



Data article

KDmarkers: A biomarker database for investigating epigenetic methylation and gene expression levels in Kawasaki disease



Wei-Sheng Wu^{a,1}, Tzu-Hsien Yang^{b,1}, Kuang-Den Chen^c, Po-Heng Lin^a, Guan-Ru Chen^a,
Ho-Chang Kuo^{c,d,e,f,g,*}

^a Department of Electrical Engineering, National Cheng Kung University, University Road, 701 Tainan, Taiwan

^b Department of Information Management, National University of Kaohsiung, Kaohsiung University Rd, 811 Kaohsiung, Taiwan

^c Institute for Translational Research in Biomedicine, Liver Transplantation Center and Department of Surgery, Kaohsiung Chang Gung Memorial Hospital and Chang Gung University College of Medicine, Kaohsiung 83301, Taiwan

^d Department of Pediatrics, Kaohsiung Chang Gung Memorial Hospital and Chang Gung University College of Medicine, Kaohsiung 83301, Taiwan

^e Kawasaki Disease Center, Kaohsiung Chang Gung Memorial Hospital, Kaohsiung 83301, Taiwan

^f Department of Respiratory Therapy, Kaohsiung Chang Gung Memorial Hospital, Kaohsiung 83301, Taiwan

^g College of Medicine, Chang Gung University, Taoyuan City 33302, Taiwan

ARTICLE INFO

Article history:

Received 25 September 2021

Received in revised form 16 February 2022

Accepted 28 February 2022

Available online 10 March 2022

Keywords:

Kawasaki disease

DNA methylation

Gene expression

ABSTRACT

Kawasaki disease (KD) is a form of acute systemic vasculitis that primarily affects children and has become the most common cause of acquired heart disease. While the etiopathogenesis of KD remains unknown, the diagnostic criteria of KD have been well established. Nevertheless, the diagnosis of KD is currently based on subjective clinical symptoms, and no molecular biomarker is yet available. We have previously performed and combined methylation array (Illumina HumanMethylation450 BeadChip) and transcriptome array (Affymetrix GeneChip Human Transcriptome Array 2.0) to identify genes that are differentially methylated/expressed in KD patients compared with control subjects. We have found that decreased methylation levels combined with elevated gene expression can indicate genes (e.g., toll-like receptors and CD177) involved in the disease mechanisms of KD. In this study, we constructed a database called KDmarkers to allow researchers to access these valuable potential KD biomarkers identified via methylation array and transcriptome array. KDmarkers provides three search modes. First, users can search genes differentially methylated and/or differentially expressed in KD patients compared with control subjects. Second, users can check the KD patient groups in which a given gene is differentially methylated and/or differentially expressed. Third, users can explore the DNA methylation levels and gene expression levels in all samples (KD patients and controls) for a particular gene of interest. We further demonstrated that the results in KDmarkers are strongly associated with KD immune responses. All analysis results can be downloaded for downstream experimental designs. KDmarkers is available online at <https://cosbi.ee.ncku.edu.tw/KDmarkers/>.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Kawasaki disease (KD) is an acute febrile systemic vasculitis that was first described by Kawasaki *et al.* in 1974 [1]. It mainly

* Corresponding author at: Kawasaki Disease Center and Department of Pediatrics, Kaohsiung Chang Gung Memorial Hospital and Chang Gung University College of Medicine, Kaohsiung 83301, Taiwan.

E-mail addresses: wessonwu@mail.ncku.edu.tw (W.-S. Wu), thyangza1025@nuk.edu.tw (T.-H. Yang), dennis8857@gmail.com (K.-D. Chen), bb932156bb@gmail.com (P.-H. Lin), asd02138787@gmail.com (G.-R. Chen), dr.hckuo@gmail.com (H.-C. Kuo).

¹ Co-first authors.

affects children younger than five years old, especially in Asian countries. In developed countries, it is the leading cause of acquired heart diseases in children. The most severe complication of KD is the occurrence of coronary artery lesions (CAL), including coronary artery dilatation, coronary artery aneurysm, and coronary artery fistula formation [2]. These symptoms may lead to myocardial infarction, ischemic heart disease, or even sudden death. Therefore, broader attention has been given to KD in recent years.

Although the clinical features of KD are recognizable, its underlying immuno-pathogenic mechanisms are still under investigation, particularly with regard to CAL development. Diagnosis of KD is currently based on subjective clinical symptoms. These diag-

<https://doi.org/10.1016/j.csbj.2022.02.032>

2001-0370/© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

nosis rules include (1) fever for more than five days and (2) the occurrence of four of the five major clinical symptoms (diffuse mucosal inflammation, bilateral conjunctivitis, neck lymphadenopathy, indurative edema of the hands and feet associated with fingertip peeling, and polymorphic skin rashes). However, no molecular diagnosis kit is currently available for KD. We have recently found that intravenous immunoglobulin (IVIG) treatment can significantly alter the methylation landscape and expression levels of specific genes in KD patients. For example, in gamma Fc region receptor II-a, toll-like receptors, matrix metalloproteinase 9, macrophage marker expression profiles, CD177, S100A gene family, and β -catenin, IVIG administration considerably increases the gene methylation levels in KD [3–10]. These findings provide epigenetic evidence regarding KD molecular markers. To search for KD molecular markers, we further conducted experiments using Illumina Human Methylation450 (M450K) BeadChip and Affymetrix GeneChip Human Transcriptome Array 2.0 (HTA 2.0) in order to assess the genome-wide methylation patterns of known CpG islands and gene expression from the blood samples of KD patients and control subjects. In general, the DNA CpG methylation level alteration indicates a corresponding gene expression change in the opposite direction [11]. By considering the differential methylation levels and expression fold changes between KD patients and fever control subjects, we identified the up-regulation of NLRC4, NLRP12, and IL-1 β in KD [12]. All these findings suggest the possibility of genome-wide investigation for potential molecular markers of KD.

To provide a more panoramic view for the epigenetic methylation and expression analysis for KD, we constructed the KDmarkers database based on the M450K BeadChip and HTA 2.0 GeneChip experiments conducted. The main focus of the KDmarkers database is to boost the comprehensive examination for potential molecular KD biomarkers. We first performed suitable data processing for the methylation chip experiments and the genome-wide microarray on the blood samples collected from KD patients and control subjects. Then an easy-to-use KD marker investigation database interface was constructed to deposit these genome-wide gene expression data and CpG methylation patterns. Researchers can search for genes with differential epigenetic methylation patterns and/or differential gene expression levels between KD patients and control subjects. These extracted genes can serve as potential KD biomarkers. Three functions were implemented in KDmarkers to facilitate researchers and clinicians's investigation of KD-related genes: (i) Search for KD gene markers; (ii) Search potentially affected KD patient groups for a given gene; (iii) Raw data visualization of the methylation and expression experiment results of patient blood samples. Elevated methylation levels combined with decreased mRNA expression measurement can help indicate functional genes in KD diagnosis, treatment, and prognostic studies. We further demonstrated that some of the results provided by KDmarkers are closely related to KD immune responses. We believe that KDmarkers can provide novel methylation and gene expression array data in KD molecular mechanism research. KDmarkers is available at <https://cosbi.ee.ncku.edu.tw/KDmarkers/>.

2. Materials and methods

In this research, we constructed the KDmarkers database to record the genome-wide methylation and expression probing experiments from the blood samples of KD patients and control fever subjects. The construction of KDmarkers can be divided into 3 different parts (Fig. 1): (I) Patient sample collection; (II) Methylation and expression experimental data collection; and (III) Data processing and differential value comparison of patient groups.

These three parts are further elucidated in the following subsections.

