npg

ORIGINAL ARTICLE

A functional variant in the cystathionine β-synthase gene promoter significantly reduces congenital heart disease susceptibility in a Han Chinese population

Jian-Yuan Zhao[1, 7, *], Xue-Yan Yang[1, *], Kai-Hu Shi[2, *], Shu-Na Sun[3], Jia Hou[3], Zhi-Zhou Ye[1], Jue Wang[1], Wen-Yuan Duan[4], Bin Qiao[4], Yi-Jiang Chen[5], Hong-Bing Shen[6], Guo-Ying Huang[3], Li Jin[1, 8], Hong-Yan Wang[1, 3, 8]

[1]The State Key Laboratory of Genetic Engineering and MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, 220 Handan Road, Shanghai 200433, China; [2]The Second Hospital of Anhui Medical University, 678 Furong Road, Hefei, Anhui 230601, China; [3]Children's Hospital of Fudan University, 399 Wanyuan Road, Shanghai 201102, China; [4]Institute of Cardiovascular Disease, General Hospital of Jinan Military Region, 8 Lashan Road, Jinan, Shandong 250022, China; [5]Department of Thoracic and Cardiovascular Surgery, The First Affiliated Hospital of Nanjing Medical University, 140 Hanzhong Road, Nanjing, Jiangsu 210029, China; [6]Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, 140 Hanzhong Road, Nanjing, Jiangsu 210029, China; [7]Institute of Sports Science and Technology, Administration of Sports of Anhui Province, 97 Wuhu Road, Hefei, Anhui 230001, China; [8]The Institutes of Biomedical Sciences, Fudan University, 138 Yixueyuan Road, Shanghai 200032, China

Homocysteine is an independent risk factor for various cardiovascular diseases. There are two ways to remove homocysteine from embryonic cardiac cells: remethylation to form methionine or transsulfuration to form cysteine. Cystathionine β-synthase (CBS) catalyzes the first step of homocysteine transsulfuration as a rate-limiting enzyme. In this study, we identified a functional variant −4673C>G (rs2850144) in the *CBS* gene promoter region that significantly reduces the susceptibility to congenital heart disease (CHD) in a Han Chinese population consisting of 2 340 CHD patients and 2 270 controls. Individuals carrying the heterozygous CG and homozygous GG genotypes had a 15% (odds ratio (OR) = 0.85, 95% confidence interval (CI) = 0.75-0.96, $P = 0.011$) and 40% (OR = 0.60, 95% CI = 0.49-0.73, $P = 1.78 \times 10^{-7}$) reduced risk to develop CHD than the wild-type CC genotype carriers in the combined samples, respectively. Additional stratified analyses demonstrated that *CBS* −4673C>G is significantly related to septation defects and conotruncal defects. *In vivo* detection of *CBS* mRNA levels in human cardiac tissues and *in vitro* luciferase assays consistently showed that the minor G allele significantly increased *CBS* transcription. A functional analysis revealed that both the attenuated transcription suppressor SP1 binding affinity and the *CBS* promoter hypomethylation specifically linked with the minor G allele contributed to the remarkably upregulated *CBS* expression. Consequently, the carriers with genetically increased *CBS* expression would benefit from the protection due to the low homocysteine levels maintained by CBS in certain cells during the critical heart development stages. These results shed light on unexpected role of CBS and highlight the importance of homocysteine removal in cardiac development.

Keywords: congenital heart disease; cystathionine β-synthase; non-coding variant; homocysteine

Cell Research (2013) **23**:242-253. doi:10.1038/cr.2012.135; published online 18 September 2012

Introduction

Congenital heart disease (CHD) is the most common birth defect, affecting 9.1 per 1 000 live births worldwide [1]. The prevalence of CHD is increasing in China. The etiology of CHD is complex, involving both genetic and environmental factors. Decades of epidemiologic studies have suggested that pre-conception administration of folic acid protects newborns from various congenital abnormalities, including a 40%-60% decreased risk of

CHD [2, 3]. Thus, association studies on genetic variants in the folate metabolism pathway genes and the risk of CHD flourished, especially the intensively studied *MTH-FR* C677T. However, these studies yielded controversial results [4].

As a metabolite in the folic acid pathway, homocysteine can be reversely regulated by folate supplementation and is an independent risk factor for CHD. Increased maternal homocysteine levels are associated with an increased risk of CHD in the offspring [5, 6]. Mouse and chicken embryo studies demonstrated that exposure to exogenous homocysteine during the critical period of cardiac development increased the incidence of CHD, especially septation defects [7]. Our previous study identified a functional variant in the first intron of methionine synthase reductase (*MTRR*) gene, c.56+781A>C that significantly increased the CHD risk in a large-scale, case-control study in a Han Chinese population [8]. MTRR is required for the activation of methionine synthase (MTR), which catalyzes homocysteine remethylation to methionine. This result strongly supported the hypothesis that homocysteine removal was crucial to an embryo's cardiac development because MTR/MTRR activity is essential for the adequate remethylation of homocysteine. In addition to being remethylated to methionine, homocysteine in embryo cells can be removed through the transsulfuration pathway to form cysteine. Cystathionine β-synthase (CBS, EC 4.2.1.22) catalyzes the first irreversible step from homocysteine to cystathionine in the transsulfuration pathway.

The *CBS* gene is located on the human chromosome 21q22.3. *CBS* deficiency is the most common cause of classical homocystinuria (HCU, OMIM236200), an inherited autosomal recessive metabolic disease [9]. Although CBS is presented at a low level in the fetus compared to adults, its expression is concentrated in the neural and cardiac tissues, especially in the endocardium cells, which implies a potential function of CBS in embryo cardiac development [10, 11]. The few association studies that focused on the *CBS* coding region variants and CHD risk obtained negative results [12, 13].

In this study, the non-coding variants in the *CBS* gene were investigated in three independent case-control studies of 2 340 CHD patients and 2 270 controls from a Han Chinese population. We identified a *CBS* promoter variant, −4673C>G (rs2850144, NC_000021.8:g44496976C>G), which increases *CBS* gene expression and is significantly associated with reduced CHD risk in all three case-control pairs and in the combined dataset.

Results

The CBS 5′ regulatory variant −4673C>G significantly reduces the risk of CHD

The *CBS* gene spans over 30 kb and consists of 23 exons [14]. The human *CBS* gene encodes different mRNAs, a result of the use of five alternative non-coding exons (designated -1a to -1e) and a constant exon 0. Transcripts containing exons -1a or -1b appear to be the most abundant and are found in an assortment of adult and fetal tissues. However, usage of exons -1c, -1d, and -1e appears rare [14, 15]. In this study, the detection region covers the *CBS* 5′ regulatory region upstream of the first exon-1a (−6 251 to −3 623, numbered from ATG), together with the entire 3′ UTR fragment (Figure 1A).

In total, 3 common polymorphisms in the *CBS* gene regulatory region were identified with the minor allele frequency > 0.1, including rs2850144 (NC_000021.8: g44496976C>G, −4673C>G at the −4673 of ATG) in the promoter and rs1051319, rs706208 in the 3′ UTR region.

In the first stage of the association study, we genotyped the 3 SNPs in 270 cases versus 552 controls in Shanghai group. The genotype distribution of only one SNP (−4673C>G, rs2850144) was significantly different between the case and control subjects. The minor allele G at site −4673 was associated with a reduced risk of CHD. Similar results were observed in the validation study of 259 cases versus 324 controls in the Shandong group (Supplementary information, Table S1). The variant −4673C>G in the promoter of *CBS* has not been reported to be associated with any diseases so far. To further validate the protective effect of allele G at site −4673, we performed a second-stage association study by expanding the sample size to 602 cases and 660 controls in the Shanghai group and 735 cases and 564 controls in the Shandong Group. Additionally, 1 003 cases and 1 046 controls sampled from Jiangsu were added for validation.

