A network-based drug repurposing approach identifies new treatment opportunities for the systemic chronic inflammation underlying multiple complex diseases

Stephanie L Hickey[1,#], Alexander McKim[2,#], Christopher A Mancuso[2], and Arjun Krishnan[1,2,*]

[1]Department of Biochemistry and Molecular Biology and [2]Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, MI 48824, USA

[#]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors. **To whom correspondence should be addressed. Email: arjun@msu.edu

Abstract

Complex diseases are associated with a wide range of cellular, physiological, and clinical phenotypes. To advance our understanding of disease mechanisms and our ability to treat these diseases, it is critical to delineate the molecular basis and therapeutic avenues of specific disease phenotypes, especially those that are associated with multiple diseases. Inflammatory processes constitute one such prominent phenotype, being involved in a wide range of health problems including ischemic heart disease, stroke, cancer, diabetes mellitus, chronic kidney disease, non-alcoholic fatty liver disease, and autoimmune and neurodegenerative conditions. While hundreds of genes might play a role in the etiology of each of these diseases, isolating the genes involved in the specific phenotype (e.g. inflammation “component”) could help us understand the genes and pathways underlying this phenotype across diseases and predict potential drugs to target the phenotype. Here, we present a computational approach that integrates gene interaction networks, disease-/trait-gene associations, and drug-target information to accomplish this goal. We apply this approach to isolate gene signatures of complex diseases that correspond to chronic inflammation and prioritize drugs to reveal new therapeutic opportunities.

1 Introduction

Inflammation is an organism's response to invasion by pathogens or to cellular damage caused by injury, and inflammatory responses fall into two categories: acute and chronic. During an acute inflammatory response, resident immune cells detect a pathogen or damaged cells in a tissue and release proinflammatory compounds, cytokines and chemokines, which recruit additional immune cells circulating in the blood to the site of injury or infection. These cells ultimately contain or neutralize the original offending stimulus, limiting further damage to the host, and clear dead cells and debris promoting tissue repair[1]. Systemic chronic inflammation (CI) occurs when inflammatory

responses do not resolve, resulting in persistent, low-grade immune activation that causes collateral damage to the affected tissue over time[2].

Complex disorders like cardiovascular diseases, diabetes, cancer, and Alzheimer's disease are among the leading causes of death and disability among adults over 50 years of age, and all are associated with underlying systemic inflammation [2,3]. Patients with systemic inflammation caused by autoimmune disorders are more likely to have another CI disorder like cardiovascular disease, type 2 diabetes mellitus, and certain types of cancer[3–5]. Further, treating one chronic-inflammatory disease can reduce the risk of contracting another, suggesting a common underlying pathway[7]. For example, treating rheumatoid arthritis with tumour necrosis factor (TNF) antagonists lowers the incidence of Alzheimer's disease and type II diabetes[8,9].

Common treatments for systemic inflammation, including non-steroidal anti-inflammatory drugs (NSAIDs), corticosteroids, and biologics like tumour necrosis factor (TNF) antagonists, can cause adverse effects when used long term. For instance, patients treated with corticosteroids or TNF antagonists have increased risk of infection[10–12], and corticosteroid use increases both the risk of fracture[13,14] and the risk of developing type II diabetes[15]. NSAIDs present a unique set of side effects, particularly in elderly patients, including gastrointestinal problems ranging from indigestion to gastric bleeding, and kidney damage[16–18]. Therefore, the search for better treatment options for CI is ongoing.

Here, we present a computational approach that uses a network-based drug repurposing strategy to isolate CI gene "signatures" specific to each disorder and predict potential new treatment opportunities specifically targeting the 'inflammation' component of complex diseases. Importantly, our strategy can be used to pinpoint the gene signature of any phenotype underlying a complex disease, and identify drugs to treat that specific phenotype.

2 Methods

2.1 Disease selection and disease-associated seed genes

2.2.1 Complex and autoimmune diseases

We searched the literature[2,4–6,8] and selected 17 complex diseases associated with chronic inflammation (CI) and 9 autoimmune diseases. Some of these diseases are quite broad (i.e "Malignant neoplasm of lung"). To add more narrowly defined diseases to our list, we used the Human Disease Ontology[21] to identify child terms of the diseases. For each original disease and child disease, we used curated annotations from DisGeNET to find disease-associated genes. To ensure that our disease gene sets

were largely non-overlapping, we only retained one representative from every group of gene sets with ≥ 0.6 overlap ($|A \cap B|/min(|A|, |B|)$). This resulted in 10 autoimmune diseases and 37 complex diseases.

2.1.2 Non-disease traits

~100 non-disease-traits that are unlikely to be related to SNPs associated with CI (i.e. handedness, coffee intake, and average household income) were hand selected from UK Biobank[22] traits to be used as negative controls. Based on GWAS summary statistics from the Neale group[23], we used Pascal[24] (upstream and downstream windows of 50 KB with the sum-of-chi-squared statistics method; only autosomal variants) to associate genes with the non-disease traits. Genes with $p < 0.001$ were included as seed genes for that trait.

2.2 GenePlexus

To predict new genes associated with a set of input seed genes, we used GenePlexus, a tool that builds an L2-regularized logistic regression model using features from a gene interaction network[25]. Possible input features include the adjacency matrix, influence matrix, or embedded matrix from the STRING[26], BioGRID[27], or tissue-naive GIANT[28] network. For predicting disease genes, positive examples were disease/trait seed genes and negative example genes were generated by: (i) finding the union of all genes annotated to all diseases in DisGeNET[29], (ii) removing genes annotated to the given seed genes, and (iii) removing genes annotated to any disease in the collection which significantly overlapped with the given seed genes ($p < 0.05$ based on the one-sided Fisher's exact test)[25]. We tested the performance of all nine of the above features for predicting new genes associated with our diseases and traits of interest using three-fold cross validation. Only diseases/traits with ≥ 15 associated genes present in a network were included. Features from the STRING adjacency matrix resulted in the highest median $log2(auPRC/prior)$ (data not shown). Diseases and traits were included in the subsequent analyses if their average $log2(auPRC/prior)$ from the three-fold cross validation was ≥ 1 (i.e. the area under the precision-recall curve '$auPRC$' is at least twice as much as expected by random chance '$prior$'[25]).

