2006

Single Nucleotide Polymorphisms in the Folypoly-gamma-glutamate synthetase Gene

Nadine Thompson
Virginia Commonwealth University

Single Nucleotide Polymorphisms in the *Folylpoly-gamma-glutamate synthetase* Gene

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

by

Nadine Thompson,
Bachelor of Science
College of William and Mary, May 2003

Director: Richard G. Moran, Ph.D.
Professor, Pharmacology and Toxicology

Virginia Commonwealth University
Richmond, Virginia
August, 2006

Acknowledgment

I would like to thank my husband, Kenneth H. Carlton and family for giving me the love and support I needed to complete this project. I would also like to thank Dr. Richard Moran and the members of the Moran/Taylor laboratory for all their help and guidance with this project. I would also like to thank Dr. George Ford for being an excellent advisor during this period. And last but not least, I would like to thank Carleton Garrett, M.D, Catherine Dumur, Ph.D and Amy Ladd Ph.D for providing the patient samples needed to successfully complete this project.

Table of Contents

List of Tables

List of Figures

Abbreviations

(FPGS), Folypoly-γ-glutamate synthetase

(SNPs), single nucleotide polymorphisms

(PCR), Polymerase Chain Reaction

(A), adenine

(T), thymine

(G), guanine

(C), cytosine

(*MTHFR,)*methylenetetrahydrofolate reductase

(*mtr),* 5-methyltetrahydrofolate-homocysteine methyltransferase

(*mthfr),* methylenetetrahydrofolate reductase

(SAM), S-adenosylmethionine

(hFR), human folate receptor

(THF), tetrahydrofolate

(dUMP), 2'-deoxyuridine-5-monophosphate

(TMP), 2'-deoxythymidine-5'-monophosphate

(TS), thymidylate synthetase

(UTR), untranslated region

(TE), Tris-EDTA buffer

(MNC), Mononuclear cells

(dNTPs),deoxyribonucleotide triphosphates

(EDTA), ethylenediaminetetraacetic acid

Abstract

SINGLE NUCLEOTIDE POLYMORPHISMS IN THE *FOLYLPOLY- GAMMA-GLUTAMATE SYNTHETASE* GENE

By Nadine Thompson, B.S.

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

Virginia Commonwealth University, 2006.

Major Director: Richard G. Moran, Ph.D.

Professor, Pharmacology and Toxicology

Folic acid is an essential vitamin utilized in the one-carbon metabolism pathway for the synthesis of purine and thymidine nucleotides, which are necessary for cell growth and proliferation. As a result, the enzymes that participate in the metabolism of folic acid have been good targets for cancer chemotherapy. Folylpoly-γ-glutamate synthetase (FPGS) is an enzyme in the folate metabolism pathway that catalyzes the addition of glutamic acid to the naturally occurring folates, thereby allowing the retention of folate cofactors in cells. Similarly, in the case of cancer chemotherapy, antifolates, such as Lometrexol and Tomudex are retained in cells through the activity of FPGS. Consequently, any single nucleotide polymorphisms (SNPs) that exist in the *fpgs* gene may decrease or increase the cytotoxicity of antifolates and, ultimately, the clinical response rate to antifolate therapy. The goal of this project is to define the position and frequency of single nucleotide polymorphisms (SNPs) in the mRNA made from the *fpgs* gene from peripheral blood of one hundred normal individuals. Six Polymerase Chain Reaction (PCR) primers were designed to amplify the gene as three overlapping pieces and four primers were designed for sequencing of the three PCR products. In this study, we found polymorphic sites at nucleotides 64, 123, 253, 423, 1334 and 1781. The

majority of the samples (49/88) expressed mRNA with point mutations on at least one allele at base 64, while 8 samples had a SNP at base 123. At nucleotide 123, 6 samples expressed the heterozygote G/A genotype, and one sample expressed the homozygote A/A allele at this site. At nucleotide 423, two samples expressed a G allele and also the common C allele. While the SNPs at nucleotide 64, 123, and 423 caused a silent or conservative mutation in the gene, in sample 82, a mutation C253T produced an amino acid change from an arginine to tryptophan, which may cause a functional change in the fpgs protein, due to the significant change in size and charge of the wild type amino acid. Similarly, sample 26 expessed a homozygote T/T allele at nucleotide 1334 instead of the common C/C allele expressed in the remaining samples. This point mutation caused a valine to alanine amino acid change. We also detected a SNP that is expressed after the stop codon in sample 40.

Chapter 1

INTRODUCTION:

Single Nucleotide Polymorphisms

Single Nucleotide Polymorphisms (SNPs) are point mutations that exist within the genetic code of every individual **(1).** The four nucleotides, adenine (A), thymine (T), guanine (G), and cytosine (C), serve as the building blocks for the DNA sequence coding for the RNA message during transcription. A Single Nucleotide Polymorphism occurs when one nucleotide, such as adenine for example, is replaced by another nucleotide such as thymine, guanine or cytosine. This change in the DNA sequence may appear on one or both alleles of a gene, producing a homozygous (altered sequence of both alleles) or heterozygous (altered sequence of one allele) mutation. As one can imagine, a significant change in the sequence of a gene that alters the protein sequence encoded by the RNA, ultimately affects protein size and/or function, which can have dramatic effect on cellular activity and function. Since all amino acids are coded by more than one codon, it is possible to have a SNP that creates a new codon that translates to the same amino acid. For instance, the nucleotide triplets AAA and AAG both result in insertion of a glutamic acid to a growing peptide codon. This 'silent' substitution normally occurs with a single base change in the third position of the codon. In some instances, however, a base change in the third position may cause a nonconservative change, where one amino acid replaces another amino acid with dissimilar chemical characteristics. For instance, mutation of the triplet, TGT to TGG would change the encoded amino acid from cysteine to tryptophan. Similarly, base changes in the first and second position of the codon can

result in several types of different mutations. A nonsense mutation occurs when a SNP causes an amino acid change that results in a stop codon. A missense mutation, on the other hand, occurs when a SNP results in the replacement of an amino acid by one of similar chemical nature, which does not alter protein function. For example, mutation of the nucleotide triplet GAC to GAA would cause an amino change from aspartic acid to glutamic acid. This phenomenon, of amino acid substitution and protein function modification, is particularly important in disease diagnosis and drug therapy, where a point mutation may signify a disease state or may alter protein function in a way that enables cells to become resistant or insensitive to certain drugs. It is difficult to determine the number of single nucleotide polymorphisms that exist in each gene within the human population, information that would require a complete sequence analysis of the entire human genome across the entire human population. In fact, the complete sequence of the entire genome of a single individual has only recently been completed **(2).** Surprisingly, the human genome consists only of about 30,000 genes spanning 23 pairs of chromosomes, far below the original estimates of 150,000-250,000 **(3).** This number is surprisingly low when compared to the number of genes found in species that are much less complex than humans. The genome of *C. elegans*, for example, consists of 19,099 genes while *D. melanogaster* and *S. cerevisiae* have only 13,061 and 6,034 genes, respectively **(3).** The estimated number of genes for *M. musculus* is roughly the same as in humans **(2).** It appears that the mechanism by which the human body maintains it complexity with a limited number of genes involves gene splicing, gene rearrangement, and/or expression of multiple transcriptional start sites that allows one gene to code for multiple proteins or multiple forms of one protein performing entirely different functions.

Multiple transcriptional start sites in the folypoly-γ-glutamate synthetase gene, for example, allows the transcription of two forms of FPGS mRNA corresponding to two isoforms of the FPGS enzyme **(4)** (**See Figure 1**).

SNPs in the genes of folate metabolism

The folylpolyglutamate synthetase enzyme (FPGS) plays a central role in intracellular accumulation of folate and antifolate in the cytosolic and mitochondrial compartments of the cell **(5).** Two translational start sites, one for the mitochondrial FPGS and the other for the cytosolic FPGS, exist within the *fpgs* coding region, allowing the gene to code for two proteins that function in two different compartments (cytosol and mitochondria) of the cell. The cytosolic and mitochondrial proteins are extremely important for cell growth and survival, and so as one can imagine, if a single nucleotide polymorphism exists in an essential part of the coding region of the *fpgs* gene preventing correct translation of the mitochondrial or cytosolic protein, then cell death could occur. Recent studies have also shown a variation in the frequency of single nucleotide polymorphisms across different populations of people **(4)**. For example, the methylenetetrahydrofolate reductase (*MTHFR)* gene was shown to have several polymorphic sites that appear more frequently in certain populations of people. A C→T polymorphism at nucleotide 677 of the *mthfr* gene caused a change of the alanine to a valine amino acid, while an A→C polymorphism at nucleotide 1298 caused a change in sequence from alanine to glutamate at this site **(7)**. While only 1 percent of the Black population in the United States, sub-Saharan Africa, and South America, for example, expressed the *TT* genotype of *mthfr,* 20 percent of US Hispanics, Colombians, and Amerindians in Brazil expressed this genotype

(8). Polymorphisms at nucleotide 1298 in *mthfr* are fairly common. White populations in Europe, North America, and Australia showed 8-20 percent expression of the *TT* genotype for the *mthfr* gene, and 12 percent of the Japanese were homozygous for this allele **(8)**. Another important protein of folate metabolism, 5-methyltetrahydrofolate-homocysteine methyltransferase (*mtr),* expressed an A→G polymorphism at nucleotide 2756 in codon 919, which caused the replacement of aspartic acid with glycine, changing the function of this enzyme, and ultimately causing a decrease in plasma homocysteine level **(9)**. This residue occurs in the substrate-binding region of the protein. A small percentage (2–3) of Japanese, Chinese, and Korean populations expressed the *GG* genotype at the *mtr* gene, while, in the African-American population, six percent had the *GG* genotype. Of Canadian Caucasian children and their mothers and in the Caucasian population in Hawaii, 10–11 percent were *GG* homozygotes at nucleotide 2756 in the *mtr* locus **(10)**. Findings from studies such as these that analyze genetic variation across different populations of people are necessary to understand why a certain population is more susceptible to certain diseases and how these patients will react to varying degrees of drug therapies. In the case of the methylenetetrahydrofolate reductase (*mthfr)* gene, for example, it was found, that heterozygotes expressing the C→T variant at nucleotide 677 on one chromosome have 10 percent lower red cell folate levels compared with *CC* homozygotes, and *TT* homozygotes produced 18 percent lower red cell folate levels **(11)**. Individuals expressing the *TT* genotype also had lowered vitamin B_{12} and plasma folate levels. In the case of the MTR protein, where the A/G polymorphism at position 2756 replaces aspartic acid with glycine in the protein substrate-binding region, studies show a lowered plasma homocysteine level in individuals expressing the *GG* allele than those

expressing the more frequent *A* allele. Another study found that individuals expressing the GG genotype in the *mtr* gene appear to produce significantly higher plasma folate levels, than those expressing the AA genotype **(12)**. Variation in expression levels of different proteins, especially those targeted for drug therapy, can affect the efficiency of drug interaction with their target proteins.

Enzymes of the Folate Metabolism

There are currently a number of different chemotherapy drugs on the market that target several proteins in the folate metabolism pathway. This pathway consists of a series of one-carbon transfer reactions that result in the production of purine and thymidine nucleotides necessary for DNA synthesis, amino acid metabolism, and the formation of the methyl-donating agent, S-adenosylmethionine (SAM) (**See Figure 2**) **(8)**. Once folates are absorbed from the diet in the gastrointestinal tract, they pass through the liver and blood stream. Folates then cross the cellular membrane through an active carrier-mediated transport system, which binds reduced folate compounds. There are two forms of carrier-mediated transport system utilized by the cell for intracellular transport of reduced folates. The first is the low affinity reduced folate carrier, which has affinity constants in the micromolar range for its endogenous substrates 5-methyltetrahydrofolate, 5-formyltetrahydrofolate, and the antifolate, methotrexate **(13).** The second transporter system employs a folate-binding protein, called the human folate receptor (hFR), which is associated with the outer membrane of the cell **(14).** This system has a much higher affinity for reduced folates, with affinity constants in the nanomolar range. Once folic acid enters the cell, it is reduced to dihydrofolate, which is then reduced to

tetrahydrofolate (THF) through the catalytic activity of the enzyme, dihydrofolate reductase (**See Figure 2**). Another reduced form of folates, 10-formyltetrahydrofolate, serves as the donor of the single carbon group in the de novo biosynthesis of purine nucleotides. Intracellular retention of folates in the cells occurs through the catalytic activity of the enzyme, folylpolyglutamate synthetase (FPGS). Once folates are polyglutamated and reduced to THF, they are converted to 5,10 methylenetetrahydrofolate. The 5,10 methyleneTHF then undergoes two separate enzymatic reactions. In the first reaction, 5,10 methyleneTHF is reduced to 5-methyl-THF by methylene tetrahydrofolate reductase (MTHFR), subsequently allowing the transfer of the N-5 methyl group from 5-methyl-THF to homocysteine, which results in the synthesis of methionine. Methionine synthesis requires the coenzyme vitamin B-12, the acceptor of the methyl group from 5-methyl-THF, and the enzyme methionine synthase. Once methionine is synthesized, it converts to S-adenosylmethionine (SAM), which serves as a source of a methyl group during DNA methylation and other methyltransferases reactions. The second enzymatic reaction involving the 5,10 methyleneTHF cofactor results in the synthesis of thymidylate by the enzyme thymidylate synthase. In this reaction, 5,10 methyleneTHF serves as the donor of the CH_2 unit, which is passed to the 5^{th} position of deoxyuridylate and reduced in the thymidylate synthase reaction. Thus, thymidylate synthase functions as the enzyme that catalyzes the transformation reaction of 2'-deoxyuridine-5-monophosphate (dUMP) to 2'-deoxythymidine-5'-monophosphate (TMP) which is then phosphorylated by cellular kinases and integrated into DNA. Thymidylate synthetase also causes the oxidation of 5,10 methyleneTHF to dihydrofolate, which must then be converted back to

tetrahydrofolate by DHFR. These one-carbon transfer reactions of the folate metabolism pathway are all required for cell growth, and, hence, this pathway is a good target for antiproliferative drug therapy.

Inhibitors of Folate Metabolism

Antifolates are folate analogs that inhibit various points of the folate metabolism pathway. Some of the classical antifolates, which are inhibitors of dihydrofolate reductase and thymidylate synthetase, are methotrexate, pemetrexed, and tomudex (**See Figure 3**) **(15)**. Methotrexate, a competitive reversible inhibitor of dihydrofolate reductase, differs from folic acid in that an amino group is substituted for the hydroxyl group at the 4^{th} position of the pteridine ring. This structural change increases the binding affinity of the drug to the folate-binding pocket of DHFR, making it more difficult to displace the drug from the enzyme **(16)**. Binding methotrexate to dihydrofolate reductase causes a decrease in the intracellular reduced folate pools and an increase in dihydrofolate. The retention of methotrexate inside the cell is due to activity of folylpolyglutamate synthetase, which attaches up to six glutamate molecules to the pteridine ring of the antifolate, causing a net trapping effect of the drug inside the cell. Tomudex, another antifolate, is a selective inhibitor of thymidylate synthetase (TS) **(17)**. Tomudex, which interferes with the enzymatic activity of thymidylate synthetase, causing a decrease in thymidylate production and DNA synthesis, and is currently being used in therapy of colorectal cancer **(17)**. Pemetrexed is also a potent inhibitor of thymidylate synthetase and DHFR that accumulates inside the tumor and is retained intracellularly as a result of polyglutamation by folylpolyglutamate synthetase **(18)**.

Folate and Antifolate Polyglutamation

Reduced folates bind more tightly to their corresponding enzymes when in a polyglutamated state than in a monoglutamated state (**15**). Similarly, the binding affinity of antifolates to their target enzymes increases, as they are polyglutamated by FPGS. The structure of the FPGS enzyme with its folate and antifolate binding sites is illustrated in Figure 4. Consequently, as these drugs are retained inside the cells, their cytoxicity is prolonged and their ability to inhibit cell growth and proliferation improves. Furthermore, previous studies have shown that tumor cells that have partially inactivating mutations in FPGS are resistant to most antifolates **(19)**. For these reasons, FPGS plays central role in trapping of intracellular antifolates, and is viewed as critical to the efficacy of antifolate chemotherapy.

Folylpolyglutamate Synthase

Human folylpolyglutamate synthetase gene is located on human chromosome 9q34 **(20)**. The mRNA consists of 2475 base pairs spliced from 15 exons and the entire gene spans 11.2 kb of genomic DNA **(21)**. FPGS cDNA isolated from T-acute lymphocytic leukemia cells (CEM) encodes a protein with 545 amino acids that has a MW of 60 Kd. The transcription of the human *fpgs* gene results in the production of two separate mRNA variants that code for a cytosolic and a mitochondrial isoform of the FPGS protein **(4)**. The cytosolic isoform has a shorter N-terminus than the mitochondrial isoform, and lacks the upstream mitochondrial leader sequence found in the mitochondrial isoform. The transcriptional leader sequence of the mitochondrial construct allows the transport of

FPGS protein to the mitochondrial matrix compartment of the cell, where FPGS is required for mitochondrial folate accumulation.

