Investigation of Gene Specific DNA Methylation and Folate Polymorphisms and their Relevance for Human Health and Disease

Thesis Submission for the Degree of

Doctor of Philosophy

by

Mari Ozaki M.Sc.

Under the supervision of

Dr. Anne Parle-McDermott

School of Biotechnology

Dublin City University

October 2014

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Ph.D. 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.: _____ Date: _____

Table of Contents

Acknowledgements

I would like to take this opportunity to thank all those who have helped, inspired and guided me throughout the course of my PhD. Firstly, I would like to thank my supervisor Dr. Anne Parle-McDermott for accepting me into her lab and giving me this wonderful opportunity. My deepest gratitude goes out to her for her unwavering guidance, motivation and immense help throughout the four years. I am also extremely grateful to everyone in the Parle-McDermott lab, both past and present, Dr. Linda Hughes, Dr. Grainne McEntee, Dr. Stefano Minguzzi, Alan Harrison and Aoife MacCooey for all their assistance, advice and friendship throughout my four years. Special thanks to my collaborators Prof. Anne Molly, Dr. James Mills, Dr. Lawrence Brody, Dr. Faith Pangilinan, Prof. Helene McNulty, Dr. Kristina Pentieva, Dr. Paula Meleady and Mr. Michael Henry for helping me and advising me with various aspects of my project. I am also extremely grateful to the Health Research Board for funding my studies, without whom this PhD would not have been possible. Finally, a massive thank you to all my family and friends for your continued love, support and encouragement, I could not have got through all the highs and lows without you.

Abbreviations

ATM	Ataxia Telangiectasia Mutated
ATP	Adenosine Triphosphate
ATP5F1	ATP synthase H^+ transporting, mitochondrial F_0 complex, subunit B1
Att	Attachment
A^{vy}	Agouti / viable yellow
CHD	Congenital Heart Defects
CHO	Chinese Hamster Ovary
Cp	Crossing Point
CVD	Cardiovascular Disease
DFE	Dietary Folate Equivalents
DHF	Dihydrofolate
DHFR	Dihydrofolate Reductase
DHFRL1	Dihydrofolate Reductase Like 1
DMR	DNA Methylated Region
DNMT	DNA Methyltransferase
DSB	Double-Strand Breaks
dTMP	Thymidylate
dUMP	Deoxyuridylate
EIF2C3	Eukaryotic Translation Initiation Factor 2C, 3
FASSTT	Folic Acid Supplementation during the Second and Third Trimester
FBS	Fetal Bovine Serum
FPGS	Folypoly-γ-Glutamate Synthase
FS-DMRs	Folate Sensitive –DNA Methylation Region
GEEN	Genome Editing with Engineered Nucleases
GST	Glutathione S Transferase
Hcy	Homocysteine
HEK	Human Embryonic Kidney
HEP	Human Epigenome Project
HR	Homologous Recombination
IAP	Intracisternal A Particle
IPA	Ingenuity Pathway Analysis
IRS	Insulin Resistance Syndrome
LB	Luria Bertani
LC-MS	Liquid Chromatography – Mass Spectrometry
MS-HRM	Methylation Sensitive-High Resolution Melting
MTX	Methotrexate
MVP	Methylation Variable Positions
NHEJ	Non-Homologous End-Joining
NIDDM	Non-Insulin-Dependent Diabetes Mellitus
NTD	Neural Tube Defect
RASSF1A	Ras Association Domain Family Member 1 Alpha
RBC	Red Blood Cell
RCF	Red Cell Folate
RDA	Recommended Dietary Allowance
RPMI	Roswell Park Memorial Institute
RT-qPCR	Real Time – quantitative Polymerase Chain Reaction
SAM	S-Adenosyl Methionine
SDM	Site-Directed Mutagenesis
SHMT	Serine Hydroxymethyltransferase
SMART-MSP	Sensitive Melting Analysis after Real Time-Methylation Specific PCR

SNP	Single Nucleotide Polymorphism
SUMO	Small Ubiquitin-Like Modifier
Ta	Annealing Temperature
TDT	Transmission Disequilibrium Test
TET1	Tet Methylcytosine Dioxygenase 1
TEV	Tobacco Etch Virus
tHcy	Total Plasma Homocysteine
THF	Tetrahydrofolate
Tm	Melting Temperature
TSS	Trinity Student Study
TYMS	Thymidylate Synthase
UTR	Untranslated Region
WT	Wild Type
5hmC	5-Hydroxymethylcytosine
5mC	5-Methylcytosine

Abstract

Folate mediated one-carbon metabolism is a complex biochemical pathway that is essential to the cell. Therefore, any impairments or deficiencies can have serious health consequences. Examination of the relationship between folate status and DNA methylation in pregnant women showed differential methylation patterns between individuals in the *DHFR* gene, however, the differences observed was not associated with folate status. The 19 bp deletion / insertion polymorphism (DIP) within the *DHFR* gene, an enzyme that catalyses the reduction of dihydrofolate to tetrahydrofolate has been studied extensively in association with both neural tube defects and folate metabolites but the results are controversial. Screening the *DHFR* 19 bp DIP in the largest cohort yet to address this issue and assessment of the DIP with folate metabolites and folate status showed no association between the DIP and the folate metabolites or folate status. DHFRL1 is a newly discovered reductase enzyme that participates in thymidylate biosynthesis in the mitochondria. Recombinant and endogenous DHFRL1 polymorphisms examined for functionality showed the polymorphisms to be capable of altering enzyme activity and to be associated with neural tube defects. By altering DHFRL1 expression numerous genes have been shown to be affected. The findings above have demonstrated the importance of the genes involved in folate metabolism and their relevance for human health and disease.

Chapter 1

Introduction

1.1 Overview

Nutritional genetics is a field of science, whereby the relationship between the human genome, nutrition and health is examined. Research fields mainly focus on how different foods can affect our health by altering our genes and how genetic differences between individuals can affect the way we respond to the different nutrients we eat, helping us to determine what genetic differences may play a role in disease risk. The Parle-Mc Dermott lab focuses on a nutrient called folate, a nutrient that is essential for numerous cellular functions and one in which, deficiencies have been associated with a broad range of clinical diseases (discussed in Section 1.5). At present, genes / proteins involved in the folate mediated one-carbon metabolism are considered to be the prime candidates for disease association.

My thesis focuses on exploring the molecular mechanism of the role of folate in human health and disease with particular focus on folate-sensitive DNA methylation regions, genetic polymorphisms within genes involved in the folate mediated one-carbon metabolism pathway and the effects of altered DHFRL1 expression, a newly discovered gene involved in folate metabolism (Anderson et al., 2011, McEntee et al., 2011) and their association with disease.

This chapter starts off with a general overview of folate, the folate mediated one-carbon metabolism pathway and its association with disease, followed by an overview on the impact of nutrition on the epigenome, with particular focus on the relationship between folate and DNA methylation, continued by a brief introduction on the *DHFR* gene family and concludes with the aims and objectives of my thesis.

1.2 Folate

Folate is an essential water soluble B_9 vitamin that is involved in numerous important metabolic pathways such as DNA replication, DNA repair, DNA methylation and nucleotide biosynthesis (Iyer and Tomar, 2009). Folate cannot be synthesised by mammalian cells and therefore, it must be obtained through the diet. Folate was first isolated from spinach leaves by Mitchell *et al.* in 1941. It can be found naturally in green leafy vegetables such as spinach and turnip greens, citrus fruits, dried beans and peas (Rampersaud et al., 2003, Ifergan and Assaraf, 2008, Ohrvik and Witthoft, 2011). Folic acid is the synthetic, oxidised, chemically stable form of folate and is found in supplements and fortified foods such as bread and breakfast cereals. The Recommended Dietary Allowance (RDA), the average daily intake that is recommended to meet the nutrient requirement of healthy individuals as approved by the European Union is 400 µg/d of dietary folate equivalents

(DFE) in adults (McPartlin et al., 1993, **(WHO)**, 2002, Food and Nutrition Boards, 2004), and 600 µg/d DFE for women who are planning a pregnancy (www.ods.od.nih.gov). However, as these requirements were not being met, widespread fortification of many foods has occurred and currently, there are 80 countries worldwide that have undertaken mandatory fortification of their flour with folic acid (Food Fortification Initiative http://www.ffinetwork.org/). Folate is essential in the body and must be converted from dihydrofolate to its biologically active form, tetrahydrofolate (THF), before it can be employed in the folate mediated one-carbon metabolism pathway (Fox and Stover, 2008) (Figure 1.1).

1.3 Folate Metabolism

Folate mediated one-carbon metabolism is a complex biochemical pathway that is essential to the cell and can be divided into two main cycles; the DNA replication and repair cycle and the methylation cycle (Figure 1.1) (Fox and Stover, 2008). Both these cycles require THFs. THFs are water-soluble B-vitamins and are made up of three different components, a fully reduced pterin ring, a p-aminobenzoyl group and a poly-γ-glutamate tail (Figure 1.2) (Shane, 1995). THFs are the active form of folates and can be absorbed directly into the small intestine. However, folate found in natural foods must be hydrolysed into monoglutamate forms before they can be absorbed. This is carried out by an enzyme called gamma-glutamyl hydrolase / human conjugase (Beaudin and Stover, 2009). Once hydrolysed, these monoglutamates can be transported across the intestinal epithelium by proton-coupled folate transporters, by saturable pH-dependent processes or via a non-saturable passive diffusion process. Once inside, the monoglutamates as well as folic acid, which can be absorbed directly into the intestine, can be reduced to their biologically active form, THF by an enzyme called dihydrofolate reductase (DHFR) (Iyer and Tomar, 2009, Stover and Field, 2011).

Folates circulate the body as monoglutamate derivatives and enter the cells via reduced folate carrier proteins (Sirotnak and Tolner, 1999), through endocytosis by the folate receptors (Antony, 1996) or by passive diffusion (Suh et al., 2001). Once folates have been imported into the cell, they are converted into polyglutamate derivatives by the addition of a γ-linked polyglutamate peptide, ranging from two to nine glutamate residues by an enzyme called folypoly-γ-glutamate synthase (FPGS) to retain the folates in the cell (Zhao et al., 2009).

1.4 Compartmentalisation of Folate Mediated One-Carbon Metabolism

Folate polyglutamates have been found to be present in all subcellular organelles in all cell types, where they are involved in activating one-carbon units. Folate mediated one-carbon

metabolism is highly compartmentalised between the cytoplasm, mitochondria and nucleus (Figure 1.1) in eukaryotes, with the majority of folate polyglutamates been shown to be circulated between the cytoplasm and the mitochondria, with low levels also seen in the nucleus (Stover and Field, 2011). However, the type of folate polyglutamates found in the cytoplasm and the mitochondria are different. In the cytoplasm, the majority of folate is found in the form of 5-methyl-tetrahydrofolate, whereas in the mitochondria, the majority of folate is found in the form of 10-formyl-tetrahydrofolate (Tibbetts and Appling, 2010).

In the cytoplasm, folate mediated one-carbon metabolism can be divided into three distinct pathways, the *de novo* purine biosynthesis pathway, the *de novo* thymidylate (dTMP) biosynthesis pathway and the homocysteine re-methylation pathway (Figure 1.1). The three pathways are interlocked and compete with each other for folate cofactors. In the mitochondria, folate metabolism is required to generate formate, which is utilised in cytoplasmic folate mediated one-carbon metabolism (Stover and Field, 2011), for the interconversion of glycine and serine (Fox and Stover, 2008) and for thymidylate biosynthesis (Anderson et al., 2011), and in the nucleus, it is required to synthesise thymidylate from uridylate and serine (Stover and Field, 2011, Anderson et al., 2012a).

The methylation cycle is responsible for the regular supply of methionine to the cell by homocysteine re-methylation. In the presence of adenosine triphosphate (ATP), methionine is adenosylated to S-adenosyl methionine (SAM), a universal methyl donor, which donates methyl groups for the majority of the methylation reactions within the cell including those required for the correct regulation of epigenetic gene activity (Miranda and Jones, 2007).

The *de novo* dTMP biosynthesis is responsible for producing thymidylate from the conversion of deoxyuridylates (dUMPs) to dTMPs. This reaction is catalysed by three enzymes, thymidylate synthase (TYMS), serine hydroxymethyltransferase (SHMT) and dihydrofolate reductase (DHFR) (Stover and Field, 2011). Until recently, it had been understood that all dTMP biosynthesis took place in the cytoplasm and the nucleus. However, in 2011, Anderson *et al*. showed that a *de novo* thymidylate biosynthesis pathway exists in the mitochondria, with the newly discovered gene, *DHFRL1* (Anderson et al., 2011, McEntee et al., 2011) taking the place of *DHFR*.

Recent findings however, have shown that in the nucleus, localisation of TYMS, SHMT and DHFR to the nucleus only occurs during the S and G_2/M phase of the cell cycle, indicating that

the *de novo* thymidylate synthesis pathway translocates to the nucleus for DNA replication and repair (Anderson et al., 2012a). These enzymes are transported to the nucleus from the cytoplasm by a post-translational modification known as sumoylation, which adds a small ubiquitin-like modifier (SUMO) to the enzymes (Anderson et al., 2007, Woeller et al., 2007).

Although folate mediated one-carbon metabolism is compartmentalised, the compartments are connected through the predominantly unidirectional transport of one-carbon donors such as formate, serine and glycine, which are transported across the mitochondrial membrane into the cytoplasm, i.e. the majority of the one-carbon units that are required for reactions that take place in the cytoplasm are derived from the mitochondria (Tibbetts and Appling, 2010).

1.5 Folate and its Association with Disease Risk

Folate mediated one-carbon metabolism is essential to the cell, and any impairment or deficiency can have serious health consequences. Impairments in folate mediated one-carbon metabolism can be caused by a number of different factors such as reduced folate status resulting from malnutrition, low intake of folate-containing foods, malabsorption, certain drugs or severe alcoholism, polymorphisms in genes that are involved in folate metabolism, and deficiencies in secondary micronutrients such as iron and other B vitamins that can alter folate status (Stover and Garza, 2002). Folate deficiency has been associated with a broad range of clinical diseases, ranging from sub-optimal folate deficiencies such as neural tube defects and cancer as well as severe folate deficiencies such as megaloblastic anaemia.

1.5.1 Sub-optimal folate deficiencies

Sub-optimal folate deficiencies are common in the general population and contribute to numerous multi-factorial diseases. Those who have suboptimal folate deficiencies are not folate deficient but have folate levels that are slightly below the average.

1.5.1.1 Folate and Neural Tube Defects

The most commonly known disease associated with folate deficiency are neural tube defects (NTDs) (Scott et al., 1994). Neural tube defects are the most frequent congenital abnormality of the central nervous system. They are birth defects of the brain and spinal cord and occur when the neural tube, the embryonic precursor of the brain and spinal cord, fails to close on around day 28 of pregnancy (van der Put et al., 2001). Although several environmental and genetic factors are thought to contribute to the development of NTDs, numerous intervention studies have

shown that folate status is a major determinant of NTD risk, with peri-conceptional folic acid supplementation demonstrated to reduce / prevent the risk of NTDs in the majority of pregnancies (1991, Czeizel and Dudás, 1992, Berry et al., 1999). However, the exact mechanism underlying how folate reduces the risk of NTDs has not yet been established.

As pregnancy is a time associated with the growth of an embryo, increased nucleotide synthesis is required for cell division and DNA replication to occur. Both of these are one-carbon transfer reactions, and therefore, a considerable increase in folate is required during pregnancy in order to ensure a sufficient supply. With this in mind, Barber *et al*. (1999) hypothesised that if neuroepithelial cells do not have an adequate internal supply of nucleotides, cellular replication will slow down and the development of the neural folds will be retarded.

Folates are also involved in the methylation pathway, and therefore, a disruption in the methylation pathway has also been suggested to be responsible for the relation between folate and NTDs. Studies in mouse models have shown that by disturbing the methylation cycle in mouse embryos by cycloleucine or methionine administration during cranial neurulation resulted in an increase prevalence of NTDs (Dunlevy et al., 2006a, Dunlevy et al., 2006b). Chick embryo models treated with methylation cycle inhibitors also showed a delay in the closure of the neural tube (Afman et al., 2005). However, the cellular, molecular or genetic mechanisms by which methylation interferes with neurulation are still largely unknown but recent work from the laboratory identified a range of genes that are differentially expressed in the presence of a methylation inhibitor compared to control cells (Carroll et al., 2012).

Polymorphisms in genes such as *MTHFR*, *MTHFD1*, *MTHFD1L* involved in the folate-mediated one carbon metabolism pathway have also been associated with NTD risk (van der Put et al., 1995, Botto and Yang, 2000, Brody et al., 2002, Parle-McDermott et al., 2006, Parle-McDermott et al., 2009). The *MTHFR* 677 C>T polymorphism is one of the most widely studied polymorphisms that is associated with NTDs. The *MTHFR* 677 C>T SNP causes an alanine to valine amino acid substitution, resulting in an enzyme that is ~50-60 % less active in individuals who are homozygous TT for the SNP (Yamada et al., 2001). Studies have also shown an elevated risk of ~50-70 % for maternal TT genotype and an 80-90 % for fetal genotype (Whitehead et al., 1995, Shields et al., 1999, Blom et al., 2006). *MTHFD1* R653Q polymorphism is another polymorphism that has been strongly associated with maternal risk of NTDs. Although the functional effect of this polymorphism has not yet been shown, it is thought to increase the risk of

NTDs by altering metabolic outcomes as no affect in folate or homocysteine levels have been reported, i.e. it may be the result of a negative effect on the *de novo* nucleotide biosynthesis pathway (Parle-McDermott et al., 2006). Therefore, although the exact mechanism underlying how folate reduces the risk of NTDs has not yet been established, evidence to date suggests that NTDs are a multi-factorial disease, caused by the interaction of gene polymorphisms, nutrients and folate enzymes (Moyers and Bailey, 2001, De Marco et al., 2011).

1.5.1.2 Folate and Congenital Heart Defects

Congenital heart defects (CHD) are one of the most common types of birth defects and are caused by a defect in the structure of the heart and / or vessels, which are present at birth. Numerous types of defects can occur, most of which either obstruct blood flow in the heart or vessels near it, or cause blood to flow through the heart in an abnormal pattern. The heart's rhythm can also be affected. Numerous studies have shown that folic acid supplementation significantly reduced CHD, most obviously in ventricular septal defects and conotruncal defects (Czeizel and Dudás, 1992, Czeizel et al., 1996, Botto et al., 2003, Czeizel et al., 2004). Although it is not known how folate reduces the risk of CHD, folic acid antagonist drugs that target and inhibit DHFR, an enzyme required for DNA synthesis (Figure 1.1) increased the risk of CHD in the children of pregnant women (Czeizel et al., 2013) suggesting that an inadequate supply of nucleotides may be a cause of CHD. An association between high levels of plasma homocysteine caused by a polymorphism in the *MTHFR* gene and the risk of CHD was also found. Meta-analysis studies showed that both infant and maternal *MTHFR* C677T polymorphisms may contribute to the risk of CHD (van Beynum et al., 2006).

1.5.1.3 Folate and Cancer

The association between folic acid supplementation and cancer is highly controversial, with some studies suggesting that folic acid supplementation can increase the risk of cancer, while others suggesting that it has a protective role. High folic acid concentrations have been associated with certain types of cancer such as prostate, lung and ovarian cancers (Stolzenberg-Solomon et al., 2006). On the other hand, an association between folate deficiency and the development of certain cancers such as colon, breast, ovary, pancreas, brain, lung and cervical cancer (McNulty and Scott, 2008, Duthie, 2011) as well as cognitive defects (Blount et al., 1997) has been shown. The most convincing data are for colorectal cancers with numerous epidemiologic studies suggesting that high folate intake may reduce the risk of colorectal cancer (Giovannucci et al., 1998, Choi and Mason, 2002, Leahy et al., 2005, Van Guelpen et al., 2006, Pompei et al., 2007), although one

intervention trial did show that folic acid supplementation increases the risk of colorectal cancer (Cole et al., 2007). Recent meta-analysis study carried out on randomised trials have also found that overall folic acid supplementation has no significant effects on total cancer incidence, colorectal cancer, gastrointestinal cancer, prostate cancer, genitourinary cancer, lung cancer, breast cancer, haematological malignancy and total cancer mortality, although a significantly reduced risk was seen for melanoma (Qin et al., 2013). The evidence above suggests folate may have a dual function as it seems to both increase the risk in some cases, while having a protective role in others. Animal studies that have been carried out also support this theory. They suggest that the effect of folate consumption on carcinogenesis may depend on the dose and the timing of folate exposure, with folate deficiency enhancing the risk of carcinogenesis in normal healthy cells, while folate supplementation enhancing the progression of existing tumour cells (Kim, 2003, Ulrich and Potter, 2007).

Folate deficiency could influence cancer risk by numerous mechanisms. Severe folate deficiency can provoke widespread chromosome damage, fragile site expression, micronucleus formation and increased uracil levels in bone marrow cell DNA. When folate is deficient, it has been hypothesised that the availability of 5-10 methylenetetrahydrofolate decreases, reducing thymidylate synthesis, causing DNA polymerase-mediated uracil misincorporation, resulting in DNA strand breaks and chromosome instability (Blount et al., 1997, Giovannucci, 2002). Folate deficiency could also result in aberrant DNA methylation patterns, resulting in the altered expression of critical proto-oncogenes and tumour suppressor genes (Choi and Mason, 2000). Polymorphisms in folate related genes such as *MTHFR* (Izmirli, 2013) and *DHFR* (Xu et al., 2007, Gemmati et al., 2009) may also influence cancer risk.

1.5.1.4 Folate and Neurological Diseases

Folate deficiency has also recently been associated with a number of neurological diseases such as Alzheimer's disease (Tchantchou and Shea, 2008), Parkinson's disease (Mattson and Shea, 2003) and depression (Bottiglieri, 2005). One study showed that depressed and schizophrenic patients receiving 5 mg of methyl-folate per day demonstrate significant improvements in both clinical and social recovery (Godfrey et al., 1990). Although the exact mechanism of how folate reduces the risk of neurological disease is unknown, it is thought that as there is normally an intrinsic balance between folate and homocysteine levels within the cell, when folate is deficient, homocysteine, a cytotoxic sulphur containing amino acid cannot be re-methylated back to methionine, due to reduced levels of SAM, and therefore, homocysteine levels are high. SAM is a

key player in mood enhancement and therefore, folic acid supplementation is thought to reduce depression (Iyer and Tomar, 2009). Reduced levels of SAM also results in an increase in homocysteine levels. This increase in homocysteine levels can bring about abnormalities in neural cell proliferation, which have been related to neurodegenerative conditions (Tchantchou and Shea, 2008). Plasma homocysteine levels have also been shown to increase in Alzheimer's disease patients, with low levels of folate and SAM observed in the spinal fluid (Mattson and Shea, 2003) An accumulation of amyloid β–peptide is a known contributing factor to the development of Alzheimer's disease and there is some evidence to suggest that increased homocysteine concentrations can lead to increased amyloid β–peptide concentration leading to neurodegeneration (Ho et al., 2001). Parkinson's disease has also been associated with elevated homocysteine levels. In a neuron cell culture model, both folate deprivation and homocysteine made the cells more sensitive to damage and death, speeding up the onset and progression of the disease (Duan et al., 2002). These studies suggest that elevated homocysteine levels are associated with neurological diseases.

1.5.1.5 Folate and Cerebrovascular and cardiovascular disease

Cardiovascular diseases (CVD) are one of the most widespread diseases in the Western population and are disorders of the heart and blood vessels. These include diseases such as heart attacks, stroke and high blood pressure, as well as cerebrovascular diseases which refer to a group of diseases caused by problems in the blood vessels that supply the brain. Numerous factors are thought to contribute to the development of CVD. The associations between folate status and CVD and cerebrovascular disease have been highly controversial. Many studies have argued that inadequate folate intake increases the risk of cerebrovascular and cardiovascular disease, as folate deficiency decreases folate-dependent methylation of homocysteine, resulting in elevated levels of plasma homocysteine, a cytotoxic sulphur containing amino acid (Blount et al., 1997). High levels of homocysteine, however can also be caused by polymorphisms in a number of folate related genes such as MTHFR, where individuals with the TT MTHFR 677 C>T genotype have been shown to exhibit moderate hyperhomocysteinemia (Dwivedi et al., 2011). High levels of homocysteine have been used as an indicator for CVD in the past (Collaboration, 2002). However, recent studies have shown that folic acid supplementation does not reduce the risk of cardiovascular disease in secondary prevention although it does reduce homocysteine levels (Bazzano et al., 2006, Clarke et al., 2012). Nonetheless, there is still no conclusive evidence that links or doesn't link folic status / homocysteine levels with CVD or cerebrovascular disease.

1.5.1.6 Effects of Excess Folic Acid on Health and Disease

Although all of the above diseases seem to be a result of folate deficiency, recent studies have suggested that excess folate / folic acid supplementation can also increase the risk of disease, raising many concerns. As mentioned earlier (Section 1.5.1.2), randomised clinical trials where participants with a recent history of a colorectal adenoma were either given a folic acid supplement or a placebo showed that those who took the supplement had a higher risk of having more adenomas (Cole et al., 2007). However, a recent study has shown that folic acid supplementation has no effect on colorectal cancer (Figueiredo et al., 2011). Similar findings were also seen in the case of prostate cancer (Figueiredo et al., 2009, Vollset et al., 2013). Apart from cancer, folic acid supplementation during pregnancy has also been linked to the increased risk of asthma in children, although results here have also been controversial (Whitrow et al., 2009, Martinussen et al., 2012).

1.5.2 Severe Folate Deficiencies

1.5.2.1 Folate and Megaloblastic Anaemia

Megaloblastic anaemia is a disorder, which is characterised by the presence of megaloblasts, very large red blood cells in which the inner contents of each cell is not entirely developed. Red blood cells have a high turnover rate, and therefore, when folate levels are low, DNA synthesis is reduced, resulting in the prolongation of the synthesis phase of cell division. This causes the red blood cells to become enlarged. When folate is severely deficient, the megaloblasts accumulate in the bone marrow, reducing cell division, as well as the number of white blood cells and platelets resulting in anaemia (Bailey and Gregory, 1999).

From the above, we can see that folic acid supplementation can both increase and decrease the risk of disease. However, the exact mechanisms of how folate increases / reduces the risk of specific diseases is not fully known. Nevertheless, a number of mechanisms such as cell proliferation rates, prevention of uracil misincorporation and alteration to DNA methylation patterns have been proposed, with the relationship between folate and DNA methylation discussed below.

1.6 Epigenetics and DNA Methylation

Epigenetics is the study of mitotically or meiotically heritable changes in gene expression or phenotype that do not result from an alteration in the DNA sequence itself, *i.e*. regulation post-DNA level. Epigenetic modifications have been up to recently grouped into three main categories, DNA methylation, histone modifications and chromatin remodelling that interplay with each other

(Portela and Esteller, 2010). However, in recent times, they have also included non-coding RNA and microRNA gene regulation. These epigenetic modifications can change throughout a person's lifetime and can be greatly influenced by environmental factors such as diet, chemical exposures, smoking, prenatal and parental care, prenatal nutrition, infectious agents and stress (Anderson et al., 2012c).

DNA methylation is the most commonly studied form of epigenetic modification and the one that is the type of modification that may show trans-generational effects following *in utero* exposure. DNA methylation is the addition of a methyl group to a cytosine residue and is mediated by DNA methyltransferases (DNMTs) (Figure 1.3) and usually occurs in CpG islands, i.e. CpG rich regions in the promoter region of a given gene. However, recent studies have shown that methylation can occur in a non-CpG context and were found to be prevalent in embryonic stem cells (Lister et al., 2009b). When DNA is methylated in a region of a gene, it is generally associated with the silencing of gene expression, whereas when DNA is not methylated, it is associated with activation of gene expression. The precise mechanism of how DNA methylation affects gene control has not yet been elucidated but it is thought to involve the interference of methylated cytosine residues with the binding of the RNA polymerase complex and associated transcription factors through an interaction with chromatin (Klose and Bird, 2006) (Figure 1.4). However, more recent studies have shown that DNA methylation is not always associated with transcriptional repression (Ball et al., 2009) and that the relationship with gene control is much more complex than previously thought. Methylation occurring in the bodies of genes has recently been shown to be positively associated with active transcription on the active X chromosome (Hellman and Chess, 2007, Jones, 2012). The exact function of gene body methylation is yet unknown, but it was originally thought to be a mechanism for silencing the transcription of repetitive DNA elements, such as retroviruses, LINE1 elements and Alu elements, while allowing gene transcription to continue. A number of more recent whole-genome studies have however, shown alternate functions for gene body methylation. These studies have shown that exons are more highly methylated than introns, and that methylation differences occur at exon–intron boundaries, suggesting that gene body methylation plays a role in regulating splicing (Jones, 2012).

It has also recently been found that DNA methylation does not always have to occur in the context of a CpG as had been previously thought, but that it can also occur in a non-CpG context. This was first shown by Lister *et al.* in 2009. They examined the DNA methylation patterns of human stem cells and fetal fibroblasts and found that non-CpG methylation was mostly enriched in

the coding regions of genes in stem cells. However, this non-CpG methylation begun to be lost during cellular differentiation and was found to be a specific feature of stem cells. In fibroblasts, reduced methylation levels were seen to be associated with lower transcriptional activity for specific genes, supporting previous research that showed that DNA hypomethylation does not always have to be associated with high transcriptional activity levels (Lister et al., 2009b, Parle-McDermott and Ozaki, 2011).

DNA methylation can be used in a variety of different ways by cells to control gene expression. These range from tissue-specific gene expression to inactivation of imprinted genes, the X-chromosome and transposable elements. DNA methylation is best known for its role in imprinted genes during differentiation where gene expression is permanently switched on or off. This epigenetic mode of regulation is stable and is passed on from parents to offspring (Jirtle and Skinner, 2007). DNA methylation can also be metastable, subject to change in response to environmental factors. No definitive term has been assigned to describe these sites but they are sometimes referred to as epialleles (Dolinoy et al., 2007). Due to these, changes in DNA methylation have been associated with a number of disease states such as imprinted disorders (Paulsen and Ferguson-Smith, 2001), tumourigenesis (Esteller, 2005), neurological disorders (Urdinguio et al., 2009), cardiovascular disease (Kim et al., 2010) and autoimmune diseases (Richardson, 2003).

1.7 The Methylome

DNA methylation patterns have been shown to vary between tissues within individuals but also between individuals. However, the range and variability of these patterns has not yet been established. At present, the Human Epigenome Project (HEP) (Eckhardt et al., 2004) and the NIH Roadmap Epigenomics (http://www.roadmapepigenomics.org/) projects are trying to decipher the variation in DNA methylation patterns by creating a catalogue of the variety of human epigenetic marks including Methylation Variable Positions (MVPs), DNA methylation sites that vary in different tissues, between individuals and in disease states. This will hopefully in the not so distant future, provide us with an insight into human epigenome patterns. Reviewed in (Parle-McDermott and Ozaki, 2011) – See Appendix.

Although 5-methylcytosine (5mC) has always been deemed to be the ‘fifth’ DNA base, a ‘sixth’ base has recently been discovered in the form of 5-hydroxymethylcytosine (5hmC) (Kriaucionis and Heintz, 2009, Tahiliani et al., 2009). 5mC is catalysed to 5hmC by an enzyme known as tet methylcytosine dioxygenase 1 (TET1), a 2-oxoglutarate and Fe(II) dependent enzyme,

which can bind throughout the genome in embryonic stem cells as well as in neurons and the brain and binds particularly at high density CpG sites in the gene promoters and within specific genes (Kriaucionis and Heintz, 2009). 5hmC modifications showed a similar pattern to TET1 binding, which is in contrast to 5mC, which is primarily found in low density CpG sites. However, 5mC and 5hmC are similar in that they both play a role in transcriptional silencing (Williams et al., 2011). 5hmC is also thought to play roles in DNA methylation regulation (Tahiliani et al., 2009), prompting DNA de-methylation (Hackett et al., 2013) as well as in the 'priming' of specific genes for rapid activation (Kriaucionis and Heintz, 2009).

Even though there have been rapid developments made in deciphering the epigenome in recent times, comprehending the relevance of these epigenetic marks and how they differ between tissues and individuals is not yet understood. Many of the studies that have been carried out to date have also only examined the environmental impacts on 5mC although it is likely that 5hmC levels are also influenced and altered by environmental factors, due to the fact that the majority of the DNA methylation techniques available at present are unable to distinguish the difference between 5mC and 5hmC. Therefore, these studies may need to be re-evaluated in order to take the newly identified 'sixth' base into consideration (Parle-McDermott and Ozaki, 2011).

1.8 Fetal Origins of Disease

DNA methylation patterns of an individual are known to change throughout their lifetime. This was first shown by a twin study carried out by Fraga *et al*. who showed how the epigenome of identical twins became divergent as they aged (Fraga et al., 2005). More recently, the evidence showing that the human *in utero* environment can impact on DNA methylation patterns and disease risk of the subsequent offspring has been increasing. This relates back to the "fetal origins of adult disease" theory originally proposed by Barker (Barker et al., 1989). Barker *et al*. initially examined the association between fetal growth impairment with cardiovascular disease (Barker and Martyn, 1992, Barker, 1993, Barker, 1997) and the relationship between birth weight and impaired glucose tolerance (Hales et al., 1991). From these studies, he proposed the 'fetal origins of disease' hypothesis, which states that susceptibility to adulthood cardiovascular disease (CVD), non-insulin-dependent diabetes mellitus (NIDDM), and the insulin resistance syndrome (IRS) is programmed *in utero* in response to fetal under-nutrition.

Different environmental stimuli / constraints such as unbalanced nutrition in early life have also been shown to be associated with an increased risk of many other diseases later in adult life and

have been shown to cause metabolic and structural alterations that result in phenotypic changes in the offspring. These changes and effects caused by environmental stimuli have been examined and documented at length in epidemiological studies (Godfrey and Barker, 2001), in whole animals (Bertram and Hanson, 2001), in specific cells types (Heywood et al., 2004) and in early embryos (Kwong et al., 2000), with a wealth of literature supporting the 'fetal origins of human disease' concept (Barker et al., 1989).

In addition to the above studies, studies such as the Dutch Hunger Winter, have given us an insight into the importance of nutrition status during a critical period of life on long-term effects. The Dutch Hunger Winter was a period during the winter of 1944, when the Nazi's occupied parts of the Netherlands. During this time, a famine was forced upon people who were previously well-nourished. The studies carried out on the Dutch cohort examined the outcomes and consequences of maternal under-nutrition on offspring on a long lasting scale. The studies found that *in utero* exposure to famine was associated with increased risks of obesity (men) (Ravelli et al., 1976), schizophrenia (Hoek et al., 1996), prevalence of obstructive airways disease (Lopuhaä et al., 2000), impaired glucose and lipid homeostasis, coronary heart disease, breast cancer (women) and reduced renal function (Roseboom et al., 2006), 4-5 decades later in life, in a manner that is dependent on gestational age at the time of exposure to famine. The worst effects were seen in the children who were born to mothers who experienced severe under-nutrition during the first trimester of pregnancy, suggesting that the first trimester of pregnancy is a particularly vulnerable period that can have severe consequences of disease later in adult life. With fetal programming for other adulthood diseases being examined subsequent to the original studies, it is now hypothesised that the *in utero* environment has a significant role in an individual's risk of contracting a number of diseases later in their adult life.

However, the molecular mechanism underlying these changes has not yet been identified. On the other hand, the mechanisms proposed in relation to how environmental influences / the *in utero* environment can have an effect on an individual's risk of disease later in life are numerous, with DNA methylation as a mediator of these effects being a popular theory. Despite this, studies that provide direct evidence to support epigenetic modifications as a mechanism for this theory have only recently begun to emerge. These include prenatal exposure to tobacco smoke and exposure to airborne polycyclic aromatic hydrocarbons and risk of childhood asthma. Two recent studies have also shown a strong link between folate availability *in utero* and risk of insulin resistance (Yajnik et

al., 2008) and childhood asthma (Håberg et al., 2009, Whitrow et al., 2009). However, whether and how folate is mediating these effects via changes in DNA methylation has not been studied at length.

1.9 Folate and DNA methylation

The relationship between folate and DNA methylation in fetal programming was initially clearly demonstrated by the classical study of the *agouti* gene in viable yellow agouti (A^{vy}) mice by Waterland and Jirtle (Waterland and Jirtle, 2003). The *agouti* gene is a gene that is expressed in mice hair follicles during a short period of hair development and growth and encodes a paracrine signalling molecule, which is responsible for the production of the yellow pigment, pheomelanin. In normal wild type mice, a yellow band appears on the otherwise brownish hair. A metastable allele due to the spontaneous insertion of an intracisternal A particle (IAP) retrotransposon upstream of the *agouti* gene results in the viable yellow agouti mice. A cryptic promoter within the IAP is what causes the permanent and constitutive agouti expression in all of the tissues in A^{vy} mice that result in the yellow coat. These mice are also extremely obese due to ectopic agouti expression in the hypothalamus. This insertion of the IAP into the agouti gene causes epigenetic dysregulation, causing inter-individual variability in CpG methylation at the A^{vy} locus. Therefore, within a single litter of genetically identical A^{vy}/a mice, depending on the methylation status of the 7 CpG dinucleotides found in the promoter at the proximal end of the A^{vy} IAP, the *agouti* gene can produce a range of phenotypes, from yellow and obese to brown/black and slim. Waterland and Jirtle tested whether the metastable methylation status of the specific transposable element insertion sites in agouti mice are epigenetically labile to early methyl donor nutrition. The results showed that a maternal diet supplemented with methyl donating dietary vitamins, such as folic acid, vitamin B_{12}, choline and betaine during pregnancy can alter the phenotype of their genetically identical offspring via increased CpG methylation at the A^{vy} locus. Hypomethylated A^{vy} alleles allowed maximal ectopic *agouti* expression causing the mice to have a yellow coat whereas hypermethylation of the A^{vy} alleles silences the ectopic *agouti* expression in pseudoagouti animals, giving them a brown coat (Waterland and Jirtle, 2003) (Figure 1.5).

Similar findings have been confirmed in other experimental models, such as the effect of maternal methyl donor supplementation on tail kinking due to the metastable *axin-fused* epiallele in mice (Waterland et al., 2006a) and the effect of a low protein diet during pregnancy on the methylation status of the hepatic *peroxisome proliferator activated receptor γ* and the *glucocorticoid receptor* gene in rats (Lillycrop et al., 2005, Choudhuri et al., 2010). Research that has examined the effect of methyl donor deficiencies in the mothers, rather than the offspring, as

well as those investigating the effect of post-natal diet have also been carried out. Methyl donor deficient diets in both cases showed differences in the methylation status of specific genes, *p53* in mothers (Pogribny et al., 1995) and *Igf2* (Waterland et al., 2006b) in the post-natal diet study. A recent study has also looked at the impact of a folate deficient diet on methylation in female mice, prior to, during pregnancy and lactation in blood, liver and kidney cells (McKay et al., 2011). Methylation changes that were tissue specific were seen. The above findings demonstrate that dietary supplementation influences the establishment of epigenetic gene regulation both trans-generationally, as well as at the organism level in animals and also suggest that dietary supplementation may influence the establishment of epigenetic gene regulation in humans.

Apart from the DNA methylation cycle, the DNA synthesis pathway mentioned in Section 1.3 also uses THF for its reaction and plays an important role in the folate mediated one-carbon metabolism pathway. One important enzyme involved in this pathway is *DHFR*.

1.10 The *Dihydrofolate Reductase* Gene Family

DHFR plays an important role in health and disease, and has therefore been researched at length and remains to be a gene of great interest. The *DHFR* gene family consists of five genes, which are highly homologous and are dispersed across the genome. *DHFR* is located on chromosome 5 (Chen et al., 1984), *DHFRP1* on chromosome 18 (Anagnou et al., 1984, Anagnou et al., 1988), *DHFRP2* on chromosome 6 (Shimada et al., 1984, Mighell et al., 2000), *DHFRP3* on chromosome 2 (Shimada et al., 1984) and *DHFRP4/DHFRL1* on chromosome 3 (Shimada et al., 1984, Mighell et al., 2000) (Figure 1.6). The first functional human *DHFR* gene was characterised by Chen *et al*. in 1983 (Chen et al., 1984). The other four members of the *DHFR* gene family were up until recently thought to be pseudogenes, i.e. genes which are similar to one or more paralogous genes, but without any function, i.e. they are not transcribed or translated into a functional protein. However, in 2011, the *DHFRP4/DHFRL1* gene was in fact shown to be a retrogene - expressed and functional (McEntee et al., 2011).

1.11 *Dihydrofolate Reductase*

DHFR is an important enzyme that plays an essential role in folate metabolism and is found in all eukaryotes and prokaryotes. DHFR catalyses the reduction of dihydrofolate (DHF) and folic acid to its biologically active metabolite, THF, which accepts and donates one-carbon groups for reactions that are necessary for DNA synthesis, cell proliferation and cellular methylation. This vital reaction occurs in the presence of NADPH. The reaction involves the transfer of a hydride ion

from NADPH to the C6 atom of the pterin ring of DHF, together with a protonation reaction to produce THF (Figure 1.7). Therefore, DHFR plays a central role in maintaining the cellular pools of THF and its derivatives, which are required for purine and thymidylate synthesis as well as cellular methylation (Schnell et al., 2004). However, there is a limit to the capacity to which DHFR can convert folic acid to THF and therefore if folic acid is ingested in doses higher than ~300 µg, unmetabolised folic acid appears to be found in the circulation. Although folic acid lacks any coenzyme activity, *in vitro* studies have shown that it can act as an inhibitor of DHFR, further restricting its conversion to the biologically active THF. Individuals have also been shown to have high variation in their DHFR activity, suggesting that folic acid supplementation could result in variable production of folate forms, which in turn could have inconsistent impacts on DNA methylation (Bailey and Ayling, 2009). High levels of folic acid in the circulation have also recently been found to be associated with colorectal tumorigenesis and prostate cancer, causing concern. However, the effect of unmetabolised folic acid on DNA methylation has never been investigated to the best of my knowledge.

In normal cells, *DHFR* expression is tightly controlled at the transcriptional, translational and post-translational level. During the cell cycle, *DHFR* transcription is controlled by the transcription factors Sp1 and E2F, with high levels of *DHFR* expression seen in late G_1 / early S phase. E2F sites have been shown to repress transcription in the G_0 and early G_1 phase, whereas Sp1 sites have been shown to mediate the induction of *DHFR* transcription in the late G_1 phase (Jensen, 1997). *DHFR* translation is controlled by the binding of the DHFR protein to its own mRNA, i.e. an auto-regulatory mechanism (Tai et al., 2004). *DHFR* is also subject to post-translational modifications, such as monoubiquitination and sumoylation. DHFR has been shown to undergo sumoylation *in vitro* by a protein called SUMO-1. Sumoylation is thought to occur during the S phase, and is believed to facilitate the translocation of DHFR to the nucleus, where it is involved in the *de novo* nuclear thymidylate synthesis pathway (Figure 1.1) (Anderson et al., 2007).

As *DHFR* plays a crucial role in the folate pathway, *DHFR* has been targeted for many different anti-folate drug therapies, as the inhibition of *DHFR* can limit the growth and proliferation of cells that are characteristic of cancer. *DHFR* inhibition can also kill prokaryotes such as *Staphylococcus aureus* and *Mycobacterium*, as well as unicellular eukaryotes such as *Plasmodium falciparum* (Hecht et al., 2011). Chemotherapeutic drugs include those such as methotrexate (MTX), which has been used in the treatment of leukaemia, non-Hodgkin's lymphoma, breast cancer, osteosarcoma and also the autoimmune disease rheumatoid arthritis (Goto et al., 2001, Schnell et al.,

2004, Walling, 2006) for over 50 years. Other drugs such as trimethoprim and pyrimethamine target bacterial and protozoal *DHFR* with pyrimethamine being an effective treatment of malaria (Roper et al., 2004). However, in response to the extensive use of antibiotics and anti-malarial drugs, *DHFR* has shown that it has the ability to resist / has developed resistance towards many of the anti-folate drugs in many organisms, *via DHFR* gene amplifications, mutations in *DHFR*, a decrease in the uptake of drugs among others, rendering them ineffective as chemotherapeutics (Hecht et al., 2011).

1.12 Dihydrofolate Reductase Like-1

Recent research has shown that *DHFRL1* is in fact a functioning retrogene (Anderson et al., 2011, McEntee et al., 2011) that localises to the mitochondria and takes the place of *DHFR* in mitochondrial *de novo* thymidylate biosynthesis (Anderson et al., 2011). *DHFRL1* has a 92% sequence homology to *DHFR* and the amino acids required for both catalytic activity and those involved in the binding of folate and NADPH in DHFR are also highly conserved in *DHFRL1* with just three differences at positions 24 (W24R), 54 (K54R) and 81 (V81I). The W24R substitution is of significance, as previous studies have shown that a tryptophan at position 24 of *DHFR* is important for substrate binding (Beard et al., 1991, McEntee et al., 2011).

McEntee *et al*. also showed that recombinant *DHFRL1* can complement a *DHFR*-negative phenotype in both bacterial and mammalian cells and that it is has enzyme activity although its specific activity was ~30% lower in comparison to *DHFR*. The Km values for NADPH were similar in both enzymes, whereas *DHFRL1* has a higher Km for DHF than *DHFR*. The difference in specific activity / Km is thought to be due to the W24R substitution in the catalytic site. Like *DHFR*, *DHFRL1* can also regulate its own translation by binding to its own mRNA, resulting in translational repression, as well as to the mRNA of *DHFR* and vice versa. This may have implications on the use of anti-folate drugs, particularly if the two enzymes show to have different binding affinities for different anti-folate drugs. In comparison, the localisation of *DHFRL1* to the mitochondria suggests that mitochondrial *DHFR* activity may be optimal with a lowered affinity for DHF and that its primary role is to support mitochondrial DNA synthesis and replication (McEntee et al., 2011).

The literature presented above demonstrates the importance of the different aspects of the folate mediated one-carbon metabolism pathway and its association with human health and disease. These findings form the basis of my thesis, of which the aims and objectives are outlined below.

1.13 Aims & Objectives

The aim of this project is to explore the molecular mechanism of the role of folate in human health. This will be achieved through three specific **Aims** (Figure 1.8):

- Investigation of the impact of folic acid supplementation on gene-specific DNA methylation patterns during pregnancy.
- Understanding the functional relevance of specific polymorphisms from the *DHFR* gene family and their relevance for disease.
- Investigation of the effect of altered DHFRL1 expression on genes / proteins.

The aims will be achieved through the following **Objectives** (Figure 1.8):

1. Screening of candidate human folate sensitive DNA methylation sites identified from the literature in a folic acid intervention pregnancy study using Methylation Sensitive-High Resolution Melting (MS-HRM) curve analysis.
2. Assessment of DNA methylation changes in the promoter regions of *DHFR* and *DHFRL1* in a folic acid intervention pregnancy study using Sensitive Melting Analysis after Real Time-Methylation Specific PCR (SMART-MSP) analysis.
3. Correlation of *DHFR* 19 bp intron polymorphism with circulating metabolites.
4. Characterisation of endogenous DHFRL1 and its polymorphisms in the mitochondria.
5. Characterisation of recombinant DHFRL1 and its polymorphisms.
6. Proteomic analysis of DHFRL1 expression.

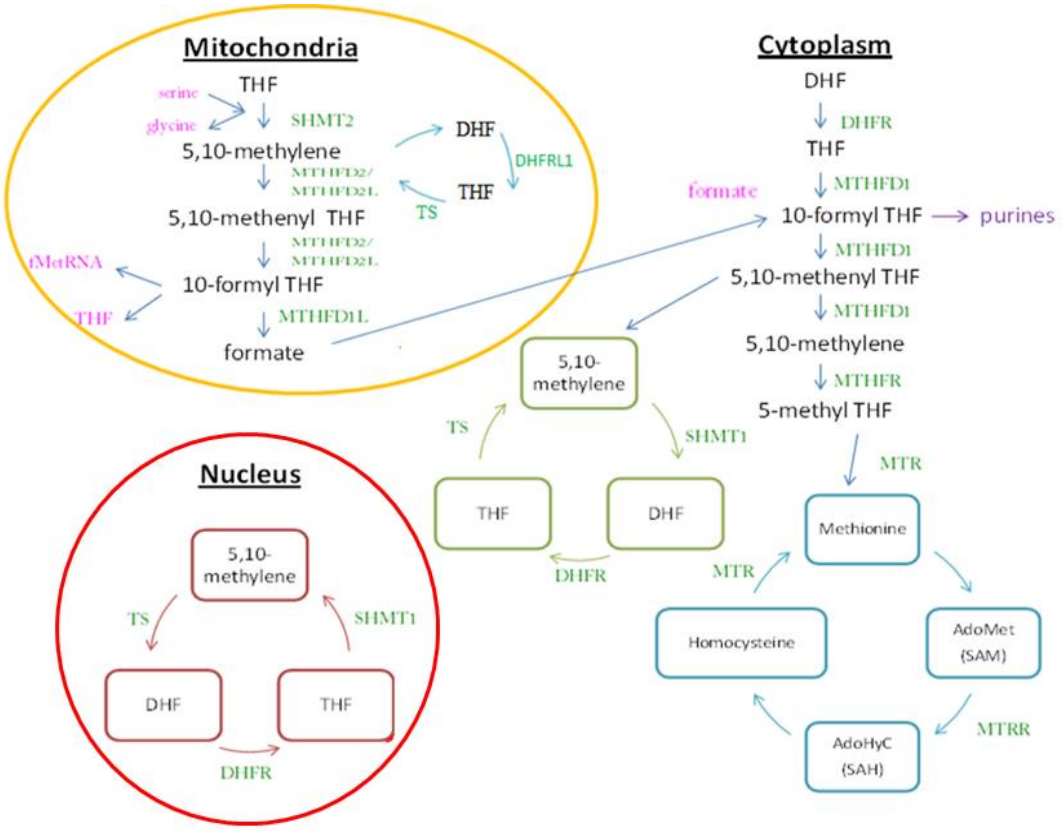


Figure 1.1 Compartmentalisation of folate mediated one-carbon metabolism. Folate mediated one-carbon metabolism is a complex biochemical pathway that is essential to the cell. It is compartmentalised between the cytoplasm, mitochondria and the nucleus. In the cytoplasm, folate mediated one-carbon metabolism can be divided into three distinct pathways, the *de novo* purine biosynthesis pathway, the *de novo* thymidylate (dTMPs) biosynthesis pathway and the homocysteine (Hcy) re-methylation pathway (methyl cycle). In the mitochondria, the folate mediated one-carbon metabolism pathway is required to generate formate, which is utilised in cytoplasmic folate mediated one-carbon metabolism as well as supplying 5,10-methyleneTHF for *de novo* thymidylate synthesis from uridylate (dUMP) and serine and in the nucleus, the pathway is required to synthesise thymidylate from dUMP and serine. Image adapted from (Tibbetts and Appling, 2010).

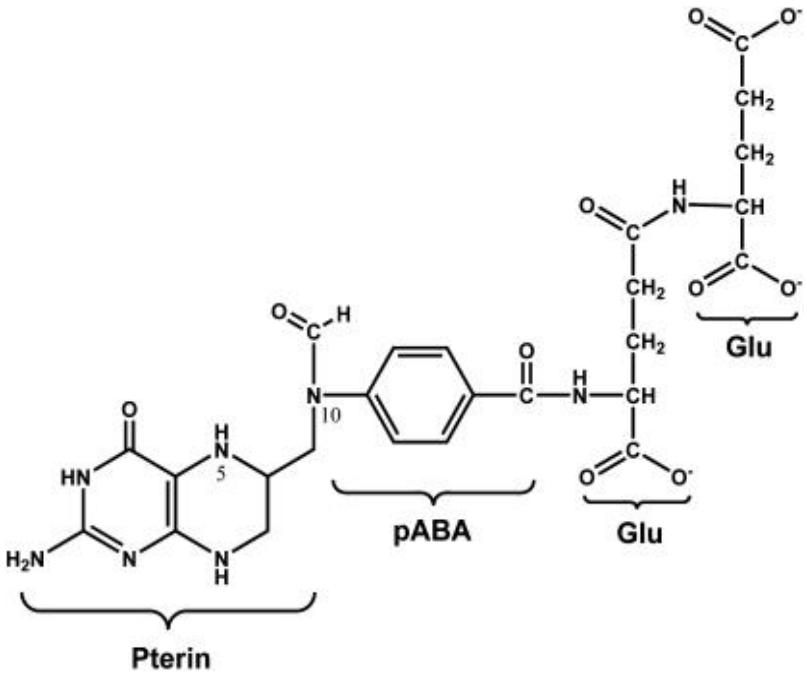


Figure 1.2 Structure of tetrahydrofolate (THF). THFs are made up of a fully reduced pterin ring, a p-aminobenzoyl group (pABA) and a poly-γ-glutamate tail (Glu). Image taken from (Stover and Field, 2011).

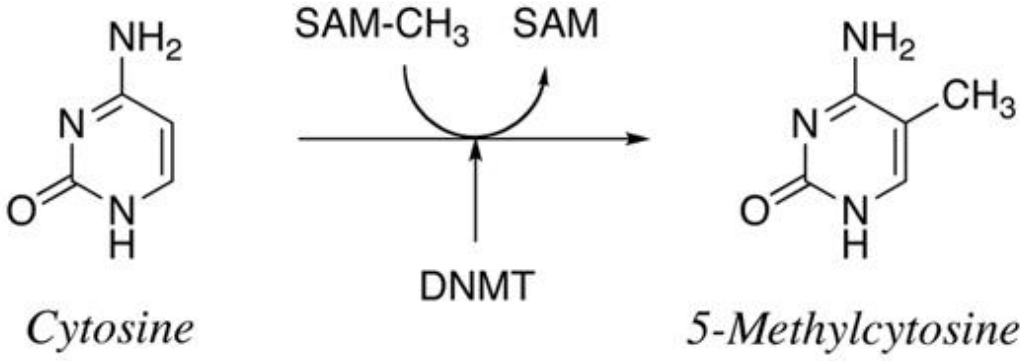


Figure 1.3. Overview of DNA methylation. DNA methylation is the addition of a methyl group (CH_3) to a cytosine residue. DNMT catalyses the transfer of the methyl group from SAM to the 5-carbon position of cytosine. Image adapted from (Klose and Bird, 2006).

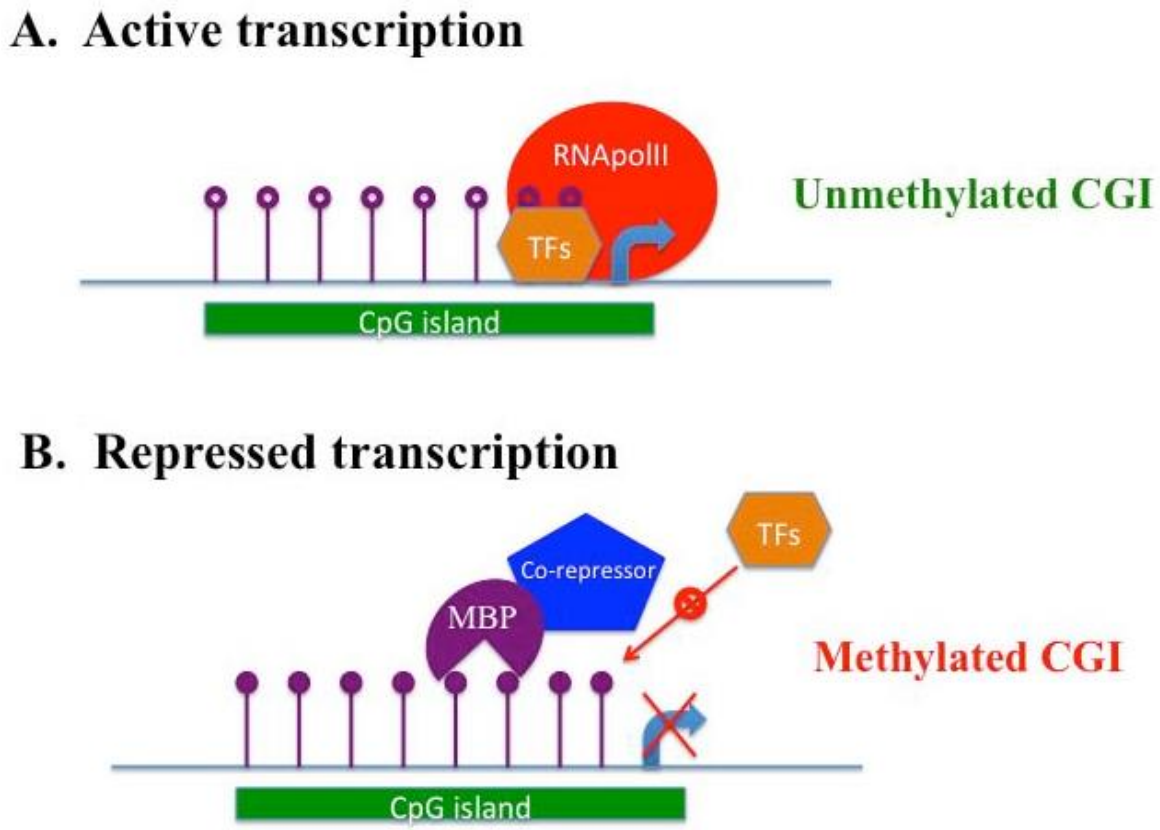


Figure 1.4. The role of CpG island methylation in the control of gene transcription. The illustration shows unmethylated and methylated CpG islands (CGI) in the promoter region of a given gene. Open circle lollipops indicate unmethylated cytosines, closed circle lollipops indicate methylated cytosines. (A) Unmethylated CGIs are usually in a transcriptionally active state. Transcription factors (TFs) recognise specific DNA motifs in the promoter and facilitate binding of the RNA polymerase II (RNApolII) complex to initiate transcription at the transcription start site (curved arrow). (B) Methylated CGIs are associated with a transcriptionally repressed state. This is thought to be due to the binding of methyl-binding protein (MBP), which recognises methylated cytosines, which subsequently recruit co-repressor complexes that ultimately prevent binding by TFs. An alternative mechanism is that methylated cytosines directly prevent binding of TFs (Parle-McDermott and Ozaki, 2011). Image taken from (Parle-McDermott and Ozaki, 2011).

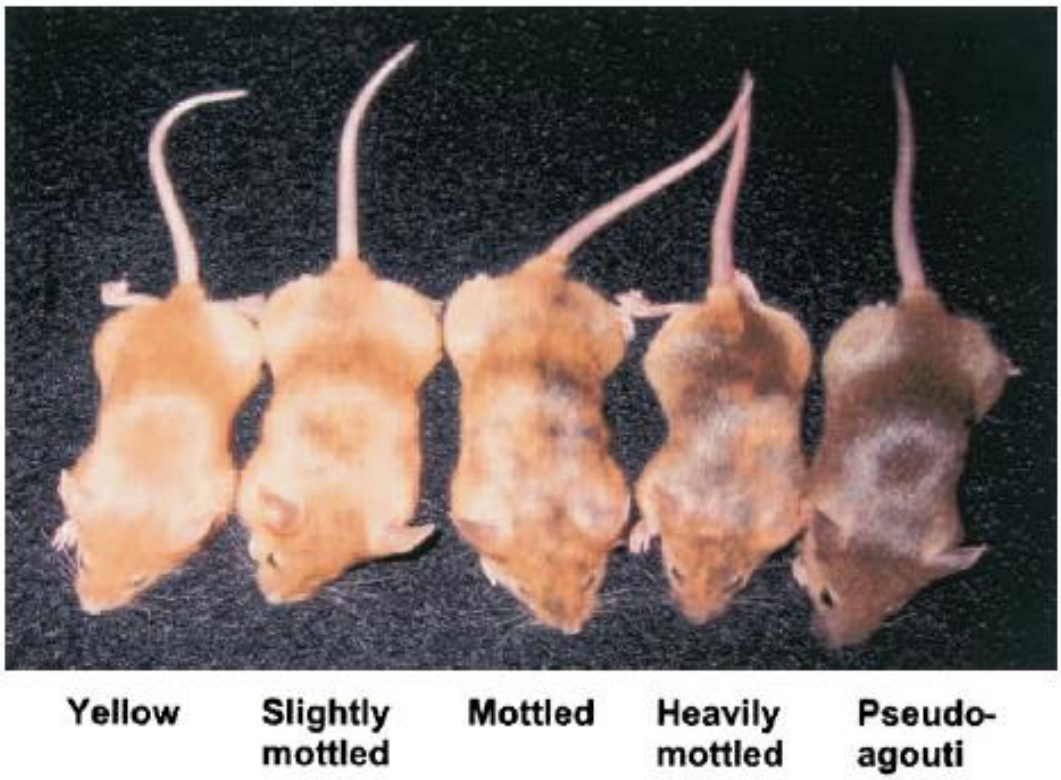


Figure 1.5. Coat colour phenotypes of A^{vy}/a offspring whose mothers were supplemented with different doses of folic acid. The offspring mice above represent the five coat colour classes used to classify the agouti phenotype. The A^{vy} alleles of yellow mice are hypomethylated, allowing maximal expression of the agouti gene. When the A^{vy} alleles are hypermethylated, it silences the expression of the agouti gene, seen in the pseudoagouti animals, and recapitulates the agouti brown/black phenotype. (Image taken from Waterland & Jirtle, *Mol Cell Biol,2003Aug;* 23**,** 5293-300.)

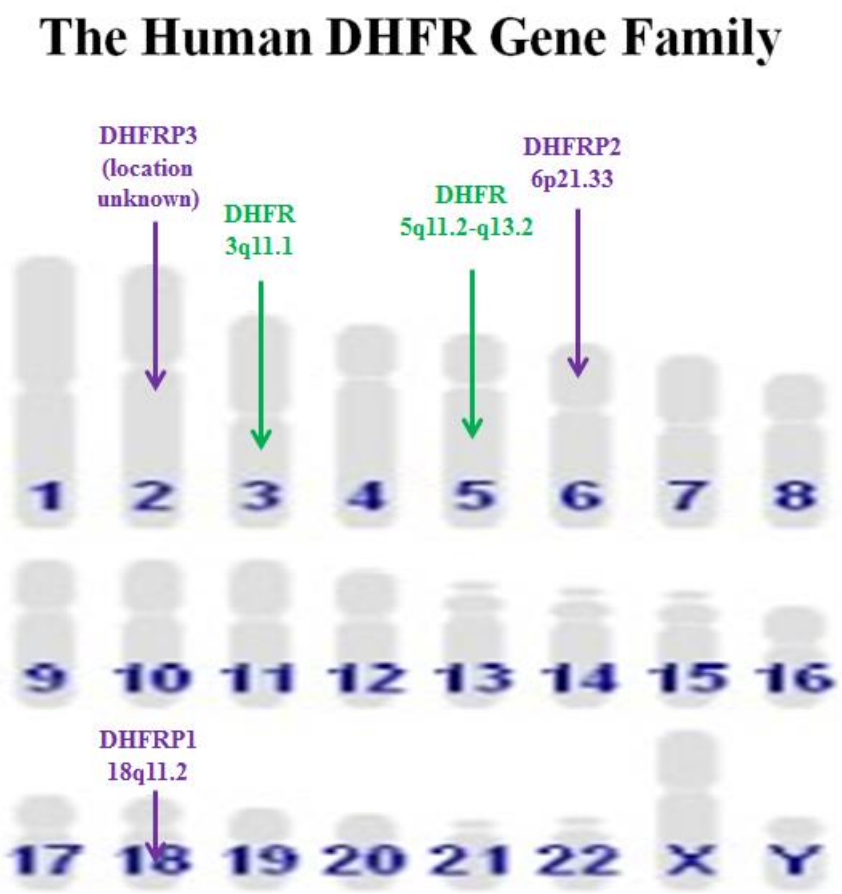


Figure 1.6. The location of the members of the human *DHFR* gene family. There are five genes that make up the human *DHFR* gene family. *DHFR* is located on chromosome 5, *DHFRP1* is located on chromosome 18, *DHFRP2* is located on chromosome 6, *DHFRP3* is located on chromosome 2 and *DHFRP4/DHFRL1* is located on chromosome 3. The genes in green represent functional genes and the genes in purple represent pseudogenes (Image by Parle-McDermott).

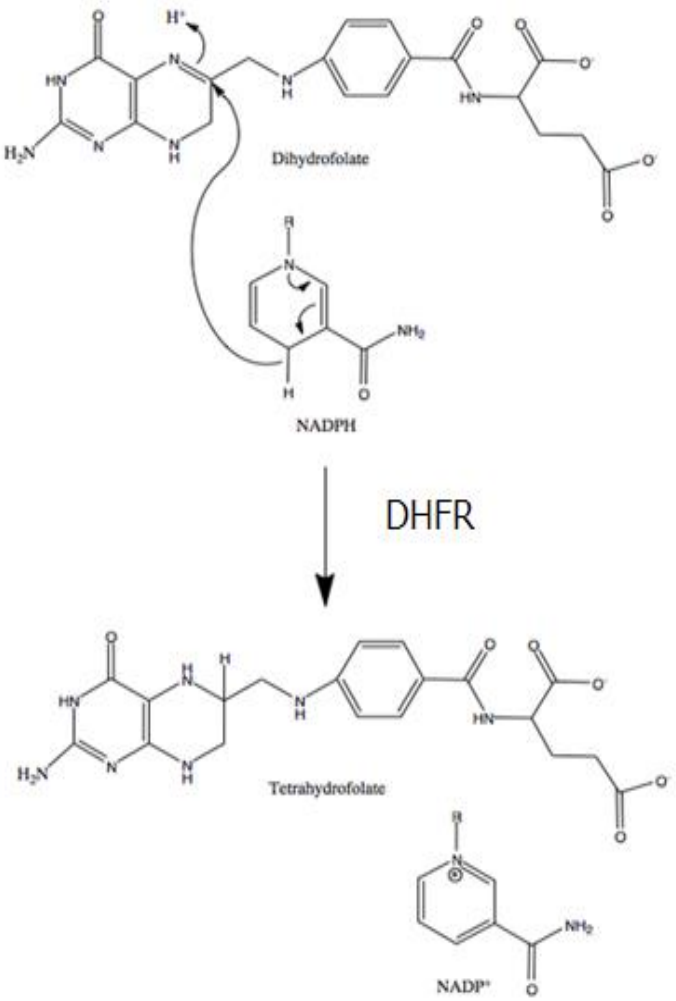


Figure 1.7. Conversion of Dihydrofolate into Tetrahydrofolate. Dihydrofolate (7,8-DHF) must be reduced to 5,6,7,8-THF before it can be utilised in the folate mediated one-carbon metabolism pathway. This vital reaction is catalysed by the enzyme called DHFR and occurs in the presence of NADPH. The reaction involves the transfer of a hydride ion from NADPH to the C6 atom of the pterin ring of DHF, together with a protonation reaction to produce THF (Image adapted from (Bailey and Ayling, 2009)).

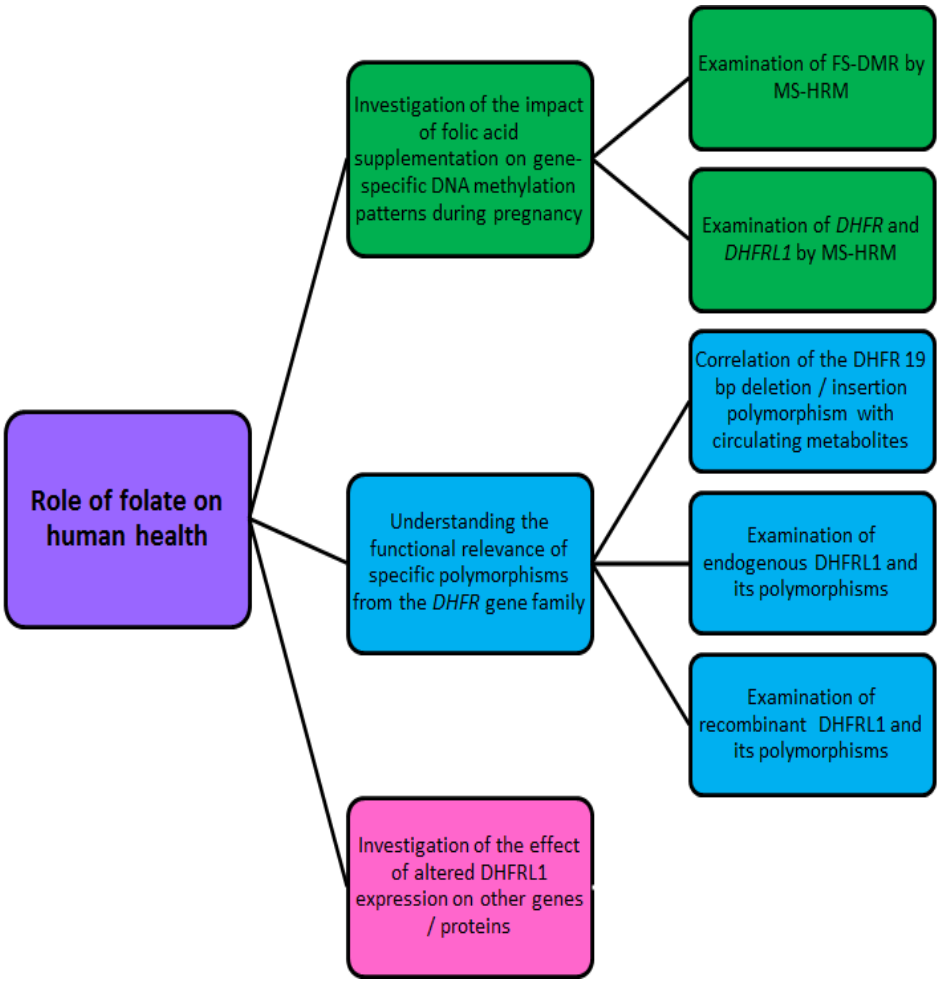


Figure 1.8. A schematic representation of the project carried out. The aim of this project was to explore the molecular mechanism of the role of folate in human health and disease and was carried out through three specific aims as shown above.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Reagents

Solid chemicals were weighed using a MettlerToledo AG204 electronic balance (Mason Technology). Gilson pipettes were used for the transfer of liquid volumes of up to 1 mL (Gilson S.A., France) and an electronic pipette was used for volumes that were greater than 1 mL with disposable plastic tips (Starstedt Ltd., Wexford, Ireland).

Mammalian Cell Culture

BioSciences Roswell Park Memorial Institute (RPMI) 1640 Medium (Cat No. 21875-034)

Cruinn Tissue Culture flasks 75cm^2 (Cat No. 658175CI), Tissue Culture flasks 25cm^2 (Cat No. 690175CI)

Coriell Institute Coriell Lymphoblast Cells: 17201 (Cat No. GM17201), 17243 (Cat No. GM17243)

Gibco Trypan Blue Stain 0.4% (Cat No. 15250-061)

Invitrogen Lipofectamine 2000 Reagent (Cat No. 11668500)

Lennox Puromycin Dihydrochloride Biochemica (Cat No. CA2856.0010)

Sarstedt 10 mL pipettes, 25 mL pipettes, 6-well plates, 100 mm plates/dishes

Sigma Aldrich Cell Freezing Medium – DMSO 1X (Cat No. C6164), Dulbecco's Modified Eagle's Medium (Cat No. D5030), Fetal Bovine Serum (FBS) (Cat No. F9665), G418 Disulfate Salt (Cat No. A1720), L-Glutamine Solution (Cat No. G7513), Sodium Pyruvate Solution (Cat No. S8636), Trypsin-EDTA Solution (Cat No. T4049-500mL)

Molecular Biology

Affymetrix ExoSAP-IT (Cat No. 78250 40 UL)

Agilent Technologies QuikChange Lightning Multi Site-Directed Mutagenesis Kit (Cat No. 210515), QuikChange II XL Site-Directed Mutagenesis Kit (Cat No. 200522)

Bioline Alpha-Select Gold Efficiency Cells (Cat No. BIO-85027), ISOLATE Plasmid Mini Kit (Cat No. BIO-52026), HyperPAGE Prestained Protein Marker (Cat No. BIO-33065),

Cambridge BioScience Human Methylated and Non-methylated DNA (Cat No. D5014)

Coriell Institute African American DNA Samples (Cat No. HD50AA), Caucasian DNA Samples (Cat No. HD50CAU)

Eurofins Sensor and Anchor probes

Fisher Scientific Acetic Acid (Cat No. 10304980), Glutathione Agarose (Cat No. 16101), Methanol (Cat No. 10785484), Pierce 96-well Microplates (Cat No. DIS-210-170F), Protease Inhibitor Cocktail I (Cat No. 12801640), TEMED (Cat No. T/P190/04),

Integrated DNA Technologies Primers, Gene Synthesis (DHFR and DHFRL1)

Invitrogen BL21-A1™ One Shot® Cells (Cat No. C6070-03), MAX Efficiency DH5α Competent Cells (Cat No. 18258-012), NuPage LDS sample buffer (Cat No. NP0008)

Merck Millipore Amicon Ultra-15 Centrifugal Filter Unit with Ultracel-3 membrane (Cat No. (UFC900324), Coomassie Brilliant Blue G250 (Cat No. 1154440025)

MyBio Endoproteinase Lys-C (Cat No. V1071), Runblue Prestained Marker (Cat No. NXA05160), Trypsin Gold, Mass Spectrometry Grade (Cat No. V5280)

New England Biolabs (NEB) BL21 (DE3) Competent *E. coli* cells (Cat No. C2527H), Quick-Load 1kB DNA Ladder (Cat No. N0468S), 100bp Ladder (Cat No. N3231)

Promega dNTP Mix (Cat No. U1515)

Qiagen EpiTect Fast DNA Bisulfite Kit (Cat No. 59826), FlexiGene DNA Kit (Cat No. 51206). Qproteome Mitochondria Isolation Kit (Cat No. 37612)

Roche Applied Science Lightcycler® 480 96 multi-well clear plates (Cat No. 05102413001), Lightcycler® 480 Genotyping Master 5X (Cat No. 04707524001), Lightcycler® 480 High Resolution Master (Cat No. 04909631001)

Sigma Aldrich Acrylamide / bis-acrylamide 40% (Cat No. A9926), Agar (Cat No. A5306), Agarose (Cat No. A9539), Ammonium Persulfate (Cat No. A3678-100G), Ampicillin Sodium salt (Cat No. A0166), Boric Acid (Cat No. B6768), Bovine Serum Albumin (Cat No. A3294), Bradford Reagent (Cat No. B6916), ColorBurst Electrophoresis Marker (Cat No. C1992-1VL), *DHFR* Assay Kit (Cat No. CS0340), Dimethyl Sulfoxide (DMSO) (Cat No. D8418), EDTA (Cat No. EDS), Ethanol (Cat No. E7023), Ethidium Bromide (Cat No. E13585), Glycine (Cat No. G8898), HIS-Select Nickel Affinity Gel (Cat No. P6611), Imidazole (Cat No. I5513), Imprint DNA Modification Kit (Cat No. MOD50) Isopropanol (Cat No. 278475-1L), Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Cat No. I6758), L-Arabinose (Cat No. A3256), Orange G (Cat No. O3756), PCR Core Kit with Taq DNA Polymerase (Cat No. CORET), Potassium Chloride (Cat No. P9541), Proteinase K (Cat No. P3354), RNA Sample Loading Buffer (R4268), Sodium Chloride (Cat No. S7653), Sodium Dodecyl Sulphate (Cat No. 71725), Sucrose (CatNo. 84100), Triton™X-100 (Cat No. X100), Trizma Base (Cat No. T1503), Tryptone (Cat No. T7293), Tween® 20 (Cat No. P9416-100 mL), Yeast Extract (Cat No. Y1333), 2-Mercaptoethanol (Cat No. M3148)

Terumo Neolus Hyperdermin needles (Cat No. NN-2719R).

Thermo Scientific Carbenicillin (Disodium Salt) (Cat No. BP2648-5), Coomaisse Brilliant Blue G250 Dye (Cat No 20279), PageRuler Plus Prestained Protein Ladder (Cat. No 26619), Pierce® BCA Protein Assay Kit (Cat No. 23227), Pierce 1-Step Transfer Buffer (Cat No. 84731), PVDF Transfer Membrane (Cat No. 88518), Restore™ PLUS Western Blot Stripping Buffer (Cat No. 46430), SuperSignal West Femto Maximum Sensitivity Substrate (Cat No. 34095), Western Blotting Filter Paper (Cat No. 84783)

2.1.2 Stock Solutions

- Bacterial Protein Lysis Buffer – 50 mM Potassium Phosphate (pH 7.8), 400 mM NaCl, 100 mM KCl, 10 % Glycerol, 0.5 % Triton X-100, 10 mM Imidazole (pH 7.8)
- Blocking Buffer – 5 g Marvel, 50 µL Tween® 20, 50 mL 10 X TBS, 450 mL dH_2O
- Buffer B – 1.5 M (145.05 g) Tris-HCl, 0.4 % (1 g) SDS, 250 mL dH_2O (pH 6.8)
- Coomaisse Blue De-stain – 100 mL Acetic Acid, 300 mL Methanol, 600 mL dH_2O
- Coomaisse Blue Stain – 2.5 g Coomaisse Brilliant Blue G250 Dye, 100 mL Acetic Acid, 300 mL Methanol, 600 mL dH_2O
- LB Broth – 2.5 g Tryptone, 2.5 g NaCl, 1.25 g Yeast Extract, 250 mL dH_2O
- LB Agar – 15 g of agar per litre of LB Broth
- Loading Buffer (Western) (4 X) – 2.4 mL Buffer B, 4 mL Glycerol, 1 g SDS, 4 mg Bromophenol blue, 14.4 M (500 µL) beta-mercaptoethanol, 3.1 mL dH_2O
- Orange G (10 X) – 0.1 g Orange G, 20 g Sucrose, 50 mL dH_2O
- PBS (10 X) – 80 g NaCl, 2 g KCl, 14.4 g Na_2HPO_4, 2.4 g KH_2PO_4, 1 L dH_2O (pH7.4)
- TBE (10 X) – 48.44 g Tris HCl, 12.37 g Boric Acid, 1.5 g EDTA, 500 mL dH_2O (pH 8.2)
- TBS (10 X) – 0.1M (6.06 g) Tris Base, 1.5 M (43.83 g) NaCl, 500 mL dH_2O (pH 8)
- TBST – 250 µL Tween® 20, 50 mL 10 X TBS, 450 mL dH_2O
- Tris – Glycine Running Buffer (10X) – 0.25 M (15.14 g) Tris Base, 1.9 M (71.3 g) Glycine, 1 % (5 g) SDS, 500 mL dH_2O (pH8.3)
- Transfer Buffer (1 X) – 100 mL 10 X Transfer Buffer, 200 mL Methanol, 700 mL dH_2O
- 4 X Sample buffer – 2.4 mL 1M Tris / HCl (pH 6.8), 4 mL Glycerol, 1 g SDS, 4 mg Bromophenol blue, 500 µL betamercaptoethanol, 3.1 mL dH_2O (pH 6.8)

2.2 Molecular biology methods

2.2.1 Agarose Gel Electrophoresis

PCR products were analysed by agarose gel electrophoresis. A 1-2 % w/v agarose gel was made by adding 0.5-1 g of agarose powder to 50 mL 1 X TBE buffer. The agarose was dissolved completely in the TBE by heating the solution in the microwave for approximately 2 min. The agarose solution was then allowed to cool to 55-60^{o}C and 10 µL ethidium bromide (10 µg/mL) was then added and mixed. The gel casting apparatus was prepared in a mould by sealing the ends of the gel chamber and the solution was poured onto a gel electrophoresis rig gel plate. Two 8 well combs were inserted and any bubbles were removed from the gel using a pipette tip. The gel was left to solidify at room temperature for 40 min. The combs were then carefully removed from the gel and the gel was submerged in 1 X TBE buffer, with the wells facing the negative terminal. A 2 µL volume of 1 kb molecular weight DNA ladder and a 5 µL volume of each sample added to 2 µL of 10 X Orange G loading dye were loaded into the wells. Electrophoresis was then carried out, keeping the voltage constant at 90 V for 45-60 min. The PCR products were then visualised, and examined under UV light using a DNR Mini-Bis Pro Bio-Imaging System. Images were taken using the VisionworksLS software program and printed on the Fujifilm Thermal Imaging System FTI-500 for recording in laboratory books and saved as an image file for thesis / publication purposes.

2.2.2 PCR Product Purification / Clean Up

A 2 µL volume of ExoSAP-IT was added to 5 µL of PCR product in a PCR tube. The tubes were centrifuged briefly and incubated in a thermal cycler for 15 min at 37 °C and then for 15 min at 80°C.

2.2.3 DNA Isolation from Buffy Coat

Frozen buffy coat samples were thawed quickly in a 37°C water bath. A 20 µL volume of proteinase K was added to each sample in order to homogenise the blood clots found in the samples before DNA extraction.

The Qiagen® FlexiGene DNA Kit composed of Buffer FG1 (lysis buffer), Buffer FG2 (denaturation buffer), Buffer FG3 (hydration buffer) and Qiagen protease was used and the protocol was carried out according to the manufacturers' instructions with slight alterations.

A 500 µL volume of Buffer FG1 was pipetted into a 1.5 mL centrifuge tube. A 200 µL volume of buffy coat was added and mixed by inverting the tube 5 times. The cells were pelleted by

centrifugation at 10,000 X g for 20 s. The supernatant was removed and the tube was left inverted on a clean sheet of absorbent paper for 2 min, ensuring that the pellet remained in the tube. A 200 µL volume of Buffer FG2 / Qiagen protease master mix (Table 2.1) was added to the tube and the tube was vortexed immediately until the pellet was completely homogenised. The tube was centrifuged for 3-5 s and incubated in a water bath for 10 min at 65°C. A 300 µL volume of 100 % isopropanol was added to the tube and the tube was mixed by inversion, until the DNA precipitate became visible as white threads / clumps. The solution was pelleted by centrifugation at 10,000 X g for 3 min. The supernatant was removed and the tube was left inverted on a clean sheet of absorbent paper, ensuring that the pellet remained in the tube. A 200 µL volume of 70 % ethanol was added to the tube and the tube was vortexed for 5 s. The solution was pelleted by centrifugation at 10,000 X g for 3 min. The supernatant was removed and the tube was left inverted on a clean sheet of absorbent paper for 5 min, ensuring that the pellet remained in the tube. The DNA pellet was air dried for 5 min, until all the liquid had evaporated. A 50 µL volume of Buffer FG3 was added to the tube, and the tube was vortexed at low speed for 5 s. The DNA was dissolved by incubating the tube in a water bath for 1-2 h at 65°C. The DNA samples were analysed by agarose gel electrophoresis, as described in Section 2.2.1.

Table 2.1 Buffer FG2 / Qiagen Protease master mix

Component	Volume per sample (µL)
Buffer FG2	200
Qiagen Protease	2

2.2.4 Quantifying DNA Concentration

DNA concentrations were measured using the NanoDrop 1000 Spectrophotometer (Thermo Scientific). The NanoDrop was blanked with the same buffer that was previously used to dissolve the DNA. The DNA parameter was chosen on the NanoDrop instrument and 1 µL of each sample was loaded onto the optical pedestal and the sample was then drawn into a column and measured. UV-Vis spectrum results were displayed using the NanoDrop 1000 Spectrophotometer software and the DNA concentrations were measured as ng/µL with an expected A260/280 ratio of 1.8-2.1, indicating acceptable DNA purity within the samples. The samples were then stored at -20°C.

2.2.5 Plasmid DNA preparation

A LB agar plate containing an appropriate selecting agent was streaked overnight at 37°C in an incubator with the clone of interest. A 5 mL volume of LB broth containing the appropriate selecting agent, which was warmed to room temperature, was pipetted into a 50 mL tube. A single colony was isolated / picked from the plate with a sterile pipette tip and grown in the broth at 37°C overnight with shaking (220 rpm).

2.2.6 Plasmid DNA Isolation (Small Scale)

The Isolate Plasmid Mini Kit, composed of re-suspension buffer, lysis buffer P, neutralisation buffer, wash buffer AP, wash buffer BP and elution buffer was used to isolate plasmid DNA from the bacterial culture. The protocol was carried out according to the manufacturers' instructions with slight alterations (Bioline).

A 5 mL volume of overnight culture was added to a 15 mL tube. The tube was centrifuged at 15,000 X g for 10 min and the supernatant was discarded. The pellet was re-suspended in 250 µL of re-suspension buffer and was vortexed thoroughly. A 250 µL volume of lysis buffer P was added to the tube and the solution was mixed carefully by inversion. A 350 µL volume of neutralisation buffer was added to the tube and the solution was mixed by inversion. The tube was centrifuged for 10 min at 15,000 X g and the supernatant was transferred to a spin column, which was placed in a 2 mL collection tube. The tube was centrifuged for 1 min at 12,000 X g and the filtrate was discarded. A 500 µL volume of wash buffer AP was added and the tube was centrifuged at 12,000 X g for 1 min. The filtrate was discarded. A 700 µl volume of wash buffer BP was added and the tube was centrifuged at 12,000 X g for 1 min. The filtrate was discarded. The tube was centrifuged at 15,000 X g for 2 min to remove ethanol and the collection tube was discarded. The spin column was placed into a 1.5 mL eppendorf tube. A 50 µL volume of molecular grade dH_2O was added directly to the column membrane and the tubes were incubated at room temperature for 1 min. The tube was centrifuged at 12,000 X g for 1 min to elute the plasmid DNA. The isolated DNA was stored at -20°C. A 1 % w/v agarose gel as described in Section 2.2.1 was run to ensure that DNA was present and intact.

2.2.7 Plasmid DNA Isolation (Large Scale)

The Qiagen Plasmid Maxi Kit, composed of buffer P1, buffer P2, buffer P3, buffer QBT, buffer QC, buffer QF and RNase A was used to isolate plasmid DNA from large bacterial cultures. The protocol was carried out according to the manufacturers' instructions with slight alterations.

A starter culture of 5 mL of LB broth containing carbenicillin (50 µg/mL) was inoculated with a single colony and grown overnight at 37°C with vigorous shaking (300 rpm). The starter culture was diluted 1/500 LB broth containing carbenicillin (50 µg/mL), i.e. a 100 mL volume of LB broth was inoculated with 100 µl of starter culture. The culture was grown at 37°C for 12–16 h with vigorous shaking (300 rpm). The bacterial cells were harvested by centrifugation at 2,300 X g for 30 min at 4°C. The pellet was re-suspended in 4 mL of Buffer P1 and was vortexed thoroughly. A 4 mL volume of Buffer P2 was added to the tube and the solution was mixed thoroughly by vigorously inverting the tube 4–6 times. The tube was incubated at room temperature for 5 min. A 4 mL volume of chilled Buffer P3 was added to the tube and mixed immediately and thoroughly by vigorously inverting the tube 4–6 times. The tubes were incubated on ice for 15 min. The solution was split in 6 x 2 mL tubes and the tubes were centrifuged at ≥20,000 X g for 30 min at 4°C. The supernatant containing the plasmid DNA was transferred into new 2 mL tubes promptly. The supernatant was centrifuged again at ≥20,000 X g for 15 min at 4°C. The supernatant containing the plasmid DNA was removed promptly. A QIAGEN-tip 500 was equilibrated by applying 10 mL of Buffer QBT, and allowing the column to empty by gravity flow. The supernatant was applied to the QIAGEN-tip and was allowed to enter the resin by gravity flow. The QIAGEN-tip was washed with 2 x 30 mL of Buffer QC. The DNA was eluted with 5 mL of Buffer QF. The eluate was collected in a 15 mL tube. The DNA was precipitated by adding 3.5 mL of room-temperature isopropanol to the eluted DNA. The DNA was mixed and centrifuged immediately at 2,300 X g for 90 min at 4°C. The supernatant was carefully decanted. The DNA pellet was washed with 2 mL of room-temperature 70 % ethanol, and transferred into a 2 mL tube. The tube was centrifuged at ≥15,000 x g for 10 min. The supernatant was carefully decanted without disturbing the pellet. The pellet was air-dried for 5–10 min, and the DNA was re-dissolved in 200 µL of dH_2O.

2.2.8 Gateway Cloning

For the His tag expression clones, primers were designed to incorporate a TEV cleavage site at the N-terminal and a His tag at the C-terminal as well as att-sites for both *DHFR* and *DHFRL1*. Primers for the *DHFR* and *DHFRL1* rare codon usage optimised clones were designed to incorporate a TEV cleavage site at the N-terminal as well as att-sites. Primers were designed in accordance to the Gateway Cloning Manual. Primer sequences are shown in Table 2.2. PCR products of the genes of interest were created by amplifying HEK 293 cDNA with the primers. Master mix for the amplification is shown in Table 2.3.

Table 2.2 Primer Sequences for Gateway Cloning

Gene	Primer Sequences
DHFR His TAG	Fwd: 5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAAATCTGTACTTCCAGGGGATGGTTGGTTCGCTAAACTG -3′ Rev: 5′-GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAATGATGATGATGATGATGATCATTCTTCTCATATACTT -3′
DHFRL1 His TAG	Fwd: 5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAAATCTGTACTTCCAGGGGATGTTTCTTTTGCTAAACTG -3′ Rev: 5′-GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAATGATGATGATGATGATGATCATCCTTCTCACATACTT -3′
DHFR Optimised	Fwd: 5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAAATCTGTACTTCCAGGGGATGGTTGGTAGCCTGAACTG-3′ Rev: 5′-CCCCACCACTTTGTACAAGAAAGCTGGGTCTTAATCGTTCTTTTCATACACCT-3′
DHFRL1 Optimised	Fwd: 5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAAATCTGTACTTCCAGGGGATGTTTCTGCTGTTAAATG-3′ Rev: 5′-CCCCACCACTTTGTACAAGAAAGCTGGGTCTTAATCGTCTTTTTCGCACACCT-3′

Table 2.3 Master Mix for PCR Amplification

Component	Volume (µL)
Forward primer (10 pmol/µL) (IDT)	1
Reverse primer (10 pmol/µL) (IDT)	1
10X buffer	5
Taq (5U/µL)	1
dH_2O	41

A 1 µL volume of a 1/160 dilution of HEK 293 cDNA (100 ng/µL) was used per reaction, giving a final volume of 50 µL. A negative-template control using 1 µL H_2O instead of DNA was also included. Template DNA was initially denatured at 95°C for 3 min, followed by 31 amplification cycles, in an automated ABI 9700 DNA Thermal Cycler (Applied Biosystems). Each cycle consisted of template denaturation (95°C for 1 min), primer annealing (60°C for 1 min) and extension (72°C for 2 min). This was followed by an elongation step to complete the amplification cycle (72°C for 10 min). Samples were maintained at 4°C. PCR products were run on a 2 % agarose gel, as described in Section 2.2.1.

Entry clones were created by carrying out a BP recombination reaction using the PCR product and the pDONR™ 211 vector, (See Appendix A for vector map).

Table 2.4 BP Reaction

Component	Volume (µL)
PCR product	4
pDONR 221 (150 ng/µL)	1
H_2O	13
BP Clonase	2

The reaction was incubated at 25°C overnight. A 1 µL volume of proteinase K was added to the BP reaction to inactivate the enzyme. The BP reaction was then transformed into alpha-select gold efficiency cells (F^- *deo*R *end*A1 *rec*A1 *rel*A1 *gyr*A96 *hsd*R17(rk-, mk^+) *sup*E44 thi-1 *pho*A Δ(*lac*ZYA-argF)U169 Φ80*lac*ZΔM15 λ^-). Alpha-select gold efficiency cells were thawed on ice. A

1 µL volume of the BP reaction was added to a tube containing 50 µL of the cells. The tubes were swirled gently and incubated on ice for 30 min. After incubation, the tubes were heat-pulsed in a 42°C water-bath for 30 s. A 450 µL volume of room temperature LB broth was added to the tubes and the tubes were incubated at 37°C for 1 h with shaking at 230 rpm. A 40 µL volume and a 100 µL volume of each transformation reaction were spread on a LB agar plate containing kanamycin (50 µg/mL). The plates were incubated at 37°C overnight. A single colony was isolated and the colony was grown overnight in 5 mL of LB broth containing kanamycin (50 µg/mL) in a 50 mL tube at 37°C. A 700 µL volume of overnight culture and a 300 µL volume of 50 % glycerol was added to a 1.5 mL cryovial tube and stored at -80°C. Plasmid DNA was isolated from the remainder of the overnight culture as described in Section 2.2.6. The DNA was sent off for sequencing with the M13 forward primer to confirm that the HIS tag and the TEV cleavage site had been inserted.

An expression clone was then created by performing an LR reaction between plasmid DNA from the entry clone and the pDEST™ 15 vector, (See Appendix B for vector map).

Table 2.5 LR Reaction

Component	Volume (µL)
Entry Clone (100 ng/µL)	2
pDEST 15 (150 ng/µL)	1.5
TE Buffer	4.5
LR Clonase	2

The reaction was incubated at 25°C overnight. A 1 µL volume of proteinase K was added to the LR reaction and was incubated at 37°C for 10 min to inactivate the enzyme. The LR recombination reaction was then transformed into Library Efficiency® DH5α cells (F^- φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(r_k^-, m_k+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ^-). Library Efficiency® DH5α cells were thawed on ice. A 1 µL volume of the BP reaction was added to a tube containing 50 µL of the cells. The tubes were swirled gently and incubated on ice for 30 min. After incubation, the tubes were heat-pulsed in a 42°C water-bath for 30 s. A 450 µL volume of room temperature LB broth was added to the tubes and the tubes were incubated at 37°C for 1 h with shaking at 230 rpm. A 40 µL volume and a 100 µL volume of each transformation reaction were spread on a LB agar plate containing ampicillin (100 µg/mL). The plates were incubated at 37°C overnight. A single colony was isolated and the colony was grown overnight in 5

mL of LB broth containing ampicillin (1000 µg/mL) in a 50 mL tube at 37°C. A 700 µL volume of overnight culture and a 300 µL volume of 50 % glycerol was added to a 1.5 mL cryovial tube and stored at -80°C. Plasmid DNA was isolated from the remainder of the overnight culture as described in Section 2.2.6. The DNA was sent off for sequencing with the T7 reverse primer to confirm the identity of the recombinant plasmid.

2.2.9 Site Directed Mutagenesis (SDM)

The QuikChange II XL SDM Kit was used to carry out SDM. This kit allows site-specific mutation in virtually any double-stranded plasmid and was therefore used.

Primer Design

Primers were designed according to the QuikChange II XL SDM Kit's protocol (Agilent Technologies). Primer design was also checked using Stratagene's web-based QuikChange Primer Design Program (*www.agilent.com/genomics/qcpd*)

Table 2.6 Primer Sequences for Site Directed Mutagenesis

Gene	Primer Sequences
rs17855824 G->A (Val->Ile)	Fwd: 5′- TGCCTGAATACCCAGGTATTCTCTCTGATGTCCAG -3′ Rev: 5′- CTGGACATCAGAGAGAATACCTGGGTATTCAGGCA -3′
rs61739170 C->G (Pro->Ala)	Fwd: 5′- CTCAGCAGAGAACTCAAGGAAGCTCCACAAGGAGCTCATTTTC -3′ Rev: 5′- GAAAATGAGCTCCTTGTGGAGCTTCCTTGAGTTCTCTGCTGAG -3′
Gene	**Primer Sequences for Optimised DHFRL1 Gene Sequence**
rs17855824 G->A (Val->Ile)	Fwd: 5′- TGCCGGAATACCCGGGCATTCTCAGTGACGTCCAG -3′ Rev: 5′- CTGGACGTCACTGAGAATGCCCGGGTATTCCGGCA -3′
rs61739170 C->G (Pro->Ala)	Fwd: 5′- GAGTCGCGAACTTAAGGAAGCACCGCAGGGGGCACACTT -3′ Rev: 5′- AAGTGTGCCCCCTGCGGTGCTTCCTTAAGTTCGCGACTC -3′

Mutant Strand Synthesis Reaction

Once primers were designed, mutant strand synthesis was carried out. The control reaction and sample reaction were prepared as indicated in Table 2.7 and Table 2.8 and were thermal cycled.

Table 2.7 Control Reaction

Component	Volume (µL)
10X Reaction Buffer	5
pWhitescript 4.5 kb control plasmid (5 ng/µL)	2
Oligonucleotide control primer #1 [34-mer (100 ng/µL)]	1.25
Oligonucleotide control primer #2 [34-mer (100 ng/µL)]	1.25
dNTP mix	1
QuikSolution Reagent	3
ddH_2O	36.5
PfuUltra HF DNA polymerase (2.5 U/µL)	1

Table 2.8 Sample Reaction

Component	Volume (µL)
10X Reaction Buffer	5
dsDNA	1
Forward primer (100 ng/µL) (IDT)	1.25
Reverse primer (100 ng/µL) (IDT)	1.25
dNTP mix	1
QuikSolution Reagent	3
ddH_2O	37.5
PfuUltra HF DNA polymerase (2.5 U/µL)	1

The reactions were cycled as follows; 95°C for 1 min, followed by 18 cycles, in an automated ABI 9700 DNA Thermal Cycler. Each cycle consisted of 95°C for 50 s, 60°C for 50 s and 68°C for 5 min (control) / 7.5 min (sample). This was followed by 68°C for 7 min. Samples were maintained at 4°C. Following temperature cycling, a 1 µL volume of the *Dpn*1 restriction enzyme (10 U/µl) was added to each amplification reaction. Each reaction mixture was mixed gently and thoroughly by pipetting the solution up and down several times. The reaction mixture was spun down at 12,000 X g in a microcentrifuge for 1 min. The reactions were incubated immediately at 37°C for 1 h to digest the parental supercoiled dsDNA.

2.3 Protein Purification Methods

2.3.1 Transformation into XL10 Cells (SDM Plasmid DNA)

XL10 Gold Ultra competent cells (Tetr Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F′ proAB lacIqZΔM15 Tn10 (Tetr) Amy Camr]) were thawed on ice. A 45 µL volume of the cells were aliquoted to chilled 1.5 mL eppendorf tubes. A 2 µL volume of 2-Mercaptoethanol was added to the tube and the contents of the tubes were swirled gently. The tubes were then incubated on ice for 10 min, swirling gently every 2 min. A 2 µL volume (10 ng) of *Dpn*I treated DNA was then added to the tube containing the cell aliquot. The transformation reaction was swirled gently and was incubated on ice for 30 min. LB broth was pre-heated in a 42°C water-bath. After incubation, the tubes were heat-pulsed in a 42°C water-bath for 30 s and then incubated on ice for 2 min. A 0.5 mL volume of the pre-heated LB broth was added to the tube and the tubes were incubated at 37°C for 1 h with shaking at 230 rpm. A 250 µL volume of each transformation reaction was plated on a LB agar plate containing ampicillin (100 µg/mL). The plates were incubated at 37°C overnight. A single colony was isolated and the colony was grown overnight in 5 mL of LB broth containing ampicillin (100 µg/mL) in a 50 mL tube at 37°C. A 700 µL volume of overnight culture and a 300 µL volume of 50 % glycerol was added to a 1.5 mL tube and stored at -80°C. Plasmid DNA was isolated from the remainder of the overnight culture as described in Section 2.2.6. The DNA was sent off for sequencing to Source BioScience to confirm the identity of the recombinant plasmid.

2.3.2 Transformation into BL21-AI™ One Shot® Cells

A 5 µL volume (10 ng) of plasmid DNA from the expression clones and from each of the transformed clones were added to a vial of BL21-AI™ One Shot ® Cells (F⁻*ompT hsdS*B (rB⁻mB⁻) *gal dcm ara*B::*T7RNAP*⁻*tetA*) and were incubated on ice for 30 min. After incubation, the cells were heat-shocked for 30 s at 42^{O}C and were then transferred immediately back on to the ice. A 250 µL volume of LB broth was added to the tubes and the tubes were incubated at 37°C for 1 h with shaking at 230 rpm. A 100 µL volume of cells were spread on a set of LB plates containing ampicillin (100 µg/mL). The plates were incubated at 37°C overnight. Plates were then stored at 4°C.

2.3.3 Transformation into BL21 (DE3) Competent *E. Coli* Cells

BL21 (DE3) Competent *E. Coli* cells (*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*) were thawed on ice for 10 min. A 1-5 µL volume (100 ng) of plasmid DNA from the expression clone and from each of the

transformed clones was added to a vial of BL21 (DE3) Competent *E. Coli* cells. The tubes were flicked 4-5 times to mix the DNA with the cells and the tubes were then incubated on ice for 30 min. After incubation, the cells were heat-shocked for 10 s at 42°C and were then transferred immediately back on to the ice for 5 min. A 950 µL volume of S.O.C. media was added to each of the tubes and the tubes were incubated at 37°C for 1 h with shaking at 250 rpm. A 50 µL volume and a 150 µL volume of cells were spread on a set of LB plates containing carbenicillin (50 µg/mL). The plates were incubated at 37°C overnight. Plates were then stored at 4°C.

2.3.4 Induction of recombinant proteins with L-Arabinose / IPTG

Single colonies were isolated from plates and the colonies were grown overnight with shaking in LB broth containing carbenicillin (50 µg/mL) at 37°C. The overnight cultures were diluted by 1/20 by adding fresh LB broth containing carbenicillin (50 µg/mL). This was left to grow at 37ºC with shaking at 220 rpm for ~ 2 h, until an OD_{600} of 0.4 (optimal growth stage) was reached. Each culture was then split in two. L-arabinose / IPTG was added to one culture to give a final concentration of 0.04 % (v/v) (L-arabinose) or 50 µM (IPTG) to induce recombinant protein expression. A sample of both cultures was taken at various time points. At each time point, the cells were centrifuged for 5 min at 4470 X g, the supernatant was removed, and the pellets were weighed and were then frozen at -20 º C.

2.3.5 Sample Fractionation

Once pellets had been obtained for all time points, each pellet was thawed and lysed in bacterial lysis buffer (ratio of 0.4 g to 8 mL of bacterial lysis buffer used). Once the pellets were re-suspended, they were put through a needle (0.4 x 20 mm) and syringe in order to lyse the cells. The cells were then separated into soluble and insoluble fractions by freezing the samples in liquid nitrogen and thawing in a 42^{o}C water bath three times. The lysed cells were then centrifuged for 10 min at 4470 X g at 4^{o}C, to pellet insoluble proteins. The supernatant was transferred into 50 mL tubes. The insoluble fractions were lysed with 5 mL lysis buffer. The protein lysates were stored at -20ºC or -80ºC.

2.3.6 SDS-PAGE Gel Preparation

Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) was carried out in order to separate the proteins within each sample according to molecular weight (size). A 10 % (v/v) separating gel was composed of 3.5 mL 40 % Acrylamide, 2.5 mL Tris-HCL (pH 8.8), 100 µL 10 % (w/v) SDS, 100 µL 10 % (w/v) fresh Ammonium Persulfate (APS), 4 µL N, N, N', N'

tetramethylethylenediamine (TEMED) and the volume was brought up to a total of 10 mL with dH_2O. TEMED initiates the polymerisation reaction and therefore is added just before the gel is poured. Before adding the TEMED, the plates for the gel were prepared by cleaning them with isopropanol in one direction, placing the seal onto the plate with the lip facing upwards and by fastening the two plates together using bulldog clips. Once the TEMED was added, the gel was mixed and poured between the two plates using a Pasteur pipette up to 2/3 full. Ethanol was then applied to the top of the separating gel to remove any air bubbles and to stop atmospheric oxygen from inhibiting the polymerisation reaction. The gel was left to polymerise for 30 min. The stacking gel is cast over the separating gel and has larger pores. The stacking gel was made up with 830 µL 40 % Acrylamide, 630 µL Tris-HCL (pH 6.8), 50 µL 10 % (w/v) SDS, 50 µL 10 % (w/v) fresh APS and 5 µL TEMED and the volume was brought up to a total of 5 mL with dH_2O. Before adding TEMED, and when the separating gel was fully polymerised, the ethanol layer was poured off onto a tissue and the stacking gel was then overlaid on top of the separating gel. A 12 tooth comb was introduced into the stacking gel and the gel was allowed to polymerise for 30 min. Once polymerised, the bulldog clips and seal were removed and the gel was transferred to an electrophoresis tank and fixed, ensuring that the grooved side of the gel is placed facing towards the inside of the chamber. The central chamber was then filled with 1X Tris-Glycine running buffer. The comb was then removed, the wells straightened and blown out, using a Gilson pipette and a needle. Air bubbles were also removed from the base of the gels by tilting the apparatus to one side and refilling the central chamber.

2.3.7 Protein Sample Preparation

An 18 µL volume of protein sample was added to 6 µL of NuPage 4X LDS sample buffer / 4 X sample buffer and the mixture was heated for 5 min at 99ºC. A 5 µL volume of Protein Marker was applied to the first well and 20 µL of each sample was applied to the wells. The gels were run at 140 V until the dye front had electrophoresed off the end of the gel (~ 90 min). The gel was then removed carefully.

2.3.8 Protein Transfer to PVDF

Following the separation of the protein samples in a SDS-PAGE gel, the gels were carefully disassembled with a scalpel and were transferred to polyvinylidene fluoride (PVDF) membranes. One piece of PVDF membrane and two pieces of MM filter paper were cut to the appropriate size. The PVDF membrane was dipped into 100 % methanol for 30 s to activate it and then rinsed in cold 1 X transfer buffer. The MM filter papers, the PVDF membrane, along with the two sponges were

soaked in cold 1 X transfer buffer for 30 min. The SDS-PAGE gel was also soaked in cold 1 X transfer buffer. The transfer cassette was prepared by loading the transfer pad with a MM filter paper, then the PVDF membrane, followed by the gel containing the electrophoresed proteins, which were then covered with the remaining MM filter paper. The two pads, filter papers, gel and membrane were then sandwiched together, making sure there was no air bubbles caught between any of the layers. The transfer cassette was placed into the holder with the black plates facing towards the centre. The transfer apparatus (Mini-Trans Blot Cell, BioRad) was filled with cold 1 X transfer buffer. The electrophoretic transfer of proteins was performed by transferring the gels at 100 V at maximum Amps for 1 h, ensuring that the membrane was placed at the anode end, the gel at the cathode. The transfer was performed with cooling and buffer circulation. A representative semi-dry transfer layout is shown in Figure 2.1.

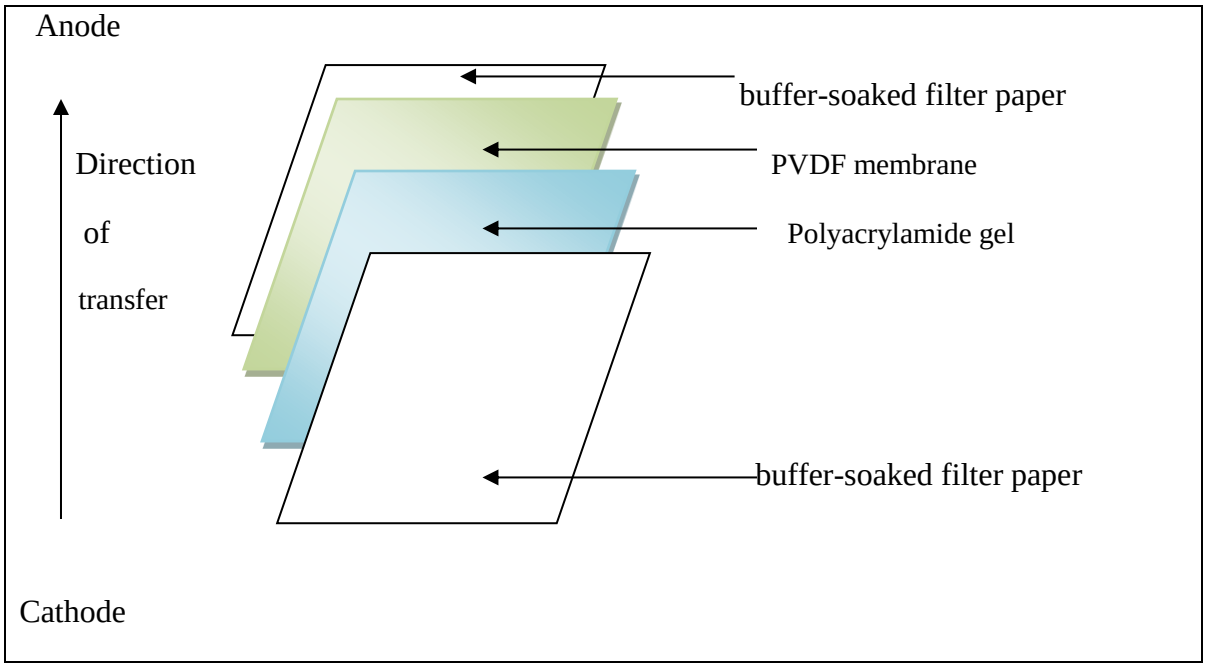


Figure 2.1. Representative layout of the semi-dry transfer. The base of the transfer apparatus was pre-soaked in 1 X transfer buffer. A pre-soaked filter paper was placed on top of this, followed by the methanol-activated, transfer buffer soaked PVDF membrane. The polyacrylamide gel was placed directly on top of this, followed by pre-soaked filter paper. The direction of protein transfer occurs from the gel (cathode) to the membrane (anode).

2.3.9 Transfer of proteins using the G2 Fast Blotter (Thermo Scientific Pierce)

Following the separation of the protein samples in a SDS-PAGE gel, the gels were carefully disassembled with a scalpel and were transferred to PVDF membranes. Four sheets of 0.83 mm thick western blotting filter paper and one sheet of PVDF membrane were cut to the same size. The PVDF membrane was dipped into 100 % methanol for 30 s to activate it and then the filter papers and the PVDF membrane were equilibrated in Thermo Scientific Pierce 1-step transfer buffer for at least 5 min. After electrophoresis, the gel was removed from the cassette and briefly placed into the tray containing the transfer buffer to ensure even wetting, to facilitate in proper gel placement and to improve gel contact with the membrane. The cassette was then assembled as before, with the anode on the bottom of the cassette. A blot roller was used to remove any trapped air bubbles. The top of the cassette (cathode) was locked into place and was slid into the cassette unit. The pre-programmed method was selected, as well as the number of gels, the molecular weight and the start button was pressed.

2.3.10 Staining gels with Coomassie Brilliant Blue

Following protein electrophoresis and the transfer of proteins, gels were stained in Coomassie brilliant blue and de-stained in Coomassie de-stain until the protein bands were clearly visible and background staining had been fully eliminated. The gel was then visualised under white light using a DNR Mini-Bis Pro Bio-Imaging System.

2.3.11 Western Blotting (Immunoblotting / Antibody binding)

Following the transfer of proteins, membranes were blocked with 50 mL blocking buffer for 2 h at room temperature on a rotating platform. Following this, the blocking solution was discarded. The membrane was then incubated in primary antibody, diluted in blocking buffer (See Table 2.9 for antibody concentrations) in a total volume of 10 mL and left overnight at 4°C with gentle agitation on a rotating platform. It was ensured that the membrane was completely covered so that it does not dry out. Following incubation, the membrane was washed three times for 15 min each in TBST on a rotating platform. The secondary antibody, (See Table 2.9 for antibody concentrations) was prepared in 20 mL TBST. The membrane was placed in the secondary antibody solution at room temperature for 1 h on a rotating platform. After incubation, the membrane was washed three times for 15 min each in TBST on a rotating platform. The SuperSignal West Femto Reagents were used to detect the HRP-labelled secondary antibody. Equal volumes of reagent A and reagent B were mixed together. The blot was placed on a sheet protector with the outside clean and the inside free of air bubbles and the reagent mixture was applied to the membrane. The excess

liquid was removed from the blot and the blot was imaged using the Gene Genome instrument (SynGene).

Table 2.9. Antibody Concentrations used for Western blotting

Primary Antibody	Primary Antibody Concentration	Secondary Antibody	Secondary Antibody Concentration
DHFR Abcam (Cat No. ab49881)	1:10,000	HRP conjugated anti-rabbit	1:50,000
DHFR (mitochondrial fractions)	1:5,000	HRP conjugated anti-rabbit	1:25,000
DHFRL1 Abcam (Cat No. ab171837)	1:250	HRP conjugated anti-rabbit	1:50,000
GST Sigma Aldrich (Cat No. G7781)	1:2,000	HRP conjugated anti-rabbit	1:200,000
GAPDH Sigma Aldrich (Cat No. G8795)	1:1,000	HRP conjugated anti-rabbit	1:10,000
PDH MitoSciences (Cat No. 110330)	1:1,000	HRP conjugated anti-mouse	1:100,000

2.3.12 Concentrating and Clarifying Protein Samples

Protein lysates were concentrated using 30 kDa centricons. Protein samples were spun in the centricons at 4470 X g for 1 h at 4 ^{o}C. This was carried out not only to concentrate the protein samples but also to eliminate free GST and small proteins. Protein lysates were then diluted 1 in 5 in bacterial protein lysis buffer and was filtered through a 0.45 µM filter in order to clarify the protein.

2.3.13 Protein Purification of GST-tagged Protein Using a Gravity Free Column

Once protein expression had been confirmed by SDS-PAGE analysis / Western blot analysis, protein purification was carried out. All samples were kept on ice at all times. The column was first washed with 1 X PBS. The column was then packed with 1 mL of Glutathione Agarose and the storage buffer was drained from the resin by gravity flow. Once the storage buffer had been drained off, the volume of resin was reduced by half. The column was then equilibrated with 10

resin bed volumes (10 mL) of Equilibration / Wash Buffer (50 mM Tris, 150 mM NaCl, pH 8.0), allowing the buffer to drain from the resin at a flow rate of 0.5 mL/min. Protein samples were prepared by diluting the protein lysates 1 in 2 in Equilibration / Wash Buffer. The prepared protein lysates were then added to the column using a Pasteur pipette, ensuring not to disturb the resin. The flow through was collected in 50 mL tubes and was re-applied to the column to maximise binding. The resin was washed with 10 resin bed volumes (10 mL) of Equilibration / Wash Buffer and the flow through was collected. This step was repeated using new collection tubes until the absorbance of the flow through fraction at 280 nm approaches baseline. The GST-tagged protein was then eluted from the resin with two resin-bed volumes (2 mL) of Elution Buffer (50 mM Tris, 150 mM NaCl, pH 8.0 containing 10 mM reduced glutathione). This step was carried out twice. Each fraction was collected in separate tubes. The protein elution was monitored by measuring the absorbance of the fractions at 280 nm. All purification fractions were analysed by SDS-PAGE and Western blot.

2.3.14 Protein Purification of HIS-tagged Protein

Once protein expression had been confirmed, protein purification was carried out. Samples were kept on ice at all times. A 50 µL volume of HIS-Select Nickel Affinity Gel suspension was added to a microcentrifuge tube and centrifuged for 30 s at 5,000 X g. The supernatant was removed and discarded. A 200 µL volume of Equilibration / Wash buffer (50 mM sodium phosphate, pH 8.0, 0.3 M sodium chloride) was added to the tube and mixed well. The tube was centrifuged for 30 s at 5,000 X g and the supernatant was discarded. A 100 µL volume of clarified protein was added to the tube and mixed gently for 1 min. The mixture was centrifuged for 30 s at 5,000 X g and the supernatant was saved for analysis. The affinity gel was washed twice with 500 µL of Wash buffer by mixing the affinity gel for 10 s and centrifuging for30 s at 5,000 X g. The supernatant (washes) were saved for analysis. The target protein was eluted with a 50 µL volume of Elution buffer (50 mM sodium phosphate, pH 8.0, 0.3 M sodium chloride, 250 mM imidazole). All purification fractions were analysed by SDS-PAGE and Western blot.

2.3.15 Mitochondrial Isolation from Eukaryotic Cell Lysates

The Qiagen® Qproteome Mitochondria Isolation Kit composed of lysis buffer, disruption buffer, mitochondrial storage buffer and protease inhibitor solution was used to extract mitochondria from Cell Lines. The protocol was carried out according to the manufacturers' instructions with slight alterations. All steps were carried out at 4°C. Lysis buffer and disruption buffer were supplemented with protease inhibitor solution.

A cell suspension containing 5 x 10^6 to 2 x 10^7 cells were transferred into 25 mL tubes and were centrifuged at 500 X g for 10 min at 4°C. The supernatant was discarded. A 1 mL volume of sterile PBS was added to the pellet and the pellet was re-suspended in order to wash the cells. The solution was transferred into 1.5 mL eppendorf tubes and was centrifuged at 500 X g for 3 min. The supernatant was discarded. A 500 µL volume of ice-cold lysis buffer was added to the tube and the pellet was re-suspended by pipetting up and down. The tube was incubated on ice for 10 min with gentle mixing every 30 s. The lysate was centrifuged at 1000 X g for 10 min at 4°C and the supernatant was removed to a clean eppendorf tube. This is the cytosolic fraction. A 500 µL volume of ice-cold disruption buffer was added to the pellet and the pellet was re-suspended by pipetting up and down. Cells were the completely disrupted using a blunt-ended needle (0.4 x 20 mm) and a syringe by drawing the lysates into the syringe slowly and ejecting the lysates in one stroke. This was repeated 10 times. The lysate was then centrifuged at 1000 X g for 10 min at 4°C and the supernatant was transferred into a clean eppendorf tube. The supernatant was centrifuged at 6000 X g for 10 min at 4°C and the supernatant was discarded. A 1 mL volume of mitochondrial storage buffer was added to the pellet and the pellet was re-suspended by pipetting up and down. The tube was then centrifuged at 6000 X g for 20 min at 4°C. The mitochondrial pellet was re-suspended in 100 µL of mitochondrial storage buffer and was stored at -20°C.

2.3.16 Protein Concentration Quantification by BCA Assay

The Pierce® BCA Protein Assay Kit composed of Bicinchoninic Acid (BCA) Protein Assay Reagent A (an alkaline bicarbonate solution), BCA Protein Assay Reagent B (a copper sulphate solution) and Albumin Stranded Ampules (2 mg/mL) was used and the protocol was carried out according to the manufacturers' instructions.

Preparation of Standards and Working Reagent

Preparation of Diluted Bovine Serum Albumin (BSA) Standards

A range of standards from 25 – 2000 µg of BSA was made. The dilution scheme for the Standard Microplate Procedure is shown in Table 2.10. dH_2O was used as the diluent.

Preparation of the BCA Working Reagent (WR)

The following formula was used to determine the total volume of WR required.

(# standards + # unknowns) x (# replicates) x (volume of WR per sample) = total volume of WR required. WR was prepared by mixing 50 parts of Reagent A with 1 part of Reagent B.

2.3.17 Microplate Procedure

A 10 µL volume of each standard and 10 µL of unknown sample replicate was pipetted into a 96-well microplate well. A 100 µL volume of WR was then added to each well and the plate was mixed on a rotating platform. The plate was covered with tinfoil and incubated at 37°C for 30 min. The plate was then cooled to room temperature and the absorbance was measured at a wavelength of 562 nm on the Tecan i600 spectrophotometer. The average 562 nm absorbance measurement of the blank standard replicates was subtracted from the average 562 nm absorbance measurements of all the other standards and unknown samples. A standard curve was then prepared by plotting the average blank-corrected 562 nm absorbance measurement for each BSA standard vs. concentration in µg/mL. The equation of the line ($y = mx + c$) was then used to determine the protein concentration of each of the unknown samples, where x is the unknown concentration of the sample, y is the absorbance of the sample, m is the slope of the line and c is a constant.

Table 2.10. Dilution Scheme for BSA

Vial	Volume of Diluent	Volume and Source of BSA	Final BSA Conc.
A	0	300 µL of stock	2,000 µg/mL
B	125 µL	375 µL of stock	1,500 µg/mL
C	325 µL	325 µL of stock	1,000 µg/mL
D	175 µL	175 µL of vial B	750 µg/mL
E	325 µL	325 µL of vial C	500 µg/mL
F	325 µL	325 µL of vial E	250 µg/mL
G	325 µL	325 µL of vial F	125 µg/mL
H	400 µL	100 µL of vial G	25 µg/mL
I	400 µL	0	0 µg/mL

2.3.18 Protein Concentration Quantification by Bradford Assay

The Sigma Bradford Reagent was used in order to determine the concentration of protein samples in which the background interfered with the Pierce® BCA Protein Assay Kit. A set of standards were prepared as described above in Table 2.10.

2.3.19 Microplate Procedure

A 5 µL volume of each standard, unknown sample and background medium was pipetted into a 96-well microplate well. All samples were carried out in duplicate. A 250 µL volume of Bradford Reagent was then added to each well and the plate was mixed on a rotating platform for 30 s. The plate was covered with tinfoil and incubated at room temperature for 10 min. The absorbance was measured at a wavelength of 595 nm using the Tecan i600 spectrophotometer. The protein concentration of each of the unknown samples was determined as described in Section 2.3.17.

2.3.20 Protein Concentration Quantification using Image J

Image J software (http://rsb.info.nih.gov/ij/index.html) was used to compare the density of the bands observed on the Western blots in order to quantify protein concentrations for Enzyme analysis.

2.3.21 Enzyme Activity Measurement for Endogenous and Recombinant Protein

Enzyme activity was tested using a Dihydrofolate Assay Kit. The assay was performed according to manufacturer's instructions, with slight alterations. All volumes in the protocol were divided by two in order to have enough protein to carry out the experiments. Assays were performed in 1 ml UVette cuvettes. In order to examine enzyme activity, methotrexate stock solutions were made up and used to ensure that any enzyme activity that is observed is due to DHFR and not any other enzymes in the cells or from the GST tag that use NADPH as a cofactor, i.e. methotrexate was used to remove any background activity. Methotrexate concentrations that were four fold higher than the concentration used for inhibition reactions of the purified enzyme was used. Samples were prepared as shown in Table 2.11 and Table 2.12. The samples were mixed after each component was added by covering the cuvette with parafilm and inverting.

The substrate was added just before enzyme activity was measured on a spectrophotometer. The absorbance was read at 340 nm using the kinetic mode reading every 15 s for 2.5 - 5 min. One unit of enzyme is defined as the amount which reduces 1 µmol of dihydrofolate per minute based on

the molar extinction coefficient of 12,300 $M^{-1}cm^{-1}$ at 340 nm. A graph was then prepared by plotting the 340 nm absorbance measurement vs. time in s. The decrease in ΔOD over time was measured as ΔOD/min and the specific activity was calculated using the following equation:

$$\text{Units / mg P} = \frac{(\Delta OD_{\text{min. sample}} - \Delta OD_{\text{min. blank}}) \times d}{12.3 \times V \times \text{mg P / ml}}$$

Where d = dilution factor of the enzyme sample

12.3 = extinction coefficient reaction at 340 nm

V = volume of enzyme used in ml

mg P / ml = concentration of enzyme of the original sample

Units / mg P = specific activity in µmol/min/mg protein

However, for the endogenous protein, the final value was multiplied by 1000 to obtain the specific activity value in nmol/min/mg protein. Also in contrast to purified protein, the exact concentration of your protein of interest in cell lysates / cytoplasmic / mitochondrial fractions is unknown and therefore the total concentration of the fractions was used in the calculations.

Table 2.11. Sample preparation for the DHFR control assay

	Blank A (µL)	Blank B (µL)	DHFR Control (µL)	DHFR + MTX (µL)
1 X Assay Buffer	492.6	493.1	490.1	488.1
DHFR	4.4	4.4	4.4	4.4
MTX 40 nM	0	0	0	2
NADPH	3	0	3	3
DHF	0	2.5	2.5	2.5

Table 2.12. Sample preparation for the sample assay

	Mt Fraction (µL)	Mt Fraction + MTX (µL)
1 X Assay Buffer	449.5	447.5
Protein Sample	45	45
MTX 40 nM	0	2
NADPH	3	3
DHF	2.5	2.5

2.4 Methylation Methods

2.4.1 Bisulfite Treatment for MS-HRM

The Imprint® DNA Modification Kit composed of DNA modification powder, DNA modification solution, balance solution, capture solution, cleaning solution and elution solution was used to bisulfite treat DNA samples and the protocol was carried out according to the manufacturers' instructions with slight alterations.

A 1.1 mL volume of DNA modification solution was added to a vial of DNA modification powder. The vial was vortexed for 2 min until the solution was clear. A 40 µl volume of balance solution was added to the vial and vortexed briefly. DNA with a concentration of 500 ng was added to a 1.5 mL micro-centrifugation tube, adjusting the total volume to 10 µL with dH_2O. A 110 µL volume of the prepared DNA modification solution was added into the tube. The tube was incubated at 99°C for 6 min, immediately followed by incubation at 65°C for 90 min. A spin column was placed into a capless 2 mL collection tube for each sample that was modified. A 300 µL volume of capture solution was added to the column and the solution was allowed to sit on the column for 1 min. The modified DNA solution was utilised onto the column already containing the capture solution. The column was centrifuged at 12,000 X g for 45 s. The flow through was discarded. A 200 µL volume of the ethanol-diluted cleaning solution was added to the column and the column was centrifuged at 12,000 X g for 45 s. A 50 µL volume of balance/ethanol wash solution was added to the bottom of the column, ensuring that no air bubbles were impeding the liquid flow to the column filter. The column was incubated at 8 min at room temperature. After incubation, the column was centrifuged at 12,000 X g for 45 s and the flow-through was discarded. A 200 µL volume of 90 % ethanol solution was added to the column and the column was centrifuged at 12,000 X g for 45 s and the flow-through was discarded. A 200 µL volume of 90 % ethanol solution was added to the column and the column was centrifuged at 12,000 X g for 1 min. The column was placed into a 1.5 mL collection tube and the capless collection tube was discarded. A 15 µL volume of elution solution was added to the bottom of the column. The solution was allowed to incubate for 1 min. The column was centrifuged at 12,000 X g for 45 s. The column was removed and discarded. The eluted solution is the modified DNA, which was stored at -20°C.

2.4.2 Methylation-Sensitive High-Resolution Melting (MS-HRM)

MS-HRM curve analysis was carried out in Lightcycler® 480 96 multi-well clear plates on the Lightcycler® 480. A master mix of PCR reagents was prepared per sample as described in Table 2.13. Primer sequences and annealing temperatures are shown in Table 2.14.

A 1 µL volume of bisulfite treated DNA (500 ng concentration) was used per reaction, giving a final volume of 10 µL. A negative-template control using 1 µL H_2O instead of DNA was also included. Template DNA was initially denatured at 95°C for 10 min, followed by 45 amplification cycles, followed by high-resolution melting analysis on the Lightcycler® 480. Each cycle consisted of template denaturation (95°C for 10 s), primer annealing (Tm for 15 s) and extension (72°C for 15 s). This was followed by high-resolution melting analysis. High-resolution melting settings consisted of 95°C for 1 min, 40°C for 1 min, 70°C for 5 s and continuous acquisition to 95°C (ramp rate 0.02 °C/s), followed by cooling at 40°C for 10 s. Melting temperature for the DNA samples were assigned automatically by the Lightcycler ® 480 Software.

Table 2.13. Master Mix for MS-HRM

Component	Volume (µL)
Forward primer (5 ρmol/µL) (IDT)	0.4
Reverse primer (5 ρmol/µL) (IDT)	0.4
25mM $MgCl_2$	1.2
Lightcycler® 480 HRM Master Mix	5
dH_2O	2.0

Table 2.14. Primer sequences and annealing temperatures used for MS-HRM

Gene	Primer Sequences	Annealing Temp.
ATM	Fwd: 5′-CGAAGAGGGTGGGTGAGAGTTT-3′ Rev: 5′-ACGCCATATCCACCAATAACCAAC-3′	58°C
RASSF1	Fwd: 5′-GTTTTAGATGAAGTCGTTATAGAGGT-3′ Rev: 5′-CCCCACGACAACTAATCCCTAA-3′	58°C
EIF2C3	Fwd: 5′-TTAGGAAGGAGGGTGGTTTTATT-3′ Rev: 5′-AAACTTACCTAAAAAATAACCTACCTAAAA-3′	52°C
ATP5F1	Fwd: 5′- CGTTGGGGGCGGTATAGGGG-3′ Rev: 5′-CCGAACTATCCCCAACCCTACCC-3′	60°C

2.4.3 Bisulfite Treatment for SMART-MSP

The Qiagen Epitect Fast Bisulfite Conversion Kit composed of Bisulfite Solution, DNA Protect Buffer, RNase-Free Water, MinElute Spin Columns, Collection Tubes, Buffer BL, Buffer BW, Buffer BD, Buffer EB and Carrier RNA was used to bisulfite treat DNA samples. The protocol was carried out according to the manufacturers' instructions with slight alterations. Bisulfite reactions were prepared in PCR tubes per sample as described in Table 2.15.

Table 2.15. Bisulfite Reaction

Component	Volume (µL)
DNA Sample (500 ng)	1-40
RNase-free water	40 - DNA sample
Bisulfite solution	85
DNA Protect Buffer	15

The PCR tube was closed and the reaction was mixed thoroughly. DNA bisulfite conversion was carried out on a thermal cycler. DNA was denatured at 95°C for 5 min, followed by incubation at 60°C for 10 min, denaturation at 95°C for 5 min, incubation at 60°C for 10 min and an indefinite hold at 20°C. The bisulfite converted DNA was then cleaned up. The PCR tube was centrifuged briefly and the reaction was transferred to a 1.5 ml eppendorf tube. A 310 µL volume of Buffer BL was added and the solution was mixed by vortexing. A 250 µL volume of 100 % ethanol was added and the sample was mixed by pulse vortexing for 15 s and centrifuged briefly. A MinElute DNA spin column was placed into a 2 mL collection tube and the entire solution from the tube was transferred into the column. The column was centrifuged at 15,000 X g for 1 min and the flow-through was discarded. A 500 µL volume of Buffer BW was added to the column, the column was centrifuged at 15,000 X g for 1 min. and the flow-through was discarded. A 500 µL volume of Buffer BD was added to the column, the lid was closed (to avoid acidification) and the column was incubated at room temperature for 15 min. The column was centrifuged at 15,000 X g for 1 min and the flow-through was discarded. A 500 µL volume of Buffer BW was added to the column, the column was centrifuged at 15,000 X g for 1 min and the flow-through was discarded. This step was repeated. A 250 µL volume of 100 % ethanol was added to the column and the column was centrifuged at 15,000 X g for 1 min. The column was transferred into new 2 mL collection tubes and centrifuged at 15,000 X g for 1 min. The column was placed into 1.5 mL eppendorf tube and the capless collection tube was discarded. A 10 µL volume of Buffer EB was added onto the centre

of the column membrane, the lid was closed and the column was incubated at room temperature for 1 min. The column was centrifuged at 15,000 X g for 1 min to elute the DNA. The modified DNA was stored at -20°C.

2.4.3 Sensitive Melting Analysis after Real Time-Methylation Specific PCR (SMART-MSP)

SMART-MSP curve analysis was carried out in Lightcycler® 480 96 multi-well white plates. A master mix of PCR reagents was prepared per sample as described in Table 2.16 and 2.17. Primer sequences and annealing temperatures are shown in Table 2.18.

A 2.5 µL volume (25 ng) of bisulfite treated DNA (theoretical amount) was used per reaction, giving a final volume of 10 µL. A positive 100 % methylated control and negative-template control using 2.5 µL H_2O instead of DNA was also included. All samples were carried out in duplicate. Template DNA was initially denatured at 95°C for 10 min, followed by 45 amplification cycles, followed by high-resolution melting analysis on the Lightcycler® 480. Each cycle consisted of template denaturation (95°C for 10 s), primer annealing (Tm for 15 s) and extension (72°C for 15 s). This was followed by high-resolution melting analysis. High-resolution melting settings consisted of 95°C for 1 min, 40°C for 1 min, 70°C for 5 s and continuous acquisition to 95°C (ramp rate 0.02 °C/s), followed by cooling at 40°C for 10 s. Crossing point (Cp) values and melting temperature for the DNA samples were assigned automatically by the Lightcycler ® 480 Software. Cp values were averaged and methylation values were quantified using the $2^{(-\Delta\Delta Cp)}$ quantification approach.

Table 2.16. Master Mix for *COL2A1* SMART-MSP

Component	Volume (µL)
Forward primer (5 ρmol/µL) (IDT)	0.8
Reverse primer (5 ρmol/µL) (IDT)	0.8
25mM $MgCl_2$	1
Lightcycler® 480 HRM Master Mix	5
dH_2O	0.5

Table 2.17. Master Mix for *DHFR* and *DHFRL1* SMART-MSP

Component	Volume (µL)
Forward primer (5 ρmol/µL) (IDT)	0.4
Reverse primer (5 ρmol/µL) (IDT)	0.4
25mM $MgCl_2$	1.2
Lightcycler® 480 HRM Master Mix	5
dH_2O	0.5

Table 2.18. Primer sequences and annealing temperatures used for MS-HRM

Gene	Primer Sequences	Annealing Temp.
COL2A1	Fwd: 5′-GTAATGTTAGGAGTATTTTGTGGGTA-3′ Rev: 5′-CTACCCCAAAAAAACCCAATCCTA-3′	63°C
DHFR	Fwd: 5′-GAGGGTTTTCGTTTTCGTTCG-3′ Rev: 5′-CAAATAAACCCTAACGCTACAACG-3′	54°C
DHFRL1	Fwd: 5′-GGGTTAGGGTTAAAGCGATTTTCGT-3′ Rev: 5′-AAACAAAAACCGCACAAACGCG-3′	58°C

2.5 Genotyping Methods

2.5.1 *DHFR* 19bp Genotyping Assay – Melt Curve Analysis

Melt curve analysis was carried out in Lightcycler ® 480 96 multi-well clear plates on the Lightcycler® 480. A master mix of reagents was prepared as shown in Table 2.19. Primer and probe sequences are shown in Table 2.20.

A 1-4 µL volume of DNA (100 ng) was used per reaction, giving a final volume of 20 µL. Template DNA was initially denatured at 95°C for 10 min, followed by 45 amplification cycles, followed by melt curve analysis on the Lightcycler® 480. Each cycle consisted of template denaturation (95°C for 10 s), primer annealing (59°C for 15 s) and extension (72°C for 15 s). This was followed by melt curve analysis. Melt curve settings consisted of 95°C for 1 min, 30°C for 1 min and 75°C for continuous (ramp rate 0.1 °C/s), followed by cooling at 40°C for 30 s. Genotypes for the DNA samples were assigned automatically by the Lightcycler ® 480 Software and manually

Table 2.19. Master Mix for *DHFR* 19bp deletion Melt Curve Analysis

Component	Volume (µL)
Forward primer (5 ρmol/µL) (IDT)	3
Reverse primer (5 ρmol/µL) (IDT)	1
Sensor probe (10 ρmol/µL) (Eurofins)	1
Anchor probe (10 ρmol/µL) (Eurofins)	1
DMSO	3
Lightcycler ® 480 Genotyping Master 5X	4
dH_2O	3-6

Table 2.20. *DHFR* 19bp deletion primer and probe sequences

Forward primer	5′-TGGGCATCGGCAAGAAC-3′
Reverse primer	5′-TCTGGCCCCATCCTCTC-3′
Sensor probe	5′-CCAGGTACCCCGACCGTG- Fluorescein -3′
Anchor probe	5′- BODIPY 630/650 - CAGCCTGCGCCCGTTTGGG- Phosphate -3′

2.5.2 *DHFRL1* rs17855824 SNP Melt Curve Analysis

Melt curve analysis was carried out in Lightcycler ® 480 96 multi-well clear plates on the Lightcycler® 480. A master mix of reagents was prepared as shown in Table 2.21. Primer and probe sequences are shown in Table 2.22.

Table 2.21. Master Mix for *DHFRL1* rs17855824 SNP Melt Curve Analysis

Component	Volume (µL)
Forward primer (5 ρmol/µL) (IDT)	2
Reverse primer (5 ρmol/µL) (IDT)	1
Sensor probe (10 ρmol/µL) (Eurofins)	1
Anchor probe (10 ρmol/µL) (Eurofins)	1
Lightcycler® 480 Genotyping Master 5X	4
dH_2O	10

A 1 µL volume of DNA was used per reaction, giving a final volume of 20 µL. Template DNA was initially denatured at 95°C for 10 min, followed by 45 amplification cycles, followed by melt curve analysis on the Lightcycler® 480. Each cycle consisted of template denaturation (95°C for 10 s), primer annealing (52°C for 15 s) and extension (72°C for 15 s). This was followed by melt curve analysis. Melt curve settings consisted of 95°C for 1 min, 40°C for 1 min and 75°C for continuous (ramp rate 0.1 °C/s), followed by cooling at 40°C for 30 s. Genotypes for the DNA samples were assigned automatically by the Lightcycler® 480 Software.

Table 2.22. DHFRL1 SNP 1 (rs17855824) primer and probe sequences

Forward primer	5′-GGAAGCCATGAATCACCTAG-3′
Reverse primer	5′-TACTTCAAATTTGTACTTGATGTGTTT-3′
Sensor probe	5′-ACATCAGAGAGAACACCTGGGTA- Fluorescein -3′
Anchor probe	5′- BODIPY 630/650 - CAGGCAGAAGTTTATATTTCTCCAAGTCAATTTCT- Phosphate -3′

2.5.3 Genotyping by Gel Electrophoresis

DHFRL1 Polymorphisms

A master mix of PCR reagents was prepared per sample as described in Table 2.23. *DHFRL1* primer sequences are shown in Table 2.24.

Table 2.23. Master Mix for *DHFR*L1 SNP Polymorphisms

Component	Volume (µL)
Forward primer (5 ρmol/µL) (IDT)	1
Reverse primer (5 ρmol/µL) (IDT)	1
10X buffer	2.5
2.5mM dNTPs	2
$MgCl_2$	1.5
Taq (5U/µL)	0.1
dH_2O	15.9

A 1 μL volume of DNA (100 ng) was used per reaction, giving a final volume of 25 μL. A negative-template control using 1 μL H_2O instead of DNA was also included. Template DNA was initially denatured at 95°C for 3 min, followed by 30 amplification cycles, in an automated ABI 9700 DNA Thermal Cycler. Each cycle consisted of template denaturation (94°C for 30 s), primer annealing (51°C for 1 min) and extension (72°C for 1 min). This was followed by an elongation step to complete the amplification cycle (72°C for 5 min). Samples were maintained at 4°C overnight. PCR products were run on a 2 % agarose gel, as described in Section 2.2.1.

Table 2.24. *DHFRL1* polymorphism primer sequences

Gene	Primer Sequences
rs17855824	Fwd: 5′- GGAAGCCATGAATCACCTAG -3′ Rev: 5′- TACTTCAAATTTGTACTTGATGTGTTT -3′
rs61739170	Fwd: 5′- AATCTGGTGATTATGGGTAGGA -3′ Rev: 5′- AGTTTAAGATGGCCTAGGTGATT -3′
rs114936057	Fwd: 5′- AACCGCTGCTGTCATGTTTCTTT -3′ Rev: 5′- CAGTAAGTTTTAAGGCATCATCCAAA -3′

2.6 Cell Culture Methods

2.6.1 Cell Culture

All cell culture was carried out in a laminar flow hood whilst adhering to strict sterile techniques. The hoods were switched on at least 15 min before use. All medium, flasks, pipettes, tips and the inside of the hoods were sprayed and wiped down with ethanol (70 % v/v) before and after use to maintain aseptic conditions in the work area.

2.6.2 Cell Lines

Coriell cell lines, Epstein Barr Virus (EBV) transformed lymphoblastoid cultures from peripheral human blood mononuclear cells and human embryonic kidney (HEK) cells were used.

2.6.3 Maintenance of Cell Lines

2.6.3.1 Coriell Cell Lines

All Coriell cell lines were maintained in Roswell Park Memorial Institute (RPMI) 1640 medium containing 2 mM L-Glutamine, and supplemented with heat-inactivated Fetal Bovine Serum (FBS) (15 % v/v) (hereafter known as complete medium). The cells were grown in 25 cm^2

vented sterile tissue culture flasks in an upright position in an incubator at 37°C with 5 % CO_2, 95 % humidified air.

2.6.3.2 HEK 293 Cell Lines

All HEK 293 cell lines were maintained in Dulbecco's Modified Eagle Medium (DMEM), supplemented with heat-inactivated FBS (10 % v/v), 2 mM L-Glutamine and 1 mM sodium pyruvate (hereafter known as complete medium). The HEK 293 DHFRL1 stably transfected overexpressed cell pools were grown in complete supplemented with 500 µg/mL G418. The HEK DHFRL1 shRNA down-regulated cell pools were grown in complete medium supplemented with 5 µg/mL puromycin, which are then treated with 500 µM IPTG. The HEK 293 cells were grown in 75 cm^2 vented sterile tissue culture flasks in a horizontal position in an incubator at 37°C with 5 % CO_2, 95 % humidified air.

2.6.4 Cell Line Subculture

2.6.4.1 Coriell Cell Line Subculture

Coriell cell lines were inspected daily, manually and via cell count (Section 2.6.7) using a haemocytometer and an Olympus CKX31 inverted microscope. All culture reagents were heated to 37°C before use unless otherwise stated. Cells were sub-cultured or re-fed with fresh medium every couple of days depending on how fast the cell line grew. The cell suspension was transferred into a 20 mL tube and the cells were spun down at 500 X g for 5 min. The cells were re-suspended in a volume of complete media to yield a minimum concentration of 200,000 viable cells / ml. Once cells had been sub-cultured three times, they could be used for experiments.

2.6.4.2 HEK 293 Cell Line Subculture

HEK 293 cell lines were inspected daily using an Olympus CKX31 inverted microscope. Once cell cultures had reached 60-70 % confluency, they were either used in experiments or were sub-cultured. All culture reagents were heated to 37°C before use unless otherwise stated. The residual growth medium was poured off from the flasks into a waste bottle. A 2 mL volume of trypsin-EDTA was added to the flask, in order to rinse the cells. The trypsin was removed from the flask and a 3 mL volume of trypsin-EDTA was added to the flask, in order to detach the cells from the flasks. The flasks were incubated at 37°C until the cells had detached from the flask surface (~10 min). Trypsin was inactivated by the addition of 5 mL of complete medium. To ensure a single cell suspension, cells were re-suspended thoroughly using a pipette. The cell suspension was transferred into a 20 mL tube and the cells were spun down at 500 X g for 5 min. The cells were re-

suspended in a volume of complete medium based on the cell density required. A 9 mL volume of complete medium was then added into a new cell culture flask and 1 mL of the cell suspension was added. Cells were maintained at 37°C in CO_2 5 %, humidified air 95 %.

2.6.5 Preparation of Frozen Stocks

Coriell cell line stocks were maintained and stored as follows. Cell clumps were dissociated by trituration and the number of viable cells was counted. The cell suspension was transferred into a 20 mL tube and the cells were pelleted by centrifugation at 500 X g for 5 min. The supernatant was discarded. The cell pellet was re-suspended in the appropriate volume of cold (4-10°C) freeze medium (RPMI 1640 with 20 % FBS and 6 % DMSO) to yield five million viable cells/mL. The cell suspension was dispensed in 1 mL aliquots into sterile, labelled cryotubes. The samples were stored at - 20°C for 20 min and then stored at - 80°C overnight. The samples were then transferred into liquid nitrogen.

HEK 293 cell line stocks were maintained and stored as follows. Cells were collected by trypsinisation as described in Section 2.6.4. The entire cell suspension was transferred into a 20 mL tube, and the cells were pelleted by centrifugation at 500 X g for 5 min. The supernatant was discarded. The cell pellet was re-suspended in the appropriate volume of cold cell freezing medium to yield 1 X 10^6 viable cells/mL. The cell suspension was dispensed in 1 mL aliquots into sterile, labelled cryotubes. The samples were stored at - 20°C for 20 min and then stored at - 80°C overnight. The samples were then transferred into liquid nitrogen.

2.6.6 Reconstitution of Frozen Cell Stocks

The cryovials containing Coriell cells were removed from the liquid nitrogen and the frozen cells were thawed out with agitation as quickly as possible in a water bath at 37°C. The cells were then transferred into sterile 20 mL tubes and 5 mL complete medium was added to the tube. Cells were counted, seeded and maintained in an upright position at 37ºC with CO_2 5 %, humidified air 95 % in 25 cm^2 tissue culture flasks. When cell lines reached confluency, they were passaged as described in Section 2.6.4.

The cryovials containing HEK cells were removed from the liquid nitrogen and the frozen cells were thawed out with agitation as quickly as possible in a water bath at 37°C. The cells were then transferred into sterile 20 mL tubes and 5 mL complete medium was added to the tube. Cells were spun down at 500 X g for 5 min and the supernatant was discarded. The pellet was re-suspended in 5 mL of complete medium and the cells were counted, seeded and maintained in a

horizontal position at 37ºC with CO_2 5 %, humidified air 95 % in 25 cm^2 or 75 cm^2 tissue culture flasks. When cell lines reached confluency, they were passaged as described in Section 2.6.4.

