Biomarkers of Blood from Patients with Atherosclerosis Based on Bioinformatics Analysis

Yongjiang Qian
Affiliated Hospital of Jiangsu University
Lili Zhang
Affiliated Hospital of Jiangsu University
Zhen Sun
Affiliated Hospital of Jiangsu University
Guangyao Zang
Affiliated Hospital of Jiangsu University
Yalan Li
Affiliated Hospital of Jiangsu University
Zhongqun Wang
Affiliated Hospital of Jiangsu University
Lihua Li (tsmc01@163.com)
Jiangsu University Hospital https://orcid.org/0000-0002-2089-8140

Research article

Keywords: Blood, Bioinformatics, Biomarker, Atherosclerosis.

Posted Date: May 20th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-425622/v2

Abstract

Background: atherosclerosis is a multifaceted disease characterized by the formation and accumulation of plaques that fix to the arteries and causes some cardiovascular disease and vascular embolism. A range of diagnostic techniques, including selective coronary angiography, stress tests, CT, and nuclear scans allow assessment of cardiovascular disease risk and treatment targets. However, there is not a very simple blood biochemical index or biological target for the diagnosis of atherosclerosis at present. So it would be interesting to find a blood biochemical marker for atherosclerosis.

Methods: Three datasets from Gene Expression Omnibus (GEO) database were analyzed to obtain differentially expressed genes (DEG) and the results were integrated using Robustrankaggreg algorithm. The genes considered more important by Robustrankaggreg algorithm were put into their own data set and the data set system with cell classification information for verification.

Results: 21 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

Conclusions: In this study, we identified *BTNL8* and *BLNK* as having good clinical significance. These results will contribute to the study of the underlying genes involved in the progression of atherosclerosis and provide insights for the discovery of new diagnostic and evaluation methods.

Background

Atherosclerosis (AS) is the main cause of coronary heart disease, peripheral vascular disease and cerebral infarction[1]. The development of atherosclerotic lesions may be caused by low-density lipoprotein, a lipoprotein that carries cholesterol through the bloodstream. Other risk factors of atherosclerosis and its thrombotic complications include smoking, diabetes and high blood pressure[2]. A growing evidence also points out that a role of the immune system, as emerging risk factors include inflammation and clonal hematopoiesis. A range of diagnostic techniques, including 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 not a very simple blood biochemical index or biological target for the diagnosis of atherosclerosis at present, but more ultrasonographic screening or angiography are used[3]. So it would be valuable to find a blood biochemical marker for atherosclerosis.

With the development of omics, and due to the availability of clinical blood samples, many studies have focused on blood transcriptome of patients with atherosclerosis. Transcriptome analysis of circulating mononuclear cells from carefully matched atherosclerotic and control patients will potentially provide insights into the pathophysiology of atherosclerosis and supply biomarkers for diagnostic purposes[4–7]. 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[8], Other studies have looked at the transcriptome of

peripheral blood and the transcriptional expression of circulating cells in patients with acute myocardial infarction or artery plaque[9], They suggest that genes that mediate immune response, inflammation, apoptosis, stress response, phosphorylation, hemostasis, platelet activation and platelet aggregation may play an important role[10], it also provides some ideas for the subsequent experimental research.

In this study, after the detection of different 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*.

Method

Retrieve

Keywords “atherosclerosis” and “blood” were searched in the GEO database and the specie was limited in “Homo sapiens”. 59 data sets were retrieved, and then we manually excluded the mRNA chip data sets unrelated to the blood of atherosclerosis patients and not clearly grouped, and finally three data sets were screened out (Table.1).

Quality control

We assessed the basic data using R language assessment and quality control, all the expression of matrix through log2 processing, and use ggpolt2 package for drawing, has carried on the related PCA (Principal Component Analysis)[11]. PCA is a common way of data Analysis and often used for dimensionality reduction of high-dimensional data to extract the main characteristics of the Component. Firstly, the data are preprocessed, and then the co-prevention matrix of matrix X is calculated. Then, the eigenvalues and eigenvectors of the co-prevention difference matrix are calculated. On the hand, the expression density, box diagram of each expression quantity were drawn to sure each data has good comparability (Fig.s1, Fig.s2 and Fig.s3).

Differentially expressed genes (DEGs)

The samples were divided into case and control groups according to the information on GEO. The LIMMA package of R language was used to analyze the differential genes[12], and the logFc values of most genes were between − 1 and 1. We screened the upregulated genes with a logFc value greater than 0.5 and a p value less than 0.01, and the down-regulated genes with a logFc value less than − 0.5 and p-value less than 0.01.

Robustrankaggreg

Robustrankaggreg R package was used to integrate the up-down-regulated genes[13], respectively. 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 (Fig. 1).

