Biomarkers of Blood from Patients with Atherosclerosis Based on Bioinformatics Analysis

Evolutionary Bioinformatics
Volume 17: 1–6

DOI: 10.1177/11769343211046020

Yongjiang Qian[1], Lili Zhang[1], Zhen Sun[1], Guangyao Zang[1], Yalan Li[1], Zhongqun Wang[1] and Lihua Li[2]

[1]Department of Cardiology, Affiliated Hospital of Jiangsu University, Zhenjiang, China.
[2]Department of Pathology, Affiliated Hospital of Jiangsu University, Zhenjiang, China.

ABSTRACT: Atherosclerosis is a multifaceted disease characterized by the formation and accumulation of plaques that attach to arteries and cause cardiovascular disease and vascular embolism. A range of diagnostic techniques, including selective coronary angiography, stress tests, computerized tomography, and nuclear scans, assess cardiovascular disease risk and treatment targets. However, there is currently no simple blood biochemical index or biological target for the diagnosis of atherosclerosis. Therefore, it is of interest to find a biochemical blood marker for atherosclerosis. Three datasets from the Gene Expression Omnibus (GEO) database were analyzed to obtain differentially expressed genes (DEG) and the results were integrated using the Robustrankaggreg algorithm. The genes considered more critical by the Robustrankaggreg algorithm were put into their own data set and the data set system with cell classification information for verification. Twenty-one possible genes were screened out. Interestingly, we found a good correlation between *RPS4Y1*, *EIF1AY*, and *XIST*. In addition, we know the general expression of these genes in different cell types and whole blood cells. In this study, we identified *BTNL8* and *BLNK* as having good clinical significance. These results will contribute to the analysis of the underlying genes involved in the progression of atherosclerosis and provide insights for the discovery of new diagnostic and evaluation methods.

KEYWORDS: Blood, bioinformatics, biomarker, atherosclerosis

RECEIVED: June 6, 2021. **ACCEPTED:** August 24, 2021.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported as follows: the National Natural Science Foundation of China (82070455, 81770450); the related Foundation of Jiangsu Province (BK20201225); the Open Project Program of Guangxi Key Laboratory of Centre of Diabetic Systems Medicine (GKLCDSM-20210101-02); Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX20_2881).

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Lihua Li, Department of Pathology, Affiliated Hospital of Jiangsu University, Jiefang Road 438, Zhenjiang, Jiangsu 212000, China. Email: tsmc01@126.com

Background

Atherosclerosis (AS) is the leading cause of peripheral vascular disease, coronary heart disease, and cerebral infarction.[1] The development of atherosclerotic lesions may be caused by low-density lipoprotein, a lipoprotein that carries cholesterol into peripheral tissue cells and can be oxidized to become oxidized low-density lipoprotein. Other risk factors contribute to atherosclerosis and its thrombotic complications, including diabetes, smoking, and high blood pressure.[2] Growing evidence also indicates a role for emerging risk factors, including clonal hematopoiesis and inflammation. A range of auxiliary examination methods, both invasive (such as selective coronary angiography) and non-invasive (such as nuclear scans, CT, stress tests, and blood biomarkers), allow assessment of cardiovascular disease risk and treatment targets. However, there is no simple blood biochemical index or biological target for the diagnosis of atherosclerosis at present; instead, more ultrasonographic screening or angiography are used.[2] Therefore, it would be valuable to identify a biochemical blood marker for atherosclerosis.

With the development of omics and the availability of clinical blood samples, many studies have focused on the blood transcriptome of patients with atherosclerosis. Transcriptome analysis of blood cells, divided into those of atherosclerotic patients and matched controls, will potentially supply biomarkers for diagnostic purposes and provide insights into the mechanism of atherosclerosis.[3–6] One study focused on differences in various cells in the blood of patients with AS to explore the biological functions of macrophages and CD34 cells[7]; other studies have examined the transcriptome of peripheral blood and the transcriptional expression of circulating cells in patients with acute myocardial infarction or artery plaque.[8] Meanwhile, with the development of high-throughput sequencing and bioinformatics analysis techniques, a bioinformatics gene analysis related to the increased risk of atherosclerosis due to familial hypercholesterolemia provides a basis for the development of therapies for atherosclerosis.[9] In addition, bioinformatics analysis of oncology,[10,11] endocrine diseases,[12] and respiratory diseases[13] drives basic research and provides directions for treating patients.

In this study, after the detection of differential expression genes in multiple data sets, the Robust rank aggregation algorithm was used for integration evaluation, and 21 possible genes were screened out as potential biomarkers for biological diagnostic screening. We looked at the expression of these genes in different circulating cells. Interestingly, we found a good correlation between *RPS4Y1*, *EIF1AY*, and *XIST*.

