

A Common Variant in *MTHFD1L* is Associated With Neural Tube Defects and mRNA Splicing Efficiency

Anne Parle-McDermott,^{1*} Faith Pangilinan,² Kirsty K. O'Brien,¹ James L. Mills,³ Alan M. Magee,¹ James Troendle,³ Marie Sutton,⁴ John M. Scott,⁵ Peadar N. Kirke,⁴ Anne M. Molloy,⁶ and Lawrence C. Brody²

¹Nutritional Genomics Group, School of Biotechnology, Dublin City University, Dublin 9, Ireland; ²Molecular Pathogenesis Section, Genome Technology Branch, National Human Genome Research Institute, Bethesda, Maryland; ³Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, Maryland; ⁴Child Health Epidemiology Unit, Health Research Board, Dublin, Ireland; ⁵School of Biochemistry & Immunology, Trinity College Dublin, Dublin 2, Ireland; ⁶School of Medicine, Trinity College Dublin, Dublin 2, Ireland

Communicated by Stephen Chanock

Received 15 May 2009; accepted revised manuscript 4 August 2009.

Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/humu.21109

ABSTRACT: Polymorphisms in folate-related genes have emerged as important risk factors in a range of diseases including neural tube defects (NTDs), cancer, and coronary artery disease (CAD). Having previously identified a polymorphism within the cytoplasmic folate enzyme, *MTHFD1*, as a maternal risk factor for NTDs, we considered the more recently identified mitochondrial paralogue, *MTHFD1L*, as a candidate gene for NTD association. We identified a common deletion/insertion polymorphism, rs3832406, c.781-6823ATT(7–9), which influences splicing efficiency and is strongly associated with NTD risk. Three alleles of rs3832406 were detected in the Irish population with varying numbers of ATT repeats: Allele 1 consists of ATT₇, whereas Alleles 2 and 3 consist of ATT₈ and ATT₉, respectively. Allele 2 of this triallelic polymorphism showed a decreased case risk as demonstrated by case-control logistic regression ($P = 0.002$) and by transmission disequilibrium test (TDT) ($P = 0.001$), whereas Allele 1 showed an increased case risk. Allele 3 showed no influence on NTD risk and represents the lowest frequency allele (0.15). Additional single nucleotide polymorphism (SNP) genotyping in the same genomic region provides additional supportive evidence of an association. We demonstrate that two of the three alleles of rs3832406 are functionally different and influence the splicing efficiency of the alternate *MTHFD1L* mRNA transcripts. *Hum Mutat* 30:1–7, 2009. © 2009 Wiley-Liss, Inc.

KEY WORDS: *MTHFD1L*; NTD; splicing; polymorphism; association; folate; mitochondria

Introduction

Low maternal folate levels and common genetic variation are recognized as important risk factors for a group of common birth malformations known as neural tube defects (NTDs) [Kirke and

Scott, 2005]. The role of folate in NTD causation was definitively demonstrated by a number of intervention trials showing up to 70% reduction in NTD-affected pregnancies when women ingest a folic acid supplement in the periconceptional period [Czeizel and Dudas, 1992; MRC Vitamin Study Research Group, 1991]. This pointed toward genes involved in the transport and metabolism of folate as prime candidates for contributing to the genetic risk of NTDs. However, only two such genes have been demonstrated so far as playing a definitive role in NTD-association in the Irish population. These genes encode two enzymes involved in the cytoplasmic metabolism of folate and are known as Methyl-tetrahydrofolate Reductase (*MTHFR*; MIM# 607093) and C₁-Tetrahydrofolate Synthase (*MTHFD1*; MIM# 172460), and have been confirmed in replicate studies [Botto and Yang, 2000; Brody et al., 2002; Parle-McDermott et al., 2006]. We considered another folate metabolic gene, *MTHFD1L* (MIM# 611427), as a candidate for association with NTDs. Although the *MTHFD1* gene encodes the cytoplasmic C₁-Tetrahydrofolate Synthase enzyme and possesses three enzymatic activities, the *MTHFD1L* gene encodes the mitochondrial localized C₁-Tetrahydrofolate Synthase enzyme and appears to be monofunctional. The mere existence of a mitochondrial C₁-Synthase was controversial for many years until its existence was clearly demonstrated by recent publications [Christensen et al., 2005; Prassannan et al., 2003; Walkup and Appling, 2005]. The strong association of the cytoplasmic C₁-Synthase (*MTHFD1*) with NTDs prompted us to consider this newly discovered mitochondrial localized enzyme, *MTHFD1L*, as an important candidate for NTD association.

Our initial approach was to identify candidate polymorphisms within *MTHFD1L* with a likely functional effect. We identified an intronic deletion/insertion polymorphism (DIP) within intron 7 (rs3832406, c.781-6823ATT[7–9]), which is in close proximity to an alternatively spliced exon (Fig. 1A). *MTHFD1L* produces two transcripts of 3.6 and 1.1 kb; the shorter transcript includes an alternative exon 8a that generates an mRNA containing a premature stop codon [Prassannan et al., 2003], resulting in a protein that lacks synthetase activity (Fig. 1B). We hypothesized that rs3832406 alters splicing efficiency as it occurs within a polypyrimidine tract (PPT) that forms part of the “splicing code” [Wang and Cooper, 2007]. Recent research has recognized the important role of common genetic variants on gene expression patterns [Morley et al., 2004] and alternative splicing efficiency [Hull et al., 2007]. The significance of an impact on splicing efficiency in relation to *MTHFD1L* is an increased or decreased amount of functional folate enzyme depending on genotype.

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Anne Parle-McDermott, Nutritional Genomics Group, School of Biotechnology, Dublin City University, Dublin 9, Ireland.

