Individualized Drug Repositioning For Rheumatoid Arthritis Using Weighted Kolmogorov–Smirnov Algorithm

This article was published in the following Dove Press journal:
Pharmacogenomics and Personalized Medicine

Ru-Yin Hu1–3,*
Xiao-Bin Tian1–3,*
Bo Li3
Rui Luo3
Bin Zhang3
Jin-Min Zhao1

1Department of Orthopaedics, Guangxi Medical University, Nanning 530021, People’s Republic of China; 2Department of Orthopaedics, The First Affiliated Hospital of Guangxi Medical University, Nanning 530021, People’s Republic of China; 3Department of Orthopaedics, Guizhou Provincial People’s Hospital, Guiyang 550002, People’s Republic of China

*These authors contributed equally to this work

Background: Existing drugs are far from enough for investigators and patients to administrate the therapy of rheumatoid arthritis. Drug repositioning has drawn broad attention by reusing marketed drugs and clinical candidates for new uses.

Purpose: This study attempted to predict candidate drugs for rheumatoid arthritis treatment by mining the similarities of pathway aberrance induced by disease and various drugs, on a personalized or customized basis.

Methods: We firstly measured the individualized pathway aberrance induced by rheumatoid arthritis based on the microarray data and various drugs from CMap database, respectively. Then, the similarities of pathway aberrances between RA and various drugs were calculated using a Kolmogorov–Smirnov weighted enrichment score algorithm.

Results: Using this method, we identified 4 crucial pathways involved in rheumatoid arthritis development and predicted 9 underlying candidate drugs for rheumatoid arthritis treatment. Some candidates with current indications to treat other diseases might be repurposed to treat rheumatoid arthritis and complement the drug group for rheumatoid arthritis.

Conclusion: This study predicts candidate drugs for rheumatoid arthritis treatment through mining the similarities of pathway aberrance induced by disease and various drugs, on a personalized or customized basis. Our framework will provide novel insights in personalized drug discovery for rheumatoid arthritis and contribute to the future application of custom therapeutic decisions.

Keywords: rheumatoid arthritis, drug repositioning, individualized pathway aberrance, differential pathway

Introduction
Rheumatoid arthritis (RA) is a chronic autoimmune disease characterized by joint synovial tissue inflammation associated with the disability of affected joints.1 Patients with RA have an increased mortality, and the expected survival of RA patients is likely to decrease 3–10 years.2 Early diagnosis and effective therapy are critical to prevent joint deterioration and unfavorable disease outcome. Currently, the treatment of RA primarily rests on the use of disease-modifying antirheumatic drugs, and has improved outcomes in RA patients significantly.3–5 Despite significant therapeutic advances in improving the lives of RA patients, RA remains a hard clinical problem because of the accumulated and persistent disease.1 The administration of RA patients needs new drugs for preventative or curative therapies.
Presently, drug repositioning has drawn broad attention from the pharmaceutical companies and research institutes. Relative to the traditional drug development process, drug repositioning replenishes the drying out drug pipelines by reusing marketed drugs and clinical candidates for new uses, such as treating another disease. These repositioned drugs with known bioavailability, safety profiles and well-characterized pharmacology can enter clinical trials for alternative indications more rapidly and less risk. Currently, multiple computational approaches have been established for drug repositioning. Connectivity map (CMap) has been widely used in drug repositioning by measuring the similarity in gene expression profiles between compounds in mammalian cell lines. While current methods mainly focus on discovering drug candidates targeting huge populations, personalized therapeutic decisions are scarce.

Human genetics provides insight into disease pathogenesis and guides drug discovery for complex traits. A large body of evidence points out the influence of inherited genetic factors on both susceptibility and resistance to the disease. Currently, high-throughput genome-wide association studies have resulted in a paradigm shift in the way that researchers treat complex diseases. Several lines of evidence have revealed numerous genes influencing the likelihood of developing RA by genome-wide analysis. While most of the biological functionality of the cell arises from complex interactions among genes, and interpreting the consequences on a pathway level has more powerful in understanding how gene activity perturbations account for disease. Moreover, personalized pathway analysis has been proposed to perform personalized or customized interpretation of disease data, making it possible to develop personalized therapeutic decisions.

