Published in final edited form as:

Birth Defects Res A Clin Mol Teratol. 2012 January ; 94(1): 52–56. doi:10.1002/bdra.22855.

Genetic Studies of the Cystathionine Beta-Synthase Gene and Myelomeningocele

Melissa M. Tilley, MD¹, Hope Northrup, MD², and Kit Sing Au, PhD²

¹Division of Neonatal-Perinatal Medicine, Department of Pediatrics The University of Texas Medical School at Houston, Houston, TX 77030, USA

²Division of Medical Genetics, Department of Pediatrics, The University of Texas Medical School at Houston, Houston, TX 77030, USA

Abstract

BACKGROUND—Among infants born with spina bifida, the most common defect is myelomeningocele (MM). The prevention of MM by maternal periconceptual folic acid (FA) supplementation has been extensively studied. The protective effect provided by FA suggests that the genes involved in folate metabolism, such as cystathionine beta-synthase (*CBS*), warrant further investigation.

METHODS—This study sequenced the DNA from 96 patients with MM to identify novel potential disease causing variants across the 17 exons of the *CBS* gene. The frequencies of known SNPs were identified and sequences that differed from the reference sequences were considered novel variants. Statistical analysis was performed using 2-sided Fisher's exact test to compare frequencies of SNPs between groups of patients and the known population frequencies.

RESULTS—We found a new variant in Exon 3 in one patient that result in a G/A change subsequently encoding a stop codon. Additionally, we found a new variant in the 3'-UTR of exon 17. Allele frequencies for ten known SNPs were determined: rs234706, rs72058776, rs1801181, rs6582281, rs71872941, rs12613, rs706208, rs706209, rs73906420, and rs9982921. Of the remaining 48 known SNPs, all tested DNAs were homozygous for the major allele.

CONCLUSION—We identified a previously undescribed variant in Exon 3 that encodes a stop codon, thus halting downstream translation of the *CBS* protein. According to the Human Splice Finder, the 3'-UTR variant found in Exon 17 is predicted to abolish the recognition sites for two splice binding factors, SRp40 and SF2/ASF. The functional significance of the 3'-UTR mutation needs to be investigated.

Keywords

neural tube defects (NTDs); spina bifida (SB); myelomeningocele (MM); folic acid (FA); *cystathionine-beta synthase (CBS)* gene; single nucleotide polymorphisms (SNPs)

INTRODUCTION

With an occurrence of 2/10,000 births, neural tube defects (NTDs) are one of the most common birth defects in the US (CDC, 2009). The process of neural tube formation occurs during the third and fourth week after fertilization. A disruption in this complex process can cause failure of closure resulting in NTDs, including anencephaly and spina bifida (SB)

(Sadler, 2005). Myelomeningocele (MM), an open defect, is the most common, non-lethal form of SB. Although anencephalic infants die in utero or shortly after birth, most infants born with SB survive as a result of extensive medical and surgical care. Medical problems occur due to the neurologic defect, its repair, or as a sequela of the neurologic deficit (Botto, et al., 1999). The estimated lifetime (up to 65 years of age) total cost of medical care for an individual born with SB is ~\$829,000 in 2003 dollars (Ouyang et al., 2007).

The etiology of NTDs is multifactorial, including genetic and environmental factors. In the US, Hispanic women have the highest prevalence of NTD-affected pregnancies compared with non-Hispanic whites and non-Hispanic Blacks (Williams et al., 2005; Boulet et al., 2008). The reason for this disparity remains unknown, although diet and genetic factors have both been suggested as possibly contributing to the etiology of SB.

