

# Sparse Representation Based Genetic Biomarker Evaluation for Congenital Heart Defects

Peng Zhou<sup>1</sup>, Benjamin H Lehrman<sup>2</sup>, Hongping Zheng<sup>3\*</sup>

<sup>1</sup> Department of Biomedical Engineering, Tianjin University, Tianjin 300072, China;

<sup>2</sup> Rush Medical College, 600 S Paulina St, Chicago, IL 60612, USA;

<sup>3</sup> Laboratory of Human Carcinogenesis, NCI, NIH, 9000 Rockville Pike, Bethesda, MD 20892, USA.

**\*Corresponding Author:** Hongping Zheng, Ph.D, Laboratory of Human Carcinogenesis, NCI, NIH, 9000 Rockville Pike, Bethesda, MD 20892, USA. Email: hongping.zheng@nih.gov; Tel: 301-496-7279.

## ABSTRACT

**Background:** Congenital heart defects (CHD) are the most common type of birth defect, affecting approximately 8 in 1,000 newborns. Hundreds of genes have been reported as CHD candidate genes. Nevertheless, each patient/patient group may demonstrate unique etiologic characteristics requiring personalized treatment.

**Methods:** We proposed a sparse representation-based variable selection (SRVS) approach to select disease-related genetic markers from a huge disease candidate gene pool acquired from ResNet relation database. The proposed approach was used to evaluate 167 CHD candidate genes and was followed by validation on a microarray expression data set. Pathway enrichment analysis (PEA), sub-network enrichment analysis (SNEA), and network connectivity analysis (NCA) were conducted to study the functional profile of the variables selected by SRVS and compare them with previous reported genetic markers.

**Results:** A significantly high disease prediction accuracy of 81.40 % was obtained (permutation  $p$ -value < 0.0002) using the top 24 SRVS-selected genes, which had been enriched within multiple pathways and sub-networks that had been previously implicated with CHD. Using the most frequently reported genes out of the 167 CHD candidate genes, the highest accuracy of 69.77 % was obtained with a permutation  $p$ -value = 0.017. Enrichment analysis and NCA showed that the top genes selected by the proposed SRVS approach were strongly related to the frequently reported CHD genes, although functional differences were present.

**Conclusion:** Our study suggests that SRVS is an effective method



<http://mo.qingres.com>

 OPEN ACCESS

DOI: 10.20900/mo.20160016

Received: May 15, 2016

Accepted: July 9, 2016

Published: August 25, 2016

**Copyright:** ©2016 Cain *et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

in data driven variable selection for CHD and that frequently reported CHD candidate genes may not be the best biomarkers for a specific CHD patient/patient group.

**Key Words:** Congenital heart defects; Sparse Representation; Variables selection; ResNet Database

## 1 INTRODUCTION

Congenital heart defects (CHD) are heart and large blood vessel anatomical abnormalities occurring during embryonic development<sup>[1]</sup>. CHD can result from genetic factors, or environmental factors; however, a combination of both is typical<sup>[2, 3]</sup>. Recently, there have been an increased number of articles reporting hundreds of genes/proteins related to CHD, many of which were suggested as candidate genes for the disease. However, every patient/patient group has unique human genome variations that requires treatment based on their predicted response or risk of disease<sup>[4]</sup>.

Sparse representation has recently received significant attention in applications such as signal recovery and significant components identification<sup>[5, 6]</sup>. However, in the cases of large variable and small sample number applications, specific modulation is required to fulfil the variable selected task. In many biomedical problems (e.g., genomic data, image data) samples are far outnumbered by variables.

This study proposes a sparse representation-based variable selection (SRVS) algorithm for selecting significant biomarkers at different detection resolutions, which has previously been effective in selecting variables from SNP data and fMRI data<sup>[7]</sup>. Instead of selecting a specific number of variables, this data-driven method ranks all variables by generating a sparse regression weight for each of them<sup>[7]</sup>.

## 2 MATERIALS AND METHODS

In this section, the proposed SRVS algorithm (Section 2.1) is described and then applied to a CHD candidate genetic biomarker selection problem (Section 2.2). Finally SRVS-selected variables are studied in terms of pathway enrichment analysis (PEA), sub-network enrichment analysis (SNEA), and network connectivity analysis (NCA) (Section 2.3).