Here, we attempted to predict candidate drugs for RA treatment from CMap database by mining the similarities of pathway aberrance induced by disease and various drugs, on a personalized or customized basis. Our study will provide novel insights into personalized drug discovery for RA.

Materials And Methods

Data Retrieve

Transcriptome Data Of RA

Here we retrieved the transcriptome data of RA from ArrayExpress database (http://www.ebi.ac.uk/arrayexpress/), under the accession number of E-GEOD-15573. In the study of Teixeira et al, a complete genome-wide transcript profiling of peripheral blood mononuclear cells from 18 RA patients and 15 controls was conducted using the Illumina Human-6v2 Expression BeadChips. The raw data and the annotations were obtained from the manufacturer’s documents, and the probes were re-annotated to genes symbols.

CMap Data

The CMap is a collection of genome-wide transcriptome data from cultured human cells treated with bioactive small molecules and simple pattern-matching algorithms that discover connections between drugs, genes expression changes and disease phenotypes. The CMap contains more than 7000 gene expression profiles for 1309 compounds. All raw data and the annotations were downloaded from the CMap and the expression values for all samples were calculated by affy package R with MAS 5.0 normalization. The probes were re-annotated to genes symbols using Brainarray CDF packages. Finally, the samples corresponding to the same drugs were merged, and the gene-drug matrix was obtained for subsequent analysis.

Pathway Data

Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/) is a knowledge database for systematic analysis of gene functional information. All 300 human pathways (covering 6919 genes) were downloaded from KEGG database.

In the present study, we retrieved transcriptome data of RA, CMap data and KEGG pathway data. Prior to analysis, a preprocessing procedure was performed. Firstly, the genes obtained from three data were intersected to gain the common genes. Then, KEGG pathways containing <5 genes or >100 genes were removed, because pathways with too many genes might be too complex to understand, and pathways with too few genes may not have sufficient biological content. Finally, a total of 6919 genes, 888 drugs, and 281 pathways were selected for subsequent analysis.

iPAS For RA

In this section, the pathway levels in each sample were calculated using iPAS algorithm by making use of the accumulated normal data. In this study, 15 normal control samples were combined and regarded as references (nRef). For individual RA cases, an uniformly normalization was performed after combining the single RA data with all nRef samples.

The gene expression value of individual RA sample was standardized by mean and standard deviation (SD) of the reference. For each gene $a$, we calculated the gene expression level as follows:
where \( g_a \) represented the expression value of gene \( a \) in an individual RA subject, \( \text{mean}(g_{n\text{Ref}}) \) stood for the mean expression value of genes \( a \) in all \( n\text{Ref} \) cases, \( \text{SD}(g_{n\text{Ref}}) \) stood for SD of the reference.

To evaluate the iPAS by \( n\text{Ref} \), Average Z algorithm was employed, which presented well in highlighting pathway aberrance and in revealing clinical importance. A vector \( Z = (z_1, z_2, \ldots, z_n) \) represented the expression status of a pathway, where \( z_a \) stood for the standardized expression value of the \( a \)-th gene, and \( n \) represented the gene number in the specific pathway. The iPAS value of a pathway was calculated as follows:

\[
i\text{PAS} = \text{Average } Z = \frac{\sum_{a}^{n} z_a}{n}
\]

Then, the expression matrix (281 pathways \( \times \) 18 RA subjects) was obtained for each pathway in each individualized RA subject from the \( n\text{Ref} \).

The pathway statistics for each individual RA subject was calculated by wilcoxon-test and false discovery rate (FDR) was used to adjust the p-value. The pathways with p-value <0.05 were defined as differential pathways.

