OXFORD

Briefings in Bioinformatics, 2023, **24(6)**, 1–10
https://doi.org/10.1093/bib/bbad387
Problem Solving Protocol

COMO: a pipeline for multi-omics data integration in metabolic modeling and drug discovery

Brandt Bessell†, Josh Loecker†, Zhongyuan Zhao, Sara Sadat Aghamiri, Sabyasachi Mohanty, Rada Amin, Tomáš Helikar and Bhanwar Lal Puniya

Corresponding authors: Tomáš Helikar, Department of Biochemistry, University of Nebraska-Lincoln, 1901 Vine Street, Lincoln, Nebraska 68588, USA. Tel.: +1-402-472-3530; E-mail: thelikar2@unl.edu; Bhanwar Lal Puniya, Department of Biochemistry, University of Nebraska-Lincoln, 1901 Vine Street, Lincoln, Nebraska 68588, USA. E-mail: bpuniya2@unl.edu

†Brandt Bessell and Josh Loecker contributed equally to this work.

Abstract

Identifying potential drug targets using metabolic modeling requires integrating multiple modeling methods and heterogeneous biological datasets, which can be challenging without efficient tools. We developed Constraint-based Optimization of Metabolic Objectives (COMO), a user-friendly pipeline that integrates multi-omics data processing, context-specific metabolic model development, simulations, drug databases and disease data to aid drug discovery. COMO can be installed as a Docker Image or with Conda and includes intuitive instructions within a Jupyter Lab environment. It provides a comprehensive solution for the integration of bulk and single-cell RNA-seq, microarrays and proteomics outputs to develop context-specific metabolic models. Using public databases, open-source solutions for model construction and a streamlined approach for predicting repurposable drugs, COMO enables researchers to investigate low-cost alternatives and novel disease treatments. As a case study, we used the pipeline to construct metabolic models of B cells, which simulate and analyze them to predict metabolic drug targets for rheumatoid arthritis and systemic lupus erythematosus, respectively. COMO can be used to construct models for any cell or tissue type and identify drugs for any human disease where metabolic inhibition is relevant. The pipeline has the potential to improve the health of the global community cost-effectively by providing high-confidence targets to pursue in preclinical and clinical studies. The source code of the COMO pipeline is available at https://github.com/HelikarLab/COMO. The Docker image can be pulled at https://github.com/HelikarLab/COMO/pkgs/container/como.

Keywords: immunometabolism; autoimmune diseases; drug discovery; computational pipeline

INTRODUCTION

Human diseases often alter cellular metabolism, making metabolic pathways potential targets for treatment [1]. Genome-scale metabolic models (GSMMs) are powerful computational tools to investigate the effects of drugs on a cell's phenotype and generate testable hypotheses [2, 3]. In recent years, the COBRA Toolbox and COBRApy have facilitated the construction and analysis of metabolic models [4, 5]. However, constructing tissue- and cell-type-specific GSMMs and performing systems-level studies often require various tools based on different programming languages, APIs, data resources and data formats. For example, R/Bioconductor is well suited for analyzing high-throughput datasets, while Matlab or Python are often used for metabolic modeling and analysis. In addition, model integration with external databases such as DrugBank [6] to run simulations under different conditions often requires additional manual, tedious and error-prone steps. The complexity and diversity of tools and data resources handled with different programming languages with manual steps could pose significant challenges for researchers, potentially leading to errors, inefficiencies and delays in using metabolic models for therapeutic interventions and other applications. Furthermore, as the amount of disease-specific omics data has increased in recent years, there has been a growing need for efficient bioinformatics tools to interpret and connect these data. [7]

To address these challenges, we developed COMO (Constraint-based Optimization of Metabolic Objectives), a pipeline that integrates data sources and analysis tools in a single, easy-to-use platform for researchers to harness the ever-growing volume of publicly available omics data. The pipeline allows users to analyze and integrate outputs of multi-omics data to build GSMMs for various tissue and cell types in different biological contexts. It combines transcriptomics and proteomics data from user-provided or publicly available databases using standardized meta-analysis

Brandt Bessell is a PhD student at University of Michigan Medical School. His research interests are genomic structural variation and mosaicism in neurodegenerative disease.
Josh Loecker is a PhD student at University of Nebraska-Lincoln. His research interests are bioinformatics and systems biology.
Zhongyuan Zhao is a postdoctoral researcher at Rice University. His research interests are wireless systems, machine learning and network science.
Sara Sadat Aghamiri is a researcher at University of Nebraska-Lincoln. Her research interest is computational systems biology.
Sabyasachi Mohanty is a PhD student at University of Nebraska-Lincoln. His research interest is bioinformatics.
Rada Amin is a Senior Research Associate at University of Nebraska-Lincoln. Her research interests are immunology and computational biology.
Tomáš Helikar is a Professor at University of Nebraska-Lincoln. His research interests are in the development and application of medical digital twins in drug discovery and development and health optimization.
Bhanwar Lal Puniya is a Research Assistant Professor at University of Nebraska-Lincoln. His research interests include computational biology, data analytics, systems immunology and drug discovery.
Received: June 21, 2023. **Revised:** September 4, 2023. **Accepted:** October 5, 2023

