

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

3 Abhinandan Devaprasad¹, Timothy RDJ Radstake¹, Aridaman Pandit^{1*}

4 ¹Center for Translation Immunology, University Medical Center Utrecht, Utrecht, The
5 Netherlands

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

7

8 **Abstract**

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

22 **Introduction**

23 The genetic and epigenetic heterogeneity has been known to play a major role in the
24 development and progression of complex diseases. The past two decades have seen a major
25 surge in studies that characterize genes and loci associated with disease. The use of high-
26 throughput omics technology and functional screenings have boosted our knowledge about
27 genetic, epigenetic and metabolic factors underlying complex diseases¹. As a result of these
28 genetic and epigenetic screenings, we now know that the majority of complex diseases and
29 genes/loci have a many-to-many relationship meaning that a complex disease is linked to
30 many different genes and a gene/loci is associated with many different genes².

31 Large high-throughput screening studies have typically used bulk tissue or whole blood to
32 study disease associated genes (DAGs). However, the expression of each gene is known to
33 vary between tissues and cell types^{3,4}. Thus, bulk tissue- or blood-based studies on DAGs do
34 not consider the role played by different cells and tissues in the disease biology. To improve
35 the understanding and molecular basis of complex diseases, a large number of research
36 groups and consortiums have started to functionally identify disease associated cells (DACs)
37 or tissue types³⁻⁷. The Genotype-Tissue Expression (GTEx) is one such valuable project,
38 which maps gene expression profiles of 54 different human tissue types and the
39 corresponding expression quantitative trait loci (eQTLs)⁵⁻⁷. Furthermore, the growth of single
40 cell technologies have advanced our understanding of DACs and have helped in identifying
41 cell types associated with complex diseases including cancer⁸, Alzheimer's⁹, rheumatoid
42 arthritis¹⁰, among others.

43 The immune system is known to play a key role in the development and progression of
44 immune-mediated as well as non-immune mediated chronic diseases. A large number of
45 association and functional studies have shown that multiple DAGs are expressed in immune
46 cells and perturbing these DAGs can modulate immune cell functions¹¹. However, very few
47 studies have explored the impact of DAGs on specific cell types and even fewer on immune
48 cells, many of which focus on limited number of cell subsets¹²⁻¹⁶. Recently Schmiedel *et al.*
49 studied the effect of genetic variants on the expression of genes in 13 different immune cell
50 types¹⁷. However, this study largely focused on the analysis of genetic variants and their
51 impact on a total of 13 immune cell types: monocytes (classical and non-classical), NK cells,
52 naïve B-cells and nine sub-populations of T-cells.

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

67

68 **Methods**

69 **Transcriptome data - *Immunome***

70 The transcriptome data consists of RNA-sequencing datasets of 40 different immune cell
71 types curated using 316 samples from a total of 27 publicly available datasets (see
72 **Supplementary Table 1** for list of GEO datasets and samples used). The 40 different
73 immune cells cover the entire hematopoietic stem cell differentiation tree comprising of 9
74 progenitors, 19 lymphoid, and 12 myeloid cell types. The samples used here were manually
75 curated considering only the unstimulated (except for macrophages, that were monocyte
76 derived) immune cells that were sorted using Fluorescence-activated cell sorting (FACS) and
77 were isolated from either blood, bone marrow or cord blood from healthy donors. All the
78 selected datasets were downloaded as FASTQ files using the fastq-dump tool from
79 sratoolkit¹⁸. The “—split-files” option was given if the library type was paired end
80 sequencing. FASTQ files were then aligned to reference genome (GRCH.Hg38.79) using
81 STAR aligner¹⁹. The result is a SAM file which was then converted into a sorted BAM file
82 using the samtools program²⁰. These were then used to calculate the count of aligned reads
83 using the HTSeq program²¹ with the mode option “intersection non-empty”. HTSeq was run

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

86 The data was then filtered by removing all genes that had less than 20 read counts in 95
87 percent of the samples using R programming. The filtered data was then lane normalized
88 using the “betweenLaneNormalization” function from the RUVSeq package²². The RUVr
89 method from RUVSeq was used to identify residual factors contributing to the batch effect.
90 The resulting filtered, batch corrected and normalized data had expression for 34,906 genes
91 that was void of any observable batch effect. We calculated counts per million (cpm) for the
92 filtered genes and used cpm as the gene expression measure. We then used the median gene
93 expression for each cell type for the rest of the analysis. This processed, batch corrected,
94 normalized and median representative data of 40 immune cells is referred to as the
95 *immunome*.

96

97 **Disease gene network from DisGeNet**

98 The full disease gene association network from DisGeNet²³ was downloaded from the
99 DisGeNet database (www.disgenet.org/downloads). All HLA associated genes was removed
100 from the network, this was done to ensure that bias towards myeloid cells and B cells are
101 removed, since the HLA genes are largely expressed by these cells. The resulting network
102 was further filtered to include only those genes that were present in the *immunome*. The final
103 network comprised of 15367 diseases and 13510 DAGs.

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

110

111 **Other disease gene networks - EBI GWAS data**

112 In addition to the DisGeNet, we also used a refined GWAS based dataset from the EBI²⁴. The
113 GWAS catalogue of Version 1.0, e89, was downloaded from the EBI website, which

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

120

121 **Mapping disease gene network to *Immunome* data**

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

$$126 \quad \dim(X_D) = \text{length}(DAG_D) \times 40,$$

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

129

130 **Using NMF to cluster X_D into k classes**

131 We used the NMF package²⁵ in R and applied the non-negative matrix factorization method
132 using Brunet's²⁶ algorithm to the expression matrix (X_D) to factor it into two matrices namely
133 W_D and H_D such that.

$$134 \quad X_D \approx W_D H_D,$$

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

$$137 \quad \dim(W_D) = \text{length}(DAG_D) \times k,$$

$$138 \quad \dim(H_D) = k \times 40,$$

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

142 clusters (in each row). We used Brunet *et al.* method to identify the ideal k value using the
 143 cophenetic correlation coefficient method²⁶.

144

145 **Identifying the key DAG_D and DAC_D from W_D and H_D**

146 The NMF algorithm clusters the data into k clusters such that, in each cluster ‘ i ’, where $i \in$
 147 $(1, \dots, k)$, the genes that have high values in w_D^i are constitutively expressed by the cells that
 148 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
 149 each cluster i , we chose the cell-gene pairs that were in the top 25th percentile range of their
 150 corresponding w_D^i and h_D^i values. These cells and their corresponding gene pairs are regarded
 151 as the key DAC_D and DAG_D respectively. The cell-gene pairs were extracted from all the
 152 clusters and were compiled together. The resulting cell-gene pairs of all the clusters form the
 153 edges of the **Disease-gene to IMMune cell Expression network**, hereby referred as the **DIME**
 154 network.