2.1. Collection of patient blood samples

The KD patients' genome-wide methylation data and mRNA expression data were measured from their whole blood samples. The diagnosis of KD patients conformed to the American Heart Association KD criteria [13]. Patients diagnosed with KD were given a high-dose intravenous immunoglobulin (IVIG) treatment (2 g/kg) over 12 h. We refer to those patients within 24 h before IVIG treatment as pre-IVIG KD patients and those recruited at least 3 weeks after IVIG administration as post-IVIG patients. The blood samples used for genome-wide methylation level probing were collected from 24 pre-IVIG KD patients, 12 post-IVIG KD patients, 12 healthy children, and 12 fever control subjects. The blood samples used for expression level analysis were gathered from 50 pre-IVIG KD patients, 18 post-IVIG KD patients, 18 healthy children, and 18 fever control subjects. Post-IVIG treatment KD samples were collected from the same pre-IVIG KD patients. The fever control group in DNA methylation experiments consisted of child fever diagnosed to be caused by acute gastroenteritis, acute otitis media, allergic purpura, acute bronchitis, acute bronchiolitis, acute tonsillitis, or acute gastritis. The fever patient samples used in the expression analysis were diagnosed to result from acute enteritis, urinary tract infection, bronchopneumonia, volume depletion, acute bronchitis, acute bronchiolitis, acute tonsillitis, or acute gastritis. All of the sample symptoms were ensured to be unrelated to KD development. For KD patients treated with IVIG, no steroids, immunosuppressive agents or biologics were simultaneously prescribed. The demographic profiles of the patients recruited for the genome-wide methylation and expression microarrays are listed in Table 1 and Table 2, respectively. As summarized in Table 1 and Table 2, genders are matched within each patient group, and the overall ages of the recruited patients all fall around two years old. Finally, all patients involved in the methylation microarray experiments and expression analysis are from the Taiwan Han Chinese population. Patient sample collection was approved by the Institutional Review Board of Chang Gung Memorial Hospital (IRB No.:101-4618A3 and 101-0680A3). Written informed consent from the parents/guardians of the patients was also obtained. Collectively, the samples can be categorized into four groups: normal controls, fever controls, KD samples before IVIG treatment (pre-IVIG patient group), and KD samples after IVIG treatment (post-IVIG patient group).

2.2. DNA methylation and gene expression probing

Patient blood samples were first enriched through white blood cell enrichment [4,14]. The samples gathered for genome-wide methylation profiling were then treated with bisulfite to convert non-methylated cytosine into uracil for methylated DNA extraction. These DNA extracts underwent genomic DNA methylation probing using Illumina HumanMethylation450 (M450K) BeadChip, which contains around 450,000 human CpG marker probe beads at single-nucleotide resolution. For the samples collected for expression level measurement, we used an isolation kit (mirVana miRNA Isolation Kit, Catalog No. AM1560, Life Technologies, Carlsbad, CA) to extract the RNAs from the samples. The quality and quantity of the total RNAs for the samples gathered for expression level analysis were estimated by Bioanalyzer (ABI) and Qubit (Thermo) to ensure that the RIN values were larger than 7. Around six RNA samples of the same group were evenly pooled to obtain unbiased expression measurement, resulting in 3 healthy pools, 3 fever pools, 9 pre-IVIG pools, and 3 post-IVIG pools. These sample pools were hybridized on the GeneChip Human Transcriptome Array 2.0

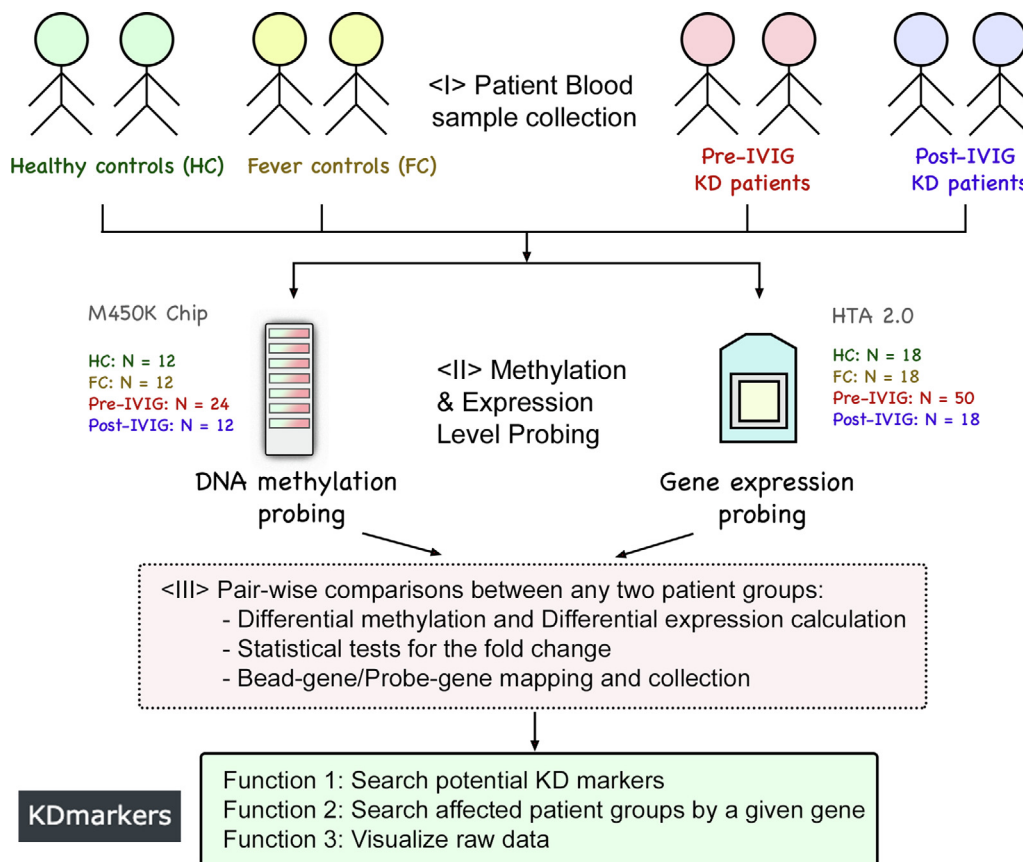


Fig. 1. The overall construction of the KDmarkers database. Pre-IVIG KD patients refer to KD patients within 24 h before IVIG treatment, and post-IVIG KD patients indicate KD patients at least 3 weeks after IVIG treatment.

Table 1

The demographic information of the patient cohorts (Taiwan Han Chinese population) used for the DNA methylation probing experiments collected in KDmarkers.

	Healthy Controls	Fever Controls	Pre-IVIG KD Patients	Post-IVIG KD Patients
Patient Number	N = 12	N = 12	N = 24	N = 12
Gender	6 male, 6 female	6 male, 6 female	13 male, 11 female	6 male, 6 female
Age	2.25 ± 0.2	1.93 ± 0.23	1.95 ± 0.31	1.52 ± 0.27

Table 2

The demographic information of the patient cohorts (Taiwan Han Chinese population) used for the expression level probing experiments gathered in KDmarkers.

	Healthy Controls	Fever Controls	Pre-IVIG KD Patients	Post-IVIG KD Patients
Patient Number	N = 18	N = 18	N = 50	N = 18
Gender	9 male, 9 female	8 male, 10 female	27 male, 23 female	9 male, 9 female
Age	2.46 ± 0.35	1.97 ± 0.26	1.36 ± 0.14	1.90 ± 0.32

(HTA 2.0, Affymetrix, Santa Clara). This microarray covers more than 285,000 full-length human transcripts. The microarray experiment quality control was subjected to the criteria proposed in previous studies [7]. The DNA methylation and expression microarray data were also deposited in NCBI GEO (GSE109430, GSE109351).

2.3. Identification of differentially methylated/expressed genes between different patient groups

Genomic methylation profiling for the four sample groups was measured using M450K BeadChip, as described above. The methylation level of one CpG probe bead is computed using the β value, which is defined by

$$\beta = \frac{M}{M + U + \alpha}$$

where M is the methylated cytosine intensity, U is the unmethylated cytosine probing intensity, and α is the calibration number. The raw data is processed using Partek. The gene expression profiling for pooled sample groups was measured with Affymetrix HTA 2.0 GeneChip. The mRNA expression levels of different human genes were also analyzed and calculated by the commercial microarray data analysis tool Partek. Default settings were used in the Partek analysis.

We compared the differential methylation/expression among the patient sample groups. β values and expression measurements are used to calculate the differential methylation level significance

and the differential expression analysis, respectively. The comparison was performed for every two sample groups from the five experimental patient sample groups (the healthy subject group, the fever patient group, the healthy and fever pooled patient controls, the pre-IVIG KD patient group, and the post-IVIG KD patient group). In KDmarkers, three statistical test results are available for users: the one-tailed independent *t*-test, the one-tailed rank-sum test, and the one-tailed Kolmogorov–Smirnov test. We calculated the one-sided *p*-value against the null hypothesis that the average/median/distribution of Group 1 is equal to the corresponding statistic of Group 2, respectively [15–17]. Multiple hypotheses biases in the analysis process were corrected using two different approaches: the FDR approach [18] and the Bonferroni method [19]. After calculating the differential methylation/expression of each bead/probe on M450K/HTA2.0, we used the Infinium annotation file/HTA annotation file to map the value reads back to the human genome. Notice that multiple read values may be mapped to one gene, and one bead read value may be annotated to multiple genes. This many-to-many relationship between beads/probes and genes comes from the BeadChip/GeneChip design in which many short segments were extracted from one gene/transcript for comprehensive genome-wide coverage.