2.6.7 Cell Enumeration

The trypan blue dye exclusion test is based on the ability of viable cells to actively exclude dye, due to an intact cell membrane. As dead cells are unable to exclude the trypan blue dye, they appear blue, rather than clear, when examined under the microscope. A 500 µL volume of cells were transferred to a 1.5 mL tube eppendorf tube. A 100 µL volume of trypan blue was added to the cell suspension. The solution was mixed by vortexing and was left to sit for 5 min. After 5 min, the cell suspension was vortexed and a 10 µL volume was applied to the counting chamber of a haemocytometer (Hausser Scientific, PA, USA). The cells were then counted under 40 X magnification. The number of cells in the 4 outer chambers, made up of 16 squares each and the cells on 2 sides were counted. Four independent determinations were carried out and the mean was used to calculate the mean cell culture density.

Cell density was determined using the following formula:

$$\text{Cells/mL} = (N/4) \times 1.2 \times 10^4$$

Where,

N =cell number counted 4 = number of fields counted

1.2 = dilution factor 10^4 = constant

2.6.8 Transfection of HEK 293 cells using Lipofectamine

Cells were seeded at a cell density of 2 x 10^5 cells per well in complete medium in a 6-well plate and incubated at 37°C for 24 h. Plasmid DNA (4 µg) was added to a final volume of 50 µL of serum free medium in an eppendorf tube. A 10 µL volume of Lipofectamine 2000 was added to 40 µL of serum free medium in an eppendorf tube. Both tubes were incubated at room temperature for 5 min. The two tubes were mixed together and the mixture was incubated at room temperature for 20 min. The tubes were mixed well by vortexing and the mixture was added drop wise into the well. The plate was incubated at 37°C for 48 h. The transfection complexes were removed and replaced with 2 mL of medium containing the selection agent.

2.6.9 RNA Isolation from Cultured Cells

The Bioline Isolate II RNA Mini Kit composed of Lysis Buffer RLY, Wash Buffer RW1, Wash Buffer RW2, Membrane Desalting Buffer (MEM), Reaction Buffer for DNase I (RDN),

DNase I and RNase-free Water was used to isolate RNA from cultured cells. The protocol was carried out according to the manufacturers' instructions with slight alterations.

A total number of 5 X 10^6 mammalian cells were trypsinised and the pellets were collected by centrifugation. A 350 µL volume of lysis buffer RLY and 3.5 µL of β-ME were added to the cell pellet, and the cell pellet was re-suspended with a pipette and vortexed vigorously. The ISOLATE II Filter was placed in a collection tube and the lysate was loaded and centrifuged for 1 min at 11,000 X g. The ISOLATE II Filter was discarded and 350 µL of 70 % ethanol was added to the homogenised lysates. The lysate was mixed by pipetting up and down 5 times. The ISOLATE II RNA Mini Column was placed in a collection tube and the lysate was loaded on to the column. The column was centrifuged for 30 s at 11,000 X g. The column was placed into a new collection tube. A 350 µL volume of MEM was added and the column was centrifuged at 11,000 X g for 1 min. A DNase I reaction mixture was prepared by adding 230 µL of water to the DNase I and incubating at room temperature for 1 min. A 10 µL volume of reconstituted DNase I was added to 90 µL of RDN and was mixed by gentle flicking. A 95 µL volume of the DNase I reaction mixture was loaded directly on to the centre of the silica membrane and incubated at room temperature for 15 min. A 200 µL volume of Wash Buffer RW1 was added to the ISOLATE II RNA Mini Column and centrifuged for 30 s at 11,000 X g. The column was then placed into a new collection tube. A 600 µL volume of Wash Buffer RW2 was added to the ISOLATE II RNA Mini Column and centrifuged for 30 s at 11,000 X g. The flow through was discarded and the column was placed back into the collection tube. A 250 µL volume of Wash Buffer RW2 was added to the ISOLATE II RNA Mini Column and centrifuged for 2 min at 11,000 X g. The column was placed into a nuclease-free 1.5 mL collection tube. A 50 µL volume of RNase-free water was added directly to the column and the tubes were incubated at room temperature for 1 min. The tube was then centrifuged at 11,000 X g for 1 min. The isolated RNA was stored at -80°C.

2.6.10 Quantifying RNA Concentration

RNA concentrations were measured using the NanoDrop 1000 Spectrophotometer (Thermo Scientific). The NanoDrop was blanked with RNase-free water that was previously used to elute the RNA. The RNA parameter was chosen on the NanoDrop instrument and 1 µL of each sample was loaded onto the optical pedestal and the sample was then drawn into a column and measured. UV-Vis spectrum results were displayed on the computer and the RNA concentrations were measured as ng/µL with an expected A260/280 ratio of 2, indicating acceptable RNA purity within the samples. The samples were then stored at -80°C.

2.6.11 DNase Treatment of RNA

A 2 µg concentration of RNA was added to nuclease-free water to give a final volume of 8 µL in a PCR tube. A 1 µL volume of Sigma 10 X reaction buffer and a 1 µL volume of Sigma amplification grade DNase I was added to the tube. The reaction solution was mixed gently by flicking and the tube was then incubated at room temperature for 15 min. A 1 µL volume of Sigma stop solution was added to the reaction in order to bind calcium and magnesium ions and to inactivate the DNase I. The reaction was then incubated at 70 °C for 10 min on a thermal cycler. The reaction was then chilled on ice / stored at -20 °C.

2.6.12 RNA Sample Preparation for Agarose Gel Electrophoresis

RNA samples were kept on ice at all times. A 1 µg concentration (one volume) of RNA sample was added to two volumes of RNA sample loading buffer in a PCR tube. The tubes were then heated to 65°C for 10 min in a thermal cycler to denature the RNA. Samples were then loaded onto as agarose gel was as described in Section 2.2.1.

2.6.13 Reverse Transcription Polymerase Chain Reaction (cDNA Synthesis)

High quality RNA was reverse transcribed to cDNA by the following method. A 1 µg (5.5 µL) concentration of DNase treated RNA was transferred into a PCR tube. A 2 µL volume of Bioline random hexamers primer mix (50 ng/µL) and 4 µL of Bioline Oligo (dT) 18 primer mix (50 µM) was added to each sample and the tubes were centrifuged briefly to ensure that the solution was mixed. The samples were then heated at 70°C for 5 min on the thermal cycler to denature the RNA. The samples were then cooled on ice for at least 1 min. A master mix of reverse transcription reagents was prepared per sample as described in Table 2.25.

Table 2.25. Master Mix for reverse transcription

Component	Volume (µL)
Nuclease free H_2O	1.5
RNAsin (RNaseOUT Ribonuclease Inhibitor) (40µg/µL)	1
10 mM dNTPs	1
5 x Reaction Buffer	4
BioScript (Reverse Transcriptase Enzyme) (200U/µL)	1

An 8.5 µL volume of master mix was added to each sample. The samples were centrifuged briefly to homogenise the mixture. cDNA synthesis was then carried out on a thermal cycler. The samples were heated to 25°C for 10 min, followed by 42°C for 60 min and 70°C for 15 min. Samples were stored at 4°C.

2.6.14 Genomic Contamination of RNA Samples Test – MTHFD1 R653Q PCR Assay

The MTHFD1 R653Q assay is an intron spanning assay, producing a PCR product of 330 bp when amplifying genomic DNA and a PCR product of 232 bp when amplifying cDNA. A master mix of PCR reagents was prepared per sample as described in Table 2.26. Primer sequences are shown in Table 2.27.

Table 2.26. Master Mix for MTHFD1 R653Q PCR Assay

Component	Volume (µL)
Forward primer (5 ρmol/µL) (IDT)	0.3
Reverse primer (5 ρmol/µL) (IDT)	0.4
10X buffer	5
2.5mM dNTPs	4
25 mM $MgCl_2$	3
Taq (5U/µL)	0.2
dH_2O	36.1

Table 2.27. MTHFD1 R653Q Polymorphism Primer sequences

Gene	Primer Sequence
MTHFD1 R653Q	Fwd: 5′- CACTCCAGTGTTTGTCCATG -3′ Rev: 5′- GCATCTTGAGAGCCCTGAC -3′

A 1 µL volume of cDNA / genomic DNA was used per reaction, giving a final volume of 50 µL. A negative-template control using 1 µL H_2O instead of DNA was also included. Template DNA was initially denatured at 95°C for 3 min, followed by 35 amplification cycles, in an automated ABI 9700 DNA Thermal Cycler. Each cycle consisted of template denaturation (94°C for 30 s), primer annealing (58°C for 1 min) and extension (72°C for 1 min). This was followed by

an elongation step to complete the amplification cycle (72°C for 10 min). Samples were maintained at 4°C overnight. PCR products were run on a 2 % agarose gel, as described in Section 2.2.1.

2.6.15 Real Time – quantitative Polymerase Chain Reaction

RT-qPCR assays were carried out in Lightcycler® 480 96 multi-well white plates on the Lightcycler® 480 and were designed using the Universal Probe Library (Roche). Primer and probe sequences are shown in Table 2.28. A master mix of PCR reagents was prepared per sample as described in Table 2.29.

Table 2.28 Primer sequence and probe number for RT-qPCR

Assay Target	Primer Sequences
DHFRL1	Fwd: 5′- CGGACCTTAGAAAGTCACACATC-3′ Rev: 5 ′-TCACAGGAGAATCACTTCAACC-3′ Probe #89
GAPDH	Fwd: 5′-CTCTGCTCCTCCTGTTCGAC-3′ Rev: 5′-ACGACCAAATCCGTTGACTC-3′ Probe #60

Table 2.29. Master Mix for RT-qPCR assay

Component	Volume (µL)
Forward primer (20 ρmol/µL) (IDT)	0.16
Reverse primer (20 ρmol/µL) (IDT)	0.16
Probe	0.16
SensiMixII	8
dH_2O	5.52

A 1 µL volume of cDNA was used per reaction, giving a final volume of 15 µL. All samples were run in triplicate. A negative-template control using 1 µL H_2O instead of DNA was also included. Template DNA was initially denatured at 95°C for 5 min, followed by 45 amplification cycles. Each cycle consisted of template denaturation (95°C for 30 s), primer annealing (60°C for 30 s) and extension (72°C for 1 s), with signal acquisition. Results were calculated by applying the comparative E-method (Tellmann, 2006) to measure the fold change in target gene expression relative to the reference gene.

2.6.16 Proteomics Analysis

Label-free LC-MS Analysis was carried out by our collaborators, Dr. Paula Meleady and Mr. Michael Henry from the National Institute for Cellular Biology (NICB) in DCU.

2.6.16.1Sample Preparation

Cell pellets were lysed with 300 µL of lysis buffer (7 M urea, 2M thiourea, 30 mM Tris, 4 % CHAPS, pH 8.5) and were homogenised by passing the samples through a 20 gauge needle five times. Samples were then placed on a shaker for 1 h to allow extraction to take place. Samples were then centrifuged at 10,000 X g for 15 min at 4°C to remove any insoluble material. The supernatant was transferred into a new tube and protein concentrations were determined using the thiourea-compatible Quick Start Bradford Assay. A100 µg concentration of the samples was then cleaned using the Ready Prep 2-D clean up kit to remove any trace of detergents. The cell pellet was then re-suspended in detergent free buffer (6M urea, 2 M thiourea, 10 mM Tris, pH 8.0). Protein concentrations were determined again using the thiourea-compatible Quick Start Bradford Assay. Protein concentrations were adjusted to a concentration of 2 µg/µL using the detergent free buffer. Proteins were then digested by adding 10 µg of protein (5 µL) to 35.4 µL of 50 mM ammonium bicarbonate and 0.5 µL of 1 % Protease MAX Surfactant. The proteins were reduced with DTT (final concentration of 5 mM) by incubating the samples at 56°C for 20 min and then allowing them to cool to room temperature for 20 min. Samples were alkylated by adding Iodoacetamide (final concentration 15 mM) and incubating the samples at room temperature in a dark room for 15 min. Digestion was carried out with sequence grade Lys-C at a ratio of 1:20 (w/w) Lys-C:Protein at 37°C for 6 h, followed by a second digestion with sequence grade trypsin at a ratio of 1:25 (w/w) Trypsin:Protein at 37°C overnight. The digestion was stopped by adding 2 µL of a 50 % trifluoroacetic acid (TFA) / 50 % water solution. Samples were then cleaned up using Pep-clean C18 spin column dried under a vacuum and stored at -80°C. Prior to mass spectrometry, the dried peptides were re-suspended in 50 µL of 0.1 % TFA in 2 % acetonitrile (CAN), vortexed and sonicated to ensure an even suspension.

2.6.16.2Label-free LC-MS Quantitative Profiling

Nano LC-MS/MS analysis was carried out using an Ultimate 3000 nano LC system (Dionex) coupled to a hybrid linear ion trap/Orbitrap mass spectrometer (LTQ Orbitrap XL; Thermo Fisher Scientific). A 5 µL volume of digest was loaded onto a C18 trap column (C18 PepMap, 300 µm ID x 5 cm, 5 µm particle size, 100 Å pore size; Thermo Fisher) and desalted for

10 min using a flow rate of 25 µL/min in 0.1 % TFA. The trap column was then switched online with the analytical column (PepMap C18, 75 µm ID x 500 mm, 3 µm particle and 100 Å pore size; Thermo Fisher) and peptides were eluted with the following binary gradients of solvent A and B: 0-25 % solvent B in 240 min and 25-50 % solvent B in a further 60 min, where solvent A consisted of 2 % CAN and 0.1 % formic acid in water and solvent B consisted of 80 % CAN and 0.08 % formic acid in water. Column flow rate was set to 300 nL/min. Data was acquired with Xcalibur software, version 2.0.7 (Thermo Fisher Scientific). The mass spectrometer was operated in data-dependent mode and externally calibrated. Survey MS scans were acquired in the Orbitrap in the 300-2000 m/z range with the resolution set to a value of 60,000 at m/z 400. Up to seven of the most intense ions (1+, 2+ and 3+) per scan were CID fragmented in the linear ion trap. A dynamic exclusion window was applied within 40 s. All tandem mass spectra were collected using a normalised collision energy of 35 %, an isolation window of 3 m/z, and one microscan.

Chapter 3

Examination of Candidate Folate Sensitive Differentially Methylated Regions (FS-DMRs) in the FASSTT study

3.1 Introduction

DNA methylation is the most commonly studied form of epigenetic modification in the mammalian genome and is thought to play a part in regulating gene expression. The aim of this chapter is to identify candidate folate-sensitive differentially methylated regions (FS-DMR) in the literature that have previously been shown to be subjected to DNA methylation changes due to changes in one carbon metabolism markers and to screen these candidate sites in a folic acid intervention trial of pregnant women. This will be carried out in order to assess whether prolonged exposure to folic acid supplementation during pregnancy can result in changes in DNA methylation patterns. This will be carried out using a technique known as MS-HRM. The folic acid intervention study will also be assessed for the DNA methylation patterns in the promoter regions of *DHFR* and *DHFRL1*, two important genes that are involved in *de novo* thymidylate biosynthesis in the folate mediated one-carbon metabolism pathway (Figure 1.1) (See Section 1.11 and 1.12). *DHFR* is involved in *de novo* thymidylate biosynthesis /folate metabolism in both the nucleus and the cytoplasm whereas *DHFRL1* localises to the mitochondria and takes the place of *DHFR* in *de novo* thymidylate biosynthesis in the mitochondria. Promoter methylation will be assessed using a technique known as SMART-MSP.

3.1.1 Nutrition and DNA Methylation

Nutrition is well-known to play a major role in health and disease, with a wealth of literature supporting the "fetal origins of disease" hypothesis (Gluckman et al., 2008) that proposes that the *in utero* environment has a significant role in an individual's risk of a number of diseases later in their adult life (Barker et al., 1989). Although the mechanism behind how the *in utero* environment can influence disease is yet unknown, DNA methylation has been a proposed mechanism. However, DNA methylation is a new area of research, and therefore nutrition sensitive methylation sites in humans have only recently begun to be identified (Parle-McDermott and Ozaki, 2011).

Studies such as the Dutch Hunger Winter study provide an excellent opportunity to examine the effects of nutrition on health and disease (Lumey et al., 1993). The Dutch Hunger Winter was a time during the end of World War II, when food restrictions were put in place in the western part of the Netherlands. In spite of this, excellent medical records were kept, allowing those who were exposed to the famine prenatally to be traced (Schulz, 2010). Research carried out on this cohort of samples by Heijmans *et al.* (2008), showed that those individuals who were exposed to the famine prenatally had significantly less DNA methylation at the imprinted *Igf2* gene in comparison

to their unexposed, same sex siblings 60 years later. However, individuals who were exposed to the famine during the late gestation period showed to have no effect, suggesting that it is an early developmental effect (Heijmans et al., 2008). Another study examined additional loci on this same cohort of samples (Tobi et al., 2009). This study found that the *INSIGF* gene had decreased methylation, while the *IL10, LEP, ABCA1, GNASAS* and *MEG3* genes had increased methylation in those individuals who were exposed to the famine prenatally. However, late gestational exposure only showed DNA methylation changes in the *GNASAS* and *LEP* genes in men only (Tobi et al., 2009). These studies indicate that nutritionally induced changes are not just limited to the peri-conceptional period, and show both increases and decreases in methylation as well as displaying some sex-specificity (Parle-McDermott and Ozaki, 2011).

Aside from the Dutch Hunger Winter study, a number of other observational and intervention studies have been carried out, which have mostly focused on methyl donors. Correlations have been found between folate status and global DNA methylation levels in peripheral blood mononuclear DNA (Friso et al., 2002). The study showed that individuals who have the *MTHFR* 677 TT genotype have a significantly lower level of global DNA methylation in comparison to those who have the *MTHFR* 677 CC genotype. As the *MTHFR* 677C>T polymorphism is a well-known disease associated variant that converts 5, 10-Methylenetetrahydrofolate to 5-Methyltetrahydrofolate at a slower rate when folate status is low, it was thought that the supply of methyl groups may be compromised in individuals with the *MTHFR* 677 TT genotype and this was shown to correlate with global DNA methylation patterns (Friso et al., 2002, Parle-McDermott and Ozaki, 2011). Other folic acid intervention studies have also found differences in global DNA methylation status (Jacob et al., 1998, Rampersaud et al., 2000), as well as DNA methylation changes in specific genes such as *Igf2* (Steegers-Theunissen et al., 2009).

The most relevant study to our project has been carried out by Fryer *et al*. (2011), who examined human cord blood samples using the Illumina Infinium Methylation 27K BeadArray, a large step towards genome wide analysis. They also measured serum folate and homocysteine levels, an inverse indicator of folate status and methyl supply. They analysed 27,578 CpG loci that were associated with 14,496 genes. Similar to previous findings, they found that the majority of CpGs within the CpG islands were hypomethylated while those outside of the CpG islands were methylated. From their analysis, they identified that DNA methylation of 12 genes (*EIF2C3*, *ZBTB11*, *BDH2*, *ZNF187*, *RUNX1T1*, *C9orf64*, *PDE2A*, *MGC33486*, *AMN*, *ZPBP2*, *FBN3* and PVRL2) was directly correlated to homocysteine levels, while DNA methylation of five genes

(*ATP5F1*, *CYP26C1*, *FSTL3*, *MDS032* and *BMX*) was inversely correlated (Fryer et al., 2011). As the cord blood samples were taken at birth, it suggests that late gestation could be an important period where nutrition can influence / induce DNA methylation changes.

Despite the fact that the numbers of studies that have been carried out to date in order to identify nutrient sensitive DMRs in humans have been restricted, due to the availability of samples and to ethical issues, a number of observational and intervention studies have reported changes in DNA methylation in response to nutrient status, with the above studies revealing how folate status can play a role in influencing the pattern of DNA methylation both *in utero* and in adults. Although a number of candidate methylation sites have emerged from these studies, we are only beginning to decipher the human epigenome and identify those DNA methylation sites that are altered by nutrition. However, the pattern that seems to be emerging from the studies carried out to date, is that nutrient mediated DNA methylation changes are gene specific, site specific, tissue specific and age specific. In light of this, it is therefore, much more significant to examine specific known regions of the genome or to carry out a genome wide screen rather than measuring global DNA methylation patterns, measuring the methylation levels of multiple DNA repetitive sites, which seems to have been the case in most of the previous studies carried out to date, which have limited relevance. Technologies such as DNA microarray based technology, next generation sequencing after sodium bisulfite treatment, SMART-MSP and MS-HRM have now been developed and are available to measure genome wide or gene specific DNA methylation patterns. Therefore, the identification of nutrient-sensitive DMRs is likely to explode in the coming years (Parle-McDermott and Ozaki, 2011). We aim to begin to assess the impact of folic acid supplementation on DNA methylation patterns using a candidate gene / region approach.

3.1.2 The FASSTT Study

It is well known that folate supplementation pre-conception and during the first term of pregnancy reduces the risk of NTDs (1992). However, no official folic acid recommendations exist after the first trimester of pregnancy. Therefore, McNulty *et al*., (2013) carried out a study in order to investigate the effect of continued folic acid supplementation after the first trimester of pregnancy on maternal folate and homocysteine responses and related effects in the new born (McNulty et al., 2013). The study was carried out on the Folic Acid Supplementation during the Second and Third Trimester (FASSTT) samples, a 26 week, double blind, randomised placebo controlled intervention trial of pregnant women during the second and third trimester of pregnancy, and consists of a total of 357 blood samples. This study was led by Prof. Helene McNulty and Dr.

Kristina Pentieva of the Northern Ireland Centre for Food and Health, School of Biomedical Sciences, University of Ulster, Northern Ireland. Women who took part in this study were between 18-35 years of age with singleton pregnancies with no known complications between 2005 and 2006. All women took folic acid supplements in the first trimester of their pregnancy. A total number of 119 women completed the study. After the first trimester, the women were divided into two groups. A total of 59 women received 400 µg of folic acid supplement per day and a total of 60 women received a placebo. Blood samples were taken at baseline (after the first trimester) and post-intervention (week 36) as well as from the cord blood at delivery (Figure 3.1). McNulty *et al*., (2013) found that those women who did not continue on with the folic acid supplementation after the first trimester had a decrease in their serum folate levels and an increase in their plasma homocysteine levels, whereas those women who were on the supplementation had a significant increase in their red blood cell folate levels. Cord blood folate levels were also found to be significantly higher in the folic acid supplementation group when compared to the placebo group. Whether these effects have benefits for pregnancy outcomes or early childhood still needs to be determined (McNulty et al., 2013).

In our study, genomic DNA samples were isolated from the buffy coats of all the blood samples and the DNA samples were used to screen candidate FS-DMRs, as well as *DHFR* and *DHFRL1* by two different DNA methylation analysis methods as described below (Section 3.1.3 and Section 3.1.4).

3.1.3 Methylation Sensitive - High Resolution Melting (MS-HRM)

MS-HRM is a PCR based assay that has been specifically adapted to study DNA methylation following sodium bisulfite treatment (Wojdacz et al., 2008). High resolution melting is a technique that was initially developed for genotyping by comparing the melting profiles of sequences that differ in their nucleotide composition. Each double-stranded DNA molecule has a specific melting temperature, a temperature at which the DNA helix separates into two single strands. During high resolution melting, the PCR products, which are bound to an intercalating dye, are exposed to an increasing temperature gradient, emitting fluorescence. When the melting temperature of the PCR product is reached, the DNA separates into two single strands, preventing the binding and fluorescing of the dye. This results in a sharp drop in fluorescence, allowing the melting profile of the PCR products to be determined (Wittwer et al., 2003).

In MS-HRM, sodium bisulfite treatment, a chemical modification that converts all the unmethylated cytosine residues to uracil, while leaving the methylated cytosines as cytosines has to take place before high resolution melting (Figure 3.2) as methylation marks are removed from genomic DNA by DNA polymerase during PCR amplification. Therefore, bisulfite treatment introduces specific changes in the DNA sequence that depends on the methylation status of the individual cytosine residues. As methylated and unmethylated DNA acquire different sequences after bisulphite treatment, this results in PCR products with noticeably different melting temperatures and profiles (Wojdacz et al., 2008).

Before PCR amplification can take place, PCR primers need to be designed carefully in order to ensure unbiased amplification of either the methylated or the unmethylated DNA. This is critical, as preferential amplification of either template (PCR bias) will lead to misinterpretation of the results. Amplification of DNA standards should result in one PCR product that is relatively rich in GC content originating from the methylated DNA, where methylated cytosines are kept and one PCR product that has little GC content, originating from unmethylated templates where all the cytosines have been converted into uracils. Once the PCR product is amplified, melt curve analysis can take place. MS-HRM allows the melt curve of your DNA of interest to be compared to a standard curve of DNA methylation standards so that the methylation status of your sample can be determined (Wojdacz et al., 2008).

3.1.4 Sensitive Methylation Analysis after Real Time-Methylation Specific PCR

Sensitive methylation analysis after real time–methylation specific PCR (SMART-MSP) is also a probe free PCR based assay that detects the level of DNA methylation at a single locus following sodium bisulfite treatment (Figure 3.2). However, in contrast to MS-HRM, SMART-MSP quantifies the levels of methylation using the data obtained from the real-time PCR, rather than that obtained from the melting profiles of the DNA samples, although the melting profiles are used to evaluate the amplicon; to identify false positives due to incomplete bisulfite conversion or false priming, i.e. the methylation levels are quantified based on the crossing point (C_p) values, the number of cycles required for the fluorescent signal to cross the threshold (Kristensen et al., 2008, Kristensen et al., 2009).

In the same way as MS-HRM, PCR primers for SMART-MSP must be designed carefully before PCR amplification can take place. In contrast to MS-HRM primers, where the primers are designed to amplify the target, independent of its methylation status, SMART-MSP primers must be

designed to only amplify up methylated DNA templates during PCR amplification. This can be achieved by including CpG sites at the 3′ end of the primers.

3.1.5 Objectives

In our study, assessment of candidate FS-DMRs, *DHFR* and *DHFRL1* in the FASSTT study was examined using the following methodology.

- ❖ Examination of the literature and choosing folate-sensitive gene candidates.
- ❖ Generation of MS-HRM assay for the candidate folate-sensitive genes.
- ❖ DNA extraction from the buffy coats of the FASSTT samples.
- ❖ Screening of FASSTT samples for all candidate FS-DMRs using MS-HRM.
- ❖ Generation of SMART-MSP assays for *DHFR* and *DHFRL1*.
- ❖ Screening of FASSTT samples for *DHFR* and *DHFRL1* using SMART-MSP.

3.2 Results

3.2.1 Examination of the literature and choosing folate-sensitive gene candidates

A literature survey was carried out on all human methylation sites that were reported to change in response to any environmental stimuli and a number of those that had been reported to change in response to folate-related markers were chosen to be examined. The FS-DMRs that were chosen were:

- *RASSF1A – ras association domain family member 1 alpha*
  - Folic acid supplementation increases *RASSF1A* methylation levels in blood samples (Vineis et al., 2011) (Examined by pyrosequencing).
- *EIF2C3 - eukaryotic translation initiation factor 2C, 3*
  - DNA methylation levels directly correlated with homocysteine levels in cord blood (Fryer et al., 2011) (Examined by Illumina bead array).
- *ATP5F1 – ATP synthase H^+ transporting, mitochondrial F_0 complex, subunit B1*
  - DNA methylation levels inversely correlated with homocysteine levels in cord blood (Fryer et al., 2011) (Examined by Illumina bead array).

As the MS-HRM assays proved to be quite difficult to develop and optimise, a MS-HRM assay was firstly optimised for the *ATM – ataxia telangiectasia mutated* gene, a gene whose primers had been previously designed by Wojdacz *et al*. (2008) to examine tumour methylation by MS-HRM (Wojdacz et al., 2008). While *ATM* is not considered to be a candidate FS-DMR, the feasibility of adopting the MS-HRM method to our laboratory could be assessed through this assay.

3.2.2 DNA extraction from buffy coats

DNA was extracted from the buffy coats of a total of 255 samples. Buffy coats are a fraction of whole blood that lies between the red blood cells and the plasma once the whole blood has been spun down and contain the white blood cells and platelets. DNA was extracted from the buffy coats using the Qiagen® FlexiGene DNA Kit as described in Section 2.2.3. The DNA proved to be difficult to extract as the buffy coats obtained were highly viscous. In order to ensure we had extracted clean DNA, all samples were run on a 1 % agarose gel (Figure 3.3) and were also nanodropped in order to determine the concentration of the DNA.

3.2.3 Screening of Candidate FS-DMR genes by MS-HRM

3.2.3.1 Generation of MS-HRM assays for the candidate folate-sensitive genes

The first step of MS-HRM assay development was to design primers that allow correction for PCR bias as the quantity of amplification achieved by methylated and unmethylated DNA controls can be compromised by PCR bias due to the fact that they have different base compositions. Wojdacz *et al*. (2008) showed that the inclusion of a limited number of CpGs into the primer sequence and optimisation of the annealing temperature compensates for PCR bias towards the unmethylated GC-poor strand in most cases by enabling selective binding of the primer to the methylated sequence, therefore enabling compensation for PCR bias that would otherwise normally favour amplification of the unmethylated template. Inserting CpGs into the primer also increases the sensitivity of the melting assay. Therefore, primers with a limited number of CpGs were designed carefully and in accordance to the guidelines given by Wojdacz *et al*. (2008) in order to ensure unbiased amplification of either the methylated or unmethylated DNA.

Primers for *ATM, RASSF1A, EIF2C3* and *ATP5F1* were designed in the proximal promoter CpG islands of each gene according to the guidelines (See Table 2.7 for primer sequences and Appendix C-F for primer sites in the proximal promoter CpG islands). Once primers were designed, methylated and unmethylated DNA standards were bisulfite treated using the Sigma Imprint DNA Modification Kit. To ensure that the primers were designed correctly and that the DNA standards had been bisulfite treated and could be amplified, the bisulfite treated DNA standards were amplified with *ATM* primers (Figure 3.4) (See Section 2.4.2)

Once amplification was confirmed, a range of methylation concentrations (0 %, 5 %, 10 %, 50 % & 100 %) were made up by a series of dilutions, mixing the relevant volume of methylated

DNA template to unmethylated DNA template to be used for the standard curve. All assays were optimized initially for annealing temperature (Ta), then for primer and $MgCl_2$ concentration (See Section 2.4.2). The most important aspect for optimization was for the amplification to be specific and therefore, lower primer concentrations worked best.

Although Wojdacz *et al*. (2008) had published the sequence of the *ATM* primers they had used in their study, optimal conditions for the MS-HRM assay were not given and therefore, had to be optimised in the lab. An optimised standard curve of the *ATM* MS-HRM assay is shown in Figure 3.5, with five different methylation concentrations (0 %, 5 %, 10 %, 50 % & 100 %). The *RASSF1A* assay was also optimised using the same five different methylation concentrations. An optimised standard curve of the *RASSF1A* MS-HRM assay is shown in Figure 3.6. The *EIF2C3* assay was also optimised using the five different methylation concentrations. However, the lower methylation ranges of 5 % and 10 % proved difficult to reproduce and were therefore left out of subsequent standard curves. An optimised standard curve of the *EIF2C3* MS-HRM assay is shown in Figure 3.7. The *ATP5F1* assay was also initially optimised using the five different methylation concentrations. However, the lower methylation ranges of 5 % and 10 % also proved difficult to reproduce and were therefore left out of the standard curve. An optimised standard curve of the *ATP5F1* MS-HRM assay is shown in Figure 3.8.

3.2.3.2 Screening of FASSTT samples for all candidate folate-sensitive genes

A total of 95 FASSTT Study samples, including samples taken from both pre- and post-folic acid intervention and those who had been given folic acid and those who had been given a placebo after the first trimester of pregnancy were chosen and were examined. A 500 ng concentration of DNA from each sample was bisulfite treated as described in Section 2.4.1 and the methylation levels of the *ATM, RASSF1A, EIF2C3* and *ATP5F1* gene were examined by MS-HRM (Figures 3.9-3.13). All samples were screened along with three methylation standards (0 % methylated, 50 % methylated and 100 % methylated) in order to produce a DNA methylation standard curve. This allowed us to determine the methylation status of the FASSTT samples by comparing the melt curve of the FASSTT samples to the standard curve of DNA methylation standards.

FASSTT samples, including both pre- and post-intervention samples with those given a placebo and those given a daily folic acid supplementation were screened for all four genes (the *ATM* gene, the *RASSF1* gene, the *EIF2C3* gene and the *ATP5F1* gene) by MS-HRM. All samples

were found to be unmethylated for the *ATM* gene, the *RASSF1* gene and the *EIF2C3* gene (Figure 3.9 – Figure 3.11). The FASSTT samples that were screened for the *ATP5F1* gene by MS-HRM were also found to be mostly unmethylated (Figure 3.12). However, a slight methylation was seen in a few of the samples as can be seen by the small methylation peaks seen in Figure 3.13. On the other hand, the small peaks seen may also be due to background noise. These need to be confirmed by sequencing.

3.2.4 Screening of *DHFR* and *DHFRL1* promoters by SMART-MSP

DHFR and *DHFRL1* promoters were screened by a method known as SMART-MSP as described in Section 3.1.4. SMART-MSP was chosen as the method of choice, as MS-HRM assays proved to be difficult to design and optimise as described above.

3.2.4.1 Generation of SMART-MSP assays for a control gene

The first step of SMART-MSP assay development was to develop a control assay. A control assay is required in order to account for the amount of input DNA going into the PCR after bisulfite treatment, so that we can normalise for the DNA input for each of our samples during quantification. The *COL2A1* gene was the gene chosen for our control assay as it had previously been used as a control gene for SMART-MSP and primers for this assay had been designed by Kristensen *et al*. (2008). The primers were designed so that any bisulfite treated DNA would be amplified regardless of its methylation status, i.e. this is a control gene for the amount of input DNA for subsequent normalisation. The *COL2A1* assay was initially optimized for melting temperature (Tm), then for primer and $MgCl_2$ concentration using bisulfite treated methylated and un-methylated DNA controls (See Section 2.4.4). Once these parameters had been optimised, the *COL2A1* assay was optimised using four different methylation concentrations (0 %, 10 %, 50 % & 100 %) in order to determine the sensitivity of the assay. A negative control was also included in the assay (See Section 2.4.4). All samples were carried out in duplicate. The C_p values of an optimised *COL2A1* SMART-MSP assay are shown in Table 3.1. C_p values are inversely proportional to the amount of DNA present in the sample, i.e. the lower the C_p value, the greater the amount of DNA in the sample. The *COL2A1* assay was initially optimised using methylated and un-methylated DNA standard samples. The melting curves of an optimised *COL2A1* SMART-MSP assay is shown in Figure 3.14., where a clean single melting peak is observed as expected for both methylated and un-methylated control DNA samples.

3.2.4.2 Generation of SMART-MSP assays for *DHFR* and *DHFRL1* genes

The first step of SMART-MSP assay development was to design primers that are selective as much as possible for methylated DNA templates, so that only methylated DNA templates will be amplified during PCR amplification. This was achieved by the inclusion of at least two CpGs, of which at least one was located at or near the 3′ end of the primer sequence. Non-CpG cytosines were also included near the 3′ end of the primer sequences in order to select against incompletely converted sequences. The primer sequences were also examined to ensure that no single nucleotide polymorphisms (SNPs) were present within the sequence as SNPs can interfere with the melting profile of the PCR product (Kristensen et al., 2008).

Primers for the *DHFR* gene were designed in the 5′ upstream and exon 1 region (promoter region) of the *DHFR* gene proximal promoter in accordance to the guidelines mentioned above. (See Table 2.7 for primer sequences and Appendix G for primer sites). Primers for the *DHFRL1* gene were designed by firstly examining the difference between the two *DHFRL1* transcripts (See Appendix H). The two transcripts differ in their 5′ UTRs but both produce the same protein sequence. Although transcript 1 is 111 bp longer than transcript 2, transcript 1 is missing 123 bp in a region of exon 1, which may be of importance as previous data generated in the lab has shown that *DHFRL1* transcript 2 is the transcript, which is being more highly expressed (McEntee et al., 2011). Therefore, SMART-MSP primers for *DHFRL1* were designed in this 123 bp region in accordance to the guidelines mentioned above. (See Table 2.7 for primer sequences and Appendix H for primer sites).

Once primers were designed, the assays were initially optimized for Ta, then for primer and $MgCl_2$ concentration using bisulfite treated methylated and unmethylated DNA controls. The most important aspect for optimization was to ensure that only methylated DNA templates were being amplified. Once the assay was optimised, a range of methylation concentrations (0%, 10%, 50% & 100%) were made up by a series of dilutions, mixing the relevant volume of methylated DNA template to unmethylated DNA template. These were then used to assess the quantitative properties and sensitivity of the assays.

The *DHFR* assay was optimised using four different methylation concentrations (0 %, 10 %, 50 % & 100 %). Although we did try lower methylation concentrations, we found that the reproducibility of samples with low levels of DNA methylation was not great and that sometimes only one of the replicates amplified. Therefore, we used a 10 % cut off point as our level of %

methylation detection. A negative control was also included in the assay. The Cp values of an optimised *DHFR* SMART-MSP assay are shown in Table 3.2. The *DHFR* assay was initially optimised using methylated and unmethylated DNA standard samples. The amplification and melting curves of an optimised *DHFR* SMART-MSP assay are shown in Figure 3.15. All methylated samples are represented by a clean singe peak at the same melting temperature as expected.

The *DHFRL1* assay was optimised using four different methylation concentrations (0 %, 10 %, 50 % & 100 %). A negative control was also included in the assay. The Cp values of an optimised *DHFRL1* SMART-MSP assay are shown in Table 3.3. The *DHFRL1* assay was initially optimised using methylated and unmethylated DNA standard samples. The amplification and melting curves of an optimised *DHFRL1* SMART-MSP assay are shown in Figure 3.16. All methylated samples are represented by a clean singe peak at the same melting temperature as expected.

3.2.4.3 Screening of FASSTT samples for *COL2A1, DHFR* and *DHFRL1*

A total of 252 FASSTT Study samples, including samples taken from both pre- and post-folic acid intervention and those who had been given folic acid and those who had been given a placebo after the first trimester of pregnancy were chosen and were examined. These samples were chosen so that we could compare the pre- and post-intervention methylation level differences within each woman, i.e. their A versus B sample. We felt that this was important due to the likely methylation differences between individuals and therefore thought it would be better to analyse the difference between pre- and post-intervention first, rather than grouping the methylation analyses of all the A's together and comparing it to the methylation analyses of all the B's grouped together. A 1 µg concentration of DNA from each sample was bisulfite treated using the Qiagen EpiTect Fast Bisulfite Kit as described in Section 2.4.3 and the methylation levels of the *COL2A1, DHFR* and *DHFRL1* gene were examined by SMART-MSP. All samples were carried out in duplicate. An example of each of these assays is shown in Figures 3.17 - 3.19 (See Appendix I & J for raw C_p values). All samples were screened along with the 100 % methylation standard as well as a negative control in order to ensure that the assay was working correctly and suitable for methylation quantification. Samples were screened twice for the *COL2A1* assay in order to ensure that similar results were being obtained, regardless of what day / what plates the samples were being run on.

3.2.4.4 Analysis of SMART-MSP data for *COL2A1, DHFR* and *DHFRL1* in the FASSTT samples

Out of the total 252 FASSTT samples screened for the *COL2A1*, *DHFR* and *DHFRL1* gene, a total of 66 samples were excluded from the analysis as either the pre- or post-intervention samples were missing, the concentration of DNA of the sample was too low to bisulfite treat or as the assay replicates could not be repeated. Therefore, our final analysis included a total of 196 samples (98 people).

The methylation levels for each of the samples for each gene was quantified relative to the methylated DNA template using the $2^{-\Delta\Delta Cp}$ real-time PCR quantification approach (Kristensen et al., 2008), where $\Delta\Delta Cp$ = unknown sample ($Cp_{target\ gene} - Cp_{control\ gene}$) – 100 % methylated DNA ($Cp_{target\ gene} - Cp_{control\ gene}$). This value was then put into the $2^{-\Delta\Delta Cp}$ equation and multiplied by 100 to get the methylation percentage of each sample (See Appendix I & J for raw analysis).

3.2.4.4.1 Analysis of SMART-MSP data for *DHFR* in the FASSTT samples

Methylation levels were found to differ highly between individuals for the promoter region of the *DHFR* gene that we examined, pre- and post-intervention as expected (Figure 3.20-3.21). Therefore, for our analysis, it was decided to examine the difference in methylation % between individuals in their pre- and post-intervention samples. A 10 % methylation cut off was used and the samples were grouped into one of three categories, 1) above a 10 % increase in methylation change after intervention, 2) no change in % methylation after intervention ($-10 \leq x \leq +10$) and 3) above a 10 % decrease in methylation change after intervention (Figure 3.22). These categories were chosen in order to take into account the limits of the SMART-MSP assay for methylation detection.

The samples were then divided up into those who were given a placebo after intervention and those who were given a daily folic acid supplement after intervention. The two groups were then subdivided into the three categories as above (Figure 3.23-3.24).

The samples that showed a methylation change of over 10 %, either an increase or a decrease after intervention were plotted as a graph in order to determine whether prolonged exposure to folic acid supplementation during pregnancy can result in changes in DNA methylation patterns (Figure 3.25-3.26). The bar charts indicated that prolonged exposure to folic acid supplementation during pregnancy does not result in changes in *DHFR* DNA methylation patterns

as differential methylation is seen regardless whether the women were given a placebo or a daily folic acid supplement after intervention (p-value 0.404), suggesting that DNA methylation patterns for *DHFR* differ between individuals irrespective of folic acid status.

3.2.4.4.2 Analysis of SMART-MSP data for *DHFRL1* in the FASSTT samples

Methylation for the *DHFRL1* gene was analysed in a similar fashion to that of the *DHFR* gene. The % methylation change pre- and post-intervention was calculated. However, when this was carried out, all samples regardless of whether the women were given a placebo or a daily folic acid supplement after intervention were found to lie in the no change category (Figure 3.27 - 3.28). The scatter plot and the bar chart indicate that the *DHFRL1* gene is most likely un-methylated in the region examined and that prolonged exposure to folic acid supplementation during pregnancy does not result in changes in *DHFRL1* DNA methylation patterns as no change in methylation is seen regardless of whether the women were given a placebo or a daily folic acid supplement after intervention.

3.3 Discussion

The *in utero* environment has been shown to have a significant role in influencing an individual's risk of a number of diseases later in their adult life (Barker et al., 1989). However, the mechanism behind how the *in utero* environment can impact on the risk of disease has yet to be deciphered but DNA methylation is thought to play a key role, with nutrition sensitive DMRs in humans beginning to be identified in recent years. The first aim of our study was to investigate the impact of folic acid supplementation during pregnancy on DNA methylation patterns by examining previously identified FS-DMRs in DNA from women who consumed a daily folic acid throughout their entire pregnancy and those who only consumed folic acid for the first trimester of pregnancy.

From examining the literature, we chose to examine the DNA methylation patterns of three genes, *RASSF1*, *EIF2C3* and *ATP5F1* in maternal DNA taken from both pre- and post-folic acid intervention, including both those who were given a daily folic acid supplement and those who were given a placebo after the first trimester of pregnancy. Although *ATM* had not previously been identified as a FS-DMR, it was also chosen to be examined as MS-HRM assays proved to be difficult to optimise and primers for the *ATM* region had previously been described (Wojdacz et al., 2008). Optimisation of the MS-HRM assays proved to be a difficult task, with optimisation of numerous variables (primer sequence, primer concentration, annealing temperature, $MgCl_2$ concentration) requiring to be optimised. The choice of gene region and primer design also caused a

lot of trouble and required a lot of optimisation. Primers had to be redesigned a couple of times in order to ensure that there were no non-specific peaks on the high resolution melting scans as well as to eliminate bias towards either the methylated or unmethylated DNA sequences.

Once the MS-HRM assays had been optimised, we analysed the methylation status of 95 FASSTT DNA samples, taken from both pre- and post-folic acid intervention, including both those who were given a daily folic acid supplement and those who were given a placebo after intervention, across the four genes (*ATM, RASSF1A, EIF2C3* and *ATP5F1),* with *ATM* used as a control assay. *RASSF1A* is a gene that has been associated with a tumour suppressor function and is required for death receptor-dependent apoptosis. *RASSF1A* interacts with a protein called CDC20, which is an activator of the anaphase-promoting complex. The interaction between *RASSF1A* and CDC20 results in the inhibition of APC activity and mitotic progression, resulting in the inhibition of cell proliferation by negatively regulating cell cycle progression at the G_1/S phase transition (Amin and Banerjee, 2012). *EIF2C3* is a gene that encodes a member of the Argonaute family of proteins and is thought to play a role in short-interfering-RNA (siRNA) mediated gene silencing (Carmell et al., 2002). *ATP5F1* is a gene that encodes a subunit of mitochondrial ATP synthase. *ATP5F1* catalyses the synthesis of ATP from ADP in the presence of an electrochemical gradient of protons across the inner membrane during oxidative phosphorylation.

From our analysis, we found that there was no change in the DNA methylation patterns in both groups, pre- and post-folic acid intervention. All the samples examined were also found to be unmethylated in the regions examined in the *ATM, RASSF1A* and *EIF2C3* gene (Figure 3.9-3.11). However, the *ATP5F1* gene proved to be somewhat interesting. Although the majority of the samples were also unmethylated in the *ATP5F1* gene region examined (Figure 3.12), there were a few samples that were seen to be slightly methylated in the *ATP5F1* region examined indicated by the small methylated peaks, (Figure 3.13) suggesting that it may be folate sensitive and affected by folic acid supplementation. However, the minor methylation patterns seen in these samples may also be due to background noise. This gene nonetheless, may be of interest and needs to be investigated further, possibly by sequencing a number of the methylated and non-methylated samples by Sanger sequencing. Apart from these few samples, the rest of the samples were found to be unmethylated for all four of the gene regions examined. Therefore, in order to preserve the precious sample set and to keep costs down, it was decided that the remaining FASSTT samples would not be screened and analysed.

Although the other genes investigated in our study were previously shown to be potentially folate sensitive in other studies (Fryer et al., 2011, Vineis et al., 2011), they showed no changes in DNA methylation pattern in our study. This could have been due to a possible number of reasons. Firstly, Fryer *et al*. (2011) examined methylation patterns in cord blood samples whose mothers were either on a folic acid supplemented diet or a non-folic acid supplemented diet during pregnancy and not the methylation patterns of the mothers before, during or after pregnancy. Therefore we do not know if folic acid supplementation had / has an effect on the DNA methylation patterns of mothers who are taking folic acid or whether the effect of folic acid supplementation is an *in utero* effect and only impacts on the DNA methylation pattern of their offspring. Another recent finding has also shown an association between pregnant women whom consumed folic acid supplement after their first trimester and epigenetic changes in the cord blood samples, with increased DNA methylation observed in the paternal copy of the *IGF2* gene and decreased methylation in the paternally imprinted *PEG3* gene (Haggarty et al., 2013). As we only examined the maternal FASSTT DNA samples, the changes in DNA methylation patterns seen by Fryer *et al*. (2011) in the cord blood samples may not have been seen. Examining the FASSTT cord blood samples may give us an insight into this and this may possibly be done as a future step of our research. Secondly, Fryer *et al*. (2011) assessed the methylation patterns of the genes using the Infinium Methylation 27K beadchip assay. However, the methylation patterns observed by the Infinium Methylation 27K beadchip were not confirmed by another method and therefore, these genes require further validation, especially as they only examined 12 cord samples in total, which increases the risk of false positive results. Thirdly, Vineis *et al*. (2011) examined DNA methylation patterns in blood samples from both lung cancer subjects and controls together and found that *RASSF1A* DNA methylation increased in response to folic acid supplementation in individuals who have never smoked. This indicates that DNA methylation patterns may be influenced by a number of dietary / lifestyle factors acting together and also on the blood sample type taken, i.e. blood from cancer patients vs. healthy subjects. *RASSF1A* is also a gene whose loss or altered expression is associated with the pathogenesis of a variety of different cancers, suggesting that it has a tumour suppressor function. The inactivation of this gene has been shown to be correlated with hypermethylation of its promoter region. As our samples were all healthy individuals, this may in part be another explanation as to why no methylation was found in any of our samples. Lastly, as the Infinium Methylation 27K beadchip only examines ~ 2 CpG sites per CpG island and the region examinable by MS-HRM only covers a small region of the gene, ~ 8-14 CpG sites, (See Appendix C-F), the CpG sites in the genes of interest that we examined may not have covered the same region examined by Fryer *et al*. (2011). The CpG sites examined by Fryer *et al*. (2011) are not indicated in

their paper and therefore we could not determine whether the regions they examined overlapped the regions we examined. In the case of the *RASSF1A* gene, the CpG sites examined by Vineis *et al*. (2011) differ to those we examined. This may explain the difference in the methylation patterns observed between the different samples. However, the methylation patterns of the genes observed above are in concordance with those within the literature, who have proposed that the genes above are hypomethylated in their promoter regions (Kristensen et al., 2012). No methylation was observed in the *RASSF1A* promoter region in normal peripheral blood mononuclear cells (Kristensen et al., 2012).

In conclusion, we can conclude that the region of the four genes (*ATM, RASSF1, EIF2C3* and *ATP5F1*) that we examined do not show any change in DNA methylation patterns between the pre- and post-folic acid intervention, including both those who were given folic acid and those who were given a placebo after the first trimester of pregnancy. In the future, we could examine subsequent genes that have been shown to be potential folate sensitive genes to further investigate the impact of folic acid supplementation on DNA methylation patterns, however, the quantity of DNA that was available to us was limiting and therefore could not be carried out as part of my project.

The second aim of our study was to investigate the effect of folic acid supplementation during pregnancy on the DNA methylation patterns in the promoter regions of *DHFR* and *DHFRL1*, two important genes that are involved in the folate mediated one-carbon metabolism pathway in DNA from women who consumed a daily folic acid throughout their entire pregnancy and those who only consumed folic acid for the first trimester of pregnancy.

As mentioned in Section 1.11, *DHFR* is an important enzyme in the folate mediated one carbon metabolism pathway and catalyses the reduction of DHF and folic acid to its biologically active metabolite, THF, in both the cytoplasm and the nucleus. THF can then accept and donate one-carbon groups for reactions that are necessary for DNA synthesis, cell proliferation and cellular methylation. On the other hand, *DHFRL1* localises to the mitochondria and takes the place of *DHFR* in mitochondrial *de novo* thymidylate biosynthesis as mentioned in Section 1.12 (Figure 1.1). As both these enzymes play an important role in the folate pathway, we wanted to determine whether prolonged exposure to folic acid would have an effect on the methylation patterns of these two genes. This was carried out by a method known as SMART-MSP.

For the SMART-MSP assays, we had to use a control assay, the *COL2A1* assay, an assay which is designed so that all bisulfite treated DNA will be amplified in order to normalise for the amount of input DNA after bisulfite conversion as the concentration of bisulfite treated DNA cannot be quantified by nanodrop. In order to ensure that the *COL2A1* assay is a true representation of the amount of input DNA, bisulfite treated FASSTT DNA samples were screened for this assay twice, on different plates and on different days in order to ensure that similar results were being obtained for both runs. The sensitivity of the SMART-MSP assays was tested using a standard dilution series of methylated DNA into unmethylated DNA. Although some papers have reported that their SMART-MSP assays are sensitive to as low as 0.1 % and 0.05 % DNA methylation (Kristensen et al., 2012), our assays proved to differ. We found that the reproducibility of the samples with low levels of DNA methylation were inconsistent and that sometimes only one of the replicates amplified. Therefore, we used a 10 % cut off point as our level of % methylation detection.

Once the SMART-MSP assays had been optimised, we analysed the methylation status of 255 FASSTT DNA samples, taken from both pre- and post-folic acid intervention, including both those who were given a daily folic acid supplement and those who were given a placebo after intervention, across the two genes (*DHFR* and *DHFRL1*). From our analysis, we found that the methylation patterns did change in the individuals' pre- and post-folic acid intervention, in both those who were given a placebo and those who were given a daily folic acid supplementation for the *DHFR* gene (Figure 3.20-3.21). However, as we expected methylation levels to differ between individuals, we worked out the percentage methylation difference in the pre- versus post-intervention sample for each individual first i.e., each woman's post-intervention DNA sample was only compared to her own pre-intervention DNA sample.

The *DHFR* SMART-MSP results (Figure 3.22-3.24) indicate that different individuals have different levels of % DNA methylation change in the promoter of their *DHFR* gene. Regardless of whether the women were given a placebo or a daily folic acid supplement after intervention, changes in the percentage of promoter *DHFR* methylation were seen between individuals in both the intervention and placebo groups suggesting that folic acid status does not affect *DHFR* DNA methylation patterns. Although, the exact reason for the difference in DNA methylation patterns observed is unknown, there are many factors that could have attributed to the differences observed. It may be due to defects in methyl metabolism or may just simply be due to the environment which can affect DNA methylation patterns (Kaminsky et al., 2009). As regions in the genome can also

autonomously determine DNA methylation states, this may be the cause of the differences seen (Lienert et al., 2011). The methylation observed in this region may also not be functional and what we are observing may just be passive methylation differences, DNA methylation that is a consequence of gene regulation (Gutierrez-Arcelus et al., 2013). The differences in the blood cell type between each sample may also be of concern as DNA methylation variation in human populations have recently showed that genetic variation can influence DNA methylation levels in different cell types (Gibbs et al., 2010, Bell et al., 2011) and therefore, the differences which we observed may simply be differences in blood cell type.

Previous studies carried out to determine the methylation status of the *DHFR* gene have shown that *DHFR* is under-methylated at its 5′-end promoter region, while completely methylated throughout the rest of the gene (Stein et al., 1983, Shimada and Nienhuis, 1985). The DNA methylation patterns observed were found to be similar in both sperm and all somatic tissue, a contrast to many of the tissue-specific genes that are highly methylated in sperm DNA and under-methylated in the tissues in which they are expressed. However, as *DHFR* is a key enzyme that is involved in the biosynthesis of purines and pyrimidines, it would be expected to be active and expressed in all cells and it has been shown to be constitutively expressed at low levels in all cell types. Although DNA methylation has been associated with transcriptional silencing, *DHFR* has been shown to be completely methylated throughout its coding regions and 3′-end, and is still actively expressed in all cells. This suggests that DNA methylation most likely exercises its major effects on sites that can be altered, with those located at and around the 5′ promoter region / exon 1 of the gene, i.e. promoter methylation can inhibit subsequent gene expression, whereas methylation of the coding region does not have an effect on transcription. This is consistent in other genes, such as the human gamma globin gene, which shows a strong correlation between its activity in certain differentiated cell types and the degree of demethylation around their 5′ sequences (Stein et al., 1983). The level of expression is said to be modulated during the cell cycle, increases with growth stimulation and decreases with terminal cell differentiation when the cells cease to proliferate (Hendrickson et al., 1980). As described in Section 1.6, the gene body could also be methylated as methylation occurring in the bodies of genes has recently been shown to be positively associated with active transcription (Hellman and Chess, 2007, Lister et al., 2009a, Jones, 2012).

However, our results suggest that the *DHFR* gene is in fact methylated at low percentages in the 5′ promoter region. However, SMART-MSP assays can only analyse a small number of CpG sites and therefore, we would have to design numerous assays in order to determine the methylation

status of the whole promoter. Sanger sequencing could also be used to determine the methylation status of all of the CpG residues in the *DHFR* promoter region.

The *DHFRL1* SMART-MSP results (Figure 3.27-3.28) indicate that all the individuals who took part in the FASSTT study have similar levels of *DHFRL1* methylation and that no change in methylation patters were found in both groups, pre- or post-folic acid intervention, regardless of whether the women were given a placebo or a daily folic acid supplement after intervention. All the FASSTT samples examined were also found to be practically unmethylated in the region examined pre- and post-intervention. This suggests that folic acid status does not have an effect on *DHFRL1* DNA methylation patterns. As *DHFRL1* plays a key role in mitochondrial *de novo* thymidylate biosynthesis, it is likely to be constitutively active in all cells and therefore, one would expect to see no methylation in its 5′ promoter region.

From the methylation studies carried out above by both MS-HRM and SMART-MSP, we can conclude that folic acid supplementation during the second and third trimester of pregnancy did not alter the DNA methylation patterns of individuals in the gene regions we examined, with most of the genes we examined being unmethylated pre- and post-intervention. However, this is not to say that folic acid supplementation does not result in DNA methylation pattern changes and it does have the potential to alter DNA methylation patterns in different genes / in different regions of the genes we examined and needs to be investigated further.

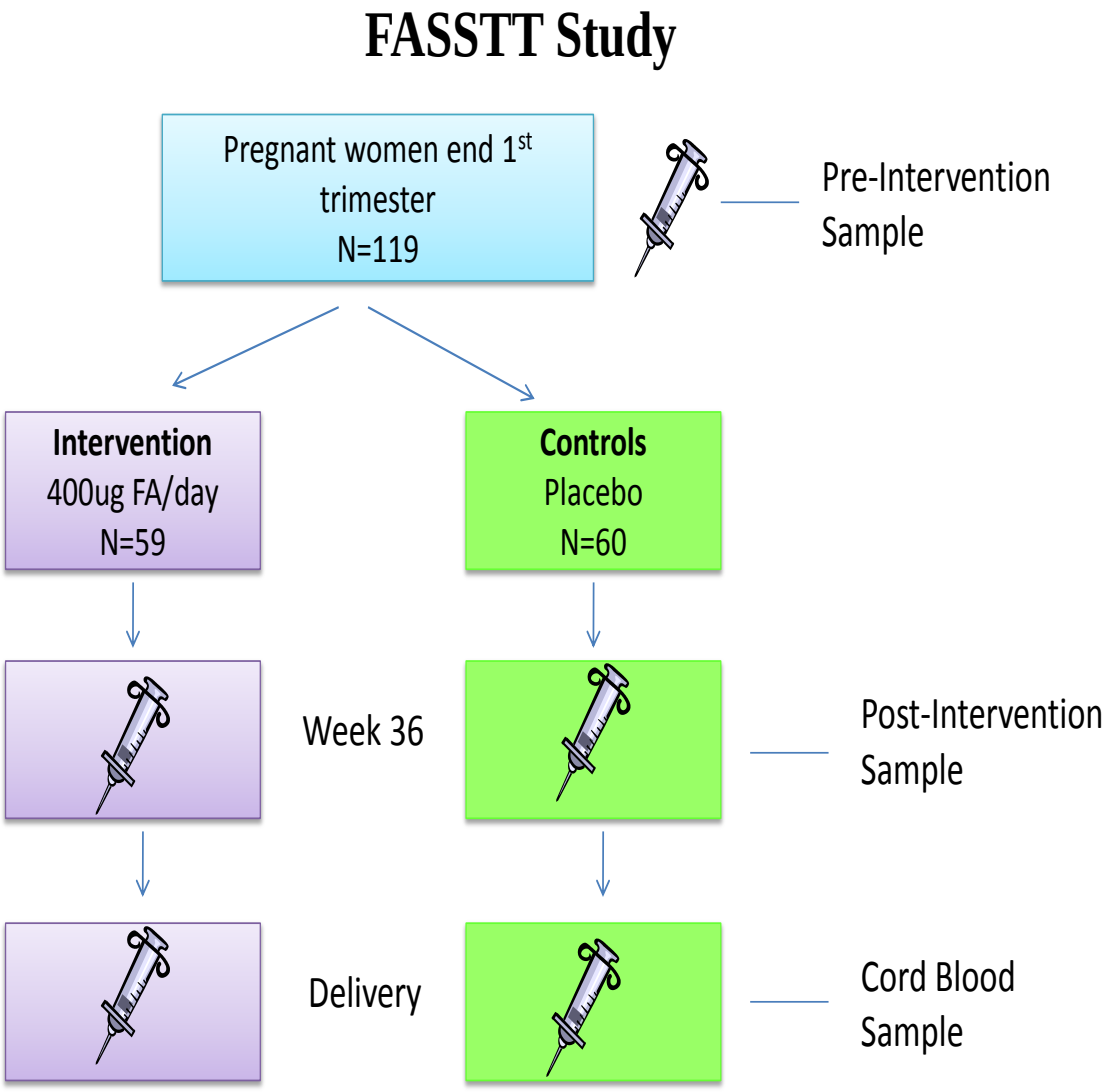


Figure 3.1. Overview of the FASSTT study. All women took folic acid supplements throughout their first trimester of pregnancy. Intervention began at the beginning of the second trimester, where one group was given 400 µg of folic acid per day, whereas the other group was given a placebo. Blood samples were taken at baseline and at week 36 and also from the cord blood at the time of delivery. Image adapted from Breige McNulty *et al*., unpublished.

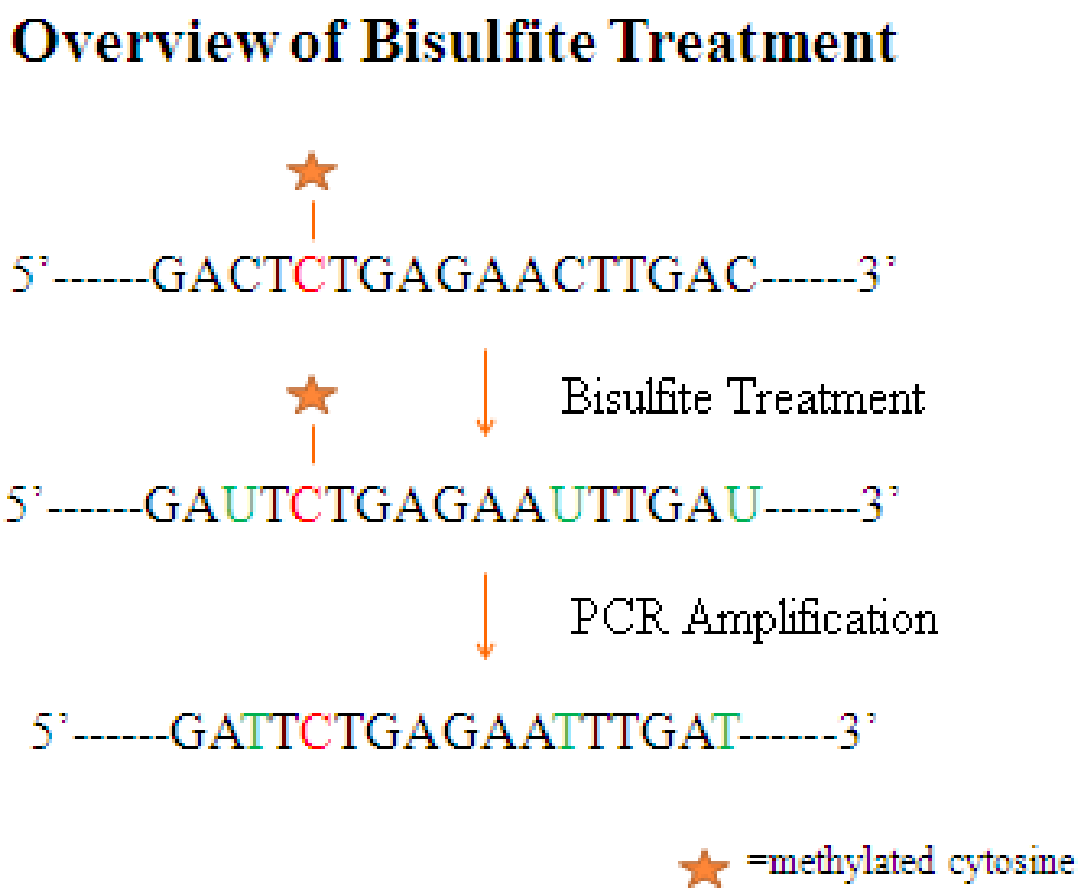


Figure 3.2. An overview of sodium bisulfite treatment. Bisulfite treatment converts the unmethylated cytosines to uracils while keeping the methylated cytosines as cytosines so that methylated and unmethylated cytosines can be distinguished after PCR amplification.

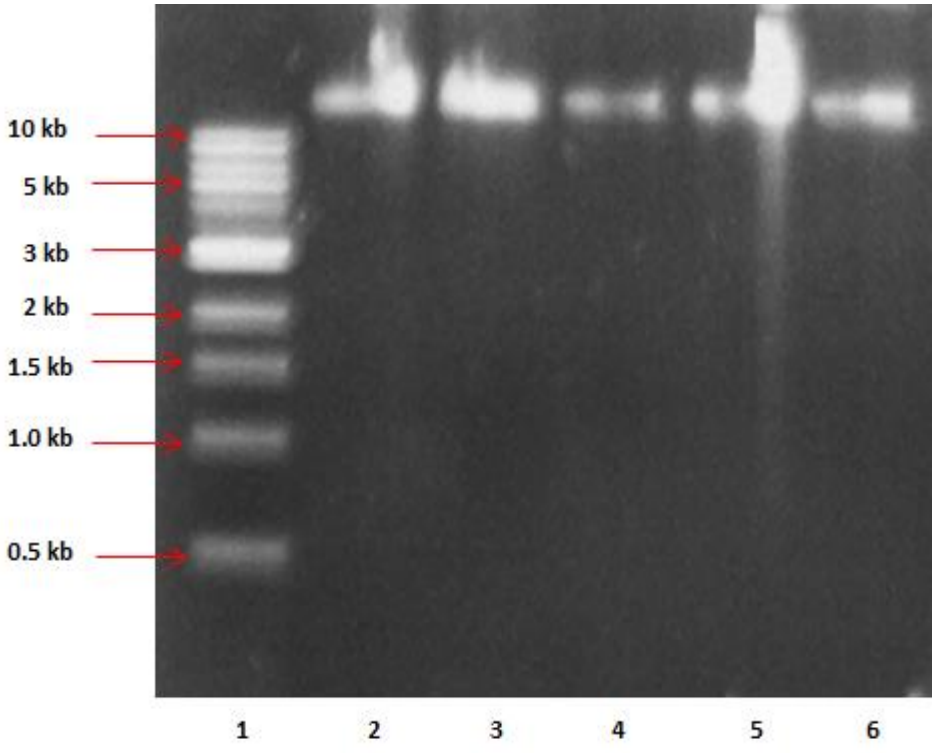


Figure 3.3. Gel electrophoresis of an example of the DNA extracted from buffy coats on a 1 % agarose gel. Lane 1 – 1 kb DNA ladder, Lane 2 – DNA Sample 1, Lane 3 – DNA Sample 2, Lane 4 – DNA Sample 3, Lane 5 – DNA Sample 4, Lane 6 – DNA Sample 5.

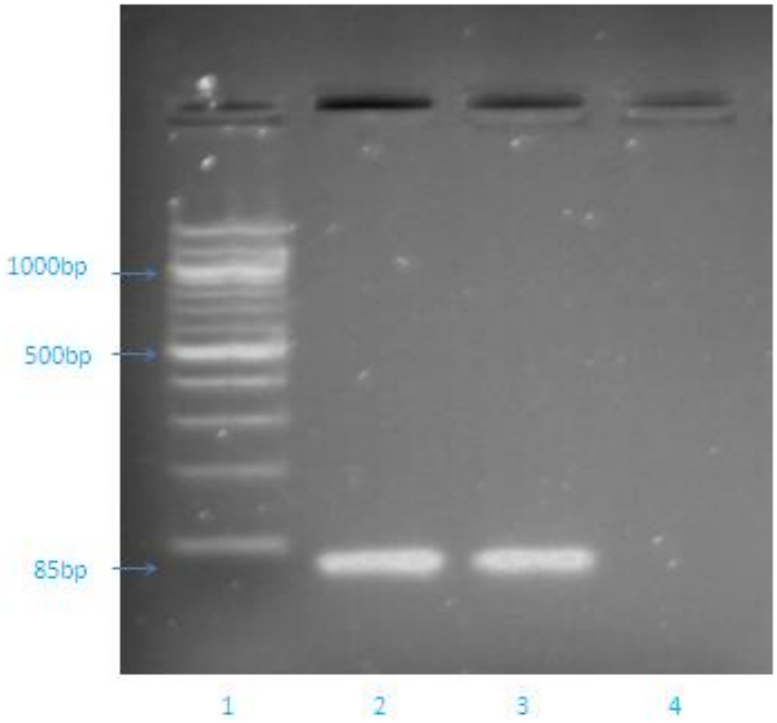


Figure 3.4. Gel electrophoresis of bisulfite treated DNA standards (500 ng) amplified with *ATM* primers on a 2 % agarose gel. Lane 1 – 100 bp DNA ladder, Lane 2 - methylated DNA standard, Lane 3 – unmethylated DNA standard, Lane 4 – negative control. Clear bands of the correct band (85 bp) were seen indicating that DNA amplification had occurred, confirming that the DNA standards had been bisulfite treated and that the primers were designed correctly.

ATM MS-HRM STANDARD CURVE

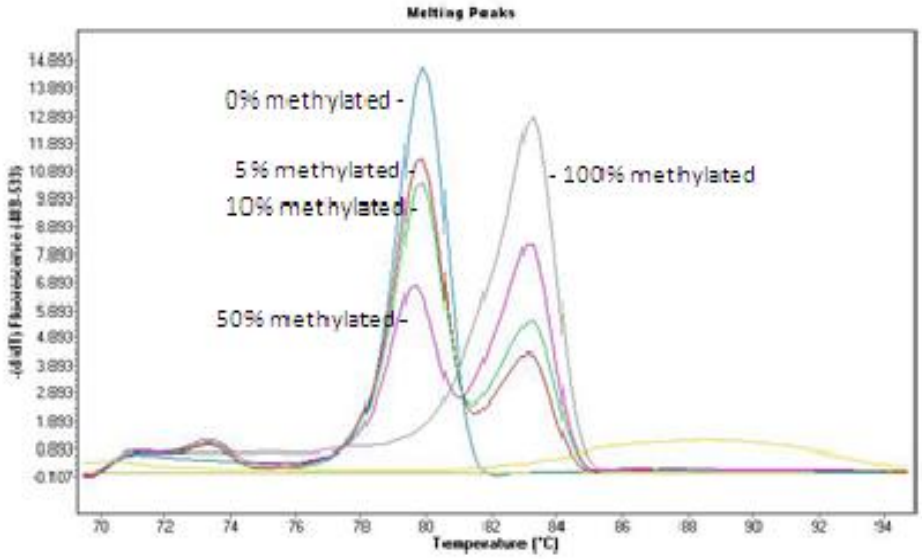


Figure 3.5. Standard MS-HRM curve for the *ATM* gene. The single blue peak represents non-methylated (0 % methylated) DNA, the two red peaks represent 5 % methylated DNA, the two green peaks represent 10 % methylated DNA, the two pink peaks represent 50 % methylated DNA and the single grey peak represents 100 % methylated DNA. The annealing temperature used for the *ATM* assay was 58°C.

RASSF1A MS-HRM STANDARD CURVE

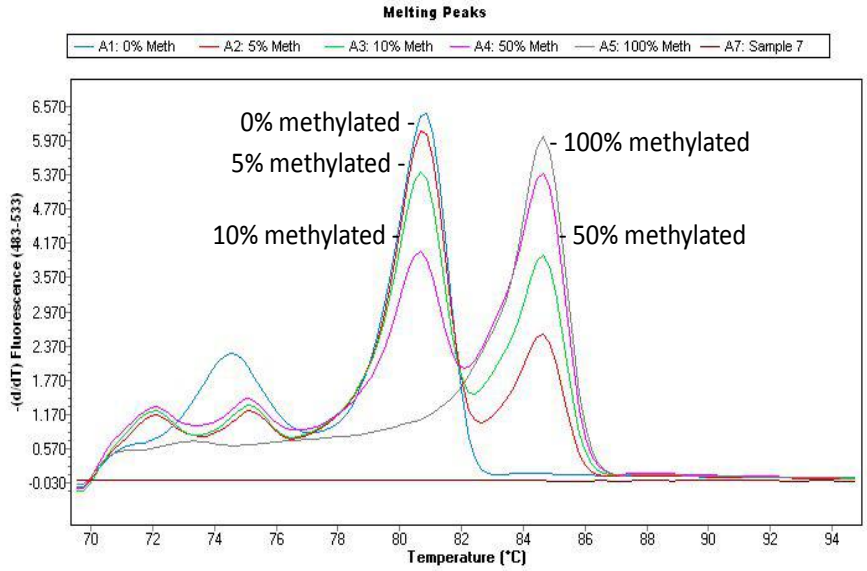


Figure 3.6. Standard MS-HRM curve for the *RASSF1A* gene. The single blue peak represents non-methylated (0 % methylated) DNA, the two red peaks represent 5 % methylated DNA, the two green peaks represent 10 % methylated DNA, the two pink peaks represent 50 % methylated DNA and the single grey peak represents 100 % methylated DNA. The small peaks seen prior to the non-methylated and methylated peaks may be due to background noise. The annealing temperature used for the *RASSF1* assay was 58°C.

EIF2C3 MS-HRM STANDARD CURVE

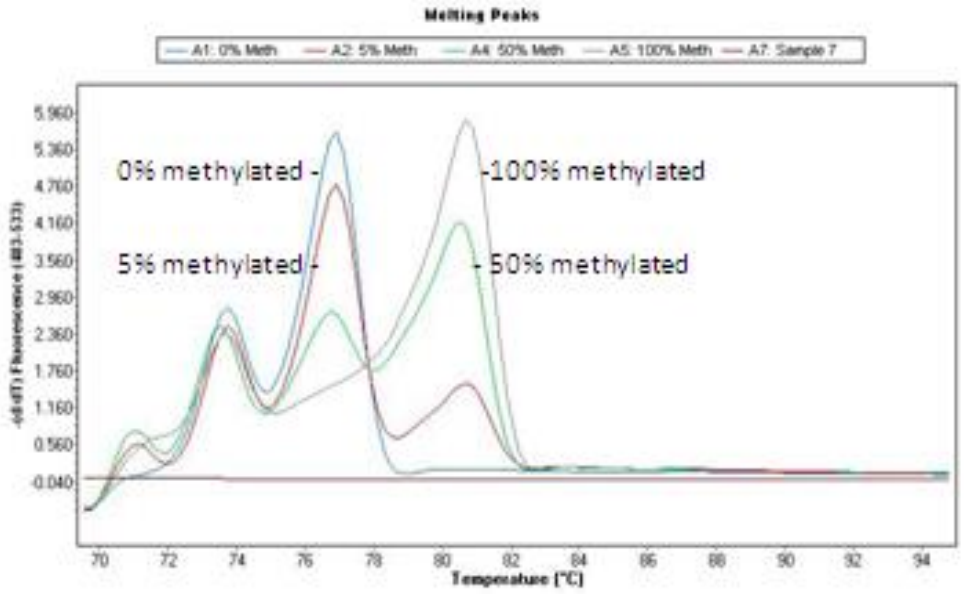


Figure 3.7. Standard MS-HRM curve for the *EIF2C3* gene. The single blue peak represents non-methylated (0 % methylated) DNA, the two red peaks represent 5 % methylated DNA, the two green peaks represent 50 % methylated DNA and the single grey peak represents 100 % methylated DNA. The small peaks seen prior to the non-methylated and methylated peaks may be due to background noise. The annealing temperature used for the *EIF2C3* assay was 52°C.

ATP5F1 MS-HRM STANDARD CURVE

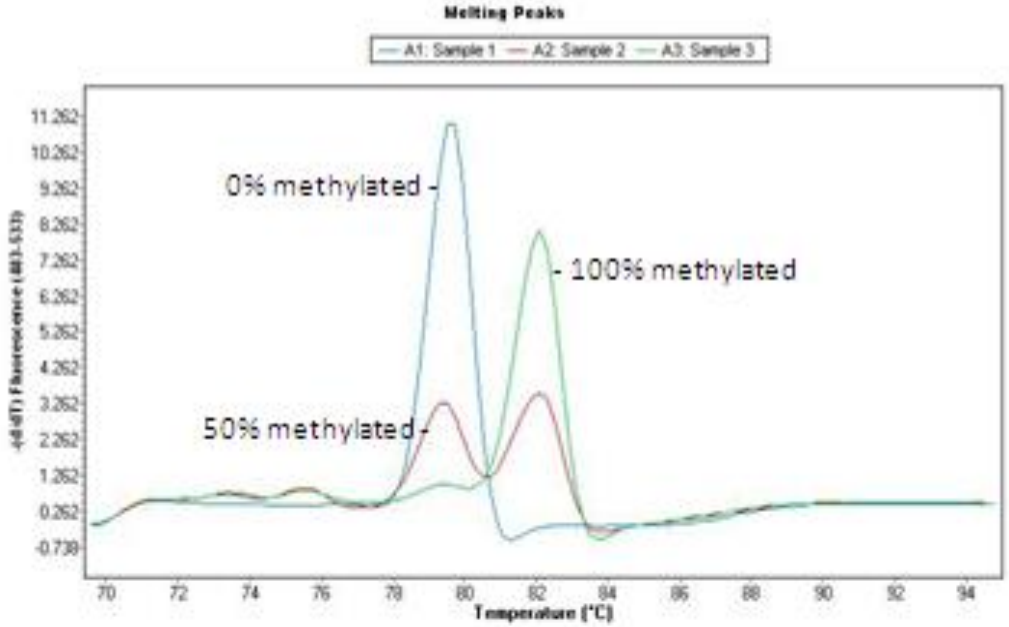


Figure 3.8. Standard MS-HRM curve for *ATP5F1*. The single blue peak represents non-methylated (0 % methylated) DNA, the two red peaks represent 50 % methylated DNA and the single green peak represents 100 % methylated DNA. The annealing temperature used for the *ATP5F1* assay was 60°C.

ATM MS-HRM FOR FASSTT SAMPLES

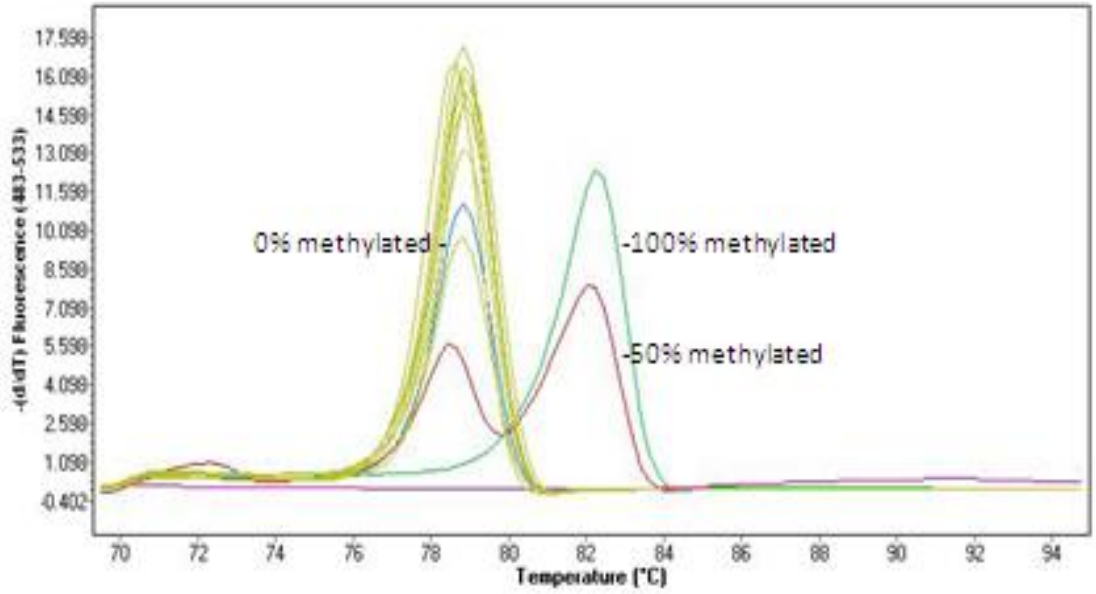


Figure 3.9. ***ATM*** **MS-HRM for FASSTT Samples.** The single blue peak represents the non-methylated (0 % methylated) DNA standard, the two red peaks represent the 50 % methylated DNA standard and the single green peak represents the 100 % methylated DNA standard. The single lime green peaks seen represent the FASSTT samples that were screened. All FASSTT samples were unmethylated in the *ATM* region examined. The annealing temperature used for the *ATM* assay was 58°C

RASSF1 MS-HRM FOR FASSTT SAMPLES

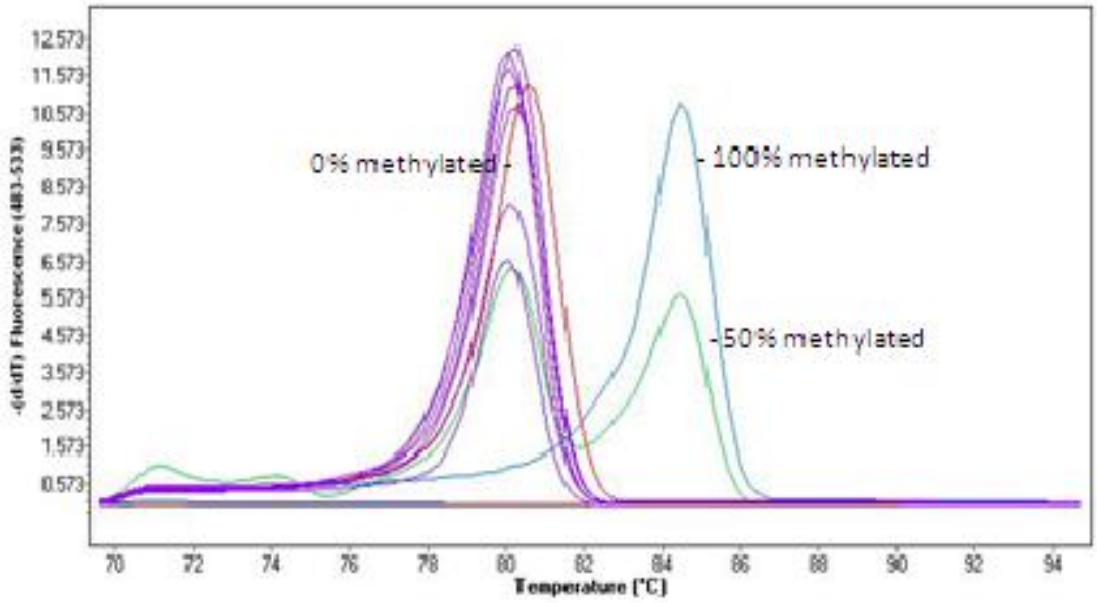


Figure 3.10. ***RASSF1A*** **MS-HRM for FASSTT Samples.** The single red peak represents the non-methylated (0 % methylated) DNA standard, the two green peaks represent the 50 % methylated DNA standard and the single blue peak represents the 100 % methylated DNA standard. The single purple peaks seen represent the FASSTT samples that were screened. All FASSTT samples were unmethylated in the *RASSF1* region examined. The annealing temperature used for the *RASSF1* assay was 58°C.

EIF2C3 MS-HRM FASSTT SAMPLES

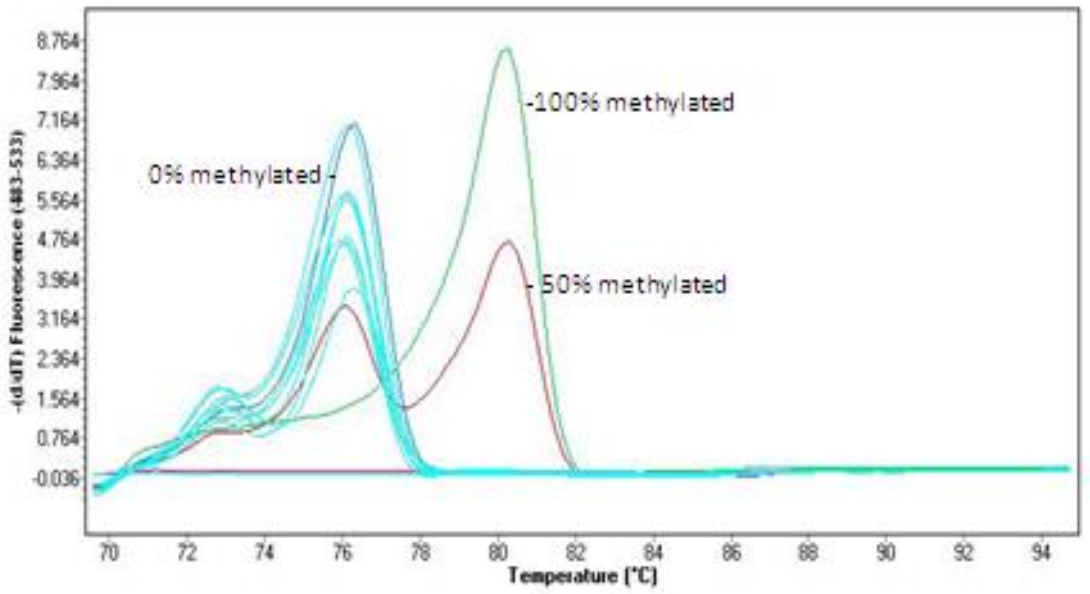


Figure 3.11. *EIF2C3* MS-HRM for FASSTT Samples. The single blue peak represents the non-methylated (0 % methylated) DNA standard, the two red peaks represent the 50 % methylated DNA standard and the single green peak represents the 100 % methylated DNA standard. The single light blue peaks seen represent the FASSTT samples that were screened. All FASSTT samples were unmethylated in the *EIF2C3* region examined. The annealing temperature used for the *EIF2C3* assay was 52°C.

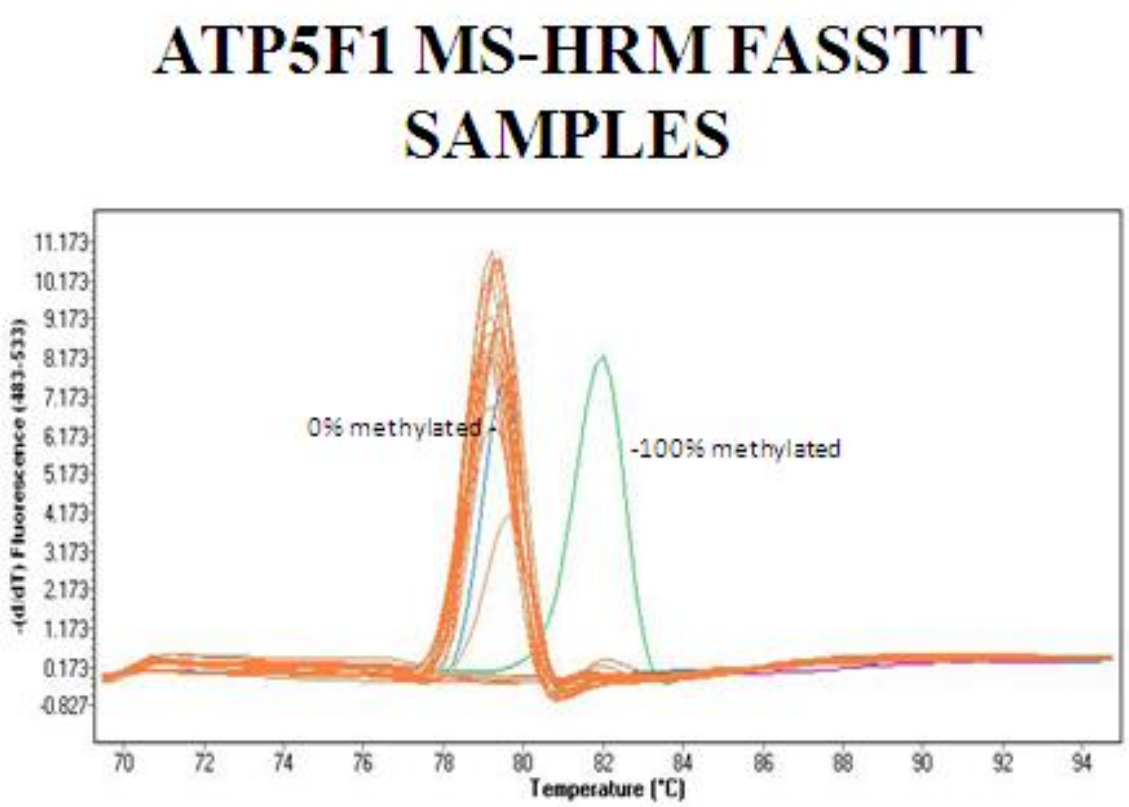


Figure 3.12. *ATP5F1* MS-HRM for FASSTT Samples. The single blue peak represents the non-methylated (0 % methylated) DNA standard and the single green peak represents the 100 % methylated DNA standard. The single orange peaks seen represent the FASSTT samples that were screened. FASSTT samples were unmethylated in the *ATP5F1* region examined. The annealing temperature used for the *ATP5F1* assay was 60°C.

ATP5F1 MS-HRM FASSTT SAMPLES

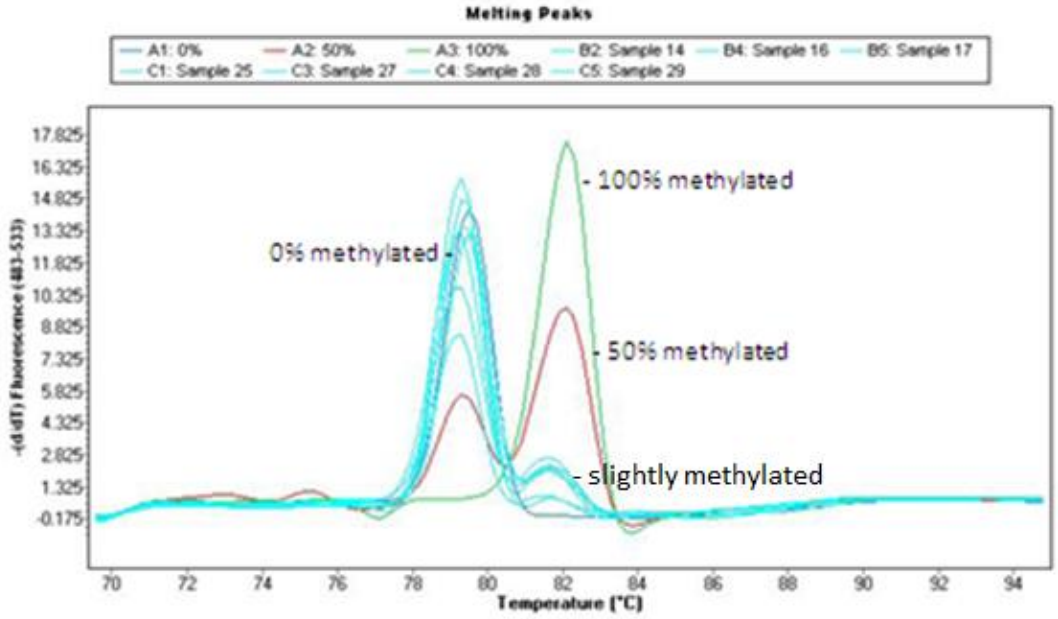


Figure 3.13. *ATP5F1* MS-HRM for FASSTT Samples. The single blue peak represents the non-methylated (0 % methylated) DNA standard, the two red peaks represent the 50 % methylated DNA standard and the single green peak represents the 100 % methylated DNA standard. The single light blue peaks seen represent the FASSTT samples that were screened. Some of the FASSTT samples were seen to be slightly methylated in the *ATP5F1* region examined. The annealing temperature used for the *ATP5F1* assay was 60°C.

COL2A1 SMART-MSP Melting Peaks

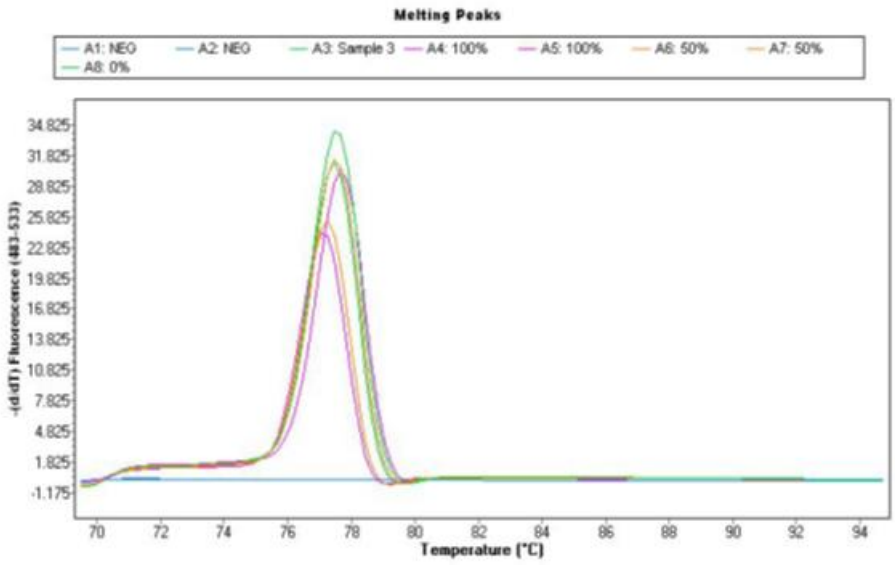


Figure 3.14. Standard SMART-MSP melting peaks for the *COL2A1* gene. The green peak represents non-methylated (0 % methylated) DNA, the orange peaks represent 50% methylated DNA, the pink peaks represent 100 % methylated DNA and the blue line represents the negative control. The same clean single melting peak is observed as expected for both methylated and un-methylated control DNA samples.The annealing temperature used for the *COL2A1* assay was 63°C.

DHFR SMART-MSP Amplification Curves & Melting Peaks

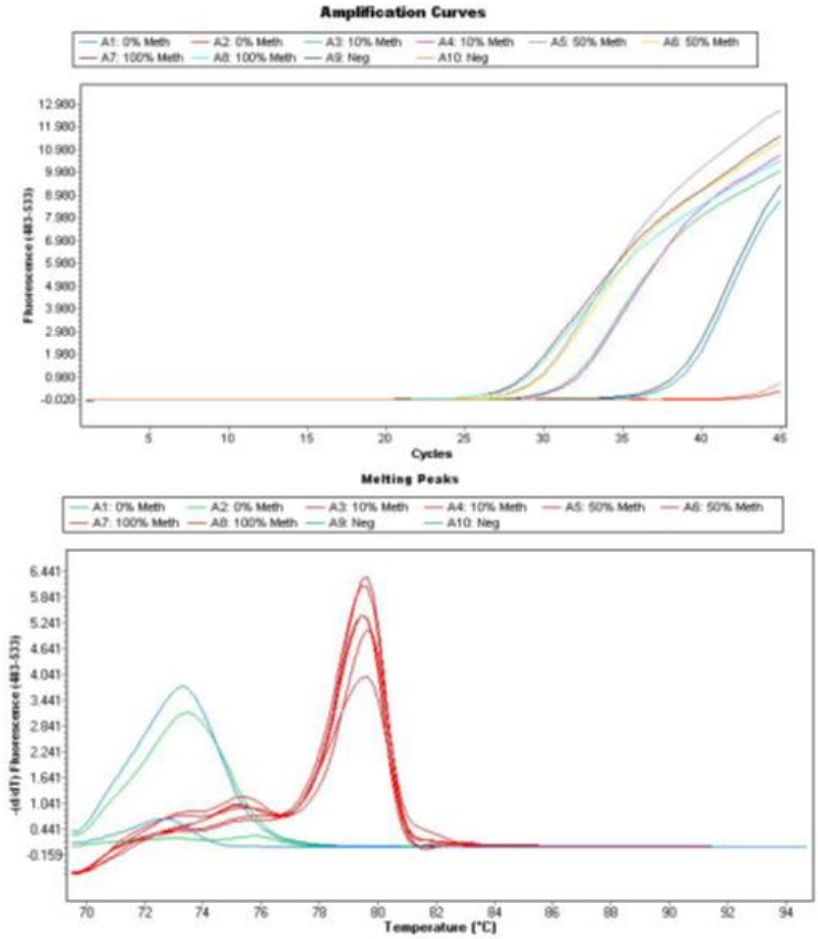


Figure 3.15. Standard SMART-MSP amplification curve and melting peaks for the *DHFR* gene. The top figure represents the amplification curves for the *DHFR* gene, with all samples being carried out in duplicate. The bottom figure represents the melting peaks for the *DHFR* gene. The blue peak represents non-methylated (0 % methylated) DNA, the red peaks represent 10 % methylated DNA, the purple peaks represent 50 % methylated DNA, the brown peak represents 100 % methylated DNA and the green peak represents the negative control. The same clean single melting peak is observed for all the methylated control DNA samples as expected. The annealing temperature used for the *DHFR* assay was 54°C.

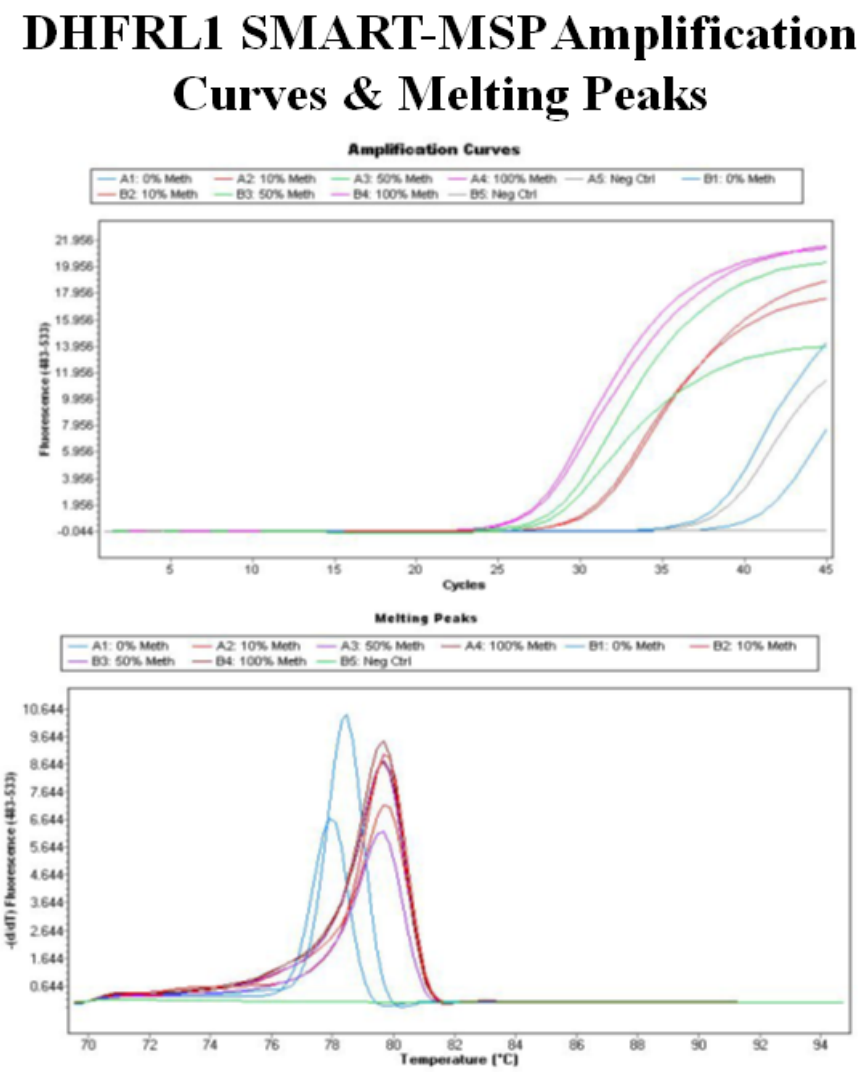


Figure 3.16. Standard SMART-MSP amplification curve and melting peaks for the *DHFRL1* gene. The top figure represents the amplification curves for the *DHFRL1* gene, with all samples being carried out in duplicate. The bottom figure represents the melting peaks for the *DHFRL1* gene. The blue peak represents non-methylated (0 % methylated) DNA, the red peaks represent 10 % methylated DNA, the purple peaks represent 50 % methylated DNA, the brown peak represents 100 % methylated DNA and the green line represents the negative control. The same clean single melting peak is observed for all the methylated control DNA samples as expected. The annealing temperature used for the *DHFRL1* assay was 58°C.

COL2A1 SMART-MSP FASSTT Samples

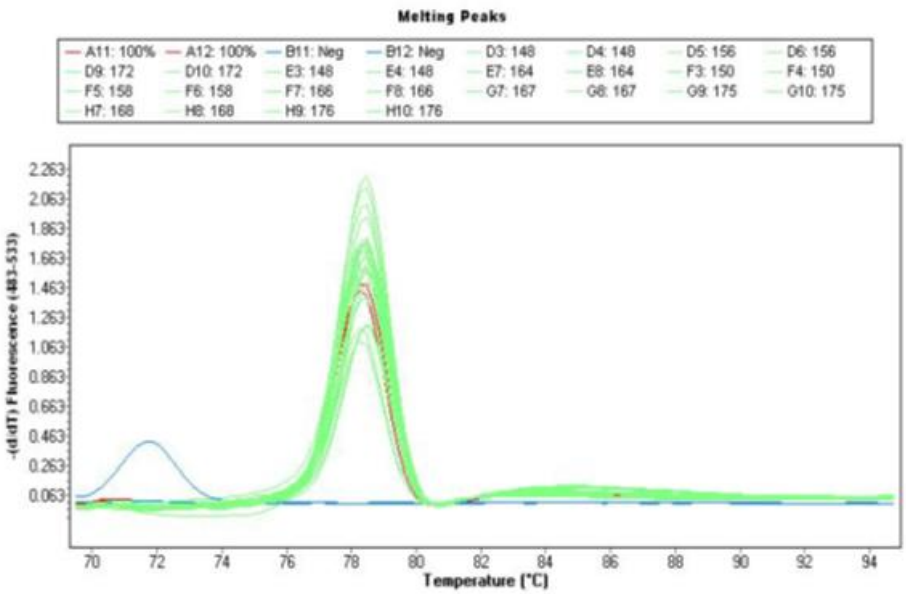


Figure 3.17. SMART-MSP Screen of FASSTT Samples for the *COL2A1* gene. The red peaks represent the 100 % methylated DNA control sample, the green peaks represent the FASSTT samples and the blue peak represents the negative control. The melting temperature used for the *COL2A1* assay was 63°C.

DHFR SMART-MSP FASSTT Samples

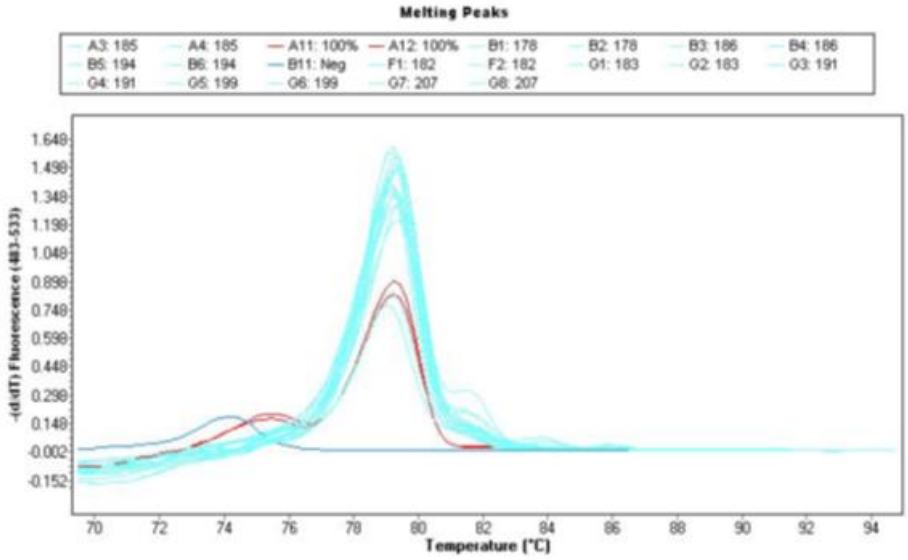


Figure 3.18. SMART-MSP Screen of FASSTT Samples for the *DHFR* gene. The red peaks represent the 100 % methylated DNA control sample, the light blue peaks represent the FASSTT samples and the blue peak represents the negative control. The melting temperature used for the *DHFR* assay was 54°C.

DHFRL1 SMART-MSP FASSTT Samples

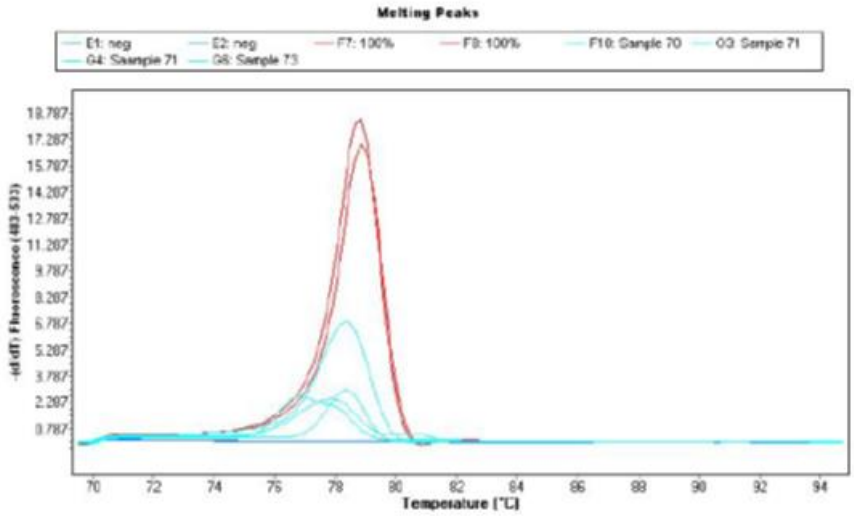


Figure 3.19. SMART-MSP Screen of FASSTT Samples for the *DHFRL11* gene. The red peak represents the 100 % methylated DNA control sample, the light blue peaks represent the FASSTT samples and the blue line represents the negative control. The melting temperature used for the *DHFRL1* assay was 58°C.

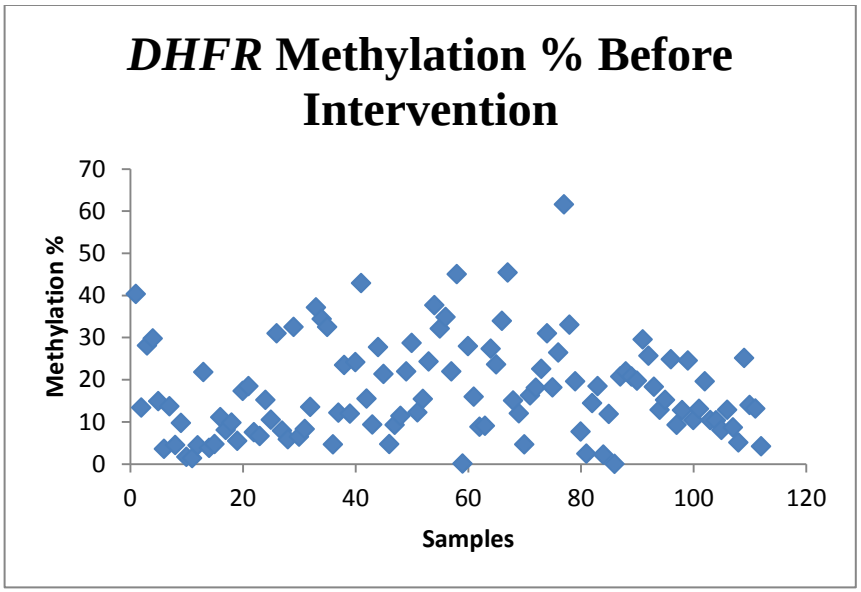


Figure 3.20. *DHFR* % Methylation pre-intervention in all of the FASSTT Samples. The starting % methylation of all the FASSTT samples was plotted as a scatter plot. % methylation differs highly between individuals.

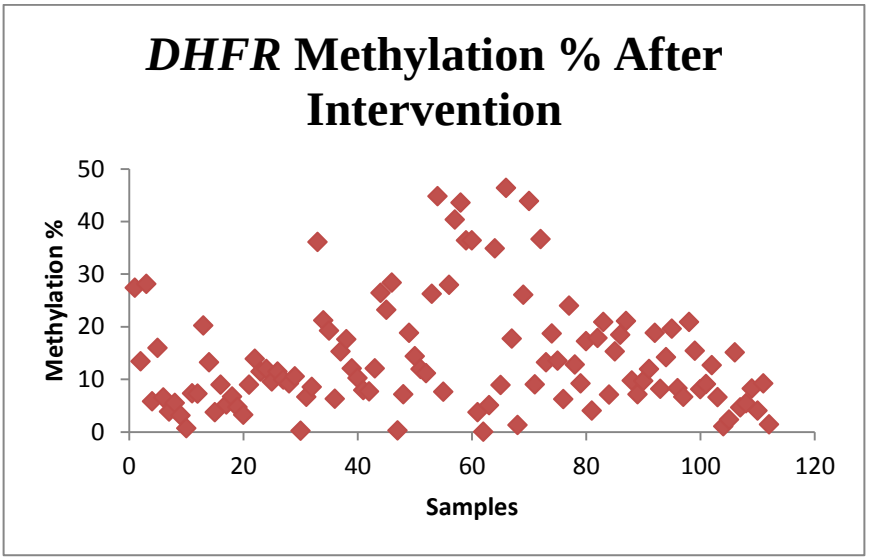


Figure 3.21. *DHFR* % Methylation post-intervention in all of the FASSTT Samples. The % methylation of all the FASSTT samples after intervention was plotted as a scatter plot. % methylation differs highly between individuals.

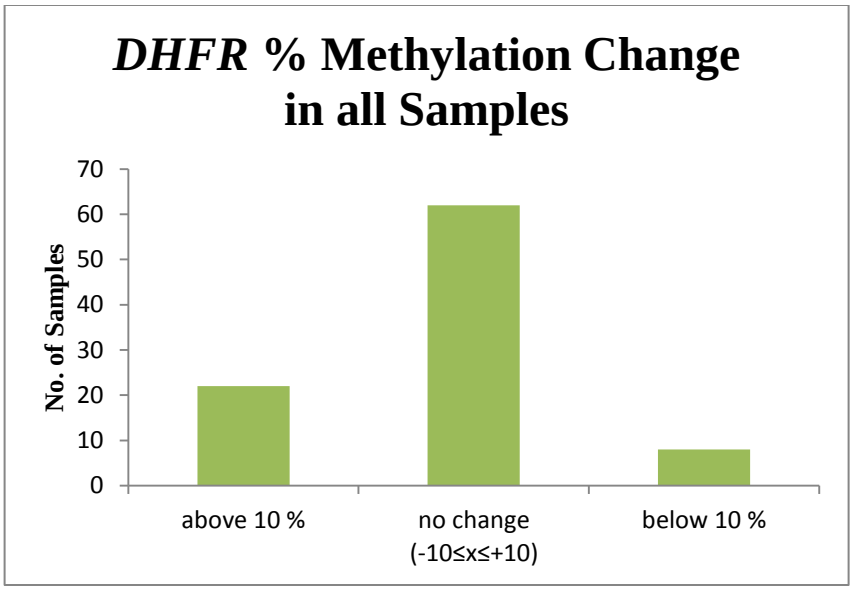


Figure 3.22. Change in *DHFR* % Methylation after Intervention in all of the FASSTT Samples. The samples were grouped into three categories depending on whether the methylation change after intervention increased by over 10 %, no change (-10 ≤x ≤+10) and methylation change decreased over 10 %. The majority of samples fell in the no change category.

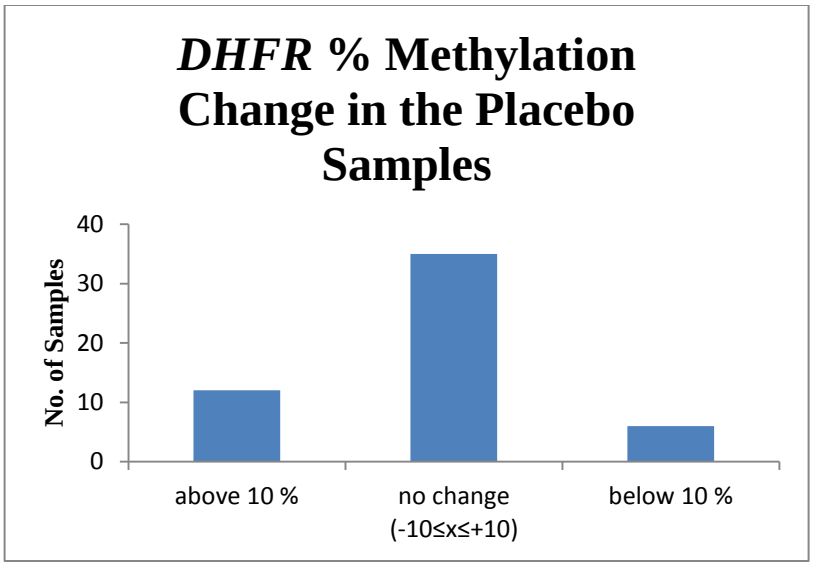


Figure 3.23. Change in *DHFR* % Methylation after Intervention in the Placebo FASSTT Samples. The placebo samples were grouped into three categories depending on whether the methylation change increased by over 10 %, no change (-10 ≤x≤ +10) and methylation change decreased over 10 %.

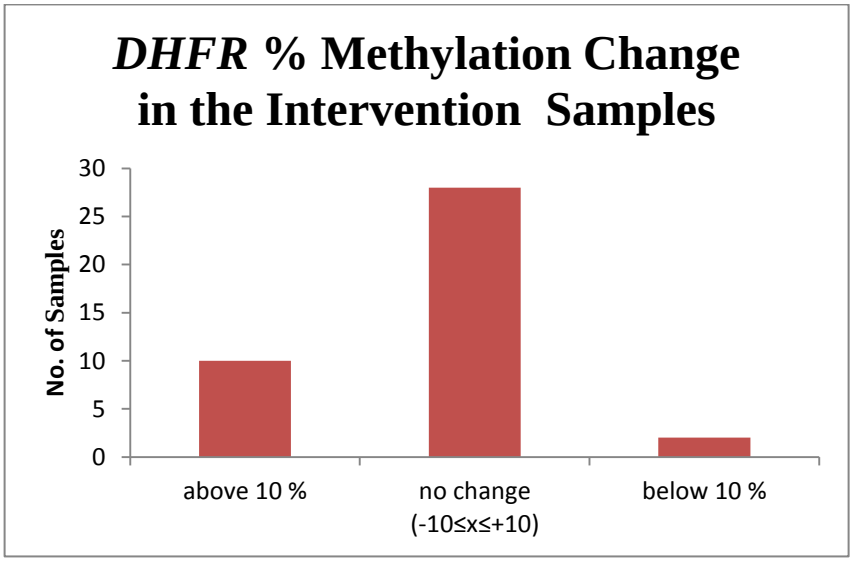


Figure 3.24. Change in *DHFR* % Methylation after Intervention in the Intervention FASSTT Samples. The intervention samples were grouped into three categories depending on whether the methylation change increased by over 10 %, no change (-10 ≤x≤ +10) and methylation change decreased over 10 %.

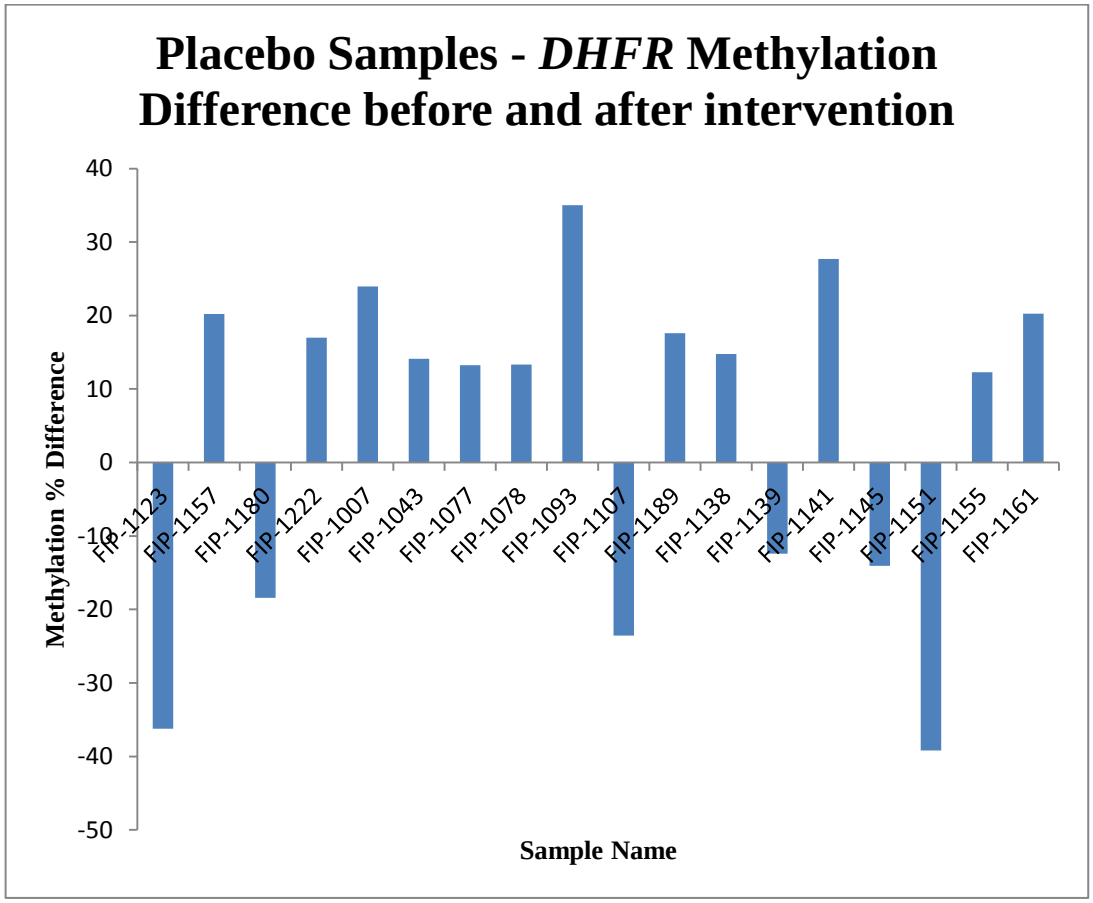


Figure 3.25. Change in *DHFR* % Methylation in the Placebo FASSTT Samples that have a Difference of over 10 % Methylation. Differential *DHFR* DNA methylation changes are seen regardless of the fact that all these women were given a placebo after intervention.

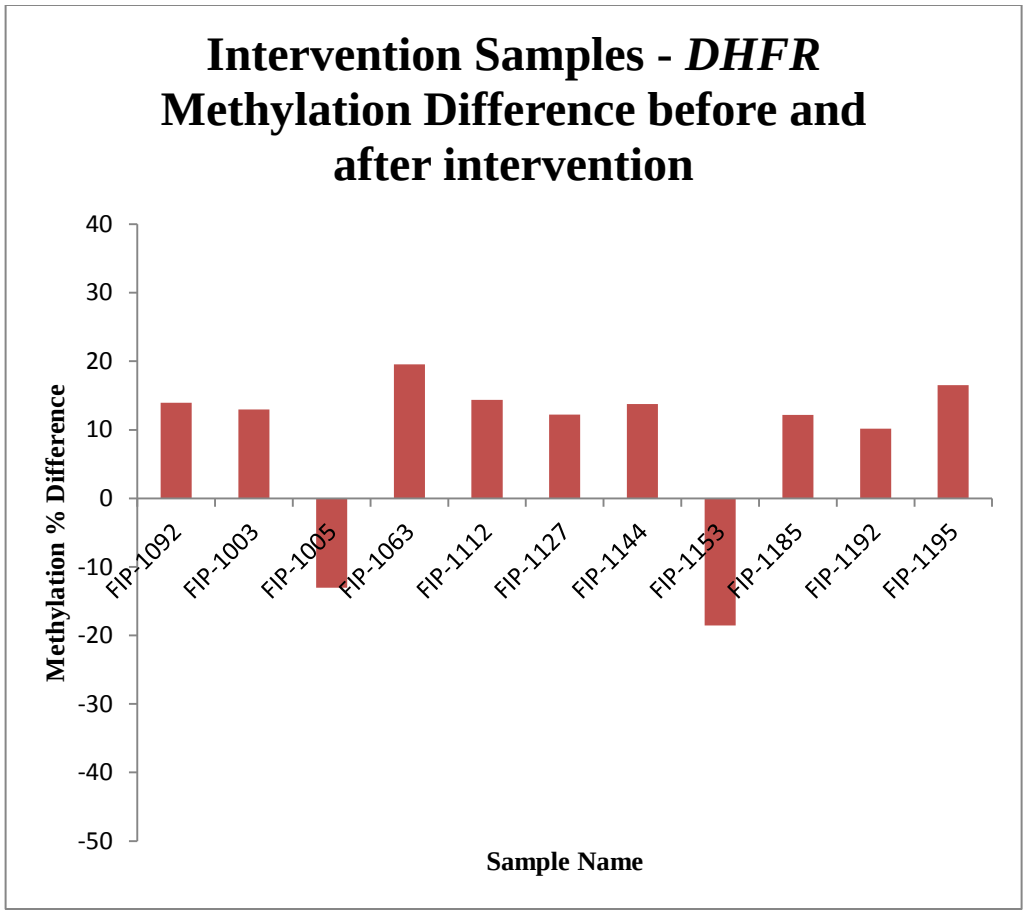


Figure 3.26. Change in *DHFR* % Methylation in the Intervention FASSTT Samples that have a Difference of over 10 % Methylation. Differential *DHFR* DNA methylation changes are seen regardless of the fact that all these women were given a daily folic acid supplement after intervention.

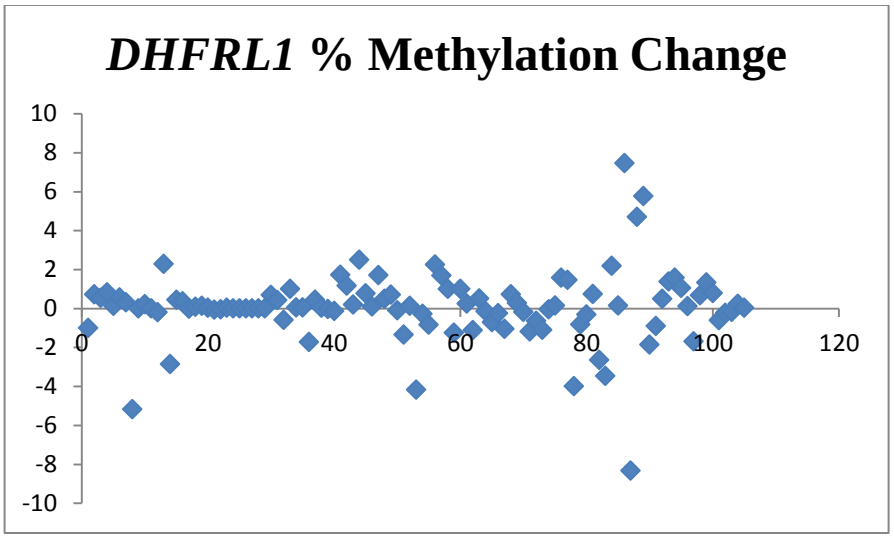


Figure 3.27. Scatter Plot of the Change in *DHFRL1* % Methylation in all of the FASSTT Samples. All samples, regardless of whether they were given a placebo or a daily folic acid supplement after intervention showed no difference (-10 ≤x≤ +10) in % methylation change before and after intervention.

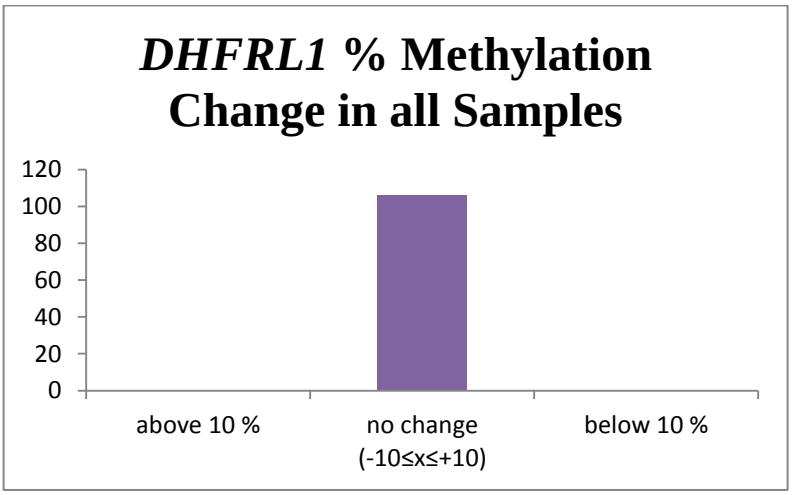


Figure 3.28. Change in *DHFRL1* % Methylation in all of the FASSTT Samples. The samples were grouped into three categories depending on whether the methylation change after intervention increased by over 10 %, no change (-10 ≤x≤ +10) and methylation change decreased over 10 %. All samples, regardless of whether they were given a placebo or a daily folic acid supplement after intervention showed no difference (-10 ≤x≤ +10) in % methylation change before and after intervention.

Table 3.1 Cp values obtained from the *COL2A1* Standards

DNA Standards	Cp Value	Cp Value	Mean Cp Value	Standard Deviation
0 % Methylated	30.80	30.45	30.62	0.24
50 % Methylated	33.24	32.35	32.80	0.62
100 % Methylated	31.80	31.15	31.47	0.46
Negative Control	-	-	-	-

Table 3.2 Cp values obtained from the *DHFR* Standards

DNA Standards	Cp Value	Cp Value	Mean Cp Value	Standard Deviation
0 % Methylated	38.64	40	39.32	0.96
10 % Methylated	31.46	31.71	31.585	0.18
50 % Methylated	29.39	29.46	29.425	0.05
100 % Methylated	28.11	28.18	28.145	0.06
Negative Control	38.3	40	39.15	1.2

Table 3.3 Cp values obtained from the *DHFRL1* Standards

DNA Standards	Cp Value	Cp Value	Mean Cp Value	Standard Deviation
0 % Methylated	40	37.8	38.9	1.55
10 % Methylated	30.31	30.65	30.48	0.24
50 % Methylated	28.46	28.12	28.29	0.24
100 % Methylated	27.07	27.16	27.115	0.07
Negative Control	38.23	-	38.23	

Chapter 4

Investigation of the impact of the *DHFR* 19 bp deletion / insertion polymorphism on circulating folate metabolites

4.1 Introduction

DHFR is an important enzyme that plays an essential role in folate mediated one-carbon metabolism (Fox and Stover, 2008). A 19 bp deletion /insertion polymorphism (rs70991108) within this *DHFR* gene and its association with circulating folate metabolites has been an area of great interest. However, there has been a lot of controversy around the results found (Johnson et al., 2004, Parle-McDermott et al., 2007, van der Linden et al., 2007). The aim of this chapter is to genotype the 19 bp deletion / insertion polymorphism by Melting Curve analysis and to elucidate the decisive impact of the *DHFR* 19 bp deletion / insertion polymorphism on circulating folate metabolites in a cohort of healthy individuals from the Trinity Student Study, the largest cohort (n= 2,507) to address this issue to date.

4.1.1 *Dihydrofolate reductase* (*DHFR*)

DHFR, as mentioned in Section 1.10 and 1.11 is an important enzyme that is expressed in all organisms and plays an essential role in folate metabolism (Fox and Stover, 2008, Litwack, 2008). It is located on chromosome 5q11.2-13.2 and is about 30 kb in length, consisting of six exons, which are separated by five introns. However, the *DHFR* coding sequence is only 564 nucleotides in length. *DHFR* is expressed in three mRNA isoforms that differ in their 3' untranslated regions (UTRs). These three mRNA species are of 800, 1,000 and 3,800 nucleotides in length (Morandi et al., 1982). *DHFR* is regulated by a TATA-less promoter at the level of transcription. This promoter is controlled by a number of different transcription factors, including Sp1 and E2F, which are key for its regulation during the cell cycle (Askari and Krajinovic, 2010). DHFR catalyses the reduction of DHF and folic acid to its biologically active metabolite, THF, which accepts and donates one-carbon groups for reactions that are necessary for DNA synthesis and homocysteine re-methylation (Fox and Stover, 2008) (Figure 1.1).

As *DHFR* plays a vital role in health and disease, it has been studied extensively and remains to be a gene of great interest. Due to this, many *DHFR* polymorphisms have been identified and examined (Goto et al., 2001, Mishra et al., 2006, Mishra et al., 2007, Parle-McDermott et al., 2007), with many thought to be functional, having the ability to alter / influence *DHFR* gene expression / activity. Although polymorphisms can occur within both the coding region and non-coding regions of genes, no polymorphisms have been found within the coding region of the *DHFR* gene (Litwack, 2008) . This may be due to the fact that *DHFR* plays a critical role in the folate mediated one-carbon metabolism pathway (Blakley and Sorrentino, 1998, Banerjee et al., 2002). However, polymorphisms found within the non-coding region of a gene have the ability to change

the expression of a gene, affect its RNA splicing or RNA stability, have the potential to influence disease and some have been found to be associated with disease susceptibility, mainly cancer (Litwack, 2008). Therefore, *DHFR* polymorphisms that are associated with increasing the levels of *DHFR* expression may play a role in the protection of cancer due to higher levels of 5,10-methylene THF needed for thymidylate synthesis. On the other hand, a change in the 5-methyl-THF pool may alter methylation reactions and subsequently increase the risk of cancer through epigenetic regulation. Variations in the gene could also alter the effect of therapeutic responses to antifolates, resulting in reduced responses to drugs / treatment (Askari and Krajinovic, 2010).

4.1.2 *DHFR* Polymorphisms

A number of *DHFR* polymorphisms have been identified in the *DHFR* gene (Figure 4.1), with several of them found to be located in the 3′ UTR of the gene. The 3′ UTR is located after the translation termination codon and contain poly(A) tails. The 3′ UTR functions principally to regulate RNA stability and translational efficiency but also function to protect the RNA from being degraded by 3′ to 5′ exonucleases, and are required for export from the nucleus (Day and Tuite, 1998). One of the first polymorphisms to be discovered in the *DHFR* 3′ UTR was the 829 C>T SNP (rs1677669), located 233 base pairs downstream of the stop codon and positioned between the first and second polyadenylation site. The 829 C>T polymorphism was shown to influence *DHFR* expression (Goto et al., 2001, Mishra et al., 2007) by RT-qPCR, with those with the 829 T/T genotype producing the highest *DHFR* mRNA expression. Although the polymorphism was examined in leukaemia and control patients, no significant association was found between disease susceptibility and the polymorphism. However, the study did show that the 829 C>T polymorphism is associated with a positive role in gene expression, i.e. increased the levels of *DHFR* mRNA expression and provides evidence that the 829 C>T SNP located in the 3′ UTR of the *DHFR* gene is a functional SNP that is involved in gene expression (Goto et al., 2001). The increased mRNA levels caused by the 829 C>T SNP was shown to be due to better RNA stability, resulting in higher levels of DHFR protein in DHFR deficient Chinese hamster cell lines by comparing cells that were transfected with the DHFR 829 TT variant and those transfected with the DHFR 829 CC variant (Mishra et al., 2006). Further research carried out on this polymorphism at a later stage, showed that this increase in *DHFR* gene expression also affects the efficiency of MTX, with those with the 829 TT genotype having a higher resistance to MTX (Mishra et al., 2007). In addition, the SNP was found to be located near the microRNA (miR) -24 3’ UTR binding site. The 829 C>T polymorphism influences the expression of *DHFR* by interfering with the function of miR-24. Cells with the T allele were found to bind miR-24 less efficiently, causing the half-life of *DHFR* mRNA

to increase, resulting in higher *DHFR* mRNA and protein levels. This in turn has an impact on MTX sensitivity, with cells with the T allele being 4-fold more resistant to MTX in comparison to those without the allele (Mishra et al., 2007). Other polymorphisms identified in the *DHFR* 3′ UTR include 35289 A>G (rs1232027) (Chandran et al., 2010) and 1171 A>T (rs7387) (Sharma et al., 2009), both of which have been shown to negatively affect (decrease) the response to MTX treatment in psoriatic arthritis and rheumatoid arthritis respectively. A 9 bp repeat polymorphism (rs3045983) has been reported in the 5′ upstream region of the *DHFR* gene that partially overlaps with the 5′ UTR. The impact of this polymorphism is not known but it is thought that it may affect mRNA stability or the efficiency of translation (Gellekink et al., 2007). A number of polymorphisms have also been shown in the upstream region from the minor promoter transcription initiation site, 317 A>G (rs408626) and 1610 C>G/T (rs1650694). Both these polymorphisms have been shown to be associated with worse acute lymphoblastic leukaemia, most likely due to higher levels of *DHFR* expression (Dulucq et al., 2008). Polymorphisms have also been identified in intron 3 of the *DHFR* gene, 8890 A>G (rs1643659) and 10372 A>C (rs1677639), both of which have been associated with a decrease in colorectal cancer risk in non-users of multivitamin supplements (Levine et al., 2010). Two recent studies have also identified rare coding mutations (not polymorphisms) in *DHFR* as causative for megaloblastic anaemia and cerebral folate deficiency (Banka et al., 2011, Cario et al., 2011). However, no non-synonymous polymorphisms have been identified within the coding region of the *DHFR* gene to date.

4.1.3 *DHFR* 19 bp deletion / insertion polymorphism

The most extensively studied *DHFR* polymorphism is a 19 bp deletion / insertion (rs70991108) polymorphism. This 19 bp deletion / insertion polymorphism in intron-1 of the *DHFR* gene, located 60 bases from the splice donor site (Figure 4.2) has been associated with maternal risk of NTDs (Johnson et al., 2004, van der Linden et al., 2007, Parle-McDermott et al., 2007). NTDs are a group of common birth defects of the brain and spinal cord, with a prevalence of 1 per 1,000 in Europe and the USA, caused by the failure of the neural tube to close and include spina bifida and anencephaly. As neural tube closure takes place during a period of rapid cellular proliferation, *DHFR* activity is most likely to be crucial in maintaining optimal DNA synthesis during this time. As previous studies have shown and it is now well known that low serum folate and high homocysteine levels are associated with an increased risk of NTDs, (van der Put et al., 2001) it is not surprising that changes / polymorphisms in the *DHFR* gene has been considered as a candidate for association with NTDs. However, the association of the 19 bp deletion/insertion polymorphism with NTDs and the biological function of this polymorphism are highly controversial. In 2004,

Johnson *et al*. (2004) reported that the *DHFR* 19 bp deletion allele in mothers is associated with an increased risk of having a baby born with spina bifida in a heterogeneous cohort of n=35 as the Sp1 transcription factor binding site is located in the deletion sequences (Johnson et al., 2004). As mentioned in Section 1.11, Sp1 sites mediate the induction of *DHFR* transcription in the late G_1 phase of the cell cycle (Jensen et al., 1997). Therefore, deletion of this site could alter *DHFR* transcription, altering the level of *DHFR* expression, which could result in an increased risk of having a baby born with NTDs. On the other hand, Parle-McDermott *et al*. (2007) suggested that the 19 bp deletion allele has a protective effect in a larger cohort (n = 283), decreasing the risk of having a baby with spina bifida while others have proposed that it has no effect at all (n = 101) (van der Linden et al., 2007). Thus, whether the *DHFR* 19 bp deletion allele is associated with NTDs remains to be elucidated. However the *DHFR* 19 bp deletion allele has been shown to have a protective role in adult acute lymphoblastic leukaemia (ALL) patients (n = 245) (Gemmati et al., 2009) and has also been shown to be a risk of breast cancer in multivitamin users (n = 1062) (Xu et al., 2007).

A number of studies have also been carried out that examine the impact of the 19 bp deletion / insertion polymorphism on circulating folate metabolites, with conflicting results. One study who examined 20 individuals ranging from 20-90 years of age reported a 2.5 % and a 14.4 % decrease in total plasma homocysteine (tHcy) for the heterozygous and homozygous genotypes respectively in comparison to the wild type genotype ($p = 0.016$) but no differences were seen with serum or red blood cell (RBC) folate levels, with adjustments for age and sex not changing the result (Gellekink et al., 2007), whereas another group found the 19 bp deletion (del/del) genotype to be modestly associated with elevated levels of RBC and serum folate in women but not in men, with no differences seen in tHcy levels, although the effects were annulled by smoking (Stanisławska-Sachadyn et al., 2008). They carried out the study on 430 young Northern Irish adults ranging from 20-26 years of age. Although only small samples numbers were used in both these studies, both these findings suggest that the deletion allele may have a protective effect against NTDs as maternal low folate / high homocysteine levels have previously been shown to be associated with the increased risk of NTDs in offspring (Mills et al., 1995). In contrast, another group who investigated the effect of the 19 bp deletion / insertion polymorphism and plasma concentrations of circulating folic acid with plasma total homocysteine, total folate and RBC folate concentrations reported that the functional effect of the *DHFR* polymorphism depends on the level of folic acid intake (Kalmbach et al., 2008). They examined 1215 individuals from the Framingham Offspring Study and reported that folic acid intake of ≥500 μg/d increased the level of high

circulating unmetabolised folic acid in individuals who had the deletion (del/del) genotype (47.0%) in comparison to those who had the heterozygous (WT/del) genotype (21.4%) and the insertion (WT/WT) genotypes (24.4%) (P = 0.03). An interaction between the *DHFR* polymorphism and folic acid intake was also seen with respect to RBC folate (P = 0.01). When folic acid intake was <250 μg/d, the del/del genotype was associated with significantly lower RBC folate (732.3 nmol/L) compared to the WT/WT genotype (844.4 nmol/L). No interaction was found between the *DHFR* genotype and folic acid intake with tHcy and plasma total folate.

The larger two studies described above (Kalmbach et al., 2008, Stanisławska-Sachadyn et al., 2008) suggest that the *DHFR* deletion allele may have an impact on circulating unmetabolised folic acid and RBC folate but that folic acid intake and sex may also need to be considered. The aim of our study is to elucidate the decisive impact of the *DHFR* 19 bp deletion / insertion polymorphism on circulating folate metabolites in the Trinity Student Study, the largest cohort (n= 2,507) to address this issue to date. This Chapter is in collaboration with Prof. Anne Molly from Trinity College Dublin and Dr. James Mills from the NIH.

4.1.4 The TSS Study

The Trinity Student Study (TSS) is a study, which is funded by the National Institute of Health in the USA and the Health Research Board of Ireland. Young males and females between the ages of 19 and 36, attending Trinity College Dublin were recruited between February 2003 and 2004. A total of 3569 students primarily took part in the study. Of these 3569 students, the 2524 individuals who had Irish grandparents and no major medical problems at the time were invited to continue with the study. Blood samples were taken from the subjects and they were also asked to fill out a questionnaire (lifestyle and nutritional information). Of these subjects, complete blood data could not be obtained for 2 students and 15 students failed to return the questionnaire and were therefore excluded from the data analysis. This left a total of 2507 subjects for the study. All samples were made anonymous before analysis. Ethical approval was attained from the Dublin Federated Hospitals Research Ethics Committee, which is associated with the University of Dublin, Trinity College, and the study was reviewed by the Office of Human Subjects Research at the National Institutes of Health. Written informed consent was also obtained from the participants when they enrolled.

4.1.5 Melt Curve Analysis

Melt curve analysis was the method chosen to genotype our DNA samples, mainly based on instrument availability. Melt curve analysis is a method that is based on looking at the fluorescence of the sensor and anchor probes and is principally designed to genotype SNPs. In melting curve genotyping experiments, labelled sequence-specific probes are added to the PCR reaction, and are allowed to anneal to the reaction products during the amplification procedure. Melting curves are acquired by heating the probe-amplicon heteroduplex and measuring the changes in fluorescence that occurs. Even a single mismatch between the probe and amplicon can show a reduced melting temperature when compared to the perfect match (Lyon and Wittwer, 2009). The fluorescence data acquired from the melting curves are converted into melting peaks by the LightCycler Data Analysis software that compensates for colour and removes the background fluorescence and plots the negative derivative of fluorescence with respect to temperature (Pont-Kingdon and Lyon, 2003) (Figure 4.3). The information obtained from the melting peaks can be used to identify which SNP base is present on the allele, and can therefore be used in applications such as clinical diagnosis and monitoring.

In our study, we used HybProbes to design our melt curve genotyping assays. HybProbes are based on fluorescence resonance energy transfer (FRET), whereby two sequence-specific oligonucleotide probes labelled with different dyes, a sensor (donor) probe labelled with Fluorescein and an anchor (acceptor) probe labelled with BODIPY 630/650, are added to the reaction mix along with the PCR primers. The sensor probe binds to the region including the variable target sequence, with the adjacent anchor probe having a distinctly higher melting temperature. During the annealing phase, the HybProbes bind to the target DNA sequence in a head-to-tail fashion, causing the two dyes to be in close proximity to each other. The donor dye (Fluorescein) is excited by the blue LED (absorption wavelength of 495 nm, emission wavelength of 520 nm), which in turn excites the acceptor dye (BODIPY 630/650) (absorption wavelength of 620 nm, emission wavelength of 640 nm) as long as the two dyes are close to each other (within 15 nucleotides), which then emits fluorescent light at a different wavelength. At low temperature, the probe hybridises to both alleles but as the temperature increases, the probe first denatures from the mismatched allele, then from the perfect matched allele. The fluorescence is directly proportional to the amount of target DNA generated during PCR. HybProbe probes are displaced during the elongation and denaturation steps. (Figure 4.4)

4.1.6 Objectives

Assessment of the *DHFR* 19 bp deletion / insertion polymorphism in the Trinity Student Study samples was examined using the following methodology.

- Development of a genotyping assay on the LightCycler 480 for the *DHFR* 19 bp deletion / insertion polymorphism.
- Screening the *DHFR* 19 bp deletion / insertion polymorphism in the TSS samples.
- Correlation analysis of the *DHFR* 19 bp deletion / insertion polymorphism with folate metabolites and folate intake in the TSS samples.

4.2 Results

4.2.1 Development of a genotyping assay for the 19 bp deletion / insertion polymorphism.

The *DHFR* 19 bp deletion/insertion polymorphism was genotyped using melt curve analysis with HybProbes on the Roche Lightcycler 480. As melting curve genotyping is principally designed for SNPs, it was necessary to develop a novel assay in order to genotype the *DHFR* 19 bp deletion / insertion polymorphism. For this assay, the sensor probe was designed so that it would bind directly to the alleles with the 19 bp insertion, resulting in alleles with the insertion to have a higher melting temperature in comparison to those with the deletion alleles due to greater complementary base pairing between them (Figure 4.5).

Once the HybProbes had been designed to incorporate the deletion / insertion polymorphism, primers were also designed. A multi-sequence analysis of the *DHFR* gene family was carried out using clustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/) and primers were designed to ensure that only the *DHFR* gene was being amplified (See Table 2.9 for primer and probe sequences and Appendix K for *DHFR* gene sequence analysis and primer and probe location). Once probes and primers were designed, the 19 bp deletion / insertion assay, was optimised by running primer and probe curves to determine the optimal concentrations required as well as altering annealing temperatures. As the polymorphism is located in a GC-rich area of the gene, the assay proved to be difficult to optimise, and in the end required the use of DMSO. The *DHFR* 19 bp deletion / insertion genotyping assay was developed and optimised successfully (Figure 4.6).

4.2.2 Screening the *DHFR* 19 bp deletion/insertion polymorphism in the TSS study

Once the 19 bp deletion / insertion polymorphism genotyping assay had been developed and optimised, ~ 2500 TSS DNA samples were screened and analysed for the polymorphism. Approximately 10 % of the DNA samples were repeated in order to ensure that the assay was

working efficiently and that the software was calling the correct genotypes consistently. The genotype (numbers and percentages) results as well as the allele frequencies obtained are shown in Table 4.1. Genotypes were analysed for both males and females together as well as separately.

