Journal of Genetics, Vol. 97, No. 1, March 2018, pp. 173–178
https://doi.org/10.1007/s12041-018-0897-9

RESEARCH ARTICLE

Investigating multiple dysregulated pathways in rheumatoid arthritis based on pathway interaction network

XIAN-DONG SONG[1], XIAN-XU SONG[2], GUI-BO LIU[3], CHUN-HUI REN[4], YUAN-BO SUN[5], KE-XIN LIU[1], BO LIU[1], SHUANG LIANG[4] and MIN ZHU[4]*

[1] *Department of Orthopaedics, Hongqi Hospital of Mudanjiang Medical University, Mudanjiang 157000, Heilongjiang, People's Republic of China*
[2] *Department of General Surgery, Second Affiliated Hospital of Mudanjiang Medical University, Mudanjiang 157000, Heilongjiang, People's Republic of China*
[3] *Department of Anatomy, Mudanjiang Medical University, Mudanjiang 157000, Heilongjiang, People's Republic of China*
[4] *Department of MRI, Hongqi Hospital of Mudanjiang Medical University, Mudanjiang 157000, Heilongjiang, People's Republic of China*
[5] *Department of Kidney Internal Medicine, Hongqi Hospital of Mudanjiang Medical University, Mudanjiang 157000, Heilongjiang, People's Republic of China*
*For correspondence. E-mail: zhuminbio@126.com.

Received 2 March 2017; revised 11 July 2017; accepted 13 July 2017; published online 9 March 2018

Abstract. The traditional methods of identifying biomarkers in rheumatoid arthritis (RA) have focussed on the differentially expressed pathways or individual pathways, which however, neglect the interactions between pathways. To better understand the pathogenesis of RA, we aimed to identify dysregulated pathway sets using a pathway interaction network (PIN), which considered interactions among pathways. Firstly, RA-related gene expression profile data, protein–protein interactions (PPI) data and pathway data were taken up from the corresponding databases. Secondly, principal component analysis method was used to calculate the pathway activity of each of the pathway, and then a seed pathway was identified using data gleaned from the pathway activity. A PIN was then constructed based on the gene expression profile, pathway data, and PPI information. Finally, the dysregulated pathways were extracted from the PIN based on the seed pathway using the method of support vector machines and an area under the curve (AUC) index. The PIN comprised of a total of 854 pathways and 1064 pathway interactions. The greatest change in the activity score between RA and control samples was observed in the pathway of epigenetic regulation of gene expression, which was extracted and regarded as the seed pathway. Starting with this seed pathway, one maximum pathway set containing 10 dysregulated pathways was extracted from the PIN, having an AUC of 0.8249, and the result indicated that this pathway set could distinguish RA from the controls. These 10 dysregulated pathways might be potential biomarkers for RA diagnosis and treatment in the future.

Keywords. rheumatoid arthritis; dysregulated pathways; pathway interaction network.

Introduction

Rheumatoid arthritis (RA) is a chronic inflammatory joint disease, characterized by the pain, swollen joints and decrease in physical activity (Michaud and Wolfe 2007). Usually, patients with RA are at a high risk of mortality, with a decrease in survival rate by 3–10 years (Alamanos and Drosos 2005). Although many biological agents have been developed to treat this disease, the curative treatment is still not discovered (Andersen *et al.* 2016). Moreover, immune-modulatory properties of drugs also increase the risk of potential adverse events, such as nosocomial infections, congestive heart failure, and malignancy, which put a significant burden on the healthcare systems (Yamada *et al.* 2016). The risk of adverse effects and expensive treatment for RA patients have driven to seek for predictive signatures that can be used to detect and treat early stages of RA.