155

156 **Identifying the key cluster**

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

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

$$161 \quad X_D \approx W_D H_D = \begin{bmatrix} | & | & | & | \\ w_D^1 & w_D^2 & \dots & w_D^k \\ | & | & | & | \end{bmatrix} \begin{bmatrix} - & h_D^1 & - \\ - & h_D^2 & - \\ - & \vdots & - \\ - & h_D^k & - \end{bmatrix} = \sum_{i=1}^k w_D^i h_D^i.$$

162

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

$$166 \quad c = \operatorname{argmax}(\|w_D^i h_D^i\|_F); i \in \{1, \dots, k\},$$

167

168 where, the c^{th} cluster represents that cluster which maximally captures/represents the
169 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
170 DAC_D .

$$171 \quad \quad \quad DAG_D \text{ score} = w_D^c,$$

$$172 \quad \quad \quad DAC_D \text{ score} = h_D^c,$$

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

175

176 **Pleiotropic associations**

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

183

184 **Integrating drug-gene network**

185 The drug to gene target database from DGIdb was downloaded²⁷. The data was filtered to
186 contain only the ChEMBL interactions and only those pertaining to the drugs approved by
187 the food and drug administration (FDA) of USA. This FDA approved drug to gene target
188 interaction serves as the drug-gene network in this study. To identify potential drug
189 candidates from the drug-gene network for disease D , we choose its key DAG_D as identified
190 in the previous step and extracted its corresponding target drugs from the drug-gene network.
191 This network between the drugs and the key DAG_D is referred to as the DIME-drug network
192 for disease D . The DIME-drug network represents all the drugs that target the key DAG_D
193 identified in the given DIME network. To identify potential common acting drugs between
194 different diseases, we used their corresponding DIME-pleiotropy network and extracted the
195 drug-gene connections between the cell-gene and the drugs in drug-gene network. The drugs
196 identified using this approach represent those that act on the common mechanisms between
197 the two diseases and can be potential candidates for drug repurposing.

198

199 **Rationale for restricting analysis to only diseases with ≥ 100 DAGs**

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

211

212 **DIME shinyapp**

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

218

219 **Results**

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

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

232

233 **Common DAGs of phenotypically distinct diseases**

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

245 We further studied the DAG patterns across MeSH terms and found that TP53 was
246 preferentially associated with diseases categorized under the neoplasm MeSH term (**Figure**
247 **1C**). Interestingly, TP53 was also associated with numerous diseases from various other
248 MeSH terms such as immune system diseases, nervous system diseases, and skin and
249 connective tissue diseases. Similarly, TNF was associated with numerous diseases from
250 various MeSH terms including immune system diseases (**Figure 1D**). APOE was largely
251 associated with nervous system diseases, and ACE was largely associated with cardiovascular
252 diseases (**Figure 1E and F**). TLR4 and CXCL8 were prevalently associated to several MeSH
253 terms (**Figure 1G and H**). The above-mentioned genes (**Figure 1C-1F**) are those that are
254 either the top represented (degree) genes across all diseases or for disease within specific
255 MeSH terms. More examples of DAGs associated with specific MeSH terms and their degree
256 can be found in the **Supplementary Figure 1**. Majority of the DAGs mapped to two or more
257 MeSH terms (**Figure 1I**) or diseases (**Figure 1J**) also hinting towards shared disease
258 mechanisms between phenotypically distinct diseases. These genes that are associated to
259 many diseases/MeSH terms (i.e. having high degree) were regarded as the hub genes of the

260 disease-gene network. We ranked the DAGs based on their degree and found TP53, TNF,
261 VEGFA, BCL2, IL1B as the top 5 DAGs each being associated with >750 diseases (**Figure**
262 **1K** and **Supplementary Figure 1**).

263

264 **Relevance of the immune system**

265 We further evaluated the expression of the hub genes identified in the previous step, in the
266 *immunome* data. The rationale was to assess if the hub genes representing the core of the
267 disease-gene network were important to the immune system. Indeed, we found that many of
268 the hub genes of the disease-gene network were expressed constitutively by either all immune
269 cells or by some specific immune cells, as seen in **Figure 1K**. For example, the nervous
270 system associated gene, APOE, important to many of neurological diseases such as
271 Alzheimer's and Parkinson²⁸ were found to be expressed specifically by macrophages, as
272 seen in **Figure 1L**. Studies have shown that genetic polymorphisms in APOE protein leads to
273 defective clearance of the A β plaques by macrophages²⁹⁻³¹. Thus, conferring macrophages as
274 one of the key players of the disease. Such links between DAGs, to observing altered
275 function in a cell type, questions the need to study DAGs by integrating cell-specific
276 expression information.

277

278 **Identifying disease associated cell types**

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

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

293 We constructed the DIME network for different diseases individually, the DIME network
294 consists of the top DACs and their respective DAGs, (see methods). The DIME network
295 represents the key cell-gene mechanisms that can be drawn from the given set of DAGs. As
296 proof of concept, we present here the DIME network of lymphoid leukemia (**Figure 2A**), a
297 group of blood cancer that typically affects the lymphocytes. The DIME network revealed 4
298 DAC clusters for lymphoid leukemia. The key DAC cluster comprised of lymphoid
299 progenitor cells as well as myeloid progenitor cells. The key DAGs contributing to this
300 cluster included genes associated with hematopoiesis such as RPS14, HSP90AA1, MPO,
301 ETV6, ATF4 and TAL1. The other key DAC cluster comprised of all the subsets of T cells,
302 primarily the CD4⁺ T cells. The key DAGs contributing to this cluster included ETS1,
303 CXCR4, IKZF1, ATM, LCK, and KMT2A. The pathway enrichment (**Supplementary**
304 **Figure 4B**) of these genes revealed pathways associated to TCR and BCR signaling and
305 PI3K/AKT signaling. Interestingly, these pathways have been shown to be important for
306 survival of cancer cells and are targets for anti-cancer drugs in acute and chronic lymphoid
307 leukemia³². Thus the DIME network of lymphoid leukemia revealed the key DAGs and
308 DACs implicated in the disease.

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

324 organization and neutrophil degranulation pathways (**Supplementary Figure 3A**). Of which,
325 evidently, TGF beta is known to play an essential role in the pathogenesis of fibrosis³³.
326 Likewise, the neutrophil degranulation pathway or otherwise known as NETosis, is the
327 mechanism by which neutrophils exhibit defensive mechanisms to trap and kill foreign
328 bodies³⁴. And this pathway has been found to be implicated in other autoimmune diseases,
329 and is now being currently investigated in clinical trials³⁵. Interestingly, we found neutrophils
330 and its NETosis associated genes to be in the top cluster of the DIME network for systemic
331 scleroderma.

332 We then constructed the DIME network of another fibrotic disease, pulmonary fibrosis. The
333 DIME network revealed 4 DAC clusters for pulmonary fibrosis (**Figure 2C**). Similar to
334 systemic scleroderma, we found a DAC cluster comprising of T-cells and NK cells enriched
335 in DAGs associated with TGF beta signaling (**Supplementary Figure 3B**). Interestingly, we
336 also found that for pulmonary fibrosis neutrophils-granulocytes, macrophages,
337 BDCA1⁺CD14⁺ cells, pDCs and myeloid progenitors were among the top DACs similar to
338 that of systemic scleroderma, (**Supplementary Figure 3B**). We also found an enrichment of
339 NLRP3 pathway genes, which are known to play a role in the collagen deposition
340 mechanisms commonly dysregulated in fibrosis³⁶.