2.3 Identifying disease-specific clusters of interacting genes

We added genes with a GenePlexus prediction probability of ≥ 0.80 to the original disease or trait seed gene list to create our final set of associated genes for each disease or trait. We formed disease/trait-specific networks by subsetting BioGRID[27] to include only the disease/trait associated genes and any edges connecting those genes based on direct physical interactions of their protein products (**Fig. 1A**). We then used the Leiden algorithm[30] to partition the disease/trait-specific networks into clusters.

Specifically, we used the *leiden_find_partition* function from the leidenbase R package (v 0.1.3) (https://github.com/cole-trapnell-lab/leidenbase) with the default parameters and CPMVertexPartition as the partition type. We retained clusters containing ≥ 5 genes.

2.4 Subcluster GOBP enrichment analysis

We used the R package topGO[31] (v 2.44.0) to find enrichment of genes annotated to GO biological processes (min size = 5, max size = 100) among disease gene clusters. The annotations were taken from the Genome wide annotation for Human bioconductor annotation package, org.Hs.eg.db[32] (v 3.13.0). The background gene set included all human genes present in BioGRID.

2.5 Isolating CI-associated disease clusters

2.5.1 Defining chronic-inflammation-associated genes

The CI-associated seed genes used in this study include the 18 human genes from the GO biological process (GOBP) term GO:0002544, “chronic inflammatory response”. As with the disease genes, we predicted additional chronic-inflammation-associated genes using GenePlexus with features from the STRING adjacency matrix. Negative examples for GenePlexus were derived from non-overlapping GOBP terms. We added genes with a prediction probability of ≥ 0.80 to the seed gene list to create our final set of CI-associated genes.

2.5.2 Creating Random Traits

After running GenePlexus to predict new genes for each trait, the gene lists for each trait were used to generate 5,000 random gene lists that have matching node degree distributions to the original traits. That is, a random gene list was generated for a given trait by replacing each of its genes in the BioGRID network with a (randomly chosen) gene that has the same node degree, or a gene that has a close node degree if there are a small number of genes with the exact node degree. We clustered the random traits as described in section 2.3. Only real traits for which all 5,000 associated fake traits clustered without error were included in the study: 29 complex diseases, 7 autoimmune diseases, 50 non-disease traits.

2.5.3 Finding CI-gene enriched disease clusters

For each disease and random trait cluster containing ≥ 5 genes, we calculated an enrichment score score, $E = log_2(\frac{(CG \cap CI)/CG}{CI/(DG \cup CI)})$ where CG are the genes in a disease cluster, CI are the CI genes, and DG are all of the disease genes. For each real disease or trait cluster, we used the matching random trait clusters to calculated a p-value,

$p = \frac{n\ random\ cluster\ score \geq real\ \ cluster\ score}{n\ random\ \ clusters\ + 1}$. We corrected for multiple comparisons across clusters within a disease using the Benjamini-Hochberg procedure. Clusters with an $FDR < 0.01$ were considered chronic-inflammation-associated disease clusters and were deemed to represent the CI signature of the disease.

2.6 Predicting novel treatment opportunities

2.6.1 Identifying expert-curated drug-target associations

The known drug-gene interactions used in this study are the subset of the interactions present in the DrugCentral database[19] that are also among the expert curated interactions in the Drug-Gene Interaction database (DGIdb)[20]. Specifically, we used the DGIdb API to retrieve only drug-gene interactions that were marked "*Expert curated*" (based on the source trust levels endpoint). Intersecting these interactions with those in DrugCentral (through a list of drug synonyms from DrugCentral) resulted in the final list of expert-curated drug-gene pairs.

2.6.2 Treatment prediction and scoring

We predicted treatment opportunities for the inflammatory component of complex diseases by first searching the CI-associated disease clusters for gene targets of drugs present in our list of expert-curated drug-gene interactions. If a target gene of a drug is in a CI-associated disease cluster, the drug is a potential candidate for treatment of CI in that disease. To test if the CI-associated disease clusters contained more targets for a given drug $Drug\ X$ than expected by chance, for each $Disease\ Y$, we first calculated the fraction of $Drug\ X$'s targets that fall within $Disease\ Y$'s CI-associated clusters.

Then, we calculated the same score for $Drug\ X$ (i.e., fraction of its targets in disease clusters) against 5,000 random traits generated for the $Disease\ Y$ (see section 2.5) based on its clusters with ≥ 5 genes. These clusters are searched for drug targets irrespective of them being significantly overlapping with the CI genes. These 5,000 fake drug-disease scores per disease form an empirical null distribution. We then used this distribution to calculate a permutation p-value for the the real drug-disease score where $p = \frac{n\ random\ Disease\ Y\ to\ Drug\ X\ score \geq real\ \ Disease\ Y\ to\ Drug\ X\ score}{n\ random\ Disease\ Y + 1}$. The p-values were corrected for multiple comparisons within each disease across drugs using the Benjamini-Hochberg procedure. Predicted treatments are disease-drug pairs with $FDR < 0.05$.

2.6.3 Enrichment of predicted drug-disease pairs among previously indicated drug-disease pairs

Previously indicated drug-disease pairs (both approved and off-label) were obtained from DrugCentral. Only drugs with expert curated target genes were included (see

section *2.6.1*). The Unified Medical Language System (UMLS) Concept Unique Identifiers (CUI) were limited to diseases (T047) and neoplastic processes (T191), and our diseases were matched to diseases in DrugCentral using UMLS CUI ids. To check for an enrichment of predicted drug-disease pairs among previously indicated drug-disease pairs for each disease, we tallied the total number of unique drugs previously indicated to any disease, the number of previously indicated drugs indicated to the disease of interest, the number of drugs predicted to treat the disease by our method, and the number of drugs predicted to treat the disease by our method that were also previously indicated for that disease. We calculated a p-value using a one tailed Fisher's exact test, and corrected for multiple comparisons within each disease across drugs using the Benjamini-Hochberg procedure.

2.6.4 Clinical trial hypergeometric test

To determine if our predicted disease-drug pairs were enriched among disease-drug pairs tested in a clinical trial, we used the database for Aggregate Analysis of Clinical Trials (AACT)[33]. AACT reports the Medical Subject Headings (MeSH) vocabulary names for diseases. We used disease vocabulary mapping provided by DisGeNET to translate UMLS CUI ids for our diseases to MeSH vocabulary names, further restricted to only those that were present in AACT. We filtered AACT for trials with "Active, not recruiting", "Enrolling by invitation", "Recruiting", or "Completed" status. For the background number for the hypergeometric test, we calculated all possible disease-gene pairs by multiplying the number of unique diseases (by MeSH name) studied in any clinical trial by the unique number of drugs studied in any clinical trial that also had expert-curated target genes. We performed the same analysis filtering for drug-disease pairs that made it to a Phase IV trial.