Accumulation of folates inside the cytosolic and mitochondrial compartments of the cell allows for the survival and maturation of proliferating cells. Any significant mutation in the *fpgs* gene might prevent intracellular accumulation of folates, which affects cell growth and cell survival. In the case of antifolate drugs, a mutation in the *fpgs* gene will negatively impact protein/drug interactions, thereby preventing intracellular accumulation of the antifolates **(19)**.

Previous reports of SNPs in the human *fpgs* gene

There are currently a number of single nucleotide polymorphisms in the *fpgs* gene that have been reported in the NCBI SNP database. The NCBI SNP database, which was established by the National Human Genome Research Institute and the National Center for Biotechnology Information, is a readily accessible repository of single nucleotide polymorphisms from genes that have been sequenced worldwide. The genetic polymorphisms (including nucleotide substitutions, deletion, and insertions) that have been submitted to the SNP database come from random sequencing performed manually or robotically in different laboratories. Because many of these randomly submitted SNPs have not yet been confirmed by further studies, it is difficult to distinguish true SNPs from sequencing errors. Table 2 is a compilation of all 37 unique SNPs that have been submitted to the NCBI database since July 2006. Of the 38 SNPs submitted, 27 of them were detected in the introns and 8 were detected in the exons of the gene. While two of the SNPs submitted were never detected on the *fpgs* locus by Blast search, four of the

intronic SNPs were from unknown regions of the intron. Of the 8 exonic SNPs reported, four were detected in exon 15, one in exon 16, one in exon 11, and two were detected in exon 1. Three of the SNPs detected in exon 15 were from the 3'untranslated region (UTR) of the gene. Only two of these SNPs, located in the 3'untranslated region of the *fpgs* locus, were later confirmed in other studies. SNP entry # rs4451422 was later confirmed in a study performed on 354 samples. The heterozygote A/C allele and the homozygote A/A and C/C alleles were detected at a frequency ratio of .065 /.234/ .333 respectively. The remaining .367 allelic frequencies were unspecified in the Database (see NCBI Database). SNP entry # rs10106 was also later confirmed in a study conducted using 421 samples. The heterozygote A/G allele and the homozygote A/A and G/G alleles were detected at a frequency ratio of .48 / .22 / .30 respectively. The remaining SNP, founded in exon 15 (entry # rs12686275), which caused an amino acid changed from valine to aspartic acid was never confirmed in a study done using 209 samples. For this reason, SNP entry # rs12686275 appears to be the result of an error in the sequence. Similarly, SNP entry number rs1042304, of uknown intron, was checked in a study done using 204 samples and was never detected. This SNP also seems to be a sequencing error. Of the three SNPs detected in exon 1, SNP entry # rs17855899 and # rs10760502, found in codon 22 and which caused an isolucine to valine amino acid change, was submitted two times to the NCBI database. This SNP was later detected in 55% of the samples used in this study. SNP entry #rs2230270 of exon 11 that resulted in a threonine to alanine amino acid conversion has also been confirmed by a study done using 23 samples. The alleles detected from this study were homozygote A/A and heterozygote A/G at a frequency of 95.7% and 4.3% respectively. Entry # rs11554717 was another SNP that

was detected in exon 1 of codon 21. This SNP caused an A to C polymorphism, which did not change the glycine amino acid of the protein. This SNP has not yet been confirmed. Similarly, SNP # rs17855900, which caused a C to T polymorphism that changed the valine to an alanine and has not yet been confirmed. Both these unconfirmed SNPs appeared as a result of random sequencing. The intronic SNPs that have been confirmed are entry # rs7856096 and rs955171. SNP rs7856096 was confirmed in a study performed on 209 individuals. In this study the alleles detected were A/A, A/G, and GG at a frequency ratio of .885 / .091 / .024 respectively. Similarly, SNP rs955171 was confirmed in four studies. Three of these studies had 100% C/Cs in all 30 samples while the other study had 100% G/G in all 47 samples. Reasons for the discrepancy in these two results are unknown.

In the current experiment, we have detected a number of single nucleotide polymorphisms in the *fpgs* gene that may cause significant functional differences in the FPGS protein. The first portion of these experiments involves the isolation of RNA using the Trizol method and sequencing of PCR products from three control cell lines with a known polymorphic site at nucleotide 64 in exon 1 of the gene. The next portion of the experiment involves isolation of RNA from 100 patient samples using the Quiagen method, which we found to be less time consuming, and direct sequencing of the PCR products to determine the presence of polymorphisms. For the 100 patient samples, RNA integrity and concentration was determined using the microchip on the Agilant Bioanalyzer. Sequencing was done at the DNA core lab at the Virginia Commonwealth University and at the SeqWright and Fisher Scientific Laboratory, and single nucleotide polymorphisms were detected using the CodonCode Aligner computer program. All

polymorphisms presented have been confirmed through repeated experiments across a number of different samples.

Fig 1. Transcriptional start site for human fpgs

The entire human *fpgs* mRNA construct consisting of 2487 nucleotides is displayed in Figure 1. Two translational methionine start sites (as shown in blue, along with the stop codon) were detected in this region of the human *fpgs* mRNA by PCR-derived RACE **(4).** The first methionine site signals the initiation of translation of the leader sequence of the mitochondrial FPGS protein, while the second methionine site signals the start of translation of the cyosolic FPGS enzyme. The 3' untranslated region of the mRNA construct consists of 50 nucleotides and the 5' untranslated region consists of 673 nucleotides. The mitochondrial *fpgs* mRNA consists of 1765 nucleotides and the cytosolic *fpgs* mRNA consists of 1643 nucleotides of the coding **(21).** These data were reproduced from GenBank entry NM 003056 (www.ncbi.gov).

1 gcggggcgtc tcccgcccgg gcctagagcg ctgccggggg cgccgggact atgtcgcggg
61 cgcggagcca cctgcgcgcc gctctattcc tggcagcggc gtctgcgcgc ggcataacga
121 cccaggtcgc ggcgcggcgg ggcttgagcg cgtggccggt gccgcaggag ccgagcatgg
181 agtaccagga tgccgtgcgc atgctcaata ccctgcagac caatgccggc tacctggagc
241 aggtgaagcg ccagcggggt gaccctcaga cacagttgga agccatggaa ctgtacctgg
301 cacggagtgg gctgcaggtg gaggacttgg accggctgaa catcatccac gtcactggga
361 cgaaggggaa gggctccacc tgtgccttca cggaatgtat cctccgaagc tatggcctga
421 agacgggatt ctttagctct ccccacctgg tgcaggttcg ggagcggatc cgcatcaatg
481 ggcagcccat cagtcctgag ctcttcacca agtacttctg gcgcctctac caccggctgg
541 aggagaccaa ggatggcagc tgtgtctcca tgccccccta cttccgcttc ctgacactca
601 tggccttcca cgtcttcctc caagagaagg tggacctggc agtggtggag gtgggcattg
661 gcggggctta tgactgcacc aacatcatca ggaagcctgt ggtgtgcgga gtctcctctc
721 ttggcatcga ccacaccagc ctcctggggg atacggtgga gaagatcgca tggcagaaag
781 ggggcatctt taagcaaggt gtccctgcct tcactgtgct ccaacctgaa ggtcccctgg
841 cagtgctgag ggaccgagcc cagcagatct catgtcctct atacctgtgt ccgatgctgg
901 aggccctcga ggaaggggggg ccgccgctga ccctgggcct ggaggggggag caccagcggt
961 ccaacgccgc cttggccttg cagctggccc actgctggct gcagcggcag gaccgccatg
1021 gtgctgggga gccaaaggca tccaggccag ggctcctgtg gcagctgccc ctggcacctg
1081 tgttccagcc cacatcccac atgcggctcg ggcttcggaa cacggagtgg ccgggccgga
1141 cgcaggtgct gcggcgcggg cccctcacct ggtacctgga cggtgcgcac accgccagca
1201 gcgcgcaggc ctgcgtgcgc tggttccgcc aggcgctgca gggccgcgag aggccgagcg
1261 gtggccccga ggttcgagtc ttgctcttca atgctaccgg ggaccgggac ccggcggccc
1321 tgctgaagct gctgcagccc tgccagtttg actatgccgt cttctgccct aacctgacag
1381 aggtgtcatc cacaggcaac gcagaccaac agaacttcac agtgacactg gaccaggtcc
1441 tgctccgctg cctggaacac cagcagcact ggaaccacct ggacgaagag caggccagcc
1501 cggacctctg gagtgccccc agcccagagc ccggtgggtc cgcatccctg cttctggcgc
1561 cccacccacc ccacacctgc agtgccagct ccctcgtctt cagctgcatt tcacatgcct
1621 tgcaatggat cagccaaggc cgagacccca tcttccagcc acctagtccc ccaaagggcc
1681 tcctcaccca ccctgtggct cacagtgggg ccagcatact ccgtgaggct gctgccatcc
1741 atgtgctagt cactggcagc ctgcacctgg tgggtggtgt cctgaagctg ctggagcccg
1801 cactgtccca gtagccaagg cccggggttg gaggtgggag cttcccacac ctgcctgcgt
1861 tctccccatg aacttacata ctaggtgcct tttgtttttg gctttcctgg ttctgtctag
1921 actggcctag gggccagggc tttgggatgg gaggccggga gaggatgtct tttttaaggc
1981 tctgtgcctt ggtctctcct tcctcttggc tgagatagca gaggggctcc ccgggtctct
2041 cactgttgca gtggcctggc cgttcagcct gtctccccca acaccccgcc tgcctcctgg
2101 ctcaggccca gcttattgtg tgcgctgcct ggccaggccc tgggtcttgc catgtgctgg
2161 gtggtagatt tcctcctccc agtgccttct gggaagggag agggcctctg cctgggacac
2221 tgcgggacag agggtggctg gagtgaatta aagcctttgt tttttaaaga aatggcaaag
2281 ccttcgactg accettgacc ccctgctccc tcagcagaga cggagggagg ggctgctggt
2341 ggtcagggac ctgcactgtg tagagggagc ctggctgtgt ggcctggaac aagtccctcc
2401 ctccctgtgc gcctcaggtg gcctgtctgt gagatgagaa gaagaccaga ctgaagcctg
2461 ttcaccatat gccaggcagt gctttct

Fig. 1

Fig 2. Folate Metabolism Pathway

Folic Acid is utilized in the folate metabolism pathway, which consists of a series of one-carbon transfer reactions that result in DNA synthesis and DNA methylation. Folic acid enters the folate metabolism pathway as its reduced form, dihydrofolate, which ultimately reduces to 5,10-methyltetrahydrofolate through the enzyme 5,10-methyltetrahydrofolate reductase. 5,10-methyltetrahydrofolate then enters the thymidylate synthase pathway where it plays a central role in the conversion of deoxyuridine monophosphate to deoxythymidine monophosphate in the synthesis of purine and thymidine nucleotides. Similarly, 5-methyltetrahydrofolate (a reduced form of tetrahydrofolate) plays a central role in the formation of the methyl donating agent, S-adenosylmethionine (SAM) in the DNA methylation pathway. This diagram was modified from Sharp, L et al **(6)**.

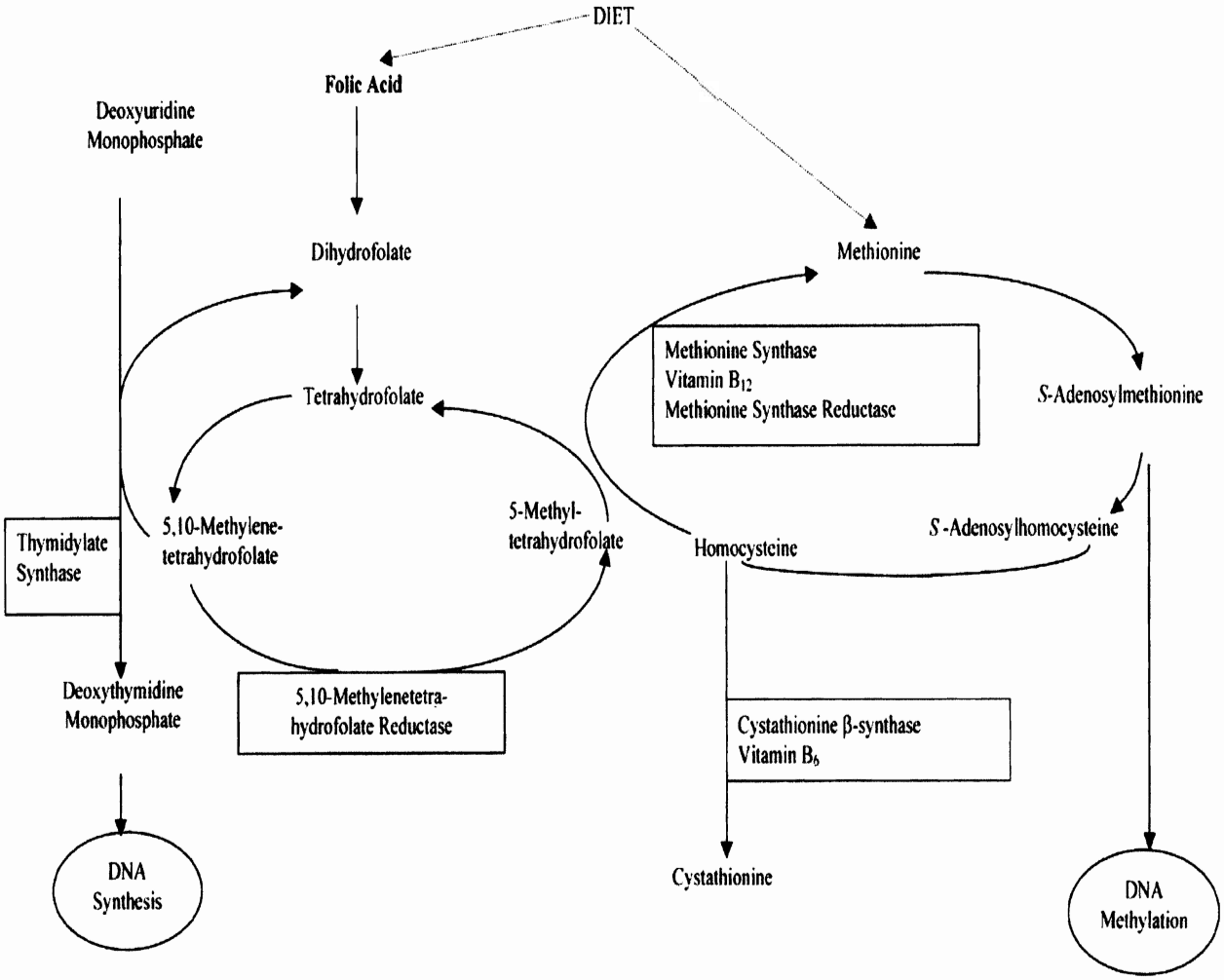
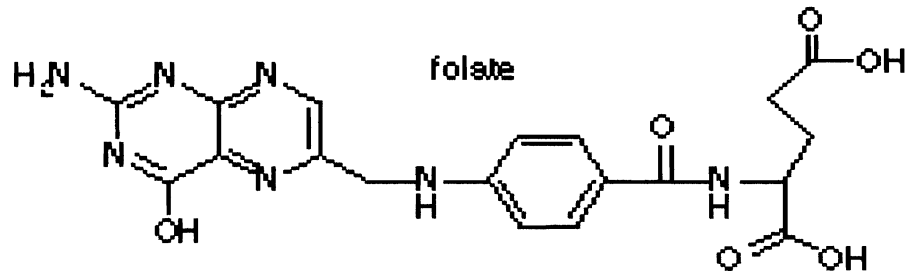
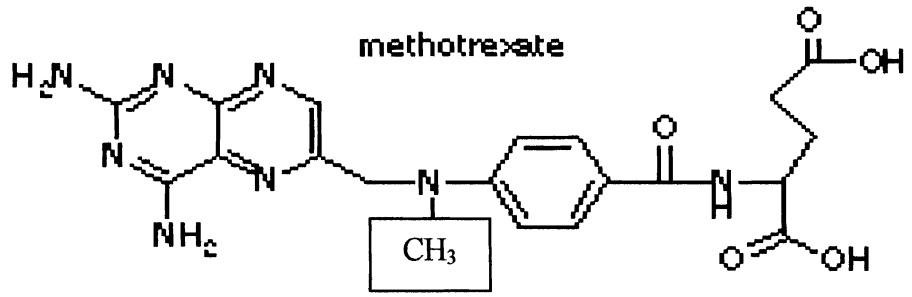

DIET
Folic Acid
Deoxyuridine Monophosphate
Dihydrofolate
Methionine
Methionine Synthase
Vitamin B_{12}
Methionine Synthase Reductase
Tetrahydrofolate
S-Adenosylmethionine
Thymidylate Synthase
5,10-Methylene-tetrahydrofolate
5-Methyl-tetrahydrofolate
Homocysteine
S-Adenosylhomocysteine
Deoxythymidine Monophosphate
5,10-Methylenetetra-hydrofolate Reductase
Cystathionine β-synthase
Vitamin B_6
DNA Synthesis
Cystathionine
DNA Methylation

Fig. 2

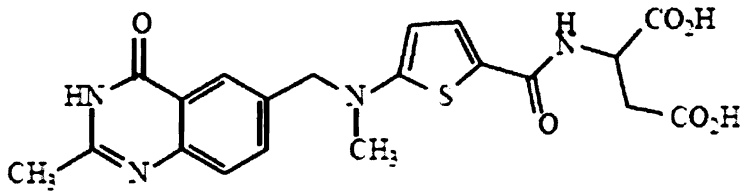
Fig 3. Classical Antifolates

Three classical antifolates used for cancer chemotherapy are methotrexate, pemetrexed, and tomudex. Due to their structural similarities to folates, these folate analogs serve as effective inhibitors of folate metabolism. Methotrexate is a competitive reversible inhibitor of dihydrofolate reductase, whose increased binding affinity to dihydrofolate reductase causes a decrease in the intracellular reduced folate pools. Similarly, tomudex when bound to thymidylate synthase, inhibits its enzymatic activity, thereby decreasing thymidylate production and DNA synthesis. Pemetrexed, also an antifolate that is structurally similary to folic acid, inhibits both thymidylate synthase and dihydrofolate reductase, preventing DNA and RNA synthesis **(15)**.