### 2.1 SRVS algorithm

In general, a sparse representation model can be presented as Eq. (1).

$$y = x\delta + \varepsilon \quad (1)$$

where  $y \in R^{n \times 1}$  is the observation vector;  $x \in R^{n \times p}$  are the data measurements and  $p \geq n$ .  $\varepsilon \in R^{n \times 1}$  is the noise-caused measurement error. The goal is to reconstruct the unknown vector  $\delta \in R^{p \times 1}$  based on  $y$  and  $x$ .

To best approximate  $y$  by choosing a small number of non-zeros entries of  $\delta$  for the model given by Eq. (1), we consider the following  $L_p$  minimization problem (P0):

$$(P0) \min \|\delta\|_p \text{ subject to } \|y - x\delta\|^2 \leq \varepsilon \quad (2)$$

where  $\|\cdot\|_p$  is the  $L_p$  norm, and  $p \in [0, 1]$ . The following algorithm is designed to solve the minimization problem (P0) given by Eq. (2) and detect the columns of  $x$  relevant to  $y$ .

#### SRVS Algorithm:

1. Initial  $\delta(0) = 0$ ;
2. For the Step  $i$ , randomly choose  $k$  columns from  $x = \{x_1, \dots, x_p\}$  to construct a  $n \times k$  sub-matrix denoted as  $x_i \in R^{n \times k}$ ; and mark the selected columns' indexes as  $I_i \in R^{1 \times k}$ ;
3. Solve the following  $L_p$  minimization problem to find the optimal sparse solution  $\delta_i \in R^{k \times 1}$ :
 
$$\min \|\delta_i\|_p \text{ s.t. } \|y - x_i\delta_i\|^2 \leq \varepsilon \quad (3)$$
4. Update  $\delta^{(i)}$  with  $\delta_i$ :  $\delta^{(i)}(I_i) = \delta^{(i-1)}(I_i) + \delta_i$ ; where  $\delta^{(i)}(I_i)$  and  $\delta^{(i-1)}(I_i)$  denote the  $I_i$ th entries in  $\delta^{(i)}$  and  $\delta^{(i-1)}$  respectively;
5. If  $\|\delta^{(i)}/I_i - \delta^{(i-1)}/(I_i-1)\|^2 > a$ , where  $a$  is a predefined constant, update  $i = i+1$ , and go to Step 2. Otherwise, set  $\delta = \delta^{(i)}/I_i$ . The non-zero entries in  $\delta$  correspond to the column vectors selected.

In Step 3, there are many proposed methods for solving the  $L_p$  minimization problem, such as the Homotopy method<sup>[8]</sup> for  $p = 1$ , and the orthogonal matching pursuit (OMP) algorithm<sup>[9]</sup> for  $p = 0$ .

2.2 CHD Candidate Genes for Evaluation

A 167-CHD-candidate gene pool was acquired from the CHD-Gene relation data, which was also included in a gene expression data set (GEO: GSE34457). CHD-Gene relation data was acquired from Pathway Studio (PS) ResNet 11 Mammalian database updated July 2016. The ResNet® Mammalian database is one part of PS ResNet Databases, which are a group of real-time update network databases, including curated signaling, cellular process, and metabolic pathways, ontologies, and annotations, as well as molecular interactions and functional relationships extracted from 35M+ references covering the entire PubMed abstract and Elsevier full text journals. More information about the PS ResNet Mammalian databases is available at <http://pathwaystudio.gousinfo.com/ResNetDatabase.html>.

The gene expression profile was acquired from the cell lines of 43 Down syndrome patients, among whom 21 were with CHD (case) and 22 without (controls). The original data includes 48,701 probes. All probes with null data entries were removed, resulting in an overlap of 167 genes with the CHD candidate gene pool.

2.3 Validation of the SRVS method

To test the validity of the proposed method, we studied the SRVS selected genes through four approaches: CHD predication, PEA, SNEA and NCA. For comparison purposes, we also compared the performance of the top frequently referenced CHD candidate genes.

2.3.1 Definition of two scores

The reference number underlying a gene-disease relationship is defined as the gene's reference score (Rscore). The SRVS approach generated weights for each gene is defined as the SRVS score (Sscore).

2.3.2 Validation using disease prediction

We hypothesize that significant CHD candidate genes/gene set should contribute to distinguishing CHD patients from healthy controls. A Euclidean, distance-based, multivariate classification<sup>[7]</sup> of the gene expression data set (GEO: GSE34457), followed by a leave-one-out (LOO) cross-validation, using the overall gene set and the sub-sets selected by Sscore and Rscore as tentative markers, was performed to validate selected gene effectiveness and the proposed SRVS approach. Next a 5,000 run permutation was conducted to test the hypothesis that a randomly-selected gene set of the same size can reach an equal, or higher, classification accuracy (CR).

