Published in final edited form as:
Hum Genet. 2009 April ; 125(3): 247–256. doi:10.1007/s00439-008-0616-3.

Analysis of the *MTHFD1* promoter and risk of neural tube defects

Nicola Carroll[1], **Faith Pangilinan**[2], **Anne M. Molloy**[3], **James Troendle**[4], **James L. Mills**[4], **Peadar N. Kirke**[5], **Lawrence C. Brody**[2], **John M. Scott**[6], and **Anne Parle-McDermott**[1,*]

[1]Nutritional Genomics Group, School of Biotechnology, Dublin City University, Dublin 9, Ireland [2]Molecular Pathogenesis Section, Genome Technology Branch, National Human Genome Research Institute, Building 50, Room 5306, 50 South Drive, MSC 8004, Bethesda, MD 20892-8004, USA [3]School of Medicine, Trinity College Dublin, Dublin 2, Ireland [4]Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD, USA [5]Child Health Epidemiology Division, Health Research Board, Dublin, Ireland [6]School of Biochemistry & Immunology, Trinity College Dublin, Dublin 2, Ireland

Abstract

Genetic variants in MTHFD1 (5,10-methylenetetrahydrofolate dehydrogenase/ 5,10-methenyltetrahydrofolate cyclohydrolase/ 10-formyltetrahydrofolate synthetase), an important folate metabolic enzyme, are associated with a number of common diseases, including neural tube defects (NTDs). This study investigates the promoter of the human *MTHFD1* gene in a bid to understand how this gene is controlled and regulated. Following a combination of *in silico* and molecular approaches, we report that *MTHFD1* expression is controlled by a TATA-less, Initiator-less promoter and transcription is initiated at multiple start sites over a 126bp region. We confirmed the presence of three database polymorphisms (dbSNP) by direct sequencing of the upstream region (rs1076991 C>T, rs8010584 G>A, rs4243628 G>T), with a fourth (dbSNP rs746488 A>T) not found to be polymorphic in our population and no novel polymorphisms identified. We demonstrate that a common SNP rs1076991 C>T within the window of transcriptional initiation exerts a significant effect on promoter activity *in vitro*. We investigated this SNP as a potential risk factor for NTDs in a large homogenous Irish population and determined that it is not an independent risk factor, but, it does increase both case ($\chi^2 = 11.06$, $P = 0.001$) and maternal ($\chi^2 = 6.68$, $P = 0.01$) risk when allele frequencies were analysed in combination with the previously identified disease-associated p.R653Q (c.1958 G>A; dbSNP rs2236225) polymorphism. These results provide the first insight into how *MTHFD1* is regulated and further emphasise its importance during embryonic development.

Keywords

MTHFD1; NTD; Functional; SNP; R653Q; Promoter

Introduction

Folate, or vitamin B9, is an essential nutrient in our diet. Folate deficiency and polymorphisms within folate-dependent enzymes have been extensively associated with a number of common complex diseases and disorders, particularly, neural tube defects (NTDs). Non-syndromic NTDs are among the most common congenital disorders, occurring at an average rate of approximately 1 in 1000 pregnancies per year (Busby *et al.*, 2005). The aetiology of NTDs is

*Correspondence to: Dr. Anne Parle-McDermott, Nutritional Genomics Group, School of Biotechnology, Dublin City University, Dublin 9, Ireland. E-mail: anne.parle-mcdermott@dcu.ie. Facsimile: 00353 (0) 1 7005412.

multifactorial, with both genetic and environmental factors, and the interaction between them playing a critical role. It is widely known and accepted that periconceptional folate supplementation can prevent up to 70% of all NTDs (MRC 1991; Czeizel & Dudas, 1992). The exact mechanism of this protective effect is not completely understood but it is most likely by overcoming disruptions in folate metabolism, partially caused by underlying genetic variation in folate-related genes (van der Linden *et al.*, 2006).

One such genetic variant occurs in the gene for 5,10-methylenetetrahydrofolate (methyleneTHF) dehydrogenase; 5,10-methenylTHF cyclohydrolase; 10-formylTHF synthetase (*MTHFD1*; GenBank accession no. NM_005956.2). This gene encodes a 100kDa nicotinamide adenine diphosphate (NADP+)-dependent trifunctional cytoplasmic enzyme that plays an important role in folate metabolism (Hum *et al.*, 1988). The enzyme consists of two functional domains; an amino-terminal portion (33kDa) containing the dehydrogenase and cyclohydrolase activities, and a larger (67kDa) synthetase domain in the carboxy-terminal region. MTHFD1 activity is essential for DNA synthesis, providing 10-formylTHF and 5,10-methyleneTHF for *de novo* purine and thymidylate synthesis. In addition to its enzymatic activity, biochemical evidence also suggests that MTHFD1 plays a role as a structural component in a multi-enzyme purine synthesising complex (Smith *et al.*, 1980; Barlowe & Appling, 1990). Therefore, MTHFD1 is a vital enzyme, especially in rapidly dividing cells such as those of the developing embryo, where purines and pyrimidines are in constant demand for *de novo* DNA synthesis. A common single nucleotide polymorphism (SNP) at nucleotide 1958 of the *MTHFD1* gene causes a G to A transition, which results in an arginine to glutamate substitution at amino acid position 653 in the synthetase domain of the enzyme (dbSNP ID: rs2236225; Hol *et al.*, 1998). This variant (usually referred to as R653Q) has been identified as a maternal risk factor for NTDs in the Irish population, with an excess of QQ homozygotes found in mothers of children with an NTD (Brody *et al.*, 2002; Parle-McDermott *et al.*, 2006). Association with NTD risk was also reported in the Italian (de Marco *et al.*, 2006), but not in the Dutch, population (Hol *et al.*, 1998; van der Linden *et al.*, 2007). The R653Q polymorphism has also been identified as a risk factor for severe abruptio placentae and mid-trimester miscarriage in Irish mothers (Parle-McDermott *et al.*, 2005(a); Parle-McDermott *et al.*, 2005(b)), intrauterine growth restriction in Australian mothers (Furness *et al.*, 2008) and congenital heart defects (CHD) in Canadian children (Christensen *et al.*, 2008); although the association with CHD was not identified in a Chinese population (Cheng *et al.*, 2005). It has also been associated with increased risk for gastric cancer in a Chinese population (Wang *et al.*, 2007), as well as, increased risk for bipolar disorder and schizophrenia in a male Polish population (Kempisty *et al.*, 2007). Despite extensive evidence of its association with numerous common diseases, very little is known of the underlying functional effect of this SNP. A recent study showed the polymorphism had no effect on synthetase enzyme activity under normal assay conditions, but it was shown to affect enzyme thermostability and to diminish its capacity for *de novo* purine synthesis (Christensen *et al.*, 2008). Although the coding region and 3′ untranslated region (UTR) of the *MTHFD1* gene have been investigated for novel polymorphisms (Parle-McDermott *et al.*, 2006) the 5′ region, including the promoter region had not been investigated prior to this study.

