Journal of the American Heart Association

ORIGINAL RESEARCH

Deep Phenotypic Analysis for Transposition of the Great Arteries and Prognosis Implication

Huayan Shen, BS*; Qiyu He, MD, BS*;Xinyang Shao, BS; Shoujun Li, MD; Zhou Zhou, MD, PhD

BACKGROUND: Transposition of the great arteries (TGA) consists of about 3% of all congenital heart diseases and 20% of cyanotic congenital heart diseases. It is always accompanied by a series of other cardiac malformations that affect the surgical intervention strategy as well as prognosis. In this study, we comprehensively analyzed the phenotypes of the patients who had TGA with concordant atrioventricular and discordant ventriculoarterial connections and explored their association with prognosis.

METHODS AND RESULTS: We retrospectively reviewed 666 patients with a diagnosis of TGA with concordant atrioventricular and discordant ventriculoarterial connections in Fuwai Hospital from 1997 to 2019. Under the guidance of the Human Phenotype Ontology database, patients were classified into 3 clusters. The Kaplan-Meier method was used to analyze the prognosis, and the Cox proportional regression model was used to investigate the risk factors. In this 666-patient TGA cohort, the overall 5-year survival rate was 94.70% (92.95%–96.49%). Three clusters with distinct phenotypes were obtained by the Human Phenotype Ontology database. Kaplan-Meier analysis revealed a significant difference in freedom from reintervention among 3 clusters (P<0.001). To eliminate the effect of surgeries, we analyzed patients who only received an arterial switch operation and still found a significant difference in reintervention (P=0.019).

CONCLUSIONS: We delineated a big cardiovascular phenotypic profile of an unprecedentedly large TGA cohort and successfully risk stratified them to reveal prognostic significance. Also, we reported the outcomes of a large TGA population in China.

Key Words: human phenotype ontology ■ prognosis ■ risk stratification ■ surgery ■ transposition of the great arteries

Transposition of the great arteries (TGA) accounts for ≈3% of all congenital heart diseases and 20% of the cyanotic congenital heart diseases.[1] In most patients with TGA, ventriculoarterial discordance is the major culprit in that the aorta arises from the morphological right ventricle, and the pulmonary artery arises from the morphological left ventricle.[2] Once being diagnosed, patients have to be operated on promptly after birth, otherwise deferred surgical repair will cause higher morbidity, neurological development defects, and psychosocial problems.[3–5] As recently as a decade ago, the total hospital cost for 1 patient who underwent an arterial switch operation (ASO) had reached $55,000,[6] and additional costs may be incurred by the possibility of subsequent adverse cardiovascular events, putting a huge economical strain on both families and society.

As a complex congenital heart disease, TGA is frequently accompanied by other cardiac malformations, the most common of which are ventricular septal defect (VSD) and pulmonic stenosis (PVS). These additional complex and diverse cardiac phenotypes have

Correspondence to: Shoujun Li, MD, Pediatric Cardiac Surgery Centre, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing 100037, China. E-mail: drlishoujun@yahoo.com and Zhou Zhou, MD, PhD, Department of Laboratory Medicine, State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100037, China. E-mail: zhouzhou@fuwai.com

*H. Shen and Q. He contributed equally.

Supplemental Material for this article is available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.121.023181

For Sources of Funding and Disclosures, see page 10.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- Phenotypic risk stratification of patients using the Human Phenotype Ontology database was applied for the first time in a large transposition of the great arteries cohort, highlighting the importance of concomitant cardiac malformations for the outcome of patients with transposition of the great arteries.
- Clustering based on phenotypic similarity may serve as a supplement for researchers to establish a unbiased phenotype cohort to more accurately analyze the impact of other factors on prognosis.

What Are the Clinical Implications?

- Patients with complex transposition of the great arteries (with ventricular septal defect or pulmonic stenosis) should continue to pay close attention to their physical condition after surgical repair, ensure regular review, and receive timely medical intervention if necessary.
- The arterial switch operation should be performed on all patients with favorable anatomy, and the timing of surgery should be emphasized to obtain a good prognosis.

Nonstandard Abbreviations and Acronyms

ASO	arterial switch operation
DRT	double root translocation
HPO	Human Phenotype Ontology
LVOTO	left ventricular outflow tract obstruction
PVS	pulmonic stenosis
TGA	transposition of the great arteries

an impact on the patients' surgical strategies.[7–9] In addition, concomitant heart malformations can significantly affect patient outcomes, such as increasing the risk of death and reintervention.[10–13] However, phenotypes considered in the current studies are relatively limited, and patients are mainly classified by VSD and left ventricular outflow tract obstruction (LVOTO),[14] without considering the influence of other rare or less-severe phenotypes. Consequently, bias might exist.

We used the Human Phenotype Ontology (HPO) database, a comprehensive collection of systematically defined and logically organized human phenotypes,[15] to stratify a large TGA cohort based on cardiovascular phenotypes. In combination with surgical strategies, we analyzed the correlation between phenotypes and prognosis in patients whose major diagnosis was TGA with concordant atrioventricular and discordant ventriculoarterial connections. We hypothesized that the 3 distinct clusters obtained by the HPO database could reveal valuable prognostic value, and aimed to figure out the phenotypes that affect the prognosis, and explore whether the HPO database can be administered as a powerful tool for risk stratification in patients with complex congenital heart disease to help clinical decision making and prognosis analysis.

METHODS

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Study Population

We collected a total of 773 patients whose diagnoses were TGA with concordant atrioventricular and discordant ventriculoarterial connections and who received surgical treatment in Fuwai Hospital. The admission period of the patients in our hospital ranged from 1997 to 2019. The information extracted from electronic medical records of these patients was recorded, double-checked by 2 authors, and confirmed by a senior pediatric cardiac surgeon. The main content included basic information (date of birth, sex, height and weight at the time of surgery, family history), diagnostic information (admitting diagnosis, surgical diagnosis, and discharge diagnosis), examination information (echocardiogram and electrocardiograph), and detailed surgical notes (anatomy from the surgical view). For the coronary artery patterns, they were confirmed by a combination of echocardiography and surgical view under the Leiden convention.[16] The coronary artery pattern of 1LCx-2R was considered as the usual pattern, whereas any other coronary artery patterns were treated as abnormal patterns. The 2 reviewers had no disputes about the TGA diagnosis of the included patients. Patients with more complex intracardiac malformations (4 patients with double outlet right ventricle, 6 patients with single atrium, 17 patients with single ventricle, 3 patients with cor triatrium, 11 patients with pulmonary atresia, and 7 patients with atrial isomerism) were excluded, because these patients might have a more sophisticated treatment process and poorer prognosis, which might bias our analysis. Eventually, we enrolled 738 patients, none of whom had severe extracardiac disease. Because the study was retrospective, informed consent was waived. The study was approved by the Institutional Review Board of the Fuwai Hospital of Chinese Academy of Medical Science and Peking Union Medical College and the ethics committee of Fuwai Hospital (No. 2020-1402).

Phenotype Standardization

Phenotypic terms were extracted from the patient's diagnostic information (surgical diagnosis combined with primary diagnosis as the primary criteria), and phenotypic terms were standardized by the HPO database (version: HPO released June 2021).

HPO-Based Clustering

The clustering method has been previously reported.[17–19] First, the pairwise phenotypic similarity was calculated based on the frequency of each phenotype (p[T]) in the HPO database. The following formula was used to analyze the similarity between any 2 phenotype terms (eg, T1 and T2):

$$\mathrm{Sim}(T1,T2) = \max_{v \in \mathrm{anc}(T1) \cap \mathrm{anc}(T2)} -\log p(v).$$

$v \in \mathrm{anc}(T1) \cap \mathrm{anc}(T2)$: the set of common ancestor terms of T1 and T2.

Every patient may have at least 1 annotated HPO term, and we then calculated the similarity matrix (sim_mat) in pairwise patients (eg, P1 and P2) based on the between-term set similarities by the following equation:

$$\mathrm{sim}(P1,P2) = \frac{1}{2|P_1|}\sum_{T1\in P1} \max_{T2\in P2} \mathrm{Sim}(T1,T2) + \frac{1}{2|P_2|}\sum_{T1\in P2} \max_{T2\in P1} \mathrm{Sim}(T1,T2)$$

We calculated a distance matrix (max[sim_mat]–sim_mat) based on similarity matrix (sim_mat, the similarity in any 2 of the patients) guided by the R package ontologySimilarity (https://rdrr.io/cran/ontologySimilarity/f/vignettes/ontologySimilarity-examples.Rmd). According to the distance matrix, unsupervised hierarchical clustering was performed by the R pheatmap package. For the selection of the parameter in the pheatmap function, the complete linkage method was used for hierarchical clustering by default. The distance between 2 clusters is the maximal distance between any 2 elements in each cluster, and cutree_col was set to be 3 to obtain 3 phenotypically heterogeneous clusters (cutree_col was set to be 6 to construct subgroups of cluster 1).

The packages of OntologySimilarity, OntologyIndex, and OntologyPlot in R were used to analyze the above analysis.

Revisit Records Collection and Follow-Up

For all enrolled patients, we collected all revisit records in our hospital, including physical examination, echocardiography, electrocardiogram, and magnetic resonance imaging if available. In addition, we conducted a follow-up telephone interview of all patients, in which 492 patients were reached. For the remaining patients who were lost to telephone interviews, we treated the last revisit records in our hospital as the primary outcome judgment. We inquired about the patient's survival, readmission, revisit records at other hospitals, medication status, exercise tolerance, and family history of congenital heart disease. We defined all-cause mortality as the primary end point and reintervention (any heart-related surgery after definitive surgery) as the secondary end point. Eventually, 72 patients with no follow-up information were excluded, and the remaining 666 patients were included for further analysis. The characteristics of 72 patients lost to follow-up were as follows: age at definitive surgery was 0.28 years (25th–75th percentile: 0.05 years–1.38 years), female patients (n=20, 27.8%), body mass index (14.30±2.50 kg/m^2), surgery ([ASO] n=55, 76.4%; double root translocation [DRT], n=5, 6.9%; "other," n=12, 16.7%), atrial septal defect ([ASD] n=42, 58.3%), VSD (n=44, 61.1%), patent ductus arteriosus (PDA, n=43, 59.7%), patent foramen ovale (n=16, 22.2%), PVS (n=15, 20.8%), cardiomegaly (n=21, 29.2%), with no statistical difference compared with 666 patients included in the study.

Statistical Analysis

All statistical analyses involved in this study were performed by SPSS Statistics version 23.0 (IBM, Armonk, NY) and R software version 3.6.2 (R Foundation for Statistical Computing, Vienna, Austria). Categorical variables were summarized as frequency (percentage) and compared via χ^2 test or Fisher exact test when group size <10. Continuous variables were summarized as mean±standard deviation or median (25th–75th percentiles). An independent *t* test was used in normal distribution, and the Wilcoxon rank sum test was used when skew distribution occurred.

We used the Kaplan-Meier method to estimate the survival and freedom from adverse events, and the log-rank test was used to compare the difference between clusters. We set the survival time of the enrolled patients at definitive surgery and ended at end points (death, reintervention, or the last follow-up). Stratification of the Kaplan-Meier analysis was based on different clusters classified by ontology and different surgical strategies (ASO, DRT, and others).

To investigate the risk factors contributing to prognosis, the Cox proportional hazard regression method was used. After clustering, significant covariates (P<0.05 in Table 1) between clusters with a percentage >5% were tested in the univariate analysis. Covariates with a P value <0.05 were included in the multivariate analysis model. For those covariates that did not show significance but did show clinical relevance, they were also included in the multivariate analysis. The multivariate Cox regression model was used in both the overall cohort and subgroups analysis as a coherent approach.

