

CASP5 and CR1 as Potential Biomarkers for Kawasaki Disease: An Integrated Bioinformatics-Experimental Study

Yazdan Rahmati

Iran University of Medical Sciences and Health Services: Tehran University of Medical Sciences

Hasan Mollanoori

Iran University of Medical Sciences: Tehran University of Medical Sciences

Sajad Najafi

Iran University of Medical Sciences: Tehran University of Medical Sciences

Sajad Esmaeili

Kermanshah University of Medical Sciences

Mohammad-Reza Alivand (✉ mohammadreza_alivand@yahoo.com)

Tabriz University of Medical Sciences <https://orcid.org/0000-0002-5847-3594>

Research article

Keywords: Gene Expression, Hub Genes, RT-PCR, ROC Curve, WGCNA

Posted Date: March 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-279444/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Pediatrics on December 1st, 2021. See the published version at <https://doi.org/10.1186/s12887-021-03003-5>.

Abstract

Kawasaki disease (KD) is a pediatric inflammatory disorder causes coronary artery complications. The disease overlapping manifestations with a set of symptomatically like diseases such as bacterial and viral infections, juvenile idiopathic arthritis, Henoch-Schönlein purpura, infection of unknown etiology, group-A streptococcal and adenoviral infections, and incomplete KD could lead to misdiagnosis of the disease. In the present study, we applied weighted gene co-expression network analysis (WGCNA) to identify network modules of co-expressed genes in GSE73464 and also, limma package was used to identify the differentially expressed genes (DEGs) in KD expression arrays composed of GSE73464, GSE18606, GSE109351, and GSE68004. By merging the results of WGCNA and limma, we detected hub genes. Then, analyzed the peripheral blood mononuclear cells (PBMCs) of 16 patients and 8 control subjects using Real-Time Polymerase Chain Reaction (RT-PCR) to evaluate the previous results. We assessed the diagnostic potency of the screened genes by plotting the area under curve (AUC). We finally identified 2 genes *CASP5* and *CR1* which were shown to potentially discriminate KD from other similar diseases and also from healthy people. The results of RT-PCR and AUC confirmed the diagnostic potentials of two suggested biomarkers for KD.

1. Introduction

Kawasaki Disease (KD) is a rare systemic vasculitis disorder which affects the medium and small arteries particularly coronary arteries in infants and young children (1). The disease is diagnosed globally with the highest incidence reported for the eastern Asia countries including, in order of magnitude, Japan, South Korea, and Taiwan; in non-Asian countries, however, substantial and meaningful differences in the incidence rate have been demonstrated (1–4). Generally, the difference has been linked with the racial composition of the societies (5). A functional single-nucleotide polymorphism (*itpkc-3*) in the inositol 1,4,5-trisphosphate 3-kinase C (ITPKC) gene has been associated with enhanced susceptibility to the disease and developing coronary artery aneurisms, so that C allele of *itpkc-3* gene can conduce to immune hyper-reactivity in KD (e.g., by increase in interleukin-2 [IL-2] transcript level) (6). Nonetheless, the etiology of KD remains enigmatic (5–7); accordingly, it is required to further explore the molecular mechanisms of the disease to achieve a decisive diagnosis.

The clinical manifestations of KD include a series of symptoms such as fever lasting longer than five days, bilateral non-purulent conjunctivitis and cervical lymphadenopathy, rashes, lip fissures, erythema and edema in oral mucosa (characteristic strawberry tongue) and peripheral extremities, and etc. which are similar to those of other types of neonatal illnesses (8) such as epistaxis, juvenile idiopathic arthritis, scarlet fever, and etc., and this can lead to misdiagnosis of the condition (9). Delayed or missed diagnosis can pose the patient at higher risk of coronary artery abnormalities (10). till now, a large number of biomarkers including those designed for inflammatory, proteomics, gene expression profiles, and micro-RNA characteristics have been failed in diagnostic approaches due to unacceptable sensitivity and specificity (10–13), and so, they may couldn't suggest confirmation in the diagnostic decision (14). Therefore, introduction of a novel, decisive, prognostic, or diagnostic biomarker would be an inevitable

necessity for the KD as a complicated condition requiring immediate diagnosis. The genetic basis known for KD has been interesting for the researchers and clinicians to develop a genetic approach for diagnosis and prediction of prognosis in patients with KD. The genetic approaches might be classified into 'candidate gene approaches' and 'genome-wide approaches' (14–17). Chaudhary et al. (16) have enlisted the studies conducted on the genetic markers of KD. They, however, have stressed on irreproducibility of the results among different nations. So, one drawback may be concluded from the insufficient genes list is that it doesn't make possible to accurately diagnose and predict prognosis of the disease.

Over the past decade, substantial pathogenetic clues have been found for various diseases such as immune reactions via the genome-wide approach (18, 19). In this regard, using the DNA microarray method, alteration in expression levels including up- and down-regulation of thousands of genes and also pathological mechanisms can be detected in a single chip such as experiments conducted for KD(20). Weighted gene co-expression network analysis (WGCNA), likewise, empowers the scientists to explore network alterations and basic mechanisms among highly correlated genes and to also help find new biomarkers from disease associated genes cluster (21–23).

In the present study, we investigated the co-expressed genes in KD patients using WGCNA package to explore network modules. Then, four microarray datasets of KD from the Gene Expression Omnibus (GEO) repository were integrated to find the differentially expressed genes (DEGs) in the patient' samples compared to control groups. By merging the results, we screened 35 genes and then through evaluation of their aberrant expression with symptomatically like diseases including bacterial and viral infections, JIA (juvenile idiopathic arthritis), HSP (Henoch-Schönlein purpura), infection of unknown etiology, GAS (group A streptococcal) infection, human adenovirus (HAdV) infection, and incomplete KD, two genes were selected as potential biomarkers. Eventually, we employed Real-Time Polymerase Chain Reaction (RT-PCR) to substantiate the selected genes.

2. Materials And Methods

2.1. Data preprocessing and weighted co-expression network construction

Following search of "Kawasaki disease" and "*Homo sapiens*" keywords at <https://www.ncbi.nlm.nih.gov/geo>, GEO series GSE73464 was selected for weighted co-expression construction. To normalize the dataset, we used quantile normalization method in limma (linear models for microarray data) R package (24) and log₂ transformation. An only single measure was included for each gene by the aggregate function in S4Vectors package. To construct the weighted gene co-expression network of GSE73464, we used WGCNA R package. Constructing the weighted gene network is necessarily accompanied by choosing the soft thresholding power β , which culminates in adjacency calculation through creating co-expression similarity. We conducted the analysis of the topology network and the improved selection of a proper soft-thresholding power using pickSoftThreshold function. The adjacency results were converted into Topological Overlap Matrix (TOM) to minimize the impacts of

noise and spurious associations. The results of TOM and cutreeDynamic function act respectively as inputs and branch cutting to produce the gene-based dendrogram. The minimum module size was set at 30, and we used the module detection sensitivity deep Split 2 in blockwiseConsensusModules function to construct the network. To determine the module(s) containing the significant genes, we assessed: 1) the relationships of individual genes and clinical data by defining Gene Significance, and 2) the correlation between the eigengene module and gene expression profile for each module as a quantitative criterion of modular membership.