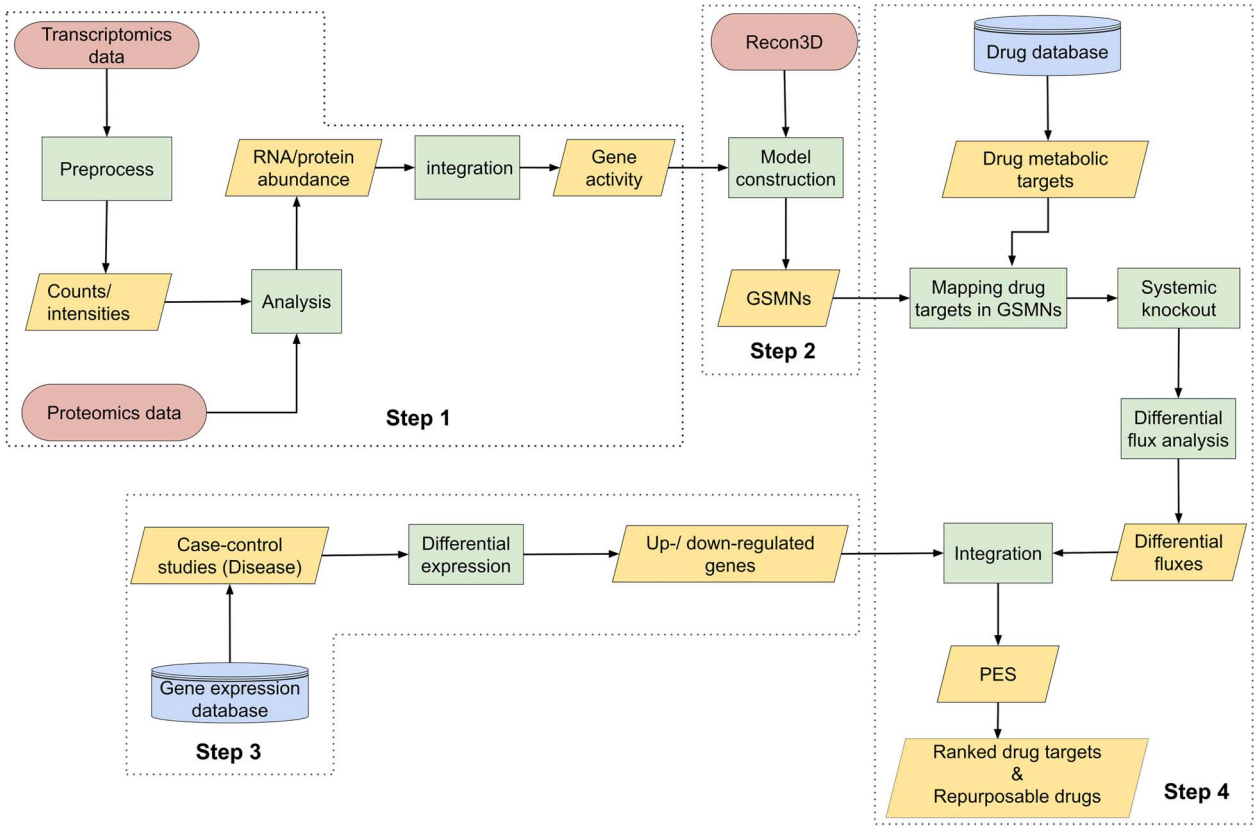

Figure 1. Flow chart of the COMO pipeline. GSMM refers to genome-scale metabolic networks.

techniques to synthesize constraint-based metabolic models and leverages existing knowledge from drug databases to identify potential therapeutic targets for user-specified diseases. COMO allows rapid, customizable, context-specific, constraint-based model generation, disease analysis and drug perturbation analysis. It comes packaged with popular standard tools such as Memote [8], Escher [9] and UMAP [10] for exploratory analysis and clustering. COMO is a general-purpose tool to develop GSMMs spanning organisms and biological scales. The drug discovery steps can be applicable to any disease area in which metabolic inhibition of specific cells and tissues is coherent.

In this study, we showed an application of the novel COMO pipeline to construct metabolic models of B cells and use them for drug target identification against autoimmune diseases: rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE). RA affects approximately 18 million individuals worldwide, while SLE impacts around 3.41 million people globally [11, 12], and both of these diseases lack a definitive cure. B cell metabolism has been shown to play a role in the onset of autoimmune diseases. For example, B cells from SLE patients have altered metabolism because of increased activity of the mTOR pathway [13], and B cells have already been explored as therapeutic targets against SLE [14]. Characterizing the naive B cell metabolism and understanding its role in autoimmune diseases could lead to new therapeutic strategies to restore normal immune function, reduce B cell activity and prevent the production of harmful antibodies.

METHODS

Pipeline design

COMO consists of four major steps for identifying drug targets (Figure 1).

The first step involves data analysis and consists of three substeps: (i) if bulk RNA-seq data are used for constructing the models, preprocessing the data and preparing the configuration files; (ii) analysis of any combinations of bulk-RNA-seq, single-cell RNA-seq, microarrays and proteomics data, and generation of a list of active genes for each data type; and (iii) checking for consensus among user-provided data types according to the desired level of rigor and merging them into a single set of active genes.

The second step uses the list of active genes generated in the first step and a reference model, for example, Recon3D [15], provided in the pipeline to build cell-type-specific models. Users can manually refine these models and provide them in Step 4. Users can also provide their version of the reference model and input datasets, but these must have the same gene annotations (e.g. Entrez IDs or HGNC symbols) if both are user-provided.

The third step analyzes disease-specific data from a user-provided case–control transcriptomics study using bulk RNA-seq or microarrays. COMO uses a platform-based standard pipeline to preprocess, normalize and identify differentially expressed genes by comparing control and patient groups. Ideally, the disease datasets come from the same cell types as the GSMM models, but if data for the specific cell types are unavailable, it advises mixed-cell populations data such as PBMCs for T cell lymphocytes.

The fourth step performs the drug perturbation simulation for drug discovery and consists of several substeps: (i) mapping of drug targets obtained from the repurposing tool of the ConnectivityMap database [16] to metabolic genes in the model; (ii) systematic knockouts of each mapped gene; (iii) comparison of flux profiles from perturbed and control models to identify differential fluxes; (iv) comparison of differentially regulated fluxes with differentially expressed genes identified in Step 3 to determine the number of genes up- and down-regulated in a disease whose

fluxes are in the opposite direction after perturbation with the mapped drug; and (v) computation of the perturbation effect score (PES) [17] for each drug target. The pipeline output includes PES-based ranks of drug targets and mapped repurposable drugs for the studied disease.

File preparation and omics data analysis

RNA-seq analysis

For RNA-seq analysis, COMO requires a matrix of gene counts from aligned FastQ reads for each context to build a context-specific constraint-based metabolic model. Users must place all the properly formatted files in designated user data directories to provide input to the COMO pipeline. The outputs of the pipeline will be saved in designated directories. COMO can integrate multiple types of RNA-seq data. The gene counts for each sample are normalized, filtered, binarized and combined. The output of this step is a CSV file containing all genes in all libraries with binary values for each gene, 0 for inactive and 1 for active (Supplementary Methods available online at http://bib.oxfordjournals.org/).

Microarray analysis

COMO can use microarray datasets from various platforms, including Affymetrix, Agilent and Illumina. The user provides a configuration file containing the GEO database's accession numbers grouped by cell types. The pipeline will automatically download datasets from the GEO database, normalize them and analyze them. Data across multiple samples are combined based on binarized expression values and user-defined ratios (Supplementary Methods available online at http://bib.oxfordjournals.org/).

Proteomics analysis

Protein abundance data can be used on its own to build context-specific GSMMs or integrated with transcriptomics data to enhance the model. Proteomics abundance data should be imported into COMO as a CSV or Excel file. Similar to the RNA-seq analysis, binarized activity states can be combined across samples using user-specified ratios (Supplementary Methods available online at http://bib.oxfordjournals.org/).

Combining activity of different 'omics' data sources

COMO merges activity states for any combination of available 'omics' data sources using binarized data. Users define a minimum activity requirement, which indicates the minimum number of provided sources in which a gene should be active for the gene to be active in the model (Supplementary Methods available online at http://bib.oxfordjournals.org/).

Context-specific model extraction

Context-specific models can be seeded from a general genome-scale model, such as Recon3D for human tissue models [15] or iMM1865 for mouse tissue models [18]. COMO currently supports GLPK (GNU Linear Programming Kit) [19] and GUROBI [20] solvers to subset and solve a context-specific model from the reference global model using a reconstruction algorithm (i.e. GIMME, iMAT and FASTCORE) implemented with Troppo [21] (Supplementary Methods available online at http://bib.oxfordjournals.org/).

Identifying disease-related genes

Disease-related genes can be determined using differential gene expression analysis (currently with transcriptomics data). COMO requires a configuration file specifying sample names and annotations for patient and control samples. This step outputs differentially expressed genes (Supplementary Methods available online at http://bib.oxfordjournals.org/).

Identifying drug targets and repurposable drugs

COMO maps drug targets obtained from the ConnectivityMap database [22] to metabolic genes in the model and performs systematic gene knockouts. It identifies differential fluxes by comparing flux profiles of perturbed and control models and comparing them with differentially expressed disease genes. This comparison determines the number of up- and down-regulated genes in a disease reversed by each mapped drug. These values are input for the PES [17] that can be calculated for each drug target (see Supplementary Methods available online at http://bib.oxfordjournals.org/). The output of this step includes PES-based ranks of drug targets and mapped repurposable drugs for the disease of interest.

