Integration of immunome with disease gene network reveals pleiotropy and novel drug repurposing targets

Abhinandan Devaprasad[1], Timothy RDJ Radstake[1], Aridaman Pandit[1*]

[1] Center for Translation Immunology, University Medical Center Utrecht, Utrecht, The Netherlands

***Correspondence:** Aridaman Pandit: a.pandit@umcutrecht.nl

Abstract

Immune system is crucial for the development and progression of immune-mediated and non-immune mediated complex diseases. Studies have shown that multiple complex diseases are associated with several immunologically relevant genes. Despite such growing evidence, the effect of disease associated genes on immune functions has not been well explored. Here, we curated the largest *immunome* (transcriptome profiles of 40 different immune cells) and integrated it with disease gene networks and drug-gene database, to generate a Disease-gene IMmune cell Expression network (DIME). We used the DIME network to: (1) study 13,510 genes and identify disease associated genes and immune cells for >15,000 complex diseases; (2) study pleiotropy between various phenotypically distinct rheumatic and other non-rheumatic diseases; and (3) identify novel targets for drug repurposing and discovery. We implemented DIME as a tool (https://bitbucket.org/systemsimmunology/dime) that allows users to explore disease-immune-cell associations and disease drug networks to pave way for future (pre-) clinical research.

Introduction

The genetic and epigenetic heterogeneity has been known to play a major role in the development and progression of complex diseases. The past two decades have seen a major surge in studies that characterize genes and loci associated with disease. The use of high-throughput omics technology and functional screenings have boosted our knowledge about genetic, epigenetic and metabolic factors underlying complex diseases[1]. As a result of these genetic and epigenetic screenings, we now know that the majority of complex diseases and genes/loci have a many-to-many relationship meaning that a complex disease is linked to many different genes and a gene/loci is associated with many different genes[2].

Large high-throughput screening studies have typically used bulk tissue or whole blood to study disease associated genes (DAGs). However, the expression of each gene is known to vary between tissues and cell types[3,4]. Thus, bulk tissue- or blood-based studies on DAGs do not consider the role played by different cells and tissues in the disease biology. To improve the understanding and molecular basis of complex diseases, a large number of research groups and consortiums have started to functionally identify disease associated cells (DACs) or tissue types[3–7]. The Genotype-Tissue Expression (GTEx) is one such valuable project, which maps gene expression profiles of 54 different human tissue types and the corresponding expression quantitative trait loci (eQTLs)[5–7]. Furthermore, the growth of single cell technologies have advanced our understanding of DACs and have helped in identifying cell types associated with complex diseases including cancer[8], Alzheimer's[9], rheumatoid arthritis[10], among others.

The immune system is known to play a key role in the development and progression of immune-mediated as well as non-immune mediated chronic diseases. A large number of association and functional studies have shown that multiple DAGs are expressed in immune cells and perturbing these DAGs can modulate immune cell functions[11]. However, very few studies have explored the impact of DAGs on specific cell types and even fewer on immune cells, many of which focus on limited number of cell subsets[12–16]. Recently Schmiedel *et al.* studied the effect of genetic variants on the expression of genes in 13 different immune cell types[17]. However, this study largely focused on the analysis of genetic variants and their impact on a total of 13 immune cell types: monocytes (classical and non-classical), NK cells, naïve B-cells and nine sub-populations of T-cells.

In this study, we mapped the largest available and expert curated disease-gene network (from the DisGeNet curated from 16 different databases) on the largest *immunome* data curated by us comprising gene expression profiles of 40 different immune cell types. We then quantified the effects of 13,510 DAGs on the *immunome*, to identify DACs for 15,367 different diseases in the DisGeNet. Using the DACs and the DAGs, we constructed the Disease-gene IMmune cell Expression (DIME) network. We use the DIME network to: (1) study the underlying cell-specific mechanisms of complex diseases; (2) identify cell-specific targets for complex disease; (3) identify networks of genes and cells that are commonly associated with different pairs of diseases; and (4) predict drug repurposing targets towards identified disease mechanisms shared between different diseases. We further built a user-friendly shinyapp called DIME (https://bitbucket.org/systemsimmunology/dime), which can be used to identify DACs and construct DIME network for: (1) diseases from the DisGeNet, (2) diseases from the EBI genome wide association study (GWAS) catalogue, or (3) custom set of genes defined by the user.

Methods

Transcriptome data - *Immunome*

The transcriptome data consists of RNA-sequencing datasets of 40 different immune cell types curated using 316 samples from a total of 27 publicly available datasets (see **Supplementary Table 1** for list of GEO datasets and samples used). The 40 different immune cells cover the entire hematopoietic stem cell differentiation tree comprising of 9 progenitors, 19 lymphoid, and 12 myeloid cell types. The samples used here were manually curated considering only the unstimulated (except for macrophages, that were monocyte derived) immune cells that were sorted using Fluorescence-activated cell sorting (FACS) and were isolated from either blood, bone marrow or cord blood from healthy donors. All the selected datasets were downloaded as FASTQ files using the fastq-dump tool from sratoolkit[18]. The "—split-files" option was given if the library type was paired end sequencing. FASTQ files were then aligned to reference genome (GRCH.Hg38.79) using STAR aligner[19]. The result is a SAM file which was then converted into a sorted BAM file using the samtools program[20]. These were then used to calculate the count of aligned reads using the HTSeq program[21] with the mode option "intersection non-empty". HTSeq was run

for all possible stranded mode options, the count file with the maximum counts was chosen as the respective count file for the sample.

The data was then filtered by removing all genes that had less than 20 read counts in 95 percent of the samples using R programming. The filtered data was then lane normalized using the "betweenLaneNormalization" function from the RUVSeq package[22]. The RUVr method from RUVSeq was used to identify residual factors contributing to the batch effect. The resulting filtered, batch corrected and normalized data had expression for 34,906 genes that was void of any observable batch effect. We calculated counts per million (cpm) for the filtered genes and used cpm as the gene expression measure. We then used the median gene expression for each cell type for the rest of the analysis. This processed, batch corrected, normalized and median representative data of 40 immune cells is referred to as the *immunome*.

Disease gene network from DisGeNet

The full disease gene association network from DisGeNet[23] was downloaded from the DisGeNet database (www.disgenet.org/downloads). All HLA associated genes was removed from the network, this was done to ensure that bias towards myeloid cells and B cells are removed, since the HLA genes are largely expressed by these cells. The resulting network was further filtered to include only those genes that were present in the *immunome*. The final network comprised of 15367 diseases and 13510 DAGs.

The DisGeNet consists of expert curated disease-gene interactions from 16 different databases: UNIPROT, CGI, ClinGen, Genomics England, CTD, PsyGeNET, Orphanet, RGD, MGD, CTD, Human Phenotype Ontology, Clinvar, GWAS catalogue, GWAS DB, LHDGN and BeFree. The DisGeNet is the largest and most comprehensive disease-gene association network available in the literature that was known to us. We also tested our methods on more specific disease networks such as those from the EBI GWAS database.

Other disease gene networks - EBI GWAS data

In addition to the DisGeNet, we also used a refined GWAS based dataset from the EBI[24]. The GWAS catalogue of Version 1.0, e89, was downloaded from the EBI website, which

contained information on the disease associated SNP for about 1900 diseases/traits. The reported p-value of all the disease associated SNP in the catalogue was ≤ 0.05. The catalogue also provided the corresponding mapped gene information for all the SNP which was used to construct the disease to gene association network. We further filtered this network using the same filtering criteria that was used for the DisGeNet. The EBI GWAS dataset was used to infer SNP based disease cell associations.

Mapping disease gene network to *Immunome* data

For a given disease *D* and its DAG_D, we first extracted the corresponding *Immunome* expression matrix. This expression matrix (X_D) comprised the gene expression of the DAG_D across the 40 cells forms as the input data upon which further analysis was performed. Thus, the dimension of each X_D was given as:

$$dim(X_D) = length(DAG_D) \times 40,$$

where, $length(DAG_D)$ is the number of DAGs in disease *D* and 40 corresponds to the number of cell types in the *immunome* data.

Using NMF to cluster X_D into *k* classes

We used the NMF package[25] in R and applied the non-negative matrix factorization method using Brunet's[26] algorithm to the expression matrix (X_D) to factor it into two matrices namely W_D and H_D such that.

$$X_D \approx W_D H_D,$$

where, W_D and H_D are the basis and coefficient matrices computed by the NMF. The dimensions of W_D and H_D are given as:

$$dim(W_D) = length(DAG_D) \times k,$$

$$dim(H_D) = k \times 40,$$

where, *k* is the number of classes/clusters that splits the data, such that it satisfies the above NMF equation. The W_D matrix comprises of the weights of the DAGs across the *k* clusters (in each column) and the H_D matrix comprises of the weights of the cells in the corresponding *k*

clusters (in each row). We used Brunet *et al.* method to identify the ideal *k* value using the cophenetic correlation coefficient method[26].

Identifying the key DAG_D and DAC_D from W_D and H_D

The NMF algorithm clusters the data into *k* clusters such that, in each cluster '*i*', where $i \in (1, \ldots, k)$, the genes that have high values in w_D^i are constitutively expressed by the cells that have high values in h_D^i. Where, w_D^i is the i^{th} column of W_D and h_D^i is the i^{th} row of H_D. For each cluster *i*, we chose the cell-gene pairs that were in the top 25th percentile range of their corresponding w_D^i and h_D^i values. These cells and their corresponding gene pairs are regarded as the key DAC_D and DAG_D respectively. The cell-gene pairs were extracted from all the clusters and were compiled together. The resulting cell-gene pairs of all the clusters form the edges of the **D**isease-gene to **IM**mune cell **E**xpression network, hereby referred as the **DIME** network.