2.2. Identification of clinically significant modules

We measured the relationships between individual genes and clinical data through defining Gene Significance (GS) as the absolute value. For each module, a quantitative criterion of module membership (MM) was considered the correlation between the eigengene module and gene expression profile through which similarity of all genes on the array was measured and determined to every module. By employment of GS and MM, it is feasible to identify interesting module(s) containing the genes with the great significance of both clinical data and module membership.

2.3. Protein-Protein interaction (PPI) Network Construction

We constructed the protein-protein interaction (PPI) network of the selected module using STRING and cytoscape (25). The cytoscape molecular complex detection (MCODE) plugin was used to identify the finest clusters (26). ClueGO v2.5.3 (27) was utilized to perform pathway enrichment analysis of the genes, and the most important signaling pathways based upon KEGG database were detected (28).

2.4. Expression array datasets

Three other expression arrays including GEO series GSE18606, GSE109351, and GSE68004 from the GEO (containing a total number of 104 control and 196 patient samples), then were normalized, and eventually DEGs were screened under the settings of a cutoff defined as $p\text{-value} < 0.05$ and $|\log_2\text{fold change}| \geq 1$.

2.5. Comparison of diagnostic values of detected hub genes in KD and symptomatically like diseases

We used the GSE73464 expression array dataset to find the gene distinctions between KD and a set of other symptomatically like diseases including bacterial and viral infections, JIA, HSP, and infection of unknown etiology, and also GSE68004 dataset was used to explore other distinctive genes between KD and another group of similar infections such as GAS, HAdV, and incomplete KD. The DEGs between KD and other diseases, and also compared to healthy controls (defined as \log_2 fold change > 1 in KD vs all the comparative groups) were, as well, regarded as biomarkers. To determine the diagnostic value of the hub genes, we plotted receiver operating characteristic (ROC) curve and the area under the curve (AUC). P-values of < 0.05 were considered for statistical significance.

2.6. Experimental Methods

2.6.1. Subjects and sample preparation

Blood samples were collected in the tubes containing EDTA from the children referred to Ali-Asghar Hospital of Iran University of medical sciences, Tehran, Iran diagnosed with KD. Regarding the approved codes of Ethics Committee form the Iran University of Medical Sciences as well as the declaration of Helsinki, the written informed consent forms were signed for each sampling. The participants included 16 patients and 8 healthy children aged less than five years old, out of which 62.5% (20 subjects) were male and remaining 37.5% (12 subjects) were female.

2.6.2. Peripheral blood mononuclear cell (PBMC) Isolation

Each blood sample was diluted with the phosphate buffered saline (PBS) in 1:1 ratio. After adding 3 ml of ficoll solution, the tubes were centrifuged at 400 g for 30 minutes. Then the layer containing peripheral blood mononuclear cells (PBMCs) was separated from other blood cell components, washed thrice with PBS buffer to achieve higher purity, and eventually centrifuged again at 200g for 10 minutes.

2.6.3. RNA extraction and cDNA production

Total RNA was extracted from the PBMCs in each group by Total RNA Miniprep Purification Kit (Cat 17200, GeneMark Diagnostics Inc., Georgia, and USA). The quality of extracted RNA was confirmed by measuring the optical density of OD₂₆₀/OD₂₈₀ ratio and viewing on 2% gel electrophoresis. Furthermore, concentration of the extracted RNA was measured using NanodropND-1000 spectrophotometer (Thermo Fisher Scientific, Wilmington, USA). Then we converted the RNA to cDNA using 2X RT-PCR Pre-Mix (Taq) Kit (BioFACT, South Korea).

2.6.4. Primer design

The primer's sequences were designed using Genscript online tool (available at <https://www.genscript.com/tools/real-time-pcr-tagman-primer-design-tool>). To evaluate their quality, the primers were assessed on online oligonucleotide calculator tool (<http://biotools.nubic.northwestern.edu/OligoCalc.html>), and Gene Runner software (Hastings Software, Colorado, and USA). The primers are available in Supplementary Table 2.

2.6.5. Real-Time PCR

The lyophilized powder of the primers (Cinaclone, Iran) was dissolved by adding sterile distilled water to each tube (according to the manufacturer's instructions); then were stored as stock at -20°C. StepOne RT-PCR system (Applied Biosystems, Forster City, CA, USA) and SYBR green PCR kit (Takara Biotech, China) were utilized to conduct the quantitative RT-PCR (qRT-PCR) for measuring the gene expression levels. The results were then normalized based on GAPDH expression levels. The final solution volume was 15 µL, comprised of 1 µl cDNA, 1 µl of each forward and reverse primers (2 µl in total), 10 µl Taq polymerase, and 2 µl DEPC water. A number of forty cycles were chosen to conduct the PCR under the following condition: 95°C (2 min), 95°C (30 sec), and 60°C (40 sec). A negative control for each individual gene was

provided at each time to examine the possible contaminations. Finally, the expression fold change of the examined genes was assessed using the Threshold Cycle (CT) method by the following formulae:

$$R = 2^{- (\Delta\Delta CT)}$$

$$\Delta\Delta CT = (CT_{\text{target}} - CT_{\text{reference}})_{\text{healthy}} - (CT_{\text{target}} - CT_{\text{reference}})_{\text{patient}}$$

3. Results

3.1. WGCNA and Modules identification

We constructed the co-expression networks (modules) of GSE73464 dataset via WGCNA. The clinical trait and gene expression aberrations between the patient and control groups were thoroughly fit in the dendrogram of the samples and included in the expression dataset (Fig. 1A). According to the approximate scale-free topology as the crucial criterion to opt the most precise soft thresholding power, we selected power 6 for the rest of analysis (Fig. 1B). Then, the clustering dendrogram of genes was plotted with a large minimum module size of 30 using the medium sensitivity (deepSplit = 2) for cluster splitting (Fig. 1C). Clustering dendrogram of GSE73464 included nine modules (Fig. 1D). Among all created modules, blue (correlation = 0.82) was selected and accentuated per dataset (Fig. 1E). Finally, the ultimate gene list containing 253 genes was achieved in blue module.

