Identifying Therapeutic Targets for Joint Specific Rheumatoid Arthritis Using Bayesian Network Models

Melissa Wei, Richard Ainsworth, and Wei Wang

The Department of Chemistry & Biochemistry, University of California San Diego, 9500 Gilman drive, La Jolla, California 92093

Abstract

Rheumatoid arthritis has a diverse pathogenesis, and its complex reactions to treatments may contribute to difficulty in finding treatments. Here we searched for therapeutic targets for joint-specific rheumatoid arthritis in the knees and hips and found that several gene perturbation recipes were more effective in certain joints. We used single cell RNA seq. data from rheumatoid arthritis patients to train 20 directed acyclic graphs and conducted Bayesian inference on them. In order to find the most efficient reprogramming recipes, we simulated the knockdown and overexpression of joint specific rheumatoid arthritis state genes in the network and found the resulting correlation to the osteoarthritis state. The identification of these therapeutic targets could help treat joint-specific rheumatoid arthritis.

Introduction

Rheumatoid arthritis, an autoimmune joint disease, is the most common type of inflammatory arthritis. Although advances in rheumatoid arthritis have been made, the elaborate pathways and interactions between cell states and genes makes it particularly hard to treat. The disease is primarily symmetrical and typically affects diarthrodial, or synovial joints [1, 2, 3] . It starts from smaller joints and spreads to larger joints, such as hips and knees, as the disease progresses [3, 4] . Recently it was proposed that epigenetic differences in fibroblast-like synoviocytes (FLS) in synovial joints contributed to the complexity of rheumatoid arthritis and its pathogenic pathways [1,4] . This was supported by the findings of differentially methylated genes in RA knee and hip joint FLS [4] .
In order to have a systematic and reliable model that will allow perturbations of certain genes, we decided to use genetic networks, in this case Bayesian models, which have been shown to accurately predict protein interactions from genomic expression data and identify therapeutic targets [5,6,7] . These studies have shown the potential of genetic networking in the finding of biological targets and reprogramming recipes, and we decided to use this approach in order to identify such genes for rheumatoid arthritis. Based on these findings, we hypothesize that that the epigenetic differences in FLS could lead to potentially identifying unique therapeutic target combinations that may improve treatment efficiency for joint-specific rheumatoid arthritis.

Methods

We used single cell normalized RNA seq data set from Ai *et al.* [4] that includes 5 samples of rheumatoid arthritis (RA) knee, 5 samples of RA hip, and 10 samples of osteoarthritis (OA), which we regarded as the normal state. The top 25 differentiated RA/OA genes identified from from Ai *et al*. [4], and top 25 most differentially expressed RA knee and RA hip genes from its data set were selected as a 50 gene list. These genes were selected as they are most likely to be the greater causes of the pathway differences in joint specific RA as well as RA and OA. We discretized the expression levels of each gene, shown in Figure 1a, and learned a Bayesian network model purely from quantitative data and did not incorporate prior knowledge.

$$\text{if } \log_2 TPM \geq \tilde{x}\text{ , then =1, else = 0}$$

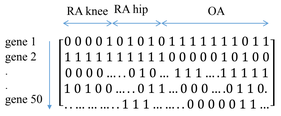


Figure 1a. Data processing was done by discretizing over $\log_2 TPM$ (transcripts per million) on the median. Genes are considered expressed (1) or not expressed (0).

In order to train the highest-scoring Directed Acyclic Graph (DAG) which is a necessary component for the Bayesian network, we used a greedy algorithm which maximizes the BIC (Bayesian Information Criterion) score in order to improve running complexity.
BIC is calculated:

$$BIC(G:\mathrm{D}) = -2\ln(P(\mathrm{D}|G)) - \frac{d}{2}\ln(m),$$

where P(D|G) is the probability of the data given the DAG.

$$P(\mathrm{D}|G) = \prod_{i=1}^{n}\prod_{j=1}^{q_i}\frac{\Gamma(N_{ij})}{\Gamma(N_{ij}+M_{ij})}\prod_{k=1}^{r_i}\frac{\Gamma(a_{ijk}+s_{ijk})}{\Gamma(a_{ijk})},$$

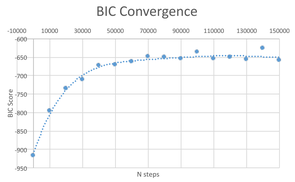
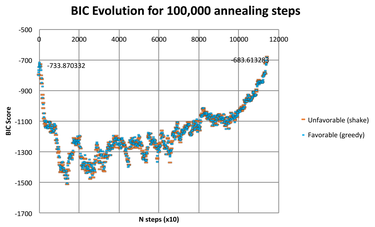


Figure 1b. Annealing by an acceptable range (which decreases as the number of steps increases) shows unfavorable drops in order to achieve a high final result (-733.8703 vs -683.6132).
Figure 1c. BIC score (averages of 5 trials each) convergence is shown as the number of steps increases.



