Genomics 115 (2023) 110565

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Genome-wide DNA methylation analysis of discordant monozygotic twins reveals consistent sites of differential methylation associated with congenital heart disease

Xi Yuan [a,1], Jiayu Huang [b,1], Li Wen [a,1], Boris Novakovic [c,d,1], Mark D. Kilby [e,f], Chao Tong [a,***], Hongbo Qi [a,g,*], Richard Saffery [c,d,**], Philip N. Baker [a,h,****]

[a] State Key Laboratory of Maternal and Fetal Medicine of Chongqing Municipality, The First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, China
[b] Reproductive Medicine Center, The First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, China
[c] Molecular Immunity, Murdoch Children's Research Institute, Royal Children's Hospital, Melbourne, Victoria, Australia
[d] Department of Pediatrics, University of Melbourne, Parkville, VIC, Australia
[e] Fetal Medicine Centre, Birmingham Women's & Children's Foundation Trust, Birmingham B15 2TG, UK
[f] Institute of Metabolism & Systems Research, College of Medical & Dental Sciences, University of Birmingham, Birmingham B15 2TT, UK
[g] Department of Obstetrics, Women and Children's Hospital of Chongqing Medical University, Chongqing 401147, China
[h] College of Life Sciences, University of Leicester, Leicester LE1 7RH, UK

ARTICLE INFO

Keywords:
Congenital heart disease
Monozygotic twins
Umbilical cord blood
DNA methylation

ABSTRACT

Background: Despite being essentially genetically identical, monozygotic (MZ) twins can be discordant for congenital heart disease (CHD), thus highlighting the importance of in utero environmental factors for CHD pathogenesis. This study aimed to identify the epigenetic variations between discordant MZ twin pairs that are associated with CHD at birth.
Methods: Cord blood of CHD-discordant MZ twins from the Chongqing Longitudinal Twin Study Cohort was subjected to whole-genome bisulfite sequencing, then validated by MeDIP-qPCR and qRT-PCR.
Results: 379 DMRs mapped to 175 differentially methylated genes (DMGs) were associated with CHD. Functional enrichment analysis identified these DMGs are involved in histone methylation, actin cytoskeleton organization, the regulation of cell differentiation, and adrenergic signaling in cardiomyocytes. Of note, *SPESP1* and *NOX5* were hypermethylated in CHD, and associated with lower gene expression levels.
Conclusions: Specific DNA methy (DNAm) variations in cord blood were associated with CHD, thus illustrating new biomarkers and potential interventional targets for CHD.
Trial registration: ChiCTR-OOC-16008203, registered on 1 April 2016 at the Chinese Clinical Trial Registry.

Abbreviations: CHD, Congenital heart disease; LoTiS, Chongqing Longitudinal Twin Study; MZ, Monozygotic; DZ, Dizygotic; WGBS, Whole-genome bisulfite-sequencing; TOF, Tetralogy of Fallot; VSD, Ventricular septal defect; ASD, Atrial septal defect; DORV, Double outlet right ventricle; DNAm, DNA methylation; DMRs, Differentially methylated regions; DMGs, Differentially methylated genes; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; TSS, Transcription start site; TTS, Transcription termination site; UTR, Untranslated region.

* Correspondence to: H. Qi, State Key Laboratory of Maternal and Fetal Medicine of Chongqing Municipality, The First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, China.
** Correspondence to: R. Saffery, Molecular Immunity, Murdoch Children's Research Institute, Royal Children's Hospital, Melbourne, Victoria, Australia.
*** Corresponding author.
**** Correspondence to: P.N.Baker, College of Life Sciences, University of Leicester, Leicester LE1 7RH, UK.
E-mail addresses: chaotongcqmu@163.com (C. Tong), qihongbo728@163.com (H. Qi), richard.saffery@mcri.edu.au (R. Saffery), philip.baker@leicester.ac.uk (P.N. Baker).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.ygeno.2023.110565
Received 1 October 2022; Received in revised form 4 January 2023; Accepted 19 January 2023
Available online 20 January 2023

1. Introduction

Congenital heart disease (CHD) is one of the most common birth defects and comprises primarily ventricular septal defects (VSDs), atrial septal defects (ASDs), patent ductus arteriosus and tetralogy of Fallot (TOF) [1]. Approximately 9 per 1000 live births worldwide are complicated with various types of CHD [2]; thus, CHD is a severe challenge for public health systems worldwide. CHD can generally be diagnosed by echocardiography in the second trimester of pregnancy, at approximately 24 weeks gestation. However, this method has limited accuracy and is not yet widely used internationally [3]. The early diagnosis of CHD has improved neonatal outcomes, but diagnostic biomarkers of CHD in early pregnancy are lacking.

While the etiology of CHD remains obscure, several studies have suggested that epigenetic variation in early development may play a critical role in CHD pathogenesis [4]. DNA methylation (DNAm) is the most widely studied epigenetic modification [5]. The remodeling of DNAm patterns is an essential part of early embryonic development [6], including the specification of cardiomyocytes and heart tissues. Therefore, it is plausible that the disruption of this process could promote a range of heart pathologies [7]. Previous studies have reported that TOF was associated not only with increased DNAm at *NKX2–5* and *HAND1* but also with decreased methylation at *TBX20* [8] and *LINE-1* repeats [9] in heart tissue from patients whose ages ranged from 1 to 48 months. Furthermore, *BRG1* was hypomethylated in the myocardium of CHD patients whose ages ranged from 1 to 156 months [10]. Aberrant methylation of the *CITED2* gene in the myocardium of CHD patients can dramatically decrease the transcription of this gene, which is closely associated with heart development [11]. These data suggest that variations in DNAm may play a role in the development of CHD.

Twins generally display the same range of phenotypes associated with complex diseases as singletons but offer considerable advantages for investigating the origin of such conditions [12]. Dizygotic (DZ) twins, which arise from two separate fertilization events, share half of their genetic variation, on average [13]. However, monozygotic (MZ) twins, which result from the mitotic division of a single embryo, are genetically identical [14]. Despite this genetic identity, adult MZ twins often show variations in the phenotypes of a range of common diseases [15]. Since MZ twins share the same genetic information, the discordant MZ study design has tremendous value for understanding the relative contributions of genetic and nongenetic determinants of disease. This includes the gene–environment interactions associated with a range of epigenetic modifications, as demonstrated by the growing number of studies profiling DNAm variations in MZ twins discordant for conditions such as systemic lupus erythematosus (SLE), rheumatoid arthritis (RA) and dermatomyositis (DM) [16,17].

MZ twins discordant for CHD are relatively rare but offer a unique opportunity to gain insight into the etiology of CHD. A previous study identified the hypermethylation of the *ZIC3* and *NR2F2* promoters in whole blood samples from a single pair of MZ twins discordant for double outlet right ventricle (DORV) [18].

Variations in DNAm levels early in development may correlate with CHD discordance in MZ twins. In this study, we analyzed the genome-wide DNAm profiles of MZ twins discordant for CHD and identified DNAm variations associated with CHD. We hope that our work can provide a new perspective for subsequent studies on the diagnosis and treatment of this disease.