Tomudex



Pemetrexed

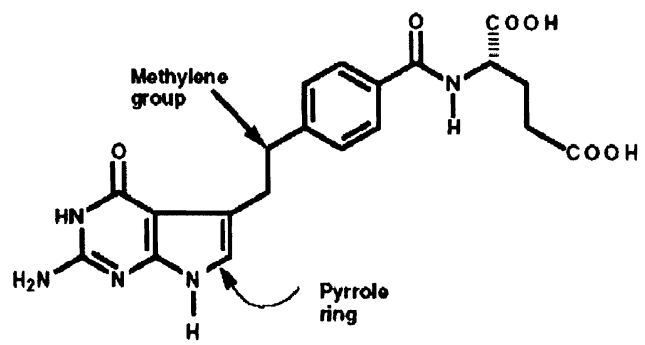


Fig. 3

Fig 4. Ribbon structure of the *Lactobacillus casei* FPGS enzyme

In order to fully understand the clinical significance of SNPs in the *fpgs* gene and its effect on antifolate therapy, extensive research on the structure of the FPGS enzyme must be conducted. The position at which amino acid changes occur in the polypeptide chain determines the degree by which they would affect the enzymatic activity of the protein. An amino change in the active site, for example, would greatly affect the catalytic activity of the enzyme versus an amino change on the outer surface of the enzyme. Figure 4 illustrates the ribbon structure of the *Lactobacillus casei* FPGS enzyme. The blue area of the structure represents the N-terminal of the FPGS protein, which is composed of a ras-like domain and a three-stranded β-sheet (shown in green) **(22)**. The C-terminal of the enzyme, which appears to contain the folate-binding cavity, is illustrated by red portion of the structure. The ATP-binding P-loop motif of the enzyme is highlighted in magenta, and next to the ATP-binding P-loop motif lies the Ω loop, which serves as the active site of the enzyme, and is highlighted in cyan. Thin black lines represent the joining of the breaks in the polypeptide chain, and the alph-helix A10 of the enzyme is highlighted in pink **(22)**.

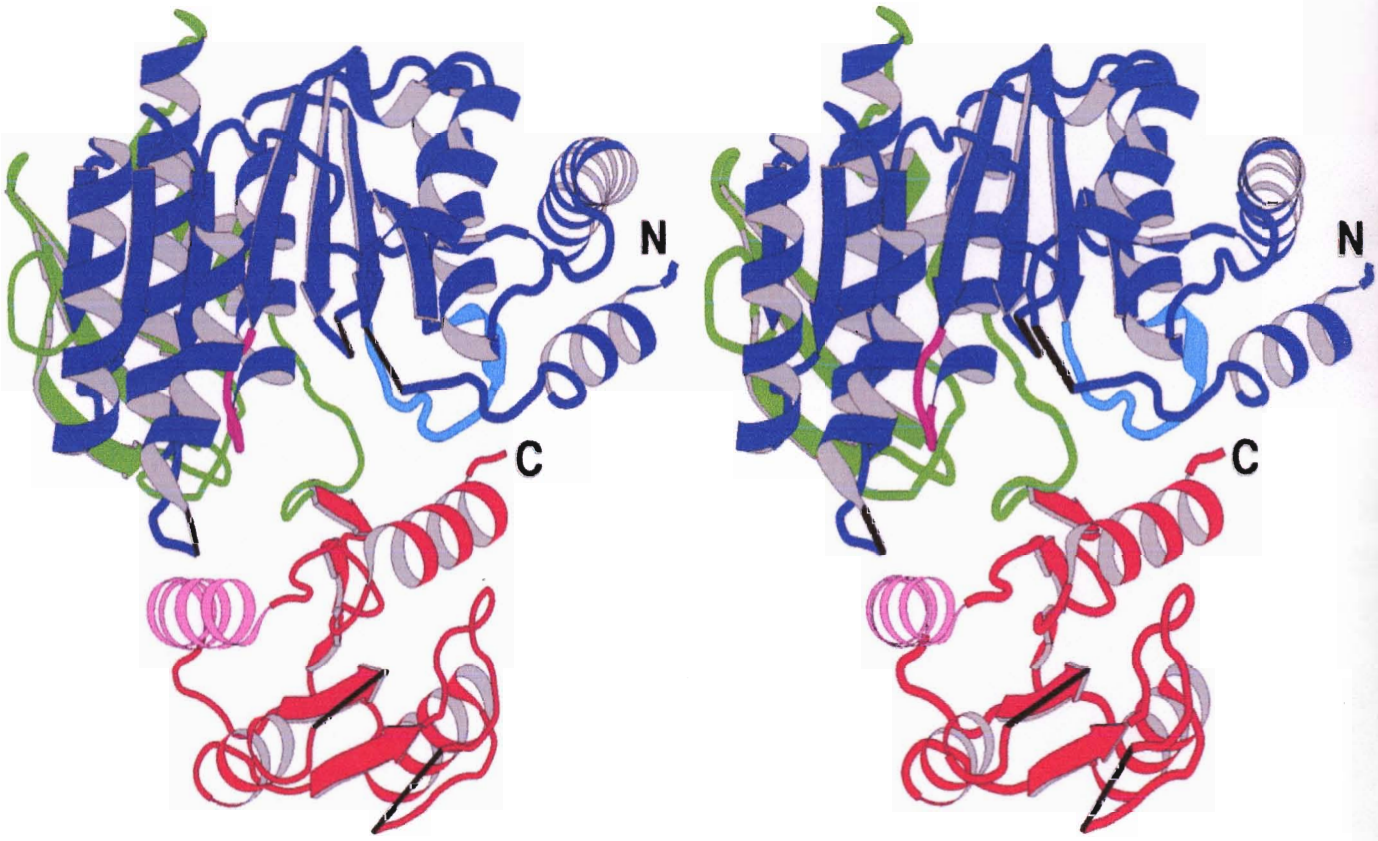


Fig. 4

Chapter 2

MATERIALS AND METHODS

Cell Line and Culture

Three human tumor cell lines, CEM (lymphoblastic leukemia cells), MCF-7 (breast cancer cells), and WIDR (colon carcinoma cells) were grown in RPMI-1640 medium supplemented with 10% fetal bovine serum and maintained at 37° in 5% CO_2. Cell counts were determined using tryptan blue and a hemocytometer.

RNA Isolation: Control Samples

RNA from three control cell lines, CEM, MCF-7, and WIDR, was extracted using Trizol reagents from Invitrogen Life Technologies. In a 15 ml tube, Trizol was added to cell pellets at a ratio of 1ml of Trizol / 1.4 x 10^7 cells and the cell suspension was vortexed for 1 minute to homogenize the cells and extract the nucleic acids. The mixture was then incubated for 3 minutes at room temperature and 0.2 ml of chloroform/1ml of Trizol was added. The mixture was vortexed for 15 seconds, and incubated for 15 minutes at room temperature. The samples were then centrifuged at 12,000 g for 15 minutes and the aqueous phase was transferred to a new Eppendorf tube. Isopropanol was added and the tubes were centrifuged at full speed in a Beckman microfuge for 10 minutes at 4^0C. Each cell pellet was then washed with 500 µl of 70% ethanol and the suspension was centrifuged at full speed for 2 minutes in a cold room. The pellet was air dried for 10 minutes and then dissolved in 50 µl of diethyl pyrocarbonate (DEPC) water. The RNA concentration and RNA purity were estimated by spectrophotometry (using a Spectronic Genesys 5 spectrophotometer), by adding 5 µl of sample and + 95 µl Tris-EDTA buffer (TE) and determining the absorbancy at 260 nm and 280 nm. RNA concentration was

determined using the Beer-Lambert law which state that A= ε Cl (**A**=Absorbancy at a particular wavelenth, ε = Extinction coefficient, **C** =Concentration and **l** = path length of spectrophotometer cuvette in centimeters). An A_{260} reading of 1.0 is equivalent to 40 µg/µl of RNA. An aliquot (1 µl of the 50 µl solution) of each RNA sample was then run on a 1% denaturing agarose gel to assess RNA integrity.

RNA Isolation: Patient Samples

Samples of peripheral blood were obtained from Carleton Garrett, M.D, Catherine Dumur, PhD and Amy Ladd, PhD, of the Molecular Diagnostics Laboratory of Virginia Commonwealth University. All samples were drawn with informed consent and were supplied for this study anonymously by Carleton Garrett, M.D. Peripheral blood was collected from 100 normal patients and mononuclear cells (MNC) were isolated using the Ficoll-Paque Plus technique. RNA was extracted from the cell pellets using the Qiagen RNAse easy kit and following the protocol suggested by the manufacturer. Cells were lysed using 350 µl of Buffer RLT, a buffer containing guanidinium isothiocyanate and ß-mercaptoethanol, and the cell suspension was vortexed until no cell clumps remained in the sample. The lysate was then washed using 350 µl of 70% ethanol to ensure sufficient RNA binding to an RNeasy silica-gel membrane of the mini column (used as a device to remove RNA contaminants). Following washing of the lysate with ethanol, 700 µl of sample was then added to the RNeasy mini column and the mini column was centrifuged for 15 seconds at 12,000 rpm. To remove contaminants from the RNA, additional washing of the RNA on the mini column was done using 700 µl of RW1 and 500 µl of Buffer RPE and the mini column was centrifuged for 15 seconds at 12,000 rpm after each addition. The columns were then removed and placed in a collection tube and the mini

column was then centrifuged for 2 min at maximum speed to dry the membrane. For complete elution of the RNA from the columns, 25 µl RNase-free water was applied directly to the filter of the column, let stand for 5 minutes at room temperature and centrifugation at 12,000 rpm for 1 minute. An aliquot (1µl) of Rnase inhibitor was added to each sample to prevent RNA degradation. The RNA concentration and RNA purity was estimated by spectrophotometry by adding 5 ul of sample and + 95 µl TE and determining absorbancy at 260 nm and 280 nm.

Quantification of RNA and Assaying RNA Integrity

To assess the integrity of total RNA, 1ul from each sample was loaded and ran on a chip of the Agilent 2100 Bioanalyzer. The Agilent 2100 Bioanalyzer produces an electropherogram that displays two peaks corresponding to the 18S and 28S ribosomal RNAs that one usually observes on denaturing agarose gel **(See Figure 6)**. We were able to determine the integrity of the RNA based on the height and quality of peaks representing the 18S and 28S ribosomal RNA displayed by the Agilent 2100 Bioanalyzer. Two distinct peaks appearing on the electropherogram represent pure RNA while multiple small peaks represent RNA degredation.

First-strand cDNA Synthesis

First-strand cDNA synthesis was done using the Invitrogen Superscript II kit. Two reaction mixtures were prepared in a 0.5 ml tube. The first reaction mixture (RNA/Primer mix) consists of total RNA template (3 ug), random hexamers (.2 ug), 10mM deoxyribonucleotide triphosphates (dNTPs), plus an added volume of nuclease-free water to achieve total desired RNA concentration. This mixture was then incubated at 65°C for

5 minutes and chilled on ice for at least 1minute. The second mixture (cDNA synthesis mix) consisted of 2 μl of 10x RT Buffer (consisting of 0.5 M Tris-HCl , 0.75 M KCl, 0.03 M $MgCl_2$), 4 μl of 25 mM $MgCl_2$, 2 μl of .1 M DTT, and 1μl RNaseOUT (4U / μl). An aloquot of the cDNA synthesis mixture (9 μl) was then added to each tube of RNA/Primer mix and the tubes were gently mixed, briefly centrifuged, and incubated at 25^0C for 2 minutes. A 1μl aliquot of Superscript II (200 units) was then added to each tube, and the tubes incubation at 25^0C for 8 minutes. RNA mixtures are then incubated at 50^0C for 50 minutes to activate the reverse transcriptase enzyme and begin reverse transcribing RNA to cDNA, then the reactions were terminated at 70^0C for 5 minutes to inactivate the enzyme. The tubes were then chilled on ice. Reactions were then collected by brief centrifugation, followed by the addition of 1 μl of *E. coli* Rnase H to each tube and the tubes were incubated at 37^0C for 20 minutes. The cDNA synthesis reaction mixes were then stored at -20^0C until use in the Polymerase Chain Reaction.

Polymerase Chain Reaction (Control samples)

DNA from the three control samples, CEM, MCF-7, and WIDR, was amplified using the Polymerase Chain Reaction as two separate fragments using four primers. Due to the high GC rich region of the 5' UTR region of the *fpgs* gene (*cctagagcgc tgccgggggc gccgggact),* which complicates the specificity of primer binding, amplification of the gene as a single 1.8 kb fragment was not feasible. As an alternative, the entire coding region of the fpgs gene was amplified as two overlapping PCR fragments (**See Figure 6**). Two primers were designed for amplification of the first fragment (.8 kb fragment) and another set of primers were designed for the second fragment (1 kb fragment). Reagents used per 25 μl reactions were 2.5X Enhancer and 2.5X Buffer solution from InVitrogen

Life Technologies, 2 mM $MgSO_4$, 1 uM primer, an aliquot dimethyl sulfoxide, 1 unit of Taq DNA Polymerase, .2 mM dNTPs, 2.5 units of Pfu DNA Polymerase, 2 μl of cDNA, and 13.5 μl of HPLC water. PCR condition were set at 94°C for 30 seconds (denaturing), 37 - 72°C for 30 seconds (annealing), 72°C for 30 seconds (synthesis), 35 cycles only, 72°C for 5 minutes (final elongation) followed by 4°C storage.

Polymerase Chain Reaction (patient samples)

Optimum PCR condition for the first fragment was performed using primers FOR3 and REVSEQ800 (See Figure 11) at 94°C for 30 seconds, 52°C for 30 seconds, 72°C for 30 seconds, 35 cycles only, 72°C for 5 minutes, followed by 4°C storage. For the second fragment PCR was performed using primers FORNEW and REVNEW at 94°C for 30 seconds, 55°C for 30 seconds, 72°C for 30 seconds, 35 cycles, 72°C for 5 minutes, followed by 4°C for precipitation. For the third fragment PCR was performed using primers FPGS RP2 and RP2. PCR conditions were set at 94°C for 30 seconds, 60°C for 30 seconds, 72°C for 30 seconds, 35 cycles, 72°C for 5 minute, and 4°C storage.

Agarose Gel Electrophoresis

Once PCR was completed, samples were loaded on 1% agarose gels for viewing of amplication products. To constitute these gels, 1.2 g of agarose powder was mixed with 120 ml of 1X Tris Acetate Buffer (40 mM Tris Base, 20 mM Glacial Acetic Acid, and 1 mM EDTA with pH ~ 8.0) and heated in a microwave for 1.5 minutes (mixing every 15 seconds to dissolve agarose powder). Ethidium bromide (7.5 μl) at a concentration of 5 μg/μl is added to the gel mixture. The mixture was then poured into a plastic tray, sealed with a casting tray containing a 25 well gel comb, and the gel was allowed to

polymerize at room temperature. The comb was then removed and the plastic tray, along with the agarose gel, was transferred to a horizontal electrophoresis chamber containing Tris Acetate Buffer, needed for optimum DNA mobility. Loading buffer (2 μl) was then added to 20 μl of each sample containing PCR products and each samples was loaded into a well of the agarose gel. Once the lid of the electrophoresis chamber was closed, 80 mA current was applied to the gel for 30 minutes, allowing the PCR products to migrate towards the positive electrode (care is taken to make sure that the products migrate towards the cathode). The DNA of each sample was then briefly visualized on an ultraviolet transluminator.