Table 1. Baseline Demographics of Patients at Definitive Surgery and Follow-Up

Variables	All cohort, n=666	Cluster 1, n=422	Cluster 2, n=205	Cluster 3, n=39	*P* value
Age at definitive surgery, y	0.22 (0.07–1.52)	0.11 (0.05–0.29)	1.85 (0.82–4.34)	0.39 (0.19–1.44)	<0.001*
Age at follow-up, y	6.27 (3.26–9.34)	5.27 (2.49–8.53)	8.08 (4.62–11.33)	7.08 (4.43–11.52)	<0.001*
Follow-up, y	4.91 (2.10–7.96)	4.85 (1.92–7.89)	4.84 (2.23–7.78)	6.35 (2.67–10.34)	0.159
Female patients	174 (26.13%)	97 (22.99%)	65 (31.71%)	12 (30.77%)	0.052
BMI	14.49±2.39	14.02±2.40	15.33±2.13	15.18±2.29	<0.001*
Surgical strategy					<0.001*
ASO	504 (75.68%)	408 (96.68%)	60 (29.27%)	36 (92.31%)	
DRT	80 (12.01%)	6 (1.42%)	73 (35.61%)	1 (2.56%)	
Other	82 (12.31%)	8 (1.90%)	72 (35.12%)	2 (5.13%)	
Associated anomalies					
VSD	433 (65.02%)	221 (52.37%)	195 (95.12%)	17 (43.59%)	<0.001*
ASD	369 (55.41%)	261 (61.84%)	83 (40.49%)	25 (64.10%)	<0.001*
PDA	357 (53.60%)	309 (73.22%)	42 (20.49%)	6 (15.38%)	<0.001*
PVS	205 (30.78%)	16 (3.79%)	187 (91.22%)	2 (5.13%)	<0.001*
POF	191 (28.68%)	145 (34.36%)	39 (19.02%)	7 (17.95%)	<0.001*
Cardiomegaly	95 (14.26%)	44 (10.43%)	39 (19.02%)	12 (30.77%)	<0.001*
Coronary artery abnormality	180 (27.03%)	98 (23.22%)	47 (22.93%)	35 (89.74%)	<0.001*
ACAO	71 (10.66%)	46 (10.90%)	13 (6.34%)	12 (30.77%)	<0.001*
SCA	63 (9.46%)	30 (7.11%)	16 (7.80%)	17 (43.59%)	<0.001*
CPAF	1 (0.15%)	0	1 (0.49%)	0	0.366
Other	46 (6.91%)	23 (5.45%)	17 (8.29%)	6 (15.38%)	0.044*
LSVC	26 (3.90%)	5 (1.18%)	15 (7.32%)	6 (15.38%)	<0.001*
LSVCC	7 (1.05%)	2 (0.47%)	3 (1.46%)	2 (5.13%)	0.032*
OSVC	1 (0.15%)	0	0	1 (2.56%)	0.059
MAPCA	19 (2.85%)	0	18 (8.78%)	1 (2.56%)	<0.001*
Aortic abnormality	23 (3.45%)	9 (2.13%)	12 (5.85%)	2 (5.13%)	0.038*
RAA	14 (2.10%)	1 (0.24%)	12 (5.85%)	1 (2.56%)	<0.001*
HAA	4 (0.60%)	4 (0.95%)	0	0	0.458
CoA	4 (0.60%)	4 (0.95%)	0	0	0.458
IAA	1 (0.15%)	1 (0.24%)	0	0	1.000
OA	1 (0.15%)	0	0	1 (2.56%)	0.057
PS	13 (1.95%)	3 (0.71%)	8 (3.90%)	2 (5.13%)	0.005*
Tricuspid valve abnormality	16 (2.40%)	1 (0.24%)	14 (6.83%)	1 (2.56%)	<0.001*
TR	9 (1.35%)	1 (0.24%)	7 (3.41%)	1 (2.56%)	0.003*
OTV	4 (0.60%)	0	4 (1.95%)	0	0.034*
TA	2 (0.30%)	0	2 (0.98%)	0	0.208
HTV	1 (0.15%)	0	1 (0.49%)	0	0.366
Aortic valve abnormality	5 (0.75%)	1 (0.24%)	4 (1.95%)	0	0.083
AR	3 (0.45%)	1 (0.24%)	2 (0.98%)	0	0.375
BAV	2 (0.30%)	0	2 (0.98%)	0	0.208
QAV	1 (0.15%)	0	1 (0.49%)	0	0.366
Mitral valve abnormality	3 (0.45%)	0	3 (1.46%)	0	0.057
MR	1 (0.15%)	0	1 (0.49%)	0	0.366
MS	1 (0.15%)	0	1 (0.49%)	0	0.366
Other	1 (0.15%)	0	1 (0.49%)	0	0.366
PI	3 (0.45%)	0	3 (1.46%)	0	0.055
PAPVC	2 (0.30%)	0	2 (0.98%)	0	0.208

(*Continued*)

Table 1. Continued

Variables	All cohort, n=666	Cluster 1, n=422	Cluster 2, n=205	Cluster 3, n=39	*P* value
TAPVC	1 (0.15%)	0	0	1 (2.56%)	0.059
PAD	1 (0.15%)	0	1 (0.49%)	0	0.366
HRH	8 (1.20%)	0	7 (3.41%)	1 (2.56%)	<0.001*
HLH	3 (0.45%)	0	0	3 (7.69%)	<0.001*
RVOTO	6 (0.90%)	2 (0.47%)	3 (1.46%)	1 (2.56%)	0.142
Dextrocardia	11 (1.65%)	4 (0.95%)	6 (2.93%)	1 (2.56%)	0.108
Mesocardia	1 (0.15%)	0	1 (0.49%)	0	0.366
CCH	4 (0.60%)	1 (0.24%)	3 (1.46%)	0	0.210
AVB	3 (0.45%)	0	3 (1.46%)	0	0.055

All values are presented as mean±SD, median (25th–75th percentiles), or n (%). ACAO indicates abnormal coronary artery origin; AR, aortic regurgitation; ASD, atrial septal defect; ASO, atrial septal defect; AVB, atrioventricular block; BAV, bicuspid aortic valve; BMI, body mass index; CCH, criss-cross heart; CoA, coarctation of aorta; CPAF, coronary-pulmonary artery fistula; DRT, double root translocation; HAA, hypoplastic aortic arch; HLH, hypoplastic left heart; HRH, hypoplastic right heart; HTV, hypoplastic tricuspid valve; IAA, interrupted aortic arch; LSVC, persistent left superior vena cava; LSVCC, left superior vena cava draining to coronary sinus; MAPCA, aortopulmonary collateral arteries; MR, mitral regurgitation; MS, mitral stenosis; OA, overriding aorta; OSVC, obstruction of the superior vena cava; OTV, overriding tricuspid valve; PAD, pulmonary artery dilatation; PAPVC, partial anomalous pulmonary venous return; PDA, patent ductus arteriosus; PI, pulmonary insufficiency; POF, patent foramen ovale; PS, pulmonary artery stenosis; PVS, pulmonic stenosis; QAV, quadricuspid aortic valve; RAA, right aortic arch; RVOTO, right ventricular outflow tract obstruction; SCA, single coronary artery origin; TA, tricuspid atresia; TAPVC, total anomalous pulmonary venous return; TR, tricuspid regurgitation; and VSD, ventricular septal defect.

*Statistical significance.

RESULTS

Baseline Characteristics

Detailed baseline characteristics of this cohort are documented in Table 1. The median follow-up period of this cohort was 4.91 years (25th–75th percentile: 2.10 years–7.96 years), and no significant difference could be observed among the 3 clusters (P=0.159). Age at definitive surgery was 0.22 years (25th–75th percentile: 0.07 years–1.52 years) with a significant difference (P<0.001). Female patients comprised 26.13% of the whole cohort, ranging from 22.99% to 31.71% among the 3 clusters (P=0.052). Body mass index was significantly different among the 3 clusters (P<0.001), with the lowest in cluster 1 (14.02±2.40 kg/m^2) and the highest in cluster 2 (15.33±2.13 kg/m^2).

Standardization and Annotation for Phenotypes

We derived cardiovascular disease-related diagnostic terms from the patients' clinical records and annotated them into the HPO database. The following 4 situations occurred in this process: (1) The diagnostic term could be fully matched to the HPO database, and it was supported for analysis in the R package OntologySimilarity. (2) The diagnostic term could be fully matched to the HPO database, but it was not supported in the R package OntologySimilarity, and we then used its superior term instead. (3) The diagnostic term could not be matched to the HPO database, but the diagnosis had been reported in other literature and had a standard name, and we then categorized it accurately and replaced it with a corresponding term in HPO. (4) For the vague diagnostic description, which did not exist in the HPO database and did not correspond to a specific disease with a standard name, we categorized it accurately and replaced it with a corresponding term in HPO. According to the above criteria, we annotated the patients' diagnosis to 42 HPO terms; detailed information is documented in Table S1. We made a tree plot representing the hierarchy and distribution of terms annotated in our cohort, and the shade of color represented the frequency information of the terms in the HPO database (Figure 1). It can be seen that the cardiovascular phenotypes of patients with TGA are complex and varied, with categorized distribution in cardiac morphology, vascular abnormalities, and arrhythmias. In this cohort, each patient had at least 1 additional cardiovascular phenotype, and 4 patients had up to 7 comorbidities. The detailed phenotypic combinations of the patients are documented in Table S2.

HPO Database Clustering for Patients With TGA

We defined a large cohort of the TGA population by a set of diagnostic terms and used unsupervised hierarchical clustering to classify patients based on their phenotypic similarity. Eventually, the patients were divided into 3 clusters, with a sample size of 422, 205, and 39, respectively (Figure 2). Each cluster had phenotypic characteristics that were significantly different from the other groups (Table 1). According to significant differences in phenotypic distribution, our clustering modality was mainly dominated by ASD, VSD, patent foramen ovale, PDA, PVS, coronary artery abnormality, and cardiomegaly. We defined cluster 1 as mainly ASD and PDA, cluster 2 was associated with VSD and PVS,

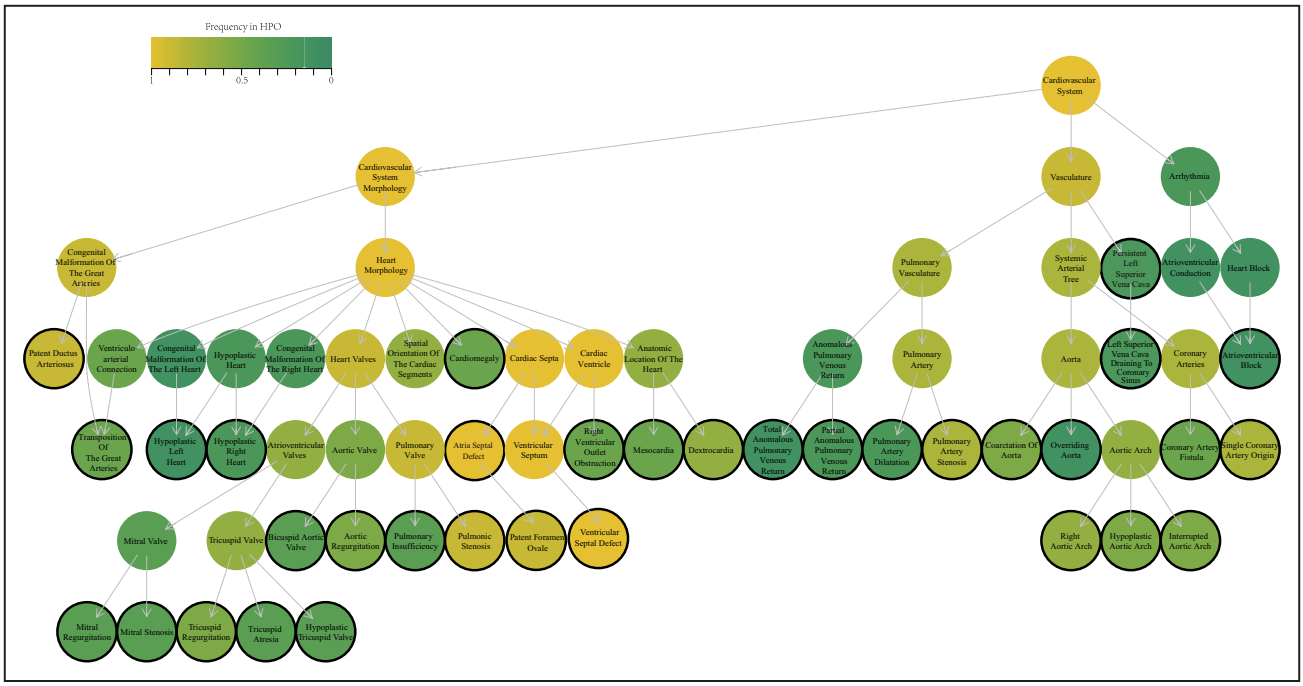

Figure 1. Tree plot of all annotated terms in the whole cohort.
The tree plot shows the relationship of all phenotypic terms. Circles with borders indicate the presence in our cohort. The shade of color indicates the frequency of terms in the Human Phenotype Ontology (HPO) database (with reference of the color key on top). The arrows represent the affiliation relationship ("is a") between terms in the ontology.

and cluster 3 was a relatively small marginal group with a high proportion of coronary artery abnormalities. In addition to differences in the distribution of phenotypic types, we found differences in the distribution of the number of phenotypes among the 3 groups. The average number of terms carried per patient in clusters 1, 2, and 3 were 2.659, 3.585, and 3.179, respectively, which suggested that the patients in cluster 2 had more complex and diverse phenotypes and might be more clinically challenging. Tree diagrams showing the distribution of phenotypes in different clusters are displayed in Figure S1.