Methods

Retrieve

Keywords "atherosclerosis" and "blood" were searched in the GEO database and the species was limited to "Homo sapiens." Fifty-nine data sets were retrieved, and then we manually

Table 1. Information of data sets.

GEO	PLATFORM	NORMAL	PATIENT	REFERENCE NUMBER
GSE27034	GPL570	18	19	Masud et al[5]
GSE90074	GPL6480	50	93	Ravi et al[6]
GSE12288	GPL96	112	110	Sinnaeve et al[7]

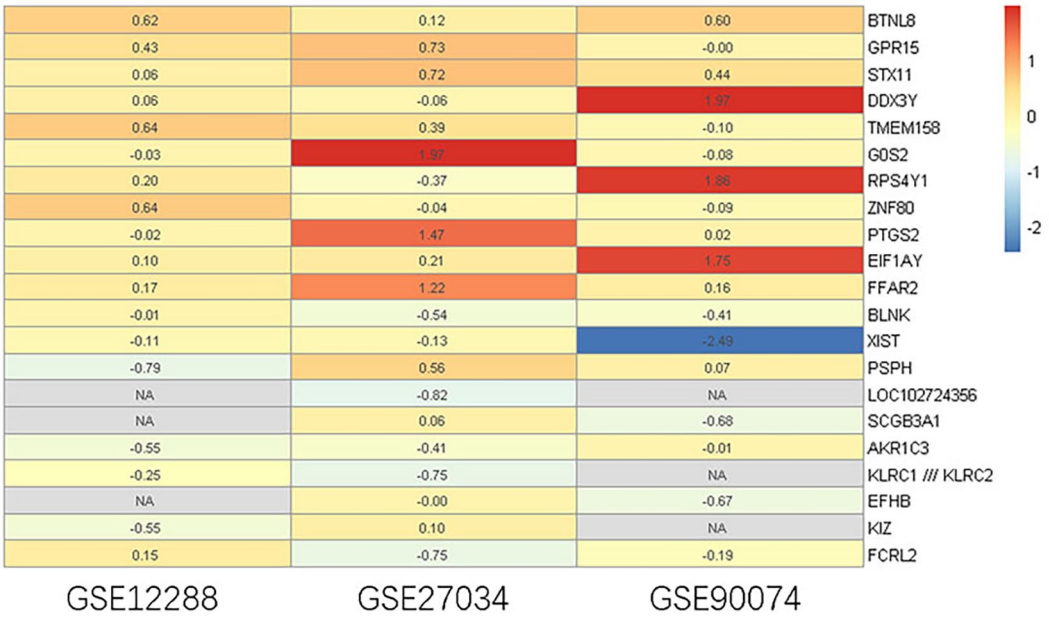

Figure 1. LogFC of Genes were identified in 3 datasets, red represents high values, and blue represents low values.

excluded the mRNA chip data sets unrelated to the blood of atherosclerosis patients and not clearly grouped, and finally 3 data sets were screened out (Table 1).

Differentially expressed genes (DEGs) analysis, network analysis, and functional enrichment analysis. We assessed the primary data using R language assessment and quality control, all the expression of matrix through log2 processing, and use ggpubr package for draw violin plots (Supplemental Figure S1). The samples were divided into case and control groups according to the information on GEO. The LIMMA package was used to analyze the differential genes[14] (Supplemental Tables S1-S3 and Figure S2). We screened genes with *P*-value less than .01 and $|\text{LogFc}| > 0.5$. We drew the network diagram based on the strings database and analyzed the path of the network diagram (Supplemental Figure S3).

Robustrankaggreg

Robustrankaggreg R package was used to integrate the up-down-regulated genes,[15] respectively. RRA is a rank aggregation method based on sequential statistics, which can achieve the purpose of removing the noise of individual experimental results while increasing the signal and reducing the proportion of false-positive results in high-throughput data integration. There were n rank vectors and normalized the rank vectors were sorted from small to large. When the l rank vector was greater than or equal to the mean of all rank vectors, the row $\sum_{k=l}^{n}\binom{n}{k}x^k(1-x)^{n-k}$ of all rank vectors was calculated, and the minimum value was taken as the score. Genes with a score less than 0.05 were screened out as the marker genes we considered, and a heat map of logFC in different datasets was drawn (Figure 1).

Genetic alignment and correlation analysis

The expression matrices of the identified genes were selected from the original data set and GSE9820,[3] and the unclustered and clustered heat maps were constructed with pheatmap function (Figures 2–6). Correlation analysis was performed for all identified markers (Supplemental Tables S4-S6), and regression analysis was performed for the most interesting genes, and scatter plots and residual plots were plotted (Figures 3–5 and Supplemental Figure S4). The genes of interest were plotted in a scatter plot. P value$<.05$.

