Inferring novel lncRNA associated with Ventricular septal defect by DNA methylation interaction network

Min Zhang[1], Yue Gu[1], Mu Su[2], Shumei Zhang[1], Chuangeng Chen[1], Wenhua Lv[1], Yan Zhang[1,2*]

[1]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, 150081, China.

[2]School of Life Science and Technology, Harbin Institute of Technology, Harbin 150001, China

*Corresponding author

E-mail: tyozhang@ems.hrbmu.edu.cn)

Abstract

Ventricular septal defect (VSD) is one of the most common types of congenital heart disease. To find more and more molecular alteration is conducive to explore the mechanism and biomarker in VSD. Herein we devised a predictive strategy to uncover novel lncRNA of VSD integrating DNA methylation, gene expression and lncRNA expression of early embryo and VSD by profiles from GEO database. In totally, 175 lncRNAs, 7290 genes and 3002 DNA methylation genes were obtained by logistic regression analysis associated with embryonic development. Moreover, 7304 DMGs were significant differential methylated by Wilcoxon rank test and Student's test in VSD. We constructed the lncRNA-mRNA co-expression network in embryo (LMCNe). Then, a reconstructed co-expression weighted network (RCWN) was built integrated LMCNe and the DNA methylation associated network（DMAN）based on the correlation of the DNA methylation level and protein interaction network between embryonic development and VSD. We extracted top 10 lncRNAs with higher score performing DRaWR from the weight network, which as potential VSD related lncRNAs. Six lncRNAs showed a high level of expression in the heart tissue recorded in the NONOCOND database. Furthermore, associated lncRNA genes DCAF8L1, NIT1, SH2D7 and DOCK9-AS2 in validated samples showed a prominently association with VSD. These outcomes provide a reference for lncRNA involved in VSD initialization and a new insight for studies of VSD-associated lncRNAs.

Key words: Ventricular septal defect; DNA methylation; lncRNA; network

Author Summary

Ventricular septal defect (VSD) is one of the most common types of congenital heart disease and has a high mortality rate in infants. Many molecular markers have proved effective as biomarker in VSD like DNA methylation and lncRNA. lncRNA is a type of non-coding RNA which has important effect in regulation gene expression and disease occurrence. VSD is an embryonic stage developmental disease. Therefore we hypothesized that lncRNA which was associated with DNA methylation and mRNA in early embryonic development may also affect the occurrence of VSD. So in this work, from the perspective of embryonic development, we devised a predictive strategy to uncovering novel lncRNA of VSD. In our result, four lncRNA associated genes were found differential expressed in VSD and normal samples by qPCR validation. The identification of lncRNA associated with ventricular septal defect is beneficial to further study the mechanism of VSD from the molecular level and also provides a good molecular marker for clinical therapeutic and diagnosis. At the same time, it also provides a new insight for the researches of lncRNA associated with VSD.

1. Introduction

The heart is the first developing organ in embryonic development which can provide oxygen and nutrients for the baby and many genes and biological processes are involved in this procedure. A slight disruption in the process of embryonic development of the heart could cause damage to the embryo, leading to congenital heart disease(1, 2). Congenital heart disease is one of the most common defects in infants accompanied with a higher mortality rate(3). Some reports show that 4-14 of every 1,000 newborns have congenital heart disease, and congenital heart disease accounts for 40% of infant defects(4-7). Congenital heart disease could be caused by a variety of factors, including environmental, genetic, epigenetic effects(8). VSD is considered to be one of the most common types of congenital heart disease, accounting for about 30% of congenital heart disease(9). VSD could exist in isolation and could as part of double outlet right ventricle and other cardiac anomalies(10).

DNA methylation is an important epigenetic modification that plays an important role in cell differentiation, regulation of gene expression, disease occurrence and development and participates in biological processes of progressive organs. DNA methylation often occurs on the fifth carbon atom of cytosine, formed by DNA methyltransferase. Hypermethylation in gene promoter can contribute tumor suppressor silencing leading to disease. DNA methylation is reprogrammed during the zygote phase of early embryos and rewrite into the genome during cell division and differentiation (11, 12). Epigenetic marks have a sustained effect on individuals and progeny and DNA methylation pattern is stable in heredity(13, 14).Furthermore, most

genes in cell were arranged by DNA methylation in the period of early embryo development and abnormal DNA methylation could lead to defect in births(15). Folate is a B-vitamin providing methyl group for methylation reactions in embryonic cell including DNA methylation and could reduce the risks of congenital heart disease (16). Recently, many studies have been investigated in understanding the role of DNA methylation in congenital heart disease. For example, higher abundance of methylation biomark S-adenosylhomocysteine (SAH) in blood increased the risk of CHD (17). The hypermethylation of RAB43 and KIAA0310 leads to ER-te-Golgi dysfunction and eventually contributes to the occurrence of VSD and abnormal embryonic heart development(18). What is more, the hypermethylation of NOX5 frequently occurs in the pathogenesis of VSD(19).

In recent years, the human genome is divided into coding and non-coding regions(20). With the advanced genome sequencing technology, many mRNA-like functional transcripts have been found in noncoding regions which are non-coding RNAs (21-23). Non-coding RNA have vital influence in regulating gene expression instead of being a protein template(24). LncRNA is a type of non-coding RNA whose length is greater than 200bp having divided into five categories: sense, antisense, bidirectional, intronic, and intergenic(25, 26). Accumulating evidence suggested that lncRNA has a wide range of biological functions involved in the regulation of gene expression networks, such as chromosome printing, cell growth and differentiation and tumorigenesis(27-29). However, lncRNA associated with VSD is rarely reported.

Herein, by analyzing the DNA methylation, gene expression and lncRNA profile of

early embryo and DNA methylation data of VSD. We constructed an integrated RCWN modified by DNA methylation correlation based on protein interaction network. This network not only illustrated intricate relationship between coding genes and non-coding lncRNA in embryonic development, but also revealed the DNA methylation-mediated correlation between embryonic development and VSD. Then we selected the top 10 lncRANs with higher score as VSD associated lncRANs after performing DRaWR algorithm in the RCWN. Six lncRNA were found expressed in heart tissue included in NONCODE database. Furthermore, we investigated the expression of lncRNA gene were apparently associated with VSD by comparing the normal and VSD cardiac tissue. These lncRNAs were mainly involved in the function of regulation of growth heart force and so on, which would be a reference for further analysis.

2. Results

2.1 Identification of differential methylation genes(DMGs) in VSD

VSD is a disease that caused by abnormal events occurring in embryonic development. DNA methylation plays an important role in embryonic development, and may affect VSD. Therefore, we devised an integrated prediction network to predict lncRNA associated with VSD. We integrated MBD-Seq DNA methylation data containing four normal right atrium and eight VSD samples and DNA methylation data lncRAN expression profile and gene expression profile of early embryos. The LMCNe was built in embryonic firstly. Then， a reconstructed co-expression weighted network (RCWN) was built integrated LMCNe and DMAN

based on the correlation of the DNA methylation level and protein-protein interaction between embryonic development and VSD. By performing Discriminative Random Walk with Restarts (DRaWR), selecting the highest score of 10 lncRNA as a candidate lncRNA Figure 1.

For DNA methylation data of VSD, 18,322,642 CpG sites were obtained by data filtering and deletion (see methods). 741,027 CpG sites were considered as differential methylated sites by Wilcoxon rank test. Subsequently, these differential DNA methylation CpG sites were matched to the promoter region of 9,274 refseq genes and the 7,292 differential DNA methylation genes (DMG) were screened using the Student test with setting threshold $p<0.05$ and districted DNA methylation level greater than 0.2 between normal and VSD figure 2 A. The overall trend of the differential DNA methylation genes were that DNA methylation levels of most genes were upregulated, with only a small fraction of the downregulation in figure 2 B. Moreover, the DNA methylation level of DMGs in VSDs were significantly higher than that in the normal samples figure 2 C. The functional enrichment of these DMGs showed that they were associated with signal transduction, organ development, Rap1 signaling pathways and other basic functions which has great regulatory for bodies (figure 2 D).

2.2 Identification of embryonic development-related lncRNA, mRNA and DNA methylation gene

We treated six stages of early embryonic DNA methylation data including oocytes, Zygotes, 2-cell-stage, 4-cell-stages, 8-cell-stage, morulae stage. In the following

analysis, methylated promoter of 14,304 refseq gene were reserved. 8,123 lncRNA expression and 10,176 mRNA expression were available from the document with RPKM (see Method). Subsequently, we fitted an Ordered Rogers regression model to screen for lncRNAs and genes associated with embryonic development. In belief, 3,002 associated DNA methylation genes (AMGs) were regarded as associated with embryonic development and most of the genes were in a low DNA methylation state. With further embryo developmental stages, DNA methylation level of AMGs exhibited a tendency of demethylation which was consistent with previous reports figure 3 A(30, 31). These AMGs were mainly involved in cell differentiation, cell adhesion, DNA template regulation, protein transcription, RNA transcription pathways, and other pathways and biological process that had played an important role in embryonic development and organ growth(figure 3 B). However, for the 7304 gene expression associated with embryonic development, their expression showed a slight increase in figure 3 C. The function of these genes were mostly enriched in transcription, DNA-templated, cell division and other biological processes and pathways(figure 3D). Similarly, the expression of 175 embryonic associated lncRANs were identified and we studied protein coding genes within upstream and downstream 100kbp of lncRNAs and found that they were primarily focused on transcription, DNA-templated, cellular protein localization and signaling pathways of regulating pluripotency of stem cells(figure 3 E and F). Additionally, as displayed in figure S 1, overlapped embryonic development-related DNA methylation genes and gene expression and found 566 overlapping genes indicating that DNA methylation levels

of the promoters of these genes have a negative regulatory effect on gene expression, in other words, higher DNA methylation level are obviously associated with down-regulated gene expression.

2.3 Construction of LMCNe

In order to illustrate the relationship between lncRNA and protein coding genes in embryonic development. We established a LMCNe on the basis of embryo-associated genes and lncRNAs in the figure 3 with Pearson correlation coefficient more than 0.8 between mRNA and lncRNA. LMCNe contained 3,344,593 linkages and 7,304 mRNAs connected with 175 lncRNAs which indicated that many protein-coding were regulated by the same lncRNA and also governed by differential lncRNA at the same time figure S 2. The distribution of node degree evenly appeared to pow-law distribution suggested that LMCNe was a proper network with biology implication figure S2

2.4 Construction of DMAN

In recent years, protein-protein interaction network (PPI) has dominated the network and revealed more complex biological mechanisms and biological processes. It is typically used in network analysis. Here, we constructed the DMAN of early embryos and VSD by two steps in the background of PPI. First, we acquired significant correlation DNA methylation gene pairs in embryo and VSD methylation genes set separately (see Materials and Methods). In the second step, we used the PPI network as the background network, and the two significant co-methylation gene set were to be incorporated into a DMAN through commons nodes and edges.

At the first stage, we found 1,464,756 co-methylated DNA methylation gene pairs (DMP) in embryo and 1,515,563 DMPs in VSD set. We observed that some DMPs were attended in both embryo set and VSD set suggested that some DMPs simultaneously affects VSD occurrence and embryo growth. In order to enhance the correlation between two DNA methylation pair sets, we introduced the protein-protein interaction network as the background network, and mapped the DMPs to the PPI network respectively. Only genes that existed in the PPI network were retained and merged into a DMAN by common nodes and co-methylation relationship. DMAN consists of 7,925 nodes and 94,945 edges with three categories of interactions: DMPs occurred in embryo or in VSD's or in both figure 4 A. Moreover, the node-degree distribution of DMAN obeyed the power-law distribution revealed that the DMAN was a scale-free network that only a few nodes were with high connectivity and most nodes are low in connectivity figure 4 B. The interaction in DMAN were established by the absolute value of Pearson correlation coefficient greater than 0.8 and to confirm intensity of DMAN, we randomized the DNA methylation level of the nodes in the network for 100 times and calculated the Pearson correlation coefficient. It was found that the Pearson correlation coefficient of DMAN was much higher than the average stochastic suggested that DMPs were not random events figure 4 C.

2.5 Construction of RCWN

The RCWN was rebuilt by integrating LMCNe and the DMAN which was as the weight of edges attribute and included 3,439,296 links and 14,147 nodes figure 4 D. 100 times permutation of Pearson coefficient for each linkage were performed by

disorganizing the expression value of lncRNAs and mRNAs. Obviously, the correlation of co-expression networks was greater than the random average correlation coefficient figure 4 F. Obviously, the degree distribution of nodes in this network was also subject to power-law distribution indicating that some genes may regulated by both lncRNA and methylation figure 4 E.