Gel Purification of PCR Products

DNA band were carefully excised from an agarose gel using a scalpel and each band was placed in a 1.5 ml centrifuge tube. Buffer QG (300 μl), containing guanidine thiocyanate, was added to gel slice to facilitate melting of the agarose gel when incubated at 60°C for 10 minutes. The gel containing the DNA band was vortexed every three minutes during heating until the gel was completely dissolved. Once the gel melted, 100 μl of isopropanol was added to the melted gel mixture, which ultimately allows binding of the DNA to the column. Next, the mixture was transferred to the mini column in a centrifuge tube, which served as a filtering device for elution of the DNA from the melted gel mixture, followed by brief centrifugation at maximum speed. The eluted gel mixture was then discarded and the columns were placed back in collection tubes. To wash the bound DNA, 0.75 ml of Buffer PE (containing ethanol) was added to the mini columns, and the columns were incubated at room temperature for 2 minutes, and then centrifuged at

13,0000 rpm speed for 1 minute. Again, the flow-through was discarded and columns are placed back in collection tubes and centrifuged at maximum speed for 1 minute to eliminate any residual ethanol from the Buffer PE. Once the gel mixture has completely filtered through the column and the DNA was washed free of contaminants, the mini columns were then transferred to a clean 1.5 ml collection tube and a 40 µl of 10 mM Tris-HCl, pH 8.5 was added to the center of the column membrane, incubated for 1 minute at room temperature, and centrifuged at maximum speed for 1 minute. This final step allows complete elution of clean DNA from the mini columns into the collection tube.

Sequences and Analysis of PCR Products

An aliquot (10 ng) of each sample was sequenced using 2 µl of 5 uM Primers. Sequencing was done at the DNA core lab at the Medical College of Virginia and SeqWright and Fisher Scientific Laboratory, and Single Nucleotide Polymorphisms were detected using the CodonCode Aligner computer program

Chapter 3

RESULTS

Preparation of RNA from control cell lines

Total RNA was isolated from the three control cell lines using the Trizol method (Protocol was obtained from the Institute of Genomic Research website at http://pga.tigr.org/sop/M019.pdf). Following isolation, 1 μg of total RNA from each cell line was loaded and run on a 1% denaturing agarose gel to assess its integrity. As was expected of normal intact RNA, the gel image in Figure 5 displays two distinct bands. The lower band represents the18S ribosomal subunit and the upper band represents the 28S ribosomal subunit. The light smear displayed above the upper band and between the two main bands represents other RNA species, including all of the mRNAs and also any minor RNA degradation. The prominent bands on this gel show stable RNA and adequate RNA concentration for our experiments.

Preparation of RNA from patient samples

For patient samples 1-6, RNA isolation was performed using the Quiagen method which required considerable less time when compared to the Trizol method (Protocol was obtained from the Institute of Genomic Research website at http://pga.tigr.org/sop/M019.pdf). Residual blood samples (1-2 ml) were obtained from the hematology laboratory at the Virginia Commonwealth University Hospital. Samples were collected in purple top tubes containing ethylenediaminetetraacetic acid (EDTA) to prevent blood clot formation. Samples were collected from patients with a white blood cell count of at least four, to ensure that sufficient RNA would be available to perform

multiple PCR reactions. Once RNA was extracted from samples 1-6, 1μl of 50 μl of total RNA was loaded and run on a nanochip of the Agilent 2100 Bioanalyzer. The Agilent 2100 Bioanalyzer then displayed a virtual gel, which represents a 1% denaturing agarose gel as it would appear with the same concentration of RNA from each sample. A representative sample of the virtual gels is illustrated in the upper left hand corner of Figure 6, with the 150 ng/μl ladder shown in the first lane. Like the1% denaturing agarose gel, the Agilent 2100 Bioanalyzer produces an electropherogram displaying two peaks, the shorter peak representing the 18S ribosomal unit and higher peak representing 28S ribosomal subunit. The first peak to the far left of each electropherogram represents the 25-nucleotide marker added to each sample. RNA integrity and concentration was determined according to the height of each peak, as well as by the amount of small peaks between the two major peaks. Sample 4, for example, has less RNA degradation than sample 5, as is evident by the increased amount of smaller peaks that exist between the two major peaks of sample 5. **(See Figure 6).**

Gel PCR products of control samples (800 bp and 1 kb)

Once RNA was isolated from the control samples and first stand cDNA synthesis was performed, amplification of the *fpgs* gene was performed using the polymerase chain reaction (PCR). DNA from the three control samples was amplified into two separate fragments using two sets of primers. Due to the high GC rich region of the 5' UTR region of the *fpgs* gene, amplification of the gene as a single 1.8 kb fragment was not feasible. A 24 μl reaction of the PCR products from each cell line was loaded and run on a 1% agarose gel for 30 minutes. In Figure 7, the first 800 bp fragment amplified from

each control cell line is displayed in lanes 1-6, with 24 µl reactions from each cell line loaded into two wells (for example, 24 µl PCR products from the CEM cell line was loaded into lanes 1 and 2). The bands representing the 1 kb fragments for each cell lines in lanes 7-9 are not as abundant as those representing the 800 bp fragments. Furthermore, the lower bands in lanes 7-12 represent primer dimer formation, which show decreased primer specificity for the DNA templates. However, due to the fact that I loaded 48 µl reaction for each cell line (24 µl in each well), I had sufficient PCR products for sequencing.

PCR and Sequencing primers used for 800 bp and 1 kb fragments

The primers used for amplification and sequencing of the two fragments of the control samples are shown in Figure 8. Two sets of primers were used for amplification and sequencing of each fragment. The first 800 bp fragment was amplified using forward primer FOR3, which starts 32 nucleotides upstream of the mitochondrial start site and ends at nucleotide 3. The reverse primer, MR2 starting at nucleotide 885 and ending at nucleotide 864 was also used to amplify the first 800 bp fragment. The 1 kb fragment was amplified using forward primer MF1, which starts at nucleotide 809 and ends at nucleotide 826. Thus, the sequences of the two PCR fragments overlap by 38 nucleotides. The reverse primer, RP2 starts at nucleotide 1814 and ends at nucleotide 1837. Sequencing of the 800 bp fragment was performed using primers For6 and F21asn. Primers FORSEQ 1 kb and REVSEQ 1 kb were used to sequence the 1 kb fragment.

Amplification of samples 1-6

Following amplification and sequencing of the two fragments of the control samples, PCR was then performed on patient samples 1-6. Figure 9 displays the PCR products of the first 800 bp fragment of samples 1-6. Again, 24 μl of PCR product from each patient sample was loaded into 2 separate wells. Sample 2 was loaded into 3 separate wells following repeated PCR. Primer dimers (low molecular weight bands formed on the bottom of the agarose gel) were removed during gel purification. The patient samples proved to be more difficult than the control samples. Because human peripheral blood cells have a limited amount of the *fpgs* mRNA, I first experienced difficulty amplifying the gene in 2 fragments, then I experienced further difficulty sequencing the PCR products. Before we began this project, we anticipated some difficulty obtaining sufficient fpgs RNA from human peripheral blood. We then decided to amplify the fpgs gene into 3 fragments with the first fragment at 465 nt, the second at 964 nt and the third at 615 nt in length. Six primers were designed for PCR and 4 primers were designed for sequencing (**See Figure 11**). This solved the PCR amplification problem, as well as the sequencing problem. An example of a good versus poor sequencing run is illustrated in Figure 10.

Amplification of 465 bp fragment of the *fpgs* gene

Once I began amplification of the *fpgs* gene into three fragments, I obtained better results. Figure 13 is an image of the most upstream 465 bp PCR product from samples 1-13. An aliquot of 10 μl reactions was loaded on to the 1% agarose gel and 80 mA of current was applied to the gel for 30 minutes. While some of the PCR products produced sharp bands, others such as samples 1 and 9 were later repeated to obtain a greater

concentration of PCR products. Sequencing of the PCR products was performed using reverse primer, REVSEQ800, as illustrated in Figure 11.

Amplification of the 964 bp fragment

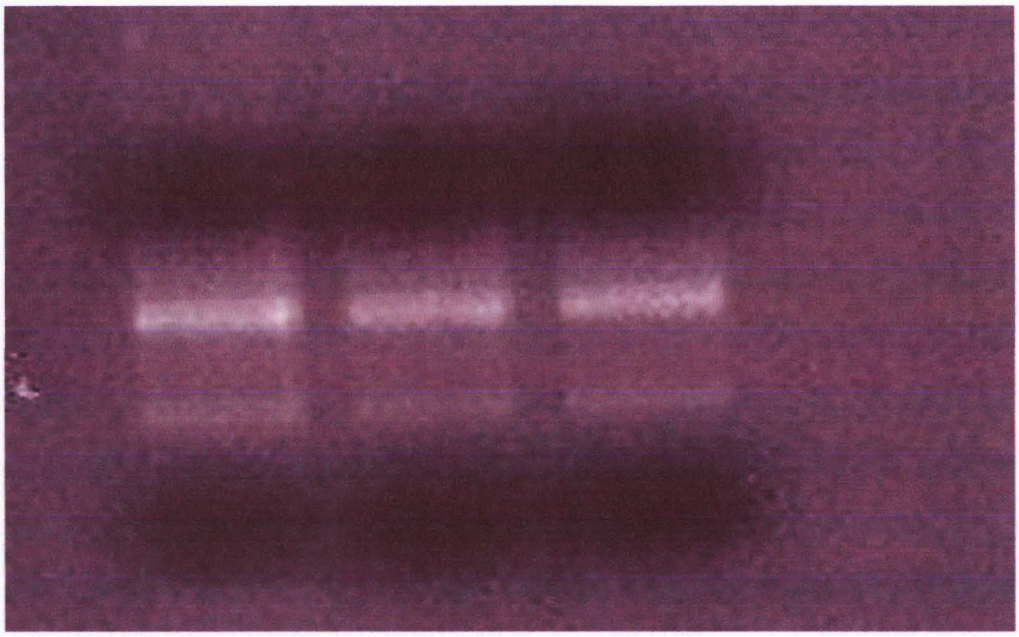
The 964 bp fragment of the fpgs gene was amplified using forward primer, FORNEW and reverse primer REVNEW. Sequencing was performed using MF1 primer and RMD primer. Figure 14 is an illustration of the middle 964 bp fragment of samples 1-18, loaded and run on the 1% agarose gel following PCR. Primer dimers are removed during gel purification. PCR products of samples, such as sample 3 and 5, were later repeated to obtain a greater amount of product needed to obtain good sequencing results.

Amplification of downstream 615 bp fragment

Figure 15 displays the downstream 615 bp PCR product of samples 1-8, loaded and run on the 1% agarose gel. This fragment of the fpgs gene was amplified using forward primer FPGS RP2 and reverse primer RP2. Sequencing was performed using reverse primer RP2. Despite the primer dimers that formed in each PCR reaction, we were able to obtain sufficient amount of the PCR products for sequencing. Again, primer dimers are removed during gel purification.

Fig 5. RNA of all 3 control cell lines

Total RNA (1 μg) of the three control cell lines, CEM, MCF-7, and WiDr, were loaded and run on a 1% denaturing agarose gell at 80 mV for 30 minutes. Upper bands represent 18S ribosomal RNA subunit and lower bands represent 28S ribosomal RNA subunit. The appearance of 2 distinct bands shows intact stable RNA.



CEM MCF-7 WiDr

Fig. 5

Fig 6. Ribosomal RNA subunits of samples 1 – 6 displayed by Agilent 2100 Bioanalyzer

Electropherogram display of peaks representing 18S and 28S ribosomal RNA subunit of total RNA from patient samples 1- 6. The first peak of each electropherogram represents an aliquot of marker that was loaded into each sample. The lower peak represents the 18s ribosomal RNA subunit and the higher peak represent the 28S ribosomal RNA subunit. The smallest minor peaks shown between the 18S and 28S peaks represent any RNA degradation in the samples. The image below is an illustration, displayed by the Agilent 2100 Bioanalyzer, of a virtual denaturing gel of total RNA from samples 1-6. The left most lane represents the 150 ng/μl ladder and lanes 1-6 represents samples 1-6. The ladder has 6-7 peaks of different sizes, which calibrates the instrument, allowing correct measurement of the 18S and 28S ribosomal RNA peaks in each samples.

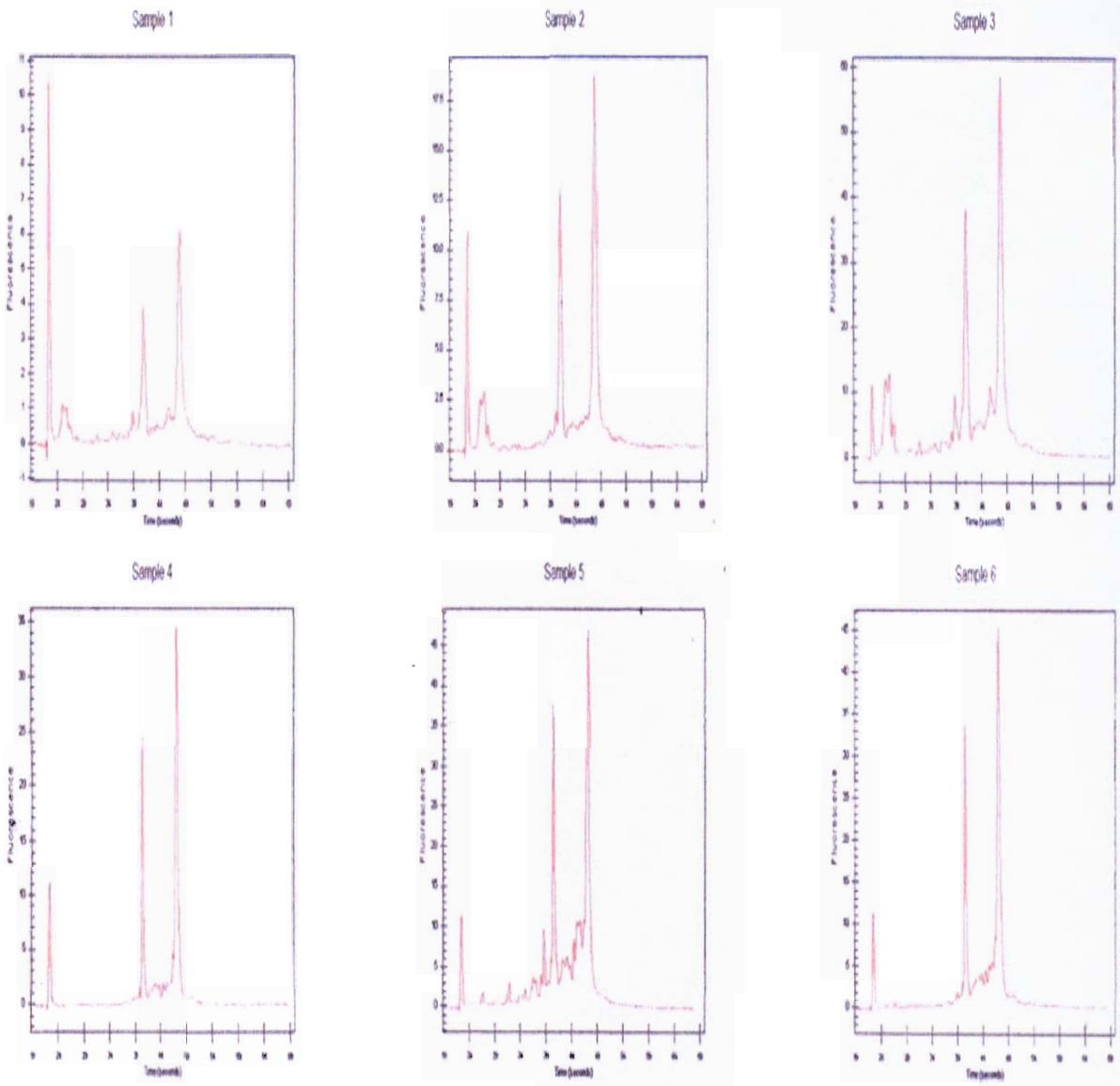
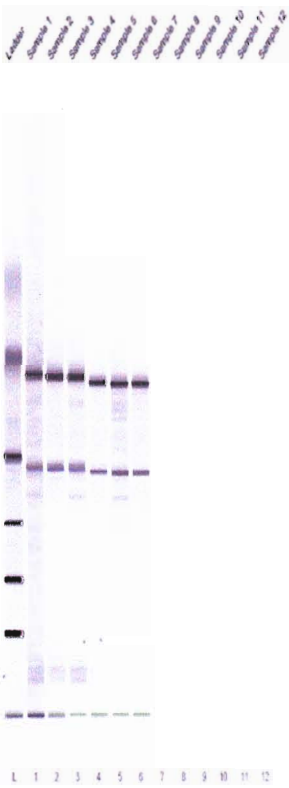


Fig. 6

Assay: Eukaryote Total RNA Nano
Data Path: C:\Program Files\Agilent 2100 Bioanalyzer\Bio Sizing\Data\2005-05-25

Read: 25/05/2005 14:19:40 (A.02.12 SI292)
Modified: 25/05/2005 14:36:07 (A.02.12 SI292)



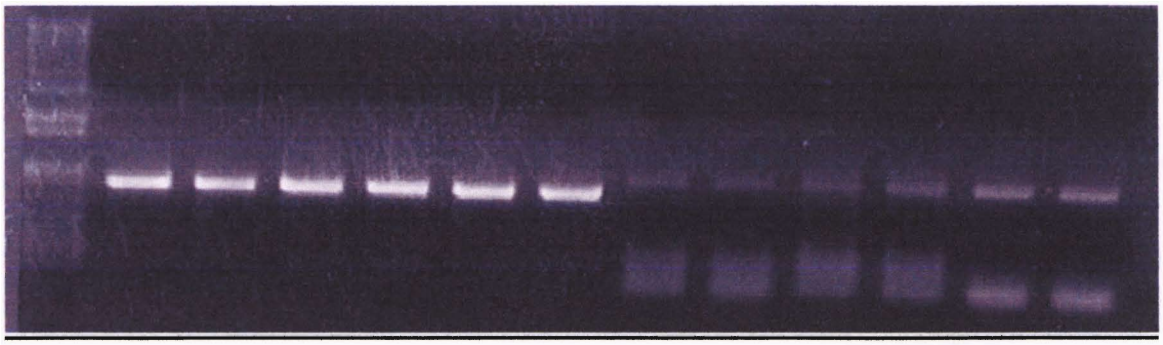
Instrument:	G2938B, Serial# DE20901552, Firmware Version A.01.16
Assay:	C:\Program Files\Agilent 2100 Bioanalyzer\Bio Sizing\Assays\RNA\Eukaryote Total RNA Nano.asy
Title:	Eukaryote Total RNA Nano
Version:	1.4
Comments:	Copyright 1999 - 2002 Caliper Technologies Corp.