3 Results

3.1 Expanding lists of disease-related genes and identifying disease-specific gene subnetworks

Our first goal was to establish a comprehensive list of genes associated with the complex diseases of interest and resolve the genes linked to each disease into subsets of tightly-connected genes in an underlying molecular network. Towards this goal, we selected 37 complex diseases associated with underlying systemic inflammation (see *Methods*). To ensure that we correctly isolate chronic inflammation (CI) signatures, we devised a set of positive and negative controls. We selected 10 autoimmune disorders as positive controls because autoimmune disorders are characterized by CI and should have an easily identifiable CI gene signature. For negative controls, we selected ~100

traits from UK Biobank[22] that are unlikely to be associated with CI (i.e. handedness, coffee intake, and average household income). **Table S1** contains the full list of diseases and traits used in this analysis.

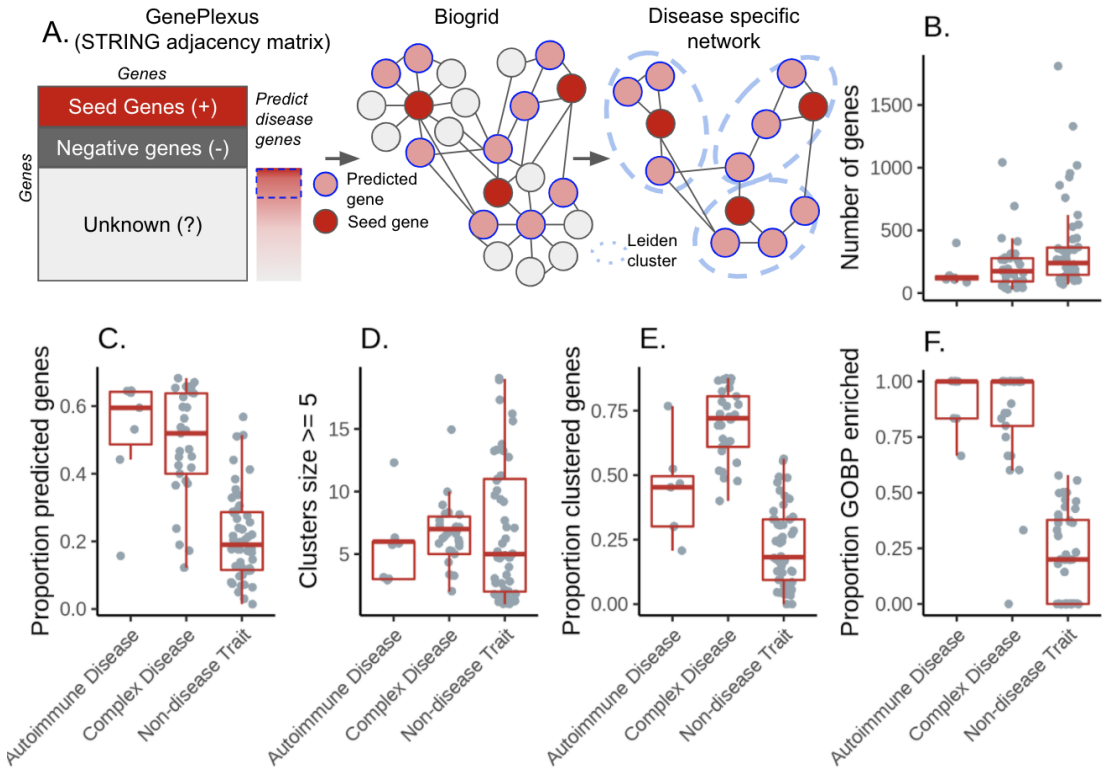

Figure 1: **A**. Schematic outline of the expansion of disease-/trait-associated gene lists with GenePlexus and the clustering of trait specific networks. **B**. Number of genes per disease/trait. **C**. Proportion of the genes per disease/trait that were predicted by GenePlexus. **D**. Number of clusters per disease/trait containing at least 5 genes. E. Proportion of total genes assigned to a cluster containing at least 5 genes. F. Proportion of clusters per disease/trait enriched with genes from at least one GO biological process.

While thousands of genes may play a role in the etiology of a chronic disease, it is unlikely that all of these genes have been cataloged in available databases such as DisGeNET or identified by GWAS. Hence, we expanded the lists of disease-/trait-associated genes using GenePlexus[25]. Briefly, GenePlexus performs supervised machine learning using network-based features to predict novel genes related to a set of input seed genes. Here, we built one GenePlexus model per disease using disease-associated genes from DisGeNET or trait-associated genes from UK Biobank GWAS as seed genes (positive examples) and the rows of the STRING[34] adjacency matrix as feature vectors (**Fig. 1A**). Genes predicted by the GenePlexus model with a probability ≥ 0.80 were added to the seed genes to create an expanded list of disease-/trait-associated genes. Though the number of genes in the expanded list for the non-disease traits are slightly higher than those for the autoimmune and complex

diseases (**Fig. 1B**), the proportion of the expanded list predicted by GenePlexus is substantially higher for autoimmune and complex disease (**Fig.1C**). This observation indicates that genes associated with a specific autoimmune/complex disease tend to have similar network neighborhoods. All disease-associated genes are listed in **Table S2**.

Next, for each disease/trait, we clustered the expanded lists of genes based on their interaction in BioGRID[27] (i.e. based on direct physical interactions of their protein products) (**Fig.1A** and **D**; **Table S2**). The complex diseases had the highest proportion of genes grouped into clusters of ≥ 5 genes, followed by autoimmune diseases and non-disease traits (**Fig. 1E**). To assess whether clusters are biologically meaningful, we performed an enrichment analysis between every cluster and hundreds of GO Biological Process (GOBP) gene sets. We theorize that significant enrichment of a cluster with a GOBP means the genes in the cluster likely function together to carry out a specific cellular process or pathway. For autoimmune and complex diseases, the median proportion of GOBP enriched clusters is 1, and for most autoimmune and complex diseases, more than half of the clusters are enriched with a GOBP (**Fig. 1F**). In contrast, most clusters in non-disease traits are not enriched for a GOBP (**Fig. 1F**).