We found a consistent protection effect against CHD from the −4673G allele compared to the C allele in both the Shanghai group (per allele adjusted OR = 0.75, 95% Cl = 0.63-0.87, *P* = 0.001) and the Shandong group (per allele adjusted OR = 0.80, 95% Cl = 0.68-0.95, *P* = 0.009). In the independently validated Jiangsu group, results similar to those in the Shandong and Shanghai groups were observed (per allele adjusted OR = 0.82, 95% Cl = 0.72-0.93, *P* = 0.002) (Table 1). All genotype frequencies were in accordance with the Hardy-Weinberg expectation among control subjects (*P* > 0.05).

Combined samples of the three groups generated a 20% decreased CHD risk for the −4673 G allele (OR = 0.80, 95% CI = 0.73-0.87, *P* = 2.22 × 10^{-7}), a 15% decreased CHD risk for the CG genotype (OR = 0.85, 95% CI = 0.75-0.96, *P* = 0.011) and a 40% decreased CHD
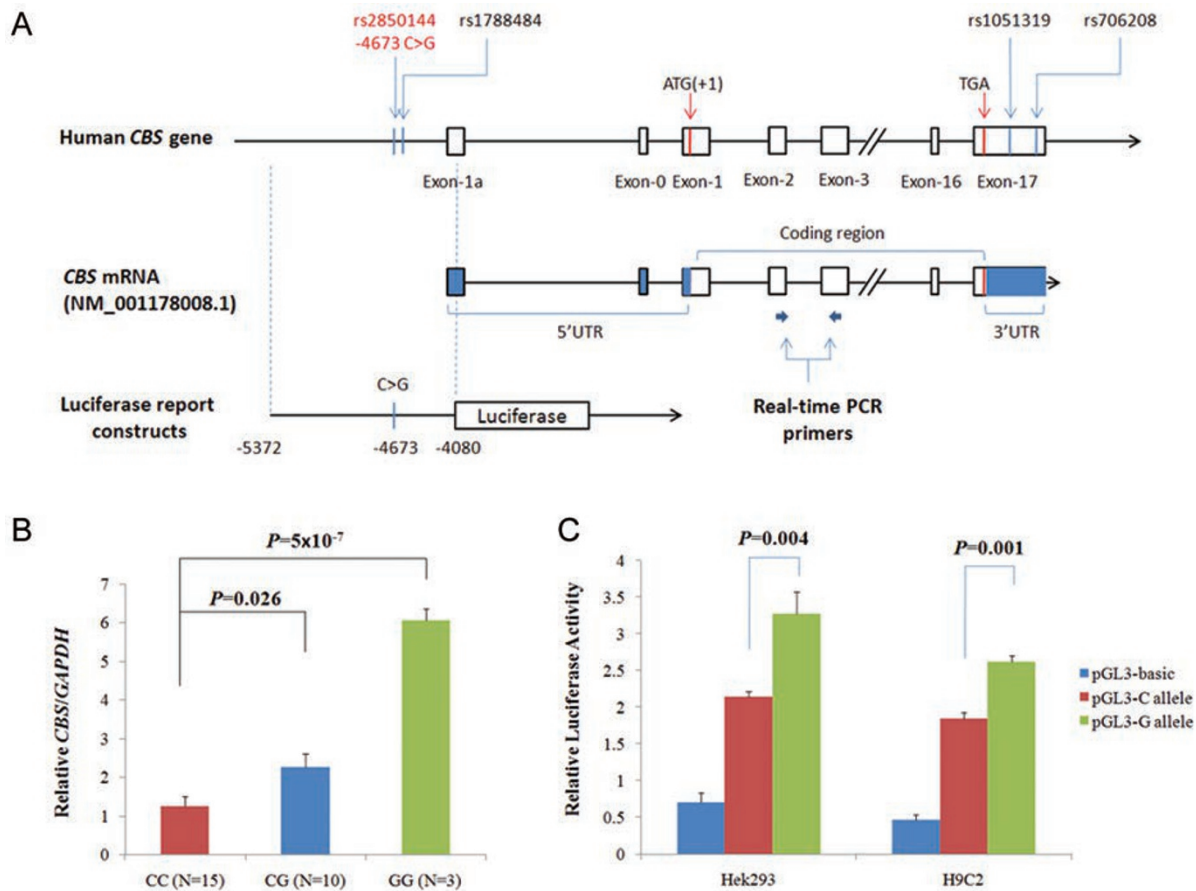
Figure 1 The *CBS* −4673C>G polymorphism upregulates *CBS* expression at the transcriptional level. **(A)** Schematic graph indicates the position of *CBS* −4673C>G polymorphism and the constructs for reporter gene assays in the *CBS* promoter. **(B)** Quantitative real-time PCR analysis of *CBS in vivo* mRNA level in 28 cardiovascular tissue samples with different −4673C>G genotypes. The actual values for each genotype group were as follows: CC = 1.26 ± 0.98; CG = 2.26 ± 1.10; GG = 6.04 ± 0.54. All values were normalized to the level of *GAPDH* and represented means ± SD of three independent experiments. **(C)** Luciferase expression was significantly increased in the minor G allelic construct compared with the major C construct in different cells (52% increases in HEK293 and 42% in H9C2 cells). The actual values in HEK293 cells were as follows: pGL3-basic = 0.70 ± 0.38; pGL3-C allele = 2.14 ± 0.23; and pGL3-G allele = 3.27 ± 0.90. The values in H9C2 cells were the following: pGL3-basic = 0.46 ± 0.22; pGL3-C allele = 1.85 ± 0.22; and pGL3-G allele = 2.63 ± 0.22. Each value represents mean ± SD of three independent experiments, and each experiment was performed in triplicate.

risk for the GG genotype (OR = 0.60, 95% CI = 0.49-0.73, $P = 1.78 \times 10^{-7}$) compared to the CC genotype.

We checked the SNP frequency of *CBS* −4673 site in different ethnic groups (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=2850144). The allele frequency varies remarkably in different ethnic population at *CBS* −4673 site (YRI: C/G = 0.127/0.873; CEU: C/G = 0.375/0.625; CHB+JPT: C/G = 0.408/0.592). It is interesting that the protective G allele is more abundant in Africa than in Europe or Asia, and Africa is reported to have the lowest total CHD birth prevalence (1.9 per 1 000 live births) comparing with Europe (8.2 per 1 000 live births) and Asia (9.3 per 1 000 live births) [1].