2.4. Implementation of KDmarkers

The data pre-processing steps were carried out using the programming language Python (version 2.7.6 by Python Software Foundation, an open-source software). The KDmarkers database user interface is facilitated by JavaScript and the Python package Django (version 1.11.1 by Django Software Foundation, an open-source software). Data used in KDmarkers were deposited using the MySQL database management system (version 5.7.19 by Oracle Corporation, Santa Clara, California, U.S).

3. Results and Discussions

KDmarkers integrates methylation and expression data to provide functions for potential Kawasaki disease biomarker exploration. Three functionalities were implemented in KDmarkers to facilitate genome-wide KD marker investigation: (I) Function 1: Search KD Markers. Differential methylation patterns and expression levels of human genes can help evaluate their suitability as potential KD molecular diagnostic kits. (II) Function 2: Search Affected Patient Groups. Users can obtain the KD patient group pairs that bear differential expression or methylation levels for an input gene. (III) Function 3: Raw Data Visualization. Boxplots are drawn to visualize the gene expression or methylation levels among different patient groups. These three functions help researchers compare the expression and methylation levels between KD patients and the control subjects. In the database construction, abbreviations for the KD patient groups are given as follows: HC refers to the healthy controls, FC stands for the fever subjects, KD1 is the KD patients within 24 h before IVIG treatment, and KD2 represents the KD patient samples collected at least 3 weeks after IVIG administration.

3.1. Function 1: search KD markers

Users can find differentially methylated/expressed genes between KD patients and normal children within this function. In KDmarkers, any two of the five patient groups (healthy children, fever control subjects, pooling group for healthy and fever control subjects, pre-IVIG KD patients, and post-IVIG KD patients) can be selected to obtain the gene list with differential methylation levels and/or differential expression levels. First, the search conditions

and the comparison groups need to be specified (Fig. 2-a). After choosing the comparison groups, users can type in the fold change threshold, statistical test (one-tailed independent *t*-test, one-tailed rank-sum test, and one-tailed Kolmogorov–Smirnov test), and the significance level. Upon clicking the search button, the results are shortly queried out in a tabular format (Fig. 2-b). Bead IDs and their corresponding genes satisfying the specified differential methylation/expression levels between the specified two groups are listed in the table. A heat map of the methylation/expression levels between the two groups is also provided to visualize the individual sample measurements between the two groups (Fig. 2-c). In this function, genes that are differentially methylated, differentially expressed, or both differentially methylated and expressed are summarized in the generated tables. Users can click on the tabs to switch the table listing into differentially methylated genes, differentially expressed genes, and genes that are both differentially methylated and expressed (Fig. 2-d). A detail page of the bead probes and the measurement results can be found by clicking on the bead/probe ID or the detail link (Fig. 2-e). Moreover, the generated tables and heat maps can be downloaded by clicking the download button in their upper-right corners.

3.2. Function 2: search affected patient groups by a given gene

In this function, users can find potentially affected patient groups of a given gene. Users first input the gene of interest and the differential conditions (fold change threshold and the significance level) that the given gene should satisfy (Fig. 3-a). Then any two patient group pairs that conform to these differential conditions are extracted (Fig. 3-b). These patient group pairs are categorized into three genres: group pairs that satisfy the methylation differential thresholds, group pairs that meet the expression differential thresholds, and group pairs that tally all user-specified methylation and expression conditions. In the resulting patient group pairs, users can choose the patient group pair of interest. Then the measurement summary for the selected patient group pair can be investigated in a table and box-plot (Fig. 3-c). Users can click on the tabs to switch the table listing into differential methylation evidence and differential expression evidence (Fig. 3-d). Details of the bead probes/expression measurement results can be found by pressing the hyperlink on the bead/probe ID (Fig. 3-e).

3.3. The Detail page

After clicking on the methylation bead ID or the expression microarray probe ID, the detail page for the bead ID or the probe ID is presented. Here, we describe the detail page of a methylation bead as an example (Fig. 4). The detail page can be divided into three different parts. In the first part, the details of the selected bead in M450K BeadChip are shown, and the user-specified fold change threshold and the significance level are re-printed (Fig. 4-a). Then the average methylation levels of the selected bead and its corresponding mapped gene between the two specified patient groups, the statistical test *p*-value, and the fold change between these two patient groups are listed in the differential methylation summary table (Fig. 4-b). Measurements of the individual samples within these two groups are further visualized through a boxplot and a heatmap (Fig. 4-c). These visualization plots can be downloaded by clicking the button in the upper-right corner (Fig. 4-d). Similar contents are presented in the expression detail page.

3.4. Function 3: visualize raw data

Sometimes, users may need to query the raw methylation levels or the expression levels of a certain gene between different KD

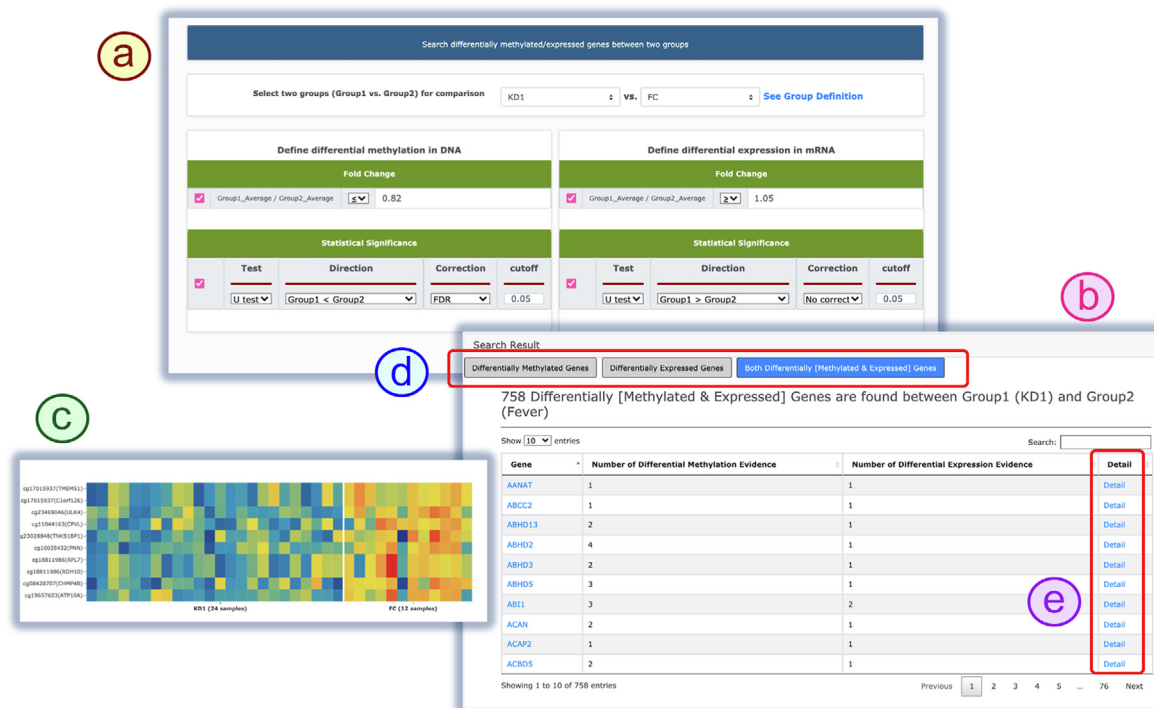


Fig. 2. Function 1 of KDmarkers. Users can search potential KD biomarkers through specified differential methylation and/or expression levels. (a) Users need to specify two sample groups and the search conditions to search for differentially methylated/expressed genes between these two groups. (b) Differentially methylated/expressed genes between the selected two groups. (c) Heat map of the methylation/expression levels between the two groups. (d) Users can click on the tabs to switch the table listing into differentially methylated genes, differentially expressed genes, and genes that are both differentially methylated and expressed. (e) A detail page containing the experimental results of the bead/probe/gene can be found by clicking the bead/probe ID or the detail link.

patient groups. Therefore, we implemented a raw data visualization function in KDmarkers. In this function, users can input the gene of interest and obtain the raw measurements of the methylation/expression levels of this gene among different sample groups. Users first key in the gene of interest and click the search button to initiate the query (Fig. 5-a). Then the listing of the CpG beads/expression probes mapped to this gene are tabulated (Fig. 5-b). Detailed information on the chip bead/probe can be viewed by clicking on the bead/probe ID. Comparison visualization boxplots among different patient groups for this gene are also generated (Fig. 5-c). In addition, the ANOVA test *p*-value is provided to indicate whether the group expression or methylation levels are statistically different from one another for this gene. All these raw methylation levels, expression measurements, and the comparison boxplots for this gene can be downloaded by either clicking the download button on the upper-right corner (Fig. 5-d) or through the download page.

