A guide for developing comprehensive systems biology maps of disease mechanisms: planning, construction and maintenance

Alexander Mazein[1]*, Marcio Luis Acencio[1], Irina Balaur[1], Adrien Rougny[2], Danielle Welter[1], Anna Niarakis[3,4], Diana Ramirez Ardila[5], Ugur Dogrusoz[6], Piotr Gawron[1], Venkata Satagopam[1,7], Wei Gu[1,7], Andreas Kremer[5], Reinhard Schneider[1,7] and Marek Ostaszewski[1,7]*

[1]Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Esch-sur-Alzette, Luxembourg, [2]Independent Researcher, Massy, France, [3]Université Paris-Saclay, Laboratoire Européen de Recherche Pour la Polyarthrite Rhumatoïde−Genhotel, University Evry, Evry, France, [4]Lifeware Group, Inria Saclay-Ile de France, Palaiseau, France, [5]ITTM Information Technology for Translational Medicine, Esch-sur-Alzette, Luxemburg, [6]Computer Engineering Department, Bilkent University, Ankara, Türkiye, [7]ELIXIR Luxembourg, Belvaux, Luxembourg

As a conceptual model of disease mechanisms, a disease map integrates available knowledge and is applied for data interpretation, predictions and hypothesis generation. It is possible to model disease mechanisms on different levels of granularity and adjust the approach to the goals of a particular project. This rich environment together with requirements for high-quality network reconstruction makes it challenging for new curators and groups to be quickly introduced to the development methods. In this review, we offer a step-by-step guide for developing a disease map within its mainstream pipeline that involves using the CellDesigner tool for creating and editing diagrams and the MINERVA Platform for online visualisation and exploration. We also describe how the Neo4j graph database environment can be used for managing and querying efficiently such a resource. For assessing the interoperability and reproducibility we apply FAIR principles.

KEYWORDS

disease mechanisms, curation, pathway biology, systems biology, translational research

1 Introduction

Disease maps have been introduced in systems biomedicine as conceptual models of disease mechanisms (Mazein et al., 2018; Ostaszewski et al., 2018). In a disease map, interconnected pathway diagrams depict events from the level of molecular processes to higher levels like communication between cells and organs (Mazein et al., 2018; Ostaszewski et al., 2018; Mazein et al., 2021a). Approaches to the construction of disease maps advanced during past years, supported by the activity of the Disease Maps community (https://disease-maps.io) and formalised into best practices (Kondratova et al., 2018). However, there is still a lack of clear, formal and well-structured guidelines for main steps of developing a disease map in collaborative projects.

The need for such guidelines became evident during the development of the COVID-19 Disease Map (Ostaszewski et al., 2021), a large-scale project supported by an extensive community with more than 200 volunteer participants. The efforts of COVID-19 Disease Map Community would be greatly advanced with step-by-step guidelines to accelerate learning about the approach and creating high-quality resources.

There are a number of publications that can be used as a reference when constructing disease maps (Kondratova et al., 2018; Mazein et al., 2018). Moreover, brief and informative guideline articles about biocuration (Tang et al., 2019) and pathway construction (Hanspers et al., 2021) offer valuable introductory material for constructing disease-oriented resources. In particular, an article on construction of a Reactome-based specific knowledge repository (Varusai et al., 2020) proposes instructions on reuse of pathway diagrams in a specific context. However, this subject still lacks a comprehensive summary, encompassing the entire lifecycle of a disease map resource.

In this article, we offer guidelines for building diagrammatic systems biology knowledge repositories of disease mechanisms that are explorable online and transformable to computational models. In particular, we discuss planning and design of a disease map including best practices for defining the scope of the resource and communication with domain experts. We discuss the process of biocuration, the required granularity of the map and the required standard languages. Finally, we outline aspects of visual exploration and possible downstream bioinformatic analyses, and the strategy for maintaining the resource following FAIR principles (Wilkinson et al., 2016). We review the literature and examples for a particular area of disease map construction, structuring them into steps of a reference workflow. The methods described in this review are built on the experience of the previously developed disease maps including recently published resources (Ravel et al., 2020; Mazein et al., 2021a; Pereira et al., 2021; Zerrouk et al., 2022). Definitions of domain-specific terms are available at https://disease-maps.io/glossary.

2 Disease map development workflow

The major phases of map development are illustrated in Figure 1. The planning phase includes i) determining map purpose and type, ii) defining the coverage of biological functions, iii) choosing a graphical standard (data model) and an editor, and iv) designing map architecture (Figure 1, Steps 1–4). While these steps are shown in a sequence, they are often realised in parallel. Next, in the curation phase, available knowledge from literature and databases is encoded into a diagram with accompanying annotation of map entities (Figure 1, Steps 5 and 6). When the map is ready for publication, it is made available online in a navigable tool and updated based on feedback from domain experts—the publishing phase (Figure 1, Steps 7 and 8). The application phase depends on the goals of the project, expertise and resources available, and can include data visualisation, drug target exploration, computational modelling and hypothesis generation (Figure 1, Step 9).

2.1 Determining map purpose and type

At the initial step, it is important to define goals of the project and to determine the map type and its applications. Later, it will influence map's format, design, size and coverage. The key questions at this stage are:

– What is the main focus? Is it one disease or multiple related disorders?
– What is the intended use of the map?
– What resources are available or needed?
– What is the timeline?
– Which domain experts (clinicians and biologists) can advise and revise the curation?

