Research paper

Next-generation sequencing identifies novel genes with rare variants in total anomalous pulmonary venous connection

Xin Shi [a,1], Tao Huang [b,1], Jing Wang [a,1], Yulai Liang [b], Chang Gu [c], Yuejuan Xu [a], Jing Sun [a], Yanan Lu [a], Kun Sun [a,*], Sun Chen [a,*], Yu Yu [a,*]

[a] Department of Pediatric Cardiovascular, Xin Hua Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200092, China
[b] Institute of Health Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China
[c] Department of Thoracic Surgery, Shanghai Pulmonary Hospital, Tong Ji University School of Medicine, Shanghai 200433, China

ABSTRACT

Background: Total anomalous pulmonary venous connection (TAPVC) is recognized as a rare congenital heart defect (CHD). With a high mortality rate of approximately 80%, the survival rate and outcomes of TAPVC patients are not satisfactory. However, the genetic aetiology and mechanism of TAPVC remain elusive. This study aimed to investigate the underlying genomic risks of TAPVC through next-generation sequencing (NGS).
Methods: Rare variants were identified through whole exome sequencing (WES) of 78 sporadic TAPVC cases and 100 healthy controls using Fisher's exact test and gene-based burden test. We then detected candidate gene expression patterns in cells, pulmonary vein tissues, and embryos. Finally, we validated these genes using target sequencing (TS) in another 100 TAPVC cases.
Findings: We identified 42 rare variants of 7 genes (*CLTCL1, CST3, GXYLT1, HMGA2, SNAI1, VAV2, ZDHHC8*) in TAPVC cases compared with controls. These genes were highly expressed in human umbilical vein endothelial cells (HUVECs), mouse pulmonary veins and human embryonic hearts. mRNA levels of these genes in human pulmonary vein samples were significantly different between cases and controls. Through network analysis and expression patterns in zebrafish embryos, we revealed that *SNAI1, HMGA2* and *VAV2* are the most important genes for TAPVC.
Interpretation: Our study identifies novel candidate genes potentially related to TAPVC and elucidates the possible molecular pathogenesis of this rare congenital birth defect. Furthermore, *SNAI1, HMGA2* and *VAV2* are novel TAPVC candidate genes that have not been reported previously in either humans or animals.
Fund: National Natural Science Foundation of China.

1. Introduction

Congenital heart diseases (CHDs) are the most common birth defects in humans, affecting approximately 1% of the population, and are the leading cause of birth defect-related infant mortality [1,2]. Total anomalous pulmonary venous connection (TAPVC) is a rare CHD in which all 4 pulmonary veins fail to link to the left atrium correctly but make abnormal connections to the right atrium or systemic venous system [3]. TAPVC accounts for approximately 1–3% of all CHDs, with a morbidity of approximately 1 out of 15,000 live births [4]. In addition, the mortality of TAPVC patients without proper intervention is nearly 80% [5].

Since the first case of TAPVC was described in 1960, the embryology of TAPVC has been the subject of increasing exploration. However, the molecular mechanism underlying pulmonary vein morphogenesis remains unknown. Only a few genes have been identified as candidate genes for TAPVC pathogenesis. Bleyl et al. revealed that loci on human chromosome 4q12 are involved in TAPVC using genetic linkage analysis, and the candidate genes in this region include *VEGFR2* and *PDGFR2* [6]. Cinquetti et al. observed increased *ANKRD1* gene expression in lymphoblastoid cells derived from a TAPVC patient, indicating that *ANKRD1* is related to TAPVC [7]. Karl et al. demonstrated *SEMA3D* as a crucial gene in pulmonary venous connection because $SEMA3D^{-/-}$ mice displayed TAPVC or partial APVC (PAPVC) phenotypes [8]. These genes explain only a small fraction of the molecular mechanism underlying TAPVC pathogenesis, and comprehensive genomic data obtained via massive parallel sequencing are still lacking.

* Corresponding authors at: Department of Pediatric Cardiology, Xin Hua Hospital, 1665 Kongjiang Road, Shanghai 200092, China
E-mail addresses: sunkun@xinhuamed.com.cn (K. Sun), chensun@xinhuamed.com.cn (S. Chen), yuyu@xinhuamed.com.cn (Y. Yu).
[1] Co-first authors

Research in context

Evidence before this study

We reviewed systematically the literature on the pathogenesis of TAPVC from 1960 to 2017 in PubMed using the search terms TAPVC and aetiology. We obtained over 80 articles, of which 14 were related to the molecular and cellular mechanisms of TAPVC. To date, only a few genes have been identified as candidate genes for TAPVC pathogenesis, and these genes are just a partial explanation for some patients.

Added value of this study

Since the pathogenesis of TAPVC remains elusive, we performed next-generation sequencing in 178 unrelated TAPVC cases and 100 healthy controls. To our knowledge, this study represents the largest series of NGS in TAPVC cases reported to date. Our research opens new avenues of investigation into TAPVC pathology and provides novel insights into pulmonary vein development.

Implications of all the available evidence

Without proper intervention in early life, TAPVC can lead to a high mortality rate of approximately 80%. Our findings demonstrate that as a rare congenital heart defect, TAPVC in some sporadic cases may be attributed to rare variants. Moreover, we identified novel candidate genes in several rare damage variants that have never been reported in connection to pulmonary vein development.

Li et al. analysed whole exome sequencing (WES) data from 6 sporadic TAPVC cases and 81 non-TAPVC counterparts, providing evidence for *ACVRL1* as a known causative gene and for *SGCD* as a candidate TAPVC gene [9]. Nash et al. used WGS analysis to identify a nonsynonymous variant in *RBP5* gene that was predicted to be deleterious and overrepresented in TAPVC [10]. Therefore, we applied WES technology to identify the likely rare damaging variants and putative candidate genes in 78 TAPVC cases and 100 non-TAPVC controls, and these variants were then validated using target sequencing (TS) in a replication cohort of 100 TAPVC patients. The expression patterns of candidate genes were analysed in human umbilical vein endothelial cells (HUVECs), mouse pulmonary veins, human pulmonary veins, human embryonic hearts, and zebrafish embryos. Finally, we identified 7 candidate genes, especially *SNAI1*, *HMGA2* and *VAV2*, that most likely underlie TAPVC pathogenesis. These results have improved our understanding of the diagnostic yield of next generation sequencing (NGS) for TAPVC, and assisted in the identification of candidate genes for TAPVC. To our knowledge, this is the largest series of NGS in TAPVC cases reported to date.

2. Materials and methods

2.1. Study population

Our discovery cohort included 78 unrelated TAPVC cases, 100 healthy controls and a validation cohort included 100 additional unrelated TAPVC cases. Patients with identified chromosomal or syndromic disorder or situs anomaly were excluded. Detailed cardiac and extra-cardiac features were assessed by reviewing medical records, imaging, and dysmorphology assessments. TAPVC cases were further divided into 4 subtype groups according to where the anomalous veins drain:

supra-cardiac, cardiac, infra-cardiac and mixed [11,12]All patients were recruited via Xin Hua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine. The study was conducted in accordance with the Declaration of Helsinki, and the protocol used to collect human samples of blood, pulmonary veins, and embryonic hearts was approved by the Ethics Committee of Xinhua Hospital. Written informed consents were also obtained from all subjects before the study protocol.

2.2. Whole-exome sequencing

DNA extraction from blood samples was carried out using the QIAamp™ DNA and Blood Mini kit (Qiagen) according to the manufacturer's instructions. Total DNA concentration and quantity was assessed by measuring the absorbance at 260 nm with a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific). WES was performed using the Agilent Sure Select Target Enrichment kit (V6 58 Mb; Agilent Technologies) for sequence capture and the Illumina HiSeq2500 for sequencing (Illumina) to a target depth of 100×.

We conducted bioinformatics analysis pipeline to call SNPs. Firstly, to exclude sequence artifacts from the FASTQ files. The cleaned FASTQ files were used for downstream analysis. Cleaned sequencing data was mapped to the reference human genome (UCSC hg19) by Burrows-Wheeler Aligner (BWA) software[13] to get the original mapping results stored in BAM format. Then, SAMtools[14], Picard (http://broadinstitute.github.io/picard/), and GATK[15] were used to sort BAM files and do duplicate marking, local realignment, and base quality recalibration to generate final BAM file for computation of the sequence coverage and depth. Samtools mpileup and bcftools were used to do variant calling and identify SNPs and InDels [16]. ANNOVAR was performed to do annotation for VCF (Variant Call Format) obtained in the previous effort[17]. We further filtered the SNPs using the American College of Medical Genetics (ACMG) criteria guidelines as follows:

i. Removed untranslated regions (UTRs) as well as synonymous and intronic variants.
ii. Extremely low allele frequency compared to that in the control (minor allele <0.005) and public databases (1000 Genomes Project, ExAC and ESP; minor allele <0.005).
iii. Variants were scored with SIFT, PolyPhen2 and Mutation Taster. Any variant predicted to be pathogenic in at least 1 program was deemed damaging.

