Twin Research and Human Genetics
Volume 20 ▌ Number 6 ▌ pp. 521–532 ▌ ▌ doi:10.1017/thg.2017.57

Copy Number Variants and Exome Sequencing Analysis in Six Pairs of Chinese Monozygotic Twins Discordant for Congenital Heart Disease

Yuejuan Xu,[1,#] Tingting Li,[1,#] Tian Pu,[1] Ruixue Cao,[1] Fei Long,[2] Sun Chen,[1] Kun Sun,[1] and Rang Xu[3]
[1]Department of Pediatric Cardiology, Xinhua Hospital, affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China
[2]School of Biomedical Science, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China
[3]Scientific Research Center, Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China

Congenital heart disease (CHD) is one of the most common birth defects. More than 200 susceptibility loci have been identified for CHDs, yet a large part of the genetic risk factors remain unexplained. Monozygotic (MZ) twins are thought to be completely genetically identical; however, discordant phenotypes have been found in MZ twins. Recent studies have demonstrated genetic differences between MZ twins. We aimed to test whether copy number variants (CNVs) and/or genetic mutation differences play a role in the etiology of CHDs by using single nucleotide polymorphism (SNP) genotyping arrays and whole exome sequencing of twin pairs discordant for CHDs. Our goal was to identify mutations present only in the affected twins, which could identify novel candidates for CHD susceptibility loci. We present a comprehensive analysis for the CNVs and genetic mutation results of the selected individuals but detected no consistent differences within the twin pairs. Our study confirms that chromosomal structure or genetic mutation differences do not seem to play a role in the MZ twins discordant for CHD.

■ **Keywords:** congenital heart disease, discordant monozygotic twins, genetic variation difference, SNP genotyping array, exome sequencing

Congenital heart diseases (CHDs) are the most commonly identified human birth defects. Approximately 0.8% of all live-born fetuses have some type of CHDs (Hoffman et al., 2004). CHDs arise from abnormal embryonic heart development. Generally, CHDs are induced either by environmental influences (teratogens, maternal exposures, and infectious agents), by altered gene function or amount, or by combinations of those factors (Breckpot et al., 2012; Bruneau, 2008; Jenkins et al., 2007). Although CHDs seldom exhibit a clear familial inheritance pattern, epidemiological studies strongly suggest that genetic factors are the predominant cause of CHD (Gelb & Chung, 2014). There is more evidence for familial aggregation of CHD and a higher risk of recurrence in the offspring (Marian, 2014). Animal models and human genetic studies have implicated mutations and copy number variations of numerous genes in CHDs (Breckpot et al., 2012; Bruce et al., 2013). Collectively, these genes are still unable to account for the population prevalence of CHD (Bruce et al., 2013). Furthermore, the cause of >80% of CHDs remains unexplained (Gelb & Chung, 2014).

Monozygotic (MZ) twins have long served as a model to estimate the contribution of genetic and environmental factors for many complex traits (Kimani et al., 2009). Although MZ twins originate from the same zygote, the discordance for structural malformations (including CHDs) has a high rate of 80% in MZ twins (Hajdu et al., 2006). The cause of phenotypic discordance between monozygotic twins is unknown. It is often attributed to differential environmental exposure (Wong et al., 2005). However, the underlying genetic difference arising during the twinning process and/or embryonic development could be involved, such as chromosomal mosaicism (e.g., aneuploid) and de novo somatic

RECEIVED 13 June 2017; ACCEPTED 11 September 2017

ADDRESS FOR CORRESPONDENCE: Kun Sun, Department of Pediatric Cardiology, Xinhua Hospital, Shanghai 200092, China. E-mail: sunkun@xinhuamed.com.cn; Rang Xu, Scientific Research Center, Xinhua Hospital, Shanghai 200092, China. E-mail: rangxu@shsmu.edu.cn
[#] These authors contributed equally to this work.

mutations (first shown for Van der Woude syndrome; Dal et al., 2014; Gilbert et al., 2002; Kondo et al., 2002; Li et al., 2014). The identification of molecular genetic differences between discordant MZ twins suggests that the utility of discordant MZ twin pairs could be beyond heritability studies for gene discovery (Kimani et al., 2009).

The association between CHDs and twin pregnancies has been reported. CHDs have been found at a higher rate in twin pregnancies (Hajdu et al., 2006). The risk of CHDs in MZ twins without twin-to-twin transfusion syndrome (TTTS) is seven-fold higher compared to the general population, and CHDs usually occur in only one twin out of a twin pair (AlRais et al., 2011; Karatza et al., 2002). Based on identical genetic backgrounds, the unaffected twin in a CHD-discordant MZ twin pair provides a well-matched control. The genetic differences identified in the affected twin but not in the unaffected twin may lead to disease. In this study, we applied this strategy to the investigation of CHDs. We hypothesized that mutations in gene or chromosomal regions might contribute to the discordance in MZ twin pairs with CHDs. To test this hypothesis, we investigated six CHD discordant MZ twin pairs using copy number variant (CNV) and whole exome sequencing (WES) analysis.

Materials and Methods

Patients and DNA Purification

The study setup was approved by the Medical Ethics Committee of Xinhua Hospital. All experiments were carried out in accordance with the approved guidelines. Informed written consent was obtained from all participants or their parents prior to study enrolment.

The twin pairs discordant for CHDs were recruited from the pediatric cardiology department of Xinhua Hospital from 2011 to 2015. Prenatal ultrasound was reviewed to exclude TTTS. The cardiac phenotype was ascertained by echocardiography and/or angiocardiography. CHDs were excluded in all non-affected twin members by transthoracic echocardiography. Medical history of the non-affected twin members was reviewed and extra-cardiac anormalies, dysmorphism, and developmental delay were all excluded. Familial history was negative with respect to congenital anomalies and developmental delay. All of the participants were of Han ethnicity and were born to non-consanguineous parents.

Peripheral blood was taken from all of the six twin pairs, and DNA was extracted from the blood samples using the QIAamp DNA Blood Mini Kit (Qiagen, Duesseldorf, Germany) according to the manufacturer's instrunctions. RNA digestion was performed at 37°C for 1 hour using RNase (Qiagen, Duesseldorf, Germany). Purity of the DNA samples was assessed with a NanoDrop spectrophotometer (Nanodrop 2000, Thermo Scientific, USA) and 1% agarose gel electrophoresis.

To confirm that the twin pairs were monozygous, the AmpFℓSTR® Identifiler® Plus PCR Amplification Kit (Applied Biosystems, Inc., Foster City, USA) was used according to the manufacturer's recommendations. Sixteen different short tandem (STR) loci were amplified and the polymerase chain reaction (PCR) products were sequenced using an ABI 3100 sequencer (Applied Biosystems, Inc., Foster City, USA). Dizygotic twins were excluded and six MZ twin pairs were enrolled.