Numerous studies have shown reduction in NTDs in the US since the implementation of mandatory folic acid fortification of enriched grain products in 1998 (Honein et al., 2001; Mathews et al., 2002). Unfortunately, there is not yet a complete understanding of the relationship between the molecular mechanism regulating folate status and NTD risk (Afman, et al., 2003). Evidence of altered homocysteine metabolism has been implicated as a risk factor for NTDs. Folate is an important element in homocysteine metabolism. Folate, via remethylation of homocysteine (Hcy) to methionine, lowers Hcy levels in blood. Another major pathway of homocysteine metabolism is the transsulfuration pathway that results in Hcy being irreversibly condensed with serine to cystathionine by the enzyme cystathionine beta-synthase (CBS) (Afman, et al., 2003). Mutations of the *CBS* gene, such as the 844ins68 allele, have been described (Sebastio et al., 1995). There have been conflicting studies regarding the potential interaction among alleles of different folate-related genes, such as the 677C-T allele of the methylene tetrahydrofolate reductase (*MTHFR*) gene and the 844ins68 allele of *CBS*, and their combined contribution to risk of SB (Ramsbottom et al., 1997; Speer et al., 1999; de Franchis et al., 2002; Houcher et al., 2009).

We recently reported that the *CBS* gene was implicated in MM risk. We found rs5742905 (c. 833T>C) that results in an amino acid change from isoleucine to threonine and resides in the 844ins68 region of *CBS*, to be significantly associated to MM risk (Martinez et al., 2009). As a follow-up study, we performed deep sequencing of the exons of the *CBS* gene among a selected population of subjects with MM to identify known SNPs as well as novel potential disease causing gene variants.

MATERIALS AND METHODS

Study Population

We selected 96 MM subjects (48 North American Caucasians and 48 Mexican Americans) from a cohort enrolled in our MM studies between 1996 and 2006 at three different sites (Houston, TX, Toronto, ON, Canada, and Los Angeles, CA). The subjects were born between 1968 and 1993, prior to US mandated folate fortification of food crops in January 1998. The project was approved by the University of Texas Health Science Center at Houston Institutional Review Board.

Sample Preparation

Blood and/or saliva samples were obtained from the affected subjects as well as their unaffected parents. Using the Puregene DNA kit (Gentra Systems Inc., Minneapolis, MN), genomic DNA was extracted from blood lymphocytes; the Oragene DNA collection kit (DNA Genotek, Inc., Ottawa, ON, Canada) was used to extract DNA from saliva. DNA concentrations were measured using Nanodrop Spectrophotometer (Thermo Fisher

Scientific Inc., Waltham, MA) and dilutions were made to 10ng/uL for all experiments in the study.

DNA Sequencing

Forward and reverse PCR primers were designed for the 17 exons of the *CBS* gene (NM 000071.2 and NM 001178008.1), including 50-100 base pairs of intron sequences flanking each exon. After PCR amplification, gel electrophoresis was utilized for verification of exon size. The amplified product underwent Exonuclease I and Shrimp Alkaline Phosphatase treatment to remove excess primers and nucleotides. Digested product was subjected to Sanger sequencing utilizing the Big Dye Terminator Protocol (Applied Biosystems Inc, Foster City, CA; ABI). Sequencing products were separated and recorded using the ABI 3100 Genetic Analyzer.

Genotyping

To genotype the 68bp insertion polymorphism (rs72058776, c.844ins68bp), PCR primer pairs (forward 5'-CACCACCCACAGGCAGATC-3' and reverse 5'-CCAGTGTGAGGGTGAGTTAC-3') flanking exon 10 of *CBS* and the 86bp insertion polymorphism preceding exon 10 were used. The PCR products were separated using a 1.8% agarose gel in Tris-borate EDTA buffer (pH8.0).

Statistical Analysis

The sequences were compared to the reference sequences of *CBS* from GenBank and to one another both manually and with BioLign 4.0.6.2 gene alignment software to identify known SNPs and novel mutations. Novel variants were verified using reverse sequencing and sequencing of parental DNA when available to verify variants as *de novo*.

Frequencies of SNPs were compared based on subject ethnicity and to known reference population frequencies published in public SNP databases (dbSNP and HapMap SNPs). Statistical analysis included 2-sided Fisher's exact test to compare frequencies of alleles for each SNP between groups of subjects and the known population frequencies. Two-tailed p values <0.05 were considered statistically significant.

