Establishment of a Combined Diagnostic Model of Abdominal Aortic Aneurysm with Random Forest and Artificial Neural Network

YIXUAN DUAN
Xi'an Jiaotong University Second Affiliated Hospital
Enrui Xie
Xi'an Jiaotong University Second Affiliated Hospital
Chang Liu
Xi'an Jiaotong University Second Affiliated Hospital
Jingjing Sun
Xi'an Jiaotong University Second Affiliated Hospital
Jie Deng (jie.deng@mail.xjtu.edu.cn)
Xi'an Jiaotong University Second Affiliated Hospital

Research

Keywords: abdominal aortic aneurysm, bioinformatics, diagnostic model, biomarkers, machine learning

Posted Date: September 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-864615/v1

Version of Record: A version of this preprint was published at BioMed Research International on March 7th, 2022. See the published version at https://doi.org/10.1155/2022/7173972.

Abstract

Background Abdominal aortic aneurysm (AAA), a disease with high mortality, is limited by the current diagnostic methods in the early screening. This study aimed to construct a diagnostic model for AAA by using a novel machine learning method, i.e., an ensemble of the random forest (RF) algorithm and an artificial neural network (ANN) (RF-ANN), to identify potential AAA-associated genetic biomarkers.

Methods Through a search of the Gene Expression Omnibus (GEO) database, two large-sample gene expression datasets (GSE57691 and GSE47472) were identified and downloaded. The differentially expressed genes (DEGs) between the AAA and normal control samples were identified, followed by Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis using the Database for Annotation, Visualization, and Integrated Discovery (DAVID). Then, RF-ANN was used to identify the key genes from the DEGs, and an AAA diagnostic model was established. Finally, the diagnostic performance of the model was assessed using the area under the receiver operating characteristic curve (AUC) with GSE47472 as a test dataset.

Results Using GSE57691, we obtained 2486 DEGs, 52 biological process annotations, 17 cellular component annotations, 17 molecular function annotations, and 13 significantly enriched KEGG pathways. Out of these DEGs, we further identified 74 key candidate feature genes by using the RF machine learning algorithm. The weight of each key gene was calculated by the ANN with GSE57691 as a training dataset to construct an AAA diagnostic model. A transcription factor (TF) regulatory network of key genes was constructed. Finally, GSE47472 was used to validate the model. The AUC value was 0.786, indicating that the model had a highly satisfactory diagnostic performance.

Conclusion Potential AAA-associated gene biomarkers were identified, and a diagnostic model of AAA was established. This study may provide a valuable reference for early clinical diagnosis and the search for therapeutic targets of AAA.

1. Background

AAA refers to a permanent and irreversible enlargement of the abdominal aorta to a diameter of 3 cm or larger, exceeding the normal diameter by more than 50% [1]. Although it is usually asymptomatic before enlargement, AAA is naturally progressive, leading to a high risk of irreversible aneurysmal growth and unpredictable rupture at any time. Therefore, it is a life-threatening disease. An aneurysm ruptures when the aortic wall cannot withstand the pressure in the aneurysmal cavity, and rupture leads to a high mortality rate of up to 80% [2]. Approximately 150,000-200,000 deaths are associated with AAA worldwide every year [3]. With the aging of the population as well as the improvements in living standard and diagnostic techniques, an increasing incidence of AAA has been reported in recent years. Early diagnosis of AAA before rupture can reduce the risk of death associated with this disease.

Conventionally, AAA is diagnosed based on imaging findings that confirm the presence of an aneurysm, and the first-choice imaging method for AAA screening is abdominal ultrasound. However, the accuracy

of ultrasound diagnosis depends on the operator's experience and skill. The measured diameter of an aneurysmal cavity varies with the direction of the scanning plane, resulting in difficulty obtaining the details of the aneurysm. Other factors, such as patient compliance, obesity, or intestinal gas accumulation, can also significantly affect the accuracy of ultrasound diagnosis [4]. The application of computed tomography angiography and magnetic resonance angiography is limited to AAA screening due to such disadvantages as their use of contrast agents, radiation damage, and high cost. A previous study showed that early screening for AAA can reduce the AAA-associated mortality by approximately 50% in men [5]. In 2005, the guidelines on screening for AAA published by the American Heart Association recommended that men aged 60 or older who have siblings or offspring with AAA should be given a physical examination for AAA [6]. Considering the lack of effective examination methods for early AAA screening and diagnosis, as well as the lack of sensitive and specific biomarkers that can be used in clinical practice, it is crucial to develop a model for early diagnosis and screening of AAA [7].

The etiology of AAA is complex. AAA is associated with various factors, including smoking, male sex, advanced age, atherosclerosis, hyperlipidemia, race, chronic obstructive pulmonary disease, and family history [8], etc. It may result from both genetic and environmental factors [9]. The risk of AAA nearly doubles if the patient has a family genetic history[10]. Previous AAA studies have confirmed several susceptibility genes that can help diagnose AAA, including *CTLA4*, *NKTR*, *CD8A*, *CANX*, *CD44*, *DAXX*, and *STAT1* [11–13]. Therefore, the search for AAA susceptibility genes has become an important research direction for AAA screening and diagnosis.

With the rapid progress of gene sequencing technology in recent years, massive data have been generated for gene-related research. The advent of “big data” has brought new opportunities for disease diagnosis, as well as challenges in how to apply these data effectively and efficiently. With their continuous optimization, machine learning algorithms have become powerful tools for data utilization thanks to their high classification accuracy and convenient use. Among machine learning methods, random forest (RF) [14] algorithms and artificial neural networks (ANNs) [15] have shown particularly strong computing power. This study aimed to use a novel method, i.e., an RF-ANN ensemble, for AAA risk factor screening and establishment of an AAA diagnostic model. The findings of this study provide potential biomarkers for early clinical screening of AAA.

2. Results

2.1 Screening of DEGs

Differential expression analysis derived 2486 DEGs in GSE57691 and 1464 in GSE47472 when using false discovery rate (FDR) < 0.001 as the threshold (Supplementary Table 1 and Supplementary Table 2). There were 178 DEGs shared by both datasets. Figure 1(a) and Fig. 1(b) are the heatmaps of DEGs in GSE57691 and GSE47472, respectively. Both show satisfactory separation of gene expression. Figure 1(c) shows the volcano plot of average gene expression levels.

2.2 GO and KEGG enrichment analysis of DEGs

All 2486 DEGs were imported into DAVID 6.8 for functional enrichment analysis (Supplementary Table 3 and Supplementary Table 4). GO analysis of the DEGs yielded 86 enriched annotations, including 52 biological process (BP), 17 cellular component (CC), and 17 molecular function (MF), as well as 13 KEGG enriched pathways. DEGs were significantly enriched for negative regulation of growth (FDR) = 5.41E-02), positive regulation of cell death (FDR = 1.52E-01), and cellular response to calcium channels (FDR = 3.05E-01). In the GO enrichment analysis, protein binding was the subcategory with the highest number of DEGs; in KEGG pathway analysis, metabolic pathways included the highest number of DEGs (Fig. 2).

