

RESEARCH

Open Access



scDrugPrio: a framework for the analysis of single-cell transcriptomics to address multiple problems in precision medicine in immune-mediated inflammatory diseases

Samuel Schäfer^{1,2}, Martin Smelik³, Oleg Sysoev⁴, Yelin Zhao³, Desiré Eklund¹, Sandra Lilja^{1,5}, Mika Gustafsson⁶, Holger Heyn^{7,8}, Antonio Julia⁹, István A. Kovács^{10,11}, Joseph Loscalzo¹², Sara Marsal⁹, Huan Zhang¹, Xinxu Li³, Danuta Gawel^{5†}, Hui Wang^{3,13†} and Mikael Benson^{3*†} 

Abstract

Background Ineffective drug treatment is a major problem for many patients with immune-mediated inflammatory diseases (IMIDs). Important reasons are the lack of systematic solutions for drug prioritisation and repurposing based on characterisation of the complex and heterogeneous cellular and molecular changes in IMIDs.

Methods Here, we propose a computational framework, scDrugPrio, which constructs network models of inflammatory disease based on single-cell RNA sequencing (scRNA-seq) data. scDrugPrio constructs detailed network models of inflammatory diseases that integrate information on cell type-specific expression changes, altered cellular crosstalk and pharmacological properties for the selection and ranking of thousands of drugs.

Results scDrugPrio was developed using a mouse model of antigen-induced arthritis and validated by improved precision/recall for approved drugs, as well as extensive *in vitro*, *in vivo*, and *in silico* studies of drugs that were predicted, but not approved, for the studied diseases. Next, scDrugPrio was applied to multiple sclerosis, Crohn's disease, and psoriatic arthritis, further supporting scDrugPrio through prioritisation of relevant and approved drugs. However, in contrast to the mouse model of arthritis, great interindividual cellular and gene expression differences were found in patients with the same diagnosis. Such differences could explain why some patients did or did not respond to treatment. This explanation was supported by the application of scDrugPrio to scRNA-seq data from eleven individual Crohn's disease patients. The analysis showed great variations in drug predictions between patients, for example, assigning a high rank to anti-TNF treatment in a responder and a low rank in a nonresponder to that treatment.

Conclusions We propose a computational framework, scDrugPrio, for drug prioritisation based on scRNA-seq of IMID disease. Application to individual patients indicates scDrugPrio's potential for personalised network-based drug screening on cellulome-, genome-, and drugome-wide scales. For this purpose, we made scDrugPrio into an easy-to-use R package (<https://github.com/SDTC-CPMed/scDrugPrio>).

[†]Danuta Gawel, Hui Wang and Mikael Benson are joint last authors.

*Correspondence:

Mikael Benson

mikael.benson@ki.se

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Keywords Single-cell RNA sequencing, scRNA-seq, Immune-mediated inflammatory disease, Drug prioritisation, Drug repurposing, Drug prediction

Background

Immune-mediated inflammatory diseases (IMIDs), such as rheumatoid arthritis, Crohn's disease, and psoriatic arthritis, affect millions of people worldwide and can cause chronic pain, disability, and reduced quality of life [1]. While new classes of therapies are transforming the management of IMIDs, it is still a general problem that many patients do not achieve remission with mono- [2, 3] or combinatorial therapy [3]. This may be due to drug development involving testing drugs on large groups of patients, with the assumption that the drug will work similarly on all patients. Such an approach does not take into account the fact that each individual's genetic makeup and environment are unique, leading to significant variations in drug efficacy and side effects.

Given that IMIDs are known to involve thousands of genes that are variably expressed in different cell types and show temporal and interindividual differences [4, 5], single-cell RNA sequencing (scRNA-seq) provides a promising foundation for the identification of suitable drug treatments [6]. Indeed, one pioneering case report described scRNA-guided therapy of one patient with an inflammatory disease [7]. The case report described successful outcomes in a patient who did not respond to standard treatment. A limitation was that drug selection was empirical rather than based on systems-level understanding of the relative importance of disease-associated cell types, pathways, and genes.

Several systematic prediction models for drug selection in cancer exist, in which omics data are leveraged to determine the chemotherapies' "killing potential" of tumour cells [8–11]. However, these models are not immediately translatable to IMIDs as they are (1) trained on large public drug-response data (e.g. GDSC database [12] and PRISM [13]), which are thus far unavailable for IMIDs, and (2) pursuing the eradication of disease-associated cell types. Rather few methodologies are applicable to IMIDs, including (1) identification of all druggable targets [14, 15], (2) targeting enriched pathways [7, 15], (3) network-based proximity calculations [6, 16] or (4) matching of transcriptomic signatures as by Connectivity Map (CMap) [17]. A limitation of these approaches is that they are developed using bulk transcriptomics or genetic variants and hence do not possess inherent solutions for rank aggregation for parallel analyses of several cell types, which limits their applicability to scRNA-seq.

Aiming to create a systematic framework for scRNA-seq-based drug prioritisation and repositioning in

inflammatory diseases, we hypothesised that the limitations of previous methodologies could be overcome by transposing network-based approaches [6, 16] to a systematic and scalable strategy for network-based virtual drug screening of multicellular disease models (MCDMs). Therefore, we composed a computational framework henceforth referred to as scDrugPrio (Fig. 1). Using scRNA-seq-derived differentially expressed genes (DEGs) of either (1) one individual or (2) a group comparison between patients and controls, scDrugPrio starts by identifying cell type-specific drug candidates by considering both proximity in a protein–protein interaction network and biopharmacological criteria. To rank drug candidates, scDrugPrio calculates two measures, intracellular and extracellular centrality. We used these two measures to capture two important drug properties, namely, (1) proficiency in targeting key disease-associated expression changes in a cell type and (2) the relative importance of the targeted cell type. These measures are then aggregated over all cell types to provide a final drug ranking.

Because of the complexity and heterogeneity of IMIDs, we started by developing scDrugPrio using scRNA-seq data from a mouse model of antigen-induced arthritis. This reduced heterogeneity since the mice are inbred and the disease induced in a standardised way. Moreover, the mouse model allowed extensive *in vitro* and *in vivo* validation studies. To illustrate some potential case-of-use scenarios, we next applied scDrugPrio to cerebrospinal fluid from multiple sclerosis patients and intestinal biopsies from Crohn's disease (CD) patients. Our analyses demonstrated drug selection and ranking capabilities through (1) the prioritisation of known drugs. For antigen-induced arthritis, drug selection was also supported by (2) experimental validation of repurposed drugs and (3) favourable comparison with previous methods. Next, we applied scDrugPrio to paired biopsies from individual CD patients, revealing its ability to capture the significant heterogeneity in individualised therapeutic prioritisations.

Methods

scDrugPrio's computational framework

As indicated in Additional file 1: Fig. S1, scDrugPrio requires (1) an adjusted scRNA-seq matrix, (2) disease-associated differentially expressed genes (DEGs) for each cell type from either group-based comparison of healthy

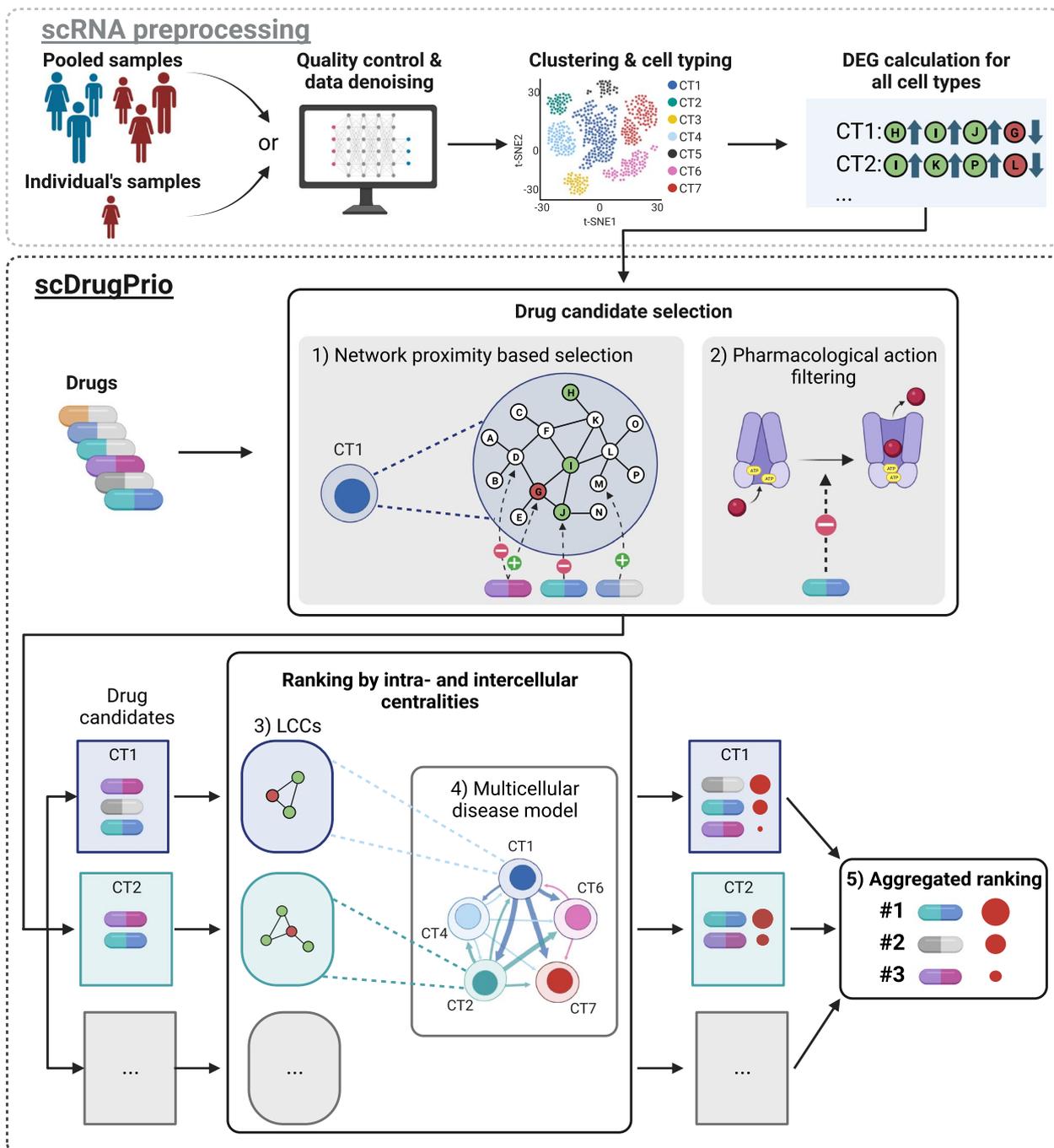


Fig. 1 Overview of the scDrugPrio workflow. Samples were derived from inflamed tissue in patients (red) and healthy tissues in controls (blue) or, for individual predictions, sampling included inflamed and uninflamed tissue taken from the same patient. Single-cell RNA-sequencing (scRNA-seq) data were preprocessed by undergoing quality control, denoising, clustering, cell typing and differentially expressed gene (DEG) calculation. DEGs for each cell type were calculated between healthy and sick samples. Using DEGs alongside information on drugs, scDrugPrio selects drug candidates (for each cell type; CT) whose gene targets are (1) in network proximity to DEGs (network proximity based selection) and (2) who counteract disease-associated expression changes (pharmacological action filtering). These cell type-specific drug candidates are next ranked using intracellular and intercellular centrality. (3) Intracellular centrality is computed based on the centrality of drug targets in the largest connected component (LCC) formed by DEGs and functions as a proxy for drug target importance. (4) Intercellular centrality measures centrality in disease-associated cellular crosstalk networks called multicellular disease models (MCDMs). (5) To derive a final ranking that aggregated cell type-specific drug selection and ranking into one list, drug candidates were ranked using a composite score of intra- and intercellular centralities (Additional file 1: Fig. S1). Drugs with identical targets and mechanism of action were given identical rankings

and sick samples or from inflamed and noninflamed samples of one individual, (3) a protein–protein interaction network (PPIN) and (4) drug–target information. scDrugPrio then utilises this information for cell type-specific drug selection, calculation of drug ranking measures and finally rank aggregation. Unless otherwise stated, analysis was performed in R 3.6.3.

For drug selection, scDrugPrio first computes the mean closest network distance (d_c) between cell type-specific DEGs and drug targets in the PPIN for each cell type–drug combination [16]. To calculate z-scores (z_c) for network distance, permutation tests (1000 iterations) were performed in which both cell type-specific DEGs and drug targets were randomised in a bin-adjusted manner [16] before the mean closest distance was calculated. The minimal bin size for randomisation was set at 100 genes. Drugs that did not have any target in the interactome were removed from the analysis ($n=4$ for literature-curated PPIN). Based on network distance, we selected only drugs that targets were significantly close ($z_c < -1.64$ corresponding to one-sided $P < 0.05$) to DEGs and that frequently targeted DEGs directly ($d_c < 1$). These cut-offs were chosen based on our empirical observations (e.g. Additional file 1: Fig. S2h, j; Additional file 2) and previous knowledge [18]. As the significance of network proximity can depend on the number of DEGs relative to the size of the network, cut-offs allowing only the top significant DEGs to enter analysis were implemented when needed. Cell type-specific drug candidates were selected further by requiring drugs to counteract the fold change of at least one targeted DEG. This criterion intuitively removes drugs that likely will not help to restore transcriptomic homeostasis. For this purpose, the pharmacological action of the drugs on their targets was determined. Binary drug action (activating/enhancing or inhibiting) on the drug target was recorded for each drug (Additional files 3, 4, 5, 6 and 7). If the pharmacological effect of the drug on the target had not been specified explicitly in DrugBank [19], a literature search was performed using the drug name and gene symbol of the targeted DEG as search terms in PubMed and Google Scholar. Additional information gathered from the literature can be found in Additional files 3, 4, 5, 6 and 7. In case the pharmacological effect of the drug on a target, despite a literature search, could not be classified as enhancing/activating or inhibiting, the drug target was assumed to not counteract fold-change.

Drug ranking by intra- and intercellular centralities was motivated by empirical observations (Fig. 3e; Additional file 1: Fig. S3) in our study as well as previous indications of the biological importance of disease modules [6, 20] and central cell types in MCDMs [6]. For calculation of intracellular centrality, disease modules for each cell type

were defined as the largest connected component (LCC) formed by a cell type's DEGs in the PPIN. For LCC identification, the igraph R package [21] was utilised. To avoid overparameterisation, the eigenvector centrality [22] of DEGs in the LCC was calculated using the CINNA R package [23]. For each drug, the intracellular centrality was calculated as the geometric mean of its differentially expressed target centrality scores in the cell type-specific LCCs. If a drug did not target any DEG included in the LCC, intracellular centrality was set to zero.