Genetic alignment and correlation analysis

The expression matrices of the identified genes were selected from the original data set and GSE9820[4], and the unclustered and clustered heat maps were constructed (Fig. 2, Fig. 3, Fig. 4,. Figure 5 and Fig. 6) The genes of interest were plotted in a scatter plot. P.value < 0.05.

Results

1. Genes detected according to the integrated DEGs

Although the logFC values of most of the differential genes identified in the whole blood transcriptome were between − 1 and 1, through the analysis and screening of the three datasets, we still obtained 21 genes 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 significant difference in GSE90074. These genes may be used for screening and evaluation of AS or vascular plaques.

2. 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, after clustering the heat maps, we found an interesting phenomenon for the first time: *XIST* is negatively correlated with RPS4Y1 in all the three 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.

3. Validation in different cell types.

We picked up the expression of these selected genes in the data set of GSE9820[4], which is a sequencing data of Mononuclear Cell Transcriptomes, and identified five 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 five 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 five kinds of cells. *RPS4Y1* and *EIF1AY* were not tissue 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 three of coronary atherosclerosis in patients with blood samples mrna array dataset to filter possible coronary atherosclerosis possible genetic detection object in the blood. we found there are 21 genes may have certain significance and also discussed these gene expressions between different cells in the blood. this study first reported *RPS4Y1*, *EIF1AY*, correlation between XIST.

Many of these genes are associated with inflammation and immunity. *BTNL8* which have the best score may stimulate primary immune response, acts on T-cell stimulated sub-optimally through the TCR/CD3 complex stimulating their proliferation and cytokine production[14]. *G0S2*, G0/G1 switch protein 2, promotes apoptosis by binding to *BCL2*, resulting in preventing the formation of protective Bcl2-Bax heterodimers[15]. *GPR15L* is chemotactic factor that mediates lymphocytes recruitment to epithelia through binding and activation of the G-protein coupled receptor GPR15 seems to be epithelia related[16]. *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[17]. What is more, *BLNK* plays a role in the activation of *ERK/ EPHB2*, *MAP kinase p38* and *JNK*. Modulates *AP1*, BCR-mediated *PLCG1*, Ca2 + mobilization, *PLCG2*, *NF-kappa-B* and *NFAT*. It plays a critical role in orchestrating the pro-B cell to pre-B cell transition[18]. May play an important 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 mammal, which may also part in the inflammatory response[19]. *EIF1AY*, Eukaryotic translation initiation factor 1A,seems to be required for maximal rate of protein biosynthesis. enhances ribosome dissociation into subunits and stabilizes the binding of the initiator Met-tRNA(I) to 40 S ribosomal subunits. ribosomal subunits[20]. *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, revelation, transcription, and they are identified by the DEG algorithm[21]. The basic blood metabolism of AS patients has certain differences, and this may be correlated with risk factor clonal hematopoiesis.

Conclusion

These mRNA molecules are still lack of 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 further understand the central principle. In general, we have only scratched the surface, which

provides some targets for subsequent cohort studies, and the associations we have found may also be useful for more basic studies of biological function.

Abbreviations

Gene Expression Omnibus (GEO)

differentially expressed genes (DEG)

Atherosclerosis (AS)

Declarations

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

Consent for publication

Not applicable

Conflict of Interest

No conflict of interest.

Sources of Funding

This work was supported as follows: the National Natural Science Foundation of China (82070455, 81770450, 81370408); the related Foundation of Jiangsu Province (BK20201225, WSN-044 LGY2018092, QNRC2016836); the Open Project Program of Guangxi Key Laboratory of Centre of Diabetic Systems Medicine (GKLCDSM-20210101-02); Research and Innovation Funding Project for College Students in Experimental Animal Center of Jiangsu University; Postgraduate Research Practice Innovation Program of Jiangsu Province (KYCX20_2881, SJKY19_2585); Zhenjiang Cardiovascular Clinical Research Center Project (SS2018008)

Author Contribution Statement

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.

Acknowledgments

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

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/;).

References

[1] S.S. Virani, A. Alonso, E.J. Benjamin, M.S. Bittencourt, C.W. Callaway, A.P. Carson, A.M. Chamberlain, A.R. Chang, S. Cheng, F.N. Delling, L. Djousse, M.S.V. Elkind, J.F. Ferguson, M. Fornage, S.S. Khan, B.M. Kissela, K.L. Knutson, T.W. Kwan, D.T. Lackland, T.T. Lewis, J.H. Lichtman, C.T. Longenecker, M.S. Loop, P.L. Lutsey, S.S. Martin, K. Matsushita, A.E. Moran, M.E. Mussolino, A.M. Perak, W.D. Rosamond, G.A. Roth, U.K.A. Sampson, G.M. Satou, E.B. Schroeder, S.H. Shah, C.M. Shay, N.L. Spartano, A. Stokes, D.L. Tirschwell, L.B. VanWagner, C.W. Tsao, Heart Disease and Stroke Statistics-2020 Update: A Report From the American Heart Association, Circulation, 141 (2020) e139-e596.