2. Methods

2.1. Patients and samples

The three MZ twin pairs discordant for CHD were participants in the Chongqing Longitudinal Twin Study (LoTiS), a prospective mother–offspring cohort study conducted in Western China. The co-twins and parents were confirmed to be healthy (Fig. 1A). The detailed clinical information of the three MZ twin pairs is summarized in Table 1. Cord blood from each neonate was collected immediately after delivery. It was processed for buffy coat and plasma isolation and then stored at −80 °C for further use. CHD was diagnosed by ultrasound results according to published guidelines [19,20]. Human fetal heart samples diagnosed with CHD ($n = 5$) were obtained from 18 to 24 weeks of gestational age. The control group (n = 5) was recruited after elective surgical abortions between 10 and 18 weeks of gestational age. In order to prevent confounding by sex, the samples analyzed were all from females. All fetal heart samples were obtained immediately after fetal delivery and stored in liquid nitrogen until further use. The study was approved by the Ethics Committee of the First Affiliated Hospital of Chongqing Medical University (No. 2021–646). Written informed consent was obtained from the parents of all study subjects. All procedures were performed in accordance with the principles stated in the Declaration of Helsinki.

2.2. Quantitative real-time PCR

Total RNA was extracted with TRIzol reagent (Invitrogen) from each sample according to the manufacturer's instructions. The quality and concentration of the total RNA were measured by using ultraviolet spectroscopy (Nano Drop 2000, Thermo). cDNA was synthesized using the PrimeScript™ RT Reagent Kit with gDNA Eraser (TaKaRa) following the manufacturer's protocol. qRT–PCR was conducted using TB Green™ Premix Ex Taq™ (Takara) on the Bio-Rad CFX Connect™ Real-Time System (Bio-Rad, USA). All the primers used for qRT–PCR were synthesized by Sangon Biotech Co. Ltd. (Shanghai, China), and the primer sequences and product lengths are listed in Appendix Table I. The target mRNA was analyzed by the $2^{-\Delta\Delta CT}$ method to evaluate the relative gene expression levels of the NC group and CHD group. Statistical analysis of the data was performed with Prism 9.0 (GraphPad Software, La Jolla, CA, USA) using Student's *t*-tests. Differences were considered significant at $p < 0.05$.

2.3. Methylated DNA Immunoprecipitation (MeDIP) qPCR

Methylated DNA was immunoprecipitated with the EpiQuicK™ Tissue MeDIP Kit (Epigentek, Farmingdale, USA) according to the manufacturer's instructions. Briefly, genomic DNA was extracted from each heart tissue sample and then fragmented by ultrasonication to a size range of 200–1000 bp. A portion of the genomic DNA fragment sample was preserved as input control (whole-genome reference) for normalization. The rest of the sonicated DNA was denatured at 95 °C for ten minutes, incubated with anti-5-methylcytosine (5-mC) antibodies (Epigentek, Farmingdale, USA) at 4 °C overnight, and then washed twice with antibody buffer and wash buffer, respectively. The beads with the captured immunoprecipitated DNA were treated with proteinase K for 1 h at 65 °C, and the methylated DNA was then purified and eluted from the beads for qPCR amplification. The validity of the MeDIP method was determined by qPCR to amplify the imprinting control region (ICR) at the H19 locus (as a control methylated gene) and GAPDH (as a control unmethylated gene). The sequences of the primers are listed in Table 2. The relative methylation levels of the target regions and their percent methylation in the heart tissue was calculated as IP DNA/Input DNA = $2^{-\Delta CT(IP-Input)}$.

2.4. DNA extraction and quantification

Genomic DNA was extracted from buffy coat samples by using the QIAamp DNA Blood Mini Kit (Qiagen, Cat no. 51104) according to the manufacturer's recommendations. Genomic DNA quality was assessed by agarose gel electrophoresis and a NanoPhotometer® spectrophotometer (IMPLEN, CA, USA). DNA concentrations were precisely measured by a Qubit® DNA Assay Kit in Qubit® 2.0 Fluorometer (Life Technologies, CA, USA).

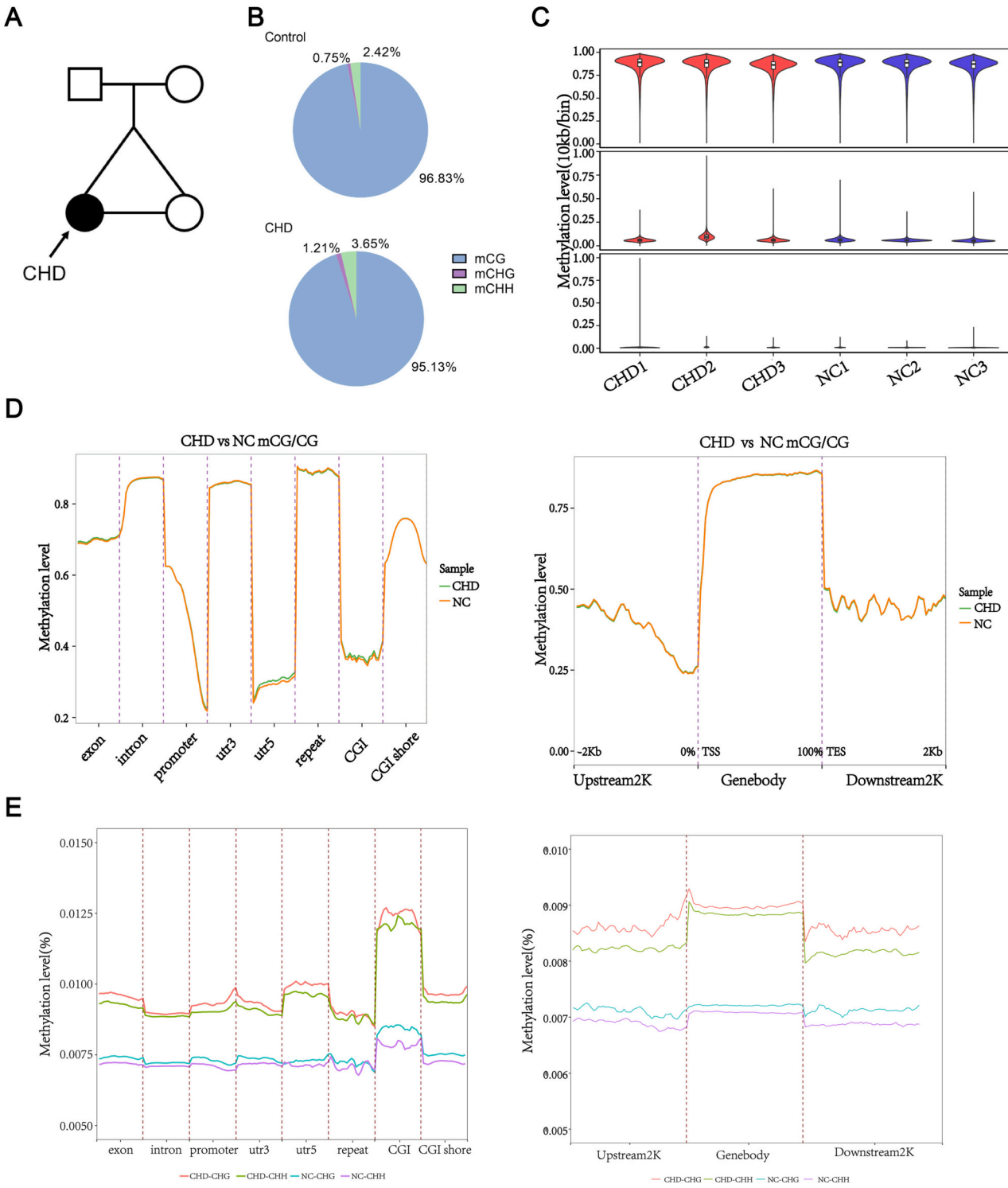

Fig. 1. Identification of discordant CHD twins and summary of differential DNAm.