We sought to analyse the human *MTHFD1* promoter by utilising both bioinformatics and experimental approaches to provide an understanding of the mechanisms responsible for regulation at the transcriptional level. Our investigation included a search for novel polymorphisms that may impact on gene expression and thus, could be related to the pathogenesis of a developmental defect, such as an NTD. We report here that *MTHFD1* expression is controlled by a typical TATA-less, Initiator (Inr)-less promoter (Smale, 1997) with transcription initiated at multiple start sites. A common SNP located within the window of transcriptional initiation significantly affects promoter activity *in vitro*. We found that this novel functional SNP is not an independent risk factor for NTDs in the Irish population, but it

is significantly associated with both case and maternal risk when analysed in combination with the R653Q (SNP rs2236225) polymorphism.

Materials & Methods

Transcription Start Site Identification

Prediction of the transcription start site (TSS) was determined initially by utilising sequence databases such as DBTSS (http://dbtss.hgc.jp) and dbEST (www.ncbi.nih.gov/EST). The TSS was assessed experimentally using the First Choice® 5′ RLM-RACE kit (Ambion®, UK). The 5′ RLM-RACE procedure was performed on 10μg total RNA extracted (Ultraspec II, Biotecx, TX, USA) from Coriell lymphoblast cell lines (Coriell Institute for Medical Research, Camden, New Jersey, US) and 250ng human placental Poly (A) RNA (Ambion®). Subsequent PCR products were cloned into pBluescript® II SK(+) and directly sequenced using a Big Dye Terminator® Sequencing Kit, Version 2.2 and an ABI PRISM™ 377 DNA Sequencer (Applied Biosystems, UK). Primer sequences used for 5′ RLM-RACE are available in the Supplementary document.

In Silico Analysis of Sequences

A CpG island plot was obtained using CpG Island Searcher (www.cpgislands.com). Putative transcription factor (TF) binding sites were identified using MatInspector, part of the Genomatix suite of bioinformatics tools (www.genomatix.de), AliBaba2, part of the Gene Regulation suite (www.gene-regulation.com), and the CONSITE algorithm was employed to identify TF binding sites that are evolutionary conserved (www.phylofoot.org/consite).

Reporter Gene Constructs

A series of overlapping PCR products spanning 2kb upstream of the translational start site of the *MTHFD1* gene were generated using either genomic DNA from Coriell® lymphoblast cell lines or a larger clone as template. The primers utilized are detailed in the Supplementary document. PCR products were cloned by conventional ligation into the pGL3 Basic vector (Promega, UK) or using Gateway® cloning (Invitrogen, UK) by employing a Gateway® converted pGL3 Basic vector (a kind gift of Glenn Maston, University Massachusetts Medical School, USA). Reporter gene constructs representing either allele of SNP rs1076991 were generated following PCR amplification of 0.59kb of the promoter region using genomic DNA from Coriell lymphoblast lines isolated from individuals that were homozygous for either allele. The sequences of these constructs were identical except for the polymorphism as verified by direct sequencing.

Reporter Gene Assays

Firefly luciferase reporter gene assays were carried out on transiently transfected Human Embryonic Kidney (HEK)-293 cells. Cells were grown in DMEM (Dulbecco's Modified Eagle's Medium) supplemented with 10% fetal calf serum, 1% Penicillin/Streptomycin (10000U: 10mg/ml) and 1% L-Glutamine (200mM). Cells were seeded at a density of 1×10^5 cells/ml, 24 hours prior to transfection. An optimised concentration of 100ng plasmid DNA was transfected using GeneJuice® Transfection Reagent (Novagen, USA) and incubated for 24 hours before assay. All cells were co-transfected with 40ng *Renilla* luciferase plasmid (pRL-TK; Promega, UK) to normalise for transfection efficiency. Cells were lysed using Passive Lysis Buffer (Promega, UK) and assayed for luminescence following incubation with either firefly or *Renilla* luciferase substrate. Luminescence was measured using either a Mediators PhL (ImmTech Inc., New Windsor, MD, USA) or a Glomax™ (Promega, UK) luminometer. Each assay was performed in triplicate and each experiment was performed at least three times.

All values were normalised to pRL-TK control values and expressed relative to empty pGL3 Basic values, with the mean, standard deviation and coefficient of variation (CV) calculated.

Polymorphism Screening

The putative regulatory region of *MTHFD1* was screened for novel polymorphisms by designing six separate overlapping PCR assays of between 320-400bp in size spanning 2kb upstream of the translational start site. An assay to PCR amplify the first 400bp of Intron 1 was also designed. All primers sequences and optimised assay conditions are available in the Supplementary document. Genomic DNA from 22 Irish individuals was used as template for all PCR assays. The PCR products from each sample were cycle sequenced bidirectionally using Big Dye Terminator® chemistry and the sequence traces were analysed for variations using the Seqman program, as part of the DNASTAR, Lasergene® suite. Haplotype block analysis was performed using the default algorithm of Haploview, version 3.2 (www.broad.mit.edu/personal/jcbarret/haploview).

Genotyping of SNP rs1076991

Genotyping of SNP rs1076991 was carried out by Matrix Assisted Laser Desorption Ionisation-Time-of-Flight (MALDI-TOF) mass spectrometry of primer extension products using the Homogenous MassEXTEND® Assay (hME®; Sequenom®, CA, USA). The following primers were used to amplify an 89bp PCR product containing the SNP: forward [5′ ACGTTGGATGGGCGCAGGCGCAGTAGTGT 3′] and reverse [5′ ACGTTGGATGAGCCAAGCAGGACAACCCAA 3′]. An extension primer [5′ GCAGTAGTGTGATCCCC 3′] was designed to anneal directly adjacent to the SNP site. Genotypes were called using SpectroTYPER® software. A PCR-Restriction Fragment Length Polymorphism (RFLP) method was also used to re-genotype >10% of DNA samples in an external quality control assay. A 334bp region surrounding the SNP was amplified by PCR and the product was digested with *Msp* I (20U) restriction enzyme, resulting in a 233bp and a 99bp product in the presence of the C allele.

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 (n = 594), consisting of both complete and incomplete triads (case, mother, father), formed the NTD cohort. Genotype information for MTHFD1 SNP rs2236225 was available for the majority of these samples from previous studies (Brody *et al.*, 2002; Parle-McDermott *et al.*, 2006).