Agilent CGH Array 4x180K

For twin pair 1 and 2, array CGH analysis was performed using an Agilent 4x180K array platform at a median resolution of 13kb (Human Genome CGH Microarray, Agilent Technologies, Santa Clara, CA, USA). Peripheral blood-derived DNA from both twins was hybridized to the microarray according to the manufacturer's recommendations. To identify inter-twin CNV differences, DNA from each twin was hybridized against its co-twin. Data analysis was performed using Cytogenomics Software (v.2.0.6.0). All arrays passed initial QC and were submitted for CNV discovery. High-confidence CNVs were detected by aberration detection method 2 (ADM-2) with the minimum marker 5 and threshold 6.0. The data was filtered, and only those regions larger than 100kb comprising at least 50 contiguous markers were retained (Cao et al., 2015). CNVs were discarded if they showed significant overlap of $\geq$50% with copy number polymorphisms in the human genome, based on the Database of Genomic Variants (DGV; http://dgv.tcag.ca/dgv/app/home).

SNP Genotyping Arrays

For twin pairs 3, 4, 5 and 6, CNV analysis was performed using the Affymetrix Cytoscan HD chromosome microarray platform (Affymetrix, Santa Clara, CA, USA), which interrogates 2.7 million of genetic markers (749,157 single nucleotide polymorphism (SNP) markers and 1,953,246 copy number markers). Hybridizations were performed according to the manufacturer's protocols. The raw data were processed using the Affymetrix Chromosome Analysis Suite (ChAS) Software. All of the arrays passed initial QC and were submitted for CNV discovery based on the annotations of genome version GRCh37 (hg19). CNVs $<$500 kb with no RefSeq genes, events located entirely within segmental duplications, and CNVs with $\geq$50% overlap with control CNVs in the DGV were excluded (Conroy et al., 2014). In addition, SNP genotyping was performed on the twin pair 1 and 2 to validate the CGH results and to detect detailed information of CNVs.

CNVs were classified into four different subgroups according to the classification system used by Bartnik et al. (2012) and Conroy et al. (2014). Group A CNVs are of clinical relevance (containing known CHD causative genes). Group B CNVs are of likely pathogenic effect (CNVs $>$ 1 Mb in length and/or are discordant duplications

in the MZ twin pairs). Group C type is CNVs > 500 kb containing no RefSeq genes, or CNVs > 300 kb with at least one RefSeq gene. Group D comprises all remaining CNVs. CNVs that overlapped by ≥50% in each twin were classified as concordant CNVs.

Exome Sequencing and Data Analysis

The genomic DNA samples of the twins were used for exome capture and sequencing. The whole exome capture was performed using an Agilent SureSelect Human All Exon v5 kit (50 Mb) (Agilent Technologies, Inc., Santa Clara, CA). Briefly, the genomic DNA samples were randomly fragmented using the Covaris S2 system, and then subjected to library preparation (perform end repair, add 'A' bases to the 3′ end, and ligate the adapters) according to the manufacturer's protocol. Exome enrichment was performed for the shotgun libraries. The enriched shotgun libraries were sequenced using the Illumina Hiseq 2,500 platform, and paired-end reads were generated. Raw image data and base calling were processed by the integrated primary analysis software (Real Time Analysis [RTA]) using the default parameters. After removing reads with sequence matching the sequencing adaptors and low-quality reads with more than five unknown bases, data were then aligned to hg19 using the Burrows-Wheeler Alignment (BWA) software (version 0.7.10; Li & Durbin, 2010). PCR duplications were excluded by SAMtools-0.1019.

Single nucleotide and indel differences between CHD and non-CHD twins were detected using SAMtools v0.1.19 and GATK (McKenna et al., 2010). The union of single nucleotide variants called by all of the variant callers was taken forward for annotation, filtering, and manual review. Variants were annotated with dbSNP build 137 and Ensembl version 74_37, and grouped by functional category as either coding (tier 1, all variants in the coding regions of annotated exons, canonical splice sites, and RNA genes), or regulatory (tier 2, all variants in the UTR, upstream and downstream of genes), or non-coding (tier 3, all variants identified in non-repeat masked regions; Mardis et al., 2009; Meltz Steinberg et al., 2015). The SNPs with a minor allele frequency (MAF) of more than 1% (dbSNP137, 1,000 Genomes, esp6500) were all excluded. To detect discordant variants, we compared side by side at each locus of the variants. Discordant SNVs were called with the sequencing depth ≥ 20X and Genotype quality ≥ 10. Coding and regulatory variants identified as discordant between twin pairs were manually reviewed in IGV version 2.3.40.

Validation of CNV and Mutation

For CNV analysis, subsequent confirmation experiments for the presence of the CNVs were done by real-time quantitation (qPCR) using SYBR-Green method as described previously (Conroy et al., 2014). Each experiment was performed in triplicate, and the predicted copy number was calculated based on the equation: CN = 2*(2^(−ΔΔCt)). When results differed between CNV analysis and qPCR, the qPCR results were used.

For exome sequencing, the variants predicted to be discordant between the twin pairs were prioritized for validation by Sanger sequencing. The primers flanking exons of candidate genes were designed by Primer 5 (provided when request).

Results

Clinical Features

Six MZ twins presenting discordant phenotypes were recruited to help identify genetic factors that might explain the occurrence of congenital heart defects of one of the twins but not the other. Detailed medical histories were obtained by interviewing other family members. No other congenital abnormalities were found upon physical examination in either the affected individuals or other family members. Furthermore, none of the patients had a family history of CHDs. The clinical manifestations of the twin pairs are depicted in Table 1.

CNV Results

With 170,334 probes spread throughout the genome, the Agilent CGH 4x180K array provided a media resolution of approximately 13 kb. No obvious inter-twin discordant CNV (≥100 kb) was found in twin pairs 1 and 2 (Table 1).

SNP genotyping arrays were performed on twin pairs 1, 2, 3, 4, 5, and 6 (Table 1). No obvious inter-twin discordant CNVs (≥100 kb) were found in these CHD-discordant twin pairs. In twin pair 2, a single group CNV was identified (chrX:16,985,921-17,729,022) (Table 2, Figure 1). The 743-kb duplication of Xp22.2-p22.13 in twin 2_1 contains 3 RefSeq genes: REPS2, NHS, and MIR4768. A lack of function NHS (OMIM 300457) mutant protein is an established cause of Nancy-Horan syndrome (congenital cataracts and dental anomalies; Brooks et al., 2004, 2010; Burdon et al., 2003). However, the patient's twin brother 2_2 also carried the same 0.7 Mb duplication and he did not have CHD or other congenital anomalies.