Therefore, several scholars focussed on the differentially expressed genes (DEGs) (Diogo *et al.* 2013) to reveal the aetiology and explore the effective treatment modalities for RA. Of note, DEGs are generally detected on the basis

of the given statistical criteria, for example, *p*-value, *q*-value and so on (Hu *et al.* 2009). Thus,, genes having the smallest *p*-values are selected from all significant genes. Nevertheless, biological experts realize that the magnitude of differential expression dose not prove the biological significance (Hu *et al.* 2009). Accordingly, these approaches are far away from perfect, since a set of genes interact with each other, rather than independent entities (Glazier *et al.* 2002; Merikangas *et al.* 2006). Further, for the same disorder, many of the gene biomarkers extracted in one dataset have not been found to work efficiently in another dataset (Braga-Neto and Dougherty 2004). Based on the poor performance of gene biomarkers, many scholars focussed on the identification of pathways associated with disease, because when the pathways are considered as biomarkers, they would improve the accuracy compared with gene signatures. For example, Miao *et al.* (2013) found that Wnt signalling pathway played an important role in the pathogenesis of RA. Traditional pathway analysis mainly focussed on the single dysregulated pathways, but the most important point neglected was the interactions between them (Khatri *et al.* 2012). Generally, the functions related pathways are interconnected with each other, and any of the dysregulated pathways could affect the activities of the other. Remarkably, detecting and understanding the interactions between pathways are beneficial for exploring the molecular mechanisms of the disease (Li and Agarwal 2009). Notably, network-based biology is broadly used to analyse interactions, in turn, further shedding lights on the action mechanism of how cellular systems operate (Barabasi and Oltvai 2004; Xia *et al.* 2004). As a result, a pathway interaction network (PIN) was built to identify dysregulated pathways that considered the interactions between pathways (Liu *et al.* 2012).

Here, a computational method was used to detect significant pathways by constructing a PIN. The findings demonstrate that this method is beneficial in predicting biomarkers and even drug targets for RA in a more robust fashion.

Materials and methods

Retrieving datasets and preprocessing

***Gene expression profile data*:** In this study, the gene expression profile data (E-GEOD-57405) (Rosenberg *et al.* 2014) were obtained from the ArrayExpress database, which were deposited in the A-MEXP-1171-Illumina HumanHT-12v3.0 Expression BeadChip platform. E-GEOD-57405 comprised of 46 samples, including 19 normal and 27 RA samples. To enhance the quality of the gene expression profile data more accurate, the standard preprocessing steps (including scale transformations, management of missing values, replicated handing, flat pattern filtering and pattern standardization) were conducted (Herrero *et al.* 2003). Further, the probes were mapped to the human genomics to identify the gene symbols. Ultimately, a total of 7352 genes were obtained for further study. Next, the expression levels of all genes were standardized.

***Protein–protein interaction (PPI) data*:** PPIs provide a valuable framework for a better understanding of the functional organization of proteomes and are crucial for all biological processes (Stelzl *et al.* 2005). All human PPI data were obtained from the search tool for the retrieval of interacting genes/proteins (STRING, http://string-db.org/), which were designed to assemble, evaluate and disseminate PPI information, in a user-friendly and comprehensive manner (Franceschini *et al.* 2013). Based on the STRING database, 16,730 genes and 787,896 interactions were retrieved. To increase the reliability of these PPIs, only those interactions with a confidence score > 0.5 were selected to establish the informative PPI network. Further, after taking the common part of genes between informative PPI network and gene expression profile, a total of 26,855 protein interactions among 4890 genes were obtained to construct the background PPI network for further analysis.

***Pathway data*:** The biological pathways for human beings were obtained from the Reactome (http://www.reactome.org) pathway database. This database is actually an expert-authored, peer-reviewed knowledgebase of human reactions and pathways, which functions as data mining resource and electronic textbook (Matthews *et al.* 2009). Totally, 1675 pathways were collected from this database. Only pathways with a gene set size ranging from 5 to 100 were reserved as the study objectives owing to the fact that different pathways consist of different number of genes, some pathways with considerably less genes may not have enough biological information, and some pathways having too many genes may be too generic (Ahn *et al.* 2014). Hence, finally, a total of 854 pathways were selected.

