Int J Clin Exp Med 2018;11(8):8054-8064
www.ijcem.com /ISSN:1940-5901/IJCEM0067893

Original Article

Prediction of key genes and metabolic biomarkers in ankylosing spondylitis based on a multi-omics composite network

Ting Wang[1], Xuesong Wang[2], Lei Peng[3], Yanan Cong[1], Chunyan Du[1]

[1]*Department of Spine, The Affiliated Hospital of Qingdao University, Shandong, P.R. China; Departments of* [2]*Spine,* [3]*Library, No. 2 Affiliated Hospital of Qingdao University, Shandong, P.R. China*

Received October 25, 2017; Accepted March 29, 2018; Epub August 15, 2018; Published August 30, 2018

Abstract: In this current study, we screened optimal key genes and metabolic biomarkers for ankylosing spondylitis (AS) based on differentially expressed genes (DEGs) network and multi-omics composite network. The OMIM (https://omim.org/) ID of AS was 106300. DEGs were identified from a total of 32 samples that were divided into two groups, the first group (16 samples, whole blood, control) and second group (16 samples, whole blood, affected). Known disease metabolites were extracted from human metabolite database (HMDB). Through integration of six data sets to build a multi-group composite network, it can be expressed as six networks. A multi-omics composite network was then constructed using the above six network data. In order to obtain preferred candidate metabolites in the complex network, random walk with restart (RWR) method was used to extend screening to a multithreaded composite network. A total of 579 DEGs, associated with AS, were identified. Fifty AS metabolites were screened. In summary, our analysis based on a network multi-omics composite not only identified 579 DEGs in AS but also importantly discovered 50 optimal disease metabolites related to AS. Our study may provide important potential therapeutic targets for AS.

Keywords: Ankylosing spondylitis, disease metabolites, multi-omics, differentially expressed genes, random walk with restart

Introduction

Ankylosing spondylitis (AS), a chronic, systemic, inflammatory, and rheumatic disease, affects axial bones and sacroiliac joints, causing characteristic inflammatory dorsal pain and leading to varying degrees of structural dysfunction [1]. Studies have reported that AS affects about 0.1% of the general population and is more common in men than women [2]. Moreover, AS may affect surrounding joints, skin, eyes, and intestines and increase the risk of cardiovascular or lung morbidity [3]. AS is difficult to diagnose and patients can endure symptoms for many years before receiving appropriate treatment [4]. Treatment for AS is mainly through use of drugs such as disease repair antirheumatic drugs (DMARDs) and non-steroidal anti-inflammatory drugs (NSAIDs). There are also non-drug interventions such as physical therapy [5]. Although these interventions may have some effect on spinal pain, peripheral pain, and physical function, they do not show a change in progression of AS. Despite the use of various methods, treatment of AS has not been ideal. Therefore, it is necessary to further study the mechanism of AS in an attempt to find a new treatment.

To fill these gaps, better understand the pathological process of AS metabolism, and improve clinical diagnosis of human medicine status, identification of related metabolites is very important. Metabolites are the end product of the cell regulation process. They are generally considered to be the ultimate response of a biological system to genetic or environmental changes [6]. Metabolite levels can directly reflect the physiological state of the human body [7]. Identification of disease-related metabolites is important, not only for a better understanding of metabolic pathological pro-

cesses, but also for improving clinical diagnosis [8]. With development of metabolomics technology, hundreds of metabolites can be identified by liquid chromatography, gas chromatography-mass spectrometry (GC-MS), nuclear magnetic resonance (NMR), and so forth [9]. Identifying and determining the priority of high-risk disease metabolites has become a challenging task. Some new high-risk diseases (such as AS) lack known metabolite information.

In this work, we proposed a method based on comprehensive multi-omics information to predict and optimize AS candidate metabolites. Comprehensive multithreading information consisted of genomes, phenotypes, metabolites, and interactions.

Materials and methods

Obtaining expression data and disease information

In the present study, the OMIM (https://omim.org/) ID of AS was 106300. Combined with information entered and data collected by this platform, the disease-related seed genes were HLA-B and PHEX but related seed metabolites were not inquired.

A total of 32 samples were divided into two groups, the first group (16 samples, whole blood, control) and second group (16 samples, whole blood, affected). Each group was then analyzed. After pretreatment, expression profiles of 11,586 genes were obtained by mapping between the probe and gene. Differentially expressed genes were calculated using the Limma package. Student's t-test and F-test were performed by gene expression matrix. We used the lmfit function to linearly fit the data and eBayes statistics and false discovery rate (FDR) to correct raw *p* values. We sorted by *p* value and the top 5% were selected.

Getting comprehensive multi-omics information

Known disease metabolites were extracted from human metabolite database (HMDB). HMDB collects detailed information on human small molecule metabolites and disease phenotype information described in the OMIM entry. Known disease gene information was obtained from the morbid map file of OMIM database, containing a comprehensive description of human genes and phenotypes and their relationships.

Through integration of six data sets to build a multi-group composite network, it can be expressed as six networks which were i) a genetic network, ii) a metabolic network, iii) a phenotypic network, iv) a gene-metabolite action network, v) a phenotype-gene action network, and vi) a phenotype-metabolic action network.

i) Gene Network (A_G). We obtained data from the human protein interaction network (containing 1,048,576 interactions) from STRING. Protein ID was transformed with gene name. We then removed repeated interactions and a PPI gene network containing 16,785 nodes and 1,515,370 pairs of interactions was obtained.

ii) Metabolic network (A_M). First, the metabolic pathways were collected from KEGG and HMDB. Human metabolites pathways were collected from Reactome, MSEA, and SMPDB. Then, we collected 4,994 human metabolites. Next, we collected human metabolites and interrelationships between human metabolites from STITCH, which must be included in 4,994 human metabolites. Eventually, we obtained 3,764 human metabolites and 74,667 interactions with human metabolites (not all metabolites are associated in STITCH).