[2] W. Herrington, B. Lacey, P. Sherliker, J. Armitage, S. Lewington, Epidemiology of Atherosclerosis and the Potential to Reduce the Global Burden of Atherothrombotic Disease, Circulation research, 118 (2016) 535-546.

[3] P. Libby, J.E. Buring, L. Badimon, G.K. Hansson, J. Deanfield, M.S. Bittencourt, L. Tokgozoglu, E.F. Lewis, Atherosclerosis, Nat Rev Dis Primers, 5 (2019) 56.

[4] S.H. Schirmer, J.O. Fledderus, A.M. van der Laan, T.C. van der Pouw-Kraan, P.D. Moerland, O.L. Volger, J.M. Baggen, M. Böhm, J.J. Piek, A.J. Horrevoets, N. van Royen, Suppression of inflammatory signaling in monocytes from patients with coronary artery disease, Journal of molecular and cellular cardiology, 46 (2009) 177-185.

[5] T.C. van der Pouw Kraan, S.H. Schirmer, J.O. Fledderus, P.D. Moerland, J.M. Baggen, T.A. Leyen, A.M. van der Laan, J.J. Piek, N. van Royen, A.J. Horrevoets, Expression of a retinoic acid signature in circulating CD34 cells from coronary artery disease patients, BMC Genomics, 11 (2010) 388.

[6] R. Masud, K. Shameer, A. Dhar, K. Ding, I.J. Kullo, Gene expression profiling of peripheral blood mononuclear cells in the setting of peripheral arterial disease, J Clin Bioinforma, 2 (2012) 6.

[7] S. Ravi, R.N. Schuck, E. Hilliard, C.R. Lee, X. Dai, K. Lenhart, M.S. Willis, B.C. Jensen, G.A. Stouffer, C. Patterson, J.C. Schisler, Clinical Evidence Supports a Protective Role for CXCL5 in Coronary Artery Disease, Am J Pathol, 187 (2017) 2895-2911.

[8] P.R. Sinnaeve, M.P. Donahue, P. Grass, D. Seo, J. Vonderscher, S.D. Chibout, W.E. Kraus, M. Sketch, Jr., C. Nelson, G.S. Ginsburg, P.J. Goldschmidt-Clermont, C.B. Granger, Gene expression patterns in peripheral

blood correlate with the extent of coronary artery disease, PLoS One, 4 (2009) e7037.

[9] E.D. Muse, E.R. Kramer, H. Wang, P. Barrett, F. Parviz, M.A. Novotny, R.S. Lasken, T.A. Jatkoe, G. Oliveira, H. Peng, J. Lu, M.C. Connelly, K. Schilling, C. Rao, A. Torkamani, E.J. Topol, A Whole Blood Molecular Signature for Acute Myocardial Infarction, Scientific reports, 7 (2017) 12268.

[10] M. Kiliszek, B. Burzynska, M. Michalak, M. Gora, A. Winkler, A. Maciejak, A. Leszczynska, E. Gajda, J. Kochanowski, G. Opolski, Altered gene expression pattern in peripheral blood mononuclear cells in patients with acute myocardial infarction, PLoS One, 7 (2012) e50054.

[11] T. Konishi, Principal component analysis for designed experiments, BMC bioinformatics, 16 Suppl 18 (2015) S7.

[12] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, G.K. Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic acids research, 43 (2015) e47.

[13] R. Kolde, S. Laur, P. Adler, J. Vilo, Robust rank aggregation for gene list integration and meta-analysis, Bioinformatics (Oxford, England), 28 (2012) 573-580.

[14] A.I. Chapoval, G. Smithson, L. Brunick, M. Mesri, F.L. Boldog, D. Andrew, N.V. Khramtsov, E.A. Feshchenko, G.C. Starling, P.S. Mezes, BTNL8, a butyrophilin-like molecule that costimulates the primary immune response, Molecular immunology, 56 (2013) 819-828.

[15] C. Welch, M.K. Santra, W. El-Assaad, X. Zhu, W.E. Huber, R.A. Keys, J.G. Teodoro, M.R. Green, Identification of a protein, G0S2, that lacks Bcl-2 homology domains and interacts with and antagonizes Bcl-2, Cancer research, 69 (2009) 6782-6789.

[16] T. Suply, S. Hannedouche, N. Carte, J. Li, B. Grosshans, M. Schaefer, L. Raad, V. Beck, S. Vidal, A. Hiou-Feige, N. Beluch, S. Barbieri, J. Wirsching, N. Lageyre, F. Hillger, C. Debon, J. Dawson, P. Smith, V. Lannoy, M. Detheux, F. Bitsch, R. Falchetto, T. Bouwmeester, J. Porter, B. Baumgarten, K. Mansfield, J.M. Carballido, K. Seuwen, F. Bassilana, A natural ligand for the orphan receptor GPR15 modulates lymphocyte recruitment to epithelia, Science signaling, 10 (2017).