3.2 Isolating CI-associated disease subnetworks

Our next goal was to identify which subnetwork(s) of a disease corresponds to its CI component. For this analysis, similar to the expansion of disease-/trait-associated genes, we used GenePlexus to take the set of 18 human genes annotated to the "chronic inflammatory response" GO biological process term (GO:0002544) and expand it to a list to 45 CI genes (**Table S2**). We then scored the enrichment of CI genes in each disease cluster and performed a permutation test using 5,000 random gene sets for each disease to determine the significance of the enrichment score (see *Methods* section 2.5, **Table S3**). As expected, we were able to identify clusters enriched for CI genes in the majority of the autoimmune disorders surveyed (6/7), while finding no CI-enriched clusters among the non-disease traits (**Fig. 2A** and **B**). Importantly, we identified at least one CI-enriched cluster in 7 of the complex diseases (**Fig. 2A** and **C**).

3.3 Predicting novel treatment opportunities

Our final goal was to leverage the CI-associated disease cluster we discovered to find potential avenues for repurposing approved drugs to therapeutically targeting systemic inflammation underlying the complex diseases. Towards this goal, we mined DrugCentral and the Drug-Gene Interaction Database[20] and looked for gene targets of drugs among the genes in the CI-associated disease clusters. To prioritize drugs for a given disease $Disease\ Y$, we developed a system to derive a score for each drug

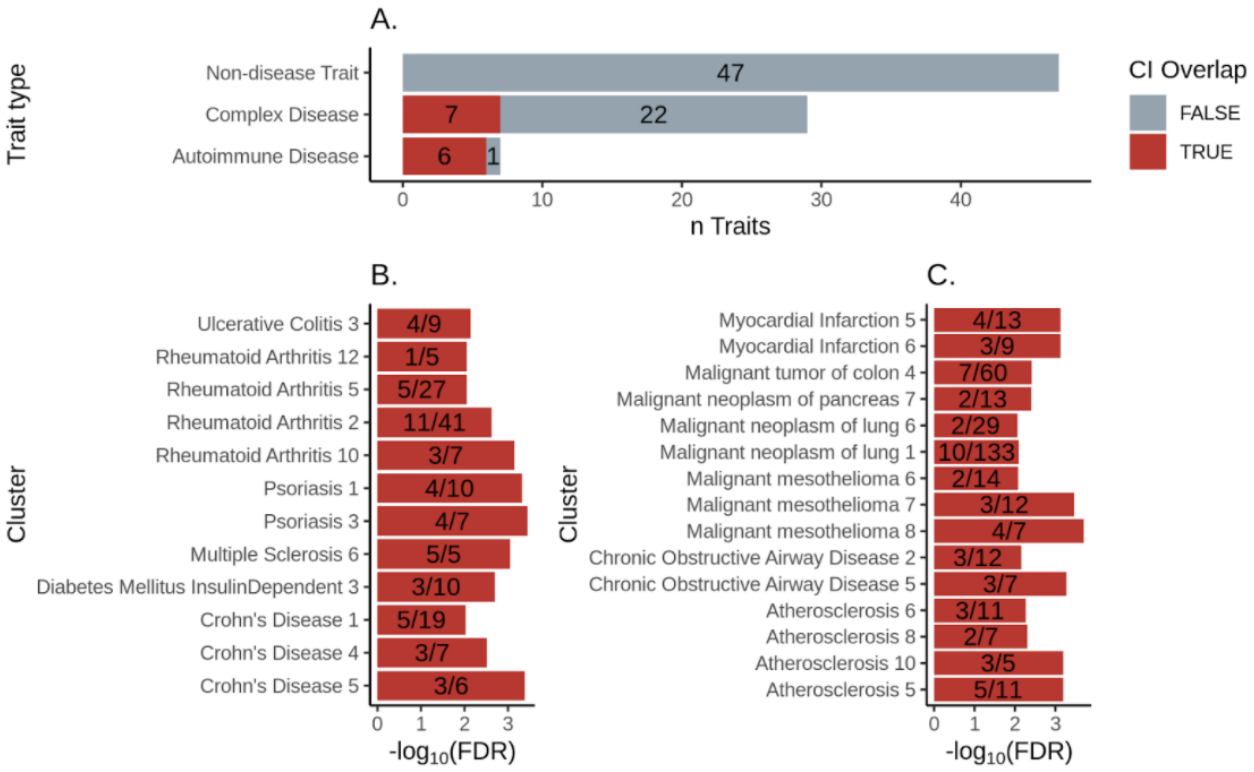

Figure 2: **A**. Number of diseases/traits with a least one cluster overlapping the expanded chronic inflammation (CI) geneset (red), out of the total number of diseases/traits. **B**. Autoimmune disease clusters significantly enriched with genes in the expanded chronic inflammation gene set. The x-axis is the -log10(FDR) of the permutation test. The numbers in the bar are the number of chronic inflammation genes in the cluster / the total number of genes in the cluster. **C**. Same as B for complex diseases.

$Drug\ X$ (**Fig. 3A**) by i) calculating the fraction of $Drug\ X$'s targets present in $Disease\ Y$'s CI-associated clusters, and ii) estimating the significance of this overlap based on $Drug\ X$'s targets in clusters of 5,000 fake gene lists, each list matching the genes of $Disease\ Y$ in their node degree distribution.

With this scoring system, we found drugs associated with all 6 autoimmune disorders that have clusters enriched for CI genes (**Fig. 3B**). We then compared the drugs identified by us to the drugs already indicated for each disease or recommended for off-label use (from DrugCentral[19]; referred to here as "previously indicated"). Our method identified previously indicated drugs for 3 out of the 6 autoimmune disorders, with significant enrichment among drug predictions for rheumatoid arthritis and ulcerative colitis (**Fig. 3B**). Of note, 6 of the 44 drugs predicted by our method for psoriasis are previously indicated for a related disease, plaque psoriasis, lending additional support to our analysis as well as highlighting the limitations in interpreting the results based on known drugs for the exact disease under consideration.

Importantly, we identified novel treatments for 5 of the 7 complex diseases assayed (**Fig. 3C**).