iii) Phenotype network (A_P). Using a phenotypic network from van Driel, it stores 5,080 phenotypes and their interactions [10].

iv) Gene-metabolite network (A_{GM}). Extracted chemical substances, human genes, and STITCH-related information, according to 4,994 kinds of human metabolites. We filtered out metabolites not in metabolic network and genes not in gene network. 12,342 genes and 3,278 metabolites and 192,763 pairs of gene-metabolites interactions were obtained.

v) Phenotype-gene Interaction Network (A_{GP}). Phenotypic-gene association was obtained from planned morbid map file of OMIM. Filtering phenotypes not in our phenotypic network and genes not in our gene network, we obtained 1,886 phenotypes, 1,715 genes, and 2,603 interactions.

vi) Phenotypic-metabolic action network (A_{PM}). Phenotypic-metabolite association was obtained from HMDB. Similarly, 149 phenotypes, 388 metabolites, and 664 pairs of interactions were retained after filtration.

Construction of multi-omics composite network

In order to construct a multi-omics composite network, the above six network data were integrated into composite network A.

$$A = \begin{bmatrix} A_G & A_{GP} & A_{GM} \\ A_{GP}^T & A_P & A_{PM} \\ A_{GM}^T & A_{PM}^T & A_M \end{bmatrix}$$

W was the transition matrix of composite network A, which was a column normalized adjacency matrix and can be inferred from adjacency matrix A.

$$W = \begin{vmatrix} W_G & W_{GP} & W_{GM} \\ W_{PG} & W_P & W_{PM} \\ W_{MG} & W_{MP} & W_M \end{vmatrix}$$

W_{ij} represents the probability of jump from node i to node j. x, y, and z are the probabilities of jump between gene network and phenotype network, gene network and metabolic network, and phenotypic network and metabolic network. The default values were 1/3. The probability of gene i (g_i) to gene j (g_j) in the genetic network can be defined as

$$W_G(i,j) = P(g_j \mid g_i) = \begin{cases} (1-x-y)A_G(i,j) \Big/ \sum_j A_G(i,j), & \text{if } \sum_j A_{GP}(i,j) \neq 0 \text{ and } \sum_j A_{GM}(i,j) \neq 0 \\ (1-x)A_G(i,j) \Big/ \sum_j A_G(i,j), & \text{if } \sum_j A_{GP}(i,j) \neq 0 \text{ and } \sum_j A_{GM}(i,j) = 0 \\ (1-y)A_G(i,j) \Big/ \sum_j A_G(i,j), & \text{if } \sum_j A_{GP}(i,j) = 0 \text{ and } \sum_j A_{GM}(i,j) \neq 0 \\ A_G(i,j) \Big/ \sum_j A_G(i,j), & \text{if } \sum_j A_{GP}(i,j) = 0 \text{ and } \sum_j A_{GM}(i,j) = 0 \end{cases}$$

Similarly, the probability from gene i (g_i) to phenotype j (p_j) was defined as

$$W_{GP}(i,j) = P(p_j \mid g_i) = \begin{cases} xA_{GP}(i,j) \Big/ \sum_j A_{GP}(i,j), & \text{if} \sum_j A_{GP}(i,j) \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

The probability from gene i (g_i) to metabolite j (m_j) was defined as

$$W_{GM}(i,j) = P(m_j \mid g_i) = \begin{cases} yA_{GM}(i,j) \Big/ \sum_j A_{GM}(i,j), & \text{if} \sum_j A_{GM}(i,j) \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

In the phenotypic network, the probability from phenotype i (p_i) to phenotype j (p_j) was defined as

$$W_P(i,j) = P(pj \mid pi) = \begin{cases} (1-x-z)A_P(i,j) \Big/ \sum_j A_P(i,j), & \text{if} \sum_j A_{PM}(i,j) \neq 0 \text{ and } \sum_j A_{GP}(j,i) \neq 0 \\ (1-z)A_P(i,j) \Big/ \sum_j A_P(i,j), & \text{if} \sum_j A_{PM}(i,j) \neq 0 \text{ and } \sum_j A_{GP}(j,i) = 0 \\ (1-x)A_P(i,j) \Big/ \sum_j A_P(i,j), & \text{if} \sum_j A_{PM}(i,j) = 0 \text{ and } \sum_j A_{GP}(j,i) \neq 0 \\ A_P(i,j) \Big/ \sum_j A_P(i,j), & \text{if} \sum_j A_{PM}(i,j) = 0 \text{ and } \sum_j A_{GP}(j,i) = 0 \end{cases}$$

The probability from phenotype i (p_i) to gene j (g_j) was defined as

$$W_{PG}(i,j) = P(g_j \mid p_i) = \begin{cases} xA_{GP}(j,i) \Big/ \sum_j A_{GP}(j,i), & \text{if} \sum_j A_{GP}(j,i) \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

The probability from phenotypic i (p_i) to metabolite j (m_j) was defined as

$$W_{PM}(i,j) = P(m_j \mid p_i) = \begin{cases} zA_{PM}(i,j) \Big/ \sum_j A_{PM}(i,j), & \text{if} \sum_j A_{PM}(i,j) \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

In the metabolite network, the probability from metabolite i (m_i) to metabolite j (m_j) was defined as

$$W_M(i,j) = P(m_j \mid m_i) = \begin{cases} (1-y-z)A_M(i,j) \Big/ \sum_j A_M(i,j), & \text{if} \sum_j A_{GM}(j,i) \neq 0 \text{ and } \sum_j A_{PM}(j,i) \neq 0 \\ (1-y)A_M(i,j) \Big/ \sum_j A_M(i,j), & \text{if} \sum_j A_{GM}(j,i) \neq 0 \text{ and } \sum_j A_{PM}(j,i) = 0 \\ (1-z)A_M(i,j) \Big/ \sum_j A_M(i,j), & \text{if} \sum_j A_{GM}(j,i) = 0 \text{ and } \sum_j A_{PM}(j,i) \neq 0 \\ A_M(i,j) \Big/ \sum_j A_M(i,j), & \text{if} \sum_j A_{GM}(j,i) = 0 \text{ and } \sum_j A_{PM}(j,i) = 0 \end{cases}$$