A, Patient pedigree. B, The average whole-genome DNAm ratios in the CHD and NC groups. The methylated cytosines in the CG context (mCG), methylated cytosines in the CHG context (mCHG) and methylated cytosines in the CHH context (mCHH) are indicated by blue, purple and green colors, respectively. C, Genome-wide distribution of methylated C in different contexts (CG, CHG and CHH) between CHD (red) and NC (blue) cotwins. The x-axis represents the different samples. The y-axis represents the level of methylation. The calculation method was 10 Kbp/bin, and the methylation level in each bin was calculated. D, The distribution of the CG methylation levels (y-axis) within genomic features (x-axis) according to gene location, repeat DNA or CG context (CpG island CGI; or CpG shore). Gene-associated features include exon, intron, promoter, UTR3', and UTR5' regions relative to the position of nearby genes (right). Average distribution of CG methylation sites over the gene body, upstream region (−2000 bp upstream of TSS) and downstream region (+2000 bp downstream of TTS) in the CHD (green) and NC (orange) groups (left). E, The distribution of non-CG methylation in functional regions (right). Average distribution of non-CG methylation in the gene body, upstream and downstream regions (left).

TSS, transcription start site. TTS, transcription termination site. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Table 1
Clinical manifestations in the three pairs of MZ twins discordant for CHD.

	Pair-1		Pair-2		Pair-3	
Mothers						
Prepregnancy age, years	24		34		39	
Prepregnancy BMI	16.03		22.77		20.57	
Chorionicity	Monochorionic		Monochorionic		Monochorionic	
Past or family history of heart disease	No		No		No	
Medication history	No		No		No	
Parity	0		0		1	
Gestational weight gain, kg	14		18		21	
Gestational age, weeks	36		35		37	
Cesarean delivery	Yes		Yes		Yes	
Fetuses						
Sex	Female	Female	Female	Female	Female	Female
Birth weight, g	2020	1900	2600	2430	1800	2090
Birth length, cm	42	42	46	45	46	46
Apgar						
1 min	9	7	10	10	9	9
5 min	10	9	10	10	10	10
10 min	10	9	10	10	10	10
Phenotype	Normal	Tetralogy of Fallot Atrial septal defect Patent ductus arteriosus	Normal	Endocardial cushion defect Mirror dextrocardia Persistent truncus arteriosus	Normal	Atrial septal defect Ventricular septal defect Patent ductus arteriosus Pulmonary hypertension

Table 2
Primers used for MeDIP qPCR.

Genes	Forward primer	Reverse primer	Product size (bp)
SPESP1/NOX5	CCTTTCGGCCTTGAGGTTCC	GCCATAGCAAAAGCGCAACTA	117
CERS1/GDF1	AGCACAGTCCAACCCTGAAG	GCTGTAATCGGGAAGAGGCA	165
AFF2	GCCGCTCCCACTAGCTGATT	CACGTTATCATCCCCGGATCT	103
PLAGL1	CAAAGCCAAGGTCGCCCAG	GGAGACTTCGGCTAGCAGGC	89
GNAS	GAGACACCGTTGAAATGTGCG	GGTCCCGCGGCTTAATTGT	154
JADE3	CCGAGAGCGTGTGTGGG	CCGCAACAGTTATGGCGCTC	156
GAPDH	CCACAGTCCAGTCCTGGGAACC	GAGCTACGTGCGCCCGTAAAA	183
H19-ICR	CAGGTCGGGCATTATCCAC	GCTGTCCTTAGACGGAGTCG	175

2.5. Whole-genome bisulfite sequencing (WGBS), library preparation and quantification

A total of 5.2 μg genomic DNA mixed with 26 ng lambda DNA (negative control) was fragmented by sonication to fragments of 200–300 bp with a Covaris S220 Focused-ultrasonicator (Covaris, Woburn, MA, USA), followed by end repair and adenylation. Cytosine-methylated barcodes were ligated to sonicated DNA according to the manufacturer's instructions. Then, these DNA fragments were treated twice with bisulfite using an EZ DNAm-GoldTM Kit (Zymo Research, CA, USA) before PCR amplification of the resulting single-strand DNA fragments using KAPA HiFi HotStart Uracil + ReadyMix (2×) (KAPA Biosystems, MA, USA). The library concentration was quantified by Qubit® 2.0 Fluorometer (Life Technologies, CA, USA) and quantitative PCR (Q-PCR). The insert size was assayed on an Agilent Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA, USA) and subsequently sequenced by the Illumina HiSeq 2500/4000.

2.6. Quality control

First, we used FastQC (fastqc-v0.11.5) to perform basic statistics on the quality of the raw reads. Then, the read sequences produced by the Illumina pipeline in FASTQ format were preprocessed using Trimmomatic (Trimmomatic-0.36) software. The remaining reads that passed all the filtering steps were counted as clean reads, and all subsequent analyses were based on these reads. Finally, we used FastQC to perform basic statistics on the quality of the clean reads.

2.7. Read mapping to the reference genome

Bismark software version 0.16.3 was used to perform alignments of bisulfite-treated reads to a reference genome. The reference genome was first transformed into a bisulfite-converted version (C-to-T and G-to-A converted) and then indexed using Bowtie2 [21]. Sequence reads were also transformed into fully bisulfite-converted versions (C-to-T and G-to-A converted) before they were aligned to similarly converted versions of the genome in a directional manner. Sequence reads that produced a unique best alignment from the two alignment processes (original top and bottom strand) were then compared to the normal genomic sequence, and the methylation states for all cytosine positions in the read were inferred. Identical reads that aligned to the same regions of the genome were regarded as duplicates.

2.8. Estimating the methylation level

To identify true methylation sites, a binomial (Bln) distribution test was performed for each cytosine site using the following formula:

$$mC \sim Bln(mC + umC^{*}r)$$

where u represents the read depth of the site and r is the bisulfite non-conversion rate.

For the identified methylation sites, the methylation levels were calculated with the following formula:

$$ML(C) = \frac{reads(mC)}{reads(mC) + reads(C)}$$

where ML is the methylation level, and mC and umC represent the

number of methylated cytosines and unmethylated cytosines, respectively. The calculated ML was further corrected with the following formula by referring to previous studies (Lister et al., 2013):

$$ML(corrected) = \frac{ML - r}{1 - r}$$

where r is the bisulfite nonconversion rate.

We obtained a mean of 391,563,070 (370,105,814-440,111,314) raw reads for each sample. After filtering out the low-quality data, an average of 328,514,621 (357,203,127–432,005,207) clean reads were generated for each sample. Then, the reads were uniquely mapped to the human reference genome (hg38), with genome mapping rates ranging from 85.84% to 86.95%. The C site mean sequencing depth of all the samples was 11.3 (10.8–13.1) reads. The average percentage of C sites with a sequencing depth at least 5-fold was 89.08% and 89.87% in the CHD and NC groups, respectively (Appendix Table II).