E-mail: anne.parle-mcdermott@dcu.ie

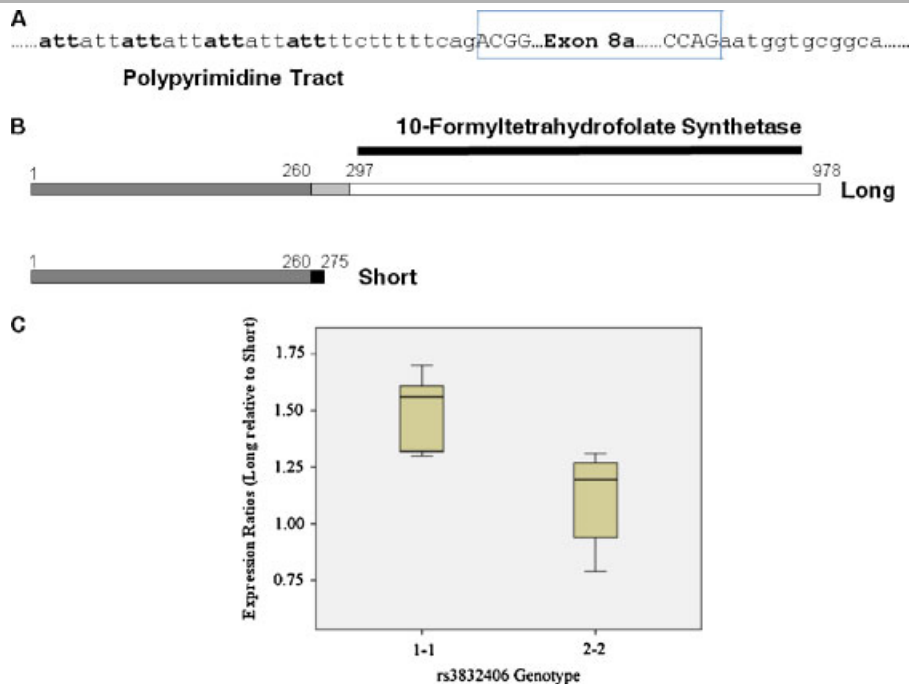


Figure 1. Analysis of the NTD associated rs3832406 and alternative splicing. **A:** Sequence context of the intron 7 DIP within the polypyrimidine tract adjacent to alternative exon 8a. Allele 1, that is, (ATT)₇ repeats are illustrated here. Exon 8a is illustrated as the boxed area as defined by Prassannan et al. [2003]. **B:** The potential protein products resulting from translation of the *MTHFD1L* long and short mRNAs. The protein products are identical up to amino acid residue 260. The longer transcript produces a monofunctional mitochondrial enzyme with 10-formyltetrahydrofolate synthetase activity localized to the C-terminal end as illustrated by the black bar. The shorter transcript differs at the extreme C-terminus due to inclusion of alternative exon 8a. The shorter protein lacks a synthetase domain. Based on a figure from Prassannan et al. [2003]. **C:** The ratio of *MTHFD1L* long and short transcripts differs by rs3832406 genotype in Coriell[®] lymphoblast cell lines. Homozygotes for Allele 1 have a significantly greater proportion (mean is 35% more) of the long transcript relative to the short transcript compared to homozygotes for Allele 2 as assessed by Mann-Whitney *U* test (two tailed; *P* = 0.006).

A polymorphism with such a functional impact may also contribute to NTD risk. We addressed this hypothesis by examining the relative ratio of the long-to-short transcript by quantitative reverse transcription-polymerase chain reaction (RT-PCR) in cell lines of known rs3832406 genotype. NTD association was assessed by case-control and triad family-based analyses by genotyping rs3832406 plus 118 single nucleotide polymorphism (SNP) markers spanning the entire *MTHFD1L* gene. Our data confirms our hypothesis, that is, the rs3832406 impacts on alternative splicing efficiency of the *MTHFD1L* mRNA transcripts and increases case risk of NTDs.

Materials and Methods

Quantitative RT-PCR

Quantitative RT-PCR (Q RT-PCR) was performed on RNA extracted from the following Coriell[®] lymphoblast cell lines (Coriell Institute for Medical Research, Camden, NJ, USA): NA17124, NA17142, NA17146, NA17147, NA17158, NA17165, NA17201, NA17214, NA17218, NA17229, NA17246, and NA17291. These lines represent six homozygotes of Allele 1 or Allele 2 for rs3832406 and each genotype group had an equal number of African American or Caucasian individuals. The cell lines were cultured in RPMI1640 with 10% fetal bovine serum and 1% penicillin/streptomycin (10,000 U:10 mg/ml) at 37°C at 5% CO₂. All lines were routinely tested for Mycoplasma utilizing a

PCR-based assay and were found to be uncontaminated. RNA was extracted using Qiagen RNeasy kit (Cat. No. 74104, UK) and Qiagen QiaShredders (Cat. No. 79654, UK). DNase 1 treatment to remove potential contaminating genomic DNA was carried out by “on-column” treatment as described in the Qiagen manual. RNA quality was verified by measurement of A260/A280 nm ratios using a Nanodropper and by resolution on a 1% agarose gel. RNA was synthesized to cDNA using Superscript II (Invitrogen, UK) and a combination of oligodT and random hexamers according to the manufacturer’s instructions. Q RT-PCR was performed on the Roche Lightcycler[®] 480 instrument using assays designed to specifically detect the long or short transcript of *MTHFD1L*. Assays were designed utilizing the ProbeFinder software and the Universal Probe library as follows: *MTHFD1L* long: forward 5’ GAGCTCTGAAGARGCATGGAG 3’ reverse 5’ TGCTTCTGGA-GGTTACAGCA 3’ universal probe #42; *MTHFD1L* short: forward 5’ ACGCCAGCTTCAAAGCAA 3’ reverse 5’ TCACAGGAGAAT-CACTTCAACC 3’ and universal probe #13. The PCR efficiencies for each assay was assessed using the pooled sample of the experimental cDNAs and was 1.83 for the short assay and 2.0 for the long assay. All assays were intron spanning and were performed using the Probes Master Mix (Roche, UK) as recommended by the manufacturer. The assays were carried out in duplicate incorporating “minus Superscript” and PCR negative controls and replicated several times on separate RNA extractions. The relative ratio of the *MTHFD1L* long transcript relative to the short transcript was generated by the Roche Lightcycler[®] 480 Relative Quantification software