2.3. TAPVC-associated SNPs detected by Fisher's exact test

For 78 TAPVC cases and 100 healthy controls, all 325,353 SNPs called in TAPVC were tested with Fisher's exact test. The SNP status was encoded as 0 or 1, where 0 indicated that no SNP alleles were found, and 1 indicated that at least 1 SNP allele was detected. The SNP status and sample class labels in which 1 indicated TAPVC patients and 0 indicated control samples were used to obtain the 2×2 confusion table for Fisher's exact test, and $P < 3e\text{-}5$ was considered statistically significant.

2.4. Gene-based burden test

Similarly, we aggregated the SNP data at the gene level. Genes with at least 1 rare mutant allele (damaging missense or loss-of-function variants) in the combined cohort, including 78 TAPVC patients and 100 healthy controls, were considered in the burden test and recorded as 1, and all other genes were recorded as 0. The SNP status of each gene was compared with the sample classes, and Fisher's exact test P values were calculated at the gene level. The gene level P values were adjusted with the FDR method, where FDR < 0.05 was considered statistically significant.

Table 1

Summary of the demographic and clinical information for 178 TAPVC patients.

Patient characteristics	Discovery cohort	Validation cohort
Mean age at diagnosis (years)	0.95 ± 1.87	1.88 ± 1.21
BMI (kg/m^2)	14.92 ± 3.05	19.87 ± 2.85
Male (n, %)	45(57.7)	54(54)
Mortality (n, %)	4(5.1)	7(7)
TAPVC Anatomical type (n, %)	is	
Supracardiac	43(55.1)	50(50)
Cardiac	21(26.9)	28(28)
Infracardiac	8(10.3)	12(12)
Mixed	6(7.7)	10(10)
Associated cardiac lesion (n, %)		
Vascular malformation	4(5.1)	9(9)
Valvular malformation	18(23.1)	22(22)
Compound malformation	6(7.7)	9(9)
Conduction block	7(9.0)	17(17)

2.5. Gene selection

We classified initial candidate genes into 3 types according to ACMG standards and guidelines [18]. First, significant variant genes were selected from TAPVC-associated SNPs using Fisher's exact test and the gene-based burden test. These genes were filtered as category I candidate genes. Category II genes included TAPVC pathogenic or likely pathogenic genes derived from the literature and publicly available databases. Category III genes included genes associated with human cardiac development, CHDs and vascular development in previous literature. Using the above screening process, we finally selected these 3 types of genes as our initial candidate genes.

2.6. Gene expression detection using the RT-qPCR assay

RNA was extracted from HUVECs and pulmonary vein samples from humans and mice. HUVECs were obtained from the Chinese Academy of Sciences (Shanghai, China). C57 mice were purchased from the Central Institute for Experimental Animals (Shanghai, China). Procedures involving animals and their care were conducted in accordance with National Institutes of Health (NIH) guidelines (NIH pub. no. 85–23, revised 1996) and approved by the Animal Care and Use Committee of Shanghai Xinhua Hospital. The human samples consisted of vertical veins from 5 patients and pulmonary veins from 5 controls harvested from Shanghai Xinhua Hospital. Total RNA extraction and the RT-qPCR were performed as described previously[19]. The RT-qPCR primers are listed in Table S1.
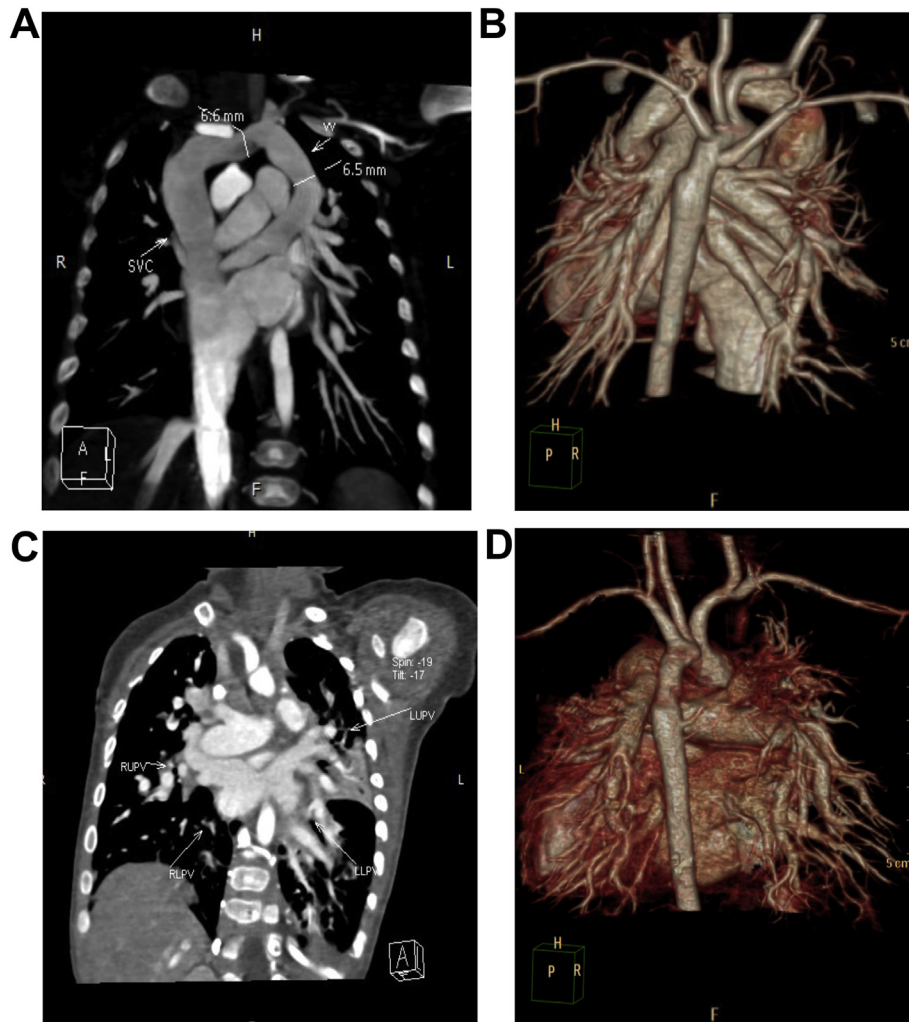

Fig. 1. Representative CT angiography graphs of supra-cardiac and cardiac TAPVC. CT angiography in a 2-month-old boy with supra-cardiac TAPVC; longitudinal section image (A) and volume-rendered image (B) showing 4 individual pulmonary veins joining in a retrocardiac venous confluence and draining into the superior vena cava (SVC) via a vertical vein (VV). CT angiography in a 6-month-old boy with cardiac TAPVC; longitudinal section image (C) and volume-rendered image (D) showing 4 individual pulmonary veins flowing into the CS. ASD: atrial septal defect; PV: pulmonary vein; SVC: superior vena cava; RLPV: right low pulmonary vein; RUPV: right upper pulmonary vein; LUPV: left upper pulmonary vein; LLPV: left low pulmonary vein; CS: coronary sinus; VV: vertical vein.

2.7. Expression patterns of the selected genes during human embryonic heart development

We collected human embryonic heart samples from Carnegie stages 11 through 15. TissueLyserII (Qiagen) and the RNeasy MinElute Cleanup Kit (Qiagen) were utilized for RNA extraction. The integrity and purity of the RNA was detected by the Experion automated gel electrophoresis system (Bio-Rad) and the NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific). The time course expression patterns of the selected TAPVC candidate genes were measured using an Affymetrix HTA 2.0 microarray.

2.8. SNP validation using targeted sequencing

We used multiplex PCR to amplify our target regions. >1 target sequence was amplified by >1 pair of primers in the reaction. Then, we sequenced the target regions of 7 candidate genes with Illumina MiSeq using DNA isolated from 100 additional TAPVC patients. We evaluated the assay performance using a 113-amplicon assay for the coding regions of the 7 candidate genes. All PCR primers were designed using OLIGO 6.2-assisted primer design, and all these primers were highly efficient and sensitive for detection [20]. The TS primers are listed in Table S2.

