International Journal Of Public Mental Health And Neurosciences
ISSN: 2394-4668
(An Official publication of Sarvasumana Association)

In-Silico analysis for the identification of differential expressed genes and candidate biomarkers in Kawasaki disease

Jayanti Shukla,
Biotechnology, GEU
Dehradun, India
Shuklajayanti12@gmail.com

Bhairvi Pant,
CSE, GEU
Dehradun, India

Neema Tufchi,
Obs & Gynae, Shri Ganga Ram hospital
New Delhi, India

Gracy Chand Paul
GEU
Dehradun, India

Kumud Pant,
Biotechnology, GEU
Dehradun, India

Manu Pant
GEU
Dehradun, India

Srishti Chaudhary
GEU
Dehradun, India

Diksha Panwar
GEU
Dehradun, India

Somya Sinha,
Biotechnology, GEU
Dehradun, India

***Abstract*— Kawasaki illness is an acute febrile systemic vasculitis with a complex aetiology involving genetics and environmental factors that affects children younger than five years. Genetic association and genome-wide association studies (GWAS) have paved the way for better understanding of the molecular mechanisms underlying Kawasaki's disease. In this study, 1,354 DEGs for Kawasaki's disease were found in the Kawasaki disease's group, with 20 genes upregulated and 15 genes downregulated. According to the GO analysis performed, the DEGs for Kawasaki's disease have been grouped together in the categories of immune response, inflammatory response, cellular response to lipopolysaccharide, positive regulation of nf-kappab transcription factor, positive regulation of T-cell proliferation, positive regulation of inflammatory response, innate immune response, positive regulation of inter-leukin-12 production, aging, and bacterium defense response. Furthermore, KEGG pathway analysis has revealed that Kawasaki's disease shares top enriched pathways, including Malaria, Leishmaniasis, Rheumatoid arthritis, Tuberculosis, Chagas's disease, Allograft rejection, Inflammatory bowel disease, Cytokine-Cytokine receptor interaction, Graft-versus-host disease and Phagosome. The current work was based on using an integrated and assimilated analysis strategy to find the DEGs as well as other biological pathways and functions that are shared by KD, thereby improved the understanding of the disease's pathophysiology. Furthermore, these findings might lead to identification of potential and probable biomarkers for the differential diagnosis of KD, as well as therapeutic targets for the development and expansion of new Kawasaki disease's treatments.**

***Index Terms*— Kawasaki's disease, GWAS, KEGG pathway analysis, inflammatory response, Transcription factors, enrichment pathways.**

I. INTRODUCTION

Mucocutaneous lymph node syndrome or Kawasaki disease (KD), is an inflammatory condition that affects the arteries, veins, and capillaries. This disease is common in children and usually accompanied with fever, bilateral nonexudative conjunctivitis, erythema of the lips and oral mucosa, changes in the extremities, rash, and cervical lymphadenopathy. It also affects the lymph nodes and creates nasal, mouth, and throat problems. It is the most predominant cause of heart dis-ease in

© IJPMN, Volume 9, Issue 3, December-2022
(This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution License citing the original author and source)

children, but it's also an atypical, rare and lethal condition. In roughly 15 to 25 percent of untreated children with the illness, coronary artery aneurysms or ectasia develop, which can lead to myocardial infraction, sudden death, ischemic heart disease, and other complications [1]. KD has several stages. It is characterized by a high fever that lasts five or more days, bloodshot eyes, bright red, swollen lips, a "strawberry" tongue that appears shiny and bright with red spots on it, rash on torso, swollen hands and feet, red palms and soles of the feet, swollen lymph nodes [2]. Heart problems may also appear during this time. After two weeks of fever, further symptoms begin. The affected child's skin on his or her hands and feet may begin to peel and come off in sheets. During this stage, some children may experience transient arthritis or joint pain. Abdominal pain, emesis, diarrhea, bladder hypertrophy, and temporary threshold shift (temporary hearing impairment) are some of the other indications and symptoms [3].Despite significant progress in understanding the physiopathology of KD, early detection, therapeutic and remedial intercession, and understanding the underlying molecular pathways of KD are difficult [4]. With the advancement of bioinformatics and molecular biology, microarray technology has become increasingly popular for studying the molecular mechanisms of a variety of disorders. Such studies can be judiciously employed to ascertain the underlying mechanisms of KD. The current study gave additional insight into the pathogenesis of KD at the molecular level by studying the biological functions and pathways shared by KD patients [5], which may aid in the disease's subsequent diagnosis and treatment

.

