MEDICAL SCIENCE MONITOR

LAB/IN VITRO RESEARCH

e-ISSN 1643-3750
© Med Sci Monit, 2016; 22: 4152-4158
DOI: 10.12659/MSM.897792

Received: 2016.01.26
Accepted: 2016.03.15
Published: 2016.11.02

A Novel Method for Pathway Identification Based on Attractor and Crosstalk in Polyarticular Juvenile Idiopathic Arthritis

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

BCDEFG 1,2 Yuanji Wang
BCDEFG 1,2 Shunhua Lin
BCDEFG 1,2 Changhui Li
CDEFG 1 Yizhao Li
ABCDEFG 1 Lei Chen
ABCDEFG 2 Yingzhen Wang

1 Department of Orthopaedics, The People's Hospital of Rizhao, Rizhao, Shandong, P.R. China
2 Department of Joint Surgery, The Affiliated Hospital of Qingdao University, Qingdao, Shandong, P.R. China

Corresponding Author: Yingzhen Wang, e-mail: 18661808238@163.com
Source of support: Departmental sources

Background: Juvenile idiopathic arthritis (JIA) is one of the most common inflammatory disorders of unknown etiology. We introduced a novel method to identify dysregulated pathways associated with polyarticular JIA (pJIA).

Material/Methods: Gene expression profiling of 61 children with pJIA and 59 healthy controls were collected from E-GEOD-13849; 300 pathways were obtained from Kyoto Encyclopedia of Genes and Genomes (KEGG) database and 787,896 protein-protein interaction sets were gathered from the Retrieval of Interacting Genes. Attractor and crosstalk were designed to complement each other to increase the integrity of pathways assessment. Then, impact factor was used to assess the interactions inter-pathways, and RP-value was used to evaluate the comprehensive influential ability of attractors.

Results: There were seven attractors with $p<0.01$ and 14 pathways with RP<0.01. Finally, two significantly dysfunctional pathways were found, which were related to pJIA progression: p53 signaling pathway (KEGG ID: 04115) and non-alcoholic fatty liver disease (NAFLD) (KEGG ID: 04932).

Conclusions: A novel approach that identified the dysregulated pathways in pJIA was constructed based on attractor and crosstalk. The new process is expected to be efficient in the upcoming era of medicine.

MeSH Keywords: Arthritis, Juvenile • Critical Pathways • Receptor Cross-Talk

Full-text PDF: http://www.medscimonit.com/abstract/index/idArt/897792

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS] [Index Copernicus]

Background

Juvenile idiopathic arthritis (JIA) is one of the most common inflammatory disorder with unknown etiology, presenting in children younger than 16 years of age [1]. Polyarticular JIA (pJIA) is defined as disease involving more than five joints [2]. The incidence and prevalence is 0.07 to 4.01 per 1,000 children varying by geography and ethnicity [3,4]. The treatment of JIA is time-consuming and frequently causes complications after managed with drugs [5,6]. Previous studies reported the composition of phospholipid fatty acids [7], and immunological investigations [8] in children with juvenile chronic arthritis.

The current discrepancy in the prevalence of JIA may be affected by genetic factors.

Identifying differentially expressed genes (DEG) and dysregulated pathways using high-throughput experimental data between normal and JIA disease groups can provide insights [9].

The DEG and pathways are easy to understand high-level functions that can be used to infer potential molecular and functional insights. Recent reports claimed that gene expression patterns could identify biomarkers of arthritis, and highlighted the relevance of pJIA through dysregulated pathways [10–13].

Pathway databases, such as Kyoto Encyclopedia of Genes and Genomes (KEGG), can exploit useful pathway topology information. It is well known that Kauffman' attractor can find many well-defined ensembles of model networks whose statistical features matched those of real cells and organisms [14]. Mar et al. [15] reported that *attract* could be a new approach that leverages both existing DEG and pathways among three different cell phenotypes. We employed *attract* to screen differentially expressed pathways for the purpose of narrowing the number of correlated dysregulated pathways.