2.9. Network analysis

The Search Tool for the Retrieval of Interacting Genes (STRING) database (http://string-db.org) can critically assess and integrate protein–protein interactions (PPI), including both direct (physical) and indirect (functional) associations [21]. To detect potential relationships among our initial candidate genes, we mapped all the genes to the STRING network and visualized the network using Cytoscape. Since many genes did not directly interact, we used Dijkstra's algorithm [22] to discover the shortest paths between candidate genes. Elucidating the genes on the shortest path between candidate genes can reveal the possible genetic mechanism underlying TAPVC. The Dijkstra's algorithm procedure was
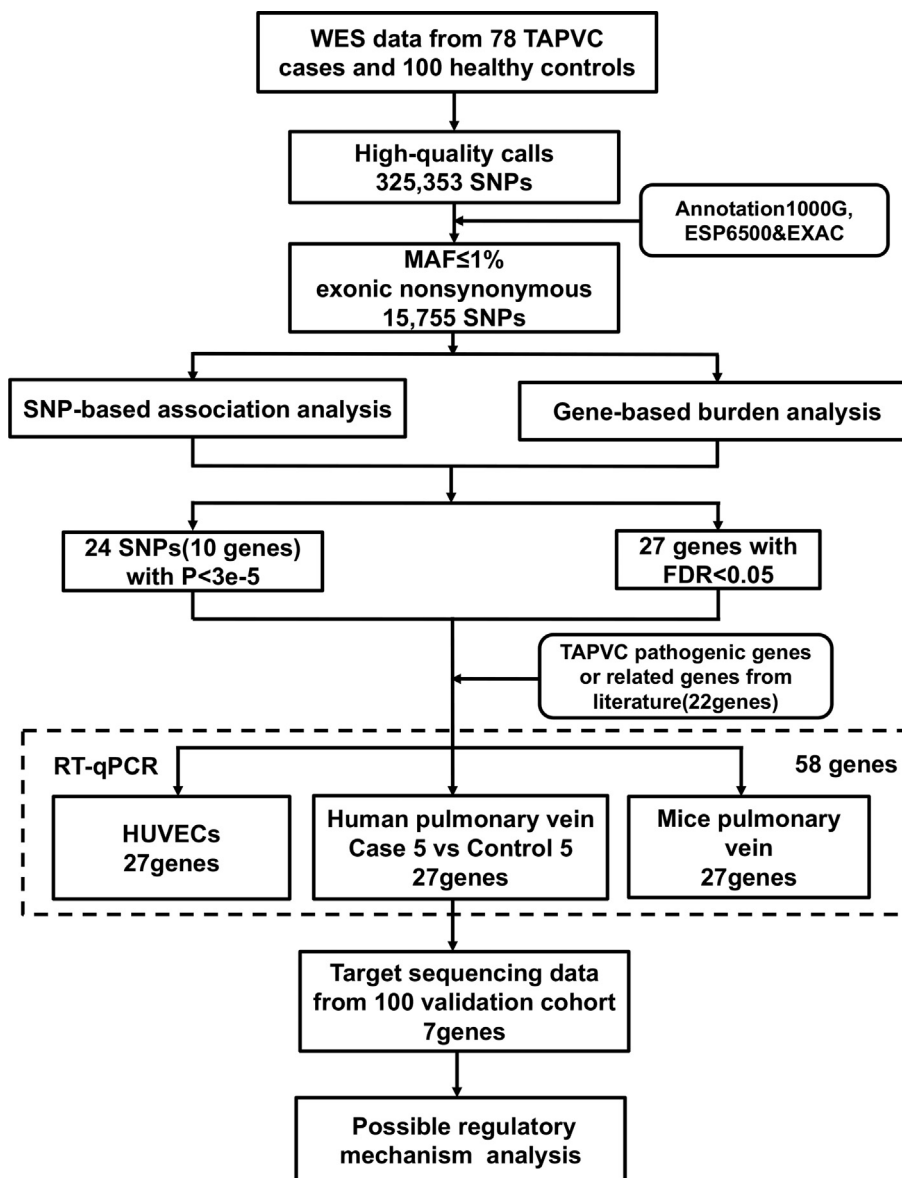

Fig. 2. The analytical strategy workflow for variant filtration. A schematic overview of the different steps taken during next-generation sequencing analysis with gene expression detection is shown. After variant calling and annotation, variants were filtered via SNP-associated analysis and the gene-based burden test. Candidate genes were collected by the detection of mRNA expression, validation of potential variants and network analysis. SNP, single nucleotide polymorphism; WES, whole-exome sequencing; MAF, minor allele frequency; FDR, false discovery rate; HUVECs, human umbilical vein endothelial cells; RT-qPCR, real-time quantitative polymerase chain reaction.

as follows [22–24]: let $G = (V,E,w)$ be a weighted graph; $V$is the set of vertices that includes the mapped TAPVC genes and other genes in the STRING network, E is the edges that correspond to the interactions in the STRING database, and w is the weight function, i.e., 1 minus the confidence score from the STRING database. If u_0 and v_0 are assumed to be 2 vertices in G, the shortest path between them can be discovered using the following procedures:

1) Let S = $\{u_0\}$, $\overline{S} = v - \{u_0\}$, $l(u_0) = 0$, and $l(v) = \infty$ for any vertex $v \in S - \{u_0\}$;
2) For each vertex, $v \in S'$such that $u'v \in E$, where $u' \in S$. If $l(v) \leq l(u') + w (u'v)$, then continue; otherwise, $l(v) = l(u') + w(u'v)$ and Parent(v) = u';
3) Find a vertex $v' \in \overline{S}$ such that $l(v') = \min \{l(v) \mid v \in S'\}$;
4) $S = S \cup \{v'\}$ and $\overline{S} = \overline{S} - \{v'\}$;
5) If $v_0 \in S$, then continue; otherwise, return to step 2;
6) The label *Parent* was used to find the shortest path from to.

2.10. Expression patterns of the candidate genes in zebrafish embryos at different stages

To confirm whether candidate genes were expressed during the development of zebrafish embryos, we designed specific probes for *SNAI1*, *HMGA2*, and *VAV2* to perform whole-mount in situ hybridization (WISH) [25]. Digoxigenin (DIG)–labelled RNA probes were transcribed from linearized DNA templates with T7 or SP6 RNA polymerases (Ambion). Embryos were hybridized with the DIG-labelled RNA probes overnight, and the concentration of the probes was 1 ng/ml. Then, embryos were incubated with anti-DIG antibody (Roche; 1: 8000), stained and observed under a fluorescent stereomicroscope (MZ FLIII, Leica).

3. Results

3.1. Descriptions of the study cohort

The discovery cohort consisted of 78 unrelated patients with TAPVC ($n = 78$, males: 45, females: 33, average age: 0.95 ± 1.87 years) and 100 healthy controls ($n = 100$, males: 52, females: 48, average age: 3.08 ± 0.86 years) collected from Shanghai Xinhua Hospital between June 2013 and May 2017. Another 100 TAPVC patients (n = 100, males: 54, females: 46, average age: 1.27 ± 0.82 years) were collected as the validation cohort (Table 1). As supra-cardiac and cardiac TAPVC comprised most of the 4 subtypes, representative graphs of the supra-cardiac and cardiac TAPVC patients are shown in Fig. 1.

3.2. TAPVC-associated SNPs detected by Fisher's exact test

In the discovery cohort, 325,353 SNPs were found in 78 samples using WES and BWA + GATK analyses. Meanwhile, 363,641 SNPs were found in the 100 control samples using the same analysis methods. To identify the candidate TAPVC pathogenic genes, we proposed an analytical strategy workflow (Fig. 2). Genomic information regarding the SNPs of the 78 patient samples was displayed using a Manhattan plot (Fig. S1). Only 15,755 rare nonsynonymous SNPs associated with TAPVC were tested by Fisher's exact-test. The 24 SNPs in 10 genes with $P < 3e\text{-}5$ are shown in Table S3.

3.3. TAPVC-associated gene-based burden test

As mentioned above, the significant SNPs seemed to locate in the same genes densely. Therefore, we tested the associations of genes with significant TAPVC variants using Fisher's exact-test by aggregating SNP data at the gene level. When we focused on the gene level, the TAPVC patients had significantly more genes with SNPs than controls, with a t test P value of 4.39e − 31. The number of genes with SNPs in the 100 controls was approximately 16,085 $\pm$ 160, while that in the

78 patients was approximately 16,356 $\pm$ 77. The boxplot of the numbers of genes with SNPs in the TAPVC patients and controls is shown in Fig. S2. In addition, the 27 genes that had an FDR < 0.05 based on the gene-based burden test are shown in Table 2.

