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 Xu3

1Department of Pediatric Cardiology, Xinhua Hospital, affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China
2School of Biomedical Science, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China
3Scientific 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
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 (ARais 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 angiocardiology. CHDs were excluded in all non-affected twin members with no RefSeq genes, events located entirely within segmental duplications, and CNVs with ≥50% overlap with control CNVs in the DGV were excluded (Conroy et al., 2014). In addition, SNP genotyping was performed on the Affymetrix Chromosome Analysis Suite (ChAS) Software. All of the 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 ≥50% with copy number polymorphisms in the human genome, based on the Database of Genomic Variants (DGV; http://dgv.tcgag.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 ≥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 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. 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, esp56000) 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 CNV (≥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 discordant 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

<table>
<thead>
<tr>
<th>Twin pair no.</th>
<th>Sex</th>
<th>Age (mon)</th>
<th>Cardiac phenotype</th>
<th>ECG</th>
<th>Extracardiac anomalies</th>
<th>Array CGH</th>
<th>Cytoscan</th>
<th>Exom sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1_1</td>
<td>M</td>
<td>51</td>
<td>HRHS/PA/ASD/PDA</td>
<td>I’AVB</td>
<td>Nil</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>1_2</td>
<td>M</td>
<td>51</td>
<td>N</td>
<td>Nil</td>
<td>Nil</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>2_1</td>
<td>M</td>
<td>42</td>
<td>LC-RV fistula</td>
<td>Nil</td>
<td>Nil</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>2_2</td>
<td>M</td>
<td>42</td>
<td>N</td>
<td>Nil</td>
<td>Nil</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>3_1</td>
<td>M</td>
<td>19</td>
<td>VSD</td>
<td>Nil</td>
<td>Nil</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>3_2</td>
<td>M</td>
<td>19</td>
<td>N</td>
<td>Nil</td>
<td>Nil</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>4_1</td>
<td>F</td>
<td>32</td>
<td>ASD</td>
<td>Nil</td>
<td>Nil</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>4_2</td>
<td>F</td>
<td>32</td>
<td>N</td>
<td>Nil</td>
<td>Nil</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>5_1</td>
<td>F</td>
<td>48</td>
<td>N</td>
<td>Nil</td>
<td>Nil</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>5_2</td>
<td>F</td>
<td>48</td>
<td>N</td>
<td>Nil</td>
<td>Nil</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>6_1</td>
<td>F</td>
<td>1d</td>
<td>AP window/PS/double Aortic arch</td>
<td>Nil</td>
<td>Nil</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>6_2</td>
<td>F</td>
<td>1d</td>
<td>N</td>
<td>Nil</td>
<td>Nil</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

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; I’AVB = atrioventricular block; 1d = 1 day.

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

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

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

Downloaded from https://www.cambridge.org/core. IP address: 54.70.40.11, on 02 Aug 2019 at 22:09:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms . https://doi.org/10.1017/thg.2017.57
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**