Results

Genes detected according to the integrated DEGs

Deg analysis was performed on all data sets, and specific DEG results can be seen in the Supplemental Data. Only 1 pathway

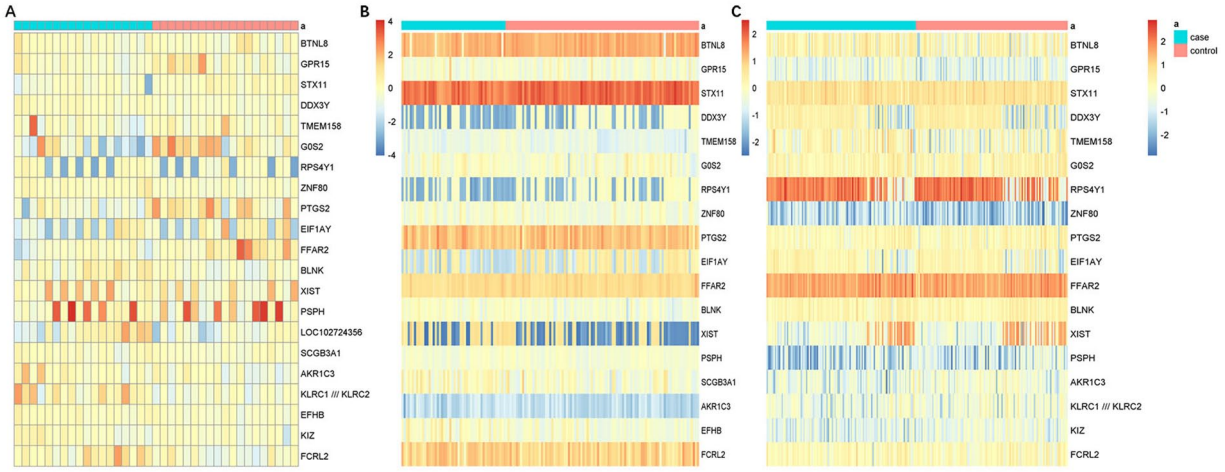

Figure 2. Gene expression in 3 datasets. Red represents high expression, blue represents low expression, each column represents a sample, and each row represents a gene. Unclustered heat map of gene expression in GSE27034 (A) unclustered heat map of gene expression in GSE90074 (B) unclustered heat map of gene expression in GSE12288 (C).