341 We then studied the DIME network of SLE, an autoimmune disease. The DIME network
342 revealed 2 DAC clusters for SLE (**Figure 2D**). The key cluster comprised of myeloid cells
343 like neutrophils-granulocytes, macrophage M1, BDCA1⁺CD14⁺ and monocytes as the DACs.
344 The key DAGs contributing to this cluster included CD74, FOS, LYZ, SOD2 and
345 HSP90AB1. The other cluster comprised of CD4⁺ TEMRA, CD4⁺ TEM, CD4⁺ TCM and
346 CD4⁺ TH1 as their DACs. The key DAGs contributing to this cluster included B2M, IGHG3,
347 IL7R, ETS1, RPS19, and TNFAIP3. The pathway enrichment for the SLE DIME network
348 includes pathways that are heavily described in the literature for SLE, such as the neutrophil
349 degranulation pathways or NETosis³⁷⁻⁴¹, interleukin 4 and interleukin 13 signaling^{42,43}, TLR
350 signaling pathway⁴⁴, translocation of ZAP70 to immunological synapsis⁴⁵, and immune-
351 regulatory interactions between lymphoid and non-lymphoid cells (**Supplementary Figure**
352 **4A**).

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

356 including those related to the neutrophil degranulation, degradation of ECM, interleukin and
357 TGF beta signaling pathways. These pathways and common mechanisms observed in the
358 DIME networks could serve as additional layers to understand the pathogenesis of these type
359 I interferon driven diseases⁴⁶. We further explored such common mechanisms based on the
360 cell-gene relationships in the next section.

361

362 **Pleiotropy based on DIME**

363 Pleiotropy is when one gene affects two or more diseases⁴⁷. Pleiotropy has been observed in
364 several studies for many different diseases based on gene mutation to phenotype
365 associations⁴⁸. Unlike most pleiotropy studies performed in the past that looks at only the
366 common DAGs between two diseases, we used a different approach. Since the DIME
367 networks from the previous analysis (for example, between SSc and pulmonary fibrosis)
368 revealed several overlapping cell-gene connections, we extended this approach to look for
369 cell-gene connections that are common between the DIME networks of all possible pairs of
370 diseases. Hence, we define pleiotropy in this study as the cell-gene connections that are found
371 in one or more diseases. Using this approach, we constructed the pleiotropy network as
372 shown in **Figure 3A**. The pleiotropy network consists of diseases as nodes. The nodes are
373 connected if there exists a significant ($JI \geq 0.1$ and FET p-value ≤ 0.01) number of common
374 cell-gene connections between their corresponding DIME networks, (see methods). The
375 network shown in **Figure 3A** is trimmed to contain only those nodes with degree ≥ 2 . The
376 node colors represent the MeSH term (as shown in **Figure 1B**) of the disease. The grey
377 colored nodes represent those diseases for which the MeSH term ontology was not available
378 from the DisGeNet. As seen from **Figure 3A**, several cancer related diseases (shown by
379 orange color for neoplasm MeSH term) cluster together. Same can be observed for the eye
380 related disease (shown by green color MeSH term) and the chemically induced disorders
381 (shown by pink and yellow color MeSH terms) seen in the bottom of the network **Figure 3A**.
382 These clusters together highlight the similar mechanisms (cell-gene connections) between
383 these diseases.

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

388 DIME-pleiotropy network that had a JI similarity of ≥ 0.1 . In the DIME-pleiotropy network
389 (**Figure 3B**) of Crohn's disease (MeSH: digestive system disease) and psoriasis (MeSH: skin
390 disease), we found cell-gene networks of lymphoid ($CD4^+$, NK and ILC's) and myeloid cells.
391 The pathway enrichment of the DAGs in this DIME-pleiotropy network revealed pathways
392 related to TCR signaling, interleukin signaling, neutrophil degranulation and regulation of
393 TLR (**Supplementary Figure 5A**). As observed in the DIME-pleiotropy network (**Figure**
394 **3B**), $CD4^+$ TH1 cells have been implicated in the pathogenesis of both Crohn's disease and
395 psoriasis^{49,50}. Likewise, in the DIME-pleiotropy network (**Figure 3C**) of SSc (MeSH: skin
396 disease) and pulmonary fibrosis (MeSH: respiratory tract disease), we found cell-gene
397 networks between progenitors and pDCs, NK cells, $CD4^+$ TEMRA and $CD4^+$ T regulatory,
398 and other myeloid cells. The pathway enrichment of the DAGs in this DIME-pleiotropy
399 network revealed pathways (**Supplementary Figure 5B**) related to TGF beta receptor
400 signaling, interleukin, ECM and integrin cell surface interactions. Both of these being fibrotic
401 diseases, the involvement of TGF beta signaling and ECM is well represented in the DIME-
402 pleiotropy network and known to be implicated in literature⁵¹⁻⁵³. Interestingly, we also
403 observed cell-gene network between macrophages and genes like TGF-B1, MMP9 and
404 TIMP1. Evidently, macrophages may exacerbate pulmonary fibrosis by TGF beta production
405 or cause ECM degradation via matrix metalloproteinase (MMP) activities⁵⁴. The DIME-
406 pleiotropy network has captured the intricate network of these key DAGs and the cells
407 (macrophages) accurately.

408 Patients with rheumatoid arthritis (RA) have an increased risk of cancer due to the severe
409 regimen of disease modifying anti-rheumatic drugs⁵⁵. However, RA patients seem to have
410 lower risk of colon cancer in comparison to the general population^{55,56}. To explore the factors
411 responsible for the protective effect against colon cancer in RA patients, we constructed the
412 DIME-pleiotropy network (**Figure 3D**) of RA (MeSH: immune system) and colon carcinoma
413 (MeSH: neoplasm). We found cell-gene networks of $CD4^+$ T cells and many of the myeloid
414 cells. The pathway enrichment of the DAGs in this DIME-pleiotropy network revealed
415 pathways (**Supplementary Figure 5C**) related to TLR signaling, interleukin signaling,
416 neutrophil degranulation, ECM organization, FCER1 and EGFR signaling. Although clear
417 signatures of the protective effect were not observed, we did find the presence of PTGS1 and
418 PTGS2 (also known as COX1 and COX2) in the DIME-pleiotropy network of RA and colon
419 carcinoma. These genes are targets of the non-steroidal anti-inflammatory drugs (NSAIDs)
420 that are frequently taken by RA patients. Evidently, NSAIDs like aspirin have been shown to

421 confer protective affect against colorectal cancer even in lower doses⁵⁷. Perhaps the missing
422 link in the protective effect of RA in colon cancer lies in the anti-inflammatory role of the
423 NSAIDs taken by the RA patients that target the pro-inflammatory mediators⁵⁸ such as
424 PTGS1 and PTGS2 in both diseases, thus conferring protection.

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

430

431 **Immunome mediated drug repurposing**

432 Using our immunome data, the DisGeNet and the NMF, we constructed the DIME network
433 and the DIME-pleiotropy network (as described before and as depicted in **Figure 4A**, part 1
434 and 2). We extended this approach to construct drug-gene network to identify drug targets
435 based on the DIME network (**Figure 4A**, part 3), which are referred to as the DIME-drug
436 network. We then used the pleiotropy method used before to then identify common drug
437 targets between the diseases (**Figure 4A**, part 4) based on the DIME-drug networks, which
438 forms the basis of the immunome mediated drug repurposing, (see methods). **Figure 4B**
439 shows an example of the DIME-drug network of Crohn’s disease that represents the
440 connections between the top DAGs of Crohn’s disease as the potential drug targets and their
441 associated drugs. The DIME-drug network of Crohn’s disease identified some known drugs
442 like the corticosteroids (such as prednisone, methylprednisolone, hydrocortisone and
443 budesonide, targets of NR3C1) and the aminosalicylates (such as sulfasalazine and
444 mesalamine, targets of PTGS2 and ALOX5) that are current line of drugs used in treatment of
445 Crohn’s disease⁶⁰. In addition to the known drugs and drug targets, the DIME-drug network
446 of Crohn’s disease also revealed some novel and potentially interesting drugs and drug
447 targets, such as lifitegrast that target the integrins, ITGB2 and ITGAL. This is particularly
448 interesting because integrin based therapies (such as natalizumab and vedolizumab) are
449 already in use for Crohn’s disease⁶¹. Exploring other integrin based therapies for Crohn’s
450 disease may be beneficial since both ITGAL and ITGB2 show in the top DAGs and are also
451 implicated in Crohn’s disease^{62,63}.

