Rare Variants in Novel Candidate Genes Associated With Nonsyndromic Patent Ductus Arteriosus Identified With Whole-Exome Sequencing

Ying Gao†, Dan Wu†, Bo Chen2, Yinghui Chen3, Qi Zhang3 and Pengjun Zhao3*

1Department of Pediatric, Shidong Hospital, Shanghai, China, 2Department of Cardiothoracic Surgery, School of Medicine, Heart Center, Shanghai Children’s Medical Center, School of Medicine, Shanghai Jiao Tong University, Shanghai, China, 3Department of Pediatric Cardiology, Xin Hua Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

Background: Patent ductus arteriosus (PDA) is one of the most common congenital heart defects causing pulmonary hypertension, infective endocarditis, and even death. The important role of genetics in determining spontaneous ductal closure has been well-established. However, as many of the identified variants are rare, thorough identification of the associated genetic factors is necessary to further explore the genetic etiology of PDA.

Methods: We performed whole-exome sequencing (WES) on 39 isolated nonsyndromic PDA patients and 100 healthy controls. Rare variants and novel genes were identified through bioinformatic filtering strategies. The expression patterns of candidate genes were explored in human embryo heart samples.

Results: Eighteen rare damaging variants of six novel PDA-associated genes (SOX8, NES, CDH2, ANK3, EIF4G1, and HIPK1) were newly identified, which were highly expressed in human embryo hearts.

Conclusions: WES is an efficient diagnostic tool for exploring the genetic pathogenesis of PDA. These findings contribute new insights into the molecular basis of PDA and may inform further studies on genetic risk factors for congenital heart defects.

Keywords: congenital heart defects, patent ductus arteriosus, whole-exome sequencing, rare variants, single-nucleotide polymorphism

INTRODUCTION

The ductus arteriosus (DA) is a normal fetal structure that connects the pulmonary artery and descending aorta to maintain blood circulation during the fetal period (Benitz, W. E. et al., 2016). From the perspective of cardiac development, the DA functionally shuts down 15 h after birth in healthy, full-term infants (Crockett, S. L. et al., 2019). This process involves abrupt contraction of the muscular wall of the DA, which is associated with a proper balance among neurohumoral factors. An increase in the levels of contractile elements, such as peroxidase O2 and endothelin-1, and decrease in the levels of relaxants, such as prostaglandin E2 and nitric oxide, are the main events causing closure of the DA (Crockett, S. L. et al., 2019). Neural crest-derived cells migrate into the subendothelial space under the action of these hormones and transform into vascular smooth muscle cells (VSMCs). With contraction of the medial membrane and circular muscle in the DA, the lumen is shortened and
Failure of the DA to close after birth is termed patent DA (PDA), which is one of the most common heart defects, affecting approximately 1 in 2000 full-term infants and 8 in 1000 premature infants (Hoffman, J. I. et al., 2002). Persistent ductal shunting may lead to pulmonary overcirculation and induce systemic hypoperfusion, thereby increasing the risk of pulmonary hypertension, infective endocarditis, heart failure, and even death (Mitra, S. et al., 2018). However, its etiology and pathogenesis remain unclear.

PDA has both inherited and acquired causes. The preliminary understanding of the genetic mechanism of PDA was based on the studies in patients with syndrome. Previous studies have confirmed the association of several chromosomal syndromes, including Turner (45, XO), Kartagener, and Klinefelter (47, XXX), with PDA (Groth, K. A. et al., 2013; Gravholt, C. H. et al., 2019; Yang, D. et al., 2019). In addition to chromosomal rearrangements, a single gene mutation can also cause syndromic PDA, including Noonan (PTPN11 mutation), Holt-Oram (TBX5 mutation), and char (TFAP2B mutation) syndrome (Satoda, M. et al., 2000; Pannone, L. et al., 2017; vanlerbergh, C. et al., 2019). However, the genetic mechanism of nonsyndromic PDA (isolated findings without other abnormalities) remains unclear. Rare damaging mutations in MYH11 and TFAP2B were detected in several isolated nonsyndromic PDA patients (Harakalova, M. et al., 2013). Erdogan et al. (Erdogan, F. et al., 2008) performed an array comparative genome hybridization analysis of 105 patients with congenital heart defects and identified a 1.92 Mb deletion of chromosome 1q21.1 (CJA5) in a PDA patient. Genetic determinants of nonsyndromic PDA is still unknown.

Therefore, in this study, we recruited 39 unrelated nonsyndromic PDA patients and 100 healthy children for WES. Using a series of bioinformatics filtering steps, we identified 18 rare damaging variants in six candidate PDA-associated genes (SOX8, NES, CDH2, AKN3, EIF4G1, and HIPK1). Notably, these candidate genes were also highly expressed in human embryonic hearts. This identification of new pathogenic genes could help to elucidate the detailed underlying mechanism of PDA and promote further experimental analyses.