The probability from metabolite i (m_i) to gene j (g_j) was defined as

$$W_{MG}(i,j) = P(g_j \mid m_i) = \begin{cases} yA_{GM}(j,i) \Big/ \sum_j A_{GM}(j,i), & \text{if} \sum_j A_{GM}(j,i) \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

The probability from metabolite i (m_i) to phenotype j (p_j) was defined as

$$W_{MP}(i,j) = P(p_j \mid m_i) = \begin{cases} zA_{PM}(j,i) \Big/ \sum_j A_{PM}(j,i), & \text{if} \sum_j A_{PM}(j,i) \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

Random walk with restart (RWR)

In order to obtain preferred candidate metabolites in the complex network, RWR method [11] was used to extend screening to a multithreaded composite network. This method selects preferred candidate metabolites based on the proximity of each candidate to seed candidate metabolite (the known metabolite) and simulates a random walk process from the seed node. Each step of the walk moves from the current node to its immediate neighbor at probability 1-α, or returns to seed node at probability α. The calculated formula was as follows:

$$P^{k+1} = (1-\alpha)\, WP^k + \alpha P^0$$

P^0 was the initial probability vector. P^k represents the probability vector remains at node i where the i-th element at step k.

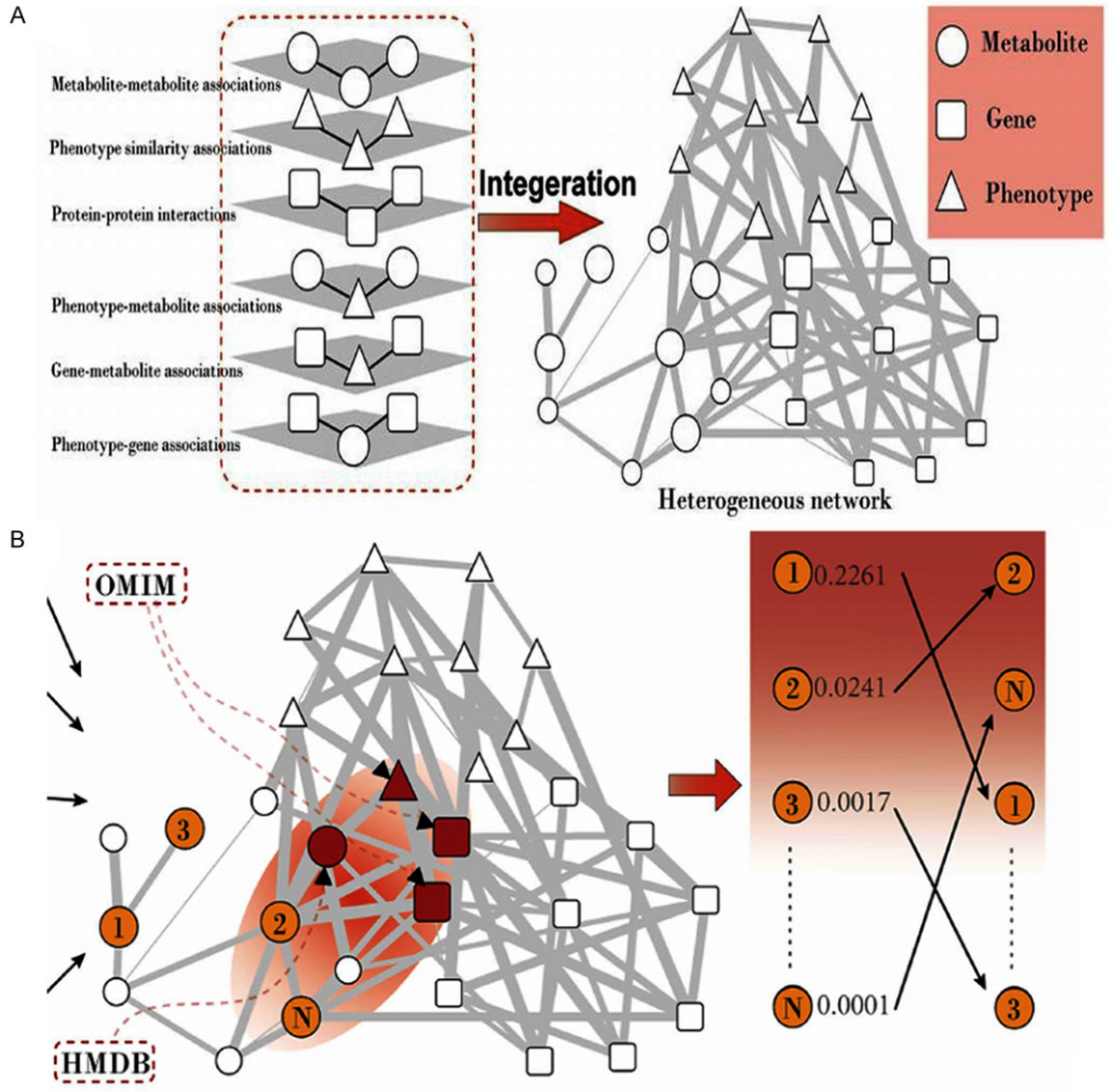

Figure 1. Flow chart of MetPriCNet. A. Construction of the multi-omics composite network. The multi-omics composite network is composed of six sub-networks. White circles indicate metabolites, white squares indicate genes, and white triangles indicate phenotypes. The thickness of an edge indicates the weight score. B. Flow chart of MetPriCNet to optimize the candidate metabolite. First, the interested candidate metabolites and seed nodes were mapped to the multi-omics composite network. Then, a global extended RWR method was used to score the candidate metabolites according to their proximity to seed nodes. Finally, the candidate metabolites were ranked by scores. Orange circles represent candidate metabolites of interest. Red triangles indicate the disease phenotype of interest (phenotype seed) from the OMIM data base, red squares represent known disease genes (gene seeds) from the OMIM database, and red circles indicate known disease metabolites (metabolite seeds) from the HMDB database.