[17] K. Pieper, B. Grimbacher, H. Eibel, B-cell biology and development, The Journal of allergy and clinical immunology, 131 (2013) 959-971.

[18] C. Fu, C.W. Turck, T. Kurosaki, A.C. Chan, BLNK: a central linker protein in B cell activation, Immunity, 9 (1998) 93-103.

[19] M. Strehle, M. Guttman, Xist drives spatial compartmentalization of DNA and protein to orchestrate initiation and maintenance of X inactivation, Current opinion in cell biology, 64 (2020) 139-147.

[20] T.E. Dever, C.L. Wei, L.A. Benkowski, K. Browning, W.C. Merrick, J.W. Hershey, Determination of the amino acid sequence of rabbit, human, and wheat germ protein synthesis factor eIF-4C by cloning and

chemical sequencing, The Journal of biological chemistry, 269 (1994) 3212-3218.

[21] C.A.P. Joazeiro, Ribosomal Stalling During Translation: Providing Substrates for Ribosome-Associated Protein Quality Control, Annual review of cell and developmental biology, 33 (2017) 343-368.

Figures

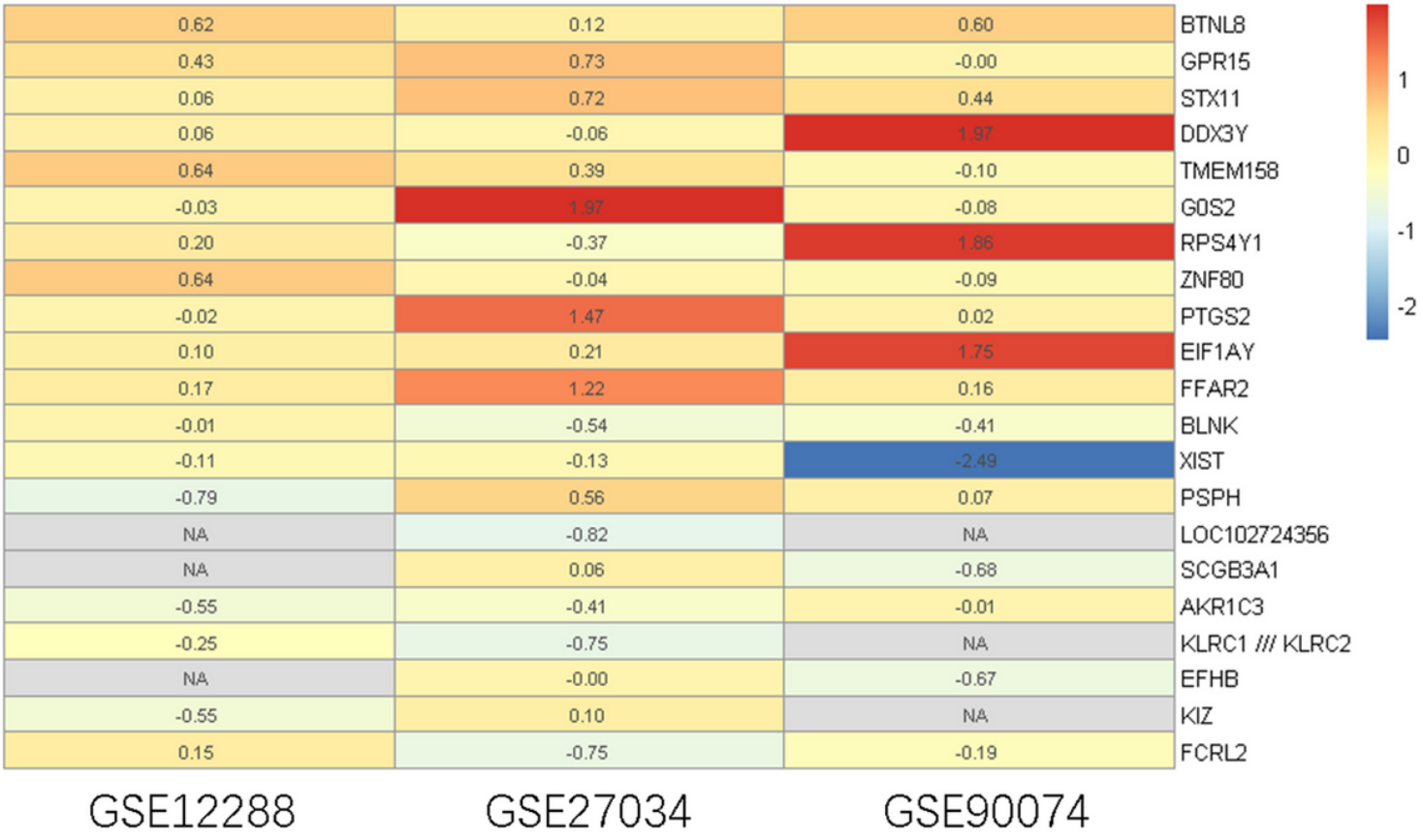

Figure 1

LogFC of Genes were identified in three datasets

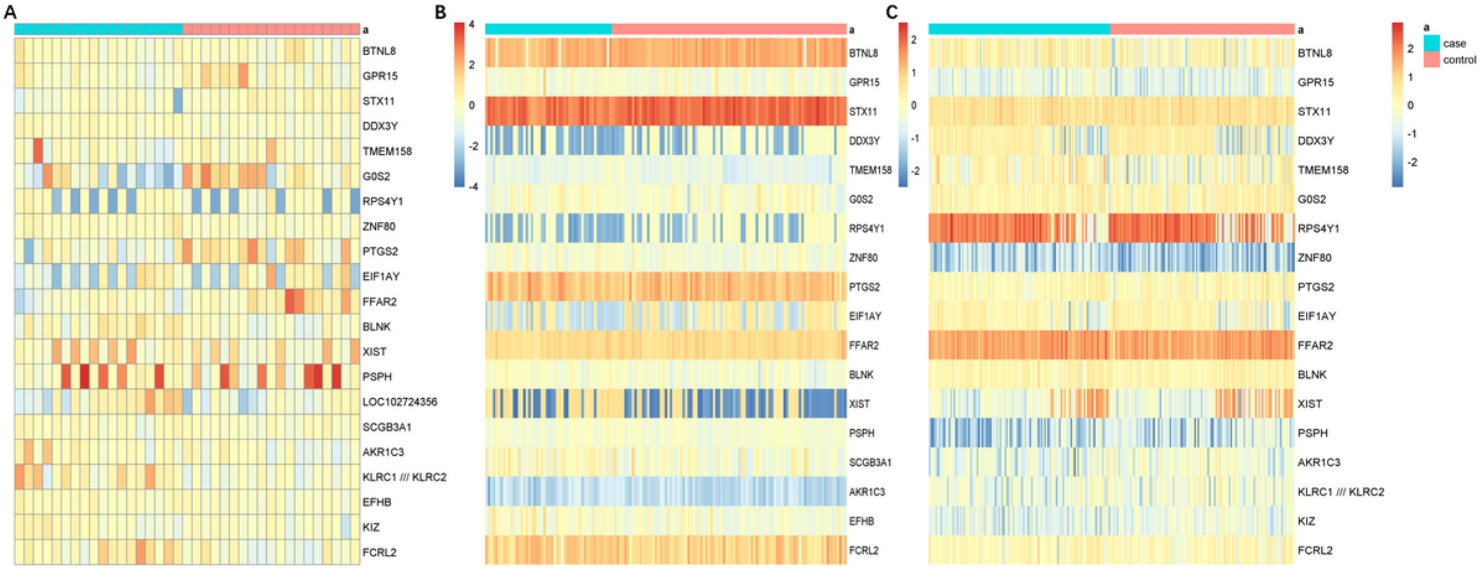

Figure 2

Gene expression in three datasets. Unclustered heat map of gene expression in GSE27034(A) Unclustered heat map of gene expression expression in GSE90074(B) Unclustered heat map of gene expression expression in GSE12288(C)