2.3 Diagnostic feature genes identified and classified by RF

To identify DEGs that were more reliable, the R package RandomForest was used to further screen the 2486 DEGs, and its classification performance was validated using the GSE47472 data. The classification was optimal when the number of variables was three and the optimal tree number was set at 100. A mean decrease accuracy > 0.001 and mean decrease Gini > 0.05 were key thresholds used for screening, which yielded 74 DEGs (Supplementary Table 5 and Fig. 3). In the training dataset, all 74 genes were clustered satisfactorily except in one control sample; in the validation dataset, the clustering of these genes was fully satisfactory.

2.4 Construction of a transcription factors (TFs) regulatory network of feature genes

The 74 feature genes selected by RF were used to construct a network (Fig. 4). The 74 DEGs formed 1084 interaction pairs (Supplementary Table 6). The WGCNA package of R was used to calculate the pairwise correlations between the 74 genes and conduct the WGCNA. The relevant genes with correlation values > 0.1 were selected to construct the TF network (Supplementary Table 7).

2.5 Construction and validation of the ANN model

The 74 DEGs selected by RF were used to construct a neural network with the GSE57691 dataset. The weight of each gene was calculated for optimal differentiation between the AAA and control samples. A diagnostic model was then constructed based on the weights of the genes and the neural network (Fig. 5). Prediction by the model had an AUC of 0.786 in GSE47472 and 1 in the original dataset GSE57691 (Supplementary Table 8 and Supplementary Table 9), suggesting that the ANN is highly stable in diagnosing AAA (Fig. 6). These findings show that we successfully constructed an AAA diagnostic model through the differential gene expression between AAA and control samples.

2.6 Comparison of the proposed model with existing diagnostic models

Most of the existing diagnostic models for AAA are rupture risk prediction models, e.g., models that predict rupture risk by simulating fluid-structure interaction [16]. To our knowledge, there are no models for early clinical screening of AAA. Computational modeling and computer hardware have become important tools in all aspects of health care, including identification of AAAs with a high risk of rupture. Rupture risk can be predicted by calculating relevant physical parameters (e.g., shear stress) after three-dimensional reconstruction by computed tomography or magnetic resonance imaging [17]. Several researchers developed an *in vitro* numerical validation model, which they used to confirm the locations of ruptures in silicone rubber models simulating AAAs, based on high-speed photography that captured the moment of rupture and finite element analysis [18]. The experimental observations and the model's calculations were highly consistent, and finite element analysis accurately predicted the rupture location in 90% of the models. The disadvantage was that the prediction performance highly depended on the quality of initial images. In the present study, for the first time, we applied the machine learning method to create a predictive model for early screening and diagnosis of AAA and tested that model's applicability to early AAA prevention and screening. This model can contribute to the primary prevention of the disease and reduce the incidence of risk events.

With regard to the existing biomarkers, a meta-analysis has summarized all existing evidence on the association between hemostatic markers (such as fibrinogen) and the presence and size of AAAs [19]. A significant correlation has been found between aortic diameter and D-dimer concentration in the blood [7]. While blood markers can only be used to predict the relationship with the occurrence and diameter of AAA, the model constructed in the present study possesses advantages for early screening.

3. Discussion

AAA is an important public health problem. The mortality of patients with a ruptured aneurysm that goes untreated is extremely high, as evidenced by a randomized controlled study reporting that the mortality rate was 80% in all patients who were sent to hospitals and 50% in patients who received emergency repair surgery [20]. Therefore, clinical management of AAA is critical. Clinical management of AAA currently focuses on three key aspects: screening, diagnosis, and monitoring and surgical intervention. A number of studies have proposed predictive models for monitoring the progression of AAA. For example, Liang *et al*. [21] used a machine learning method to predict the rupture risk of ascending aortic aneurysm based on the shape characteristics of the aneurysm. This machine learning–based method was much faster than finite element analysis. Another study developed an auxiliary tool to assess the possibility of AAA rupture and predict the progression of AAA through machine learning [22]. Even so, screening and diagnosing aortic aneurysms remain challenging. One reason is that aneurysms are asymptomatic: they usually remain clinically silent unless the aneurysmal cavity grows larger rapidly or is acutely ruptured, or thrombosis has occurred in a distal artery [23]. Another reason is that the existing examination technologies have certain limitations when applied for early screening and diagnosis of AAA, and there are no good predictive clinical indicators. Therefore, to achieve early diagnosis and treatment of AAA before rupture, it is necessary to develop new diagnostic models to supplement the existing technologies.

This study aimed to develop a diagnostic model based on gene expression data. We used as many samples as possible from the GEO database and ensured that the samples of the selected dataset had come from the same sequencing platform, which minimized the effect of confounding factors to a certain extent. First, 2486 DEGs were selected from the GSE57691 dataset, and then GO enrichment analysis and KEGG pathway analysis were conducted. Cell growth and death regulation were the most significantly enriched GO terms. The association of cell growth and death with aneurysm development has been investigated before. A meta-analysis showed that 263 relevant genes have been associated with AAA. Most of these previous studies focused on three categories of genes, i.e., genes encoding the structural components of the aortic wall, genes encoding the enzymes responsible for degrading the aortic wall structure (e.g., matrix metalloproteinases and their inhibitors) [24, 25], and genes encoding the proteins involved in the immune response [26]. For example, the growth and proliferation of smooth muscle cells affects aneurysmal progression [27], and the formation of aneurysms involves chronic inflammatory cell infiltration into the tunica adventitia and tunica media along with elastin rupture, degeneration, and attenuation [28]. H19 promotes apoptosis and suppresses the proliferation of smooth muscle cells, resulting in aortic enlargement [29]. Among the KEGG pathways, we found that metabolic pathways had the most genes. A number of studies have also reported metabolic pathway changes in the aneurysmal wall compared with the normal arterial wall; for example, BAF60a deficiency in vascular smooth muscle cells can prevent the occurrence and progression of AAA by reducing inflammation and extracellular matrix degradation [30]. The findings from GO enrichment analysis and KEGG pathway analysis in the present study can contribute to the discovery of novel diagnostic indicators and therapeutic targets for AAA. Further, 74 key genes were identified using the RF algorithm, providing more susceptibility genes as targets of AAA research. Nine genes (*ZBED5*, *VEZF1*, *CLASP1*, *ARPP19*, *CTBP1*, C12orf16, *PUM1*, *CXXC5*, and *CSNK2A2*) with correlation values > 0.1 were obtained through WGCNA. Based on these genes with stronger correlations, a regulatory network of TFs was constructed, through which the pathogenesis underlying AAA can be further determined. Finally, using the ANN to calculate the weight of each key gene, a diagnostic model for AAA was established. The accuracy of the model was verified in an independent dataset, which had a prediction AUC of 0.786. The high AUC indicates that the constructed model reliably distinguishes AAA samples from normal samples.

