lead to a rise in DHFR RNA and protein. Moreover, drops in DHFR enzymatic levels signal back to DHFR2, which, by incrementing its transcription, prompts an upturn in DHFR RNA levels, thus coordinating a regulatory response. This operating system reveals a feedback loop mechanism in which the principal regulator, DHFR2, is in its turn regulated by its target, DHFR. The HCR localisation experiments and the results of the EMSA test described by McEntee (2011) indicate that DHFR2 RNA (201) has the capacity to bind the DHFR enzyme. In addition, DHFR2 201 was also found colocalising with DHFR RNA. We hypothesise that DHFR2 could be an additional player in the DHFR auto-regulatory mechanism, or alternatively, interact with DHFR RNA and/or protein independently; additional regulatory factors could also be involved in the regulatory mechanism. However, the available data is not sufficient to determine how DHFR2 regulates DHFR. Despite the delineation of the DHFR2 mechanism of action requiring further investigation, the discovery of DHFR2 RNA regulatory function on DHFR could

open the way toward more natural and balanced therapies to control and modulate DHFR activity in diseases like cancer or rheumatoid arthritis.

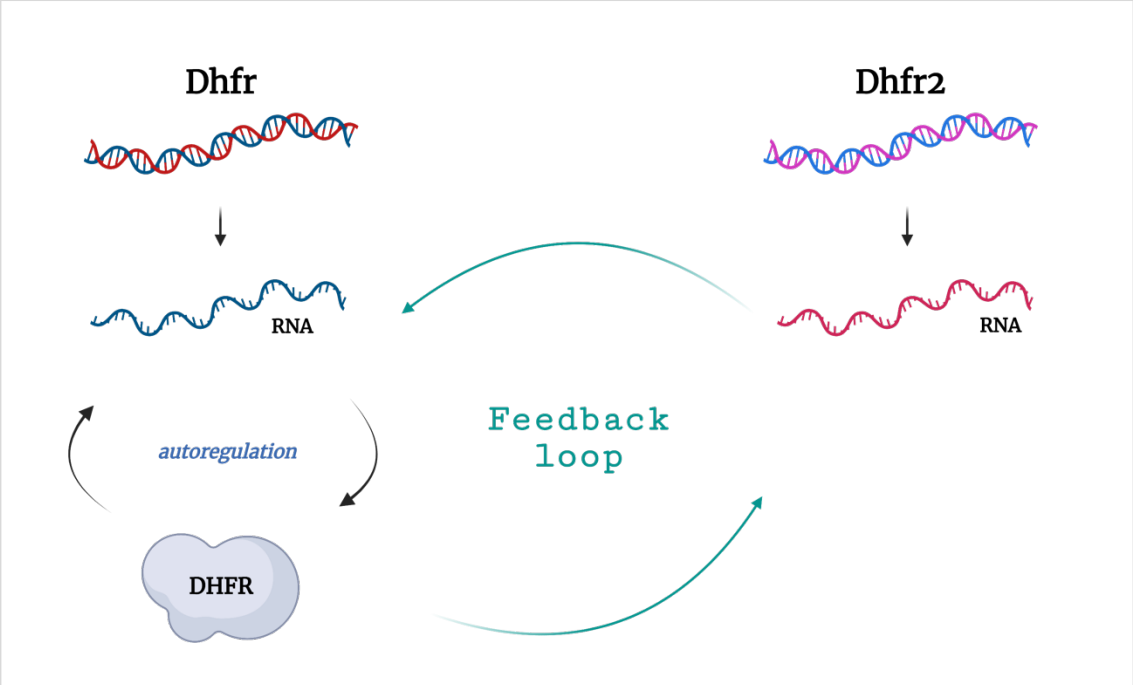


Figure 8.6 DHFR/DHFR2 feedback loop regulation. DHFR2 RNA regulates DHFR RNA (and protein, consecutively). Low levels of DHFR enzyme trigger an increase in DHFR2 which, in turn, leads to a DHFR RNA rise. DHFR enzyme level are intrinsically regulated by the protein ability to bind its own RNA in an auto-regulatory mechanism. DHFR2 RNA might be involved in the auto-regulatory mechanism or act on DHFR RNA/protein independently.

The DHFR2 protein could not be identified using targeted mass spectrometry in the following adult tissues: HepG2, HepG2 DHFR knock-down, HepG2 DHFR2 knockout, undifferentiated iPSC, iPS-derived hepatocytes (101B), HuH7 (hepatocyte-derived carcinoma), IMR32 (neuroblastoma), NEP (neuroectodermal progenitor stem cells). Kidney, heart, lung, liver and brain tissues from human embryos (Carnegie stage 21-22) and foetuses at weeks 9-10 and 15-17 were analysed using the same methodology, resulting in identifying DHFR2 protein only in embryonic and foetal hearts. Similarly, the recombinant DHFR2 protein was detected, at low concentrations, in the complemented line. Despite considering potential technical constraints relative to the instruments' sensitivity limit, which could have led to not detecting the DHFR2 protein in adult tissues, we are confident to suggest that DHFR2 plays a critical regulatory role in DHFR regulation in the form of lncRNA in HepG2 cells.

However, the discovery of the DHFR2 protein in embryonic heart opens the possibility of a dual function, as described already for other genes (Konina et al., 2021; Nam et al., 2016): reductase enzyme during embryonic heart development and DHFR regulatory lncRNA in adult tissues. This possibility highlights the versatility and multifaceted nature of some genes, including DHFR2.

The RNA sequence could be the key to a gene function, especially when that function is strictly associated with the ribonucleic form. The sequence itself is intrinsically related to structure, equally to post-transcriptional modifications. The RNA sequence, with or without certain post-transcriptional modifications, folds into specific structures that mediate certain functional roles. The function is often carried out through interaction with other molecules. Therefore, investigating DHFR2 RNAs localisation and interactions could help unravel their specific functions. The sequence analysis of DHFR2 led to the identification of particular secondary structures at the 5'UTR and A-to-I modifications within the 3'UTR Alu elements (as discussed in Chapter 7).

The two main DHFR2 isoforms, 201 and 202, are identical except for ~100 bp at the 5' end. This minor difference is reflected in a more substantial structural variation, which, in turn, could be the cause for a differential association with ribosomes (Chapter 7, Section 7.3.1). These data, primarily found via sucrose cushion ultracentrifugation of cycloheximide-blocked ribosomes, were confirmed by the HCR IHC + RNA-FISH. In both experimental settings, DHFR2 201 was found to be moderately associated with ribosomes, compared to DHFR2 202, which showed little to no association (Chapter 7, Section 7.3.3). Despite the apparent DHFR2 201 ribosomal binding, no protein was identified. The hypothesis of ubiquitin-based degradation of DHFR2 protein was tested using a proteasome inhibitor with negative results (Bookey et al., unpublished).

The discovery of DHFR2 as one of the 333 human genes that undergo A-to-I modification in their 3'UTR makes it a potential candidate for paraspeckle retention. Again, by testing the localisation and interactions of 201 and 202, we could infer that 201 is in part associated with PSPC1, one of the main components of paraspeckles. On the other hand, DHFR2 202 does not seem to associate with paraspeckle, having a predominantly cytoplasmic localisation (Chapter 7, Section 7.3.3).

Finally, the interaction with DHFR was tested as key to the regulatory role. We hypothesised that DHFR2 RNA (either isoform) would associate with DHFR, as already demonstrated by McEntee (McEntee et al., 2011). To our surprise, DHFR2 201 seems to closely co-localise with DHFR RNA (Chapter 7, Section 7.3.3). At the same time, DHFR RNA was found colocalising with DHFR protein confirming the autoregulatory mechanism by which DHFR traps its own RNA until it binds its DHF substrate (Abali et al., 2008). These two events can find significance in light of DHFR2 being a non-coding RNA. In fact, among the various mechanisms of action, some lncRNA work by simultaneously interacting with RNAs and proteins, for example, by anchoring themselves to the RNA and summoning factors to the near protein (Sebastian-delaCruz et al., 2021).

8.2 Conclusion and Future work

To conclude, we discovered that DHFR2 works as a direct regulator of DHFR (in HepG2 cells), possibly in the form of a long non-coding RNA. We established that DHFR2 201 colocalises with DHFR RNA and to a minor extent with PSPC1 (paraspeckles) and ribosomes. None of the tested targets seems to interact with DHFR2 202, prompting the hypothesis of it being the isoform responsible for producing the DHFR2 protein in embryonic heart and subject to some repressive regulation in the rest of the tissues. Despite these overall promising results, we just started exploring the many possible scenarios that would explain DHFR2 regulation and function. Several questions remain unanswered at present; therefore, the following studies are recommended.

- Mass spectrometry-based proteomics analysis of adult heart tissues.
  Heart tissue was not included in the targeted proteomics investigation. A mass spectrometry analysis must be performed on adult heart tissues to assess if DHFR2 is a tissue-specific protein, i.e., cardiac protein or a protein required exclusively at developmental stages.
- Testing recombinant DHFR2 activity (NADPH vs NADH; DHF vs BH2; MTX).
  While assessing if it is possible to obtain heart protein extracts, the DHFR2 recombinant protein can be tested for functionality. In addition to the standard

enzymatic reaction (which was already performed by McEntee et al., 2011), it is advisable to test the coenzyme NADH, as the three amino acid differences at the binding sites (NADPH and DHF) between DHFR and DHFR2 could be indicative of a different coenzyme usage. On the same note, the main substrate of DHFR2 could be dihydrobiopterin (BH2). Recent studies have found evidence of a secondary function of DHFR, which can convert BH2 to BH4 (tetrahydrobiopterin) (Crabtree et al., 2009). The minor structural differences in the active sites could make BH2 the perfect substrate for DHFR2. Testing the reductase activity of DHFR2 using BH2 as a substrate could help clarify this possibility. Finally, DHFR2 inhibition by methotrexate should be investigated, following the similarity with DHFR.

- Validating DHFR2 201 paraspeckle retention with RIP experiments.

  The HCR experiment provided partial information about the nuclear retention of DHFR2 201. A more specific investigation, such as RNA immunoprecipitation (RIP), could address this question. By using a PSPC1 (or $p54^{nrb}$) antibody anchored to magnetic beads, a cell lysate could be filtered for paraspeckles. The retained enriched RNA pool can be screened for the presence of DHFR2 201. Multiple cell types can be screened with this method.

- Testing *in vitro* association of DHFR, DHFR RNA, and DHFR2 RNA (+ MTX).

  The interaction between DHFR (RNA/protein) and DHFR2 can be proven *in vitro* using an EMSA test (electrophoretic mobility-shift assay). The DHFR/DHFR2 RNA:RNA interaction and the triple association among DHFR protein, RNA and DHFR2 RNA can all be investigated. In the hypothesis of a positive result of the triple association, the action of MTX in disrupting the complex can also be evaluated.

- Bioinformatic investigation of post-transcriptional modification and associated secondary structure.

  As discussed in Chapter 7, RNA sequence and structure are intimately correlated. Investigating possible sites of post-transcriptional modifications could indicate changes in secondary structure and associated interactions. Understanding how the structure mediates the contact with other molecules (DNA, RNA, protein) in the context of lncRNA functions would unveil the mechanism of action and

function of DHFR2, whether involving DHFR alone or, more likely, multiple factors.

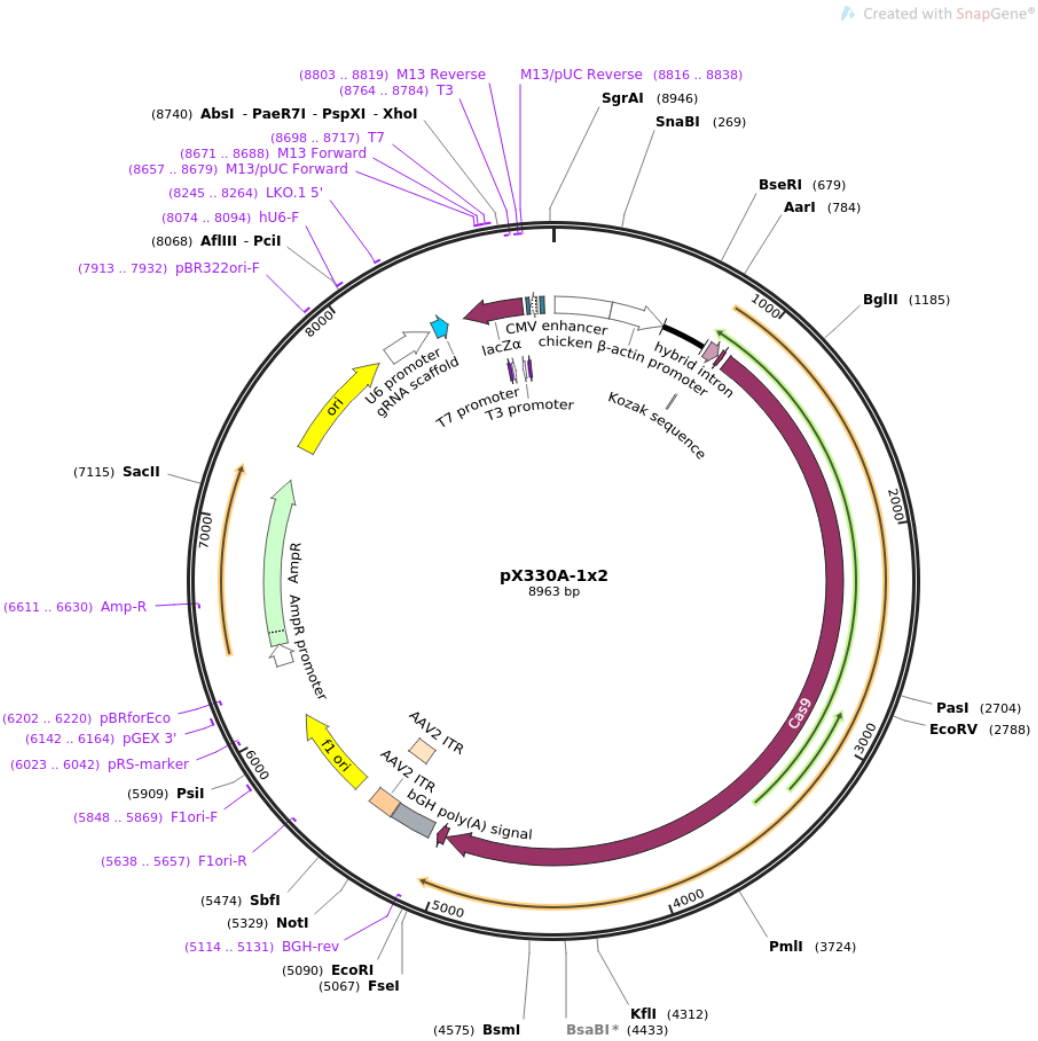
- Bioinformatic investigation of DHFR2-associated miRNAs.
  LncRNAs have been reported to have roles at multiple regulatory levels (X. Zhang et al., 2019). One of them involves micro RNAs, with lncRNAs acting like sponges. A hypothesis is that DHFR2 could act as endogenous target mimics, sequestering miRNAs which otherwise would target DHFR. Alternatively, DHFR2 could be the precursor of miRNA and siRNA itself. Perhaps, DHFR2 201 and 202 present diverse functional mechanisms. A bioinformatic investigation on DHFR2-associated miRNAs could help add supplementary data to unravel the function of DHFR2.

Appendices

A. pX330A-1x2, CRISPR/Cas9 vector backbone.

Plasmid containing the Cas9 gene and the gRNA scaffold. Used to construct the DHFR-gRNA plasmid by annealed oligo cloning (http://n2t.net/addgene:58766).

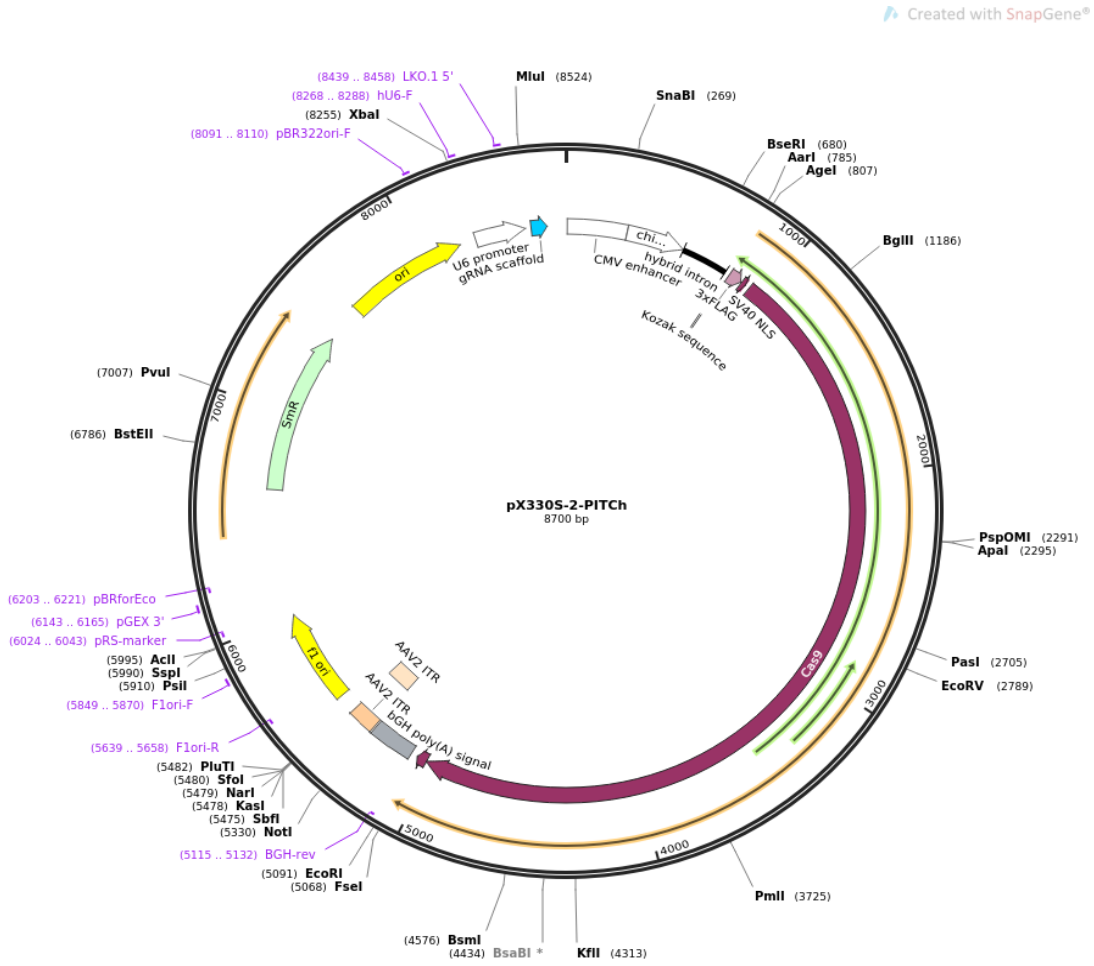
pX330A-1x2	
Vector backbone	pUC ori vector
Vector type	Mammalian Expression, CRISPR
Selectable markers	n/a
Bacterial Resistance	Ampicillin
Growth Temperature	37°C
Growth Strain	Stbl3
Copy number	High Copy
Gene/Insert name	humanized *S.pyogenes* Cas9 nuclease
Alt name	SpCas9
Species	S. pyogenes
Insert Size (bp)	4272
Promoter	CBh



B. pX330S-2-PITCh, CRISPR/Cas9 plasmid targeting the extremities of the EGFP-2A-PuroR cassette.

Plasmid containing the Cas9 gene and a PITCh gRNA to target the edges of the gene cassette. The PITCh gRNA was excised and cloned into the DHFR gRNA plasmid to build the multiplexed All-in-One vector (http://n2t.net/addgene:63670).

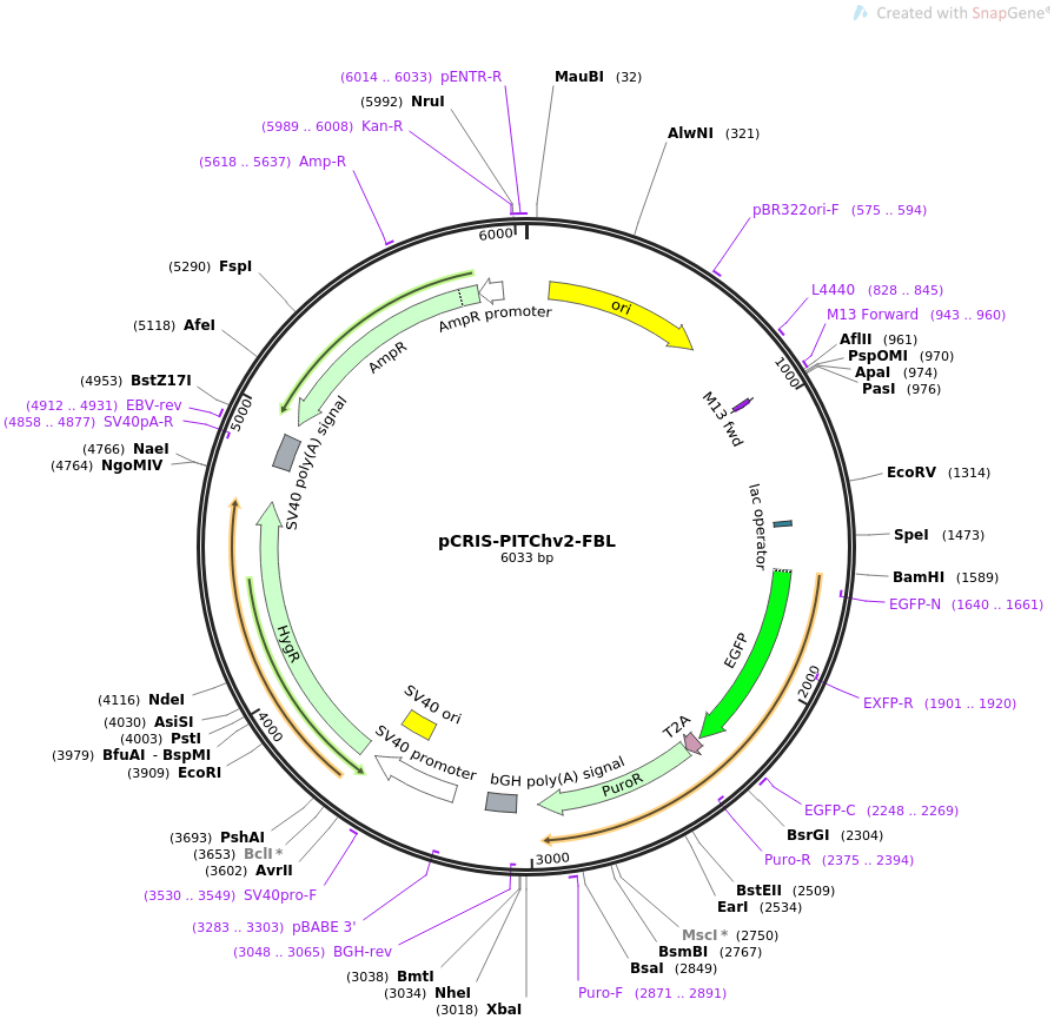
pX330S-2-PITCh	
Vector backbone	pUC ori vector
Vector type	Mammalian Expression, CRISPR
Bacterial Resistance	Spectinomycin/Streptomycin
Growth Temperature	37°C
Growth Strain	Stbl3
Copy number	High Copy
Gene/Insert name	humanized *S.pyogenes* Cas9
Species	S. pyogenes
Insert Size (bp)	4272
Promoter	CBh



C. pCRIS-PITChv2-FBL, PITCh donor vector for EGFP-2A-PuroR insertion into human FBL locus.

Plasmid used directly as a positive control in the transfection experiment for DHFR knock out. Also, used as a template to construct the μΩ vector by replacing FBL microhomologies with DHFR-specific microhomologies (http://n2t.net/addgene:63672).