Intercellular centrality was calculated using MCDMs that modelled disease-associated cellular crosstalk. For the creation of MCDMs, first, cell type interactions were predicted using NicheNet [24]. Briefly, NicheNet predicts and ranks ligand–target links between interacting cells by combining their expression data with prior knowledge on signalling and gene regulatory networks. As suggested by Browaeys et al. [24], Pearson correlation was used to measure each ligand's ability to predict the gene expression of genes in the *gene set of interest* compared to *background genes* in the receiving cell type. This means that a ligand has a strong positive correlation coefficient if its cognate receptor and the downstream genes of that receptor are all differentially expressed in the downstream cell type. We downloaded the human ligand–target model as well as the human ligand–receptor network (downloaded from <https://zenodo.org/record/3260758> April 2020). Cell type-specific DEGs constituted the *gene set of interest*. A set of *potentially active ligands* was defined as the intersection of ligands included in the downloaded human ligand–target model and ligands among respective cell type DEGs. *Background genes* for each cell type were defined as genes (i) in the denoised single-cell expression matrix D of k cell type-associated cells that showed a mean aggregate expression, $Ea(i)$, over $Ea(i) \geq 0.2$. This definition of background genes was similar to definitions by Browaeys et al. [24] and Puram et al. [25]. At the chosen cut-off, we identified ca. 10,000 background genes that corresponded to the recommended amount for NicheNet calculations [24].

$$Ea(i) = \log_2 \left(\sum_{j=1}^k 10^{D_{ij}} / k \right)$$

All genes were translated to human Entrez gene symbols using human–mouse orthologues downloaded from NCBI (August 2019). Ligand activity analysis in NicheNet was performed for all possible cell type pairs, including self-interactions, excluding cell types that did not express DEGs. In the next step of MCDM construction, directed cell type interactions were derived from ligand activity results and weighted by NicheNet-derived Pearson coefficients. Only ligand interactions with a positive Pearson correlation coefficient were considered negative Pearson

coefficients that reflected the association of a ligand with background genes and therefore were not biologically relevant. The resulting MCDM was visualised using Cytoscape 3.6.1 [26], and for visualisation purposes, the sum of Pearson coefficients that described the directed interaction between two cell types was used. Supplemental analysis supported the relevance of identified ligand interactions (Additional file 2). Eigenvector centrality was calculated for each cell type based on the weighted, directed interactions in the MCDM using the *igraph* and *CINNA* R package [21, 23]. The intercellular centrality of each drug was computed as the sum of MCDM centralities of the cell types that had selected the drug as a candidate. While eigenvector centrality is well tailored to capture central disease-associated cell types in the MCDM, considering both direct and indirect node connections [22], multiple centrality measures are available. We evaluated several of them, finding them to yield similar results to eigenvector centrality (as described in Additional file 2 and listed in Additional files 3, 4, 5, 6 and 7).

Final rank aggregation involved the calculation of a drug's compounded intra- and intercellular centrality. For this, we calculated combined intracellular centrality for each drug as the sum of drug-specific intracellular centralities in all cell types. The centrality compound score consisted of a drug *intercellular centrality* + $0.1 \times$ *combined intracellular centrality* and thereby emphasised the importance of intercellular centrality over intracellular centrality. Intracellular centrality effectively worked as a tiebreaker. Drugs were ranked based on centrality compound scores, using the average position for ties.

Drug data

We retrieved data on 13,339 drugs from DrugBank [19] (downloaded July 2019) and selected only drugs that had been or currently FDA approved ($n=4021$) were indicated for use in humans ($n=1964$) and had at least one human protein target ($n=1864$). Of those, drug targets could be translated to human Entrez IDs for 1844 drugs. The drug-target interactions used are provided in Additional file 3. Sets of drugs that are approved for each disease were identified according to DrugBank's [19] "Indication" category (as described in Additional file 2 and listed in Additional files 3, 4, 5 & 7). Unless otherwise specified, precision is calculated using these disease-specific sets of FDA-approved drugs as *relevant drugs* or true positives.

For validation of ranked drug candidates, we also downloaded data from www.clinicaltrials.gov (September 2023). Data included information on 465,269 clinical trials registered from September 17th, 1999, to September 7th, 2023. Clinical trials ($n=70,396$) for 1085 of the included 1844 drugs were found. To derive information

on the disease relevance of drug candidates, we filtered clinical trials further by MESH terms, resulting in sets of 724, 494, 532, and 140 drugs that had been tried for rheumatoid arthritis, multiple sclerosis, Crohn's disease and psoriatic arthritis, respectively (Additional files 3, 4, 5, 6 and 7). Even though the outcome of such trials is largely unknown, using drugs registered for clinical trials alongside approved drugs for calculation of precision tests *scDrugPrio*'s ability to capture the pharmacological consensus of the medical community on drugs with an expected effect.

For further validation of drug ranking, we also performed a literature search for the top 100 ranked drug candidates of each data set. We systematically searched PubMed and Google Scholar between June 2020 and December 2022 using the specific disease denotation and the drug name as search terms. No restrictions or filters were applied. The relevance of the identified articles was screened by title and abstract. When no relevant articles were identified, the drug name was replaced by the substance name, and another search was conducted. To be eligible, studies had to (1) include a control group, (2) be a human clinical study or rodent experiment, (3) measure inflammatory activity and (4) be accessible. When several studies were identified that reported contradictory results, the drug was labelled as having a previously reported effect, reasoning that it would be impossible to determine the evidence and accuracy level in every such instance. In Additional files 3, 4, 5 and 6, a summary on the nature of the identified article and a full reference is provided, listed by drug.

Precision and recall

$$precision = \frac{|\{\text{relevant drugs}\} \cap \{\text{drugs in selection}\}|}{|\{\text{drugs in selection}\}|}$$

$$recall = \frac{|\{\text{relevant drugs}\} \cap \{\text{drugs in selection}\}|}{|\{\text{relevant drugs}\}|}$$

When referring to precision among the top 100 candidates, we refer to all candidates with rank ≤ 100 .

Protein–protein interaction network

The human interactome was derived from do Valle et al. [27]. The literature-curated interactome included 351,444 protein–protein interactions (PPIs) connecting 17,706 unique proteins and was annotated using Entrez Gene IDs. The largest connected component included 351,393

PPIs and 17,651 proteins. Only the largest connected component was used for further analysis.

Antigen-induced arthritis mouse model

Antigen-induced arthritis (AIA) was triggered in six 8-week-old, anaesthetised female 129/Sve mice by intra-articular injection of methylated bovine serum (mBSA) in the left knee joint after having presensitised mice to mBSA. The left knee joints of four naïve mice were injected with phosphate-buffered saline (PBS, 20 µL) and used as a negative control. One week after intra-articular triggering of AIA, mice were euthanized with carbon dioxide, and joints were either used for immunohistochemistry or scRNA-seq. Histochemical preparation was performed as previously described [6], and specimens were examined in a blinded manner for pannus formation, cartilage and subchondral bone destruction, and synovial hypertrophy on an arbitrary scale, 0–3, as described by Magnusson et al. [28]. For the scRNA-seq experiment, joint tissue was minced to ~1-mm³ pieces, which were digested by collagenase IV (1.5 mg/mL) and DNase I (100 µg/mL) at 37 °C. Dissociated cells were passed through a 70-µm cell strainer. Single-cell suspensions were resuspended in RPMI-1640 at a density of 1 × 10⁵ cells/mL for cell loading. One mouse in which AIA had been triggered developed only mild arthritis (arthritis score 0.5) and was therefore excluded from further analysis. All mice were housed with a 12-h light/dark cycle and fed standard chow diets in a specific pathogen-free animal facility at Linköping University.

scRNA sequencing was performed using the Seq-Well technique [29] following a described protocol [6]. Briefly, prepared single-cell suspensions were counted and coloaded with barcoded and functionalised oligo-dT beads (Chemgenes, Wilmington, MA, USA, cat. No. MACOSKO-2011–10) on microwell arrays synthesised as described by Gierahn et al. [29]. For each sample, 20,000 live cells were loaded per array to bind with oligo-dT beads. Beads were collected for capturing mRNA and preparing the library following cell lysis, hybridisation, reverse transcription and transcriptome amplification. Except for one library, which was sequenced alone, libraries from three samples were pooled for sequencing (Additional file 2: Table S1), resulting in a coverage of 6.6 reads per base. Four libraries were prepared for each sample using the Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA; cat. No. FC-131–1096) according to the manufacturer's instructions. Each library was sequenced once, except for one library, which was sequenced twice using the NextSeq 500/550 system.

The single-cell data were processed into digital gene expression matrices following James Nemesh, McCarroll's lab Drop-seq Core Computational Protocol (version

1.0.1, <http://mccarrolllab.com>) using *bcl2fastq* Conversion [30] and Picard software [31]. To increase the read depth for the cells, each sample was sequenced multiple times (Additional file 2: Table S1), and the fastq files for each sample were merged before further alignment steps. The indexed reference for alignment of reads was generated from GRCm38 (June 2017, Ensembl) using STAR software (2.5.3) [32]. Only primary alignments towards the reference genome were considered during downstream analyses, according to the mapping quality using STAR software.

Sampling and sequencing of psoriatic arthritis patients

Sampling

Psoriatic arthritis (PsA) patients and controls were recruited by the Immune-Mediated Inflammatory Diseases Consortium (IMIDC) [33]. PsA patients were recruited from different rheumatology departments from university hospitals belonging to the IMIDC. All PsA patients were diagnosed according to the CASPAR diagnostic criteria for PsA [34] with >1 year of disease evolution and >18 years old at the time of recruitment. Exclusion criteria for PsA included the presence of any other form of inflammatory arthritis, rheumatoid factor levels greater than twice the normality threshold or confirmed presence of an inflammatory bowel disease. PBMCs were sampled prior to treatment with anti-TNF or anti-IL17 and cryopreserved. Treatment response was classified at week 12 according to the EULAR response [35] (Additional file 7). For the anti-TNF study, 6 males and 10 females were included. The corresponding figures for anti-IL-17 treatment were 3 males (2 responders) and 13 females (6 responders). Simultaneously, healthy age- and sex-matched control subjects (Additional file 7) were recruited from healthy volunteers recruited through the Vall d'Hebron University Hospital in Barcelona (Spain). All the controls were screened for the presence of any autoimmune disorder, as well as for first-degree family occurrence of autoimmune diseases. None were found to be positive. Four males and four females were included.

Cell thawing

PBMCs cryopreserved at –80 °C were thawed in a 37 °C water bath and transferred with a bored tip to a 15-ml Falcon tube containing 14 ml of 37 °C prewarmed RPMI medium supplemented with 10% FBS (Thermo Fisher Scientific). Samples were centrifuged at 300 × g for 10 min at (room temperature) RT, the supernatant was removed, and pellets were resuspended in 1 ml of 1 × PBS (Thermo Fisher Scientific) supplemented with 1% BSA (PN 130–091-376, Miltenyi Biotec) and 10 µL of DNase I (PN LS002007, Worthington-Biochem). After incubation at RT for 10 min with periodic shaking, the cells

were filtered with a 20- μ m strainer (PN 43–10,020-70, Cell Strainer) into a new 15-ml falcon on ice, and the filter was washed by adding 9 mL of cold 1 \times PBS. Samples were concentrated afterwards by centrifugation at 300 \times g for 10 min at 4°C and resuspended in 1 \times PBS with 0.05% BSA for further assessment of cell numbers and viability with the TC20™ Automated Cell Counter (Bio-Rad). Samples balanced by responders and nonresponders for each treatment were mixed in pools of 8 patients at a 50:50 ratio and concentrated by centrifugation in an appropriate volume of 1 \times PBS-0.05% BSA to obtain a final cell concentration >4000 cells/ μ L, suitable for 10 \times Genomics scRNA-sequencing. The suspension was filtered again with a 20- μ m strainer, and the cell concentration was verified by counting with the TC20™ Automated Cell Counter.

Cell encapsulation and library preparation

Cells were partitioned into Gel BeadInEmulsions (GEMs) by using the Chromium Controller system (10 \times Genomics). Each pooled sample was loaded into two channels with a target recovery of 35,000 cells per channel to ensure a minimum final recovery of 2000 cells per sample condition. After GEM-RT incubation, the resulting cDNAs were purified with SPRI beads. To ensure maximal cDNA recovery, a second Sylane bead purification was performed on the supernatant from the first purification, and both products were eluted together and preamplified for 13 cycles, following the 10 \times Genomics protocol. cDNA was quantified on an Agilent Bioanalyzer High Sensitivity chip (Agilent Technologies), and 100 ng was used for library preparation. Gene Expression (GEX) libraries were indexed with 13 cycles of amplification using the Dual Index Plate TT Set A (10 \times Genomics; PN-3000431). The size distribution and concentration of full-length GEX libraries were verified on an Agilent Bioanalyzer High Sensitivity chip. Finally, sequencing of GEX libraries was carried out on a NovaSeq 6000 sequencer (Illumina) using the following sequencing conditions: 28 bp (Read 1) + 10 bp (i7 index) + 10 bp (i5 index) + 90 bp (Read 2) to obtain approximately >20,000 paired-end reads per cell.

3' single-cell RNA sequencing (scRNA-seq)

PBMC samples from 32 patients and 8 healthy controls were evenly mixed in pools of 8 donors per library following a multiplexing approach based on donor genotype, as in Kang et al. [36], for a more cost- and time-efficient strategy. Importantly, libraries were designed to pool samples together from the same treatment (anti-TNF or anti-IL17) but mixing patients with a different response to treatment. With this approach, we aimed to avoid technical artefacts that could mask subtle biological

differences between responders and nonresponders. To profile the cellular transcriptome, we processed the sequencing reads with 10X Genomics Inc. software package CellRanger v6.1.1 and mapped them against the human GRCh38 reference genome.

Library demultiplexing

The donor's genotypes (VCF format) were simplified by removing SNPs that were unannotated or located in the sexual Y, pseudoautosomal XY or mitochondrial chromosomes (chr 0, 24, 25 and 26, respectively). As genotypes were assembled using the human GRCh19 reference genome, we converted them to the same genome assembly used to map the sequencing reads, the human GRCh38 reference genome, using the USCS LiftOver (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>) command line executable. To meet the LiftOver required format (BED format), we used an available wrapper script (`liftOver_vcf.py`) to support input/output from VCF format [37]. The library demultiplexing by donor was performed with `cellsn-lite v1.2.2` in Mode 1a [37], which allows genotyping single-cell GEX libraries by piling up the expressed alleles based on a list of given SNPs. To do so, we used a list of 7.4 million common SNPs in the human population (MAF > 5%) published by the 10,000 Genome Project consortium and compiled by Huang et al. [37]. Importantly, we used the default parameters, setting the MAF > 5% (`-minMAF 0.05`) and requesting genotyping in addition to counting (`-genotype`). Then, we performed donor deconvolution with `vireo v0.5.6` [38], which assigns the deconvoluted samples to their donor identity using known genotypes while detecting doublets and unassigned cells. Finally, we discarded detected doublets and unassigned cells before moving on to the downstream processing steps.

scRNA-seq data sets and preprocessing

Below, we describe all of the scRNA-seq data sets and the preprocessing steps of the current work as outlined in Additional file 1: Fig. S1. As most preprocessing steps were applied to scRNA-seq expression matrices of all data sets, we will describe them jointly.

