Research Paper

Comprehensive analysis of gene expression and DNA methylation datasets identify valuable biomarkers for rheumatoid arthritis progression

Gang Fang[1], Qing Huai Zhang[1], Qianqian Tang[2], Zuling Jiang[3], Shasha Xing[2], Jianying Li[1] and Yuzhou Pang[1]

[1]Laboratory of Zhuang Medicine Prescriptions Basis and Application Research, Guangxi University of Chinese Medicine, Nanning, China

[2]Department of Rheumatism, Ruikang Hospital Affiliated to Guangxi University of Chinese Medicine, Nanning, China

[3]Department of Zhuang Medicine, The First Affiliated of Guangxi University of Chinese Medicine, Nanning, China

Correspondence to: Yuzhou Pang, *email:* 2017pangyuzhou@sina.com

ABSTRACT

Rheumatoid arthritis (RA) represents a common systemic autoimmune disease which lays chronic and persistent pain on patients. The purpose of our study is to identify novel RA-related genes and biological processes/pathways. All the datasets of this study, including gene expression and DNA methylation datasets of RA and OA samples, were obtained from the free available database, i.e. Gene Expression Omnibus (GEO). We firstly identified the differentially expressed genes (DEGs) between RA and OA samples through the limma package of R programming software followed by the functional enrichment analysis in the Database for Annotation, Visualization and Integrated Discovery (DAVID) for the exploring of potential involved biological processes/pathways of DEGs. For DNA methylation datasets, we used the IMA package for their normalization and identification of differential methylation genes (DMGs) in RA compared with OA samples. Comprehensive analysis of DEGs and DMGs was also conducted for the identification of valuable RA-related biomarkers. As a result, we obtained 394 DEGs and 363 DMGs in RA samples with the thresholds of $|log2fold\ change| > 1$ and p-value < 0.05, and $|delta\ beta| > 0.2$ and p-value < 0.05 respectively. Functional analysis of DEGs obtained immune and inflammation associated biological processes/pathways. Besides, several valuable biomarkers of RA, including BCL11B, CCDC88C, FCRLA and APOL6, were identified through the integrated analysis of gene expression and DNA methylation datasets. Our study should be helpful for the development of novel drugs and therapeutic methods for RA.

INTRODUCTION

Rheumatoid arthritis (RA) is a common inflammatory disorder induced disease with a morbidity of ~1% [1]. RA is characterized by the chronic arthritis with the result of irreversible joint damage, and put persistent pain on patients [2]. Besides, RA is a multi factor related disease, which is influenced by age, race, territory, smoking and so on [3]. Marked improvement has been achieved in the classification and treatment of RA, several

biologic response-modifying drugs were also proposed for the inhibition of the inflammatory response, however, most patients don't obtained any remission [4–7]. While, the poor understanding of its mechanisms has limited the development of effective drugs and therapeutic methods.

DNA methylation plays an important role in the development of complex diseases. It has been well known that hper-methylation of tumor suppressor genes' promoter as a crucial factor for progression of cancers [8–10]. The rapid development of high-throughput

sequencing and gene microarray accelerate the research of DNA modification, including DNA methylation, which is closely associated with RA [11–15]. Through genome-wide DNA methylation profiling in RA, Glossop *et al*. [16] identified RA-related methylation changes that distinct in T- and B-lymphocyte populations. Yuan *et al*. [17] also reported important roles of DNA methylation in fibroblast-like synoviocytes for RA progression. Gene expression could be silenced by DNA methylation through adding methyl to cytosine of CG site and prevent binding of transcription factor to specific regions. So combined analysis of gene expression and DNA methylation profiles should be helpful for the understanding of disease mechanisms. There have been many this type studies, particular in cancers, such as Bapat *et al*. conducted integrated analysis of epigenomic and genomic changes by DNA methylation and obtained several novel biomarkers for prostate cancer [18]. Haller also identified SPP1 as an independent prognostic factor for gastrointestinal stromal tumors through combined DNA methylation and gene expression profiling [19]. While, there has none this type of study for RA.