Blood samples were obtained from a population of 56,049 pregnant women attending the three main maternity hospitals in the Dublin area between 1986 and 1990 (Kirke *et al.*, 1993). The control group (n = 999) was randomly selected from those women who did not have an NTD-affected pregnancy and had no previous NTD history. An additional control group (n = 216) from the same collection was also included as data relating to red cell folate (RCF) and homocysteine levels were available for these samples. Informed consent and ethical approval was obtained for all human samples used in this study.

Statistical Analyses

Differences in activity of the promoter constructs were assessed using a two-tailed, unpaired Student's t-test. A χ^2 test was used to compare allele frequencies between NTD triad and control groups. A homozygous TT genotype effect was investigated by calculation of an odds ratio (OR) with 95% confidence intervals. OR is computed as the ratio of carriers to non-carriers in cases compared to controls and, thereby, estimates the risk conferred by the particular genotype.

Log-linear analysis (Weinberg *et al.*, 1998; Wilcox *et al.*, 1998) was performed using the SAS PROC GENMOD program. This uses a likelihood ratio test to examine the joint transmission of alleles from parents to the affected offspring, which enables detection of indirect genetic influence *i.e.* the maternal genotype and interactive genotype effects. A TDT analysis was also performed on informative triads *i.e.*, those with at least one heterozygous parent, to investigate any deviation in normal allele segregation from parents to affected child. Complete triads were analysed using the TDT/STDT program Version 1.1, which has a χ^2 distribution with 1 degree of freedom. SNP-SNP interactions in NTD mothers and cases were assessed by logistic regression using SAS PROC GENMOD software. The OR for combined genotype or allele frequencies of both SNPs was estimated to assess a possible increased risk due to having a specific combination of alleles and/or genotypes in cases or mothers compared to controls. A complementary test for interaction was performed utilising log linear modelling and employed all members of the triad data rather than comparing each group to controls. These approaches for detecting SNP-SNP interactions allows for testing combined effects on risk and are not biased towards previously established independent SNP effects. Significance was set at $P < 0.05$ for all statistical testing.

Results

Transcription Start Site Identification

A multi-site pattern of transcriptional initiation was predicted *in silico* and then experimentally confirmed using the 5′ RLM-RACE method. At least 28 different TSSs were experimentally identified in a region spanning from -30 to -156bp, relative to the ATG start codon (Supplementary Figure 1). Within this 126bp window of initiation three TSSs, at positions -68, -72, and -100, were used most frequently and have been designated as the major TSSs of the *MTHFD1* gene. An alternate upstream exon was not identified and there were no significantly different patterns of transcriptional initiation between individuals or between lymphocytes and placental cells. Interrogation of the EST database (dbEST) confirmed the absence of an alternative upstream exon and the lack of a definitive transcription start site (Supplementary Figure 2).

Bioinformatics & Reporter Gene Assays Define the *MTHFD1* Promoter region

In silico analysis revealed that the *MTHFD1* promoter is both TATA-less and Inr-less. Regions of regulatory importance are indicated by the presence of a 1.38kb CpG island spanning a region from 1kb upstream extending into Intron 1. Numerous putative TF binding sites were identified within the upstream region; far too many to all have biological relevance. Therefore, comparison to previously characterised promoters of similar folate-related/TATA-less genes was employed to identify those likely to be functional. Putative Sp1, E2F, and NRF-1 TF binding sites were identified, as well as, a consensus E-box (CACGTG) that is conserved across human, rat and mouse species (Figure 1). Experimental investigation of promoter activity revealed significant levels of transcriptional activity in constructs ranging from 0.26 to 2kb upstream of the ATG start codon (Figure 2). A construct of 0.59kb displayed the highest level of activity but was similar to the 0.47kb and 1kb constructs. Promoter activity was reduced to basal levels in a 0.11kb construct, demonstrating that the region most important for activated *MTHFD1* gene transcription occurs between 0.11kb and 0.47kb upstream. The 1.94kb construct showed a drop in luciferase activity, possibly due to the presence of a repressor element. The mean intra-assay coefficient of variation (CV) for these experiments was 10.2% and the mean inter-assay CV was 9.6%.

Polymorphism Screen

A sequencing screen of the *MTHFD1* regulatory region, spanning from 2kb upstream of the translational start site to the first 400bp of Intron 1, was performed with no novel

polymorphisms identified. Three dbSNP-listed SNPs were identified in the upstream region: -105 C>T (rs1076991), -1470 G>A (rs8010584), and -1474 G>T (rs4243628). A fourth SNP present in dbSNP (-473 A>T; rs746488) was not identified in our screen, suggesting this SNP is either a rare variant and was not detected in this small screen, or it is not variable in the Irish population. Haplotype analysis revealed that all three identified SNPs are in LD with each other in the same haplotype block. SNP rs1076991 and rs8010584 are in complete LD ($D' = 1$; $R^2 = 0.76$) and SNP rs4243628 is in strong LD with both ($D' = 0.87$; $R^2 = 0.51$ and $D' = 0.79$; $R^2 = 0.48$, respectively). These SNPs are not in LD with SNP rs2236225 ($D' = 0.2$; $R^2 = 0.01$). Common haplotypes for the three upstream SNPs in this population are: GGC = 0.41; TAT = 0.4; GAT = 0.1.

Functional Characterisation of SNP rs1076991

The presence of the common SNP rs1076991 C>T in the window of transcriptional initiation has a significant effect on *MTHFD1* promoter activity *in vitro*, with the 'T' 0.59kb promoter construct only 38% as transcriptionally active as the 'C' 0.59kb promoter construct in a luciferase reporter gene assay ($P = 0.04$; Figure 3). No other sequence variation was present between the two constructs, since both contained the more common allele for SNP rs746488 while SNPs rs8010584 and rs4243628 are localised 5′ to the 0.59kb construct.

The existence of a biochemical phenotype for SNP rs1076991 was investigated following genotyping of a control group, by correlating rs1076991 genotype with red cell folate and homocysteine levels. A significant difference in RCF and/or homocysteine levels was not observed between different genotype groups. Mean RCF levels (ng/ml) with 95% confidence interval for the CC, CT, and TT genotypes were: 349 [302-389], 324 [295-347], and 316 [288-347], respectively, which are not significantly different ($P = 0.54$). Mean homocysteine levels were 7.57 [7.0-8.2], 8.07 [7.6-8.6], and 7.69 [7.2-8.2] for the CC, CT, and TT genotypes, respectively, which again are not significantly different ($P = 0.34$).

Genotyping of SNP rs1076991 and Association with NTDs

A total of 2462/2561 (96%) DNA samples were genotyped successfully using the mass spectrometry based assays. Of these samples, 284 were re-genotyped as a quality control measure, with a 94% agreement rate. A further 277 samples were re-genotyped in an external quality control assay and the rate of agreement was 96%. Genotype and allele frequencies for SNP rs1076991 in each of the study groups are displayed in Table 1.