Our study also has some limitations. First, it is difficult to obtain abdominal aorta specimens, which may limit the clinical application scenarios of this diagnostic model. Second, the etiology of AAA involves both genetic and environmental factors; because many environmental factors are associated with AAA, they may interfere with the diagnostic performance of our model that was constructed based on susceptibility genes. Third, the diagnosis of AAA using an ANN based on gene expression data depends highly on the source of the samples: Diagnosis based vascular samples from other locations would have a lower accuracy than diagnosis based on samples from the abdominal aortic segment in patients with AAA. This diagnostic model will have certain significance in scientific research and can guide the clinical screening and diagnosis of AAA, but its clinical application still needs to be confirmed by a large amount of laboratory data.

4. Conclusion

In this study, the genetic biomarkers associated with AAA were identified and used to construct an AAA diagnostic model. This study provides a valuable reference for the early clinical screening of AAA, sheds new light on the pathogenesis of AAA, and offers new targets for the clinical treatment of AAA.

5. Materials And Methods

5.1 Research design

Figure 7 is the flowchart of this study. Two large-sample gene expression datasets (GSE57691 and GSE47472) were obtained through a search of the Gene Expression Omnibus database (GEO, https://www.ncbi.nlm.nih.gov/geo/) with “Abdominal aortic aneurysm” as the keyword. The differentially expressed genes (DEGs) between the two sample groups (AAA and non-AAA) in each dataset were identified by using the limma package of R software (Step 1). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed through the Database for Annotation, Visualization, and Integrated Discovery (DAVID) based on the DEGs in GSE57691, followed by functional classification of these genes (Step 2). The DEGs identified in GSE57691 were subject to RF analysis using the RandomForest package of R, through which the genes with a mean decrease accuracy > 0.001 and mean decrease Gini > 0.05 were determined (Step 3). Further, the genes identified by weighted gene co-correlation network analysis (WGCNA) combined with the Enrichr database were used to construct a weighted gene coexpression network, and a regulatory network was then made of the genes with correlation values > 0.1 and their related TFs (Step 4). A diagnostic model was constructed using the neuralnet package based on the DEGs identified in Step 3 and was validated by in GSE47472 (Step 5). Finally, the performance of the constructed model was compared with the performance of existing diagnostic models (Step 6).

5.2 Data downloading and preprocessing

The Illumina HumanHT-12 V4.0 expression beadchip data of GSE57691 and GSE47472 AAA and normal control samples were downloaded from the GEO database. In the end, 59 samples (10 control samples and 49 AAA samples) in GSE57691 were selected for study. The constructed diagnostic model for AAA was validated using 8 control samples and 14 AAA samples obtained from GSE47472 (Table 1). The datasets downloaded from the database were normalized by Genespring GX version 11.5.1 software for luminal single-color arrays. After mapping the probes to genes, the unidentifiable probes were removed; if multiple probes could be mapped to the same gene, the expression level of the gene was represented by the maximum mean expression value for subsequent analysis.

Table 1
Source of datasets

Dataset	Platform	AAA samples	Control samples
GSE57691	GPL10558	49	10
GSE47472	GPL10558	14	8

5.3 Screening for DEGs

The limma package of R was used to analyze DEGs in the GSE57691 and GSE47472 datasets, with a FDR < 0.001 as a threshold. The DEGs were visualized with volcano plots and heatmaps.

5.4 Functional enrichment analysis of DEGs

The DEGs determined in GSE57691 were subject to GO enrichment, which categorizes genes into BP, CC, MF, and KEGG analysis, using DAVID 6.8 (https://david.ncifcrf.gov/home.jisp). The results were visualized by R software.

5.5 RF analysis to further screen DEGs

The DEGs in GSE57691 were further screened with the R package RandomForest. First, we conducted cyclical computing and obtained the out-of-bag (OOB) error rates when using different numbers of DEGs as a variable number, through which the optimal number of variables was determined based on the lowest OOB error. The OOB errors when the number of trees ranged from 1 to 3000 were calculated, and the optimal number of decision trees was determined by considering both OOB error and stability. Finally, based on the parameters determined, an RF model was constructed, and the candidate genes for AAA diagnosis were determined according to the mean decrease accuracy and mean decrease Gini.

5.6 Construction of a TF regulatory network

Pairwise correlations between genes identified from the RF screening were calculated and WGCNA was conducted by using the WGCNA package in R. The relevant genes with correlation values > 0.1 were subject to TF–mRNA regulatory relationship analysis using the Enrichr database (http://amp.pharm.mssm.edu/Enrichr/), through which the TFs that regulated the DEGs were identified for Cytoscape-aided construction of a TF regulatory network.

5.7 Construction and validation of an ANN model

GSE57691 was used for training, and GSE47472 was used for validation. According to the DEGs selected by RF, an ANN model was constructed by using the R package neuralnet based on the training dataset. The model was validated in the validation dataset, and its diagnostic performance was assessed by calculating the area under the AUC.

Abbreviations

abdominal aortic aneurysm (AAA)

random forest (RF)

artificial neural network (ANN)

differentially expressed genes (DEGs)

Gene Ontology (GO)

Kyoto Encyclopedia of Genes and Genomes (KEGG)

Database for Annotation, Visualization, and Integrated Discovery (DAVID)

area under the receiver operating characteristic curve (AUC)

false discovery rate (FDR)

Gene Expression Omnibus database (GEO)

weighted gene co-correlation network analysis (WGCNA)

transcription factors (TFs)

biological process (BP)

cellular component (CC)

molecular function (MF)

out-of-bag (OOB)

Declarations

Ethics approval and consent to participate

The study and all research materials do not involve ethics.

Consent for publication

Not applicable.

Availability of data and materials

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

Competing interests

The authors declare that they have no competing interests. The authors declare that there are no conflicts of interest.

Funding

This research was funded by Shannxi Social Development Funding (grant no.2017SF-134) and Shannxi Science Funding (grant no.2020JQ-553).

Authors' contributions

DYX and XER designed and supervised the implementation of the research, conducted preliminary analysis, drafted preliminary papers, and participated in investigations. XER participated in the revision of the intellectual content of the paper. LC and SJJ conducted the analysis in the early stages of the research and participated in the review and editing. DJ participated in research design and implementation, manuscript revision, submission and fund acquisition. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Authors' information

Affiliations:

Department of Cardiology, Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China, 710004

Duan Yixuan, Xie Enrui, Liu Chang, Sun Jingjing and Deng Jie.

Corresponding author

Correspondence to Deng Jie.

Supplementary Materials

Supplementary Table1: 2486 differentially expressed genes in GSE57691 dataset.

Supplementary Table2: 1464 differentially expressed genes in GSE47472 dataset.

Supplementary Table 3: Significantly enriched GO terms in BP, CC, and MF from GSE57691 dataset.

Supplementary Table 4: KEGG Analysis with DEGs from GSE57691 dataset.

Supplementary Table 5: Random Forest Selected Genes.

Supplementary Table 6: TF-mRNA Network.

Supplementary Table 7: WGCNA Correlation Table.

Supplementary Table 8: ANN in GSE57691 dataset.

Supplementary Table 9: ANN in GSE47472 dataset.