452 As we discovered common cell-gene mechanisms in the DIME-pleiotropy network in the
453 previous analysis (**Figure 3**), we extended this approach to discover plausible drugs and drug
454 targets between the disease pair comparison to find new candidates for drug repurposing, we
455 refer to these analysis as the pleiotropy-drug network analysis (**Figure 4A**, part 4). In
456 addition to the drug targets revealed from the DIME-drug network of Crohn's disease, the
457 pleiotropy drug network of Crohn's disease and psoriasis (**Figure 4C**), comprised some
458 known drug targets such as IL6R for psoriasis and IL1B for Crohn's disease. The associated
459 drug of IL6R, tocilizumab is known to be used in the treatment for psoriasis⁶⁴. Interestingly,
460 anti IL6 therapy has been shown to have promising clinical response for Crohn's disease as
461 well⁶⁵. Similarly, anti-IL1 therapies have also been explored for psoriasis and have shown
462 beneficial clinical outcome⁶⁴. Thus, DIME-network identifies established and novel potential
463 targets for drug repurposing.

464 Studying the pleiotropy-drug network of fibrotic disease systemic scleroderma and
465 pulmonary fibrosis (**Figure 4D**), we found the DNMT1 (a DNA methyl transferase enzyme)
466 targeting drugs decitabine and azacitidine. Epigenetic modulation as a therapy has been
467 explored as a treatment option in SSc⁶⁶. Similarly, in the pleiotropy drug network of SSc and
468 myocardial infarction (MeSH: Cardiovascular), **Figure 4E**, we found canakinumab an IL1B
469 targeting drug. Anti-IL drugs have also been tested on patients with myocardial infarction and
470 were shown to have fewer cardiovascular events than placebo⁶⁷. The pleiotropy drug network
471 of rheumatoid arthritis and colon carcinoma was found to be larger (**Figure 4F**), due to the
472 larger overlap between the two diseases, see **Figure 3D**. The network shows possibilities of
473 using anti-inflammatory drugs for treatment or prevention of colon carcinoma as discussed
474 before. With drugs that target genes such as IL2RG, TYK2, PTGS1, PTGS2, etc. that are
475 widely used to treat anti-inflammatory or immune mediated diseases, these can be used as
476 preventive care drugs or for use along with chemotherapy⁶⁸.

477 We also looked at the known drug targets that are present across the DIME networks of the
478 613 diseases analyzed by us. We found the top 5 drug targets to be BCL2 (**Figure 4B, C, E,**
479 **and F**), PTGS2 (**Figure 4B, 4E-4F**), PIK3CD (**Figure 4D, 4E**), CXCR4 (**Figure 4B-4F**) and
480 IL1B (**Figure 4B, C, E, and F**) that were implicated in the DIME networks of more than 200
481 out of the 613 diseases analyzed by us. The CXCR4 targetable by plerixafor is present in all
482 of the pleiotropy-drug networks shown here (**Figure 4B-4F**). Plerixafor is a drug intended for
483 use in cancer after stem cell transplantation to initiate migration of stem cells in the
484 bloodstream⁶⁹. This drug is now in clinical trials (NCT01413100) to be evaluated for use

485 after autologous transplant in patients with scleroderma⁷⁰. Such trials may potentially be
486 extended to other immune diseases like psoriasis, Crohn's disease, rheumatoid arthritis, and
487 potentially for a variety of other phenotypically distinct rheumatic diseases that are driven by
488 CXCR4 mediated dysregulation of immune system. Many such potential drug repurposing
489 targets (enlisted in **Supplementary Table 2**) can be similarly evaluated in future studies and
490 trials.

491

492 **Discussion**

493 Despite decades of experimental data on the understanding of the molecular mechanism of
494 diseases, we know little about the perturbations in the niche cell compartments that are
495 specific to the disease. To address this gap several efforts at the tissue⁵⁻⁷ and immune cell^{16,17}
496 level have been performed to identify these disease specific compartments. However,
497 previous studies have concentrated at whole tissues, not distinguishing different immune cell
498 subsets or, in contrast, focused on a few immune cell subsets thereby likely missing the
499 complex molecular network underpinning immune mediated diseases. Additionally, it is
500 important to understand which of the gene subsets contribute to a mechanism within the
501 different cell populations. A disease may have perturbations occurring at a single cell type or
502 at multiple cell types with completely different or similar genes being involved. To truly
503 understand a disease, it is essential to capture these cell-gene networks as holistically as
504 possible.

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

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

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

538

539 REFERENCES

- 540 1. Sun, Y. V & Hu, Y.-J. Integrative Analysis of Multi-omics Data for Discovery and
541 Functional Studies of Complex Human Diseases. *Adv. Genet.* **93**, 147–90 (2016).
- 542 2. Li, Y. & Agarwal, P. A Pathway-Based View of Human Diseases and Disease
543 Relationships. *PLoS One* **4**, e4346 (2009).
- 544 3. Camp, J. G., Platt, R. & Treutlein, B. Mapping human cell phenotypes to genotypes
545 with single-cell genomics. *Science (80-.)*. **365**, 1401–1405 (2019).
- 546 4. Rozenblatt-Rosen, O., Stubbington, M. J. T., Regev, A. & Teichmann, S. A. The
547 Human Cell Atlas: from vision to reality. *Nature* **550**, 451–453 (2017).
- 548 5. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**,
549 580–585 (2013).