Pathway activity calculation and PIN construction

***Pathway activity calculation*:** After filtering the procedures for pathways, only those genes that were mapped to the 854 pathways were reserved for further analysis. After the genes were aligned to the 854 pathways, an activity score for each pathway was defined as the sum of the expression levels of all genes enriched in this given pathway using the principal component analysis (PCA) method. PCA method is a mathematical algorithm that reduces the dimensionality of data, while retaining most of the variation in the datasets (Ringn and Eacute 2008; Hotelling 2010). In particular, the first principal component was determined as the activity score for the corresponding pathway. Thus, there were different activity scores for each pathway in the disease samples and controls. This means that the activity score

for a corresponding pathway between RA and controls was different, and the difference implied the correlation to RA development. A bigger difference in activity score may lead to a closer correlation of the pathway to the disease. Therefore, the pathway with the biggest change in activity score between the two groups was regarded as the most important pathway and was denoted as the seed pathway.

***PIN construction*:** The interactions between the correlative pathways constituted a network, called PIN. In PIN, each node represent a pathway, and one edge was laid between the two pathways if it shared at least one gene or if there were interactions between genes from the two pathways on the basis of PPIs. The PIN is important not only to understand the drug response but also for the development of novel drugs and therapy in human disease (Song *et al.* 2014). For constructing the PIN, two conditions should be met.

Firstly, based on the gene expression data and student's *t*-test, DEGs were determined using P value < 0.05 between RA and controls. In the process of establishing the PIN, it was necessary that there was at least one common DEG between two pathways.

Secondly, to measure whether the two genes that code a pair of interacting proteins used to lay an edge between two pathways was highly coexpressed, we calculated the Pearson correlation coefficient (PCC) for all PPIs between RA and control samples. The distribution of PCC was obtained after the correlation strength between the two pathways was evaluated. The PCC determines the strength of linear association between two variables, and it is measured on a scale with no units and can take a value from -1 through 0 to $+1$ (Sedgwick 1996). The absolute difference in PCC values for the PPIs in RA and control groups was then calculated. In this study, the weight score for a pathway interaction was determined as the total |PCC| values of all genes. The other condition was that the two genes coding a pair of interacting proteins employed to lay an edge between two pathways were highly coexpressed (the absolute value of PCC > 0.8) between genes in the two pathways. If not, the edges between the two pathways were removed.

Based on the above two conditions, an original PIN was built. To simplify the pathway network, the score values of each pathway interactions in PIN, e.g. the sum of the absolute values of PCC for the PPIs in every two pathways were computed and then the pathway interactions were ranked in a descending order based on the score values. Next, the top 5% of pathway interactions were selected to construct an informative PIN for RA to further identify dysregulated pathways.

***Identifying dysregulated pathway from the informative PIN*:** The informative PIN that was extracted increased the difficulty in distinguishing the diseases from the controls, which had a large number of genes and pathways. Hence, to make the differentiation more accurate, detecting a minimum set of pathways (MSP) which were considered to be more possibly dysregulated pathways from the PIN may be the best way to discriminate RA samples from controls.

As mentioned above, the seed pathway that had the biggest change activity score was the first pathway biomarker used to discriminate between diseases and controls. Subsequently, the second pathway was selected from the available pathways and was added to the first pathway to get a better classification ability, which interacted with the first pathway. This process was repeated to add new pathways to extract pathway biomarkers until no more pathways could be added to enhance the accuracy of the classification. The final selected pathway sets were regarded as potential MSP in RA. During the process, support vector machines, a widely used kernel-based method, especially useful for a smaller number of samples with high dimensional variables (Liu *et al.* 2012) was utilized to select the dysregulated pathways. Utilizing a fivefold cross-validation, the classification performance was evaluated, and an area under the curve (AUC) value was adopted as the classification performance index. With the goal of obtaining robust results, the five-fold cross-validation was repeated 100 times and the mean value of classification accuracy was defined as the final result.

