CANCER-TESTIS GENE EXPRESSION AS A BIOMARKER OF THE GENETIC VARIATION IN THE ONE CARBON METABOLIC PATHWAY

A THESIS SUBMITTED TO
THE DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS
AND THE INSTITUTE OF ENGINEERING AND SCIENCES OF
BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE

BY
AHMET RASİM BARUTÇU
AUGUST 2008

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Ali Osmay Güre

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. Bensu KARAHALİL

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Zeynep KALAYLIOĞLU-WHEELER

Approved for the Institute of Engineering and Science

Director of Institute of Engineering and Science

Prof. Dr. Mehmet Baray

CANCER-TESTIS GENE EXPRESSION AS A BIOMARKER OF THE GENETIC VARIATION IN THE ONE CARBON METABOLIC PATHWAY

Ahmet Rasim Barutçu
MSc. in Molecular Biology and Genetics
Supervisor: Assist. Prof. Dr. Ali Osmay Güre
August 2008, 87 Pages

ABSTRACT

S-adenosyl methionine (SAM) is the sole methyl donor for all biological reactions in humans. Folate consumption, required for SAM generation, is also essential for dTMP synthesis and both events occur via enzymes of the one-carbon pathway. Frequently occurring alleles of these enzymes have occasionally been associated with several diseases including cancer. However, the cumulative effects of the polymorphic variants of these enzymes on S-adenosylmethionine production have not been studied. The identification of a biomarker that can reflect the collective effect of these allelic variants is critical in moving the field forwards. We hypothesized that Cancer-Testis (CT) genes, whose expression strongly correlates with DNA hypomethylation, could be such a biomarker. In this study, we have pursued an extensive correlation of CT expression and allelic variants of the several one-carbon pathway enzyme genes , including methyltetrahydrofolatereductase (MTHFR), methionine synthase (MS), reduced folate carrier (RFC) and methionine synthase reductase (MTRR) in non-small cell lung cancer. Our results revealed linkage disequilibrium among alleles as well as correlations between given genootypes and CT gene expression, and illuminate the critical next steps that need to be pursued.

BİR KARBON METABOLİK YOLUNDAKİ GENETİK FARKLILIKLARIN BELİRLEYİCİSİ OLARAK KANSER-TESTİS GEN İFADESİ

Ahmet Rasim Barutçu

Moleküler Biyoloji ve Genetik Yüksek Lisansı

Tez Yöneticisi: Yard. Prof. Dr. Ali Osmay Güre

Ağustos 2008, 87 Sayfa

ÖZET

İnsanlarda bütün biyolojik tepkimelerde kullanılan tek metil grubu vericisi S-adenozilmetiyonin (SAM)'dir. SAM üretimi için gereken folik asit tüketimi, deoksitimidin monofosfat (dTMP) sentezi için gereklidir ve her iki olay da bir karbon metabolik yolundaki enzimler sayesinde gerçekleşir. Bu enzimlerde sıkça görülen aleller kanser dâhil birçok hastalık ile ilişkilendirilmiştir. Ancak, bu enzimlerde görülen polimorfizmlerin SAM üretimine olan kümülatif etkisi şu ana kadar çalışılmamıştır. Bir karbon metabolik yolundaki enzim alellerinin kümülatif etkisine duyarlı ve buna cevap verebilecek biyolojik bir belirleyicinin teşhisi, bu alanda ilerleme kaydedilmesi için önemlidir. Hipotezimize göre transkripsiyonları DNA demtilasyonu ile ilişkili olan Kanser-Testis (KT) genleri, böyle bir biyolojik belirteç olabilir. Bu çalışmada, küçük hücre dışı akciğer kanserinde (KHDAK) KT gen ifadesi ile bir karbon metabolik yolundaki, metilentetrahidrofolat redüktaz (MTHFR), metiyonin sentaz (MS), indirgenmiş folat taşıyıcısı (RFC) ve metiyonin sentaz redüktaz (MTRR) enzim alelleri ilişkiyi araştırdık. Sonuçlarımız, enzim alelleri arasında linkage disequilibrium olduğunu göstermiş, ayrıca belli bir haplotip ile CT gen ifadesi arasındaki ilişkiyi açığa çıkarmış ve izlenecek adımlarda önemli noktaları aydınlatmıştır.

TO MY FAMILY

ACKNOWLEDGEMENTS

First of all, I would like to thank Assist. Prof. Dr. Ali Osmay Güre for his endless support, supervision and guidance. I am grateful for his patience and this thesis could not have been prepared without him. I feel very lucky and honored to work with Dr. Ali Güre for I believe I have gained the maximum experience in my master study by working with him.

We would never be able to finalize my results without the help of Assist. Prof. Dr. Zeynep Kalaylıoğlu-Wheeler. I would like to thank her for the time and effort she has dedicated. She performed critical statistical analyses without which interpretation of some of our results might have been impossible. She has been a genuine support for my study.

I would also like to thank Prof. Dr. Bensu Karahalil for her valuable knowledge and guidance. She graciously made precious contributions to my study.

I was delighted to have such a wonderful and amiable team of colleagues. They were not only colleagues, but marvelous friends in and out of the lab. I will never forget our monthly and weekly "meetings" and "feasts". I would like thank Aydan Bulut, Duygu Akbaş, Şükrü Atakan, Derya Dönertaş, Esen Oktay and Ender Avcı for their support. I do not know if I will have such a laboratory team in my academic life. I would also like to thank the MBG family for creating such a laboratory environment.

Last but never the least; I would like to thank my father for raising such an individual and I wish he could see me making this work come true. I would like to thank my dear family and my companion for their never-ending love and support.

During this the course of this study, towards my Master's degree, I received a Scholarship 2210 from TÜBİTAK.

Table of contents

ABSTRACT…….....III

ÖZET…….……..IV

DEDICATION PAGE……..…….V

List of tables

List of Figures

Abbreviations

DHF - Dihydrofolate

MTA- Methylthioadenosine

MTHFR- Methylenetetrahydrofolate reductase

MTR – Methionine synthase

MTRR - Methionine synthase reductase

RFC – Reduced Folate Carrier

RFLP – Restriction Fragment Length Polymorphism

SAM – S-adenosyl methionine

SAH – S-adenosyl homocysteine

SNP – Single nucleotide polymorphism

THF - Tetrahydrofolate

TS – Thymidylate synthase

CHAPTER1. INTRODUCTION

1.1 DNA methylation

The inheritance of information which is not based on DNA sequence is known as epigenetics [1]. DNA methylation is a crucial determinant in gene expression, DNA stability and chromatin modifications. The vitality of DNA methylation for vertebrates has not only been shown in embryonic lethality of DNA methyl transferase-1 (DNMT1) knockout mice, but also by its deregulation in various diseases, especially cancer [2]. DNA methylation is thought to be evolved to silence viral sequences and transposable elements as well as to regulate gene transcription [3]. DNA methylation is a heritable, tissue and cell specific modification of cytosine residues in CpG sequences. The methyl group can be attached at the N^4 or C^5 positions of the cytosine residues of prokaryotic or eukaryotic genomic DNA [4]. The distribution of the CpG residues in the genome is not equal, some regions such as the Alu repeats and transcription start sites contain a higher frequency of CpG residues than other regions [84].

Most of the mammalian genome consists of extended regions that are deficient for CpG's. CpG residues are at 20% of their predicted frequency in the mammalian genome. This is thought to be a direct consequence of an increased mutation rate of the 5-methylcytosine residues found in CpG sequences from cytosine to thymine. Most methylated CpG's are found in clusters, called a CpG island which corresponds to 1% of the human genome. The term CpG island refers to a 500 base pair window with an increased G:C content of at least 55%, and an observed to expected CpG frequency of at least 0.65. These islands may span the 5' regions of approximately half of the human genes including exons, promoters and untranslated regions.

Up to 80% of all cytosine residues, which are not related to CpG islands, are normally methylated. In contrast, the CpG residues in CpG islands, especially at the gene promoter regions of actively transcribed genes, are usually unmethylated. The major genomic regions which have methylated CpG islands are the inactive X chromosome in female and silenced alleles of parentally imprinted genes [5]. The CpG methylation also occurs at the sites with a low frequency of CpG's such as repeat DNA sites, heterochromatin, telomeres, non-coding regions and exons. Methylation of the "bulk" of the genome enables the silencing of these non-coding

regions, which prevents the transcription of repeat elements and parasitic DNA sequences. DNA methylation has also attracted a considerable interest in the cell differentiation and tissue specific gene expression process [6].

DNA methylation is a dynamic process in which involves *DNA methyltransferases* (Dnmt1, Dnmt3a and Dnmt3b), methyl-binding proteins, histone modifying enzymes, chromatin remodeling factors and their molecular complexes [84]. It has been previously shown that cyclical and dynamic methylation/demetylation of the CpG residues in a model containing the *ps2* gene promoter occurs with the presence of MeCP2 (a methyl-binding protein), SWI/SNF (a histone remodeling complex), DNMT1 and DNMT3a and DNMT3b [7]. DNMT1 is responsible for the maintenance methylation of the genome after each round of replication. During the replication of eukaryotic genomic DNA, approximately 40 million CpG dinucleotides are converted into the hemimethylated state in the newly synthesized DNA strand. These hemimethylated CpG sites must be methylated precisely to maintain the original DNA methylation pattern. DNMT1 is located at the replication fork and methylates newly biosynthesized DNA strands directly after the replication. DNMT3a and DNMT3b methyltransferases methylate CpG dinucleotides without preference for hemimethylated DNA, and are responsible for the *de novo* methylation of DNA [8].

1.2 DNA Methylation and Cancer

For over a decade, abnormalities of DNA methylation in cancer cells have been recognized [9]. In cancer, the usual pattern of methylation is observed to be genomic hypomethylation and gene-specific hypermethylation. Hypermethylation, which is also related to transcriptional silencing, is a common mechanism for the inactivation of several tumor suppressor genes in cancer [10]. Methylations of such genes seem to occur early in carcinogenesis, and in some cases increasing progressively leading to malignant phenotypes [11]. Hypermethylation occurs not only at the promoter of tumor suppressor genes, but also at the promoter regions of other genes involved in cancer progression such as DNA repair, cell cycle regulation, apoptosis, hormonal response and cell adherence genes [12]. Previously, it has

been suggested that a tumor-type specific profile of DNA hypermethylation exists and thus allows us to use these hypermethylated loci as biomarkers of tumorigenesis [13], [11].

On the other hand, global DNA hypomethylation is also seen in early carcinogenesis [14]. It has been previously shown that DNA hypomethylation causes genomic instability, thus resulting in increased levels of DNA damage, mutation rates, copy number alterationd and loss of heterozygosity. Chen *et al* have shown that hypomethylation in murine embryonic stem cells lacking Dnmt1, at the same time significantly elevating mutation rates in two model genes [15]. It has been suggested that there is a direct correlation between the methylation capacity of the cell and genetic instability [16]. Abnormal DNA methylation has been associated with tumor aggressiveness and poor prognosis [17]. Precancerous cells showing aberrant DNA methylation profiles designate increased malignancy [18]. The reason for the instability of genomic DNA observed along DNAhypomethylation might be the lack of a splice variant of Dnmt3b which results in chromosomal instability through hypomethylation of pericentromeric satellite regions [19]. In addition, Dnmt1 over-expression observed in cancer cells has been observed to associate with increased CpG island methylation in a cancer-specific manner [19].

In normal cells, DNA methylation allows the condensation of the chromatin structure through the recruitment of the chromatin-organizing proteins such as histone remodeling complexes or polycomb group proteins [20]. In a DNA-hypomethylated cell, this arrangement is lost which results in chromatin decondensation and chromosomal rearrangements [21].

1.3 Cancer Testis (CT) Genes

CT genes are a group of genes which are consistently expressed in spermatogonia, oogonia and trophoblast cells, but not expressed in any other healthy tissues except cancer cells [22]. CT genes can be categorized as families based on sequence similary. Some families such as MAGE-A and SSX contain more than 8 members, while others like NY-ESO-1 consist of only two. As will be explained below, despite the dissimilarity of their sequences, there is overwhelming evidence that tumor-specific aberrant CT gene expression occurs in a coordinate fashion. More than half of 40 CT genes, so far identified, are located on the X chromosome [22],[23]. CT gene promoters (NY-ESO-1, MAGE-A1, SSX4 and SSX 7) lack a TATA box

[24], {unpublished data}. An interesting aspect of CT genes is that most of them are thought to have risen due to duplications [23] and thus, they are located juxtameric to each other forming gene clusters.

1.3.1 CT Gene Expression Regulation

Since CT genes are diverse, sequence-wise, it can be envisioned that they are possibly regulated via unique transcription factors. Yet, the fact that all CT genes, studied to date, are co-expressed, suggests the presence of a general mechanism behind their expression in cancer. DNA methylation is thought to be the major control mechanism for the expression and silencing of CT genes. This assumption was previously verified by a detailed analysis of the CpG islands of a model CT gene, MAGEA1 [24]. It has been shown that the critical CpG islands at the promoter region remain significantly methylated *in vivo* where the gene is not expressed [24]. Similary, expression of all other CT genes, studied to date, correlates with the demethylation of their promoter-proximal DNA regions [24](our unpublished data).

An example to such an experiment is detailed below SSX expression correlates with its methylation status. The southern blot figure below shows the change in expression of SSX genes, along with their methylation status (Figure 1). The general demethylation observed for the SSX gene region seems to occur in parallel to L1 repeat demetylation, which in turn, is considered to reflect the general methylation state of the genome.

The probes used in the assay were SSX 1, 2, 3, 4 and 5 (a mixture of SSX genes) and L1. SSX probe recognizes all of the SSX exons whereas L1 probe recognizes all of the L1 repeat sites. In this experiment, the genomic DNA from different cancer cell lines was cleaved by *MspI* and *HspII* enzymes. Then, the digested products were separated by agarose gel electrophoresis. Following the transfer of the DNA to the nitrocellulose membrane, the membrane wass exposed to a hybridization probe, a single DNA fragment with a specific sequence whose presence in the target DNA is to be determined. Then by autoradiography, the pattern of hybridization was visualized.

Both *MspI* and *HspII* cut at CCGG, yet *MspI* enzyme cuts DNA regardless of the methylation status while *HspII* cuts only the unmethylated DNA. Thus, the methylated SSX or L1 sequence will not be digested by *HspII*, as seen for the Calu-3 cell line. On the other hand, the unmethylated SSX or L1 sequences will be digested by *HspII* and, as observed for the SK-LC17 cell line. . In fact, for this cell line, the digestion pattern observed for the two enzymes is identical suggesting complete demetylation of SSX gene proximal CCGG sites. Below the southern blot figure, semi-quantitative PCR results show that the expression of SSX genes increase in parallel to the decrease of methylation within the SSX , as well as in L1 genomic DNA.

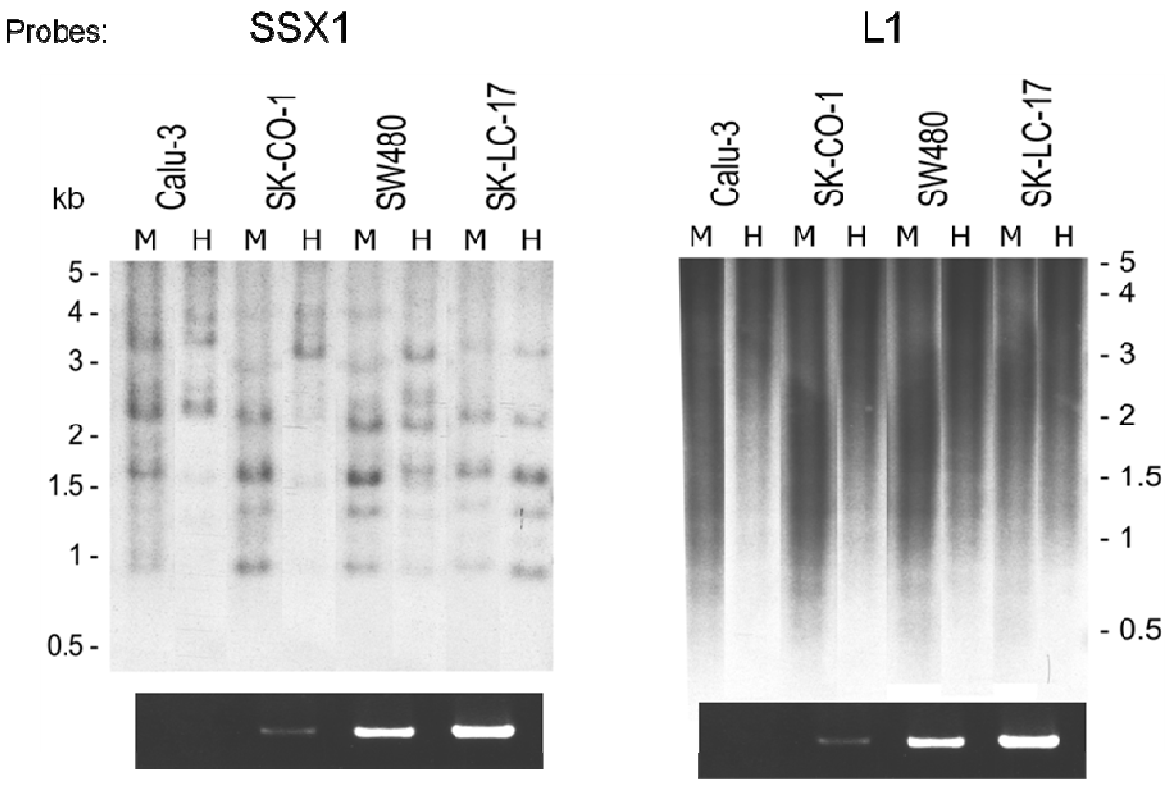


Figure 1. Southern blot and PCR showing the expression levels along with their methylation status of SSX genes and L1 repeats. Probes: SSX1 and L1. M: *Msp1*, H: *HspII*

Thus, there are many examples demonstrating a clear relationship between hypomethylation of genomic DNA and CT gene induction in cancer. CT genes can be upregulated by DNA methyl transferase inhibitors [25] and histone deacetylase inhibitors as well [22]. In 2005, Gure *et al* reported that in NSCLC (non-small cell lung cancer), CT gene expression is frequently observed. In the same study, CT genes were found to be coordinately expressed. Among 9 CT genes studied, expression of every single CT was found to correlate significantly with any other [26].

Moreover, CT genes were found to be significantly associated with less differentiated, higher grade of tumors, later stages of cancer and worse outcome. Larger tumor size and invasion capacity are also correlated with CT expression [26]. It was also observed that CT

expression increases as the tumors progress [26]. Therefore, one can presume that the hypomethylation observed in cancer cells causes demethylation in the promoter regions of CT genes, thus increasing their expression. . Most importantly, in the same study, overall survival of patients without (or with low levels of) CT gene expression was found to be significantly better than those with high level CT gene expression in their tumors. Thus, given this relationship, if the CT expression levels of tumors can be decreased, it can be hypothesized that this would result in improved survival. In this study, I have addressed this question with reference to the one-carbon pathway.

