European Review for Medical and Pharmacological Sciences 2018; 22: 6396-6402

Disturbed ribosome-related modules were associated with Kawasaki disease

X.-J. ZHANG[1], X. LI[2], Y. SUN[1], M. XU[1]

[1]Department of Pediatrics, Jinan Maternity and Child Care Hospital, Jinan, Shandong Province, P.R. China
[2]Neonatal Intensive Care Unit, Jinan Maternity and Child Care Hospital, Jinan, Shandong Province, P.R. China

Abstract. – OBJECTIVE: Kawasaki disease (KD) is an acute vasculitis in young children, with ambiguous etiology. Early diagnosis will decrease the risk of coronary artery aneurysms and contributes to the favourable prognosis. This work aimed to identify disease modules that accurately predict clinical outcome using a systemic module tracking method.

PATIENTS AND METHODS: Based on the transcriptional data and protein-protein interaction (PPI) data, we constructed the differential co-expression network for KD. Then, a systemic module tracking method was performed to extract KD-related modules that accurately predict clinical outcome from the differential co-expression network, according to two steps: key genes identification and module inference by key gene expansion.

RESULTS: 16 key genes were identified based on their importance in differential co-expression network and most of them were ribosomal protein-related genes. With each key gene as initial gene, we identified 10 disease modules with high predictive accuracy. Function analysis found that these disease modules were related to one common pathway-ribosome pathway.

CONCLUSIONS: Our study for the first time indicated that disturbed ribosome-related disease modules might contribute to the development of KD. These modules could be considered as novel contributors to the progression of KD, and potential diagnostic biomarkers for predicting the clinical outcome.

Key Words:
Kawasaki disease, Gene module, Ribosomal protein, Differential co-expression network.

Introduction

Kawasaki disease (KD), an acute vasculitis mostly affecting young children, is recognized as a serious and potentially life-threatening condition. Approximately 25.0% cases untreated with intravenous immunoglobulin would develop coronary artery aneurysms1. In the developed world, KD is the predominant cause of pediatric acquired heart disease[2]. In North America, Australia and Europe, present incidence is about 4-25 per 100,000 children younger than 5 years old, and presents a stable incidence from 2000 to 2012[3,4]. In China, the incidence is high and increasing during the last 10-20 years based on studies conducted in different provinces; however, there is still no accurate national epidemiology[4]. Despite advances on epidemiology, genetic susceptibility and pathogenesis of KD, the etiology of this enigmatic disease remains ambiguous[5,6]. Because of the absence of specific and sensitive diagnostic test, the symptomatic diagnosis and treatment may be not completely reliable and be delayed. A comparison study has indicated that delayed diagnosis of KD increases the risk of coronary artery aneurysms[7]. Current documents believe that KD is a complex interplay of genetic factors, infections, and immunity[8,9]. Hence, a better understanding of the underlying genetic factors may contribute to early diagnosis and efficient therapeutic interventions for KD.

The development of high-throughput technology and bioinformatics has driven the investigators to shift the way of recognizing complex diseases. Previous transcriptome analyses have yielded many genes influencing the likelihood of developing and treating KD[10,11]. While disease development always refers to the dysregulation of a set of genes, interpreting that the consequences by gene sets may facilitate a better understanding of how gene perturbations are responsible for disease[12]. It is well known that genes with related functions are often co-expressed and interact with each other to form a biological network[13]. Network analysis has been widely used to predict pathogenic genes and gene sets involved in complex diseases[14]. Moreover, several computational algorithms have been