Identifying the key cluster

We then identified the largest weighted cluster among the *k* clusters identified by the NMF. That is, the subset of genes and cells of X_D that can capture most of its expression pattern. We did this by using the following approach.

Since X_D, W_D and H_D can be represented as below:

$$X_D \approx W_D H_D = \begin{bmatrix} | & | & | & | \\ w_D^1 & w_D^2 & \ldots & w_D^k \\ | & | & | & | \end{bmatrix} \begin{bmatrix} - & h_D^1 & - \\ - & h_D^2 & - \\ - & \vdots & - \\ - & h_D^k & - \end{bmatrix} = \sum_{i=1}^{k} w_D^i \, h_D^i.$$

We calculated the Frobenius norm of each $w_D^i h_D^i$ for all values of *i*. We then identified the cluster (represented as *c*) for which $||w_D^i h_D^i||_F$ is the maximum. This can be mathematically represented as:

$$c = \text{argmax}\left(||w_D^i h_D^i||_F\right);\ i \in \{1, \ldots, k\},$$

where, the c^{th} cluster represents that cluster which maximally captures/represents the expression matrix X_D. We used the w_D^c as the scores for the DAG_D and h_D^c as the scores for the DAC_D.

$$DAG_D \text{ score} = w_D^c,$$

$$DAC_D \text{ score} = h_D^c,$$

The scores were scaled between 0 and 1, with 1 representing the cell or gene with the highest score.

Pleiotropic associations

To identify common mechanisms between two diseases, we looked at their overlapping cell-gene connections in their corresponding DIME networks. Jaccards' index (JI) was used to measure the overlap between the two diseases with Fisher's exact test (FET) used to obtain confidence p-value for the given overlap. The overlapping genes were used to calculate JI and statistical significance of overlap using FET. The pleiotropy based overlapping cell-gene network between the two diseases is referred to as the DIME-pleiotropy network.

Integrating drug-gene network

The drug to gene target database from DGIdb was downloaded[27]. The data was filtered to contain only the CHEMBL interactions and only those pertaining to the drugs approved by the food and drug administration (FDA) of USA. This FDA approved drug to gene target interaction serves as the drug-gene network in this study. To identify potential drug candidates from the drug-gene network for disease *D*, we choose its key DAG_D as identified in the previous step and extracted its corresponding target drugs from the drug-gene network. This network between the drugs and the key DAG_D is referred to as the DIME-drug network for disease D. The DIME-drug network represents all the drugs that target the key DAG_D identified in the given DIME network. To identify potential common acting drugs between different diseases, we used their corresponding DIME-pleiotropy network and extracted the drug-gene connections between the cell-gene and the drugs in drug-gene network. The drugs identified using this approach represent those that act on the common mechanisms between the two diseases and can be potential candidates for drug repurposing.

Rationale for restricting analysis to only diseases with ≥ 100 DAGs

The filtered disease-gene network of the DisGeNet comprised of DAGs for about 15,367 diseases. This could make the pleiotropy analysis between diseases extremely cumbersome, with the number of disease-to-disease comparisons reaching as large as 118,064,661. Additionally, to have sufficient number of DAC-DAG associations in the DIME-pleiotropy network and also have sufficient number of drug-target gene associations in the DIME-drug network, we used a smaller subset of diseases but with larger DAGs associated to them. We found by looking at the distribution of DAGs across the diseases, that the number of diseases with ≥ 100 DAGs were fewer in number, i.e., 613 diseases. This was sufficiently large DAC-DAG associations for the pleiotropy analysis and less cumbersome to analyze the comparisons (187,578). Hence, we choose those diseases with ≥ 100 DAGs. All analysis presented in this study have been performed on this disease subset.

DIME shinyapp

To construct the DIME and DIME-drug network for other diseases not mentioned in this study or using a custom gene input, we built a tool in R/Bioconductor called DIME that is available as a shiny app (https://bitbucket.org/systemsimmunology/dime). The app can be used to identify the key DAC, DAG and the DIME-drug network for the diseases from the DisGeNet, the EBI-GWAS catalogue or for custom set of genes from the user.

Results

The aim of this study is to identify disease associated cell types (DACs) based on the existing disease gene network, and to further identify disease associated gene (DAG) subsets that may perturb the associated immune cells. To achieve this, we integrated our *immunome* data with the disease network from DisGeNet and used the non-negative matrix factorization (NMF) method to identify those subsets of genes and cells that maximally represent the **D**isease gene **IM**mune cell **E**xpression (DIME) network, as illustrated in **Figure 1A**. The constructed immunome dataset comprises 40 different cell types and is the largest bulk RNASeq meta-dataset of the immune cell types known to us. The filtered disease-gene network from

DisGeNet consists of associations between 15,367 diseases and 13,510 DAGs. In this massive disease network, numerous DAGs were found to be common between both phenotypically similar and distinct diseases. We explore these common DAG patterns in more detail in the next section.

Common DAGs of phenotypically distinct diseases

The 15,367 diseases in the DisGeNet belong to 29 different disease MeSH (Medical Subject Headings) terms (**Figure 1B**). The MeSH-MeSH network in **Figure 1B** depicts the connections between the different MeSH terms, where the thickness of the connections represents the number of DAGs common between the different MeSH terms. The neoplasm MeSH term was the most well connected disease category in the network. The highest number of common DAGs (6,959 DAGs) between two different MeSH terms was observed between neoplasm and digestive system diseases. Other top MeSH-MeSH connections that had more than 5,000 common DAGs include those between neoplasm and that of the skin and connective tissue, nervous system, congenital, endocrine system, and female urogenital diseases and pregnancy complications. Thus, the MeSH-MeSH gene network revealed the shared DAGs across phenotypically distinct diseases belonging to different MeSH terms.

We further studied the DAG patterns across MeSH terms and found that TP53 was preferentially associated with diseases categorized under the neoplasm MeSH term (**Figure 1C**). Interestingly, TP53 was also associated with numerous diseases from various other MeSH terms such as immune system diseases, nervous system diseases, and skin and connective tissue diseases. Similarly, TNF was associated with numerous diseases from various MeSH terms including immune system diseases (**Figure 1D**). APOE was largely associated with nervous system diseases, and ACE was largely associated with cardiovascular diseases (**Figure 1E** and **F**). TLR4 and CXCL8 were prevalently associated to several MeSH terms (**Figure 1G** and **H**). The above-mentioned genes (**Figure 1C-1F**) are those that are either the top represented (degree) genes across all diseases or for disease within specific MeSH terms. More examples of DAGs associated with specific MeSH terms and their degree can be found in the **Supplementary Figure 1**. Majority of the DAGs mapped to two or more MeSH terms (**Figure 1I**) or diseases (**Figure 1J**) also hinting towards shared disease mechanisms between phenotypically distinct diseases. These genes that are associated to many diseases/MeSH terms (i.e. having high degree) were regarded as the hub genes of the

disease-gene network. We ranked the DAGs based on their degree and found TP53, TNF, VEGFA, BCL2, IL1B as the top 5 DAGs each being associated with >750 diseases (**Figure 1K** and **Supplementary Figure 1**).

Relevance of the immune system

We further evaluated the expression of the hub genes identified in the previous step, in the *immunome* data. The rationale was to assess if the hub genes representing the core of the disease-gene network were important to the immune system. Indeed, we found that many of the hub genes of the disease-gene network were expressed constitutively by either all immune cells or by some specific immune cells, as seen in **Figure 1K**. For example, the nervous system associated gene, APOE, important to many of neurological diseases such as Alzheimer's and Parkinson[28] were found to be expressed specifically by macrophages, as seen in **Figure 1L**. Studies have shown that genetic polymorphisms in APOE protein leads to defective clearance of the Aβ plaques by macrophages[29–31]. Thus, conferring macrophages as one of the key players of the disease. Such links between DAGs, to observing altered function in a cell type, questions the need to study DAGs by integrating cell-specific expression information.

Identifying disease associated cell types

We further identified such disease associated cells (DACs) based on the DAGs and report DACs for about 600 diseases in the disease-gene network (see Methods). The immune cells can be broadly categorized as myeloids, lymphoids and progenitors. Using the DAC profiles of about 600 diseases, we observed that in most diseases, DAGs do not associate with all immune cells but tend to associate with specific immune cells or with specific category of immune cells (**Supplementary Figure 2**). We observed from these DAC profiles that several phenotypically different diseases had similar DAC profiles. For example, diseases such as skin carcinoma, muscle degeneration, juvenile arthritis, epilepsy, etc. clustered together, showing progenitors as their top DACs. Similarly, peripheral T-cell lymphoma, celiac disease, atopic dermatitis, malignant glioma, basal cell carcinoma, etc. clustered together, with lymphoid cells as the top DACs. Interestingly, systemic lupus erythematosus (SLE), heart failure, myocardial infarction, colitis, type 1 diabetes, etc. cluster together showing

myeloid cells as their top DACs. Thus the DAC profiles like the DAG profiles showed that phenotypically different diseases had similar DAC pattern.