RESULTS

We identified two new variants, not previously reported in the dbSNP for Caucasians or Mexican Americans. One variant is located within exon 3 at c.162G>A (human *CBS* RefSeq NM_000071) resulting in a stop codon (p.Trp54stop) (Fig.1a). Additional testing revealed the same variant in the subject's mother. Paternal DNA was unavailable for testing. The other variant located in the 3'-UTR in exon 17 (c.*540G>A) was found in three subjects (Fig.1b).

In the 17 exon regions of the *CBS* gene sequenced, the most recent dbSNP (snp132) includes 58 SNPs with 29 having a reported minor allele frequency. Among our MM subjects, we observed the minor allele for 10 of the 58 known SNPs but only the reference allele for the remaining 48.

Among those 58 SNPs, 41 do not have allele frequencies reported for the reference populations and we did not observe variation in our MM subjects. Seventeen SNPs have allele frequencies available for the reference population that were compared to those observed in our tested population using Fisher's exact test (Table 1). We found statistically significantly higher frequencies of reference alleles among three SNPs (rs12329764, rs6586282, and rs1051319) for Caucasian subjects compared to the reference CEU

population of dbSNP. We found three SNPs (rs12329764, rs6586282, and rs1005584) with significantly higher reference alleles and two SNPs (rs706208 and rs706209) with significantly higher rare alleles compared to the reference HapMap MEX population (indicated in bold Table 1).

Among those known SNPs in the *CBS* gene, we identified one MM subjects with a homozygous 844ins68bp insertion in exon 10 (rs72058776) (Fig.1c) and MM subjects who were heterozygous for the insertion. The PCR product without the insertion is shown as a 318bp band compared to the 386bp band with the 68bp insertion. Overall, the frequency of the 844ins86bp for Caucasian MM subjects was 0.13 and for Mexican American MM subjects was 0.1. We identified 20 MM subjects with a 5bp deletion (rs71872941) in the 3'UTR in Exon 17 (no reference population frequencies available).

DISCUSSION

We utilized DNA sequencing to examine the 17 exons of the *CBS* gene for 96 MM subjects. We identified two new variants and quantified allele frequencies of 58 known SNPs (snp 132). Completion of the 1000 Genomes Project will allow us to identify whether these new variants are present in the general population. Once these variants are validated, future population-based studies will provide allele frequency information to evaluate potential disease association of the variants.

A novel variant (c.162G>A) in exon 3 was identified in one Caucasian MM subject. The unaffected mother also has the variant. This mutation causes the premature formation of a stop codon at amino acid 54, predicted to stop translation of the CBS protein downstream from the change. The subject and mother are carriers for the disease classic homocystinuria. Classic homocystinuria is a rare autosomal recessive disease with a carrier frequency ranging from ~1/250 to 1/1000 (Skovby et al., 2010). There is no described phenotype for individuals who are carriers. Interestingly, individuals of Irish descent are reported to have a high homocystinuria carrier frequency and coincidentally have a high NTD incidence. We do not know the significance of the novel nonsense variant (p.Trp52stop) in exon 3 to MM observed in our study. The same subject also carries the 844ins68bp variation in exon 10. It is unknown whether the two variants are *cis* or *trans* in our subject. Unfortunately, without the paternal DNA, we cannot address the question by reconstructing haplotypes.

The other novel variant (c.*540G>A) in the 3'-UTR of exon 17 was noted in three affected subjects, two Caucasian and one Hispanic. A most recent data release of the early phase 1000 Genome Project included the c.*540G>A variant, referred to as rs111386779, as detected in one West African YRI sample of 118 chromosomes examined. The frequencies for the reference population CEU or MEX samples are not available. The mutation is predicted by the Human Splicing Finder online tool (HSF; <http://www.umd.be/HSF/>) to abolish the recognition sites for two splice binding factors, SRp40 and SF2/ASF. The functional significance of this mutation needs to be defined through functional studies.