employing the E (Efficiency)-Method. The E-Method compensates for differences in PCR efficiency of the target and reference genes, in particular sample sets, and is thought to provide a more accurate estimate of relative quantitative data than the $\Delta\Delta C_T$ method.

Study Population

Families affected by an NTD pregnancy were recruited throughout the Republic of Ireland with the assistance of the Irish Association for Spina Bifida and Hydrocephalus (IASBAH) and the Irish Public Health Nurses from 1993 to 2005. These families, consisting of both complete and incomplete triads (case, mother, father), formed the NTD cohort. Control samples for NTD association were selected from a population of 56,049 pregnant women attending the three main maternity hospitals in the Dublin area between 1986 and 1990. These women had no history of an NTD-affected pregnancy. Details of these cohorts have been previously published [Brody et al., 2002]. In addition, genotype and blood metabolite data were available from a cohort of 2,524 healthy, ethnically Irish individuals consisting of university students aged between 18 and 25 years old and recruited over a period of one academic year (TSS cohort). Informed consent and ethical approval were obtained for all human samples used in this study.

Genotyping

Genomic DNA was extracted from all samples using a Qiagen QIAamp DNA Blood Mini Kit or a Qiagen DNeasy Kit (Qiagen, UK). The rs3832406 polymorphism (c.781-6823ATT[7-9]) was genotyped by PCR amplification using fluorescently labeled primers that flanked the DIP (forward primer: 5' 6-FAM TTCTCTTCTTAGCCCCACG 3'; reverse primer: 5' AGAGCT-TGCAGTGAGCCTAGA 3') and products were resolved and scored on an ABI 377 or 3100 using Genescan 3.1.2 software. For NTD association the rs3832406 DIP was genotyped in 860 controls, and the following samples from the NTD cohort: complete triads $n = 439$, mother and case only $n = 55$; mother and father only $n = 34$; case only $n = 42$; mother only $n = 166$; father only $n = 2$. Additional rs3832406 genotyping included all 2,524 samples from the TSS cohort and a panel of Coriell[®] lymphoblast cell lines isolated from African American and Caucasian individuals. Quality control for rs3832406 genotyping consisted of repeat genotyping of at least 10% of samples by the same assay (99% agreement), repeat genotyping of an additional 10% of samples by a different assay (94% agreement) and a genotyping success rate of >95%. Discrepant genotype calls were resolved by re-genotyping or were left out of the final analysis. SNPs within *MTHFD1L* were genotyped within a 1,536 custom Illumina[®] Goldengate assay (see Supp. Table S1) on a subset of the NTD cohort and controls as follows: 277 complete triads and 340 controls. The call rate for the Illumina[®] Goldengate assay was 98% with a blind duplicate concordance rate of 99.99%.

Haplotype Analysis

A linkage disequilibrium plot of *MTHFD1L* was generated using Haploview (<http://www.broad.mit.edu/mpg/haploview/>) [Barrett et al., 2005]. Haplotype blocks for genomic region spanning introns 7–10 were defined using the Solid Spine of linkage disequilibrium (LD) and included the following SNPs: rs803422, rs17080461, rs3832406, rs2295083, rs712210, rs6905272, rs71208,

rs12195069, rs17080476, rs1771845, rs175862, rs9397365, rs2295084, rs175853, and rs803456. Haplotype frequencies of the blocks were estimated using PHASE 2.1.1 [Stephens and Donnelly, 2003; Stephens et al., 2001]. Additionally, PHASE 2.1.1 was used to compare haplotype frequency distributions in controls and NTD cases or NTD mothers using a permutation test.

Metabolite Measurements

Nonfasting blood samples were processed within 2 hours of collection. Full blood count data including mean corpuscular volume (MCV) were determined on fresh EDTA blood samples using a Sysmex F-800 Microcell counter. Serum and red cell hemolysates in 1% ascorbic acid were stored at -40°C until analyzed for total folate by microbiological assay [Molloy and Scott, 1997]. Plasma total homocysteine (tHcy) was determined by immunofluorescence using an Abbott IMX instrument [Leino, 1999].

Statistical Analyses

Genotype frequencies of rs3832406 were compared between each sample group, that is, mother, father, or cases versus control by a χ^2 test. An effect of each genotype was compared to the combined frequency of the other genotypes for each comparison (data not shown). Associations with an NTD were tested in cases/controls and separately in mothers/controls by logistic regression with a continuous term indicating the number of alleles of a given type. The transmission of alleles from parents to affected NTD cases was assessed by Transmission Disequilibrium Test (TDT) using SAS PROC GENMOD. Case and maternal effects were also assessed by a two degree of freedom log-linear model using SAS PROC GENMOD [Weinberg et al., 1998; Wilcox et al., 1998].

Q RT-PCR relative ratios (as described above) were tested for statistical significance by a Mann-Whitney *U*-test using SPSS 15.0 for Windows by stratifying the fold change by genotype. Values below $P = 0.05$ were considered significant for all analyses.