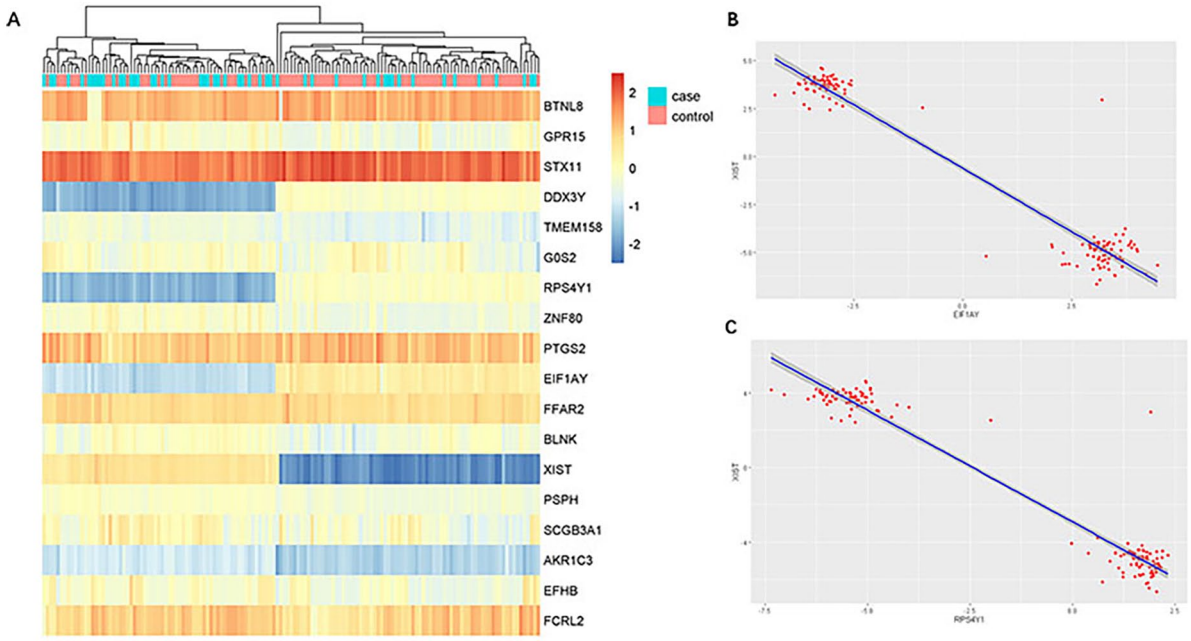

Figure 3. Clustered heat map of gene expression in GSE90074 (A) scatter diagram and regression line of XIST and EIF1AY, regression equation: $y=-1.32619x-0.61110$, residual standard error: 1.075 on 141 degree of freedom, n:143, multiple R-squared: 0.9411, adjusted R-squared: 0.9407, F-statistic: 2253 on 1 and 141 DF, P-value: $<2.2e^{-16}$ (B) scatter diagram and regression line of XIST and RPS4Y1, regression equation: $y=-1.19600x-2.91454$, residual standard error: 1.009 on 141° of freedom, n:143, multiple R-squared: 0.9482, adjusted R-squared: 0.9478, F-statistic: 2579 on 1 and 141 degree of freedom, P-value: $<2.2e^{-16}$ (C).

"Cytokine Signaling in Immune system" was enriched after network analysis and pathway analysis of the genes considered significant. We still obtained 21 genes based on RRA algorithm integration with good scores, including up-regulated genes: *BTNL8*, *GPR15*, *STX11*, *DDX3Y*(*DBY*), *TMEM158*. *G0S2*, *PS4Y1* (*RPS4Y*), *ZNF80*, *PTGS2*, *EIF1AY* (*IF1AY*), and *FFAR2*. Among them, *BTNL8*, *GPR15*, *STX11*, and *TMEM158* have relatively high logFC in multiple data sets, while *DDX3Y(DBY)*, *G0S2*, *PS4Y1(RPS4Y)*, *PTGS2*, *EIF1AY(IF1AY)*, and *FFAR2* have relatively high logFC in a single data set. The down-regulated genes included *BLNK*, *XIST*, *PSPH*, *LOC10272435*, *SCGB3A1*, *AKR1C3*, *KLRC1*, *EFHB*, *KIZ*, and *FCRL2*, among them BLNK showed significant differences in multiple data sets, while *XIST* showed a considerable difference in GSE90074. These genes may be used for screening and evaluation of AS or vascular plaques.

The correlation between RPS4Y1, XIST, and EIF1AY

Because the logFC value is low, the difference between the case and control groups is not visible to the naked eye. However,