MATERIAL AND METHODS

Patients and Consent

Thirty-nine isolated nonsyndromic PDA patients of Han Chinese ethnicity and 100 healthy children (aged between 2 months and 13 years) were recruited from Xinhua Hospital affiliated with Shanghai Jiao Tong University (Shanghai, China). The structural heart phenotypes of all participants were assessed using echocardiography or cardiac catheterization. A diagnosis of PDA was made in the patient group by cardiac catheterization or surgery. Patients with a history of complex congenital heart disease were excluded from the study. The study protocol and ethics were approved by the Medical Ethics Committee of Xinhua Hospital. Informed consent was obtained from the parents of all participants. The study was conducted in accordance with the Declaration of Helsinki and the International Ethical Guidelines for Health-Related Research Involving Humans.

DNA Extraction and Whole-Exome Sequencing

The genomic DNA of all participants was extracted from blood samples using QIAamp DNA Blood Mini Kit (QIAGEN, Germany). DNA samples were stored at −80°C until further use. Genomic DNA was eluted, purified, amplified by ligation-mediated polymerase chain reaction, and then subjected to DNA sequencing on an Illumina platform. The target depth of the DNA sequencing was x100. Qualified DNA samples from the PDA and control groups were subjected to WES to detect rare variations. Read quality was checked using Fastp software (Chen, S. et al., 2018) and raw sequence data were aligned to human genome (human_glk_v37) using BWA (v0.7.12-r1039). Duplicated and low-quality reads (Per base sequence quality <20) were removed by using Picard software (https://broadinstitute.github.io/picard). Alignment quality was assessed using qualimap software (Okonechnikov, K. et al., 2016).

Single-Nucleotide Polymorphism Identification and Quality Filtering

Single nucleotide polymorphisms (SNPs) account for much of the phenotypic diversity among individuals.

SNPs and insertions/deletions were detected using the HaplotypeCaller module of GATK4 software (Mckenna, A. et al., 2010), based on sequence alignment of the clinical samples to the reference genome. Before detection, we recalibrated the base qualities using the BaseRecalibrator module of GATK4 software (Mckenna, A. et al., 2010) to improve variant detection accuracy based on the quality with a depth (QD) criterion >2. The resulting BAM files were then sorted, indexed, and processed using base quality score recalibration (Okonechnikov, K. et al., 2016). The GATK HaplotypeCaller module was then used for variant calling. We used ANNOVAR53 (Wang, K. et al., 2010) to annotate the variants for functional and population frequency information with the 1000 Genomes (Clarke, L. et al., 2012), Refseq (O’Leary, N. A. et al., 2016), ExAC (Karczewski, K. J. et al., 2017), ESP6500 (Liang, Y. et al., 2019), gnomAD, SIFT (Flanagan, S. E. et al., 2010), clinvar (Landrum, M. J. et al., 2020), PolyPhen (Flanagan, S. E. et al., 2010), MutationTaster (Steinhaus, R. et al., 2021), COSMIC (Forbes, S. A. et al., 2011), gwasCatalog, and OMIM databases (Amberger, J. S. et al., 2017). All potentially damaging variants of the candidate genes were classified into five groups: pathogenic, likely pathogenic, variant of uncertain significance, likely benign, and benign (Richards, S. et al., 2015). Finally, the rare damaging variants were filtered according to the American College of Medical Genetics criteria guidelines (Figure 1).
Variant Filtering Based on Fisher’s Exact Test and Burden Analysis

The difference in allele frequency for each SNP between cases and controls was compared using the Fisher’s exact test with R statistical software packages; a p-value < 0.05 was considered statistically significant. Subsequently, we aggregated the SNP data based on gene expression levels and conducted a gene-based burden analysis to increase statistical power. Candidate pathogenic genes were filtered based on the results of burden analysis according to the following criteria: 1) p-value or false-discovery rate (FDR) < 0.05, 2) hit for at least one variant in three cases, and 3) not found in any sample of the control group. We then prioritized genes based on the p value of Fisher’s exact test and burden analysis.

Functional Enrichment and Network Analysis

To further filter the candidate genes associated with PDA, we performed functional enrichment analysis to identify the functions of candidate genes identified through the aforementioned filtering steps. Pathway analysis of the candidate gene profiling results was performed using Gene Ontology (GO; version 30.10.2017) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway (http://www.genome.jp/kegg/pathway.html) mapping within the web-based tool Database for Annotation, Visualization, and Integrated Discovery (Gene Ontology, C. 2015; Kanehisa, M. et al., 2017). GO terms represent a network of biological processes that overlap in space and are clustered according to their relationships (Gene Ontology, C. 2015). The threshold was set to an adjusted p-value < 0.05. In addition, we prioritized these genes based on functional enrichment analysis. Furthermore, to detect the relationship between the candidate genes and known disease-causing genes, we constructed the protein–protein interaction (PPI) network (Brohee, S. et al., 2008) using Cytoscape software based on the STRING database.

Tissue Collection and Expression Detection

In addition to the genes prioritized using the steps described above, we further prioritized genes according to their expression levels in the human embryonic heart. Previous studies have divided eight embryonic weeks (56 days) into 23 internationally accepted Carnegie stages (O’rahilly, R. 1987). To further investigate the potential function of our candidate genes, human embryonic hearts in different Carnegie stages (S10–S16) were collected after medical termination of pregnancy from patients at Xinhua Hospital. RNA was extracted and purified using the Experion automated gel electrophoresis system and RNeasy MinElute Cleanup Kit. The expression patterns of candidate genes were subsequently detected using the Affymetrix HTA 2.0 microarray.