- 550 6. Ardlie, K. G. *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis:
551 Multitissue gene regulation in humans. *Science (80-.)*. **348**, 648–660 (2015).
- 552 7. GTEx Consortium *et al.* Genetic effects on gene expression across human tissues.
553 *Nature* **550**, 204–213 (2017).
- 554 8. Tirosh, I. & Suvà, M. L. Deciphering Human Tumor Biology by Single-Cell
555 Expression Profiling. *Annu. Rev. Cancer Biol.* **3**, 151–166 (2019).
- 556 9. Keren-Shaul, H. *et al.* A Unique Microglia Type Associated with Restricting
557 Development of Alzheimer’s Disease. *Cell* **169**, 1276-1290.e17 (2017).
- 558 10. Kuo, D. *et al.* HBEGF+ macrophages in rheumatoid arthritis induce fibroblast
559 invasiveness. *Sci. Transl. Med.* **11**, eaau8587 (2019).
- 560 11. Knight, J. C. Genomic modulators of the immune response. *Trends Genet.* **29**, 74–83
561 (2013).
- 562 12. Raj, T. *et al.* Polarization of the Effects of Autoimmune and Neurodegenerative Risk
563 Alleles in Leukocytes. *Science (80-.)*. **344**, 519–523 (2014).
- 564 13. Quach, H. *et al.* Genetic Adaptation and Neandertal Admixture Shaped the Immune
565 System of Human Populations. *Cell* **167**, 643-656.e17 (2016).
- 566 14. Lee, M. N. *et al.* Common Genetic Variants Modulate Pathogen-Sensing Responses in
567 Human Dendritic Cells. *Science (80-.)*. **343**, 1246980–1246980 (2014).
- 568 15. Fairfax, B. P. *et al.* Innate Immune Activity Conditions the Effect of Regulatory
569 Variants upon Monocyte Gene Expression. *Science (80-.)*. **343**, 1246949–1246949
570 (2014).
- 571 16. Chen, L. *et al.* Genetic Drivers of Epigenetic and Transcriptional Variation in Human
572 Immune Cells. *Cell* **167**, 1398-1414.e24 (2016).
- 573 17. Schmiedel, B. J. *et al.* Impact of Genetic Polymorphisms on Human Immune Cell
574 Gene Expression. *Cell* **175**, 1701-1715.e16 (2018).
- 575 18. SRA-Tools - NCBI. <http://ncbi.github.io/sra-tools/>.
- 576 19. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21
577 (2013).
- 578 20. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,
579 2078–2079 (2009).
- 580 21. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-
581 throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
- 582 22. Risso, D., Ngai, J., Speed, T. P. & Dudoit, S. Normalization of RNA-seq data using
583 factor analysis of control genes or samples. *Nat. Biotechnol.* **32**, 896–902 (2014).
- 584 23. Piñero, J. *et al.* DisGeNET: a comprehensive platform integrating information on
585 human disease-associated genes and variants. *Nucleic Acids Res.* **45**, D833–D839
586 (2017).
- 587 24. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide
588 association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**,
589 D1005–D1012 (2019).

- 590 25. Gaujoux, R. & Seoighe, C. A flexible R package for nonnegative matrix factorization.
591 *BMC Bioinformatics* **11**, 367 (2010).
- 592 26. Brunet, J.-P., Tamayo, P., Golub, T. R. & Mesirov, J. P. Metagenes and molecular
593 pattern discovery using matrix factorization. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 4164–
594 9 (2004).
- 595 27. Cotto, K. C. *et al.* DGIdb 3.0: a redesign and expansion of the drug–gene interaction
596 database. *Nucleic Acids Res.* **46**, D1068–D1073 (2018).
- 597 28. Vergheze, P. B., Castellano, J. M. & Holtzman, D. M. Roles of Apolipoprotein E in
598 Alzheimer’s Disease and Other Neurological Disorders. *Lancet Neurol.* **10**, 241
599 (2011).
- 600 29. Costarelli, L., Malavolta, M., Giacconi, R. & Provinciali, M. Dysfunctional
601 macrophages in Alzheimer Disease: another piece of the ‘macroph-aging’
602 puzzle? *Aging (Albany, NY)*. **9**, 1865–1866 (2017).
- 603 30. Strittmatter, W. J. *et al.* Apolipoprotein E: high-avidity binding to beta-amyloid and
604 increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc.*
605 *Natl. Acad. Sci. U. S. A.* **90**, 1977–81 (1993).
- 606 31. Bell, R. D. *et al.* Transport Pathways for Clearance of Human Alzheimer’s Amyloid β -
607 Peptide and Apolipoproteins E and J in the Mouse Central Nervous System. *J. Cereb.*
608 *Blood Flow Metab.* **27**, 909–918 (2007).
- 609 32. Sanchez, V. E., Nichols, C., Kim, H. N., Gang, E. J. & Kim, Y.-M. Targeting PI3K
610 Signaling in Acute Lymphoblastic Leukemia. *Int. J. Mol. Sci.* **20**, (2019).
- 611 33. Biemacka, A., Dobaczewski, M. & Frangogiannis, N. G. TGF- β signaling in fibrosis.
612 *Growth Factors* **29**, 196–202 (2011).
- 613 34. Yipp, B. G. & Kubes, P. NETosis: how vital is it? *Blood* **122**, 2784–94 (2013).
- 614 35. Neutrophil Extracellular Traps in Systemic Sclerosis - Full Text View -
615 ClinicalTrials.gov. <https://clinicaltrials.gov/ct2/show/NCT03374618>.
- 616 36. M. Artlett, C. The Role of the NLRP3 Inflammasome in Fibrosis. *Open Rheumatol. J.*
617 **6**, 80–86 (2012).
- 618 37. Yu, Y. & Su, K. Neutrophil Extracellular Traps and Systemic Lupus Erythematosus. *J.*
619 *Clin. Cell. Immunol.* **4**, (2013).
- 620 38. Lande, R. *et al.* Neutrophils Activate Plasmacytoid Dendritic Cells by Releasing Self-
621 DNA–Peptide Complexes in Systemic Lupus Erythematosus. *Sci. Transl. Med.* **3**,
622 73ra19-73ra19 (2011).
- 623 39. Garcia-Romo, G. S. *et al.* Netting Neutrophils Are Major Inducers of Type I IFN
624 Production in Pediatric Systemic Lupus Erythematosus. *Sci. Transl. Med.* **3**, 73ra20-
625 73ra20 (2011).
- 626 40. Villanueva, E. *et al.* Netting Neutrophils Induce Endothelial Damage, Infiltrate
627 Tissues, and Expose Immunostimulatory Molecules in Systemic Lupus Erythematosus.
628 *J. Immunol.* **187**, 538–552 (2011).
- 629 41. Leffler, J. *et al.* Neutrophil Extracellular Traps That Are Not Degraded in Systemic
630 Lupus Erythematosus Activate Complement Exacerbating the Disease. *J. Immunol.*