Case/control analysis revealed that SNP rs1076991 is not associated with NTD risk in the Irish population, since there is no significant difference in genotype/allele frequencies between NTD family groups and the control group (Table 1). Likewise, log-linear analysis did not reveal any risk association with case genotype ($\chi^2 = 0.23$, $P = 0.89$), with maternal genotype ($\chi^2 = 4.53$, $P = 0.1$), or with case-maternal interaction ($\chi^2 = 0.86$, $P = 0.86$). A TDT analysis was carried out on complete triads (n = 353). A total of 359 informative genotypes *i.e.* heterozygous parents, were identified and transmission analysed. No preferential transmission of either allele was observed [C = 185 (52%), T = 174 (48%), $\chi^2 = 0.34$, $p = 0.53$].

A potential SNP-SNP interaction between SNP rs1076991 and SNP rs2236225 in relation to NTD-risk was also investigated. Two-locus genotype frequencies for SNP rs1076991 and SNP rs2236225 were calculated for NTD family groups and controls (Table 2). Comparison of combined allele frequencies revealed an interaction in both the case group ($\chi^2 = 11.06$, $P = 0.001$) and the maternal group ($\chi^2 = 6.68$, $P = 0.01$). Analysis of genotype combinations revealed that both NTD cases (0.11) and mothers (0.11) have a higher frequency of the rs1076991 TT/rs2236225 AA genotype compared to controls (0.05). Logistic regression analysis of these genotype combinations revealed a significant interaction between SNP

rs1076991 and SNP rs2236225 in cases ($\chi^2 = 20.15$, $P = 0.001$) but not in mothers ($\chi^2 = 7.01$, $P = 0.14$). Maternal and case genotype interactive effects were also tested using the triad data only. This analysis complements the maternal/cases versus control analyses but has less power due to incomplete triads. Statistical significance was not reached for maternal ($P = 0.10$) or case genotype interactions ($P = 0.36$).

Discussion

MTHFD1 is a logical candidate gene for investigation in relation to disease risk; thus, we have undertaken the primary study of the promoter controlling its expression, and report the existence of a functional promoter SNP, rs1076991 C>T, which affects gene expression *in vitro*. Our results demonstrate that the *MTHFD1* gene is regulated by a TATA-less, Inr-less promoter that directs transcriptional initiation at multiple start sites within a 126bp initiation window in the upstream region. Three major TSSs were identified at positions 68, 72 and 100bp upstream of the translation start site. An alternate upstream exon is not present and TSS usage is similar between different individuals and tissue types investigated (lymphocyte and placenta). This type of TATA-less promoter with multiple start sites resembles that found in other folate-related genes, including thymidylate synthase (Dong *et al.*, 2000) and reduced folate carrier (Gong *et al.*, 1999), and is also similar to the promoter controlling rat *Mthfd1* expression (Howard *et al.*, 2003). A 1.38kb CpG island spans the promoter region, and is associated with the presence of multiple GC boxes, suggesting that Sp1 is an important TF for gene regulation (Brandeis *et al.*, 1994). It has been proposed that these Sp1 sites may serve to maintain the hypomethylated state of such promoters, indicating that the gene will be constitutively expressed (Pugh & Tjian, 1990). *In silico* prediction and previously reported empirical evidence from large-scale promoter studies suggest the E2F family of TFs will also play an important role in *MTHFD1* gene expression (Cam *et al.*, 2004; Weinmann *et al.*, 2002). Therefore, it is likely that *MTHFD1* will be cell cycle-regulated and, although predicted to be constitutively expressed, an up-regulation of its expression during S phase would be expected given its role in DNA synthesis, followed by a down-regulation in the G_0 and G_1 phases. *In silico* analysis also indicates that c-Myc-Max binding to a conserved E-box motif regulates *MTHFD1* expression. Cross-species conservation signifies the importance of this promoter feature, since nucleotide sequences that are actively conserved during evolution are likely to be of biological importance. Evidence of functionality is also supported by identification of cMyc-Max binding to the *MTHFD1* upstream region in another large-scale promoter binding study (Mao *et al.*, 2003). The c-Myc TF can play a role in recruiting factors necessary for the initiation of transcription to the core promoter (Hermann *et al.*, 2001; McEwan *et al.*, 1996) and could be responsible for this process in the absence of a TATA box/Inr in the *MTHFD1* promoter. Co-ordinated binding of these TFs, as well as other predicted ones such as NRF-1, is likely responsible for the high levels of activated transcription measured from the *MTHFD1* promoter, especially within the first 0.47kb of the upstream region, which supports the highest level of transcriptional activity. Activated transcription was not induced by a promoter construct of 0.11kb, demonstrating the absence of essential regulatory elements and indicating that the minimal promoter region for activated transcription of this gene is between 0.11kb and 0.26kb upstream. The drop in activity observed in the 1.94kb construct may be due to a yet to be identified repressor element. Our analysis was confined to the 2kb region upstream of the translational start site of MTHFD1, thus, we cannot rule out the role of additional regulatory elements further upstream of this.

Promoter function and normal gene expression can be significantly affected by polymorphisms in important regulatory regions (Hoogendoorn *et al.*, 2003). SNP rs1076991 C>T is located within the window of initiation and was shown to have a significant impact on promoter function *in vitro*. Transcriptional activity of the 0.59kb ‘T’ promoter construct was shown to be about 1.6 fold less than that of the 0.59kb wildtype ‘C’ construct. If this effect is translated

in vivo, a decrease in *MTHFD1* gene expression could result in lower levels of the MTHFD1 enzyme available for purine and thymidylate synthesis; a situation that could be detrimental under certain conditions, especially during times of increased demand on *de novo* DNA synthesis, such as embryogenesis. However, results obtained from *in vitro* reporter gene studies should be interpreted with caution, since gene regulation and expression in the natural genomic environment *in vivo* is undoubtedly more complex than seen in the cell line model. Bioinformatic analysis did not reveal an alteration of any consensus TF binding site that would explain the observed difference in activity between the two genotypes. The loss of a DNA methylation site is also possible, but is unlikely to explain the reduced expression of MTHFD1 *in vitro*. The most likely explanation for the functional impact of SNP rs1076991 on MTHFD1 gene expression is through the loss or gain of binding to a non-consensus binding site, the identification of which would require further investigation.