3.5. A walk-through example

We have previously reported that human toll-like receptor 6 (TLR6) shows hypomethylation and upregulation in KD patients [7]. We use TLR6 here as a walk-through example for using KDmarkers. In Function 3 (visualize raw data) of KDmarkers, using TLR6 as the input gene, we can obtain the output data showing 11 CpG sites with methylation group difference when comparing the methylation levels among different KD patient groups (See Fig. 5-a, 5-b). For example, the cg02221520 bead shows significantly (one-way ANOVA $p = 7.9E-8$) lower β values in pre-IVIG KD patients than healthy control subjects and post-IVIG KD patients (See Fig. 6-a). The gene expression array (TC04002917) also reveals a significant expression difference (one-way ANOVA $p = 6.6E-3$) among different patient groups for TLR6 (See Fig. 6-b).

Furthermore, the differences in the methylation array probing results and expression array values go in the opposite direction, showing higher methylation levels and higher expression levels in KD patients before IVIG treatment. We can further confirm the pair-wise group comparisons through Function 2 (search affected patient groups) of KDmarkers (Fig. 3-a). As shown in the query result table of Function 2, TLR6 has five CpG sites with significantly lower β values (FDR threshold = 0.05) in pre-IVIG KD patients than healthy/fever controls (See Fig. 3-c). The pre-IVIG KD patients also demonstrate 4 CpG sites with significantly lower β values in TLR6 than the KD patients after IVIG treatment. Similarly, in the pair-wise differential expression evidence of Function 2, the pre-IVIG KD sample group reveals significantly higher expression in TLR6 than the healthy/fever control group and the post-IVIG KD sample group. From these results, we can infer that TLR6 may trigger the immunopathogenesis of KD, thus serving as a potential molecular marker for KD diagnosis.

3.6. An example analysis of the results deposited in KDmarkers

KDmarkers can provide insightful information for KD researchers. To search possible KD markers, we set up the β fold change threshold of 0.82 and expression fold change of 1.05 (Mann-Whitney U/rank-sum test with $FDR \leq 0.05$) between KD patients before IVIG treatment (pre-IVIG KD patient group) and fever control subjects in Function 1 of KDmarkers. These thresholds were selected to achieve the top 10% cutoff fold changes in the analysis. The methylation level and expression comparison results between KD patients before IVIG treatment and fever control subjects extract a total of 758 differentially methylated and expressed genes. We further used the DAVID program [20] to perform functional enrichment analysis on these differentially expressed genes (DEGs). The Gene Ontology categorization showed that the vast majority of

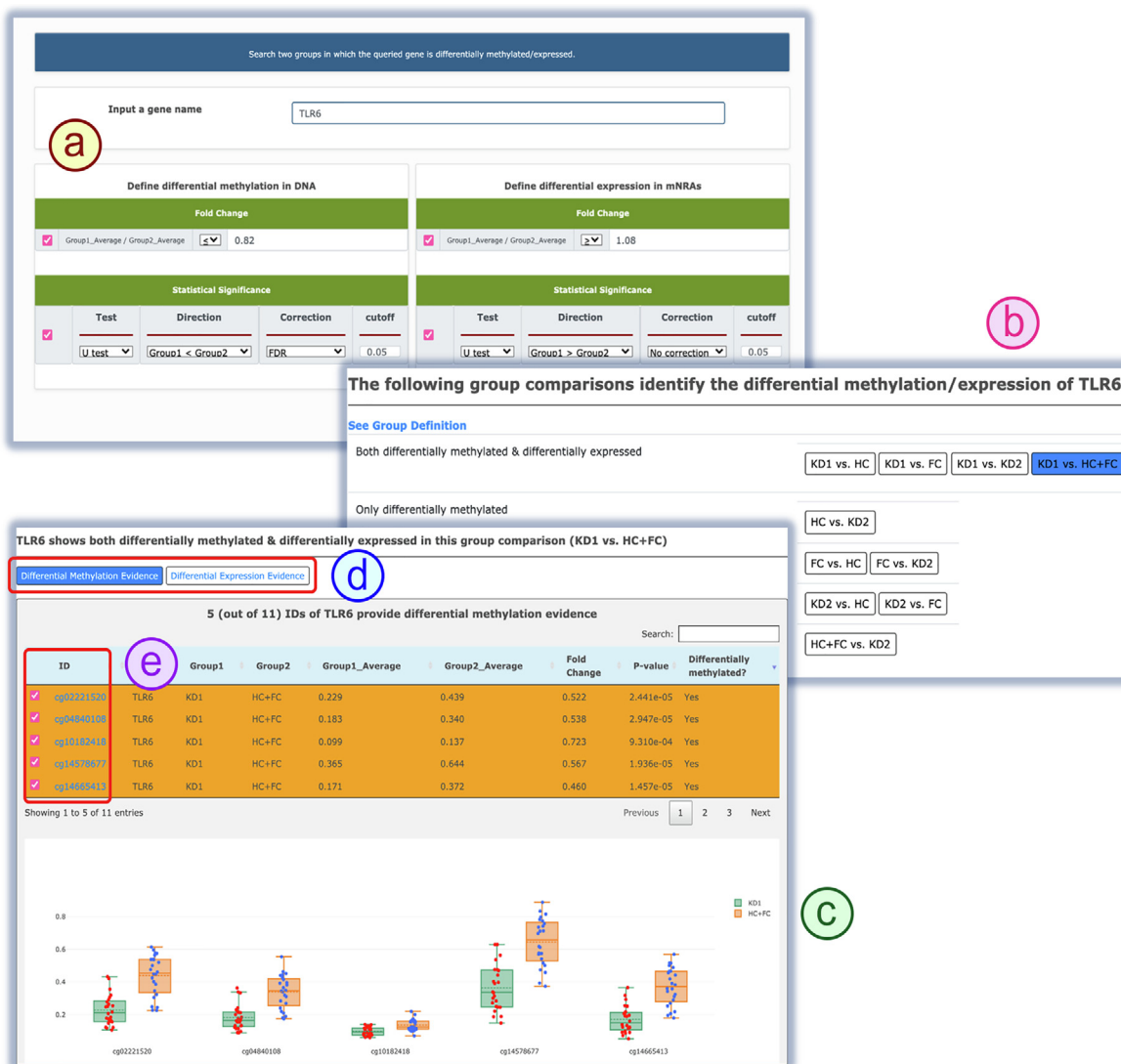


Fig. 3. Function 2 of KDmarkers. Users can search the affected patient groups by the differential methylation and/or expression of a given gene. (a) Users can type in the gene of interest and the search conditions to find potentially affected patient groups. (b) Patient group pairs that bear differential methylation/expression levels on the given gene. (c) The list and the boxplot of the experimental results of the gene between the selected two groups. (d) Users can click the tabs to switch the table and boxplot between differential methylation evidence and differential expression evidence. (e) The detail page of the results for the bead/probe can be found by clicking the bead/probe ID.

these genes are encoded for a series of membrane-associated proteins in the lipid bilayers of plasma, lysosomal, endosomal, peroxisomal membranes, membrane-bound Golgi apparatus, endoplasmic reticulum, phagocytic vesicle, and extracellular exosome (See Fig. 7). Moreover, the KEGG pathway analysis reveals that these DEGs are significantly related to signaling pathways of mitogen-activated protein kinase (MAPK), tumor necrosis factor (TNF), toll-like receptor, Foxo, Nod-like receptor, and endocytosis. By performing INTERPRO enrichment analysis, the 758 differentially methylated and expressed genes also bear patterns in the protein kinase-like domain, serine/threonine-protein kinase, and toll/interleukin 1 receptor homology (TIR) domain. Finally, the top-ranked enriched UniProtKB keywords specific to the 758 DEGs are alternative splicing, phosphoprotein, membrane, transferase, endosome, apoptosis, and innate immunity. Taking these enrichment analyses altogether, these up-regulated genes with significant hypomethylation in KD patients have annotation results that indicate a strong association with membrane-associated proteins. Further, these genes are prone to get involved in kinase and transferase activities for MAPK/TNF/Foxo signaling,

receptors/modifiers for inflammatory responses, and cell death programs controlled via alternatively spliced variants.