<table>
<thead>
<tr>
<th></th>
<th>Twin pair_1</th>
<th>Twin pair_2</th>
<th>Twin pair_3</th>
<th>Twin pair_4</th>
<th>Twin pair_5</th>
<th>Twin pair_6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1_1</td>
<td>1_2</td>
<td>2_1</td>
<td>2_2</td>
<td>3_1</td>
<td>3_2</td>
</tr>
<tr>
<td>Total yield (base pair) (Mb)</td>
<td>11,964</td>
<td>13,587</td>
<td>13,167</td>
<td>13,012</td>
<td>15,352</td>
<td>13,330</td>
</tr>
<tr>
<td>Percent of paired reads on target region (%)</td>
<td>81.42</td>
<td>83.86</td>
<td>81.81</td>
<td>82.29</td>
<td>78.24</td>
<td>76.44</td>
</tr>
<tr>
<td>Mean read depth of target regions</td>
<td>137.65</td>
<td>120.06</td>
<td>134.68</td>
<td>129.71</td>
<td>133.76</td>
<td>124.02</td>
</tr>
<tr>
<td>Total number of variants</td>
<td>53,255</td>
<td>53,306</td>
<td>52,976</td>
<td>52,967</td>
<td>54,120</td>
<td>54,118</td>
</tr>
<tr>
<td>Ti/Tv ratio</td>
<td>2.62</td>
<td>2.62</td>
<td>2.58</td>
<td>2.58</td>
<td>2.6</td>
<td>2.6</td>
</tr>
<tr>
<td>Ti_CDS/Tv_CDS</td>
<td>3.18</td>
<td>3.18</td>
<td>3.14</td>
<td>3.13</td>
<td>3.18</td>
<td>3.18</td>
</tr>
<tr>
<td>NS/S ratio</td>
<td>0.87</td>
<td>0.87</td>
<td>0.86</td>
<td>0.86</td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td>Total number of SNVs</td>
<td>42,147</td>
<td>42,156</td>
<td>41,761</td>
<td>41,763</td>
<td>42,418</td>
<td>42,419</td>
</tr>
<tr>
<td>Total number of indels</td>
<td>5,174</td>
<td>5,153</td>
<td>5,224</td>
<td>5,254</td>
<td>5,359</td>
<td>5,339</td>
</tr>
<tr>
<td>Synonymous SNVs</td>
<td>9,741</td>
<td>9,740</td>
<td>9,779</td>
<td>9,779</td>
<td>9,965</td>
<td>9,964</td>
</tr>
<tr>
<td>Non-synonymous SNVs</td>
<td>8,490</td>
<td>8,496</td>
<td>8,411</td>
<td>8,414</td>
<td>8,476</td>
<td>8,474</td>
</tr>
<tr>
<td>Stopgain SNVs</td>
<td>53</td>
<td>53</td>
<td>47</td>
<td>47</td>
<td>56</td>
<td>55</td>
</tr>
<tr>
<td>Stoploss SNVs</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Non-frameshift indels</td>
<td>220</td>
<td>223</td>
<td>207</td>
<td>203</td>
<td>227</td>
<td>224</td>
</tr>
<tr>
<td>Frameshift indels</td>
<td>110</td>
<td>110</td>
<td>115</td>
<td>117</td>
<td>120</td>
<td>120</td>
</tr>
<tr>
<td>Splice-site variants</td>
<td>114</td>
<td>114</td>
<td>103</td>
<td>103</td>
<td>110</td>
<td>110</td>
</tr>
<tr>
<td>Upstream, downstream</td>
<td>532</td>
<td>531</td>
<td>520</td>
<td>522</td>
<td>531</td>
<td>530</td>
</tr>
<tr>
<td>Variants in UTR</td>
<td>1,314</td>
<td>1,317</td>
<td>1,269</td>
<td>1,270</td>
<td>1,312</td>
<td>1,308</td>
</tr>
<tr>
<td>Intronic variants</td>
<td>22,439</td>
<td>22,454</td>
<td>22,773</td>
<td>22,756</td>
<td>22,828</td>
<td>22,833</td>
</tr>
<tr>
<td>Intergenic variants</td>
<td>1,885</td>
<td>1,892</td>
<td>1,866</td>
<td>1,868</td>
<td>2,035</td>
<td>2,027</td>
</tr>
</tbody>
</table>
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 SCML1, and RA12 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 cataractas and tetralogy of Fallot who had a 2.8 Mb microdeletion on Xp22.2-Xp22.13, including the NHS and RA12 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 explain the discordant phenotype in six CHD-discordant MZ twin pairs. We therefore sequenced the whole exomes of the six 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 placentaion 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 monzygotic 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

### Table 4

<table>
<thead>
<tr>
<th>Twin pair</th>
<th>Twin pair</th>
<th>Twin pair</th>
<th>Twin pair</th>
<th>Twin pair</th>
<th>Twin pair</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total discordant variants</td>
<td>3,262</td>
<td>3,411</td>
<td>3,887</td>
<td>4,130</td>
<td>4,257</td>
</tr>
<tr>
<td>Coding and regulatory</td>
<td>743</td>
<td>774</td>
<td>1,079</td>
<td>1,156</td>
<td>1,154</td>
</tr>
<tr>
<td>Manual review</td>
<td>7</td>
<td>5</td>
<td>8</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>Sanger validated</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4: Filtering Protocol and Results for Variants Used to Identify Differences of Functional Variants Between Twin Siblings
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 TTPS, 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