All the available features in the COMO pipeline are given in Table 1.

B cell data preprocessing and analysis

We applied the COMO pipeline to drug discovery in B cells. We collected RNA-seq published in GEO [23] and processed the RNA-seq datasets using our accompanying alignment and quality control pipeline. We selected three RNA-seq datasets based on criteria defined in Supplementary Methods available online at http://bib.oxfordjournals.org/ for COMO to analyze and merge to construct context-specific models (Supplementary Table 1 available online at http://bib.oxfordjournals.org/). We also used published quantitative proteomics data from high-resolution mass spectrometry (Supplementary Table 1 available online at http://bib.oxfordjournals.org/).

We imported RNA-seq read counts and protein copy number data to the COMO container and processed them. The function 'rnaseq_preprocess.py' was used to generate count matrix files. The automatically generated configuration .xlsx files were modified to exclude RNA-seq datasets, which failed quality control checks (Supplementary Table 2 available online at http://bib.oxfordjournals.org/). The parameters used in B cell data analysis are shown in Supplementary Table 3 available online at http://bib.oxfordjournals.org/. The RNA-seq counts and protein copy number values were transformed to Z scores. Weights for RNA-seq-based Z-transformed scores and proteomics-based Z-transform scores were 6 and 10, respectively, to follow the 0.6 correlation coefficient found between transcription and protein expression [24].

RESULTS

Constructing metabolic models

We used the combined Z-transformed values based on transcriptomics and proteomics data of naive B cells (Figure 2A) to seed a model using the iMAT and GIMME algorithms. We used biomass maintenance for the objective function of naive B cells since they do not proliferate [25]. We defined exchange reactions based on literature about B cells' nutrient preferences (Supplementary Table 4 available online at http://bib.oxfordjournals.org/). Using combined RNA-seq and proteomics datasets, we generated many models with combined Z-transformed value cutoffs ranging from −6 to 4. To assess the functionality of the default zFPKM cutoff of −3 to −2 proposed by Hart *et al.* [26], we generated calibration models using GIMME and iMAT methods. We started from the lowest functional cutoff of >−4 created from binarized

Table 1: Available features in the COMO pipeline

Features in COMO	
Data integration (Docker container)	**Model seeding (Docker container)**
Microarray	iMAT
Bulk RNA-seq	FastCORE
Single-cell RNA-seq	GIMME
Protein abundance	
Peripheral analysis (Docker container)	**Peripheral analysis (HPC Snakemake pipeline)**
Differential gene expression	Fastq_screen integration
Drug target analysis (inhibition)	FastQC integration
Sample clustering	Trim_galore integration
Memote integration	Picard RNA-seqMetrics integration
Escher mapping integration	MultiQC integration
	STAR alignment integration

data for GIMME models. For iMAT calibration models, we used >−5 as a high threshold in combined Z-transform scores. A lower threshold of >−5 was used to prefer using genes that did not show near-zero expression. This will fulfill a minimum flux requirement through reactions associated with genes expressed above the higher threshold. The GIMME algorithm results in models with infeasible and dead-end reactions, which must be made fixed using an algorithm such as FASTCC [27]. The COBRApy version of FASTCC can be used in COMO. We used sanity checks [4] to determine baseline characteristics for each calibration model, including metabolic tasks and reactions contributing to biomass normalized by total feasible reactions.

We used results from physiological tests and sanity checks to evaluate the models. To choose the best model among the models generated from different multi-omics data cutoffs, we investigated (i) the number of feasible reactions (Figure 2B) and the number of metabolic tasks passed (Figure 2C) and (ii) the number of reactions in the final model that contributed to maximizing the objective functions (Figure 2C and Supplementary Table 5 available online at http://bib.oxfordjournals.org/). The ratio of metabolic tasks versus feasible reactions can help us identify the contextualized model with better retention of global metabolic functions while ensuring its reduction compared to the reference model (as the previously published CD4+ T cell models). A higher number of reactions contributing to the objective functions can indicate greater connectivity of reactions within the network. A Z-transformed value of −3 generated the best iMAT model with 4528 feasible reactions and 212 metabolic tasks passed (Supplementary Table 5 available online at http://bib.oxfordjournals.org/). A −3 cutoff has also generated the best GIMME model with 6999 reactions and 74 metabolic tasks passed (Supplementary Table 5 available online at http://bib.oxfordjournals.org/). Among iMAT and GIMME models, iMAT models showed better performance regarding the number of completed metabolic tasks and ATP production from different carbon sources (Supplementary Data 1 available online at http://bib.oxfordjournals.org/). For further analysis, we chose a model based on the iMAT algorithm, considering both the performance in metabolic tasks and the proportion of reactions contributing to the objective function relative to the set of feasible reactions.

Metabolic model of naive B cell

The selected iMAT-based naive B cell model passed all the sanity checks and completed 212 metabolic tests out of global 460 human metabolic tasks (Supplementary Table 5 available online at http://bib.oxfordjournals.org/). The model performance on metabolic tasks aligns with previously documented outcomes observed in other cell-specific metabolic models [28]. This model comprises 4528 reactions and 1463 genes (Supplementary Data 2 available online at http://bib.oxfordjournals.org/). 4064 reactions were internal enzyme–catalyzed and transport reactions distributed across 93 metabolic pathways in different compartments (extracellular, cytosol, mitochondria, endoplasmic reticulum, Golgi apparatus, lysosome, nucleus and peroxisome). Most reactions involved extracellular transport, followed by fatty acid oxidation, mitochondrial transport and peptide metabolism (Supplementary Figure 1 available online at http://bib.oxfordjournals.org/).

The model's performance was assessed through global metabolic tasks, but it is imperative to conduct further validation involving B cell–specific metabolic functions. We performed a thorough literature search to collect information available on naive B cell metabolism and validated the naive B cell model by comparing it. We used flux balance analysis (FBA) to identify active pathways and minimization of metabolic adjustment (MoMA) to simulate knockouts. The results presented in Table 2 highlight the agreement between our model and the established biological scenarios in the literature. Among the tested seven complex scenarios, five were in full agreement, and two partially agreed, consistent with previously published literature [17, 29]. Several major pathways agreed with the literature, including glycolysis, lactate production, the TCA cycle and glucose transport (Figure 3). Additionally, in agreement with the literature, our model did not show any impact of glucose removal from the exchange media on growth (Figure 4A). The literature has shown that inhibition of HIF1 induces apoptosis and decreases mature B cells [30]. HIF-1 increases the expression of glycolytic enzymes [31]. We knocked out its target genes to simulate HIF-1's effect on B cells. Among all tested genes, GAPDH knockout reduced B cell growth to zero, while other enzymes affected growth but not significantly (Figure 4B). This approach allowed us to validate our models against known biological scenarios of naive B cell metabolic pathways.