2.6 DRaWR analysis for VSD related lncRNA

With respect to DNA methylation plays an important role in embryonic development and abnormal DNA methylation programming could increase the risk of VSD in neonates (32). Therefore, we hypothesized that VSD was an embryonic stage developmental disease. LncRNAs which were associated with DNA methylation and mRNA in early embryonic development may also affect the occurrence of VSD. Thence we used the DRaWR algorithm to analyze the RCWN and predicted the lncRNA with VSD figure 5 A (see the materials and methods). In our study, we selected 120 genes that simultaneously affect embryonic development and VSD initialization as 'query set' figure 5 B. DNA methylation level of these 120 genes presented a demethylation pattern figure 5 C during developmental stages. We retained the top 10 lncRNAs with higher scores as VSD potential lncRNA lnc-POLE4-8, lnc-DCAF8L1-1, lnc-RAB1A-1, lnc-NIT1-1, lnc-SH2D7-1, lnc-POTEB-15, LINC01467, lnc-ECI2-5, DOCK9-AS2, LINC01622 as showed in figure 6 A . At the same time, we also studied 10 genes, MTO1, DNAJB12, MRFAP1, MAD2L2, MTG1, LPIN2, FBXL22, PEBP1 and POR, which were connected to the 10 lncRNAs in the network with the highest rank figure 6 B. MTO1

mutations were associated with hypertrophic cardiomyopathy and PEBP1 could affect heart failure and lactic acidosis and cause respiratory chain deficiency in humans and yeast (38,39). FBXL22 is a cardiac-enriched sarcomere protein and is essential for maintenance of normal contractile function in vivo (40). LPIN2 is likewise a cardiac gene (41).We also analyzed protein-coding genes located 100 kbp near lncRNA, such as APOA2, a gene that encodes high-density lipoprotein binding protein, who was associated with cardiovascular risk(37). The mutation of NDUFS2 and USF1 gene variants was often occurred in cardiomyopathy and has been a biomark in cardiovascular disease(38-40) and its promoter hypermethylation is associated with congenital heart disease which were the neighbors of lnc-NIT1-1(41).

2.7 Validation of lncRNAs in VSD

To explore whether the discovered lncRNAs were indeed associated with VSD, through the NONCODE database, we found that six lncRNAs were highly expressed in the heart tissue and were validated by quantitative PCR in 13 heart samples (see method) figure 7A. Four lncRNAs associated genes exhibited high expression level than control group especially DCAD8L1 and SH2D7 measured by students test. (figure 7) and their function was mainly focused on the regulation of birth, cardiac systolic, cell cycle, and other important pathways related to cardiac development figure 6 C. Some lncRNAs like lnc-POLE4-8 were not only involved in the regulation of growth but also in the contraction of the heart.

3 Discussion

VSD is one of the most common defect in newborns, affected by multiple factors

such as genetic factors and epigenetic factors. DNA methylation as the most apparent modification plays an important role in the embryonic development, cell growth, differentiation gradually becoming a diagnostic marker. Hypermethylation of gene promoter region regulated gene expression.

In recent years, the crucial role of DNA methylation in congenital heart disease has been gradually found, for example, NOX5 promoter hypermethylation occurred more frequently in VSD, abnormal hypermethylation of gene promoter leads to gene silencing. In this study, we first obtained DNA methylation gene that was significantly different with normal samples by compared with the VSD sample, and also obtained DNA methylation genes associated with embryonic development. By Pearson correlation analysis, the significant DMPs in the VSD were excavated, and the DMPs in embryonic development were also investigated. In order to further increase the relationship of DMPs, we introduced functional correlation protein-protein interaction network. After constructing a DNA methylation associated network, Protein-protein interaction network had increased the robustness of DMAN. Looking for VSD associated related lncRNAs, we used DRaWR algorithm for network analysis and 120 genes were as seed nodes, the methylation of these genes were not only related with embryonic development, but also VSD. We considered top 10 lncRNAs as candidate lncRNA. Most of the lncRNAs are lincRAN and adjacent to the heart disease gene. For example, lnc-POLE4-8 is adjacent to HK2 which is Hexokinases phosphorylate glucose to produce glucose-6-phosphat was found to be associated with ventricular function in mice (33).

However there also exist some limitations, firstly it did not have enough data of VSD to use to support our study, we downloaded MBD-seq methylation data of VSD containing 8 cases and 4 controls covered only part of CpG sites. Secondly, due to the use of lncRNA and gene expression data from the early embryonic stage, some values were somewhat lower and there may exits false positive rate.

4 Method and Materials

4.1 Samples and data process

VSD methylation data was download from GEO database (https://www.ncbi.nlm.nih.gov/geo/) with the accession number GSE62629 including four normal right atrium and eight VSD samples. Methylation dataset of embryo with six stages was also download from GEO and accession number is GSE49828. As for DNA methylation data only the CpG sites with read coverages more than five times were taken into account. When we quantified the average DAN methylation level of a gene, we calculated the average DNA methylation level of total number of all CpG site located within 1500bp upstream and 500bp downstream of transcription start site as gene promoter DNA methylation level. Gene expression and lncRNA expression profile of embryo with RPKM were reported by Yan (34) . All of genes and lncRNA with an RPKM>0.0 was considered in the following analysis .The hg19 Refseq genes and lncRNA annotation were download from the UCSC Genome Browser (http://genome.ucsc.edu) and NONCODE database separately(35). The validate samples were extracted from The first Affiliated Hospital of Harbin Medical University and the study was approved by The Ethics Committee of First Affiliated

Hospital of Harbin Medical University.

4.2 Total RNA extraction and purification

The total tissue for RNA extraction varied between 10 and 15 mg from each animal. In order to ensure unbiased analysis of tissue response, total RNA was isolated from randomly sectioned (weighting 10- 15 mg) heart. RNA was isolated from 13 to 8 individual samples from each of the treatment and control groups using TRIzol reagent (inveitrogen, Carlsbad, CA, USA) and purified using RNeasy Plus Mini kits (Qiagen, Mississauga, ON, Canada) according to the manufacturer's instruction. Total RNA concentration was measured using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific Inc., Wilmington, DE, USA), and RNA quality and integrity were assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc. Mississauga, ON, Canada) according to the manufacturer's instruction. All samples showed RNA integrity numbers of 7 and above, indicating high quality RNA, and were used to conduct qPCR experiments.

4.3 Real-time quantitative PCR

Total RNA were used to perform reverse transcription using PrimeScriptTM RT reagent Kit (TaKaRa, Tokyo, Japan). Real-time polymerase chain reaction amplification of cDNA was performed with SYBR Premix Ex TaqTM II (TaKaRa, Tokyo, Japan). The sequences of the primer sets were listed as follows:

DCAF8L1 forward, 5'- TCAAAGTAGTTTCCAGGCAATCTTGTG-3' and reverse, 5'- CTGGGATTTGGATGGAGTCGTT-3'; RAB1A forward, 5'- CAGTCACGAGGCTCTCCGAA-3' and reverse, 5'-

AGTTCCTCATAGGCTATTGGACTGA-3'; NIT1 forward, 5'-GGAAGACCAGATTATGCGAA-3' and reverse, 5'-CATTATTCCCCAGTTGTCCA-3'; DOCK9-AS2 forward, 5'-AGAGAGCAAAACACCTCCGTT-3' and reverse 5'-AGGAGGAAATGGGTTAGTGTGT-3'; SH2D7 forward, 5'-TCCCCTGGGTTACTCCATTC-3' and reverse 5'-CTCATTACTCTCCCCAAAACCTG-3'; □-actin forward, 5'-GCTCCAAGCAGATGCAGCA-3' and reverse 5'-CCGGATGTGAGGCAGCAG-3' Quantification of gene expression was determined by comparative quantity, using □-actin gene expression as inner control.

4.4 Statistical analysis

Due to the multiple sorted stages of embryo, ordered logistic Regression model was used to estimate correlation between development stage of embryo and DNA methylation performing with R packages MASS plor function. Y donated early stages of embryonic development, X was delegated the DNA methylation level of each gene or lncRNA expressiong or gene expression level.

$$\text{logit}(Y) = \beta_0 + \beta_1 X \quad (1)$$

For DNA methylation data in VSD, significant differential methylated CpG sites were identified using Wilcoxon rank test in R software. Bedtools was used to calculate total CpG sites in Refseq gene promoter. To analyze differences in DNA methylation between normal tissue and VSD, p value was calculated by Student's t test. Co-expression relationships among the lncRNA, gene expression and DNA

methylation were estimated by Pearson correlation test and correlation coefficient greater than 0.8 were used in fallowing analysis.

4.5 Go and pathways analysis

Enrichment analysis was analyzed in DAVID database. For associated-lncRNA in embryo, protein-coding genes between 100kbp upstream and 100kbp downstream for each lncRNA were used for enrichment analysis. Go terms and pathways with the threshold p-value<0.05 were selected as significant enrichment results and visualized top 10 terms of biological process, cellular component, molecular function and pathways of KEGG in Cytoscape3.5 enrichment plug-in.

4.6 DNA Methylation associated network (DMAN)

In this study, we introduced a human protein-protein interaction network as background network for the further analysis. The network integrates the protein interaction relationship for six database, including HPRD ,DIP ,MINT,BIND, IntAct (36) and STRING database. A DNA methylation associated network was extracted by mapping DMG pairs into the background network. DMGs were used as the node of the network and interaction between DMGs serves as network edges formed by Pearson correlation analysis with threshold p value<0.05 and coefficient>0.8. Then 1000 random permutations were performed for DMG pairs calculating Pearson correlation coefficient.

4.7 RCWN

In order to identify lncRNA associated with VSD, we merged LMCNe and DMAN as RCWN. The degree of node refers to the number of neighbor nodes and the weight of

an edge was defined as following principles:

M_E refers to co-methylation gene pair dataset in embryo,G_E is defined as co-expression gene pair in embryo.M_V donated significant co-methylation gene pairs in VSD. GL_E means significant gene-lncRNA co-expression pair in embryo.

$$\text{weight(edge)} = \begin{cases} 1 \; if \; gene \; pair \; occure \; in \; M_E \; or \; G_E \; or \; M_V \; or \; GL_E \\ 2 \quad if \; gene \; pair \; occure \; in \; M_E \; and \; G_E \\ 2 \quad if \; gene \; pair \; occure \; in \; M_E \; and \; M_V \\ 2 \quad if \; gene \; pair \; ocure \; in \; G_E \; and \; M_V \\ 3 \quad if \; gene \; pair \; occure \; in \; M_E \; , G_E \; and \; M_V \end{cases}$$

4.8 Discriminative random walk with Restarts (DRaWR)

Discriminative random walk with Restarts is an algorithm to rank genes for their relevance to a given 'query' gene set based on "guilt by association". The method performs two times of Random walking with restart (RWR). In the first stage of DRaWR, a proper subnetwork was extracted from a large original network including only seed-relevant node. Then ranking all genes according to the relevance to query gene set in the second stage of DraWR in the light of subnetwork(37).

In this study, gene set $\mathrm{Q} = \{q_1, q_2 \ldots q_n\}$ refers to the seed gene set and $\mathrm{G} = \{g_1, g_2, \ldots g_m\}$ is the total node of the RCWN. D denotes an adjacency matrix of network and D_{ij} is the link between g_i and g_j .

$$\mathrm{D} = \begin{bmatrix} D_{11} & D_{12} & \cdots & D_{1m} \\ D_{21} & D_{22} & \cdots & D_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ D_{n1} & D_{n2} & \ldots & D_{nm} \end{bmatrix} \quad (2)$$

Then normalized the adjacency matrix N:

$$N_{ij} = \frac{D_{ij}}{\sum_{ij} D_{ij}} \quad (3)$$

Next normalized each columns of then N to creat a transition probability matrix T

Using following equation:

$$T_{ij} = \frac{N_{ij}}{\sum_{ij} N_{ij}}. \quad (4)$$

T_{ij} is the probability the walker according to an edge transition from node i to node j. Moreover follows the equation to perform the RWR :

$$W_{t+1} = (1-r)TW_t + r\alpha \quad (5)$$

where r is the restart parameter, the walker reset the walk by moving directly to other genes in query gene set. W_t refers to the probability distribution of all node in the network since t steps of RWR algorithm. Given a cutoff $|W_{t+1} - W_t| < 0.05$ ranking all node in W by probability distribution. Notably, in the first stage, W_Q and W_G are calculated respectively, and the nodes with the largest difference between W_Q and W_G are reserved to form a subnet. In the second stage of DRaRW , all nodes were ranked by repeating RWR on the subnet until $|W'_{t+1} - W'_t| < 0.05$. DRaWR is a R packages and performed in R3.3.1 software.

Acknowledgments

This work was supported by National Natural Science Foundation of China [grant number 81573021].

References

1. Satou Y, Satoh N. Gene regulatory networks for the development and evolution of the chordate heart. Genes & development. 2006;20(19):2634-8.

2. McQuillen PS, Miller SP. Congenital heart disease and brain development. Annals of the New York Academy of Sciences. 2010;1184:68-86.

3. Hoffman JI, Kaplan S. The incidence of congenital heart disease. Journal of the American College of Cardiology. 2002;39(12):1890-900.