- 631 **188**, 3522–3531 (2012).
- 632 42. Yang, W.-C. *et al.* Interleukin-4 Supports the Suppressive Immune Responses Elicited
633 by Regulatory T Cells. *Front. Immunol.* **8**, 1508 (2017).
- 634 43. Wang, R. *et al.* Association of interleukin 13 gene polymorphisms and plasma IL 13
635 level with risk of systemic lupus erythematosus. *Cytokine* **104**, 92–97 (2018).
- 636 44. Gao, W. *et al.* Notch signalling pathways mediate synovial angiogenesis in response to
637 vascular endothelial growth factor and angiopoietin 2. *Ann. Rheum. Dis.* **72**, 1080–8
638 (2013).
- 639 45. Moulton, V. R. & Tsokos, G. C. Abnormalities of T cell signaling in systemic lupus
640 erythematosus. *Arthritis Res. Ther.* **13**, 207 (2011).
- 641 46. Crow, M. K. & Rönnblom, L. Report of the inaugural Interferon Research Summit:
642 interferon in inflammatory diseases. *Lupus Sci. Med.* **5**, e000276 (2018).
- 643 47. Wagner, G. P. & Zhang, J. The pleiotropic structure of the genotype–phenotype map:
644 the evolvability of complex organisms. *Nat. Rev. Genet.* **12**, 204–213 (2011).
- 645 48. Yang, C., Li, C., Wang, Q., Chung, D. & Zhao, H. Implications of pleiotropy:
646 challenges and opportunities for mining Big Data in biomedicine. *Front. Genet.* **6**, 229
647 (2015).
- 648 49. Casciano, F., Pigatto, P. D., Secchiero, P., Gambari, R. & Reali, E. T Cell Hierarchy in
649 the Pathogenesis of Psoriasis and Associated Cardiovascular Comorbidities. *Front.*
650 *Immunol.* **9**, 1390 (2018).
- 651 50. Li, N. & Shi, R.-H. Updated review on immune factors in pathogenesis of Crohn’s
652 disease. *World J. Gastroenterol.* **24**, 15–22 (2018).
- 653 51. Todd, N. W., Luzina, I. G. & Atamas, S. P. Molecular and cellular mechanisms of
654 pulmonary fibrosis. *Fibrogenesis Tissue Repair* **5**, 11 (2012).
- 655 52. Wynn, T. A. Integrating mechanisms of pulmonary fibrosis. *J. Exp. Med.* **208**, 1339–50
656 (2011).
- 657 53. Pattanaik, D., Brown, M., Postlethwaite, B. C. & Postlethwaite, A. E. Pathogenesis of
658 Systemic Sclerosis. *Front. Immunol.* **6**, 272 (2015).
- 659 54. Kolahian, S., Fernandez, I. E., Eickelberg, O. & Hartl, D. Immune Mechanisms in
660 Pulmonary Fibrosis. *Am. J. Respir. Cell Mol. Biol.* **55**, 309–322 (2016).
- 661 55. Wilton, K. M. & Matteson, E. L. Malignancy Incidence, Management, and Prevention
662 in Patients with Rheumatoid Arthritis. *Rheumatol. Ther.* **4**, 333–347 (2017).
- 663 56. Simon, T. A., Thompson, A., Gandhi, K. K., Hochberg, M. C. & Suissa, S. Incidence
664 of malignancy in adult patients with rheumatoid arthritis: a meta-analysis. *Arthritis*
665 *Res. Ther.* **17**, 212 (2015).
- 666 57. Din, F. V. N. *et al.* Effect of aspirin and NSAIDs on risk and survival from colorectal
667 cancer. *Gut* **59**, 1670–1679 (2010).
- 668 58. Dubois, R. N. Role of inflammation and inflammatory mediators in colorectal cancer.
669 *Trans. Am. Clin. Climatol. Assoc.* **125**, 358–72; discussion 372-3 (2014).
- 670 59. Chesmore, K., Bartlett, J. & Williams, S. M. The ubiquity of pleiotropy in human

- 671 disease. *Hum. Genet.* **137**, 39–44 (2018).
- 672 60. Akobeng, A. K. Crohn's disease: current treatment options. *Arch. Dis. Child.* **93**, 787–
673 92 (2008).
- 674 61. Ley, K., Rivera-Nieves, J., Sandborn, W. J. & Shattil, S. Integrin-based therapeutics:
675 biological basis, clinical use and new drugs. *Nat. Rev. Drug Discov.* **15**, 173–83
676 (2016).
- 677 62. Ellinghaus, D. *et al.* Analysis of five chronic inflammatory diseases identifies 27 new
678 associations and highlights disease-specific patterns at shared loci. *Nat. Genet.* **48**,
679 510–518 (2016).
- 680 63. Bernstein, C. N., Sargent, M. & Rector, E. Alteration in expression of beta 2 integrins
681 on lamina propria lymphocytes in ulcerative colitis and Crohn's disease. *Clin.*
682 *Immunol.* **104**, 67–72 (2002).
- 683 64. Tsai, Y.-C. & Tsai, T.-F. Anti-interleukin and interleukin therapies for psoriasis: current
684 evidence and clinical usefulness. *Ther. Adv. Musculoskelet. Dis.* **9**, 277–294 (2017).
- 685 65. Ito, H. *et al.* A pilot randomized trial of a human anti-interleukin-6 receptor
686 monoclonal antibody in active Crohn's disease. *Gastroenterology* **126**, 989–996
687 (2004).
- 688 66. O'Reilly, S. Epigenetic modulation as a therapy in systemic sclerosis. *Rheumatology*
689 **58**, 191–196 (2019).
- 690 67. Ridker, P. M. *et al.* Antiinflammatory Therapy with Canakinumab for Atherosclerotic
691 Disease. *N. Engl. J. Med.* **377**, 1119–1131 (2017).
- 692 68. Terzić, J., Grivennikov, S., Karin, E. & Karin, M. Inflammation and Colon Cancer.
693 *Gastroenterology* **138**, 2101-2114.e5 (2010).
- 694 69. Hübel, K. *et al.* Plerixafor with and without chemotherapy in poor mobilizers: results
695 from the German compassionate use program. *Bone Marrow Transplant.* **46**, 1045–
696 1052 (2011).
- 697 70. Scleroderma Treatment With Autologous Transplant (STAT) Study - Full Text View -
698 ClinicalTrials.gov. <https://clinicaltrials.gov/ct2/show/study/NCT01413100>.
- 699
- 700