Results

The *MTHFD1L* rs3832406 DIP is Situated within a Polypyrimidine Tract and Significantly Correlates with Altered Splicing Efficiency

Initial genotyping of rs3832406 in control samples from the Irish population identified a total of three alleles that differ in the length of a repeated “ATT” sequence within the polypyrimidine tract of the alternatively spliced exon 8a of *MTHFD1L* (Fig. 1). The “ATT” sequence occurs as 7 (ATT₇), 8 (ATT₈), or 9 (ATT₉) repeats, which are referred to as Alleles 1, 2, or 3, respectively. The ratio of the long to short transcript of *MTHFD1L* was assessed by Quantitative Reverse Transcription (RT)-PCR in a panel of Coriell[®] cell lines that were homozygous for the most common alleles, that is, Alleles 1 or 2. The ratio of the *MTHFD1L* long transcript relative to the short transcript was generated by the Roche Lightcycler[®] 480 Relative Quantification software employing the E (Efficiency)-Method. Q RT-PCR relative ratios were tested for statistical significance by a Mann-Whitney *U* test using SPSS 15.0 for Windows by stratifying the fold change by genotype. The result of this analysis (Fig. 1c; Supp. Table S2) showed that the genotype of rs3832406 is associated with splicing efficiency. Allele 1 is associated with having an approximately 1.4-fold higher proportion of the long transcript relative to the short transcript compared to Allele 2 ($P = 0.006$). Our panel of Coriell[®] cell lines

included a single line that was homozygous for Allele 3, and thus, could not be included in the final data analysis.

The *MTHFD1L* rs3832406 DIP is Associated with NTDs

Our association study consisted of NTD triads (mother, father, and affected case) and controls from the Irish population. Not all triad families were complete, that is, samples from all three family members were not always available. The rs3832406 DIP was genotyped in a total of 1,705 samples from both complete and incomplete NTD triads and 860 controls. Associations with an NTD were tested in cases/controls and separately in mothers/controls by logistic regression with a continuous term indicating the number of alleles of a given type. The transmission of alleles from parents to affected NTD cases was assessed by TDT using SAS PROC GENMOD. Case and maternal effects were also assessed by a two degree of freedom log-linear model using SAS PROC GENMOD [Weinberg et al., 1998; Wilcox et al., 1998]. The rs3832406 polymorphism showed strong evidence for a case association. As described above, this polymorphism is a repeated “ATT” sequence that has three common alleles 7 (ATT₇), 8 (ATT₈), or 9 (ATT₉) referred to as Alleles 1, 2, and 3, respectively. Case-control logistic regression and TDT analysis (Table 1) revealed that carriers of Allele 1 are associated with an increased

risk of having an NTD (TDT, $P = 0.016$), while Allele 2 carriers appear to have a decreased risk (TDT, $P = 0.001$). A 2 degree of freedom log-linear analysis confirmed these case associations (Table 1). Allele 3 showed no evidence of an association and has the lowest frequency. We also tested whether rs3832406 correlated with circulating folate or homocysteine in 2,524 healthy students and found no evidence of an association (data not shown).

SNP Genotyping Data Supports the *MTHFD1L* rs3832406 DIP as a Novel Risk Factor for NTDs

We genotyped 118 SNPs by Illumina[®] Goldengate assay to ensure appropriate coverage of the *MTHFD1L* gene (Supp. Table S1). A total of 277 complete NTD triad families (831 samples) and 340 controls were genotyped. Data analysis revealed several case and maternal association signals from three SNP clusters within *MTHFD1L* (Fig. 2). These SNP clusters appear to be separated by recombination hotspots based on Phase II HapMap data as reported previously [Samani et al., 2007]. This suggests that polymorphisms from three separate regions of the *MTHFD1L* gene have independent associations with NTDs. We focused our attention on the cluster of association signals spanning genomic region intron 7 to 10, which is marked as region “a” in Figure 2 and is shown in more detail in Figure 3. The three allele rs3832406 data were collapsed into a two allele format to incorporate the DIP into the LD map (Figs. 2 and 3). The linkage disequilibrium plot of *MTHFD1L* was generated using Haploview [Barrett et al., 2005] and the predicted haplotypes for this region (introns 7–10; Fig. 3) are described in Table 2. Haplotype blocks were defined using the Solid Spine of LD. Haplotype frequencies of these blocks were estimated using PHASE 2.1.1 [Stephens and Donnelly, 2003; Stephens et al., 2001]. There are six additional SNP markers in this region showing NTD disease association; all are intronic and show significant association for case-control or mother-control by logistic regression or log linear analysis. The case-control-associated SNPs are as follows: rs17080461 (logistic regression [LR] odds ratio [OR] 1.16, $P = 0.05$, minor allele frequency (MAF) = 0.126), rs2295083 (LR OR 1.45, $P = 0.02$, MAF = 0.15), rs712208 (log linear (LL) relative risk (RR) 4.3, $P = 0.05$, MAF = 0.20), rs17080476 (LL RR 0.73, $P = 0.002$, MAF = 0.17). The mother-control-associated SNPs are as follows: rs712210 (LL RR 0.55, $P = 0.05$, MAF = 0.50), rs175853 (LL RR 0.44, $P = 0.04$, MAF = 0.33). SNP rs17080476:A>G showed the most significant association for an NTD case effect as revealed by case-control logistic regression ($P = 0.009$), by TDT ($P = 0.05$) and by log linear analysis ($P = 0.002$). The D' values between rs3832406 and rs17080476 is 0.61 with an r^2 of 0.13. SNP rs712208 had the highest case relative risk of 4.3. The D' values between rs3832406 and rs712208 is 0.902 with an r^2 of 0.39. The D' values between rs3832406 and the other disease-associated SNPs from this region range from 0.56 to 1, and none of the polymorphisms share strong r^2 values ($r^2 \leq 0.43$). This provides supporting evidence that the genomic region of *MTHFD1L* incorporating introns 7 to 10 harbors a disease-associated polymorphism. Whether the disease-causing variant is acting independently or in combination with a haplotype is difficult to determine; however, our haplotype analyses did not identify an association with a specific haplotype (Table 2). All polymorphisms showing a significant association with NTD risk from this genomic region are intronic. Our evidence to date supports rs3832406 as the most plausible variant within this region to contribute to disease causation as we proposed and