Research has shown that alternative splicing plays a crucial role in regulating the innate immune response. The innate immune response is activated by infectious challenges to provide diversity within the proteome of immune cells [21,22]. Approximately one-fifth of the genes expressed in dendritic cells undergo alternative splicing in response to bacterial challenges and participate directly in antimicrobial defense [23]. Here we use the TLR gene family, which includes parts of the 758 genes, as an example. It has been reported that TLR signaling pathways are subjected to extensive post-transcriptional regulation, in which more than 250 alternatively processed transcripts encode variants of receptors, adaptors, modifiers, and signaling regulators [24–29]. Each human TLR gene has numerous spliced variants within immune cells. From this detailed analysis of the 758 genes extracted by KDmarkers, these gene targets provide excellent potential markers for further Kawasaki disease immune regulation investigation. Overall, we conclude that KDmarkers can help suggest possible genes that help identify novel

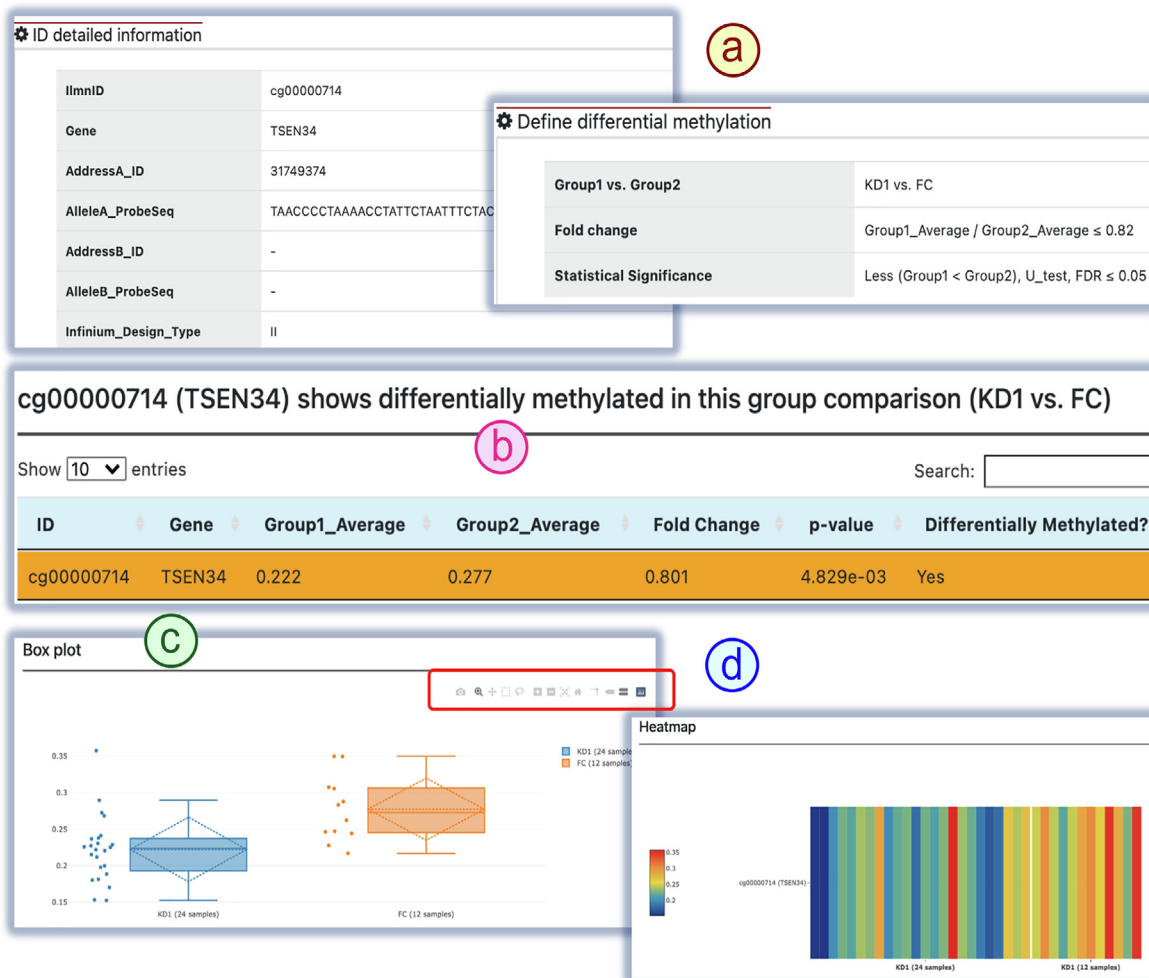


Fig. 4. Detail pages in KDmarkers. The detailed methylation information of the selected bead can be viewed in the detail page. The expression detail page contains similar results to the methylation detail page. (a) Information of the selected bead ID, the specified fold change threshold, and the significance level. (b) The measurement results between the specified two patient groups. (c) The boxplot and heatmap visualization for the methylation levels of individual samples in each of the specified patient groups. (d) The heatmap and the boxplot can be downloaded via the buttons in the upper-right corner.

diagnostic targets for the therapeutic development of Kawasaki disease treatment.

3.7. Case study

KD patients are usually at risk of cardiovascular complications that can cause morbidity and mortality [30]. Several molecules are now known to be closely related to cardiovascular diseases. We take the matrix metalloproteinase 9 (MMP9) gene as a case study example to elucidate the applicability of KDmarkers. Researchers have discovered that MMPs may assist with the maintenance of the structures and functions of coronary vascular walls [31]. Among the MMP protein family, MMP9 is notorious for triggering coronary artery aneurysms [32], leading to 20% of coronary artery aneurysms in children suffering from these symptoms if not properly treated [33]. We investigated the results of MMP9 among the healthy control group, KD patients before IVIG treatment, and KD blood samples after IVIG treatment using KDmarkers. According to KDmarkers, the gene methylation level of MMP9 in KD patients is significantly lower than the level of the healthy group (Fig. 8-a). Furthermore, MMP9 reveals higher expression in KD patients than in the healthy control group (Fig. 8-b). After IVIG treatment, the expression level of MMP9 returns to the expected level of the healthy group while its methylation level increases

(Fig. 8-a and 8-b). By combining these results, epigenetic hypomethylation and up-regulation of MMP9 can be observed in KD patients, indicating the potential molecular diagnosis of KD. And in previous real-time quantification polymerase chain reaction (RT-qPCR) experiments on a different cohort of patients, the mRNA levels of MMP9 were indeed verified to be higher in the white blood cells of KD patients and to be closely related to coronary artery lesion formation [10]. From this example, we demonstrated the bio-medical applicability of KDmarkers in KD biomarker investigation.

3.8. Comparison with related analysis works on Kawasaki disease

We developed KDmarkers to facilitate researchers to investigate potential KD biomarkers. Currently, two works related to KD molecular research are available in the literature: one database created by Park *et al.* [19] and one meta-analysis study carried out by Chang *et al.* [34]. Since these two works targeted similar aims as KDmarkers, we have compared them with KDmarkers in this subsection.

Park *et al.* [19] reported the first KD database, called Kawasaki Disease Database (KDD), which includes 1,292 patient cases and 1,283 genomic samples from 449 KD Korean patients from 2016. KDD includes clinical data of IVIG resistance, CAL formation, labo-