Quality cut-offs

Starting from a raw scRNA-seq gene expression matrix, the quality of cells was assured by application of quality cut-offs that aimed to filter out low-quality cells (few genes, low read depth), dying cells (high expression of mitochondrial genes) and doublets (unexpectedly high reads and large number of genes). While arbitrary, these specific cut-offs were adapted to the corresponding data

set and reported in Additional files 3, 4, 5, 6 and 7 [39]. Genes that were expressed in less than three cells were excluded from further analysis.

Batch correction

In case a high degree of interindividual expression differences existed, batch correction was performed according to a previously established pipeline [40]. In short, we used Seurat's function *findIntegrationAnchors()* [41] for the list of objects that corresponded to each individual. These anchors were later used by *IntegrateData()* [41] to integrate the data from individuals to correct for patient-specific differences as suggested in [42].

Denoising

Next, data for all cells were processed by a deep count autoencoder (DCA) model [43], which is a neural network performing a nonlinear principal component analysis (PCA). The DCA method is initiated by computing a library size, log- and z-score normalised expression matrix, which is taken as an input to the neural network, and the output of the neural network log₁₀ transformed), and denoised single-cell expression matrix *D*, which has the same features as the original data but is corrected for various sources of noise in the data. The DCA method also outputs a representation of the original single-cell data in a latent space. This representation has many fewer features than the original data, which is particularly important for performing accurate cluster analyses. The intercellular expression differences are generally better represented in this latent space than in purely linear PCA models, and the latent space representation is also corrected for single-cell data artefacts such as dropouts and varying library sizes.

Clustering analysis was performed using the Seurat v3.1 package [41] on the DCA-derived latent representation. A shared nearest neighbour graph was constructed, and neighbourhood overlap between every cell and its *k*-nearest neighbours was calculated based on the Jaccard index using the *FindNeighbors()* function on all supplied latent features. Next, clusters were identified through application of the Louvain algorithm to the shared nearest neighbour graph using the *FindClusters()* function along with a specified resolution setting. The resolution parameter and *k* for *k*-nearest neighbour analysis were tailored to each data set and are reported below. Clusters were visualised through *RunTSNE()*.

Analyses of interindividual molecular heterogeneity

After denoising and clustering, heterogeneity among samples was determined. For this, we trained a flexible machine learning model that attempts to find a decision boundary between the given groups of cells. If this

model results in a high misclassification rate for test data, it indicates that groups are highly mixed. More specifically, the data were randomly divided in half and used for training and testing the model. A random forest classifier [44] was used to classify cells from sick samples based on what patient the sample was derived from. Cross-validation with tenfold and grid search [45] was used to find the most appropriate hyperparameters of the random forest. The bootstrap [46] percentile method [46] was used to construct the 95% confidence intervals for training and test misclassification rates. The Scikit package [47] from Python (3.7.9) was used to perform the analysis.

Furthermore, patient heterogeneity was explored through comparison of cell type proportions and examination of latent features of the non-batch-corrected data. Interindividual differences in cell type proportions were explored by the application of the chi-square test to the proportions of cell types in sick samples. The contingency table for the chi-square test included individuals in columns and cell types in rows. Latent feature comparison was conducted visually through tSNE visualisation of the latent features of each patient. The results for these analyses are found in Additional file 2.

Cell typing

While cell typing is not crucial for scDrugPrio (which might be performed on unlabelled clusters), we cell typed clusters to enhance biological interpretation. Cell types were assigned to each cluster based on the relative coexpression of several known cell type marker genes (Additional file 3). For AIA data, each gene's expression was expressed as a fraction of a cell's total gene expression score. For visualisation of gene expression differences between clusters, z-scores were calculated. Z-scores for single cells were derived by comparison of one cell's gene fraction to all other cells' gene fractions. Z-scores for gene expression of clusters were derived by comparison of the average gene fraction in a cluster to the cluster-averaged gene fractions of cells in other clusters. Murine cell type-specific marker genes for the RA data were derived from the online resources of the R&D systems (www.rndsystems.com/research-area; accessed July 2020). Cell typing of the human data sets was performed using DCA denoised gene fractions and utilised combinations of marker genes (Additional files 4 & 5).

Differentially expressed genes

For each cell type separately, differentially expressed genes (DEGs) were calculated by comparing denoised gene expression of cells derived from healthy samples vs. cells from sick samples. For this purpose, the *FindMarkers()* function in Seurat [41] was used to deploy a scRNA-seq-tailored hurdle model supplied by the MAST

package [48]. Genes were considered significantly differentially expressed when they showed an absolute log fold change greater than or equal to 1.5 and a Bonferroni-adjusted $P < 0.05$. The fold change cut-off was motivated by previous studies [49] and aimed to decrease the number of DEGs for later network calculations.

Antigen-induced arthritis

Quality cut-offs resulted in a total of 16,751 cells (Additional file 3). Genes were annotated as murine NCBI Gene Symbols. Following denoising, clustering using $k=20$ and a resolution of 0.6 resulted in the identification of 20 clusters that were cell typed. Heterogeneity analysis used $n_estimators=1500$, $max_depth=15$, $min_samples_split=50$, and $min_samples_leaf=50$ and showed no significant heterogeneity. DEGs and the denoised expression matrix were translated to human Entrez gene IDs using human-mouse orthologues downloaded from NCBI (August 2019).

Multiple sclerosis

A unique molecular identifier (UMI) matrix [50] for cerebrospinal fluid (CSF) of five human multiple sclerosis (MS) patients and five human patients with idiopathic intracerebral hypertension (IIH) were downloaded from Gene Expression Omnibus (GEO), GSE138266 [51]. Gene annotation was translated from human Ensembl gene IDs to human Entrez gene IDs and symbols using the HUGO Gene Nomenclature Committee (HGNC) database [52] (downloaded November 2020). After the application of quality cut-offs (Additional file 4), we derived 33,848 cells. Initial preprocessing was performed without batch correction using cluster parameters $k=10$ and $resolution=0.2$ after DCA denoising to derive 17 clusters. Interindividual heterogeneity was assessed as described below using the following hyperparameters: $n_estimators=500$, $max_depth=30$, $min_samples_split=50$, $min_samples_leaf=25$. Since we noticed substantial patient-related heterogeneity, preprocessing was repeated, including batch correction, DCA denoising, and clustering using $k=15$ and $resolution=0.35$ to derive 21 clusters that were cell typed using known marker genes derived from the original publication [50] (Additional file 4). The number of DEGs ranged from 0 to 10,076.

Crohn's disease

A unique molecular identifier (UMI) matrix [53] for eleven human Crohn's disease (CD) patients was downloaded from GEO (GSE134809) [54]. Data for each patient included intestinal biopsies from one inflamed site and one uninflamed site. After application of quality cut-offs (Additional file 5), we derived 77,416 cells. Gene

annotation was translated from human Ensembl gene IDs to human Entrez gene IDs and symbols using the HGNC database (2020–11-08) [52]. Initially, data were DCA denoised without applying batch correction. Clustering was performed using $k=15$ and $resolution=0.8$. Interindividual molecular heterogeneity was assessed as described below using the following hyperparameters: $n_estimators=1000$, $max_depth=20$, $min_samples_split=50$, $min_samples_leaf=25$. Since there was substantial interindividual heterogeneity, preprocessing was repeated now batch-correcting before DCA denoising. Clustering was again performed using $k=15$ and $resolution=0.8$.

Individual Crohn's patients

For individual patient predictions, we used the same quality cut-offs as for the pooled analysis of CD patients. As interindividual heterogeneity does not affect the predictions made for individual patients, these calculations were performed on non-batch-corrected data. DCA denoising was applied to the joint data, and gene annotation was translated to Entrez gene IDs. Thereafter, scRNA-seq data were separated by patient, and cells from each patient were clustered individually. An individual patient cluster was assigned a cell type based on which cluster it most resembled in the joined CD analysis, as measured by the number of shared cell identifiers. DEGs were then calculated between cells from sick and healthy samples. Visualisations of data were in part created using BioRender.com.

Psoriatic arthritis

Data sets were divided into one anti-TNF and one anti-IL17 data set, including responders (R), nonresponders (NR) and healthy controls. We filtered out the doublet and unassigned cells as well as those that did not meet the quality cut-off criteria (Additional file 7) and derived 78,610 cells with 5088 mean reads for the anti-TNF data set and 72,472 cells with 5343 mean reads for the anti-IL17 data set. In both data sets, 19,415 cells were derived from healthy controls. Data sets were batch-corrected, and DCA was performed. For anti-TNF, clustering was performed using $k=25$ and $resolution=0.25$. The corresponding parameters for anti-IL17 were $k=15$ and $resolution=0.45$. The remaining downstream analysis was performed for responders and nonresponders separately, which meant that DEGs were calculated between responders and healthy controls and between nonresponders and healthy controls. For the anti-TNF data set, the number of DEGs ranged from 0 to 5989 for R and from 0 to 5877 for NR. The

corresponding figures for anti-IL17 were 0 to 3097 and 0 to 3284. Next, scDrugPrio was applied to DEGS from anti-TNF R and NR as well as anti-IL17 R and NR separately.

In vitro validation of potential novel drugs

To validate the predicted novel drugs, in vitro culture of murine and human B-cells upon activation with the indicated stimuli was employed to assess the effects of the predicted drugs on B-cell survival, activation, proliferation, and antibody production. Three doses for each predicted drug were used to challenge in vitro cultured B-cells (Additional file 2: Table S2). For the assessment of potential novel drugs on murine B-cell survival and activation, 300,000 murine naïve B-cells ($\text{Lin}^- \text{B220}^+ \text{CD43}^-$) were enriched by flow cytometric sorting and cultured in the presence of AffiniPure F(ab')₂ Fragment goat anti-mouse anti-IgM (10 µg/mL, CAT: 115–006-075, Jackson ImmunoResearch), anti-mouse CD40 (10 µg/mL, Clone:1C10, Biolegend), or LPS (10 µg/mL) for 24 h. B-cell survival was determined by flow cytometric analysis of propidium iodide (PI)⁺ cells. Surface CD69, CD86, and MHC-II were used as readouts for assaying B-cell activation. For the analysis of B-cell proliferation, purified B-cells were stained with carboxyfluorescein succinimidyl ester (CFSE) (1 µM) before in vitro culture for 3 days. To determine the effects of the predicted drugs on antibody production, 200,000 purified murine naïve B-cells were stimulated with anti-CD40 (10 µg/mL) + IL-4 (10 ng/mL), LPS (10 µg/mL) + IL-4 (10 ng/mL), or LPS (10 µg/mL) + IFN-γ (10 ng/mL) for 6 days.

For the analysis of novel drugs predicted to regulate the biology of human B-cells, peripheral blood mononuclear cells (PBMCs) were isolated from the buffy coat as previously described [55, 56]. Human naïve B-cells were subsequently pre-enriched by MACS sorting using a B-Cell Isolation Kit II (Miltenyi) and further purified by flow cytometric sorting of CD19⁺CD27⁻ B-cells. Purified human naïve B-cells were cultured in 96-well plates in the presence of AffiniPure F(ab')₂ Fragment goat anti-human IgG + IgM (5 µg/mL, CAT: 109–006-127, Jackson ImmunoResearch), anti-human CD40 (5 µg/mL, Clone: G28.5, Bio X Cell), and IL-21 (10 ng/mL, PeproTech).

B-cell survival was determined by flow cytometric analysis of propidium iodide (PI) +—cells. Surface CD69, CD86, and MHC-II were used as readouts for assaying murine B-cell activation. Surface CD69 was assayed for the measurement of human B-cell activation. For the analysis of B-cell proliferation, purified B-cells were prestained with carboxyfluorescein succinimidyl ester (CFSE) (1 µM) before in vitro culture for 3 days. Murine IgG2a and IgG1, as well as human IgG in the supernatant,

were determined by enzyme-linked immunosorbent assay (ELISA) using goat anti-mouse Ig, goat-anti-mouse IgG1-HRP and goat-anti-human IgG2a-HRP, goat anti-human Ig, and goat-anti-human IgG-HRP (SouthernBiotech) as previously described [57].

In vivo validation of predicted drugs

Amrinone was tested for treating collagen-induced arthritis (CIA). For this, male DBA1/J mice purchased from GemPharmatech (China) were immunised intradermally with 100 µg of chicken type II collagen (2 mg/mL, Chondrex, USA) emulsified with complete Freund's adjuvant (CFA, 1 mg/mL) and boosted on day 21 with 100 µg of chicken type II collagen emulsified with incomplete Freund's adjuvant (IFA). Mice were i.g. given with diluent ($n=5$) or amrinone (30 mg/kg, $n=5$) daily from day 21 for 3 weeks. The rear paw thickness and the clinical arthritis score for each limb were recorded every other day from 0 to 4 with a maximal score of 16 for each mouse according to the previous protocol [58]. Mice were housed with a 12-h light/dark cycle and fed standard chow diets in a specific pathogen-free animal facility at Xuzhou Medical University.

Mice were euthanized on day 21 post drug intervention with carbon dioxide. Serum was collected for the analysis of collagen-specific autoantibodies by enzyme-linked immunosorbent assay (ELISA) as previously described [59]. Briefly, diluted serum was incubated in a 96-well ELISA plate precoated with chicken type II collagen (5 µg/mL). Goat-anti-mouse IgG1-HRP, goat-anti-mouse IgG2a-HRP, and goat-anti-mouse IgG-HRP (SouthernBiotech) were used as detection antibodies. Knee joints were fixed in 4% formaldehyde and subsequently decalcified with decalcification solution (ServiceBio, China) for 1 week. The specimens were next embedded in paraffin, and sagittal Sects. (4 µm) were cut. The sections were stained with haematoxylin and eosin (H&E) for the histological analysis of immune cell infiltration and Safranin-O for the analysis of bone erosion as previously described [57, 59].

Comparison of scRNA-seq-based screening outcomes for rheumatoid arthritis to other data types and prediction methods

Briefly, we benchmarked outcomes based on the scRNA-seq-derived DEGs against microarray data from GEO (GSE55235, <https://ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE55235>; GSE93272, <https://ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE93272>) [60–63], GWAS Catalog [64] genes and OMIM [65] genes as well as combinations of these data sets. scDrugPrio was compared to previous methods such as [1] identifying druggable

DEGs, targeting key enriched pathways [66], CMAP [67] drug predictions and the empirical drug selection of Kim et al. [7]. Predictions were also replicated using the smaller, unbiased HuRI PPIN [68] (8236 proteins, 52,150 interactions) to ensure the absence of knowledge bias. More information can be found in Additional file 2.