Surgery Description

In our cohort, 504 patients underwent ASO, 80 underwent DRT, and 82 underwent other types of surgical intervention (Table 1). ASO and DRT were identified as definitive surgery, and definitions of definitive surgery for other types of surgical interventions are shown in Table S3. The distribution of surgical strategies varied significantly among the 3 clusters; >90% of patients in cluster 1 and cluster 3 received ASO, whereas <30% of patients in cluster 2 received ASO, and most of them underwent DRT or other types of surgeries.

Prognosis Analysis

A total of 38 patients suffered all-cause mortality (including 7 hospital deaths), and 33 patients experienced reinterventions. Survival rates at 30 days, 1 year, and 5 years were 98.20% (97.19%–99.21%), 95.68% (94.12%–97.26%), and 94.70% (92.95%–96.49%), respectively (Table S4). We used univariate and multivariate Cox proportional regression analyses to analyze the risk of various factors in patients, revealing that no factor increased the risk of all-cause mortality significantly in multivariate analysis, whereas DRT (P=0.028) was the independent risk factor associated with reintervention (Table S5, Table 2). After phenotypic stratification of patients by the HPO database, patients with closer phenotypic similarity were grouped so that there might be cluster-specific risk factors. For cluster 1 with a simpler phenotype composition, age at definitive surgery significantly increased patients' mortality risk (P=0.002), which indicated that for these patients, more consideration should be given to the timing of surgery, and early treatment should be provided to achieve a more satisfactory prognosis. Cluster 2 was a heterogeneous group with complex and divergent phenotypic combinations, in which single coronary artery (P=0.009) and receiving other types of surgeries (P=0.031) were risk factors, VSD (P=0.026) was a protective factor for survival, and DRT (P=0.033) was associated with a higher reintervention risk in multivariate analysis. We did not analyze or describe cluster 3 because of its relatively limited number of patients. More details are shown in Table 2.

It would be intriguing to see whether there was a significant prognostic difference among the 3 clusters

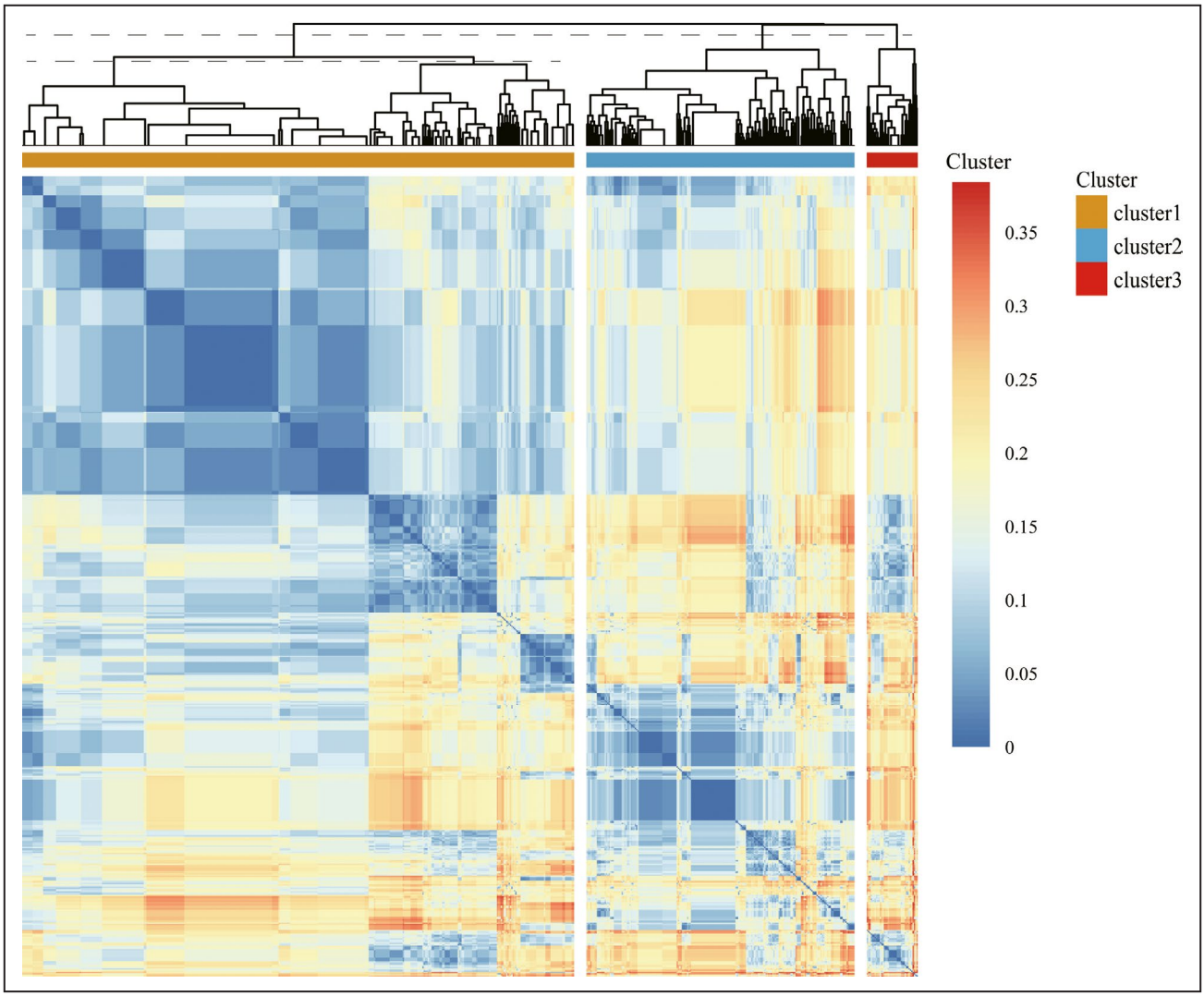

Figure 2. Heatmap of the clustering.
The heatmap was generated from a distance matrix calculated by the phenotypic similarity of all patients. The upper dashed line indicates the height to cut the tree into 3 clusters, and the lower dashed line indicates the height to cut cluster 1 into 2 subgroups. The shade of color indicates the similarity between patients (the color key is on the right).

of patients with different phenotypes. First, the Kaplan-Meier curve was used to compare survival and freedom from reintervention. It was found that there was no significant difference in overall survival among the 3 clusters (Figure 3A, P=0.330), but the risk of reintervention was significantly higher in cluster 2 than in the other 2 clusters (Figure 3B, P<0.001). The choice of the appropriate surgical strategy may also have a great impact on the prognosis of patients. Therefore, patients were classified into 3 groups according to the different surgeries received, namely the ASO group, DRT group, and "other" group. We found significant differences in both overall survival (Figure S2A, P=0.018) and freedom from reintervention (Figure S2B, P<0.001) among the 3 surgical groups. Patients in the "other" group had a worse survival rate, and patients in the DRT group had a higher risk of reintervention. To control for variables, we analyzed all patients receiving ASO among 3 phenotypic clusters, and found that there was no difference in survival (Figure S3A, P=0.620) but a significant difference in the risk of reintervention (Figure S3B, P=0.019).

It was obvious that there were 2 subgroups in cluster 1 according to the clustering heatmap (Figure 2); thus, we conducted a further subgroup analysis. The same multivariate model was used as a coherent approach, whereas the covariates that did not exist were excluded. Compared with subgroup 2, patients in subgroup 1 only had ASD, VSD, PDA, patent foramen ovale, and PVS, so they were more homogeneous in phenotypic distribution, whereas patients in subgroup 2 had a higher frequency of coronary artery abnormalities and cardiomegaly with some rare phenotypes scattering. Table S6 shows the significant phenotypes driving to these 2 subgroups. However, we

Table 2. Results of Multivariate Cox Proportional Analysis

Variables	Death		Reintervention	
	Hazard ratio (95% CI)	*P* value	Hazard ratio	*P* value
Overall cohort				
Age at definitive surgery	0.994 (0.902–1.096)	0.903	0.923 (0.795–1.071)	0.291
Female patients	0.366 (0.695–2.687)	0.366	1.167 (0.522–2.609)	0.706
PDA	0.602 (0.295–1.229)	0.164	0.679 (0.305–1.512)	0.343
ASD	1.725 (0.847–3.511)	0.133	0.738 (0.339–1.607)	0.445
VSD	0.724 (0.316–1.662)	0.447	1.209 (0.343–4.260)	0.767
PVS	1.399 (0.450–4.344)	0.562	2.293 (0.668–7.876)	0.187
SCA	1.239 (0.431–3.562)	0.691	0.962 (0.286–3.238)	0.951
Surgery				
ASO	Reference		Reference	
DRT	1.190 (0.307–4.619)	0.802	3.793 (1.154–12.465)	0.028*
Other	2.166 (0.664–7.061)	0.200	1.466 (0.388–5.537)	0.573
Cluster 1				
Age at definitive surgery	1.293 (1.100–1.520)	0.002*	1.105 (0.773–1.580)	0.582
Female patients	1.964 (0.774–4.988)	0.156	0.472 (0.058–3.852)	0.483
PDA	0.601 (0.221–1.630)	0.317	2.700 (0.317–22.971)	0.363
ASD	1.599 (0.544–4.703)	0.393	1.009 (0.260–3.908)	0.990
VSD	1.045 (0.374–2.917)	0.933	1.223 (0.262–7.700)	0.798
PVS	1.791 (0.099–32.332)	0.693	0 (0–Inf)	0.962
SCA	0 (0–Inf)	0.976	0 (0–Inf)	0.926
Surgery				
ASO	Reference		Reference	
DRT	0.984 (0.034–28.545)	0.992	$4.376e^{4}$ (0–Inf)	0.952
Other	0.978 (0.057–16.664)	0.988	0.001 (0–Inf)	0.970
Cluster 2				
Age at definitive surgery	0.781 (0.582–1.047)	0.098	0.867 (0.710–1.058)	0.160
Female patients	1.031 (0.310–3.427)	0.961	1.774 (0.692–4.546)	0.232
PDA	0.395 (0.081–1.921)	0.250	0.351 (0.077–1.596)	0.175
ASD	1.581 (0.509–4.909)	0.428	0.857 (0.304–2.417)	0.770
VSD	0.130 (0.021–0.786)	0.026*	$1.563e^{5}$ (0–Inf)	0.980
PVS	1.363 (0.111–16.666)	0.808	0.522 (0.102–2.670)	0.435
SCA	7.452 (1.660–33.456)	0.009*	1.229 (0.268–5.630)	0.791
Surgery				
ASO	Reference		Reference	
DRT	3.134 (0.506–19.426)	0.220	4.078 (1.117–14.890)	0.033*
Other	6.537 (1.190–35.896)	0.031*	1.154 (0.281–4.751)	0.842

ASD indicates atrial septal defect; ASO, arterial switch operation; DRT, double root translocation; Inf, Infinity; PDA, patent ductus arteriosus; PVS, pulmonic stenosis; SCA, single coronary artery origin; and VSD, ventricular septal defect.

