A network-based model for drug repurposing in Rheumatoid Arthritis

Short title: Network-based model for drug repurposing in RA

Ki-Jo Kim ^{1,2,3}, Navneet Rai ^{2,3}, Minseung Kim ^{2,3} and Ilias Tagkopoulos ^{2,3, §}

¹ Division of Rheumatology, St. Vincent hospital, The Catholic University of Korea, Suwon, Republic of Korea

² Department of Computer Science, University of California, Davis, California, US

³ Genome Center, University of California, Davis, California, US

§ Correspondence to:

Ilias Tagkopoulos, PhD

Genome Center, University of California, Davis, California, US

5313 Genome & Biomedical Science Facility, 451 Health Sciences Drive, University of California, Davis, CA 95616

E-mail: itagkopoulos@ucdavis.edu

Abstract

Background

The identification of drug repositioning targets through computational methods has the potential to provide a fast, inexpensive alternative to traditional drug discovery process. Diseases where complicated pathophysiology evolves over time are excellent targets for such methods. One such disease is Rheumatoid Arthritis (RA), a chronic inflammatory autoimmune disease, where the drug survival at 5-years is less than 50% for patients treated with disease-modifying anti-rheumatic drugs (DMARDs).

Methods and Findings

We have developed a network-based approach for drug repositioning that takes into account the human interactome network, proximity measures between drug targets and disease-associated genes, potential side-effects, genome-wide gene expression and disease modules that emerge through pertinent analysis. We found that all DMARDs, except for hydroxychloroquine (HCQ), were found to be significantly proximal to RA-related genes. Application of the method on anti-diabetic agents, statins and H2 receptor blockers identified anti-diabetic agents – glimepiride, sitagliptin and metformin – that have similar network signatures with the DMARDs. Subsequent *in-vitro* experiments on mouse fibroblast NIH-3T3 cells validated the findings and the down regulation of six key RA-related inflammatory genes. Our analysis further argues that the combination of HCQ and/or sulfasalazine with methotrexate (MTX) is predicted to have an additive synergistic effect in treatment based on network complementarity. Similarly, leflunomide and tofacitinib were found to be suitable alternatives upon chemoresistance to MTX-based double/triple therapy, given the complementary network signatures and overlapping critical target hubs.

Conclusions

Our results corroborate that computational methods that are based on network proximity, among other contextual information can help narrow down the drug candidates for drug repositioning, as well as support decisions for combinatorial drug treatment that is tailored to patient's needs.

Author summary

The network-based proximity between drug targets and disease genes can provide novel insights on the repercussion, interplay, and reposition of drugs in the context of disease. Disease-modifying anti-rheumatic drugs (DMARDs) located significantly close to rheumatoid arthritis (RA)-associated genes and RA-relevant pathways. Three anti-diabetic agents were identified to have an anti-inflammatory effect like DMARDs. We built RA disease module encompassing the emerging small-molecule targets and functional neighbors, which better explained the RA pathophysiology. By proximity and network robustness, tofacitinib and tocilizumab were the most potent for RA disease module as a single agent, and this is consistent with clinical observation. Side effects of clinical importance are predictable by measuring network-based proximity between drug targets and side effect protein. Network-based drug-disease proximity offers a novel and clinically actionable information about drugs and opens the new possibility to drug combination and reposition.

Introduction

Cellular state can be viewed as a well-orchestrated choreography with a touch of chaotic behavior. Genes, proteins and other chemical compounds rarely act in isolation, instead they participate in pathways and mechanisms that give rise to the various biological functions of the cellular network [1]. In the context of disease, cellular state is perturbed by the activation of common pathways through a variety of dysregulated genes and aggregated activity of compounds, leading to a similar phenotype [2, 3]. In a similar fashion, drugs exert their therapeutic effects by modulating molecular pathways and molecule(s) [4], with possible off-target effects due to the presence of system-wide mechanisms and processes [5]. Direct study of the respective pathways and biological processes is cumbersome; however, systemic analysis of large integrated datasets has recently provided opportunities to unravel novel connections, and hence bridge knowledge gaps to better understand human diseases [6, 7]. Using the proximity of the compounds and their targets to key disease genes has emerged as a possible method to identify potential compound-disease pairs. These analyses are based on the hypothesis that the closer a drug's target is to the disease-associated network, the higher the chance it will influence the disease's progression and state [8]. This principle, which is the foundation of this study, can guide drug exploration towards drug repurposing and personalized medicine.

Rheumatoid arthritis (RA) is a chronic inflammatory autoimmune disease that causes progressive joint destruction [9]. Dynamic, complex interactions between a multitude of environmental and genetic factors affect the disease development and activate pathogenic mechanisms, some of which are poorly understood [9]. Despite the introduction of new drugs targeting key mediators of RA – cytokines (TNF and IL-6) and immune cell regulators (B7 molecules and CD20), over 50% of patients only partially respond to therapy [10]. Moreover, disease flare and efficacy reduction is

common in patients with satisfactory initial response to medication [11]. Hence, there is a current and constant need for the development of new RA drugs, which can be tailored to specific targets and personalized based on the patient profile [12].

In this study, we analyzed the relationship between disease-modifying, anti-rheumatic drugs (DMARDs) and RA-associated subnetworks using proximity measures in an unsupervised and unbiased network-based framework. First, we explored the distance of key molecules that are targeted by DMARDs to RA-associated genes. Then we focused on the pertinent pathways and respective biological processes associated based on the distance metrics with a focus on the degree the drug-perturbed pathways cover pathogenic processes of RA. Finally, we created predictors of drug effectiveness and retargeting potential based on network-based signatures, which resulted in 3 potential RA drug candidates that were further validated through in vitro screening. **Fig 1** depicts the overarching methodology for our data collection and proximity analysis performed.

Materials and Methods

The human interactome

A previous study [13] constructed a model of the human interactome based on its protein-protein interaction (PPI) network built based on the experimentally documented human physical interactions from TRANSFAC [14], IntAct [15], MINT [16], BioGRID [17], HPRD [18], KEGG [19], BIGG [20], CORUM [21], PhosphoSitePlus [22], as well as an individually published large-scale signaling network [23]. This human interactome dataset contains 141,150 interactions among 13,329 proteins and was found to perform better in capturing the therapeutic effect of drugs as compared to functional networks from STRING [24] and other high-throughput binary screens [8]. We are using this human interactome network as our basis to conduct our proximity analysis.

RA-associated genes and drug targets

RA-associated genes were retrieved from PheGenI [25] and DisGeNET [26]. 147 and 219 genes were found to be associated with RA in the PheGenI and the DisGeNET database, respectively. 252 unique genes belong to human interactome elements (**S1 Table**). Information on drug-target molecules were extracted from the DrugBank database (**S2 Table**) [27].

Network-based proximity measure

To quantify the network-based relationship between drug targets and disease protein, we used the metric of *closest proximity* d_c that is defined as the average shortest path length between the drug's targets and the nearest disease protein [8]. We found that the *closest proximity* metric offered superior performance when compared to other network-based distance measures such as the shortest measure, the kernel measure, the center measure, and the separation measure [8]. Briefly,

given S , the set of disease proteins, T , the set of drug targets and $d(s, t)$, the shortest path length between nodes s and t in the network, the closest proximity d_c between S and T is given by:

$$d_{c(S,T)} = \frac{1}{\|T\|} \sum_{t \in T} \min_{s \in S} d(s, t)$$

To assess the statistical significance of the closest proximity, we introduced a measure of relative proximity z_c that captures the distribution of distances between random and RA-associated proteins by calculating its z-score. The reference distance distribution was generated by calculating the distance between RA-associated proteins and the randomly selected group of proteins matching the size and the degrees of the drug targets in the network, re-calculated over 1,000 random runs.