The genotypes and allele frequencies obtained from the TSS DNA samples were similar to those seen previously in an Irish population, (Allele 1 = 0.55, Allele 2 = 0.45) (Parle-McDermott et al., 2007), a mixed ethnicity population (Allele 1 = 0.55, Allele 2 = 0.45) (Johnson et al., 2004) and a Dutch population (Allele 1 = 0.57, allele 2 = 0.42) (van der Linden et al., 2007) and did not differ significantly from the expected frequencies based on the Hardy-Weinberg principle, which states that in a large randomly breeding population, the allelic frequencies will remain the same from generation to generation, assuming that there is no mutation, gene migration, selection or genetic drift. The Hardy-Weinberg principle is illustrated as follows: p2 + 2pq + q2 = 1, where p2 is the frequency of the WT genotype, 2pq is the frequency of the HET genotype and q2 is the frequency of the DEL genotype. Figures for genotype frequency are shown in Table 4.1. From this, the expected number of individuals for each genotype could be calculated. In our population, the expected numbers for each genotype was as follows:

WT - 0.33 *(2503) = 826
HET - 0.48 *(2503) = 1201
DEL - 0.19 *(2503) = 476

To test whether a population is in Hardy-Weinberg Equilibrium, the chi-squared test (χ^2) can be used.

$$\chi^2 = \Sigma \left[(\text{observed value} - \text{expected value})^2 / \text{expected value}\right]$$

In our population, the calculation was as follows:

$$\chi^2 = [(829-826)^2/826] + [(1204-1201)^2/1201] + [(470-476)^2/476]$$
$$= 0.09$$

The critical chi-square value for a p-value of 0.05 (5 %) significance cut off value with one degree of freedom is 3.841. Our chi-squared value of 0.09 falls below the 0.05 significance cut off value of 3.841 and therefore, we can conclude that the differences observed are caused by chance and that our population does not differ significantly from what we would expect, indicating that our population is in Hardy-Weinberg Equilibrium.

4.2.3 *DHFR* 19 bp deletion / insertion polymorphism is not associated with folate metabolites and/or folate intake

The 19 bp deletion / insertion polymorphism results that were obtained were assessed with three folate metabolites, red cell folate (RCF), serum folate and total plasma homocysteine (tHcy), whose measurements had previously been obtained from the TSS.

The characteristics of the participants in the TSS are shown in Table 4.2. As males and females differ with respect to several of the biochemical and lifestyle variables, analyses were performed separately for males and females. Information regarding age, sex, weight and smoking was previously collected using questionnaires.

To determine the correlation between the *DHFR* 19 bp deletion / insertion polymorphism and the folate metabolites, the mean serum folate, the mean RCF and the mean tHcy levels were analysed with respect to *DHFR* genotype (Table 4.3). Furthermore, as sex was shown to appear to influence the *DHFR* 19 bp deletion / insertion correlations with serum and RBC folate in a previous study (Stanisławska-Sachadyn et al., 2008), the association between the *DHFR* 19 bp deletion / insertion polymorphism with the folate metabolites was also analysed separately for males and females (Table 4.4 and Table 4.5). ANOVA analysis was carried out on the log transformed data in order to generate the p-values. From the tables / p-values, we can see that no association was observed between *DHFR* genotype and any of the three folate metabolites, even when analyses were carried out separately for males and females. Mean RCF, serum folate and tHcy values were similar regardless of *DHFR* genotype or sex.

In order to determine whether folate intake has an influence on genotype correlations as previously reported (Kalmbach et al., 2008), each individual was firstly categorised by total folic acid intake per week into one of four groups. The first group was comprised of individuals who had no folic acid intake per week, and the remainder of the groups was subdivided into three equal groups, defining the categories as low, those individuals who had a folic acid intake of 128.57 µg/week or lower, medium, those individuals who had a folic acid intake of between 128.58 µg/week $\leq x \leq$ 200 µg/week and high, those individuals who had a folic acid intake of over 200 µg/week. Out of a total of 2,503 people, 76.07 % of the individuals had no folate intake, 8.03 % had low folate intake, 8.03 % had medium folate intake and 7.87 % had high folate intake. These subgroups were then analysed in correlation with the *DHFR* 19 bp deletion / insertion polymorphism (Table 4.6).

These groups were then analysed with mean RCF, serum folate and tHcy levels (Table 4.7-4.9). From Table 4.7, we can see that RCF levels roughly correlate with folate intake as expected, with increased levels of RCF as the levels of folate intake increase. However, no association was seen between genotype, folate intake and red cell folate. From Table 4.8, we can see that serum folate levels also roughly correlate with folate intake as expected, with increased levels of serum folate as the levels of folate intake increase. However, no association was seen between genotype, folate intake and serum folate. From Table 4.9, we can see that tHcy levels are roughly inversely correlated with folate intake as expected, with decreased levels of tHcy as the levels of folate intake increase. However, no association was seen between genotype, folate intake and tHcy.

4.3 Discussion

The *DHFR* 19 bp deletion / insertion polymorphism is one of the most widely studied *DHFR* polymorphisms. Although the polymorphism has been associated with a number of different cancers (Kalmbach et al., 2008, Xu et al., 2007), there has been a lot of controversy surrounding its association with NTDs (Johnson et al., 2004, Parle-McDermott et al., 2007, van der Linden et al., 2007) and its association with circulating folate metabolites (Gellekink et al., 2007, Stanisławska-Sachadyn et al., 2008, Kalmbach et al., 2008) and its exact role has yet to be determined. *DHFR* plays a major role in the folate mediated one-carbon metabolism pathway and is the only known enzyme that is capable of reducing folic acid. Therefore the activity of this enzyme is of great importance, especially due to widespread worldwide mandatory and voluntary folic acid fortification that has occurred in order to prevent the risk of NTDs (Czeizel and Dudás, 1992). However, a recent study has shown that *DHFR* activity in the human liver varies highly between individuals and that it has a limited ability to reduce folic acid (Bailey and Ayling, 2009). Therefore, it has been hypothesized that any polymorphisms within the *DHFR* gene will alter its ability to reduce folic acid, thereby reducing the assimilation of folic acid into the endogenous forms of folate (Kalmbach et al., 2008).

The location of the 19 bp deletion / insertion polymorphism within the DHFR gene sequence also suggests that this polymorphism may be functional. The *DHFR* 19 bp deletion / insertion polymorphism is located in intron 1, a non-coding region of the *DHFR* DNA sequence (Johnson et al., 2004). This region has previously been reported to be an important site for transcription, translation and other regulatory functions in other genes (Takayanagi et al., 1992, Guérin et al., 1995, Clark et al., 1997). This suggests that the *DHFR* 19 bp deletion genotype may

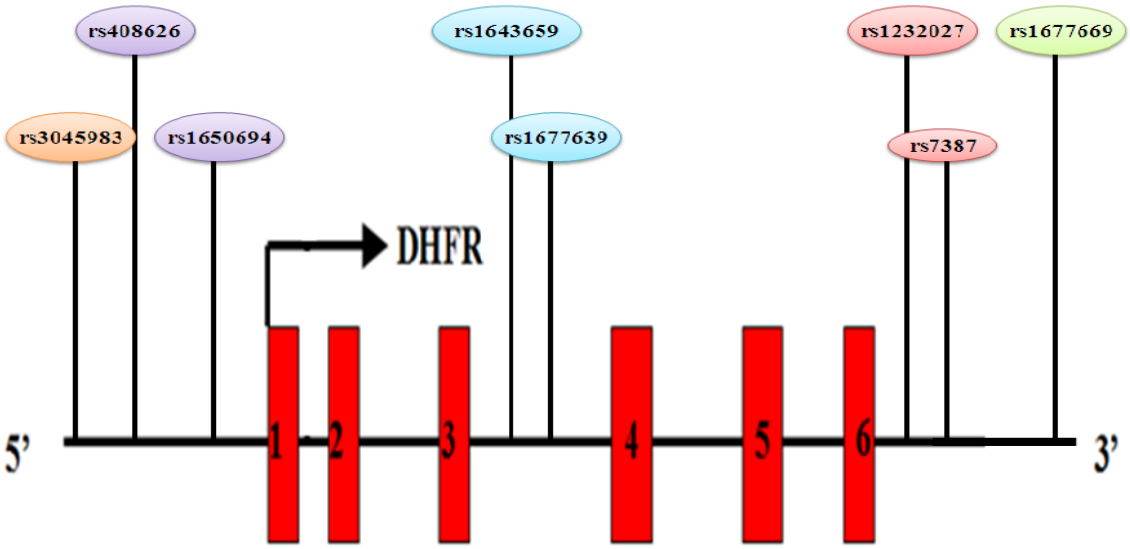
alter the levels of *DHFR* mRNA and subsequently alter the amount of enzyme. Studies that have been carried out to investigate the functionality of the *DHFR* polymorphism have shown a dose-dependent relationship between the *DHFR* deletion allele and *DHFR* mRNA levels, with those with the deletion genotype to have a 1.5-fold increase in *DHFR* mRNA levels in healthy individuals (Parle-McDermott et al., 2007) and a 4.8-fold increase in *DHFR* mRNA levels in breast cancer patients (Xu et al., 2007) in comparison to those with the insertion genotype.

The *DHFR* intron-1 has also been shown to contain transcriptional factor binding sites for transcriptional factors such as Specificity Protein 1 (Sp1), Sp3 and estrogen receptor 1 (ESR1), which may be involved in the regulation of *DHFR* transcription. The *DHFR* 19 bp insertion contains the binding sites for Sp1 and Sp3, whereas the 19 bp deletion contains ESR1 sites (Kalmbach et al., 2008). The Sp1 site is thought to be involved in growth / cell cycle regulation during *DHFR* transcription, through the recruitment and interactions with co-activators and repressors and therefore, the removal of the Sp1 transcription factor binding site may increase or decrease transcription (Kalmbach et al., 2008). ERS1 is a ligand-activated transcription factor and is composed of several domains that are important for hormone binding, DNA binding and transcriptional activation (Jensen et al., 1997).

In order to finally resolve the issue as to whether the *DHFR* 19 bp deletion / insertion polymorphism is associated with circulating folate metabolites, we examined the relationship and the effect of the *DHFR* 19 bp deletion / insertion genotype on serum folate, red cell folate and total homocysteine in the largest human sample set yet to be examined. In order to do this, we first had to design an assay that would be able to detect the genotype of the *DHFR* polymorphism in a large study set in an optimal manner. The method of choice was a melt curve assay and this was successfully designed and optimized for genotype detection (Figure 4.5). For validation purposes, we could have also validated the genotypes by gel electrophoresis or by sequencing of th PCR products and a genotyping assay by gel electrophoresis has been developed for the *DHFR* 19 bp deletion / insertion polymorphism.

The results from the analyses presented above indicate that the *DHFR* 19 bp deletion / insertion polymorphism is not associated with any of the three folate metabolites examined, serum folate, red cell folate or total homocysteine in the TSS study, whether the data from the males and females was analysed in combination or separately, which is in contrast to the studies which have been carried out previously (Gellekink et al., 2007, Kalmbach et al., 2008, Stanisławska-Sachadyn

et al., 2008). Although previous groups have reported the 19 bp deletion to be associated with lower tHcy levels in both males and females and elevated RBC and serum folate levels in women, the numbers of participants in their studies were much smaller in comparison to the number of individuals in our study, which may have influenced the results obtained, resulting in false-positive outcomes. The previous studies also suggest that the differences seen are gender dependent, while others not, resulting in conflicting results. The differences in the age of the participants may have also been a contributable factor to the difference in results obtained between our study and previous studies carried out by Gellekink *et al*. (2007), however Stanisławska-Sachadyn *et al*. (2008), used subjects of similar age and therefore the discrepancies seen between the results cannot be attributed to age difference. Another study also found that the functional effect of the *DHFR* polymorphism depends on the level of folic acid intake. We also examined the effect of folic acid intake on the *DHFR* genotype (Table 4.6) as well as analysing them with respect to each of the three folate metabolites (Table 4.7-4.9) but did not observe any associations in our population. In conclusion, our results indicate that the *DHFR* 19 bp deletion / insertion polymorphism does not have an impact on circulating folate metabolites.



SNP	Allele Change	Location
rs3045983	9 bp repeat	5' upstream region
rs408626	A>G	Minor promoter
rs1650694	C>G/T	Minor promoter
rs1643659	A>G	Intron 3
rs1677639	A>C	Intron 3
rs1677669	C>T	3'UTR
rs7387	A>T	3'UTR
rs1232027	A>G	Downstream to 3'UTR

Figure 4.1. An Overview of the DHFR Polymorphisms. The schematic above shows the rough positions of the *DHFR* polymorphisms on the *DHFR* gene. The table above shows the allele change and the locations of the different *DHFR* polymorphisms. Although there are numerous *DHFR* polymorphisms, none are located within the coding region of the gene.

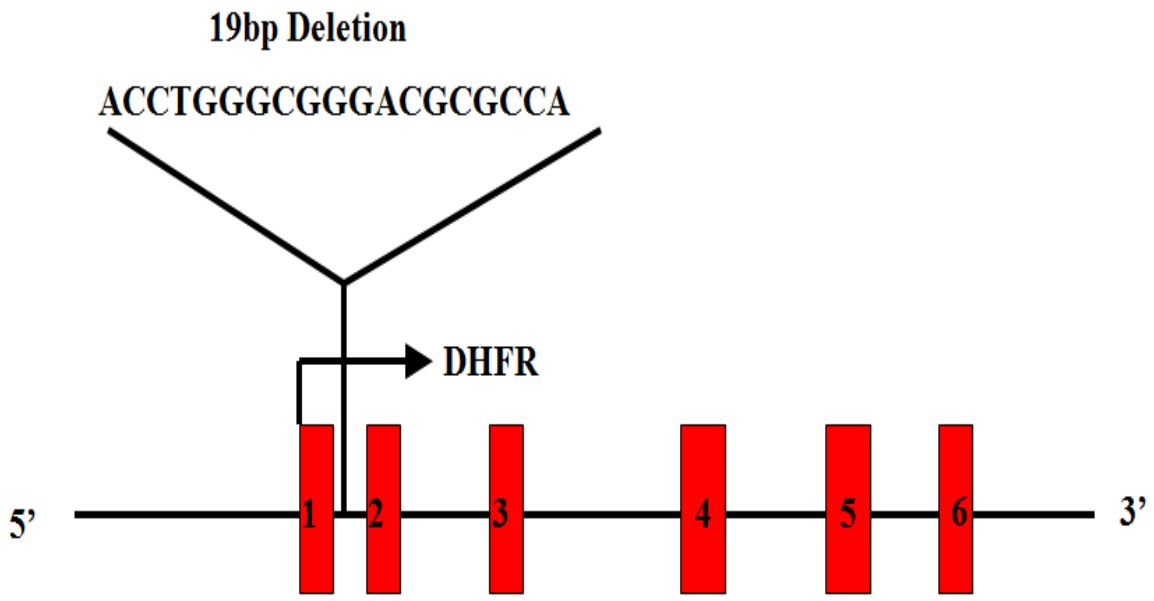


Figure 4.2. Structure and location of the *DHFR* 19 bp deletion / insertion polymorphism. The *DHFR* gene consists of six exons, indicated as red boxes above. The 19bp deletion / insertion polymorphism is located in intron 1, 60 bases downstream of the splice donor site. The nucleotide sequence of the 19bp deletion / insertion is also shown (Image taken from Parle-McDermott *et al*., unpublished).

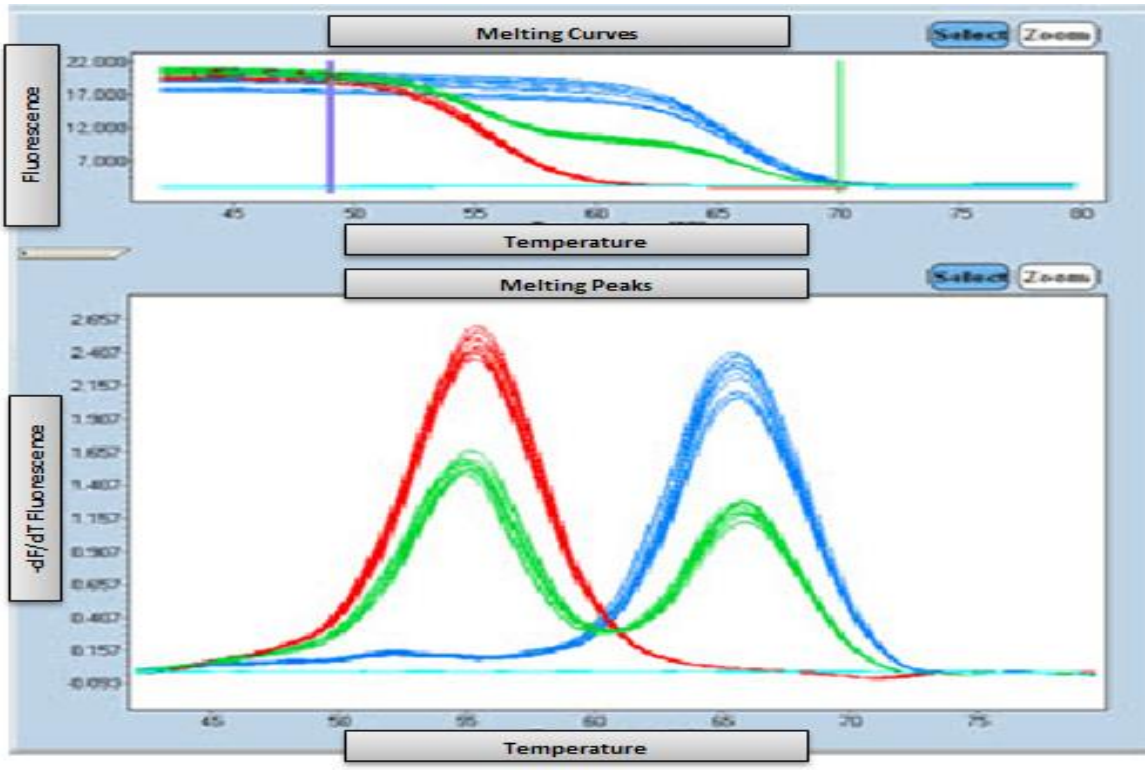


Figure 4.3. Overview of Melt Curve Analysis. Melt curve analysis analyses SNPs on specific alleles by melting in the presence of sequence-specific labelled probes that bind with different affinities to different alleles in the region containing the SNP. The raw data from the melting curves are generally represented by plotting fluorescence against temperature. This data is often converted into the negative first derivatives (-dF/dT) for convenience, which allows the melting temperatures to be seen as peaks. Data derived from homozygous wild types (single blue peaks), homozygous mutants (single red peaks), and heterozygous samples (two green peaks) differ in their peak number and positions. Matching colours for samples with the same genotype (peaks) allow for easy visualisation and interpretation of the results. (Image taken from www.roche-applied-science.com)

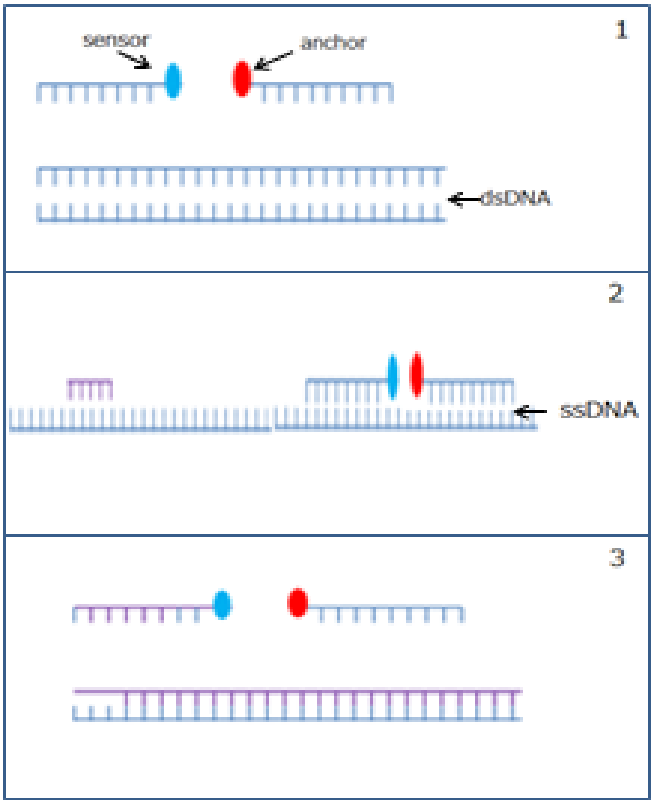


Figure 4.4. Overview of how HybProbes Function. (1) The sensor (donor) probe is labelled with a dye called Fluorescein (blue circle) at its 3' end and the anchor (acceptor) probe is labelled with a dye called BODIPY 630/650 (red circle) at its 5'end. No hybridisation occurs during PCR denaturation and therefore, no energy transfer takes place between the two dyes. (2) During the annealing phase, the HybProbes bind to the target DNA sequence in a head-to-tail fashion, causing the two dyes to be in close proximity to each other. The donor dye is excited by the blue LED light, which in turn excites the acceptor dye, which then emits fluorescent light at a different wavelength. (3) HybProbe probes are displaced during the elongation. (Images adapted from www.roche-applied-science.com)

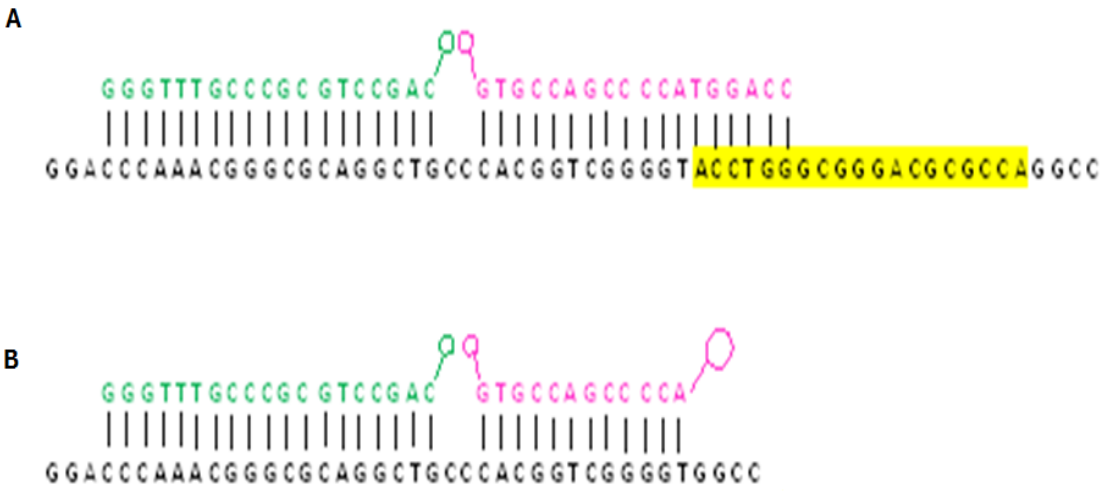


Figure 4.5. **Schematic diagram of the approach taken to design the genotyping assay for the *DHFR* 19bp deletion / insertion polymorphism.** The 19 bp deletion / insertion is highlighted in yellow. The sensor probe is represented in pink; while the anchor probe is represented in green **(A)** Perfect match between the sensor probe and the allele with the insertion. **(B)** Mismatch between the sensor probe and the allele with the deletion. Due to the greater complementary base pairing between the sensor probe and the insertion allele, those with the insertion allele will have a higher melting temperature than those with the deletion allele.

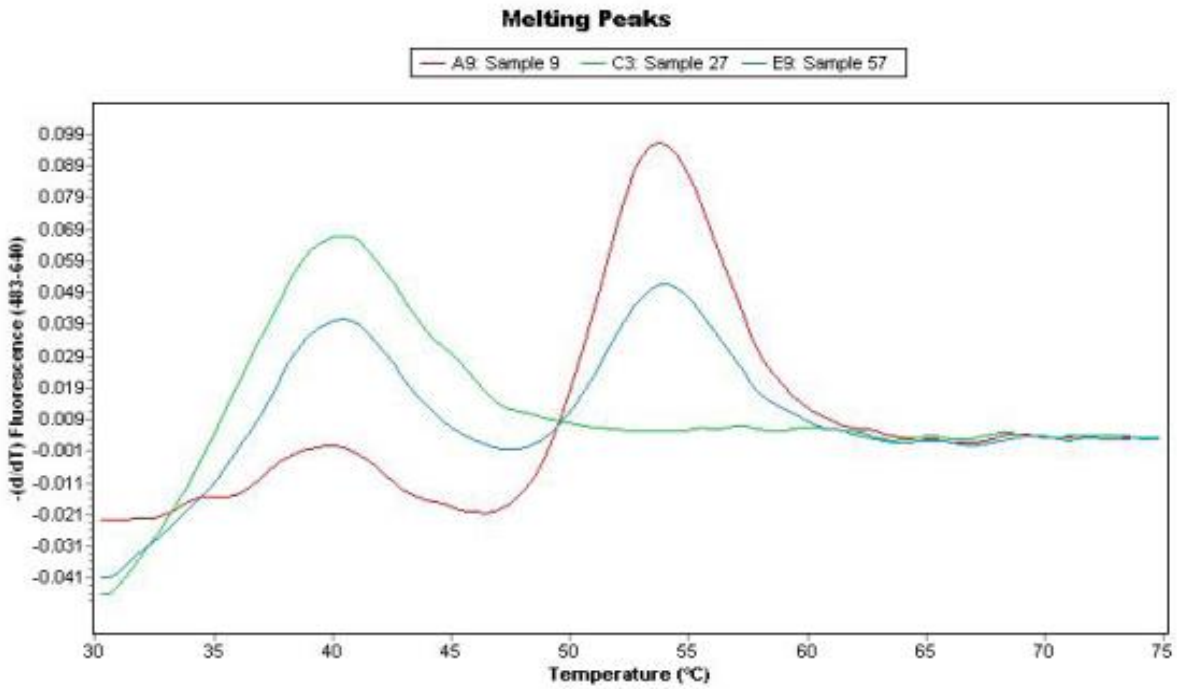


Figure 4.6. Melt curve analysis of the *DHFR* 19 bp deletion/insertion polymorphism. The single green peak at 40.5°C represents the homozygous deletion genotype. The single red peak at 54.5°C represents the homozygous insertion genotype. The two blue peaks at 40.5°C and 54.5°C represent the heterozygous genotype.

Table 4.1. Genotypes of the *DHFR* 19 bp deletion / insertion polymorphism in the TSS study analysed for both males and females together and separately.

DHFR	**Total [n (%)]**	**Males [n (%)]**	**Females [n (%)]**
WT	829 (33.1)	340 (32.8)	489 (33.3)
HET	1204 (48.1)	507 (48.9)	697 (47.5)
DEL	470 (18.9)	189 (18.2)	281 (19.2)
Allele 1 (WT)	2862 (57.2)	1187 (57.3)	1675 (57.1)
Allele 2 (DEL)	2144 (42.8)	885 (42.7)	1259 (42.9)

Table 4.2. Characteristics of the student who took part in the TSS study

	Males (n=1036)	**Females (n=1472)**
Age	22.63	22.32
BMI	23.4	22.69
Creatine	75.1	59.6
Total FA Intake	247.87	266.44
Haemoglobin	15.38	13.44
Smoking [n (%)]		
Yes	335 (32.4)	450 (30.7)
No	700 (67.6)	1018 (69.3)

Table 4.3. *DHFR* Genotype & Mean RCF, Serum folate & tHcy levels [mean±SD] in all samples.

DHFR	**RCF**	**Serum folate**	**tHcy**
WT	1083±432.21	33.92±20.68	8.71±2.96
HET	1081±432.11	34.37±20.66	8.66±2.95
DEL	1055±432.30	34.83±20.68	8.55±2.96
p-value	0.405	0.975	0.330

Table 4.4. *DHFR* Genotype & Mean RCF, Serum folate & tHcy levels [mean±SD] in males

DHFR	**RCF**	**Serum folate**	**tHcy**
WT	1108±414.65	33.02±17.18	9.35±3.32
HET	1095±414.19	31.63±17.16	9.59±3.31
DEL	1091±414.48	34.09±17.21	9.13±3.33
p-value	0.211	0.546	0.057

Table 4.5. *DHFR* Genotype & Mean RCF, Serum folate & tHcy levels [mean±SD] in females.

DHFR	**RCF**	**Serum folate**	**tHcy**
WT	1066±443.70	34.54±22.76	8.27±2.54
HET	1070±443.65	36.37±22.73	7.98±2.54
DEL	1031±443.07	35.33±22.74	6.67±2.54
p-value	0.355	0.399	0.301

Table 4.6. *DHFR* Genotype and Total Folate Intake [n (%)]

	Total Folate Intake			
DHFR	**None**	**Low**	**Medium**	**High**
WT	628 (32.98)	63 (31.34)	75 (37.31)	63 (31.98)
HET	915 (48.06)	104 (51.74)	89 (44.28)	96 (48.73)
DEL	361 (18.96)	34 (16.92)	37 (18.41)	38 (19.29)

Table 4.7. RCF [mean±SD] by *DHFR* Genotype and total folate intake

RCF (nM)	Total Folate Intake			
DHFR	**None**	**Low**	**Medium**	**High**
WT	1049±391.17	1130±425.37	1214±414.68	1221±511.63
HET	1041±425.12	1170±425.33	1200±456.38	1257±512.19
DEL	1015±413.71	1184±425.82	1172±423.20	1207±509.95
All	1035±410.00	1161±425.51	1195±431.42	1228±511.26

Table 4.8. Serum Folate [mean±SD] by *DHFR* Genotype and total folate intake

Serum Folate (nM)	Total Folate Intake			
DHFR	**None**	**Low**	**Medium**	**High**
WT	32.03±16.99	36.34±17.84	40.05±18.29	42.94±39.29
HET	32.07±17.21	36.18±17.84	42.08±20.80	47.27±39.03
DEL	33.17±19.23	39.63±17.86	39.97±18.43	41.32±39.02
All	32.42±17.81	37.38±17.85	40.70±19.17	43.84±39.11

Table 4.9. tHcy [mean±SD] divided by *DHFR* Genotype and total folate intake

tHcy (μM)	Total Folate Intake			
DHFR	**None**	**Low**	**Medium**	**High**
WT	8.92±3.66	8.22±2.72	7.93±3.36	8.07±2.71
HET	8.84±2.69	8.02±2.72	8.15±2.10	8.01±2.70
DEL	8.72±3.14	8.03±2.72	7.39±2.98	8.59±2.70
All	8.88±2.95	8.10±2.95	7.82±2.96	8.22±2.70

Chapter 5

Identification of *DHFRL1* polymorphisms, functionality testing and their relevance in human health and disease

5.1 Introduction

DHFRL1 or Dihydrofolate reductase like-1 is a gene that belongs to the *DHFR* gene family. Until recently, *DHFRL1* was thought to be a pseudogene but has now been characterised to be expressed and functional (Anderson et al., 2011, McEntee et al., 2011). As *DHFR* plays such an important role in folate mediated one-carbon metabolism, no non-synonymous polymorphisms have been found within the coding region of the gene. However, non-synonymous polymorphisms have been found to be present in the *DHFRL1* gene. These may contribute to health and disease. The aim of this chapter is to identify and validate *DHFRL1* polymorphisms and to determine their functionality and relevance in human health and disease.

5.1.1 DNA Polymorphisms

The human genome contains millions of different DNA polymorphisms (Sachidanandam et al., 2001). DNA polymorphisms are differences in the DNA sequence when compared to a reference standard, and are found to be present in at least 1-2 % of the population. They can range from a change in a single nucleotide to thousands of bases. Known polymorphisms include SNPs, insertions/deletions, short tandem repeats (1-8bp), variable number tandem repeats (8 to >50bp) and copy number variations in nucleotide repeat motifs. These polymorphisms may or may not have a phenotypic effect, depending on their position within the DNA sequence (Schork et al., 2000).

The most common type of polymorphism is a SNP, the change of a single nucleotide. SNPs can act as a landmark / marker for locating other genes / genetic regions and are often used to form dense genetic maps. However, they also have the potential to be functional and are often associated with both common and complex genetic diseases, as well as playing a role in altering an individual's response to drugs. SNPs that are located within the coding region of a gene can introduce variations that result in missense substitutions or stop codons, resulting in premature termination of the gene product, while SNPs that are located in the non-coding regions can influence gene expression by altering specific regulatory elements (Goto et al., 2001). Therefore, it is important to identify SNPs in the human genome and to evaluate the role they play in gene expression and drug response.

5.1.2 *DHFRL1* and its polymorphisms

In 2011, *DHFRL1*, a member of the *DHFR* gene family, located on chromosome 3 was found to be functional (Anderson et al., 2011, McEntee et al., 2011) with similar properties to *DHFR*, with the ability of complementing a *DHFR* negative phenotype in both bacterial and mammalian cells. Like *DHFR*, *DHFRL1* can also regulate its own translation by binding to its own

mRNA. Although *DHFR* and *DHFRL1* share many characteristics, they also have distinct differences. *DHFRL1* has a lower binding affinity for dihydrofolate (DHF) and is localized to the mitochondria (McEntee et al., 2011), the site of energy production in eukaryotic cells. In the mitochondria, *DHFRL1* takes the place of *DHFR* in the *de novo* thymidylate synthesis pathway (Anderson et al., 2011). As *DHFR* has been shown to play a key role in health and disease (Section 1.10 and 1.11), *DHFRL1* is also thought to have the possibility of playing a vital role in health and disease. Therefore, it is important to examine the functional characteristics not only of *DHFRL1* but also to examine the functional characteristics of any polymorphisms within the *DHFRL1* gene sequence.

Numerous synonymous (those that do not change the amino acid sequence of the protein) and non-synonymous (those that alter the amino acid sequence of the protein) *DHFRL1* polymorphisms have been identified and entered into the NCBI SNP database (http://www.ncbi.nlm.nih.gov/SNP). Of these, three non-synonymous polymorphisms, rs17855824, rs61739170 and rs114936057 within the coding region of the gene were chosen to be examined. The rs17855824 polymorphism results in a valine to isoleucine change, located at position 166 in the amino acid sequence. The rs61739170 polymorphism results in a proline to alanine change, located at position 83 in the amino acid sequence. The rs114936057 polymorphism results in a methionine to isoleucine change, located at position 38 in the amino acid sequence (Figure 5.1). Although none of the amino acids that are involved in the polymorphisms are located in the active site of *DHFRL1*, the changes in amino acids could alter the structure of the protein, causing differences in enzyme activity, resulting in alterations in the function of the enzyme. An Irish NTD (Section 5.1.3) was also screened for a number of *DHFRL1* polymorphisms, some of which are located in the coding region and others which are located both upstream and downstream of the coding regions, in order to determine the relevance of *DHFRL1* polymorphisms in health and disease.

5.1.3 Irish Neural Tube Defect (NTD) cohort

The Irish NTD cohort are a set of DNA samples from NTD affected triads (mother, father and their NTD-affected child) that were recruited throughout Ireland from 1993-2005 with the assistance of the Irish Association for Spina Bifida and Hydrocephalus (Shields et al., 1999, Brody et al., 2002, Parle-McDermott et al., 2009). The complete Irish NTD case population consists of a total of 586 families. However, triad families were not always complete as samples from all three family members were not always available. The NTD cohort consists of 442 full family triads, 53 case–mother pairs, 5 case–father pairs, 46 cases only and 40 families with parents or single parent

only. The control population are samples that were obtained from 56,049 pregnant women who did not give birth to an NTD-affected child and had no previous history of an NTD-affected pregnancy and were attending one of the three main maternity hospitals, The National Maternity Hospital Holles Street, The Coombe Women's Hospital or The Rotunda Hospital in Dublin between 1986-1990 (Kirke et al., 1993, Daly et al., 1995, Mills et al., 1995). These NTD samples and a random sample of the control samples were used in our study in order to determine the relevance of *DHFRL1* polymorphisms in health and disease.

5.1.4 Objectives

As mentioned earlier, recombinant DHFRL1 (McEntee et al., 2011) and endogenous DHFRL1 in cancer cells (Anderson et al., 2011) has been shown to localise to the mitochondria. We wanted to assess the presence of endogenous DHFRL1 in the mitochondria of non-cancerous cells, and to determine whether or not endogenous DHFRL1 and DHFRL1 polymorphisms harbour enzyme activity in the mitochondria in order to test for functionality. *DHFRL1* polymorphisms will also be screened in an NTD cohort in order to determine their relevance in health and disease.

Assessment of the functionality of *DHFRL1* polymorphisms were examined using the following methodology

- Validation of candidate *DHFRL1* polymorphisms by gel electrophoresis and sequencing.
- Development of a genotyping assay on the LightCycler 480 for the *DHFRL1* SNPs.
- Screening the *DHFRL1* rs17855824 SNP in a panel of Coriell DNA samples.
- Culturing of Coriell and HEK 293 cell lines.
- Extraction of cytoplasmic and mitochondrial protein fractions from the cell lines.
- Confirmation of endogenous mitochondrial DHFRL1 and its variants by Western blot.
- Enzyme kinetic analysis of endogenous mitochondrial DHFRL1 and its polymorphisms.
- Screening and analysis of *DHFRL1* polymorphisms in an Irish NTD cohort

5.2 Results

5.2.1 Validation of candidate *DHFRL1* polymorphisms

The NCBI SNP database (http://www.ncbi.nlm.nih.gov/snp) was examined in order to determine what polymorphisms had been reported in the *DHFRL1* gene. Four polymorphisms within the coding region of the *DHFRL1* gene had been identified when we examined the database early in 2011. Of these four DHFRL1 polymorphisms, three were non-synonymous polymorphisms,

the *DHFRL1* rs17855824 SNP (G->A), the rs61739170 SNP (C->G) and the rs114936057 (G->T). These three SNPs were chosen to be examined (Figure 5.1).

In order to validate these polymorphisms, a multi-nucleotide sequence analysis of the *DHFR* gene family was first carried out using clustalW to facilitate primer design (Figure 5.2) (See Appendix L for full clustalW alignment). ClustalW alignment showed high similarity / homology between *DHFRL1* and all of the other genes in the *DHFR* gene family. As there are not many differences between all of the genes, validation of the polymorphisms proved to be a big challenge. Primers had to be designed specifically for each of the non-synonymous polymorphisms (Table 2.10) (See Appendix L for primer location), ensuring that the primers only amplified the regions of the *DHFRL1* gene under investigation, without the possibility of amplifying other homologous regions in the *DHFR* gene or the other pseuodogenes in the *DHFR* gene family.

A set of Coriell DNA samples (DNA from human lymphoblasts) from both African American and European American descent as well as a number of DNA samples from different cancer cell lines were chosen to be examined and were amplified by PCR. Once PCR conditions had been optimized, amplification was firstly confirmed by running the PCR products on a 2 % agarose gel (Figure 5.3 – 5.5). The PCR products were then cleaned up with ExoSAP-IT, a reagent which removes any unconsumed dNTPs or primers remaining in the PCR product mixture as described in Section 2.2.2. The PCR products were then sent off for Sanger sequencing (Source BioScience). Due to the homology between the members of the *DHFR* gene family, the chromatograms had to be analysed with a fine tooth comb and the sequences were aligned and examined to determine the genotype of the DNA samples (Table 5.1) (See Appendix M for sequence alignments). An example of a 'dirty' sequencing run, a trace that not only picks up the *DHFRL1* gene but also other members of the *DHFR* gene family (Figure 5.6) and that of a clean sequencing run, one that picks up only the *DHFRL1* gene (Figure 5.7) are shown below (See Appendix N for Chromatograms of each of the three polymorphisms).

Although one of our assays, the *DHFRL1* rs61739170 (C->G) assay seemed to be susceptible to producing a 'dirty' assay, the mixed signal was not coming from the SNP and therefore, after careful analysis of the traces, we concluded that the three polymorphisms were 'real' variants. However, direct sequencing and chromatogram analysis of the samples for the *DHFRL1* rs114936057 (G>T) polymorphism found this polymorphism not to be frequent. Therefore, we

focused our analysis on the two common variants *DHFRL1* rs61739170 and rs17855824, both of which result in an amino acid change.

5.2.2 Development of a genotyping assay for two *DHFRL1* polymorphisms.

Once the *DHFRL1* polymorphisms had been confirmed to be real variants, we first had to develop a genotyping assay for the *DHFRL1* SNPs using melt curve analysis with HybProbes on the Roche Lightcycler® 480 as described in Section 4.1.5 and 4.2.1 in order to be able to screen a large panel of DNA samples accurately. This would allow us to identify the genotype of the DNA samples so that the polymorphisms could be characterised. The rs61739170 SNP is located in a region of high homology with other members of the *DHFR* gene family and although primers and probes were designed carefully in order to try and only amplify up the region of the SNP, a genotyping assay on the Roche Lightcycler could not be designed. However, a genotyping assay was successfully designed and optimised for the rs17855824 SNP (Figure 5.8). The sensor probe was designed so that it would bind fully complementary to the wild type alleles, resulting in those with the wild type alleles having a higher melting temperature in comparison with those who had the alleles with the polymorphism due to greater complementary base pairing between them.

5.2.3 Screening and analysis of the *DHFRL1* rs17855824 polymorphism in a panel of Coriell DNA samples.

Once the *DHFRL1* rs17855824 polymorphism genotyping assay had been developed and optimised on the LightCycler® 480, a panel of 169 Caucasian and African American Coriell DNA samples were screened for the polymorphism. Samples were genotyped twice in order to ensure that the assay was working efficiently and that the software was calling the correct genotypes as we had issues with contamination in our negative controls. The genotype results (numbers and percentages) as well as the allele frequencies obtained from the samples are shown in Table 5.2.

5.2.4 Examination of the presence of endogenous mitochondrial DHFRL1

Previous work carried out in the lab had determined the presence of recombinant DHFRL1 in the mitochondria (McEntee et al., 2011). Therefore, the presence of endogenous DHFRL1 in the mitochondrial needed to be determined before examining the functionality of the DHFRL1 rs17855824 polymorphism endogenously. HEK 293 and DHFRL1 over-expressed HEK 293 cells were cultured as described in Section 2.6.4.2. Mitochondrial and cytoplasmic protein fractions were extracted from the HEK 293 and DHFRL1 over-expressed HEK 293 cells, pooling four 75 cm^3

flasks together as described in Section 2.3.15 in order to obtain enough mitochondrial protein to carry out the experiments.

A Bradford assay (as described in Section 2.3.18) was carried out on a set of standard protein samples (Table 5.3, Figure 5.9) and the extracted mitochondrial and cytoplasmic protein fractions in order to determine the concentration of the extracted protein fractions. An example of the concentrations observed from the mitochondrial and cytoplasmic protein fractions obtained are shown in Table 5.4.

SDS-PAGE gels were then run and Western blots were carried out on the cytoplasmic and mitochondrial fractions in order to determine the presence or absence of DHFRL1 in the mitochondria and DHFR in the cytoplasm (Figure 5.10-5.11). In order to ensure that our mitochondrial protein fractions were clean, i.e. did not contain any cytoplasmic protein, the mitochondrial protein fractions were probed with GAPDH, a positive control for cytoplasmic protein. Cytoplasmic protein samples were also probed for GAPDH at the same time in order to ensure the antibody was working sufficiently. No bands were seen in our mitochondrial protein fractions confirming that the mitochondrial protein fractions were clean. An example of both cytoplasmic and mitochondrial protein fraction samples probed with the GAPDH antibody is shown in Figure 5.10. Once it was confirmed that our mitochondrial protein fractions were pure, both cytoplasmic and mitochondrial protein fractions were examined for the presence of DHFR in the cytoplasm and DHFRL1 in the mitochondria (a 21 kDa band). Figure 5.11 shows an example of a Western blot carried out on the cytoplasmic protein fractions and Figure 5.12 shows an example of a Western blot carried out on the mitochondrial protein fractions. The Western blots (Figure 5.11-Figure 5.12) showed that DHFR is present in the cytoplasmic protein fractions and that DHFRL1 is present in the mitochondrial protein fractions. As Figure 5.9 confirms that the mitochondrial protein fractions are clean, we can confirm that the DHFRL1 bands seen in the mitochondrial protein fractions are real and are not contaminated with any cytoplasmic protein.

5.2.5 Examination of endogenous mitochondrial DHFRL1 enzyme activity

Once the presence of endogenous DHFRL1 in the mitochondria was confirmed by Western blot, the presence / absence of endogenous DHFRL1 enzyme activity in the mitochondria was examined. In order to determine DHFRL1 enzyme activity in the mitochondria, the Dihydrofolate Assay Kit was used as described in Section 2.3.21. The DHFR control and blanks were tested first

in order to confirm that all the reagents were working correctly within the region of detection by the spectrophotometer (Figure 5.13-5.16).

Once it was confirmed that the reagents and the assay was working well, the assay was carried out on the mitochondrial protein fractions. The assay was carried out on both HEK 293 and DHFRL1 over expressed mitochondrial protein samples. An example of each is shown in Figure 5.17–5.20. Assays were repeated twice with different batches of mitochondrial protein extracted from the cells at different times. The decrease in ΔOD / min (the decrease in absorbance per minute) was then calculated for each of the samples. In order to obtain the true ΔOD of DHFR / DHFRL1, the ΔOD / min of the sample + MTX was taken away from the ΔOD / min of the sample. This is to work out the proportion of NADPH loss that is due to DHFR activity in endogenous samples. After deduction, the true ΔOD / min of DHFRL1 in HEK 293 cells was determined to be 0.0089. The specific activity was then calculated as mentioned in Section 2.3.21. The specific activity of DHFRL1 in HEK293 cells was found to be 14.88 nmol/min/mg protein, indicating the presence of endogenous DHFRL1 activity in the mitochondria of normal human cell lines. The true ΔOD / min of DHFRL1 was also calculated in the DHFRL1 over-expressed HEK 293 cells and was determined to be 0.01, indicating the presence of endogenous DHFR / DHFRL1 mitochondrial activity. The specific activity was then calculated as mentioned in Section 2.3.21. The specific activity of DHFRL1 in DHFRL1 over-expressed HEK293 cells was found to be 14.11 nmol/min/mg protein, also indicating the presence of endogenous DHFRL1 activity in the mitochondria of normal human cell lines.

5.2.6 Examination of the presence of endogenous mitochondrial DHFRL1 in Coriell Cell Lines

Once the genotypes of the Coriell DNA samples had been identified by melt curve analysis, two Coriell cell lines for each of the DHFRL1 rs17855824 genotypes were cultured as mentioned in Section 2.6.4.1. Mitochondrial and cytoplasmic protein fractions were extracted from the Coriell cell lines, pooling six to eight 25 cm^3 flasks together. Protein concentrations were determined by Bradford assay as before. However, the protein concentration values that were obtained from the Coriell mitochondrial protein fractions were less than half, and in some samples, less than a quarter of that when compared to those obtained for the HEK 293 and DHFRL1 over-expressed HEK 293 mitochondrial protein fractions. Nonetheless, Western blots were carried out on the mitochondrial protein fractions and were probed with both the DHFR and GAPDH antibodies as before (Figure 5.21-5.22).

No bands were seen in our mitochondrial protein fractions when they were probed with the GAPDH antibody, confirming that the mitochondrial protein fractions were clean, indicating that the DHFRL1 bands seen in the Coriell mitochondrial protein fractions are real for all three of the DHFRL1 rs17855824 genotypes.

5.2.7 Examination of endogenous mitochondrial DHFR / DHFRL1 enzyme activity in the Coriell samples

Once the presence of endogenous DHFRL1 was confirmed in the Coriell mitochondrial protein fractions by Western blot, the presence / absence of endogenous DHFRL1 enzyme activity in the Coriell mitochondrial protein fractions was examined in order to try and determine whether the DHFRL1 rs17855824 polymorphism causes a change in DHFRL1 enzyme activity. In order to determine DHFRL1 enzyme activity in the mitochondria, the Dihydrofolate Assay Kit was used as described in Section 2.3.21. As before, the DHFR control and blanks were tested first in order to confirm that all the reagents were working correctly within the region of detection by the spectrophotometer.

Once it was shown that the controls were working correctly, enzyme kinetic analysis was carried out on the Coriell mitochondrial protein fractions. As mentioned above (Section 5.2.6), the protein concentrations of the Coriell mitochondrial protein fractions were very low in comparison to those of the HEK 293 and DHFRL1 over-expressed HEK 293 mitochondrial protein fractions. Therefore, when enzyme activity analysis was carried out on the different Coriell mitochondrial protein samples, no enzyme activity could be detected in any of the mitochondrial protein samples. We did however; generate recombinant protein versions of the polymorphisms and this is described in Chapter 6.

5.2.8 Examination of DHFRL1 polymorphisms within an NTD cohort

Although, the enzyme activity of the different genotypes of the rs17855824 polymorphism could not be determined due to the concentration of the mitochondrial fractions being too low, the fact that polymorphisms existed within the *DHFRL1* gene was interesting in itself and therefore, *DHFRL1* polymorphisms, including the rs17855824 SNP were examined and screened in an NTD cohort by our collaborators Dr. Faith Pangilinan and Dr. Larry Brody at the National Human Genome Research Institute–National Institutes of Health (NHGRI-NIH) and Prof. Anne Molloy in TCD. This was carried out in order to determine whether or not any of the *DHFRL1* polymorphisms

are associated with NTDs in an Irish cohort. All the analyses described below were carried out by our collaborators.

Seven *DHFRL1* polymorphisms (rs10454213, rs11927165, rs17855824, rs61739170, rs7644176, rs7645522 and rs7653521) were genotyped in the Irish NTD cohort as well as the controls. The genotype frequency results (numbers and percentages) for each of the *DHFRL1* polymorphisms are shown in Table 5.5.

After careful analysis, three of the *DHFRL1* polymorphisms, rs7645522, rs10454213 and rs17855824 were found to be in linkage disequilibrium with each other, i.e. the alleles were non-randomly associated at the three loci, in the Irish NTD cohort, with an r^2 value of 1 (Table 5.6). Linkage disequilibrium was also found in the Irish controls, with the rs7645522, rs10454213 and rs17855824 *DHFRL1* polymorphisms having an r^2 value of > 0.99, while the rs11927165 and rs7644176 polymorphism had an r^2 value of 0.93. The positions of the DHFRL1 polymorphisms are also shown in Table 5.6. The rs7645522 and rs10454213 SNPs are located ~ >10 kb and ~ 5 kb from the *DHFRL1* gene, while the rs17855824 SNP is located within the coding regions of the *DHFRL1* gene.

A transmission disequilibrium test (TDT), a family based association test for the presence of genetic linkage between a genetic marker and a trait was then carried out on the data. The TDT test showed that the rs7645522, rs10454213 and rs17855824 *DHFRL1* polymorphisms are in high linkage disequilibrium with each other and that they are associated with NTDs, with significant p-values (Table 5.7). No significant p-values were observed for the other *DHFRL1* polymorphisms. The table also shows the allele which is being over-transmitted for all seven *DHFRL1* polymorphisms

5.3 Discussion

The aim of this chapter was to identify and validate *DHFRL1* polymorphisms and to determine their functionality and relevance in human health and disease. By examining the dbSNP database, we were able to identify a number of *DHFRL1* polymorphisms within the coding region of the *DHFRL1* gene. *DHFRL1*, a member of the *DHFR* gene family, which was until recently thought to be a pseudogene is highly homologous to *DHFR*. *DHFR* has been studied extensively and no polymorphisms have been found within its coding regions (Litwack, 2008).Therefore, the identification of polymorphisms within the *DHFRL1* coding region was of great interest to us.

Three *DHFRL1* polymorphisms found within the coding region of *DHFRL1* were validated *via* PCR amplification and Sanger sequencing in both Coriell DNA and DNA from various cancer cell lines. From this, it was detected that two of the polymorphisms were common in the population and therefore of most interest to us. In order to functionally characterise these polymorphisms, we wanted to examine these polymorphisms endogenously. Although recombinant DHFRL1 was previously examined in the lab (McEntee et al., 2011), we first had to demonstrate endogenous mitochondrial DHFRL1 activity in our cell lines of interest.

HEK 293 and DHFRL1 over-expressed HEK 293 cell lines were cultured and both cytoplasmic and mitochondrial protein fractions were extracted. Western blots carried out on the samples detected DHFR in the cytoplasm and DHFRL1 in the mitochondria. The extracted mitochondrial protein fractions were shown to be pure, with no cytoplasmic contamination by immunoblotting against GAPDH, a positive control for cytoplasmic protein, indicating that DHFRL1 is present in the mitochondria. Enzyme kinetic analysis carried out on the mitochondrial fractions showed that DHFRL1 has enzyme activity in the mitochondria in these cell lines.

Therefore, we wanted to determine whether the DHFRL1 rs17855824 polymorphism could alter DHFRL1 enzyme activity. A genotyping assay was successfully designed to determine the genotypes of a large cohort of Caucasian and African American Coriell DNA. This was carried out so that the different rs17855824 genotypes could be identified and so that Coriell cell lines with the different genotypes could be cultured. Western blots carried out on the extracted mitochondrial protein fractions confirmed again that DHFRL1 is present in the mitochondria. However, due to the low protein concentrations obtained for the Coriell mitochondrial protein fractions, no enzyme activity could be detected in any of the Coriell samples.

In order to obtain enough mitochondrial protein (~ the same concentrations that were being obtained from the HEK 293 and DHFRL1 over-expressed HEK 293 cells) from the Coriell cell lines, we would have had to quadruple the amount of cells that were cultured in order to carry out the enzyme kinetic analysis experiments. As six - eight 25 cm^3 flasks were already being pooled together, it was not practical to grow 32 flasks for each of the genotypes in order to just carry out the experiment just once. Cost restraints would have also prevented us to carry this out. The low protein concentrations observed for the Coriell cell lines may be due to a number of reasons. Coriell cells lines are suspension cells unlike the adherent HEK 293 cells and are also typically grown in 25

cm^3 flasks. Therefore, in the future, it may be possible to adapt the Coriell cell lines so that they would grow in 75 cm^3 flasks, so that a larger number of cells could be collected, hopefully resulting in higher protein concentrations.

The sensitivity of the DHFR assay kit was also unknown and therefore, even if enough mitochondrial protein could have been obtained, the DHFR assay kit may not have been specific enough to measure the difference in enzyme activity between the different *DHFRL1* rs17855824 polymorphism genotypes, as the SNP only results in a single base change with unknown enzyme impact. Enzyme analysis carried out on the HEK 293 and DHFRL1 over-expressed HEK 293 cells showed that DHFRL1 is not the only enzyme in the mitochondria that utilises NADPH as part of their reaction system as a decrease in absorbance was still observed even after methotrexate, an inhibitor of DHFR was added to the reaction, indicating the presence of other enzymes that utilise NADPH (as expected). Therefore, if the Coriell cell lines could be adapted to grow in larger volumes and larger amounts of mitochondrial protein could be collected, one could purify out the DHFRL1 protein from the mitochondrial fractions by immunoprecipitation, allowing us to carry out the enzyme kinetic analysis on a purified endogenous protein. However, high concentration yields would be required. In order to carry out immunoprecipitation, one requires ~300-500 µg of total protein. As the protein yield obtainable from the mitochondrial extraction kit is 20-80 µg of total protein from 5 X 10^6 - 2 X 10^7 cells, pooling of large amounts of mitochondrial protein fractions would be required in order to obtain enough protein to carry out immunoprecipitation.

Another alternative method that could be used in the future to examine the functionality of polymorphisms could be using newer technologies such as genome editing with engineered nucleases (GEEN). GEEN is a form of engineering in which DNA can be inserted, replaced or removed from a genome using artificially engineered nucleases. These nucleases are programmable and function by creating specific double-stranded breaks (DSBs) at the location of interest in the genome and then make use of the cells endogenous mechanisms to repair the induced break by natural processes such as homologous recombination (HR) and non-homologous end-joining (NHEJ) (Kim and Kim, 2014). HR relies on a homologous sequence to repair the DSBs. By exploiting this system, a sequence of interest can be inserted within a sequence that is homologous to the flanking sequences of a DSB, leading to the formation of the desired change within the genomic region of interest (Rouet et al., 1994). NHEJ uses a variety of enzymes to directly join the DNA ends in a DSB. NHEJ is error prone and has been shown to cause mutations at the repair site in ~ 50 % of DSBs. Therefore, if one could create a DSB at a desired region of the gene in multiple

samples, it would be very likely for the mutation to be generated at the region of interest in at least some of the samples (Bibikova et al., 2002). In this way, if one can create specific DSBs at specific locations, the cell's own repair system can be exploited to insert specific mutations at regions of interest (Figure 5.23).

Engineered nucleases include zinc finer nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), RNA-guided engineered nucleases (RGENs) and engineered meganuclease re-engineered homing endonucleases. ZFNs, TALENs and RGENs are all based on the same principal and work by the same mechanism; they cleave the DNA in a site-specific manner, which then triggers the cells endogenous repair mechanisms resulting in targeted genome modification (Kim and Kim, 2014). The choice of nuclease and their success rate depends on the cell type and delivery method, i.e. transient transfection by electroporation, liposome transfection, microinjection, lentivirus vectors, adenoviruses etc. However, to date, there are no reliable rules to predict which nuclease will work best before experimental confirmation.

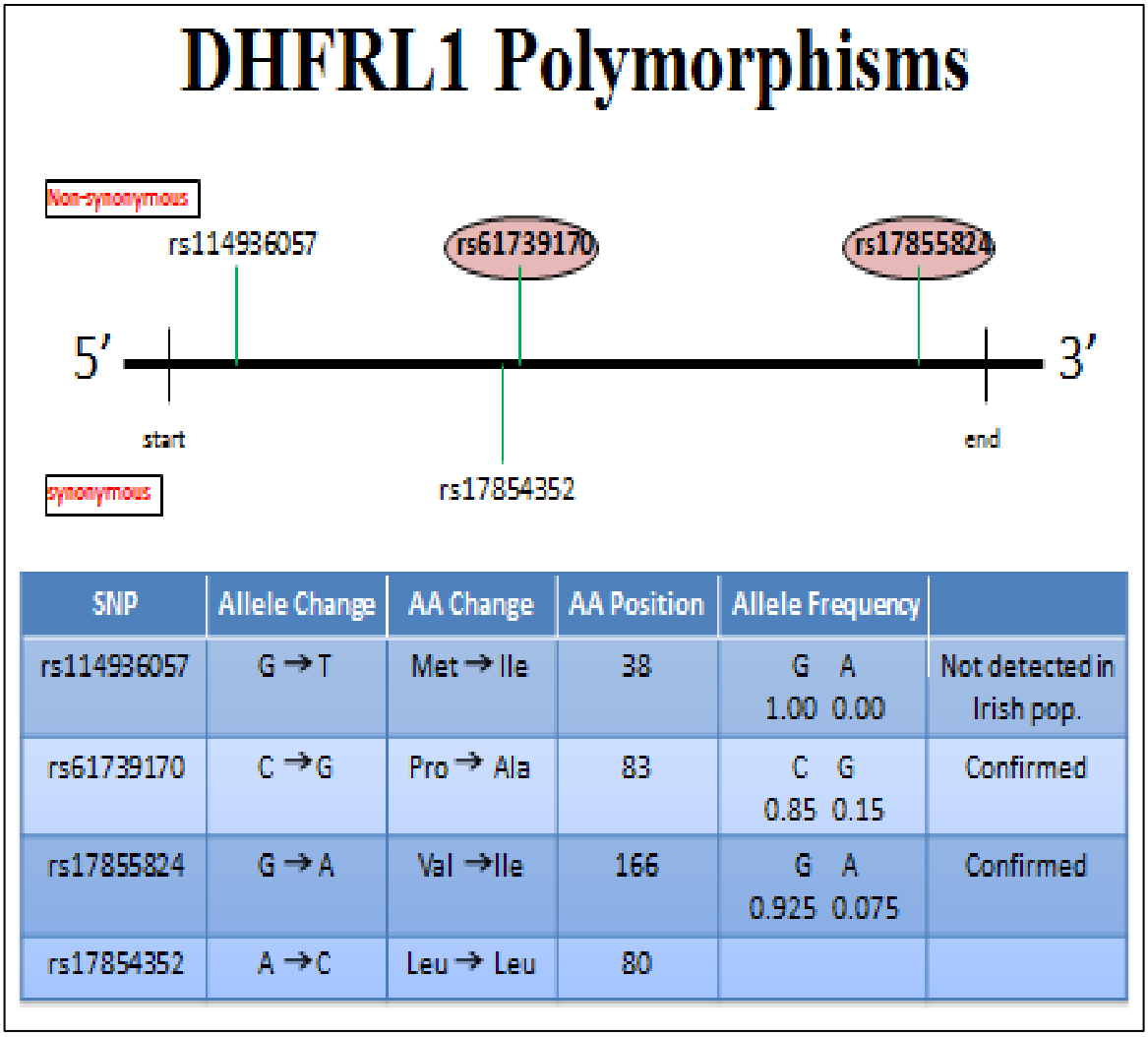
These new methods could be used to try and edit the endogenous sequence in order to incorporate our polymorphism of interest into HEK 293 cells, which would allow us to produce and extract higher protein concentrations, enabling us to both purify and carry out enzyme kinetic analysis on the mitochondrial protein fractions, allowing us to determine whether different genotypes alter the enzyme activity of a specific gene.

Although we could not obtain enough mitochondrial protein to carry out enzyme kinetic analysis in order to determine the effect of the *DHFRL1* rs1785824 polymorphism, we have confirmed that endogenous DHFRL1 is present in the mitochondria in both HEK 293 and Coriell cell lines and that it has enzyme activity in normal human cells (HEK 293 cells). Endogenous mitochondrial DHFRL1 enzyme activity has been previously shown in HepG2 cells, a cancerous liver cell line (Anderson et al., 2011). Anderson *et al*., treated HepG2 cells with scrambled control siRNA and siRNA specific to the 5′ UTR of *DHFRL1* that is not present in the *DHFR* transcript. They showed that HepG2 mitochondrial protein extracted from the cells that had been treated with the control siRNA had a DHFR enzyme activity of 32.4±4.0 nmol/min/mg protein, while those which had been treated with the siRNA that was specific to DHFRL1 had an activity of 0.17±0.0008 nmol/min/mg protein, indicating that DHFRL1 is active in the mitochondria. The activity was significantly different with a p value of 0.0002 (n=3). Likewise to our experiments, extracted mitochondrial protein fractions were tested for purity by immunoblotting against Lamin A

(nuclear) and GAPDH (cytoplasmic) and mitochondrial protein samples were shown to exhibit no nuclear or cytoplasmic contamination. Mitochondrial protein fractions that were collected from the cells which had been treated with the siRNA specific to DHFRL1, i.e. in which DHFRL1 had been knocked down, no DHFRL1 protein was found to be present, with COXIV being used as a positive mitochondrial control. DHFR activity was examined in the cytoplasmic protein fractions from those extracted from both the scrambled and DHFRL1 specific siRNA treatment cells and were found to be unaffected by the siRNAs with specific activities of 4.5 ± 2.6 and 4.5 ± 2.8 μmol/min/mg protein observed, indicating that the DHFRL1 knockdown did not affect DHFR activity.

Although enough mitochondrial protein could not be obtained from the Coriell cell lines in order to determine the effect of the *DHFRL1* rs1785824 polymorphism on enzyme activity, genotyping of the Irish NTD cohort with seven *DHFRL1* polymorphisms carried out by our collaborators in the NHGRI-NIH and TCD have shown that three out of the seven *DHFRL1* polymorphisms analysed, including the rs17855824 polymorphism are in high linkage disequilibrium and are associated with the risk of NTDs in an Irish population, which is of great interest. How the *DHFRL1* polymorphisms are affecting NTD risk is yet unknown and further analyses will have to be carried out in order to elucidate the exact mechanism as to how these *DHFRL1* polymorphisms affect NTDs.

Our findings have shown that HEK 293 cells and DHFRL1 over-expressed HEK 293 cells harbour enzyme activity with an enzyme activity of 14.88 nmol/min/mg protein and 14.11 nmol/min/mg protein respectively. The enzyme activity observed in our mitochondrial protein fractions is lower than that found for the HepG2 mitochondrial protein fractions by Anderson *et al*. However, Anderson *et al*. examined DHFRL1 activity in HepG2 cells, a cancerous liver cell line, and therefore, one would expect it to have a higher proliferation rate than that of normal human cell lines, and therefore, most likely to have a higher enzyme activity. This is the first time that it has been shown that endogenous DHFRL1 is functionally active in the mitochondria of normal healthy human cells. Although DHFRL1 has been shown to be active, the exact function that DHFRL1 plays in the mitochondria is yet unknown, however, it could be one of great importance in human health and disease especially as *DHFRL1* polymorphisms have been shown to be associated with NTDs.



SNP	Allele Change	AA Change	AA Position	Allele Frequency	
rs114936057	G → T	Met → Ile	38	G A 1.00 0.00	Not detected in Irish pop.
rs61739170	C → G	Pro → Ala	83	C G 0.85 0.15	Confirmed
rs17855824	G → A	Val → Ile	166	G A 0.925 0.075	Confirmed
rs17854352	A → C	Leu → Leu	80		

Figure 5.1. An overview of the *DHFRL1* polymorphisms. The schematic above shows the rough positions of the four *DHFRL1* polymorphisms on the *DHFRL1* gene. The three polymorphisms above the black line represent non-synonymous polymorphisms while the polymorphism below the black line is synonymous. The table shows the characteristics of each of the polymorphisms. The rs114936057 SNP results in a methionine to isoleucine change at position 38, the rs61739170 SNP results in a proline to alanine change at position 83, and the rs17855824 SNP results in a valine to isoleucine change at position 166.

DHFR Gene Family Alignment

Figure 5.2. ClustalW alignment of the *DHFR* gene family. The clustalW alignment shows that there is high homology with very little differences between *DHFRL1* and all of the other genes in the *DHFR* gene family. The rs61739170 SNP is shown in dark yellow, while the primers for the rs61739170 SNP are shown in light blue. Primers were designed so that they would only amplify up the region of interest.

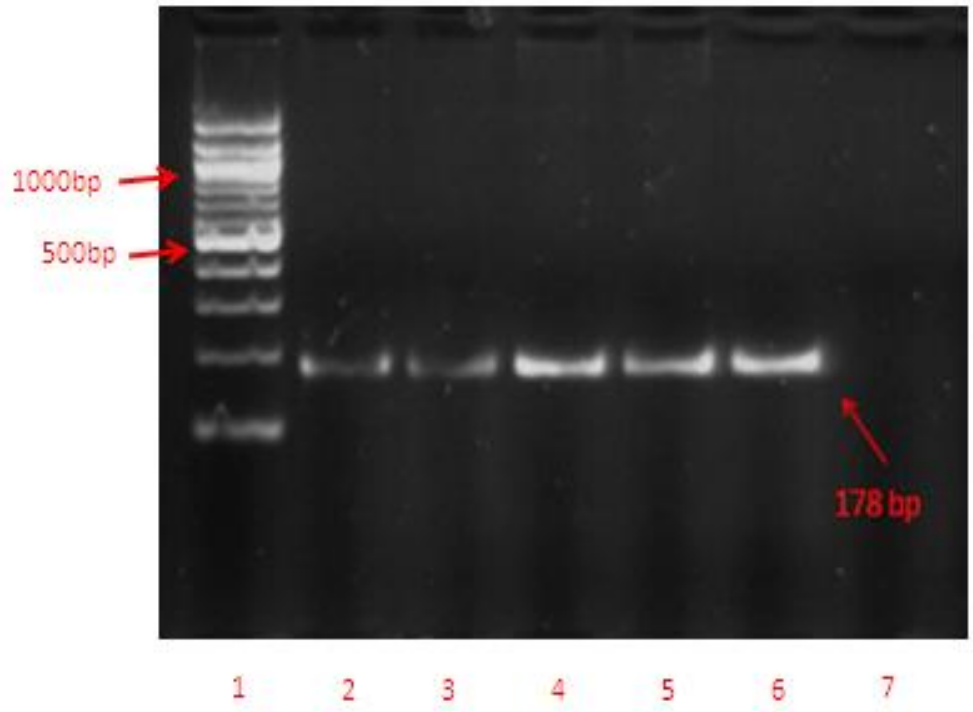


Figure 5.3. Gel electrophoresis of DNA samples amplified by specific *DHFRL1* primers designed to amplify the region containing the rs17855824 polymorphism. Lane 1 – 100 bp DNA ladder, Lane 2 - Coriell sample 17229, Lane 3 - Coriell sample 17214, Lane 4 – Coriell sample 17147, Lane 5 – Coriell sample 17146, Lane 6 Coriell samples 17234, Lane 7 – negative control. The gel shows bands of the correct size (178 bp), indicating that the DNA samples have been amplified by the primers.

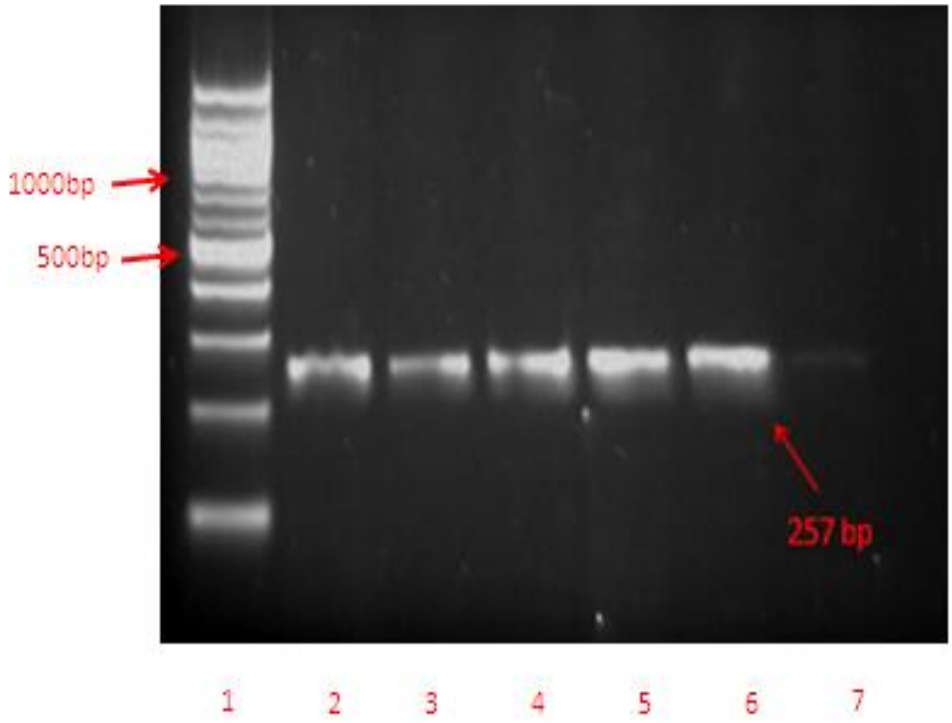


Figure 5.4. Gel electrophoresis of DNA samples amplified by specific *DHFRL1* primers designed to amplify the region containing the rs61739170 polymorphism. Lane 1 – 100 bp DNA ladder, Lane 2 - Coriell sample 17265, Lane 3 - Coriell sample 17133, Lane 4 – Coriell sample 17219, Lane 5 – Coriell sample 17102, Lane 6 Coriell samples 17259, Lane 7 – negative control. The gel shows bands of the correct size (257 bp), indicating that the DNA samples have been amplified by the primers.

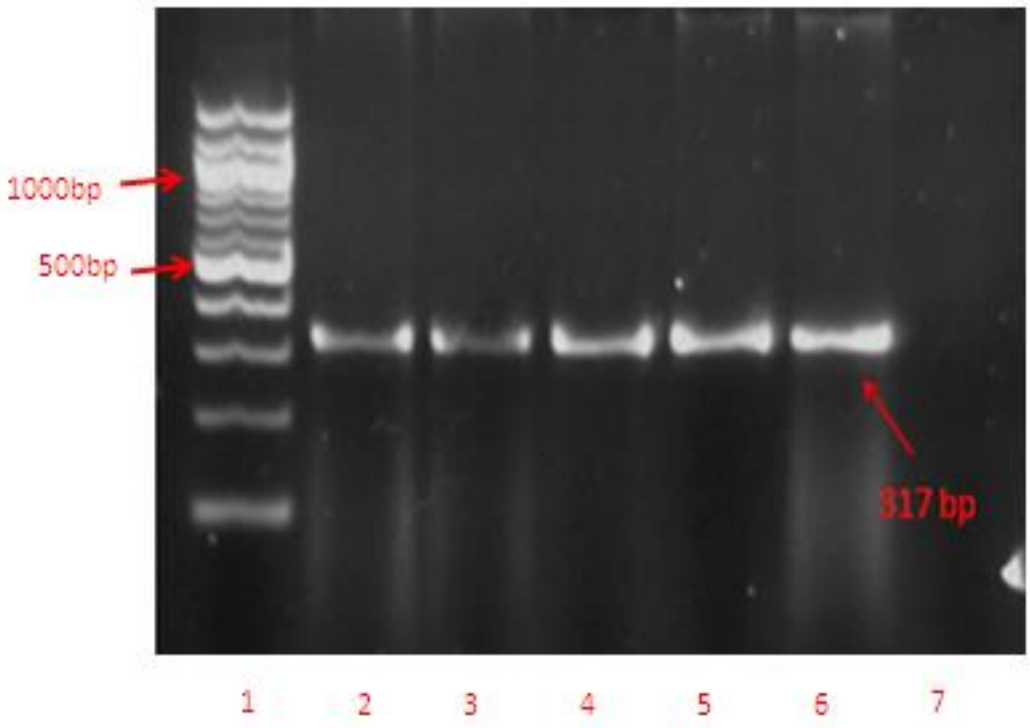


Figure 5.5. Gel electrophoresis of DNA samples amplified by specific *DHFRL1* primers designed to amplify the region containing the rs114936057 polymorphism. Lane 1 – DNA ladder, Lane 2 - Coriell sample 17206, Lane 3 – SW480, Lane 4 – Coriell sample 17145, Lane 5 – Coriell sample 17158, Lane 6 Coriell samples 17165, Lane 7 – negative control. The gel shows bands of the correct size (317 bp), indicating that the DNA samples have been amplified by the primers.

Dirty Assay

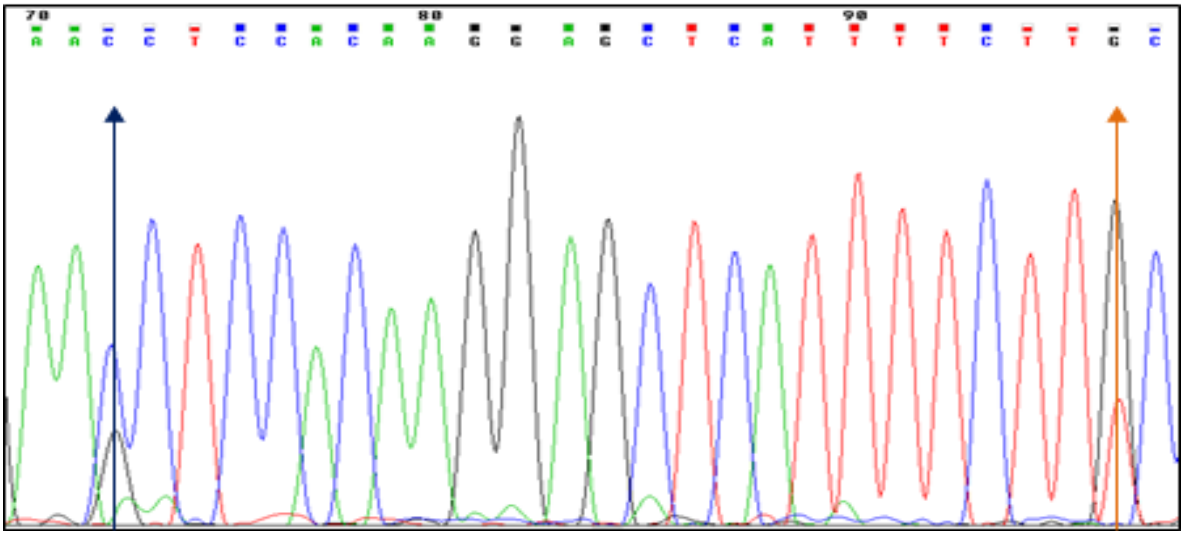


Figure 5.6. An example of a dirty chromatogram from a Sanger sequencing run of one of the rs61739170 (C>G) polymorphism DNA samples. The navy arrow indicates the location of the polymorphism, showing that this individual is heterozygous for the *DHFRL1* rs61739170 (C->G) polymorphism. However, we can see that the trace not only picks up the *DHFRL1* gene but also other members of the DHFR gene family, indicated by the two peaks observed at the orange arrow. The chromatogram shows evenly spaced peaks, with little baseline noise, indicating that our template and primers were good, as well as allowing us to determine the base at each of the polymorphisms.

Clean Assay

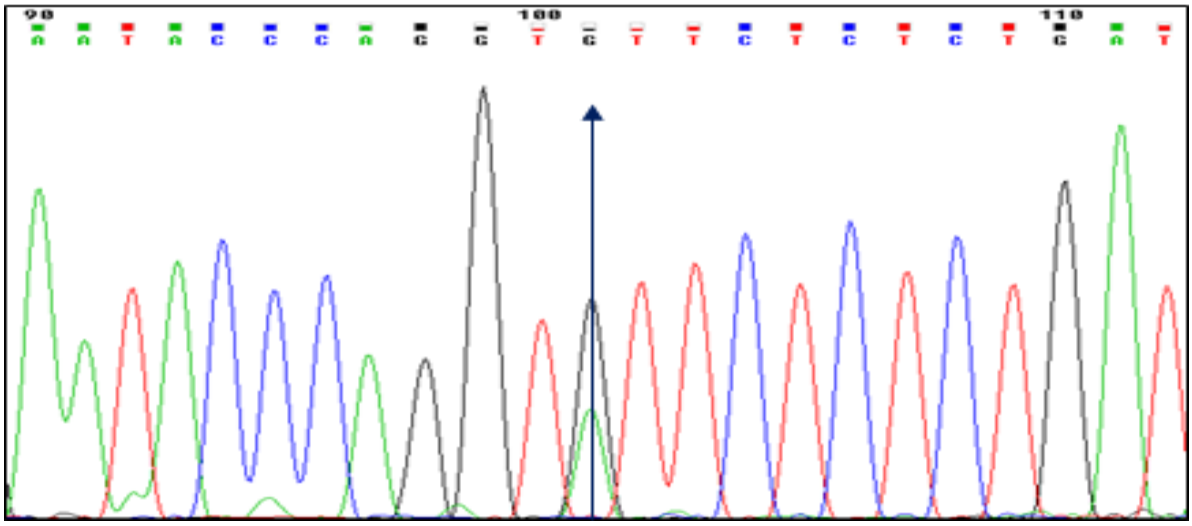


Figure 5.7. An example of a clean chromatogram from a Sanger sequencing run of one of the rs17855824 (G>A) polymorphism DNA samples. The navy arrow indicates the location of the polymorphism, showing that this individual is heterozygous for the *DHFRL1* rs17855824 (G->A) polymorphism. This trace only picks up the *DHFRL1* gene as indicated by the single peaks observed at each base The chromatogram shows evenly spaced peaks, with little baseline noise, indicating that our template and primers were good, as well as allowing us to determine the base at each of the polymorphisms.

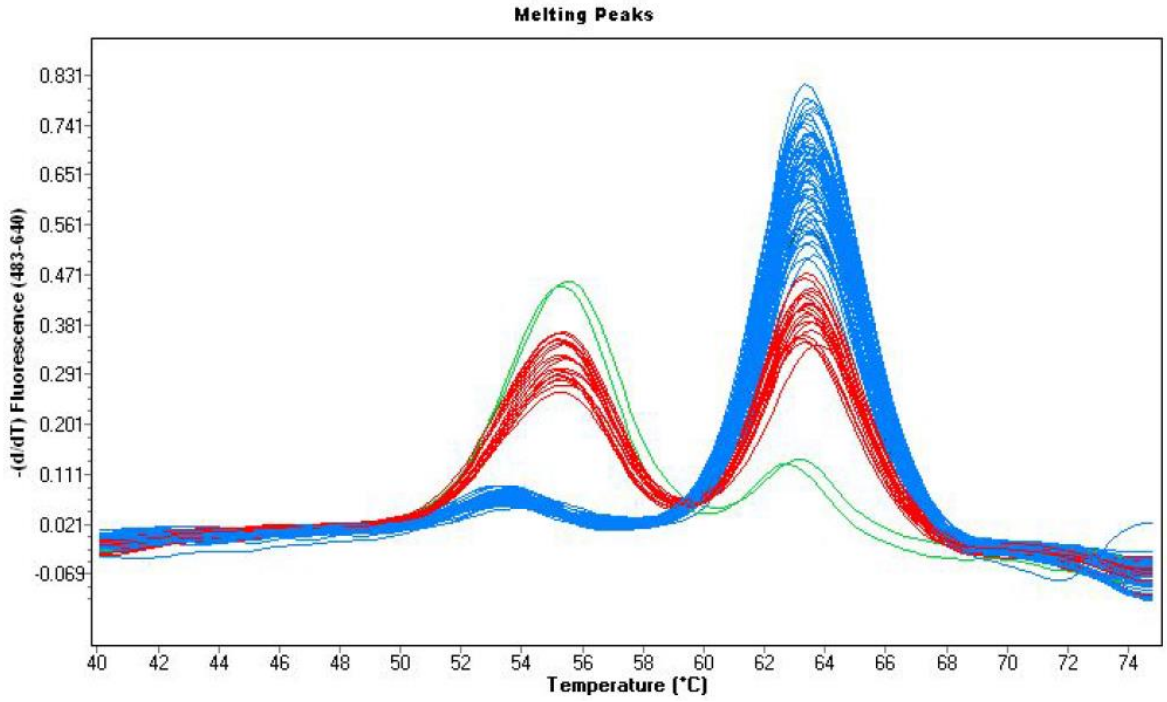


Figure 5.8. Melt curve analysis of the *DHFRL1* rs17855824 Val->Ile polymorphism. The single green peak at 55.5°C represents the homozygous polymorphism A/A genotype. The single blue peak at 63.5°C represents the homozygous wild type G/G genotype. The two red peaks at 55.5°C and 63.5°C represent the heterozygous G/A genotype.

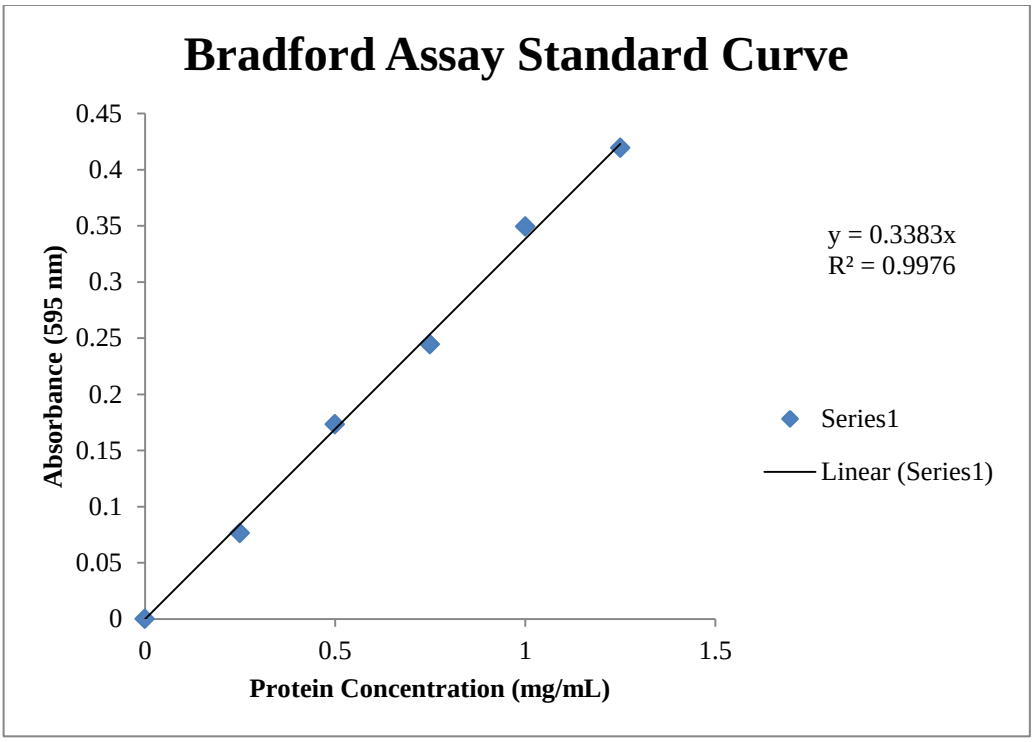


Figure 5.9. Bovine Serum Albumin (BSA) Standard Curve. The BSA Standard Curve was obtained by plotting the protein standards against the absorbance measured at 595 nm. Mitochondrial and cytoplasmic protein concentrations were obtained from the standard curve using the equation of the line. An R^2 value of over 0.98 indicated good linearity of the assay and precision of the standard preparation. .

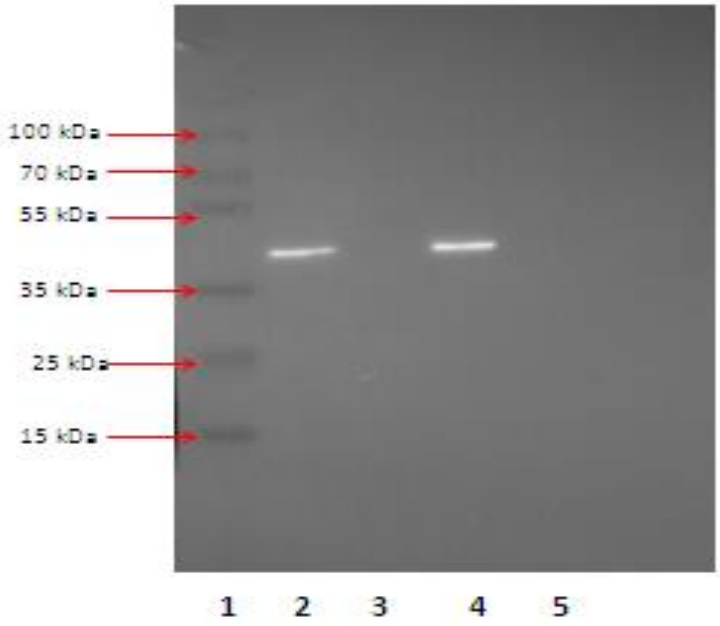


Figure 5.10. Western blot of HEK 293 and DHFRL1 over-expressed cytoplasmic and mitochondrial protein fractions probed with the GAPDH antibody. Lane 1 – protein ladder, Lane 2 –HEK 293 cytoplasmic protein fraction, Lane 3 – HEK 293 mitochondrial protein fraction, Lane 4 – DHFRL1 over-expressed HEK 293 cytoplasmic protein fraction, Lane 5 – DHFRL1 over-expressed HEK 293 mitochondrial protein fraction. All samples were equally loaded as determined by Bradford assay (Table 5.3). Bands of the correct size (40 kDa) were seen for all of the cytoplasmic protein samples.

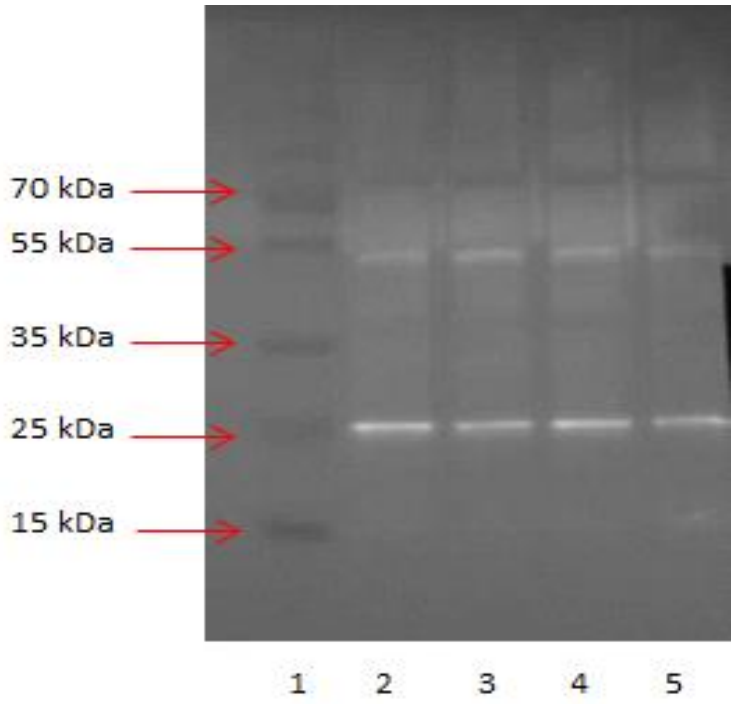


Figure 5.11. Western blot of HEK 293 and DHFRL1 over-expressed HEK 293 cytoplasmic protein fractions probed with the DHFR antibody. Lane 1 – protein ladder, Lane 2 –HEK 293 cytoplasmic protein fraction, Lane 3 – DHFRL1 over-expressed HEK 293 cytoplasmic protein fraction, Lane 4 - HEK 293 cytoplasmic protein fraction, Lane 5 – DHFRL1 over-expressed HEK 293 cytoplasmic protein fraction. All samples were equally loaded. Bands of the correct size (21 kDa) were seen for all of the cytoplasmic protein samples.

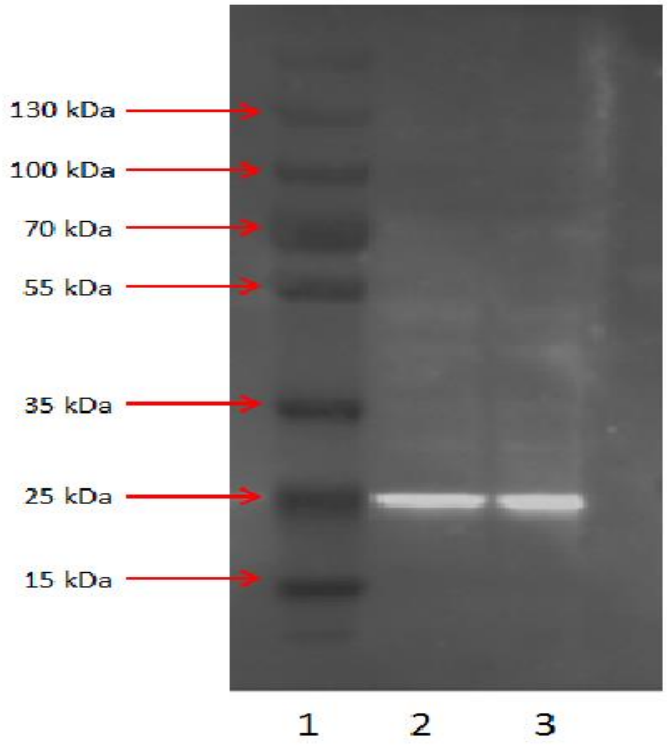
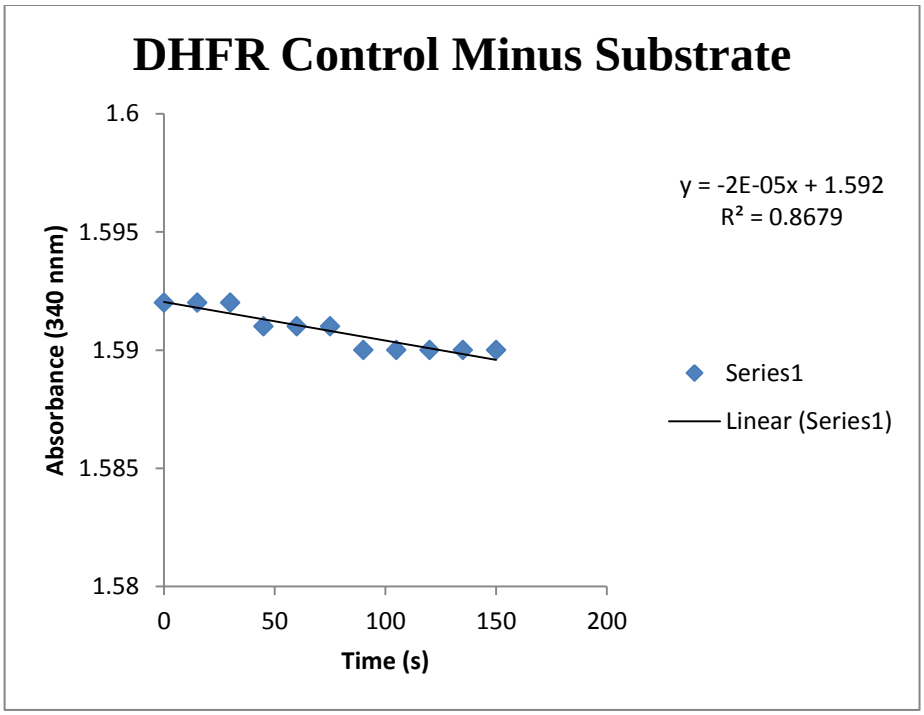
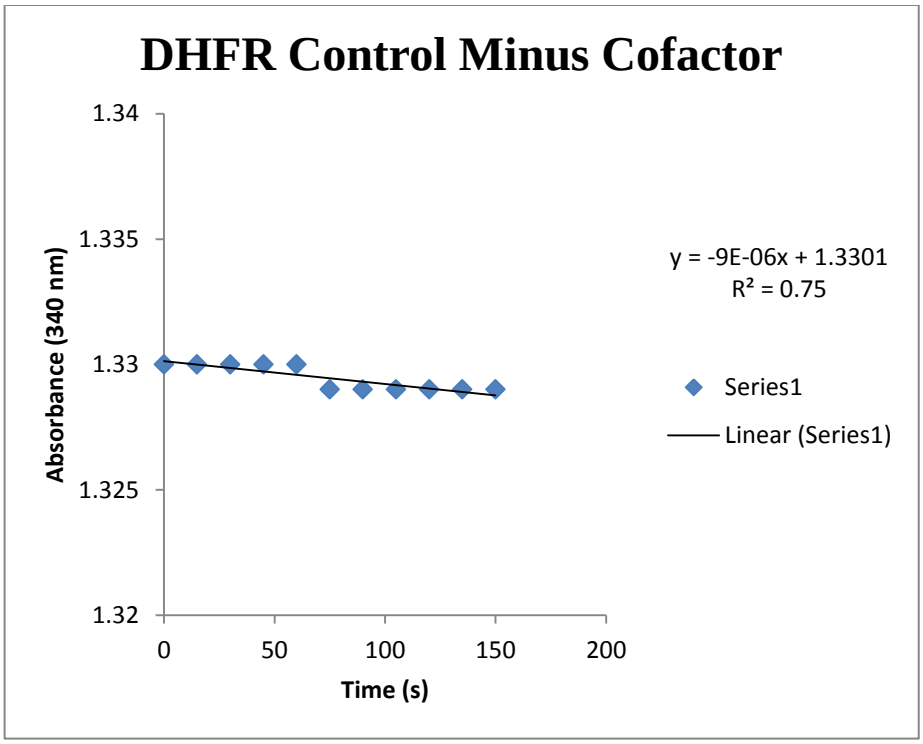


Figure 5.12. Western blot of HEK 293 and DHFRL1 over-expressed HEK 293 mitochondrial protein fractions probed with the DHFR antibody. Lane 1 – protein ladder, Lane 2 – HEK 293 mitochondrial protein fraction, Lane 3 – DHFRL1 over-expressed HEK 293 mitochondrial protein fraction. All samples were equally loaded. Bands of the correct size (21 kDa) were seen for all of the mitochondrial protein samples.



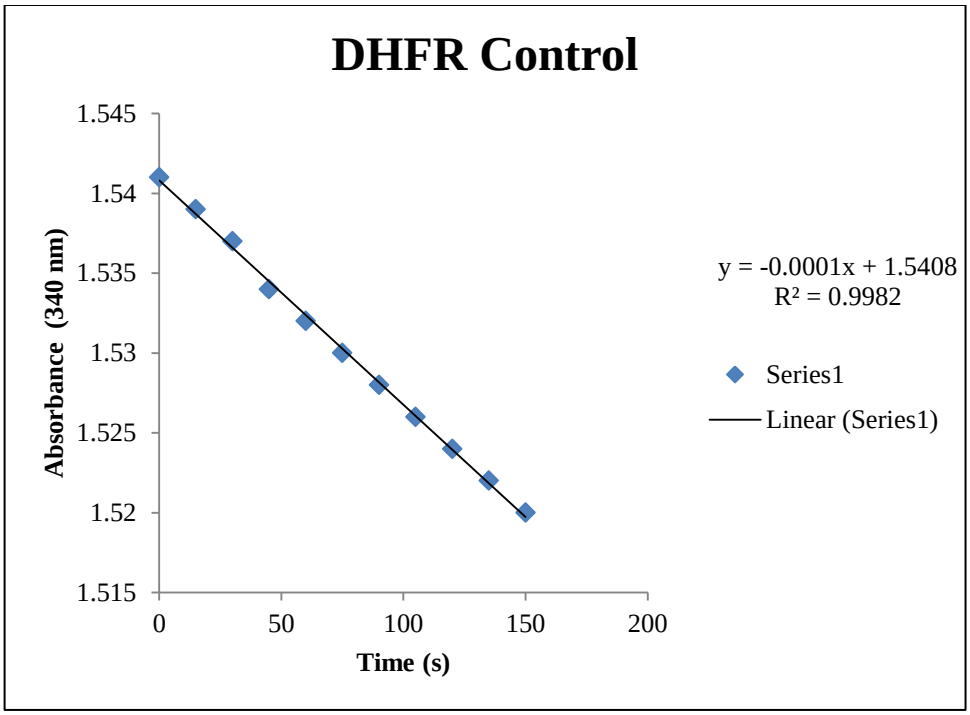
ΔOD/Min = 0.0008

Figure 5.13. DHFR Control without DHF (substrate). The graph above was obtained by plotting the 340 nm absorbance measurement vs. time in s. From the graph, it can be confirmed that there is no DHFR activity, indicated by the roughly straight horizontal line, i.e. when one of the key components of the reaction, DHF, which is the substrate for DHFR, is absent, the reaction cannot take place.



ΔOD/Min = 0.0004

Figure 5.14. DHFR control without NADPH (cofactor). The graph above was obtained by plotting the 340 nm absorbance measurement vs. time in s. From the graph, it can be confirmed that there is no DHFR activity, indicated by the straight horizontal line, i.e. when one of the key components of the reaction, NADPH, which is a cofactor for DHFR, is absent, the reaction cannot take place.



ΔOD/Min = 0.008

Figure 5.15. DHFR Control. The graph above was obtained by plotting the 340 nm absorbance measurement vs. time in s. From the graph, it can be confirmed that once all the necessary components that are involved in the DHFR pathway are present, there is DHFR activity, indicated by the decrease in absorbance over time. The decrease in ΔOD over the 2.5 min was measured as ΔOD/min. An R^2 value of over 0.98 indicated good linearity of the assay and precision of the standard preparation.

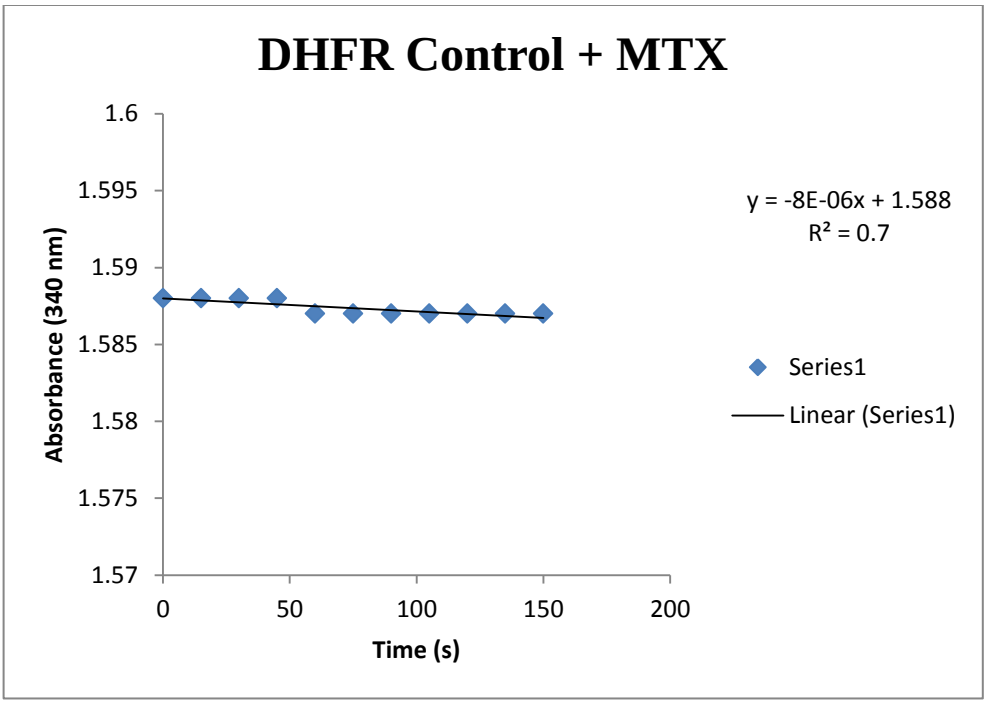
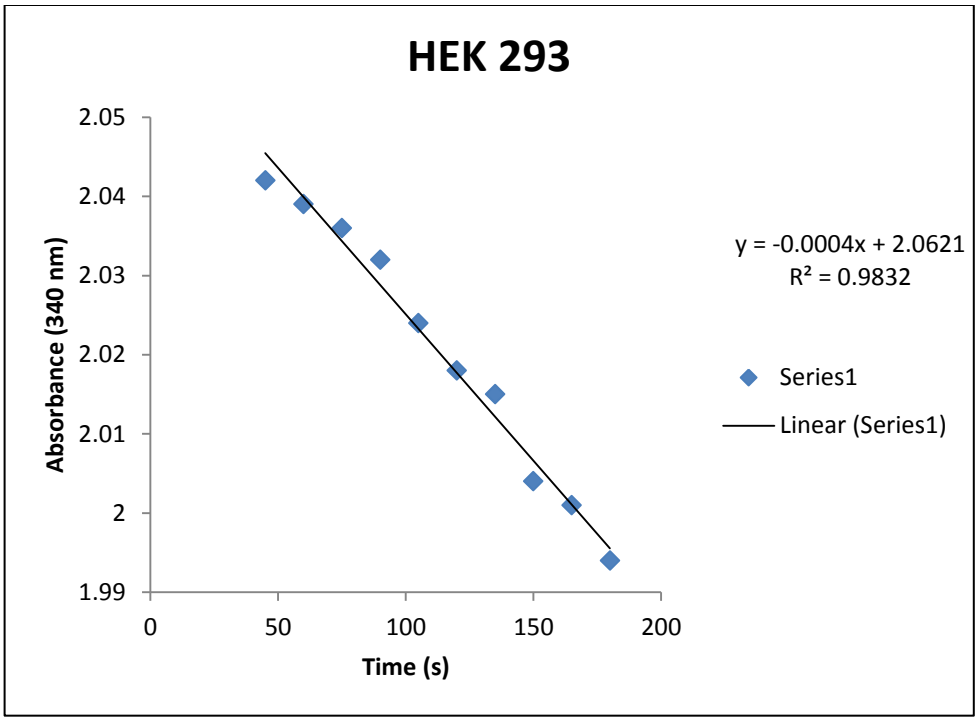
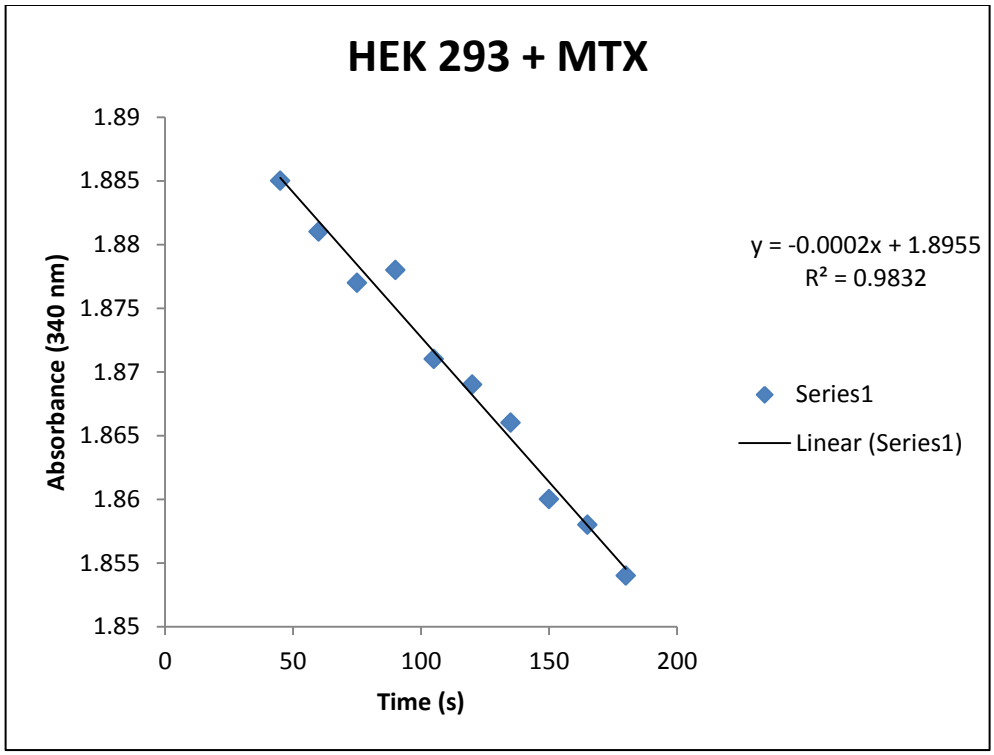


Figure 5.16. DHFR control plus methotrexate. The graph above was obtained by plotting the 340 nm absorbance measurement vs. time in s. From the graph, it can be confirmed that methotrexate is inhibiting DHFR activity and that the concentration of methotrexate used is high enough to inhibit all DHFR enzyme activity, indicated by the straight horizontal line.



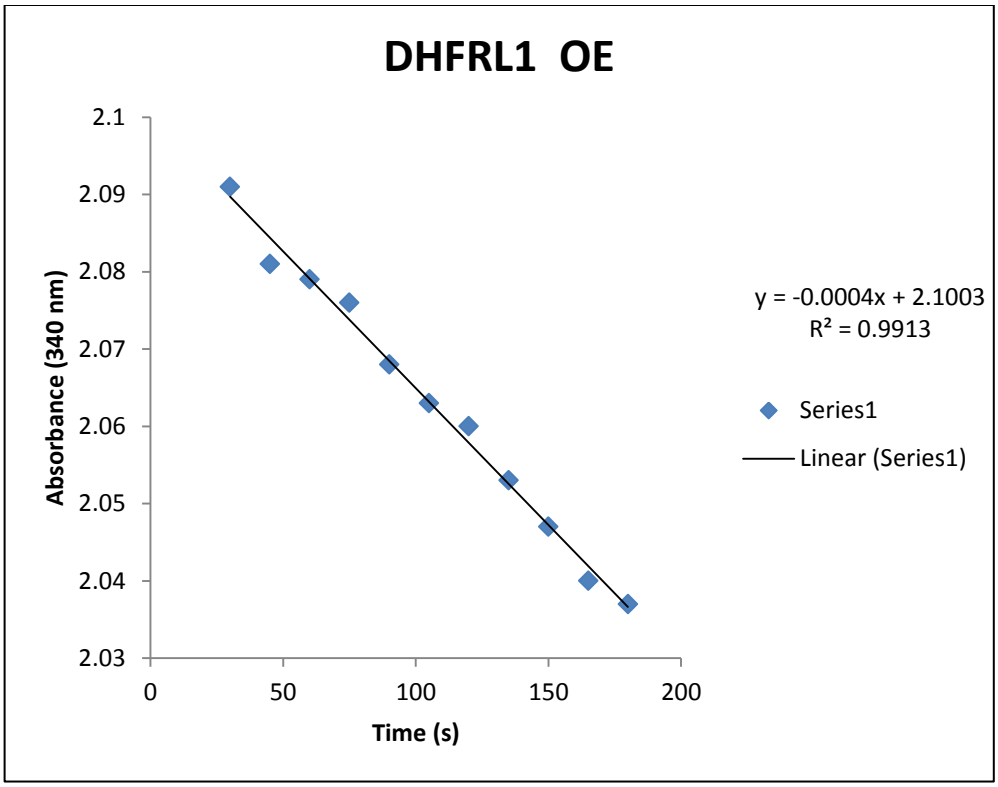
ΔOD/Min = 0.021

Figure 5.17. HEK 293 Mitochondrial Fraction. The graph above was obtained by plotting the 340 nm absorbance measurement vs. time in s. From the graph, it can be confirmed that there is enzyme activity occurring in the mitochondria, indicated by the decrease in absorbance over time. The decrease in ΔOD over the 2.5 min was measured as ΔOD/min. An R^2 value of over 0.98 indicated good linearity of the assay and precision of the standard preparation.



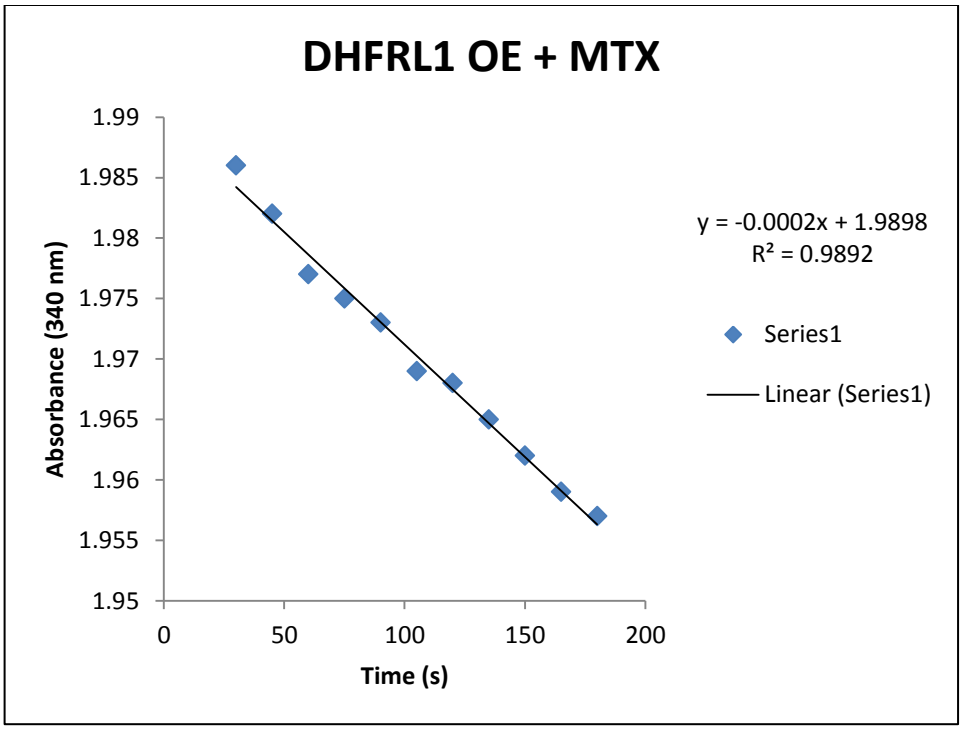
ΔOD/Min = 0.012

Figure 5.18. HEK 293 Mitochondrial Fraction plus Methotrexate. The graph above was obtained by plotting the 340 nm absorbance measurement vs. time in s. The graph above demonstrates the background activity present in the mitochondrial samples, as the enzyme activity caused by DHFR / DHFRL1 should be inhibited by the methotrexate. An R^2 value of over 0.98 indicated good linearity of the assay and precision of the standard preparation.



ΔOD/Min = 0.0216

Figure 5.19. DHFRL1 Over-Expressed HEK 293 Mitochondrial Fraction. The graph above was obtained by plotting the 340 nm absorbance measurement vs. time in s. From the graph, it can be confirmed that there is enzyme activity occurring in the mitochondria, indicated by the decrease in absorbance over time. The decrease in ΔOD over the 2.5 min was measured as ΔOD/min. An R^2 value of over 0.98 indicated good linearity of the assay and precision of the standard preparation.



ΔOD/Min = 0.0116

Figure 5.20. DHFRL1 Over-Expressed HEK 293 Mitochondrial Fraction plus Methotrexate. The graph above was obtained by plotting the 340 nm absorbance measurement vs. time in s. The graph above demonstrates the background activity present in the mitochondrial samples, as the enzyme activity caused by DHFR / DHFRL1 should be inhibited by the methotrexate. An R^2 value of over 0.98 indicated good linearity of the assay and precision of the standard preparation.

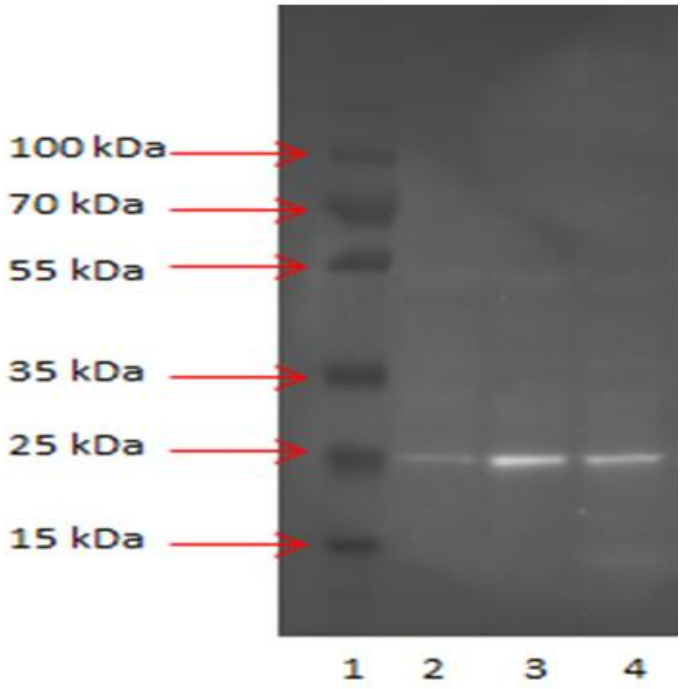


Figure 5.21. Western blot of Coriell mitochondrial protein fractions probed with the DHFR antibody. Lane 1 – protein ladder, Lane 2 – Coriell 17205 mitochondrial protein fraction (G/G genotype, wild type), Lane 3 – Coriell 17189 mitochondrial protein fraction (G/A genotype, heterozygous), Lane 4 - Coriell 17234 mitochondrial protein fraction (A/A genotype, variant). Bands of the correct size (21 kDa) were seen for all of the mitochondrial protein samples.

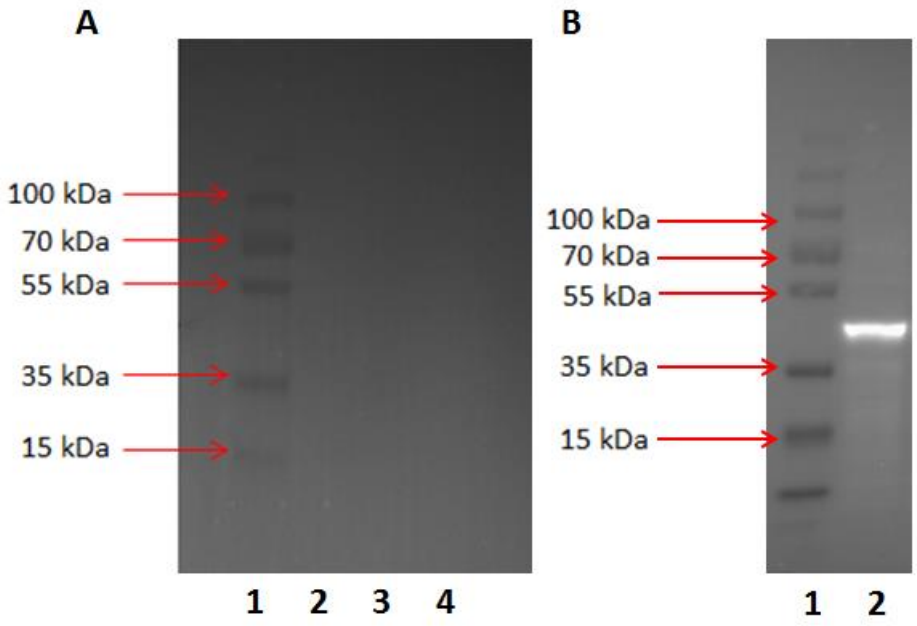


Figure 5.22. Western blot of Coriell mitochondrial protein fractions probed with the GAPDH antibody. (A). Lane 1 – protein ladder, Lane 2 – Coriell 17205 mitochondrial protein fraction (G/G genotype, wild type), Lane 3 – Coriell 17189 mitochondrial protein fraction (G/A genotype, heterozygous), Lane 4 - Coriell 17234 mitochondrial protein fraction (A/A genotype, variant). **(B).).** Lane 1 – protein ladder, Lane 2 – Coriell 17205 cytoplasmic protein fraction (control)

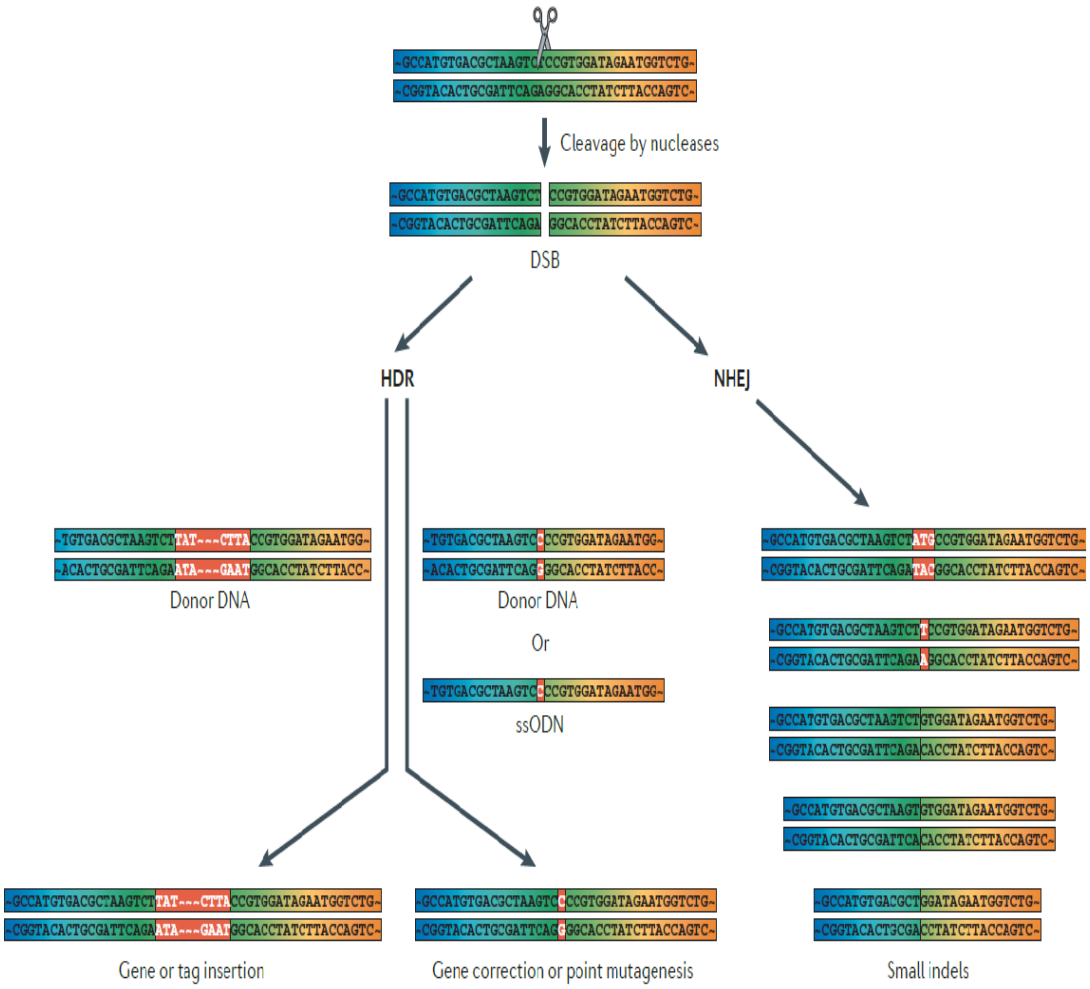


Figure 5.23. Genome Editing using Programmable Nucleases. Nuclease-induced DSBs can lead to sequence insertions, nucleotide correction or change (red box) through HR in the presence of a donor DNA or a single-stranded oligodeoxynucleotide (ssODN), both of which contain homology arms. DSBs can also be repaired through error-prone NHEJ, which does not require donor DNA or ssODN and consequently often leads to small insertions and deletions (indels). Image and legend taken from (Kim and Kim, 2014).

Table 5.1. Genotyping results from sequence and chromatogram analyses for each of the three *DHFRL1* polymorphisms

Samples Sequenced	G->A (rs17855824)	C->G (rs61739170)	G->T (rs114936057)
Coriell 17229	G/A	C/C	G/G
Coriell 17214	G/G	C/C	G/G
Coriell 17147	G/G	C/C	G/G
Coriell 17146	G/G	C/G	G/G
Coriell 17234	G/G	C/G	G/G
Coriell 17265	G/G	C/C	G/G
Coriell 17133	G/G	C/C	G/G
Coriell 17219	G/G	C/C	G/G
Coriell 17102	G/G	C/G	G/G
Coriell 17259	G/G	C/G	G/G
Coriell 17206	G/A	C/G	G/G
SW480	G/G	C/C	G/G
Coriell 17145	G/G	C/C	G/G
Coriell 17158	G/G	C/C	G/G
Coriell 17165	G/G	C/C	G/G
Coriell 17291	G/G	C/C	G/G
Coriell 17218	G/A	C/C	G/G
BT474	G/G	C/G	G/G
NCI-H1299	G/G	C/C	G/G
HT29	G/G	C/C	G/G

Table 5.2. Genotypes of the *DHFRL1* rs17855824 polymorphism in a panel of Coriell DNA Samples

rs17855824 Genotype	Total [n (%)]
G/G (wild type) G/A (heterozygous) A/A (polymorphism)	144 (85.2) 23 (13.6) 2 (1.18)
Allele 1 (WT) Allele 2 (SNP)	311 (92) 27 (8)

Table 5.3. Protein Standard Concentration and Absorbance Values

Concentration (mg/mL)	Absorbance		Average Absorbance	Average – Blank
0	0.3071	0.3103	0.3087	0
0.25	0.3858	0.38415	0.384975	0.076275
0.5	0.47819999	0.4859	0.482049995	0.173349995
0.75	0.54729998	0.5592	0.55324999	0.24454999
1	0.63120002	0.6847	0.65795001	0.34925001
1.25	0.70209998	0.7544	0.72824999	0.41954999

Table 5.4. Mitochondrial and Cytoplasmic Protein Absorbance and Concentration Values

Sample	Absorbance		Average Absorbance	Average - Blank	Protein Concentration (mg/mL)
HEK 293 Cytoplasmic	0.7134	0.7731	0.74325	0.4369	1.291457286
HEK 293 Mitochondrial	0.71249	0.7188	0.715645	0.409795	1.211336092
DHFRL1 OE^{+} Cytoplasmic	0.7993	0.8311	0.8152	0.50885	1.504138339
DHFRL1 OE Mitochondrial	0.4923	0.5048	0.49855	0.1927	0.56961277
Lysis Buffer	0.305	0.3077	0.30635		
Storage Buffer	0.3068	0.3049	0.30585		

$^{+}$Over-expressed

Table 5.5. DHFRL1 polymorphism genotyping results in NTD triads (fathers, mothers and cases) and controls.

DHFRL1 rs7653521 SNP								
	Fathers		Mothers		Cases		Controls	
	n	%	N	%	n	%	N	%
CC	91	20.00	86	17.10	97	18.95	182	18.69
CT	200	43.96	234	46.52	248	48.44	468	48.05
TT	164	36.04	183	36.38	167	32.62	324	33.26
Total	455		503		512		974	

rs10454213								
	Fathers		Mothers		Cases		Controls	
	n	%	n	%	n	%	N	%
AA	339	74.83	380	74.80	404	77.99	732	75.78
GA	105	23.18	121	23.82	108	20.85	215	22.26
GG	9	1.99	7	1.38	6	1.16	19	1.97
Total	453		508		518		966	

rs7644176								
	Fathers		Mothers		Cases		Controls	
	n	%	n	%	n	%	N	%
GG	434	94.55	480	93.93	495	95.19	910	93.43
TG	24	5.23	31	6.07	25	4.81	64	6.57
TT	1	0.22	0	0.00	0	0.00	0	0.00
Total	459		511		520		974	

rs11927165								
	Fathers		Mothers		Cases		Controls	
	N	%	n	%	n	%	N	%
CC	2	0.43	0	0.00	0	0.00	0	0.00
CT	23	4.99	31	6.03	26	4.97	67	6.88
TT	436	94.58	483	93.97	497	95.03	909	93.33
Total	461		514		523		976	

rs61739170								
	Fathers		Mothers		Cases		Controls	
	N	%	n	%	n	%	N	%
CC	24	5.38	26	5.17	25	4.96	57	5.91
CG	149	33.41	179	35.59	179	35.52	344	35.65
GG	273	61.21	298	59.24	300	59.52	564	58.45
Total	446		503		504		965	

rs17855824								
	Fathers		Mothers		Cases		Controls	
	N	%	n	%	n	%	N	%
CC	339	75.67	374	74.35	395	79.16	724	76.29
CT	101	22.54	123	24.45	99	19.84	211	22.23
TT	8	1.79	6	1.19	5	1.00	14	1.48
Total	448		503		499		949	

rs7645522								
	Fathers		Mothers		Cases		Controls	
	N	%	n	%	n	%	N	%
AA	337	74.56	377	73.92	403	77.95	736	75.64
GA	106	23.45	126	24.71	109	21.08	222	22.82
GG	9	1.99	7	1.37	5	0.97	15	1.54
Total	452		510		517		973	

Table 5.6. Location of each of the *DHFRL1* polymorphisms and the linkage disequilibrium test results carried out on the *DHFRL1* polymorphisms.

Irish NTDs	CHR	Position_b36	Region	HapMap CEU
rs7645522.25	3	95257618	DHFRL1_10kb	r^2=1 with rs10454213
rs11927165.25	3	95258101	DHFRL1_5kb	r^2=1 with rs7644176
rs10454213.25	3	95259426	DHFRL1_5kb	r^2=1 with rs7645522
DHFRL1_start	3	95262482	DHFRL1_start	
rs17855824.25	3	95262550	DHFRL1	r^2=1
rs61739170.25	3	95262799	DHFR1	
DHFRL1_end	3	95263045	DHFRL1_end	
rs7644176.25	3	95266787	DHFRL1_5kb	r^2=1 with rs11927165
rs7653521.25	3	95267249	DHFRL1_10kb	Singleton

Table 5.7. TDT analysis for the *DHFRL1* polymorphisms.

Name	Over-transmitted	T:U	Chi square	P value
DHFRL1_rs7645522	A	107:71	7.281	**0.007**
DHFRL1_rs11927165	T	24:19	0.581	0.4458
DHFRL1_rs10454213	A	105:66	8.895	**0.0029**
DHFRL1_rs17855824	C	97:62	7.704	**0.0055**
DHFRL1_rs61739170	C	130:125	0.098	0.7542
DHFRL1_rs7644176	G	25:18	1.14	0.2858
DHFRL1_rs7653521	C	177:158	1.078	0.2992

Chapter 6

Investigation of the impact of *DHFRL1* polymorphisms on the activity of recombinant enzyme

6.1 Introduction

Non-synonymous polymorphisms have been identified in the coding region of the *DHFRL1* gene as previously mentioned (Section 5.1.2). In Chapter 5, we have shown that endogenous DHFRL1 is present in the mitochondria and that it harbours enzyme activity. The presence of polymorphisms within the *DHFRL1* coding region was also confirmed. This in itself is highly interesting as no common polymorphisms are known to exist within the coding region of its paralogous gene, *DHFR*. Although the enzyme activity of the DHFRL1 rs17855824 polymorphism could not be determined endogenously, due to low mitochondrial concentration yields, the two *DHFRL1* polymorphisms that we have identified could alter the activity of the enzyme, which could be of great importance and may contribute to health and disease. As the *DHFRL1* rs17855824 polymorphism, along with two other *DHFRL1* polymorphisms (rs7645522 and rs10454213) showed to be associated with NTDs in Chapter 5, it is of great interest to try and determine the function of these *DHFRL1* polymorphisms. Therefore, the aim of this chapter was to assess the enzyme activity of recombinant DHFRL1, which was generated by Gateway Cloning and its recombinant variants, which were generated by Site Directed Mutagenesis, in order to complement our findings found in endogenous mitochondrial protein as well as to try and determine the function of the DHFRL1 polymorphisms.

6.1.1 Gateway Cloning

Gateway cloning is a universal cloning technology that is based on the integration of the site-specific recombination properties of bacteriophage lambda into the *E. coli*. chromosome, relying on the switch between the lytic and lysogenic pathways (Landy, 1989, Ptashne, 1992). Lambda integration into the *E. coli* chromosome occurs *via* intermolecular DNA recombination that is mediated by a mixture of lambda and *E. coli* encoded recombination proteins. Recombination proteins involved differ depending on whether lambda utilises the lytic or lysogenic pathway. Lambda recombination occurs between specific attachment (*att*) sites on the interacting DNA molecules, i.e. the DNA segments flanking the recombination sites are switched, and after recombination, the *att* sites are hybrid sequences comprised of sequences donated by each parental vector (e.g. *attL* sites are comprised of sequences from *attB* and *attP* sites). The strand exchange occurs within a core region that is common to all *att* sites and the recombination can occur between DNAs of any topology, i.e. supercoiled, linear or relaxed. Lambda recombination is conservative (there is no net gain or loss of nucleotides) and requires no DNA synthesis (https://tools.lifetechnologies.com/content/sfs/manuals/gatewayman.pdf.)

In Gateway cloning, the components of the lambda recombination system have been modified in order to improve the specificity and efficiency of the system (Bushman et al., 1985). Mutations have been made to the core regions of the *att* sites to eliminate stop codons and to ensure specificity of the recombination reactions to maintain orientation and reading frame as well as to the short (5 bp) regions flanking the 15 bp core regions of the *attB* sites to minimise secondary structure formation in single stranded forms of *attB* plasmids. A 43 bp portion of the *attR* site has also been removed to make the *in vitro attL* x *attR* reaction irreversible and more efficient (Bushman et al., 1985). Site specific point mutations have also been made to some *att* sites in order to increase recombination efficiency.

Gateway cloning involves three main steps. 1) Designing *attB* PCR primers and amplifying up the sequence of interest, 2) BP recombination reaction with the *attB* PCR product and a donor vector to generate an entry clone and 3) LR recombination reaction with the entry clone and a Gateway destination vector to generate an expression clone. These reactions are carried out by proteins that mediate the recombination reaction (Clonase^{TM} enzyme mix, which contain a mixture of lambda and *E. coli* encoded recombination proteins) (Figure 6.1).

6.1.2 Site Directed Mutagenesis (SDM)

SDM is a molecular technique that is used to introduce a specific mutation at a specific site in the DNA. This will allow the relationship between protein structure and function to be determined. Site-directed mutagenesis functions by utilising a supercoiled double-stranded DNA vector with the insert of interest and two synthetic primers, both of which contain the desired mutation. The primers are complementary to opposite strands of the vector and are extended by *Pfu*Ultra high fidelity DNA polymerase during elongation, generating a mutated plasmid containing staggered nicks. Once the mutated plasmid is produced, it is digested with *Dpn* I, an endonuclease that is specific for methylated and hemimethylated DNA, and is used to digest the parental DNA and select for mutation-containing synthesized DNA. (DNA isolated from almost all *E. coli* strains is dam methylated and therefore susceptible to *Dpn* I digestion) (Figure 6.2). Once the vectors containing the mutations are produced, they are transformed into XL10 Gold Ultracompetent cells (Section 2.3.1).

6.1.3 Objectives

Assessment of enzyme activity in *DHFRL1* and its polymorphisms was examined using the following methodology.

- Prediction of DHFRL1 and variant DHFRL1 protein structures
- Generation of DHFRL1 variant GST tagged proteins via site-directed mutagenesis.
- Production and optimisation of DHFRL1 and DHFRL1 variant GST tagged proteins.
- Generation of GST and His tagged DHFR, DHFRL1 and DHFRL1 variant expression clones by Gateway Cloning and site-directed mutagenesis.
- Production and optimisation of DHFR, DHFRL1 and DHFRL1 variant GST and His tagged proteins.
- Rare codon optimisation of DHFR and DHFRL1.
- Generation of rare codon optimised GST tagged DHFR and DHFRL1 expression clones by Gateway Cloning.
- Production and optimisation of rare codon optimised GST tagged DHFR and DHFRL1 proteins.
- Purification of DHFRL1 proteins.
- Production and optimisation of recombinant DHFRL1 variant proteins.
- Purification of DHFRL1 variant proteins.
- Enzyme kinetic analysis of purified *DHFRL1* and its polymorphisms.

6.2 Results

6.2.1 Protein Structure Prediction

The predicted tertiary structures of DHFRL1, DHFRL1 rs17855824 and DHFRL1 rs61739170 polymorphisms were generated using the Protein Structure Prediction Server software from the Molecular Bioinformatics Centre, National Chiao Tung University (Chen et al., 2006) and the STRAP-NT software (Figure 6.3), which predicts the structure of the protein from an amino acid sequence. By comparing the structures of DHFRL1, DHFRL1 rs17855824 and DHFRL1 rs61739170 to each other, we can see that all three proteins are highly similar, with all major α helices and β sheets present but there are slight differences in the folding of their loops and coils.

6.2.2 Generation of recombinant DHFRL1 GST tagged variant clones

A DHFR and DHFRL1 recombinant protein had previously been generated in the lab by recombining a PCR product that had been amplified from HEK 293 cDNA into a pDONR™ 211 vector, to form an entry clone. Plasmid DNA from the entry clone was then cloned into the Gateway Cloning vector pDEST 15 (Invitrogen), which contains a glutathione S-transferase (GST) tag (26kDa) at its N terminus. Clones were verified by Sanger sequencing. The pDEST 15 vectors

uses the pET system to induce high levels of recombinant protein expression in *Escherichia coli* (*E. coli*) by taking advantage of the high activity and specificity of the bacteriophage T7 RNA polymerase to control expression (Studier et al., 1990). The *DHFR* and *DHFRL1* ORF was inserted into the vector at position 799 (Figure 6.4), with the region between position 799 and 2482 of the original vector sequence being removed to allow entry. As the *DHFRL1* ORF is only 564 bases long, it altered the overall size of pDEST 15 from 7013bp to 5893bp. A tobacco etch virus (TEV) recognition site was also included upstream of the DHFRL1 protein but downstream of the GST tag at this stage. The vector construct was then transformed into DH5α *E-coli* cells, as described in Section 2.2.8 which can be used to maintain a library of the clone.

DHFR and *DHFRL1* clones were streaked out onto LB agar plates with carbenicillin and plasmid DNA was extracted from the *DHFR* and *DHFRL1* overnight cultures (Section 2.2.6) and were run on a 1 % w/v agarose gel to ensure that DNA was present (Figure 6.5). Once the presence of DNA was confirmed, SDM was carried out (Section 2.2.9). Primers were designed according to the QuikChange II XL SDM Kit (Section 2.2.9) to incorporate the rs17855824 and rs61739170 polymorphisms respectively and SDM was carried out. Once the vector containing the mutations was produced, they were transformed into XL10 Gold Ultracompetent cells, cells which can be used to maintain a library of the clone (Section 2.3.1) and plasmid DNA was extracted (Section 2.2.6). To confirm the presence of DNA and to verify that the correct mutation had been inserted, the purified plasmid DNA was run on a 1 % w/v agarose gel (Figure 6.6.) and Sanger sequenced (Figure 6.7) (See Appendix O for complete mutant clone chromatograms).

6.2.3 Production of recombinant DHFRL1 and DHFRL1 polymorphism proteins

DHFRL1 WT, *DHFRL1* rs17855824 and *DHFRL1* rs61739170 recombinant DNA samples were used to transform BL21-AI *E-coli* cells. This cell line has been modified to ensure that the protein of interest is expressed optimally by exploiting the use of an inducer system. Transformed BL21-AI cells were induced using L-arabinose to give a final concentration of 0.04 % as described in Section 2.3.4. T7 expression is controlled by the araBAD promoter, which in turn is controlled by araC, a transcriptional regulator which can form a complex with L-arabinose. When L-arabinose binds to araC, it activates the T7 RNA polymerase, which activates the araBAD genes, releasing the protein. Therefore, L-arabinose acts as an inducer of recombinant DHFRL1 in this system. An overnight culture was sub-cultured and grown until an OD_{600} of 0.4 was reached. The culture was then induced with L-arabinose and the protein was harvested at specific time points (0 h, 2 h, 4 h, 6 h and 24 h) and fractionated into soluble and insoluble fractions by multiple freeze thaw cycles.

SDS-PAGE analysis was carried out to make sure that the protein of interest was expressed, to determine at which time point the protein was best induced at and to ensure that the protein of interest was present in the soluble fraction. The 24 h time point showed the best results, with a strong band seen in the induced in comparison to the un-induced sample (Figure 6.8).

Results shown in Figure 6.8 illustrate a strong band at ~ 48 kDa in the induced 24 hr time point in the DHFRL1 and polymorphism protein samples and show that the protein of interest is both strongly expressed and present in the soluble fraction allowing protein purification to take place. A strong band is also seen at ~ 33 kDa in all of the induced proteins, indicating that other proteins may also be induced by this inducer system or it may be a truncated version of the fusion protein. Once it was confirmed that the protein of interest was present in the soluble fraction, the GST tagged protein was purified by binding the GST tag to glutathione agarose. GST-tagged fusion proteins can be purified by binding its substrate glutathione. When reduced glutathione (GSH) is immobilized by binding to a solid support, such as glutathione agarose beads, it can be used to capture GST-tagged proteins *via* the enzyme-substrate binding reaction. The purified protein was eluted using 35mM glutathione, which competitively displaces the GST to bind to the glutathione agarose, allowing the fusion protein to emerge from the affinity column. The protein purification fractions were analysed by 10 % w/v SDS-PAGE. Results are shown in Figure 6.9. Bands of the correct size in the elution fraction were seen at 48 kDa, indicating that GST-tagged fusion protein is present. Results shown in Figure 6.9 illustrate that the protein of interest at ~ 48 kDa is being purified out. However, the ~ 33 kDa band is also being co-purified out. This was of concern as the band at ~ 33 kDa is stronger than that of our protein of interest.

6.2.4 Optimisation of recombinant DHFRL1 & DHFRL1 polymorphism proteins: Elimination and identification of the 33 kDa band

Although the protein of interest at 48 kDa was being induced and could be purified out, a strong 33 kDa band was also seen to be induced at the same time and on repetition, the 33kDa was stronger than that of the band of interest at 48 kDa (Figure 6.10).

In order to try and eliminate this band at ~ 33 kDa, the cultures were grown at different temperatures in order to try and optimise the growth of the cultures so that the protein of interest would be produced most optimally. Cultures were grown at 25°C, 30°C and 37°C for 24 h. The protein fractions were fractionated as before and were analysed by 10 % w/v SDS-PAGE (Figure 6.11). Strong bands were still observed at ~33 kDa

6.2.5 Generation of GST and His tagged recombinant DHFR, DHFRL1 and DHFRL1 variant clones.

The strong band seen at ~ 33 kDa was initially thought to be a truncated version of the protein of interest and therefore, in order to only purify out the full length protein, a His tag was added to the C-terminus of the protein. This was carried out by designing primers that would incorporate a TEV cleavage site at the N-terminus and a His tag (6 x histidine amino acids) at the C-terminus as well as att-sites for both DHFR and DHFRL1 (Section 2.2.8) (Figure 6.12) and by PCR amplification with these primers using HEK 293 cDNA as a template (Figure 6.13). Once the PCR product was obtained, recombinant DHFR and DHFRL1 protein was generated by Gateway cloning (Section 2.2.8). Generated clones were plasmid prepped and verified by Sanger sequencing (Figure 6.14) (See Appendix P-W for sequences and chromatograms).

Once the His tag had been incorporated, SDM was carried out on the DHFRL1 plasmid DNA as before in order to generate the recombinant DHFRL1 polymorphisms (See Section 2.2.9). Generated clones were plasmid prepped and verified by Sanger sequencing (Figure 6.15) (See Appendix X-AA for sequences and chromatograms).

6.2.6 Production and optimisation of DHFR and DHFRL1 GST and His tagged proteins

Once the correct clones had been obtained, the proteins were induced as before and the protein was purified using a His column (Section 2.3.14). The protein samples were then run on 10 % w/v SDS-PAGE gels (Figure 6.16). However, two bands, one at 48 kDa, the size of the protein of interest and one at 33kDa or just one band at the lower size was observed, depending on the protein sample.

Therefore, in order to try and eliminate the lower strong band at ~ 33 kDa, L-Arabinose concentrations were re-optimised, time points were altered, the temperature at which the culture were grown were altered. The cell strain in which the cells were transformed into was also changed from BL21-AI cells to BL21 (DE3) cells. The BL21 (DE3) *E. coli* cells contain the DE3 lysogen that carries the gene for the T7 RNA polymerase, which unlike the BL21-AI cells is under the control of the lacUV5 promoter, and therefore IPTG, rather than L-Arabinose is required to induce the expression of the T7 RNA polymerase. However, changing the growth conditions caused more problems and eventually induced bands could only be obtained in certain samples and no consistency could be observed on the SDS-PAGE gels (Figure 6.17).

6.2.7 Identification of 33 kDa band

In order to try and determine what was happening with our samples and in order to try and identify what the initial two strong bands were, a Western blot was carried out on those samples that had previously shown the two strong induction bands and those samples in which induction did not occur at all. As the samples had not been purified, a Bradford assay (See Section 2.3.17) was carried out on the samples in order to determine the concentration of the protein samples. A serial dilution of one of the protein samples was carried out and run on an SDS-PAGE gel in order to determine the best concentration to use for Western blot analysis. The gel was run so that the two strong bands would be separated as much as possible, i.e. the lower bands of the ladder were run off the gel (Figure 6.18).

Figure 6.18 shows that when the bands were separated out, the two strong induction bands seen previously were in actual fact made up of numerous bands rather than a single strong band. From Figure 6.18, we can also determine that the best concentration of protein to use for Western blot analysis is ~ 2.82 µg of total protein. Once the best concentration was determined, the GST and DHFR primary and secondary antibodies were optimised by varying the concentration of the antibodies (Figure 6.19).

Once the Western blots had been optimised, a Western blot was carried out on those samples that had previously shown the two induction bands and those samples in which induction did not occur at all. All lanes were equally loaded for protein (Figure 6.20). Figure 6.20 shows no 33 kDa band in any of the samples indicating that the 33 kDa bands observed in the SDS-PAGE gels (Figure 6.16) are not our protein of interest nor is it a truncated version of our protein. The Figure also shows that our protein of interest is not only present in the induced fraction samples but also in the un-induced fraction samples. This is most likely due to a leaky promoter.

6.2.8 Rare codon optimisation of DHFR and DHFRL1

We ruled out that the 33kDa protein was a truncated version of our fusion protein. Our main difficulty was the inconsistency in the induction of our fusion protein from one experiment to the next. From our Western blots, (Figure 6.20), we can conclude that the clones were originally producing our protein of interest, but that the clones stopped producing DHFR and DHFRL1 for an unknown reason. One cause could be attributed to the fact that different colonies were picked from the LB agar plates each time the experiment was carried out, indicating that the production of the

recombinant protein could be colony specific, with some colonies producing the protein while others not, although this is not known for sure. Therefore, we decided to optimise our DHFR and DHFRL1 clones for rare codon usage in bacteria by generating another set of recombinant clones. The *DHFR* and *DHFRL1* genes were optimised for codon usage in bacteria using the IDT codon optimisation web tool (https://eu.idtdna.com/CodonOpt) and the GenScript rare codon analysis tool (http://www.genscript.com/cgi-bin/tools/rare_codon_analysis). Sequences optimised for rare codon usage were obtained from the IDT website and the optimised sequences were analysed for the best fit using the GenScript website. The best DHFR sequence optimised for rare codon usage obtained from the IDT website is shown in Figure 6.21, with the characteristics for the best fit shown in Figure 6.22–6.24. The best DHFRL1 sequence optimised for rare codon usage obtained from the IDT website is shown in Figure 6.25, with the characteristics for the best fit shown in Figure 6.26–6.28.

6.2.9 Generation of rare codon optimised GST tagged DHFR and DHFRL1 expression clones by Gateway Cloning

Once the *DHFR* and *DHFRL1* genes were optimised for rare codon usage in *E. coli*, the optimised gene was synthesised by IDT using their Gene Synthesis Technology. Primers were then designed to amplify up the optimised *DHFR* and *DHFRL1* gene. Once the PCR product was obtained (Figure 6.29), recombinant DHFR and DHFRL1 protein was generated by Gateway cloning (Section 2.2.8). Generated clones were plasmid prepped and verified by Sanger sequencing (See Appendix AB-AI for whole sequences and chromatograms).