Table 1. Analysis of Genotype and Allele Frequencies of rs3832406 in NTD Groups and Controls

Genotypes	Cases	Fathers	Mothers	Controls
1-1	254 (.48)	182 (0.39)	305 (0.45)	338 (0.41)
1-2	110 (.21)	128 (0.27)	158 (0.23)	237 (0.29)
1-3	105 (.20)	96 (0.20)	134 (0.20)	160 (0.19)
2-2	13 (.02)	19 (0.04)	20 (0.03)	31 (0.04)
2-3	33 (.06)	26 (0.06)	39 (0.06)	44 (0.05)
3-3	17 (.03)	21 (0.04)	25 (0.04)	19 (0.02)
Total	532 (97.1%)	472 (98.5%)	681 (97.6%)	829 (96.4%)
Case/controls ^a		Odds ratio	CI	P-value
Allele 1		1.16	0.98–1.36	0.082
Allele 2		0.72	0.58–0.88	0.002
Allele 3		1.13	0.91–1.39	0.272
Alleles	Cases	Fathers	Mothers	Controls
1	723 (0.68)	588 (0.62)	902 (0.66)	1073 (0.65)
2	169 (0.16)	192 (0.20)	237 (0.17)	343 (0.21)
3	172 (0.16)	164 (0.17)	223 (0.16)	242 (0.15)
TDT ^b	Passed	Not passed	P-value	
1	211 (56%)	164 (44%)	0.016	
2	102 (39%)	159 (61%)	0.001	
3	107 (49%)	112 (51%)	0.736	
Log linear ^c	Case χ^2	P-value	Maternal χ^2	P-value
1	7.64	0.022	3.78	0.151
2	12.63	0.002	2.5	0.287
3	1.54	0.462	0.01	0.997

^aOdds of being a case for each allele present compared to someone with none, assessed by logistic regression.

^bSummary of transmission of alleles from all heterozygous parents as assessed by TDT.

^cLog linear analysis modeling for case and maternal effects.

NTD = neural tube defects; CI = confidence interval; TDT = transmission disequilibrium test.

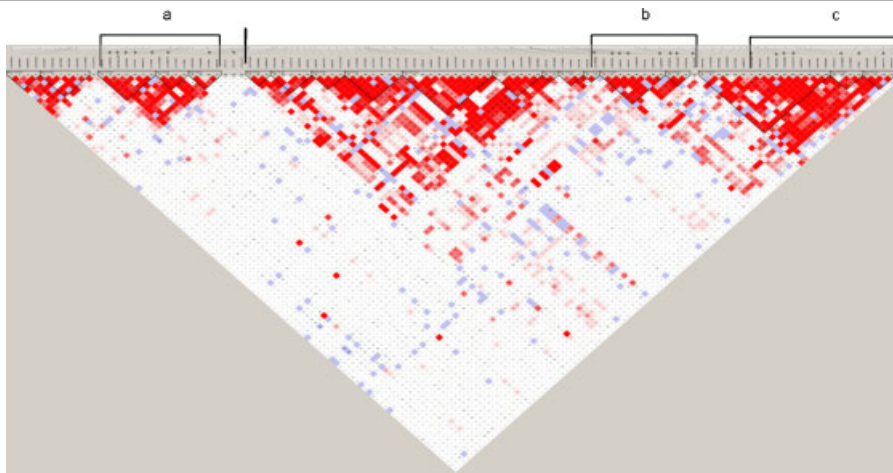


Figure 2. LD plot of pairwise values of D' for 119 markers within and adjacent to *MTHFD1L*. The *MTHFD1L* gene (~250 kb) is depicted by an arrow indicating the direction of transcription. The LD plot was constructed using genotype data from 338 Irish controls. Brackets (a, b, c) mark three regions in which single markers were found to be associated with NTD risk. Region “a” contains the rs3832406 DIP. Each individual polymorphism that showed an association with NTD are indicated by an asterisk (*). The single vertical line indicates the location of the SNP rs6922269:A > G (associated with coronary artery disease risk), which was not included in this LD plot.