3.2. Enrichment Analysis

A total number of 253 genes were selected in the blue module for enrichment analysis and then their PPI network was constructed using the STRING database and cytoscape software. Among them, 50 genes were selected as hub genes through screening the top candidates with the highest number of interactions, Closeness and Betweenness Centrality (Fig. 2). According to the network sizes, MMP9, ITGB2, SPI1, SELL, and CD68 play an important role in pathogenesis of KD. We determined the interactive relationships between the hub genes in the network using the MCODE plugin. Two clusters were found with the following characteristics: 23 nodes and 199 edges in cluster 1, and 10 nodes and 26 edges in cluster 2. They were identified from the MCODE based on a scoring system (cutoff k-score = 12). The screened genes were enriched and analyzed through the Cytoscape plug-in ClueGo based on KEGG database. As it can be concluded from the pie chart, chemokine signaling pathway, tuberculosis, yersinia infection, pathogenic *Escherichia coli* infection, and fc gamma r-mediated phagocytosis are the most important signaling pathways (Fig. 4A, and 4B).

3.3. Identification of DEGs

Four expression arrays (GSE73464, GSE18606, GSE109351, and GSE68004) were analyzed to evaluate the alterations of gene expression levels between healthy control and patient samples using Limma package. We screened all the DEGs using $|\log_2FC| > 1$ and $p\text{-value} < 0.05$ as the threshold and then were

showed in volcano plots (Fig. 5A-D). Following overlapping, we identified 70 upregulated and 17 downregulated common genes (Fig. 5E, F). Through merging the blue module with DEGs, 35 genes were identified that could discriminate KD from healthy controls.

3.4. Biomarkers to differentiate KD from other symptomatic-overlapping febrile conditions

To date, there is no definite DEG-based diagnostic test for KD. Accordingly, we included two expression array datasets (GSE73464, and GSE68004) in our investigation aimed to help distinguish KD from other resembling febrile conditions through discerning the most informative and distinctive gene expression signature. According to our criterion, by comparing the resemblance and dissimilarity of expressional patterns of our hub genes which were specified via their logFC values (defined as log₂ fold change > 1 in KD vs all of the comparator groups), two genes signature comprised of *CASP5* and *CR1* were identified (Fig. 6).

3.5. Real-Time PCR

To experimentally verify the differential expression of *CR1* and *CASP5* in KD patients, qRT-PCR assay was used, the results showed significantly increased expression in KD patients compared to healthy controls (Fig. 7A, B). According to RT-PCR data, ROC curve analysis showed that both *CASP5* and *CR1* could distinguish KD from the common febrile diseases (Fig. 7C, D).

3.6. Hub genes validation

A ROC curve was generated to verify the diagnostic performance of *CASP5* and *CR1* based on the disease vs healthy control expression array databases (GSE73464, GSE68004, GSE18606, and GSE109351). The AUC showed that *CASP5* and *CR1* indicated perfect diagnostic efficiency for KD compared to normal samples. The AUC of *CASP5* related to GEO series GSE73464, GSE68004, GSE18606, and GSE109351 was 0.949, 0.982, 0.955, and 1 respectively (Fig. 8A-D). Same measures of AUC for *CR1* were 0.866, 0.713, 0.883, and 1 respectively (Fig. 9A-D). AUC was also employed to evaluate the diagnostic power of *CASP5* and *CR1* to distinguish KD from other similar febrile conditions (Fig. 10).

4. Discussion

Despite more than fifty years since Dr. Tomisaku Kawasaki primarily introduced Kawasaki disease (29), no confirmatory test has been developed for the clinical diagnosis of KD. Moreover, the pathophysiological mechanisms of systemic vasculitis in the disease have not yet been completely understood. In histopathological assessments, it has been found that vasculitis, as an inflammatory cardiovascular disorder develops in KD, damages coronary arteries and cardiac sequelae leading to coronary failure and formation of aneurysms (30). Genetically, as well, the genes included in the early detection might not be applied to differentiate KD from symptomatically similar diseases.

Wright et al. (31) introduced a signature composed of a set of genes, in order of priority regarding logistic regression coefficient, as specific biomarkers to distinguish KD from the similar diseases including CACNA1E, DDIAS, KLHL2, PYROXD2, SMOX, ZNF185, LINC02035, CLIC3, S100P, IFI27, HS.553068, CD163, and RTN. The authors revealed that these biomarkers expressed differentially in samples taken from the patients with KD. Using a bio-signature of 13 transcripts in a parallel regularized regression model, Wright et al. [11] established their criteria for differential diagnosis as follows: the $\log_{2}FC > 1$ in KD compared to the control group, and $\log_{2}FC$ difference of at least '1' when comparing KD to the similar diseases. Some uncertainties are seen in the criteria regarding the precise detection of the differential genes in KD especially if an almost identical gene expression pattern similar to KD's condition obscures the experiment.

In the present study, comparing the expression levels of 35 mentioned genes in KD compared to other symptomatically like diseases mainly infections of bacterial or viral causes, juvenile idiopathic arthritis, Henoch-Schönlein purpura, infection of unknown etiology, group A streptococcal and adenoviral infections, and incomplete KD, we revealed two upregulated genes including *CR1* and *CASP5* in samples of patients with KD compared to other. Both the genes were found to be significantly over-expressed in KD patients ($\log_{2}FC > 1$).

The complement receptor type 1 (*CR1*) processes and clears opsonized complement safety complexes (32). Several roles have been attributed to this protein including negative regulation of complement cascades, mediating the immune compliance and phagocytosis, and inhibiting both classical and alternative pathways of apoptosis. Significant upregulation of *CR1* in KD besides the role of *CR1* in immunity, for instance in innate immune response, might elucidate the inflammatory symptoms in (30).

Caspase-5 together with caspase-1 and caspase-4 plays role in apoptosis and contributes significantly to the immunity (33). Caspase-1 expresses constitutively, whereas caspase-5 (the human orthologue of caspase-11) expression is induced by bacterial lipopolysaccharide (LPS) (34). Caspase-5 is found in a limited number of tissues including peripheral blood lymphocytes, liver, placenta, spleen, and colon. Conversely, caspase-4 exists in all body tissues (35, 36). Mutations in *CASP5* gene have been associated with several types of cancer such as leukemia, endometrial cancer, lung cancer, gastrointestinal tract and colorectal cancers (37). As well, a direct relationship might exist between skin inflammation and *CASP5* upregulation (37). Caspase-5 and caspase-1 may functionally synergize to regulate the processing of interleukin-1 β (IL-1 β), pro-IL18, and pyroptosis (38).

Conclusions

Using a series of experiments aim to evaluate the expression level of genes in KD patients compared to healthy control samples and also between patient's samples before and after treatment by intravenous immunoglobulin (IVIG), we identified two genes *CASP5* and *CR1* as specific biomarkers for KD. The results of our study suggest that KD could be distinguished from a group of diseases with similar manifestations by evaluating the expression level of the candidate genes. Designing a diagnostic test

based on the gene expression signature could provide early diagnosis and therefore, help prescribe immediate treatment which is critical in KD patients and consequently can prevent life-threatening cardiac complications in the affected children.

Declarations

Ethics approval and consent to participate: Regarding the approved codes of Ethics Committee at Iran University of Medical Sciences as well as the declaration of Helsinki, the written informed consent forms were signed for each sampling.