Disease gene analysis

We collected a publicly available case–control RNA-seq dataset (GSE110999) of naive B cells for RA and SLE) (Supplementary Table 6 available online at http://bib.oxfordjournals.org/). We used COMO's count matrix and configuration files and identified differentially expressed genes (adjusted P-values < 0.05)

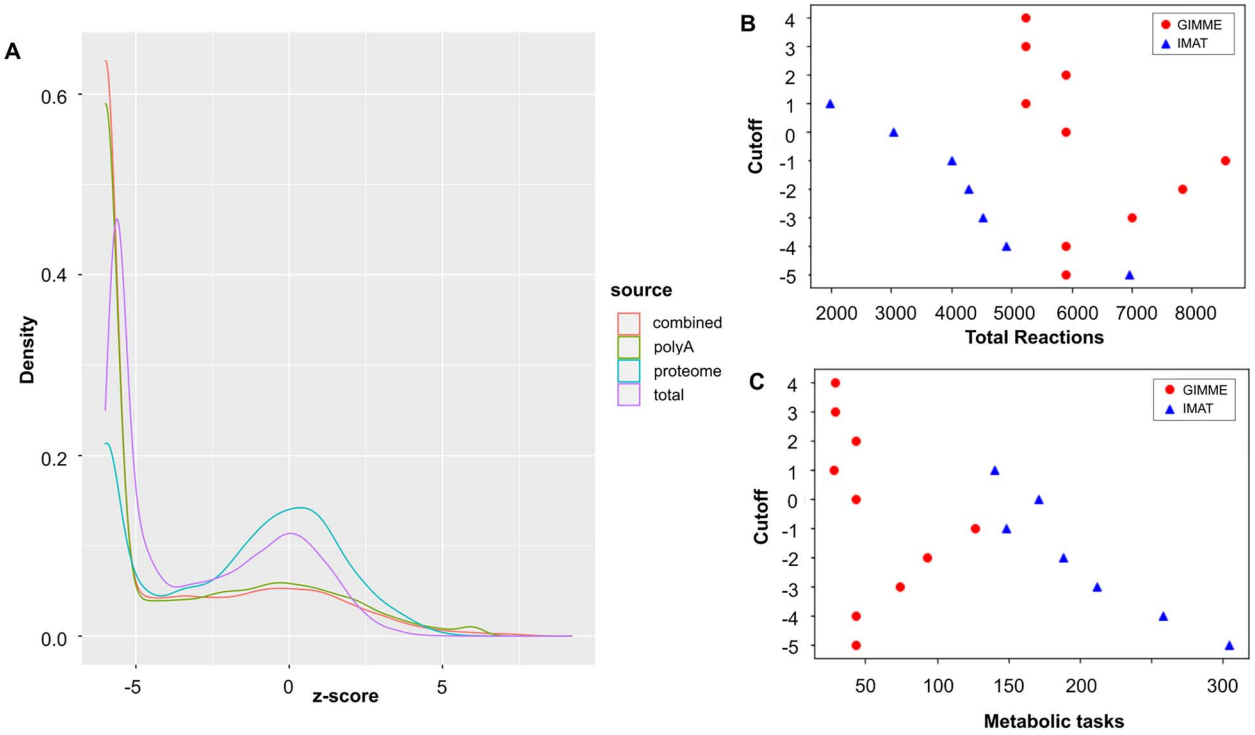

Figure 2. Combined Z-transform scores and their effect on model building and performance. (**A**) Combined Z-transformed scores for different B cell datasets. (**B**) Effect of Z-score cutoffs on the total number of reactions in the model constructed using GIMME and iMAT algorithms. (**C**) Number of metabolic tasks completed by GIMME- and iMAT-based models constructed with different Z-score cutoffs.

Table 2: Model validation using specific behaviors

#	Observation from literature	*In silico* experiment	Simulation result	Agreement
1	Glucose restriction has no impact on naive B cell function [55].	Decreased glucose uptake rate and studied its effect on naive B cells growth	No effect on growth	Yes
2	Glucose distributes through G6P, F16BP, G3P and 3PG [55].	Simulated the model and checked the flux through reactions producing these metabolites	The model produced all the metabolites	Yes
3	Naive B cells produce lactate [55].	Checked flux through reactions producing lactate	Model produced lactate	Yes
4	In naive B cells, citrate, glutamate, glutamine, α-ketoglutarate, succinate, fumarate and malate are active [55].	Checked the flux inTCA pathway producing these metabolites	Model produced all the metabolites of TCA cycle	Yes
5	Untreated or anti-IgM treated B cells showed no significant increase in the ECAR in response to the glucose addition [30, 56].	Increased glucose uptake and showed its effect on the naive B cells lactate production, which is directly related to ECAR	Lactate production was unaffected by more availability of glucose	Yes
6	B cell survival requires proper maintenance of glycolytic and oxidative glucose pathways. Inhibition of HIF1 induces apoptosis and decreases mature B cells [30].	Inhibited glycolytic enzymes that are targets of HIF1 and checked their effect on B cell growth	Only GAPD and PGK significantly reduced the B cell growth	Yes (partial)
7	Naive B cells seem to rely on FA as their main fuel [57].	Removed fatty acids uptake and investigated the effect on B cell growth	A slight decrease was observed in growth when fatty acid uptake was blocked, specifically under glucose-depleted conditions	Yes (partial)

(Supplementary Table 3 available online at http://bib.oxfordjournals.org/).

Drug target analysis

We mapped drug targets in the B cell model using connectivity-Map's list of drugs and their targets. We simulated knockouts by inhibiting existing drugs' metabolic targets, compared the simulation results with genes up- and down-regulated in RA and SLE and computed the PES [17] (Supplementary Methods available online at http://bib.oxfordjournals.org/). Finally, we used genes with a positive PES to predict potential targets and their associated drugs as potential candidates for repurposing that were validated through literature mining.

A positive PES indicates that reactions associated with differentially expressed genes—when inhibited by drugs—have fluxes in directions opposite to the differential expression. Thus, reactions