Examples of motivation, objectives and primary users for several disease map projects are shown in Table 1. Often they are aligned

FIGURE 1
The map development workflow [based on Kondratova et al., 2018 (Kondratova et al., 2018)]. The workflow includes nine steps in four phases (please see in the text). Curators and domain experts are involved in planning and designing the map, then curators develop the map using a diagram editor of choice, advised by domain experts. After publishing online, data analysts, modellers and other users have access to the map for exploration and applications in translational projects, as well as domain experts suggest further evolving the map.
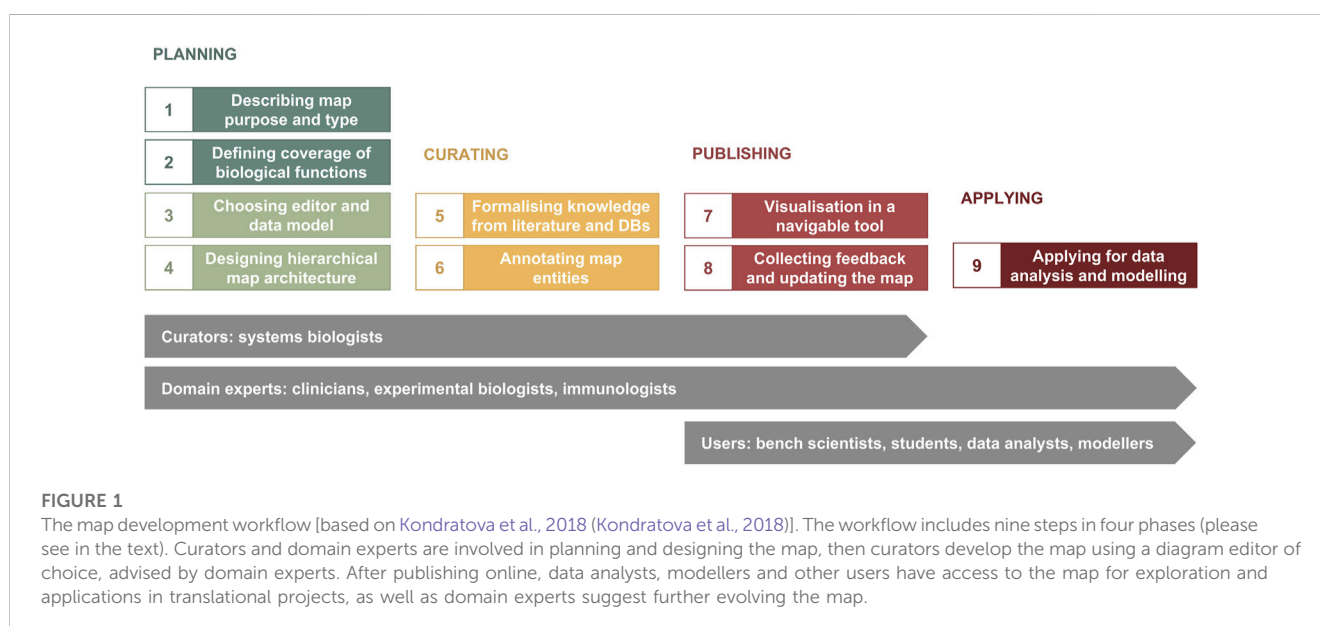
TABLE 1 Examples of disease map purpose, intended use and expected primary users.

Primary purpose	Details	Examples	Primary users
Integrating knowledge	Visualisation, identifying gaps, educational purposes	Atlas of Cancer Signalling Networks Kuperstein et al. (2015a), Parkinson's Disease Map Fujita et al. (2013), RA-Map Singh et al. (2020)	Research centres, students, modellers
Combining knowledge with external databases and tools	Exploring drug targets, text mining	COVID-19 Disease Map Ostaszewski et al. (2021)	COVID-19 research community, data analysts
Data analysis	Network-based analysis, omics and multi-omics data visualisation, interpretation and hypothesis generation	AsthmaMap Mazein et al. (2021a)	IMI U-BIOPRED (No 115010, https://www.imi.europa.eu) community, IMI eTRIKS (No 115446, https://www.etriks.org) community, data analysts
Computational modelling	Using maps as templates for building computational models, refining hypotheses and making predictions	Atherosclerosis Map Parton et al. (2019)	Research centres, modellers
Computational modelling combined with data analysis	Using maps as templates for building Boolean and hybrid modelling and data analysis	RA-Atlas Miagoux et al. (2021), Aghakhani et al. (2022); Zerrouk et al. (2022)	Research centres, modellers, data analysts

with the focus of a research centre so the developed mam supports related research. It also can be part of a large-scale translational project for integrating knowledge and future applications in interpreting data generated by the project.

Intended map purpose might include one or several of the following objectives:

– integrating available knowledge:
– reviewing and visualising disease mechanisms,
– identifying gaps in knowledge,
– educational purposes;
– combining knowledge with external databases and tools:
– exploring drug targets,
– text mining;
– analysis:
– network-based analysis,
– omics and multi-omics data visualisation, interpretation and hypothesis generation;
– computational modelling:
– using maps as the basis/template for different types of models,
– refining hypothesis,
– making predictions.

Clarifying the motivation, objectives and intended applications is needed for the following steps about the coverage of biological functions. This can be documented in a brief project description document (https://disease-maps.io/template), or in a web page (e.g., https://disease-maps.io/projects), or in a project management platform such as FAIRDOMHub (https://fairdomhub.org).

2.2 Defining coverage of biological functions

Exchange with the disease domain experts allows to define the scope and boundaries of the map. Also, key reviews and figures that can help to create an initial skeleton of the map. The main tasks at this stage include:

– Identify priority pathways;
– Identify cell types, tissues and organs relevant to the pathology mechanisms;
– List the molecules known to be involved and drugs used in a given condition
– Define disease outcomes on the molecular, cellular and tissue levels, including subdivision into disease stages and subtypes, if applicable.

Such lists can be stored in a document shared and discussed with the domain experts, and maintained throughout the map development, e.g., https://disease-maps.io/template.

Often this information is available in relevant databases, like disease-focused pathways of Reactome (Reactome:R-HSA-1643685) including cancers (Reactome:R-HSA-5663202) and infectious diseases (Reactome:R-HSA-5663205, Reactome:R-HSA-9694516). Disease pathways of KEGG (https://www.genome.jp/kegg/disease), include Parkinson's disease (KEGG:hsa05012) and COVID-19 (KEGG:hsa05171). The Open Targets Platform (https://platform.opentargets.org) [18] lists molecules relevant for various diseases, e.g., targets associated with allergic asthma (https://platform.opentargets.org/disease/MONDO_0004784/associations).