Screened attractors are an efficient means to identify target functions. However, they invariably focus on internal effects of a single pathway in isolation and fail to consider the inherent dependency inter-pathways. Pathway crosstalk refers to the interaction or cooperation between pathways. The construction of a pathway crosstalk network (PCN) inter-pathways is convenient to master the comprehensive interactions in pJIA [16]. In our study we used a scoring scheme to identify these pathways, taking into account both attractors of internal pathway effects and crosstalk inter-pathways.

In our study, we introduced a novel approach to identify dysregulated pathways associated with pJIA. What we want is significantly dysfunctional pathways with strong interactions which are related to pJIA progression. The new method may be meaningful in seeking influential pathways by reinforcement of attractor and crosstalk, which served as therapy targeting markers.

Material and Methods

Gene expression datasets

The transcription profile was obtained from EMBI-EBI ArrayExpress [17]. Gene expression profiling of 61 children with pJIA and 59 healthy controls were collected from E-GEOD-13849 [18]. The platform was A-AFFY-44 - Affymetrix GeneChip Human Genome U133 Plus 2.0 (HG-U133_Plus_2).

The gene chip data were read using the affy package [19]. The Linear Models for Microarray Data (LIMMA) [20] was then used to preprocess data. Background adjustment and quantile data normalization were performed by robust multi-array average [21]. To protect against outlier probes we used a robust procedure, median polish, to estimate model parameters. The average value of a gene symbol with multiple probes was calculated. To screen DEG, $p \leq 0.01$ and |log fold change (FC) | ≥ 2 were set as the threshold levels.

Pathway data

Information from gene sets representing biological human pathways was obtained from KEGG database [22], which provides copious pathway information [23,24]. A set of pathways with gene set size >100 or <5 were filtered. After size cutoffs were set, 300 pathways were obtained for downstream analysis.

Protein interaction data

The human protein-protein interaction (PPI) sets representing biological genes were obtained from the Retrieval of Interacting Genes (STRING; v. 9.0) [25]. After removing self-interactions, we had 787,896 PPI sets.

Attractor analysis within pathways

Based on attractor theory [14], *attract* was used to screen differentially expressed pathways related to pJIA from 300 KEGG pathways.

To test data of 300 KEGG pathways, GSEA-ANOVA was employed as a gene set enrichment algorithm; which was different from other methods in multiple classes [15]. The obtained differences among the expression profile of samples were identified as attractors. From the ANOVA model, we compute the *F*-statistic for gene *i*:

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS] [Index Copernicus]

$$F^{(i)} = \frac{MSS_i}{RSS_i} \quad (1)$$

where MSS_i represents the mean treatment sum of squares, and captures the amount of variation due to group-specific effects:

$$MSS_i = \frac{1}{K-1}\sum_{k=1}^{k} r_k \left[y_k^{(i)} - y^{(i)}\right]^2 \quad (2)$$

and RSS_i represents the residual sum of squares:

$$RSS_i = \frac{1}{N-K}\sum_{k=1}^{K}\sum_{j=1}^{r_j}\left[y_{jk}^{(i)} - y^{(i)}\right]^2 \quad (3)$$

where N is the total number of samples, and the overall mean is given by:

$$y^{(i)} = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{1}{r_k}\sum_{j=1}^{r_k} y_{jk}^{(i)}\right) \quad (4)$$

The *F*-statistic captures the strength of different expression observed in genes of pJIA. Large values of the *F*-statistic mean a strong association with JIA-specific expression changes.