We compared allele frequencies of SNPs found in MM subjects to those of ethnically-matched reference populations. In our previous study, we demonstrated no significant differences in population stratification between the MM subjects and the reference population samples used in the study (Au et al., 2008). Comparing allele frequency between MM subjects to ethnically matched reference populations suggested significant differences for three SNPs (rs123329764, rs6586282 and rs1051319) in Caucasian MM subjects and five SNPs (rs123329764, rs6586282, rs1005584, rs706208 and rs706209) for Mexican American MM subjects (Table 1). The rare alleles for two SNPs (rs706208 and rs706209 in 3'-UTR) are significantly higher in Mexican American MM subjects than HapMap MEX

reference population suggesting potential risk for MM. Similar higher rare allele frequency for the two SNPs was shown in Caucasian MM subjects but did not reach a statistically significant level. The rare allele T of rs706209 is predicted to eliminate SRP40 binding using the HSF online tool. All SNPs that showed significance appear to be in linkage disequilibrium (LD) in Caucasian individuals tested in HapMap with $D' \geq 0.8$. Similar LD has been reported in Mexican individuals reported in HapMap except between rs706209 and rs6586282. Other than for rs706208 and rs706209, the correlation coefficient r^2 is less than 0.8. Functional significance for the observation remains to be determined. We are aware that the observed results in this study are based on a small sample size and that some of the SNPs have a low minor allele frequency, thus limiting the power of the study to compare genotype distributions between the patient and reference populations. Our finding suggests further investigation of the observed significant SNPs for association with MM is needed.

Additionally, we observed a higher than reference frequency for 844ins68bp (rs72058776) among MM subjects but the observed differences did not reach significance. In the public SNP databases, there is no allele frequency for 844ins68bp (rs72058776). However, an NHANEIII study conducted at the Center for Disease Control (CDC) reported that the frequency of 844ins68bp for Caucasian Americans is 0.08 and for Mexican Americans is 0.063 (Chang et al., 2010). The rs5742905 (c.833T>C, p.I278T) has been shown to be evolutionarily conserved and in *cis* with 844ins68bp (Romano et al., 2008). We previously demonstrated association to MM for rs5742905 (Martinez et al., 2009). Unfortunately, we could not clearly identify the 833T variant that superimposed with the inserted 68 bases in a heterozygous sequencing design. Previous genotyping data showed the presence of 833T with subjects who have the 844ins68bp allele. We were not able to observe significant differences for the 844ins68bp compared to reference population in this study possibly because of the limited sample size.

As in many SNP studies, we observed differences in allele frequencies between subjects of different ethnicity. For example, in our population, the rs1801181 minor T allele and the rs234706 minor T allele were more frequent in Caucasians than Mexicans (34% vs. 17%, $p=0.008$ and 39% vs. 24%, $p=0.044$, respectively), as can be anticipated due to their expected allele frequency from the SNP database. Our group is the first to describe allele frequencies for three known SNPs (rs72058776, rs71872947, and rs73906420) that do not have reported frequencies.

In summary, this study provides a much more thorough examination of the *CBS* gene in MM subjects. We identified a novel nonsense variant (c.162G>A, p.Trp52Stop) in one MM subject and a novel 3'-UTR variant (c.*540G>A) in three other MM subjects. A higher frequency of a 68bp insertion polymorphism (rs72058776) than reference population was observed although the finding did not reach statistical significance. Finally, we found six SNPs with demonstrated significance for association to MM. Our study provides additional evidence that the *CBS* gene may play a role in MM susceptibility. Further evaluation of *CBS* gene for its association with MM is warranted.

Acknowledgments

We would like to thank the patients and their families for their participation in this study. We are grateful to Phong X. Tran for technical assistance and data management and to Michelle R. O'Byrne, PhD, for technical assistance.

This work was supported by grants from the National Institutes of Health (P01 HD35946-06A2) and Shriners Hospital for Children (project 8580) to H.N. and K.S.A.