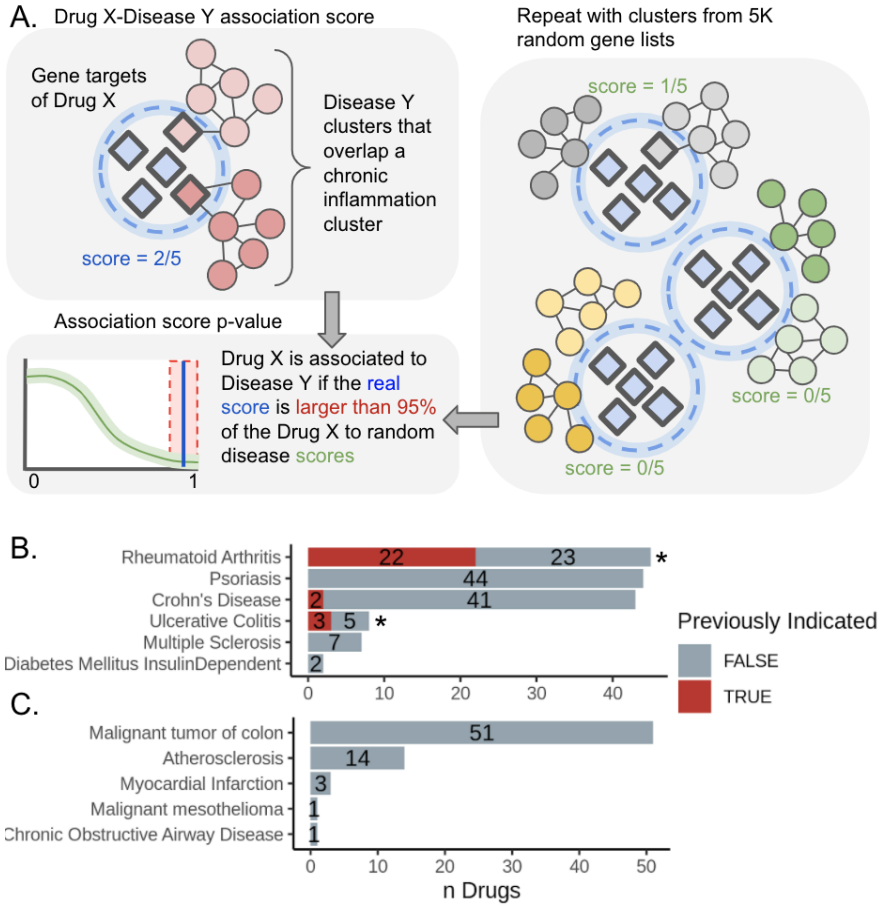

Figure 3: **A**. Schematic showing the drug scoring pipeline. **B**. The number of drugs identified by the current method that were previously indicated for autoimmune diseases (red) out of the total number of drugs identified by the current method. *Denotes significant enrichment for drugs previously indicated to treat the disease or off-label use to treat the disease (FDR < .01, BH corrected, one-tailed Fisher's exact test) **C**. Same as B for complex diseases.

To determine if the drug-disease associations predicted by our method are actionable, we compared them with drugs tested for efficacy for diseases in clinical trials[33]. We found a significant enrichment of drug-disease pairs in an active, recruiting, or completed trial among our results ($p < 0.001$; **Table S4**), suggesting that these are the types of drug-diseases pairings experts would consider for a clinical trial, whether or not that trial was successful. We next tested the overlap between disease-drug pairs identified by us and those that were ever in a Phase IV clinical trial. Drugs in Phase VI are those that have already been proved effective for treating the disease (in Phase III) and are being monitored for long-term safety and efficacy. Drugs identified by us are significantly enriched for drugs reaching Phase IV trial for the associated disease ($p < 0.001$; **Table S4**). The top drug-disease pairs (based on FDR) identified by our

approach for complex diseases are shown in **Figure 4**. All the predictions are listed in **Table S5**.

4 Discussion

Complex diseases exhibit a staggering amount of heterogeneity, being associated with hundreds of genes and with a range of phenotypes. Therefore, to continue advancing our understanding of disease mechanisms and our ability to treat these diseases, it is critical to deconvolve disease heterogeneity by: a) resolving subset of disease genes (and cellular processes/pathways) that underlie specific disease-associated phenotypes, and b) identifying avenues to diagnostically and/or therapeutically target those specific phenotypes.

Here, we present a computational data-driven approach to address this critical need (**Fig. 1A**). By applying this approach to study chronic inflammation (CI) — a major phenotype in several common, complex diseases — we have identified novel treatment options to target the inflammatory components of five diseases. These treatment prioritizations are based on earlier steps in our approach that generate comprehensive lists of (known and predicted) disease-associated genes and identify the CI signal among these genes.

A key aspect of our approach is ensuring its sensitivity to detect CI disease signatures using autoimmune diseases as positive controls. In autoimmune diseases, the immune system mistakenly attacks healthy tissue causing long-term systemic inflammation. Thus, we expect that the underlying CI disease signatures would be easily identifiable by a valid approach. Indeed, in six of the seven autoimmune diseases analyzed, our approach isolated gene clusters enriched for CI genes (**Fig. 2A**), and identified drugs already used to treat rheumatoid arthritis, Crohn's disease, and ulcerative colitis (**Fig. 3B**). In fact, approximately half of the drugs we predicted to treat rheumatoid arthritis and ulcerative colitis were already indicated for those diseases (**Fig. 3B**). This finding is encouraging given that we conservatively matched drugs to diseases only based on expert-curated drug-target data from DGIDb[20] rather than using all drug-target information in DrugCentral[19].

Moreover, among all our predicted disease-drug pairs, there is an enrichment for disease-drug pairs that have been part of a clinical trial, indicating that our novel drug repurposing results are reasonable predictions to be tested experimentally and clinically. Out of the 32 drug-disease pairs predicted by our method that have been tested in a clinical trial, 6 are for complex diseases, demonstrating that our method extends beyond autoimmune disorders. This result is encouraging because, while systemic inflammation is prominent in autoimmune disorders, it is one among a plethora of phenotypes in complex diseases, which presents the challenge of teasing apart the CI signal and associating that signal with drugs.