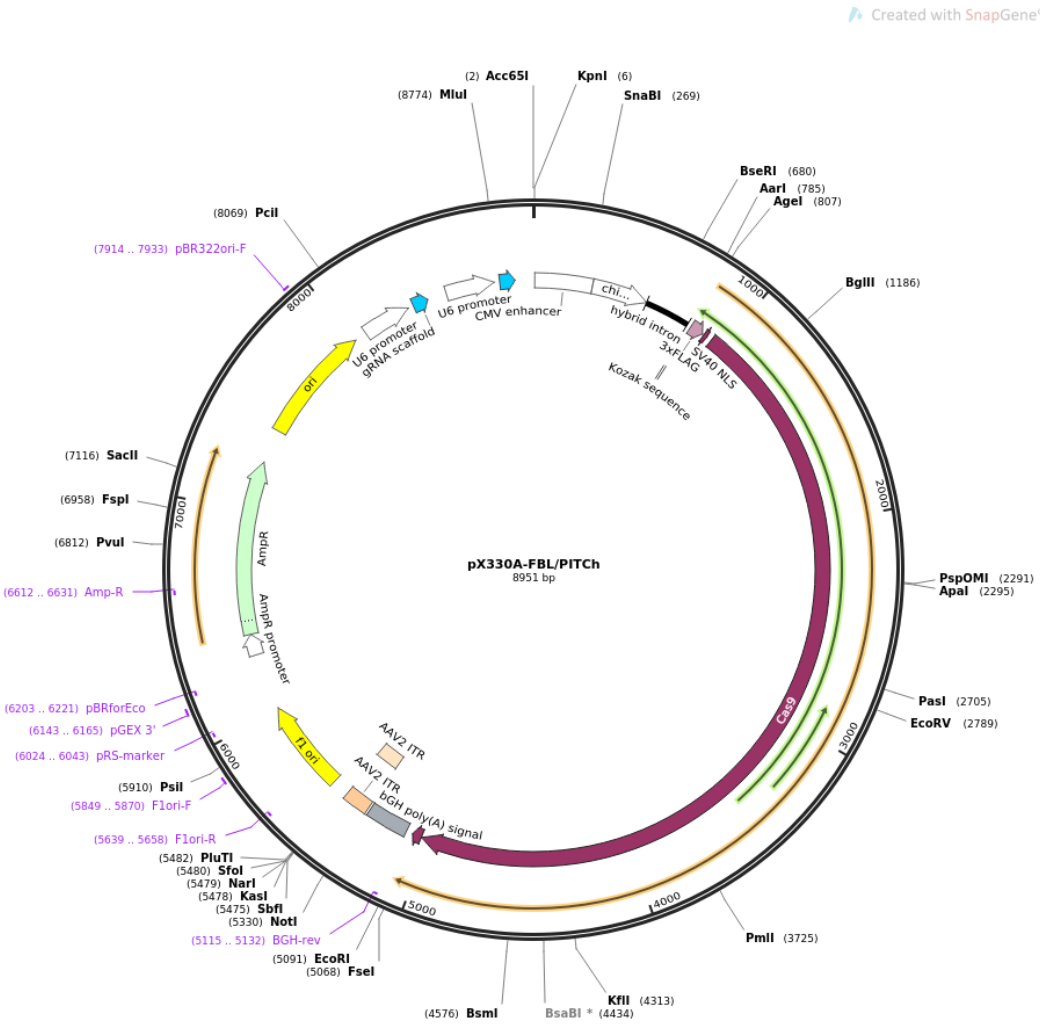
pCRIS-PITChv2-FBL	
Vector backbone	pUC ori vector
Vector type	CRISPR
Selectable markers	Hygromycin
Bacterial Resistance	Ampicillin
Growth Temperature	37°C
Growth Strain	Stbl3
Copy number	High Copy
Gene/Insert name	EGFP-2A-PuroR
Species	Synthetic
Promoter	Promoterless



D. pX330A-FBL/PITCh, CRISPR/Cas9 plasmid targeting FBL.

Plasmid used uniquely for positive control in the DHFR knockout experiment, along with pCRIS-PITChv2-FBL. Plasmid containing the Cas9 gene, an FBL-specific gRNA and a PITCh gRNA to target the edges of the gene cassette (http://n2t.net/addgene:63671).

pX330A-FBL/PITCh	
Vector backbone	pUC ori vector
Vector type	Mammalian Expression, CRISPR
Bacterial Resistance	Ampicillin
Growth Temperature	37°C
Growth Strain	Stbl3
Copy number	High Copy
Gene/Insert name	humanized *S.pyogenes* Cas9
Species	S. pyogenes
Insert Size (bp)	4272
Promoter	CBh



E. PCR Primers used to construct CRISPR/Cas Recombinant Clones.

DHFR-gRNA plasmid	**Top DHFR insert**	5'-[Phos]CACCGCGGCCCGGCAGATACCTGAG-3'
	Bottom DHFR insert	5'- [Phos]AAACCTCAGGTATCTGCCGGGCCGC-3'
	U6Fwd_Control DHFR insert	5'-GAGGGCCTATTTCCCATGAT-3'
Golden Gate Assembly	**CRISPR-step2-Fwd**	5'-GCCTTTTGCTGGCCTTTTGCTC-3'
	CRISPR-step2-Rev	5'-CGGGCCATTTACCGTAAGTTATGTAACG-3'
μΩ vector	**5#-rev DonorVector**	5'-TGCTATGTAACGCGGAACTCCATATATGGG-3'
	3#-fwd DonorVector	5'-CAAACACGTACGCGTACGATGCTCTAGAATG-3'
	5Fwd DonorVect Mo	5'-CCGCGTTACATAGCATCGTACGCGTACGTGTTTGGGACCTGCCCTGGCCACCGCTCCCCGGATCCATGGTGAGCAAGGG-3'
	3rev DonorVect Mo	5'-ACGCGTACGTGTTTGGCCCCGGCCCGGCAGATACCTTCAGGCACCGGGCTTGCG-3'
	Contr DonorUpstr	5'-GCCCTTAATTGTGAGCGGATAAC-3'
	Contr DonorDownst	5'-CACCAGGGCAAGGGTCTG-3'

F. PCR Primers used to control the insertion of the EGFP-2A-Puro cassette.

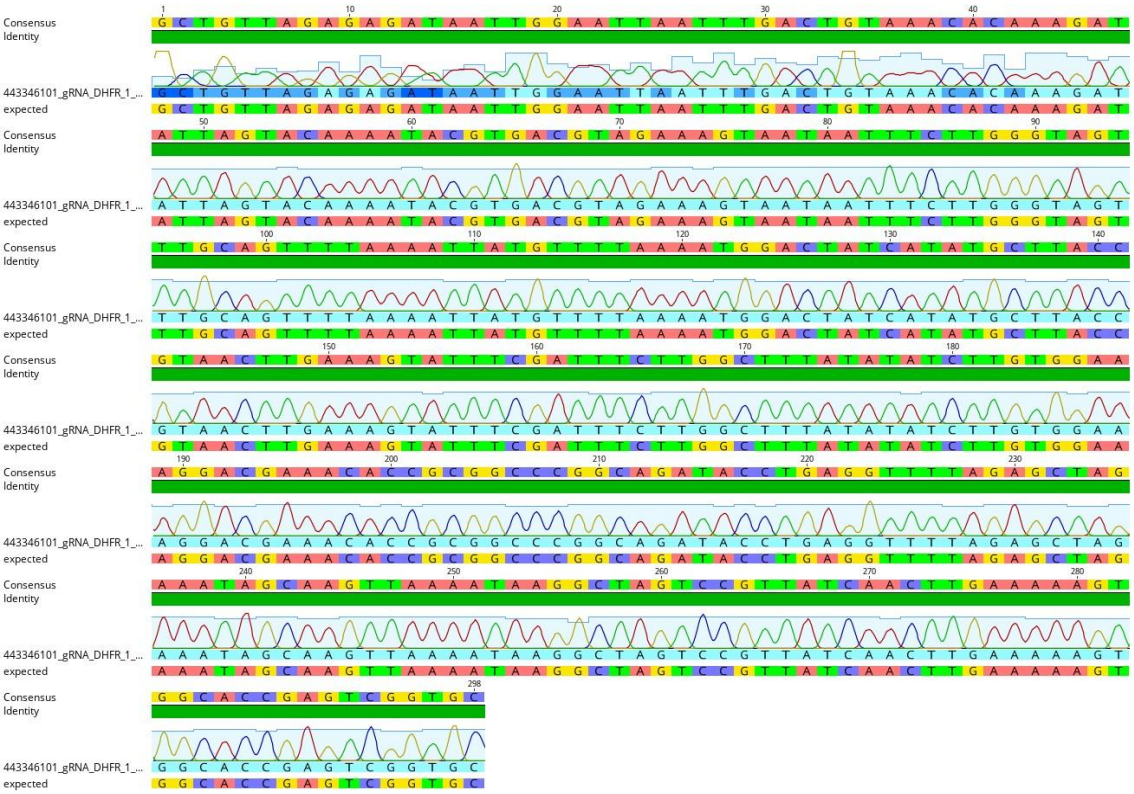
FBL Knock out	**Fwd FBL KIn**	5'-CTGGGGACCCTCCTTCATCAC-3'
	Rev FBL KIn	5'-CGCAATCCTGACAGCGCTGAAC-3'
DHFR knock out	**DHFR KO Fwd**	5'-GTCGCTGTGTCCCAGAAC-3'
	DHFR KO Rev	5'-GCAGAAATCAGCAACTGGG-3'
	Invert cassette Fwd	5'-CGCAGCAACAGATGGAAGG-3'
	Invert cassette Rev	5'-AACTTGTGGCCGTTTACG-3'

G. PCR Primers used to check for genomic DNA contamination of RNA samples (MTHFD1 R653Q Assay).

MTHFD1 R653Q Assay	**MTHFD1 Fwd**	5'-CACTCCAGTGTTTGTCCATG-3'
	MTHFD1 Rev	5'-GCATCTTGAGAGCCCTGAC-3'

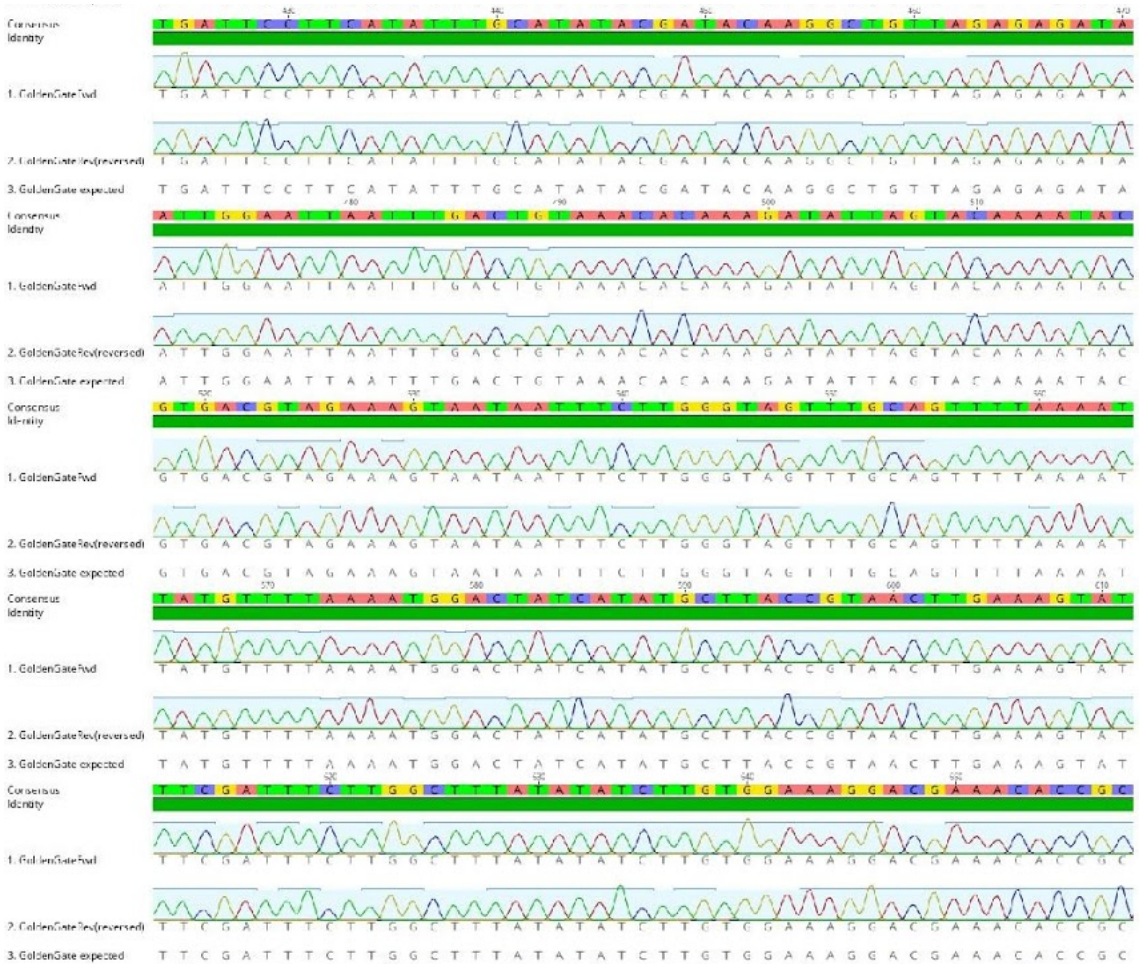
H. Sanger sequencing of DHFR gRNA.

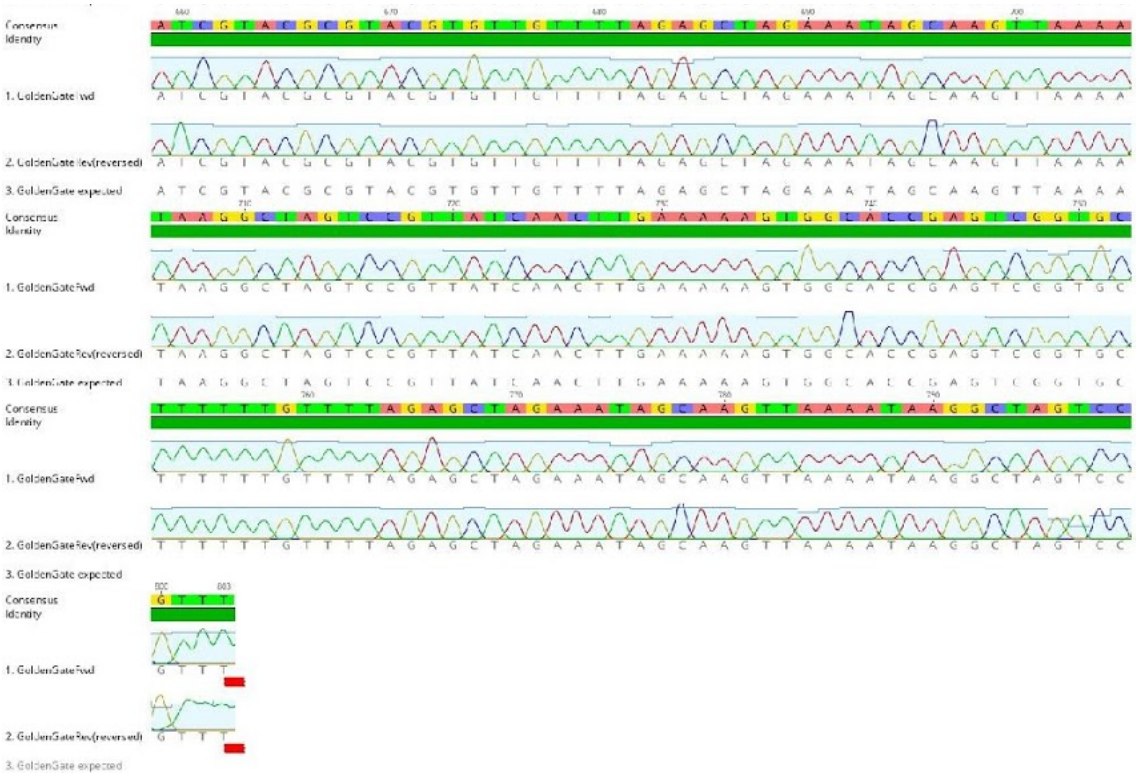
DHFR-specific gRNA inserted by annealed oligo cloning between U6 promoter and gRNA scaffold into pX330A-1x2 plasmid, thus creating DHFR-gRNA plasmid. The green bar “identity” shows the level of similarity between the sequenced product and the expected sequence (Geneious version 2019.2 created by Biomatters).



I. Sanger sequencing of the All-in-One vector.

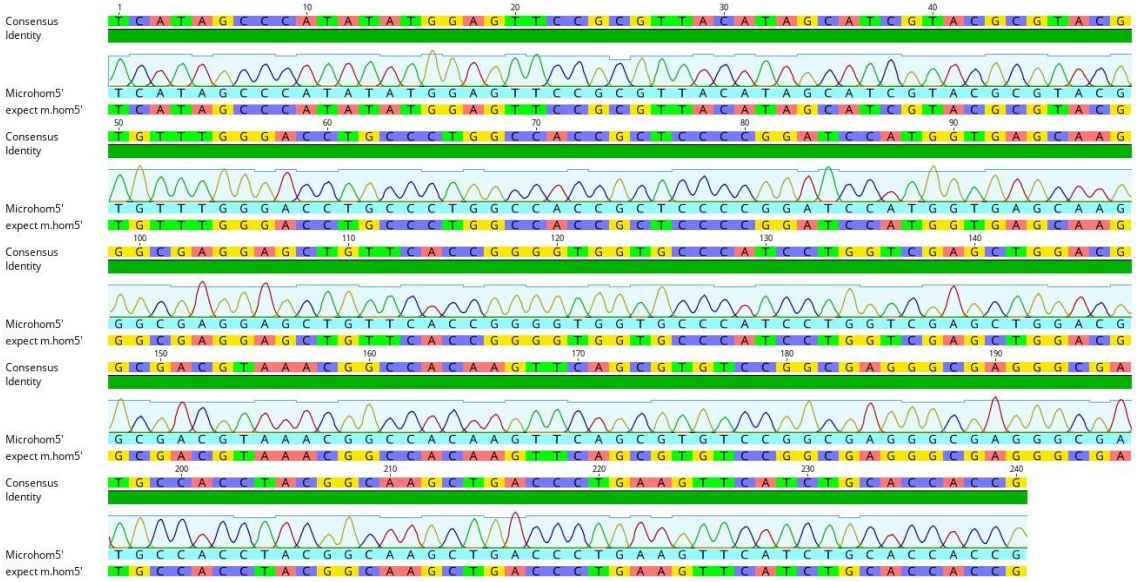
The All-in-One vector was obtained by Golden Gate Cloning, inserting the PITCh gRNA from pX330S-2-PITCh plasmid into DHFR-gRNA plasmid. The resulting All-in-One vector harbours the DHFR gRNA, followed by the PITCh gRNA and the *Cas9* gene (Geneious version 2019.2 created by Biomatters).




Consensus
Identity
1. GoldenGate fwd
2. GoldenGate Rev(reversed)
3. GoldenGate expected
Consensus
Identity
1. GoldenGate fwd
2. GoldenGate Rev(reversed)
3. GoldenGate expected
Consensus
Identity
1. GoldenGate fwd
2. GoldenGate Rev(reversed)
3. GoldenGate expected
Consensus
Identity
1. GoldenGate fwd
2. GoldenGate Rev(reversed)
3. GoldenGate expected

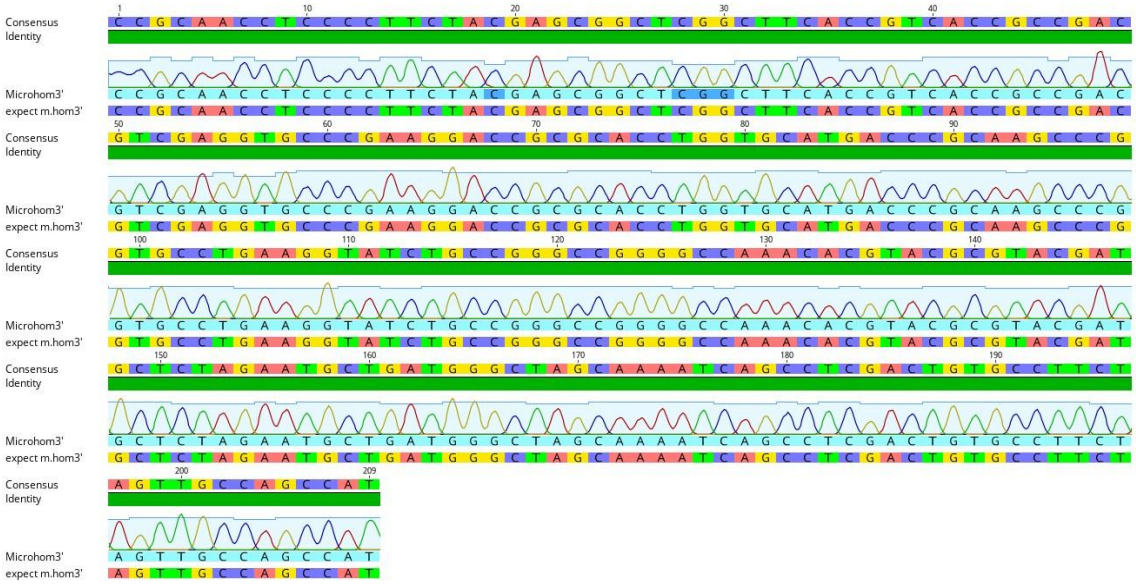
J. Sanger sequencing of the 5'end of the Ωμ vector.

DHFR-specific microhomologies replaced the FBL-specific microhomologies of pCRIS-PITChv2-FBL by PCR, thus obtaining the Ωμ vector (Geneious version 2019.2 created by Biomatters).



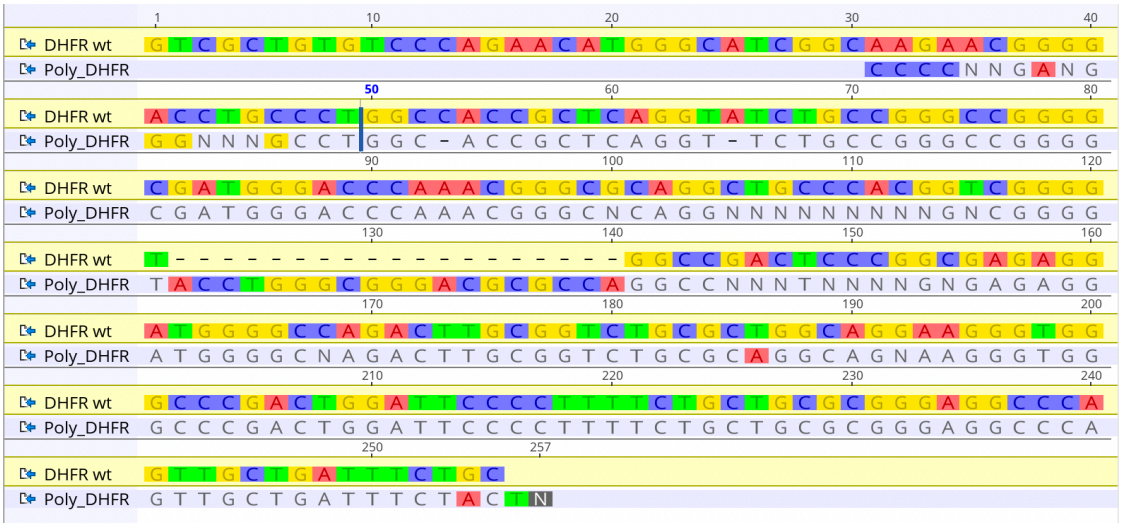
K. Sanger sequencing of the 3'end of the Ωμ vector.

The FBL-specific microhomologies of pCRIS-PITChv2-FBL were replaced by DHFR-specific microhomologies by PCR, thus obtaining the Ωμ vector (Geneious version 2019.2 created by Biomatters).



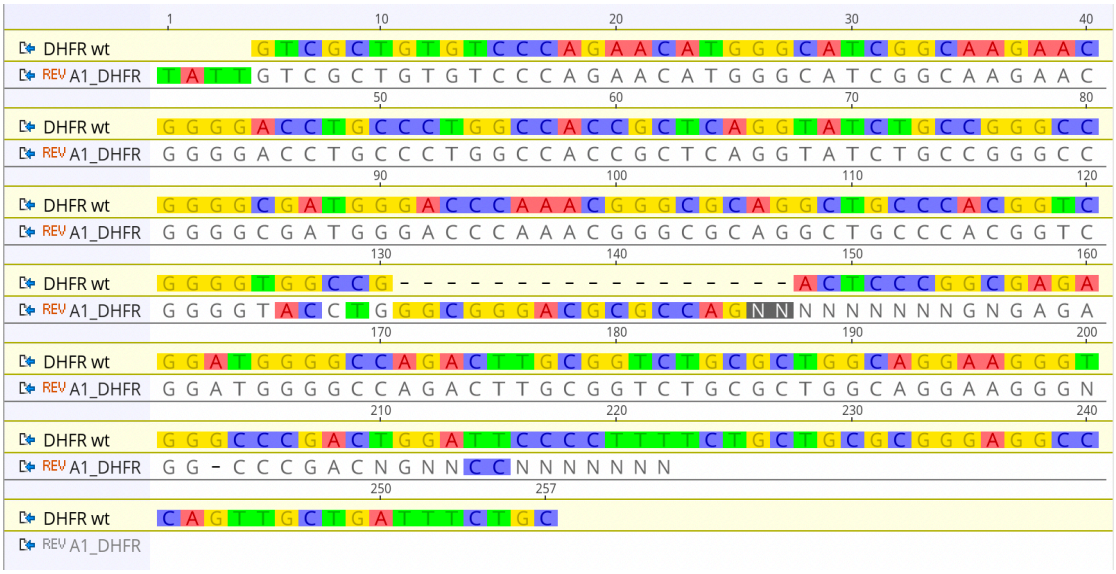
L. Sanger sequencing of DHFR cleavage area from the polyclonal line and alignment with DHFR wild type.

DNA extracted from the DHFR KO polyclonal line. The amplicon, obtained via the DHFR full-length assay, was purified and Sanger sequenced. The alignment between the sample (Poly_DHFR) and the DHFR wild type reference sequence is shown. The mismatches are highlighted in the Poly_DHFR sequence. Gaps are indicated with hyphens in either of the two sequences. Ambiguous bases are marked with an N (Geneious Prime® 2022.1.1).



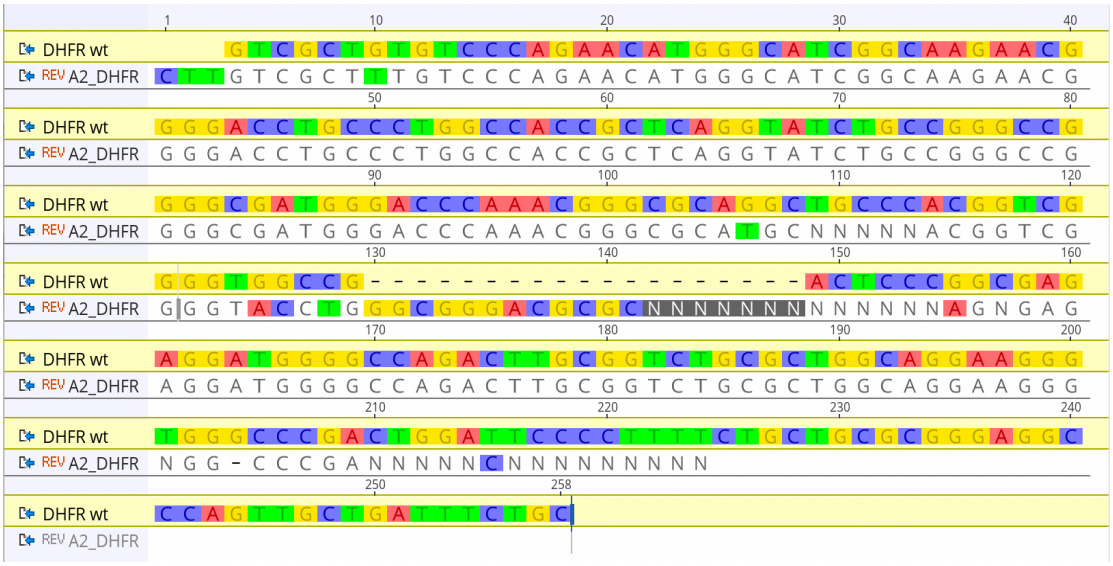
M. Sanger sequencing of DHFR cleavage area from the A1 monoclonal line and alignment with DHFR wild type.

DNA extracted from the DHFR KO A1 monoclonal line. The amplicon, obtained via the DHFR full-length assay, was purified and Sanger sequenced. The alignment between the sample (A1_DHFR) and the DHFR wild type reference sequence is shown. The mismatches are highlighted in the A1_DHFR sequence. Gaps are indicated with hyphens in either of the two sequences. Ambiguous bases are marked with an N (Geneious Prime® 2022.1.1).



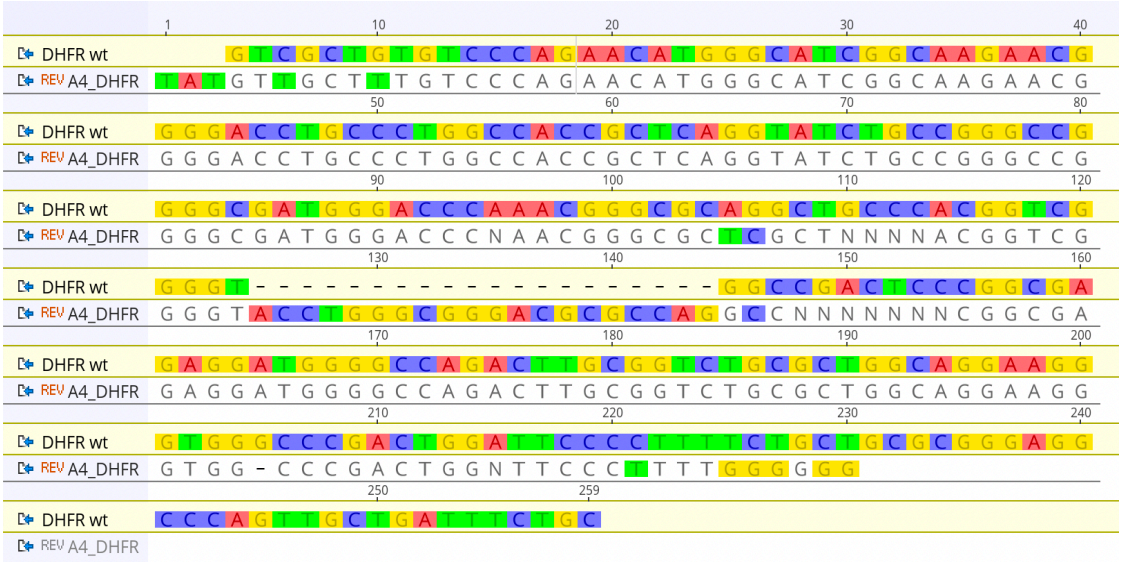
N. Sanger sequencing of DHFR cleavage area from the A2 monoclonal line and alignment with DHFR wild type.