1.4 S-Adenosylmethionine: The universal methyl donor

S-adenosylmethionine (SAM) is a coenzyme involved in methyl group transfers. It is synthesized from ATP and methionine by *methionine adenosyltransferase*. In this reaction, the tri-phosphate group is cleaved from ATP and methionine is covalently attached. Figure 2 shows the chemical structure of SAM.

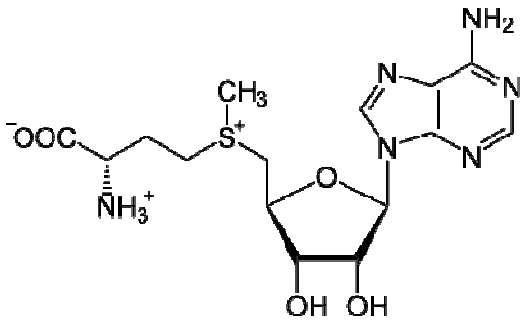


Figure 2.The chemical structure of S-adenosylmethionine. The methyl group attached to the sulfur atom is reactive and is donated to an acceptor, forming S-adenosylhomocysteine., adapted from **[27]**.

The methyl group attached to the sulfur atom is chemically reactive and thus allows the donation of this methyl group to an acceptor substrate in the methylation reactions. The product

following these transmethylation reactions, is *S-adenosylhomocysteine* (SAH) is formed by the demethylation of SAM.

Since SAM is the only methyl donor in the cell, the efficiency of SAM production might possibly effect methylation reactions including DNA and histone methylation. A recent study showed that rats fed with a methyl-deficient diet displayed significantly reduced levels of SAM, reduced ratios of SAM to SAH (SAM's more stable metabolite) and that this correlated with elevated levels of DNA hypomethylation [28]. Since CT gene expression is controlled by DNA methylation, a deficiency in SAM production could in turn be crucial in the regulation of CT genes. The low production of SAM might result in DNA hypomethylation, causing CT up-regulation and poor prognosis in cancer. For this reason, we have focused on the one carbon pathway in order to investigate the possible mechanisms which consequently results in reduced SAM production and DNA hypomethylation.

1.5 One carbon pathway

One carbon metabolism is an intervention of two pathways enabling the cross-talk between epigenetic and genetic processes, which involves DNA methylation and DNA synthesis. One carbon metabolism is vitally important for the maintenance of methionine cycle, nucleotide synthesis and biological methylation reactions. Folate is the most important input substrate utilized in the one carbon pathway. Folate is not only an essential co-factor for the de novo biosynthesis of purines and pyrimidines, it also plays an important role in DNA synthesis, stability and integrity [29]. Figure 3 summarizes those particular events in the one carbon pathway which we chose to focus on.

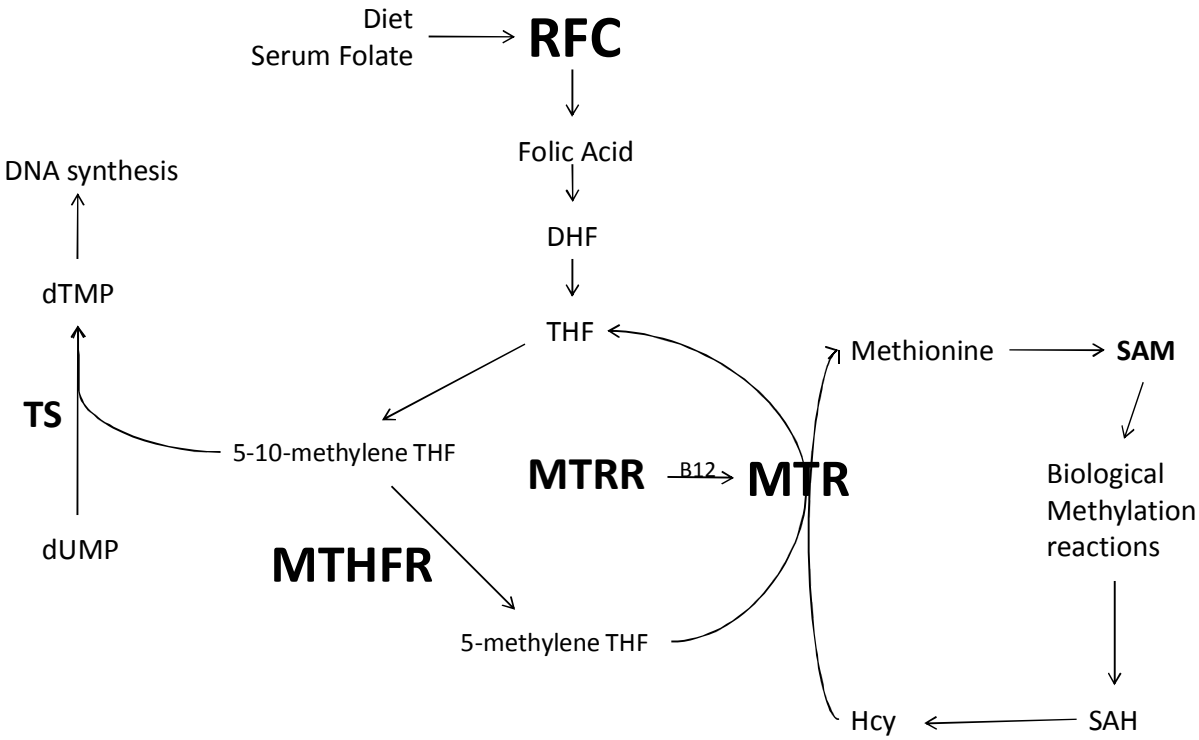


Figure 3. The one carbon pathway. Annotations: MTHFR- Methylenetetrahydrofolate reductase, MTRR- Methionine synthase reductase, MTR- Methionine synthase, RFC- Reduced Folate Carrier, TS- Thymidylate synthase, DHF- Dihydrofolate, THF-Tetrahydrofolate, SAM- S-adenosylmethionine, SAH- S-adenosylhomocysteine, Hcy- homocysteine

Folate may be gathered from two sources; either from foods or from blood by serum folate. Because folate cannot pass through the cell membrane when its glutamate tail is longer than 3, it is absorbed in the small intestine after the hydrolysis of polyglutamate chain by glutamate carboxypeptidase II (GCPII) [30]. Two receptors can transport folate from blood to into the cells. The *folate receptor* (FR), and the *reduced folate carrier* (RFC). The FR has a higher affinity for oxidized folate when compared to RFC. When taken into the cell, folate is first converted to dihydrofolate, and then to tetrahydrofolate (THF) by the enzyme *dihydrofolate reductase* (DHFR). The following reaction generates the origin of one carbon units by the breakage of β-carbon of

serine. In this reaction catalayzed by *serine hydroxymethyltransferase* (SHMT), THF is converted to N^5-N^{10}-methylene-THF while glycine is produced. N^5-N^{10}-methylene-THF is a crucial intermediate factor for the direction of the pathway [30].

For thymidine synthesis, deoxythymidylate monophosphate (dTMP) is synthesized from deoxyuridylate monophophate (dUMP) by *thymidylate synthase* (TS). TS transfers a methyl group from N^5-N^{10}-methylene-tetrahydrofolate. The non-reversible methylation of dUMP to dTMP results in the oxidation of N^5-N^{10}-methylene-THF to the inactive dihydrofolate, which can be converted back to THF by DHFR [31].

For *de novo* methionine biosynthesis and methylation reactions, *methylene tetrahydrofolate reductase* (MTHFR) converts N^5-N^{10}-methylene-THF to 5-methyl-THF. This step creates the only source of 5-methyl-THF in the one carbon pathway. *Methionine synthase* (MTR or MS), a cobalamin dependent enzyme which is activated by *methionine synthase reductase* (MTRR or MSR), transfers one methyl group from 5-methyl-THF in order to convert homocysteine (Hcy) into methionine. Dimethyl glycine is also produced from choline and betaine for methionine generation. Methionine synthesis ensures the provision of the universal methyl *donor* S-adenosyl methionine (SAM), with the activation of methionine by *methionine adenosyl transferase* and ATP. SAM is used to methylate more than 80 important biomolecules such as DNA, RNA, and proteins including histones. During this process, SAM is converted into S-adenosyl-homocysteine (SAH), which is either hydrolyzed to homocysteine (Hcy) to initiate a new remethylation cycle or transsulphurated to cysteine by *cystathionine β-synthase* [30].

In a cell, every methylation process requires a methyl group transfer from SAM. All of the metabolic methylation reactions are under the effect of SAM, which directs the utilization of Hcy and indicates the level of methylation. As explained before, the SAM/SAH ratio in a cell determines the cellular methylation potential. SAH is known to be an inhibitor of methyltransferases [23]. The level of SAM in the cell is regulated by two mechanisms. First of all, SAM suppresses the synthesis of N^5-methylene-THF by inhibiting the MTHFR enzyme. Thus, methionine level and therefore Hcy level decreases when SAM increases. On the other hand, homocysteine can also be converted to methionine through *SAH hydrolase*. Therefore, N^5-

methylene-THF is an important molecule determining the pathway between remethylation and transsulphuration reactions. If there is sufficient N^5-methylene-THF entering the methylation pathway, the direction of the SAH favors cysteine synthesis; if not, SAH is used to produce methionine [32]. The source of N^5-methylene-THF is folate acquired by diet. Folate, as well as B_{12} and B_6 vitamins, which function as co-enzymes for MTR and cysthation synthase, respectively, regulate the homocysteine removal and in turn, prevent SAH accumulation [33].

1.6 Folate deficiency and DNA damage

There is accumulating evidence that folate deficiency, either due to low dietary folate intake or due to low blood folate levels, is related to tumorigenesis [1]. Perturbations of the one carbon metabolism are thus, thought to play vital roles in neoplastic development by affecting gene regulation through DNA methylation and genome integrity through DNA synthesis and repair [34].Folate deficiency has been shown to induce hepatocellular carcinoma development in rats [1]. Diets deficient in various methyl donor groups (folic acid, choline, methionine and vitamin B_{12}) have been shown to induce DNA hypomethylation, site-specific DNA hypermethylation, double strand breaks, and upregulation of DNMT's [1].

Low dietary folate intake is strongly associated with DNA damage through uracil misincorporation or by DNA hypomethylation leading to genome instability [35]. Decreased cytosolic levels of N^5-N^{10}-methylene-THF caused by low folate status decreases the dTMP synthesis and increase the dUMP/dTMP ratio [35]. The increase of uracil pool in the cell elevates the rate of DNA-polymerase-mediated dUTP misincorporation into the DNA [36]. Uracil is excised from DNA by uracil-DNA glycosylase and apyrimidinic endonuclease, generating nicks which will be ligated by DNA ligase. However, if these nicks (single strand breaks) are located on two opposite strands, less repairable and hazardous double strand breaks may form. Uracil mediated double strand breakage is the major cause of deletions, duplications, chromosome breaks, micronucleus formation, chromatid recombinations, fragile sites and translocations which play important roles in tumorigenesis [35]. In cell culture, folate depletion can increase the

intracellular dUMP/dTMP ratio up to 10 fold. Folate depletion induced DNA breakage has been observed in lymphocytes and rodent cell lines in vitro [37], [38].

Decreasing folate levels can cause upregulation of DNA damage and cell cycle checkpoint related genes such as p53, p16 and p21 as well [37]. The normal levels of dNTPs for normal DNA/RNA synthesis are directly dependent on intracellular folate availability. Maintenance of folate concentration from diet or supplements is correlated with a protective effect and reduced incidence of number of cancers [39].

On the other hand, genomic instability caused by global hypomethylation is characterized by an elevated rate of DNA sequence changes, aneuploidy, chromosome translocations and gene amplifications [40]. A balance between the two critical reactions (methylation and DNA synthesis/repair) must be reached so one does not compromise the other [41].

1.7 One carbon pathway enzyme variants

Table 1 summarizes studies where polymorphic variants of the one carbon pathway enzyme activities were investigated utilizing various assays. From the enzymatic activity data obtained from the literature, one can infer that the variations of the enzymes partaking in SAM generation can create disturbances in the methylation process.

1.7.1 Methylenetetrahydrofolate reductase (MTHFR)

All four enzymes which I focused on in this study have allelic variants in the population and are associated with cancer**.** Different studies have focused on each of these alleles individually, despite the possibility that the net cumulative effect of various alleles is what ultimately determines SAM production efficiency. A large number of allelic variants of the one carbon pathway enzymes described. For example, there are 65 SNPs found on the MTHFR. However, among these, only 10 are gene non-synonymous SNPs that are found in more than

10% of the population are only 2 for MTHFR [42]. For some of these alleles, an association with decreased enzymatic activity has been described [42], (Table 1). Cancer risk generally has been linked to such hypomorphic alleles of one carbon pathway enzymes [33]. The major variant of the MTHFR gene, C677T transition (rs1801133), resulting in an Ala to Val change in aa.222, was found to decrease the activity of the MTHFR enzyme by %70 [43], (Table 1). This common variant of the MTHFR gene is associated with an increased risk of cancer [44]. MTHFR 677 TT homozygosity causes a significant increase in the homocysteine levels and DNA hypomethylation when compared to the wild-type allele [45].

Apart from the C677T transition, A1298C transition (rs1801131), converting Glu 429 to Ala, of the MTHFR enzyme causes a 15% reduction in the activity and results in increased levels of plasma homocysteine levels [45], **Table 1**. The impaired activity of MTHFR, especially along with the low folate status, results in less production of N^5-methylene-THF and methionine, thus proceeds to hypomethylation [45]. In another study, carriers of A1298C polymorphism do not appear to have high levels of plasma Hcy **(Table 1)** but have a lesser degree of MTHFR activity reduction compared to C677T variants, when tested by biochemical assays [46]. However, when combined with other variants, A1298C polymorphism becomes effective in determining the Hcy and folate levels. Individuals heterozygote for both alleles for both the C677T and A1298C variants have significantly increased plasma Hcy levels when compared to 677 CC / 1298 AA homozygotes [47]. Although in several studies, the 677TT variant of MTHFR was found preferentially associated with cancer [48], in other studies, this variant was found to reduce cancer risk by directing the pathway to thymidine synthesis and thus preventing uracil accumulation [49]. In the MTHFR gene, as high as 65 SNP's exist in a population of 240 individuals. Interestingly, the combinations of different polymorphisms in the gene coding MTHFR result in an additive effect of reduction in the enzyme activity [42]. So, it becomes an important aspect to study the polymorphisms collectively to overcome the possibility of missing the effect of a polymorphism that can be observed only when within the context of a particular genotype.

1.7.2 Methionine synthase reductase (MTRR)

Methionine synthase reductase (MTRR), an intermediate methyl carrier during the remethylation of Hcy to methionine, may be another important determinant of SAM production. In the past decade, an A to G polymorphism (A66G) in the 66th base pair (rs1801394), resulting in an isoleucine to methionine substitution in the 22nd aminoacid, was found to be associated with neural tube defects and cancer, especially in the absence of B_{12} vitamin and in the presence of the MTHFR C677T variant [48],[50]. In one study where a biochemical assay was used, demonstrated that the 66 GG variant of the MTRR enzyme had only 25% enzymatic activity compared to the wild-type protein [51], **(Table 1)**. This low enzymatic activity results in an inefficient production of SAM [52]. This is likely due to the decreased capacity of MTRR to activate MTR, in turn would lead to impaired production of methionine from Hcy. Not independently, but when combined with the 677T MTHFR variant, a cell carrying the MTRR 66G variant will have a lower capacity of remethylation, which will direct to DNA hypomethylation, instability and damage [48].

1.7.3 Methionine synthase (MTR)

A variation in the *methionine synthase* (MTR) enzyme (A2756G), which results in a subsitution of aspartic acid to glycine in the 919th aminoacid, also decreases the rate of methylation of Hcy to methionine (rs1805087). Although an increased plasma Hcy concentration is not observed in neural tube defect patients carrying the hypomorphic allele, it is thought to have an additive effect along with other variants of other enzymes, and especially the MTRR variants [53]. This aspartic acid to glycine substitution occurs in the vicinity of the binding domain of vitamin B_{12} of the MTR [54]. It was observed that plasma homocysteine levels are lower in MTR 2756 GG individuals when compared to the more common AA genotype [48], **(Table 1).**

1.7.4 Reduced Folate Carrier (RFC)

Another important variant in the one carbon pathway is in the *reduced folate carrier* (RFC) enzyme. The RFC gene codes an integral membrane protein which intakes the folate into the cell. A guanine to adenine transition (rs1051266) at the 80^{th} base pair (G80A) of RFC is related to a moderate degree of elevated plasma Hcy and plasma folate levels [55]. This arginine to histidine transition results in a higher affinity of RFC for folate impairing its subsequent intracellular release [55]. A recent study pointed out that this polymorphism reduces N^5-methyl-tetrahydrofolate transport efficiency to 54% compared to the wild-type allele [56], **(Table 1).**

1.7.5 Thymidylate synthase (TYMS)

The enhancer region of the thymidylate synthase gene contains a number of 28-base-pair tandem repeats. These repeats are either two repeats or three repeats, while three repeats occurring most frequently. The triple repeat has been associated with a 2.6 fold increase in the expression of TYMS [57]. One study showed that 3rpt/3rpt subjects have lower plasma folate and higher plasma Hcy levels than those with other genotypes [58]. This tandem repeat polymorphism, combined with reduced folate intake, has been associated with lung adenocarcinomas, [59]. Conversely, among those individuals with the 3rpt/3rpt genotype, higher folate intake was correlated with a 50% reduced risk of cancer [60].On the other hand, in individuals carrying a 2rpt/2rpt genotype, higher folate intake is correlated with a 50% increased cancer risk [60].Similar results were shown for vitamin B_{12}, but not with vitamin B_6, methionine or alcohol intake [60].

Table 1. The one-carbon pathway enzyme allele activity differences*.