Here, we proposed to obtain novel methylation and expression signatures for RA through the combined analysis of DNA methylation and gene expression datasets from GEO. Intersect analysis obtained several genes that differential expression and methylation in RA samples simultaneously, which should be potential targets. Functional analysis indicated biological processes/pathways participate in the development of RA, including immune and inflammatory related processes.

RESULTS

DEGs and DMGs Figure 1A and 1B illustrated overall expression profiles in all samples before and after normalization. Significant improvement in the comparable of the expression profiles among all samples was observed, which indicated that the normalized expression values are suitable for the following analysis. We totally obtained 313 DEGs in RA samples compared with OA samples, which including 232 up-regulated genes and 81 down-regulated genes. Hierarchical clustering of DEGs and samples were shown in Figure 1C, in which the control and case samples were clustered into their own groups. Furthermore, 772 differential methylated sites, which correspond to 363 genes, were identified.

Enriched functions

Functional enrichment analysis identified 60 GO terms satisfied the thresholds of *P*-value < 0.05, which are mainly associated with immune and inflammation. Cluster analysis of GO terms through the enrichment MAP plugin [25] of Cytoscape [26] was shown in Figure 2, in which

node size represents gene number contained in the GO term, color represents significance of GO term. Besides, 4 immune and inflammatory related pathways, including primary immunodeficiency, cytokine-cytokine receptor interaction, hematopoietic cell lineage and T cell receptor signaling pathway were also significantly enriched in the DEGs.

Integrated analysis of DEGs and DMGs

Intesected analysis of DEGs and DMGs identified 7 overlaps, among which 4 genes were up-regulated and hypo-methylated simultaneously, i.e. APOL6, BCL11B, CCDC88C and FCRLA. Figure 3 illustrated the fold change (log2 scale) of those 7 overlaps.

DISCUSSION

RA is a multifactorial disease which is also closely associated with genetic. In this study, we performed DNA methylation and gene expression analysis of RA and oA samples to identify potential biomarkers and biological processes/pathways involved in the progression of RA. This should provide comprehensive landscape for RA which is helpful for the development of its diagnosis and treatment.

Functional enrichment analysis of DEGs in RA compared with OA samples identified significantly enriched GO terms and KEGG pathways that are closely related to inflammatory and immune response. It is easy to interpret this results for the close association between inflammatory and immune with RA progression [27–30]. Accordingly, many anti-inflammatory drugs were also developed, such as morin [31], JAK inhibitors [32], TNF inhibitors [33], and so on. Al-Okbi *et al*. even thought that the anti-inflammatory activity of nutraceuticals as a complementary therapy for RA for the severe side effects of anti-inflammatory drugs [34]. Screening of specific targets for drugs has also been an important spot for the treatment of RA.

DNA methylation represents one of the most common epigenetics that could induce gene expression repression through prevent transcription factor from binding the target regions. It is believed that DNA methylation play an important role in RA and several valuable methylation signatures of RA were also obtained [14, 35, 36]. In contrast to most of cancers, in which hyper-methylation is a widespread character, RA tends to be hypo-methylation in its associated tissues and cells [37]. In this study, we obtained 4 genes, i.e. FCRLA, CCDC88C, BCL11B and APOL6, that found to be hypo-methylated and down-regulated in RA compared with OA samples simultaneously. The differential methylation sites in CCDC88C located in promoter region, which should indicate that CCDC88C is more likely participate