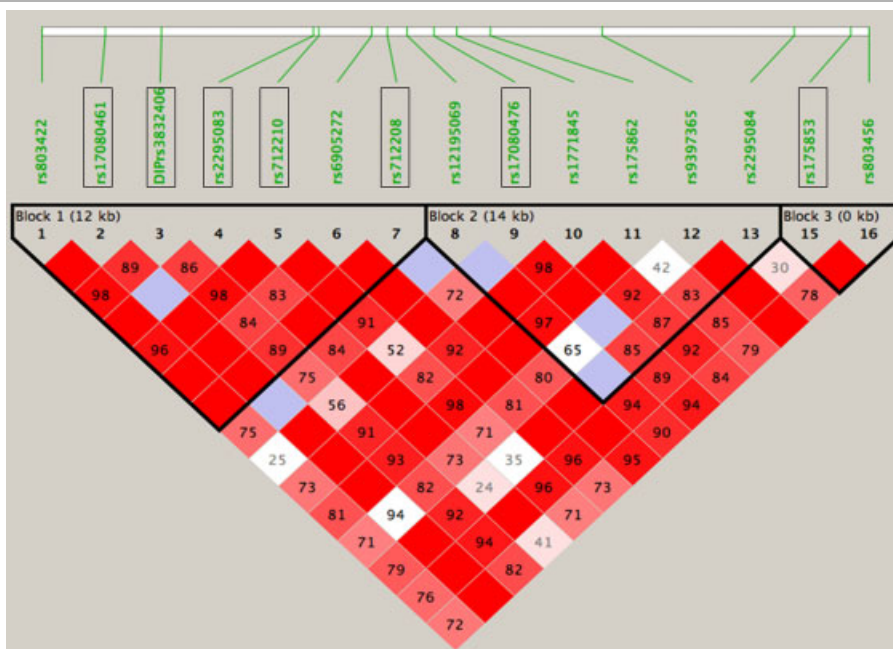


Figure 3. Linkage disequilibrium (LD) map of one associated region of the *MTHFD1L* Gene (marked region “a” in Fig. 2). LD plot of pairwise values of D' for the 15 markers within the region spanning introns 7–10 of the *MTHFD1L* gene. Boxed markers were associated with NTD risk. The rs3832406 DIP is the third marker from the left.

provided evidence for a functional effect of this variant. However, given the nature of LD and association studies, we cannot rule out a yet to be identified polymorphism from this genomic region as contributing to NTD risk.

Discussion

Our functional analysis and association study of the *MTHFD1L* gene has identified rs3832406 as a risk factor for NTDs by impacting on alternative splicing efficiency. This interesting polymorphism resides within the PPT of intron 7; an important element that is recognized by the splicing machinery. The

consensus “UCUU” within a pyrimidine-rich sequence provides the optimal binding site for the polypyrimidine tract binding protein (PTB) [Perez et al., 1997]. PTB, also known as hnRNP-I, acts as a splicing repressor through interference with a necessary component of the spliceosome, U2AF [Izquierdo et al., 2005; Sharma et al., 2005]. Splicing repression via PTB involves binding at sites both upstream and downstream of the exon [Amir-Ahmady et al., 2005]. The upstream site adjacent to the 3' splice site appears to bind PTB with high affinity, that is, equivalent to the PTB site adjacent to rs3832406 in *MTHFD1L*. Cooperative binding with weaker PTB binding site(s) is necessary for splicing repression. *MTHFD1L* contains at least four consensus

Table 2. Haplotype Frequency Estimates of MTHFD1L Blocks^a as Defined by a Solid Spine of LD

	Controls	NTD cases	NTD mothers
<i>Block 1^b</i>			
AG1GAAG	0.31	0.28	0.27
GG1GAAG	0.17	0.18	0.18
GG2GGGG	0.14	0.16	0.16
GG1AGAG	0.13	0.19	0.18
GA2GGAA	0.12	0.08	0.09
GG2GGAA	0.08	0.06	0.05
GG1GGGG	0.01	0.01	0.02
AG2GAAG	<0.01	<0.01	0.02
GG2GAAG	<0.01	<0.01	0.01
GA1GGAA	<0.01	<0.01	0.01
<i>P</i> -value ^c		0.5	0.2
<i>Block 2</i>			
GAAAGG	0.22	0.23	0.23
GGAAGG	0.17	0.13	0.11
GAGGGG	0.16	0.16	0.17
GAGAGG	0.12	0.12	0.13
GAGAAG	0.11	0.13	0.14
AAGGGG	0.09	0.09	0.08
GAGAGA	0.08	0.11	0.11
GAGGAG	0.03	0.02	0.02
GAAAAG	<0.01	<0.01	0.01
<i>P</i> -value		0.3	0.7
<i>Block 3</i>			
CC	0.52	0.52	0.51
AA	0.33	0.32	0.33
CA	0.15	0.16	0.16
<i>P</i> -value		0.4	0.5

^aBlocks are shown in Figure 3 and consist of the following SNPs in order: Block 1: rs803422, rs17080461, rs3832406, rs2295083, rs712210, rs6905272, rs71208. Block 2: rs12195069, rs17080476, rs1771845, rs175862, rs9397365, rs2295084. Block 3: rs175853, rs803456.

^bThe intron 7 DIP rs3832406 is the third marker in this haplotype block (1 = ATT₇, 2 = other).

^cCase-control and mother-control comparisons of haplotype frequency distribution were performed for each block using a permutation test and PHASE 2.1.1. NTD = neural tube defects; LD = linkage disequilibrium.

“UCUU” sites located both upstream and downstream of alternate exon 8a. Our data indicates that a change in the length of the polypyrimidine tract at the high-affinity PTB binding site interferes with the efficiency of PTB-mediated repression, possibly by weakening cooperative binding of multiple PTBs. A shortened polypyrimidine tract as in Allele 1 (ATT₇) appears to result in less efficient splicing of exon 8a, whereas extending the tract by three bases as in Allele 2 (ATT₈) results in more efficient splicing of exon 8a. The functional consequences of this are that Allele 1 carriers, particularly in the homozygous state, have a higher proportion of functional MTHFD1L mRNA compared to Allele 2 carriers (the shorter alternatively spliced mRNA lacks enzyme activity). Direct assessment of endogenous MTHFD1L protein level and activity is complicated by the difficulty of separating the mitochondrial form of this enzyme from the more abundant cytoplasmic form as illustrated by the controversy that surrounded the existence of MTHFD1L initially [Prassannan et al., 2003].