2.9. Differentially methylated region identification

Differentially methylated regions (DMRs) are genomic regions with distinct DNAm levels under different conditions. We used dispersion shrinkage for sequencing data (DSS) software [22] to identify the DMRs. Briefly, differentially methylated loci (DML) were first identified with the Wald test based on the site methylation level data. Hypothesis testing for DML was performed as follows: H0: μi1 = μi2, where μi1 and μi2 represent the methylation level of specific C sites in the case and control groups, respectively. The DML in the present study are defined as the C sites with $|\mu i1 - \mu i2| \geq 0.1$ and an adjusted *P* value less than 1e-05 (the *P* values are adjusted by the Benjamini–Hochberg method). For DMR identification, the standards were as follows: the percentage of DML in this region was above 50%; the numbers of C sites in this region were above 3; the length of DMR was longer than 50 bp; and two DMRs were merged if the distance was <100 bp. DMRs were identified by using the callDMR function in DSS software, with the following parameters: smoothing. Span = 200, delta = 0, p.threshold = 1e05, minlen = 50, minCG = 3, dis.merge = 100, and pct.sig = 0.5. We used Fisher's test to identify the DMRs.

2.10. Annotation of DMR-related genes

The DMRs in the genome were functionally annotated as promoter, exon, intron, transcription start site (TSS) or transcription end site (TES). Based on the distribution of the DMRs, any differentially methylated gene (DMG) whose gene body region (from the TSS to the TES) or promoter region (2 kb upstream from the TSS) overlapped with the DMR was defined as a gene related to that DMR. Then, the significant DMRs were screened according to the criteria of a differentially methylated level (diff. Methy) > 0.20 and areaStat >250.

2.11. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of DMR-related genes

KEGG is a useful database resource for understanding the high-level functions and utilities of biological systems from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies. We used Database for Annotation, Visualization, and Integrated Discovery (DAVID) software to conduct GO and KEGG analyses of the selected DMR-related genes. GO enrichments were performed using the GOTERM_BP_FAT database under default parameters, and the pathway enrichments were performed using the KEGG database. Bioinformatic analysis was performed using the OmicStudio tools at https://www.omicstudio.cn/tool.

3. Results

3.1. Global DNAm profiles and differences in MZ twins

There was no difference in the mean total cytosine methylation level between the CHD and NC groups (mean 4.55 ± 0.26% in the CHD groups vs. 4.44 ± 0.08% in the control group, $p = 0.5223$) (Fig. 1B and Appendix Table II).

Overall, there was a similar distribution of methylation levels between CHD and NC individuals, with much higher DNAm levels in the CG context than at CHG and CHH (where H is A, C, or T) sites (Fig. 1C). To more comprehensively characterize the distribution of DNAm in the CHD group relative to the NC group, we assigned the methylation sites to different functional genomic features. As anticipated, the methylation levels were higher in gene introns, 3' UTRs and repeat regions than in gene promoters, 5'UTRs and CpG islands (CGIs) in all individuals (Fig. 1D). The gene terminal methylation density both upstream and downstream of transcription initiation and termination sites (TSS and TTS, respectively) was lower than that in the gene body region in CG contexts, but this measure did not differ between the groups (Fig. 1D). In contrast, the non-CG DNAm level was higher in the CHD group than in the NC group (Fig. 1E).

3.2. Differentially methylated regions (DMRs) and genes (DMGs) associated with CHD: between-pair analysis

DMRs are genomic regions containing multiple contiguous/clustered sites of differential DNAm variation between groups. Many studies have previously associated variations in DNAm at DMRs with gene expression and/or the etiology of various diseases [23].

A total of 379 significant DMRs were identified between CHD and NC individuals (average DMR difference 27%) following between-pair analysis. These DMRs included 299 hypermethylated and 80 hypomethylated regions in CHD individuals relative to NC individuals (Fig. 2A). Of these, 153 DMRs comprised methylation solely in the CG context, 26 DMRs comprised CHG sites, and 200 DMRs comprised CHH sites (Fig. 2B). More DMRs were more frequently located in gene promoter, exon and intron regions than in the 3' UTRs (Fig. 2C).

DMRs were next assigned to specific genes based on their location relative to annotated gene loci. Of the 379 total identified DMRs, 175 could be assigned to specific genes (132 hyper and 43 hypo-DMRs, Appendix Table III).

To probe the gene functions of CHD-associated DMGs, GO and KEGG pathway analyses were performed. The GO analysis revealed that DMGs were significantly enriched in biological processes involved in histone methylation, actin cytoskeleton organization, regulation of cell differentiation, and transcription (Fig. 2D). Adrenergic signaling in cardiomyocytes, the regulation of actin cytoskeleton and signaling pathways (e.g., Wnt, PI3K-Akt, Rap1, MAPK) were also found to be enriched through KEGG analysis. Each of these pathways have previously been linked to cardiac development. Furthermore, the DMGs were enriched in certain cellular functions, including focal adhesion, gap junction and tight junction (Fig. 2E).

3.3. Identification of CHD-associated DMRs and DMGs between cotwins: within-pair analysis

Subsequently, we analyzed each MZ pair in detail to further explore the role of DNAm in the development of CHD and control for other confounding factors, such as genetics. Comparing the DNAm level between each MZ twin within a pair, we found that both the hyper and hypo-DMRs were differentially distributed in relation to genomic features, including genes (Appendix Fig. 1A), and the majority of these DMRs were located in the promoter, exon and intron regions (Appendix Fig. 1B). In total, we identified 220 hypermethylated and 422 hypomethylated DMGs within Pair 1, 1160 hypermethylated and 75