6.2.10 Production and optimisation of rare codon usage optimised GST tagged DHFR and DHFRL1 proteins

Once the correct DHFR and DHFRL1 clones had been obtained, the plasmid DNA was transformed into BL21 (DE3) cells (Section 2.3.3) and were sub-cultured and grown as before. The cultures were then optimised for time (0 h, 0.5 h, 1 h, 1.5 h, 2 h, 2.5 h, 3 h, 3.5 h, 4 h, 6 h and 24 h), IPTG concentration (0 µM, 10 µM, 25µM, 50µM and 500 µM) and temperature (37°C, 30°C and 20°C). Samples were harvested for each optimisation step. Protein samples were then lysed with bacterial lysis buffer, put through a needle and syringe, and the samples were fractionated into soluble and insoluble fractions by multiple freeze thaw cycles using liquid nitrogen. Samples were then run on SDS-PAGE gels and Western blot analysis was carried out to make sure that the protein of interest was expressed, to determine at which time point the protein was best induced at and to ensure that the protein of interest was present in the soluble fraction. Proteins had to be confirmed

by Western blot analysis as although a band of the correct size was sometimes seen on the SDS-PAGE gel, there are numerous proteins of the same size as our protein of interest that are induced under our conditions (Figure 6.30). Although stronger bands were always observed in our insoluble fractions for all our proteins (Figure 6.31–6.34), the goal of our optimisation steps was to obtain the highest possible amount of our protein of interest in the soluble fraction. For the DHFR clone, the induced 1 h time point at 30°C showed the best results, with a strong band observed at 48 kDa (Figure 6.31). For the DHFRL1 clone, the un-induced 1 h time point at 30°C showed the best results (Figure 6.32), with the un-induced samples producing higher amounts of protein in the soluble fractions in comparison to the induced samples. This may be due to as mentioned earlier, (Section 6.2.6), a leaky promoter and / or the fact that as the new clones have been optimised for bacteria, they are producing the protein too fast when the protein is induced, resulting in the formation of protein aggregates (inclusion bodies), which are located in the insoluble protein fraction.

For the DHFRL1 polymorphisms clones, we went back to our original GST clones and re-optimised them using the BL21 (DE3) cells. For the DHFRL1 rs61739170 (C->G) clone, the un-induced 4 h time point at 30°C showed the best results, with a strong band seen (Figure 6.33) and for DHFRL1 rs17855824 (G->A) clone, the induced (final concentration 50 µM IPTG) 4 h time point at 30°C showed the best results (Figure 6.34).

6.2.11 Purification of DHFR and DHFRL1 proteins

Once the clones had been optimised for expression in a small scale culture (50-100 mL), large scale cultures (2 L) were then carried out to ensure that the expression of the protein did not change. Once this had been confirmed, the protein samples were spun down in 30 kDa centricons at 4°C for ~ 1 h at 4470 X g in order to concentrate the protein as well as to eliminate small proteins. The protein samples were then clarified in lysis buffer and put through a 0.45 µM filter. The GST tagged protein was then purified as described in Section 2.3.13. The eluted fractions were then pooled together and concentrated using the 30 kDa centricons by centrifugation at 4°C for ~ 30 min at 4470 X g. The protein purification fractions were run on a 10 % w/v SDS-PAGE and Western blot analysis was carried out to ensure our protein of interest had been purified (Figure 6.35–6.37). Bands of the correct size in the elution fraction were seen at 48 kDa, indicating that the GST-tagged fusion protein is present and purified.

6.2.12 Enzyme kinetic analysis of purified DHFRL1 and its polymorphisms

Once the presence of the protein of interest was confirmed by Western blot analysis, the eluted concentrated protein of interest was quantified. Protein concentrations could not be quantified by Bradford assay as when the SDS-PAGE gel corresponding to the Western blot was analysed *via* Coomassie blue, a strong glutathione band (~ 28 kDa) as well as our band of interest (48 kDa) as well as other bands were observed (Figure 6.38). Therefore, the concentration of our protein of interest was quantified using our Western blot images via Image J, a semi-quantitative method for determining protein concentration (Figure 6.39) (Table 6.1). The concentrations of our purified recombinant proteins were determined by loading a known concentration (176 ng) of the DHFR control onto each of the gels and normalising against the control. The DHFR control also allowed us to estimate the amount of protein required for the enzyme activity assays and to ensure that the same amount of protein was being used for each enzyme assay. This was important for calculating accurate specific activity values. Image J quantifications indicated that we had 4.4 ng of protein/μL for the DHFRL1 protein, 3.5 ng of protein/μL for the rs61739170 polymorphism protein and 5.55 ng of protein/μL for the rs17855824 polymorphism protein (Table 6.1).

Enzyme kinetic analyses were then carried out on the purified recombinant WT DHFRL1, DHFRL1 rs17855824 and DHFRL1 rs61739170 polymorphisms in order to determine the specific activity of each of the proteins. Specific activity measures the amount of product formed by the enzyme per min per milligram of total protein i.e. how efficient the enzyme is at breaking down the substrate. The higher the specific activity of the enzyme, the more efficient the enzyme is at breaking down the substrate.

In order to determine the enzyme activity of our purified protein, the Dihydrofolate Assay Kit, a colorimetric assay, was used as described in Section 2.3.21. The assay kit functions on the basis that the absorbance at 340 nm will decrease over time, due to the decrease in NADPH concentration (cofactor) as it is utilised by the enzyme, with DHF acting as the substrate for the reaction. The DHFR control and blanks were tested first in order to confirm that all the reagents were working correctly within the region of detection by the spectrophotometer (Figure 6.40-6.43).

Once it was confirmed that the reagents and the assay was working well, the assay was carried out on the purified protein fractions. An example of each is shown in Figure 6.44–6.49. It was important to carry out the enzyme assays with our purified protein with the methotrexate in order to ensure / eliminate any possible involvement of the GST tag, which may lead to

overestimation of the specific activity of the enzyme. The decrease in ΔOD / min was then calculated for each of the samples. In order to obtain the true ΔOD of the protein samples, the ΔOD / min of the sample + MTX (the background activity, which was practically zero in our case) was taken away from the ΔOD / min of the sample. The specific activity was then calculated as mentioned in Section 2.3.21. The specific activity measurements are based on the molar extinction coefficient of 12,300 $M^{-1}cm^{-1}$ at 340nm. As mentioned earlier, it is crucial that protein concentrations are estimated correctly in order to calculate the specific activity of the proteins. The specific activity measurements showed that both the DHFRL1 polymorphisms were more efficient at breaking down the substrate than the WT. The specific activity calculations showed the recombinant DHFRL1 protein to have a specific activity of 2.11 μnmol/min/mg protein, the DHFRL1 rs61739170 protein to have a specific activity of 7.10 μnmol/min/mg protein and the DHFRL1 rs17855824 protein to have a specific activity of 3.45 μnmol/min/mg. The specific enzyme activity fold difference was then calculated between the DHFRL1 wild type and the polymorphisms. The DHFRL1 rs61739170 protein showed to have a 3.36 fold increase and the DHFRL1 rs17855824 protein showed to have a 1.64 fold increase in specific enzyme activity when normalised to the DHFRL1wild type (Figure 6.50).

6.3 Discussion

The discovery of *DHFRL1* as a functional gene in 2011 (McEntee et al., 2011) has great implications and may offer us further insights not only into folate-mediated one carbon metabolism but may also have serious implications for disease risk as shown in Chapter 5. We generated recombinant DHFRL1 and DHFRL1 polymorphism proteins *via* Gateway cloning in combination with SDM and optimised them for protein expression. Although the DHFRL1 polymorphisms only differ from the DHFRL1 wild type protein by a single base / amino acid, optimisation conditions for protein expression seemed to have varied between all three proteins, indicating that the proteins may have different effects.

A lot of optimisation was required in order to produce enough recombinant DHFRL1 protein and variant DHFRL1 proteins in the soluble fraction of the cells to carry out the enzyme kinetic analysis. It appeared that if the bacterial cultures were grown at too high a temperature or if the cultures were grown for too long, our proteins of interest moved into the insoluble fractions of the cells. Localisation of the proteins from the soluble fraction into the insoluble fraction is most likely caused by the protein being produced too quickly, resulting in the aggregation of the proteins, which will then result in the likelihood of the protein being mis-folded. This would then most likely

affect our downstream experiments of both protein purification and enzyme kinetic analysis as protein mis-folding could result in the GST tag being mis-folded, preventing binding to the GST column as well as altering the activity of the protein. Even after optimisation, a lot of the recombinant protein appeared to remain in the insoluble fraction, however, enough remained in the soluble fraction to carry out an assessment of the specific activity but not enough to calculate the Km.

We examined the effect of two non-synonymous polymorphisms within the *DHFRL1* gene on enzyme activity. Although the two polymorphisms chosen are not located in the regions required for catalytic activity or are the residues known to be important for folate, NADPH and/or MTX binding (Figure 6.51), the results above demonstrate that they have an impact on enzyme activity. On the other hand, our predicted protein structures for DHFRL1 and its variants showed slight differences in the folding of their loops and coils, (Figure 6.3) suggesting that the polymorphisms may have a potential effect on the 3D structure of the protein, resulting in enzyme activity differences.

The DHFRL1 rs61739170 polymorphism, a non-synonymous polymorphism, which results in a proline to alanine amino acid change, showed that it can efficiently bind to the co-factor NADPH but also showed to have a much greater specific activity in comparison to the WT DHFRL1, indicating that it is much more efficient at reducing DHF. Proline and alanine are both linker molecules and can alter the stability of proteins as well as having an effect on protein fold rates (George and Heringa, 2002). However, there are many characteristic differences between the two amino acids. Prolines are the only amino acids whose side chain is connected to the protein backbone twice, and therefore does not allow for much flexibility. As a result, this restricts proline to occupy many of the main chain conformations that are easily adapted by all the other amino acids. This causes proline residues to have rigid structures, and therefore they often either act as scaffolds to prevent unfavourable interactions between folding domains or they function to introduce kinks into alpha helices, as it is unable to adopt a normal helical conformation. On the other hand, alanine is a much more flexible amino acid, allowing for movement within the helices (Solan et al., 2002). Therefore this proline to alanine amino acid change in the DHFRL1 rs61739170 polymorphism may be what is having an effect on the specific activity and the binding affinity of the protein.

The DHFRL1 rs17855824 polymorphism, a non-synonymous polymorphism, which results in a valine to isoleucine amino acid change also showed that it can efficiently bind to the co-factor

NADPH and also showed to have a higher specific activity in comparison to the WT DHFRL1, although to a far lesser extent to that of the DHFRL1 rs61739170 polymorphism. In contrast to the amino acids involved in the DHFRL1 rs61739170 polymorphism, valine and isoleucine are both aliphatic hydrophobic amino acids and have very similar characteristics. They are both C-beta branched and contain two non-hydrogen substituents attached to its C-beta carbon, making them bulky near the protein backbone. Although valine and isoleucine seem to have similar properties, differences in protein structure can been seen from the predicted structures in Figure 6.3, implying that it may have an effect on functional impact. In Chapter 5, an association between the DHFRL1 rs17855824 polymorphism and NTDs was observed. The association may be linked with the differences observed between the specific activity of the DHFRL1 WT and that of the DHFRL1 rs17855824 polymorphism but will need to be investigated further.

Although the exact function of these two polymorphisms is yet unclear and unknown, they seem to impact the activity of the enzyme and should be explored further. *DHFR* has been associated with many different diseases such as NTDs and cancer, and with *DHFRL1* been previously shown to have similar properties to *DHFR*, it may also have implications in disease risk. When the specific activity and affinity for the substrate were examined in DHFR and DHFRL1, DHFR was shown to have a much higher specific activity and binding affinity in comparison to DHFRL1 (McEntee et al., 2011). However, both the DHFRL1 rs61739170 C->G polymorphism (proline to alanine change) and the DHFRL1 rs17855824 G->A polymorphism (valine to isoleucine) seems to increase the specific activity and binding affinity of DHFRL1, suggesting that it may be a significant factor in terms of disease risk. As the rs61739170 G allele and the rs17855824 A allele have been found to be frequent within the European population (approximately 15 % and 7.5 % respectively), a large number of the population may be exhibiting increased activity levels of DHFRL1, which in turn may be increasing / decreasing their risk of disease by altering the folate-mediated one carbon metabolism pathway resulting in people who harbour this polymorphism to be at an increased / decreased risk of disease.

In the future, two other specific kinetic measurements could be examined Vmax and Km, in order to determine the enzymatic properties of the WT DHFRL1 protein and each of the DHFRL1 polymorphisms, which would hopefully allow us to understand better the function of the proteins. Vmax is the maximum rate of reaction that can be achieved by the enzyme and measures the amount of time it takes for the substrate to bind all the enzymes active sites. Km (Michaelis-Menten constant) measures the substrate concentration required to reach half the Vmax. The Km

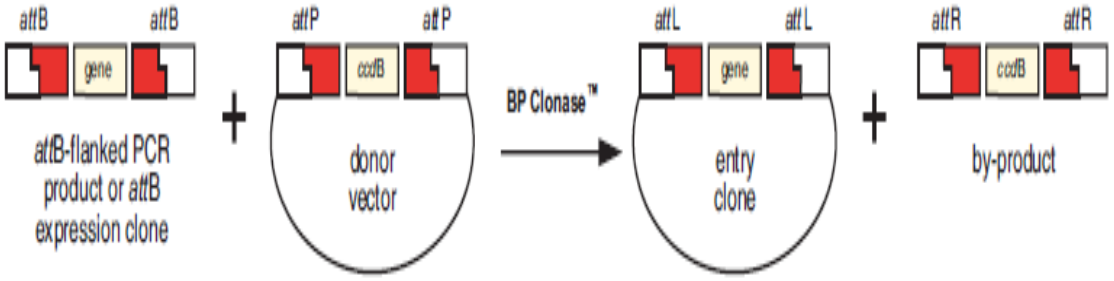
characterises the enzyme's affinity for the substrate. The higher the Km of the enzyme, the lower it's affinity for its substrate (Figure 6.52). The Km and Vmax values for the substrate DHF could be determined by keeping the cofactor, NADPH constant at 60 μM, a concentration that was suggested by the DHFR assay kit and varying the concentration of DHF. However, in order to carry out these enzyme kinetic analyses, one would require a much higher concentrations of total recombinant protein than what we were obtaining. At present, we are obtaining ~ 3.5 μg of total protein from a 2 L bacterial culture based on our Image J quantifications and in order to carry out the enzyme kinetics to obtain the Km and Vmax values; one would require at least three times the amount. In order to try and obtain higher protein concentration yields, one could try and solubilise the insoluble protein aggregates (inclusion bodies) and then re-fold the protein in order to obtain a functionally active protein. Protein solubilisation could be carried out by using denaturants such as urea or guanidine hydrochloride at high concentrations, in combination with reducing agents such as β-mercaptoethanol. Once the protein is solubilised, it can be re-folded by the slow removal of the denaturing agent in the presence of an oxidising agent, which removes the secondary structures, resulting in the formation of an active protein (Sirawaraporn et al., 1993, Singh and Panda, 2005).

In order to produce higher yields of purified recombinant protein, one could also try expressing the protein of interest in a mammalian expression system, such as Chinese hamster ovary (CHO) cells, which have been adapted into growing in suspension. A target protein sequence encoded with a signalling sequence for secretion could be transfected into the CHO cells and grown until a desired cell density is obtained. Extracellular production of the protein of interest would allow the recovery of the protein from the supernatant during the collection process. This method has been used routinely in biopharmaceutical companies for the last few decades in order to obtain high yields of protein with low cellular protein contamination (Lory, 1992, Wurm, 2004).

As previously mentioned, Bailey *et al*. found that there was a high variability of *DHFR* activity among individuals, however the cause of this variability is yet unknown. Polymorphisms within the non-coding region of *DHFR*, have been thought to be related to the variability of *DHFR* activity (Bailey and Ayling, 2009) although no definitive cause has yet been discovered. In our study, we found that both the *DHFRL1* polymorphisms that we examined had an impact of enzyme activity. This suggests that if variability of *DHFRL1* activity is also found between different individuals, that these polymorphisms may be the cause of the differences observed.

DHFRL1 has also been shown to localise to the mitochondria, suggesting that polymorphisms in the *DHFRL1* gene may have an effect on mitochondrial diseases, such as Rett Syndrome, Sudden Infant Death syndrome, Leigh Syndrome and Diabetes Mellitus and Deafness. If these two polymorphisms prove to be functional, they could have a huge impact on disease especially as they are located within the coding region of the *DHFRL1* gene and no polymorphisms have been found in the coding region of the *DHFR* gene to date.

A



B

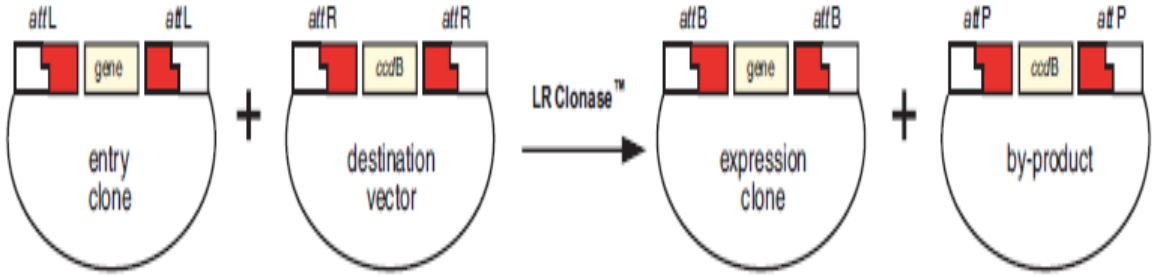


Figure 6.1. Gateway Recombination Reactions. A. The BP Reaction. The BP reaction facilitates the recombination of an *att*B substrate (*att*B-PCR product) with an *att*P substrate (donor vector) to create an *att*L-containing entry clone. This reaction is catalysed by the BP Clonase™ enzyme mix, which contains the bacteriophage lambda integrase (Int) and the *E. coli* integration host factor (IHF) proteins. **B. The LR Reaction.** The LR reaction facilitates the recombination of an *att*L substrate (entry clone) with an *att*R substrate (destination vector) to create an *att*B-containing expression clone. This reaction is catalysed by the LR Clonase™ enzyme mix, which contains the bacteriophage lambda integrase (Int) and Excisionase (Xis) and E. coli integration host factor (IHF) proteins. The BP and LR enzymes bind to the specific *att* sites, bring together the target sites, cleave them and covalently attach the DNA (Figure taken from the Gateway Technology Manual, Invitrogen).

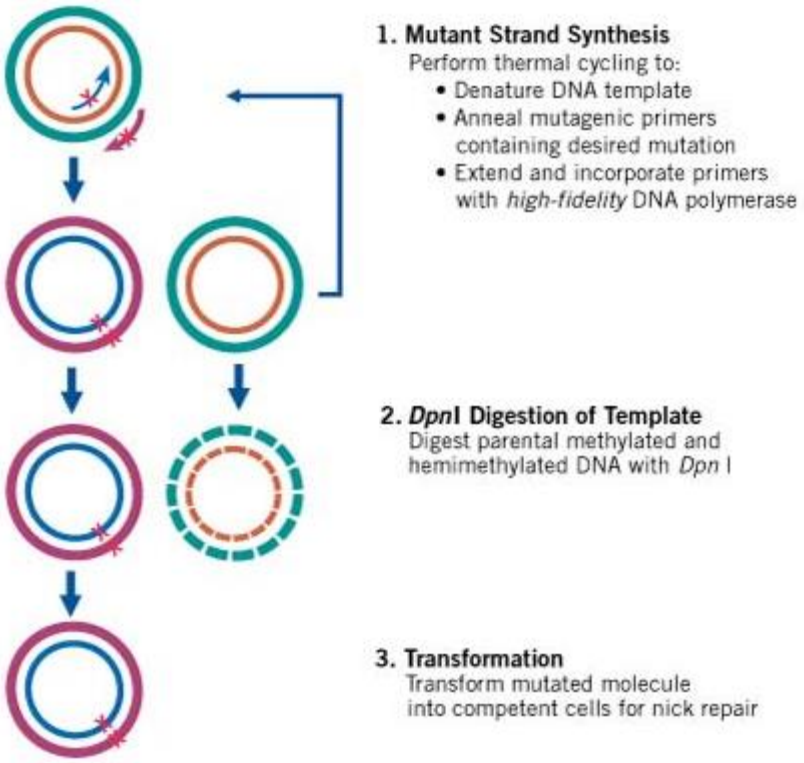


Figure 6.2. Overview of Site-Directed Mutagenesis. Site-directed mutagenesis functions by utilising a supercoiled double-stranded DNA vector with the insert of interest and two synthetic primers, both of which contain the desired mutation. The primers are annealed to the DNA and are extended by *Pfu*Ultra high fidelity DNA polymerase during elongation, generating a mutated plasmid. Once the mutated plasmid is produced, it is digested with *Dpn* I, an endonuclease that digests parental methylated and hemimethylated DNA. Mutant DNA is transformed in to XL10 Gold Cells (Image taken from www.genomics.agilent.com).

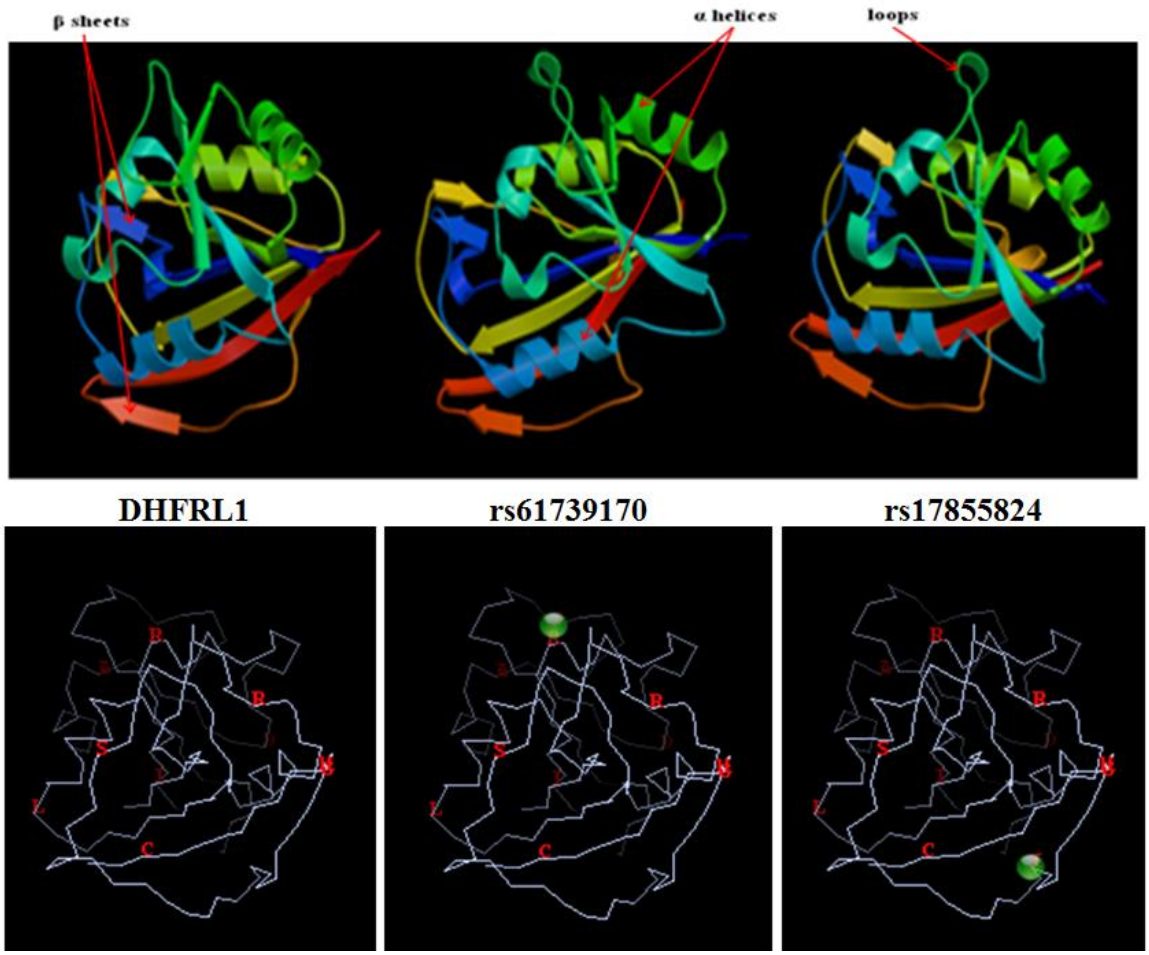


Figure 6.3. Tertiary Protein Structures of wild type DHFRL1, DHFRL1 rs61739170 (Pro->Ala) polymorphism and DHFRL1 rs17855824 (Val->Ile) polymorphism. **(A).** Protein structures were generated using the Protein Structure Prediction Server software from the Molecular Bioinformatics Centre, National Chiao Tung University. All three proteins are highly similar, with all main α helices and β sheets present but there are slight differences in the folding of their loops and coils. These differences may have an impact on the enzyme activity of the proteins, which in turn may have an impact on function. (B). Protein structures were generated using STRAP-NT software (http://www.charite.de/bioinf/strap/). The green circles indicate the location of the SNPs.

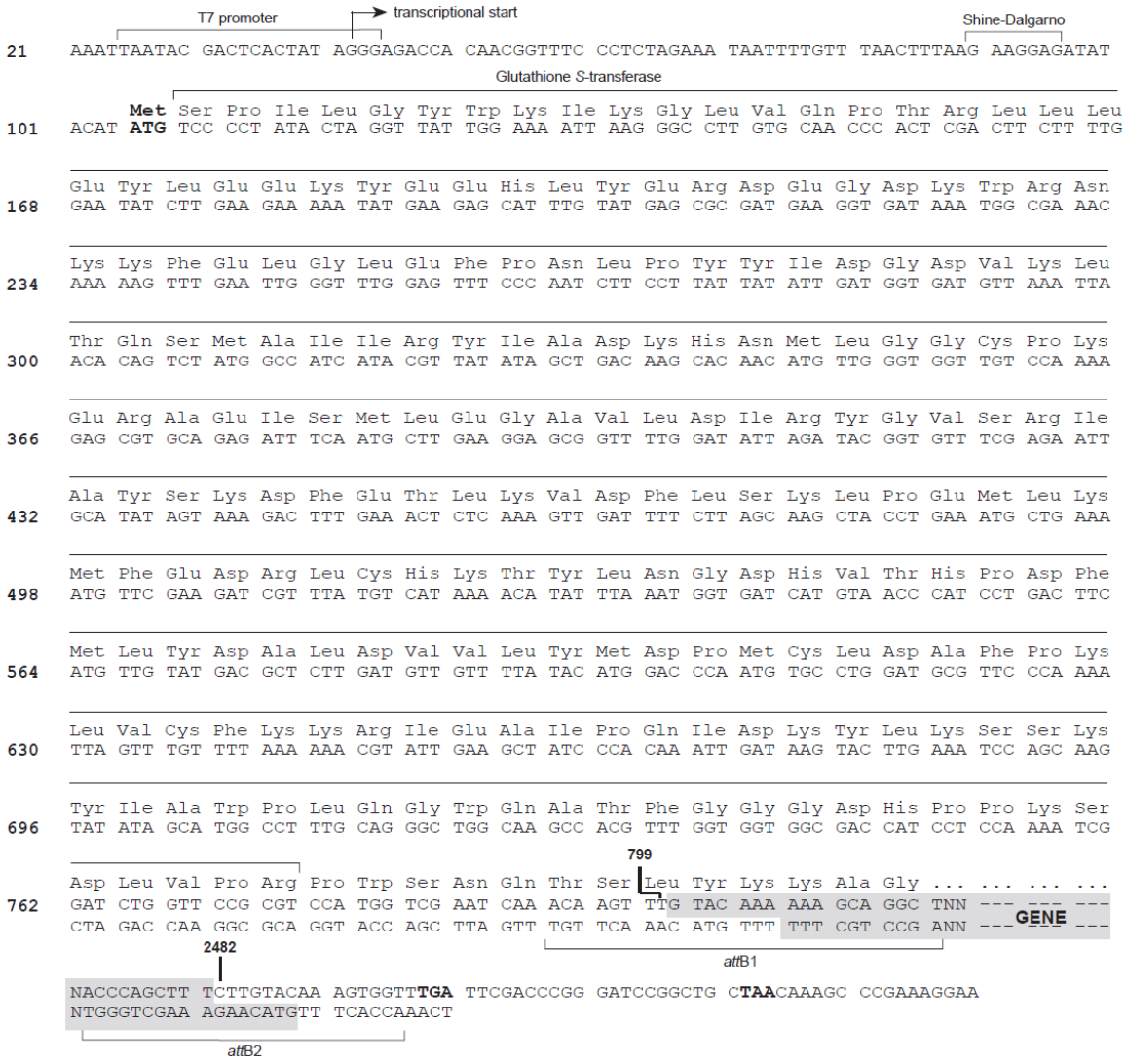


Figure 6.4. Sequence of the pDEST 15 vector. The highlighted region above shows the position where the *DHFR* / *DHFRL1* ORF was inserted into the vector (at position 799) and the region of the original sequence which was removed to allow for the entry of the *DHFRL1* ORF. The position at which the original vector sequence joins to the end of the *DHFRL1* ORF is also marked (at position 2482). (Image taken from www.invitrogen.com)

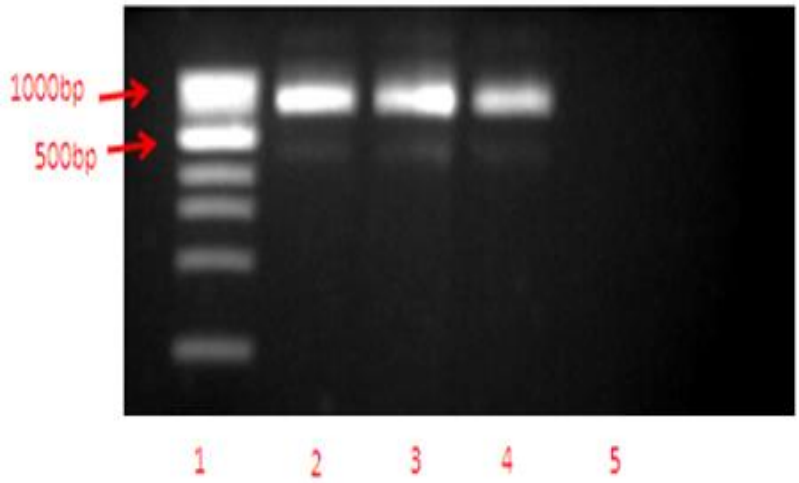


Figure 6.5. Gel electrophoresis of *DHFRL1* plasmid DNA. Lane 1 – 1 kb DNA ladder, Lane 2 - *DHFRL1* clone 1, Lane 3 – *DHFRL1* clone 2, Lane 4 – *DHFR* clone 1, Lane 5 – negative control. The bands seen on the gel indicates that DNA is present and that it has been extracted correctly.

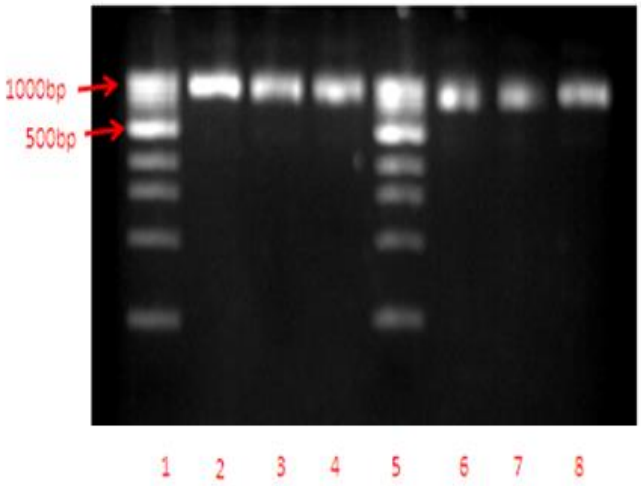


Figure 6.6. Gel electrophoresis of plasmid DNA. Lane 1 – 1 kb DNA ladder, Lane 2 - *DHFRL1* rs17855824 clone 1, Lane 3 – *DHFRL1* rs17855824 clone 2, Lane 4 – *DHFRL1* rs17855824 clone 3, Lane 5 – 1 kb DNA ladder, Lane 6 - *DHFRL1* rs61739170 clone 1, Lane 7 – *DHFRL1* rs61739170 clone 2, Lane 8 - *DHFRL1* rs61739170 clone 3. The bands seen on the gel indicates that DNA is present and that it has been extracted correctly.

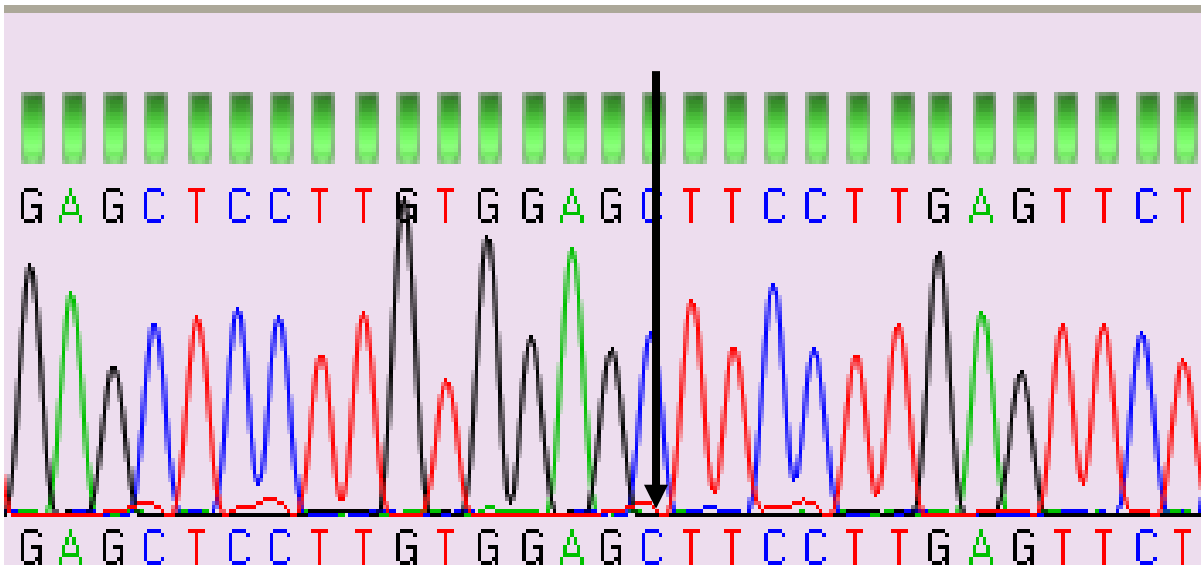


Figure 6.7. An example of a chromatogram from a sequencing run of one the successful mutant clones (reverse sequence). The arrow indicates confirmation of the base change of the rs61739170 polymorphism resulting in a C->G (Proline to Alanine amino acid) change. The chromatograms show evenly spaced peaks, with little baseline noise, indicating that our templates were of good quality. Direct sequencing and chromatogram results confirmed that the correct mutation had been inserted into the correct site.

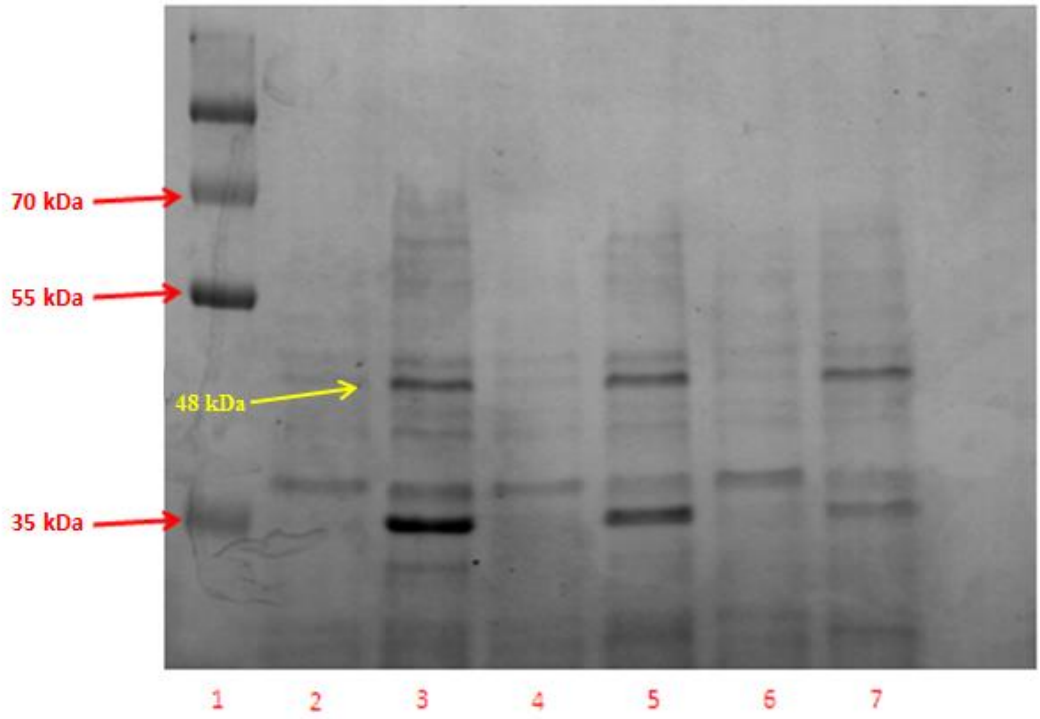


Figure 6.8. SDS-PAGE analysis of BL21-AI soluble fraction after 24 h. Protein was isolated from both induced and un-induced samples at 24 h and was separated into soluble and insoluble fractions after incubation at 37°C overnight. The soluble fraction was run on a 10 % w/v SDS-PAGE gel. Lane 1 – PageRuler pre-stained protein ladder, Lane 2 – DHFRL1 WT un-induced sample at 24 h, Lane 3 – WT induced sample at 24 h, Lane 4 - DHFRL1 C->G (rs61739170) polymorphism un-induced sample at 24 h, Lane 5 - DHFRL1 C->G (rs61739170) polymorphism induced sample at 24 h, Lane 6 - DHFRL1 G->A (rs17855824) polymorphism un-induced sample at 24 h and Lane 7 DHFRL1 G->A (rs17855824) polymorphism induced sample at 24 h. The protein of interest is clearly present at 48 kDa in the induced samples after 24 h and not in the un-induced samples.

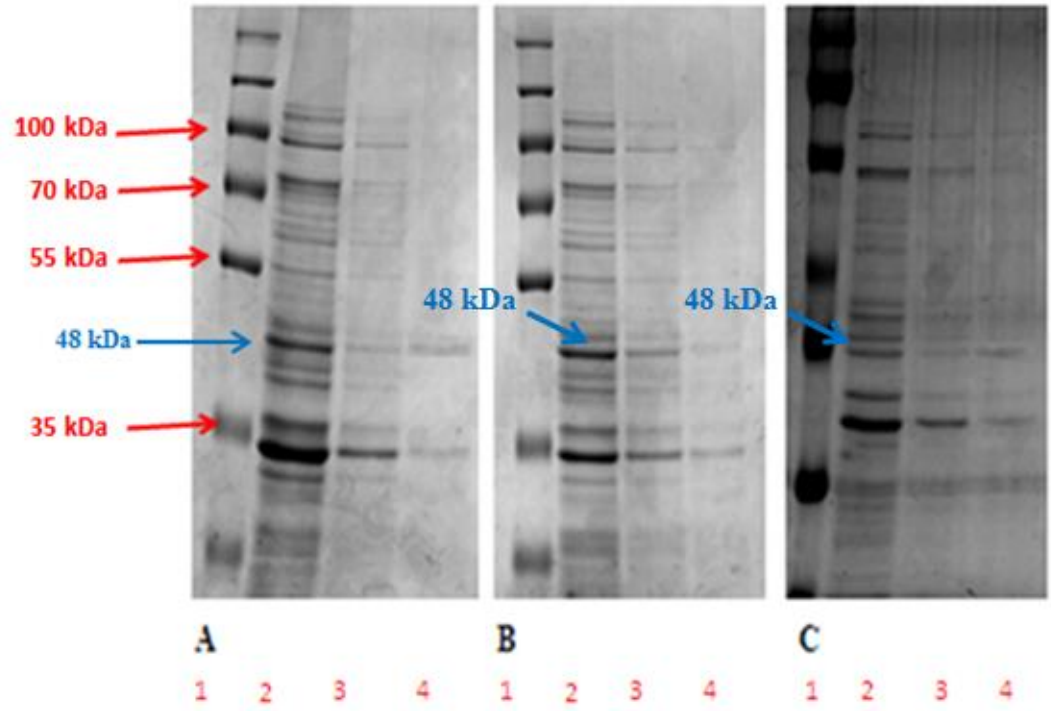


Figure 6.9. SDS-PAGE analysis of protein purification fractions. The GST tagged protein was purified by binding the GST tag to glutathione agarose. The purified protein was eluted using 35 mM glutathione and the samples were analysed by running the samples on a 10 % w/v SDS-PAGE gel. **(A)** GST-DHFRL1 WT purification. Lane 1 – PageRuler pre-stained protein ladder, Lane 2 - flow-through, Lane 3 - wash step, Lane 4 - elution. **(B)** GST-DHFRL1 C->G (rs61739170) polymorphism purification. Lane 1 - PageRuler pre-stained protein ladder, Lane 2 - flow-through, Lane 3 - wash step, Lane 4 - elution 1. **(C)** GST-DHFRL1 G->A (rs17855824) polymorphism purification. Lane 1 - PageRuler pre-stained protein ladder, Lane 2 - flow-through, Lane 3 - wash step, Lane 4 - elution 1. Clear bands of the protein of interest can be seen at 48 kDa in the elution steps.

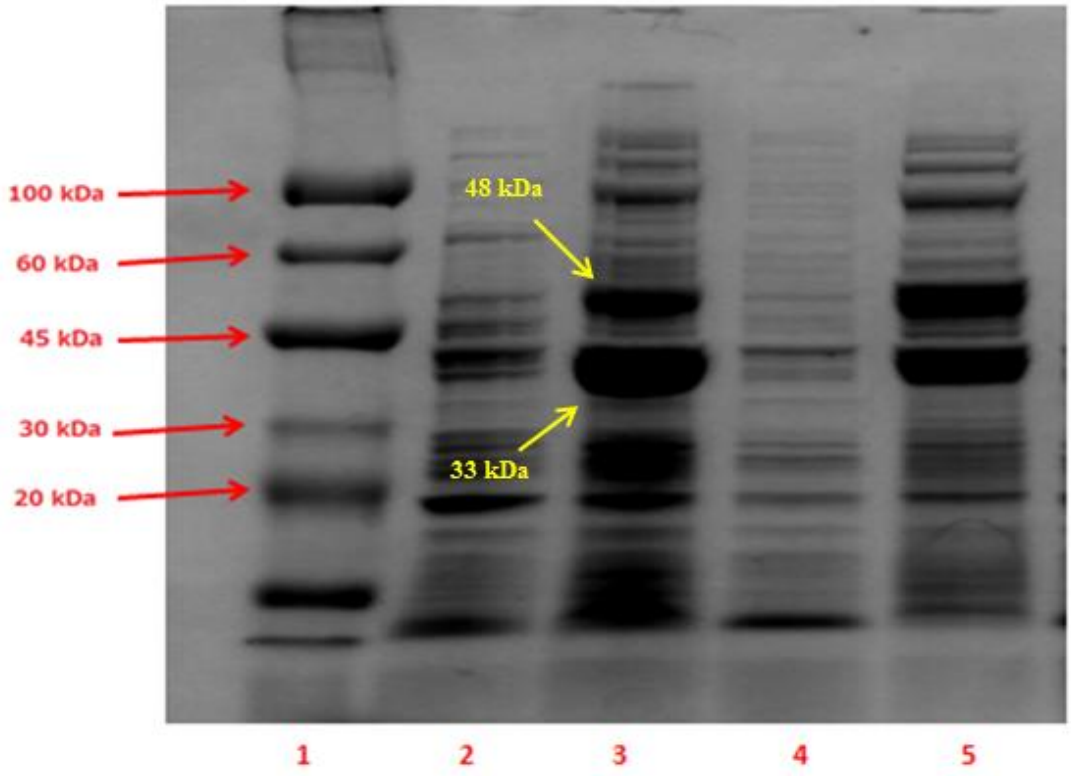


Figure 6.10. SDS-PAGE analysis of BL21-AI soluble fraction after 24 h. Protein was isolated from both induced and un-induced samples at 24 h and was separated into soluble and insoluble fractions after incubation at 37°C overnight. The soluble fraction was run on a 10 % w/v SDS-PAGE gel. Lane 1 – Colorburst Electrophoresis Marker protein ladder, Lane 2 – DHFRL1 WT un-induced sample at 24 h, Lane 3 – WT induced sample at 24 h, Lane 4 - DHFRL1 C->G (rs61739170) polymorphism un-induced sample at 24 h, Lane 5 - DHFRL1 C->G (rs61739170) polymorphism induced sample at 24 h. The protein of interest is clearly present at ~ 48 kDa in the induced samples after 24 h and not in the un-induced samples, however, a strong band is also seen below the 48 kDa band at ~ 33 kDa.

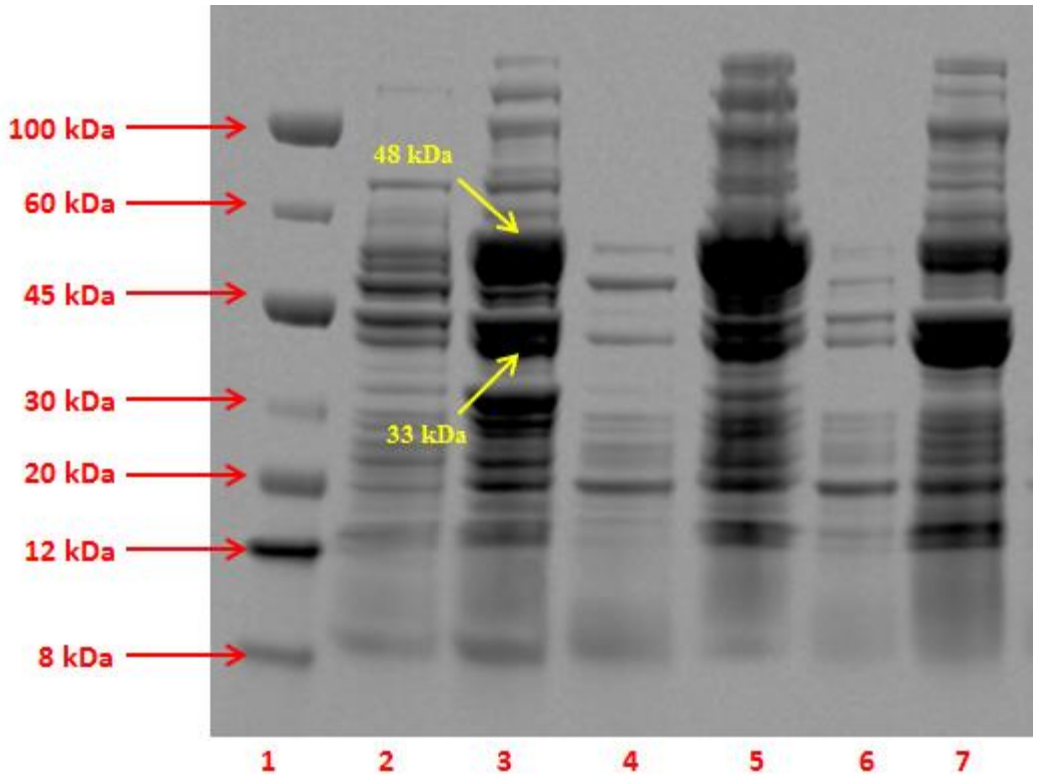


Figure 6.11. SDS-PAGE analysis of DHFRL1 BL21-AI soluble fraction after 24 h grown at different temperatures. Protein was isolated from both DHFRL1 induced and un-induced samples at 24 h and was separated into soluble and insoluble fractions after incubation at different temperatures overnight. The soluble fraction was run on a 10 % w/v SDS-PAGE gel. Lane 1 – Colorburst Electrophoresis Marker protein ladder, Lane 2 – DHFRL1 WT un-induced sample at 25°C after 24 h, Lane 3 – DHFRL1 WT induced sample at 25°C after 24 h, Lane 4 - DHFRL1 WT un-induced sample at 30°C after 24 h, Lane 5 - DHFRL1 WT induced sample at 30°C after 24 h. Lane 6 - DHFRL1 WT un-induced sample at 37°C after 24 h, Lane 7 - DHFRL1 WT induced sample at 37°C after 24 h. The protein of interest is clearly present at ~ 48 kDa in the induced samples after 24 h and not in the un-induced samples, however, a strong band is still present below the 48 kDa band at ~ 33 kDa.

ACAAGTTTGTACAAAAAAGCAGGCTTCGAAAATCTGTACTTCCAGGGGATGTTTC
TTTTGCTAAACTGCATCGTCGCTGTGTCCCAAAACATGGGCATCGGCAAGAACGGGG
ACCTGCCCAGGCCGCCGCTCAGGAATGAATTCAGGTATTTCCAGAGAATGACCACAAC
TTCTTCAGTAGAGGGTAAACAGAATCTGGTGATTATGGGTAGGAAGACCTGGTTCTCC
ATTCCTGAGAAGAATCGACCTTTAAAGGATAGAATTAATTTAGTTCTCAGCAGAGAAC
TCAAGGAACCTCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGATGCCTTAAAA
CTTACTGAACGACCAGAATTAGCAAATAAAGTAGACATGATTTGGATAGTTGGTGGCA
GTTCTGTTTATAAGGAAGCCATGAATCACCTAGGCCATCTTAAACTATTTGTGACAAGG
ATCATGCAGGACTTTGAAAGTGACACGTTTTTTTCAGAAATTGACTTGGAGAAATATAA
ACTTCTGCCTGAATACCCAGGTGTTCTCTCTGATGTCCAGGAGGGGAAACACATCAAGT
ACAAATTTGAAG**TATGTGAGAAGGATGATCATCATCATCATCATCATTAGGACCCA
GCTTTCTTGTACAAAGTGGT**

Figure 6.12. Primer design for His tag incorporation for the DHFRL1 clone. The letters in bold represent the forward and reverse primers. The att sites are represented in green, the TEV cleavage site in blue, the His tag in pink and the stop codon in yellow.

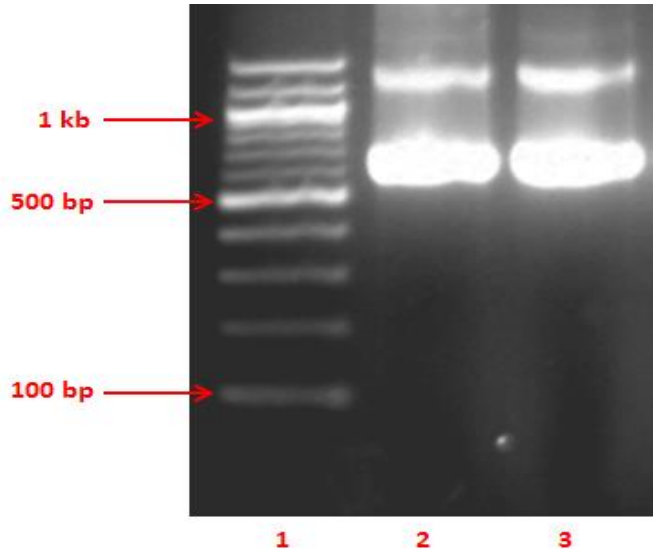


Figure 6.13. Gel electrophoresis of PCR products amplified with specific *DHFR* and *DHFRL1* primers designed to incorporate a TEV cleavage site and a His tag using HEK 293 cDNA as a template. Lane 1 – 100 bp DNA ladder, Lane 2 – DHFR PCR product, Lane 3 –DHFRL1 PCR product. The gel shows bands of the correct size, 646 bp indicating that the target products have been amplified.

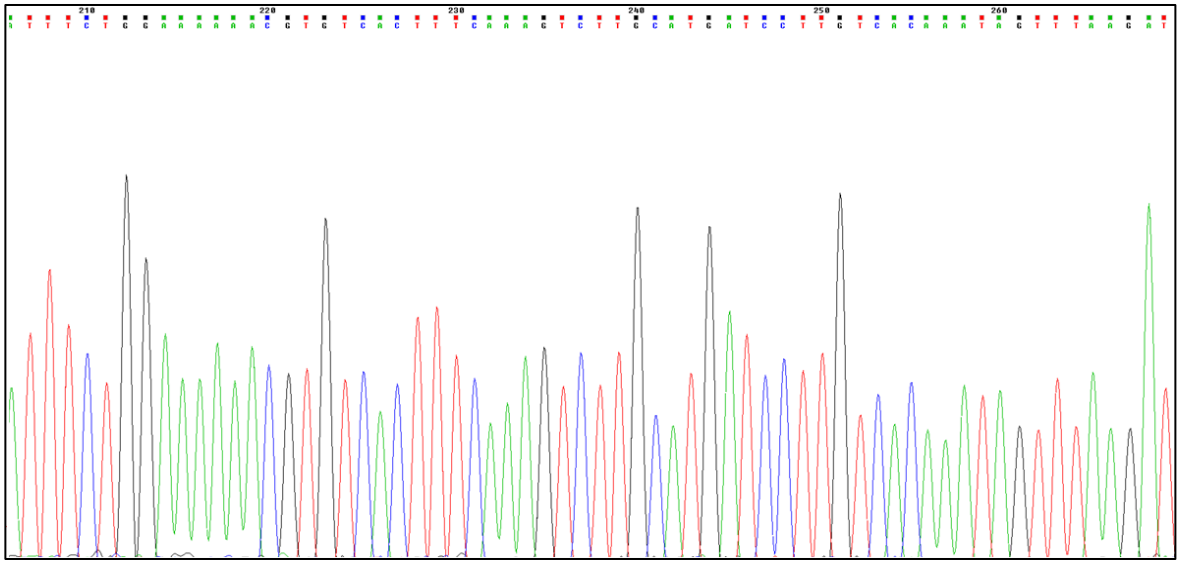


Figure 6.14. An example of a chromatogram from a Sanger sequencing run of one of the HIS incorporated *DHFR* recombinant plasmid DNA samples. The chromatogram shows evenly spaced peaks, with little baseline noise, indicating that our template and primers were good.

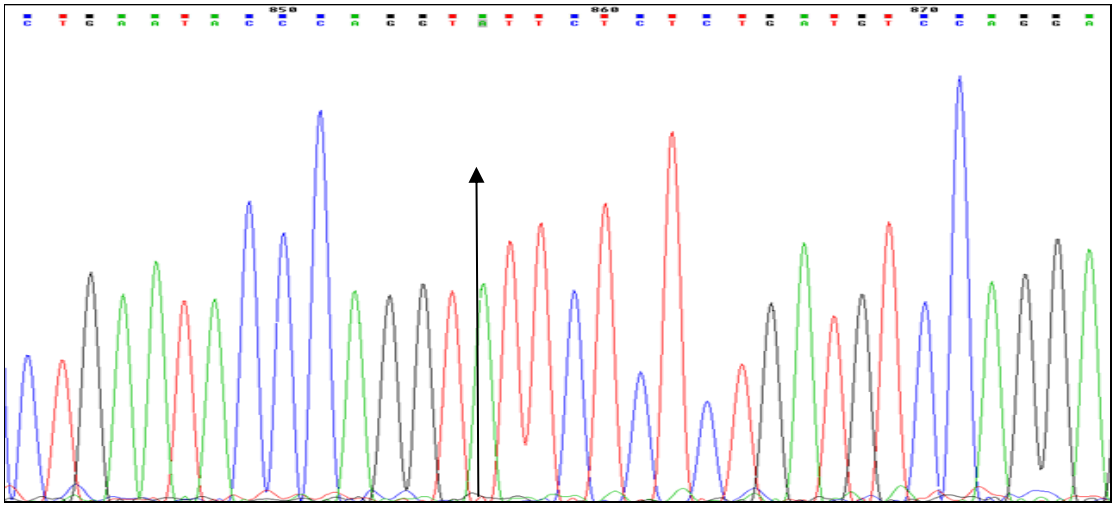


Figure 6.15. An example of a chromatogram from a sequencing run of one the successful mutant clones. The arrow indicates confirmation of the base change (G->A) of the rs17855824 polymorphism resulting in a Valine to Isoleucine amino acid change**.** The chromatogram shows evenly spaced peaks, with little baseline noise, indicating that our template and primers were of good quality. This demonstrated that the site directed mutagenesis was successful.

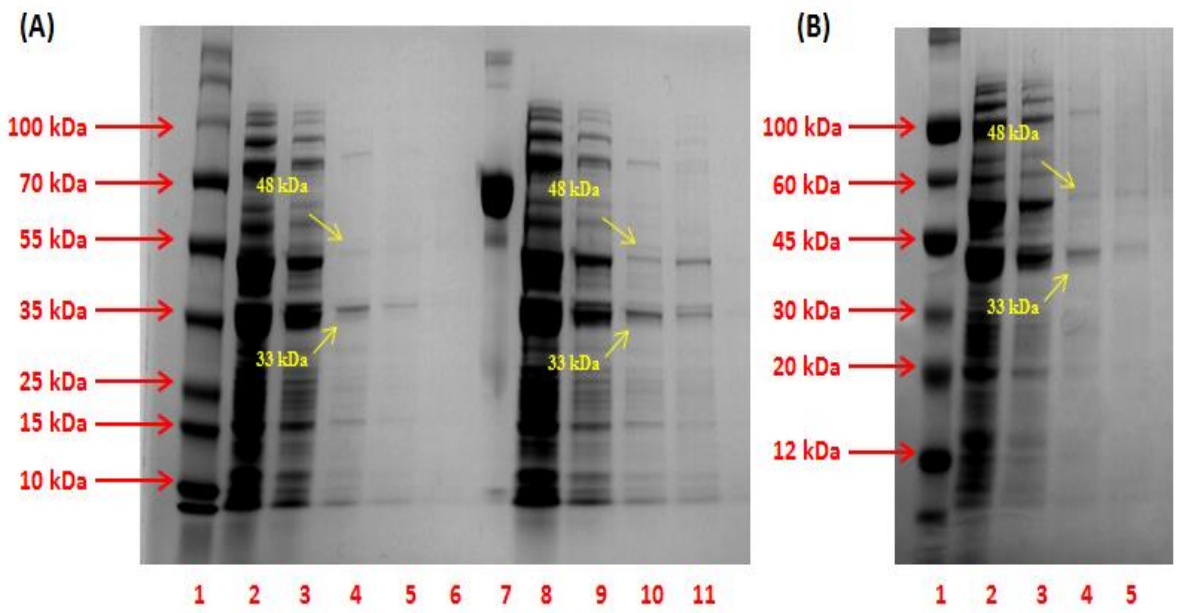


Figure 6.16. SDS-PAGE analysis of protein purification fractions. The GST and His tagged protein was purified by binding the His tag to a nickel affinity gel. The purified protein was eluted using 250 mM imidazole and the samples were analysed by running the samples on a 10 % w/v SDS-PAGE gel. **(A)** GST-DHFRL1 WT and GST-DHFRL1 C->G (rs61739170) polymorphism purification. Lane 1 –PAGE Ruler plus prestained protein marker protein ladder, Lane 2 – WT flow-through, Lane 3 – WT wash step, Lane 4 – WT elution 1, Lane 5 – WT elution 2, Lane 6 – blank, Lane 7 – BSA control, Lane 8 – C->G flow-through, Lane 9 – C->G wash step, Lane 10 – C->G elution 1, Lane 11 – C->G elution 2. **(B)** GST-DHFRL1 G->A (rs17855824) polymorphism purification. Lane 1 - Colorburst Electrophoresis Marker protein ladder, Lane 2 – G->A flow-through, Lane 3 – G->A wash step, Lane 4 – G->A elution 1, Lane 5 – G-> elution 2. A clear but faint band of the protein of interest can be seen at 48 kDa in the elution steps with a stronger band seen at the lower 33 kDa size.

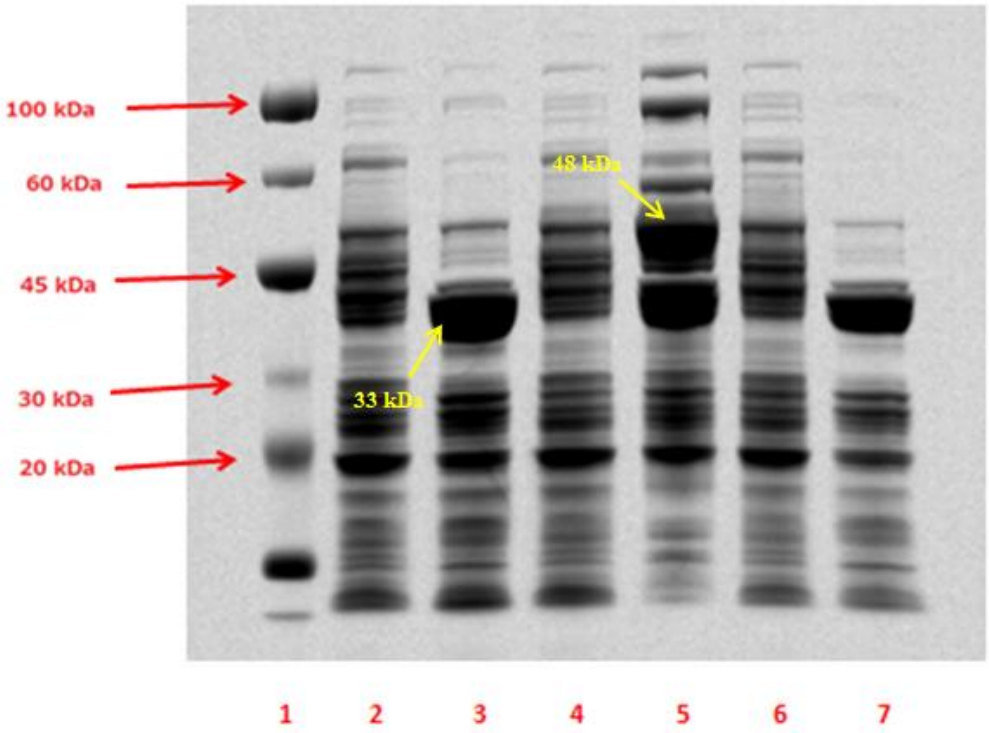


Figure 6.17. SDS-PAGE analysis of recombinant clones expressed in BL21-AI separated into soluble and insoluble fractions after 24 h. Protein was isolated from both induced and un-induced samples at 24 h and was separated into soluble and insoluble fractions after incubation at 37°C overnight. The soluble fraction was run on a 10 % w/v SDS-PAGE gel. Lane 1 – Colorburst Electrophoresis Marker protein ladder, Lane 2 – DHFRL1 WT un-induced sample at 24 h, Lane 3 – WT induced sample at 24 h, Lane 4 - DHFRL1 C->G (rs61739170) polymorphism un-induced sample at 24 h, Lane 5 - DHFRL1 C->G (rs61739170) polymorphism induced sample at 24 h, Lane 6 - DHFRL1 G->A (rs17855824) polymorphism un-induced sample at 24 h and Lane 7 DHFRL1 G->A (rs17855824) polymorphism induced sample at 24 h.. The protein of interest at ~ 48 kDa is only clearly induced in Lane 5, however, the 33 kDa is strongly induced in all of the induced samples after 24 h.

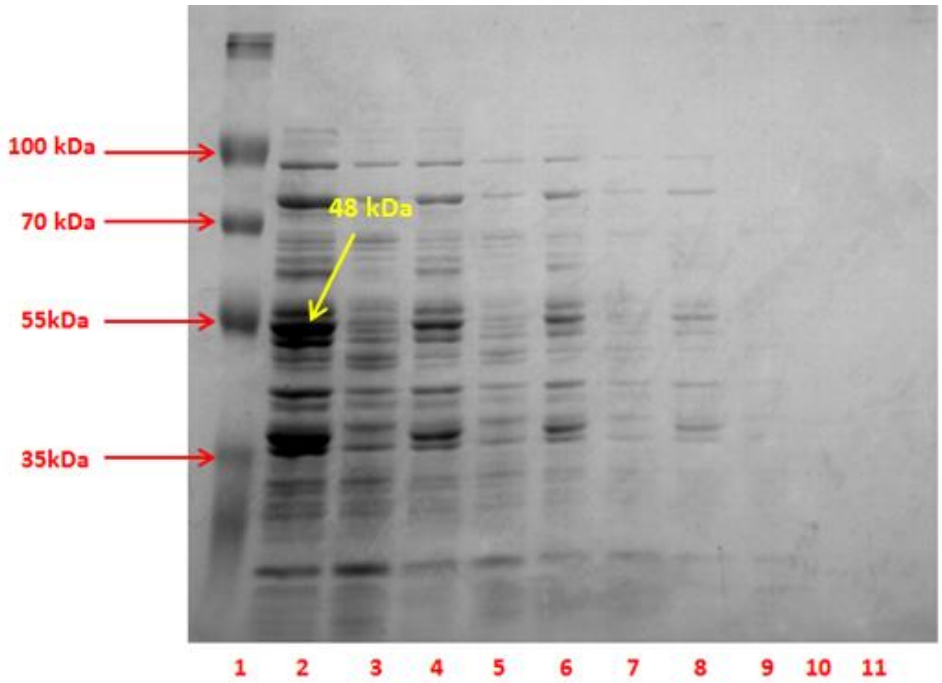


Figure 6.18. SDS-PAGE analysis of serial dilution of the rs61739170 (C->G) induced and un-induced protein samples. Lane 1 – PAGE Ruler plus protein ladder, Lane 2 - 11.3 µg of C->G induced sample, Lane 3 - 11.3 µg of C->G un-induced sample, Lane 4 - 5.6 µg of C->G induced sample, Lane 5 - 5.6 µg of C->G un-induced sample, Lane 6 - 2.8 µg of C->G induced sample, Lane 7 - 2.8 µg of C->G un-induced sample, Lane 8 - 1.4 µg of C->G induced sample, Lane 9 - 1.4 µg of C->G un-induced sample, Lane 10 - 0.7 µg of C->G induced sample, Lane 11 - 0.7 µg of C->G un-induced sample.

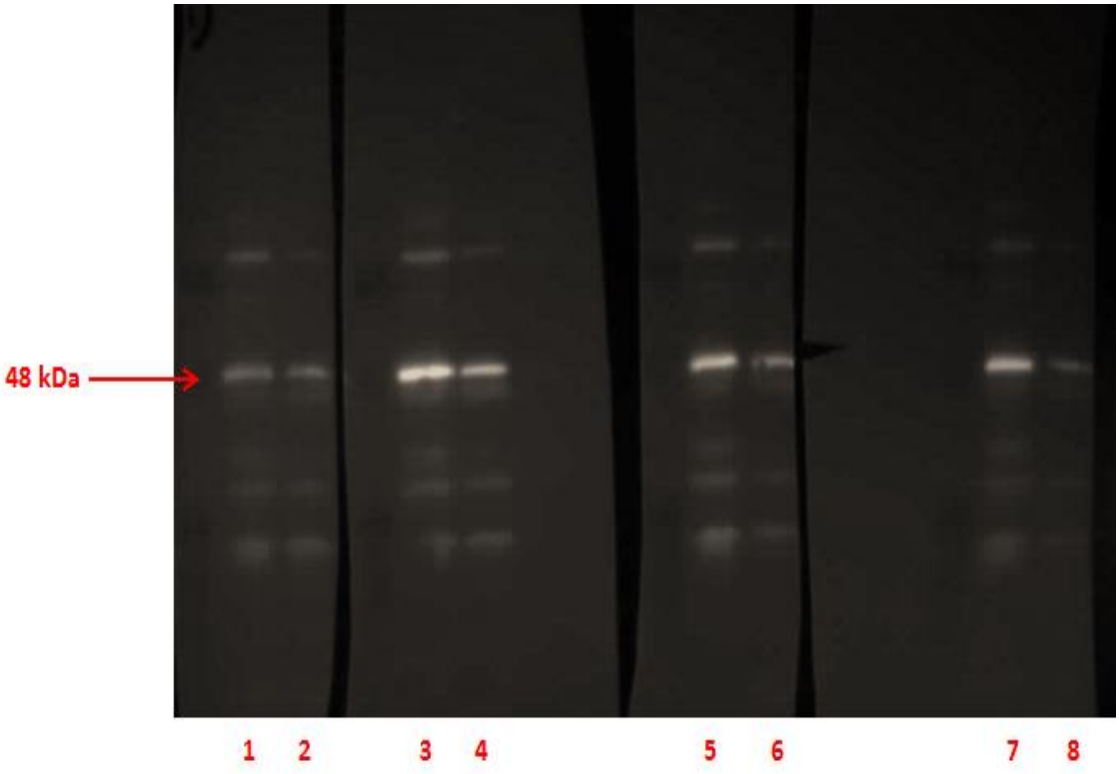


Figure 6.19. Optimisation of GST antibody by Western blot analysis on 2.8 µg of total protein. Lane 1 - C->G induced sample, 1:2,000 primary and 1:100,000 secondary antibody, Lane 2 - C->G un-induced sample, 1:2,000 primary and 1:100,000 secondary antibody Lane 3 - C->G induced sample, 1:3,000 primary and 1:100,000 secondary antibody, Lane 4 - C->G un-induced sample, 1:3,000 primary and 1:100,000 secondary antibody, Lane 5 - C->G induced sample, 1:2,000 primary and 1:150,000 secondary antibody, Lane 6 - C->G induced sample, 1:2,000 primary and 1:150,000 secondary antibody, Lane 7 - C->G induced sample, 1:2,000 primary and 1:200,000 secondary antibody, Lane 8 - C->G induced sample, 1:2,000 primary and 1:200,000 secondary antibody.

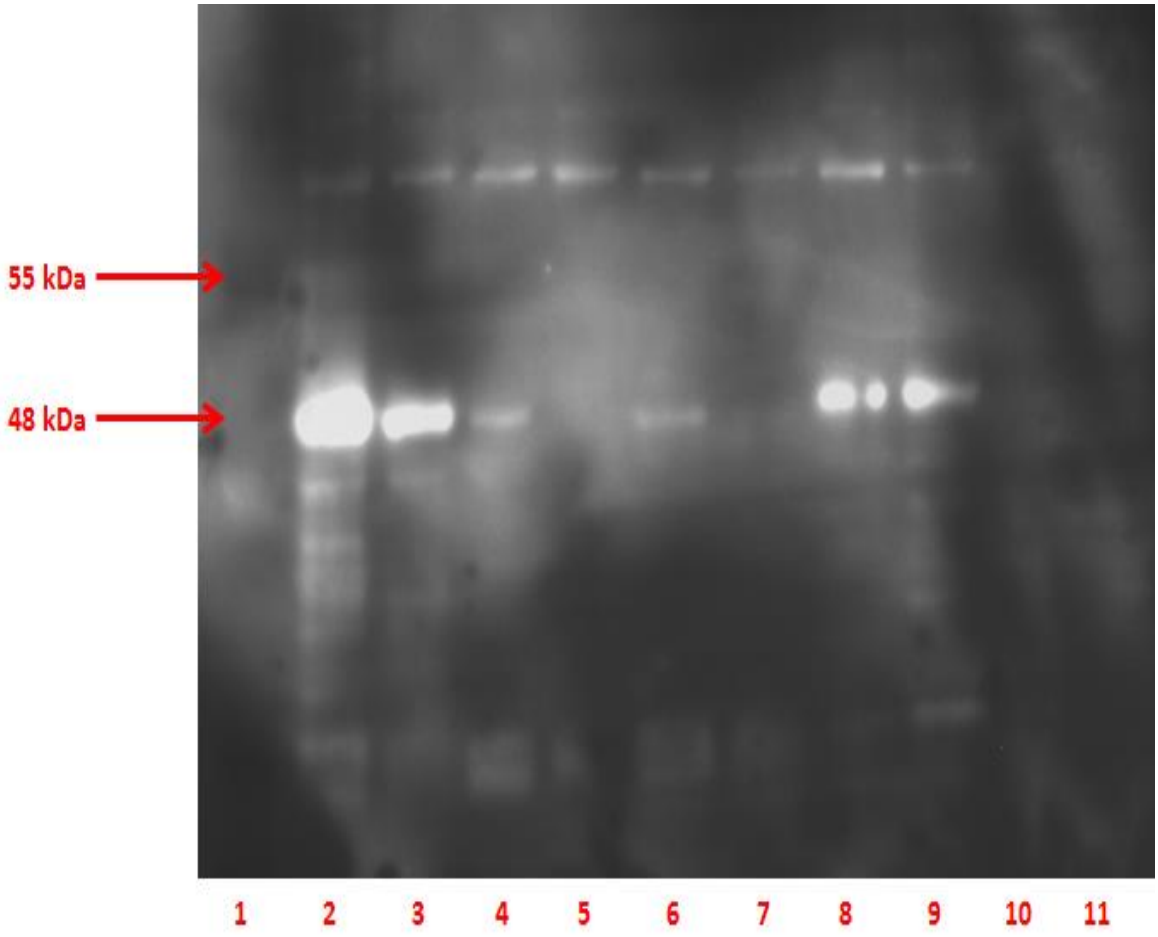


Figure 6.20. Western blot analysis of soluble protein samples expressing recombinant DHFR / DHFRL1 where induction was observed or not observed (by Coomassie staining), probed with the GST antibody. Lane 1 – PAGE Ruler plus protein ladder, Lane 2 – C->G induced, Lane 3 – C->G un-induced, Lane 4 – G->A induced, Lane 5 G->A un-induced, Lane 6 DHFRL1 GST 24 h induced, Lane 7 – DHFRL1 GST 24 h un-induced, Lane 8 – DHFR His 24 h induced, Lane 9 – DHFR His 24 h un-induced, Lane 10 – untransformed control

GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAAATCTGTACTTCCAGGGGATGGTTGGTAGCCTGAACT
GCATTGTCGCGGTATCTCAGAATATGGGCATTGGCAAAAACGGCGACTTACCGTGGCCGCCGTTGCGTAAT
GAATTTCGCTATTTTCAGCGTATGACGACTACTTCATCAGTTGAAGGGAAACAAAACCTCGTGATTATGGGC
AAGAAGACCTGGTTCTCTATTCCCGAGAAAAATCGTCCGTTAAAGGGCCGTATTAACCTTGTGTTATCTCGC
GAGCTGAAGGAGCCGCCGCAGGGGGCCCATTTCTTATCCCGTTCCCTGGATGACGCACTGAAGCTGACCGA
ACAACCTGAACTGGCTAATAAAGTCGACATGGTGTGGATTGTAGGCGGTAGCTCGGTTTATAAAGAAGCGA
TGAACCACCCCGGGCACTTAAAACTGTTCGTAACTCGCATCATGCAAGACTTTGAAAGTGATACGTTTTTCC
CCGAAATCGACCTGGAAAAGTATAAACTGCTGCCGGAATATCCGGGGGTTTTGTCCGATGTTCAGGAGGAA
AAAGGTATCAAATACAAGTTTGAGGTGTATGAAAAGAACGATTAAGACCCAGCTTTCTTGTACAAAGTGGT
GGGG

Figure 6.21. DHFR sequence optimised for rare codon usage. The bases highlighted in yellow represent the G-clamp, the bases highlighted in pink represent the attB sites and the bases highlighted in green represent the TEV Cleavage site.

Organism	*Escherichia*
CDS length	564

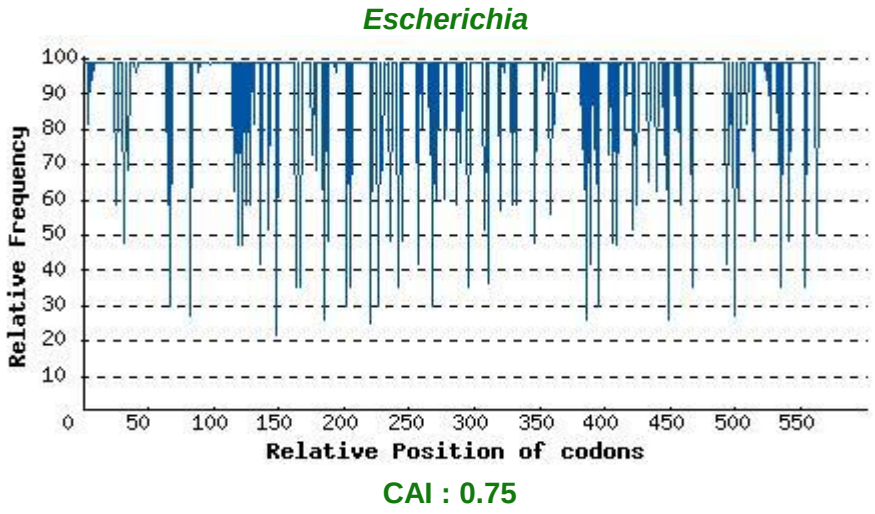


Figure 6.22. The distribution of codon usage frequency along the length of the DHFR gene to be expressed in the target host organism. The Codon Adaptation Index (CAI) is a value that is correlated with the possibility of high protein expression level in a desired expression organism, where a CAI of 1.0 is considered ideal. Our gene has a CAI value of 0.75. The lower the CAI number, the higher the chance that the gene of interest will be expressed poorly (www.genscript.com).

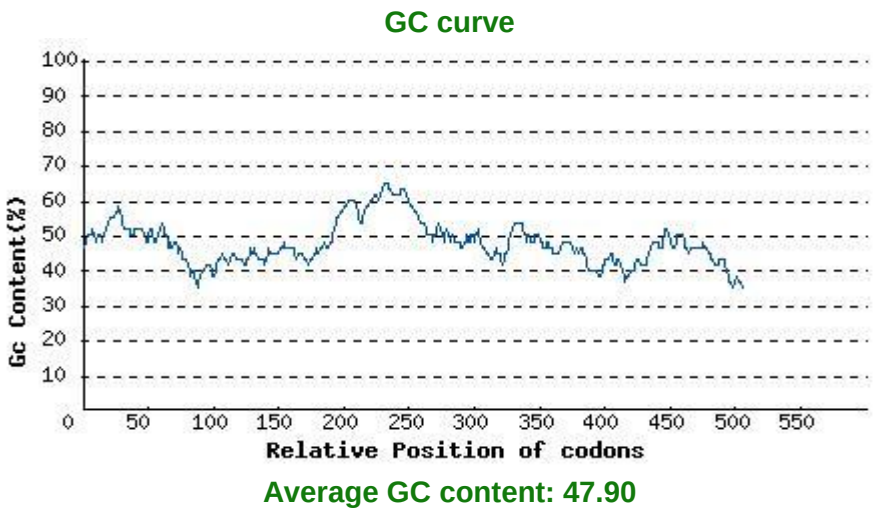


Figure 6.23. A curve showing the average GC content of the DHFR gene. The GC content of the DHFR gene is 47.90%. The ideal % range of GC content is between 30% and 70%. Any peaks outside of this range can adversely affect transcriptional and translational efficiency (www.genscript.com).

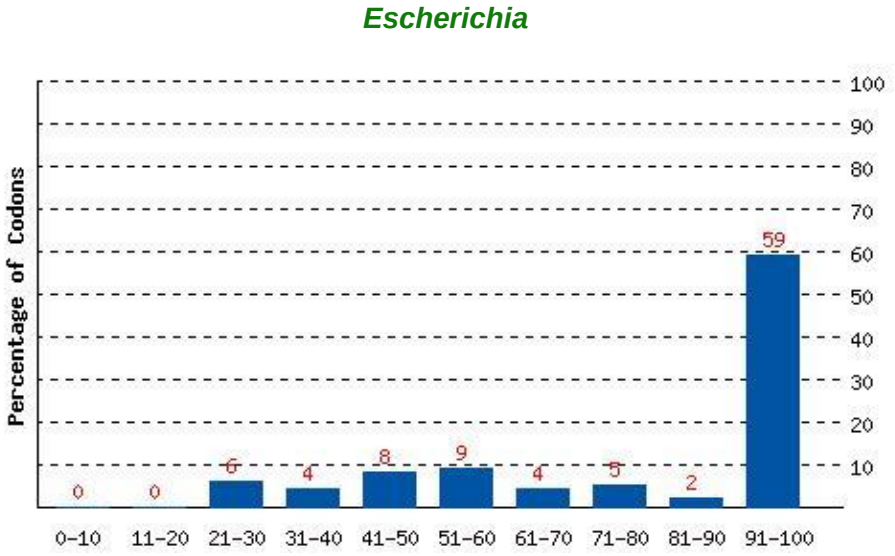


Figure 6.24. The percentage distribution of codons in computed codon quality groups. The value of 100 is set for the codon with the highest usage frequency for a given amino acid in the desired expression organism. Codons with values lower than 30 are likely to hamper the expression efficiency. The percentage of low frequency (<30%) codons based on our target host organism is 6 % (www.genscript.com).

GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAAATCTGTACTTCCAGGGGATGTTTCTGCTGTTAAATT
GCATCGTTGCCGTCTCCCAGAACATGGGCATTGGTAAGAATGGTGACCTGCCACGCCCGCCGCTGCGCAAC
GAATTTCGTTACTTTCAACGGATGACAACTACCAGTTCTGTAGAAGGCAAACAGAACTTGGTGATCATGGG
CCGTAAAACCTGGTTCAGTATCCCGGAAAAGAACCGCCCGTTAAAAGATCGGATTAATCTGGTGTTGAGTC
GCGAACTTAAGGAACCACCGCAGGGGGCACACTTTCTGGCCCGTTCACTGGATGATGCTCTGAAATTAACA
GAACGTCCAGAATTAGCGAACAAAGTCGATATGATCTGGATTGTCGGTGGGAGTAGCGTCTACAAAGAGG
CGATGAATCATCTGGGCCACTTGAAGCTCTTTGTCACCCGGATCATGCAGGATTTTGAGTCAGATACATTCT
TTTCTGAAATCGATCTGGAAAAATATAAACTTCTGCCGGAATACCCGGGCGTTCTCAGTGACGTCCAGGAA
GGTAAGCACATCAAATATAAATTTGAGGTGTGCGAAAAAGACGATTAAGACCCAGCTTTCTTGTACAAAGT
GGTGGGGG

Figure 6.25. DHFRL1 sequence optimised for rare codon usage. The bases highlighted in yellow represent the G-clamp, the bases highlighted in pink represent the attB sites and the bases highlighted in green represent the TEV Cleavage site.

Organism	*Escherichia*
CDS length	564

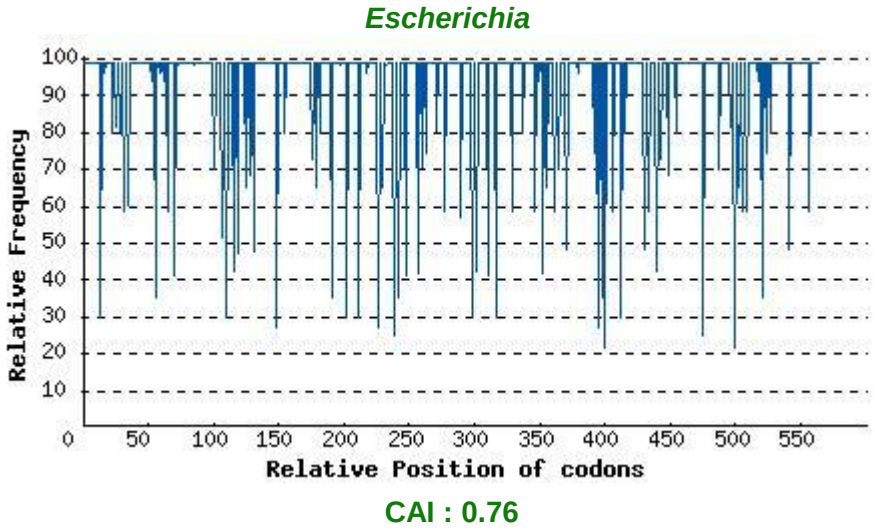


Figure 6.26. The distribution of codon usage frequency along the length of the DHFRL1 gene to be expressed in the target host organism. The Codon Adaptation Index (CAI) is a value that is correlated with the possibility of high protein expression level in a desired expression organism, where a CAI of 1.0 is considered ideal. Our gene has a CAI value of 0.76. The lower the CAI number, the higher the chance that the gene of interest will be expressed poorly (www.genscript.com).

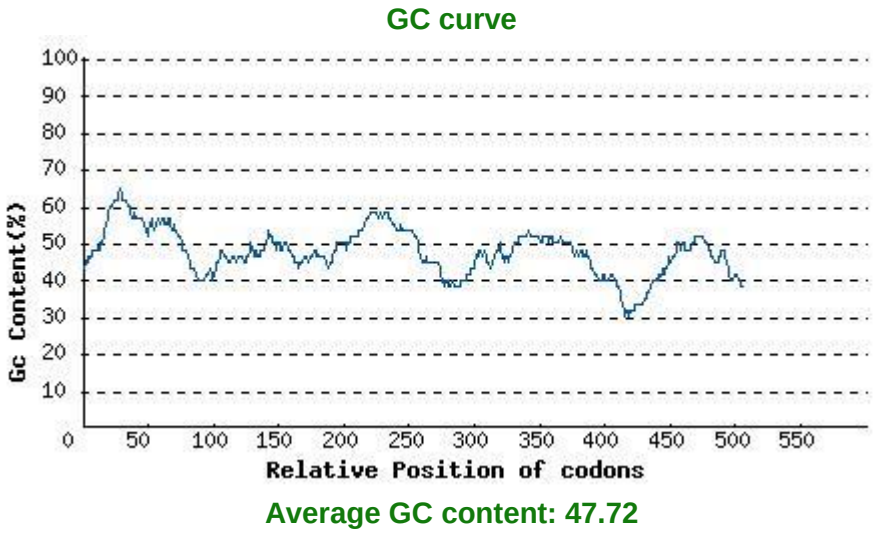


Figure 6.27. A curve showing the average GC content of the DHFR gene. The GC content of the DHFR gene is 47.72%. The ideal % range of GC content is between 30% and 70%. Any peaks outside of this range can adversely affect transcriptional and translational efficiency (www.genscript.com).

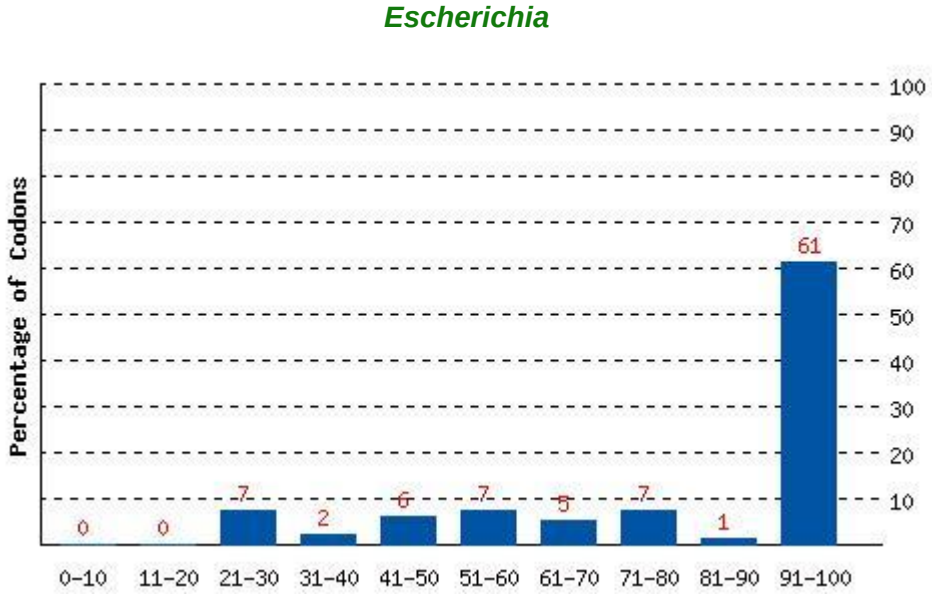


Figure 6.28. The percentage distribution of codons in computed codon quality groups. The value of 100 is set for the codon with the highest usage frequency for a given amino acid in the desired expression organism. Codons with values lower than 30 are likely to hamper the expression efficiency. The percentage of low frequency (<30%) codons based on our target host organism is 7 % (www.genscript.com).

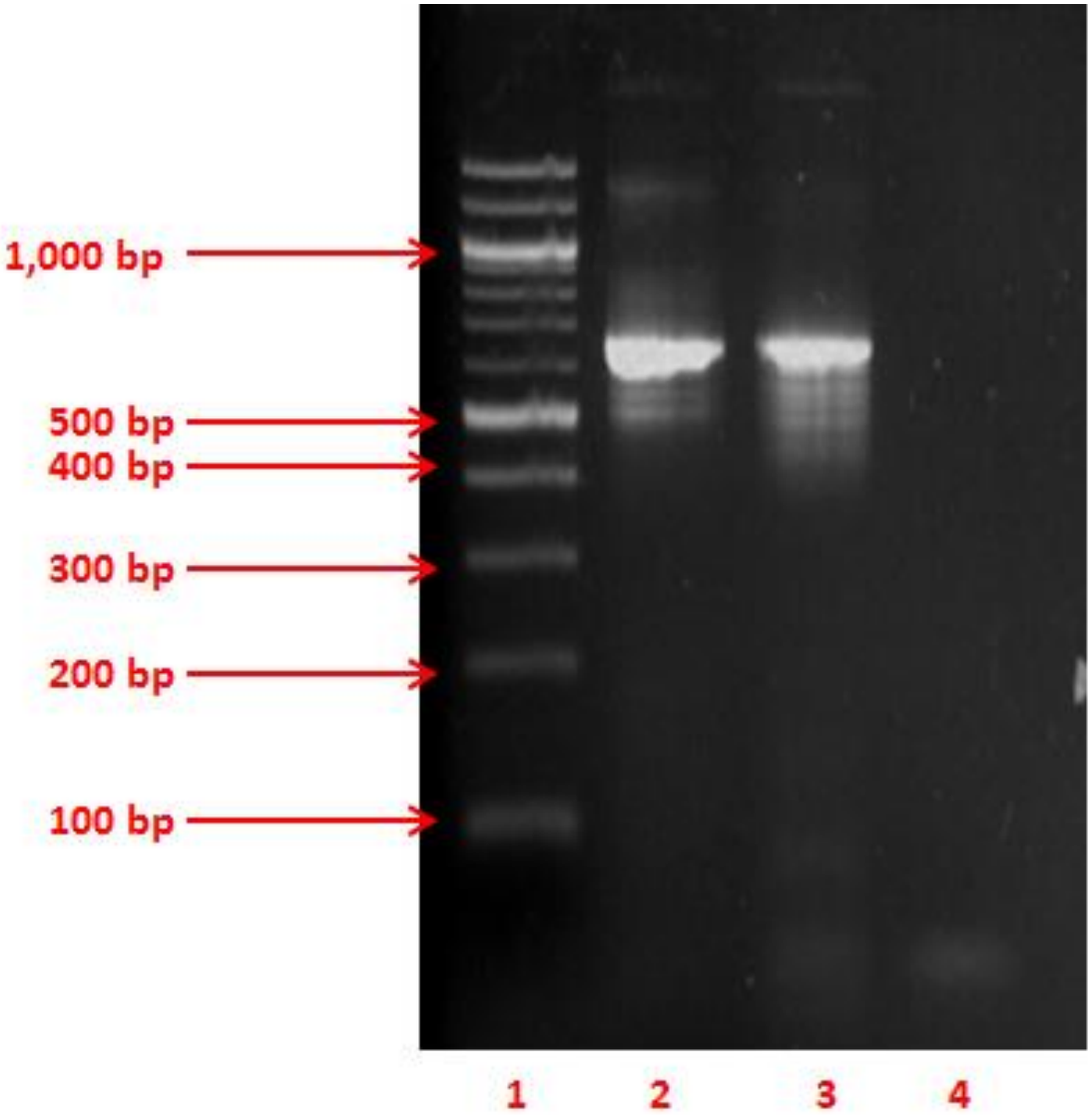


Figure 6.29. Gel Electrophoresis of amplified DHFR and DHFRL1 PCR products. Lane 1 – 100 bp DNA ladder, Lane 2 – DHFR PCR product, Lane 3 – DHFRL1 PCR product, Lane 4 – negative control. PCR products of the correct size (646 bp) were obtained indicating that the DNA samples have been amplified.

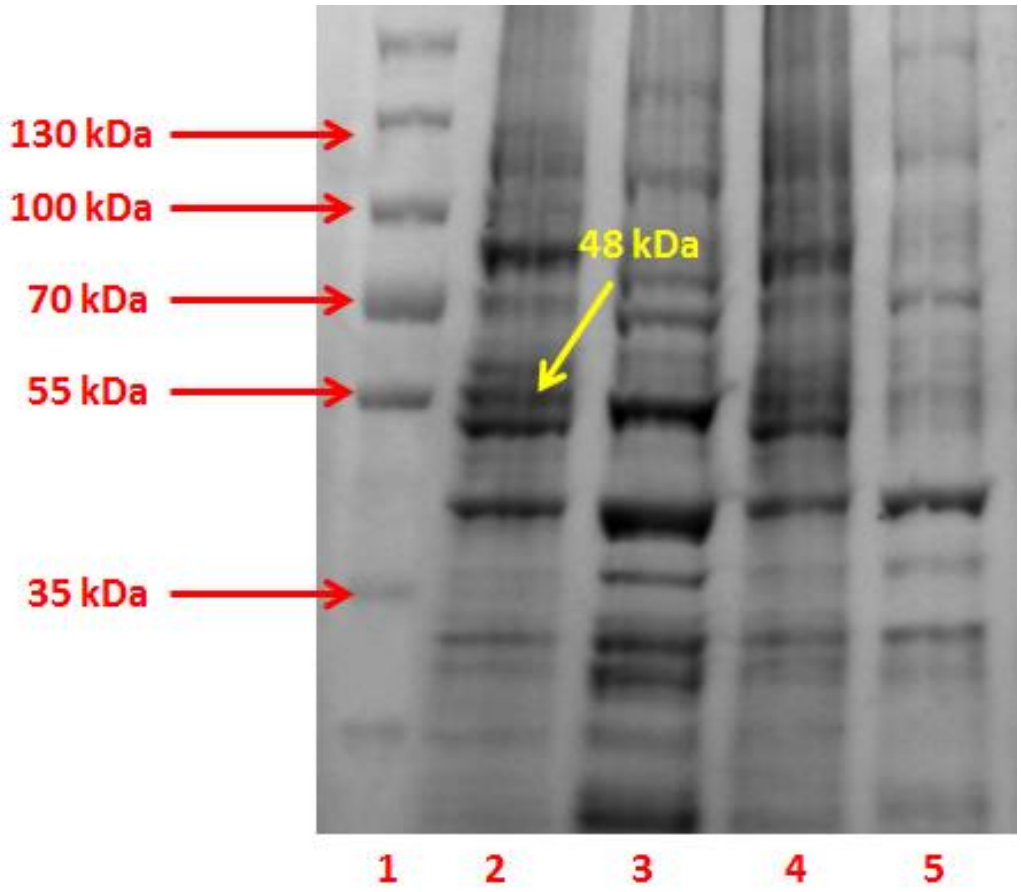


Figure 6.30. An example of an SDS-PAGE gel showing numerous proteins of the same size being expressed. Lane 1 – PAGE Ruler Plus protein ladder, Lane 2 – 1 h induced soluble, Lane 3 – 1 h induced insoluble, Lane 4 – 1 h un-induced soluble, Lane 5 – 1 h un-induced insoluble. Cultures were grown at 30°C for 1 h. Numerous bands can be seen at ~ 48 kDa, preventing us to identify our protein of interest.

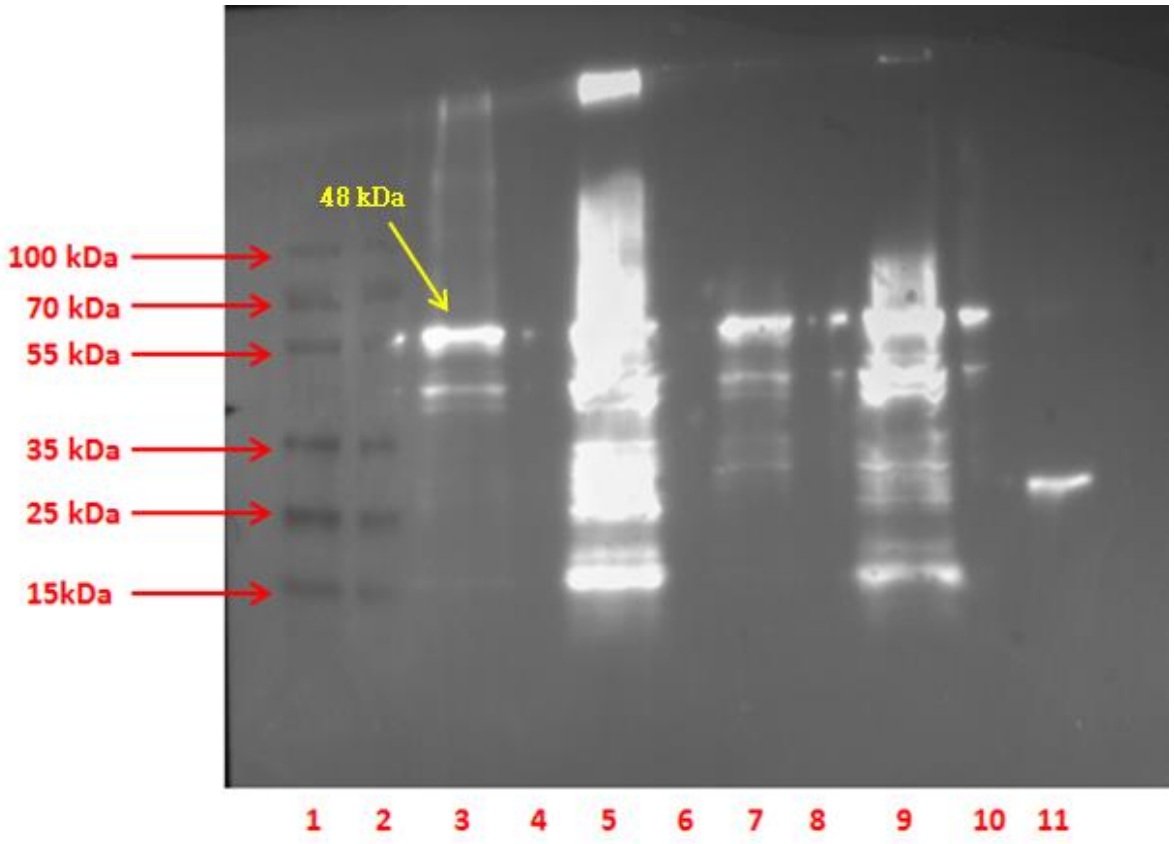


Figure 6.31. Western blot analysis of rare codon optimised DHFR clone optimised for protein expression probed with the DHFR antibody. Lane 1 – PAGE Ruler Plus protein ladder, Lane 2 – blank, Lane 3 – 1 h induced soluble, Lane 4 – blank, Lane 5 – 1 h induced insoluble, Lane 6 – blank, Lane 7 – 1 h un-induced soluble, Lane 8 – blank, Lane 9 – 1 h un-induced insoluble, Lane 10 – blank, Lane 11 – DHFR control (150 ng). Cultures were grown at 30°C for 1 h. A band of the correct size is present at ~ 48 kDa, indicating that our protein of interest is in the soluble fraction.

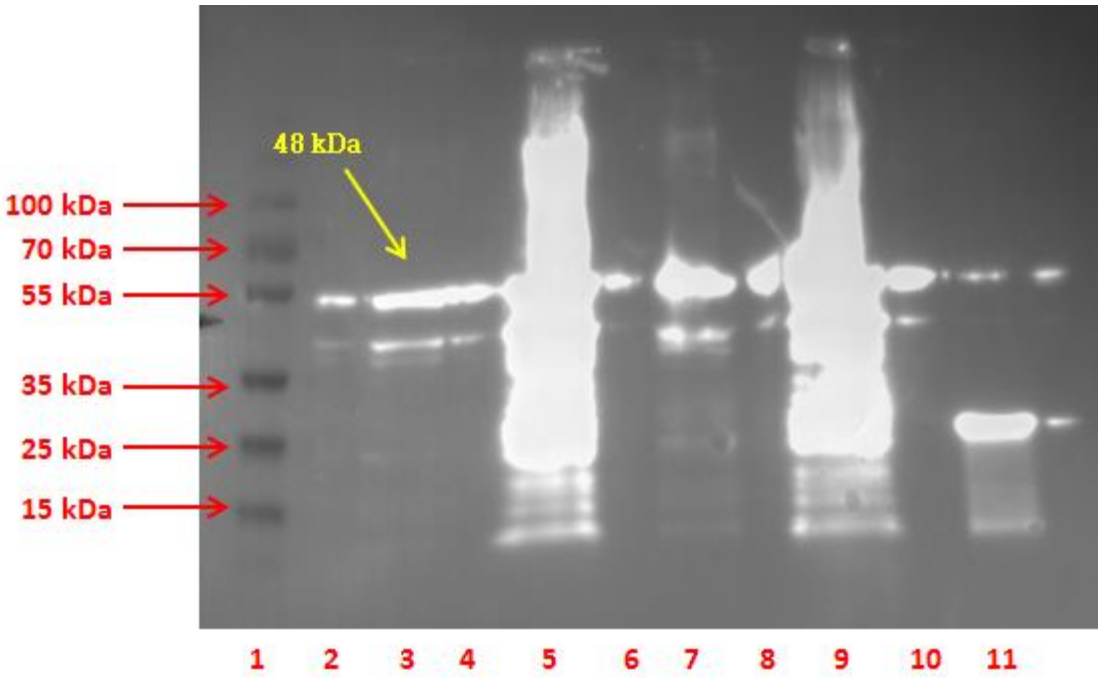


Figure 6.32. Western blot analysis of rare codon optimised DHFRL1 clone optimised for protein expression probed with the DHFR antibody. Lane 1 – PAGE Ruler Plus protein ladder, Lane 2 – blank, Lane 3 – 1 h induced soluble, Lane 4 – blank, Lane 5 – 1 h induced insoluble, Lane 6 – blank, Lane 7 – 1 h un-induced soluble, Lane 8 – blank, Lane 9 – 1 h un-induced insoluble, Lane 10 – blank, Lane 11 – DHFR control (150 ng). Cultures were grown at 30°C for 1 h. A band of the correct size is present at ~ 48 kDa, indicating that our protein of interest is in the soluble fraction.

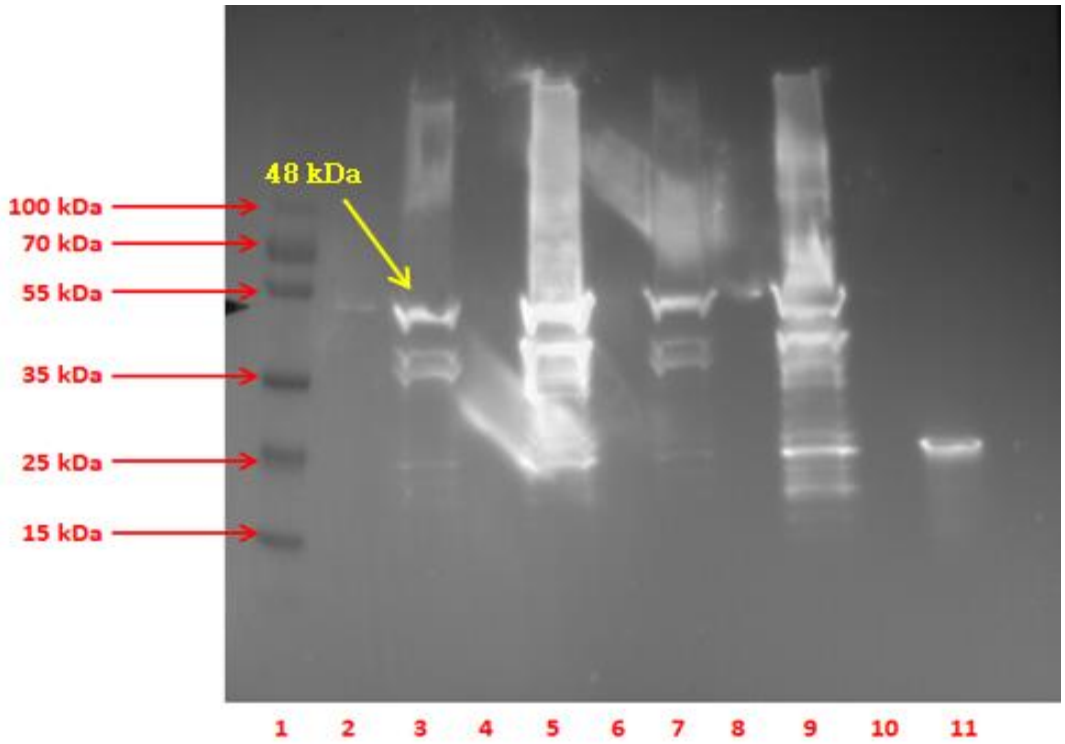


Figure 6.33. Western blot analysis of the GST tagged DHFRL1 rs61739170 polymorphism clone optimised for protein expression probed with the DHFR antibody. Lane 1 – PAGE Ruler Plus protein ladder, Lane 2 – blank, Lane 3 – 4 h induced soluble, Lane 4 – blank, Lane 5 – 4 h induced insoluble, Lane 6 – blank, Lane 7 – 4 h un-induced soluble, Lane 8 – blank, Lane 9 – 4 h un-induced insoluble, Lane 10 – blank, Lane 11 – DHFR control (150 ng). Cultures were grown at 30°C for 4 h. A band of the correct size is present at ~ 48 kDa, indicating that our protein of interest is in the soluble fraction.

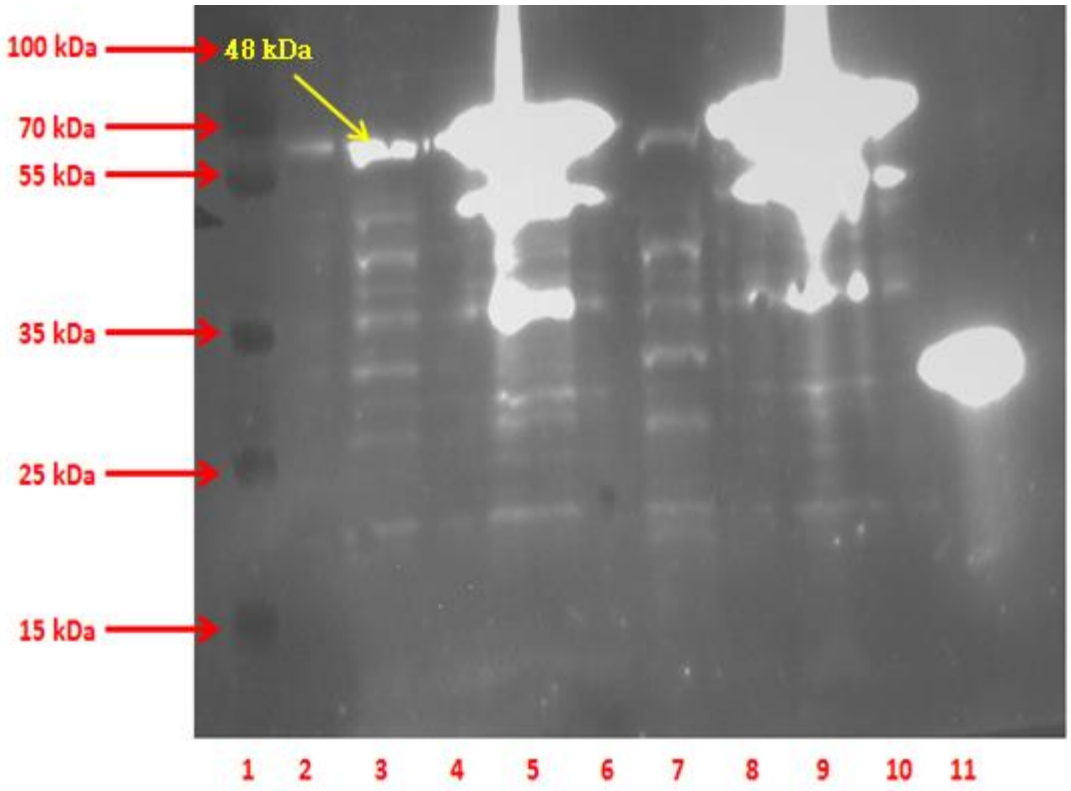


Figure 6.34. Western blot analysis of the GST tagged DHFRL1 rs17855824 polymorphism clone optimised for protein expression probed with the DHFR antibody. Lane 1 – PAGE Ruler Plus protein ladder, Lane 2 – blank, Lane 3 – 4 h induced soluble, Lane 4 – blank, Lane 5 – 4 h induced insoluble, Lane 6 – blank, Lane 7 – 4 h un-induced soluble, Lane 8 – blank, Lane 9 – 4 h un-induced insoluble, Lane 10 – blank, Lane 11 – DHFR control (150 ng). Cultures were grown at 30°C for 4 h. A band of the correct size is present at ~ 48 kDa, indicating that our protein of interest is in the soluble fraction.

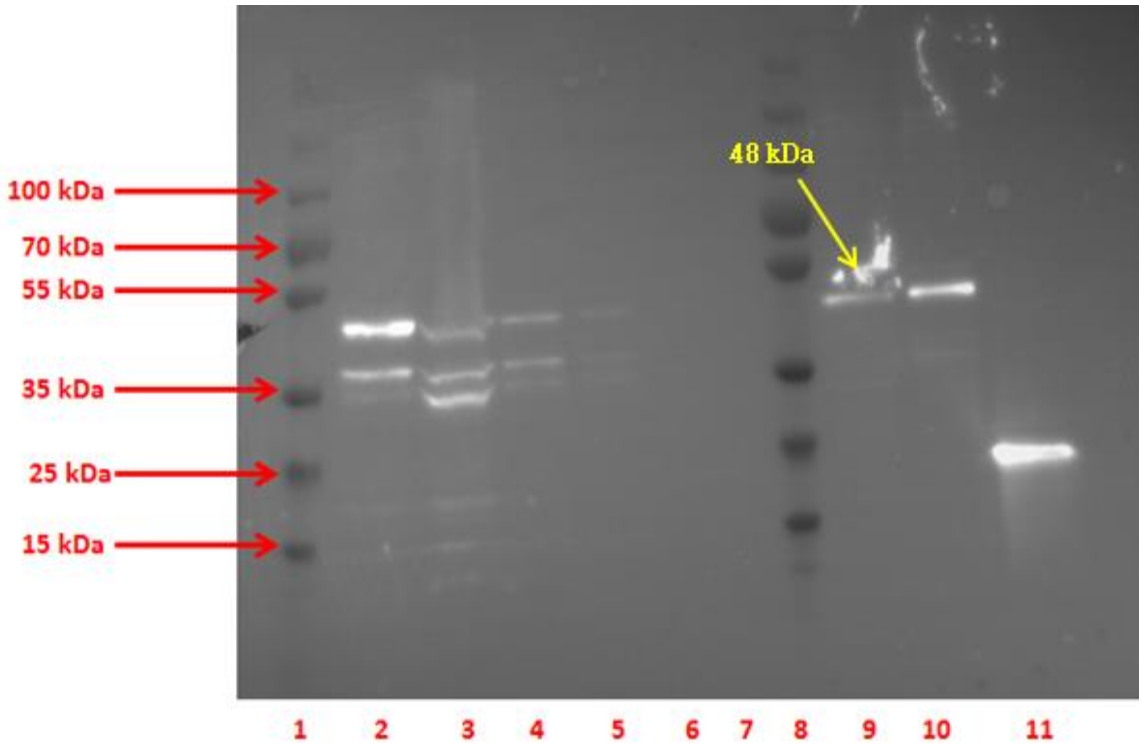


Figure 6.35. Western blot analysis of the DHFRL1 protein purification fractions probed with the DHFR antibody. Lane 1 – PAGE Ruler Plus protein ladder, Lane 2 – DHFRL1 optimised soluble fraction, Lane 3 - DHFRL1 optimised concentrated soluble fraction, Lane 4 – DHFRL1 optimised clarified soluble fraction, Lane 5 - DHFRL1 optimised flow through, Lane 6 – DHFRL1 optimised wash, Lane 7 – blank, Lane 8 – PAGE Ruler Plus protein ladder, Lane 9 – DHFRL1 optimised concentrated elution 1, Lane 10 – DHFRL1 optimised concentrated elution 2, Lane 11 – DHFR control. Bands of the correct size (48 kDa) were seen in the elution fraction, indicating that the GST-tagged fusion protein is present.

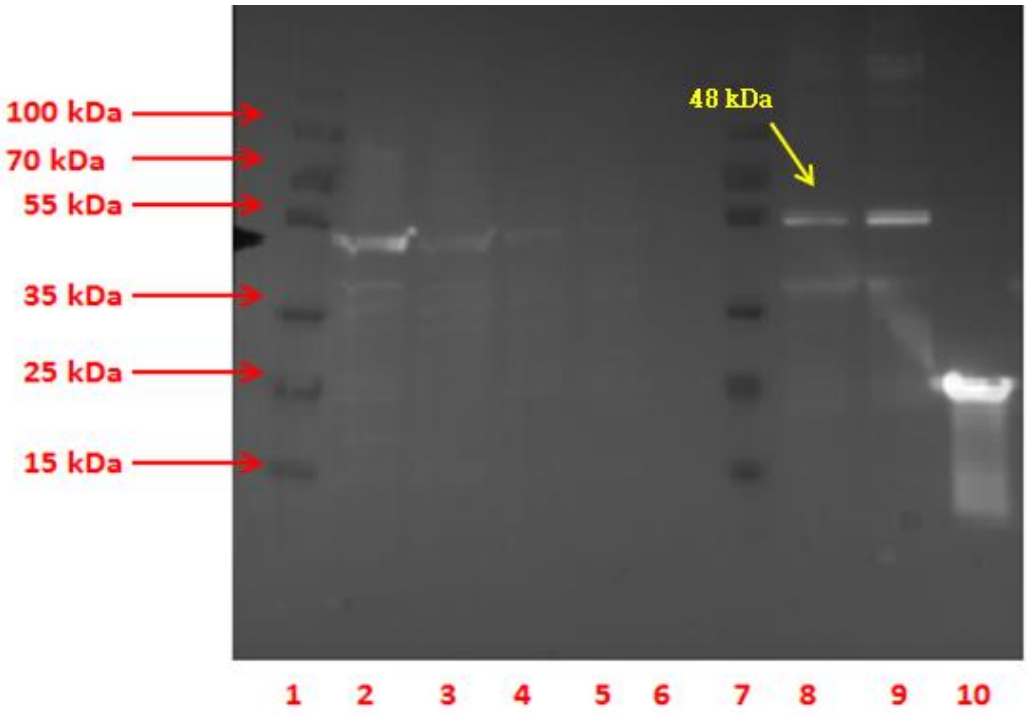


Figure 6.36. Western blot analysis of the DHFRL1 rs61739170 polymorphism protein purification fractions probed with the DHFR antibody. Lane 1 – PAGE Ruler Plus protein ladder, Lane 2 – DHFRL1 rs61739170 optimised soluble fraction, Lane 3 - DHFRL1 rs61739170 optimised concentrated soluble fraction, Lane 4 – DHFRL1 rs61739170 optimised clarified soluble fraction, Lane 5 - DHFRL1 rs61739170 optimised flow through, Lane 6 – DHFRL1 rs61739170 optimised wash, Lane 7 – blank, Lane 8 – PAGE Ruler Plus protein ladder, Lane 9 – DHFRL1 rs61739170 optimised concentrated elution 1, Lane 10 – DHFRL1 rs61739170 optimised concentrated elution 2, Lane 11 – DHFR control. Bands of the correct size (48 kDa) were seen in the elution fraction, indicating that the GST-tagged fusion protein is present.

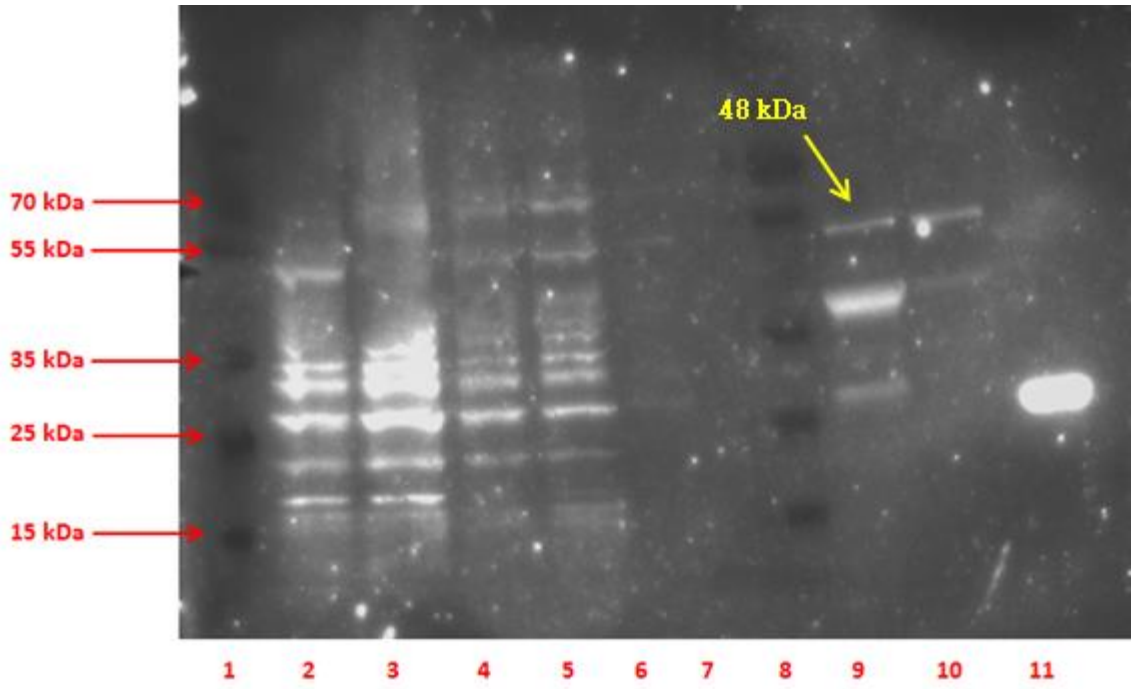


Figure 6.37. Western blot analysis of the DHFRL1 rs17855824 polymorphism protein purification fractions probed with the DHFR antibody. Lane 1 – PAGE Ruler Plus protein ladder, Lane 2 – DHFRL1 rs17855824 optimised soluble fraction, Lane 3 - DHFRL1 rs17855824 optimised concentrated soluble fraction, Lane 4 – DHFRL1 rs17855824 optimised clarified soluble fraction, Lane 5 - DHFRL1 rs17855824 optimised flow through, Lane 6 – DHFRL1 rs17855824 optimised wash, Lane 7 – blank, Lane 8 – PAGE Ruler Plus protein ladder, Lane 9 – DHFRL1 rs17855824 optimised concentrated elution 1, Lane 10 – DHFRL1 rs17855824 optimised concentrated elution 2, Lane 11 – DHFR control. Bands of the correct size (48 kDa) were seen in the elution fraction, indicating that the GST-tagged fusion protein is present.

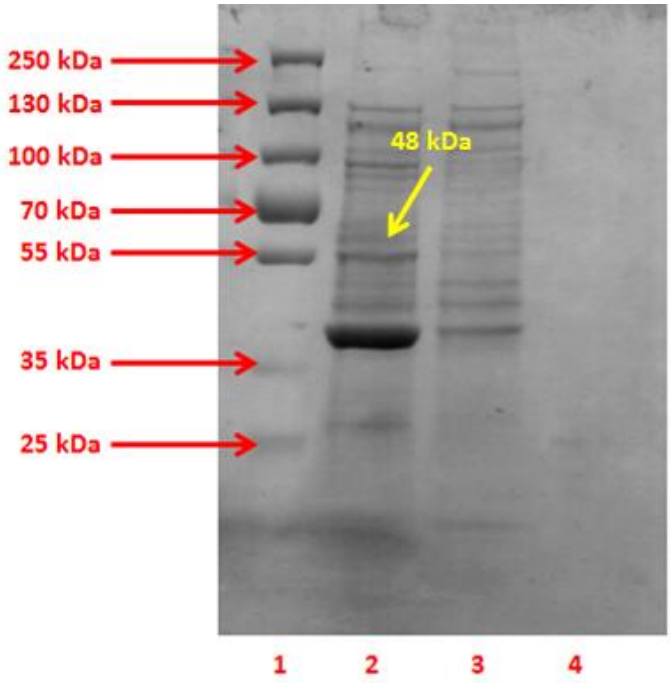


Figure 6.38. SDS-PAGE gel of the DHFRL1 rs17855824 polymorphism purified eluted protein fractions. Lane 1 – PAGE Ruler Plus protein ladder, Lane 2 - DHFRL1 rs17855824 optimised concentrated elution 1, Lane 3 – DHFRL1 rs17855824 optimised concentrated elution 2, Lane 4 – DHFR control. A strong glutathione band, our band of interest (48 kDa) as well as other bands were observed on the Coomassie stained SDS-PAGE gel.

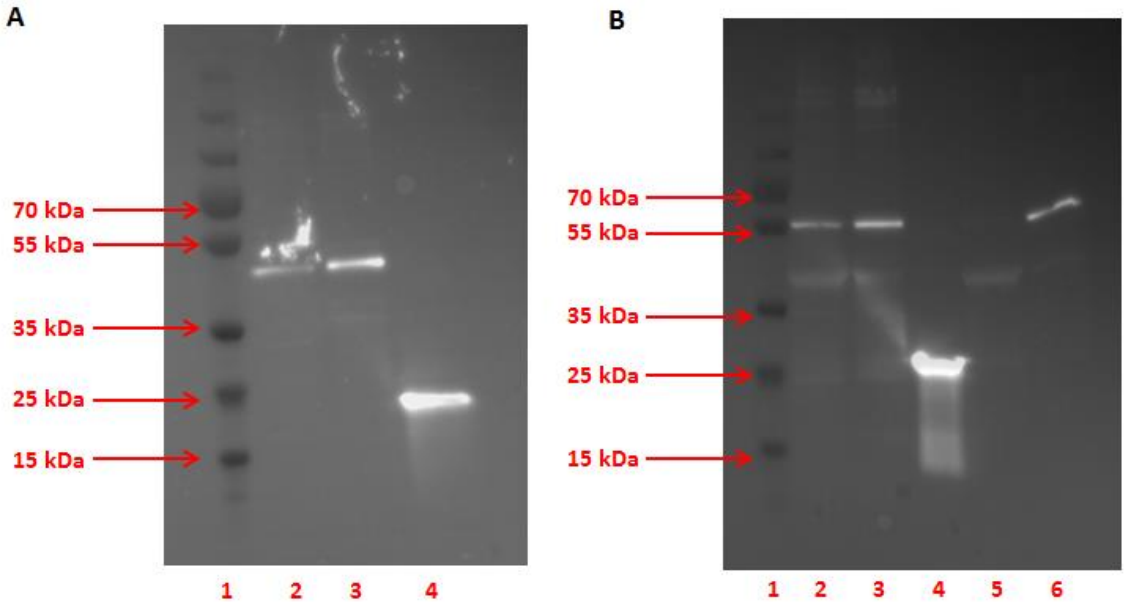
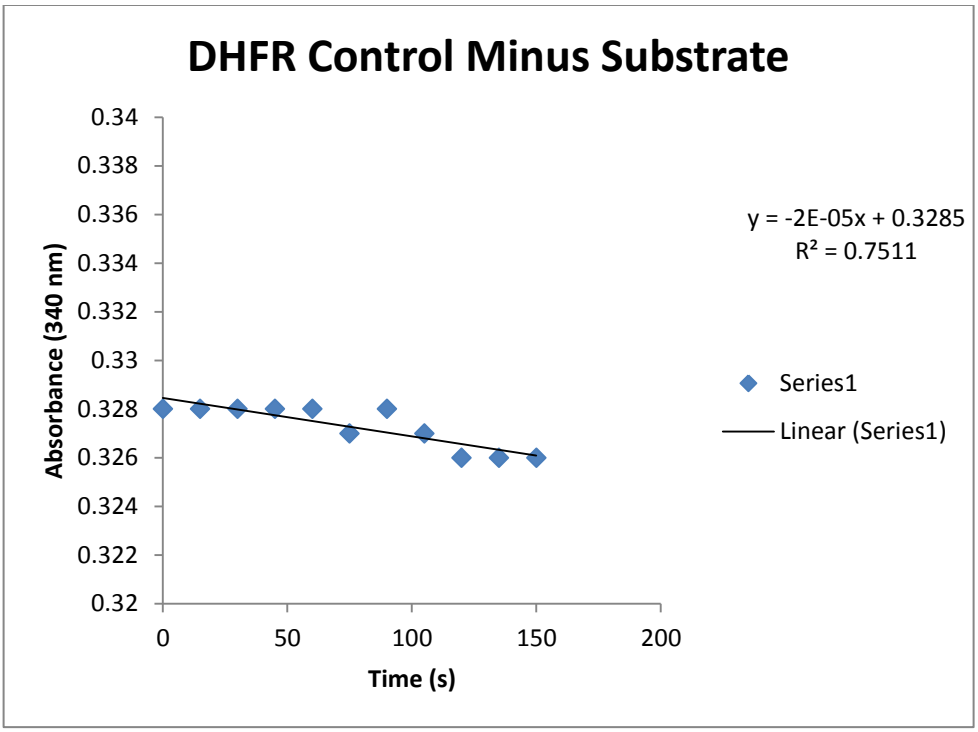
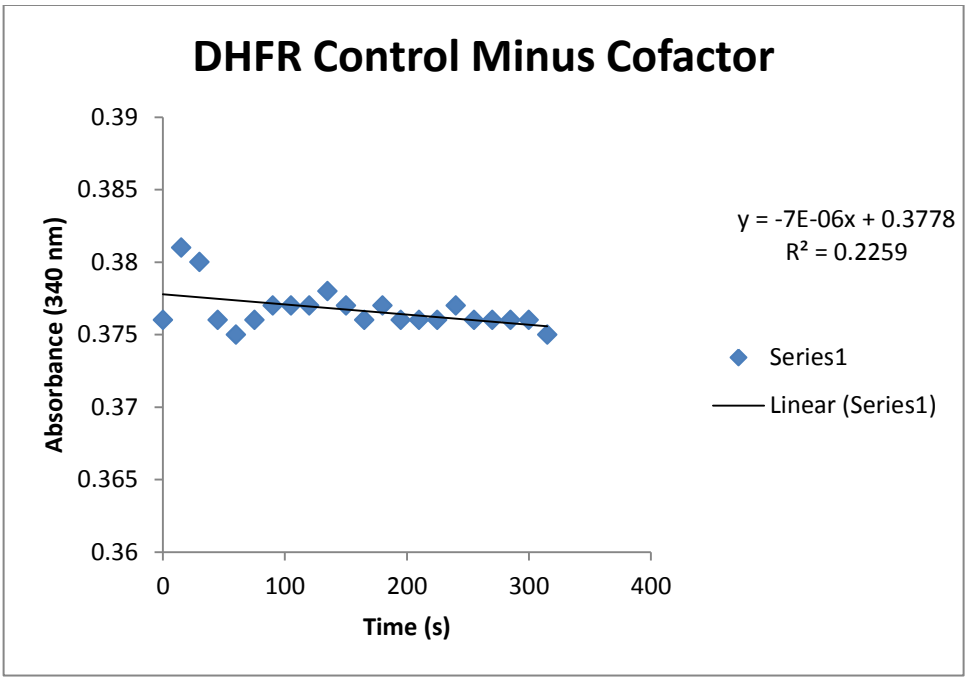


Figure 6.39. Quantification of purified protein samples via Image J using Western blot analysis of the DHFRL1 and DHFRL1 polymorphism protein purification fractions probed with the DHFR antibody. Recombinant proteins were quantified by normalising against a known concentration of DHFR control. **(A)** Lane 1 – PAGE Ruler Plus protein ladder, Lane 2 – DHFRL1 optimised concentrated elution 1, Lane 3 – DHFRL1 optimised concentrated elution 2, Lane 4 – 176 ng DHFR control. **(B)** Lane 1 - PAGE Ruler Plus protein ladder, Lane 2 – DHFRL1 rs61739170 optimised concentrated elution 1, Lane 3 – DHFRL1 rs61739170 optimised concentrated elution 2, Lane 4 – 176 ng DHFR control, Lane 5 – DHFRL1 rs17855824 optimised concentrated elution 1, Lane 6 – DHFRL1 rs17855824 optimised concentrated elution 2.



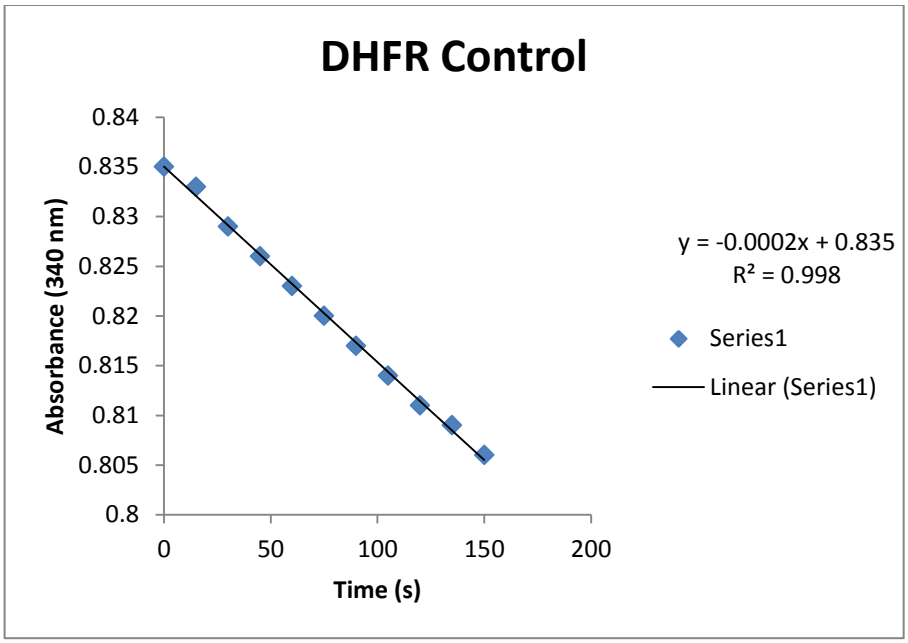
ΔOD/Min = 0.0008

Figure 6.40. DHFR Control without DHF (substrate). The graph above was obtained by plotting the 340 nm absorbance measurement vs. time in s. From the graph, it can be confirmed that there is no DHFR activity, indicated by the roughly straight horizontal line, i.e., when one of the key components of the reaction, DHF, which is the substrate for DHFR, is absent, the reaction cannot take place.



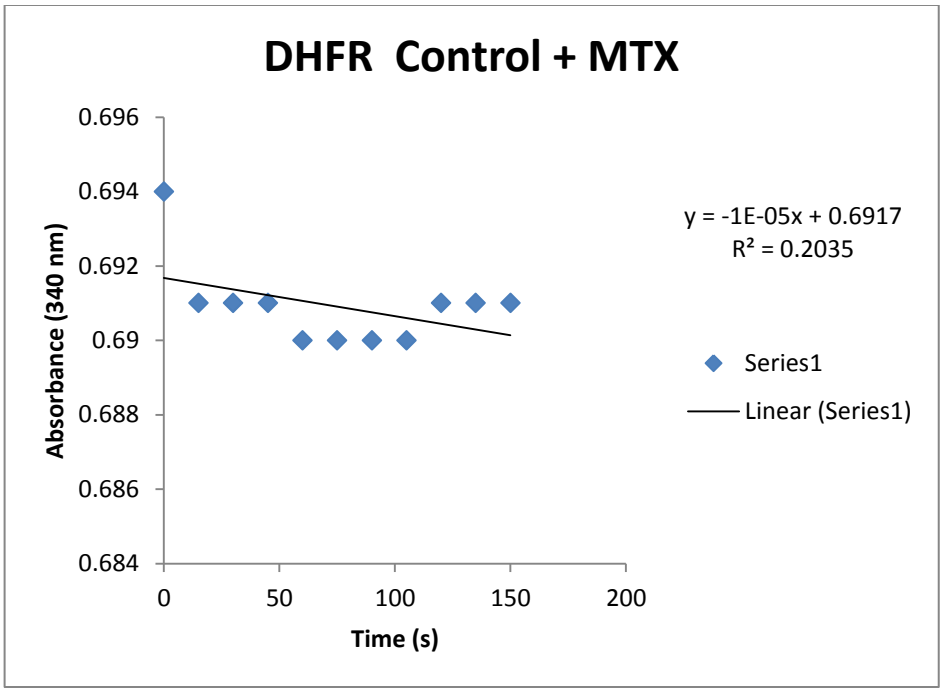
ΔOD/Min = 0.0002

Figure 6.41. DHFR control without NADPH (cofactor). The graph above was obtained by plotting the 340 nm absorbance measurement vs. time in s. From the graph, it can be confirmed that there is no DHFR activity, indicated by the straight horizontal line, i.e., when one of the key components of the reaction, NADPH, which is a cofactor for DHFR, is absent, the reaction cannot take place.



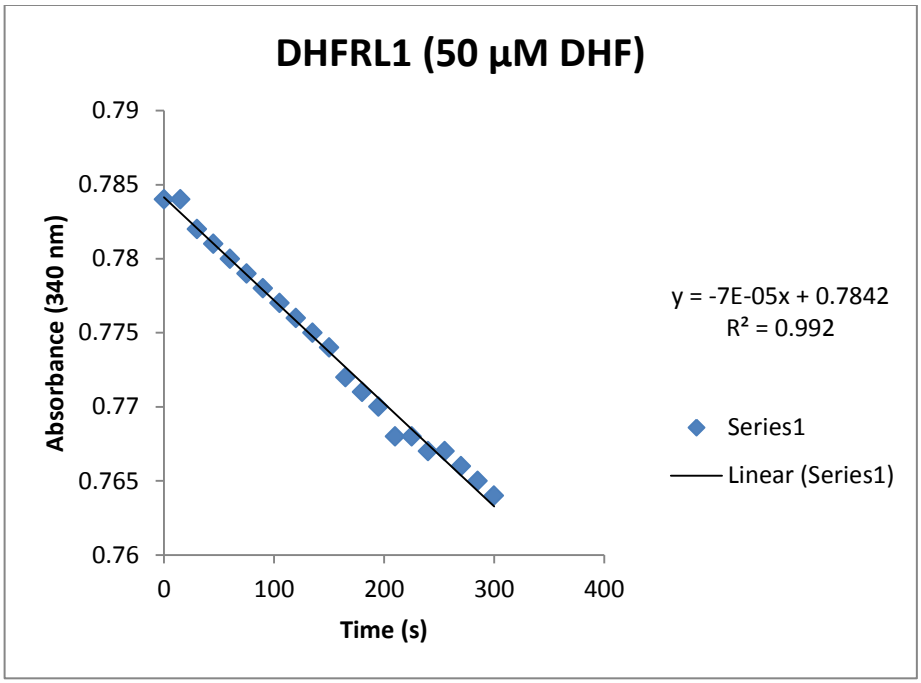
ΔOD/Min = 0.0116

Figure 6.42. DHFR Control. The graph above was obtained by plotting the 340 nm absorbance measurement vs. time in s. From the graph, it can be confirmed that once all the necessary components that are involved in the DHFR pathway are present, there is DHFR activity, indicated by the decrease in absorbance over time. The decrease in ΔOD over the 2.5 min was measured as ΔOD/min. An R^2 value of over 0.98 indicated good linearity of the assay and precision of the standard preparation.



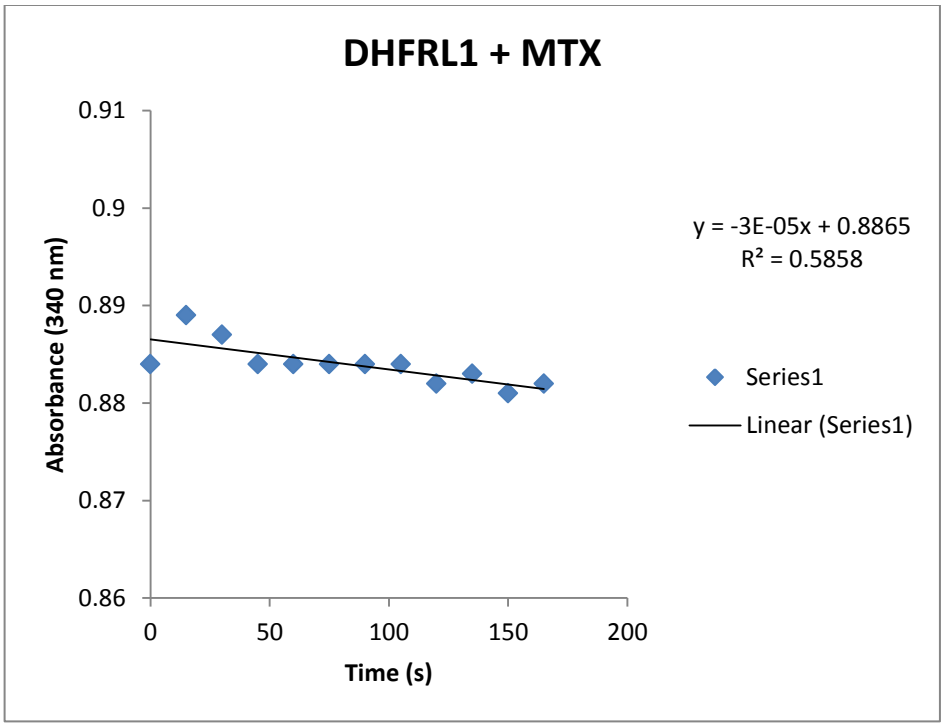
ΔOD/Min = 0.0012

Figure 6.43. DHFR control plus methotrexate. The graph above was obtained by plotting the 340 nm absorbance measurement vs. time in s. The graph above demonstrates that the change in NADPH absorbance that we observed in Figure 6.40 is due to reductase activity as the majority of activity is lost in the presence of the dihydrofolate reductase inhibitor, Methotrexate. From the graph, it can be confirmed that methotrexate (40 nM) is inhibiting DHFR activity and that the concentration of methotrexate used is high enough to inhibit all DHFR enzyme activity, indicated by the straight horizontal line.



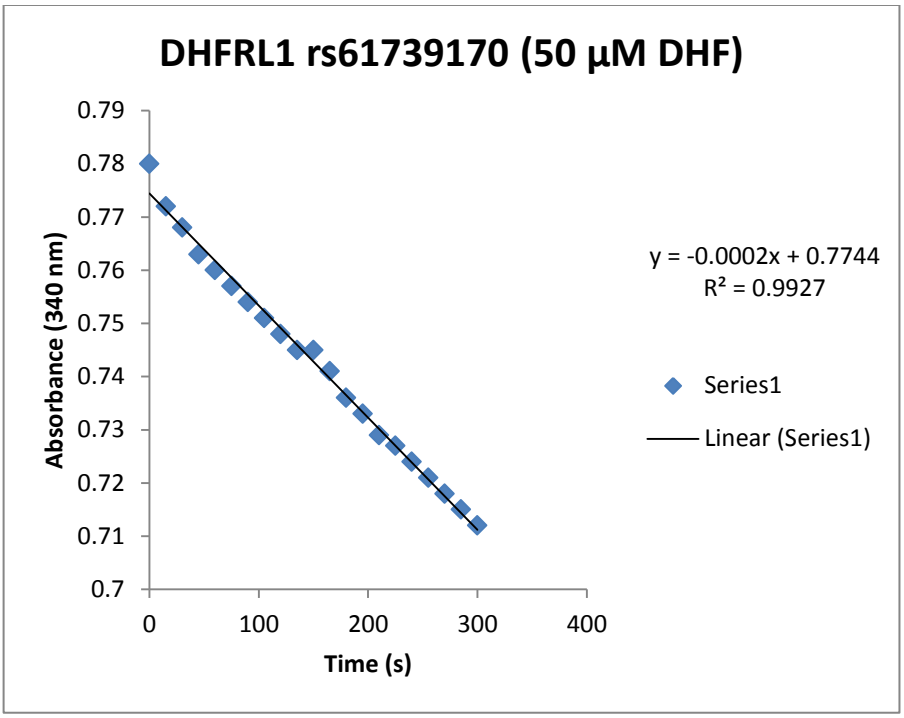
ΔOD/Min = 0.004

Figure 6.44. Purified DHFRL1 protein sample. The graph above was obtained by plotting the 340 nm absorbance measurement vs. time in s. From the graph, it can be confirmed that our purified DHFRL1 protein is active, indicated by the decrease in absorbance over time. The decrease in ΔOD over the 5 min was measured as ΔOD/min. An R^2 value of over 0.98 indicated good linearity of the assay and precision of the standard preparation.



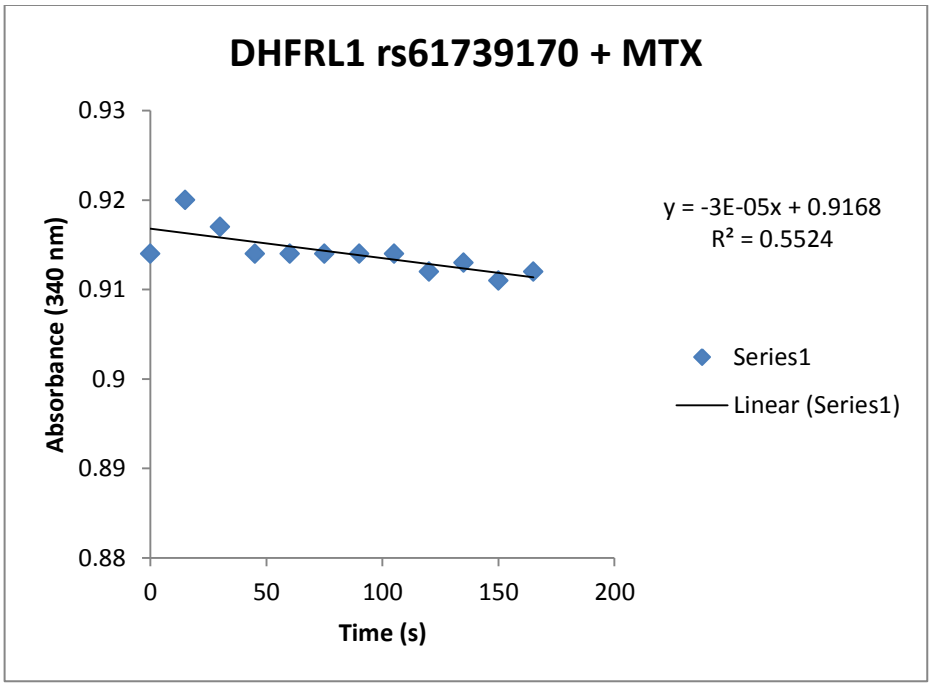
ΔOD/Min = 0.0007

Figure 6.45. Purified DHFRL1 protein sample plus Methotrexate. The graph above was obtained by plotting the 340 nm absorbance measurement vs. time in s. The graph above demonstrates that the change in NADPH absorbance that we observed in Figure 6.42 is due to reductase activity as the majority of activity is lost in the presence of the dihydrofolate reductase inhibitor, Methotrexate, which is also known to inhibit DHFRL1 (at a higher concentration) based on previous experiments carried out in the laboratory. From the graph, it can be confirmed that methotrexate (40 nM) is inhibiting DHFR activity and that the concentration of methotrexate used is high enough to inhibit all DHFR / DHFRL1 enzyme activity, indicated by the straight horizontal line.



