Yazdan Rahmati1, Soodabeh Shafiee1,2, Hasan Mollanoori1, Sajjad Esmaeili3*

1 Department of Medical Genetics, Iran University of Medical Sciences, Tehran, Iran
2 Barati medical diagnostic laboratory, Gachsaran, Iran
3 Medical Biology Research Center, Health Technology Institute, Kermanshah University of Medical Sciences, Kermanshah, Iran

CORRESPONDENCE AUTHOR
Mr. Sajjad Esmaeili
E-mail: sajad.smaeili@kums.ac.ir

CITATION

ABSTRACT
Kawasaki disease (KD) is an acute and self-limited vasculitis which has unknown etiology with little described genetic mechanism. To achieve further insight into the molecular mechanisms underlying KD and identify new therapeutic targets for acute KD, we compared the acute and convalescent KD transcriptional profiles by bioinformatics tools. Recently, a novel co-expression algorithm, weighted gene co-expression network analysis (WGCNA), has provided more comprehensive data analysis. In the present study, the WGCNA package was applied to the datasets of KD patients obtained from the Gene Expression Omnibus (GEO). We identified network modules of co-expressed genes in the acute phase subject dataset and verified their preservation in the convalescent dataset. Then, in the non-preserved modules, we selected hub genes that distinguishing acute from convalescent phase and performed functional enrichment analyses of genes involved in these modules. We identified several genes that may play essential role in KD pathogenesis and discriminating acute from convalescent phase patients. Functional enrichment analysis revealed that most of them contribute to innate immune system defense against infectious agents. Neutrophil mediated immunity by antimicrobial humoral response against infectious agents. In conclusion our results indicated several genes essential for KD pathogenesis using WGCNA algorithm. These genes may be used as potential novel therapeutic targets.

Keywords: Kawasaki disease, WGCNA, innate immune response, antimicrobial peptides, acute phase, convalescent phase
1. INTRODUCTION
KD is an inflammatory vasculitis with unknown etiology that occurs in infantile and toddlers. Its main complication is the development of coronary artery lesions (CALS) including coronary arterial dilation, stenosis, and aneurysms [1-3]. Kawasaki disease (KD) exhibits many signs such as high temperature, erythema which occurs in the mucosa of lips or mouth, alterations in the body organs, and large lymph nodes in the neck [4-6]. KD among Japanese and Japanese American children is clearly more common (265/100,000) than other populations like other Asian countries (51 to 194/100,000), Europe (8.39/100,000) or American children (20.8/100,000) [7-12]. To date, the etiology of KD has remained an undetermined question and the causative agents are also unexplained [13]. However, based on research studies, genetic predisposition considered a cause for the development of KD and also the interaction with an unknown infectious source can predispose to this disease [14]. The observation of familial occurrence and raised frequency in Asian populations, taken together suggest the existence of a genetic basis [15, 16]. Associations with genetic variants in several genes such as BLK, CASP3, CD40, FCGR2A, IPTKC, and HLA class II have been ascertained in diverse populations. Also, it is known that the increased risk for the development of coronary aneurysms in populations with European ethnicity is correlated with the genetic variation in the TGF pathway (TGFβ2, TGFβR2, SMAD3) [17].

Heart damage is visible in a significant number of untreated KD patients which can lead to the development of myocardial infarction, unexpected death, or ischemic heart disorder from ectasia (Coronary artery aneurysms) [18, 19]. Early diagnosis is crucial for effective treatment which will result in abolition of the inflammatory process and reduces the risk of coronary artery aneurysms (CAA) rates to approximately 5-10% [20]. CAA is the most important complication of KD patients. In untreated KD children, the disease-associated inflammation modifies the arterial wall which leads to CAA in 25% of them [21]. In developed countries, KD has been reported as the most prevalent cause of acquired heart disease in children [22]. KD shares certain features with other childhood febrile conditions, including, infectious (e.g. staphylococcal and streptococcal toxic shock syndromes, measles and other viral illnesses) and inflammatory conditions which makes differential diagnosis difficult [20]. Although nowadays there are guidelines to assist diagnosis based on clinical signs and symptoms, such as echocardiography, and laboratory variables, distinguishing KD in early stages from other mimicking conditions for treatment and impediment of CAA development remains a momentous mission [23].

Recently, analysis algorithms for differential co-expression network advanced and applied to study the expression data of genes and microRNAs [24, 25]. There is a new functional strategy in systems biology, the weighted gene co-expression network analysis (WGCNA) algorithm, which identifies the most important genes in co-expressed genes in association with a sample trait [26, 27]. Based on similarities in expression profiles of samples, WGCNA makes modules, sub-network regions, and detects those of which contain associated genes [28, 29]. Through analysis of these modules in two different conditions, acute against convalescent, we aimed to identify genes in relation to the acute state that may be used as new therapeutic targets for KD.