Corresponding Author: Xia Li, MD; e-mail: li1437746033@126.com

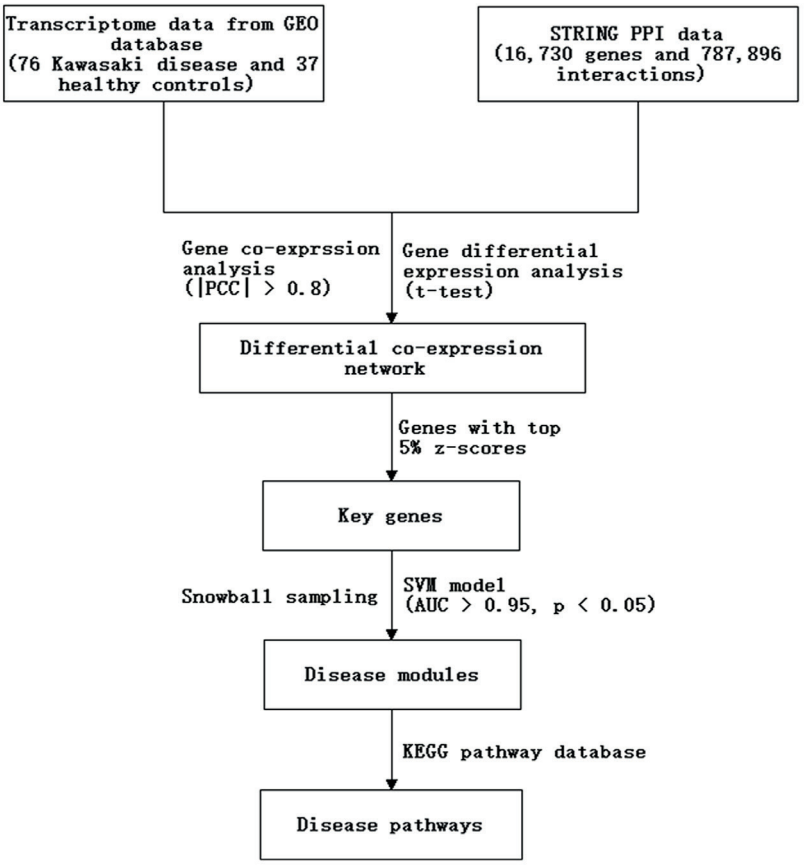

Figure 1. A schematic diagram for the systemic module tracking method.

developed to excavate significant sub-networks for predicting clinical phenotypes[15,16]. While existing algorithms are usually heuristic, the derived sub-networks are ambiguous without a formal topological feature. It is necessary to develop a sub-network prediction method that not only embody disease-related genes but also predict clinical outcome accurately.

Here, we proposed a systemic module tracking method to extract KD-related modules that could accurately predict clinical outcome from the differential co-expression network. A schematic diagram was described in Figure 1. This study might give a better understanding of the underlying molecular mechanism of KD and contribute to early diagnosis and therapy for KD.

Materials and Methods

Transcriptome Data

In this work, the whole blood transcriptional data of KD were recruited from GEO database (http://www.ncbi.nlm.nih.gov/geo/), under the accessing number of GSE68004. The microarray data were presented on Illumina HumanHT-12 V4.0 expression beadchip. In dataset GSE68004, whole blood RNA samples were obtained from 76 pediatric patients with complete KD, 13 with incomplete KD, 19 patients with adenovirus infection, 17 patients with Group A streptococcal disease, 18 patients with fever of undetermined origin, and 37 age- and sex-matched healthy controls. To extract KD-related modules that could accurately predict clinical outcome, only 76 pediatric patients with complete KD and 37 age- and sex-matched healthy controls were included in this study. The whole blood transcriptional data and annotation files of KD subjects and healthy controls were downloaded and preprocessed via standard procedure[17,18]. After mapping gene expression data onto probe level into gene symbol level, a total of 21,043 genes and their expression data were obtained. This study was approved by the ethics committee of Jinan Maternity and Child Care Hospital.

General Network

The general protein-protein interaction (PPI) network was constructed on basis of the PPI data from the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database (http://string-db.org/). In STRING database, there are a total of 1,048,576 human interactions. The self-loops and duplicated PPIs were removed from this study, and 787,896 interactions were remained for subsequent analysis.