Enzymatic activity	*Assay*	*Genotypes*			*Reference*
		CC	CT	TT	
MTHFR C677T	*Micobiological*	100%	65%	30%	[53]
	Biochemical	100%		55%	[42]
	RBC folate	100%	90%	82%	[61]
	RBC folate	100%	90%	**112%**	[45]
	Plasma folate	100%	96%	**75%**	[45]
	Total Hcy	100%	101%	121%	[55]
	Total Hcy	100%	108%	**123%**	[45]
	DNA hypomethylation	100%	98%	**121%**	[45]
	CpG Hypermethylation	100%	82%	**28%**	[62]
MTHFR A1298C					
		AA	AC	CC	
	Biochemical	100%		102%	[42]
	Plasma folate	100%	112%	114%	[46]
	Plasma folate	100%	110%	107%	[45]
	Total Hcy	100%	**75%**	**77%**	[46]
	Total Hcy	100%	**90%**	**115%**	[45]
	Genomic DNA methylation	100%	**231%**	**182%**	[46]
	DNA hypomethylation	100%	102%	119%	[45]
	CpG Hypermethylation	100%	107%	**24%**	[62]
MTR A2756G					
		AA	AG	GG	
	CpG Hypermethylation	100%	**28%**	**3%**	[62]
		AA	AG	GG	
MTRR A66G	*Biochemical*	100%		25%	[51]
		GG	GA	AA	
RFC G80A	*Plasma (total) Hcy*	100%	103%	124%	[55]
	5-CHO-mH4F transport	100%		57%	[56]

*Bold numbers show differences that were found to be statistically significant

1.8 Cancer and one carbon pathway enzyme variants

Epidemiologic studies where these alleles were studied individually and in pairs, suggested that they could be associated with cancer. Additive and synergistic effects of one carbon pathway enzyme variants and related dietary factors, such as folic acid and methionine

consumption, are complex and in order to clarify the association of each allelic variant with cancer, these, as well as their interactions with dietary factors need to be studies carefully. On the basis of functional effects of these polymorphic variants, there ought to be a correlation between the incidence of cancer and the hypoactive alleles of these enzymes.

However, there seems to be an inverse association between the MTHFR 677TT and cancer, especially in situations with sufficient folate intake [49]. A study has shown that the 2rpt/2rpt phenotype of TYMS and MTHFR 677 TT together result in a statistically decreased risk of hepatocellular carcinoma [41], MTHFR 677TT and 1298CC, together, are also associated with a reduced risk of cancer [41]. It seems likely that in folate deficiency, MTHFR 677TT becomes almost totally incapable of producing N^5-methylene-THF and thus, these cells have an impaired methylation capacity [52]. On the other hand, even when there is sufficient folate in the one carbon pool, the hypomorphic MTHFR protein will not be able to convert N^5-N^{10}-methylene-THF effectively; so, TYMS will use it for thymidine synthesis. One can envision that because this will prevent the accumulation of uracil and therefore, prevent uracil misincorporation, the "T" allele will protect against tumorigenesis by avoiding DNA damage caused by dUMP. In contrast, the TT allele was also shown to increase the cancer risk by 2.64 fold in folate deficiency and 1.6 fold in folate sufficiency [44]. Friso *et* al have shown that the MTHFR 677 TT allele associates with a significantly lower level of methyl cytosine in DNA, when compared to 677 CC and CT tissues, but only under conditions of low folate status [63]. At higher folate levels, the methyl cytosine levels did not differ from that among the 677 "CC" individuals [63]. That the 677 TT individuals have low levels of methyl cytosine in their DNA and elevated levels of plasma Hcy suggests that the MTHFR TT allele results in insufficient amounts of N^5-methylene-THF that can't meet the demands for the *de novo* methylation. In another study, the MTHFR 677TT and the MTRR 66GG variant combination showed the highest amount of DNA damage, calculated by micronucleus formation [47]. Vaughn *et al* showed that the MTRR alleles have no effect on plasma folate and plasma vitamin B12, in the presence of a 677 CC or CT allele. However, in those individuals homozygous for the 677 T allele, the existence of MTRR AG or GG allele resulted in a significant increase in plasma homocysteine

levels [64]. This observation suggests that the functional MTRR 66 A/A allele compensates for a hypoactive MTHFR 677 TT allele.

It is worthwhile to point out that the association of MTHFR 677 variants is different in different stages and types of cancer. While the 677 CC and CT alleles are related to a reduced risk in renal cell carcionoma [65], they are related to an elevated risk in colorectal cancer [48]]. The effect may be different at different stages of the oncogenesis process. Supporting this is the observation that folate deficiency predisposes to colon cancer, but once neoplastic lesions are present, folate supplementation actually accelerates colon cancer transformation {Kim, 2003}. Folate supplementation after a certain stage of tumorigenesis may enhance DNA synthesis in an already transformed cell.

In a study implemented by Chango *et al*, the RFC allele does not seem to affect plasma homocysteine levels when the MTHFR 677 allele is CC or CT. However, in the presence of the 677 TT allele, RFC GA and AA will influence the utilization of folate [55].

It is a challenging task to determine whether uracil misincorporation or errors in methylation process causes oncogenesis. Moreover, how the genes in the one carbon metabolism respond to folate deficiency is largely unknown. For instance, in the HCT116 cell line, folate deficiency appeared to preferentially shuttle the flow of one carbon units to the methionine cycle to protect methylation reactions and thereby suppress DNA synthesis. However in Caco2 cells in the same experimental setup, the metabolic priority in response to folate depletion was to shuttle the available folate pools to the nucleotide biosynthesis pathway at the expense of the methionine cycle [31].

The studies of the one carbon pathway enzyme variants suggest that folate has an important role in modulating epigenetic features of the DNA that controls gene expression [66]. Since the end product of the one carbon pathway, S-adenosylmethionine, will be utilized for the *de novo* methylation of DNA, the efficiency of the production of SAM will affect gene expression via DNA methylation. Folate depletion is thought to cause tumorigenesis either by impairing methylation or by hindering DNA synthesis {Kim, 2003}. When the effects of one the

carbon metabolism enzymes are combined with the effects of low folate intake, significant results on the alterations of DNA methylation are obtained suggesting the roles of both events.

.

1.9 The aim of the study

The role of DNA methylation in differentiation, disease and tissue specific gene expression has received considerable attention. In many circumstances, DNA methylation can on its own be an effective mechanism for gene silencing [24]. Methylated silent genes can be re-activated by using demethylating agents [67]. It has been proposed that the methylation of CpG sites at gene promoter regions inhibits the recruitment of the transcription machinery and thus strongly represses the expression [24]. DNA methylation is proposed to be the basic mechanism for the regulation of CT genes such as MAGE [24] and SSX (Gure, 2005}. Interestingly, spermatogonia cells undergo epigenetic reprogramming, which involves genomic hypomethylation [68].

Since genomic hypomethylation causes an increase in CT expression which is associated with developed stages of cancer and poor outcome (Gure, 2005}, the reversal of the genomic hypomethylation might improve prognosis. Because the methylation process requires the effective production of SAM by the one carbon pathway, we decided to investigate the relationship between CT gene expression and the allelic variants of the one carbon pathway, which we think are the primary effectors of SAM production.

It is already known that a variety of enzyme variants in the one carbon pathway cause perturbations in the methylation cycle. Several epidemiologic studies have shown that variants of one carbon pathway enzymes are correlated with oncogenesis in different ways [69]. One can envision that the perturbations of the one carbon pathway hindering the methylation capacity of the cell may result in an additive effect resulting in DNA hypomethylation and might lead thus, to CT positivity and poor prognosis. We hypothesized that there could be a correlation between the allelic variants of the one carbon pathway enzymes and CT expression. The importance of such correlation would be that if CT positivity can be predicted by the allele variants a cancer

patient, the possible conversion of high CT expression to a low state by the addition of dietary supplements such as folic acid could improve the prognosis of the patient.

One can envision that typing for the one carbon enzyme alleles among individuals who have a high-risk of developing cancer might help us suggest precautions towards restoring the patient's impaired methylation capacity, such as the fortification or the supplementation of the diet by folic acid.

In this study, I determined the one carbon pathway enzyme genotypes of lung cancer tumors and cell lines, and studied the correlation of these genotypes with CT expression levels. The allelic variants I chose to study were the non-synonymous, highly prevalent and for which there were a large number of epidemiologic studies indicating a correlation with cancer. The selected variants include MTHFR C677T, MTHFR A1298C, MTRR A66G, MTR A2756G and RFC G80A single nucleotide polymorphisms.

I, thus, determined the one carbon enzyme genotypes of individuals who were previously typed for CT expression, using the restriction fragment length polymorphism assay (RFLP). In this method, the region containing the single nucleotide polymorphism (SNP) is amplified by PCR. The PCR product is subsequently digested with a specific endonuclease, which can recognize and cut DNA wherever a specific short sequence exists. This process is called a *restriction digestion*. This method can only be used if the allelic variant can be distinguished by a specific restriction endonuclease. Thus, restriction enzymes were selected such that the PCR product of one allele would be digested whereas the other allele would not. When the digestion products are analyzed by agarose gel electrophoresis, the allele is determined according to the molecular weights of the digestion products.

In summary, apart from many epigenetic mechanisms which may be perturbed in cancer, DNA hypomethylation seem to be the major actor for the ectopic expression of CT genes. By elucidating the underlying mechanisms which might thus cause CT expression, we potentially will have described a novel biomarker useful for the determination of tumorigenesis as well as improving the prognosis and lifespan of a cancer patient.

CHAPTER 2. MATERIALS AND METHODS

2.1 The PCR Method

The standart reaction mixture contained 1µg of genomic DNA, 1µl of 10mM dNTP mix (Finnzymes, Cat. No: F-506L), 2µl of 25mM primer mix, 2.5 µl of 10X Buffer (Finnzymes Dnazyme II HotStart Reaction Buffer, Cat. No: F-522), 0.5 µl Hot Start Taq Polymerase (Finnzymes, Cat. No: F-504L) added up to 25 µl with ddH_2O. Then, the tubes are transferred to the Perkin Elmer (PE9700) thermal cycler for 35 cycles. Before the reaction, initial denaturation was performed at 94 ^oC for 10mins. Each cycle consisted of 30sec at 94 ^oC, 30sec at Tm ^oC, and 30 sec at 72 ^oC. Then, the mixture was incubated at 72 ^oC for 7mins for final annealing and a final hold of 4 ^oC. See Table 2 for primer pairs, Tm values and Table 3 for nested primers.

Table 2. PCR primers used for RFLP analysis

Gene Name and SNP	Gene ID	Ref. Seq.	Primers	Tm	Product Size(bp)
MTHFR C677T	4524	AY338232.1	5' – TTTGAGGCTGACCTGAAGCAC – 3' (sense) IIF 3' – GACCTGAGAGGAGATCTGG – 5' (antisense) IIR	60C	288 (8,700nt-8,988nt)
MTHFR A1298C	4524	AY338232.1	5' – GAGGAGCTGCTGAAGATGTG -3' (sense) IIF 3' - TGGAGGTCTCCCAACTTACC -5'(antisense) IIR	65C	261 (10,605nt-10,866nt)
MTR A2756G	4548	NT_004836	5' - TGTTATCAGCATTGACCATTACTACAC -3'(sense) IIF 3' - ACTGTTTCAGCACCTGTTTCCC -5' (antisense) IIR	65C	498 (105,261nt-105,759nt)
MTRR G66A	4552	NT_006576	5' - GCAAAGGCCATCGCAGAAGACAT -3'(sense) IF 3' - CACTTGTTCTCACAGCCACCC -5'(antisense) IIR	60C	380 (7,860,967nt-7,861,347nt)
RFC G80A	6573	AL163302.2	5' – CTCCCGCGTGAAGTTCTT -3' (sense) IF 3' – AGCGTCACCTTCGTCCCCTC -5' (antisense) IIR	60C	231 (236,569nt-236,800nt)
TS Polymorphism	7298	NC_000018.8	5' - GTTCCCGGGTTTCCTAAGAC -3' (sense) IIF 3'- GATCTGCCCCAGGTACTGCA -5' (antisense) IIR	65C	390
TS Expression	7298	NC_000018.8	5' - GCAGATCCAACACATCCTCC - 3' (sense) 3' - CCATTGGCATCCCAGATTTTCAC -5' (antisense)	60C	236

Table 3. Nested PCR primers used for RFLP analysis

Gene Name and SNP	Ref. Seq.	Nested Primers	Product size (bp) nested	Tm
MTHFR C677T	AY338232.1	5'- TGAAGGAGAAGGTGTCTGCGGGA – 3' (sense nested) IF 3'- AGGACGGTGCGGTGAGAGTG -5' (antisense nested) IR	198 (8723nt-8921nt)	60°C
MTHFR A1298C	AY338232.1	5' – GAGGAGCTGACCAGTGAAG -3'(sense nested) IF 3'- GGTAAGTTGGGAGACCTCCA -5' (antisense nested) IR	136 (10.629nt-10.765nt)	60°C
MTR A2756G	NT_004836	5'- TGTTCCCAGCTGTTAGATGAAAATC -3'(sense nested) IF 3'- GATCCAAAGCCTTTTACACTCCTC – 5'(antisense nested) IR	211(105.334nt-105.545nt)	65°C
MTRR G66A	NT_006576	3'-GTGAAGATCTGCAGAAAATCCATGTA -5'(antisense nested) IR*	66bp (7,860,967nt-7,861,032nt)	60°C
RFC G80A	AL163302.2	5'- AGCCGTAGAAGCAAAGGTAGC – 3'(sense nested) IIF 3' – TGCATTCGTCTCCAGGGTG – 5'(antisense nested) IR	154bp (236,646nt-236,667nt)	55°C
TS	NC_000018.8	5'-AGCAGGAAGAGGCGGAGC-3'(sense nested) 3'- CCGGCCACAGGCATG -5'(antisense nested)	172bp	60°C

*The 5'(forward) nested primer used for MTRR was identical to that shown in Table 2.

2.2 DNA Samples

Tumor samples from 763 lung cancer patients were previously collected and typed for CT expression (Gure, 2005). All of the samples were scored for the positivity of CT expression. The scores corresponding to different expression levels of each CT gene ranged between a negative (-) to three positives (+++). We then selected, genomic DNA prepared from these tumor samples with the highest and the lowest amount of CT expression and used them for one carbon pathway allele genotype analysis. Table 4 shows the label, cancer type and the CT expression scores of the samples. Table 4a and Table 4b the CT(+) and CT (-) lung cancer samples used to for this study respectively.

Table 4. CT gene expression profile of CT (+) tumor samples

Lu CT(+)	HISTOLOGY	NY-ESO-1	LAGE-1	MAGE-A1	MAGE-A3	MAGE-A4	MAGE-A10	CT-7	SSX-2	SSX-4
31	SQCC	(+++)	(+/-)	(+)	(+++)	(+++)		(+)		
68	Adeno	(+++)		(+)	(+)			(-)		
87	Adeno	(-)	(-)	(+++)	(+++)			(+++)		
89	Adeno	(+++)	(+++)	(+++)	(+++)	(+/-)	(+)	(+/-)		
111	Lgcell	(-)	(-)	(+++)	(+++)	(+++)		(+++)		
131	SQCC	(+++)	(+/-)	(+)	(+++)	(+++)				
168	SQCC	(+++)	(+++)	(+++)	(+++)	(+++)		(+)		
185	NSCLC	(+++)	(+++)	(+++)	(+++)	(+/-)	(-)	(-)	(++)	(-)
186	Adeno	(+++)	(+++)		(+++)					
219	SQCC	(+)	(+)	(+++)	(+++)	(+++)		(-)	(+/-)	(+++)
223	SQCC	(+++)	(+++)		(+++)	(-)				
649	AdenoBAC	(+++)	(+/-)	(+++)	(+++)	(+/-)	(+++)	(-)	(-)	
652	SQCC	(+++)	(+++)	(+++)	(+++)	(+++)	(+++)	(-)	(-)	
658	AdenoBAC	(+/-)	(+++)	(+++)	(+++)	(+++)	(+/-)		(+++)	
726	Adeno	(+++)	(+++)	(+++)	(+++)	(+++)	(+++)		(++)	(+++)
736	Adeno	(+++)	(+++)	(+++)	(+++)	(+++)	(+++)		(+/-)	(-)
739	Lgcell	(+++)	(+++)	(+++)	(+++)	(+++)	(+++)		(++)	(+++)
745	SQCC	(+++)	(+++)	(+++)	(+++)	(+++)	(+/-)		(++)	(-)
752		(+++)	(+++)	(+++)	(+++)	(+++)	(+/-)		(-)	(-)
753		(++)	(+/-)	(+++)	(+++)	(+++)	(+/-)			(-)
759		(+++)	(+++)	(+++)	(+++)	(+++)	(+/-)			(-)

Table 5. CT gene expression profile of CT(-) tumor samples

Lu CT(-)	HISTOLOGY	NY-ESO-1	LAGE-1	MAGE-A1	MAGE-A3	MAGE-A4	MAGE-A10	CT-7	SSX-2	SSX-4
69	BAC	(-)	(-)	(-)	(-)			(-)		
77	SQCC	(-)	(-)	(-)	(-)			(-)		
88	Adeno	(-)	(-)	(-)	(-)			(-)		
90	Adeno	(-)	(-)	(-)	(-)			(-)		
108	Adeno	(-)	(-)	(-)	(-)			(-)		
110	Adeno	(-)	(-)	(-)	(-)			(-)		
112	SQCC	(-)	(-)	(-)	(-)			(-)		
180	Adeno	(-)	(-)	(-)	(-)			(-)	(-)	(-)
183	BAC	(-)	(-)	(-)	(-)			(-)	(-)	(-)
191	Adeno	(-)	(-)	(-)	(-)	(-)		(-)		
221	Adeno	(-)	(-)	(-)		(-)		(-)		
225	Adeno	(-)	(-)	(-)	(-)	(-)		(-)	(-)	(-)
639	AdenoBAC	(-)	(-)	(-)	(-)	(-)	(-)	(-)		
656	AdenoBAC	(-)	(-)	(-)	(-)	(-)	(-)			
657	SQCC	(-)	(-)	(-)	(-)	(-)	(-)			
670	AdenoBAC	(-)	(-)	(-)	(-)	(-)	(-)			
692	AdenoBAC	(-)	(-)	(-)	(-)	(-)	(-)		(-)	(-)
693		(-)	(-)	(-)	(-)	(-)	(-)		(-)	(-)
694		(-)	(-)	(-)	(-)	(-)	(-)		(-)	(-)
698	Adeno	(-)	(-)	(-)	(-)	(-)	(-)		(-)	(-)
706	Adeno	(-)	(-)	(-)	(-)	(-)	(-)		(-)	(-)
707	Adeno	(-)	(-)	(-)	(-)	(-)	(-)		(-)	(-)
713	AdenoBAC	(-)	(-)	(-)	(-)	(-)	(-)		(-)	(-)
716	Adeno	(-)	(-)	(-)	(-)	(-)	(-)		(-)	(-)
718	AdenoBAC	(-)	(-)	(-)	(-)	(-)	(-)		(-)	(-)
728	AdenoBAC	(-)	(-)	(-)	(-)	(-)	(-)		(-)	(-)
748	AdenoBAC	(-)	(-)	(-)	(-)	(-)	(-)		(-)	(-)
749	Adeno	(-)	(-)	(-)	(-)	(-)	(-)		(-)	(-)
751	AdenoBAC	(-)	(-)	(-)	(-)	(-)	(-)		(-)	(-)
763		(-)	(-)	(-)	(-)	(-)	(-)			

2.3 Restriction Fragment Length Polymorphism (RFLP) Analysis

RFLP analysis was done as previously described [43]. Following the PCR reaction, 5 µl of PCR product was mixed with 1 µl of restriction endonuclease enzyme (New England Biolabs), 1 µl of Buffer 2 (New England Biolabs), added up to 10 µl with ddH_2O. The mixture was incubated at 37°C overnight. Table 4 shows the expected sizes of the digestion products for each primer pair.