DNA extracted from the DHFR KO A2 monoclonal line. The amplicon, obtained via the DHFR full-length assay, was purified and Sanger sequenced. The alignment between the sample (A2_DHFR) and the DHFR wild type reference sequence is shown. The mismatches are highlighted in the A2_DHFR sequence. Gaps are indicated with hyphens in either of the two sequences. Ambiguous bases are marked with an N (Geneious Prime® 2022.1.1).



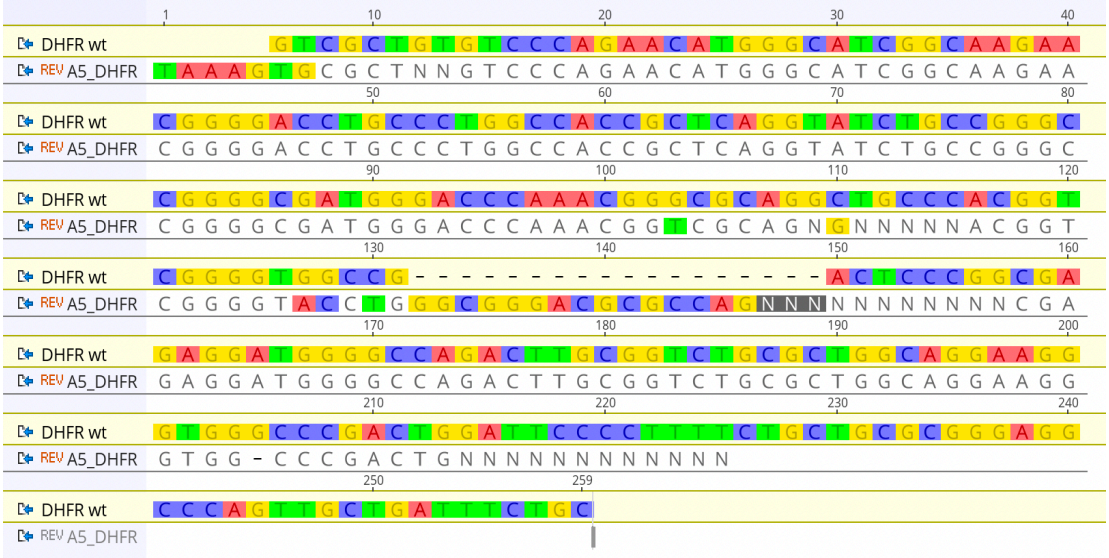
O. Sanger sequencing of DHFR cleavage area from the A4 monoclonal line and alignment with DHFR wild type.

DNA extracted from the DHFR KO A4 monoclonal line. The amplicon, obtained via the DHFR full-length assay, was purified and Sanger sequenced. The alignment between the sample (A4_DHFR) and the DHFR wild type reference sequence is shown. The mismatches are highlighted in the A4_DHFR sequence. Gaps are indicated with hyphens in either of the two sequences. Ambiguous bases are marked with an N (Geneious Prime® 2022.1.1).



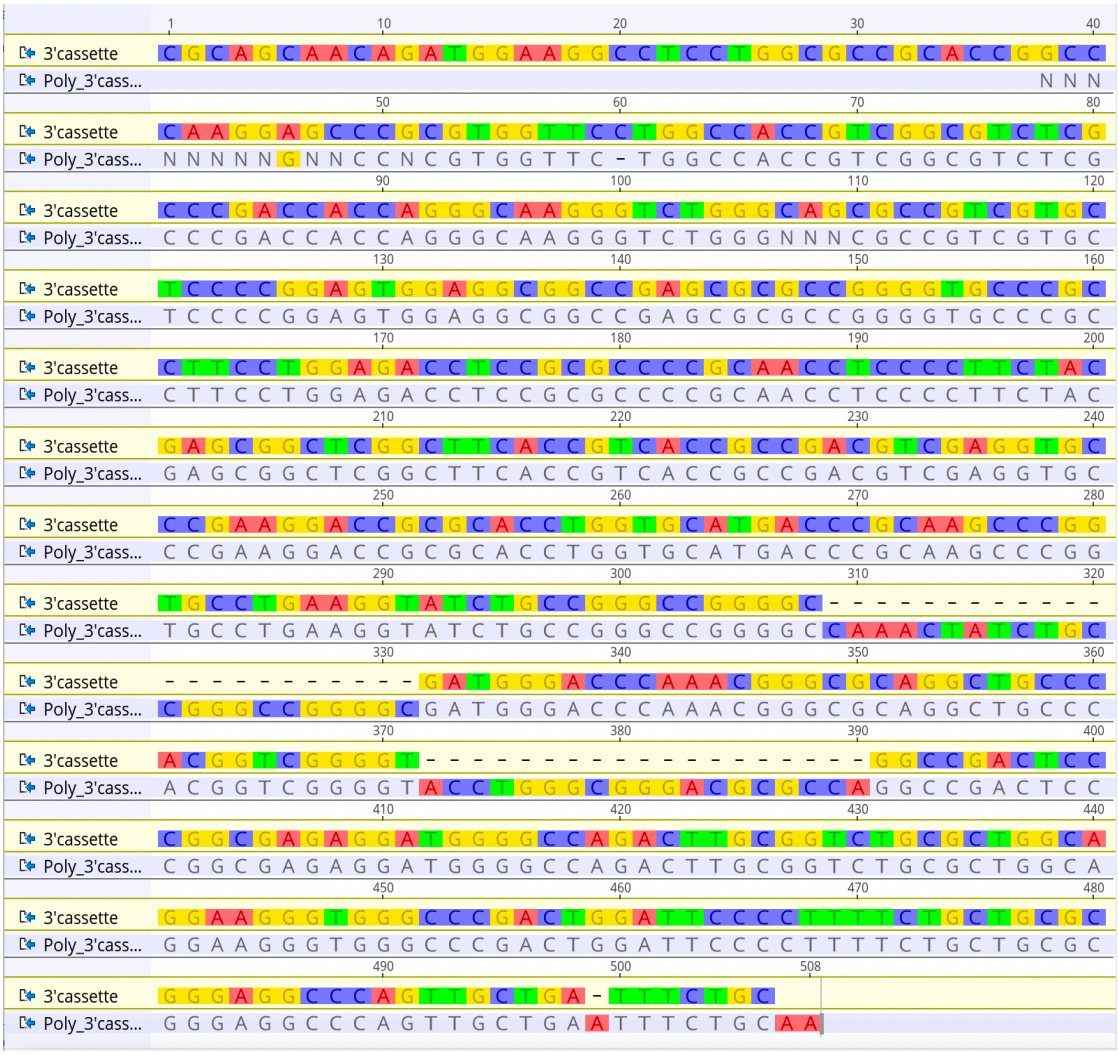
P. Sanger sequencing of DHFR cleavage area from the A5 monoclonal line and alignment with DHFR wild type.

DNA extracted from the DHFR KO A5 monoclonal line. The amplicon, obtained via the DHFR full-length assay, was purified and Sanger sequenced. The alignment between the sample (A5_DHFR) and the DHFR wild type reference sequence is shown. The mismatches are highlighted in the A5_DHFR sequence. Gaps are indicated with hyphens in either of the two sequences. Ambiguous bases are marked with an N (Geneious Prime® 2022.1.1).



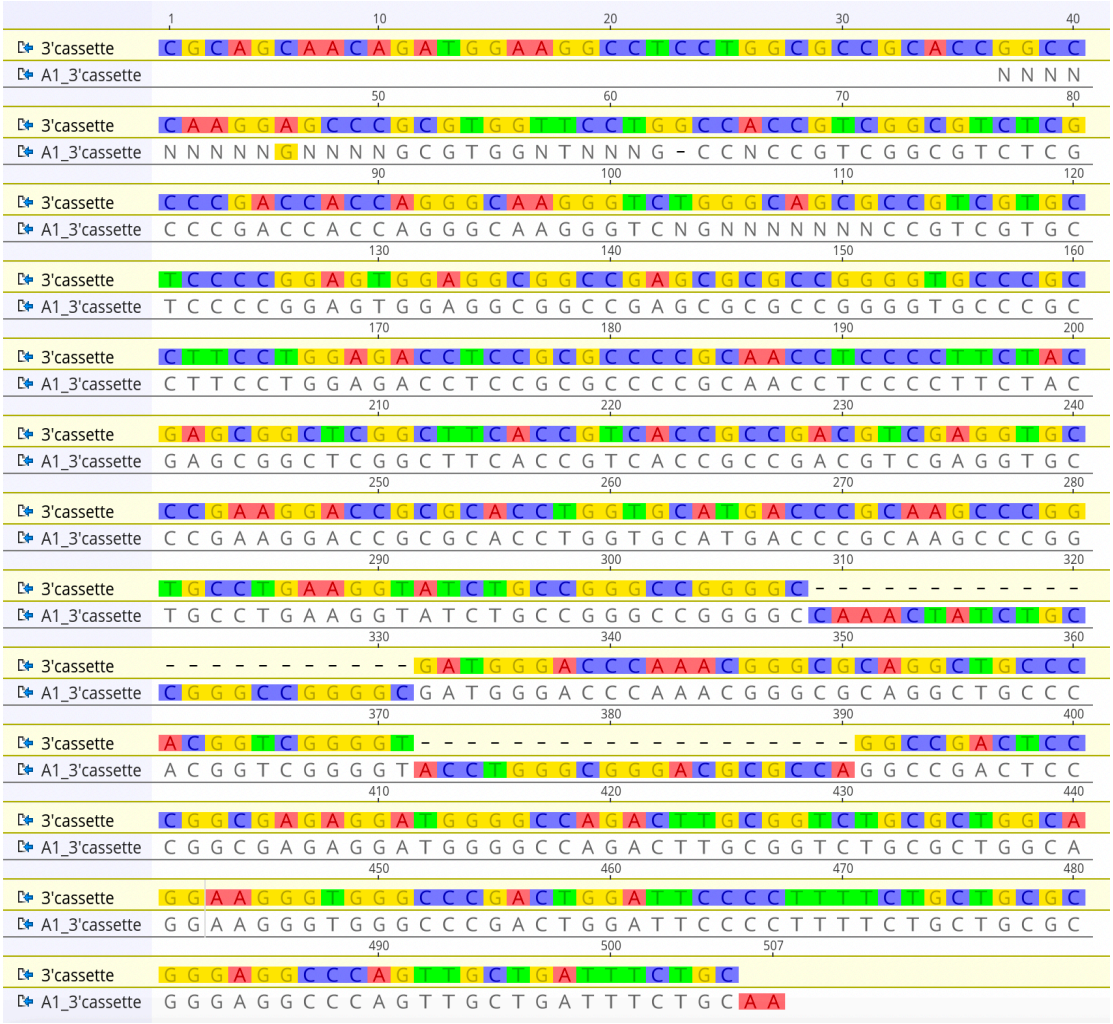
Q. Sanger sequencing of gene cassette insertion at its 3’ end on DHFR from the polyclonal line and alignment with expected amplicon.

DNA extracted from the DHFR KO polyclonal line. The amplicon, obtained via the 3’ end assay, was purified and Sanger sequenced. The alignment between the sample (Poly_3’cassette) and the reference sequence (3’cassette) is shown. Gaps are indicated with hyphens. Ambiguous bases are marked with an N (Geneious Prime® 2022.1.1).



R. Sanger sequencing of gene cassette insertion at its 3' end on DHFR from the A1 monoclonal line and alignment with expected amplicon.

DNA extracted from the DHFR KO A1 monoclonal line. The amplicon, obtained via the 3' end assay, was purified and Sanger sequenced. The alignment between the sample (A1_3'cassette) and the reference sequence (3'cassette) is shown. Gaps are indicated with hyphens. Ambiguous bases are marked with an N (Geneious Prime® 2022.1.1).



S. R Studio Analysis on the effects of HT/NEAA on HepG2 cell growth.

A preliminary test on cell growth has been conducted to assess the influence of HT/NEAA supplementation on native cells. Cell count has been performed every day on two groups of cells, supplemented and non-supplemented, over 9 days. HepG2 doubling time is 48-hours, so the expected growth has also been considered. R Studio was used to visualise the growth trends (normal and linear regression model) and assess whether statistically relevant differences were present.

Graphics: how to visualise growth curves in a graph

Count measurements expressed as vectors

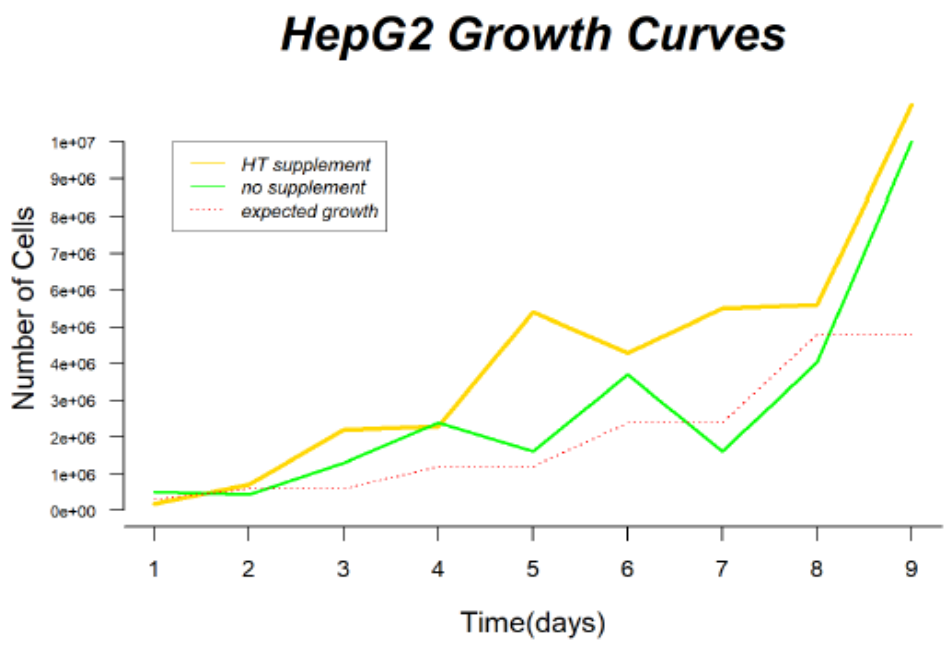
```
days=c(1:9)
expectedgrowth=c(300000, 600000, 600000, 1200000, 1200000, 2400000, 2400000, 4800000, 4800000)  #predicted
noHT_tot=c(520000, 440000, 1300000, 2400000, 1600000, 3700000, 1600000, 4030000, 10000000)      #un-treated

HT_tot= c(1.8e+05, 7.1e+05, 2.2e+06, 2.3e+06, 5.4e+06, 4.3e+06, 5.5e+06, 5.6e+06, 1.1e+07)      #treated
```

Assembling a table

```
t=data.frame(days,expectedgrowth,noHT_tot,HT_tot)
t
```

Visualising the data

```
plot(t$HT_tot, type="l",ylim = range(t),lwd=3,col="gold",
     main="HepG2 Growth Curves", col.main="black", cex.main=2,
     font.main=4, xlab="Time(days)", ylab="Number of Cells",
     col.lab="black", cex.lab=1.25,xaxt = "n", yaxt = "n", bty="n")
axis(side=1, at=c(0:10))
axis(side=2, at=seq(0, 10000000, by=1000000),cex.axis= 0.7, las=2)
lines(t$noHT_tot, type = "l", lwd=2, col= "green")
points(t$expectedgrowth,type="l",lty=3,col=450)
legend(1.2,10e+06 , legend=c("HT supplement", "no supplement", "expected growth"),
       col=c("gold", "green", "450"), lty=c(1,1,3), cex=0.8, text.font=3, box.lty=1, box.lwd = .5)
```



Welch test: when the samples are independent

```
t.test(my_data$cells_no[my_data$group=="expect"],
       my_data$cells_no[my_data$group=="no_HT"])
```

```
##
##  Welch Two Sample t-test
##
## data:  my_data$cells_no[my_data$group == "expect"] and my_data$cells_no[my_data$group == "no_HT"]
## t = -0.70784, df = 12.91, p-value = 0.4916
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3283930  1663930
## sample estimates:
## mean of x mean of y
##   2033333   2843333
```

```
t.test(my_data$cells_no[my_data$group=="expect"],
       my_data$cells_no[my_data$group=="HT"])
```

```
##
##  Welch Two Sample t-test
##
## data:  my_data$cells_no[my_data$group == "expect"] and my_data$cells_no[my_data$group == "HT"]
## t = -1.6879, df = 12.103, p-value = 0.117
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4805633   607855
## sample estimates:
## mean of x mean of y
##   2033333   4132222
```

```
t.test(my_data$cells_no[my_data$group=="no_HT"],
       my_data$cells_no[my_data$group=="HT"])
```

```
##
##  Welch Two Sample t-test
##
## data:  my_data$cells_no[my_data$group == "no_HT"] and my_data$cells_no[my_data$group == "HT"]
## t = -0.87158, df = 15.815, p-value = 0.3965
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4426793  1849016
## sample estimates:
## mean of x mean of y
##   2843333   4132222
```

ANOVA: a more comprehensive analysis

```
library(dplyr)
```

```
library("ggpubr")
```

Import data, summarise and calculate mean and sd

```
my_data <- read.csv("newtable.csv")
show(my_data)
```

```
group_by(my_data, group) %>%
  summarise(
    count = n(),
    mean = mean(cells_no, na.rm = TRUE),
    sd = sd(cells_no, na.rm = TRUE)
  )
```

Computing the analysis of variance

```
res.aov <- aov(cells_no ~ group, data = my_data)
summary(res.aov)
```

```
##             Df    Sum Sq   Mean Sq F value Pr(>F)
## group        2 2.017e+13 1.008e+13   1.333  0.282
## Residuals   24 1.815e+14 7.564e+12
```

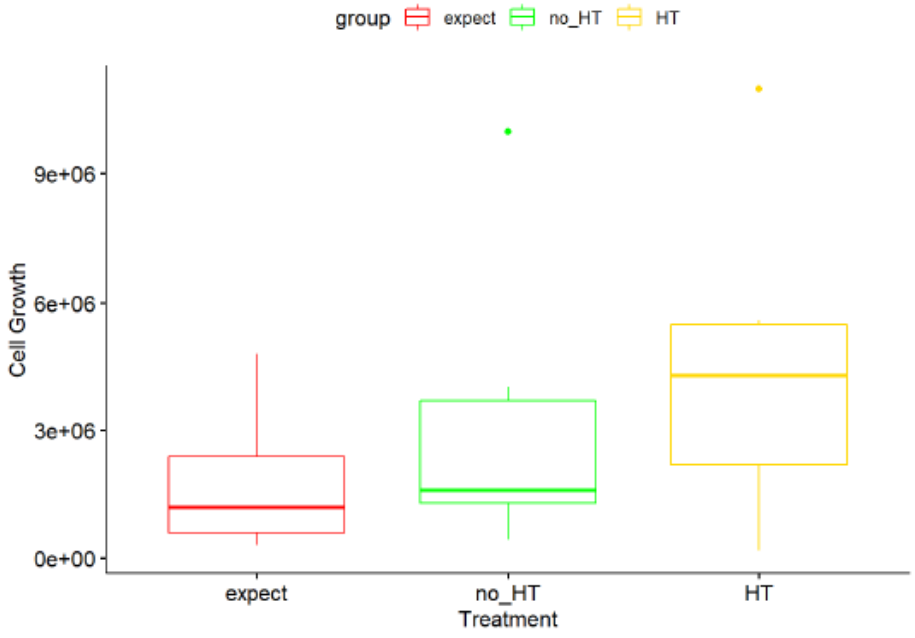
Tukey multiple comparisons of means

```
TukeyHSD(res.aov)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = cells_no ~ group, data = my_data)
##
## $group
##                  diff      lwr     upr     p adj
## HT-expect     2098889 -1138796 5336574 0.2572515
## no_HT-expect   810000 -2427685 4047685 0.8080659
## no_HT-HT     -1288889 -4526574 1948796 0.5875646
```

Box plots: Plot growth by group and color by group

```
ggboxplot(my_data, x = "group", y = "cells_no",
          color = "group", palette = c("450", "green", "gold"),
          order = c("expect", "no_HT", "HT"),
          ylab = "Cell Growth", xlab = "Treatment")
```



Linear Regression: statistical analysis and graphics

To better compare the trend of growth of the treated population (HT/NEAA) and untreated one, a simple linear regression has been applied. The aim is to summarise the data and study the relationship between time (independent variable) and number of cells (dependent variable).

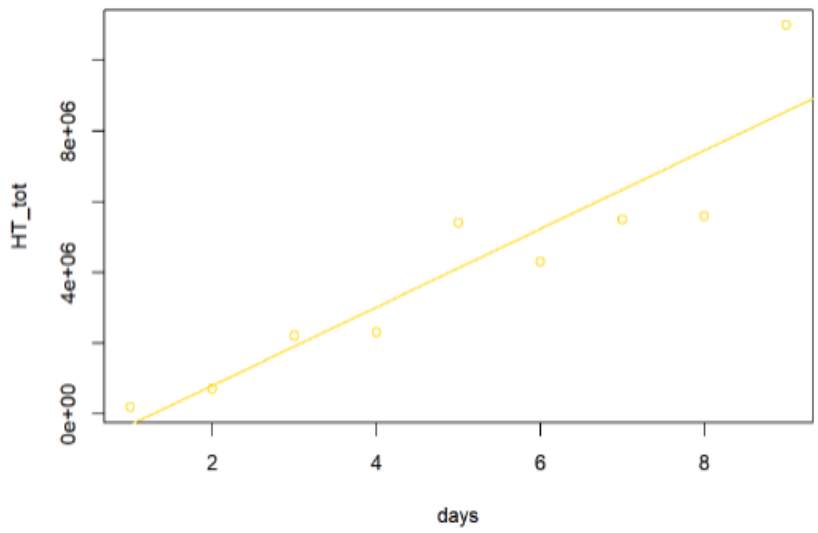
```
regression=lm(HT_tot~days)
summary(regression)
```

```
##
## Call:
## lm(formula = HT_tot ~ days)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1859722  -850556   -94722   484444  2431111
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1413611    1006424  -1.405 0.202929
## days         1109167     178846   6.202 0.000444 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1385000 on 7 degrees of freedom
## Multiple R-squared:  0.846,  Adjusted R-squared:  0.824
## F-statistic: 38.46 on 1 and 7 DF,  p-value: 0.0004445
```

```
confint(regression)
```

```
##                 2.5 %    97.5 %
## (Intercept) -3793426.0  966203.8
## days          686262.5 1532070.9
```

```
df <- data.frame(days,HT_tot)
plot(days,HT_tot,col="gold")
abline(lm(HT_tot~days),col="gold")
```



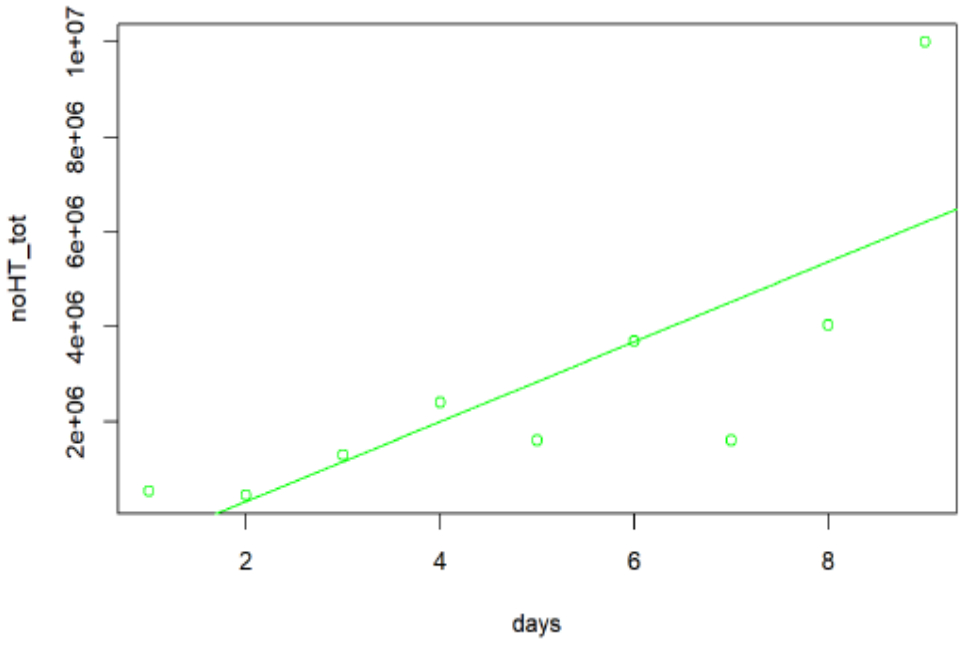
```
regression=lm(noHT_tot~days)
summary(regression)
```

```
##
## Call:
## lm(formula = noHT_tot ~ days)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2929667 -1243333   126167   399833  3784000
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1372500    1441196  -0.952   0.3726
## days          843167     256107   3.292   0.0133 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1984000 on 7 degrees of freedom
## Multiple R-squared:  0.6076, Adjusted R-squared:  0.5515
## F-statistic: 10.84 on 1 and 7 DF,  p-value: 0.01326
```

```
confint(regression)
```

```
##                 2.5 %  97.5 %
## (Intercept) -4780387.7 2035388
## days          237569.1 1448764
```

```
df_noHT <- data.frame(days,noHT_tot)
plot(days,noHT_tot,col="green")
abline(lm(noHT_tot~days),col="green")
```



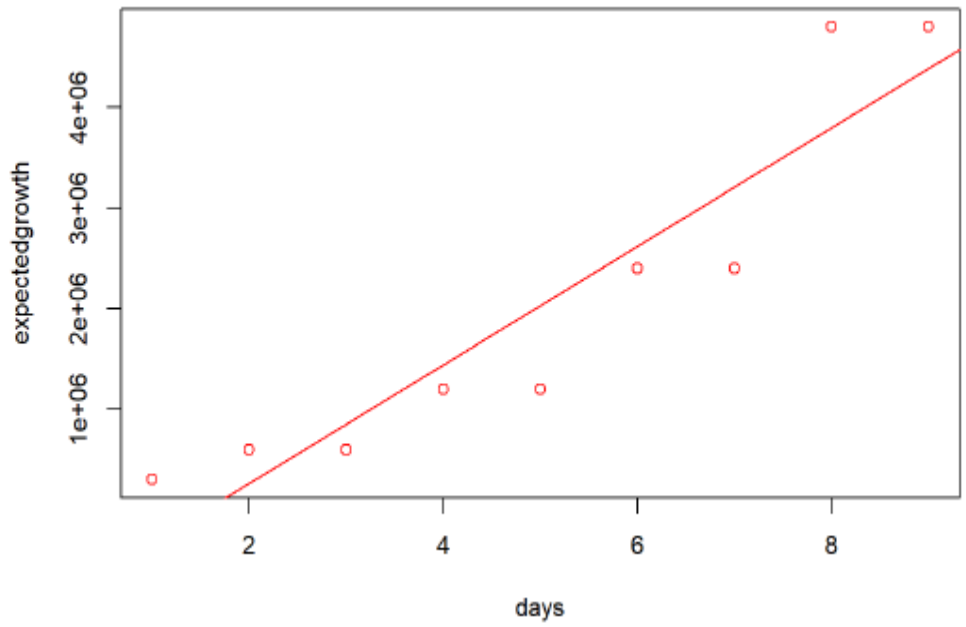
```
regression=lm(expectedgrowth~days)
summary(regression)
```

```
##
## Call:
## lm(formula = expectedgrowth ~ days)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -833333 -253333 -223333  406667  996667
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -916667     490732  -1.868 0.103989
## days          590000      87205   6.766 0.000261 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 675500 on 7 degrees of freedom
## Multiple R-squared:  0.8674, Adjusted R-squared:  0.8484
## F-statistic: 45.77 on 1 and 7 DF,  p-value: 0.0002613
```

```
confint(regression)
```