2.3.3 Enrichment and connectivity analysis

To better understand the underlying functional profile of the genes selected by Sscore and Rscore, we also conducted a PEA and an SNEA on the top genes selected by the two scores. An NCA was also conducted on the gene subsets using Pathway Studio, to identify connectivity between given genes/proteins. The weight of an edge is the number of scientific references underlying a reported gene-gene interaction.

3 RESULTS

3.1 CHD candidate genes for validation

An analysis of CHD-Gene relation data revealed a CHD gene pool of 684 genes, supported by 3,237 articles (Supplementary Table S1). Of these 684 CHD candidate genes, 167 were evaluated using the proposed SRVS algorithm with an independent gene expression data (GEO: GSE34457). Fig. 1 presents the CHD candidate genes. The complete 167 gene list and related information, including Sscore and Rscore, is in Supplementary Table S2.

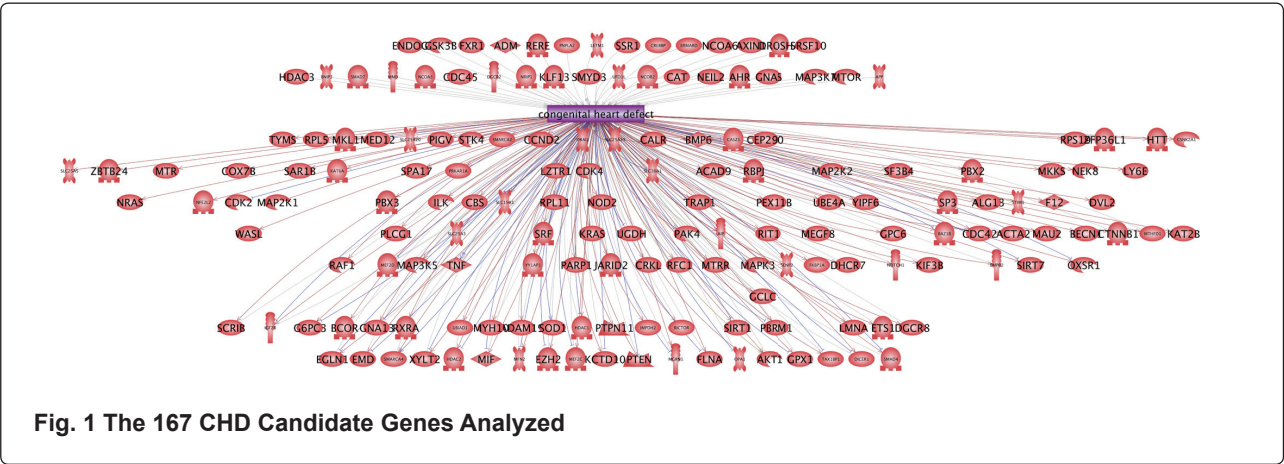


Fig. 1 The 167 CHD Candidate Genes Analyzed

3.2 Validation in CHD prediction

The effectiveness of the SRVS-generated metrics, the Sscore, was evaluated using a case/control classification and a LOO cross-validation on an RNA microarray dataset (GEO: GSE34457). This was followed by a 5,000 run permutation test. The Rscore was tested for comparison purposes.

In the LOO cross validation, the 167 genes were ranked by different metric scores, and then the top  $n$  ( $n = 1, 2, \dots$ ) genes were used as input variables for classification and LOO cross-validation. Fig. 2 presents the results with maximum classification ratios (CRs) marked at the position of corresponding number of genes.