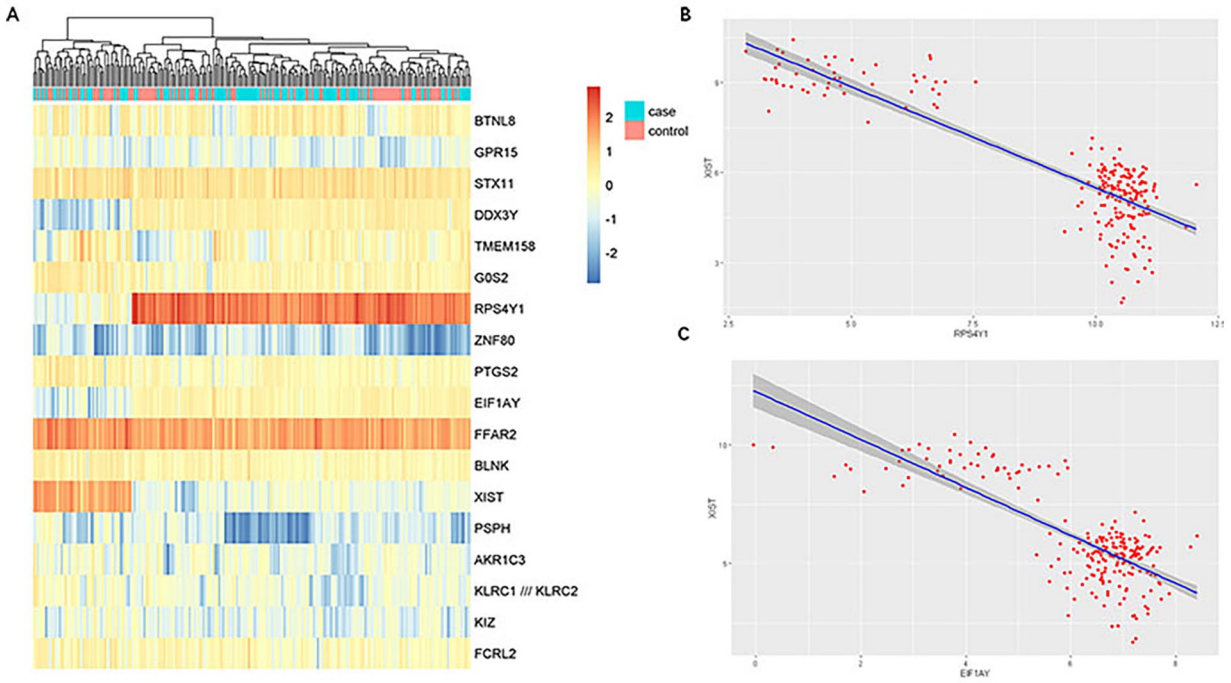

Figure 4. Clustered heat map of gene expression in GSE12288 (A) scatter diagram and regression line of XIST and RPS4Y1, regression equation: $y = -0.67326x + 12.23007$, residual standard error: 0.9911 on 220 degree of freedom, N:222, multiple *R*-squared: 0.7398, adjusted *R*-squared: 0.7386, *F*-statistic: 625.5 on 1 and 220 DF, *P*-value: $<2.2e^{-16}$ (B) scatter diagram and regression line of XIST and EIF1AY, regression equation: $y = -1.01055x + 12.24370$, residual standard error: 1.239 on 220 degree of freedom, N:222, multiple *R*-squared: 0.5933, adjusted *R*-squared: 0.5914, *F*-statistic: 320.9 on 1 and 220 DF, *P*-value: $<2.2e^{-16}$ (C).

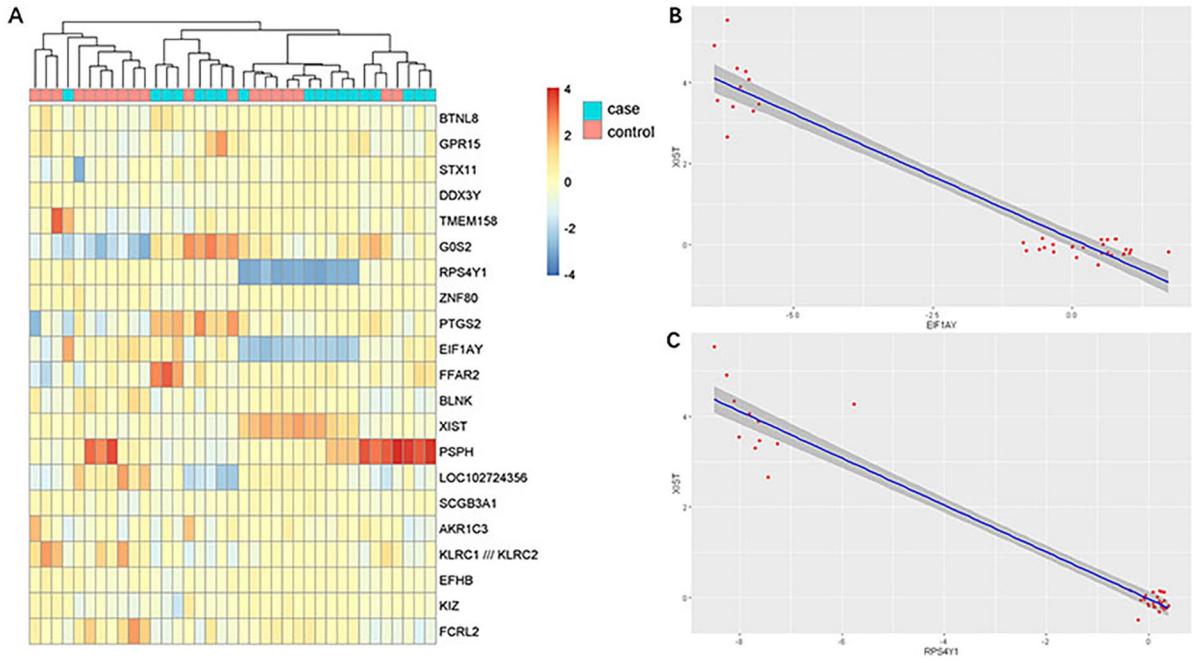

Figure 5. Clustered heat map of gene expression in GSE27037 (A) scatter diagram and regression line of XIST and EIF1AY, regression equation: $y = -0.61712x + 0.14227$, residual standard error: 0.5456 on 35 degree of freedom n:37, multiple *R*-squared: 0.9219, adjusted *R*-squared: 0.9197, *F*-statistic: 413.2 on 1 and 35 DF, *P*-value: $<2.2e^{-16}$ (B) scatter diagram and regression line of XIST and RPS4Y1, regression equation: $y = -0.51754x - 0.02932$, residual standard error: 0.4404 on 35 degree of freedom, n:37, multiple R-squared: 0.9491, adjusted *R*-squared: 0.9477, F-statistic: 652.8 on 1 and 35 DF, *P*-value: $<2.2e^{-16}$ (C).

after clustering the heat maps, we found an interesting phenomenon for the first time: *XIST* is negatively correlated with RPS4Y1 in all 3 data sets, and *XIST* is negatively correlated with *EIF1AY*. The sample expressing *XIST*, *RPS4Y1*, and *EIF1AY* are basically not expressed, and vice versa. This mechanism may also be involved in atherosclerosis.