Differential Co-Expression Network

This work attempted to identify KD-related modules from a differential co-expression network view. Thus, we combined genes in transcriptional data with the general PPI network, and obtained a KD-related network, including 8387 genes and 52,425 PPIs. To obtain the differential co-expression network for KD, the gene co-expression values were firstly calculated using Pearson correlation coefficients (PCC)[19] based on the transcriptional data, and only the highly co-expressed PPIs (|PCC| > 0.8) under KD condition were remained for the differential co-expression network. Then, the gene differential expression analysis was performed to determine the differential levels between KD and healthy control conditions using one-side *t*-test, and the *p*-value was adjusted using Benjamini-Hochberg fault detection rate (FDR). Based on the co-expression and differential values, each interaction was given a weight value to reflect the degrees of co-expression and differential expression. Given an interaction between genes i and j, the weight value was calculate as follows:

$$\begin{cases} w_{ij} = \dfrac{\sqrt{logp_i + logp_j}}{\sqrt{2 * max_{l \in V} |logp_l|}}, if\ |PCC| > 0.8 \\ w_{ij} = 0, if\ |PCC| \leq 0.8 \end{cases}$$

where p_i and p_j represented *p*-values of differential expression for genes i and j. V stood for the gene set of the co-expression network. In this case, we constructed the KD-related differential co-expression network with each interaction assigned a weight value.

Disease Modules

Here a systemic module tracking method was presented to extract KD-related modules that could accurately predict clinical outcome from the differential co-expression network, according two steps: key genes identification, and module inference by key gene expansion.

Key Genes Identification

In this study, the key genes were determined based on the topological parameter of the genes in the differential co-expression network. The differential co-expression network was visualized as an adjacency matrix $A = (a_{ij})_{n \times n}$, and the importance of gene i in the differential co-expression network was defined as,

$$g(i) = \sum_{j \in N(i)} A'_{ij} g(j)$$

where N(i) was the neighbor set of gene i in the differential co-expression network, A' represented the degree normalized adjacency matrix, and A'g stood for the information propagation on network through the interactions, reflecting that the importance of a gene was determined by the neighbor numbers, co-expression values and the importance of its neighbors. Then, we calculated the z-scores[20] according to the importance of genes in the differential co-expression network. Based on the z-scores, all genes in the differential co-expression network were ranked in descending order, and the genes with top 5% z-scores were defined as key genes for KD.

Module Inference

In this study, module inference was performed using a snowball sampling strategy[21]. Specifically, with each key gene as initial gene, genes were involved in the level-one network based on how well the collection of genes predicted the clinical outcome. Then, it progressively spread outward from the initial gene to scan more genes in the predictive model. Through the iterative procedure, module inference by key gene expansion stopped until the prediction accuracy dropped. Here, support vector machines (SVM)[22], a machine-learning method, was implemented to assess the power of the disease module to predict clinical outcome. For prediction accuracy, the area under the receiver operating characteristic curve (AUC) was used as the metric to evaluate the performance. Modules with AUC > 0.95 and gene number > 5 were considered as candidate modules for KD. Finally, module significance was determined by permutation test. Modules with Benjamini-Hochberg adjusted *p*-value < 0.05 were defined as disease modules for KD.

Function Inference

Generally, co-expressed genes are inclined to involve in similar functions. Thus, after identifying disease modules, we performed the function inference for each disease module. Here, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database was utilized to reveal biological functions related to KD. In this study, the statistical significance was determined by Fisher's exact test, and pathways with Benjamini-Hochberg adjusted p-values < 0.05 were regarded as significant pathways. The most significant pathway was defined as the optimal function for each disease module.

Statistical Analysis

The gene differential expression analysis between two groups was performed using one-side t-test. Module significance and function inference significance were determined by permutation test and Fisher's exact test, respectively. p-value < 0.05 were regarded to be statistically significant.

Results

Differential Co-Expression Network

From transcriptional data of KD, we obtained the expression data of 21,043 genes under KD and healthy conditions. From STRING, we screened a total of 787,896 human interactions (covering 16,730 genes). Integrating genes in transcriptional data with the general PPI network, a KD-related network, including 8387 genes and 52,425 PPIs, was obtained. Via PCC-based co-expression analysis, only interactions with |PCC| > 0.8 were remained for the co-expression network. Then gene differential expression analysis between KD and healthy control conditions was performed using one-side t-test. The differential co-expression network was constructed based on the co-expression and differential expression analyses, in which each interaction was allotted a weight value to reflect the degrees of co-expression and differential expression.

Key Genes

From the differential co-expression network, key genes were identified based on the importance of genes in the network. Based on the z-scores, all genes were ranked in descending order, and the top 5% genes were defined as key genes for KD. In this study, a total of 16 key genes were determined, as shown in Table I. Among them, RPS13 showed the highest z-score (z = 433.8), indicating its most important role in the differential co-expression network.

