Utilizing Bioinformatics Process to Compare the Single-Cell Sequencing Results of Normal and Ascending Thoracic Aortic Aneurysm Tissue Samples to Obtain Key Genes from Holistic Level

Yang Yu
The First Affiliated Hospital of China Medical University
Zheng Ding
The First Affiliated Hospital of China Medical University
Enyi Shi
The First Affiliated Hospital of China Medical University
Tianxiang Gu (sey2004@sina.com)
The First Affiliated Hospital of China Medical University

Research Article

Keywords: Ascending thoracic aortic aneurysm, Gene network, Protein-protein interactions, Single cell sequencing, Functional enrichment analysis, Bioinformatics analysis

Posted Date: February 24th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1321051/v1

Abstract

Background

The thoracic aortic aneurysms especially ascending aortic aneurysms are lethal and tend to be rupture and dissected. The prognosis of untreated aortic aneurysms is worse compared with surgical treatment, however surgical treatment is still a challenging operation with massive trauma and hard to perform on the appropriate time. Medical treatment is mainly to treat the risk factors but not the pathogenesis. With the exploration of the mechanisms of the diseases from onset to development till deterioration or even death leading, the medicine treatment of aortic aneurysms will become more well-targeted and effective.

The single cell RNA sequencing is widely used to reveal the function of an individual cell in the context of its microenvironment. After the gene differences study at the cell clusters level, we attempt to obtain the overall differential expression analysis through the algorithm on this basis. This is not only an overall exploration of the mechanism of this disease, but also an attempt to propose a new analytical method to incorporate both local and holistic results from the single-cell sequencing outcome.

Methods

All DEGs are obtained by starting from the processing of the original samples. After ensuring no statistical significance, we extracted gene expression array and finally to the results intersection of the edgeR and DESeq models. The following part is bioinformatics analysis process, including enrichment analysis such as GO analysis and KEGG pathways, the presentation of PPI network, and the target prediction associated with network analysis used to select the hub genes.

Results

By comparing the gene expression of normal ascending aorta tissue and ascending aorta tissue of ATAA patients, we screened 1390 DEGs, including 526 up-regulated DEGs and 864 down-regulated DEGs. The selected hub genes are: *CDC20, CCNB2, BUB1B, BIRC5, PTTG1, ESPL1, PPP1R12B, MYLK, MYL9, MYH14, MYH11*. In the classification process using k-means clustering algorithm, *MYLK* and *MYH11* gene showed the characteristics of being in the center and difficult to be classified, indicating that they have a certain pivotal effect. It is consistent with the observation of the characteristics of the small world network. The results obtained by this method is highly compatible with the consensus and can explain the mechanisms of the aortic aneurysms progressing.

Conclusion

CDC20, CCNB2, BUB1B, BIRC5, PTTG1, ESPL1 genes as down-regulated hub genes, *PPP1R12B, MYLK, MYL9, MYH14, MYH11* were up-regulated hub genes associated with ATAA. *MYLK* and *MYH11* are in the strong group in the consensus of experts. Also, we provided a method to convert single-cell mRNA sequencing results to whole-exome sequencing results, thereby enabling further thinking about molecular pathways of disease at the global and local level in single cell sequencing result.

1. Introduction

Ascending thoracic aortic aneurysm (ATAA) is aortic enlargement caused by various reasons [1]. And surgery is the only treatment method. In some literatures, it was called the 18th most common cause of death in adults[2, 3]. According to the findings in recent years, ATAA has a certain family genetic tendency[4]. If genetic screening can be performed early, which will greatly reduce the risk of ATAA rupture and then, to keep patients alive.

In the field of cardiac surgery, within our knowledge, we have not found the use of bioinformatics analysis process to screen the hub genes from general original samples. Therefore, not only do we try to analyze the single-cell sequencing results in a holistic level, but further, we are using a sequencing result to jointly analyze the whole and the parts, so that there will be a more in-depth discussion of the microscopic mechanism of a disease.

Our screening step is to use a series of sample processing algorithms to obtain gene expression with no statistical significance of cell clusters, then, use edgeR model[5] and DESeq model[6] to obtain differentially expressed genes (DEGs) respectively, and after intersection, perform GO[7] and KEGG enrichment analysis on DEGs[8]. Next step is using PPI network to present interaction of DGEs. Finally, the hub genes are determined based on the connection degree of the nodes and the results of KEGG pathway, and our screening results are verified according to an expert consensus on genetics[9].

2. Material And Methods

2.1 Data resource

We searched all databases in the National Center for Biotechnology Information(NCBI)[10], downloaded original samples, which contain ascending thoracic aortic aneurysm (ATAA) (GSM4704934、GSM4704935、GSM4704936、GSM4704937、GSM4704938、GSM4704939、GSM4704940、GSM47049341) and normal samples(GSM4704931、GSM4704932、GSM4704933) from Sequence Read Archive(SRA) database (https://www.ncbi.nlm.nih.gov/sra). All samples are used platform GPL24676 Illumina NovaSeq 6000 to obtain high-throughput sequencing outcome.