We constructed the DIME network for different diseases individually, the DIME network consists of the top DACs and their respective DAGs, (see methods). The DIME network represents the key cell-gene mechanisms that can be drawn from the given set of DAGs. As proof of concept, we present here the DIME network of lymphoid leukemia (**Figure 2A**), a group of blood cancer that typically affects the lymphocytes. The DIME network revealed 4 DAC clusters for lymphoid leukemia. The key DAC cluster comprised of lymphoid progenitor cells as well as myeloid progenitor cells. The key DAGs contributing to this cluster included genes associated with hematopoiesis such as RPS14, HSP90AA1, MPO, ETV6, ATF4 and TAL1. The other key DAC cluster comprised of all the subsets of T cells, primarily the $CD4^+$ T cells. The key DAGs contributing to this cluster included ETS1, CXCR4, IKZF1, ATM, LCK, and KMT2A. The pathway enrichment (**Supplementary Figure 4B**) of these genes revealed pathways associated to TCR and BCR signaling and PI3K/AKT signaling. Interestingly, these pathways have been shown to be important for survival of cancer cells and are targets for anti-cancer drugs in acute and chronic lymphoid leukemia[32]. Thus the DIME network of lymphoid leukemia revealed the key DAGs and DACs implicated in the disease.

We then explored the DIME network of different kinds of rheumatic and/or fibrotic diseases, such as systemic scleroderma (MeSH: skin disease), pulmonary fibrosis (MeSH: respiratory tract disease) and SLE (MeSH: immune system disease). Moreover, we aimed to characterize the diseases on the basis of the DAGs, DACs and the accompanying DIME networks. The DIME network of systemic scleroderma, an autoimmune rheumatic condition with chronic inflammation and fibrotic phenotype, revealed a complex relationship between their DAGs and their DACs (**Figure 2B**). For systemic scleroderma, myeloid cells (neutrophils, granulocytes, $BDCA1^+CD14^+$, and CD11c myeloid dendritic cells), and lymphoid cells (NK, $CD4^+$ T regulatory, ILC3, and ILC2) were identified as key cluster of DACs (see methods). The key DAGs contributing to this cluster included PTPRC, FOS, SRRM2, MSN, JUNB, CXCR4, ITGB2, and TNFAIP3 genes. For systemic scleroderma, the other key DAC cluster comprised myeloid cells (macrophages and pDCs) and myeloid progenitor cells. The key DAGs contributing to this cluster included HSP90AB1, SPP1, HSP90AA1, MMP9, WNK1, HIF1A and IRF8. The two systemic scleroderma associated DAC clusters together were enriched in DAGs from interleukin signaling, TGF beta signaling, TLR signaling, ECM

organization and neutrophil degranulation pathways (**Supplementary Figure 3A**). Of which, evidently, TGF beta is known to play an essential role in the pathogenesis of fibrosis[33]. Likewise, the neutrophil degranulation pathway or otherwise known as NETosis, is the mechanism by which neutrophils exhibit defensive mechanisms to trap and kill foreign bodies[34]. And this pathway has been found to be implicated in other autoimmune diseases, and is now being currently investigated in clinical trials[35]. Interestingly, we found neutrophils and its NETosis associated genes to be in the top cluster of the DIME network for systemic scleroderma.

We then constructed the DIME network of another fibrotic disease, pulmonary fibrosis. The DIME network revealed 4 DAC clusters for pulmonary fibrosis (**Figure 2C**). Similar to systemic scleroderma, we found a DAC cluster comprising of T-cells and NK cells enriched in DAGs associated with TGF beta signaling (**Supplementary Figure 3B**). Interestingly, we also found that for pulmonary fibrosis neutrophils-granulocytes, macrophages, $BDCA1^{+}CD14^{+}$ cells, pDCs and myeloid progenitors were among the top DACs similar to that of systemic scleroderma, (**Supplementary Figure 3B**). We also found an enrichment of NLRP3 pathway genes, which are known to play a role in the collagen deposition mechanisms commonly dysregulated in fibrosis[36].

We then studied the DIME network of SLE, an autoimmune disease. The DIME network revealed 2 DAC clusters for SLE (**Figure 2D**). The key cluster comprised of myeloid cells like neutrophils-granulocytes, macrophage M1, $BDCA1^{+}CD14^{+}$ and monocytes as the DACs. The key DAGs contributing to this cluster included CD74, FOS, LYZ, SOD2 and HSP90AB1. The other cluster comprised of $CD4^{+}$ TEMRA, $CD4^{+}$ TEM, $CD4^{+}$ TCM and $CD4^{+}$ TH1 as their DACs. The key DAGs contributing to this cluster included B2M, IGHG3, IL7R, ETS1, RPS19, and TNFAIP3. The pathway enrichment for the SLE DIME network includes pathways that are heavily described in the literature for SLE, such as the neutrophil degranulation pathways or NETosis[37–41], interleukin 4 and interleukin 13 signaling[42,43], TLR signaling pathway[44], translocation of ZAP70 to immunological synapsis[45], and immune-regulatory interactions between lymphoid and non-lymphoid cells (**Supplementary Figure 4A**).

The DIME networks revealed immune system mediated mechanisms for the different diseases. As observed in the **Figure 2**, **Supplementary Figure 3** and **4**, the rheumatic and/or fibrotic diseases such as SLE and SSc, had overlap in specific cell-gene mechanisms

including those related to the neutrophil degranulation, degradation of ECM, interleukin and TGF beta signaling pathways. These pathways and common mechanisms observed in the DIME networks could serve as additional layers to understand the pathogenesis of these type I interferon driven diseases[46]. We further explored such common mechanisms based on the cell-gene relationships in the next section.

Pleiotropy based on DIME

Pleiotropy is when one gene affects two or more diseases[47]. Pleiotropy has been observed in several studies for many different diseases based on gene mutation to phenotype associations[48]. Unlike most pleiotropy studies performed in the past that looks at only the common DAGs between two diseases, we used a different approach. Since the DIME networks from the previous analysis (for example, between SSc and pulmonary fibrosis) revealed several overlapping cell-gene connections, we extended this approach to look for cell-gene connections that are common between the DIME networks of all possible pairs of diseases. Hence, we define pleiotropy in this study as the cell-gene connections that are found in one or more diseases. Using this approach, we constructed the pleiotropy network as shown in **Figure 3A**. The pleiotropy network consists of diseases as nodes. The nodes are connected if there exists a significant (JI $\geq$ 0.1and FET p-value $\leq$ 0.01) number of common cell-gene connections between their corresponding DIME networks, (see methods). The network shown in **Figure 3A** is trimmed to contain only those nodes with degree $\geq$ 2. The node colors represent the MeSH term (as shown in **Figure 1B**) of the disease. The grey colored nodes represent those diseases for which the MeSH term ontology was not available from the DisGeNet. As seen from **Figure 3A**, several cancer related diseases (shown by orange color for neoplasm MeSH term) cluster together. Same can be observed for the eye related disease (shown by green color MeSH term) and the chemically induced disorders (shown by pink and yellow color MeSH terms) seen in the bottom of the network **Figure 3A**. These clusters together highlight the similar mechanisms (cell-gene connections) between these diseases.

Within this pleiotropy network, we further looked at diseases that belong to different MeSH terms and searched for common cell-gene mechanistic patterns between them. To do so, we constructed the DIME-pleiotropy network of several pairs of phenotypically distinct rheumatic diseases within the pleiotropy network. We present here some examples of the

DIME-pleiotropy network that had a JI similarity of ≥ 0.1. In the DIME-pleiotropy network (**Figure 3B**) of Crohn's disease (MeSH: digestive system disease) and psoriasis (MeSH: skin disease), we found cell-gene networks of lymphoid ($CD4^+$, NK and ILC's) and myeloid cells. The pathway enrichment of the DAGs in this DIME-pleiotropy network revealed pathways related to TCR signaling, interleukin signaling, neutrophil degranulation and regulation of TLR (**Supplementary Figure 5A**). As observed in the DIME-pleiotropy network (**Figure 3B**), $CD4^+$ TH1 cells have been implicated in the pathogenesis of both Crohn's disease and psoriasis[49,50]. Likewise, in the DIME-pleiotropy network (**Figure 3C**) of SSc (MeSH: skin disease) and pulmonary fibrosis (MeSH: respiratory tract disease), we found cell-gene networks between progenitors and pDCs, NK cells, $CD4^+$ TEMRA and $CD4^+$ T regulatory, and other myeloid cells. The pathway enrichment of the DAGs in this DIME-pleiotropy network revealed pathways (**Supplementary Figure 5B**) related to TGF beta receptor signaling, interleukin, ECM and integrin cell surface interactions. Both of these being fibrotic diseases, the involvement of TGF beta signaling and ECM is well represented in the DIME-pleiotropy network and known to be implicated in literature[51–53]. Interestingly, we also observed cell-gene network between macrophages and genes like TGF-B1, MMP9 and TIMP1. Evidently, macrophages may exacerbate pulmonary fibrosis by TGF beta production or cause ECM degradation via matrix metalloproteinase (MMP) activities[54]. The DIME-pleiotropy network has captured the intricate network of these key DAGs and the cells (macrophages) accurately.

Patients with rheumatoid arthritis (RA) have an increased risk of cancer due to the severe regimen of disease modifying anti-rheumatic drugs[55]. However, RA patients seem to have lower risk of colon cancer in comparison to the general population[55,56]. To explore the factors responsible for the protective effect against colon cancer in RA patients, we constructed the DIME-pleiotropy network (**Figure 3D**) of RA (MeSH: immune system) and colon carcinoma (MeSH: neoplasm). We found cell-gene networks of $CD4^+$ T cells and many of the myeloid cells. The pathway enrichment of the DAGs in this DIME-pleiotropy network revealed pathways (**Supplementary Figure 5C**) related to TLR signaling, interleukin signaling, neutrophil degranulation, ECM organization, FCER1 and EGFR signaling. Although clear signatures of the protective effect were not observed, we did find the presence of PTGS1 and PTGS2 (also known as COX1 and COX2) in the DIME-pleiotropy network of RA and colon carcinoma. These genes are targets of the non-steroidal anti-inflammatory drugs (NSAIDs) that are frequently taken by RA patients. Evidently, NSAIDs like aspirin have been shown to

confer protective affect against colorectal cancer even in lower doses[57]. Perhaps the missing link in the protective effect of RA in colon cancer lies in the anti-inflammatory role of the NSAIDs taken by the RA patients that target the pro-inflammatory mediators[58] such as PTGS1 and PTGS2 in both diseases, thus conferring protection.