Table I. Key genes with top 5% z-scores for Kawasaki disease.

Key gene	z-score	Key gene	z-score
RPS13	433.8	EEF1B2	289.5
RPL6	411.8	RPL19	283.3
RPL18A	408.6	RPL4	277.6
RPS5	387.6	RPS6	272.8
RPL36	363. 1	RPL22	271.3
RPL35	338.7	RPS8	269.2
RPL18	323.7	RPL27	262.5
RPS16	316.3	RPL7A	261.3

Disease Modules

To uncover the hidden genes with no significance by themselves, we performed a module inference procedure to find genes as a collection that could accurately predict the clinical outcome. With each key gene as initial gene, the module inference was performed by a snowball sampling strategy. Starting from 16 key genes, we identified 16 candidate disease modules. Among them, there were a total of 11 modules with AUC > 0.95 and gene number > 5. Permutation test identified 10 modules with p-value < 0.05, that were defined as disease modules for KD, named as module 1-module 10. The details of disease modules was shown in Table II. Among them, module 1 and module 2 showed a predictive accuracy of 1, indicating that they could accurately predict the clinical outcome. The graph of module 1 and module 2 was shown in Figure 2.

Pathway Inference

After obtaining disease modules, each disease module was given a related pathway by function inference. Based on KEGG pathway database, the most significant pathway of each disease module was defined as the optimal function. Our result showed that ten disease modules were related to one common pathway-ribosome pathway (Table III), including RPL30, RPS16, RPL18, RPL18A and so forth, indicating its important role in the development of KD. Especially, RPL30 and RPS16 were involved in module 1 and module 2, implying they might be significant to KD progression.

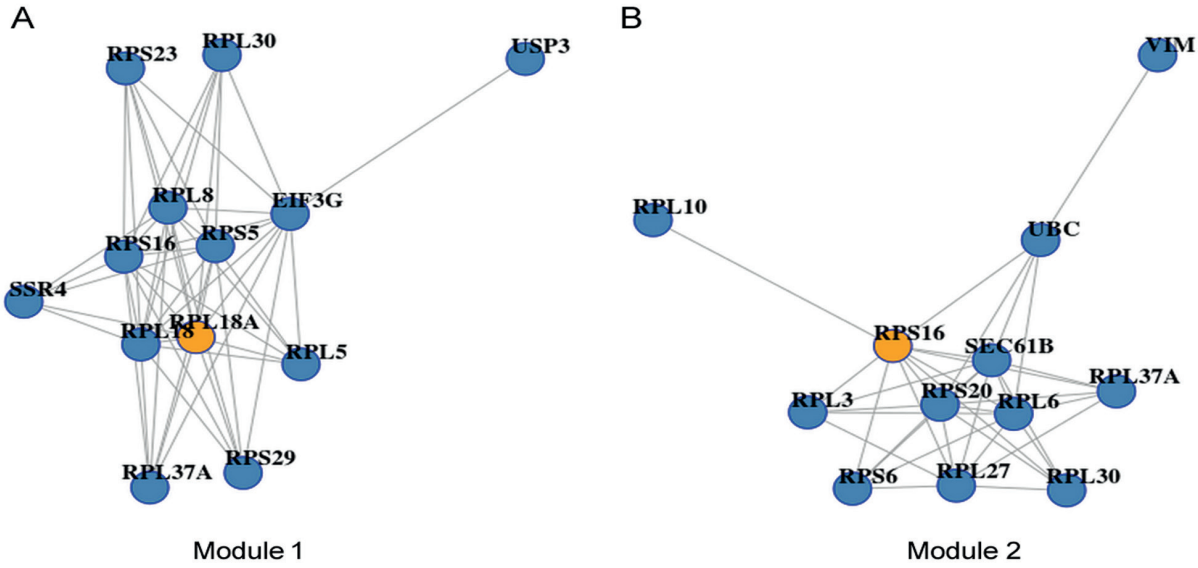

Figure 2. The Kawasaki disease-related modules with a predictive accuracy of 1. Yellow node represented key gene.