Outcomes of this step guide the curation efforts in steps 5 and 6 (Figure 1), and the coverage of the biological functions may change depending on available knowledge. In effect, this step is often revisited, so it is important to maintain versioning of the map planning document (https://disease-maps.io/template).

2.3 Choosing a graphical standard and an editor

Majority of disease maps are created in Systems Biology Graphical Notation (SBGN) (Le Novère et al., 2009) using the CellDesigner editor (https://www.celldesigner.org), using the SBGN Process Description (PD) and Activity Flow (AF) languages (Le Novère et al., 2009) (Figure 2). The Entity Relationship (ER) language (Le Novère, 2015) is not currently employed in developing disease maps.
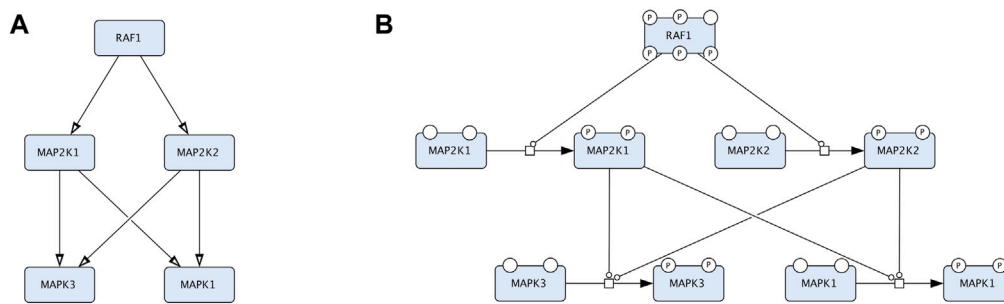
FIGURE 2
An example that compares two representations in CellDesigner that correspond to Activity Flow (Reduced Notation palette in CellDesigner) and Process Description (default palette in CellDesigner). The two diagrams represent the same biological events but in two conceptually different languages. **(A)**. The Process Description representation of the RAF-MEK-ERK signalling: the process of MEK1/2 phosphorylation is catalysed by RAF1 and the process of ERK1/2 phosphorylation is catalysed by the phosphorylated MEK1/2. **(B)**. The Activity Flow representation of the RAF-MEK-ERK signalling: the activity of RAF1 stimulates the activity of MEK1/2 (MAP2K1 and MAP2K2 in official HGNC names), and the activity of MEK1/2 stimulates the activity of ERK1/2 (MAPK3 and MAPK1 in official HGNC names).

Process Description (PD) (Rougny et al., 2019) and Activity Flow (AF) (Mi et al., 2015) are the two most used languages of the Systems Biology Graphical Notation (SBGN) standard (www.sbgn.org) (Le Novère et al., 2009) (Figure 2). More information about the SBGN standard are available in the "SBGN Learning" materials at https://sbgn.github.io/learning and the "Figure to SBGN" page at https://sbgn.github.io/figuretosbgn.

CellDesigner (https://www.celldesigner.org) is a diagram editor that is frequently used for developing disease maps. It supports extensive diagrams and follows SBGN logic and makes it possible to draw SBGN-compatible diagrams in Process Description and Activity Flow type of languages using the default CellDesigner's palette or its Reduced Notation palette. Alternative solutions include the open-source web-based Newt Editor (Balci et al., 2020), the open-source VANTED editor with its SBGN-ED add-on (Czauderna et al., 2010) and the freely available yEd Graph Editor from yWorks (https://www.yworks.com/products/yed). Each tool has its own advantages. Newt offers support for the latest 0.3 SBGN-ML (Bergmann et al., 2020), SBGN bricks (Rougny et al., 2021), advanced automatic layout (Balci and Dogrusoz, 2022), semantic validation of maps, conversion to and from CellDesigner (Balaur et al., 2020) and SBML formats, and an easy switch to alternative colour schemes. SBGN-ED allows map validation (Czauderna et al., 2010), conversion from KEGG and SBML formats and PD-to-AF conversion (Vogt et al., 2013). yEd has an SBGN palette and SBGN layout tool (Siebenhaller et al., 2018), offers intuitive drawing and can support very extensive graphs. The AsthmaMap and the Cystic Fibrosis Map (Pereira et al., 2021) were developed in the yEd Graph Editor, then converted to the SBGN-ML format via the ySBGN tool (Balaur et al., 2022) and then to CellDesigner and MINERVA.

The choice of an editor is also connected to the visualisation platform used (Figure 1, step 7). The MINERVA and NaviCell (Kuperstein et al., 2013; Bonnet et al., 2015) platforms by design are compatible to CellDesigner's XML. The MINERVA Platform also accepts SBGN-ML files and SBML files with layout (Hoksza et al., 2019a), and can therefore be used with Newt, SBGN-ED, as well as with yEd in combination with the ySBGN converter. Importantly, the platform also supports export to these formats, allowing to review, improve and reupload disease map diagrams.

2.4 Designing map architecture

Depending on the goals of the project, map coverage and the resources available, there are several possibilities for designing map architecture and for step-by-step progress including starting from PD or AF layer and working with a single diagram or a set of sub-pathways (Table 2).

Suggested starting design is a PD diagram of mechanisms with accompanying top-level view of disease mechanisms. This design focuses on the prioritised pathways, allows reuse of the PD layer from other projects, and helps communicating with domain experts via the top-level view (Figure 3). Please note that each map version (A, B, C and D) is a functional, complete resource. Potentially the AF layer can be created automatically from the PD layer using CaSQ (Aghamiri et al., 2020) or a PD-to-AF converter (Vogt et al., 2013). More options for building a multi-layered structure and example architectures of the published maps are available at https://disease-maps.io/architecture.