RESULTS

Population

Among the 39 patients, 28% had common cardiac defects, including atrial septal defect (n = 7), ventricular septal defect
Table 1: Characteristics of 39 PDA patients.

| Characteristics          | Numbers   |
|--------------------------|-----------|
| Age (year)               | 2.92 ± 2.44 |
| Male-to-female ratio (%) | 62%       |
| BMI (kg/m²)              | 16.58 ± 4.34 |
| PDA size (mm)            | 2.87 ± 1.68 |
| Birth weight (kg)        | 2.96 ± 0.73 |
| Gestational age (week)   | 39.04 ± 1.46 |
| Associated cardiac defect n (%) | 2 (5%) |
| VSD n (%)                | 2 (5%)    |
| ASD n (%)                | 7 (18%)   |
| Others n (%)             | 2 (5%)    |

All values are expressed as mean ± SD or n (%).

Substitutions accounted for the majority of single-base mutations compared with other types (Figure 2). Based on these mutations, we adopted a bioinformatics filtering strategy to identify candidate genes associated with PDA.

Variants Identified Based on Fisher’s Exact Test

Based on the results of Fisher’s exact test, we identified 44 variants that were more frequently detected in the PDA group than in the control group (FDR <0.05, p < 0.05), as presented in Table 2 (p < 0.01). We then prioritized these variants based on the p-value; the top 10 variants with statistical significance are shown in Figure 3. Notably, we found that the SNPs rs103826685 and rs32552095 located in SLC9B1 and HLA-DRB1, respectively, had the most significantly different frequencies between the patient and control groups (p < 0.0001).

Candidate Genes Identified Based on Burden Analysis

To further increase statistical power, we aggregated the SNP data at the gene level and performed burden analysis. Under a significance threshold of 0.05, we observed 57 genes with potential pathogenicity as candidate PDA-associated genes (Table 3 (p < 0.01)). We then prioritized these genes based
on the p-value from the burden analysis; the top 10 genes with statistical significance are displayed as a heatmap in Figure 4. The top three genes with high confidence were NPIPB5, SLC9B1, and HLA-DRB1. Notably, SLC9B1 and HLA-DRB1 were also in the top significant genes based on Fisher’s exact test.

**Functional Analysis**

Functional enrichment analysis of the 101 candidate differentially expressed genes identified through Fisher’s exact test and burden analysis revealed that the main enriched GO terms in the upregulated gene set were thiol-dependent ubiquitinyl hydrolase activity (TermID: GO:0036459), peptide antigen binding (TermID: GO:0042605), and ubiquitin-dependent protein catabolic process (TermID: GO:0006511). Particular focus was placed on terms representing pro stagnalgin, apoptosis, and heart development (Figure 5). Moreover, KEGG analysis of the direct gene targets in PDA patients revealed enrichment in pathways related to cell adhesion molecules (TermID: path: hsa04514, p < 0.001), viral myocarditis (TermID: path: hsa05416, p = 0.0035), and asthma (TermID: path: hsa05310, p = 0.01; Figure 6). Based on functional enrichment analysis, 29 pathway genes related to cardiovascular development were screened.

**Network Analysis**

To further explore their roles, the 29 candidate genes were mapped to construct a PPI network along with 240 known pathogenic genes involved in cardiovascular development (Supplementary Table S1). The 240 known genes from the literature were divided into two groups related to cardiovascular development and PDA, respectively. In the network, the candidate genes NES and CDH2 showed the most direct and strongest relationship with known pathogenic genes in both groups. Moreover, CDH2 and NES had the highest molecular weights and were located at the center of the PPI network (Figures 7, 8). Therefore, based on the degree of correlation, we screened out 11 candidate genes for final verification.

**Detection of Candidate Gene Expression in the Human Embryonic Heart**

To further investigate the potential function of our candidate genes, we detected the expression levels of the 11 screened out genes in human embryonic hearts at different Carnegie stages. After prioritizing the candidate genes based on expression levels, the final six pathogenic genes (SOX8, NES, CDH2, ANK3, EIF4G1, and HIPK1) were identified (Figure 9). Among them, CDH2 was the most highly expressed in the embryonic heart (Figure 10).

### TABLE 2 | SNP filtering Based on Fisher Exact Test.

| Chromosome | Gene    | Mutation position | Mutation type | p-value |
|------------|---------|-------------------|---------------|---------|
| 1          | LRRC8C  | 90179703          | T > G         | 0.006   |
| 1          | NES     | 156640657         | A > C         | 0.000   |
| 11         | LRRC4C  | 40136434          | C > T         | 0.006   |
| 14         | SLC7A8  | 23612372          | T > G         | 0.001   |
| 16         | SOX8    | 1034733           | A > C         | 0.006   |
| 16         | NPIPA1  | 15045634          | T > C         | 0.000   |
| 16         | NPIPB5  | 22545668          | A > C         | 0.000   |
| 16         | NPIPB5  | 22546505          | G > T         | 0.000   |
| 19         | MAPK10  | 40719910          | C > G         | 0.006   |
| 3          | ZNF717  | 75786264          | G > T         | 0.001   |
| 3          | EIF4G1  | 184033821         | G > C         | 0.000   |
| 3          | MUC4    | 195506722         | G > A         | 0.001   |
| 3          | MUC4    | 195506723         | T > G         | 0.001   |
| 4          | MUC4    | 195514174         | G > T         | 0.000   |
| 4          | USP17L20| 9217567           | C > A         | 0.006   |
| 4          | USP17L17| 9246041           | C > A         | 0.006   |
| 4          | SLC9B1  | 103826685         | T > A         | 0.000   |
| 4          | LRBA    | 151770608         | A > C         | 0.006   |
| 5          | VARS    | 31746821          | G > A         | 0.006   |
| 6          | HLA-DRB5| 32487344          | T > C         | 0.000   |
| 6          | HLA-DRB1| 32552095          | C > T         | 0.000   |
| 7          | TOAF2   | 143400090         | G > A         | 0.001   |