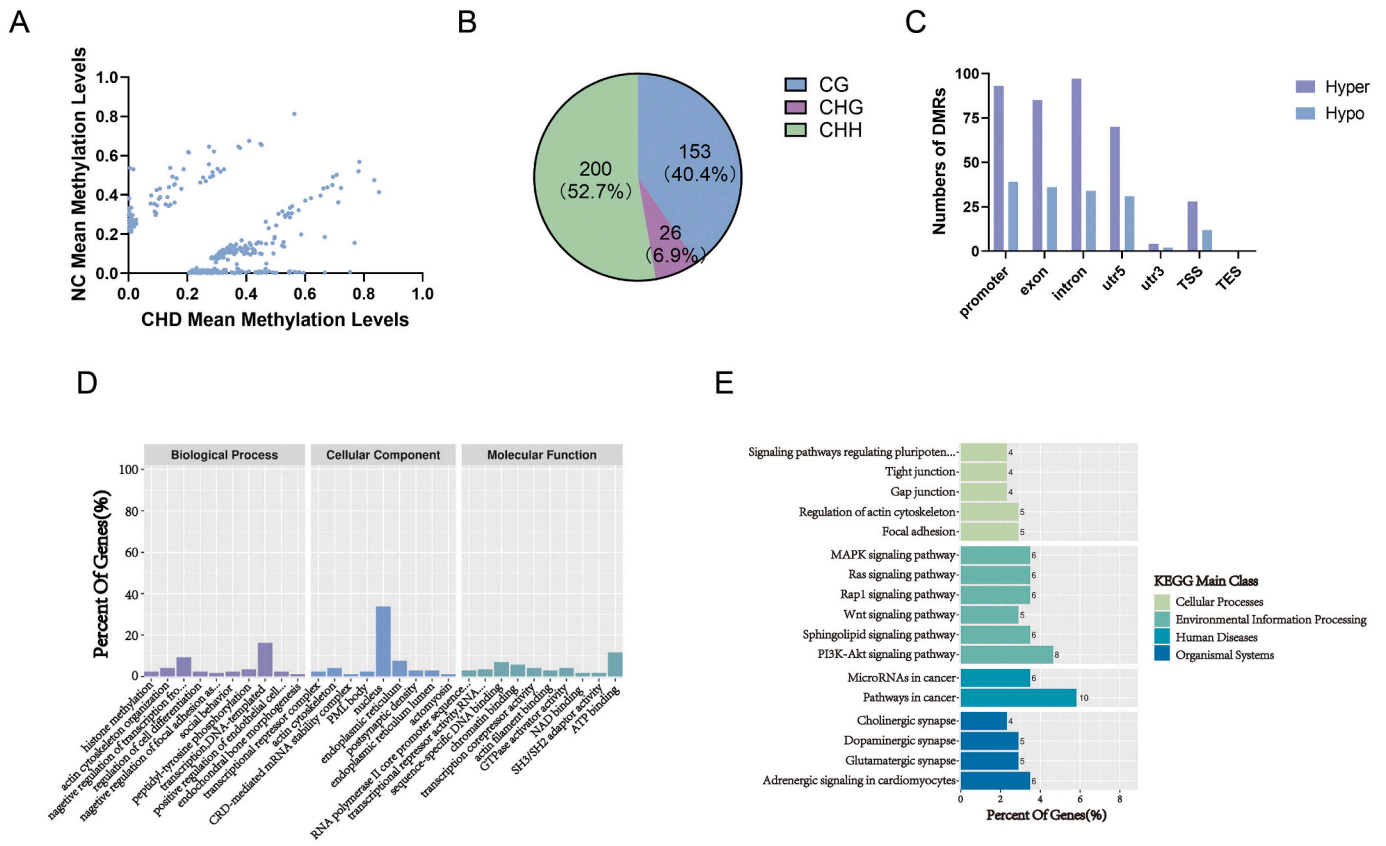

Fig. 2. Comparison of the DMRs and functional analysis of the DMGs of the cotwins.
A, Scatter plot of DMRs with differences >0.20 and areaStat >250. B, The number and ratio of identified DMRs in CG (blue), CHG (purple) and CHH (green) sites. C, Characteristics of genomic elements subject to differential methylation (higher and lower) within the MZ twins. D, Gene Ontology analysis of the differentially methylated genes. A total of 10 significantly enriched biological process, cell component and molecular function terms are shown. E, Kyoto Encyclopedia of Genes and Genomes pathway analysis of the differentially methylated genes. The Kyoto Encyclopedia of Genes and Genomes pathways were divided into the following subcategories: cellular processes, environmental information processing, human diseases, and organismal system. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

hypomethylated DMGs within Pair 2, and 294 hypermethylated and 132 hypomethylated DMGs within Pair 3 (Appendix Table IV). Next, the hyper- and hypomethylated DMGs in each pair were subjected to GO analysis. The hypermethylated DMGs in Pair 1 were involved in various stages of cardiovascular system development and heart development. Those in Pair 2 were enriched in several biological processes, including aorta morphogenesis, cell migration and differentiation, and angiogenesis; and more hypermethylated DMGs were enriched in cardiac conduction and ion transport in Pair 3 (Appendix Fig. 1C). Moreover, we found that the hypomethylated DMGs in each twin pair were mostly associated with histone modifications, DNA methylation, heart rate regulation, myocardial tissue development and angiogenesis (Appendix Fig. 1C).

3.4. CHD-associated differential methylation is common in all discordant twin pairs

We then focused on genes previously identified in CHD with aberrant DNAm that have been reported to play a critical role in morphogenesis and the establishment of cardiac development. We found that *TBX1* (T-box transcription factor 1), *GRB10* (Growth factor receptor-bound protein 10), and *NR2F2* (Nuclear receptor subfamily 2 group F member) showed a higher DNAm in the CHD group, while *RUNX1* (Runt-related transcription factor 1), *MEST* (mesoderm specific transcript), and *CITED1* (Cbp/p300 interacting transactivator with Glu/Asp rich carboxy-terminal domain 1) were hypomethylated in the CHD group (Appendix Table V). Utilizing the ENCODE Project, which includes the public data of embryo and adult heart tissues (Accession IDs: ENCSR863BUL, ENCSR047LLJ, ENCSR693CSQ, and ENCSR635GTY), to identify the possible role of these genes in cardiac development. We found that these genes were expressed in the heart tissue throughout life (Appendix Table V). This finding implies DNAm may contribute to CHD by regulating the expression of genes implicated in embryonic and cardiac development.

To further characterize the relationship between specific aberrant cord blood DNAm sites and CHD, we compared the significant DMGs in each MZ pair and found that 8 hypermethylated DMGs and 1 hypomethylated DMG were present in all three discordant MZ twin pairs (Fig. 3A). The difference in the methylation level of the *DLGAP2* gene was not statistically significant ($p = 0.2774$) (Appendix Fig. 2). Ultimately, we identified 8 candidate CHD-related DMGs, including *SPESP1*, *NOX5*, *GDF1*, *CERS1*, *GNAS*, *PLAGL1*, *AFF2* and *JADE3*. Of these genes, 7 showed significantly higher methylation in the MZ twins with CHD than in the control individuals, and one showed evidence of lower methylation in association with CHD (*JADE3*) (Appendix Fig. 2). In addition, *SPESP1*, *NOX5*, *GNAS*, *PLAGL1*, *AFF2* and *JADE3* were located in the promoter regions, and two other DMGs (*GDF1* and *CERS1*) were located in the gene body.

3.5. Aberrant methylation and gene expression of DMGs in additional CHD fetuses

To further confirm the dysregulation of these genes in CHD, we collected heart tissues from CHD fetuses and normal fetuses. Using