2.4.1 Modularisation

Modular design of a disease map means developing diagrams that are manageable and reusable (Mazein et al., 2021b) for faster curation, reuse in other projects and downstream use in computational modelling (Cooling et al., 2016; Niarakis et al., 2021; Niarakis et al., 2022). A module is a diagram part that is self-contained and can be developed and maintained independently while being included in a larger network. Modularisation can be applied at different scales including signalling subnetworks, e.g., the MAPK cascade, and on a larger scale, e.g., cell-specific pathways in mast cells, eosinophils and macrophages in asthma (AsthmaMap, https://asthma-map.org/mr).

2.4.2 Examples

AlzPathway (http://alzpathway.org) (Mizuno et al., 2012; Ogishima et al., 2016) catalogues signalling pathways in Alzheimer's disease as the first comprehensive map of a particular disease. It is a single PD diagram in CellDesigner with an overview image highlighting the included pathways.

TABLE 2 Examples of map architecture from a single diagram to a hierarchically-organised multi-layered structure.

#	Disease map	Architecture	Format	Top-level view
a	AlzPathway Mizuno et al. (2012), Ogishima et al. (2016), RA-Map Singh et al. (2020)	Single diagram	Process Description	None
b	FluMap Matsuoka et al. (2013)	Single diagram	Process Description	None
c	Cystic Fibrosis Map Pereira et al. (2021)	Single diagram	Activity Flow	None
d	Parkinson's Disease Map Fujita et al. (2013); ACSN Kuperstein et al. (2015a)	Hierarchical	Process Description, Activity Flow	Illustrative
e	ACSN Kuperstein et al. (2015a); RA-Atlas Zerrouk et al. (2022), COVID-19 Disease Map Ostaszewski et al. (2021)	Hierarchical	Process Description	Integrating diagram[a]
f	AsthmaMap Mazein et al. (2021a)	Hierarchical	Process Description, Activity Flow	Integrating diagram[a]

[a]The integrating diagram has elements that represent lower-layer sub-pathways and this way unites them into a single virtual map.
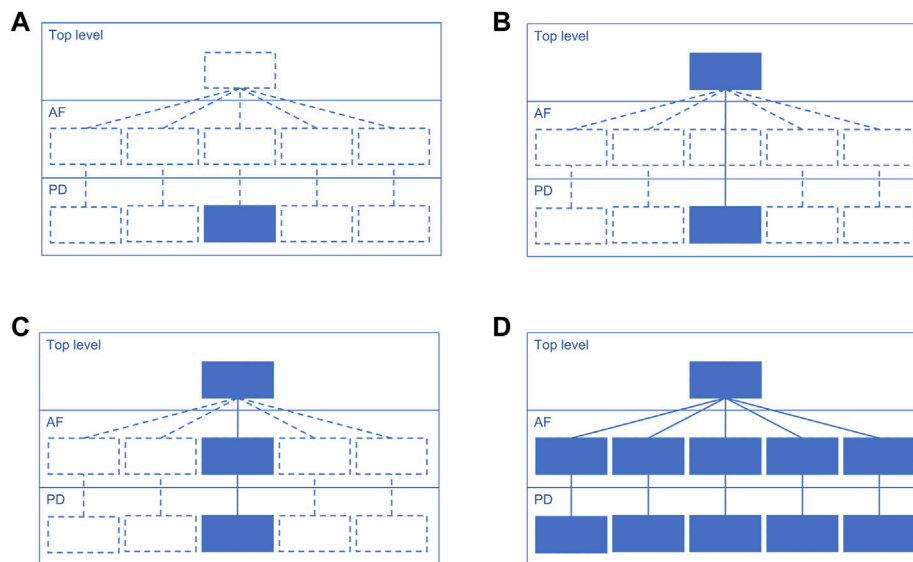

FIGURE 3
A recommended way to gradually build a hierarchical structure starting from a single PD diagram and step by step making the map architecture more complex. Each map version works as a fully functional complete resource. **(A)**. Starting with a priority sub-pathway in PD. **(B)**. Building a top-level view with key mechanisms partly or entirely represented in the PD diagram. **(C)**. Automatically or manually creating an AF diagram that matches the content of the PD diagram. **(D)**. Extending the content to multiple diagrams with one integrating top-level view diagram that represents more detailed AF and PD layers.

The Parkinson's Disease Map (Fujita et al., 2013) was originally designed as a single diagram integrating disease-relevant pathways, e.g., mitochondrial dysfunction, calcium homeostasis, synaptic pathobiology, α-synuclein misfolding, failure of protein degradation systems, apoptosis and neuroinflammation. Later a top-level view was added in MINERVA for easier overview and navigation ("SHOW OVERVIEW" button), and a number of pathways were added as underlying PD diagrams (https://pdmap.uni.lu/minerva).

The FluMap (Matsuoka et al., 2013) is a comprehensive map of the influenza A virus replication cycle. It is available as a single PD diagram with a simplified version in an AF-compatible format developed in the Reduced Notation in CellDesigner.

The Atlas of Cancer Signalling Networks (ACSN) (Kuperstein et al., 2015a) is a collection of large-scale PD diagrams that cover mechanisms

of cancer and include cell cycle and DNA repair, cell survival, regulated cell death, telomere maintenance, epithelial-mesenchymal transition (EMT) and senescence, invasion and motility, and many more (https://acsn.curie.fr/ACSN2/maps.html). These modules are also available as part of a single extensive map (https://acsn.curie.fr).

The AsthmaMap includes three layers: the detailed Biochemical Mechanisms layer in PD (https://asthma-map.org/bm), the intermediate Molecular Relations layer in AF (https://asthma-map.org/mr) and the top-layer Cellular Interactions view in AF-compatible format (https://asthma-map.org/ci) (Mazein et al., 2021a). The three layers are interconnected and presented in a hierarchically-organised map in MINERVA (https://asthma.uni.lu).