In twin pair 5, one inter-twin concordant group B CNV was identified (chrX:6,460,521-8,053,641) (Table 2, Figure 2). The 1.6 Mb deletion of Xp22.31 contains 4 genes (HDHD1, STS, VCX, PNPLA4) and a mircoRNA-4767. It is the typical Xp22.31 deletion that causes an STS-deficiency or X-linked ichthyosis (XLI), which affects 1: 2,000–6,000 males (Cuevas-Covarrubias & Gonzalez-Huerta, 2008; Van Esch et al., 2005). However, no obvious abnormalities in the skin were found in the heterozygous females twins.

Exome Sequencing Results

We sequenced the exomes of the six twin pairs and generated ~1.1 Gbp of paired-end sequencing data per individual, which yielded an average depth of ~100× in target regions with an average of 95.4% of the target base

TABLE 1

Twin Samples Analyzed in this Study

	Twin sample clinical data				Analysis method			
Twin pair no.	Sex	Age (mon)	Cardiac phenotype	ECG	Extracardiac anomalies	Array CGH 4x180K	Cytoscan HD	Exom sequencing
1_1	M	51	HRHS(PA/TS/ASD/PDA)	I°AVB	Nil	√	√	√
1_2	M	51	N	N	Nil	√	√	√
2_1	M	42	LC-RV fistula	N	Nil	√	√	√
2_2	M	42	N	N	Nil	√	√	√
3_1	M	19	VSD	N	Nil		√	√
3_2	M	19	N	N	Nil		√	√
4_1	F	32	ASD	N	Nil		√	√
4_2	F	32	N	N	Nil		√	√
5_1	F	48	VSD	N	Nil		√	√
5_2	F	48	N	N	Nil		√	√
6_1	F	1d[a]	AP window/PS/double Aortic arch	N	Nil		√	√
6_2	F	1d[a]	N	N	Nil		√	√

Note: *n*_1 = CHD affected twin; *n*_2 = CHD unaffected twin; M = male; F = female; mon = month; N = normal; LC = left coronary; RV = right ventricle; HRHS = hypoplastic right heart syndrome; PA = pulmonary atresia; TS = tricuspid stenosis; ASD = atrial septal defect; PDA = patent ductus arteriosus; VSD = ventricular septal defect; AP window = aortic-pulmonary window; PS = pulmonary stenosis; ECG = electrocardiography; AVB = atrioventricular block; 1d = 1 day.
[a]The cardiac defects of the twin 6_1 were detected during the pregnancy, and identified after birth.

TABLE 2

Summary of CNVs Concordant Between Twin Pairs

Twin pair	Sex	Type	Region	CN state	Start, bp	End, bp	Size, kb	Probe count	Genes
2_1	Male	Gain	Xp22.2-p22.13	2	16,985,921	17,729,022	743	1,212	*REPS2, NHS, MIR4768*
2_2	Male	Gain	Xp22.2-p22.13	2	16,985,921	17,729,022	743	1,212	*REPS2, NHS, MIR4768*
5_1	Female	Loss	Xp22.31	1	6,460,521	8,053,641	1,598	4,216	*HDHD1, STS, VCX, PNPLA4,MIR4767*
5_2	Female	Loss	Xp22.31	1	6,460,521	8,053,816	1,593	4,214	*HDHD1, STS, VCX, PNPLA4,MIR4767*

pairs having at least 30× coverage (Table 3). The transition/transversion (Ti/Tv) ratio is generally used to evaluate the quality of SNP calls and is reported to be between 2–2.2 and 2.8–3.0 for SNPs anywhere in the genome and in the coding region, respectively (1000 Genomes Project Consortium et al., 2010, 2012; Carson et al., 2014). While the Ti/Tv ratio for our exome-wide SNVs ranged between 2.57 and 2.63, the Ti_CDS/Tv_CDS ratio matched these values. We identified an average of 53,000 variants per individual, 98.5% of which were concordant between each twin pair. These results additionally confirmed the monozygosity of all analyzed twin pairs. For each twin pair, we called single nucleotide variants and indels on targeted regions, and we classified each variant as either discordant (genotypes differ within a twin pair) or concordant (genotypes were the same within a twin pair).

No Discordant Variants Found in CHD-Discordant Twin Pairs

An average of 2,174 SNVs and 1,584 indels were present in the twin with CHD but not in the unaffected sibling, that is, these appeared to be discordant variants (Table 4). However, many of these discordant variants were germline variants that were under-called in one individual of the twin pair due to low coverage at the end of reads, while others were identified as false-positives due to cryptic paralogous sequences (sequences that are not annotated in the genome as segmental duplications or repetitive elements but are highly similar). After manual review to correct for these false discordant variants, 41 coding and 1 regulatory discordant variant remained. However, none of these manually reviewed variants could be validated using Sanger sequencing. Therefore, no convincing discordant single nucleotide or indel variants were found in coding or regulatory regions in any of our CHD phenotypically discordant twin siblings.

Discussion

The hypothesis we proposed in this study is that post-twinning copy number variants, single nucleotide variants, and short indel variants could be responsible for the CHDs discordance in MZ twin pairs. CNV profiles of peripheral blood-derived DNA from six MZ twin pairs discordant for CHDs were generated using Affymetrix Cytoscan HD chromosome microarray analysis. No CNV difference was validated in the six MZ twin pairs. We have validated a 743-kb duplication of Xp22.2-p22.13 and a 1.6 Mb deletion of Xp22.31 in twin pairs 2 and 5, respectively (Table 2, Figures 1 and 2). The *NHS* gene, located in the 743-kb