400 4. Mendieta-Alcantara GG, Santiago-Alcantara E, Mendieta-Zeron H, Dorantes-Pina R,
401 Ortiz de Zarate-Alarcon G, Otero-Ojeda GA. [Incidence of congenital heart disease and
402 factors associated with mortality in children born in two Hospitals in the State of Mexico].
403 Gaceta medica de Mexico. 2013;149(6):617-23.
404 5. Acharya G, Sitras V, Maltau JM, Dahl LB, Kaaresen PI, Hanssen TA, et al. Major
405 congenital heart disease in Northern Norway: shortcomings of pre- and postnatal diagnosis.
406 Acta obstetricia et gynecologica Scandinavica. 2004;83(12):1124-9.
407 6. Ransom J, Srivastava D. The genetics of cardiac birth defects. Seminars in cell &
408 developmental biology. 2007;18(1):132-9.
409 7. Jenkins KJ, Correa A, Feinstein JA, Botto L, Britt AE, Daniels SR, et al. Noninherited risk
410 factors and congenital cardiovascular defects: current knowledge: a scientific statement from
411 the American Heart Association Council on Cardiovascular Disease in the Young: endorsed
412 by the American Academy of Pediatrics. Circulation. 2007;115(23):2995-3014.
413 8. Eskedal LT, Hagemo PS, Eskild A, Froslie KF, Seiler S, Thaulow E. A population-based
414 study relevant to seasonal variations in causes of death in children undergoing surgery for
415 congenital cardiac malformations. Cardiology in the young. 2007;17(4):423-31.
416 9. Ko JM. Genetic Syndromes associated with Congenital Heart Disease. Korean
417 circulation journal. 2015;45(5):357-61.
418 10. de Hemptinne Q, De Becker B, Unger P. Ventricular septal defect: an unusual cause of
419 paradoxical low-gradient aortic stenosis. European heart journal cardiovascular Imaging.
420 2017;18(5):609.
421 11. Altorok N, Tsou PS, Coit P, Khanna D, Sawalha AH. Genome-wide DNA methylation

422 analysis in dermal fibroblasts from patients with diffuse and limited systemic sclerosis reveals
423 common and subset-specific DNA methylation aberrancies. Annals of the rheumatic diseases.
424 2015;74(8):1612-20.
425 12. Gilsbach R, Preissl S, Gruning BA, Schnick T, Burger L, Benes V, et al. Dynamic DNA
426 methylation orchestrates cardiomyocyte development, maturation and disease. Nature
427 communications. 2014;5:5288.
428 13. Hanson M, Godfrey KM, Lillycrop KA, Burdge GC, Gluckman PD. Developmental
429 plasticity and developmental origins of non-communicable disease: theoretical considerations
430 and epigenetic mechanisms. Progress in biophysics and molecular biology.
431 2011;106(1):272-80.
432 14. Reik W. Stability and flexibility of epigenetic gene regulation in mammalian development.
433 Nature. 2007;447(7143):425-32.
434 15. Kim KC, Friso S, Choi SW. DNA methylation, an epigenetic mechanism connecting folate
435 to healthy embryonic development and aging. The Journal of nutritional biochemistry.
436 2009;20(12):917-26.
437 16. Hernandez-Diaz S, Werler MM, Walker AM, Mitchell AA. Folic acid antagonists during
438 pregnancy and the risk of birth defects. The New England journal of medicine.
439 2000;343(22):1608-14.
440 17. Obermann-Borst SA, van Driel LM, Helbing WA, de Jonge R, Wildhagen MF, Steegers
441 EA, et al. Congenital heart defects and biomarkers of methylation in children: a case-control
442 study. European journal of clinical investigation. 2011;41(2):143-50.
443 18. Zhu C, Yu ZB, Chen XH, Pan Y, Dong XY, Qian LM, et al. Screening for differential

444 methylation status in fetal myocardial tissue samples with ventricular septal defects by
445 promoter methylation microarrays. Molecular medicine reports. 2011;4(1):137-43.
446 19. Zhu C, Yu ZB, Chen XH, Ji CB, Qian LM, Han SP. DNA hypermethylation of the NOX5
447 gene in fetal ventricular septal defect. Experimental and therapeutic medicine.
448 2011;2(5):1011-5.
449 20. Morris KV, Mattick JS. The rise of regulatory RNA. Nature reviews Genetics.
450 2014;15(6):423-37.
451 21. Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, et al. Ab initio
452 reconstruction of cell type-specific transcriptomes in mouse reveals the conserved
453 multi-exonic structure of lincRNAs. Nature biotechnology. 2010;28(5):503-10.
454 22. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative
455 annotation of human large intergenic noncoding RNAs reveals global properties and specific
456 subclasses. Genes & development. 2011;25(18):1915-27.
457 23. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE
458 v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and
459 expression. Genome research. 2012;22(9):1775-89.
460 24. Qiu MT, Hu JW, Yin R, Xu L. Long noncoding RNA: an emerging paradigm of cancer
461 research. Tumour biology : the journal of the International Society for Oncodevelopmental
462 Biology and Medicine. 2013;34(2):613-20.
463 25. Mattick JS. The genetic signatures of noncoding RNAs. PLoS genetics.
464 2009;5(4):e1000459.
465 26. Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. Cell.

2009;136(4):629-41.

27. Vance KW, Sansom SN, Lee S, Chalei V, Kong L, Cooper SE, et al. The long non-coding RNA Paupar regulates the expression of both local and distal genes. The EMBO journal. 2014;33(4):296-311.

28. Clark MB, Mattick JS. Long noncoding RNAs in cell biology. Seminars in cell & developmental biology. 2011;22(4):366-76.

29. Li J, Xuan Z, Liu C. Long non-coding RNAs and complex human diseases. International journal of molecular sciences. 2013;14(9):18790-808.

30. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nature reviews Genetics. 2013;14(3):204-20.

31. Lee MT, Bonneau AR, Giraldez AJ. Zygotic genome activation during the maternal-to-zygotic transition. Annual review of cell and developmental biology. 2014;30:581-613.

32. Wijnands KP, Chen J, Liang L, Verbiest MM, Lin X, Helbing WA, et al. Genome-wide methylation analysis identifies novel CpG loci for perimembranous ventricular septal defects in human. Epigenomics. 2017;9(3):241-51.

33. Del Duca D, Tadevosyan A, Karbassi F, Akhavein F, Vaniotis G, Rodaros D, et al. Hypoxia in early life is associated with lasting changes in left ventricular structure and function at maturity in the rat. International journal of cardiology. 2012;156(2):165-73.

34. Yan L, Yang M, Guo H, Yang L, Wu J, Li R, et al. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nature structural & molecular biology. 2013;20(9):1131-9.

488 35. Bu D, Yu K, Sun S, Xie C, Skogerbo G, Miao R, et al. NONCODE v3.0: integrative
489 annotation of long noncoding RNAs. Nucleic acids research. 2012;40(Database
490 issue):D210-5.
491 36. Liu H, Su J, Li J, Liu H, Lv J, Li B, et al. Prioritizing cancer-related genes with aberrant
492 methylation based on a weighted protein-protein interaction network. BMC systems biology.
493 2011;5:158.
494 37. Blatti C, Sinha S. Characterizing gene sets using discriminative random walks with
495 restart on heterogeneous biological networks. Bioinformatics. 2016;32(14):2167-75.
496
497 Figure legends
498 Figure 1. An overview of predicting strategy of VSD-associated lncRNA.
499 Figure 2 Differential DAN methylation gene in VSD. (A) Volcano plot of P value as a
500 function of the differential DNA methylation level in normal tissue and VSD sample.
501 Red dots represent that DNA methylation level of genes are not significantly
502 different. Green dots represent that DNA methylation level of genes are significantly
503 different. (B) Heatmap of DNA methylation level for 7292 DMGs. (C) Distinction
504 DNA methylation level between normal tissue and VSD sample. Significance was
505 tested using a Wilcoxon test. (D) DAVID enrichment analysis for DMGs in VSD
506 Figure 3. An overview of DNA methylation lncRAN and gene expression to
507 embryonic development. (A) Heatmap and DAVID enrichment analysis for
508 embryonic methylation genes.(B) Heatmap and DAVID enrichment analysis for
509 embryonic gene expression. (C) Heatmap and DAVID enrichment analysis for

510 embryonic lncRAN.

511 Figure 4. RCWN. (A)DMAN. (B) The degree distribution of DMAN follows a
512 power-low distribution. (C) Comparison of Pearson correlation coefficient between
513 real network and random network.(D) DNA methylation-lncRNA –mRNA
514 co-expression weighted network. (E-F) The degree distribution of co-expression
515 network and comparison of pearson correlation coefficient with permutation.

516 Figure 5 Prediction of VSD-associated lncRAN by DRaWR. (A) An overview of
517 DRaWR. (B) Overlap of gene expression and DNA methylation of embryo and VSD
518 as the 'query set'. (C) The DNA methylation level of 'query set' in six stages of
519 embryo VSD and normal tissue.

520 Figure 6 VSD-associated lncRNAs. (A) The genomic position of the top 10 lncRNAs
521 and genes, the green line means the interaction between gene and gene. The blue line
522 represent the interaction between gene and lncRAN and red ones mean
523 lncRNA-lncRNA co-expression. (B)The subnetwork of top 10 lncRNAs with highest
524 score genes extracted from extracted from global co-expression network(C) The
525 function of lncRNA. Black blocks represent that the lncRNA has the function in
526 NONCODE database.

527 Figure 7 Validation of lncRNA related genes in VSD cardiac tissue. (A)
528 Electropherogram of validated sampes. (B) Bar plot of lncRNA-related gene
529 expression in normal and VSD.

530 Figure S1 : Negative regulation of DNA methylation to gene expression.

531 (A) Venn diagram of detected embryonic DNA methylation gene and embryonic gene

532 expression. (B)The scatterplot model shows that the methylation negatively regulates

533 the gene.

534 Figure S2: lncRNA-mRNA co-expression network in embryo(LMCNe) (A) An

535 overview of (LMCNe) (B) The degree distribution of (LMCNe) follows a power-low

536 distribution

537

538

Embryo

lncRNA expression gene expression DNA methylation

logistic regression test

p<0.05

stages~lncRNA
stages~gene expression
stages~DNA methylation

Ventricular septal defect

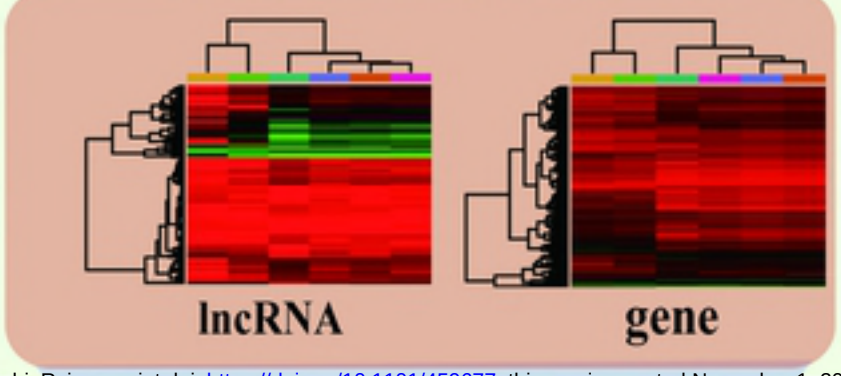
DNA methylation MBD-seq

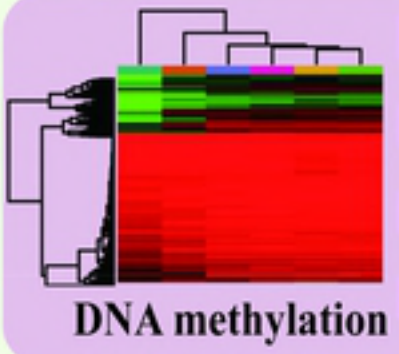
differential methylation CpGs
wilcoxon test p<0.05