Results

PIN establishment

RA is a chronic autoimmune disease, characterized by cytokine-mediated inflammation of the synovial lining of joints. By using student's *t*-test, a total of 807 DEGs were obtained with a P value < 0.05. To establish the PIN, the edges between any two pathways were randomly selected, provided that at least one of the common gene in both pathways was differentially expressed in the two groups, or the two genes coding a pair of interacting proteins used to lay an edge between the two pathways were highly coexpressed (PCC absolute value > 0.8). Finally, an original PIN covering a total of 21,281 pathway–pathway interactions was established. However, due to the complicated scale of the network, the sum of the absolute values of PCC for the PPIs in every pair of pathways was calculated, and only the top 5% of pathway interactions were extracted to construct the informative PIN to identify dysregulated pathways. Finally, a total of 1064 pathway interactions were collected to build the informative PIN.

Seed pathway

After analysing 854 pathways using the PCA, it was discovered that different activity scores existed in pathways between the disease and normal controls. To identify the most important pathway, referred to as the pathway related to the disease, the activity scores of 854 pathways between

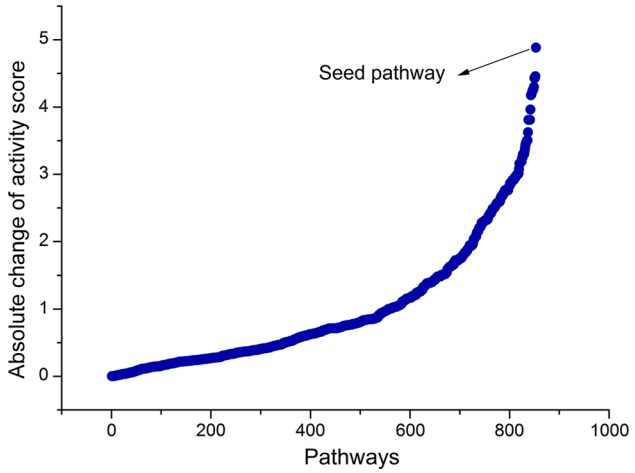

Figure 1. Distribution of the absolute changes in the activity score of 854 pathways. The pathway with the greatest change in activity score between RA sample and normal control was considered to be the seed pathway.

disease and normal controls were analysed. Meanwhile, to make the changes in activity scores more intuitively, the distribution of the activity score is illustrated in figure 1. This figure showed that the greatest absolute change of activity score in the pathways was located on the top point, which was the pathway of the epigenetic regulation of gene expression (absolute change of activity score = 4.8836) and was defined as the seed pathway.

Identification of dysregulated pathways

Taking the seed pathway of epigenetic regulation of gene expression as a start, dysregulated pathways were detected based on the increased classification accuracy. Finally, an MSP (including 10 dysregulated pathways) with an AUC of 0.8249 was obtained from the informative PIN, which was regarded as the most possibly dysregulated pathways. Good performance indicated that the identified dysregulated pathways may have acted as robust biomarkers of RA. The illustrative diagram of the network of these 10 pathways are shown in figure 2.

Additionally, the genes annotated in the identified dysregulated pathways were compared with the DEGs (table 1). From this table, it can be seen that only a small fraction (3.5–17.6%) of the genes enriched in the dysregulated pathways overlapped with the DEGs. This phenomenon further suggested that the pathway as an entity may have a better propensity to diagnose complex diseases rather than individual genes, even if the genes enriched in the pathway were not differentially expressed. From this table, we find that the seed pathway of epigenetic regulation of gene expression owned the largest number of DEGs (n = 9), and the pathway of DNA replication had the second largest number of DEGs (n = 7). Another dysregulated pathway, toll-like receptor 2 (TLR2) cascade was enriched by relatively much more DEGs (n = 6) and