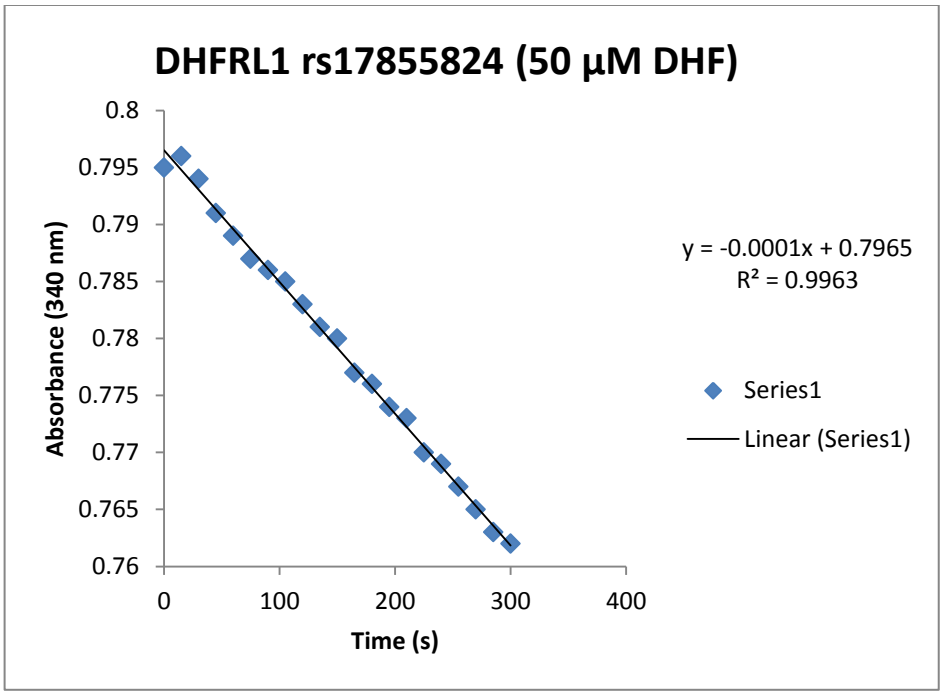
ΔOD/Min = 0.0136

Figure 6.46. Purified DHFRL1 rs61739170 protein sample. The graph above was obtained by plotting the 340 nm absorbance measurement vs. time in s. From the graph, it can be confirmed that our purified DHFRL1 rs61739170 protein is active, indicated by the decrease in absorbance over time. The decrease in ΔOD over the 5 min was measured as ΔOD/min. An R^2 value of over 0.98 indicated good linearity of the assay and precision of the standard preparation.



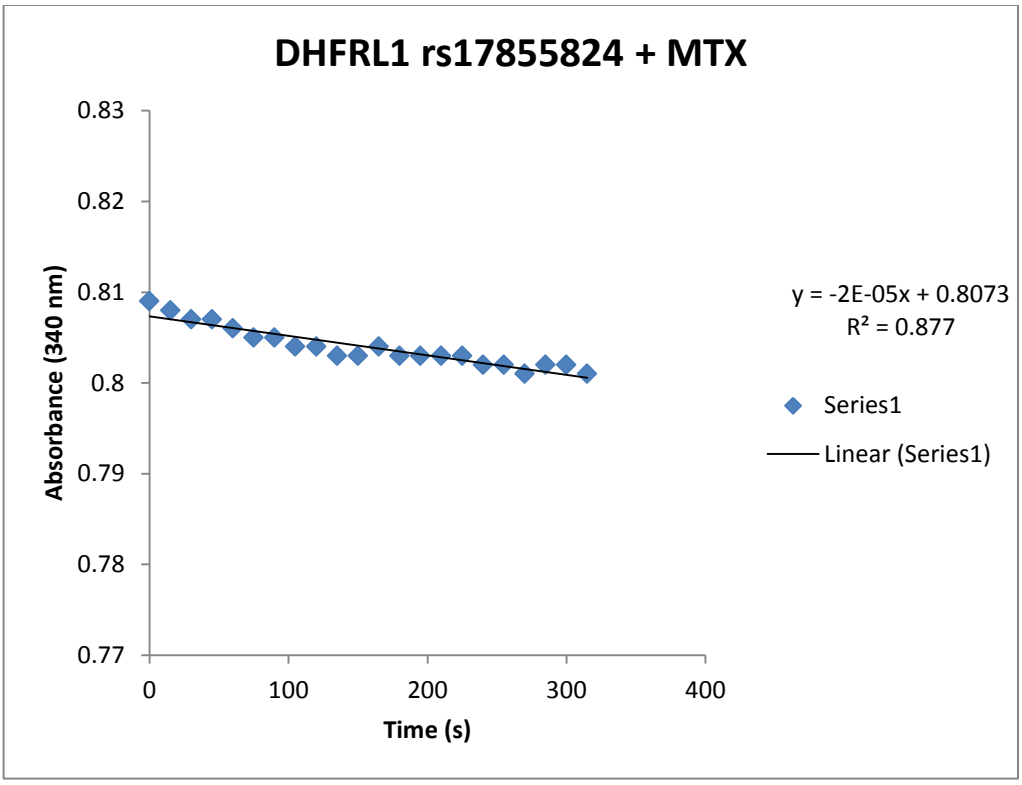
ΔOD/Min = 0.0007

Figure 6.47. Purified DHFRL1 rs61739170 protein sample plus Methotrexate. The graph above was obtained by plotting the 340 nm absorbance measurement vs. time in s. The graph above demonstrates that the change in NADPH absorbance that we observed in Figure 6.44 is due to reductase activity as the majority of activity is lost in the presence of the dihydrofolate reductase inhibitor, Methotrexate, which is also known to inhibit DHFRL1 (at a higher concentration) based on previous experiments carried out in the laboratory. From the graph, it can be confirmed that methotrexate (40 nM) is inhibiting DHFR activity and that the concentration of methotrexate used is high enough to inhibit all DHFR / DHFRL1 enzyme activity, indicated by the straight horizontal line.



ΔOD/Min = 0.0066

Figure 6.48. Purified DHFRL1 rs17855824 protein sample. The graph above was obtained by plotting the 340 nm absorbance measurement vs. time in s. From the graph, it can be confirmed that our purified DHFRL1 rs17855824 protein is active, indicated by the decrease in absorbance over time. The decrease in ΔOD over the 5 min was measured as ΔOD/min. An R^2 value of over 0.98 indicated good linearity of the assay and precision of the standard preparation.



ΔOD/Min = 0.001

Figure 6.49. Purified DHFRL1 rs17855824 protein sample plus Methotrexate. The graph above was obtained by plotting the 340 nm absorbance measurement vs. time in s. The graph above demonstrates that the change in NADPH absorbance that we observed in Figure 6.46 is due to reductase activity as the majority of activity is lost in the presence of the dihydrofolate reductase inhibitor, Methotrexate, which is also known to inhibit DHFRL1 (at a higher concentration) based on previous experiments carried out in the laboratory. From the graph, it can be confirmed that methotrexate (40 nM) is inhibiting DHFR activity and that the concentration of methotrexate used is high enough to inhibit all DHFR / DHFRL1 enzyme activity, indicated by the straight horizontal line.

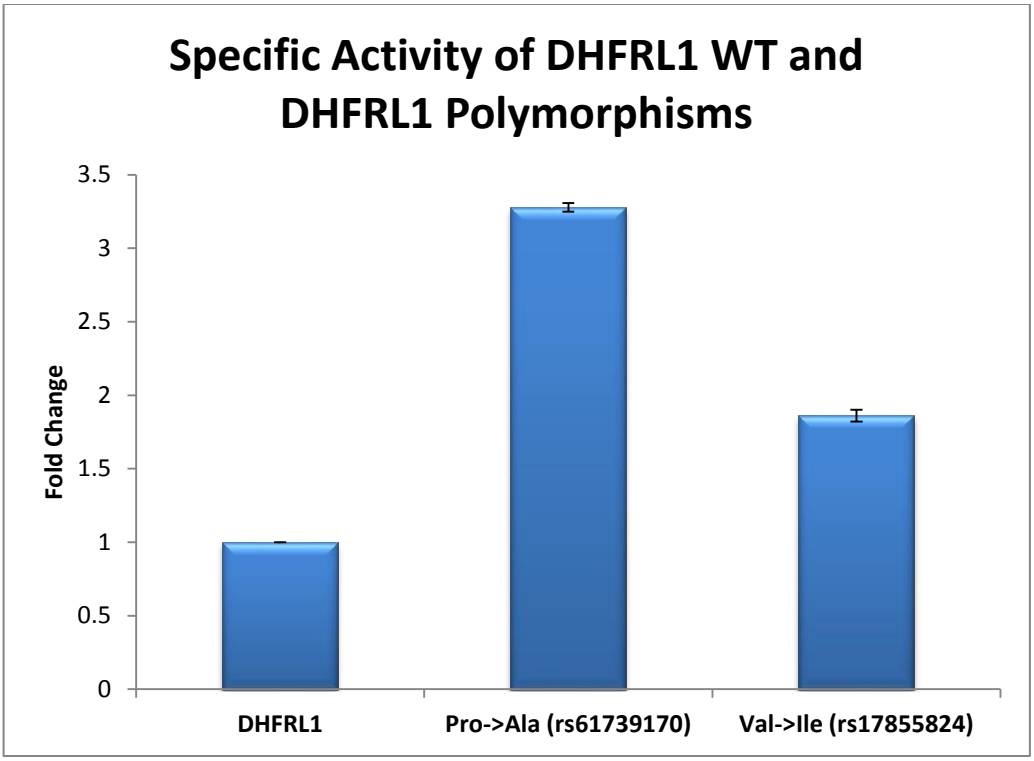


Figure 6.50. Specific activity of DHFRL1 and DHFRL1 polymorphisms. The bar chart above shows specific enzyme activity fold difference with respect to the DHFRL1 wild type. The DHFRL1 rs61739170 polymorphism has a 3.36 fold increase in comparison to the wild type and the DHFRL1 rs17855824 polymorphism has a 1.64 fold increase when compared to the DHFRL1.

DHFRL1 Protein Sequence (187 aa):
MFLLLNCIVAVSQNMGIGKNGDLPRPPLRNEFRYFQRMTTTSSVEGKQNLVIMGRKT WFSIPEKNRPLKDRINLVLSRELKEPPQGAHFLARSLDDALKLTERPELANKVDMIWI VGGSSVYKEAMNHLGHLKLFVTRIMQDFESDTFFSEIDLEKYKLLPEYPGVLSDVQEG KHIKYKFEVCEKDD

Figure 6.51. DHFRL1 Protein Sequence. The four motifs that are required for *DHFRL1* catalytic activity are highlighted in yellow. Specific residues important for folate, NADPH and/or MTX binding are underlined. Specific amino acids important for DHFR mRNA binding are shown in red. Image adapted from (McEntee et al., 2011). P shows the location of the *DHFRL1* rs61739170 polymorphism and the V shows the location of the *DHFRL1* rs17855824 polymorphism.

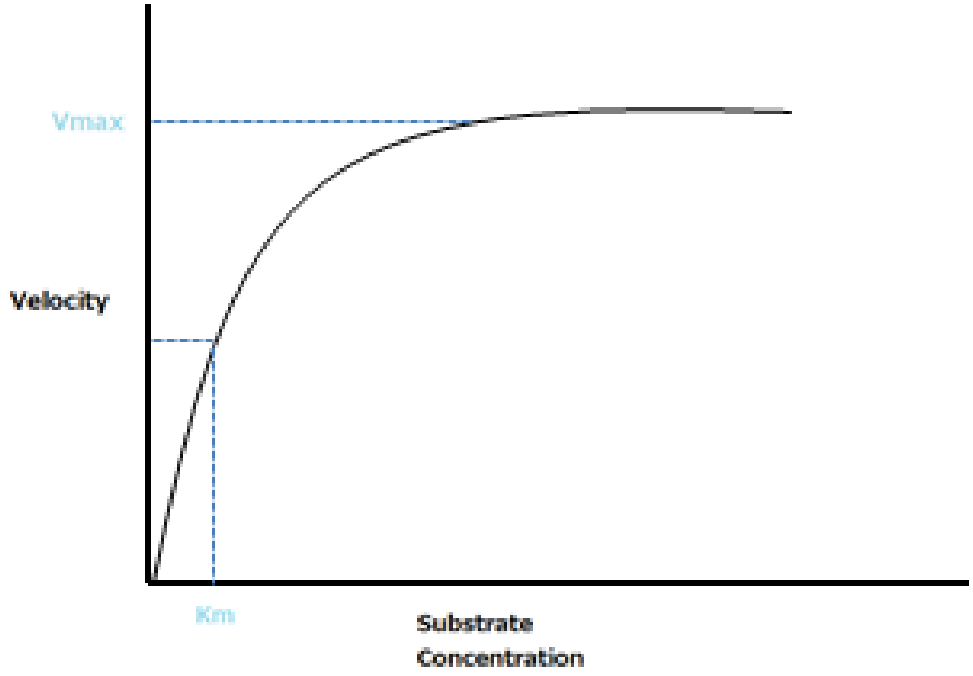


Figure 6.52. Overview of Enzyme Kinetics. Vmax is the maximum velocity of the enzyme and measures the amount of time it takes for the substrate to bind all the enzymes active sites. Km (Michaelis-Menten constant) measures the substrate concentration required to reach half the Vmax. The Km characterises the enzyme's affinity for the substrate.

Table 6.1. Purified Protein Quantification via Image J Software.

	Quantification for protein	Quantification for control (176 ng)	Fold Difference	Amount of protein (ng) loaded (18 µL)	Amount of protein (ng) in 1 µL
DHFRL1	**7084**	**15691**	**2.21**	**79.84**	**4.4**
rs61739170	**8689**	**24542**	**2.82**	**62.4**	**3.5**
rs17855824	**13928**	**24542**	**1.76**	**100**	**5.55**

Chapter 7

Proteomic Analysis of Altered DHFRL1 Expression

7.1 Introduction

The term proteomics generally refers to the large-scale analysis of proteins and it is often specifically used to refer to protein purification and mass spectrometry. Proteomic tools have been used to help us to extend our knowledge of disease mechanisms, by helping us to develop novel biomarkers for diagnosing disease, to identify new targets for therapeutics and to study altered protein expression (Hanash, 2003). *DHFRL1*, as mentioned in Section 1.12 is a newly discovered gene and therefore, studies on the gene have only recently begun in order to determine its functionality. Therefore, there is not much knowledge on how it may affect the expression of other genes or what effect it has on health and disease, though preliminary data and analysis obtained in Chapter 5 and Chapter 6, did show that polymorphisms within the coding region of *DHFRL1* are associated with NTDs and that they can alter the activity of the gene. Therefore the aim of this chapter is to investigate the impact of up-regulating and down-regulating DHFRL1 expression by proteomic analysis, namely LC-MS. This will hopefully allow us to have a better understanding of DHFRL1 and help us to determine whether DHFRL1 expression levels could have an impact on human health and disease.

7.1.1 Liquid Chromatography–Mass Spectrometry (LC-MS)

LC-MS is a powerful analytical technique that combines HPLC with MS. As the name suggests, the first part of LC-MS is HPLC. Reversed phase liquid chromatography is used for most cases of LC-MS, whereby the chemicals in the samples are separated by forcing the sample through a column that is packed with a hydrophilic solvent by a liquid at high pressure in order to allow the sample to bind to the column. The different metabolites are then eluted off the column by a more hydrophobic solvent such as methanol. Once the chemicals have been separated physically by RP-LC, the metabolites enter the mass detector, where MS takes place. MS is a technique that measures the mass-to-charge ratio of charged particles. Therefore, once the metabolites enter the mass detector, the solvents are removed and the metabolites are ionised. Solvents need to be removed as ions only fly through a very good vacuum. The mass detector then scans the molecules it sees by mass and produces a full high-resolution spectrum, separating all ions that have different masses (Pitt, 2009).

7.1.2 Objectives

In this chapter, the role of altered DHFRL1 expression was examined. Assessment of the impact of altered DHFRL1 expression levels were examined using the following methodology:

- Confirmation of DHFRL1 down-regulation in DHFRL1 down-regulated cells by RT-qPCR and Western blot.
- Generation of DHFRL1 over-expressing cell pools.
- Confirmation of DHFRL1 over-expression by RT-qPCR and Western blot.
- Examination of altered DHFRL1 expression by LC-MS.
- Proteomic Analysis of LC-MS data by Progenesis 3.1 Software.
- Pathway analysis of genes affected by altered DHFRL1 expression by the Ingenuity Pathway Analysis (IPA) software.

7.2 Results

7.2.1 Confirmation of DHFRL1 Down-Regulation

A stably transfected cell pool model of targeted DHFRL1 knock-downs (shRNA B1 and shRNA C1) in HEK 293 cells were previously generated in the lab by Dr. Stefano Minguzzi, using the MISSION^{TM} TRC lentiviral shRNA vectors. As severe down-regulation may have a detrimental effect on the cells, inducible IPTG vectors were used in order to allow a certain control on the down-regulation of DHFRL1. Five shRNA constructs (Table 7.1) were originally tested with various concentrations of IPTG and the two shRNAs that showed the best down-regulation were selected, i.e. shRNA B1 and shRNA C1 for the down-regulation experiments. First of all, in order to confirm down-regulation of these cell pools before proteomic analysis was carried out, the cells were cultured as described in Section 2.6.3 and Section 2.6.4 and RNA was extracted from the DHFRL1 down-regulated cells as well as those from parental HEK 293 cells and HEK 293 shRNA control cell lines as described in Section 2.6.9 and run on an agarose gel in order to ensure that the RNA was intact (Figure 7.1). RNA was then quantified and cDNA was synthesised as described in Section 2.6.13

Once cDNA was synthesised, down-regulation by the two shRNA constructs was confirmed at the mRNA level by RT-qPCR as described in Section 2.6.15 (Figure 7.2 and Figure 7.3). Figure 7.2 shows that when the DHFRL1 down-regulated samples are normalised to the non-targeting shRNA control or to the normal parental HEK 293 cells, they do not show

a lower expression of DHFRL1 as expected, i.e., the induced shRNA cells showed a similar level of DHFRL1 expression to the shRNA control and the parental HEK 293 cells. The un-induced cells appeared to have a higher level of DHFRL1 expression compared to the other cells and therefore, it was decided to compare our IPTG induced samples to our un-induced samples (Figure 7.3). Figure 7.3 shows that when the IPTG induced DHFRL1 down-regulated samples are normalised to their un-induced controls, both DHFRL1 shRNA B1 and shRNA C1 cells are down-regulated by 24 % and 57 % respectively.

Mitochondrial protein fractions were also isolated from the cells as described in Section 2.3.15 and Western blot analysis was carried in order to determine DHFRL1 down-regulation at the protein level. However Western blot failed to confirm DHFRL1 down-regulation at the protein level, with only very subtle changes seen between the induced and un-induced samples. Quantification of DHFRL1 at the protein level however, has been complicated by recent findings in the Parle-McDermott lab. Up until now, DHFR has been considered to be entirely cytoplasmic. However, the Parle-McDermott lab have recently shown that both DHFR and DHFRL1 fused GFP proteins localise to the mitochondria by Western blotting and confocal microscopy, though whether the GFP is driving the localisation is yet to be determined. Therefore, quantification of DHFRL1 at the protein level is complicated by the likely cross-reactivity of the antibody between DHFRL1 and its paralogue DHFR as DHFRL1 and DHFR differ only in a total of 16 amino acids. An antibody that can specifically distinguish between the two proteins is difficult to produce.

7.2.2 Generation of DHFRL1 Over-expressing Cell Lines

DHFRL1 was cloned into the mammalian expression vector pcDNA3.2/V5-DEST *via* Gateway cloning as described in Section 2.2.8 and plasmid DNA was isolated and sequenced to ensure the correct sequence. Once the correct sequence was confirmed, HEK 293 cells were transfected with pcDNA^{TM}3.2-*DHFRL1* and pcDNA^{TM}3.2/V5-DEST in a 6-well plate as described in Section 2.6.8 The *DHFRL1* HEK 293 over-expressed cells attached and grew on the plates. However, the pcDNA^{TM}3.2/V5-DEST (empty vector) was not compatible with the cells and the cells died after transfection and were therefore abandoned. After 48 h of transfection, the medium was replaced with DMEM containing 500 µg/mL G418 antibiotic and the resistant cells were selected for two weeks. The stably transfected cells were then cultured for a further three passages before stocks were made / before they were used in experiments.

Un-transfected HEK 293 cells were also treated with DMEM with 500 µg/mL G418 antibiotic. All these cells died within ten days.

7.2.3 Confirmation of DHFRL1 Up-Regulation

The *DHFRL1* over-expressed and the HEK 293 cells were selected for the proteomics experiment. First of all, in order to confirm up-regulation of the DHFRL1 over-expressing cell pool before proteomic analysis was carried out, cells were cultured as described in Section 2.6.3.2 and Section 2.6.4.2 and RNA was extracted from the DHFRL1 over-expressed cells as well as those from the HEK 293 control cells as described in Section 2.6.9 and the RNA samples were run on an agarose gel (Figure 7.4). RNA was then quantified and cDNA was synthesised as described in Section 2.6.13.

Once cDNA was synthesised, DHFRL1 over-expression was confirmed at the mRNA level by RT-qPCR as described in Section 2.6.15 (Figure 7.5). Figure 7.5 shows that when the DHFRL1 over-expressed samples are normalised to the HEK 293 control cells, the DHFRL1 over-expressed cells are up-regulated by ~ 5.5 fold.

Mitochondrial protein fractions were also isolated from the DHFRL1 over-expressed HEK 293 and HEK 293 cells as described in Section 2.3.15 and Western blot analysis was carried out as described in Section 2.3.11 to determine the over-expression of DHFRL1 at the protein level. Western blots were probed with both the DHFR and PDH antibody in order to 1) confirm the presence of DHFRL1 in the mitochondria and 2) so that the DHFRL1 expression levels could be determined. Western blots were analysed using the Image J software, which confirmed a 1.5 fold increase in DHFRL1 expression at the protein level (Figure 7.6). However, as mentioned earlier (Section 7.2.1), the localisation of DHFR to the mitochondria complicates our protein quantification. Therefore, quantification of DHFRL1 via Image J may not be highly accurate, and the utilisation of Image J at the level of a 1.5 fold difference may not be considered truly quantitative.

7.2.4 Sample Preparation of Cell Lysates for Proteomic analysis

Once DHFRL1 down-regulation and over-expression was confirmed by RT-qPCR / Western blot, the DHFRL1 HEK 293 knock-down cells, the DHFRL1 HEK 293 over-expressed cells and the control cells were set up for proteomic analysis. For the DHFRL1 HEK 293 down-

regulated cells, six flasks were seeded with 2 x 10^6 cells. Half of the flasks were induced with 500 µM IPTG. All six flasks were incubated at 37ºC with CO_2 5 %, humidified air 95 % and the cells were left to grow for five days. For the DHFRL1 HEK 293 over-expressed and the HEK 293 control cells, three flasks were seeded with 2 x 10^6 cells. The flasks were incubated at 37ºC with CO_2 5 %, humidified air 95 % and the cells were left to grow for five days. After five days, cell pellets were collected from all flasks, were washed with 1 mL of sterile PBS five times and were then stored at -80°C until they were processed for proteomic analysis.

7.2.5 Proteomic analysis on the effects of altering DHFRL1 expression

DHFRL1 down-regulated and over-expressed whole cell lysates were analysed by LC-MS by our collaborators Dr. Paula Meleady and Mr. Michael Henry from the National Institute for Cellular Biology (NICB) in DCU. Label-free LC-MS analysis was carried out using the Progenesis label-free LC-MS software version 3.1 (NonLinear Dynamics), as recommended by the manufacturer (see www.nonlinear.com for more information on alignment, normalisation, calculation of peptide abundance, etc.). The raw data was processed by the software in two main steps. First of all, each sample run was subjected to an alignment, which involves aligning the data based on the LC retention time of each of the samples, compensating for any between-run variation. The sample run that yielded the most features (i.e. peptide ions) was used as the reference run, to which retention of all of the other runs were aligned and peak intensities were normalised. The Progenesis peptide quantification algorithm calculates the abundance of the peptides as the sum of the peak area within its isotope boundaries. Each abundance value is then transformed to a normalised abundance value by applying a global scaling factor. Protein abundance was calculated as the sum of the abundances of all the peptide ions, which have been identified as coming from the same protein. A number of criteria were used to filter the data before exporting the MS/MS output files to MASCOT (www.matrixscience.com) for protein identification; peptide features with ANOVA < 0.05 between experimental groups, mass peaks with charge states from +1 to +3, and greater than three isotopes per peptide. All MS/MS spectra were exported from the Progenesis software as a MASCOT generic file (mgf) and used for peptide identification with MASCOT (version 2.2) searched against the UniProKB-SwissProt database (taxonomy, Mammals). The search parameters that were used are as follows: peptide mass tolerance set to 20 ppm, MS/MS mass tolerance set at 0.5 Da; up to two missed cleavages allowed, carbamidomethylation set as a fixed modification and

methionine oxidation set as a variable modification. Only peptides with ion scores of 30 and above were considered and re-imported back into the Progenesis LC-MS software for further analysis. An ANOVA p-value of ≤ 0.05 between experimental groups was used as a cut-off.

7.2.5.1 Proteomic Analysis on the DHFRL1 Down-Regulation Experiment

LC-MS data analysed with the Progenesis 3.1 software reported a large number of proteins to be up-regulated and down-regulated when DHFRL1 was knocked-down. A total of 310 significant proteins were found for the DHFRL1 shRNA B1 construct (Table 7.2) and 24 significant proteins for the DHFRL1 shRNA C1 construct (Table 7.3). Network analysis was performed on these proteins using the IPA software (Qiagen). Network analysis identified many of these proteins to be associated with cell morphology and embryonic development (Figure 7.7). Protein analysis of the data set also identified a large number of the proteins to be associated with nucleic acid metabolism, cell cycle, DNA replication, recombination and repair and cell death and survival (Table 7.4) as well as with glycolysis, oxidative phosphorylation, electron transport, ATP synthesis and the malate / aspartate shuttle (Table 7.9).

However, when the data sets from the two different DHFRL1 shRNA constructs were combined and analysed together, the majority of the proteins proved not to be consistent between the two sample sets. This could be due to the fact that the two shRNA constructs down-regulated DHFRL1 to different extents or it may be due to off-target effects of shRNA regulation. When both data sets were analysed together, a total of six proteins were found to be present in both sets. However, of these, only three were found to have similar effects. The three proteins that were seen to alter protein expression in the same direction in both sets of samples were Serpin H1 (SERPINH1), which was found to be up-regulated, heterogeneous nuclear ribonucleoprotein K (HNRNPK), which was found to be down-regulated and dihydrolipoyl dehydrogenase, mitochondria (DLD), which was also found to be down-regulated (Table 7.5). Although the XRCC5, the RPL8 and the RPL18a proteins were also found in both samples sets they had opposing effects and were therefore removed.

7.2.5.2 Proteomic Analysis on the DHFRL1 Over-Expression Experiment

LC-MS data analysed with the Progenesis 3.1 software reported a large number of proteins to be up-regulated and down-regulated when DHFRL1 was over-expressed. The software reported a total of 79 significant proteins of which 48 were up-regulated (Table 7.6)

and 31 were down-regulated (Table 7.7). Protein function and network analysis was performed on these proteins using the IPA software. The proteins which appeared in our dataset were analysed and matched for different networks in which they are involved / associated in. The two networks generated, which included the most number of our proteins are shown in Figure 7.8 and Figure 7.9. The figures show that DHFRL1 over-expression has an impact on proteins that are associated with cell proliferation, cancer, cell cycle, cell death and survival and DNA replication, recombination and repair. Using the IPA software, proteins which are of interest to us, i.e. those involved in DNA replication, recombination and repair, nucleic acid metabolism, cell death and survival, cell proliferation and cell cycle (Table 7.8) as well as those involved in glycolysis, oxidative phosphorylation, electron transport, ATP synthesis and the malate / aspartate shuttle (Table 7.9) could also be identified.

7.2.5.3 Proteomic Analysis on the DHFRL1 Down-Regulation and Over-Expression Experiment Analysed Together

When the proteomics data from both the DHFRL1 down-regulation and over-expression experiments were examined together, numerous proteins were found to be present in both data sets, with some proteins having opposing effects, while others having similar effects (Table 7.10).

7.3 Discussion

The aim of this chapter was to analyse the effects of down-regulating and up-regulating DHFRL1 in HEK 293 cells in order to determine the effects that altered DHFRL1 expression has on the expression of other proteins by proteomic analysis, namely LC-MS, a method which is being highly used in order to produce databases in order to help us broaden and strengthen our knowledge of disease mechanisms. This is of great interest to us as, as mentioned earlier (Section 1.12), *DHFRL1* is a newly discovered member of the DHFR gene family and plays an important role in *de novo* thymidylate biosynthesis in the mitochondria along with *SHMT2* and *TYMS* (Anderson et al., 2011). This would suggest that like *DHFR*, *DHFRL1* plays an important role in DNA replication and repair / cell proliferation. In order for us to examine what effects DHFRL1 has on other proteins, we first had to generate stable cell pools, in which, DHFRL1 was either down-regulated or over-expressed. These were generated by transfection of HEK 293 cells (Section 2.6.8).

In the case of our DHFRL1 down-regulated stable cell pools, inducible shRNA constructs were used in order to down-regulate DHFRL1. As DHFRL1 is involved in DNA replication and repair, knocking-out the gene completely may have had severe consequences on cell stability. Therefore, it was important not to completely knock-out DHFRL1 but to down-regulate it enough so that substantial proteomic changes could be observed. When we examined the down-regulation of the DHFRL1 down-regulated cells, we found that when we normalised the DHFRL1 down-regulated samples to the non-targeting shRNA control or to the normal HEK 293 cells, the down-regulated cells were not down-regulated, with the non-targeting shRNA control and the normal HEK 293 cells having similar / lower DHFRL1 gene expression in comparison to both induced and un-induced DHFRL1 down-regulated shRNA constructs (Figure 7.2). However, when we normalised the IPTG induced down-regulated samples against the un-induced samples, down-regulation was observed with the DHFRL1 shRNA B1 samples down-regulated by 24 % and the shRNA C1 samples down-regulated by 57 % (Figure 7.3). Although this model is not 100 % ideal, as the insertion of the shRNA constructs appears to increase the level of DHFRL1 expression, as the shRNA constructs have been transfected into both induced and un-induced samples, we are still examining like with like with the only difference being the induction of the DHFRL1 shRNA so that it is being down-regulated in one of the samples and not the other. Therefore, we can say that an artificial DHFRL1 down-regulated system has been generated that will allow us to gather preliminary data on the relevance of DHFRL1 expression level for global cellular functions. The increase in DHFRL1 expression levels observed by the insertion of the shRNA constructs may be due to the fact that DHFRL1 is involved in DNA synthesis as mentioned above. The insertion of the shRNA construct may cause damage to the cells and therefore, DHFRL1 expression may be up-regulated as a result in order to repair the damage caused; an interesting observation in itself. In order to try and avoid the change in DHFRL1 expression levels after transfection, one could try transfecting the HEK 293 cells with shRNA constructs in different vectors so as to try and obtain one that does not alter the expression levels of DHFRL1 or switch to a completely different system such as the newly developed genome editing approaches as discussed in Section 5.3. This will have to be carried out in the future in order to confirm our findings that we obtained from our preliminary analysis.

For our *DHFRL1* over-expressed experiments, the over-expressed cells (pcDNA^{TM}3.2-*DHFRL1*) were normalised against HEK 293 cells. Although it would have been more ideal to

normalise our DHFRL1 over-expressed cells with HEK 293 cells in which the empty vector (pcDNA^{TM}3.2/V5-DEST) had been transfected into, as mentioned in Section 7.2.2, the empty vector was not compatible with the cells and therefore, HEK 293 cells were used as our control. The reason why the HEK 293 cells transfected with the empty vector died may be due to the fact that the empty vector contains the ccdB gene, a toxic gene for bacteria (See Appendix AJ for vector map), whose product may also inhibit mammalian cell division (Hiraga et al., 1985). In order to try and resolve this problem and in order to confirm our preliminary findings, it may be possible to remove the ccdB gene from the vector using a restriction enzyme approach in the future.

Despite the fact that altered DHFRL1 expression could not be determined conclusively at the protein level by Western blot analysis as mentioned in Section7.2.1 and 7.2.3, due to the possible confounding localisation of DHFR to the mitochondria, as altered DHFRL1 was confirmed by RT-qPCR, whole cell lysates were collected and analysed by LC-MS.

Once LC-MS data was obtained, analysis was first carried out with the Progenesis 3.1 software for both DHFRL1 down-regulated and over-expressed protein samples, in order to apply cut offs of $p<0.05$. After careful analysis of the data, we found numerous genes that are involved in DNA replication, recombination and repair, cell death and survival, cell cycle, cell proliferation and nucleic acid metabolism to be affected when DHFRL1 was both down-regulated (Table 7.4) and over-expressed (Table 7.8). This is not surprising as, as mentioned earlier, *DHFRL1* plays a major role in DNA replication and repair (Anderson et al., 2011). Although many of the effects that we observed were minor, i.e. only a small up-regulation or down-regulation, many of the proteins that were observed function in the same pathway or in pathways which are interconnected with each other, and therefore the combinational effect of these proteins together may have a large impact on pathway function. These minor effects however, may also be random effects and therefore, further analysis is required in order to confirm these findings. A number of proteins, however, such as PARP1, an enzyme involved in the base excision repair pathway (Beck et al., 2014), TRIM28, a gene known as the master regulator of the genome, involved in mediating gene silencing, cell growth, apoptosis and DNA repair (Iyengar and Farnham, 2011), MYH9, an ATP hydrolysis site, DHX9, an enzyme that catalyses the ATP-dependent unwinding of double-stranded DNA and RNA (Chakraborty and Grosse, 2011) and XRCC6, an enzyme involved in non-homologous end joining required for

the repair of double-strand breaks (Ramsden and Gellert, 1998) were seen to be highly down-regulated by over-expressing DHFRL1 and proteins such as SEPT9, HINT1 and SOD1, all of which play a role in cell growth or cell death were found to be up-regulated when DHFRL1 was down-regulated. A protein of great interest to us, serine hydroxymethyltransferase, mitochondrial (SHMT2), a protein that is involved in the *de novo* thymidylate biosynthesis pathway, as well as the inter-conversion of serine and glycine (Anderson et al., 2011) was also found to be up-regulated when DHFRL1 was over-expressed and down-regulated when DHFRL1 was knocked-down (Table 7.10). It was also interesting to see that the majority of proteins that were affected by both knocking-down and over-expressing DHFRL1 were involved in DNA replication, recombination and repair, cell death and survival and the cell cycle (Table 7.10).

As we know, *DHFR* plays a critical role in the folate mediated one-carbon metabolism pathway, playing an important role in the regulation of the amount of THF in the cell (Figure 1.1). THF is essential DNA synthesis, which in turn is required for cell proliferation and cell growth. As folate / THF are required by rapidly dividing cells, such as cancer cells, *DHFR* has been targeted in the treatment of cancer (Schweitzer et al., 1990). As *DHFRLI* also regulates the level of THF in the cell, inhibiting / targeting *DHFRL1* with drugs, such as a modified version of methotrexate could prevent / reduce the growth and proliferation of cells that are characteristic of cancer. As *DHFRL1* takes the place of *DHFR* in the mitochondria, it is likely that *DHFRL1* plays a similar role to *DHFR*. Therefore, it is not surprising that a large number of genes that were altered after over-expressing DHFRL1 are involved in cancer (Figure 7.9), with many of the genes involved in apoptosis in cancer being down-regulated (DHX9, GPI, MYH9, PARP1, XRCC6, YWHAG, PRKDC and RPA1) and those involved in proliferation / nucleotide biosynthesis / replication being up-regulated (ALDOA1, FUBP1, GNB2L1, BSG, IMPDH2, CSNK2B, PDIA3, SHMT2, and GAPDH) in response to DHFRL1 over-expression.

From the findings obtained in the previous chapters, *DHFRL1* is also thought to be associated with NTDs, birth defects that occur during the first trimester of pregnancy. Therefore, it was highly interesting to observe numerous proteins that are involved in cell morphology and embryonic development to be altered when DHFRL1 was down-regulated (Figure 7.7). Examination of the proteins that were identified to be involved in embryonic development showed that the majority of the genes that were shown to be down-regulated are

involved in growth regulation (STIM1, PA2G4, ITGB1, PHB, SMG1, SLC3A2, BSG and EIF2S1). These proteins, in association with those identified to be involved in DNA replication, recombination and repair, nucleic acid metabolism and cell death and survival further support our theory that *DHFRL1* is associated with NTDs.

It has recently been shown that folate metabolism also plays a role in NADPH production in both the cytoplasm and the mitochondria (Fan et al., 2014). Tracing of one-carbon metabolism by LC-MS and carbon-labelling by Fan *et al*., (2014) showed that the oxidation of methylene-THF to 10-formyl-THF is coupled with the reduction of $NADP^+$ to NADPH. In order to confirm this finding, they also knocked-down the MTHFD genes, in both the cytoplasm and the mitochondria. Depletion of either of the MTHFD isoenzymes resulted in the reduction of cellular NADPH / $NADP^+$, indicating a clear connection between the folate pathway and the NADPH pathways. Therefore, it was interesting to see that a number of proteins that are involved in glycolysis, oxidative phosphorylation, electron transport, ATP synthesis and the malate / aspartate shuttle to be affected by both down-regulating and over-expressing DHFRL1 (Table 7.9).

Finally, when the two DHFRL1 down-regulated shRNA construct data was analysed together, as mentioned in Section 7.2.5.1, we found three genes, SERPINH1, HNRNPK and DLD to be affected (Table 7.5). Of these three genes, the DLD gene, which was observed to be down-regulated, is of interest to us. As mentioned in Section 1.3, one of the functions of folate metabolism in the mitochondria is the inter-conversion of glycine and serine (Fox and Stover, 2008) (Figure 1.1). The DLD gene plays a major role in the glycine cleavage system (GCS) (Kikuchi et al., 2008) as well as a vital role in energy metabolism (Vaubel et al., 2011). The GCS is a reversible system where a series of enzymes are triggered in response to high glycine concentrations in order to catalyse the degradation of glycine (Kikuchi et al., 2008) and is made up of three enzymes, the P protein (GLDC), the T protein (AMT) and the L protein (DLD) as well as a carrier protein, the H protein. These proteins form an enzyme complex that is loosely associated with the mitochondrial inner membrane in order to break down glycine (Kikuchi et al., 2008). Glycine catabolism involves the oxidative cleavage of glycine to CO_2, NH_4^+ and CH_2, a methylene group, which is accepted by THF to form 5,10-methylene-THF, which is a one-carbon donor for the biosynthesis of a number of cellular molecules such as purines, thymidylate and methionine (Figure 1.1). Therefore, it was highly interesting that when

DHFRL1 was down-regulated, that the DLD protein was also down-regulated. As DHFRL1 is known to be involved in thymidylate biosynthesis in the mitochondria, down-regulation of this gene would suggest a reduction at the rate in which thymidylate is synthesised, and therefore, a reduced need for one-carbon donors. As a result, there would be less of a requirement for glycine degradation to occur and therefore, it is not surprising that the DLD protein is down-regulated when DHFRL1 is down-regulated. However, how the mechanism of loss of expression of DLD in response to DHFRL1 down-regulation is open to be elucidated. Why DHFRL1 down-regulation is affecting the other two proteins is not clear but the HNRNPK protein is thought to play a role in cell cycle progression by acting as a transcriptional co-activator of p53 for cell-cycle arrest (Lee et al., 2012) as well as apoptosis in the mitochondria (Marchenko et al., 2000). In response to apoptosis inducing stress signals, p53 is known to translocate to the mitochondria, where it interacts with anti- and pro-apoptotic Bcl-two family members to either inhibit or activate apoptosis, resulting in the permeabilisation of the outer mitochondrial membrane, releasing enzymes such as caspases that are involved in apoptosis (Marchenko et al., 2000). These findings are of interest as it has recently been shown that the enzymes involved in nuclear thymidylate biosynthesis form a complex and translocate to the nuclear lamina in a cell cycle dependent manner, with translocation occurring during the S and G_2/M phase of the cell cycle (Anderson et al., 2012b), with the complex also shown to be enriched at sites of DNA replication and repair, suggesting that thymidylate biosynthesis occurs at the sites of DNA synthesis (Anderson et al., 2012b).

All in all, these preliminary novel findings above indicate that down-regulating or over-expressing DHFRL1 can have an impact on altering the expression levels of a numerous number of proteins, particularly those that are involved in DNA replication, recombination and repair, cancer and embryonic development as well as those involved in NADPH pathways. We have pointed out the connection between DHFRL1 and many of the proteins identified, however, the relationship between DHFRL1 and the other proteins remain to be elucidated. Although these novel findings will provide a basis for future work that will be carried out in this area, there were a number of issues associated with both generating the DHFRL1 down-regulated and over-expressed stable cell pools and confirming the altered DHFRL1 expression levels. Therefore, the data obtained from our proteomic analysis will have to repeated / carried out and confirmed by another system.

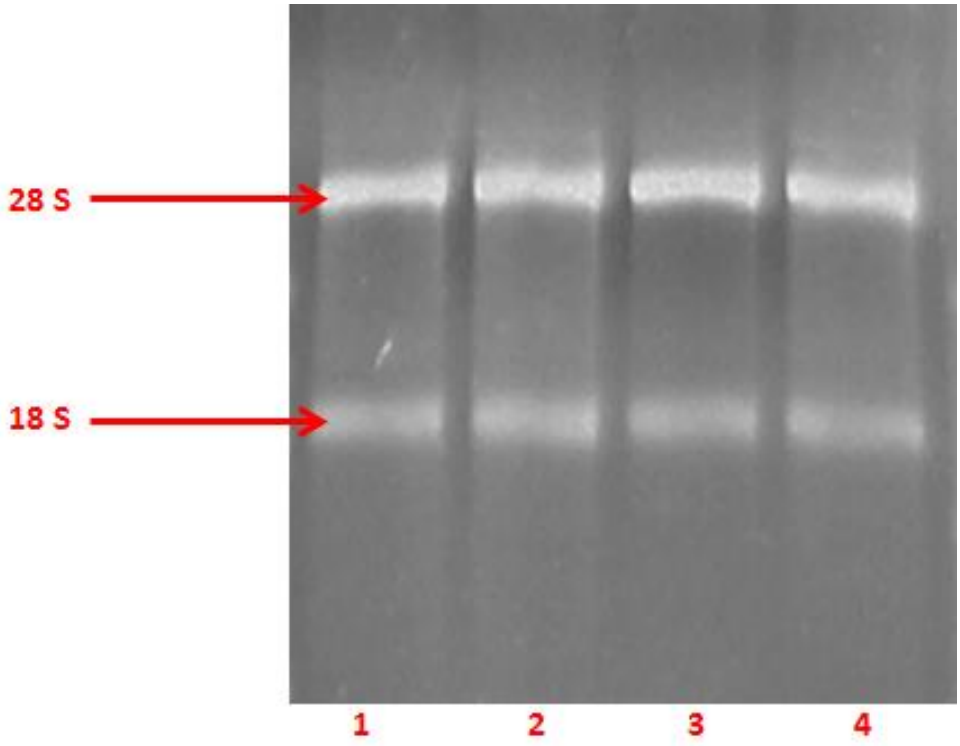


Figure 7.1. Gel Electrophoresis of extracted RNA on a 1 % agarose gel. Lane 1 – HEK 293 DHFRL1 knock down un-induced RNA, Lane 2 – HEK 293 DHFRL1 knock down induced RNA, Lane 3 – HEK 293 shRNA 1-1 control RNA, Lane 4 – HEK 293 RNA. A 1 µg concentration of RNA was loaded into each well. Sharp, clear 28S and 18S rRNA bands can be seen for all samples, with the 28S rRNA band having a stronger intensity that the 18S rRNA band, indicating that the RNA is intact.

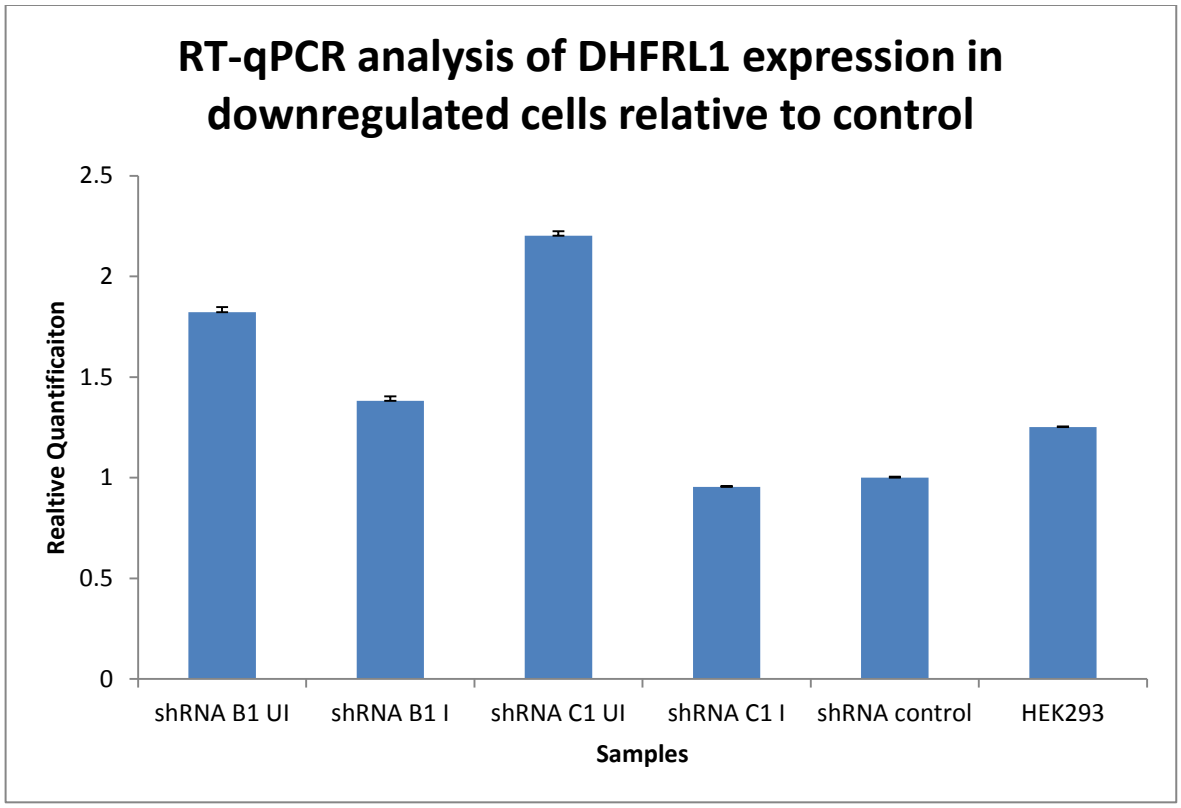


Figure 7.2. Validation of DHFRL1 down-regulation by RT-qPCR in HEK 293 stable transfected cell pools. DHFRL1 gene expression was analysed in two cell pools that differed by shRNA used to target DHFRL1, i.e., shRNA B1 and shRNA C1, at the mRNA level by RT-qPCR to confirm its down-regulation. Relative DHFRL1 gene expression was analysed and normalised to cDNA generated from the control cell line, i.e., cells transfected with a non-specific shRNA control. Relative expression was calculated using GUS as the reference gene. The bar chat shows that when the DHFRL1 down-regulated samples are normalised to the non-targeting shRNA control or to the normal parental HEK 293 cells, they do not show a lower expression of DHFRL1 as expected, i.e., the induced shRNA cells showed a similar level of expression of DHFRL1 compared to the shRNA control line and the parental HEK 293 line. UI = un-induced, I = induced

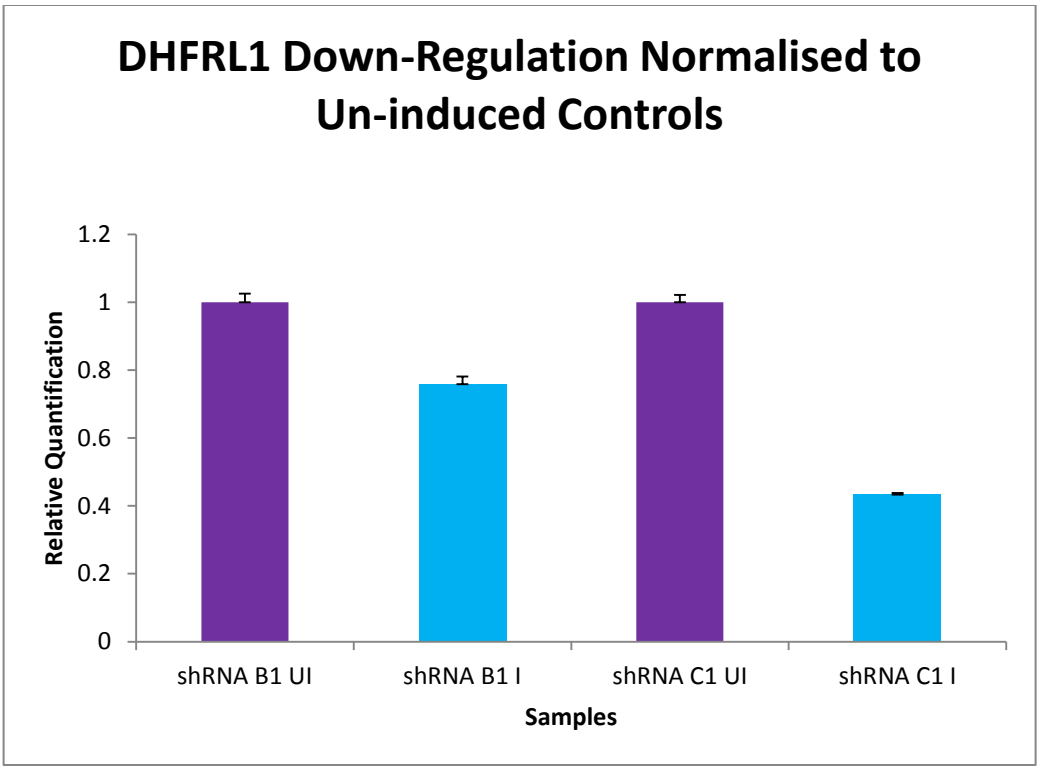


Figure 7.3. Validation of DHFRL1 down-regulation by RT-qPCR in HEK 293 stable transfected cell pools. DHFRL1 gene expression was analysed in two cell pools that differed by shRNA used to target DHFRL1, i.e., shRNA B1 and shRNA C1, at the mRNA level by RT-qPCR to confirm its down-regulation. Relative DHFRL1 gene expression regulation was analysed and confirmed by comparing the un-induced (UI) (negative control) samples to the induced (I) samples, normalising to the un-induced DHFRL1 samples. The B1 cells show a down-regulation of 24 % and the C1 cells show a down-regulation of 57 %. Relative expression was calculated using GUS as the reference gene. UI = un-induced, I = induced

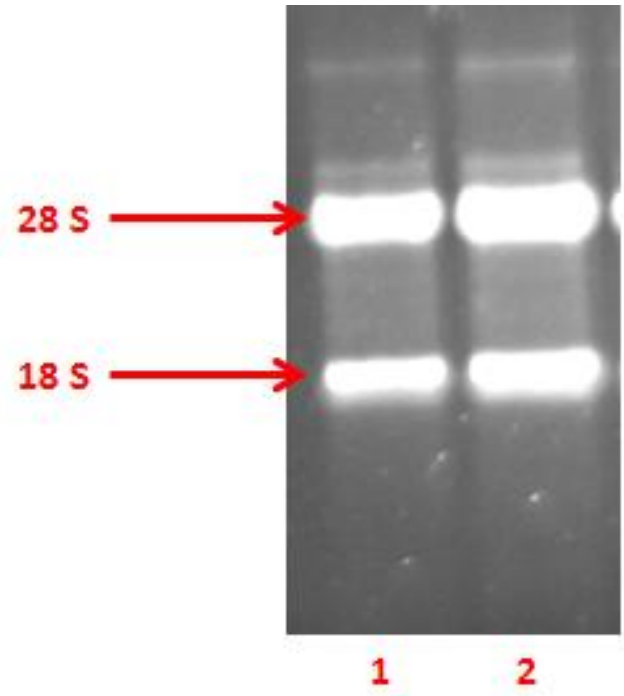


Figure 7.4. Gel Electrophoresis of extracted RNA on a 1 % agarose gel. Lane 1 – HEK 293 DHFRL1 over-expressed RNA, Lane 2 – HEK 293 control RNA. A 1 µg concentration of RNA was loaded into each well. Sharp, clear 28S and 18S rRNA bands can be seen for all samples, with the 28S rRNA band having a stronger intensity that the 18S rRNA band, indicating that the RNA is intact.

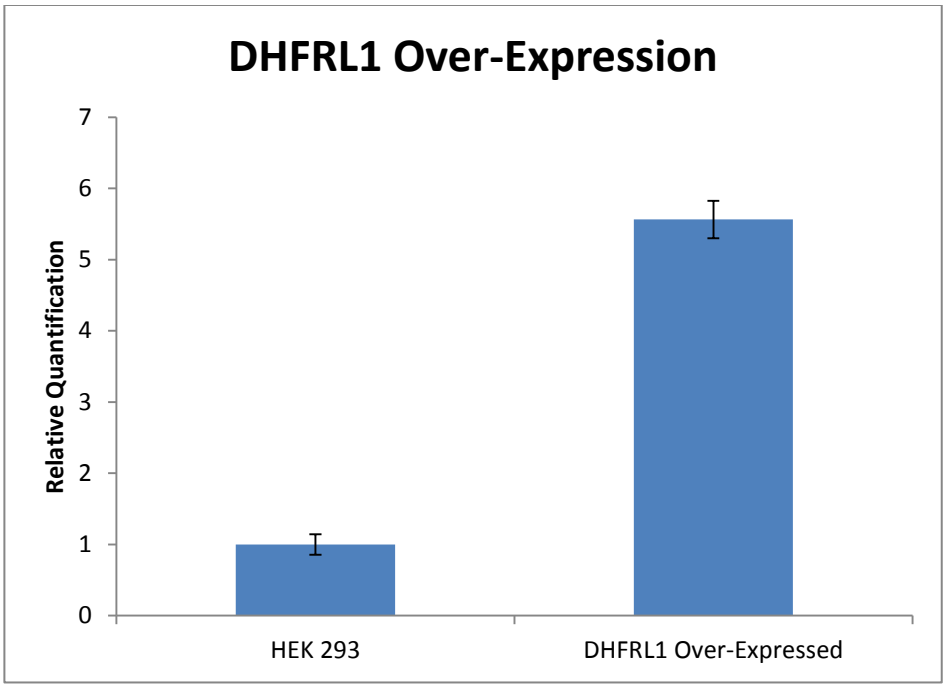


Figure 7.5. Validation of DHFRL1 over-expression by RT-qPCR in HEK 293 stable transfected cell pools. DHFRL1 gene expression was analysed in both HEK 293 and DHFRL1 over-expressed cells at the mRNA level by RT-qPCR to confirm up-regulation. The over-expressed cells show an up-regulation of ~ 5.5 fold. Relative expression was normalised against the HEK 293 cells. GUS was used as the reference gene.

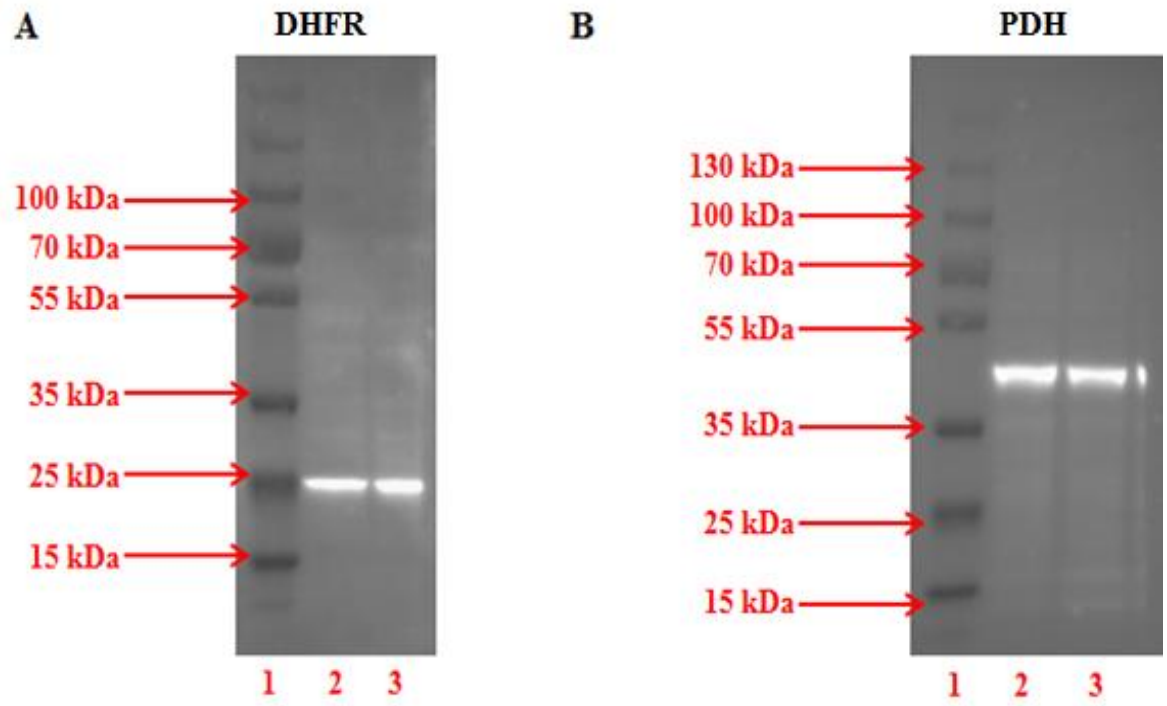


Figure 7.6. Validation of DHFRL1 over-expression by Western blot analysis for proteomic analysis. (A) Western blot of HEK 293 and DHFRL1 over-expressed mitochondrial protein fractions probed with the DHFR antibody. Lane 1 – protein ladder, Lane 2 – HEK 293 mitochondrial protein sample, Lane 3 – DHFRL1 over-expressed HEK 293 mitochondrial sample. All samples were equally loaded. Bands of the correct size (21 kDa) were seen for all of the mitochondrial protein samples. **(B) Western blot of HEK 293 and DHFRL1 over-expressed mitochondrial protein fractions probed with the PDH antibody.** Lane 1 – protein ladder, Lane 2 – HEK 293 mitochondrial protein sample, Lane 3 – DHFRL1 over-expressed HEK 293 mitochondrial sample. All samples were equally loaded. Bands of the correct size (43 kDa) were seen for all of the mitochondrial protein samples.

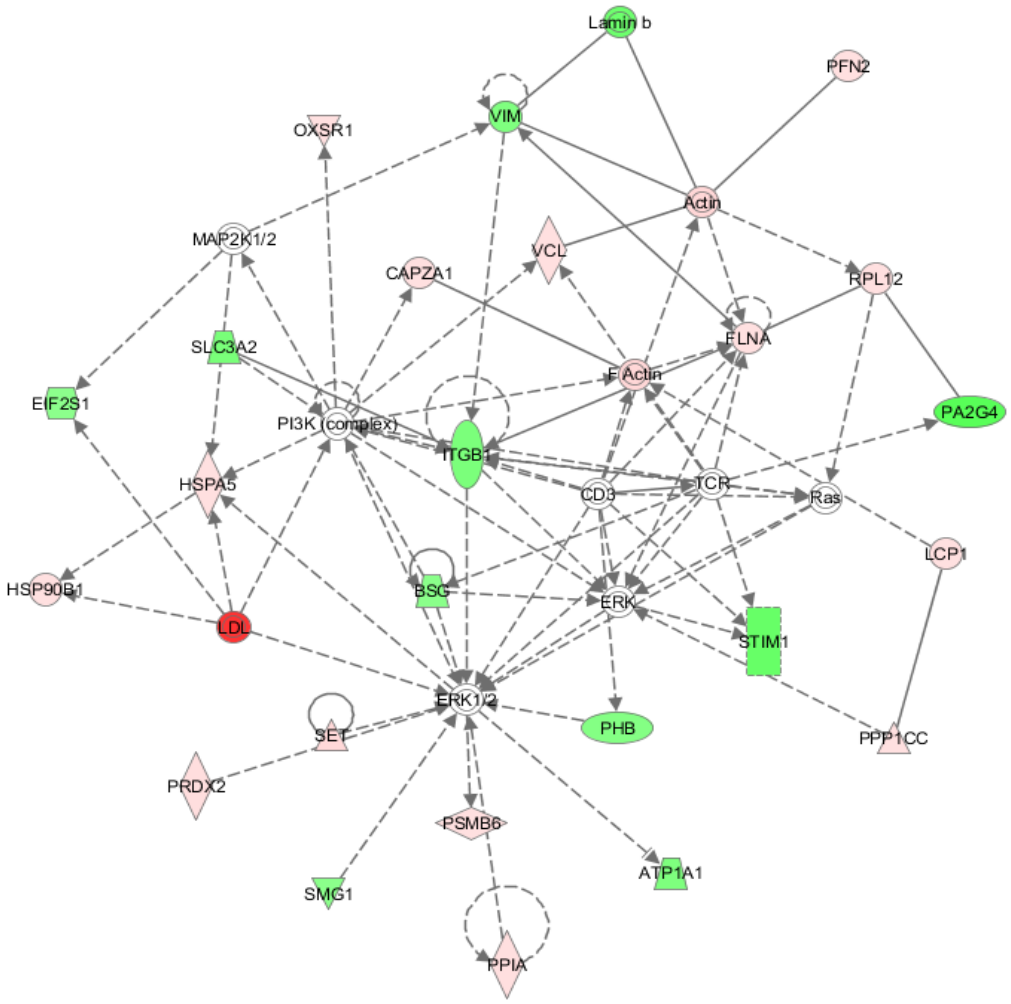


Figure 7.7. Network Analysis of Proteins Classified as Involved in Cell Morphology and Embryonic Development Drawn Using the IPA Software. Molecules in green represent those that are down-regulated, molecules in light green represent those that are slightly down-regulated, molecules in red represent those that are up-regulated, molecules in light red represent those that are slightly up-regulated and molecules in white represent those that were added from the Ingenuity Knowledge Base. Molecules represented by vertical diamonds are enzymes, those represented by horizontal diamonds are peptidases, those represented by horizontal ovals are transcription regulators, those represented by triangles are phosphatases, those represented by inverted triangles are kinases, those represented by squares are cytokines, those represented by rectangles are G-protein coupled receptors, those represented by trapezoids are transporters, those represented by double circles are complexes and those represented by circles are other.

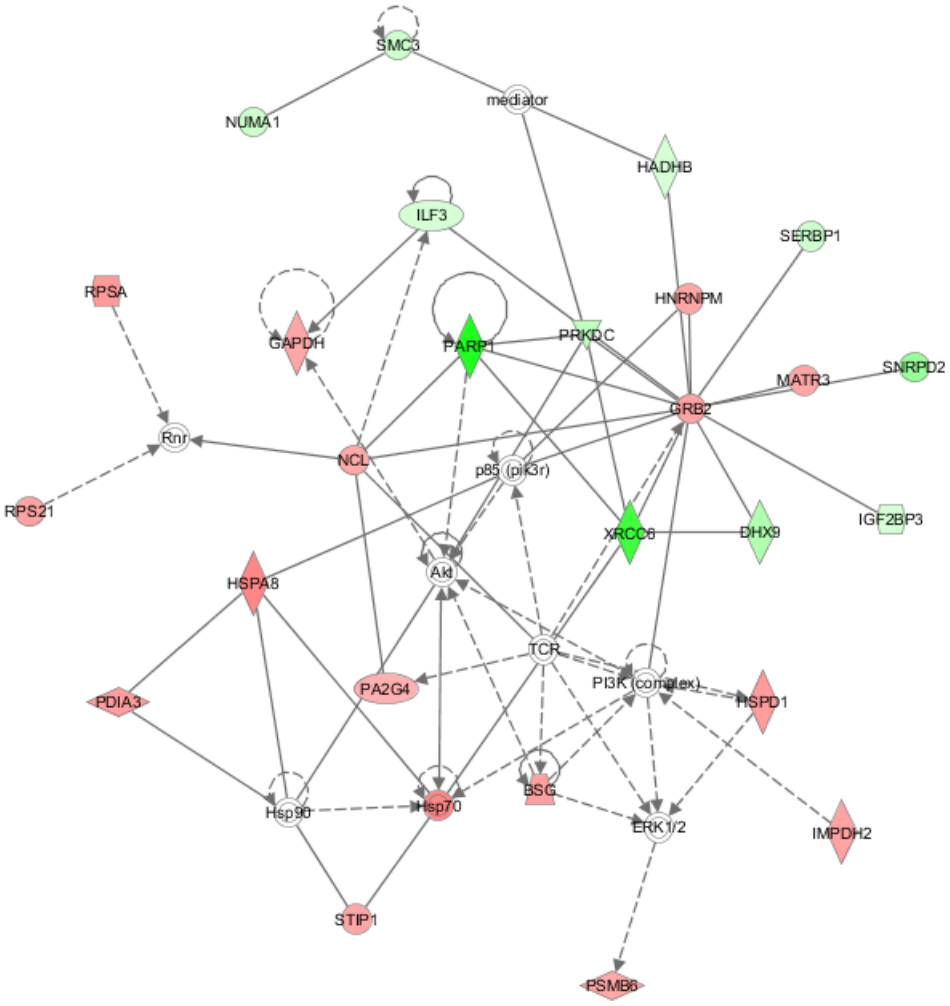


Figure 7.8. Network Analysis of Proteins Classified as Involved in Cellular Growth and Proliferation, Cancer and Cell Cycle Drawn Using the IPA Software. Molecules in green represent those that are down-regulated, molecules in light green represent those that are slightly down-regulated, molecules in light red represent those that are slightly up-regulated and molecules in white represent those that were added from the Ingenuity Knowledge Base. Molecules represented by vertical diamonds are enzymes, those represented by horizontal diamonds are peptidases, those represented by horizontal ovals are transcription regulators, those represented by triangles are kinases, those that are represented by squares are cytokines, those that are represented by double circles are complexes and those represented by circles are other.

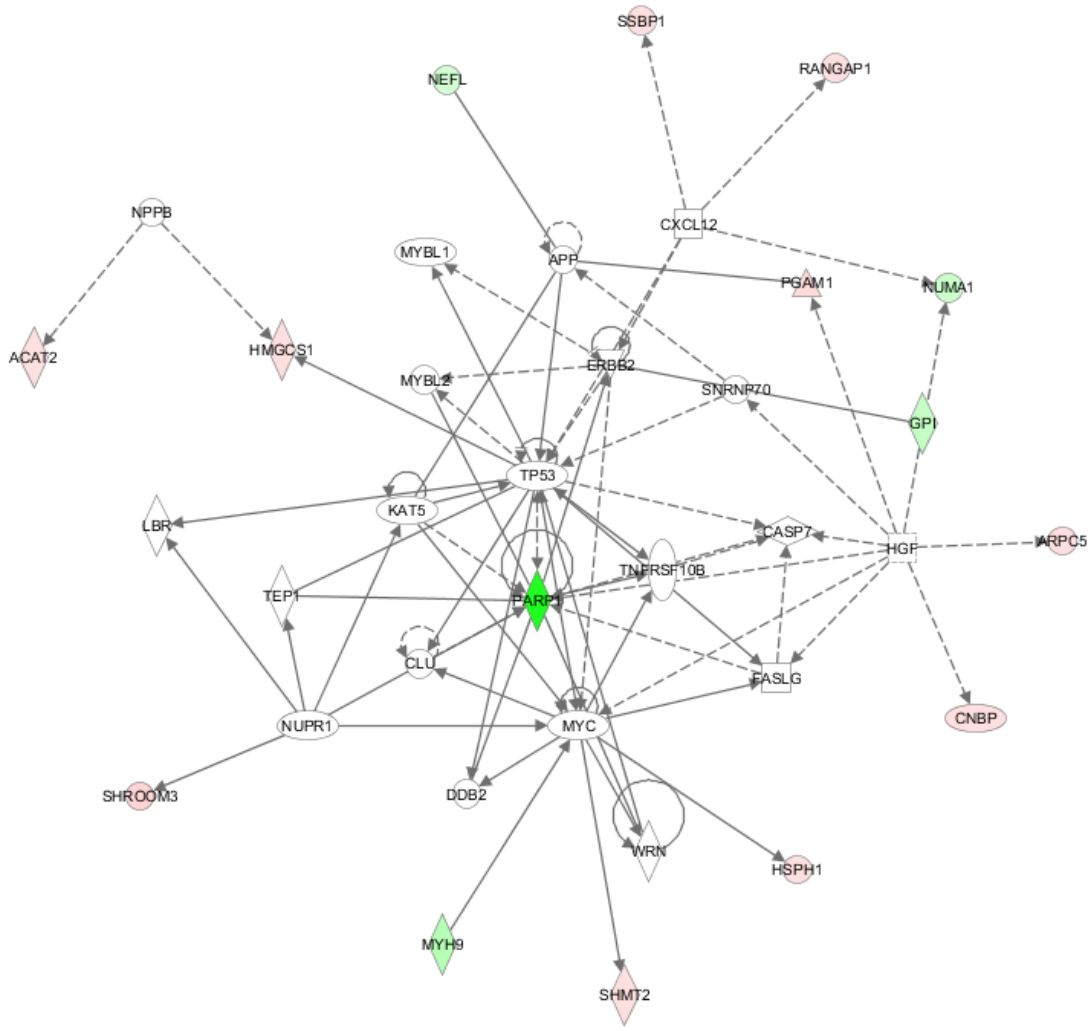


Figure 7.9. Network Analysis of Proteins Classified as Involved in Cell Death and Survival, Cell Cycle and DNA Replication, Recombination and Repair Using the IPA Software. Molecules in green represent those that are down-regulated, molecules in light green represent those that are slightly down-regulated, molecules in light red represent those that are slightly up-regulated and molecules in white represent those that were added from the Ingenuity Knowledge Base. Molecules represented by vertical diamonds are enzymes, those represented by horizontal ovals are transcription regulators, those represented by vertical ovals are transmembrane receptors, those that are represented by squares are cytokines and those represented by circles are other.

7.1 shRNA constructs used to generate DHFRL1 down-regulated cell pools.

	shRNA Construct	Vector	Mismatches with DHFR
1	Control	pLKO.1	
A	28 base shRNA†	1 x Lac O	7
B	19 base shRNA†	3 x Lac O	7
C	TRCN0000160428	3 x Lac O	4
D	TRCN0000161799‡	3 x Lac O	0
E	TRCN00001658880	1 x Lac O	5

† custom shRNA

‡selected to target both DHFR and DHFRL1

Table 7.2. List of proteins which altered in expression following down-regulation of DHFRL1 by shRNA B1 construct in HEK 293 cells.

Accession	Gene	Description	Peptides	Score	Anova (p)	Fold
P07311	ACYP1	Acylphosphatase-1	1	2.99	9.28E-05	1.41
P50454	SERPINH1	Serpin H1	1	3.64	0.00015	1.26
O95793	STAU1	Double-stranded RNA-binding Staufen homolog 1	1	3.61	0.00018	1.7
Q96HE7	ERO1L	ERO1-like protein alpha	1	4.77	0.00043	1.28
Q8NBS9	TXNDC5	Thioredoxin domain-containing protein 5	3	13.17	0.00066	1.47
P28072	PSMB6	Proteasome subunit beta type-6	2	5.59	0.00078	1.38
P37802	TAGLN2	Transgelin-2	2	6.66	0.00104	1.53
P20962	PTMS	Parathymosin	1	3.53	0.00135	1.7
O43852	CALU	Calumenin	1	2.62	0.00147	1.54
P14854	COX6B1	Cytochrome c oxidase subunit 6B1	1	2.73	0.00168	1.2
P16949	STMN1	Stathmin	4	10.87	0.00175	1.62
P15531	NME1	Nucleoside diphosphate kinase A	1	2.65	0.00236	1.48
P30085	CMPK1	UMP-CMP kinase	1	3.14	0.00243	1.39
P60174	TPI1	Triosephosphate isomerase	9	33.03	0.00244	1.37
P60174	PIWIL1	Piwi-like protein 1	1	2.28	0.00354	1.6
P06744	GPI	Glucose-6-phosphate isomerase	1	4.02	0.00372	1.29
Q9NZL9	MAT2B	Methionine adenosyltransferase 2 subunit beta	1	5.19	0.00381	1.3
P67936	TPM4	Tropomyosin alpha-4 chain	2	9.9	0.00388	1.52
P22314	UBA1	Ubiquitin-like modifier-activating enzyme 1	5	16.91	0.00451	1.23
Q13263	TRIM28	Transcription intermediary factor 1 beta	6	18.44	0.00499	1.4
P13489	RNH1	Ribonuclease inhibitor	2	7.74	0.00523	1.24
Q13347	EIF3I	Eukaryotic translation initiation factor	1	3.15	0.00536	1.28
Q96CT7	CCDC124	Coiled-coil domain-containing protein 124	2	5.65	0.00607	1.46

O94760	DDAH1	N[G], N[G]-dimethylarginine dimethylaminohydrolase 1	1	3.38	0.00611	1.2
P49327	FASN	Fatty acid synthase	2	6.53	0.00668	1.27
Q99615	DNAJC7	Dna J homolog subfamily C member 7	2	5.8	0.00673	1.38
P49915	GMPS	GMP synthase (glutamine hydrolysing)	1	2.75	0.00695	1.52
P15259	PGAM2	Phosphoglycerate mutase 2	1	3.83	0.00727	1.48
O00299	CLIC1	Chloride intracellular channel protein 1	1	3.01	0.00786	1.42
P55327	TPD52	Tumour protein D52	1	3.93	0.00867	1.61
Q6ZQQ2	SPATA31D1	Spermatogenesis associated protein	1	2.22	0.0089	1.24
O75390	CS	Citrate synthase, mitochondrial	1	2.56	0.00909	1.41
O43399	TPD52L2	Tumour protein D54	1	2.57	0.00933	1.45
P40925	MDH1	Malate dehydrogenase, cytoplasmic	2	8.22	0.00955	1.18
P54577	YARS	Tyrosine-tRNA ligase, cytoplasmic	1	2.32	0.01	1.68
O00170	AIP	AH receptor-interacting protein	1	2.52	0.01	1.51
O60664	PLIN3	Perilipin-3	1	3.38	0.01	1.36
O43423	ANP32C	Acidic leucine-rich nuclear phosphoprotein 32 family member C	2	6.55	0.01	1.36
P43487	RANBP1	Ran-specific GTPase-activating protein	1	2.77	0.01	1.42
P08107	HSPA1A	Heat shock 70kDa protein 1A/1B	3(2)	9.73	0.01	1.27
P14625	HSP90B1	Endoplasmin	2	5.4	0.01	1.26
Q9H444	CHMP4B	Charged multivescular body protein 4b	2	5.95	0.01	1.65
P15374	UCHL3	Ubiquitin carboxyl-terminal hydrolase isozyme L3	1	2.27	0.01	1.69
Q99832	CCT7	T-complex protein 1 subunit eta	1	3.94	0.01	1.2
P52907	CAPZA1	F-actin capping protein subunit alpha-1	1	3.9	0.01	1.41
P61956	SUMO2	Small ubiquitin-related modifier 2	2	6.31	0.01	1.25
O00154	ACOT7	Cytosolic acyl coenzyme A thioester hydrolase	1	5.26	0.01	1.39
P07900	HSP90AA1	Heat shock protein HSP90-alpha	5(4)	14.17	0.01	1.23
P29692	EEF1D	Elongation factor 1-delta	1	4.32	0.01	1.3
O75533	SF3B1	Splicing factor 3B subunit 1	1	2.99	0.01	1.52
Q02790	FKBP4	Peptidyl-prolyl cis-trans isomerase	2(1)	7.25	0.01	1.3
P32119	PRDX2	Peroxiredoxin 2	1	2.62	0.02	1.53
P13667	PDIA4	Protein disulfide isomerase A4	3	10.17	0.02	1.24

P62826	RANBP1	GTP-binding nuclear protein Ran	1	2.71	0.02	1.21
P14314	PRKCSH	Glucosidase 2 subunit beta	2	10.75	0.02	1.46
Q04637	EIF4G1	Eukaryotic translation initiation factor 4 gamma 1	1	2.73	0.02	1.66
P78527	PRKDC	DNA-dependent protein kinase catalytic subunit	2	7.05	0.02	1.75
P62937	PPIA	Peptidyl-prolyl cis-trans isomerase A	1	4.14	0.02	1.48
P08238	HSP90AB1	Heat shock protein HSP90-beta	2(1)	6.48	0.02	1.28
P07737	PFN1	Profilin 1	2	5.57	0.02	1.33
P00390	GSR	Glutathione reductase, mitochondrial	1	2.87	0.02	1.18
P08240	SRPR	Signal recognition particle receptor subunit alpha	1	2.43	0.02	1.56
P48643	CCT5	T-complex protein 1 subunit epsilon	1	3.48	0.02	1.33
P14618	PKM	Pyruvate kinase isozymes M1/M2	1	4.73	0.02	1.96
P60709	ACTB	Actin, cytoplasmic 1	1	2.28	0.02	1.87
P29401	TKT	Transketolase	1	2.4	0.02	1.52
C9JLR9	C11orf95	Uncharacterised protein C11 or95	1	1.92	0.02	1.76
P34932	HSPA4	Heat shock 70kDA protein 4	3(2)	12.66	0.02	1.28
P04075	ALDOA	Fructose bisphosphate aldolase A	1	2.81	0.02	1.49
Q9UNN5	FAF1	FAS-associated factor 1	1	2.63	0.02	1.46
Q15773	MLF2	Myeloid leukeia factor 2	1	2.37	0.02	1.48
P48147	PREP	Prolyl endopeptidase	1	3.19	0.02	1.18
P07108	DBI	Acyl-CoA binding protein	1	2.74	0.02	9.69
P18206	VCL	Vinculin	2	5.98	0.02	1.26
Q8TAW3	ZNF671	Zinc finger protein 671	1	2.33	0.02	3.16
P23396	RPS3	40S ribosomal protein S3	1	2.41	0.03	1.55
P49321	NASP	Nuclear autoantigenic sperm protein	1	3.24	0.03	1.33
Q9NTK5	OLA1	Obg-like ATPase 1	1	3.47	0.03	1.18
P50990	CCT8	T-complex protein 1 subunit theta	2	6.45	0.03	1.27
Q8WUW1	BRK1	Protein BRICK1	1	2.53	0.03	1.72
P06733	ENO1	Alpha-enolase	2	8.03	0.03	1.77
P31150	GDI1	Rab GDP dissocation inhibitor alpha	1	3.11	0.03	1.41
O15355	PPM1G	Protein phosphatase 1G	1	3.33	0.03	1.47

P35244	RPA3	Replcation protein A 14kDa subunit	1	2.47	0.03	1.25
Q16181	SEPT7	Septin 7	1	3.27	0.03	1.38
Q9H773	DCTPP1	dCTP pyrophosphatase 1	1	2.96	0.03	1.44
Q15019	SEPT2	Septin 2	1	4.84	0.03	1.25
P61626	LYZ	Lysozyme C	1	2.07	0.03	3267.94
Q15293	RCN1	Reticulocalbin-1	2	8.52	0.03	1.44
P11021	HSPA5	78kDa glucose-regulated protein	2(1)	7.97	0.03	1.13
P62633	CNBP	Cellular nucleic acid binding protein	1	2.48	0.03	1.45
O43665	RGS10	Regulator of G-protein signaling 10	1	2.34	0.03	1.73
Q9UHR4	BAIAP2L1	Brain-specific angiogenesis inhibitoyr 1-associated protein 2-like protein 1	1	2.05	0.03	1.17
P55884	EIF3B	Eukaryotic translation initiation factor 3 subunit B	1	4.15	0.03	1.25
P42167	TMPO	Lamina-assocaited polypeptide 2, isoforms beta/gamma	2	6.74	0.03	1.38
P04406	GAPDH	Glyceraldehyde 3-phosphate dehydrogenase	2	8.72	0.03	1.29
P21333	FLNA	Filamin-A	1	2.53	0.03	1.25
P11586	MTHFD1	C-1-tetrahydrofolate sybthase, cytoplasmic	1	2.21	0.03	1.2
Q9UHD8	SEPT9	Septin-9	2(1)	5.36	0.03	2.24
P47756	CAPZB	F-actin-capping protein subunit beta	1	3.76	0.03	1.54
P55036	PSMD4	26S proteasome non ATPase regulatory subunit 4	1	2.35	0.03	1.38
Q9UGP8	SEC63	Translocation protein SEC63 homolog	1	2.47	0.04	1.77
P07237	P4HB	Protein disulfide isomerase	1	5.06	0.04	1.4
P49773	HINT1	Histidine triad nuceotide-binding protein 1	1	3.04	0.04	2.86
Q01105	SET	Protein SET	1	2.94	0.04	1.82
P62258	YWHAE	14-3-3 protein epsilon	1	2.4	0.04	1.13
P40227	CCT6A	T-complex protein 1 subunit zeta	1	4.48	0.04	2.02
P61758	VBP1	Prefoldin subunit 3	1	2.82	0.04	1.46
O43242	PSMD3	26S proteasome non ATPase regulatory subunit 3	1	2.8	0.04	1.92
P33176	KIF5B	Kinesin-1 heavy chain	1	3.8	0.04	1.57
P36873	PPP1CC	Serine/threonine protein phosphatase PP1-gamma catalytic subunit	1	2.64	0.04	1.15
Q15056	EIF4H	Eukaryotic translation initiation factor 4H	1	2.21	0.04	1.83

Q13162	PRDX4	Peroxiredoxin-4	1	6.29	0.04	1.41
P35579	MYH9	Myosin-9	2	6.99	0.04	1.43
O43765	SGTA	Small glutamine-rich tetratricopeptide repeat-containing protein alpha	1	2.39	0.04	1.25
P13796	LCP1	Plastin-2	2	6.15	0.04	1.28
P00441	SOD1	Superoxide dismutase [Cu-Zn]	1	3.76	0.04	4.08
Q8WW12	PCNP	PEST proteolytic signal-containing nuclear protein	1	2.79	0.04	1.44
P61981	YWHAG	14-3-3 protein gamma	1	3.01	0.04	4.4
P35080	PFN2	Profilin-2	1	4.17	0.04	1.18
P35580	MYH10	Myosin-10	1	4.48	0.05	1.51
Q32MZ4	LRRFIP1	Leucine-rich repeat flightless-interacting protein 1	1	2.58	0.05	1.43
P68104	EEF1A1	Elongation factor 1-alpha 1	1	4.04	0.05	1.17
O95747	OXSR1	Serine/threonine protein kinase OSR1	1	3.88	0.05	1.22
O75874	IDH1	Isocitrate dehydrogenase [NADP] cytoplasmic	1	2.45	0.05	1.41
Q9H2E6	SEMA6A	Semaphorin-6A	1	2.25	0.05	1.79
Q99471	PFDN5	Prefoldin subunit 5	1	5.72	0.05	1.59
Q9UKY7	CDV3	Protein CDV3 homolog	1	4.69	0.05	1.38
P27816	MAP4	Microtubule-associaed protein 4	1	2.58	0.05	1.39
P00338	LDHA	L-lactate dehydrogenase A chain	1	3.91	0.05	1.19
Q13247	SRSF6	Serine/arginine rich splicing factor 6	2	5.38	6.45E-05	1.29
Q96PK6	RBM14	RNA-binding protein 14	3	7.35	0.000138	1.3
Q15007	WTAP	Pre-mRNA splicing regulator WTAP	1	3.19	0.00022	1.57
Q9Y5B9	SUPT16H	FACT complex subunit SPT16	5	14.67	0.000243	1.51
O00148	DDX39A	ATP-dependent RNA helicase DDX39A	6	17	0.000248	1.44
Q8NBN7	RDH13	Retinol dehydrogenase 13	2	6.66	0.00115	1.34
P31040	SDHA	Succinate dehydrogenase flavopotein subunit, mitochondrial	2	6.52	0.00144	1.22
Q14108	SCARB2	Lysosome membrane protein 2	1	2.59	0.00145	1.39
Q9BPW8	NIPSNAP1	Protein NipSnap homolg 1	1	3.31	0.0015	1.17
Q92841	DDX17	Probable ATP-dependent RNA helicase DDX17	5	13.74	0.00159	1.51
Q13813	SPTAN1	Spectrin alpha chain, non-erythrocytic 1	5	13.13	0.00163	1.37
Q9NVX2	NLE1	Notchless protein homolg 1	1	2.47	0.00167	1.3

P00505	GOT2	Aspartate aminotransferase, mitochondrial	1	2.33	0.00174	1.38
Q9P032	NDUFAF4	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex assembly factor 4	1	3.04	0.00208	1.33
P61978	HNRNPK	Heterogeneous nuclear ribonucleoprotein K	2	4.82	0.00226	1.31
P51659	HSD17B4	Peroxisoma multifunctional enzyme type 2	3	10.41	0.00275	1.29
Q96TA2	YME1L1	ATP-dependent zinc metalloprotease	2	6.88	0.00309	1.24
Q15393	SF3B3	Splicing factor 3B subunit 3	3	10.89	0.00326	1.24
O75439	PMPCB	Mitochondrial-processing peptidase subunit beta	2	6.3	0.00327	1.25
Q01082	SPTBN1	Spectrin beta chain, non-erythrocytic 1	6(5)	17.26	0.00355	1.38
P30049	ATP5D	ATP synthase subunit delta, mitochondrial	1	4.18	0.00374	1.19
P23246	SFPQ	Splicing factor, proline and glutamine rich	6	22.63	0.00376	1.61
P21796	VDAC1	Voltage-dependent anion selective channel protein 1	3	12.56	0.00437	1.21
P09622	DLD	Dihydrolipoyl dehydrogenase, mitochondrial	1	3.79	0.00441	1.4
O75955	FLOT1	Flotillin	1	3.63	0.00453	1.26
P05388	RPLP0	60S acidic ribosomal protein P0	1	3.18	0.00457	1.32
Q15233	NONO	Non-POU domain-containing octamer-binding protein	10(9)	28.47	0.00506	1.56
O00571	DDX3X	ATP-dependent RNA helicase DDX3X	4	13.23	0.00507	1.29
P18085	ARF4	ADP-ribosyltaion factor 4	1	2.75	0.0052	1.47
O95218	ZRANB2	Zinc finger Ran-binding domain-containing protein 2	1	4.89	0.00539	3.96
P52272	HNRNPM	Heterogeneous nuclear ribonucleoprotein M	4	8.51	0.00541	1.44
P22087	FBL	rRNA 2'-O-methyltransferase fibrillarin	1	4.67	0.00582	1.44
P05556	ITGB1	Integrin beta-1	1	3.8	0.00601	1.39
P50213	IDH3A	Isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial	1	2.81	0.00615	1.11
P20700	LMNB1	Lamin-B1	3	8.2	0.00622	1.54
Q06265	EXOSC9	Exosome complex component RRP45	1	2.33	0.0064	1.44
P61353	RPL27	60S ribosomal protein L27	1	2.78	0.00643	1.44
Q9UMS4	PRPF19	Pre-mRNA processing factor 19	1	2.72	0.00701	1.29
P55072	VCP	Transitional endoplasmic reticulum ATPase	1	3.34	0.00756	1.17
P35637	FUS	RNA-binding protein FUS	3	11.02	0.00768	1.44
O43707	ACTN4	Alpha-actinin-4	2	5.96	0.0078	1.27

Q00839	HNRNPU	Heterogeneous nuclear ribonucleoprotein U	1	1.92	0.00825	1.31
P14866	HNRNPL	Heterogeneous nuclear ribonucleoprotein L	3	12.14	0.00828	1.3
P42704	LRPPRC	Leucine-rich PPR motif-containing protein, mitochondrial	2	5.86	0.00832	1.19
Q15029	EFTUD2	116 kDa US small nuclear ribonucleoprotein component	1	2.32	0.00839	1.29
Q92547	TOPBP1	DNA topoisomerase 2-binding protein 1	1	2.05	0.0085	1.78
Q00610	CLTC	Clathrin heavy chain 1	1	4.83	0.00885	1.29
Q9Y2W1	THRAP3	Thyroid hormone receptor-associated protein 3	3	8.31	0.00901	1.76
Q9BZE1	MRPL37	39S ribosomal protein L37, mitochondrial	1	5.07	0.00911	1.26
Q92552	MPPS27	28S ribosomal protein S27, mitochondrial	1	6.4	0.00927	1.18
Q12788	TBL3	Transducin beta-like protein 3	1	2.68	0.00937	1.44
Q96JB5	CDK5RAP3	CDK5 regulatory subunit-associated protein 3	1	3.61	0.00937	1.24
P46776	RPL27A	60S ribosomal protein L27a	2	4.97	0.00962	1.44
O94905	ERLIN2	Erlin-2	1	3.5	0.00969	1.28
Q9P2R7	SUCLA2	Succinyl-CoA ligase [ADP-forming] subunit beta, mitochondrial	1	3.34	0.00974	1.16
P53582	METAP1	Methionine aminopeptidase 1	1	2.98	0.01	1.33
Q92945	KHSRP	Far upstream element-binding protein 2	3	14.16	0.01	1.55
P31943	HNRNPH1	Heterogeneous nuclear ribonucleoprotein H	1	4.54	0.01	1.29
P28838	LAP3	Cytosol aminopeptidase	1	3.5	0.01	1.07
Q13423	NNT	NAP(P) transhydrogenase, mitochondrial	2	7.29	0.01	1.2
Q9UKV3	ACIN1	Apoptotic chromatin condensation inducer in the nucleus	1	3.52	0.01	1.38
Q15075	EEA1	Early endosome antigen 1	1	2.77	0.01	1.37
Q9UQ80	PA2G4	Proliferation-associated protein 2G4	1	2.21	0.01	1.76
Q12931	TRAP1	Heat shock protein 75kDa, mitochondrial	1	3.49	0.01	1.23
Q9HB40	SCPEP1	Retinoid-inducible serine carboxypeptidase	1	3.28	0.01	1.15
Q9BVP2	GNL3	Guanine nucleotide-binding protein-like3	1	5.52	0.01	1.54
Q9UHB6	LIMA1	LIM domain and actin-binding protein 1	1	3.26	0.01	1.48
O00410	IPO5	Importin 5	1	2.44	0.01	1.18
P19338	NCL	Nucleolin	4	9.87	0.01	1.46
Q16891	IMMT	Mitochondrial inner membrane protein	3	10.28	0.01	1.24
Q9Y277	VDAC3	Voltage-dependent anion selective channel protein 3	3	9.42	0.01	1.3

Q5JTZ9	AARS2	Alanine-tRNA ligase, mitochondrial	1	2.99	0.01	1.2
P62917	RPL8	60S ribosomal protein L8	1	4.38	0.01	1.85
P20340	RAB6A	Ras-related protein Rab-6A	4	11.16	0.01	1.3
P25705	ATP5A1	ATP synthase subunit alpha, mitochondrial	2	6.72	0.01	1.31
Q9UQ35	SRRM2	Serine/arginine repetitve matrix protein 2	5	15.03	0.01	1.48
Q8NFH5	NUP35	Nucleoporin	1	2.51	0.01	1.26
P08195	SLC3A2	4F2 cll-surface antigen heavy chain	1	4.06	0.01	1.41
P28331	NDUFS1	NADH-ubiquinone oxidoreductase 75kDa subunit, mitochondrial	2	6.44	0.01	1.29
P06239	LCK	Tyrosine-protein kinase Lck	1	3.23	0.01	1.34
P08134	RHOC	Rho-related GTP-binding protein RhoC	1	1.98	0.01	1.35
O00422	SAP18	Histone deacetylase complex subunit SAP18	1	2.59	0.01	1.39
Q7Z4V5	HDGFRP2	Hepatoma derived growth factor-related protein 2	1	4.34	0.01	1.33
Q96DI7	SNRNP40	U5 small nuclear ribonucleoprotein 40kDa protein	1	3.19	0.02	1.4
Q2TAY7	SMU1	WD40 repeat-containing protein SMU1	1	2.25	0.02	1.6
Q96I24	FUBP3	Far upstream element-binding protein 3	1	4.06	0.02	1.32
Q02543	RPL18A	60S ribosomal protein L18a	1	2.98	0.02	1.53
Q08211	DHX9	ATP-dependent RNA helicase A	3	10.61	0.02	1.32
P35613	BSG	Basigin	1	5.63	0.02	1.23
P20339	RAB5A	Ras-related protein Rab-5A	1	2.83	0.02	1.51
Q9UJS0	SLC25A13	Calcium-bindin mitochondrial carrier protein Aralar 2	1	3.82	0.02	1.2
P38432	COIL	Coilin	1	3.22	0.02	2.36
P31937	HIBADH	3-hydroxyisobutyrate dehydrogenase, mitochondrial	1	3.34	0.02	1.11
Q12905	ILF2	Interleukin enhancer-binding factor 2	2	6.61	0.02	1.26
P00367	GLUD1	Glutamate dehydrogenase 1, mitochondrial	1	2.74	0.02	1.17
Q96A26	FAM162A	Protein FAM162A	1	2.96	0.02	1.4
Q9H1E3	NUCKS1	Nuclear ubiquitous casein and cyclin-dependent kinase substrate	1	6.28	0.02	2.3
Q8N1F7	NUP93	Nuclear pore complex protein Nup93	1	2.2	0.02	1.21
O95831	AIFM1	Apoptosis-inducing factor 1, mitochondrial	1	2.16	0.02	1.29
P05023	ATP1A1	Sodium/potassium-transporting ATPase subunit alpha-1	2	5.14	0.02	1.36
P27144	AK$	Adenylate kinase isoenzyme 4, mitochondrial	1	2.57	0.02	1.29

P08754	GNAI3	Guanine nucleotide-binding protein G(k) subunit alpha	1	3.44	0.02	1.28
O95167	NDUFA3	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 3	1	1.94	0.02	1.35
P35250	RFC2	Replication factor C subunit 2	1	4.12	0.02	1.17
P51149	RAB7A	Ras-related protein Rab-7a	1	2.98	0.02	1.66
Q96SB4	SRPK1	SRSF protein kinase 1	1	4.11	0.02	1.57
Q8IYB3	SRRM1	Serine/arginine repetitve matrix protein 1	1	2.89	0.02	1.36
P0C0S5	H2AFZ	Histone H2A.Z	1	2.12	0.02	1.29
P38919	EIF4A3	Eukaryotic initiation factor 4A-III	2	5.06	0.02	1.19
O43684	BUB3	Mitotic checkpoint protein	1	3.91	0.02	1.16
Q14978	NOLC1	Nucleolar and coiled-body phosphprotein 1	2	5.78	0.02	1.55
O96008	TOMM40	Mitochondrial import receptor subunit TOM40 homolog	1	4.65	0.02	1.29
Q07955	SRSF1	Serine/arginine-rich splicing factor 1	3	8.2	0.02	1.39
Q9Y3B9	RRP15	RRP15-like protein	1	3.16	0.02	1.14
P63244	GNB2L1	Guanine nucleotide-binding protein subunit beta 2-like 1	3	12.49	0.02	1.31
P45880	VDAC2	Voltage-dependent anion-selective channel protein 2	1	2.35	0.03	1.47
P31930	UQCR1	Cytochrome b-c1 complex subunit 1, mitochondrial	1	3.13	0.03	1.21
Q15717	ELAVL1	ELAV-like protein 1	1	3.51	0.03	1.18
P35232	PHB	Prohibition	1	2.28	0.03	1.3
P24752	ACAT1	Acetyl-CoA acetyltransferase, mitochondrial	2	7.1	0.03	1.27
P03247	E1B	E1B protein, small T-antigen	1	2.35	0.03	1.22
Q9NUJ1	ABHD10	Mycophenoic acid acyl-glucuronide esterase, mitochondrial	1	4.86	0.03	1.25
P84095	RHOG	Rho-related GTP-binding protein RhoG	1	4.15	0.03	1.87
P34897	SHMT2	Serine hydroxymethyltransferase, mitochondrial	1	3.12	0.03	1.24
O75643	SNRNP200	U5 small nuclear ribonucleoprotein 200kDa helicase	1	3.11	0.03	1.31
Q9Y305	ACOT9	Acyl-coenzyme A thioesterase 9, mitochondrial	1	2.97	0.03	1.44
Q01130	SRSF2	Serine/arginine-rich splicing factor 2	2	4.18	0.03	1.65
Q92667	AKAP1	A-kinase anchor protein 1, mitochondrial	1	3.87	0.03	1.44
O00541	PES1	Pescadillo homlog	1	2.97	0.03	1.52
Q9H3P7	ACBD3	Golgi resident protein GCP60	1	4.44	0.03	1.19
P49755	TMED10	Transmembrane emp24 domain-containing protein 10	1	3.8	0.03	1.44

P13804	ETFA	Electron transfer flavoprotein subunit alpha, mitochondrial	1	2.74	0.03	1.09
Q8TEV9	SMCR8	Smith-Magenis syndrome chromosomal region candidate gene 8 protein	1	2.29	0.03	1.59
P04181	OAT	Ornithine aminotransferase, mitochondrial	2	7.55	0.03	1.23
Q15800	MSMO1	Methylsterol monooxygenase 1	1	5.02	0.03	1.3
P02545	LMNA	Prelamin-A/C	1	3.18	0.03	1.31
P46777	RPL5	60S ribosomal protein L5	1	2.8	0.03	1.28
Q07065	CKAP4	Cytoskeleton-associated protein 4	1	2.31	0.03	1.28
P31942	HNRNPH3	Heterogeneous nuclear ribonucleoprotein H3	1	3.87	0.03	1.13
Q969V3	NCLN	Nicalin	1	4.18	0.03	1.27
P40926	MDH2	Malate dehydrogenase, mitochondrial	1	3.42	0.03	1.34
P48735	IDH2	Isocitrate dehydrogenase [NADP], mitochondrial	1	3.06	0.03	1.27
P13010	XRCC5	X-ray repair cross-complementing protein 5	1	2.27	0.04	1.19
Q96Q15	SMG1	Serine/threonine protein kinase SMG1	1	2.35	0.04	1.22
Q9H074	PAIP1	Polyadenylate-binding protein interacting protein 1	1	2.96	0.04	1.27
Q9UKL3	CASP8AP2	CASP8-assocaited protein 2	1	2.21	0.04	1.69
P11498	PC	Pyruvate carboxylase, mitochondrial	1	3.52	0.04	1.28
P38117	ETFB	Electron transfer flavoprotein subunit beta, mitochondrial	1	2.08	0.04	1.38
P38646	HSPA9	Stress-70 protein, mitochondrial	1	3.44	0.04	1.13
Q9NS69	TOMM22	Mitochondrial import receptor subunit TOM22 homolog	1	4.14	0.04	1.33
P11177	PDHB	Pyruvate dehydrogenase E1 component subunit beta, mitochondrial	1	2.72	0.04	1.35
P46778	RPL21	60S ribosomal protein L21	1	2.76	0.04	1.24
Q14444	CAPRIN1	Caprein-1	1	3.2	0.04	1.45
Q8WXF0	SRSF12	Serine/arginine-rich splicing factor 12	1	3.06	0.04	1.18
P53992	SEC24C	Protein transport protein Sec24C	1	3.37	0.04	1.33
P09543	CNP	2',3'-cyclic-nucleotide 3'-phosphodiesterase	1	3.7	0.04	1.43
A6NIZ1		Ras-related protein Rap-1-b like protein	1	2.37	0.04	1.39
P54709	ATP1B3	Sodium/potassium-transporting ATPase subunit beta	1	3.01	0.04	1.23
P39023	RPL3	60S ribosomal protein L3	2	5.37	0.04	1.34
P63092	GNAS	Guanine nucleotide-binding protein G(s) subunit alpha isoforms short	1	3.68	0.04	1.23

Q99714	HSD17B10	3-hydroxyacyl-CoA dehydrogenase type-2	1	2.01	0.04	1.81
P62316	SNRPD2	Small nuclear ribonucleoprotein Sm D2	1	3.15	0.04	1.31
P52815	MRPL12	39S ribosomal protein L12, mitochondrial	1	2.86	0.04	1.09
P11233	RALA	Ras-related protein Ral-A	1	4.1	0.04	1.32
P62306	SNRPF	Small nuclear ribonucleoprotein F	1	3.36	0.04	1.53
P51649	ALDH5A1	Succinate semialdehyde dehydrogenase, mitochondrial	1	3.04	0.04	1.24
Q9NYF0	DACT1	Dapper homolog 1	1	2.53	0.05	1.64
Q8WWY3	PRPF31	U4/U6 small nuclear ribonucleoprotein Prp31	1	4.39	0.05	1.16
Q13586	STIM1	Stromal interation molecule 1	1	2.73	0.05	1.6
P08670	VIM	Vimentin	1	2.25	0.05	1.35
Q15365	PCBP1	Poly(rC)-binding protein 1	1	2.49	0.05	1.55
O75964	ATP5L	ATP synthase subunit g, mitochondrial	1	2.56	0.05	1.93
Q8WZ42	TTN	Titin	1	1.99	0.05	1.71
P32969	RPL9	60S ribosomal protein L9	1	3.13	0.05	1.32
P37268	FDFT1	Squalene synthase	1	3.27	0.05	1.23
Q9NYF8	BCLAF1	Bcl-2-associated transcription factor 1	1	3.34	0.05	1.3
Q13535	ATR	Serine/threonine protein kinase ATR	1	2.34	0.05	1.17
Q6PIW4	FIGNL1	Fidgetin-like protein 1	1	2.74	0.05	1.15
P05198	EIF2S1	Eukaryotic translation initiation factor 2 subunit 1	1	3.35	0.05	1.21
P62910	RPL32	60S ribosomal protein L32	1	5.12	0.05	1.31

Table 7.3. List of proteins which altered in expression following down-regulation of DHFRL1 by shRNA C1 construct in HEK 293 cells.

Accession	Gene	Description	Peptides	Score	Anova (p)	Fold
Q02878	RPL6	60S ribosomal protein L6	2	4.08	0.00634	1.42
Q02543	RPL18A	60S ribosomal protein L18a	1	2.05	0.00848	1.54
P30050	RPL12	60S ribosomal protein L12	1	2.1	0.00969	1.23
P26599	PTBP1	Polypyrimidine tract-binding protein 1	1	2.26	0.01	1.09
Q6ZU65	UBN2	Ubinuclein-2	1	2.3	0.02	1.89
P15880	RPS2	40S ribosomal protein S2	1	2.03	0.02	1.32
P62917	RPL8	60S ribosomal protein L8	1	3.14	0.03	1.35
P13010	XRCC5	X-ray repair cross-complementing protein 5	1	3.17	0.04	1.39
P50454	SERPINH1	Serpin H1	1	3.21	0.04	1.24
Q14684	RRP1B	Ribosomal RNA processing protein 1 homolog B	1	2.73	0.04	1.29
P09622	DLD	Dihydrolipoyl dehydrogenase, mitochondrial	1	3.19	0.04	1.36
Q14684	RRP1B	Ribosomal RNA processing protein 1 homolog B	1	2.73	0.04	1.29
Q9UQE7	SMC3	Structural maintenance of chromosomes protein 3	1	2.98	0.0076	1.22
Q7Z3U7	MON2	Protein MON2 homolog	1	2.57	0.01	1.9
Q9UBQ0	VPS29	Vacuolar protein sorting-associated protein 29	1	3.77	0.01	1.63
P49756	RBM25	RNA-binding protein 25	1	3.09	0.02	1.15
P62993	GRB2	Growth factor receptor-bound protein 2	1	2.71	0.02	1.07
P61978	HNRNPK	Heterogeneous nuclear ribonucleoprotein K	1	3.39	0.02	1.4
O95782	AP2A1	AP-2 complex subunit alpha-1	1	2.61	0.03	1.33
Q99536	VAT1	Synaptic vesicle membrane protein VAT-a homolog	1	3.95	0.03	1.17
Q9NYF3	FAM53C	Protein FAM53C	1	2.22	0.04	1.44
Q9BQE3	TUBA1C	Tubulin alpha-1C chain	1	2.03	0.04	1.17
Q96QK1	VPS35	Vacuolar protein sorting-associated protein 35	1	3.11	0.04	1.23
Q8NI60	ADCK3	Chaperone activity of bc1 complex-like, mitochondrial	1	2.41	0.05	1.16

Table 7.4. List of proteins that are involved in DNA Replication, Recombination and Repair, Nucleic Acid Metabolism, Cell Death and Survival and Cell Cycle that are affected by down-regulating DHFRL1 in HEK 293 cells.

Gene	Function	Fold
TXNDC5	Cell Death & Survival	1.47
TAGLN2	Cell Death & Survival	1.53
STMN1	DNA Replication, Recombination & Repair / Cell Death & Survival	1.62
NME1	DNA Replication, Recombination & Repair / Cell Cycle / Cell Death & Survival	1.48
GPI	Cell Death & Survival	1.29
UBA1	Cell Death & Survival	1.23
TRIM28	DNA Replication, Recombination & Repair / Cell Death & Survival	1.4
FASN	Cell Cycle / Cell Death & Survival	1.27
TPD52	Cell Death & Survival	1.61
YARS	Cell Death & Survival	1.68
PLIN3	DNA Replication, Recombination & Repair	1.36
HSP90B1	Cell Death & Survival	1.26
SUMO2	DNA Replication, Recombination & Repair	1.25
HSP90AA1	Nucleic Acid Metabolism / DNA Replication, Recombination & Repair / Cell Death & Survival	1.23
SF3B1	DNA Replication, Recombination & Repair	1.52
PRDX2	Cell Death & Survival	1.53
EIF4G1	Cell Death & Survival	1.66
PRKDC	DNA Replication, Recombination & Repair / Cell Death & Survival	1.75
PPIA	DNA Replication, Recombination & Repair	1.48
HSP90AB1	Cell Death & Survival	1.28
PFN1	Cell Cycle	1.33
GSR	Cell Death & Survival	1.18
CCT5	Nucleic Acid Metabolism / DNA Replication, Recombination & Repair	1.33
PKM	Nucleic Acid Metabolism / Cell Death & Survival	1.96

HSPA4	Cell Death & Survival	1.28
ALDOA	Cell Death & Survival	1.49
FAF1	Cell Death & Survival	1.46
OLA1	Nucleic Acid Metabolism / DNA Replication, Recombination & Repair	1.18
CCT8	Nucleic Acid Metabolism / DNA Replication, Recombination & Repair	1.27
ENO1	Cell Death & Survival	1.77
PPM1G	Cell Death & Survival	1.47
RPA3	DNA Replication, Recombination & Repair	1.25
SEPT7	Cell Cycle	1.38
LYZ	Cell Death & Survival	3267.94
HSPA5	Cell Death & Survival	1.13
GAPDH	Cell Death & Survival	1.29
FLNA	Cell Cycle / Cell Death & Survival	1.25
SEPT9	Cell Cycle	2.24
P4HB	Cell Death & Survival	1.4
HINT1	Cell Death & Survival	2.86
SET	Cell Death & Survival	1.82
YWHAE	Cell Death & Survival	1.13
PPP1CC	Cell Cycle / Cell Death & Survival	1.15
MYH9	Cell Cycle / Cell Death & Survival	1.43
SOD1	DNA Replication, Recombination & Repair / Cell Cycle / Cell Death & Survival	4.08
YWHAG	Cell Cycle / Cell Death & Survival	4.4
MYH10	Cell Cycle	1.51
EEF1A1	DNA Replication, Recombination & Repair	1.17
OXSR1	DNA Replication, Recombination & Repair	1.22
SEMA6A	Cell Death & Survival	1.79
LDHA	Nucleic Acid Metabolism / Cell Death & Survival	1.19
WTAP	Cell Death & Survival	-1.57
RDH13	Cell Death & Survival	-1.34
SDHA	Cell Death & Survival	-1.22

DDX17	Nucleic Acid Metabolism / DNA Replication, Recombination & Repair / Cell Death & Survival	-1.51
SPTAN1	Cell Cycle	-1.37
NDUFAF4	Cell Death & Survival	-1.33
HNRNPK	Cell Death & Survival	-1.31
SFPQ	DNA Replication, Recombination & Repair / Cell Death & Survival	-1.61
VDAC1	Cell Death & Survival	-1.21
RPLP0	Cell Death & Survival	-1.32
DDX3X	Nucleic Acid Metabolism / DNA Replication, Recombination & Repair / Cell Death & Survival	-1.29
ARF4	Cell Death & Survival	-1.47
ITGB1	Cell Cycle / Cell Death & Survival	-1.39
LMNB1	Cell Cycle / Cell Death & Survival	-1.54
PRPF19	DNA Replication, Recombination & Repair / Cell Cycle / Cell Death & Survival	-1.29
VCP	Nucleic Acid Metabolism / DNA Replication, Recombination & Repair / Cell Death & Survival	-1.17
FUS	DNA Replication, Recombination & Repair / Cell Cycle	-1.44
HNRNPU	DNA Replication, Recombination & Repair / Cell Death & Survival	-1.31
TOPBP1	Cell Cycle / Cell Death & Survival	-1.78
TBL3	DNA Replication, Recombination & Repair	-1.44
CDK5RAP3	Cell Cycle / Cell Death & Survival	-1.24
RPL27A	Cell Death & Survival	-1.44
ACIN1	DNA Replication, Recombination & Repair / Cell Cycle / Cell Death & Survival	-1.38
TRAP1	Cell Death & Survival	-1.23
GNL3	Cell Cycle	-1.54
NCL	DNA Replication, Recombination & Repair / Cell Death & Survival	-1.46
IMMT	Cell Death & Survival	-1.24
VDAC3	Nucleic Acid Metabolism	-1.3
ATP5A1	Nucleic Acid Metabolism / DNA Replication, Recombination & Repair / Cell Death & Survival	-1.31
SRRM2	DNA Replication, Recombination & Repair	-1.48
SLC3A2	Cell Death & Survival	-1.41
NDUFS1	Nucleic Acid Metabolism / DNA Replication, Recombination & Repair	-1.29
LCK	DNA Replication, Recombination & Repair / Cell Death & Survival	-1.34

RHOC	Cell Cycle / Cell Death & Survival	-1.35
SAP18	Cell Death & Survival	-1.39
SMU1	DNA Replication, Recombination & Repair	-1.6
DHX9	Nucleic Acid Metabolism / DNA Replication, Recombination & Repair	-1.32
BSG	DNA Replication, Recombination & Repair / Cell Death & Survival	-1.23
RAB5A	Cell Death & Survival	-1.51
SLC25A13	Nucleic Acid Metabolism	-1.2
ILF2	Cell Death & Survival	-1.26
GLUD1	Cell Death & Survival	-1.17
FAM162A	Cell Death & Survival	-1.4
NUP93	Cell Death & Survival	-1.21
AIFM1	DNA Replication, Recombination & Repair / Cell Death & Survival	-1.29
ATP1A1	Nucleic Acid Metabolism / DNA Replication, Recombination & Repair / Cell Death & Survival	-1.36
AK4	Nucleic Acid Metabolism / DNA Replication, Recombination & Repair	-1.29
GNAI3	Cell Cycle	-1.28
SRPK1	DNA Replication, Recombination & Repair / Cell Death & Survival	-1.57
EIF4A3	Nucleic Acid Metabolism / DNA Replication, Recombination & Repair / Cell Death & Survival	-1.19
SRSF1	Cell Death & Survival	-1.39
GNB2L1	Cell Death & Survival	-1.31
VDAC2	Cell Death & Survival	-1.47
ELAVL1	Cell Cycle / Cell Death & Survival	-1.18
PHB	Cell Death & Survival	-1.3
RHOG	Cell Death & Survival	-1.87
SNRNP200	Nucleic Acid Metabolism / DNA Replication, Recombination & Repair	-1.31
SRSF2	Cell Death & Survival	-1.65
TMED10	Cell Death & Survival	-1.44
LMNA	Cell Cycle / Cell Death & Survival	-1.31
RPL5	Cell Cycle	-1.28
XRCC5	DNA Replication, Recombination & Repair / Cell Cycle / Cell Death & Survival	-1.19
SMG1	Cell Death & Survival	-1.22

CASP8AP2	Cell Cycle / Cell Death & Survival	-1.69
HSPA9	Cell Death & Survival	-1.13
GNAS	Cell Death & Survival	-1.23
HSD17B10	Cell Death & Survival	-1.81
RALA	Cell Cycle / Cell Death & Survival	-1.32
SNRPF	DNA Replication, Recombination & Repair	-1.53
STIM1	Cell Death & Survival	-1.6
ATP5L	Nucleic Acid Metabolism / DNA Replication, Recombination & Repair	-1.93
FDFT1	Cell Death & Survival	-1.23
BCLAF1	DNA Replication, Recombination & Repair / Cell Death & Survival	-1.3
ATR	DNA Replication, Recombination & Repair / Cell Cycle / Cell Death & Survival	-1.17
FIGNL1	DNA Replication, Recombination & Repair	-1.15
EIF2S1	Cell Death & Survival	-1.21
SMC3	DNA Replication, Recombination & Repair	-1.22
RBM25	Cell Death & Survival	-1.15

Table 7.5. List of proteins that had altered expression following down-regulation of DHFRL1 in HEK 293 cells with both shRNA constructs.

Accession	Gene	Description	Peptides	Score	Anova (p)	Fold		Average
P50454	SERPINH1	Serpin H1	1	3.64	0.000158	+1.26	+1.24	1.25
P61978	HNRNPK	Heterogeneous nuclear ribonucleoprotein K	2	4.82	0.00226	-1.31	-1.4	-1.355
P09622	DLD	Dihydrolipoyl dehydrogenase, mitochondrial	1	3.79	0.00441	-1.4	-1.36	-1.38

Table 7.6. List of proteins which increased in expression following over-expression of DHFRL1 in HEK293 cells.

Accession	Gene	Description	Peptides	Score	Anova (p)	Fold
Q96AE4	FUBP1	Far upstream element-binding protein 1	1	56.66	1.46E-03	1.27
Q92598	HSPH1	Heat shock protein 105 kDa	1	43.57	2.92E-03	1.64
O75947	ATP5H	ATP synthase subunit d, mitochondrial	1	40.01	3.26E-03	1.59
P52272	HNRNPM	Heterogeneous nuclear ribonucleoprotein M	3	191.01	3.85E-03	1.23
P04075	ALDOA	Fructose-bisphosphate aldolase A	1	53.39	4.50E-03	1.36
P11142	HSPA8	Heat shock cognate 71 kDa protein	2	97.21	8.70E-03	1.67
Q9BWD1	ACAT2	Acetyl-CoA acetyltransferase, cytosolic	1	62.45	8.94E-03	1.47
O60637	TSPAN3	Tetraspanin-3	1	43.83	9.26E-03	1.59
P10599	TXN	Thioredoxin	1	41.28	0.01	1.26
P63220	RPS21	40S ribosomal protein S21	1	61.29	0.01	1.27
P55036	PSMD4	26S proteasome non-ATPase regulatory subunit 4	1	113.48	0.01	1.35
P34897	SHMT2	Serine hydroxymethyltransferase, mitochondrial	1	64.09	0.01	1.37
P63244	GNB2L1	Guanine nucleotide-binding protein subunit beta-2-like 1	1	41.38	0.01	1.35
O60884	DNAJA2	DnaJ homolog subfamily A member 2	1	59.92	0.01	1.42
P28072	PSMB6	Proteasome subunit beta type-6	1	72.19	0.02	1.2
P00505	GOT2	Aspartate aminotransferase, mitochondrial	1	59.43	0.02	1.23
P18669	PGAM1	Phosphoglycerate mutase 1	1	48.01	0.02	2.36
O15511	ARPC5	Actin-related protein 2/3 complex subunit 5	1	53.7	0.02	1.43
P40926	MDH2	Malate dehydrogenase, mitochondrial	1	45.23	0.02	1.13
P10809	HSPD1	60 kDa heat shock protein, mitochondrial	1	51.44	0.02	1.37
P13639	EEF2	Elongation factor 2	1	41.69	0.02	1.45
P19338	NCL	Nucleolin	1	46.9	0.03	1.19
P46060	RANGAP1	Ran GTPase-activating protein 1	1	44.04	0.03	1.61
Q04837	SSBP1	Single-stranded DNA-binding protein, mitochondrial	1	55.65	0.03	1.19
P67870	CSNK2B	Casein kinase II subunit beta	1	60.17	0.03	1.37
P09497	CLTB	Clathrin light chain B	1	59.61	0.03	1.33

P31948	STIP1	Stress-induced-phosphoprotein 1	1	53.35	0.03	1.22
P23588	EIF4B	Eukaryotic translation initiation factor 4B	1	43.02	0.03	1.35
P62805	HIST1H4A	Histone H4	1	58.13	0.03	13.58
P62633	CNBP	Cellular nucleic acid-binding protein	1	61.27	0.03	1.46
Q04760	GLO1	Lactoylglutathione lyase	2	86	0.03	1.16
Q9UQ35	SRRM2	Serine/arginine repetitive matrix protein 2	1	66.77	0.03	1.53
O95218	ZRANB2	Zinc finger Ran-binding domain-containing protein 2	1	45.24	0.03	1.98
P04406	GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	1	91.59	0.03	1.22
P43243	MATR3	Matrin-3	1	53.14	0.03	1.23
Q4VXU2	PABPC1L	Polyadenylate-binding protein 1-like	1	42.91	0.04	1.33
P12268	IMPDH2	Inosine-5'-monophosphate dehydrogenase 2	1	73.26	0.04	1.29
Q01581	HMGCS1	Hydroxymethylglutaryl-CoA synthase, cytoplasmic	1	44.52	0.04	1.28
P35613	BSG	Basigin	1	61.47	0.04	1.32
P16152	CBR1	Carbonyl reductase [NADPH] 1	1	62.77	0.04	1.35
Q8TF72	SHROOM3	Protein Shroom3	1	41.96	0.04	2.85
P08865	RPSA	40S ribosomal protein SA	1	55.38	0.04	1.41
Q15417	CNN3	Calponin-3	1	79	0.05	1.38
P30101	PDIA3	Protein disulfide-isomerase A3	1	46.78	0.05	1.35
Q9NQ29	LUC7L	Putative RNA-binding protein Luc7-like 1	1	54.59	0.05	1.2
P62993	GRB2	Growth factor receptor-bound protein 2	1	40.88	0.05	1.26
Q9UQ80	PA2G4	Proliferation-associated protein 2G4	1	45.05	0.05	1.09
Q99832	CCT7	T-complex protein 1 subunit eta	1	67.22	0.05	1.33

Table 7.7. List of proteins which decreased in expression following over-expression of DHFRL1 in HEK 293 cells.

Accession	Gene	Description	Peptides	Score	Anova (p)	Fold
P07196	NEFL	Neurofilament light polypeptide	3	154.41	6.62E-05	-1.31
Q8NC51	SERBP1	Plasminogen activator inhibitor 1 RNA-binding protein	1	49.12	1.79E-03	-1.46
P07197	NEFM	Neurofilament medium polypeptide	2	109.91	1.98E-03	-1.45
P31040	SDHA	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mt	1	67.27	2.84E-03	-1.18
P09874	PARP1	Poly [ADP-ribose] polymerase 1	1	66.55	3.67E-03	-6.84
Q13263	TRIM28	Transcription intermediary factor 1-beta	1	89.53	4.42E-03	-2.62
O75131	CPNE3	Copine-3	1	41.58	6.23E-03	-2.08
Q9UQE7	SMC3	Structural maintenance of chromosomes protein 3	1	52.08	7.36E-03	-1.5
Q08211	DHX9	ATP-dependent RNA helicase A	1	64.9	7.53E-03	-2.38
O00425	IGF2BP3	Insulin-like growth factor 2 mRNA-binding protein 3	1	49.66	7.89E-03	-1.23
P06744	GPI	Glucose-6-phosphate isomerase	1	43.91	8.89E-03	-1.76
P52209	PGD	6-phosphogluconate dehydrogenase, decarboxylating	1	58.61	9.96E-03	-1.21
P78527	PRKDC	DNA-dependent protein kinase catalytic subunit	1	315.64	0.01	-1.95
Q8NE71	ABCF1	ATP-binding cassette sub-family F member 1	1	46.9	0.01	-1.86
P62316	SNRPD2	Small nuclear ribonucleoprotein Sm D2	1	67.05	0.01	-3.12
Q14980	NUMA1	Nuclear mitotic apparatus protein 1	2	98.73	0.01	-1.55
Q12906	ILF3	Interleukin enhancer-binding factor 3	1	40.31	0.02	-1.24
P35579	MYH9	Myosin-9	1	47.87	0.02	-2.23
P35637	FUS	RNA-binding protein FUS	1	40.54	0.03	-1.66
Q13428	TCOF1	Treacle protein	1	45.91	0.03	-3.04
P24752	ACAT1	Acetyl-CoA acetyltransferase, mitochondrial	1	41.19	0.03	-1.25
P27694	RPA1	Replication protein A 70 kDa DNA-binding subunit	1	61.31	0.03	-1.4
P61981	YWHAG	14-3-3 protein gamma	1	43.44	0.03	-1.52
Q3ZCQ8	TIMM50	Mitochondrial import inner membrane translocase subunit TIM50	1	60.98	0.04	-1.37
P12956	XRCC6	X-ray repair cross-complementing protein 6	1	40.08	0.04	-5.87
P60468	SEC61B	Protein transport protein Sec61 subunit beta	1	51.98	0.04	-2.98

P55084	HADHB	Trifunctional enzyme subunit beta, mitochondrial	1	64.52	0.04	-1.23
O95831	AIFM1	Apoptosis-inducing factor 1, mitochondrial	1	44.13	0.04	-1.4
P56182	RRP1	Ribosomal RNA processing protein 1 homolog A	1	44.02	0.05	-1.86
Q01082	SPTBN1	Spectrin beta chain, non-erythrocytic 1	1	48.34	0.05	-1.73
P30419	NMT1	Glycylpeptide N-tetradecanoyltransferase 1	1	61.54	0.05	-5.08

Table 7.8. List of proteins that are involved in DNA Replication, Recombination and Repair, Nucleic Acid Metabolism, Cell Death and Survival, Cellular Growth and Proliferation and Cell Cycle that are affected by over-expressing DHFRL1 in HEK 293 cells.

Gene	Function	Fold
FUBP1	Cell Death & Survival / Cell Cycle	1.27
ATP5H	Nucleic Acid Metabolism	1.59
HNRNPM	Cellular Growth & Proliferation	1.23
ALDOA	Nucleic Acid Metabolism / Cell Death & Survival	1.36
HSPA8	DNA Replication, Recombination & Repair / Cell Death & Survival	1.67
TXN	DNA Replication, Recombination & Repair / Cell Death & Survival / Cellular Growth & Proliferation	1.26
GNB2L1	Cell Death & Survival / Cellular Growth & Proliferation	1.35
DNAJA2	Cellular Growth & Proliferation	1.42
HSPD1	Nucleic Acid Metabolism / Cell Death & Survival / Cellular Growth & Proliferation	1.37
NCL	Cell Death & Survival / Cellular Growth & Proliferation	1.19
SSBP1	DNA Replication, Recombination & Repair / Cellular Growth & Proliferation	1.19
CSNK2B	Cellular Growth & Proliferation	1.37
STIP1	Nucleic Acid Metabolism	1.22
EIF4B	Cell Death & Survival / Cellular Growth & Proliferation	1.35
GLO1	Cell Death & Survival	1.16

GAPDH	Cell Death & Survival	1.22
IMPDH2	Cell Death & Survival / Cellular Growth & Proliferation	1.29
BSG	DNA Replication, Recombination & Repair / Cell Death & Survival / Cellular Growth & Proliferation	1.32
CBR1	Cell Death & Survival	1.35
RPSA	Cellular Growth & Proliferation	1.41
PDIA3	Cellular Growth & Proliferation	1.35
GRB2	Cellular Growth & Proliferation	1.26
PA2G4	Cellular Growth & Proliferation	1.09
CCT7	Cellular Growth & Proliferation	1.33
NEFL	Cell Cycle	-1.31
SDHA	Cell Death & Survival	-1.18
PARP1	DNA Replication, Recombination & Repair / Cell Death & Survival / Cellular Growth & Proliferation / Cell Cycle	-6.84
TRIM28	DNA Replication, Recombination & Repair / Cell Death & Survival / Cellular Growth & Proliferation	-2.62
SMC3	DNA Replication, Recombination & Repair	-1.5
DHX9	DNA Replication, Recombination & Repair / Nucleic Acid Metabolism	-2.38
IGF2BP3	Cellular Growth & Proliferation	-1.23
GPI	Cell Death & Survival / Cellular Growth & Proliferation	-1.76
PGD	Nucleic Acid Metabolism	-1.21
PRKDC	DNA Replication, Recombination & Repair / Cell Death & Survival / Cell Cycle	-1.95
NUMA1	Cellular Growth & Proliferation / Cell Cycle	-1.55
ILF3	Cellular Growth & Proliferation	-1.24
MYH9	Cell Death & Survival / Cellular Growth & Proliferation	-2.23
FUS	DNA Replication, Recombination & Repair / Cellular Growth & Proliferation / Cell Cycle	-1.66
ACAT1	Cellular Growth & Proliferation	-1.25
RPA1	DNA Replication, Recombination & Repair / Nucleic Acid Metabolism / Cell Death & Survival / Cell Cycle	-1.4
YWHAG	Cell Death & Survival / Cellular Growth & Proliferation	-1.52
TIMM50	Cell Death & Survival	-1.37
XRCC6	DNA Replication, Recombination & Repair / Cell Death & Survival / Cellular Growth & Proliferation	-5.87

HADHB	Cellular Growth & Proliferation	-1.23
AIFM1	DNA Replication, Recombination & Repair / Cell Death & Survival	-1.4

Table 7.9. List of proteins that are involved in glycolysis, oxidative phosphorylation, electron transport, ATP synthesis, and malate / aspartate shuttle that are affected by altering DHFRL1 expression in HEK 293 cells.

Gene	Description	Function	DHFRL1 KD[+]	DHFRL1 OE[ɣ]
ATP5H	ATP synthase subunit d, mitochondrial	ATP synthesis		1.59
NDUFAF4	NADH dehydrogenase 1 alpha subcomplex assembly factor 4	ATP synthesis	-1.33	
NDUFS1	NADH-ubiquinone oxidoreductase 75kDa subunit, mitochondrial	ATP synthesis	-1.29	
NDUFA3	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 3	ATP synthesis	-1.35	
ATP5L	ATP synthase subunit G, mitochondrial	ATP metabolism	-1.93	
DLD	Dihydrolipoyl dehydrogenase, mitochondrial	Glycine Cleavage System	-1.36	
ALDOA	Fructose-bisphosphate aldolase A	Glycolysis		1.36
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	Glycolysis	1.29	1.22
GPI	Glucose-6-phosphate isomerase	Glycolysis		-1.76
TPI1	Triosephosphate isomerase	Glycolysis	1.37	
PGAM1	Phosphoglycerate mutase 1	Glycolysis		2.36
MDH2	Malate dehydrogenase, mitochondrial	Malate / aspartate shuttle		1.13
GOT2	Aspartate aminotransferase, mitochondrial	Malate / aspartate shuttle	-1.38	1.23
SDHA	Succinate dehydrogenase flavoprotein subunit, mitochondrial	Electron transport		-1.18
ALDH5A1	Succinate semialdehyde dehydrogenase, mitochondrial	Electron transport	-1.24	
COX6B1	Cytochrome c oxidase subunit 6B1	Electron transport	1.2	
ACAT1	Acetyl-CoA acetyltransferase, mitochondrial	Oxidative phosphorylation	-1.27	-1.25

[+]KD = knocked-down
[ɣ]OE = over-expressed

Table 7.10. List of proteins that showed that showed altered expression (some opposing, some similar effects) in both DHFRL1 over-expressing and DHFRL1 down-regulated HEK 293 cells.

Gene	Description	L1 OE	L1 KD
GPI	Glucose-6-phosphate isomerase	-1.76	1.29
TRIM28	Transcription intermediary factor 1-beta	-2.62	1.4
PRKDC	DNA-dependent protein kinase catalytic subunit	-1.95	1.75
MYH9	Myosin-9	-2.23	1.43
YWHAG	14-3-3 protein gamma	-1.52	4.4
GOT2	Aspartate aminotransferase, mitochondrial	1.23	-1.38
ZRANB2	Zinc finger Ran-binding domain-containing protein 2	1.98	-3.96
HNRNPM	Heterogeneous nuclear ribonucleoprotein M	1.23	-1.44
PA2G4	Proliferation-associated protein 2G4	1.09	-1.76
NCL	Nucleolin	1.19	-1.46
SRRM2	Serine/arginine repetitive matrix protein 2	1.53	-1.48
BSG	Basigin	1.32	-1.23
GNB2L1	Guanine nucleotide-binding protein subunit beta-2-like 1	1.35	-1.31
SHMT2	Serine hydroxymethyltransferase, mitochondrial	1.37	-1.24
MDH2	Malate dehydrogenase, mitochondrial	1.13	-1.34
GRB2	Growth factor receptor-bound protein 2	1.26	-1.07
SMC3	Structural maintenance of chromosomes protein 3	-1.5	-1.22
SDHA	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	-1.18	-1.22
SPTBN1	Spectrin beta chain, non-erythrocytic 1	-1.73	-1.38
FUS	RNA-binding protein FUS	-1.66	-1.44
DHX9	ATP-dependent RNA helicase A	-2.38	-1.32
AIFM1	Apoptosis-inducing factor 1, mitochondrial	-1.4	1.29
ACAT1	Acetyl-CoA acetyltransferase, mitochondrial	-1.25	-1.27

SNRPD2	Small nuclear ribonucleoprotein Sm D2	-3.12	-1.31
PSMB6	Proteasome subunit beta type-6	1.2	1.38
CCT7	T-complex protein 1 subunit eta	1.33	1.2
ALDOA	Fructose-bisphosphate aldolase A	1.36	1.49
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	1.22	1.29
PSMD4	26S proteasome non-ATPase regulatory subunit 4	1.35	1.38
CNBP	Cellular nucleic acid-binding protein	1.46	1.45

Chapter 8

General Discussion and Future Work

8.1 Discussion

This project focused on investigating the molecular mechanism of the role of folate in human health by 1) examining the impact of folic acid supplementation on gene-specific DNA methylation patterns during pregnancy; 2) exploring the functional relevance of specific polymorphisms from the *DHFR* gene family and 3) investigating the effect of altered DHFRL1 expression on genes / proteins. A schematic representation of the project is shown in Figure 1.8. As mentioned earlier (Section 1.3), the folate mediated one-carbon metabolism pathway is a complex biochemical pathway that is compartmentalised between the cytoplasm, the mitochondria and the nucleus (Section 1.4) (Figure 1.1) (Tibbetts and Appling, 2010, Stover and Field, 2011). It involves numerous enzymes and is essential to the cell and is made up of three main interdependent metabolic pathways; the *de novo* dTMP biosynthesis pathway, the *de novo* purine biosynthesis pathway, which are involved in DNA synthesis and the homocysteine re-methylation pathway (Figure 1.1). All three of these pathways use the cofactor THF to carry the one carbon units that are required for biosynthetic reactions within the cell (Stover, 2004).

The first part of my project focused on the homocysteine re-methylation pathway, which uses 5-methyltetrahydrofolate for the re-methylation of homocysteine to methionine. Methionine is then converted to SAM, which serves as a cofactor for numerous methylation reactions, such as DNA methylation (Anderson et al., 2012c). DNA methylation patterns have been shown to be capable of changing throughout a person's lifetime, especially during reprogramming events associated with development and aging (Waterland and Jirtle, 2003, Fraga et al., 2005, Martin, 2005, Faulk and Dolinoy, 2011), as well as cancer (Feinberg, 2007). Increasing evidence shows that the human *in utero* environment can impact DNA methylation patterns and disease risk of the subsequent offspring. This relates back to the "fetal origins of adult disease" theory originally proposed by Barker (Barker et al., 1989) (Section 1.8). As DNA methylation is dependent on several enzymes and the presence of dietary micronutrients such as folate as cofactors (Anderson et al., 2012c), there is a key association between the two. Therefore, nutri-genetics is emerging as an area of great interest and the role of dietary influences on gene expression is an area of large research interest as DNA methylation events and dietary practices, particularly micronutrient intake may ultimately influence disease phenotype.

Due to this, we examined the DNA methylation patterns of pregnant women, half of whom had taken folic acid supplements throughout their entire pregnancy and those who had only taken folic acid for their first trimester. In our study, we chose to examine a small number of genes, which had previously been shown to be candidates that could potentially be altered in response to folate as well as two genes, *DHFR* and *DHFRL1* that are known to play a major role

in the folate mediated one-carbon metabolism pathway. However, when we examined the *DHFRL1* gene and those genes that had been previously shown to be folate-sensitive in our FASSTT samples, we found that the CpG sites that we examined in these genes were not folate sensitive / were 100 % unmethylated and no DNA methylation pattern differences were seen between the samples. Nonetheless, we examined only a small number of genes and the CpG sites that we examined were not the same as those that had been examined previously. The *DHFR* gene also showed not to be folate sensitive in regards to methylation, however methylation changes were observed between individuals as well as between themselves, pre- and post- intervention, indicating that DNA methylation patterns can change within individuals (Fraga et al., 2005).

Gene-specific methylation studies are a relatively new area of research and have only recently begun to be carried out with very few studies carried out to date examining the relationship between folate and DNA methylation patterns of specific genes. In order for us to examine the relationship between folate and the DNA methylation of specific genes, novel assays were designed using newly developed DNA methylation profiling techniques such as MS-HRM and SMART-MSP. Although the results we obtained from our methylation assays were negative, i.e., no correlation was found between women who took folic acid throughout their entire pregnancy and those who only took folic acid for their first trimester, for the gene regions we examined, these techniques only allow for the analysis of a small region of the gene at a time, ~100-200 bp, and therefore are not perfect assays. As mentioned in Chapter 1, epigenetics is a new a field of science that is emerging and examining the epigenome / DNA methylation is of vital importance in order to help us understand and to account for the diversity observed between the population, in addition to the genome, as well as its relevance for human health and disease. Therefore, new techniques are currently being developed in order to try and assess DNA methylation in the best possible manner. Although, there is no gold standard as of yet, it is most likely to be a sequenced based assay, which would allow genes to be examined at single base resolution. Once these techniques have been developed, it would be of great interest to examine the FASSTT study samples for a large number of genes in order to determine if any DNA methylation differences exist between the samples, allowing us to elucidate the effects of folate on DNA methylation patterns.

The other part of my project focused on the DNA synthesis pathway, which is made up of the *de novo* dTMP biosynthesis pathway and the *de novo* purine biosynthesis pathway (Figure 1.1). As previously discussed (Section 1.11), *DHFR* is one of the most important enzymes that is involved in the folate-mediated one carbon metabolism pathway and plays a major role in purine and dTMP synthesis. Therefore, any polymorphisms found within this gene

are often hypothesised to have severe consequences on the folate pathway, which in turn may impact on an individual's risk to health and disease. Although no polymorphisms have been found within the *DHFR* coding region to date (Litwack, 2008), a 19 bp deletion / insertion polymorphism, located in intron-1 has been studied at length and has been associated with NTDs (Johnson et al., 2004, van der Linden et al., 2007, Parle-McDermott et al., 2007). However, the biological function of this polymorphism as well as its association with circulating metabolites has been controversial throughout the years, and has yet to be determined (Gellekink et al., 2007, Stanisławska-Sachadyn et al., 2008). Therefore, we examined the effect of this 19 bp deletion / insertion polymorphism on circulating folate metabolites and folate intake in the largest population yet to be examined. In our study, we found that the polymorphism was not associated with any of the circulating folate metabolite levels or with folate intake. The reason behind why there are no polymorphisms located within the *DHFR* coding region is likely to relate to the critical role this enzyme plays in cellular function and a dramatic impact on enzyme activity will have severe consequences. This theory is further supported by two recent studies that have identified rare coding mutations (not polymorphisms) in *DHFR* as causative for megaloblastic anaemia and cerebral folate deficiency (Banka et al., 2011, Cario et al., 2011). In relation to the 19 bp deletion / insertion polymorphism; previous work from the laboratory suggested that it might impact on mRNA levels (Parle-McDermott et al., 2007), although the effect was marginal. The lack of any impact on circulating metabolites may indicate that the effect of this polymorphism is not dramatic enough to impact on the overall one carbon balance but doesn't rule out an influence on maternal supply to the embryo during very early development. Further investigation is required to elucidate this. However, a recent finding has shown that individuals with the combined genotype of the *DHFR* 19 bp deletion and the *SHMT* C1420T TT genotypes have increased folate concentrations (Mendes et al., 2013), suggesting that it may be important to examine the combination effects of gene polymorphisms rather than examining them individually.

It has also recently been shown that there is an almost five-fold variation in *DHFR* activity between individuals (Bailey and Ayling, 2009). However, the cause of this variability is yet unknown. Several factors that are thought to be associated with the cause of this variability are MDM2 and *DHFR* polymorphisms. The oncogene product MDM2 has been shown to decrease *DHFR* activity, by binding to and mono-ubiquitinating *DHFR*. This was shown in cell lines that either over-expressed MDM2 or were transiently transfected with MDM2 (Bailey and Ayling, 2009). Polymorphisms within the *DHFR* promoter region as well as the 19 bp deletion / insertion polymorphism have also been candidates and at present, combinations of these non-coding variants in addition to the post-translational modification, ubiquitination are thought to be sufficient in explaining the variability of *DHFR* activity among individuals (Bailey and

Ayling, 2009). *DHFR* promoter methylation may also be a mechanism which may cause variability in *DHFR* activity. *DHFR* promoter methylation was investigated in Chapter 3 and our findings indicated that individuals do differ in their *DHFR* promoter methylation patterns. This could be interesting in itself as it may contribute to the variability of *DHFR* activity among individuals but will have to be investigated further.

On the other hand, the discovery of *DHFRL1* as a functional gene (Anderson et al., 2011, McEntee et al., 2011) has opened up a window of opportunities for investigating the effects of polymorphisms within another folate related gene. *de novo* dTMPs are synthesised through the catalytic conversion of dUMPs to dTMPs by three folate enzymes, SHMT, TYMS and DHFR (Appling, 1991). This reaction takes place in the nucleus during the S phase of the cell cycle (Woeller et al., 2007). dTMP synthesis is required for DNA replication in both the nucleus and the mitochondria and up until recently it was thought that dTMP synthesis in the cytoplasm supported both. However Anderson *et al*. (2011), recently identified that *DHFRL1* takes the place of *DHFR* in the *de novo* thymidylate biosynthesis pathway in the mitochondria (Anderson et al., 2011)

Although it is not yet known how *DHFRL1* is transported and enters the mitochondria, it highlights the importance of mitochondrial folate mediated one-carbon metabolism. The *DHFRL1* gene has similar properties to *DHFR* but in contrast has been shown to have much lower activity in comparison to the *DHFR* gene in a recombinant system (McEntee et al., 2011). However, unlike *DHFR*, examination of the NCBI SNP database indicated that there are non-synonymous polymorphisms within the coding region of *DHFRL1*, which is of great interest. Therefore, we examined two of these polymorphisms, the rs61739170 proline to alanine (C->G) polymorphism and the *DHFRL1* rs17855824 valine to isoleucine (G->A) polymorphism in detail. Analysis of these two *DHFRL1* non-synonymous polymorphisms indicated that they have the ability to impact enzyme activity, with both polymorphisms increasing the specific enzyme activity of the protein in comparison to that of the WT DHFRL1 protein.

The novel finding of the existence of non-synonymous polymorphisms within the *DHFRL1* gene is highly interesting in itself, but the fact that the polymorphisms can alter the enzyme activity of the gene is of great significance, as the majority of polymorphisms found within the coding region of a gene are typically synonymous. This suggests that the polymorphisms that we examined could be functional polymorphisms, which could have major impacts on the function of the gene at the transcriptional and / or the translational level. This theory is further supported by our protein structure prediction (Section 6.2.1), which showed the polymorphisms to alter the 3D structure / folding of DHFRL1. As mentioned earlier, as the

activity of *DHFR* is highly variable among individuals, the same may be the case for *DHFRL1* and these polymorphisms may play a role in this variability. The newly discovered polymorphisms may also help us predict an individual's response to certain drugs, susceptibility to environmental factors such as toxins as well as the risk of developing a particular disease. However, the exact function of these polymorphisms remains to be elucidated.

These polymorphisms may also play a role in relation to disease risk and in Chapter 5, the rs17855824 SNP along with two other DHFRL1 SNPs were shown to be associated with NTDs in an Irish cohort by our collaborators in the NIH and TCD for the first time. This novel finding is of great significance, especially as the polymorphisms that were examined were found to be frequent in the Caucasian population, and will hopefully help us to elucidate the mechanisms behind NTDs. It also further implicates the importance of looking at genetic variation as a risk factor for diseases such as NTDs, with genes involved in the folate-mediated one-carbon metabolism pathway as prime candidates for association studies. Although the mechanism behind how these polymorphisms are affecting NTDs are yet unknown, the polymorphisms may affect NTDs by altering the expression of the gene / protein or as already shown it may be due to the fact the polymorphisms altering the activity of the enzyme. In order to try and determine the functions of the polymorphisms and to elucidate their association with NTDs, they will be investigated further in relation to their function in the laboratory and they will also be screened for risk of NTDs in an UK NTD cohort in collaboration with the NIH and TCD in order to confirm our findings observed in the Irish NTD cohort.

The final part of my project involved the generation of stable DHFRL1 knocked-down and over-expressed cell pools for the first time so that the effects of altered DHFRL1 expression could be examined by proteomic analysis. Data obtained from the proteomic analysis, demonstrated that altered DHFRL1 expression level has an effect on numerous proteins, with a large number of proteins that were identified as being involved in DNA replication, recombination and repair, cell growth and cell death as well as those involved in cancer. Numerous proteins that are involved in embryonic development were also discovered, further supporting our theory that DHFRL1 is associated with NTDs. Proteins that are also involved in NADPH pathways were also identified, strengthening the recent finding by Fan *et al*. (2014), that folate metabolism also plays a role in NADPH production in both the cytoplasm and the mitochondria (Fan et al., 2014). As NADPH and its derivatives play an important role in energy metabolism (Atkinson, 1977), our findings will hopefully help us to identify the relationship between folate metabolism and NADPH pathways. These novel findings have demonstrated that altering the expression of DHFRL1 not only has an effect on proteins that are involved in the folate mediated one carbon metabolism pathway but also on numerous proteins that are

involved in a large number of different biological pathways, signifying the importance of examining the relationship of proteins in numerous pathways.

In conclusion, we have investigated for the first time, the DNA methylation patterns in candidate folate sensitive genes as well as in *DHFR* and *DHFRL1* in order to determine whether prolonged exposure to folic acid supplementation during pregnancy can result in changes in DNA methylation patterns, finding out that the regions we examined are not folate-sensitive and how folate polymorphisms as well as altering the expression of a gene could have an impact in human health and disease by altering the activity / expression of other proteins.

8.2 Future Work

Future work to be carried out in the laboratory will focus on the following areas:

- Further functional characterization of *DHFRL1* polymorphisms

  As mentioned in Chapter 6, the two DHFRL1 polymorphism proteins examined showed to have an increase in specific activity when compared to the WT DHFRL1 protein. The next step would be to determine the Km and Vmax's of the two polymorphisms and compare them to that of the WT in order to further characterise the function of the polymorphisms.

- Impact of recombinant DHFRL1 polymorphic variants on anti-folates, such as methotrexate and aminopterin – what impact do they have on treatment.

  Previous studies that have been carried out in the Parle-McDermott laboratory have shown that DHFRL1 has a lower binding capacity to the anti-folate drug, methotrexate in comparison to DHFR (unpublished data). Therefore, it would be interesting to see whether DHFRL1 polymorphisms have an effect on the binding of methotrexate by enzyme inhibition assays. If the polymorphisms are found to alter the binding affinity of the drugs, this may have a large impact on treating cancer patients with anti-folates in the future.

- How DHFRL1 localises to the mitochondria

  Neither DHFR nor DHFRL1 have mitochondrial targeting sequences. Aligning DHFR and DHFRL1 protein sequences show 16 amino acid differences between the two proteins (Figure 8.1). Mutating these amino acids back to those of DHFR one at a time / in combination may determine how DHFRL1 is localised to the mitochondria. Post-translational modifications such as sumoylation and acetylation could also be examined.

- Examination of thymidylate biosynthesis complexes in the mitochondria

  It has recently been shown that SHMT1, TYMS, DHFR and other proteins in the nucleus form a thymidylate biosynthesis complex that is associated with the nuclear

lamina (Anderson et al., 2012a). It would be interesting to determine whether a similar complex is also formed in the mitochondria. This could be carried out by using co-immunoprecipitation and tandem affinity purification assays, using monoclonal antibodies for DHFRL1 to pull down DHFRL1 / DHFRL1 enzyme complexes. These complexes could then be analysed by LC-MS in order to identify proteins which have formed a complex.