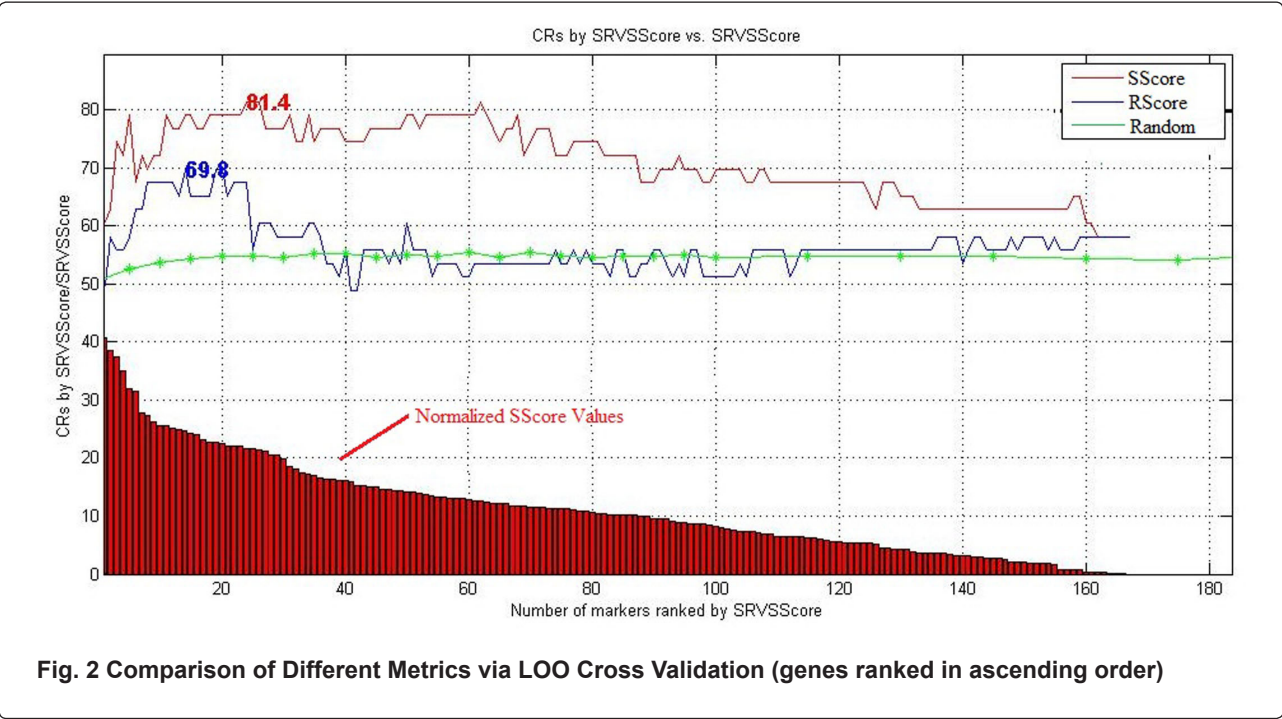


Table 1. LOO cross validation and permutation results

	Sscore	Rscore	167 Genes
MaxCRs	81.40	69.77	58.14
#Genes	24	14	167
pvalue	< 0.0002	0.017	0.30

Fig. 2 suggests that the top genes using both the Sscore and the Rscore leads to significant better classification accuracies as compared to CRs generated by randomly selected gene sets of same size. Moreover, the highest CRs were acquired by using only the top genes selected by different scores, (See Fig. 2 and Table 1). Adding more genes with lower scores do not help to get a better CR, suggesting the validity of both Sscore and Rscore.

To note, the Sscore led to much higher CRs

with lower permutation  $p$ -values, demonstrating the proposed method’s effectiveness. The top 24 Sscore are in Table 2. The top 24 Rscore genes are provided for comparison purposes. The full lists are in Supplementary Table S2.