2.2 Data processing

i. Unzip: Under Windows 10 system, by using sratoolkit tools (version 2.11.0-win64) [11] to translate files with SRA format to fastq format. Cause all samples are double paired, so we used “—split -e” commend in sratoolkit package to obtain the front-end and back-end sequencing files separately.

ii. Statistical significance validation: We pooled the mean and standard deviation of each cell ratio in the normal and ATAA samples in the tissue and used a fixed-effects model[12-15] to restore the cell ratio of the original tissue as much as possible. After Z-test[16], we finally proved that the two There was no statistical difference between the tissues (p=0.36), and the following methods were used to

calculate and screen the expression levels of the differential genes, to obtain the differential genes of the whole tissue.

iii. End sequence splicing: Under linux system (The version is Ubuntu 21.04. Unless otherwise specified, all the following tools are performed under this linux system.), according to parameters in Illumina platform, we used trim_galore(version TrimGalore-0.4.5, http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) [17] package with default parameters and "—paired" commend to cut each single sequence before and after modified suffix in the "fastq" file. Finally, we got file with "fq" format.

iv. Quality assessment: By using "fastqc" software (FastQC version: 0.11.9)[18]to "fq" samples mentioned before, the quality assessment results obtained are all above moderate. Therefore, we don't need to delete our samples.

v. Index establish: We used BWA software(version: bwa-0.7.17)[19] with index tool to construct index based on the human genome reference "hg38.fa"(GRCh38 Genome Reference Consortium Human Reference 38, downloaded from https://hgdownload.cse.ucsc.edu/goldenpath/hg38/bigZips/.), then using "mem" tool in bwa package to compare and combine two "fq" files with index file through the mapping process, acquired combination outcome in "sam" format. Here, we acquired files with chaos single sequence, after this step, we need to put each sequence in right place and furthermore, to calculate each gene's expression mounts so we can do analysis better.

vi. Sequence sort: Consider about capability of disk, we compressed the "sam" format files to "bam" format, which will reduce the required disk space to 1/6. Then, we used "sort" tool in samtools package[20] to sort files, the purpose is shaping sequences to right position of chromosome. So, we can count gene expression data properly. The final step is to construct file index of samples, and we got "bai" format files, these are used to see in IGV software (Version: IGV_2.10.2).

vii. Feature counts: Ultimately, using the FeatureCounts function in subread package (Version: subread-2.0.2)[21] to calculate gene expression under the instruction of hg38.ncbiRefSeq.gtf file (https://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/genes/) which is gene reference sequence. Then, we got the quantitative analysis outcome with the help of DESeq2 package in R language (Version: 4.1.0).

viii. Inclusion Criteria: After using DESeq2 model and edgeR model to calculate feature counts dataset, we defined the inclusion criteria as $P<0.05$ and $|\text{fold change}|>2$ in DESeq2 model and $P<0.05$ in edgeR model to find DEGs. The final DEGs included is the result of gene screening of the two strategies after the intersection.

2.3 Enrichment analysis

The DAVID online analysis website (https://david.ncifcrf.gov/)[22] was used to analyze the DEGs, including Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. Besides, $P < 0.05$ was considered to have significant statistical significance.

2.4 Protein–protein interaction (PPI) network construction and hub gene identification

The STRING database (https://string-db.org/)[23] is used to generate the PPI network connection. After exporting the node gene connection results, the genes with the combined score > 0.9 are selected and imported into Cytoscape software (version: 3.8.2), and the Molecular Complex Detection (MCODE)[24] plug-in is used for cluster identification. The parameter settings are as follows: degree cutoff = 2, node score cutoff = 0.2, k-score = 2, max-depth = 100. We excluded clusters with immune function and included genes that exist in both the KEGG pathway and clusters as hub genes.

2.5 Gene validation based on expert consensus

To avoid the contingency of the results, all the steps used the currently commonly bioinformatics algorithms and DEGs screening and extraction strategies. Finally, the results were compared with the genes of the ATAA strongly related genome included in the previously published expert consensus articles to see if there contained overlap, which was used to verify our results.

3. Results

3.1 Identification of DEGs

The featurecounts of gene expression were obtained. After the intersection of the two differential gene calculation model results, a total of 1390 DEGs were obtained, including 526 up-regulated DEGs and 864 down-regulated DEGs. In addition, we selected the 50 most up-regulated and down-regulated genes (100 DEGs in total) to make a heat map using the R language, which is shown in Fig. 1. Red represents high expression, and blue represents low expression.

3.2 Enrichment analysis

In the GO and KEGG analysis, we obtained statistically significant 264 GO pathways and 33 KEGG pathways, and screened some pathways related to the pathogenesis of ATAA, due to the occurrence of cascade reactions in our cognition affected by many factors, we selected some pathways closely related to the onset of ATAA for presentation. The results are shown in Table 1.

Table 1
GO enrichment analysis and KEGG pathways of ascending aortic aneurysm.