These DIME-pleiotropy analyses highlighted the common cell-gene relationships between the different diseases (**Figure 3B-D**). Thus, making such pleiotropic relationships less "ubiquitous"[59] but rather based on the similar cell-gene mechanism implicated between the two diseases. We used this to further capture the plausible drug targets based on the immune system mechanism as we identified with RA and colon cancer.

Immunome mediated drug repurposing

Using our immunome data, the DisGeNet and the NMF, we constructed the DIME network and the DIME-pleiotropy network (as described before and as depicted in **Figure 4A**, part 1 and 2). We extended this approach to construct drug-gene network to identify drug targets based on the DIME network (**Figure 4A**, part 3), which are referred to as the DIME-drug network. We then used the pleiotropy method used before to then identify common drug targets between the diseases (**Figure 4A**, part 4) based on the DIME-drug networks, which forms the basis of the immunome mediated drug repurposing, (see methods). **Figure 4B** shows an example of the DIME-drug network of Crohn's disease that represents the connections between the top DAGs of Crohn's disease as the potential drug targets and their associated drugs. The DIME-drug network of Crohn's disease identified some known drugs like the corticosteroids (such as prednisone, methylprednisolone, hydrocortisone and budesonide, targets of NR3C1) and the aminosalicylates (such as sulfasalazine and mesalamine, targets of PTGS2 and ALOX5) that are current line of drugs used in treatment of Crohn's disease[60]. In addition to the known drugs and drug targets, the DIME-drug network of Crohn's disease also revealed some novel and potentially interesting drugs and drug targets, such as lifitegrast that target the integrins, ITGB2 and ITGAL. This is particularly interesting because integrin based therapies (such as natalizumab and vedolizumab) are already in use for Crohn's disease[61]. Exploring other integrin based therapies for Crohn's disease may be beneficial since both ITGAL and ITGB2 show in the top DAGs and are also implicated in Crohn's disease[62,63].

As we discovered common cell-gene mechanisms in the DIME-pleiotropy network in the previous analysis (**Figure 3**), we extended this approach to discover plausible drugs and drug targets between the disease pair comparison to find new candidates for drug repurposing, we refer to these analysis as the pleiotropy-drug network analysis (**Figure 4A**, part 4). In addition to the drug targets revealed from the DIME-drug network of Crohn's disease, the pleiotropy drug network of Crohn's disease and psoriasis (**Figure 4C**), comprised some known drug targets such as IL6R for psoriasis and IL1B for Crohn's disease. The associated drug of IL6R, tocilizumab is known to be used in the treatment for psoriasis[64]. Interestingly, anti IL6 therapy has been shown to have promising clinical response for Crohn's disease as well[65]. Similarly, anti-IL1 therapies have also been explored for psoriasis and have shown beneficial clinical outcome[64]. Thus, DIME-network identifies established and novel potential targets for drug repurposing.

Studying the pleiotropy-drug network of fibrotic disease systemic scleroderma and pulmonary fibrosis (**Figure 4D**), we found the DNMT1 (a DNA methyl transferase enzyme) targeting drugs decitabine and azacitidine. Epigenetic modulation as a therapy has been explored as a treatment option in SSc[66]. Similarly, in the pleiotropy drug network of SSc and myocardial infarction (MeSH: Cardiovascular), **Figure 4E**, we found canakinumab an IL1B targeting drug. Anti-IL drugs have also been tested on patients with myocardial infarction and were shown to have fewer cardiovascular events than placebo[67].The pleiotropy drug network of rheumatoid arthritis and colon carcinoma was found to be larger (**Figure 4F**), due to the larger overlap between the two diseases, see **Figure 3D**. The network shows possibilities of using anti-inflammatory drugs for treatment or prevention of colon carcinoma as discussed before. With drugs that target genes such as IL2RG, TYK2, PTGS1, PTGS2, etc. that are widely used to treat anti-inflammatory or immune mediated diseases, these can be used as preventive care drugs or for use along with chemotherapy[68].

We also looked at the known drug targets that are present across the DIME networks of the 613 diseases analyzed by us. We found the top 5 drug targets to be BCL2 (**Figure 4B, C, E,** and **F**), PTGS2 (**Figure 4B, 4E-4F**), PIK3CD (**Figure 4D, 4E**), CXCR4 (**Figure 4B-4F**) and IL1B (**Figure 4B, C, E,** and **F**) that were implicated in the DIME networks of more than 200 out of the 613 diseases analyzed by us. The CXCR4 targetable by plerixafor is present in all of the pleiotropy-drug networks shown here (**Figure 4B-4F**). Plerixafor is a drug intended for use in cancer after stem cell transplantation to initiate migration of stem cells in the bloodstream[69]. This drug is now in clinical trials (NCT01413100) to be evaluated for use

after autologous transplant in patients with scleroderma[70]. Such trials may potentially be extended to other immune diseases like psoriasis, Crohn's disease, rheumatoid arthritis, and potentially for a variety of other phenotypically distinct rheumatic diseases that are driven by CXCR4 mediated dysregulation of immune system. Many such potential drug repurposing targets (enlisted in **Supplementary Table 2**) can be similarly evaluated in future studies and trials.

Discussion

Despite decades of experimental data on the understanding of the molecular mechanism of diseases, we know little about the perturbations in the niche cell compartments that are specific to the disease. To address this gap several efforts at the tissue[5–7] and immune cell[16,17] level have been performed to identify these disease specific compartments. However, previous studies have concentrated at whole tissues, not distinguishing different immune cell subsets or, in contrast, focused on a few immune cell subsets thereby likely missing the complex molecular network underpinning immune mediated diseases. Additionally, it is important to understand which of the gene subsets contribute to a mechanism within the different cell populations. A disease may have perturbations occurring at a single cell type or at multiple cell types with completely different or similar genes being involved. To truly understand a disease, it is essential to capture these cell-gene networks as holistically as possible.

To address the above-mentioned gap, we used the systems immunology approach of dissecting diseases using the *immunome* comprising of 40 immune cells, the vast literature on the available disease network and computational methods to compute and construct DIME networks. The unique integration of these parts resulted in the novel mechanisms being captured by our method. In this report, we highlight some of the known mechanisms we capture from the DIME networks. For example, the role of NETosis and granulocytes in many of the immune diseases, the role of TGF beta signaling in fibrotic diseases, the role of BCR signaling and PI3K/AKT pathways in cancer, etc. We further outlined methods to capture pleiotropy between diseases using the combination of cell-gene commonalities between the DIME networks of the diseases, to ensure the robust capture of common mechanism between diseases. We further extend this approach to identify immune mechanism based drug targets that provides additional support and rationale for drug

repurposing. For example, we found from the DIME drug network that several specific interleukins are involved in Crohn's disease, psoriasis and myocardial infarction. Given the success of anti-IL therapies in many of the immune related diseases, our results indicate that the anti-IL therapies might also be a promising option for these immune related diseases. Likewise, our results indicate that using NSAIDs might be a promising option for cancer prevention and treatment. However, additional functional experiments and extensive clinical trials have to be done to support this approach.

The DACs and DAGs of the DIME network make it more robust to pinpoint which mechanism and in which cell type. And this makes it a useful tool to dissect diseases. Hence we built a shiny app to identify and construct DIME networks for all the diseases in the DisGeNet, the GWAS network and also for user defined set of genes. The tool also identifies potential drug targets based on the DIME networks. A caveat is that we focused on 40 different immune cells in this study, certain diseases may have manifestations that may or may not be perturbed in the immune system or the progenitor cells that we have looked at. Such diseases should be analyzed with additional data to identify the DACs and the underlying mechanism. We have built the tool to accommodate and plugin such futuristic or existing data consisting of gene expression information (coming from single cell or bulk) on the different cell types apart from those found in the immune system. We believe that this tool will aid scientist to increase the understanding of disease pathology and facilitate drug development by better determining drug targets, thereby mitigating risk of failure in late clinical development.

REFERENCES

1. Sun, Y. V & Hu, Y.-J. Integrative Analysis of Multi-omics Data for Discovery and Functional Studies of Complex Human Diseases. *Adv. Genet.* **93**, 147–90 (2016).

2. Li, Y. & Agarwal, P. A Pathway-Based View of Human Diseases and Disease Relationships. *PLoS One* **4**, e4346 (2009).

3. Camp, J. G., Platt, R. & Treutlein, B. Mapping human cell phenotypes to genotypes with single-cell genomics. *Science (80-.).* **365**, 1401–1405 (2019).

4. Rozenblatt-Rosen, O., Stubbington, M. J. T., Regev, A. & Teichmann, S. A. The Human Cell Atlas: from vision to reality. *Nature* **550**, 451–453 (2017).

5. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).

6. Ardlie, K. G. *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science (80-.).* **348**, 648–660 (2015).

7. GTEx Consortium *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).

8. Tirosh, I. & Suvà, M. L. Deciphering Human Tumor Biology by Single-Cell Expression Profiling. *Annu. Rev. Cancer Biol.* **3**, 151–166 (2019).