3.4. TAPVC-associated gene selection

After screening the TAPVC-associated SNPs and genes from the WES data using Fisher's exact test, *SNAI1* was found in both gene lists. Therefore, 36 genes were ultimately defined as category I genes. We then further reviewed the literature systematically to select genes associated with TAPVC identified in previous studies and databases. We identified 12 pathogenic/likely pathogenic genes as being category II (*KDR, SMAD1, NRP1, PDGFRA, GJA1, ACVRL1, NKX2–5, ZIC3, SGCD, SEMA3D, ANKRD1, RBP5*) (Table S4). We also deemed 10 interesting genes related to vascular development in the literature as being category III genes (*ANGPT1, ANGPT2, ANKRD6, ATM, EGR1, ENG, FGF2, SMG1, TYMP, VEGFRA*); these genes were also related to human cardiac development and CHDs. Taken together, we considered the aforementioned 58 genes as initial TAPVC candidate genes and subjected them to further expression validation.

3.5. Detection of TAPVC-associated gene expression

We detected mRNA levels of the aforementioned 58 genes in HUVECs, mouse pulmonary veins and human pulmonary veins harvested from 5 TAPVC patients and 5 controls (Fig. 3A-C). Then, we further filtered these genes according to whether they were highly expressed in different tissues. Finally, 27 genes were selected from the 58 initial candidate genes, and 7 were from the list of TAPVC-associated genes. Of the 7 genes, *GXYLT1, HMGA2, SNAI1* and *VAV2* were significantly down-regulated in TAPVC patients ($P < .05$), while *CST3, CLTCL1, ZDHHC8* were significantly up-regulated ($P < .05$) compared with that in the control cohort. Another 20 genes perfectly

Table 2
TAPVC-associated genes detected using the gene-based burden test.

Gene symbol	Gene name	FDR	P value
CRISP1	Cysteine rich secretory protein 1	1.93E-27	8.94E-32
HOXD9	Homeobox D9	3.59E-16	3.33E-20
WDR54	WD repeat domain 54	2.46E-14	3.41E-18
VAV2	Vav guanine nucleotide exchange factor 2	2.04E-11	3.77E-15
MARCH9	Membrane associated ring-CH-type finger 9	1.53E-09	4.24E-13
CA14	Carbonic anhydrase 14	1.75E-09	5.66E-13
ZDHHC8	Zinc finger DHHC-type containing 8	2.03E-09	7.53E-13
SDHAP3	SDHA C-Terminal Like	3.73E-08	1.90E-11
ASB9	Ankyrin repeat and SOCS box containing 9	9.71E-08	5.39E-11
MNT	MAX network transcriptional repressor	6.78E-06	4.52E-09
SNAI1	Snail family transcriptional repressor 1	6.78E-06	4.70E-09
CLTCL1	Clathrin heavy chain like 1	1.31E-05	1.03E-08
CST3	Cystatin C	0.000100455	9.76E-08
RHCG	Rh family C glycoprotein	0.000254069	2.82E-07
SERF1A	Small EDRK-rich factor 1A	0.001780581	2.72E-06
SERF1B	Small EDRK-rich factor 1B	0.001780581	2.72E-06
RAC3	Ras-related C3 botulinum toxin substrate 3	0.00197802	3.11E-06
ANKRD7	Ankyrin repeat domain 7	0.005872844	9.78E-06
WWTR1	WW domain containing transcription regulator 1	0.006274163	1.07E-05
ESX1	ESX homeobox 1	0.006567915	1.18E-05
RGPD5	RANBP2-like and GRIP domain containing 5	0.013392925	2.73E-05
RGPD6	RANBP2-like and GRIP domain containing 6	0.013392925	2.73E-05
NUS1	Nogo-B Receptor	0.017510764	3.64E-05
NME5	NME/NM23 family member 5	0.01776352	3.86E-05
ABCG5	ATP binding cassette subfamily G member 5	0.017879142	3.97E-05
HMGA2	High mobility group AT-hook 2	0.019242641	4.36E-05
ARX	Aristaless related homeobox	0.049588497	0.0001376

satisfied our criteria for being pathogenic/likely pathogenic genes or related genes, and their expression patterns are shown in Fig. S3.

We then studied the temporal expression patterns of the selected TAPVC genes during embryonic heart development using an Affymetrix HTA 2.0 microarray (Fig. 3D). The expression levels of *ANKRD1*, *KDR*, *SEMA3D* in embryonic hearts were higher than other known TAPVC pathogenic genes. Comparing these known TAPVC pathogenic genes, all 7 candidate genes had relatively high expression, especially

GXYLT1. Among the pathogenic or likely pathogenic genes, *ANKRD1* was significantly up-regulated ($P < .05$), while *GJA1*, *KDR*, *SEMA3D* were significantly down-regulated ($P < .05$); these patterns were identical to those reported in the literature [6,8,10,26,27]. Altered expression was also detected in some related genes, such as *ANGPT1* and *TYMP* ($P < .05$), but well qualified SNPs were not found in our WES data. In conclusion, we considered *CLTCL1*, *CST3*, *GXYLT1*, *HMGA2*, *SNAI1*, *VAV2*, and *ZDHHC8* as our candidate genes.
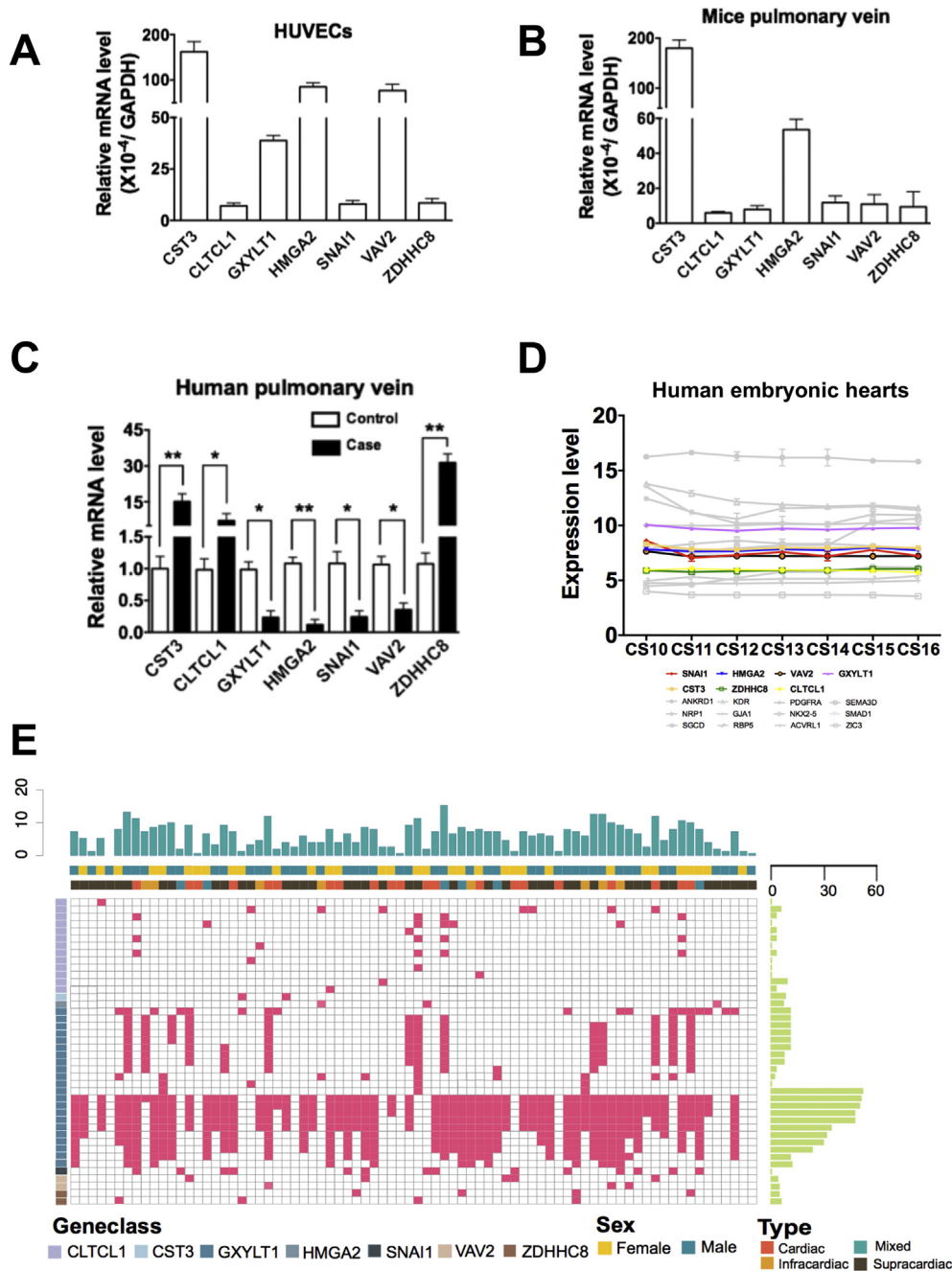