Polymorphisms that exert a functional effect, such as SNP rs1076991, are those most likely to be involved in common disease. The link between disruptions to folate metabolism and NTD risk is well established and, more specifically, variation in the MTHFD1 folate enzyme has previously been associated with NTD risk in the Irish population (Brody *et al.*, 2002; Parle-McDermott *et al.*, 2006). Therefore, SNP rs1076991 was investigated as a candidate polymorphism for NTD risk in the Irish population in a large association study using both case/control and family triad-based analysis methods. SNP rs1076991 was not associated with NTD risk/protection in this study, nor did it have an effect on RCF or homocysteine levels analysed in a separate control group. However, SNP-SNP interaction analysis with MTHFD1 SNP rs2236225 (R653Q) revealed a highly significant association with NTD risk in both case (genotype and allele frequencies) and maternal groups (allele frequencies only). These two SNPs are not in LD with each other and, therefore, the identified interaction cannot be attributed to simple co-segregation. Therefore, it seems that the SNP rs1076991, while not an independent risk factor for NTDs, in some way contributes to the risk associated with SNP rs2236225 and homozygosity for these two SNPs confers a greater risk than either one in isolation. SNP-SNP interaction is particularly relevant in the aetiology of common complex diseases, where it is likely that the mechanism of action of one variant may be influenced by the presence or absence of another. Further investigations of SNP-SNP interaction in relation to abruptio placentae, mid-trimester miscarriage risk, and would be of particular interest, since the rs2236225 AA genotype is a known risk factor (Parle-McDermott *et al.*, 2005(a); Parle-McDermott *et al.*, 2005(b)) for both these conditions and it is possible that the presence of SNP rs1076991 confers an even greater risk. However, we do acknowledge the limitations of our data set and confirmation of this interaction in a second NTD cohort would be desirable.

We are the first to investigate the promoter of *MTHFD1* and to identify a novel functional SNP rs1076991 that may have disease relevance. The results reported here marks a step in the direction toward understanding the underlying molecular pathways and disruptions involved in folate-related disease development and progression. This is necessary to achieve the fundamental goal of elucidating the aetiology of these complex diseases and, eventually, optimising individual folate status to prevent or overcome them.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We wish to thank the Irish Association for Spina Bifida and Hydrocephalus and the Public Health Nurses for assistance with subject recruitment and the following Dublin maternity hospitals for control recruitment: National Maternity Hospital Holles street, the Coombe Womens' Hospital and the Rotunda Hospital. This research was supported by the Intramural Research Program of the NICHD, NIH and the Health Research Board of Ireland. Approval was obtained

from the Research Ethics Committee of the Health Research Board, Ireland and the Institutional Review Board at the National Institutes of Health. All research participants provided informed consent.

References

Barlowe CK, Appling DR. Molecular genetic analysis of Saccharomyces cerevisiae C1-tetrahydrofolate synthase mutants reveals a noncatalytic function of the ADE3 gene product and an additional folate-dependent enzyme. Mol Cell Biol 1990;10:5679–87. [PubMed: 2233711]

Brandeis M, Frank D, Keshet I, Siegfried Z, Mendelsohn M, Nemes A, Temper V, Razin A, Cedar H. Sp1 elements protect a CpG island from de novo methylation. Nature 1994;371:435–8. [PubMed: 8090226]

Brody LC, Conley M, Cox C, Kirke PN, McKeever MP, Mills JL, Molloy AM, O'Leary VB, Parle-McDermott A, Scott JM, Swanson DA. A polymorphism, R653Q, in the trifunctional enzyme methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase/formyltetrahydrofolate synthetase is a maternal genetic risk factor for neural tube defects: report of the Birth Defects Research Group. Am J Hum Genet 2002;71:1207–15. [PubMed: 12384833]

Busby A, Abramsky L, Dolk H, Armstrong B. Preventing neural tube defects in Europe: population based study. Bmj 2005;330:574–5. [PubMed: 15760997]

Cam H, Balciunaite E, Blais A, Spektor A, Scarpulla RC, Young R, Kluger Y, Dynlacht BD. A common set of gene regulatory networks links metabolism and growth inhibition. Mol Cell 162004;:399–411.411 [PubMed: 15525513]; Supplemental Data

Cheng J, Zhu WL, Dao JJ, Li SQ, Li Y. Relationship between polymorphism of methylenetetrahydrofolate dehydrogenase and congenital heart defect. Biomed Environ Sci 2005;18(1):58–64. [PubMed: 15861780]

Christensen KE, Rohlicek CV, Andelfinger GU, Michaud J, Bigras JL, Richter A, Mackenzie RE, Rozen R. The MTHFD1 p.Arg653Gln variant alters enzyme function and increases risk for congenital heart defects. Hum Mutat. 2008Sep 2 Epub

Czeizel AE, Dudas I. Prevention of the first occurrence of neural-tube defects by periconceptional vitamin supplementation. N Engl J Med 1992;327:1832–5. [PubMed: 1307234]

De Marco P, Merello E, Calevo MG, Mascelli S, Raso A, Cama A, Capra V. Evaluation of a methylenetetrahydrofolate-dehydrogenase 1958G>A polymorphism for neural tube defect risk. J Hum Genet 2005;51(2):98–103. [PubMed: 16315005]

Dong S, Lester L, Johnson LF. Transcriptional control elements and complex initiation pattern of the TATA-less bidirectional human thymidylate synthase promoter. J Cell Biochem 2000;77:50–64. [PubMed: 10679816]

Furness DL, Fenech MF, Khong YT, Romero R, Dekker GA. One-carbon metabolism enzyme polymorphisms and uteroplacental insufficiency. Am J Obstet Gynecol 2008;199(3):276.e1–8. [PubMed: 18771981]

Gong M, Cowan KH, Gudas J, Moscow JA. Isolation and characterization of genomic sequences involved in the regulation of the human reduced folate carrier gene (RFC1). Gene 1999;233:21–31. [PubMed: 10375617]

Hermann S, Berndt KD, Wright AP. How transcriptional activators bind target proteins. J Biol Chem 2001;276:40127–32. [PubMed: 11514548]

Hol FA, van der Put NM, Geurds MP, Heil SG, Trijbels FJ, Hamel BC, Mariman EC, Blom HJ. Molecular genetic analysis of the gene encoding the trifunctional enzyme MTHFD (methylenetetrahydrofolate-dehydrogenase, methenyltetrahydrofolate-cyclohydrolase, formyltetrahydrofolate synthetase) in patients with neural tube defects. Clin Genet 1998;53:119–25. [PubMed: 9611072]

Hoogendoorn B, Coleman SL, Guy CA, Smith K, Bowen T, Buckland PR, O'Donovan MC. Functional analysis of human promoter polymorphisms. Hum Mol Genet 2003;12:2249–54. [PubMed: 12915441]

Howard KM, Muga SJ, Zhang L, Thigpen AE, Appling DR. Characterization of the rat cytoplasmic C1-tetrahydrofolate synthase gene and analysis of its expression in liver regeneration and fetal development. Gene 2003;319:85–97. [PubMed: 14597174]

Hum DW, Bell AW, Rozen R, MacKenzie RE. Primary structure of a human trifunctional enzyme. Isolation of a cDNA encoding methylenetetrahydrofolate dehydrogenase-methenyltetrahydrofolate

cyclohydrolase-formyltetrahydrofolate synthetase. J Biol Chem 1988;263:15946–50. [PubMed: 3053686]