![FIGURE 3](image3.png) Single SNP allele frequency and genotype frequency p-values were obtained using the Fisher exact test. X-axis represents the position of each SNP (represented in circles) on human chromosome, Y-axis is the-log p-value of Fisher Exact test. Top 10 variants in our study were represented in the figure.
TABLE 3 | Gene filtering based on Burden analysis.

| Gene     | Case mutation | Case normal | Control mutation | Control normal | p-value |
|----------|---------------|-------------|------------------|----------------|---------|
| ASIC3    | 5             | 34          | 0                | 100            | 0.001   |
| CFAP45   | 4             | 35          | 0                | 100            | 0.006   |
| CYP21A2  | 4             | 35          | 0                | 100            | 0.006   |
| EV5      | 4             | 35          | 0                | 100            | 0.006   |
| HIPK1    | 4             | 35          | 0                | 100            | 0.006   |
| HLA-DRB1 | 25            | 14          | 0                | 100            | 0.000   |
| HLA-DRB5 | 6             | 33          | 0                | 100            | 0.000   |
| KRT39    | 5             | 34          | 0                | 100            | 0.001   |
| LRRC4C   | 4             | 35          | 0                | 100            | 0.006   |
| MAP3K10  | 4             | 35          | 0                | 100            | 0.006   |
| NPIPA1   | 13            | 26          | 0                | 100            | 0.000   |
| NPIP85   | 31            | 8           | 0                | 100            | 0.000   |
| POTEE    | 5             | 34          | 0                | 100            | 0.001   |
| SLC9B1   | 29            | 10          | 0                | 100            | 0.000   |
| SLX4     | 4             | 35          | 0                | 100            | 0.006   |
| SOX8     | 4             | 35          | 0                | 100            | 0.006   |
| TBC1D3F  | 4             | 35          | 0                | 100            | 0.006   |
| TCAF2    | 5             | 34          | 0                | 100            | 0.001   |
| USP17L11 | 5             | 34          | 0                | 100            | 0.001   |
| USP17L17 | 6             | 33          | 0                | 100            | 0.000   |
| USP17L18 | 5             | 34          | 0                | 100            | 0.001   |
| USP17L2  | 6             | 33          | 0                | 100            | 0.000   |
| USP17L20 | 5             | 34          | 0                | 100            | 0.001   |
| VARS     | 4             | 35          | 0                | 100            | 0.006   |
| ZNF717   | 6             | 33          | 0                | 100            | 0.000   |

FIGURE 4 | Heatmap representing the top 10 genes identified in Burden analysis. Heatmap that shows the mutational burden (p-value<0.05) of the top 10 gene based on gene-based burden analysis in PDA patients. The heatmap was generated by using R package, the mutation values were normalized per gene over all PDA samples. Each box in the heatmap represent a single variant in a case, with the dark red indicating high gene mutation ration in gene-based Burden analysis.
DISCUSSION

The underlying molecular genetic mechanisms of PDA remain largely unknown as one of the most common congenital heart defects. In this study, we explored the clinical characteristics of 39 PDA patients and 100 healthy controls by performing WES to identify rare variants and candidate PDA-associated genes. Through a series of bioinformatic filtering strategies, we prioritized the candidate genes via Fisher’s exact test, mutation burden analysis, gene network construction, and expression levels in embryonic hearts. Finally, we identified 18 rare damaging variants in six novel candidate genes (SOX8, NES, CDH2, ANK3, EIF4G1, and HIPK1) associated with PDA. Among these, CDH2 was highly expressed in the human embryonic heart and appears to be the most important candidate gene identified in our study.

CDH2 encodes N-cadherin, a member of a protein family regulating cadherin-mediated cell–cell adhesion in multiple tissues. The structure comprises a single transmembrane domain, cytoplasmic domain, and five conserved extracellular cadherin domains (ECI–V) (Alimperti, S. et al., 2015). We found two variants (rs25565020 and rs25532304) in CDH2 in four patients with PDA. In addition, CDH2 had the highest molecular weight and was located at the center of the PPI network, both among known CHD- and PDA-related genes. Further investigation showed that CDH2 is highly expressed in human embryonic hearts. Previous studies in mice have also noted the importance of CDH2 in the proper development of the heart, brain, and skeletal structures (Radice, G. L. et al., 1997). Moreover, genetic analyses in zebrafish revealed that mutation in the EC-I or EC-IV domains of cdh2 play important role in embryonic development (Masai, I. et al., 2003). Mayosi, B. M. et al. (2017) used WES to detect novel rare variants in patients with arrhythmogenic cardiomyopathy and found that CDH2 mutation changes the conserved amino acids of CDH2 protein. Since the relationship between CDH2 and PDA is unclear, additional studies are needed to determine how genetic perturbations of CDH2 contribute to PDA.