Consent for publication: Not Applicable

Availability of data and materials: The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Competing interests: The authors declare no conflict of interest.

Funding: This study was funded by Iran University of medical sciences, grant number 96-03-30-30735.

Authors' contributions: Yazdan Rahmati designed the study. Mohammad Reza Alivand and Hasan Mollanoori, Sajad Esmaili and Sajad Najafi wrote the first draft of the manuscript. Yazdan Rahmatia and Mohammad Reza Alivand supervised the study. All authors read and approved the final version of the manuscript.

Acknowledgements: This study was funded by Iran University of medical sciences, grant number 96-03-30-30735.

References

1. Jennette JC, Falk R, Bacon P, Basu N, Cid M, Ferrario F, et al. 2012 revised international chapel hill consensus conference nomenclature of vasculitides. 2013.
2. Holman RC, Belay ED, Christensen KY, Folkema AM, Steiner CA, Schonberger LBJTPidj. Hospitalizations for Kawasaki syndrome among children in the United States, 1997–2007. 2010;29(6):483-8.
3. Lin M-C, Lai M-S, Jan S-L, Fu Y-CJJotCMA. Epidemiologic features of Kawasaki disease in acute stages in Taiwan, 1997–2010: effect of different case definitions in claims data analysis. 2015;78(2):121-6.
4. Makino N, Nakamura Y, Yashiro M, Sano T, Ae R, Kosami K, et al. Epidemiological observations of Kawasaki disease in Japan, 2013–2014. 2018;60(6):581-7.

5. Rowley AH, Shulman STJFip. The epidemiology and pathogenesis of Kawasaki disease. 2018;6:374.
6. Onouchi Y, Gunji T, Burns JC, Shimizu C, Newburger JW, Yashiro M, et al. ITPKC functional polymorphism associated with Kawasaki disease susceptibility and formation of coronary artery aneurysms. 2008;40(1):35-42.
7. Rowley AHJlJord. Is Kawasaki disease an infectious disorder? 2018;21(1):20-5.
8. Pinna GS, Kafetzis DA, Tselkas OI, Skevaki CLJCOiID. Kawasaki disease: an overview. 2008;21(3):263-70.
9. Komatsu H, Tateno AJJop, health c. Failure to distinguish systemic-onset juvenile idiopathic arthritis from incomplete Kawasaki disease in an infant. 2007;43(10):707-9.
10. Singh S, Jindal AK, Pilania RKJlJord. Diagnosis of Kawasaki disease. 2018;21(1):36-44.
11. Sato S, Kawashima H, Kashiwagi Y, Hoshika AJIJoRD. Inflammatory cytokines as predictors of resistance to intravenous immunoglobulin therapy in Kawasaki disease patients. 2013;16(2):168-72.
12. Korematsu S, Uchiyama S-i, Miyahara H, Nagakura T, Okazaki N, Kawano T, et al. The characterization of cerebrospinal fluid and serum cytokines in patients with Kawasaki disease. 2007;26(8):750-3.
13. Rasouli M, Heidari B, Kalani MJIL. Downregulation of Th17 cells and the related cytokines with treatment in Kawasaki disease. 2014;162(1):269-75.
14. Engelberg R, Martin M, Wrotniak BH, Hicar MDJPR. Observational study of Interleukin-21 (IL-21) does not distinguish Kawasaki disease from other causes of fever in children. 2017;15(1):32.
15. Onouchi YJlJord. The genetics of Kawasaki disease. 2018;21(1):26-30.
16. Chaudhary H, Nameirakpam J, Kumrah R, Pandiarajan V, Suri D, Rawat A, et al. Biomarkers for Kawasaki disease: clinical utility and the challenges ahead. 2019;7:242.
17. Rahmati Y, Mollanoori H, Kakavandi N, Nateghian A, Sayyahfar S, Babaei V, et al. Gene expression analysis in Kawasaki disease; bioinformatics and experimental approach. 2020;20:100423.
18. Daetwyler HD, Villanueva B, Woolliams JAJPo. Accuracy of predicting the genetic risk of disease using a genome-wide approach. 2008;3(10):e3395.
19. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. 2018;50(9):1219-24.
20. Popper SJ, Shimizu C, Shike H, Kanegaye JT, Newburger JW, Sundel RP, et al. Gene-expression patterns reveal underlying biological processes in Kawasaki disease. 2007;8(12):1-12.
21. Giulietti M, Occhipinti G, Principato G, Piva FJCO. Weighted gene co-expression network analysis reveals key genes involved in pancreatic ductal adenocarcinoma development. 2016;39(4):379-88.
22. Giulietti M, Occhipinti G, Righetti A, Bracci M, Conti A, Ruzzo A, et al. Emerging biomarkers in bladder cancer identified by network analysis of transcriptomic data. 2018;8:450.
23. Rahmati Y, Mollanoori H, Shafiee S, Esmaeili S. Identification of 7 key age-related genes involved in Kawasaki disease, an integrated study by metaDE and weighted gene co-expression network

analysis.

24. Smyth GK. Limma: linear models for microarray data. *Bioinformatics and computational biology solutions using R and Bioconductor*: Springer; 2005. p. 397-420.
25. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*. 2003;13(11):2498-504.
26. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC bioinformatics*. 2003;4(1):2.
27. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. 2009;25(8):1091-3.
28. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, et al. KEGG for linking genomes to life and the environment. *Nucleic acids research*. 2007;36(suppl_1):D480-D4.
29. Rowley AH. Incomplete (atypical) Kawasaki disease. *The Pediatric infectious disease journal*. 2002;21(6):563-5.
30. Durongpisitkul K, Soongswang J, Laohaprasitiporn D, Nana A, Prachuabmoh C, Kangkagate C. Immunoglobulin failure and retreatment in Kawasaki disease. *Pediatric cardiology*. 2003;24(2):145-8.
31. Wright VJ, Herberg JA, Kaforou M, Shimizu C, Eleftherohorinou H, Shailes H, et al. Diagnosis of Kawasaki disease using a minimal whole-blood gene expression signature. *JAMA pediatrics*. 2018;172(10):e182293-e.
32. Krych-Goldberg M, Atkinson JP. Structure–function relationships of complement receptor type 1. *Immunological reviews*. 2001;180(1):112-22.
33. Martinon F, Tschopp JJCD, Differentiation. Inflammatory caspases and inflammasomes: master switches of inflammation. 2007;14(1):10-22.
34. Bouchier-Hayes L, Martin SJ. CARDINAL roles in apoptosis and NFκB activation. *Vitamins & Hormones*. 67: Elsevier; 2004. p. 133-47.
35. Lin XY, Choi MSK, Porter AGJJoBC. Expression analysis of the human caspase-1 subfamily reveals specific regulation of the CASP5 gene by lipopolysaccharide and interferon-γ. 2000;275(51):39920-6.
36. Munday NA, Vaillancourt JP, Ali A, Casano FJ, Miller DK, Molineaux SM, et al. Molecular cloning and pro-apoptotic activity of ICERelll and ICERellll, members of the ICE/CED-3 family of cysteine proteases. 1995;270(26):15870-6.
37. Gao J, Wells JA. Caspase-4 and Caspase-5. *Handbook of Proteolytic Enzymes*: Elsevier; 2013. p. 2265-9.
38. Clore GM, Appella E, Yamada M, Matsushima K, Gronenborn AMJB. Three-dimensional structure of interleukin 8 in solution. 1990;29(7):1689-96.