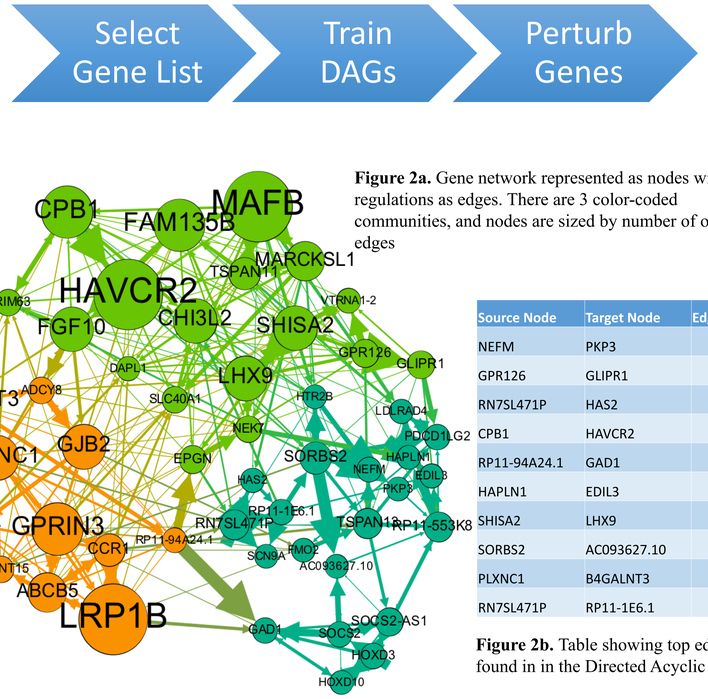


Figure 2a. Gene network represented as nodes with regulations as edges. There are 3 color-coded communities, and nodes are sized by number of out-edges

Source Node	Target Node	Edges
NEFM	PKP3	20
GPR126	GLIPR1	18
RN7SL471P	HAS2	16
CPB1	HAVCR2	16
RP11-94A24.1	GAD1	16
HAPLN1	EDIL3	16
SHISA2	LHX9	15
SORBS2	AC093627.10	15
PLXNC1	B4GALNT3	15
RN7SL471P	RP11-1E6.1	14

Figure 2b. Table showing top edges found in in the Directed Acyclic Graphs.

Next we conducted dynamic Bayesian inference for each Directed Acyclic Graph using the Monte Carlo Markov Chain (MCMC) method using the initial probabilities P(1|E), when E is a given condition for each cell state. The probabilities obtained from the wild type (unperturbed state) of each of the three cell states corresponds to a potential minimum. To simulated the knockdown or overexpression of certain genes, we clamped the activity levels of those genes. Gene overexpression was simulated by clamping gene activity levels to 1, and knockdowns to 0. The inference would calculate the probabilities of the other genes given the perturbations of certain genes and the connections (DAG) between them. The resultant gene expression levels were compared to the initial probability levels to OA and using Pearson correlation.

Results

Half of our gene list were the top genes that were RA differentiated from OA, and the other half were RA knee and hip differentiated genes. The starting Pearson correlations of these RA state genes to OA state genes were very strongly negatively correlated. In fact, they were -0.9465 and -0.9637 for hip and knee, respectively, when compared to OA. The starting RA knee and hip correlation was quite highly correlated, 0.8260. We trained 20 Directed Acyclic Graphs from our 50 gene list and their discretized data and clamped gene activity levels (161,700 possibilities). We compared the resultant probabilities of the genes of each perturbation to the OA gene state and took the Pearson correlation in order to rank them since we wanted the cell states to be similar to the OA. The top 5 recipes of each joint-to-OA correlations are shown Figure 3; a graphical representation of cell state improvement is shown in Figure 4.

Knee Gene Perturbation Recipes	Pearson (R)	Hip Gene Perturbation Recipes	Pearson (R)
OvExp FGF10_OvExp LRP1B_KnD SHISA2	0.704856449	OvExp PLXNC1_OvExp LRP1B_OvExp FAM135B	0.69612477
OvExp FGF10_OvExp LRP1B_KnD LHX9	0.695968112	OvExp LRP1B_OvExp FAM135B_KnD LHX9	0.691949814
OvExp LRP1B_OvExp FAM135B_KnD LHX9	0.687616224	OvExp FGF10_OvExp LRP1B_KnD SHISA2	0.690867825
OvExp PLXNCl_OvExp LRP1B_KnD LHX9	0.681777857	OvExp LRP1B_OvExp FAM135B_KnD SHISA2	0.681886245
OvExp LRP1B_OvExp FAM1358_KnD SHISA2	0.681161734	OvExp PLXNC1_OvExp LRP1B_OvExp MAFB	0.679994853

Figure 3. Table of the top 5 knee and hip recipes, different recipes are identified for each joint; FGF10, LRP1B, and SHISA2 for knee, and PLXNC1, LRP1B, and FAM135B for hip.