Category	Term	Count	%	P-Value
GOTERM_BP_DIRECT	GO:0010628 ~ positive regulation of gene expression	25	1.941	2.64E-04
GOTERM_BP_DIRECT	GO:0006936 ~ muscle contraction	13	1.009	0.001659
GOTERM_BP_DIRECT	GO:0010667 ~ negative regulation of cardiac muscle cell apoptotic process	5	0.388	0.002829
GOTERM_BP_DIRECT	GO:0048739 ~ cardiac muscle fiber development	4	0.311	0.003451
GOTERM_BP_DIRECT	GO:0007507 ~ heart development	17	1.32	0.004218
GOTERM_BP_DIRECT	GO:0010460 ~ positive regulation of heart rate	5	0.388	0.004615
GOTERM_BP_DIRECT	GO:0060048 ~ cardiac muscle contraction	7	0.543	0.010473
GOTERM_BP_DIRECT	GO:0055008 ~ cardiac muscle tissue morphogenesis	4	0.311	0.011969
GOTERM_BP_DIRECT	GO:0060045 ~ positive regulation of cardiac muscle cell proliferation	5	0.388	0.012029
GOTERM_BP_DIRECT	GO:0001525 ~ angiogenesis	18	1.398	0.012322
GOTERM_BP_DIRECT	GO:0086064 ~ cell communication by electrical coupling involved in cardiac conduction	4	0.311	0.015085
GOTERM_BP_DIRECT	GO:0007512 ~ adult heart development	4	0.311	0.015085
GOTERM_CC_DIRECT	GO:0005615 ~ extracellular space	92	7.143	8.50E-07
GOTERM_CC_DIRECT	GO:0030175 ~ filopodium	11	0.854	5.51E-04
GOTERM_CC_DIRECT	GO:0030027 ~ lamellipodium	15	1.165	0.005583
GOTERM_CC_DIRECT	GO:0030485 ~ smooth muscle contractile fiber	3	0.233	0.009306
GOTERM_MF_DIRECT	GO:0008307 ~ structural constituent of muscle	10	0.776	3.88E-05
GOTERM_MF_DIRECT	GO:0003779 ~ actin binding	21	1.63	0.010037
GOTERM_MF_DIRECT	GO:0005516 ~ calmodulin binding	16	1.242	0.010518
KEGG_PATHWAY	hsa04060: Cytokine-cytokine receptor interaction	35	2.717	4.93E-09
KEGG_PATHWAY	hsa04660: T cell receptor signaling pathway	19	1.475	5.82E-07

Category	Term	Count	%	P-Value
KEGG_PATHWAY	hsa05340: Primary immunodeficiency	11	0.854	1.94E-06
KEGG_PATHWAY	hsa04640: Hematopoietic cell lineage	15	1.165	4E-05
KEGG_PATHWAY	hsa04650: Natural killer cell mediated cytotoxicity	17	1.32	0.000147
KEGG_PATHWAY	hsa04514: Cell adhesion molecules (CAMs)	17	1.32	0.000835
KEGG_PATHWAY	hsa04270: Vascular smooth muscle contraction	15	1.165	0.000968
KEGG_PATHWAY	hsa04080: Neuroactive ligand-receptor interaction	26	2.019	0.001006

3.3 PPI network

The PPI network we obtained using string database is shown in Figure 2. After exporting the node's interaction situation, by setting the parameter to combine score> 0.9, a total of 582 genes were selected to be included in the cluster identification. After processing by the MCODE plug-in, we had a total of 6 clusters. After excluding immune-related clusters, there were three more clusters which are shown in Figure 3.

3.4 Hub gene selection

After we selected genes that exist in both the KEGG pathway and clusters, we got the following genes included: *CDC20, CCNB2, BUB1B, BIRC5, PTTG1, ESPL1, PPP1R12B, MYLK, MYL9, MYH14, MYH11*. By searching data sources, we found that the *CDC20, CCNB2, BUB1B, BIRC5, PTTG1, ESPL1* genes were specifically expressed as down-regulation, and *PPP1R12B, MYLK, MYL9, MYH14, MYH11* were up-regulated. The five genes whose expression was up-regulated constituted in one of the clusters together, the down-regulated genes are all located in the other cluster, and the two clusters are connected by the MYLK gene. Furthermore, compared with the expert consensus results, we found that MYLK and MYH11 are located in the A2 group. The expression network composed of these genes with their first neighbors are shown in the Fig. 4.

Discussion

We analyzed a total of 11 original samples with SRA format, including three normal aorta tissues and eight ATAA tissues. Through sample processing, differential gene screening, enrichment analysis, and PPI network construction, we finally screened the following hub genes: *CDC20, CCNB2, BUB1B, BIRC5, PTTG1, ESPL1* genes as down-regulated hub genes, *PPP1R12B, MYLK, MYL9, MYH14, MYH11* were up-regulated hub genes identified, among which *MYLK* and *MYH11* genes are strongly related genes in the consensus

of experts that these genes have a clear and important role of smooth muscle cell functions and phenotypic transformation during aneurysm formation[25–27].