II. Methodology

A. *Microarray data*

The aim and objective of the current study was to identify gene expression signatures in patients with KD (Kawasaki disease). The gene ex-pression profiles of the GSE109430 datasets were obtained from the Gene Expression Omnibus data-base. A total of 36 samples were included in this dataset, 24 KD patients and 12 controls. GEO Datasets were used to search for Kawasaki disease. The database stores all original submitter-supplied Platform, Sample and Series records, and curated gene expression Dataset records. Retrievals include the title, summary, organ-ism, and accession for each record.

B. *Identification of DEGs*

Data pre-processing was performed using GEO2R [6] and was applied to screen DEGs between the following groups: KD vs. Controls [7]. Identification of differentially expressed genes, associated functional terms pathways, and candidate diagnostic biomarkers in inflammatory bowel diseases by bioinformatics analysis. GEO2R is an interactive tool that enables the analysis of approximately 90% of GEO Series as soon as they are released. The GSE (number) was pasted, that was GSE109430 and clicked to analysis, this step of Analysis was main [8].

C. *GO and KEGG pathway enrichment analyses*

It is intended to methodically extract biological meaning from a substantial inventory of genes or proteins using the Database for Annotation, Visualization, and Integrated Discovery [7] which is a program accessible on the internet. Using DAVID [7], GO enrichment and KEGG path-way analyses were conducted to obtain and analyze the screened DEGs at the practical extent. There are several tools available with DAVID. These tools provide an extensive and diverse set of functional annotation techniques for understanding and comprehending the bio-logical meaning behind large inventory of gene lists. With DAVID, you can identify functionally related gene groups with enriched sequences. You can also visualize and conjure genes on KEGG pathway maps along with exploring their names in batch [9]. It combines GO and KEGG pathways and gives out the best possible result of genes.

D. *PPI network analysis*

Additionally, to assess and continue to evaluate the operative interactions among DEGs at the biochemical level, a PPI network was con-structed. The DEGs were plotted and portrayed using the Search Tool for the Retrieval of Interacting Genes [7,10]. A STRING database is freely accessible database which contains information from many sources such as experimental data, public text collection, computational prediction methods, etc. [11]. The goal of STRING DATABASE is to achieve a diversified and impartial global network, including indirect (functional) inter-actions and including direct (physical), by aiming to collect, score and integrate all publicly available sources of protein–protein interaction in-formation (PPI), and to complement these with computational predictions those are done before-hand [12].

III. Results and Discussion

The present study employed in silico techniques which include bioinformatic tools such as databases, pharmacophores, homology models, quantitative structure-activity relationships, and other molecular modelling approaches, machine learning, network analysis tools, data mining and types of data analysis tools. Available in vitro data on KD was used to construct and test the model using in silico approaches. Such models have been widely used in the discovery, clarification, and physicochemical characterization of absorption, distribution, metabolism, excretion, and toxicity features [13]. The Gene Expression Omnibus (GEO) database was used for the purpose of retrieving a gene expression dataset (GSE109430) for this study [14]. When KD patients were compared to controls, the DEGs were discovered. As a result of combining Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) studies in the DAVID data-base, we have

(This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution License citing the original author and source)

been able to determine the functional enrichment and major signaling path-ways associated with DEGs. In addition, to identify the essential genes, a protein-protein interaction (PPI) network comprising common DEGs was built.

A. Identification of DEGs

The gene expression dataset GSE109430 was inputted from the GEO database. GEO2R tool were used to determine the DEGs between the disease samples and the controls. A total of 24 KDs were recognized in the UC group using the threshold of P<0.05 and |log2FC| >1, including 20 upregulated genes and 15 downregulated genes in the KD vs. control samples.

The top 10 up- and downregulated genes for KD versus the control are listed in table. Studies conducted before have illustrated that the precursory GO terms can be of great significance to the pathogenesis of KD as well as previous studies have shown that there is overlap between them. To perform a systematic and structured characterization of each DEG for functional annotations, and pathway analyses, including GO and KEGG analyses, were carried out using DAVID to further explore the biological functions of each DEG. The top 10 GO terms in the category biological process are presented in Fig. 2. The outcome of the GO analysis has elucidated that most of the DEGs in the KD were simultaneously enriched in immune response, inflammatory response, innate immune response, Cellular response to Lipopolysaccharide, Positive regulation of NF- kappaB transcription factor activity, aging, defense response to bacterium, Positive regulation of T CELL PROLIFERATION, Positive regulation of inflammatory response and positive regulation of interleukin-12 production as shown in fig. (9). Similarly, KEGG pathway analysis indicated that the DEGs in KD were primarily enriched in Cytokine-cytokine receptor interaction, Tuberculosis, Chagas disease, Malaria, Rheumatoid arthritis, Phagosome, Leishmaniasis, Inflammatory bowel disease, Allograft rejection and Graft-versus-host disease. Fig. (8). Studies and researches conducted before have illustrated that communicable microbes also play a role in developing KD.