To show that our method was not erroneously uncovering CI signals where there were none, we identified UK Biobank traits not patently associated with CI (along with their genes) to use as negative controls. Following this analysis, we found that the median fraction of trait-associated genes predicted by GenePlexus and the median fraction of genes assigned to sizable clusters were lower for these traits than for autoimmune and complex diseases (**Fig. 1D** and **E**). Given that GenePlexus is a method that leverages network connectivity for predicting new genes belonging to a set, these results suggest that the genes associated with non-disease traits may not be as highly connected to one another in either STRING or BioGRID as the autoimmune and complex disease genes. Moreover, a smaller proportion of the non-disease trait clusters were enriched with genes annotated to GO biological processes. Taken together, these results suggest that these chosen traits serve as reasonable negative controls and offer a way to meaningfully contrast the results from complex diseases. Ideally, diseases or traits with no underlying inflammatory component but with associated genes that cluster in a network (as well as the autoimmune and complex disease) will serve as better negative controls. Given how common inflammatory processes are in disease, however, such diseases are difficult to definitively identify.

Unlike the non-disease traits, most of the autoimmune and complex diseases have more than half of their associated genes stem from predictions by GenePlexus[25], rather than mostly coming directly from the DisGeNET database[29]. The curated gene-disease associations collected in DisGeNET originate from over a dozen repositories including ClinGen[35], Genomics England (https://panelapp.genomicsengland.co. uk/), Orphanet[36], and OMIM[37]. While DisGeNET also includes gene-disease associations extracted from text-mining of literature, and inferred from GWAS studies and animal models, the “seed genes” used in this study came from the curated sources only[29]. Approximately 60% of the target genes associating drugs to diseases in our study are predicted to be associated to the disease by GenePlexus, including all but one of the target genes for the top novel drug-disease pairs shown in **Figure 4** (*IL1B* is a seed gene for Malignant tumor of colon). Therefore, using GenePlexus to expand our gene lists was a crucial component of our approach.

There is literature support for a number of GenePlexus predicted disease-gene associations among the top drug-disease pairs. For example, our approach predicted that several anticoagulants including bivalirudin and dabigatran etexilate could be used to treat the inflammatory component of atherosclerosis. Atherosclerosis is a thickening of the arteries caused by a buildup of fats and cholesterol (plaque) on the artery walls, and inflammation plays an important role in the plaque formation process[38]. When the vascular endothelium is disrupted during atherosclerosis, it enters an proinflammatory state that leads to coagulation caused in part by an increase in the serine protease thrombin. Thrombin itself then increases inflammation by triggering the production of cytokines and chemokines[39]. Thrombin is encoded by the *F2* gene[39], a member of

CI-enriched Atherosclerosis cluster 8 (**Fig. 2C** and **Fig. 4**), and the anticoagulants predicted by our method to treat atherosclerosis are all thrombin inhibitors, meaning they may also decrease vascular inflammation.

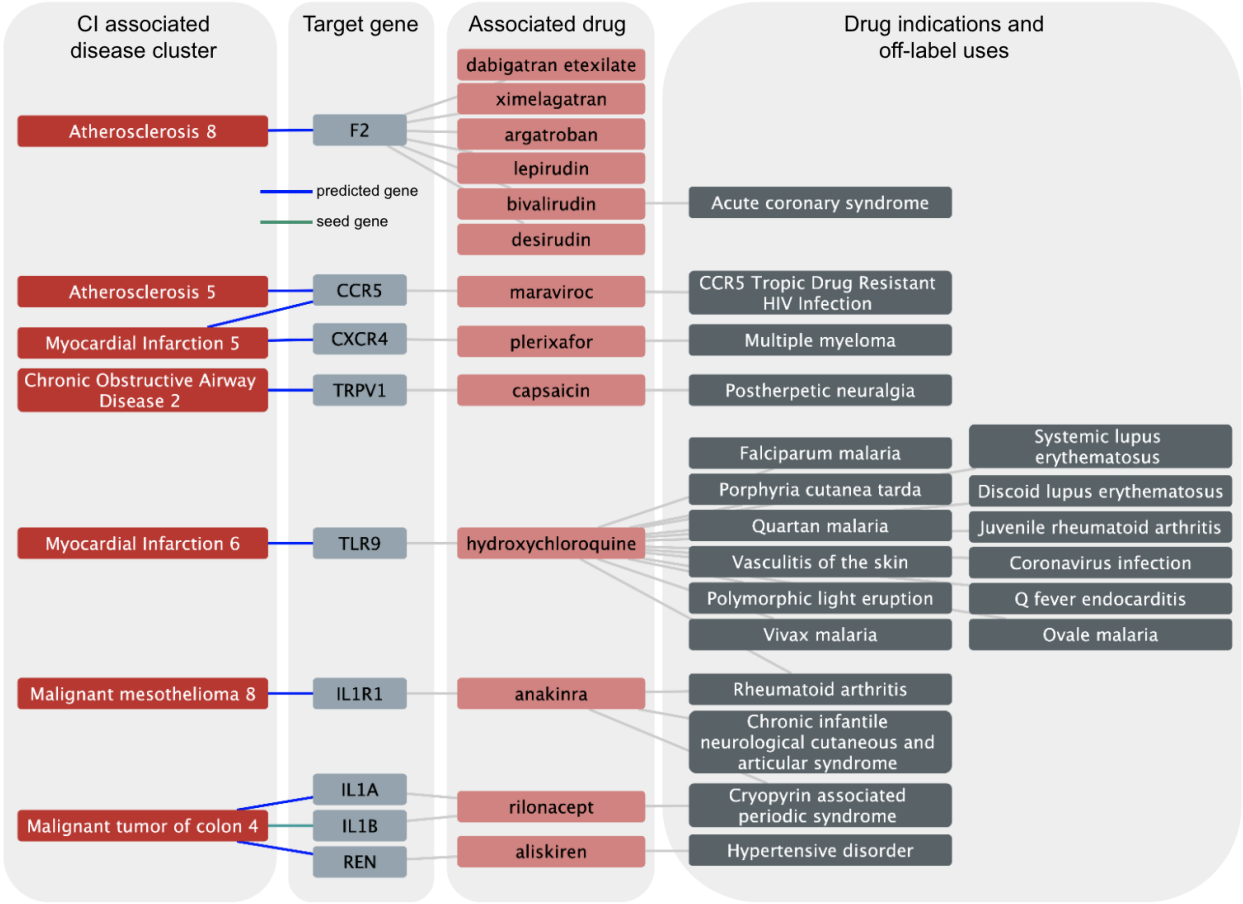

Figure 4: Top drug-disease associations by FDR, along with the target gene from the disease cluster and drug indications and off-label uses.