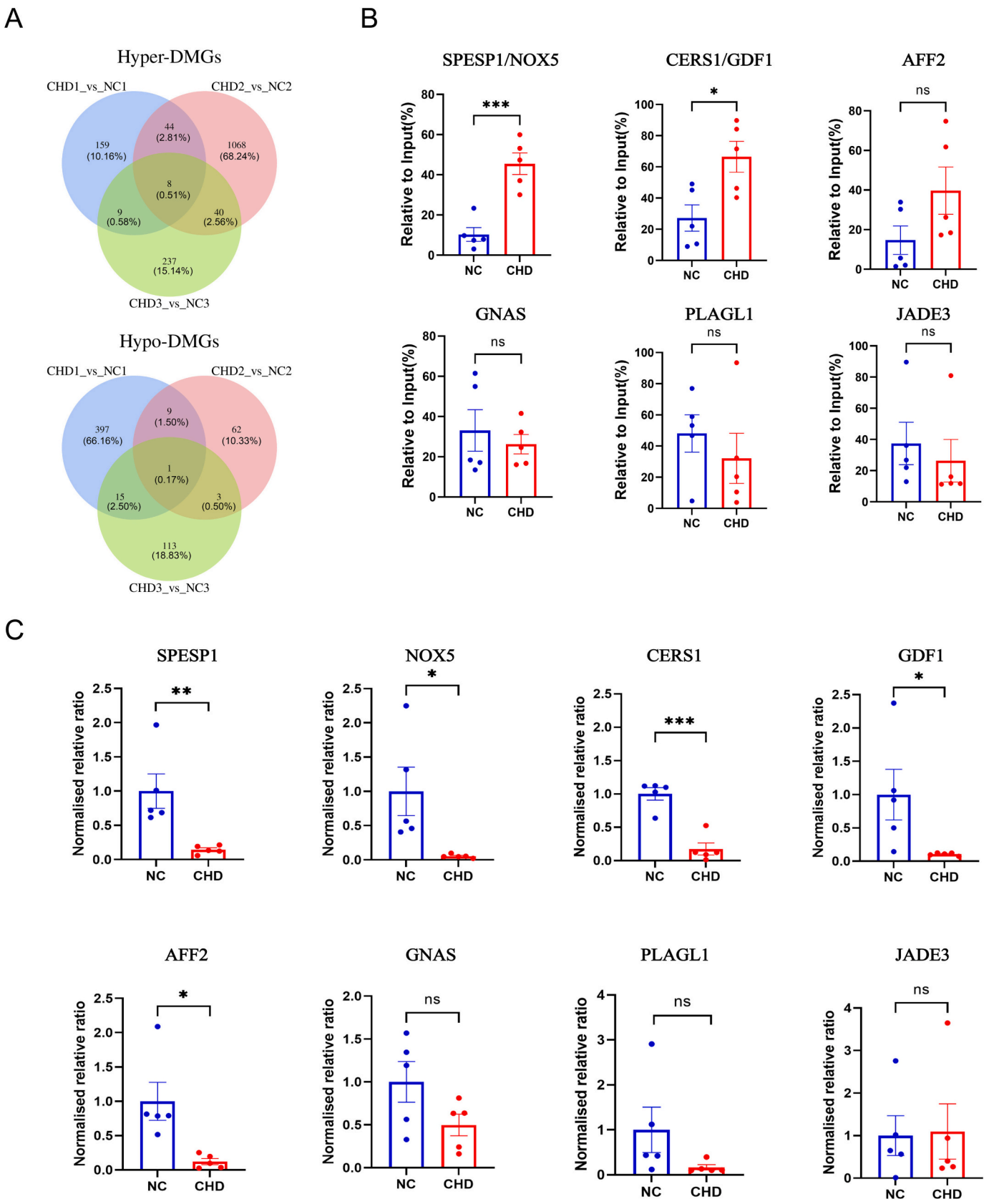

Fig. 3. Candidate DMGs in MZ twins discordant for CHD.

A, Venn diagram comparing the hypermethylated and hypomethylated DMGs among each MZ twin pair. B, The relative methylation level of candidate DMGs in additional individuals quantified by MeDIP qPCR. Data are presented as the mean ± SEM, and 2-tailed *t*-tests were used. *$P < 0.05$; ***, $P < 0.005$; ns, not significant, $P>0.05$. C, The relative gene expression levels of the DMGs in the CHD and NC specimens. Data are presented as the mean ± SEM, and 2-tailed t-tests were used. *$P < 0.05$; ** $P < 0.01$; ***, $P < 0.005$; ns, not significant, $P>0.05$.

MeDIP qPCR, we confirmed that *SPESP1/NOX5* and *CERS1/GDF1* were hypermethylated in the CHD group compared to the normal group (Fig. 3B). Correspondingly, gene expression analysis for the 8 candidate CHD-related DMGs showed that five out of seven hypermethylated DMGs, *SPESP1*, *NOX5*, *GDF1, CERS1* and *AFF2,* had significantly lower expression levels in the CHD group than in the NC group ($p < 0.05$). While the expression level of *JADE3*, the only hypomethylated DMG, was higher in the CHD group than in the NC group, this difference was not statistically significant ($p \geq 0.05$) (Fig. 3C). These results demonstrated that lower gene expression levels of *SPESP1* and *NOX5* are associated with aberrant hypermethylation of these genes in CHD patients.

4. Discussion

Accumulating data suggest that genome methylation changes may be a potential mechanism in the development of a range of congenital malformations, including cardiac defects [24]. In pregnancy, maternal folic acid supplementation may impact the risk of CHD by regulating DNAm [25]. Atypical DNAm patterns have previously been identified in neonatal blood spots, placental tissues and heart tissues from CHD patients and may be a biomarker for the prediction of CHD [26,27]. DNAm of newborn cord blood has been shown to be associated with a range of in utero exposures and offspring outcomes [28,29]. However, the genome-wide DNAm profile of cord blood in CHD patients has not been reported.

Several longitudinal cohort studies have previously linked patterns of DNAm in peripheral blood with the risk of CHD [30]. Furthermore, methylation variations can be detected in peripheral leukocytes before vascular lesions occur [31]. During embryonic development, the mesoderm forms the progenitor cells of the musculoskeletal system (bone, cartilage and muscle), cardiovascular system (heart, blood and blood vessels), and connective tissues throughout our bodies [32,33]. A common mesodermal origin of all cardiovascular lineages provides the developmental context to tie these interdependent cell fates together [33]. We investigated whether the blood methylation profile reflects that of the developing heart and whether it is potentially related to the origin of mesoderm. The present study profiled whole-genome DNAm in the umbilical cord blood from CHD-discordant MZ twins using WGBS to test for evidence of a common epigenetic origin of these conditions. The methylation level and gene expression level of the related genes was then validated in human fetal heart tissue from individuals diagnosed with CHD.

Here, we reported that the proportion of methylated cytosine across the genome (reflecting average genomic methylation) was similar between CHD and NC individuals, suggesting that CHD is potentially not associated with global changes in DNAm status. As anticipated based on previous studies, the overall levels of CHG and CHH methylation were far lower than those in the CG context in all cases. The overall levels of CHG and CHH sites were found to be significantly lower in individuals with CHD. A previous report using WGBS detection of hypertrophic myocardium tissue in spontaneously hypertensive rats found that the loss of methylation at CHG and CHH sites was decreased, which may be related to cardiac homeostasis [34]. More interestingly, the tissues enriched in non-CG DNAm have previously been found to be associated with cardiomyocyte differentiation [35]. However, the potential role of CHG and CHH methylation in heart development remains to be further explored.

Using between-pair analysis, we identified 379 DMRs overall by comparing the whole-genome methylation of the CHD and NC groups, and these DMRs map to 175 meaningful DMGs. In addition, these DMGs were significantly enriched for adrenergic signaling in cardiomyocytes, cardiac development functions, and cellular functions, such as focal adhesion, gap junctions and tight junctions. DMGs associated with CHD from within-pair analysis were significantly enriched in GO terms directly related to cardiovascular system development, heart development and functions such as cell migration, differentiation, and adhesion.