Ladder Concentration:	150 ng/ul	Baseline:	19 Seconds
Min Peak Height:	0.5 (above baseline)	Filter Width:	1 Seconds
Slope Threshold:	0.8 /Second	Baseline Plateau:	0.5 Seconds
Min Peak Width:	0.5 Seconds	Polynomial Order:	6
Start Time:	20 Seconds		
End Time:	69 Seconds		

Notes:

Fig 7. PCR product of control cell lines (800 bp and 1 kb)

A Display of the 800 bp and 1 kb PCR products of control cell lines, CEM, MCF-7, and WiDr is shown in figure 7. The 1 kb ladder (8 μl) was loaded into the first lane and the 800 bp PCR products of the three control lines were loaded into lanes 7 – 12. The low molecular weight bands formed at the bottom of the agarose gel illustrate primer dimers formed during PCR.



L 1 2 3 4 5 6 7 8 9 10 11 12

Lane L = Ladder
Lanes 1,2 = CEM (800 bp)
Lanes 3,4 = MCF-7 (800 bp)
Lanes 5,6 = WiDr (800 bp)
Lanes 7,8 = CEM (1 kb)
Lanes 9,10 = MCF-7 (1 kb)
Lanes 11,12 = WiDr (1 kb)

Fig. 7

Fig 8. Primers used for PCR and sequencing of two fragments of the control samples

A set of primers were used for sequencing of the 800 bp and 1 kb fragments of the *fpgs* gene. The 800 bp fragment was amplified using forward primer FOR3 and reverse primer MR2. The 1 kb fragment was amplified using forward primer MF1 and reverse primer RP2. A separate set of primers was used to sequence the 800 bp and 1 kb PCR products of the *fpgs* gene. The 800 bp fragment was sequenced using forward primer For6 and reverse primer F21asn. The 1 kb fragment was sequenced using forward primer FORSEQ 1 kb and reverse primer REVSEQ 1 kb.

Primer	Sequence	Reverse or Forward	Position in Sequence
800 bp fragment			
For3	5'GCCGGGGGCGCCGGGACTATG 3'	Forward	-32-3
MR2	5' CTCTATACCTGTGCCGATGC 3'	Reverse	864-885
1 KB fragment			
MF1	5' CCAGCAGATCTCATGTC 3'	Forward	809-826
RP2	5' GCACCTAGTATGTAAGTTCATGG 3'	Reverse	1814-1837

Sequencing primers:

800 bp fragment			
For6	5' CCGAGCATGGAGTACCA 3'	Forward	121-137
F21asn	5' GGTACAGTTCCATGGCTTCC 3'	Reverse	228-247
1 KB fragment			
FORSEQ 1 kb	5' AACCTGACAGAGGTGTCATCC 3'	Forward	1321-1341
REVSEQ 1 kb	5' CCAGTGTCACTGTGAAGTTC 3'	Reverse	1361-1381

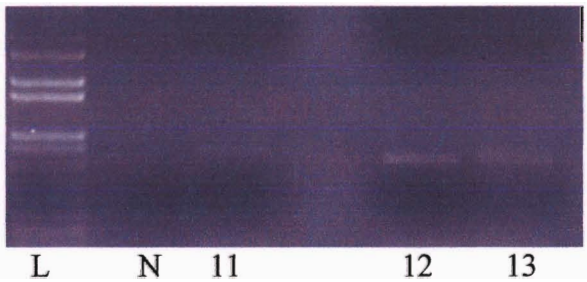
Fig. 8

Fig 9. PCR products of samples 1-6 (800 bp)

The 800 bp PCR fragments from samples 1 – 6 were loaded and run on 1% agarose gel at 80 mV for 30 minutes. The 1 kb ladder (8 µl) was loaded into the first lane and 24 µl of the 800 bp PCR products of each sample was loaded into 2 separate wells in lanes 1 – 10. Sample 2 was repeated and loaded into three separate wells. The negative control was loaded into the second lane of the bottom gel (lane N).

L 1 2 3 4 5 6 7 8 9 10

Lanes 1,2 = sample 1 (800 bp)
Lanes 3,4 = sample 3 (800 bp)
Lanes 5,6 = sample 4 (800 bp)
Lanes 7,8 = sample 5 (800 bp)
Lanes 9,10 =sample 6 (800 bp)
Lane L = Ladder



L N 11 12 13

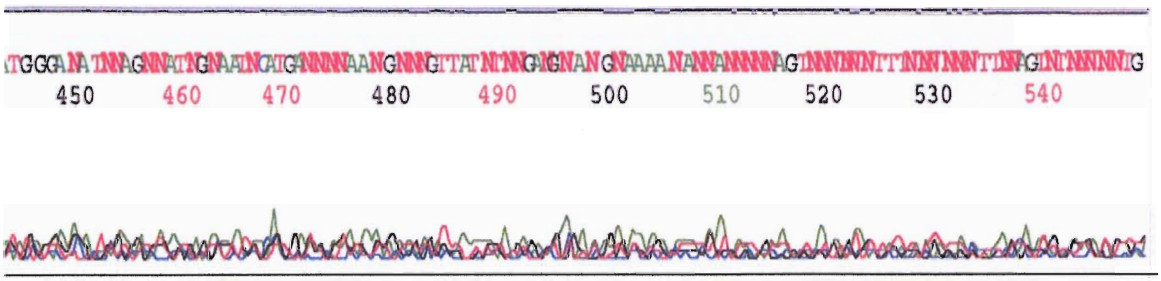
Lane N= negative

Lanes 11,12, 13 = sample 2 (800 bp)

Fig. 9

Fig 10. Good and poor sequencing results of 1 kb fragment of samples 1-6

The top figure is an illustration of a chromatogram of the 1 kb PCR product of one sample that was poorly sequenced. The lower chromatogram figure shows an example of a successful sequencing run performed on the 800 bp PCR product of one of the samples. In the bottom chromatogram, the peaks of the raw data are distinguishable in height and color. Unlike the top chromatogram, the program was able to successfully label each peak with the appropriate nucleotide. Furthermore, there is little to no background peaks in the bottom chromatogram, which eliminates any chance of misinterpreting sequencing errors for a single nucleotide polymorphism.


450
460
470
480
490
500
510
520
530
540

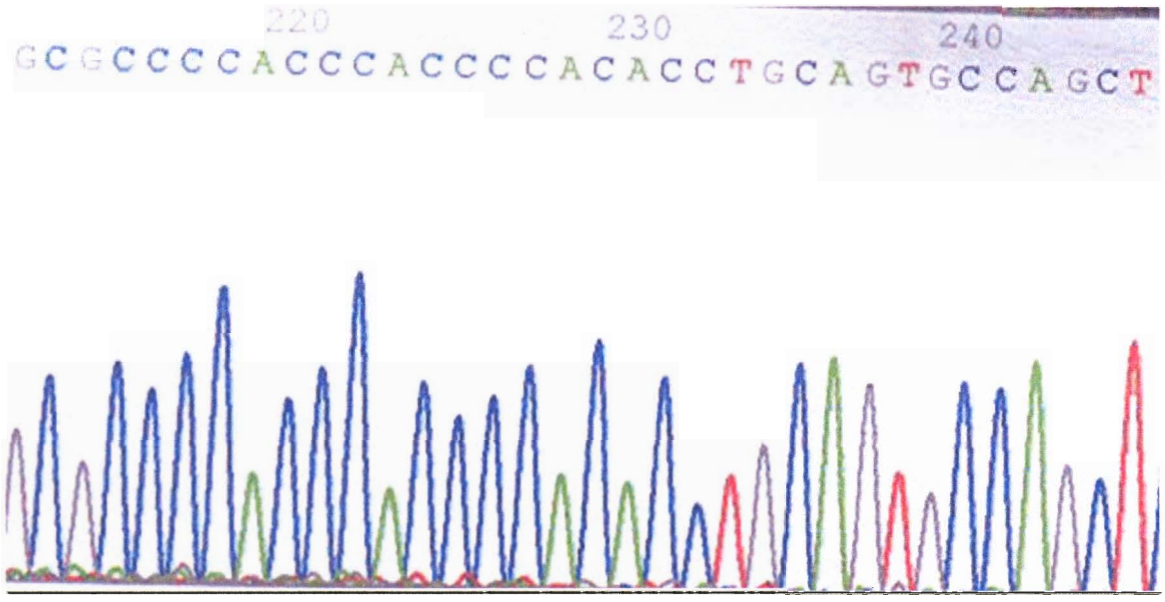

220
230
240
GCGCCCCACCCACCCCACACCTGCAGTGCCAGCT

Fig. 10

Fig 11. Primers used for PCR and sequencing of the three fragments of the *fpgs* gene

Three sets of primers were used for sequencing of the 465 bp, 964 bp, and 615 bp fragments of the *fpgs* gene. The 465 bp fragment was amplified using forward primer FOR3 and reverse primer RevSeq800. The middle 964 bp fragment was amplified using forward primer FORNEW and reverse primer REVNEW. The downstream 615 bp fragment was amplified using forward primer FPGS RP2 and reverse primer RP2. A separate set of primers was used to sequence the three PCR fragments of the *fpgs* gene. The upstream 465 bp fragment was sequenced using reverse primer RevSeq800 and the middle 964 bp fragment was sequenced using forward primer MF1 and reverse primer RMD. The downstream 615 bp fragment was sequenced using reverse primer RP2.

PCR Primers

Primer	Sequence	Reverse or Forward	Position in Sequence
Upstream 465 bp			
For3	GCCGGGGGCGCCGGGACTATG	Forward	-32-3
RevSeq800	TGAAGAGCTCAGGACTGATGG	Reverse	437-458
Middle 964bp			
FORNEW	AAGCTATGGCCTGAAGACGGGATT	Forward	657-380
REVNEW	AGGGCAGAAGACGGCATAGTCAAA	Reverse	1297-1320
Downstream 615 bp			
FPGS RP2	GTTCGAGTCTTGCTCTTCAATGCTAC	Forward	1322-1347
RP2	GCACCTAGTATGTAAGTTCATGG	Reverse	1814-1837

Sequencing Primers

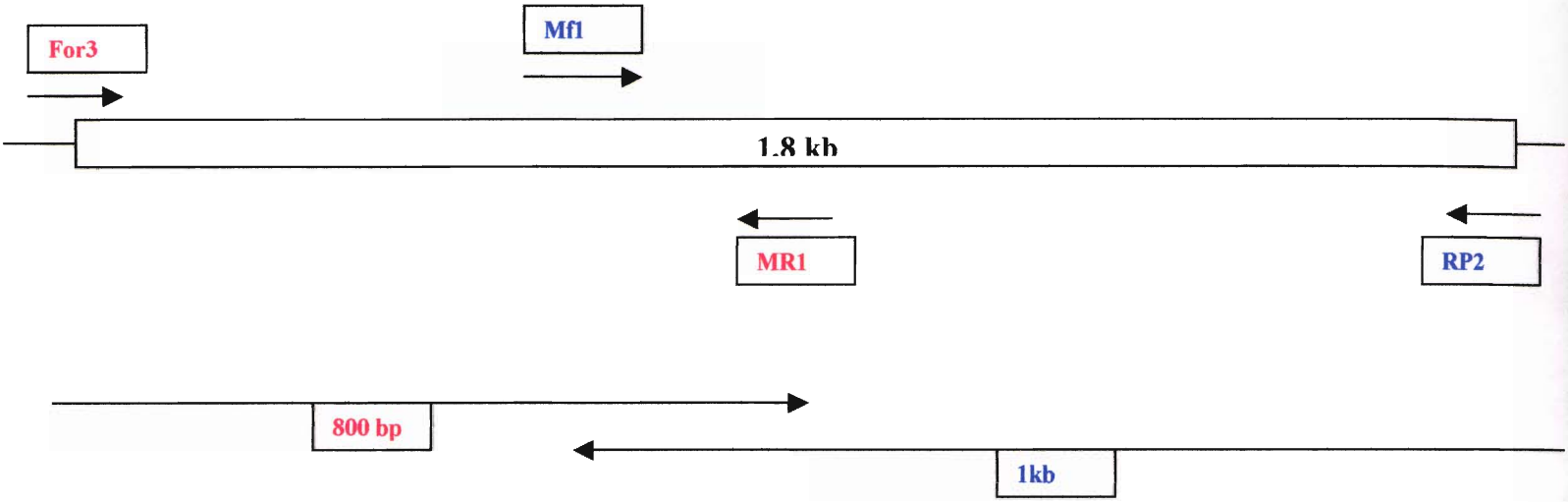
Upstream 465 bp			
RevSeq800	TGAAGAGCTCAGGACTGATGG	Reverse	437-458
Middle 964 bp			
MF1	CCAGCAGATCTCATGTC	Forward	809-826
RMD	GTTCGAGTCTTGCTCTTCAATGCTAC	Reverse	969-990
Downstream 615 bp			
RP2	GCACCTAGTATGTAAGTTCATGG	Reverse	1814-1837

Fig. 11

Fig 12. Diagram of primers used for amplification of the gene into 2 and 3 fragments

A schematic of the PCR primers used to amplify the fpgs gene into 2 and 3 fragments is illustrated in Figure 12. The top figure represents the 1.8 kb cDNA construct of the *fpgs* when amplified into 2 fragments, an 800 bp and 1 kb fragment. The primers used to amplify the 800 bp fragment are forward primer For3 (shown in red in the top figure) and reverse primer MR1 (also shown in red in the top figure). The 1 kb PCR fragment was amplified using forward primer MF1 (shown in blue in the top figure) and reverse primer RP2 (also shown in blue in the top figure). Following unsuccessful amplification of the *fpgs* gene into 2 fragments using patient samples, attempts were made to amplify the gene into three fragments. The bottom figure represents the 1.8 kb cDNA construct amplified into 3 fragments, a 465 bp, a 964 bp and downstream 615 bp fragment. The primers used to amplify the upstream 465 bp fragment are forward primer For3 (shown in red in the top figure) and reverse primer Revseq800 (shown in red in the bottom figure). The 964 bp PCR fragment was amplified using forward primer FORNEW (shown in blue in the bottom figure) and reverse primer REVNEW (also shown in blue in the bottom figure). The primers used to amplify the downstream 615 bp fragment are forward primer FPGSRP2 (shown in brown in the bottom figure) and reverse primer RP2 (shown in brown in the bottom figure).

Overlapping PCR primer sequences of the 2 fragments



Overlapping PCR primer sequences of 3 framents.