The Rheumatoid Arthritis Map (RA-Map) was initially designed as a single Process Description map (Singh et al., 2020) and then

TABLE 3 Comparison of the top-level view vs. Activity Flow vs Process Description layers in a hierarchically-organised map. Resources: KEGG: Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000), OmniPath (Türei et al., 2016), SIGnaling Network Open Resource (SIGNOR) (Licata et al., 2020), Reactome (Gillespie et al., 2022), Recon and ReconMap (Thiele et al., 2013; Noronha et al., 2017), PANTHER (Mi et al., 2019).

Features	Top layer: top-level view	Intermediate layer: AF	Detailed layer: PD
SBGN language	Activity Flow or free format[b]	Activity Flow	Process Description
Connectivity	Interactions	Interactions	Processes
Granularity	Least detailed: focus on key molecules and cell types	Intermediate level of details	Most detailed, the level of processes
Compatible databases	Not available[a]	KEGG signalling pathways, BioCarta, OmniPath, SIGNOR	Reactome, PANTHER, Recon X, KEGG metabolic pathways
Application/Use	Communication with domain experts for defining disease hallmarks	omics visualisation, network analysis, Boolean modelling, integration with machine learning inferred networks	omics visualisation, network analysis, ODE modelling, Boolean modelling (via CaSQ)

[a]The closest similar visualisations are figures with an overview of disease mechanisms in review papers, representations with cell types, tissues and organs shown.
[b]By free-format here we mean flexible representation with one necessary function: all lower-layer sub-pathways in a hierarchically-organised structure are represented by an element on the top-level view and this way integrate them into a single virtual map.

evolved into the Rheumatoid Arthritis Atlas (RA-Atlas), a collection of interconnected diagrams with an access point via a top-level view (Zerrouk et al., 2022). The RA-Atlas includes an updated version of the global RA-Map (Singh et al., 2020) covering relevant metabolic pathways and cell-specific molecular interaction maps for CD4[+] Th1 cells, fibroblasts, and M1 and M2 macrophages.

The COVID-19 Disease Map is a collection of interconnected Process Description diagrams developed by a community-driven effort in the answer to the SARS-CoV-2 pandemic (https://covid.pages.uni.lu). The map gathers diagrams developed by curators using CellDesigner and VANTED software, describing mechanisms of virus cycle in the host cells, its interactions with host molecular pathways, and the mechanisms of host response to the infection. These diagrams are organised into one hierarchical map via a top-level view diagram (https://covid19map.elixir-luxembourg.org/minerva). Notably, the COVID-19 Disease Map Community involves diagram curators of the Reactome and WikiPathways platforms, hosting their SARS-CoV-2-related diagrams.

2.4.3 Choice and implementation of a map architecture

The choice of a map design should match the biological coverage of the map (Figure 1, step 2) and the overall progress of the project. The design should be adapted in case of a change in the coverage, timeline or available resources, as the stages introduced in Figure 3 are flexible. Considering the discussed examples, a multi-layered and modularised hierarchically-organised map should be considered as a default design (Figure 3D). Depending on a project, the focus can be only on one layer, PD or AF, and the top-level view is optional (Figures 3A, B). Though, with adequate resources available, it is recommended to work on a multi-layered hierarchical map as the state-of-the-art practice. The main advantages of this hierarchically-organised structure (Table 3):

– rich content and detailed representation supports different types of visualisation and different types of computational modelling;
– a possibility of reusing available resources;
– compatibility with similar disease map projects, map parts are easy to reuse in new projects;

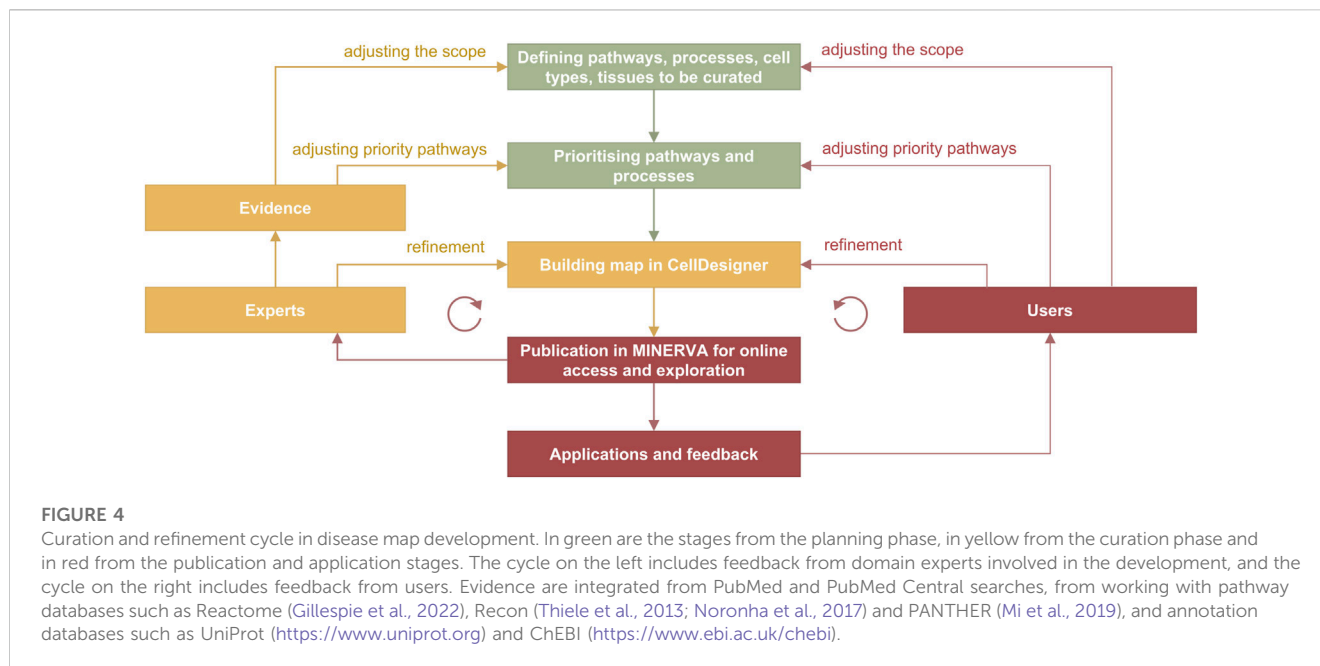– a top-level view of disease hallmarks helps communicating with domain experts.