2. METHODS
2.1. Data selection, preprocessing and detection differentially expressed genes (DEGs) between acute and convalescent samples
In the present study, one expression array, GSE63881, was obtained from NCBI Gene Expression Omnibus (GEO). GSE63881 consists of 341 samples, 171 out of which are acute, and 170 are convalescent. This dataset was produced using Illumina HumanHT-12 V4.0 expression beadchip. Normalizing DFGs of this dataset was performed by quantile normalization method in limma package. For more qualified results, we included only single measurement of each gene, with the aggregate function in the S4Vectors package, which gives an average measurement for each gene’s probes. Using
Sajjad Esmaeili et al.

After preprocessing and quality assessment, using limma package DEGs with cutoff $|\log_{2} FC| > 0.5$ and P.value < 0.05 as the threshold, we detected 374 up-regulated and 687 downregulated genes. The DEGs of this dataset was shown as volcano plot in Figure 1.
Figure 1: The volcano plot of differentially expressed genes of GSE63881 between acute and convalescent samples.

3.2. WGCNA prerequisites, outlier samples, batch effects, comparability

One of the key points in using the WGCNA package is the number of samples, the number of samples for construction of weighted gene co-expression network should be at least 15 while the number of samples included in current study was 171 and 170 for acute and convalescent status, respectively. Since only one dataset with one platform and one laboratory condition was used, the batch effect was not observed. Fulfilling the WGCNA package requirements, two and five outlier samples, based on the Z.K score, were excluded from the acute and convalescent datasets respectively (Figure 2). Furthermore, datasets comparability was evaluated by the softConnectivity function which provides positive correlation values when there is high comparability between two datasets. Our datasets were comparable as the overall gene expression correlation (cor=0.93, P.value<1e−200), and the overall gene connectivity (cor=0.7, P.value<1e−200) were significant.( Figure 3)

Figure 2: Clustering dendrogram of samples based on their euclidean distance and detecting outlier samples. (A) Acute samples, (B) Convalescent samples.

Figure 3: Results of softConnectivity. Correlations are positive and the p-values are significant for both datasets. In addition correlations and p-values are better for expression than for connectivity.
3.3. Identification of modules able to distinguish acute from convalescent datasets

Acute and convalescent datasets exhibited a scale-free topology with the powers equal to 6 for both acute and convalescent samples and the Scale-free Topology Fit Index reached values above 0.9 for low powers (< 30). These results also show that the batch-effects were not present in our datasets (Figure 4). After constructing weighted gene co-expression networks for each acute and convalescent expression dataset, the modules of these datasets were identified. Based on the WGCNA package, a module is a group of strongly co-expressed genes, which have similar biochemical and functional properties or belong to similar pathways. By hierarchical clustering, 14 modules were identified for the acute dataset as the reference network and also 15 modules for convalescent samples as the test network. All of which had different number of genes which are labelled by different colors and are shown in Figure 5. By considering convalescent dataset a test network, we evaluated the preservation level across the two networks by finding preserved and non-preserved modules in two datasets. As it was expected, except for one module, they were well preserved between two mentioned expression datasets. Therefore, this module was potentially related to the KD disease pathogenesis. Qualifying the module preservation level was conducted through taking advantage of modulePreservation function from WGCNA package. For each module, a Z-score representing the preservation level was calculated. Modules with Z-score higher than 10 were considered well preserved between two datasets, whereas modules with Z-scores <10 were assumed less preserved and distinguishing acute from the convalescent condition. (Figure 6)
3.4. Identification of hub genes and functional analysis

Since we found tan module non-preserved between acute and convalescent datasets, it seemed that tan module-related genes play an important role in the acute phase of disease. We set out to identify genes related to this module (Table 1). According to the results of gene ontology, the most important signaling pathways pertained to the innate immune defense against the infectious agent, and in particular, in the process that neutrophils, as professional killers of invading pathogens to the human organism act by releasing various antimicrobial proteins during degranulation (Figure 7A). The cenplot that shows the relation between signaling pathways and tan module genes is shown in Figure 7B. The PPI network of detected genes was constituted by STRING (Figure 8A), 36 nodes and 102-edged related hub genes in PPI were recognized. After PPI network establishment, the PPI data imported into Cytoscape software and the interactions of hub genes were detected. (Figure 8B). The genes including, BPI, CAMP, DEFA1, DEFA3, DEFA4, ELANE, LCN2, MMP8, and ORM2 were the most interacting genes identified as hub genes.