Table 2. Top 24 Genes Reported Associations with CHD Ranked by Different Scores

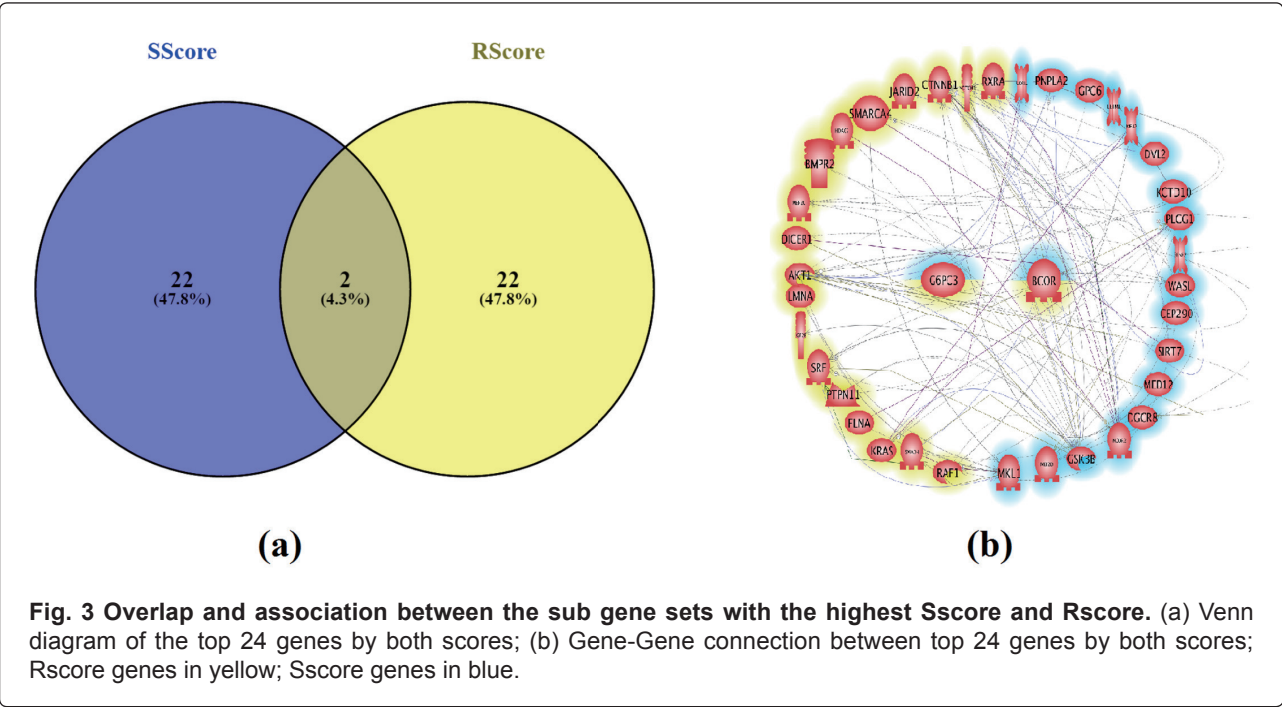
Genes by SScore	MEGF8; GSK3B; MED12; WASL; NCOR 2; SIRT7; PNPLA2; CEP290; MKL1; GPC6; PLCG1; MEF2D; ORAI1; KCTD10; DVL2; SENP2; DGCR8; ALG13; RPS19; BCOR; RERE; MFN2; G6PC3; LETM1
	PTPN11; NOTCH1; LMNA; G6PC3; MTRR; RXRA; SRF; BMPR2; HDAC2; SLC19A1; AKT1; KRAS; CTNNB1; DICER1; RAF1; FLNA; RPL5; MEF2C; BCOR; JARID2; SMARCA4; NEK8; IGF2R; SMAD4
Genes by RScore	



3.3 Compare Top Genes by Sscore and Rscore

To better understand the profile of the genes selected by SRVS approach, the two groups of top genes selected by Sscore and Rscore (Table 2) were compared using PEA and NCA approaches. A two-gene overlap: BCOR and G6PC3 (Fig. 3 (a)) was

identified. Nevertheless, a strong relation between the two gene groups is suggested by an NCA analysis, which detected 94 relations between 18/24 Sscore genes and 18/24 Rscore genes (Fig. 3 (b)), supported by more than 3,000 references (Supplementary Table S3).



3.4 Enrichment Analysis

PEA and SNEA results for the different groups listed in Table 3 are presented in this section. At the same enrichment *p*-value threshold (< 0.001), 152 pathways are enriched by the top 24 Rscoree

genes, while only 17 enriched pathways for the Sscore group were identified. The top 10 pathways/ gene sets, by different scores, appear in Table 3. Complete results are in Supplementary Table S4a and S4b.

Table 3. Pathways/groups enriched by 24 genes with the highest Sscore and Rscore

Pathway/gene set Name		GO ID	p-value
The first 10 pathways/ gene sets enriched by top 24 genes with highest Sscore	TCTN-B9D complex	0036038	8.93E-05
	Wnt signaling pathway, planar cell polarity pathway	0060071	9.70E-05
	Heart development	0007511	1.43E-04
	Canonical Wnt signaling pathway	0060070	1.59 E-04
	Histone deacetylase binding	0042826	2.92 E-04
	TC 2.A.97.1	Pathway Studio Ontology	4.27 E-04

	Positive regulation of transcription, DNA-templated	0045941	6.03E-04
	Histone deacetylase complex	0000118	6.33E-04
	Establishment or maintenance of cell polarity	0007163	6.88E-04
	Phospholipase C gamma	Pathway Studio Ontology	8.54 E-04
The first 10 pathways/ gene sets enriched by top 24 genes with highest Rscore	Heart development	0007511	3.01E-12
	Positive regulation of transcription from RNA polymerase II promoter	0010552	5.84E-09
	Regulation of myelination	0031641	6.29E-09
	Liver development	0001889	6.77E-09
	polymerase II promoter	0000122	7.31E-09
	transition	0010718	3.53E-08
	Embryonic hindlimb morphogenesis	0035116	6.73E-08
	Transcription factor binding	0008134	1.07E-07
	Chromatin binding	0003682	1.2E-07
	Protein complex	0043234	1.36E-07

Table 3 shows that top Sscores genes and those with the top Rscores gene were enriched in different groups of pathways (only one overlap: GOID 0007511), with different *p*-values (Sscore group: 8.54 E-04 - 8.93 E-05; Rscore group: 1.36E-07 - 3.01E-12), indicating that the top genes selected by SRVS were functionally different from the most frequently reported ones.