For pathway *P* consisting of g_p genes, the *T*-statistic takes the following form:

$$T_p = \frac{\left[\frac{1}{g_p}\sum_{i=1}^{g_p} F^{(i)}\right] - \left[\frac{1}{G}\sum_{j=1}^{G} F^{(j)}\right]}{\sqrt{\left(\frac{s_p^2}{g_p}\right) + \left(\frac{s_G^2}{G}\right)}} \quad (5)$$

where *G* represents the total number of genes with a pathway annotation and the sample variances S_p^2 and S_G^2 are defined as:

$$s_p^2 = \frac{1}{g_p - 1}\sum_{j=1}^{g_p}\left(F^{(j)} - \frac{1}{g_p}\sum_{i=1}^{g_p} F^{(i)}\right)^2 \quad (6)$$

$$s_G^2 = \frac{1}{G-1}\sum_{j=1}^{G}\left(F^{(j)} - \frac{1}{G}\sum_{i=1}^{G} F^{(i)}\right)^2 \quad (7)$$

and the degrees of freedom are specified by the Welch-Satterwhaite equation:

$$v = \frac{\left(\frac{s_p^2}{g_p} + \frac{s_G^2}{G}\right)^2}{\frac{s_p^4}{g_p^2(g_p - 1)} + \frac{s_G^4}{G^2(G-1)}} \quad (8)$$

Attractors were ranked according to the significance of difference.

Crosstalk analysis of inter-pathways

Background analysis

The pathway crosstalk network (PCN) of the control group was constructed using the Li et al. [16] method. The value of weight of the background PCN was defined as the number of PPI sets.

1) Fish Exact test was performed to evaluate gene overlap between any pair of 300 pathways [26]. Raw *p*-values were adjusted by false discovery rate (FDR) [27]. Pathway pairs with adjusted $p<0.05$ were removed.
2) The number of PPI sets was counted between any pair of pathways. For each pathway pair, count was for all interactions after removing genes shared in both pathways.
3) Background distribution of PPI sets counted in each pair of pathways was estimated. Every pathway was randomized, repeating 1,000 times. When a gene in the pathway had interactions, it was considered as crosstalk between pathways. First, we count the number of genes it interacts with, and randomly draws a gene from the PPI database which interacts with a similar number of genes. Then the original gene was replaced with this newly selected gene. Once both pathways were randomized, Step 2 was performed to count the number of interactions between them.
4) The one-sided Fisher Exact test on all pathway pairs was performed using the 2×2 contingency table. *P*-values of Fisher Exact test were adjusted using FDR BH procedure and empirical *p*-value was calculated by counting the number of permutations in which the count of random interactions was higher than or equal to that of true interactions.
5) All pathway pairs with adjusted Fisher $p<0.05$ were used to construct the PCN, where a node represented a pathway and an edge was crosstalk between two pathways. To clean up the network, two types of 'redundant' edges were removed: (a) edges with significant gene overlap identified in Step 1 were removed from the network; (b) the two edges between two overlapping pathways were considered redundant.

Network of pJIA

Base on the original method of crosstalk [16], the network of pJIA could be constructed. In Step 3, we modified it to narrow the number of edges in the network.

A gene in the pathway had interactions when it met one of the two conditions: (1) Spearman correlation coefficients of every PPI set were calculated in the control and normal group. When the absolute value of the different value between them was >0.7, the edge remained. Geometric mean of the absolute value was defined as the value of the weight between the two pathways (2) The two genes in an interaction were DEGs. The $p<0.01$ and |log fold change (FC) | >1 were set as the threshold levels for the identification of DEG.