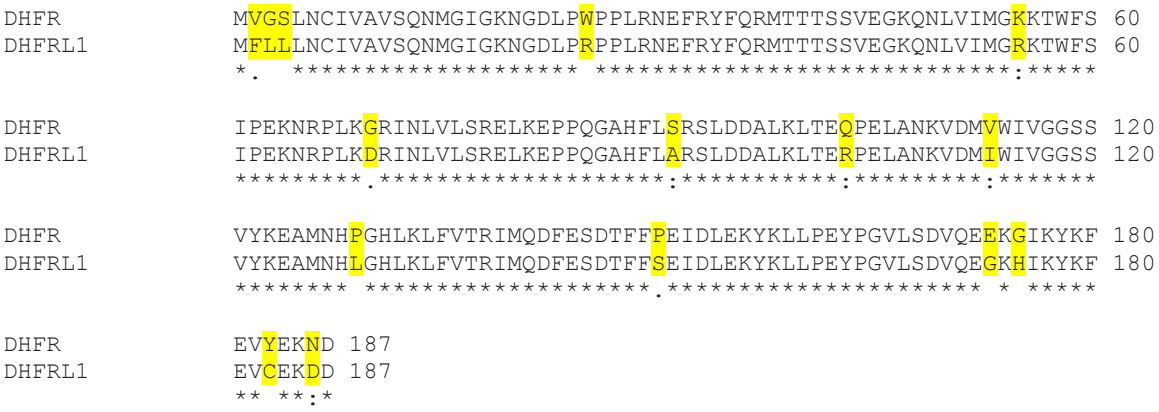


Figure 8.1. A protein sequence alignment of human DHFR and DHFRL1 generated using clustalW2. Amino acids highlighted in yellow indicate those that differ between the two proteins.

References

References

1991. Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. MRC Vitamin Study Research Group. *Lancet,* 338**,** 131-7.

1992. Recommendations for the use of folic acid to reduce the number of cases of spina bifida and other neural tube defects. *MMWR Recomm Rep,* 41**,** 1-7.

(WHO), F. A. A. O. O. T. U. N. F. A. W. H. O. 2002. Human vitamin and mineral requirements: report of a joint FAO/WHO expert consultation. Rome, Italy.

AFMAN, L. A., BLOM, H. J., DRITTIJ, M. J., BROUNS, M. R. & VAN STRAATEN, H. W. 2005. Inhibition of transmethylation disturbs neurulation in chick embryos. *Brain Res Dev Brain Res,* 158**,** 59-65.

AMIN, K. S. & BANERJEE, P. P. 2012. The cellular functions of RASSF1A and its inactivation in prostate cancer. *J Carcinog,* 11**,** 3.

ANAGNOU, N. P., ANTONARAKIS, S. E., O'BRIEN, S. J., MODI, W. S. & NIENHUIS, A. W. 1988. Chromosomal localization and racial distribution of the polymorphic human dihydrofolate reductase pseudogene (DHFRP1). *Am J Hum Genet,* 42**,** 345-52.

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. *Proc Natl Acad Sci U S A,* 81**,** 5170-4.

ANDERSON, D. D., EOM, J. Y. & STOVER, P. J. 2012a. Competition between sumoylation and ubiquitination of serine hydroxymethyltransferase 1 determines its nuclear localization and its accumulation in the nucleus. *J Biol Chem,* 287**,** 4790-9.

ANDERSON, D. D., QUINTERO, C. M. & STOVER, P. J. 2011. Identification of a de novo thymidylate biosynthesis pathway in mammalian mitochondria. *Proc Natl Acad Sci U S A,* 108**,** 15163-8.

ANDERSON, D. D., WOELLER, C. F., CHIANG, E. P., SHANE, B. & STOVER, P. J. 2012b. Serine hydroxymethyltransferase anchors de novo thymidylate synthesis pathway to nuclear lamina for DNA synthesis. *J Biol Chem,* 287**,** 7051-62.

ANDERSON, D. D., WOELLER, C. F. & STOVER, P. J. 2007. Small ubiquitin-like modifier-1 (SUMO-1) modification of thymidylate synthase and dihydrofolate reductase. *Clin Chem Lab Med,* 45**,** 1760-3.

ANDERSON, O. S., SANT, K. E. & DOLINOY, D. C. 2012c. Nutrition and epigenetics: an interplay of dietary methyl donors, one-carbon metabolism and DNA methylation. *J Nutr Biochem,* 23**,** 853-9.

ANTONY, A. C. 1996. Folate receptors. *Annu Rev Nutr,* 16**,** 501-21.

APPLING, D. R. 1991. Compartmentation of folate-mediated one-carbon metabolism in eukaryotes. *FASEB J,* 5**,** 2645-51.

ASKARI, B. S. & KRAJINOVIC, M. 2010. Dihydrofolate reductase gene variations in susceptibility to disease and treatment outcomes. *Curr Genomics,* 11**,** 578-83.

ATKINSON, D. E. 1977. *Cellular Energy Metabolism and Its Regulation*, Academic Press Inc.

BAILEY, L. B. & GREGORY, J. F. 1999. Folate metabolism and requirements. *J Nutr,* 129**,** 779-82.

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. *Proc Natl Acad Sci U S A,* 106**,** 15424-9.

BALL, M. P., LI, J. B., GAO, Y., LEE, J. H., LEPROUST, E. M., PARK, I. H., XIE, B., DALEY, G. Q. & CHURCH, G. M. 2009. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol,* 27**,** 361-8.

BANERJEE, D., MAYER-KUCKUK, P., CAPIAUX, G., BUDAK-ALPDOGAN, T., GORLICK, R. & BERTINO, J. R. 2002. Novel aspects of resistance to drugs targeted to dihydrofolate reductase and thymidylate synthase. *Biochim Biophys Acta,* 1587**,** 164-73.

BANKA, S., BLOM, H. J., WALTER, J., AZIZ, M., URQUHART, J., CLOUTHIER, C. M., RICE, G. I., DE BROUWER, A. P., HILTON, E., VASSALLO, G., WILL, A., SMITH, D. E., SMULDERS, Y. M., WEVERS, R. A., STEINFELD, R., HEALES, S., CROW, Y. J., PELLETIER, J. N., JONES, S. & NEWMAN, W. G. 2011. Identification and characterization of an inborn error of metabolism caused by dihydrofolate reductase deficiency. *Am J Hum Genet,* 88**,** 216-25.

BARKER, D. J. 1993. Fetal origins of coronary heart disease. *Br Heart J,* 69**,** 195-6.

BARKER, D. J. 1997. Fetal nutrition and cardiovascular disease in later life. *Br Med Bull,* 53**,** 96-108.

BARKER, D. J. & MARTYN, C. N. 1992. The maternal and fetal origins of cardiovascular disease. *J Epidemiol Community Health,* 46**,** 8-11.

BARKER, D. J., WINTER, P. D., OSMOND, C., MARGETTS, B. & SIMMONDS, S. J. 1989. Weight in infancy and death from ischaemic heart disease. *Lancet,* 2**,** 577-80.

BAZZANO, L. A., REYNOLDS, K., HOLDER, K. N. & HE, J. 2006. Effect of folic acid supplementation on risk of cardiovascular diseases: a meta-analysis of randomized controlled trials. *JAMA,* 296**,** 2720-6.

BEARD, W. A., APPLEMAN, J. R., HUANG, S. M., DELCAMP, T. J., FREISHEIM, J. H. & BLAKLEY, R. L. 1991. Role of the conserved active site residue tryptophan-24 of human dihydrofolate reductase as revealed by mutagenesis. *Biochemistry,* 30**,** 1432-40.

BEAUDIN, A. E. & STOVER, P. J. 2009. Insights into metabolic mechanisms underlying folate-responsive neural tube defects: a minireview. *Birth Defects Res A Clin Mol Teratol,* 85**,** 274-84.

BECK, C., ROBERT, I., REINA-SAN-MARTIN, B., SCHREIBER, V. & DANTZER, F. 2014. Poly(ADP-ribose) polymerases in double-strand break repair: Focus on PARP1, PARP2 and PARP3. *Exp Cell Res*.

BELL, J. T., PAI, A. A., PICKRELL, J. K., GAFFNEY, D. J., PIQUE-REGI, R., DEGNER, J. F., GILAD, Y. & PRITCHARD, J. K. 2011. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol,* 12**,** R10.

BERRY, R. J., LI, Z., ERICKSON, J. D., LI, S., MOORE, C. A., WANG, H., MULINARE, J., ZHAO, P., WONG, L. Y., GINDLER, J., HONG, S. X. & CORREA, A. 1999. Prevention of neural-tube defects with folic acid in China. China-U.S. Collaborative Project for Neural Tube Defect Prevention. *N Engl J Med,* 341**,** 1485-90.

BERTRAM, C. E. & HANSON, M. A. 2001. Animal models and programming of the metabolic syndrome. *Br Med Bull,* 60**,** 103-21.

BIBIKOVA, M., GOLIC, M., GOLIC, K. G. & CARROLL, D. 2002. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics,* 161**,** 1169-75.

BLAKLEY, R. L. & SORRENTINO, B. P. 1998. In vitro mutations in dihydrofolate reductase that confer resistance to methotrexate: potential for clinical application. *Hum Mutat,* 11**,** 259-63.

BLOM, H. J., SHAW, G. M., DEN HEIJER, M. & FINNELL, R. H. 2006. Neural tube defects and folate: case far from closed. *Nat Rev Neurosci,* 7**,** 724-31.

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. *Proc Natl Acad Sci U S A,* 94**,** 3290-5.

BOTTIGLIERI, T. 2005. Homocysteine and folate metabolism in depression. *Prog Neuropsychopharmacol Biol Psychiatry,* 29**,** 1103-12.

BOTTO, L. D., MULINARE, J. & ERICKSON, J. D. 2003. Do multivitamin or folic acid supplements reduce the risk for congenital heart defects? Evidence and gaps. *Am J Med Genet A,* 121A**,** 95-101.

BOTTO, L. D. & YANG, Q. 2000. 5,10-Methylenetetrahydrofolate reductase gene variants and congenital anomalies: a HuGE review. *Am J Epidemiol,* 151**,** 862-77.

BRODY, L. C., CONLEY, M., COX, C., KIRKE, P. N., MCKEEVER, M. P., MILLS, J. L., MOLLOY, A. M., O'LEARY, V. B., PARLE-MCDERMOTT, A., SCOTT, J. M. & SWANSON, D. A. 2002. A polymorphism, R653Q, in the trifunctional enzyme methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase/formyltetrahydrofolate synthetase is a maternal genetic risk factor for neural tube defects: report of the Birth Defects Research Group. *Am J Hum Genet,* 71**,** 1207-15.

BUSHMAN, W., THOMPSON, J. F., VARGAS, L. & LANDY, A. 1985. Control of directionality in lambda site specific recombination. *Science,* 230**,** 906-11.

CARIO, H., SMITH, D. E., BLOM, H., BLAU, N., BODE, H., HOLZMANN, K., PANNICKE, U., HOPFNER, K. P., RUMP, E. M., AYRIC, Z., KOHNE, E., DEBATIN, K. M., SMULDERS, Y. & SCHWARZ, K. 2011. Dihydrofolate reductase deficiency due to a homozygous DHFR mutation causes megaloblastic anemia and cerebral folate deficiency leading to severe neurologic disease. *Am J Hum Genet,* 88**,** 226-31.

CARMELL, M. A., XUAN, Z., ZHANG, M. Q. & HANNON, G. J. 2002. The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev,* 16**,** 2733-42.

CARROLL, N., HUGHES, L., MCENTEE, G. & PARLE-MCDERMOTT, A. 2012. Investigation of the molecular response to folate metabolism inhibition. *J Nutr Biochem,* 23**,** 1531-6.

CHAKRABORTY, P. & GROSSE, F. 2011. Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair (Amst),* 10**,** 654-65.

CHANDRAN, V., SIANNIS, F., RAHMAN, P., PELLETT, F. J., FAREWELL, V. T. & GLADMAN, D. D. 2010. Folate pathway enzyme gene polymorphisms and the efficacy and toxicity of methotrexate in psoriatic arthritis. *J Rheumatol,* 37**,** 1508-12.

CHEN, C. C., HWANG, J. K. & YANG, J. M. 2006. (PS)2: protein structure prediction server. *Nucleic Acids Res,* 34**,** W152-7.

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. *J Biol Chem,* 259**,** 3933-43.

CHOI, S. W. & MASON, J. B. 2000. Folate and carcinogenesis: an integrated scheme. *J Nutr,* 130**,** 129-32.

CHOI, S. W. & MASON, J. B. 2002. Folate status: effects on pathways of colorectal carcinogenesis. *J Nutr,* 132**,** 2413S-2418S.

CHOUDHURI, S., CUI, Y. & KLAASSEN, C. D. 2010. Molecular targets of epigenetic regulation and effectors of environmental influences. *Toxicol Appl Pharmacol,* 245**,** 378-93.

CLARK, I. M., ROWAN, A. D., EDWARDS, D. R., BECH-HANSEN, T., MANN, D. A., BAHR, M. J. & CAWSTON, T. E. 1997. Transcriptional activity of the human tissue inhibitor of metalloproteinases 1 (TIMP-1) gene in fibroblasts involves elements in the promoter, exon 1 and intron 1. *Biochem J,* 324 (Pt 2)**,** 611-7.

CLARKE, R., BENNETT, D. A., PARISH, S., VERHOEF, P., DÖTSCH-KLERK, M., LATHROP, M., XU, P., NORDESTGAARD, B. G., HOLM, H., HOPEWELL, J. C., SALEHEEN, D., TANAKA, T., ANAND, S. S., CHAMBERS, J. C., KLEBER, M. E., OUWEHAND, W. H., YAMADA, Y., ELBERS, C., PETERS, B., STEWART, A. F., REILLY, M. M., THORAND, B., YUSUF, S., ENGERT, J. C., ASSIMES, T. L., KOONER, J., DANESH, J., WATKINS, H., SAMANI, N. J., COLLINS, R., PETO, R. & GROUP, M. S. C. 2012. Homocysteine and coronary heart disease: meta-analysis of MTHFR case-control studies, avoiding publication bias. *PLoS Med,* 9**,** e1001177.

COLE, B. F., BARON, J. A., SANDLER, R. S., HAILE, R. W., AHNEN, D. J., BRESALIER, R. S., MCKEOWN-EYSSEN, G., SUMMERS, R. W., ROTHSTEIN, R. I., BURKE, C. A., SNOVER, D. C., CHURCH, T. R., ALLEN, J. I., ROBERTSON, D. J., BECK, G. J., BOND, J. H., BYERS, T.,

MANDEL, J. S., MOTT, L. A., PEARSON, L. H., BARRY, E. L., REES, J. R., MARCON, N., SAIBIL, F., UELAND, P. M., GREENBERG, E. R. & GROUP, P. P. S. 2007. Folic acid for the prevention of colorectal adenomas: a randomized clinical trial. *JAMA,* 297**,** 2351-9.

COLLABORATION, H. S. 2002. Homocysteine and risk of ischemic heart disease and stroke: a meta-analysis. *JAMA,* 288**,** 2015-22.

CZEIZEL, A. E. & DUDÁS, I. 1992. Prevention of the first occurrence of neural-tube defects by periconceptional vitamin supplementation. *N Engl J Med,* 327**,** 1832-5.

CZEIZEL, A. E., DUDÁS, I., VERECZKEY, A. & BÁNHIDY, F. 2013. Folate deficiency and folic acid supplementation: the prevention of neural-tube defects and congenital heart defects. *Nutrients,* 5**,** 4760-75.

CZEIZEL, A. E., PUHÓ, E., SØRENSEN, H. T. & OLSEN, J. 2004. Possible association between different congenital abnormalities and use of different sulfonamides during pregnancy. *Congenit Anom (Kyoto),* 44**,** 79-86.

CZEIZEL, A. E., TÓTH, M. & ROCKENBAUER, M. 1996. Population-based case control study of folic acid supplementation during pregnancy. *Teratology,* 53**,** 345-51.

DALY, L. E., KIRKE, P. N., MOLLOY, A., WEIR, D. G. & SCOTT, J. M. 1995. Folate levels and neural tube defects. Implications for prevention. *JAMA,* 274**,** 1698-702.

DAY, D. A. & TUITE, M. F. 1998. Post-transcriptional gene regulatory mechanisms in eukaryotes: an overview. *J Endocrinol,* 157**,** 361-71.

DE MARCO, P., MERELLO, E., CAMA, A., KIBAR, Z. & CAPRA, V. 2011. Human neural tube defects: genetic causes and prevention. *Biofactors,* 37**,** 261-8.

DOLINOY, D. C., DAS, R., WEIDMAN, J. R. & JIRTLE, R. L. 2007. Metastable epialleles, imprinting, and the fetal origins of adult diseases. *Pediatr Res,* 61**,** 30R-37R.

DUAN, W., LADENHEIM, B., CUTLER, R. G., KRUMAN, I. I., CADET, J. L. & MATTSON, M. P. 2002. Dietary folate deficiency and elevated homocysteine levels endanger dopaminergic neurons in models of Parkinson's disease. *J Neurochem,* 80**,** 101-10.

DULUCQ, S., ST-ONGE, G., GAGNÉ, V., ANSARI, M., SINNETT, D., LABUDA, D., MOGHRABI, A. & KRAJINOVIC, M. 2008. DNA variants in the dihydrofolate reductase gene and outcome in childhood ALL. *Blood,* 111**,** 3692-700.

DUNLEVY, L. P., BURREN, K. A., CHITTY, L. S., COPP, A. J. & GREENE, N. D. 2006a. Excess methionine suppresses the methylation cycle and inhibits neural tube closure in mouse embryos. *FEBS Lett,* 580**,** 2803-7.

DUNLEVY, L. P., BURREN, K. A., MILLS, K., CHITTY, L. S., COPP, A. J. & GREENE, N. D. 2006b. Integrity of the methylation cycle is essential for mammalian neural tube closure. *Birth Defects Res A Clin Mol Teratol,* 76**,** 544-52.

DUTHIE, S. J. 2011. Folate and cancer: how DNA damage, repair and methylation impact on colon carcinogenesis. *J Inherit Metab Dis,* 34**,** 101-9.

DWIVEDI, M. K., TRIPATHI, A. K., SHUKLA, S., KHAN, S. & CHAUHAN, U. K. 2011. Homocysteine and cardiovascular disease. *Biotechnology and Molecular Biology Review,* 5(5)**,** 101-107.

ECKHARDT, F., BECK, S., GUT, I. G. & BERLIN, K. 2004. Future potential of the Human Epigenome Project. *Expert Rev Mol Diagn,* 4**,** 609-18.

ESTELLER, M. 2005. Aberrant DNA methylation as a cancer-inducing mechanism. *Annu Rev Pharmacol Toxicol,* 45**,** 629-56.

FAN, J., YE, J., KAMPHORST, J. J., SHLOMI, T., THOMPSON, C. B. & RABINOWITZ, J. D. 2014. Quantitative flux analysis reveals folate-dependent NADPH production. *Nature,* 510**,** 298-302.

FAULK, C. & DOLINOY, D. C. 2011. Timing is everything: the when and how of environmentally induced changes in the epigenome of animals. *Epigenetics,* 6**,** 791-7.

FEINBERG, A. P. 2007. Phenotypic plasticity and the epigenetics of human disease. *Nature,* 447**,** 433-40.

FIGUEIREDO, J. C., GRAU, M. V., HAILE, R. W., SANDLER, R. S., SUMMERS, R. W., BRESALIER, R. S., BURKE, C. A., MCKEOWN-EYSSEN, G. E. & BARON, J. A. 2009. Folic acid and risk of prostate cancer: results from a randomized clinical trial. *J Natl Cancer Inst,* 101**,** 432-5.

FIGUEIREDO, J. C., MOTT, L. A., GIOVANNUCCI, E., WU, K., COLE, B., GRAINGE, M. J., LOGAN, R. F. & BARON, J. A. 2011. Folic acid and prevention of colorectal adenomas: a combined analysis of randomized clinical trials. *Int J Cancer,* 129**,** 192-203.

FOOD AND NUTRITION BOARDS, I. O. M. I. 2004. Dietary reference intakes (DRIs) : recommended intakes for individuals, vitamins. Washington D.C.: National Academy of Sciences.

FOX, J. T. & STOVER, P. J. 2008. Folate-mediated one-carbon metabolism. *Vitam Horm,* 79**,** 1-44.

FRAGA, M. F., BALLESTAR, E., PAZ, M. F., ROPERO, S., SETIEN, F., BALLESTAR, M. L., HEINE-SUÑER, D., CIGUDOSA, J. C., URIOSTE, M., BENITEZ, J., BOIX-CHORNET, M., SANCHEZ-AGUILERA, A., LING, C., CARLSSON, E., POULSEN, P., VAAG, A., STEPHAN, Z., SPECTOR, T. D., WU, Y. Z., PLASS, C. & ESTELLER, M. 2005. Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A,* 102**,** 10604-9.

FRISO, S., CHOI, S. W., GIRELLI, D., MASON, J. B., DOLNIKOWSKI, G. G., BAGLEY, P. J., OLIVIERI, O., JACQUES, P. F., ROSENBERG, I. H., CORROCHER, R. & SELHUB, J. 2002. A common mutation in the 5,10-methylenetetrahydrofolate reductase gene affects genomic DNA methylation through an interaction with folate status. *Proc Natl Acad Sci U S A,* 99**,** 5606-11.

FRYER, A. A., EMES, R. D., ISMAIL, K. M., HAWORTH, K. E., MEIN, C., CARROLL, W. D. & FARRELL, W. E. 2011. Quantitative, high-resolution epigenetic profiling of CpG loci identifies associations with cord blood plasma homocysteine and birth weight in humans. *Epigenetics,* 6.

GELLEKINK, H., BLOM, H. J., VAN DER LINDEN, I. J. & DEN HEIJER, M. 2007. Molecular genetic analysis of the human dihydrofolate reductase gene: relation with plasma total homocysteine, serum and red blood cell folate levels. *Eur J Hum Genet,* 15**,** 103-9.

GEMMATI, D., DE MATTEI, M., CATOZZI, L., DELLA PORTA, M., SERINO, M. L., AMBROSIO, C., CUNEO, A., FRISO, S., KRAMPERA, M., ORIOLI, E., ZERI, G. & ONGARO, A. 2009. DHFR 19-bp insertion/deletion polymorphism and MTHFR C677T in adult acute lymphoblastic leukaemia: is the risk reduction due to intracellular folate unbalancing? *Am J Hematol,* 84**,** 526-9.

GEORGE, R. A. & HERINGA, J. 2002. An analysis of protein domain linkers: their classification and role in protein folding. *Protein Eng,* 15**,** 871-9.

GIBBS, J. R., VAN DER BRUG, M. P., HERNANDEZ, D. G., TRAYNOR, B. J., NALLS, M. A., LAI, S. L., AREPALLI, S., DILLMAN, A., RAFFERTY, I. P., TRONCOSO, J., JOHNSON, R., ZIELKE, H. R., FERRUCCI, L., LONGO, D. L., COOKSON, M. R. & SINGLETON, A. B. 2010. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet,* 6**,** e1000952.

GIOVANNUCCI, E. 2002. Epidemiologic studies of folate and colorectal neoplasia: a review. *J Nutr,* 132**,** 2350S-2355S.

GIOVANNUCCI, E., STAMPFER, M. J., COLDITZ, G. A., HUNTER, D. J., FUCHS, C., ROSNER, B. A., SPEIZER, F. E. & WILLETT, W. C. 1998. Multivitamin use, folate, and colon cancer in women in the Nurses' Health Study. *Ann Intern Med,* 129**,** 517-24.

GLUCKMAN, P. D., HANSON, M. A., COOPER, C. & THORNBURG, K. L. 2008. Effect of in utero and early-life conditions on adult health and disease. *N Engl J Med,* 359**,** 61-73.

GODFREY, K. M. & BARKER, D. J. 2001. Fetal programming and adult health. *Public Health Nutr,* 4**,** 611-24.

GODFREY, P. S., TOONE, B. K., CARNEY, M. W., FLYNN, T. G., BOTTIGLIERI, T., LAUNDY, M., CHANARIN, I. & REYNOLDS, E. H. 1990. Enhancement of recovery from psychiatric illness by methylfolate. *Lancet,* 336**,** 392-5.

GOTO, Y., YUE, L., YOKOI, A., NISHIMURA, R., UEHARA, T., KOIZUMI, S. & SAIKAWA, Y. 2001. A novel single-nucleotide polymorphism in the 3'-untranslated region of the human dihydrofolate reductase gene with enhanced expression. *Clin Cancer Res,* 7**,** 1952-6.

GUTIERREZ-ARCELUS, M., LAPPALAINEN, T., MONTGOMERY, S. B., BUIL, A., ONGEN, H., YUROVSKY, A., BRYOIS, J., GIGER, T., ROMANO, L., PLANCHON, A., FALCONNET, E., BIELSER, D., GAGNEBIN, M., PADIOLEAU, I., BOREL, C., LETOURNEAU, A., MAKRYTHANASIS, P., GUIPPONI, M., GEHRIG, C., ANTONARAKIS, S. E. & DERMITZAKIS, E. T. 2013. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *Elife,* 2**,** e00523.

GUÉRIN, S. L., LECLERC, S., VERREAULT, H., LABRIE, F. & LUU-THE, V. 1995. Overlapping cis-acting elements located in the first intron of the gene for type I 3 beta-hydroxysteroid dehydrogenase modulate its transcriptional activity. *Mol Endocrinol,* 9**,** 1583-97.

HACKETT, J. A., SENGUPTA, R., ZYLICZ, J. J., MURAKAMI, K., LEE, C., DOWN, T. A. & SURANI, M. A. 2013. Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science,* 339**,** 448-52.

HAGGARTY, P., HOAD, G., CAMPBELL, D. M., HORGAN, G. W., PIYATHILAKE, C. & MCNEILL, G. 2013. Folate in pregnancy and imprinted gene and repeat element methylation in the offspring. *Am J Clin Nutr,* 97**,** 94-9.

HALES, C. N., BARKER, D. J., CLARK, P. M., COX, L. J., FALL, C., OSMOND, C. & WINTER, P. D. 1991. Fetal and infant growth and impaired glucose tolerance at age 64. *BMJ,* 303**,** 1019-22.

HANASH, S. 2003. Disease proteomics. *Nature,* 422**,** 226-32.

HECHT, D., TRAN, J. & FOGEL, G. B. 2011. Structural-based analysis of dihydrofolate reductase evolution. *Mol Phylogenet Evol,* 61**,** 212-30.

HEIJMANS, B. T., TOBI, E. W., STEIN, A. D., PUTTER, H., BLAUW, G. J., SUSSER, E. S., SLAGBOOM, P. E. & LUMEY, L. H. 2008. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A,* 105**,** 17046-9.

HELLMAN, A. & CHESS, A. 2007. Gene body-specific methylation on the active X chromosome. *Science,* 315**,** 1141-3.

HENDRICKSON, S. L., WU, J. S. & JOHNSON, L. F. 1980. Cell cycle regulation of dihydrofolate reductase mRNA metabolism in mouse fibroblasts. *Proc Natl Acad Sci U S A,* 77**,** 5140-4.

HEYWOOD, W. E., MIAN, N., MILLA, P. J. & LINDLEY, K. J. 2004. Programming of defective rat pancreatic beta-cell function in offspring from mothers fed a low-protein diet during gestation and the suckling periods. *Clin Sci (Lond),* 107**,** 37-45.

HIRAGA, S., OGURA, T., MORI, H. & TANAKA, M. 1985. Mechanisms essential for stable inheritance of mini-F plasmid. *Basic Life Sci,* 30**,** 469-87.

HO, P. I., COLLINS, S. C., DHITAVAT, S., ORTIZ, D., ASHLINE, D., ROGERS, E. & SHEA, T. B. 2001. Homocysteine potentiates beta-amyloid neurotoxicity: role of oxidative stress. *J Neurochem,* 78**,** 249-53.

HOEK, H. W., SUSSER, E., BUCK, K. A., LUMEY, L. H., LIN, S. P. & GORMAN, J. M. 1996. Schizoid personality disorder after prenatal exposure to famine. *Am J Psychiatry,* 153**,** 1637-9.

HÅBERG, S. E., LONDON, S. J., STIGUM, H., NAFSTAD, P. & NYSTAD, W. 2009. Folic acid supplements in pregnancy and early childhood respiratory health. *Arch Dis Child,* 94**,** 180-4.

IFERGAN, I. & ASSARAF, Y. G. 2008. Molecular mechanisms of adaptation to folate deficiency. *Vitam Horm,* 79**,** 99-143.

IYENGAR, S. & FARNHAM, P. J. 2011. KAP1 protein: an enigmatic master regulator of the genome. *J Biol Chem,* 286**,** 26267-76.

IYER, R. & TOMAR, S. K. 2009. Folate: a functional food constituent. *J Food Sci,* 74**,** R114-22.

IZMIRLI, M. 2013. A literature review of MTHFR (C677T and A1298C polymorphisms) and cancer risk. *Mol Biol Rep,* 40**,** 625-37.

JACOB, R. A., GRETZ, D. M., TAYLOR, P. C., JAMES, S. J., POGRIBNY, I. P., MILLER, B. J., HENNING, S. M. & SWENDSEID, M. E. 1998. Moderate folate depletion increases plasma homocysteine and decreases lymphocyte DNA methylation in postmenopausal women. *J Nutr,* 128**,** 1204-12.

JENSEN, D. E., BLACK, A. R., SWICK, A. G. & AZIZKHAN, J. C. 1997. Distinct roles for Sp1 and E2F sites in the growth/cell cycle regulation of the DHFR promoter. *J Cell Biochem,* 67**,** 24-31.

JIRTLE, R. L. & SKINNER, M. K. 2007. Environmental epigenomics and disease susceptibility. *Nat Rev Genet,* 8**,** 253-62.

JOHNSON, W. G., STENROOS, E. S., SPYCHALA, J. R., CHATKUPT, S., MING, S. X. & BUYSKE, S. 2004. New 19 bp deletion polymorphism in intron-1 of dihydrofolate reductase (DHFR): a risk factor for spina bifida acting in mothers during pregnancy? *Am J Med Genet A,* 124A**,** 339-45.

JONES, P. A. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet,* 13**,** 484-92.

KALMBACH, R. D., CHOUMENKOVITCH, S. F., TROEN, A. P., JACQUES, P. F., D'AGOSTINO, R. & SELHUB, J. 2008. A 19-base pair deletion polymorphism in dihydrofolate reductase is associated with increased unmetabolized folic acid in plasma and decreased red blood cell folate. *J Nutr,* 138**,** 2323-7.

KAMINSKY, Z. A., TANG, T., WANG, S. C., PTAK, C., OH, G. H., WONG, A. H., FELDCAMP, L. A., VIRTANEN, C., HALFVARSON, J., TYSK, C., MCRAE, A. F., VISSCHER, P. M., MONTGOMERY, G. W., GOTTESMAN, I. I., MARTIN, N. G. & PETRONIS, A. 2009. DNA methylation profiles in monozygotic and dizygotic twins. *Nat Genet,* 41**,** 240-5.

KIKUCHI, G., MOTOKAWA, Y., YOSHIDA, T. & HIRAGA, K. 2008. Glycine cleavage system: reaction mechanism, physiological significance, and hyperglycinemia. *Proc Jpn Acad Ser B Phys Biol Sci,* 84**,** 246-63.

KIM, H. & KIM, J. S. 2014. A guide to genome engineering with programmable nucleases. *Nat Rev Genet,* 15**,** 321-34.

KIM, M., LONG, T. I., ARAKAWA, K., WANG, R., YU, M. C. & LAIRD, P. W. 2010. DNA methylation as a biomarker for cardiovascular disease risk. *PLoS One,* 5**,** e9692.

KIM, Y. I. 2003. Role of folate in colon cancer development and progression. *J Nutr,* 133**,** 3731S-3739S.

KIRKE, P. N., MOLLOY, A. M., DALY, L. E., BURKE, H., WEIR, D. G. & SCOTT, J. M. 1993. Maternal plasma folate and vitamin B12 are independent risk factors for neural tube defects. *Q J Med,* 86**,** 703-8.

KLOSE, R. J. & BIRD, A. P. 2006. Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci,* 31**,** 89-97.

KRIAUCIONIS, S. & HEINTZ, N. 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science,* 324**,** 929-30.

KRISTENSEN, L. S., MIKESKA, T., KRYPUY, M. & DOBROVIC, A. 2008. Sensitive Melting Analysis after Real Time- Methylation Specific PCR (SMART-MSP): high-throughput and probe-free quantitative DNA methylation detection. *Nucleic Acids Res,* 36**,** e42.

KRISTENSEN, L. S., RAYNOR, M. P., CANDILORO, I. & DOBROVIC, A. 2012. Methylation profiling of normal individuals reveals mosaic promoter methylation of cancer-associated genes. *Oncotarget,* 3**,** 450-61.

KRISTENSEN, L. S., WOJDACZ, T. K., THESTRUP, B. B., WIUF, C., HAGER, H. & HANSEN, L. L. 2009. Quality assessment of DNA derived from up to 30 years old formalin fixed paraffin embedded (FFPE) tissue for PCR-based methylation analysis using SMART-MSP and MS-HRM. *BMC Cancer,* 9**,** 453.

KWONG, W. Y., WILD, A. E., ROBERTS, P., WILLIS, A. C. & FLEMING, T. P. 2000. Maternal undernutrition during the preimplantation period of rat development causes

blastocyst abnormalities and programming of postnatal hypertension. *Development,* 127**,** 4195-202.
LANDY, A. 1989. Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu Rev Biochem,* 58**,** 913-49.
LEAHY, S. C., HIGGINS, D. G., FITZGERALD, G. F. & VAN SINDEREN, D. 2005. Getting better with bifidobacteria. *J Appl Microbiol,* 98**,** 1303-15.
LEE, S. W., LEE, M. H., PARK, J. H., KANG, S. H., YOO, H. M., KA, S. H., OH, Y. M., JEON, Y. J. & CHUNG, C. H. 2012. SUMOylation of hnRNP-K is required for p53-mediated cell-cycle arrest in response to DNA damage. *EMBO J,* 31**,** 4441-52.
LEVINE, A. J., FIGUEIREDO, J. C., LEE, W., CONTI, D. V., KENNEDY, K., DUGGAN, D. J., POYNTER, J. N., CAMPBELL, P. T., NEWCOMB, P., MARTINEZ, M. E., HOPPER, J. L., LE MARCHAND, L., BARON, J. A., LIMBURG, P. J., ULRICH, C. M. & HAILE, R. W. 2010. A candidate gene study of folate-associated one carbon metabolism genes and colorectal cancer risk. *Cancer Epidemiol Biomarkers Prev,* 19**,** 1812-21.
LIENERT, F., WIRBELAUER, C., SOM, I., DEAN, A., MOHN, F. & SCHÜBELER, D. 2011. Identification of genetic elements that autonomously determine DNA methylation states. *Nat Genet,* 43**,** 1091-7.
LILLYCROP, K. A., PHILLIPS, E. S., JACKSON, A. A., HANSON, M. A. & BURDGE, G. C. 2005. Dietary protein restriction of pregnant rats induces and folic acid supplementation prevents epigenetic modification of hepatic gene expression in the offspring. *J Nutr,* 135**,** 1382-6.
LISTER, R., PELIZZOLA, M., DOWEN, R. H., HAWKINS, R. D., HON, G., TONTI-FILIPPINI, J., NERY, J. R., LEE, L., YE, Z., NGO, Q. M., EDSALL, L., ANTOSIEWICZ-BOURGET, J., STEWART, R., RUOTTI, V., MILLAR, A. H., THOMSON, J. A., REN, B. & ECKER, J. R. 2009a. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature,* 462**,** 315-22.
LISTER, R., PELIZZOLA, M., DOWEN, R. H., HAWKINS, R. D., HON, G., TONTI-FILIPPINI, J., NERY, J. R., LEE, L., YE, Z., NGO, Q. M., EDSALL, L., ANTOSIEWICZ-BOURGET, J., STEWART, R., RUOTTI, V., MILLAR, A. H., THOMSON, J. A., REN, B. & ECKER, J. R. 2009b. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature,* 462**,** 315-22.
LITWACK, G. 2008. Folic Acid and Folates. USA: Elseiver Inc.
LOPUHAÄ, C. E., ROSEBOOM, T. J., OSMOND, C., BARKER, D. J., RAVELLI, A. C., BLEKER, O. P., VAN DER ZEE, J. S. & VAN DER MEULEN, J. H. 2000. Atopy, lung function, and obstructive airways disease after prenatal exposure to famine. *Thorax,* 55**,** 555-61.
LORY, S. 1992. Determinants of extracellular protein secretion in gram-negative bacteria. *J Bacteriol,* 174**,** 3423-8.
LUMEY, L. H., RAVELLI, A. C., WIESSING, L. G., KOPPE, J. G., TREFFERS, P. E. & STEIN, Z. A. 1993. The Dutch famine birth cohort study: design, validation of exposure, and selected characteristics of subjects after 43 years follow-up. *Paediatr Perinat Epidemiol,* 7**,** 354-67.
LYON, E. & WITTWER, C. T. 2009. LightCycler technology in molecular diagnostics. *J Mol Diagn,* 11**,** 93-101.
MARCHENKO, N. D., ZAIKA, A. & MOLL, U. M. 2000. Death signal-induced localization of p53 protein to mitochondria. A potential role in apoptotic signaling. *J Biol Chem,* 275**,** 16202-12.
MARTIN, G. M. 2005. Epigenetic drift in aging identical twins. *Proc Natl Acad Sci U S A,* 102**,** 10413-4.
MARTINUSSEN, M. P., RISNES, K. R., JACOBSEN, G. W. & BRACKEN, M. B. 2012. Folic acid supplementation in early pregnancy and asthma in children aged 6 years. *Am J Obstet Gynecol,* 206**,** 72.e1-7.

MATTSON, M. P. & SHEA, T. B. 2003. Folate and homocysteine metabolism in neural plasticity and neurodegenerative disorders. *Trends Neurosci,* 26**,** 137-46.
MCENTEE, G., MINGUZZI, S., O'BRIEN, K., BEN LARBI, N., LOSCHER, C., O'FÁGÁIN, C. & PARLE-MCDERMOTT, A. 2011. The former annotated human pseudogene dihydrofolate reductase-like 1 (DHFRL1) is expressed and functional. *Proc Natl Acad Sci U S A,* 108**,** 15157-62.
MCKAY, J. A., XIE, L., HARRIS, S., WONG, Y. K., FORD, D. & MATHERS, J. C. 2011. Blood as a surrogate marker for tissue-specific DNA methylation and changes due to folate depletion in post-partum female mice. *Mol Nutr Food Res,* 55**,** 1026-35.
MCNULTY, B., MCNULTY, H., MARSHALL, B., WARD, M., MOLLOY, A. M., SCOTT, J. M., DORNAN, J. & PENTIEVA, K. 2013. Impact of continuing folic acid after the first trimester of pregnancy: findings of a randomized trial of Folic Acid Supplementation in the Second and Third Trimesters. *Am J Clin Nutr,* 98**,** 92-8.
MCNULTY, H. & SCOTT, J. M. 2008. Intake and status of folate and related B-vitamins: considerations and challenges in achieving optimal status. *Br J Nutr,* 99 Suppl 3**,** S48-54.
MCPARTLIN, J., HALLIGAN, A., SCOTT, J. M., DARLING, M. & WEIR, D. G. 1993. Accelerated folate breakdown in pregnancy. *Lancet,* 341**,** 148-9.
MENDES, C. C., RAIMUNDO, A. M., OLIVEIRA, L. D., ZAMPIERI, B. L., MARUCCI, G. H., BISELLI, J. M., GOLONI-BERTOLLO, E. M., EBERLIN, M. N., HADDAD, R., RICCIO, M. F., VANNUCCHI, H., CARVALHO, V. M. & PAVARINO, É. 2013. DHFR 19-bp deletion and SHMT C1420T polymorphisms and metabolite concentrations of the folate pathway in individuals with Down syndrome. *Genet Test Mol Biomarkers,* 17**,** 274-7.
MIGHELL, A. J., SMITH, N. R., ROBINSON, P. A. & MARKHAM, A. F. 2000. Vertebrate pseudogenes. *FEBS Lett,* 468**,** 109-14.
MILLS, J. L., MCPARTLIN, J. M., KIRKE, P. N., LEE, Y. J., CONLEY, M. R., WEIR, D. G. & SCOTT, J. M. 1995. Homocysteine metabolism in pregnancies complicated by neural-tube defects. *Lancet,* 345**,** 149-51.
MIRANDA, T. B. & JONES, P. A. 2007. DNA methylation: the nuts and bolts of repression. *J Cell Physiol,* 213**,** 384-90.
MISHRA, P. J., HUMENIUK, R., LONGO-SORBELLO, G. S., BANERJEE, D. & BERTINO, J. R. 2007. A miR-24 microRNA binding-site polymorphism in dihydrofolate reductase gene leads to methotrexate resistance. *Proc Natl Acad Sci U S A,* 104**,** 13513-8.
MISHRA, P. J., LONGO, G. S., MENON, L. G., ABALI, E. E., HUMENIUK, R., COLE, P. D., KAMEN, B. A., BANERJEE, D. & BERTINO, J. R. 2006. The 829CT single nucleotide polymorphism in the 3' UTR of the dihydrofolate reductase gene results in methotrexate resistance and is rare among non-Japanese American patients. AACR Meeting Abstacts.
MORANDI, C., MASTERS, J. N., MOTTES, M. & ATTARDI, G. 1982. Multiple forms of human dihydrofolate reductase messenger RNA. Cloning and expression in Escherichia coli of their DNA coding sequence. *J Mol Biol,* 156**,** 583-607.
MOYERS, S. & BAILEY, L. B. 2001. Fetal malformations and folate metabolism: review of recent evidence. *Nutr Rev,* 59**,** 215-24.
OHRVIK, V. E. & WITTHOFT, C. M. 2011. Human folate bioavailability. *Nutrients,* 3**,** 475-90.
PARLE-MCDERMOTT, A., KIRKE, P. N., MILLS, J. L., MOLLOY, A. M., COX, C., O'LEARY, V. B., PANGILINAN, F., CONLEY, M., CLEARY, L., BRODY, L. C. & SCOTT, J. M. 2006. Confirmation of the R653Q polymorphism of the trifunctional C1-synthase enzyme as a maternal risk for neural tube defects in the Irish population. *Eur J Hum Genet,* 14**,** 768-72.
PARLE-MCDERMOTT, A. & OZAKI, M. 2011. The impact of nutrition on differential methylated regions of the genome. *Adv Nutr,* 2**,** 463-71.
PARLE-MCDERMOTT, A., PANGILINAN, F., MILLS, J. L., KIRKE, P. N., GIBNEY, E. R., TROENDLE, J., O'LEARY, V. B., MOLLOY, A. M., CONLEY, M., SCOTT, J. M. & BRODY, L. C. 2007. The 19-

bp deletion polymorphism in intron-1 of dihydrofolate reductase (DHFR) may decrease rather than increase risk for spina bifida in the Irish population. *Am J Med Genet A,* 143A**,** 1174-80.

PARLE-MCDERMOTT, A., PANGILINAN, F., O'BRIEN, K. K., MILLS, J. L., MAGEE, A. M., TROENDLE, J., SUTTON, M., SCOTT, J. M., KIRKE, P. N., MOLLOY, A. M. & BRODY, L. C. 2009. A common variant in MTHFD1L is associated with neural tube defects and mRNA splicing efficiency. *Hum Mutat,* 30**,** 1650-6.

PAULSEN, M. & FERGUSON-SMITH, A. C. 2001. DNA methylation in genomic imprinting, development, and disease. *J Pathol,* 195**,** 97-110.

PITT, J. J. 2009. Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *Clin Biochem Rev,* 30**,** 19-34.

POGRIBNY, I. P., POIRIER, L. A. & JAMES, S. J. 1995. Differential sensitivity to loss of cytosine methyl groups within the hepatic p53 gene of folate/methyl deficient rats. *Carcinogenesis,* 16**,** 2863-7.

POMPEI, A., CORDISCO, L., AMARETTI, A., ZANONI, S., MATTEUZZI, D. & ROSSI, M. 2007. Folate production by bifidobacteria as a potential probiotic property. *Appl Environ Microbiol,* 73**,** 179-85.

PONT-KINGDON, G. & LYON, E. 2003. Rapid detection of aneuploidy (trisomy 21) by allele quantification combined with melting curves analysis of single-nucleotide polymorphism loci. *Clin Chem,* 49**,** 1087-94.

PORTELA, A. & ESTELLER, M. 2010. Epigenetic modifications and human disease. *Nat Biotechnol,* 28**,** 1057-68.

PTASHNE, M. 1992. *A Genetic Switch: Phage (Lambda) and Higher Organisms*, Cell Press: Blackwell Scientific Publications.

QIN, X., CUI, Y., SHEN, L., SUN, N., ZHANG, Y., LI, J., XU, X., WANG, B., HUO, Y. & WANG, X. 2013. Folic acid supplementation and cancer risk: a meta-analysis of randomized controlled trials. *Int J Cancer,* 133**,** 1033-41.

RAMPERSAUD, G. C., KAUWELL, G. P. & BAILEY, L. B. 2003. Folate: a key to optimizing health and reducing disease risk in the elderly. *J Am Coll Nutr,* 22**,** 1-8.

RAMPERSAUD, G. C., KAUWELL, G. P., HUTSON, A. D., CERDA, J. J. & BAILEY, L. B. 2000. Genomic DNA methylation decreases in response to moderate folate depletion in elderly women. *Am J Clin Nutr,* 72**,** 998-1003.

RAMSDEN, D. A. & GELLERT, M. 1998. Ku protein stimulates DNA end joining by mammalian DNA ligases: a direct role for Ku in repair of DNA double-strand breaks. *EMBO J,* 17**,** 609-14.

RAVELLI, G. P., STEIN, Z. A. & SUSSER, M. W. 1976. Obesity in young men after famine exposure in utero and early infancy. *N Engl J Med,* 295**,** 349-53.

RICHARDSON, B. 2003. DNA methylation and autoimmune disease. *Clin Immunol,* 109**,** 72-9.

ROPER, C., PEARCE, R., NAIR, S., SHARP, B., NOSTEN, F. & ANDERSON, T. 2004. Intercontinental spread of pyrimethamine-resistant malaria. *Science,* 305**,** 1124.

ROSEBOOM, T., DE ROOIJ, S. & PAINTER, R. 2006. The Dutch famine and its long-term consequences for adult health. *Early Hum Dev,* 82**,** 485-91.

ROUET, P., SMIH, F. & JASIN, M. 1994. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol,* 14**,** 8096-106.

SACHIDANANDAM, R., WEISSMAN, D., SCHMIDT, S. C., KAKOL, J. M., STEIN, L. D., MARTH, G., SHERRY, S., MULLIKIN, J. C., MORTIMORE, B. J., WILLEY, D. L., HUNT, S. E., COLE, C. G., COGGILL, P. C., RICE, C. M., NING, Z., ROGERS, J., BENTLEY, D. R., KWOK, P. Y., MARDIS, E. R., YEH, R. T., SCHULTZ, B., COOK, L., DAVENPORT, R., DANTE, M., FULTON, L., HILLIER, L., WATERSTON, R. H., MCPHERSON, J. D., GILMAN, B., SCHAFFNER, S., VAN ETTEN, W. J., REICH, D., HIGGINS, J., DALY, M. J., BLUMENSTIEL, B., BALDWIN, J.,

STANGE-THOMANN, N., ZODY, M. C., LINTON, L., LANDER, E. S., ALTSHULER, D. & GROUP, I. S. M. W. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature,* 409**,** 928-33.

SCHNELL, J. R., DYSON, H. J. & WRIGHT, P. E. 2004. Structure, dynamics, and catalytic function of dihydrofolate reductase. *Annu Rev Biophys Biomol Struct,* 33**,** 119-40.

SCHORK, N. J., FALLIN, D. & LANCHBURY, J. S. 2000. Single nucleotide polymorphisms and the future of genetic epidemiology. *Clin Genet,* 58**,** 250-64.

SCHULZ, L. C. 2010. The Dutch Hunger Winter and the developmental origins of health and disease. *Proc Natl Acad Sci U S A,* 107**,** 16757-8.

SCHWEITZER, B. I., DICKER, A. P. & BERTINO, J. R. 1990. Dihydrofolate reductase as a therapeutic target. *FASEB J,* 4**,** 2441-52.

SCOTT, J., KIRKE, P., MOLLOY, A., DALY, L. & WEIR, D. 1994. The role of folate in the prevention of neural-tube defects. *Proc Nutr Soc,* 53**,** 631-6.

SHANE, B. 1995. *Folate in Health and Disease,* New York, Bailey, L. B., ed.

SHARMA, S., DAS, M., KUMAR, A., MARWAHA, V., SHANKAR, S., SINGH, P., RAGHU, P., ANEJA, R., GROVER, R., ARYA, V., DHIR, V., GUPTA, R., KUMAR, U., JUYAL, R. C. & K, T. B. 2009. Purine biosynthetic pathway genes and methotrexate response in rheumatoid arthritis patients among north Indians. *Pharmacogenet Genomics,* 19**,** 823-8.

SHIELDS, D. C., KIRKE, P. N., MILLS, J. L., RAMSBOTTOM, D., MOLLOY, A. M., BURKE, H., WEIR, D. G., SCOTT, J. M. & WHITEHEAD, A. S. 1999. The "thermolabile" variant of methylenetetrahydrofolate reductase and neural tube defects: An evaluation of genetic risk and the relative importance of the genotypes of the embryo and the mother. *Am J Hum Genet,* 64**,** 1045-55.

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**,** 1-8.

SHIMADA, T. & NIENHUIS, A. W. 1985. Only the promoter region of the constitutively expressed normal and amplified human dihydrofolate reductase gene is DNase I hypersensitive and undermethylated. *J Biol Chem,* 260**,** 2468-74.

SINGH, S. M. & PANDA, A. K. 2005. Solubilization and refolding of bacterial inclusion body proteins. *J Biosci Bioeng,* 99**,** 303-10.

SIRAWARAPORN, W., PRAPUNWATTANA, P., SIRAWARAPORN, R., YUTHAVONG, Y. & SANTI, D. V. 1993. The dihydrofolate reductase domain of Plasmodium falciparum thymidylate synthase-dihydrofolate reductase. Gene synthesis, expression, and anti-folate-resistant mutants. *J Biol Chem,* 268**,** 21637-44.

SIROTNAK, F. M. & TOLNER, B. 1999. Carrier-mediated membrane transport of folates in mammalian cells. *Annu Rev Nutr,* 19**,** 91-122.

SOLAN, A., RATIA, K. & FAIRMAN, R. 2002. Exploring the role of alanine in the structure of the Lac repressor tetramerization domain, a ferritin-like Alacoil. *J Mol Biol,* 317**,** 601-12.

STANISŁAWSKA-SACHADYN, A., BROWN, K. S., MITCHELL, L. E., WOODSIDE, J. V., YOUNG, I. S., SCOTT, J. M., MURRAY, L., BOREHAM, C. A., MCNULTY, H., STRAIN, J. J. & WHITEHEAD, A. S. 2008. An insertion/deletion polymorphism of the dihydrofolate reductase (DHFR) gene is associated with serum and red blood cell folate concentrations in women. *Hum Genet,* 123**,** 289-95.

STEEGERS-THEUNISSEN, R. P., OBERMANN-BORST, S. A., KREMER, D., LINDEMANS, J., SIEBEL, C., STEEGERS, E. A., SLAGBOOM, P. E. & HEIJMANS, B. T. 2009. Periconceptional maternal folic acid use of 400 microg per day is related to increased methylation of the IGF2 gene in the very young child. *PLoS One,* 4**,** e7845.

STEIN, R., SCIAKY-GALLILI, N., RAZIN, A. & CEDAR, H. 1983. Pattern of methylation of two genes coding for housekeeping functions. *Proc Natl Acad Sci U S A,* 80**,** 2422-6.

STOLZENBERG-SOLOMON, R. Z., CHANG, S. C., LEITZMANN, M. F., JOHNSON, K. A., JOHNSON, C., BUYS, S. S., HOOVER, R. N. & ZIEGLER, R. G. 2006. Folate intake, alcohol use, and postmenopausal breast cancer risk in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. *Am J Clin Nutr,* 83**,** 895-904.

STOVER, P. J. 2004. Physiology of folate and vitamin B12 in health and disease. *Nutr Rev,* 62**,** S3-12; discussion S13.

STOVER, P. J. & FIELD, M. S. 2011. Trafficking of intracellular folates. *Adv Nutr,* 2**,** 325-31.

STOVER, P. J. & GARZA, C. 2002. Bringing individuality to public health recommendations. *J Nutr,* 132**,** 2476S-2480S.

STUDIER, F. W., ROSENBERG, A. H., DUNN, J. J. & DUBENDORFF, J. W. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol,* 185**,** 60-89.

SUH, J. R., HERBIG, A. K. & STOVER, P. J. 2001. New perspectives on folate catabolism. *Annu Rev Nutr,* 21**,** 255-82.

TAHILIANI, M., KOH, K. P., SHEN, Y., PASTOR, W. A., BANDUKWALA, H., BRUDNO, Y., AGARWAL, S., IYER, L. M., LIU, D. R., ARAVIND, L. & RAO, A. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science,* 324**,** 930-5.

TAI, N., SCHMITZ, J. C., LIU, J., LIN, X., BAILLY, M., CHEN, T. M. & CHU, E. 2004. Translational autoregulation of thymidylate synthase and dihydrofolate reductase. *Front Biosci,* 9**,** 2521-6.

TAKAYANAGI, A., KANEDA, S., AYUSAWA, D. & SENO, T. 1992. Intron 1 and the 5'-flanking region of the human thymidylate synthase gene as a regulatory determinant of growth-dependent expression. *Nucleic Acids Res,* 20**,** 4021-5.

TCHANTCHOU, F. & SHEA, T. B. 2008. Folate deprivation, the methionine cycle, and Alzheimer's disease. *Vitam Horm,* 79**,** 83-97.

TELLMANN, G. 2006. The E-Method: a highly accurate technique for gene-expression analysis. *Nature Methods| Application Notes,* i-iii.

TIBBETTS, A. S. & APPLING, D. R. 2010. Compartmentalization of Mammalian folate-mediated one-carbon metabolism. *Annu Rev Nutr,* 30**,** 57-81.

TOBI, E. W., LUMEY, L. H., TALENS, R. P., KREMER, D., PUTTER, H., STEIN, A. D., SLAGBOOM, P. E. & HEIJMANS, B. T. 2009. DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Hum Mol Genet,* 18**,** 4046-53.

ULRICH, C. M. & POTTER, J. D. 2007. Folate and cancer--timing is everything. *JAMA,* 297**,** 2408-9.

URDINGUIO, R. G., SANCHEZ-MUT, J. V. & ESTELLER, M. 2009. Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies. *Lancet Neurol,* 8**,** 1056-72.

VAN BEYNUM, I. M., KAPUSTA, L., DEN HEIJER, M., VERMEULEN, S. H., KOUWENBERG, M., DANIËLS, O. & BLOM, H. J. 2006. Maternal MTHFR 677C>T is a risk factor for congenital heart defects: effect modification by periconceptional folate supplementation. *Eur Heart J,* 27**,** 981-7.

VAN DER LINDEN, I. J., NGUYEN, U., HEIL, S. G., FRANKE, B., VLOET, S., GELLEKINK, H., DEN HEIJER, M. & BLOM, H. J. 2007. Variation and expression of dihydrofolate reductase (DHFR) in relation to spina bifida. *Mol Genet Metab,* 91**,** 98-103.

VAN DER PUT, N. M., STEEGERS-THEUNISSEN, R. P., FROSST, P., TRIJBELS, F. J., ESKES, T. K., VAN DEN HEUVEL, L. P., MARIMAN, E. C., DEN HEYER, M., ROZEN, R. & BLOM, H. J. 1995. Mutated methylenetetrahydrofolate reductase as a risk factor for spina bifida. *Lancet,* 346**,** 1070-1.

VAN DER PUT, N. M., VAN STRAATEN, H. W., TRIJBELS, F. J. & BLOM, H. J. 2001. Folate, homocysteine and neural tube defects: an overview. *Exp Biol Med (Maywood),* 226**,** 243-70.

VAN GUELPEN, B., HULTDIN, J., JOHANSSON, I., HALLMANS, G., STENLING, R., RIBOLI, E., WINKVIST, A. & PALMQVIST, R. 2006. Low folate levels may protect against colorectal cancer. *Gut,* 55**,** 1461-6.

VAUBEL, R. A., RUSTIN, P. & ISAYA, G. 2011. Mutations in the dimer interface of dihydrolipoamide dehydrogenase promote site-specific oxidative damages in yeast and human cells. *J Biol Chem,* 286**,** 40232-45.

VINEIS, P., CHUANG, S. C., VAISSIÈRE, T., CUENIN, C., RICCERI, F., JOHANSSON, M., UELAND, P., BRENNAN, P., HERCEG, Z. & INVESTIGATORS, T. G.-E. 2011. DNA methylation changes associated with cancer risk factors and blood levels of vitamin metabolites in a prospective study. *Epigenetics,* 6.

VOLLSET, S. E., CLARKE, R., LEWINGTON, S., EBBING, M., HALSEY, J., LONN, E., ARMITAGE, J., MANSON, J. E., HANKEY, G. J., SPENCE, J. D., GALAN, P., BØNAA, K. H., JAMISON, R., GAZIANO, J. M., GUARINO, P., BARON, J. A., LOGAN, R. F., GIOVANNUCCI, E. L., DEN HEIJER, M., UELAND, P. M., BENNETT, D., COLLINS, R., PETO, R. & COLLABORATION, F. T. B.-V. T. T. 2013. Effects of folic acid supplementation on overall and site-specific cancer incidence during the randomised trials: meta-analyses of data on 50 000 individuals. *Lancet*.

WALLING, J. 2006. From methotrexate to pemetrexed and beyond. A review of the pharmacodynamic and clinical properties of antifolates. *Invest New Drugs,* 24**,** 37-77.

WATERLAND, R. A., DOLINOY, D. C., LIN, J. R., SMITH, C. A., SHI, X. & TAHILIANI, K. G. 2006a. Maternal methyl supplements increase offspring DNA methylation at Axin Fused. *Genesis,* 44**,** 401-6.

WATERLAND, R. A. & JIRTLE, R. L. 2003. Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol Cell Biol,* 23**,** 5293-300.

WATERLAND, R. A., LIN, J. R., SMITH, C. A. & JIRTLE, R. L. 2006b. Post-weaning diet affects genomic imprinting at the insulin-like growth factor 2 (Igf2) locus. *Hum Mol Genet,* 15**,** 705-16.

WHITEHEAD, A. S., GALLAGHER, P., MILLS, J. L., KIRKE, P. N., BURKE, H., MOLLOY, A. M., WEIR, D. G., SHIELDS, D. C. & SCOTT, J. M. 1995. A genetic defect in 5,10 methylenetetrahydrofolate reductase in neural tube defects. *QJM,* 88**,** 763-6.

WHITROW, M. J., MOORE, V. M., RUMBOLD, A. R. & DAVIES, M. J. 2009. Effect of supplemental folic acid in pregnancy on childhood asthma: a prospective birth cohort study. *Am J Epidemiol,* 170**,** 1486-93.

WILLIAMS, K., CHRISTENSEN, J., PEDERSEN, M. T., JOHANSEN, J. V., CLOOS, P. A., RAPPSILBER, J. & HELIN, K. 2011. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature,* 473**,** 343-8.

WITTWER, C. T., REED, G. H., GUNDRY, C. N., VANDERSTEEN, J. G. & PRYOR, R. J. 2003. High-resolution genotyping by amplicon melting analysis using LCGreen. *Clin Chem,* 49**,** 853-60.

WOELLER, C. F., ANDERSON, D. D., SZEBENYI, D. M. & STOVER, P. J. 2007. Evidence for small ubiquitin-like modifier-dependent nuclear import of the thymidylate biosynthesis pathway. *J Biol Chem,* 282**,** 17623-31.

WOJDACZ, T. K., DOBROVIC, A. & HANSEN, L. L. 2008. Methylation-sensitive high-resolution melting. *Nat Protoc,* 3**,** 1903-8.

WURM, F. M. 2004. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotechnol,* 22**,** 1393-8.

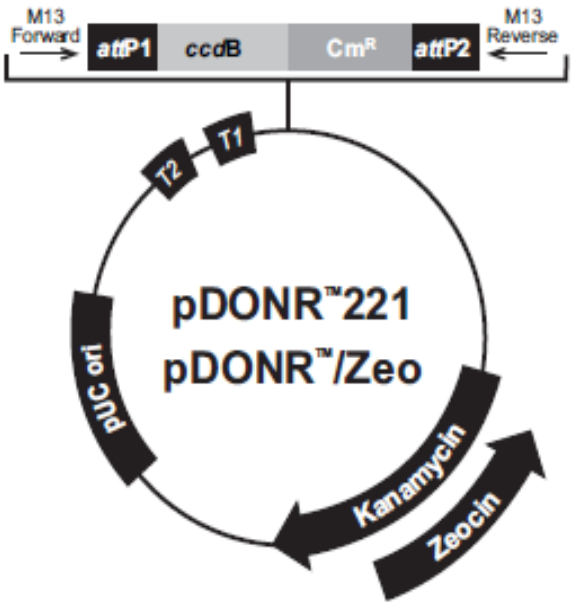
XU, X., GAMMON, M. D., WETMUR, J. G., RAO, M., GAUDET, M. M., TEITELBAUM, S. L., BRITTON, J. A., NEUGUT, A. I., SANTELLA, R. M. & CHEN, J. 2007. A functional 19-base pair deletion polymorphism of dihydrofolate reductase (DHFR) and risk of breast cancer in multivitamin users. *Am J Clin Nutr,* 85**,** 1098-102.

YAJNIK, C. S., DESHPANDE, S. S., JACKSON, A. A., REFSUM, H., RAO, S., FISHER, D. J., BHAT, D. S., NAIK, S. S., COYAJI, K. J., JOGLEKAR, C. V., JOSHI, N., LUBREE, H. G., DESHPANDE, V. U., REGE, S. S. & FALL, C. H. 2008. Vitamin B12 and folate concentrations during pregnancy and insulin resistance in the offspring: the Pune Maternal Nutrition Study. *Diabetologia,* 51**,** 29-38.

YAMADA, K., CHEN, Z., ROZEN, R. & MATTHEWS, R. G. 2001. Effects of common polymorphisms on the properties of recombinant human methylenetetrahydrofolate reductase. *Proc Natl Acad Sci U S A,* 98**,** 14853-8.

ZHAO, R., MATHERLY, L. H. & GOLDMAN, I. D. 2009. Membrane transporters and folate homeostasis: intestinal absorption and transport into systemic compartments and tissues. *Expert Rev Mol Med,* 11**,** e4.

Appendices

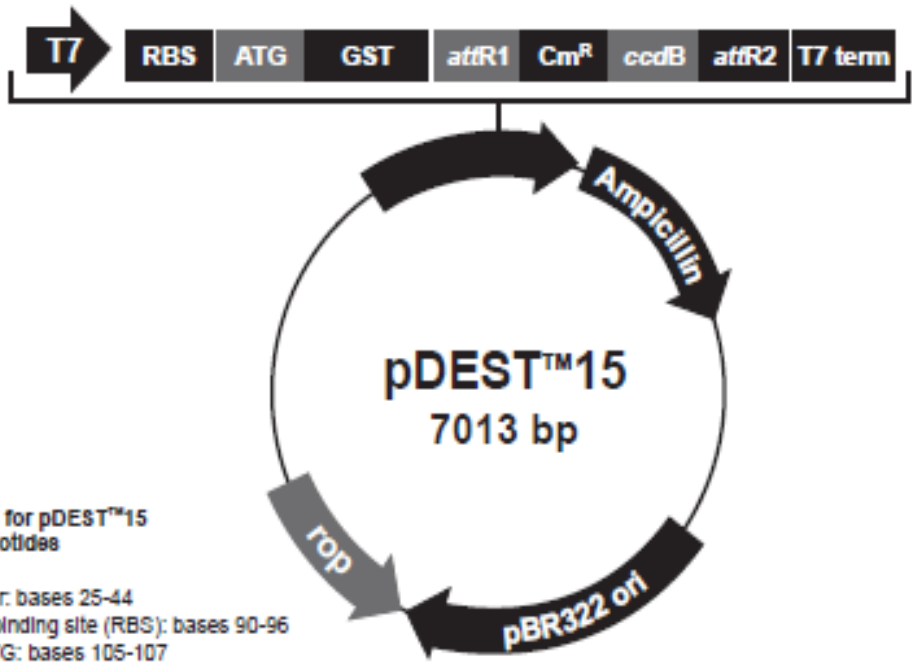


Comments for:	pDONR™221 4761 nucleotides	pDONR™/Zeo 4291 nucleotides
*rrn*B T2 transcription termination sequence (c):	268-295	268-295
rrnB T1 transcription termination sequence (c):	427-470	427-470
M13 Forward (-20) priming site:	537-552	537-552
*att*P1:	570-801	570-801
*ccd*B gene (c):	1197-1502	1197-1502
Chloramphenicol resistance gene (c):	1825-2505	1847-2506
*att*P2 (c):	2753-2984	2754-2985
M13 Reverse priming site:	3026-3042	3027-3043
Kanamycin resistance gene:	3155-3964	—
EM7 promoter (c):	—	3486-3552
Zeocin resistance gene (c):	—	3111-3485
pUC origin:	4085-4758	3615-4288

(c) = complementary strand

invitrogen™

Appendix A. Vector map for pDONR 221. Image taken from Invitrogen



Comments for pDEST™15
7013 nucleotides

T7 promoter: bases 25-44
Ribosome binding site (RBS): bases 90-96
Initiation ATG: bases 105-107
GST tag: bases 108-776
attR1: bases 792-916
Chloramphenicol resistance gene (Cm^R): bases 1025-1684
ccdB gene: bases 2026-2331
attR2: bases 2372-2496
T7 transcription termination region: bases 2518-2646
bla promoter: bases 3134-3232
Ampicillin (*bla*) resistance gene: bases 3233-4093
pBR322 origin: bases 4238-4911
ROP ORF: bases 5282-5473 (C)

C=complementary strand

Invitrogen™
life technologies

Appendix B. Vector map for pDEST15. Image taken from Invitrogen

ATM 5′ flanking sequence(1000 bp)

TTGGTCCCAACACTGAACCAAGGAAATTGCGCCCAAGGTTGAAGCACACGCAATCCGTCCGTCA
TGCAGACAAATGCTCTTTCAGGGGTCCTAATTAAGTGTGGAGGCCTGACTTTCCTTCCGAATCC
GCCCCAGGTGTTCCACCCCGAGCTTCCCCTCGGGCTTCCCCTCGCCACGCAGCCTCTGAAGAGA
GGAGCATCTACATACAAAGAGGCTTAAACTGCCCAGAACCTCCGAATGACGAAGAATCACCGCC
AGTCTCAACTCGTAAGCTGGGAGGCAAAACCCCAAAGCTTCCCTACCAAGGGAAAACCTTTGGC
CTCAAAGGTCCTTCTGTCCAGCATAGCCGGGTCCAATAACCCTCCATCCCGCGTCCGCGCTTAC
CCAATACAAGCCGGGCTACGTCCGAGGGTAACAACATGATCAAAACCACAGCAGGAACCACAAT
AAGGAACAAGACTCAGGTTAAAGCAAACACAGCGACAGCTCCTGCGCCGCATCTCCTGGTTCCA
GTGGCGGCACTGAACTCGCGGCAATTTGTCCCGCCTCTTTCGCTTCACGGCAGCCAATCGCTTC
CGCCAGAGAAAGAAAGGCGCCGAAATGAAACCCGCCTCCGTTCGCCTTCGGAACTGTCGTCACT
TCCGTCCTCAGACTTGGAGGGGCGGGGATGAGGAGGGCGGGGAGGACGACGAGGGCGAAGAGGG
TGGGTGAGAGCCCCGGAGCCCGAGCCGAAGGGCGAGCCGCAAACGCTAAGTCGCTGGCCATTGG
TGGACATGGCGCAGGCGCGTTTGCTCCGACGGGCCGAATGTTTTGGGGCAGTGTTTTGAGCGCG
GAGACCGCGTGATACTGGATGCGCATGGGCATACCGTGCTCTGCGGCTGCTTGGCGTTGCTTCT
TCCTCCAGAAGTGGGCGCTGGGCAGTCACGCAGGGTTTGAACCGGAAGCGGGAGTAGGTAGCTG
CGTGGCTAACGGAGAAAAGAAGCCGTGGCCGCGGGAGGAG

ATM primers used for MS-HRM in Wojdacz paper

Fwd primer: CGAAGAGGGTGGGTGAGAGTTT (22) 58.8°C
NON CpG Cs CONVERTED TO Ts
ATM sequence: CGAAGAGGGTGGGTGAGAGCCC

Rev primer: ACGCCATATCCACCAATAACCAAC (24) 57.5°C
Rev complement: GTTGGTTATTGGTGGATATGGCGT
NON CpG Cs CONVERTED TO Ts
ATM sequence: GCTGGCCATTGGTGGACATGGCGC

PCR PRODUCT = 85 bp

Appendix C. *ATM* MS-HRM primer sequence and location

RASSF1A 5′ Flanking Sequence (1000 bp)

TTTGTGATGTCCTCTTGTACCGGATGAGAGGGTCTCCATGCACACACAGACCTGGGACACTATC
CATCCACAAGTTCCTAAATAGGCCAGAGCAGTGATGCTCAACCCAGACTCCATGTTACAATAAT
TTGGGGAGTTTTTAAAATTTACTGATGCCTAGGGTCCACTCCCAGCAGTTGATTCAACAGGTCT
GCGGTGGGATCCAGGCTAGCGGGGAGGACTGTAAAAGCACCCCTGGTGATTCCAGCTGGTGTCT
ACCCAGGGGAGAGCAACCTTTGCTTGCTGGCGATTCCCAGGGGTGCAGAAGGACTGCTGGGTGT
GTGGCTGCGTGCATATTTTAGCATCTGATTCACTGGGTCAGAAAAGGGTGTTTGCTAAATAAAG
ACTCAACAAAACTCCTGCTTGCAGGGGGCCCACCAAAGGTTCTAAATTTTTCCAGGCTCCCTCC
CATAGGTGGTAATTTCCCTTCACCCTAAAGGTTCTGGAGGGGGTCATGAGTGTTTGAGAAGAGG
CAAGCCTGGGAAGATGGACTCCGAGGACAGTAGGCACAAACCCTTTCTCAAGAAGGGCCAAGGC
ATTTTAAAGATAAGAAACTTAAAATCAGCGTATTTTTACATATAAGCAGCCACCTCTGCTCATC
TGTGGCCCAGATACGAGTGGAGTGCGACAAGGGATAAACCATTTTCGCGCACTCTTCAGCGATG
GGGCGAAAGTAACGGACCTAGTCCTCGGGAGCTGTCCCCGCCGACCCCCTCTGCCGCGACTTGA
CCCGCGGCGACTGCGCTGCCCCTTGGCTGCCCCTTCCGCTCTCGTAGGCGCGCGGGGCCACTAC
TCACGCGCGCACTGCAGGCCTTTGCGCACGACGCCCCAGATGAAGTCGCCACAGAGGTCGCACC
ACGTGTGCGTGGCGGGCCCCGCGGGCTGGAAGCGGTGGCCACGGCCAGGGACCAGCTGCCGTGT
GGGGTTGCACGCGGTGCCCCGCGCGATGCGCAGCGCGTTG

Fwd primer: GTTTTAGATGAAGTCGTTATAGAGGT (26) 52.5°C
NON CpG Cs CONVERTED TO Ts
RASSF1A sequence: GCCCCAGATGAAGTCGCCACAGAGGT

Rev primer: CCCCACACGACAACTAATCCCTAA (24) 56.3°C
Rev complement: TTAGGGATTAGTTGTCGTGTGGGG
NON CpG Cs CONVERTED TO Ts
RASSF1A sequence: CCAGGGACCAGCTGCCGTGTGGGG

PCR PRODUCT = 100 bp

Appendix D. *RASSF1A* MS-HRM primer sequence and location

EIF2C3 5′ flanking sequence (1000bp)

TCTCAGATGTCTTTCGTTGAGACGCAAAAAAAAGCTGTGGGGTAGGGCGTAAGGGGAGGGTGGA
GAAGAGAGAGAGGGTTCCTGGCCAAGGGAACCAAAAGCAGAGAAGCTCCAAGTCAGAAAAAAGC
AGCGGTTGGCCCGTTGGGAGAATTGGAAGCCAACCAGTATAGCTGTAGCACAGGCGGGGCCCAT
TTACGCCGTGCGTTTTCACCCTGGAAAGGAGTTTGGGTTTTCAAGTGTGACGAGGCGCTATTAA
AGAGTTTTAAACATGGAAATTACTCGATCTGATGTTTGTGTAGTGTCCTGTTTGTTAGTCATCT
TTTCCAAACTATAATACATTTTTTTTTTCCTAAAAATGGTCGTTAAAGCAGCCTAGGACTTAGG
TCAATTGCGGTGGTCACTTGACCGCCTGTGAGGAGGCGGCGCGTGGGGTGGATCGGACTGGGCG
TGGCGGGGGCTCAGGAAGGAGGGTGGCCCTACCCCAGCGGGCTCGGCTCGGGGCCTCCGCGGCA
GTTCGGGGTCCTTCACCCGCCGGCTCCAGGTAGGCTACTCCTCAGGTAAGCCCCGCCGCCAGCC
GCGACGTCGTCGCAGACAGGCACCGCCCCCACTCGTGCGGCGCGAGTAGTCCTGCCCCTCCGCG
TTGTCTCCGGCCGGCACGGCCCGGCGGGGTACGGCCGAGCCCGCCGCATGTGGCCCGGCTCCCG
GACACCTCCCCGGCGTCCTCCGCGCCGGCCGCTCCTGCCCCGACGTCGCTCCGGCACGGCTCGG
GGCCCAGAGGCGAGGCGAGGACGCCGGGCAAGCCAGGCAGCGGAACTGACGCCGGCGAGCTTCC
GGGGCGGCCCCGGGCAGGTCGGCGGCGGCGGCCCGCAGTCGTGGAGGAGCGGTGGGAGCGTCGG
CGGCCGCGGGCGATGCAACTTCCGGACGGGACTCCCCTCTGTCCGCGCCTCACATCTCCCCTTC
CTCTCGCCTAGTCCTGTGCCGTTTTCCGTCCGCGACTCTT

Fwd primer: TTAGGAAGGAGGGTGGTTTTATT (23) 53.7°C
NON CpG Cs CONVERTED TO Ts
EIF2C3 sequence: TCAGGAAGGAGGGTGGCCCTACC

Rev primer: AAACTTACCTAAAAAATAACCTACCTAAA (29) 51.5°C
Rev complement: TTTAGGTAGGTTATTTTTTAGGTAAGTTT
NON CpG Cs CONVERTED TO Ts
EIF2C3 sequence: TCCAGGTAGGCTACTCCTCAGGTAAGCCC

PCR PRODUCT = 107 bp

Appendix E. *EIF2C3* MS-HRM primer sequence and location

ATP5F1 5′ flanking sequence (1000 bp)

TTAACACTTCCCCAATATTTGGACACTAGTAAGTATTTGCTGAACTAGTTGAATAAATGTCTTG
CAGAGTCTTTAAAAATATGTTGAATAATTGAACGTCTTGCACAGTGACTGGTTAAGTATAGAGA
ATCCGCCAAAAGTTGTCTTGATTATGTAGTGTTCAAGCTCAACTGTCAGAGAGGTTAAATCTAT
CAAAAACTGGGACTTTAAACTTCCATGCAATGCCGGCCAACTTCCAAATTCTGATTTTAAAAAA
AAAAAAAGCCAAAGCCTGCATCGCTAAGCCTAACAGTCAACAAAAAGAAAGCAGAATGAAGACA
AGAAAGCGGTAAGAAATGGGAAGAGACATGTAGCACAGATTGGCTTCAATGTCATCTTAAAATC
AAGATTATGCTTGGTCTGGGCTGTGGGGGTTAAGGGGCAGGTTACACAGAAGAAAAATGGTTCC
AATCCGTTGCTTAATTTTGTTACGGATTTTTCTAGGAAATCTCACCAGACAATTTAGAATTTTA
GAAAAGGTATTTCCCACCTACACAAGAGTTTGCATTCGGGGCCAACACTCTCAGGCTCCTCCAA
GCACTGAGTATAGGACGTTAGATATTGAGCTCAACCCTACGGTACCCCAAGTCTCCATTCTACC
GTTCCACCCGGGGGTGGGGGTGGATAATGGGATAGTGACTCACCTGAATCGGAGGCCACTAGAA
TACCTCTCTCCCCAACCCAAGTGAGGTCAGCCACTCCAGCCTCCGTTTGGACTCCGGCGGAGCA
GAAGCCTTCGTTGGGGGCGGCACAGGGGTCCTTAAAAAGCCAGAGGGAGCCGGCCCAGCAGCGC
CCACTCAGGCTGGAGGCCCCGAGGAGAAGCGCCCCATCTACGGGGGGTACAAGCTGTCAGCGCG
TGAAGGGTAGAAGGGTTCGCCTCCATGGAGACCGCACAGAACAGCGTTCCGGCCGGGTCTGGAT
CTGAGCTCAGGAACGCGGGGCAGGGCTGGGGACAGCTCGG

Fwd sequence: CGTTGGGGGCGGTATAGGGG 62.8°C
NON CpG Cs CONVERTED TO Ts
ATP5F1 sequence: CGTTGGGGGCGGCACAGGGG

Rev sequence: CCGAACTATCCCCAACCCTACCC 60.9°C
Rev complement: GGGTAGGGTTGGGGATAGTTCGG
NON CpG Cs CONVERTED TO Ts
ATP5F1 sequence: GGGCAGGGCTGGGGACAGCTCGG

PCR PRODUCT = 223 bp

Appendix F. *ATP5F1* MS-HRM primer sequence and location

DHFR 5′ Upstream Sequence ~ intron 1-2

ttataatctaaactacaaaaagagttcactggggaactgcacaatatgactgcttttaa
ccgtagtgatttcaaatattgagccatgctgttgcagtcttaaaaactggagacctaag
ggcagctttcttctagtcacccaatccagcacttttttaaaaaatcagtaaaactcttc
gaccaccaaggaaaaaaaaaaaggatggaggttaaaagacgcacccttgcccacaagc
cccctcatcagaatgggagtcaggagacctgagttcctgtctcaggcctgccattaaaa
acctgcataacctttgcctatctcctcaaacggaagtactaaaacctcagcgcttcacc
caatttgtagccccggctgggctcttcccaccttcccccttcttcagcccgccccttcct
cctccagccctatcatcgggcggagggtccccgcctccgcccgccttacccacaagccc
cgcccccccagccccgatggccctgcccAGTCCCAGACAGAACCTACTATGTGCGGCGG
CAGCTGGGGCGGGAAGGCGGGAGCTGGGGGCGCTGGGGGCGCTGCGGCCGCTGCGGCCG
CTGCAGCCGCTGCAGCGCCAGGGTCCACCTGGTCGGCTGCACCTGTGGAGGAGGAGGTG
GATTTCAGGCTTCCCGTAGACTGGAAGAATCGGCTCAAAACCGCTTGCCTCGCAGGGGC
TGAGCTGGAGGCAGCGAGGCCGCCCGACGCAGGCTTCCGGCGAGACATGGCAGGGCAAG
GATGGCAGCCCGGCGGCAGGGCCTGGCGAGGAGCGCGAGCCCGCGGCCGCAGTTCCCAG
GCGTCTGCGGGCGCGAGCACGCCGCGACCCTGCGTGCGCCGGGGCGGGGGGGCGGGGCC
TCGCCTGCACAAATGGGGACGAGGGGGGCGGGGCGGCCACAATTTCGCGCCAAACTTGA
CCGCGCGTTCTGCTGTAACGAGCGGGCTCGGAGGTCCTCCCGCTGCTGTCATGGTTGGT
TCGCTAAACTGCATCGTCGCTGTGTCCCAGAACATGGGCATCGGCAAGAACGGGGACCT
GCCCTGGCCACCGCTCAGgtatctgccgggccggggcgatgggacccaaacgggcgcag
gctgcccacggtcggggt**ACCTGGGCGGGACGCGCCA**ggccgactcccggcgagaggat
ggggccagacttgcggtctgcgctggcaggaagggtgggcccgactggattccccttttt
ctgctgcgcgggaggcccagttgctgatttctgcccggattctgctgcccggtgaggtc
tttgccctgcggcgccctcgcccagggcaaagtcccagccctggagaaaacacctcacc
cctacccacagcgctccgtttgtcaggtgccttagagctcgagcccaagggataatgtt
tcgagtaacgctgtttctctaacttgtag

5′ Upstream Sequence
Exon 1
Intron 1
19 bp deletion

```
Fwd sequence: GAGGGTCCCCGCCTCCGCCCG       54.6°C
     NON CpG Cs CONVERTED TO Ts
DHFR sequence: GAGGGTTTTCGTTTTCGTTCG

Rev sequence: CGCTGCAGCGCCAGGGTCCACCTG    54.4°C
Rev complement: CGTTGTAGCGTTAGGGTTTATTTG
     NON CpG Cs CONVERTED TO Ts
DHFR sequence: CAAATAAACCCTAACGCTACAACG

PCR PRODUCT = 186 bp
```

Appendix G. *DHFR* SMART-MSP primer sequence and location

DHFRL1 transcripts - multiple sequence alignment

```
DHFRL1_Transcript1
ATCCCAGCACCCTCCCTCGCCACCTCGGCCGCGTCCAGAAGCGTCTCATTCAGATAGAGGTACCCAGCCGGGCCCCAGGTCTCACCTGGGTCAGCATTGTCCCGC
TGAGAGCTTTCAAACAACAGGTGTACGCGAATCAGTATCAAAAAGCCTCCAGGGAACTGAAAAAAACAGTCTGACTTCCGGGATTCGCGGGCATCCACTTTCATT
CTCACGTTTCCTCTTCAATTTCTCGC

DHFRL1_Transcript1      GAGATACTGATTTGCTCGCGCTATAGAATTCTTGTCGAAGAGGCTTAGCTCCCGCGGGAG 60
DHFRL1_Transcript2      GAGATACTGATTTGCTCGCGCTATAGAATTCTTGTCGAAGAGGCTTAGCTCCCGCGGGAG 60
                        ************************************************************

DHFRL1_Transcript1      ATTTCGTGCGCAGTTACTTCCGGTAGCTGGTAAAGGCTGATACTTCCCAGGACGCGGAG- 119
DHFRL1_Transcript2      ATTTCGTGCGCAGTTACTTCCGGTAGCTGGTAAAGGCTGATACTTCCCAGGACGCGGAGG 120
                        ************************************************************

DHFRL1_Transcript1      ------------------------------------------------------------
DHFRL1_Transcript2      TAACGGGCCAGGGCCAAAGCGACTTTCGCTACTTGGATTGGTCGGCGTAGCTTTGGGCGG 180


DHFRL1_Transcript1      ------------------------------------------------------------
DHFRL1_Transcript2      CCGGACCTTAGAAAGTCACACATCTGCGCGCCTGTGCGGCCCCTGCTTCTGCGGATGCTG 240


DHFRL1_Transcript1      --GCACGTAAAAAAATTTGAAGAAGGGAATTTCGCGGCATTCTTGGCCTGGCTTCCTGGC 177
DHFRL1_Transcript2      AGGCACGTAAAAAAATTTGAAGAAGGGAATTTCGCGGCATTCTTGGCCTGGCTTCCTGGC 300
                          ************************************************************

DHFRL1_Transcript1      GTAGCCAGCAAGTTCGGAGGTGTTAACCGCTGCTGTCATGTTTCTTTTGCTAAACTGCAT 237
DHFRL1_Transcript2      GTAGCCAGCAAGTTCGGAGGTGTTAACCGCTGCTGTCATGTTTCTTTTGCTAAACTGCAT 360
                        ************************************************************

DHFRL1_Transcript1      CGTCGCTGTGTCCCAAAACATGGGCATCGGCAAGAACGGGGACCTGCCCAGGCCGCCGCT 297
DHFRL1_Transcript2      CGTCGCTGTGTCCCAAAACATGGGCATCGGCAAGAACGGGGACCTGCCCAGGCCGCCGCT 420
```

```
                         ************************************************************

DHFRL1_Transcript1       CAGGAATGAATTCAGGTATTTCCAGAGAATGACCACAACTTCTTCAGTAGAGGGTAAACA 357
DHFRL1_Transcript2       CAGGAATGAATTCAGGTATTTCCAGAGAATGACCACAACTTCTTCAGTAGAGGGTAAACA 480
                         ************************************************************


Fwd sequence: GGGCCAGGGCCAAAGCGACTTTCGC     57.5°C
      NON CpG Cs CONVERTED TO Ts
DHFR sequence: GGGTTAGGGTTAAAGCGATTTTCGT

Rev sequence: CGCCTGTGCGGCCCCTGCTT    58.0°C
Rev complement: CGCGTTTGTGCGGTTTTTGTTT
      NON CpG Cs CONVERTED TO Ts
DHFR sequence: AAACAAAAACCGCACAAACGCG

PCR PRODUCT = 81 bp
```

Appendix H. *DHFRL1* Transcript Alignment and SMART-MSP primer sequence and location

TUBE NUMBER	PCR	COL2A1 Mean Cp	DHFR Mean Cp	DHFR Cp [100%(target - control)]	DHFR ΔCp (target - control)	DHFR ΔΔCp	DHFR 2^-ΔΔCp	DHFR % methylation
1	**FIP-1074 A**	28.21	30.54	-0.56	2.33	2.89	0.135	13.49
2	**FIP-1074 B**	28.13	30.14	-1.54	2.01	3.55	0.085	8.54
3	**FIP-1085 A**	27.84	30.32	-0.56	2.48	3.04	0.122	12.16
4	**FIP-1085 B**	28.42	30.57	-0.56	2.15	2.71	0.153	15.28
5	**FIP-1010 A**	28.32	30.14	-0.56	1.82	2.38	0.192	19.21
6	**FIP-1010 B**	-	37.45					
7	**FIP-1156 A**	31.34	33.24	-0.56	1.9	2.46	0.182	18.17
8	**FIP-1156 B**	25.7	27.61	-0.98	1.91	2.89	0.135	13.49
9	**FIP-1099 A**	28.66	30.54	-1.54	1.88	3.42	0.093	9.34
10	**FIP-1099 B**	25.75	28.24	-0.56	2.49	3.05	0.121	12.07
11	**FIP-1061 A**	29.6	31.76	-0.56	2.16	2.72	0.152	15.18
12	**FIP-1061 B**	25.53	28.04	-0.56	2.51	3.07	0.119	11.91
13	**FIP-1069 A**	31.27	34.82	-0.4	3.55	3.95	0.065	6.47
14	**FIP-1069 B**	26.01	34.93	-0.4	8.92	9.32	0.002	0.16
15	**FIP-1160 A**	32.66	32.96	-0.4	0.3	0.7	0.616	61.56
16	**FIP-1160 B**	28.04	29.7	-0.4	1.66	2.06	0.240	23.98
17	**FIP-1092 A**	31.68	33.33	-0.4	1.65	2.05	0.241	24.15
18	**FIP-1092 B**	28.84	31.73	-0.4	2.89	3.29	0.102	10.22
19	**FIP-1123 A**	27.15	36.76	-0.4	9.61	10.01	0.001	0.10
20	**FIP-1123 B**	34.42	35.48	-0.4	1.06	1.46	0.363	36.35
21	**FIP-1132 A**	29.76	33.35	0.12	3.59	3.47	0.090	9.02
22	**FIP-1132 B**	25.76	29.66	-0.4	3.9	4.3	0.051	5.08
23	**FIP-1157 A**	29.96	31.48	-0.4	1.52	1.92	0.264	26.43
24	**FIP-1157 B**	27.47	31.08	-0.4	3.61	4.01	0.062	6.21
25	**FIP-1082 A**	30.85	34.45	-0.82	3.6	4.42	0.047	4.67
26	**FIP-1082 B**	27.86	31.46	-0.4	3.6	4	0.062	6.25

27	**FIP-1168 A**	30.24	35.21	-0.4	4.97	5.37	0.024	2.42
28	**FIP-1168 B**	25.09	28.9	-0.82	3.81	4.63	0.040	4.04
29	**FIP-1174 A**	34.88	40	-0.4	5.12	5.52	0.022	2.18
30	**FIP-1174 B**	25.8	28.8	-0.82	3	3.82	0.071	7.08
31	**FIP-1180 A**	25.07	37.26	-0.4	12.19	12.59	0.000	0.02
32	**FIP-1180 B**	27.76	29.38	-0.82	1.62	2.44	0.184	18.43
33	**FIP-1222 A**	32.65	34.24	-0.4	1.59	1.99	0.252	25.17
34	**FIP-1222 B**	25.91	29.12	-0.4	3.21	3.61	0.082	8.19
35	**FIP-1224 A**	31.24	33.68	-0.4	2.44	2.84	0.140	13.97
36	**FIP-1224 B**	26.41	30.64	-0.4	4.23	4.63	0.040	4.04
37	**FIP-1003 A**	30.96	32.18	-0.09	1.22	1.31	0.403	40.33
38	**FIP-1003 B**	27.56	29.34	-0.09	1.78	1.87	0.274	27.36
39	**FIP-1004 A**	34.04?	31.05	-1.62	#VALUE!	#VALUE!	#VALUE!	#VALUE!
40	**FIP-1004 B**	26.04	28.85	-0.09	2.81	2.9	0.134	13.40
41	**FIP-1005 A**	29.47	28.9	-3.3	-0.57	2.73	0.151	15.07
42	**FIP-1005 B**	31.28	29.81	-3.3	-1.47	1.83	0.281	28.13
43	**FIP-1007 A**	29.05	27.5	-3.3	-1.55	1.75	0.297	29.73
44	**FIP-1007 B**	28.93	29.74	-3.3	0.81	4.11	0.058	5.79
45	**FIP-1008 A**	29.78	29.23	-3.3	-0.55	2.75	0.149	14.87
46	**FIP-1008 B**	27.1	29.66	-0.09	2.56	2.65	0.159	15.93
47	**FIP-1009 A**	28.76	30.27	-3.3	1.51	4.81	0.036	3.56
48	**FIP-1009 B**	29.2	29.84	-3.3	0.64	3.94	0.065	6.52
49 b	**FIP-1012 A**	28.4	30.93	-0.34	2.53	2.87	0.137	13.68
50	**FIP-1012 B**	27.9	29.31	-3.3	1.41	4.71	0.038	3.82
51	**FIP-1018 A**	27.63	28.82	-3.3	1.19	4.49	0.045	4.45
52	**FIP-1018 B**	25.87	29.97	-0.09	4.1	4.19	0.055	5.48
53	**FIP-1019 A**							
54	**FIP-1019 B**	25.77	28.55	-0.09	2.78	2.87	0.137	13.68

55	**FIP-1020 A**	28.81	28.87	-3.3	0.06	3.36	0.097	9.74
56	**FIP-1020 B**	30.56	32.27	-3.3	1.71	5.01	0.031	3.10
57	**FIP-1022 A**	28.11	30.68	-3.3	2.57	5.87	0.017	1.71
58	**FIP-1022 B**	29.72	33.44	-3.55	3.72	7.27	0.006	0.65
59	**FIP-1023 A**	28.59	31.49	-3.3	2.9	6.2	0.014	1.36
60	**FIP-1023 B**	26.64	30.32	-0.09	3.68	3.77	0.073	7.33
61 b	**FIP-1024 A**	28.48	32.64	-0.34	4.16	4.5	0.044	4.42
62	**FIP-1024 B**	30.62	31.11	-3.3	0.49	3.79	0.072	7.23
63	**FIP-1028 A**	32.39	31.29	-3.3	-1.1	2.2	0.218	21.76
64	**FIP-1028 B**	31.18	30.19	-3.3	-0.99	2.31	0.202	20.17
65	**FIP-1030 A**	23.58	28.21	-0.09	4.63	4.72	0.038	3.79
66	**FIP-1030 B**	28.22	31.05	-0.09	2.83	2.92	0.132	13.21
67 b	**FIP-1031 A**	29.75	33.36	-0.09	3.61	3.7	0.077	7.69
68	**FIP-1031 B**							
69	**FIP-1032 A**							
70	**FIP-1032 B**							
71	**FIP-1033 A**	25.7	29.38	-0.73	3.68	4.41	0.047	4.70
72	**FIP-1033 B**	27.1	31.12	-0.73	4.02	4.75	0.037	3.72
73	**FIP-1035 A**	29.8	32.63	-0.34	2.83	3.17	0.111	11.11
74	**FIP-1035 B**	27.04	30.18	-0.34	3.14	3.48	0.090	8.96
75	**FIP-1037 A**	33.19	35.29	-1.54	2.1	3.64	0.080	8.02
76	**FIP-1037 B**	27.7	31.63	-0.34	3.93	4.27	0.052	5.18
77	**FIP-1041 A**	29.59	32.6	-0.34	3.01	3.35	0.098	9.81
78	**FIP-1041 B**	29.11	32.67	-0.34	3.56	3.9	0.067	6.70
79	**FIP-1042 A**	30.86	34.71	-0.34	3.85	4.19	0.055	5.48
80	**FIP-1042 B**	29.04	33.12	-0.34	4.08	4.42	0.047	4.67
81	**FIP-1043 A**	29.46	31.65	-0.34	2.19	2.53	0.173	17.31
82	**FIP-1043 B**	27.25	31.87	-0.34	4.62	4.96	0.032	3.21

83	**FIP-1050 A**	29.05	31.15	-0.34	2.1	2.44	0.184	18.43
84	**FIP-1050 B**	25.16	28.3	-0.34	3.14	3.48	0.090	8.96
85	**FIP-1052 A**	27.37	30.76	-0.34	3.39	3.73	0.075	7.54
86	**FIP-1052 B**	27.65	30.16	-0.34	2.51	2.85	0.139	13.87
87	**FIP-1053 A**	28.26	30.64	-1.54	2.38	3.92	0.066	6.61
88	**FIP-1053 B**	27.7	30.49	-0.34	2.79	3.13	0.114	11.42
89	**FIP-1062 A**	26.81	29.72	-0.34	2.91	3.25	0.105	10.51
90	**FIP-1062 B**	27.59	30.64	-0.34	3.05	3.39	0.095	9.54
91	**FIP-1063 A**	31.32	32.67	-0.34	1.35	1.69	0.310	30.99
92	**FIP-1063 B**	27.58	30.37	-0.34	2.79	3.13	0.114	11.42
93	**FIP-1064 A**	29.5	32.83	-0.34	3.33	3.67	0.079	7.86
94	**FIP-1064 B**	27.27	30.26	-0.34	2.99	3.33	0.099	9.94
95	**FIP-1066 A**	30.25	34	-0.34	3.75	4.09	0.059	5.87
96	**FIP-1066 B**	28.93	30.46	-1.91	1.53	3.44	0.092	9.21
97	**FIP-1067 A**	29.42	31.32	0.28	1.9	1.62	0.325	32.53
98	**FIP-1067 B**	30.17	33.7	0.28	3.53	3.25	0.105	10.51
99	**FIP-1070 A**	24.76	26.94	-1.42	2.18	3.6	0.082	8.25
100	**FIP-1070 B**	27.23	29.72	-1.42	2.49	3.91	0.067	6.65
101	**FIP-1076 A**	30.19	31.9	0.28	1.71	1.43	0.371	37.11
102	**FIP-1076 B**	28.51	30.26	0.28	1.75	1.47	0.361	36.10
103	**FIP-1077 A**	30.17	31.87	0.28	1.7	1.42	0.374	37.37
104	**FIP-1077 B**	27.8	30.32	0.16	2.52	2.36	0.195	19.48
105	**FIP1078 A**	28.29	30.19	0.28	1.9	1.62	0.325	32.53
106	**FIP-1078 B**	27.05	29.71	0.28	2.66	2.38	0.192	19.21
107	**FIP1087 A**	30.43	32.27	0.28	1.84	1.56	0.339	33.92
108	**FIP-1090 A**	29.74	32.11	0.28	2.37	2.09	0.235	23.49
109	**FIP-1090 B**	29.28	32.07	0.28	2.79	2.51	0.176	17.56
110	**FIP-1091 A**	27.73	31.08	0.28	3.35	3.07	0.119	11.91

111	**FIP-1091 B**	25.82	29.15	0.28	3.33	3.05	0.121	12.07
112	**FIP-1093 A**	29.57	31.07	0.28	1.5	1.22	0.429	42.93
113	**FIP-1093 B**	24.75	28.69	0.28	3.94	3.66	0.079	7.91
114	**FIP-1095 A**	30.52	33.49	0.28	2.97	2.69	0.155	15.50
115	**FIP-1095 B**	28.43	30.72	-1.42	2.29	3.71	0.076	7.64
116	**FIP-1100 A**	28.74	30.87	0.28	2.13	1.85	0.277	27.74
117	**FIP-1100 B**	30.37	32.57	0.28	2.2	1.92	0.264	26.43
118	**FIP-1102 A**	27.82	30.33	0.28	2.51	2.23	0.213	21.32
119	**FIP-1102 B**	27.98	30.37	0.28	2.39	2.11	0.232	23.16
120	**FIP-1107 A**	27.98	30.96	-1.42	2.98	4.4	0.047	4.74
121	**FIP-1107 B**	28.29	30.39	0.28	2.1	1.82	0.283	28.32
122	**FIP-1108 A**	28.49	32.2	0.28	3.71	3.43	0.093	9.28
123	**FIP-1108 B**	27.79	36.82	0.28	9.03	8.75	0.002	0.23
124	**FIP-1109 A**	29.96	31.67	-1.42	1.71	3.13	0.114	11.42
125	**FIP-1109 B**	26.9	29.17	-1.54	2.27	3.81	0.071	7.13
126	**FIP-1110 A**	29.51	31.98	0.28	2.47	2.19	0.219	21.92
127	**FIP-1110 B**	26.95	29.64	0.28	2.69	2.41	0.188	18.82
128	**FIP-1112 A**	30.26	32.34	0.28	2.08	1.8	0.287	28.72
129	**FIP-1112 B**	24.94	28.02	0.28	3.08	2.8	0.144	14.36
130	**FIP-1113 A**	27.15	30.46	0.28	3.31	3.03	0.122	12.24
131	**FIP-1113 B**	27.47	30.82	0.28	3.35	3.07	0.119	11.91
132	**FIP-1114 A**	29.74	32.72	0.28	2.98	2.7	0.154	15.39
133	**FIP-1114 B**	27.22	30.67	0.28	3.45	3.17	0.111	11.11
134	**FIP-1114 B**	27.28	27.95	-2.55	0.67	3.22	0.107	10.73
135	**FIP-1115 A**	31.04	31.54	-1.54	0.5	2.04	0.243	24.32
136	**FIP-1115 B**	25.88	28.09	0.28	2.21	1.93	0.262	26.24
137	**FIP-1116 A**	29.95	31.52	0.16	1.57	1.41	0.376	37.63
138	**FIP-1116 B**	28.07	29.39	0.16	1.32	1.16	0.448	44.75

139	**FIP-1118 A**	29.28	31.08	0.16	1.8	1.64	0.321	32.09
140	**FIP-1118 B**	22.64	26.52	0.16	3.88	3.72	0.076	7.59
141	**FIP-1119 A**	27.5	29.18	0.16	1.68	1.52	0.349	34.87
142	**FIP-1119 B**	26.3	28.3	0.16	2	1.84	0.279	27.93
143	**FIP-1120 A**	28.75	31.1	0.16	2.35	2.19	0.219	21.92
144	**FIP-1120 B**	27.4	28.87	0.16	1.47	1.31	0.403	40.33
145	**FIP-1122 A**	29.59	30.9	0.16	1.31	1.15	0.451	45.06
146	**FIP-1122 B**	26.91	28.27	0.16	1.36	1.2	0.435	43.53
147	**FIP-1126 A**	29.91	31.91	0.16	2	1.84	0.279	27.93
148	**FIP-1126 B**	28.04	29.66	0.16	1.62	1.46	0.363	36.35
149	**FIP-1127 A**	26.68	29.49	0.16	2.81	2.65	0.159	15.93
150	**FIP-1127 B**	26.76	31.68	0.16	4.92	4.76	0.037	3.69
151	**FIP-1131 A**	27.63	29.28	-1.86	1.65	3.51	0.088	8.78
152	**FIP-1131 B**	28.57	39.66	-3.44	11.09	14.53	0.000	0.00
153	**FIP-1134 B**	28.58	29.61	0.16	1.03	0.87	0.547	54.71
154	**FIP-1135 A**	28.32	30.17	0.16	1.85	1.69	0.310	30.99
155	**FIP-1136 A**	30.37	32.4	0.16	2.03	1.87	0.274	27.36
156	**FIP-1136 B**	27.69	29.37	0.16	1.68	1.52	0.349	34.87
157	**FIP-1138 A**	29.75	31.99	0.16	2.24	2.08	0.237	23.65
158	**FIP-1138 B**	25.21	25.57	-2.87	0.36	3.23	0.107	10.66
159	**FIP-1139 A**	30.51	32.23	0.16	1.72	1.56	0.339	33.92
160	**FIP-1139 B**	27.81	29.08	0.16	1.27	1.11	0.463	46.33
161	**FIP-1141 A**	30.91	32.21	0.16	1.3	1.14	0.454	45.38
162	**FIP-1141 B**	24.99	27.65	0.16	2.66	2.5	0.177	17.68
163	**FIP-1144 A**	24.1	26.99	0.16	2.89	2.73	0.151	15.07
164	**FIP-1144 B**	27.75	34.19	0.16	6.44	6.28	0.013	1.29
165	**FIP-1145 A**	29.27	32.49	0.16	3.22	3.06	0.120	11.99
166	**FIP-1145 B**	26.91	29.01	0.16	2.1	1.94	0.261	26.06

167	**FIP-1151 A**	28.81	31.38	-1.86	2.57	4.43	0.046	4.64
168	**FIP-1151 B**	29.88	31.23	0.16	1.35	1.19	0.438	43.83
169	**FIP-1152 A**	30.03	32.81	0.16	2.78	2.62	0.163	16.27
170	**FIP-1152 B**	27.65	31.28	0.16	3.63	3.47	0.090	9.02
171	**FIP-1153 A**	30.39	33.02	0.16	2.63	2.47	0.180	18.05
172	**FIP-1153 B**	26.84	28.45	0.16	1.61	1.45	0.366	36.60
173	**FIP-1154 A**	31.24	33.55	0.16	2.31	2.15	0.225	22.53
174	**FIP-1154 B**	25.44	28.52	0.16	3.08	2.92	0.132	13.21
175	**FIP-1155 A**	28.09	29.94	0.16	1.85	1.69	0.310	30.99
176	**FIP-1155 B**	27.58	30.16	0.16	2.58	2.42	0.187	18.69
177	**FIP-1161 A**	31.37	32.95	-0.02	1.58	1.6	0.330	32.99
178	**FIP-1161 B**	26.44	29.39	-0.02	2.95	2.97	0.128	12.76
179	**FIP-1163 A**	31.08	33.41	-0.02	2.33	2.35	0.196	19.61
180	**FIP-1163 B**	27.26	30.68	-0.02	3.42	3.44	0.092	9.21
181	**FIP-1166 A**	25.61	29.3	-0.02	3.69	3.71	0.076	7.64
182	**FIP-1166 B**	27.68	30.2	-0.02	2.52	2.54	0.172	17.19
183	**FIP-1171 A**	26.97	29.74	-0.02	2.77	2.79	0.145	14.46
184	**FIP-1171 B**	27.64	30.11	-0.02	2.47	2.49	0.178	17.80
185	**FIP-1173 A**	26.03	28.45	-0.02	2.42	2.44	0.184	18.43
186	**FIP-1173 B**	27.25	29.49	-0.02	2.24	2.26	0.209	20.88
187	**FIP-1177 A**	27.75	30.81	-0.02	3.06	3.08	0.118	11.83
188	**FIP-1177 B**	27.34	30.03	-0.02	2.69	2.71	0.153	15.28
189	**FIP-1183 A**	**29.86**	32.02	-0.11	2.16	2.27	0.207	20.73
190	**FIP-1183 B**	28.91	31.14	-0.02	2.23	2.25	0.210	21.02
191	**FIP-1185 A**	27.72	29.89	-0.02	2.17	2.19	0.219	21.92
192	**FIP-1185 B**	29.44	30.57	-1	1.13	2.13	0.228	22.85
193	**FIP-1185 B**	26.17	29.51	-0.02	3.34	3.36	0.097	9.74
194	**FIP-1187 A**	**29.77**	31.93	-0.11	2.16	2.27	0.207	20.73

195	**FIP-1187 B**	26.17	29.96	-0.02	3.79	3.81	0.071	7.13
196	**FIP-1188 A**	**31.14**	33.37	-0.11	2.23	2.34	0.198	19.75
197	**FIP-1188 B**	27.65	31	-0.02	3.35	3.37	0.097	9.67
198	**FIP-1189 A**	33.22	34.96	-0.02	1.74	1.76	0.295	29.52
199	**FIP-1189 B**	29.29	32.34	-0.02	3.05	3.07	0.119	11.91
200	**FIP-1190 A**	30.14	32.08	-0.02	1.94	1.96	0.257	25.70
201	**FIP-1190 B**	29.72	32.11	-0.02	2.39	2.41	0.188	18.82
202	**FIP-1192 A**	32.43	34.86	-0.02	2.43	2.45	0.183	18.30
203	**FIP-1192 B**	28.91	32.51	-0.02	3.6	3.62	0.081	8.13
204	**FIP-1193 A**	33	35.94	-0.02	2.94	2.96	0.129	12.85
205	**FIP-1193 B**	29.22	32.02	-0.02	2.8	2.82	0.142	14.16
206	**FIP-1194 A**	31.67	34.37	-0.02	2.7	2.72	0.152	15.18
207	**FIP-1194 B**	30.31	32.64	-0.02	2.33	2.35	0.196	19.61
208	**FIP-1195 A**	31.29	33.28	-0.02	1.99	2.01	0.248	24.83
209	**FIP-1195 B**	28.94	32.51	-0.02	3.57	3.59	0.083	8.30
210	**FIP-1197 A**	31.64	35.05	-0.02	3.41	3.43	0.093	9.28
211	**FIP-1197 B**	29.98	33.87	-0.02	3.89	3.91	0.067	6.65
212	**FIP-1198 A**	30.57	34.09	-0.02	3.52	3.54	0.086	8.60
213	**FIP-1199 A**	32.91	35.86	-0.02	2.95	2.97	0.128	12.76
214	**FIP-1199 B**	32.6	34.84	-0.02	2.24	2.26	0.209	20.88
215	**FIP-1200 A**					0		
216	**FIP-1200 B**	30.2	32.67	-0.02	2.47	2.49	0.178	17.80
217	**FIP-1201 A**	32.81	28.65	-6.19	-4.16	2.03	0.245	24.49
218	**FIP-1201 B**	31.42	27.93	-6.19	-3.49	2.7	0.154	15.39
219	**FIP-1203 A**					0		
220	**FIP-1203 B**	31.77	28.06	-6.19	-3.71	2.48	0.179	17.92
221	**FIP-1204 A**							
222	**FIP-1204 B**	31.55	29.01	-6.19	-2.54	3.65	0.080	7.97

223	**FIP-1205 A**							
224	**FIP-1205 B**	30.61	26.5	-6.19	-4.11	2.08	0.237	23.65
225	**FIP-1206 A**	29.48	26.56	-6.19	-2.92	3.27	0.104	10.37
226	**FIP-1206 B**	29.65	27.09	-6.19	-2.56	3.63	0.081	8.08
227	**FIP-1209 A**	32.68	29.43	-6.19	-3.25	2.94	0.130	13.03
228	**FIP-1209 B**	30.91	28.18	-6.19	-2.73	3.46	0.091	9.09
229	**FIP-1210 B**	30.66	28.5	-6.19	-2.16	4.03	0.061	6.12
230	**FIP-1211 A**							
231	**FIP-1211 B**	31.12	28.47	-6.19	-2.65	3.54	0.086	8.60
232	**FIP-1212 A**	32.77	28.93	-6.19	-3.84	2.35	0.196	19.61
233	**FIP-1212 B**	30.34	27.13	-6.19	-3.21	2.98	0.127	12.67
234	**FIP-1214 A**	31.18	28.25	-6.19	-2.93	3.26	0.104	10.44
235	**FIP-1214 B**	29.92	27.66	-6.19	-2.26	3.93	0.066	6.56
236	**FIP-1215 A**	32.36	29.45	-6.19	-2.91	3.28	0.103	10.29
237	**FIP-1215 B**	26.66	27.08	-6.19	0.42	6.61	0.010	1.02
238	**FIP-1216 A**	30.68	28.12	-6.19	-2.56	3.63	0.081	8.08
239	**FIP-1216 B**	28.37	27.61	-6.19	-0.76	5.43	0.023	2.32
240	**FIP-1217 A**	30.06	26.83	-6.19	-3.23	2.96	0.129	12.85
241	**FIP-1217 B**	31.42	27.96	-6.19	-3.46	2.73	0.151	15.07
242	**FIP-1218 A**	33.73	31.53	-6.19	-2.2	3.99	0.063	6.29
243	**FIP-1218 B**							
244	**FIP-1219 A**	33.87	31.21	-6.19	-2.66	3.53	0.087	8.66
245	**FIP-1219 B**	31.06	29.3	-6.19	-1.76	4.43	0.046	4.64
246	**FIP-1220 A**	32.79	30.89	-6.19	-1.9	4.29	0.051	5.11
247	**FIP-1220 B**	31.8	29.81	-6.19	-1.99	4.2	0.054	5.44
248	**FIP-1225 A**	33.73	30.47	-6.19	-3.26	2.93	0.131	13.12
249	**FIP-1225 B**	31.34	28.61	-6.17	-2.73	3.44	0.092	9.21
250	**FIP-1226 A**	30.55	28.95	-6.19	-1.6	4.59	0.042	4.15

251	**FIP-1226 B**	30.16	28.36	-6.17		6.17	0.014	1.39
252	**FIP-2001 A**	31.52	29.67	-6.19	-1.85	4.34	0.049	4.94
253	**FIP-2002 A**							
254	**FIP-2003 A**	30.88	29.82	-6.19	-1.06	5.13	0.029	2.86
255	**FIP-2003 A**	25.53	26.38	-3.44	0.85	4.29	0.051	5.11

Appendix I. Raw Cp values and analysis for the *COL2A1* and *DHFR* SMART-MSP Assays

TUBE NUMBER	PCR	COL2A1 Mean Cp	L1 Mean Cp	L1 Cp [100%(target - control)]	L1 ΔCp (target - control)	L1 ΔΔCp	L1 2^ΔΔCp	L1 % methylation
1	**FIP-1074 A**	28.21	35.88	1.17	7.67	6.5	0.011	1.10
2	**FIP-1074 B**	28.13	34.88	1.17	6.75	5.58	0.021	2.09
3	**FIP-1085 A**	27.84	35.01	1.17	7.17	6	0.016	1.56
4	**FIP-1085 B**	28.42	36.5	1.17	8.08	6.91	0.008	0.83
5	**FIP-1010 A**	28.32	34.14	1.17	5.82	4.65	0.040	3.98
6	**FIP-1010 B**		37.65					0.00
7	**FIP-1156 A**	31.34	37.54	1.17	6.2	5.03	0.031	3.06
8	**FIP-1156 B**	25.7	32.18	1.17	6.48	5.31	0.025	2.52
9	**FIP-1099 A**	28.66	35.15	1.17	6.49	5.32	0.025	2.50
10	**FIP-1099 B**	25.75	32.82	1.17	7.07	5.9	0.017	1.67
11	**FIP-1061 A**	29.6	36.61	1.17	7.01	5.84	0.017	1.75
12	**FIP-1061 B**	25.53	32.66	1.17	7.13	5.96	0.016	1.61
13	**FIP-1069 A**	31.27	40	1.33	8.73	7.4	0.006	0.59
14	**FIP-1069 B**	26.01	39.39	1.33	13.38	12.05	0.000	0.02
15	**FIP-1160 A**	32.66	37.45	1.33	4.79	3.46	0.091	9.09
16	**FIP-1160 B**	28.04	34.18	1.33	6.14	4.81	0.036	3.56
17	**FIP-1092 A**	31.68	38.32	1.33	6.64	5.31	0.025	2.52
18	**FIP-1092 B**	28.84	35.67	1.33	6.83	5.5	0.022	2.21
19	**FIP-1123 A**	27.15	38.6	1.33	11.45	10.12	0.001	0.09
20	**FIP-1123 B**	34.42	40	1.33	5.58	4.25	0.053	5.26
21	**FIP-1132 A**	29.76	38.85	-1.13	9.09	10.22	0.001	0.08
22	**FIP-1132 B**	25.76	37.38	1.33	11.62	10.29	0.001	0.08
23	**FIP-1157 A**	29.96	38.1	1.33	8.14	6.81	0.009	0.89
24	**FIP-1157 B**	27.47	36.02	1.33	8.55	7.22	0.007	0.67
25	**FIP-1082 A**	30.85	40	1.33	9.15	7.82	0.004	0.44
26	**FIP-1082 B**	27.86	37.03	1.33	9.17	7.84	0.004	0.44

27	**FIP-1168 A**	30.24	40	1.33	9.76	8.43	0.003	0.29
28	**FIP-1168 B**	25.09	34.13	1.33	9.04	7.71	0.005	0.48
29	**FIP-1174 A**	34.88	40	0.12	5.12	5	0.031	3.13
30	**FIP-1174 B**	25.8	34.03	1.33	8.23	6.9	0.008	0.84
31	**FIP-1180 A**	25.07	40	1.33	14.93	13.6	0.000	0.01
32	**FIP-1180 B**	27.76	34.22	1.33	6.46	5.13	0.029	2.86
33	**FIP-1222 A**	32.65	40	1.33	7.35	6.02	0.015	1.54
34	**FIP-1222 B**	25.91	33.76	1.33	7.85	6.52	0.011	1.09
35	**FIP-1224 A**	31.24	40	1.33	8.76	7.43	0.006	0.58
36	**FIP-1224 B**	26.41	36.65	1.33	10.24	8.91	0.002	0.21
37	**FIP-1003 A**	30.96	39.39	-4.82	8.43	13.25	0.000	0.01
38	**FIP-1003 B**	27.56	34.27	-4.82	6.71	11.53	0.000	0.03
39	**FIP-1004 A**	34.04?	33.81	-3.14	#VALUE!	#VALUE!	#VALUE!	#VALUE!
40	**FIP-1004 B**	26.04	34.69	-4.82	8.65	13.47	0.000	0.01
41	**FIP-1005 A**	29.47	33.96	-4.82	4.49	9.31	0.002	0.16
42	**FIP-1005 B**	31.28	37.01	-4.82	5.73	10.55	0.001	0.07
43	**FIP-1007 A**	29.05	33.19	-4.82	4.14	8.96	0.002	0.20
44	**FIP-1007 B**	28.93	34.71	-4.82	5.78	10.6	0.001	0.06
45	**FIP-1008 A**	29.78	36.04	-4.82	6.26	11.08	0.000	0.05
46	**FIP-1008 B**	27.1	35.99	-4.82	8.89	13.71	0.000	0.01
47	**FIP-1009 A**	28.76	37.84	-4.82	9.08	13.9	0.000	0.01
48	**FIP-1009 B**	29.2	34.8	-4.82	5.6	10.42	0.001	0.07
49 b	**FIP-1012 A**	28.4	38.5	-4.82	10.1	14.92	0.000	0.00
50	**FIP-1012 B**	27.9	34.18	-4.82	6.28	11.1	0.000	0.05
51	**FIP-1018 A**	27.63	34.5	-4.82	6.87	11.69	0.000	0.03
52	**FIP-1018 B**	25.87	37.8	-4.82	11.93	16.75	0.000	0.00
53	**FIP-1019 A**							
54	**FIP-1019 B**	25.77	35.2	-4.82	9.43	14.25	0.000	0.01

55	**FIP-1020 A**	28.81	36.09	-4.82	7.28	12.1	0.000	0.02
56	**FIP-1020 B**	30.56	35.82	-4.82	5.26	10.08	0.001	0.09
57	**FIP-1022 A**	28.11	37.48	-4.82	9.37	14.19	0.000	0.01
58	**FIP-1022 B**	29.72	40	-4.82	10.28	15.1	0.000	0.00
59	**FIP-1023 A**	28.59	39.16	-4.82	10.57	15.39	0.000	0.00
60	**FIP-1023 B**	26.64	36.97	-4.82	10.33	15.15	0.000	0.00
61 b	**FIP-1024 A**	28.48	36.91	-4.82	8.43	13.25	0.000	0.01
62	**FIP-1024 B**	30.62	39.38	-4.82	8.76	13.58	0.000	0.01
63	**FIP-1028 A**	32.39	38.55	-4.82	6.16	10.98	0.000	0.05
64	**FIP-1028 B**	31.18	37.35	-4.82	6.17	10.99	0.000	0.05
65	**FIP-1030 A**	23.58	36.44	-4.82	12.86	17.68	0.000	0.00
66	**FIP-1030 B**	28.22	40	-4.82	11.78	16.6	0.000	0.00
67 b	**FIP-1031 A**	29.75	40	-4.82	10.25	15.07	0.000	0.00
68	**FIP-1031 B**							
69	**FIP-1032 A**							
70	**FIP-1032 B**							
71	**FIP-1033 A**	25.7	36.39	-2.25	10.69	12.94	0.000	0.01
72	**FIP-1033 B**	27.1	38.87	-2.25	11.77	14.02	0.000	0.01
73	**FIP-1035 A**	29.8	37.88	0.6	8.08	7.48	0.006	0.56
74	**FIP-1035 B**	27.04	35.64	0.6	8.6	8	0.004	0.39
75	**FIP-1037 A**	33.19	40	0.6	6.81	6.21	0.014	1.35
76	**FIP-1037 B**	27.7	35.57	0.6	7.87	7.27	0.006	0.65
77	**FIP-1041 A**	29.59	37.19	0.6	7.6	7	0.008	0.78
78	**FIP-1041 B**	29.11	37.86	0.6	8.75	8.15	0.004	0.35
79	**FIP-1042 A**	30.86	40	0.6	9.14	8.54	0.003	0.27
80	**FIP-1042 B**	29.04	36.52	0.6	7.48	6.88	0.008	0.85
81	**FIP-1043 A**	29.46	36.49	0.6	7.03	6.43	0.012	1.16
82	**FIP-1043 B**	27.25	37.25	0.6	10	9.4	0.001	0.15

83	**FIP-1050 A**	29.05	36.54	0.6	7.49	6.89	0.008	0.84
84	**FIP-1050 B**	25.16	32.76	0.6	7.6	7	0.008	0.78
85	**FIP-1052 A**	27.37	35.58	0.6	8.21	7.61	0.005	0.51
86	**FIP-1052 B**	27.65	36.03	0.6	8.38	7.78	0.005	0.45
87	**FIP-1053 A**	28.26	35.91	0.6	7.65	7.05	0.008	0.75
88	**FIP-1053 B**	27.7	33.64	0.6	5.94	5.34	0.025	2.47
89	**FIP-1062 A**	26.81	34.2	0.6	7.39	6.79	0.009	0.90
90	**FIP-1062 B**	27.59	35.97	0.6	8.38	7.78	0.005	0.45
91	**FIP-1063 A**	31.32	40	0.6	8.68	8.08	0.004	0.37
92	**FIP-1063 B**	27.58	36.39	0.6	8.81	8.21	0.003	0.34
93	**FIP-1064 A**	29.5	37.65	0.6	8.15	7.55	0.005	0.53
94	**FIP-1064 B**	27.27	35.35	0.6	8.08	7.48	0.006	0.56
95	**FIP-1066 A**	30.25	39.77	0.6	9.52	8.92	0.002	0.21
96	**FIP-1066 B**	28.93	37.74	0.6	8.81	8.21	0.003	0.34
97	**FIP-1067 A**	29.42	33.49	-1.49	4.07	5.56	0.021	2.12
98	**FIP-1067 B**	30.17	36.67	-1.49	6.5	7.99	0.004	0.39
99	**FIP-1070 A**	24.76	28.91	-1.49	4.15	5.64	0.020	2.01
100	**FIP-1070 B**	27.23	32.63	-1.49	5.4	6.89	0.008	0.84
101	**FIP-1076 A**	30.19	33.46	-1.49	3.27	4.76	0.037	3.69
102	**FIP-1076 B**	28.51	31.86	-1.49	3.35	4.84	0.035	3.49
103	**FIP-1077 A**	30.17	33.5	-1.49	3.33	4.82	0.035	3.54
104	**FIP-1077 B**	27.8	32.78	-1.61	4.98	6.59	0.010	1.04
105	**FIP1078 A**	28.29	31.88	-1.49	3.59	5.08	0.030	2.96
106	**FIP-1078 B**	27.05	31.08	-1.49	4.03	5.52	0.022	2.18
107	**FIP1087 A**	30.43	34.38	-1.49	3.95	5.44	0.023	2.30
108	**FIP-1090 A**	29.74	33.89	-1.49	4.15	5.64	0.020	2.01
109	**FIP-1090 B**	29.28	35.03	-1.49	5.75	7.24	0.007	0.66
110	**FIP-1091 A**	27.73	33.04	-1.49	5.31	6.8	0.009	0.90

111	**FIP-1091 B**	25.82	31.29	-1.49	5.47	6.96	0.008	0.80
112	**FIP-1093 A**	29.57	33.34	-1.49	3.77	5.26	0.026	2.61
113	**FIP-1093 B**	24.75	30.06	-1.49	5.31	6.8	0.009	0.90
114	**FIP-1095 A**	30.52	35.14	-1.49	4.62	6.11	0.014	1.45
115	**FIP-1095 B**	28.43	33.65	-1.49	5.22	6.71	0.010	0.96
116	**FIP-1100 A**	28.74	33.29	-1.49	4.55	6.04	0.015	1.52
117	**FIP-1100 B**	30.37	37.39	0.08	7.02	6.94	0.008	0.81
118	**FIP-1102 A**	27.82	32.55	-1.49	4.73	6.22	0.013	1.34
119	**FIP-1102 B**	27.98	32.61	-1.49	4.63	6.12	0.014	1.44
120	**FIP-1107 A**	27.98	32.99	-1.49	5.01	6.5	0.011	1.10
121	**FIP-1107 B**	28.29	32.15	-1.49	3.86	5.35	0.025	2.45
122	**FIP-1108 A**	28.49	35.34	-1.49	6.85	8.34	0.003	0.31
123	**FIP-1108 B**	27.79	37.55	0.6	9.76	9.16	0.002	0.17
124	**FIP-1109 A**	29.96	34.64	-1.49	4.68	6.17	0.014	1.39
125	**FIP-1109 B**	26.9	31.88	0.81	4.98	4.17	0.056	5.56
126	**FIP-1110 A**	29.51	33.98	-1.49	4.47	5.96	0.016	1.61
127	**FIP-1110 B**	26.95	31.19	-1.49	4.24	5.73	0.019	1.88
128	**FIP-1112 A**	30.26	34.87	-1.49	4.61	6.1	0.015	1.46
129	**FIP-1112 B**	24.94	29.87	-0.51	4.93	5.44	0.023	2.30
130	**FIP-1113 A**	27.15	31.13	-0.51	3.98	4.49	0.045	4.45
131	**FIP-1113 B**	27.47	32.47	-0.51	5	5.51	0.022	2.19
132	**FIP-1114 A**	29.74	34.55	-0.51	4.81	5.32	0.025	2.50
133	**FIP-1114 B**	27.22	33.66	-0.51	6.44	6.95	0.008	0.81
134	**FIP-1114 B**	27.28	33.27	-0.63	5.99	6.62	0.010	1.02
135	**FIP-1115 A**	31.04	35.05	-0.51	4.01	4.52	0.044	4.36
136	**FIP-1115 B**	25.88	30.27	-0.51	4.39	4.9	0.033	3.35
137	**FIP-1116 A**	29.95	34.92	-0.95	4.97	5.92	0.017	1.65
138	**FIP-1116 B**	28.07	32.23	-0.95	4.16	5.11	0.029	2.90

139	**FIP-1118 A**	29.28	34.74	-0.95	5.46	6.41	0.012	1.18
140	**FIP-1118 B**	22.64	30.8	-0.95	8.16	9.11	0.002	0.18
141	**FIP-1119 A**	27.5	32.18	-0.95	4.68	5.63	0.020	2.02
142	**FIP-1119 B**	26.3	31.17	-0.95	4.87	5.82	0.018	1.77
143	**FIP-1120 A**	28.75	34.87	-0.95	6.12	7.07	0.007	0.74
144	**FIP-1120 B**	27.4	32.21	-0.95	4.81	5.76	0.018	1.85
145	**FIP-1122 A**	29.59	34.19	-0.95	4.6	5.55	0.021	2.13
146	**FIP-1122 B**	26.91	31.92	-0.95	5.01	5.96	0.016	1.61
147	**FIP-1126 A**	29.91	36.78	-0.95	6.87	7.82	0.004	0.44
148	**FIP-1126 B**	28.04	33.59	-0.95	5.55	6.5	0.011	1.10
149	**FIP-1127 A**	26.68	33.01	-0.95	6.33	7.28	0.006	0.64
150	**FIP-1127 B**	26.76	32.81	-0.95	6.05	7	0.008	0.78
151	**FIP-1131 A**	27.63	32.92	-0.95	5.29	6.24	0.013	1.32
152	**FIP-1131 B**	28.57	31.64	-3.14	3.07	6.21	0.014	1.35
153	**FIP-1134 B**	28.58	32.65	-0.95	4.07	5.02	0.031	3.08
154	**FIP-1135 A**	28.32	33.3	-0.95	4.98	5.93	0.016	1.64
155	**FIP-1136 A**	30.37	36.23	-0.95	5.86	6.81	0.009	0.89
156	**FIP-1136 B**	27.69	32.71	-0.95	5.02	5.97	0.016	1.60
157	**FIP-1138 A**	29.75	38.23	-0.95	8.48	9.43	0.001	0.14
158	**FIP-1138 B**	25.21	32.3	-0.95	7.09	8.04	0.004	0.38
159	**FIP-1139 A**	30.51	38	-0.95	7.49	8.44	0.003	0.29
160	**FIP-1139 B**	27.81	33.08	-0.95	5.27	6.22	0.013	1.34
161	**FIP-1141 A**	30.91	36.35	-0.95	5.44	6.39	0.012	1.19
162	**FIP-1141 B**	24.99	31.76	-0.95	6.77	7.72	0.005	0.47
163	**FIP-1144 A**	24.1	30.99	-0.95	6.89	7.84	0.004	0.44
164	**FIP-1144 B**	27.75	36.1	-0.95	8.35	9.3	0.002	0.16
165	**FIP-1145 A**	29.27	35.92	-0.95	6.65	7.6	0.005	0.52
166	**FIP-1145 B**	26.91	33.11	-0.95	6.2	7.15	0.007	0.70

167	**FIP-1151 A**	28.81	35.99	-0.95	7.18	8.13	0.004	0.36
168	**FIP-1151 B**	29.88	34.95	-0.95	5.07	6.02	0.015	1.54
169	**FIP-1152 A**	30.03	37.4	-1.93	7.37	9.3	0.002	0.16
170	**FIP-1152 B**	27.65	34.28	-0.36	6.63	6.99	0.008	0.79
171	**FIP-1153 A**	30.39	35.46	-1.93	5.07	7	0.008	0.78
172	**FIP-1153 B**	26.84	30.65	-1.93	3.81	5.74	0.019	1.87
173	**FIP-1154 A**	31.24	36.26	-1.93	5.02	6.95	0.008	0.81
174	**FIP-1154 B**	25.44	30.38	-1.93	4.94	6.87	0.009	0.85
175	**FIP-1155 A**	28.09	32.5	-1.93	4.41	6.34	0.012	1.23
176	**FIP-1155 B**	27.58	32.18	-1.93	4.6	6.53	0.011	1.08
177	**FIP-1161 A**	31.37	34.31	-2.54	2.94	5.48	0.022	2.24
178	**FIP-1161 B**	26.44	31.14	-2.54	4.7	7.24	0.007	0.66
179	**FIP-1163 A**	31.08	40	3.29	8.92	5.63	0.020	2.02
180	**FIP-1163 B**	27.26	32.22	-2.54	4.96	7.5	0.006	0.55
181	**FIP-1166 A**	25.61	35.73	3.29	10.12	6.83	0.009	0.88
182	**FIP-1166 B**	27.68	35.33	3.29	7.65	4.36	0.049	4.87
183	**FIP-1171 A**	26.97	35.84	3.29	8.87	5.58	0.021	2.09
184	**FIP-1171 B**	27.64	36.04	3.29	8.4	5.11	0.029	2.90
185	**FIP-1173 A**	26.03	29.69	-2.54	3.66	6.2	0.014	1.36
186	**FIP-1173 B**	27.25	36.45	3.29	9.2	5.91	0.017	1.66
187	**FIP-1177 A**	27.75	36.77	3.29	9.02	5.73	0.019	1.88
188	**FIP-1177 B**	27.34	37.09	3.29	9.75	6.46	0.011	1.14
189	**FIP-1183 A**	**29.86**	39.25	3.2	9.39	6.19	0.014	1.37
190	**FIP-1183 B**	28.91	36.84	3.29	7.93	4.64	0.040	4.01
191	**FIP-1185 A**	27.72	36.11	3.29	8.39	5.1	0.029	2.92
192	**FIP-1185 B**	29.44	35.07	2.31	5.63	3.32	0.100	10.01
193	**FIP-1185 B**	26.17	33.43	3.29	7.26	3.97	0.064	6.38
194	**FIP-1187 A**	**29.77**	38.17	3.2	8.4	5.2	0.027	2.72

195	**FIP-1187 B**	26.17	37.01	3.29	10.84	7.55	0.005	0.53
196	**FIP-1188 A**	**31.14**	40	3.2	8.86	5.66	0.020	1.98
197	**FIP-1188 B**	27.65	36.72	3.29	9.07	5.78	0.018	1.82
198	**FIP-1189 A**	33.22	40	3.29	6.78	3.49	0.089	8.90
199	**FIP-1189 B**	29.29	38.71	3.29	9.42	6.13	0.014	1.43
200	**FIP-1190 A**	30.14	34.06	-2.54	3.92	6.46	0.011	1.14
201	**FIP-1190 B**	29.72	36.41	3.29	6.69	3.4	0.095	9.47
202	**FIP-1192 A**	32.43	40	3.29	7.57	4.28	0.051	5.15
203	**FIP-1192 B**	28.91	40	3.29	11.09	7.8	0.004	0.45
204	**FIP-1193 A**	33	40	3.29	7	3.71	0.076	7.64
205	**FIP-1193 B**	29.22	32.42	-2.54	3.2	5.74	0.019	1.87
206	**FIP-1194 A**	31.67	40	3.29	8.33	5.04	0.030	3.04
207	**FIP-1194 B**	30.31	37.95	3.29	7.64	4.35	0.049	4.90
208	**FIP-1195 A**	31.29	33.98	-2.54	2.69	5.23	0.027	2.66
209	**FIP-1195 B**	28.94	37.04	3.29	8.1	4.81	0.036	3.56
210	**FIP-1197 A**	31.64	40	3.29	8.36	5.07	0.030	2.98
211	**FIP-1197 B**	29.98	38.6	3.29	8.62	5.33	0.025	2.49
212	**FIP-1198 A**	30.57	40	3.29	9.43	6.14	0.014	1.42
213	**FIP-1199 A**	32.91	40	3.29	7.09	3.8	0.072	7.18
214	**FIP-1199 B**	32.6	40	3.29	7.4	4.11	0.058	5.79
215	**FIP-1200 A**							
216	**FIP-1200 B**	30.2	36.12	3.29	5.92	2.63	0.162	16.15
217	**FIP-1201 A**	32.81	32.68	-5.13	-0.13	5	0.031	3.13
218	**FIP-1201 B**	31.42	32.3	-5.13	0.88	6.01	0.016	1.55
219	**FIP-1203 A**							
220	**FIP-1203 B**	31.77	32.03	-5.13	0.26	5.39	0.024	2.38
221	**FIP-1204 A**				0			
222	**FIP-1204 B**	31.55	32.65	-5.13	1.1	6.23	0.013	1.33

223	**FIP-1205 A**							
224	**FIP-1205 B**	30.61	31.04	-5.13	0.43	5.56	0.021	2.12
225	**FIP-1206 A**	29.48	29.79	-5.13	0.31	5.44	0.023	2.30
226	**FIP-1206 B**	29.65	30.85	-5.13	1.2	6.33	0.012	1.24
227	**FIP-1209 A**	32.68	33.69	-5.13	1.01	6.14	0.014	1.42
228	**FIP-1209 B**	30.91	32.04	-5.13	1.13	6.26	0.013	1.30
229	**FIP-1210 B**	30.66	32.48	-5.13	1.82	6.95	0.008	0.81
230	**FIP-1211 A**							
231	**FIP-1211 B**	31.12	32.44	-5.13	1.32	6.45	0.011	1.14
232	**FIP-1212 A**	32.77	33.41	-5.13	0.64	5.77	0.018	1.83
233	**FIP-1212 B**	30.34	30.04	-5.13	-0.3	4.83	0.035	3.52
234	**FIP-1214 A**	31.18	32.01	-5.13	0.83	5.96	0.016	1.61
235	**FIP-1214 B**	29.92	31.52	-5.13	1.6	6.73	0.009	0.94
236	**FIP-1215 A**	32.36	33.3	-5.13	0.94	6.07	0.015	1.49
237	**FIP-1215 B**	26.66	30.86	-5.13	4.2	9.33	0.002	0.16
238	**FIP-1216 A**	30.68	32.13	-5.13	1.45	6.58	0.010	1.05
239	**FIP-1216 B**	28.37	31.89	-5.13	3.52	8.65	0.002	0.25
240	**FIP-1217 A**	30.06	30.71	-5.13	0.65	5.78	0.018	1.82
241	**FIP-1217 B**	31.42	31.66	-5.13	0.24	5.37	0.024	2.42
242	**FIP-1218 A**	33.73	34.85	-5.13	1.12	6.25	0.013	1.31
243	**FIP-1218 B**							
244	**FIP-1219 A**	33.87	36.05	-5.13	2.18	7.31	0.006	0.63
245	**FIP-1219 B**	31.06	32.81	-5.13	1.75	6.88	0.008	0.85
246	**FIP-1220 A**	32.79	35.08	-5.13	2.29	7.42	0.006	0.58
247	**FIP-1220 B**	31.8	33.73	-5.13	1.93	7.06	0.007	0.75
248	**FIP-1225 A**	33.73	33.99	-5.13	0.26	5.39	0.024	2.38
249	**FIP-1225 B**	31.34	31.75	-5.13	0.41	5.54	0.021	2.15
250	**FIP-1226 A**	30.55	32.95	-5.13	2.4	7.53	0.005	0.54

251	**FIP-1226 B**	30.16	32.64	-5.13	2.48	7.61	0.005	0.51
252	**FIP-2001 A**	31.52	33.2	-5.13	1.68	6.81	0.009	0.89
253	**FIP-2002 A**							
254	**FIP-2003 A**	30.88	33.58	-5.13	2.7	7.83	0.004	0.44
255	**FIP-2003 A**	25.53	29.78	-2.4	4.25	6.65	0.010	1.00

Appendix J. Raw Cp values and analysis for the *COL2A1* and *DHFRL1* SMART-MSP Assays

DHFR 19bp Deletion

CTGCAGCGCCAGGGTCCACCTGGTCGGCTGCACCTGTGGAGGAGGAGGTGGATTTCAGGCTTCC
CGTAGACTGGAAGAATCGGCTCAAAACCGCTTGCCTCGCAGGGGCTGAGCTGGAGGCAGCGAGG
CCGCCCGACGCAGGCTTCCGGCGAGACATGGCAGGGCAAGGATGGCAGCCCGGCGGCAGGGCCC
GGCGAGGAGCGCGAACCCGCGGCCGCAGTTCCCAGGCGTCTGCGGGCGCGAGCACGCCGCGACC
CTGCGTGCGCCGGGGCGGGGGGGCGGGGCCTCGCCTGCACAAATAGGGACGAGGGGGCGGGGCG
GCCACAATTTCGCGCCAAACTTGACCGCGCGTTCTGCTGTAACGAGCGGGCTCGGAGGTCCTCC
CGCTGCTGTCATGGTTGGTTCGCTAAACTGCATCGTCGCTGTGTCCCAGAACATGGGCATCGGC
AAGAACGGGGACCTGCCCTGGCCACCGCTCAGGTATCTGCCGGGCCGGGGCGATGGGACCCAAA
CGGGCGCAGGCTGCCACGGTCGGGGTACCTGGGCGGGACGCGCCAGGCCGACTCCCGGCGAGA
GGATGGGGCCAGACTTGCGGTCTGCGCTGGCAGGAAGGGTGGGCCCGACTGGATTCCCCTTTTC
TGCTGCGCGGGAGGCCCAGTTGCTGATTTCTGCCCGGATTCTGCTGCCCGGTGAGGTCTTTGCC
CTGCGGCGCCCTCGCCCAGGGCAAAGTCCCAGCCCTGGAGAAAACACCTCACCCCTACCCACAG
CGCTCCGTTTGTCAGGTGCCTTAGAGCTCGAGCCCAAGGGATAATGTTTCGAGTAACGCTGTTT
CTCTAACTTGTAGGAATGAATTCAGATATTTCCAGAGAATGACCACAACCTCTTCAGTAGAAGG
TAATGTGGGATTAAGTAGGGTCTTGCTTGATGAAGTTTACCAGTGCAAATGTTAGTTAAATGGA
AAGTTTTCCGTGTTAATCTGGGACCTTTTCTCTTATTATGGATCTGTATGATCTGTATGCAGTT
CCCAAGGTTCATTTACCATTATTAAAAAATTTTTGTCTTAGAAATTTTATGTATGTCAACGCAC
GAGCAAATTATCAGGCATGGGGCAGAATTGGCAACTGGGTGGAGGCTTCGGTGGAGGTTAGCAC
TCCGAAAGGAAAACAGAGTAGGCCTTTGGAACAGCTGCTGGAAGAGATAAGGCCTGAACAAGGG
CAGTGGAGAAGAGAGGGTAAAAATTTTTTAAGGTTACATGACCCTGGATTTTGGAGATC

ATG – translational start site
Primers
Sensor probe
Anchor probe
19 bp insertion / deletion

Appendix K. Sequence of *DHFR* gene indicating the location of primers and probes used for genotyping the *DHFR* 19 bp deletion / insertion polymorphism.