(1) The initial probability vector P^0.

u_0, v_0, and w_0 were supposed to be the initial probabilities of the genetic network, phenotype network, and metabolic network, respectively. For a phenotype (ie disease), the seed nodes consisted of i) phenotypes, ii) corresponding known metabolites, and iii) known genes. Initial probability of the genetic network u_0 was calculated by giving an equal probability to gene nodes in the gene network. The sum was equal to 1, meaning that the random walker starts at the same probability from each seed node. Similarly, initial probabilities v_0 and w_0 were calculated, among them a = 1/3, b = 1/3.

Table 1. Statistical information of the composite network

Statistics of the composite Network	Node	Edge
Gene network	516	3325
Metabolite network	3764	74667
Phenotype network	5080	10140046
Gene-metabolite association network	516 genes, 3764 metabolites	5966
Phenotype-gene association network	5080 phenotypes, 516 genes	73
Phenotype-metabolite association network	149 phenotypes, 388 metabolites	664
All	9360	10224741

Table 2. List of top 10 differentially expressed genes (DEGs)

Genes	|log (fold change)|	False discovery rate (FDR)
PTPN1	0.541708409	0.006354508
LAMTOR2	0.456185459	0.006354508
IL27RA	0.479655086	0.006354508
XPC	0.499038398	0.006354508
FAM222B	0.350544393	0.006354508
TCEAL9	0.317543	0.006354508
AAMDC	0.290132672	0.006354508
MRI1	0.391011129	0.007153737
HOXB1	0.306503564	0.007508667
DGKQ	0.681929291	0.007921708

Table 3. List of top 10 preferred metabolites

Rank	Metabolite CID	Metabolite Name	Score
1	1567	2-mercaptoethanol	0.006739923
2	888	Magnesium ion	0.006243695
3	123762	Aerobactin	0.003147908
4	5283588	SM (d18:1/18:0)	0.002561754
5	23964	TUNGSTEN	0.001647578
6	439503	Salicin	0.001243961
7	5280453	Rocaltrol (TN)	0.000934286
8	5280795	Colecalciferol (INN)	0.000805469
9	2526	Beta-Glycerophosphoric acid	0.000755857
10	33032	Gulutamine (USP)	0.000709295

$$P^0 = \begin{vmatrix} a * \mu 0 \\ b * v0 \\ (1 - a - b) * \omega 0 \end{vmatrix}$$

(2) k-step probability vector P^k.

After multiple iterations, the change between P^{k+1} and P^k was less than 10^{-10}. At this moment, the probability reached a steady state and iteration stopped.

Identifying preferred metabolites and their co-expression genes

After the random walk to steady state, each metabolite in the complex network had a probability corresponding to it. After removal of seed metabolite nodes and sequencing the probability, the top 50 metabolites were screened out as preferred metabolites. We identified genes that interacted with preferred metabolites and analyzed their scores. The top 100 genes were found, with scores greater than the mean. These were recognized as co-expression genes.

Results

MetPriCNet

The general idea of MetPriCNet is depicted in **Figure 1**. First, we constructed a composite network which integrated omics data including the genome, phenome, metabolome, and interactome. Then, we proposed a global computational method to capture interaction information between the multi-omics composite network to prioritize candidate metabolites according to their proximity with known disease seed nodes. Details are given below. Statistical information of the composite network is shown in **Table 1**.

Candidate differential genes

First, we identified differentially expressed genes (DEGs) between AS and normal samples using t-test and F-test. Based on FDR < 0.05 and |log FC| ≥ 0.05, a total of 579 candidate genes were screened. We then built a genetic network. The top 10 DEGs are shown