A non-coding variant c.56+781A>C in the *MTRR* gene was significantly associated with CHD in the same cohort [8]. So the possible gene-gene interaction between *CBS* (−4673C>G) and *MTRR* (c.56+781A>C) were queried through logistic regression model and epistasis analysis in PLINK. However, there are no any interactions observed by PLINK epistasis analysis between these two variants either in additive model (coefficient = −0.034, *P* = 0.595), or in recessive model (coefficient = 0.15, *P* = 0.585), or in dominant model (coefficient = −0.125, *P* = 0.314), or in genotype model (*CBS* −4673C>G interacts with *MTRR* c.56+781A>C. CG/CA: coefficient = −0.135, *P* = 0.332; GG/CA: coefficient = −0.037, *P* = 0.862; CG/

Table 1 Associations between *CBS* −4673C>G and the risk of CHD in 3 separate groups

Group	Genotype	Case	Control	P value[1]	OR (95% CI)[3]	P value	HWE test[2]
	C/C	295 (49%)	273 (41.4%)		1.00		
	C/G	255 (42.4%)	299 (45.3%)	0.003	0.79 (0.62-1.00)	0.047	0.67
Shanghai	G/G	52 (8.6%)	88 (13.3%)		0.53 (0.36-0.78)	0.001	
	C	845 (70.2%)	845 (64.0%)		1.00		
	G	359 (29.8%)	475 (36.0%)	0.001	0.75 (0.63-0.87)	0.001	
	C/C	350 (47.6%)	243 (43.1%)		1.00		
	C/G	322 (43.8%)	244 (43.3%)	0.011	0.91 (0.72-1.15)	0.44	0.23
Shandong	G/G	63 (8.6%)	77 (13.7%)		0.57 (0.39-0.82)	0.0028	
	C	1 022 (69.5%)	730 (64.7%)		1.00		
	G	448 (30.5%)	398 (35.3%)	0.009	0.80 (0.68-0.95)	0.009	
	C/C	465 (46.4%)	430 (41.1%)		1.00		
	C/G	430 (42.9%)	464 (44.4%)	0.01	0.86 (0.71-1.03)	0.10	0.14
Jiangsu	G/G	108 (10.8%)	152 (14.5%)		0.66 (0.50-0.87)	0.003	
	C	1 360 (67.8%)	1 324 (63.3%)		1.00		
	G	646 (32.2%)	768 (36.7%)	0.002	0.82 (0.72-0.93)	0.002	
	C/C	1 110 (47.4%)	946 (41.7%)		1.00		
	C/G	1 007 (43%)	1 007 (44.4%)	6.62×10^{-7}	0.85 (0.75-0.96)	0.011	0.062
Combined	G/G	223 (9.5%)	317 (14%)		0.60 (0.49-0.73)	1.78×10^{-7}	
	C	3 227 (69.0%)	2 899 (63.9%)		1.00		
	G	1 453 (31.0%)	1 641 (36.1%)	2.17×10^{-7}	0.80 (0.73-0.87)	2.22×10^{-7}	

[1]Genotype frequencies in case and control participants were compared using χ^2 test with 2 degrees of freedom (df).
[2]*P* value for Hardy-Weinberg equilibrium test in the control subjects.
[3]Adjusted for age, sex.

CC: coefficient = −0.247, *P* = 0.203; GG/CC: coefficient = 0.036, *P* = 0.909).

CBS −4673C>G is strongly related with septation defects and conotruncal heart disease

According to the CHD standard classification [16], a stratified analysis of −4673C>G was performed. The most significant protective effect was observed for septation defects (1 652 cases, $P = 1.29 \times 10^{-6}$) and conotruncal defects (386 cases, *P* = 0.002) (Table 2). Additionally, in the sub-grouped CHD, we observed that *CBS* −4673C>G is significantly related with ventricular septal defect (VSD, 1 220 cases, $P = 3.18 \times 10^{-5}$), atrial septal defect (ASD, 235 cases, *P* = 0.0006) and tetralogy of Fallot (TOF, 291 cases, *P* = 0.0016).

These affected CHD sub-classifications are highly consistent with the *CBS* expression pattern in both human and mouse embryos [10, 11]. *CBS* mRNA was observed in a 36-day-old human embryo (CS15) in the endocardium cells and was additionally found in cells derived from the neural crest forming the myocardial walls of the bulbus cordis [11]. Then, *CBS* mRNA was

detectable in the endocardial tissue and in the atrial and ventricle myocardial walls at a later stage [11]. In the mouse cardiovascular system, *CBS* expression was additionally detected from E12.5 in the endocardial cells [10].

Meanwhile, nonsyndromic isolated CHD showed a significant association with the *CBS* −4673C>G polymorphism (2 025 cases, $P = 3.2 \times 10^{-7}$), but non-isolated CHD did not (315 cases, *P* = 0.25).

The −4673C>G variant increases CBS transcriptional activity

Because the associated polymorphism was located in the promoter region of the *CBS* gene, we speculated that it might affect *CBS* transcription. The real-time RT-PCR results using 28 cardiovascular tissue samples showed that the −4673CG samples displayed 2-fold and −4673GG samples nearly 6-fold upregulated expression compared with the samples of the CC genotype (Figure 1B).

To further investigate whether the increased *CBS* mRNA level was caused by the −4673C>G variant, we performed luciferase assays. As expected, the plasmid

Table 2 Stratification analysis of *CBS* −4673C>G genotypes according to CHD classification and phenotype. The used controls were the total 2 270 combined controls

Variable	Case number	*P* value	Association [OR (95% CI)][1]	
			CG vs CC	GG vs CC
CHD Classification I[2]				
Conotruncal defects	386	0.002	0.69 (0.55-0.88)	0.63 (0.44-0.91)
Septation defects	1 652	1.29×10^{-6}	0.92 (0.81-1.05)	0.57 (0.46-0.71)
LVOTO	47	0.51	0.69 (0.36-1.30)	0.85 (0.36-2.01)
RVOTO	75	0.98	0.97 (0.59-1.59)	0.92 (0.45-1.91)
APVR	14	0.006	0.08 (0.01-0.66)	0.54 (0.12-2.47)
Complex CHDs	34	0.01	0.30 (0.13-0.71)	0.69 (0.26-1.85)
Other CHDs	132	0.62	0.93 (0.64-1.35)	0.75 (0.42-1.35)
CHD Classification II				
Isolated nonsyndromic CHD	2 025	3.2×10^{-7}	0.86 (0.75-0.97)	0.58 (0.47-0.70)
Syndromic CHD	315	0.25	0.84 (0.64-1.08)	0.77 (0.53-1.14)
Isolated phenotype				
ASD (atrial septal defect)	235	0.0006	0.89 (0.67-1.19)	0.37 (0.22-0.65)
VSD (ventricular septal defect)	1 220	3.18×10^{-5}	0.93 (0.80-1.08)	0.58 (0.45-0.74)
TOF (tetralogy of Fallot)	291	0.0016	0.66 (0.50-0.85)	0.59 (0.39-0.89)

[1]Adjusted for age, sex;
[2]Classification described in [16].

containing the protective G allele displayed a significantly higher luciferase expression than the wild-type C allele with a 52% increase in the human embryonic kidney 293 (HEK293) cells and a 42% increase in the rat cardiac myocyte (H9C2) cells (Figure 1C). These consistent *in vivo* and *in vitro* results confirmed that the promoter −4673C>G variant functionally increases the transcription of *CBS*.

The variant −4673C>G attenuates the transcription repressor SP1 binding affinity

The *CBS* variant −4673C>G is in complete linkage with another nearby variant, rs1788484 (−4638G>A). A surface plasmon resonance (SPR) analysis was conducted to test the DNA-protein binding capacity of both variants. In the direct SPR assay using the HEK293 nuclear extraction, the SPR-binding activity of the major −4673C allele was more than 1 000-fold higher than that of the minor G allele (Figure 2A). However, there were no binding activity differences between the linked variant −4638G allele and the A allele (Supplementary information, Figure S1). We performed the competition SPR assay for −4673C>G by adding 5- or 10-fold excess non-biotinylated C/G probes as competitors. The SPR responses in both competition assays confirmed the direct assay results that the probe of the −4673C allele has a significantly higher affinity to certain transcription factors than the −4673G allele probe (Figure 2B).