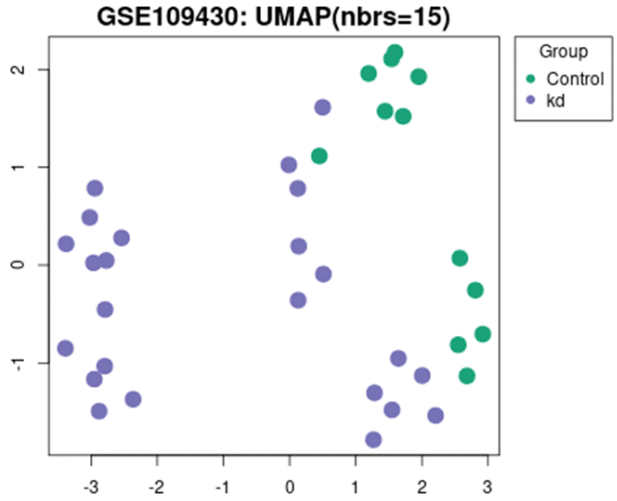

Fig. 1. GO and pathway enrichment analysis of DEGs

B. Construction of PPI network and identification of hub genes

The PPI networks were constructed on the basis of interaction of DEGs to under-stand and explore the relationship and association of DEGs at biochemical and protein level. The KD's DEGs and the overlapping DEGs were mapped into PPI networks using STRING server. The pre-defined criterion of a combined score of >0.7 was the standard, and a PPI network of the overlapping DEGs consisting of 172 nodes and 349 edges was obtained [15].

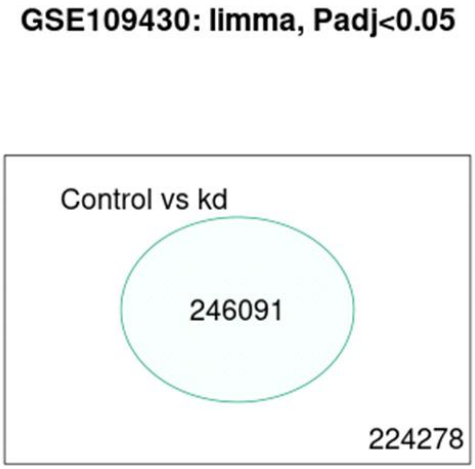

Fig. 2. Control vs. KD

(This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution License citing the original author and source)