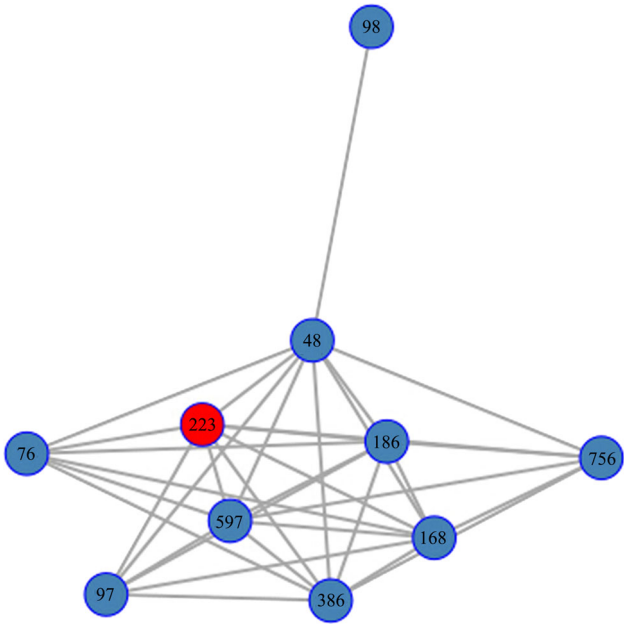

Figure 2. Altered pathway interaction network in RA, involving 10 dysregulated pathways, which were assembled into a pathway network on the basis of the interactions. Each node stood for a pathway. Red node denoted seed pathway. Blue ones was the dysregulated pathways interacted with the seed pathway. The number represented the pathway ID defined in this paper, in alphabetical order.

served as an immunological pathway. Significantly, TLRs, particularly TLR2, play a key role in RA (Seibl *et al.* 2003), which is specific for the immune-related diseases.

Discussion

RA is a common systemic autoimmune disorder characterized by inflammation of the synovial tissue (Hair *et al.* 2014). Although it has been reported that T cells, B cells, the orchestrated interaction of proinflammatory cytokines and Wnt signalling pathway significantly participate in the RA pathogenesis (Choy 2012; Miao *et al.* 2013), the aetiology of this disease is still unknown (Miao *et al.* 2013). In addition, this disease is associated with the risk of morbidity and mortality, especially an increase in mortality rate due to the increased risk of cardiovascular events (Avinazubieta *et al.* 2012; Roubille *et al.* 2015). With the advent of the era of genes, understanding the diseases at the molecular level will help us to pinpoint the root cause of these diseases. Hence, it is considered that biological functions do not just rely on a single gene or an individual pathway, but also the interactions between genes or pathways. Meanwhile, as the interactions between the function-related genes or pathways constitute a large network, it has led to confusion among scholars as to which pathways are closely related to the disease. Consequently, to determine the most closely related pathways about RA, the MSP was extracted from the PIN, which was denoted as the dysregulated pathways. Compared with the traditional

Table 1. Dysregulated pathways extracted from the informative PIN, and the DEGs involed in dysregulated pathways.

Index	Pathway	Number of genes	Number of DEG
223	Epigenetic regulation of gene expression	51	9
597	Removal of licensing factors from origins	37	3
186	DNA replication	57	7
168	Degradation of beta-catenin by the destruction complex	48	4
76	Autodegradation of cdh1 by cdh1:APC/C	34	2
48	Antigen processing-cross presentation	48	4
386	M/G1 transition	46	4
97	Cdc20: phospho-APC/C mediated degradation of cyclin A	38	5
756	Toll like receptor 2 (TLR2) cascade	39	6
98	CDK-mediated phosphorylation and removal of Cdc6	29	1

PIN, pathway interaction network; DEG, differentially expressed genes; 'pathway IDs (indexes)' are defined based on alphabetical order.

methods that focussed on a single pathway (Begovich *et al.* 2004; Choy 2012), our method considered the interactions between pathways and extracted the minimum set of interactions in the pathways from the PIN. These would serve as better biomarkers for the disease and would help us to diagnose the disease (Okada *et al.* 2014).