References

1. Flm, A., et al., *Management of Abdominal Aortic Aneurysms Clinical Practice Guidelines of the European Society for Vascular Surgery - ScienceDirect.* European Journal of Vascular and Endovascular Surgery, 2011. **41**. DOI: 10.1016/j.ejvs.2010.09.011.
2. Golledge, J. and P.E. Norman, *Current status of medical management for abdominal aortic aneurysm.* Atherosclerosis, 2011. **217**(1): p. 57-63. DOI: 10.1016/j.atherosclerosis.2011.03.006.
3. Sampson, U., et al., *Global and Regional Burden of Aortic Dissection and Aneurysms:Mortality Trends in 21 World Regions, 1990 to 2010.* Global Heart, 2014. **9**(1): p. 171-180.e10. DOI: 10.1016/j.gheart.2013.12.010.
4. Solheim, K., . *Abdominal aortic aneurysm.* Proceedings, 1993. **365**(9470): p. 1577-1589. DOI: 10.1016/0140-6736(93)90395-W.
5. Fleming, C., et al., *Screening for Abdominal Aortic Aneurysm: A Best-Evidence Systematic Review for the U.S. Preventive Services Task Force.* Annals of Internal Medicine, 2005. **142**(3): p. p.203-211. DOI: 10.7326/0003-4819-142-3-200502010-00012.
6. Calonge, N., *Screening for abdominal aortic aneurysm: recommendation statement.* Annals of Internal Medicine, 2005. **14**(11): p. 15-16.DOI: 10.7326/0003-4819-142-3-200502010-00011.
7. Stather, P.W., et al., *Meta-analysis and meta-regression analysis of biomarkers for abdominal aortic aneurysm.* British Journal of Surgery, 2015. **101**(11): p. 1358-1372. DOI: 10.1002/bjs.9593.
8. Toghill, B.J., A. Saratzis, and M.J. Bown, *Abdominal aortic aneurysm—an independent disease to atherosclerosis?* Cardiovascular pathology: the official journal of the Society for Cardiovascular Pathology, 2017. **27**: p. 71. DOI: 10.1016/j.carpath.2017.01.008.
9. Tejas, C., et al., *On the assessment of abdominal aortic aneurysm rupture risk in the Asian population based on geometric attributes.* Proceedings of the Institution of Mechanical Engineers Part H Journal of Engineering in Medicine, 2018. **232**: p. 095441191879472. DOI: 10.1177/0954411918794724.
10. Larsson, E., et al., *A population-based case-control study of the familial risk of abdominal aortic aneurysm.* Journal of Vascular Surgery, 2009. **49**(1): p. 47-51. DOI: 10.1016/j.jvs.2008.08.012.
11. Biros, E., et al., *Differential gene expression in human abdominal aortic aneurysm and aortic occlusive disease.* Oncotarget, 2015. **6**(15). DOI: 10.18632/oncotarget.3848.
12. Lenk, G.M., et al., *Whole genome expression profiling reveals a significant role for immune function in human abdominal aortic aneurysms.* Bmc Genomics, 2007. **8**(1): p. 237. DOI: 10.1186/1471-2164-8-237.

13. Liu, Y., et al., *Identification of key genes and pathways in abdominal aortic aneurysm by integrated bioinformatics analysis.* The Journal of international medical research, 2019: p. 030006051989443. DOI: 10.1177/0300060519894437.
14. Liaw, A. and M. Wiener, *Classification and regression by randomForest. R News 2:18-22.* 2001.
15. Judith, E. and J.M. Deleo, *Artificial neural networks.* Cancer, 2001. **91**(S8): p. 1615-1635. DOI: doi:http://dx.doi.org/.
16. Xenos, M., et al., *Patient-Based Abdominal Aortic Aneurysm Rupture Risk Prediction with Fluid Structure Interaction Modeling.* Annals of Biomedical Engineering, 2010. **38**(11): p. 3323-37. DOI: 10.1007/s10439-010-0094-3.
17. Doyle, B.J. and T.M. Mcgloughlin, *Computer-Aided Diagnosis of Abdominal Aortic Aneurysms*. 2011: Biomechanics and Mechanobiology of Aneurysms.
18. Doyle, B.J., et al., *Identification of rupture locations in patient-specific abdominal aortic aneurysms using experimental and computational techniques.* Journal of Biomechanics, 2010. **43**(7): p. 1408-1416. DOI: 10.1016/j.jbiomech.2009.09.057.
19. David A. Sidloff BSc Hons, M., MRCS a, et al., *A systematic review and meta-analysis of the association between markers of hemostasis and abdominal aortic aneurysm presence and size.* Journal of Vascular Surgery, 2014. **59**(2): p. 528-535. DOI: 10.1111/j.1600-065X.1997.tb00994.x.
20. Cosford, P.A., G.C. Leng, and J. Thomas, *Screening for abdominal aortic aneurysm.* Bmj British Medical Journal, 2009. **338**(7710): p. 1509-1510. DOI: 10.1002/bjs.4140.
21. Liang, et al., *A machine learning approach to investigate the relationship between shape features and numerically predicted risk of ascending aortic aneurysm.* Biomechanics and modeling in mechanobiology, 2017. DOI: 10.1007/s10237-017-0903-9.
22. Lee, R., et al., *Applied Machine Learning for the Prediction of Growth of Abdominal Aortic Aneurysm in Humans.* Ejves Short Reports, 2018: p. S2405655318300094. DOI: 10.1007/s10237-017-0903-9.
23. Kent, K.C., *Clinical practice. Abdominal aortic aneurysms.* New England Journal of Medicine, 2014. **371**(22): p. 2101-8.DOI: 10.1056/NEJMcp1401430.
24. *Functional matrix metalloproteinase-9 polymorphism (C-1562T) associated with abdominal aortic aneurysm.* Journal of Vascular Surgery, 2003. **38**(6): p. 1363-1367. DOI: 10.1016/S0741-5214(03)01027-9.
25. Toru, et al., *Genetic analysis of polymorphisms in biologically relevant candidate genes in patients with abdominal aortic aneurysms.* Journal of Vascular Surgery, 2005. **41**(6): p. 1036-1042. DOI: 10.1016/j.jvs.2005.02.020.
26. Rasmussen, T.E., et al., *Genetic similarity in inflammatory and degenerative abdominal aortic aneurysms: A study of human leukocyte antigen class II disease risk genes - ScienceDirect.* Journal of Vascular Surgery, 2001. **34**(1): p. 84-89. DOI: 10.1067/mva.2001.115603.
27. Hoshina, K., et al., *Aortic wall cell proliferation via basic fibroblast growth factor gene transfer limits progression of experimental abdominal aortic aneurysm.* Journal of Vascular Surgery, 2004. **40**(3): p. 512-518. DOI: 10.1016/j.jvs.2004.06.018.

28. Shimizu, K., R.N. Mitchell, and P. Libby, *Inflammation and Cellular Immune Responses in Abdominal Aortic Aneurysms.* Arteriosclerosis Thrombosis & Vascular Biology, 2006. **26**(5): p. 987-994. DOI: 10.1016/j.jvs.2004.06.018.
29. Li, D.Y., et al., *H19 Induces Abdominal Aortic Aneurysm Development and Progression.* Circulation, 2018: p. CIRCULATIONAHA.117.032184. DOI: 10.1161/CIRCULATIONAHA.117.032184.
30. Chang, Z., et al., *BAF60a Deficiency in Vascular Smooth Muscle Cells Prevents Abdominal Aortic Aneurysm by Reducing Inflammation and ECM (Extracellular Matrix) Degradation.* Arteriosclerosis Thrombosis and Vascular Biology, 2020. **40**(10). DOI: 10.1161/ATVBAHA.120.314955.