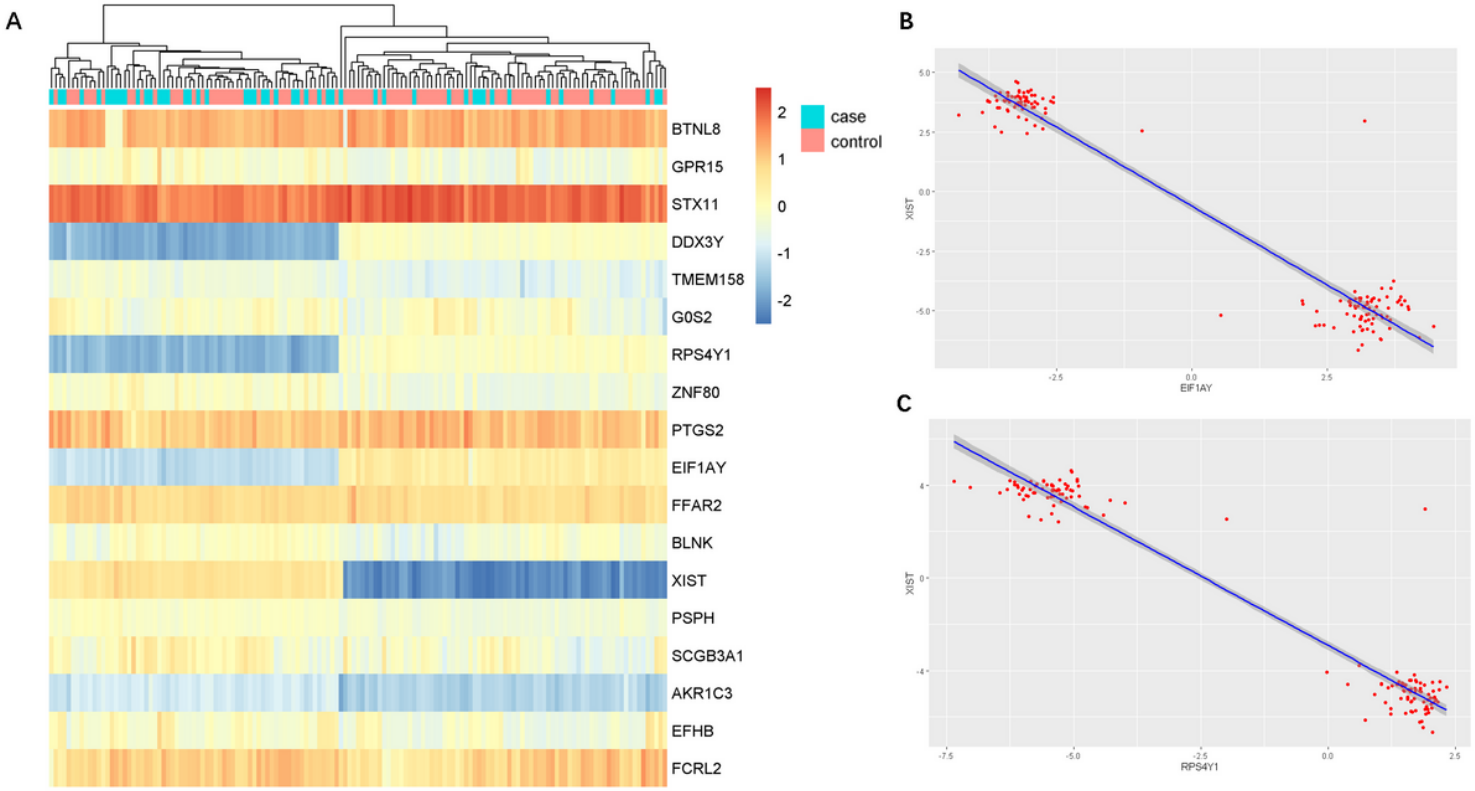

Figure 3

Gene expression in GSE90074 and scatter diagrams Clustered heat map of gene expression in GSE90074(A) Scatter diagram of XIST and EIF1AY (B) Scatter diagram of XIST and RPS4Y1 (C)