*Statistical significance.

found no significant discrepancy in prognosis (Figure S4, *P*=0.250 for survival, *P*=0.090 for reintervention). Age at definitive surgery was the risk factor for survival in subgroup 1 (*P*=0.006). Detailed information about the subgroup of cluster 1 is documented in Table S7.

Sex diversity has been a profoundly established factor of clinical concern. In our cohort, the male/female ratio was about 3:1, which is consistent with an earlier report.[20] The incidence of adverse outcomes was not significantly different between men and women (Figure S5, *P*=0.230 for survival, *P*=0.690 for reintervention). For male patients, receiving other types of surgeries was an independent risk factor to death, whereas age at definitive surgery and DRT were the risk factors for reintervention for female patients. Because the female population was relatively small, our results need to be verified. More information can be obtained in Table S8.

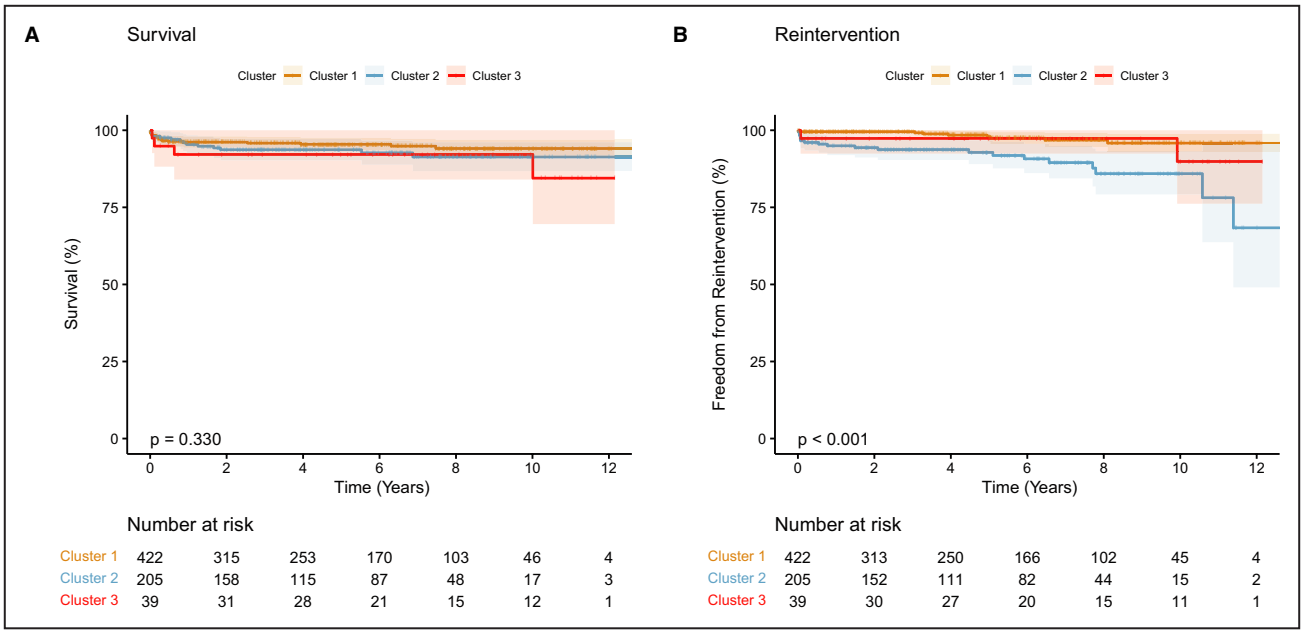

Figure 3. Kaplan-Meier curve for outcomes classified by clusters.
A, Kaplan-Meier curve showing the overall survival rate among 3 clusters; no significance is observed (*P*=0.330). **B**, Kaplan-Meier curve showing the freedom from reintervention among 3 clusters; a significant reintervention rate is observed in cluster 2 compared with the other 2 clusters (*P*<0.001).

DISCUSSION

TGA is complex congenital heart disease with a high degree of genetic correlation,[21] so there is a high risk of accompanied heart defects. For accurate preoperative risk stratification, operative planning, and family counseling, we should pay more attention to additional anatomical malformations.[13] Heterogeneous phenotypic distributions of patients were observed in our cohort, influencing the clinicians' decisions and leading to different outcomes for patients. The top 5 most common phenotypes were VSD, ASD, PDA, PVS, and patent foramen ovale, with the frequency of 65.02%, 55.41%, 53.60%, 30.78%, and 28.68%, respectively.

Based on phenotypic similarity, we successfully divided a large population of patients with TGA into 3 clusters. Each cluster has its distinct characteristics and the complexity of phenotypic distribution varieties. Cluster 1 has a more homogeneous phenotypic combination, whereas cluster 2 is heterogeneous, with a total occurrence of 34 additional terms and 3.585 terms carried by each patient. Cluster 1 is characterized by some simple concomitant malformations, such as septal defects and PDA. More than 95% of patients received ASO, and their outcomes were satisfying. We noticed that the time for definitive surgery intervention in cluster 1 was much earlier than the other 2 groups, which might be attributable to the simpler cardiac anatomic malformation. The patients in cluster 3 had no distinct features but a higher percentage of coronary abnormalities in which ≈90% of patients were affected. Most of them received ASO and acquired an acceptable prognosis. The pattern of coronary arteries has always been a concern by surgeons, and transfer of the coronary artery is key to a successful ASO,[11] and abnormal coronary artery patterns affect the prognosis of patients.[10] However, some studies have shown that the coronary artery pattern is no longer an independent risk factor of early mortality,[22] and its effect will dissipate over time.[23] Our results indicated that the effect of coronary artery on prognosis was not significant, although ≈30% of patients in the entire cohort had coronary malformations. However, intriguingly, single coronary artery was a specific risk factor to survival for patients in cluster 2, which meant coronary patterns might be more important for complex TGA. Cluster 2, a group of patients with heterogeneous phenotypes featuring VSD and PVS, had undergone a wide range of surgeries and had shown a worse prognosis. The commonly accompanied phenotypes for TGA, VSD, and LVOTO are the most clinically challenging factors.[24] Surgical management for patients with anomalies of ventriculoarterial connection associated with VSD as well as pulmonary outflow tract obstruction continues to be a challenge because the anatomic correction of these lesions requires a complete reconstruction of the biventricular outflow tract.[25] VSD was reported as the risk factor for in-hospital mortality, and patients who had TGA with VSD, or TGA with VSD plus LVOTO had a 3-fold greater risk for late mortality than patients who had TGA with intact ventricular septal.[26] A

meta-analysis also indicated that patients with complex TGA, commonly defined in the literature as TGA with VSD, PVS, or other LVOTO, had a higher risk for late mortality.[12] Furthermore, complex TGA, VSD, coronary anomalies, aortic coarctation, LVOTO, and moderate pulmonary stenosis are strong predictors for poor overall free-reoperation survival,[27] which is in part consistent with our results. Therefore, patients with TGA with VSD or PVS should continue to pay close attention to their physical condition after surgical repair and ensure regular revisits and timely medical intervention if needed.

After the success of the first ASO, Jatene et al have ushered in a new era of TGA repair, in which patient outcomes have greatly improved.[28] In our cohort, 5-year survival after ASO is as high as 95.66% (93.84%–97.50%). The ASO should be performed on all patients with suitable indications, and is decided by a favorable anatomy.[7] For patients with unfavorable malformations contraindicated for the ASO, DRT or other types of surgeries were alternatives, but the prognosis was not as satisfactory as with the ASO. Our results indicated the increasing need for reintervention after DRT and other types of surgeries leading to increased mortality without considering other factors. Collectively, the choice of a specific surgical procedure and different prognoses are attributable to the distinct phenotypes, resulting in the varying physiological status of patients.

In general, TGA could be classified into 3 categories: TGA with intact ventricular septum, TGA with VSD, and TGA with VSD and LVOTO.[14] Our patients in cluster 2 were similar to those who had TGA with VSD and LVOTO. However, because the majority of our patients had VSD, VSD was distributed across 3 clusters and did not mainly contribute to the classification results. In addition, all cardiovascular phenotypes of the patients were included in the analysis, which made our clustering modality different and novel. It should be noted that our purpose was mainly to risk stratify patients, rather than to define and group patients specifically. Moreover, because of the HPO algorithm, our clustering results will change with the different patient populations included, and only when the cohort is large enough will it tend to be stable.

Our risk stratification is more accurate in identifying patients with high-risk phenotypes than previous general classification according to our prognostic analysis. We selected patients who had TGA with intact ventricular septum, TGA with VSD, or TGA with VSD and LVOTO from each of the 3 clusters, and compared the outcomes of conventionally classified patients in different clusters. The results showed that patients with TGA and intact ventricular septum in cluster 2, patients with TGA and VSD in cluster 2, and patients with TGA and VSD plus LVOTO in cluster 1 had worse outcomes compared with patients in the same traditional group in other clusters (Figure S6). Therefore, if the prognosis of the patient needs to be analyzed, only VSD and LVOTO as the classification basis are not enough. Our clustering, which is based on phenotypic similarity, may also serve as a supplement for researchers to establish an unbiased phenotype cohort to more accurately analyze the impact of other factors on prognosis.

Our study has some limitations that should be considered. First, the nature of a single-center study might lead to the inapplicability of our results to patients with TGA in all regions. Second, during the process of clustering, we did not fully consider the subtypes and severity of phenotypes, such as the position and number of VSDs and the degree of PVS. Third, our median follow-up time was 4.92 years, so that the analysis for long-term postoperative adverse events, such as neo-aortic valve regurgitation, neo-aortic root dilation, and coronary artery disease after ASO, might not be adequate. Fourth, we did not consider patients with more severe and complex phenotypes (double outlet right ventricle, single atrium, single ventricle, pulmonary atresia, cor triatrium, and atrial isomerism), which might cause bias. Fifth, because of the noncooperation of the patients' families, we were unable to obtain the cause of death of all patients, which might affect the accuracy of our results. Last, the "other" surgery group, which consists of 11 different operations, was not analyzed separately.

CONCLUSIONS

Here, we successfully delineated a big picture of cardiovascular phenotypes and used the HPO database to perform risk stratification based on phenotypic similarity in an unprecedentedly large TGA population. In addition, we reported postoperative survival situations in a large Chinese population.

ARTICLE INFORMATION

Received July 8, 2021; accepted December 9, 2021.

Affiliations

Department of Laboratory Medicine, State Key Laboratory of Cardiovascular Disease (H.S., X.S., Z.Z.) and Pediatric Cardiac Surgery Center (Q.H., S.L.), Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China.

Acknowledgments

The authors thank all patients enrolled in this study.

Sources of Funding

The study was supported by the Chinese Academy of Medical Sciences Initiative for Innovative Medicine (CAMS-I2 M) (number 2016-I2 M-1-016, Z.Z.), National Natural Science Foundation of China (number 82070326, Z.Z.), National Key R&D Program of China (number 2017YFC1308100, S.L.), and Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences (CIFMS, 2020-I2 M-C&T-A-009, S.L.).