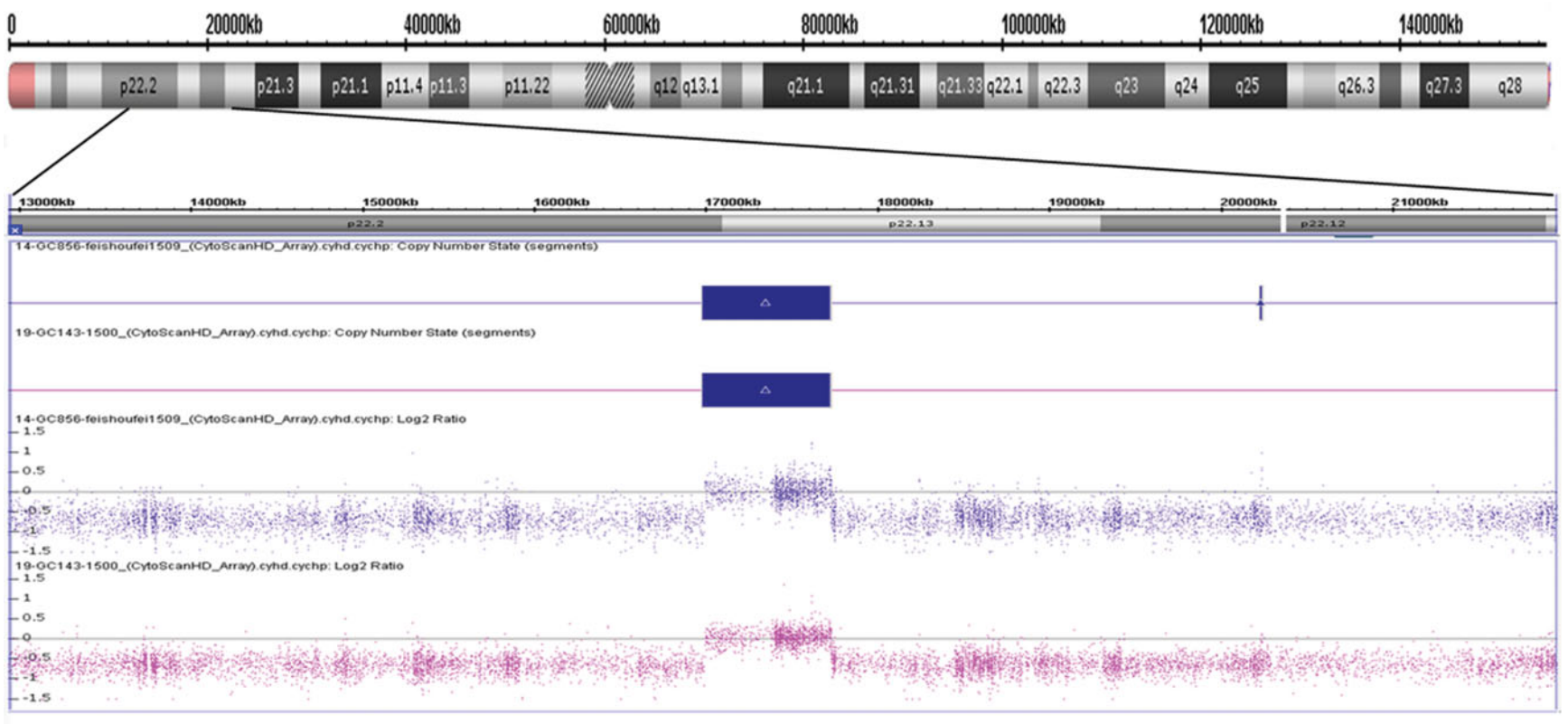

FIGURE 1

(Colour online) Chromosomal view of the 743 kb duplication at chromosome bands Xp22.2-p22.13 (chrX: [16,985,921−17,729,022] × 2) [GRCh37 (hg19)] in twin pair 2.

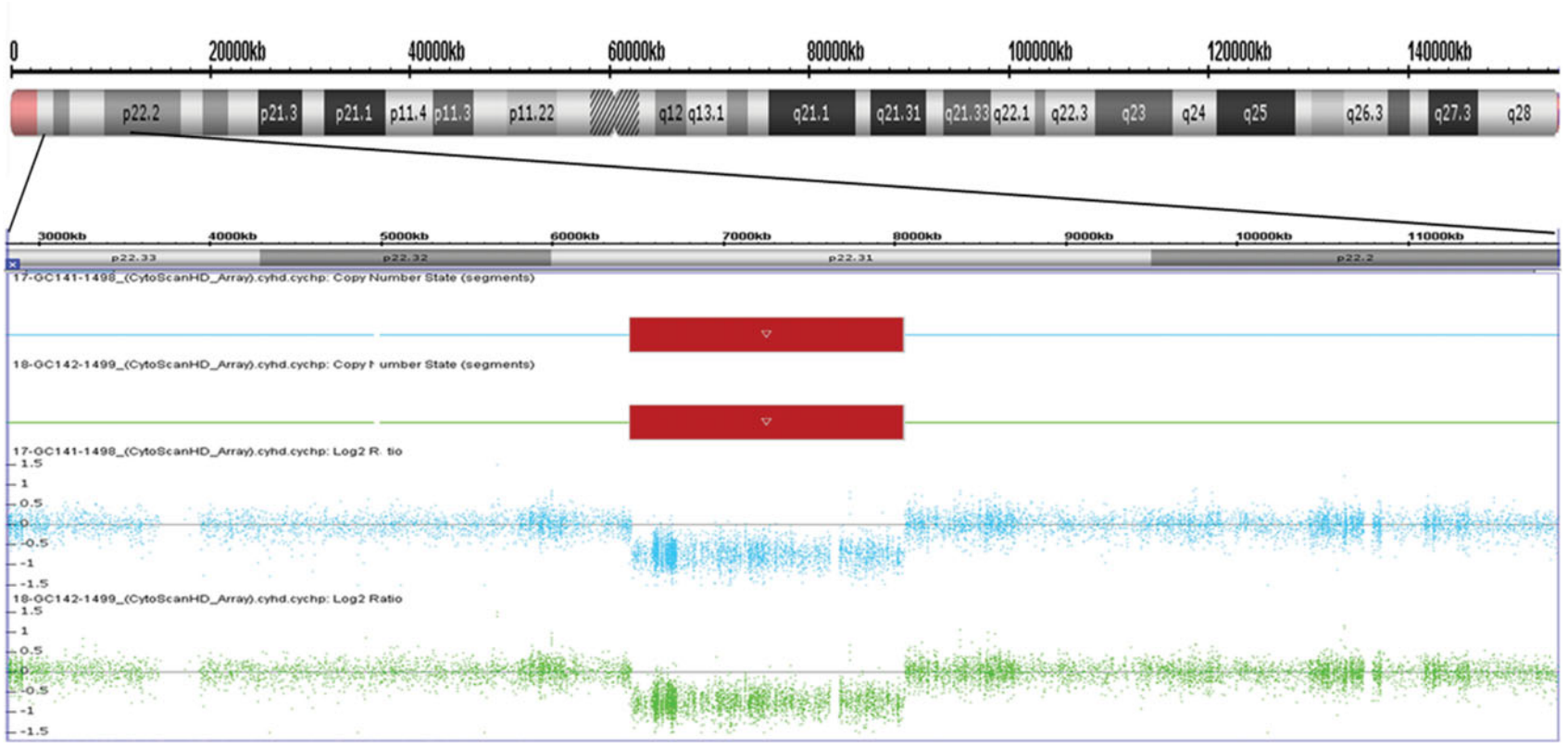

FIGURE 2

(Colour online) Chromosomal view of the 1.6 Mb deletion at chromosome bands Xp22.31 (chrX: [6,460,521−8,053,641] × 1) [GRCh37 (hg19)] in twin pair 5.