Results

The scDrugPrio framework

Aiming to create an analytical framework for scRNA-seq-driven drug prioritisation, we constructed scDrugPrio, which consists of three main modules: (1) drug candidate selection based on cell type-specific DEGs, (2) drug candidate ranking and (3) aggregated ranking of drug candidates from all cell types. For this scDrugPrio requires two components (Fig. 1): (a) differentially expressed genes (DEGs) between sick and healthy samples for each cell type and (b) drug data, including information on gene drug targets and pharmacological effects.

Preprocessing includes the calculation of DEGs based on scRNA expression from one or more healthy and one or more sick samples (Additional file 1: Fig. S1). Preprocessing starts with quality control, batch correction (if needed) and data denoising of all scRNA-seq data, followed by clustering and cell typing. DEGs are computed per cell type by using sick versus healthy expressions. In the analysis below, we computed DEGs in the two modes. In the first mode, we use data from groups of sick and healthy individuals, while in the second mode, we use scRNA-seq data from paired samples of sick and adjacent healthy tissue for personalised drug prioritisations for one patient of interest.

scDrugPrio starts by computing the mean closest distance between DEGs and gene targets of drugs (henceforth referred to as drug targets) in the protein–protein interaction network (PPIN) for each cell type and each drug candidate. Intuitively, the “closer” the drug targets are to DEGs, the better is the chance for the drug to affect the disease-associated genes [16]. Specifically, a relative proximity measure (z_c) capturing the statistical significance of the observed closest distance (d_c) was calculated based on a comparison of d_c to the random expectation. Furthermore, scDrugPrio assumed that drugs that counteract the fold-change of at least one DEG will have a better chance to reverse disease-associated expression changes in the targeted cell type (following a similar idea as CMap [17]). To determine whether a drug counteracted a DEG fold change, we considered (1) the direction of the fold change (upregulated or downregulated) and (2) pharmacological action (e.g. agonistic, antagonistic) on the targeted DEG. For each cell type, a list of drug candidates was derived by filtering out drugs with low network proximity ($d_c \geq 1$, $z_c \geq -1.64$ corresponding

to one-sided $P < 0.05$) and drugs that did not exert counteracting pharmacological action against at least one targeted DEG.

In the next step, scDrugPrio computed intra- and intercellular centrality measures that were later used to aggregate the prediction and rank drugs. To compute intracellular centralities per cell type, scDrugPrio determines the largest connected component formed by DEGs in the PPIN and next computes a drug’s centrality score based on the centrality of the drug targets within this component. Intracellular centrality hence presents an approximation of a drug’s target relevance for disease-associated expression changes in a cell type. For intercellular centrality, scDrugPrio constructs a multicellular disease model (MCDM). In short, MCDMs were based on predicted molecular interactions between differentially expressed upstream regulatory genes in any cell type and their downstream genes in any other cell type using NicheNet [24]. The resulting MCDM is a network in which cell types were nodes connected by directed, weighted ligand-target interactions. Intercellular centrality refers to the centrality of cell types in the MCDM.

Finally, scDrugPrio aggregates drug predictions through aggregation of cell type-specific drug predictions and intracellular and intercellular centralities into a compound score (Fig. 1). The resulting scores were used to rank drugs. Hence, overall drug ranking prioritised drugs that targeted key disease-associated expression changes in the most important cell types.

scDrugPrio development and evaluation in antigen-induced arthritis

scRNA-seq data were generated from whole joints of five inbred AIA mice and four naïve mice. After application of quality criteria (Additional file 3), data included 16,751 cells with 132,459 mean reads per cell. Data for all mice were denoised jointly using a deep count autoencoder network (DCA) [43], and 20 clusters were cell typed using marker gene expression (Fig. 2a; Additional file 1: Fig. S2). Comparison of cells from AIA and naïve mice identified DEGs in 17 cell types (Fig. 2; Additional file 2).

We retrieved drug target information from DrugBank [19] for 13,339 drugs (Additional file 3). From those drugs, we selected all of which had been FDA approved for human use and had at least one target in the human interactome ($n=1840$). According to the indication information in DrugBank, 57 drugs were FDA approved for RA and hence treated as true positives for the calculation of precision and recall.

Applying scDrugPrio, the final list of candidates included 334 out of 1840 drugs; 32 drugs were established RA drugs, of which 22 ranked among the top 100 candidates (Fisher exact $P < 10^{-6}$; Fig. 3; Additional

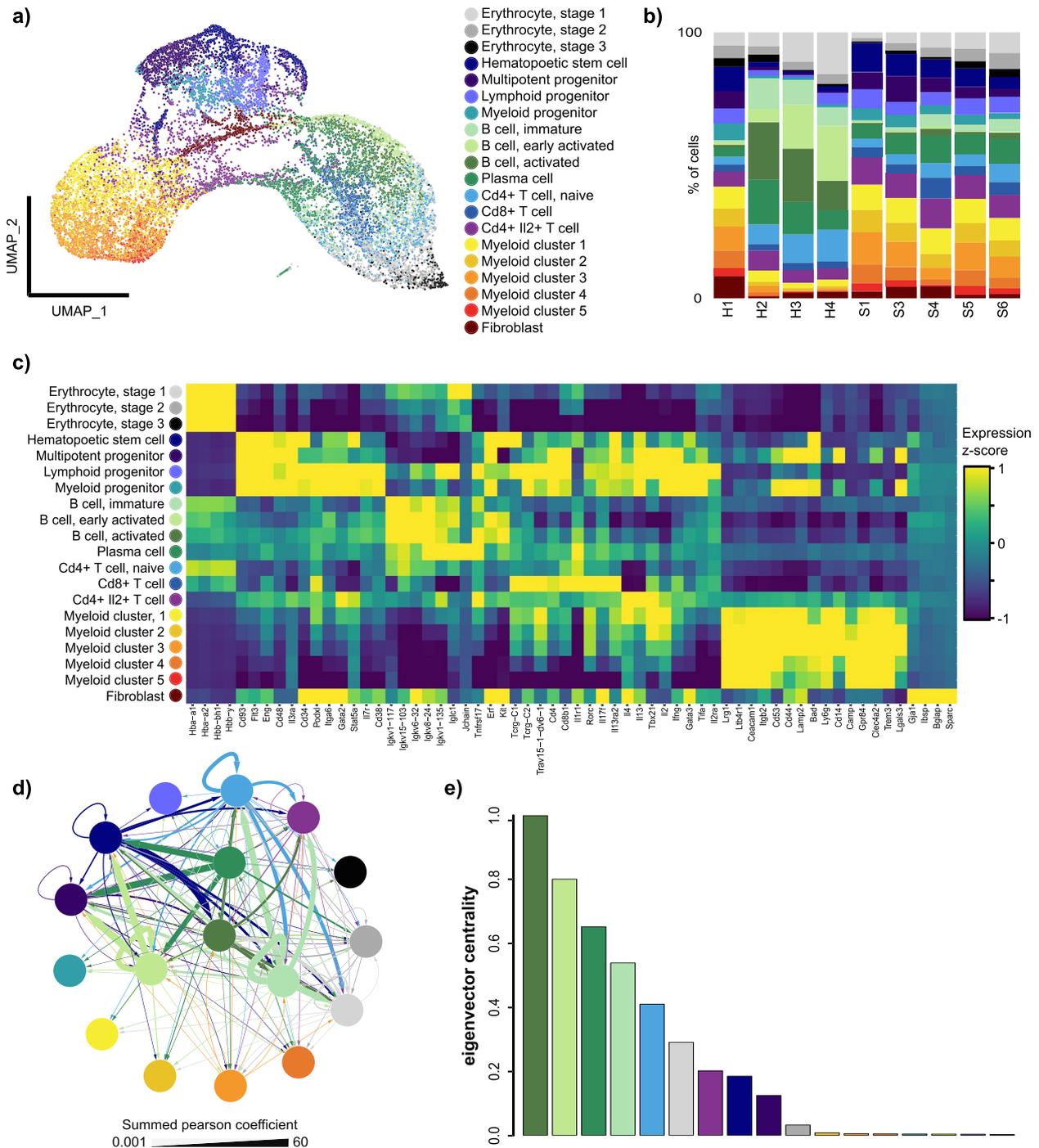


Fig. 2 Construction and analysis of a multicellular disease model (MCDM) based on scRNA-seq analysis of a mouse model of antigen-induced arthritis (AIA). **a** tSNE of cells ($n = 16,751$) pooled from all samples. **b** Cell type proportions (%) for individual AIA (S1 to S6) and control (H1 to H4) mice. **c** Heatmap of z-scores of the cluster-based means of normalised, denoised gene expression for cell type markers. **d** MCDM, in which cell types were represented by nodes connected by the predicted interactions between upstream regulatory genes in any cell type and their downstream target genes in any other cell type [24]. Edge width corresponds to the sum of Pearson coefficients for all interactions between two cell types, with arrows directed from upstream to downstream cell types. Edge colour corresponds to the upstream cell type. As indicated by more central positions in the MCDM, the four B-cell clusters were the most central cell types. **e** Bar plot depicting eigenvector cell type centrality in MCDM. Colours in all plots correspond to cell type clusters in (a)

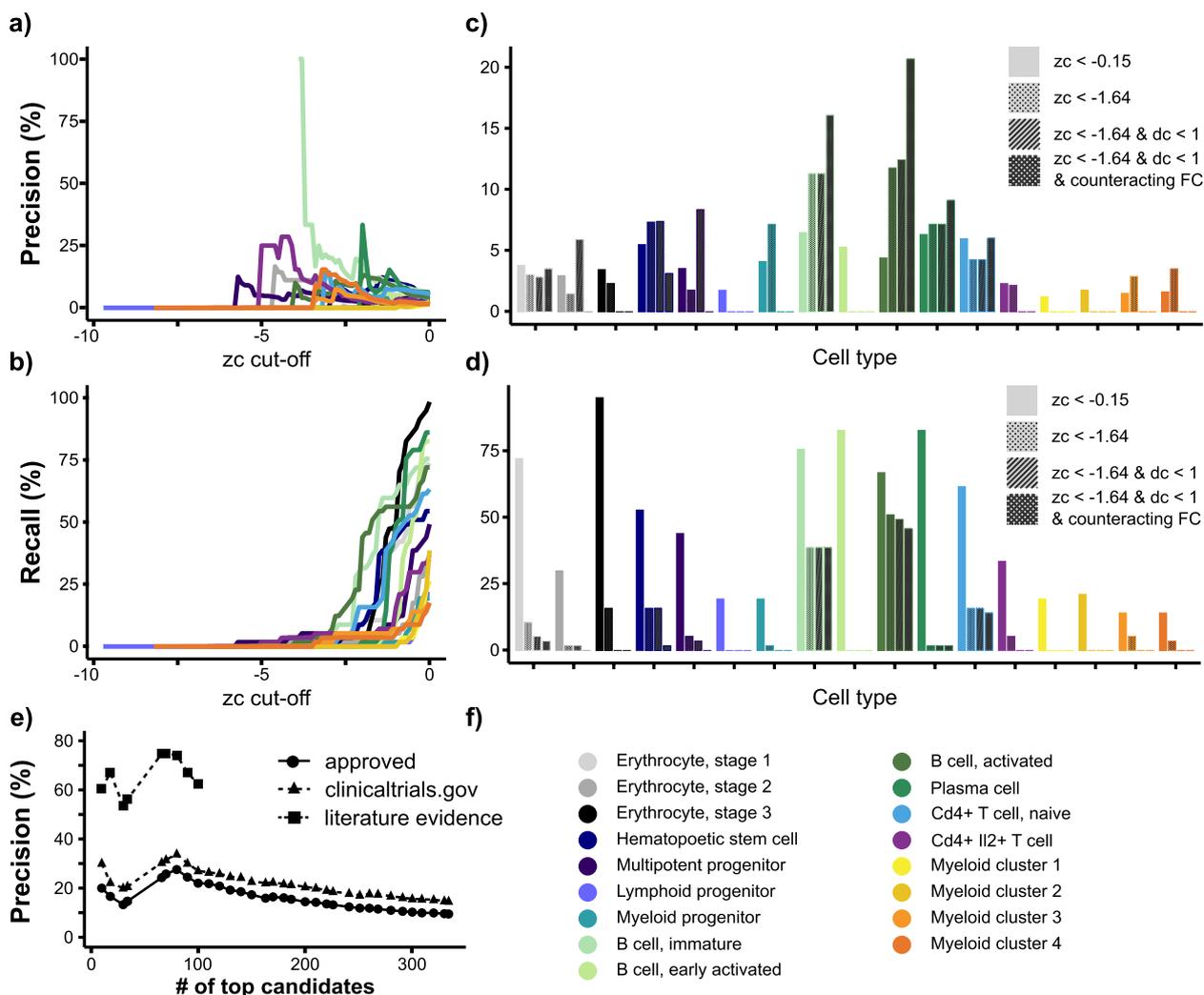


Fig. 3 Precision and recall in relation to drug selection and ranking criteria. **a, b** Precision and recall at different z_c cut-offs. In most cell types, precision increased with decreasing z_c . Too stringent z_c cut-offs, by contrast, led to the exclusion of almost all candidates, including approved RA drugs. **c, d** Precision and recall after stepwise application of drug selection criteria. Bars with no pattern represent a z_c cut-off that Guney et al.¹⁰ had previously found to offer good coverage of known drug-disease pairs. The dotted pattern represents precision at $z_c < -1.64$ (corresponding to one-sided $P < 0.05$). The striped pattern represents precision among candidates with $z_c < -1.64$ and $d_c < 1$ (in other words, candidates passing network proximity-based selection). The crosshatch pattern represents candidates passing network proximity and pharmacological action selection. The application of selection criteria substantially increased the precision among central cell types. **e** Precision for approved RA drugs for which clinical trials had been registered for RA, and literature evidence (only for the top 100 ranking drugs) among candidates. Drug ranking included rank ties. Literature evidence was gathered for the top 100 ranked candidates and is presented as triangles with a dashed line. The colour legend for (a–d) is depicted in (f)

file 3). To further evaluate the candidates' relevance, we collected clinical trial data from clinicaltrials.gov to capture the medical community's interest in the identified candidates as RA medications. We also performed a literature review of the top 100 drugs to evaluate whether candidates had shown promise when tested in human RA or murine/rat RA models. Through a literature review, evidence for the relevance of 40 additional drugs was identified, whereas three of the top 100 ranking drugs

had not shown effects in previous trials. Hence, 62.0% of the top 100 ranking candidates were either approved or had successful experimental validation in prior literature (Fig. 3e), and 95.4% of previously studied candidates (among the top 100 candidates) had shown promise.