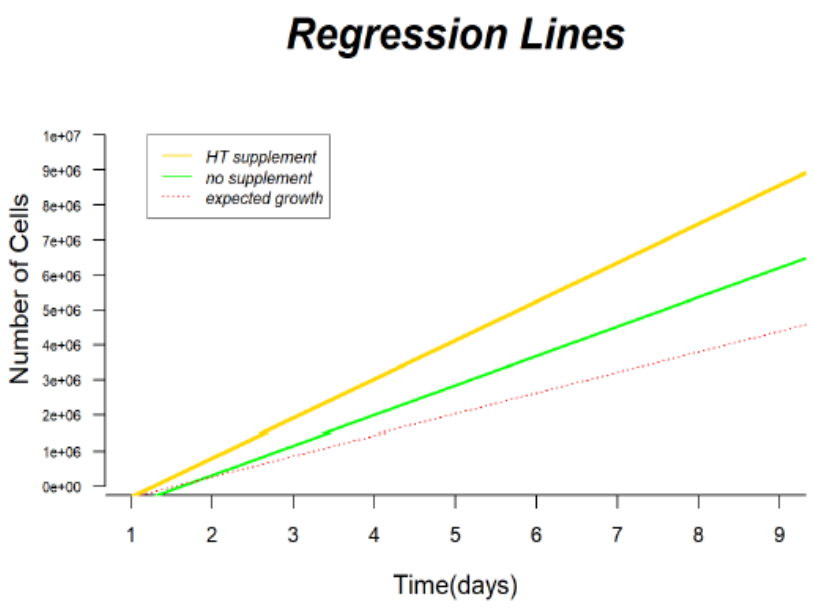
```
##                 2.5 %   97.5 %
## (Intercept) -2077062.4 243729.1
## days          383792.3 796207.7
```

```
df_expect <- data.frame(days,expectedgrowth)
plot(days,expectedgrowth,col="450")
abline(lm(expectedgrowth~days),col="450")
```



Plotting the regression lines together

```
plot(days,HT_tot,col="white", pch=20, main="Regression Lines", col.main="black", cex.main=2,
     font.main=4, xlab="Time(days)", ylab="Number of Cells",
     col.lab="black", cex.lab=1.25,xaxt = "n", yaxt = "n", bty="n")
axis(side=1, at=c(0:10))
axis(side=2, at=seq(0, 10000000, by=1000000),cex.axis= 0.7, las=2)
points(df_noHT$noHT_tot, col="white", pch=20)
points(df_expect$expectedgrowth, col="white", pch=20)

abline(lm(HT_tot~days),col="gold", lwd=3)
abline(lm(noHT_tot~days),col="green", lwd=2)
abline(lm(expectedgrowth~days),col="450",lty=3, lwd=1)

legend(1.2,10e+06 , legend=c("HT supplement", "no supplement", "expected growth"),
       col=c("gold", "green", "450"), lty=c(1,1,3), cex=0.8, text.font=3, box.lty=1, box.lwd = .5)
```



T. Direct PCR of genomic DNA from mammalian cells.

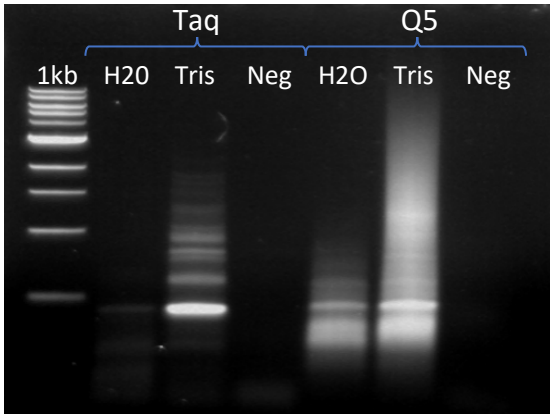
Direct PCR was employed to facilitate the screening of multiple cell populations in search of DHFR2 KO positive transformants. The use of this method allows skipping the DNA extraction step, which can only be performed on cell amounts larger than those typical of 96 well plates. The cell suspension was boiled in either water or Tris, buffer, ice shocked and centrifuged. The supernatant was directly added to the PCR reaction mix as follows:

DHFR2 Deletion Detection Assay - Taq		**DHFR2 Deletion Detection Assay – Q5**	
5 µl	5X Q5 Buffer	2.5 µl	10X Taq Buffer
0.5 µl	10 mM dNTPs	2.5 µl	2.5 mM dNTPs
1.25 µl	10 µM Fwd primer DHFR2 Fb	2 µl	25 mM $MgCl_2$
1.25 µl	10 µM Rev primer DHFR2 Re	0.5 µl	10 µM Fwd primer DHFR2 Fb
0.25 µl	Q5 DNA Pol	0.5 µl	10 µM Rev primer DHFR2 Re
2 µl	Supernatant (DNA)	0.25 µl	Taq DNA Pol
14.75 µl	H_2O	2 µl	Supernatant (DNA)
		14.75	H_2O

PCR cycling conditions

98°C, 30 seconds		94°C, 5 minutes	
98°C, 10 seconds		94°C, 1 minute	
56.9°C, 30 seconds	35 cycles	59.6°C, 1 minute	35 cycles
72°C, 1.5 minutes		72°C, 2.5 minutes	
72°C, 2 minutes		72°C, 10 minutes	

All reactions produced bands at the expected length (453bp), even though the Tris buffer treatment coupled with Taq amplification seems to give the most promising results. However, the assay required further optimisation.



U. Primers used for the DHFR2 Deletion detection Assay (and direct PCR).

DHFR2 Deletion Detection	**DHFR2 Fb**	5'-TACCAAATGCGTGAAGACCA-3'
	DHFR2 Re	5'-GGTTGTTCCATTGCACTCCG-3'

V. RT-PCR primers used for the 201-202 Assay (Detection of DHFR2 201 & 202 isoforms).

201-202 Assay	**FPDHFRL1Var1&2**	5'-CTTCCGGTAGCTGGTAAAGG-3'
	RPDHFRL1Var1&2	5'-CCCATGTTTTGGGACACAG-3'

W. RT-qPCR primers used to test the DHFR2 knockout and DHFR knock-down cell lines.

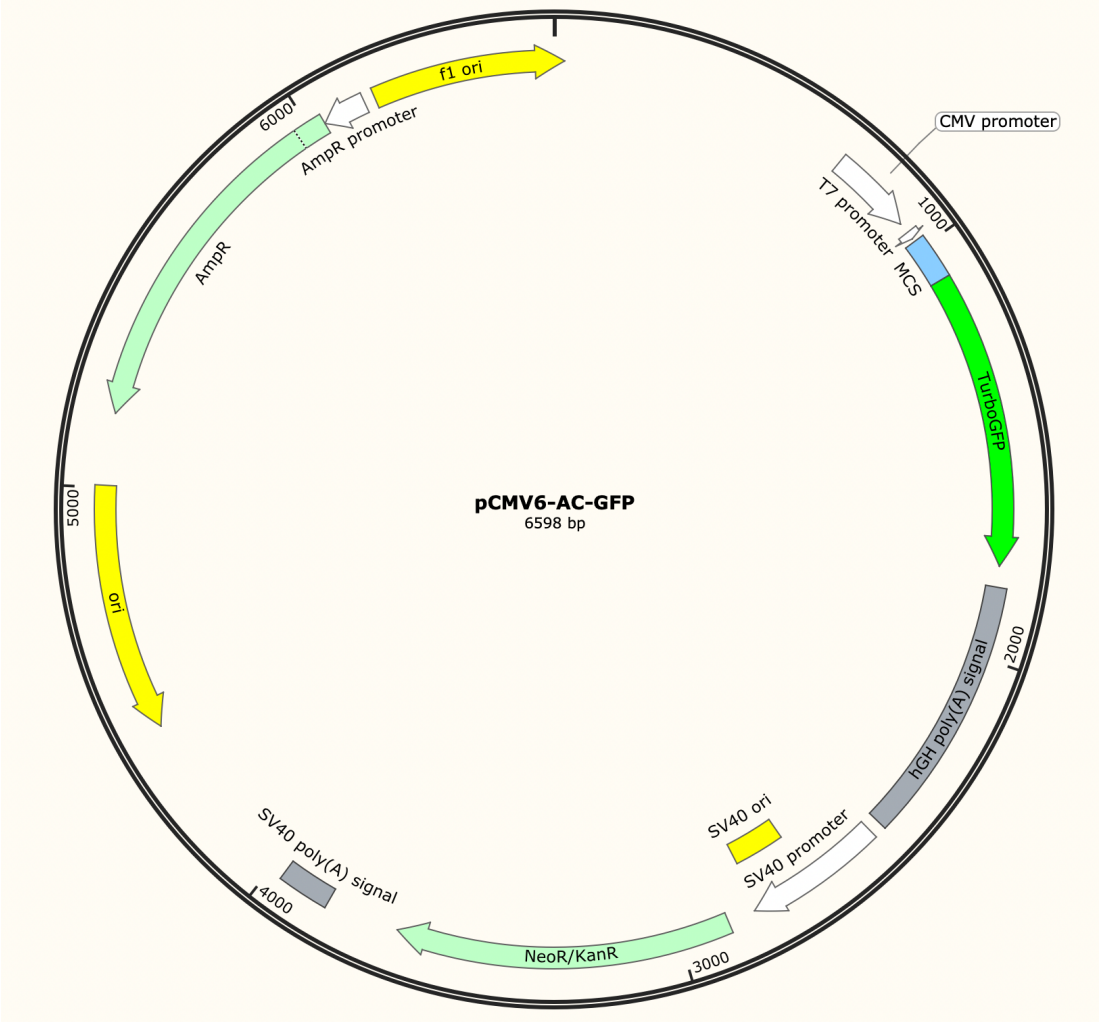
		PRIMER SEQUENCE	UPL PROBE
GUS	**Fwd**	5′ -GGTACGAACGGGAGGTGAT - 3′	#73
	Rev	5′ -CACGATGGCATAGGAATGG - 3′	
GAPDH	**Fwd**	5′- CTCTGCTCCTCCTGTTCGAC -3′	#60
	Rev	5′- ACGACCAAATCCGTTGACTC -3′	
TBP	**Fwd**	5′- TTGGGTTTTCCAGCTAAGTTCT -3′	#24
	Rev	5′- CCAGGAAATAACTCTGGCTCA -3′	
RPS13	**Fwd**	5′- GGTTGAAGTTGACATCTGACGA -3′	#68
	Rev	5′- TGTGCAACACCATGTGAATCT -3′	
DHFR2_All	**Fwd**	5′- AATTTCGCGGCATTCTTG -3′	#72
	Rev	5′- GGTTAACACCTCCGAACTTGC -3′	
DHFR mains	**Fwd**	5′- CGCGAGCACGCCGCGACCCTGCGT -3′	#87
	Rev	5′- CGCCCCCCTCGTCCCCATT -3′	
SHMT1	**Fwd**	5′- TGGTGTAGAAATGGCCTCCT -3′	#58
	Rev	5′- TCTCACACCAGGATGGGACT-3′	
SHMT2	**Fwd**	5′- GATCCTGAGATGTGGGAGTTG -3′	#1
	Rev	5′- CAGCTCGGCTGCAGAAGT-3′	
TYMS	**Fwd**	5′- CCCAGTTTATGGCTTCCAGT-3′	#43
	Rev	5′- GCAGTTGGTCAACTCCCTGT-3′	

ALDH1L1	**Fwd**	5′- ACCGCAACCTGACCTTGA-3′	#12
	Rev	5′- TTCCAGGACAGCATCATCAG-3′	
SARDH	**Fwd**	5′- CACCGAGAGTGACCTGACTG-3′	#83
	Rev	5′- GCCCATGGCCAGGTAGTA-3′	
NSUN3	**Fwd**	5′- CGGAGATGCCTGGAATACA-3′	#56
	Rev	5′- CAAAAGGATAATTGAATCGGTTAAG-3′	
ARL13B	**Fwd**	5′- TTGAGCCTCTTGGTGAAACAC-3′	#16
	Rev	5′- TCACTGTTAGGTCGCTGTGG-3′	
tGFP	**Fwd**	5′- ACGATCTGGATGGCAGCTTC-3′	#41
	Rev	5′- TCCACCACGGAGCTGTAGTA-3′	

X. pCMV6-AC-GFP. Mammalian expression vector (Origene).

Plasmid used in complementation experiment as quality control.

C-Tag	TurboGFP
E. coli Selection	Ampicillin
Cell Selection	Neomycin
Features	• TurboGFP (excitation/ emission max = 482/ 502 nm). • Structually it forms dimers. • Such clones are the best for monitoring protein expression, localization, translocation and co-localization with other protein(s).



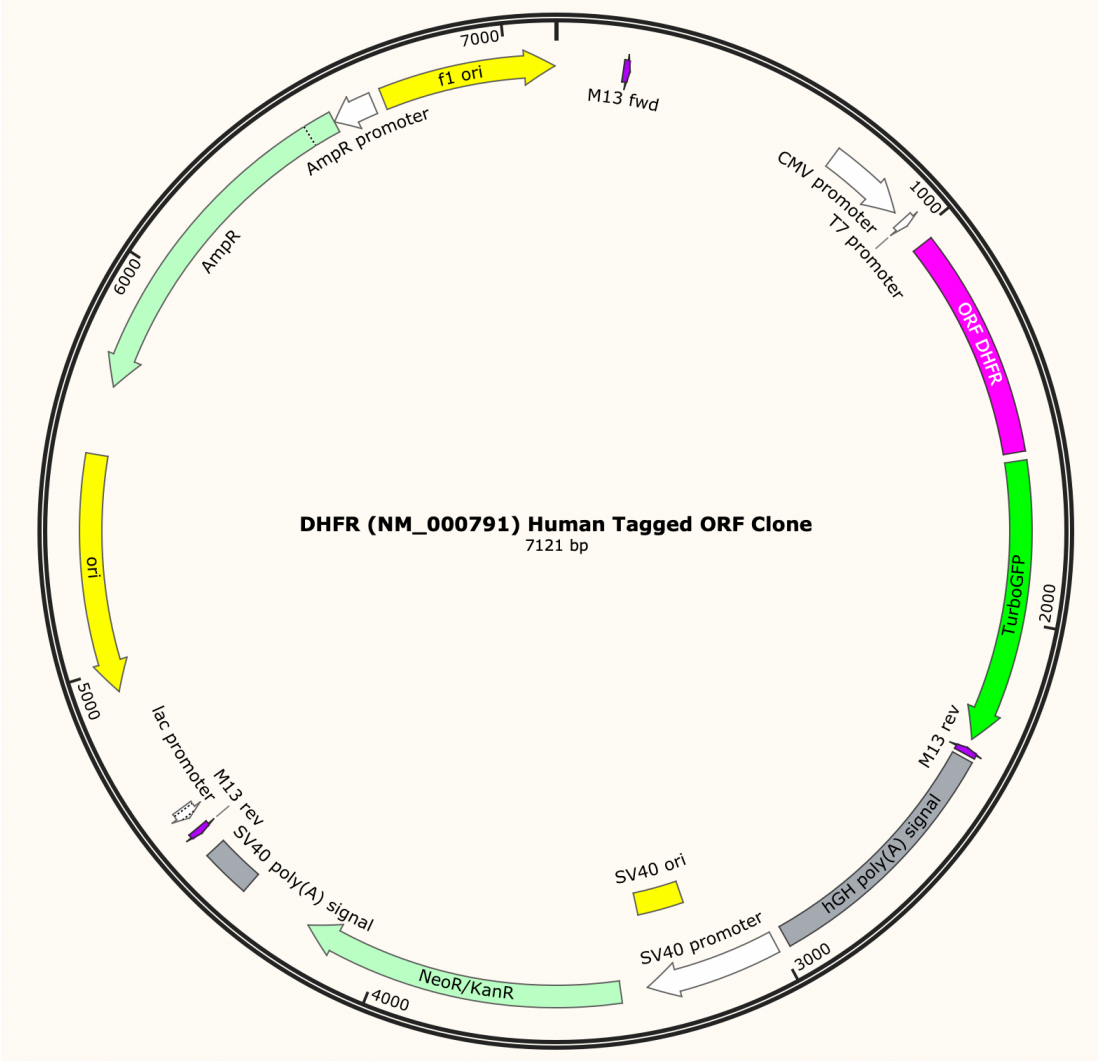
Y. pCMV6-AC-DHFR2-GFP (variant 1). Mammalian expression vector (Origene).
DHFR2 (NM_001195643) Human Tagged ORF Clone. Plasmid used to complement DHFR2 knockout cell line.

Type	Human Tagged ORF Clone
Tag	TurboGFP
Symbol	DHFR2
Synonyms	DHFRL1; DHFRP4
Vector	pCMV6-AC-GFP
E. coli Selection	Ampicillin (100 ug/mL)
Cell Selection	Neomycin
ACCN	NM_001195643

Z. pCMV6-AC-DHFR-GFP. Mammalian expression vector (Origene).

DHFR (NM_000791) Human Tagged ORF Clone. Plasmid used to complement DHFR knock-down cell line.

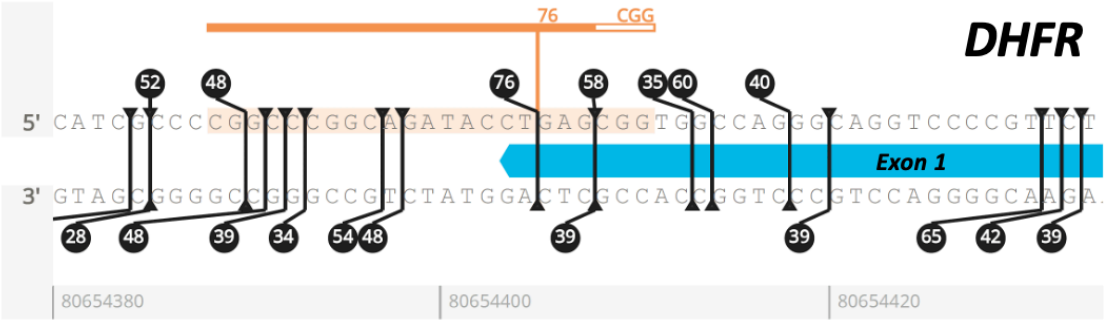
Tag	TurboGFP
Symbol	DHFR
Synonyms	DHFRP1; DYR
Vector	pCMV6-AC-GFP
E. coli Selection	Ampicillin (100 ug/mL)
Cell Selection	Neomycin
ACCN	NM_000791
ORF Size	561 bp



AA. Coefficients of intensity correlation of colocalizing objects in a dual-colour image. Analysis of HCR IHC + RNA-FISH on cultured HepG2 cells. The probe pairs are indicated in the first column. Pearson's and Mander's coefficients indicate the degree of correlation and co-occurrence, respectively. The statistical significance of Pearson's coefficient is calculated via Costes' test.

	Pearson's coefficient	*Manders' coefficients*		*Costes' test*
		M1	M2	
RPS3 + DHFR2 201	0.379	0.998	0.728	0.0
RPS3 + DHFR2 202	0.05	0.385	0.434	0.251
PSPC1 + DHFR2 201	0.683	0.998	0.8	0.473
PSPC1 + DHFR2 202	0.118	0.469	0.116	0.581
DHFR + DHFR2 201	0.617	0.999	0.576	0.384
DHFR + DHFR2 202	0.087	0.923	0.165	0.0
DHFR RNA + DHFR2 201	0.989	0.998	0.992	0.877
DHFR RNA + DHFR2 202	0.491	0.323	0.053	0.94
DHFR RNA + DHFR	0.94	0.893	0.999	0.889

BB. Screening of suitable PAM sites across DHFR gene (zoom on exon1-intron1).
DHFR gene screened via DESKGen online tool to selected the most suitable PAM/gRNA for knockout. The image shows a section of the gene where the best PAM site was located, followed by the list of other suitable PAMs. The chosen PAM and relative gRNA sit on the splice junction between exon 1 and intron1; it presents the highest on- and off-target activity.



	Sequence	PAM	Exon	Activity	Off-target
☑	CGGCCCGGCAGATACCTGAG	CGG	1	76	90
☐	ATCTGCCGGGCCGGGGCGAT	GGG	-	38	unscored
☐	CCGTTTGGGTCCCATCGCCC	CGG	-	52	unscored
☐	TATCTGCCGGGCCGGGGCGA	TGG	-	28	unscored
☐	TGGGTCCCATCGCCCCGGCC	CGG	-	48	unscored
☐	CTCAGGTATCTGCCGGGCCG	GGG	-	48	unscored
☐	GCTCAGGTATCTGCCGGGCC	GGG	-	39	unscored
☐	CGCTCAGGTATCTGCCGGGC	CGG	-	34	unscored
☐	CCACCGCTCAGGTATCTGCC	GGG	-	54	unscored
☐	GCCACCGCTCAGGTATCTGC	CGG	-	48	unscored

References

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). Academic Press. https://doi.org/10.1016/S0083-6729(08)00409-3

Acebron, S. P., Karaulanov, E., Berger, B. S., Huang, Y.-L., & Niehrs, C. (2014). Mitotic Wnt Signaling Promotes Protein Stabilization and Regulates Cell Size. *Molecular Cell*, *54*(4), 663–674. https://doi.org/10.1016/j.molcel.2014.04.014

Adesina, A. S., Luk, L. Y. P., & Allemann, R. K. (2021). Cryo-kinetics Reveal Dynamic Effects on the Chemistry of Human Dihydrofolate Reductase. *ChemBioChem*, *22*(14), 2410–2414. https://doi.org/10.1002/cbic.202100017

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. *Nature Communications*, *9*(1), 1911. https://doi.org/10.1038/s41467-018-04252-2

Agudelo, D., Duringer, A., Bozoyan, L., Huard, C. C., Carter, S., Loehr, J., Synodinou, D., Drouin, M., Salsman, J., Dellaire, G., Laganière, J., & Doyon, Y. (2017). Marker-free coselection for CRISPR-driven genome editing in human cells. *Nature Methods*, *14*(6), 615–620. https://doi.org/10.1038/nmeth.4265

Aida, T., Nakade, S., Sakuma, T., Izu, Y., Oishi, A., Mochida, K., Ishikubo, H., Usami, T., Aizawa, H., Yamamoto, T., & Tanaka, K. (2016). Gene cassette knock-in in mammalian cells and zygotes by enhanced MMEJ. *BMC Genomics*, *17*(1), 979. https://doi.org/10.1186/s12864-016-3331-9

Albrecht, L. V., Bui, M. H., & Robertis, E. M. D. (2019). Canonical Wnt is inhibited by targeting one-carbon metabolism through methotrexate or methionine deprivation. *Proceedings of the National Academy of Sciences*, *116*(8), 2987–2995. https://doi.org/10.1073/pnas.1820161116

Albrecht, L. V., Ploper, D., Tejeda-Muñoz, N., & De Robertis, E. M. (2018). Arginine methylation is required for canonical Wnt signaling and endolysosomal trafficking. *Proceedings of the National Academy of Sciences*, *115*(23), E5317–E5325. https://doi.org/10.1073/pnas.1804091115

Al-Dosari, M. S., & Gao, X. (2009). Nonviral Gene Delivery: Principle, Limitations, and Recent Progress. *The AAPS Journal*, *11*(4), 671. https://doi.org/10.1208/s12248-009-9143-y

Alonzo, J. R., Venkataraman, C., Field, M. S., & Stover, P. J. (2018). The mitochondrial inner membrane protein MPV17 prevents uracil accumulation in mitochondrial DNA. *The Journal of Biological Chemistry*, *293*(52), 20285–20294. https://doi.org/10.1074/jbc.RA118.004788

Amenyah, S. D., McMahon, A., Ward, M., Deane, J., McNulty, H., Hughes, C. F., Strain, J. J., Horigan, G., Purvis, J., Walsh, C. P., & Lees-Murdock, D. J. (2020). Riboflavin supplementation alters global and gene-specific DNA methylation in adults with

the MTHFR 677 TT genotype. *Biochimie*, *173*, 17–26. https://doi.org/10.1016/j.biochi.2020.04.007

Amenyah, S. D., Ward, M., McMahon, A., Deane, J., McNulty, H., Hughes, C., Strain, J. J., Horigan, G., Purvis, J., Walsh, C. P., & Lees-Murdock, D. J. (2021). DNA methylation of hypertension-related genes and effect of riboflavin supplementation in adults stratified by genotype for the MTHFR C677T polymorphism. *International Journal of Cardiology*, *322*, 233–239. https://doi.org/10.1016/j.ijcard.2020.09.011

An, S., Kumar, R., Sheets, E. D., & Benkovic, S. J. (2008). Reversible compartmentalization of de novo purine biosynthetic complexes in living cells. *Science (New York, N.Y.)*, *320*(5872), 103–106. https://doi.org/10.1126/science.1152241

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. https://doi.org/10.1073/pnas.81.16.5170

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 of the United States of America*, *108*(37), 15163–15168. https://doi.org/10.1073/pnas.1103623108

Anderson, D. D., & Stover, P. J. (2009). SHMT1 and SHMT2 Are Functionally Redundant in Nuclear De novo Thymidylate Biosynthesis. *PLoS ONE*, *4*(6). https://doi.org/10.1371/journal.pone.0005839

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 Chemistry and Laboratory Medicine*, *45*(12), 1760–1763. https://doi.org/10.1515/CCLM.2007.355

Appling, D. R. (1991). Compartmentation of folate-mediated one-carbon metabolism in eukaryotes. *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, *5*(12), 2645–2651. https://doi.org/10.1096/fasebj.5.12.1916088

Ash, J., Liao, W. C., Ke, Y., & Johnson, L. F. (1995). Regulation of mouse thymidylate synthase gene expression in growth-stimulated cells: Upstream S phase control elements are indistinguishable from the essential promoter elements. *Nucleic Acids Research*, *23*(22), 4649–4656.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

Athanasiadis, A., Rich, A., & Maas, S. (2004). Widespread A-to-I RNA Editing of Alu-Containing mRNAs in the Human Transcriptome. *PLOS Biology*, *2*(12), e391. https://doi.org/10.1371/journal.pbio.0020391

Ayoubi, T. A. Y., & Van De Yen, W. J. M. (1996). Regulation of gene expression by alternative promoters. *The FASEB Journal*, *10*(4), 453–460. https://doi.org/10.1096/fasebj.10.4.8647344

Ayusawa, D., Shimizu, K., Koyama, H., Kaneda, S., Takeishi, K., & Seno, T. (1986). Cell-cycle-directed regulation of thymidylate synthase messenger RNA in human diploid fibroblasts stimulated to proliferate. *Journal of Molecular Biology*, *190*(4), 559–567. https://doi.org/10.1016/0022-2836(86)90241-X

Baggott, J. E., & Tamura, T. (2015). Folate-Dependent Purine Nucleotide Biosynthesis in Humans. *Advances in Nutrition (Bethesda, Md.)*, *6*(5), 564–571. https://doi.org/10.3945/an.115.008300

Bahn, J. H., Lee, J.-H., Li, G., Greer, C., Peng, G., & Xiao, X. (2012). Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Research*, *22*(1), 142–150. https://doi.org/10.1101/gr.124107.111

Bailey, L. B., Barton, L. E., Hillier, S. E., & Cerda, J. J. (1988). Bioavailability of mono and polyglutamyl folate in human subjects. *Nutrition Reports International (USA)*. https://scholar.google.com/scholar_lookup?title=Bioavailability+of+mono+and+polyglutamyl+folate+in+human+subjects&author=Bailey%2C+L.B.&publication_year=1988

Bailey, L. B., & Caudill, M. A. (2012). Folate. In *Present Knowledge in Nutrition* (pp. 321–342). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119946045.ch21

Bailey, L. B., Cerda, J. J., Bloch, B. S., Busby, M. J., Vargas, L., Chandler, C. J., & Halsted, C. H. (1984). Effect of age on poly- and monoglutamyl folacin absorption in human subjects. *The Journal of Nutrition*, *114*(10), 1770–1776. https://doi.org/10.1093/jn/114.10.1770

Bailey, L. B., & Gregory, J. F., III. (1999). Folate Metabolism and Requirements. *The Journal of Nutrition*, *129*(4), 779–782. https://doi.org/10.1093/jn/129.4.779

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. (2015).