TABLE 3

Summary of Sequencing Results From Exome Sequencing Data

	Twin pair_1		Twin pair_2		Twin pair_3		Twin pair_4		Twin pair_5		Twin pair_6	
	1_1	1_2	2_1	2_2	3_1	3_2	4_1	4_2	5_1	5_2	6_1	6_2
Total yield (base pair)	11,964 Mb	13,587 Mb	13,167 Mb	13,012 Mb	15,352 Mb	13,330 Mb	10,683 Mb	11,162 Mb	12,230 Mb	10,682 Mb	12,480 Mb	11,061 Mb
Percent of paired reads on target region (%)	81.42	83.86	81.81	82.29	78.24	76.44	78.33	77.21	77.08	78.07	80.84	83.03
Mean read depth of target regions	137.65	120.06	134.68	129.71	133.76	124.02	102.85	104.41	111.33	98.32	131.2	125.58
% coverage of target regions (>10×)	99.39	99.35	99.41	99.39	99.45	99.37	99.12	99.08	99.23	99.09	99.07	98.93
Total number of variants	53,255	53,306	52,976	52,967	54,120	54,118	54,684	54,475	53,956	53,953	52,986	52,918
Ti/Tv ratio	2.62	2.62	2.58	2.58	2.6	2.6	2.57	2.57	2.63	2.63	2.6	2.6
Ti_CDS/Tv_CDS:	3.18	3.18	3.14	3.13	3.18	3.18	3.13	3.12	3.21	3.21	3.15	3.14
NS/S ratio	0.87	0.87	0.86	0.86	0.85	0.85	0.86	0.87	0.85	0.85	0.86	0.86
Total number of SNVs	42,147	42,156	41,761	41,763	42,418	42,419	42,941	42,944	42,582	42,576	42,418	42,414
Total number of indels	5,174	5,153	5,224	5,254	5,359	5,339	5,341	5,321	5,240	5,289	5,214	5,163
Synonymous SNVs	9,741	9,740	9,779	9,779	9,965	9,964	9,872	9,868	9,935	9,941	9,810	9,814
Non-synonymous SNVs	8,490	8,496	8,411	8,414	8,476	8,474	8,531	8,536	8,471	8,471	8,410	8,414
Stopgain SNVs	53	53	47	47	56	55	54	54	50	51	48	48
Stoploss SNVs	7	7	9	9	8	8	10	10	7	7	9	9
Non-frameshift indels	220	223	207	203	227	224	220	220	217	215	212	214
Frameshift indels	110	110	115	117	120	120	123	123	116	116	125	124
Stopgain indels	2	4	4	5	8	7	3	3	2	3	4	4
Splice-site variants	114	114	103	103	110	110	109	110	111	110	102	102
Upstream, downstream	532	531	520	522	531	530	528	532	535	530	554	551
Variants in UTR	1,314	1,317	1,269	1,270	1,312	1,308	1,350	1,345	1,325	1,330	1,296	1,303
Intronic variants	22,439	22,454	22,773	22,756	22,828	22,833	23,151	23,114	22,863	22,902	22,854	22,789
Intergenic variants	1,885	1,892	1,866	1,868	2,035	2,027	2,145	2,165	2,023	2,019	2,048	2,042

TABLE 4

Filtering Protocol and Results for Variants Used to Identify Differences of Functional Variants Between Twin Siblings

	Twin pair_1	Twin pair_2	Twin pair_3	Twin pair_4	Twin pair_5	Twin pair_6
Total discordant variants	3,262	3,411	3,887	4,130	4,257	3,598
Coding and regulatory	743	774	1,079	1,156	1,154	707
Manual review	7	5	8	11	5	6
Sanger validated	0	0	0	0	0	0

duplication of Xp22.2-p22.13, was reported to play a role in mouse heart development (Burdon et al., 2003). Furthermore, Coccia et al. (2009) reported that four of six X-linked cataract male patients harbored a 0.8 Mb duplication–triplication segment encompassing the *NHS*, *SCML1*, and *RAI2* genes and had congenital heart defects (ductus arteriosus, tetralogy of Fallot, VSD and stenosis of a major cardiac vessel). Van Esch et al. (2007) also reported a 10-month-old male infant with congenital cataracts and tetralogy of Fallot who had a 2.8 Mb microdeletion on Xp22.2-Xp22.13, including the *NHS* and *RAI2* genes. Our male patient twin 2_1 with a 0.7 Mb duplication containing the NHS gene also had a congenital heart defect called LC-RV fistula, but congenital cataract or dental anomalies were not observed. These results support the hypothesis that aberrant transcription or a dosage effect of NHS may lead to cardiac anomalies.

Though the 0.7 Mb CNV contained an *NHS* gene that might be involved in the pathogenesis of CHD, they were concordant between the twin siblings. Previous studies have demonstrated the existence of CNV discordances in MZ twins, which suggests that CNV discordance might explain phenotypic MZ discordance (Abdellaoui et al., 2015; Bruder et al., 2008; Ehli et al., 2012; Ketelaar et al., 2012). However, we failed to find any discordant CNV that could explain the discordant phenotype in six CHD-discordant MZ twin pairs. These results correlate with previous studies that tried to highlight CNV differences between phenotypically discordant MZ twins (Abdellaoui et al., 2015; Baudisch et al., 2013; Breckpot et al., 2012). Breckpot and his colleagues have validated three small CNV differences in CHD-positive twin siblings, but they also failed to find any causal CNV differences (Breckpot et al., 2012). Baudisch and his colleagues also did not confirm any disease-causing CNVs in MZ twin pairs discordant for urorectal malformations (Baudisch et al., 2013). To date, most studies looking for CNV discordances in MZ twins (even with discordant phenotypes) did not detect reproducible post-twinning CNV mutations, which indicates that relatively large CNV discordances between MZ twins are a considerably rare phenomenon (Abdellaoui et al., 2015).

We therefore sequenced the whole exomes of the six MZ CHD-discordant twin pairs to identify post-twinning SNVs and short indel differences between the twins that could make one of them susceptible to CHD. Unfortunately, we also did not detect any discordant coding or regulatory SNV or indel variants that would have been likely to cause CHD to appear in one twin sibling only (Table 4). We are not alone in failing to find post-zygotic single nucleotide and short indel mutations between MZ twins using WES. Several previous studies have failed to find any discordant disease-causing SNVs or indels between MZ twins discordant for Crohn's disease (Petersen et al., 2014), congenital cataracts (Wei et al., 2015), amyotrophic lateral sclerosis (Meltz Steinberg et al., 2015), cleft lip and/or palate (Kimani et al., 2009) using whole genome or exome sequencing. Such negative results suggest that the DNA sequence changes between MZ twins are very rare (Chaiyasap et al., 2014), or are at least hard to detect, even among phenotypically discordant twins. Nevertheless, discordant genetic variants between MZ twins do exist. Reumers et al. (2011) have identified two discordant SNVs in MZ twins discordant for schizophrenia. Vadlamudi et al. (2010) detected discordant *SCN1A* gene mutations in MZ twin pairs discordant for Dravet syndrome. Li et al. (2014) also verified two discordant mutations from two distinct twin pairs. We did not validate most of the intronic and intergenic DNA variants detected in the WES. It remains possible that variants in these regions could differ between twin pairs and therefore contribute to the disease.