Describing scDrugPrio results in more detail, network proximity selection ($d_c < 1$, $z_c < -1.64$) yielded no drug candidates for six cell types and an average of 67 drug candidates (min=1, max=226) for the remaining

fourteen cell types. Precision among cell type-specific candidates ranged from 0.0 to 12.8%. However, precision of cell types that were central in the MCDM, such as activated, immature, and plasma B-cells (12.8, 11.4, and 6.3%, respectively), outperformed random selection from the 1840 included drugs ($57/1,840=3.1\%$; Fisher exact $P < 10^{-12}$, $P < 10^{-8}$, $P=0.4$, respectively). Following network proximity calculations that were performed in the absence of information on a drug's effect on the target, we found tasonermin, a synthetic version of TNF, among the top-ranking candidates for several B-cell subtypes. Since overexpression of TNF has a crucial role in RA pathogenesis [69], this ranking supported the network proximity criterion. However, because tasonermin mimics the effects of TNF, it could worsen the disease. This finding exemplified the importance of pharmacological action selection. Pharmacological action selection resulted in removal of all drug candidates in two cell types and an average of 43 drug candidates per cell type (min=2, max=137) in the remaining twelve cell types. Overall precision decreased to a median [range] of 1.56% [0.00–20.63], although it increased in activated, immature, and plasma B-cells (20.6, 16.1, and 9.1%, respectively) as well as in T cells (Fig. 3c).

Having identified drug candidates for every cell type individually, intra- and intercellular centrality were calculated. The use of intra- and intercellular centrality measures for composite ranking was motivated by our findings that (1) intercellular centrality correlated with the significance of GWAS enrichment among cell type-specific DEGs (Pearson's r [95% CI]=0.62 [0.21–0.85]; $P < 0.01$), (2) intercellular centrality correlated with the precision among cell type-specific drug candidates (Pearson's r [95% CI]=0.77 [0.46–0.91]; $P < 10^{-3}$; Additional file 1: Fig. S2j), (3) drugs that targeted more than one cell type were more likely to be known RA drugs (Additional file 1: Fig. S2k) and (4) intracellular centrality could improve the mean rank of known drugs more than expected by chance (Additional file 2).

Using the AIA and human RA data, we benchmarked scDrugPrio against previous methodologies (Additional file 2) and performed extensive additional testing, demonstrating the advantage of scRNA-seq-driven analysis over genetic variations or bulk transcriptomics. Furthermore, we evaluated drug selection criteria and performed robustness analysis (Additional file 2).

Experimental validation of scDrugPrio

To further validate scDrugPrio, five high-ranking drugs that were not found to have any prior literature support for their efficacy in RA were chosen alongside the highest-ranking RA drug (auranofin #6) serving as a positive control. We first evaluated the five drugs by in vitro

studies of B-cells. This cell type was selected because of its crucial role in the pathogenesis of RA and its central position in MCDM [70]. We used previously described in vitro models of B-cell functions [70] measuring murine B-cell survival, activation, proliferation, and antibody production upon in vitro stimulation with selected drugs at various concentrations. Auranofin dramatically suppressed B-cell functions, including cell viability, proliferation, and immunoglobulin production (Fig. 4). Additionally, two of the five novel drugs (#114 amrinone and #100 adapalene) showed concentration-dependent in vitro effects on B-cell viability, proliferation, and immunoglobulin production (Fig. 4). Adapalene also greatly inhibited murine B-cell activation (Fig. 4b). The other three drugs (#91 irbesartan, #98 isosorbide, and #134 dimethyl furamate) showed little to no effect on murine B-cell functions (Additional file 1: Fig. S4). Next, we examined the effects of drug candidates on the function of human B-cells. Similarly, auranofin, adapalene, and amrinone inhibited human B-cell viability, activation, proliferation, and IgG production (Additional file 1: Fig. S5). Thus, two out of five candidate drugs prioritised by scDrugPrio, namely, adapalene and amrinone, were successfully validated by in vitro studies.

Valid drug candidates should arguably be transposable and replicable. As drug prediction was performed using data from the AIA model of arthritis, we deployed the collagen-induced arthritis (CIA) model for further in vivo study. We selected only amrinone for further study, as adapalene was designed for topical skin use and not for systemic delivery. CIA mice were administered 30 mg/kg amrinone i.g. for 3 weeks. This treatment significantly reduced the paw thickness (Fig. 4e) and clinical scores of CIA mice (Fig. 4f). Analysis of collagen-specific serum autoantibodies revealed a significant inhibitory effect (Fig. 4g). Considering that it even reduced immune cell infiltration (Fig. 4h) and bone erosion (Fig. 4i), in vivo studies confirmed the relevance of amrinone treatment and thereby further supported scDrugPrio.

Application of scDrugPrio to multiple sclerosis

We next applied scDrugPrio to human IMIDs using scRNA-seq data of cerebrospinal fluid (CSF) from multiple sclerosis (MS) patients and idiopathic intracranial hypertension that served as controls [50]. After application of quality cut-offs (Additional file 4), the data included 33,848 cells with 47,332 mean reads per cell. Comparing MS samples with controls, DEGs were calculated from batch-corrected, normalised expression scores. DEGs were found in 19 of 20 cell types. scDrugPrio identified on average 59 (min=1, max=270) drugs in these 19 cell types (Fig. 5; Additional file 1: Fig. S6). Aggregated ranking of 417 drug candidates (Additional

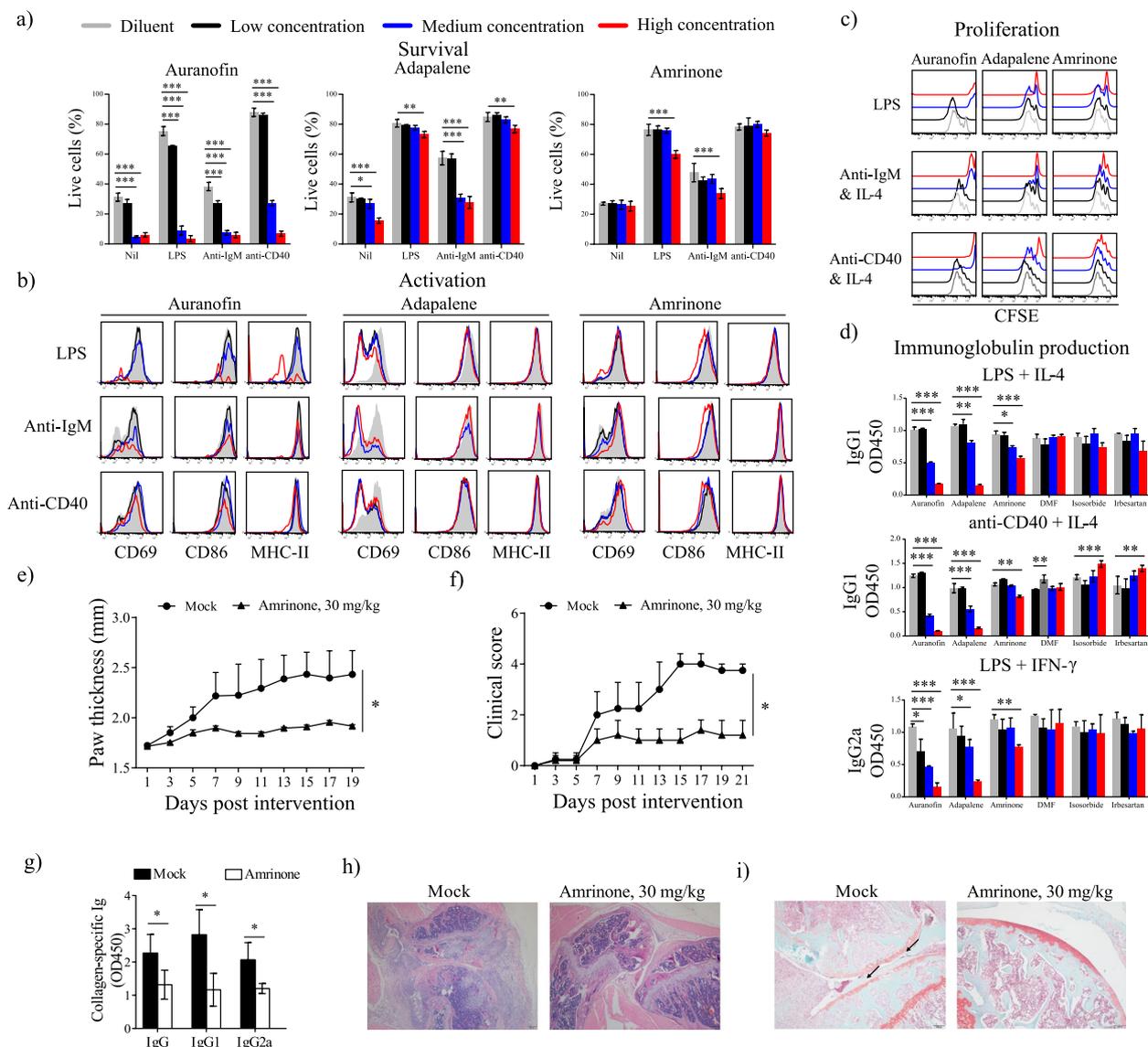


Fig. 4 Experimental validation of amrinone in suppressing murine B-cell function and the pathogenesis of the CIA mouse model. Drug effect of the selected drugs and the positive control on in vitro murine (a) B-cell survival, (b) activation, (c) proliferation, and (d) immunoglobulin production. Colours represent the responses to diluent and different drug concentrations (low, medium, high); for specific concentrations, see Additional file 2: Table S2. Having successfully validated adapalene and amrinone in vitro, we conducted in vivo experiments for amrinone. Mice with collagen-induced arthritis (CIA) were treated with diluent ($n=5$) or amrinone (30 mg/kg, $n=5$) for 3 weeks. The (e) rear paw thickness, (f) clinical scores, and (g) collagen-specific serum autoantibodies were measured. Furthermore, drug efficacy was assessed by analysis of (h) joint immune cell infiltration using H&E staining and (i) bone erosion using Safranin-O staining. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Dimethyl furamatate, DMF

file 1: Fig. S7; Additional file 4) included ten out of 17 approved MS drugs. Among the top 100 ranking candidates, four approved MS drugs, as well as biosimilars rituximab and obinutuzumab, were identified, which outperformed random expectation (precision 0.9%). A literature search revealed that an additional 22 of the top 100 ranking drugs had shown effects in previous studies, resulting in a precision of 28.3% (Fig. 5d). Seventy-one

drugs had not yet been validated in previous studies, and one evaluated drug had not shown promise in previous studies. Hence, 96.7% of the studied candidates (among the top 100 candidates) showed efficacy. Taken together, these data supported the potential of scDrugPrio to predict the response to drugs approved for that disease, as well as for repurposing other drugs.

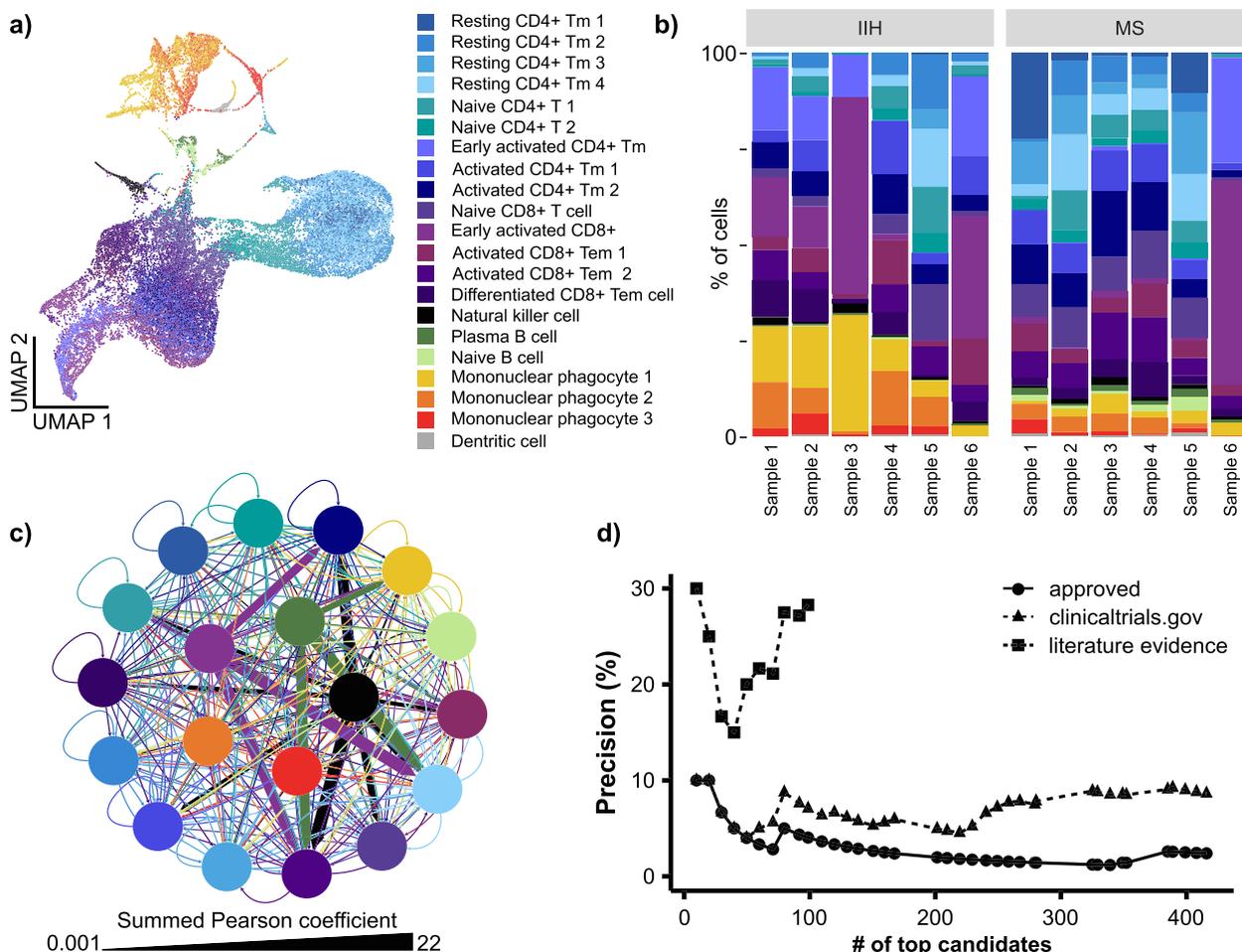


Fig. 5 scDrugPrio applied to scRNA-seq data from cerebrospinal fluid samples of MS patients and controls. **a, b** Clustering and cell type proportions (%) for CSF samples of idiopathic intracranial hypertension (IIH) and multiple sclerosis patients (MS). Cell type proportions showed significant interindividual differences among MS patients. **c** The MCDM for MS was more complex than that of AIA mice. The most central cells included mononuclear phagocyte 3, natural killer cells, mononuclear phagocyte 2, early activated CD8+T cells and plasma B-cells (in order of eigenvector centrality). **d** Precision among ranked candidates for approved MS drugs, drugs that had been registered for clinical trials for MS in the ClinicalTrial database, and literature evidence (only for top 100 ranking drugs). Abbreviations: Tm, T memory; Tem, T effector memory

The value of above rank aggregation was supported as precision among cell type-specific drug candidates only ranging from 0.0 to 12.5%. The MCDM (Fig. 5c; Additional file 4) indicated plasma cells, mononuclear phagocytes, natural killer cells, and activated CD8+T cells to be of central importance, which is consistent with the current pathophysiological understanding [50, 71]. Notably, the MS MCDM was more complex than the AIA MCDM regarding the number of cell types and interactions.