in the progression of RA. CCDC88C (coiled-coil domain containing 88C) encodes a member of the hook-related proteins that involved in the regulation of the Wnt signaling pathway [38], which could control inflammatory response induced by multiple factors, such as pathogenic bacteria, Toll-like receptor [39, 40]. So we inferred that CCDC88C could influence RA development through Wnt signaling pathway. There is no study about the direct role of APOL6 in RA, while it has been reported that APOL6 could induce immune response in HIV-associated
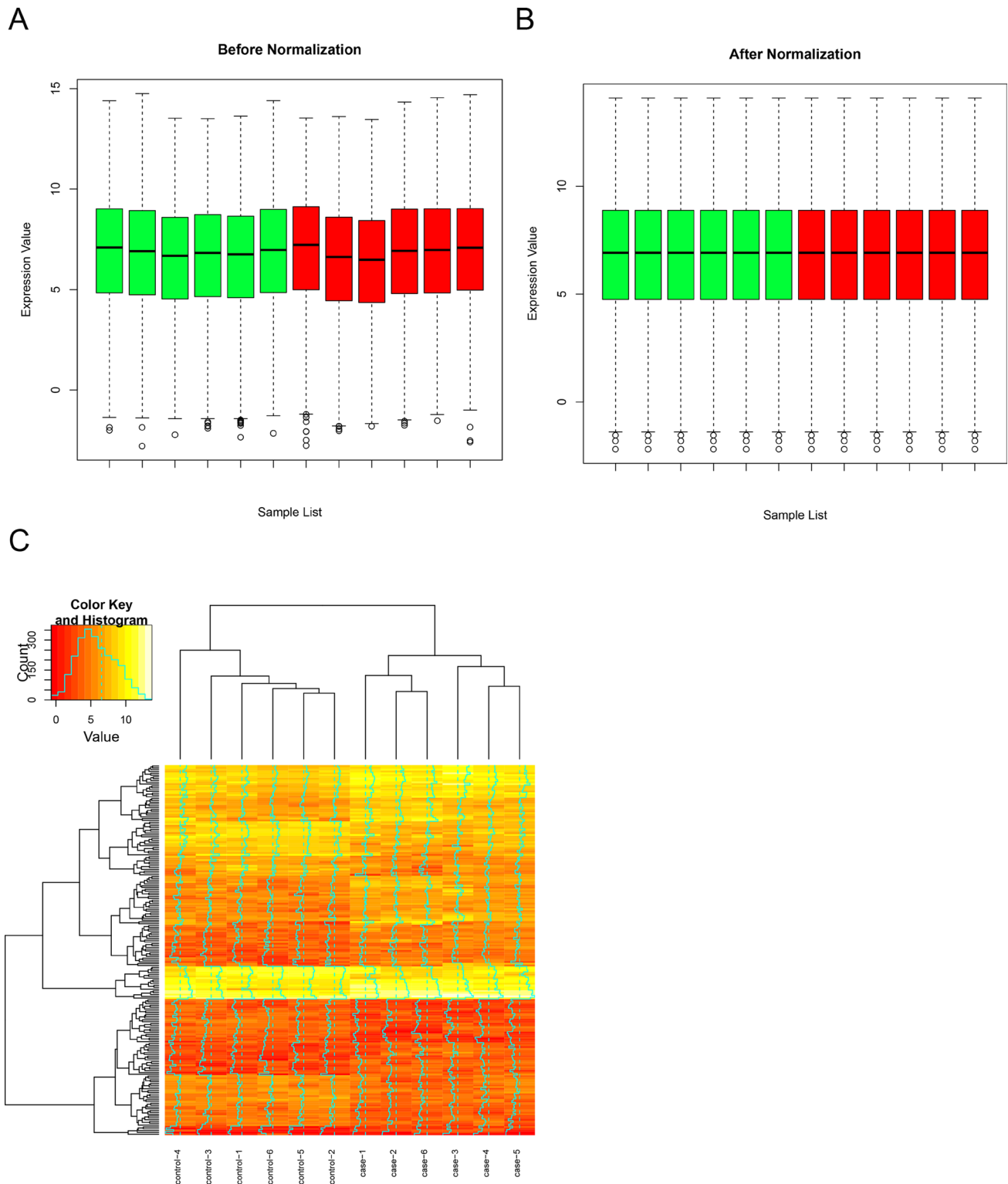