Fig. 3. mRNA expression levels of the 7 candidate genes (*CST3*, *CLTCL1*, *GXYLT1*, *HMGA2*, *SNAI1*, *VAV2*, *ZDHHC8*) in different samples (A) mRNA expression levels of the 7 candidate genes in HUVECs; $n = 3$. (B) mRNA abundance of the 7 candidate genes in mouse pulmonary veins; n = 3. (C) mRNA differential expression in human pulmonary veins between TAPVC patient and control samples; $n = 5$. Statistical significance was calculated by Student's *t*-test, where *$P < .05$ and **$P < .01$. (D) Expression patterns of the candidate genes in human embryonic hearts at different time points. The coloured lines represent the expression patterns of the candidate genes, while the grey lines represent known pathogenic genes. CS, Carnegie stage. (E) Heat map showing 42 rare nonsynonymous variants in 7 candidate genes from 78 TAPVC cases. Significantly mutated genes are listed on the left. The percentage of each gene with the variants detected is noted on the right. The upper histogram shows the variant rates in each of the 78 patients. Samples are divided into 4 subtypes (supracardiac, cardiac, infracardiac and mixed); gender information for the patients is also shown on the heat map. The red box shows the patients that had the variants. The middle region of the heat map shows details regarding variants in the sequenced samples.

3.6. Rare variants of 7 candidate genes in the TAPVC discovery cohort

By assessing gene expression in different samples, we finally obtained 7 candidate genes. We found 42 rare damage-associated nonsynonymous variants in our WES data (Fig. 3E). GXYLT1 variants (22/42 52.3%) and CLTCL1 variants (13/42 30.9%) accounted for a substantial proportion of variants of the 7 candidate genes, while VAV2 (p. R816C and p.D532G) and ZDHHC8 (p.T184A and p.R540C) had 2 rare nonsynonymous variants, and HMGA2 (p.S105I), SNAI1 (p.R224P) and CST3 (p.R79S) had only 1 rare nonsynonymous variant (Table 3). To elucidate the inner relationship between these SNPs, we performed linkage disequilibrium (LD block) analysis of the 42 rare damage-associated nonsynonymous SNPs from the 7 candidate genes (Fig. S4). LD block analyses were performed for the chromosomal regions with multiple significant SNPs clustered around genome-wide significant SNPs. The LD blocks were defined using Haploview (version 4.2) and criteria established by Gabriel et al. [28]. LD block analysis explained why the rare GXYLT1 and CLTCL1 variants accounted for such a large proportion of variants of the 7 candidate genes, as many SNPs of the same block indicated that 1 SNP mutation can lead to other SNP mutations, and they may have similar effects on TAPVC incidence.

3.7. TAPVC-associated variants were confirmed by TS

Additionally, we performed TS by multiplex PCR in the additional 100 cases to validate the rare damage-associated variants of the 7 candidate genes. We found 42 rare damage-associated nonsynonymous variants in the WES data, and many significant SNPs were in genes, such as GXYLT1 and CLTCL1, while 56 rare damage-associated nonsynonymous variants were identified from the TS data (Table S5). Moreover, CLTCL1, SNAI1, ZDHHC8, HMGA2, and VAV2 showed the same variants in both the WES data and the TS data, while the CST3 and GXYLT1 variants differed in the 2 datasets. A heat map was used to illustrate the TS rare variant results (Fig. S5).

3.8. Regulatory network of the TAPVC candidate genes

We used Dijkstra's algorithm to find the shortest path to explore the protein interaction networks of candidate genes. Since the CLTCL1 showed no direct interaction with other genes in the STRING database, we showed 6 candidate genes without CLTCL1. We plotted these shortest paths between the different gene sets using Cytoscape. Gene regulatory networks revealed that our candidate genes not only interacted with each other but also closely associated with CHD-related genes and known pathogenic genes (Fig. 4). SNAI1, VAV2 and HMGA2 were at the centre of the regulatory network, suggesting that these three genes may play more important roles in TAPVC pathogenesis.

3.9. The expression pattern of SNAI1, HMGA2, and VAV2 in zebrafish embryos

Then, we investigated whether SNAI1, HMGA2, and VAV2 could affect the development of embryonic vessels and hearts. We analysed the

Table 3
Rare variants of 7 candidate genes associated with TAPVC.

Chromosome	Position	Gene	Base change	Amino acid change	avsnp142	1000G	ExAC	ESP6500
chr9	136,633,707	VAV2	G > A	p.R816C	rs191274326	0.0006	0.0002	0.0002
chr9	136,649,478	VAV2	T > C	p.D532G	rs191239028	0.0008	0.0001	.
chr12	42,499,690	GXYLT1	T > C	p.Y265C	rs79044728	.	.	.
chr12	42,499,694	GXYLT1	A > T	p.Y264N	.	.	.	.
chr12	42,499,701	GXYLT1	C > A	p.R261S	rs74583427	.	.	.
chr12	42,499,711	GXYLT1	C > A	p.R258L	rs76555438	.	.	.
chr12	42,499,714	GXYLT1	T > C	p.N257S	rs78536827	.	.	.
chr12	42,499,738	GXYLT1	T > C	p.E249G	.	.	.	.
chr12	42,499,739	GXYLT1	C > T	p.E249K	rs77582546	.	.	.
chr12	42,499,763	GXYLT1	T > C	p.I241V	rs80202058	.	0.0002	.
chr12	42,499,802	GXYLT1	C > A	p.D228Y	rs78540738	.	.	.
chr12	42,499,825	GXYLT1	A > C	p.I220S	rs76034661	.	.	.
chr12	42,499,826	GXYLT1	T > C	p.I220V	rs75241273	.	.	.
chr12	42,499,857	GXYLT1	T > A	p.E209D	.	.	.	.
chr12	42,538,334	GXYLT1	C > A	p.G39 W	rs181558534	.	0.0001	.
chr12	42,538,340	GXYLT1	C > A	p.G37C	.	.	.	.
chr12	42,538,349	GXYLT1	T > C	p.T34A	.	.	.	.
chr12	42,538,366	GXYLT1	A > T	p.V28E	.	.	.	.
chr12	42,538,367	GXYLT1	C > T	p.V28 M	.	.	.	.
chr12	42,538,406	GXYLT1	C > T	p.G15S	.	.	.	.
chr12	42,538,412	GXYLT1	C > T	p.A13T	.	.	.	.
chr12	42,538,415	GXYLT1	C > G	p.V12 L	.	.	.	.
chr12	42,538,423	GXYLT1	A > G	p.V9A	.	.	.	.
chr12	42,538,435	GXYLT1	A > G	p.L5P	.	.	.	.
chr12	66,308,093	HMGA2	G > T	p.S105I	.	.	.	.
chr20	23,618,265	CST3	G > T	p.R79S	rs574152075	0.0014	0.0003	.
chr20	48,604,469	SNAI1	G > C	p.R224P	.	.	.	.
chr22	20,127,408	ZDHHC8	A > G	p.T184A	rs200408305	0.0022	0.0005	.
chr22	20,130,771	ZDHHC8	C > T	p.R540C	.	.	.	.
chr22	19,168,250	CLTCL1	C > T	p.V1633 M	.	.	0.0001	.
chr22	19,170,956	CLTCL1	C > T	p.V1592 M	rs2073738	0.032	0.017	0.0057
chr22	19,188,928	CLTCL1	C > T	p.R1226H	.	.	.	.
chr22	19,196,497	CLTCL1	T > C	p.N1126S	.	.	.	.
chr22	19,207,480	CLTCL1	G > A	p.R945C	.	.	.	.
chr22	19,209,603	CLTCL1	C > T	p.R811Q	rs112636081	0.004	0.0037	.
chr22	19,209,604	CLTCL1	G > A	p.R811W	rs12628529	0.0002	0.0001	0.0003
chr22	19,212,999	CLTCL1	A > T	p.F702Y	rs182917063	0.002	0.0006	.
chr22	19,213,059	CLTCL1	C > T	p.C682Y	.	.	.	.
chr22	19,230,318	CLTCL1	G > A	p.R221C	rs376289636	0.0004	0.0001	.
chr22	19,241,585	CLTCL1	A > C	p.M139R	rs187925949	0.0002	.	.
chr22	19,241,684	CLTCL1	T > C	p.E106G	.	.	.	.
chr22	19,279,131	CLTCL1	G > T	p.H12N	.	.	.	.

expression pattern of these 3 genes by WISH (Fig. 5). *SNAI1* expression was observed in the primitive veins and entire head at 20 hpf (hours post fertilization), and sustained expression was found in the posterior cardinal vein (PCV) at 24 hpf and 30 hpf. At 48 hpf and 72 hpf, the expression of *SNAI1* was mainly observed in the heart. *HMGA2* expression patterns mirrored those of *SNAI1*. At 48 hpf and 72 hpf, *VAV2* was expressed mainly in the heart. These findings indicated that these 3

genes are important for cardiovascular development and might be the most crucial candidate genes for TAPVC.