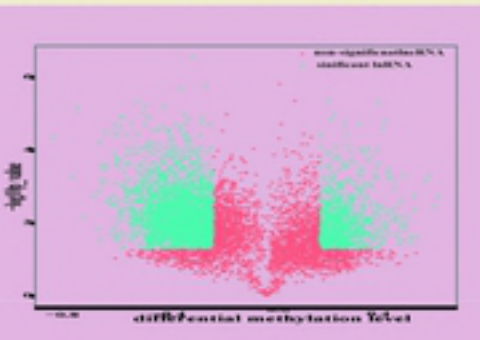
$|M_{(VSD)} - M_{(normal)}| > 0.2$
Student's test p<0.05

differential methylation genes

associated genes and lncRNA







lncRNA-lncRNA

lncRNA-mRNA

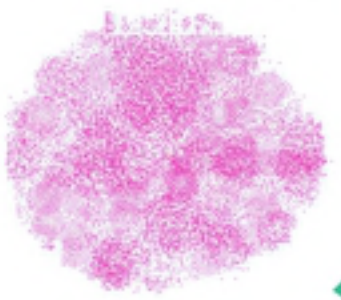
mRNA-mRNA

Pearson correlation analysis

protein-protein interaction network

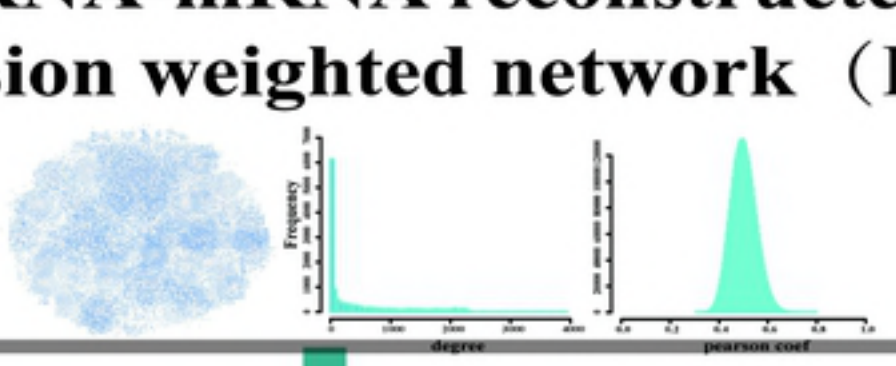
lncRNA-mRNA co-expression network（LMCNe）

DNA methylaiton associated network(DMAN)



weight of edges

lncRNA-mRNA reconstructed co-expression weighted network（RCWN）



inferring VSD-associated lncRNAs by DRaWR

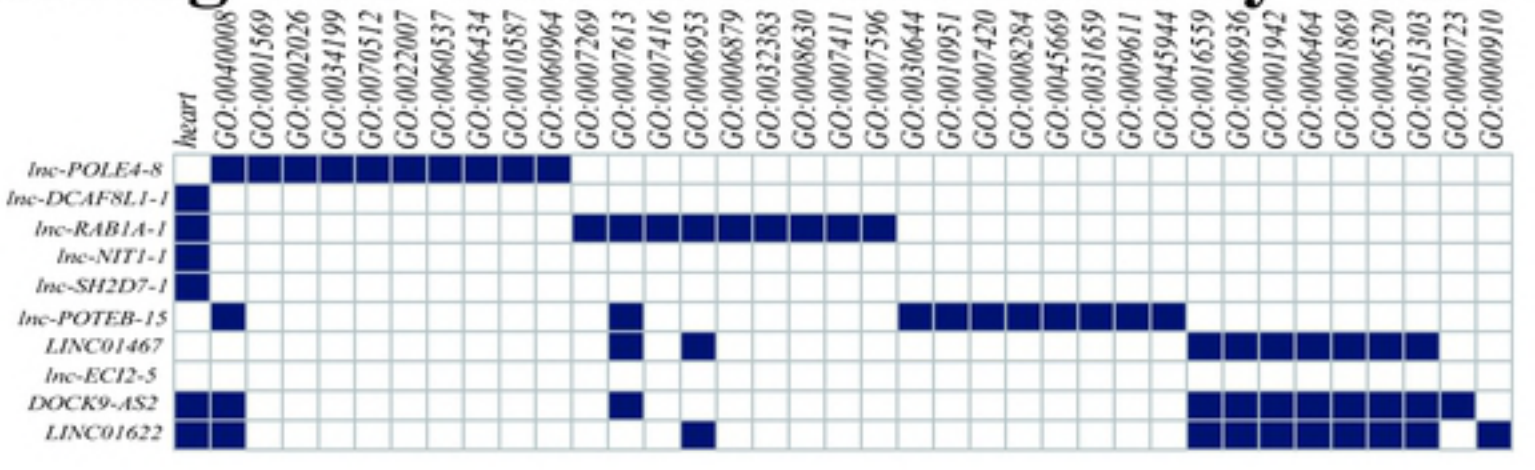
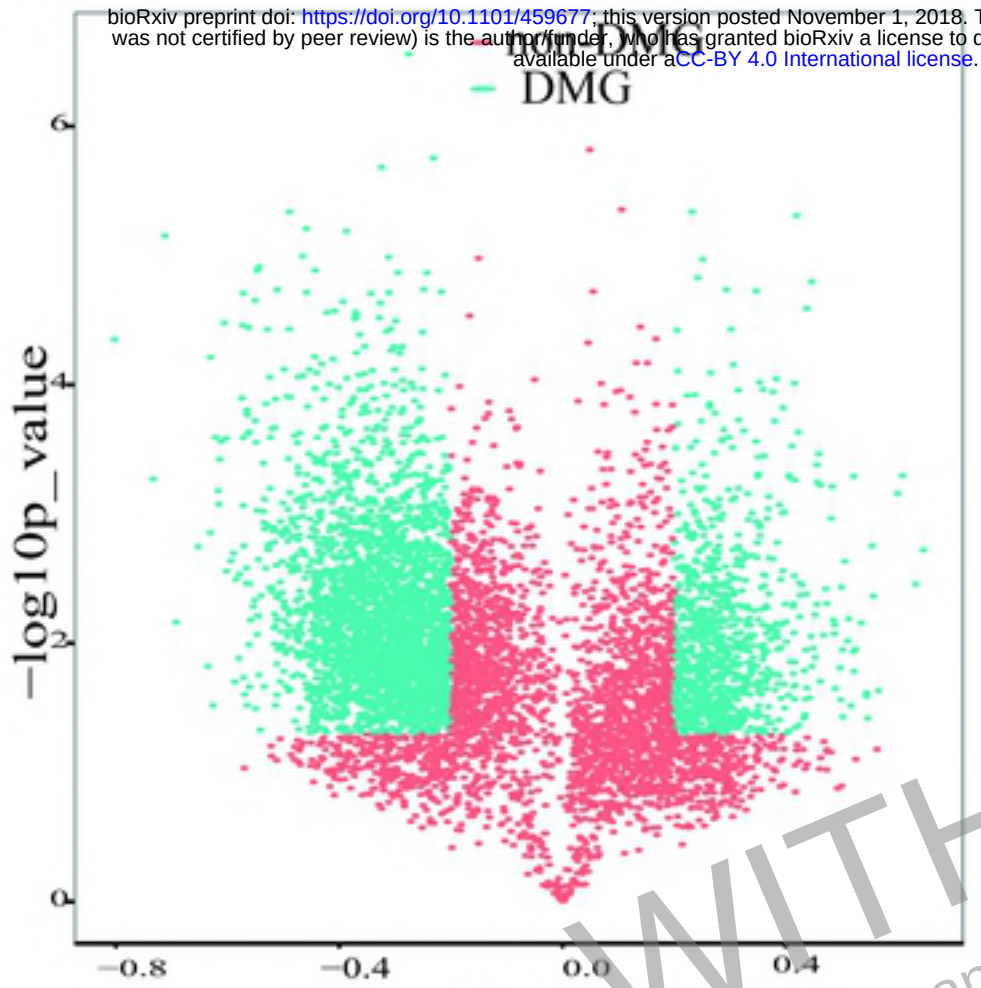
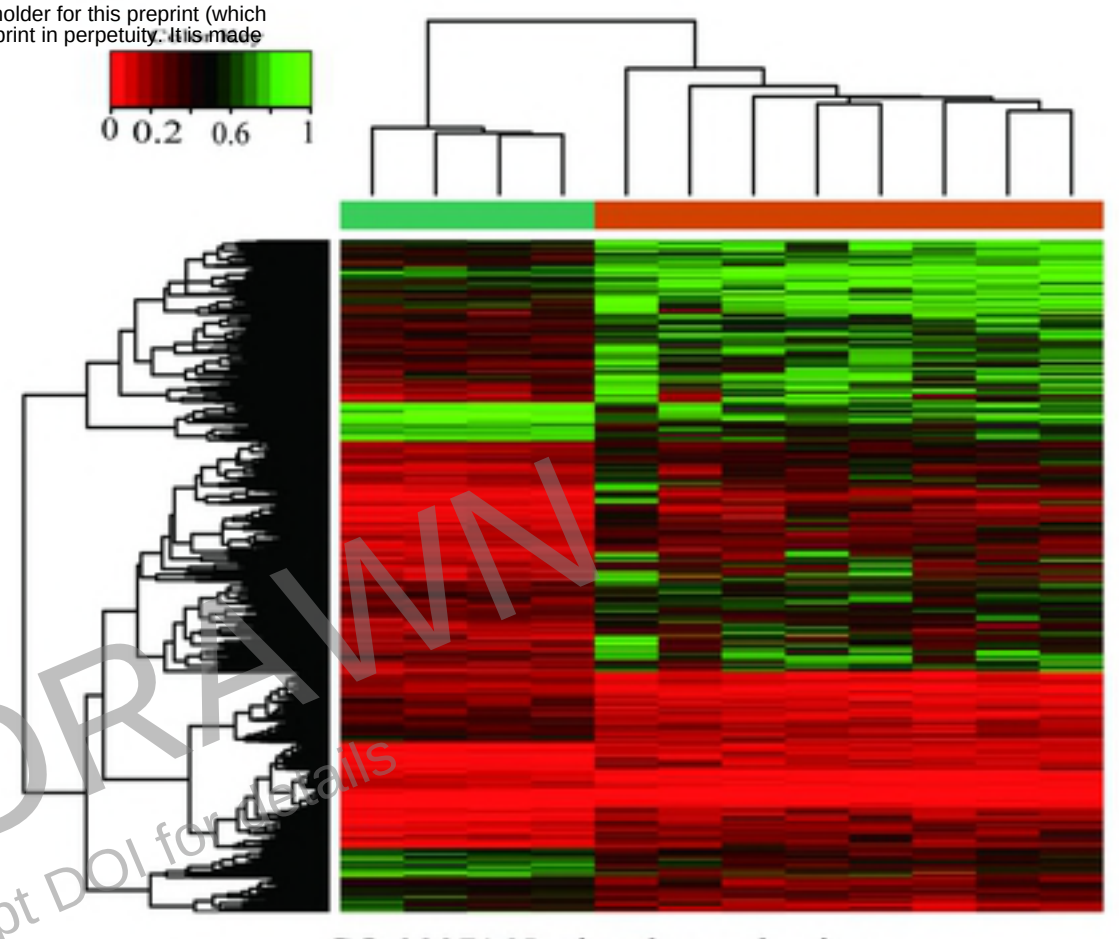