Figure 1: Gene expression microarray analysis. Overall expression profiles before (**A**) and after (**B**) normalization. (**C**) Hierarchical clustering of DEGs and samples by euclidean distance. RA and OA samples were clustered into their own group respectively.
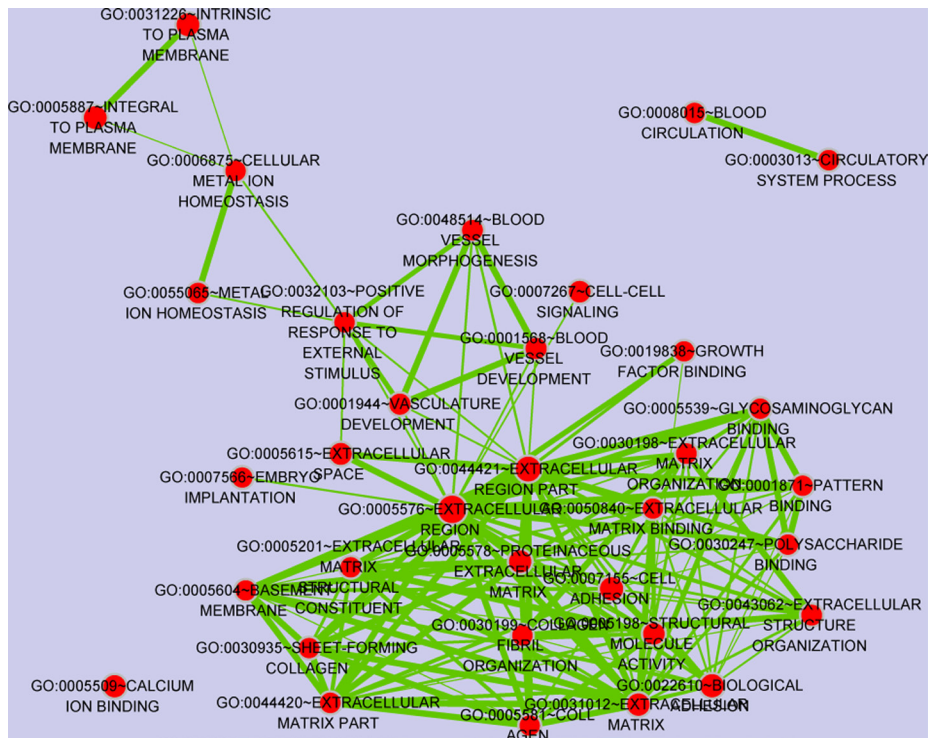
Figure 2: GO terms clustering via the enrichment map plugin of cytoscape software. Links between any two GO terms indicates overlapping genes and more thick indicates more overlaps. Node size represents gene number contained in the GO terms and color represents significance.
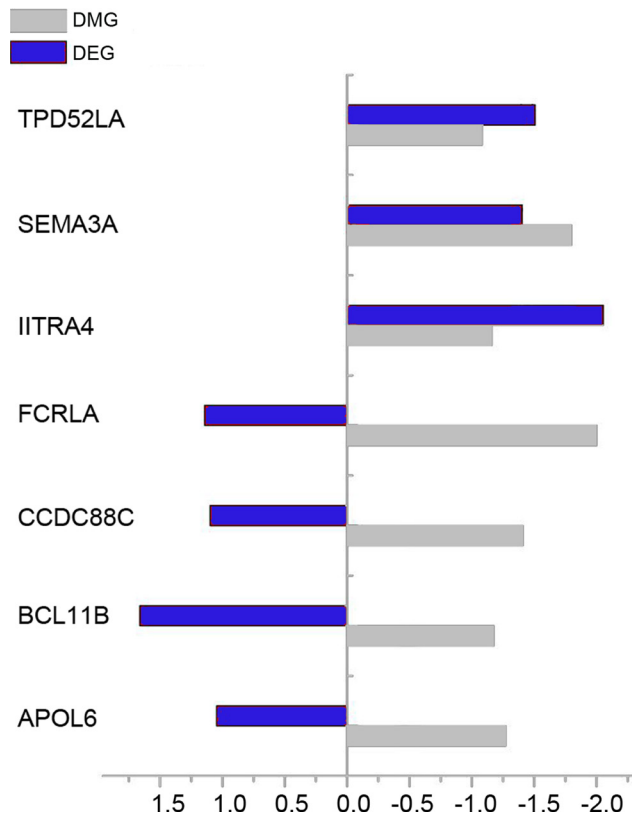
Figure 3: The fold change (log2 scale) of DEGs and DMGs in RA compared with OA samples. Blue bar and gray bar represent DEG and DMG respectively.