Figures

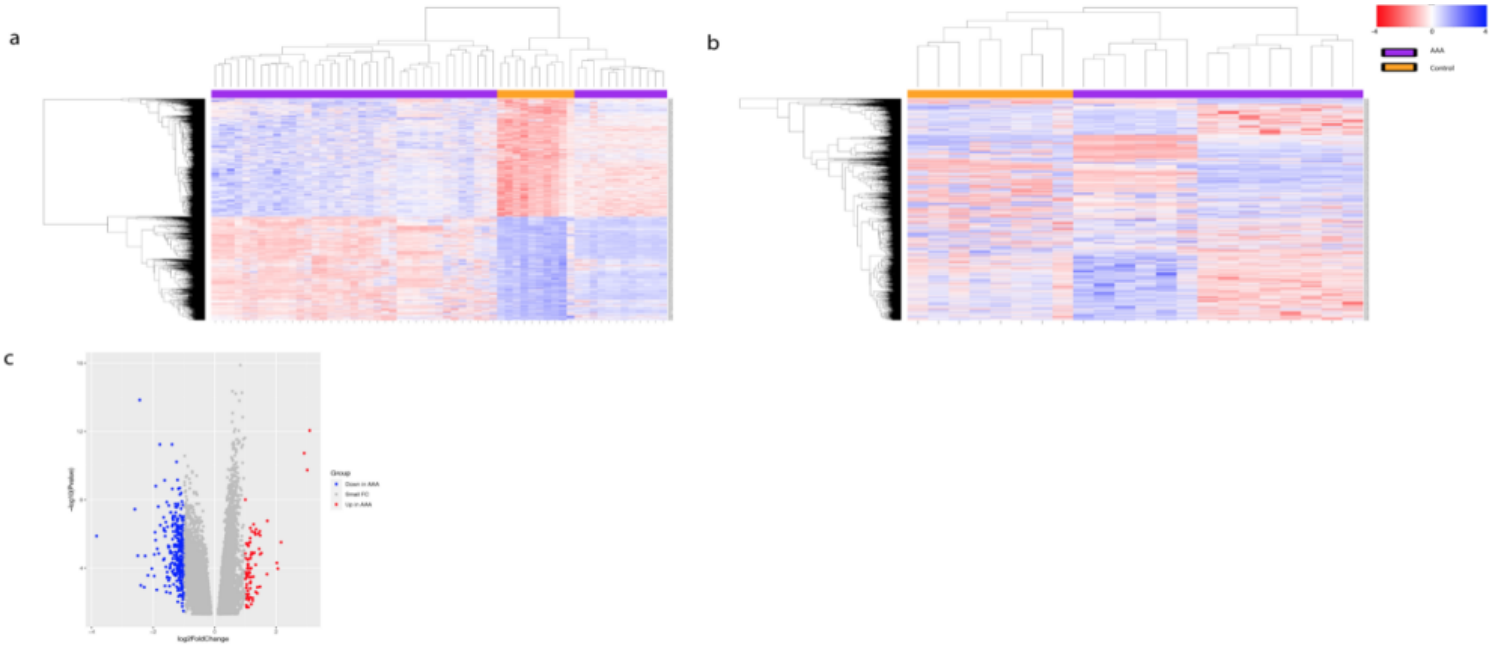

Figure 1

Screening of DEGs in the datasets. (a) The heatmap of 2486 DEGs in GSE57691, which was derived from clustering analysis of gene expression data in 49 AAA and 10 control samples. (b) The heatmap of 1464 DEGs in GSE47472, which was derived from clustering analysis of gene expression data in 14 AAA and 8 control samples. (c) Volcano plots demonstrating the distribution of DEGs in GSE57691. The x-axis shows -log10 (p value), the y-axis refers to |log2(fold change)|, and the cutoff value is |log2FC|≥1.