**iPAS For Drugs**

The CMap contains more than 7000 gene expression profiles for 1309 compounds. Each drug presents a specific drug-induced gene expression changes of human cells, enabling us to identify the pathway aberrance. After data preprocessing, a total of 6919 genes, 888 drugs, and 281 pathways were selected for subsequent analysis. To identify the drug-induced pathway aberrance, iPAS algorithm was utilized to estimate the pathway levels. For each drug, we calculated the specific drug-induced iPAS status of each pathway using Average Z algorithm. Then, the expression matrix (281 pathways \( \times \) 888 drugs) was obtained for each pathway in each drug.

**Prediction Of Candidate Drugs**

After the above treatment, we obtained RA-induced pathway aberrances and drug-induced pathway aberrances, respectively. Then, we systematically estimate the similarities between RA-induced pathway aberrances and drug-induced pathway aberrances using to a select drugs that might mimic or suppress RA. Prior to similarity analysis, we firstly built Prototype Ranked Lists (PRLs) by merging all the samples corresponding to the same drug, after converting iPAS values to ranks (the iPAS value was used as a ranking procedure in our analysis). The expression matrix (281 pathways \( \times \) 888 drugs) of PRLs was obtained for further analysis.

Next, the pathway-drug PRL matrix was converted to a subject-oriented matrix. Here, a rank-based pattern-matching Enrichment Score (ES) strategy that was based on the weighted Kolmogorov–Smirnov (KS) statistic in Drug Set Enrichment Analysis (DSEA) was employed to perform the converted procedure. Given a PRL \( x \) and an RA subject \( y \), the ES\(_{xy}\) was calculated through DSEA approach. The KS-weighted ES could quantitatively measure the enrichment of signatures in the top/bottom ranked region. The ES value gives a range from 0 to 1. ES value tending to 1 indicates complete similarity and the value tending to 0 indicates the complete opposite. Finally, we built an ES matrix (888 drugs \( \times \) 18 RA samples), the row corresponded to drug and the column represented RA subjects.

By calculating the ES values, the drugs tending to mimic or suppress RA were quantified. Based on the drug-subject matrix, we sorted each row \( x \) according to the ES\(_{xy}\) values of the drug \( x \) across the \( y = 1, \ldots, Y \) RA subjects, and obtained a rank-based drug matrix \( R \). Given an element \( R_{xy} \) in \( R \), it represented the rank of drug \( x \) according to its effect on RA subject \( y \). In this case, the ES could sign whether a drug was a mimic or inhibitor in the development of RA.

The significance of a drug for RA subjects was assessed by applying a nonparametric, rank-based procedure. For each disease subject \( y \), the rank value of drug \( x \) was defined as:

\[
\text{Rank}_{xy} = \frac{\sum(\text{length_{total}} > |ES_{xy}|)}{\text{length_{total}}}
\]

The larger the rank value, the greater the likelihood of a suppressant; the smaller the rank value, the greater the likelihood of a good mimic. In the present study, the top 1% drugs with larger rank value were predicted as therapeutic drugs, and the top 1% drugs with smaller rank value could be considered as good mimics.

**Results**

**iPAS For RA**

In the present study, 15 healthy subjects were denoted as \( n\text{Ref} \)s (reference) of 18 subjects diagnosed RA. The genes were subjected to quantile normalization to evaluate the
gene-level statistics. Meanwhile, a total of 281 pathways were screened from the KEGG pathway database after data preprocessing. By iPAS algorithm, we obtained the pathway aberrance scores of individual RA subjects. Using the mean value of iPAS as the pathway aberrance level of RA, 170 pathways were up-regulated and 111 pathways were down-regulated in RA. The conditions of the altered pathways in all RA subjects were elucidated by individual pathway analysis. The pathway statistics for each individual was tested by wilcoxon-test and the p-value was adjusted by FDR. Under p-value <0.05, a total of 4 pathways were regarded as differential pathways (Table 1). In our study, cardiac muscle contraction pathway showed the highest altered frequency, which altered in 14 of 18 RA subjects, followed by amoebiasis, amino sugar and nucleotide sugar metabolism, and protein processing in endoplasmic reticulum pathway.

iPAS For Drugs
Similar to the identification of pathway aberrance in RA, we analyzed the function aberrance induced by each drug in CMap using iPAS algorithm. For each drug, we calculated the specific gene expression profiles to detect the drug-induced iPAS status of each pathway, using untreated human cells as nRefs (reference). After data preprocessing, we obtained 888 drugs from CMap database and 281 pathways from KEGG database for further analysis. After drug iPAS analysis, we obtained a 281 pathways × 888 drugs matrix.