DHFR Gene Family Multiple Sequence Alignment

```
DHFR            TCCCAGACAGAACCTACTATGTGCGGCGGCAGCTGGGGCGGGAAGGCGGGAGCTGGGGGC 60
DHFRP1          ------------------------------------------------------------
DHFRL1          ------------------------------------------------------------
DHFRP2          ------------------------------------------------------------

DHFR            GCTGGGGGCGCTGCGGCCGCTGCGGCCGCTGCAGCCGCTGCAGCGCCAGGGTCCACCTGG 120
DHFRP1          ------------------------------------------------------------
DHFRL1          ------------------------------------------------------------
DHFRP2          ------------------------------------------------------------

DHFR            TCGGCTGCACCTGTGGAGGAGGAGGTGGATTTCAGGCTTCCCGTAGACTGGAAGAATCGG 180
DHFRP1          ------------------------------------------------------------
DHFRL1          ------------------------------------------------------------
DHFRP2          ------------------------------------------------------------

DHFR            CTCAAAACCGCTTGCCTCGCAGGGGCTGAGCTGGAGGCAGCGAGGCCGCCCGACGCAGGC 240
DHFRP1          ------------------------------------------------------------
DHFRL1          ------------------------------------------------------------
DHFRP2          ------------------------------------------------------------

DHFR            TTCCGGCGAGACATGGCAGGGCAAGGATGGCAGCCCGGCGGCAGGGCCTGGCGAGGAGCG 300
DHFRP1          ------------------------------------------------------------
DHFRL1          ------------------------------------------------------------
DHFRP2          ------------------------------------------------------------

DHFR            CGAGCCCGCGGCCGCAGTTCCCAGGCGTCTGCGGGCGCGAGCACGCCGCGACCCTGCGTG 360
DHFRP1          ------------------------------------------------------------
DHFRL1          ------------------------------------------------------------
```

```
DHFRP2          ------------------------------------------------------------

DHFR            CGCCGGGGCGGGGGGGCGGGGCCTCGCCTGCACAAATGGGGACGAGGGGGGCGGGGCGGC 420
DHFRP1          ------------------------------------------------------------
DHFRL1          --------------------------------AAAAATTTGAAGAAGGG------------ 17
DHFRP2          ------------------------------------------------------------

DHFR            CACAATTTCGCGCCAAACTTGACCGCGCGTTCTGCTGTAACGAGCGGGCTCGGAGGTCCT 480
DHFRP1          -ACAATTTCGCGCCAAACTTGACCGCGCGTTCTGCTGTAACGAGCGGGCTCGGAGGTCCT 59
DHFRL1          ---AATTTCGCGGCATTCTTGGCCTGGCTTCCTGGCGTAGCCAGCAAGTTCGGAGGTGTT 74
DHFRP2          -----TTCCG-GCCAATCTTGGCCGCGCGTCCTGGTGTAGCGAGCGGGCTTGGAGGAGCT 54
                     ** ** * **  **** **  ** * ***  *** * ***  * * *****   *

DHFR            C-CCGCTGCTGTCATGGTTGGTTCGCTAAACTGCATCGTCGCTGTGTCCCAGAACATGGG 539
DHFRP1          C-CCGCTGCTGTCATGGTTGGTTCGCTAAACTGCATCGTCGCTGTGTCCCAGAACATGGG 118
DHFRL1          AACCGCTGCTGTCATGTTTCTTTTGCTAAACTGCATCGTCGCTGTGTCCCAAAACATGGG 134
DHFRP2          CACCGCTGCTGTCATGGTTCGTTTGCTAAACTGCATCGTCGCTGTGTCTCAGAACATGGG 114
                  ************** **  ** ************************ ** ********

DHFR            CATCGGCAAGAACGGGGACCTGCCCTGGCCACCGCTCAGGAATGAATTCAGATATTTCCA 599
DHFRP1          CATCGGCAAGAACGGGGACCTGCCCTGGCCACCGCTCAGGAATGAATTCAGATATTTCCA 178
DHFRL1          CATCGGCAAGAACGGGGACCTGCCCAGGCCGCCGCTCAGGAATGAATTCAGGTATTTCCA 194
DHFRP2          CATCGGCAAGAACGGGGACCTGCCCTGACCGCCGCTCAGGAATGAATTTAGGTATTTCCA 174
                ************************ * ** ***************** ** ********

DHFR            GAGAATGACCACAACCTCTTCAGTAGAAGGTAAACAGAATCTGGTGATTATGGGTAAGAA 659
DHFRP1          GAGAATGACCACAACCTCTTCAGTAGAAGGTAAACAGAATCTGGTGATTATGGGTAAGAA 238
DHFRL1          GAGAATGACCACAACTTCTTCAGTAGAGGGTAAACAGAATCTGGTGATTATGGGTAGGAA 254
DHFRP2          GAGAATGAC-ACAATCTCTTCAGTAGAAGGTAAACAGAATCTGGTGATTTTGTGTAGGAA 233
                ********* ****  *********** ********************** ** *** ***

DHFR            GACCTGGTTCTCCATTCCTGAGAAGAATCGACCTTTAAAGGGTAGAATTAATTTAGTTCT 719
DHFRP1          GACCTGGTTCTCCATTCCTGAGAAGAATCGACCTTTAAAGGGTAGAATTAATTTAGTTCT 298
DHFRL1          GACCTGGTTCTCCATTCCTGAGAAGAATCGACCTTTAAAGGATAGAATTAATTTAGTTCT 314
DHFRP2          GACCTGGTTCTCCATTCCTGGGAAGAATCGACCTTCAAAGGATAGAATTAATTTAGTTCT 293
```

```
                ******************** ************** ***** ******************


DHFR            CAGCAGAGAACTCAAGGAACCTCCACAAGGAGCTCATTTTCTTTCCAGAAGTCTAGATGA 779
DHFRP1          CAGCAGAGAACTCAAGGAACCTCCACAAGGAGCTCATTTTCTTTCCAGAAGTCTAGATGA 358
DHFRL1          CAGCAGAGAACTCAAGGAACCTCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGA 374
DHFRP2          CAGCAGAGAACTCAAGGAACCTCCACAACGAGCTCATTTTCTTGCCAGAAGTCTGGTTGA 353
                **************************** ************** ******** * * ***

DHFR            TGCCTTAAAACTTACTGAACAACCAGAATTAGCAAATAAAGTAGACATGGTCTGGATAGT 839
DHFRP1          TGCCTTAAAACTTACTGAACAACCAGAATTAGCAAATAAAGTAGACATGGTCTGGATAGT 418
DHFRL1          TGCCTTAAAACTTACTGAACGACCAGAATTAGCAAATAAAGTAGACATGATTTGGATAGT 434
DHFRP2          GATCTTAAAACTTACTTAACAACCAGAATTAGCAAATAGAGTAGACATGATTTGGATAGT 413
                   ************* *** ***************** ********** * ********

DHFR            TGGTGGCAGTTCTGTTTATAAGGAAGCCATGAATCACCCAGGCCATCTTAAACTATTTGT 899
DHFRP1          TGGTGGCAGTTCTGTTTATAAGGAAGCCATGAATCACCCAGGCCATCTTAAACTATTTGT 478
DHFRL1          TGGTGGCAGTTCTGTTTATAAGGAAGCCATGAATCACCTAGGCCATCTTAAACTATTTGT 494
DHFRP2          TGGTGGCAGTTCTGTTTATAAGGAAGCCATGAGTCGCCCAGGCCATCTTAAACTATTTGT 473
                ******************************** ** ** ********************

DHFR            GACAAGGATCATGCAAGACTTTGAAAGTGACACGTTTTTTCCAGAAATTGATTTGGAGAA 959
DHFRP1          GACAAGGATCATGCAAGACTTTGAAAGTGACACGTTTTTTCCAGAAATTGATTTGGAGAA 538
DHFRL1          GACAAGGATCATGCAGGACTTTGAAAGTGACACGTTTTTTTCAGAAATTGACTTGGAGAA 554
DHFRP2          GACAAAGATCATGCAGGACTTTGAAAGTGACACGTTTTTTCCAGAAATTGATTTGGAGAA 533
                ***** ********* ************************ ********** ********

DHFR            ATATAAACTTCTGCCAGAATACCCAGGTGTTCTCTCTGATGTCCAGGAGGAGAAAGGCAT 1019
DHFRP1          ATATAAACTTCTGCCAGAATACCCAGGTGTTCTCTCTGATGTCCAGGAGGAGAAAGGCAT 598
DHFRL1          ATATAAACTTCTGCCTGAATACCCAGGTGTTCTCTCTGATGTCCAGGAGGGGAAACACAT 614
DHFRP2          ATATAAAATTCTGCCAGAATATCCAGGCGTTCTCTCGGATGT----------------AT 577
                ******* ******* ***** ***** ******** *****                ** 

DHFR            TAAGTACAAATTTGAAGTATATGAGAAGAATGATTAATATGAAGGTGTTTTCTAGTTTA- 1078
DHFRP1          TAAGTACAAATTTGAAGTATATGAGAAGAATGATTAATATGAAGGTGTTTTCTAGTTTA- 657
DHFRL1          CAAGTACAAATTTGAAGTATGTGAGAAGGATGATTAATATGAAGGTGTTTTCTGGTTTA- 673
DHFRP2          TAAGTACAAATTTCAAGTATATGAGAAGAATGATTAATATGAAGGTGTTTTCTGGTTTAT 637
```

```
                 ************ ****** ******* ************************ *****

DHFR             ----AGTTGTTCCCCCTCCCTCTGAAAAAA-GTATGTATTTTTACATTAGAAAAGG-T-T 1131
DHFRP1           ----AGTTGTTCCCCCTCCCTCTGAAAAAA-GTATGTATTTTTACATTAGAAAAGG-T-T 710
DHFRL1           ----AGTTGTTCCCCCTCCCTCTGAGAAAA-GTATGCATTTTTACATTAGAAAAGG-GAC 727
DHFRP2           TTTAAGTTGTTCCCCCTCCCTCTGAAAAAAAGTTTATATTTTTACATTAGAAAAAAAGAC 697
                     ********************** **** ** *  *****************

DHFR             TTTTGTTGACTTTAGATCTAT----AATTATTTCTAAGCAACTAGTTTTTATTCCCCACT 1187
DHFRP1           TTTTGTTGACTTTAGATCTAT----AATTATTTCTAAGCAACTAGTTTTTATTCCCCACT 766
DHFRL1           TTTTGTTGACTTCAGATCTATGGATAATTATTTCTAAGCAACGTGTTTTTATTCCTCACT 787
DHFRP2           TTTTGTTGACTTTAGATCTTTGGATAATTATTTCTAAGCAACATGTTTTTACTCCCCACT 757
                 ************ ****** *    *****************  ******* *** ****

DHFR             ACTCTTGTCTCTATCAGATACCATTTATGAGACATTCTTGCTATAACTAAGTGCTTCTCC 1247
DHFRP1           ACTCTTGTCTCTATCAGATACCATTTATGAGACATTCTTGCTATAACTAAGTGCTTCTCC 826
DHFRL1           AATCTTGGCTATATCAGATACCATTTATGAAACATTCTTGCTATAACT----GTCTCTCC 843
DHFRP2           AATCTTGACTATATCAGATACCACTTATGAAATATTCTTGCTATAATTAAGTGCCTCTCC 817
                 * ***** ** ************ ****** * ************* *    *  *****

DHFR             AAGACCCCAACTGAGTCCCCAGCACCTGCTACAGTGAGCTGCCATTCCACACCCATCACA 1307
DHFRP1           AAGACCCCAACTGAGTCCCCAGCACCTGCTATAGTGAGCTGCCATTCCACACCCATCACA 886
DHFRL1           AAGACCCCGACTGAGTCCCCAGCACCTGCTACAGTGAGCT-------------------- 883
DHFRP2           AAGACCCTGACTGAGTCCCCAGCACCTGCTACAGTGAGCTGCCATTCCACACCCATCGCA 877
                 *******  ********************** ********

DHFR             TGTGGCACTCTTGCCAGTCCTTGACATTGTCGGGCTTTTCACATGTTGGTAATATTTATT 1367
DHFRP1           TGTGGCACTCTTGCCAGTCCTTGACATTGTCGGGCTTTTCACATGTTGGTAATATTTATT 946
DHFRL1           ------------------------------------------------------------
DHFRP2           TATGGGACTCTTGCCAGTCCTTGACATTGTCGGGCTTTTCAAATGTTGGTAGTATTTCTT 937

DHFR             AAAGATGAAGATCCACATACCCTTCAACTGAGCAGTTTCACTAGTGGAAAT 1418
DHFRP1           AAAGATGAAGATCCACATAC------------------------------ 966
DHFRL1           -------------------------------------------------
DHFRP2           AAAGATGAAGATGCACATACCCTCCAACT---------------------- 966
```

GTT->ATT Val->Ile SNP rs17855824 primers

CCT->GCT Pro->Ala SNP rs61739170 primers

ATG->ATT Met->Ile SNP rs114936057 primers

SNPS IN RED

Appendix L. A Multiple Sequence Alignment of the *DHFR* Gene Family and the location of the *DHFRL1* polymorphisms and primers used.

Genotyping Sequence Alignments for the 3 *DHFRL1* polymorphisms in the Coriell samples examined

SNP1 rs17855824 polymorphism

```
D               -NNNNNNNNNNNNNNNNNNNNNNNTGCNGGNCTTTGAAGTGACACGTTTTTTTCNGAAAT 59
E               ---NNNNNNNNNNNNNNNNNNNCATGCNGGNCTTTGAAGTGACACGTTTTTTTCNGAAAT 57
G               -NNNNNNNNNNNNNNNNNNNNNCATGCAGGACTTTGAAGTGACACGTTTTTTTCNGAAAT 59
C               ----NNNNNNNNNNNNNNNNNCATGCAGGNNTTTGAAGTGACACGTTTTTTTCNGAAAT 56
F               -NNNNNNNNNNNNNNNNNNNNNCATGCAGGNNTTTGAAGTGACACGTTTTTTTCNGAAAT 59
B               NNNNNNNNNNNNNNNNNNNGNNCATGCAGGNNTTTGAAGTGACACGTTTTTTTCAGAAAT 60
H               --NNNNNNNNNNNNNNNNNGNNCATGCAGGNNTTTGAAGTGACACGTTTTTTTCAGAAAT 58
A               -NNNNNNNNNNNNNNNNNNGNNCATGCAGGNCTTTGAAGTGACACGTTTTTTTCAGAAAT 59
I               ----NNNNNNNNNGNNNNNGANCATGCAGGACTTTGAAGTGACACGTTTTTTTCAGAAAT 56
J               ----NNNNNNNNNGNNNNNGANCATGCAGGACTTTGAAGTGACACGTTTTTTTCAGAAAT 56
                    ********* *****  *  *** **  ********************** *****

D               TGACTTGGAGAAATATAAACTTCTGCCTGAATACCCAGGTGTTCTCTCTGATGTCCAGGA 119
E               TGACTTGGAGAAATATAAACTTCTGCCTGAATACCCAGGTGTTCTCTCTGATGTCCAGGA 117
G               TGACTTGGAGAAATATAAACTTCTGCCTGAATACCCAGGTGTTCTCTCTGATGTCCAGGA 119
C               TGACTTGGAGAAATATAAACTTCTGCCTGAATACCCAGGTGTTCTCTCTGATGTCCAGGA 116
F               TGACTTGGAGAAATATAAACTTCTGCCTGAATACCCAGGTGTTCTCTCTGATGTCCAGGA 119
B               TGACTTGGAGAAATATAAACTTCTGCCTGAATACCCAGGTGTTCTCTCTGATGTCCAGGA 120
H               TGACTTGGAGAAATATAAACTTCTGCCTGAATACCCAGGTGTTCTCTCTGATGTCCAGGA 118
A               TGACTTGGAGAAATATAAACTTCTGCCTGAATACCCAGGTGTTCTCTCTGATGTCCAGGA 119
I               TGACTTGGAGAAATATAAACTTCTGCCTGAATACCCAGGTGTTCTCTCTGATGTCCAGGA 116
J               TGACTTGGAGAAATATAAACTTCTGCCTGAATACCCAGGTGTTCTCTCTGATGTCCAGGA 116
                ************************************************************

D               GGGGAAACACATCAAGTACAAATTTGAAGTAA- 151
E               GGGGAAACACATCAAGTACAAATTTGAAGTAAN 150
G               GGGGAAACACATCAAGTACAAATTTGAAGTAA- 151
C               GGGGAAACACATCAAGTACAAATTTGAAGTAA- 148
F               GGGGAAACACATCAAGTACAAATTTGAAGTAA- 151
B               GGGGAAACACATCAAGTACAAATTTGAAGTAA- 152
H               GGGGAAACACATCAAGTACAAATTTGAAGTAA- 150
A               GGGGAAACACATCAAGTACAAATTTGAAGTAA- 151
```

```
I               GGGGAAACACATCAAGTACAAATTTGAANNA-- 147
J               GGGGAAACACATCAAGTACAAATTTGAANNA-- 147
                **************************  *
```

SNP 2 rs61739170 polymorphism

```
A               --NNNNNNNNNNTTCNNANNNATCGACCTTTAAGGNTAGAATTAATTTAGTTCTCAGCA 58
E               --NNNNNNNNNNNTCNNNNNNATCGACCTTTAAGGATAGAATTAATTTAGTTCTCAGCA 58
G               ---CNNNNNNNCNNNCNNNNNNATCGACCTTTAAGGATAGAATTAATTTAGTTCTCAGCA 57
I               ---CNNNNNNN-NNTCNNNNANATCGACCTTTAAGGATAGAATTAATTTAGTTCTCAGCA 56
D               ---NNNNNNNNNNNNNNNNNAGATCGACCTTTAAGGATAGAATTAATTTAGTTCTCAGCA 57
H               -NNNNNNNNNNNNNNNNNNNNNATCGACCTTTAAGGANAGAATTAATTTAGTTCTCAGCA 59
J               CNNNNNNNNNCNNTNCNNNNNNATCGACCTTTAAGGANAGAATTAATTTAGTTCTCAGCA 60
F               NNNNNNNNNNNNNNNNCNNANNNATCGACCTTTAAGGATAGAATTAATTTAGTTCTCAGCA 60
B               ---NNNNNNNNNNNNCNGANNNANCGACCTTTAAGGATAGAATTAATTTAGTTCTCAGCA 57
C               ---NNNNNNNNNNNNCNNANNNATCGACCTTTAAGGANAGAATTAATTTAGTTCTCAGCA 57
                    ******  *.. *. * .*.************ .**********************

A               GAGAACTCAAGGAACCTCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGATGCCT 118
E               GAGAACTCAAGGAACCTCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGATGCCT 118
G               GAGAACTCAAGGAACCTCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGATGCCT 117
I               GAGAACTCAAGGAAGCTCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGATGCCT 116
D               GAGAACTCAAGGAACCTCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGATGCCT 117
H               GAGAACTCAAGGAACCTCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGATGCCT 119
J               GAGAACTCAAGGAAGCTCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGATGCCT 120
F               GAGAACTCAAGGAACCTCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGATGCCT 120
B               GAGAACTCAAGGAACCTCCACAAGGAGCTCATTTTCTTGCCAGAAGTCTGGATGATGCCT 117
C               GAGAACTCAAGGAACCTCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGATGCCT 117
                ************** ******************************** ************

A               TAAAACTTACTGAACGACCAGAATTAGCAAATAAAGTAGACATGNTNTGGATAGTTGGTG 178
E               TAAAACTTACTGAACGACCAGAATTAGCAAATAAAGTAGACATGGTNTGGATAGTTGGTG 178
G               TAAAACTTACTGAACGACCAGAATTAGCAAATAAAGTAGACATGATTTGGATAGTTGGTG 177
I               TAAAACTTACTGAACGACCAGAATTAGCAAATAAAGTAGACATGATTTGGATAGTTGGTG 176
D               TAAAACTTACTGAACGACCAGAATTAGCAAATAAAGTAGACATGGTTTGGATAGTTGGTG 177
H               TAAAACTTACTGAACGACCAGAATTAGCAAATAAAGTAGACATGATTTGGATAGTTGGTG 179
```

```
J               TAAAACTTACTGAACGACCAGAATTAGCAAATAAAGTAGACATGATTTGGATAGTTGGTG 180
F               TAAAACTTACTGAACGACCAGAATTAGCAAATAAAGTAGACATGATTTGGATAGTTGGTG 180
B               TAAAACTTACTGAACAACCAGAATTAGCAAATAAAGTAGACATGGTCTGGATAGTTGGTG 177
C               TAAAACTTACTGAACGACCAGAATTAGCAAATAAAGTAGACATGATTTGGATAGTTGGTG 177
                **************.*************************** * *************


A               GCAGTTCTGTTTATAAGGAAGCCATGAATCACCTAGGCCATCTTAAACTAN 229
E               GCAGTTCTGTTTATAAGGAAGCCATGAATCACCTAGGCCATCTTAAACTAN 229
G               GCAGTTCTGTTTATAAGGAAGCCATGAATCACCTAGGCCATCTTAAACTAN 228
I               GCAGTTCTGTTTATAAGGAAGCCATGAATCACCTAGGCCATCTTAAACTAA 227
D               GCAGTTCTGTTTATAAGGAAGCCATGAATCACCTAGGCCATCTTAAACTN- 227
H               GCAGTTCTGTTTATAAGGAAGCCATGAATCACCTAGGCCATCTTAAACTAN 230
J               GCAGTTCTGTTTATAAGGAAGCCATGAATCACCTAGGCCATCTTAAANN-- 229
F               GCAGTTCTGTTTATAAGGAAGCCATGAATCACCTAGGCCATCTTAAANN-- 229
B               GCAGTTCTGTTTATAAGGAAGCCATGAATCACCTAGGCCATCTTAANNN-- 226
C               GCAGTTCTGTTTATAAGGAAGCCATGAATCACCTAGGCCATCTNNNNNNNN 228
                ****************************************.    .
```

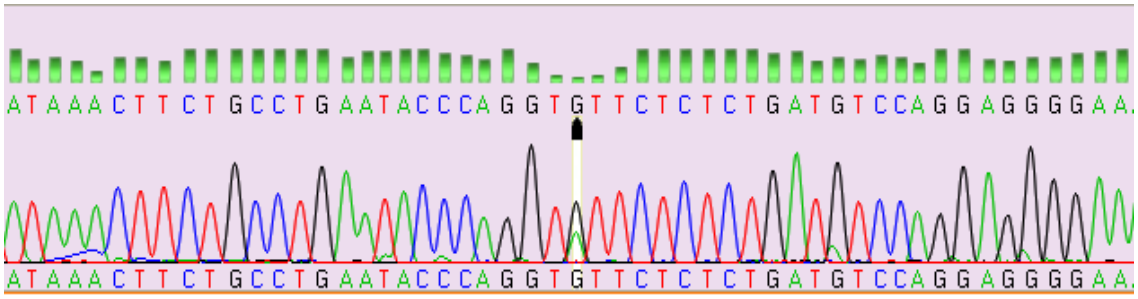
SNP 3 rs114936057 polymorphism

```
A               -------NNNNNNNNNNNNNCNCNNNNNNNNNNCATGGGCATCGGCAAGAACGGGGACCT 53
G               -------NNNNNNNNNCNNNNNCNNNGTNNNANCATGGGCATCGGCAAGAACGGGGACCT 53
C               NNNNNNNNNNNNNNNNNNNNNNNCNNNNNNNNNAACATGGGCATCGGCAAGAACGGGGACCT 60
J               ------NNNNNNNNNNNNNNNCNCNNNNNCNNAACATGGGCATCGGCAG-AACGGGGACCT 53
D               -NNNNNNNNNNNNNNNNNNNNNNCNNNNN-NNAACATGGGCATCGGCAAGAACGGGGACCT 58
E               -------NNNNNNNNNNNNNNNNCNNNGNCNAAACATGGGCATCGGCAG-AACGGGGACCT 52
I               ------NNNNNNNNNNNNNNNNNCNNNGNCNAAACATGGGCATCGGCAG-AACGGGGACCT 53
F               -----NNNNNNNNNNNNCNNNNCNNNNTCNAAACATGGGCATCGGCAAGAACGGGGACCT 55
B               ---NNNNNNNNNNNNTNNNNNNCNNNNNCNAAACATGGGCATCGGCAAGAACGGGGACCT 57
H               ----NNNNNNNTNNNNNTNNNNNGCTGTGNCAACATGGGCATCGGCAG-AACGGGGACCT 55
                       ****.***.  ** * . ... *   **************. ***********

A               GCCCAGGCCGCCGCTCAGGAATGAATTCAGGTATTTCCAGAGAATGACCACAACTTCTTC 113
```

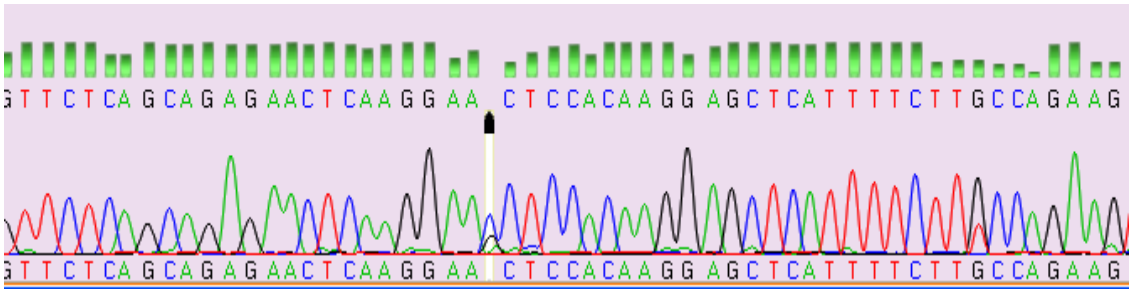
```
G               GCCCAGGCCGCCGCTCAGGAATGAATTCAGGTATTTCCAGAGAATGACCACAACTTCTTC 113
C               GCCCAGGCCGCCGCTCAGGAATGAATTCAGGTATTTCCAGAGAATGACCACAACTTCTTC 120
J               GCCCAGGCCGCCGCTCAGGAATGAATTCAGGTATTTCCAGAGAATGACCACAACTTCTTC 113
D               GCCCAGGCCGCCGCTCAGGAATGAATTCAGGTATTTCCAGAGAATGACCACAACTTCTTC 118
E               GCCCAGGCCGCCGCTCAGGAATGAATTCAGGTATTTCCAGAGAATGACCACAACTTCTTC 112
I               GCCCAGGCCGCCGCTCAGGAATGAATTCAGGTATTTCCAGAGAATGACCACAACTTCTTC 113
F               GCCCAGGCCGCCGCTCAGGAATGAATTCAGGTATTTCCAGAGAATGACCACAACTTCTTC 115
B               GCCCAGGCCGCCGCTCAGGAATGAATTCAGGTATTTCCAGAGAATGACCACAACTTCTTC 117
H               GCCCAGGCCGCCGCTCAGGAATGAATTCAGGTATTTCCAGAGAATGACCACAACTTCTTC 115
                ************************************************************

A               AGTAGAGGGTAAACAGAATCTGGTGATTATGGGTAGGAAGACCTGGTTCTCCATTCCTGA 173
G               AGTAGAGGGTAAACAGAATCTGGTGATTATGGGTAGGAAGACCTGGTTCTCCATTCCTGA 173
C               AGTAGAGGGTAAACAGAATCTGGTGATTATGGGTAGGAAGACCTGGTTCTCCATTCCTGA 180
J               AGTAGAGGGTAAACAGAATCTGGTGATTATGGGTAGGAAGACCTGGTTCTCCATTCCTGA 173
D               AGTAGAGGGTAAACAGAATCTGGTGATTATGGGTAGGAAGACCTGGTTCTCCATTCCTGA 178
E               AGTAGAGGGTAAACAGAATCTGGTGATTATGGGTAGGAAGACCTGGTTCTCCATTCCTGA 172
I               AGTAGAGGGTAAACAGAATCTGGTGATTATGGGTAGGAAGACCTGGTTCTCCATTCCTGA 173
F               AGTAGAGGGTAAACAGAATCTGGTGATTATGGGTAGGAAGACCTGGTTCTCCATTCCTGA 175
B               AGTAGAGGGTAAACAGAATCTGGTGATTATGGGTAGGAAGACCTGGTTCTCCATTCCTGA 177
H               AGTAGAGGGTAN-CAGAATCTGGTGATTATGGGTAGGAAGACCTGGTTCTCCATTCCTGA 174
                ***********  *********************************************

A               GAAGAATCGACCTTTAAAGGATAGAATTAATTTAGTTCTCAGCAGAGAACTCAAGGAACC 233
G               GAAGAATCGACCTTTAAAGGATAGAATTAATTTAGTTCTCAGCAGAGAACTCAAGGAACC 233
C               GAAGAATCGACCTTTAAAGGATAGAATTAATTTAGTTCTCAGCAGAGAACTCAAGGAACC 240
J               GAAGAATCGACCTTTAAAGGATAGAATTAATTTAGTTCTCAGCAGAGAACTCAAGGAANC 233
D               GAAGAATCGACCTTTAAAGGATAGAATTAATTTAGTTCTCAGCAGAGAACTCAAGGAAGC 238
E               GAAGAATCGACCTTTAAAGGATAGAATTAATTTAGTTCTCAGCAGAGAACTCAAGGAAGC 232
I               GAAGAATCGACCTTTAAAGGATAGAATTAATTTAGTTCTCAGCAGAGAACTCAAGGAANC 233
F               GAAGAATCGACCTTTAAAGGATAGAATTAATTTAGTTCTCAGCAGAGAACTCAAGGAACC 235
B               GAAGAATCGACCTTTAAAGGATAGAATTAATTTAGTTCTCAGCAGAGAACTCAAGGAACC 237
H               GAAGAATCGACCTTTAAAGGATAGAATTAATTTAGTTCTCAGCAGANAACTCAAGGAACC 234
                *********************************************.************ *

A               TCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGATGCCTTAAAACTTACTGN-- 290
G               TCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGATGCCTTAAAACTTACTGN-- 290
```

```
C               TCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGATGCCTTAAAACTTACTGAA- 298
J               TCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGATGCCTTAAAACTTACTGN-- 290
D               TCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGATGCCTTAAAACTTACTGAN- 296
E               TCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGATGCCTTAAAACTTACTGAN- 290
I               TCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGATGCCTTAAAACTTACTGAN- 291
F               TCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGATGCCTTAAAACTTACTGAN- 293
B               TCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGATGCCTTAAAACTTACTGAN- 295
H               TCCNCAAGGAGCTCANTTTCTNGNCAGAANNTTNNATGATGCCTN-AAACNTACNNNNN 292
                *** ***********.*****.* *****..**..*********. ****.***..
```

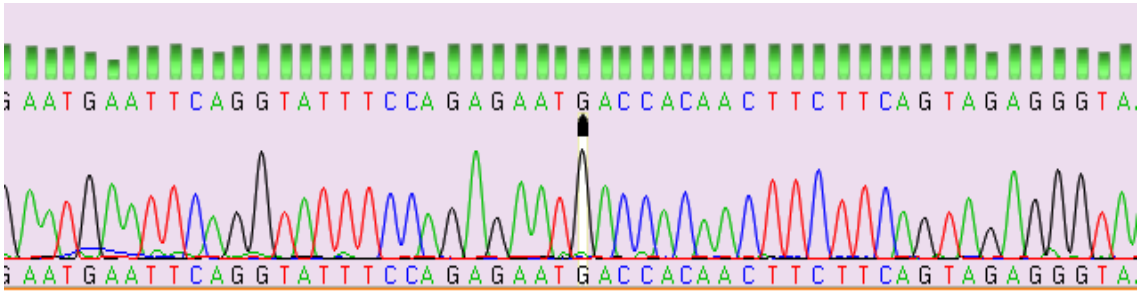
Appendix M. Examples of Genotyping Sequence Alignments for the three DHFRL1 polymorphisms in the Coriell samples examined



Chromatogram of *DHFRL1* G->A (rs17855824) polymorphism. The arrow indicates the SNP position.

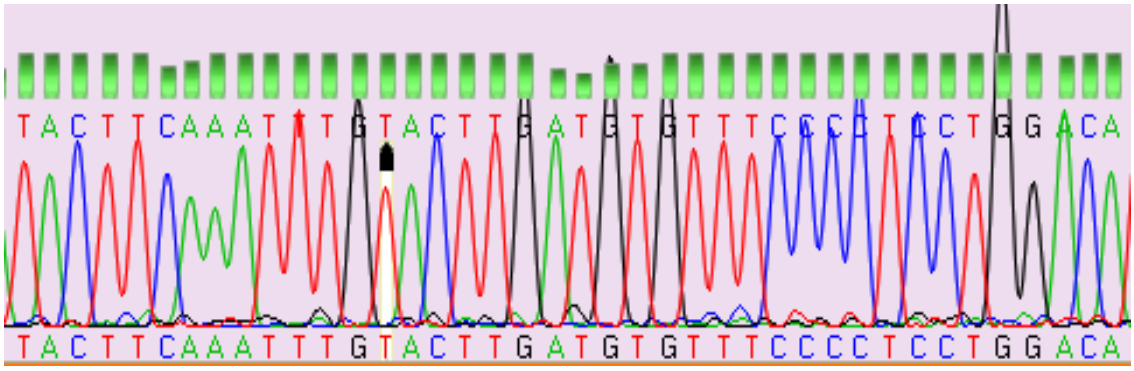


Chromatogram of *DHFRL1* C->G (rs61739170) polymorphism. The arrow indicates the SNP position.

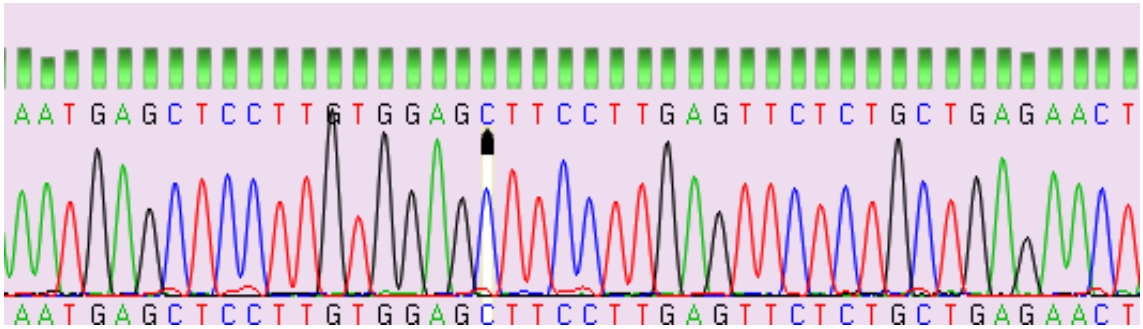


Chromatogram of *DHFRL1* G->T (rs114936057) polymorphism. The arrow indicates the SNP position.

Appendix N. Chromatograms for the 3 *DHFRL1* polymorphisms that were Sanger sequenced to determine the genotypes of the samples



Chromatogram from a sequencing run of one of the successful *DHFRL1* G->A (rs17855824) mutant clones. The arrow indicates the SNP position.



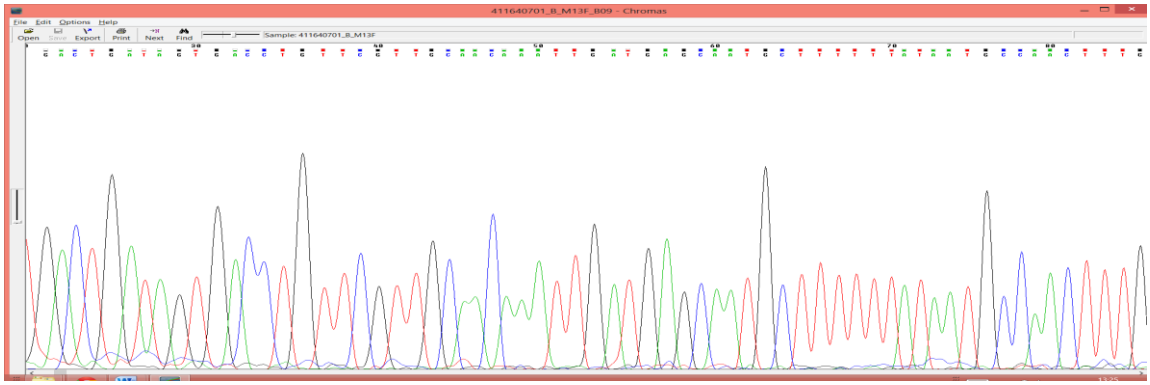
Chromatogram from a sequencing run of one of the successful *DHFRL1* C->G (rs61739170) mutant clones. The arrow indicates the SNP position.

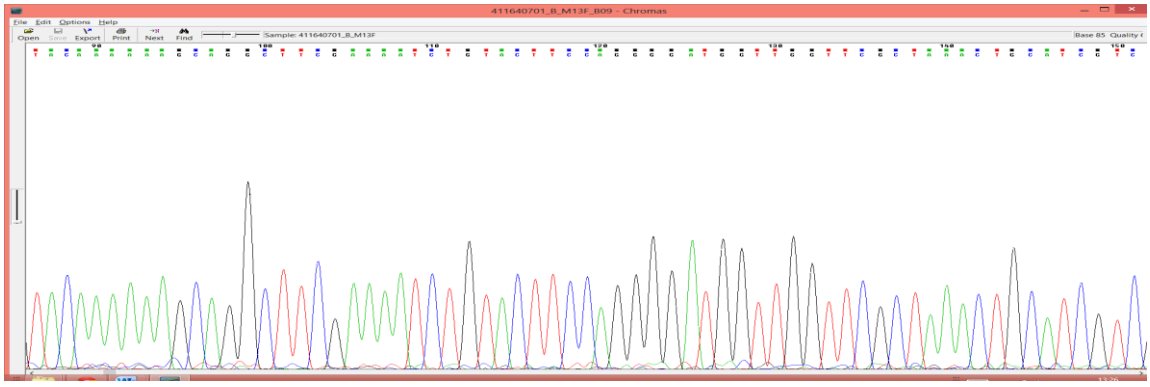
Appendix O. Chromatograms from Sanger sequencing runs of the two *DHFRL1* mutant clones to verify that the correct mutation had been inserted into the DNA samples after site directed mutagenesis.

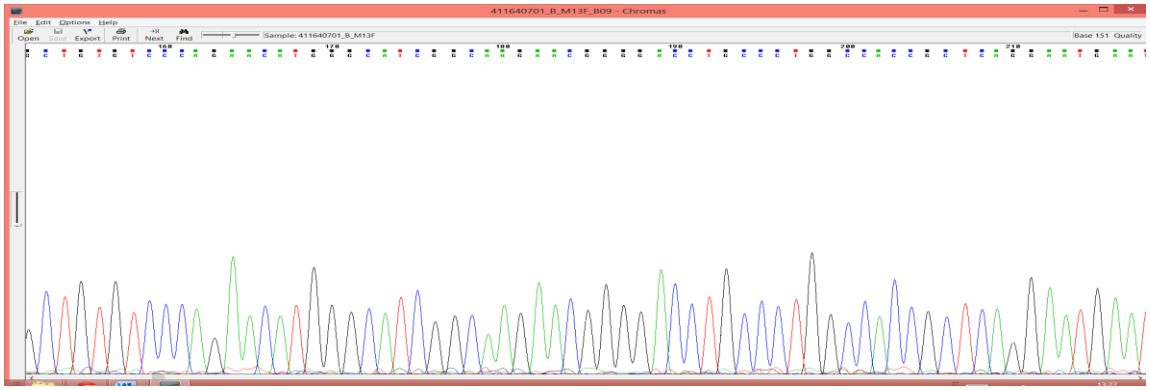
```
DHFR_HIS_TAG              --------------------------------------------------
DHFR_HIS_TAG_BP_CLONE     NNNNNNNNNGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAA 50


DHFR_HIS_TAG              ----------------------GGGGACAAGTTTGTACAAAAAAGCAGGC 28
DHFR_HIS_TAG_BP_CLONE     TTGATGAGCAATGCTTTTTTATAATGCCAACTTTGTACAAAAAAGCAGGC 100
                                              .. *.*** ******************

DHFR_HIS_TAG              TTCGAAAATCTGTACTTCCAGGGGATGGTTGGTTCGCTAAACTGCATCGT 78
DHFR_HIS_TAG_BP_CLONE     TTCGAAAATCTGTACTTCCAGGGGATGGTTGGTTCGCTAAACTGCATCGT 150
                          **************************************************

DHFR_HIS_TAG              CGCTGTGTCCCAGAACATGGGCATCGGCAAGAACGGGGACCTGCCCTGGC 128
DHFR_HIS_TAG_BP_CLONE     CGCTGTGTCCCAGAACATGGGCATCGGCAAGAACGGGGACCTGCCCTGGC 200
                          **************************************************

DHFR_HIS_TAG              CACCGCTCAGGAATGAATTCAGATATTTCCAGAGAATGACCACAACCTCT 178
DHFR_HIS_TAG_BP_CLONE     CACCGCTCAGGAATGAATTCAGATATTTCCAGAGAATGACCACAACCTCT 250
                          **************************************************

DHFR_HIS_TAG              TCAGTAGAAGGTAAACAGAATCTGGTGATTATGGGTAAGAAGACCTGGTT 228
DHFR_HIS_TAG_BP_CLONE     TCAGTAGAAGGTAAACAGAATCTGGTGATTATGGGTAAGAAGACCTGGTT 300
                          **************************************************

DHFR_HIS_TAG              CTCCATTCCTGAGAAGAATCGACCTTTAAAGGGTAGAATTAATTTAGTTC 278
DHFR_HIS_TAG_BP_CLONE     CTCCATTCCTGAGAAGAATCGACCTTTAAAGGGTAGAATTAATTTAGTTC 350
                          **************************************************

DHFR_HIS_TAG              TCAGCAGAGAACTCAAGGAACCTCCACAAGGAGCTCATTTTCTTTCCAGA 328
DHFR_HIS_TAG_BP_CLONE     TCAGCAGAGAACTCAAGGAACCTCCACAAGGAGCTCATTTTCTTTCCAGA 400
                          **************************************************

DHFR_HIS_TAG              AGTCTAGATGATGCCTTAAAACTTACTGAACAACCAGAATTAGCAAATAA 378
```

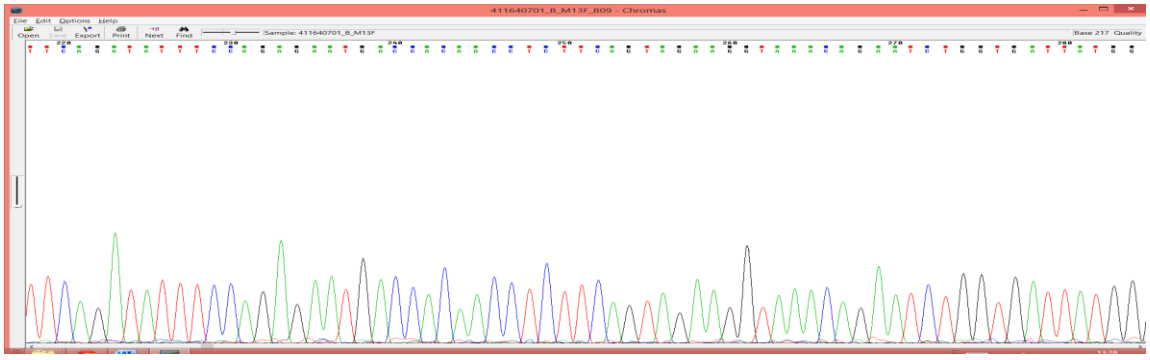
```
DHFR_HIS_TAG_BP_CLONE      AGTCTAGATGATGCCTTAAAACTTACTGAACAACCAGAATTAGCAAATAA 450
                           **************************************************

DHFR_HIS_TAG               AGTAGACATGGTCTGGATAGTTGGTGGCAGTTCTGTTTATAAGGAAGCCA 428
DHFR_HIS_TAG_BP_CLONE      AGTAGACATGGTCTGGATAGTTGGTGGCAGTTCTGTTTATAAGGAAGCCA 500
                           **************************************************

DHFR_HIS_TAG               TGAATCACCCAGGCCATCTTAAACTATTTGTGACAAGGATCATGCAAGAC 478
DHFR_HIS_TAG_BP_CLONE      TGAATCACCCAGGCCATCTTAAACTATTTGTGACAAGGATCATGCAAGAC 550
                           **************************************************

DHFR_HIS_TAG               TTTGAAAGTGACACGTTTTTTCCAGAAATTGATTTGGAGAAATATAAACT 528
DHFR_HIS_TAG_BP_CLONE      TTTGAAAGTGACACGTTTTTTCCAGAAATTGATTTGGAGAAATATAAACT 600
                           **************************************************

DHFR_HIS_TAG               TCTGCCAGAATACCCAGGTGTTCTCTCTGATGTCCAGGAGGAGAAAGGCA 578
DHFR_HIS_TAG_BP_CLONE      TCTGCCAGAATACCCAGGTGTTCTCTCTGATGTCCAGGAGGAGAAAGGCA 650
                           **************************************************

DHFR_HIS_TAG               TTAAGTACAAATTTGAAGTATATGAGAAGAATGATCATCATCATCATCAT 628
DHFR_HIS_TAG_BP_CLONE      TTAAGTACAAATTTGAAGTATATGAGAAGAATGATCATCATCATCATCAT 700
                           **************************************************

DHFR_HIS_TAG               CATTAGGACCCAGCTTTCTTGTACAAAGTGG------------------- 659
DHFR_HIS_TAG_BP_CLONE      CATTAGGACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCATTG 750
                           ***************************** *

DHFR_HIS_TAG               --------------------------------------------------
DHFR_HIS_TAG_BP_CLONE      CTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCA 800
```

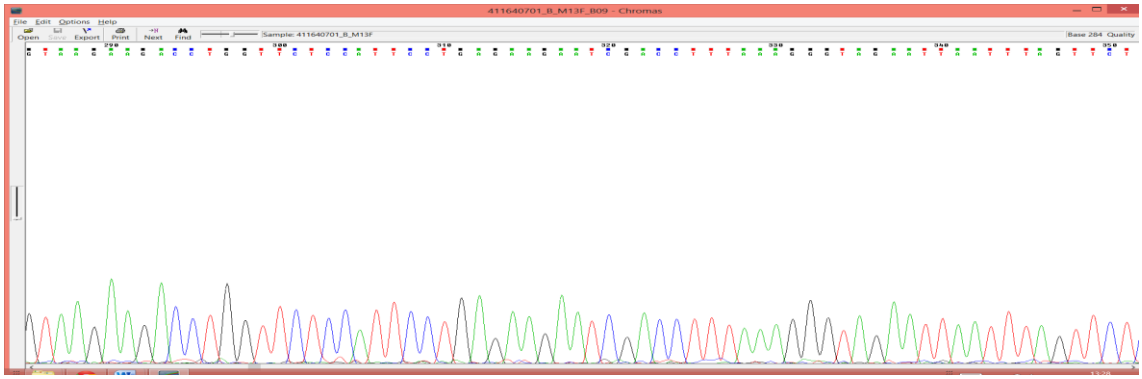
Appendix P. Clustal Sequence Alignment of the *DHFR* BP Clone Incorporating the HIS Tag

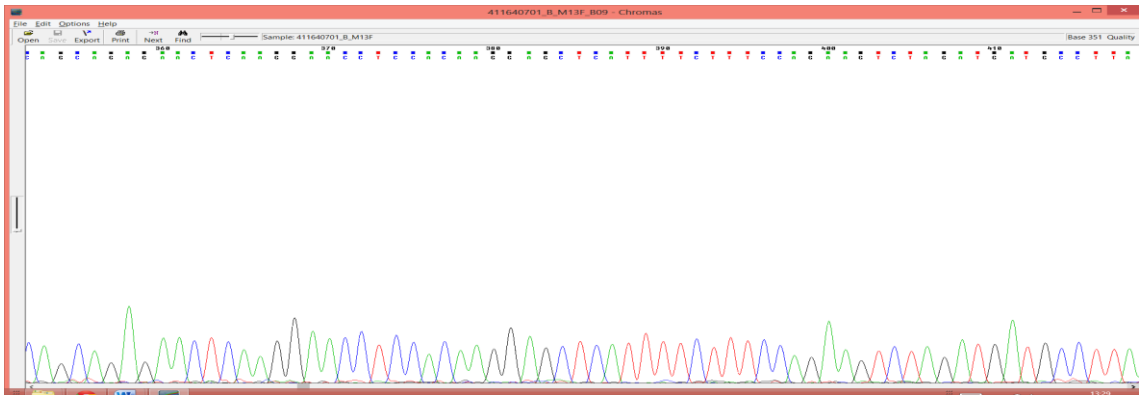

411640701_B_M13F_809 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411640701_B_M13F

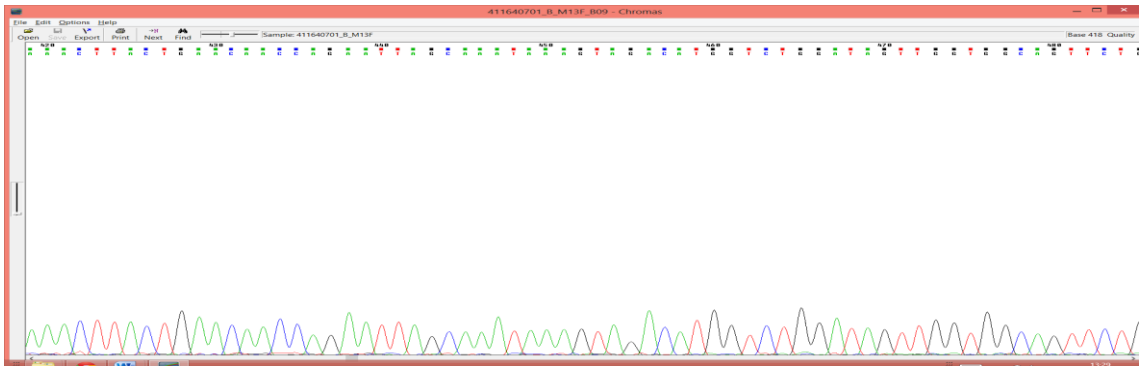

411640701_B_M13F_809 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411640701_B_M13F

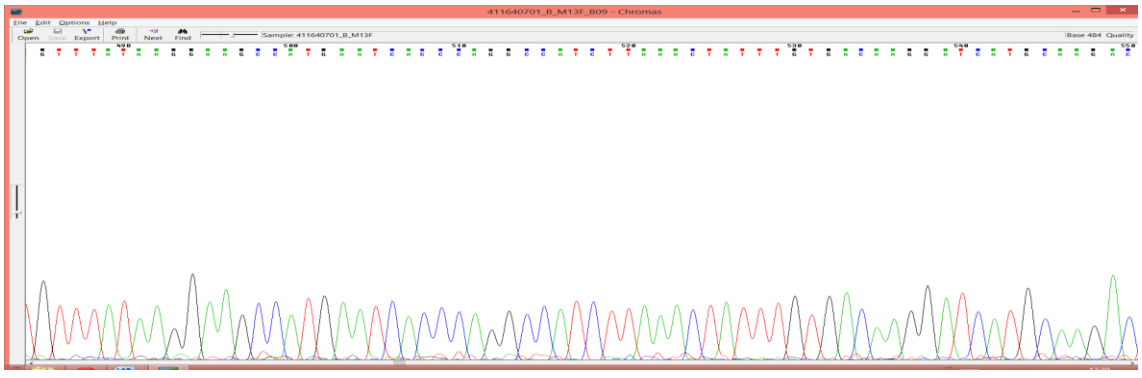

411640701_B_M13F_809 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411640701_B_M13F

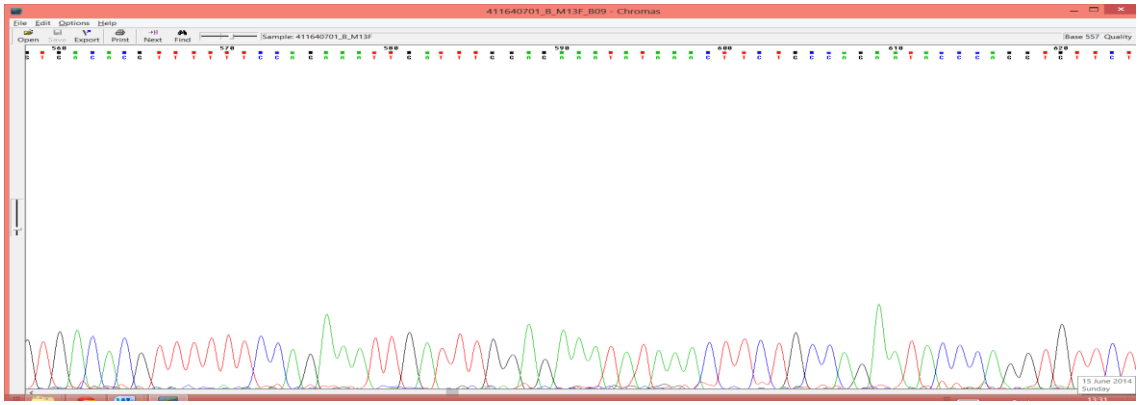

411640701_B_M13F_809 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411640701_B_M13F

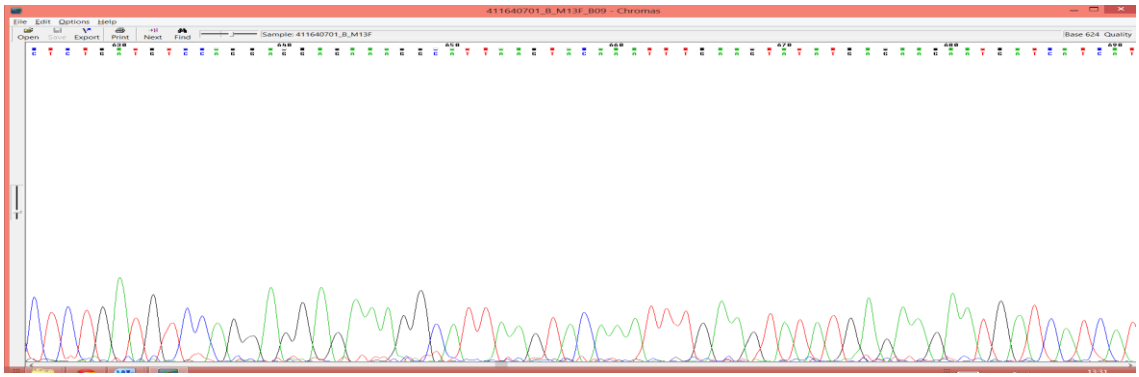

411640701_B_M13F_809 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411640701_B_M13F
Base 284 Quality

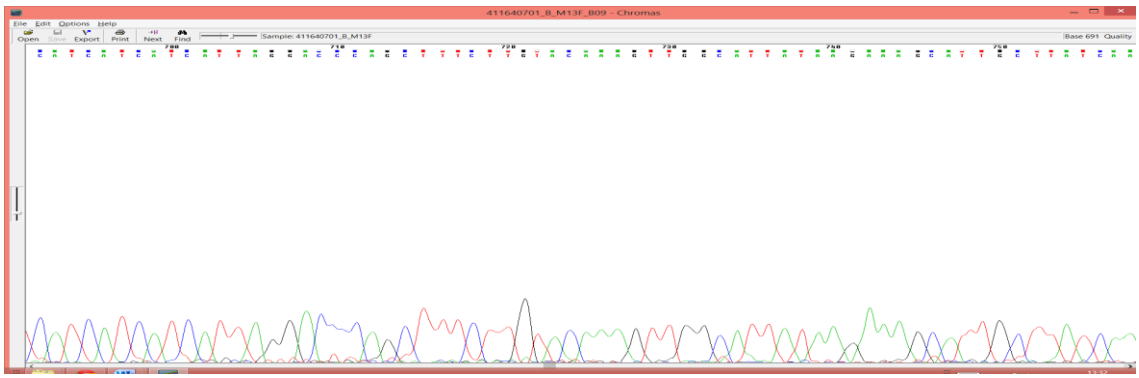

411640701_B_M13F_809 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411640701_B_M13F
Base 351 Quality

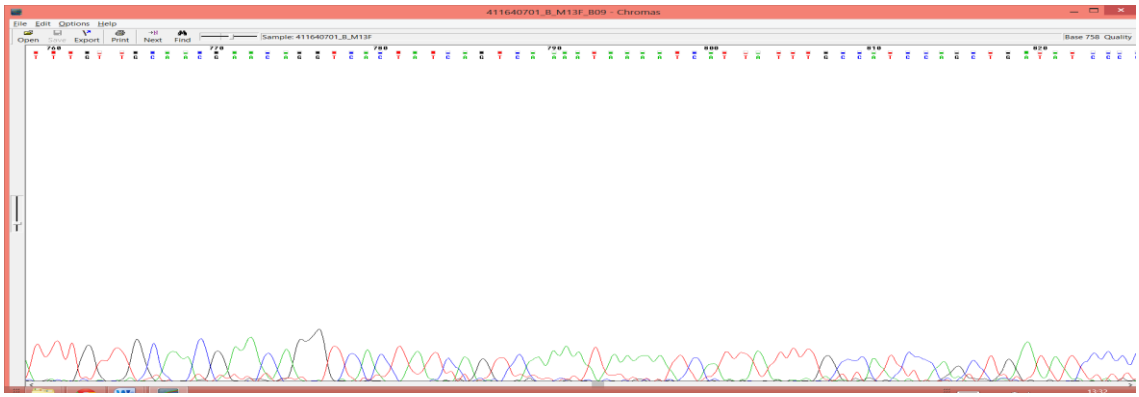

411640701_B_M13F_809 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411640701_B_M13F
Base 418 Quality


411640701_B_M13F_809 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411640701_B_M13F
Base 484 Quality







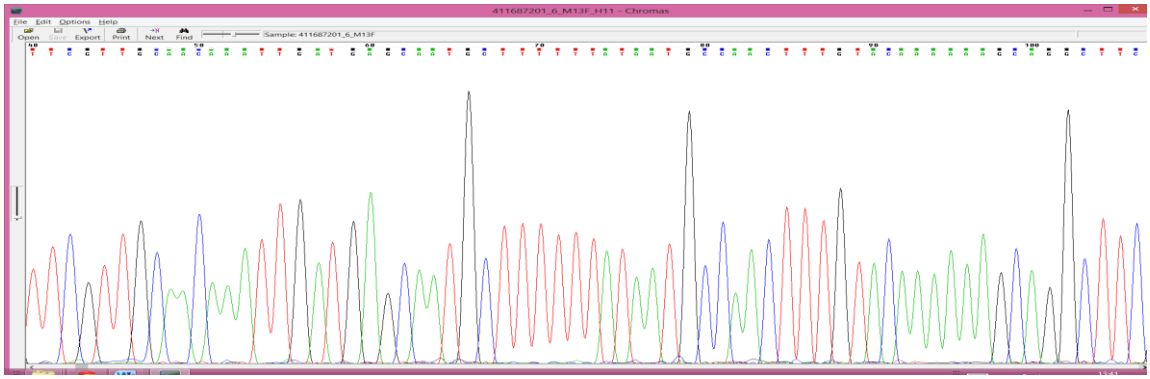


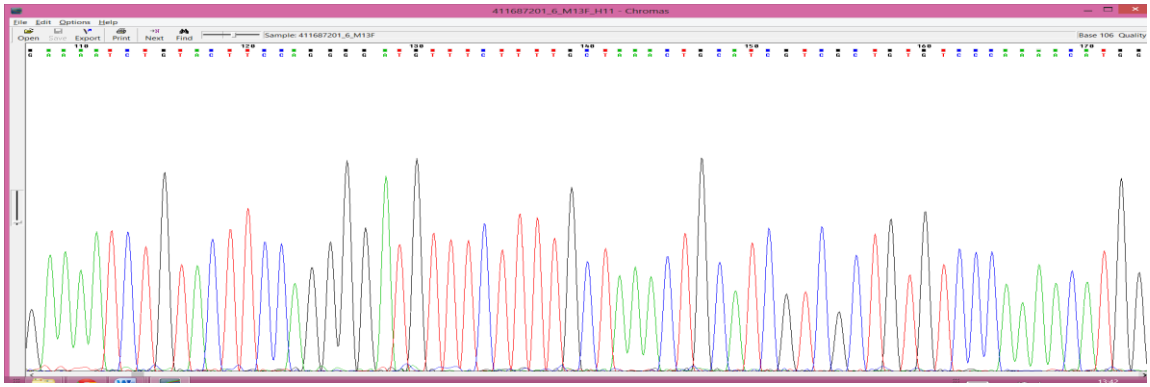
Appendix Q. Chromatogram Sequence of the *DHFR* BP Clone Incorporating the HIS Tag

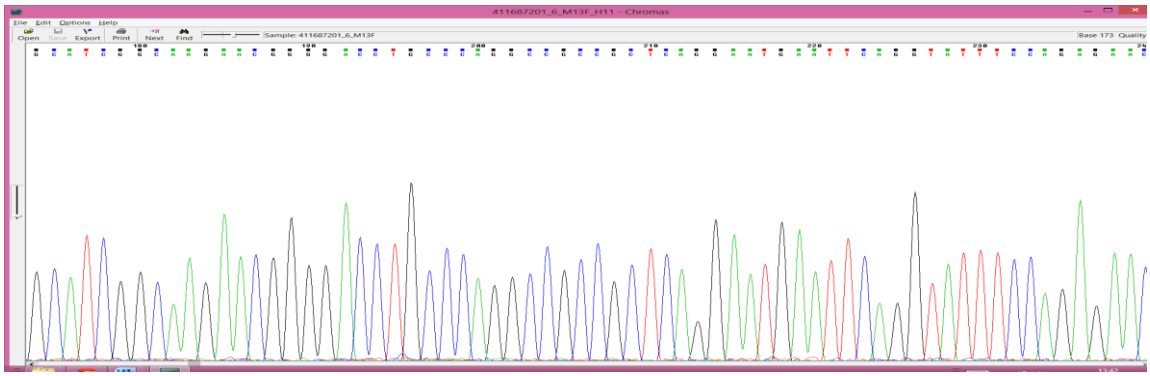
```
DHFRL1_HIS_TAG              --------------------------------------------------
DHFRL1_HIS_TAG_BP_Clone     NNNNNNNNNNNNGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAAC 50


DHFRL1_HIS_TAG              ----------------------------ACAAGTTTGTACAAAAAAGCA 21
DHFRL1_HIS_TAG_BP_Clone     AAATTGATGAGCAATGCTTTTTTATAATGCCAACTTTGTACAAAAAAGCA 100
                                                       .*** ***************

DHFRL1_HIS_TAG              GGCTTCGAAAATCTGTACTTCCAGGGGATGTTTCTTTTGCTAAACTGCAT 71
DHFRL1_HIS_TAG_BP_Clone     GGCTTCGAAAATCTGTACTTCCAGGGGATGTTTCTTTTGCTAAACTGCAT 150
                            **************************************************

DHFRL1_HIS_TAG              CGTCGCTGTGTCCCAAAACATGGGCATCGGCAAGAACGGGGACCTGCCCA 121
DHFRL1_HIS_TAG_BP_Clone     CGTCGCTGTGTCCCAAAACATGGGCATCGGCAAGAACGGGGACCTGCCCA 200
                            **************************************************

DHFRL1_HIS_TAG              GGCCGCCGCTCAGGAATGAATTCAGGTATTTCCAGAGAATGACCACAACT 171
DHFRL1_HIS_TAG_BP_Clone     GGCCGCCGCTCAGGAATGAATTCAGGTATTTCCAGAGAATGACCACAACT 250
                            **************************************************

DHFRL1_HIS_TAG              TCTTCAGTAGAGGGTAAACAGAATCTGGTGATTATGGGTAGGAAGACCTG 221
DHFRL1_HIS_TAG_BP_Clone     TCTTCAGTAGAGGGTAAACAGAATCTGGTGATTATGGGTAGGAAGACCTG 300
                            **************************************************

DHFRL1_HIS_TAG              GTTCTCCATTCCTGAGAAGAATCGACCTTTAAAGGATAGAATTAATTTAG 271
DHFRL1_HIS_TAG_BP_Clone     GTTCTCCATTCCTGAGAAGAATCGACCTTTAAAGGATAGAATTAATTTAG 350
                            **************************************************

DHFRL1_HIS_TAG              TTCTCAGCAGAGAACTCAAGGAACCTCCACAAGGAGCTCATTTTCTTGCC 321
DHFRL1_HIS_TAG_BP_Clone     TTCTCAGCAGAGAACTAAAGGAACCTCCACAAGGAGCTCATTTTCTTGCC 400

                            ***************.**********************************
```

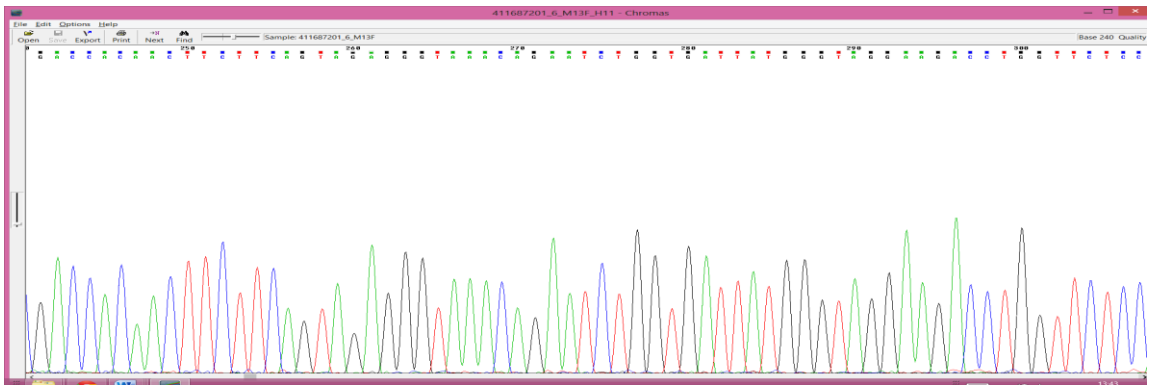
```
DHFRL1_HIS_TAG              AGAAGTTTGGATGATGCCTTAAAACTTACTGAACGACCAGAATTAGCAAA 371
DHFRL1_HIS_TAG_BP_Clone     AGAAGTTTGGATGATGCCTTAAAACTTACTGAACGACCAGAATTAGCAAA 450
                            **************************************************

DHFRL1_HIS_TAG              TAAAGTAGACATGATTTGGATAGTTGGTGGCAGTTCTGTTTATAAGGAAG 421
DHFRL1_HIS_TAG_BP_Clone     TAAAGTAGACATGATTTGGATAGTTGGTGGCAGTTCTGTTTATAAGGAAG 500
                            **************************************************

DHFRL1_HIS_TAG              CCATGAATCACCTAGGCCATCTTAAACTATTTGTGACAAGGATCATGCAG 471
DHFRL1_HIS_TAG_BP_Clone     CCATGAATCACCTAGGCCATCTTAAACTATTTGTGACAAGGATCATGCAG 550
                            **************************************************

DHFRL1_HIS_TAG              GACTTTGAAAGTGACACGTTTTTTTCAGAAATTGACTTGGAGAAATATAA 521
DHFRL1_HIS_TAG_BP_Clone     GACTTTGAAAGTGACACGTTTTTTTCAGAAATTGACTTGGAGAAATATAA 600
                            **************************************************

DHFRL1_HIS_TAG              ACTTCTGCCTGAATACCCAGGTGTTCTCTCTGATGTCCAGGAGGGGAAAC 571
DHFRL1_HIS_TAG_BP_Clone     ACTTCTGCCTGAATACCCAGGTGTTCTCTCTGATGTCCAGGAGGGGAAAC 650
                            **************************************************

DHFRL1_HIS_TAG              ACATCAAGTACAAATTTGAAGTATGTGAGAAGGATGATCATCATCATCAT 621
DHFRL1_HIS_TAG_BP_Clone     ACATCAAGTACAAATTTGAAGTATGTGAGAAGGATGATCATCATCATCAT 700
                            **************************************************

DHFRL1_HIS_TAG              CATCATTAGGACCCAGCTTTCTTGTACAAAGTGGT--------------- 656
DHFRL1_HIS_TAG_BP_Clone     CATCATTAGGACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCA 750
                            ******************************** *

DHFRL1_HIS_TAG              -------------------------------------------------
DHFRL1_HIS_TAG_BP_Clone     TTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAA 800
```

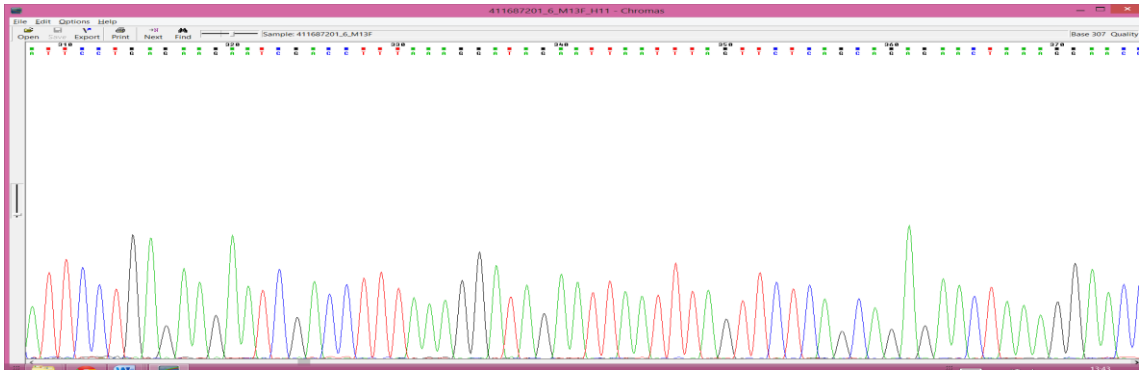
Appendix R. Clustal Sequence Alignment of the *DHFRL1* BP Clone Incorporating the HIS Tag

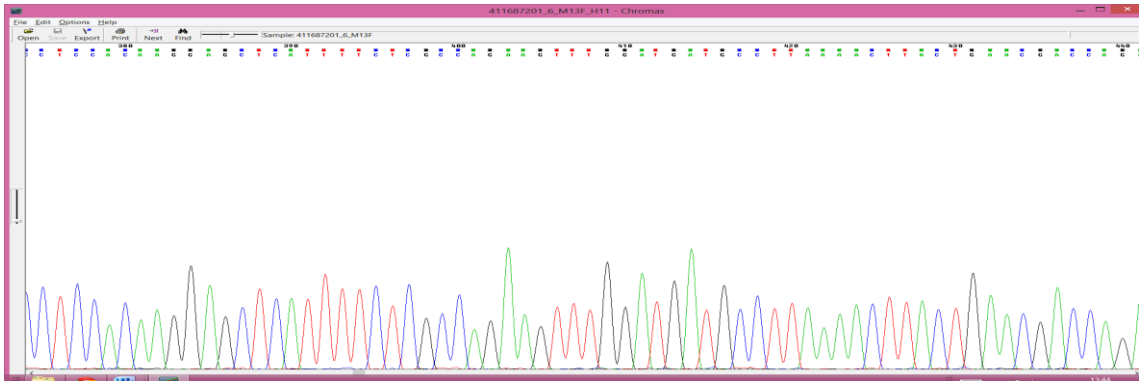

411687201_6_M13F_H11 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411687201_6_M13F

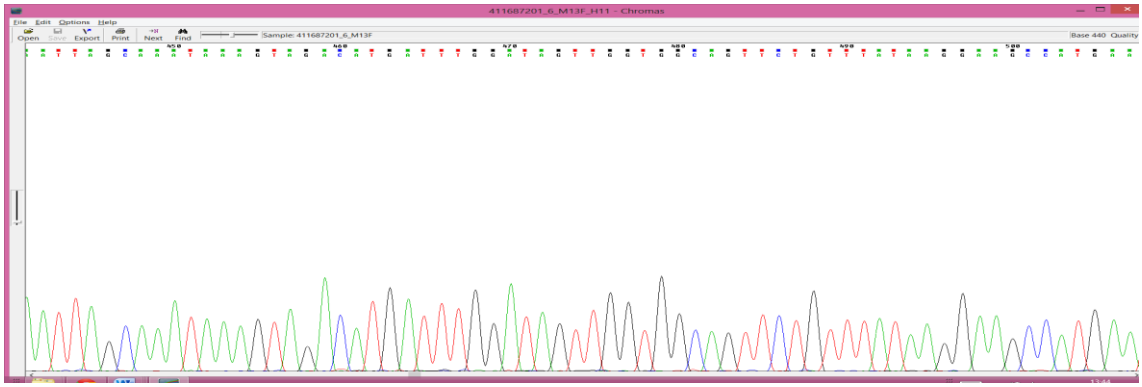

411687201_6_M13F_H11 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411687201_6_M13F
Base 106 Quality

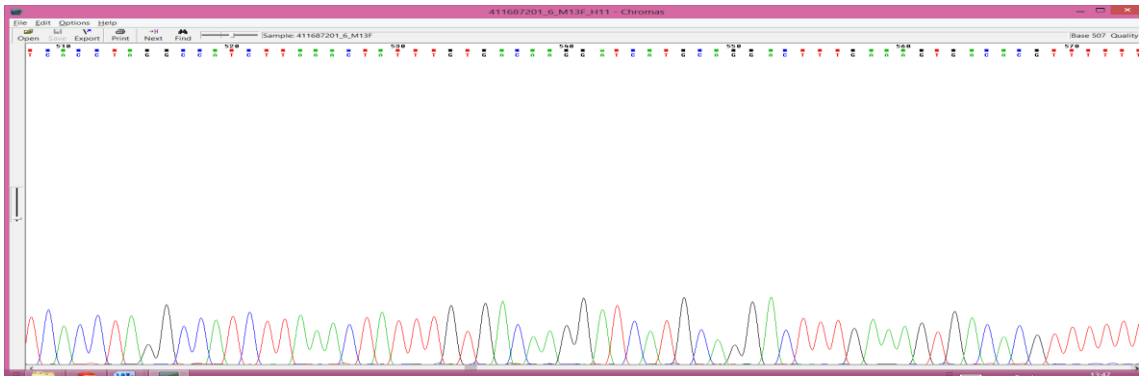

411687201_6_M13F_H11 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411687201_6_M13F
Base 173 Quality

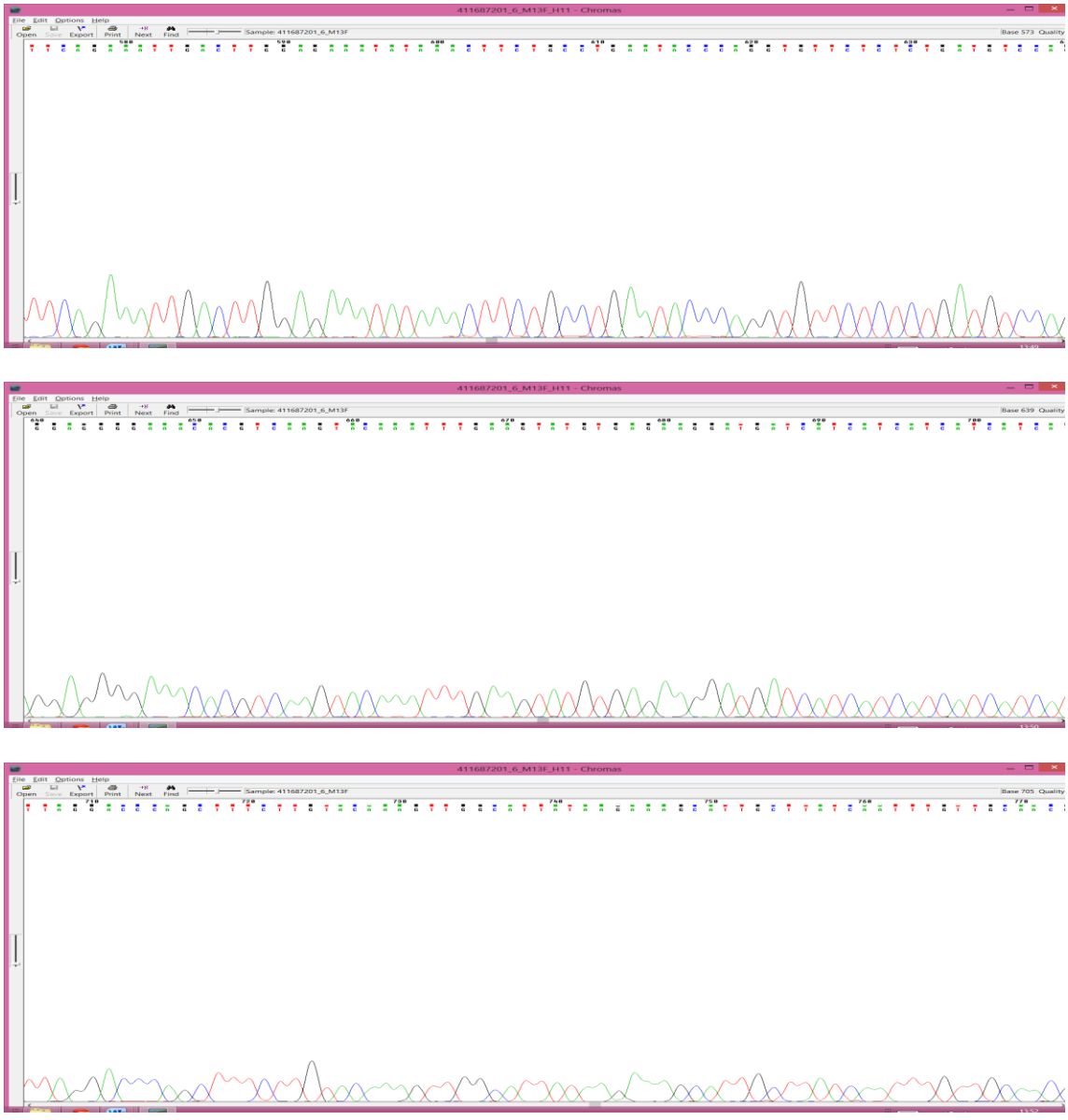

411687201_6_M13F_H11 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411687201_6_M13F
Base 240 Quality


411687201_6_M13F_H11 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411687201_6_M13F
Base 307 Quality


411687201_6_M13F_H11 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411687201_6_M13F


411687201_6_M13F_H11 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411687201_6_M13F
Base 440 Quality


411687201_6_M13F_H11 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411687201_6_M13F
Base 507 Quality

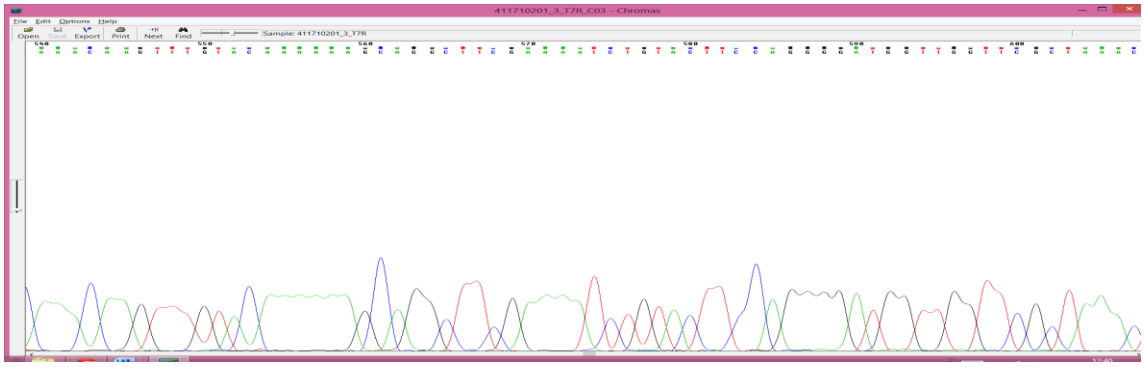


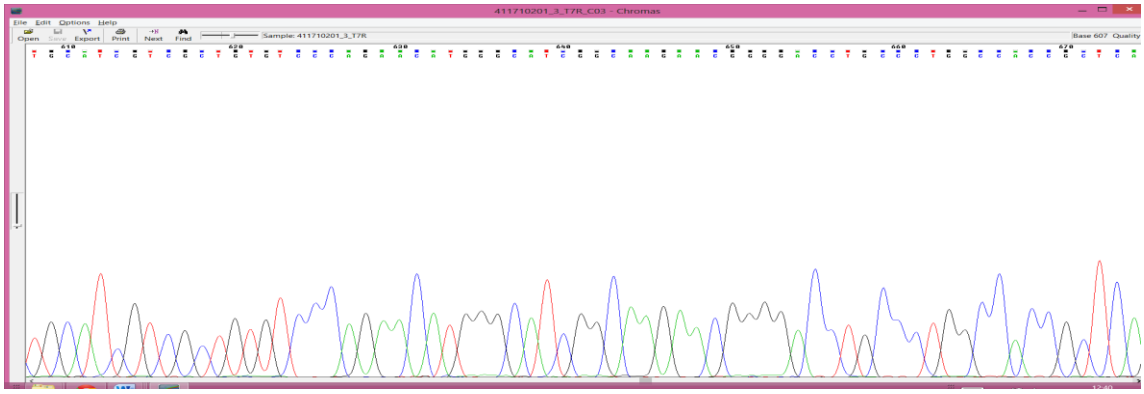
Appendix S. Chromatogram Sequence of the *DHFRL1* BP Clone Incorporating the HIS Tag

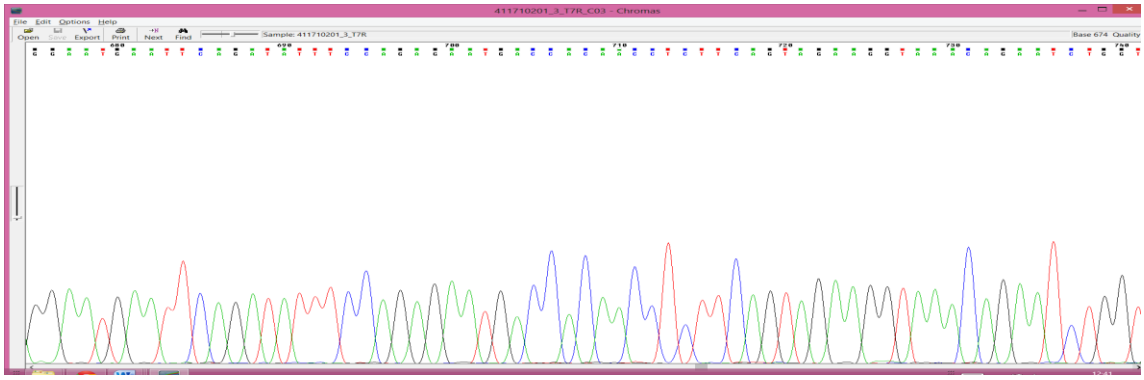
```
DHFR_HIS_TAG          ------------------------------------------------------------
DHFR_HIS_TAG_Clone    TGATAAGTACTTGAAATCCAGCAAGTATATAGCATGGCCTTTGCAGGGCTGGCAAGCCAC 480

DHFR_HIS_TAG          ------------------------------------------------------------
DHFR_HIS_TAG_Clone    GTTTGGTGGTGGCGACCATCCTCCAAAATCGGATCTGGTTCCGCGTCCATGGTCGAATCA 540

DHFR_HIS_TAG          ------TTTGTACAAAAAAGCAGGCTTCGAAAATCTGTACTTCCAGGGGATGGTTGGTTC 54
DHFR_HIS_TAG_Clone    AACAAGTTTGTACAAAAAAGCAGGCTTCGAAAATCTGTACTTCCAGGGGATGGTTGGTTC 600
                            **************************************************

DHFR_HIS_TAG          GCTAAACTGCATCGTCGCTGTGTCCCAGAACATGGGCATCGGCAAGAACGGGGACCTGCC 114
DHFR_HIS_TAG_Clone    GCTAAACTGCATCGTCGCTGTGTCCCAGAACATGGGCATCGGCAAGAACGGGGACCTGCC 660
                      ************************************************************

DHFR_HIS_TAG          CTGGCCACCGCTCAGGAATGAATTCAGATATTTCCAGAGAATGACCACAACCTCTTCAGT 174
DHFR_HIS_TAG_Clone    CTGGCCACCGCTCAGGAATGAATTCAGATATTTCCAGAGAATGACCACAACCTCTTCAGT 720
                      ************************************************************

DHFR_HIS_TAG          AGAAGGTAAACAGAATCTGGTGATTATGGGTAAGAAGACCTGGTTCTCCATTCCTGAGAA 234
DHFR_HIS_TAG_Clone    AGAAGGTAAACAGAATCTGGTGATTATGGGTAAGAAGACCTGGTTCTCCATTCCTGAGAA 780
                      ************************************************************

DHFR_HIS_TAG          GAATCGACCTTTAAAGGGTAGAATTAATTTAGTTCTCAGCAGAGAACTCAAGGAACCTCC 294
DHFR_HIS_TAG_Clone    GAATCGACCTTTAAAGGGTAGAATTAATTTAGTTCTCAGCAGAGAACTCAAGGAACCTCC 840
                      ************************************************************

DHFR_HIS_TAG          ACAAGGAGCTCATTTTCTTTCCAGAAGTCTAGATGATGCCTTAAAACTTACTGAACAACC 354
DHFR_HIS_TAG_Clone    ACAAGGAGCTCATTTTCTTTCCAGAAGTCTAGATGATGCCTTAAAACTTACTGAACAACC 900
                      ************************************************************

DHFR_HIS_TAG          AGAATTAGCAAATAAAGTAGACATGGTCTGGATAGTTGGTGGCAGTTCTGTTTATAAGGA 414
```

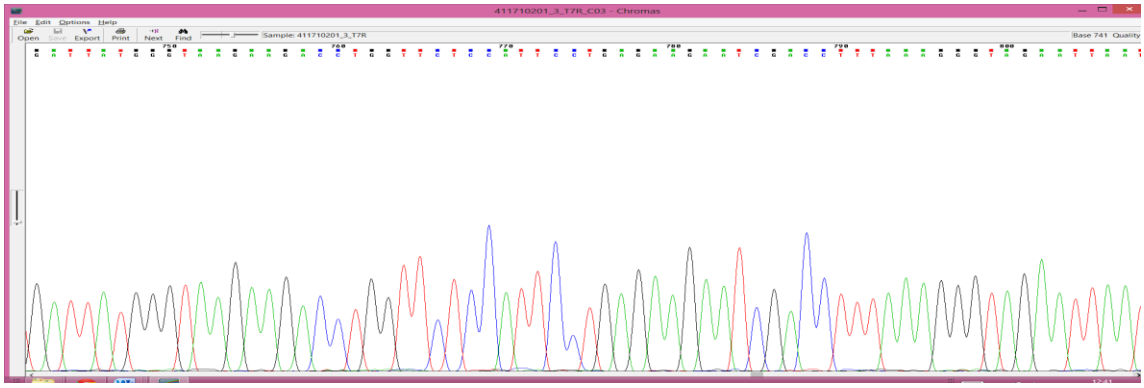
```
DHFR_HIS_TAG_Clone      AGAATTAGCAAATAAAGTAGACATGGTCTGGATAGTTGGTGGCAGTTCTGTTTATAAGGA 960
                        ************************************************************

DHFR_HIS_TAG            AGCCATGAATCACCCAGGCCATCTTAAACTATTTGTGACAAGGATCATGCAAGACTTTGA 474
DHFR_HIS_TAG_Clone      AGCCATGAATCACCCAGGCCATCTTAAACTATTTGTGACAAGGATCATGCAAGACTTTGA 1020
                        ************************************************************

DHFR_HIS_TAG            AAGTGACACGTTTTTTCCAGAAATTGATTTGGAGAAATATAAACTTCTGCCAGAATACCC 534
DHFR_HIS_TAG_Clone      AAGTGACACGTTTTTTCCAGAAATTGATTTGGAGAAATATAAACTTCTGCCAGAATACCC 1080
                        ************************************************************

DHFR_HIS_TAG            AGGTGTTCTCTCTGATGTCCAGGAGGAGAAAGGCATTAAGTACAAATTTGAAGTATATGA 594
DHFR_HIS_TAG_Clone      AGGTGTTCTCTCTGATGTCCAGGAGGAGAAAGGCATTAAGTACAAATTTGAAGTATATGA 1140
                        ************************************************************

DHFR_HIS_TAG            GAAGAATGATCATCATCATCATCATTAGGACCCAGCTTTCTTGTACAAAGT------ 648
DHFR_HIS_TAG_Clone      GAAGAATGATCATCATCATCATCATTAGGACCCAGCTTTCTTGTACAAAGTGGTTTG 1200
                        *****************************************************

DHFR_HIS_TAG            ---------------------------------------------
DHFR_HIS_TAG_Clone      ATTCGACCCGGGATCCGGCTGCTAACAAAGCCCGAANGANNNNNNNNNN 1249
```

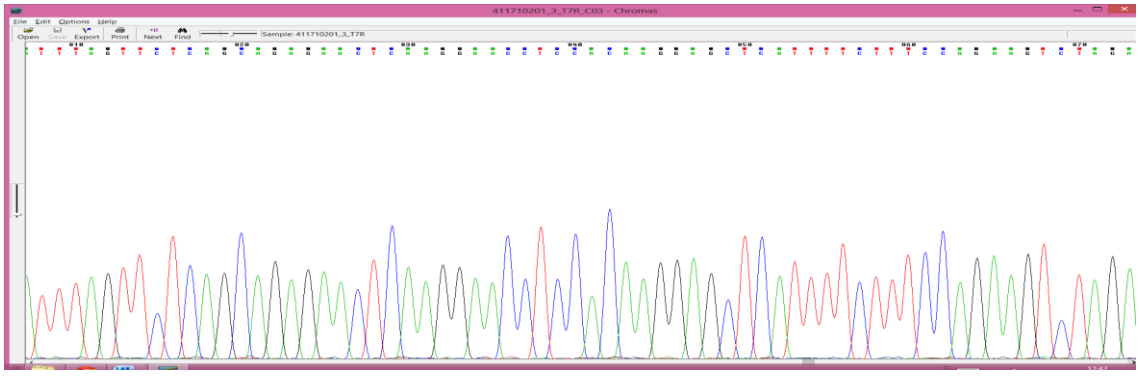
Appendix T. Clustal Sequence Alignment of Recombinant *DHFR* (LR clone) showing that the HIS Tag has Been Incorporated after Gateway Cloning

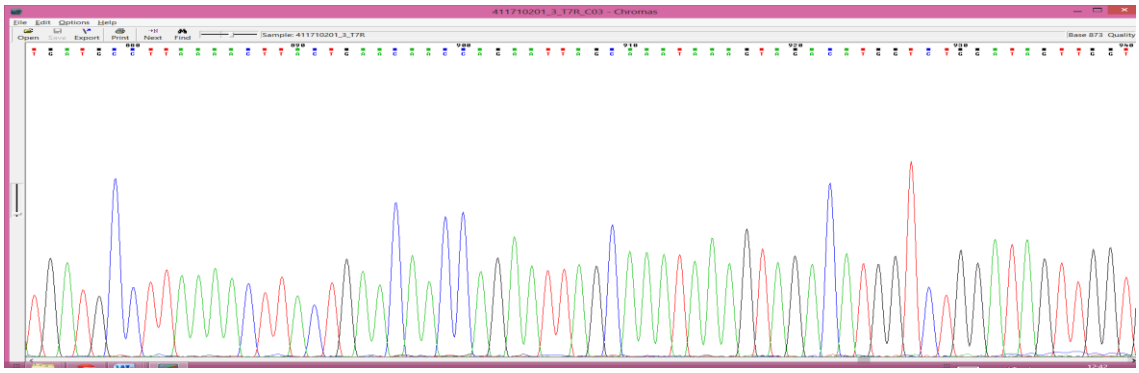

411710201_3_T7R_C03 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411710201_3_T7R

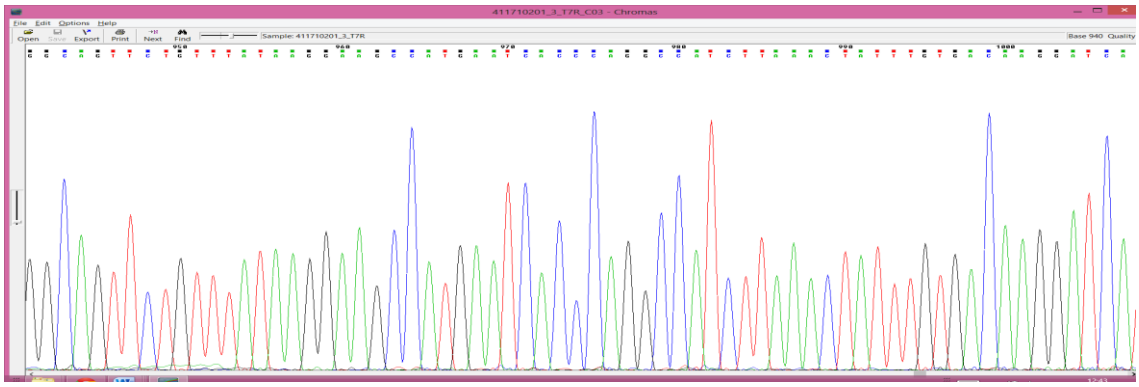

411710201_3_T7R_C03 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411710201_3_T7R
Base 607 Quality

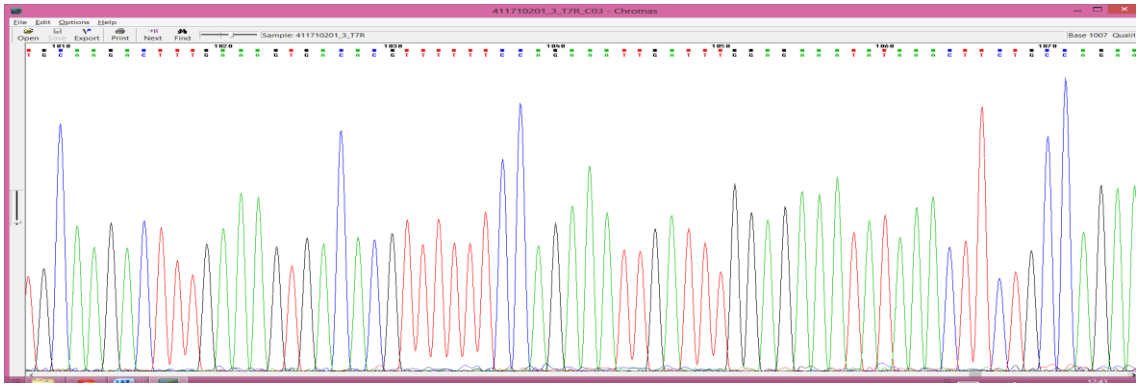

411710201_3_T7R_C03 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411710201_3_T7R
Base 674 Quality

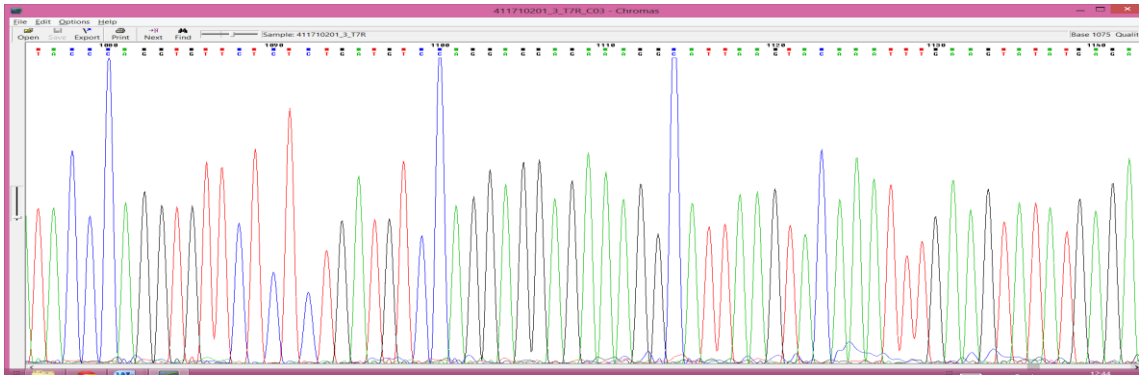

411710201_3_T7R_C03 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411710201_3_T7R
Base 741 Quality

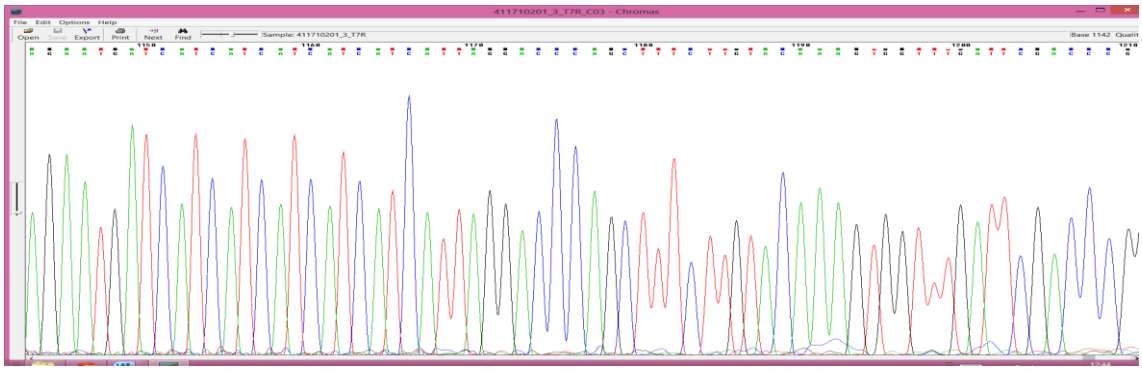

411710201_3_T7R_C03 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411710201_3_T7R


411710201_3_T7R_C03 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411710201_3_T7R


411710201_3_T7R_C03 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411710201_3_T7R


411710201_3_T7R_C03 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411710201_3_T7R



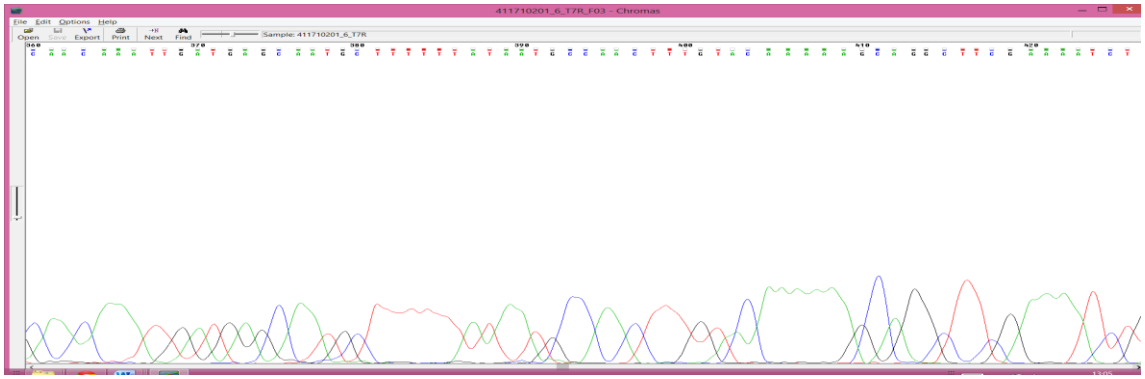


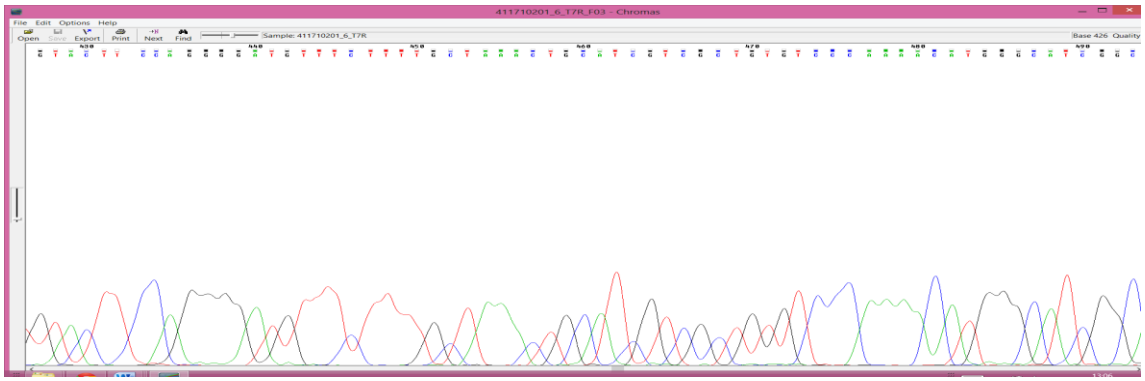
Appendix U. Chromatogram Sequence of Recombinant *DHFR* (LR clone) showing that the HIS Tag has Been Incorporated after Gateway Cloning

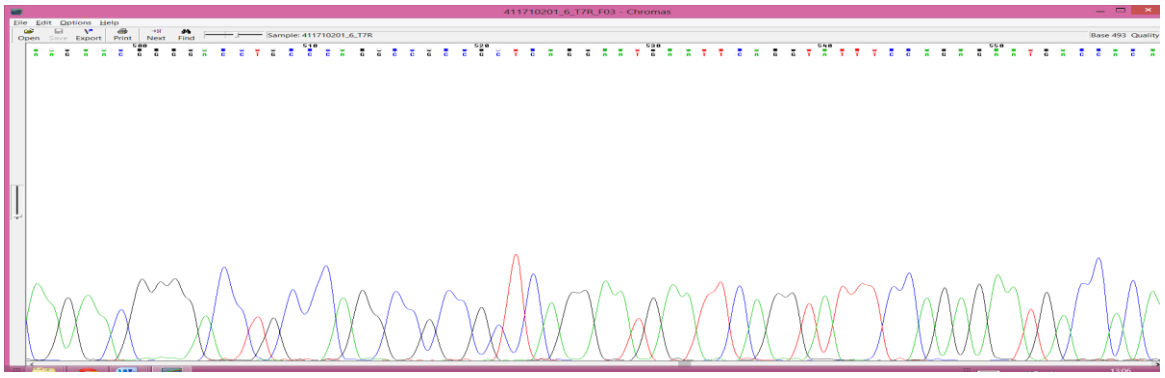
```
DHFRL1_HIS_TAG          ------------------------------------------------------------
DHFRL1_HIS_TAG_Clone    NTACTNTNGNNNACGCTAGCATGGATGTTTNCCCAGTCNNGACGTNGTAAACGACNNCCA 300


DHFRL1_HIS_TAG          ------------------------------------------------------------
DHFRL1_HIS_TAG_Clone    GTCTTAAGCTCGGNCCCCAAATAATGATTTNATTTTGACTGNNAGNNACCTGTNNGTTGC 360


DHFRL1_HIS_TAG          --------------------------------ACAAGTTTGTACAAAAAAGCAGGCTTCG  28
DHFRL1_HIS_TAG_Clone    AACAAATTGATGAGCAATGCTTTTTTATAATGCCAACTTTGTACAAAAAG-CAGGCTTCG 419
                                                       .*** ************. *********

DHFRL1_HIS_TAG          AAAATCTGTACTTCCAGGGGATGTTTCTTTTGCTAAACTGCATCGTCGCTGTGTCCCAAA  88
DHFRL1_HIS_TAG_Clone    AAAATCTGTACTTCCAGGGGATGTTTCTTTTGCTAAACTGCATCGTCGCTGTGTCCCAAA 479
                        ************************************************************

DHFRL1_HIS_TAG          ACATGGGCATCGGCAAGAACGGGGACCTGCCCAGGCCGCCGCTCAGGAATGAATTCAGGT 148
DHFRL1_HIS_TAG_Clone    ACATGGGCATCGGCAAGAACGGGGACCTGCCCAGGCCGCCGCTCAGGAATGAATTCAGGT 539
                        ************************************************************

DHFRL1_HIS_TAG          ATTTCCAGAGAATGACCACAACTTCTTCAGTAGAGGGTAAACAGAATCTGGTGATTATGG 208
DHFRL1_HIS_TAG_Clone    ATTTCCAGAGAATGACCACAACTTCTTCAGTAGAGGGTAAACAGAATCTGGTGATTATGG 599
                        ************************************************************

DHFRL1_HIS_TAG          GTAGGAAGACCTGGTTCTCCATTCCTGAGAAGAATCGACCTTTAAAGGATAGAATTAATT 268
DHFRL1_HIS_TAG_Clone    GTAGGAAGACCTGGTTCTCCATTCCTGAGAAGAATCGACCTTTAAAGGATAGAATTAATT 659
                        ************************************************************

DHFRL1_HIS_TAG          TAGTTCTCAGCAGAGAACTCAAGGAACCTCCACAAGGAGCTCATTTTCTTGCCAGAAGTT 328
DHFRL1_HIS_TAG_Clone    TAGTTCTCAGCAGAGAACTAAAGGAACCTCCACAAGGAGCTCATTTTCTTGCCAGAAGTT 719
                        ******************.*****************************************

DHFRL1_HIS_TAG          TGGATGATGCCTTAAAACTTACTGAACGACCAGAATTAGCAAATAAAGTAGACATGATTT 388
```

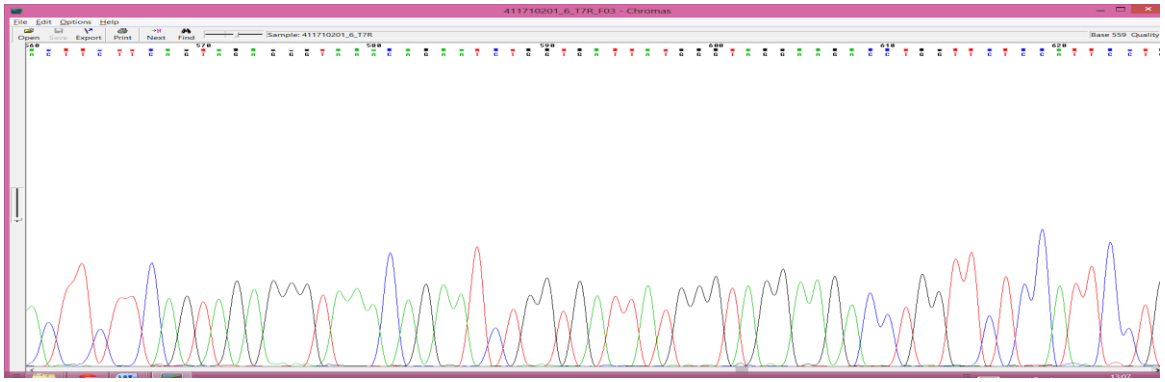
```
DHFRL1_HIS_TAG_Clone      TGGATGATGCTCTAAAACTTACTGAACGACCAGAATTAGCAAATAAAGTAGACATGATTT 779
                          **********  ********************************************

DHFRL1_HIS_TAG           GGATAGTTGGTGGCAGTTCTGTTTATAAGGAAGCCATGAATCACCTAGGCCATCTTAAAC 448
DHFRL1_HIS_TAG_Clone     GGATAGTTGGTGGCAGTTCTGTTTATAAGGAAGCCATGAATCACCTAGGCCATCTTAAAC 839
                          ************************************************************

DHFRL1_HIS_TAG           TATTTGTGACAAGGATCATGCAGGACTTTGAAAGTGACACGTTTTTTTCAGAAATTGACT 508
DHFRL1_HIS_TAG_Clone     TATTTGTGACAAGGATCATGCAGGACTTTGAAAGTGACACGTTTTTTTCAGAAATTGACT 899
                          ************************************************************

DHFRL1_HIS_TAG           TGGAGAAATATAAACTTCTGCCTGAATACCCAGGTGTTCTCTCTGATGTCCAGGAGGGGA 568
DHFRL1_HIS_TAG_Clone     TGGAGAAATATAAACTTCTGCCTGAATACCCAGGTGTTCTCTCTGATGTCCAGGAGGGGA 959
                          ************************************************************

DHFRL1_HIS_TAG           AACACATCAAGTACAAATTTGAAGTATGTGAGAAGGATGATCATCATCATCATCATCATT 628
DHFRL1_HIS_TAG_Clone     AACACATCAAGTACAAATTTGAAGTATGTGAGAAGGATGATCATCATCATCATCATCATT 1019
                          ************************************************************

DHFRL1_HIS_TAG           AGGACCCAGCTTTCTTGTACAAAGTGGT-------------------------------- 656
DHFRL1_HIS_TAG_Clone     AGGACCCAGCTTTCTTGTACAAAGTGGTTTGATTCGACCCGGGATCCGGCTGCTAACAAA 1079
                          ****************************

DHFRL1_HIS_TAG           ---------------------
DHFRL1_HIS_TAG_Clone     GCCCGAAAGNAGNNNNNNNNN 1100
```

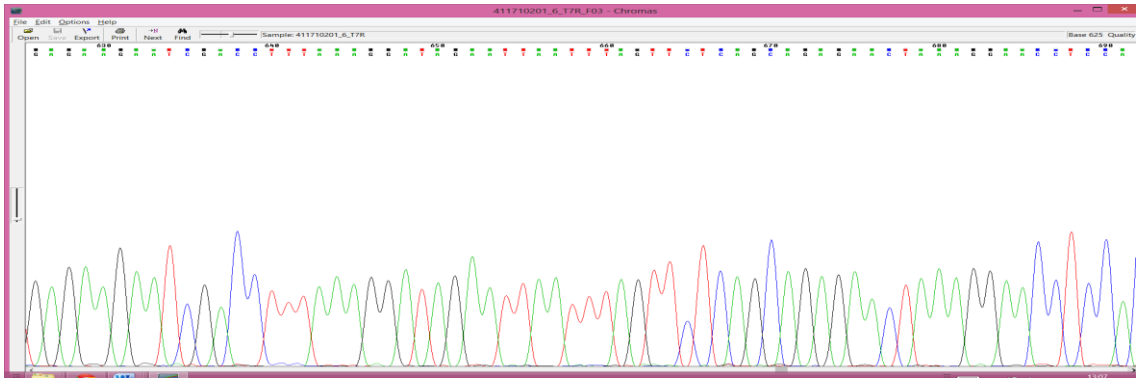
Appendix V. Clustal Sequence Alignment of Recombinant *DHFRL1* (LR clone) showing that the HIS Tag has Been Incorporated after Gateway Cloning

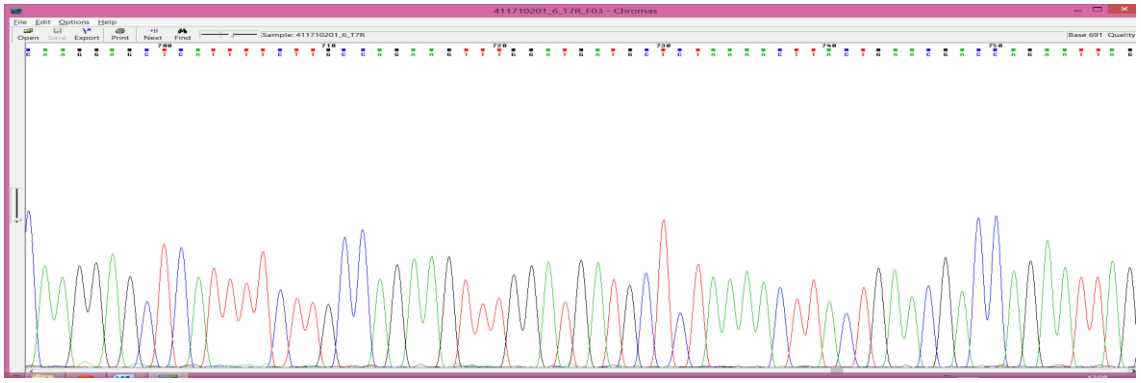

411710201_6_T7R_F03 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411710201_6_T7R

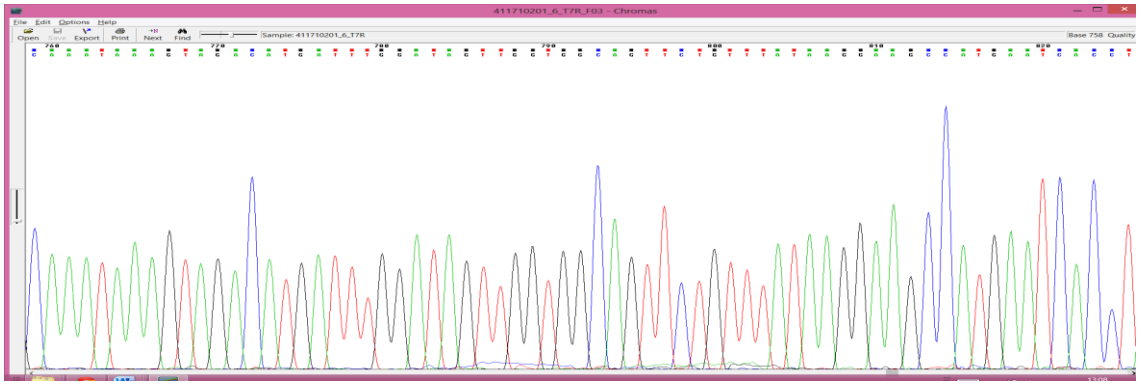

411710201_6_T7R_F03 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411710201_6_T7R
Base 426 Quality

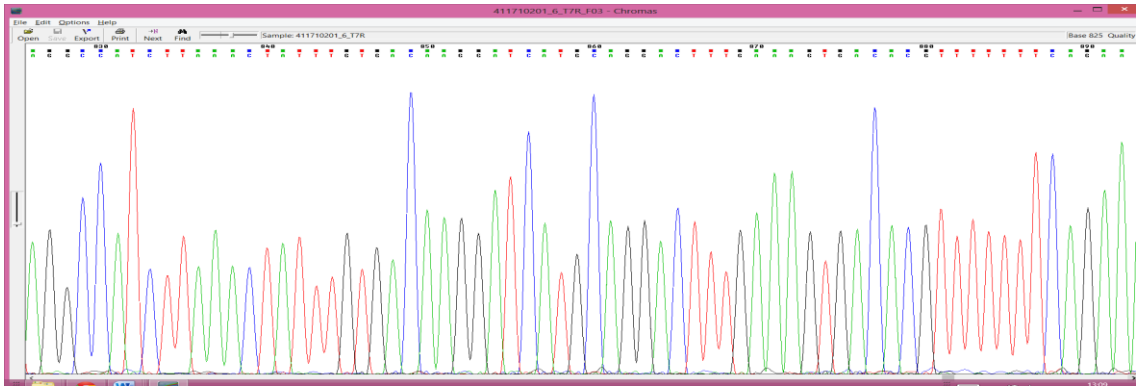

411710201_6_T7R_F03 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411710201_6_T7R
Base 493 Quality

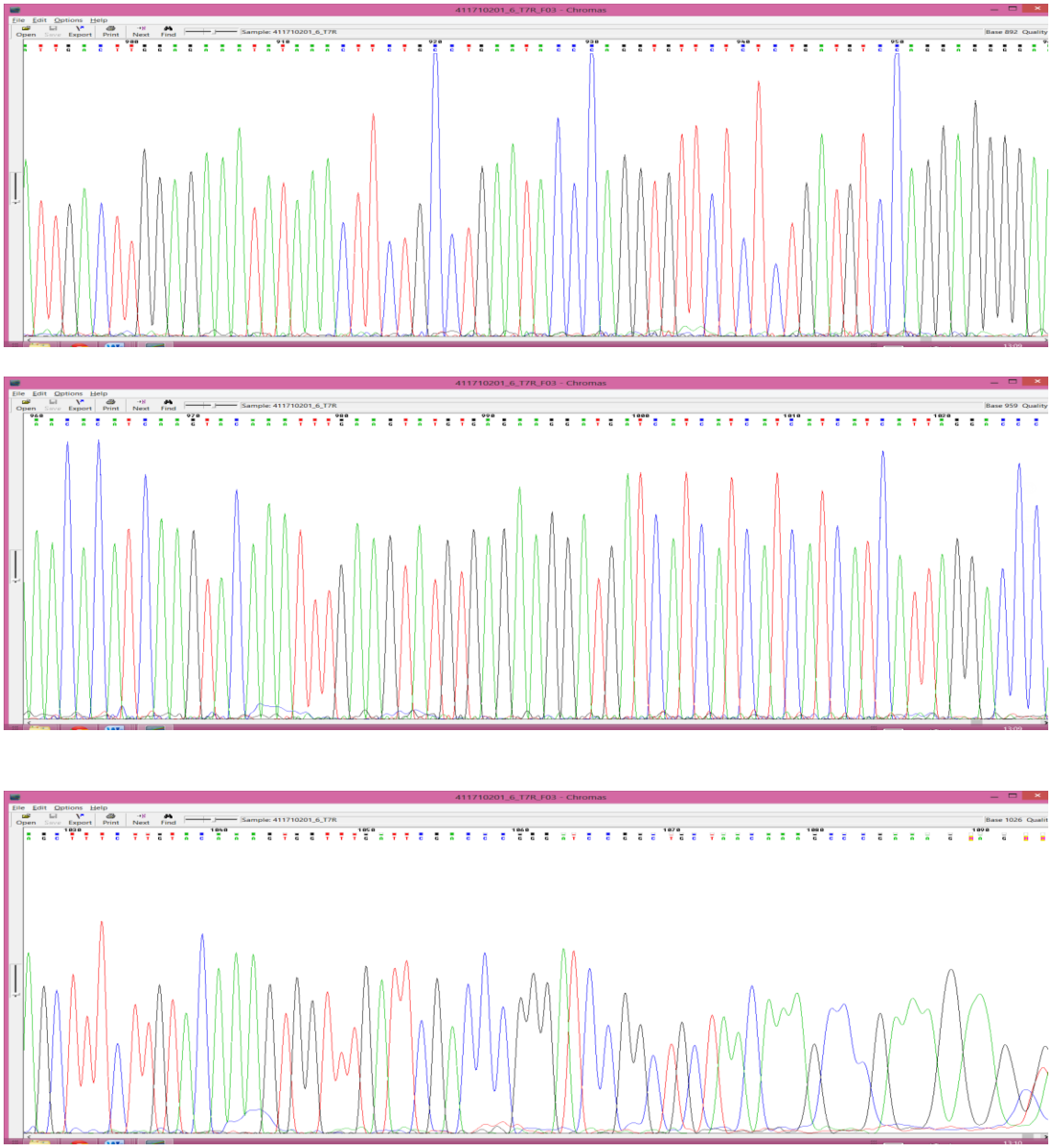

411710201_6_T7R_F03 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411710201_6_T7R
Base 559 Quality


411710201_6_T7R_F03 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411710201_6_T7R
Base 625 Quality


411710201_6_T7R_F03 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411710201_6_T7R
Base 691 Quality


411710201_6_T7R_F03 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411710201_6_T7R
Base 758 Quality


411710201_6_T7R_F03 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 411710201_6_T7R
Base 825 Quality

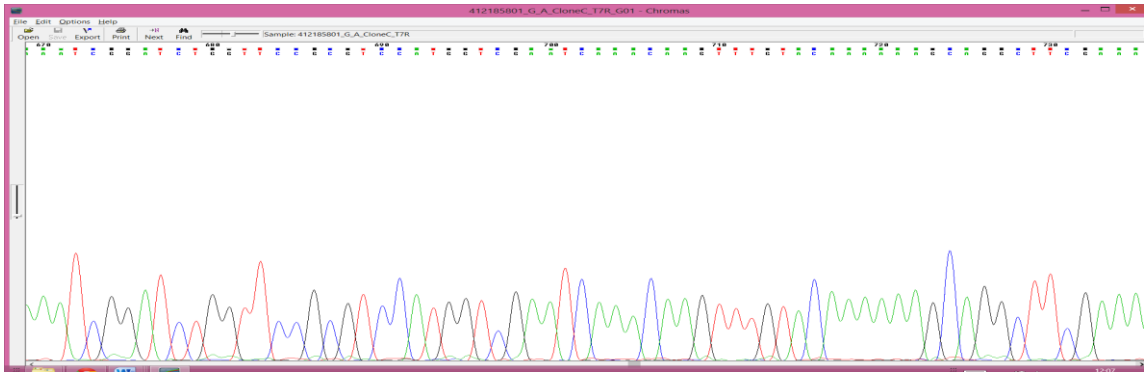


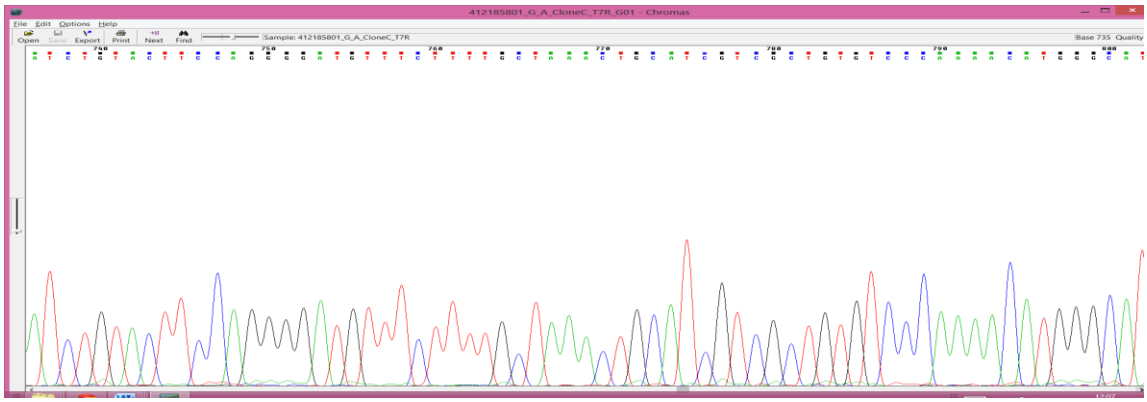
Appendix W. Chromatogram Sequence of Recombinant *DHFRL1* (LR clone) showing that the HIS Tag has Been Incorporated after Gateway Cloning

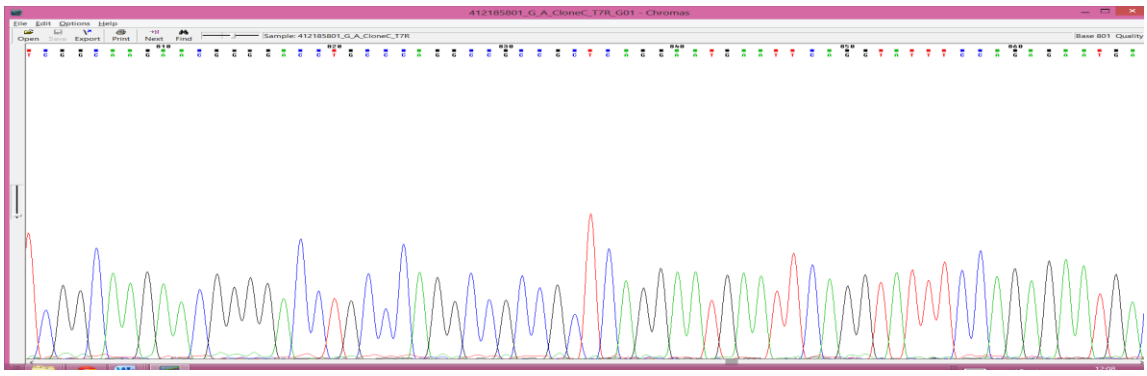
```
DHFRL1_HIS_TAG        ------------------------------------------------------------
DHFRL1_G-A_Clone      AATTAGTTTGTTTTAAAAAACGTATTGAAGCTATCCCACAAATTGATAAGTACTTGAAAT 600

DHFRL1_HIS_TAG        ------------------------------------------------------------
DHFRL1_G-A_Clone      CCAGCAAGTATATAGCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACC 660

DHFRL1_HIS_TAG        ---------------------------------------------ACAAGTTTGTACAAAA 16
DHFRL1_G-A_Clone      ATCCTCCAAAATCGGATCTGGTTCCGCGTCCATGGTCGAATCAAACAAGTTTGTACAAAA 720
                                                                    ****************

DHFRL1_HIS_TAG        AAGCAGGCTTCGAAAATCTGTACTTCCAGGGGATGTTTCTTTTGCTAAACTGCATCGTCG 76
DHFRL1_G-A_Clone      AAGCAGGCTTCGAAAATCTGTACTTCCAGGGGATGTTTCTTTTGCTAAACTGCATCGTCG 780
                      ************************************************************

DHFRL1_HIS_TAG        CTGTGTCCCAAAACATGGGCATCGGCAAGAACGGGGACCTGCCCAGGCCGCCGCTCAGGA 136
DHFRL1_G-A_Clone      CTGTGTCCCAAAACATGGGCATCGGCAAGAACGGGGACCTGCCCAGGCCGCCGCTCAGGA 840
                      ************************************************************

DHFRL1_HIS_TAG        ATGAATTCAGGTATTTCCAGAGAATGACCACAACTTCTTCAGTAGAGGGTAAACAGAATC 196
DHFRL1_G-A_Clone      ATGAATTCAGGTATTTCCAGAGAATGACCACAACTTCTTCAGTAGAGGGTAAACAGAATC 900
                      ************************************************************

DHFRL1_HIS_TAG        TGGTGATTATGGGTAGGAAGACCTGGTTCTCCATTCCTGAGAAGAATCGACCTTTAAAGG 256
DHFRL1_G-A_Clone      TGGTGATTATGGGTAGGAAGACCTGGTTCTCCATTCCTGAGAAGAATCGACCTTTAAAGG 960
                      ************************************************************

DHFRL1_HIS_TAG        ATAGAATTAATTTAGTTCTCAGCAGAGAACTCAAGGAACCTCCACAAGGAGCTCATTTTC 316
DHFRL1_G-A_Clone      ATAGAATTAATTTAGTTCTCAGCAGAGAACTAAAGGAACCTCCACAAGGAGCTCATTTTC 1020
                      ********************************.***************************

DHFRL1_HIS_TAG        TTGCCAGAAGTTTGGATGATGCCTTAAAACTTACTGAACGACCAGAATTAGCAAATAAAG 376
```

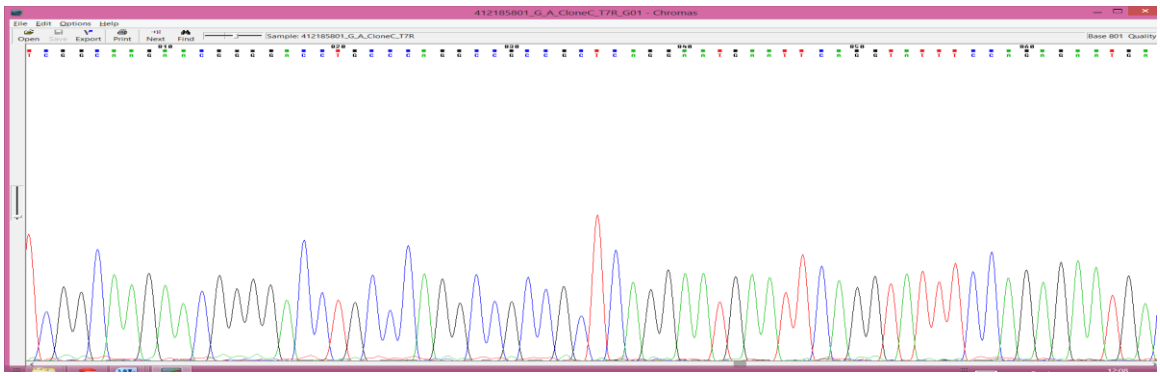
```
DHFRL1_G-A_Clone      TTGCCAGAAGTTTGGATGATGCTCTAAAACTTACTGAACGACCAGAATTAGCAAATAAAG 1080
                      **********************  ***********************************

DHFRL1_HIS_TAG        TAGACATGATTTGGATAGTTGGTGGCAGTTCTGTTTATAAGGAAGCCATGAATCACCTAG 436
DHFRL1_G-A_Clone      TAGACATGATTTGGATAGTTGGTGGCAGTTCTGTTTATAAGGAAGCCATGAATCACCTAG 1140
                      ************************************************************

DHFRL1_HIS_TAG        GCCATCTTAAACTATTTGTGACAAGGATCATGCAGGACTTTGAAAGTGACACGTTTTTTT 496
DHFRL1_G-A_Clone      GCCATCTTAAACTATTTGTGACAAGGATCATGCAGGACTTTGAAAGTGACACGTTTTTTT 1200
                      ************************************************************

DHFRL1_HIS_TAG        CAGAAATTGACTTGGAGAAATATAAACTTCTGCCTGAATACCCAGGTGTTCTCTCTGATG 556
DHFRL1_G-A_Clone      CAGAAATTGACTTGGAGAAATATAAACTTCTGCCTGAATACCCAGGTATTCTCTCTGATG 1260
                      ***********************************************.************

DHFRL1_HIS_TAG        TCCAGGAGGGGAAACACATCAAGTACAAATTTGAAGTATGTGAGAAGGATGATCATCATC 616
DHFRL1_G-A_Clone      TCCAGGAGGGGAAACACATCAAGTACAAATTTGAAGTATGTGAGAAGGATGATCATCATC 1320
                      ************************************************************

DHFRL1_HIS_TAG        ATCATCATCATTAGGACCCAGCTTTCTTGTACAAAGTGGT-------------------- 656
DHFRL1_G-A_Clone      ATCATCATCATTAGGACCCAGCTTTCTTGTACAAAGTGGTTTGATTCGACCCGGGATCCG 1380
                      ****************************************

DHFRL1_HIS_TAG        ----------------------------------
DHFRL1_G-A_Clone      GCTGCTAACAAAGCNCGAANGNNGNNNNNNNNNNN 1415
```

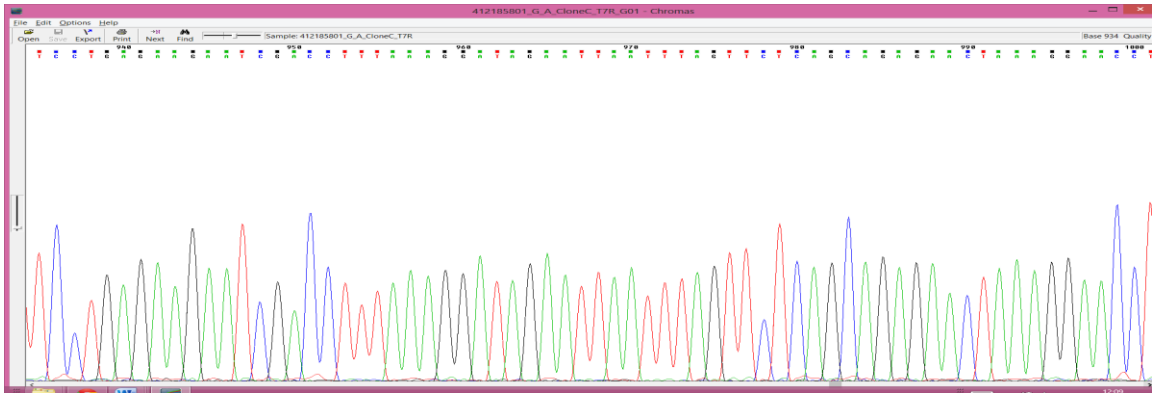
Appendix X. Clustal Sequence Alignment of Recombinant *DHFRL1* clone showing that the G->A rs17855824 (Valine->Isoleucine) SNP has Been Incorporated by SDM

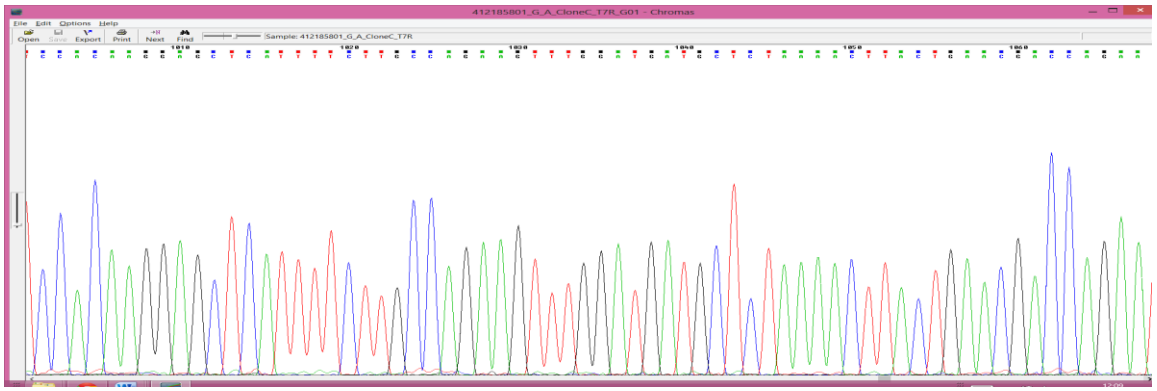

412185801_G_A_CloneC_T7R_G01 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 412185801_G_A_CloneC_T7R

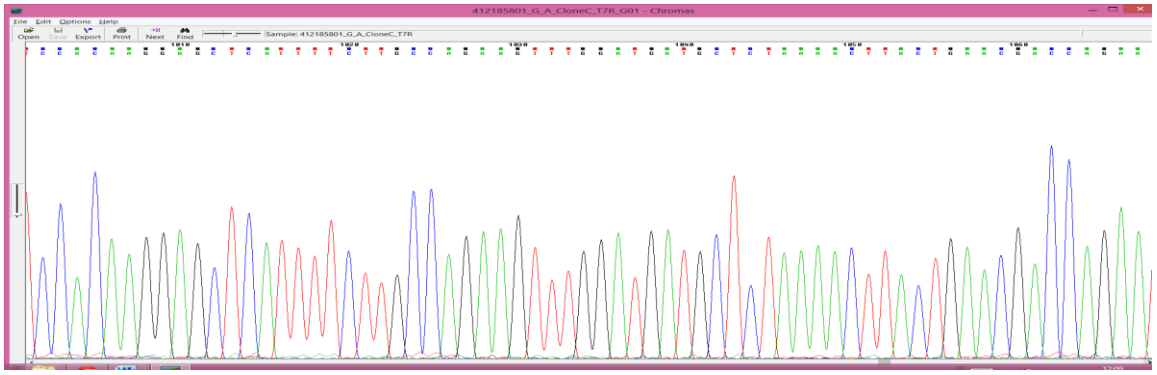

412185801_G_A_CloneC_T7R_G01 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 412185801_G_A_CloneC_T7R
Base 735 Quality

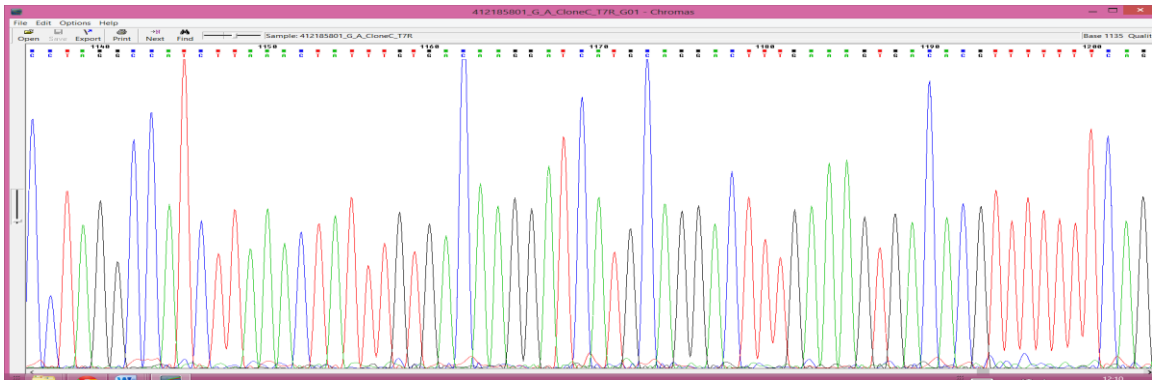

412185801_G_A_CloneC_T7R_G01 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 412185801_G_A_CloneC_T7R
Base 801 Quality

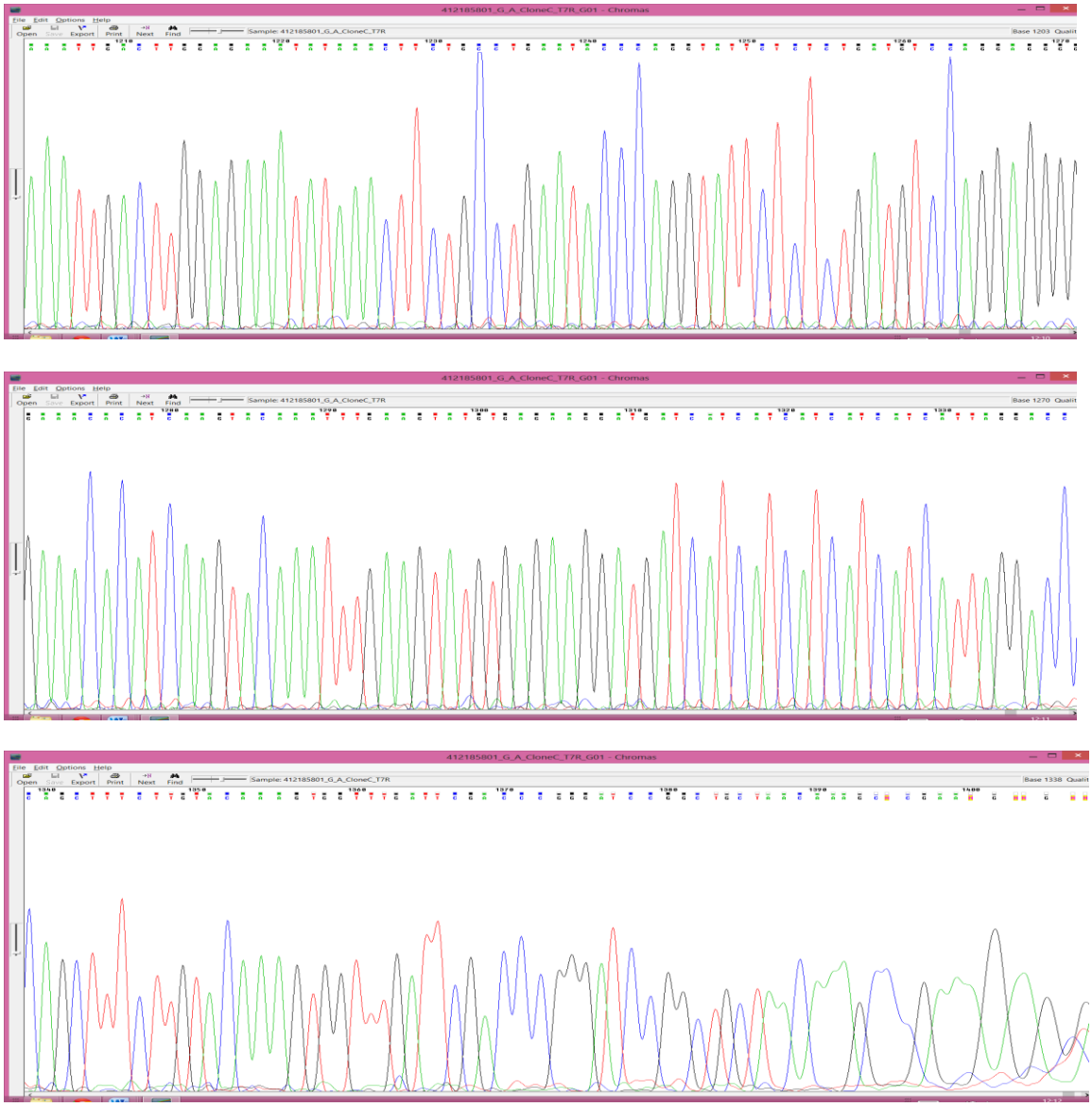

412185801_G_A_CloneC_T7R_G01 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 412185801_G_A_CloneC_T7R
Base 801 Quality


412185801_G_A_CloneC_T7R_G01 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 412185801_G_A_CloneC_T7R
Base 934 Quality


412185801_G_A_CloneC_T7R_G01 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 412185801_G_A_CloneC_T7R


412185801_G_A_CloneC_T7R_G01 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 412185801_G_A_CloneC_T7R


412185801_G_A_CloneC_T7R_G01 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 412185801_G_A_CloneC_T7R
Base 1135 Quality

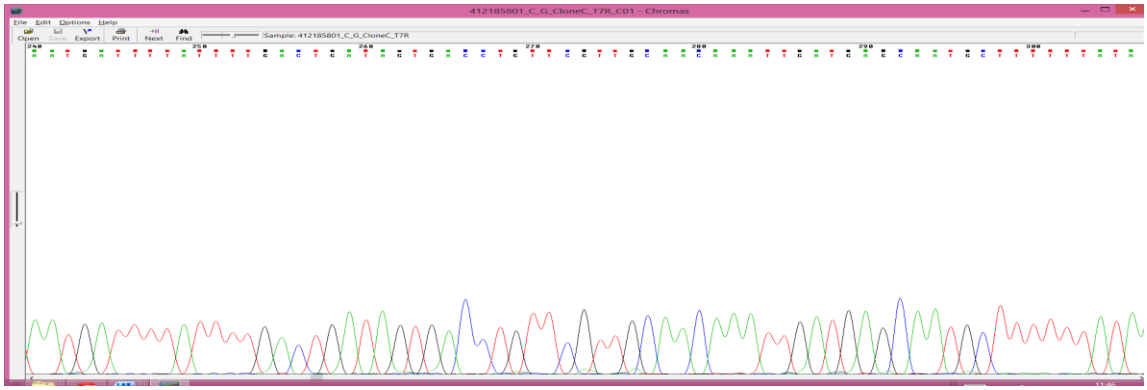


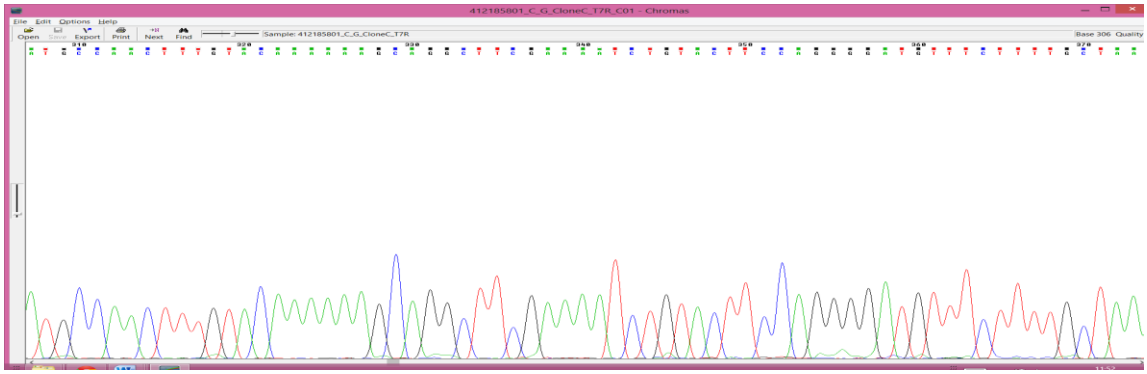
Appendix Y. Chromatogram Sequence of Recombinant *DHFRL1* showing that the G->A rs17855824 (Valine->Isoleucine) SNP has Been Incorporated by SDM

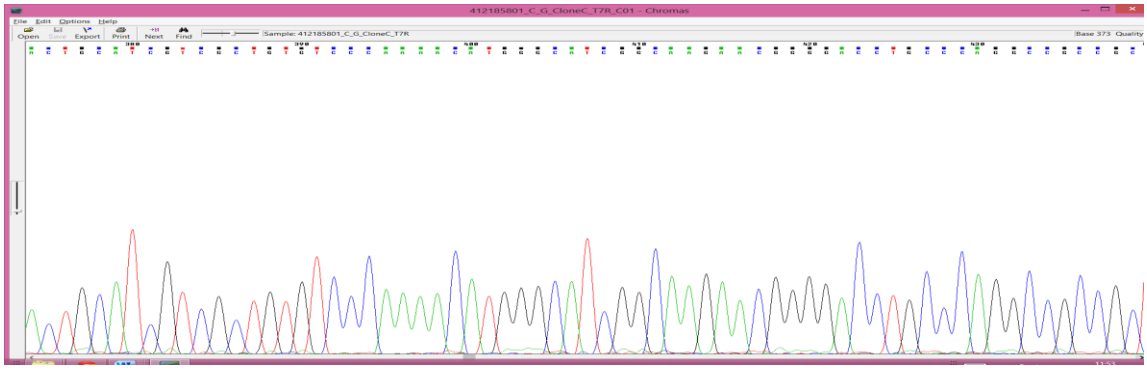
```
DHFRL1_HIS_TAG        ------------------------------------------------------------
DHFRL1_C-G_Clone      GAGCCTTTCGTTTTATTTGATGCCTGGCAGTTCCCTACTCTCGCGTTAACGCTAGCATGG 180

DHFRL1_HIS_TAG        ------------------------------------------------------------
DHFRL1_C-G_Clone      ATGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTCTTAAGCTCGGGCCCCAAATA 240

DHFRL1_HIS_TAG        ------------------------------------------------------------
DHFRL1_C-G_Clone      ATGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATGAGCAATGCTTT 300

DHFRL1_HIS_TAG        ---------ACAAGTTTGTACAAAAAAGCAGGCTTCGAAAATCTGTACTTCCAGGGGATG 51
DHFRL1_C-G_Clone      TTTATAATGCCAACTTTGTACAAAAAAGCAGGCTTCGAAAATCTGTACTTCCAGGGGATG 360
                             .*** ******************************************

DHFRL1_HIS_TAG        TTTCTTTTGCTAAACTGCATCGTCGCTGTGTCCCAAAACATGGGCATCGGCAAGAACGGG 111
DHFRL1_C-G_Clone      TTTCTTTTGCTAAACTGCATCGTCGCTGTGTCCCAAAACATGGGCATCGGCAAGAACGGG 420
                      ************************************************************

DHFRL1_HIS_TAG        GACCTGCCCAGGCCGCCGCTCAGGAATGAATTCAGGTATTTCCAGAGAATGACCACAACT 171
DHFRL1_C-G_Clone      GACCTGCCCAGGCCGCCGCTCAGGAATGAATTCAGGTATTTCCAGAGAATGACCACAACT 480
                      ************************************************************

DHFRL1_HIS_TAG        TCTTCAGTAGAGGGTAAACAGAATCTGGTGATTATGGGTAGGAAGACCTGGTTCTCCATT 231
DHFRL1_C-G_Clone      TCTTCAGTAGAGGGTAAACAGAATCTGGTGATTATGGGTAGGAAGACCTGGTTCTCCATT 540
                      ************************************************************

DHFRL1_HIS_TAG        CCTGAGAAGAATCGACCTTTAAAGGATAGAATTAATTTAGTTCTCAGCAGAGAACTCAAG 291
DHFRL1_C-G_Clone      CCTGAGAAGAATCGACCTTTAAAGGATAGAATTAATTTAGTTCTCAGCAGAGAACTCAAG 600
                      ************************************************************

DHFRL1_HIS_TAG        GAACCTCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGATGCCTTAAAACTTACT 351
```

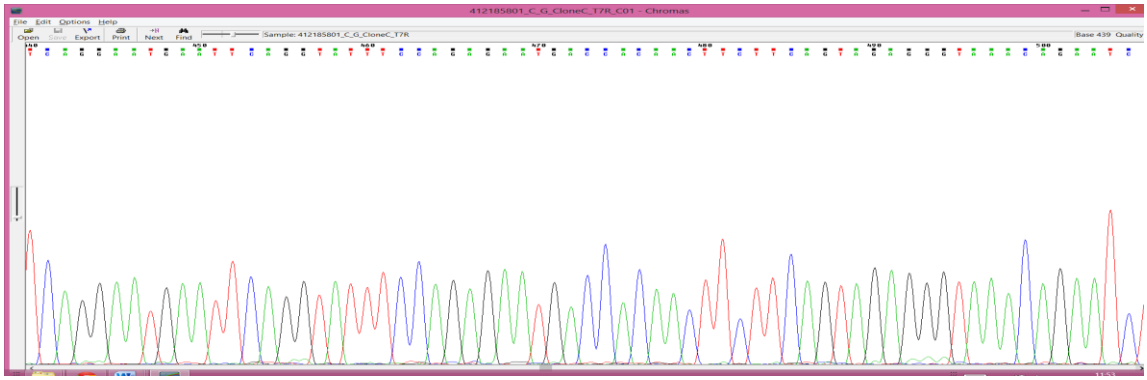
```
DHFRL1_C-G_Clone      GAAGCTCCACAAGGAGCTCATTTTCTTGCCAGAAGTTTGGATGATGCTCTAAAACTTACT 660
                      *** ****************************************  ***********

DHFRL1_HIS_TAG        GAACGACCAGAATTAGCAAATAAAGTAGACATGATTTGGATAGTTGGTGGCAGTTCTGTT 411
DHFRL1_C-G_Clone      GAACGACCAGAATTAGCAAATAAAGTAGACATGATTTGGATAGTTGGTGGCAGTTCTGTT 720
                      ************************************************************

DHFRL1_HIS_TAG        TATAAGGAAGCCATGAATCACCTAGGCCATCTTAAACTATTTGTGACAAGGATCATGCAG 471
DHFRL1_C-G_Clone      TATAAGGAAGCCATGAATCACCTAGGCCATCTTAAACTATTTGTGACAAGGATCATGCAG 780
                      ************************************************************

DHFRL1_HIS_TAG        GACTTTGAAAGTGACACGTTTTTTTCAGAAATTGACTTGGAGAAATATAAACTTCTGCCT 531
DHFRL1_C-G_Clone      GACTTTGAAAGTGACACGTTTTTTTCAGAAATTGACTTGGAGAAATATAAACTTCTGCCT 840
                      ************************************************************

DHFRL1_HIS_TAG        GAATACCCAGGTGTTCTCTCTGATGTCCAGGAGGGGAAACACATCAAGTACAAATTTGAA 591
DHFRL1_C-G_Clone      GAATACCCAGGTGTTCTCTCTGATGTCCAGGAGGGGAAACACATCAAGTACAAATTTGAA 900
                      ************************************************************

DHFRL1_HIS_TAG        GTATGTGAGAAGGATGATCATCATCATCATCATCATTAGGACCCAGCTTTCTTGTACAAA 651
DHFRL1_C-G_Clone      GTATGTGAGAAGGATGATCATCATCATCATCATCATTAGGACCCAGCTTTCTTGTACAAA 960
                      ************************************************************

DHFRL1_HIS_TAG        GTGGT---------------------------------------------------- 656
DHFRL1_C-G_Clone      GTGGTTTGATTCGACCCGGGATCCGGCTGCTAACAAAGCCCGNNTNNNNGNNNNNNNNN 1019
                      *****
```