The association analysis of rs3832406 in our NTD study cohort provides strong evidence that this common variant increases the case risk of an NTD. The case-control analysis association was strongly supported by the TDT analysis, which identified opposite effects on risk for two of the alleles. Allele 1 was associated with an increase of NTD risk, whereas Allele 2 was associated with a decreased risk (Table 1). Allele 3 appeared to have no impact on NTD risk, possibly by not significantly influencing the ratio of the long to short mRNA form of *MTHFD1L*; but this requires

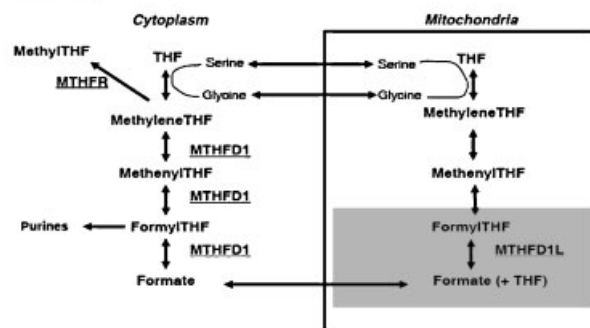


Figure 4. A simplified schematic of folate metabolism in the mitochondria and cytoplasm. Based on a figure from Prassannan et al. [2003]. THF = tetrahydrofolate. Enzyme abbreviations are underlined adjacent to the reaction(s) that they catalyze. Not all enzymes are included. MTHFD1L: monofunctional enzyme that catalyses the reversible synthesis of 10-formyltetrahydrofolate to formate and tetrahydrofolate. Reaction is shaded in gray. MTHFD1: trifunctional enzyme that catalyses the reversible synthesis of 10-formylTHF to formate and tetrahydrofolate in the cytoplasm and interconverts 5,10-methenylTHF and 5,10-methyleneTHF via its cyclohydrolase and dehydrogenase activities. MTHFR: carries out the irreversible conversion of 5,10-methyleneTHF to 5-methylTHF that provides one-carbons for S-adenosylmethionine synthesis (SAM). SAM provides the methyl group required for cellular methylation reactions.

further investigation. The association of rs3832406 prompted an assessment of other variants within the gene region. Our screen of an additional 118 SNP markers in a subset of our NTD study cohort confirmed the region surrounding and including intron 7 as harboring variant(s) that influence the risk of NTDs in the Irish population (Figs. 2 and 3). Haplotype analysis did not detect a significant association with any of the haplotypes and risk of NTDs but haplotype “GG1AGAG” showed a frequency of 13% in controls compared to 19% in cases, whereas haplotype “GA2GGAA” showed a frequency of 12% in controls compared to 8% in cases. The association of the *MTHFD1L* gene region incorporating introns 7 to 10 with NTD risk may be due to rs3832406 itself, the DIP in the context of other variants in the haplotype, or there may be an unidentified disease causing polymorphism(s) that remain to be identified. However, the location of the rs3832406 within the PPT adjacent to exon 8a, the impact on alternative splicing efficiency, and NTD disease association points toward this triallelic polymorphism as the strongest candidate for directly contributing to disease causation. Our analysis also identified two other regions of the *MTHFD1L* gene as harboring additional variants that are associated with NTDs.

Evidence is accumulating that the *MTHFD1L* gene is not just important for risk of NTDs. A recent genome-wide association study also identified *MTHFD1L* as a risk factor of coronary artery disease (CAD) in both UK and German populations [Samani et al., 2007]. A lead-positive SNP in their analysis, rs6922269:A>G, resides in intron 11, with a recombination hotspot occurring between it and rs3832406. Although not genotyped in the current study, its physical location excludes it from the three SNP clusters showing NTD association (Fig. 2). Thus, it appears that rs6922269:A>G and rs3832406 represent separate risks in their respective disease associations.

Approximately 41% of populations of North European descent are homozygous for Allele 1. Our Q RT-PCR data predict that these individuals could have up to 70% more functional MTHFD1L, and this somehow alters their risk of having an

NTD. How does having more MTHFD1L increase ones risk of having an NTD? Studies have shown that folate metabolism is compartmentalized between the cytoplasm, mitochondrion [Appling, 1991] and more recently the nucleus [Anderson et al., 2007]. This compartmentalization is thought to facilitate the different metabolic roles within the cell [Anderson et al., 2007]. The mitochondrial folate pathway is believed to play an important role during embryogenesis by ensuring an adequate supply of formate and glycine [Christensen and MacKenzie, 2006]. Formate is the preferred one-carbon donor for purine synthesis. Mitochondrial C₁-Tetrahydrofolate Synthase, encoded by *MTHFD1L*, supplies this formate by catalyzing the reversible synthesis of 10-formyltetrahydrofolate to formate and tetrahydrofolate (Fig. 4). Increased production of formate in Allele 1 homozygotes may disrupt the one-carbon flux through the mitochondria, and thus interfere with cellular proliferation. Alternatively, a higher level of formate itself may be toxic to cells, which also results in disrupted cellular production during embryogenesis.

In conclusion, we have identified an *MTHFD1L* functional polymorphism, which appears to influence NTD disease risk by affecting splicing efficiency. We acknowledge that replication of this association in another population is required to demonstrate whether this gene has relevance for NTDs outside of Ireland. The *MTHFD1L* gene is now implicated in two distinct common diseases, that is, NTDs and CAD. The role of MTHFD1L in disease risk highlights the importance of folate metabolism in maintaining health.