Kempisty B, Sikora J, Lianeri M, Szczepankiewicz A, Czerski P, Hauser J, Jagodzinski PP. MTHFD 1958G>A and MTR 2756A>G polymorphisms are associated with bipolar disorder and schizophrenia. Psychiatr Genet 2007;17(3):177–181. [PubMed: 17417062]

Kirke PN, Molloy AM, Daly LE, Burke H, Weir DG, Scott JM. Maternal plasma folate and vitamin B12 are independent risk factors for neural tube defects. Q J Med 1993;86:703–8. [PubMed: 8265769]

Mao DY, Watson JD, Yan PS, Barsyte-Lovejoy D, Khosravi F, Wong WW, Farnham PJ, Huang TH, Penn LZ. Analysis of Myc bound loci identified by CpG island arrays shows that Max is essential for Myc-dependent repression. Curr Biol 2003;13(10):882–886. [PubMed: 12747840]

McEwan IJ, Dahlman-Wright K, Ford J, Wright AP. Functional interaction of the c-Myc transactivation domain with the TATA binding protein: evidence for an induced fit model of transactivation domain folding. Biochemistry 1996;35:9584–93. [PubMed: 8755740]

MRC. Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. MRC Vitamin Study Research Group. Lancet 1991;338:131–7. [PubMed: 1677062]

Parle-McDermott A, Kirke PN, Mills JL, Molloy AM, Cox C, O'Leary VB, Pangilinan F, Conley M, Cleary L, Brody LC, Scott JM. 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 2006;14:768–72. [PubMed: 16552426]

Parle-McDermott A, Mills JL, Kirke PN, Cox C, Signore CC, Kirke S, Molloy AM, O'Leary VB, Pangilinan FJ, O'Herlihy C, Brody LC, Scott JM. MTHFD1 R653Q polymorphism is a maternal genetic risk factor for severe abruptio placentae. Am J Med Genet A 2005a;132:365–8. [PubMed: 15633187]

Parle-McDermott A, Pangilinan F, Mills JL, Signore CC, Molloy AM, Cotter A, Conley M, Cox C, Kirke PN, Scott JM, Brody LC. A polymorphism in the MTHFD1 gene increases a mother's risk of having an unexplained second trimester pregnancy loss. Mol Hum Reprod 2005b;11:477–80. [PubMed: 16123074]

Pugh BF, Tjian R. Mechanism of transcriptional activation by Sp1: evidence for coactivators. Cell 1990;61:1187–97. [PubMed: 2194667]

Smale. Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. Biochimica et Biophysica Acta 1997;1351:73–88. [PubMed: 9116046]

Smith GK, Mueller WT, Wasserman GF, Taylor WD, Benkovic SJ. Characterization of the enzyme complex involving the folate-requiring enzymes of de novo purine biosynthesis. Biochemistry 1980;19:4313–21. [PubMed: 7417406]

Van der Linden IJ, Afman LA, Heil SG, Blom HJ. Genetic variation in genes of folate metabolism and neural-tube defect risk. Proc Nutr Soc 2006;65(2):204–215. [PubMed: 16672082]

Van der Linden IJ, Heil SG, Kouwenberg IC, den Heijer M, Blom HJ. The methylenetetrahydrofolate dehydrogenase (MTHFD1) 1958G>A variant is not associated with spina bifida risk in the Dutch population. Clin Genet 2007;72(6):599–600. [PubMed: 17894836]

Wang L, Ke Q, Chen W, Wang J, Tan Y, Zhou Y, Hua Z, Ding W, Niu J, Shen J, Zhang Z, Wang X, Xu Y, Shen H. Polymorphisms of MTHFD, plasma homocysteine levels, and risk of gastric cancer in a high-risk Chinese population. Clin Cancer Res 2007;13(8):2526–2532. [PubMed: 17438114]

Weinberg CR, Wilcox AJ, Lie RT. 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 1998;62:969–78. [PubMed: 9529360]

Weinmann AS, Yan PS, Oberley MJ, Huang TH, Farnham PJ. Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. Genes Dev 2002;16:235–44. [PubMed: 11799066]

Wilcox AJ, Weinberg CR, Lie RT. Distinguishing the effects of maternal and offspring genes through studies of “case-parent triads”. Am J Epidemiol 1998;148:893–901. [PubMed: 9801020]