As such, the relative proximity z_c is given by:

$$z_{c(S,T)} = \frac{d_{c(S,T)} - \mu_{d(S,T)}}{\sigma_{d(S,T)}}$$

, where $\mu_{d(S,T)}$ and $\sigma_{d(S,T)}$ are the mean and standard deviation of the reference distribution. If $z_c \leq -0.15$, a drug was considered as proximal to the disease. To select the threshold we performed parameter sweeps (grid-search) and also investigated the sensitivity of the final results based on the threshold neighborhood. We observed that for thresholds close to -0.15 , both sensitivity and specificity have optimal values, drug-disease association coverage is in agreement with expert opinion and known knowledge and that interactome-based proximity is not changed significantly by repeating the analyses by using either National Drug Files or the KEGG database, in agreement with a previous study [8].

Drugs used in this study

The DMARD drug class here includes the following drugs: methotrexate (MTX), hydroxychloroquine (HCQ), sulfasalazine (SSZ), leflunomide (LFNM), tofacitinib (TFCN),

adalimumab (ADLM), etanercept (ETAN), infliximab (IFXM), tocilizumab (TCZM), rituximab (RTXM), and abatacept (ABCT). For comparison, selected as controls were some of the anti-diabetic agents (ADA), statins, H2 receptor blockers (H2RB): metformin (MFOM), gliclazide (GCZ), glimepiride (GMPD), pioglitazone (PGTZ), sitagliptin (STGT), canagliflozin (CNGZ), exenatide (EXNT), simvastatin (SMVT), atorvastatin (ATVT), pravastatin (PRVT), fluvastatin (FLVT), rosuvastatin (RSVT), cimetidine (CMTD), and ranitidine (RNTD).

Network-based pathway and side-effect proximity analysis

To identify the biological pathways affected by a drug in the human interactome, we used the closest distance measure to assess the proximity between drugs and pathways. The drug-pathway proximity is the normalized distance calculated between the drug targets and proteins belonging to a given pathway. Similar to our calculation of drug-disease proximity, 1000 randomly selected protein sets matching the original protein sets in size and network degree were used to calculate the mean and the standard deviation of the z -score. We used gene sets of all KEGG [28] and Reactome pathways [29] provided in MSigDB [30].

To check if a drug was adjacent to the proteins provoking certain side effects, we calculated the network-based proximity of drug targets to side-effect proteins (same for disease and pathway proteins). Side-effect proteins were referred to the result by a Fisher's exact test-based enrichment analysis [31] and 99 proteins and 230 side effects were paired. Drug-specific side effects were imported from the SIDER database [32]. Since there is redundancy in the types of side effects for different proteins, if multiple proteins were candidates for drug-associated side effects, a protein more proximal to the drug target was selected.

Disease Module Detection Method

To obtain the RA-specific disease module in the human interactome, we used the Disease Module Detection (DIAMOnD) algorithm [33, 34], which is robust over a wide range of noise and superior performance as compared to the other methods. This algorithm is based on the observation that potential disease genes have a propensity to interact with the known disease genes and prioritizes the proteins in the network based on their topological proximity to seed proteins [13]. Briefly, if there is a number of s seed proteins associated to a certain disease in the network, the probability for a protein with a total of k links to have exactly k_s links to seed proteins is given by the hypergeometric distribution:

$$P(X = k_s) = \frac{\binom{s}{k_s} \binom{N-s}{k-k_s}}{\binom{N}{k}}$$

The connectivity p -value, i.e. the cumulative probability for the observed or any higher number of connections is given by:

$$p - \text{value} = \sum_{n=k_s}^k p(X = n)$$

The protein with the highest rank (i.e. lowest P value) is added to the basic set of seed genes and the gene encoding the protein added by the algorithm are designated as ‘DIAMOnD gene’. In the next iteration, the updated set of seed genes is used to identify the highest ranking protein, until the entire network is completed. The analysis identified statistically significant enrichment for RA-relevant genes among the identified DIAMOnD genes. Since there was no significant further gain beyond 550 iteration steps, we considered the first 550 DIAMOnD genes to have the most significant RA association (**S3 Table**). The module that starts at 85 directly connected seed genes was added by 550 DIAMOnD genes and was further connected to a number of isolated seed genes

(103 genes) through DIAMOnD genes. As a result, a putative RA disease module of 738 genes in total was created.

Validation of DIAMOnD genes

To determine the cutoff size of DIAMOnD genes, we used four RA-relevant validation data – RA gene expression data, gene sets from KEGG, Reactome pathways, and GO biological process terms. For RA gene expression data, we searched the NCBI Gene Expression Omnibus (GEO, January 2017, <http://www.ncbi.nlm.nih.gov/geo/>) and ArrayExpress database (<https://www.ebi.ac.uk/arrayexpress/>) for all gene expression dataset using ‘rheumatoid arthritis’ and ‘synovial tissue’ as a search term, resulting in 14 public available human microarray data. Since there was a trade-off between the number of studies to include and the number of genes that are within the intersection from all datasets, we optimized the product of the two by selecting the point where these two trends cross. We finally selected 11 expression data with GEO series (GSE) ids: GSE77298, GSE48780, GSE45867, GSE36700, GSE24742, GSE15602, GSE21537, GSE12021, GSE55584, GSE55457, GSE55235 and the united collection included the 256 RA, 41 osteoarthritis, and 36 normal samples. After normalizing and removing the batch effect, we performed differentially expression analysis and selected the gene candidates that were identified in three independent methods (empirical Bayes, significance analysis of microarray, and ranked product) and finally filtered the 2212 shared up-regulated genes. The gene sets of KEGG, Reactome pathways or GO terms that are significantly enriched within the given set of seed genes were obtained using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) version 6.8 software and Reactome Analysis tool [29, 35] (P value < 0.05). For each DIAMOnD gene, it is considered as a “hit” if it has at least one annotation that is in the gene sets

significantly enriched within the seed genes (Fisher's exact test, P value < 0.05). To compensate for the dependence of P values on the underlying set size, we used a sliding-window approach: At each iteration step i , we consider all DIAMOnD genes in the interval $[i - 252/2, i + 252/2]$, thereby obtaining sets of the same size as the seed genes that can be compared to each other.

Assessment of network robustness

To assess the robustness of the networks against damage, we used the iterative 'random' and 'targeted' removal of nodes [36, 37]. Nodes were deleted randomly from the graph irrespective of the degree of the node. Or only the nodes targeted by the drugs were selectively removed. After each deletion, we computed the evolution of the size of the largest connected component (LCC), which means minimum path length under a random or targeted attack of a graph. This can tell us how much the drug can disrupt the network of disease module.

Functional enrichment analysis

We performed functional enrichment analysis focusing on the genes of interest using DAVID software [35]. Terms were regarded significant if the P value (EASE score) is lower than 0.01, the gene count more than 3, and the fold enrichment larger than 1.5. The enrichment results were visualized with the Enrichment Map format, where nodes represent gene-sets and weighted links between the nodes represent an overlap score depending on the number of genes two gene-sets share (Jaccard coefficient) [38]. To intuitively identify redundancies between gene sets, the nodes were connected if their contents overlap by more than 10%.

Single sample gene-set enrichment analysis

To test for gene enrichment in individual samples from RA patients (rather than in a group of samples from RA subjects as the former is clinically more relevant), we used a single sample version of gene-set enrichment analysis (ssGSEA), which defines an enrichment score as the degree of absolute enrichment of a gene set in each sample within a given data set [39]. The gene expression values for a given sample were rank-normalized, and an enrichment score was produced using the Empirical Cumulative Distribution Functions (ECDF) of the genes in the signature and the remaining genes. This procedure is similar to the GSEA technique, but the list is ranked by absolute expression in one sample.