We attempted to revert single-cell sequencing to whole-exome sequencing in tissues, which can complement single-cell sequencing results. To the best of our knowledge, it is much more difficult for a developed targeted drug to target a certain type of cell-expressing product in a tissue than for any cell that expresses the product. Therefore, by comparing the statistical differences, a negative result is obtained. After the results, the obtained differential genes are the differential expression of the whole tissue. This is of great interest for complementing single-cell sequencing information and developing low-cost targeted drugs.

In a general sense, many scholars use the MCODE score to screen hub genes, which means that they are scored according to the degree of connection of each gene. The higher degree of connection the more important of the genes, but we consider the cascading effect of gene pathways and vulnerability, which means that gene pathways are easily affected by many factors. Besides, these samples are all from older people (the details are shown in literature[28]), and this is consistent with the distribution of ATAA, so it is inevitable that we cannot rule out immunity response caused by other reasons, that's why we exclude immune-related genes first. Afterward, all the genes of the remaining clusters were included in the KEGG pathway to find the pathways related to ATAA. At the same time, the genes in the clusters and in the KEGG pathway are considered hub genes.

We also use the k-means cluster algorithm[29] to classify all 1930 DEGs. The number of clusters ranges from 3 to 13. It is found that the positions of the MYLK and MYH11 genes are in the central region of all gene classifications, and as the number of clusters increases, they move back and forth at the edge of each cluster approaching the center. That is to say, these two genes are nodes that are difficult to classify in machine learning. Such nodes generally have some of the characteristics of multiple clusters. It further illustrates the high correlation between the two genes MYLK and MYH11 and other clusters and plays a key role in the whole gene network. At the same time, in the Cytoscape software, we have successively selected MYLK and MYH11 genes to observe their first neighbors with their first neighbors' connections and found that most of the first neighbors of these two genes can light up almost all nodes of the cluster where the first neighbors are located. This may explain some problems to a certain extent. This reminds us of the characteristics of small-world networks[30]. For example, the nodes in the cluster are highly connected to each other. However, the nodes between different clusters do not have such a high degree of connection but are transmitted information to each other through several highly connected nodes[31].

Our statistical difference comparison is based on the proportion of each cell. If the weight of the expressed product in each cell can be considered, the theoretical result will be clearer, but due to the limitation of our own computer computing power, there is no way to complete this process.

Conclusion

Overall, by using mRNA single-cell sequencing gene arrays, we pooled the sequencing results according to the exome sequencing method with no statistically significant conditions. Finally, we got *CDC20, CCNB2, BUB1B, BIRC5, PTTG1, ESPL1* genes as down-regulated hub genes, *PPP1R12B, MYLK, MYL9, MYH14, MYH11* were up-regulated hub genes associated with ATAA. This not only proves the important role of the two genes *MYLK* and *MYH11*, but also proves the judgment results of these two genes in the expert consensus. More importantly, to a certain extent, we elucidate the molecular pathways of ATAA pathogenesis from a holistic perspective, which can facilitate the prediction and diagnosis of ATAA. Of course, future experiments of targeting biology at the protein level are required to further clarify mechanism of ATAA.

Declarations

Funding information

Not applicable.

Data availability statement

The data that support the findings of this study, all the KEGG and GO pathways are available in the supplementary material of this article.

Author contributions

GTX, YY and DZ designed the study; SEY, YY and DZ collected the data and carried out the computation and generated figures; YY and DZ analyzed the data and wrote the manuscript.

Ethics declarations

Ethics approval and consent to participate

All methods were performed in accordance with the relevant guidelines and regulations.

Consent for publication

Not applicable.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

We appreciate Dr. Gaoxing Zheng from Zhongshan First Hospital Affiliated to Fudan University for his time and advice to debugging our code.