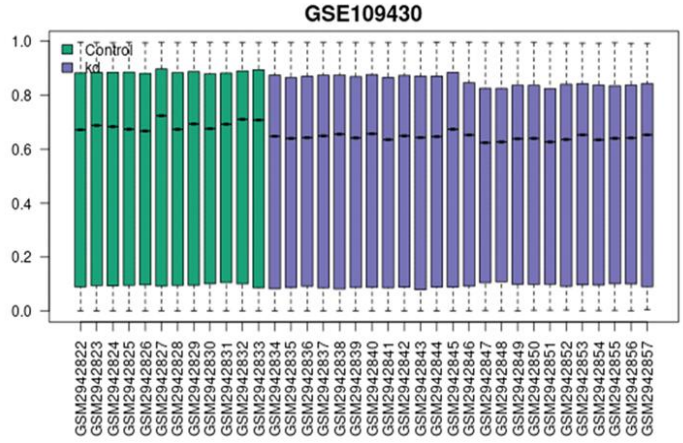

Fig. 3. GSE109430

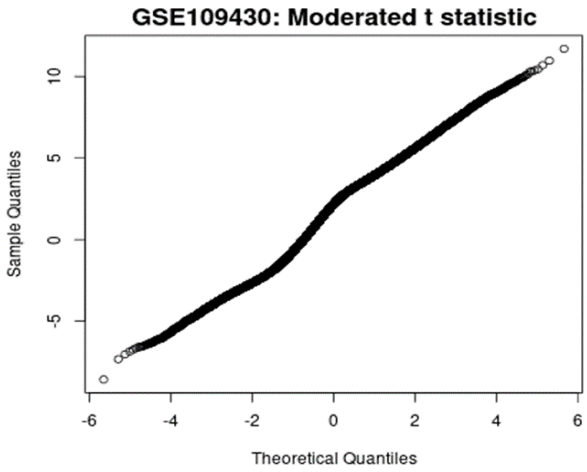

Fig. 6. Moderated t statistic

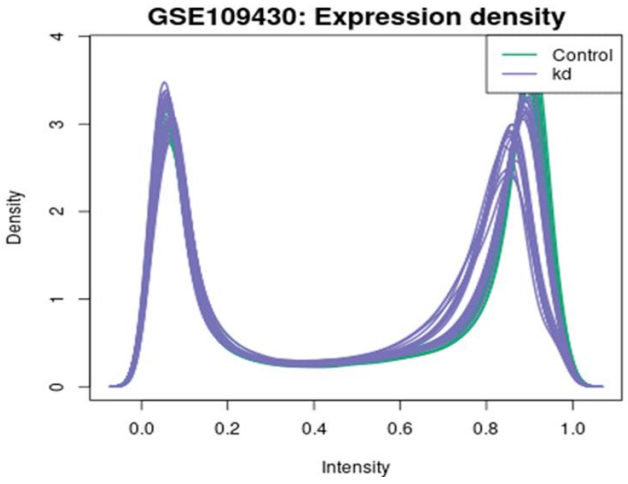

Fig. 4. Expression density

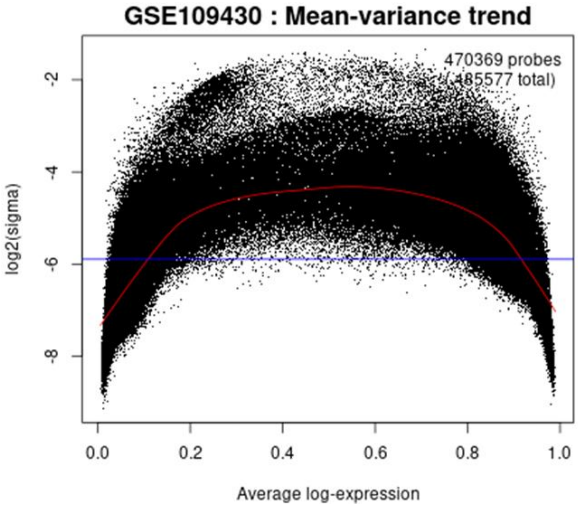

Fig. 7. Mean Variance band

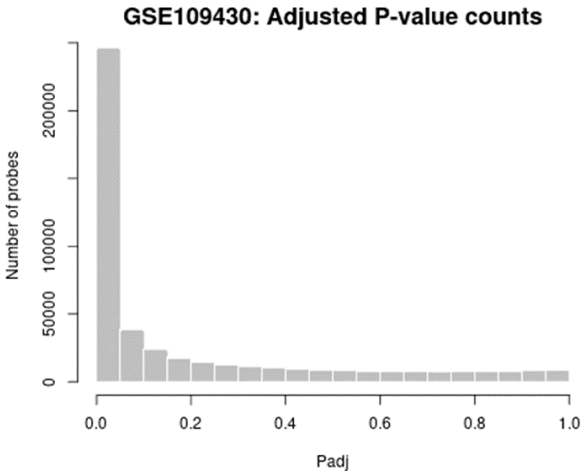

Fig. 5. Adjusted P value counts

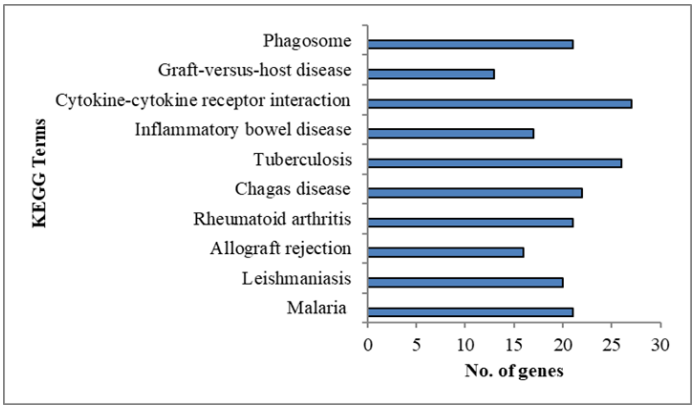

Fig. 8. KEGG pathway analysis

(This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution License citing the original author and source)