In our study, 16 patients (42%) had the same variant (rs156646936) in NES. In the network analysis, we observed a strong correlation between NES and known pathogenic genes. NES belongs to the human tissue kallikrein family of secreted serine proteases (Luo, L. et al., 1998), which play an important role in carcinogenesis, including in breast, prostate, and testicular cancers, and leukemia (Luo, L. Y. et al., 2001). Further experimental evidence suggests that the function of NES as a tumor suppressor...
may be achieved by hypermethylation of the CpG islands (Li, B. et al., 2001). However, this is the first report of NES mutations in PDA. ANK3 is a member of the ankyrin family, which is expressed in several different isoforms in many tissues. ANK3 plays key roles in cell motility, activation, proliferation, contact, and the maintenance of specialized membrane domains. In our study, eight patients (10%) had variants in ANK3. ANK3 variants have previously been associated with schizophrenia, autism, epilepsy, and intellectual disability (Leussis, M. P. et al., 2013; Wirgemes, K. V. et al., 2014). Studies from knockout mouse models have revealed that loss of ANK3 function leads to defects in cardiac calcium handling and arrhythmias (Mohler, P. J. et al., 2004). Although the roles of NES and ANK3 in the pathogenesis of PDA are supported by bioinformatic analyses, our study was limited by the lack of experimental evidence to validate the deleteriousness of the variants.

EIF4G1 encodes a protein, that is, a component of the multi-subunit protein complex EIF4F. EIF4G plays a crucial role in translation initiation and serves as a scaffolding protein that binds several initiation factors (the cap-binding protein eIF4E, the RNA helicase eIF4A, and eIF3) (Haimov, O. et al., 2018). In our study, 15 patients had three types of variants in EIF4G1, and the same variant (rs184033621) was detected in 14 patients. EIF4G1 modulates the proliferation, apoptosis, and angiogenesis of most tumor types by limiting steps during the initiation phase of protein synthesis and interacting with ubiquitin-specific protease 10 (USP10) (Cao, Y. et al., 2016). Moreover, EIF4G1 phosphorylation specifically activates the PKC-Ras-ERK

FIGURE 7 | Interaction between our candidate genes and known CHD-related genes. PPI network was generated by Cytoscape software and our candidate pathogenic genes and the known CHD-related genes were uploaded in STRING database. Each node represents one gene, and each edge represents the protein-protein interaction collected from BioGRID.
signaling pathway, which is involved in the control of cell growth and proliferation (Dobrikov, M. et al., 2011). Diseases associated with EIF4G1 include Parkinson’s disease, nonsmall cell lung carcinoma, and prostate cancer (Cao, Y. et al., 2016). Although the relationship between EIF4G1 and cardiovascular development remains unknown, our results suggest that EIF4G1 might be potentially pathogenic in terms of PDA.

HIPK1 belongs to the Ser/Thr family of protein kinases as part of the HIPK subfamily. HIPK1 is related to pathways involved in the regulation of TP53 activity and cardiac conduction. The homeodomain-interacting protein kinases HIPK1 and HIPK2 play key roles in embryonic development by regulating transforming growth factor β-dependent angiogenesis (Aikawa, Y. et al., 2006; Shang, Y. et al., 2013). HIPK1 loss-of-function conditional knockout mice exhibit defects in primitive/definitive hematopoiesis, vasculogenesis, angiogenesis, and neural tube closure (Shang, Y. et al., 2013). In addition, HIPK1 can interact with homeobox proteins and other transcription factors to regulate various biological processes, including signal transduction, apoptosis, embryonic development, and retinal vascular dysfunction (Aikawa, Y. et al., 2006). In our study, only two HIPK1 variants (rs114516009 and rs114506069) were detected in four individuals with PDA; these are novel variants that have not been reported previously. Further investigation showed that HIPK1 is highly expressed in human embryonic hearts. However, additional experiments are needed to determine the genetic mechanism by which HIPK1 contributes to PDA.

SOX8 is a member of the SRY-related HMG-box (SOX) family of transcription factors, which are involved in the regulation of embryonic development and in determining cell fate (Haseeb, A. et al., 2019). In our study, the same rare variant (rs1034733) was detected in three patients with PDA. SOX8 expression is essential in the developing heart, which correlates with heart septation and differentiation of the connective tissue of the valve leaflets (Montero, J. A. et al., 2002). Moreover, a previous study revealed that SOX8 overexpression might be associated with hypoxia-induced cell injury by activating the PI3K/AKT/mTOR and MAPK pathways (Gong, L. C. et al., 2017). Interestingly, DA closure after birth is closely related to the blood oxygenation level, and hypoxia can lead to an increase in endogenous PGE2 release, which directly leads to opening of the DA (Benitz, W. E. et al., 2016). Therefore, SOX8 may be a novel candidate gene involved in the pathogenesis of PDA.
In conclusion, through a series of bioinformatics filtering steps, we identified 18 rare damaging variants in six novel candidate genes (SOX8, NES, CDH2, ANK3, EIF4G1, and HIPK1) associated with PDA. The discovery of these genes opens up a new field for genetic research on PDA and provides new ideas for understanding the pathogenesis of PDA. Nevertheless, our study has some limitations. The lack of parental samples and the small sample size limited our ability to identify the detailed genetic background of PDA. Thus, more fundamental research is needed to determine candidate genes that contribute to PDA. We hope to confirm these findings with larger sample sizes.