The 17 pathways/gene sets enriched with the 24 Sscore genes (*p*-value < 0.00091, with 16/24 unique genes; Supplementary Table 4a), held 1 gene set related to heart development (GO: 0007511; *p*-value = 0.00014, overlap: 4). These sets also held 2 gene sets related to transcription factors for positive DNA-templated transcription regulation (GO: 0045941; *p*-value = 0.0006, overlap: 5) and positive regulation of transcription on the RNA polymerase II promoter

(GO: 0010552; *p*-value = 0.00091, overlap: 6). Both have been previously implicated with CHD (Clark *et al.* 2016).

In addition to PEA, a SNEA using Pathway Studio for identifying a selected gene's pathogenic significance to other disorders, possibly related to CHD, was performed. Full results appear in Supplementary Table S5a and S5b. Table 4 shows the top 10 disease-related sub-networks enriched by the two gene groups.

Table 4 indicates that both groups enriched some other heart defect related sub-networks, as well as other congenital and genetic mutation related sub-networks. Two sub-networks were observed for both groups: congenital malformation and leukemia.

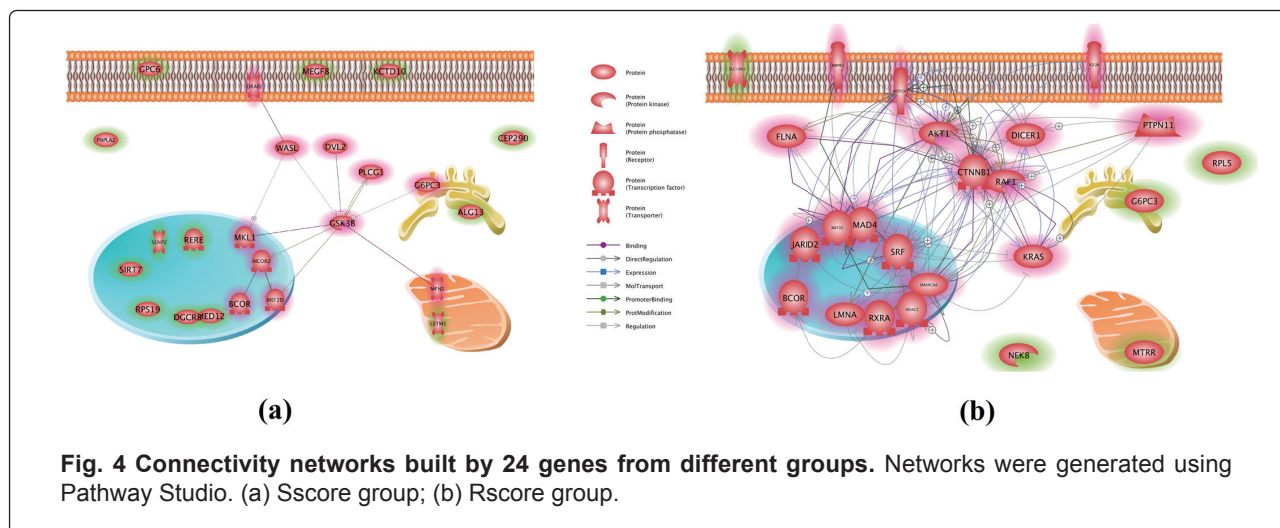
**Table 4. SNEA Results by 63 Genes with the Highest NScore and Rscore**