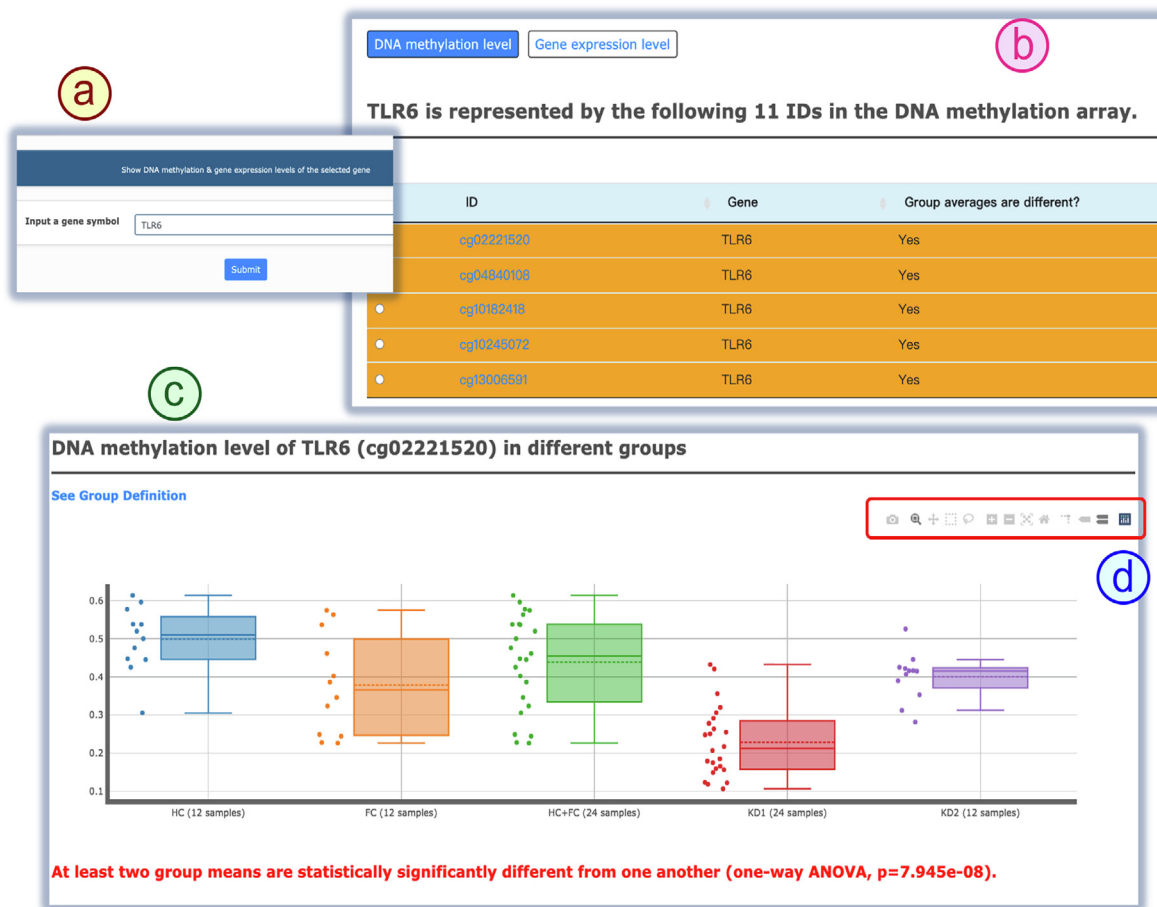


Fig. 5. Function 3 of KDmarkers: raw data visualization. (a) The input form of Function 3. (b) The listing of the beads/probes mapped to this gene. (c) A comparison visualization on the methylation/expression levels of the selected bead. (d) The comparison boxplot can be downloaded via the upper-right corner button.

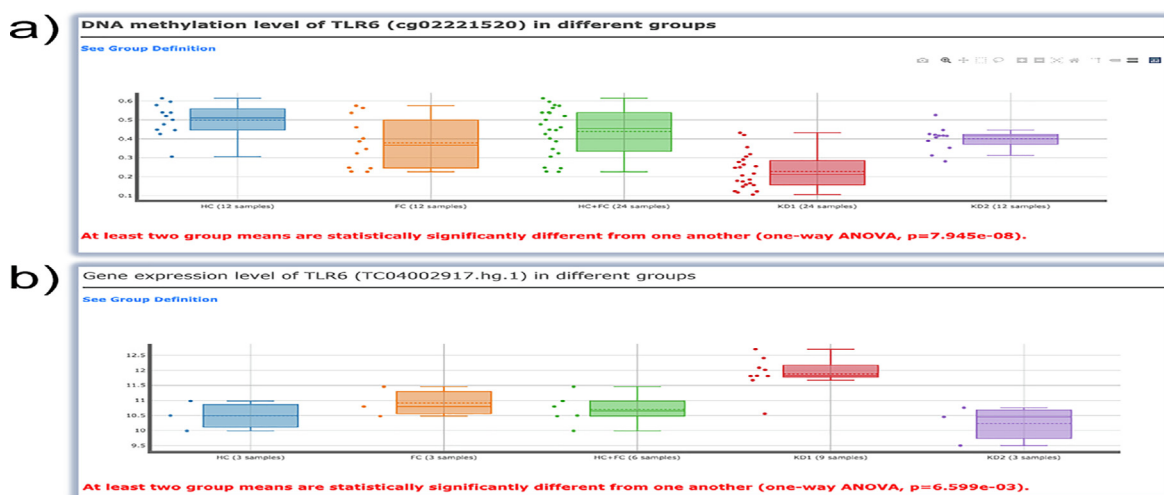


Fig. 6. (a)/(b) The methylation/expression levels of TLR6 among KD patient groups using Function 3 of KDmarkers. HC refers to healthy controls, FC stands for fever subjects, KD1 is the KD patients before IVIG treatment, and KD2 represents the KD samples collected at least 3 weeks after IVIG administration.

ratory data, genomic DNA data, plasma sample data, and KD genotype data. The KDD database helps researchers analyze the effects resulting from different KD genotypes, clinical lab data, and disease outcomes. However, KDD does not provide epigenetic and expression data among KD patients before and after IVIG treatment.

Therefore, it can not offer any aids for KD prognostic and diagnostic biomarker investigation.

The previous meta-analysis study carried out by Chang *et al.* [34] utilizes a combination of DNA methylation and gene expression data. In their analysis, two up-regulated and hypomethylated

Category	ID	Term	Count	P-value	FDR
GOTERM_CC_DIRECT	GO:0016020	membrane	126	3.82E-06	5.65E-04
GOTERM_CC_DIRECT	GO:0005765	lysosomal membrane	28	7.85E-06	8.34E-04
GOTERM_CC_DIRECT	GO:0010008	endosome membrane	22	9.39E-06	8.34E-04
GOTERM_CC_DIRECT	GO:0005794	Golgi apparatus	60	1.34E-05	9.89E-04
GOTERM_BP_DIRECT	GO:0006954	inflammatory response	38	6.37E-07	9.99E-04
GOTERM_CC_DIRECT	GO:0070062	extracellular exosome	149	2.71E-05	1.72E-03
GOTERM_CC_DIRECT	GO:0005886	plasma membrane	200	1.64E-04	8.09E-03
GOTERM_CC_DIRECT	GO:0005925	focal adhesion	31	2.75E-04	1.22E-02
GOTERM_CC_DIRECT	GO:0016324	apical plasma membrane	25	3.91E-04	1.58E-02
GOTERM_CC_DIRECT	GO:0005769	early endosome	21	5.68E-04	2.10E-02
GOTERM_CC_DIRECT	GO:0005778	peroxisomal membrane	9	1.00E-03	2.96E-02
GOTERM_CC_DIRECT	GO:0005783	endoplasmic reticulum	51	1.09E-03	3.03E-02
GOTERM_CC_DIRECT	GO:0030136	clathrin-coated vesicle	9	1.61E-03	4.21E-02
GOTERM_CC_DIRECT	GO:0030670	phagocytic vesicle membrane	9	1.80E-03	4.45E-02
INTERPRO	IPR011009	Protein kinase-like domain	49	4.16E-08	5.10E-05
INTERPRO	IPR008271	Serine/threonine-protein kinase, active site	29	3.18E-05	1.30E-02
INTERPRO	IPR000157	Toll/interleukin-1 receptor homology (TIR) domain	8	4.28E-05	1.31E-02
KEGG_PATHWAY	hsa04010	MAPK signaling pathway	29	2.00E-05	2.97E-03
KEGG_PATHWAY	hsa04668	TNF signaling pathway	17	3.66E-05	2.97E-03
KEGG_PATHWAY	hsa04620	Toll-like receptor signaling pathway	16	1.22E-04	3.09E-03
KEGG_PATHWAY	hsa04068	FoxO signaling pathway	17	5.38E-04	9.50E-03
KEGG_PATHWAY	hsa04621	NOD-like receptor signaling pathway	10	1.11E-03	1.69E-02
KEGG_PATHWAY	hsa04144	Endocytosis	23	2.20E-03	2.91E-02
UP_KEYWORDS		Alternative splicing	506	2.34E-21	8.60E-19
UP_KEYWORDS		Phosphoprotein	412	2.48E-18	4.55E-16
UP_KEYWORDS		Membrane	360	6.06E-12	7.41E-10
UP_KEYWORDS		Transferase	113	2.99E-10	2.75E-08
UP_KEYWORDS		Endosome	48	6.07E-10	4.46E-08
UP_KEYWORDS		Apoptosis	41	1.22E-05	2.80E-04
UP_KEYWORDS		Innate immunity	25	2.70E-05	5.50E-04
UP_KEYWORDS		Inflammatory response	18	3.51E-05	6.14E-04
UP_KEYWORDS		Metal-binding	163	1.79E-03	1.73E-02
UP_KEYWORDS		Acetylation	155	1.52E-03	1.51E-02
UP_KEYWORDS		Coiled coil	145	2.51E-04	3.07E-03
UP_KEYWORDS		Disease mutation	127	1.21E-04	1.78E-03
UP_KEYWORDS		Ubl conjugation	88	5.50E-04	5.93E-03
UP_KEYWORDS		Lysosome	22	5.79E-04	6.08E-03
UP_SEQ_FEATURE		sequence variant	511	3.14E-05	1.06E-02
UP_SEQ_FEATURE		splice variant	361	1.03E-08	2.45E-05

Fig. 7. DAVID identifications of enriched GO terms, InterPro classes, KEGG pathways, and UniProtKB keywords for the 758 up-regulated genes with significant hypomethylation in KD are tabulated. P-values here indicate the thresholds of EASE Scores, or modified Fisher Exact *p*-values, for gene-enrichment analysis. FDR in DAVID requests adaptive linear step-up to adjust *p*-values to control the false discovery rate against multiple hypotheses biases.