**DATA AVAILABILITY STATEMENT**

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, SRP288538.
ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Medical Ethics Committee of Xinhua Hospital. Written informed consent to participate in this study was provided by the participants’ legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

PZ contributed to design of the study and performed the statistical analysis. YC, BC, and QZ collected the blood samples from all subjects. YG and DW wrote the first draft of the manuscript and contributed to this study equally. PZ revised the manuscript. All authors have read and agreed to the published version of the manuscript.

FUNDING

This study received financial supports from National Natural Science Foundation of China (82070386), the Project of Shanghai Municipal Health Commission (Grant No.201940393).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.921925/full#supplementary-material

REFERENCES

Aikawa, Y., Nguyen, L. A., Isomoto, K., Takakura, N., Tagata, Y., Schmitz, M. L., et al. (2006). Roles of HIPK1 and HIPK2 in AML1- and P300-dependent Transcription, Hematopoiesis and Blood Vessel Formation. EMBO J. 25 (17), 3955–3965. doi:10.1038/sj.emboj.7601273

Alimperti, S., and Andreadis, S. T. (2015). CDH2 and CDH11 Act as Regulators of Stem Cell Fate Decisions. Stem Cell Res. 14 (3), 270–282. doi:10.1016/j.scr.2015.02.002

Amberger, J. S., and Hamosh, A. (2017). Searching Online Mendelian Inheritance in Man (OMIM): A Knowledgebase of Human Genes and Genetic Phenotypes. Curr. Protoc. Bioinformatics. 58, 1–12. doi:10.1002/cpbi.27

Benit, W. E., Watterberg, K. L., Cummings, S. J. J., Eichenwald, E. C., Goldsmith, J., Poindeix, B. B., et al. (2016). Patent Ductus Arteriosus in Preterm Infants. Pediatrics 137 (1). doi:10.1542/peds.2015-3730

Brodée, S., Faust, K., Lima-Mendes, G., Vanderstocken, G., and van Helden, J. (2008). Network Analysis Tools: from Biological Networks to Clusters and Pathways. Nat. Protoc. 3 (10), 1616–1629. doi:10.1038/nprot.2008.110

Cao, Y., Wei, M., Li, B., Yu, Y., Tang, Z., et al. (2016). Functional Role of Eukaryotic Translation Initiation Factor 4 Gamma 1 (eIF4G1) in NSCLC. Oncotarget 7 (17), 24224–24251. doi:10.18632/oncotarget.8168

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: an Ultra-fast All-In-One FASTQ Preprocessor. Bioinformatics 34 (17), i884–i890. doi:10.1093/bioinformatics/bty560

Clarke, L., Zheng-Bradley, X., Zheng-Bradley, X., Smith, R., Kulesha, E., Xiao, C., et al. (2012). The 1000 Genomes Project: Data Management and Community Access. Nat. Methods 9 (5), 459–462. doi:10.1038/nmeth.1974

Crockett, S. L., Berger, C. D., Shelton, E. L., and Reese, J. (2019). Molecular and Mechanical Factors Contributing to Ductus Arteriosus Patency and Closure. Congenit. Heart Dis. 14 (1), 15–20. doi:10.1111/chd.12714

Dobrikov, M., Dobrikova, E., Shveygert, M., and Gromeier, M. (2011). Phosphorylation of Eukaryotic Translation Initiation Factor 4G1 (eIF4G1) by Protein Kinase Ca Regulates eIF4G1 Binding to Mrn1. Mol. Cell. Biol. 31 (14), 2947–2959. doi:10.1128/MCB.05589-11

Erdogan, F., Larsen, L. A., Zhang, L., Turner, Z., Tommerup, N., Chen, W., et al. (2008). High Frequency of Submicroscopic Genomic Aberrations Detected by Tiling Path Array Comparative Genome Hybridisation in Patients with Isolated Congenital Heart Disease. J. Med. Genet. 45 (11), 704–709. doi:10.1136/jmg.2008.058776

Flanagan, S. E., Patch, A. M., and Ellard, S. (2010). Using SIFT and PolyPhen to Predict Loss-Of-Function and Gain-Of-Function Mutations. Genet. Test. Mol. Biomarkers 14 (4), 533–537. doi:10.3109/108365710.00036

Forbes, S. A., Bindal, N., Bamford, S., Cole, C., Kok, C. Y., Bearé, D., et al. (2011). COSMIC: Mining Complete Cancer Genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res. 39 (Database issue), D945–D950. doi:10.1093/nar/gkr929

Gong, L.-C., Xu, H.-M., Guo, G.-L., Zhang, T., Shi, J.-W., and Chang, C. (2017). Long Non-coding RNA H19 Protects H9c2 Cells against Hypoxia-Induced Injury by Targeting MicroRNA-139. Cell. Physiol. Biochem. 44 (3), 857–869. doi:10.1159/000485354