701 **FIGURE LEGENDS**

702

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

711

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

716

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

724

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

731

732

733 **Supplementary Information**

734

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

737

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

740

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

743

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

747

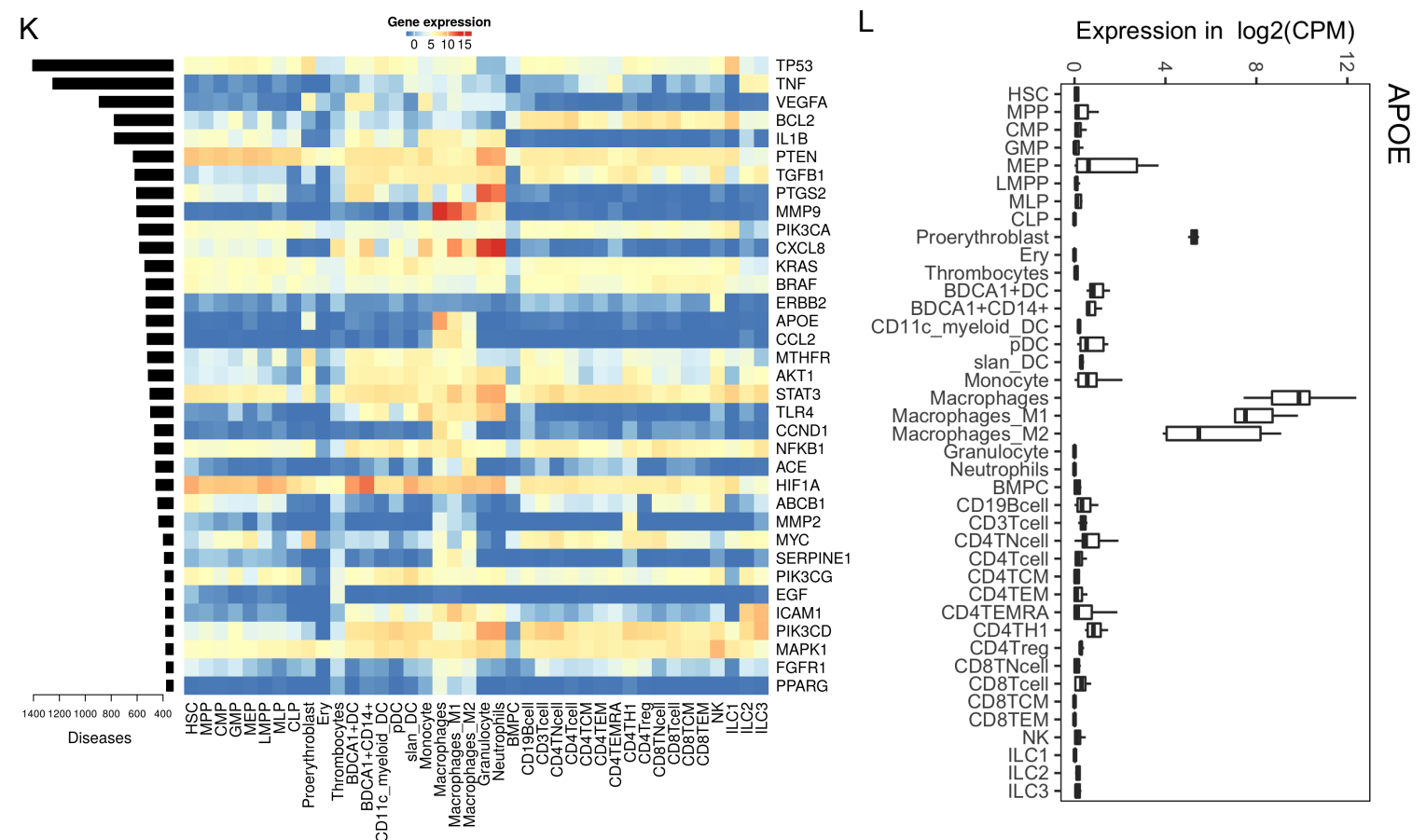
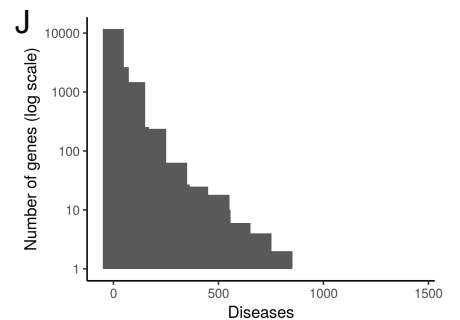
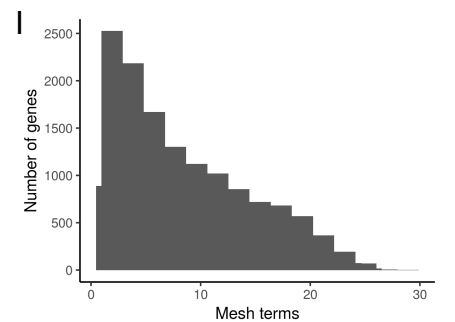
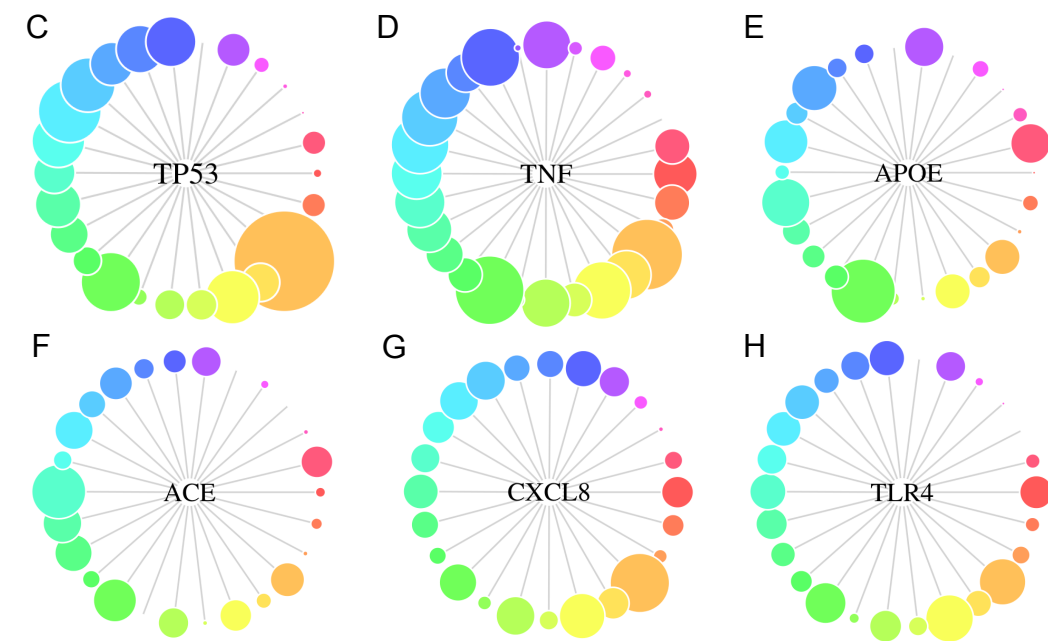
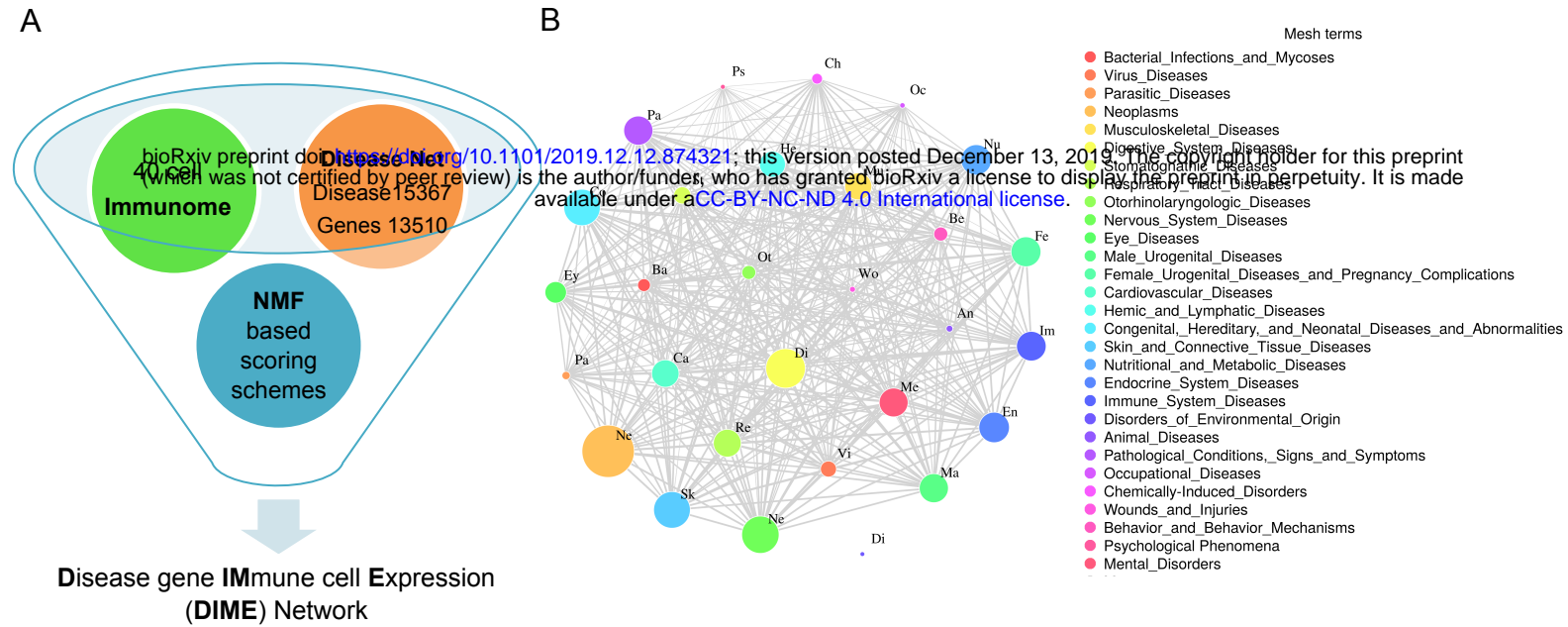
748 **Supplementary Figure 5:** Reactome pathway enrichment analysis of pleiotropy DIME
749 network of **A** Crohn's disease and psoriasis, **B** systemic scleroderma and pulmonary fibrosis,
750 and **C** rheumatoid arthritis and colon carcinoma.