9. Keren-Shaul, H. *et al.* A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell* **169**, 1276-1290.e17 (2017).

10. Kuo, D. *et al.* HBEGF+ macrophages in rheumatoid arthritis induce fibroblast invasiveness. *Sci. Transl. Med.* **11**, eaau8587 (2019).

11. Knight, J. C. Genomic modulators of the immune response. *Trends Genet.* **29**, 74–83 (2013).

12. Raj, T. *et al.* Polarization of the Effects of Autoimmune and Neurodegenerative Risk Alleles in Leukocytes. *Science (80-.).* **344**, 519–523 (2014).

13. Quach, H. *et al.* Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. *Cell* **167**, 643-656.e17 (2016).

14. Lee, M. N. *et al.* Common Genetic Variants Modulate Pathogen-Sensing Responses in Human Dendritic Cells. *Science (80-.).* **343**, 1246980–1246980 (2014).

15. Fairfax, B. P. *et al.* Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression. *Science (80-.).* **343**, 1246949–1246949 (2014).

16. Chen, L. *et al.* Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell* **167**, 1398-1414.e24 (2016).

17. Schmiedel, B. J. *et al.* Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell* **175**, 1701-1715.e16 (2018).

18. SRA-Tools - NCBI. http://ncbi.github.io/sra-tools/.

19. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

20. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

21. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

22. Risso, D., Ngai, J., Speed, T. P. & Dudoit, S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.* **32**, 896–902 (2014).

23. Piñero, J. *et al.* DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* **45**, D833–D839 (2017).

24. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).

25. Gaujoux, R. & Seoighe, C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* **11**, 367 (2010).

26. Brunet, J.-P., Tamayo, P., Golub, T. R. & Mesirov, J. P. Metagenes and molecular pattern discovery using matrix factorization. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 4164–9 (2004).

27. Cotto, K. C. *et al.* DGIdb 3.0: a redesign and expansion of the drug–gene interaction database. *Nucleic Acids Res.* **46**, D1068–D1073 (2018).

28. Verghese, P. B., Castellano, J. M. & Holtzman, D. M. Roles of Apolipoprotein E in Alzheimer's Disease and Other Neurological Disorders. *Lancet Neurol.* **10**, 241 (2011).

29. Costarelli, L., Malavolta, M., Giacconi, R. & Provinciali, M. Dysfunctional macrophages in Alzheimer Disease: another piece of the "macroph-aging" puzzle? *Aging (Albany. NY).* **9**, 1865–1866 (2017).

30. Strittmatter, W. J. *et al.* Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 1977–81 (1993).

31. Bell, R. D. *et al.* Transport Pathways for Clearance of Human Alzheimer's Amyloid β-Peptide and Apolipoproteins E and J in the Mouse Central Nervous System. *J. Cereb. Blood Flow Metab.* **27**, 909–918 (2007).

32. Sanchez, V. E., Nichols, C., Kim, H. N., Gang, E. J. & Kim, Y.-M. Targeting PI3K Signaling in Acute Lymphoblastic Leukemia. *Int. J. Mol. Sci.* **20**, (2019).

33. Biernacka, A., Dobaczewski, M. & Frangogiannis, N. G. TGF-β signaling in fibrosis. *Growth Factors* **29**, 196–202 (2011).

34. Yipp, B. G. & Kubes, P. NETosis: how vital is it? *Blood* **122**, 2784–94 (2013).

35. Neutrophil Extracellular Traps in Systemic Sclerosis - Full Text View - ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/NCT03374618.

36. M. Artlett, C. The Role of the NLRP3 Inflammasome in Fibrosis. *Open Rheumatol. J.* **6**, 80–86 (2012).

37. Yu, Y. & Su, K. Neutrophil Extracellular Traps and Systemic Lupus Erythematosus. *J. Clin. Cell. Immunol.* **4**, (2013).

38. Lande, R. *et al.* Neutrophils Activate Plasmacytoid Dendritic Cells by Releasing Self-DNA–Peptide Complexes in Systemic Lupus Erythematosus. *Sci. Transl. Med.* **3**, 73ra19-73ra19 (2011).

39. Garcia-Romo, G. S. *et al.* Netting Neutrophils Are Major Inducers of Type I IFN Production in Pediatric Systemic Lupus Erythematosus. *Sci. Transl. Med.* **3**, 73ra20-73ra20 (2011).

40. Villanueva, E. *et al.* Netting Neutrophils Induce Endothelial Damage, Infiltrate Tissues, and Expose Immunostimulatory Molecules in Systemic Lupus Erythematosus. *J. Immunol.* **187**, 538–552 (2011).

41. Leffler, J. *et al.* Neutrophil Extracellular Traps That Are Not Degraded in Systemic Lupus Erythematosus Activate Complement Exacerbating the Disease. *J. Immunol.*

188, 3522–3531 (2012).

42. Yang, W.-C. *et al*. Interleukin-4 Supports the Suppressive Immune Responses Elicited by Regulatory T Cells. *Front. Immunol.* **8**, 1508 (2017).

43. Wang, R. *et al.* Association of interleukin 13 gene polymorphisms and plasma IL 13 level with risk of systemic lupus erythematosus. *Cytokine* **104**, 92–97 (2018).

44. Gao, W. *et al*. Notch signalling pathways mediate synovial angiogenesis in response to vascular endothelial growth factor and angiopoietin 2. *Ann. Rheum. Dis*. **72**, 1080–8 (2013).

45. Moulton, V. R. & Tsokos, G. C. Abnormalities of T cell signaling in systemic lupus erythematosus. *Arthritis Res. Ther.* **13**, 207 (2011).

46. Crow, M. K. & Rönnblom, L. Report of the inaugural Interferon Research Summit: interferon in inflammatory diseases. *Lupus Sci. Med.* **5**, e000276 (2018).

47. Wagner, G. P. & Zhang, J. The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. *Nat. Rev. Genet.* **12**, 204–213 (2011).

48. Yang, C., Li, C., Wang, Q., Chung, D. & Zhao, H. Implications of pleiotropy: challenges and opportunities for mining Big Data in biomedicine. *Front. Genet.* **6**, 229 (2015).

49. Casciano, F., Pigatto, P. D., Secchiero, P., Gambari, R. & Reali, E. T Cell Hierarchy in the Pathogenesis of Psoriasis and Associated Cardiovascular Comorbidities. *Front. Immunol.* **9**, 1390 (2018).

50. Li, N. & Shi, R.-H. Updated review on immune factors in pathogenesis of Crohn's disease. *World J. Gastroenterol.* **24**, 15–22 (2018).

51. Todd, N. W., Luzina, I. G. & Atamas, S. P. Molecular and cellular mechanisms of pulmonary fibrosis. *Fibrogenesis Tissue Repair* **5**, 11 (2012).

52. Wynn, T. A. Integrating mechanisms of pulmonary fibrosis. *J. Exp. Med.* **208**, 1339–50 (2011).

53. Pattanaik, D., Brown, M., Postlethwaite, B. C. & Postlethwaite, A. E. Pathogenesis of Systemic Sclerosis. *Front. Immunol.* **6**, 272 (2015).

54. Kolahian, S., Fernandez, I. E., Eickelberg, O. & Hartl, D. Immune Mechanisms in Pulmonary Fibrosis. *Am. J. Respir. Cell Mol. Biol.* **55**, 309–322 (2016).

55. Wilton, K. M. & Matteson, E. L. Malignancy Incidence, Management, and Prevention in Patients with Rheumatoid Arthritis. *Rheumatol. Ther.* **4**, 333–347 (2017).

56. Simon, T. A., Thompson, A., Gandhi, K. K., Hochberg, M. C. & Suissa, S. Incidence of malignancy in adult patients with rheumatoid arthritis: a meta-analysis. *Arthritis Res. Ther.* **17**, 212 (2015).

57. Din, F. V. N. *et al*. Effect of aspirin and NSAIDs on risk and survival from colorectal cancer. *Gut* **59**, 1670–1679 (2010).

58. Dubois, R. N. Role of inflammation and inflammatory mediators in colorectal cancer. *Trans. Am. Clin. Climatol. Assoc.* **125**, 358–72; discussion 372-3 (2014).

59. Chesmore, K., Bartlett, J. & Williams, S. M. The ubiquity of pleiotropy in human

disease. *Hum. Genet.* **137**, 39–44 (2018).

60. Akobeng, A. K. Crohn's disease: current treatment options. *Arch. Dis. Child.* **93**, 787–92 (2008).

61. Ley, K., Rivera-Nieves, J., Sandborn, W. J. & Shattil, S. Integrin-based therapeutics: biological basis, clinical use and new drugs. *Nat. Rev. Drug Discov.* **15**, 173–83 (2016).

62. Ellinghaus, D. *et al.* Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat. Genet.* **48**, 510–518 (2016).

63. Bernstein, C. N., Sargent, M. & Rector, E. Alteration in expression of beta 2 integrins on lamina propria lymphocytes in ulcerative colitis and Crohn's disease. *Clin. Immunol.* **104**, 67–72 (2002).

64. Tsai, Y.-C. & Tsai, T.-F. Anti-interleukin and interleukin therapies for psoriasis: current evidence and clinical usefulness. *Ther. Adv. Musculoskelet. Dis.* **9**, 277–294 (2017).

65. Ito, H. *et al.* A pilot randomized trial of a human anti-interleukin-6 receptor monoclonal antibody in active Crohn's disease. *Gastroenterology* **126**, 989–996 (2004).

66. O'Reilly, S. Epigenetic modulation as a therapy in systemic sclerosis. *Rheumatology* **58**, 191–196 (2019).

67. Ridker, P. M. *et al.* Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. *N. Engl. J. Med.* **377**, 1119–1131 (2017).