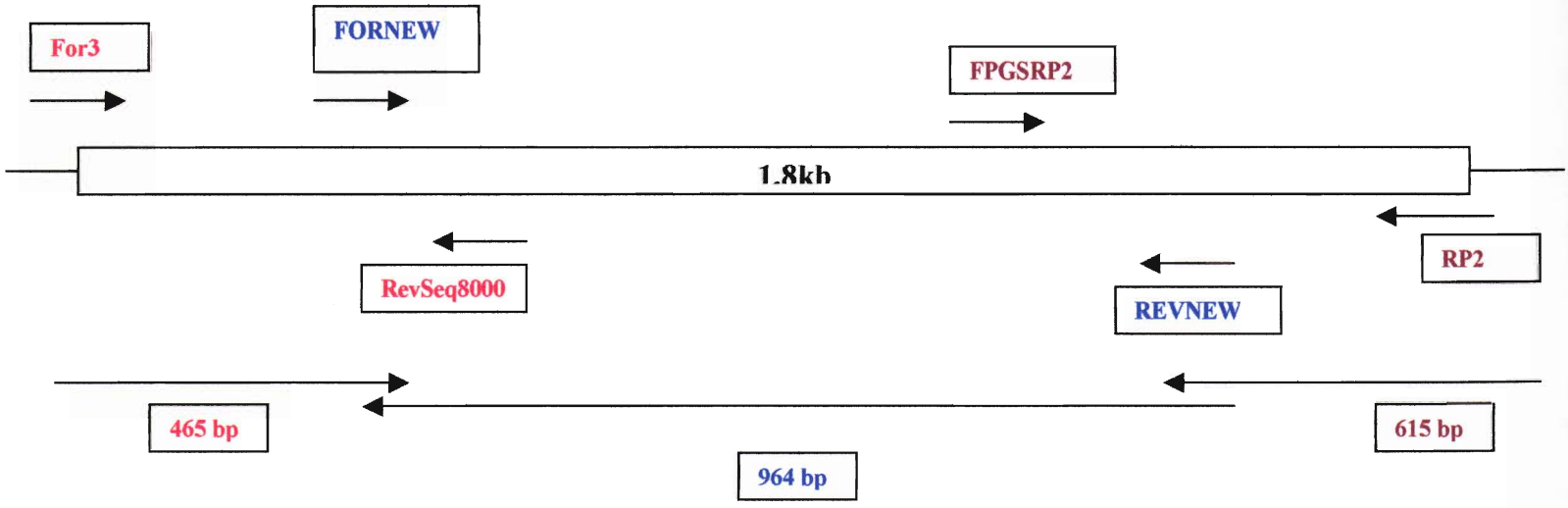
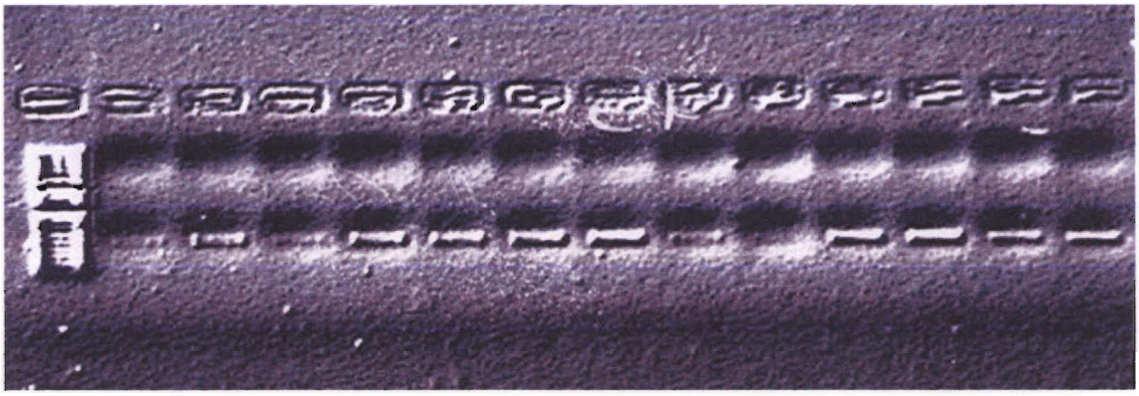


Fig. 12

Fig 13. Example of upstream 465 bp PCR products

The 465 bp PCR fragments from ten of the samples were loaded and run on 1% agarose gel at 80 mV for 30 minutes. The 1 kb ladder (8 μl) was loaded into the first lane and 24 μl of the 465 bp PCR products of samples 1 - 13 were loaded into each well in lanes 1 – 13. Sample 26 and 38 were later repeated to obtain a higher concentration.



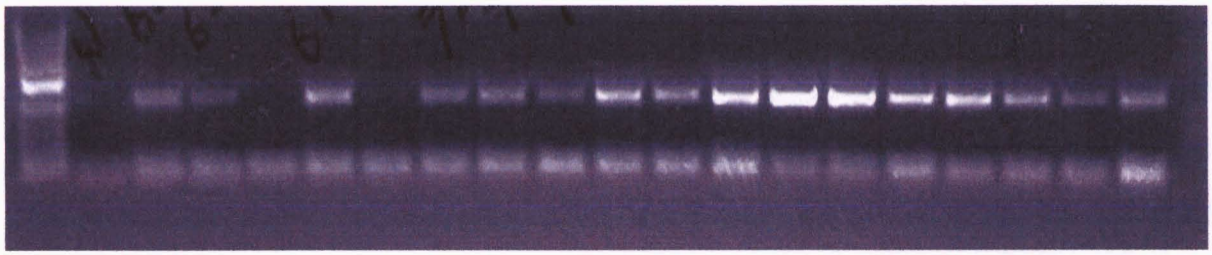
L 1 2 3 4 5 6 7 8 9 10 11 12 13

Lane 1 = sample 25
Lane 2 = sample 26
Lane 3 = sample 27
Lane 4 = sample 28
Lane 5 = sample 29
Lane 6 = sample 30
Lane 7 = sample 31
Lane 8 = sample 32
Lane 9= sample 38
Lanes 10 = sample 39
Lane 11 = sample 40
Lane 12 = sample 41
Lane 13 = sample 42
Lane L = Ladder

Fig. 13

Fig 14. Example of 964 bp PCR products

964 bp PCR fragments from 18 of the samples were loaded and run on 1% agarose gel at 80 mV for 30 minutes. The 1 kb ladder (8 μl) was loaded into the first lane and 24 μl of negative control was loaded into the second lane. The 964 bp PCR products of these samples (24 μl) were loaded into each well in lanes 1 – 18. Sample 66 and 68 were later repeated to obtain higher concentrations.



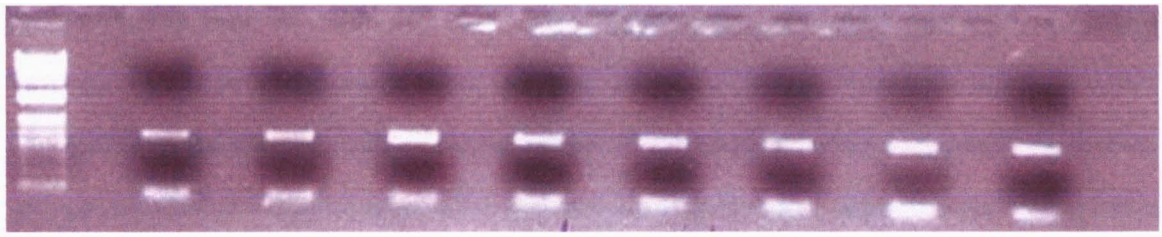
L N 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

Lane N = Negative control
Lane 1 = sample 61
Lane 2 = sample 62
Lane 3 = sample 66
Lane 4 = sample 67
Lane 5 = sample 68
Lane 6 = sample 72
Lane 7 = sample 73
Lane 8 = sample 75
Lane 9 = sample 76
Lane 10 = sample 80
Lane 11 = sample 82
Lane 12 = sample 83
Lane 13= sample 86
Lane 14 = sample 87
Lane 15= sample 88
Lane 16 = sample 89
Lane 17 = sample 90
Lane 18 = sample 91
Lane L = Ladder

Fig. 14

Fig 15. Example of downstream 615 bp PCR products

The downstream 615 bp PCR fragments from eight of the samples were loaded and run on 1% agarose gel at 80 mV for 30 minutes. The 1 kb ladder (8 μl) was loaded into the first lane (lane L) and 24 μl of the 615 bp PCR products of each sample was loaded into each well in lanes 1 – 8. All eight samples were successfully amplified.



L 1 2 3 4 5 6 7 8

Lane 1 = sample 44
Lane 2 = sample 45
Lane 3 = sample 47
Lane 4 = sample 49
Lane 5 = sample 51
Lane 6 = sample 52
Lane 7 = sample 53
Lane 8 = sample 54
Lane L = Ladder

Fig. 15

Chapter 4

Detection of Single Nucleotide Polymorphisms in *Fpgs* Gene

Use of the CodonCode Aligner to detect SNPs

The CodonCode Aligner is a multifunctional user-friendly software program designed by the CodonCode Corporation. This program is capable of aligning sequences, editing sequences, and detecting mutations. Once we obtained a license for use of this program, sequences were imported and aligned as traces into the program folder. Once sample traces were aligned, we clipped the beginning and end of the sequences to remove undecipherable peaks (See Figure 10 for an example of unreadable sequences). The program also allowed viewing of the sequences in reverse complement forms, which optimized sequence analysis of the reverse strand. Upon alignment of the sequences, the program determines the consensus sequence, which it displays at the bottom of the screen, labled contig 1 (See Figure 16). We then instructed the program to detect mutations beginning at a predetermined start codon and ending at a predetermined stop codon of the sequence (users may determine where to begin and end sequence analysis). Following sequence analysis, the CodonCode Aligner program tagged various areas of each sequence indicating sequence quality and mutations types, such as single nucleotide polymorphisms (See figure 16). The tags, which are color coded, can also be modified in intensity and color by the user (See figure 16). If a sequence differed by one or more nucleotides from the consensus sequence, the program tagged that base in the consensus sequence, labeled contig1 in figure 16 (See figure 16). In Figure 16, for example, the consensus base is A/A, which is highlighted in red in sequence 29C and the consensus

sequence (contig 1). In the bottom sequence labeled 40C, however the G/G homozygote allele is expressed at this position, which the program highlights in the consensus sequence using a red tag. So essentially, one method of detecting a SNP using this program was to scan the consensus sequence for tagged bases. Another method of detected SNPs was to scan every single sequence for tags and then analyze every tagged area for possible SNPs. Though extremely tedious and time consuming, this method proved to be the most effective form of sequence analysis. The program not only tagged mutations and poor quality peaks, but it also tagged every position where peaks were not as high as adjacent peaks or where peaks were not equidistance to adjacent peaks. For this, we spent a great deal of time analyzing tagged areas of the sequences that were ultimately accepted as normal sequence. It was important to analyze all tagged areas to ensure that no SNPS were overlooked. Figure 23, for example, shows an important SNP at nucleotide 1334 that sequestered in a poor quality region of the sequence. Following sequence analysis, repeated PCR, and repeated sequence analysis, we eliminated a large number of false positive SNPs. An example of a false positive is illustrated in Figure 17. This apparent SNP was not observed in repeat sequences over this region. Ultimately, it was determined that a SNP was real if the base at the position where two nucleotides were observed, decreased in height when compared to the base at the corresponding position in the wild type sequence. For example, in Figure 18 base G in the A/G SNP at nucleotide 64 deceased in height by approximately one-half when compared to the G/G wild type sequence below. Similarly Figure 19 base G in the G/A SNP at nucleotide 123 deceased in height when compared to the G/G wild type sequence.

SNPs Detected

Due to the fact that the *fpgs* coding sequence was amplified in three pieces, some fragments were more successfully amplified in some samples than in others. For example, 88 samples were successfully amplified for the first 465 nucleotides of the fpgs gene, while 78 samples were successfully amplified for the middle fragment spanning 964 nucleotides. The last 615 base pair fragment was successfully amplified in 85 of the samples. For the remaining 15-20 samples, PCR or sequencing was unsuccessful due to RNA/ DNA instability or limited amount of RNA availability (all samples are unique to each patient and were available in very limited supply). Following sequencing of approximately 78-88 samples, polymorphic sites were found at nucleotides 64, 123, 253, 423, 1334 and 1781 (See Figure 16-21). These SNPs are summarized in table 1. Sequencing of the first 465 nucleotides was performed on 88 total samples. Of the first PCR fragment sequenced, in a total of 88 samples, 48 samples express a heterozygote G/A genotype at nucleotide 64 (~ 55 %), 1 sample expressed the homozygote A/A at this site (~ 1 %), and the remaining samples expressed the homozygote G/G genotype at this position (~ 44 %). At nucleotide 123, 6 samples (samples 12, 45, 88, 90, 97, 99) expressed the heterozygote G/A genotype (~ 7 %), and one sample (sample 74) expressed the homozygote A/A allele at this site (~ 1 %), and 81 (~ 92 %), samples express the wild type G/G allele. Of the 88 samples sequenced, only one sample (sample 80) expressed a heterozygote C/T allele at nucleotide 253 (~ 1 %). SNPs that were detected in less than 10 % of the samples were confirmed by repeated amplification and repeated sequencing cDNA from both directions. For the second fragment, at nucleotide 423, two samples (32 & 97) expressed a heterozygous C/G allele (~ 3 %), while the remaining samples

expressed a homozygous C/C allele (~ 97 %). At nucleotide 1334, one sample (26) expressed T/C allele while the remaining samples were C/C at this site (~ 1 %). Again, this SNP was found by reamplifying and resequencing cDNA from this sample. The homozygous change at nucleotide 64 encodes a conservative amino acid change from a valine to isoleucine. The G123A change expressed in six of the samples produced no change in the proline at this codon. Similarly, at nucleotide 423 the SNP observed was silent, that is, both wild type codon (CGC) and the mutant codon (CGG) encoded the amino acid, arginine. At nuclotide 253, however, a C to T base change produced a change from arginine to tryptophan. The heterozygote T/T allele at nucleotide 1334, expessed in the *fpgs* gene of sample 26, caused a valine to alanine amino acid change. The G1781A SNP was detected downstream from the stop codon in sample 40 (~ 1 %). A summary of these results is displayed in Table 1.

Fig 16. Example of tagged SNPs in consensus sequence.

A portion of the downstream 615 bp sequence of sample 29 and 40 is displayed in Figure 16. The consensus sequence of this portion of the gene is also displayed at the bottom of the figure, labeled Contig1. Because the two sequences of samples 29 and 40 differ by one nucleotide at the position highlighted by the program in red, the program also highlights that nucleotide position in the consensus sequence displayed at the bottom of the screen (shown by the arrow).

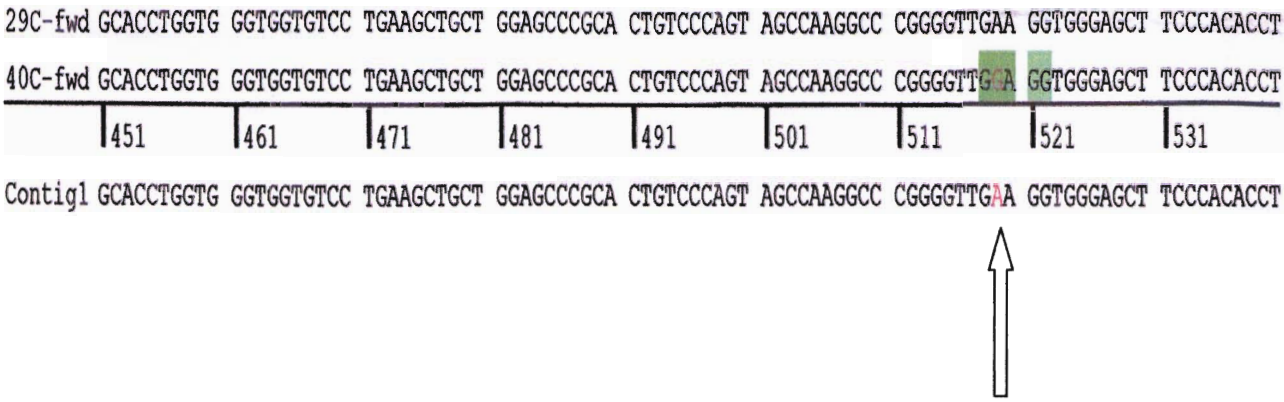

29C-fwd GCACCTGGTG GGTGGTGTCC TGAAGCTGCT GGAGCCCGCA CTGTCCCAGT AGCCAAGGCC CGGGGTTGAA GGTGGGAGCT TCCCACACCT
40C-fwd GCACCTGGTG GGTGGTGTCC TGAAGCTGCT GGAGCCCGCA CTGTCCCAGT AGCCAAGGCC CGGGGTTGCA GGTGGGAGCT TCCCACACCT
451
461
471
481
491
501
511
521
531
Contig1 GCACCTGGTG GGTGGTGTCC TGAAGCTGCT GGAGCCCGCA CTGTCCCAGT AGCCAAGGCC CGGGGTTGAA GGTGGGAGCT TCCCACACCT

Fig. 16

Fig 17. Example of false positive SNPs

False positive SNPs frequently appeared during sequence analysis. This false positive shown in the top chromatogram and highlighted in green by the CodonCode Aligner program originally appeared to be a SNP because of the double peak at this position (shown by the arrow) as well as the lack of background sequencing error in adjacent peaks. This SNP, however, did not appear during repeated sequence analysis of repeated PCR products.