Our method also identified maraviroc, a C-C chemokine receptor type 5 (CCR5) antagonist as a potential therapy for atherosclerosis and myocardial infarction. CCR5 was first known as a co-receptor for HIV infection of host CD4+ T cells, and maraviroc is FDA approved for clinical use in HIV-1-infected patients carrying the CCR5 tropism for whom traditional antiretroviral therapies are ineffective[40]. Patients with the *CCR5-Δ32* variant of the *CCR5* gene, which encodes an inactive version of CCR5[41], have a lower incidence of acute myocardial infarction[42], and animal models of atherosclerosis have better outcomes when the gene is knocked out[43]. Moreover, in a pilot study of the efficacy of maraviroc in treating atherosclerosis in HIV positive patients, the drug significantly improved markers of cardiovascular risk[44]. While the idea of treating the inflammatory components of cardiovascular diseases with chemokine receptor

antagonists may not be novel[45], the presence of these drugs among our results lends validity to our drug repurposing strategy.

Similarly, the role of the interleukin-1 (IL-1) family of cytokines in carcinogenesis and tumor progression has long highlighted the potential of interleukin-1 inhibitors for treating cancer[46]. Recently, an additional analysis of the Canakinumab Anti-inflammatory Thrombosis Outcomes Study (CANTOS)[47] suggests that treatment with canakinumab, a monoclonal antibody that targets IL-1β, could reduce incident lung cancer and lung cancer mortality[48]. The genes encoding both IL-1α (*IL1A*) and IL-1β (*IL1B*) belong to the CI-enriched malignant tumor of colon cluster 4 (**Fig. 2C** and **Fig. 4**). Through these target genes, our approach predicted the IL-1 inhibitor, rilonacept, as a potential treatment for colon cancer. Additionally, we associated the gene encoding Interleukin-1 Receptor Type 1 (*IL1R1*), a receptor for IL-1α and IL-1β, with malignant mesothelioma, and our approach predicted anakinra, an interleukin-1 receptor antagonist, as a potential treatment for this disease. Anakinra is currently in clinical trials for treating cancers including chronic lymphocytic leukemia (NCT04691765), pancreatic cancer (NCT02550327, NCT04926467), and relapsed or refractory advanced cancers (NCT01624766) among others.

While our method's top predictions are largely reasonable, there are limitations. Our approach extracts information about gene-disease and gene-drug associations through databases that record the presence of associations but not the nature of those associations, which can lead to mismatches in drug-disease pairs. For example, our method predicted capsaicin, the chemical compound that makes chilli peppers spicy, as a treatment for chronic obstructive airway disease (COPD). As a drug, capsaicin is used topically to treat musculoskeletal pain, postherpetic neuralgia, and diabetic neuropathy, and is an activator of TRPV1 (transient receptor potential vanilloid 1), a non-selective excitatory cation channel expressed in sensory neurons that promote neurogenic inflammation[49]. As opposed to being therapeutic for COPD, TRPV1 activation may contribute negatively to the respiratory symptoms associated with the disease[49]. Therefore, TRPV1 antagonists, not agonists like capsaicin, might be candidates for disease treatment. While our method correctly identified an association between TRPV1 and COPD[50], as well as TRPV1 and capsaicin, treating COPD with capsaicin would clearly be counterproductive. Text mining approaches that would predict the nature of the gene-disease and gene-drug associations and check for concordance between the two would be a useful addition to our approach.

Overall, we have shown that our method is capable of isolating the chronic inflammation gene signature of a complex disease using a network-based strategy and, by integrating information across multiple complementary sources of data, it can predict and prioritize potential therapies for the systemic inflammation involved in that disease specifically. Importantly, our approach provides a blueprint for identifying and prioritizing therapeutic opportunities for any disease endophenotype.

Acknowledgements

We thank the members of the Krishnan Lab for helpful discussions.

Funding

This work was supported by the US National Institutes of Health (NIH) grant R35 GM128765 to AK and NIH Fellowship F32 GM134595 to CAM.

Author Contributions

SLH, AM, and AK conceived and designed the approach and experiments and wrote the manuscript; SLH and AM implemented the approach and performed the experiments. CAM wrote the GenePlexus code used in the experiments. All authors read, edited, and approved the final manuscript.
The authors have no conflicts of interest.

Availability of data, materials, and code

The data, main and supplemental results, and code used to reproduce this study are freely available at https://github.com/krishnanlab/chronic-inflammation.

References

1. Rock, K. L. & Kono, H. The inflammatory response to cell death. *Annu Rev Pathol* **3**, 99–126 (2008).
2. Furman, D. *et al.* Chronic inflammation in the etiology of disease across the life span. *Nat Med* **25**, 1822–1832 (2019).
3. Vos, T. *et al.* Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet* **396**, 1204–1222 (2020).
4. Dregan, A., Charlton, J., Chowienczyk, P. & Gulliford, M. C. Chronic Inflammatory Disorders and Risk of Type 2 Diabetes Mellitus, Coronary Heart Disease, and Stroke. *Circulation* **130**, 837–844 (2014).
5. Armstrong, A. W., Harskamp, C. T. & Armstrong, E. J. Psoriasis and the Risk of Diabetes Mellitus: A Systematic Review and Meta-analysis. *JAMA Dermatology* **149**, 84–91 (2013).
6. Yashiro, M. Ulcerative colitis-associated colorectal cancer. *World J Gastroenterol* **20**, 16389–16397 (2014).
7. Fullerton, J. N. & Gilroy, D. W. Resolution of inflammation: a new therapeutic

frontier. *Nat Rev Drug Discov* **15**, 551–567 (2016).