Biomarkers of Nutrition for Development-Folate Review. *The Journal of Nutrition*, *145*(7), 1636S-1680S. https://doi.org/10.3945/jn.114.206599

Bailey, S. W., & Ayling, J. E. (2009). The extremely slow and variable activity of dihydrofolate reductase in human liver and its implications for high folic acid intake. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(36), 15424–15429. https://doi.org/10.1073/pnas.0902072106

Baldwin, T., Sakthianandeswaren, A., Curtis, J. M., Kumar, B., Smyth, G. K., Foote, S. J., & Handman, E. (2007). Wound healing response is a major contributor to the severity of cutaneous leishmaniasis in the ear model of infection. *Parasite Immunology*, *29*(10), 501–513. https://doi.org/10.1111/j.1365-3024.2007.00969.x

Bao, C., Loerch, S., Ling, C., Korostelev, A. A., Grigorieff, N., & Ermolenko, D. N. (2020). MRNA stem-loops can pause the ribosome by hindering A-site tRNA binding. *ELife*, *9*, e55799. https://doi.org/10.7554/eLife.55799

Bao, X. R., Ong, S.-E., Goldberger, O., Peng, J., Sharma, R., Thompson, D. A., Vafai, S. B., Cox, A. G., Marutani, E., Ichinose, F., Goessling, W., Regev, A., Carr, S. A., Clish, C. B., & Mootha, V. K. (2016). Mitochondrial dysfunction remodels one-carbon metabolism in human cells. *ELife*, *5*, e10575. https://doi.org/10.7554/eLife.10575

Barrangou, R. (2015). The roles of CRISPR–Cas systems in adaptive immunity and beyond. *Current Opinion in Immunology*, *32*, 36–41. https://doi.org/10.1016/j.coi.2014.12.008

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., & Horvath, P. (2007). CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science*, *315*(5819), 1709–1712. https://doi.org/10.1126/science.1138140

Bartha, I., di Iulio, J., Venter, J. C., & Telenti, A. (2018). Human gene essentiality. *Nature Reviews. Genetics*, *19*(1), 51–62. https://doi.org/10.1038/nrg.2017.75

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*. https://doi.org/10.3791/52118

Bazak, L., Haviv, A., Barak, M., Jacob-Hirsch, J., Deng, P., Zhang, R., Isaacs, F. J., Rechavi, G., Li, J. B., Eisenberg, E., & Levanon, E. Y. (2014). A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Research*, *24*(3), 365–376. https://doi.org/10.1101/gr.164749.113

Beck, C. F., & Warren, R. A. (1988). Divergent promoters, a common form of gene organization. *Microbiological Reviews*, *52*(3), 318–326.

Bennett, E. P., Petersen, B. L., Johansen, I. E., Niu, Y., Yang, Z., Chamberlain, C. A., Met, Ö., Wandall, H. H., & Frödin, M. (2020). INDEL detection, the 'Achilles heel' of precise genome editing: A survey of methods for accurate profiling of gene editing induced indels. *Nucleic Acids Research*, *48*(21), 11958–11981. https://doi.org/10.1093/nar/gkaa975

Benson, D. A., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2015). GenBank. *Nucleic Acids Research*, *43*(Database issue), D30–D35. https://doi.org/10.1093/nar/gku1216

Bertino, J. R., Donohue, D. M., Simmons, B., Gabrio, B. W., Silber, R., & Huennekens, F. M. (1963). THE "INDUCTION" OF DIHYDROFOLIC REDUCTASE ACTIVITY IN LEUKOCYTES AND ERYTHROCYTES OF PATIENTS TREATED WITH AMETHOPTERIN*. *Journal of Clinical Investigation*, *42*(4), 466–475.

Bertino, J. R., Donohue, D. R., Gabrio, B. W., Silber, R., Alenty, A., Meyer, M., & Huennekens, F. M. (1962). Increased Level of Dihydrofolic Reductase in Leucocytes of Patients Treated with Amethopterin. *Nature*, *193*(4811), 140–142. https://doi.org/10.1038/193140a0

Blakley, R. L. (2006, November 22). *Eukaryotic Dihydrofolate Reductase*. 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

Blomen, V. A., Májek, P., Jae, L. T., Bigenzahn, J. W., Nieuwenhuis, J., Staring, J., Sacco, R., van Diemen, F. R., Olk, N., Stukalov, A., Marceau, C., Janssen, H., Carette, J. E., Bennett, K. L., Colinge, J., Superti-Furga, G., & Brummelkamp, T. R. (2015). Gene essentiality and synthetic lethality in haploid human cells. *Science (New York, N.Y.)*, *350*(6264), 1092–1096. https://doi.org/10.1126/science.aac7557

Blount, B. C., Mack, M. M., Wehr, C. M., MacGregor, J. T., Hiatt, R. A., Wang, G., Wickramasinghe, S. N., Everson, R. B., & Ames, B. N. (1997). Folate deficiency causes uracil misincorporation into human DNA and chromosome breakage: Implications for cancer and neuronal damage. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(7), 3290–3295. https://doi.org/10.1073/pnas.94.7.3290

Bo, Y., Zhu, Y., Tao, Y., Li, X., Zhai, D., Bu, Y., Wan, Z., Wang, L., Wang, Y., & Yu, Z. (2020). Association Between Folate and Health Outcomes: An Umbrella Review of Meta-Analyses. *Frontiers in Public Health*, *8*, 550753. https://doi.org/10.3389/fpubh.2020.550753

Bolte, S., & Cordelières, F. P. (2006). A guided tour into subcellular colocalization analysis in light microscopy. *Journal of Microscopy*, *224*(3), 213–232. https://doi.org/10.1111/j.1365-2818.2006.01706.x

Bond, C. S., & Fox, A. H. (2009). Paraspeckles: Nuclear bodies built on long noncoding RNA. *The Journal of Cell Biology*, *186*(5), 637–644. https://doi.org/10.1083/jcb.200906113

Boo, S. H., & Kim, Y. K. (2020). The emerging role of RNA modifications in the regulation of mRNA stability. *Experimental & Molecular Medicine*, *52*(3), 400–408. https://doi.org/10.1038/s12276-020-0407-z

Bookey, N., Drago, P., & Parle-McDermott, A. (unpublished).

Bravo, F. E., Parra, N. C., Camacho, F., Acosta, J., González, A., Toledo, J. R., & Sanchez, O. (2020). Fluorescence-assisted sequential insertion of transgenes (FASIT): An approach for increasing specific productivity in mammalian cells. *Scientific Reports*, *10*(1), 12840. https://doi.org/10.1038/s41598-020-69709-1

Brooks, H. L., & Lindsey, M. L. (2018). Guidelines for authors and reviewers on antibody use in physiology studies. *American Journal of Physiology-Heart and Circulatory Physiology*, *314*(4), H724–H732. https://doi.org/10.1152/ajpheart.00512.2017

Brosnan, J. T., & Brosnan, M. E. (2020). Formate and its role in amino acid metabolism. *Current Opinion in Clinical Nutrition & Metabolic Care*, *23*(1), 23–28. https://doi.org/10.1097/MCO.0000000000000611

Brown, T. A. (2002). Understanding a Genome Sequence. In *Genomes. 2nd edition*. Wiley-Liss. https://www.ncbi.nlm.nih.gov/books/NBK21136/

Burger, A., Lindsay, H., Felker, A., Hess, C., Anders, C., Chiavacci, E., Zaugg, J., Weber, L. M., Catena, R., Jinek, M., Robinson, M. D., & Mosimann, C. (2016). Maximizing mutagenesis with solubilized CRISPR-Cas9 ribonucleoprotein complexes. *Development (Cambridge, England)*, *143*(11), 2025–2037. https://doi.org/10.1242/dev.134809

Burren, K. A., Mills, K., Copp, A. J., & Greene, N. D. E. (2006). Quantitative analysis of s-adenosylmethionine and s-adenosylhomocysteine in neurulation-stage mouse embryos by liquid chromatography tandem mass spectrometry. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences*, *844*(1), 112–118. https://doi.org/10.1016/j.jchromb.2006.07.012

Caliskan, N., Wohlgemuth, I., Korniy, N., Pearson, M., Peske, F., & Rodnina, M. V. (2017). Conditional Switch between Frameshifting Regimes upon Translation of dnaX mRNA. *Molecular Cell*, *66*(4), 558-567.e4. https://doi.org/10.1016/j.molcel.2017.04.023

Camarillo, C., Swerdel, M., & Hart, R. P. (2011). Comparison of microarray and quantitative real-time PCR methods for measuring microRNA levels in MSC cultures. *Methods in Molecular Biology (Clifton, N.J.)*, *698*, 419–429. https://doi.org/10.1007/978-1-60761-999-4_30

Capecchi, M. R. (2005). Gene targeting in mice: Functional analysis of the mammalian genome for the twenty-first century. *Nature Reviews. Genetics*, *6*(6), 507–512. https://doi.org/10.1038/nrg1619

Carter, M., & Shieh, J. C. (2010). Chapter 13—Cell Culture Techniques. In M. Carter & J. C. Shieh (Eds.), *Guide to Research Techniques in Neuroscience* (pp. 281–296). Academic Press. https://doi.org/10.1016/B978-0-12-374849-2.00013-6

Carugo, O., & Argos, P. (1997, May 1). *NADP-Dependent enzymes. II: Evolution of the mono- and dinucleotide binding domains*. Proteins: Structure, Function, and Bioinformatics. https://doi.org/10.1002/(SICI)1097-0134(199705)28:1<29::AID-PROT3>3.0.CO;2-E

Caudill, M. A. (2010). Folate bioavailability: Implications for establishing dietary recommendations and optimizing status1234. *The American Journal of Clinical Nutrition*, *91*(5), 1455S-1460S. https://doi.org/10.3945/ajcn.2010.28674E

Chandrasekaran, A. P., Song, M., Kim, K.-S., & Ramakrishna, S. (2018). Different Methods of Delivering CRISPR/Cas9 Into Cells. *Progress in Molecular Biology and Translational Science*, *159*, 157–176. https://doi.org/10.1016/bs.pmbts.2018.05.001

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

Chasin, L. A., Feldman, A., Konstam, M., & Urlaub, G. (1974). Reversion of a Chinese Hamster Cell Auxotrophic Mutant. *Proceedings of the National Academy of Sciences of the United States of America*, *71*(3), 718–722.

Chełkowska-Pauszek, A., Kosiński, J. G., Marciniak, K., Wysocka, M., Bąkowska-Żywicka, K., & Żywicki, M. (2021). The Role of RNA Secondary Structure in Regulation of Gene Expression in Bacteria. *International Journal of Molecular Sciences*, *22*(15), 7845. https://doi.org/10.3390/ijms22157845

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., DeCerbo, J. N., & Carmichael, G. G. (2008). 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.

Chen, P.-Y., Chang, W.-S. W., Chou, R.-H., Lai, Y.-K., Lin, S.-C., Chi, C.-Y., & Wu, C.-W. (2007). Two non-homologous brain diseases-related genes, SERPINI1 and PDCD10, are tightly linked by an asymmetric bidirectional promoter in an evolutionarily conserved manner. *BMC Molecular Biology*, *8*(1), 2. https://doi.org/10.1186/1471-2199-8-2

Chial, H., & Craig, J. (2008). *MtDNA and Mitochondrial Diseases | Learn Science at Scitable*. http://www.nature.com/scitable/topicpage/mtdna-and-mitochondrial-diseases-903

CHO/dhFr- 94060607 | Sigma-Aldrich. (n.d.). Retrieved 2 May 2022, from http://www.sigmaaldrich.com/

Choi, H. M. T., Schwarzkopf, M., Fornace, M. E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., & Pierce, N. A. (2018). Third-generation in situ hybridization chain reaction: Multiplexed, quantitative, sensitive, versatile, robust. *Development*, *145*(12), dev165753. https://doi.org/10.1242/dev.165753

Chon, J., Stover, P. J., & Field, M. S. (2017). Targeting nuclear thymidylate biosynthesis. *Molecular Aspects of Medicine*, *53*, 48–56. https://doi.org/10.1016/j.mam.2016.11.005

Christensen, K. E., & MacKenzie, R. E. (2006). Mitochondrial one-carbon metabolism is adapted to the specific needs of yeast, plants and mammals. *BioEssays*, *28*(6), 595–605. https://doi.org/10.1002/bies.20420

Christofi, T., & Zaravinos, A. (2019). RNA editing in the forefront of epitranscriptomics and human health. *Journal of Translational Medicine*, *17*(1), 319. https://doi.org/10.1186/s12967-019-2071-4

Clare, C. E., Brassington, A. H., Kwong, W. Y., & Sinclair, K. D. (2019). One-Carbon Metabolism: Linking Nutritional Biochemistry to Epigenetic Programming of Long-Term Development. *Annual Review of Animal Biosciences*, *7*, 263–287. https://doi.org/10.1146/annurev-animal-020518-115206

Clifford, A. J., Heid, M. K., Müller, H. G., & Bills, N. D. (1990). Tissue distribution and prediction of total body folate of rats. *The Journal of Nutrition*, *120*(12), 1633–1639. https://doi.org/10.1093/jn/120.12.1633

Cochrane, K. M., Mayer, C., Devlin, A. M., Elango, R., Hutcheon, J. A., & Karakochuk, C. D. (2020). Is natural (6S)-5-methyltetrahydrofolic acid as effective as synthetic folic acid in increasing serum and red blood cell folate concentrations during pregnancy? A proof-of-concept pilot study. *Trials*, *21*(1), 380. https://doi.org/10.1186/s13063-020-04320-3

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

Cooper, G. M. (2000). DNA Repair. *The Cell: A Molecular Approach. 2nd Edition*. https://www.ncbi.nlm.nih.gov/books/NBK9900/

Cortese, C., & Motti, C. (2001). MTHFR gene polymorphism, homocysteine and cardiovascular disease. *Public Health Nutrition*, *4*(2B), 493–497. https://doi.org/10.1079/phn2001159

Cowan, K. H., Goldsmith, M. E., Ricciardone, M. D., Levine, R., Rubalcaba, E., & Jolivet, J. (1986). Regulation of dihydrofolate reductase in human breast cancer cells and in mutant hamster cells transfected with a human dihydrofolate reductase minigene. *Molecular Pharmacology*, *30*(1), 69–76.

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

Crider, K. S., Bailey, L. B., & Berry, R. J. (2011). Folic Acid Food Fortification—Its History, Effect, Concerns, and Future Directions. *Nutrients*, *3*(3), 370–384. https://doi.org/10.3390/nu3030370

Crider, K. S., Yang, T. P., Berry, R. J., & Bailey, L. B. (2012). Folate and DNA Methylation: A Review of Molecular Mechanisms and the Evidence for Folate's Role. *Advances in Nutrition*, *3*(1), 21–38. https://doi.org/10.3945/an.111.000992

CRISPR-Cas9 guide RNA design checker | IDT. (n.d.). Retrieved 18 July 2019, from https://eu.idtdna.com/site/order/designtool/index/CRISPR_SEQUENCE

Cubbon, A., Ivancic-Bace, I., & Bolt, E. L. (2018). CRISPR-Cas immunity, DNA repair and genome stability. *Bioscience Reports*, *38*(5), BSR20180457. https://doi.org/10.1042/BSR20180457

Cunningham, F., Allen, J. E., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Austine-Orimoloye, O., Azov, A. G., Barnes, I., Bennett, R., Berry, A., Bhai, J., Bignell, A., Billis, K., Boddu, S., Brooks, L., Charkhchi, M., Cummins, C., Da Rin Fioretto, L., … Flicek, P. (2022). Ensembl 2022. *Nucleic Acids Research*, *50*(D1), D988–D995. https://doi.org/10.1093/nar/gkab1049

Czeizel, A. E., & Dudás, I. (1992). Prevention of the First Occurrence of Neural-Tube Defects by Periconceptional Vitamin Supplementation. *New England Journal of Medicine*, *327*(26), 1832–1835. https://doi.org/10.1056/NEJM199212243272602

Daneshtalab, N., Doré, J. J. E., & Smeda, J. S. (2010). Troubleshooting tissue specificity and antibody selection: Procedures in immunohistochemical studies. *Journal of Pharmacological and Toxicological Methods*, *61*(2), 127–135. https://doi.org/10.1016/j.vascn.2009.12.002

Davies, J. F., Delcamp, T. J., Prendergast, N. J., Ashford, V. A., Freisheim, J. H., & Kraut, J. (1990). Crystal structures of recombinant human dihydrofolate reductase complexed with folate and 5-deazafolate. *Biochemistry*, *29*(40), 9467–9479. https://doi.org/10.1021/bi00492a021

DeCerbo, J., & Carmichael, G. G. (2005). SINEs point to abundant editing in the human genome. *Genome Biology*, *6*(4), 216. https://doi.org/10.1186/gb-2005-6-4-216

Deininger, P. L., & Batzer, M. A. (2002). Mammalian retroelements. *Genome Research*, *12*(10), 1455–1465. https://doi.org/10.1101/gr.282402

DESKGEN Cloud | Guide Picker. (n.d.). Retrieved 2 July 2019, from https://www.deskgen.com/guide-picker/#/plot/GRCh38.81/DHFR

Dickerman, H. W., Steers, E., Redfield, B. G., & Weissbach, H. (1967). Methionyl Soluble Ribonucleic Acid Transformylase: I. PURIFICATION AND PARTIAL CHARACTERIZATION. *Journal of Biological Chemistry*, *242*(7), 1522–1525. https://doi.org/10.1016/S0021-9258(18)96123-5

Donato, M. T., Tolosa, L., & Gómez-Lechón, M. J. (2015). Culture and Functional Characterization of Human Hepatoma HepG2 Cells. *Methods in Molecular Biology (Clifton, N.J.)*, *1250*, 77–93. https://doi.org/10.1007/978-1-4939-2074-7_5

Du, L.-L. (2020). Resurrection from lethal knockouts: Bypass of gene essentiality. *Biochemical and Biophysical Research Communications*, *528*(3), 405–412. https://doi.org/10.1016/j.bbrc.2020.05.207

Ducker, G. S., Chen, L., Morscher, R. J., Ghergurovich, J. M., Esposito, M., Teng, X., Kang, Y., & Rabinowitz, J. D. (2016). Reversal of Cytosolic One-Carbon Flux Compensates for Loss of the Mitochondrial Folate Pathway. *Cell Metabolism*, *23*(6), 1140–1153. https://doi.org/10.1016/j.cmet.2016.04.016

Ducker, G. S., & Rabinowitz, J. D. (2017). One-Carbon Metabolism in Health and Disease. *Cell Metabolism*, *25*(1), 27–42. https://doi.org/10.1016/j.cmet.2016.08.009

Dunn, K. W., Kamocka, M. M., & McDonald, J. H. (2011). A practical guide to evaluating colocalization in biological microscopy. *American Journal of Physiology - Cell Physiology*, *300*(4), C723–C742. https://doi.org/10.1152/ajpcell.00462.2010

Ebara, S. (2017). Nutritional role of folate. *Congenital Anomalies*, *57*(5), 138–141. https://doi.org/10.1111/cga.12233

Eisenberg, E., & Levanon, E. Y. (2018). A-to-I RNA editing—Immune protector and transcriptome diversifier. *Nature Reviews Genetics*, *19*(8), 473–490. https://doi.org/10.1038/s41576-018-0006-1

Elso, C. M., Roberts, L. J., Smyth, G. K., Thomson, R. J., Baldwin, T. M., Foote, S. J., & Handman, E. (2004). Leishmaniasis host response loci (lmr1-3) modify disease

severity through a Th1/Th2-independent pathway. *Genes and Immunity*, *5*(2), 93–100. https://doi.org/10.1038/sj.gene.6364042

Ercikan, E., Banerjee, D., Waltham, M., Schnieders, B., Scotto, K. W., & Bertino, J. R. (1993). Translational Regulation of the Synthesis of Dihydrofolate Reductase. In J. E. Ayling, M. G. Nair, & C. M. Baugh (Eds.), *Chemistry and Biology of Pteridines and Folates* (pp. 537–540). Springer US. 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

Ezaki, R., Ichikawa, K., Matsuzaki, M., & Horiuchi, H. (2022). Targeted Knock-in of a Fluorescent Protein Gene into the Chicken Vasa Homolog Locus of Chicken Primordial Germ Cells using CRIS-PITCh Method. *The Journal of Poultry Science*, *59*(2), 182–190. https://doi.org/10.2141/jpsa.0210067

Featherstone, C., & Jackson, S. P. (1999). DNA double-strand break repair. *Current Biology*, *9*(20), R759–R761. https://doi.org/10.1016/S0960-9822(00)80005-6

Fei, Z., Gao, Y., Qiu, M., Qi, X., Dai, Y., Wang, S., Quan, Z., Liu, Y., & Ou, J. (2016). Down-regulation of dihydrofolate reductase inhibits the growth of endothelial EA.hy926 cell through induction of G1 cell cycle arrest via up-regulating p53 and p21waf1/cip1 expression. *Journal of Clinical Biochemistry and Nutrition*, *58*(2), 105–113. https://doi.org/10.3164/jcbn.15-64

Fernández, A., Josa, S., & Montoliu, L. (2017). A history of genome editing in mammals. *Mammalian Genome*, *28*(7), 237–246. https://doi.org/10.1007/s00335-017-9699-2

Fiddler, J. L., Blum, J. E., Castillo, L. F., Thalacker-Mercer, A. E., & Field, M. S. (2022). *Impairments in SHMT2 expression or cellular folate availability reduce oxidative phosphorylation and pyruvate kinase activity* (p. 2022.04.07.487511). bioRxiv. https://doi.org/10.1101/2022.04.07.487511

Fiddler, J. L., Blum, J., Thalacker-Mercer, A., & Field, M. (2021). Loss of SHMT2 and Folate Deficiency Impair Energy Metabolism in Mouse Embryonic Fibroblasts Cells. *Current Developments in Nutrition*, *5*(Suppl 2), 1311. https://doi.org/10.1093/cdn/nzab059_012

Fiddler, J. L., Xiu, Y., Blum, J. E., Lamarre, S. G., Phinney, W. N., Stabler, S. P., Brosnan, M. E., Brosnan, J. T., Thalacker-Mercer, A. E., & Field, M. S. (2021a). *Reduced Shmt2 expression impairs mitochondrial folate accumulation and respiration, and leads to uracil accumulation in mouse mitochondrial DNA* (p. 2021.04.12.439270). bioRxiv. https://doi.org/10.1101/2021.04.12.439270

Fiddler, J. L., Xiu, Y., Blum, J. E., Lamarre, S. G., Phinney, W. N., Stabler, S. P., Brosnan, M. E., Brosnan, J. T., Thalacker-Mercer, A. E., & Field, M. S. (2021b). Reduced Shmt2 Expression Impairs Mitochondrial Folate Accumulation and Respiration, and Leads to Uracil Accumulation in Mouse Mitochondrial DNA. *The Journal of Nutrition*, *151*(10), 2882–2893. https://doi.org/10.1093/jn/nxab211

Field, M. S., Anderson, D. D., & Stover, P. J. (2011). Mthfs is an Essential Gene in Mice and a Component of the Purinosome. *Frontiers in Genetics*, *2*, 36. https://doi.org/10.3389/fgene.2011.00036

Field, M. S., Anguera, M. C., Page, R., & Stover, P. J. (2009). 5,10-Methenyltetrahydrofolate Synthetase Activity is Increased in Tumors and Modifies the Efficacy of Antipurine LY309887. *Archives of Biochemistry and Biophysics*, *481*(2), 145–150. https://doi.org/10.1016/j.abb.2008.11.001

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., Chon, J., & Stover, P. J. (2018). Nuclear Folate Metabolism. *Annual Review of Nutrition*, *38*, 219–243. https://doi.org/10.1146/annurev-nutr-071714-034441

Field, M. S., Kamynina, E., Watkins, D., Rosenblatt, D. S., & Stover, P. J. (2015). Human mutations in methylenetetrahydrofolate dehydrogenase 1 impair nuclear de novo thymidylate biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(2), 400–405. https://doi.org/10.1073/pnas.1414555112

Field, M. S., & Stover, P. J. (2018). Safety of folic acid. *Annals of the New York Academy of Sciences*, *1414*(1), 59–71. https://doi.org/10.1111/nyas.13499

Field, M. S., Stover, P. J., & Kisliuk, R. (2016a). Thymidylate Synthesis. In *ELS* (pp. 1–7). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470015902.a0001397.pub3

Field, M. S., Stover, P. J., & Kisliuk, R. (2016b). Thymidylate Synthesis. In *ELS* (pp. 1–7). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470015902.a0001397.pub3

Field, M. S., Szebenyi, D. M. E., & Stover, P. J. (2006). Regulation of de novo purine biosynthesis by methenyltetrahydrofolate synthetase in neuroblastoma. *The Journal of Biological Chemistry*, *281*(7), 4215–4221. https://doi.org/10.1074/jbc.M510624200

Finglas, P. M., & Wright, A. J. A. (2002). Folate bioavailability and health. *Phytochemistry Reviews*, *1*(2), 189–198. https://doi.org/10.1023/A:1022559417212

Fischer, M., Stronati, M., & Lanari, M. (2017). Mediterranean diet, folic acid, and neural tube defects. *Italian Journal of Pediatrics*, *43*(1), 74. https://doi.org/10.1186/s13052-017-0391-7

Fischer, M., Thöny, B., & Leimkühler, S. (2010). 7.17—The Biosynthesis of Folate and Pterins and Their Enzymology. In H.-W. (Ben) Liu & L. Mander (Eds.), *Comprehensive Natural Products II* (pp. 599–648). Elsevier. https://doi.org/10.1016/B978-008045382-8.00150-7

Fodale, V., Pierobon, M., Liotta, L., & Petricoin, E. (2011). Mechanism of Cell Adaptation. *Cancer Journal (Sudbury, Mass.)*, *17*(2), 89–95. https://doi.org/10.1097/PPO.0b013e318212dd3d

Foster, A. J., Martin-Urdiroz, M., Yan, X., Wright, H. S., Soanes, D. M., & Talbot, N. J. (2018). CRISPR-Cas9 ribonucleoprotein-mediated co-editing and counterselection in the rice blast fungus. *Scientific Reports*, *8*, 14355. https://doi.org/10.1038/s41598-018-32702-w

Fowler, B. (2001). The folate cycle and disease in humans. *Kidney International. Supplement*, *78*, S221-229. https://doi.org/10.1046/j.1523-1755.2001.59780221.x

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., & Stover, P. J. (2008). Folate-mediated one-carbon metabolism. *Vitamins and Hormones*, *79*, 1–44. https://doi.org/10.1016/S0083-6729(08)00401-9

Francis, K., & Kohen, A. (2014). Protein Motions and the Activation of the C-H Bond Catalyzed by Dihydrofolate Reductase. *Current Opinion in Chemical Biology*, *0*, 19–24. https://doi.org/10.1016/j.cbpa.2014.03.009