Table 6. Expected RFLP product sizes

Polymorphism and Restriction endonuclease	Primers			
	IIF-IIR	IIF-IR	IF-IIR	IF-IR
MTHFR C677T C/C (HinF1I)	288	220	266	198
MTHFR C677T C/T (HinF1I)	288, 241, 47	220, 173, 47	266, 241, 25	198, 173, 25
MTHFR C677T T/T (HinF1I)	241, 47	173, 47	241, 25	173, 25
MTHFR A1298C C/C (MboII)	266	160	237	136
MTHFR A1298C C/A (MboII)	261, 209, 28, 24	160, 108, 28, 24	237, 209, 28	136, 108, 28
MTHFR A1298C A/A(MboII)	209, 28, 24	208, 28, 24	209, 28	108, 28
MTR A2756G A/A (HaeIII)	498	284	425	211
MTR A2756G A/G (HaeIII)	498, 345, 153	284, 153, 131	425, 345, 80	211, 131, 80
MTR A2756G G/G (HaeIII)	345, 153	153, 131	345, 80	131, 80
MTRR G66A G/G (Nde1)	-	-	381	66
MTRR G66A G/A (Nde1)	-	-	381, 317, 39, 25	66, 42, 24
MTRR G66A A/A (Nde1)	-	-	317, 39, 25	42, 24
RFC G80A A/A (HinP1I)	287	154	364	231
RFC G80A A/G (HinP1I)	287, 257, 30	154, 124, 30	364, 257, 70, 37	231, 124, 70,37
RFC G80A G/G (HinP1I)	257, 30	124, 30	257, 70, 37	124, 70, 37

2.3.1 Digestion Products

2.3.1.1 MTHFR C677T

The MTHFR C677T polymorphism creates a *HinF1I* restriction site. Table 6 shows the transition of the protein and the nucleotide sequence and the digestion products of all alleles. Upon digestion, *HinF1I* will be able to cut the "T" allele whereas the "C" allele will remain uncut.

Table 7. A222V (C677T) variant of the MTHFR gene

- *Amino acid sequence A222V (Ref Seq: NM_005957):*

151- KNIMALRGDP IGDQWEEEEG GFNYAVDLVK HIRSEFGDYF DICVAGYPKG

201- HPEAGSFEAD LKHLKEKVSA G**A**DFIITQLF FEADTFFRFV KACTDMGITC

251- PIVPGIFPIQ GYHSLRQLVK LSKLEVPQEI KDVIEPIKDN DAAIRNYGIE

- *Nucleotide Sequence C677T (Ref Seq: NM_005957):*

GANTC – *HinF1I* restriction site

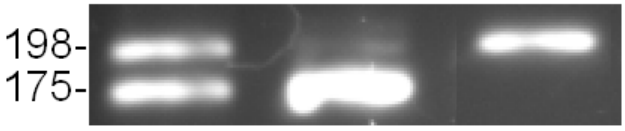
781 aggccacccc gaagcaggga gctttgaggc tgacctgaag cacttgaagg agaaggtgtc

841 tgcgg**gag**C**c** gatttcatca tcacgcagct ttctttgag gctgacacat tcttccgctt

901 tgtgaaggca tgcaccgaca tgggcatcac ttgccccatc gtccccggga tctttcccat

961 ccagggctac cactcccttc ggcagcttgt gaagctgtcc aagctggagg tgccacagga

1021 gatcaaggac gtgattgagc caatcaaaga caacgatgct gccatccgca actatggca



C/T T/T C/C

2.3.1.2 MTHFR A1298C

Upon the transition of adenine to cytosine at the 1298th position of the MTHFR gene, the recognition site of the *MboII* enzyme becomes abolished. Therefore, the undigested products will indicate the presence of an "C" allele whereas the "A" allele will result in digestion. The amino acid and the nucleotide alterations are designated in Table 7.

Table 8. E429A (A1298C) variant of the MTHFR gene

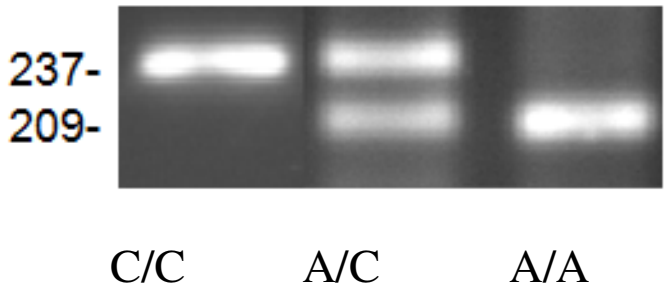
Amino acid sequence E429A (Ref Seq: NM_005957):

```
351- LSAHPKRREE DVRPIFWASR PKSYIYRTQE WDEFPNGRWG NSSSPAFGEL

401- KDYYLFYLKS KSPKEELLKM WGEELTSEES VFEVFVLYLS GEPNRNGHKV

451- TCLPWNDEPL AAETSLLKEE LLRVNRQGIL TINSQPNING KPSSDPIVGW
```

- *Nucleotide Sequence A1298C (Ref Seq: NM_005957):*

  GAAGA…N^{8} –*MboII* restriction site

```
1321 ggagtgggac gagttcccta acggccgctg gggcaattcc tcttcccctg cctttgggga
1381 gctgaaggac tactacctct tctacctgaa gagcaagtcc cccaaggagg agctgctgaa
1441 gatgtggggg gaggagctga ccagtgaaga aagtgtcttt gaagtcttcg ttctttacct
1501 ctcgggagaa ccaaaccgga atggtcacaa agtgacttgc ctgccctgga acgatgagcc
1561 cctggcggct gagaccagcc tgctgaagga ggagctgctg cgggtgaacc gccagggcat
1621 cctcaccatc aactcacagc ccaacatcaa cgggaagccg tcctccgacc ccatcgtggg
```



2.3.1.3 MTR A2756G

The A to G transition of this polymorphism creates a *HaeIII* restriction site. Thus, the G allele will be able to be digested. The information about the transition is indicated in Table 8.

Table 9. D919G (A2756G) variant of the MTR gene

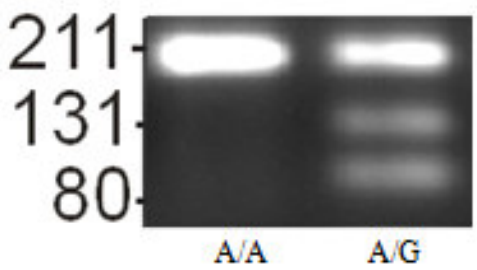
- *Amino acid sequence D919G (Ref Seq: NP_000245):*

```
781 kgdvhdigkn ivgvvlgcnn frvidlgvmt pcdkilkaal dhkadiigls glitpsldem
841 ifvakemerl airiplligg attskthtav kiaprysapv ihvldasksv vvcsqllden
901 lkdeyfeeim eeyedirqdh yeslkerryl plsqarksgf qmdwlsephp vkptfigtqv
961 fedydlqklv dyidwkpffd vwqlrgkypn rgfpkifndk tvggearkvy ddahnmlntl
```

- *Nucleotide Sequence A2756G (Ref Seq: NC_000001):*

GGCC – *HaeIII* restriction site

```
89641 attgaccatt actacaccag ttttatcatc ttttgctcat ctatggctat cttgcatttt
89701 cagtgttccc agctgttaga tgaaaatcta aaggatgaat actttgagga aatcatggaa
89761 gaatatgaag atattagaca ggaccattat gagtctctca aggtaagtgg tagaaacaga
89821 tttttgcttg tttttaatgt gactgttttt tatgatccta gtttttaatg tgacttttta
89881 aaatggtttt gaggagtgta aaaggctttg gatcatttta gagaatttct gtcttctagt
```



2.3.1.4 MTRR A66G

The A66G transition in the MTRR gene, upon with a substitution of a base with modified primers, creates a restriction site for the *NdeI* enzyme. The "A" allele will produce digested products where as the "G" allele will remain uncut.

Table 10. I22M (A66G) variant of the MTRR gene

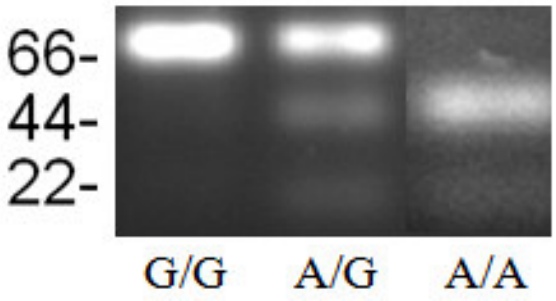
- *Amino acid sequence I22M (Ref Seq: NP_002445):*

```
  1 mrrflllyat qqgqakaiae eiceqavvhg fsadlhcise sdkydlktet aplvvvvstt
 61 gtgdppdtar kfvkeiqnqt lpvdffahlr ygllglgdse ytyfcnggki idkrlqelga
121 rhfydtghad dcvglelvve pwiaglwpal rkhfrssrgq eeisgalpva spassrtdlv
```

- *Nucleotide Sequence A66G (Ref Seq: NC_000005):*

CATATG – *Nde1* restriction site gcaaaggccatcgcagaagcaat -primer with a modified base

```
1681 gccttgaagt gatgaggagg tttctgttac tatatgctac acagcaggga caggcaaagg
1741 ccatcgcaga agaatatatgt gagcaagctg tggtacatgg attttctgca gatcttcact
1801 gtattagtga atccgataag gttagagccg ttacagtgga ttttaccgtt ttgtgctttg
1861 aagaattttg gttgggaagt gatatttatg aaacaaaagg acactaatac caccacatag
1921 tctttgtttt ttaacagaaa tgtgtttgtt caatggtata gtaagatatc accagcattt
```



2.3.1.5 RFC G80A

The G to A transition at the 80th base of the RFC gene abolishes a *HinP1I* restriction site. So, the G allele will be recognized and cut by *HinP1I* whereas the A allele will be undigested.

Table 11. R27H (G80A) variant of the RFC gene

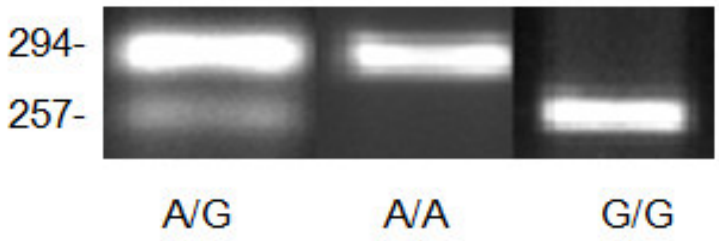
- *Amino acid sequence R27H (Ref Seq: NP_919231):*

```
  1 mvpsspavek qvpvepgpdp elrswrhlvc ylcfygfmaq irpgesfitp yllgpdknft
 61 reqvtneitp vlsysylavl vpvflltdyl rytpvlllqg lsfvsvwlll llghsvahmq
121 lmelfysvtm aariayssyi fslvrparyq rvagysraav llgvftssvl gqllvtvgrv
```

- *Nucleotide Sequence G80A (Ref Seq: NC_000021):*

GCGC – *HinP1I* restriction site

```
4441 ccttcgtccc ctccggagct gcacgtggcc tgagcaggat ggtgccctcc agcccagcgg
4501 tggagaagca ggtgcccgtg gaacctgggc ctgaccccga gctccggtcc tggcggcacc
4561 tcgtgtgcta cctttgcttc tacggcttca tggcgcagat acggccaggg gagagcttca
4621 tcaccccctа cctcctgggg cccgacaaga acttcacgcg ggagcaggca tgtgggtgcc
```



2.4 c-DNA Synthesis

c-DNA synthesis was carried out with a c-DNA synthesis kit (Finnzymes DyNAmo cDNA synthesis kit, #F-470L). The master mix contained 10 µl of 2X RT Buffer, 1 µl of random hexamer primer set with a concentration of 300µg/µl, 2 µl of M-MuLV Rnase H^+ reverse transcriptase, maximum 1 µg of template RNA added up to 20 µl with ddH_2O. After transferring to the Perkin Elmer (PE9700) thermal cycler, the mixture was incubated in the following conditions: 25 ^oC for 10 mins, 37 ^oC for 30mins, 85 ^oC for 5 mins and a final hold of 4 ^oC.

2.5 Cell Culture

Cells were grown at a confluence of approximately 80% in RPMI with 10% fetal calf serum (Hycole), supplemented with glutamine, streptomycin and penicillin. The old medium was removed and the cells were washed with 1X PBS. After removal of the 1X PBS, Trypsin/EDTA solution was added and the cells were left for incubation at 37C for 3 - 4 minutes. When the cells were detached from the surface of the plate, trypsin was inactivated by adding fresh growth medium. The detached cells were suspended using a pipettor and were placed in a falcon tube. The cells then were centrifuged at 800rpm for 5mins and the supernatant containing the Trypsin/EDTA solution was removed. The pellet was re-suspended with fresh medium and cell suspension was plated to a new flask. Fresh medium is added to the flask and placed in 37 ^oC, %5 CO_2 incubator.

2.6 RNA Isolation

Cells were trypsinized as previously described in the *Cell Culture* section. Cold media is added and the cells are mixed. The cells are transferred into a 15ml falcon tube and centrifuged at 800rpm for 5 mins, at 4 ^oC. After centrifugation, the media was removed and washed with 1X PBS and 1ml Trizol (Invitrogen) was added. The cells were mixed until they homogenize and transferred into an eppendorf tube. 200µl chloroform was added and the tube was vigorously shaken for 15 seconds. After incubating for 3 minutes at room temperature, the cells were

centrifuged at 13.000rpm for 15 mins, at 4 °C. After this step, the supernatant was removed by a pipettor and placed into a new tube. 500 µl isopropanol was added and the supernatant is mixed. The mixture was incubated for 10 mins at room temperature and centrifuged at 13.000 rpm for 10 minutes, at 4 °C. Then the isopropanol was completely removed and 1ml 75% EtOH was added following 8 minutes of centriguation at 8000rpm, at 4 °C. The 75% EtOH wassucked and 1ml 99.8% EtOH was added following the same centrifugation step of 8 mins at 8000rpm, 4 °C. 99.8% EtOH was removed and the pellet is dried under laminar flow for 5-10mins. Lastly, the pellet is dissolved in 20µl-25µl DNase-RNase free H_2O and mixed

2.7 Real Time PCR

The real-time RT-PCR assays were done with the iCycler instrument (BioRad Laboratories) using the Finnzymes Dynamo SYBR Green qPCR kit (Finnzymes Cat #F-410). The primers used for TS expression were 5' -GCAGATCCAACACATCCTCC-3' (sense) and 3'-CCATTGGCATCCCAGATTTTCAC-5' (antisense). β-actin was used as a loading control. The PCR reactions were set up in a volume of 20 µl, containing 2 µl of sample cDNA, 10 µl of 2X Master mix, 5 pmol from each *TS* specific primers, added up to 20 µl with RNase-DNase free water. The cycling conditions were as follows: 95 °C for 1 min, 58 °C for 1 min, and 72 °C for 1min for 40 cycles with initial melting at 95 °C for 10 min.

Relative expression levels were calculated using the PCR threshold cycle number (C_T) for each sample and control sample (both of which were normalized according to β-actin mRNA for differences in amount of total RNA added to the reaction), using the formula $2^{-(\Delta C_{T\,\text{sample}} - \Delta C_{T\,\text{control}})}$ [70], [71]. ΔC_T represents the difference in C_T values between the target and β-actin transcripts. RT-PCR was performed in triplicates for each sample and average C_T values were calculated

2.8 Statistical Analysis

For the significance tests used in the percentage tables, chi-square test was used. The significance level was determined as p=0.05. For the sample number calculation tests, univariate logistic regression model analysis (Wald's test) was performed by the Power and Precision software. Multivariate logistic regression models were done by applying an algorithm in R software [83]. For both of the tests the significance level was set to α=0.05. The event rate is defined as the probability of a genotype being CT (+). The power of the analysis is defined as the ability of the test to differentiate the difference between the alleles in terms of CT expression. The odds ratio in multivariate tests indicates the ratio of the probability of a genotype being a CT (+) sample divided by the probability of the same genotype being a CT (-) sample and the division of the probabilities of the other genotypes being CT (+) and CT (-) samples, respectively. An exemplary formula showing this ratio for the MTHFR 677 polymorphism is indicated below. The estimate values in the multivariate test is calculated as log(OR).

$$\textbf{Odds Ratio (OR)} = \frac{\textbf{P(CT=+[G677=CC])/ P(CT=-[G677=CC])}}{\textbf{P(CT=+[G677=CT, TT])/ P(CT=-[G677=CT, TT])}}$$

CHAPTER 3. RESULTS

3.1 One carbon pathway enzyme genotype frequencies of lung cancer patients

All 50 lung cancer samples were typed for five allelic variants, except for MTR A2756G and MTRR G66A, for which 1 and 4 samples could not be typed, respectively. We observed a frequency of 44 % CC, 34% CT and 11% TT for the MTHFR C677T polymorphism, 48% AA, 46% AC and 6% CC for MTHFR A1298C, 52% AA and 48% AG for MTR A2756G, 38.1% AA, 47.6% AG and 14.3% GG for MTRR A66G and 22% GG, 56% AG and 22% GG genotypes for RFC G80A polymorphisms..

Next, we checked to see whether our lung cancer sample panel was in Hardy-Weinberg equilibrium. For this reason, we calculated the expected frequencies of each genotype and evaluated chi-square analysis for observed/expected ratio. The genotype frequencies for all the polymorphisms were in accordance with the Hardy–Weinberg equilibrium: MTHFR C677T (p=0.4), MTHFR A1298C (p=0.7), MTR A2756G (p=0.08), MTRR A66G (p=1) and RFC G80A (p=0.9). **Table 12** summarizes our observations and expected values for each genotype.

Table 12 One carbon enzyme distributions in lung cancer patients and Hardy-Weinberg expectations.

Polymorphism	***Genotype***	***Observed***	***Observed/ Expected***	***χ^2 value***	***p****
MTHFR677	CC	21	1.1		
	CT	19	0.8		
	TT	10	1.3	2.0	0.4
MTHFR1298	AA	24	1		
	AC	23	1.1		
	CC	3	0.7	0.7	0.7
MTR2756	AA	25	0.9		
	AG	24	1.3		
	GG	0	-	5.2	0.08
	Undetermined (1)				
MTRR66	GG	16	1		
	GA	20	1		
	AA	6	1	0.004	1
	Undetermined (4)				
RFC80	GG	11	0.9		
	GA	27	1.1		
	AA	12	0.9	0.3	0.9

*Chi square

3.2 Distribution of one carbon enzyme genotypes among CT (+) and CT (-) lung cancer patients I

Among CT (+) samples the distribution was 44% CC, 26% CT and 30 % TT, while CT (-) samples' distribution was 31% CC, 41% CT and 28% TT.