figure1

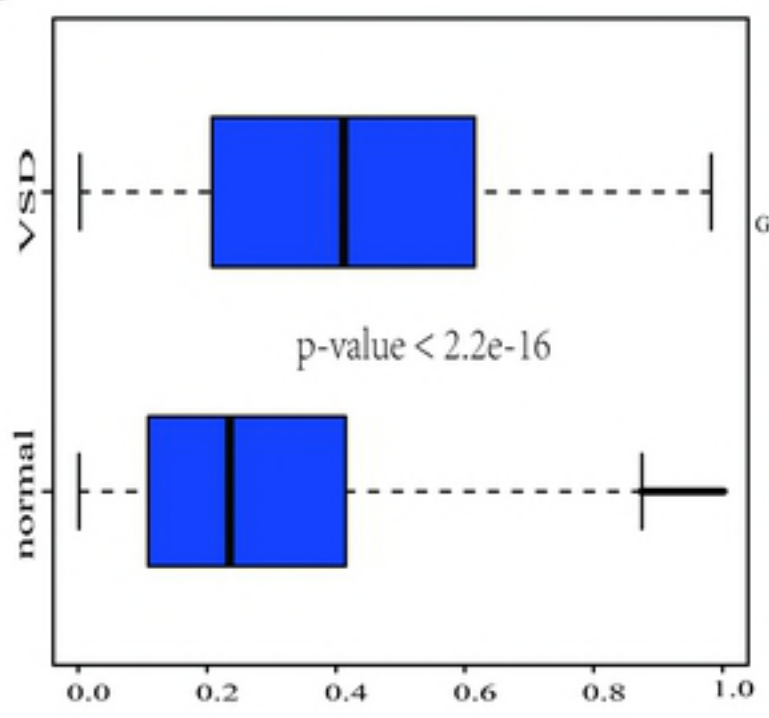
A



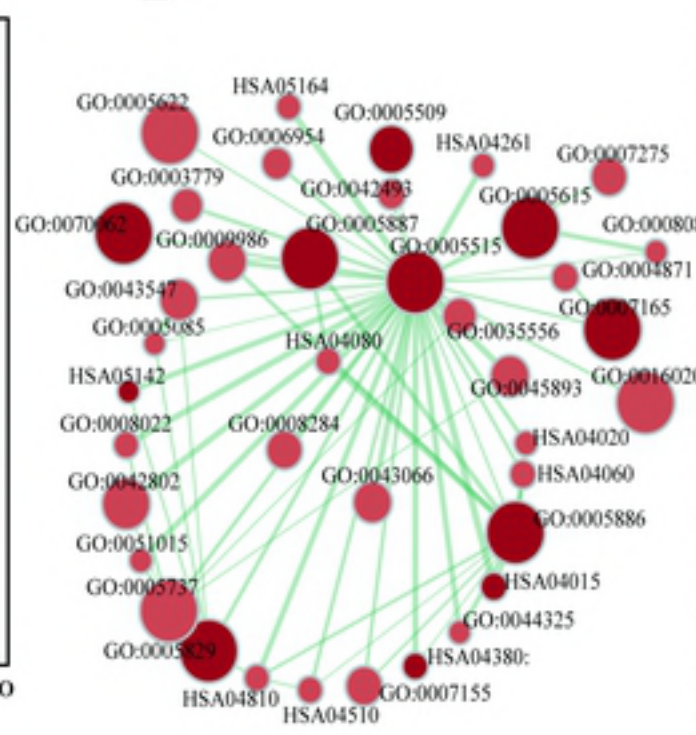
B



C

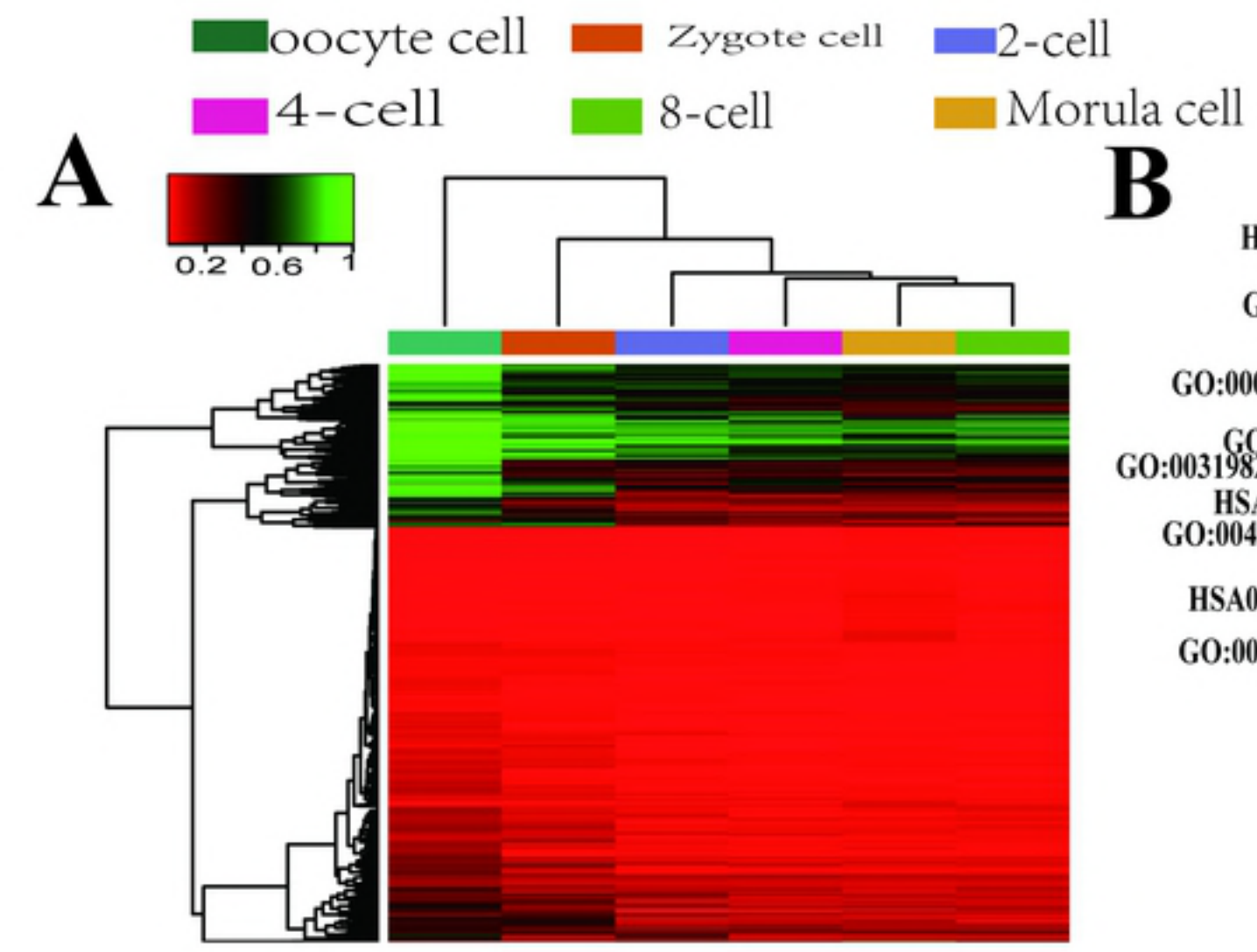


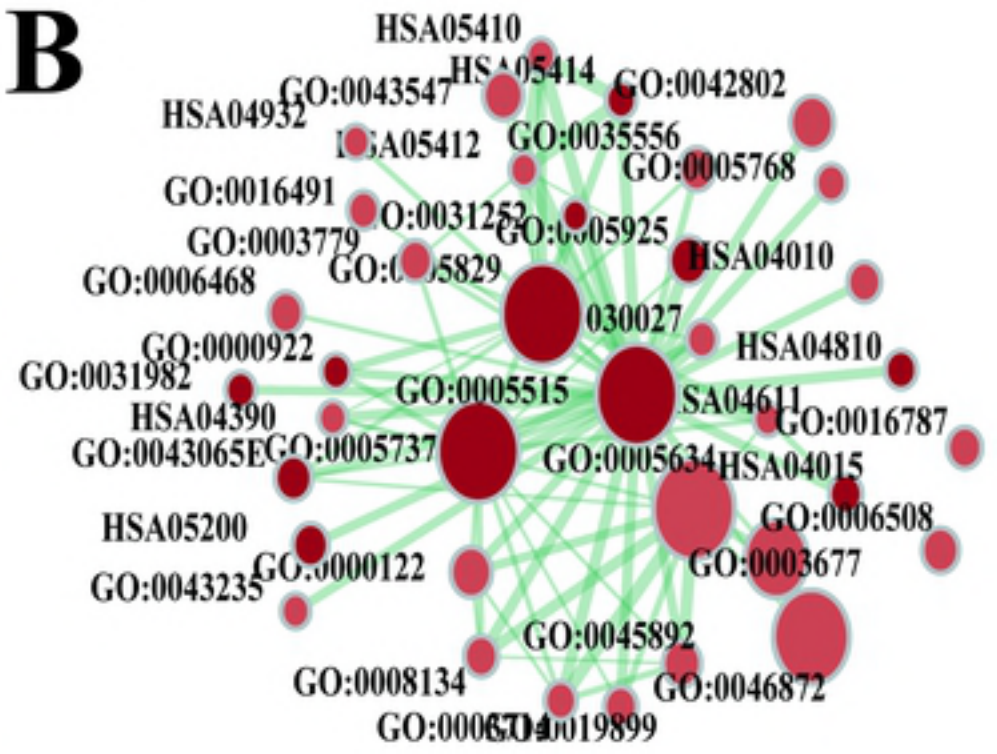
D

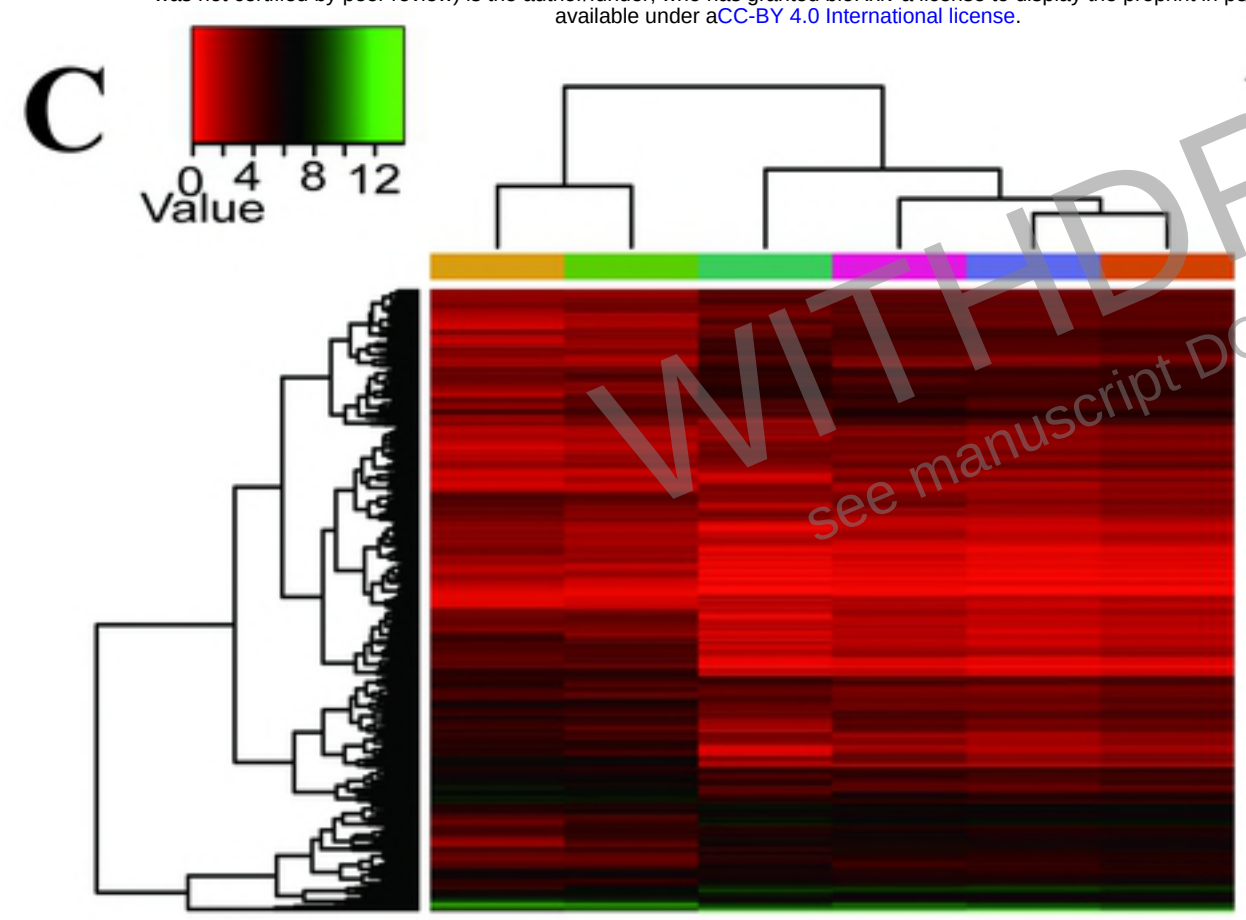


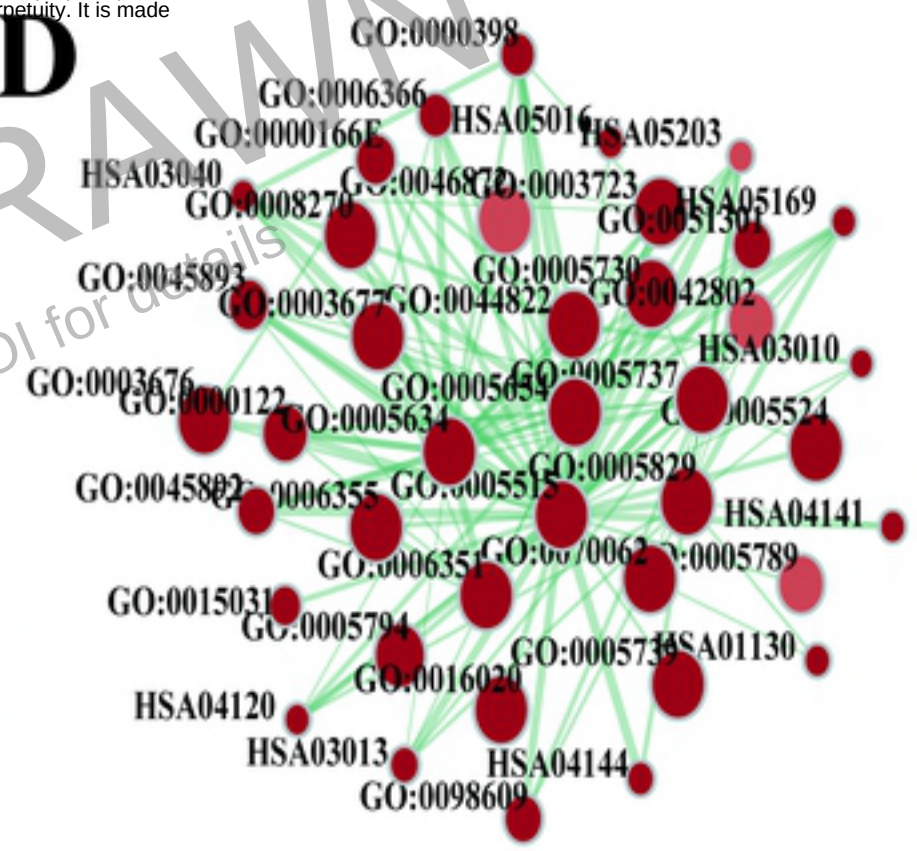
GO:0007165~signal transduction
GO:0043547~positive regulation of GTPase activity
GO:0045893~positive regulation of transcription, DNA-templated
GO:0007275~multicellular organism development
GO:0007155~cell adhesion
GO:0008284~positive regulation of cell proliferation
GO:0043066~negative regulation of apoptotic process
GO:0035556~intracellular signal transduction
GO:0006954~inflammatory response
GO:0042493~response to drug
hsa04080:Neuroactive ligand-receptor interaction
hsa04060:Cytokine-cytokine receptor interaction
hsa04015:Rap1 signaling pathway
hsa04810:Regulation of actin cytoskeleton
hsa04510:Focal adhesion
hsa05164:Influenza A
hsa04020:Calcium signaling pathway
hsa04261:Adrenergic signaling in cardiomyocytes
hsa04380:Osteoclast differentiation
hsa05142:Chagas disease (American trypanosomiasis)