It is noteworthy that the risk of CHDs in twin gestations is higher than in the general population, and CHDs usually occur in only one of the twin siblings. Given the lack of genetic differences in the SNP-based array analysis and exome sequencing to explain the discordance of CHD in our twin pairs, other reasons need to be considered for why only one sibling of an MZ twin pair suffers from the disease. Approximately two-thirds of MZ twins are monochorionic. It has been reported that monochorionic placentation is a risk factor for CHDs (Bahtiyar et al., 2007; Karatza et al., 2002). A proportion of CHDs are acquired due to altered hemodynamics, particularly in the recipient twin affected by TTTS (Manning, 2008). In our study, TTTS had been already excluded from the six twin pairs through a detailed sonographic evaluation. Other possible reasons include the monozygotic twinning process itself or a postzygotic unequal division of the embryonic cell mass that allows for unequal potential for development (Bahtiyar et al., 2007).

Another possible explanation for discordance between twin pairs for CHD could be epigenetic differences. Evidence from animal models and humans confirmed the intrauterine period as a sensitive time for the establishment

of epigenetic variability (Gluckman et al., 2008; Gordon et al., 2012; Hanson and Gluckman, 2008; Ollikainen et al., 2010). Previous studies demonstrated that MZ twin pairs exhibit epigenetic differences by comparing DNA methylation or histone acetylation in different tissues (peripheral lymphocytes, buccal epithelial cells, gut biopsies, and placental tissue; Fraga et al., 2005; Kaminsky et al., 2009; Ollikainen et al., 2010). Additionally, epigenetic differences between MZ twins emerged early in life and were observed in the perinatal epigenome (Castillo-Fernandez et al., 2014; Gordon et al., 2012; Ollikainen et al., 2010). The intrauterine environment differences may cause epigenetic differences in MZ twins. It has been reported that DNA methylation profiles are less similar within pairs of monochorionic (shared placenta) MZ twins compared to pairs of dichorionic (non-shared placenta) MZ twins. This observation suggests that sharing a placenta may cause an imbalance under in-utero conditions along with more discordant epigenetic profiles (Castillo-Fernandez et al., 2014). These epigenetic differences may account for phenotypic discordance in MZ twins. Multiple studies have explored the epigenetic changes in twins discordant for a range of congenital diseases, including autism spectrum disorder (Nguyen et al., 2010; Wong et al., 2014), birth weight (Souren et al., 2013; Tsai et al., 2015), congenital renal agenesis (Jin et al., 2014), and congenital cataracts (Wei et al., 2015). Recently, epigenetic aberrations have been sought in CHDs (Serra-Juhé et al., 2015; Vecoli et al., 2015) but not in CHD-discordant twins. Additional studies in this area are necessary.

It is noteworthy that the blood-derived DNA we used would be chimeric between MZ twin pairs. Shared blood circulation during embryogenesis is found in most MZ twin pregnancies, especially in monochorionic MZ twin pairs in which both twins are fed by a single placenta (Erlich, 2011). Under this condition, hematopoietic stem cells can be transferred between the twins and chimeric hematopoietic systems can be created (Chaiyasap et al., 2014; Erlich, 2011). Such a system will mask the underlying discordant variations that cause phenotypic differences in MZ twins because it is possible that post-twinning genetic variations that arise in one twin can be detected in the blood system of their co-twin. Although none of the twin pairs in our study had TTTS, which is caused by excessive blood sharing, the exact degree of sharing cannot be ascertained. It is important to sample the tissue that shows the discordant phenotype for DNA extraction. Unfortunately, cardiac tissue is difficult to obtain, especially in healthy individuals. Buccal and skin cells are easier to obtain, but they are ectoderm-derived and not suitable for our study.

Our research also has certain limitations due to the small sample size and the heterogeneity in the heart defects of the affected twin. Thus, the presented findings might be confounded by extraneous contribution factors. Additional studies that test a larger sample set of more similar CHD sub-phenotypes will help to further our understanding.

In conclusion, our study confirms that large CNV (>100 kb) or exome DNA differences are very rare in MZ twin pairs. Our results correlated with the outcomes of previous studies (Breckpot et al., 2012; Chaiyasap et al., 2014; Hui et al., 2016; Zhang et al., 2016). These results suggest that further efforts are needed to identify other mechanisms that could contribute to the occurrence of CHDs, such as environmental agents, epigenetic changes, variants in non-coding DNA, or oligogenic mechanisms (signal pathway-related variants).

Acknowledgments

We thank all the patients for their participation in this study. We also thank Yi-Fan Zhu and AJE for the article proofreading. This work was supported by the grants from the National Natural Science Foundation of China (81270233/H0204, 81300068/H0201), the grant (SHDC12015102) from the new frontier technology joint project of Shanghai ShenKang Hospital Development Center, the Joint Health Research Program for major diseases in Shanghai (2013ZYJB0016), and the Shanghai Educational committee (15ZZ055).

Conflict of Interest

None.

References

1000 Genomes Project Consortium, Abecasis, G. R., Altshuler, D., Auton, A., Brooks, L. D., Durbin, R. M., … McVean, G. A. (2010). A map of human genome variation from population-scale sequencing. *Nature, 467*, 1061–1073.

100 Genomes Project Consortium, Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., … McVean, G. A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, *491*, 56–65.

Abdellaoui, A., Ehli, E. A., Hottenga, J. J., Weber, Z., Mbarek, H., Willemsen, G., … Boomsma, D. I. (2015). CNV concordance in 1,097 MZ twin pairs. *Twin Research and Human Genetics, 18*, 1–12.

AlRais, F., Feldstein, V. A., Srivastava, D., Gosnell, K., & Moon-Grady, A. J. (2011). Monochorionic twins discordant for congenital heart disease: A referral center's experience and possible pathophysiologic mechanisms. *Prenatal Diagnosis, 31*, 978–984.

Bahtiyar, M. O., Dulay, A. T., Weeks, B. P., Friedman, A. H., & Copel, J. A. (2007). Prevalence of congenital heart defects in monochorionic/diamniotic twin gestations: A systematic literature review. *Journal of Ultrasound in Medicine*, *26*, 1491–1498.

Bartnik, M., Szczepanik, E., Derwinska, K., Wiśniowiecka-Kowalnik, B., Gambin, T., Sykulski, M., … Stankiewicz, P. (2012). Application of array comparative genomic hybridization in 102 patients with epilepsy and additional neurodevelopmental disorders. *American Journal of*

Medical Genetics Part B: Neuropsychiatric Genetics, *159*, 760–771.

Baudisch, F., Draaken, M., Bartels, E., Schmiedeke, E., Bagci, S., Bartmann, P., … Reutter, H. (2013). CNV analysis in monozygotic twin pairs discordant for urorectal malformations. *Twin Research and Human Genetics, 16*, 802–807.