We noticed that the recipes for knee and hip RA are different. Most of the common genes in the top recipes also happened to be major nodes in genetic network (Figure 2a). Expectedly, similar genes (FGF10, PLXNC1, LRP1B) are seen in both joint recipes; however, the combinations of genes are unique for each joint.

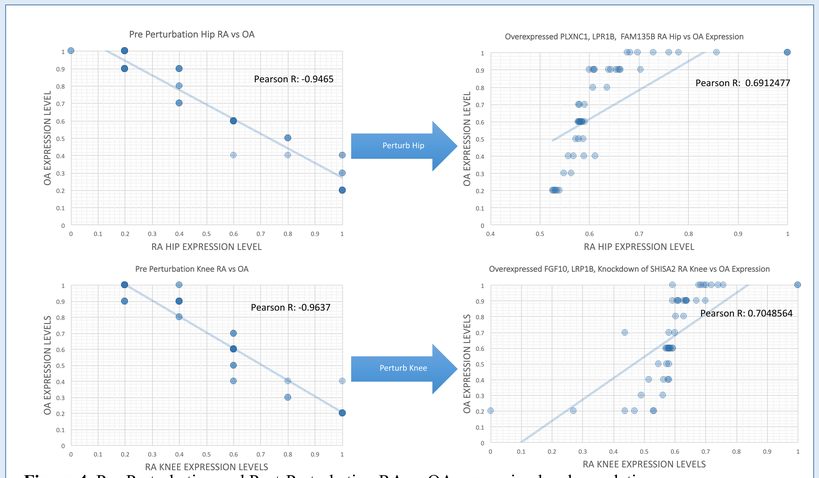


Figure 4. Pre-Perturbation and Post-Perturbation RA vs OA expression level correlation.
The top two graphs compare RA knee to OA expression levels, while the bottom two compare RA hip to OA expression levels. It can be seen that there is a strong negative correlation in the left two graphs, which show the pre perturbation relations between the RA and OA states. The right graphs show the new correlations for the two joint types after being perturbed by their respective top recipe.

Discussion and Conclusions

In this study we used dynamic Bayesian networks (DBN) in order to model the complex interactions between selected genes that were shown to be differentiated either by knee-hip rheumatoid arthritis or osteoarthritis-rheumatoid arthritis. Because of this, we had a very strong starting negative correlation when comparing expression levels of RA to OA FLS states. The gene list had 20 samples with 5 RA-knee, 5 RA-hip, and 10 OA samples. When we perturbed combinations of genes in the RA state by simulating the knocking out or overexpressing them in the network, we noticed that certain recipes increased the correlation to the OA state. However, we also saw that the recipes, including some of the top recipes, were different for RA knee and RA hip, which means that the joint-specific FLS expression could extend to therapeutic importance. Some of the genes in the top recipes, such as FGF10, have been biologically validated to be linked to rheumatoid arthritis [8] . Further research is needed to biologically validate the recipes, to find therapeutic targets for different FLS that are not from knees and hips, and to find more efficient recipes with the current method (for example, more than 3 genes could be perturbed as this will likely improve efficiency of the recipes), or by a different method. It may also be useful develop therapeutics for these candidate genes as realization of joint specific gene recipes may treat patients with joint specific rheumatoid arthritis.

References

1. Majithia V, Geraci SA: Rheumatoid arthritis: diagnosis and management. Am J Med 120:936–939, 2007.
2. Neumann E, Lefevre S, Zimmermann B, Gay S, Muller-Ladner U. Rheumatoid arthritis progression mediated by activated synovial fibroblasts. Trends Mol Med. 2010;16(10):458–68.
3. Firestein, G. S. The disease formerly known as rheumatoid arthritis. *Arthritis Res. Ther.* **16**,114 (2014).
4. Ai, R. *et al*. Joint-specific DNA methylation and transcriptome signatures in rheumatoid arthritis identify distinct pathogenic processes. *Nat. Commun.* 7:11849 doi: 10.1038/ncomms11849 (2016).
5. N. Friedman, M. Linial, I. Nachman, D. Pe'er Using Bayesian networks to analyze expression data. Journal of Computational Biology, 7 (3-4) (2000), pp. 601–620
6. Chen J, Aronow BJ, Jegga AG Disease candidate gene identification and prioritization using protein interaction networks. BMC Bioinformatics 10: 73. (2009)
7. Jansen R, Yu H, Greenbaum D, et al. A Bayesian networks approach for predicting protein-protein interactions from genomic data. Science2003;302(5644):449–53.
8. Geetha-Loganathan P, Nimmagadda S, Scaal M. Wnt signaling in limb organogenesis. Organogenesis.2008;4:109–115.

Acknowledgments

I want to thank Dr. Komives for directing the Research Scholars program, Dr. Wang for welcoming me into his lab, and Dr. Ainsworth for being such a great mentor.

UCSD | Academic Connections

NSF