Gravholt, C. H., Viuff, M. H., Brun, S., Stochholm, K., and Andersen, N. H. (2019). Turner Syndrome: Mechanisms and Management. Nat. Rev. Endocrinol. 15 (10), 601–614. doi:10.1038/s41574-019-0224-4

Groth, K. A., Skakkebæk, A., Host, C., Gravholt, C. H., and Bojesen, A. (2013). Klinefelter Syndrome–A Clinical Update. J. Clin. Endocrinol. Metabolism 98 (1), 20–30. doi:10.1210/jc.2012-2382

Haimov, O., Shcharaw, U., Tabarkin-Ben Harush, A., Bahat, A., Uzoniy, A., Will, A., et al. (2018). Dynamic Interaction of Eukaryotic Initiation Factor 4G1 (eIF4G1) with eIF4E and eIF1 Underlying Scanning-dependent and -Independent Translation. Mol. Cell. Biol. 38 (18). 1. doi:10.1128/MCB.00139-18

Harakalova, M., van der Smagt, J., de Kovel, C. G. F., Van’t Slot, R., Poot, M., Nijman, I. J., et al. (2013). Incomplete Segregation of MYH11 Variants with Thoracic Aortic Aneurysms and Dissections and Patent Ductus Arteriosus. Eur. J. Hum. Genet. 21 (5), 487–493. doi:10.1038/ejhg.2012.206

Haseeb, A., and Lefebvre, V. (2019). The SOXE Transcription Factors-SOX8, SOX9 and SOX10-Share a Bi-partite Transactivation Mechanism. Nucleic Acids Res. 47 (13), 6917–6931. doi:10.1093/nar/gkz523

Hoffman, J. I. E., and Kaplan, S. (2002). The Incidence of Congenital Heart Disease. J. Am. Coll. Cardiol. 39 (12), 1890–1900. doi:10.1016/s0735-1097(02)01886-7

Kanehsia, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs. Nucleic Acids Res. 45 (D1), D355–D361. doi:10.1093/nar/gkw1092

Karczewski, K. J., Weisburd, B., Thomas, B., Soloromon, M., Ruderfer, D. M., Kavanagh, D., et al. (2017). The ExAC Browser: Displaying Reference Data Information from over 60 000 Exomes. Nucleic Acids Res. 45 (D1), D840–D845. doi:10.1093/nar/gkw971

Landrum, M. J., Chitipiralla, S., Brown, G. R., Chen, C., Gu, B., Hart, J., et al. (2020). ClinVar: Improvements to Accessing Data. Nucleic Acids Res. 48 (D1), D835–D844. doi:10.1093/nar/gka972

Leusiss, M. P., Berry-Scott, E. M., Saito, M., Huang, H., de Haan, G., Alkan, O., et al. (2013). The ANK3 Bipolar Disorder Gene Regulates Psychiatric-Related Behaviors That Are Modulated by Lithium and Stress. Biol. Psychiatry 73 (7), 683–690. doi:10.1016/j.biopsych.2012.10.016

Li, B., Goyal, J., Dhar, S., Dimri, G., Evron, E., Sukumar, S., et al. (2001). CpG Methylation as a Basis for Breast Tumor-specific Loss of NES1/kallkrein 10 Expression. Cancer Res. 61 (21), 8014–8021.

Li, N., Subrahmanyan, L., Smith, E., Yu, X., Zaidi, S., Choi, M., et al. (2016). Mutations in the Histone Modifier PRDM6 Are Associated with Isolated Nonsyndromic Patent Ductus Arteriosus. Am. J. Hum. Genet. 98 (6), 1082–1091. doi:10.1016/j.ajhg.2016.03.022
Liang, Y., Jiang, L., Zhong, X., Hochwald, S. N., Wang, Y., Huang, L., et al. (2019). Discovery of Aberrant Alteration of Genome in Colorectal Cancer by Exome Sequencing. *Am. J. Med. Sci.* 358 (5), 340–349. doi:10.1016/j.amjms.2019.07.012

Luo, L.-Y., Metys, E. R.-D., Jung, K., and Diamandis, E. P. (2001). Expression of the Normal Epithelial Cell-specific 1 (NES1; KLK10) Candidate Tumour Suppressor Gene in Normal and Malignant Testicular Tissue. *Br. J. Cancer* 85 (2), 220–224. doi:10.1054/bjoc.2001.1870

Luo, L., Herbrick, J.-A., Scherer, S. W., Beatty, B., Squier, J., and Diamandis, E. P. (1998). Structural Characterization and Mapping of the Normal Epithelial Cell-specific 1 Gene. *Biochem. Biophysical Res. Commun.* 247 (3), 580–586. doi:10.1006/bbrc.1998.8793

Masai, I., Iele, Z., Yamaguchi, M., Komori, A., Nakata, A., Nishiwaki, Y., et al. (2003). N-cadherin Mediates Retinal Lamination, Maintenance of Forebrain Compartments and Patterning of Retinal Neurites. *Development* 130 (11), 2479–2494. doi:10.1242/dev.004465