68. Terzić, J., Grivennikov, S., Karin, E. & Karin, M. Inflammation and Colon Cancer. *Gastroenterology* **138**, 2101-2114.e5 (2010).

69. Hübel, K. *et al.* Plerixafor with and without chemotherapy in poor mobilizers: results from the German compassionate use program. *Bone Marrow Transplant.* **46**, 1045–1052 (2011).

70. Scleroderma Treatment With Autologous Transplant (STAT) Study - Full Text View - ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/study/NCT01413100.

FIGURE LEGENDS

Figure 1: A Brief overview of methods, highlighting the integration of the 3 layers. **B** MeSH-MeSH network highlighting different MeSH (disease category) terms as nodes and the edges represent the genes common between the MeSH terms. Node names in network represent first two letters of the complete MeSH term as shown in the legend on right. **C-H** Gene prevalence over different MeSH terms. The degree distribution of genes across **I** MeSH terms and **J** diseases shown as histogram. **K** heatmap of gene expression across immunome of high degree genes in the DisGeNet. **L** Gene expression of APOE in the immunome. In **B, E-J**, size of node represents number of diseases in MeSH term.

Figure 2: DIME network of **A** lymphoid leukemia, **B** systemic scleroderma, **C** pulmonary fibrosis and **D** systemic lupus erythematosus. Green nodes represent genes, blue represents cell types and red represents diseases. Size of nodes is proportional to DAG score in genes and DAC score in cell types.

Figure 3: **A** Top pleiotropy network of the disease subset of the DisGeNet. Nodes are diseases that have a minimum degree of 2 nodes. Edges between diseases exist if the JI ≥ 0.1 and FET p-value ≤ 0.01 of the common DIME network between the diseases. Pleiotropy analysis between **B** Crohn's disease and psoriasis, **C** systemic scleroderma and pulmonary fibrosis, and **D** rheumatoid arthritis and colon carcinoma. **B, C, D** Venn diagrams represent overlap of cell-gene connections between the disease comparisons. JI and FET p-value are calculated for overlap in the genes in the DIME network of the diseases.

Figure 4: **A** Schema showing the different analysis performed in this study from building the DIME network (part1) to the pleiotropy (part 2) to the DIME-drug network (part3) and to the pleiotropy-drug network (part 4). **B** DIME-drug network of Crohn's disease. Pleiotropy-drug network for **C** Crohn's disease and psoriasis, **D** systemic scleroderma and pulmonary fibrosis, **E** systemic scleroderma and myocardial infarction, and **F** rheumatoid arthritis and colon carcinoma.

Supplementary Information

Supplementary Figure 1: Shows the degree of the top 10 high degree genes for different MeSH term categories and for all diseases.

Supplementary Figure 2: Heatmap of DAC scores for top weighted feature of 613 diseases from the DisGeNet.

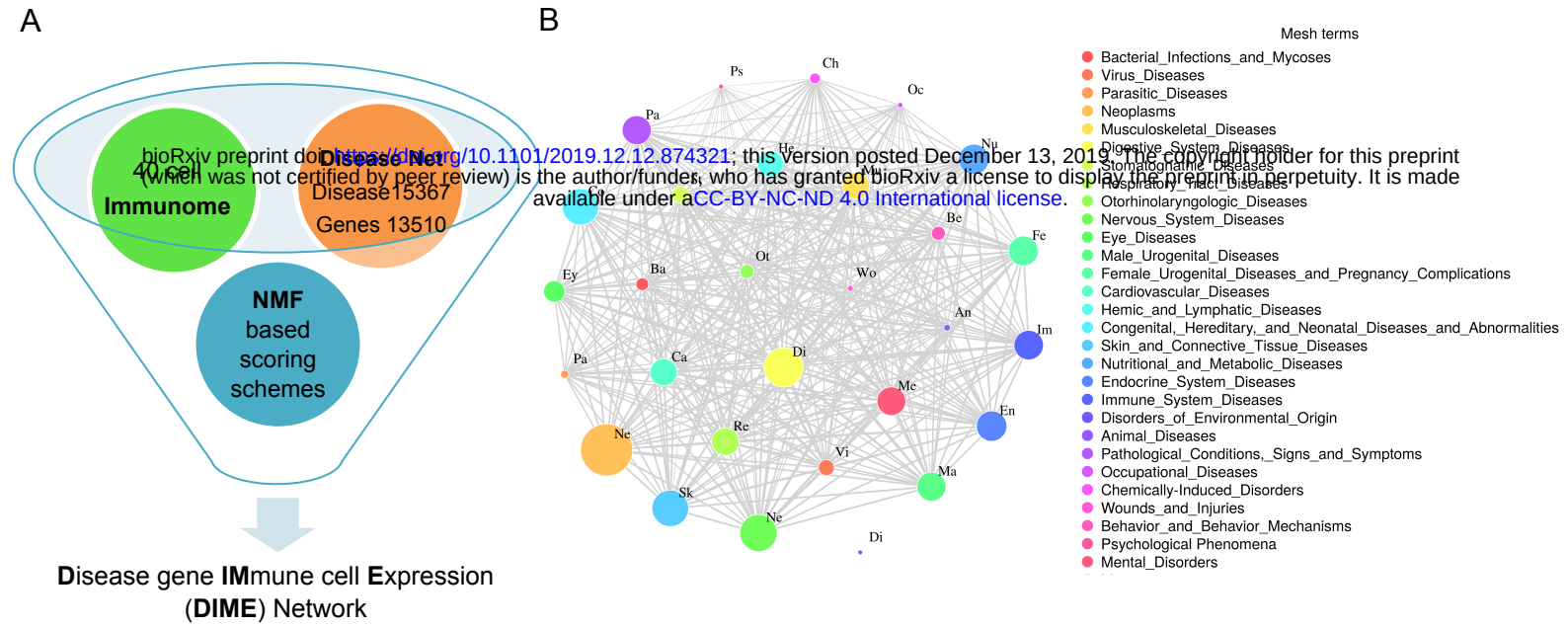
Supplementary Figure 3: Reactome pathway enrichment analysis of top DAGs in the different clusters of the DIME network of **A** systemic scleroderma and **B** pulmonary fibrosis.

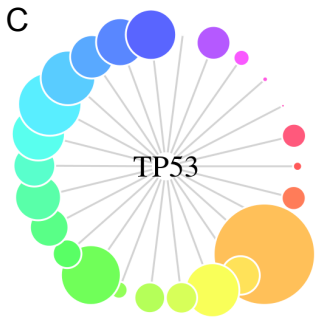
Supplementary Figure 4: Reactome pathway enrichment analysis of top DAGs in the different clusters of the DIME network of **A** systemic lupus erythematosus and **B** lymphoid leukemia.

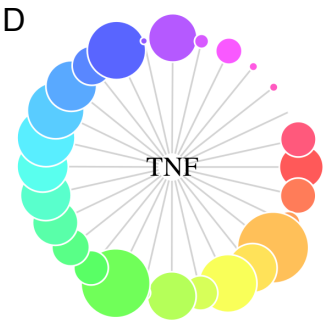
Supplementary Figure 5: Reactome pathway enrichment analysis of pleiotropy DIME network of **A** Crohn's disease and psoriasis, **B** systemic scleroderma and pulmonary fibrosis, and **C** rheumatoid arthritis and colon carcinoma.

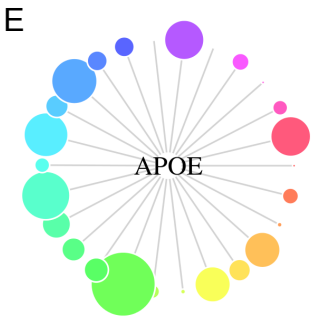
Supplementary Table 1: Represents all GEO datasets and samples used to construct the *immunome*.

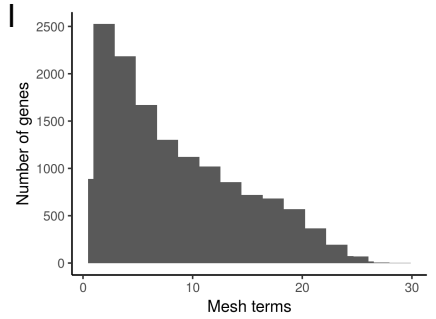
Supplementary Table 2: List of genes as potential drug targets identified from the pleiotropy-drug network analysis. The degree represents the number of diseases in which the gene was found to be in the DIME-drug network. A total of 613 diseases were analyzed.