In-vitro validation

Materials. The mouse fibroblast NIH-3T3 cell line was used throughout the experiments with the Dulbecco's Modified Eagle's Medium with L-Glutamine, 4.5g/L Glucose and sodium pyruvate (Corning Inc.), HyClone Newborn Bovine Calf Serum (GE Healthcare), Gliclazide and Sulfasalazine (ACROS Organics), Metformin hydrochloride (Mp Biomedicals Inc.), Cimetidine (Alfa Aesar), Sitagliptin phosphate (Biotang Inc.), Penicillin/Streptomycin and Methotrexate hydrate (Sigma). The RevertAid First strand cDNA Synthesis kit, VeriQuest SYBR Green qPCR Master Mix, and Nunc EasYFlask 75cm² were purchased from Thermo Scientific.

Cell culture and treatment. The expression of RA related inflammatory genes (IL-6, IL-1 β , MCP-1, RANTES, CXCL1, and I κ B) was measured on the NIH-3T3. The cells were treated with following immunomodulatory agents: gliclazide (10 μ g/mL), metformin (2 mM), sitagliptin (0.1 mM), methotrexate (10 μ M), sulfasalazine (1 μ M), cimetidine (10 μ g/mL) and TNF- α (10 ng/mL). Gene expression in untreated NIH-3T3 cells was used as the baseline. Two different strategies

were used to measure the efficacy of these drugs on the inflammatory genes. In first strategy, NIH-3T3 cells were treated with above mentioned immunomodulatory agents for 24 h at 37°C in an incubator saturated with 5% CO₂, and samples were preserved in RLT buffer (Qiagen) at -80°C, as described in the RNeasy kit (Qiagen) manual. In second strategy, NIH-3T3 cells were first treated with above mentioned immunomodulatory agents, excluding TNF- α , for 2 h and then were treated with the TNF- α for 24 h at 37°C in an incubator saturated with 5% CO₂. Samples were preserved for RNA-isolation as described before.

RNA isolation and Real Time PCR. Total RNA was isolated using RNeasy kit (Qiagen) as described in the manual. Quantification of total RNA was done using Qubit (Life Technologies). Subsequently, random hexamer assisted first strand of cDNA was synthesized from 5 μ g of total RNA using RevertAid cDNA kit. Real time gene expression was measured using Veriquest SYBR mix on ABI Viia 7 RT-PCR and primer sets described in the table below. GAPDH was used as endogenous control (reference gene). Ct values were calculated in Quant Studio Real Time PCR Software with default settings.

Statistical analysis

For continuous distributed data, between-group comparisons were performed using the paired or unpaired *t*-test. Categorical or dichotomous variables were compared using the chi-squared test or Fisher's exact test. All analyses were conducted in *R* (The R Project for Statistical Computing, www.r-project.org).

Results

Proximity between DMARDs and RA-associated genes in the interactome

The drug-target information was gathered from DrugBank database [27] (**S2 Table**) and bipartite network between RA-associated genes and DMARDs is depicted in **Fig 2A**. The relative proximity between DMARDs and RA-associated genes is plotted in **Fig 2B**. All DMARDs except for HCQ were found to be significantly proximal to RA by $z_c < -0.15$. ABCT was most proximal to a group of RA-associated genes ($z_c = -3.09$) and LFNM was least proximal ($z_c = -0.22$). MTX had a proximity of $z_c = -1.61$ though it was defined to target a single molecule, dihydrofolate reductase. In contrast, molecules targeted by statins and H₂-receptor blockers placed much far from RA-associated genes ($z_c > 0.0$).

Identification of three ADA drugs as RA drug candidates

Interestingly, a significant proximity to RA was found in three ADAs – GCZ, MFOM, and STGT, indicating their potential as RA drug candidates. To validate if GCZ, MFOM, and STGT can be repurposed to treat the RA, *in-vitro* experiments were performed on mouse fibroblast NIH-3T3 cells by quantifying the expression of six RA associated inflammatory genes RANTES, IL-6, CXCL-1, IL-1 β , I κ B and MCP-1, which are the key genes induced in the joints of RA patients [40]. We found that the computationally identified ADAs- GCZ, MFOM, and STGT were repressing the expression of these inflammatory genes. We found this effect to be similar to current commercial anti-RA drugs, methotrexate and sulfasalazine (positive control). Cimetidine, a non-inflammatory drug, appears to have no effect on the expression of inflammatory genes (negative control). TNF- α , a potent inducer of inflammatory genes, actively induced the expressions of inflammatory genes (**Fig 2C** and **Materials and Methods**).

Proximity between DMARDs and pathways

To investigate further the network-based mechanism of drug action, we examined the pathways that are proximal to DMARDs using the gene sets of KEGG and Reactome pathways. **Fig 3A** depicts the heatmap analysis for the RA-relevant pathways. MTX was found to be located close to MAPK-, P53-, PI3K-Akt-, TGF β -, and WNT signaling pathways. HCQ is adjoined to TLR signaling pathways, which is expected given that the target molecules are TLR7 and TLR9, but is also close to NF- κ B- and VEGF signaling pathways. Type I interferon and VEGF signaling pathways in proximity to SSZ and LFNM is adjacent to a broad range of pathways. TFCN was contiguous to type I and II interferon pathways, in which JAK1 and JAK2 are the key mediators. Despite TNF inhibitors, ADLM and ETAN showed strong proximity to Fc ϵ RI-, Jak-STAT-, and NF- κ B signaling pathways rather than TNF signaling pathways.

A large size of collateral targets other than TNF might dilute the direct effect of drugs on TNF signaling pathway since network-based proximity formula equivalently integrates all the distance between target proteins and a set of genes in interest. At a close range for TNF signaling pathway was IFXM, as regarded to target uniquely TNF. TCZM, ABCT, and RTX were placed in the immediate vicinity of IL-6-related (chemokine- and interferon-), T cell receptor-, and B cell receptor signaling pathways in conformity with their specific target(s).

MTX is an anchor drug of widespread use and known clinical efficacy. However, the mechanisms by which low-dose methotrexate exerts its therapeutic effect in RA remain only partially understood and cannot be explained simply by folate antagonism [41]. To gain a better insight on the MTX-involved pathways, we measured the pathway proximity by including MTX-interacting molecules provided in other sources (DGIdb, PharmGKB, and STITCH) [42-44] as a set of targets

(S1 Fig). MAPK-, P53-, PI3K-Akt-, TGF β -, and WNT signaling pathways were the proximal pathways in common across the sources. In PharmGKB and STITCH data providing more adjunctive interactors, TNF, interferon type I and II signaling pathways were identified to approximate to MTX.

Disease module identification

Known disease-associated genes tend to be investigated more extensively, which might introduce a bias and be incomplete to fully explain the pathogenesis of disease. To ease this limitation, we made the disease module based on the established RA-associated genes. A cluster of 85 highly interconnected seed genes was observed on the interactome of the 252 RA associated genes from two sources [25, 26] shown in **Fig 2A**, which we call the “proto-module”. The terms most strongly enriched within the largest connected component (LCC) of the proto-module were T cell receptor signaling pathway (p -value = 3.29×10^{-9}), positive regulation of T cell proliferation (p -value = 1.89×10^{-8}), positive regulation of transcription from RNA polymerase II promoter (p -value = 2.73×10^{-8}), humoral immune response (p -value = 3.87×10^{-7}), positive regulation of smooth muscle cell proliferation (p -value = 5.27×10^{-7}), positive regulation of gene expression (p -value = 7.21×10^{-7}). The rest of the seed proteins either formed small clusters (2,3 and 11 genes) or they had no interaction registered (about 54.4% of the seed proteins, 137/252).