Application of scDrugPrio to Crohn's disease

Next, scDrugPrio was applied to patients with Crohn's disease (CD) from whom scRNA-seq data [53] from paired inflamed and uninfamed intestinal tissue biopsies

were available. After application of quality criteria (Additional file 5), 77,416 cells were included with 3591 mean unique molecular identifiers (UMIs) per cell. Following preprocessing, analysis was performed on batch-corrected, pooled data from all patients in which DEGs were calculated through comparison of inflamed and uninfamed samples (Additional file 1: Fig. S8 & S9; Additional file 5). The aggregated ranking included 343 drug candidates, of which five were known CD drugs. These five consisted of sulfasalazine (#91), mesalazine (#169.5), rifaximin (#238), and two anti-TNF drugs (adalimumab and infliximab, tied rank #292.5). Apart from sulfasalazine, literature evidence suggested the effectiveness of 13 additional top-ranking 100 drugs, resulting in a precision of 14%.

Patient heterogeneity in human disease

An important difference between human data and mouse data was substantial interindividual heterogeneity in human patients. To illuminate such differences, we (1) applied chi-square test to compare cell type proportions among sick samples, (2) trained a flexible machine learning model (random forest) with the aim to separate cells from different individuals based on gene expression profiles, and (3) examined latent features of the non-batch-corrected data. Mice, as expected, showed no differences in cell type proportions (chi-square $P=0.9919$; Fig. 2c), gene expression (random forest, 95% CI of misclassification rate 0.504–0.532 in the training data and 0.638–0.655 in the test data), or latent features (Additional file 1: Fig. S2c). However, both MS and CD patients showed great interindividual heterogeneity in cell type composition (chi-square $P < 10^{-15}$, respectively). Heterogeneity in cell type composition, although decreased, can still be observed in the batch-corrected data (Fig. 5b; Additional file 1: S7i). A patient effect was also observed in the non-batch-corrected gene expression for MS (random forest, 95% CI of misclassification rate 0.075–0.084 in training data, and 0.124–0.132 in testing data) and CD patients (random forest; 95% CI of misclassification rate 0.270–0.283 in training data, and 0.378–0.385 in testing data) as well as latent features derived from non-batch-corrected data (Additional file 1: Fig. S6 & S9). Taken together, such patient effects necessitated batch correction for pooled prediction using the human data sets above.

Potential for individualised predictions

Even though batch correction had been applied appropriately, a potential limitation of the above MS and CD analyses was that scDrugPrio was applied to pooled data derived from heterogeneous patients and controls. Patient effects might form the bases for the responder/nonresponder dichotomy, and we therefore evaluated scDrugPrios potential for individualised drug prioritisation. For this, we applied scDrugPrio to individual CD patients, using similar preprocessing as for CD data above with the following exceptions: (1) data were not batch-corrected and (2) following denoising, cells from inflamed and uninfamed samples of each patient were clustered separately. DEGs were derived through comparison of individual patient inflamed and uninfamed cells in each cluster. DEGs showed that the eleven patients expressed important CD drug targets differently (Additional file 1: Fig. S10a, b). To investigate whether such molecular differences could affect drug prediction outcomes, scDrugPrio was applied to all patients separately (Additional file 1: Fig. S10-12; Additional file 6).

Strikingly, individualised drug predictions of nine out of eleven patients (such as patient 1: 19.0%; patient 10:

20.5%, Fig. 6; Additional file 1: Fig. S13) outperformed the precision of pooled patient analysis (14.0%). Among the top 10 candidates, precisions for individualised predictions (20–70%) outperformed precision of pooled patient analysis in seven patients and equalled that of pooled patient analysis in four patients (10%). All predictions outperformed random chance (1.5%). More detailed analysis revealed interindividual differences in cell type proportions and network properties in the MCDM (Additional file 1: Fig. S12) as well as different drug rankings (Additional file 1: Fig. S14). Taken together, these findings supported that scDrugPrio presents a valid strategy for personalised drug prioritisation.

To exemplify the potential of scDrugPrio for individual patients, we next compared two patients who previously had been classified as an anti-TNF responder (patient 10) and nonresponder (patient 1) based on a cellular signature score [53]. In agreement with the previous classification of anti-TNF response [53], TNF had a more central role in the MCDM of patient 10 (Additional file 1: Fig. S11m, n). Hence, it was not surprising that aggregated drug ranking ranked adalimumab (anti-TNF) higher for patient 10 (#15.5) than for patient 1 (#658). As expected, adalimumab was the highest-ranking approved CD drug in patient 10. For patient 1, scDrugPrio prioritised other immunomodulatory drugs over anti-TNF treatment, namely, natalizumab (#19), human intravenous immune globulin (#21), basiliximab (#25), sarilumab (#29), and other approved CD drugs, such as methotrexate (#188) and sulfasalazine (#202).

Application of scDrugPrio to nonresponder/responder data from patients with psoriatic arthritis highlighted the importance of local tissue samples

Since the previous analyses supported scDrugPrio's potential for another case-of-use scenario, namely, to distinguish drug responders from nonresponders. To explore this potential, we collected peripheral mononuclear blood cells (PBMCs) from patients with psoriatic arthritis (PsA) as well as healthy controls. PBMCs were chosen because the analysis of blood samples is clinically more tractable than the analysis of biopsies. Samples were cryopreserved before treatment with either anti-TNF or anti-IL17. Treatment response was later assessed by a rheumatologist according to EULAR response criteria [35] (Additional file 7). We next selected 32 patients, of whom eight were classified as responders (R) and eight as nonresponders (NR) to either of the two drugs, as well as eight healthy controls. Cryopreserved PBMCs from these patients were analysed with scRNA-seq. For preprocessing, data were divided into two data sets by treatment regimen, each data set containing the corresponding eight R and NR along with the eight healthy

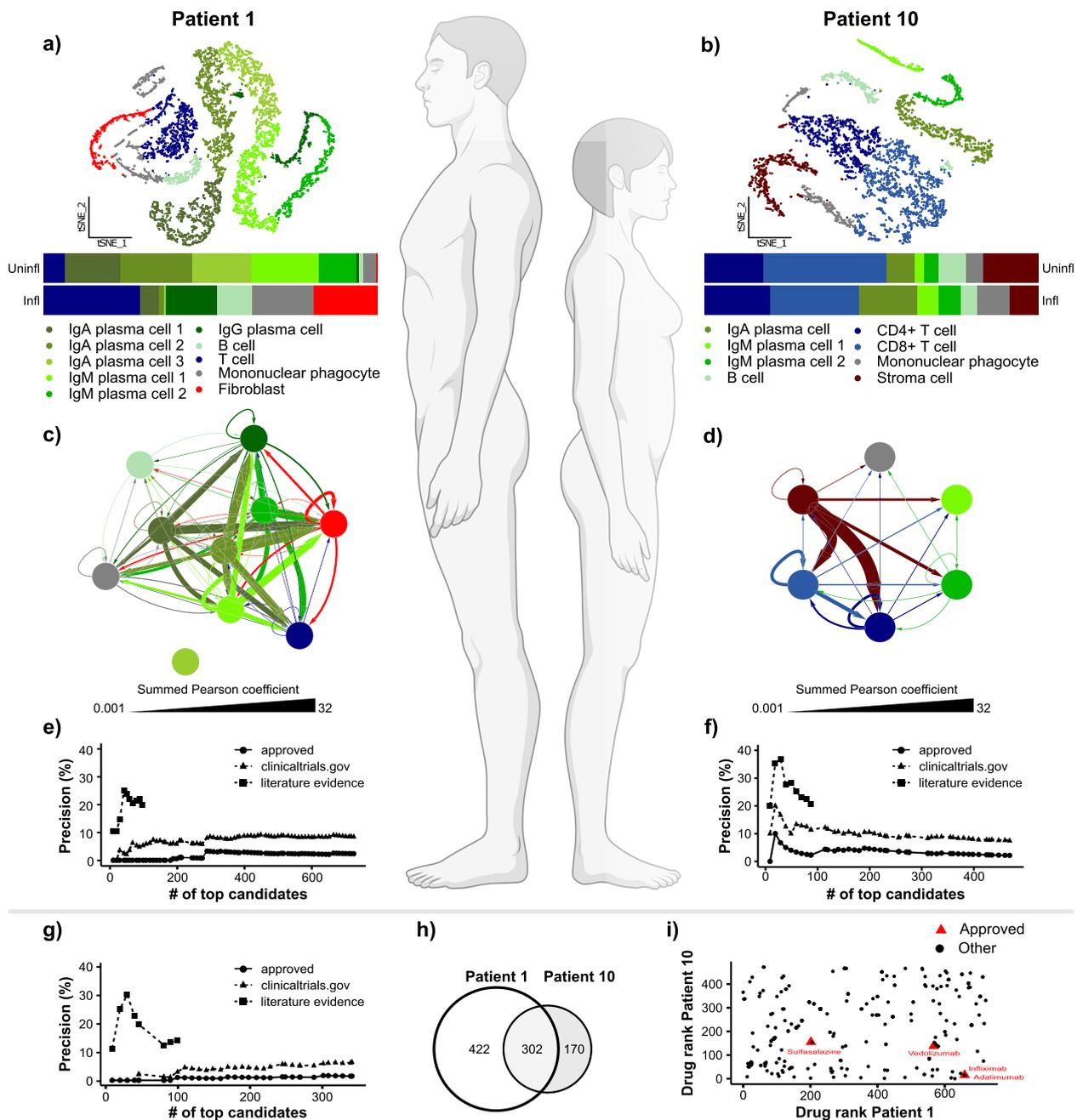


Fig. 6 scDrugPrio for individual drug prediction. **a, b** Cell type proportions differed greatly between two CD patients, as shown in the horizontally stacked bar plots representing paired biopsies from inflamed (infl) and uninflamed (uninfl) lesions that were taken from each patient. **c, d** Patients also show differences in the composition and interconnectivity (representing ligand interactions) of the MCDMs. Patient 1 had a cell type for which no ligand-target interactions could be found with any other cell types in the MCDM. **e, f** Precision among ranked candidates for approved CD drugs, drugs for which clinical trials had been registered for CD and literature evidence (only for top 100 ranking drugs). The precision for the ranked drug candidates for patient 1 was low for approved CD drugs, while literature evidence supported the top-ranking drugs, of which many are anti-inflammatory. In contrast, patient 10 had several approved CD drugs among the top 100 candidates, and a curated literature search confirmed the validity of many more candidates. **g** Precision for prediction based on pooled patient data was poor. **h** Venn diagram presenting the overlap of considered drug candidates for patients 1 and 10. **i** Interindividual differences between patients 1 and 10 were reflected in the prediction outcome, as no correlation existed between the drug rank of drugs that were candidates in both patients

controls. After application of quality cut-offs, the data included 78,610 cells with 5088 mean reads for anti-TNF analysis and 72,472 cells with 5343 mean reads for anti-IL17 analysis. Data were batch-corrected and DCA denoised before clusters were identified and cell-typed using marker genes (Additional file 1: Fig. S15 & S16; Additional file 7).

For each data set, DEGs were calculated through comparison of cells from healthy controls to either R or NR. The precision for approved PsA drugs among the top 100 candidates in the respective aggregated ranking was 0% for anti-IL17 NR, 4% for anti-IL17 R, 1% for anti-TNF NR and 1% for anti-TNF R (Additional file 1: Fig. S15 & S16). Unexpectedly, anti-TNF treatment received a low rank in anti-TNF R but not in NR (#333 and #88, respectively), while anti-IL-17 was not considered a candidate in either R or NR.

Further analyses of our PBMC data from all R and NR patients showed that the TNF signalling pathway was significant in only 8 and 8% of clusters, respectively. The corresponding figures for the IL-17 signalling pathway were 17 and 13% in the anti-IL17 R and NR groups, respectively. In those cell types, most pathways were downregulated, including those regulated by *TNF* and *IL17* (Additional file 1: Fig. S17). This result contrasted with previous studies of skin and synovium from PsA, which showed increased mRNA expression of *TNF* and *IL17*, as well as their pathways [72, 73]. A similar dichotomy between local inflamed tissues and cells in the blood in autoimmune diseases has been previously described [74]. This dichotomy can be explained by the physiological need to localise inflammatory responses to inhibit systemic and possibly fatal responses. The general clinical implication may be that drug predictions should ideally be based on local tissue samples [35].

Discussion

The main problem in therapeutics, which serves as the basis for this study, is the large number of IMID patients who do not respond to treatment [2, 3, 75]. Previous virtual drug screening methods for inflammatory diseases are based on genetic variance or bulk RNA sequencing [16, 76, 77] and hence do not consider variations in gene expression across different cell types, biopharmacological properties, or individual variations between patients with the same diagnosis. While harnessing the daunting complexity and heterogeneity for personalised treatment may seem impossible by health care standards today, this challenge should be put in the context of the suffering and costs resulting from ineffective drug treatment. Many IMIDs cause life-long morbidity and increased mortality. The yearly cost of treating an individual IMID patient

may be hundreds of thousands of dollars for drugs and hospital care [1].

Recent efforts for drug toxicity screening [8, 78] support the feasibility of scRNA-seq to capture relevant cellular information. However, systematic solutions for drug prioritisation for IMIDs based on scRNA-seq remain to be devised. We therefore propose a computational framework, scDrugPrio, that extends on existing bioinformatic tools [6, 16, 24] by providing a framework for data integration, enabling drug ranking based on a multifaceted understanding of cell type-specific disease mechanisms, altered cellular crosstalk and pharmacological effects. We demonstrate that scDrugPrio yields relevant and robust drug prioritisations, outperforms previous methods [7, 67], and holds potential for individualised as well as pooled drug prioritisation and repurposing.