8. Chou, R. C., Kane, M., Ghimire, S., Gautam, S. & Gui, J. Treatment for Rheumatoid Arthritis and Risk of Alzheimer's Disease: A Nested Case-Control Analysis. *CNS Drugs* **30**, 1111–1120 (2016).
9. Antohe, J. I. *et al.* Diabetes mellitus risk in rheumatoid arthritis: Reduced incidence with anti–tumor necrosis factor α therapy. *Arthritis Care & Research* **64**, 215–221 (2012).
10. Shah, M. *et al.* Cumulative Burden of Oral Corticosteroid Adverse Effects and the Economic Implications of Corticosteroid Use in Patients With Systemic Lupus Erythematosus. *Clinical Therapeutics* **35**, 486–497 (2013).
11. Rosenblum, H. & Amital, H. Anti-TNF therapy: Safety aspects of taking the risk. *Autoimmunity Reviews* **10**, 563–568 (2011).
12. Murdaca, G. *et al.* Infection risk associated with anti-TNF-α agents: a review. *Expert Opinion on Drug Safety* **14**, 571–582 (2015).
13. Kanis, J. A. *et al.* A Meta-Analysis of Prior Corticosteroid Use and Fracture Risk. *Journal of Bone and Mineral Research* **19**, 893–899 (2004).
14. Mitra, R. Adverse Effects of Corticosteroids on Bone Metabolism: A Review. *PM&R* **3**, 466–471 (2011).
15. Blackburn, D., Hux, J. & Mamdani, M. Quantification of the Risk of Corticosteroid-induced Diabetes Mellitus Among the Elderly. *Journal of General Internal Medicine* **17**, 717–720 (2002).
16. Griffin, M. R. Epidemiology of Nonsteroidal Anti-inflammatory Drug–Associated Gastrointestinal Injury. *The American Journal of Medicine* **104**, 23S-29S (1998).
17. Marcum, Z. A. & Hanlon, J. T. Recognizing the Risks of Chronic Nonsteroidal Anti-Inflammatory Drug Use in Older Adults. *Ann Longterm Care* **18**, 24–27 (2010).
18. Griffin, M. R., Yared, A. & Ray, W. A. Nonsteroidal antiinflammatory drugs and acute renal failure in elderly persons. *Am J Epidemiol* **151**, 488–496 (2000).
19. Avram, S. *et al.* DrugCentral 2021 supports drug discovery and repositioning. *Nucleic Acids Research* **49**, 1160–1169 (2021).
20. Freshour, S. *et al.* Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. *Nucleic Acids Research* **49**, 1144–1151 (2021).
21. Schriml, L. M. *et al.* Human Disease Ontology 2018 update: classification, content and workflow expansion. *Nucleic Acids Res* **47**, D955–D962 (2019).
22. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine* **12**, (2015).
23. Abbot, L. *et al.* UK Biobank GWAS. *nealelab* http://www.nealelab.is/uk-biobank/.

24. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. *PLOS Computational Biology* **12**, (2016).
25. Liu, R., Mancuso, C. A., Yannakopoulos, A., Johnson, K. A. & Krishnan, A. Supervised-learning is an accurate method for network-based gene classification. *Bioinformatics* (2020) doi:10.1093/bioinformatics/btaa150.
26. Szklarczyk, D. *et al.* The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res* **45**, D362–D368 (2017).
27. Stark, C. *et al.* BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* **34**, D535–D539 (2006).
28. Wong, A. K., Krishnan, A. & Troyanskaya, O. G. GIANT 2.0: genome-scale integrated analysis of gene networks in tissues. *Nucleic Acids Res* **46**, W65–W70 (2018).
29. Piñero, J. *et al.* The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Research* **48**, D845–D855 (2020).
30. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports* **9**, 5233 (2019).
31. Alexa, A. & Rahnenfuhrer, J. *topGO: Enrichment Analysis for Gene Ontology*. (Bioconductor version: Release (3.14), 2022). doi:10.18129/B9.bioc.topGO.
32. Carlson, M. *org.Hs.eg.db: Genome wide annotation for Human*. (2019).
33. AACT Database | Clinical Trials Transformation Initiative. https://aact.ctti-clinicaltrials.org/.
34. Szklarczyk, D. *et al.* STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* **43**, D447–D452 (2015).
35. Rehm, H. L. *et al.* ClinGen — The Clinical Genome Resource. *N Engl J Med* **372**, 2235–2242 (2015).
36. Rath, A. *et al.* Representation of rare diseases in health information systems: The orphanet approach to serve a wide range of end users. *Human Mutation* **33**, 803–808 (2012).
37. Amberger, J. S., Bocchini, C. A., Scott, A. F. & Hamosh, A. OMIM.org: leveraging knowledge across phenotype–gene relationships. *Nucleic Acids Research* **47**, D1038–D1043 (2019).
38. Russell, R. Atherosclerosis — An Inflammatory Disease. *The New England Journal of Medicine* 12 (1999).
39. Kalz, J., ten Cate, H. & Spronk, H. M. H. Thrombin generation and atherosclerosis. *J Thromb Thrombolysis* **37**, 45–55 (2014).

40. Parra, R. G. *et al.* Reconstructing complex lineage trees from scRNA-seq data using MERLoT. *Nucleic Acids Res* doi:10.1093/nar/gkz706.
41. Liu, R. *et al.* Homozygous Defect in HIV-1 Coreceptor Accounts for Resistance of Some Multiply-Exposed Individuals to HIV-1 Infection. *Cell* **86**, 367–377 (1996).
42. González, P. *et al.* Genetic variation at the chemokine receptors CCR5/CCR2 in myocardial infarction. *Genes Immun* **2**, 191–195 (2001).
43. Zhao, Q. Dual targeting of CCR2 and CCR5: therapeutic potential for immunologic and cardiovascular diseases. *Journal of Leukocyte Biology* **88**, 41–55 (2010).
44. Francisci, D. *et al.* Maraviroc Intensification Modulates Atherosclerotic Progression in HIV-Suppressed Patients at High Cardiovascular Risk. A Randomized, Crossover Pilot Study. *Open Forum Infectious Diseases* **6**, ofz112 (2019).
45. Noels, H., Weber, C. & Koenen, R. R. Chemokines as Therapeutic Targets in Cardiovascular Disease. *Arteriosclerosis, Thrombosis, and Vascular Biology* **39**, 583–592 (2019).
46. Dinarello, C. A. Why not treat human cancer with interleukin-1 blockade? *Cancer Metastasis Rev* **29**, 317–329 (2010).
47. Ridker, P. M. *et al.* Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. *N Engl J Med* **377**, 1119–1131 (2017).
48. Ridker, P. M. *et al.* Effect of interleukin-1β inhibition with canakinumab on incident lung cancer in patients with atherosclerosis: exploratory results from a randomised, double-blind, placebo-controlled trial. *Lancet* **390**, 1833–1842 (2017).
49. Geppetti, P., Materazzi, S. & Nicoletti, P. The transient receptor potential vanilloid 1: Role in airway inflammation and disease. *European Journal of Pharmacology* **533**, 207–214 (2006).
50. Belvisi, M. G. & Birrell, M. A. The emerging role of transient receptor potential channels in chronic lung disease. *European Respiratory Journal* **50**, (2017).