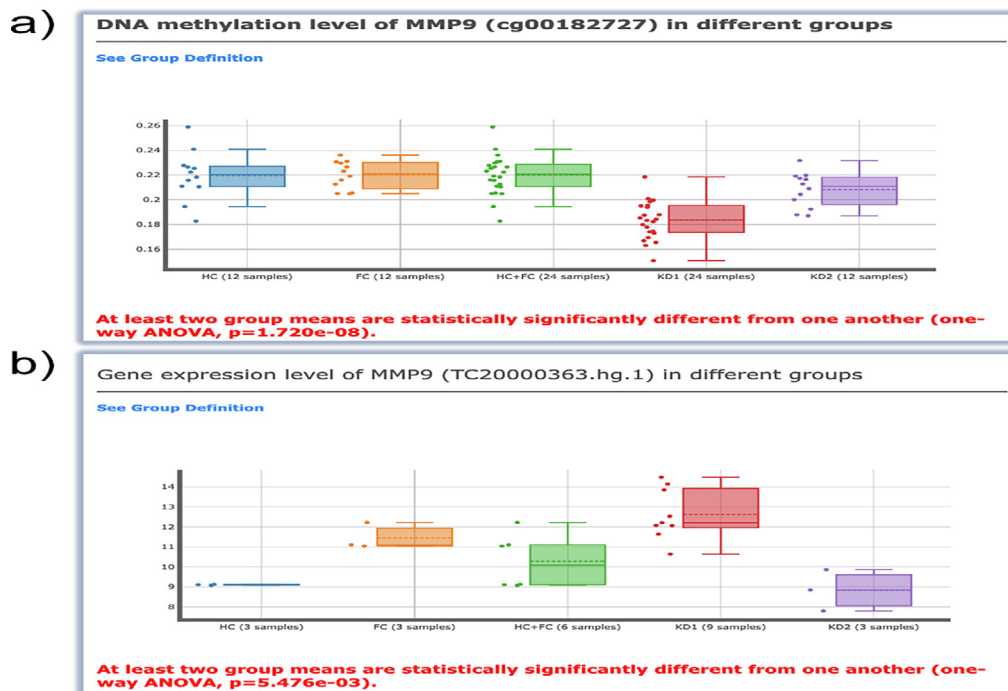


Fig. 8. The methylation/expression level comparison of MMP9 among KD patient groups. (a) The gene methylation level of MMP9 in KD patients is significantly lower than the level of the healthy group. (b) MMP9 reveals higher expression in KD patients than in the healthy control group. After IVIG treatment, the expression level of MMP9 returns to the expected level of the healthy group while its methylation level increases. HC refers to healthy controls, FC stands for fever subjects, KD1 is the KD patients before IVIG treatment, and KD2 represents the KD samples collected at least 3 weeks after IVIG administration.

genes, *MAPK14* and *PHLPP1*, were identified in the acute phase of KD patients using both the gathered DNA methylation and gene expression datasets. Chang *et al.* further highlighted that these two genes are associated with innate immunity. However, this analysis still has weaknesses. First, the gathered gene expression microarray datasets and DNA methylation datasets in the meta-analysis are ethnically diverse. Chang *et al.* downloaded the DNA methylation data and the expression data from our previous study

[5] and the work of Jaggi *et al.* [35], respectively. By further investigating the demographic information, the DNA methylation microarrays were performed on samples from the Taiwan Han Chinese population, while the expression probing experiments were conducted for patients from White, Hispanic, and other races. While datasets with a mix of racial groups might be beneficial for finding common molecular mechanisms involved in KD pathogenesis, the mechanisms underlying the more frequent incidence

of KD in children of Asian or Pacific Island ancestry may be excluded. Second, the meta-analysis results by Chang *et al.* are not readily available. Further, no integrated platform has been built for easy comparison investigation on their results, which can hinder advanced analysis of the results [36]. In contrast, we constructed the KDmarkers database for researchers to compare different patient groups and extract diagnosis candidate markers. In addition, our platform is based on DNA methylation and gene expression datasets of the same race (Taiwan population). We also verified that KDmarkers provides consistent results with the previously identified 14 hub genes (genes that are both differentially methylated and expressed) provided in Chang's study.

In summary, KDmarkers is the first platform established based on the methylation array and the gene expression microarray for functional genetic analysis. Researchers can investigate the functional effect of a specific gene by comparing its methylation and expression levels among different KD patient groups (normal controls, fever subjects, the pre-IVIG patient group, and the post-IVIG patient group) at the same time to extract possible KD biomarkers. In summary, KDmarkers broadens the genomic investigation facility for the study of Kawasaki disease and can better benefit the community.

4. Conclusions

In this research, we performed genome-wide methylation level and expression level experiments for four different groups of Kawasaki disease (KD) patients (healthy controls, fever control subjects, pre-IVIG KD patients, and post-IVIG KD patients) and constructed the KDmarkers database. KDmarkers aims to provide potential KD biomarkers to supplement the current subjective KD clinical symptom diagnosis criteria. Comprehensive data analysis was conducted on the methylation array and the gene expression probing results using the blood samples gathered from KD patients. Based on these results, three main functions were implemented in KDmarkers. Users can find those differentially methylated and/or differentially expressed genes among different KD patient groups. Furthermore, the KD patient group pairs that carry differential methylation levels and/or differential expression levels can be identified. Finally, the experimental results of the methylation array and the gene expression array for a given gene can be explored in KDmarkers. Compared with the existing tool KDD and the analysis results of Chang *et al.*, KDmarkers can additionally provide differential genetic clues for identifying KD diagnosis and molecular prognosis kits. We believe that KDmarkers paves the way for KD-related molecular basis identification and potentially sheds light on the cure for KD.

Data Availability

KDmarkers and the related datasets are freely available at <https://cosbi.ee.ncku.edu.tw/KDmarkers/>.

CRedit authorship contribution statement

Wei-Sheng Wu: Conceptualization, Investigation, Supervision, Project administration, Visualization, Writing - original draft, Writing - review & editing. **Tzu-Hsien Yang:** Investigation, Visualization, Writing - original draft, Writing - review & editing. **Kuang-Den Chen:** Investigation, Writing - original draft, Writing - review & editing. **Po-Heng Lin:** Software, Visualization. **Guan-Ru Chen:** Software, Visualization. **Ho-Chang Kuo:** Conceptualization, Supervision, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National University of Kaohsiung, National Cheng Kung University, the Ministry of Science and Technology of Taiwan [MOST 107-2218-E-390-009-MY3, MOST 110-2222-E-390-001, MOST 107-2221-E-006-225-MY3, MOST 108-2628-E-006-004-MY3, MOST 108-2314-B-182-037-MY3, and MOST 110-2221-E-006-198-MY3], and Chang Gung Memorial Hospital [CORPG8L0321, CMRPG8L0021, CMRPG8L1241, CORPG8M0091, and CPRPG8H0051-2]. Conflict of interest statement: None declared.