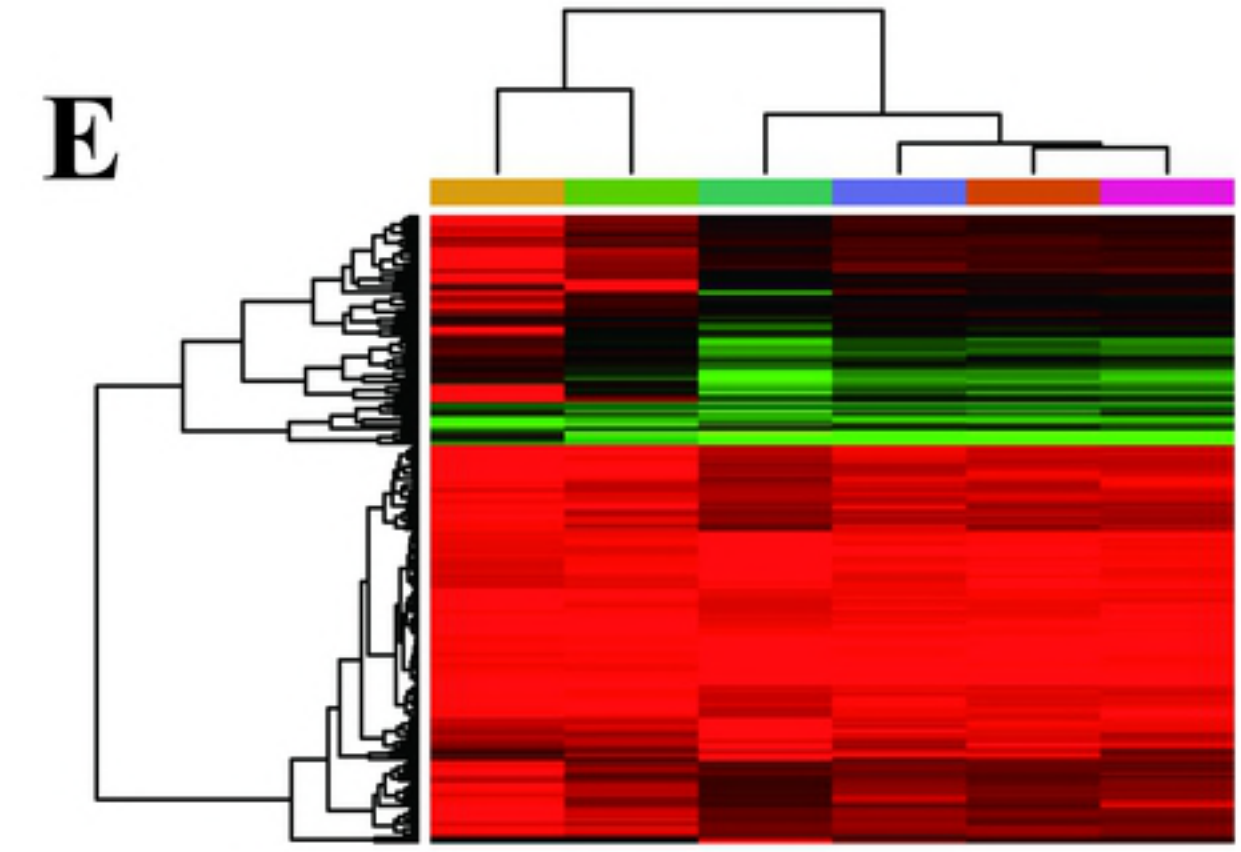
figure2











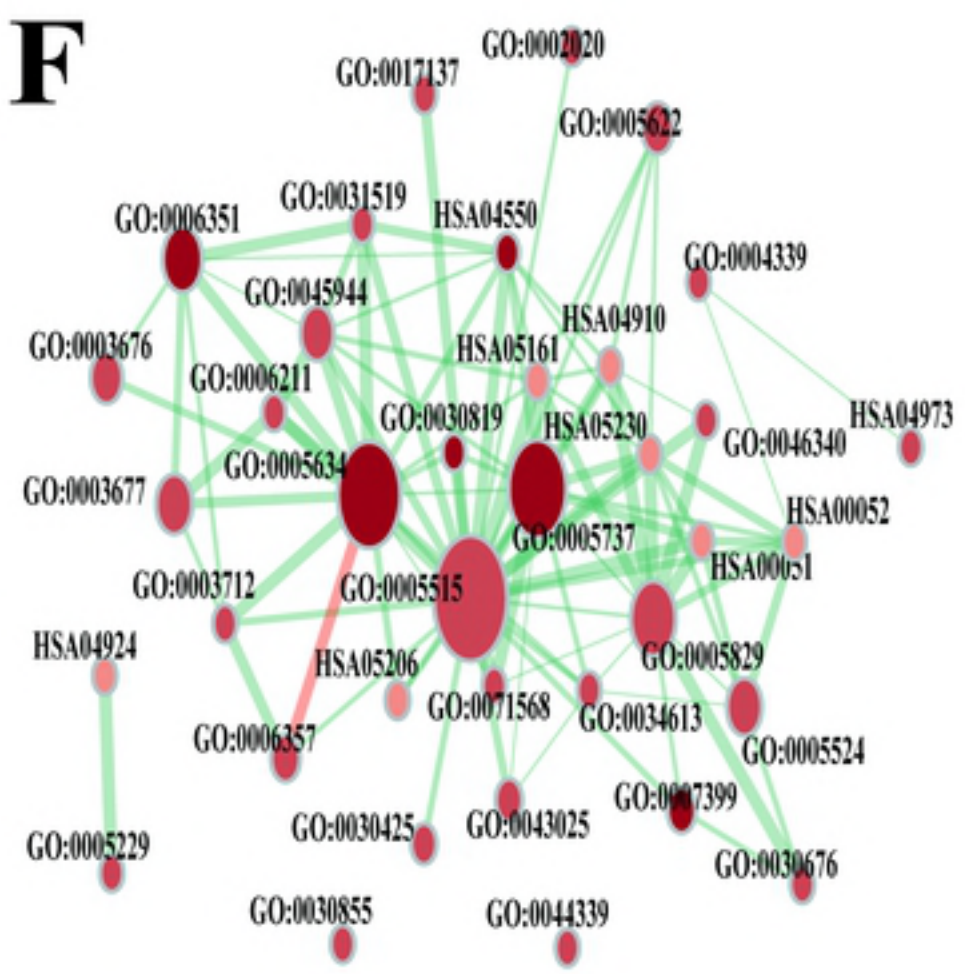


figure 3

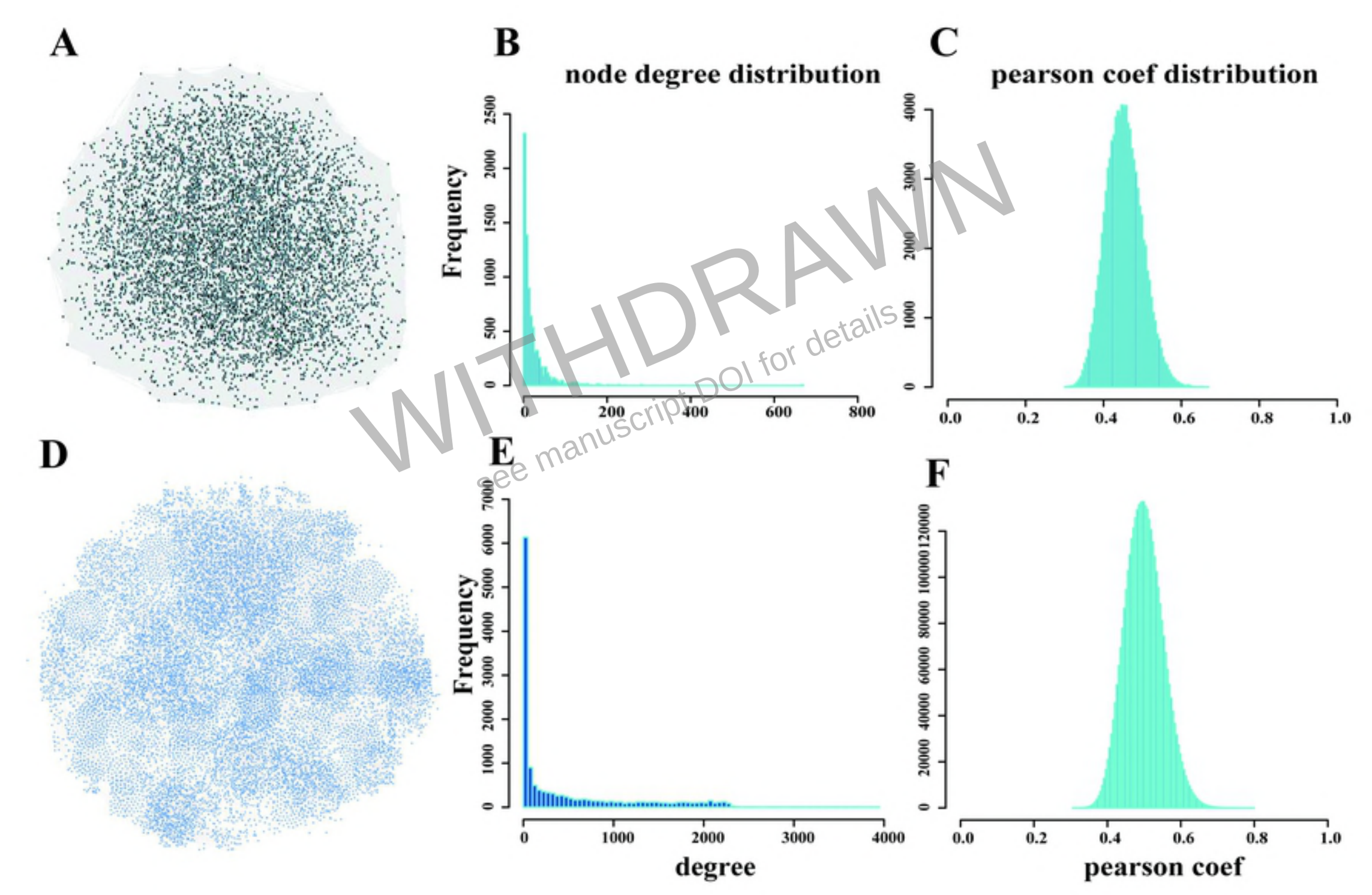
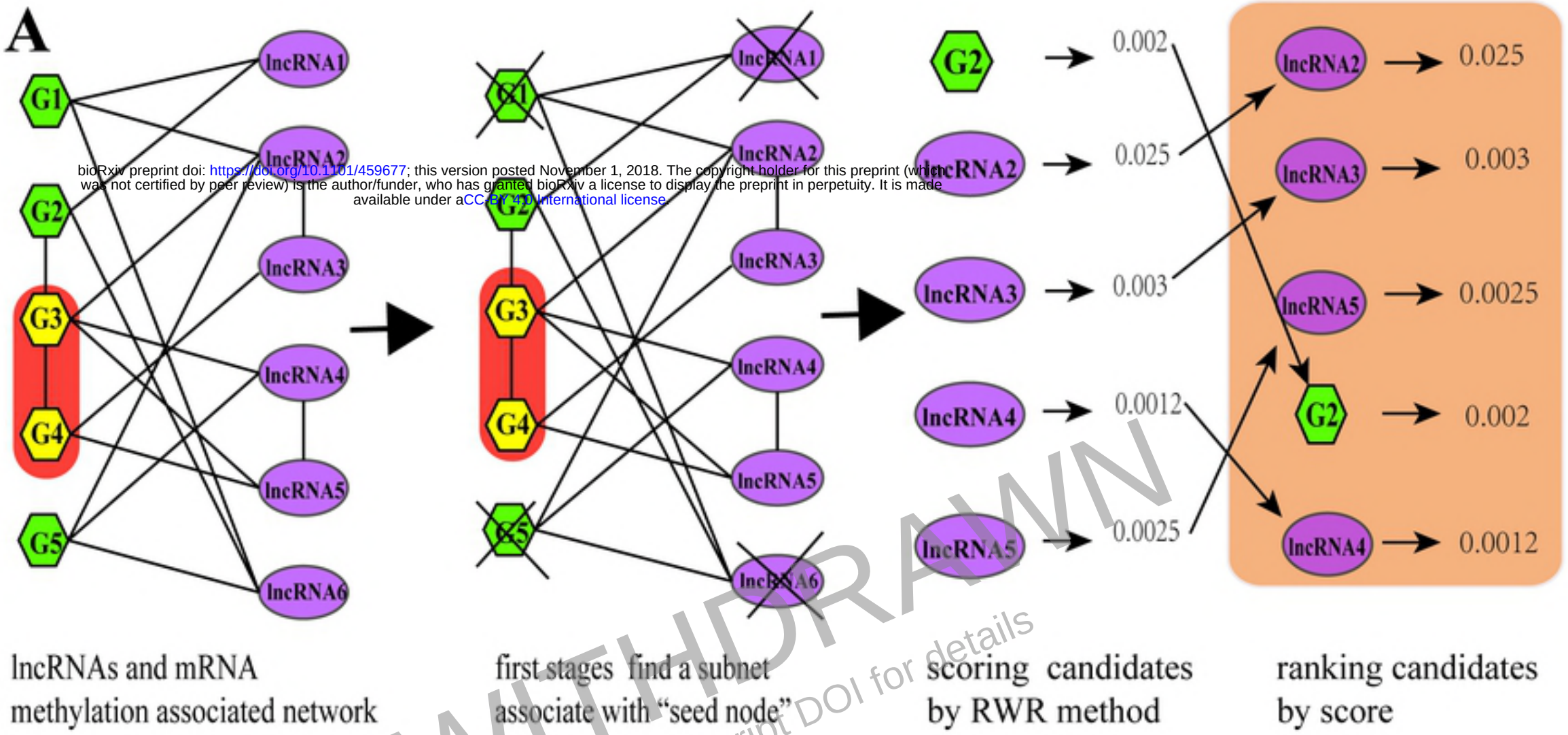