The distribution of other polymorphisms was not significantly different when samples were stratified for CT gene expression. The MTHFR A1298C polymorphism, the distribution was 57% AA, 38% AC and 5% CC for CT (+) lung cancer samples and, 41% AA, 52% AC and 7% CC among the CT (-) samples.

The observed genotype frequencies for the MTR A2756G polymorphism were 63% AA and 37% AG among the CT (+); 55% AA and 45% AG among the CT (-) samples. For the MTRR A66G polymorphism, our observation for the CT (+) lung cancer samples was a frequency of 35% AA, 55% AG and 10% GG genotypes. On the other hand, CT (-) lung cancer samples were 35% AA, 46% AG and 19% GG.

We found 29% GG, 48% AG and 24% AA genotype frequency in CT (+) samples for the RFC G80A polymorphism. Among the CT (-) samples, the distribution was 17% GG, 59% AG and 24% AA.

Table 13. Distribution of 1-carbon enzyme genotypes among CT-positive and -negative lung cancer patients I: Chi-square test.

Polymorphism	Genotype	CT(+)	CT(-)	p*
MTHFR677	CC	12 (*44*%)	9 (31%)	
	CT	7 (26%)	12 (41%)	
	TT	2 (30%)	8 (28%)	0.02
MTHFR1298	AA	12 (57%)	12 (41%)	
	AC	8 (38%)	15 (52%)	
	CC	1 (5%)	2 (7%)	0.2
MTR2756	AA	10 (63%)	16 (55%)	
	AG	11 (37%)	13 (45%)	
	GG	0 (0%)	0 (0%)	0.3
MTRR66	GG	7 (35%)	9 (35%)	
	GA	11 (55%)	12 (46%)	
	AA	2 (10%)	5 (19%)	0.3
	Undetermined	4		
RFC80	GG	6 (29%)	5 (17%)	
	GA	10 (48%)	17 (59%)	
	AA	5 (24%)	7 (24%)	0.2

*Chi square

Next, for each polymorphism, we grouped heterozygotes with the homozygote where both showed a similar bias towards a given CT expression phenotype, and thus generated 2 by 2 charts. The odds ratio calculated from these charts demonstrated borderline significance for the MTHFR677 polymorphism, with an odds ratio of 2.96, when other polymorphisms, when thus evaluated, were not significantly different when CT(+) and (-) samples were compared (**Table 14**). Both analyses suggest that the hypoactive MTHFR677 CT/TT phenotypes preferentially associate with lack of CT gene expression. .

Table 14. Distribution of 1-carbon enzyme genotypes among CT (+) and CT (-) lung cancer patients II: Odds ratios

Polymorphism	Genotype	CT(+)	CT(-)	OR (95% CI)	p*
MTHFR677	CC	12	9		
	CT/TT	9	20	2.96 (0.92-9.53)	0.07
MTHFR1298	AA	12	12		
	AC/CC	9	17	1.89 (0.61-5.89)	0.3
MTR2756	AA	10	16		
	AG/GG	11	13	0.85 (0.28-2.61)	0.6
MTRR66	GG	7	9		
	GA/AA	13	17	0.9 (0.26-3.2)	0.8
	Undetermined	4			
RFC80	GG	6	5		
	GA/AA	15	24	1.92 (0.5-7.41)	0.3

*Fisher's exact test (2-sided)

3.3 One carbon enzyme allele distribution among CT-positive and -negative lung cancer patient.

Since the previous analyses suggested an association between the MTHFR677 CT and TT genotypes with the lack of CT gene expression, we tested whether a similar association could be found for a given allele. As shown in **Table 15**, the hypoactive T allele of the MTHFR677 polymorphism was strongly associated with the lack of CT gene expression as well. We did not observe a significant association between CT gene expression and the presence of any other allele (**Table 15**).

Table 15. One carbon enzyme allele distribution among CT-positive and -negative lung cancer patients.

Polymorphism	Allele	CT(+)	CT(-)	OR (95% CI)	p*
MTHFR677	C	31	11		
	T	30	28	2.63 (1.11-6.21)	0.025
MTHFR1298	A	32	10		
	C	39	19	1.56 (0.64-3.82)	0.3
MTR2756	A	31	11		
	G	45	13	0.81 (0.32-2.05)	0.7
MTRR66	G	25	15		
	A	30	22	1.22 (0.53-2.84)	0.6
RFC80	G	22	20		
	A	27	31	1.26 (0.57-2.8)	0.6

*Fishers exact test (2-sided)

3.4 CT expression associations with one carbon enzyme genotype combinations in lung cancer patients

Since combinations of genotypes are known to result in inefficient utilization of folate, as well as decreased enzymatic activity, we tested whether a given combination of two genotypes would result in a significant stratification of samples based on CT gene expression. Among all possible two-by-two combinations, we found borderline significance between the lack of CT

expression and the presence of both the MTHFR677 "CC or TT" genotype, and the RFC80 "GG or AA" genotype (**Table 16**); as well as when the MTHFR1298 "AC or CC" genotype was combined with the RFC80 "GG or AA" genotype (**Table 17**). This is in line with what we observed for individual genotype associations with CT gene expression, therefore, that hypomorphic genotypes in the 1-carbon pathway associate with the lack of CT gene expression in tumors.

Table 16. CT expression associations with 1-carbon enzyme genotype combinations in lung cancer patients I: MTHFR677 C>T and RFC80 G>A.

RFC80	MTHFR677	CT(+)	CT(-)	OR (95% CI)	p*
GG	CC	8	9	1.9 (0.51-7.05)	0.05
GG	CT/TT	7	15		
GA/AA	CC	4	0	-	0.3
GA/AA	CT/TT	2	5		

*Fisher's exact test (2-sided)

Table 17. CT expression associations with 1-carbon enzyme genotype combinations in lung cancer patients II: MTHFR1298 A>C and RFC80 G>A

RFC80	MTHFR1298	CT(+)	CT(-)	OR (95% CI)	p*
GG	AA	2	4	0.13 (0.01-2)	0.2
GG	AC/CC	4	1		
GA/AA	AA	10	8	4 (1.01-15.71)	0.06
GA/AA	AC/CC	5	16		

*Fishers exact test (2-sided)

3.5 One carbon enzyme genotype associations in lung cancer patients

We tested whether the presence of genotype correlated with that of another. When all lung cancer samples were evaluated, we observed a significant negative correlation between the MTHFR677 CC and MTRR66 GG genotypes, between the MTHFR1298 AA and MTHFR66 GG genotypes, and co-occurrence of RFC80 "GG or GA" genotype with MTHFR1298 AA, as well as with MTHFR677 CC (**Table 18**). A positive correlation was also present between the RFC80 AA genotype and MTRR2756 AA. When samples were stratified according to CT gene expression, the correlation between RFC80 AA and MTRR2756 AA genotypes was only observed among CT(+) samples (**Table 18**). Among the CT(-) samples, the RFC80 "GG or GA" correlation with MTHFR1298 AA was observed, in addition to a novel negative correlation between MTRR66 GG and RFC80 GA (**Table 18**).

Table 18. One carbon enzyme genotype associations in lung cancer patients

Genotype 1	Genotype 2	Correlation Coefficient (Spearman's rho)			p*
		All	CT(+)	CT(-)	
MTHFR677 CC	MTRR66 GG	-0.417	N.C.[&]	N.C.	0.006
MTHFR1298 AA	MTRR66 GG	-0.354	N.C	N.C.	0.02
MTHFR1298 AA	RFC80 GG/GA	0.45	N.C.	0.482	0.001/0.01[#]
MTRR2756 AA	RFC80 AA	0.338	0.63	N.C.	0.02/0.001
MTHFR677 CC	RFC80 GG/GA	0.281	N.C.	N.C.	0.05
MTRR66 GG	RFC80 GA	N.C.	N.C.	-0.452	0.03

*Fishers exact test (2-sided); [&]N.C.: no correlation; #p values correspond to the first and second correlation coefficients, respectively.

3.6 Univariate power analysis of CT expression and one carbon enzyme genotype associations

We then calculated the minimum sample size we would required to differenciate samples according to their CT gene expression status, if genotypes were studied as two groups. Using the Power and Precision software, we performed Wald's test and a univariate logistic regression model. It is important to point out that the polymorphisms are analyzed independently, regardless of the effect of other variants studied. Our analysis revealed that as few as 80 samples would help us distinguish samples by their CT expression pattern if they were classified into MTFF66 AA and "AG or GG" genotype groups. Similarly, including less than 200 samples in a prospective study is predicted to help distinguish samples according to their RFC80 as well as MTHFR677 polymorphisms (**Table 19**). It is important to point out that the polymorphisms are analyzed independently, regardless of the effect of other variants studied.

Table 19. Univariate power analysis*

Genotype	Event rates**	Sample size
MTRR66 AA vs. AG/GG	0.20 vs. 0.51	80
RFC80GG vs. GA/AA	0.67 vs. 0.42	128
MTHFR677 CC vs. CT/TT	0.29 vs. 0.51	159
MTHFR1298 AA vs. AC/CC	0.56 vs. 0.41	352
MTHFR2756 AA vs. AG/GG	0.45 vs. 0.50	3135

* Logistic regression model, Wald's test, α=0.05, power=80; **Probability of a sample being CT (+) for the given genotype

3.7 Multivariate power analysis of CT expression and one carbon enzyme genotype associations

The univariate power analysis model assumes that the distribution of genotypes of polymorphisms other than that for which the calculations are made occur at random frequencies. However, we already determined that genotypes co-exist among lung cancer samples, as well as within CT(+) and CT(-) groups, as explained above. We wanted to determine the minimal sample number that could distinguish CT (+) from CT(-) samples for a given genotype would be, if the others behaved similar to our observations. We, therefore, performed a multivariate power analysis for each genotype controlling for the others. The minimum sample sizes calculated are shown in **Table 20**. The beta variant in this table is defined as the logarithm of the odds ratio defined on page 47. Thus, the MTRR66 AA versus AG/GG genotype distribution has a coefficient of -1.43 and requires at least 132 samples to distinguish CT (+) and CT (-) samples, given the genotypes of other enzymes are held constant as indicated (**Table 20**) . The beta variant in this table is defined as the logarithm of the odds ratio defined on page 47. Thus, the MTRR66 AA versus AG/GG genotype distribution has a coefficient of -1.43 and requires at

least 132 samples to distinguish CT (+) and CT (-) samples, given the genotypes of other enzymes are held constant as indicated (**Table 20**) .

Table 20. Multivariate power analysis*

Genotype	Adjusting for	Beta	Sample size
MTRR66 AA MTRR66 AG/GG	MTHFR677=CC, MTHFR1298=AA, MTR2756=AA, RFC80=GG	-1.43	132
MTHFR677 CC MTHFR677 CT/TT	MTHFR1298=AA, MTR2756=AA, MTRR66=AA, RFC80=GG	1.02	162
RFC80GG RFC80GA/AA	MTHFR677=CC, MTHFR1298=AA, MTR2756=AA, MTRR66=AA	1.03	165
MTHFR1298 AA MTHFR1298 AC/CC	MTHFR677=CC, MTR2756=AA, MTRR66=AA, RFC80=GG	0.61	480
MTHFR2756 AA MTHFR2756 AG/GG	MTHFR677=CC, MTHFR1298=AA, MTRR66=AA, RFC80=GG	-0.2	5253

* Logistic regression model, Wald's test, α=0.05, power=80

Note that there is a difference between the sample sizes found in the univariate and the multivariate tests. This finding constitutes the backbone of our study, which argues that it is important to study these alleles comprehensively in order to overcome the possibility of missing the effect of a polymorphism which is disregarded.

From the results derived from the multivariate analysis mentioned above, we have, with an 80% power, calculated the standard errors and p-values in the condition where the sample size is large enough to find a significant difference. For example, if we studied ≥132 samples, we would be able to distinguish the CT (+) and CT (-) lung cancer samples by testing for MTRR 66 AA and AG/GG genotypes.

These statistical analyses give us an idea about the size of the sample and the genotypes to be studied in our future experiments. For instance, since the MTR 2756 genotype requires 5253 samples, it is very likely that we would not include the typing of this genotype in a larger epidemiologic study. Besides, we have observed that when the combined effects of the polymorphisms are taken into consideration, we obtain a different result when they are not collectively analyzed and this indicates the importance of studying these alleles collectively.

CHAPTER 4. DISCUSSION AND FUTURE PERSPECTIVES

The study conducted by Gure *et al* in 2005 revealed that CT genes were coordinately expressed and associated with poor prognosis [26]. Data from the same group and others showed that CT gene expression was primarily regulated by DNA methylation [24]. The outcome of these studies constituted the basis of the current study. S-adenosylmethionine is the one and only methyl donor in the cell and therefore it is likely that an event altering SAM production would affect CT expression. The polymorphisms in the one carbon pathway enzyme genes, which are thought to cause reductions in SAM production, have already been shown to correlate with DNA hypomethylation and increased SAH levels. Therefore, we decided to type lung cancer samples whose CT gene expression status had been identified previously for the one carbon pathway enzyme variants and check whether a given haplotype could predict the existence of CT expression so that further studies for silencing the CT expression to improve the prognosis can be conducted.

In this study, the typing of the polymorphisms was conducted by the *restriction fragment length polymorphism* (RFLP) method. In order to verify our results, we tried to reproduce our data by a number of methods. While a total of 127 experiments were performed only once, 36 experiments were repeated two times and 13 experiments were repeated three times or more. We found inconsistent results for a total of 9 samples. 18 samples within our lung cancer panel were re-typed by sequencing analysis. Besides the lung cancer samples, genomic DNA samples from cancer cell lines were also typed with the same method. A number of inconsistencies remain unresolved at presence. and are summarized in **Table 21**.

Table 21. Genotype typing inconsistencies

Genotype/Allele	*Experiment 1*	*Experiment 2*
MTHFR 677		
Sample #726	*C/T (RFLP)*	*C/C (RFLP)*
Sample #745	*C/T (RFLP)*	*T/T (sequencing)*
Sample #639	*C/T (RFLP)*	*C/C (sequencing)*
MTRR 66		
Sample #693	*A/A (RFLP)*	*G/G (RFLP)*
RFC 80		
SKCO1 cell line	*A/A (RFLP)*	*A/G (sequencing)*

For the MTHFR C677T genotype, with the RFLP analysis, we observed contradictory results for 3 samples and typing with sequencing analysis generated inconsistent data for 2 out of the 5 samples. There were 5 findings which we were unable to re-obtain by RFLP analysis for the MTRR A66G polymorphism and 2 discordant results optained by sequencing for the RFC G80A variant. These samples remain to be further analyzed.

The below experiment demonstrates and innate problem of RFLP, namely that incomplete enzymatic digestion can cause erroneous results (**Figure 4**). The indicated Lung cancer sampleswere subjected to *Hinp1I* digestion overnight, following PCR amplification. Lu-183 DNA digestion generated a 154 bp band (as well as a smaller 124bp band not visible on the figure) and therefore was classified as a GG homozygote. Lu 191 is a heterozygote, and so is Lu 726. However, despite the extended digestion time, Lu 718 and Lu 728 DNA digestion is most likely incomplete.If incomplete digestion took place for the other samples as well, we could be misinterpreting the results; for example all the above samples could be AA homozygotes. Despite several repeats Lu 726 typing remains unresolved.

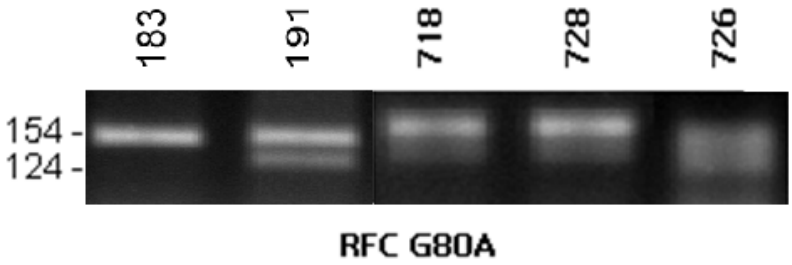


Figure 4. RFPL analysis of RFC G80A polymorphism.

Consequently, these results require us to reproduce our data by a more reliable and ideally, a high-throughput method. One dependable and relatively new technique for SNP analysis is the probe based real time quantitative polymerase chain reaction (Q-PCR) [72].

Q-PCR allows quantification of polymorphic DNA regions and genotyping of single nucleotide polymorphisms in one run. The RFLP method, which involves 3 steps, requires the amplification of the region around a given SNP and digestion of the PCR product by a restriction enzyme with ensuing gel electrophoresis. However, with a single Q-PCR run, both quantification and genotyping can be performed simultaneously. Online monitoring of the amplification process as well as genotyping by melting curve analysis is possible by the use of hybridization probes. The hybridization probes are sequence-specific oligonucleotides labeled by fluorescence dyes. For the genotyping of the SNP, two hybridization probes are required. One covers the polymorphic region (sensor hybridization probe) and carries a florescent dye that emits green light. The second probe (the anchor probe) binds to a site close in proximity to the sensor hybridization probe and carries a red light emitting dye. During the amplification process, the hybridization probe anneals to the amplified DNA and emits green light, which then excites the red dye in the anchor probe. The energy transferred from the sensor probe to the anchor probe is called *florescence resonance energy transfer* (FRET). The FRET signal, which is detected by the Q-PCR machine, is a direct measure of the DNA copy.

After amplification, the SNP's can be detected by a melting-curve analysis. At this stage, the temperature is raised from 40^{o}C to 75^{o}C. Then, the temperature at which the hybridization

probes are melted off the DNA strand (T_m), which is an indicator of the presence or the absence of the mutation, is calculated. If the hybridization probe fits perfectly to the template DNA, then it will have a higher T_m. If there is a single nucleotide mismatch, then the probe will melt at a lower temperature. The melting curve analysis can yield three results. One is a curve with a single early peak, second is a curve with a single late peak, and the last one is a curve with two peaks. The results obtained by this method are quite reliable that the melting points for a specific SNP is always at the same point. However, if there is another SNP within the region of the hybridization probe binding site, then the melting points for each genotype will deviate. A deviation more than 1.5^{o}C in the melting curve step, is an indication of an additional mutation within the binding region of the hybridization probe [73]. For instance, a recent study has shown that an additional mutation causes deviations in the detection of the MTHFR A1298C SNP [72]. An exemplary graph shows the possible alleles and an unexpected nearby mutation curves (**Figure 5**). Notice the heterozygote allele has a lower florescence value than the homozygote alleles.