The computational analysis (using Alibaba2 software) predicted that the promoter variant −4673C>G could influence the binding ability of the transcriptional factors SP1 and GR (Supplementary information, Figure S2). However, the chromatin immunoprecipitation (ChIP) sequencing results in the online database (Biobase: www.biobase-international.com; GenomeUCSC: http://genome.ucsc.edu/cgi-bin/hgTrackUi?db=hg18&g=wgEncodeYaleChIPseq) and other studies suggested that the *CBS* promoter could bind to the transcriptional factor SP1 but not GR [17, 18]. ChIP assays using either HEK293 cells or cardiac tissue samples demonstrated that site −4673 of the *CBS* promoter was occupied by SP1 (Figure 2C). The *CBS* promoter fragment with site −4673 could be specifically precipitated with an SP1 antibody, but not with antibodies against CEBPα or a nonspecific rabbit IgG, which was confirmed by PCR amplification. The co-transfection experiment showed that the overexpressed SP1 decreased *CBS* transcription generally, and the variant −4673C/G amplified the promoter function disparity (Figure 2D). Taken together, SP1 acts as a transcription repressor and the −4673C>G variant attenuates the binding affinity with SP1, which finally results in the elevated expression of *CBS*.

The variant −4673C>G is correlated with the different allelic promoter methylation

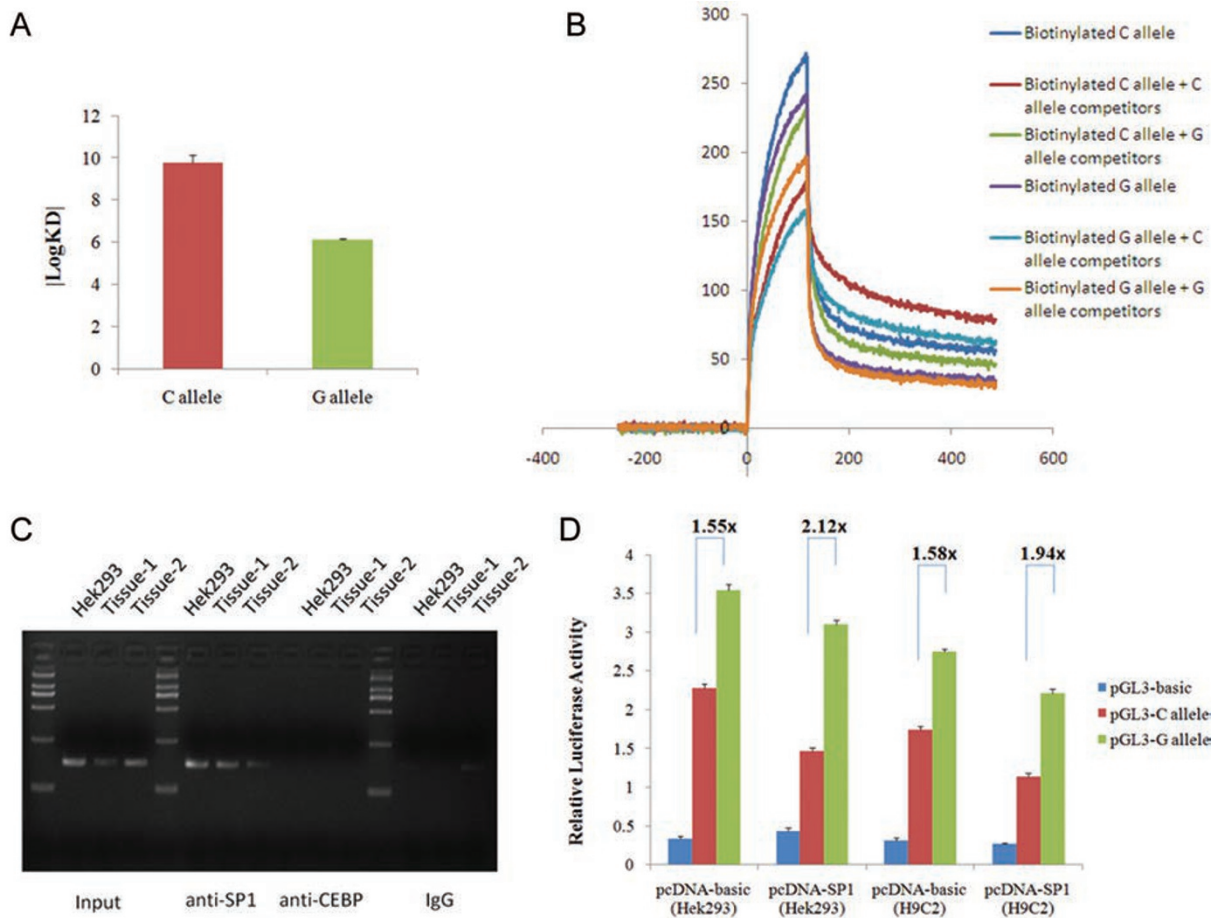The *CBS* promoter is rich in GC, and the variant

Figure 2 The *CBS* −4673 G allele attenuates transcription factor binding affinity. **(A)** SPR analysis comparing the binding affinity of nuclear extracts to DNA probes containing either the −4673 C or G alleles. The SPR-binding activity of the A-allele ($|logKD|$ = 9.75 ± 0.65) was more than 1 000-fold higher than that of the C-allele ($|logKD|$ = 6.12 ± 0.05). Each value represents the mean ± SD of three experiments. **(B)** Competition SPR assays performed in the presence of 5-fold excess non-biotinylated C- or G-allele probes. **(C)** ChIP assays using HEK293 and cardiovascular tissue samples. The presence of the SP1-binding *CBS* promoter was verified by PCR. **(D)** A luciferase construct containing either the C or G allele was co-transfected with pcDNA3.1 (control) or pcDNA3.1-SP1 expression plasmids. The actual values in HEK293 cells were as follows: pcDNA3.1-basic group: pGL3-basic = 0.34 ± 0.08, pGL3-C allele = 2.27 ± 0.16, and pGL3-G allele = 3.54 ± 0.23; pcDNA3.1-SP1 group: pGL3-basic = 0.43 ± 0.13, pGL3-C allele = 1.46 ± 0.12, and pGL3-G allele = 3.10 ± 0.16. The actual values in H9C2 cells were the following: pcDNA3.1 group: pGL3-basic = 0.31 ± 0.08, pGL3-C allele = 1.73 ± 0.14, and pGL3-G allele = 2.74 ± 0.10; pcDNA3.1-SP1 group: pGL3-basic = 0.27 ± 0.03, pGL3-C allele = 1.14 ± 0.12, and pGL3-G allele = 2.21 ± 0.18. Each value represented mean ± SD of three experiments, and each experiment was performed in triplicate.

−4673C>G is located within a region with high CpG content. To determine whether the *CBS* expression was regulated by methylation status, we inhibited DNA methylation in HEK293 cells using 5-aza-20-deoxycytidine (5-Aza). After the treatment, *CBS* mRNA increased significantly compared with the control group (DMSO treatment), indicating that *CBS* transcription was influenced by methylation status (Figure 3A). To explore how *CBS* expression was influenced by the promoter methylation status, we analyzed 20 potential methylation sites around site −4673 using 28 cardiac tissue samples (Figure 3B).