Francis, K., Stojkovic, V., & Kohen, A. (2013). Preservation of protein dynamics in dihydrofolate reductase evolution. *The Journal of Biological Chemistry*, *288*(50), 35961–35968. https://doi.org/10.1074/jbc.M113.507632

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

Gagliardi, M., & Matarazzo, M. R. (2016). RIP: RNA Immunoprecipitation. *Methods in Molecular Biology (Clifton, N.J.)*, *1480*, 73–86. https://doi.org/10.1007/978-1-4939-6380-5_7

Gagnon, J. A., Valen, E., Thyme, S. B., Huang, P., Akhmetova, L., Ahkmetova, L., Pauli, A., Montague, T. G., Zimmerman, S., Richter, C., & Schier, A. F. (2014). Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale

assessment of single-guide RNAs. *PloS One*, *9*(5), e98186. https://doi.org/10.1371/journal.pone.0098186

Gaj, T., Gersbach, C. A., & Barbas, C. F. (2013). ZFN, TALEN and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology*, *31*(7), 397–405. https://doi.org/10.1016/j.tibtech.2013.04.004

Gallagher, D. N., & Haber, J. E. (2018). Repair of a site-specific DNA cleavage: Old-school lessons for Cas9-mediated gene editing. *ACS Chemical Biology*, *13*(2), 397–405. https://doi.org/10.1021/acschembio.7b00760

Gao, J., Cui, W., Du, Y., & Ji, M. (2013). Insight into the molecular mechanism about lowered dihydrofolate binding affinity to dihydrofolate reductase-like 1 (DHFRL1). *Journal of Molecular Modeling*, *19*(12), 5187–5198. https://doi.org/10.1007/s00894-013-2018-2

Garcia-Martinez, L. F., & Appling, D. R. (1993). Characterization of the folate-dependent mitochondrial oxidation of carbon 3 of serine. *Biochemistry*, *32*(17), 4671–4676. https://doi.org/10.1021/bi00068a027

García-Montero, C., Fraile-Martínez, O., Gómez-Lahoz, A. M., Pekarek, L., Castellanos, A. J., Noguerales-Fraguas, F., Coca, S., Guijarro, L. G., García-Honduvilla, N., Asúnsolo, A., Sanchez-Trujillo, L., Lahera, G., Bujan, J., Monserrat, J., Álvarez-Mon, M., Álvarez-Mon, M. A., & Ortega, M. A. (2021). Nutritional Components in Western Diet Versus Mediterranean Diet at the Gut Microbiota–Immune System Interplay. Implications for Health and Disease. *Nutrients*, *13*(2), 699. https://doi.org/10.3390/nu13020699

Gaughan, D. J., Barbaux, S., Kluijtmans, L. A., & Whitehead, A. S. (2000). The human and mouse methylenetetrahydrofolate reductase (MTHFR) genes: Genomic organization, mRNA structure and linkage to the CLCN6 gene. *Gene*, *257*(2), 279–289. https://doi.org/10.1016/s0378-1119(00)00392-9

Giacinti, C., & Giordano, A. (2006). RB and cell cycle progression. *Oncogene*, *25*(38), 5220–5227. https://doi.org/10.1038/sj.onc.1209615

Gong, J., Shao, D., Xu, K., Lu, Z., Lu, Z. J., Yang, Y. T., & Zhang, Q. C. (2018). RISE: A database of RNA interactome from sequencing experiments. *Nucleic Acids Research*, *46*(D1), D194–D201. https://doi.org/10.1093/nar/gkx864

Good, L., Dimri, G. P., Campisi, J., & Chen, K. Y. (1996). Regulation of dihydrofolate reductase gene expression and E2F components in human diploid fibroblasts during growth and senescence. *Journal of Cellular Physiology*, *168*(3), 580–588. https://doi.org/10.1002/(SICI)1097-4652(199609)168:3<580::AID-JCP10>3.0.CO;2-3

Greasley, S. E., Horton, P., Ramcharan, J., Beardsley, G. P., Benkovic, S. J., & Wilson, I. A. (2001). Crystal structure of a bifunctional transformylase and cyclohydrolase

enzyme in purine biosynthesis. *Nature Structural Biology*, *8*(5), 402–406. https://doi.org/10.1038/87555

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

Gregory, J. F., Bhandari, S. D., Bailey, L. B., Toth, J. P., Baumgartner, T. G., & Cerda, J. J. (1991). Relative bioavailability of deuterium-labeled monoglutamyl and hexaglutamyl folates in human subjects. *The American Journal of Clinical Nutrition*, *53*(3), 736–740. https://doi.org/10.1093/ajcn/53.3.736

Gross, A., Schoendube, J., Zimmermann, S., Steeb, M., Zengerle, R., & Koltay, P. (2015). Technologies for Single-Cell Isolation. *International Journal of Molecular Sciences*, *16*(8), 16897–16919. https://doi.org/10.3390/ijms160816897

Gu, T., Buaas, F. W., Simons, A. K., Ackert-Bicknell, C. L., Braun, R. E., & Hibbs, M. A. (2012). Canonical A-to-I and C-to-U RNA Editing Is Enriched at 3'UTRs and microRNA Target Sites in Multiple Mouse Tissues. *PLOS ONE*, *7*(3), e33720. https://doi.org/10.1371/journal.pone.0033720

Guiducci, G., Paone, A., Tramonti, A., Giardina, G., Rinaldo, S., Bouzidi, A., Magnifico, M. C., Marani, M., Menendez, J. A., Fatica, A., Macone, A., Armaos, A., Tartaglia, G. G., Contestabile, R., Paiardini, A., & Cutruzzolà, F. (2019). The moonlighting RNA-binding activity of cytosolic serine hydroxymethyltransferase contributes to control compartmentalization of serine metabolism. *Nucleic Acids Research*, *47*(8), 4240–4254. https://doi.org/10.1093/nar/gkz129

Gupta, R. M., & Musunuru, K. (2014). Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *The Journal of Clinical Investigation*, *124*(10), 4154–4161. https://doi.org/10.1172/JCI72992

Gurumayum, S., Jiang, P., Hao, X., Campos, T. L., Young, N. D., Korhonen, P. K., Gasser, R. B., Bork, P., Zhao, X.-M., He, L., & Chen, W.-H. (2021). OGEE v3: Online GEne Essentiality database with increased coverage of organisms and human cell lines. *Nucleic Acids Research*, *49*(D1), D998–D1003. https://doi.org/10.1093/nar/gkaa884

Haber, J. E. (2000). Partners and pathwaysrepairing a double-strand break. *Trends in Genetics: TIG*, *16*(6), 259–264. https://doi.org/10.1016/s0168-9525(00)02022-9

Haimovich, A. D., Muir, P., & Isaacs, F. J. (2015). Genomes by design. *Nature Reviews Genetics*, *16*(9), 501–516. https://doi.org/10.1038/nrg3956

Hall, B., Limaye, A., & Kulkarni, A. B. (2009). Overview: Generation of Gene Knockout Mice. *Current Protocols in Cell Biology*, *44*(1), 19.12.1-19.12.17. https://doi.org/10.1002/0471143030.cb1912s44

Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K. R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., Mero, P., Dirks, P., Sidhu, S., Roth, F. P., Rissland, O. S., Durocher, D., Angers, S., & Moffat, J. (2015). High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell*, *163*(6), 1515–1526. https://doi.org/10.1016/j.cell.2015.11.015

Hartenian, E., & Doench, J. G. (2015). Genetic screens and functional genomics using CRISPR/Cas9 technology. *The FEBS Journal*, *282*(8), 1383–1393. https://doi.org/10.1111/febs.13248

Hartman, S. C., & Buchanan, J. M. (1959). The biosynthesis of the purines. *Ergebnisse Der Physiologie Biologischen Chemie Und Experimentellen Pharmakologie*, *50*(1), 75–121. https://doi.org/10.1007/BF02269579

Hartung, N., Benary, U., Wolf, J., & Kofahl, B. (2017). Paracrine and autocrine regulation of gene expression by Wnt-inhibitor Dickkopf in wild-type and mutant hepatocytes. *BMC Systems Biology*, *11*, 98. https://doi.org/10.1186/s12918-017-0470-9

Hazra, A., Selhub, J., Chao, W.-H., Ueland, P. M., Hunter, D. J., & Baron, J. A. (2010). Uracil misincorporation into DNA and folic acid supplementation123. *The American Journal of Clinical Nutrition*, *91*(1), 160–165. https://doi.org/10.3945/ajcn.2009.28527

Henderson, J. F., & Khoo, K. Y. (1965). ON THE MECHANISM OF FEEDBACK INHIBITION OF PURINE BIOSYNTHESIS DE NOVO IN EHRLICH ASCITES TUMOR CELLS IN VITRO. *The Journal of Biological Chemistry*, *240*, 3104–3109.

Hendriks, I. A., D'Souza, R. C., Chang, J.-G., Mann, M., & Vertegaal, A. C. O. (2015). System-wide identification of wild-type SUMO-2 conjugation sites. *Nature Communications*, *6*(1), 7289. https://doi.org/10.1038/ncomms8289

Hendriks, I. A., & Vertegaal, A. C. O. (2016). A comprehensive compilation of SUMO proteomics. *Nature Reviews. Molecular Cell Biology*, *17*(9), 581–595. https://doi.org/10.1038/nrm.2016.81

Hep G2 [HEPG2] ATCC® HB-8065™. (n.d.). Retrieved 6 July 2019, from http://www.lgcstandards-atcc.org/products/all/HB-8065.aspx?geo_country=ie#characteristics

Herbig, K., Chiang, E.-P., Lee, L.-R., Hills, J., Shane, B., & Stover, P. J. (2002). Cytoplasmic Serine Hydroxymethyltransferase Mediates Competition between Folate-dependent Deoxyribonucleotide andS-Adenosylmethionine Biosyntheses *. *Journal of Biological Chemistry*, *277*(41), 38381–38389. https://doi.org/10.1074/jbc.M205000200

Herzenberg, L. A., Parks, D., Sahaf, B., Perez, O., Roederer, M., & Herzenberg, L. A. (2002). The History and Future of the Fluorescence Activated Cell Sorter and Flow

Cytometry: A View from Stanford. *Clinical Chemistry*, *48*(10), 1819–1827. https://doi.org/10.1093/clinchem/48.10.1819

Hewitt, S. M., Baskin, D. G., Frevert, C. W., Stahl, W. L., & Rosa-Molinar, E. (2014). Controls for Immunohistochemistry. *Journal of Histochemistry and Cytochemistry*, *62*(10), 693–697. https://doi.org/10.1369/0022155414545224

Hoffbrand, A. V., Tripp, E., & Lavoie, A. (1977). *Folate polyglutamate synthesis and breakdown in human cells*. https://scholar.google.com/scholar_lookup?title=Folate+polyglutamate+synthesis+and+breakdown+in+human+cells&author=Hoffbrand%2C+A.V.&publication_year=1977

Hood, J. L., & Emeson, R. B. (2012). Editing of Neurotransmitter Receptor and Ion Channel RNAs in the Nervous System. *Current Topics in Microbiology and Immunology*, *353*, 61–90. https://doi.org/10.1007/82_2011_157

Hori, T., Ayusawa, D., Shimizu, K., Koyama, H., & Seno, T. (1984). Chromosome breakage induced by thymidylate stress in thymidylate synthase-negative mutants of mouse FM3A cells. *Cancer Research*, *44*(2), 703–709.

Horne, D. W., Patterson, D., & Cook, R. J. (1989). Effect of nitrous oxide inactivation of vitamin B12-dependent methionine synthetase on the subcellular distribution of folate coenzymes in rat liver. *Archives of Biochemistry and Biophysics*, *270*(2), 729–733. https://doi.org/10.1016/0003-9861(89)90556-0

Hoshijima, K., Jurynec, M. J., Shaw, D. K., Jacobi, A. M., Behlke, M. A., & Grunwald, D. J. (2019). Highly efficient methods for generating deletion mutations and F0 embryos that lack gene function in zebrafish. *Developmental Cell*, *51*(5), 645-657.e4. https://doi.org/10.1016/j.devcel.2019.10.004

Hough, S. H., & Ajetunmobi, A. (2017). The Future of CRISPR Applications in the Lab, the Clinic and Society. In S. H. Tsang (Ed.), *Precision Medicine, CRISPR, and Genome Engineering: Moving from Association to Biology and Therapeutics* (pp. 157–178). Springer International Publishing. https://doi.org/10.1007/978-3-319-63904-8_9

Housden, B. E., Muhar, M., Gemberling, M., Gersbach, C. A., Stainier, D. Y. R., Seydoux, G., Mohr, S. E., Zuber, J., & Perrimon, N. (2017). Loss-of-function genetic tools for animal models: Cross-species and cross-platform differences. *Nature Reviews. Genetics*, *18*(1), 24–40. https://doi.org/10.1038/nrg.2016.118

Hsieh, Y.-C., Tedeschi, P., AdeBisi Lawal, R., Banerjee, D., Scotto, K., Kerrigan, J. E., Lee, K.-C., Johnson-Farley, N., Bertino, J. R., & Abali, E. E. (2013). Enhanced Degradation of Dihydrofolate Reductase through Inhibition of NAD Kinase by Nicotinamide Analogs. *Molecular Pharmacology*, *83*(2), 339–353. https://doi.org/10.1124/mol.112.080218

Huennekens, F. M. (1994). The methotrexate story: A paradigm for development of cancer chemotherapeutic agents. *Advances in Enzyme Regulation*, *34*, 397–419. https://doi.org/10.1016/0065-2571(94)90025-6

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. https://doi.org/10.1016/j.febslet.2015.05.017

Hum, D. W., Bell, A. W., Rozen, R., & MacKenzie, R. E. (1988). Primary structure of a human trifunctional enzyme. Isolation of a cDNA encoding methylenetetrahydrofolate dehydrogenase-methenyltetrahydrofolate cyclohydrolase-formyltetrahydrofolate synthetase. *The Journal of Biological Chemistry*, *263*(31), 15946–15950.

Hundley, H. A., & Bass, B. L. (2010). ADAR editing in double-stranded UTRs and other noncoding RNA sequences. *Trends in Biochemical Sciences*, *35*(7), 377–383. https://doi.org/10.1016/j.tibs.2010.02.008

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of Bacteriology*, *169*(12), 5429–5433. https://doi.org/10.1128/jb.169.12.5429-5433.1987

Iyer, R., & Tomar, S. K. (2009). Folate: A Functional Food Constituent. *Journal of Food Science*, *74*(9), R114–R122. https://doi.org/10.1111/j.1750-3841.2009.01359.x

Jacques, P. F., Bostom, A. G., Williams, R. R., Ellison, R. C., Eckfeldt, J. H., Rosenberg, I. H., Selhub, J., & Rozen, R. (1996). Relation between folate status, a common mutation in methylenetetrahydrofolate reductase, and plasma homocysteine concentrations. *Circulation*, *93*(1), 7–9. https://doi.org/10.1161/01.cir.93.1.7

Jencks, D. A., & Mathews, R. G. (1987). Allosteric inhibition of methylenetetrahydrofolate reductase by adenosylmethionine. Effects of adenosylmethionine and NADPH on the equilibrium between active and inactive forms of the enzyme and on the kinetics of approach to equilibrium. *The Journal of Biological Chemistry*, *262*(6), 2485–2493.

Jensen, D. E., Black, A. R., Swick, A. G., & Azizkhan, J. C. (1997, October 1). *Distinct roles for Sp1 and E2F sites in the growth/cell cycle regulation of the DHFR promoter*. Journal of Cellular Biochemistry. https://doi.org/10.1002/(SICI)1097-4644(19971001)67:1<24::AID-JCB3>3.0.CO;2-Y

Jiang, Bikard, D., Cox, D., Zhang, F., & Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology*, *31*(3), 233–239. https://doi.org/10.1038/nbt.2508

Jiang, F., & Doudna, J. A. (2017). CRISPR–Cas9 Structures and Mechanisms. *Annual Review of Biophysics*, *46*(1), 505–529. https://doi.org/10.1146/annurev-biophys-062215-010822

Jiménez-Badillo, S. E., Oviedo, N., Hernández-Guzmán, C., González-Mariscal, L., & Hernández-Sánchez, J. (2017). Catsper1 promoter is bidirectional and regulates the expression of a novel lncRNA. *Scientific Reports*, *7*. https://doi.org/10.1038/s41598-017-13867-2

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

Ju, Y., Gong, J., Yang, Y. T., & Zhang, Q. C. (2018). Investigation of RNA-RNA Interactions Using the RISE Database. *Current Protocols in Bioinformatics*, *64*(1), e58. https://doi.org/10.1002/cpbi.58

Kadumuri, R. V., & Janga, S. C. (2018). Epitranscriptomic Code and Its Alterations in Human Disease. *Trends in Molecular Medicine*, *24*(10), 886–903. https://doi.org/10.1016/j.molmed.2018.07.010

Kamynina, E., Lachenauer, E. R., DiRisio, A. C., Liebenthal, R. P., Field, M. S., & Stover, P. J. (2017). Arsenic trioxide targets MTHFD1 and SUMO-dependent nuclear de novo thymidylate biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(12), E2319–E2326. https://doi.org/10.1073/pnas.1619745114

Karlsson, M., Zhang, C., Méar, L., Zhong, W., Digre, A., Katona, B., Sjöstedt, E., Butler, L., Odeberg, J., Dusart, P., Edfors, F., Oksvold, P., von Feilitzen, K., Zwahlen, M., Arif, M., Altay, O., Li, X., Ozcan, M., Mardinoglu, A., … Lindskog, C. (2021). A single–cell type transcriptomics map of human tissues. *Science Advances*, *7*(31), eabh2169. https://doi.org/10.1126/sciadv.abh2169

Karran, P. (2000). DNA double strand break repair in mammalian cells. *Current Opinion in Genetics & Development*, *10*(2), 144–150. https://doi.org/10.1016/s0959-437x(00)00069-1

Kawabe, Y., Komatsu, S., Komatsu, S., Murakami, M., Ito, A., Sakuma, T., Nakamura, T., Yamamoto, T., & Kamihira, M. (2018). Targeted knock-in of an scFv-Fc antibody gene into the hprt locus of Chinese hamster ovary cells using CRISPR/Cas9 and CRIS-PITCh systems. *Journal of Bioscience and Bioengineering*, *125*(5), 599–605. https://doi.org/10.1016/j.jbiosc.2017.12.003

Kawahara, A., Hisano, Y., Ota, S., & Taimatsu, K. (2016). Site-Specific Integration of Exogenous Genes Using Genome Editing Technologies in Zebrafish. *International Journal of Molecular Sciences*, *17*(5), 727. https://doi.org/10.3390/ijms17050727

Khalil, A. M. (2020). The genome editing revolution: Review. *Journal of Genetic Engineering and Biotechnology*, *18*(1), 68. https://doi.org/10.1186/s43141-020-00078-y

Khermesh, K., D'Erchia, A. M., Barak, M., Annese, A., Wachtel, C., Levanon, E. Y., Picardi, E., & Eisenberg, E. (2016). Reduced levels of protein recoding by A-to-I RNA editing in Alzheimer's disease. *RNA*, *22*(2), 290–302. https://doi.org/10.1261/rna.054627.115

Kikuchi, G., Motokawa, Y., Yoshida, T., & Hiraga, K. (2008). Glycine cleavage system: Reaction mechanism, physiological significance, and hyperglycinemia. *Proceedings of the Japan Academy. Series B, Physical and Biological Sciences*, *84*(7), 246–263. https://doi.org/10.2183/pjab/84.246

Kim, D. W., Huang, T., Schirch, D., & Schirch, V. (1996). Properties of tetrahydropteroylpentaglutamate bound to 10-formyltetrahydrofolate dehydrogenase. *Biochemistry*, *35*(49), 15772–15783. https://doi.org/10.1021/bi9619684

Kim, S., Kim, D., Cho, S. W., Kim, J., & Kim, J.-S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Research*, *24*(6), 1012–1019. https://doi.org/10.1101/gr.171322.113

Kleter, B., van Doorn, L. J., ter Schegget, J., Schrauwen, L., van Krimpen, K., Burger, M., ter Harmsel, B., & Quint, W. (1998). Novel short-fragment PCR assay for highly sensitive broad-spectrum detection of anogenital human papillomaviruses. *The American Journal of Pathology*, *153*(6), 1731–1739. https://doi.org/10.1016/S0002-9440(10)65688-X

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

Kong, Q., Wu, M., Huan, Y., Zhang, L., Liu, H., Bou, G., Luo, Y., Mu, Y., & Liu, Z. (2009). Transgene Expression Is Associated with Copy Number and Cytomegalovirus Promoter Methylation in Transgenic Pigs. *PLoS ONE*, *4*(8), e6679. https://doi.org/10.1371/journal.pone.0006679

Konina, D., Sparber, P., Viakhireva, I., Filatova, A., & Skoblov, M. (2021). Investigation of LINC00493/SMIM26 Gene Suggests Its Dual Functioning at mRNA and Protein Level. *International Journal of Molecular Sciences*, *22*(16), 8477. https://doi.org/10.3390/ijms22168477

Kosicki, M., Allen, F., Steward, F., Tomberg, K., Pan, Y., & Bradley, A. (2022). Cas9-induced large deletions and small indels are controlled in a convergent fashion. *Nature Communications*, *13*(1), 3422. https://doi.org/10.1038/s41467-022-30480-8

Kouranova, E., Forbes, K., Zhao, G., Warren, J., Bartels, A., Wu, Y., & Cui, X. (2016). CRISPRs for Optimal Targeting: Delivery of CRISPR Components as DNA, RNA, and Protein into Cultured Cells and Single-Cell Embryos. *Human Gene Therapy*, *27*(6), 464–475. https://doi.org/10.1089/hum.2016.009

Kowalczykowski, S. C. (2015). An Overview of the Molecular Mechanisms of Recombinational DNA Repair. *Cold Spring Harbor Perspectives in Biology*, *7*(11). https://doi.org/10.1101/cshperspect.a016410

Kruman, I. I., & Fowler, A.-K. (2014). Impaired one carbon metabolism and DNA methylation in alcohol toxicity. *Journal of Neurochemistry*, *129*(5), 770–780. https://doi.org/10.1111/jnc.12677

Krupenko, N. I., Holmes, R. S., Tsybovsky, Y., & Krupenko, S. A. (2015). Aldehyde dehydrogenase homologous folate enzymes: Evolutionary switch between cytoplasmic and mitochondrial localization. *Chemico-Biological Interactions*, *234*, 12–17. https://doi.org/10.1016/j.cbi.2014.12.022

Krupenko, S. A., & Krupenko, N. I. (2018). ALDH1L1 and ALDH1L2 Folate Regulatory Enzymes in Cancer. *Advances in Experimental Medicine and Biology*, *1032*, 127–143. https://doi.org/10.1007/978-3-319-98788-0_10

Kumar, S., & Mohapatra, T. (2021). Deciphering Epitranscriptome: Modification of mRNA Bases Provides a New Perspective for Post-transcriptional Regulation of Gene Expression. *Frontiers in Cell and Developmental Biology*, *9*. https://www.frontiersin.org/article/10.3389/fcell.2021.628415

Lan, X., Field, M. S., & Stover, P. J. (2018). Cell cycle regulation of folate-mediated one-carbon metabolism. *Wiley Interdisciplinary Reviews. Systems Biology and Medicine*, *10*(6), e1426. https://doi.org/10.1002/wsbm.1426

Lander, E. S. (2016). The Heroes of CRISPR. *Cell*, *164*(1), 18–28. https://doi.org/10.1016/j.cell.2015.12.041

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., … International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921. https://doi.org/10.1038/35057062

Lang, Y.-D., & Jou, Y.-S. (2021). PSPC1 is a new contextual determinant of aberrant subcellular translocation of oncogenes in tumor progression. *Journal of Biomedical Science*, *28*(1), 57. https://doi.org/10.1186/s12929-021-00753-3

Lawrence, S. A., Hackett, J. C., & Moran, R. G. (2011). Tetrahydrofolate recognition by the mitochondrial folate transporter. *The Journal of Biological Chemistry*, *286*(36), 31480–31489. https://doi.org/10.1074/jbc.M111.272187

Lawrence, S. A., Titus, S. A., Ferguson, J., Heineman, A. L., Taylor, S. M., & Moran, R. G. (2014). Mammalian mitochondrial and cytosolic folylpolyglutamate synthetase maintain the subcellular compartmentalization of folates. *The Journal of Biological Chemistry*, *289*(42), 29386–29396. https://doi.org/10.1074/jbc.M114.593244

Ledford, H. (2020). Super-precise CRISPR tool enhanced by enzyme engineering. *Nature*. https://doi.org/10.1038/d41586-020-00340-w

Ledford, H., & Callaway, E. (2020). Pioneers of revolutionary CRISPR gene editing win chemistry Nobel. *Nature*, *586*(7829), 346–347. https://doi.org/10.1038/d41586-020-02765-9

Lee, S. Y., Baek, M., & Lee, G. M. (2021). Comprehensive characterization of dihydrofolate reductase-mediated gene amplification for the establishment of recombinant human embryonic kidney 293 cells producing monoclonal antibodies. *Biotechnology Journal*, *16*(5), 2000351. https://doi.org/10.1002/biot.202000351

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

Lepoivre, C., Belhocine, M., Bergon, A., Griffon, A., Yammine, M., Vanhille, L., Zacarias-Cabeza, J., Garibal, M.-A., Koch, F., Maqbool, M. A., Fenouil, R., Loriod, B., Holota, H., Gut, M., Gut, I., Imbert, J., Andrau, J.-C., Puthier, D., & Spicuglia, S. (2013). Divergent transcription is associated with promoters of transcriptional regulators. *BMC Genomics*, *14*(1), 914. https://doi.org/10.1186/1471-2164-14-914

Leppek, K., Das, R., & Barna, M. (2018). Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nature Reviews Molecular Cell Biology*, *19*(3), 158–174. https://doi.org/10.1038/nrm.2017.103

Leung, K.-Y., De Castro, S. C. P., Cabreiro, F., Gustavsson, P., Copp, A. J., & Greene, N. D. E. (2013). Folate metabolite profiling of different cell types and embryos suggests variation in folate one-carbon metabolism, including developmental changes in human embryonic brain. *Molecular and Cellular Biochemistry*, *378*(1), 229–236. https://doi.org/10.1007/s11010-013-1613-y

Leung, K.-Y., Pai, Y. J., Chen, Q., Santos, C., Calvani, E., Sudiwala, S., Savery, D., Ralser, M., Gross, S. S., Copp, A. J., & Greene, N. D. E. (2017). Partitioning of One-Carbon Units in Folate and Methionine Metabolism Is Essential for Neural Tube Closure. *Cell Reports*, *21*(7), 1795–1808. https://doi.org/10.1016/j.celrep.2017.10.072

Levanon, E. Y., Eisenberg, E., Yelin, R., Nemzer, S., Hallegger, M., Shemesh, R., Fligelman, Z. Y., Shoshan, A., Pollock, S. R., Sztybel, D., Olshansky, M., Rechavi, G., & Jantsch, M. F. (2004). Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nature Biotechnology*, *22*(8), 1001–1005. https://doi.org/10.1038/nbt996

Li, Q., Youn, J. Y., Siu, K. L., Murugesan, P., Zhang, Y., & Cai, H. (2019). Knockout of dihydrofolate reductase in mice induces hypertension and abdominal aortic aneurysm via mitochondrial dysfunction. *Redox Biology*, *24*, 101185. https://doi.org/10.1016/j.redox.2019.101185

Liang, X., Potter, J., Kumar, S., Zou, Y., Quintanilla, R., Sridharan, M., Carte, J., Chen, W., Roark, N., Ranganathan, S., Ravinder, N., & Chesnut, J. D. (2015). Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *Journal of Biotechnology*, *208*, 44–53. https://doi.org/10.1016/j.jbiotec.2015.04.024