Proteins that are involved in the same disease tend to interact with each other [13]. We prioritized the putative RA-relevant genes forming the disease module on the interactome using the DIAMOnD algorithm [33, 34] (**Fig 4A-C** and **S3 Table**). To define the boundary of the disease module, we analyzed the additional enrichment of DIAMOnD genes at each iteration using four different RA-relevant validation data - RA gene expression data, gene sets from KEGG, Reactome

pathways, and GO biological process terms (**Fig 4B**). **Fig 4B-C** depict the p -values and construction of RA disease module by seed and DIAMOnD genes. The top 30 hits were listed in **Fig 4D**. Three of JAK kinase family genes (JAK3, JAK1, and JAK2) and the SYK gene, the emerging small molecule targets of RA, were incorporated early by priority into DIAMOnD gene although those were not assigned as seed genes. The RA disease module by the KEGG pathways was most strongly enriched with TNF signaling pathway (p -value = 2.08×10^{-35}), followed by Fc epsilon RI signaling pathway (p -value = 2.92×10^{-29}), Toll-like receptor signaling pathway (p -value = 1.02×10^{-24}), and VEGF signaling pathway (p -value = 2.20×10^{-24}).

Perturbation of disease modules by DMARDs

To comprehend the pathways or cellular processes guided by the disease module, we performed the functional enrichment analysis for all seed and DIAMOnD genes using the DAVID software [35] and constructed the enrichment map. Graph of enriched GO terms of biological process was made of 120 nodes with 488 edges and the LCC was comprised of 107 nodes (**Fig 5A** and **S2 Fig**). To check how close a drug was to each node enough to perturb the process, we calculated the network-based proximity of drug targets to gene set of each node in the same way for disease and pathway proteins. We regarded the node significantly proximal to a certain drug ($z_c < -0.15$) as being perturbed.

In an attempt to understand the action mechanisms of the respective drugs, we analyzed the biological processes and molecular function of the GO terms in its disease network. The terms most proximal to MTX were ‘positive regulation of protein insertion into mitochondrial membrane involved in apoptotic signaling pathway’ ($z_c = -2.41$), ‘DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest’ ($z_c = -2.26$), ‘rhythmic process’ ($z_c = -2.11$),

and ‘transcription from RNA polymerase II promoter’ ($z_c = -2.04$). MTX perturbed 83 (69.2%) nodes, which were colored with red hue in **Fig 5B**. Less nodes (66, 55%) and smaller overlap with MTX was found in the case of HCQ. LFNM and TFCN were found to have a high overlap with MTX. It is noteworthy that TFCN attacked the largest size of nodes (90, 75.0%) and was the closest in proximity to 40 nodes (33.3%) with $z_c < -2$. The distribution of perturbed nodes by each DMARD is provided in **Fig 5C** and **S3 Fig**.

In a network context, robustness refers to the system’s ability to carry out its basic functions even at the absence of some network nodes or links. The connectivity of LCC is regarded as the backbone of robustness [37]. To investigate the robustness of the disease module network, we computed the evolution of the size of the LCC under random failures of network nodes [36, 37] (**Fig 5D**). Our results (1000 simulations) argue that half of the network could collapse when at least 22% of the nodes are disturbed on average. All DMARDs, except for ABCT, impacted over a quarter of the nodes and collapsed over half of the disease module network. TFCN and TCZM were found to confer the highest robustness with respect to network stability, while ABCT had the weakest stability of all.

Module-driven validation of network-based proximity by tocilizumab

To test the applicability of network-based proximity by DMARDs, we imported the gene expression dataset of RA synovial tissues, which was obtained from 12 RA patients on two separate occasions before and after a 12-week TZCM treatment (GSE45867) [45]. To compare the paired samples before and after TZCM treatment, ssGSEA was used to generate enrichment scores for gene sets of GO terms by RA disease module. Significant change (p -value < 0.05) was found in 49 GO terms and the 10 main signaling pathways or cytokine-related process of reduced scores

are identified (**Fig 6A** and **S4 Fig**), which are in agreement with the pathway proximity result of TCZM (**Fig 3A**). We applied the outcome to a binary classification (labels of “good” and “not-good” responder) by EULAR response criteria and 6 patients were classified as “good” responders. As compared between “good” and “not-good” responder, score decrease in the regulation of phosphatidylinositol 3-kinase signaling (GO:0014066) and ERK1 and ERK2 cascade (GO:0070371) were significant in the “good” responder category as opposed on the “not-good” one (p -value < 0.05; **Fig 6B**).

Proximity between DMARDs and side effect proteins

Network-based proximity does not always imply that the drug will assuage the corresponding disease. The drug could negatively affect the nearby molecules of normal physiology, leading to unwanted side effects [31]. To see whether the known side effects of DMARDs could be obtained by extrapolating from the network-based analysis, we checked the proximity of chemical DMARDs and the three candidate repurposing ADA drugs (GCZ, STGT and MFOM) to the protein sets predicted to induce the side effects, as shown in **Fig 7**. Since biologic DMARDs of synthetic immunoglobulin-like forms are deemed to interact with the defined targets and to be unable to pass through cell membrane, they were excluded for this analysis. In the case of MTX, 20 of the 64 proteins significantly proximal to MTX (31.3%) have at least one corresponding side effect that is listed in the side effects of MTX from the SIDER database (**S4 Table**). For HCQ, 7 proteins provoking at least one SIDER side effect were found in the proximal range with some of them associated with visual disturbance (CACNA1F, RHO), myopathy (RYR1), and dizziness (CACNA1A), hence need more attention during clinical practice (**S2 Table**). Only 2 proteins having SIDER side effects were observed in TFCN. This might be because its side effects were

not sufficiently surveyed due to their short experience in clinical practice. There was no protein having the corresponding side effects in the outside of proximal area except for one (ATP1A2) in LFNM.

Discussion

As expected, most of DMARDs were found to be significantly proximal to RA-associated genes. Determination of the action mechanism of low-dose MTX for the observed clinical effects in RA has been quite challenging, despite MTX being an anchor drug, [41]. Interestingly, we found that antagonism of folate-dependent processes, the defined mechanism of MTX, was closely linked to MAPK-, P53-, PI3K-AKT-, TGF- β -, and WNT signaling pathways, something that is corroborated by the interaction of MTX with the MAPK-, P53-, or WNT signaling pathways depending on cell types and experimental condition [46-48]. A significant finding of our analysis is the identification of three ADAs (GCZ, MFOM, and STGT) as candidates for drug repurposing in the case of RA treatment. Subsequent *in-vitro* experiments on mouse fibroblast NIH-3T3 cells proved the engagement of GCZ, MFOM, and STGT in the repression of six RA related inflammatory genes. We compared the proximity of DMARDs with those of controls including ADA and statins. It is intriguing that GCZ, MFOM, and STGT stand in a significant proximity to a group of RA-associated genes. In the literature, GCZ has an anti-inflammatory effect by inhibiting monocyte adhesiveness to vascular cells through reduction of oxidized low-density lipoprotein and advanced glycation end product [49]. MFOM showed a suppressive effect on TNF- α -dependent I κ B degradation and expression of proinflammatory mediators IL-6 and IL-1 β [50]. Moreover, the anti-arthritic effect of MFOM was recently demonstrated in collagen-induced arthritis model [51]. DPP4 has capability to directly activate the MAPK and NF- κ B signaling cascade [52] and DPP4 inhibitors showed additional anti-inflammatory and immunomodulatory effects in autoimmune diabetes or myocarditis animal models by recovering the number of CD4⁺CD25⁺Foxp3⁺ cells or inhibiting expression of IL-2, TNF- α and IL-6 [53, 54]. Since these drugs might be beneficial to ease off the inflammatory response, they could be considered by priority in case of a patient having

both RA and diabetes unless contraindicated. These results suggest that network-based proximity inform us of candidate drugs to be repurposed in other disease [8].