neurocognitive disorders [41]. Besides, APOL6 is also an important apoptosis related protein, which is critical for the progression of RA, so APOL6 should be a novel biomarker in RA.

MATERIALS AND METHODS

Microarray datasets

The gene expression profile data was downloaded from GEO with the accession number of GSE36700 that deposited by Toukap [20]. A total of 12 samples were contained, including 5 osteoarthritis (OA) samples and 7 rheumatoid arthritis (RA) samples. The commercial microarray, GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array was used for the quantification of mRNA abundance. The methylation profiles datasets, (accession number: GSE46650), were downloaded from GEO, which was deposited by Rica [21], which including 6 OA samples and 6 RA samples. GPL13534 Illumina Human Methylation 450 BeadChip (HumanMethylation450_15017482) was used for the detection of DNA methylation level and determination of detection P-value.

Data preprocessing

For gene expression datasets, preprocessing of raw microarray, including background correction, quantile normalization and probe summarization of the raw microarray data were all conducted by the affy package [22] of R programming software. For methylation microarray datasets, we performed preprocessing through the IMA package. After beta value normalization and peak correction, the CpG sites with detection P-value > 0.05 in more than 75% samples and samples contained more than 75% CpG sites were filtered out. Besides, we mapped the CpG sites to genes through the annotation package.

Identification of DEGs and DMGs

After preprocessing of raw datasets, DEGs in RA samples compared with OA samples were obtained through the limma package [23] with the thresholds of |log2fold change| > 1 and P-value < 0.05. For DNA methylation datasets, we obtained the differentially methylation sites (DMS) firstly through the IMA package with the thresholds of |delta beta| > 0.2 and P-value < 0.05. Then, the DMGs were obtained through mapping CpG sites to genes.

Functional enrichment analysis of DEGs

To explore enriched functions of DEGs, we conducted functional enrichment analysis of DEGs through the Database for Annotation, Visualization and Integrated Discovery, (DAVID) [24]. GO terms and KEGG pathways with FDR < 0.05 was obtained as the enriched functions of DEGs. analysis.

CONCLUSIONS

In conclusion, this study conducted combined analysis of DNA methylation and gene expression profiles and identified valuable biological processes/pathways and biomarkers for RA. Further studies are still needed for the confirmation of their roles in RA.

Author contributions

The author(s) conducted the accelerator design, analyzed the results and wrote the manuscript and doing all things!

CONFLICTS OF INTEREST

The author declare no conflict of interest.

FUNDING

REFERENCES

1. Assayag D, Lee JS, King TE Jr. Rheumatoid arthritis associated interstitial lung disease: a review. Medicina (B Aires). 2014; 74:158–65.

2. Verheul MK, Fearon U, Trouw LA, Veale DJ. Biomarkers for rheumatoid and psoriatic arthritis. Clin Immunol. 2015; 161:2–10.

3. Scott DL, Wolfe F, Huizinga TW. Rheumatoid arthritis. Lancet. 2010; 376:1094–108.

4. Koopman FA, Schuurman PR, Vervoordeldonk MJ, Tak PP. Vagus nerve stimulation: a new bioelectronics approach to treat rheumatoid arthritis? Best Pract Res Clin Rheumatol. 2014; 28:625–35.

5. Davis JM 3rd, Matteson EL. My treatment approach to rheumatoid arthritis. Mayo Clin Proc. 2012; 87:659–73.

6. Metsios GS, Stavropoulos-Kalinoglou A, Kitas GD. The role of exercise in the management of rheumatoid arthritis. Expert Rev Clin Immunol. 2015; 11:1121–30.