Disclosures

None.

Supplemental Material

Tables S1–S8
Figures S1–S6

REFERENCES

1. Reller MD, Strickland MJ, Riehle-Colarusso T, Mahle WT, Correa A. Prevalence of congenital heart defects in metropolitan Atlanta, 1998–2005. *J Pediatr*. 2008;153:807–813. doi: 10.1016/j.jpeds.2008.05.059
2. Warnes CA. Transposition of the great arteries. *Circulation*. 2006;114:2699–2709. doi: 10.1161/CIRCULATIONAHA.105.592352
3. Anderson BR, Ciarleglio AJ, Hayes DA, Quaegebeur JM, Vincent JA, Bacha EA. Earlier arterial switch operation improves outcomes and reduces costs for neonates with transposition of the great arteries. *J Am Coll Cardiol*. 2014;63:481–487. doi: 10.1016/j.jacc.2013.08.1645
4. Robson VK, Stopp C, Wypij D, Dunbar-Masterson C, Bellinger DC, DeMaso DR, Rappaport LA, Newburger JW. Longitudinal Associations between neurodevelopment and psychosocial health status in patients with repaired D-Transposition of the great arteries. *J Pediatr*. 2019;204:38–45.e1. doi: 10.1016/j.jpeds.2018.08.069
5. Bellinger DC, Wypij D, Rivkin MJ, DeMaso DR, Robertson RL, Dunbar-Masterson C, Rappaport LA, Wernovsky G, Jonas RA, Newburger JW. Adolescents with d-transposition of the great arteries corrected with the arterial switch procedure: neuropsychological assessment and structural brain imaging. *Circulation*. 2011;124:1361–1369. doi: 10.1161/CIRCULATIONAHA.111.026963
6. Pasquali SK, Sun J-L, d'Almada P, Jaquiss RDB, Lodge AJ, Miller N, Kemper AR, Lannon CM, Li JS. Center variation in hospital costs for patients undergoing congenital heart surgery. *Circ Cardiovasc Qual Outcomes*. 2011;4:306–312. doi: 10.1161/CIRCOUTCOMES.110.958959
7. Bierbach B, Arenz C, Suchowerskyj P, Schroth S, Blaschczok J, Asfour B, Schneider M, Hraška V. Current mid-term outcome with an integrated surgical strategy for correction of d-transposition of the great arteries with ventricular septal defect and left ventricular outflow tract obstruction. *Eur J Cardiothorac Surg*. 2016;50:617–625. doi: 10.1093/ejcts/ezw058
8. Hazekamp M, Portela F, Bartelings M. The optimal procedure for the great arteries and left ventricular outflow tract obstruction. An anatomical study. *Eur J Cardiothorac Surg*. 2007;31:879–887. doi: 10.1016/j.ejcts.2007.02.014
9. Honjo O, Kotani Y, Bharucha T, Mertens L, Caldarone CA, Redington AN, Van Arsdell G. Anatomical factors determining surgical decision-making in patients with transposition of the great arteries with left ventricular outflow tract obstruction. *Eur J Cardiothorac Surg*. 2013;44:1085–1094. doi: 10.1093/ejcts/ezt283
10. Khairy P, Clair M, Fernandes SM, Blume ED, Powell AJ, Newburger JW, Landzberg MJ, Mayer JE. Cardiovascular outcomes after the arterial switch operation for d-transposition of the great arteries. *Circulation*. 2013;127:331–339. doi: 10.1161/CIRCULATIONAHA.112.135046
11. Villafañe J, Lantin-Hermoso MR, Bhatt AB, Tweddell JS, Geva T, Nathan M, Elliott MJ, Vetter VL, Paridon SM, Kochilas L, et al. D-transposition of the great arteries. *J Am Coll Cardiol*. 2014;64:498–511. doi: 10.1016/j.jacc.2014.06.1150
12. Venkatesh P, Evans AT, Maw AM, Pashun RA, Patel A, Kim L, Feldman D, Minutello R, Wong SC, Stribling JC, et al. Predictors of late mortality in D-transposition of the great arteries after atrial switch repair: systematic review and meta-analysis. *J Am Heart Assoc*. 2019;8:e012932. doi: 10.1161/JAHA.119.012932
13. Gottlieb D, Schwartz ML, Bischoff K, Gauvreau K, Mayer JE. Predictors of outcome of arterial switch operation for complex D-transposition. *Ann Thorac Surg*. 2008;85:1698–1703. doi: 10.1016/j.athoracsur.2008.01.075
14. Jaggers JJ, Cameron DE, Herlong JR, Ungerleider RM. Congenital heart surgery nomenclature and database project: transposition of the great arteries. *Ann Thorac Surg*. 2000;69:205–235. doi: 10.1016/S0003-4975(99)01282-5
15. Köhler S, Gargano M, Matentzoglu N, Carmody LC, Lewis-Smith D, Vasilevsky NA, Danis D, Balagura G, Baynam G, Brower AM, et al. The human phenotype ontology in 2021. *Nucleic Acids Res*. 2021;49:D1207–D1217. doi: 10.1093/nar/gkaa1043
16. Gittenberger-de Groot AC, Koenraadt WMC, Bartelings MM, Bökenkamp R, DeRuiter MC, Hazekamp MG, Bogers AJJC, Quaegebeur JM, Schalij MJ, Vliegen HW, et al. Coding of coronary arterial origin and branching in congenital heart disease: the modified Leiden convention. *J Thorac Cardiovasc Surg*. 2018;156:2260–2269. doi: 10.1016/j.jtcvs.2018.08.009
17. Tang X, Chen W, Zeng Z, Ding K, Zhou Z. An ontology-based classification of Ebstein's anomaly and its implications in clinical adverse outcomes. *Int J Cardiol*. 2020;316:79–86. doi: 10.1016/j.ijcard.2020.04.073
18. Westbury SK, Turro E, Greene D, Lentaigne C, Kelly AM, Bariana TK, Simeoni I, Pillois X, Attwood A, Austin S, et al. Human phenotype ontology annotation and cluster analysis to unravel genetic defects in 707 cases with unexplained bleeding and platelet disorders. *Genome Med*. 2015;7:36. doi: 10.1186/s13073-015-0151-5
19. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. The human phenotype ontology: a tool for annotating and analyzing human hereditary disease. *Am J Hum Genet*. 2008;83:610–615. doi: 10.1016/j.ajhg.2008.09.017
20. Digilio MC, Casey B, Toscano A, Calabrò R, Pacileo G, Marasini M, Banaudi E, Giannotti A, Dallapiccola B, Marino B. Complete transposition of the great arteries. *Circulation*. 2001;104:2809–2814. doi: 10.1161/hc4701.099786
21. Pierpont ME, Brueckner M, Chung WK, Garg V, Lacro RV, McGuire AL, Mital S, Priest JR, Pu WT, Roberts A, et al. Genetic basis for congenital heart disease: revisited: a scientific statement from the American Heart Association. *Circulation*. 2018;138:e653–e711. doi: 10.1161/CIR.0000000000000606
22. Blume ED, Altmann K, Mayer JE, Colan SD, Gauvreau K, Geva T. Evolution of risk factors influencing early mortality of the arterial switch operation. *J Am Coll Cardiol*. 1999;33:1702–1709. doi: 10.1016/S0735-1097(99)00071-6
23. Pasquali SK, Hasselblad V, Li JS, Kong DF, Sanders SP. Coronary artery pattern and outcome of arterial switch operation for transposition of the great arteries: a meta-analysis. *Circulation*. 2002;106:2575–2580. doi: 10.1161/01.CIR.0000036745.19310.BB
24. Silberbach M, Castro WL, Goldstein MA, Lucas RV, Edwards JE. Comparison of types of pulmonary stenosis with the state of the ventricular septum in complete transposition of the great arteries. *Pediatr Cardiol*. 1989;10:11–15. doi: 10.1007/BF02328629
25. Hu S-S, Liu Z-G, Li S-J, Shen X, Wang X, Liu J, Yan F-X, Wang L, Li Y. Strategy for biventricular outflow tract reconstruction: Rastelli, REV, or Nikaidoh procedure? *J Thorac Cardiovasc Surg*. 2008;135:331–338. doi: 10.1016/j.jtcvs.2007.09.060
26. Hörer J, Schreiber C, Cleuziou J, Vogt M, Prodan Z, Busch R, Holper K, Lange R. Improvement in long-term survival after hospital discharge but not in freedom from reoperation after the change from atrial to arterial switch for transposition of the great arteries. *J Thorac Cardiovasc Surg*. 2009;137:347–354. doi: 10.1016/j.jtcvs.2008.09.015
27. Prifti E, Crucean A, Bonacchi M, Bernabei M, Murzi B, Luisi SV, Vanini V. Early and long term outcome of the arterial switch operation for transposition of the great arteries: predictors and functional evaluationq. *Thorac Surg*. 2002;22:864–873. doi: 10.1016/S1010-7940(02)00613-9
28. Jatene AD, Fontes VF, Paulista PP, Souza LCB, Neger F, Galantier M, Sousa JEMR, Zerbini EJ. Anatomic correction of transposition of the great vessels. *J Thorac Cardiovasc Surg*. 1976;72:364–370. doi: 10.1016/S0022-5223(19)40063-9

SUPPLEMENTAL MATERIAL

Table S1. HPO terms in our cohort

HPO number	Terms	Abbreviation
HP:0006704	Abnormal coronary artery morphology*	ACAM
HP:0011636	Abnormal coronary artery origin*	ACAO
HP:0001633	Abnormal mitral valve morphology*[§]	AMVM
HP:0001659	Aortic regurgitation	AR
HP:0001679	Aortopulmonary collateral arteries[†]	MAPCAs
HP:0001631	Atrial septal defect	ASD
HP:0001678	Atrioventricular block	AVB
HP:0001647	Bicuspid aortic valve	BAV
HP:0001640	Cardiomegaly	-
HP:0001680	Coarctation of aorta	COA
HP:0011641	Coronary-pulmonary artery fistula[§]	CPAF
HP:0011534	Criss-cross Heart[†]	CCH
HP:0001651	Dextrocardia	-
HP:0001669	Transposition of the great arteries	TGA
HP:0012304	Hypoplastic aortic arch	HAA
HP:0004383	Hypoplastic left heart	HLH
HP:0010954	Hypoplastic right heart	HRH
HP:0011573	Hypoplastic tricuspid valve[§]	HTV
HP:0011611	Interrupted aortic arch[§]	IAA
HP:0011670	Left superior vena cava draining to coronary sinus	LSVCC

HP:0011599	Mesocardia[§]	-
HP:0001653	Mitral regurgitation[§]	MR
HP:0001718	Mitral stenosis[§]	MS
HP:0005345	Obstruction of the superior vena cava[†§]	OSVC
HP:0002623	Overriding aorta[§]	OA
HP:0010773	Partial anomalous pulmonary venous return	PAPVC
HP:0001643	Patent ductus arteriosus	PDA
HP:0001655	Patent foramen ovale	POF
HP:0005301	Persistent left superior vena cava	LSVC
HP:0004927	Pulmonary artery dilatation[§]	PAD
HP:0004415	Pulmonary artery stenosis	PS
HP:0010444	Pulmonary insufficiency	PI
HP:0001642	Pulmonic stenosis	PVS
HP:0001646	Quadricuspid aortic valve[†§]	QAV
HP:0012020	Right aortic arch	RAA
HP:0001705	Right ventricular outlet tract obstruction	RVOTO
HP:0011640	Single coronary artery origin	SCA
HP:0005160	Total anomalous pulmonary venous return[§]	TAPVC
HP:0011662	Tricuspid atresia	TA
HP:0001702	Overriding tricuspid valve[‡]	OTV
HP:0005180	Tricuspid regurgitation	TR
HP:0001629	Ventricular septal defect	VSD

*These terms are superior terms of a group of phenotypes description that cannot be described in a standard way

†These terms are traced to their superior terms for their absence in R package "OntologySimilarity", but they can be annotated to HPO.