The *CBS* mRNA is generally upregulated with the promoter hypomethylation level, which is independent of the −4673C>G SNP. Considering the allelic methylation of variant −4673C>G, nearly 70% of C alleles have at least one CpG site methylated, while almost 80% G alleles are not methylated at any CpG sites (Figure 3D). Moreover, the *CBS* promoter of the −4673 G allele is always associated with a significantly reduced methylation level compared to the C allele promoter (Figure 3C; 15.49% vs 2.73%, P = 0.0001).

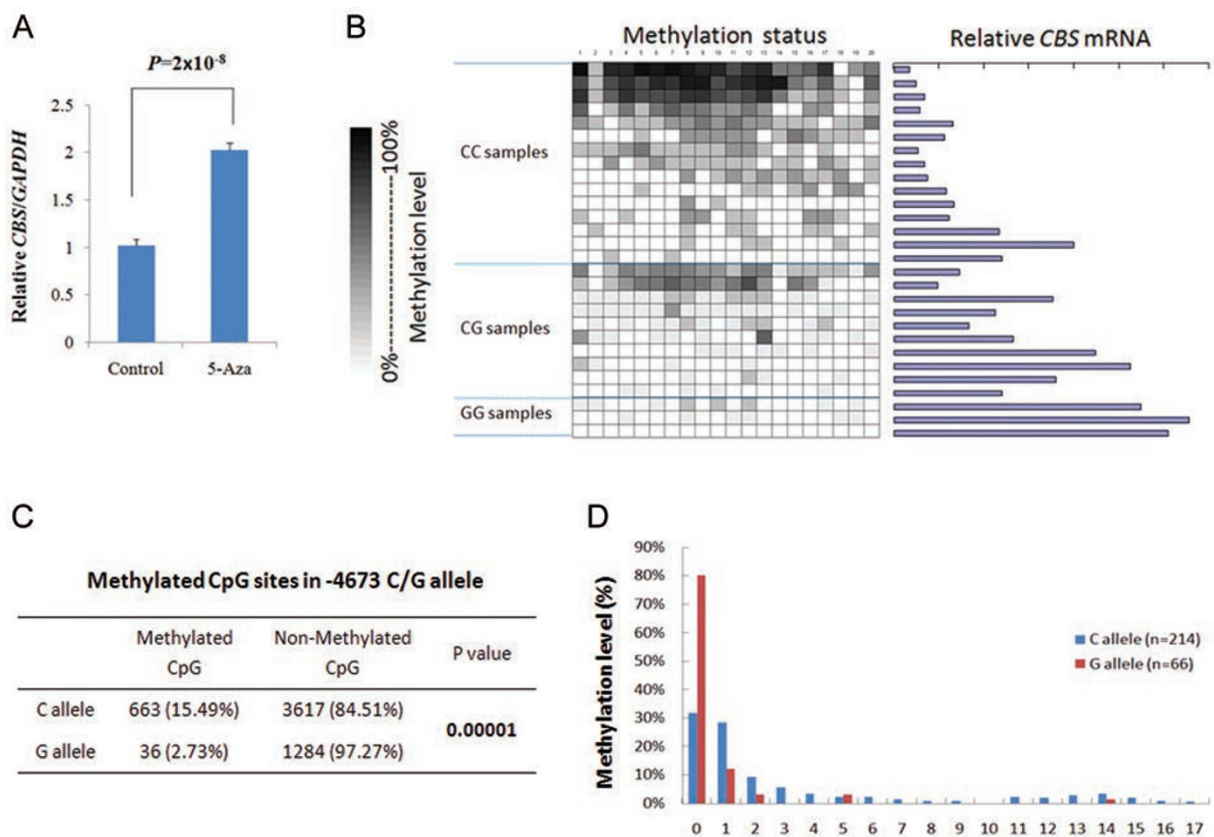Therefore, the hypomethylated minor G allele at site

Figure 3 Polymorphism *CBS* −4673C>G is related with *CBS* gene promoter hypomethylation. **(A)** After blocking DNA methylation in HEK293 cells using 5-Aza, *CBS* mRNA increased significantly compared with the control group. The actual value of 5-Aza group is 2.02 ± 0.22, and value of control group is 1.02 ± 0.19. Each value represents the mean ± SD of three experiments, and each experiment was performed in triplicate. **(B)** The methylation pattern of 20 CpG sites around the −4673 polymorphism (from −4 803 to −4 516) in 28 cardiovascular tissue samples. **(C)** Polymorphism *CBS* −4673C>G was related with allelic-specific methylation. **(D)** Nearly 70% of C alleles had at least one CpG site methylated, but almost 80% G alleles were not methylated at any CpG sites.

−4673 of *CBS*, together with the attenuated transcriptional repressor SP1 binding affinity, would eventually contribute to the elevated *CBS* expression.

CBS −4673C>G is not correlated with plasma homocysteine levels in adults

CBS plays important roles in maintaining plasma homocysteine levels, and *CBS*-deficient patients show severe hyperhomocysteinemia and homocystinuria [9]. We explored the relationship between the *CBS* −4673C>G variant and plasma homocysteine levels in 522 healthy, fasting undergraduate student volunteers using the Axis Homocysteine Enzyme Immunoassay (EIA) assay. The results showed that the −4673C>G variant is not correlated with the plasma homocysteine levels (Figure 4). Although the carriers of the GG genotype had relative lower plasma homocysteine levels than carriers of the CC and CG genotypes, the difference has no statistical
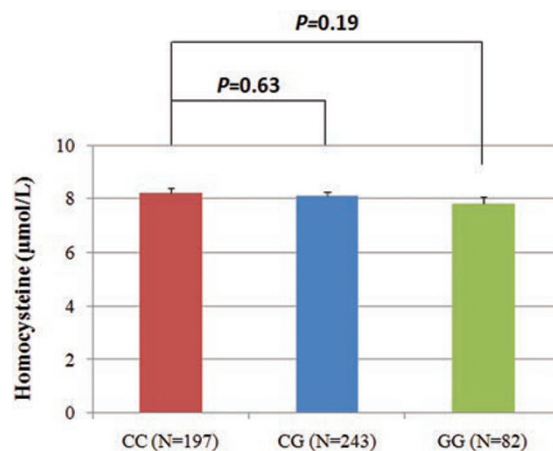

Figure 4 *CBS* −4673C>G polymorphism and human plasma homocysteine level. The actual values for each genotype group were as follows: CC = 8.21 ± 2.52; CG = 8.10 ± 2.31; and GG = 7.80 ± 2.33. Data shown were mean ± SD.

significance.

Discussion

In this study, a functional promoter variant −4673C>G in the *CBS* gene was identified to significantly protect individuals against CHD in a large-scale case-control study in a Han Chinese population. In the same case-control cohort, we also found that non-coding variants in the *MTRR* [8] and *MTR* (unpublished) were significantly associated with increased CHD risk. *CBS* and *MTRR/MTR* are responsible for homocysteine clearance with one accord and are the only 3 genes found to be significantly associated with CHD after systematically investigating 9 genes (*CBS, MTRR* [8]*, MTR, TYMS* [19]*, MTHFR, MTHFD, DNMT1, DNMT3A* and *DNMT3B*) in the folate/homocysteine metabolism pathway. These results consistently suggest that guaranteed homocysteine removal might be vital for normal heart development. Genetic variants affecting the capacity for homocysteine removal are closely associated with CHD and may contribute to the common CHD etiology. Meanwhile, our findings further accentuated the potential importance of non-coding variants in the folate pathway core genes in the etiology of CHD that was previously ignored. However, our initial observations only involved Han Chinese population, thus further confirmatory studies are demanded in other ethnic groups.