Figures

Fig.1

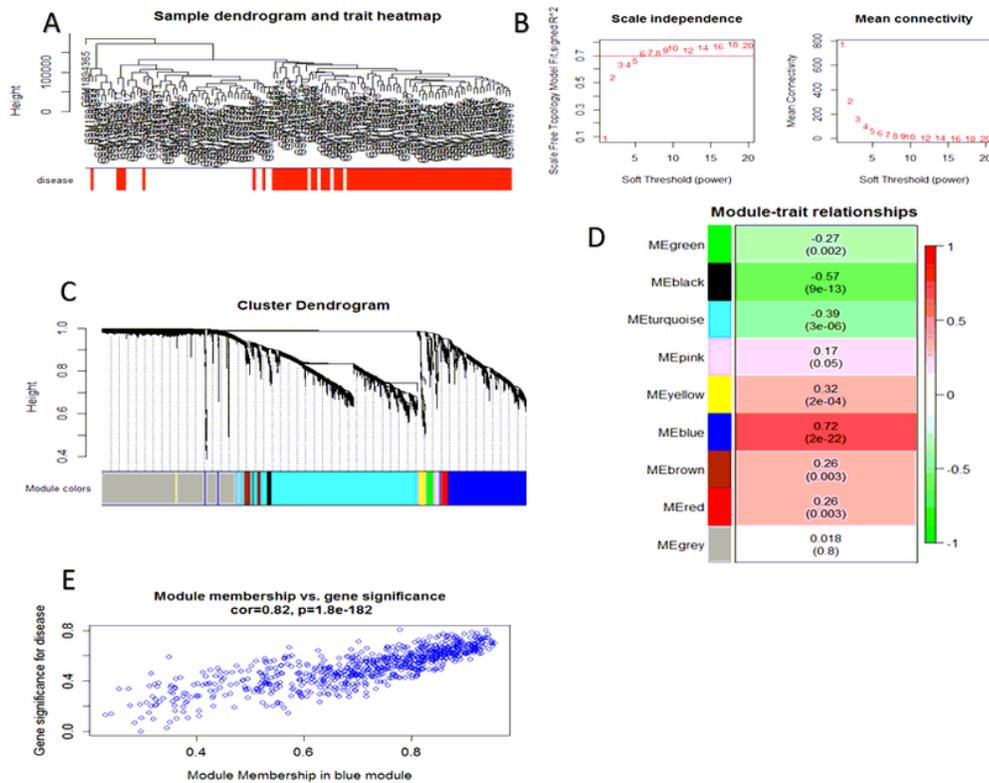


Figure 1

WGCNA data for GSE73464. (A) Clustering dendrogram of samples based on their Euclidean distance and how the clinical traits relate to the sample dendrogram, white means a low value, red a high value. (B) Analysis of network topology for various soft-thresholding powers. Scale-free fit index and the mean connectivity for various soft-thresholding powers. (C) Clustering dendrogram of DEGs with dissimilarity based on topological overlap, together with assigned module colors. Each color represents a module in the constructed gene co-expression network. (D) Module-trait associations for DEGs. Each row corresponds to a module eigengene, column to a trait. Each cell contains the corresponding correlation and p-value. The table is color-coded by correlation according to the color legend. (E) (F) scatterplot of Gene Significance (GS) for disease gene significance vs. Module Membership (MM) of blue module.

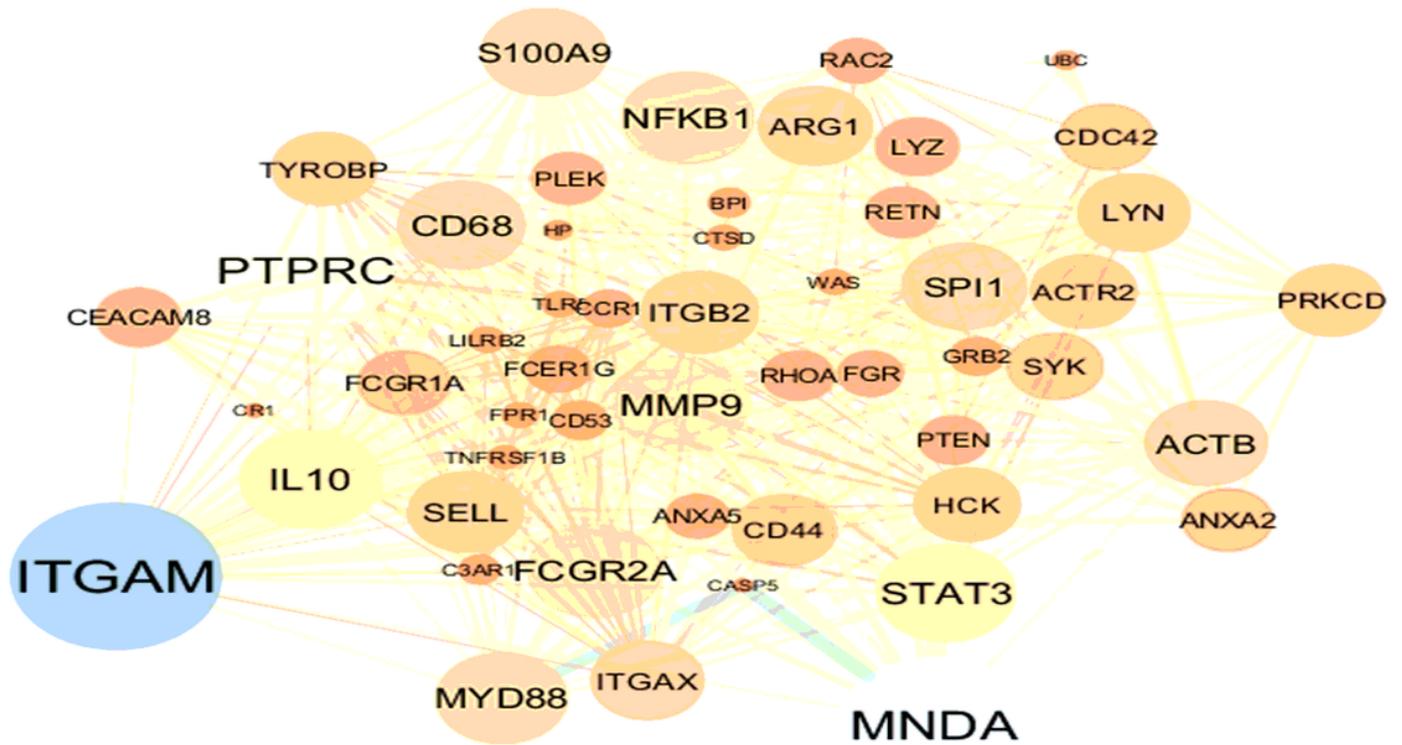


Figure 2

Protein-protein interaction network of the DEGs. Showing genes with the highest number of interactions. The gradual change in the size indicates the connectivity degree. The gradual change in the color indicates the betweenness centrality. The thickness of the edges stands for the credibility.