In this study, a total of 10 dysregulated pathways were obtained from the PIN, among which epigenetic regulation of gene expression was designated as the seed pathway. Epigenetic gene regulation refers to different states of phenotypic expression caused by differential effects of chromosome or chromatin packaging (Hendrich and Willard 1995). Emerging evidence have suggested that the epigenetics play a key role in human pathologies, including inflammatory disorders, and such epigenetic factors may be important in understanding the origins of interindividual variations in the inflammatory response (Wilson 2008). RA is the most common inflammatory disease (Firestein 2003). In addition, Karouzakis *et al.* (2009) have discovered that three main mechanisms of epigenetic control, including DNA methylation, histone modifications and microRNA activity, interact in the development of the RA synovial fibroblasts. Therefore, the results of this study suggested that the pathway of epigenetic regulation of gene expression may play an important role in the progression of RA.

Another dysregulated pathway, the TLR2 cascade, was also identified which was enriched by relatively much more DEGs, and served as an immunological pathway. TLR2 is a member of TLRs which have been demonstrated to mediate the activation of NF-κB, thereby leading the generation of mediators in the immune system, such as TNF-α and IL-1β (Takeuchi *et al.* 2000; Wang *et al.* 2001). Further, proinflammatory cytokines TNF-α and IL-1β stimulates cultured synovial fibroblasts (SFs) to cause the significant increase of TLR2 (Seibl *et al.* 2003). The activation of SFs has been found to be an important characteristic in RA. An earlier study has implicated that TLR2-dependent mechanisms induce the activation of synovial cells, possibly resulting in the destruction of cartilage, accounting for the aetiology of RA (Kyburz *et al.* 2003). In addition, Pierer *et al.* (2004) have indicated that chemokine secretion by activating SFs using TLR-2 contribute to the development of RA. Accordingly, our results support the strategy of targeting the pathway of TLR2 cascade to suppress joint inflammation in RA patients.

Consequently, unlike the previous method, this integration-based analysis has several merits. Firstly, we paid attention to the functional dependency between pathways via establishing a PIN, thereby indicating the robustness of the identified pathway biomarkers. Secondly, when pathways had marginal *P* values, they still may produce a stronger signal if these pathways could form a cluster in the PIN. The results of our study suggested that dysregulated pathways, especially epigenetic regulation of gene expression and TLR2 cascade may be important in RA initiation, development and progression. However, there were still several limitations in this study. To begin with, the sample size was small. Secondly, this study was conducted based on existing data using only bioinformatics method, and the findings lack experimental validation. Although there are limitations, the findings of this study provide some preliminary evidence to uncover alternative candidate therapeutic strategies for RA. These 10 dysregulated pathways may be the potential biomarkers for RA diagnosis and treatment in the future.

References

Ahn T., Lee E., Huh N. and Park T. 2014 Personalized identification of altered pathways in cancer using accumulated normal tissue data. *Bioinformatics* **30**, i422–i429.

Alamanos Y. and Drosos A. A. 2005 Epidemiology of adult rheumatoid arthritis. *Autoimmun. Rev*. **4**, 130–136.

Andersen M., Meyer M. K., Nagaev I., Nagaeva O., Wikberg J. E. S., Minchevanilsson L. *et al.* 2016 AB0020 the melanocortin system is responsive in disease driving immune cells in rheumatoid arthritis and may offer a pathway to curative treatment. *Ann. Rheum. Dis.* **75**, 903–904.