‡This term is traced to its superior term for the absence in R package "OntologySimilarity", and it cannot be annotated to HPO.

§These terms were present in only one patient.

Table S2. Phenotype combination of three clusters

Combination	Number of cases
Cluster 1	
ASD+PDA	67
VSD+ASD+PDA	36
ASD	32
PDA+POF	29
VSD+PDA+POF	21
VSD+POF	18
VSD+ASD	16
ASD+PDA+ACAO	11
ASD+PDA+SCA	10
VSD	10
ASD+PDA+Cardiomegaly	9
VSD+ASD+PDA+ACAO	9
VSD+ASD+PDA+PVS	8
VSD+PDA	8
VSD+PDA+PVS+POF	8
VSD+ASD+ACAO	6
VSD+ASD+PDA+Cardiomegaly	6
VSD+PDA+POF+ACAO	6
ASD+Cardiomegaly	5

ASD+PDA+POF	5
PDA+POF+ACAM	5
PDA+POF+SCA	5
VSD+ASD+PDA+ACAM	5
VSD+PDA+POF+Cardiomegaly	5
ASD+PDA+ACAM	4
PDA+POF+ACAO	4
VSD+ASD+Cardiomegaly	4
VSD+ASD+SCA	4
VSD+PDA+Cardiomegaly	4
VSD+POF+ACAO	4
VSD+ASD+PDA+POF	3
VSD+PDA+POF+SCA	3
VSD+POF+Cardiomegaly	3
PDA+POF+Cardiomegaly	2
POF	2
POF+Cardiomegaly	2
VSD+ASD+ACAM	2
VSD+PDA+ACAO	2
VSD+PDA+POF+HAA	2
VSD+PDA+POF+ACAM	2
VSD+POF+ACAM	2

VSD+POF+SCA	2
ASD+PDA+Dextrocardia	1
ASD+PDA+LSVCC	1
ASD+PDA+LSVC	1
ASD+PDA+TR	1
ASD+PDA+ACAO+ACAM	1
ASD+PDA+POF+SCA	1
PDA+POF+Dextrocardia	1
PDA+POF+PS	1
PDA+POF+LSVC	1
VSD+ASD+PDA+HAA+COA	1
VSD+ASD+PDA+RVOTO	1
VSD+ASD+PDA+PS	1
VSD+ASD+PDA+RAA	1
VSD+ASD+PDA+SCA	1
VSD+ASD+PDA+SCA+COA	1
VSD+ASD+PDA+ACAO+HAA	1
VSD+ASD+PDA+ACAO+PS	1
VSD+ASD+PDA+Cardiomegaly+COA	1
VSD+ASD+PDA+Cardiomegaly+ACAO	1
VSD+ASD+PDA+Cardiomegaly+SCA	1
VSD+ASD+PDA+POF+ACAM	1

VSD+ASD+POF	1
VSD+PDA+IAA	1
VSD+PDA+Dextrocardia	1
VSD+PDA+SCA	1
VSD+PDA+POF+AR	1
VSD+PDA+POF+RVOTO+CCH	1
VSD+PDA+POF+Dextrocardia+LSVCC	1
VSD+PDA+POF+Cardiomegaly+ACAM	1
VSD+PDA+POF+LSVC	1
VSD+PDA+POF+SCA+COA	1
Cluster 2	
VSD+PVS	34
VSD+ASD+PVS	18
VSD+PVS+POF	11
VSD+PDA+PVS	6
VSD+PVS+MAPCAs	5
VSD+PVS+Cardiomegaly	4
VSD+Cardiomegaly	3
VSD+ASD+PVS+RAA	3
ASD+PVS	2
VSD+ASD+PS	2
VSD+ASD+PDA+PVS+LSVC	2

VSD+ASD+PDA+PVS+Cardiomegaly	2
VSD+ASD+PVS+HRH+OTV	2
VSD+ASD+PVS+ACAO	2
VSD+ASD+PVS+Cardiomegaly	2
VSD+ASD+PVS+MAPCAs	2
VSD+PDA+PVS+Cardiomegaly	2
VSD+PDA+PVS+POF+SCA	2
VSD+PVS+ACAM	2
VSD+PVS+ACAM+Dextrocardia	2
VSD+PVS+LSVC	2
VSD+PVS+ACAO	2
VSD+PVS+POF+Cardiomegaly	2
VSD+PVS+POF+Cardiomegaly+SCA	2
VSD+PVS+POF+ACAM	2
ASD+PAPVR	1
ASD+PS	1
ASD+TR	1
ASD+TR+MR	1
ASD+PDA+PVS+TR	1
ASD+PDA+PVS+ACAM	1
ASD+PDA+PVS+Cardiomegaly	1
PVS+POF+ACAM	1

PVS+POF+Cardiomegaly	1
VSD+MS	1
VSD+AR	1
VSD+Cardiomegaly+TR	1
VSD+ASD+PI	1
VSD+ASD+Dextrocardia+LSVCC	1
VSD+ASD+LSVC	1
VSD+ASD+PDA+PVS+AMVM	1
VSD+ASD+PDA+PVS+PAPVR	1
VSD+ASD+PDA+PVS+HRH	1
VSD+ASD+PDA+PVS+PS	1
VSD+ASD+PDA+PVS+RAA	1
VSD+ASD+PDA+PVS+MAPCAs+Mesocardia	1
VSD+ASD+PDA+PVS+POF+ACAO	1
VSD+ASD+PDA+PVS+SCA	1
VSD+ASD+PDA+PVS+SCA+LSVC+RAA	1
VSD+ASD+PVS+HTV	1
VSD+ASD+PVS+BAV	1
VSD+ASD+PVS+AR+QAV	1
VSD+ASD+PVS+OTV	1
VSD+ASD+PVS+TR	1
VSD+ASD+PVS+ACAM	1

VSD+ASD+PVS+HRH	1
VSD+ASD+PVS+HRH+BAV	1
VSD+ASD+PVS+HRH+TA	1
VSD+ASD+PVS+LSVC	1
VSD+ASD+PVS+LSVC+RVOTO	1
VSD+ASD+PVS+PS	1
VSD+ASD+PVS+ACAO+LSVC+RVOTO	1
VSD+ASD+PVS+ACAO+MAPCAs	1
VSD+ASD+PVS+Cardiomegaly+RVOTO	1
VSD+ASD+PVS+Cardiomegaly+HRH	1
VSD+ASD+PVS+Cardiomegaly+LSVCC	1
VSD+ASD+PVS+Cardiomegaly+ACAM	1
VSD+ASD+PVS+Cardiomegaly+SCA	1
VSD+ASD+PVS+Cardiomegaly+SCA+TR+PAD	1
VSD+ASD+PVS+MAPCAs+CPAF	1
VSD+ASD+PVS+MAPCAs+LSVC	1
VSD+ASD+PVS+MAPCAs+RAA	1
VSD+ASD+PVS+POF+AVB	1
VSD+ASD+PVS+POF+MAPCAs	1
VSD+ASD+PVS+POF+ACAM	1
VSD+ASD+PVS+SCA+LSVC	1
VSD+PDA+Dextrocardia+PI	1

VSD+PDA+PVS+TA	1
VSD+PDA+PVS+ACAM	1
VSD+PDA+PVS+PS	1
VSD+PDA+PVS+ACAO	1
VSD+PDA+PVS+Cardiomegaly+SCA	1
VSD+PDA+PVS+MAPCAs	1
VSD+PDA+PVS+POF+LSVCC	1
VSD+PDA+PVS+POF+RAA	1
VSD+PDA+PVS+POF+ACAM	1
VSD+PDA+PVS+POF+ACAO	1
VSD+PDA+PVS+POF+Cardiomegaly	1
VSD+PDA+PVS+POF+Cardiomegaly+OTV	1
VSD+PDA+PVS+POF+Cardiomegaly+PS	1
VSD+PDA+PVS+POF+Cardiomegaly+ACAO	1
VSD+PDA+PVS+POF+Cardiomegaly+SCA+CCH	1
VSD+POF+PS	1
VSD+POF+LSVC	1
VSD+PVS+CCH	1
VSD+PVS+ACAM+AVB	1
VSD+PVS+ACAM+MAPCAs	1
VSD+PVS+Dextrocardia	1
VSD+PVS+Dextrocardia+CCH	1

VSD+PVS+RAA	1
VSD+PVS+ACAO+RAA	1
VSD+PVS+Cardiomegaly+PI	1
VSD+PVS+Cardiomegaly+ACAM+MAPCAs	1
VSD+PVS+Cardiomegaly+ACAM+TR	1
VSD+PVS+Cardiomegaly+ACAO+MAPCAs	1
VSD+PVS+Cardiomegaly+SCA+AVB	1
VSD+PVS+Cardiomegaly+SCA+RAA	1
VSD+PVS+MAPCAs+RAA	1
VSD+PVS+POF+ACAO	1
VSD+PVS+POF+Cardiomegaly+ACAM	1
VSD+PVS+POF+SCA	1
VSD+PVS+POF+SCA+RAA	1
VSD+PVS+SCA	1
Cluster 3	
ASD+SCA	6
ASD+ACAO	3
POF+ACAM	2
VSD+SCA	2
VSD+ASD+Cardiomegaly+ACAO	2
VSD+POF+Cardiomegaly+ACAO	2
ASD+LSVCC+OSVC	1

ASD+ACAM	1
ASD+ACAO+TAPVR	1
ASD+ACAO+LSVC	1
ASD+MAPCAs	1
ASD+Cardiomegaly+SCA+HLH	1
ASD+PDA+SCA+RAA+RVOTO	1
ASD+PVS+LSVC+Dextrocardia+HLH	1
ASD+PVS+ACAO+LSVC+HLH	1
PDA+Cardiomegaly	1
POF+ACAO	1
VSD+ACAO	1
VSD+ASD+Cardiomegaly+ACAM	1
VSD+ASD+Cardiomegaly+ACAM+TR	1
VSD+ASD+Cardiomegaly+SCA	1
VSD+ASD+PDA+Cardiomegaly+SCA+LSVC+HRH	1
VSD+ASD+SCA+PS	1
VSD+PDA+SCA+OA	1
VSD+PDA+Cardiomegaly+ACAM	1
VSD+PDA+Cardiomegaly+SCA	1
VSD+POF+SCA+LSVCC	1
VSD+POF+SCA+PS	1

Table S3. Other definitive surgeries

Operation name	Number of cases
Rastelli procedure	17
Fontan procedure	16
Modified Blalock-Taussig shunts procedure	13
REV procedure	8
Modified REV procedure	8
Bidirectional Glenn	8
Senning procedure	5
Nikaidoh procedure	3
Pulmonary banding	2
Blalock-Taussig shunts procedure	1
Bilateral bidirectional Glenn	1

Table S4. Survival rate

	30 days	1 year	5 years	10 years
Overall	98.20% (97.19%-99.21%)	95.68% (94.12%-97.26%)	94.70% (92.95%-96.49%)	93.13% (90.83%-95.49%)
Cluster 1	98.34% (97.13%-99.57%)	96.15% (94.32%-98.02%)	95.46% (93.41%-97.55%)	94.08% (91.33%-96.92%)
Cluster 2	98.05% (96.17%-99.96%)	95.38% (92.47%-98.38%)	93.67% (90.26%-97.21%)	91.33% (86.76%-96.14%)
Cluster 3	97.44% (92.60%-100%)	92.16% (84.02%-100%)	92.16% (84.02%-100%)	92.16% (84.02%-100%)
ASO	98.21% (97.06%-99.38%)	95.96% (94.24%-97.71%)	95.66% (93.84%-97.50%)	95.04% (92.88%-97.25%)
DRT	100% (100%-100%)	97.33% (93.75%-100%)	92.95% (87.17%-99.12%)	92.95% (87.17%-99.12%)
Other	96.34% (92.36-100%)	92.26% (86.49%-98.42%)	90.84% (84.58%-97.57%)	80.86% (69.45%-94.15%)

Abbreviations: ASO, arterial switch operation; DRT, double root translocation.