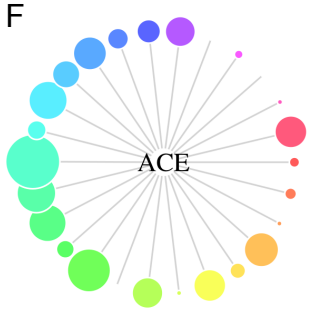


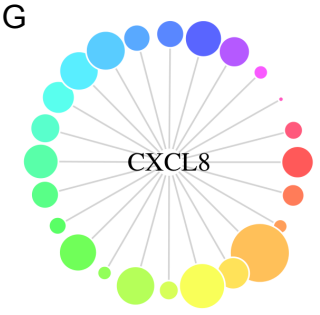


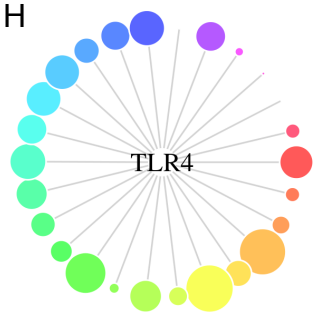


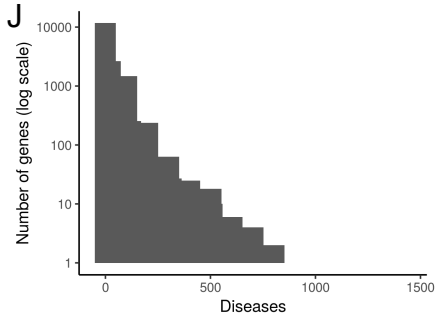


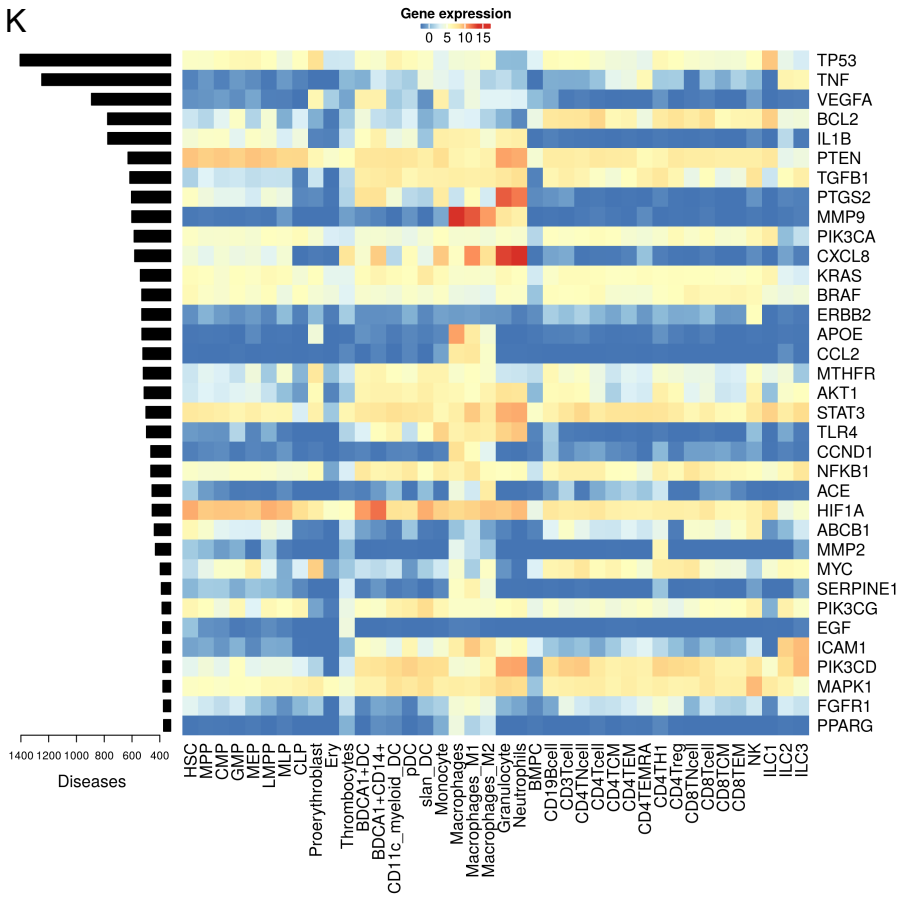


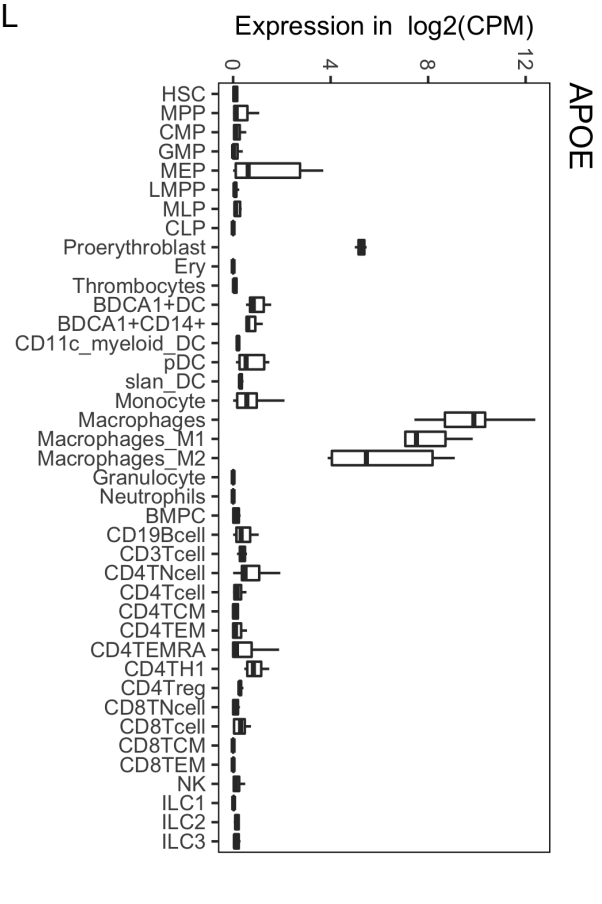




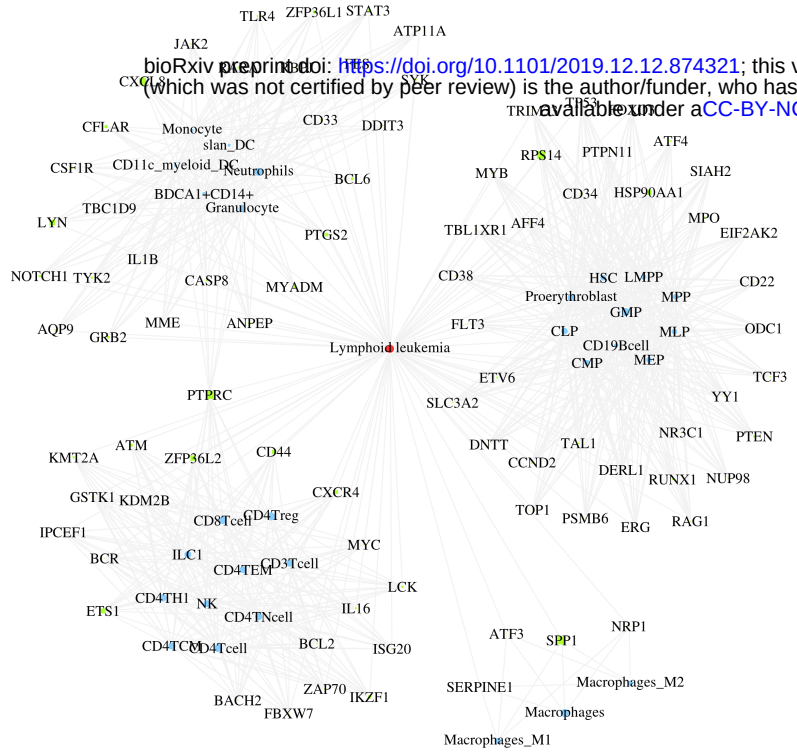




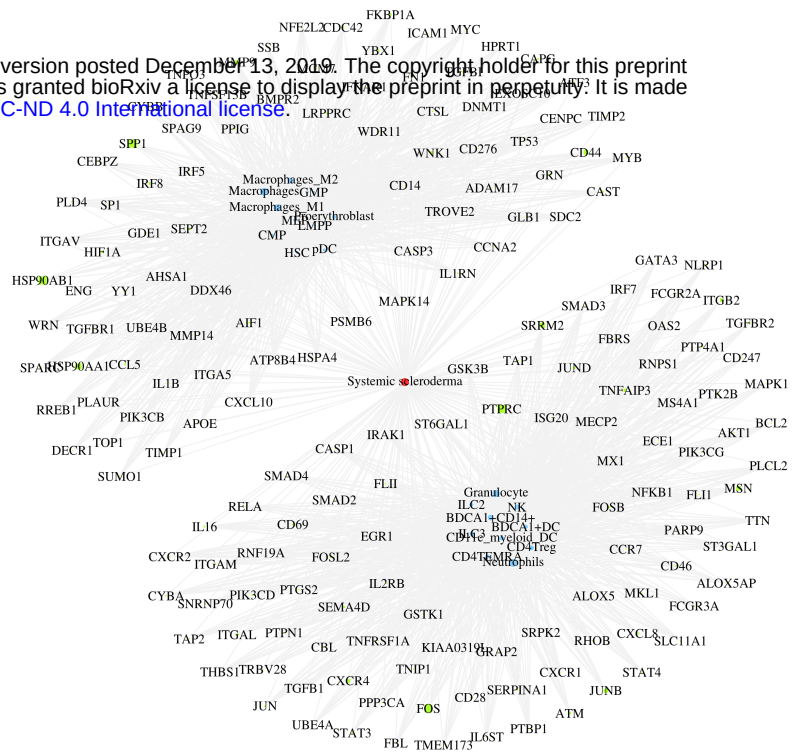




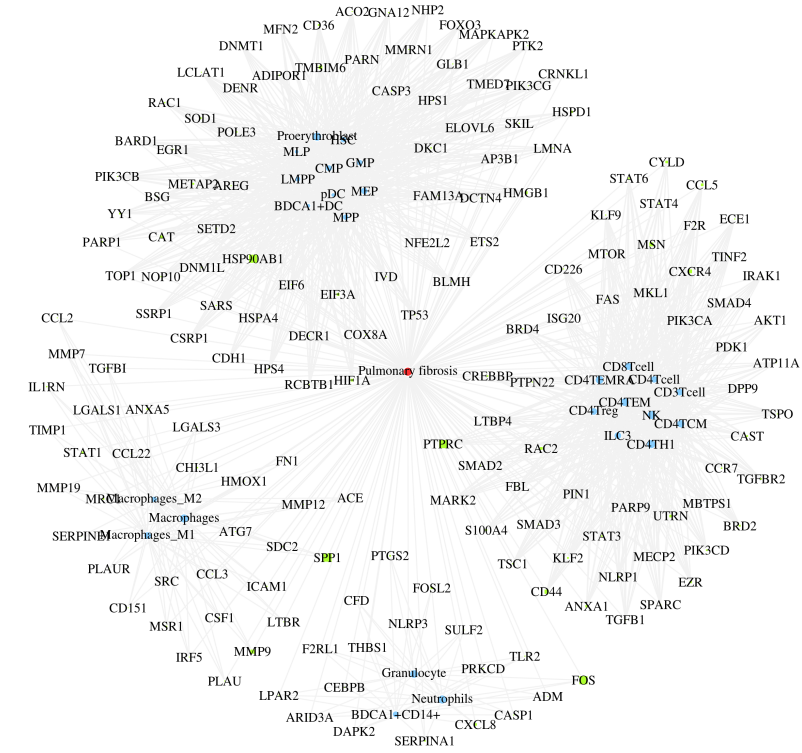
A



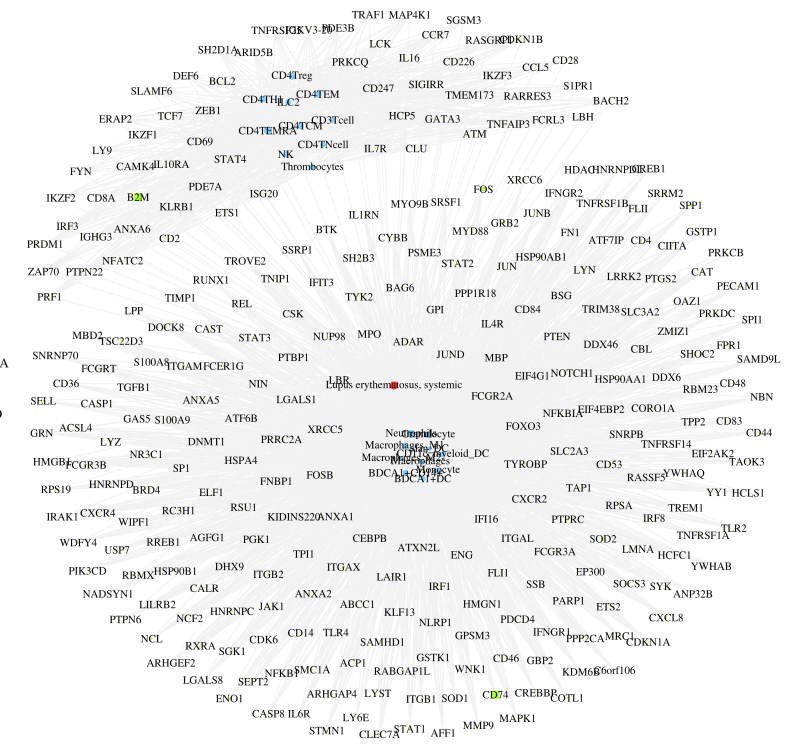