Discussion

It has been largely accepted that KD is a consequence of combined action of genetics, infections, and immune dysregulation, where genetic factors play a crucial role in the pathogenesis of KD. KD presents a racial predilection to Asian populations, also implying a genetic predisposition to this disease. In addition, a Japanese nationwide survey documented that parents with a history of KD had increased offspring and sibling risk in KD[23]. Several studies[24-26] have investigated this theory in different populations and found several genes associated with susceptibility to KD. Unfortunately, KD is still a difficult medical challenge. Current diagnosis is mainly based on clinical signs and symptoms, and there are no specific diagnostic tests for KD. It is imperative to identify diagnostic biomarkers to help clinicians correctly classify clinical phenotype or predict clinical outcomes. Thus, we attempted to identify KD-related modules that could accurately predict clinical outcome using a systemic module tracking method.

In this investigation, a total of 16 key genes were identified from the differential co-expression network, and almost all of them had direct correlations with ribosomal protein except for *EEF1B2*. EEF1B2 encodes a translation elongation factor, which is a guanine nucleotide exchange factor mediating the transfer of aminoacylated tRNAs to the ribosome. Thus, all key genes were related to ribosome. Starting from these key genes, we identified ten KD-related disease modules, which could accurately distinguish KD subjects from healthy controls. Functional inference indicated that

Table II. The details of disease modules for Kawasaki disease.

Module	Initial gene	Gene number	Accuracy	*p*-value
1	RPL18A	13	1	0
2	RPS16	12	1	0
3	RPL35	9	0.98	0.007
4	RPS5	11	0.98	0.038
5	RPS13	10	0.98	0.012
6	RPL36	8	0.98	0.013
7	RPL4	14	0.98	0.013
8	RPL22	10	0.98	0.02
9	RPL18	8	0.98	0.02
10	RPL7A	8	0.98	0.032

these disease modules were related to one common pathway, i.e., ribosome pathway, implying that these disease modules might promote the progression of KD by regulating ribosome pathway.

The ribosome, as a cellular organelle responsible for protein synthesis, is essential for all organisms. Evolutionary and genetic considerations prompt investigators to predict the roles of ribosomal protein-related genes in human disease[27]. It seems probable that defective ribosomal proteins will lead to the dysfunction of ribosome, resulting in pathological conditions. Dysregulated ribosomal function is involved in the decreased expression of ribosomal proteins, decreased protein synthesis, and increased RNA oxidation[27]. It is indicated that the decrease of protein synthesis in KD subjects is driven by decreased expression of ribosomal proteins, without increased expression of genes identified to inhibit protein synthesis[28]. It would be of great interest to uncover the specific ribosomal proteins whose aberrance disturbs normal cell function and lead to the development of KD.

KD is an acute systemic vasculitis syndrome primarily affecting infants and young children. It has been documented that suppression of ribosomal function leads to initiation of inflammatory activation by interleukin (IL)-1β and inflammation[29]. Gan et al[30] also indicated that ribosome-related genes and ribosome pathway were involved in systemic vasculitis. Hoang et al[10] illustrated that ribosomal protein-related genes were prominently downregulated in acute KD subjects relative to convalescent KD samples. Fukazawa et al[31] showed that RPL6 was differentially expressed in KD-induced coronary artery aneurysm. A genome-wide association study of Han Chinese population has identified that a mitochondrial ribosomal gene S22 showed the strongest association with KD[32]. In our study, all KD-related disease modules that could accurately predict clinical outcome were associated with ribosomal protein-related genes and ribosome pathway. The findings led to a speculation that dysregualtion of ribosome function might promote inflammation activation, thus resulting in the progression of KD.

Conclusions

We found for the first time that impairments in ribosomal proteins may be important genetic factor in the development of KD. These ribosome-related disease modules could be considered as novel contributors to the progression of KD, and potential diagnostic biomarkers for predicting the clinical outcome.

Conflict of Interest

The Authors declare that they have no conflict of interest.