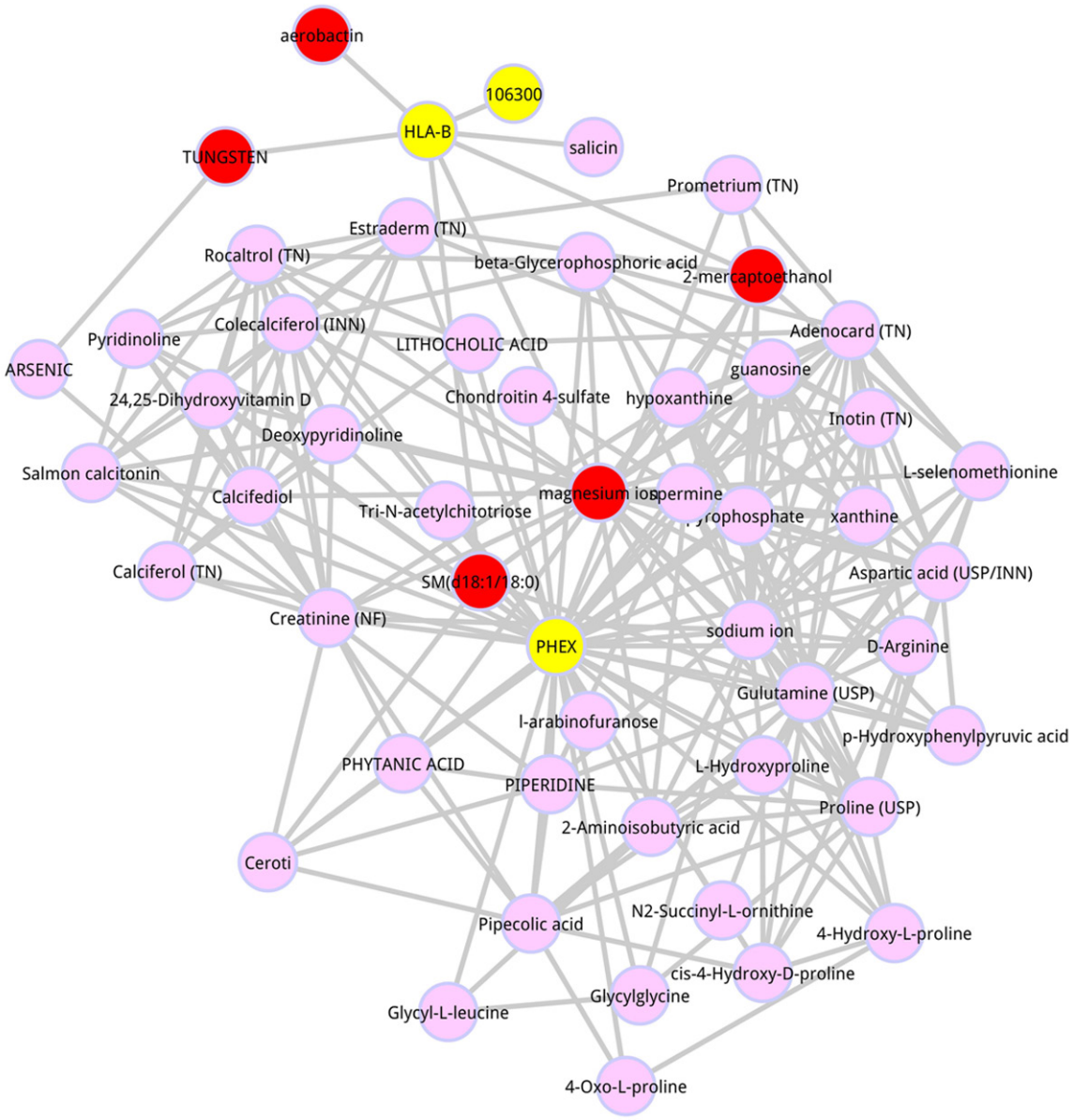

Figure 2. AS preferred metabolites complex network. The pink node represents the metabolite, red node represents the top 5 metabolites, and the yellow represents the seed node.

in **Table 2**. The most significant DEGs were PTPN1, LAMTOR2, IL27RA, XPC, and FAM22-2B.

Disease preferred metabolites complex network construction

We used RWR to screen 50 disease metabolites of AS. The top 10 are shown in **Table 3**. The most significant disease metabolites were 2-mercaptoethanol, magnesium ion, and aerobactin.

In order to further probe the biological activity of disease metabolites of AS, we built a network using the above 50 metabolites. In topological analysis of the constructed network, distribution of network nodes degree was the measure of connectivity. Assortativity examines whether nodes with similar values tend to connect with each other. Assortativity coefficient is a kind of "degree" based on Pearson's correlation coefficient, used to measure relationship between connected nodes. If the result of assortativity coefficient is positive, it repre-

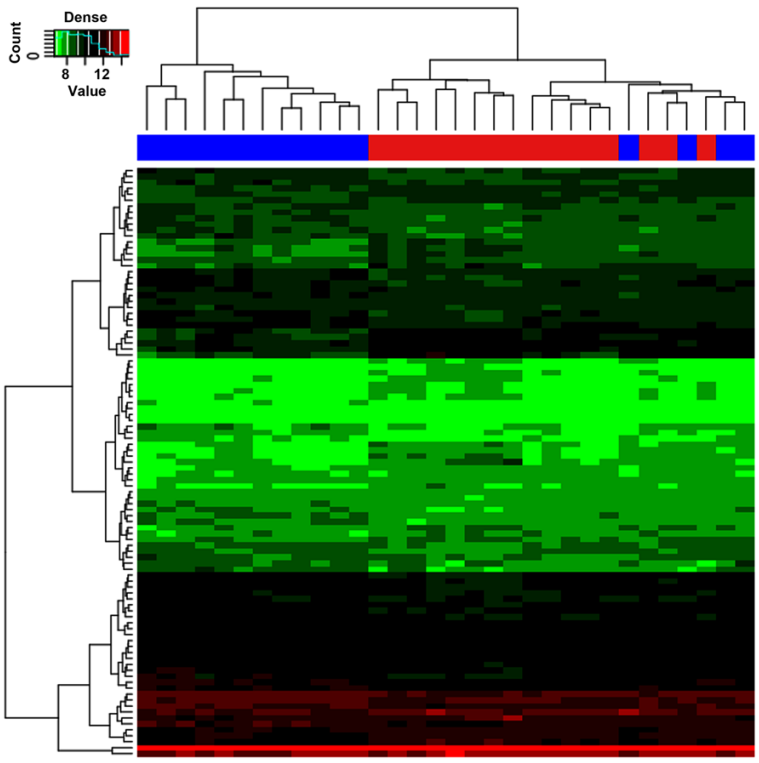

Figure 3. Heat map of AS candidate metabolite co-expression genes. Two groups of patients (columns) of the disease preferred metabolites co-expression gene (line) expression. Blue represents: Whole blood, control; Red Represents: Whole blood, affected.

sents some kind of synergy between the points with the same degree. Negative values indicate that there is some connection between the nodes with different degrees. Connection between metabolites is represented by the corresponding entry in the matrix. In general, value of the assortativity coefficient is between -1 and +1. +1 indicates that the network has a good homology, 0 indicates the network is non-assortative, and -1 indicates the network is negatively correlated (**Figure 2**). If there is a connection between nodes, value of the edge is 1. Otherwise, it is 0. The closer the value is to 0 means that the connection is weaker and the closer it is to 1 indicates a stronger connection.