figure 4



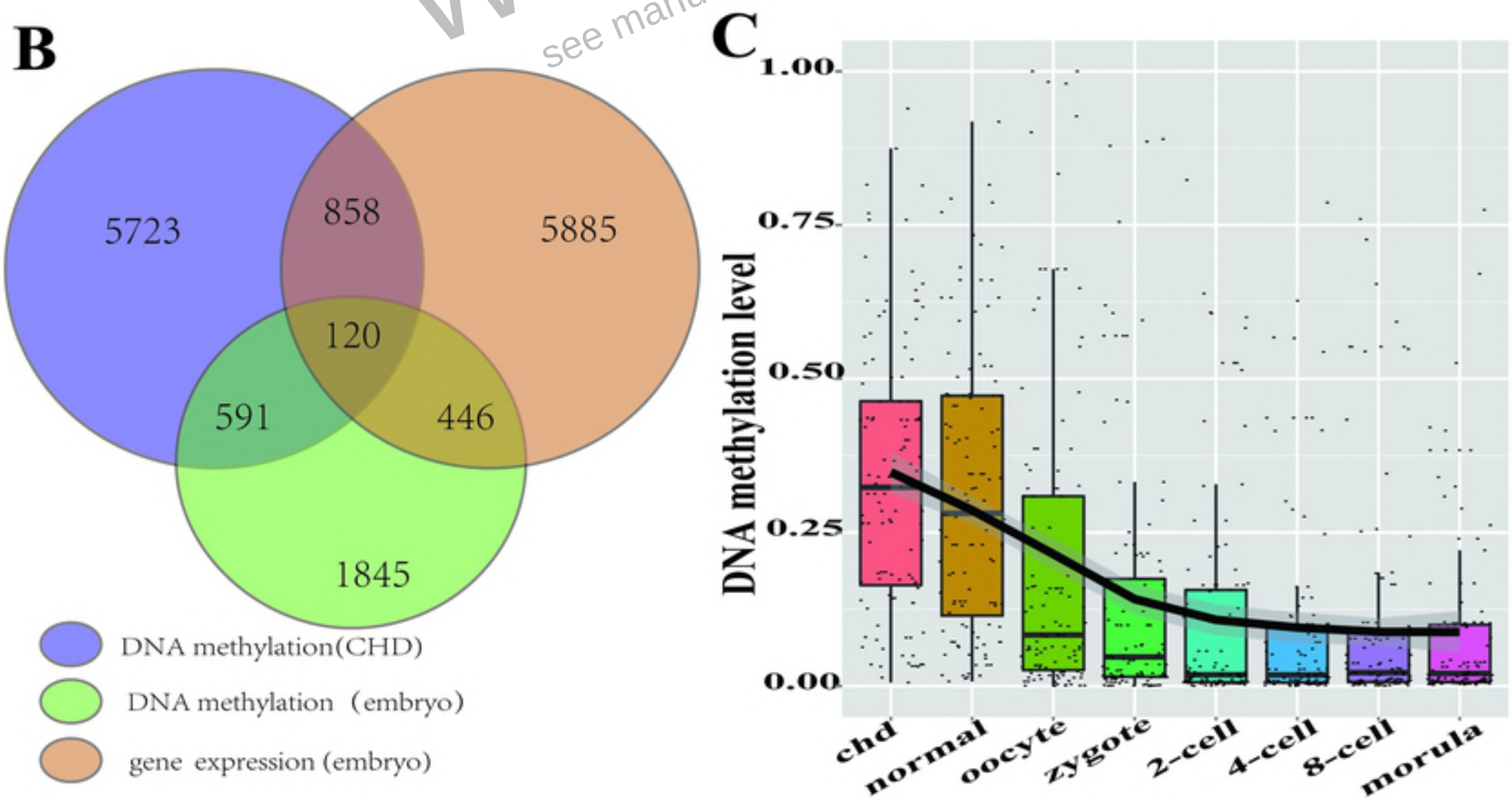
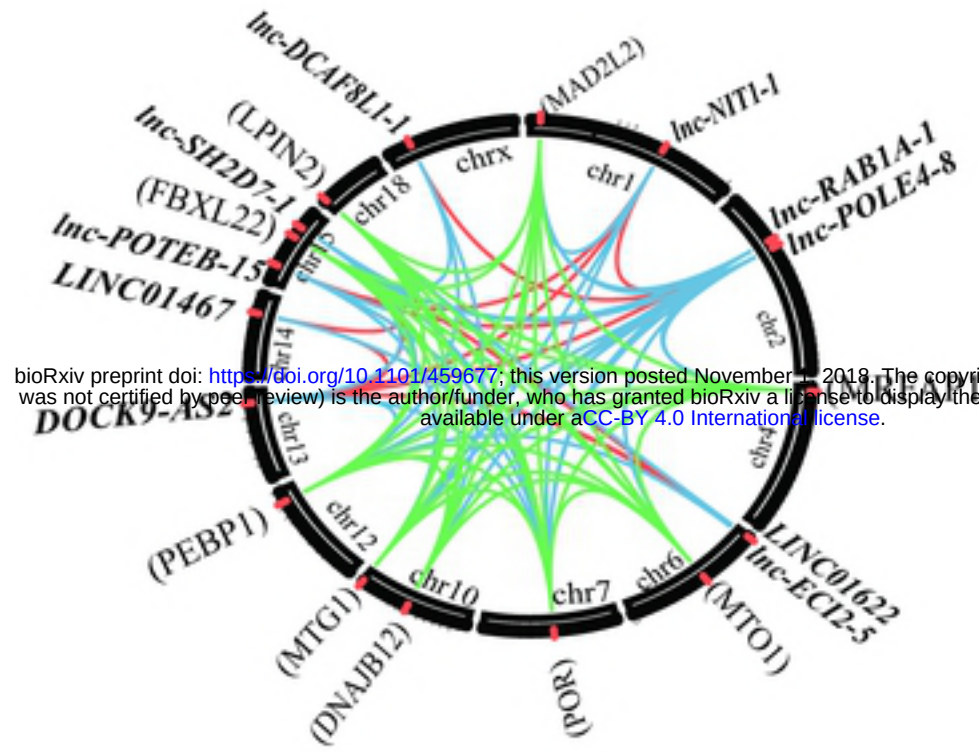
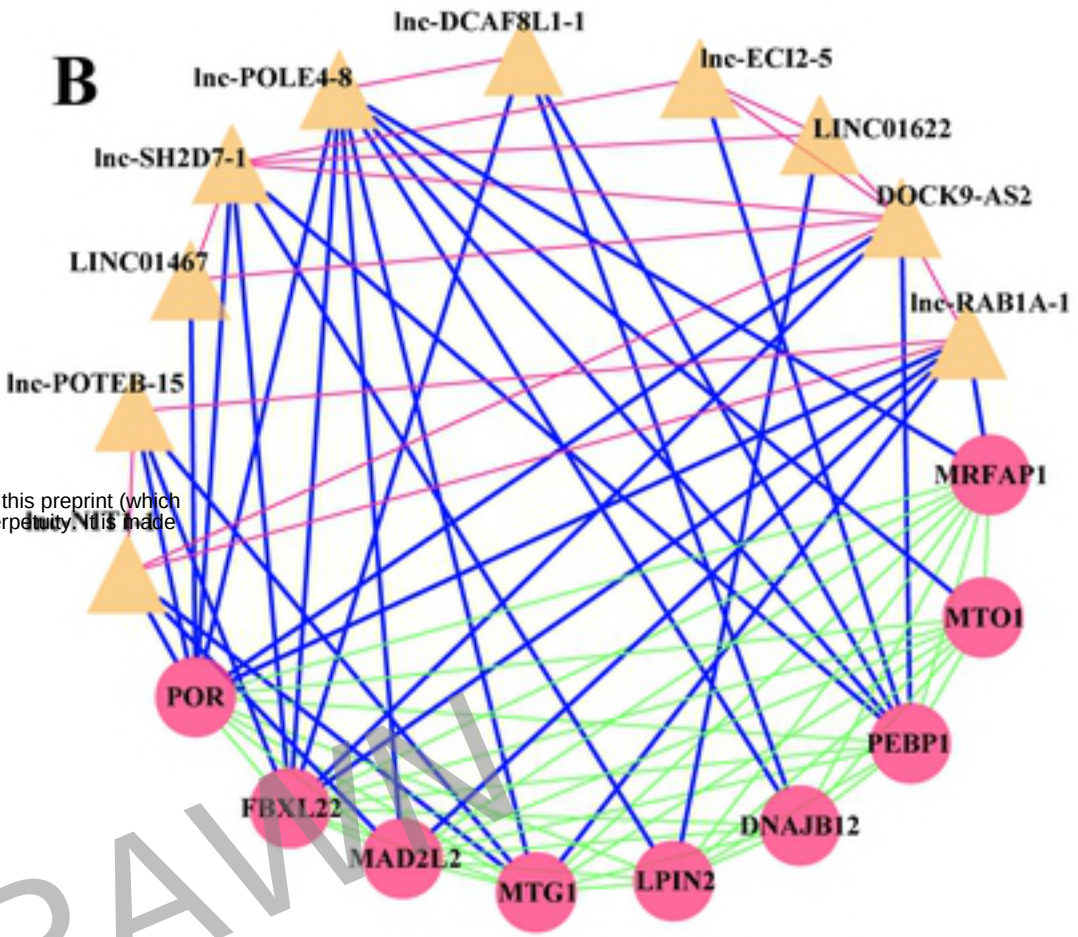