The networks of the RA-associated genes are sparse and fragmented probably due to inherent incompleteness of interactome, false positives, and missing disease genes bridging between seed genes [25, 26]. We further reconstituted the RA disease module of 738 genes from 252 seed genes (proto-module), with a distinct signature: while the proto-module largely focused on the T-cell receptor signaling pathway, the RA disease module was enriched with a list of RA-relevant signaling pathways and its functional enrichment analysis was in agreement with the known RA pathophysiology. In parallel, HCQ and LFNM have a better proximity to RA disease module rather than just seed genes and the emerging small molecule targets such as JAK family and SYK were captured at early stages.

We investigated the size and range of disease module differentially covered by each DMARD to understand its mode of action and topology in the network-based framework. We noticed a variable size and differential coverage of the disease network by the drugs. When a patient is first diagnosed with RA, MTX monotherapy is usually used as the first line of defense [55, 56]. If the patient is refractory to MTX, HCQ and/or SSZ are added to the regimen. Our analysis argues that the combination of HCQ and/or SSZ with MTX could create a good synergy effect because of their complementarity in the LLC network (**S3** and **S5 Fig**).

Our analysis shows that if MTX-based double or triple therapy fails, the regimen can be switched to LFNM or TFCN, as these are suitable substitutes given the more extensive and less overlapping to MTX sub-network they perturb, including all the critical hubs. It is well-known that efficacy of TNF inhibitors is enhanced by their combination with MTX [57]. This is inferable by the unshared subnetwork and complementary relationship between TNF inhibitors and MTX. It is also plausible

that TCZM monotherapy produces a similar effect to combination of TNF inhibitor and MTX [57] thanks to the merged coverage by TNF inhibitor and MTX. ABCT inhibits T cell activation, selectively blocking the specific interaction of CD80/CD86 receptors to CD28 and covered the least territory of RA disease module, resulting the weakest direct effect to collapse the disease module. Indeed, ABCT is criticized for being slower in reducing disease activity although the final outcome is equivalent to other biologic DMARDs [58]. However, ABCT was estimated to be a best proximity to RA-associated genes (seed genes) and this is probably because T-cell receptor signaling pathway is the main enrichment of RA-associated genes. Taken together, the disease module and the proximity metric can be used together to inform how to develop better treatment strategies that rely on drug combinations. In effect, when applied to real gene expression data, pathways proximal to TCZM were in accordance with down-regulation (**Fig 3 and 6**), with the PI3K-AKT pathway and ERK cascade might be contributing factors determining the responsiveness for their different transition by outcome.

We found that the proximity between drug targets and proteins that elicit side effects is a suitable metric for the probability of side effects in the case of DMARDs. It is noteworthy that information on retinal toxicity and myopathy was captured in the case HCQ, as these are rare but serious complications that often lead to patient death. SIDER proteins in DMARDs were more proximal as compared with those in 3 candidate ADAs (mean z_c -1.64 versus -0.58 , p -value < 0.001), possibly because of the difference in the main target molecules (inflammatory/immunologic versus metabolic). Although, the proximity metric provides a useful quantification of risk, it is just one of the parameters that should be taken into account, as many proteins that are not directly associated with side effects are impacted by these drugs, hence indirect effects are likely.

Our results argue that network-based drug-disease proximity offers a novel perspective to a drug's therapeutic effect. In this work, we investigated just one of the dimensions that need to be taken into account when determining potential drug candidates for each case, as the actual clinical outcome is determined by a plethora of variables, including dosage, drug interactions, comorbidity, metabolic capacity to drugs, genetic susceptibility, among others. Still, the network-based approach that we describe here provides a fast and efficient way to determine likely candidates for drug repurposing and understand their underlying mechanisms, with far-reaching applications to various diseases beyond RA.

Acknowledgements

We would like to thank Prof. Henry Ho for providing us the mouse fibroblast NIH-3T3 cell line, and Prof. Alyssa Panitch for sharing the animal cell culture facility.

Author Contributions

Conceptualization: Ki-Jo Kim, Ilias Tagkopoulos

Data curation, formal analysis, & investigation: Ki-Jo Kim, Navneet Rai, Minseung Kim

Methodology: Ki-Jo Kim, Navneet Rai

Visualization: Ki-Jo Kim, Navneet Rai, Minseung Kim

Supervision: Ilias Tagkopoulos

Writing – original draft: Ki-Jo Kim

Writing – review & editing: Navneet Rai, Ilias Tagkopoulos

References

1. Yi S, Lin S, Li Y, Zhao W, Mills GB, Sahni N. Functional variomics and network perturbation: connecting genotype to phenotype in cancer. *Nat Rev Genet.* 2017. Epub 2017/03/28. doi: 10.1038/nrg.2017.8. PubMed PMID: 28344341.
2. Kim YA, Wuchty S, Przytycka TM. Identifying causal genes and dysregulated pathways in complex diseases. *PLoS Comput Biol.* 2011;7(3):e1001095. Epub 2011/03/11. doi: 10.1371/journal.pcbi.1001095. PubMed PMID: 21390271; PubMed Central PMCID: PMC3048384.
3. Schadt EE. Molecular networks as sensors and drivers of common human diseases. *Nature.* 2009;461(7261):218-23. Epub 2009/09/11. doi: 10.1038/nature08454. PubMed PMID: 19741703.
4. Csermely P, Korsmaros T, Kiss HJ, London G, Nussinov R. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol Ther.* 2013;138(3):333-408. Epub 2013/02/07. doi: 10.1016/j.pharmthera.2013.01.016. PubMed PMID: 23384594; PubMed Central PMCID: PMC3647006.
5. Segura-Cabrera A, Singh N, Komurov K. An integrated network platform for contextual prioritization of drugs and pathways. *Mol Biosyst.* 2015;11(11):2850-9. Epub 2015/09/01. doi: 10.1039/c5mb00444f. PubMed PMID: 26315485; PubMed Central PMCID: PMC4814269.
6. Zhang B, Tian Y, Zhang Z. Network biology in medicine and beyond. *Circ Cardiovasc Genet.* 2014;7(4):536-47. Epub 2014/08/21. doi: 10.1161/circgenetics.113.000123. PubMed PMID: 25140061; PubMed Central PMCID: PMC4333150.

7. Philippi S, Kohler J. Addressing the problems with life-science databases for traditional uses and systems biology. *Nat Rev Genet.* 2006;7(6):482-8. Epub 2006/05/10. doi: 10.1038/nrg1872. PubMed PMID: 16682980.
8. Guney E, Menche J, Vidal M, Barabasi AL. Network-based in silico drug efficacy screening. *Nat Commun.* 2016;7:10331. Epub 2016/02/03. doi: 10.1038/ncomms10331. PubMed PMID: 26831545; PubMed Central PMCID: PMC4740350.
9. Smolen JS, Aletaha D, McInnes IB. Rheumatoid arthritis. *Lancet.* 2016;388(10055):2023-38. Epub 2016/10/30. doi: 10.1016/s0140-6736(16)30173-8. PubMed PMID: 27156434.
10. Smolen JS, Aletaha D. Rheumatoid arthritis therapy reappraisal: strategies, opportunities and challenges. *Nat Rev Rheumatol.* 2015;11(5):276-89. Epub 2015/02/18. doi: 10.1038/nrrheum.2015.8. PubMed PMID: 25687177.
11. Schipper LG, van Riel PL. Ups and downs in the treatment strategies of rheumatoid arthritis. *Rheumatology (Oxford).* 2011;50(5):818-20. Epub 2010/07/29. doi: 10.1093/rheumatology/keq235. PubMed PMID: 20663813.
12. Kim K-J, Kim M, Tagkopoulos I. Compendium of Synovial Signatures Identifies Pathologic Characteristics for Predicting Treatment Response in Rheumatoid Arthritis. *bioRxiv.* 2018. doi: 10.1101/279786.
13. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, et al. Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science.* 2015;347(6224):1257601. Epub 2015/02/24. doi: 10.1126/science.1257601. PubMed PMID: 25700523; PubMed Central PMCID: PMC4435741.

14. Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R, et al. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* 2003;31(1):374-8. Epub 2003/01/10. PubMed PMID: 12520026; PubMed Central PMCID: PMCPMC165555.
15. Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, et al. The IntAct molecular interaction database in 2012. *Nucleic Acids Res.* 2012;40(Database issue):D841-6. Epub 2011/11/29. doi: 10.1093/nar/gkr1088. PubMed PMID: 22121220; PubMed Central PMCID: PMCPMC3245075.
16. Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, Galeota E, et al. MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.* 2012;40(Database issue):D857-61. Epub 2011/11/19. doi: 10.1093/nar/gkr930. PubMed PMID: 22096227; PubMed Central PMCID: PMCPMC3244991.
17. Chatr-Aryamontri A, Oughtred R, Boucher L, Rust J, Chang C, Kolas NK, et al. The BioGRID interaction database: 2017 update. *Nucleic Acids Res.* 2017;45(D1):D369-d79. Epub 2016/12/17. doi: 10.1093/nar/gkw1102. PubMed PMID: 27980099; PubMed Central PMCID: PMCPMC5210573.
18. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, et al. Human Protein Reference Database--2009 update. *Nucleic Acids Res.* 2009;37(Database issue):D767-72. Epub 2008/11/08. doi: 10.1093/nar/gkn892. PubMed PMID: 18988627; PubMed Central PMCID: PMCPMC2686490.
19. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 2012;40(Database issue):D109-14. Epub 2011/11/15. doi: 10.1093/nar/gkr988. PubMed PMID: 22080510; PubMed Central PMCID: PMCPMC3245020.

20. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, et al. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A*. 2007;104(6):1777-82. Epub 2007/02/03. doi: 10.1073/pnas.0610772104. PubMed PMID: 17267599; PubMed Central PMCID: PMC1794290.
21. Ruepp A, Waegel B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, et al. CORUM: the comprehensive resource of mammalian protein complexes--2009. *Nucleic Acids Res*. 2010;38(Database issue):D497-501. Epub 2009/11/04. doi: 10.1093/nar/gkp914. PubMed PMID: 19884131; PubMed Central PMCID: PMC1794290.
22. Hornbeck PV, Kornhauser JM, Tkachev S, Zhang B, Skrzypek E, Murray B, et al. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res*. 2012;40(Database issue):D261-70. Epub 2011/12/03. doi: 10.1093/nar/gkr1122. PubMed PMID: 22135298; PubMed Central PMCID: PMC1794290.
23. Vinayagam A, Stelzl U, Foulle R, Plassmann S, Zenkner M, Timm J, et al. A directed protein interaction network for investigating intracellular signal transduction. *Sci Signal*. 2011;4(189):rs8. Epub 2011/09/09. doi: 10.1126/scisignal.2001699. PubMed PMID: 21900206.
24. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res*. 2017;45(D1):D362-d8. Epub 2016/12/08. doi: 10.1093/nar/gkw937. PubMed PMID: 27924014; PubMed Central PMCID: PMC1794290.
25. Ramos EM, Hoffman D, Junkins HA, Maglott D, Phan L, Sherry ST, et al. Phenotype-Genotype Integrator (PheGenI): synthesizing genome-wide association study (GWAS) data with

- existing genomic resources. *Eur J Hum Genet.* 2014;22(1):144-7. Epub 2013/05/23. doi: 10.1038/ejhg.2013.96. PubMed PMID: 23695286; PubMed Central PMCID: PMC3865418.
26. Pinero J, Queralt-Rosinach N, Bravo A, Deu-Pons J, Bauer-Mehren A, Baron M, et al. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database (Oxford).* 2015;2015:bav028. Epub 2015/04/17. doi: 10.1093/database/bav028. PubMed PMID: 25877637; PubMed Central PMCID: PMC4397996.
27. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* 2008;36(Database issue):D901-6. Epub 2007/12/01. doi: 10.1093/nar/gkm958. PubMed PMID: 18048412; PubMed Central PMCID: PMC2238889.
28. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 2017;45(D1):D353-d61. Epub 2016/12/03. doi: 10.1093/nar/gkw1092. PubMed PMID: 27899662; PubMed Central PMCID: PMC5210567.
29. Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, et al. The Reactome pathway Knowledgebase. *Nucleic Acids Res.* 2016;44(D1):D481-7. Epub 2015/12/15. doi: 10.1093/nar/gkv1351. PubMed PMID: 26656494; PubMed Central PMCID: PMC4702931.
30. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 2015;1(6):417-25. Epub 2016/01/16. doi: 10.1016/j.cels.2015.12.004. PubMed PMID: 26771021; PubMed Central PMCID: PMC4707969.

31. Kuhn M, Al Banchaabouchi M, Campillos M, Jensen LJ, Gross C, Gavin AC, et al. Systematic identification of proteins that elicit drug side effects. *Mol Syst Biol.* 2013;9:663. Epub 2013/05/02. doi: 10.1038/msb.2013.10. PubMed PMID: 23632385; PubMed Central PMCID: PMC3693830.
32. Kuhn M, Campillos M, Letunic I, Jensen LJ, Bork P. A side effect resource to capture phenotypic effects of drugs. *Mol Syst Biol.* 2010;6:343. Epub 2010/01/21. doi: 10.1038/msb.2009.98. PubMed PMID: 20087340; PubMed Central PMCID: PMC2824526.
33. Ghiassian SD, Menche J, Barabasi AL. A DIseAse MOdule Detection (DIAMOnD) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome. *PLoS Comput Biol.* 2015;11(4):e1004120. Epub 2015/04/09. doi: 10.1371/journal.pcbi.1004120. PubMed PMID: 25853560; PubMed Central PMCID: PMC4390154.
34. Sharma A, Menche J, Huang CC, Ort T, Zhou X, Kitsak M, et al. A disease module in the interactome explains disease heterogeneity, drug response and captures novel pathways and genes in asthma. *Hum Mol Genet.* 2015;24(11):3005-20. Epub 2015/01/15. doi: 10.1093/hmg/ddv001. PubMed PMID: 25586491; PubMed Central PMCID: PMC4447811.
35. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4(1):44-57. Epub 2009/01/10. doi: 10.1038/nprot.2008.211. PubMed PMID: 19131956.
36. Albert R, Jeong H, Barabasi AL. Error and attack tolerance of complex networks. *Nature.* 2000;406(6794):378-82. Epub 2000/08/10. doi: 10.1038/35019019. PubMed PMID: 10935628.
37. Barabási AL, Pálfai MÁ. *Network Science*: Cambridge University Press; 2016.