The interesting question is why the elevated *CBS* expression could protect individuals against CHD. In both Down syndrome children and the *CBS* transgenic mice, a significant decrease in plasma homocysteine levels was observed with the overexpression of *CBS* [20-24]. In contrast, both the patients with inherited homocystinuria caused by *CBS* deficiency and the *Cbs* knockout mice had significantly elevated plasma homocysteine levels [9, 25-30]. Therefore, the *CBS* expression level demonstrated a correlation with the plasma homocysteine levels, although in this study we did not observe a statistically significant difference in plasma homocysteine levels in 522 fasting undergraduates with different *CBS* −4673 genotypes. It might be due to the age of the tested subjects. In Down syndrome patients carrying three copies of *CBS*, the plasma homocysteine level increased with age. Various studies of Down syndrome children showed consistently decreased plasma homocysteine levels compared with the same age controls [20-22, 31], while trisomy 21 adults showed no difference in homocysteine levels [23], and elderly patients even had increased homocysteine levels compared with controls [32]. Thus the correlation between *CBS* −4673 genotypes and plasma homocysteine level during embryonic development cannot be truly re-

flected in the tested undergraduate students.

In both human and mouse embryos, *CBS* is expressed in endocardium cells and cardiac neural crest cells, which implies that CBS might play important roles in maintaining homocysteine level in these cells [10, 11]. Our results indicate that the *CBS* −4673C>G variant was highly associated with septation defects (1 652 cases, $P = 1.29 \times 10^{-6}$) and conotruncal defects (386 cases, $P = 0.002$), and they also indirectly supported the influence of *CBS* expression mainly on endocardium cells and cardiac neural crest cells. Exposure of chicken and mouse embryos to exogenous homocysteine also resulted in septal defects [7]. Folate supplementation has been implicated in specifically reducing conotruncal anomalies and VSDs [33]. Consequently, the carriers with genetically increased *CBS* expression would benefit from the protection due to the low homocysteine levels maintained by CBS in certain cells during the critical heart development stages.

It is notable that we only observed protective variant in *CBS* gene instead of risk variant as previously reported in *MTRR* gene in the same cohort [8]. And among the *CBS*-deficient HCU patients or *Cbs* knockout mice, there were no reports that they were affected by CHD, although they had dramatically elevated plasma homocysteine levels [9, 25-30]. We presume that it reflects the different roles of transsulfuration pathway catalyzed by CBS and remethylation pathway by MTR/MTRR in homocyestine removal in different developmental stages. The *Km* of CBS for homocysteine is estimated to be approximately 5 mmol/L, whereas the *Km* for both MTR and betaine-dependent homocysteine methyltransferase (BHMT) are in the low micromolar range [34]. Consequently, remethylation with homocysteine conservation would be favored at low concentrations of the metabolite. Conversely, cystathionine synthase can utilize homocysteine at the higher concentrations that exceeds the capacity of the methylases and is the only reaction that removes homocysteine from the methionine cycle. Thus, the *CBS* deficiency-induced excess homocysteine in certain embryonic cardiac cells could be removed by the dominant pathway (remethylation by MTR) or by maternal intervention. However, the overloaded homocysteine due to *CBS* deficiency after birth could not be compensated anymore, and CBS activity undergoes dramatic change after birth to maintain normal plasma total homocysteine (tHcy) level. Thus, HCU patients and *Cbs* knockout mice had extremely high plasma tHcy levels without any CHD phenotype [9, 25-30]. On the other hand, the decreasing MTR/MTRR activity, which plays a vital role in homocysteine removal, could not be compensated either by CBS activity in the same cells or by hepatic BHMT before fetal liver formation on day 32. Therefore, *CBS* deficiency would

not lead to CHD directly, but increasing *CBS* expression will protect certain susceptible cells from a homocysteine attack and reduce CHD risk, whereas decreasing *MTR/MTRR* expression caused by genetic variants will significantly increase CHD risk [8]. This hypothesis still needs further validation in animal models in the future.

Another concern might be that the overexpressed *CBS* has limited protection effects for CHD in the Down syndrome patients. As we know, 40%-60% of Down syndrome patients have accompanying heart defects [35]. The plausible explanation is that the protective effect of *CBS* via homocysteine removal is overshadowed by the structural defects caused by the other dosage-sensitive genes on chromosome 21. In our study, we noticed that the *CBS* −4673C>G could not show protective effects on non-isolated CHD cases ($n = 315$, $P = 0.25$) because the causative mutations for the complicated CHD are too strong to be reversed by the elevated *CBS* expression. Meanwhile, the triplicated Mmu17 mice, a Down syndrome model without CHD, indicated that the whole triplicated region covering 19 genes, including *CBS*, were excluded as the causative factor of Down syndrome CHD [36].

Folate/homocysteine metabolism is very complex, and the homocysteine level in embryonic cells is determined by both gene-gene interactions (such as *CBS, MTR, MTRR, etc.*) and gene-environment interaction (such as maternal folate status). Therefore, it will be necessary to comprehensively evaluate these interacted variants and maternal folate levels for the accurate risk prediction of the offspring's CHD outcomes.

Although folate intake has been recommended to prevent newborns from birth defects, including CHD, for decades, the underlying molecular mechanism remains elusive. In our studies, only 3 homocysteine removal genes (*CBS, MTR* and *MTRR* [8]) were strongly associated with CHD risk, which suggests that homocysteine, as a typical metabolite in the folate pathway, may play a potentially more important role than a sensitive biomarker. One possibility is that homocysteine level fluctuations, which can be genetically influenced by *CBS* or *MTR/MTRR* expression, may lead to an altered protein homocysteinylation, a type of post-translational modification. For the homocysteinylated proteins, the possible direct effects are changing protein functions, resulting in possible changes to developmental pathways. Investigation of associated protein homocysteinylation may be helpful to the interpretation of the molecular mechanisms involved in folate supplementation.

Our findings demonstrated a significant association between the *CBS* −4673C>G variant and a reduced risk of CHD in three independent Han Chinese popula-tions. Because CBS played important roles in regulating homocysteine level in certain cardiac cells, elevated *CBS* expression would certainly protect the developing embryo against CHD, especially septation defects and conotruncal defects. Functional studies revealed that the *CBS* −4673C>G variant upregulates *CBS* expression by attenuated transcriptional suppressor SP1 binding affinity and promoter methylation. Our study sheds light on an unexpected role of CBS in cardiac development and further emphasizes the functional SNPs in non-coding region of homocysteine removal genes in birth defect research. The reported findings may initiate novel prediction and prevention strategy for birth defects, especially post folic acid fortification.

Materials and Methods

Study subjects

We analyzed samples from three independent case-control groups (Shanghai, Shandong and Nanjing) consisting totally 2 340 CHD patients and 2 270 controls, and classified the 2 340 cases into seven broad categories as described previously (Supplementary information, Table S2) [8]. In order to avoid potential artifacts due to population stratification effects, we performed principle component (PC) analyses using 16 ancestry-informative markers in Shanghai and Shandong groups. In both Shanghai and Shandong groups, rs2627923, rs117832620 and rs34852119 were the PCs (Shanghai group: cumulative = 0.967, lambda = 0.365; Shandong group: cumulative = 0.966, lambda = 0.398). There was no significant difference in these ancestry-informative biomarkers among the tested cases and controls (Shanghai group: $P_{PC1} = 0.267$, $P_{PC2} = 0.974$, $P_{PC3} = 0.664$; Shandong group: $P_{PC1} = 0.595$, $P_{PC2} = 0.639$, $P_{PC3} = 0.589$).