	Gene Set Seed	Overlap	p-value	Jaccard similarity
The first 10 pathways/gene sets Enriched by top 63 genes with highest Sscore	congenital malformation	13	3.59E-10	0.008238
	cardiac hypertrophy	7	5.17E-07	0.012844
	Obesity	10	7.57E-06	0.005033
	Dystrophy	4	2.22E-05	0.020833
	cognitive impairment	5	4.99E-05	0.011236
	Leukemia	7	6.38E-05	0.006256
	craniorachischisis	2	8.28E-05	0.055556
	malignant transformation	5	0.00011	0.00956
	DiGeorge syndrome	3	0.000111	0.025641
	adenovirus infection	4	0.000126	0.013841
The first 10 pathways/gene sets Enriched by top 150 genes with highest Rscore	congenital malformation	22	1.09989E-24	0.01402167
	leukemia	16	1.28978E-16	0.014414414
	colon cancer	17	1.30808E-16	0.012203877
	aneuploidy	11	8.3047E-16	0.037288136
	dysplasia	13	1.13206E-15	0.022146508
	gastric cancer	17	1.66302E-14	0.009129968
	adenoma	13	2.59156E-14	0.017426273
	genetic disorder	16	4.65089E-14	0.009919405
	diabetes mellitus	18	4.84479E-14	0.007475083
	T-cell acute lymphoblastic leukemia	10	9.46603E-14	0.032258065

### 3.5 Connectivity Analysis

In addition to PEA and SNEA, an NCA was performed on the top 24 genes having the highest Rscores and Sscores (Table 2) to generate functional networks. In the Sscore group, 11 of the 24 genes presented 14 direct connections supported with 53

references. The Rscore group had more than 113 relations among 19/24 genes, with 1,283 literature supports (Fig. 4 (b)). This observation is consistent with the PEA and SNEA. It suggests that the highest Sscores genes were less functionally related to each other than the Rscore group genes were.



## 4 DISCUSSION AND CONCLUSION

Identification of significant biomarkers based on a small number of observations is a fundamental problem in signal processing. This study proposed a sparse representation based genetic marker selection approach, and applied it to evaluate 167 CHD candidate genes. The genes were identified from a CHD-Gene network relation data set acquired from the ResNet database. This set overlapped with a RNA gene expression data set (GEO: GSE34457). Two metric scores were generated and compared: the Sscore from SRVS analysis, and the Rscore from the CHD-Gene relation data set analysis.

LOO cross validation demonstrated that if whole 167 CHD candidate genes were used, a classification ratio of only 58.14 % was obtained with a permutation  $p$ -value of 0.30 (Table 1). Using both Sscore, and Rscore, top genes resulted in greater CRs (81.40 % and 69.77 %) with a more significant permutation  $p$ -value ( $< 0.017$ ). This suggests a need for variable selection for the candidate CHD genes tested, as well as the efficacy of both Sscore and Rscore.

The top 24 Sscore genes led to the greatest CR with the lowest permutation  $p$ -value, demonstrating the effectiveness of the proposed SRVS method in CHD variable selection. The top 24 Rscore-selected genes and the 24 genes selected by SRVS were compared to provide a clearer understanding. Analysis showed that these two groups share only two genes: BCOR and G6PC3 (Fig. 3 (a)). Enriched pathways (Table 4), associated sub-networks (Table 5), and gene-gene interactions (Fig. 4) revealed other differences. The results highlight that even though well-studied CHD candidate genes were

significant to the disease and were also effective in disease prediction (LOO permutation  $p$ -value = 0.017), they were not the best genetic markers for the subjects involved with the expression data tested (GEO: GSE34457).

Despite differences between the top genes selected by Sscore and Rscore, many of Sscore enrichment pathways previously reported to associate with CHD were identified. For example, heart development, positive regulation of transcription from the RNA polymerase II promoter, the DNA-templated positive regulation of transcription, the canonical Wnt signaling pathway, and histone deacetylase binding<sup>[11-13]</sup>. Furthermore, these genes were also identified to be the genetic basis of other CHD related diseases, such as congenital malformation, cardiac hypertrophy, dystrophy, and cancer<sup>[14, 15]</sup>. These results support the biological validity of the top genes selected by the SRVS approach.

There is direct literature support for associating CHD and the top 24 genes selected by Sscore (Supplementary Table S1). A strong functional association between the top genes selected by the Sscore and Rscore groups (Fig. 3 (b)) is supported by over 3,000 references (Supplementary Table S3). Genes with a high Rscore suggests strong support in the literature for linking it to CHD. Our observations provide indirect support for the proposition that the majority of the top genes selected by the SRVS method are functionally significant in CHD.

This study does have several limitations calling for future work. Although the algorithm was tested on 167 CHD candidate genes, there are other genes



linked to CHD that were not included in the data set and were, therefore, not analyzed. More inclusive data sets including all CHD genes should be used to test this method's accuracy. The method should also be tested on other diseases to determine its validity.