REFERENCES

- Afman LA, Lievers KJA, Kluijtmans LAJ, Trijbels FJM, Blom HJ. Gene–gene interaction between the cystathionine β -synthase 31 base pair variable number of tandem repeats and the methylenetetrahydrofolate reductase 677C>T polymorphism on homocysteine levels and risk for neural tube defects. *Mol Genet Metab*. 2003; 78:211–215. [PubMed: 12649066]
- Au KS, Tran PX, Tsai CC, O'Byrne MR, Lin JI, Morrison AC, Hampson AW, Cirino P, Fletcher JM, Ostermaier KK, Tyerman GH, Doebel S, Northrup H. Characteristics of a spina bifida population including North American Caucasians and Hispanic individuals. *Birth Defects Res A: Clin Mol Teratol*. 2008; 82:692–700. [PubMed: 18937358]
- Botto LD, Moore CA, Khoury MJ, Erickson JD. Neural-tube defects. *N Engl J Med*. 1999; 341:1509–1519. [PubMed: 10559453]
- Boulet SL, Yang Q, Mai C, Kirby RS, Collins JS, Robbins JM, Meyer R, Canfield MA, Mulinare J. Trends in the postfortification prevalence of spina bifida and anencephaly in the United States. *Birth Defects Res A: Clin Mol Teratol*. 2008; 82:527–532. [PubMed: 18481813]
- Centers for Disease Control and Prevention (CDC). Racial/ethnic differences in the birth prevalence of spina bifida-US, 1999–2005. *MMWR Morb Mortal Wkly Rep*. 2009; 57:1409–1413. [PubMed: 19129744]
- Chang M, Lindegren ML, Butler MA, Chanock SJ, Dowling NF, Gallagher M, Moonesinghe R, Moore CA, Ned RM, Reichler MR, Sanders CL, Welch R, Yesupriya A, Khoury MJ. Prevalence in the United States of selected candidate gene variants. Third national health and nutrition examination survey 1991–1994. *Amer J of Epid*. 2010; 169:54–66.
- de Franchis R, Botto LD, Sebastio G, Ricci R, Iolascon A, Capra V, Andria G, Mastroiacovo P. Spina bifida and folate-related genes: A study of gene-gene interactions. *Genet Med*. 2002; 4:126–130. [PubMed: 12180146]
- Honein MA, Paulozzi LJ, Mathews TJ, Erickson JD, Wong LY. Impact of folic acid fortification of the US food supply on the occurrence of neural tube defects. *JAMA*. 2001; 285:2981–2986. [PubMed: 11410096]
- Houcher B, Bourouba R, Djabi F, Yilmaz E, Egin Y, Akar N. Polymorphisms of 5,10-methylenetetrahydrofolate reductase and cystathionine beta-synthase genes as a risk factor for neural tube defects in Setif, Algeria. *Pediatr Neurosurg*. 2009; 45:472–477. [PubMed: 20160465]
- Martinez CA, Northrup H, Lin JI, Morrison AC, Fletcher JM, Tyerman GH, Au KS. Genetic association study of putative functional single nucleotide polymorphisms of genes in folate metabolism and spina bifida. *Am J Obstet Gynecol*. 2009; 201:394e1–e11. [PubMed: 19683694]
- Mathews TJ, Honein MA, Erickson JD. Spina bifida and anencephaly prevalence--United States, 1991–2001. *MMWR Recomm Rep*. 2002; 51:9–11. [PubMed: 12353510]
- Ouyang L, Grosse SD, Armour BS, Waitzman NJ. Health care expenditures of children and adults with spina bifida in a privately insured U.S. population. *Birth Defects Res A Clin Mol Teratol*. 2007; 79(7):552–8. PMID: 17335056. [PubMed: 17335056]
- Ramsbottom D, Scott JM, Molloy A, Weir DG, Kirke PN, Mills JL, Gallagher PM, Whitehead AS. Are common mutations of cystathionine beta-synthase involved in the aetiology of neural tube defects? *Clin Genet*. 1997; 51:39–42. [PubMed: 9084933]
- Sadler TW. Embryology of neural tube development. *Am J Med Genet C Semin Med Genet*. 2005; 135C:2–8. [PubMed: 15806586]
- Sebastio G, Sperandeo MP, Panico M, de Franchis R, Kraus JP, Andria G. The molecular basis of homocystinuria due to cystathionine beta-synthase deficiency in Italian families, and report of four novel mutations. *Am J Hum Genet*. 1995; 56:1324–1333. [PubMed: 7762555]
- Skovby F, Gaustadness M, Mudd SH. A revisit to the natural history of homocysteinuria due to cystathionine beta synthase deficiency. *Mol Genet Met*. 2010; 99:1–3.
- Speer M, Nye J, McLone D, Worley G, Melvin E, Viles K, Franklin A, Drake C, Mackey J, George T, NTD Collaborative Group. Possible interaction of genotypes at cystathionine beta-synthase and methylenetetrahydrofolate reductase (MTHFR) in neural tube defects. *Clin Genet*. 1999; 56:142–144. [PubMed: 10517251]