Clustering and AS preferred metabolites analyses

Results of 2D clustering analyses of 100 co-expression genes: As shown in **Figure 3**, control and affected patients can be largely separated into two clusters. We obtained top 50 AS preferred metabolites and top 100 co-expression genes. A matrix was obtained that used co-expression genes as row name and AS patients as column name. Relationship between genes and metabolites was represented by the corresponding entry in 2D clustering.

Disease candidate metabolites co-expression genes network construction

To further illustrate the intrinsic mode of MetPriCNet, we dissected candidate metabolites. A network was constructed by all metabolites (**Figure 4**). In this network, it is easily noticed that there are two types of nodes among the neighbors, including metabolite nodes and gene nodes. This consists with the intrinsic mechanism of MetPriCNet, which can account for interaction information of every type of node to consider similarity between candidate metabolites and seeds under the global composite network. We further inspected direct and indirect interaction (we only considered two-step neighbors) between the five top-ranked metabolites and seed nodes. The top-ranked metabolite is expected to have more and stronger direct and indirect interactions with seeds. This seems inconsistent with our methods. However, MetPriCNet is based on a global distance measure, which considers not only strong direct interactions but also indirect and weak interactions.

Discussion

Many genetic association studies have identified genes that contribute to AS susceptibility but these methods have provided little information on changes in gene activity that occur in the process of AS. Transcriptional spectroscopy results in a snapshot of sampled cell activity, thus providing an understanding of molecular processes that drive the disease process. In this work, we used multi-omics and genome-

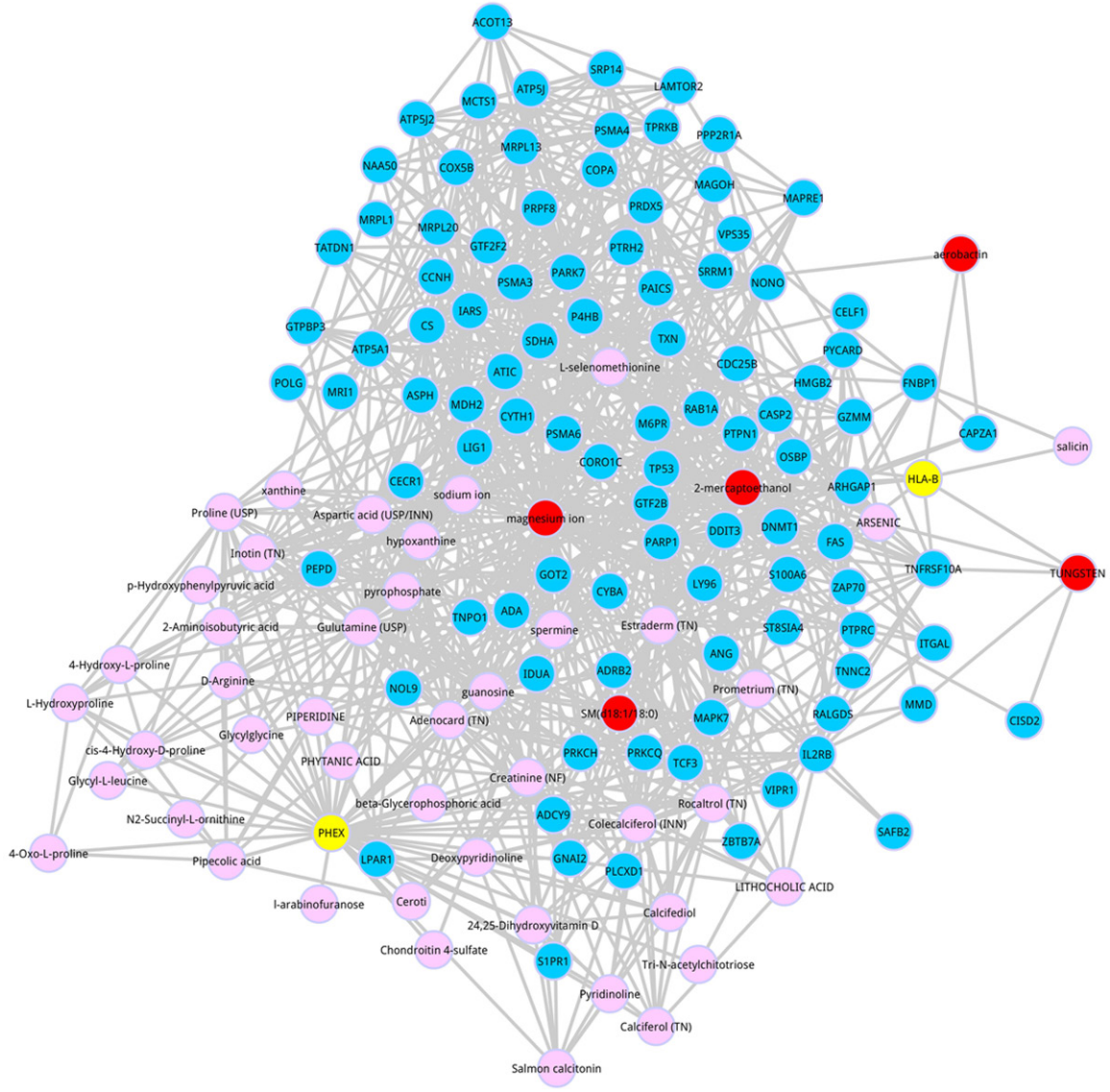

Figure 4. AS candidate metabolite co-expression gene network. The blue node represents the gene, pink node represents the metabolite, and yellow represents the seed node.

wide microarray to identify AS-related candidate genes and predict disease metabolites of AS.