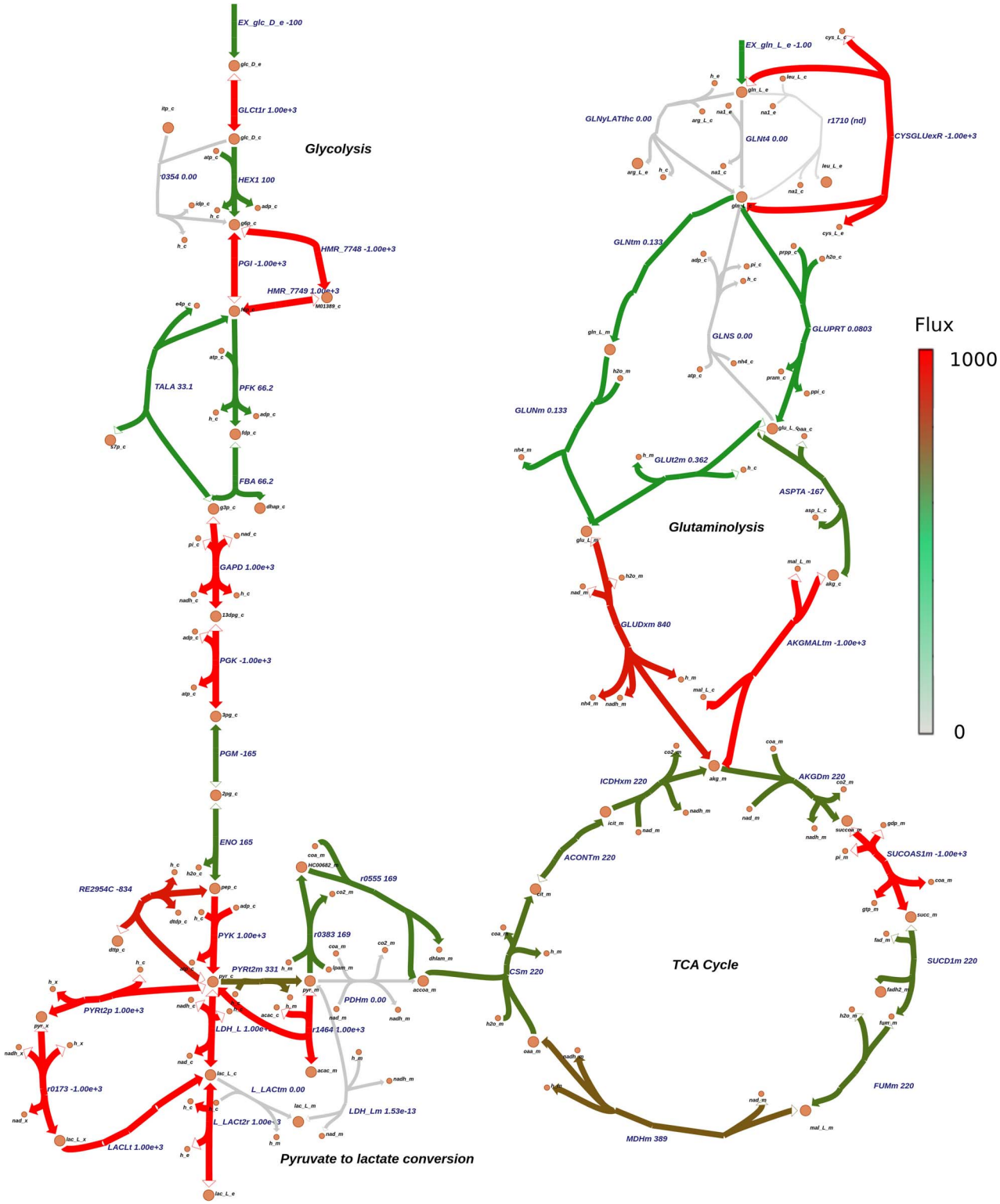

Figure 3. Flux distribution through glycolysis, TCA cycle and glutaminolysis pathways.

associated with up-regulated genes have reduced flux, while reactions associated with down-regulated genes have increased flux. A higher PES indicates that the gene's perturbation has a stronger effect on differentially expressed genes. We identified 44 genes for RA and 52 for SLE with positive PES, which were associated with 123 and 120 drugs, respectively (Supplementary Data 3 available online at http://bib.oxfordjournals.org/).

Among the potential targets identified for RA, APRT (coagulation factor inhibitor) was ranked the highest, followed by GLB1, CTSA, ODC1, TBXAS1, KHK, ADK and ALDH5A1 (Supplementary Table 7 available online at http://bib.oxfordjournals.org/). We identified five genes (TBXAS1—rank 5, ABCB1—rank 10, HPRT1—rank 13, SLC10A1—rank 33, SOAT1—rank 40) targeted by the drugs already used to treat RA or are under study for RA.

Table 3: Predicted drug targets validated with literature

Disease	Drug target	Support
RA	TBXAS1 (Thromboxane A Synthase 1)	Potential target of the anti-rheumatic drug sulfasalazine [58]
	ABCB1 (ATP binding cassette subfamily B member 1)	Target of macrolides used as RA treatment option [6, 42]
	HPRT1 (hypoxanthine phosphoribosyltransferase 1)	Target of azathioprine, and mercaptopurine used for RA treatment [59, 60]
	PPAT (phosphoribosyl pyrophosphate amidotransferase)	
	SLC10A1 (solute carrier family 10 member 1)	Target of disease modifying anti-rheumatic drug cyclosporin-A [61]
SLE	PPAT (phosphoribosyl pyrophosphate amidotransferase)	Target of the drug azathioprine used in SLE treatments. [62, 63]
	DHFR (dihydrofolate reductase)	Target of the drug methotrexate, used to treat SLE. [6]
	SLC10A1 (solute carrier family 10 member 1)	Target of immunosuppressant drug cyclosporin-A [64]

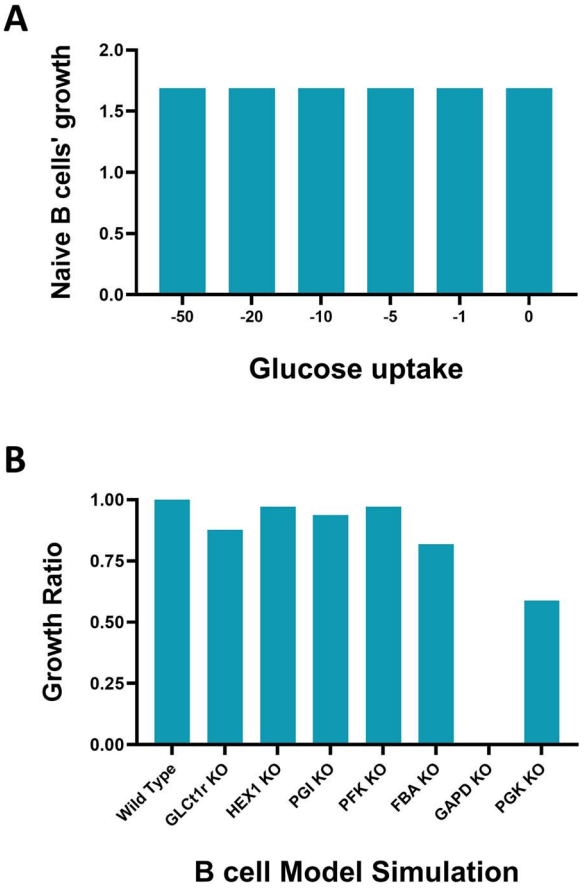

Figure 4. Validation of B cell model. (**A**) Limiting glucose in media has no impact on naive B cell growth. (**B**) Impact of glycolytic enzymes knockout (growth ratio compared to wild type). All the tested reactions showed some impact on B cell growth.

Three targets (TBXAS1, ABCB1, HPRT1) had high-positive PES. Our literature search identified seven RA drugs targeting metabolism: sulfasalazine, azathioprine, mercaptopurine, cyclosporin-A, macrolides, methotrexate and leflunomide. We identified targets of five of these as drug targets based on positive PES (Table 3). The targets of methotrexate (DHFR) and leflunomide (DHODH) had negative PES and thus were not predicted as drug targets. The new potential drug targets included GLB1 (galactosidase beta 1), ODC1 (ornithine decarboxylase) and ALDH5A1 (aldehyde dehydrogenase 5 family member A1).

For SLE, GGPS1 (geranylgeranyl diphosphate synthase 1), GSR (glutathione-disulfide reductase) and GLB1 (galactosidase beta 1) were ranked the top three drug targets (Supplementary Table 7 available online at http://bib.oxfordjournals.org/). PPAT (phosphoribosyl pyrophosphate amidotransferase; rank 7), SLC10A1 (solute carrier family 10 member 1; Rank 19) and DHFR (dihydrofolate reductase; rank 23) are the targets of drugs used for SLE (Table 3). Our literature search for SLE drugs targeting metabolic genes identified four drugs (azathioprine, methotrexate, cyclosporin-A and leflunomide). Targets of three SLE drugs (azathioprine, methotrexate and cyclosporin-A) were identified as drug targets (Table 3). The target of immunosuppressant drug leflunomide was identified with negative PES, thus not predicted as a drug target. The other drug targets included GGPS, GSR and ALDH5A1.