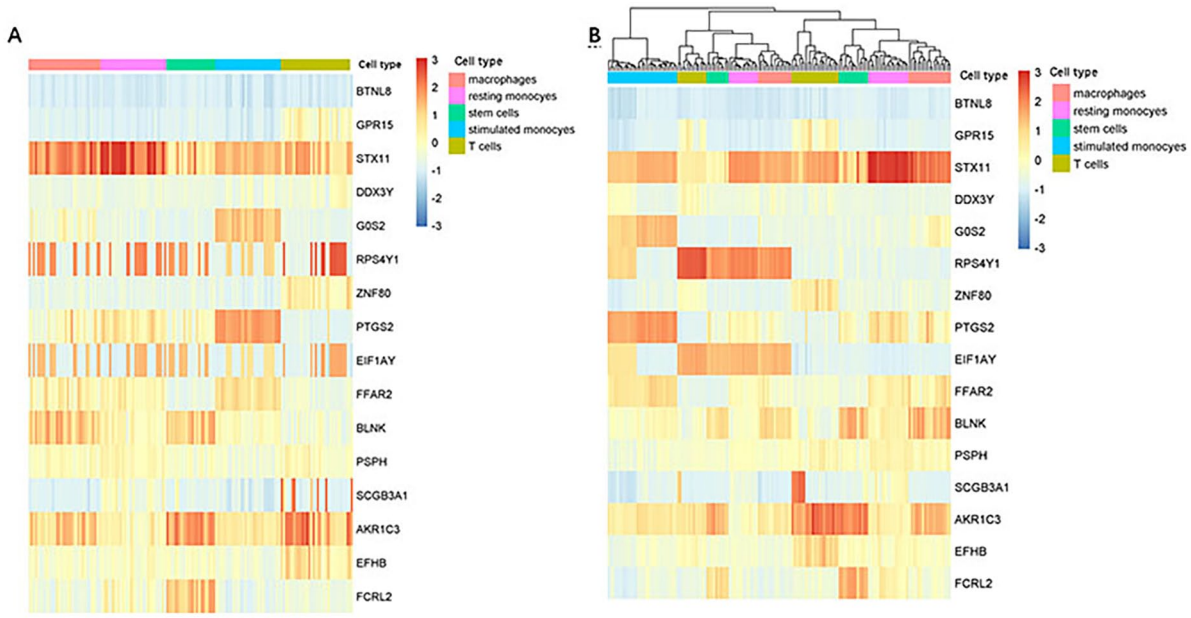

Figure 6. Gene expression in GSE9820. Unclustered heat map of gene expression in GSE9820 (A) clustered heat map of gene expression in GSE9820 (B).

Validation in different cell types

We picked up the expression of these selected genes in the data set of GSE9820,[3] which is a sequencing data of Mononuclear Cell Transcriptomes, and identified 5 kinds of cells, including $CD34^+$ stem cells, $CD4^+$ T-cells, resting $CD14^+$ monocytes, stimulated monocytes, and macrophages. It can be seen that the expression level of *BTNL8* is relatively low in these 5 kinds of cells, while it is still relatively high in other data sets, so it should be highly expressed in a cell that does not belong to these 5 kinds of cells. *RPS4Y1* and *EIF1AY* were not tissued specific, but individual specific. *GPR15* and *ZNF80* were highly expressed in T cells, *G0S2*, *PTGS2*, and *FFAR2* were highly expressed in stimulated monocytes, and stem cells mainly highly expressed *BLNK*, *AkR1C3*, and *FCRL2*. Good consistency between *RPS4Y1* and *EIF1AY* can also be seen in the cluster diagram of GSE9820.

Discussions

This study combines 3 coronary atherosclerosis in patients with blood samples mRNA array dataset to filter possible coronary atherosclerosis possible genetic detection objects in the blood. We found there are 21 genes that may have specific significance and also discussed these gene expressions between different cells in the blood. This study first reported *RPS4Y1*, *EIF1AY* own the correlation between *XIST*.