Prediction Of Candidate Drugs For RA
Based on the iPAS matrix of RA and drugs, a KS-weighted ES analysis was implemented to compare similarities between RA-induced pathway aberrances and drug-induced pathway aberrances. Then, a non-parametric, rank-based procedure was employed to yield a rank list to select the drug candidates that might mimic or treat RA. The pathway profiles for 888 drugs were merged into a specific PRLs by ranking the aberrant pathways. Then, pathway-oriented drug matrix was converted into an RA subject-oriented drug matrix by measuring the ES values. After that, each drug was given a specific ES value for each RA subject. A drug with high ES values indicated that the drug showed related genomic response to RA. The candidate drugs for RA subjects were assessed by applying the rank-based procedure, and the drug-subject matrix was sorted row-wise to select RA-related drugs from the most inhibiting one to the most mimicking one. Here, we identified 9 candidate drugs for RA treatment, as shown in Table 2. Also, nine good mimics were identified based on the similarity of pathway profiles to disease (Table 3). Some candidates with current indications to treat other diseases might be repurposed to treat RA.

Discussion
In response to the high attrition rates in the traditional drug development process, drug repositioning which recaptures marketed drugs for new indications has attracted the attention from pharmaceutical companies and medical researchers. Ashburn and Thor indicated that drug repositioning might shorten the time of drug development from 10–17 years to 3–12 years.27 Previous studies have proposed numerous methods to build predictive models and some have shown promising results.10,12,28,29 These emerging technologies enable investigators to identify candidate drugs that will prevent this disease and its complications. For example, Zhang et al30 presented a drug repositioning strategy based on “omics” data mining to screen candidates for new indications in diabetes treatment, and successfully identified 9 drugs that might have the potential to treat diabetes.

Existing drug repositioning methods mainly focus on discovering candidate drugs for a kind of disease, and are not suitable for predicting candidate drugs for an individual sample. The drug response heterogeneity makes the raise of new strategies that target genotypically well-characterized subpopulations of patients, instead of targeting huge populations. This conversion drives the researchers focus to settle personalized pathway profile by which the disease works and how to intervene on it. In this study, we proposed a computational method to predict candidate drugs from CMap database for RA, in a personalized way, contributing to revealing the molecular mechanisms and the future application of custom therapeutic decisions.

Table 1 The Differential Pathways In Rheumatoid Arthritis Based On The Individualized Pathway Aberrance Score

<table>
<thead>
<tr>
<th>Category</th>
<th>Pathway</th>
<th>p</th>
<th>Altered Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hsa00520</td>
<td>Amino sugar and nucleotide sugar metabolism</td>
<td>0.019</td>
<td>11</td>
</tr>
<tr>
<td>Hsa04260</td>
<td>Cardiac muscle contraction</td>
<td>0.021</td>
<td>14</td>
</tr>
<tr>
<td>Hsa04141</td>
<td>Protein processing in endoplasmic reticulum</td>
<td>0.024</td>
<td>10</td>
</tr>
<tr>
<td>Hsa05146</td>
<td>Amoebiasis</td>
<td>0.047</td>
<td>12</td>
</tr>
</tbody>
</table>
In the present study, we first identify the pathway aberrance of individual RA subjects using iPAS algorithm, and a total of 4 differential pathways were identified. It is well known that RA is a kind of autoimmune disorder. Our study identified 4 differential pathways (cardiac muscle contraction, amoebiasis, amino sugar and nucleotide sugar metabolism, and protein processing in endoplasmic reticulum) in RA. Among them, cardiac muscle contraction pathway showed the highest altered frequency, which altered in 14 of 18 RA subjects. Previous studies revealed that RA patients had an increased risk of premature death compared with the general population, mainly due to cardiovascular disease, and the two diseases shared genetic and environmental risk factors. Our result that identified cardiac muscle contraction as a differential pathway might affirm this declaration.