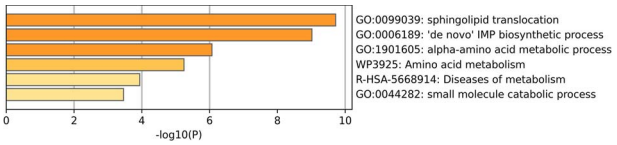

Figure 5. Biological processes and pathways enriched in common drug targets between RA and SLE.

Next, we investigated the biological processes that are common in both diseases to understand the similarities of target biological processes in B cells across two autoimmune diseases. Between RA and SLE, we identified 12 common drug targets in the top-ranked list (Supplementary Table 7 available online at http://bib.oxfordjournals.org/). The gene enrichment analysis [32] showed that the sphingolipid translocation (ABCA1, ABCC1, ABCB1, GLB1, CTSA) and the *de novo* IMP biosynthetic process (ATIC, GART, PPAT, GSR, ABCC1, GCLM) were enriched highest in the commonly found genes (Figure 5).

DISCUSSION

Dysregulation of metabolic pathways is often associated with various diseases, making them attractive targets for drug development [33, 34]. For example, disrupted metabolic pathways in autoimmune diseases lead to producing harmful immune cells or activating inflammatory pathways [34]. Computational modeling of metabolism can help simulate the effects of gene or reaction perturbations *in silico* to identify potential drug targets. However, the systematic analysis of such models requires integrating various modeling methods and biological datasets.

COMO is a user-friendly pipeline that processes and integrates transcriptomics and proteomics data, builds context-specific metabolic models, runs simulations, investigates gene inhibition's effect and identifies repurposable drugs and drug targets. COMO uses open-source solutions to facilitate cost-effective disease treatment exploration by researchers. We previously used this approach in identifying drug targets in CD4+ T cell metabolism

[17]. COMO is scalable and robust, accommodating single-cell RNA-seq, bulk RNA-seq and proteomics data, as well as microarrays. The integration of multi-omics data allows for a more comprehensive understanding of the metabolic state of a cell or tissue and can provide additional insights into the effects of perturbations on the cell's phenotype.

COMO also presents some limitations. COMO relies on the accuracy and completeness of the available omics data and reference models. If the data are noisy or incomplete, it can introduce biases leading to inaccurate predictions. The noisy data may inherit biases from the original datasets, such as over or under-representation of specific pathways, which could skew the analysis and increase the chance of false positives and false negatives in the target identification. COMO is also limited to the scope of the reference models and databases it interfaces with and will not be able to identify targets or repurposable drugs that are not included in these resources. Addressing these limitations requires rigorous validation and preprocessing of the input data to ensure it is of high quality and suitable for use in the pipeline.

GSMMs can identify potential therapeutic targets by simulating perturbations in metabolic pathways and analyzing the resulting changes in the cell's phenotype [17, 35]. To identify drug targets, COMO integrates perturbation-induced flux changes with differentially expressed genes in diseases. Finding metabolic fluxes of up- and down-regulated disease genes closer to the non-disease baseline is based on the assumption that differential regulation of such genes impacts metabolic fluxes. In previous studies, it has been found that the relationship between gene expression and metabolic fluxes is rather complex [36]. However, many studies have showcased the utility of integrating—omics data with metabolic fluxes in various contexts [17, 29, 37–39]. By using differential gene expression with changes in metabolic fluxes, our group identified drug targets in CD4 T cells against autoimmune diseases, validated by *in vitro* experiments [17]. In a recent study, Kaste and Shachar-Hill used transcriptomic and proteomic-derived gene expression levels and integrated them with flux prediction, significantly improving the model's agreement with experimental data [29]. Another study combines genetic and environmental perturbation-based differential gene expression and metabolite abundance data with differential fluxes to analyze the altered physiological states [38]. Furthermore, a new approach was proposed incorporating data on differential gene expression to assess metabolic flux differences between two distinct conditions [39].

We observed notable proportions of existing metabolic targets in the predicted drug targets compared to metabolic genes not predicted as potential drug targets for both RA and SLE. This observation validates the tool's significance in drug target prediction. In drug targets identified against RA, literature evidence supports TBXAS1 (thromboxane A synthase 1), ABCB1 (ATP binding cassette subfamily B member 1) and HPRT1 (hypoxanthine phosphoribosyltransferase 1). TBXAS1 is a target of sulfasalazine, an older disease-modifying anti-rheumatic drug prescribed as a second- or third-line defense for RA. RA patients show significantly higher levels of thromboxane B2 (TBXAS1 product) in their serum compared to healthy controls [40]. ABCB1 is targeted by macrolides like erythromycin used in RA treatment. Initially, the control of periodontal bacteria was believed to be the reason, potentially contributing to RA progression [41]. However, subsequent studies [42] suggest an anti-inflammatory mechanism, as discontinuing clarithromycin leads to a rapid rebound of RA symptoms. HPRT1 is targeted by azathioprine and mercaptopurine, effective treatments for RA and SLE. Metabolites of each prodrug enter the cell [43] and undergo breakdown in the purine salvage pathway, inhibiting purine synthesis [44], possibly through non-competitive inhibition of phosphoribosylpyrophosphate amidotransferase (PPAT) [45].

In SLE drug targets, PPAT and DHFR (dihydrofolate reductase) were supported by the literature. PPAT is involved in *de novo* purine biosynthesis and is a target of azathioprine, an immunosuppressive drug used in SLE treatments. However, this drug has not been studied in B cells. DHFR plays a role in the metabolism of folic acid and is an essential enzyme that converts the nutrient folate into its active form. DHFR is known to be a target of the drug methotrexate, used to treat SLE [46]. However, DHFR has not been investigated as a drug target for B cells in SLE.

Among RA targets, several are supported by literature experiments demonstrating the anti-inflammatory effect on various immune cells. GLB1 (galactosidase beta 1) encodes an enzyme that hydrolyzes terminal beta-linked galactose residue from ganglioside substrates and other glycoconjugates. D-Fagomine has been found to control inflammatory processes related to an overactivation of the humoral immune response [47]. ODC1 (ornithine decarboxylase) catalyzes ornithine decarboxylation. In type 3 innate lymphoid cells (ILC3s), deficiency of ODC1 reduced IL-22 production and protected mice from colitis [48]. ALDH5A1 breaks down GABA (gamma-aminobutyric acid), and inhibiting it increases available GABA, potentially suppressing the immune response as B cells produce GABA [49].

In SLE targets, many have been previously studied and linked with SLE. GGPS1 (geranylgeranyl diphosphate synthase 1) is involved in isoprenoid synthesis, a pathway up-regulated in SLE patients [50]. However, GGPS1's role as a drug target in B cells remains unexplored. GSR (glutathione-disulfide reductase) catalyzes the reduction of glutathione disulfide into glutathione (GSH), an antioxidant crucial for protecting cells from oxidative stress. A decrease in GSH increases ROS production and affects T cell differentiation [51]. ALDH5A1 (aldehyde dehydrogenase 5 family member A1) and ABAT (4-aminobutyrate aminotransferase) break down GABA. ALDH5A1 inhibition makes more GABA available, potentially suppressing the immune response [49]. LTA4H (leukotriene A4 hydrolase) converts leukotriene A4 (LTA4) to leukotriene B4 (LTB4), a potent proinflammatory mediator [52]. FADS1 (fatty acid desaturase 1) regulates eicosanoid production, proinflammatory mediators linked to SLE pathogenesis [53]. ACACB (acetyl-CoA carboxylase beta) study suggested the role of ACACB in the pathogenesis of SLE [54].