References

1. Masuda, Y., Takanashi, K., Takasu, J., Morooka, N., & Inagaki, Y. (1992). Expansion rate of thoracic aortic aneurysms and influencing factors. Chest, 102(2), 461–466. https://doi.org/10.1378/chest.102.2.461
2. Pham, T., Martin, C., Elefteriades, J., & Sun, W. (2013). Biomechanical characterization of ascending aortic aneurysm with concomitant bicuspid aortic valve and bovine aortic arch. Acta biomaterialia, 9(8), 7927–7936. https://doi.org/10.1016/j.actbio.2013.04.021
3. Elefteriades, J. A., & Farkas, E. A. (2010). Thoracic aortic aneurysm clinically pertinent controversies and uncertainties. Journal of the American College of Cardiology, 55(9), 841–857. https://doi.org/10.1016/j.jacc.2009.08.084
4. Isselbacher, E. M., Lino Cardenas, C. L., & Lindsay, M. E. (2016). Hereditary Influence in Thoracic Aortic Aneurysm and Dissection. Circulation, 133(24), 2516–2528. https://doi.org/10.1161/CIRCULATIONAHA.116.009762
5. Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics (Oxford, England), 26(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616
6. Anders, Simon, and Wolfgang Huber. 2012. 'Differential expression of RNA-Seq data at the gene level–the DESeq package', Heidelberg, Germany: European Molecular Biology Laboratory (EMBL), 10: f1000research.
7. Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. Genome biology, 11(2), R14. https://doi.org/10.1186/gb-2010-11-2-r14
8. Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research, 28(1), 27–30. https://doi.org/10.1093/nar/28.1.27
9. Renard, M., Francis, C., Ghosh, R., Scott, A. F., Witmer, P. D., Adès, L. C., Andelfinger, G. U., Arnaud, P., Boileau, C., Callewaert, B. L., Guo, D., Hanna, N., Lindsay, M. E., Morisaki, H., Morisaki, T., Pachter, N., Robert, L., Van Laer, L., Dietz, H. C., Loeys, B. L., … De Backer, J. (2018). Clinical Validity of Genes for Heritable Thoracic Aortic Aneurysm and Dissection. Journal of the American College of Cardiology, 72(6), 605–615. https://doi.org/10.1016/j.jacc.2018.04.089
10. Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S., & Soboleva, A. (2013). NCBI GEO: archive for functional genomics data sets–update. Nucleic acids research, 41(Database issue), D991–D995. https://doi.org/10.1093/nar/gks1193
11. Leinonen, R., Sugawara, H., Shumway, M., & International Nucleotide Sequence Database Collaboration (2011). The sequence read archive. Nucleic acids research, 39(Database issue), D19–D21. https://doi.org/10.1093/nar/gkq1019
12. Allison, Paul D. 2009. Fixed effects regression models (SAGE publications).

13. Wood, Simon N. 2011. 'Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models', Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73: 3-36.
14. Doi, S. A., Barendregt, J. J., Khan, S., Thalib, L., & Williams, G. M. (2015). Advances in the meta-analysis of heterogeneous clinical trials I: The inverse variance heterogeneity model. Contemporary clinical trials, 45(Pt A), 130–138. https://doi.org/10.1016/j.cct.2015.05.009
15. Borenstein, M., Higgins, J. P., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. Research synthesis methods, 8(1), 5–18. https://doi.org/10.1002/jrsm.1230
16. Zaykin D. V. (2011). Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. Journal of evolutionary biology, 24(8), 1836–1841. https://doi.org/10.1111/j.1420-9101.2011.02297.x
17. Krueger, Felix. 2015. 'Trim galore', A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, 516: 517.
18. Brown, J., Pirrung, M., & McCue, L. A. (2017). FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. Bioinformatics (Oxford, England), 33(19), 3137–3139. https://doi.org/10.1093/bioinformatics/btx373
19. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997.
20. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England), 25(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352
21. Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics (Oxford, England), 30(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656
22. Huang, d., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols, 4(1), 44–57. https://doi.org/10.1038/nprot.2008.211
23. von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., & Snel, B. (2003). STRING: a database of predicted functional associations between proteins. Nucleic acids research, 31(1), 258–261. https://doi.org/10.1093/nar/gkg034
24. Bader, G. D., & Hogue, C. W. (2003). An automated method for finding molecular complexes in large protein interaction networks. BMC bioinformatics, 4, 2. https://doi.org/10.1186/1471-2105-4-2
25. Humphrey, J. D., Schwartz, M. A., Tellides, G., & Milewicz, D. M. (2015). Role of mechanotransduction in vascular biology: focus on thoracic aortic aneurysms and dissections. Circulation research, 116(8), 1448–1461. https://doi.org/10.1161/CIRCRESAHA.114.304936

26. Milewicz, D. M., Trybus, K. M., Guo, D. C., Sweeney, H. L., Regalado, E., Kamm, K., & Stull, J. T. (2017). Altered Smooth Muscle Cell Force Generation as a Driver of Thoracic Aortic Aneurysms and Dissections. Arteriosclerosis, thrombosis, and vascular biology, 37(1), 26–34. https://doi.org/10.1161/ATVBAHA.116.303229
27. Michel, J. B., Jondeau, G., & Milewicz, D. M. (2018). From genetics to response to injury: vascular smooth muscle cells in aneurysms and dissections of the ascending aorta. Cardiovascular research, 114(4), 578–589. https://doi.org/10.1093/cvr/cvy006
28. Li, Y., Ren, P., Dawson, A., Vasquez, H. G., Ageedi, W., Zhang, C., Luo, W., Chen, R., Li, Y., Kim, S., Lu, H. S., Cassis, L. A., Coselli, J. S., Daugherty, A., Shen, Y. H., & LeMaire, S. A. (2020). Single-Cell Transcriptome Analysis Reveals Dynamic Cell Populations and Differential Gene Expression Patterns in Control and Aneurysmal Human Aortic Tissue. Circulation, 142(14), 1374–1388. https://doi.org/10.1161/CIRCULATIONAHA.120.046528
29. Na, Shi, Liu Xumin, and Guan Yong. 2010. "Research on k-means clustering algorithm: An improved k-means clustering algorithm." In 2010 Third International Symposium on intelligent information technology and security informatics, 63-67. Ieee.
30. Barrat, Alain, and Martin Weigt. 2000. 'On the properties of small-world network models', The European Physical Journal B-Condensed Matter and Complex Systems, 13: 547-60.
31. van Noort, V., Snel, B., & Huynen, M. A. (2004). The yeast coexpression network has a small-world, scale-free architecture and can be explained by a simple model. EMBO reports, 5(3), 280–284. https://doi.org/10.1038/sj.embor.7400090