4. Discussion

Because knowledge regarding the aetiology of TAPVC is lacking, genetically characterizing TAPVC has been challenging. To investigate the
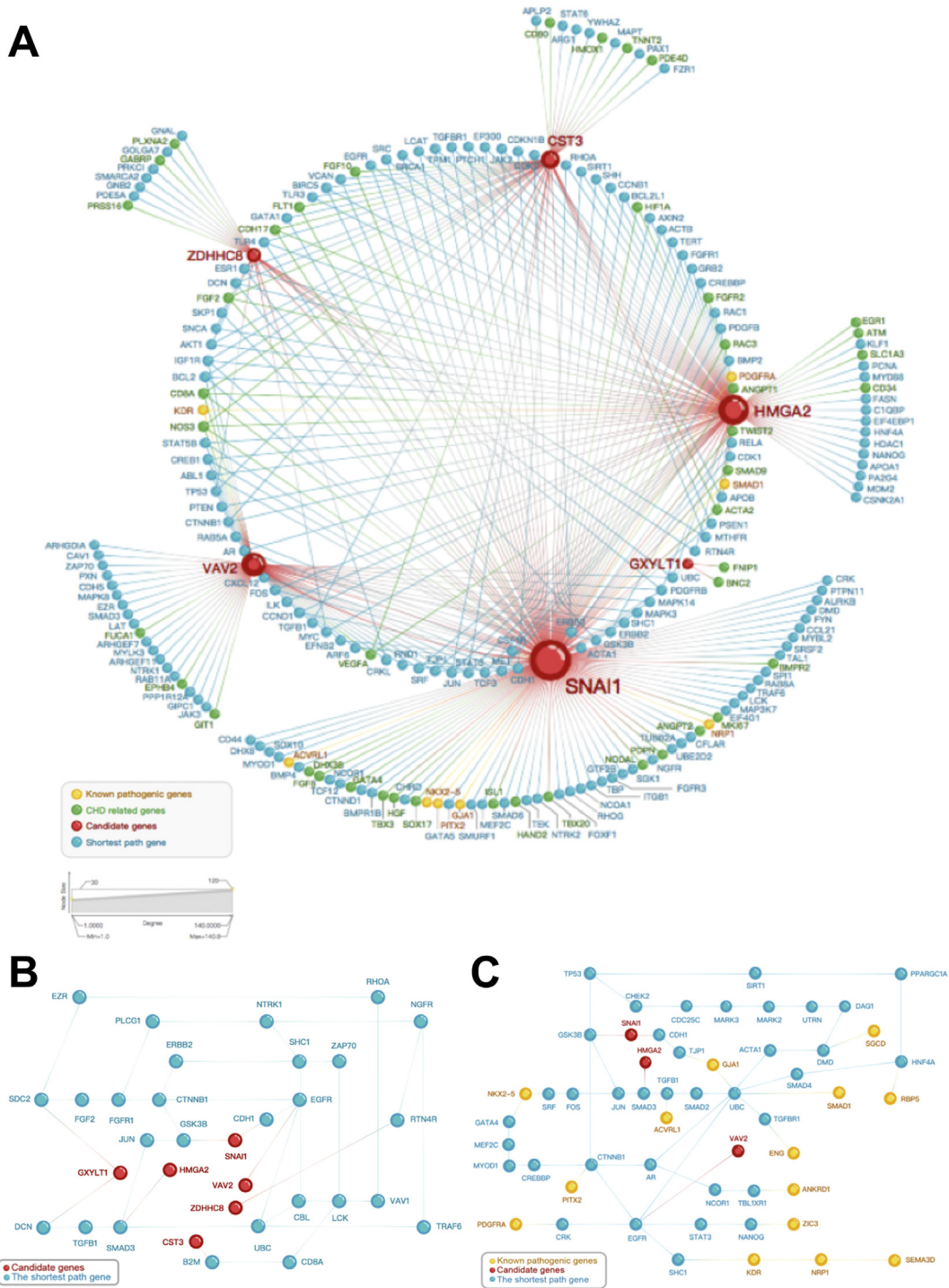
Fig. 4. Gene regulatory networks and sub-networks showing the importance of highly connected genes. The network was modelled in Cytoscape using annotations from the STRING 9.1 protein-protein interactions database. We used Dijkstra's algorithm to discover the shortest paths between these genes. (A) Network analysis indicated the protein-protein interactions among 6 candidate genes, known pathogenic genes and CHD-related genes. (B) Network analysis of internal correlations among 6 candidate genes. (C) Network analysis of relationships between SNAI1, HMGA2, VAV2 and known TAPVC pathogenic genes. The red nodes represent candidate proteins, the yellow nodes represent known pathogenic genes, the green nodes represent CHD-related genes, and the blue nodes represent the shortest path proteins that connect these genes.

genetic pathogenesis of TAPVC in Chinese population, we conducted gene analysis using next-generation sequencing in a cohort of 178 sporadic TAPVC patients and 100 healthy controls. The major findings of our study can be summarized as follows. Seven totally novel candidate genes (*CLTCL1*, *CST3*, *GXYLT1*, *HMGA2*, *SNAI1*, *VAV2*, *ZDHHC8*) were associated with TAPVC pathogenesis. STRING network analysis demonstrated that *SNAI1*, *HMGA2*, and *VAV2*, which are highly related to vascular development, appear to play an important role in the genetic mechanism of TAPVC. They were also required for cardiovascular development in zebrafish embryos. Additionally, we reviewed systematically the literature to select TAPVC-related genes identified in previous studies and databases. 12 pathogenic/likely pathogenic genes were identified (Table S4). We found some rare nonsynonymous variants among these genes, but several known pathogenic variants were not detected in our data. Reasons for this discrepancy included that previous studies focused mainly on Caucasians rather than on Chinese individuals, that the sample size of our study was larger than previous reports, or that the models and analysis methods utilized were different. Among the pathogenic or likely pathogenic genes, ANKRD1 in pulmonary vein samples up-regulated significantly, while GJA1, KDR, SEMA3D down-regulated significantly detected by RT-qPCR between case and control groups. These patterns were identical to those reported in the literature.

SNAI1 encodes a protein critical for mesoderm formation in the developing embryo [29]. Approximately 18% (14/78) of the discovery cohort patients had the same variant (p.R224P), a novel variant that is reported here for the first time. The incidence of this variant was 17% (17/100) in the validation cohort. *SNAI1* is involved in the epithelial to mesenchymal transition (EMT) as well as in the formation and maintenance of the embryonic mesoderm [30,31]. *SNAI1* also modulates the proliferation, apoptosis, angiogenesis of most tumour types by associating with VEGF, FGF18 and CDH1 [32,33]. Endothelial cell-specific *SNAI1* loss-of-function conditional knockout mice showed an early embryonic lethal phenotype with noticeable defects in vascular remodelling, morphogenesis and arterial-vein specification [34].