751

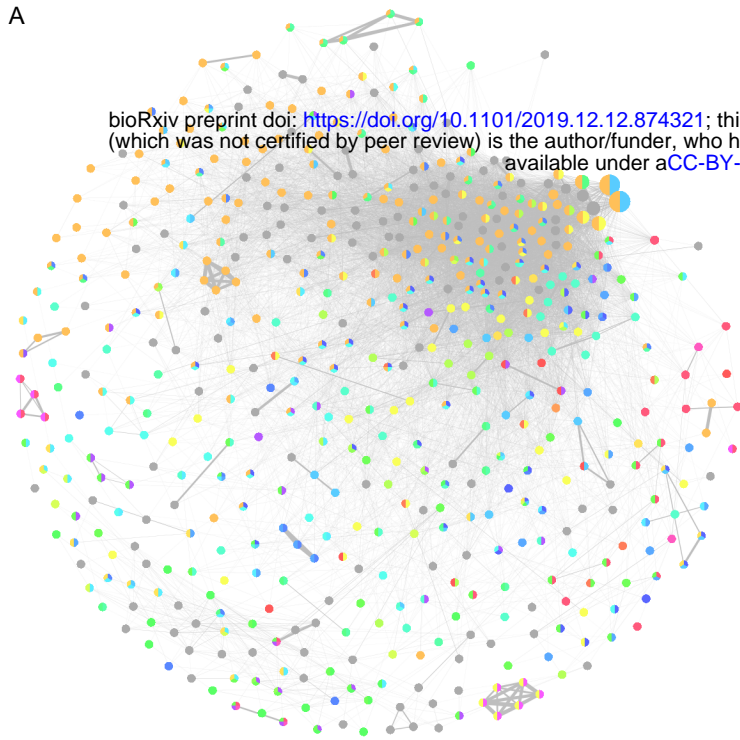
752 **Supplementary Table 1:** Represents all GEO datasets and samples used to construct the
753 *immunome*.

754

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

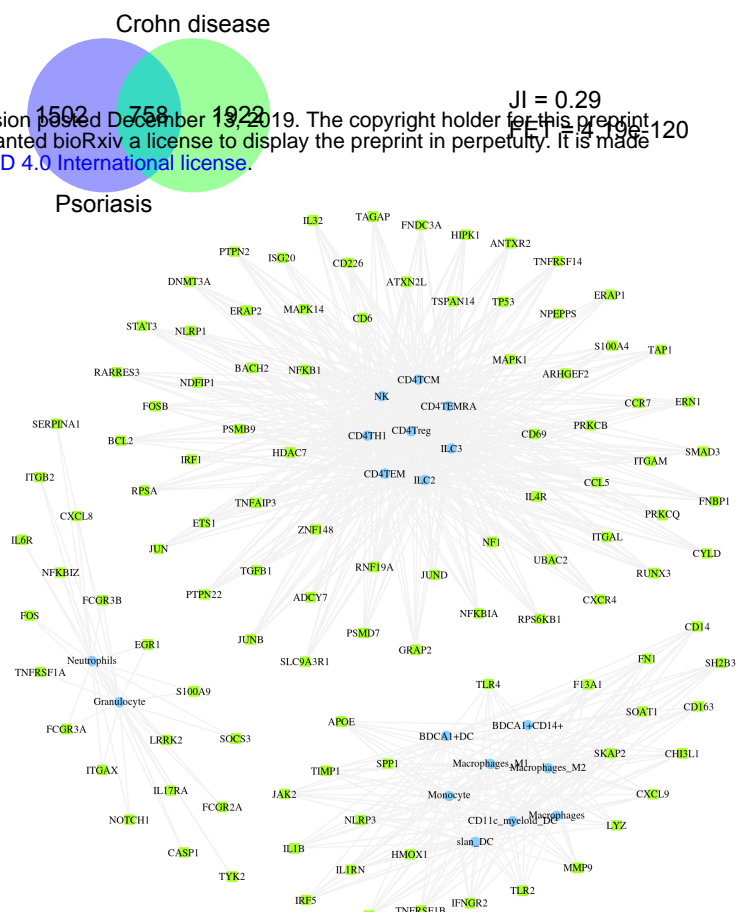


A



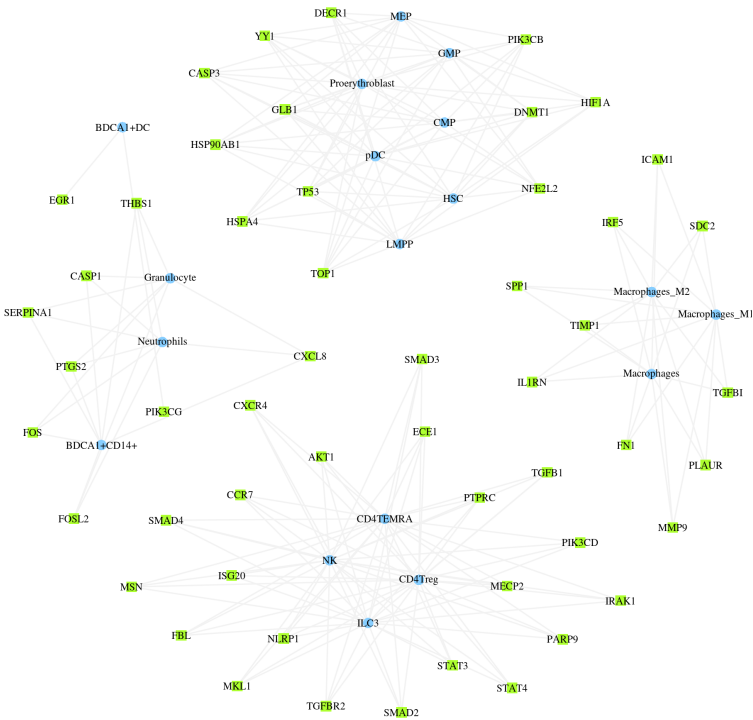
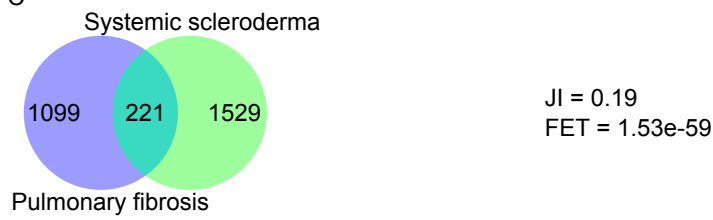
Top Pleiotropy Network
 $JI \geq 0.1$, $FET \leq 0.01$, Degree ≥ 2

B



$JI = 0.29$
 $FET = 4.19e-120$

C



D