Liberti, M. V., & Locasale, J. W. (2016). The Warburg Effect: How Does it Benefit Cancer Cells? *Trends in Biochemical Sciences*, *41*(3), 211–218. https://doi.org/10.1016/j.tibs.2015.12.001

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

Liew, S.-C., & Gupta, E. D. (2015). Methylenetetrahydrofolate reductase (MTHFR) C677T polymorphism: Epidemiology, metabolism and the associated diseases. *European Journal of Medical Genetics*, *58*(1), 1–10. https://doi.org/10.1016/j.ejmg.2014.10.004

Lin, B. F., Huang, R. F., & Shane, B. (1993). Regulation of folate and one-carbon metabolism in mammalian cells. III. Role of mitochondrial folylpoly-gamma-glutamate synthetase. *Journal of Biological Chemistry*, *268*(29), 21674–21679. https://doi.org/10.1016/S0021-9258(20)80594-8

Lino, C. A., Harper, J. C., Carney, J. P., & Timlin, J. A. (2018). Delivering CRISPR: A review of the challenges and approaches. *Drug Delivery*, *25*(1), 1234–1257. https://doi.org/10.1080/10717544.2018.1474964

Lionaki, E., Ploumi, C., & Tavernarakis, N. (2022). One-Carbon Metabolism: Pulling the Strings behind Aging and Neurodegeneration. *Cells*, *11*(2), 214. https://doi.org/10.3390/cells11020214

Liu, C. T., Hanoian, P., French, J. B., Pringle, T. H., Hammes-Schiffer, S., & Benkovic, S. J. (2013). Functional significance of evolving protein sequence in dihydrofolate reductase from bacteria to humans. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(25), 10159–10164. https://doi.org/10.1073/pnas.1307130110

Liu, M., Rehman, S., Tang, X., Gu, K., Fan, Q., Chen, D., & Ma, W. (2019). Methodologies for Improving HDR Efficiency. *Frontiers in Genetics*, *9*. https://www.frontiersin.org/article/10.3389/fgene.2018.00691

Liu, M.-S., Gong, S., Yu, H.-H., Jung, K., Johnson, K. A., & Taylor, D. W. (2020). Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA. *Nature Communications*, *11*(1), 3576. https://doi.org/10.1038/s41467-020-17411-1

Liu, S. (2020). *Comparison of traditional cloning methods vs. CloneSelect Single-Cell Printer f.sight using CHO cell lines commonly used for monoclonal antibody production*. Molecular Devices. https://www.moleculardevices.com/en/assets/app-note/bpd/comparison-of-traditional-cloning-methods-vs-cloneselect-single-cell-printer-fsight-using-cho-cell-lines

Lorenz, R., Wolfinger, M. T., Tanzer, A., & Hofacker, I. L. (2016). Predicting RNA secondary structures from sequence and probing data. *Methods*, *103*, 86–98. https://doi.org/10.1016/j.ymeth.2016.04.004

Lowe, R., Shirley, N., Bleackley, M., Dolan, S., & Shafee, T. (2017). Transcriptomics technologies. *PLoS Computational Biology*, *13*(5), e1005457. https://doi.org/10.1371/journal.pcbi.1005457

Lu, W., Kwon, Y. K., & Rabinowitz, J. D. (2007). Isotope Ratio-Based Profiling of Microbial Folates. *Journal of the American Society for Mass Spectrometry*, *18*(5), 898–909. https://doi.org/10.1016/j.jasms.2007.01.017

Lucock, M. (2000). Folic Acid: Nutritional Biochemistry, Molecular Biology, and Role in Disease Processes. *Molecular Genetics and Metabolism*, *71*(1), 121–138. https://doi.org/10.1006/mgme.2000.3027

MacFarlane, A. J., Anderson, D. D., Flodby, P., Perry, C. A., Allen, R. H., Stabler, S. P., & Stover, P. J. (2011). Nuclear Localization of de Novo Thymidylate Biosynthesis Pathway Is Required to Prevent Uracil Accumulation in DNA*. *Journal of Biological Chemistry*, *286*(51), 44015–44022. https://doi.org/10.1074/jbc.M111.307629

MacFarlane, A. J., Liu, X., Perry, C. A., Flodby, P., Allen, R. H., Stabler, S. P., & Stover, P. J. (2008). Cytoplasmic Serine Hydroxymethyltransferase Regulates the Metabolic Partitioning of Methylenetetrahydrofolate but Is Not Essential in Mice *. *Journal of Biological Chemistry*, *283*(38), 25846–25853. https://doi.org/10.1074/jbc.M802671200

Manieri, W., Moore, M. E., Soellner, M. B., Tsang, P., & Caperelli, C. A. (2007). Human Glycinamide Ribonucleotide Transformylase: Active Site Mutants as Mechanistic Probes. *Biochemistry*, *46*(1), 156–163. https://doi.org/10.1021/bi0619270

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

Martin, R. M., Ikeda, K., Cromer, M. K., Uchida, N., Nishimura, T., Romano, R., Tong, A. J., Lemgart, V. T., Camarena, J., Pavel-Dinu, M., Sindhu, C., Wiebking, V., Vaidyanathan, S., Dever, D. P., Bak, R. O., Laustsen, A., Lesch, B. J., Jakobsen, M. R., Sebastiano, V., … Porteus, M. H. (2019). Highly Efficient and Marker-free Genome Editing of Human Pluripotent Stem Cells by CRISPR-Cas9 RNP and AAV6 Donor-Mediated Homologous Recombination. *Cell Stem Cell*, *24*(5), 821-828.e5. https://doi.org/10.1016/j.stem.2019.04.001

Martin, S. A. M., Dehler, C. E., & Król, E. (2016). Transcriptomic responses in the fish intestine. *Developmental and Comparative Immunology*, *64*, 103–117. https://doi.org/10.1016/j.dci.2016.03.014

Maruvada, P., Stover, P. J., Mason, J. B., Bailey, R. L., Davis, C. D., Field, M. S., Finnell, R. H., Garza, C., Green, R., Gueant, J.-L., Jacques, P. F., Klurfeld, D. M., Lamers, Y., MacFarlane, A. J., Miller, J. W., Molloy, A. M., O'Connor, D. L., Pfeiffer, C. M., Potischman, N. A., … Zappalà, G. (2020). Knowledge gaps in understanding the metabolic and clinical effects of excess folates/folic acid: A summary, and perspectives, from an NIH workshop. *The American Journal of Clinical Nutrition*, *112*(5), 1390–1403. https://doi.org/10.1093/ajcn/nqaa259

Mauger, D. M., Cabral, B. J., Presnyak, V., Su, S. V., Reid, D. W., Goodman, B., Link, K., Khatwani, N., Reynders, J., Moore, M. J., & McFadyen, I. J. (2019). MRNA structure regulates protein expression through changes in functional half-life. *Proceedings of the National Academy of Sciences*, *116*(48), 24075–24083. https://doi.org/10.1073/pnas.1908052116

Maurer, B. J., Barker, P. E., Masters, J. N., Ruddle, F. H., & Attardi, G. (1984). Human dihydrofolate reductase gene is located in chromosome 5 and is unlinked to the related pseudogenes. *Proceedings of the National Academy of Sciences of the United States of America*, *81*(5), 1484–1488. https://doi.org/10.1073/pnas.81.5.1484

McEntee, G., Minguzzi, S., O'Brien, K., Larbi, N. B., Loscher, C., Ó'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*, *108*(37), 15157–15162. https://doi.org/10.1073/pnas.1103605108

McGarry, D. J., & Olson, M. F. (2016). Coping with Loss: Cell Adaptation to Cytoskeleton Disruption. *Developmental Cell*, *39*(1), 3–4. https://doi.org/10.1016/j.devcel.2016.09.020

McVey, M., & Lee, S. E. (2008). MMEJ repair of double-strand breaks (director's cut): Deleted sequences and alternative endings. *Trends in Genetics : TIG*, *24*(11), 529–538. https://doi.org/10.1016/j.tig.2008.08.007

Meiser, J., Tumanov, S., Maddocks, O., Labuschagne, C. F., Athineos, D., Van Den Broek, N., Mackay, G. M., Gottlieb, E., Blyth, K., Vousden, K., Kamphorst, J. J., & Vazquez, A. (2016). Serine one-carbon catabolism with formate overflow. *Science Advances*, *2*(10), e1601273. https://doi.org/10.1126/sciadv.1601273

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

Menezo, Y., Elder, K., Clement, A., & Clement, P. (2022). Folic Acid, Folinic Acid, 5 Methyl TetraHydroFolate Supplementation for Mutations That Affect Epigenesis through the Folate and One-Carbon Cycles. *Biomolecules*, *12*(2), 197. https://doi.org/10.3390/biom12020197

Mensah, E. O., Guo, X.-Y., Gao, X.-D., & Fujita, M. (2019). Establishment of DHFR-deficient HEK293 cells for high yield of therapeutic glycoproteins. *Journal of Bioscience and Bioengineering*, *128*(4), 487–494. https://doi.org/10.1016/j.jbiosc.2019.04.005

Mignone, F., Gissi, C., Liuni, S., & Pesole, G. (2002). Untranslated regions of mRNAs. *Genome Biology*, *3*(3), reviews0004.1. https://doi.org/10.1186/gb-2002-3-3-reviews0004

Mills, J. L. (2017). Strategies for Preventing Folate-Related Neural Tube Defects: Supplements, Fortified Foods, or Both? *JAMA*, *317*(2), 144–145. https://doi.org/10.1001/jama.2016.19894

Minton, D. R., Nam, M., McLaughlin, D. J., Shin, J., Bayraktar, E. C., Alvarez, S. W., Sviderskiy, V. O., Papagiannakopoulos, T., Sabatini, D. M., Birsoy, K., & Possemato, R. (2018). Serine Catabolism by SHMT2 Is Required for Proper Mitochondrial Translation Initiation and Maintenance of Formylmethionyl-tRNAs. *Molecular Cell*, *69*(4), 610-621.e5. https://doi.org/10.1016/j.molcel.2018.01.024

Misselbeck, K., Marchetti, L., Field, M. S., Scotti, M., Priami, C., & Stover, P. J. (2017). A hybrid stochastic model of folate-mediated one-carbon metabolism: Effect of the common C677T MTHFR variant on de novo thymidylate biosynthesis. *Scientific Reports*, *7*(1), 797. https://doi.org/10.1038/s41598-017-00854-w

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

Mitchell, H. K., Snell, E. E., & Williams, R. J. (1944). Folic Acid. I. Concentration from Spinach1a. *Journal of the American Chemical Society*, *66*(2), 267–268. https://doi.org/10.1021/ja01230a032

Miyaoka, Y., Berman, J. R., Cooper, S. B., Mayerl, S. J., Chan, A. H., Zhang, B., Karlin-Neumann, G. A., & Conklin, B. R. (2016). Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. *Scientific Reports*, *6*(1), 23549. https://doi.org/10.1038/srep23549

Mohamadi, S., Zaker Bostanabad, S., & Mirnejad, R. (2020). CRISPR Arrays: A Review on Its Mechanism. *Journal of Applied Biotechnology Reports*, *7*(2), 81–86. https://doi.org/10.30491/jabr.2020.109380

Molloy, A. M., & Scott, J. M. (2001). Folates and prevention of disease. *Public Health Nutrition*, *4*(2b), 601–609. https://doi.org/10.1079/PHN2001144

Moran, R. G. (1999). Roles of folylpoly-gamma-glutamate synthetase in therapeutics with tetrahydrofolate antimetabolites: An overview. *Seminars in Oncology*, *26*(2 Suppl 6), 24–32.

Morellato, A. E., Umansky, C., & Pontel, L. B. (2021). The toxic side of one-carbon metabolism and epigenetics. *Redox Biology*, *40*, 101850. https://doi.org/10.1016/j.redox.2020.101850

Morley, A. A., Trainor, K. J., & Seshadri, R. S. (1983). Cloning of human lymphocytes using limiting dilution. *Experimental Hematology*, *11*(5), 418–424.

Morscher, R. J., Ducker, G. S., Li, S. H.-J., Mayer, J. A., Gitai, Z., Sperl, W., & Rabinowitz, J. D. (2018). Mitochondrial translation requires folate-dependent tRNA methylation. *Nature*, *554*(7690), 128–132. https://doi.org/10.1038/nature25460

Mout, R., Ray, M., Yesilbag Tonga, G., Lee, Y.-W., Tay, T., Sasaki, K., & Rotello, V. M. (2017). Direct Cytosolic Delivery of CRISPR/Cas9-Ribonucleoprotein for Efficient Gene Editing. *ACS Nano*, *11*(3), 2452–2458. https://doi.org/10.1021/acsnano.6b07600

Mudd, S. H., Brosnan, J. T., Brosnan, M. E., Jacobs, R. L., Stabler, S. P., Allen, R. H., Vance, D. E., & Wagner, C. (2007). Methyl balance and transmethylation fluxes in humans. *The American Journal of Clinical Nutrition*, *85*(1), 19–25. https://doi.org/10.1093/ajcn/85.1.19

Mullenders, J., & Bernards, R. (2009). Loss-of-function genetic screens as a tool to improve the diagnosis and treatment of cancer. *Oncogene*, *28*(50), 4409–4420. https://doi.org/10.1038/onc.2009.295

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

Nachtergaele, S., & He, C. (2018). Chemical Modifications in the Life of an mRNA Transcript. *Annual Review of Genetics*, *52*, 349–372. https://doi.org/10.1146/annurev-genet-120417-031522

Naderi, N., & House, J. D. (2018). Recent Developments in Folate Nutrition. *Advances in Food and Nutrition Research*, *83*, 195–213. https://doi.org/10.1016/bs.afnr.2017.12.006

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*. https://doi.org/10.1038/ncomms6560

Nakagawa, S., Yamazaki, T., & Hirose, T. (2018). Molecular dissection of nuclear paraspeckles: Towards understanding the emerging world of the RNP milieu. *Open Biology*, *8*(10), 180150. https://doi.org/10.1098/rsob.180150

Nakanishi, T., Kuroiwa, A., Yamada, S., Isotani, A., Yamashita, A., Tairaka, A., Hayashi, T., Takagi, T., Ikawa, M., Matsuda, Y., & Okabe, M. (2002). FISH analysis of 142 EGFP transgene integration sites into the mouse genome. *Genomics*, *80*(6), 564–574. https://doi.org/10.1006/geno.2002.7008

Nam, J.-W., Choi, S.-W., & You, B.-H. (2016). Incredible RNA: Dual Functions of Coding and Noncoding. *Molecules and Cells*, *39*(5), 367–374. https://doi.org/10.14348/molcells.2016.0039

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

Nicolas, E., Roumillac, C., & Trouche, D. (2003). Balance between Acetylation and Methylation of Histone H3 Lysine 9 on the E2F-Responsive Dihydrofolate Reductase Promoter. *Molecular and Cellular Biology*, *23*(5), 1614–1622. https://doi.org/10.1128/MCB.23.5.1614-1622.2003

Nidhi, S., Anand, U., Oleksak, P., Tripathi, P., Lal, J. A., Thomas, G., Kuca, K., & Tripathi, V. (2021). Novel CRISPR–Cas Systems: An Updated Review of the Current Achievements, Applications, and Future Research Perspectives. *International Journal of Molecular Sciences*, *22*(7), 3327. https://doi.org/10.3390/ijms22073327

Nijhout, H. F., Reed, M. C., Budu, P., & Ulrich, C. M. (2004). A Mathematical Model of the Folate Cycle: NEW INSIGHTS INTO FOLATE HOMEOSTASIS *. *Journal of Biological Chemistry*, *279*(53), 55008–55016. https://doi.org/10.1074/jbc.M410818200

Nilsson, R., Nicolaidou, V., & Koufaris, C. (2019). Mitochondrial MTHFD isozymes display distinct expression, regulation, and association with cancer. *Gene*, *716*, 144032. https://doi.org/10.1016/j.gene.2019.144032

Nishikura, K. (2016). A-to-I editing of coding and non-coding RNAs by ADARs. *Nature Reviews Molecular Cell Biology*, *17*(2), 83–96. https://doi.org/10.1038/nrm.2015.4

Noé, V., MacKenzie, S., & Ciudad, C. J. (2003). An Intron Is Required for Dihydrofolate Reductase Protein Stability*. *Journal of Biological Chemistry*, *278*(40), 38292–38300. https://doi.org/10.1074/jbc.M212746200

Nusse, R. (2005). Wnt signaling in disease and in development. *Cell Research*, *15*(1), 28–32. https://doi.org/10.1038/sj.cr.7290260

Nusse, R., & Clevers, H. (2017). Wnt/β-Catenin Signaling, Disease, and Emerging Therapeutic Modalities. *Cell*, *169*(6), 985–999. https://doi.org/10.1016/j.cell.2017.05.016

Oefner, C., D'arcy, A., & Winkler, F. K. (1988). Crystal structure of human dihydrofolate reductase complexed with folate. *European Journal of Biochemistry*, *174*(2), 377–385. https://doi.org/10.1111/j.1432-1033.1988.tb14108.x

Office of Dietary Supplements—Folate. (n.d.). Retrieved 20 May 2022, from https://ods.od.nih.gov/factsheets/Folate-HealthProfessional/

Ohrvik, V. E., & Witthoft, C. M. (2011). Human Folate Bioavailability. *Nutrients*, *3*(4), 475–490. https://doi.org/10.3390/nu3040475

Olds, W., & Li, J. (2016). siRNA knockdown validation 101: Incorporating negative controls in antibody research. *F1000Research*, *5*, 308. https://doi.org/10.12688/f1000research.8159.1

Pai, Y. J., Leung, K.-Y., Savery, D., Hutchin, T., Prunty, H., Heales, S., Brosnan, M. E., Brosnan, J. T., Copp, A. J., & Greene, N. D. E. (2015). Glycine decarboxylase deficiency causes neural tube defects and features of non-ketotic hyperglycinemia in mice. *Nature Communications*, *6*, 6388. https://doi.org/10.1038/ncomms7388

Palmer, A. M., Kamynina, E., Field, M. S., & Stover, P. J. (2017). Folate rescues vitamin B12 depletion-induced inhibition of nuclear thymidylate biosynthesis and genome instability. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(20), E4095–E4102. https://doi.org/10.1073/pnas.1619582114

Pannunzio, N. R., Watanabe, G., & Lieber, M. R. (2018). Nonhomologous DNA end-joining for repair of DNA double-strand breaks. *Journal of Biological Chemistry*, *293*(27), 10512–10523. https://doi.org/10.1074/jbc.TM117.000374

Paone, A., Marani, M., Fiascarelli, A., Rinaldo, S., Giardina, G., Contestabile, R., Paiardini, A., & Cutruzzolà, F. (2014). SHMT1 knockdown induces apoptosis in lung cancer cells by causing uracil misincorporation. *Cell Death & Disease*, *5*, e1525. https://doi.org/10.1038/cddis.2014.482

Papatheodorou, I., Moreno, P., Manning, J., Fuentes, A. M.-P., George, N., Fexova, S., Fonseca, N. A., Füllgrabe, A., Green, M., Huang, N., Huerta, L., Iqbal, H., Jianu, M., Mohammed, S., Zhao, L., Jarnuczak, A. F., Jupp, S., Marioni, J., Meyer, K., … Brazma, A. (2020). Expression Atlas update: From tissues to single cells. *Nucleic Acids Research*, *48*(D1), D77–D83. https://doi.org/10.1093/nar/gkz947

Patanwala, I., King, M. J., Barrett, D. A., Rose, J., Jackson, R., Hudson, M., Philo, M., Dainty, J. R., Wright, A. J., Finglas, P. M., & Jones, D. E. (2014). Folic acid handling by the human gut: Implications for food fortification and supplementation123. *The American Journal of Clinical Nutrition*, *100*(2), 593–599. https://doi.org/10.3945/ajcn.113.080507

Pedley, A. M., & Benkovic, S. J. (2017). A New View into the Regulation of Purine Metabolism – The Purinosome. *Trends in Biochemical Sciences*, *42*(2), 141–154. https://doi.org/10.1016/j.tibs.2016.09.009

Pejchal, R., Campbell, E., Guenther, B. D., Lennon, B. W., Matthews, R. G., & Ludwig, M. L. (2006). Structural Perturbations in the Ala → Val Polymorphism of Methylenetetrahydrofolate Reductase: How Binding of Folates May Protect against Inactivation. *Biochemistry*, *45*(15), 4808–4818. https://doi.org/10.1021/bi052294c

Petrenko, N., Jin, Y., Dong, L., Wong, K. H., & Struhl, K. (2019). Requirements for RNA polymerase II preinitiation complex formation in vivo. *ELife*, *8*, e43654. https://doi.org/10.7554/eLife.43654

Pietzke, M., Meiser, J., & Vazquez, A. (2020). Formate metabolism in health and disease. *Molecular Metabolism*, *33*, 23–37. https://doi.org/10.1016/j.molmet.2019.05.012

Pike, S. T., Rajendra, R., Artzt, K., & Appling, D. R. (2010). Mitochondrial C1-tetrahydrofolate synthase (MTHFD1L) supports the flow of mitochondrial one-carbon units into the methyl cycle in embryos. *The Journal of Biological Chemistry*, *285*(7), 4612–4620. https://doi.org/10.1074/jbc.M109.079855

Pillai-Kastoori, L., Heaton, S., Shiflett, S. D., Roberts, A. C., Solache, A., & Schutz-Geschwender, A. R. (2020). Antibody validation for Western blot: By the user, for the user. *The Journal of Biological Chemistry*, *295*(4), 926–939. https://doi.org/10.1074/jbc.RA119.010472

PITCh KnockIn | PITCh designer 2.0. (n.d.). Retrieved 2 July 2019, from http://www.mls.sci.hiroshima-u.ac.jp/smg/PITChdesigner/index.html

Porath, H. T., Knisbacher, B. A., Eisenberg, E., & Levanon, E. Y. (2017). Massive A-to-I RNA editing is common across the Metazoa and correlates with dsRNA abundance. *Genome Biology*, *18*(1), 185. https://doi.org/10.1186/s13059-017-1315-y

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

Prevention of neural tube defects: Results of the Medical Research Council Vitamin Study. (1991). *The Lancet*, *338*(8760), 131–137. https://doi.org/10.1016/0140-6736(91)90133-A

Quinlivan, E. P., Davis, S. R., Shelnutt, K. P., Henderson, G. N., Ghandour, H., Shane, B., Selhub, J., Bailey, L. B., Stacpoole, P. W., & Gregory, J. F. (2005). Methylenetetrahydrofolate reductase 677C->T polymorphism and folate status affect one-carbon incorporation into human DNA deoxynucleosides. *The Journal of Nutrition*, *135*(3), 389–396. https://doi.org/10.1093/jn/135.3.389

Raghubeer, S., & Matsha, T. E. (2021). Methylenetetrahydrofolate (MTHFR), the One-Carbon Cycle, and Cardiovascular Risks. *Nutrients*, *13*(12), 4562. https://doi.org/10.3390/nu13124562

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

Rajagopalan, P. T. R., Zhang, Z., McCourt, L., Dwyer, M., Benkovic, S. J., & Hammes, G. G. (2002). Interaction of dihydrofolate reductase with methotrexate: Ensemble and single-molecule kinetics. *Proceedings of the National Academy of Sciences*, *99*(21), 13481–13486. https://doi.org/10.1073/pnas.172501499

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, *8*(11), 2281–2308. https://doi.org/10.1038/nprot.2013.143

Rath, D., Amlinger, L., Rath, A., & Lundgren, M. (2015). The CRISPR-Cas immune system: Biology, mechanisms and applications. *Biochimie*, *117*, 119–128. https://doi.org/10.1016/j.biochi.2015.03.025

Razak, M. A., Begum, P. S., Viswanath, B., & Rajagopal, S. (2017). Multifarious Beneficial Effect of Nonessential Amino Acid, Glycine: A Review. *Oxidative Medicine and Cellular Longevity*, *2017*, 1716701. https://doi.org/10.1155/2017/1716701

Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., Wadi, L., Meyer, M., Wong, J., Xu, C., Merico, D., & Bader, G. D. (2019). Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nature Protocols*, *14*(2), 482–517. https://doi.org/10.1038/s41596-018-0103-9

Rice, G. I., Kasher, P. R., Forte, G. M. A., Mannion, N. M., Greenwood, S. M., Szynkiewicz, M., Dickerson, J. E., Bhaskar, S. S., Zampini, M., Briggs, T. A., Jenkinson, E. M., Bacino, C. A., Battini, R., Bertini, E., Brogan, P. A., Brueton, L. A., Carpanelli, M., De Laet, C., de Lonlay, P., … Crow, Y. J. (2012). Mutations in ADAR1 cause Aicardi-Goutières syndrome associated with a type I interferon signature. *Nature Genetics*, *44*(11), 1243–1248. https://doi.org/10.1038/ng.2414

Robb, G. B. (2019). Genome Editing with CRISPR-Cas: An Overview. *Current Protocols Essential Laboratory Techniques*, *19*(1), e36. https://doi.org/10.1002/cpet.36

Rodríguez-Rodríguez, D. R., Ramírez-Solís, R., Garza-Elizondo, M. A., Garza-Rodríguez, M. D. L., & Barrera-Saldaña, H. A. (2019). Genome editing: A perspective on the application of CRISPR/Cas9 to study human diseases (Review). *International Journal of Molecular Medicine*, *43*(4), 1559–1574. https://doi.org/10.3892/ijmm.2019.4112

Rosenthal, J. J. C., & Seeburg, P. H. (2012). A-to-I RNA Editing: Effects on Proteins Key to Neural Excitability. *Neuron*, *74*(3), 432–439. https://doi.org/10.1016/j.neuron.2012.04.010

Rosenzweig, A., Blenis, J., & Gomes, A. P. (2018). Beyond the Warburg Effect: How Do Cancer Cells Regulate One-Carbon Metabolism? *Frontiers in Cell and Developmental Biology*, *6*. https://doi.org/10.3389/fcell.2018.00090

Roundtree, I. A., Evans, M. E., Pan, T., & He, C. (2017). Dynamic RNA modifications in gene expression regulation. *Cell*, *169*(7), 1187–1200. https://doi.org/10.1016/j.cell.2017.05.045

Rozov, S. M., Permyakova, N. V., & Deineko, E. V. (2019). The Problem of the Low Rates of CRISPR/Cas9-Mediated Knock-ins in Plants: Approaches and Solutions. *International Journal of Molecular Sciences*, *20*(13), 3371. https://doi.org/10.3390/ijms20133371

Ryan, J. A. (n.d.). *Cell Cloning by Serial Dilution in 96 Well Plates*. 3.