SNAI1 is regulated by *HMGA2* during the induction of EMT with Smads [35]. In the WES data, the same rare variant (p.S105I) in *HMGA2*, which critically functions in cardiogenesis and is essential for normal cardiac development, was detected in 3 patients [36]. p.S105I is a totally novel variant that has never been reported. Furthermore, 8 patients were determined to have the same variant in the TS data. *HMGA2* encodes a protein with structural DNA-binding domains that acts as a transcriptional regulating factor, and 12q14 microdeletion cases including *HMGA2* are reportedly associated with clinical CHD symptoms (including 2 patients with atrial septal defects and 1 patient
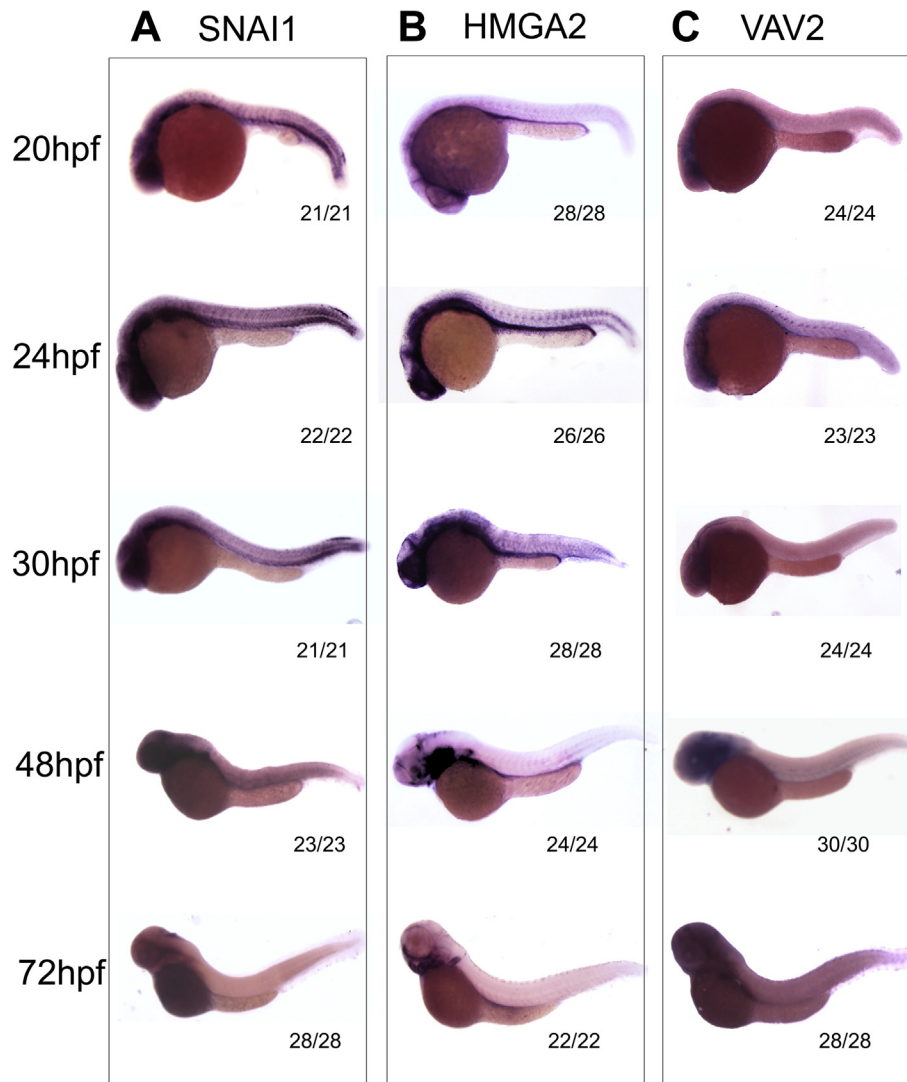

Fig. 5. Expression patterns of *SNAI1*, *HMGA2*, and *VAV2* during zebrafish embryonic development. WISH results demonstrating the expression of *SNAI1* (A), *HMGA2* (B), and *VAV2* (C) in wild-type zebrafish at the different stages (20 hpf, 24 hpf, 30 hpf, 48 hpf, and 72 hpf). The numbers shown in the bottom corner indicate the number of embryos with similar staining patterns out of the total number of embryos examined. All images are lateral views, rostral to the left.

each with pulmonary stenosis, a sub-aortic stenosis and a patent ductus) [37]. *HMGA2* highly expressed during embryogenesis and has been linked to vascular tumours, including angiomyxomas and pulmonary hamartomas, but the relationship between *HMGA2* and vascular development remains unknown [38].

VAV2, the second member of the VAV guanine nucleotide exchange factor family of oncogenes, is related to epidermal growth factor receptor binding and angiogenesis (OMIM 600428). *VAV2* and *VAV3* reportedly work together in neurons and endothelial cells to ensure proper axon guidance and angiogenic responses [39]. The *VAV2-Rac1* pathway is important for vasodilatation responses in vascular smooth muscle cells (VSMCs) [40]. Two rare variants (p.R816C, rs191274326 and p. D532G, rs191239028) were identified in 8 TAPVC individuals, 1 of which (p.D532G) was confirmed by TS in the additional patients. Although the 2 variants were previously reported in databases, their functions have never been studied.

In our network analysis, *SNAI1*, which had the highest weight, was located at the centre of the PPI. *SNAI1*, *HMGA2*, and *VAV2* had closer relationships with known TAPVC pathogenic or likely pathogenic genes than the other candidate genes. Interestingly, our zebrafish models showed that *SNAI1* and *HMGA2* had a similar expression pattern during the development of zebrafish embryos. They both were highly expressed in primitive veins, PCV, and hearts than in other organs. Although *VAV2* did not exhibit the same expression pattern, it was highly expressed in heart and blood vessels in the head. These findings indicate that SNAI, *HMGA2*, and *VAV2* might have an important impact on the development of zebrafish embryo hearts and vessels. However, there is a lack of comprehensive research demonstrating the role of these 3 genes in the foetal cardiovascular development of zebrafish embryos.

GXYLT1, which encodes enzymes that transfer xylose to the O-Glc residue bound to Notch EGF repeats in vitro and in vivo (OMIM 613321), had the most variants in our WES data. Nearly 77% (66/78) of the TAPVC patients exhibited 22 different rare damage-associated variants in *GXYLT1*. However, none of these variants were found in follow-up sequencing, while 3 different variants were observed in the 100 validation patients at follow-up. LD block analysis showed that most of these SNPs were assigned to the same block and may have the same function, which could also explain this phenomenon.

Only 2 patients had 1 novel variant (p.R79S) in *CST3*, but the variant was highly predicted to be damaging. *CST3* has been implicated as a prognostic marker in cardiovascular disease [41] and may remodel and establish the arterial wall, and *CST3* deficiency occurs in vascular disease [42]. Because we did not find the same variants in deeper and larger sequences, whether *CST3* variants play a causal or modifier role in TAPVC is still unclear.

Seven TAPVC patients (7/78 8.9%) had 2 rare nonsynonymous variants in *ZDHHC8*. The p.R816C variant was highly conserved, and p.D532G has not been previously reported in public datasets or predicted to be damaging. Seventeen patients (17/78 21.8%) were found to have 13 rare damage-associated variants of *CLTCL1*, which, like *ZDHHC8*, is located on 22q11.2, and both genes are associated with DiGeorge syndrome (DGS). The deletion of 22q11.2 may cause conotruncal heart defects, such as Tetrology of Fallot (TOF), double outlet of right ventricle (DORV) and pulmonary atresia/ventricular septal defect (PA/VSD) [43] [44]. Thus far, the functions of *ZDHHC8* and *CLTCL1* in cardiovascular development remain unknown, and they be might newly associated with TAPVC pathogenesis.

Our study did have some limitations. First, the lack of parental samples limited our ability to study the genetic backgrounds of the variants. In addition, the functions of our candidate genes need to be further verified with fundamental research. In summary, an effective analytical bioinformatics strategy allowed us to identify rare damage variants in novel genes that play a vital role in TAPVC pathology. Our candidate genes (*CLTCL1*, *CST3*, *GXYLT1*, *HMGA2*, *SNAI1*, *VAV2*, *ZDHHC8*) open new fields of investigation into TAPVC pathology and provide novel insights into pulmonary vein development.

Declaration of interests

We declare that no conflicts of interest are present.

Author contributions

YY, SC and KS conceived and designed the project and are responsible for the overall content. XS, TH, and CG performed the bioinformatics analysis of WES and TS data. XS, JW and YLL carried out all the experiments. JS, YJX, FL, QHF and YNL collected the clinical data. XS and TH prepared the manuscript. YY, SC and KS contributed to revising the manuscript. All authors have seen and approved the final manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2018.11.008.

References

[1] Hoffman JI, Kaplan S. The incidence of congenital heart disease. J Am Coll Cardiol 2002;39(12):1890–900.