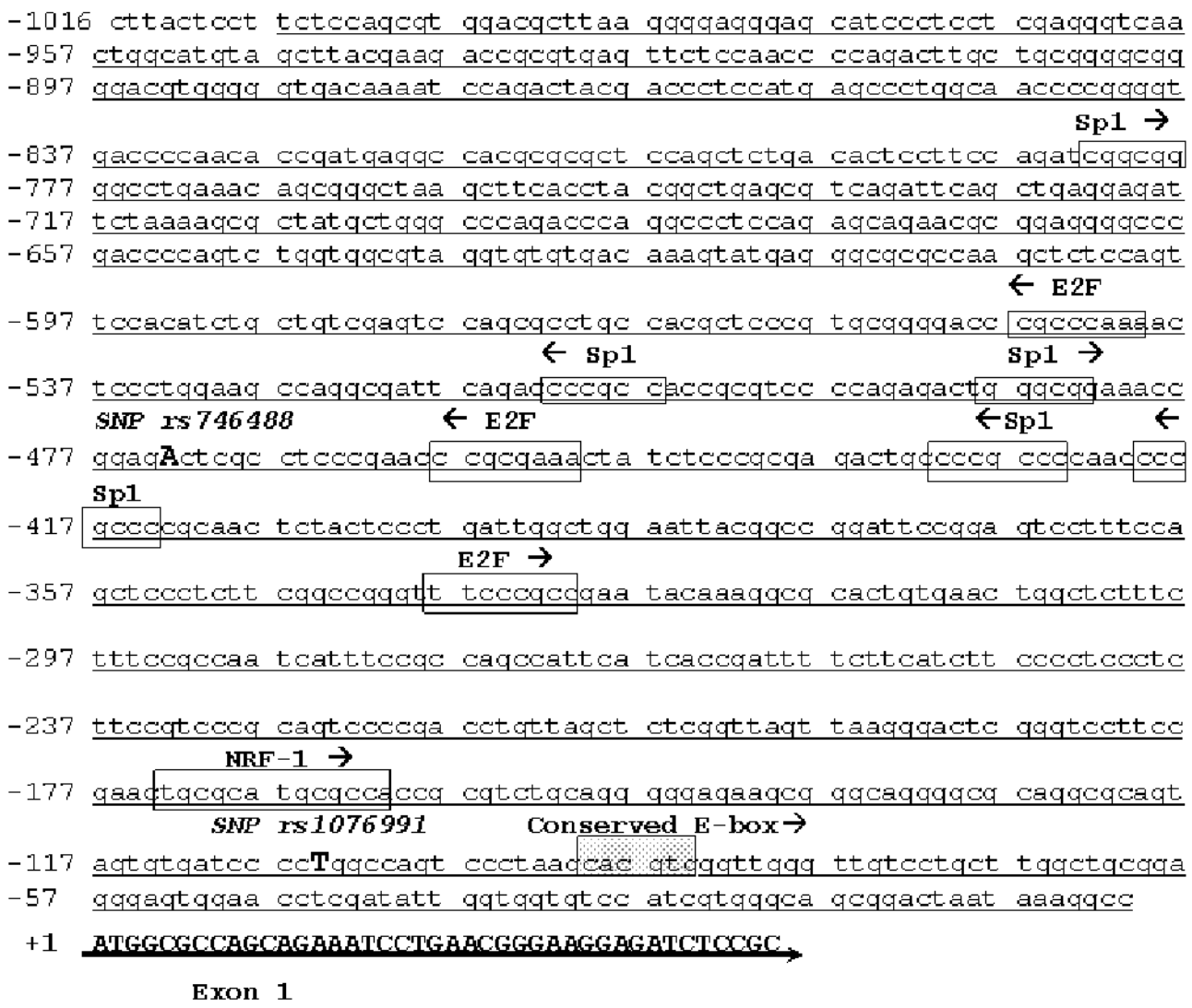

Figure 1. Sequence Analysis of the *MTHFD1* Promoter Region

A CpG island spans the promoter region and is shown underlined. Putative transcription factor binding sites in the *MTHFD1* upstream region were identified using MatInspector and AliBaba2 programs. Those likely to have biological function are shown surrounded by boxes and their direction indicated by arrows. A consensus E-box that is conserved in human, rat and mouse is illustrated by a shaded box. Database SNPs (rs1076991 C>T and rs746488 A>T) are also highlighted in bold capitals. Two additional database SNPs (rs8010584 G>A and rs4243628 G>T) are located further upstream and are not shown here.

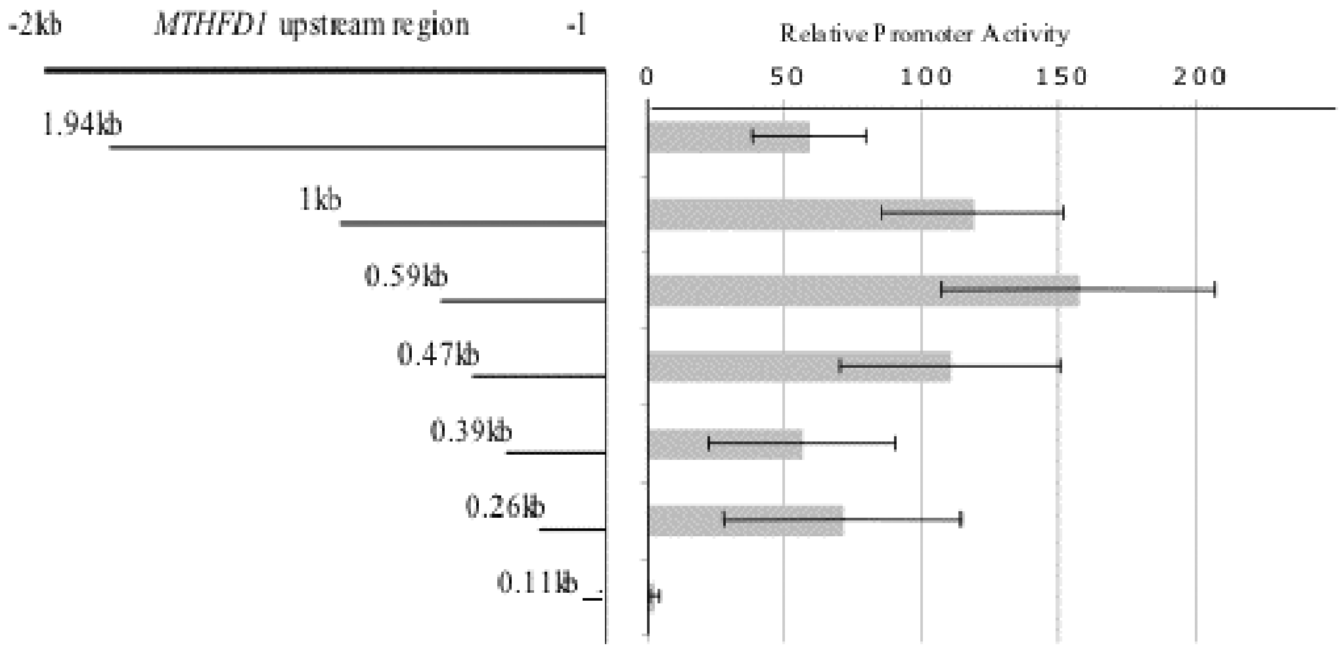

Figure 2. Relative Luciferase Activity of A Series of *MTHFD1* Promoter Constructs

MTHFD1 promoter constructs ranging in size from 0.11kb to 1.94kb were assayed for their ability to drive transcription in a luciferase reporter gene system. The -1 in the *MTHFD1* upstream region refers to the nucleotide immediately upstream of the translational start site. The 0.59kb construct was found to induce the highest levels of luciferase activity, but was similar to the activity of the 1kb and 0.47kb constructs. A lower level of promoter activity was associated with the larger 1.94kb construct while the promoter activity was almost abolished with the 0.11kb construct. Results are expressed relative to empty vector values and as the % mean activity of at least three separate experiments performed in triplicate.