Table S5. Univariable analysis for outcomes in the overall cohort.

Variables	Death		Reintervention	
	Hazard ratio	P value	Hazard ratio	P value
Univariate				
Baseline characteristics				
Age at definitive surgery	1.057 (0.982-1.137)	0.140	1.057 (0.969-1.154)	0.212
Female	1.498 (0.766-2.928)	0.238	1.170 (0.538-2.544)	0.692
BMI	0.979 (0.855-1.122)	0.764	1.120 (1.000-1.255)	0.051
Associated anomalies				
ASD	1.636 (0.837-3.200)	0.150	0.468 (0.226-0.972)	0.042*
VSD	1.001 (0.512-1.958)	0.997	3.374 (1.180-9.650)	0.023*
PDA	0.506 (0.262-0.979)	0.043*	0.435 (0.205-0.924)	0.030*
POF	0.645 (0.296-1.408)	0.271	1.331 (0.637-2.784)	0.447

PVS	1.857 (0.979-3.522)	0.058	5.259 (2.471-11.192)	<0.001*
ACAO	0.457 (0.110-1.899)	0.281	0.041 (0-4.813)	0.189
SCA	1.058 (0.375-2.986)	0.916	0.893 (0.270-2.950)	0.852
Other coronary artery abnormality	1.620 (0.575-4.564)	0.362	0.456 (0.062-3.347)	0.440
Cardiomegaly	1.079 (0.451-2.586)	0.864	0.843 (0.318-2.233)	0.731
Surgery				
ASO	Reference		Reference	
DRT	1.377 (0.523-3.625)	0.517	7.489 (3.436-16.322)	<0.001*
Other	2.803 (1.333-5.893)	0.007*	2.598 (0.924-7.309)	0.070
Surgery				
ASO	0.726 (0.276-1.912)	0.517	0.134 (0.061-0.291)	<0.001*
DRT	Reference		Reference	
Other	2.036 (0.695-5.963)	0.195	0.347 (0.122-0.989)	0.047*

*Indicated statistical significance

Abbreviations: ACAO, Abnormal coronary artery origin; ASD, Atrial septal defect; ASO, Arterial switch operation; DRT, Double root translocation; PDA, Patent ductus arteriosus; POF, Patent foramen ovale; PVS, pulmonic stenosis; SCA, Single coronary artery origin; VSD, ventricular septal defect.

Table S6. Different phenotypic distribution between two subgroups of cluster 1

Anomalies	Subgroup 1 (n=265)	Subgroup 2 (n=157)	P-value
Coronary artery abnormality			
ACAO	0	46 (29.30%)	<0.001*
SCA	0	30 (19.11%)	<0.001*
Others	0	23 (14.65%)	<0.001*
Cardiomegaly	0	44 (28.03%)	<0.001*
PVS	16 (6.04%)	0	<0.001*
HAA	0	4 (2.55%)	0.019*
CoA	0	4 (2.55%)	0.019*
Dextrocardia	0	4 (2.55%)	0.019*

Abbreviations: ACAO, Abnormal coronary artery origin; CoA, Coarctation of aorta; HAA, Hypoplastic aortic arch; PVS, Pulmonic stenosis; SCA, Single coronary artery.

Table S7. Results of multivariate Cox proportional analysis regarding subgroup of cluster 1

Variables	Death		Reintervention	
	Hazard ratio	P value	Hazard ratio	P value
Subgroup 1				
Age at definitive surgery	1.303 (1.077-1.577)	0.006*	1.141 (0.761-1.710)	0.523
Female	1.719 (0.572-5.167)	0.335	0.488 (0.058-4.115)	0.510
PDA	0.920 (0.267-3.168)	0.895	3.038 (0.329-28.053)	0.327
ASD	1.549 (0.416-5.761)	0.514	1.285 (0.287-5.754)	0.743
VSD	0.850 (0.244-2.960)	0.799	1.279 (0.248-6.612)	0.769
PVS	1.414 (0.060-33.132)	0.830	0 (0-Inf)	0.964
Surgery				
ASO	Reference		Reference	
DRT	1.061 (0.031-36.119)	0.974	$5.617e^{4}$ (0-Inf)	0.958

Other	1.126 (0.053-23.936)	0.939	0 (0-Inf)	0.972
Subgroup 2				
Age at definitive surgery	1.343 (0.933-1.933)	0.112	0.653 (0.008-56.525)	0.851
Female	2.474 (0.381-16.054)	0.343	0 (0-Inf)	0.983
PDA	0.201 (0.030-1.334)	0.097	$9.078e^{4}$ (0-Inf)	0.985
ASD	1.291 (0.189-8.799)	0.794	0 (0-Inf)	0.978
VSD	1.794 (0.184-17.455)	0.615	$1.347e^{5}$ (0-Inf)	0.980
SCA	0 (0-Inf)	0.985	0 (0-Inf)	0.984

*Indicated statistical significance

Abbreviations: ASD, Atrial septal defect; ASO, Arterial switch operation; DRT, Double root translocation; PDA, Patent ductus arteriosus; PVS, pulmonic stenosis; SCA, Single coronary artery origin; VSD, ventricular septal defect.

Table S8. Results of multivariate Cox proportional analysis regarding sex

Variables	Death		Reintervention	
	Hazard ratio	P value	Hazard ratio	P value
Female				
Age at definitive surgery	1.162 (0.949-1.423)	0.145	0.529 (0.284-0.986)	0.045*
PDA	0.553 (0.153-2.000)	0.367	0.178 (0.019-1.644)	0.128
ASD	1.799 (0.474-6.830)	0.388	0.735 (0.159-3.406)	0.694
VSD	0.458 (0.101-2.078)	0.312	0.826 (0.044-15.324)	0.898
PVS	2.533 (0.408-15.712)	0.318	1.909 (0.111-32.688)	0.656
SCA	0.698 (0.088-5.545)	0.734	1.625 (0.162-16.307)	0.680
Surgery				
ASO	Reference		Reference	
DRT	0.464 (0.059-3.652)	0.465	37.507 (2.510-560.373)	0.009*
Other	0.094 (0.004-2.062)	0.133	1.452 (0.100-21.063)	0.785

Male				
Age at definitive surgery	0.940 (0.807-1.094)	0.422	0.985 (0.854-1.136)	0.834
PDA	0.704 (0.292-1.701)	0.436	0.922 (0.369-2.304)	0.863
ASD	1.530 (0.649-3.604)	0.331	0.691 (0.278-1.717)	0.426
VSD	0.775 (0.285-2.106)	0.618	1.324 (0.321-5.473)	0.698
PVS	0.927 (0.218-3.947)	0.918	2.442 (0.575-10.367)	0.226
SCA	1.659 (0.485-5.680)	0.420	1.104 (0.253-4.825)	0.895
Surgery				
ASO	Reference		Reference	
DRT	2.248 (0.374-13.491)	0.376	2.221 (0.525-9.386)	0.278
Other	6.980 (1.589-30.651)	0.010*	1.642 (0.334-8.080)	0.542

*Indicated statistical significance

Abbreviations: ASD, Atrial septal defect; ASO, Arterial switch operation; DRT, Double root translocation; PDA, Patent ductus arteriosus; PVS, pulmonic stenosis; SCA, Single coronary artery origin; VSD, ventricular septal defect.

Figure S1. Subgraphs of the HPO terms for each cluster. Three tree plots in different colors represent the phenotypic distribution of three clusters (A for cluster 1, B for cluster 2, and C for cluster 3). The arrows represent the affiliation relationship ('is a') between terms in the ontology. The shade of color indicated the frequency of the specific term in HPO database with the referenced color key on the top of each panel. HPO, human phenotype ontology.

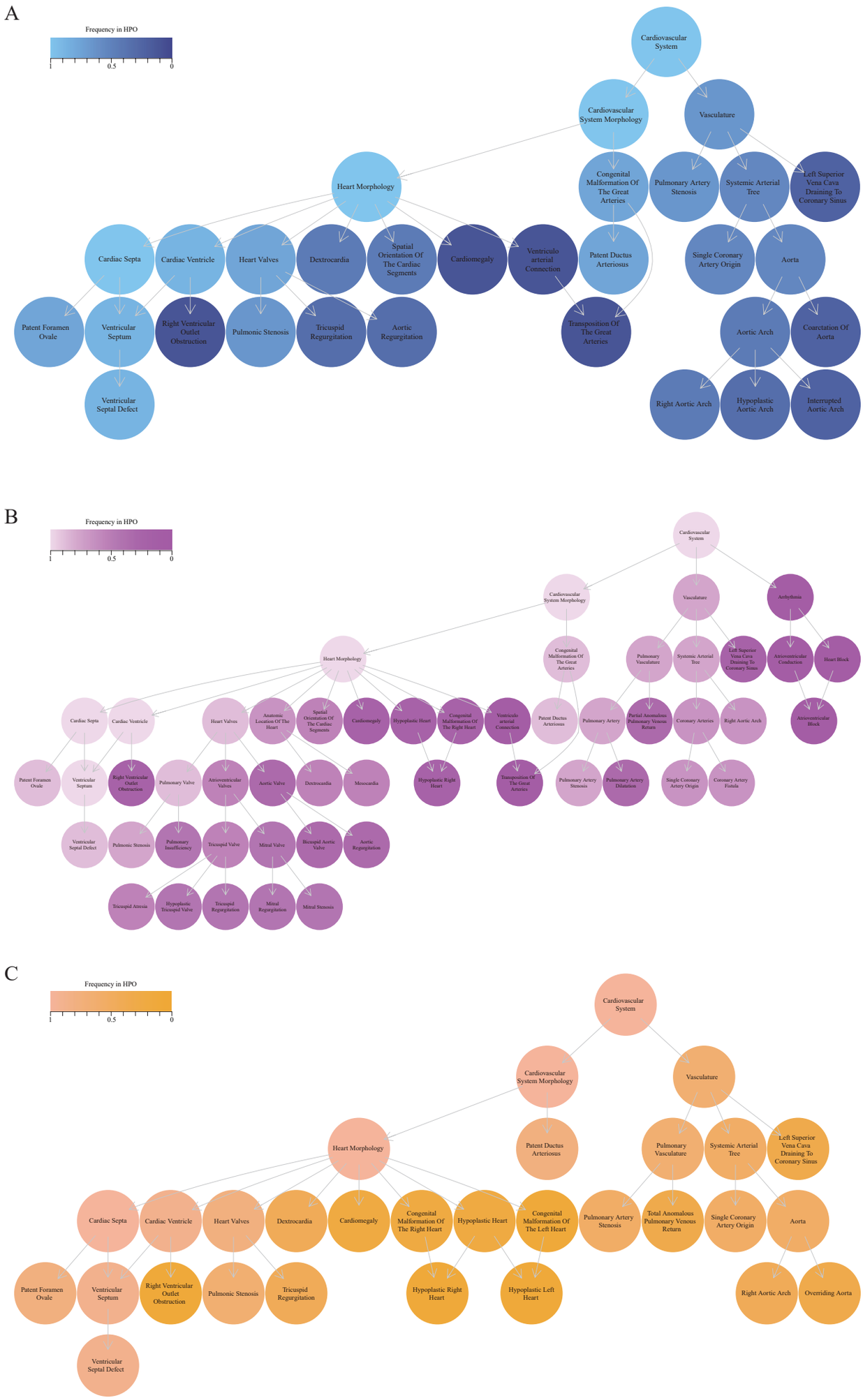

Figure S2. Kaplan-Meier curve for outcomes classified by different surgical strategies. A) Kaplan-Meier curve showed the overall survival rate among three different surgical strategies and the statistical difference could be observed (P=0.018). B) Kaplan-Meier curve showed the freedom from reintervention among three different surgical strategies and a significant difference could be observed (P<0.001).