28 human cardiac tissue samples were used for the quantitative RT-PCR assay as described previously [8]. They were obtained from CHD patients who had undergone heart catheterizations or cardiac operations between January 2010 and May 2010 at the Cardiovascular Disease Institute, General Hospital of Jinan Military Command (Jinan, Shandong Province, China). The collected cardiac tissues include 3 ventricle samples, 4 atrium samples, 2 auricle samples, 10 outlet samples, 5 aorta samples, and 4 ventricular septum samples.

All study protocols were reviewed and approved by the each medical center's ethics committee and written consent was obtained from the parents and/or patients prior to the commencement of the study.

SNP identification and genotyping

Genomic DNA was isolated from the venous blood using conventional reagents. The *CBS* 5′ regulatory region from −6 251 to −3 623 bp (2 628 bp, chr21: 43344288- 43346916, NC_000021.7, GI: 51511750) and the fragment containing the whole 3′ UTR (930 bp, chr21: 43368713-43369643, NC_000021.7, GI: 51511750) were amplified by PCR from 32 unrelated individuals randomly selected from both the Shanghai and Shandong groups for polymorphism screening by sequencing. Direct dye terminator sequencing of the PCR products was carried out using the ABI Prism

BigDye system, according to the manufacturer's instructions (ABI, Foster City, CA, USA). Selected SNPs were genotyped using SNaPshot analysis (ABI). The samples for sequencing and genotyping were run on an ABI 3730 automated sequencer and analyzed by SeqMan and Peakscan, respectively. All of the primer sequences are listed in Supplementary information, Table S3.

Plasmid constructs, host cell culture and luciferase assays

To construct the *CBS* reporter plasmid, we amplified the 1 292 bp *CBS* fragment from −5 372 to −4 080 by PCR from genomic DNA, which contains the C allele of SNP −4673C>G. The PCR products were subcloned into the *Nhe*I and *Hin*dIII restriction sites of the pGL3-Basic vector (Promega, Madison, WI, USA). The corresponding G allele plasmid was generated by site-directed mutagenesis with the MutanBEST kit (Takara, Berkeley, CA, USA) to ensure a uniform backbone sequence. We verified all of the recombinant clones by DNA sequencing. The primers utilized are presented in Supplementary information, Table S3.

HEK293 and H9C2 cells (1×10^5) were seeded in 24-well culture plates. After 24 h of culture, the cells were transfected with 1 μg of both a *CBS* promoter reporter plasmid and 20 ng of the pRL-TK plasmid (Promega) as a normalization control, half of the cells were additionally co-transfected with 50 ng of the pcDNA3.1-SP1 expression plasmid or equivalent amounts of empty pcDNA3.1 vector using Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA), according to the manufacturer's instructions. After an additional 24 h of culture, the transfected cells were assayed for luciferase activity using the Dual-Luciferase Reporter Assay System (Promega). Three independent transfection experiments were performed, and each luciferase assay was carried out in triplicate.

For treatment with 5-Aza (Sigma-Aldrich, St. Louis, MO, USA), which causes DNA demethylation or hemidemethylation, 2×10^5 HEK293 cells were seeded into a 6-well plate in 3 ml of medium. After 24 h of incubation, the medium was removed, and cells were incubated in 3 ml of fresh medium containing 10 mM 5-Aza for 24 h. After treatment, the medium was removed, and cells were subjected to additional 24 h incubation in 3 ml of fresh medium without 5-Aza, and total RNA was extracted. In the control well, 5-Aza was replaced with DMSO, and medium was changed daily.

SPR analysis

The SPR analysis was carried out using the ProteOn XPR36 Protein Interaction Array System (Bio-Rad, Hercules, CA, USA). Biotinylated duplex oligonucleotide probes representing the −4673 C or G alleles and −4638 G or A alleles (sequences listed in Supplementary information, Table S3) were immobilized on the streptavidin-modified surfaces of the different channels from DNA solutions at a fixed concentration (400 nM) to ensure identical surface density. Nuclear extracts from HEK293 cells were diluted in PBST to different concentrations and then pre-incubated with non-specific DNA for 15 min before passing across the DNA-immobilized surface. Non-biotinylated competitors in 4-fold excess were used for the competition assays. The results presented in the sensorgram were converted by BIA evaluation software. Each experiment was repeated 3 times.

ChIP assays

The ChIP assays were conducted using the EZ ChIP Kit (Up-state, Lake Placid, NY, USA). First, HEK293 cells and two cardiac tissue samples were crosslinked by 1% formaldehyde for 10 min. DNA was then sonicated into fragments with a mean length of 200 to 1 000 bp. The sheared chromatin was immunoprecipitated by incubation with antibodies against SP1, CEBPα or non-specific rabbit IgG (Santa Cruz Biotechnology, CA, USA) overnight at 4 °C. The DNA fragments were identified using PCR, and the primers used are listed in Supplementary information, Table S3.

Quantitative real-time RT-PCR

Total RNA was extracted from the human cardiovascular tissue samples preserved in RNAlater (Qiagen, Valencia, CA, USA) and converted to cDNA using random hexamers, oligo (dT) primers and Moloney murine leukemia virus reverse transcriptase (Takara). The *CBS* mRNA levels were measured by quantitative real-time RT-PCR using the ABI Prism 7900 sequence detection system with *GAPDH* as an internal reference gene. Each reaction was performed in triplicate. The primers used are listed in Supplementary information, Table S3.

Bisulfite sequencing

The treatment of genomic DNA with bisulfite was performed using the EZ DNA Methylation-Gold kit (ZYMO Research, Los Angeles, CA, USA), according to the manufacturer's instructions. The region near −4673C>G of the *CBS* promoter from −4 803 to −4 516 consisting of 20 CpG sites was amplified. The PCR products were purified and subcloned into pMD T-19 (Takara). After an overnight bacterial culture, 10 subclones from each of the 28 different PCR assays were subjected to direct sequencing on an ABI Prism 3770.

Plasma homocysteine detection

EDTA-plasma samples were obtained from fasting undergraduate volunteers in the early morning, centrifuged immediately and stored in a −80 °C freezer until subjected to homocysteine detection. The Axis® Homocysteine Enzyme Immunoassay (EIA) Kit (Axis-Shield, Norton, MA, USA) was used to determine the plasma homocysteine level according to the manufacturer's instructions. Each test was duplicated, and the mean level was used for further analysis.

Statistical analysis

Differences in demographic features and allelic or genotypic frequencies between the CHD cases and controls were compared using the χ^2 test. The Hardy-Weinberg equilibrium was additionally tested using the χ^2 test in the controls. To evaluate the associations between the genotypes and the CHD risk, ORs and 95% CIs were calculated by unconditional logistic regression analysis with adjustments for age and sex. Other differences were evaluated using the Student's *t*-test. All statistical tests were two-tailed with $P < 0.05$ set as the significance level and were performed using SPSS 15.0 software (SPSS, Chicago, IL, USA).

Acknowledgments

References

1   van der Linde D, Konings E, Slager MA, *et al*. Birth prevalence of congenital heart disease worldwide a systematic review and meta-analysis. *J Am Coll Cardiol* 2011; **58**:2241-2247.

2   Botto LD, Mulinare J, Erickson JD. Occurrence of congenital heart defects in relation to maternal mulitivitamin use. *Am J Epidemiol* 2000; **151**:878-884.

3   Botto LD, Mulinare J, Erickson JD. Do multivitamin or folic acid supplements reduce the risk for congenital heart defects? Evidence and gaps. *Am J Med Genet PART A* 2003; **121A**:95-101.