Many of these genes are associated with inflammation and immunity. *BTNL8*, which has the best score, may stimulate the primary immune response acts on T-cell stimulated sub-optimally through the TCR/CD3 complex stimulating their proliferation and cytokine production.[16] *G0S2*, G0/G1 switch protein 2, promotes apoptosis by binding to *BCL2*, resulting in preventing the formation of protective Bcl2-Bax heterodimers.[17] *GPR15L* is a chemotactic factor that mediates recruitment of lymphocytes to epithelia through binding and activation of the G-protein coupled receptor GPR15 seems to be epithelia related.[18] *BLNK*, B-cell linker protein, functions as a central linker protein downstream of the B-cell receptor (BCR), bridging the *SYK* kinase to a multitude of signaling pathways, and regulating biological outcomes of B-cell function and development.[19] What is more, *BLNK* plays a role in the activation of *ERK/EPHB2*, *MAP kinase p38*, and *JNK*. Modulates *AP1*, BCR-mediated *PLCG1*, Ca^{2+} mobilization, *PLCG2*, *NF-kappa-B*, and *NFAT*. It plays a critical role in orchestrating the pro-B cell to pre-B cell transition[20] and may play an essential role in BCR-induced B-cell apoptosis. These differentially expressed genes between patients and normal controls can explain, to some extent, the genetic susceptibility of patients and the body's response to AS.

XIST is a key initiator of X chromosome inactivation in Eutherian mammals, which may also be part of the inflammatory response.[21] *EIF1AY*, Eukaryotic translation initiation factor 1A, seems to be required for the maximal rate of protein biosynthesis. Enhances ribosome dissociation into subunits and stabilizes the binding of the initiator Met-tRNA(I) to 40S ribosomal subunits.[22] *RPS4Y1*, the ribosomal protein S4 40S ribosomal protein S4, Y isoform 1, is was extensively involved in RNA binding, multicellular organism development, nuclear-transcribed mRNA catabolic process, nonsense-mediated decay, SRP-dependent cotranslational, protein targeting to membrane, translation, translational initiation, and viral transcription. These genes are involved in the more basic biological functions of replication, translation, transcription, and they are identified by the DEG algorithm.[23] The basic blood

metabolism of AS patients has certain differences, which may be correlated with risk factor clonal hematopoiesis.

Conclusion

These mRNA molecules are still lacking clinical cohort verification, and their use as a marker of screening is still to be debated. However, the differences between normal population and AS patients to some extent can explain their correlation with AS, indicating that repeated activation of inflammation is involved in the formation and development of AS. The specific roles of *XIST*, *RPS4Y1*, and *EIF1AY* in transcription and translation and how they are related need to be verified by molecular biology, which will be of great help for us to understand the central principle further. In general, we have only scratched the surface, which provides some targets for subsequent cohort studies. Through bioinformatics analysis, our results may be beneficial for the clinical molecular diagnosis,[24,25] treatment,[26] and prognosis.[27] The associations we have found may also be helpful for more fundamental studies of biological function.[28]

Acknowledgements

We thank Dr. Jianming Zeng (University of Macau) and all the members of his bioinformatics team, biotrainee, for generously sharing their experience and codes.

Author Contributions

Lihua Li participated in the experimental design, Lili Zhang, Zhen Sun, Guangyao Zang, Yalan Li, and Zhongqun Wang participated in literature retrieval and paper writing. Yongjiang Qian conducted the data analysis.

Ethics Approval and Consent to Participate

Human studies conform to the principles outlined in the Declaration of Helsinki (1964) and was approved by the Ethical Committee of the Affiliated Hospital of Jiangsu University.

ORCID iD

Lihua Li https://orcid.org/0000-0002-2089-8140

Data Availability Statement

The datasets [GSE12288; GSE27034; GSE90074] for this study can be found in the GEO database (http://www.ncbi.nlm.nih.gov/geo/).