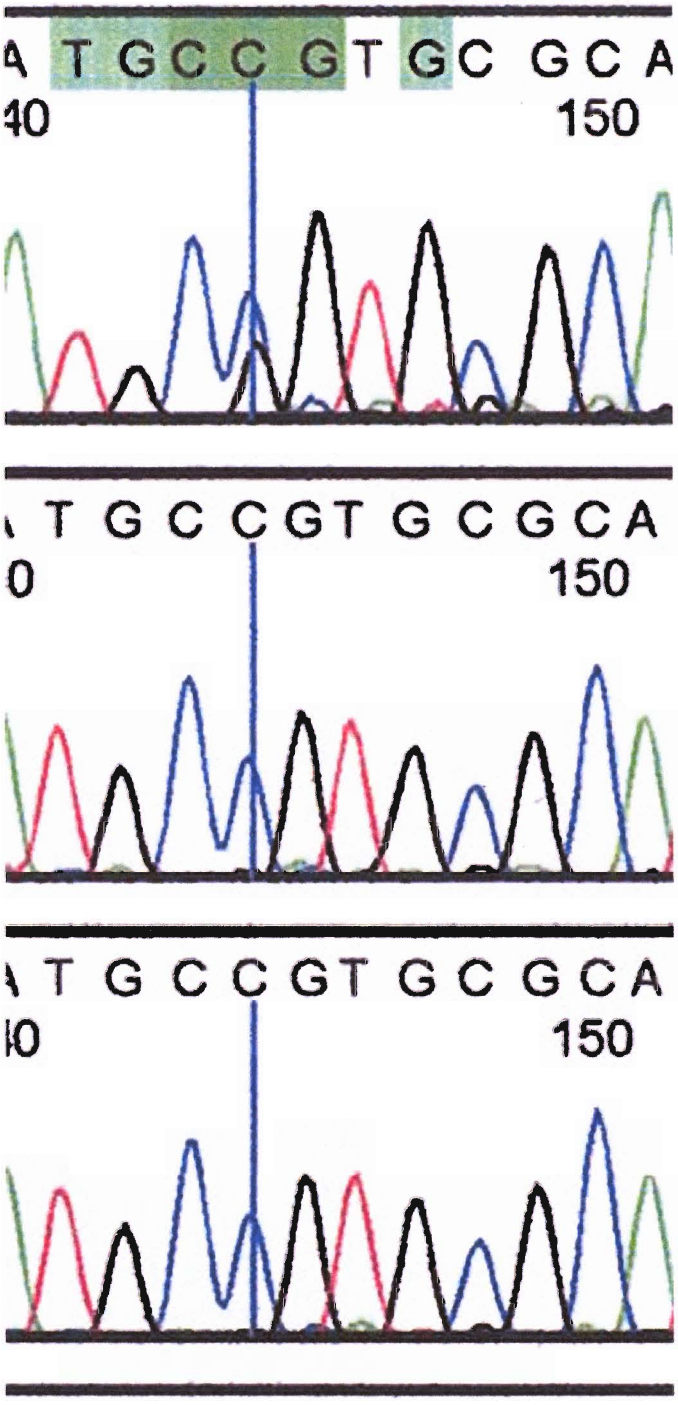
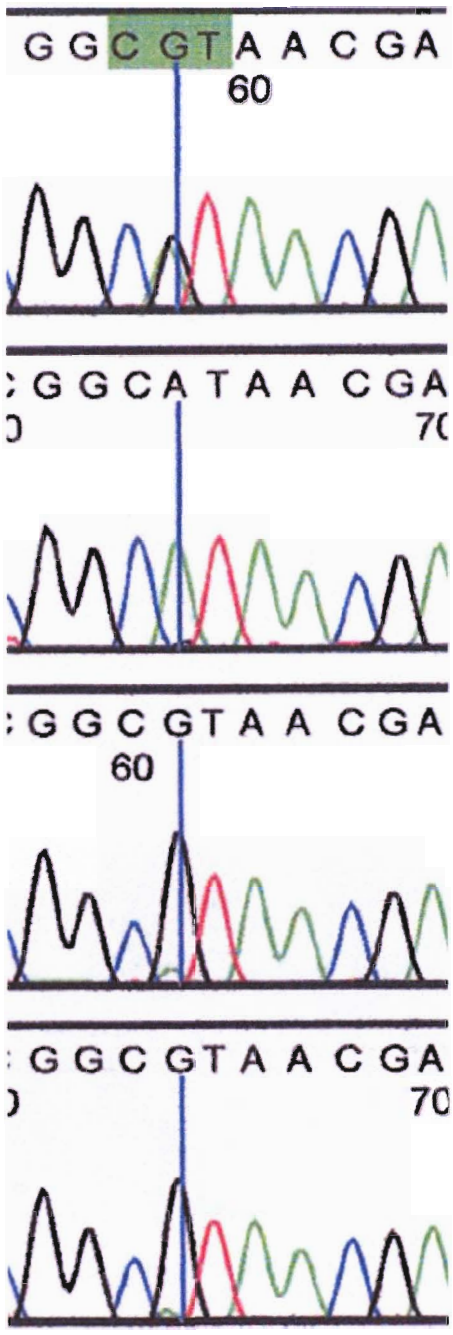


Fig. 17

Fig 18. SNP 64 A/G Valine/Isoleucine

SNP 64 is highlighted by the CodonCode Aligner program in the top chromatogram. There are no background sequencing errors, which optimizes sequence analysis and accuracy. The lower three chromatograms have a single peak at nucleotide 64, representing either the homozygote A/A or G/G Allele. The peaks of these lower four chromatograms are approximately double the height of the peak in the chromatogram of the heterozygote G/A SNP.



1. Heterozygote G/A

2. Homozygote A/A

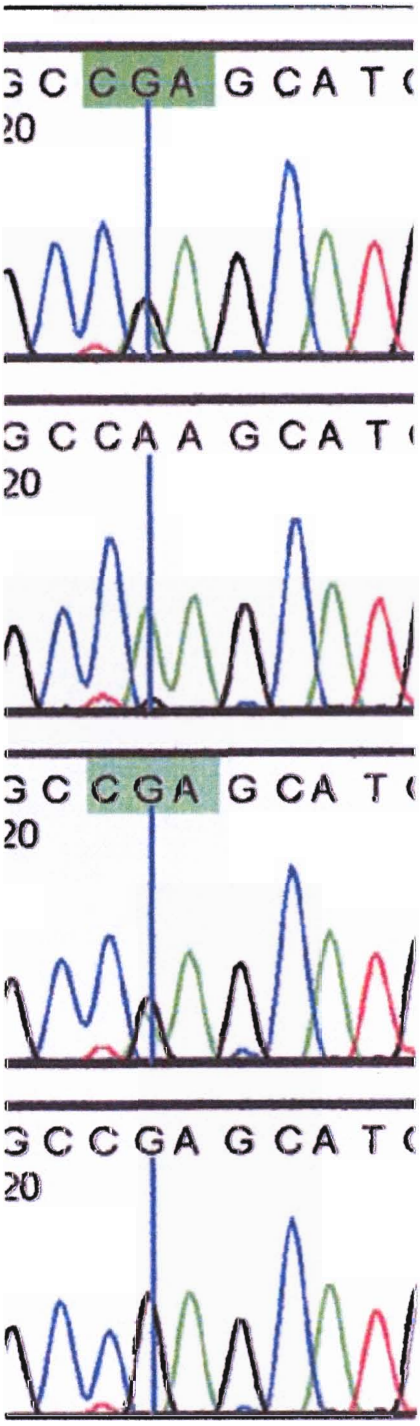
3. Homozygote G/G

4. Homozygote G/G

Fig. 18

Fig 19. SNP 123 A/G Proline/Proline

SNP 123 is highlighted in the first three chromatograms below, while the wild type homozygote G/G allele is displayed in the bottom chromatogram. The homozygote A/A SNP is shown in the second chromatogram. Again, there are no background sequencing errors, optimizing sequence analysis and accuracy, and the peaks of the homozygote alleles are higher than the peaks of the heterozygote alleles.



1. Heterozygote G/A

2. Homozygote A/A

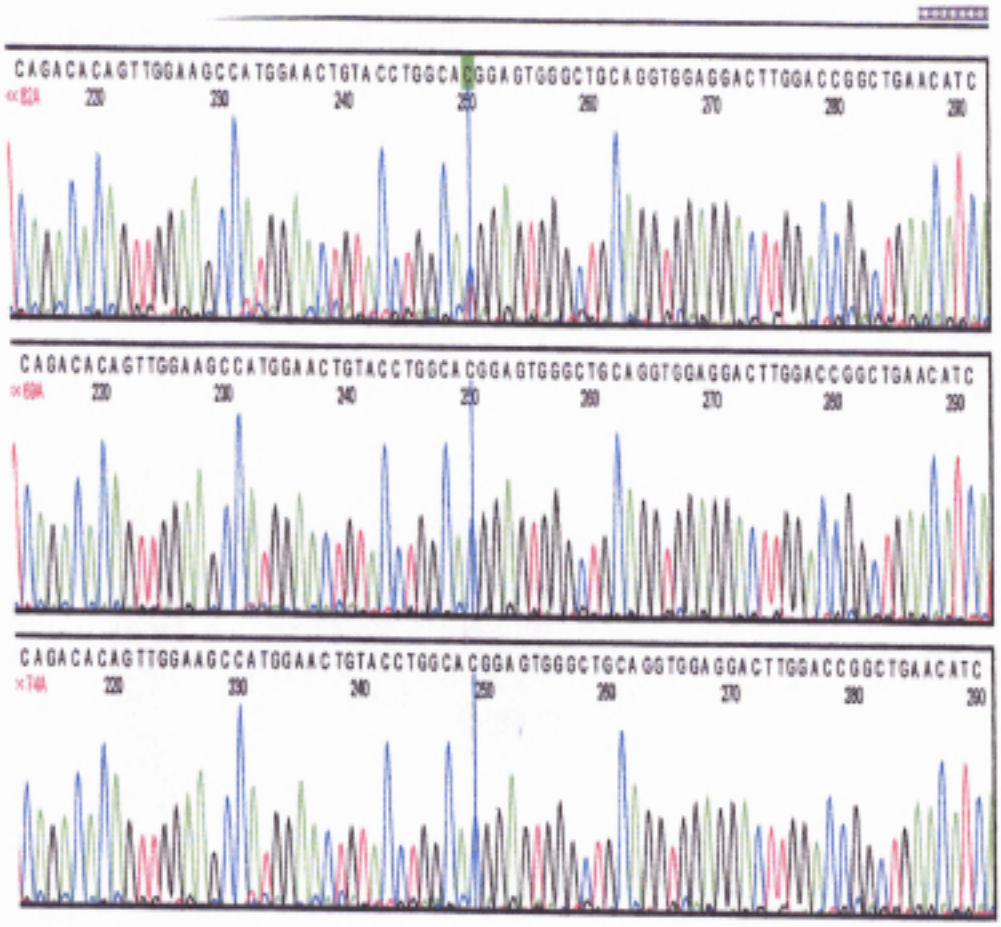
3. Heterozygote G/A

4. Homozygote G/G

Fig. 19

Fig 20. SNP 253 C/T Arginine/Tryptophan

SNP 253, which caused an arginine to tryptopan amino acid change, is displayed in chromatogram 1. This SNP caused a C to T base change in the ACG nucleotide triplet coding for the aarginine amino acid. Chromatogram 2 and 3, in the figure below, shows the homozygote T/T wild type allele.



1. Heterozygote C/T

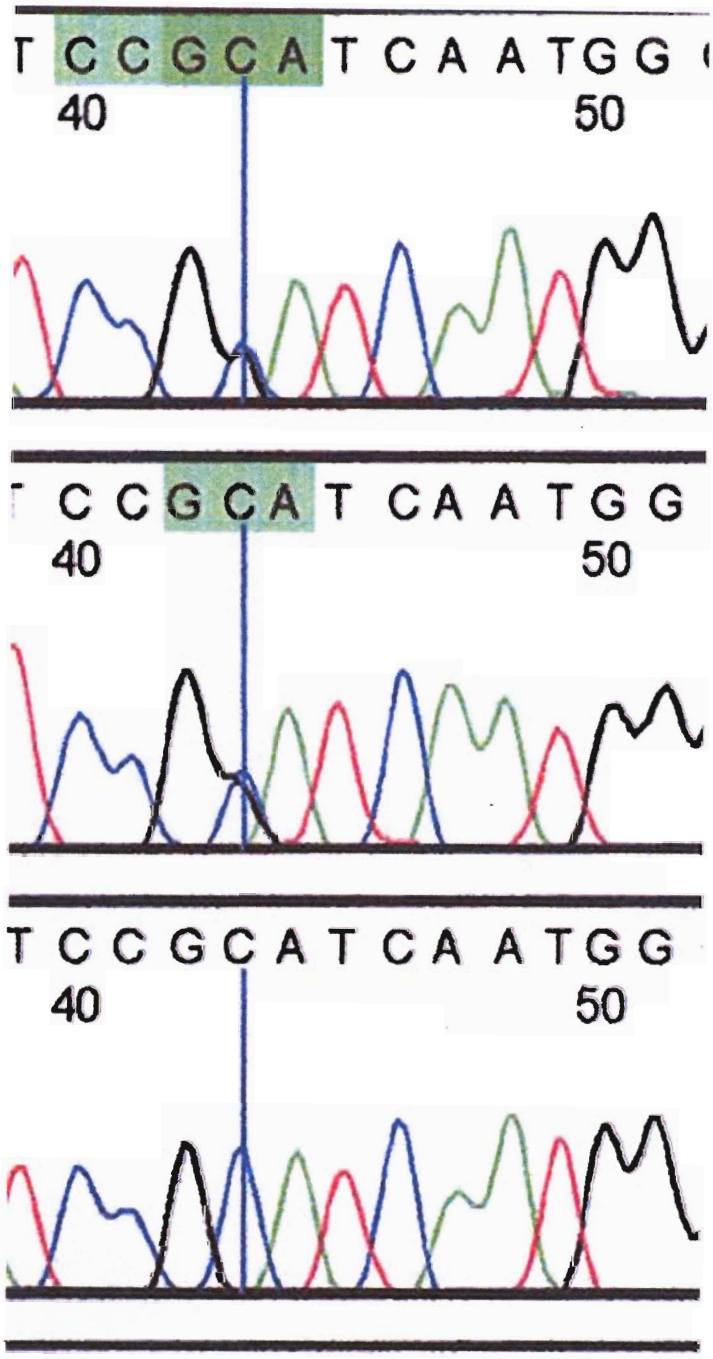
2. Homozygote C/C

3. Homozygote C/C

Fig 20.

Fig 21. SNP 423 C/G Arginine/Arginine

Chromatograms 1 and 2 in the figure below show SNP 423, which caused a C to G base change in the CGC codon. The wild type homozygote C/C allele is displayed in the chromatogram 3. The heterozygote C to G base change did not change the arginine amino acid coded by the CGC nucleotide triplet.



1. Heterozygote C/G

2. Heterozygote C/G

3. Homozygote C/C

Fig. 21

Fig 22. SNP 1334 C/T Valine/Alanine

SNP 1334, which caused a valine to alanine amino acid change is displayed in chromatogram 1. This SNP caused a C to T base change in the CTG nucleotide triplet coding for the valine amino acid. Chromatogram 2, in the figure below, shows the homozygote T/T wild type allele and chromatogram 1 shows the homozygote C/C allele.

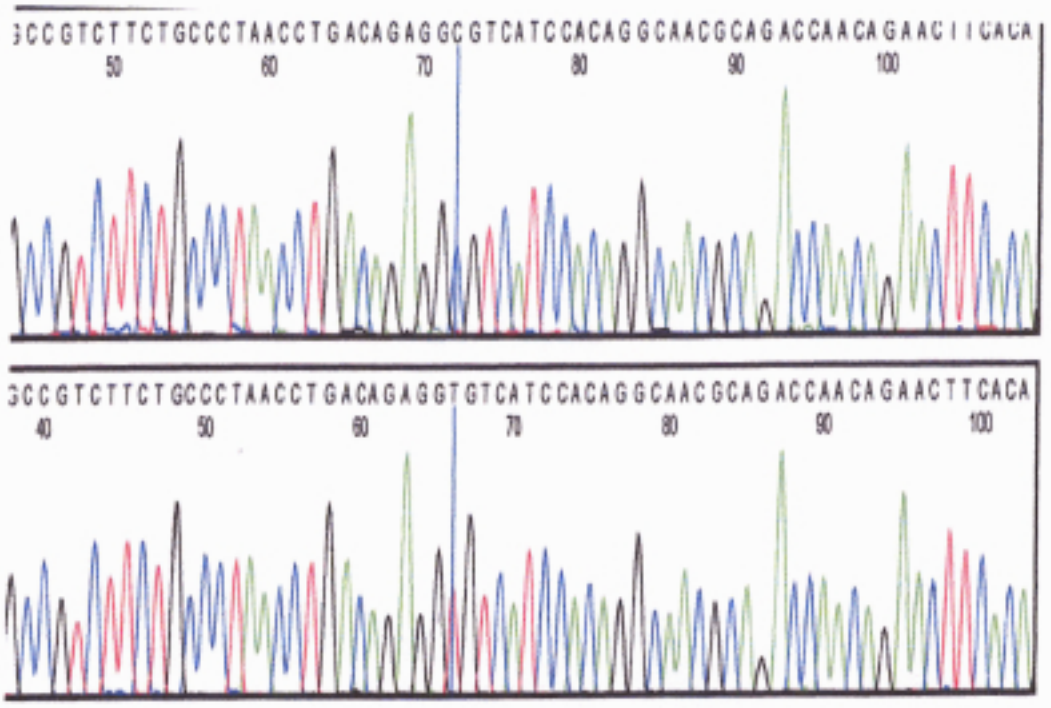

50
60
70
80
90
100
1. Homozygote C/C
40
50
60
70
80
90
100
2. Homozygote T/T

Fig. 22

Fig 23. SNP 1781 A/G

SNP 1781 in figure 23 resulted n a homozygote A to G base change. The heterozygote A/A alleles are displayed in chromatograms 1 and 4, and the wild type A/A alleles are displayed in chromatograms 2 and 3. Because this SNP is located downstream from the stop codon, no amino acids are translated by the ribosomal RNAs.