Finally, other design alternatives are possible, e.g., with more high-level layers or using other standard languages. For instance, the disease map approach supports the concept of adverse outcome pathways (AOPs) in the field of immunotoxicology (paper in preparation).

2.5 Formalising knowledge from literature and databases

In the knowledge formalisation step, relevant causal interactions and processes are taken from selected sources including disease-relevant scientific papers and existing representations in pathway databases. In case of scientific papers, causal interactions and processes are manually identified, integrated and encoded into a diagrammatic representation (Kondratova et al., 2018). In case of pathway databases, relevant causal interactions are manually reviewed and contextualised for a specific disease. Knowledge about molecular processes relevant in the context of a particular disease need to be represented in a standard way, for example, receptor activation, phosphorylation, complex formation, positive and negative regulation of transcription, transport, glycosylation, and more. Knowledge is transformed from textual descriptions and figures in scientific articles into a standardised diagrammatic representation. This representation is also machine-readable and can be further converted into executable computational models.

Knowledge about disease mechanisms is formalised in the SBGN standard within two conceptually different languages, AF and PD. AF is focusing on molecular activities and their sequential connectivity, and PD is focusing on the detailed description of the involved molecular processes (see Table 3; Figure 2). AF is compact and is suitable for overview diagrams and for highlighting key mechanisms. PD is detailed and is better for comprehensive representation of molecular events.

In the curation cycle (Figure 4) information is manually integrated from different sources such as review papers and discussions with domain experts, producing lists of relevant

FIGURE 4
Curation and refinement cycle in disease map development. In green are the stages from the planning phase, in yellow from the curation phase and in red from the publication and application stages. The cycle on the left includes feedback from domain experts involved in the development, and the cycle on the right includes feedback from users. Evidence are integrated from PubMed and PubMed Central searches, from working with pathway databases such as Reactome (Gillespie et al., 2022), Recon (Thiele et al., 2013; Noronha et al., 2017) and PANTHER (Mi et al., 2019), and annotation databases such as UniProt (https://www.uniprot.org) and ChEBI (https://www.ebi.ac.uk/chebi).

molecules, pathways and cell types. Prioritised pathways are curated in CellDesigner from literature or pathway databases, and visualised in MINERVA for use and review, leading to updates in CellDesigner.

The knowledge formalisation and integration works in a similar way to how we build a review paper but, in this case, it is structured and systematic to follow the rules of the graphical standard and to ensure the connectivity in the network. Disease map curation can be seen as a graphical review of relevant disease mechanisms and can be compared to a combination of scoping, rapid and systematic reviews (Grant and Booth, 2009). A scoping review, corresponding to step 2 of the map development process (Figure 1), determines the coverage of biological functions, defines disease hallmarks, and identifies knowledge gaps. Rapid and systematic reviews, corresponding to steps 5 and 6 (Figure 1), narrow down the focus and collect evidence from literature and databases for drawing and annotating diagrams in an editor. These steps are usually performed by a dedicated curator with advice and feedback from a panel of domain experts.

This "graphical review" requires following diagram-specific knowledge representation formats. It demands consistency in using a standardised graphical language and describing connectivity between diagram elements that may span beyond the currently curated article. For example, if a paper states that IL6 activates epithelial cells, it is important to identify the relevant protein (IL6, UniProt:P05231), identify receptors involved (IL6RA, IL6ST, IL6RB), and describe the following signalling pathway, which may require looking for additional articles.

More details on knowledge formalisation for disease maps are available at https://disease-maps.io/formalisation: the webpage provides examples of text-to-SBGN and figure-to-SBGN transformation, examples of search queries in PubMed and PubMed Central, and describes text-mining capabilities of Europe PMC (Ferguson et al., 2021).

A map project can include multiple collaboration and exploration pipelines. An extended example of map development and application ecosystem is available for the COVID-19 Disease Map community in Figure 1 in the 2021 publication (Ostaszewski et al., 2021).

2.6 Annotating map entities and interactions

Two objectives during the annotation step are identifying map components and providing evidence for interactions. Identification assigns stable identifiers or ontology terms to map components to uniquely define them. This ensures interoperability with systems biology tools and external databases. Providing evidence for interactions defines the context of a given interaction, e.g., a specific cell type, tissue, organism, or relevant pathology. Contrary to components, these annotations are not unique but interaction-specific, and require evidence from a relevant source, e.g., scientific publications, curated pathways, or interaction databases.

2.6.1 Annotating entities

Different types of map entities are shown in Table 4 together with recommended annotation.

Proteins, RNAs and genes should be named according to the official symbols of the HUGO Gene Nomenclature Committee (HGNC, https://www.genenames.org). For example, while COX1 is a commonly used name, PTGS1 symbol is recommended. This provides consistency in naming and allows skipping manual addition of other stable identifiers in CellDesigner, because HGNC symbols are recognised by the MINERVA annotation tool on upload.

Annotation of complexes is optional because they can be identified by their components. However, it is recommended to name them by listing the molecules involved, for example, "FCER1A: FCER1G:MS4A2", or using specific names if available. When annotating, complex-specific annotations are recommended,

TABLE 4 Recommended annotation of map entities. Resources: HUGO Gene Nomenclature Committee (HGNC, https://www.genenames.org), UniProt (https://www.uniprot.org), Complex Portal (https://www.ebi.ac.uk/complexportal) (Meldal et al., 2022), ChEBI (https://www.ebi.ac.uk/chebi), PubChem (https://pubchem.ncbi.nlm.nih.gov), ChEMBL (https://www.ebi.ac.uk/chembl), DrugBank (https://go.drugbank.com), Cell Ontology (https://www.ebi.ac.uk/ols/ontologies/cl), BRENDA (https://www.brenda-enzymes.org), Cellosaurus (https://www.cellosaurus.org), GO (http://geneontology.org), MeSH (https://www.ncbi.nlm.nih.gov/mesh).