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS] [Index Copernicus]

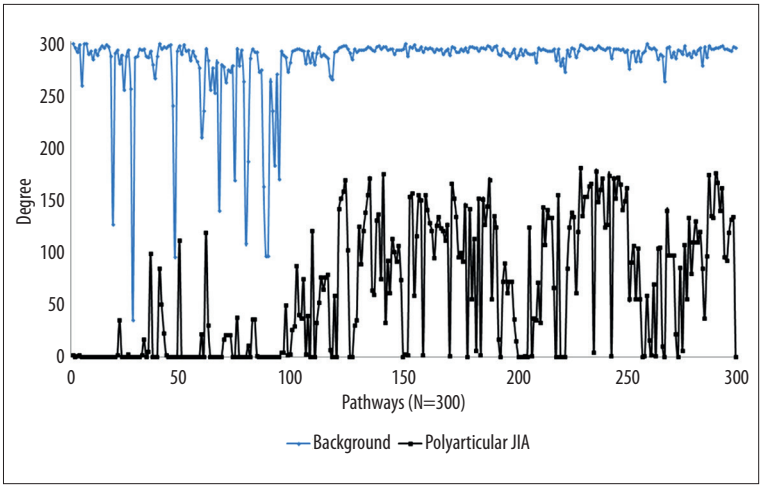

Figure 1. The crosstalk difference of background (control) and pJIA.

Table 1. 7 significant pathways identified by Kauffman' attractor (P<0.01).

KEGG ID	KEGG Pathway	Attractor P-value	Impact factor	RP-value
03010	Ribosome	9.29E-15	29	0.001878
03050	Proteasome	0.00039	32.98714	0.003689
05016	Huntington's disease	0.000548	68.96219	0.0044
04932	Non-alcoholic fatty liver disease (NAFLD)	0.000827	106.9115	0.004089
00190	Oxidative phosphorylation	0.00366	1.99268	0.011
03040	Spliceosome	0.005941	25.84552	0.011333
05012	Parkinson's disease	0.005941	15.90494	0.014

Important crosstalk pathways

The PCN was performed in topology analysis. Pathways were ordered by the degree of nodes. The scores of pathways were calculated.

$$\text{Score} = \text{degree of pJIA/degree of background} \quad (9)$$

Comprehensive analysis

Impact factor was used to assess the interactions between 2 pathways.

$$\text{Impact factor} = \text{outer} \times (1\text{-}p) \quad (10)$$

Outer means the degree of interactions from crosstalk analysis and p represents the p-value of the attractor.

RP-value was used to evaluate the comprehensive identified ability within pathways and between pathways.

$$\text{RP-value} = (\text{rank inter/total}) \times (\text{rank outer/total}) \quad (11)$$

Rank inter represents the ranking of attractor's p-value and rank outer means the ranking of interactions. Total means the sum of within and outer.

Results

Crosstalk of pJIA related pathways

The PCNs of background and pJIA were generated with gene expression profiling of 61 pJIA patients and 59 controls, respectively. The detail of PCNs was shown in the supplemental material. The crosstalk difference of background and pJIA are shown in Figure 1. In the control group, a majority of degrees in 300 pathways were between 255 and 300. The pJIA group was significantly different, with degrees lower than 180. This result gives evidence for the strong relationship between these pathways with pJIA.

Bigger value scores indicated more important crosstalk pathways. The top three important pathways were hepatitis B (KEGG ID: 05161), viral carcinogenesis (KEGG ID: 05203), and

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS] [Index Copernicus]

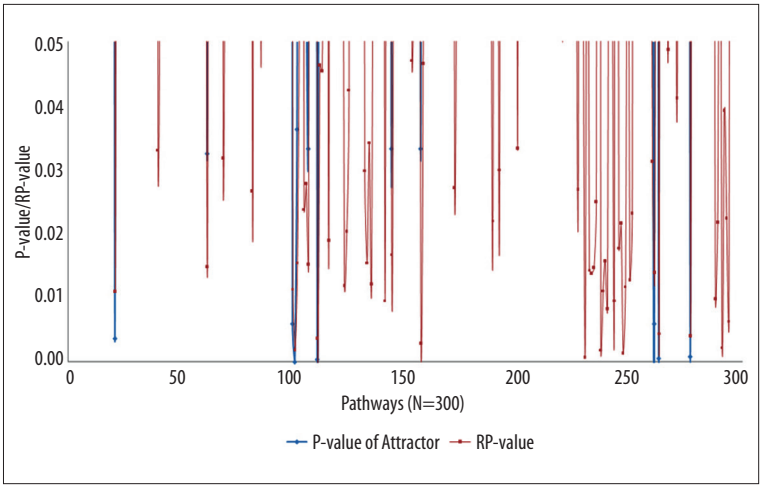

Figure 2. 300 KEGG pathways were evaluated by Kauffman' attractor and RP-value.