Sakuma, T., Nakade, S., Sakane, Y., Suzuki, K.-I. T., & Yamamoto, T. (2016). MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems. *Nature Protocols*, *11*(1), 118–133. https://doi.org/10.1038/nprot.2015.140

Sakuma, T., Nishikawa, A., Kume, S., Chayama, K., & Yamamoto, T. (2014). Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Scientific Reports*, *4*, 5400. https://doi.org/10.1038/srep05400

Sanderson, S. M., Gao, X., Dai, Z., & Locasale, J. W. (2019). Methionine metabolism in health and cancer: A nexus of diet and precision medicine. *Nature Reviews Cancer*, *19*(11), 625–637. https://doi.org/10.1038/s41568-019-0187-8

Sato, K., Kato, Y., Hamada, M., Akutsu, T., & Asai, K. (2011). IPknot: Fast and accurate prediction of RNA secondary structures with pseudoknots using integer

programming. *Bioinformatics (Oxford, England)*, *27*(13), i85-93. https://doi.org/10.1093/bioinformatics/btr215

Sauberlich, H. E., Kretsch, M. J., Skala, J. H., Johnson, H. L., & Taylor, P. C. (1987). Folate requirement and metabolism in nonpregnant women. *The American Journal of Clinical Nutrition*, *46*(6), 1016–1028. https://doi.org/10.1093/ajcn/46.6.1016

Savva, Y. A., Rieder, L. E., & Reenan, R. A. (2012). The ADAR protein family. *Genome Biology*, *13*(12), 252. https://doi.org/10.1186/gb-2012-13-12-252

Scaglione, F., & Panzavolta, G. (2014). Folate, folic acid and 5-methyltetrahydrofolate are not the same thing. *Xenobiotica*, *44*(5), 480–488. https://doi.org/10.3109/00498254.2013.845705

Schaffer, A. A., & Levanon, E. Y. (2021). ALU A-to-I RNA Editing: Millions of Sites and Many Open Questions. In E. Picardi & G. Pesole (Eds.), *RNA Editing: Methods and Protocols* (pp. 149–162). Springer US. https://doi.org/10.1007/978-1-0716-0787-9_9

Schilling, L. J., & Farnham, P. J. (1994). Transcriptional regulation of the dihydrofolate reductase/rep-3 locus. *Critical Reviews in Eukaryotic Gene Expression*, *4*(1), 19–53.

Schirch, V., & Strong, W. B. (1989). Interaction of folylpolyglutamates with enzymes in one-carbon metabolism. *Archives of Biochemistry and Biophysics*, *269*(2), 371–380. https://doi.org/10.1016/0003-9861(89)90120-3

Schmieder, V., Bydlinski, N., Strasser, R., Baumann, M., Kildegaard, H. F., Jadhav, V., & Borth, N. (2018). Enhanced Genome Editing Tools For Multi-Gene Deletion Knock-Out Approaches Using Paired CRISPR sgRNAs in CHO Cells. *Biotechnology Journal*, *13*(3), 1700211. https://doi.org/10.1002/biot.201700211

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675. https://doi.org/10.1038/nmeth.2089

Schwarzkopf, M., Liu, M. C., Schulte, S. J., Ives, R., Husain, N., Choi, H. M. T., & Pierce, N. A. (2021). Hybridization chain reaction enables a unified approach to multiplexed, quantitative, high-resolution immunohistochemistry and in situ hybridization. *Development*, *148*(22), dev199847. https://doi.org/10.1242/dev.199847

Schweitzer, B. I., Dicker, A. P., & Bertino, J. R. (1990). Dihydrofolate reductase as a therapeutic target. *The FASEB Journal*, *4*(8), 2441–2452. https://doi.org/10.1096/fasebj.4.8.2185970

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

Sebastian-delaCruz, M., Gonzalez-Moro, I., Olazagoitia-Garmendia, A., Castellanos-Rubio, A., & Santin, I. (2021). The Role of lncRNAs in Gene Expression Regulation through mRNA Stabilization. *Non-Coding RNA*, *7*(1), 3. https://doi.org/10.3390/ncrna7010003

Seila, A. C., Core, L. J., Lis, J. T., & Sharp, P. A. (2009). Divergent transcription: A new feature of active promoters. *Cell Cycle (Georgetown, Tex.)*, *8*(16), 2557–2564. https://doi.org/10.4161/cc.8.16.9305

Selhub, J. (1999). Homocysteine metabolism. *Annual Review of Nutrition*, *19*, 217–246. https://doi.org/10.1146/annurev.nutr.19.1.217

Seol, J.-H., Shim, E. Y., & Lee, S. E. (2018). Microhomology-mediated end joining: Good, bad and ugly. *Mutation Research*, *809*, 81–87. https://doi.org/10.1016/j.mrfmmm.2017.07.002

Shane, B. (1989). Folylpolyglutamate synthesis and role in the regulation of one-carbon metabolism. *Vitamins and Hormones*, *45*, 263–335. https://doi.org/10.1016/s0083-6729(08)60397-0

Shane, B. (2001). Folate Chemistry and Metabolism*. *Clinical Research and Regulatory Affairs*, *18*(3), 137–159. https://doi.org/10.1081/CRP-100108170

Shane, B. (2008). Folate and Vitamin B12 Metabolism: Overview and Interaction with Riboflavin, Vitamin B6, and Polymorphisms. *Food and Nutrition Bulletin*, *29*(2_suppl1), S5–S16. https://doi.org/10.1177/15648265080292S103

Sharp, P. A. (2009). The Centrality of RNA. *Cell*, *136*(4), 577–580. https://doi.org/10.1016/j.cell.2009.02.007

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), 1–8. https://doi.org/10.1016/0378-1119(84)90188-4

Shimada, T., Inokuchi, K., & Nienhuis, A. W. (1986). Chromatin structure of the human dihydrofolate reductase gene promoter. Multiple protein-binding sites. *Journal of Biological Chemistry*, *261*(3), 1445–1452. https://doi.org/10.1016/S0021-9258(17)36113-6

Shin, Y. S., Chan, C., Vidal, A. J., Brody, T., & Stokstad, E. L. (1976). Subcellular localization of gamma-glutamyl carboxypeptidase and of folates. *Biochimica Et Biophysica Acta*, *444*(3), 794–801. https://doi.org/10.1016/0304-4165(76)90326-3

Shulpekova, Y., Nechaev, V., Kardasheva, S., Sedova, A., Kurbatova, A., Bueverova, E., Kopylov, A., Malsagova, K., Dlamini, J. C., & Ivashkin, V. (2021). The Concept of

Folic Acid in Health and Disease. *Molecules*, *26*(12), 3731. https://doi.org/10.3390/molecules26123731

Sjöstedt, E., Zhong, W., Fagerberg, L., Karlsson, M., Mitsios, N., Adori, C., Oksvold, P., Edfors, F., Limiszewska, A., Hikmet, F., Huang, J., Du, Y., Lin, L., Dong, Z., Yang, L., Liu, X., Jiang, H., Xu, X., Wang, J., … Mulder, J. (2020). An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science*, *367*(6482), eaay5947. https://doi.org/10.1126/science.aay5947

Slansky, J. E., & Farnham, P. J. (1996, January 1). *Transcriptional regulation of the dihydrofolate reductase gene*. BioEssays. https://doi.org/10.1002/bies.950180111

Slotkin, W., & Nishikura, K. (2013). Adenosine-to-inosine RNA editing and human disease. *Genome Medicine*, *5*(11), 105. https://doi.org/10.1186/gm508

Soitu, C., Deroy, C., Castrejón-Pita, A. A., Cook, P. R., & Walsh, E. J. (2020). Using Fluid Walls for Single-Cell Cloning Provides Assurance in Monoclonality. *SLAS TECHNOLOGY: Translating Life Sciences Innovation*, *25*(3), 267–275. https://doi.org/10.1177/2472630319891135

Spizzichino, S., Boi, D., Boumis, G., Lucchi, R., Liberati, F. R., Capelli, D., Montanari, R., Pochetti, G., Paone, A., Rinaldo, S., Contestabile, R., Paiardini, A., Tramonti, A., Giardina, G., & Cutruzzolà, F. (2020). *Cytosolic localization and in vitro assembly of human de novo thymidylate synthesis complex* (p. 2020.12.23.423904). bioRxiv. https://doi.org/10.1101/2020.12.23.423904

Stark, M., Raz, S., & Assaraf, Y. G. (2021). Folylpoly-ɣ-glutamate synthetase association to the cytoskeleton: Implications to folate metabolon compartmentalization. *Journal of Proteomics*, *239*, 104169. https://doi.org/10.1016/j.jprot.2021.104169

Staszewski, R. (1990). Murphy's law of limiting dilution cloning. *Statistics in Medicine*, *9*(4), 457–461. https://doi.org/10.1002/sim.4780090416

Steenbergen, R., Oti, M., ter Horst, R., Tat, W., Neufeldt, C., Belovodskiy, A., Chua, T. T., Cho, W. J., Joyce, M., Dutilh, B. E., & Tyrrell, D. L. (2018). Establishing normal metabolism and differentiation in hepatocellular carcinoma cells by culturing in adult human serum. *Scientific Reports*, *8*(1), 11685. https://doi.org/10.1038/s41598-018-29763-2

Sternberg, S. H., & Doudna, J. A. (2015). Expanding the Biologist's Toolkit with CRISPR-Cas9. *Molecular Cell*, *58*(4), 568–574. https://doi.org/10.1016/j.molcel.2015.02.032

Stover, & Field, M. S. (2011). Trafficking of Intracellular Folates12. *Advances in Nutrition*, *2*(4), 325–331. https://doi.org/10.3945/an.111.000596

Stover, P. J. (2004). Physiology of folate and vitamin B12 in health and disease. *Nutrition Reviews*, *62*(6 Pt 2), S3-12; discussion S13. https://doi.org/10.1111/j.1753-4887.2004.tb00070.x

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. (2012). 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., & Schirch, V. (1993). The metabolic role of leucovorin. *Trends in Biochemical Sciences*, *18*(3), 102–106. https://doi.org/10.1016/0968-0004(93)90162-g

Suh, J. R., Herbig, A. K., & Stover, P. J. (2001). New perspectives on folate catabolism. *Annual Review of Nutrition*, *21*, 255–282. https://doi.org/10.1146/annurev.nutr.21.1.255

Sumner, J. S., & Matthews, R. G. (1992). Stereochemistry and mechanism of hydrogen transfer between NADPH and methylenetetrahydrofolate in the reaction catalyzed by methylenetetrahydrofolate reductase from pig liver. *Journal of the American Chemical Society*, *114*(18), 6949–6956. https://doi.org/10.1021/ja00044a001

Sun, L., Fazal, F. M., Li, P., Broughton, J. P., Lee, B., Tang, L., Huang, W., Kool, E. T., Chang, H. Y., & Zhang, Q. C. (2019). RNA structure maps across mammalian cellular compartments. *Nature Structural & Molecular Biology*, *26*(4), 322–330. https://doi.org/10.1038/s41594-019-0200-7

Sun, S., Gui, Y., Jiang, Q., & Song, H. (2011). Dihydrofolate reductase is required for the development of heart and outflow tract in zebrafish. *Acta Biochimica Et Biophysica Sinica*, *43*(12), 957–969. https://doi.org/10.1093/abbs/gmr098

Sun, S., Gui, Y., Wang, Y., Qian, L., Jiang, Q., Liu, D., & Song, H. (2007). Effect of dihydrofolate reductase gene knock-down on the expression of heart and neural crest derivatives expressed transcript 2 in zebrafish cardiac development. *Chinese Medical Journal*, *120*(13), 1166–1171.

Suzuki, M. T., & Giovannoni, S. J. (1996). Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Applied and Environmental Microbiology*, *62*(2), 625–630. https://doi.org/10.1128/aem.62.2.625-630.1996

Swarbrick, J., Iliades, P., Simpson, J. S., & Macreadie, I. (2008). Folate Biosynthesis—Reappraisal of Old and Novel Targets in the Search for New Antimicrobials. *The Open Enzyme Inhibition Journal*, *1*(1), 12–33. https://doi.org/10.2174/1874940200801010012

Symington, L. S., & Gautier, J. (2011). Double-Strand Break End Resection and Repair Pathway Choice. *Annual Review of Genetics*, *45*(1), 247–271. https://doi.org/10.1146/annurev-genet-110410-132435

Tabatabaei, R. S., Fatahi-Meibodi, N., Meibodi, B., Javaheri, A., Abbasi, H., Hadadan, A., Bahrami, R., Mirjalili, S. R., Karimi-Zarchi, M., & Neamatzadeh, H. (2022). Association of Fetal MTHFR C677T Polymorphism with Susceptibility to Neural Tube Defects: A Systematic Review and Update Meta-Analysis. *Fetal and Pediatric Pathology*, *41*(2), 225–241. https://doi.org/10.1080/15513815.2020.1775734

Tai, N., Ding, Y., Schmitz, J. C., & Chu, E. (2002). Identification of critical amino acid residues on human dihydrofolate reductase protein that mediate RNA recognition. *Nucleic Acids Research*, *30*(20), 4481–4488.

Tai, N., Schmitz, J. C., Liu, J., Lin, X., Bailly, M., Chen, T., & Chu, E. (2004). Translational autoregulation of thymidylate synthase and dihydrofolate reductase. *Frontiers in Bioscience: A Journal and Virtual Library*, *9*, 2521–2526. https://doi.org/10.2741/1413

Tammsalu, T., Matic, I., Jaffray, E. G., Ibrahim, A. F. M., Tatham, M. H., & Hay, R. T. (2014). Proteome-wide identification of SUMO2 modification sites. *Science Signaling*, *7*(323), rs2. https://doi.org/10.1126/scisignal.2005146

Taylor, R. T., & Hanna, M. L. (1982). Folate-dependent enzymes in cultured Chinese hamster ovary cells: Impaired mitochondrial serine hydroxymethyltransferase activity in two additional glycine--auxotroph complementation classes. *Archives of Biochemistry and Biophysics*, *217*(2), 609–623. https://doi.org/10.1016/0003-9861(82)90543-4

Teng, X., Dayhoff-Brannigan, M., Cheng, W.-C., Gilbert, C. E., Sing, C. N., Diny, N. L., Wheelan, S. J., Dunham, M. J., Boeke, J. D., Pineda, F. J., & Hardwick, J. M. (2013). Genome-wide Consequences of Deleting Any Single Gene. *Molecular Cell*, *52*(4), 485–494. https://doi.org/10.1016/j.molcel.2013.09.026

The Problems of Using Plasmids for CRISPR Genome Editing. (n.d.). Retrieved 17 July 2019, from https://www.synthego.com/blog/crispr-plasmid-pitfalls

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

Titus, S. A., & Moran, R. G. (2000). Retrovirally mediated complementation of the glyB phenotype. Cloning of a human gene encoding the carrier for entry of folates into mitochondria. *The Journal of Biological Chemistry*, *275*(47), 36811–36817. https://doi.org/10.1074/jbc.M005163200

Tomczak, A., Mortensen, J. M., Winnenburg, R., Liu, C., Alessi, D. T., Swamy, V., Vallania, F., Lofgren, S., Haynes, W., Shah, N. H., Musen, M. A., & Khatri, P. (2018).

Interpretation of biological experiments changes with evolution of the Gene Ontology and its annotations. *Scientific Reports*, *8*(1), 5115. https://doi.org/10.1038/s41598-018-23395-2

Torlakovic, E. E., Nielsen, S., Vyberg, M., & Taylor, C. R. (2015). Getting controls under control: The time is now for immunohistochemistry. *Journal of Clinical Pathology*, *68*(11), 879–882. https://doi.org/10.1136/jclinpath-2014-202705

Torres, R. J. (2013). Lesch–Nyhan Syndrome. In S. Maloy & K. Hughes (Eds.), *Brenner's Encyclopedia of Genetics (Second Edition)* (pp. 219–222). Academic Press. https://doi.org/10.1016/B978-0-12-374984-0.00856-1

Treuren, T. V., & Vishwanatha, J. K. (2018). CRISPR deletion of MIEN1 in breast cancer cells. *PLOS ONE*, *13*(10), e0204976. https://doi.org/10.1371/journal.pone.0204976

Trimmer, E. E., Ballou, D. P., Ludwig, M. L., & Matthews, R. G. (2001). Folate activation and catalysis in methylenetetrahydrofolate reductase from Escherichia coli: Roles for aspartate 120 and glutamate 28. *Biochemistry*, *40*(21), 6216–6226. https://doi.org/10.1021/bi002790v

Trinklein, N. D., Aldred, S. F., Hartman, S. J., Schroeder, D. I., Otillar, R. P., & Myers, R. M. (2004). An abundance of bidirectional promoters in the human genome. *Genome Research*, *14*(1), 62–66. https://doi.org/10.1101/gr.1982804

Tsang, B. L., Devine, O. J., Cordero, A. M., Marchetta, C. M., Mulinare, J., Mersereau, P., Guo, J., Qi, Y. P., Berry, R. J., Rosenthal, J., Crider, K. S., & Hamner, H. C. (2015). Assessing the association between the methylenetetrahydrofolate reductase (MTHFR) 677C>T polymorphism and blood folate concentrations: A systematic review and meta-analysis of trials and observational studies. *The American Journal of Clinical Nutrition*, *101*(6), 1286–1294. https://doi.org/10.3945/ajcn.114.099994

Tsuchida, C. A., Zhang, S., Doost, M. S., Zhao, Y., Wang, J., O'Brien, E., Fang, H., Li, C.-P., Li, D., Hai, Z.-Y., Chuck, J., Brötzmann, J., Vartoumian, A., Burstein, D., Chen, X.-W., Nogales, E., Doudna, J. A., & Liu, J.-J. G. (2022). Chimeric CRISPR-CasX enzymes and guide RNAs for improved genome editing activity. *Molecular Cell*, *82*(6), 1199-1209.e6. https://doi.org/10.1016/j.molcel.2022.02.002

Urlaub, G., & Chasin, L. A. (1980). Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. *Proceedings of the National Academy of Sciences of the United States of America*, *77*(7), 4216–4220.

Urlaub, G., Käs, E., Carothers, A. M., & Chasin, L. A. (1983). Deletion of the diploid dihydrofolate reductase locus from cultured mammalian cells. *Cell*, *33*(2), 405–412. https://doi.org/10.1016/0092-8674(83)90422-1

Vakulskas, C. A., Dever, D. P., Rettig, G. R., Turk, R., Jacobi, A. M., Collingwood, M. A., Bode, N. M., McNeill, M. S., Yan, S., Camarena, J., Lee, C. M., Park, S. H., Wiebking,

V., Bak, R. O., Gomez-Ospina, N., Pavel-Dinu, M., Sun, W., Bao, G., Porteus, M. H., & Behlke, M. A. (2018). A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nature Medicine*, *24*(8), 1216–1224. https://doi.org/10.1038/s41591-018-0137-0

Van Winkle, L. J., & Ryznar, R. (2019). One-Carbon Metabolism Regulates Embryonic Stem Cell Fate Through Epigenetic DNA and Histone Modifications: Implications for Transgenerational Metabolic Disorders in Adults. *Frontiers in Cell and Developmental Biology*, *7*. https://www.frontiersin.org/article/10.3389/fcell.2019.00300

Vandivier, L. E., Anderson, S. J., Foley, S. W., & Gregory, B. D. (2016). The Conservation and Function of RNA Secondary Structure in Plants. *Annual Review of Plant Biology*, *67*, 463–488. https://doi.org/10.1146/annurev-arplant-043015-111754

Vergis, J. M., Bulock, K. G., Fleming, K. G., & Beardsley, G. P. (2001). Human 5-aminoimidazole-4-carboxamide ribonucleotide transformylase/inosine 5'-monophosphate cyclohydrolase. A bifunctional protein requiring dimerization for transformylase activity but not for cyclohydrolase activity. *The Journal of Biological Chemistry*, *276*(11), 7727–7733. https://doi.org/10.1074/jbc.M009940200

Visentin, M., Zhao, R., & Goldman, I. D. (2012). The Antifolates. *Hematology/Oncology Clinics of North America*, *26*(3), 629–ix. https://doi.org/10.1016/j.hoc.2012.02.002

Viswanathan, M., Treiman, K. A., Kish-Doto, J., Middleton, J. C., Coker-Schwimmer, E. J. L., & Nicholson, W. K. (2017). Folic Acid Supplementation for the Prevention of Neural Tube Defects: An Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA*, *317*(2), 190. https://doi.org/10.1001/jama.2016.19193

Voth, H., Oberthuer, A., Simon, T., Kahlert, Y., Berthold, F., & Fischer, M. (2009). Co-regulated expression of HAND2 and DEIN by a bidirectional promoter with asymmetrical activity in neuroblastoma. *BMC Molecular Biology*, *10*(1), 28. https://doi.org/10.1186/1471-2199-10-28

Wagner, C. (2001). Biochemical Role of Folate in Cellular Metabolism*. *Clinical Research and Regulatory Affairs*, *18*(3), 161–180. https://doi.org/10.1081/CRP-100108171

Wan, Y., Kertesz, M., Spitale, R. C., Segal, E., & Chang, H. Y. (2011). Understanding the transcriptome through RNA structure. *Nature Reviews Genetics*, *12*(9), 641–655. https://doi.org/10.1038/nrg3049

Wang, H., & Xu, X. (2017). Microhomology-mediated end joining: New players join the team. *Cell & Bioscience*, *7*(1), 6. https://doi.org/10.1186/s13578-017-0136-8

Wang, J., Rajbhandari, P., Damianov, A., Han, A., Sallam, T., Waki, H., Villanueva, C. J., Lee, S. D., Nielsen, R., Mandrup, S., Reue, K., Young, S. G., Whitelegge, J., Saez, E., Black, D. L., & Tontonoz, P. (2017). RNA-binding protein PSPC1 promotes the differentiation-dependent nuclear export of adipocyte RNAs. *The Journal of Clinical Investigation*, *127*(3), 987–1004. https://doi.org/10.1172/JCI89484

Wang, J. Z., Ghergurovich, J. M., Yang, L., & Rabinowitz, J. D. (2020). *Methionine synthase supports tumor tetrahydrofolate pools* (p. 2020.09.05.284521). bioRxiv. https://doi.org/10.1101/2020.09.05.284521

Wang, T., Birsoy, K., Hughes, N. W., Krupczak, K. M., Post, Y., Wei, J. J., Lander, E. S., & Sabatini, D. M. (2015). Identification and characterization of essential genes in the human genome. *Science (New York, N.Y.)*, *350*(6264), 1096–1101. https://doi.org/10.1126/science.aac7041

Wang, X.-W., Liu, C.-X., Chen, L.-L., & Zhang, Q. C. (2021). RNA structure probing uncovers RNA structure-dependent biological functions. *Nature Chemical Biology*, *17*(7), 755–766. https://doi.org/10.1038/s41589-021-00805-7

Wei, W., Pelechano, V., Järvelin, A. I., & Steinmetz, L. M. (2011). Functional consequences of bidirectional promoters. *Trends in Genetics: TIG*, *27*(7), 267–276. https://doi.org/10.1016/j.tig.2011.04.002

Wills, L. (1931). TREATMENT OF "PERNICIOUS ANAEMIA OF PREGNANCY" AND "TROPICAL ANAEMIA". *British Medical Journal*, *1*(3676), 1059–1064.

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. *The Journal of Biological Chemistry*, *282*(24), 17623–17631. https://doi.org/10.1074/jbc.M702526200

Wong, T. N., & Pan, T. (2009). RNA folding during transcription: Protocols and studies. *Methods in Enzymology*, *468*, 167–193. https://doi.org/10.1016/S0076-6879(09)68009-5

Wright, A. J. A., Dainty, J. R., & Finglas, P. M. (2007). Folic acid metabolism in human subjects revisited: Potential implications for proposed mandatory folic acid fortification in the UK. *British Journal of Nutrition*, *98*(4), 667–675. https://doi.org/10.1017/S0007114507777140

Wright, A. V., Nuñez, J. K., & Doudna, J. A. (2016). Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell*, *164*(1), 29–44. https://doi.org/10.1016/j.cell.2015.12.035

Xiu, Y., & Field, M. S. (2020). The Roles of Mitochondrial Folate Metabolism in Supporting Mitochondrial DNA Synthesis, Oxidative Phosphorylation, and Cellular Function. *Current Developments in Nutrition*, *4*(10), nzaa153. https://doi.org/10.1093/cdn/nzaa153

Xu, C.-F., Chen, G.-J., Luo, Y.-L., Zhang, Y., Zhao, G., Lu, Z.-D., Czarna, A., Gu, Z., & Wang, J. (2021). Rational designs of in vivo CRISPR-Cas delivery systems. *Advanced Drug Delivery Reviews*, *168*, 3–29. https://doi.org/10.1016/j.addr.2019.11.005

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

Yadav, S., Longkumer, I., Joshi, S., & Saraswathy, K. N. (2021). Methylenetetrahydrofolate reductase gene polymorphism, global DNA methylation and blood pressure: A population based study from North India. *BMC Medical Genomics*, *14*(1), 59. https://doi.org/10.1186/s12920-021-00895-1

Yamaoka, T., Kondo, M., Honda, S., Iwahana, H., Moritani, M., Ii, S., Yoshimoto, K., & Itakura, M. (1997). Amidophosphoribosyltransferase limits the rate of cell growth-linked de novo purine biosynthesis in the presence of constant capacity of salvage purine biosynthesis. *The Journal of Biological Chemistry*, *272*(28), 17719–17725. https://doi.org/10.1074/jbc.272.28.17719

Yamaoka, T., Yano, M., Kondo, M., Sasaki, H., Hino, S., Katashima, R., Moritani, M., & Itakura, M. (2001). Feedback inhibition of amidophosphoribosyltransferase regulates the rate of cell growth via purine nucleotide, DNA, and protein syntheses. *The Journal of Biological Chemistry*, *276*(24), 21285–21291. https://doi.org/10.1074/jbc.M011103200

Yamazaki, T., & Hirose, T. (2015). The building process of the functional paraspeckle with long non-coding RNAs. *Frontiers in Bioscience (Elite Edition)*, *7*(1), 1–41. https://doi.org/10.2741/715

Yan, S., Wen, J.-D., Bustamante, C., & Tinoco, I. (2015). Ribosome Excursions during mRNA Translocation Mediate Broad Branching of Frameshift Pathways. *Cell*, *160*(5), 870–881. https://doi.org/10.1016/j.cell.2015.02.003

Yang, H., Ren, S., Yu, S., Pan, H., Li, T., Ge, S., Zhang, J., & Xia, N. (2020). Methods Favoring Homology-Directed Repair Choice in Response to CRISPR/Cas9 Induced-Double Strand Breaks. *International Journal of Molecular Sciences*, *21*(18), E6461. https://doi.org/10.3390/ijms21186461

Yook, K. (2005). Complementation. *WormBook: The Online Review of C. Elegans Biology*, 1–17. https://doi.org/10.1895/wormbook.1.24.1

Zhang, J.-P., Li, X.-L., Li, G.-H., Chen, W., Arakaki, C., Botimer, G. D., Baylink, D., Zhang, L., Wen, W., Fu, Y.-W., Xu, J., Chun, N., Yuan, W., Cheng, T., & Zhang, X.-B. (2017). Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biology*, *18*(1), 35. https://doi.org/10.1186/s13059-017-1164-8

Zhang, X., Wang, W., Zhu, W., Dong, J., Cheng, Y., Yin, Z., & Shen, F. (2019). Mechanisms and Functions of Long Non-Coding RNAs at Multiple Regulatory Levels. *International Journal of Molecular Sciences*, *20*(22), 5573. https://doi.org/10.3390/ijms20225573

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

Zhao, B. S., Roundtree, I. A., & He, C. (2017). Post-transcriptional gene regulation by mRNA modifications. *Nature Reviews. Molecular Cell Biology*, *18*(1), 31–42. https://doi.org/10.1038/nrm.2016.132

Zhao, R., Diop-Bove, N., Visentin, M., & Goldman, I. D. (2011). Mechanisms of Membrane Transport of Folates into Cells and Across Epithelia. *Annual Review of Nutrition*, *31*, 10.1146/annurev-nutr-072610–145133. https://doi.org/10.1146/annurev-nutr-072610-145133

Zheng, Y., Lin, T.-Y., Lee, G., Paddock, M. N., Momb, J., Cheng, Z., Li, Q., Fei, D. L., Stein, B. D., Ramsamooj, S., Zhang, G., Blenis, J., & Cantley, L. C. (2018). Mitochondrial One-Carbon Pathway Supports Cytosolic Folate Integrity in Cancer Cells. *Cell*, *175*(6), 1546-1560.e17. https://doi.org/10.1016/j.cell.2018.09.041