Although, a rapid evolution of the care and treatment of patients with RA has contributed to a decrease in disease and disease-related complications. Efficient and safe drugs are still pressing problem. By drug repositioning, several small molecules that might suppress RA were identified, which might be critically important for disease prevention and early treatment. Among these candidate drugs, celecoxib had a current drug indication of RA, which is a COX-2 selective nonsteroidal anti-inflammatory drug used to treat various forms of arthritis. The candidate parthenolide showed a current drug indication of anti-inflammatory. López-Franco et al indicated that parthenolide could modulate the NF-kB-mediated inflammatory responses during vascular damage. Moreover, parthenolide has a variety of reported in vitro biological activities, such as blocking lipopolysaccharide-induced osteolysis, inducing apoptosis of leukemia cells, and activity against a parasite Leishmania amazonensis. The candidate drug harmine was reported to promote differentiation of osteoblasts and chondrocytes, and inhibit osteoclastogenesis. Sulocidil and prenylamine both showed a current drug indication of vasodilator. Unfortunately, they were withdrawn from market due to liver toxicity and cardiac arrhythmias, respectively. The other candidate drugs predicted by our method had current indications to treat other diseases and might complement the drug group for RA treatment.

Given these screened potential drug for RA treatment, previous literature have reported that some of these drugs

### Table 2 The Candidate Therapeutic Drugs Of Rheumatoid Arthritis

<table>
<thead>
<tr>
<th>Candidate Drug</th>
<th>Mean ES</th>
<th>Rank Score</th>
<th>CAS Number</th>
<th>Stage</th>
<th>Current Drug Indication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sulocidil</td>
<td>0.245</td>
<td>0.783</td>
<td>54063-56-8</td>
<td>Withdrawn</td>
<td>Vasodilator</td>
</tr>
<tr>
<td>Prenylamine</td>
<td>0.245</td>
<td>0.783</td>
<td>390-64-7</td>
<td>Withdrawn</td>
<td>Vasodilator</td>
</tr>
<tr>
<td>Mebendazole</td>
<td>0.253</td>
<td>0.780</td>
<td>31431-39-7</td>
<td>Approved</td>
<td>Anthelmintic</td>
</tr>
<tr>
<td>Danazol</td>
<td>0.247</td>
<td>0.778</td>
<td>17230-88-5</td>
<td>Approved</td>
<td>Endometriosis</td>
</tr>
<tr>
<td>Piperlongumine</td>
<td>0.249</td>
<td>0.775</td>
<td>20069-09-4</td>
<td>Investigated</td>
<td>Osteogenic differentiation</td>
</tr>
<tr>
<td>Harmine</td>
<td>0.254</td>
<td>0.773</td>
<td>442-51-3</td>
<td>Investigated</td>
<td>Arthritis</td>
</tr>
<tr>
<td>Celecoxib</td>
<td>0.247</td>
<td>0.770</td>
<td>169590-42-5</td>
<td>Investigated</td>
<td>Anti-inflammatory</td>
</tr>
<tr>
<td>Sanguinarine</td>
<td>0.246</td>
<td>0.769</td>
<td>2447-54-3</td>
<td>Investigated</td>
<td>Anti-cancer</td>
</tr>
<tr>
<td>Parthenolide</td>
<td>0.250</td>
<td>0.767</td>
<td>20554-84-1</td>
<td>Investigated</td>
<td>Anti-inflammatory</td>
</tr>
</tbody>
</table>

Abbreviation: ES, enrichment score.