An important advance of scDrugPrio is that it can be applied to scRNA-seq data. The importance lies in the fact that complex diseases each involve differential expression of thousands of genes across multiple cell types [6]. A previous case report [7] described one successful example of treating an individual patient with immunological diseases based on scRNA-seq data. However, the drug choices were empirical rather than systematic. Because scRNA-seq allows transcriptome-wide analyses in each of thousands of cells, it is possible to infer disease-associated changes in individual patients preferably by comparisons with noninflamed samples from the same individual or to groups of healthy individuals. Thus, scDrugPrio has the potential to personalise the treatment of individual patients. The importance of this advance is highlighted by our results and previous findings [53, 79] showing great interindividual differences in the molecular and cellular composition of human diseases. For example, we showed that scDrugPrio ranked anti-TNF treatment high in a CD patient who was classified as a responder but not in a nonresponding patient. In the latter patient, other immunomodulatory drugs, such as natalizumab, received high ranks. Natalizumab is mainly used in MS but has, in previous studies, shown positive effects in CD [80], making it a viable recommendation. These examples emphasise that successful drug screening will need to consider variations between patients with the same diagnosis.

There are several limitations of scRNA-seq-based drug predictions in IMIDs. Many of these depend on the challenges involved in harnessing complex and heterogeneous disease-associated changes with an emerging technology such as scRNA-seq. An analogy to a historical example may illustrate how such limitations may drive scientific progress. In 1970, Needleman–Wunsch [81] and 1981, Smith–Waterman [82] published algorithms for global and local sequence alignment, which were

widely used. The limitations of those algorithms were that they were mainly useful for nucleotide but not protein sequence analyses because of limited protein sequence data and no scoring system that modelled protein evolution. During the next two decades, these problems were resolved by increasingly accurate data and methods [83]. Importantly, 42 years after publication of the Smith–Waterman algorithms for proteins, these algorithms can generate very accurate results when combined with scoring systems that were later developed [83]. We propose that the limitations of scRNA-seq that we face today will lead to a similar development of increasingly accurate technologies. One obvious limitation of scRNA-seq is that mRNA and protein levels may be poorly correlated, which can limit biological interpretability. From this perspective, the use of DEGs for scDrugPrio’s pharmacological predictions is a relative strength, as DEGs have been shown to correlate significantly better with protein levels [84] compared to mRNA levels alone and hence increase the biological relevance of our predictions. Inherent limitations of scDrugPrio also derive from the use of current interactomes, which are not comprehensive in terms of proteins, interactions, and variations across cell types [85] and are prone to investigative biases. While it is impossible to address all these concerns, we explored whether network proximity-based drug selection was influenced by investigative bias through replication of key results in a smaller yet unbiased interactome. We found that precision among candidates was slightly lower, partly due to missing drug targets in the interactome, but that results were comparable. Additionally, scDrugPrio might benefit from systematic parameter optimisation, which is currently not possible due to the limited amount of suitable scRNA-seq data sets.

Predictions were based on drug-target effects which are complex and potentially incompletely characterised [86] which could explain why some of the drug predictions could not be validated experimentally. The need for better characterised drug effects and the relative importance of drug targets is exemplified by etanercept, which inhibits TNF and its receptors but may activate IgG receptors. The TNF-inhibitory effects are beneficial in PsA, while those on IgG receptors are not clearly defined. However, because all these targets were downregulated in PBMCs from nonresponding PsA patients, etanercept (counteracting IgG downregulation) received a higher rank than in patients responding to anti-TNF treatment. While unexpected, this highlights the need for systematic information about the relative importance of drug targets. Future efforts aiming to address these limitations might find that the predictive capability of scDrugPrio can be further enhanced by integration of multiple other

factors, such as binding affinity (e.g. BindingDB) or bioactivity (e.g. ChEMBL), especially if data become more comprehensive. However, it should be emphasised that treatment decisions ultimately depend on informed discussions between patients and physicians. Future clinical studies are warranted to determine which variables should be included in a drug prediction tool like scDrugPrio, and which variables should be analysed separately depending on the specific disease-context or questions arising from patient-physician discussions.

The above example of etanercept in PsA highlights an important clinical concern, which to our knowledge has not been recognised in the context of drug prediction methods. While analyses of blood samples are often more tractable in routine clinical practice, disease-associated mechanisms may vary greatly between cells in blood and inflamed tissues. Our scRNA-seq analyses of PBMCs from PsA patients who did or did not respond to treatment with either anti-TNF or anti-IL-17 showed that predicted activation of TNF and IL-17 induced pathways was found only in a small portion of the PBMC cell types and, in fact, was downregulated in both responders and nonresponders. In contrast, previous studies [72, 73] of synovium from PsA patients have shown consistent upregulation of both signalling pathways. Additionally, one previous study also showed differences between synovium and skin from the same patients [72]. Thus, scDrugPrio should ideally be applied to local, inflamed tissue samples of the relevant tissue.

Despite these limitations, the translational relevance of scDrugPrio was supported by analyses of precision/recall for drugs that were approved for the studied diseases, as well as by *in vitro* and *in vivo* experiments. Those experiments implied two drugs, namely, adapalene (used for acne vulgaris) and amrinone (used for congestive heart failure), that had not been previously described as candidates for RA treatment. However, both have anti-inflammatory effects and could, therefore, be effective [87, 88]. This potential was supported by *in vivo* experiments in which CIA mice were treated with amrinone (adapalene is a topical skin drug and hence is not suitable for systemic treatment in this experimental system). This example also suggests a potentially important pharmacological application of scDrugPrio, namely, virtual drug repurposing by systematic screening of thousands of drugs across several inflammatory diseases, as well as in patients who do not respond to standard treatment.

Conclusions

Here, we show that scDrugPrio has the potential for individualised drug predictions. We have made data and tools freely available for this purpose. However, further

parameter optimisation and controlled, prospective clinical studies are needed for clinical translation. If successful, this approach could lead to a radical change in health care, which today is largely based on treating groups of patients with the same diagnosis with a limited number of drugs based on a limited understanding of the underlying molecular complexity and heterogeneity with limited population-based efficacy [89].

Abbreviations

AIA	Antigen-induced arthritis
BSA	Bovine serum albumin
CIA	Collagen-induced arthritis
CD	Crohn's disease
CFSE	Carboxyfluorescein succinimidyl ester
CMap	Connectivity Map
CSF	Cerebrospinal fluid
DCA	Deep count autoencoder
DEGs	Differentially expressed genes
ELISA	Enzyme-linked immunosorbent assay
EULAR	European League Against Rheumatism
FBS	Foetal bovine serum
FDA	Food and Drug Administration
GEM	Gel BeadInEmulsions
GDSC	Genomics of Drug Sensitivity in Cancer
GEO	Gene Expression Omnibus
GWAS	Genome-wide association study
H&E	Haematoxylin and eosin
HGNC	HUGO Gene Nomenclature Committee
IiH	Idiopathic intracerebral hypertension
IL	Interleukin
IMiD	Immune-mediated inflammatory disease
MACS	Magnetic-activated cell sorting
MCDM	Multicellular disease model
MHC-II	Major histocompatibility complex class II
MS	Multiple sclerosis
NCBI	National Center for Biotechnology Information
NR	Non-responder
OMiM	Online Mendelian Inheritance in Man
PBS	Phosphate-buffered saline
PBMCs	Peripheral blood mononuclear cells
PRISM	Profile of Resistance In Solid Malignancies
PsA	Psoriatic arthritis
R	Responder
RA	Rheumatoid arthritis
R&D	Research and Development
scRNA-seq	Single-cell RNA sequencing
SNP	Single-nucleotide polymorphism
TNF	Tumour necrosis factor
UMI	Unique molecular identifier
VCF	Variant Call Format

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-024-01314-7>.

Additional file 1: Supplementary figures. Including supplementary figures (Fig. S1-S22) and figure legends.

Additional file 2: Supplement. Including supplementary methods and results as well as supplementary tables (Tables S1& S2).

Additional file 3: AIA data. Includes quality cut-offs, cell type markers, DEGs, metadata for scDrugPrio application to the AIA data set, as well as DrugBank and clinicaltrials.gov related data.

Additional file 4: MS data. Includes quality cut-offs, cell type markers, DEGs, metadata for scDrugPrio application to the MS data set, as well as DrugBank and clinicaltrials.gov related data.

Additional file 5: Pooled CD data. Includes quality cut-offs, cell type markers, DEGs, metadata for scDrugPrio application to the pooled CD data set, as well as DrugBank and clinicaltrials.gov related data.

Additional file 6: Individual CD data. Includes cell typing outcomes, DEGs and metadata for scDrugPrio application to individual CD patients' data sets.

Additional file 7: PsA data. Includes quality cut-offs, cell type markers, DEGs, metadata for scDrugPrio application to the anti-TNF and anti-IL17 PsA data sets, as well as DrugBank and clinicaltrials.gov related data.

Acknowledgements

Not applicable.

Authors' contributions

SS designed and performed the bioinformatics and analyses, which were supervised by DG, OS, IAK, MG, and MB. HH, XL, and HZ performed scRNA-seq. SL performed the scRNA-seq data extraction, which was led by DG and MB. SS, DE, MS, and YZ analysed the data. SM and AJ were responsible for clinical studies of drug responses in patients with PsA. HW performed the experimental validation. SS, JL, DG, OS, and MB took the lead in writing the manuscript, and all authors provided critical feedback. All authors read and approved the final manuscript.

Funding

Open access funding provided by Karolinska Institute. This work was supported by the DocTIS project, which has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement N° 848028; Swedish Cancer Society CAN 2017/411; Coccozza Foundation; National Natural Science Foundation of China 82171791, US National Institutes of Health grants HL155107 and HL155096; American Heart Association grant 957729; European Union's Horizon 2021 Research and Innovation Programme grant 101057619 and Mag-Tarmfonden (grant 1–23). The computations were partially enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at Linköping University partially funded by the Swedish Research Council through grant agreement no. 2018–05973.

Availability of data and materials

The scRNA-seq data of the AIA mouse model that was generated and analysed during the current study is available from Gene Expression Omnibus (GEO) GSE193536 (<https://ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE193536>) [90]. The MS dataset analysed during the current study is available from GEO GSE138266 (<https://ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE138266>) [51]. The CD dataset analysed during the current study is available from GEO GSE134809 (<https://ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE134809>) [54]. The scRNA-seq of PsA patients and controls that was generated and analysed during the current study is available at FigShare (anti-IL-17 treatment: <https://doi.org/10.25452/figshare.plus.22344175>; anti-TNF treatment: <https://doi.org/10.25452/figshare.plus.22344145>) [91]. Extended meta-data for all above data sets is available from FigShare [91], <https://doi.org/10.25452/figshare.plus.c.7094116>. The codes associated with scDrugPrio analysis are available from GitHub [92], <https://github.com/SDTC-CPMed/scDrugPrio>.

Declarations

Ethics approval and consent to participate

All psoriatic arthritis patients and healthy controls provided written informed consent to participate in the study. The study was approved by Hospital Universitari Vall d'Hebron Clinical Research Ethics Committee with reference number 20/0022. Protocols were reviewed and approved by the local institutional review board of each participating centre. This research conformed to the principles of the Helsinki Declaration. All experimental procedures related to the AIA mouse model were performed according to the guidelines provided by the Swedish Animal Welfare Act and approved by the Ethical Committee for Research on Animals in Stockholm, Sweden (N271-14). All experimental procedures related to the CIA mouse model were performed in accordance

with protocols approved by the Animal Experimental Ethics Committee of Xuzhou Medical University (202012A162).

Consent for publication

Not applicable.

Competing interests

MB is the scientific founder of Mavatar, Inc. DRG is employed by Mavatar, Inc. H.H. is co-founder and shareholder of Omniscope, SAB member of Nanostring and MIRXES, and consultant to Moderna and Singularity. JL is coscientific founder of Scipher Medicine. The remaining authors declare that they do not have any competing interests.

Author details

¹Centre for Personalised Medicine, Linköping University, Linköping, Sweden. ²Department of Gastroenterology and Hepatology, University Hospital, Linköping, Sweden. ³Postal Address: LIME/Medical Digital Twin Research Group, Division of ENT, CLINTEC, Karolinska Institute, Tomtebodavägen 18A, 171 65 Solna, Stockholm, Sweden. ⁴Division of Statistics and Machine Learning, Department of Computer and Information Science, Linköping University, Linköping, Sweden. ⁵Mavatar, Inc, Stockholm, Sweden. ⁶Division for Bioinformatics, Department of Physics, Chemistry and Biology, Linköping University, Linköping, Sweden. ⁷CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), 08028 Barcelona, Spain. ⁸Universitat Pompeu Fabra (UPF), 08002 Barcelona, Spain. ⁹Grup de Recerca de Reumatologia, Institut de Recerca Vall d'Hebron, Barcelona, Spain. ¹⁰Department of Physics and Astronomy, Northwestern University, Evanston, IL 60208, USA. ¹¹Northwestern Institute On Complex Systems, Northwestern University, Evanston, IL 60208, USA. ¹²Division of Cardiovascular Medicine, Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. ¹³Jiangsu Key Laboratory of Immunity and Metabolism, Department of Pathogenic Biology and Immunology, Xuzhou Medical University, Jiangsu, China.