7. Wasserman AM. Diagnosis and management of rheumatoid arthritis. Am Fam Physician. 2011; 84:1245–52.

8. Henriksen SD, Madsen PH, Krarup H, Thorlacius-Ussing O. DNA Hypermethylation as a Blood-Based Marker for Pancreatic Cancer: A Literature Review. Pancreas. 2015; 44:1036–45.

9. Hwang JA, Lee BB, Kim Y, Hong SH, Kim YH, Han J, Shim YM, Yoon CY, Lee YS, Kim DH. HOXA9 inhibits migration of lung cancer cells and its hypermethylation is associated with recurrence in non-small cell lung cancer. Mol Carcinog. 2015; 54:E72–80.

10. Hubers AJ, Brinkman P, Boksem RJ, Rhodius RJ, Witte BI, Zwinderman AH, Heideman DA, Duin S, Koning R, Steenbergen RD, Snijders PJ, Smit EF, Sterk PJ, et al. Combined sputum hypermethylation and eNose analysis for lung cancer diagnosis. J Clin Pathol. 2014; 67:707–11.

11. Bottini N, Firestein GS. Epigenetics in rheumatoid arthritis: a primer for rheumatologists. Curr Rheumatol Rep. 2013; 15:372.

12. Frank-Bertoncelj M, Gay S. The epigenome of synovial fibroblasts: an underestimated therapeutic target in rheumatoid arthritis. Arthritis Res Ther. 2014; 16:117.

13. Klein K, Gay S. Epigenetic modifications in rheumatoid arthritis, a review. Curr Opin Pharmacol. 2013; 13:420–5.

14. Klein K, Gay S. Epigenetics in rheumatoid arthritis. Curr Opin Rheumatol. 2015; 27:76–82.

15. Glant TT, Mikecz K, Rauch TA. Epigenetics in the pathogenesis of rheumatoid arthritis. BMC Med. 2014; 12:35.

16. Glossop JR, Emes RD, Nixon NB, Haworth KE, Packham JC, Dawes PT, Fryer AA, Mattey DL, Farrell WE. Genome-wide DNA methylation profiling in rheumatoid arthritis identifies disease-associated methylation changes that are distinct to individual T- and B-lymphocyte populations. Epigenetics. 2014; 9:1228–37.

17. Yuan FL, Li X, Xu RS, Jiang DL, Zhou XG. DNA methylation: roles in rheumatoid arthritis. Cell Biochem Biophys. 2014; 70:77–82.

18. White-Al Habeeb NM, Ho LT, Olkhov-Mitsel E, Kron K, Pethe V, Lehman M, Jovanovic L, Fleshner N, van der Kwast T, Nelson CC, Bapat B. Integrated analysis of epigenomic and genomic changes by DNA methylation dependent mechanisms provides potential novel biomarkers for prostate cancer. Oncotarget. 2014; 5:7858–69. https://doi.org/10.18632/oncotarget.2313.

19. Haller F, Zhang JD, Moskalev EA, Braun A, Otto C, Geddert H, Riazalhosseini Y, Ward A, Balwierz A, Schaefer IM, Cameron S, Ghadimi BM, Agaimy A, et al. Combined DNA methylation and gene expression profiling in gastrointestinal stromal tumors reveals hypomethylation of SPP1 as an independent prognostic factor. Int J Cancer. 2015; 136:1013–23.

20. Nzeusseu Toukap A, Galant C, Theate I, Maudoux AL, Lories RJ, Houssiau FA, Lauwerys BR. Identification of distinct gene expression profiles in the synovium of patients with systemic lupus erythematosus. Arthritis Rheum. 2007; 56:1579–88.

21. de la Rica L, Urquiza JM, Gómez-Cabrero D, Islam AB, López-Bigas N, Tegnér J, Toes RE, Ballestar E. Identification of novel markers in rheumatoid arthritis through integrated analysis of DNA methylation and microRNA expression. J Autoimmun. 2013; 41:6–16.

22. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy—analysis of Affymetrix GeneChip data at the probe level. Bioinformatics. 2004; 20:307–15.