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

1. Virani SS, Alonso A, Benjamin EJ, et al. Heart disease and stroke statistics-2020 update: a report from the American Heart Association. *Circulation*. 2020;141:e139-e596.
2. Libby P, Buring JE, Badimon L, et al. Atherosclerosis. *Nat Rev Dis Primers*. 2019;5:56.
3. Schirmer SH, Fledderus JO, van der Laan AM, et al. Suppression of inflammatory signaling in monocytes from patients with coronary artery disease. *J Mol Cell Cardiol*. 2009;46:177-185.
4. van der Pouw Kraan TC, Schirmer SH, Fledderus JO, et al. Expression of a retinoic acid signature in circulating CD34 cells from coronary artery disease patients. *BMC Genomics*. 2010;11:388.
5. Masud R, Shameer K, Dhar A, Ding K, Kullo IJ. Gene expression profiling of peripheral blood mononuclear cells in the setting of peripheral arterial disease. *J Clin Bioinforma*. 2012;2:6.
6. Ravi S, Schuck RN, Hilliard E, et al. Clinical evidence supports a protective role for CXCL5 in coronary artery disease. *Am J Pathol*. 2017;187:2895-2911.
7. Sinnaeve PR, Donahue MP, Grass P, et al. Gene expression patterns in peripheral blood correlate with the extent of coronary artery disease. *PLoS One*. 2009;4:e7037.
8. Kiliszek M, Burzynska B, Michalak M, et al. Altered gene expression pattern in peripheral blood mononuclear cells in patients with acute myocardial infarction. *PLoS One*. 2012;7:e50054.
9. Udhaya Kumar S, Thirumal Kumar D, Bithia R, et al. Analysis of differentially expressed genes and molecular pathways in familial hypercholesterolemia involved in atherosclerosis: a systematic and bioinformatics approach. *Front Genet*. 2020;11:734.
10. Wan J, Jiang S, Jiang Y, et al. Data mining and expression analysis of differential lncRNA ADAMTS9-AS1 in prostate cancer. *Front Genet*. 2019;10:1377.
11. Fu D, Zhang B, Yang L, Huang S, Xin W. Development of an immune-related risk signature for predicting prognosis in lung squamous cell carcinoma. *Front Genet*. 2020;11:978.
12. Rajan B, Abunada T, Younes S, et al. Involvement of essential signaling cascades and analysis of gene networks in diabesity. *Genes*. 2020;11:1256.
13. Kumar SU, Madhana Priya N, Thirumal Kumar D, et al. An integrative analysis to distinguish between emphysema (EML) and alpha-1 antitrypsin deficiency-related emphysema (ADL)-A systems biology approach. *Adv Protein Chem Struct Biol*. 2021;127:315-342.
14. Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43:e47.
15. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics*. 2012;28:573-580.
16. Chapoval AI, Smithson G, Brunick L, et al. BTNL8, a butyrophilin-like molecule that costimulates the primary immune response. *Mol Immunol*. 2013;56:819-828.
17. Welch C, Santra MK, El-Assaad W, et al. Identification of a protein, G0S2, that lacks Bcl-2 homology domains and interacts with and antagonizes Bcl-2. *Cancer Res*. 2009;69:6782-6789.
18. Suply T, Hannedouche S, Carte N, et al. A natural ligand for the orphan receptor GPR15 modulates lymphocyte recruitment to epithelia. *Sci Signal*. 2017;10:496.
19. Pieper K, Grimbacher B, Eibel H. B-cell biology and development. *J Allergy Clin Immunol*. 2013;131:959-971.
20. Fu C, Turck CW, Kurosaki T, Chan AC. BLNK: a central linker protein in B cell activation. *Immunity*. 1998;9:93-103.
21. Strehle M, Guttman M. Xist drives spatial compartmentalization of DNA and protein to orchestrate initiation and maintenance of X inactivation. *Curr Opin Cell Biol*. 2020;64:139-147.
22. Dever TE, Wei CL, Benkowski LA, Browning K, Merrick WC, Hershey JW. Determination of the amino acid sequence of rabbit, human, and wheat germ protein synthesis factor eIF-4C by cloning and chemical sequencing. *J Biol Chem*. 1994;269:3212-3218.
23. Joazeiro CAP. Ribosomal stalling during translation: providing substrates for ribosome-associated protein quality control. *Annu Rev Cell Dev Biol*. 2017;33:343-368.
24. Kumar SU, Kumar DT, Siva R, Doss CGP, Zayed H. Integrative bioinformatics approaches to map potential novel genes and pathways involved in ovarian cancer. *Front Bioeng Biotechnol*. 2019;7:391.
25. Kumar SU, Thirumal Kumar D, Siva R, et al. Dysregulation of signaling pathways due to differentially expressed genes from the B-Cell transcriptomes of Systemic Lupus Erythematosus patients – a bioinformatics approach. *Front Bioeng Biotechnol*. 2020;8:276.
26. Kumar SU, Saleem A, Thirumal Kumar D, et al. A systemic approach to explore the mechanisms of drug resistance and altered signaling cascades in extensively drug-resistant tuberculosis. *Adv Protein Chem Struct Biol*. 2021;127:343-364.
27. Yan H, Zheng G, Qu J, et al. Identification of key candidate genes and pathways in multiple myeloma by integrated bioinformatics analysis. *J Cell Physiol*. 2019;234:23785-23797.
28. Mishra S, Shah MI, Udhaya Kumar S, et al. Network analysis of transcriptomics data for the prediction and prioritization of membrane-associated biomarkers for idiopathic pulmonary fibrosis (IPF) by bioinformatics approach. *Adv Protein Chem Struct Biol*. 2021;123:241-273.