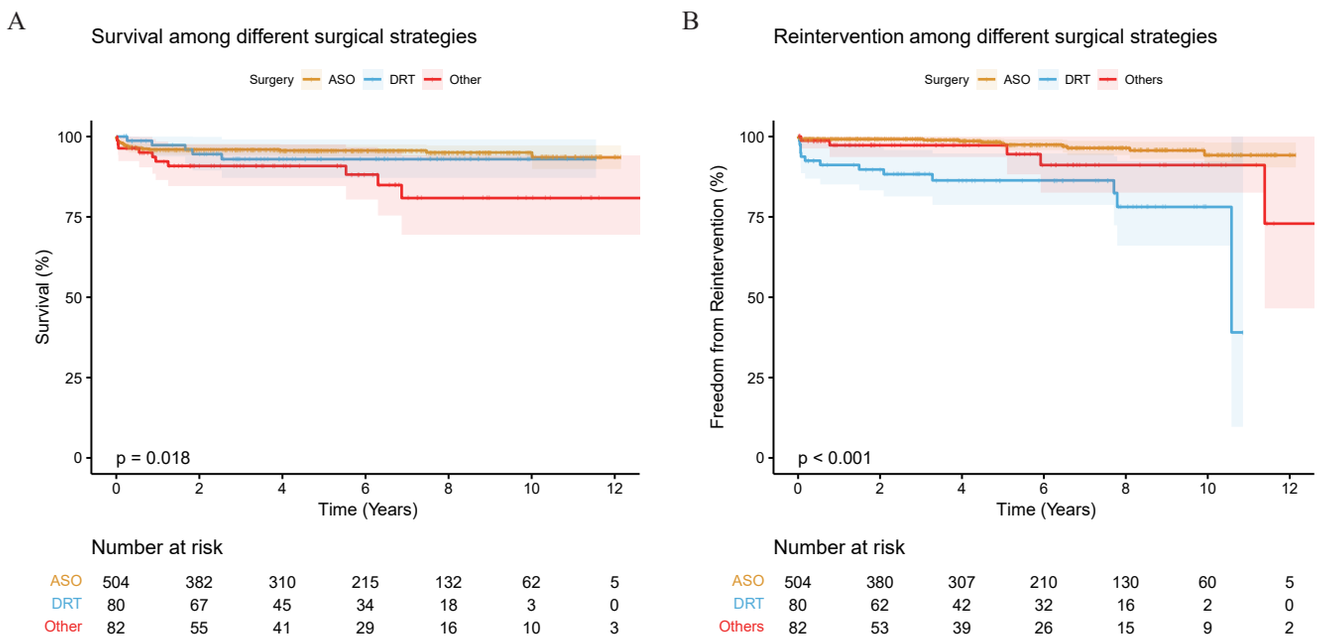

Figure S3. Kaplan-Meier curve for ASO outcomes classified by clusters. A) Kaplan-Meier curve showed the overall survival rate of patients receiving ASO among three clusters and no significance could be observed (P=0.620). B) Kaplan-Meier curve showed the freedom from reintervention of patients receiving ASO among three clusters and significance could be observed (P=0.019). ASO, arterial switch operation.

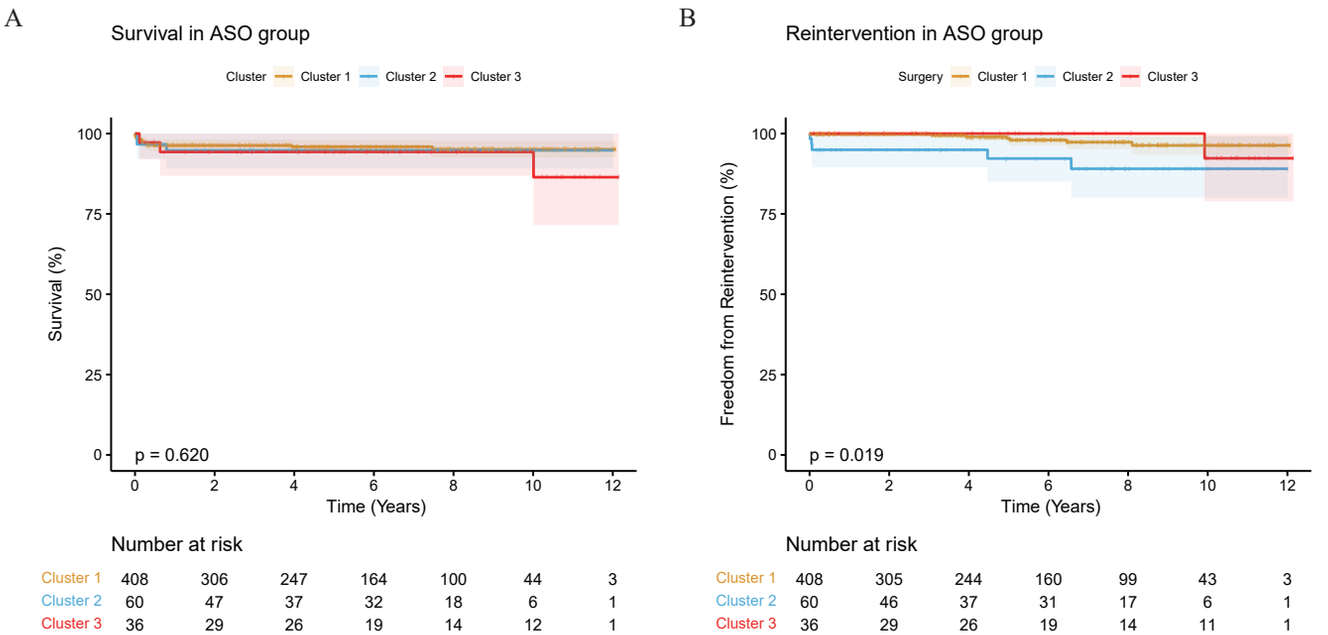

Figure S4. Kaplan-Meier curve for outcomes in subgroups of cluster 1. Kaplan-Meier curve showed no significant outcomes difference between subgroups of cluster 1 (P=0.250 for survival and P=0.090 for reintervention).

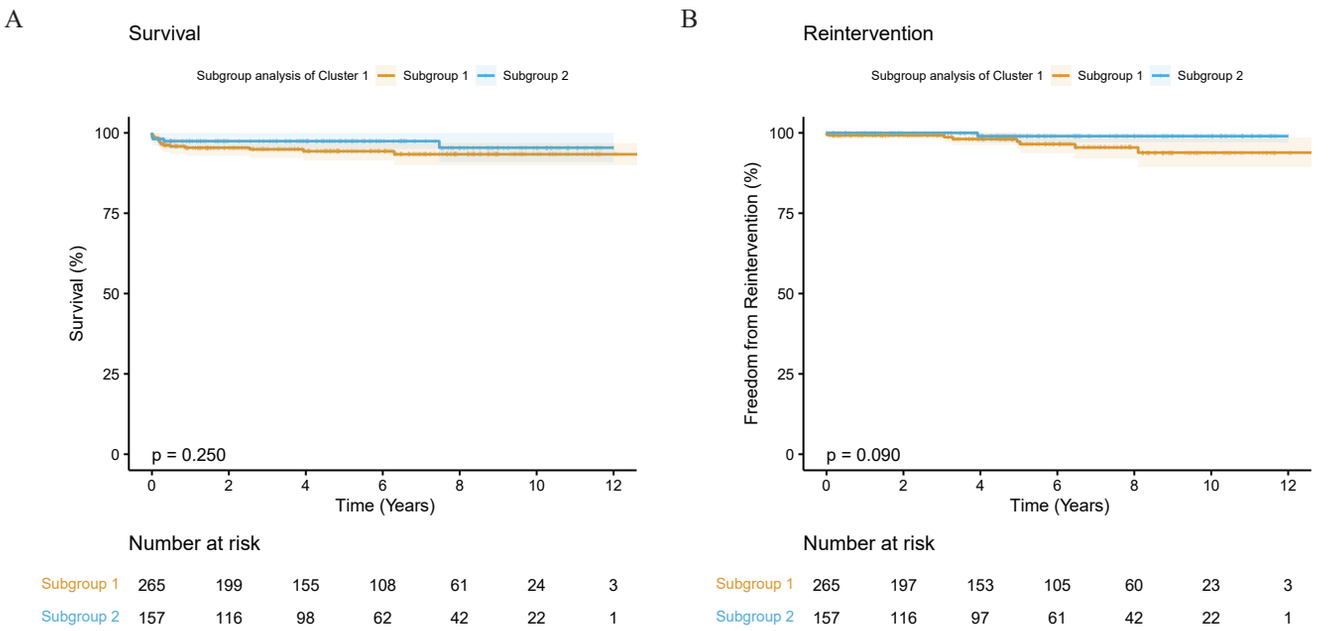

Figure S5. Kaplan-Meier curve for outcomes regarding sex Kaplan-Meier curve revealed no significant difference between male and female patients (P=0.230 for survival and P=0.690 for reintervention).

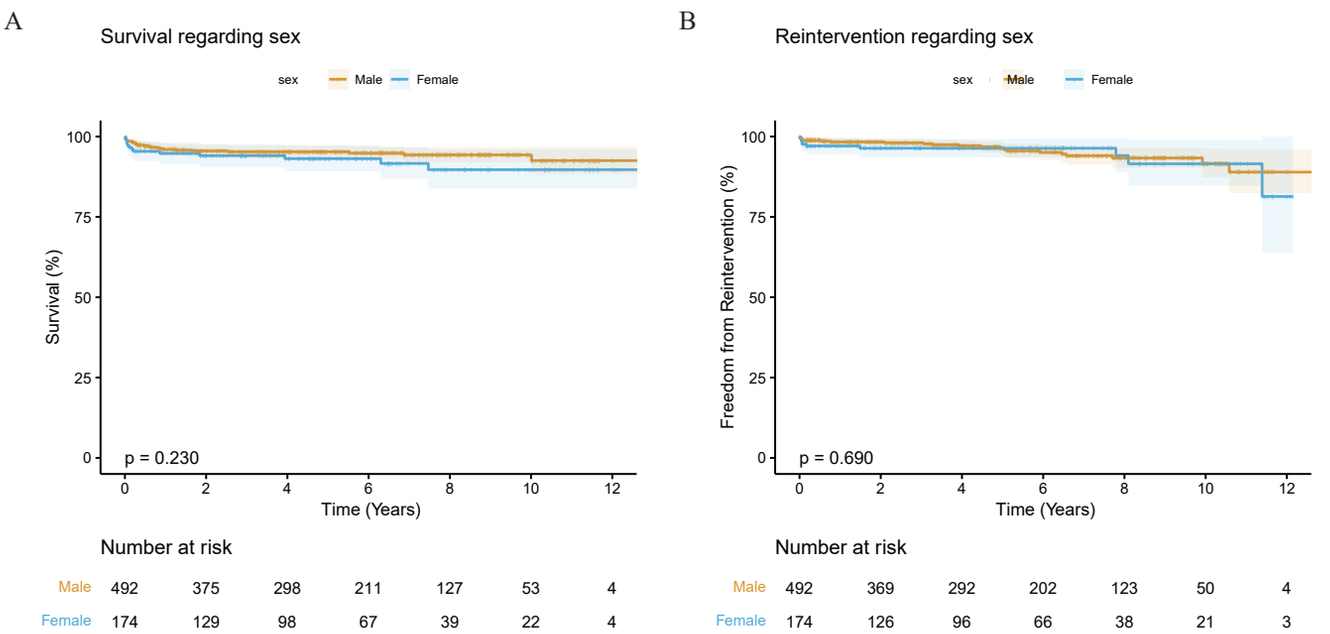

Figure S6. Kaplan-Meier curve for prognosis when considering both our novel clustering and conventional classification.