### Table 3 The Good Mimics Of Rheumatoid Arthritis

<table>
<thead>
<tr>
<th>Candidate Drug</th>
<th>Mean ES</th>
<th>Rank Score</th>
<th>CAS Number</th>
<th>Stage</th>
<th>Current Drug Indication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camptothecin</td>
<td>0.608</td>
<td>0.088</td>
<td>2114454</td>
<td>Experimental</td>
<td>Anticancer</td>
</tr>
<tr>
<td>Doxorubicin</td>
<td>0.583</td>
<td>0.104</td>
<td>23214-92-8</td>
<td>Approved</td>
<td>Anticancer</td>
</tr>
<tr>
<td>Daunorubicin</td>
<td>0.582</td>
<td>0.132</td>
<td>20830-81-3</td>
<td>Approved</td>
<td>Anti-infection</td>
</tr>
<tr>
<td>Ampicillin</td>
<td>0.531</td>
<td>0.221</td>
<td>69-53-4</td>
<td>Approved</td>
<td>Anti-infection</td>
</tr>
<tr>
<td>Cloxacillin</td>
<td>0.518</td>
<td>0.241</td>
<td>61-72-3</td>
<td>Approved</td>
<td>Anti-infection</td>
</tr>
<tr>
<td>Streptomycin</td>
<td>0.500</td>
<td>0.271</td>
<td>57-92-1</td>
<td>Approved</td>
<td>Anti-infection</td>
</tr>
<tr>
<td>Cortisone</td>
<td>0.501</td>
<td>0.271</td>
<td>53-06-5</td>
<td>Approved</td>
<td>Anti-inflammatory</td>
</tr>
<tr>
<td>Estradiol</td>
<td>0.499</td>
<td>0.271</td>
<td>50-28-2</td>
<td>Approved</td>
<td>Estrogen deficiency</td>
</tr>
<tr>
<td>Mitoxantrone</td>
<td>0.487</td>
<td>0.278</td>
<td>65271-80-9</td>
<td>Approved</td>
<td>Anticancer</td>
</tr>
</tbody>
</table>

Abbreviation: ES, enrichment score.
could exert certain relieved effect on RA. For instance, Danazol has been identified that treatment of it attenuates refractory autoimmune thrombocytopenia in rheumatic diseases successfully.\textsuperscript{41,42} Some studies have demonstrated that piperlongumine relieves RA through expansion of myeloid-derived suppressor cells (MDSCs) and the inhibition of the Th17 response and activation of fibroblast-like synoviocytes (FLS)\textsuperscript{43} represses dendritic cell maturation by decreasing production of reactive oxygen species,\textsuperscript{44} and suppresses proliferation, migration and invasion of FLS\textsuperscript{45} via animal model and patient samples. Celecoxib has been approved to treat RA and osteoarthritis (OA);\textsuperscript{46,47} it, combined with Cilostazol, impedes proinflammatory factors in FLS of RA patients.\textsuperscript{48} These results may provide more favorable evidence for the clinical use of these drugs.

Meanwhile, we predicted nine mimics based on the similarity of pathway aberrances between disease and drugs. In this study, the drugs with higher ES values showed more similar to disease in the matter of pathway aberrances. Theoretically, it happened that the drugs with high ES values might mimic the disease at the molecular level. While it may not make sense to consider that the pathway profiles of one drug is positively related to those of one disease from the biological view. Thus, this study focused on the candidate drugs for RA treatment in a personalized way.

**Conclusions**

In the present study, we predicted several candidate drugs for RA treatment, and conducted drug discovery in a personalized way contributing to the future application of custom therapeutic decisions.

**Acknowledgments**

This study was funded by the grants from Science and Technology Foundation, Health and Family Planning Commission of Guizhou Province, China (No. gzwjkj2015-1-025), and Joint Project of Guizhou Provincial Department of Science and Technology and Guizhou Provincial People’s Hospital, China (LH7006). Ru-Yin Hu and Xiao-Bin Tian are equal contributors and co-first authors for this study.

**Author Contributions**

All authors contributed to data analysis, drafting and revising the article, gave final approval of the version to be published, and agree to be accountable for all aspects of the work.

**Disclosure**

The authors report no conflicts of interest in this work.

**References**