[2] Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, et al. Heart disease and stroke statistics–2012 update: a report from the American Heart Association. Circulation 2012;125(1):e2-220.

[3] Delisle G, Ando M, Calder AL, Zuberbuhler JR, Rochenmacher S, Alday LE, et al. Total anomalous pulmonary venous connection: Report of 93 autopsied cases with emphasis on diagnostic and surgical considerations. Am Heart J 1976;91(1):99–122.

[4] Bjornard K, Riehle-Colarusso T, Gilboa SM, Correa A. Patterns in the prevalence of congenital heart defects, metropolitan Atlanta, 1978 to 2005. Birth Defects Res A Clin Mol Teratol 2013;97(2):87–94.

[5] Burroughs JT, Edwards JE. Total anomalous pulmonary venous connection. Am Heart J 1960;59:913–31.

[6] Bleyl SB, Botto LD, Carey JC, Young LT, Bamshad MJ, Leppert MF, et al. Analysis of a Scottish founder effect narrows the TAPVR-1 gene interval to chromosome 4q12. Am J Med Genet A 2006;140(21):2368–73.

[7] Cinquetti R, Badi I, Campione M, Bortoletto E, Chiesa G, Parolini C, et al. Transcriptional deregulation and a missense mutation define ANKRD1 as a candidate gene for total anomalous pulmonary venous return. Hum Mutat 2008;29(4):468–74.

[8] Degenhardt K, Singh MK, Aghajanian H, Massera D, Wang Q, Li J, et al. Semaphorin 3d signaling defects are associated with anomalous pulmonary venous connections. Nat Med 2013;19(6):760–5.

[9] Li J, Yang S, Pu Z, Dai J, Jiang T, Du F, et al. Whole-exome sequencing identifies SGCD and ACVRL1 mutations associated with total anomalous pulmonary venous return (TAPVR) in Chinese population. Oncotarget 2017;8(17):27812–27,819.

[10] Nash D, Arrington CB, Kennedy BJ, Yandell M, Wu W, Zhang W, et al. Shared Segment Analysis and Next-Generation Sequencing Implicates the Retinoic Acid Signaling Pathway in Total Anomalous Pulmonary Venous Return (TAPVR). PLoS One 2015;10(6):e0131514.

[11] Nurkalem Z, Gorgulu S, Eren M, Bilal MS. Total anomalous pulmonary venous return in the fourth decade. Int J Cardiol 2006;113(1):124–6.

[12] Vyas HV, Greenberg SB, Krishnamurthy R. MR imaging and CT evaluation of congenital pulmonary vein abnormalities in neonates and infants. Radiographics 2012;32(1):87–98.

[13] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25(14):1754–60.

[14] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. Genome Project Data Processing S: The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25(16):2078–9.

[15] DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M et al.: A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011, 43(5):491–498.

[16] Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SRF, Consortium WGS, et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat Genet 2014;46(8):912–8.

[17] Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat Protoc 2015;10(10):1556–66.

[18] Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. 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 2015;17(5):405–24.

[19] Nolan T, Hands RE, Bustin SA. Quantification of mRNA using real-time RT-PCR. Nat Protoc 2006;1(3):1559–82.

[20] Henegariu O, Heerema NA, Dlouhy SR, Vance GH, Vogt PH. Multiplex PCR: critical parameters and step-by-step protocol. Biotechniques 1997;23(3):504–11.

[21] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015;43(Database issue):D447–52.

[22] Dijkstra EW. A note on two problems in connexion with graphs. Numerische Mathematik 1959;1:269–71.

[23] Chartrand G, Oellermann OR. Applied and Algorithmic Graph Theory. Mcgraw-Hill College; 1992.

[24] Cormen TH, Leiserson CE, Stein C. Introduction to Algorithms. second ed. MIT press and Mcgraw-Hill; 2001.

[25] Thisse C, Thisse B. High-resolution in situ hybridization to whole-mount zebrafish embryos. Nat Protoc 2008;3(1):59–69.

[26] Bleyl SB, Saijoh Y, Bax NA, Gittenberger-de Groot AC, Wisse LJ, Chapman SC, et al. Dysregulation of the PDGFRA gene causes inflow tract anomalies including TAPVR: integrating evidence from human genetics and model organisms. Hum Mol Genet 2010;19(7):1286–301.

[27] Fahed AC, Gelb BD, Seidman JG, Seidman CE. Genetics of congenital heart disease: the glass half empty. Circ Res 2013;112(4):707–20.

[28] Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21(2):263–5.

[29] Nieto MA. The snail superfamily of zinc-finger transcription factors. Nat Rev Mol Cell Biol 2002;3(3):155–66.

[30] Luna-Zurita L, Prados B, Grego-Bessa J, Luxan G, del Monte G, Benguria A, et al. Integration of a Notch-dependent mesenchymal gene program and Bmp2-driven cell invasiveness regulates murine cardiac valve formation. J Clin Invest 2010;120(10):3493–507.

[31] Bai Y, Wang J, Morikawa Y, Bonilla-Claudio M, Klysik E, Martin JF. Bmp signaling represses Vegfa to promote outflow tract cushion development. Development 2013;140(16):3395–402.

[32] Timmerman LA, Grego-Bessa J, Raya A, Bertran E, Perez-Pomares JM, Diez J, et al. Notch promotes epithelial-mesenchymal transition during cardiac development and oncogenic transformation. Genes Dev 2004;18(1):99–115.

[33] Deep G, Jain AK, Ramteke A, Ting H, Vijendra KC, Gangar SC, et al. SNAI1 is critical for the aggressiveness of prostate cancer cells with low E-cadherin. Mol Cancer 2014;13:37.

[34] Wu ZQ, Rowe RG, Lim KC, Lin Y, Willis A, Tang Y, et al. A Snail1/Notch1 signaling axis controls embryonic vascular development. Nat Commun 2014;5:3998.

[35] Thuault S, Tan EJ, Peinado H, Cano A, Heldin CH, Moustakas A. HMGA2 and Smads co-regulate SNAIL1 expression during induction of epithelial-to-mesenchymal transition. J Biol Chem 2008;283(48):33437–33,446.

[36] Monzen K, Ito Y, Naito AT, Kasai H, Hiroi Y, Hayashi D, Shiojima I, Yamazaki T, Miyazono K, Asashima M et al.: A crucial role of a high mobility group protein HMGA2 in cardiogenesis. Nat Cell Biol 2008, 10(5):567–574.

[37] Lynch SA, Foulds N, Thuresson AC, Collins AL, Anneren G, Hedberg BO, et al. The 12q14 microdeletion syndrome: six new cases confirming the role of HMGA2 in growth. Eur J Hum Genet 2011;19(5):534–9.

[38] Vasan RS, Glazer NL, Felix JF, Lieb W, Wild PS, Felix SB, et al. Genetic variants associated with cardiac structure and function: a meta-analysis and replication of genome-wide association data. JAMA 2009;302(2):168–78.

[39] Cowan CW, Shao YR, Sahin M, Shamah SM, Lin MZ, Greer PL, et al. Vav family GEFs link activated Ephs to endocytosis and axon guidance. Neuron 2005;46(2):205–17.

[40] Fabbiano S, Menacho-Marquez M, Sevilla MA, Albarran-Juarez J, Zheng Y, Offermanns S, et al. Genetic dissection of the vav2-rac1 signaling axis in vascular smooth muscle cells. Mol Cell Biol 2014;34(24):4404–19.

[41] Lassus J, Harjola VP, Sund R, Siirila-Waris K, Melin J, Peuhkurinen K, et al. Nieminen MS, group F-AS: Prognostic value of cystatin C in acute heart failure in relation to other markers of renal function and NT-proBNP. Eur Heart J 2007;28(15):1841–7.

[42] Patel PC, Ayers CR, Murphy SA, Peshock R, Khera A, de Lemos JA, et al. Association of cystatin C with left ventricular structure and function: the Dallas Heart Study. Circ Heart Fail 2009;2(2):98–104.

[43] Xu YJ, Wang J, Xu R, Zhao PJ, Wang XK, Sun HJ, et al. Detecting 22q11.2 deletion in Chinese children with conotruncal heart defects and single nucleotide polymorphisms in the haploid TBX1 locus. BMC Med Genet 2011;12:169.

[44] Holmes SE, Riazi MA, Gong W, McDermid HE, Sellinger BT, Hua A, et al. Disruption of the clathrin heavy chain-like gene (CLTCL) associated with features of DGS/VCFS: a balanced (21;22)(p12;q11) translocation. Hum Mol Genet 1997;6(3):357–67.