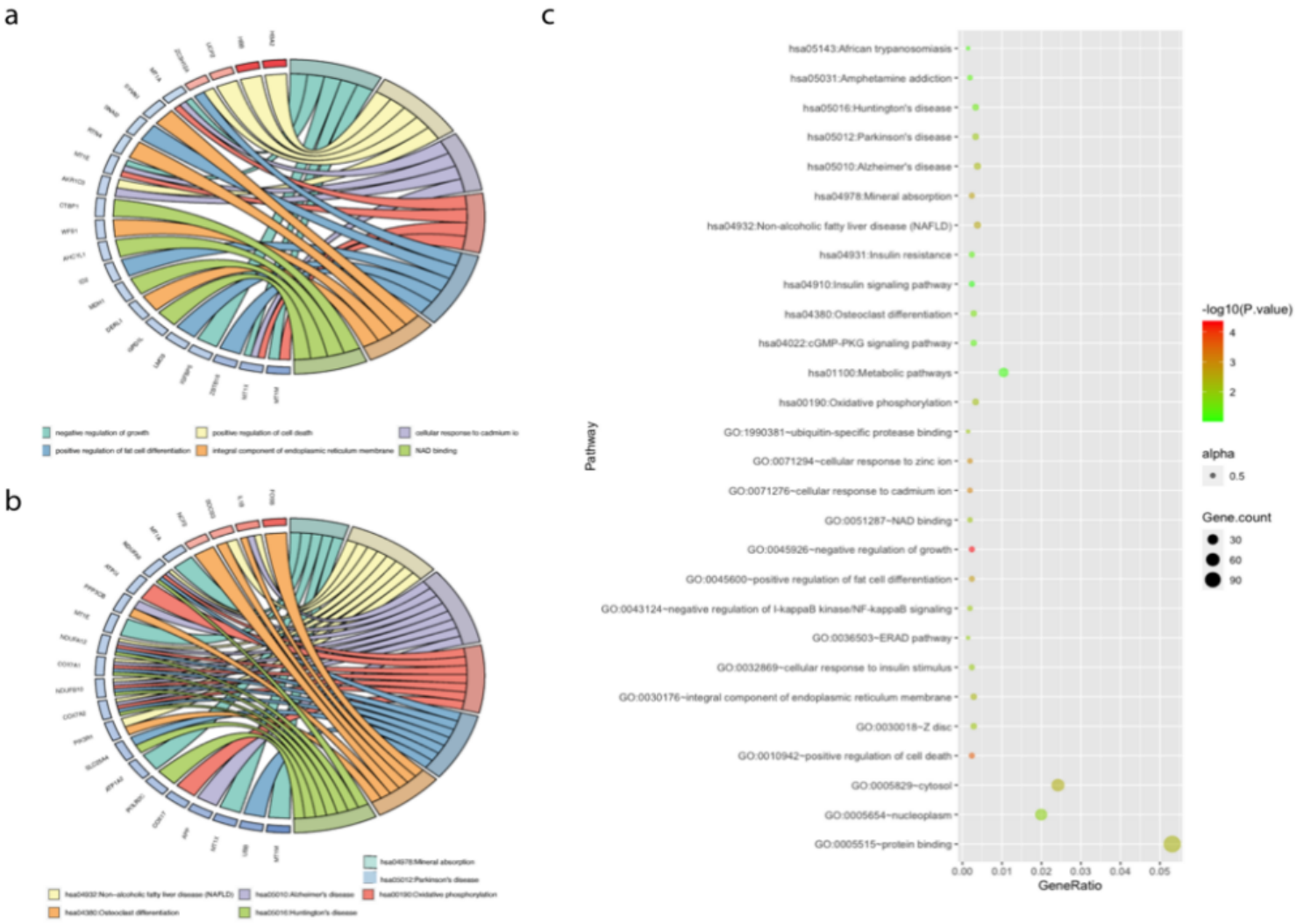

Figure 2

Functional analysis and visualization of all 2486 DEGs in DAVID. (a) Circle diagram of enriched GO functional clusters. (b) Circle diagram of enriched KEGG pathways. (c) Functional enrichment bubble diagram of the 2486 DEGs.

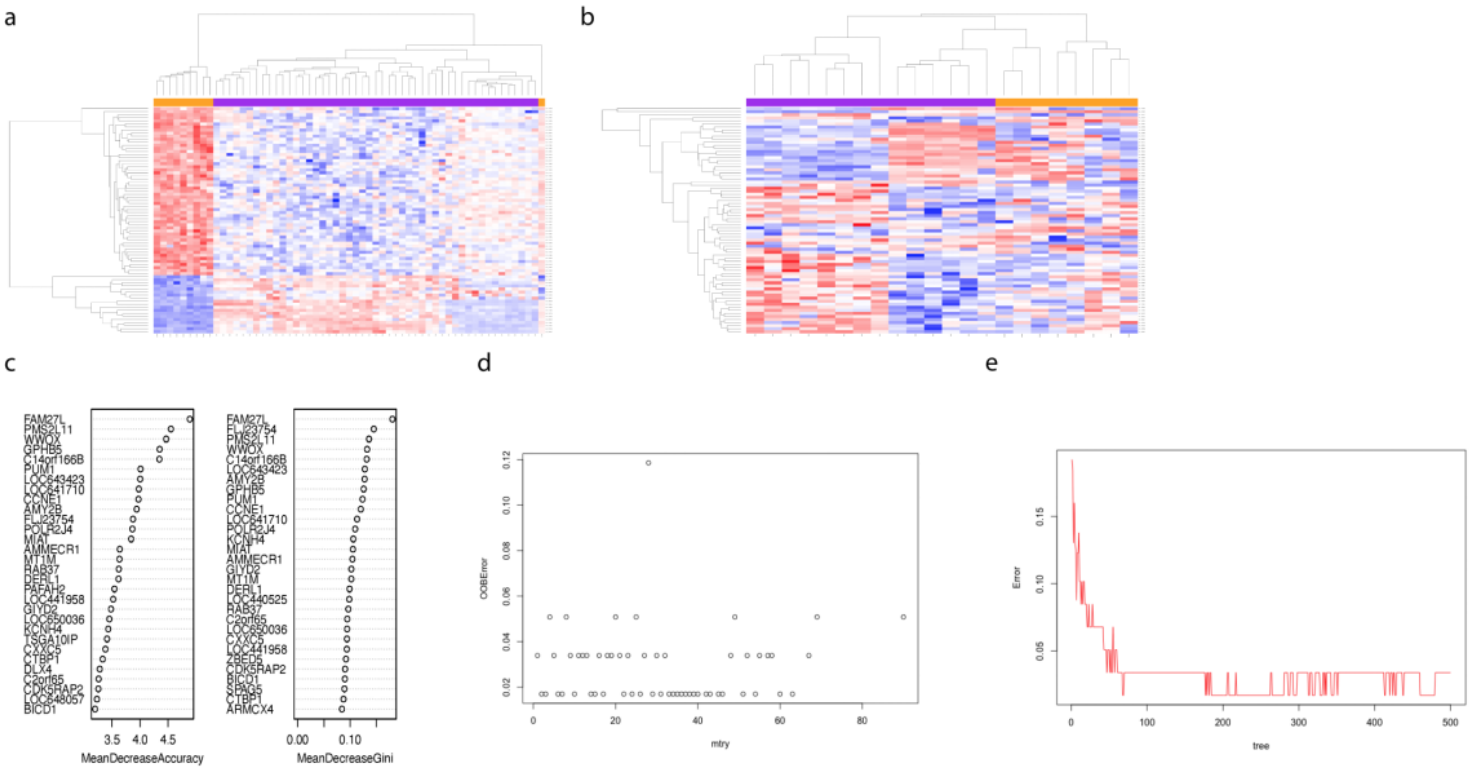

Figure 3

Heatmaps of 74 feature genes selected by RF in GSE57691 and GSE47472. (a) Clustering of the 74 genes in GSE57691. (b) Clustering of the 74 genes in GSE47472. (c) Ranks of input variables in the RF model, based on which the genes were classified in both the AAA and control groups. (d) Determination of the optimal number of feature genes. (e) The impact of the number of decision trees on the OOB error.