Breckpot, J., Thienpont, B., Gewillig, M., Allegaert, K., Vermeesch, J. R., & Devriendt, K. (2012). Differences in copy number variation between discordant monozygotic twins as a model for chromosomal mosaicism in congenital heart defects. *Molecular Syndromology*, *2*, 81–87.

Brooks, S. P., Coccia, M., Tang, H. R., Kanuga, N., Machesky, L. M., Bailly, M., … Hardcastle, A. J. (2010). The Nance-Horan syndrome protein encodes a functional WAVE homology domain (WHD) and is important for co-ordinating actin remodelling and maintaining cell morphology. *Human Molecular Genetics, 19*, 2421–2432.

Brooks, S. P., Ebenezer, N. D., Poopalasundaram, S., Lehmann, O. J., Moore, A. T., & Hardcastle, A. J. (2004). Identification of the gene for Nance-Horan syndrome (NHS). *Journal of Medical Genetics, 41*, 768–771.

Bruce, A. C., Seidman, G. J. G., & Seidman, C. E. (2013). Genetics of congenital heart disease: The glass half empty. *Circulation Research, 112*, 707–720.

Bruder, C. E., Piotrowski, A., Gijsbers, A. A., Andersson, R., Erickson, S., Diaz de Ståhl, T., … Dumanski, J. P. (2008). Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variantion profiles. *American Journal of Human Genetics, 82*, 763–771.

Bruneau, B. G. (2008). The developmental genetics of congenital heart disease. *Nature*, *451*, 943–948.

Burdon, K. P., McKay, J. D., Sale, M. M., Russell-Eggitt, I. M., Mackey, D. A., Wirth, M. G., … Craig, J. E. (2003). Mutations in a novel gene, NHS, cause the pleiotropic effects of Nance-Horan syndrome, including severe congenital cataract, dental anomalies, and mental retardation. *American Journal of Human Genetics, 73*, 1120–1130.

Cao, R., Pu, T., Fang, S., Long, F., Xie, J., Xu, Y., … Xu, R. (2015). Patients carrying 9p31.1-q32 deletion share common features with Cornelia de Lange syndrome. *Cellular Physiology and Biochemistry, 35*, 270–280.

Carson, A. R., Smith, E. N., Matsui, H., Brækkan, S. K., Jepsen, K., Hansen, J. B., & Frazer, K. A. (2014). Effective filtering strategies to improve data quality from population-based whole exome sequencing studies. *BMC Bioinformatics*, *15*, 125.

Castillo-Fernandez, J. E., Spector, T. D., & Bell, J. T. (2014). Epigenetics of discordant monozygotic twins: Implications for disease. *Genome Medicine, 6*, 60.

Chaiyasap, P., Kulawonganunchai, S., Srichomthong, C., Tongsima, S., Suphapeetiporn, K., & Shotelersuk, V. (2014). Whole genome and exome sequencing of monozygotic twins with trisomy 21, discordant for a congenital heart defect and epilepsy. *PLoS One*, *9*, e100191.

Coccia, M., Brooks, S. P., Webb, T. R., Christodoulou, K., Wozniak, I.O., Murday, V., … Hardcastle, A. J. (2009). X-linked cataract and Nance-Horan syndrome are allelic disorders. *Human Molecular Genetics, 18*, 2643–2655.

Conroy, J., McGerrigan, P. A., McCreary, D., Shah, N., Collins, K., Parry-Fielder, B., … King, M. D. (2014). Towards the identitication of genetic basis for Landau-Kleffner syndrome. *Epilepsia*, *55*, 858–865.

Cuevas-Covarrubias, S. A., & Gonzalez-Huerta, L. M. (2008). Analysis of the VCX3A, VCX2 and VCX3B genes shows that VCX3A gene deletion is not sufficient to result in mental retardation in X-linked ichthyosis. *British Journal of Dermatology, 158*, 483–486.

Dal, G. M., Erguner, B., Sagiroglu, M. S., Yuksel, B., Onat, O. E., Alkan, C., & Ozcelik, T. (2014). Early postzygotic mutations contribute to de novo variation in a healthy monozygotic twin pair. *Journal of Medical Genetics, 51*, 455–459.

Ehli, E. A., Abdellaoui, A., Hu, Y., Hottenga, J. J., Kattenberg, M., van Beijsterveldt, T., … Davies, G. E. (2012). De novo and inherited CNVs in MZ twin pairs selected for discordance and concordance on Attention Problems. *European Journal of Human Genetics, 20*, 1037–1043.

Erlich, Y. (2011). Blood ties: Chimerism can mask twin discordance in high-throughput sequencing. *Twin Research and Human Genetics, 14*, 137–143.

Fraga, M. F., Ballestar, E., Paz, M. F., Ropero, S., Setien, F., Ballestar, M. L., … Esteller, M. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 10604–10609.

Gelb, B. D., & Chung, W. K. (2014). Complex genetics and etiology of human congenital heart disease. *Cold Spring Harbor Perspectives in Medicine*, *4*, a13953.

Gilbert, B., Yardin, C., Briault, S., Belin, V., Lienhardt, A., Aubard, Y., … Lacombe, D. (2002). Prenatal diagnosis of female monozygotic twins discordant for Turner syndrome: Implications for prenatal genetic counseling. *Prenatal Diagnosis, 22*, 697–702.

Gluckman, P. D., Hanson, M. A., Cooper, C., & Thornburg, K. L. (2008). Effect of in utero and early-life conditions on adult health and disease. *New England Journal of Medicine, 359*, 61–73.

Gordon, L., Joo, J. E., Powell, J. E., Ollikainen, M., Novakovic, B., Li, X., … Saffery, R. (2012). Neonatal DNA methylation profile in human twins is specified by a complex interplay between intrauterine environmental and genetic factors, subject to tissue-specific influence. *Genome Research, 22*, 1395–1406.

Hajdu, J., Beke, A., Marton, T., Hruby, E., Pete, B., & Papp, Z. (2006). Congenital heart diseases in twin pregnancies. *Fetal Diagnosis and Therapy*, *21*, 198–203.

Hanson, M. A., & Gluckman, P. D. (2008). Developmental origins of health and disease: New insights. *Basic and Clinical Pharmacology and Toxicology, 102*, 90–93.

Hoffman, J. I., Kaplan, S., & Liberthson, P. R. (2004). Prevalence of congenital heart disease. *American Heart Journal*, *147*, 425–439.

Hui, D. S., Bonow, R. O., Stolker, J. M., Braddock, S. R., & Lee, R. (2016). Discordant aortic valve morphology in monozygotic twins: A clinical case series. *JAMA Cardiology*, *1*, 1043–1047.

Jenkins, K. J., Correa, A., Feinstein, J. A., Botto, L., Britt, A. E., Daniels, S. R., … American Heart Association Council on Cardiovascular Disease in the Young. (2007). Noninherited risk factors and congenital cardiovascular defects: Current knowledge: A scientific statement from the American Heart Association Council on Cardiovascular disease in the young: Endorsed by the American Academy of Pediatrics. *Circulation*, *115*, 2995–3014.