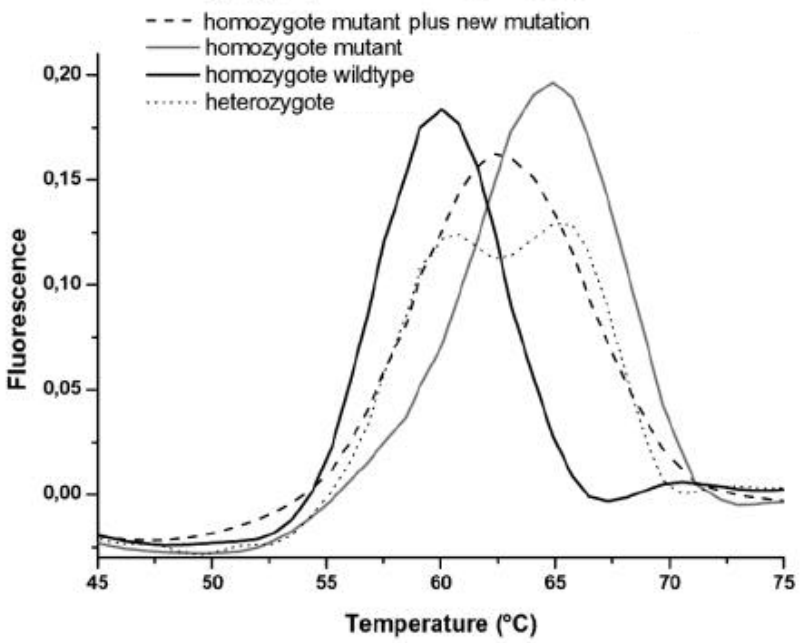


Figure 5. An exemplary RT-PCR melting-curve analysis showing the possible genotypes with an additional unexpected mutation.

Thus quantitative Q-PCR offers several advantages over RFLP, However, given the fact that there are a significantly large number of SNPs identified for each enzyme; sequencing analysis would offer further reliability.

Despite the suggestion that enzyme digestion might be incomplete in a number of experiments of ours, another possible explanation that can explain variations of band intensities we consistently observed for some samples is that there might be a copy number variation (CNV) in the genes we studied. What directed us to this conclusion was that upon the digestion procedure of the MTRR 66 polymorphism, there was a greater intensity of the G allele than the A allele in some of the heterozygote samples. This is demonstrated in **Figure 6**. The figure demonstrates the RFLP results obtained for the MTRR A66G allele. The three bands obtained for Lu 706 is equal in intensity and correspond to bands expected for a heterozygote (66, 44 and 22bp). The 44bp band is significantly more intense than the 66bp band. However, for Lu 706, 713, 716 and 658, the 66bp band is significantly more intense than the 44bp and 22bp bands. It is unlikely that incomplete digestion accounts for these, because this experiment was repeated 4 times with equal results. If alleles are amplified in the genome resulting in CNV, then the subsequent RFLP analysis would be expected to generate a result as shown in Figure 6.

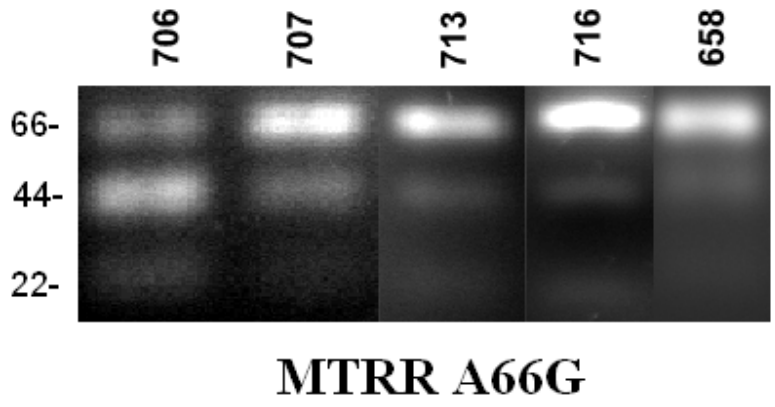


Figure 6. A possible copy number variation in the MTRR A66G polymorphism. The "G" band is significantly more intense than the "A" band in the 2nd, 3rd, 4th and 5th lanes whereas the intensity is equal in the 1st lane.

It is important to know whether there is a copy number variation (CNV) of these genes and this should be investigated by the Q-PCR method explained above or CNV microarray studies. The amplification step could reveal, the increase in the copy number of the one carbon pathway genes. If the presence of a copy number variation in the MTRR gene and in other genes is found, this possibly might have occurred towards the compensation of a reduced methyl group flow in the one carbon pathway. For instance, if the patient has a MTHFR 677 hypoactive allele (T), the cell could be amplifying the MTRR normoactive (C) allele in order to compensate for methionine synthesis through an increased activation of MTR, in the presence of low production of N^5-methyl-THF caused by the MTHFR 677 T allele. Despite the fact that methionine synthesis is N^5-methyl-THF – dependent, the amplification might still have a functional consequence. This could be novel a mechanism by which the cell maintains a specific rate of SAM production in the presence of certain mutations as there are no publications documenting CNV for these genes to date.

Our analysis of the correlations between the 1 –carbon pathway enzyme genotypes and CT expression profiles of the lung cancer genomic DNA samples demonstrated important contradictions. First of all, we found out that the CT (-) tumor samples had a higher proportion of

the MTHFR 677 T allele when compared to the CT (+) samples (see Table 21, Chapter 3). This is difficult to reconcile with our hypothesis that the CT (+) samples have a higher proportion of this allele if their expression is due to inefficient SAM production. Moreover, despite not being statistically significant, 37.9% of the CT (-) were heterozygous for MTHFR 677, while only 28.6% of CT (+) samples were heterozygous.

In line with these results we also observed that for the MTRR A66G polymorphism. MTRR 66 GG allele, thought to code for the hypoactive enzyme, was present among CT (-) lung cancer samples at a rate that was five-fold more of what we observed for the CT (+) samples. As explained above, because we hypothesized that the CT (+) would have the greater frequency of hypoactive alleles, this data obtained appeared to be just the opposite of our initial hypothesis.

Our data demonstrates a strong negative correlation between MTHFR 677 CC and MTRR 66 GG alleles, normoactive variants of these two enzymes (MTHFR 677 CC and MTRR 66 AG/AA); as well as between hypoactive variants (MTHFR 677 CT/TT and MTRR 66 GG). This would further suggest that among lung cancer patients, there are two haplotypes, one prone to efficient and the other prone to inefficient SAM production. Even though this is a plausible possibility, it makes it even more difficult to explain why the inefficient SAM producers should preferentially have CT (-) tumors. Therefore, I offer the following hypothesis that might possibly help understand this dilemma.

It is known that SAM is used for the methylation of several substrates including RNA, DNA and proteins. Some of the most important substrates of SAM are, therefore the histone proteins. Histones are modified at specific residues in their N-terminal chains. These modifications include phosphorylation, sumoylation, acetylation, ubiquitination as well as methylation. The effect of each modification is being actively studied [74]. Methylation is one of the most critical modifications regulating of histone function. Mono, di- or tri- methylation of especially the 4^{th}, 9^{th} 27^{th} and 79^{th} lysine residues of H3 closely associate with repressed or activated chromatin [75]. It has been shown that the modifications associated with active chromatin include H3K4me2, HeK4me3, H3K36me3, while those associated with repressed chromatin include H3K4me1, H3K36me2, H3K9me1 and H3K79me1 {Li, 2007}.

Methylthioadenosine (MTA), a nucleoside produced in the polyamine biosynthetic pathway by the cleavage of SAM, is a known inhibitor of methyltransferases {Mato, 2007}. MTA has been shown to inhibit H3K4 methylation [76], [77] by inhibiting Set1 methyltransferase [78]. In line with this observation, it has been shown that the addition of SAM to the RAW cell line inhibited LPS-induced TNF-α gene expression *in vitro* {Ara, 2008}. In the same study, it has also been observed by ChIP experiments that upon SAM treatment, H3K4 tri-methylation is decreased compared to the LPS treated cells alone. H3K4 tri-methylation, as explained before, is a hallmark of transcriptional activation. It has also been found that the treatment of cells with SAM or MTA reduced the levels of H3K4me1 and H3K4me2, which suggested that these agents might be playing a role in inhibiting the enzyme responsible for the trimethylation of H3K4. Same observations were also obtained by a different study. Huang *et al* have observed that MTA decreased the levels of both H3K4me2 and H3K4me3 [78]. MTA is known to inhibit SAH hydrolase, the enzyme which converts SAH to homocysteine [79], and SAH itself is a strong competitive inhibitor of almost all SAM-dependent methyltransferases [82]. The SAH levels must be kept in check because many methyltransferases have a higher affinity for SAH than SAM, and this makes SAH, as just mentioned, a potent inhibitor of many methylation reactions [82].

Apart from affecting H3K4 methylation, MTA might also affects H3K9 methylation since it is known that the G9a, the H3K9 methyltransferase is a SAM-dependent methyltransferase. MTA, which reduces the level of H3K4 methylation, might also influence H3K9 methylation. In addition to this, Mutskov *et al* conducted a study in which they showed that H3K9 methylation can occur prior to DNA methylation [80].

Considering all of these observations, I have created a model in which CT expression is primarily regulated by histone modifications, which are affected by SAM concentrations.

In this model, if the CT (-) lung cancer samples have a higher frequency of hypoactive alleles of MTHFR 677 and MTRR 66, resulting in less efficient SAM production when compared to CT (+) lung cancer samples. The inefficient production of SAM will yield decreased amounts of its metabolite, MTA. The decreased amount of MTA will allow H3K4 and

H3K9 methylations which in turn result in DNA methylation and the CT genes will be silenced (Figure 7).

On the other side, the wild-type alleles of MTHFR 677 and MTRR 66 genes will result in an adequate production of SAM and consequently MTA. Because MTA is itself an inhibitor of methyltransferases and SAH hydrolase, SAH levels will increase. Since SAH is a strong inhibitor of almost all SAM dependent methyltransferases, along with MTA, they will have a suppressive effect on histone methylation. This would result in a reduction of H3K4 trimethylation and H3K9 methylation. This will lead to disturbances in DNA methylation and CT expression since H3K9 will not be methylated (**Figure 7**).

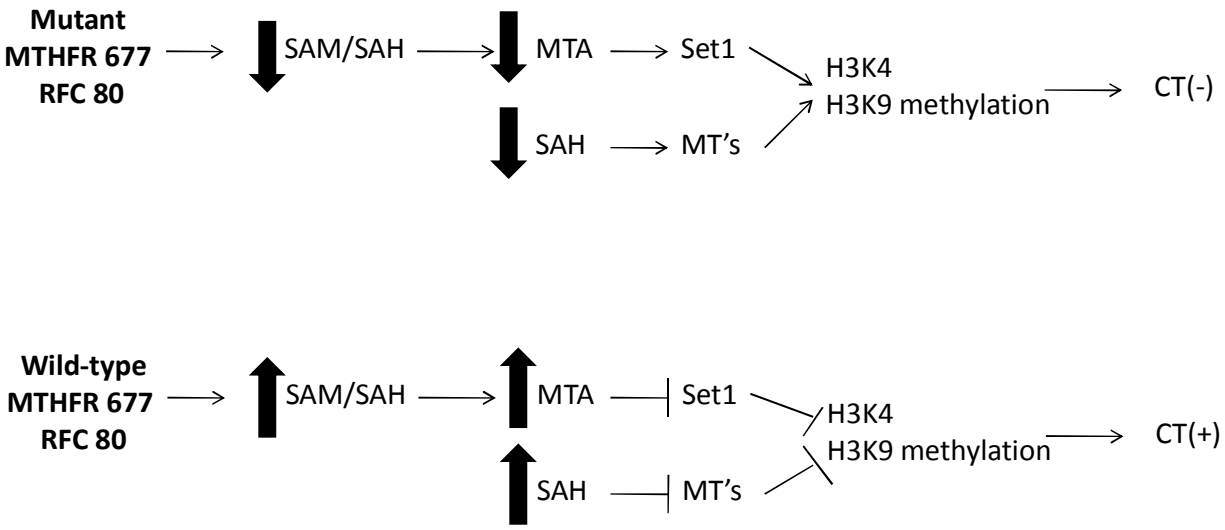


Figure 7. Proposed model adopted from the results of this study for the regulation of CT genes

As a result, a new insight about CT gene expression could emerge. In order to test this model, we first have to start by checking whether different histone methyltransferases have the same sensitivity for SAM. This aspect of histone methyltransferases is important because different sensitivities would mean that they will respond to different SAM concentrations by distinct ways, which would eventually affect the activation/suppression of the target genes. There is a procedure of determining the enzymatic activity by a method called *peptide methylation assay*. In this assay described by Rathert *et al* [81], just like the procedure in ELISA experiments, the wells on the plate are coated with a target peptide. After the addition of the desired methyltransferases and SAM mixture into the wells, a continuous read-out of the reaction process is performed. As the methyltransferase methylates the target peptide, light of a certain

wavelength is emitted and the amount of light is read and quantified. The time versus emission graph will give us the kinetic activity of the enzyme. Then, ChIP experiments could be performed in a given haplotype to see whether the polymorphisms in the one carbon enzyme genes would have an effect on the histone modifications in CT genes.

In order to regard the effect of these polymorphisms thoroughly, we have to focus on folic acid and DNA damage. As mentioned earlier, the deficiency of folic acid reduces the substrate flow for the MTHFR enzyme and the SAM production is impaired due to inefficient methionine production. Folic acid deficiency has been correlated with DNA hypomethylation. On the other hand, thymidylate synthase prevents uracil misincorporation by converting uracil into thymidine.

Another possible mechanism for the explanation of a model which could explain our results belies on the assumption that different methyltransferases have different sensitivities to SAM. In this model, because CT (-) samples have more frequent mutant genotypes, the SAM production will be inefficient. If some methyltransferases are more sensitive to SAM concentrations than others, the histone proteins might be preferentially modified. For instance, if Set1, a H3K4 methyltransferase requires higher SAM concentrations for its optimum performance than Ga9, a H3K9 methyltransferase, then preferential downregulation of H3K4 methylation, in low SAM concentrations, might result in a relatively high H3K9 methylation and thus, repression of gene transcription. On the other hand, in CT (+) samples generating adequate SAM concentrations, comparable H3K4- and H3K9-methyltransferase activities might result in preferential H3K4 tri-methylation resulting in gene activation. **Figure 8** summarizes this model.

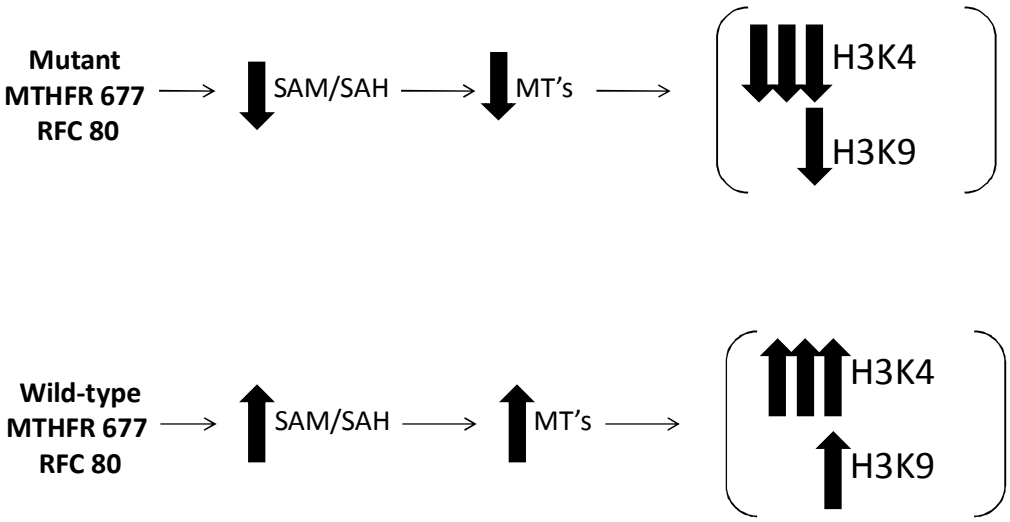


Figure 8. Proposed model for the regulation of CT genes adopted from the results of this study

For a better understanding of CT gene regulation and the effect of one carbon pathway enzyme gene polymorphisms, we will pursue our studies by selecting two cell lines with a certain a genotype which we know is associated with CT positivity. Then, by culturing these cell lines in a folic acid deficient medium we'll induce DNA damage. The DNA damage will be evaluated by using the comet assay method, which is used for the detection of single and double strand breaks on the DNA. The folic acid deficiency would result in decreased amounts of 5-methylene-THF, which is a substrate for both MTHFR and TS. So, we would be able to see the difference between a MTHFR 677 CC and MTHFR 677 TT genotype for the utilization of the 5-

methylene-THF in the methylation direction. The impaired MTHFR 677 TT genotypes are expected to have a lesser degree of DNA damage because the thymidylate synthase enzyme will have a larger amount of 5-methylene-THF for the conversion of dUMP to dTMP, preventing uracil misincorporation. Moreover, by the addition of SAM into the cell culture, the expression of CT genes will be observed.

In our subsequent studies, we might also focus on the methyltransferases and the metabolism of SAM in order to further investigate the relationship between the production of SAM, MTA and histone modifications. For this reason, we have pursued a preliminary micro-array meta analysis from data compiled from 17 independent datasets corresponding to microarray data obtained from different tumors.. **Table 22** lists the genes which are significantly up or downregulated, the probes and the number of experiments along with the average fold change.

Table 22. The microarray meta-analysis showing the average fold changes of gene expression for the listed genes in cancer.

Gene	Probe	Direction	Number of experiments	Average Fold change
TS	1	up	2	2.04
TS	2	up	11	3.44
MTR		down	5	1.53
MTRR	1	down	4	1.77
MTRR	2	up	1	1.84
DHFR	1	up	2	1.67
DHFR	2	up	7	1.92
DHFR	3	up	6	1.96
DHFR	4	up	5	1.78
Methionine adenosyltransferase		up	5	1.96
Adomet decarboxylase	1	up	7	2.68
Adomet decarboxylase	2	up	5	1.92
Spermine synthase		up	3	1.76
Spermidine synthase		up	8	1.98
Ornithine decarboxylase		up	10	2.67
5'methyladenosine phosphorylase	1	up	1	1.65
5'methyladenosine phosphorylase	2	up	1	1.52

It can be inferred from the Table 16 that different enzymes which we are not focused on might play important roles in SAM utilization and folate metabolism. The cancer cells may be upregulating these genes in order to compensate the deficiency of their substrates. Along with the thymidylate synthase expression data, obtained by Q-PCR, we can regard that different cancer cell lines have diverse TS expression values. For instance, the LS174T cell line has almost 6 times TS expression than the HCT15 cell line. These differences direct us to the one carbon pathway enzyme gene polymorphisms and the expression levels of methyltransferases of these two cell lines. These aspects of the pathway can be included in future studies.