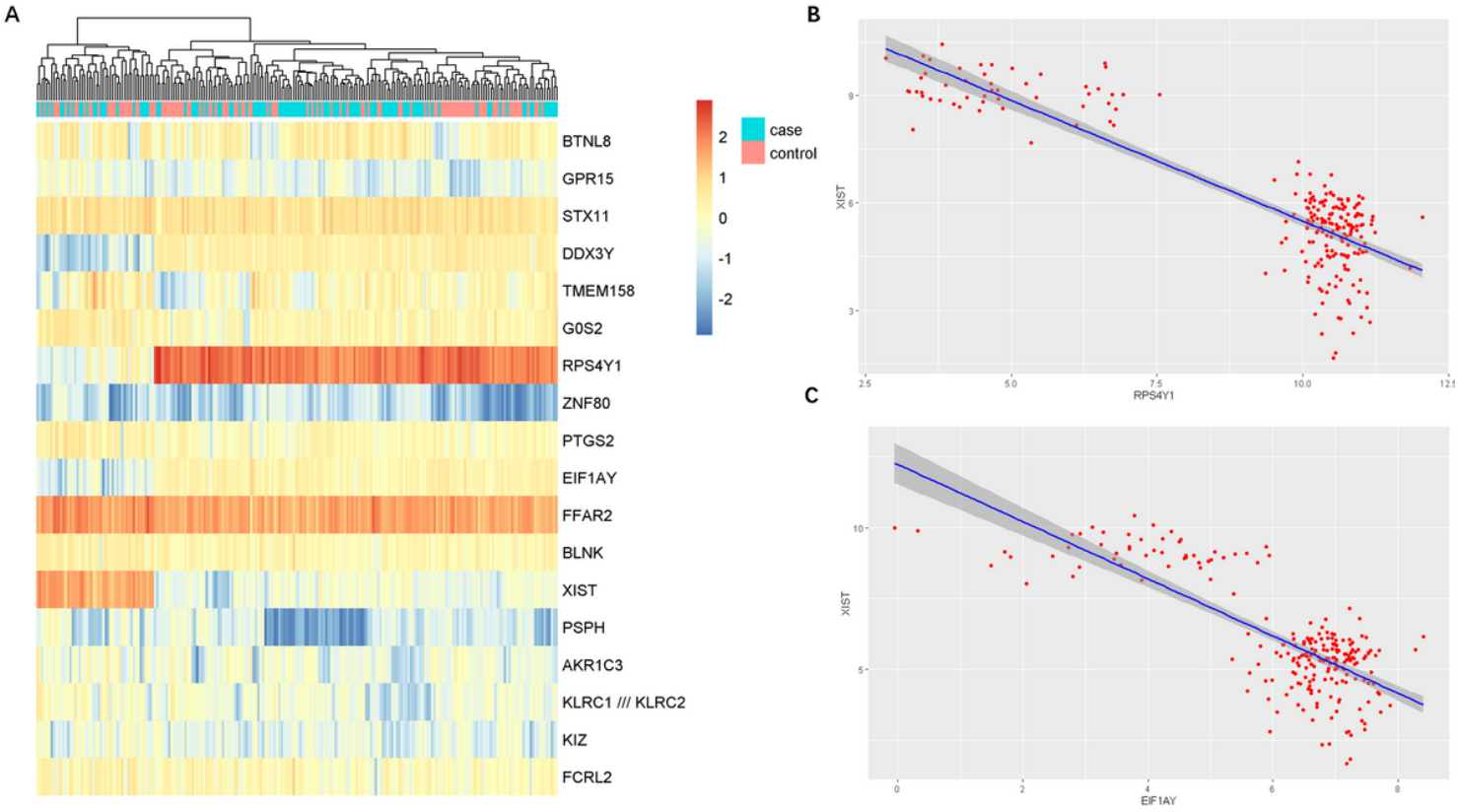

Figure 4

Gene expression in GSE12288 and scatter diagrams Clustered heat map of gene expression in GSE12288(A) Scatter diagram of XIST and RPS4Y1 (B)Scatter diagram of XIST and EIF1AY (C)