In summary, using metabolic pathways as drug targets can provide new therapies for complex diseases. COMO can aid in identifying potential drug targets and repurposable drugs. The outputs from COMO, including potential drug targets and repurposable drugs, can be integrated into drug discovery pipelines, involving initial *in silico* studies for target confirmation, followed by *in vitro* and animal model experiments for efficacy and safety assessments and advancing the candidates through clinical trials.

Key Points

- Provided a generalized pipeline for constructing metabolic models utilizing bulkRNA-seq, single-cell RNA-seq, proteomics and microarray datasets, implementing them for drug discovery.

- Developed a genome-scale metabolic model of B cells using transcriptomics and proteomics data and characterized metabolic landscapes through flux-based analysis.
- Proposed novel metabolic targets suitable for drug development in the treatment of rheumatoid arthritis and systemic lupus erythematosus, identified through combined *in silico* perturbation of models and clinical data.
- Offered a prioritized list of new drug targets for the mentioned diseases, validated through literature mining along with drugs with potential repurposing.

SUPPLEMENTARY DATA

Supplementary data are available online at https://academic.oup.com/bib.

FUNDING

This material is based upon work supported by the Defense Health Agency and U.S. Strategic Command under Contract No. FA4600-12-D-9000 and Contract No. FA4600-18-D-9001, and in part by the National Institutes of Health, grant #R35GM119770, and by the Layman seed grant by University of Nebraska Foundation. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Defense Health Agency, U.S. Strategic Command, or 55th Contracting Squadron. Similarly, the opinions or assertions contained herein are the private views of the authors and are not necessarily those of the Uniformed Services University of the Health Sciences, or the Department of Defense. The mention of specific therapeutic agents does not constitute endorsement by the U.S. Department of Defense, and trade names are used only for the purpose of clarification.

DATA AVAILABILITY

The pipeline code is available at https://github.com/HelikarLab/COMO, and the other data generated in this study are available as Supplementary Data available online at http://bib.oxfordjournals.org/.

REFERENCES

1. DeBerardinis RJ, Thompson CB. Cellular metabolism and disease: what do metabolic outliers teach us? *Cell* 2012;**148**:1132–44.
2. Jerby L, Ruppin E. Predicting drug targets and biomarkers of cancer via genome-scale metabolic modeling. *Clin Cancer Res* 2012;**18**:5572–84.
3. Uhlen M, Zhang C, Lee S, *et al.* A pathology atlas of the human cancer transcriptome. *Science* 2017;**357**:eaan2507.
4. Heirendt L, Arreckx S, Pfau T, *et al.* Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat Protoc* 2019; **14**:639–702
5. Ebrahim A, Lerman JA, Palsson BO, *et al.* COBRApy: COnstraints-based reconstruction and analysis for python. *BMC Syst Biol* 2013;**7**:74.
6. Wishart DS, Feunang YD, Guo AC, *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;**46**:D1074–82.
7. Dash S, Shakyawar SK, Sharma M, *et al.* Big data in healthcare: management, analysis and future prospects. *J Big Data* 2019; **6**:54.
8. Lieven C, Beber ME, Olivier BG, *et al.* MEMOTE for standardized genome-scale metabolic model testing. *Nat Biotechnol* 2020;**38**:272–6.
9. King ZA, Dräger A, Ebrahim A, *et al.* Escher: a web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLoS Comput Biol* 2015;**11**:e1004321.
10. McInnes L, Healy J, Saul N, *et al.* UMAP: uniform manifold approximation and projection. *J Open Source Softw* 2018;**3**:861.
11. Shi G, Liao X, Lin Z, *et al.* Estimation of the global prevalence, incidence, years lived with disability of rheumatoid arthritis in 2019 and forecasted incidence in 2040: results from the Global Burden of Disease Study 2019. *Clin Rheumatol* 2023;**42**:2297–309.
12. Tian J, Zhang D, Yao X, *et al.* Global epidemiology of systemic lupus erythematosus: a comprehensive systematic analysis and modelling study. *Ann Rheum Dis* 2023;**82**:351–6.
13. Torigoe M, Iwata S, Nakayamada S, *et al.* Metabolic reprogramming commits differentiation of human CD27+IgD+ B cells to plasmablasts or CD27-IgD- cells. *J Immunol Baltim Md* 1950;**2017**(199):425–34.
14. Atisha-Fregoso Y, Toz B, Diamond B. Meant to B: B cells as a therapeutic target in systemic lupus erythematosus. *J Clin Invest* 2021;**131**:e149095.
15. Brunk E, Sahoo S, Zielinski DC, *et al.* Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat Biotechnol* 2018;**36**:272–81.
16. Corsello SM, Bittker JA, Liu Z, *et al.* The drug repurposing hub: a next-generation drug library and information resource. *Nat Med* 2017;**23**:405–8.
17. Puniya BL, Amin R, Lichter B, *et al.* Integrative computational approach identifies drug targets in CD4 + T-cell-mediated immune disorders. *Npj Syst Biol Appl* 2021;**7**:1–18.
18. Khodaee S, Asgari Y, Totonchi M, *et al.* iMM1865: a new reconstruction of mouse genome-scale metabolic model. *Sci Rep* 2020;**10**:6177.
19. Oki E. Basics of Linear Programming. In: Eiji Oki (ed) *Linear Programming and Algorithms for Communication Networks*. Boca Raton: CRC Press, 2012.
20. Gurobi OL. *Gurobi Optimizer Reference Manual*, 2023.
21. Ferreira J, Vieira V, Gomes J, *et al.* Troppo—a python framework for the reconstruction of context-specific metabolic models. *Pract Appl Comput Biol Bioinforma 13th Int Conf* 2020;**1005**:146–53.
22. Subramanian A, Narayan R, Corsello SM, *et al.* A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 2017;**171**:1437–1452.e17.
23. Barrett T, Wilhite SE, Ledoux P, *et al.* NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res* 2013;**41**:D991–5.
24. Buccitelli C, Selbach M. mRNAs, proteins and the emerging principles of gene expression control. *Nat Rev Genet* 2020;**21**:630–44.
25. Moscardó García M, Pacheco M, Bintener T, *et al. Importance of the biomass formulation for cancer metabolic modeling and drug prediction* iScience 2021; **24**:103110.
26. Hart T, Komori HK, LaMere S, *et al.* Finding the active genes in deep RNA-seq gene expression studies. *BMC Genomics* 2013;**14**:778.