Figures

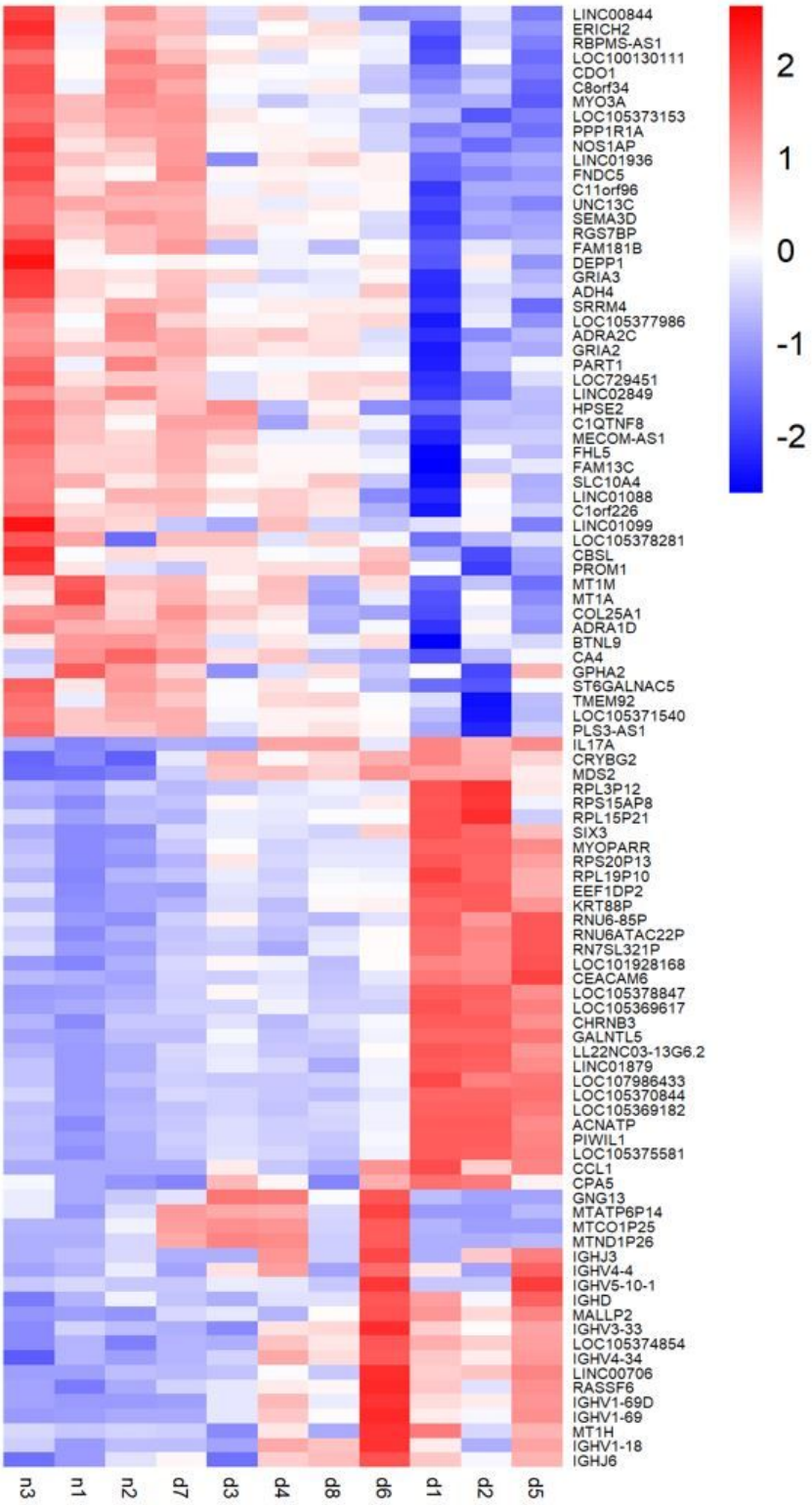

Figure 1

Top 50 upregulated and downregulated DEGs.