Mayosi, B. M., Fish, M., Shaboodien, G., Mastantuono, E., Kraus, S., Wieland, T., et al. (2017). Identification of Cadherin 2 (CDH2) Mutations in Arrhythmogenic Right Ventricular Cardiomyopathy. *Circ. Cardiovasc. Genet.* 10 (2). 1. doi:10.1161/CIRCGENETICS.116.001605

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data. *Genome Res.* 20 (9), 1297–1303. doi:10.1101/gr.107524.110

Mitra, S., Flores, I. D., Tamayo, M. E., Mbaagbow, L., Vanniyasingam, T., Veroniki, A. A., et al. (2018). Association of Placebo, Indomethacin, Buprofen, and Acetaminophen with Closure of Hemodynamically Significant Patent Ductus Arteriosus in Preterm Infants. *JAMA* 319 (12), 1221–1238. doi:10.1001/jama.2018.1896

Mohler, P. J., Splawski, I., Napolitano, C., Bottelli, G., Sharpe, L., Timothy, K., et al. (2004). A Cardiac Arrhythmia Syndrome Caused by Loss of Ankyrin-B Function. *Proc. Natl. Acad. Sci. U.S.A.* 101 (24), 9137–9142. doi:10.1073/pnas.0402546101

Montero, J. A., Giron, B., Arrechedera, H., Cheng, Y. C., Scotting, P., Chimal-Monroy, J. I., et al. (2002). Expression of Sox8, Sox9 and Sox10 in the Developing Nerves of the Embryonic Heart. *Mech. Dev.* 118 (12), 199–202. doi:10.1016/s0925-4773(02)00249-6

O’Leary, N. A., Wright, M. W., Brister, J. R., Ciufio, S., Haddad, D., McVeigh, R., et al. (2016). Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation. *Nucleic Acids Res.* 44 (D1), D733–D745. doi:10.1093/nar/gkv1189

O’Rahilly, R. (1987). Human Embryo. *Nature* 329 (6138), 385. doi:10.1038/329385a0

Okonechnikov, K., Conesa, A., and García-Alcalde, F. (2016). Qualimap 2: Advanced Multi-Sample Quality Control for High-Throughput Sequencing Data. *Bioinformatics* 32 (2), btv566–294. doi:10.1093/bioinformatics/btv566

Pannone, L., Bocchinfuso, G., Flex, E., Rossi, C., Baldassarre, G., Lissewski, C., et al. (2017). Structural, Functional, and Clinical Characterization of a NovelPTPN11Mutation Cluster Underlying Noonan Syndrome. *Hum. Mutat.* 38 (4), 451–459. doi:10.1002/humu.23175

Radlice, G. L., Rayburn, H., Matsunami, H., Knudsen, K. A., Takeichi, M., and Hynes, R. O. (1997). Developmental Defects in Mouse Embryos Lacking N-Cadherin. *Dev. Biol.* 181 (1), 64–78. doi:10.1006/dbio.1996.8443

Richards, S., Aziz, N., Bale, S., Dick, D., S., Gastier-Foster, J., et al. (2015). Standards and Guidelines for the Interpretation of Sequence Variants: a Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* 17 (5), 405–424. doi:10.1038/gim.2015.30

Satoda, M., Zhao, F., Diaz, G. A., Burn, J., Goodship, J., Davidson, H. R., et al. (2000). Mutations in TAF1P2 Cause Char Syndrome, a Familial Form of Patent Ductus Arteriosus. *Nat. Genet.* 25 (1), 42–46. doi:10.1038/755758

Shang, Y., Doan, C. N., Arnold, T. D., Lee, S., Tang, A. A., Reichardt, L. F., et al. (2013). Transcriptional Corepressors HIPK1 and HIPK2 Control Angiogenesis via TGF-β-TAK1-dependent Mechanism. *PLoS Biol.* 11 (4), e1001527. doi:10.1371/journal.pbio.1001527

Steinhaus, R., Proft, S., Schuellek, M., Cooper, D. N., Schwarz, J. M., and Seelow, D. (2021). MutationTaster2021. *Nucleic Acids Res.* 49 (W1), W446–W451. doi:10.1093/nar/gkab266

Vanlerberge, C., Jourdain, A.-S., Ghoumid, J., Frenois, F., Mezel, A., Vaksman, G., et al. (2019). Holt–Oram Syndrome: Clinical and Molecular Description of 78 Patients with TBX3 Variants. *Eur. J. Hum. Genet.* 27 (3), 360–368. doi:10.1038/s41431-018-0303-3

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data. *Nucleic Acids Res.* 38 (16), e164. doi:10.1093/nar/gkq603

Wirgenes, K. V., Tesli, M., Ingham, E., Athanasiou, L., Agartz, I., Melle, I., et al. (2014). ANK3 Gene Expression in Bipolar Disorder and Schizophrenia. *Br. J. Psychiatry* 205 (3), 244–245. doi:10.1192/bjp.bp.114.145433

Yang, D., Liu, B. C., Luo, J., Huang, T. X., and Liu, C. T. (2019). Kartagener Syndrome. *QJM* 112 (4), 297–298. doi:10.1093/qjmed/hcz242

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Gao, Wu, Chen, Chen, Zhang and Zhao. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.