Multi-omics compound networks plays a pivotal role in many important biological processes but research on functions of multi-omics compound networks in human disease is still in its infancy [8]. Therefore, it is urgent to prioritize multi-omics compound networks that are potentially associated with diseases [12]. In this work, we developed a novel algorithm, RWR, which used a multi-level composite network to prioritize candidates of the multi-omics compound network associated with diseases. By integrating genes, multi-omics, phenotypes and their associations, RWR achieves an overall performance superior to that of previous methods. Notably, RWR still performs well even when information on known disease multi-omics compound network is lacking [13]. When applied to AS, RWR identified known AS-related multi-omics compound networks, revealed novel disease metabolites candidates, and inferred their functions via pathway analysis. We further constructed human disease metabolites landscape, revealed the modularity of disease metabolites network, and identified sev-

eral metabolite hotspots. In summary, RWR is a useful tool for prioritizing disease-related multi-omics compound networks and may facilitate understanding of molecular mechanisms of human disease at the multi-omics compound network level.

According to the results of candidate genes, we identified some genes that interacted with preferred AS metabolites. PTPN1, LAMTOR2, IL27RA, and so on, were ranked more forward. Protein-tyrosine phosphatase N1 (PTPN1), also called PTP1B, specifically inactivates phosphorylated STAT5A and STAT5B, which modulates self-renewal capacity and differentiation of memory B-cells [14]. The human PTPN1 gene is located in chromosome 20q13 and has been identified in a number of studies as a region of quantitative trait associated with obesity and type 2 diabetes [15]. Recently, association studies have indicated that single nucleotide polymorphisms within PTPN1 genes are associated with metabolic disorders and obesity [16]. Adapter molecule p14 is called the late endosome/lysosomal adapter and the mitogen-activated protein kinase (MAPK) and the mTOR activator/regulatory complex 2 (LAMTOR2) in the mammalian target. It affects weight row and other gene cell processes such as growth factor signaling and immune proliferation as well as immune cells [17]. The defect of p14 is the molecular cause for development of unknown primary human immunodeficiency syndrome which shows signs of diseases such as Hermansky-Pudlak syndrome [18].

Identification of disease-related metabolites is important in obtaining a better understanding of the metabolic pathology process, in order to improve human drugs [19]. Metabolites, as end products of the cell regulatory process, can be produced by a multi-omic process [20]. In our study, based on comprehensive multi-omics information, we predicted and prioritized disease metabolites. From the results of AS preferred metabolite, we obtained some metabolites such as 2-mercaptoethanol (2-ME), magnesium ion, and aerobactin. 2-ME inhibits growth of Bacillus paraspora NRRL 3240 and formation of aflatoxin in fungi [21]. It is known that 2-ME is necessary for the activity of some enzymes and inhibits certain other enzymes [22]. Studies have reported that there are three potential areas in which 2-ME may be beneficial for disease intervention: (a) potential radioactive adjuvant to enhance effectiveness of local radiotherapy, (b) radiation protectants, (c) radiation-induced use of radiation-induced tumor progression, independent prevention [23]. As one of the important ions associated with bone fusion, magnesium is a titanium surface ion implantation method using magnesium plasma immersion into micro/nanostructures [24]. Studies have shown that magnesium ions play an active role in osteogenic differentiation of rat bone marrow mesenchymal stem cells (rBMMSCs) [25]. Studies have indicated that aerobactin contributes to spread of bacteria from the intestine to organs and body fluids [26]. In E. coli, aerobactin has been shown to be the most effective carrier for iron. Incidence of aerobactin is very high in E. coli strains that cause parenteral infection in humans and animals. The highest proportion of bio-E. coli strains is always present in blood and urine [27].

In summary, our analysis used a multi-omics compound network approach to identify disease metabolites in AS. A total of 579 DEGs and 50 preferred metabolites were identified from 32 samples. In our study, we found metabolites and key genes associated with AS, providing a new basis for treatment of AS in the future. Further research may reveal new insights into the role of metabolites in AS.

Acknowledgements

This work was supported by Shandong Province College Science and Technology Project (J15LL10, ZR2015050013) and Shandong Natural Research Foundation (ZR2016HM31).

Disclosure of conflict of interest

None.

Address correspondence to: Xuesong Wang, Department of Spine, No. 2 Affiliated Hospital of Qingdao University, 127 Siliunanlu Road, Qingdao 266042, Shandong, P.R. China. Tel: +86-0532-82911298; E-mail: Wxs010811@163.com

References

[1] Berdal G, Halvorsen S, van der Heijde D, Mowe M and Dagfinrud H. Restrictive pulmonary function is more prevalent in patients with an-