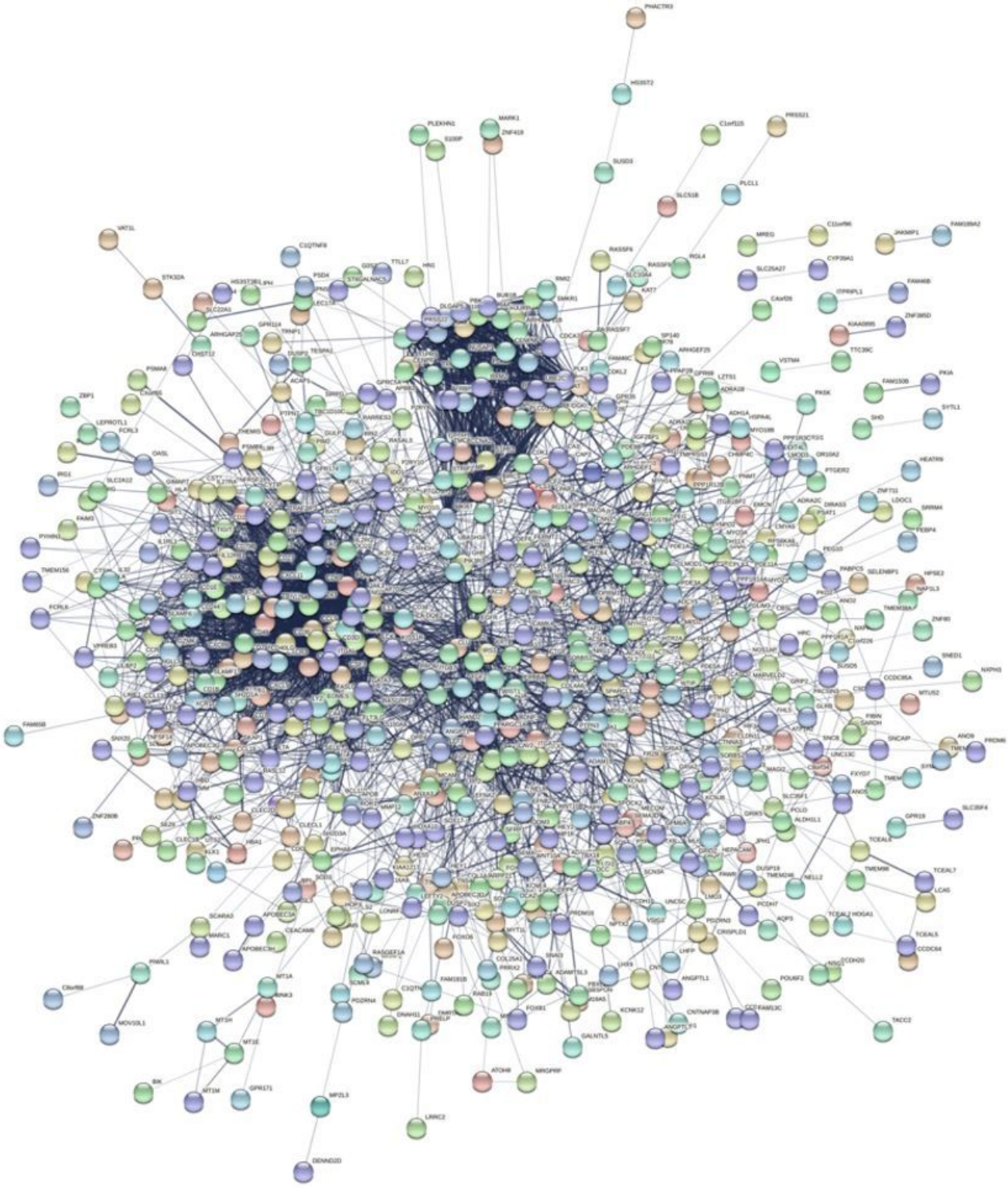

Figure 2

PPI network

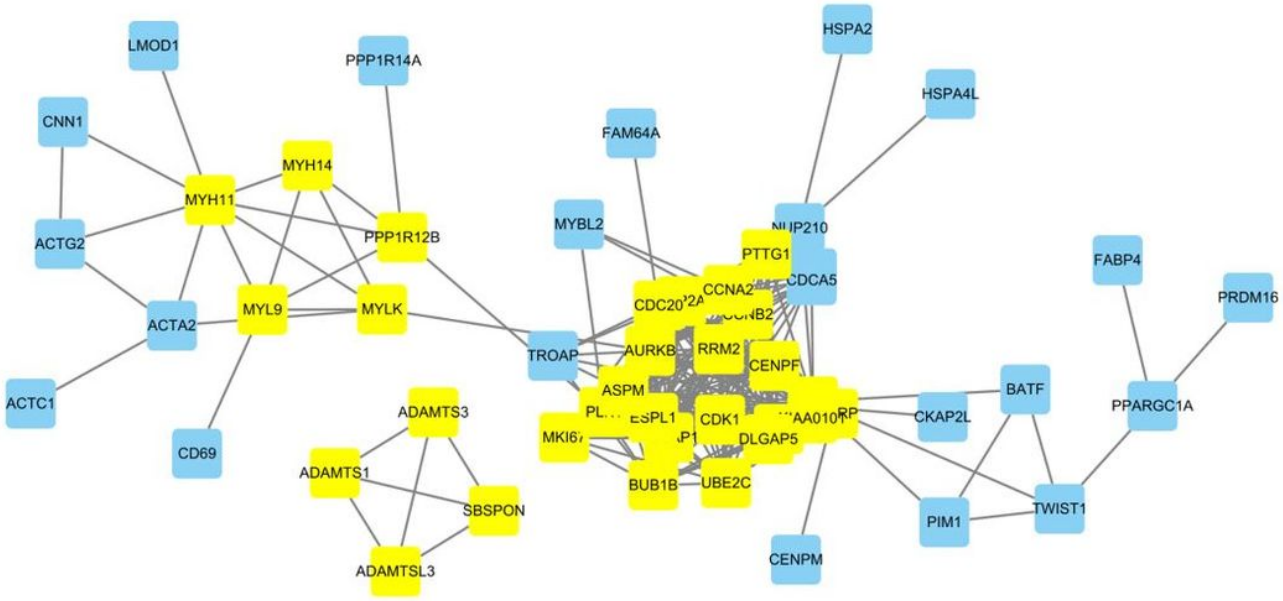

Figure 3

Clusters related to ATAA.

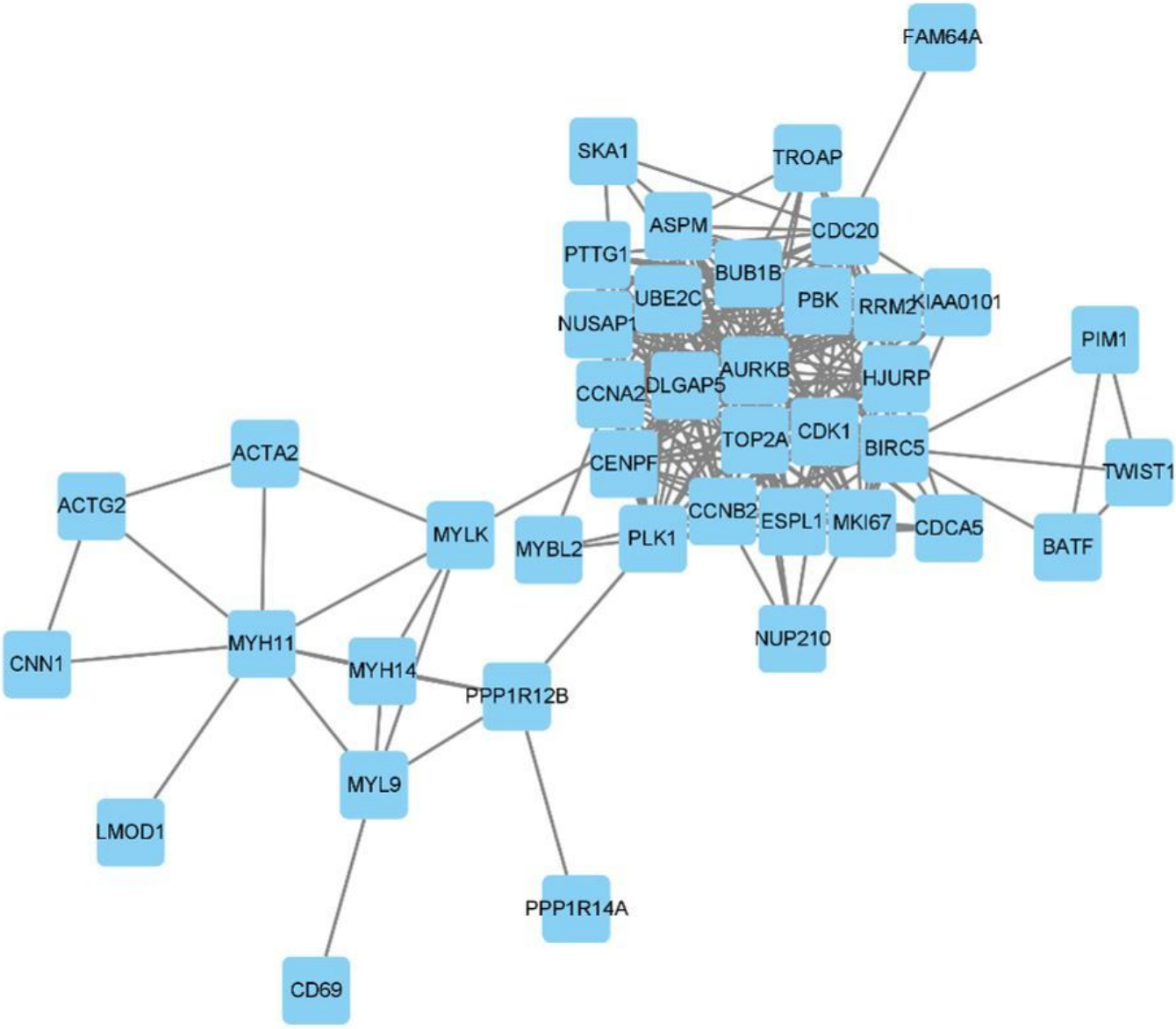

Figure 4

The expression network composed of hub genes with their first neighbors.

Supplementary Files

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

- supplementarymaterial.zip