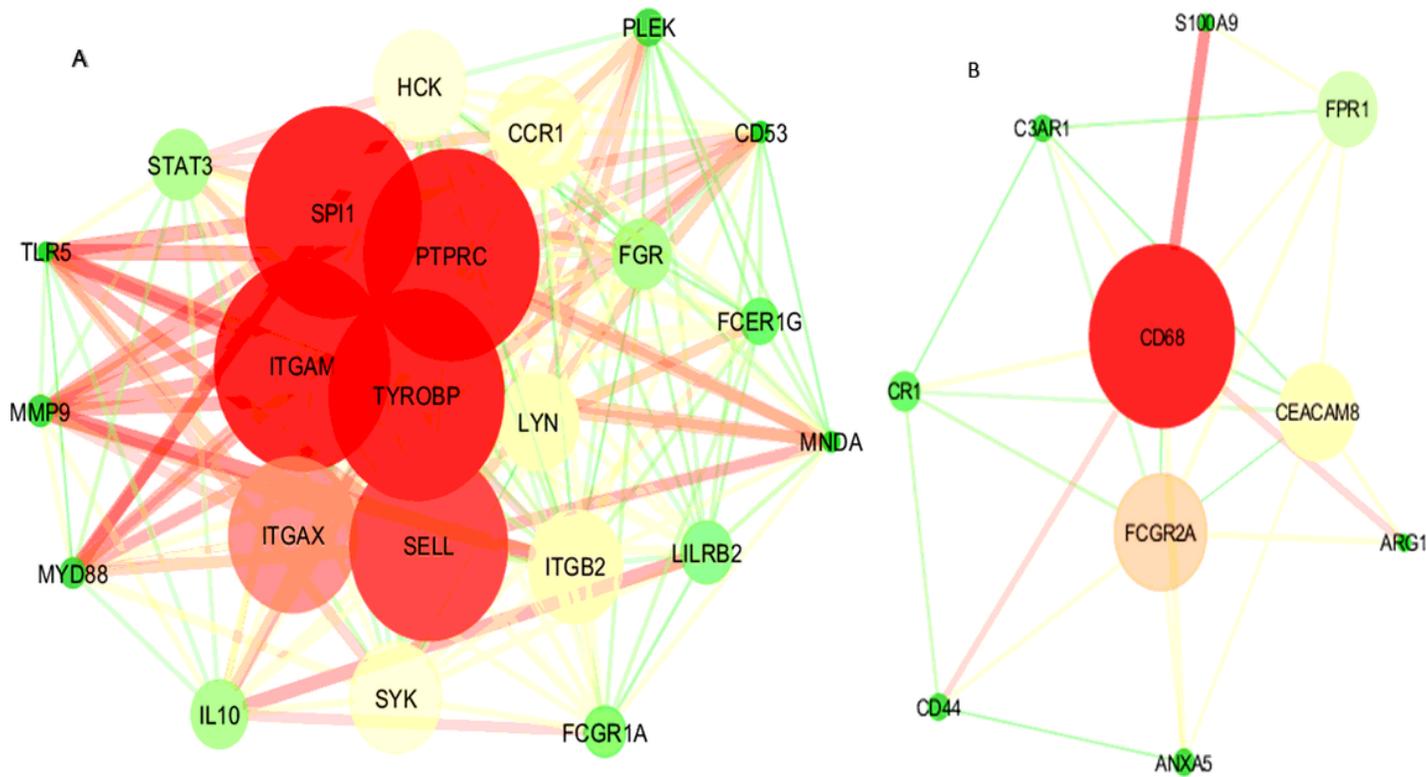


Figure 3

A total of 2 modules were identified in the PPI network using the MCODE tool in Cytoscape software. Module (A) had 23 nodes and 199 edges. Module (B) had 10 nodes and 26 edges.

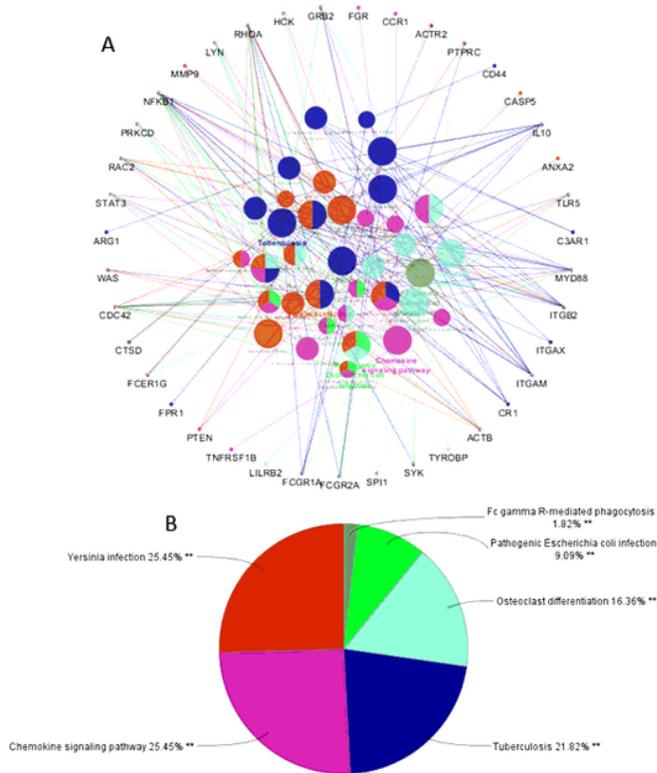


Figure 4

Enriched KEGG pathways using the ClueGo plugin of Cytoscape: (A) functionally grouped networks based on KEGG database of Genes with terms as nodes linked based on their κ score level (≥ 0.3). Only the most significant term in each group is presented. Associated proteins were visualized using CluePedia. (B) A chart with functional groups, including specific terms for the detected signaling pathways in the KD.