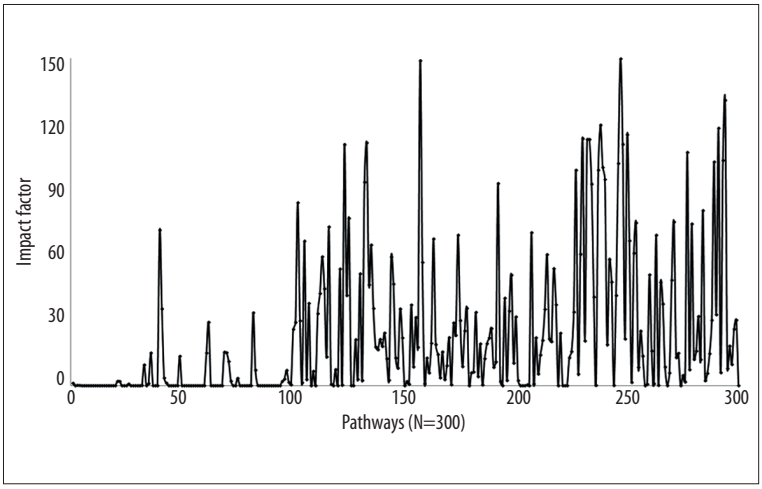

Figure 3. Interactions inter-pathways were assessed by impact factor.

pathways in cancer (KEGG ID: 05200). They provided valuable information for the mechanism of the pJIA.

Identification of KEGG pathways

300 KEGG pathways were evaluated comprehensively by Kauffman' attractor, impact factor and RP-value. There were 12 attractors with $p<0.05$ and 7 attractors with $p<0.01$ (Table 1, Figure 2); the 12 attractors were significantly different in the peripheral blood mononuclear cells (PBMC). There were some molecular alterations in the pathways themselves related to pJIA, which showed they were differentially expressed pathways.

In terms of interactions inter-pathways, impact factor was used to assess their contact. There were different values that varied from 0 to 150, shown in Figure 3, which indicated that there were different degrees of interactions inter-pathways.

RP-value was used to comprehensively assess 300 pathways, including within pathways and inter-pathways. There were 60 pathways with RP <0.05 and 14 pathways with RP <0.01 (Figure 2). Finally, screened from 12 attractors, two pathways matched with attractor $p<0.05$, big value of impact factor and RP-value <0.05. They were the p53 signaling pathway (KEGG ID: 04115) and the non-alcoholic fatty liver disease (NAFLD) (KEGG ID: 04932) pathway, which were considered to be significantly dysfunctional pathways with strong interactions. These pathways might play more important roles in the development of pJIA due to their dysfunctional expression and strong interactions.

Discussion

Attractor theory is well known as a knowledge-driven analytical way to distinguish and annotate gene-sets [14]. It was used to evaluate expression within pathways in embryonic stem cells [15]. The results are more complete than traditional DEG analysis due to narrowing the number of correlated dysregulated pathways.

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS] [Index Copernicus]

In our study, 12 attractors ($p<0.05$) with statistically significant alterations were screened. While they were differentially expressed pathways in patients, the integral influence to the system was absent. From Figure 2, we can see that the variation trend of *p*-values of attractor were not absolutely consistent with that of RP-value. Therefore, it was necessary to employ crosstalk to assess interactions across pathways. Pathways with big impact factor values were considered to have strong contact with other pathways. We found that three attractors showed big impact factor values, but nine attractors did not (impact factor <80). This suggests that pathways screened by attractors were not exactly dysregulated and influential ones. Those pathways with $p<0.05$ and small impact factor values were considered to have small effects, which should be filtered.