Perhaps the most compelling finding of this analysis was the independent evidence of common DMGs in CHD neonates relative to their siblings. In total, 8 genes were differentially methylated. Of these genes, 7 demonstrated higher DNAm levels in CHD individuals, and 1 demonstrated lower DNAm levels in CHD individuals. Some of these genes have previously been linked to CHD pathogenesis. Importantly, *SPESP1/NOX5* and *CERS1/GDF1*, 2 of the 8 DMGs, were also found to be hypermethylated in additional CHD patients. This finding further supports a link for the pathogenesis of CHD in general, as opposed to that of specific types of CHD. Next, we also assessed the gene expression of the aforementioned 8 genes in human fetal heart tissue. The expression levels of the *SPESP1*, *NOX5*, *GDF1*, *CERS1* and *AFF2* genes were lower in the CHD group than in the NC group. The gene expression levels of *GNAS*, *PLAGL1* and *JADE3* were not significantly different.

The DNAm changes in the promoter region generally act as silencers of downstream genes and are negatively associated with gene expression. However, the DNAm changes in the gene body region are more positively correlated with gene expression [36]. Therefore, after integrated analysis of the DNAm and gene expression levels of the DMGs, we finally screened two CHD-related DMGs that were shared by 2 MZ twins, namely, *SPESP1* and *NOX5*. The gene expression levels of *GNAS*, *PLAGL1*, and *JADE3*, which are located in the promoter region, were not significantly different, but the gene expression levels of the hypermethylated *GDF1* and *CERS1* genes, which are located in the gene body, were lower in the CHD group. These results indicate that different DNAm modulations exist for different genes involved in CHD development. Further studies should verify this hypothesis.

SPESP1 is known to be abundantly expressed in the testis and has been detected in small amounts in the heart muscle, according to the Human Protein Atlas database. *SPESP1* is considered to be an essential factor in initiating fusion with the egg plasma membrane during fertilization. *SPESP1* knockout male mice had significantly reduced fertility, but female knockout mice had no abnormal phenotype [37]. We cannot exclude the possibility of sex differences because our samples were all females. The role of *SPESP1* expression in the heart is unclear. In the present study, we found the hypermethylation of *SPESP1* in the promoter region and its lower gene expression in the CHD group compared to their healthy siblings. These findings suggest a potential role for *SPESP1* in the pathogenesis of CHD.

NOX5, the most recently identified member of the NADPH oxidase family, is a major source of reactive oxygen species (ROS). It functions in cell proliferation, transformation, differentiation, and apoptosis [38]. Thus far, the NOX isoforms *NOX1*, *NOX2*, and *NOX4* have been identified in cardiomyocytes and are important sources of ROS in the heart. *NOX2* and *NOX4* were reported to be involved in proapoptotic and prohypertrophic signaling in cardiomyocytes [39]. In metabolic disease in mice, *NOX1* was found to mediate endothelial activation and contribute to myocardial inflammation and remodeling [40]. As such, the NOX family has been involved in (patho)physiological processes in the heart, but knowledge of calcium-dependent *NOX5* in the heart is limited. To date, the expression of *NOX5* has been detected in human blood vessels, including coronary vessels, and its expression level is altered in cardiovascular disease [41]. Strikingly, the hypermethylation of *NOX5* at the promoter region results in transcriptional silencing and is thought to be associated with CHD [42]. The hypermethylation of *NOX5* might lead to aberrant protein expression and impede normal cardiac development. Thus, it would be worth investigating the role of *NOX5* in CHD.

In addition, this study confirmed abnormal DNAm patterns for certain genes that have previously been reported to be correlated with CHD, including *TBX1*, *GRB10*, *NR2F2*, *RUNX1*, *MEST* and *CITED1* [43]. Nevertheless, not all of the well-recognized epigenetic changes identified in CHD singletons have been observed in CHD-discordant twins, such those for *LINE-1*, *GATA4* and *SCO2*. We speculate that since the similarity of the intrauterine environment between MZ cotwins is much

higher than that among singleton individuals, epigenetic changes triggered by strong environmental stimuli may not occur in twins. Moreover, our study aimed to determine the common epigenetic changes conserved in general CHD; hence, genes involved in the development of specific CHD types might not be discoverable in the MZ twin model and the present sample size. Last, but not least, DNAm in CHD may vary among different populations [44]. The present study is the first to identify epigenetic signatures of CHD in Asian twins and elucidated different methylation patterns from previous studies, which have mainly investigated Caucasian populations. Thus, it will be interesting to further evaluate the impact of the identified DMGs (*SPESP1* and *NOX5*) on CHD development in other ethnic groups.

Our results may provide new insight into the mechanisms by which CHD is related to epigenomic factors. The main strength of this study is the use of a disease-discordant MZ twin model, which controls for genetic factors and shared environmental factors, such as maternal environment (diet, toxins, smoking, etc.). Furthermore, studying individuals early in life (at birth) reduces the impact of stochastic, age and environmental influences on DNAm, which are known as the epigenetic drift [45]. Compared with similar studies, we obtained a relatively large sample size, but there still exist sample size limitations. Further studies in singletons and different stages of development are needed to validate the relevance of these findings in cord blood to the pathogenesis of CHD.

5. Conclusions

Our results suggest that aberrant DNAm patterns and the dysregulation of *SPESP1* and *NOX5* may be linked to atypical heart development in utero. Importantly, the correlation of CHD with changes in the DNAm status of *SPESP1* is reported for the first time. These findings reveal the importance of DNAm to the normal development of heart structure and function. Therefore, this study not only reveals promising novel blood-based biomarkers for the early screening of CHD but also provides novel insights into potential interventional targets for CHD.

Disclosures

None.

Sources of funding

This work was supported by the National Natural Science Foundation of China (U21A20346, 82171662, 82101711, 82001580, and 81520108013), Chongqing Science and Technology Commission (cstc2021ycjh-bgzxm0192), Chongqing Health Committee (2019GDRC012), and Chongqing Education Commission (KJZD-K202100404).

CRediT authorship contribution statement

Xi Yuan: Methodology, Data curation, Writing – original draft. **Jiayu Huang:** Investigation, Data curation. **Li Wen:** Methodology, Data curation. **Boris Novakovic:** Writing – review & editing. **Mark D. Kilby:** Writing – review & editing. **Chao Tong:** Conceptualization, Writing – review & editing. **Hongbo Qi:** Supervision, Funding acquisition. **Richard Saffery:** Conceptualization, Writing – review & editing. **Philip N. Baker:** Conceptualization, Resources.

Data availability

Data are available upon reasonable request. Due to legal restrictions and the nature of the data (individual level data), the data cannot be placed in a public repository or provided as supporting files. No other documents are available. The Chongqing Longitudinal Twin Study welcomes collaboration with investigators interested in this research. Interested investigators should contact the principal investigators, Prof. Chao Tong (chaotongcqmu@163.com) and Prof. Hongbo Qi (qihongbo728@163.com).

Acknowledgments

We are grateful to the families who participated in the Chongqing LoTiS.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2023.110565.