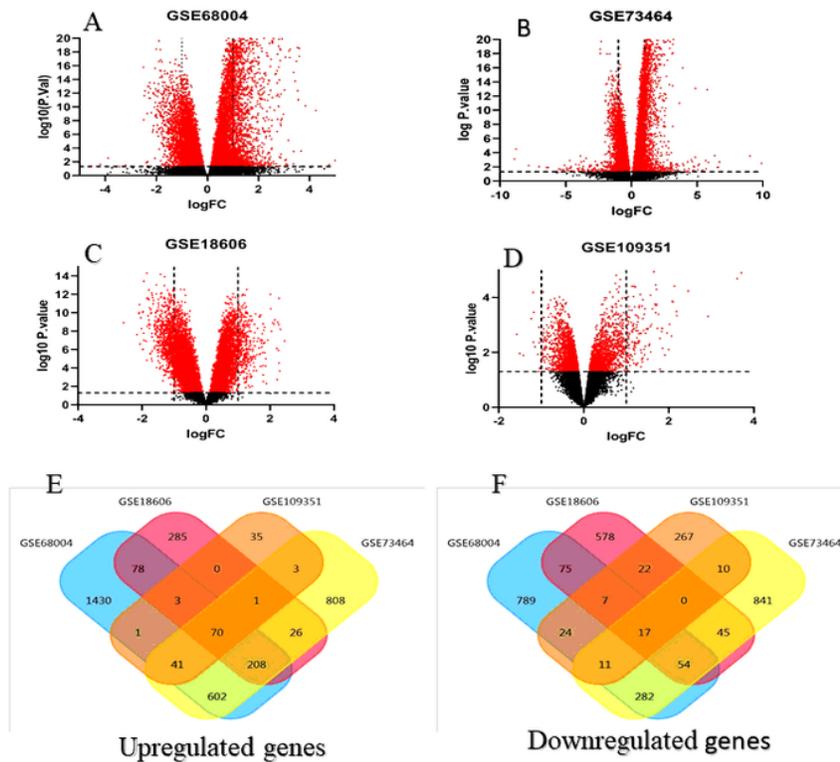


Figure 5

Differentially expressed genes and common differentially expressed genes in four datasets between KD and healthy control samples. (A, B, C, D) The volcano plots of differentially expressed genes in GSE68004, GSE73464, GSE18606, and GSE109351 respectively using $p\text{Value} < 0.05$ and $|\log_{2}\text{fold change}| \geq 1$. (E, F) Common differentially expressed genes in four datasets.

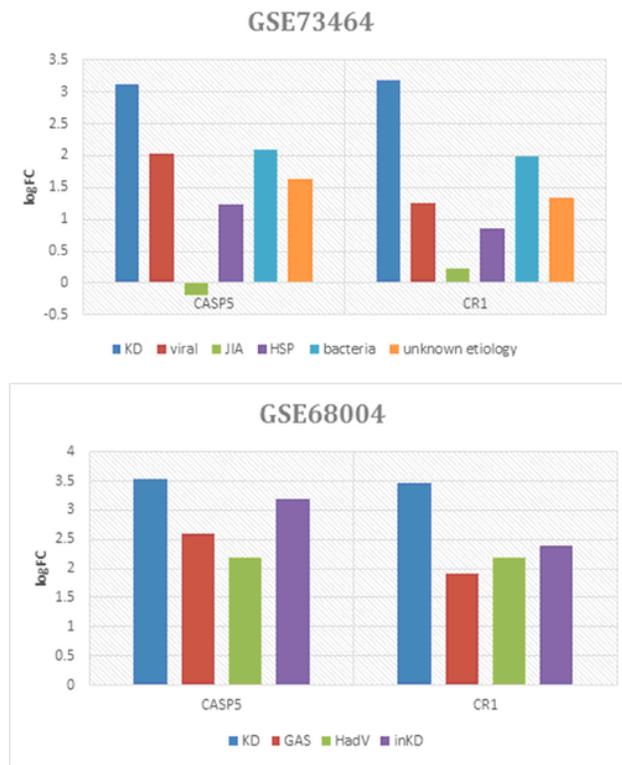


Figure 6

Expression level of CASP5 and CR1 in KD and other symptomatic-similar diseases in GSE73464 and GSE68004.

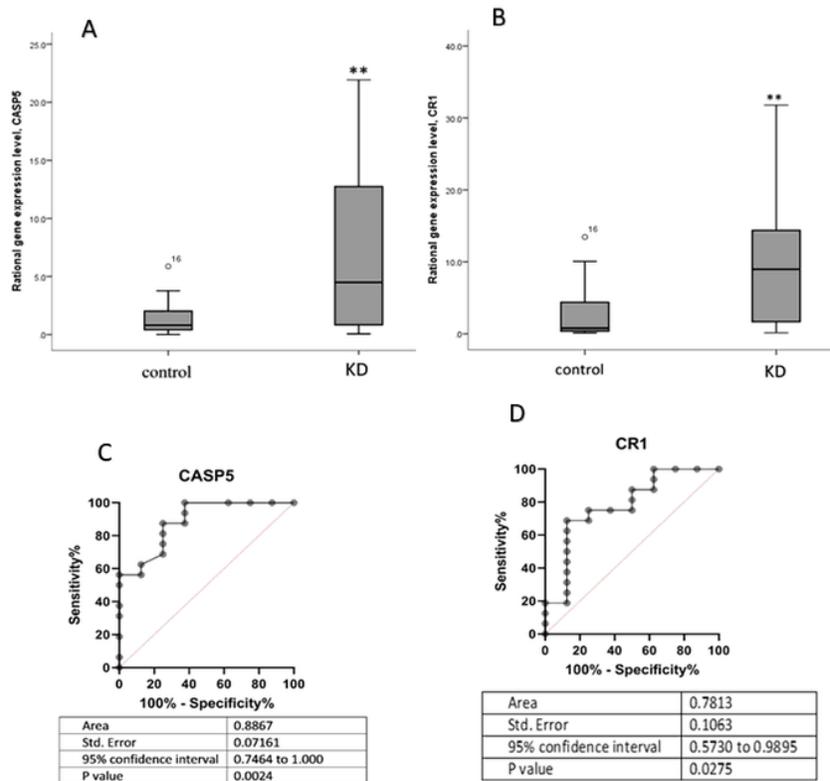


Figure 7

(A-B) Box plots of serum levels of CASP5 and CR1 in 16 KD patients compare to 8 control samples. (C-D) Receiver operating characteristic (ROC) curves of (C) CASP5, (D) CR1, based on real-time PCR results.