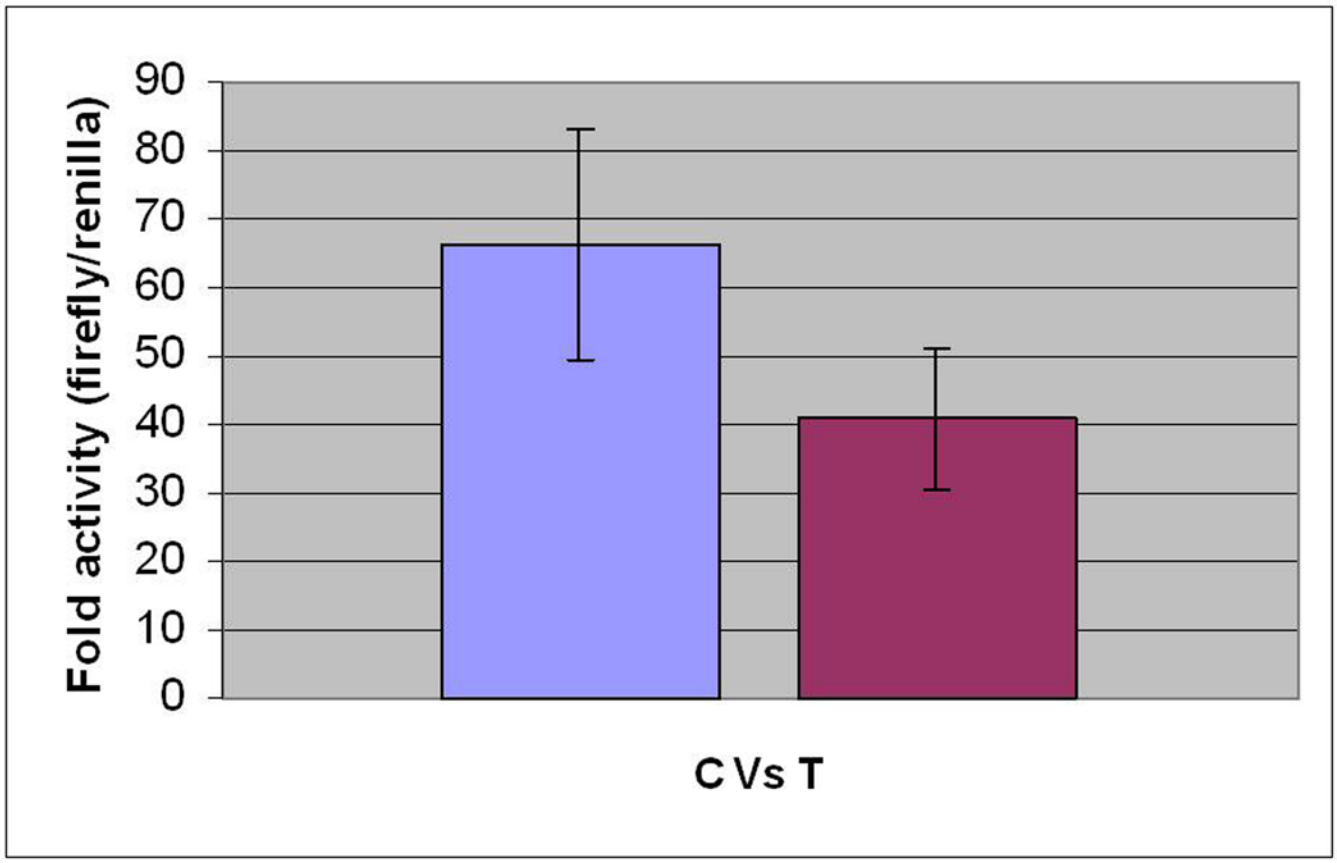

Figure 3. Relative Activity of Different SNP rs1076991 Genotype Promoter Constructs

A total of three luciferase reporter gene assays, each with a fresh preparation of 0.59kb promoter clones, were performed to investigate the effect of SNP rs1076991 on promoter function. No other polymorphic variants were present within this region; therefore, any effect is solely due to SNP rs1076991. The overall result is expressed as the mean and SD of fold luciferase activity (firefly/renilla) detected in each construct above the empty vector (pGL3basic). The 'T' construct exhibited an approximately 1.6 fold (or 62.5%) drop in activity relative to the 'C' construct. This difference was shown to be statistically significant, with $P = 0.04$ using a 2-tailed Student's unpaired T test.

Table 1

SNP rs1076991 Genotype and Allele Frequencies in NTD Triad and Control Groups

	NTD Triad Group			Control Group
Genotype	**Cases (n = 509)**	**Mothers (n = 485)**	**Fathers (n = 439)**	**Controls (n = 966)**
CC	97 (0.19)[a]	98 (0.20)	66 (0.15)	198 (0.20)
CT	250 (0.49)	225 (0.46)	234 (0.53)	468 (0.48)
TT	162 (0.32)	162 (0.33)	139 (0.32)	300 (0.31)
TT vs. CT/CC OR (LL – UL)[b]	1.04 (0.82 – 1.31); $P = 0.37$	1.11 (0.88 – 1.41); $P = 0.76$	1.03 (0.81 – 1.31); $P = 0.82$	
Allele				
C	444 (0.44)	421 (0.43)	366 (0.42)	864 (0.45)
T	574 (0.56)	549 (0.57)	512 (0.58)	1068 (0.55)
T vs. C	$\chi^2 = 0.33$; $P = 0.57$	$\chi^2 = 0.45$; $P = 0.50$	$\chi^2 = 2.26$; $P = 0.13$	

[a] Population frequencies are shown in parentheses. Values might not add to 1 due to rounding.

[b] OR (Odds Ratio); LL (Lower Limit); UL (Upper Limit).

Table 2

Interactive Effect of *MTHFD1* SNP rs1076991 C→T with SNP rs2236225 G→A in NTD triad groups and controls

	NTD Triad Group			Control Group
Genotype SNP rs1076991 C→T/ SNP rs2236225 G→A	**Cases**	**Mothers**	**Fathers**	**Controls**
CC/GG	41 (0.08) [a]	37 (0.08)	24 (0.06)	25 (0.06)
CC/GA*	43 (0.09)	46 (0.1)	23 (0.05)	60 (0.13)
CC/AA	11 (0.02)	12 (0.02)	11 (0.02)	11 (0.02)
CT/GG	71 (0.14)	55 (0.12)	83 (0.2)	59 (0.13)
CT/GA	142 (0.28)	107 (0.23)	104 (0.25)	109 (0.24)
CT/AA	36 (0.07)	58 (0.12)	38 (0.09)	45 (0.1)
TT/GG	32 (0.06)	36 (0.08)	30 (0.07)	39 (0.09)
TT/GA	69 (0.14)	70 (0.15)	72 (0.17)	75 (0.17)
TT/AA*	**55 (0.11)[b]**	**51 (0.11)[b]**	34 (0.08)	**23 (0.05)**
Maternal Interaction		**Genotype[c]**	**Allele[d]**	**Triads[e]**
		P = 0.14	P = 0.01	*P* = 0.10
Case Interaction		**Genotype[c]**	**Allele[d]**	**Triads[e]**
		P = 0.001	*P* = 0.001	*P* = 0.36

[a] Genotype frequencies are shown in parentheses. Values might not add to 1 due to rounding.

[b] The most significant differences in genotype combinations in Cases or Mothers versus controls are shown in bold.

[c] Genotype interactions were tested by logistic regression in Mothers or Cases versus Controls.

[d] Allele interactions were tested by logistic regression in Mothers or Cases versus Controls.

[e] Allele interactions were tested in a log linear model using triads.

* Odds Ratios were calculated using the combined genotype of lowest risk as the reference i.e., CC/GA in Cases or Mothers versus Controls. The highest risk was the case TT/AA genotype with an OR 3.34.