References

[1] X. Zhang, et al., Epidemiology, prenatal diagnosis, and neonatal outcomes of congenital heart defects in eastern China: a hospital-based multicenter study, BMC Pediatr. 20 (2020) 416.
[2] H. Baumgartner, et al., 2020 ESC guidelines for the management of adult congenital heart disease, Eur. Heart J. 42 (2021) 563–645.
[3] B. Ulm, et al., Diagnostic quality of 3Tesla postmortem magnetic resonance imaging in fetuses with and without congenital heart disease, Am. J. Obstet. Gynecol. 225 (2021) 189.e1–189.e30.
[4] B.N. Akerberg, et al., Genetic and epigenetic control of heart development, Cold Spring Harb. Perspect. Biol. 12 (2020).
[5] G.D. Ginder, et al., The role of the epigenetic signal, DNA methylation, in gene regulation during erythroid development, Curr. Top. Dev. Biol. 82 (2008) 85–116.
[6] P.A. Jones, Functions of DNA methylation: islands, start sites, gene bodies and beyond, Nat. Rev. Genet. 13 (2012) 484–492.
[7] M. Liu, et al., H3K4 di-methylation governs smooth muscle lineage identity and promotes vascular homeostasis by restraining plasticity, Dev. Cell 56 (2021) 2765–2782.e10.
[8] W. Sheng, et al., Association of promoter methylation statuses of congenital heart defect candidate genes with tetralogy of Fallot, J. Transl. Med. 12 (2014) 31.
[9] W. Sheng, et al., LINE-1 methylation status and its association with tetralogy of fallot in infants, BMC Med. Genet. 5 (2012) 20.
[10] Y. Qian, et al., Hypomethylation and decreased expression of BRG1 in the myocardium of patients with congenital heart disease, Birth Defects Res. 109 (2017) 1183–1195.
[11] M. Xu, et al., CITED2 mutation and methylation in children with congenital heart disease, J. Biomed. Sci. 21 (2014) 7.
[12] J. van Dongen, et al., The continuing value of twin studies in the omics era, Nat. Rev. Genet. 13 (2012) 640–653.
[13] E. Generali, et al., Lessons learned from twins in autoimmune and chronic inflammatory diseases, J. Autoimmun. 83 (2017) 51–61.
[14] M.F. Fraga, et al., Epigenetic differences arise during the lifetime of monozygotic twins, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 10604–10609.
[15] J.E. Castillo-Fernandez, et al., Epigenetics of discordant monozygotic twins: implications for disease, Genome Med. 6 (2014) 60.
[16] J.T. Bell, et al., The value of twins in epigenetic epidemiology, Int. J. Epidemiol. 41 (2012) 140–150.
[17] Q. Tan, The epigenome of twins as a perfect laboratory for studying behavioural traits, Neurosci. Biobehav. Rev. 107 (2019) 192–195.
[18] G. Lyu, et al., Genome and epigenome analysis of monozygotic twins discordant for congenital heart disease, BMC Genomics 19 (2018) 428.
[19] International Society of Ultrasound in O, et al., ISUOG Practice Guidelines (updated): sonographic screening examination of the fetal heart, Ultrasound Obstet. Gynecol. 41 (2013) 348–359.
[20] L.J. Salomon, et al., Practice guidelines for performance of the routine mid-trimester fetal ultrasound scan, Ultrasound Obstet. Gynecol. 37 (2011) 116–126.
[21] B. Langmead, et al., Fast gapped-read alignment with bowtie 2, Nat. Methods 9 (2012) 357–359.
[22] Y. Park, et al., Differential methylation analysis for BS-seq data under general experimental design, Bioinformatics. 32 (2016) 1446–1453.
[23] D.P. Chen, et al., Methods for identifying differentially methylated regions for sequence- and array-based data, Brief Funct. Genom. 15 (2016) 485–490.
[24] S. Chowdhury, et al., Maternal DNA hypomethylation and congenital heart defects, Birth Defects Res. A Clin. Mol. Teratol. 91 (2011) 69–76.
[25] S.M. González-Peña, et al., Maternal folic acid intake and methylation status of genes associated with ventricular septal defects in children: case-control study, Nutrients. 13 (2021).
[26] J. Zhou, et al., Genome-wide methylation analysis reveals differentially methylated CpG sites and altered expression of heart development-associated genes in fetuses with cardiac defects, Exp. Ther. Med. 22 (2021) 1032.
[27] N. Glezeva, et al., Targeted DNA methylation profiling of human cardiac tissue reveals novel epigenetic traits and gene deregulation across different heart failure patient subtypes, Circ. Heart Fail. 12 (2019), e005765.
[28] J.B. Herbstman, et al., Predictors and consequences of global DNA methylation in cord blood and at three years, PLoS One 8 (2013), e72824.
[29] Y. Jiang, et al., Epigenome wide comparison of DNA methylation profile between paired umbilical cord blood and neonatal blood on Guthrie cards, Epigenetics. 15 (2020) 454–461.

[30] A. Navas-Acien, et al., Blood DNA methylation and incident coronary heart disease: evidence from the strong heart study, JAMA Cardiol. 6 (2021) 1237–1246.
[31] G. Lund, et al., DNA methylation polymorphisms precede any histological sign of atherosclerosis in mice lacking apolipoprotein E, J. Biol. Chem. 279 (2004) 29147–29154.
[32] E. Ferretti, et al., Mesoderm specification and diversification: from single cells to emergent tissues, Curr. Opin. Cell Biol. 61 (2019) 110–116.
[33] K.D. Prummel, et al., The lateral plate mesoderm, Development. 147 (2020).
[34] L. Liu, et al., Regulation of DNA methylation and 2-OG/TET signaling by choline alleviated cardiac hypertrophy in spontaneously hypertensive rats, J. Mol. Cell. Cardiol. 128 (2019) 26–37.
[35] W. Wang, et al., Pten regulates cardiomyocyte differentiation by modulating non-CG methylation via Dnmt3, Adv. Sci. (Weinh). 8 (2021), e2100849.
[36] X. Yang, et al., Gene body methylation can alter gene expression and is a therapeutic target in cancer, Cancer Cell 26 (2014) 577–590.
[37] Y. Fujihara, et al., Sperm equatorial segment protein 1, SPESP1, is required for fully fertile sperm in mouse, J. Cell Sci. 123 (2010) 1531–1536.
[38] K.R. Yeung, et al., DNA methylation profiles in preeclampsia and healthy control placentas, Am. J. Physiol. Heart Circ. Physiol. 310 (2016) H1295–H1303.
[39] C.E. Murdoch, et al., NADPH oxidase-dependent redox signalling in cardiac hypertrophy, remodelling and failure, Cardiovasc. Res. 71 (2006) 208–215.
[40] L. Xu, et al., NOX1 mediates metabolic heart disease in mice and is upregulated in monocytes of humans with diastolic dysfunction, Cardiovasc. Res. 118 (2022) 2973–2984.
[41] D.J. Fulton, Nox5 and the regulation of cellular function, Antioxid. Redox Signal. 11 (2009) 2443–2452.
[42] C. Zhu, et al., DNA hypermethylation of the NOX5 gene in fetal ventricular septal defect, Exp. Ther. Med. 2 (2011) 1011–1015.
[43] J. Cao, et al., The role of DNA methylation in syndromic and non-syndromic congenital heart disease, Clin. Epigenetics 13 (2021) 93.
[44] X. Song, et al., Association of MTHFD1 gene polymorphisms and maternal smoking with risk of congenital heart disease: a hospital-based case-control study, BMC Pregnancy Childbirth. 22 (2022) 88.
[45] J. Kochmanski, et al., Environmental deflection: the impact of toxicant exposures on the aging epigenome, Toxicol. Sci. 156 (2017) 325–335.