References

1) Kato H, Sugimura T, Akagi T, Sato N, Hashino K, Maeno Y, Kazue T, Eto G, Yamakawa R. Long-term consequences of Kawasaki disease. A 10- to 21-year follow-up study of 594 patients. Circulation 1996; 94: 1379-1385.
2) McCrindle BW, Rowley AH, Newburger JW, Burns JC, Bolger AF, Gewitz M, Baker AL, Jackson MA, Takahashi M, Shah PB, Kobayashi T, Wu MH, Saji TT, Pahl E. Diagnosis, treatment, and long-term management of Kawasaki disease: a scientific statement for health professionals from the American Heart Association. Circulation 2017; 135: e927-e999.
3) Uehara R, Belay ED. Epidemiology of Kawasaki disease in Asia, Europe, and the United States. J Epidemiol 2012; 22: 79-85.
4) Singh S, Vignesh P, Burgner D. The epidemiology of Kawasaki disease: a global update. Arch Dis Child 2015; 100: 1084-1088.
5) Agarwal S, Agrawal DK. Kawasaki disease: etiopathogenesis and novel treatment strategies. Expert Rev Clin Immunol 2017; 13: 247-258.
6) Rowley AH. Kawasaki disease: novel insights into etiology and genetic susceptibility. Annu Rev Med 2011; 62: 69-77.
7) Wilder MS, Palinkas LA, Kao AS, Bastian JF, Turner CL, Burns JC. Delayed diagnosis by physicians contributes to the development of coronary artery aneurysms in children with Kawasaki syndrome. Pediatr Infect Dis J 2007; 26: 256-260.
8) Qi XL, Chen LL, Sun XG, Li XM, Zhao LH, Kong DJ. 1,25-Dihydroxyvitamin D3 regulates T lymphocyte proliferation through activation of P53 and of ERK1/2 signaling pathway in children with Kawasaki disease. Eur Rev Med Pharmacol Sci 2017; 21: 3714-3722.
9) Zhou Y, Wang S, Zhao J, Fang P. Correlations of complication with coronary arterial lesion with VEGF, PLT, D-dimer and inflammatory factor in child patients with Kawasaki diesase. Eur Rev Med Pharmacol Sci 2018; 22: 5121-5126.
10) Ogihara Y, Ogata S, Nomoto K, Ebato T, Sato K, Kokubo K, Kobayashi H, Ishii M. Transcriptional regulation by infliximab therapy in Kawasaki disease patients with immunoglobulin resistance. Pediatr Res 2014; 76: 287-293.

11) Hoang LT, Shimizu C, Ling L, Naim AN, Khor CC, Tremoulet AH, Wright V, Levin M, Hibberd ML, Burns JC. Global gene expression profiling identifies new therapeutic targets in acute Kawasaki disease. Genome Med 2014; 6: 541.

12) Xi WD, Liu YJ, Sun XB, Shan J, Yi L, Zhang TT. Bioinformatics analysis of RNA-seq data revealed critical genes in colon adenocarcinoma. Eur Rev Med Pharmacol Sci 2017; 21: 3012-3020.

13) Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL. Hierarchical organization of modularity in metabolic networks. Science 2002; 297: 1551-1555.

14) Wang JY, Wang CL, Wang XM, Liu FJ. Comprehensive analysis of microRNA/mRNA signature in colon adenocarcinoma. Eur Rev Med Pharmacol Sci 2017; 21: 2114-2129.

15) Dutkowski J, Ideker T. Protein networks as logic functions in development and cancer. PLoS Comput Biol 2011; 7: e1002180.

16) Zhu Y, Shen X, Pan W. Network-based support vector machine for classification of microarray samples. BMC Bioinformatics 2009; 10 Suppl 1: S21.

17) Ma L, Robinson LN, Towle HC. ChREBP• Mlx is the principal mediator of glucose-induced gene expression in the liver. J Biol Chem 2006; 281: 28721-28730.

18) Pepper SD, Saunders EK, Edwards LE, Wilson CL, Miller CJ. The utility of MAS5 expression summary and detection call algorithms. BMC Bioinformatics 2007; 8: 273.

19) Benesty J, Chen J, Huang Y, Cohen I. Pearson correlation coefficient. In, Noise reduction in speech processing: Springer, 2009; pp. 1-4.

20) Chen L, Xuan J, Riggins RB, Wang Y, Clarke R. Identifying protein interaction subnetworks by a bagging Markov random field-based method. Nucleic acids research 2013; 41: e42-e42.