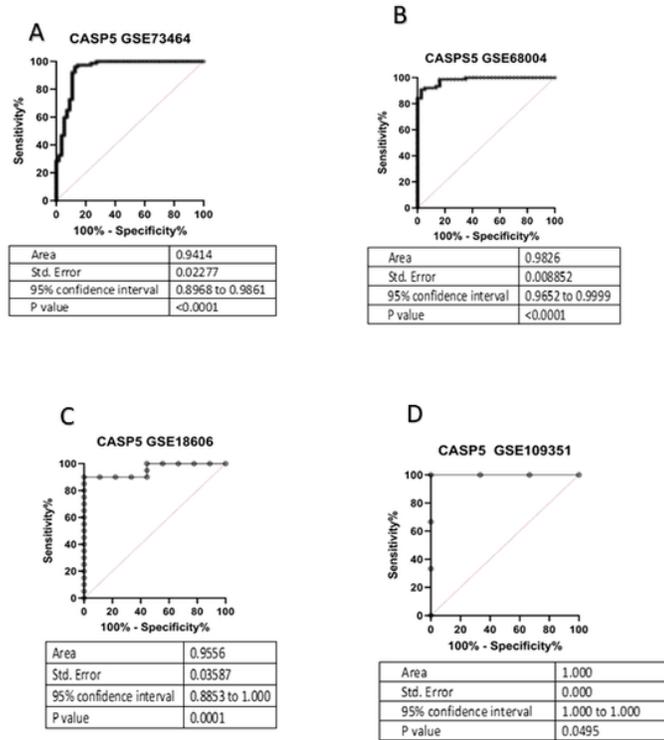


Figure 8

ROC curve analysis of CASP5 gene diagnosis. Receiver operating characteristic (ROC) curves and area under the curve (AUC) statistics are used to evaluate the capacity to discriminate KD from normal controls with excellent specificity and sensitivity in (A) GSE73464, (B) GSE68004, (C) GSE18606, (D) GSE109351.

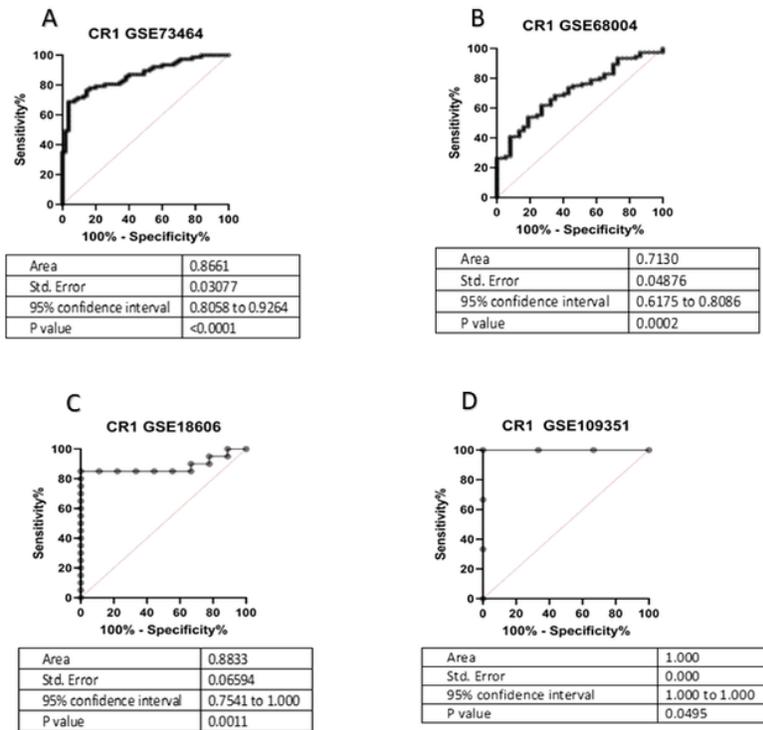


Figure 9

ROC curve analysis of CR1 gene diagnosis. Receiver operating characteristic (ROC) curves and area under the curve (AUC) statistics are used to evaluate the capacity to discriminate KD from normal controls with excellent specificity and sensitivity in (A) GSE73464, (B) GSE68004, (C) GSE18606, (D) GSE109351.

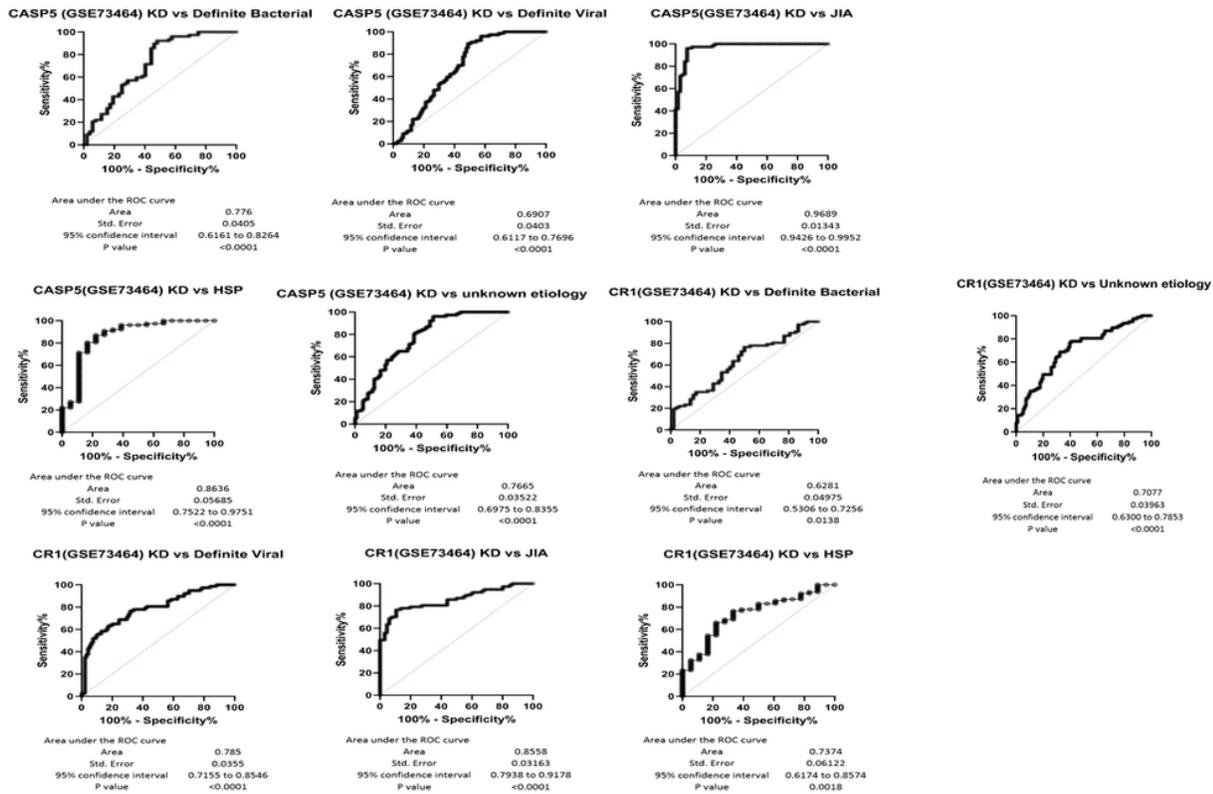


Figure 10

ROC curve analysis of CASP5 and CR1 genes diagnosis based on GSE73464. Receiver operating characteristic (ROC) curves and area under the curve (AUC) statistics are utilized to determine the capacity to discriminate KD from other febrile conditions with excellent specificity and sensitivity in GSE73464 dataset.

Supplementary Files

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

- [supplementaryfile.docx](#)