References

- [1] Kawasaki T, Kosaki F, Okawa S, Shigematsu I, Yanagawa H. A new infantile acute febrile mucocutaneous lymph node syndrome (MLNS) prevailing in Japan. *Pediatrics* 1974;54(3):271–6.
- [2] Kuo H-C. Preventing coronary artery lesions in Kawasaki disease. *Biomed J* 2017;40(3):141–6.
- [3] Kuo H-C, Chang J-C, Kuo H-C, Yu H-R, Wang C-L, Lee C-P, Huang L-T, Yang KD. Identification of an association between genomic hypomethylation of FCGR2A and susceptibility to Kawasaki disease and intravenous immunoglobulin resistance by DNA methylation array. *Arthritis Rheumatol* 2015;67(3):828–36.
- [4] Li S-C, Chan W-C, Huang Y-H, Guo MM-H, Yu H-R, Huang F-C, Kuo H-C, Kuo H-C. Major methylation alterations on the CpG markers of inflammatory immune associated genes after IVIG treatment in Kawasaki disease. *BMC Med Genomics* 2016;9(1):81–8.
- [5] Chen K-D, Huang Y-H, Guo MM-H, Lin T-Y, Weng W-T, Yang H-J, Yang KD, Kuo H-C. The human blood DNA methylome identifies crucial role of β /catenin in the pathogenesis of Kawasaki disease. *Oncotarget* 2018;9(47):28337.
- [6] Huang Y-H, Lo M-H, Cai X-Y, Liu S-F, Kuo H-C. Increase expression of CD177 in Kawasaki disease. *Pediatric Rheumatol* 2019;17(1):1–7.
- [7] Huang Y-H, Li S-C, Huang L-H, Chen P-C, Lin Y-Y, Lin C-C, Kuo H-C. Identifying genetic hypomethylation and upregulation of toll-like receptors in Kawasaki disease. *Oncotarget* 2017;8(7):11249.
- [8] Guo MM-H, Chang L-S, Huang Y-H, Wang F-S, Kuo H-C. Epigenetic regulation of macrophage marker expression profiles in Kawasaki disease. *Front Pediatrics* 2020;8:129.
- [9] Huang L-H, Kuo H-C, Pan C-T, Lin Y-S, Huang Y-H, Li S-C. Multiomics analyses identified epigenetic modulation of the S100A gene family in Kawasaki disease and their significant involvement in neutrophil transendothelial migration. *Clinical Epigenetics* 2018;10(1):1–13.
- [10] Kuo H-C, Li S-C, Huang L-H, Huang Y-H. Epigenetic hypomethylation and upregulation of matrix metalloproteinase 9 in Kawasaki disease. *Oncotarget* 2017;8(37):60875.
- [11] Bock C. Analysing and interpreting DNA methylation data. *Nat Rev Genet* 2012;13(10):705–19.
- [12] Huang Y-H, Lo M-H, Cai X-Y, Kuo H-C. Epigenetic hypomethylation and upregulation of NLR4 and NLRP12 in Kawasaki disease. *Oncotarget* 2018;9(27):18939.
- [13] McCrindle BW, Rowley AH, Newburger JW, Burns JC, Bolger AF, Gewitz M, Baker AL, Jackson MA, Takahashi M, Shah PB, et al. Diagnosis, treatment, and long-term management of Kawasaki disease: a scientific statement for health professionals from the American Heart Association. *Circulation* 2017;135(17):e927–99.
- [14] Huang Y-H, Kuo H-C, Li S-C, Cai X-Y, Liu S-F, Kuo H-C. HAMP promoter hypomethylation and increased hepcidin levels as biomarkers for Kawasaki disease. *J Mol Cell Cardiol* 2018;117:82–7.
- [15] Yang T-H, Wu W-S. Inferring functional transcription factor–gene binding pairs by integrating transcription factor binding data with transcription factor knockout data. *BMC Systems Biology* 2013;7(6):1–14.
- [16] Yang T-H. Transcription factor regulatory modules provide the molecular mechanisms for functional redundancy observed among transcription factors in yeast. *BMC Bioinformatics* 2019;20(23):1–16.
- [17] Yang T-H, Chiang Y-H, Shiu S-C, Yang Y-C, Tu K-C, Tseng Y-Y, Tseng JT, Wu W-S. Cancer DEIso: an integrative analysis platform for investigating differentially expressed gene-level and isoform-level human cancer markers. *Computational and Structural Biotechnology Journal* 2021;19:5149–59.
- [18] Yang T-H, Wang C-Y, Tsai H-C, Liu C-T. Human IRES Atlas: an integrative platform for studying IRES-driven translational regulation in humans. *Database* 2021;2021.
- [19] Y.R. Park, J.-J. Kim, Y.J. Yoon, Y.-K. Yoon, H.Y. Koo, Y.M. Hong, G.Y. Jang, S.-Y. Shin, J.-K. Lee, K.K.D.G. Consortium, et al., Establishment of Kawasaki disease database based on metadata standard, *Database* 2016 (2016).

- [20] Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009;37(1):1–13.
- [21] Carpenter S, Ricci EP, Mercier BC, Moore MJ, Fitzgerald KA. Post-transcriptional regulation of gene expression in innate immunity. *Nat Rev Immunol* 2014;14(6):361–76.
- [22] Liao K-C, Garcia-Blanco MA. Role of alternative splicing in regulating host response to viral infection. *Cells* 2021;10(7):1720.
- [23] Rodrigues R, Grosso AR, Moita L. Genome-wide analysis of alternative splicing during dendritic cell response to a bacterial challenge. *PLoS One* 2013;8(4):e61975.
- [24] Wells CA, Chalk AM, Forrest A, Taylor D, Waddell N, Schroder K, Himes SR, Faulkner G, Lo S, Kasukawa T, et al. Alternate transcription of the Toll-like receptor signaling cascade. *Genome Biol* 2006;7(2):1–17.
- [25] Jarešová I, Rožková D, Špišek R, Janda A, Brázová J, Šedivá A. Kinetics of Toll-like receptor-4 splice variants expression in lipopolysaccharide-stimulated antigen presenting cells of healthy donors and patients with cystic fibrosis. *Microbes Infect* 2007;9(11):1359–67.
- [26] Iwami K-I, Matsuguchi T, Masuda A, Kikuchi T, Musikacharoen T, Yoshikai Y. Cutting edge: naturally occurring soluble form of mouse Toll-like receptor 4 inhibits lipopolysaccharide signaling. *J Immunol* 2000;165(12):6682–6.
- [27] Stevenson BJ, Iseli C, Beutler B, Jongeneel CV. Use of transcriptome data to unravel the fine structure of genes involved in sepsis. *J Infect Dis* 2003;187(Supplement_2):S308–14.
- [28] Chang J-S, Huggett JF, Dheda K, Kim LU, Zumla A, Rook GA. Mycobacterium tuberculosis induces selective up-regulation of TLRs in the mononuclear leukocytes of patients with active pulmonary tuberculosis. *J Immunol* 2006;176(5):3010–8.
- [29] Haehnel V, Schwarzfischer L, Fenton MJ, Rehli M. Transcriptional regulation of the human Toll-like receptor 2 gene in monocytes and macrophages. *J Immunol* 2002;168(11):5629–37.
- [30] Kuo H-C, Wang C-L, Liang C-D, Yu H-R, Huang C-F, Wang L, Hwang K-P, Yang KD. Association of lower eosinophil-related T helper 2 (Th2) cytokines with coronary artery lesions in Kawasaki disease. *Pediatr Allergy Immunol* 2009;20(3):266–72.
- [31] Hopps E, Caimi G. Matrix metalloproteases as a pharmacological target in cardiovascular diseases. *Eur Rev Med Pharmacol Sci* 2015;19(14):2583–9.
- [32] Korematsu S, Ohta Y, Tamai N, Takeguchi M, Goto C, Miyahara H, Kawano T, Izumi T. Cell distribution differences of matrix metalloproteinase-9 and tissue inhibitor of matrix metalloproteinase-1 in patients with Kawasaki disease. *Pediatr Infect Dis J* 2012;31(9):973–4.
- [33] Newburger JW, Takahashi M, Burns JC, Beiser AS, Chung KJ, Duffy CE, Glode MP, Mason WH, Reddy V, Sanders SP, et al. The treatment of Kawasaki syndrome with intravenous gamma globulin. *N Engl J Med* 1986;315(6):341–7.
- [34] Chang D, Qian C, Li H, Feng H. Comprehensive analyses of DNA methylation and gene expression profiles of Kawasaki disease. *J Cell Biochem* 2019;120(8):13001–11.
- [35] Jaggi P, Mejias A, Xu Z, Yin H, Moore-Clingenpeel M, Smith B, Burns JC, Tremoulet AH, Jordan-Villegas A, Chaussabel D, et al. Whole blood transcriptional profiles as a prognostic tool in complete and incomplete Kawasaki disease. *PLoS One* 2018;13(5):e0197858.
- [36] Yang T-H, Chang H-T, Hsiao ES, Sun J-L, Wang C-C, Wu H-Y, Liao P-C, Wu W-S. iPhos: a toolkit to streamline the alkaline phosphatase-assisted comprehensive LC-MS phosphoproteome investigation. *BMC Bioinformatics* 2014;15(16):1–14.