We conclude that CHD is a complex disease with genetic causes linked to a network composed of a large gene group. Each patient/patient group may present with unique genomic variations requiring treatment based on their particular disease risk prediction, where the proposed SRVS method can be employed effectively.

## DECLARATION OF INTERESTS

The authors declare no conflict of interests.

## REFERENCES

- Hoffman JI, Kaplan S. The incidence of congenital heart disease. *J Am Coll Cardiol*. 2002; 39(12): 1890-1900.
- Sifrim A, Hitz MP, Wilsdon A, Breckpot J, Turki SH1, Thienpont B, McRae J, Fitzgerald TW, Singh T, Swaminathan GJ, Prigmore E, Rajan D, Abdul-Khaliq H, Banka S, Bauer UM, Bentham J, Berger F, Bhattacharya S, Bu'Lock F, Canham N, Colgiu IG, Cosgrove C, Cox H, Daehnert I, Daly A, Danesh J, Fryer A, Gewillig M, Hobson E, Hoff K, Homfray T; INTERVAL Study., Kahlert AK, Ketley A, Kramer HH, Lachlan K, Lampe AK, Louw JJ, Manickara AK, Manase D, McCarthy KP, Metcalfe K, Moore C, Newbury-Ecob R, Omer SO, Ouwehand WH, Park SM, Parker MJ, Pickardt T, Pollard MO, Robert L, Roberts DJ, Sambrook J, Setchfield K, Stiller B, Thornborough C, Toka O, Watkins H, Williams D, Wright M, Mital S, Daubeney PE, Keavney B, Goodship J; UK10K Consortium., Abu-Sulaiman RM, Klaassen S, Wright CF, Firth HV, Barrett JC, Devriendt K, FitzPatrick DR, Brook JD, Deciphering Developmental Disorders Study, Hurles ME. Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nat Genet*. 2016; 48(9): 1060-1065.
- Srivastava D. Making or breaking the heart: From lineage determination to morphogenesis. *Cell*. 2006; 126(6): 1037-1048.
- Lu Y, Goldstein DB, Angrist M, Cavalleri G. Personalized medicine and human genetic diversity. *Cold Spring Harb Perspect Med*. 2014; 4(9): a008581.
- Donoho DL, Elad M, Temlyakov VN. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans Inform Theory*. 2006; 52(1): 6-18.
- Cao H, Deng H, Li M, Wang Y. Classification of multicolor fluorescence in situ hybridization (M-FISH) images with sparse representation. *IEEE Trans Nanobiosci*. 2012; 11(2): 111-118.
- Cao H, Duan J, Lin D, Shugart YY, Calhoun V, Wang Y. Sparse representation based biomarker selection for schizophrenia with integrated analysis of fMRI and SNPs. *Neuroimage*. 2014; 102(1): 220-228.
- Donoho DL, Tsaig Y. Fast solution of L1-norm minimization problems when the solution may be sparse. *IEEE Trans on Informat Theory*. 2008; 54(11): 4789-4812.
- Davis G, Mallat S, Avellaneda M. Greedy adaptive approximation. *J Constr Approx*. 1997; 13(1): 57-98.
- Sivachenko AY, Yuryev A, Daraselia N, Mazol. Molecular networks in microarray analysis. *J Bioinform Comput Biol*. 2007; 5(2B): 429-456.
- Clark KL, Yutzey KE, Benson DW. Transcription factors and congenital heart defects. *Annu Rev Physiol*. 2006; 68(1): 97-121.
- Zhang R, Gui Y, Wang X. Role of the canonical Wnt signaling pathway in heart valve development. *Chin J Contemp Pediat*. 2015; 17(7): 757-762.
- Walfridsson J, Bjerling P, Thalen M, Yoo EJ, Park SD, Ekwall K. The CHD remodeling factor Hrp1 stimulates CENP-A loading to centromeres. *Nucleic Acids Res*. 2005; 33(9): 2868-2879.
- Prokofyeva E, Wilke R, Lotz G, Troger E, Strasser T, Zrenner E. Distribution and clinical peculiarities of monogenic retinal dystrophies at the Center for Ophthalmology, University of Tuebingen. *Acta Ophthalmologica*. 2008; 86 (s243): doi:10.1111/j.1755-3768.2008.6356.x.
- Lee Y, Chen Y, Jeng M, Tsao P, Yen H, Lee P, Li S, Liu C, Chen T, Chou P, Soong WJ. The risk of cancer in patients with congenital heart disease: a nationwide population-based cohort study in Taiwan. *PloS One*. 2015; 10(2): e0116844.