38. 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(11):e13984. Epub 2010/11/19. doi: 10.1371/journal.pone.0013984. PubMed PMID: 21085593; PubMed Central PMCID: PMC2981572.
39. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature*. 2009;462(7269):108-12. Epub 2009/10/23. doi: 10.1038/nature08460. PubMed PMID: 19847166; PubMed Central PMCID: PMC2783335.
40. Iwamoto T, Okamoto H, Toyama Y, Momohara S. Molecular aspects of rheumatoid arthritis: chemokines in the joints of patients. *Febs j*. 2008;275(18):4448-55. Epub 2008/07/30. doi: 10.1111/j.1742-4658.2008.06580.x. PubMed PMID: 18662305.
41. Brown PM, Pratt AG, Isaacs JD. Mechanism of action of methotrexate in rheumatoid arthritis, and the search for biomarkers. *Nat Rev Rheumatol*. 2016;12(12):731-42. Epub 2016/11/05. doi: 10.1038/nrrheum.2016.175. PubMed PMID: 27784891.
42. Szklarczyk D, Santos A, von Mering C, Jensen LJ, Bork P, Kuhn M. STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Res*. 2016;44(D1):D380-4. Epub 2015/11/22. doi: 10.1093/nar/gkv1277. PubMed PMID: 26590256; PubMed Central PMCID: PMC4702904.
43. Wagner AH, Coffman AC, Ainscough BJ, Spies NC, Skidmore ZL, Campbell KM, et al. DGIdb 2.0: mining clinically relevant drug-gene interactions. *Nucleic Acids Res*. 2016;44(D1):D1036-44. Epub 2015/11/05. doi: 10.1093/nar/gkv1165. PubMed PMID: 26531824; PubMed Central PMCID: PMC4702839.

44. Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, et al. Pharmacogenomics knowledge for personalized medicine. *Clin Pharmacol Ther.* 2012;92(4):414-7. Epub 2012/09/21. doi: 10.1038/clpt.2012.96. PubMed PMID: 22992668; PubMed Central PMCID: PMC3660037.
45. Ducreux J, Durez P, Galant C, Nzeusseu Toukap A, Van den Eynde B, Houssiau FA, et al. Global molecular effects of tocilizumab therapy in rheumatoid arthritis synovium. *Arthritis Rheumatol.* 2014;66(1):15-23. Epub 2014/01/23. doi: 10.1002/art.38202. PubMed PMID: 24449571.
46. Winter-Vann AM, Kamen BA, Bergo MO, Young SG, Melnyk S, James SJ, et al. Targeting Ras signaling through inhibition of carboxyl methylation: an unexpected property of methotrexate. *Proc Natl Acad Sci U S A.* 2003;100(11):6529-34. Epub 2003/05/17. doi: 10.1073/pnas.1135239100. PubMed PMID: 12750467; PubMed Central PMCID: PMC164480.
47. Huang WY, Yang PM, Chang YF, Marquez VE, Chen CC. Methotrexate induces apoptosis through p53/p21-dependent pathway and increases E-cadherin expression through downregulation of HDAC/EZH2. *Biochem Pharmacol.* 2011;81(4):510-7. Epub 2010/12/01. doi: 10.1016/j.bcp.2010.11.014. PubMed PMID: 21114963.
48. Georgiou KR, King TJ, Scherer MA, Zhou H, Foster BK, Xian CJ. Attenuated Wnt/beta-catenin signalling mediates methotrexate chemotherapy-induced bone loss and marrow adiposity in rats. *Bone.* 2012;50(6):1223-33. Epub 2012/04/10. doi: 10.1016/j.bone.2012.03.027. PubMed PMID: 22484100.

49. Renier G, Mamputu JC, Serri O. Benefits of gliclazide in the atherosclerotic process: decrease in monocyte adhesion to endothelial cells. *Metabolism*. 2003;52(8 Suppl 1):13-8. Epub 2003/08/27. PubMed PMID: 12939734.
50. Cameron AR, Morrison VL, Levin D, Mohan M, Forteath C, Beall C, et al. Anti-Inflammatory Effects of Metformin Irrespective of Diabetes Status. *Circ Res*. 2016;119(5):652-65. Epub 2016/07/16. doi: 10.1161/circresaha.116.308445. PubMed PMID: 27418629; PubMed Central PMCID: PMC4990459.
51. Son HJ, Lee J, Lee SY, Kim EK, Park MJ, Kim KW, et al. Metformin attenuates experimental autoimmune arthritis through reciprocal regulation of Th17/Treg balance and osteoclastogenesis. *Mediators Inflamm*. 2014;2014:973986. Epub 2014/09/13. doi: 10.1155/2014/973986. PubMed PMID: 25214721; PubMed Central PMCID: PMC4158168.
52. Wronkowitz N, Gorgens SW, Romacho T, Villalobos LA, Sanchez-Ferrer CF, Peiro C, et al. Soluble DPP4 induces inflammation and proliferation of human smooth muscle cells via protease-activated receptor 2. *Biochim Biophys Acta*. 2014;1842(9):1613-21. Epub 2014/06/15. doi: 10.1016/j.bbadis.2014.06.004. PubMed PMID: 24928308.
53. Hirakawa H, Zempo H, Ogawa M, Watanabe R, Suzuki J, Akazawa H, et al. A DPP-4 inhibitor suppresses fibrosis and inflammation on experimental autoimmune myocarditis in mice. *PLoS One*. 2015;10(3):e0119360. Epub 2015/03/15. doi: 10.1371/journal.pone.0119360. PubMed PMID: 25768281; PubMed Central PMCID: PMC4359137.
54. Yang L, Yuan J, Zhou Z. Emerging roles of dipeptidyl peptidase 4 inhibitors: anti-inflammatory and immunomodulatory effect and its application in diabetes mellitus. *Can J*

Diabetes. 2014;38(6):473-9. Epub 2014/07/19. doi: 10.1016/j.jcjd.2014.01.008. PubMed PMID: 25034244.

55. Singh JA, Saag KG, Bridges SL, Jr., Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. *Arthritis Rheumatol.* 2016;68(1):1-26. Epub 2015/11/08. doi: 10.1002/art.39480. PubMed PMID: 26545940.

56. Smolen JS, Landewe R, Bijlsma J, Burmester G, Chatzidionysiou K, Dougados M, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. *Ann Rheum Dis.* 2017;76(6):960-77. Epub 2017/03/08. doi: 10.1136/annrheumdis-2016-210715. PubMed PMID: 28264816.

57. Witte T. Methotrexate as combination partner of TNF inhibitors and tocilizumab. What is reasonable from an immunological viewpoint? *Clin Rheumatol.* 2015;34(4):629-34. Epub 2015/01/22. doi: 10.1007/s10067-015-2861-x. PubMed PMID: 25604317; PubMed Central PMCID: PMC4365184.

58. Westhovens R, Verschueren P. The efficacy and safety of abatacept in rheumatoid arthritis. *Ther Adv Musculoskelet Dis.* 2010;2(2):89-94. Epub 2010/04/01. doi: 10.1177/1759720x09360429. PubMed PMID: 22870440; PubMed Central PMCID: PMC3383476.

Supporting Information

S1 Fig. Heatmap of proximity between MTX-interacting molecules from different sources (DrugBank, DGIdb, PharmGKB, and STITCH) and RA-relevant pathways. MTX-interacting molecules from each source were listed in the upper panel.

(PDF)

S2 Fig. Functional enrichment map using DAVID for combination of RA-associated genes and DIAMOnD genes. Same as **Figure 5A** except for detailed annotation.

(PDF)

S3 Fig. Network perturbation by proximity between all DMARDs and disease module. Layout of the network is the same as the enrichment map of RA disease module (**Figure 5A**). Nodes significantly proximal to drug targets are colored by red hue depending on the degree of proximity and insignificant nodes are coated by blue color.