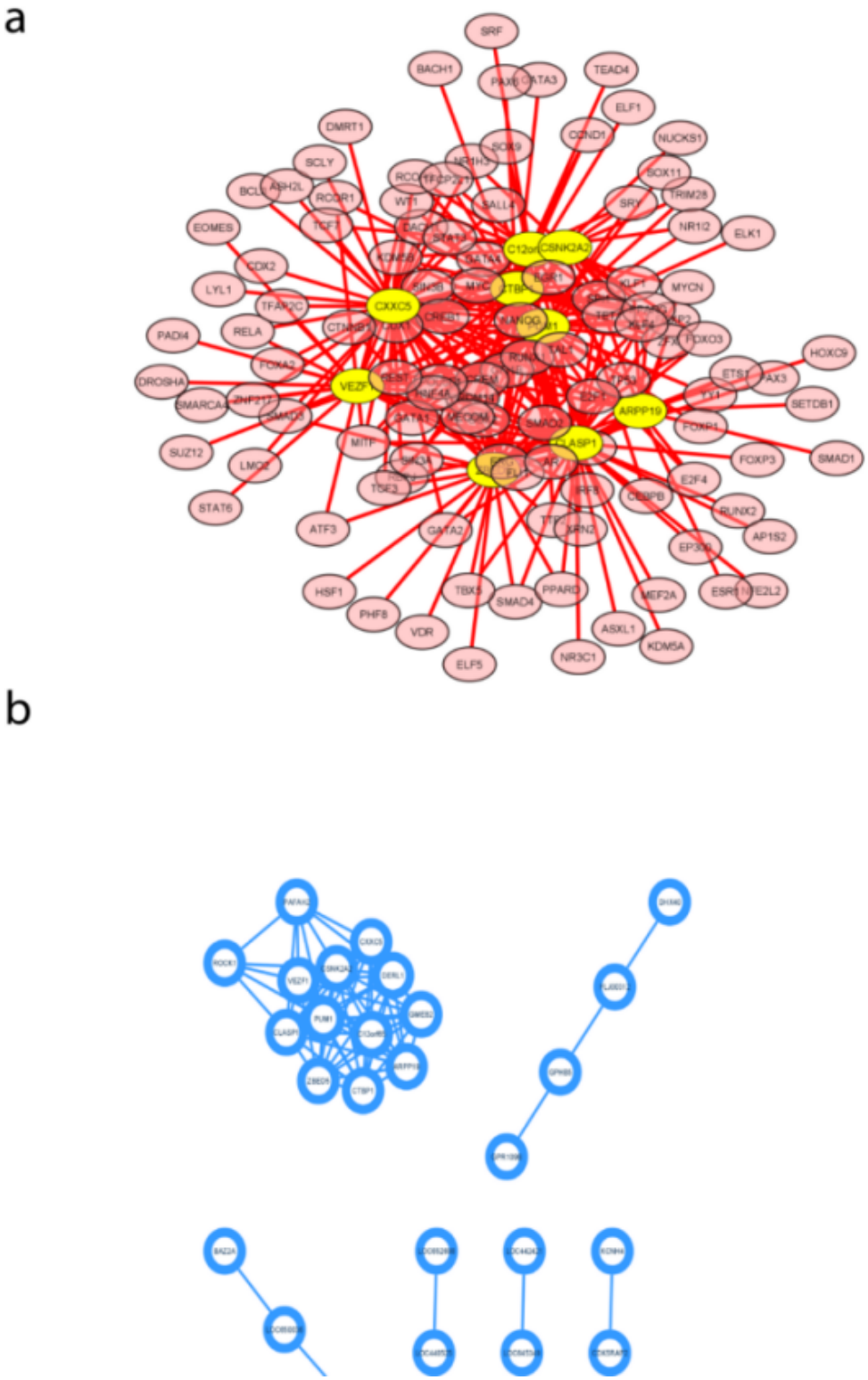

Figure 4

The TF–mRNA network and WGCNA network. (a) The 9 genes with a yellow background all have correlation values > 0.1, and the related gene-regulatory factors are marked with a red background. (b) The WGCNA network.

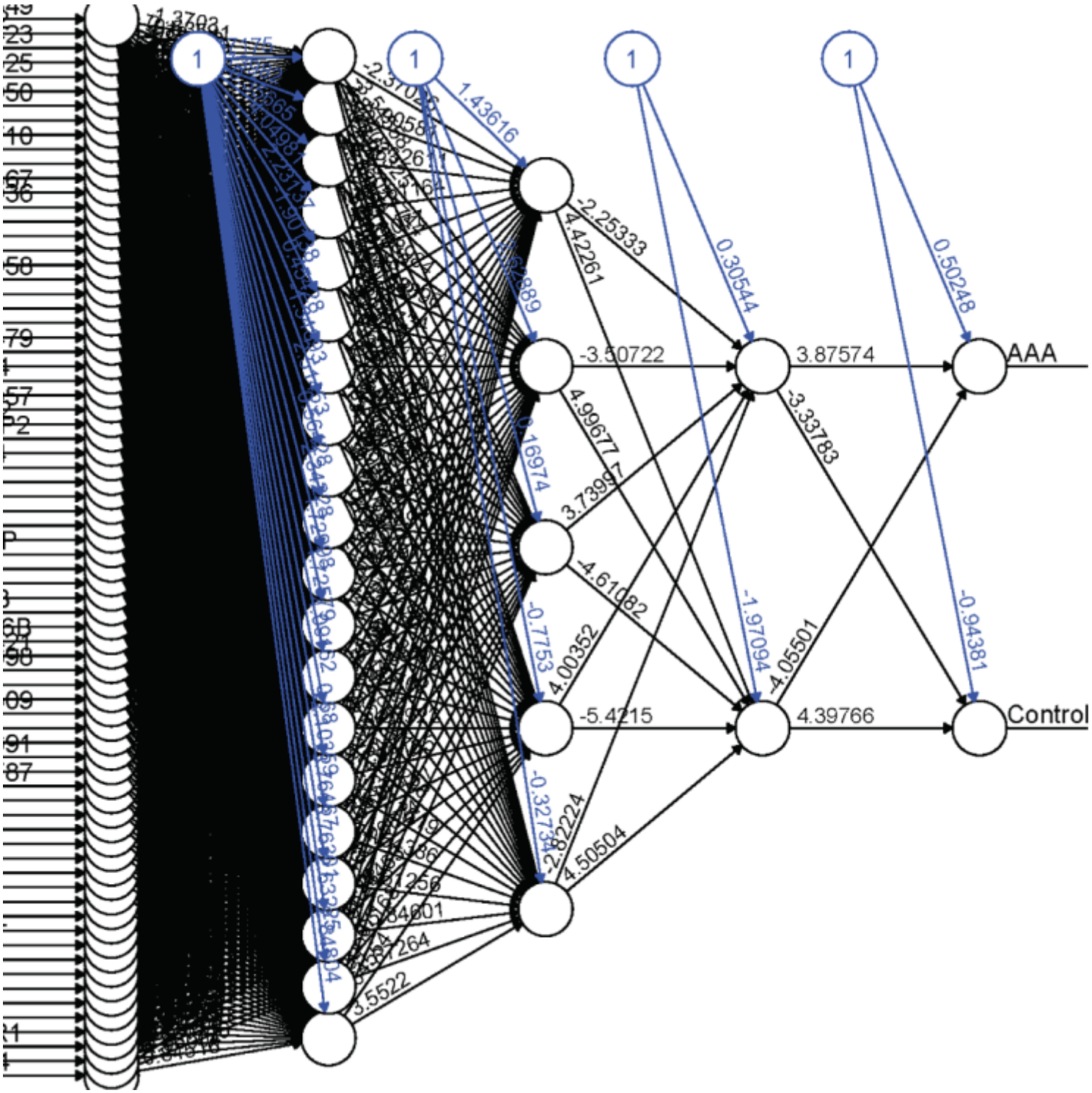

Figure 5

Construction of a neural network: the neural network topology of the dataset (GSE57691) with 1 input layer, 3 hidden layers, and 1 output layer.