Received: 2 March 2023 Accepted: 12 March 2024

Published online: 20 March 2024

References

- Björnsson B, Borrebaeck C, Elander N, Gasslander T, Gawel DR, Gustafsson M, et al. Digital twins to personalize medicine. *Genome Med.* 2019;12(1):4.
- Kayal M, Ungaro RC, Bader G, Colombel J-F, Sandborn WJ, Stalgis C. Net remission rates with biologic treatment in Crohn's disease: a reappraisal of the clinical trial data. *Clin Gastroenterol Hepatol.* 2022.
- Alic H, Amiot A, Kirchgessner J, Tréton X, Allez M, Bouhnik Y, et al. The effectiveness of either ustekinumab or vedolizumab in 239 patients with Crohn's disease refractory to anti-tumour necrosis factor. *Aliment Pharmacol Ther.* 2020;51(10):948–57.
- Shalek AK, Benson M. Single-cell analyses to tailor treatments. *Sci Transl Med.* 2017;9(408).
- Breyner C, Dresselaers T, Perrier C, Arijis I, Cremer J, Van Lommel L, et al. Unique gene expression and MR T2 relaxometry patterns define chronic murine dextran sodium sulphate colitis as a model for connective tissue changes in human Crohn's disease. *PLoS ONE.* 2013;8(7):e68876.
- Gawel DR, Serra-Musach J, Lilja S, Aagesen J, Arenas A, Asking B, et al. A validated single-cell-based strategy to identify diagnostic and therapeutic targets in complex diseases. *Genome Med.* 2019;11(1):47.
- Kim D, Kobayashi T, Voisin B, Jo J-H, Sakamoto K, Jin S-P, et al. Targeted therapy guided by single-cell transcriptomic analysis in drug-induced hypersensitivity syndrome: a case report. *Nat Med.* 2020;26(2):236–43.
- Hsieh C-Y, Wen J-H, Lin S-M, Tseng T-Y, Huang J-H, Huang H-C, et al. scDrug: From single-cell RNA-seq to drug response prediction. *Comput Struct Biotechnol J.* 2023;21:150–7.
- Suphavitai C, Chia S, Sharma A, Tu L, Da Silva RP, Mongia A, et al. Predicting heterogeneity in clone-specific therapeutic vulnerabilities using single-cell transcriptomic signatures. *Genome Med.* 2021;13(1):189.
- Fustero-Torre C, Jimenez-Santos MJ, Garcia-Martin S, Carretero-Puche C, Garcia-Jimeno L, Ivanchuk V, et al. Beyondcell: targeting cancer therapeutic heterogeneity in single-cell RNA-seq data. *Genome Med.* 2021;13(1):187.
- Ianevski A, Lahtela J, Javarappa KK, Sergeev P, Ghimire BR, Gautam P, et al. Patient-tailored design for selective co-inhibition of leukemic cell subpopulations. *Sci Adv.* 2021;7(8).
- Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, Schubert M, et al. A Landscape of Pharmacogenomic Interactions in Cancer. *Cell.* 2016;166(3):740–54.
- Corsello SM, Nagari RT, Spangler RD, Rossen J, Kocak M, Bryan JG, et al. Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. *Nature Cancer.* 2020;1(2):235–48.
- Freshour SL, Kiwala S, Cotto KC, Coffman AC, McMichael JF, Song JJ, et al. Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsourcing efforts. *Nucleic Acids Res.* 2021;49(D1):D1144–51.
- Li X, Lee EJ, Lilja S, Loscalzo J, Schäfer S, Smelik M, et al. A dynamic single cell-based framework for digital twins to prioritize disease genes and drug targets. *Genome Med.* 2022;14(1):48.
- Guney E, Menche J, Vidal M, Barabasi AL. Network-based in silico drug efficacy screening. *Nat Commun.* 2016;7:10331.
- Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, et al. Functional discovery via a compendium of expression profiles. *Cell.* 2000;102(1):109–26.
- Cheng F, Kovacs IA, Barabasi AL. Network-based prediction of drug combinations. *Nat Commun.* 2019;10(1):1197.
- Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* 2006;34(Database issue):D668–72.
- Özgür A, Vu T, Erkan G, Radev DR. Identifying gene-disease associations using centrality on a literature mined gene-interaction network. *Bioinformatics.* 2008;24(13):i277–85.
- Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal Complex Systems.* 2006;1695(5):1–9.
- Negre CFA, Morzan UN, Hendrickson HP, Pal R, Lisi GP, Loria JP, et al. Eigenvector centrality for characterization of protein allosteric pathways. *Proc Natl Acad Sci U S A.* 2018;115(52):E12201–8.
- Ashtiani M, Mirzaie M, Jafari M. CINNA: an R/CRAN package to decipher central informative nodes in network analysis. *Bioinformatics (Oxford, England).* 2018;35.
- Browaeys R, Saelens W, Saeys Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat Methods.* 2020;17(2):159–62.
- Puram SV, Tirosh I, Parkh AS, Patel AP, Yizhak K, Gillespie S, et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell.* 2017;171(7):1611–24.e24.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498–504.
- de Valle IF, Roweth HG, Malloy MW, Moco S, Barron D, Battinelli E, et al. Network medicine framework shows that proximity of polyphenol targets and disease proteins predicts therapeutic effects of polyphenols. *Nature Food.* 2021;2(3):143–55.
- Magnusson M, Zare F, Tarkowski A. Requirement of type I interferon signaling for arthritis triggered by double-stranded RNA. *Arthritis Rheum.* 2006;54(1):148–57.
- Gierahn TM, Wadsworth MH, Hughes TK, Bryson BD, Butler A, Satija R, et al. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat Methods.* 2017;14(4):395–8.
- Illumina. bcl2fastq and bcl2fastq2 Conversion Software 2013 [Available from: https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html].
- Picard toolkit. Broad Institute, GitHub repository. 2019.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15–21.
- Aterido A, Cañete JD, Tornero J, Ferrándiz C, Pinto JA, Gratacós J, et al. Genetic variation at the glycosaminoglycan metabolism pathway contributes to the risk of psoriatic arthritis but not psoriasis. *Ann Rheum Dis.* 2019;78(3).
- Taylor W, Gladman D, Helliwell P, Marchesoni A, Mease P, Mielants H. Classification criteria for psoriatic arthritis: development of new criteria from a large international study. *Arthritis Rheum.* 2006;54(8):2665–73.

35. Ritchlin CT, Kavanaugh A, Gladman DD, Mease PJ, Helliwell P, Boehncke WH, et al. Treatment recommendations for psoriatic arthritis. *Ann Rheum Dis*. 2009;68(9):1387–94.
36. Kang HM, Subramaniam M, Targ S, Nguyen M, Maliskova L, McCarthy E, et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat Biotechnol*. 2018;36(1):89–94.
37. Huang X, Huang Y. Cellsnr-lite: an efficient tool for genotyping single cells. *Bioinformatics*. 2021;37(23):4569–71.
38. Huang Y, McCarthy DJ, Stegle O. Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference. *Genome Biol*. 2019;20(1):273.
39. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol*. 2019;15(6):e8746.
40. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive Integration of Single-Cell Data. *Cell*. 2019;177(7):1888–902.e21.
41. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol*. 2018;36(5):411–20.
42. Luecken MD, Büttner M, Chaichompu K, Danese A, Interlandi M, Mueller MF, et al. Benchmarking atlas-level data integration in single-cell genomics. *Nat Methods*. 2022;19(1):41–50.
43. Eraslan G, Simon LM, Mircea M, Mueller NS, Theis FJ. Single-cell RNA-seq denoising using a deep count autoencoder. *Nat Commun*. 2019;10(1):390.
44. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32.
45. Krstajic D, Buturovic LJ, Leahy DE, Thomas S. Cross-validation pitfalls when selecting and assessing regression and classification models. *J Cheminform*. 2014;6(11):1–15.
46. Efron B, Tibshirani RJ. An introduction to the bootstrap. CRC press; 1994.
47. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. *J Machine Learn Res*. 2011;12:2825–30.
48. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol*. 2015;16(1):278.
49. Zhao B, Erwin A, Xue B. How many differentially expressed genes: a perspective from the comparison of genotypic and phenotypic distances. *Genomics*. 2018;110(1):67–73.
50. Schafflick D, Xu CA, Hartlehner M, Cole M, Schulte-Mecklenbeck A, Lautwein T, et al. Integrated single cell analysis of blood and cerebrospinal fluid leukocytes in multiple sclerosis. *Nat Commun*. 2020;11(1):247.
51. Schafflick D, Hartlenert M, Schulte-Mecklenbeck A, Lautwein T, Wolbert J, Horste GM. Integrated single cell analysis of blood and cerebrospinal fluid leukocytes in multiple sclerosis. *Gene Expression Omnibus (GEO)*. 2019. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE138266>.
52. Tweedie S, Braschi B, Gray K, Jones TEM, Seal RL, Yates B, et al. GeneNames.org: the HGNC and VGNCR resources in 2021. *Nucleic Acids Res*. 2021;49(D1):D939–46.
53. Martin JC, Chang C, Boschetti G, Ungaro R, Giri M, Grout JA, et al. Single-cell analysis of Crohn's disease lesions identifies a pathogenic cellular module associated with resistance to anti-TNF therapy. *Cell*. 2019;178(6):1493–508.e20.
54. Cho J, Martin J, Merad M, Kenigsberg E, Giri M. Single-cell analysis of Crohn's disease lesions identifies a pathogenic cellular module associated with resistance to anti-TNF therapy. *Gene Expression Omnibus (GEO)*. 2019. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE134809>.
55. Shi Y, Xu M, Pan S, Gao S, Ren J, Bai R, et al. Induction of the apoptosis, degranulation and IL-13 production of human basophils by butyrate and propionate via suppression of histone deacetylation. *Immunology*. 2021;164(2):292–304.
56. He C, Gao S, Zhao X, Shi Y, Tang Y, Cao Y, et al. An efficient and cost-effective method for the purification of human basophils. *Cytometry A*. 2022;101(2):150–8.
57. Li H, Tang Y, Ren J, Bai R, Hu L, Jia W, et al. Identification of novel B-1 transitional progenitors by B-1 lymphocyte fate-mapping transgenic mouse model Bhlhe41 (dTomato-Cre). *Front Immunol*. 2022;13:946202.
58. Brand DD, Latham KA, Rosloniec EF. Collagen-induced arthritis. *Nat Protoc*. 2007;2(5):1269–75.
59. Park SH, Rhee J, Kim SK, Kang JA, Kwak JS, Son YO, et al. BATF regulates collagen-induced arthritis by regulating T helper cell differentiation. *Arthritis Res Ther*. 2018;20(1):161.
60. Woetzel D, Huber R, Kupfer P, Pohlens D, Pfaff M, Driesch D, et al. Identification of rheumatoid arthritis and osteoarthritis patients by transcriptome-based rule set generation. *Arthritis Res Ther*. 2014;16(2):R84.
61. Woetzel D, Huber R, Kupfer P, Pohlens D, Pfaff M, Driesch D, Häupl T, Koczan D, Stiehl P, Guthke R, Kinne RW. Identification of rheumatoid arthritis and osteoarthritis patients by transcriptome-based rule set generation. *Gene Expression Omnibus (GEO)*. 2014. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE55235>.
62. Tasaki S, Suzuki K, Kassai Y, Takeshita M, Murota A, Kondo Y, et al. Multi-omics monitoring of drug response in rheumatoid arthritis in pursuit of molecular remission. *Nat Commun*. 2018;9(1):2755.
63. Tasaki S, Toyoshiba H, Takeuchi T. Whole blood gene expression of rheumatoid arthritis. *Gene Expression Omnibus (GEO)*. 2018. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE93272>.
64. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Mangano C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019;47(D1):D1005–12.
65. Amberger J, Bocchini CA, Scott AF, Hamosh A. McKusick's Online Mendelian Inheritance in Man (OMIM). *Nucleic Acids Res*. 2009;37(1):D793–6.
66. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci*. 2019;28(11):1947–51.
67. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*. 2006;313(5795):1929–35.
68. Luck K, Kim D-K, Lambourne L, Spirohn K, Begg BE, Bian W, et al. A reference map of the human binary protein interactome. *Nature*. 2020;580(7803):402–8.
69. Taylor PC, Feldmann M. Anti-TNF biologic agents: still the therapy of choice for rheumatoid arthritis. *Nat Rev Rheumatol*. 2009;5(10):578–82.
70. Rizzi M, Lorenzetti R, Fischer K, Staniek J, Janowska I, Troilo A, et al. Impact of tofacitinib treatment on human B-cells in vitro and in vivo. *J Autoimmun*. 2017;77:55–66.
71. Hausser-Kinzel S, Weber MS. The role of B cells and antibodies in multiple sclerosis, neuromyelitis optica, and related disorders. *Front Immunol*. 2019;10:8.
72. Belasco J, Louie JS, Gulati N, Wei N, Nograles K, Fuentes-Duculan J, et al. Comparative genomic profiling of synovium versus skin lesions in psoriatic arthritis. *Arthritis Rheumatol*. 2015;67(4):934–44.
73. Nerviani A, Boutet MA, Tan WSG, Goldmann K, Purkayastha N, Lajtos TA, et al. IL-23 skin and joint profiling in psoriatic arthritis: novel perspectives in understanding clinical responses to IL-23 inhibitors. *Ann Rheum Dis*. 2021;80(5):591–7.
74. Lee EJ, Lilja S, Li X, Schäfer S, Zhang H, Benson M. Bulk and single cell transcriptomic data indicate that a dichotomy between inflammatory pathways in peripheral blood and arthritic joints complicates biomarker discovery. *Cytokine*. 2020;127:154960.
75. FDA U. Paving the way for personalized medicine. FDA's Role in a new Era of Medical Product Development US Department of Health and Human Services. 2013:1–61.
76. Cheng F, Desai RJ, Handy DE, Wang R, Schneeweiss S, Barabási A-L, et al. Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat Commun*. 2018;9(1):1–12.
77. Cheng F, Zhao J, Wang Y, Lu W, Liu Z, Zhou Y, et al. Comprehensive characterization of protein–protein interactions perturbed by disease mutations. *Nat Genet*. 2021;53(3):342–53.
78. Kanemaru K, Cranley J, Muraro D, Miranda AMA, Ho SY, Wilbrey-Clark A, et al. Spatially resolved multiomics of human cardiac niches. *Nature*. 2023;619(7971):801–10.
79. Menche J, Guney E, Sharma A, Branigan PJ, Loza MJ, Baribaud F, et al. Integrating personalized gene expression profiles into predictive disease-associated gene pools. *npj Syst Biol Appl*. 2017;3(1):10.
80. Nelson SML, Nguyen TM, McDonald JWD, MacDonald JK. Natalizumab for induction of remission in Crohn's disease. *Cochrane Database of Systematic Reviews*. 2018(8).
81. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol*. 1970;48(3):443–53.

82. Smith TF, Waterman MS. Identification of common molecular subsequences. *J Mol Biol.* 1981;147(1):195–7.
83. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A.* 1992;89(22):10915–9.
84. Koussounadis A, Langdon SP, Um IH, Harrison DJ, Smith VA. Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. *Sci Rep.* 2015;5(1):10775.
85. Gillis J, Pavlidis P. The impact of multifunctional genes on "guilt by association" analysis. *PLoS ONE.* 2011;6(2):e17258.
86. Lin A, Giuliano CJ, Palladino A, John KM, Abramowicz C, Yuan ML, et al. Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. *Sci Transl Med.* 2019;11(509).
87. Rusu A, Tanase C, Pascu GA, Todoran N. Recent Advances Regarding the Therapeutic Potential of Adapalene. *Pharmaceuticals (Basel).* 2020;13(9).
88. Chanani NK, Cowan DB, Takeuchi K, Poutias DN, Garcia LM, del Nido PJ, et al. Differential effects of amrinone and milrinone upon myocardial inflammatory signaling. *Circulation.* 2002;106(12 Suppl 1):I284–9.
89. Schork NJ. Personalized medicine: time for one-person trials. *Nature.* 2015;520(7549):609–11.
90. Schäfer S, Lilja S, Smelik M, Eklund D, Li X, Sysoev O, Zhang H, Zhao Y, Kovács IA, Gustafsson M, Gawel D, Wang H, Benson M. Network analyses of single cell-based digital twins for individualized treatment of antigen induced arthritis in mice. *Gene Expression Omnibus (GEO).* 2023. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE193536>.
91. Schäfer S. scDrugPrio data sets and metadata. Figshare+ 2024. <https://doi.org/10.25452/figshare.plus.c.7094116>.
92. Schäfer S. scDrugPrio codes and vignettes. GitHub 2024. <https://github.com/SDTC-CPMed/scDrugPrio>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.