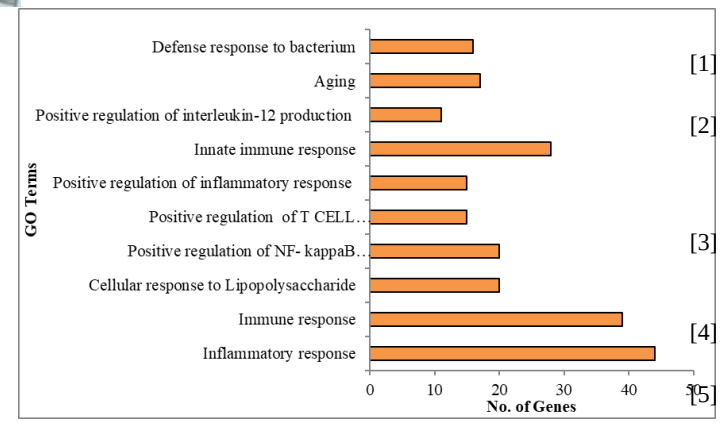

Fig. 9. GO pathway analysis

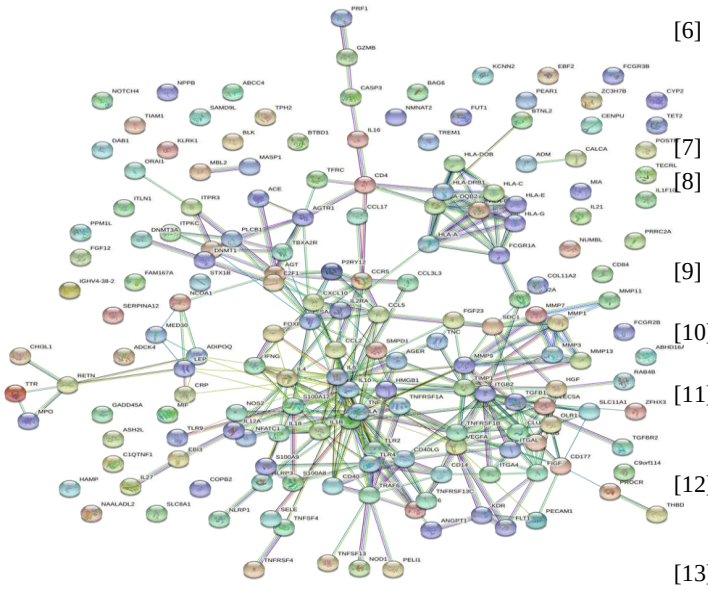

Fig. 10. PPI network

ACKNOWLEDGMENT

The authors express a deep sense of gratitude to the Department of Biotechnology, Department of Life sciences, Department of Computer science of Graphic Era (Deemed to be university) for all the support, assistance, constant encouragements to carry out this work.

REFERENCES

[1] Cox J R, & Sallis, R E, Recognition of Kawasaki disease, The Permanente Journal, 13(1), 57), 2009.

[2] Healthline-Kawasaki disease - https://www.healthline.com/health/kawasaki-dis-ase#:~:text=A%20rare%20but%20serious%20illness,of%20heart%20disease%20in%20children.

[3] Freeman, A, & Shulman, S T, Kawasaki disease: summary of the American Heart Association guidelines, American family physician, 74(7), 1141-1148, 2006.

[4] Leung, A K, Kawasaki disease, Common Problems in Ambulatory Pediatrics, 1, 385-390, 2012.

[5] Cheng, C, Hua, J, Tan, J, Qian, W, Zhang, L, & Hou, X, Identification of differentially expressed genes, associated functional terms pathways, and candidate diagnostic biomarkers in inflammatory bowel diseases by bioinformatics analysis, Experimental and therapeutic medicine, 18(1), 278-288, 2019.

[6] Clough, Emily, and Tanya Barrett. "The Gene Expression Omnibus Database." Methods in molecular biology (Clifton, N.J.) vol. 1418 (2016): 93-110. doi:10.1007/978-1-4939-3578-9_5

[7] DAVID homepage- https://david.ncifcrf.gov

[8] Ekins, S, Mestres, J, & Testa, B, In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling, British journal of pharmacology, 152(1), 9-20, 2007.

[9] Huerta, L, & Burke, M Functional genomics (II): Common technologies and data analysis methods.

[10] Nature Protocols 2009; 4(1):44 & Nucleic Acids Res.;37(1):1,2009.

[11] Clough, E, & Barrett, T, The gene expression omnibus database, In Statistical genomics Humana Press, New York, pp, 93-110, 2016.

[12] Kawasaki disease Jane W NewburgerMD, MasatoTakahashiMD Jane C BurnsMD, Journal of the American College of Cardiology, 67(14), 1738-1749, 2019.

[13] Szklarczyk, D, Morris, J H, Cook, H, Kuhn, M, Wyder, S, Simonovic, M., ... & Von Mering, C, The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic acids research, gkw937), 2017.

[14] NCBI, Home - GEO - NCBI (nih.gov)

[15] STRING database homepage; version 11.0; http://string-db.org/cgi/input.pl.

(This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution License citing the original author and source)