Acknowledgments

These studies would not be possible without the participation of the affected families, and their recruitment by the Irish Association of Spina Bifida and Hydrocephalus and the Irish Public Health Nurses. The authors also thank Peter Chines and Kristine Krebs for their computational assistance. Grant sponsors: Intramural research programs of the National Human Genome Research Institute, the National Institute of Child Health and Human Development and the Health Research Board (HRB), Ireland, including an HRB Research Project Grant (to A.P.-M.) (Grant number RP/2005/58).

References

Amir-Ahmady B, Boutz PL, Markovtsov V, Phillips ML, Black DL. 2005. Exon repression by polypyrimidine tract binding protein. *RNA* 11:699–716.

Anderson DD, Woeller CF, Stover PJ. 2007. Small ubiquitin-like modifier-1 (SUMO-1) modification of thymidylate synthase and dihydrofolate reductase. *Clin Chem Lab Med* 45:1760–1763.

Appling DR. 1991. Compartmentation of folate-mediated one-carbon metabolism in eukaryotes. *FASEB J* 5:2645–2651.

Barrett JC, Fry B, Maller J, Daly MJ. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265.

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

Brody LC, Conley M, Cox C, Kirke PN, McKeever MP, Mills JL, Molloy AM, O’Leary VB, Parle-McDermott A, Scott JM, Swanson DA. 2002. A polymorphism, R653Q, in the trifunctional enzyme methylenetetrahydrofolate dehydrogenase/methylenetetrahydrofolate cyclohydrolase/formyltetrahydrofolate synthetase is a

maternal genetic risk factor for neural tube defects: report of the Birth Defects Research Group. *Am J Hum Genet* 71:1207–1215.

Christensen KE, Patel H, Kuzmanov U, Mejia NR, MacKenzie RE. 2005. Disruption of the *Mthfd1* gene reveals a monofunctional 10-formyltetrahydrofolate synthetase in mammalian mitochondria. *J Biol Chem* 280:7597–7602.

Christensen KE, MacKenzie RE. 2006. Mitochondrial one-carbon metabolism is adapted to the specific needs of yeast, plants and mammals. *BioEssays* 28:595–605.

Czeizel AE, Dudas I. 1992. Prevention of the first occurrence of neural-tube defects by periconceptional vitamin supplementation. *N Engl J Med* 327:1832–1835.

Hull J, Campino S, Rowlands K, Chan M-S, Copley RR, Taylor MS, Rockett K, Elvidge G, Keating B, Knight J, Kwiatkowski D. 2007. Identification of common genetic variation that modulates alternative splicing. *PLoS Genet* 3:1009–1018.

Izquierdo JM, Majos N, Bonnal S, Martinez C, Castelo R, Guigo R, Bilbao D, Valcarcel J. 2005. Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. *Mol Cell* 19:475–484.

Kirke PN, Scott JM. 2005. Pre-conception nutrition and prevention of neural tube defects. In: Sadler MJ, Strain JJ, Caballero B, editors. *Encyclopaedia of human nutrition*. 2nd ed. New York: Oxford Academic Press. p 15–27.

Leino A. 1999. Fully automated measurement of total homocysteine in plasma and serum on the Abbott IMx analyser. *Clin Chem* 45:569–571.

Molloy AM, Scott JM. 1997. Microbiological assay for serum, plasma, and red cell folate using cryopreserved, microtiter plate method. *Methods Enzymol* 281:43–53.

Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG. 2004. Genetic analysis of genome-wide variation in human gene expression. *Nature* 430:743–747.

MRC Vitamin Study Research Group. 1991. Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. *Lancet* 338:131–137.

Parle-McDermott A, Kirke PN, Mills JL, Molloy AM, Cox C, O’Leary VB, Pangilinan F, Conley M, Cleary L, Brody LC, Scott JM. 2006. Confirmation of the R653Q polymorphism of the trifunctional C1-synthase enzyme as a maternal risk for neural tube defects in the Irish population. *Eur J Hum Genet* 14:768–772.

Perez I, Lin C-H, McAfee JG, Patton JG. 1997. Mutation of PTB binding sites causes misregulation of alternative 3’ splice site selection in vivo. *RNA* 3:764–778.

Prassananan P, Pike S, Peng K, Shane B, Appling DR. 2003. Human mitochondrial C1-tetrahydrofolate synthase. *J Biol Chem* 278:43178–43187.

Samani NJ, Erdmann J, Hall AS, Henstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann HE, Barrett JH, König IR, Stevens SE, Szymczak S, Tregouet DA, Iles MM, Pahlke F, Pollard H, Lieb W, Cambien F, Fischer M, Ouwehand W, Blankenberg S, Balmforth AJ, Baessler A, Ball SG, Strom TM, Braenne I, Gieger C, Deloukas P, Tobin MD, Ziegler A, Thompson JR, Schunkert H, WTCCC and the Cardiogenics Consortium. 2007. Genomewide Association Analysis of Coronary Artery Disease. *N Engl J Med* 357:443–453.

Sharma S, Falick AM, Black DL. 2005. Polypyrimidine tract binding protein blocks the 5’ splice site-dependent assembly of U2AF and the pre-spliceosomal E complex. *Mol Cell* 19:485–496.

Stephens M, Donnelly P. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 73:1162–1169.

Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68:978–989.

Walkup AS, Appling DR. 2005. Enzymatic characterization of human mitochondrial C₁-tetrahydrofolate synthase. *Arch Biochem Biophys* 442:196–205.

Wang G-S, Cooper TA. 2007. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* 8:749–761.

Weinberg CR, Wilcox AJ, Lie RT. 1998. A log-linear approach to case-parent-triad data: assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. *Am J Hum Genet* 62:969–978.

Wilcox AJ, Weinberg CR, Lie RT. 1998. Distinguishing the effects of maternal and offspring genes through studies of “case-parent triads.” *Am J Epidemiol* 148:893–901.