kylosing spondylitis than in matched population controls and is associated with impaired spinal mobility: a comparative study. Arthritis Res Ther 2012; 14: R19.
[2] Cho H, Kim T, Kim TH, Lee S and Lee KH. Spinal mobility, vertebral squaring, pulmonary function, pain, fatigue, and quality of life in patients with ankylosing spondylitis. Ann Rehabil Med 2013; 37: 675-682.
[3] Papagoras C, Markatseli TE, Saougou I, Alamanos Y, Zikou AK, Voulgari PV, Kiortsis DN and Drosos AA. Cardiovascular risk profile in patients with spondyloarthritis. Joint Bone Spine 2014; 81: 57-63.
[4] Wright C, Sibani S, Trudgian D, Fischer R, Kessler B, LaBaer J and Bowness P. Detection of multiple autoantibodies in patients with ankylosing spondylitis using nucleic acid programmable protein arrays. Mol Cell Proteomics 2012; 11: M9.00384.
[5] Betts KA, Griffith J, Song Y, Mittal M, Joshi A, Wu EQ and Ganguli A. Network meta-analysis and cost per responder of tumor necrosis factor-alpha and interleukin Inhibitors in the treatment of active ankylosing spondylitis. Rheumatol Ther 2016; 3: 323-336.
[6] Iwasaki Y, Sawada T, Hatayama K, Ohyagi A, Tsukuda Y, Namekawa K, Ito R, Saito K and Nakazawa H. Separation technique for the determination of highly polar metabolites in biological samples. Metabolites 2012; 2: 496-515.
[7] Mon A, Abe C, Durazzo TC and Meyerhoff DJ. Effects of fat on MR-measured metabolite signal strengths: implications for in vivo MRS studies of the human brain. NMR Biomed 2013; 26: 1768-1774.
[8] Yao Q, Xu Y, Yang H, Shang D, Zhang C, Zhang Y, Sun Z, Shi X, Feng L, Han J, Su F, Li C and Li X. Global prioritization of disease candidate metabolites based on a multi-omics composite network. Sci Rep 2015; 5: 17201.
[9] Nordstrom A and Lewensohn R. Metabolomics: moving to the clinic. J Neuroimmune Pharmacol 2010; 5: 4-17.
[10] van Driel MA, Bruggeman J, Vriend G, Brunner HG and Leunissen JA. A text-mining analysis of the human phenome. Eur J Hum Genet 2006; 14: 535-542.
[11] Wu X, Jiang R, Zhang MQ and Li S. Network-based global inference of human disease genes. Mol Syst Biol 2008; 4: 189.
[12] Yao Q, Wu L, Li J, Yang LG, Sun Y, Li Z, He S, Feng F, Li H and Li Y. Global prioritizing disease candidate lncRNAs via a multi-level composite network. Sci Rep 2017; 7: 39516.
[13] Huang S, Chaudhary K and Garmire LX. More is better: recent progress in multi-omics data integration methods. Front Genet 2017; 8: 84.
[14] Knecht H, Bruderlein S, Wegener S, Lichtensztejn D, Lichtensztejn Z, Lemieux B, Moller P and Mai S. 3D nuclear organization of telomeres in the Hodgkin cell lines U-HO1 and U-HO1-PTPN1: PTPN1 expression prevents the formation of very short telomeres including "t-stumps". BMC Cell Biol 2010; 11: 99.
[15] Tsou RC and Bence KK. The genetics of PTPN1 and obesity: insights from mouse models of tissue-specific PTP1B deficiency. J Obes 2012; 2012: 926857.
[16] Chiappini F, Catalano KJ, Lee J, Peroni OD, Lynch J, Dhaneshwar AS, Wellenstein K, Sontheimer A, Neel BG and Kahn BB. Ventromedial hypothalamus-specific Ptpn1 deletion exacerbates diet-induced obesity in female mice. J Clin Invest 2014; 124: 3781-3792.
[17] Sparber F, Tripp CH, Komenda K, Scheffler JM, Clausen BE, Huber LA, Romani N and Stoitzner P. The late endosomal adaptor molecule p14 (LAMTOR2) regulates TGFbeta1-mediated homeostasis of langerhans cells. J Invest Dermatol 2015; 135: 119-129.
[18] Schiefermeier N, Scheffler JM, de Araujo ME, Stasyk T, Yordanov T, Ebner HL, Offterdinger M, Munck S, Hess MW, Wickstrom SA, Lange A, Wunderlich W, Fassler R, Teis D and Huber LA. The late endosomal p14-MP1 (LAMTOR2/3) complex regulates focal adhesion dynamics during cell migration. J Cell Biol 2014; 205: 525-540.
[19] Fujimura Y and Miura D. MALDI mass spectrometry imaging for visualizing in situ metabolism of endogenous metabolites and dietary phytochemicals. Metabolites 2014; 4: 319-346.
[20] Wilson AD. Advances in electronic-nose technologies for the detection of volatile biomarker metabolites in the human breath. Metabolites 2015; 5: 140-163.
[21] Silva M, Rey V, Botana A, Vasconcelos V and Botana L. Determination of gonyautoxin-4 in echinoderms and gastropod matrices by conversion to neosaxitoxin using 2-mercaptoethanol and post-column oxidation liquid chromatography with fluorescence detection. Toxins (Basel) 2015; 8.
[22] Click RE. Obesity, longevity, quality of life: alteration by dietary 2-mercaptoethanol. Virulence 2010; 1: 509-515.
[23] Click RE. Alteration of radiation-sensitive processes associated with cancer and longevity by dietary 2-mercaptoethanol. J Cancer Res Ther 2014; 10: 127-132.
[24] Wang G, Li J, Zhang W, Xu L, Pan H, Wen J, Wu Q, She W, Jiao T, Liu X and Jiang X. Magnesium ion implantation on a micro/nanostructured titanium surface promotes its bioactivity and

osteogenic differentiation function. Int J Nanomedicine 2014; 9: 2387-2398.

[25] Ho MH, De Vivo M, Dal Peraro M and Klein ML. Understanding the effect of magnesium ion concentration on the catalytic activity of ribonuclease H through computation: does a third metal binding site modulate endonuclease catalysis? J Am Chem Soc 2010; 132: 13702-13712.

[26] Torres AG, Cieza RJ, Rojas-Lopez M, Blumentritt CA, Souza CS, Johnston RK, Strockbine N, Kaper JB, Sbrana E and Popov VL. In vivo bioluminescence imaging of Escherichia coli O104:H4 and role of aerobactin during colonization of a mouse model of infection. BMC Microbiol 2012; 12: 112.

[27] Ling J, Pan H, Gao Q, Xiong L, Zhou Y, Zhang D, Gao S and Liu X. Aerobactin synthesis genes iucA and iucC contribute to the pathogenicity of avian pathogenic Escherichia coli O2 strain E058. PLoS One 2013; 8: e57794.