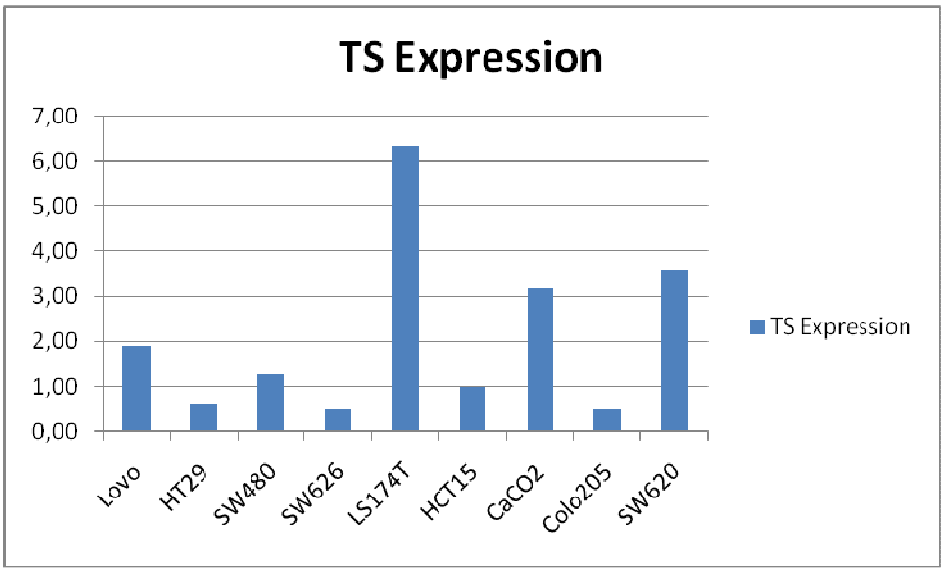


Figure 9. Q-PCR results showing the expression levels of thymidylate synthase in different cancer cell lines.

Differential TS expression in different cell lines might indicate a mechanism to compensate for a hypoactive genotype of MTHFR 677 or severe DNA damage caused by uracil

misincorporation. Thus, as explained earlier [31], different cell lines might have distinct mechanisms to maintain an efficient SAM production and DNA synthesis rates.

As a result, this study has shown that there might be a mechanistic bridge between the production of SAM and DNA methylation, which affects cancer-testis gene expression. The studies of folic acid intake, TS expression and activity should be pursued in order to regard their effects on SAM production and especially histone methyltransferases. This study has helped us gain new insight into the potential interplay between two mechanisms, the one carbon pathway and epigenetic modification, and has made it possible to envision the correct experiments that would be needed to elucidate this new aspect of CT gene regulation.

CHAPTER 5. REFERENCES

1. Kim, Y.I., *Folate and DNA methylation: a mechanistic link between folate deficiency and colorectal cancer?* Cancer Epidemiol Biomarkers Prev, 2004. **13**(4): p. 511-9.
2. Jaenisch, R. and A. Bird, *Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals.* Nat Genet, 2003. **33 Suppl**: p. 245-54.
3. Laird, P.W. and R. Jaenisch, *The role of DNA methylation in cancer genetic and epigenetics.* Annu Rev Genet, 1996. **30**: p. 441-64.
4. Turek-Plewa, J. and P.P. Jagodzinski, *The role of mammalian DNA methyltransferases in the regulation of gene expression.* Cell Mol Biol Lett, 2005. **10**(4): p. 631-47.
5. Regha, K., et al., *Active and repressive chromatin are interspersed without spreading in an imprinted gene cluster in the mammalian genome.* Mol Cell, 2007. **27**(3): p. 353-66.
6. Ginder, G.D., M.N. Gnanapragasam, and O.Y. Mian, *The role of the epigenetic signal, DNA methylation, in gene regulation during erythroid development.* Curr Top Dev Biol, 2008. **82**: p. 85-116.
7. Metivier, R., et al., *Cyclical DNA methylation of a transcriptionally active promoter.* Nature, 2008. **452**(7183): p. 45-50.
8. Margot, J.B., A.E. Ehrenhofer-Murray, and H. Leonhardt, *Interactions within the mammalian DNA methyltransferase family.* BMC Mol Biol, 2003. **4**: p. 7.
9. Brown, R. and G. Strathdee, *Epigenomics and epigenetic therapy of cancer.* Trends Mol Med, 2002. **8**(4 Suppl): p. S43-8.
10. Agrawal, A., R.F. Murphy, and D.K. Agrawal, *DNA methylation in breast and colorectal cancers.* Mod Pathol, 2007. **20**(7): p. 711-21.
11. Narayan, A., et al., *Hypomethylation of pericentromeric DNA in breast adenocarcinomas.* Int J Cancer, 1998. **77**(6): p. 833-8.
12. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.
13. Esteller, M., *Cancer epigenetics: DNA methylation and chromatin alterations in human cancer.* Adv Exp Med Biol, 2003. **532**: p. 39-49.

14. Guerrero-Preston, R., et al., *Global DNA hypomethylation in liver cancer cases and controls: a phase I preclinical biomarker development study*. Epigenetics, 2007. **2**(4): p. 223-6.
15. Chen, Y.T., et al., *Identification of multiple cancer/testis antigens by allogeneic antibody screening of a melanoma cell line library.* Proc Natl Acad Sci U S A, 1998. **95**(12): p. 6919-23.
16. Lengauer, C., K.W. Kinzler, and B. Vogelstein, *DNA methylation and genetic instability in colorectal cancer cells.* Proc Natl Acad Sci U S A, 1997. **94**(6): p. 2545-50.
17. Seng, T.J., et al., *DLEC1 and MLH1 promoter methylation are associated with poor prognosis in non-small cell lung carcinoma.* Br J Cancer, 2008. **99**(2): p. 375-82.
18. Kang, G.H., et al., *Profile of aberrant CpG island methylation along multistep gastric carcinogenesis.* Lab Invest, 2003. **83**(4): p. 519-26.
19. Kanai, Y. and S. Hirohashi, *Alterations of DNA methylation associated with abnormalities of DNA methyltransferases in human cancers during transition from a precancerous to a malignant state.* Carcinogenesis, 2007. **28**(12): p. 2434-42.
20. Mohn, F., et al., *Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors.* Mol Cell, 2008. **30**(6): p. 755-66.
21. Tuck-Muller, C.M., et al., *DNA hypomethylation and unusual chromosome instability in cell lines from ICF syndrome patients.* Cytogenet Cell Genet, 2000. **89**(1-2): p. 121-8.
22. Scanlan, M.J., et al., *Cancer/testis antigens: an expanding family of targets for cancer immunotherapy*. Immunol Rev, 2002. **188**: p. 22-32.
23. Ross, M.T., et al., *The DNA sequence of the human X chromosome.* Nature, 2005. **434**(7031): p. 325-37.
24. De Smet, C., et al., *DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter.* Mol Cell Biol, 1999. **19**(11): p. 7327-35.
25. Wischnewski, F., K. Pantel, and H. Schwarzenbach, *Promoter demethylation and histone acetylation mediate gene expression of MAGE-A1, -A2, -A3, and -A12 in human cancer cells.* Mol Cancer Res, 2006. **4**(5): p. 339-49.

26. Gure, A.O., et al., *Cancer-Testis Genes Are Coordinately Expressed and Are Markers of Poor Outcome in Non-Small Cell Lung Cancer.* Clin Cancer Res, 2005. **11**(22): p. 8055-8062.
27. Finkelstein, J.D., *Metabolic regulatory properties of S-adenosylmethionine and S-adenosylhomocysteine.* Clin Chem Lab Med, 2007. **45**(12): p. 1694-9.
28. Pogribny, I.P., et al., *Irreversible global DNA hypomethylation as a key step in hepatocarcinogenesis induced by dietary methyl deficiency.* Mutat Res, 2006. **593**(1-2): p. 80-7.
29. Stempak, J.M., et al., *Cell and stage of transformation-specific effects of folate deficiency on methionine cycle intermediates and DNA methylation in an in vitro model.* Carcinogenesis, 2005. **26**(5): p. 981-90.
30. Lucock, M., *Folic acid: nutritional biochemistry, molecular biology, and role in disease processes.* Mol Genet Metab, 2000. **71**(1-2): p. 121-38.
31. Hayashi, I., et al., *Folate deficiency induces cell-specific changes in the steady-state transcript levels of genes involved in folate metabolism and 1-carbon transfer reactions in human colonic epithelial cells.* J Nutr, 2007. **137**(3): p. 607-13.
32. Selhub, J., *Homocysteine metabolism.* Annu Rev Nutr, 1999. **19**: p. 217-46.
33. Huang, S., *Histone methyltransferases, diet nutrients and tumour suppressors.* Nat Rev Cancer, 2002. **2**(6): p. 469-76.
34. Xu, X., et al., *Polymorphisms of one-carbon-metabolizing genes and risk of breast cancer in a population-based study.* Carcinogenesis, 2007. **28**(7): p. 1504-9.
35. Blount, B.C., et al., *Folate deficiency causes uracil misincorporation into human DNA and chromosome breakage: implications for cancer and neuronal damage.* Proc Natl Acad Sci U S A, 1997. **94**(7): p. 3290-5.
36. Wickramasinghe, S.N. and S. Fida, *Bone marrow cells from vitamin B12- and folate-deficient patients misincorporate uracil into DNA.* Blood, 1994. **83**(6): p. 1656-61.
37. Crott, J.W., et al., *Moderate folate depletion modulates the expression of selected genes involved in cell cycle, intracellular signaling and folate uptake in human colonic epithelial cell lines.* J Nutr Biochem, 2008. **19**(5): p. 328-35.

38. Branda, R.F., et al., *Effects of folate deficiency on the metastatic potential of murine melanoma cells.* Cancer Res, 1988. **48**(16): p. 4529-34.
39. Melnyk, S., et al., *Uracil misincorporation, DNA strand breaks, and gene amplification are associated with tumorigenic cell transformation in folate deficient/repleted Chinese hamster ovary cells.* Cancer Lett, 1999. **146**(1): p. 35-44.
40. Rodriguez, J., et al., *Chromosomal instability correlates with genome-wide DNA demethylation in human primary colorectal cancers.* Cancer Res, 2006. **66**(17): p. 8462-9468.
41. Yuan, J.M., et al., *Genetic polymorphisms in the methylenetetrahydrofolate reductase and thymidylate synthase genes and risk of hepatocellular carcinoma.* Hepatology, 2007. **46**(3): p. 749-58.
42. Martin, Y.N., et al., *Human methylenetetrahydrofolate reductase pharmacogenomics: gene resequencing and functional genomics.* Pharmacogenet Genomics, 2006. **16**(4): p. 265-77.
43. Frosst, P., et al., *A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase.* Nat Genet, 1995. **10**(1): p. 111-3.
44. Heijmans, B.T., et al., *A common variant of the methylenetetrahydrofolate reductase gene (1p36) is associated with an increased risk of cancer.* Cancer Res, 2003. **63**(6): p. 1249-53.
45. Castro, R., et al., *5,10-methylenetetrahydrofolate reductase (MTHFR) 677C-->T and 1298A-->C mutations are associated with DNA hypomethylation.* J Med Genet, 2004. **41**(6): p. 454-8.
46. Friso, S., et al., *The MTHFR 1298A>C polymorphism and genomic DNA methylation in human lymphocytes.* Cancer Epidemiol Biomarkers Prev, 2005. **14**(4): p. 938-43.
47. Botto, N., et al., *Genetic polymorphisms in folate and homocysteine metabolism as risk factors for DNA damage*. Eur J Hum Genet, 2003. **11**(9): p. 671-8.
48. Sharp, L. and J. Little, *Polymorphisms in genes involved in folate metabolism and colorectal neoplasia: a HuGE review.* Am J Epidemiol, 2004. **159**(5): p. 423-43.

49. Stern, L.L., et al., *Genomic DNA hypomethylation, a characteristic of most cancers, is present in peripheral leukocytes of individuals who are homozygous for the C677T polymorphism in the methylenetetrahydrofolate reductase gene.* Cancer Epidemiol Biomarkers Prev, 2000. **9**(8): p. 849-53.
50. Relton, C.L., et al., *Gene-gene interaction in folate-related genes and risk of neural tube defects in a UK population.* J Med Genet, 2004. **41**(4): p. 256-60.
51. Olteanu, H., T. Munson, and R. Banerjee, *Differences in the efficiency of reductive activation of methionine synthase and exogenous electron acceptors between the common polymorphic variants of human methionine synthase reductase.* Biochemistry, 2002. **41**(45): p. 13378-85.
52. Devlin, A.M., et al., *Effect of Mthfr genotype on diet-induced hyperhomocysteinemia and vascular function in mice.* Blood, 2004. **103**(7): p. 2624-9.
53. Christensen, B., et al., *Correlation of a common mutation in the methylenetetrahydrofolate reductase gene with plasma homocysteine in patients with premature coronary artery disease.* Arterioscler Thromb Vasc Biol, 1997. **17**(3): p. 569-73.
54. Leclerc, D., et al., *Human methionine synthase: cDNA cloning and identification of mutations in patients of the cblG complementation group of folate/cobalamin disorders.* Hum Mol Genet, 1996. **5**(12): p. 1867-74.
55. Chango, A., et al., *A polymorphism (80G->A) in the reduced folate carrier gene and its associations with folate status and homocysteinemia.* Mol Genet Metab, 2000. **70**(4): p. 310-5.
56. Whetstine, J.R., et al., *Single nucleotide polymorphisms in the human reduced folate carrier: characterization of a high-frequency G/A variant at position 80 and transport properties of the His(27) and Arg(27) carriers.* Clin Cancer Res, 2001. **7**(11): p. 3416-22.
57. Horie, N., et al., *Functional analysis and DNA polymorphism of the tandemly repeated sequences in the 5'-terminal regulatory region of the human gene for thymidylate synthase.* Cell Struct Funct, 1995. **20**(3): p. 191-7.

58. Trinh, B.N., et al., *Thymidylate synthase: a novel genetic determinant of plasma homocysteine and folate levels.* Hum Genet, 2002. **111**(3): p. 299-302.
59. Suzuki, T., et al., *Impact of one-carbon metabolism-related gene polymorphisms on risk of lung cancer in Japan: a case control study.* Carcinogenesis, 2007. **28**(8): p. 1718-25.
60. Ulrich, C.M., et al., *Polymorphisms in the reduced folate carrier, thymidylate synthase, or methionine synthase and risk of colon cancer.* Cancer Epidemiol Biomarkers Prev, 2005. **14**(11 Pt 1): p. 2509-16.
61. Markus, H.S., et al., *A common polymorphism in the methylenetetrahydrofolate reductase gene, homocysteine, and ischemic cerebrovascular disease.* Stroke, 1997. **28**(9): p. 1739-43.
62. Paz, M.F., et al., *Germ-line variants in methyl-group metabolism genes and susceptibility to DNA methylation in normal tissues and human primary tumors.* Cancer Res, 2002. **62**(15): p. 4519-24.
63. Friso, S., et al., *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, 2002. **99**(8): p. 5606-11.
64. Vaughn, J.D., et al., *Methionine synthase reductase 66A->G polymorphism is associated with increased plasma homocysteine concentration when combined with the homozygous methylenetetrahydrofolate reductase 677C->T variant.* J Nutr, 2004. **134**(11): p. 2985-90.
65. Moore, L.E., et al., *Folate metabolism genes, vegetable intake and renal cancer risk in central Europe.* Int J Cancer, 2008. **122**(8): p. 1710-5.
66. Robertson, K.D. and A.P. Wolffe, *DNA methylation in health and disease.* Nat Rev Genet, 2000. **1**(1): p. 11-9.
67. Jones, P.A. and J.D. Buckley, *The role of DNA methylation in cancer.* Adv Cancer Res, 1990. **54**: p. 1-23.
68. del Mazo, J., et al., *DNA methylation changes during mouse spermatogenesis.* Chromosome Res, 1994. **2**(2): p. 147-52.

69. Shi, Q., et al., *Sex differences in risk of lung cancer associated with methylene-tetrahydrofolate reductase polymorphisms.* Cancer Epidemiol Biomarkers Prev, 2005. **14**(6): p. 1477-84.
70. Pfaffl, M.W., *A new mathematical model for relative quantification in real-time RT-PCR.* Nucleic Acids Res, 2001. **29**(9): p. e45.
71. Tanaka, S., et al., *O6-methylguanine-DNA methyltranspherase gene expression in gliomas by means of real-time quantitative RT-PCR and clinical response to nitrosoureas.* Int J Cancer, 2003. **103**(1): p. 67-72.
72. Allen, R.A., et al., *A common 1317TC polymorphism in MTHFR can lead to erroneous 1298AC genotyping by PCR-RE and TaqMan probe assays.* Genet Test, 2007. **11**(2): p. 167-73.
73. Reuter, M., et al., *Detection of new single nucleotide polymorphisms by means of real time PCR.* J Genet, 2005. **84**(3): p. 341-5.
74. Kouzarides, T., *Chromatin modifications and their function.* Cell, 2007. **128**(4): p. 693-705.
75. Barski, A., et al., *High-resolution profiling of histone methylations in the human genome.* Cell, 2007. **129**(4): p. 823-37.
76. Chau, C.M. and P.M. Lieberman, *Dynamic chromatin boundaries delineate a latency control region of Epstein-Barr virus.* J Virol, 2004. **78**(22): p. 12308-19.
77. Song, M.R. and A. Ghosh, *FGF2-induced chromatin remodeling regulates CNTF-mediated gene expression and astrocyte differentiation.* Nat Neurosci, 2004. **7**(3): p. 229-35.
78. Huang, J., et al., *Trimethylation of histone H3 lysine 4 by Set1 in the lytic infection of human herpes simplex virus 1.* J Virol, 2006. **80**(12): p. 5740-6.
79. Ferro, A.J., A.A. Vandenbark, and M.R. MacDonald, *Inactivation of S-adenosylhomocysteine hydrolase by 5'-deoxy-5'-methylthioadenosine.* Biochem Biophys Res Commun, 1981. **100**(2): p. 523-31.
80. Mutskov, V. and G. Felsenfeld, *Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9.* Embo J, 2004. **23**(1): p. 138-49.

81. Rathert, P., X. Cheng, and A. Jeltsch, *Continuous enzymatic assay for histone lysine methyltransferases*. Biotechniques, 2007. **43**(5): p. 602, 604, 606 passim.

82. Clarke S, Banfield K. *S-adenosylmethionine-dependent methyltransferases: potential targets in homocysteine-linked pathology*. In: Carmel R, Jacobsen D, eds. Homocysteine in Health and Disease. Cambridge, England: Cambridge University Press; 2001:63-78.

83. Hosmer W, Lameshow S. *Applied Logistic Regression.* Wiley Series; 2000

84. Davis, C. D. and E. O. Uthus (2004). "DNA methylation, cancer susceptibility, and nutrient interactions." Exp Biol Med (Maywood) **229**(10): 988-95.