Avinazubieta J. A., Thomas J., Sadatsafavi M., Lehman A. J. and Lacaille D. 2012 Risk of incident cardiovascular events in

patients with rheumatoid arthritis: a meta-analysis of observational studies. *Ann. Rheum. Dis.* **71**, 1524–1529.
Barabasi A.-L. and Oltvai Z. N. 2004 Network biology: understanding the cell's functional organization. *Nat. Rev. Genet*. **5**, 101–113.
Begovich A. B., Carlton V. E., Honigberg L. A., Schrodi S. J., Chokkalingam A. P., Alexander H. C. *et al.* 2004 A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am. J. Hum. Genet*. **75**, 330–337.
Braga-Neto U. M. and Dougherty E. R. 2004 Is cross-validation valid for small-sample microarray classification? *Bioinformatics* **20**, 374–380.
Choy E. 2012 Understanding the dynamics: pathways involved in the pathogenesis of rheumatoid arthritis. *Rheumatology (Oxford)*. **51**, suppl. 5, v3–v11.
Diogo D., Kurreeman F., Stahl E., Liao K., Gupta N., Greenberg J. *et al.* 2013 Rare, low-frequency, and common variants in the protein-coding sequence of biological candidate genes from GWASs contribute to risk of rheumatoid a Arthritis. *Am. J. Hum. Genet.* **92**, 15–27.
Firestein G. S. 2003 Evolving concepts of rheumatoid arthritis. *Nature* **423**, 356–361.
Franceschini A., Szklarczyk D., Frankild S., Kuhn M., Simonovic M., Roth A. *et al.* 2013 STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*. **41**, 808–815.
Glazier A. M., Nadeau J. H. and Aitman T. J. 2002 Finding genes that underlie complex traits. *Science* **298**, 2345–2349.
Hair M. J. H., Sande M. G. H., Ramwadhdoebe T. H., Hanson M., Landewe R. and Leij C. 2014 Features of the synovium of individuals at risk of developing rheumatoid arthritis: implications for understanding preclinical rheumatoid arthritis. *Arthritis Rheumatol*. **66**, 513–522.
Hendrich B. D. and Willard H. F. 1995 Epigenetic regulation of gene expression: the effect of altered chromatin structure from yeast to mammals. *Hum. Mol. Genet*. 4 spec no, 1765–1777.
Herrero J., Díazuriarte R. and Dopazo J. 2003 Gene expression data preprocessing. *Bioinformatics***19**, 655–656.
Hotelling H. 2010 Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol*. **24**, 417–441.
Hu R., Qiu X., Glazko G., Klebanov L. and Yakovlev A. 2009 Detecting intergene correlation changes in microarray analysis: a new approach to gene selection. *BMC Bioinformatics***10**, 20–28.
Karouzakis E., Gay R. E., Gay S. and Neidhart M. 2009 Epigenetic control in rheumatoid arthritis synovial fibroblasts. *Nat. Rev. Rheumatol*. **5**, 266–272.
Khatri P., Sirota M. and Butte A. J. 2012 Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput. Biol*. **8**, 1454–1459.
Kyburz D., Rethage J., Seibl R., Lauener R., Gay R. E., Carson D. A. *et al.* 2003 Bacterial peptidoglycans but not CpG oligodeoxynucleotides activate synovial fibroblasts by toll-like receptor signaling. *Arthritis Rheumatol*. **48**, 642.
Li Y. and Agarwal P. 2009 A pathway-based view of human diseases and disease relationships. *PLoS ONE* **4**, e4346–e4346.
Liu K. Q., Liu Z. P., Hao J. K., Chen L. and Zhao X. M. 2012 Identifying dysregulated pathways in cancers from pathway interaction networks. *BMC Bioinformatics* **13**, 1–11.
Matthews L., Gopinath G., Gillespie M., Caudy M., Croft D., De B. B. *et al.* 2009 Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res*. **37**, 49–61.