Biochemical entity	Naming[a]	Identifier
Protein	HGNC official symbol[b]	UniProt/HGNC
RNA	HGNC official symbol	HGNC
Gene	HGNC official symbol	HGNC
Metabolite	ChEBI/PubChem recommended name	ChEBI/PubChem
Drug	ChEBI/PubChem recommended name[c]	ChEBI/PubChem/ChEMBL/DrugBank
Complex	Specific name from literature or listing complex components: Element A:Element B	Not required. If available: Complex Portal
Compartment	Appropriate term from Cell Ontology, BRENDA, Cellosaurus, or a specific name from literature	Cell Ontology/BRENDA/Cellosaurus
Phenotype (biological processes)	Appropriate Gene Ontology (GO) Biological Process (BP) term if available[d]	GO Biological Process
Phenotype (symptom, disease state)	Appropriate MeSH term if available[d]	MeSH

[a]Flexible if an identifier is provided.
[b]This way the protein will be annotated on upload to MINERVA but UniProt will be assigned only if there are no multiple proteins that correspond to that name.
[c]Non-standard names are allowed as long as the annotation is provided: for example, "ERK1" instead of "MAPK3" (UniProt:P27361) or "PGE2" instead of "prostaglandin E2" (ChEBI:15551).
[d]As in the literature or as suggested by the domain experts.

e.g., GO Cellular Component (CC) or Complex Portal identifiers (Meldal et al., 2022) (https://www.ebi.ac.uk/complexportal). For instance, GO CC term "Fc-epsilon receptor I complex" (GO:0032998) for the complex above.

Phenotypes or other elements representing higher order processes should be annotated with Gene Ontology (GO) Biological Process (BP) terms () (Gene Ontology Consortium et al., 2021) or MeSH terms (https://www.ncbi.nlm.nih.gov/mesh). GO BP is recommended for annotating components describing biological processes and MeSH terms are recommended for diseases-related terms.

Metabolites ("simple chemical" in SBGN, or "simple molecule" in CellDesigner) are identified via the Chemical Entities of Biological Interest database (ChEBI, https://www.ebi.ac.uk/chebi) (Hastings et al., 2016). Alternatively, PubChem (https://pubchem.ncbi.nlm.nih.gov) (Kim et al., 2021), Human Metabolome Database (https://hmdb.ca) (Wishart et al., 2007) or LIPID MAPS (https://www.lipidmaps.org) (Sud et al., 2007) can be used. Automatic MINERVA annotation works for standard ChEBI names, for example, "arachidonic acid" (CHEBI:15843), however in other cases annotations need to be provided manually.

An extended description of annotation for disease maps at offers suggested annotation of diagram information and specific details of annotation fields in CellDesigner.

Other useful sources on the topic of annotation are the recommendations of the logical modelling community (Niarakis et al., 2021), the Reactome data model and pathway development practices (Varusai et al., 2020), the rules suggested for creating reusable pathway models (Hanspers et al., 2021), the documentation of the MI2CAST format (Minimum Information about a Molecular Interaction CAusal STatement) (Touré et al., 2021b) and the protocol for high-quality metabolic reconstruction (Thiele and Palsson, 2010).

2.6.2 Annotating interactions

Scientific articles contain evidence that confirm the existence of molecular interactions in the context of a specific disease. Each interaction should be annotated with stable identifiers of relevant articles, a PubMed ID or a DOI. If pathway interactions are used, their identifiers can be used instead. Although some guidelines suggest a minimal number of references confirming an interaction (Kondratova et al., 2018), this should be adapted to the literature coverage of a given disorder.

Importantly, often the relevance of molecular processes in the context of a disease can only be confirmed with a combination of references. For example, one paper can confirm that a specific process is involved in a disease, and another paper will describe details of this process. Only together they offer strong evidence for the interaction.

Some molecular processes related to the disease of interest can be detected in non-human organisms, typically in well-established rodent models for that disease. For parts built based on evidence from cell or animal models, it is important to annotate the corresponding interactions. As proteins, RNA and genes are named using HGNC symbols (human gene symbols), it is recommended to annotate interactions with the NCBI taxon ID of the non-human organism in which the molecular process was determined. For example, if the phosphorylation of STAT3 by JAK2 was determined in mice, the NCBI Taxonomy ID "10090" (NCBI:txid10090) annotation should be added to the interaction.

As most of disease-specific evidence comes from the literature, annotating map interactions requires maintaining a large repository of articles. Such a repository should be stored in a reference manager like Zotero (https://www.zotero.org) or Mendeley (https://www.mendeley.com). Articles should be archived while curating and annotating interactions, preferably in a structure mirroring the organisation of the map. Overall, it is important to check that all interactions in the diagram are annotated with evidence. Automated validation is strongly advised, for instance by using the validators available in the MINERVA Platform.

2.7 Visualising and analysing the maps

Systems biology diagrams of disease maps are developed for online interactive exploration and visual analytics to bridge the gap between the domain experts and bioinformatics experts. The maps can also be used as a part of larger bioinformatic workflows. For example, they can be transformed into a graph database for scalable exploration and analysis, for instance via Neo4j (Section 4), or integrated into a shared network repository like the NDEx platform (Pratt et al., 2015). Their content can be transformed with the use of the CaSQ tool for Boolean modelling (Section 2.4; Table 3), or for analysis of network topology with the help of tools such as Cytoscape (Shannon et al., 2003). However, the reminder of the section focuses on the visual exploration aspect of disease maps, using a dedicated online platform.

The MINERVA Platform (Gawron et al., 2016; Hoksza et al., 2019a; Hoksza et al., 2019b) enables easy online access to and visual interactive exploration of disease maps. Examples are:

– Parkinson's Disease Map (https://pdmap.uni.lu);
– AsthmaMap (https://asthma.uni.lu);
– Rheumatoid Arthritis Map (https://ramap.elixir-luxembourg.org);
– Cystic Fibrosis Map (https://cysticfibrosismap.github.io);
– COVID-19 Disease Map (https://covid19map.elixir-luxembourg.org).