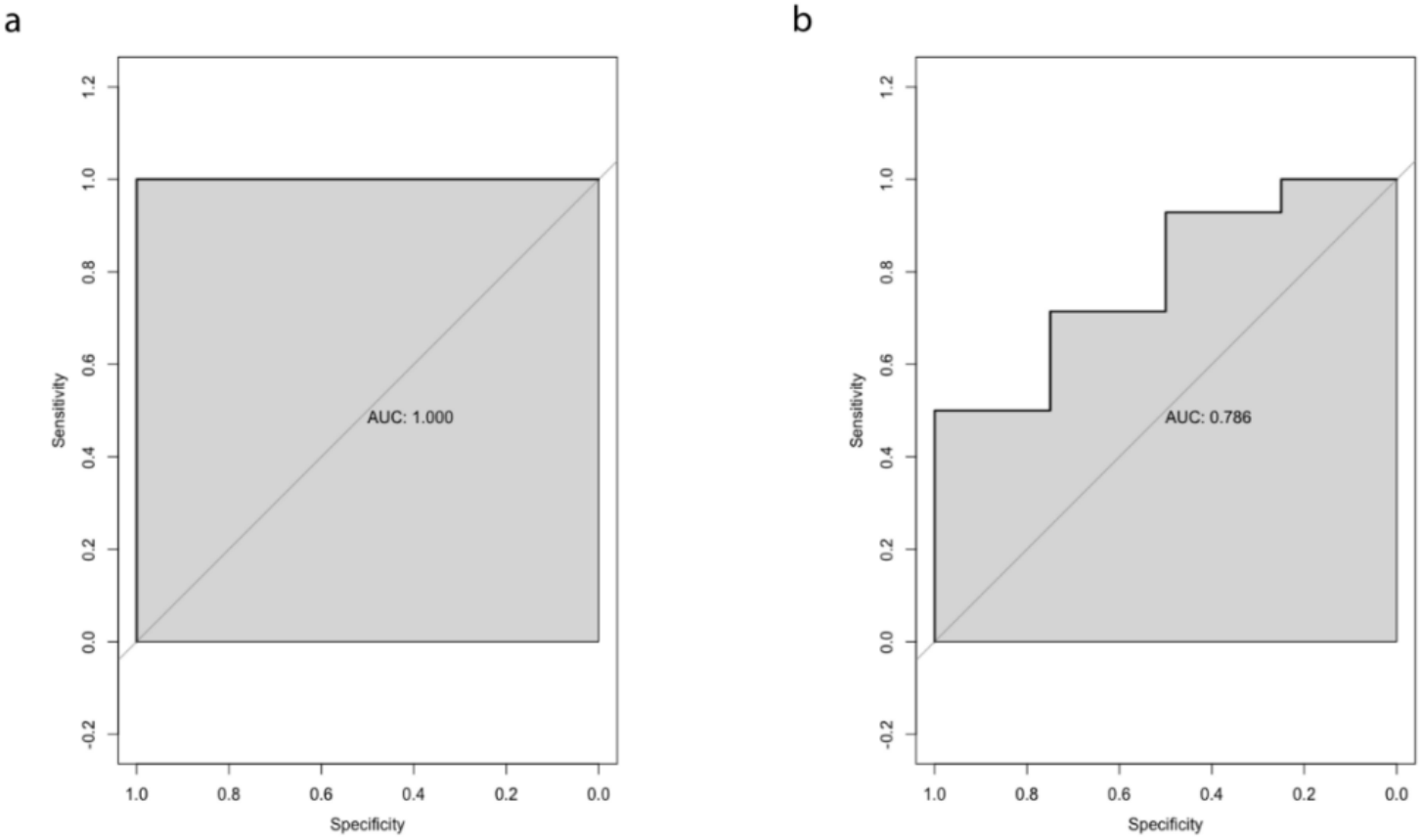

Figure 6

(a) Prediction by the constructed ANN model in the GSE57691 dataset. (b) Validation of the ANN model in the GSE47472 dataset.

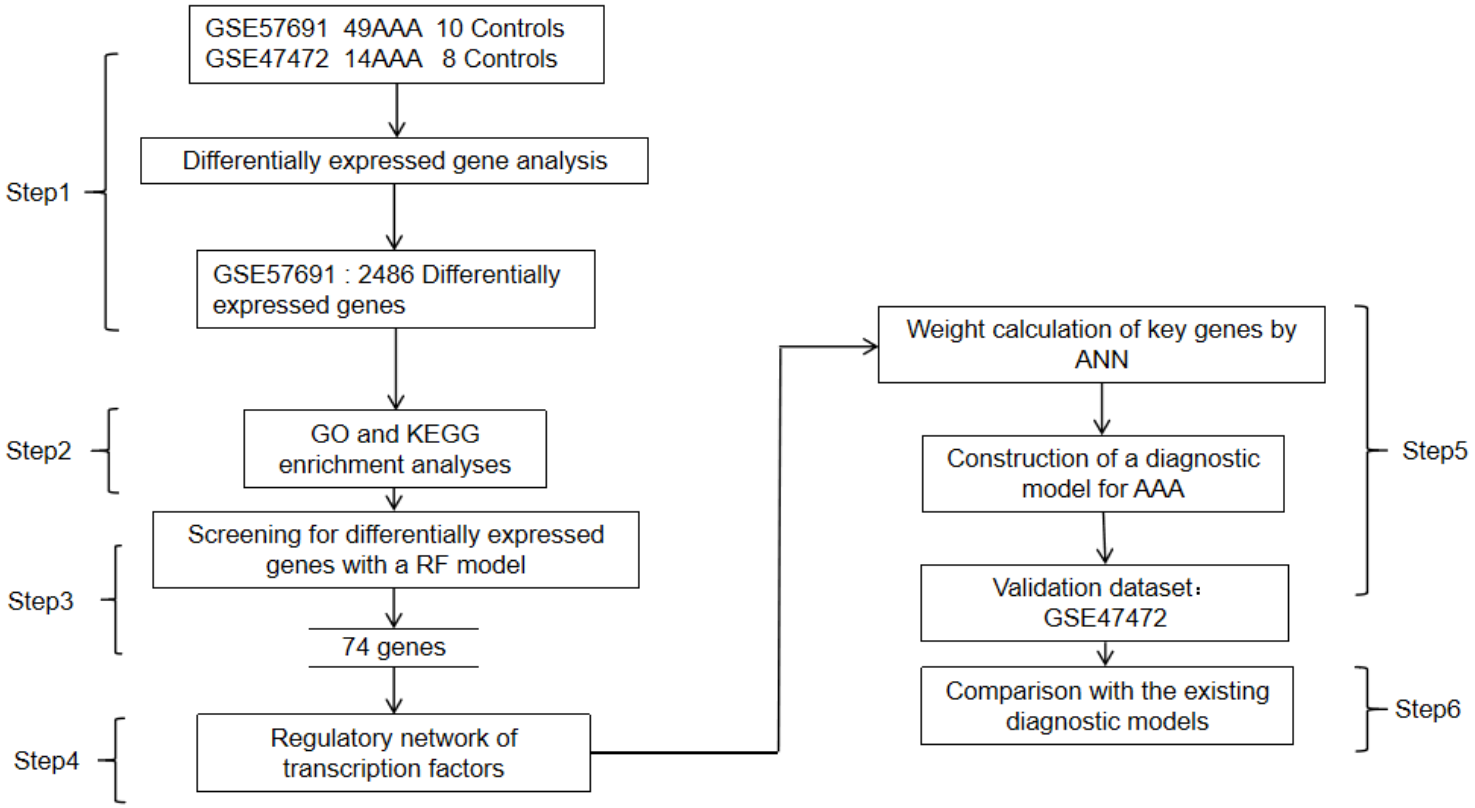

Figure 7

Schematic illustration of rearch design

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementTable.xlsx