27. Vlassis N, Pacheco MP, Sauter T. Fast reconstruction of compact context-specific metabolic network models. *PLoS Comput Biol* 2014;**10**:e1003424.
28. Thiele I, Swainston N, Fleming RMT, *et al.* A community-driven global reconstruction of human metabolism. *Nat Biotechnol* 2013;**31**:419–25.
29. Kaste JAM, Shachar-Hill Y. Accurate flux predictions using tissue-specific gene expression in plant metabolic modeling. *Bioinformatics* 2023;**39**:btad186.
30. Boothby MR, Brookens SK, Raybuck AL, *et al.* Supplying the trip to antibody production—nutrients, signaling, and the programming of cellular metabolism in the mature B lineage. *Cell Mol Immunol* 2022;**19**:352–69.
31. Taylor CT, Scholz CC. The effect of HIF on metabolism and immunity. *Nat Rev Nephrol* 2022;**18**:573–87.
32. Zhou Y, Zhou B, Pache L, *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 2019;**10**:1523.
33. Du D, Liu C, Qin M, *et al.* Metabolic dysregulation and emerging therapeutical targets for hepatocellular carcinoma. *Acta Pharm Sin B* 2022;**12**:558–80.
34. Piranavan P, Bhamra M, Perl A. Metabolic targets for treatment of autoimmune diseases. *Immunometabolism* 2020;**2**: e200012.
35. Turanli B, Zhang C, Kim W, *et al.* Discovery of therapeutic agents for prostate cancer using genome-scale metabolic modeling and drug repositioning. *EBioMedicine* 2019;**42**:386–96.
36. Tian M, Reed JL. Integrating proteomic or transcriptomic data into metabolic models using linear bound flux balance analysis. *Bioinforma Oxf Engl* 2018;**34**:3882–8.
37. Jenior ML, Moutinho Jr TJ, Dougherty BV, Papin JA. Transcriptome-guided parsimonious flux analysis improves predictions with metabolic networks in complex environments. *PLoS Comput Biol* 2020;**16**:e1007099.
38. Pandey V, Hadadi N, Hatzimanikatis V. Enhanced flux prediction by integrating relative expression and relative metabolite abundance into thermodynamically consistent metabolic models. *PLoS Comput Biol* 2019;**15**:e1007036.
39. Ravi S, Gunawan R. ΔFBA-predicting metabolic flux alterations using genome-scale metabolic models and differential transcriptomic data. *PLoS Comput Biol* 2021;**17**:e1009589.
40. Wang M-J, Huang Y, Huang R-Y, *et al.* Determination of role of thromboxane A2 in rheumatoid arthritis. *Discov Med* 2015;**19**: 23–32.
41. Ogrendik M. Antibiotics for the treatment of rheumatoid arthritis. *Int J Gen Med* 2013;**7**:43–7.
42. Saviola G, Benucci M, Cirino G. Comments on:Effects of clarithromycin in patients with active rheumatoid arthritis. *Curr Med Res Opin* 2007; **23**:2763–4
43. Fotoohi AK, Lindqvist M, Peterson C, *et al.* Involvement of the concentrative nucleoside transporter 3 and equilibrative nucleoside transporter 2 in the resistance of T-lymphoblastic cell lines to thiopurines. *Biochem Biophys Res Commun* 2006;**343**: 208–15.
44. Karran P, Attard N. Thiopurines in current medical practice: molecular mechanisms and contributions to therapy-related cancer. *Nat Rev Cancer* 2008;**8**:24–36.
45. Tay BS, *et al.* Inhibition of phosphoribosyl pyrophosphate amidotransferase from Ehrlich ascites-tumour cells by thiopurine nucleotides. *Biochem Pharmacol* 1969;**18**:936–8.
46. Bedoui Y, Guillot X, Sélambarom J, *et al.* Methotrexate an old drug with new tricks. *Int J Mol Sci* 2019;**20**:5023.
47. Segura SP. Torras MAC. D-Fagomine for the control of inflammatory processes related to an overactivation of the humoral immune response. 2014. https://patents.google.com/patent/US20150368283A1/enPub.
48. Peng V, Cao S, Trsan T, *et al.* Ornithine decarboxylase supports ILC3 responses in infectious and autoimmune colitis through positive regulation of IL-22 transcription. *Proc Natl Acad Sci U S A* 2022;**119**:e2214900119.
49. Zhang B, Vogelzang A, Miyajima M, *et al.* B cell-derived GABA elicits IL-10+ macrophages to limit anti-tumour immunity. *Nature* 2021;**599**:471–6.
50. Kurup RK, Kurup PA. Hypothalamic-mediated model for systemic lupus erythematosis: relation to hemispheric chemical dominance. *Int J Neurosci* 2003;**113**:1561–77.
51. Lian G, Gnanaprakasam JR, Wang T, *et al.* Glutathione de novo synthesis but not recycling process coordinates with glutamine catabolism to control redox homeostasis and directs murine T cell differentiation. *Elife* 2018;**7**:e36158.
52. Fourie AM. Modulation of inflammatory disease by inhibitors of leukotriene A4 hydrolase. *Curr Opin Investig Drugs Lond Engl* 2000;**2009**(10):1173–82.
53. Gromovsky AD, Schugar RC, Brown AL, *et al.* Δ-5 fatty acid desaturase FADS1 impacts metabolic disease by balancing proinflammatory and proresolving lipid mediators. *Arterioscler Thromb Vasc Biol* 2018;**38**:218–31.
54. Harley I, Schulz H, Cambier J, *et al.* De novo mutation in ACACB in childhood onset SLE highlights a novel role as modulator of nucleic acid sensor-driven type I interferon responses. ACR. *Meet Abstr* 2017;**69**(Suppl 10).
55. Waters LR, Ahsan FM, Wolf DM, *et al.* Initial B cell activation induces metabolic reprogramming and mitochondrial remodeling. *iScience* 2018;**5**:99–109.
56. Akkaya M, Traba J, Roesler AS, *et al.* Second signals rescue B cells from activation-induced mitochondrial dysfunction and death. *Nat Immunol* 2018;**19**:871–84.
57. Wilson CS, Moore DJ. B cell metabolism: an understudied opportunity to improve immune therapy in autoimmune type 1 diabetes. *Immunometabolism* 2020;**2**:e200016.
58. Stenson WF, Lobos E. Inhibition of platelet thromboxane synthetase by sulfasalazine. *Biochem Pharmacol* 1983;**32**:2205–9.
59. Dean L, Kane M. In: Victoria M. Pratt, Stuart A. Scott, Munir Pirmohamed, Bernard Esquivel, Brandi L. Kattman, Co-editor, Adriana J. Malheiro (eds) *Mercaptopurine Therapy and TPMT and NUDT15 Genotype*. Summ: Med. Genet, 2012.
60. Suarez-Almazor ME, Spooner C, Belseck E. Azathioprine for treating rheumatoid arthritis. *Cochrane Database Syst Rev* 2000; **4**:CD001461.
61. Wells GA, Haguenauer D, Shea B, *et al.* Cyclosporine for treating rheumatoid arthritis. *Cochrane Database Syst Rev* 1998;**2**:CD001083.
62. Abu-Shakra M, Shoenfeld Y. xAzathioprine therapy for patients with systemic lupus erythematosus. *Lupus* 2001;**10**:152–3.
63. Zhou Y, Zhang Y, Lian X, *et al.* Therapeutic target database update 2022: facilitating drug discovery with enriched comparative data of targeted agents. *Nucleic Acids Res* 2022;**50**:D1398–407.
64. Germano V, Picchianti Diamanti A, Ferlito C, *et al.* Cyclosporine A in the long-term management of systemic lupus erythematosus. *J Biol Regul Homeost Agents* 2011;**25**:397–403.