Williams LJ, Rasmussen SA, Flores A, Kirby RS, Edmonds LD. Decline in the prevalence of spina bifida and anencephaly by race/ethnicity: 1995–2002. *Pediatrics*. 2005; 116:580–586. [PubMed: 16140696]

Figure 1.

Three variants of *CBS* gene identified in the study. 1a shows the sequences of a novel nonsense variant (c.162G>A; p.Trp54Stop) identified in exon 3 of an MM subject. 1b shows the sequences a novel 3'-UTR variant (c.*540G>A) in exon 17 of another unrelated MM subject. 1c shows PCR products of a 68bp insertion polymorphism (rs72058776) preceding exon 10 in a panel of 16 MM subjects. The PCR product without insertion is shown as a 318bp band compared to the 386bp band with the 68bp insertion. Samples in lanes 6 and 12 showed the 68bp insertion allele (arrow). Lanes M are 100 bp DNA markers from New England Biolabs.

Table 1

Allele count for SNPs with variability from dbSNP among MM subjects

SNP	Func.	alleles	Caucasian American MM ₁	Mexican American MM	All MM	Reference population	
						CEU	MEX
rs9982921	Int 1	G/A	98/0	92/2	190/2	118/0	NA
rs73906420	T191	G/A	96/2	93/1	189/3	NA	NA
rs2298758	N212	G/A	98/0	94/0	192/0	118/0	NA
rs234706	Y233	C/T	60/38	71/23	131/61	149/72	85/15
rs12329764	Int9	G/A	98/0**	94/0*	192/0	107/13	109/7
rs72058776	Int 9	-/ins	87/11	85/9	172/20	4834/426 [‡]	3885/261 [‡]
rs2228298	T313	G/A	98/0	94/0	192/0	118/0	NA
rs1801181	A360	C/T	64/34	78/16	142/50	83/37	NA
rs11700812	R369H	C/T	98/0	94/0	192/0	212/4	113/1
rs6586282	Int 14	C/T	98/0**	94/0**	192/0	97/23	99/15
rs6586281	Int 14	C/T	80/18	75/17	155/37	97/23	NA
rs1005584	Int 14	G/A	98/0	94/0*	192/0	216/6	108/6
rs71872941	3'-UTR	-/del	74/24	78/16	152/40	NA	NA
rs1051319	3'-UTR	C/G	98/0**	94/0	192/0	102/12	NA
rs12613	3'-UTR	G/A	88/8	87/7	175/15	108/12	NA
rs706208	3'-UTR	T/C	60/36	59/35*	119/71	69/51	88/26
rs706209	3'-UTR	C/T	61/35	59/35*	120/70	67/53	87/27

Notes: Func. – functional significance; CEU-reference Caucasian population from the Utah Centre d'Étude du Polymorphisme Humain (CEPH) pedigrees; MEX-HapMap reference population of LA Mexican Americans. ins – 68bp insertion, del – deletion of GCCAG. NA – data not available. Fisher's exact test compared MM pts by ethnicity with HAPMAP reference expected frequencies. Significant result shows with

* for p<0.05 and

** for p<0.005.

[‡] indicates data from Chang et al. 2008.