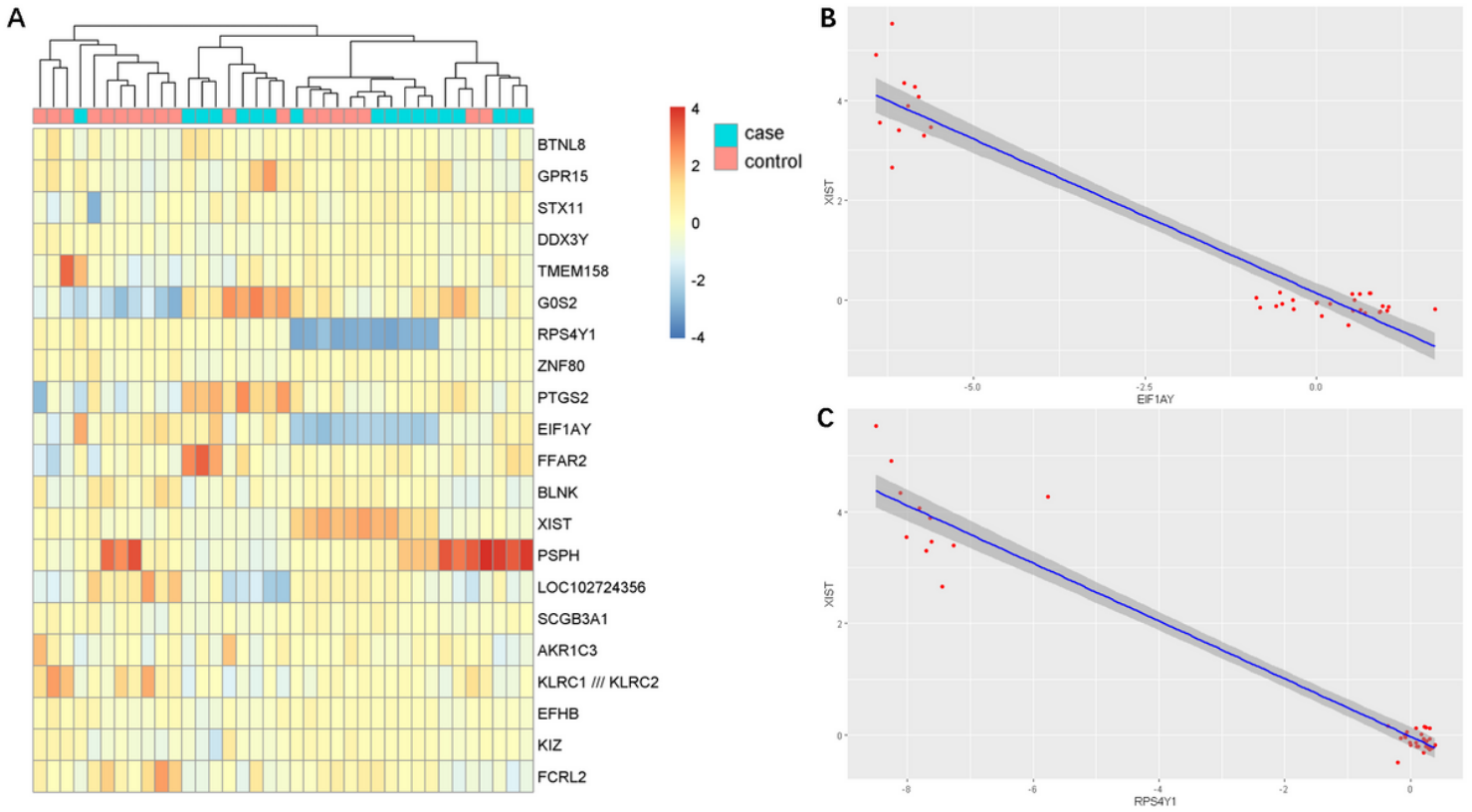

Figure 5

Gene expression in GSE27034 and scatter diagrams Clustered heat map of gene expression in GSE27037(A) Scatter diagram of XIST and EIF1AY (B) Scatter diagram of XIST and RPS4Y1 (C)

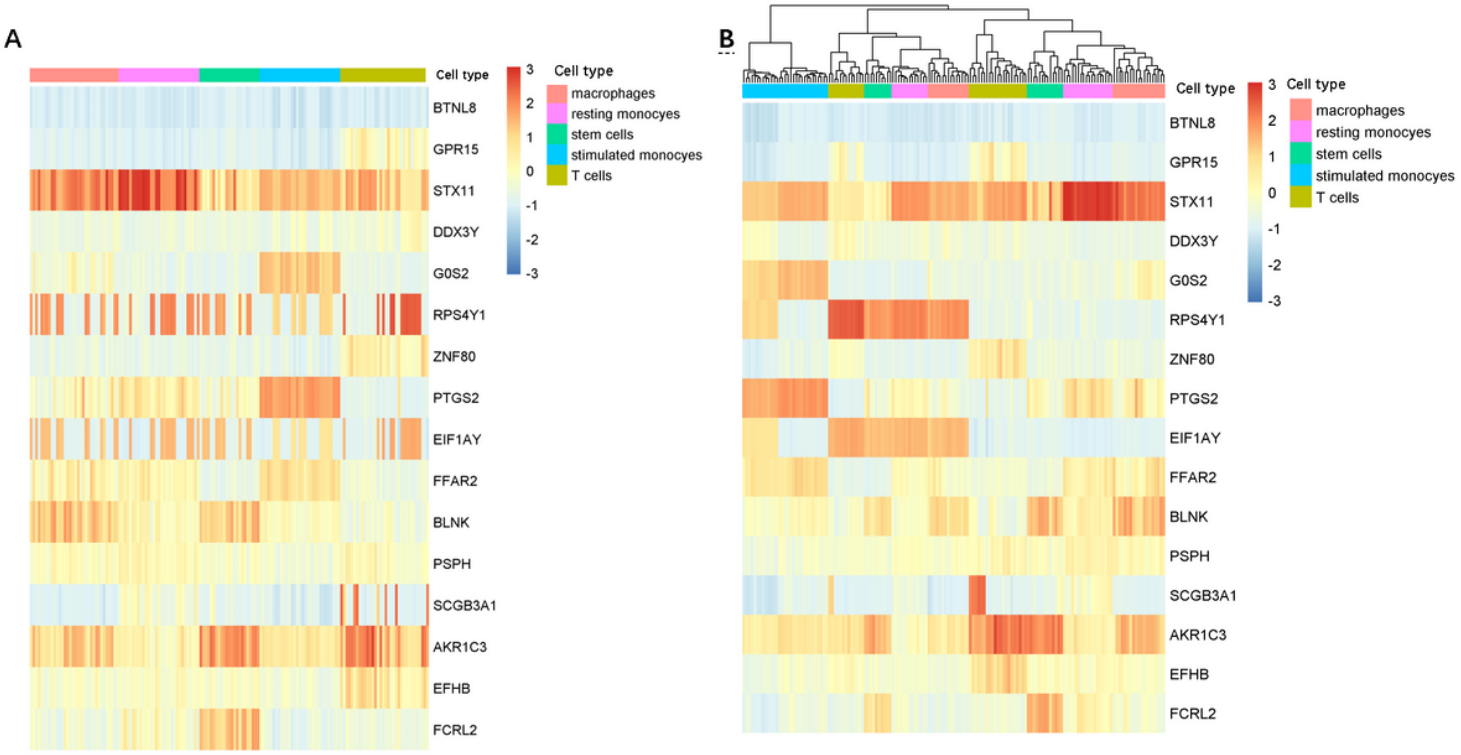

Figure 6

Gene expression in GSE9820 Unclustered heat map of gene expression in GSE90082(A) Clustered heat map of gene expression in GSE90082(B)

Supplementary Files

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

- Fig.s1.png
- Fig.s2.png
- Fig.s3.png