23. Diboun I, Wernisch L, Orengo CA, Koltzenburg M. Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma. BMC Genomics. 2006; 7:252.

24. Sherman BT, Huang da W, Tan Q, Guo Y, Bour S, Liu D, Stephens R, Baseler MW, Lane HC, Lempicki RA. DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. BMC Bioinformatics. 2007; 8:426.

25. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One, 2010. 5:e13984.

26. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13:2498–504.

27. Siebert S, Tsoukas A, Robertson J, McInnes I. Cytokines as therapeutic targets in rheumatoid arthritis and other inflammatory diseases. Pharmacol Rev. 2015; 67:280–309.

28. Kirkham BW, Kavanaugh A, Reich K. Interleukin-17A: a unique pathway in immune-mediated diseases: psoriasis, psoriatic arthritis and rheumatoid arthritis. Immunology. 2014; 141:133–42.

29. Yoshida K, Hashimoto T, Sakai Y, Hashiramoto A. Involvement of the circadian rhythm and inflammatory cytokines in the pathogenesis of rheumatoid arthritis. J Immunol Res. 2014; 2014:282495.

30. Benedetti G, Miossec P. Interleukin 17 contributes to the chronicity of inflammatory diseases such as rheumatoid arthritis. Eur J Immunol. 2014; 44:339–47.

31. Sultana F, Rasool M. A novel therapeutic approach targeting rheumatoid arthritis by combined administration of morin, a dietary flavanol and non-steroidal anti-inflammatory drug indomethacin with reference to pro-inflammatory cytokines, inflammatory enzymes, RANKL and transcription factors. Chem Biol Interact. 2015; 230:58–70.

32. Norman P. Selective JAK inhibitors in development for rheumatoid arthritis. Expert Opin Investig Drugs. 2014; 23:1067–77.

33. Kimura K, Takayanagi R, Yokoyama H, Yamada Y. Theory-based analysis of the anti-inflammatory effect of TNF inhibitors on rheumatoid arthritis. Drug Metab Pharmacokinet. 2014; 29:272–7.

34. Al-Okbi SY. Nutraceuticals of anti-inflammatory activity as complementary therapy for rheumatoid arthritis. Toxicol Ind Health. 2014; 30:738–49.

35. Viatte S, Plant D, Raychaudhuri S. Genetics and epigenetics of rheumatoid arthritis. Nat Rev Rheumatol. 2013; 9:141–53.

36. Plant D, Webster A, Nair N, Oliver J, Smith SL, Eyre S, Hyrich KL, Wilson AG, Morgan AW, Isaacs JD, Worthington J, Barton A. Differential Methylation as a Biomarker of Response to Etanercept in Patients With Rheumatoid Arthritis. Arthritis Rheumatol. 2016; 68:1353–60.

37. Karouzakis E, Gay RE, Michel BA, Gay S, Neidhart M. DNA hypomethylation in rheumatoid arthritis synovial fibroblasts. Arthritis Rheum. 2009; 60:3613–22.

38. Enomoto A, Ping J, Takahashi M. Girdin, a novel actin-binding protein, and its family of proteins possess versatile functions in the Akt and Wnt signaling pathways. Ann N Y Acad Sci. 2006; 1086:169–84.

39. Silva-Garcia O, Valdez-Alarcon JJ, Baizabal-Aguirre VM. The Wnt/beta-catenin signaling pathway controls the inflammatory response in infections caused by pathogenic bacteria. Mediators Inflamm. 2014; 2014:310183.

40. Trinath J, Holla S, Mahadik K, Prakhar P, Singh V, Balaji KN. The WNT signaling pathway contributes to dectin-1-dependent inhibition of Toll-like receptor-induced inflammatory signature. Mol Cell Biol. 2014; 34:4301–14.

41. Siangphoe U, Archer KJ. Gene Expression in HIV-Associated Neurocognitive Disorders: A Meta-Analysis. J Acquir Immune Defic Syndr. 2015; 70:479–88.