Merikangas K. R., Low N. C. and Hardy J. 2006 Commentary: understanding sources of complexity in chronic diseases–the importance of integration of genetics and epidemiology. *Int. J. Epidemiol*. **35**, 593–596.
Miao C. G., Yang Y. Y., He X., Li X. F., Huang C., Huang Y. *et al.* 2013 Wnt signaling pathway in rheumatoid arthritis, with special emphasis on the different roles in synovial inflammation and bone remodeling. *Cell. Signal*. **25**, 2069–2078.
Michaud K. and Wolfe F. 2007 Comorbidities in rheumatoid arthritis. *Best Pract. Res. Clin. Rheumatol*. **21**, 885–906.
Okada Y., Wu D., Trynka G., Raj T., Terao C., Ikari K. *et al.* 2014 Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381.
Pierer M., Rethage J., Seibl R., Lauener R., Brentano F., Wagner U. *et al.* 2004 Chemokine secretion of rheumatoid arthritis synovial fibroblasts stimulated by toll-like receptor 2 ligands. *J. Immunol*. **172**, 1256–1265.
Ringn and Eacute M. 2008 What is principal component analysis? *Nature Biotechnol*. **26**, 303–304.
Rosenberg A., Fan H., Chiu Y. G., Bolce R., Tabechian D., Barrett R. *et al.* 2014 Divergent gene activation in peripheral blood and tissues of patients with rheumatoid arthritis, psoriatic arthritis and psoriasis following infliximab therapy. *PLOS One* **9**, e110657–e110657.
Roubille C., Richer V., Starnino T., Mccourt C., Mcfarlane A., Fleming P. *et al.* 2015 The effects of tumour necrosis factor inhibitors, methotrexate, non-steroidal anti-inflammatory drugs and corticosteroids on cardiovascular events in rheumatoid arthritis, psoriasis and psoriatic arthritis: a systematic review and meta-analysis. *Ann. Rheum. Dis*. **74**, 480–489.
Sedgwick P. 1996 Pearson's correlation coefficient. *N. Z. Med. J*. **109**, 377.
Seibl R., Birchler T., Loeliger S., Hossle J. P., Gay R. E., Saurenmann T. *et al.* 2003 Expression and regulation of toll-like receptor 2 in rheumatoid arthritis synovium. *Am. J. Pathol.* **162**, 1221–1227.
Song M., Yan Y. and Jiang Z. 2014 Drug-pathway interaction prediction via multiple feature fusion. *Mol. Biosyst*. **10**, 2907–2913.
Stelzl U., Worm U., Lalowski M., Haenig C., Brembeck F. H., Goehler H. *et al.* 2005 A human protein-protein interaction metwork: a resource for annotating the proteome. *Cell* **122**, 957–968.
Takeuchi O., Hoshino K. and Akira S. 2000 Cutting edge: TLR2-deficient and MyD88-deficient mice are highly susceptible to Staphylococcus aureus infection. *J. Immunol*. **165**, 5392–5396.
Wang J., Warris A., Ellingsen E., Jørgensen P., Flo T., Espevik T. *et al.* 2001 Involvement of CD14 and toll-like receptors in activation of human monocytes by Aspergillus fumigatus hyphae. *Infect. Immun*. **69**, 2402–2406.
Wilson A. G. 2008 Epigenetic regulation of gene expression in the inflammatory response and relevance to common diseases. *J. Periodontol*. **79**, 1514–1519.
Xia Y., Yu H., Jansen R., Seringhaus M., Baxter S., Greenbaum D. *et al.* 2004 Analyzing cellular biochemistry in terms of molecular networks. *Annu. Rev. Biochem.***73**, 1051–1087.
Yamada H., Nakashima Y., Okazaki K., Mawatari T., Fukushi J. I., Kaibara N. *et al.* 2016 Comparative risk of hospitalized infection associated with biologic agents in rheumatoid arthritis patients enrolled in medicare. *Arthritis Rheumatol*. **68**, 56–66.

Corresponding editor: Rajiva Raman