(PDF)

S4 Fig. Ten main signaling pathways or cytokine-related process of significant reduction in enrichment scores after treatment (p -value < 0.05).

(PDF)

S5 Fig. Percentage of the shared, distinct, or united nodes on RA disease module by DMARDs. Each column is the baseline DMARD and rows are the counterparts in comparison. Values were calculated from the number of the corresponding nodes divided by the total number of nodes on RA disease module (n=120).

(PDF)

S1 Table. 252 RA-associated genes from PheGenI and DisGeNET.

(PDF)

S2 Table. Information on drug-target molecules from the DrugBank database.

(PDF)

S3 Table. 550 genes from DIAMOnD algorithm.

(PDF)

S4 Table. Proximity between DMARDs and side effect proteins.

(PDF)

Figure Captions

Fig 1. Overview of the computational approach. (A) The RA-associated genes are used as seed for identification of the RA disease module, by using the DIAMOnD module finding algorithm on the human interactome network. Once the RA disease module has been constructed, GO analysis is performed through DAVID for assessment of the related biological processes. (B) Overview of the network proximity data sources and methodology.

Fig 2. Gene-Drug target network and proximity between RA-associated genes and DMARDs.

(A) A graphical representation of all the interactions between RA-associated genes, DMARDs, and their targets. Red nodes are RA-associated genes that are also drug targets. Yellow nodes are disease genes, green nodes are drug targets, and blue triangles are DMARDs. (B) Given a set of disease proteins S , and the set of drug targets T , the closest distance d is the shortest path length between all members of S and T in the network. The relative proximity (z) was calculated by comparing the distance d between T and S to a reference distribution of distances between RA-associated proteins and the 10^3 groups of randomly selected proteins matching the size and the degrees of the drug targets in the network. The red dotted line corresponds to the significance threshold ($z \leq -0.15$). All DMARDs excluding HCQ were found to be significantly proximal to RA-associated genes. From non-DMARDs, PGTZ, GMPD, EXNT, statins, and H2RB were found to not be associated to RA-related genes, while GCZ, MFOM, and STGT have DMARDs-like signatures. (C) GCZ, MFOM, and STGT mediated repression of RA associated inflammatory genes in mouse fibroblast NIH-3T3 cells. Responses of inflammatory genes after 24 hours of drug treatment (left panel). Responses of inflammatory genes when cells were treated first for 2 hours

with the respective drug and then for 24 hours with the TNF- α (right panel). Refer to the **Materials and Methods** for abbreviations.

Fig 3. Pathway analysis of DMARDs. DMARDs and RA-relevant pathways based on proximity metrics. Refer to the **Materials and Methods** for abbreviations.

Fig 4. Disease module identification and validation. (A) Schematic network configuration from seed genes to the RA disease module. Seed genes (red) and their neighbor nodes (green) are identified iteratively. The proto-module consists of the seed genes only, while the final disease module incorporates all genes identified by the DIAMOnD method. (B) Validation of the identified genes through cross-reference with orthogonal data sources. The left column depicts the number of DIAMOnD genes found in the different validation datasets (RA gene expression data, gene sets from KEGG, Reactome pathways, and GO biological process terms), with the corresponding statistical significance (p -value) shown on the right column. No significant gain was noted beyond 550 iteration steps (vertical gray dot line). In right column, a sliding-window approach was used to adjust for the dependence of p -values on the underlying set size. (C) The 252 seed genes and the resulting network once the 486 DIAMOnD-identified genes are added to create the 738-gene RA disease module. (D) List of the DIAMOnD genes at the first 30 iterations.

Fig 5. RA disease module perturbation by DMARDs. (A) Enriched pathway map of the RA disease module. Nodes represent distinct GO terms. Node size is proportional to the number of genes in the respective GO term. Edge was determined by the Jaccard similarity coefficient (edge cut-off is 0.1 distance). (B) Perturbation of the disease module by the MTX drug. Nodes

significantly proximal to drug targets are red, insignificant nodes are blue ($z \leq -0.15$). (C) Ratio of nodes based on the proximity between the DMARDs and the disease module provide the specificity and generalization of the respective DMARD. (D) The percentage of the largest connected component (LCC) that remains when treated by DMARDs (points) or under random node failure (red line, 10^3 simulations). Refer to the **Materials and Methods** for abbreviations.

Fig 6. Validation of network-based proximity by tocilizumab. ssGSEA was performed for paired synovial biopsy samples obtained from naïve RA patients before and 12 weeks after TCZM treatment (GSE45867) [45]. Comparisons between the groups were made using paired or unpaired *t*-test. (A) Ten main signaling pathways or cytokine-related process of significant reduction in enrichment scores after treatment (p -value < 0.05). (B) Change of enrichment score in ‘regulation of phosphatidylinositol 3-kinase signaling’ and ‘ERK1 and ERK2 cascade’ by outcome. Score decrease in “good” responders is significant as compared with “not good” responders (p -value < 0.05).

Fig 7. Proximity between DMARDs and side effect proteins. Proteins eliciting the side effects listed in SIDER database by drugs were marked as red dots [32]. A proximity of $z \leq -0.15$ (blue dot line) is regarded to be significantly proximal to drug [8]. Refer to the **Materials and Methods** for abbreviations.

Figure 1

A. Disease module identification and module enrichment

B. Network proximity measurement

Figure 2

A

B

C

Figure 3

Figure 4

A

C

B

D

Iteration	DIAMOnD gene
1	JAK3
2	IL1R1
3	CD4
4	STAT5B
5	JAK1
6	LCK
7	CD79B
8	PTPN6
9	PIK3R1
10	SYK
11	CD79A
12	STAT3
13	STAT5A
14	GHR
15	JAK2
16	PTPN11
17	ELP2
18	CCR5
19	CSF2RB
20	SHC1
21	PDGFRB
22	CBL
23	FYN
24	LYN
25	KIT
26	EPOR
27	SH2B2
28	IL7R
29	PTK2B
30	SOCS3
...	...

Figure 5

A

B

C

D

Figure 6

A

ID	Term	Enrichment score (mean $\pm$ SE, arbitrary unit)	
		Pre-treatment	Post-treatment
GO:0002755	MyD88-dependent toll-like receptor signaling pathway	5104.2 $\pm$ 96.5	4709.1 $\pm$ 125.1
GO:0007249	I-kappaB kinase/NF-kappaB signaling	4536.6 $\pm$ 96.9	4181.9 $\pm$ 102.0
GO:0007259	JAK-STAT cascade	2664.6 $\pm$ 113.5	2223.4 $\pm$ 75.4
GO:0010803	regulation of tumor necrosis factor-mediated signaling pathway	5684.6 $\pm$ 98.3	5436.3 $\pm$ 95.9
GO:0014066	regulation of phosphatidylinositol 3-kinase signaling	235.5 $\pm$ 51.5	-38.5 $\pm$ 79.4
GO:0032722	positive regulation of chemokine production	4106.5 $\pm$ 184.9	3092.5 $\pm$ 253.7
GO:0032743	positive regulation of interleukin-2 production	3514.1 $\pm$ 188.2	2995.6 $\pm$ 172.2
GO:0060337	type I interferon signaling pathway	3174.8 $\pm$ 134.7	2680.9 $\pm$ 179.8
GO:0070098	chemokine-mediated signaling pathway	-345.3 $\pm$ 402.0	-2107.8 $\pm$ 280.9
GO:0097296	activation of cysteine-type endopeptidase activity involved in a poptotic signaling pathway	3355.7 $\pm$ 209.9	2898.8 $\pm$ 138.2

B

Figure 7