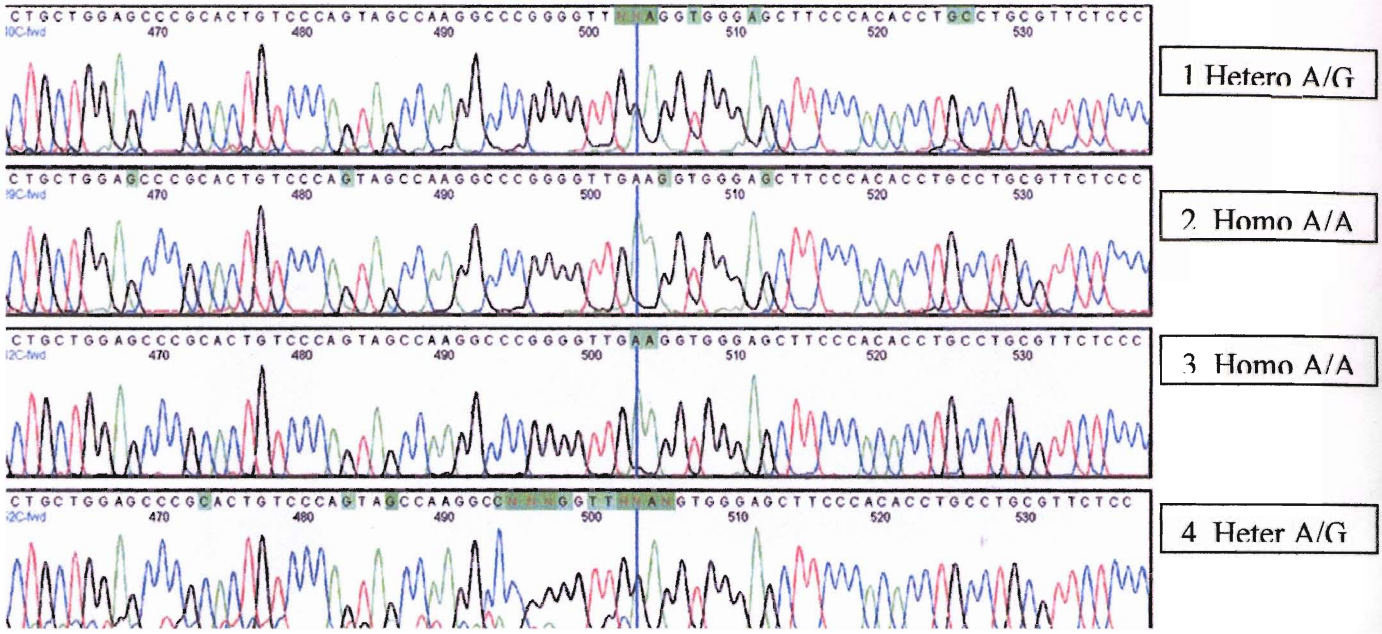


Fig. 23

Table 1.

Detected Human *fpgs* SNPs

nt	**SNP**	**Codon Change**	**Amino Acid Change**
64	A/G	**A**TA > **G**TA	**Isoleucine> Valine**
123	G/A	CC**G** > CC**A**	Proline (no change)
253	C/T	A**C**G > A**C**G	**Arginine > Tryptophan**
423	C/G	CG**C** > CG**G**	Arginine (no change)
1350	C/G	AA**C** > AA**G**	**Asparagine > Lysine**
1781	G/A	GGA > G**A**A	---

Table 2.

Human *fpgs* SNPs of the NCBI database

Entry number	Sequence	Intron/Exon	Codon	Comment
rs28570299	TAAG[A/T]TGCA	Not found on the FPGS genomic locus by Blast search		
rs17855900	AGTG[C/T]CCCC	Exon 16	489 V>A	Random sequence
rs17855899	CGGC[A/G]TAAC	Exon 1	22 I>V	nt 64
rs12686275	GCCG[A/T]CTTC	Exon 15	437 V>D	Checked in 209 samples, never found. Sequencing error
rs12005665	ggac[A/G]tggt	Intron		
rs12004468	gtct[C/G]gttc	Intron 15		
rs117900075	actg[C/T]actc	Intron		
rs11554717	GCGG[A/C]GTAA	Exon 1	21 G>G	Random sequence
rs11414149	CTAC[-/A]AAAA	Intron 14		
rs10987745	ATAG[C/T]TGGG	Intron 16		
rs10987744	ACAT[C/T]TCAG	Intron 10		
rs10987743	AATA[A/G]GGTG	Intron 5		
rs10987742	TTGC[C/T]CTAG	Intron 1		
rs10819307	gcac[A/T]tcag	Intron 5		
rs10760502	CGGC[A/G]TAAC	Exon 1	22 I>V	nt 64 (again)
rs10732395	aatt[A/G]ttag	Intron		

rs10671024	tctc[-/AA]aaaa	Intron 15		
rs10122348	ggcc[A/G]aggca	Intron 5		
rs10122345	ggga[A/G]gccg	Intron 5		
rs10122777	ggcc[A/G]aagc	Intron 5		
rs10120483	CCAG[A/G]AGTG	Intron 1		
rs10118903 gene by Blast	TGGT[C/T]GGTG	?	Not found on the FPGS analysis	
rs9802255	TTTT[G/T]GGTT	Intron 12		
rs9632913	gtct[C/T]taca	Intron 15		
rs7868117	gtgt[G/T]tttt	Intron 15		
rs7856096	ACCT[A/G]CTAT	Intron 1	Confirmed in 209 individuals. AA/AG/GG = .885/.091/.024	
rs7854264	tgtg[G/T]gttt	Intron 15		
rs7847684	cact[C/T]gaac	Intron 9		
rs7027767	GGGC[A/G]GGGC	Intron 13		
rs5900768	tttt[-/T]gaga	Intron 5		
rs4451422	ATAC[A/C]CATT	Exon 15	3'-UTR	Confirmed in 354 samples. AA/AC/CC = .065/.234/.333 The remaining .367allelic frequencies was unspecified (see NCBI Database)

rs2230270	GCTG[A/G]CCCT	Exon 11	294 T>A	Confirmed. 23 samples AA/AG = .957/.043
rs2183471	AAGG[C/T]TGGG	Intron 2		
rs1966227	gttc[A/G]agaa	Intron 5		
rs1042320	CCTT[C/T]TGGG	exon 15	3'-UTR nt 2139	
rs1042304	GATA[A/G]CAGA	Intron		Checked in 204 samples. No SNP found at this nt. Error in reference sequence
rs955171	ATCA[G/T]GGGT	Intron 1		Checked in four studies, three had 100 % CCs in all 30 samples; the other had 100 % GG in all 47 samples
rs10106	GCCA[A/G]GAGG	Exon 15	3'-UTR.	Checked in 421 individuals. AA/AG/GG = .221/ .480/ .299

DISCUSSION

Previous studies by Barredo, J et al showed that while substantial levels of the *fpgs* gene was expressed in lymphoblasts from patients with acute lymphoblastic leukemia, no expression of this gene was detected in circulating granulocytes, erythrocytes, and lymphocytes (**23**). In this experiment, however, we were able to successfully detect RNA expression in monocytes and lymphocytes. The discrepancy in both these results has much to do with the difference in techniques being used to detect *fpgs* expression. While Barredo, J et al used a protein enzyme assay technique to detect FPGS activity, in this experiment we first extracted the RNA from the mononuclear cells then used Polymerase Chain Reaction to amplify the gene. PCR is a much more sensitive technique than an enzyme assay technique for detecting RNA expression. With as little as 1 μg of DNA, we were able successfully amplify and sequence 85 % of the *fpgs* gene in the mononuclear cells of the human peripheral blood. Barredo, J et al did not have the benefit of using the PCR technique for detecting *fpgs* RNA expression in mononuclear cells because the sequence of the *fpgs* gene, which was required for designing primers, was not yet known. With the advantages of PCR and the availability of the *fpgs* gene sequence we were able to detect and confirm, by repeated testing, six single nucleotide polymorphisms expressed in the fpgs gene, of the peripheral blood, of 85 normal human subjects. These single nucleotide polymorphisms are located in a variety of different regions of the *fpgs* gene. One SNP was detected in the mitochondrial leader sequencing, another was detected downstream of the stop codon, while the other four SNPs were detected within the cytosolic coding sequence of the protein. The only frequently observed SNPs in the

coding region were in the mitochondrial leader sequence. SNPs downstream of the cytosolic start methionine were rare.

The first SNP (A/G) at nucleotide 64 caused a base change in the first position of the codon, resulting in a valine to isoleucine amino acid change. This SNP, which was previously reported by Freemantle et al. was detected in a majority of the samples (49/88). The valine to isoleucine amino acid change, caused by a conservative point mutation, is expected to have little to no effect on the structure and function of the protein. Furthermore, this SNP is located in a region upstream of the cytosolic start site, and so is not expected to significantly affect catalytic function, although it might well affect the efficiency of mitochondrial entry.

The second most frequently detected SNP, which was expressed in the third position of the codon at nucleotide 123, did not change the proline amino acid at that position of the polypeptide. Six of the samples expressed the G/A heterozygote allele at nucleotide 123, while only one sample expressed the homozygote A/A allele. Because this SNP did not change the amino acid, no change in protein structure and function would be expected. Similarly, SNP 423 also caused a base change in the third position of the codon, which did not change the arginine amino acid located at that position in the growing polypeptide. The remaining three SNPs, however, alter the codon sequence, which resulted in a nonconservative change in the amino acids. One of these SNPs, C253T, caused a base change at nucleotide 253 in the first position of the codon, replacing an arginine amino acid with tryptophan, a more bulky amino acid. This amino acid conversion, which was detected in one of the 88 samples, might affect the chemical nature of the enzyme, and would presumably alter the enzymatic activity of the fpgs

protein. Similarly, SNP T1334C, which caused a base change at the second position of the codon, resulted in a valine to alanine amino acid change in the polypeptide chain. The final SNP was detected at position 1781 (G/A) at the second position of the nucleotide triplet, falling outside of the reading frame of the fpgs gene. Normally amino acids translated outside of the reading frame of the gene do not play a role in protein function, however, SNP 1781 might affect the stability of the mRNA or the efficiency of polyadenylation of the message.

There have been limited studies done on polymorphisms in the human *fpgs* gene, and very little is known about the effects of any polymorphisms in the gene that may alter clinical response to antifolate therapy. The next major step of this project it to analyze the effects of polymorphisms detected in the fpgs gene on response rate to cancer chemotherapy. Cells, which are mutant for the fpgs gene, will be transfected with fpgs gene construct expressing known polymorphic sites. These cells with then be observed for change in enzymatic activity of the gene.

List of References

List of References

1. Kwok, P.Y, Gu, Z. Single nucleotide polymorphism libraries: why and how are we building them? *Mol Med Today* **12:** 538-43 (1999).

2. Ewing, B and Green, P. Analysis of expressed sequence tags indicates 35,000 human genes. *Nature Genetics* **25**, 232-234 (2000).

3. Pennisi, E. The Human Genome. *Science* **16:** 1177-1180 (2001)

4. Freemantly, S.J and Morgan R.G. Transcription of the Human Folylpoly-gamma-glutamate Synthetase *Gene. J. Biol. Chemistry* **40:** 25373-25379 (1997).

5. Morgan R.G. Roles of folylpoly-gamma-glutamate synthetase in therapeutics with tetrahydrofolate antimetabolites: an overview. *Semin Oncol.* **26:** 24-32 (1999).

6. Sharp, L, Little, J. Polymorphisms in Genes Involved in Folate Metabolism and Colorectal Neoplasia: A HuGE Review. *American Journal of Epidemiology* **159:** 423-443 (2004).

7. Frost P, Blom H.J, Milos R, et al. A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nat Genet* **10:** 111–113 (1995).

8. Botto L.D, Yang Q. 5,10-Methylenetetrahydrofolate reductase gene variants and congenital anomalies: a HuGE review. ***Am J Epidemiol*** **151:** 862–877 (2000).

9. Van der Put J, Van der Molen E.F, Kluijtmans L, et al. Sequence analysis of the coding region of human methionine synthase: relevance to hyperhomocysteinaemia in neural-tube defects and vascular disease. *QJM* **90:** 511–517 (1997).

10. Ma J, Stampfer M.J, Christensen B, et al. A polymorphism of the methionine synthase gene: association with plasma folate, vitamin B_{12}, homocyst(e)ine, and colorectal cancer risk. *Cancer Epidemiol Biomarkers Prev* **8:** 825–829 (1999).

11. Ulrich C, Kampman E, Bigler J, et al. Colorectal adenomas and the C677T MTHFR polymorphism: evidence for gene-environment interaction? *Cancer Epidemiol Biomarkers Prev* **8:** 659–68 (1999).

12. Ma J, Stampfer MJ, Christensen B, et al. A polymorphism of the methionine synthase gene: association with plasma folate, vitamin B_{12}, homocyst(e)ine, and colorectal cancer risk. *Cancer Epidemiol Biomarkers Prev* **8:** 825–9 (1999).

13. Weitman S, Weinberg A, Coney L, Zurawski R, Jennings S and Kamen A. Cellular localization of the folate receptor: potential role in drug toxicity and folate homeostasis. *Cancer Research* **52:** 6708-6711 (1992).

14. Anthony, A.C, Kane M.A, Portillo R.M, Elwood P.C, Kolhouse J.F. Studies of the role of a particulate folate-binding protein in the uptake of 5-methyltetrahydrofolate by cultured human KB cells. *J. Biol. Chemistry* **260:** 14911-14917 (1985).

15. Costi, M.P and Ferrari, S. Update on Antifolate Drug Targets. *Current Drug Targets* **2:** 135-166 (2001)

16. Bolin J.T, Filman D.J, Matthews D.A, Hamlin R.C and Kraut, J. Crystal structures of Escherichia coli and Lactobacillus casei dihydrofolate reductase refined at 1.7 A resolution. I. General features and binding of methotrexate. *J. Biol. Chem.* **257:** 13650-13662 (1982).

17. Farrugia, D, Ford, H, Cunningham, D,. Danenberg, K, Danenberg, P, Brabender, J, McVicar, D, Aherne, W, Hardcastle, A, McCarthy, K and Jackman A, Thymidylate synthase expression in advanced colorectal cancer predicts for response to raltitrexed.*Clin Cancer Res.* **9:** 792-801 (2003).

18. Taylor, E, Kuhnt, D, Shih, C, Rinzel, S, Grindey, G, Barredo, J, Jannatipour, M and Moran, R. A dideazatetrahydrofolate analogue lacking a chiral center at C-6, N-[4-[2-(2-amino-3,4-dihydro-4-oxo-7H-pyrrolo[2,3-d]pyrimidin-5- yl)ethyl]benzoyl]-L-glutamic acid, is an inhibitor of thymidylate synthase. ***J. Med. Chem.*35:** 4450-4454 (1992).

19. Zhao, R, Titus, S, Gao, F, Moran, R and Goldman, D. Molecular analysis of murine leukemia cell lines resistant to 5, 10-dideazatetrahydrofolate identifies several amino acids critical to the function of folylpolyglutamate synthetase. *Gene. J. Biol. Chemistry* **275:** 26599-26602 (2000).

20. Schuler, G.D., Boguski, M.S., Stewart, E.A., Stein, L.D., Gyapay, G., Rice, K., White, R.E., Rodriguez-Tome, P., Aggarwal, A., Bajorek, E., et al. 1996. A gene map of the human genome. *Science* **274:** 540–546.

21. Taylor S.M, Freemantle S.J, and Moran R,G. Structural organization of the human folypoly-gamma-glutamate synthetase gene: evidence for a single genomic locus. ***Cancer Research* 55:** 6030-6034 (1995).

22. Sun, X, Bognar, AL. Baker, Edward N, and Smith C.A. Structural Homologies with ATP-binding and Folate-binding enzymes in the structure of Folylpolyglutamate Synthetase. *Biochemistry* **95**, 6647-6652 (1998).

23. Barredo J, Moran RG. Determinants of antifolate cytotoxicity: folylpolyglutamate synthetase activity during cellular proliferation and development. *Mol Pharmacol.* **42** 687-94 (1992).