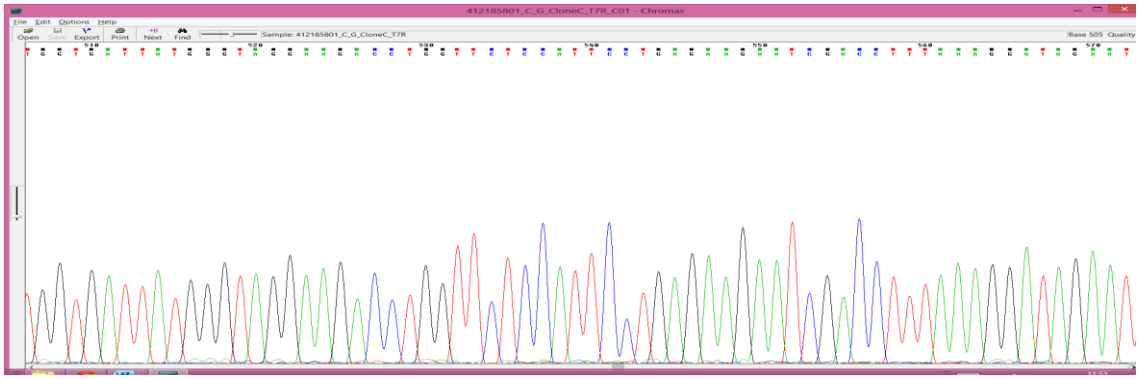
Appendix Z. Clustal Sequence Alignment of Recombinant *DHFRL1* showing that the C->G rs6139170 (Proline->Alanine) SNP has Been Incorporated by SDM

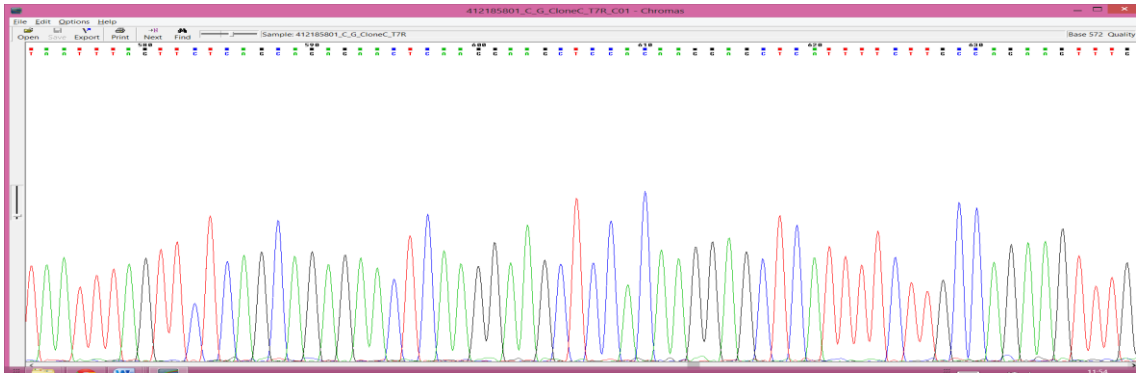

412185801_C_G_CloneC_T7R_C01 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 412185801_C_G_CloneC_T7R

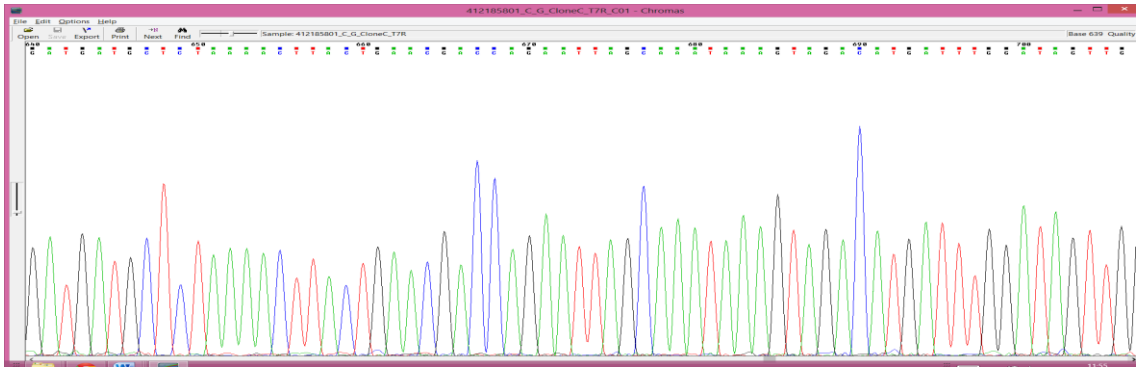

412185801_C_G_CloneC_T7R_C01 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 412185801_C_G_CloneC_T7R
Base 306 Quality

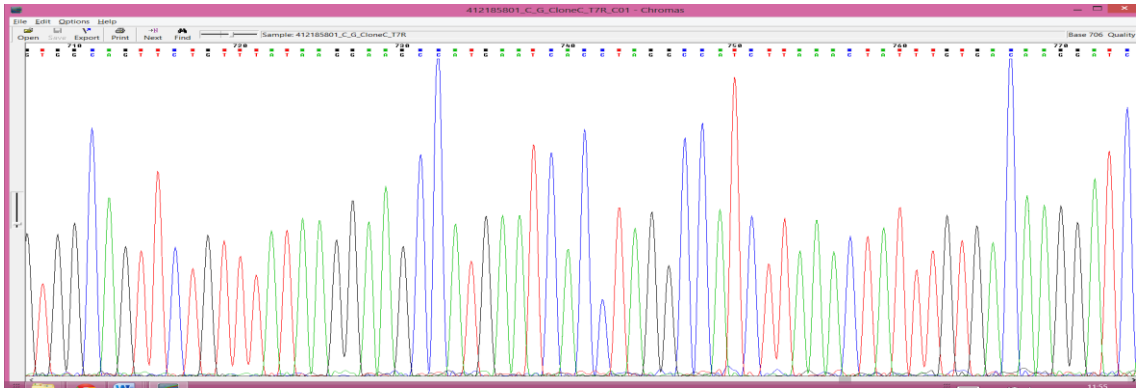

412185801_C_G_CloneC_T7R_C01 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 412185801_C_G_CloneC_T7R
Base 373 Quality

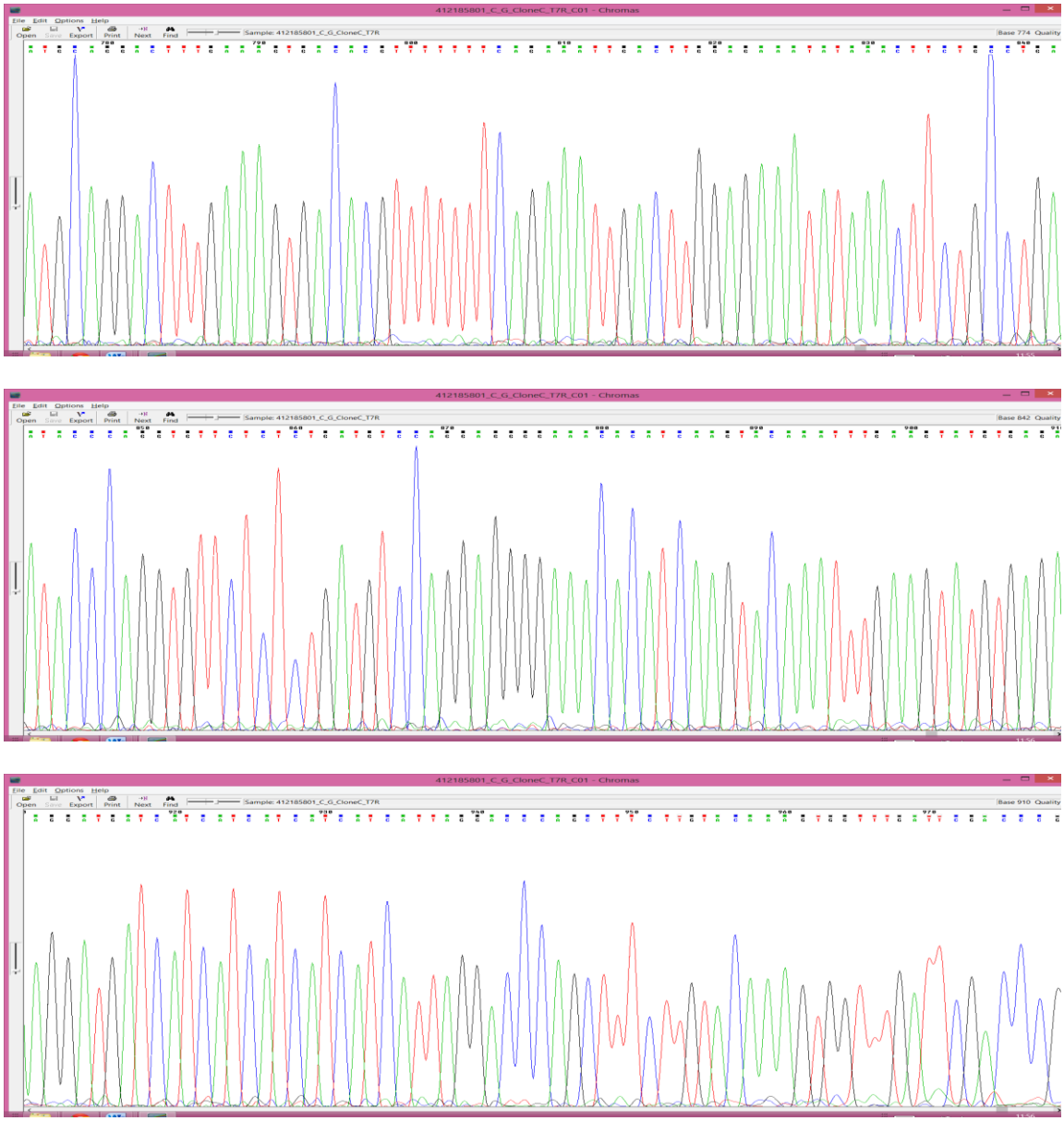

412185801_C_G_CloneC_T7R_C01 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 412185801_C_G_CloneC_T7R
Base 439 Quality


412185801_C_G_CloneC_T7R_C01 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 412185801_C_G_CloneC_T7R
Base 505 Quality


412185801_C_G_CloneC_T7R_C01 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 412185801_C_G_CloneC_T7R
Base 572 Quality


412185801_C_G_CloneC_T7R_C01 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 412185801_C_G_CloneC_T7R
Base 639 Quality


412185801_C_G_CloneC_T7R_C01 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 412185801_C_G_CloneC_T7R
Base 706 Quality

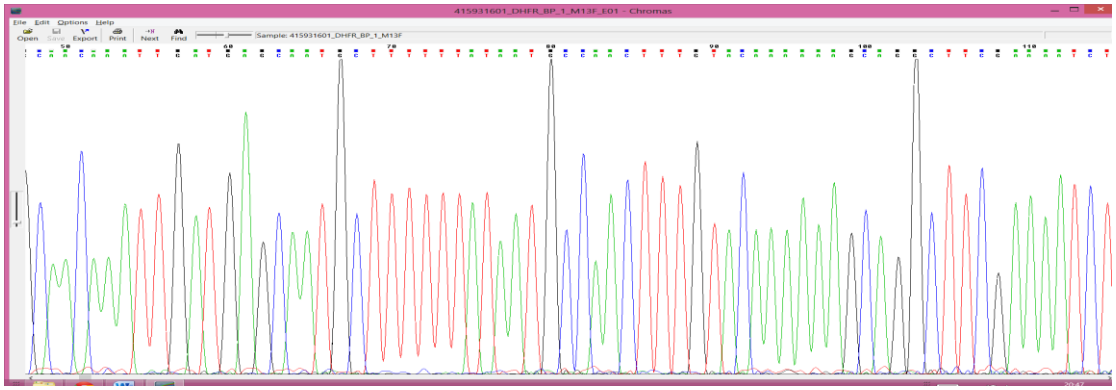


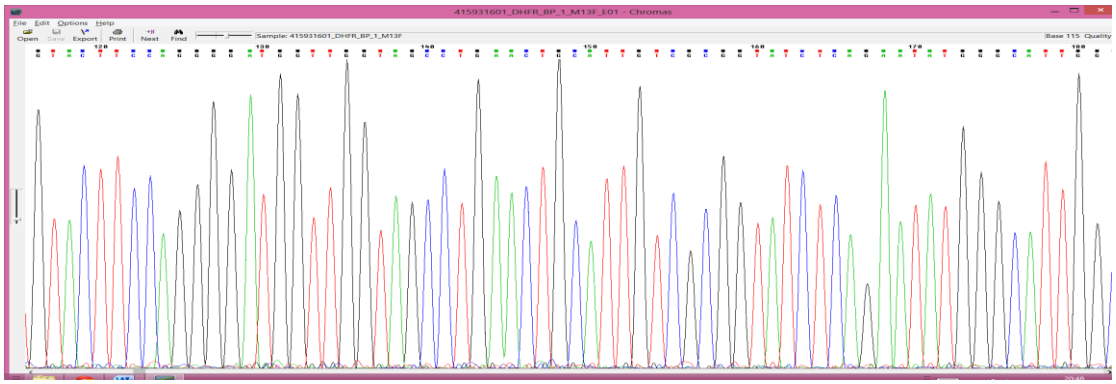
Appendix AA. Chromatogram Sequence Alignment of Recombinant *DHFRL1* showing that the C->G rs6139170 (Proline->Alanine) SNP has Been Incorporated by SDM

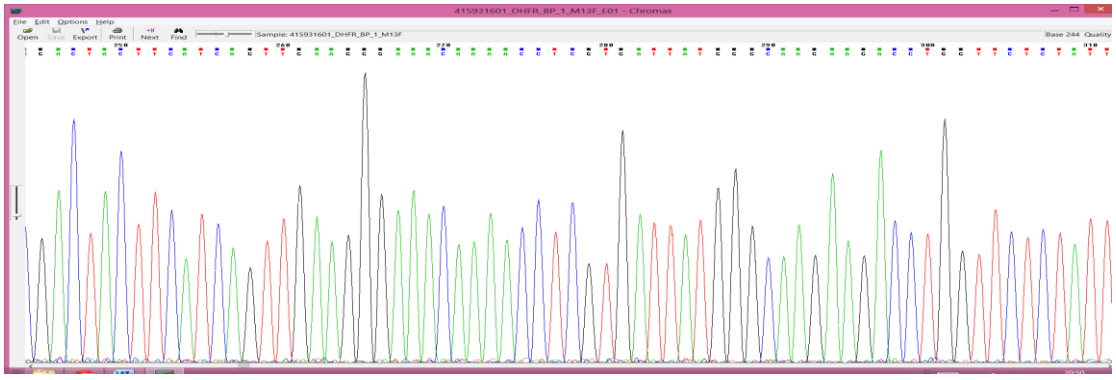
```
DHFR            ------------------------------------------------------------
DHFR_BP         NNNNNNNNNNNNNGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATG 60


DHFR            ----------------GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAAATCTGTACT 44
DHFR_BP         AGCAATGCTTTTTTATAATGCCAACTTTGTACAAAAAAGCAGGCTTCGAAAATCTGTACT 120
                             .. *.*** **********************************


DHFR            TCCAGGGGATGGTTGGTAGCCTGAACTGCATTGTCGCGGTATCTCAGAATATGGGCATTG 104
DHFR_BP         TCCAGGGGATGGTTGGTAGCCTGAACTGCATTGTCGCGGTATCTCAGAATATGGGCATTG 180
                ************************************************************


DHFR            GCAAAAACGGCGACTTACCGTGGCCGCCGTTGCGTAATGAATTTCGCTATTTTCAGCGTA 164
DHFR_BP         GCAAAAACGGCGACTTACCGTGGCCGCCGTTGCGTAATGAATTTCGCTATTTTCAGCGTA 240
                ************************************************************


DHFR            TGACGACTACTTCATCAGTTGAAGGGAAACAAAACCTCGTGATTATGGGCAAGAAGACCT 224
DHFR_BP         TGACGACTACTTCATCAGTTGAAGGGAAACAAAACCTCGTGATTATGGGCAAGAAGACCT 300
                ************************************************************


DHFR            GGTTCTCTATTCCCGAGAAAAATCGTCCGTTAAAGGGCCGTATTAACCTTGTGTTATCTC 284
DHFR_BP         GGTTCTCTATTCCCGAGAAAAATCGTCCGTTAAAGGGCCGTATTAACCTTGTGTTATCTC 360
                ************************************************************


DHFR            GCGAGCTGAAGGAGCCGCCGCAGGGGGCCCATTTCTTATCCCGTTCCCTGGATGACGCAC 344
DHFR_BP         GCGAGCTGAAGGAGCCGCCGCAGGGGGCCCATTTCTTATCCCGTTCCCTGGATGACGCAC 420
                ************************************************************


DHFR            TGAAGCTGACCGAACAACCTGAACTGGCTAATAAAGTCGACATGGTGTGGATTGTAGGCG 404
DHFR_BP         TGAAGCTGACCGAACAACCTGAACTGGCTAATAAAGTCGACATGGTGTGGATTGTAGGCG 480
                ************************************************************


DHFR            GTAGCTCGGTTTATAAAGAAGCGATGAACCACCCCGGGCACTTAAAACTGTTCGTAACTC 464
```

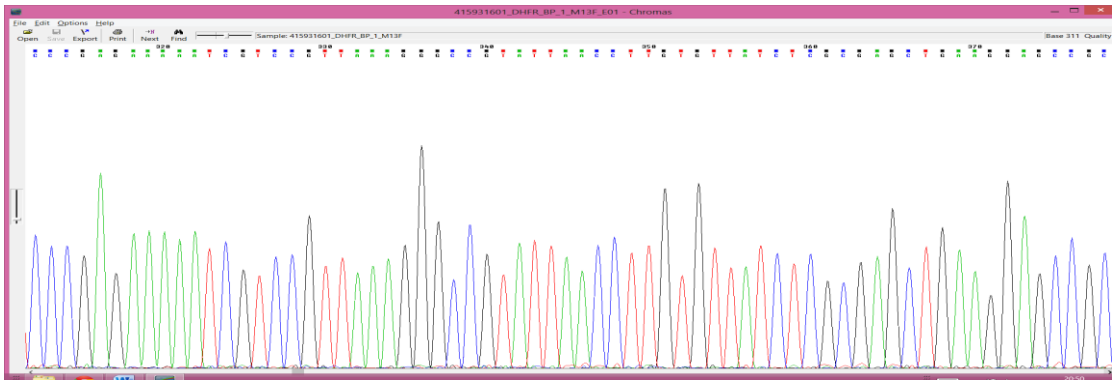
```
DHFR_BP         GTAGCTCGGTTTATAAAGAAGCGATGAACCACCCCGGGCACTTAAAACTGTTCGTAACTC 540
                ************************************************************

DHFR            GCATCATGCAAGACTTTGAAAGTGATACGTTTTTCCCCGAAATCGACCTGGAAAAGTATA 524
DHFR_BP         GCATCATGCAAGACTTTGAAAGTGATACGTTTTTCCCCGAAATCGACCTGGAAAAGTATA 600
                ************************************************************

DHFR            AACTGCTGCCGGAATATCCGGGGGTTTTGTCCGATGTTCAGGAGGAAAAAGGTATCAAAT 584
DHFR_BP         AACTGCTGCCGGAATATCCGGGGGTTTTGTCCGATGTTCAGGAGGAAAAAGGTATCAAAT 660
                ************************************************************

DHFR            ACAAGTTTGAGGTGTATGAAAAGAACGATTAAGACCCAGCTTTCTTGTACAAAGTGG--- 641
DHFR_BP         ACAAGTTTGAGGTGTATGAAAAGAACGATTAAGACCCAGCTTTCTTGTACAAAGTTGGCA 720
                ********************************************************** *

DHFR            ---TGGGG----------------------------------------------------- 646
DHFR_BP         TTATAAGAAAGCATTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATA 780
                   *..*.

DHFR            ------------------------------------------------------------
DHFR_BP         AAATCATTATTTGCCATCCAGCTGATATCCCCTATAGTGAGTCGTATTACATGGTCATAG 840
```

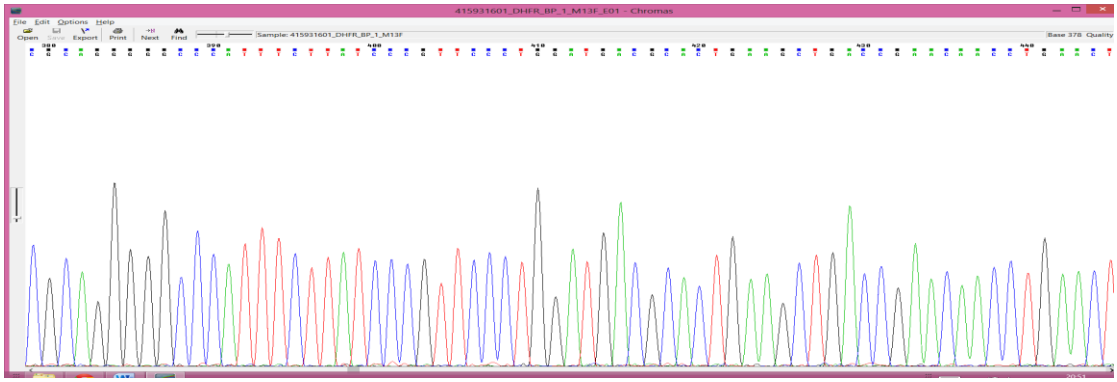
Appendix AB. Clustal Sequence Alignment of the Rare Codon Optimised *DHFR* BP Clone

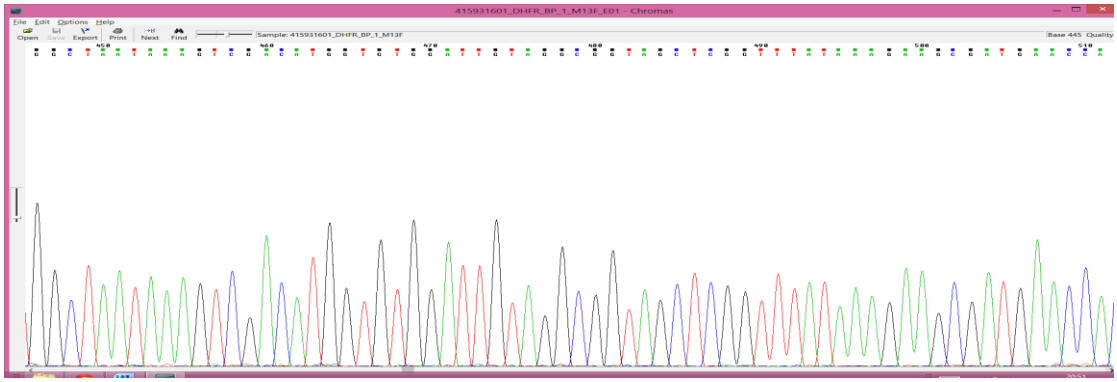

415931601_DHFR_BP_1_M13F_E01 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 415931601_DHFR_BP_1_M13F

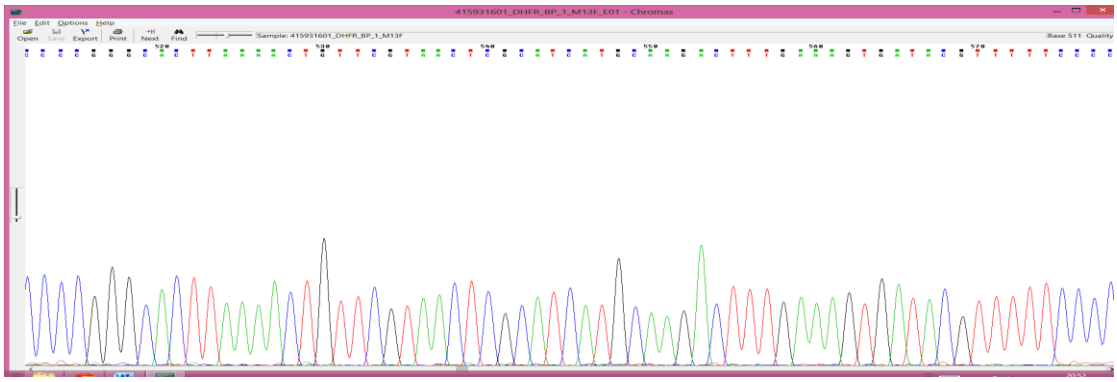

415931601_DHFR_BP_1_M13F_E01 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 415931601_DHFR_BP_1_M13F
Base 115 Quality

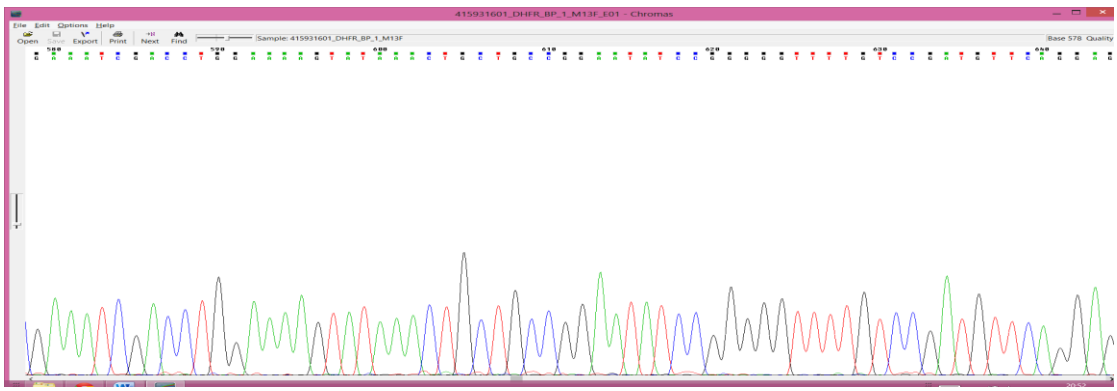

415931601_DHFR_BP_1_M13F_E01 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 415931601_DHFR_BP_1_M13F
Base 244 Quality

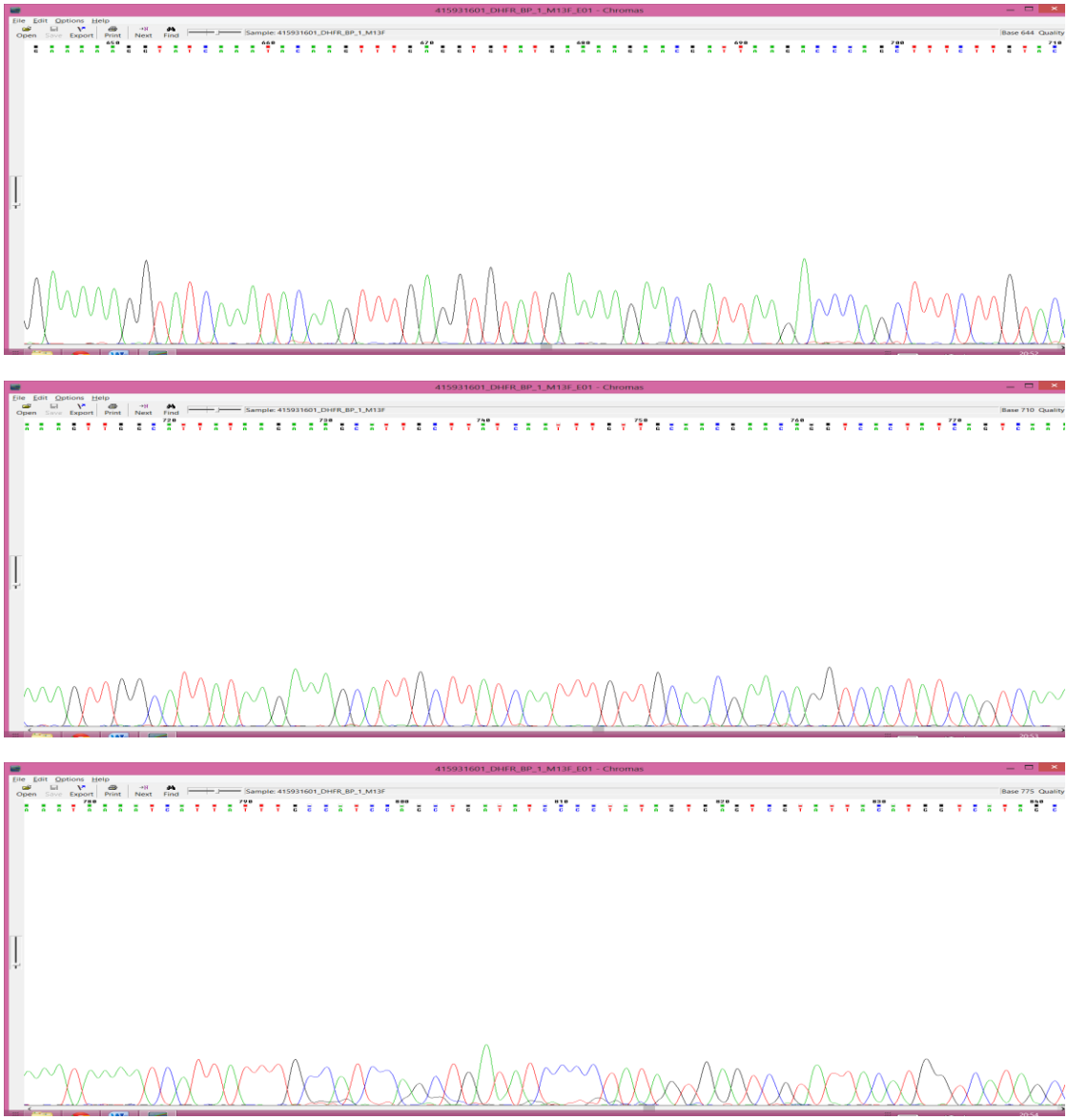

415931601_DHFR_BP_1_M13F_E01 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 415931601_DHFR_BP_1_M13F
Base 311 Quality


415931601_DHFR_BP_1_M13F_E01 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 415931601_DHFR_BP_1_M13F
Base 378 Quality


415931601_DHFR_BP_1_M13F_E01 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 415931601_DHFR_BP_1_M13F
Base 445 Quality


415931601_DHFR_BP_1_M13F_E01 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 415931601_DHFR_BP_1_M13F
Base 511 Quality


415931601_DHFR_BP_1_M13F_E01 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 415931601_DHFR_BP_1_M13F
Base 578 Quality

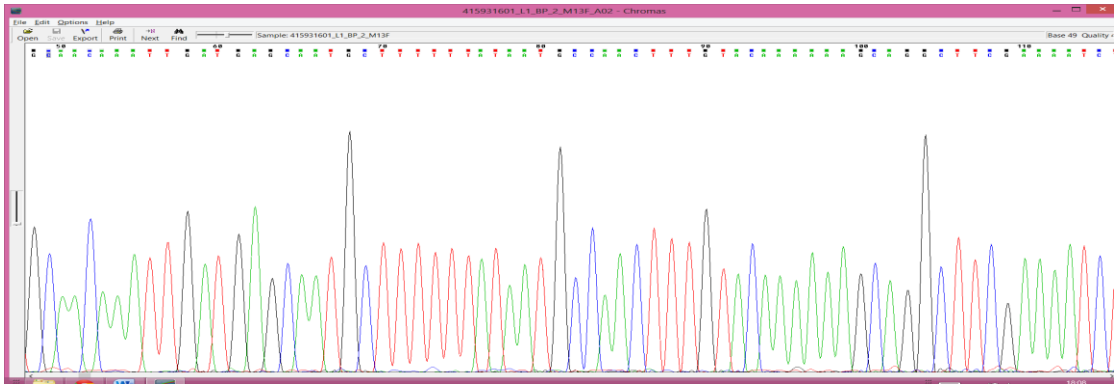


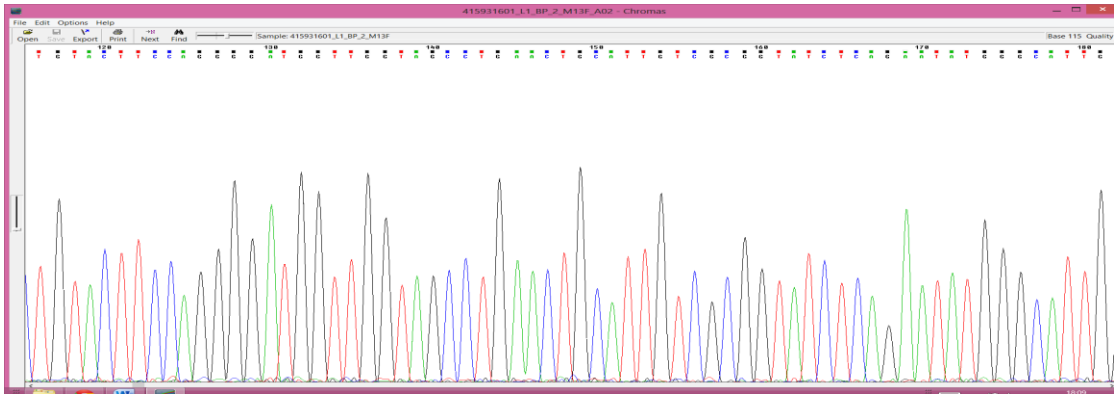
Appendix AC. Chromatogram Sequence Alignment of the Rare Codon Optimised *DHFR* BP Clone

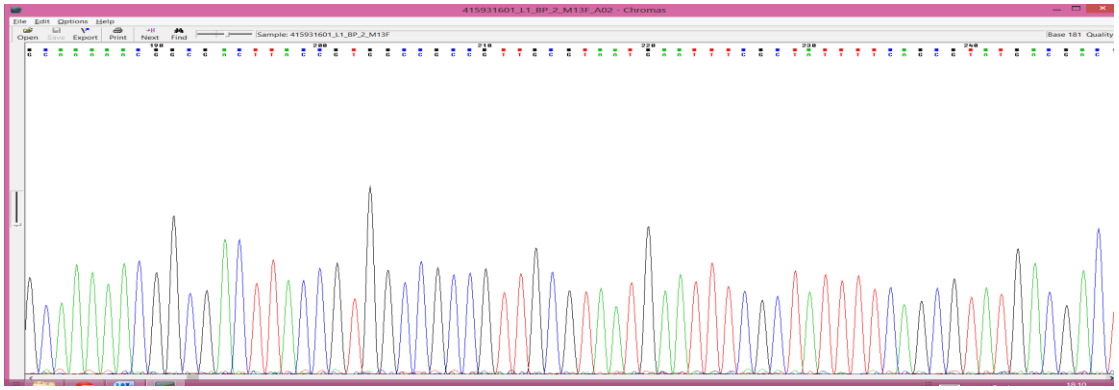
```
DHFRL1          ------------------------------------------------------------
DHFRL1_BP       NNNNNNNNNNNNNNTGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGAT 60


DHFRL1          -----------------GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAAATCTGTAC 43
DHFRL1_BP       GAGCAATGCTTTTTTATAATGCCAACTTTGTACAAAAAAGCAGGCTTCGAAAATCTGTAC 120
                              .. *.*** **********************************


DHFRL1          TTCCAGGGGATGTTTCTGCTGTTAAATTGCATCGTTGCCGTCTCCCAGAACATGGGCATT 103
DHFRL1_BP       TTCCAGGGGATGTTTCTGCTGTTAAATTGCATCGTTGCCGTCTCCCAGAACATGGGCATT 180
                ************************************************************


DHFRL1          GGTAAGAATGGTGACCTGCCACGCCCGCCGCTGCGCAACGAATTTCGTTACTTTCAACGG 163
DHFRL1_BP       GGTAAGAATGGTGACCTGCCACGCCCGCCGCTGCGCAACGAATTTCGTTACTTTCAACGG 240
                ************************************************************


DHFRL1          ATGACAACTACCAGTTCTGTAGAAGGCAAACAGAACTTGGTGATCATGGGCCGTAAAACC 223
DHFRL1_BP       ATGACAACTACCAGTTCTGTAGAAGGCAAACAGAACTTGGTGATCATGGGCCGTAAAACC 300
                ************************************************************


DHFRL1          TGGTTCAGTATCCCGGAAAAGAACCGCCCGTTAAAAGATCGGATTAATCTGGTGTTGAGT 283
DHFRL1_BP       TGGTTCAGTATCCCGGAAAAGAACCGCCCGTTAAAAGATCGGATTAATCTGGTGTTGAGT 360
                ************************************************************


DHFRL1          CGCGAACTTAAGGAACCACCGCAGGGGGCACACTTTCTGGCCCGTTCACTGGATGATGCT 343
DHFRL1_BP       CGCGAACTTAAGGAACCACCGCAGGGGGCACACTTTCTGGCCCGTTCACTGGATGATGCT 420
                ************************************************************


DHFRL1          CTGAAATTAACAGAACGTCCAGAATTAGCGAACAAAGTCGATATGATCTGGATTGTCGGT 403
DHFRL1_BP       CTGAAATTAACAGAACGTCCAGAATTAGCGAACAAAGTCGATATGATCTGGATTGTCGGT 480
                ************************************************************


DHFRL1          GGGAGTAGCGTCTACAAAGAGGCGATGAATCATCTGGGCCACTTGAAGCTCTTTGTCACC 463
```

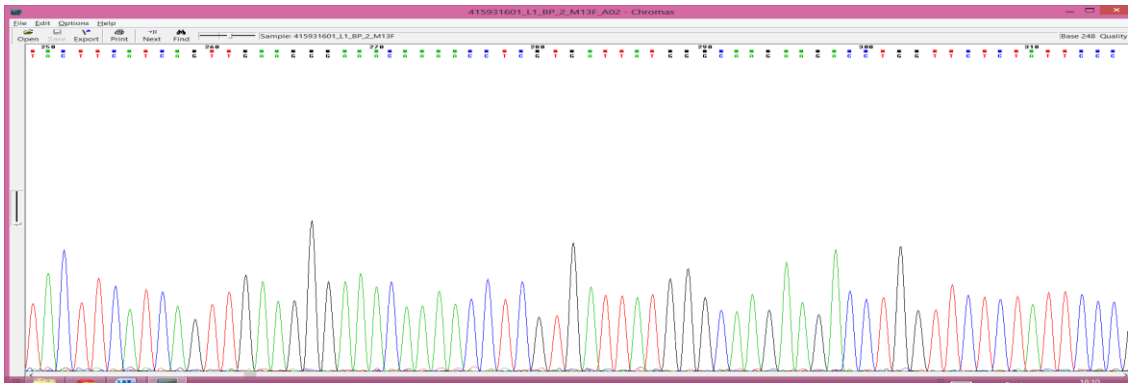
```
DHFRL1_BP       GGGAGTAGCGTCTACAAAGAGGCGATGAATCATCTGGGCCACTTGAAGCTCTTTGTCACC 540
                ************************************************************

DHFRL1          CGGATCATGCAGGATTTTGAGTCAGATACATTCTTTTCTGAAATCGATCTGGAAAAATAT 523
DHFRL1_BP       CGGATCATGCAGGATTTTGAGTCAGATACATTCTTTTCTGAAATCGATCTGGAAAAATAT 600
                ************************************************************

DHFRL1          AAACTTCTGCCGGAATACCCGGGCGTTCTCAGTGACGTCCAGGAAGGTAAGCACATCAAA 583
DHFRL1_BP       AAACTTCTGCCGGAATACCCGGGCGTTCTCAGTGACGTCCAGGAAGGTAAGCACATCAAA 660
                ************************************************************

DHFRL1          TATAAATTTGAGGTGTGCGAAAAAGACGATTAAGACCCAGCTTTCTTGTACAAAGTGG-- 641
DHFRL1_BP       TATAAATTTGAGGTGTGCGAAAAAGACGATTAAGACCCAGCTTTCTTGTACAAAGTTGGC 720
                ************************************************** *

DHFRL1          ----TGGGG---------------------------------------------------- 646
DHFRL1_BP       ATTATAAGAAAGCATTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAAT 780
                    *..*.

DHFRL1          ------------------------------------------------------------
DHFRL1_BP       AAAATCATTATTTGCCATCCAGCTGATATCCCCTATAGTGAGTCGTATTACATGGTCATA 840
```

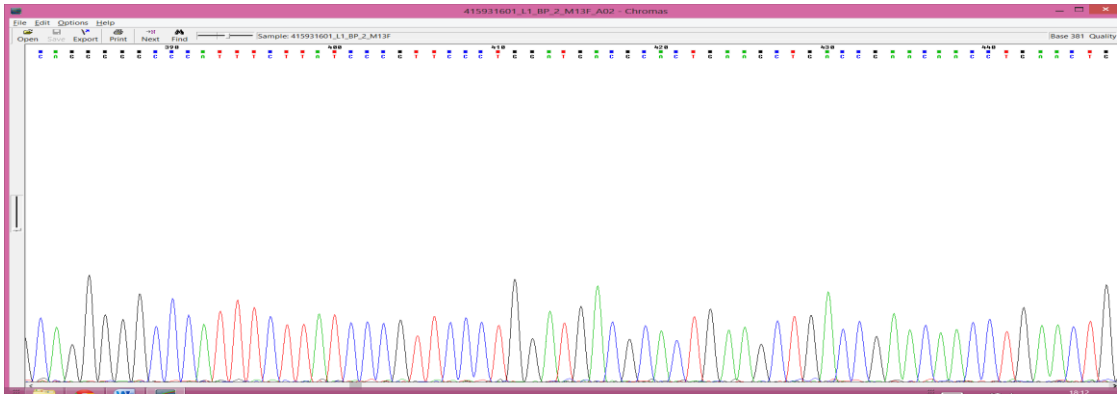
Appendix AD. Clustal Sequence Alignment of the Rare Codon Optimised *DHFRL1* BP Clone

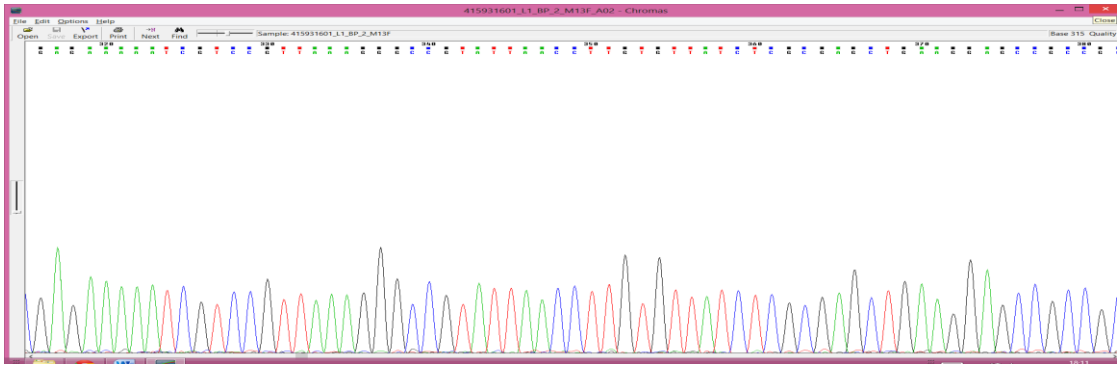

415931601_L1_BP_2_M13F_A02 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 415931601_L1_BP_2_M13F
Base 49 Quality

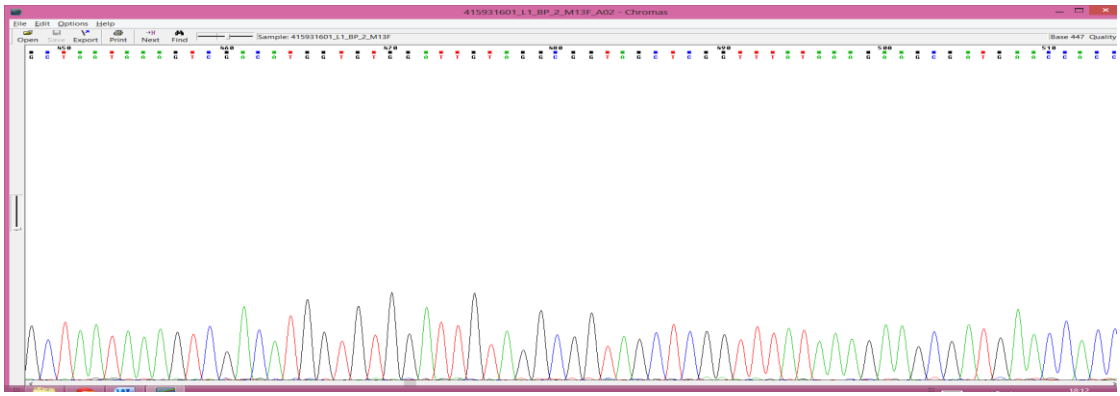

415931601_L1_BP_2_M13F_A02 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 415931601_L1_BP_2_M13F
Base 115 Quality

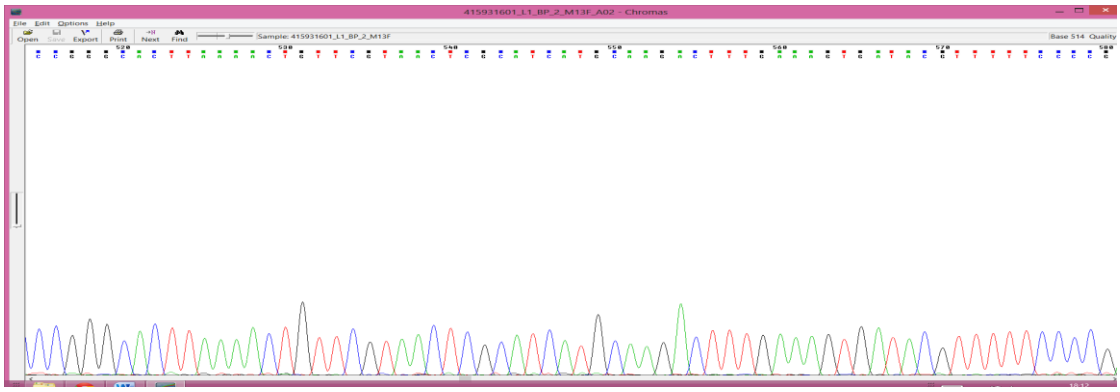

415931601_L1_BP_2_M13F_A02 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 415931601_L1_BP_2_M13F
Base 181 Quality

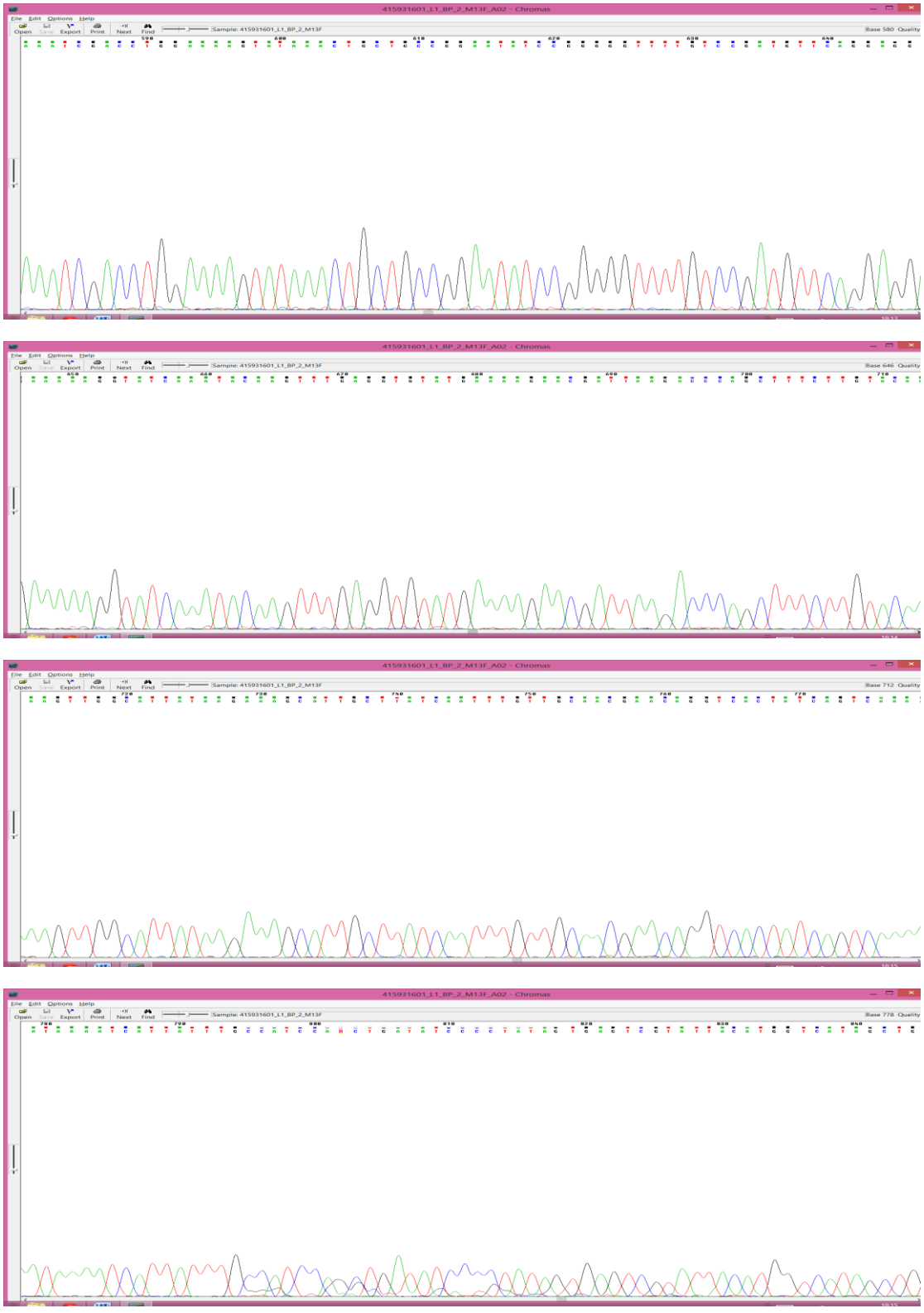

415931601_L1_BP_2_M13F_A02 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 415931601_L1_BP_2_M13F
Base 248 Quality


415931601_L1_BP_2_M13F_A02 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 415931601_L1_BP_2_M13F
Base 381 Quality


415931601_L1_BP_2_M13F_A02 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 415931601_L1_BP_2_M13F
Base 315 Quality
Close


415931601_L1_BP_2_M13F_A02 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 415931601_L1_BP_2_M13F
Base 447 Quality


415931601_L1_BP_2_M13F_A02 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 415931601_L1_BP_2_M13F
Base 514 Quality

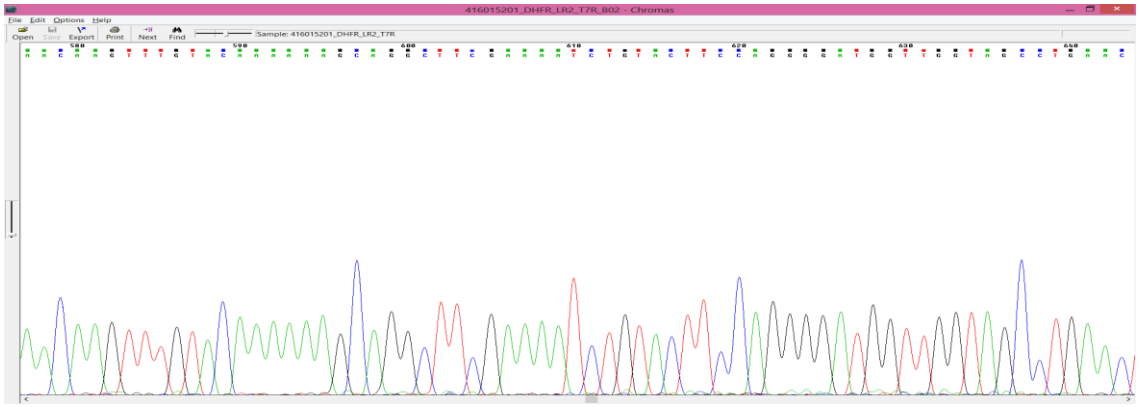


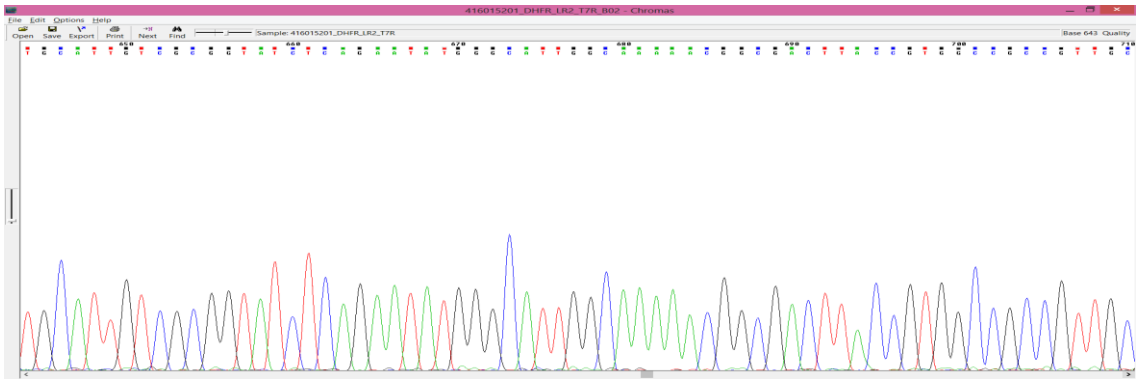
Appendix AE. Chromatogram Sequence Alignment of the Rare Codon Optimised *DHFRL1* BP Clone

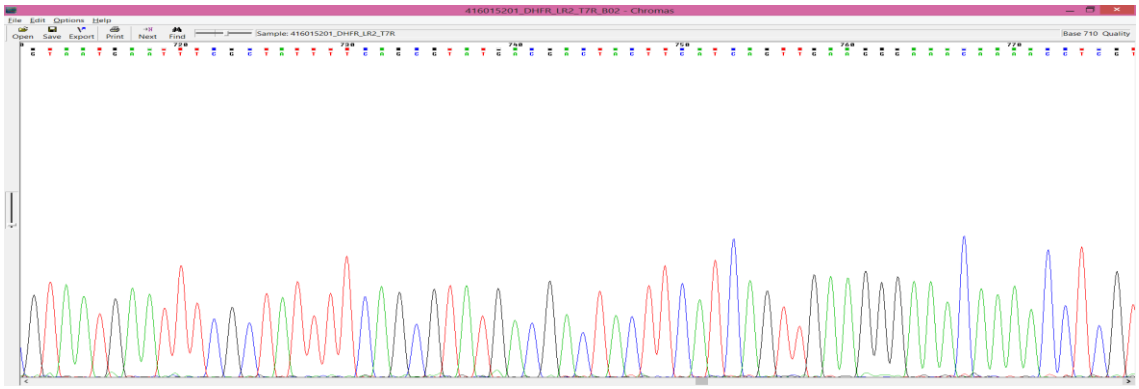
```
DHFR           ------------------------------------------------------------
DHFR_Clone     TTGTTTTAAAAAACGTATTGAAGCTATCCCACAAATTGATAAGTACTTGAAATCCAGCAA 480


DHFR           ---------------------------------------------GGGG---------- 4
DHFR_Clone     GTATATAGCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCC 540
                                                           ** *

DHFR           ------------------------------------ACAAGTTTGTACAAAAAAGCAGG 27
DHFR_Clone     AAAATCGGATCTGGTTCCGCGTCCATGGTCGAATCAAACAAGTTTGTACAAAAAAGCAGG 600
                                                   **********************

DHFR           CTTCGAAAATCTGTACTTCCAGGGGATGGTTGGTAGCCTGAACTGCATTGTCGCGGTATC 87
DHFR_Clone     CTTCGAAAATCTGTACTTCCAGGGGATGGTTGGTAGCCTGAACTGCATTGTCGCGGTATC 660
               ************************************************************

DHFR           TCAGAATATGGGCATTGGCAAAAACGGCGACTTACCGTGGCCGCCGTTGCGTAATGAATT 147
DHFR_Clone     TCAGAATATGGGCATTGGCAAAAACGGCGACTTACCGTGGCCGCCGTTGCGTAATGAATT 720
               ************************************************************

DHFR           TCGCTATTTTCAGCGTATGACGACTACTTCATCAGTTGAAGGGAAACAAAACCTCGTGAT 207
DHFR_Clone     TCGCTATTTTCAGCGTATGACGACTACTTCATCAGTTGAAGGGAAACAAAACCTCGTGAT 780
               ************************************************************

DHFR           TATGGGCAAGAAGACCTGGTTCTCTATTCCCGAGAAAAATCGTCCGTTAAAGGGCCGTAT 267
DHFR_Clone     TATGGGCAAGAAGACCTGGTTCTCTATTCCCGAGAAAAATCGTCCGTTAAAGGGCCGTAT 840
               ************************************************************

DHFR           TAACCTTGTGTTATCTCGCGAGCTGAAGGAGCCGCCGCAGGGGGCCCATTTCTTATCCCG 327
DHFR_Clone     TAACCTTGTGTTATCTCGCGAGCTGAAGGAGCCGCCGCAGGGGGCCCATTTCTTATCCCG 900
               ************************************************************

DHFR           TTCCCTGGATGACGCACTGAAGCTGACCGAACAACCTGAACTGGCTAATAAAGTCGACAT 387
```

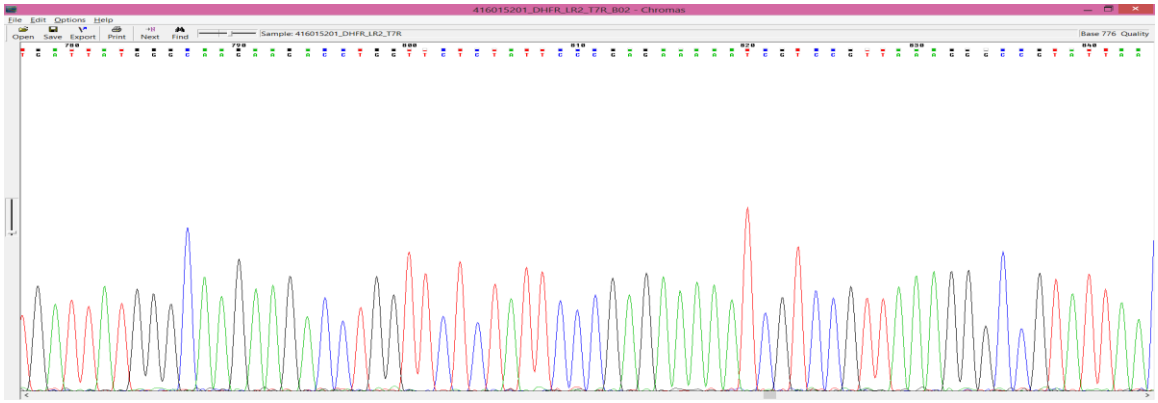
```
DHFR_Clone      TTCCCTGGATGACGCACTGAAGCTGACCGAACAACCTGAACTGGCTAATAAAGTCGACAT 960
                ************************************************************

DHFR            GGTGTGGATTGTAGGCGGTAGCTCGGTTTATAAAGAAGCGATGAACCACCCCGGGCACTT 447
DHFR_Clone      GGTGTGGATTGTAGGCGGTAGCTCGGTTTATAAAGAAGCGATGAACCACCCCGGGCACTT 1020
                ************************************************************

DHFR            AAAACTGTTCGTAACTCGCATCATGCAAGACTTTGAAAGTGATACGTTTTTCCCCGAAAT 507
DHFR_Clone      AAAACTGTTCGTAACTCGCATCATGCAAGACTTTGAAAGTGATACGTTTTTCCCCGAAAT 1080
                ************************************************************

DHFR            CGACCTGGAAAAGTATAAACTGCTGCCGGAATATCCGGGGGTTTTGTCCGATGTTCAGGA 567
DHFR_Clone      CGACCTGGAAAAGTATAAACTGCTGCCGGAATATCCGGGGGTTTTGTCCGATGTTCAGGA 1140
                ************************************************************

DHFR            GGAAAAAGGTATCAAATACAAGTTTGAGGTGTATGAAAAGAACGATTAAGACCCAGCTTT 627
DHFR_Clone      GGAAAAAGGTATCAAATACAAGTTTGAGGTGTATGAAAAGAACGATTAAGACCCAGCTTT 1200
                ************************************************************

DHFR            CTTGTACAAAGTGGTGGGG---------------------------------------- 646
DHFR_Clone      CTTGTACAAAGTGGTTTGATTCGACCCGGGATCCGGNGNGNGNNNNNNNNNNNNNNNNG 1260
                ***************  *.

DHFR            ------------
DHFR_Clone      NNNNNNNNNNNN 1272
```

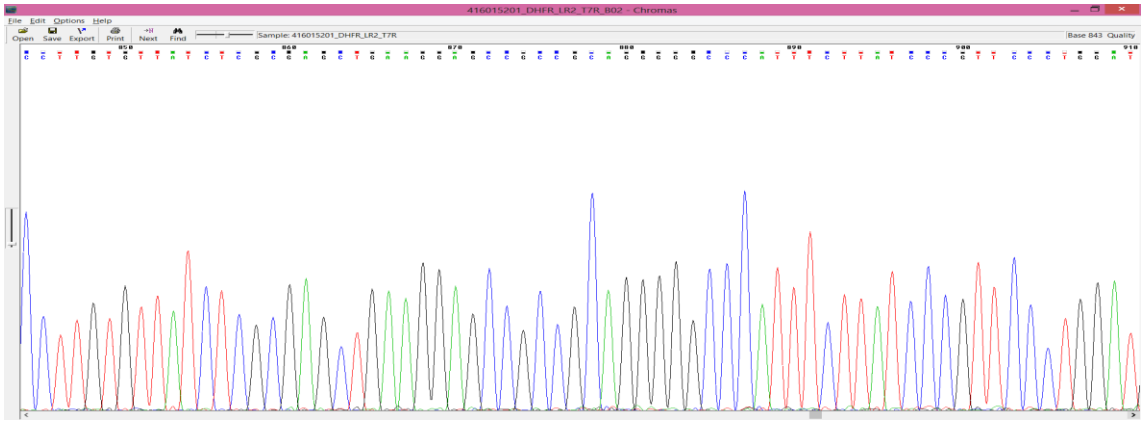
Appendix AF. Clustal Sequence Alignment of the Recombinant Rare Codon Optimised *DHFR* Clone

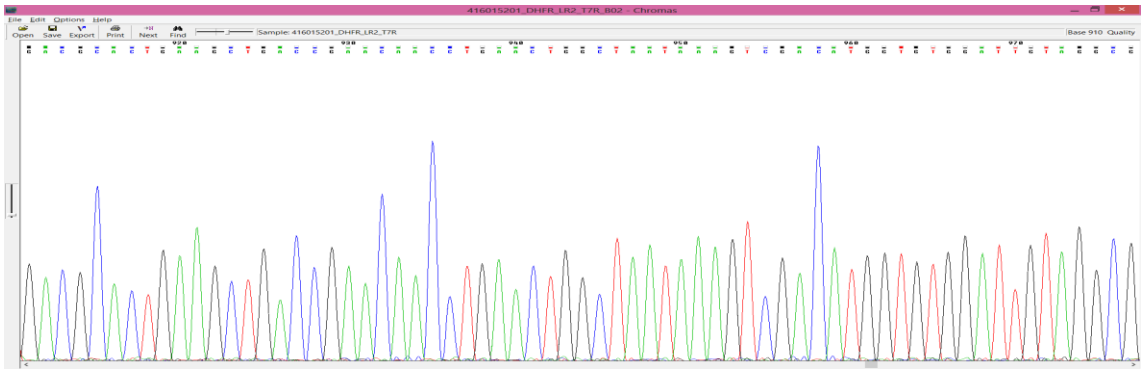

416015201_DHFR_LR2_T7R_802 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 416015201_DHFR_LR2_T7R

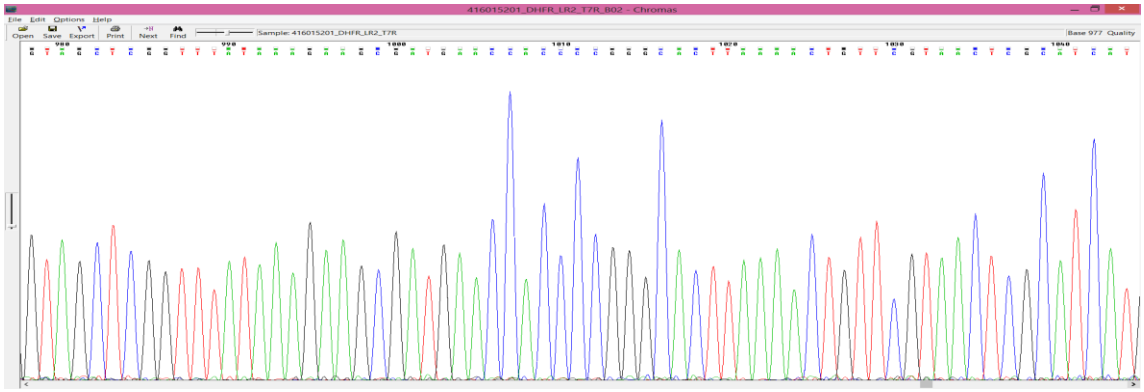

416015201_DHFR_LR2_T7R_802 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 416015201_DHFR_LR2_T7R
Base 643 Quality

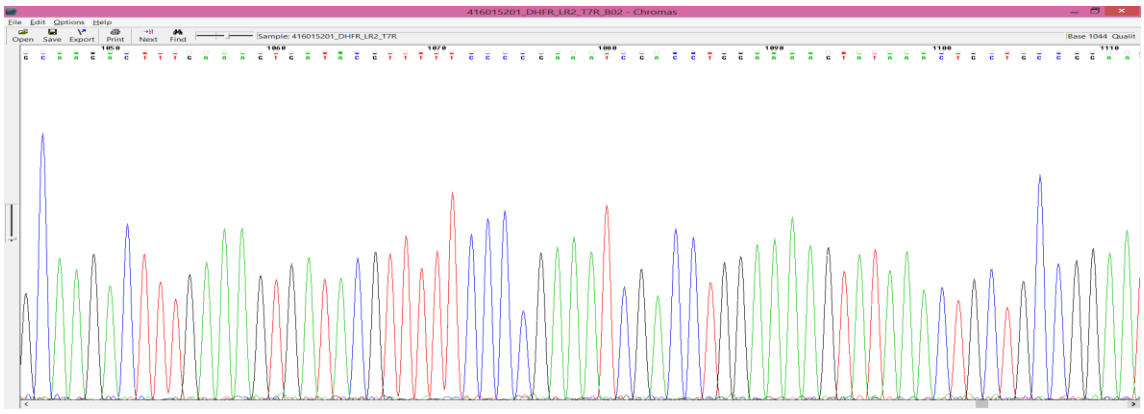

416015201_DHFR_LR2_T7R_802 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 416015201_DHFR_LR2_T7R
Base 710 Quality

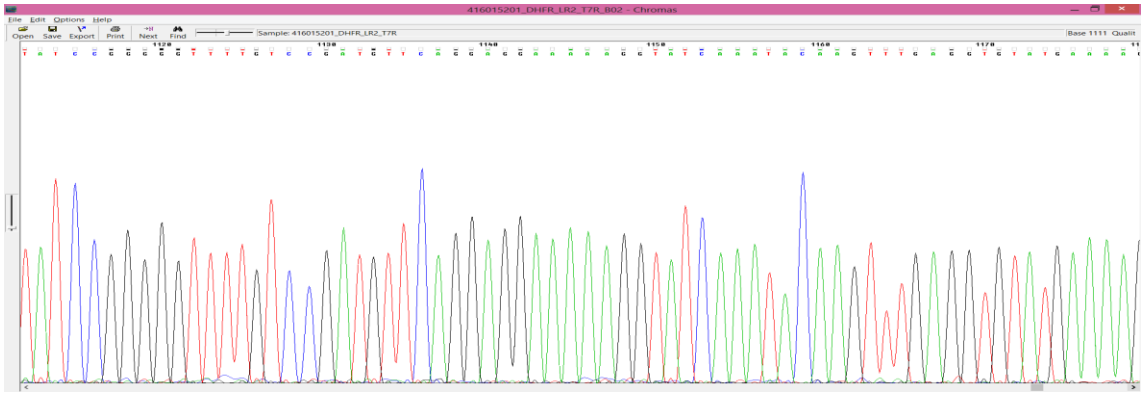

416015201_DHFR_LR2_T7R_802 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 416015201_DHFR_LR2_T7R
Base 776 Quality

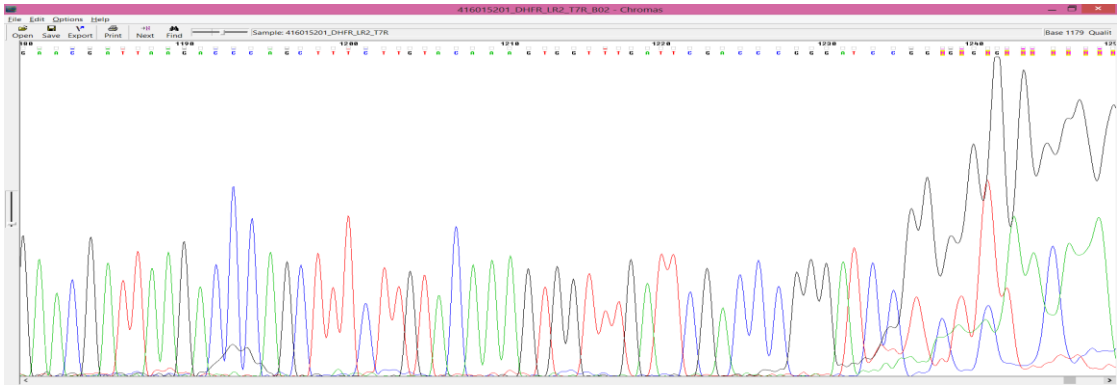

416015201_DHFR_LR2_T7R_802 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 416015201_DHFR_LR2_T7R
Base 843 Quality


416015201_DHFR_LR2_T7R_802 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 416015201_DHFR_LR2_T7R
Base 910 Quality


416015201_DHFR_LR2_T7R_802 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 416015201_DHFR_LR2_T7R
Base 977 Quality


416015201_DHFR_LR2_T7R_802 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 416015201_DHFR_LR2_T7R
Base 1044 Qualit



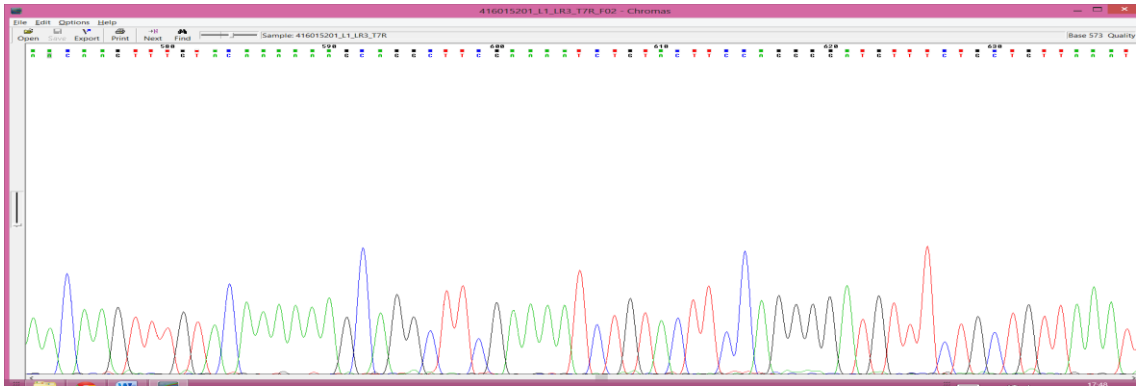


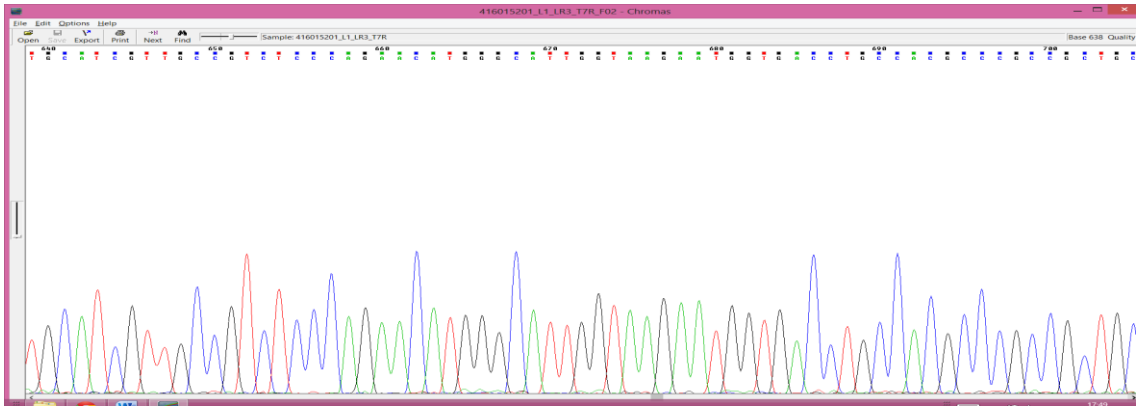
Appendix AG. Chromatogram Sequence Alignment of the Recombinant Rare Codon Optimised *DHFR* Clone

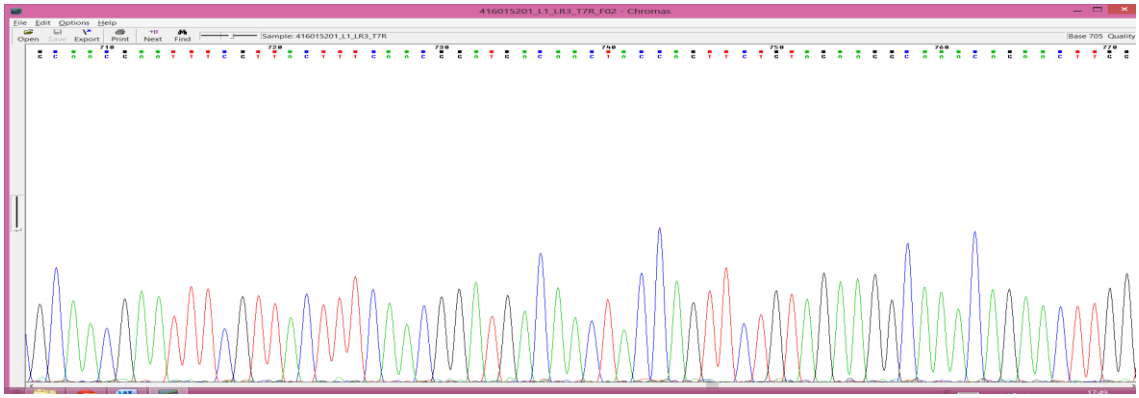
```
DHFRL1          ------------------------------------------------------------
DHFRL1_Clone    TTAAAAAACGTATTGAAGCTATCCCACAAATTGATAAGTACTTGAAATCCAGCAAGTATA 480


DHFRL1          ------------------------------------GGGG------------------ 4
DHFRL1_Clone    TAGCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAAT 540
                                                    ** *

DHFRL1          --------------------------------ACAAGTTTGTACAAAAAAGCAGGCTTCG 32
DHFRL1_Clone    CGGATCTGGTTCCGCGTCCATGGTCGAATCAAACAAGTTTGTACAAAAAAGCAGGCTTCG 600
                                               ****************************

DHFRL1          AAAATCTGTACTTCCAGGGGATGTTTCTGCTGTTAAATTGCATCGTTGCCGTCTCCCAGA 92
DHFRL1_Clone    AAAATCTGTACTTCCAGGGGATGTTTCTGCTGTTAAATTGCATCGTTGCCGTCTCCCAGA 660
                ************************************************************

DHFRL1          ACATGGGCATTGGTAAGAATGGTGACCTGCCACGCCCGCCGCTGCGCAACGAATTTCGTT 152
DHFRL1_Clone    ACATGGGCATTGGTAAGAATGGTGACCTGCCACGCCCGCCGCTGCGCAACGAATTTCGTT 720
                ************************************************************

DHFRL1          ACTTTCAACGGATGACAACTACCAGTTCTGTAGAAGGCAAACAGAACTTGGTGATCATGG 212
DHFRL1_Clone    ACTTTCAACGGATGACAACTACCAGTTCTGTAGAAGGCAAACAGAACTTGGTGATCATGG 780
                ************************************************************

DHFRL1          GCCGTAAAACCTGGTTCAGTATCCCGGAAAAGAACCGCCCGTTAAAAGATCGGATTAATC 272
DHFRL1_Clone    GCCGTAAAACCTGGTTCAGTATCCCGGAAAAGAACCGCCCGTTAAAAGATCGGATTAATC 840
                ************************************************************

DHFRL1          TGGTGTTGAGTCGCGAACTTAAGGAACCACCGCAGGGGGCACACTTTCTGGCCCGTTCAC 332
DHFRL1_Clone    TGGTGTTGAGTCGCGAACTTAAGGAACCACCGCAGGGGGCACACTTTCTGGCCCGTTCAC 900
                ************************************************************

DHFRL1          TGGATGATGCTCTGAAATTAACAGAACGTCCAGAATTAGCGAACAAAGTCGATATGATCT 392
```

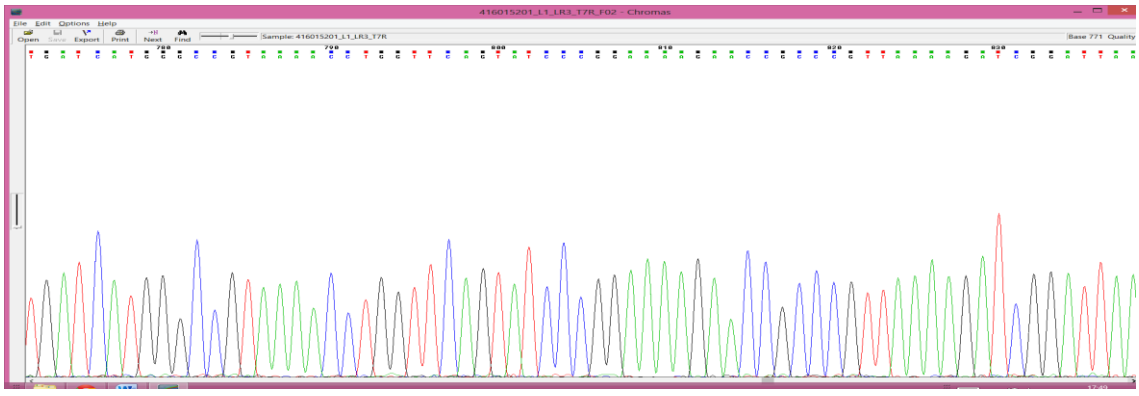
```
DHFRL1_Clone      TGGATGATGCTCTGAAATTAACAGAACGTCCAGAATTAGCGAACAAAGTCGATATGATCT 960
                  ************************************************************

DHFRL1            GGATTGTCGGTGGGAGTAGCGTCTACAAAGAGGCGATGAATCATCTGGGCCACTTGAAGC 452
DHFRL1_Clone      GGATTGTCGGTGGGAGTAGCGTCTACAAAGAGGCGATGAATCATCTGGGCCACTTGAAGC 1020
                  ************************************************************

DHFRL1            TCTTTGTCACCCGGATCATGCAGGATTTTGAGTCAGATACATTCTTTTCTGAAATCGATC 512
DHFRL1_Clone      TCTTTGTCACCCGGATCATGCAGGATTTTGAGTCAGATACATTCTTTTCTGAAATCGATC 1080
                  ************************************************************

DHFRL1            TGGAAAAATATAAACTTCTGCCGGAATACCCGGGCGTTCTCAGTGACGTCCAGGAAGGTA 572
DHFRL1_Clone      TGGAAAAATATAAACTTCTGCCGGAATACCCGGGCGTTCTCAGTGACGTCCAGGAAGGTA 1140
                  ************************************************************

DHFRL1            AGCACATCAAATATAAATTTGAGGTGTGCGAAAAAGACGATTAAGACCCAGCTTTCTTGT 632
DHFRL1_Clone      AGCACATCAAATATAAATTTGAGGTGTGCGAAAAAGACGATTAAGACCCAGCTTTCTTGT 1200
                  ************************************************************

DHFRL1            ACAAAGTGGTGGGG---------------------------------------------- 646
DHFRL1_Clone      ACAAAGTGGTTTGATTCGACCCGGGATNNNNNNNNNNNNNNNNNNNCNNNCNNNNNNNNN 1260
                  **********  *.

DHFRL1            ------------
DHFRL1_Clone      NNNNNNNNNNNN 1272
```

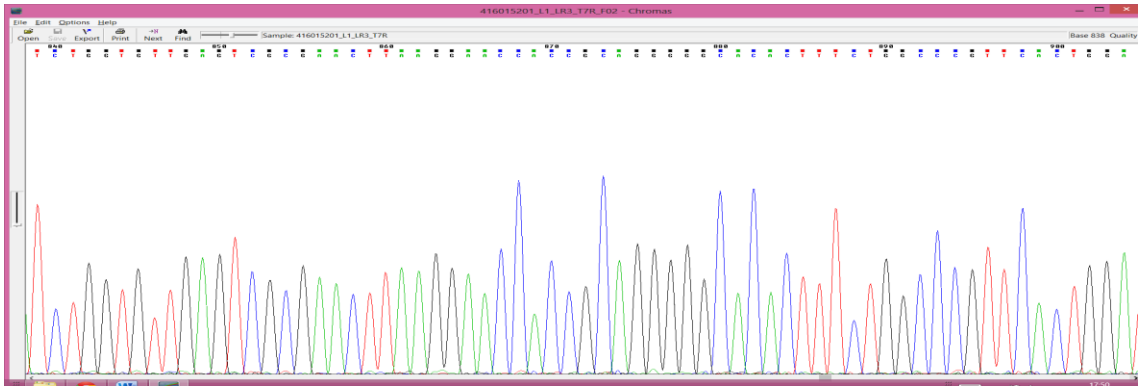
Appendix AH. Clustal Sequence Alignment of the Recombinant Rare Codon Optimised *DHFRL1* Clone

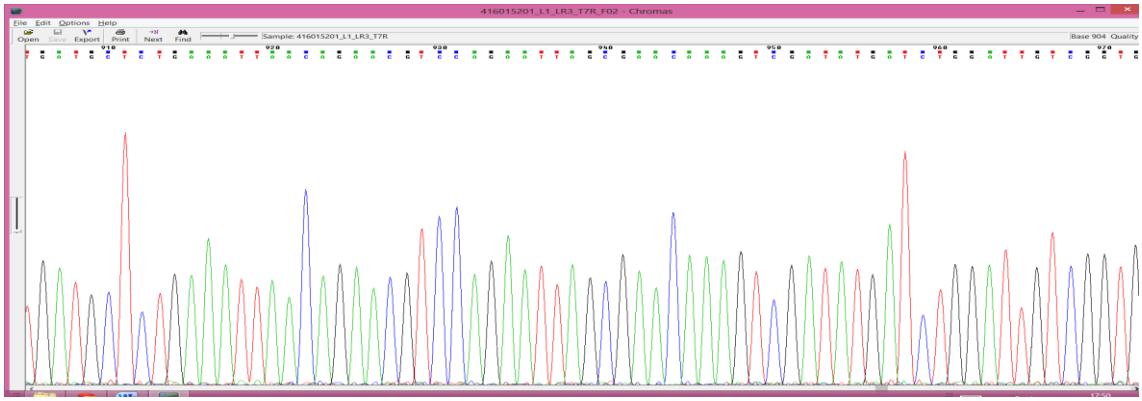

416015201_L1_LR3_T7R_F02 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 416015201_L1_LR3_T7R
Base 573 Quality

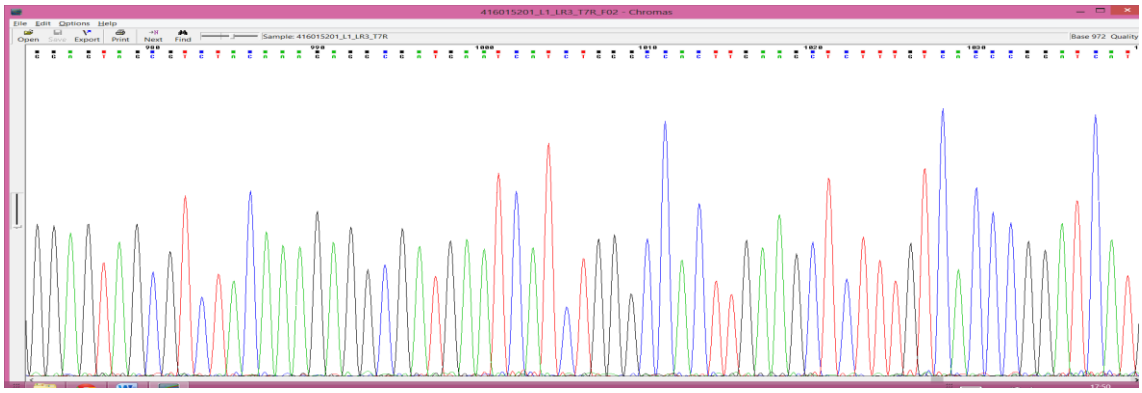

416015201_L1_LR3_T7R_F02 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 416015201_L1_LR3_T7R
Base 638 Quality

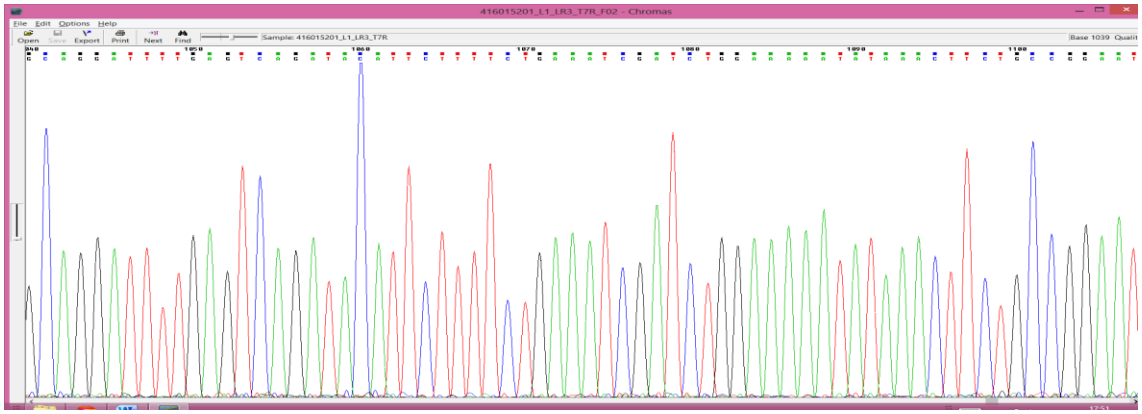

416015201_L1_LR3_T7R_F02 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 416015201_L1_LR3_T7R
Base 705 Quality

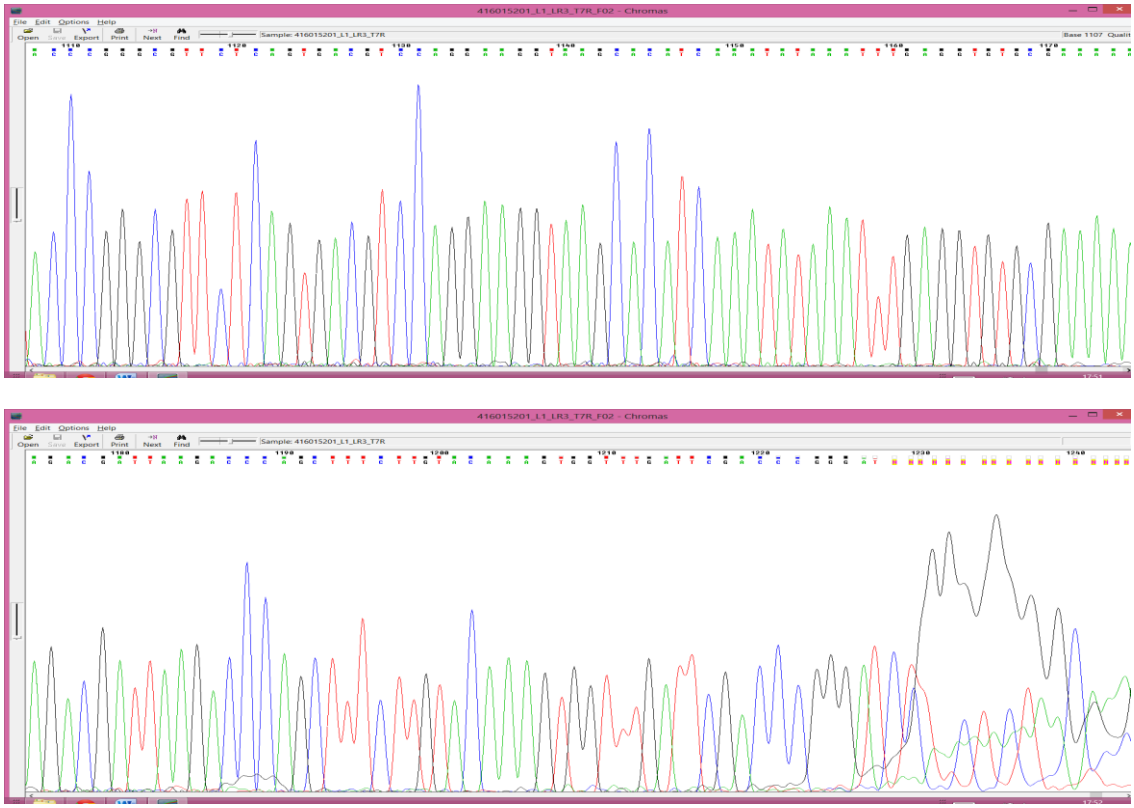

416015201_L1_LR3_T7R_F02 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 416015201_L1_LR3_T7R
Base 771 Quality


416015201_L1_LR3_T7R_F02 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 416015201_L1_LR3_T7R
Base 838 Quality

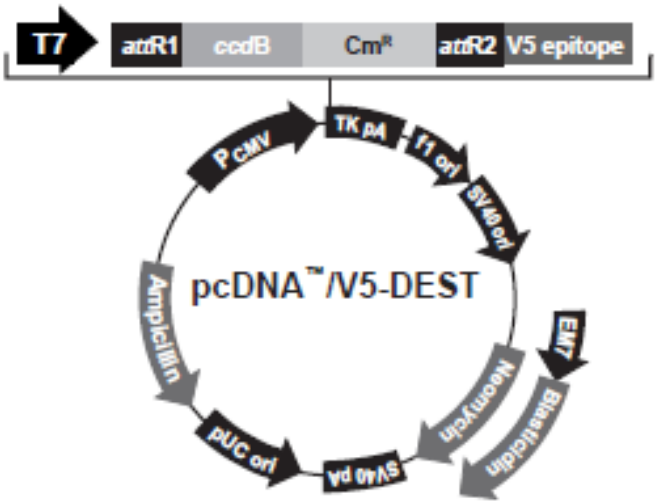

416015201_L1_LR3_T7R_F02 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 416015201_L1_LR3_T7R
Base 904 Quality


416015201_L1_LR3_T7R_F02 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 416015201_L1_LR3_T7R
Base 972 Quality


416015201_L1_LR3_T7R_F02 - Chromas
File Edit Options Help
Open Save Export Print Next Find
Sample: 416015201_L1_LR3_T7R
Base 1039 Quality



Appendix AJ. Chromatogram Sequence Alignment of the Recombinant Rare Codon Optimised *DHFRL1* Clone

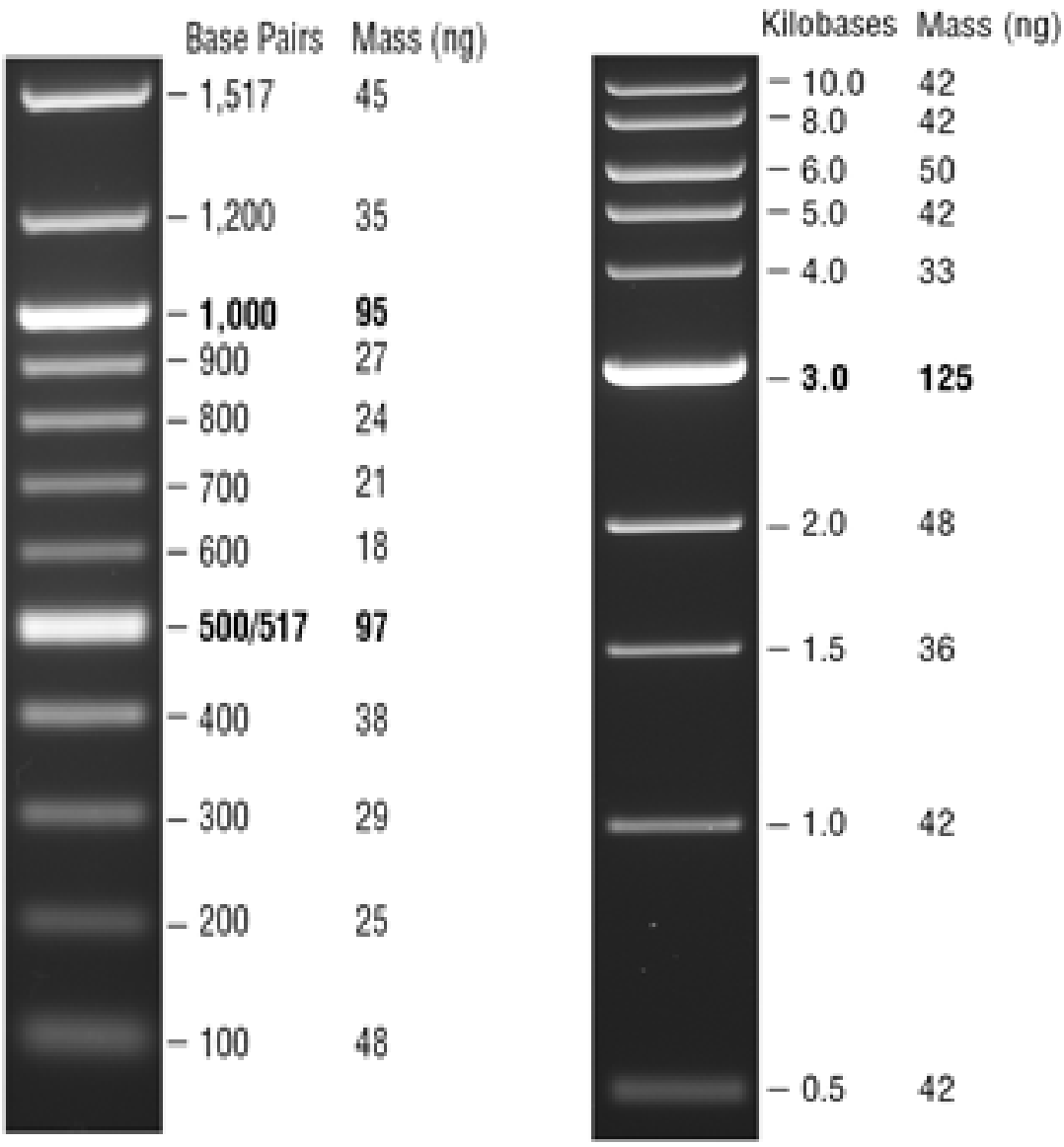


Comments for:	pcDNA™3.2/V5-DEST 7711 nucleotides	pcDNA™6.2/V5-DEST 7341 nucleotides
CMV promoter:	232-819	232-819
T7 promoter/priming site:	863-882	863-882
attR1 site:	911-1035	911-1035
ccdB gene (c):	1464-1769	1464-1769
Chloramphenicol resistance gene (c):	2111-2770	2111-2770
attR2 site:	3051-3175	3051-3175
V5 epitope:	3201-3242	3201-3242
V5 reverse priming site:	3210-3230	3210-3230
TK polyadenylation signal:	3269-3540	3269-3540
f1 origin:	3576-4004	3576-4004
SV40 early promoter and origin:	4031-4339	4031-4339
Neomycin resistance gene:	4414-5208	—
EM7 promoter:	—	4394-4460
Blasticidin resistance gene:	—	4461-4859
SV40 early polyadenylation signal:	5384-5514	5017-5147
pUC origin (c):	5897-6570	5530-6200
Ampicillin (*bla*) resistance gene (c):	6715-7575	6345-7205
bla promoter (c):	7576-7674	7206-7304

(c) = complementary strand

Invitrogen™
life technologies

Appendix AK. Vector map for pcDNA / V5-DEST. Image taken from Invitrogen



A **B**

Appendix AL. DNA Ladders. Image taken from New England BioLabs (NEB). A. 100 bp ladder. B. 1 kb ladder. The DNA ladders were loaded on to the agarose gels in order to determine the sizes of the PCR products and plasmid preps. Ladder used for each case depended on the size of the products.

Appendix AM. ColorBurst Electrophoresis Marker (Protein Ladder). Image taken from Sigma-Aldrich. The protein ladder was loaded on to SDS-PAGE gels in order to determine the sizes of the proteins for SDS-PAGE and Western blotting analysis.

The Impact of Nutrition on Differential Methylated Regions of the Genome[1]

Anne Parle-McDermott* and Mari Ozaki
School of Biotechnology, Dublin City University, Glasnevin, Dublin, Ireland

ABSTRACT

Nutrition has always played an important role in health and disease, ranging from common diseases to its likely contribution to the fetal origins of adult disease. However, deciphering the molecular details of this role is much more challenging. The impact of nutrition on the methylome, i.e., DNA methylation, has received particular attention in more recent years. Our understanding of the complexity of the methylome is evolving as efforts to catalog the DNA methylation differences that exist between different tissues and individuals continue. We review selected examples of animal and human studies that provide evidence that, in fact, specific genes and DNA methylation sites are subject to change during development and during a lifetime as a direct response to nutrition. Investigation of the methyl donors folate, choline, and methionine provide the most compelling evidence of a role in mediating DNA methylation changes. Although a number of candidate regions/genes have been identified to date, we are just at the beginning in terms of cataloging so-called nutrient-sensitive methylation variable positions in humans. *Adv. Nutr. 2: 463–471, 2011.*

Introduction

The methylome describes the modifications to DNA that potentially exist within a given genome without changes to the DNA sequence itself. In humans, traditionally these modifications mostly referred to methylation of cytosine residues (5mC),[2] usually in the context of a CpG. However, more recent work indicates that non-CpG methylation (1) and 5-hydroxymethylcytosine (2,3) also form part of the human methylome. Methylation of DNA falls under the umbrella term *epigenetics*, which describes heritable changes to DNA and chromatin that are passed on to daughter cells either mitotically or meiotically. Although histone modifications and their impact on chromatin remodeling are highly relevant and interplay with DNA modifications, they are not the focus of this review [see Choi and Friso (4)].

Research interest in DNA methylation gained momentum in the past 10 years and focused on its effect on gene expression and its association with various disease states, most notably cancer (5). Methylation of DNA in the region of a gene is generally associated with gene expression being switched off, whereas unmethylated DNA is associated with gene expression being switched on. The exact mechanism is not fully elucidated but is believed to involve the interference of methylated cytosine with binding of the RNA polymerase complex and associated transcription factors through an interaction with chromatin (6) (**Fig. 1**). Although this has been the prevailing mechanism, more recent research indicates that DNA methylation is not always associated with transcriptional repression, particularly when it occurs in gene bodies rather than in the promoters of genes (7). Cells use DNA methylation as a mechanism to control gene expression in a number of different ways. The types of DNA methylation changes as they relate to specific functions as we currently understand them range from tissue-specific gene expression to inactivation of imprinted genes, the X-chromosome, and transposable elements. DNA methylation changes have also been associated with a range of disease states including imprinting disorders (8), tumorigenesis (5), neurological disorders (9), cardiovascular disease (10), and autoimmune diseases (11). Apart from definitive pathologically induced changes, evidence to date suggests that there are also regions of the methylome that are subject to change in response to the environment including nutrition. A comprehensive overview of the type of sites that are subject to this is not currently available, but preliminary work on this indicates that it is likely to include regions that silence transposable elements, possibly some imprinted genes and other site types that have yet to be identified. A definitive term has not been assigned to describe these sites that are subject to

[1] Author disclosures: M. Ozaki and A. Parle-McDermott, no conflicts of interest.
[2] Abbreviations used: CGI, CpG island; DVR, differentially methylated region; 5hmC, 5-hydroxymethylcytosine; 5mC, 5-methylcytosine; IUGR, intrauterine growth restriction; MTHFR, methyltetrahydrofolate reductase; MVP, methylation variable position.

* To whom correspondence should be addressed. E-mail: anne.parle-mcdermott@dcu.ie

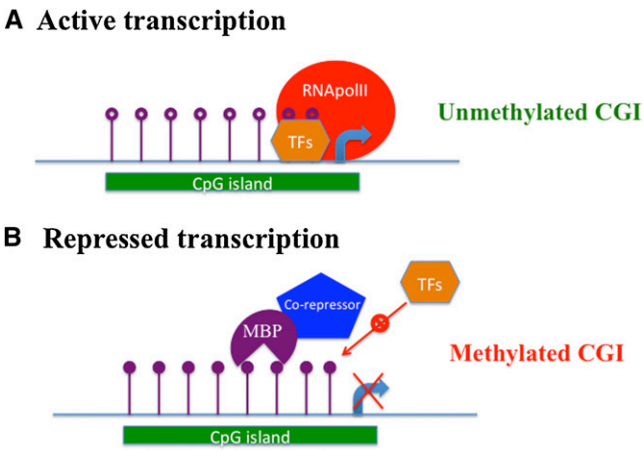


FIGURE 1 The role of CGI methylation in the control of gene transcription. Unmethylated and methylated CGIs in the promoter region of a given gene are shown. Open circle lollipops indicate unmethylated cytosines; closed circle lollipops indicate methylated cytosines. The nucleosomes of chromatin are not shown. *(A)* Unmethylated CGIs are usually in a transcriptionally permissive state in tandem with specific histone modifications. TFs recognize specific DNA motifs in the promoter and facilitate binding of the RNA polymerase II (RNApolII) complex to initiate transcription at the transcription start site (curved arrow). *(B)* Methylated CGIs are associated with a transcriptionally repressed state. This is thought to be due to the binding of MBP, which recognizes methylated cytosines, which subsequently recruit co-repressor complexes that influence the chromatin state via histone deacetylases and ultimately prevent binding by TFs. An alternative mechanism is that methylated cytosines directly prevent binding of TFs. CGI, CpG island; MBP, methyl-binding protein; TFs, transcription factors.

environmental influence, but they have been referred to as metastable alleles or epialleles (12). The Human Epigenome Project (13) uses the term MVP to describe individual DNA methylation sites that show tissue-specific changes or changes in response to a disease state. Others refer to a group of such sites as DMRs (14). Our focus is on those DNA methylation sites that are influenced by nutritional status—so-called nutrient-sensitive MVPs.

Current status of knowledge

The human methylome

Efforts such as the 1000 Genomes Project (15) have provided a comprehensive catalog of the genetic variation that exists between human individuals, but what about the methylome? It is clear that DNA methylation patterns will vary not only among tissues but also among individuals, but the extent and variability of these patterns have yet to be established. The Human Epigenome Project (13) and the NIH Roadmap Epigenomics (16) projects are bids to bridge the gap in our current knowledge of the variation in DNA methylation patterns that exist among individuals. In the same way that single nucleotide polymorphisms provide a catalog of human genetic variation, these projects aim to catalog a variety of epigenetic marks including MVPs, i.e., those DNA methylation sites that vary in different tissues, among individuals, and in disease states.

Extensive human methylome data across a number of different populations and disease states are currently being generated, as alluded to previously. Although such data sets will provide us with a unique insight into human methylome patterns, we rely on recently completed analyses to give us a glimpse of what might be revealed. The first most detailed human methylome, i.e., at single-base-resolution, was published in 2009 by Lister et al. (1). The DNA methylation patterns of human stem cells and fetal fibroblasts were assessed, and a number of interesting patterns were identified that challenged some of the previous assumptions regarding DNA methylation patterns. Previously, most cytosine methylation was thought to occur in the context of CG (or CpG). High-density CG regions are referred to as CGIs, can occur in the promoter regions of genes, and, when they do, are often unmethylated (17). The analysis by Lister et al. (1) revealed that in stem cells, one-fourth of cytosine methylation occurred in a non-CG context and were enriched in the coding regions of genes. Highly expressed genes tended to have a 3-fold higher level of non-CG methylation density than nonexpressed genes. However, this non-CG methylation started to be lost on cellular differentiation and appeared to be a specific feature of stem cells. In human fibroblasts, they noted that reduced methylation levels were associated with decreased transcriptional activity for some genes. Thus, unmethylated cytosines are not always associated with a high level of gene expression, and their impact on transcriptional activity appears to depend on their location within the gene.

5mC was always considered to be the fifth DNA base, but it appears that there is also a sixth base in the form of 5hmC. Although the existence of 5hmC in mammals was originally suggested in 1972 (18), it was only definitively shown in 2009 by Kriaucionis and Heintz (2) and Tahiliani et al. (3). The enzyme that catalyzes the conversion of 5mC to 5hmC was identified as TET1 (3). TET1 is a 2-oxoglutarate and Fe(II)-dependent enzyme that has been shown to bind throughout the genome in embryonic stem cells but particularly at high-density CpG sites in the promoters of genes and within specific genes (19). Modifications of 5hmC showed a pattern similar to that of TET1 binding, i.e., found particularly at high- and intermediate-density CpG sites, which is in contrast to 5mC, which is found predominantly at low-density CpG sites. Similarly to 5mC, however, 5hmC is thought to have a role in transcriptional repression, also mediated through TET1 (19). However, additional roles for 5hmC include the ability to "prime" specific genes for rapid activation when required (20), and it is also thought to have a role in the regulation of DNA methylation patterns (19).

Despite the rapid recent advances in interrogating the methylome, it is clear that understanding the relevance of these methylation marks and how they might vary among tissues and individuals is far from complete. We describe the current evidence demonstrating that nutrition has an impact on 5mC patterns in the following paragraphs, but there are no studies to date that have addressed this specifically at

the 5hmC level. It is likely that nutritional influences are equally relevant for 5hmC given that its formation appears to depend on the presence of 5mC (19). However, the majority of DNA methylation techniques to date are unable to distinguish between 5mC and 5hmC. Therefore, previous studies may need to be reassessed to consider the newly identified sixth DNA base.

Nutrition as a mediator of DNA methylation changes

The identification of verified nutrient-sensitive MVPs is of particular interest because it offers a potential mechanism to explain the observations that relate to the fetal origins of adult disease theory and the role of nutrition in a range of common diseases. The fetal origins of adult disease theory was originally proposed by Barker et al. (21,22) and states that the in utero environment of the embryo can have an impact on the adult disease risk of the offspring. It has also been proposed that some effects are transgenerational, i.e., can be passed on to subsequent generations. This theory was based on initial studies that demonstrated an association between fetal growth impairment and adult cardiovascular disease, diabetes, and insulin resistance (23–27). Fetal undernutrition is believed to be the key mitigating factor, and DNA methylation is one of the mechanisms that may mediate these effects. Additional studies such as the Dutch Hunger Winter found that in utero exposure to famine was associated with increased risks of obesity (men) (28) and schizophrenia (29), increased prevalence of obstructive airways disease (30) and impaired glucose and lipid homeostasis, an increase in coronary heart disease later in life, an increased risk of breast cancer (women), and reduced renal function in a manner that is dependent on gestational age at the time of exposure to famine (31). Evidence to support the fetal origins of adult disease theory has been reviewed extensively elsewhere (32). However, whether it is the changes to the methylome that are mediating these effects remains to be proven.

Apart from in utero exposure, DNA methylation patterns have been shown to change during one's lifetime. The most striking evidence of this was a twin study by Fraga et al. (33) that showed how the methylome of identical twins became more divergent as they aged. They also correlated this with changes in the expression of a range of genes. It is likely that nutrition is one of the environmental influences that have an impact on these methylome changes. Another twin study by Kaminksy et al. (34) suggested that the phenotypic similarity of monozygotic twins is attributed not only to their identical genomes, but also to sharing a similar epigenome compared with dizygotic twins. Further evidence of age-dependent DNA methylation changes has recently been reported for a number of specific genes (35).

It is clear that DNA methylation patterns can change during our development and as we age, but a role in disease development and progression has received particular focus in recent years. The most intensely studied has been changes associated with cancer (5). However, it is difficult to disentangle the sites relevant to nutrition given the complexity and mutational nature of cancer development. Other diseases that exhibit dramatic gene-specific DNA methylation changes include dementia and systemic lupus erythematosus (36).

While we await high-resolution methylome profiles of different tissues from a variety of populations, it is clear that there are regions of the genome that are subject to DNA methylation changes during development, cellular differentiation, or one's lifetime. These may be distinct or overlap with disease-associated DNA methylation changes. The challenge now is to actually identify those nutrient-sensitive MVPs that have relevance to a phenotype. For the remainder of this review, we describe the current evidence of nutrient-sensitive MVPs both in animal models and human studies.

Nutrient-sensitive MVPs

Animal models

The evidence to date suggests that the impact of nutrition on DNA methylation patterns can happen at the level of the organism, probably in a tissue-specific fashion or transgenerationally, i.e., can have an impact on the subsequent progeny. Examples of those nutrients that have been shown to influence DNA methylation patterns are described in the following and are summarized in **Table 1**. Mutant animal models are not included because they complicate the relevance of nutrition on the methylome in these animals.

Methyl donors: folate, methionine, and choline. The folate metabolic pathway consists of a plethora of enzymes and is compartmentalized between the cytoplasm, nucleus, and mitochondria (37). Folate metabolism provides the 1-carbons necessary for a number of reactions involving the synthesis of DNA and in methylation reactions (**Fig. 2**). S-adenosylmethionine is particularly relevant for DNA methylation because it provides the actual methyl group (CH_3) that is added onto cytosine to produce 5mC. S-adenosylhomocysteine is the product formed after donation of the methyl group from S-adenosylmethionine and is an inhibitor of methyltransferases including DNA methyltransferases. Thus, the suggestion that folate supply and other methyl donors such as methionine and choline can influence DNA methylation patterns of an organism has strong biological plausibility. This was elegantly demonstrated in the *agouti* mouse model (38). This mouse model is probably one of the most widely cited animal models in support of the fetal origins of adult disease theory, i.e., transgenerational effects. This study showed how a maternal diet supplemented with methyl-donating vitamins, including folic acid, resulted in hypermethylation of a specific region of the mouse genome, producing an altered coat color phenotype in the mice offspring that were genetically identical (39). Inappropriate expression at the viable yellow *agouti* locus (A^{vy}) due to changes in the methylation status of a cryptic gene promoter led to the variable observable phenotypes. This was one of the first studies that provided experimental evidence in support of DNA methylation as one of the molecular mechanisms mediating maternal diet-related phenotypic

TABLE 1. Summary of selected studies investigating nutrient sensitive DNA methylation sites in animal models[1]

Nutrient	Model	DNA methylation site(s) examined	References
Methyl donor supplementation	Maternal mouse offspring	*Agouti* gene	38–41
		Cdk5 activator binding protein	41
		Axin^{Fu} gene	42
		82 genes including *runx3*	43
Methyl donor deficiency	Maternal sheep offspring	1400 CpG islands	44
	Adult rats	*p53*	45
	Postweaning mouse	*Igf2*	46
	Pregnant mice	*Esr1*, *Igf2*, *Slc39a4CC*	47
	Rat brain	Global methylation	48
Choline deficiency	Maternal mouse offspring	*Cdkn3*, global methylation	49
	Maternal rat offspring	*Dnmt1*, *Ifg2*, global methylation	50
Protein restriction	Adult rats, F2 offspring	PPARα	52, 53
	Maternal mouse offspring	CpG islands including *Lxr*α	54
	Cultured mouse blastocysts	*H19*	55
	Paternal rat offspring	PPARα: CpG island 50kb upstream	58
	Adult mice	Leptin promoter	60
	Maternal rat offspring/leptin injections	PPARα, glucocorticoid receptor	61
High-fat diet	Adult rats	Leptin promoter	59
Genistein	Adult mice	Differential method including ribosomal DNA and desmin-binding fragment	64
Genistein supplementation	Maternal mouse offspring	*Agouti* gene	65

[1] Methyl donors include folate, methionine, and choline.

changes in offspring (39). These DNA methylation changes were also seen to persist in offspring born in the next generation who were not exposed to methyl-donating vitamins in utero (40). Decreased methylation at the A^{vy} locus was also observed in response to the chemical Bisphenol A but was reversed by supplementation of the diet with methyl donors or the phytoestrogen genistein (41). A similar finding was also observed at an additional locus, i.e., the cdk5 activator binding protein (41). Methyl-donating supplementation transgenerational effects have been reported in other animal models including increased methylation at the *Axin*^{Fu} gene in offspring (42). Hollingsworth et al. (43) identified 82 potentially differentially methylated genes in the lung tissue of mouse progeny whose mothers were supplemented with methyl donors compared with controls. Among these, additional convincing evidence identified *Runx3* as being excessively methylated. Apart from rodent supplementation, studies have also considered methyl donor deficiency diets in other animal models. Female sheep were fed a vitamin B-12, folate, and methionine–restricted diet periconceptionally (44). Their offspring were heavier and fatter and had an altered immune response and insulin resistance, and males particularly had elevated blood pressure compared with controls. Methylation of 1400 CGIs was assessed by restriction landmark genome scanning in the fetal liver of the offspring, and 4% of sites showed altered methylation status. Studies investigating the effect of methyl donor deficiency to the individual organism rather than the offspring have also been conducted. The cancer-relevant *p53* gene was demethylated at specific sites in F344 rats fed a diet deficient in the methyl donors choline, methionine, and folic acid for 9 wk. However, it was also noted that some CpG sites were resistant to demethylation (45). Mice subjected to a postweaning methyl donor–deficient diet demonstrated a loss of imprinting of the *Igf2* gene compared with those mice fed the control diet (46). This study also highlights the relevance of the postnatal diet and how it can also influence nutrient-sensitive MVPs. A more recent study (47) assessed the impact of a folate-deficient diet on female mice before and during pregnancy and lactation. DNA methylation patterns of three genes including *Esr1*, *Igf2*, and *Slc39a4CC* were assessed in the dam's blood, liver, and

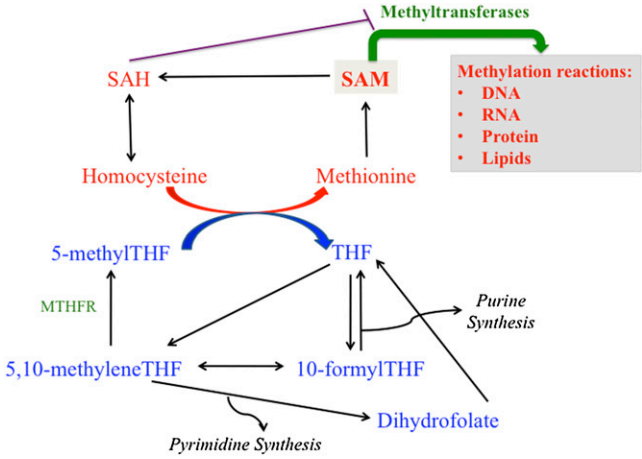


FIGURE 2 A simplified view of cytoplasmic 1C folate metabolism. Not all enzymes are shown. The reactions in blue supply the 1Cs required for DNA synthesis. The reactions in red form part of the folate pathway that produces the methyl donor SAM. A range of methyltransferase enzymes catalyze the transfer of the methyl group (CH_3) from SAM to a number of substrates including DNA, i.e., methylation reactions. SAH is an inhibitor of methyltransferase enzymes. MTHFR, methyltetrahydrofolate reductase enzyme; SAH, S-adenosylhomocysteine; SAM, S-adenosylmethionine THF, tetrahydrofolate.

kidney. Although the focus of many studies has been on the offspring, this study highlights the relevance of nutrient-sensitive MVPs in pregnant mothers, particularly during a period of high demand for methyl-donating nutrients. DNA methylation changes were observed but were found to be gene and tissue specific. The tissue specificity of DNA methylation changes was also reported in rats fed a diet lacking methionine, choline, and folic acid (48). Global DNA methylation was assessed by HPLC–MS/MS and showed hypermethylation in rat brain and hypomethylation in rat liver compared with controls. However, the biological relevance and sensitivity of measuring global DNA methylation patterns (as opposed to genomewide DNA methylation changes) are not clear. Global DNA methylation by immunohistochemistry (49) and HPLC (50) were also assessed in animal models assessing the transgenerational impact of choline specifically. Dietary choline during embryonic development is thought to have a significant role in memory function in subsequent adults (49,51). A choline-deficient diet was introduced to mouse dams during days 12–17 of pregnancy. Fetal brains were isolated on embryonic day 17, and global DNA methylation was decreased in a specific region of the brain (49). Gene-specific DNA methylation was also assessed by bisulfite sequencing and showed that choline deficiency decreased DNA methylation of the *Cdkn3* gene but had no effect on *Cdkn2b* and *Calb2* (49). A rat study also assessed choline specifically but addressed both supplementation and deficiency. Female rats were fed a diet consisting of varying levels of choline during pregnancy, and the impact of this on DNA methylation of the fetal liver and brain was assessed. Exposure to choline deficiency resulted in reduced methylation of the DNA methyltransferase enzyme Dnmt1 with a corresponding increased level of mRNA. This appeared to lead to increased global DNA methylation and hypermethylation of *Igf2* (50).

These studies illustrate that dietary methyl-donor content can influence DNA methylation patterns, both transgenerationally and at the level of the organism. The propensity to change appears to be gene specific, CpG site specific, and tissue specific. Moreover, changes to global gene expression patterns appear to have little relevance to what is happening at the DNA site–specific level, and, therefore, the relevance of such measurements is questionable. As already mentioned, the link with methyl-donor supply and DNA methylation follows a plausible path, but what about the supply of other nutrients?

Protein and amino acids. Protein restriction is thought to mimic fetal undernutrition, a key factor associated with the fetal origins of adult disease concept. A protein-restricted diet in rats during pregnancy showed loss of DNA methylation in the promoters of glucocorticoid receptor and the PPARα by methyl-sensitive PCR (52). The loss of DNA methylation correlated with an increase in mRNA levels. Interestingly, folic acid supplementation prevented the methylation changes mediated by protein restriction. However, again, gene-specificity is at play here because no change in the promoter of PPARγ, an isoform of PPARα, was observed. A follow-up of this study also showed that the hypomethylation state of the target genes observed in the F1 generation is passed onto the F2 offspring, even when the F1 offspring were fed a normal diet (53). Maternal protein restriction was also observed to affect DNA methylation of a range of genes by CGI microarray analysis in fetal livers shortly after birth (54). Further analysis of one of these genes, i.e., liver-X-receptor alpha (*Lxr*α), showed hypermethylation in protein-restricted pups that correlated with a drop in mRNA levels. Loss of imprinting has also been observed in cultured mouse blastocysts at the maternally expressed *H19* gene (55). Demethylation at the normally hypermethylated paternal allele was dependent on the culture medium used. Medium containing extra amino acids did not display loss of imprinting, highlighting the relevance of the nutritional environment for the preservation of imprinting patterns. Again, gene specificity is observed because another imprinted gene, *Snrpn*, did not display loss of imprinting in either medium. In addition to maternal transgenerational nutritional effects, evidence of paternal effects has also been shown (56,57), which highlights the importance of considering alternative mechanisms beyond maternal effects. Offspring livers of male rats on a low protein-diet showed a significant increase in DNA methylation at an intragenic CGI, $\sim$50 kb upstream of the *PPAR*α gene by bisulfite sequencing. Because the global methylation levels in low-protein and control offspring were similar, this further demonstrates the relevance of gene specificity (58).

Fat. Leptin is an adipokine that is involved in the regulation of body weight and food intake. A recent study describes how a specific CpG site within the leptin promoter showed significant methylation differences in rats fed a high-fat diet compared with controls. The change in methylation was associated with circulating leptin levels (59). Changes in the leptin promoter, i.e., the removal of methyl groups at CpG sites within the promoter, was also shown in mice fed a low-protein diet (60). The impact of leptin on DNA methylation was further investigated by Gluckman et al. (61). Adult rats that were previously injected with leptin after birth to well-nourished mothers demonstrated a differential methylation change in PPARα and glucocorticoid receptor compared with those born to undernourished mothers. This article highlights how maternal diet history can influence subsequent neonatal and adult cellular responses.

Genistein. Soy products are known to reduce the risk of cardiovascular disease and carcinogenesis (62,63). Genistein is a soy phytoestrogen that may mediate these risk reductions and was investigated for an impact on DNA methylation. Day et al. (64) used differential methylation hybridization arrays to investigate DNA methylation in mice on a genistein diet. Liver, brain, and prostate tissues were examined, but only changes in the prostate were detected using this technique. These included three novel sequences, ribosomal DNA, and a desmin-binding fragment. However, confirmation

of these changes with an additional technique was not performed, and, therefore, their relevance is unclear. Genistein supplementation did, however, increase methylation at the A^{vy} locus in the *agouti* mouse model offspring (65) in a fashion similar to methyl donor supplementation, as described previously.

Animal models have provided substantial evidence that fetal undernutrition, or indeed overnutrition, can influence DNA methylation patterns of an individual or their offspring. Are humans susceptible to the same types of changes in response to nutritional status? It is much more challenging to address this in humans.

Human studies

Although there is a wealth of literature supporting the role of nutrition in health and disease including the fetal origins of human disease concept (32), the identification of nutrient-sensitive MVPs in humans has only been a recent focus of attention. The Dutch Hunger Winter from 1944 to 1945 provided an excellent opportunity to investigate this further. Food restrictions were in place in the western part of the Netherlands toward the end of World War II. Despite this, medical records and registries were maintained, which allowed prenatally exposed individuals to be traced. Heijmans et al. (66) showed that prenatally exposed individuals displayed significantly less DNA methylation at the imprinted *IGF2* locus compared with their unexposed, same-sex siblings 60 y later. This appeared to be an early developmental effect because individuals exposed during late gestation showed no effect. The susceptibility of the *IGF2* locus to in utero exposure in humans correlates with similar findings in animal models already described. Additional loci for this cohort of samples were examined in a subsequent study that found that *INSIGF* showed decreased methylation, whereas *IL10*, *LEP*, *ABCA1*, *GNASAS*, and *MEG3* showed increased methylation in individuals exposed prenatally to the famine (67). Late gestational exposure showed DNA methylation changes in just *GNASAS* and *LEP* in men only. Thus, nutrition-induced changes appeared to be particularly relevant but not limited to the periconceptional period, showed both increased and decreased methylation, and displayed some sex specificity. Apart from the Dutch Hunger Winter samples, a number of both observational and intervention studies have been conducted in humans, focusing for the most part on methyl donors, particularly folate. We did not include those studies involving cancer tissues because epigenetic dysregulation is well documented as contributing to tumorigenesis and may cloud the role of nutrients in mediating DNA methylation changes. Friso et al. (68) reported a correlation between folate status and global DNA methylation levels in peripheral blood mononuclear DNA measured by HPLC/MS. Again, the relevance of measuring global DNA methylation without site-specific information is not clear, but the authors did report the interesting observation that individuals who were *MTHFR* 677 TT homozygous had a significantly lower level of global DNA methylation compared with 677 CC individuals. The *MTHFR* gene encodes a key enzyme in the folate metabolic pathway (**Fig. 2**), i.e., 5,10-methyltetrahydrofolate reductase. The *MTHFR* 677C > T polymorphism is a well-accepted, disease-associated variant that is thermolabile and converts 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate at a reduced rate, particularly when the folate status is low (69,70). Thus, the supply of methyl groups may be compromised in *MTHFR* 677 TT individuals, and this appears to correlate with global DNA methylation patterns. Additional observational studies focusing on folic acid include an examination of promoter methylation of estrogen receptor α and *MLH1* (71). This study found no correlation between serum and red cell folate and DNA methylation but did find a correlation between vitamin B-12 and estrogen receptor α methylation in colon tissue. Steegers-Theunissen et al. (72) found a 4.5% increase in the DNA methylation of *IGF2* DMR in the children of mothers who consumed 400 μg of folic acid periconceptionally compared with those who had not. Folic acid intervention studies of postmenopausal women showed decreased global DNA methylation in women consuming a low-folate diet measured by incorporation of a tritiated methyl group using SssI methyltransferase (73,74). However, one study reported recovery of DNA methylation in folate-replete subjects (73) but no change in another (74). Again, we question the validity of measuring global DNA methylation levels as opposed to a genome-wide analysis. A more recent report by Fryer et al. (75) applied the Illumina Infinium Methylation 27K BeadArray system to human cord blood samples in which folate and homocysteine levels were also measured. Homocysteine is inversely correlated with folate levels (**Fig. 2**). This allows analysis of 27,578 CpG loci associated with 14,496 genes and is a large step toward a genomewide analysis. Similar to previous findings [reviewed by Deaton and Bird (17)], the general pattern observed was that the majority of CpGs within CGIs were found to be hypomethylated, whereas those outside CGIs were methylated at a mid to high level. They specifically identified that DNA methylation of 12 genes (*EIF2C3*, *ZBTB11*, *BDH2*, *ZNF187*, *RUNX1T1*, *C9orf64*, *PDE2A*, *MGC33486*, *AMN*, *ZPBP2*, *FBN3*, *PVRL2*) directly correlated with homocysteine levels, whereas the methylation of 5 genes (*ATP5F1*, *CYP26C1*, *FSTL3*, *MDS032*, *BMX*) displayed an inverse correlation with homocysteine levels. Because cord blood samples were taken at term, it supports late gestation as also representing a window of opportunity for nutrition-induced DNA methylation changes. However, the methylation patterns observed by the Illumina Infinium Methylation 27K beadchip were not confirmed by another method, and, therefore, these genes require further validation. The limited number of human studies described here does implicate folate status and its associated metabolites in influencing DNA methylation patterns both in utero and in adults. Assessment of other nutrients has been limited. An intervention study of soy isoflavones in 34 healthy premenopausal women showed that the promoter region of *RARβ2* and *CCND2* in breast tissue showed hypermethylation post-treatment but no effect on *p16*, *RASSF1A*, or *ER* (76). The relevance of

TABLE 2. Summary of selected studies investigating nutrient-sensitive DNA methylation sites in humans[1]

Nutrient	Study type	DNA methylation site(s) examined	References
Calorie restriction	Maternal offspring	*IGF2*, *INSIGF* in adults	66, 67
		IL10, *LEP*, *ABCA1*, *GNASAS*, *MEG3*	67
Folate status	Adults/MTHFR 677C > T genotype	Global methylation	68
Folate, vitamin B-12 status	Adults: colon	*ERα*, *MLH1*	71
Folic acid consumption	Maternal offspring	*IGF2*	72
Folic acid intervention	Postmenopausal women	Global methylation	73,74
Folate/homocysteine status	Cord blood	14,496 genes	75
Soy isoflavone intervention	Premenopausal women	*RARβ2*, *CCND2*, *p16*, *RASSF1A*, *ER*	76

[1] MTHFR, methyltetrahydrofolate reductase.

gene specificity is again highlighted here. DNA methylation patterns and IUGR were considered using a microarray DNA methylation assay known as HELP (HpaII tiny fragment Enrichment by ligation Mediated PCR) (77). Although specific nutrients were not examined, nutrient status has been implicated as a contributing factor to IUGR (77). A modest change of 6% DNA methylation difference was observed consistently in *HNF4A*, *ATGS*, and *TADA3L* in the cord blood of IUGR neonates compared with controls. These genes were previously linked with type 2 diabetes.

The identification of nutrient-sensitive MVPs in humans is limited by sample availability and ethical issues. Despite this, a number of observational and intervention studies found DNA methylation changes in response to nutrient status, particularly in response to folate status (summarized in **Table 2**). The more relevant studies examined gene-specific DNA methylation patterns (66,67,71,72,76) or a narrow genomewide approach (75). Candidate methylation sites have emerged, but we are at the beginning in terms of deciphering the human methylome and identifying those nutrient-relevant DNA methylation sites.

Conclusions

Efforts in the past 10 y provide us with a glimpse of the complexity of the human methylome. We now know that the simple association of DNA methylation and repressed gene expression is not always the case. Mammalian DNA methylation can occur outside the CG context and within the coding regions of genes. Methylcytosine can also be converted to hydroxymethylcytosine and is recognized by a specific set of molecules that are just starting to be understood. Data just published (78) have finally begun to shed light on how demethylation of DNA occurs, i.e., by conversion of 5mC and 5hmC to 5-carboxylcytosine with subsequent processing by thymine-DNA glycosylase. Cataloging the methylomic differences between tissues and individuals has just begun, but what is clear is that nutrition has a distinct role to play in mediating DNA methylation changes during development or one's lifetime. Currently there is no comprehensive list of which sites are nutrient sensitive, but studies to date, some which are described here, provide potential candidates. These studies highlighted how it is not simply about methyl group supply because both supplementation and deficiency studies observe both increases and decreases in DNA methylation. The pattern that is emerging so far is that nutrient-mediated DNA methylation changes are gene specific, site specific, tissue specific, and age specific. In light of this, it is much more meaningful to examine specific, known regions of the genome or, indeed, a genomewide screen rather than measuring global DNA methylation patterns that have limited relevance. The technology is now available to do this, and the identification of nutrient-sensitive MVPs is likely to explode in the next few years.

Literature Cited

1. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo Q-M, et al. Human DNA methylomes at base resolution show widespread epigenominc differences. Nature. 2009;462:315–22.
2. Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science. 2009;324:929–30.
3. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009;324:930–5.
4. Choi S-W, Friso S. Epigenetics: A new bridge between nutrition and health. Adv Nutr. 2010;1:8–16.
5. Esteller M. Aberrant DNA methylation as a cancer-inducing mechanism. Annu Rev Pharmacol Toxicol. 2005;45:629–56.
6. Klose RJ, Bird AP. Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci. 2006;31:89–97.
7. Ball MP, Li JB, Gao Y, Lee J-H, LeProust EM, Park I-H, Xie B, Daley GQ, Church GM. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat Biotechnol. 2009;27:361–8.
8. Paulsen M, Ferguson-Smith AC. DNA methylation in genomic imprinting, development, and disease. J Pathol. 2001;195:97–110.
9. Urdinguio RG, Sanchez-Mut JV, Esteller M. Epigenetic mechanisms in neurological diseases: genes, syndromes and therapies. Lancet Neurol. 2009;8:1056–72.
10. Kim M, Long TI, Arakawa K, Wang R, Yu MC, Laird PW. DNA methylation as a biomarker for cardiovascular disease risk. PLoS ONE. 2010;5:e9692.
11. Richardson B. DNA methylation and autoimmune disease. Clin Immunol. 2003;109:72–9.
12. Dolinoy DC, Das R, Weidman JR, Jirtle RL. Metastable epialleles, imprinting, and the fetal origins of adult diseases. Pediatr Res. 2007;61:30R–7R.
13. Eckhardt F, Beck S, Gut IG, Berlin K. Future potential of the Human Epigenome Project. Expert Rev Mol Diagn. 2004;4:609–18.
14. Gonzalgo ML, Liang G, Spruck CH 3rd, Zing J-M, Rideout WM 3rd, Jones PA. Identification and characterization of differentially methylated regions of genomic DNA by methylation-sensitive arbitrarily primed PCR. Cancer Res. 1997;57:594–9.
15. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. Nature. 2010;467:1061–73. Erratum in: Nature. 2011;473:544.
16. Available athttp://www.roadmapepigenomics.org/

17. Deaton AM, Bird A. CpG islands and the regulation of transcription. Genes Dev. 2011;25:1010–22.
18. Penn NW, Suwalski R, O'Riley C, Boianowski K, Yura R. The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. Biochem J. 1972;126:781–90.
19. Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PA, Rappsilber J, Helin K. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature. 2011;473:343–8.
20. Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, McLoughlin EM, Brudno Y, Mahapatra S, Kapranov P, et al. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature. 2011;473:394–7.
21. Barker DJ, Winter PD, Osmond C, Margetts B, Simmonds SJ. Weight in infancy and death from ischaemic heart disease. Lancet. 1989;2:577–80.
22. Barker DJ. The fetal and infant origins of adult disease. BMJ. 1990;301:1111.
23. Barker DJ, Martyn CN. Then maternal and fetal origins of cardiovascular disease. J Epidemiol Community Health. 1992;46:8–11.
24. Barker DJ. Fetal nutrition and cardiovascular disease in later life. Br Med Bull. 1997;53:96–108.
25. Barker DJ. Fetal origins of coronary heart disease. Br Heart J. 1993;69:195–6.
26. Barker DJ. The fetal origins of type 2 diabetes mellitus. Ann Intern Med. 1999;130:322–4.
27. Hales CN, Barker DJ, Clark PM, Cox LJ, Fall C, Osmond C, Winter PD. Fetal and infant growth and impaired glucose tolerance at age 64. BMJ. 1991;303:1019–22.
28. Ravelli GP, Stein ZA, Susser MW. Obesity in young men after famine exposure in utero and early infancy. N Engl J Med. 1976;295:349–53.
29. Hoek HW, Susser E, Buck KA, Lumey LH, Lin SP, Gorman JM. Schizoid personality disorder after prenatal exposure to famine. Am J Psychiatry. 1996;153:1637–9.
30. Lopuhaä CE, Roseboom TJ, Osmod C, Barker DJ, Ravelli AC, Bleker OP, van Der Zee JS, van Der Meulen JH. Atopy, lung function, and obstructive airways disease after prenatal exposure to famine. Thorax. 2000;55:555–61.
31. Roseboom T, De Rooij S, Painter R. The Dutch famine and its long-term consequences for adult health. Early Hum Dev. 2006;82:485–91.
32. Gluckman PD, Hanson MA, Cooper C, Thornburg KL. Effect of in utero and early-life conditions on adult health and disease. N Engl J Med. 2008;359:61–73.
33. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suñer D, Cigudosa JC, Urioste M, Benitez J, et al. Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci U S A. 2005;102:10604–9.
34. Kaminsky ZA, Tang T, Wang SC, Ptak C, Oh GH, Wong AH, Feldcamp LA, Virtanen C, Halfvarson J, Tysk C, et al. DNA methylation profiles in monozygotic and dizygotic twins. Nat Genet. 2009;41:240–5.
35. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H, Campan M, Noushmehr H, Bell CG, Maxwell AP, et al. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. Genome Res. 2010;20:440–6.
36. Fernandez AF, Assenov Y, Martin-Subero JI, Balint B, Siebert R, Taniguchi H, Yamamoto H, Hidalgo M, Tan AC, Galm O, et al. A DNA methylation fingerprint of 1628 human samples. Genome Res.
37. Stover PJ, Field MS. Trafficking of intracellular folates. Adv. Nutr. 2011; 2:325–31.
38. Wolff GL, Kodell RL, Moore SR, Cooney CA. Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. FASEB J. 1998;12:949–57.
39. Waterland RA, Jirtle RL. Transposable elements: targets for early nutritional effects on epigenetic gene regulation. Mol Cell Biol. 2003;23:5293–300.
40. Cropley JE, Suter CM, Beckman KB, Martin DI. Germ-line epigenetic modification of the murine A^{vy} allele by nutritional supplementation. Proc Natl Acad Sci U S A. 2006;103:17308–12.
41. Dolinoy DC, Huang D, Jirtle RL. Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. Proc Natl Acad Sci U S A. 2007;104:13056–61.
42. Waterland RA, Dolinoy DC, Lin JR, Smith CA, Shi X, Tajiliani KG. Maternal methyl supplements increase offspring DNA methylation at Axin fused. Genesis. 2006;44:401–6.
43. Hollingsworth JW, Maruoka S, Boon K, Garantziotis S, Li Z, Tomfohr J, Bailey N, Potts EN, Whitehead G, Brass DM, et al. In utero supplementation with methyl donors enhances allergic airway disease in mice. J Clin Invest. 2008;118:3462–9.
44. Sinclair KD, Allegrucci C, Singh R, Gardner DS, Sebastian S, Bispham J, Thurston A, Huntley JF, Rees WD, Maloney CA, et al. DNA methylation, insulin resistance and blood pressure in offspring determined by maternal periconceptional B vitamin and methionine status. Proc Natl Acad Sci U S A. 2007;104:19351–6.
45. Pogribny IP, Poirier LA, James SJ. Differential sensitivity to loss of cytosine methyl groups within the hepatic p53 gene of folate/methyl deficient rats. Carcinogenesis. 1995;16:2863–7.
46. Waterland RA, Lin JR, Smith CA, Jirtle RL. Post-weaning diet affects genomic imprinting at the insulin-like growth factor 2 (Igf2) locus. Hum Mol Genet. 2006;15:705–16.
47. McKay JA, Xie L, Harris S, Wong YK, Ford D, Mathers JC. Blood as a surrogate marker for tissue-specific DNA methylation and changes due to folate depletion in post-partum female mice. Mol Nutr Food Res. 2011;55:1026–35.
48. Pogribny IP, Karpf AR, James SR, Melnyk S, Han T, Tryndyak VP. Epigenetic alterations in the brains of Fisher 344 rats induced by long-term administration of folate/methyl-deficient diet. Brain Res. 2008;1237:25–34.
49. Niculescu MD, Craciunescu CN, Zeisel SH. Dietary choline deficiency alters global and gene-specific DNA methylation in the developing hippocampus of mouse fetal brains. FASEB J. 2006;20:43–9.
50. Kovacheva VP, Mellott TJ, Davison JM, Wagner N, Lopez-Coviella I, Schnitzler AC, Blusztajn JK. Gestational choline deficiency causes global and Igf2 gene DNA hypermethylation by up-regulation of Dnmt1 expression. J Biol Chem. 2007;282:31777–88.
51. Zeisel SH. Epigenetic mechanisms for nutrition determinants of later health outcomes. Am J Clin Nutr. 2009;89:1488S–93S.
52. Lillycrop KA, Phillips ES, Jackson AA, Hanson MA, Burdge GC. Dietary protein restriction of pregnant rats induces and folic acid supplementation prevents epigenetic modification of hepatic gene expression in the offspring. J Nutr. 2005;135:1382–6.
53. Burdge GC, Slater-Jefferies J, Torrens C, Phillips ES, Hanson MA, Lillycrop KA. Dietary protein restriction of pregnant rats in the F0 generation induces altered methylation of hepatic gene promoters in the adult male offspring in the F1 and F2 generations. Br J Nutr. 2007;97:435–9.
54. van Straten EM, Bloks VW, Huijkman NC, Baller JF, Meer H, Lutjohann D, Kuipers F, Plosch T. The liver X-receptor gene promoter is hypermethylated in a mouse model of prenatal protein restriction. Am J Physiol Regul Integr Comp Physiol. 2010;298:R275–82.
55. Doherty AS, Mann MR, Tremblay KD, Bartolomei MS, Schultz RM. Differential effects of culture on imprinted H19 expression in the preimplantation mouse embryo. Biol Reprod. 2000;62:1526–35.
56. Kaati G, Bygre LO, Edvinsson S. Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period. Eur J Hum Genet. 2002;10:682–8.
57. Ng SF, Lin RC, Laybutt DR, Barres R, Owens JA, Morris MJ. Chronic high-fat diet in father programs β-cell dysfunction in female rat offspring. Nature. 2010;467:963–6.
58. Carone BR, Fauguier L, Habib N, Shea JM, Hart CE, Li R, Bock C, Li C, Gu H, Zamore PD, et al. Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. Cell. 2010;143:1084–96.
59. Milagro FI, Campión J, García-Díaz DF, Goyenechea E, Paternain L, Martínez JA. High fat diet-induced obesity modifies the methylation pattern of leptin promoter in rats. J Physiol Biochem. 2009;65:1–9.
60. Jousse C, Parry L, Lambert-Langlais S, Maurin AC, Averous J, Bruhat A, Carraro V, Tost J, Letteron P, Chen P, et al. Perinatal undernutrition affects the methylation and expression of the leptin gene in adults: implication for the understanding of metabolic syndrome. FASEB J.
61. Gluckman PD, Lillycrop KA, Vickers MH, Pleasants AB, Phillips ES, Beedle AS, Burdge GC, Hanson MA. Metabolic plasticity during mammalian development is directionally dependent on early nutritional status. Proc Natl Acad Sci U S A. 2007;104:12796–800.

62. Adlercreutz H. Phytoestrogens: epidemiology and a possible role in cancer protection. Environ Health Perspect. 1995;103:103–12.
63. Fotsis T, Pepper M, Adlercreutz H, Fleischmann G, Hase T, Montesano R, Schweigerer L. Genistein, a dietary-derived inhibitor of in vitro angiogenesis. Proc Natl Acad Sci U S A. 1993;90:2690–4.
64. Day JK, Bauer AM, Desbordes C, Zhuang Y, Kim BE, Newton LG, Nehra V, Forsee KM, MacDonald RS, Besch-Williford C, et al. Genistein alters methylation patterns in mice. J Nutr. 2002;132:2419S–23S.
65. Dolinoy DC, Weidman JR, Waterland RA, Jirtle RL. Maternal genistein alters coat color and protects Avy mouse offspring from obesity by modifying the fetal epigenome. Environ Health Perspect. 2006;114:567–72.
66. Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES, Slagboom PE, Lumey LH. Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proc Natl Acad Sci U S A. 2008;105:17046–9.
67. Tobi EW, Lumey LH, Talens RP, Kremer D, Putter H, Stein AD, Slagboom PE, Heijmans BT. DNA methylation differences after exposure to prenatal famine are common and timing-and sex-specific. Hum Mol Genet. 2009;18:4046–53.
68. Friso S, Choi SW, Girelli D, Mason JB, Dolnikowski GG, Bagley PJ, Olivieri O, Jaques PF, Rosenberg IH, Corrocher R, et al. A common mutation in the 5,10-methylenetetrahydrofolate reductase gene affects genomic DNA methylation through interaction with folate status. Proc Natl Acad Sci U S A. 2002;99:5606–11.
69. Frosst P, Blom HJ, Milos R, Goyette P, Sheppard CA, Matthews RG, Boers GJH, den Heijer M, Kluijtmans LAJ, van den Heuvel LP, et al. A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. Nat Genet. 1995;10:111–3.
70. Yamada K, Chen Z, Rozen R, Matthews RG. Effects of common polymorphisms on the properties of recombinant human methylenetetrahydrofolate reductase. Proc Natl Acad Sci U S A. 2001;98:14853–8.
71. Al-Ghnaniem R, Peters J, Foresti R, Heaton N, Pufulete M. Methylation of estrogen receptor alpha and mutL homolog 1 in normal colonic mucosa: association with folate and vitamin B-12 status in subjects with and without colorectal neoplasia. Am J Clin Nutr. 2007;86:1064–72.
72. Steegers-Theunissen RP, Obermann-Borst SA, Kremer D, Lindemans J, Siebel C, Steegers EA, Slagboom PE, Heijmans BT. Periconceptional maternal folic acid use of 400 microg per day is related to increased methylation of the IGF2 gene in the very young child. PLoS ONE. 2009;4:e784.
73. Jacob RA, Gretz DM, Taylor PC, James SJ, Pogribny IP, Miller BJ, Henning SM, Swendseid ME. Moderate folate depletion increases plasma homocysteine and decreases lymphocyte DNA methylation in postmenopausal women. J Nutr. 1998;128:1204–12.
74. Rampersaud GC, Kauwell GP, Hutson AD, Cerda JJ, Bailey LB. Genomic DNA methylation decreases in response to moderate folate depletion in elderly women. Am J Clin Nutr. 2000;72:998–1003.
75. Fryer AA, Emes RD. Ismail Km, Haworth KE, Mein C, Carroll WD, Farrell WE. Quantitative high-resolution epigenetic profiling of CpG loci identifies associations with cord blood plasma homocysteine and birth weight in humans. Epigenetics. 2011;6:86–94.
76. Qin W, Zhu W, Shi H, Hewett JE, Ruhlen RL, MacDonald RS, Rottinghaus GE, Chen YC, Sauter ER. Soy isoflavones have an antiestrogenic effect and alter mammary promoter hypermethylation in healthy premenopausal women. Nutr Cancer. 2009;61:238–44.
77. Einstein F, Thompson RF, Bhahat TD, Fazzari MJ, Verma A, Barzilai N, Greally JM. Cytosine methylation dysregulation in neonates following intrauterine growth restriction. PLoS ONE. 2010;5:e8887.
78. He YF, Li BZ, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science.