Our results suggested attractors might fail to identify pathways effectively because of incomplete information on inherent interdependency inter-pathways. This is similar to the challenge faced by other pathway-identification methods that apply topological pathway information [24]. After assessing the interactions inter-pathways by crosstalk, our novel approach enhanced attractors to identify dysregulated pathways. Recently, methods to comprehensively identify dysregulated pathways have become a major focus of study [9]. This novel method that combined attractor and crosstalk is hoped to be further applied to other diseases.

We applied RP-values to evaluate the comprehensive identified ability both within pathways and inter-pathways. What we wanted was influential dysregulated pathways with attractor $p<0.05$, big impact factor value and RP-value <0.05. From the screening of 12 attractors, two pathways matched the conditions: p53 and NAFLD pathways. The p53 signaling pathway (KEGG ID: 04115) is important in many diseases [28], in which p53 activation is induced by DNA damage, oxidative stress, and activated oncogenes [29]. The NAFLD (KEGG ID: 04932) has been identified as a significant pathway of pJIA. Previous studies have reported that JIA is related to hepatomegaly [30]. Therefore, these two pathways are likely to be important in the molecular mechanism of pJIA.

Conclusions

In our study, a novel approach that identified the dysregulated pathways of pJIA was constructed. It was based on attractor of within-pathway effects and crosstalk inter-pathways. This constructed process is expected to be efficient in the upcoming era of medicine.

References:

1. Petty RE, Southwood TR, Manners P et al: International League of Associations for Rheumatology classification of juvenile idiopathic arthritis: Second revision, Edmonton, 2001. J Rheumatol, 2004; 31(2): 390–92
2. Webb K, Wedderburn LR: Advances in the treatment of polyarticular juvenile idiopathic arthritis. Curr Opin Rheumatol, 2015; 27(5): 505–10
3. Goldmuntz EA, White PH: Juvenile idiopathic arthritis: A review for the pediatrician. Pediatr Rev, 2006; 27(4): e24–32
4. Keller H, Muller LM, Markic G et al: Is early TMJ involvement in children with juvenile idiopathic arthritis clinically detectable? Clinical examination of the TMJ in comparison with contrast enhanced MRI in patients with juvenile idiopathic arthritis. Pediatr Rheumatol Online J, 2015; 13(1): 56
5. Huang H, Qian X, Yu H et al: Clinical analysis in 202 children with juvenile idiopathic arthritis. Clin Rheumatol, 2013; 32(7): 1021–27
6. Li RJ, Tang XM, Liu W et al: [Expression and significance of interleukin-6, interferon-inducible protein-10 and interleukin-17 in serum and synovial fluid of patients with juvenile idiopathic arthritis.] Zhonghua Er Ke Za, 2013, 51(6): 472–76 [in Chinese]
7. Gorska A, Nawrocki A, Urban M, Florys B: Composition of phospholipid fatty acids in erythrocyte membranes of children with chronic juvenile arthritis: Clinical and biochemical correlations. Med Sci Monit, 2000, 6(1): 30–39
8. Sediva A, Hoza J, Nemcova D et al: Immunological investigation in children with juvenile chronic arthritis. Med Sci Monit, 2001; 7(1): 99–104
9. Han J, Li C, Yang H et al: A novel dysregulated pathway-identification analysis based on global influence of within-pathway effects and crosstalk between pathways. J R Soc Interface, 2015; 12(102): 20140937
10. Burmester GR, Blanco R, Charles-Schoeman C et al: Tofacitinib (CP-690,550) in combination with methotrexate in patients with active rheumatoid arthritis with an inadequate response to tumour necrosis factor inhibitors: A randomised phase 3 trial. Lancet, 2013; 381(9865): 451–60
11. Lee EB, Fleischmann R, Hall S et al: Tofacitinib versus methotrexate in rheumatoid arthritis. N Engl J Med, 2014; 370(25): 2377–86
12. van der Heijde D, Tanaka Y, Fleischmann R et al: Tofacitinib (CP-690,550) in patients with rheumatoid arthritis receiving methotrexate: Twelve-month data from a twenty-four-month phase III randomized radiographic study. Arthritis Rheum, 2013; 65(3): 559–70
13. van Vollenhoven RF, Fleischmann R, Cohen S et al: Tofacitinib or adalimumab versus placebo in rheumatoid arthritis. N Engl J Med, 2012; 367(6): 508–19
14. Kauffman S: A proposal for using the ensemble approach to understand genetic regulatory networks. J Theor Biol, 2004; 230(4): 581–90
15. Mar JC, Matigian NA, Quackenbush J, Wells CA: attract: A method for identifying core pathways that define cellular phenotypes. PloS One, 2011; 6(10): e25445
16. Li Y, Agarwal P, Rajagopalan D: A global pathway crosstalk network. Bioinformatics, 2008; 24(12): 1442–47
17. Parkinson H, Kapushesky M, Shojatalab M et al: ArrayExpress – a public database of microarray experiments and gene expression profiles. Nucleic Acids Res, 2007; 35(Database issue): D747–50
18. Griffin TA, Barnes MG, Ilowite NT et al: Gene expression signatures in polyarticular juvenile idiopathic arthritis demonstrate disease heterogeneity and offer a molecular classification of disease subsets. Arthritis Rheum, 2009; 60(7): 2113–23
19. Gautier L, Cope L, Bolstad BM, Irizarry RA: Affy – analysis of Affymetrix GeneChip data at the probe level. Bioinformatics, 2004; 20(3): 307–15
20. Smyth GK, Gentleman R, Carey V et al: Limma: Linear models for microarray data. In Bioinformatics and Computational Biology Solutions Using R and Bioconductor. vol. 43. New York: Springer; 2005; vol. 7: 397–420
21. Irizarry RA, Hobbs B, Collin F et al: Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics; 2003, 4(2): 249–64
22. Kanehisa M, Goto S: KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res, 2000; 28(1): 27–30

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS] [Index Copernicus]

23. Rahnenfuhrer J, Domingues FS, Maydt J, Lengauer T: Calculating the statistical significance of changes in pathway activity from gene expression data. Stat Appl Genet Mol Biol, 2004; 3: Article16
24. Hung JH, Whitfield TW, Yang TH et al: Identification of functional modules that correlate with phenotypic difference: the influence of network topology. Genome Biol, 2010; 11(2): R23
25. Franceschini A, Szklarczyk D, Frankild S et al: STRING v9.1: Protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res, 2013; 41(Database issue): D808–15
26. Al-Shahrour F, Diaz-Uriarte R, Dopazo J: FatiGO: A web tool for finding significant associations of Gene Ontology terms with groups of genes. Bioinformatics, 2004; 20(4): 578–80
27. Benjamini Y, Drai D, Elmer G et al: Controlling the false discovery rate in behavior genetics research. Behavioural Brain Res, 2001; 125(1–2): 279–84
28. Sun Y YK, Zhang P, Ma R et al: Crosstalk analysis of pathways in breast cancer using a network model based on overlapping differentially expressed genes. Exp Ther Med, 2015; 10(2): 743–48
29. Levine AJ, Hu W, Feng Z: The P53 pathway: What questions remain to be explored? Cell Death Differ, 2006; 13(6): 1027–36
30. Bedoya ME, Ceccato F, Paira S: Spleen and liver enlargement in a patient with rheumatoid arthritis. Reumatol Clin, 2015; 11(4): 227–31

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS] [Index Copernicus]