4   van Beynum IM, den Heijer M, Blom HJ, Kapusta L. The MTHFR 677C->T polymorphism and the risk of congenital heart defects: a literature review and meta-analysis. *QJM* 2007; **100**:743-753.

5   Hobbs CA, Cleves JA, Melnyk S, Zhao WZ, James SJ. Congenital heart defects and abnormal maternal biomarkers of methionine and homocysteine metabolism. *Am J Clin Nutr* 2005; **81**:147-153.

6   Kapusta L, Haagmans M, Steegers E, Cuypers M, Blom HJ, Eskes T. Congenital heart defects and maternal derangement of homocysteine metabolism. *J Pediatr* 1999; **135**:773-774.

7   van Mil NH, Oosterbaan AM, Steegers-Theunissen RP. Teratogenicity and underlying mechanisms of homocysteine in animal models: a review. *Reprod Toxicol* 2010; **30**:520-531.

8   Zhao JY, Yang XY, Gong XH, *et al*. Functional variant in methionine synthase reductase intron-1 significantly increases the risk of congenital heart disease in the Han Chinese population. *Circulation* 2012; **125**:482-490.

9   Harvey MS, Levy HL, Kraus JP. Disorders of Transsulfuration. In: Scriver CR, Beaudet AL, Sly WS, *et al*. eds. *The Metabolic and Molecular Bases of the Inherited Disease*. New York: McGraw-Hill 2001: 2007-2056.

10 Robert K, Vialard F, Thiery E, *et al*. Expression of the cystathionine beta synthase (CBS) gene during mouse development and immunolocalization in adult brain. *J Histochem Cytochem* 2003; **51**:363-371.

11 Quere I, Paul V, Rouillac C, *et al*. Spatial and temporal expression of the cystathionine beta-synthase gene during early human development. *Biochem Biophys Res Commun* 1999; **254**:127-137.

12 Song XM, Zheng XY, Zhu WL, Huang L, Li Y. Relationship between polymorphism of cystathionine beta synthase gene and congenital heart disease in Chinese nuclear families. *Biomed Environ Sci* 2006; **19**:452-456.

13 Locke AE, Dooley KJ, Tinker SW, *et al*. Variation in folate pathway genes contributes to risk of congenital heart defects among individuals with Down syndrome. *Genet Epidemiol* 2010; **34**:613-623.

14 Kraus JP, Oliveriusova J, Sokolova J, *et al*. The human cystathionine beta-synthase (CBS) gene: complete sequence, alternative splicing, and polymorphisms. *Genomics* 1998; **52**:312-324.

15 Bao L, Vlcek C, Paces V, Kraus JP. Identification and tissue distribution of human cystathionine beta-synthase mRNA isoforms. *Arch Biochem Biophys* 1998; **350**:95-103.

16 Botto LD, Lin AE, Riehle-Colarusso T, Malik S, Correa A. Seeking causes: Classifying and evaluating congenital heart defects in etiologic studies. *Birth Defects Res A Clin Mol Teratol* 2007; **79**:714-727.

17 Reddy TE, Pauli F, Sprouse RO, *et al*. Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. *Genome Res* 2009; **19**:2163-2171.

18 Cawley S, Bekiranov S, Ng HH, *et al*. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* 2004; **116**:499-509.

19 Zhao JY, Sun JW, Gu ZY, *et al*. Genetic polymorphisms of the *TYMS* gene are not associated with congenital cardiac septal defects in a Han Chinese population. *PLoS One* 2012; **7**:e31644.

20 Meguid NA, Dardir AA, El-Sayed EM, Ahmed HH, Hashish AF, Ezzat A. Homocysteine and oxidative stress in Egyptian children with Down syndrome. *Clin Biochem* 2010; **43**:963-967.

21 Pogribna M, Melnyk S, Pogribny I, Chango A, Yi P, James SJ. Homocysteine metabolism in children with Down syndrome: *in vitro* modulation. *Am J Hum Genet* 2001; **69**:88-95.

22 Al-Gazali LI, Padmanabhan R, Melnyk S, *et al*. Abnormal folate metabolism and genetic polymorphism of the folate pathway in a child with Down syndrome and neural tube defect. *Am J Med Genet* 2001; **103**:128-132.

23 Fillon-Emery N, Chango A, Mircher C, *et al*. Homocysteine concentrations in adults with trisomy 21: effect of B vitamins and genetic polymorphisms. *Am J Clin Nutr* 2004; **80**:1551-1557.

24 Wang L, Jhee KH, Hua X, DiBello PM, Jacobsen DW, Kruger WD. Modulation of cystathionine beta-synthase level regulates total serum homocysteine in mice. *Circ Res* 2004; **94**:1318-1324.

25 Maclean KN, Sikora J, Kozich V, *et al*. A novel transgenic mouse model of CBS-deficient homocystinuria does not incur hepatic steatosis or fibrosis and exhibits a hypercoagulative phenotype that is ameliorated by betaine treatment. *Mol Genet Metab* 2010; **101**:153-162.

26 Maclean KN, Sikora J, Kozich V, *et al*. Cystathionine beta-synthase null homocystinuric mice fail to exhibit altered hemostasis or lowering of plasma homocysteine in response to betaine treatment. *Mol Genet Metab* 2010; **101**:163-171.

27 Gupta S, Kuhnisch J, Mustafa A, *et al*. Mouse models of cystathionine beta-synthase deficiency reveal significant threshold effects of hyperhomocysteinemia. *FASEB J* 2009; **23**:883-893.

28 Heil SG, Riksen NP, Boers GH, Smulders Y, Blom HJ. DNA methylation status is not impaired in treated cystathionine beta-synthase (CBS) deficient patients. *Mol Genet Metab* 2007;

91:55-60.

29 Wang L, Chen X, Tang B, Hua X, Klein-Szanto A, Kruger WD. Expression of mutant human cystathionine beta-synthase rescues neonatal lethality but not homocystinuria in a mouse model. *Hum Mol Genet* 2005; **14**:2201-2208.

30 Watanabe M, Osada J, Aratani Y, *et al*. Mice deficient in cystathionine beta-synthase: animal models for mild and severe homocyst(e)inemia. *Proc Natl Acad Sci USA* 1995; **92**:1585-1589.

31 Gueant JL, Anello G, Bosco P, *et al*. Homocysteine and related genetic polymorphisms in Down's syndrome IQ. *J Neurol Neurosurg Psychiatry* 2005; **76**:706-709.

32 Licastro F, Marocchi A, Penco S, *et al*. Does Down's syndrome support the homocysteine theory of atherogenesis? Experience in elderly subjects with trisomy 21. *Arch Gerontol Geriatr* 2006; **43**:381-387.

33 Jenkins KJ, Correa A, Feinstein JA, *et al*. Noninherited risk factors and congenital cardiovascular defects: Current knowledge a scientific statement from the American Heart Association Council on cardiovascular disease in the young. *Circulation* 2007; **115**:2995-3014.

34 Finkelstein JD. The metabolism of homocysteine: pathways and regulation. *Eur J Pediatr* 1998; **1572**:S40-S44.

35 Vis JC, Duffels MG, Winter MM, *et al*. Down syndrome: a cardiovascular perspective. *J Intellect Disabil Res* 2009; **53**:419-425.

36 Liu C, Morishima M, Yu T, *et al*. Genetic analysis of Down syndrome-associated heart defects in mice. *Hum Genet* 2011; **130**:623-632.

(**Supplementary information** is linked to the online version of the paper on the *Cell Research* website.)