Jin, M., Zhu, S., Hu, P., Liu, D., Li, Q., Li, Z., … Chen, X. (2014). Genomic and epigenomic analyses of monozygotic twins discordant for congenital renal agenesis. *American Journal of Kidney Diseases, 64*, 119–122.

Kaminsky, Z. A., Tang, T., Wang, S. C., Ptak, C., Oh, G. H., Wong, A. H., … Petronis, A. (2009). DNA methylation profiles in monozygotic and dizygotic twins. *Nature Genetics, 41*, 240–245.

Karatza, A. A., Wolfenden, J. L., Taylor, M. J., Wee, L., Fisk, N. M., & Gardiner, H. M. (2002). Influence of twin-twin transfusion syndrome on fetal cardiovascular structure and function: Prospective case-control study of 136 monochorionic twin pregnancies. *Heart, 88*, 271–277.

Ketelaar, M. E., Hofstra, E. M., & Hayden, M. R. (2012). What monozygotic twins discordant for phenotype illustrate about mechanisms influencing genetic forms of neuodegeneration. *Clinical Genetics, 81*, 325–333.

Kimani, J. W., Yoshiura, K., Shi, M., Jugessur, A., Moretti-Ferreira, D., Christensen, K., & Murray, J. C. (2009). Search for genomic alterations in monozygotic twins discordant for cleft lip and/or palate. *Twin Research and Human Genetics, 12*, 462–468.

Kondo, S., Schuttee, B. C., Richardson, R. J., Bjork, B. C., Knight, A. S., Watanabe, Y., … Murray, J. C. (2002). Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. *Nature Genetics, 32*, 285–289.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics, 26*, 589–595.

Li, R., Montpetit, A., Rousseau, M., Wu, S. Y., Greenwood, C. M., Spector, T. D., … Richards, J. B. (2014). Somatic point mutations occurring early in development: A monozygotic twin study. *Journal of Medical Genetics, 51*, 28–34.

Manning, N. (2008). The influence of twinning on cardiac development. *Early Human Development*, *84*, 173–179.

Mardis, E. R., Ding, L., Dooling, D. J., Larson, D. E., McLellan, M. D., Chen, K., … Ley, T. J. (2009). Recurring mutations found by sequencing an acute myeloid leukaemia genome. *New England Journal of Medicine*, *361*, 1058–1066.

Marian, A. J. (2014). Copy number variants and the genetic enigma of congenital heart disease. *Circulation Research, 115*, 821–823.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*, 1297–1303.

Meltz Steinberg, K., Nicholas, T. J., Koboldt, D. C., Yu, B., Mardis, E., & Pamphlett, R. (2015). Whole genome analyses reveal no pathogenetic single nucleotide or structural differences between monozygotic twins discordant for amyotrophic lateral scletosis. *Amyotrophic Lateral Sclerosis & Frontotemporal Degeneration*, *16*, 385–392.

Nguyen, A., Rauch, T. A., Pfeifer, G. P., & Hu, V. W. (2010). Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum discorders and a novel autism candidate gene, RORA, whose protein product is reduced in autistic brain. *FASEB Journal, 24*, 3036–3051.

Ollikainen, M., Smith, K. R., Joo, E. J., Ng, H. K., Andronikos, R., Novakovic, B., … Craig, J. M. (2010). DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome. *Human Molecular Genetics, 19*, 4176–4188.

Petersen, B. S., Spehlmann, M. E., Raedler, A., Stade, B., Thomsen, I., Rabionet, R., … Franke, A. (2014). Whole genome and exome sequencing of monozygotic twins discordant for Crohn's disease. *BMC Genomics, 15*, 564.

Reumers, J., De Rijk, P., Zhao, H., Liekens, A., Smeets, D., Cleary, J., … Del-Favero, J. (2011). Optimized filtering reduces the error rate in detecting genomic variants by short-read sequencing. *Nature Biotechnology, 30*, 61–68.

Serra-Juhé, C., Cuscó, I., Homs, A., Flores, R., Torán, N., & Pérez-Jurado, L. A. (2015). DNA methylation abnormalities in congenital heart disease. *Epigenetics*, *10*, 167–177.

Souren, N. Y., Lutsik, P., Gasparoni, G., Tierling, S., Gries, J., … Walter, J. (2013). Adult monozygotic twins discordant for intra-uterine growth have indistinguishable genome-wide DNA methylation profiles. *Genome Biology, 14*, R44.

Tsai, P. C., Van Dongen, J., Tan, Q., Willemsen, G., Christiansen, L., Boomsma, D. I., … Bell, J. T. (2015). DNA methylation changes in the IGF1R gene in birth weight discordant adult monozygotic twins. *Twin Research and Human Genetics, 18*, 635–646.

Vadlamudi, L., Dibbens, L. M., Lawrence, K. M., Iona, X., McMahon, J. M., Murrell, W., … Berkovic, S. F. (2010). Timing of de novo mutagenesis - A twin study of sodium-channel mutations. *New England Journal of Medicine, 363*, 1335–1340.

Van Esch, H., Hollanders, K., Badisco, L., Melotte, C., Van Hummelen, P., Vermeesch, J. R., … Froyen, G. (2005). Deletion of VCX-A due to NAHR plays a major role in the occurrence of mental retardation in patients with X-linked ichthyosis. *Human Molecular Genetics, 14*, 1795–1803.

Van Esch, H., Jansen, A., Bauters, M., Froyen, G., & Fryns, J. P. (2007). Encephalopathy and bilateral cataract in a boy with an interstitial deletion of Xp22 comprising the CDKL5 and NHS genes. *American Journal of Medical Genetics Part A, 143*, 364–369.

Vecoli, C., Pulignani, S., Foffa, I., & Andreassi, M. G. (2015). Congenital heart disease: The crossroads of genetics, epigenetics and environment. *Current Genomics, 15*, 390–399.

Wei, T., Sun, H., Hu, B., Yang, J., Qiao, C., & Yan, M. (2015). Exome sequencing and epigenetic analysis of twins who are discordant for congenital cataract. *Twin Research and Human Genetics*, *18*, 393–398.

Wong, A. H., Gottesman, J., & Petronis, A. (2005). Phenotypic differences in genetically identical organisms: The epigenetic perspective. *Human Molecular Genetics*, *14*, R11–R18.

Wong, C. C., Meaburn, E. L., Ronald, A., Price, T. S., Jeffries, A. R., Schalkwyk, L. C., … Mill, J. (2014). Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Molecular Psychiatry, 19*, 495–503.

Zhang, R., Thiele, H., Bartmann, P., Hilger, A. C., Berg, C., Herberg, U., … Reutter, H. (2016). Whole-exome sequencing in nine monozygotic discordant twins. *Twin Research and Human Genetics, 19*, 60–65.