SUPPLEMENTARY FIGURES

Table 23 (Supplementary Table 1). The genotype data of the CT (+) lung cancer patients

Lu	*MTHFR C677T*	*MTHFR A1298C*	*MTR A2756G*	*MTRR A66G*	*RFC G80A*
31	CT	AC	AA	AG	AA
68	CT	AA	AG	GG	AG
87	CC	AC	AA	GG	GG
89	CT	AC	AG	AG	AG
111	CC	AC	AA	AA	GG
131	CC	AA	AA	AG	GG
168	CC	AA	AG	AG	AG
185	CT	AA	AG	AG	AG
186	CC	AC	AA	AG	GG
219	CT	AA	AA	GG	GG
223	TT	AA	AA	GG	AG
649	CC	CC	AG	GG	AA
652	CC	AA	AG	AG	AA
658	CC	AC	AG	AG	AG
726	CC	AC	AA	AG	AG
736	CC	AA	AG	AG	AG
739	TT	AA	AG	GG	AG
745	CT	AC	AA	GG	GG
752	CC	AA	AA	AG	AA
753	CC	AA	AG	AG	AA
759	CT	AA	AG	AA	AG

Table 24 (Supplementary Table 2). The genotype data of the CT (-) lung cancer patients

Lu	*MTHFR C677T*	*MTHFR A1298C*	*MTR A2756G*	*MTRR A66G*	*RFC G80A*
69	CT	AC	AA	GG	AA
77	TT	AA	AA	AG	GG
88	CT	AA	AG	GG	GG
90	CT	AA	AA	AA	AG
108	CC	AC	AG	AG	AG
110	CT	CC	AA	GG	AA
112	CT	AC	AA	GG	AG
180	CT	AC	AG	AA	AG
183	CT	AC	AA	AG	AA
191	CC	AC	AA	AG	AG
221	TT	AA	AG	GG	GG
225	CC	AC	AA	AA	AG
639	CT	AC	AA		GG
656	TT	AA	AA		GG
670	TT	AA	AG	AG	AG
692	CT	AA	AG		AG
693	TT	AC	AG	GG	AG
694	TT	AC	AG		AG
698	CC	AC	AG	AG	AG
706	CC	AA	AG		AA
707	CT	AC	AG	AG	AA
713	CC	AC	AG	AG	AG
716	TT	AC	AG	AG	AG
718	CT	AA	AA	AA	AG
728	CC	AC	AA	AA	AG
748	CT	AA	AA	GG	AG
749	CC	AA	AA	AG	AG
751	TT	AA	AA	GG	AG
763	CC	CC	AA	GG	AA

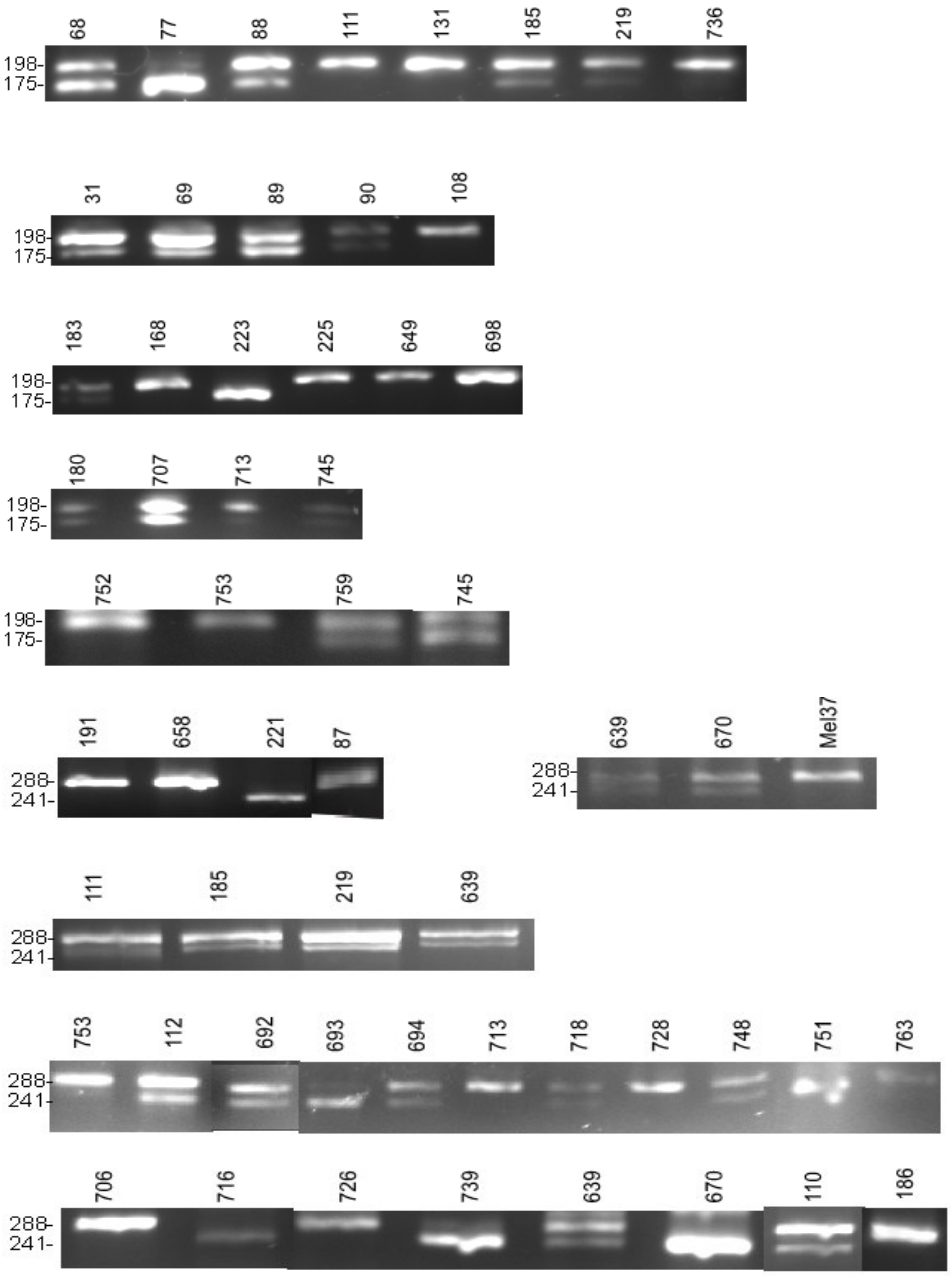


Figure 10 (Supplementary Figure 1). RFLP results of the MTHFR C677T polymorphisms in lung cancer patients.

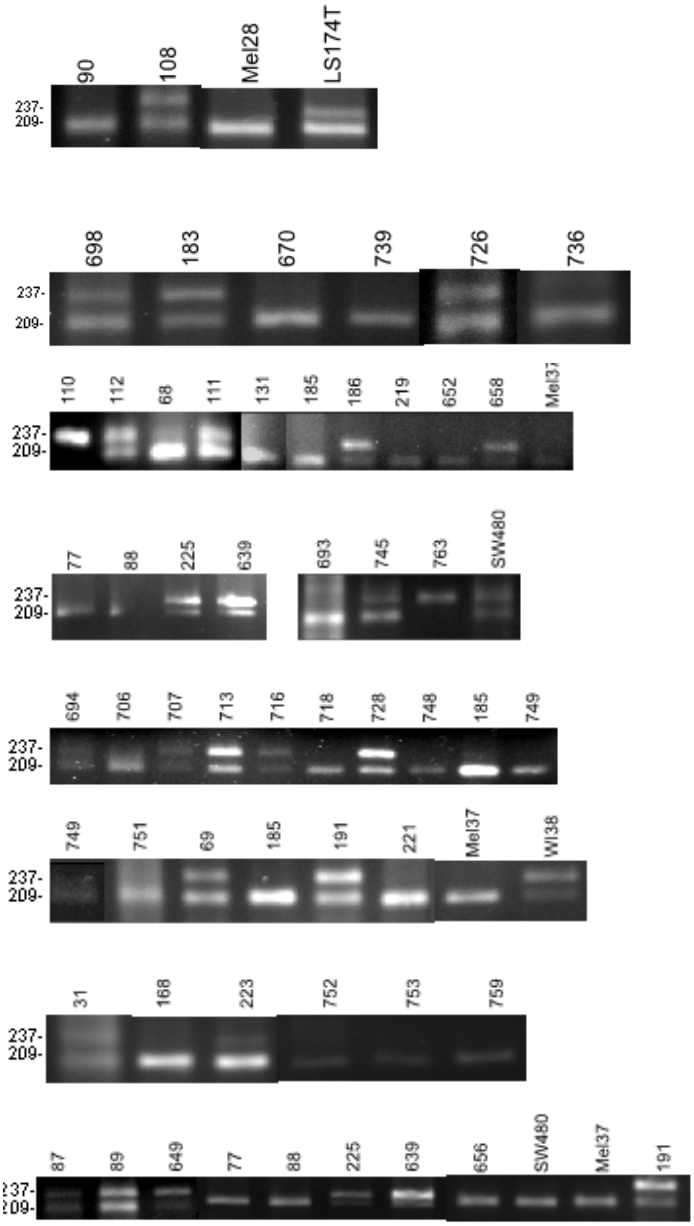


Figure 11 (Supplementary Figure 2). RFLP results of the MTHFR A1298C polymorphisms in lung cancer patients.

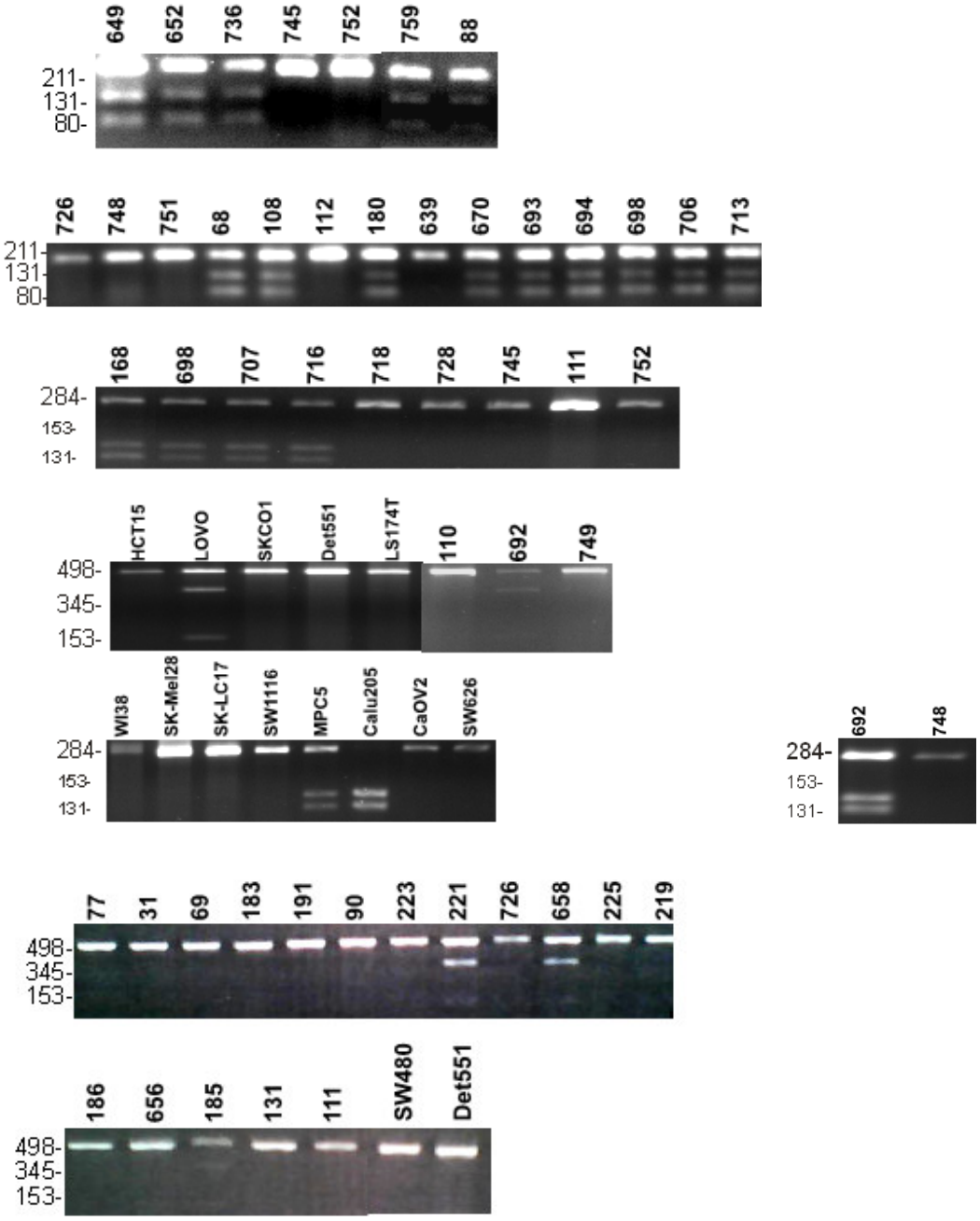


Figure 12 (Supplementary Figure 3). RFLP results of the MTR A2756G polymorphisms in lung cancer patients

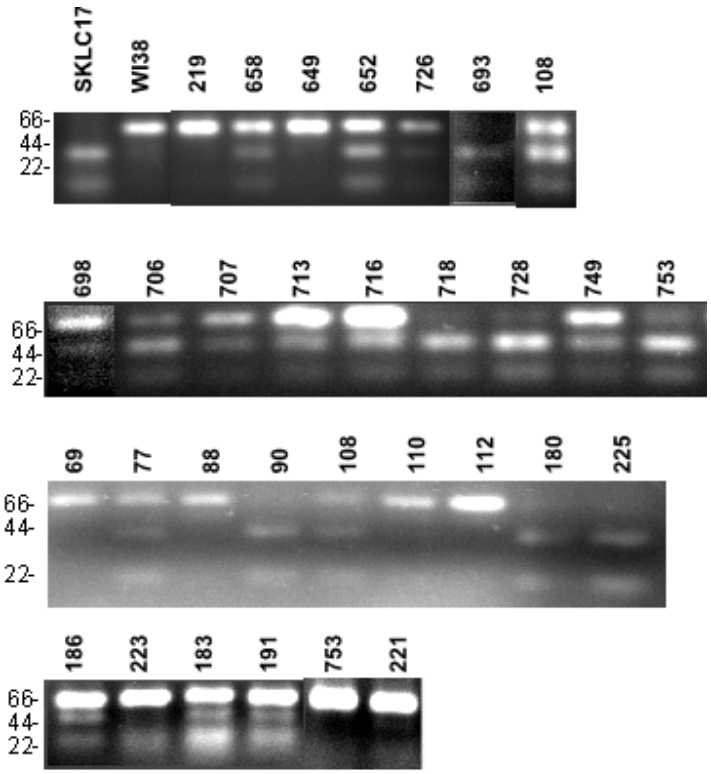
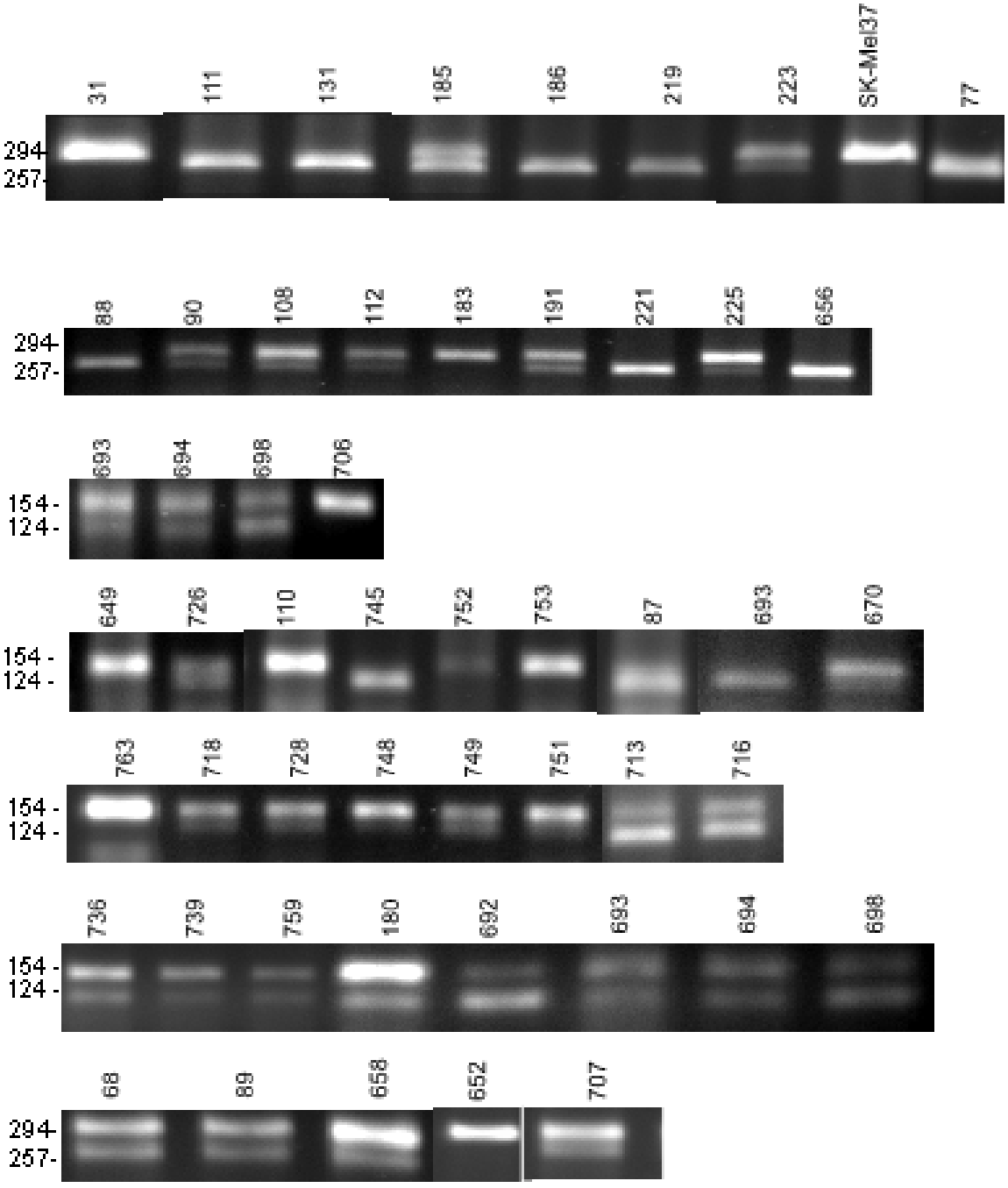


Figure 13 (Supplementary Figure 4). RFLP results of the MTRR A66G polymorphisms in lung cancer patients.



Supplementary Figure 5. RFLP results of the RFC G80A polymorphisms in lung cancer patients.