21) Goodman LA. Snowball sampling. Annals Mathematical Statistics 1961: 148-170.

22) Gillani Z, Akash MS, Rahaman MD, Chen M. CompareSVM: supervised, Support Vector Machine (SVM) inference of gene regularity networks. BMC Bioinformatics 2014; 15: 395.

23) Uehara R, Yashiro M, Nakamura Y, Yanagawa H. Parents with a history of Kawasaki disease whose child also had the same disease. Pediatr Int 2011; 53: 511-514.

24) Khor CC, Davila S, Breunis WB, Lee YC, Shimizu C, Wright VJ, Yeung RS, Tan DE, Sim KS, Wang JJ, Wong TY, Pang J, Mitchell P, Cimaz R, Dahdah N, Cheung YF, Huang GY, Yang W, Park IS, Lee JK, Wu JY, Levin M, Burns JC, Burgner D, Kuijpers TW, Hibberd ML. Genome-wide association study identifies FCGR2A as a susceptibility locus for Kawasaki disease. Nat Genet 2011; 43: 1241-1246.

25) Khor CC, Davila S, Shimizu C, Sheng S, Matsubara T, Suzuki Y, Newburger JW, Baker A, Burgner D, Breunis W, Kuijpers T, Wright VJ, Levin M, Hibberd ML, Burns JC. Genome-wide linkage and association mapping identify susceptibility alleles in ABCC4 for Kawasaki disease. J Med Genet 2011; 48: 467-472.

26) Onouchi Y, Ozaki K, Burns JC, Shimizu C, Terai M, Hamada H, Honda T, Suzuki H, Suenaga T, Takeuchi T, Yoshikawa N, Suzuki Y, Yasukawa K, Ebata R, Higashi K, Saji T, Kemmotsu Y, Takatsuki S, Ouchi K, Kishi F, Yoshikawa T, Nagai T, Hamamoto K, Sato Y, Honda A, Kobayashi H, Sato J, Shibuta S, Miyawaki M, Oishi K, Yamaga H, Aoyagi N, Iwahashi S, Miyashita R, Murata Y, Sasago K, Takahashi A, Kamatani N, Kubo M, Tsunoda T, Hata A, Nakamura Y, Tanaka T. A genome-wide association study identifies three new risk loci for Kawasaki disease. Nat Genet 2012; 44: 517-521.

27) Yoshihama M, Uechi T, Asakawa S, Kawasaki K, Kato S, Higa S, Maeda N, Minoshima S, Tanaka T, Shimizu N, Kenmochi N. The human ribosomal protein genes: sequencing and comparative analysis of 73 genes. Genome Res 2002; 12: 379-390.

28) Herberg JA, Kaforou M, Gormley S, Sumner ER, Patel S, Jones KD, Paulus S, Fink C, Martinon-Torres F, Montana G, Wright VJ, Levin M. Transcriptomic profiling in childhood H1N1/09 influenza reveals reduced expression of protein synthesis genes. J Infect Dis 2013; 208: 1664-1668.

29) Vyleta ML, Wong J, Magun BE. Suppression of ribosomal function triggers innate immune signaling through activation of the NLRP3 inflammasome. PLoS One 2012; 7: e36044.

30) Gan SJ, Ye B, Qian SX, Zhang C, Mao JQ, Li K, Tang JD. Immune- and ribosome-related genes were associated with systemic vasculitis. Scand J Immunol 2015; 81: 96-101.

31) Fukazawa R, Ikegam E, Watanabe M, Hajikano M, Kamisago M, Katsube Y, Yamauchi H, Ochi M, Ogawa S. Coronary artery aneurysm induced by Kawasaki disease in children show features typical senescence. Circ J 2007; 71: 709-715.

32) Tsai FJ, Lee YC, Chang JS, Huang LM, Huang FY, Chiu NC, Chen MR, Chi H, Lee YJ, Chang LC, Liu YM, Wang HH, Chen CH, Chen YT, Wu JY. Identification of novel susceptibility Loci for Kawasaki disease in a Han chinese population by a genome-wide association study. PLoS One 2011; 6: e16853.