figure 5

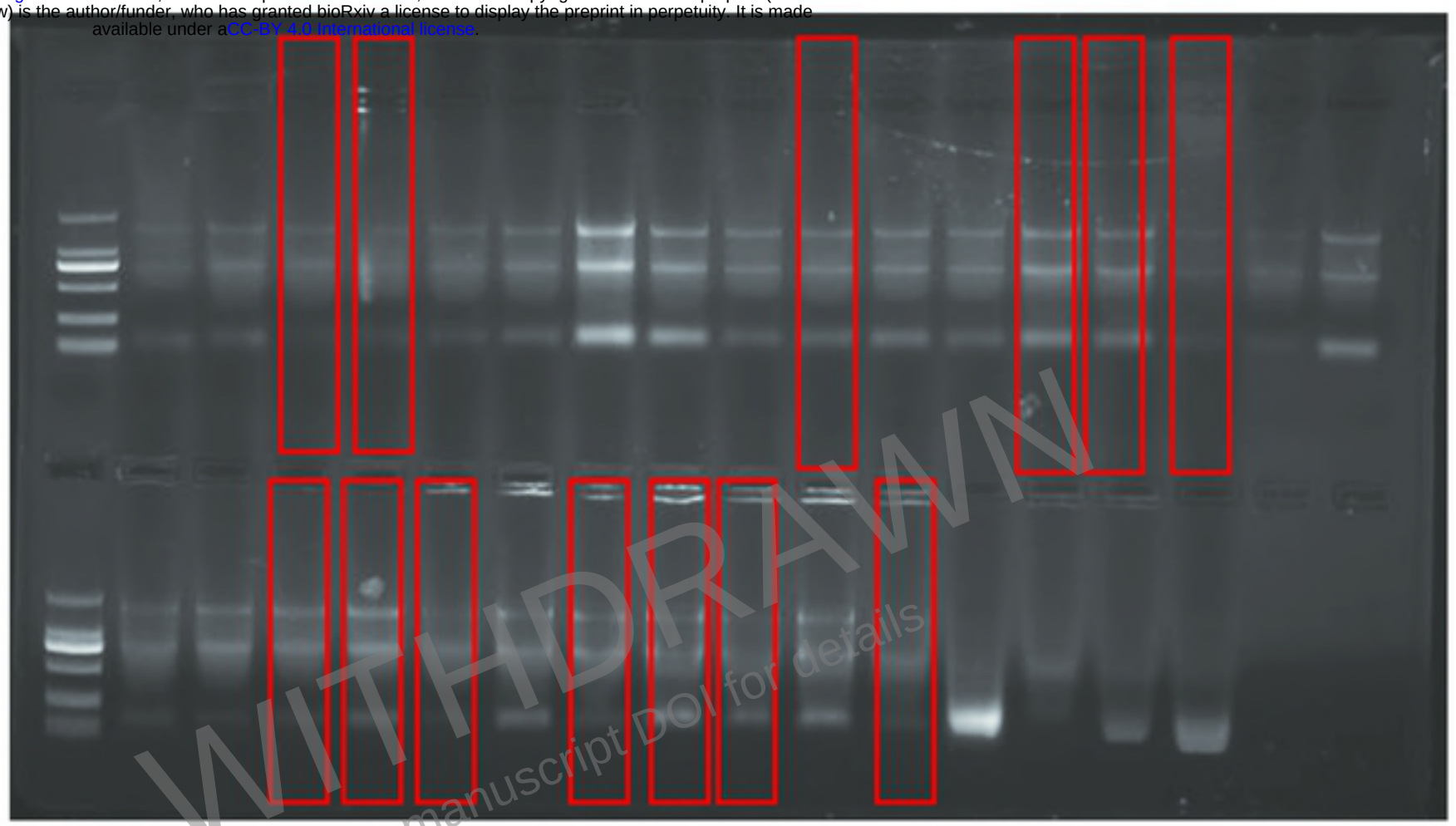




GO:0040008 regulation of growth
GO:0002026 regulation of the force of heart contraction
GO:0007613 memory
GO:0006953 acute-phase response
GO:0031659 positive regulation of cyclin-dependent protein serine/threonine kinase activity involved in G1/S transition of mitotic cell cycle
GO:0009611 response to wounding
GO:0045944 positive regulation of transcription from RNA polymerase II promoter
GO:0016559 peroxisome fission
GO:0006936 muscle contraction
GO:0006464 cellular protein modification process
GO:0001869 negative regulation of complement activation, lectin pathway
GO:0006520 cellular amino acid metabolic process
GO:0051303 establishment of chromosome localization

figure 6

A



SS

B

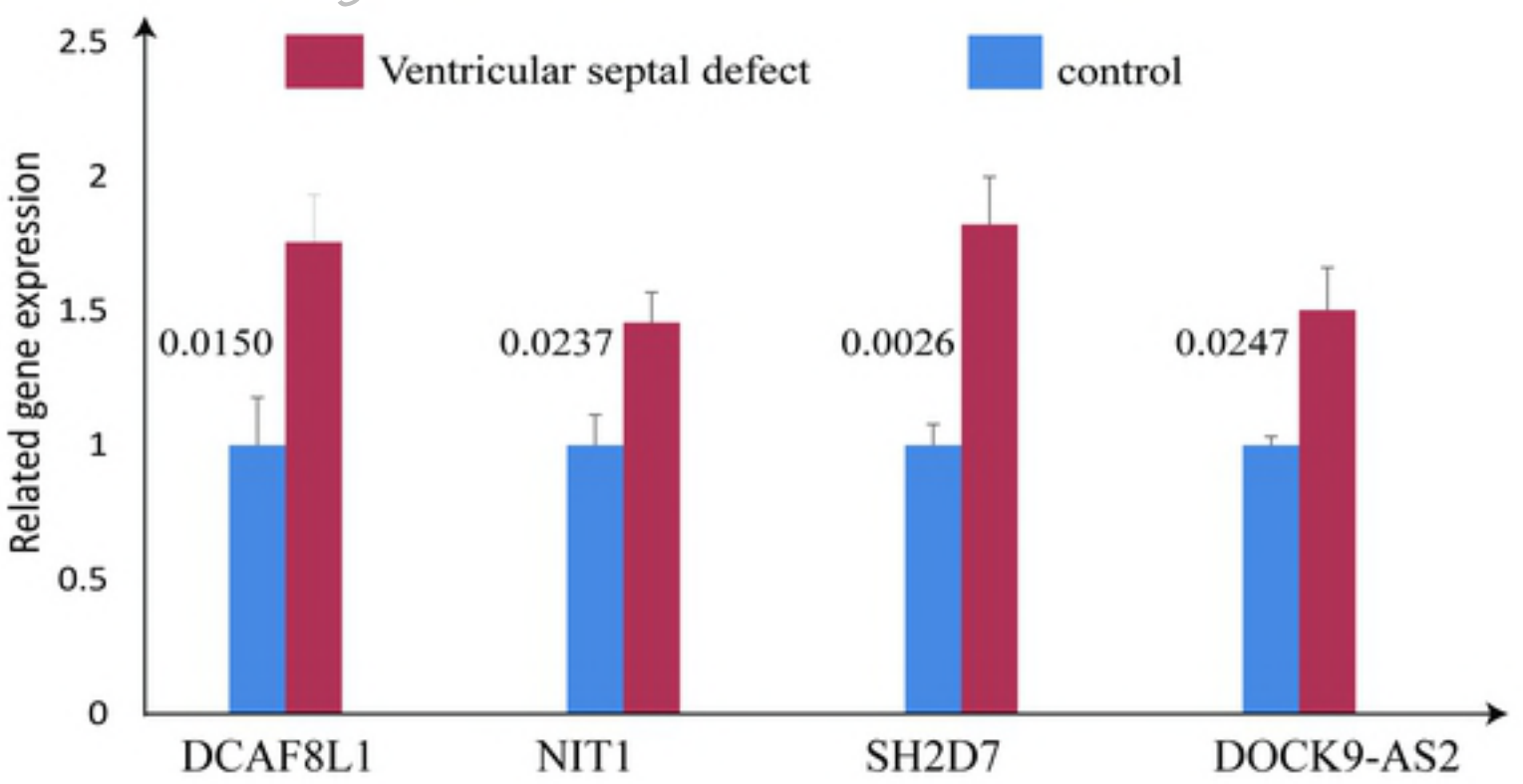


figure 7

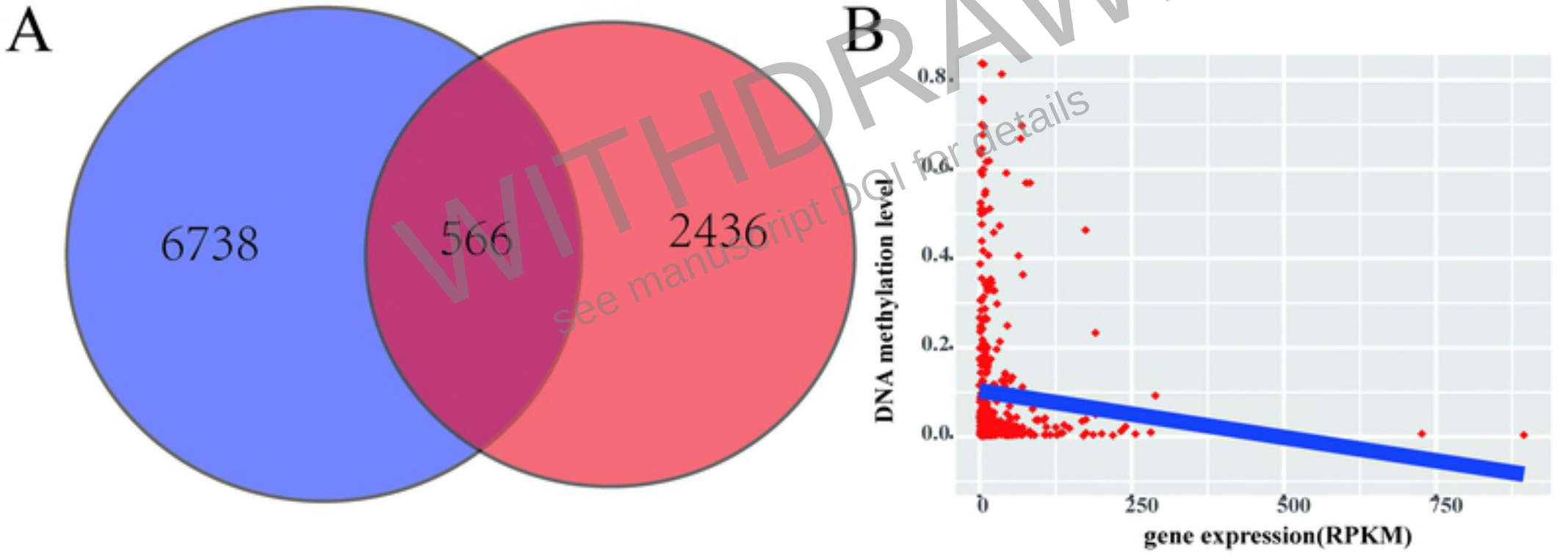


figure S1

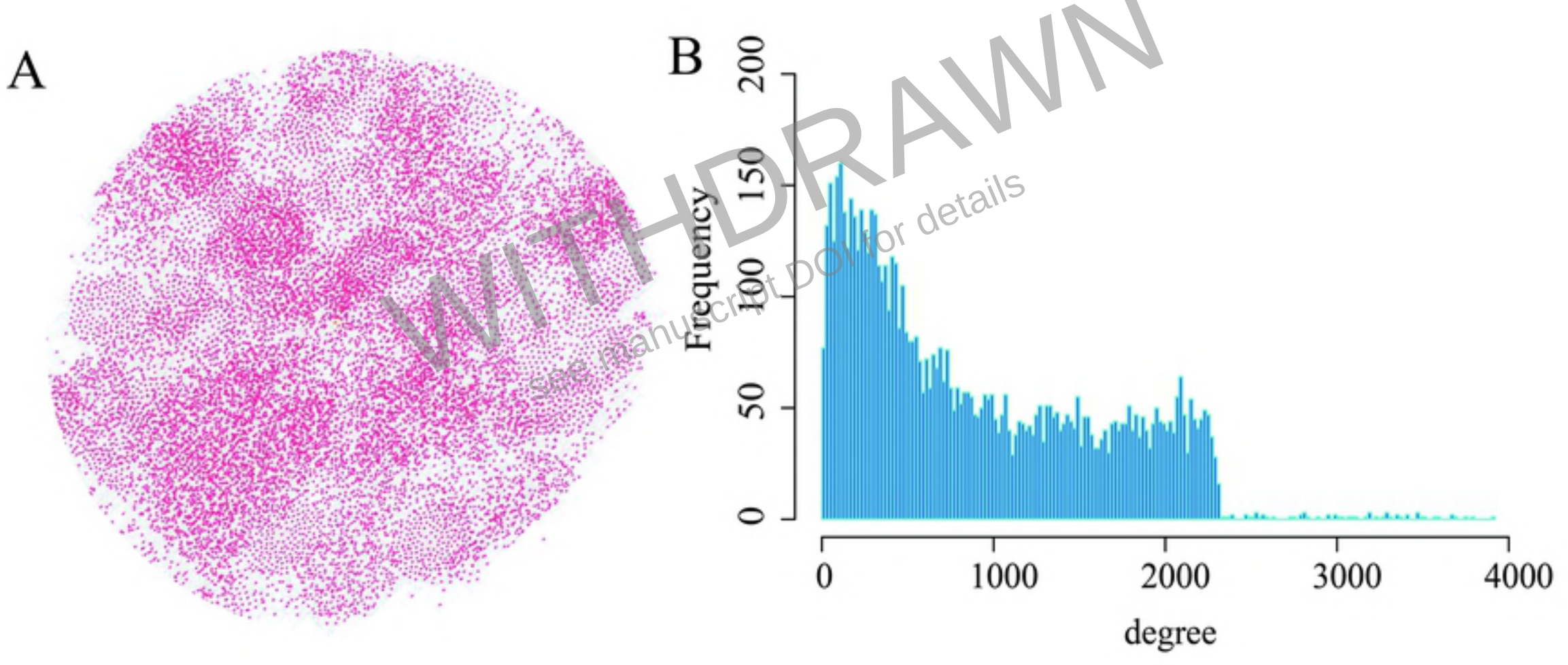


figure S2