B



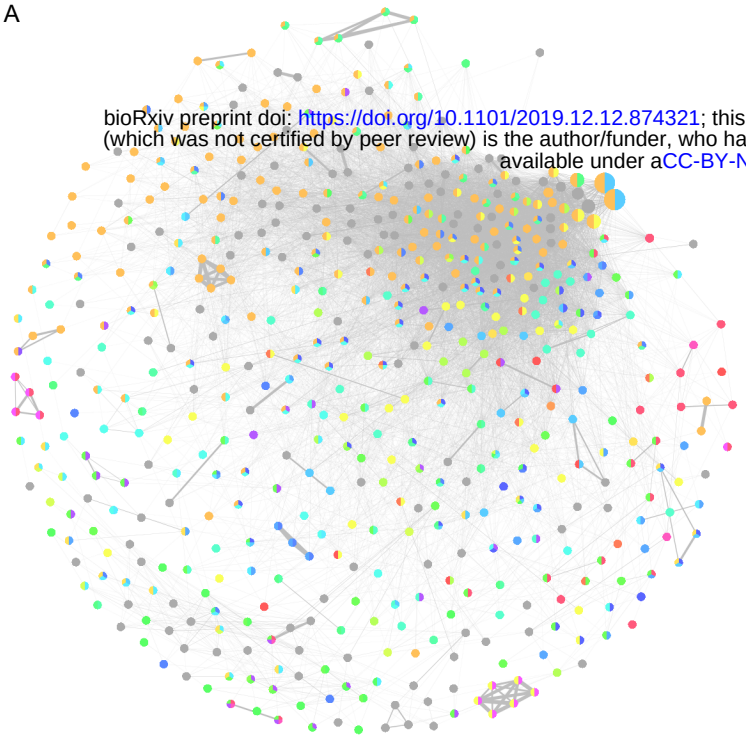
C



D

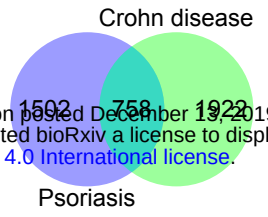


A

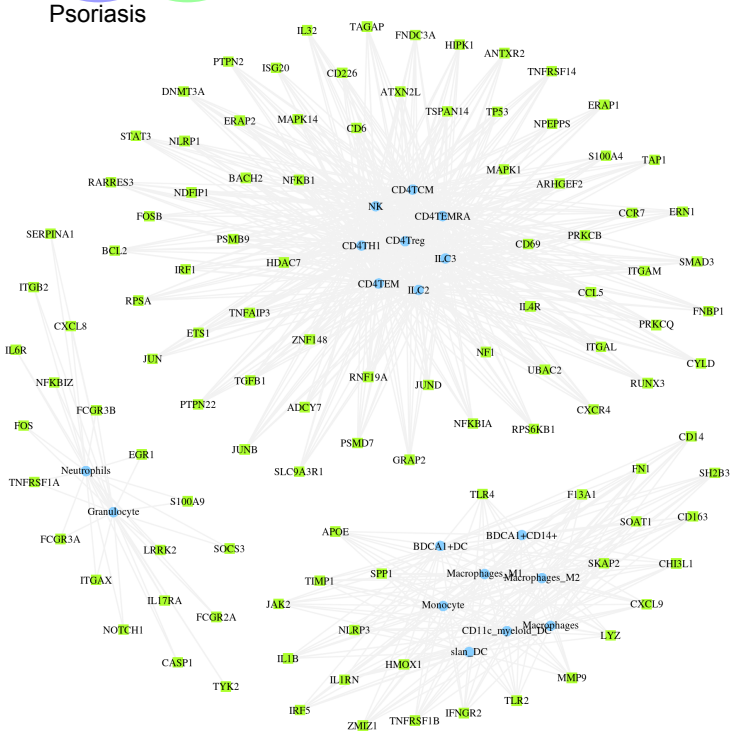


Top Pleiotropy Network
JI ≥ 0.1, FET ≤ 0.01, Degree ≥ 2

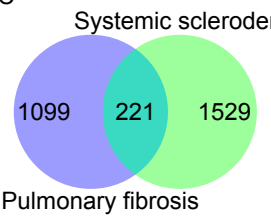
B



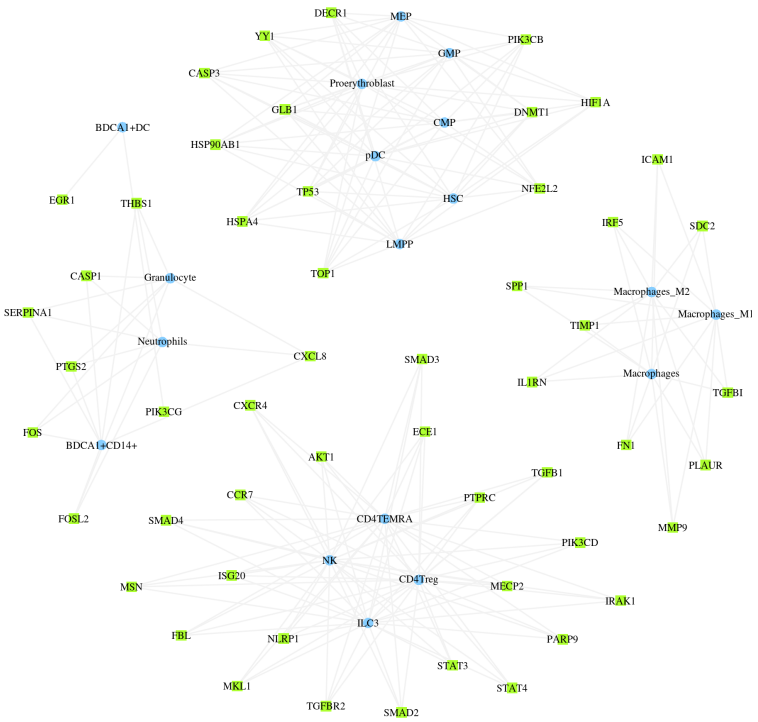
JI = 0.29
FET = 4.19e-120



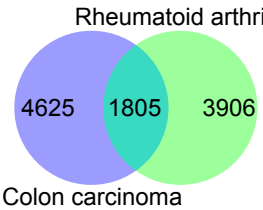
C



JI = 0.19
FET = 1.53e-59



D



JI = 0.26
FET = 2.35e-156