<table>
<thead>
<tr>
<th>Module</th>
<th>Hub Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tan</td>
<td>CAMP, AZU1, OLR1, RNASE3, CEACAM6, SLC22A16, CEACAM8, BEX1, DEFA4, MS4A3, DEFA1B, DEFA1, ELANE, DEFA3, ABCA13, LCN2, CTSG, TCN1, BPI, PRTN3, CEBPE, LTF, COL17A1, TACSTD2, PTPN20, MPO, HYAL3, ATP8B4, TFF3, OLFM4, SERPINB10, MMP8, TCTEX1D1, ORM2, CRISP3, ORM1</td>
</tr>
</tbody>
</table>
Figure 7: Functional enrichment analysis of tan module genes: (A) Gene ontology analysis of all genes in tan module. (B) cnetplot of all genes in tan module that depicts the linkages of genes and most important signaling pathways.
Figure 8: The PPI network of detected genes that was analyzed by String software. 36 nodes and 102 edged in the PPI network. BPI, CAMP, DEFA1, DEFA3, DEFA4, ELANE, LCN2, MMP8 and ORM2 are the most interacted genes and identified as hub genes.
4. DISCUSSION
To analyze complicated diseases with obscure etiology such as KD, a comprehensive study is required to cover all aspects of the disease, including molecular phenomena, pathogenicity, physiological side effects, and cellular alterations. Since there is no complete data about KD, molecular analysis alongside clinical studies can help us explain etiology of diseases and achieve new biomarkers and safe therapy [31, 32].

In this study, we employed WGCNA package on the expression data obtained from GSE63881 of the acute and convalescent phase of KD patients. By WGCNA package, we discovered that one out of 14 network modules (Tan module) in the acute network were not preserved in the convalescent KD network. This gene module may play a significant part in KD pathogenesis and consequently, it is probable for the hub gens within this module to serve as new diagnostic or therapeutic purposes. Afterward, we summarized the list of genes within this module through distinguishing only the hub genes or the most KD-associated genes according to the WGCNA package. Functional enrichment analysis of genes in Tan module revealed that the most significant GO terms are related to the innate immune response against infectious agents.

Although the etiology of KD remains unknown so far, several lines of clinical and epidemiological data suggest an environmental trigger such as epidemic occurrence, change in season, and temporal clustering of cases [33, 34]. The new hypothesis of the etiology of KD explains that an abnormal response to undisclosed infectious diseases in genetically susceptible children may be the trigger of the disease. One of the first hypotheses based on the implication of an infectious agent in KD was superantigen-mediated etiology that came from the similarities in clinical features and immunological reactions between KD, toxic shock syndrome (TSS), and streptococcal toxic shock syndrome (STSS), but further studies have failed to confirm these findings [35-38]. The scientific society has so far failed to achieve an understanding about the infectious origin of KD (viral, bacterial or fungal), or identification of the underlying immune mechanisms behind KD, accordingly, other possible environmental triggers have been assessed by various research groups to determine the role of the resident gut microbiota as a potential contributor to KD. Many pieces of research and microbiologic observations fortified this hypothesis that heterogeneity and abnormalities in the gut microbiota composition may set off or contribute to the development of KD [39].

We detected α-defensins in our detected module. Based on this result, we suggest that an unknown infectious agent or an unknown intestinal microbiota, may induce KD. α-defensins, host defense peptides that supply protection against viral, bacterial, and fungal infections, are mostly found in neutrophils and Paneth cells of the small intestine. If host primary defense mechanisms are incapable to identify and eliminate the infectious agent, chemokines and other warning signals summon circulating polymorphonuclear leukocytes (PMNs) to the Infection environment. In humans, antimicrobial peptides that belong to the α-defensin (DEFA1-4) and cathelicidin (hCAP-18/LL-37 encoded by the CAMP gene) families are the main components of the PMN apparatus. The α-defensins and hCAP-18/LL-37 are stored in the primary (azurophil) granules and secondary (specific) granules, respectively. This different storage sites is an indication for their respective scene of action because azurophil granules are preferentially derived from phagosome and specific granules are mainly in the extracellular environment [40]. The antimicrobial function of α-defensins during the viral, bacterial, protozoa (Entamoeba histolytica), and fungal infections in small intestine is well explained [39, 41-43]. Bactericidal/permeability-increasing protein (BPI), neutrophil elastase (ELANE), proteinase 3 (PRTN3), Neutrophil Gelatinase-Associated Lipocalin encoded by the LCN2 gene, and cathepsin G (CTSG) are other components of azurophil granule that participate in the killing pathogens in phagosomes [40, 44, 45]. It has been shown that systemic upregulation of LCN2 associate with the development of several inflammatory disease, such as peritonitis, atherosclerosis, and vasculitis/Kawasaki disease [46]. With regard to the NOD-like receptor signaling pathway in KEGG results, it is recognized that...
NOD2 (a pathogen recognition receptor of the NOD-like receptor family) controls the expression of a subset of paneth cell-expressed alpha-defensins, and the cumulation of both microbiota and pathogenic infection in the terminal ileum in mice [47].

CONCLUSION

In the present study, we attempted to figure out potential molecular mechanisms in KD by using a comprehensive algorithm for differential co-expression network. The non-preserved module, tan module, has possibly been associated with the pathogenesis of KD. Functional enrichment analyses demonstrated that the innate immune defense against the infectious agent, and, the processes that neutrophils kill invading pathogens by mediating antimicrobial humoral response, involved in the pathogenesis of KD. Additionally, hub genes may be used as transcriptional biomarkers or potential therapeutic targets.

Declaration of competing interest

The authors have indicated that they have no conflicts of interest regarding the content of this article.

REFERENCES


Email: info@siftdesk.org