Video tutorials are available for introducing key MINERVA functionalities with the AsthmaMap as an example: navigating the maps, adding comments, exploring drug targets and visualising data (https://asthma-map.org/tutorials).

Maps can be uploaded to MINERVA in several formats including CellDesigner SBML, SBML + layout + render, SBGN-ML and GPML (Hoksza et al., 2019a). The conversion module within the MINERVA Platform allows export to these formats as well. On upload users can choose from CellDesigner's view and SBGN view.

Automatic annotation can be selected on upload. As discussed above, with HGNC names used for proteins, mRNAs and genes, the MINERVA Platform automatically adds links to relevant external databases.

Map browsing and zooming works in MINERVA in the Google Maps manner. For multi-layered maps, interconnected diagrams are organised into a single entity via the submaps function and are visible from the top-level diagram in the SUBMAPS tab (https://asthma-map.org/tutorials/#navigating-the-asthmamap). Search functionality queries all connected maps.

Data can be uploaded as a tab-separated file (https://minerva.uni.lu/doc/examples). This way, omics data (for example, differential expressed genes) can be converted into a gradient of colours on the map (https://asthma-map.org/tutorials/#visualising-data). Multiple datasets can be visualised simultaneously and compared, including a multi-omics combination (e.g., proteomics and metabolomics), or time series data with different time points visualised as separate datasets.

Drug target search connects via online queries to DrugBank (https://www.drugbank.ca) (Wishart et al., 2018) and ChEMBL (https://www.ebi.ac.uk/chembl) (Gaulton et al., 2012; Bento et al., 2014) databases and allows highlighting all entities affected by a drug (https://asthma-map.org/tutorials/#exploring-drug-targets).

Moreover, MINERVA API and plugins (Hoksza et al., 2019b) allow the implementation of customised visual exploration and analytics pipelines.

An alternative web-based tool is NaviCell (Kuperstein et al., 2013; Bonnet et al., 2015) developed primarily to support the ACSN project. It offers functionalities for visualising omics data for performing functional analysis of the maps, and for computing aggregated values for sample groups and their visualisation on the maps (https://navicell.vincent-noel.fr). A video tutorial on ACSN in NaviCell is available at https://acsn.curie.fr/documentation.html#video.

2.8 Collecting feedback and updating the map

When the first complete version of a map is available online, it is advisable to request feedback from domain experts regarding the coverage of disease hallmarks, the depth and quality of pathways included, accuracy and adequacy of the disease mechanisms representation. This input can be collected in the form of 1) free-format notes, 2) suggested publications, or 3) directs comments on the map in MINERVA (https://asthma-map.org/tutorials/#commenting).

It is also suggested to have an automatic check of any inconsistencies in curation and annotation. This can be done using Neo4j-based tools (see the "Disease map management in the Neo4j graph database" section) or the automated verification in the MINERVA Platform (https://minerva.pages.uni.lu/doc/admin_manual/v16.0/index/#configure-automatic-verification).

2.9 Applying the resource

Map applications can play a role in the development process, as analysis and computational modelling may identify missing feedback loops or analytical endpoints, key pathways, or larger gaps in the network. Therefore, considering the use-cases of a map can help with its curation, revision and refinement (Ostaszewski et al., 2021).

One of map applications is omics data visualisation for its interpretation, hypothesis generation, and planning validation experiments. Another is exploration of drug targets in MINERVA using the corresponding functionality (https://asthma-map.org/tutorials). Finally, computational modelling allows to transform a static diagrammatic representation into a dynamic view of disease progression to be used for refining hypotheses and for predictions.

Demonstrated applications of disease maps include:

Data visualisation:
– omics data visualisation (Kuperstein et al., 2015b; Satagopam et al., 2016; Mazein et al., 2021a);
– analysis of cell-specific mechanisms using single-cell expression data (Ostaszewski et al., 2021);
– RNA-Seq-based analysis of the activity of transcription factors (Ostaszewski et al., 2021);
Network-based analysis:
– network analysis and representing map-based molecular signatures for sub-groups of patients in cancer (Kuperstein et al., 2015b; Jdey et al., 2016);
– structural network analysis together with omics data to rationalise the synergistic effects of drugs towards the design of complex disease

stage-specific druggable interventions in preclinical studies (Monraz Gomez et al., 2019);

– network-based analysis and prediction of epithelial-to-mesenchymal-like transition mechanisms followed by experimental validation with the use of an animal model, a transgenic mice (Chanrion et al., 2014);

Computational modelling:

– creating computational models based on disease maps, for example, for rheumatoid arthritis (Miagoux et al., 2021; Aghakhani et al., 2022) and atherosclerosis (Parton et al., 2019);

– creating causal-interaction networks based on a disease map (Touré et al., 2021a);

In preclinical studies with follow-up validation experiments:

– in preclinical studies in cancer (Chanrion et al., 2014; Jdey et al., 2016; Monraz Gomez et al., 2019) with validation of the proposed hypothesis by follow-up experiments;

Fusing disease maps with other solutions:

– integrating disease maps with machine learning inferred networks (Miagoux et al., 2021);

– identifying new crosstalks between pathways from combined analysis of interactions and text mining datasets (Ostaszewski et al., 2021).

3 Long-term sustainability of disease maps

Developing a disease map requires a significant amount of work thus it is important to ensure its long-term sustainability and use. First, the map should be deployed on a stable IT infrastructure for uninterrupted access to the content. This can happen via institutional resources or via a third party, such as the open-access map hosting service offered by ELIXIR Luxembourg (https://elixir-luxembourg.org/services/catalog/minerva). Second, a universal programmatic interface should be set up allowing reproducible queries to the disease map contents. The performance of the already available MINERVA API can be further improved by deployment of graph databases, supporting large and complex queries. Finally, a map should be set up as a FAIR resource (https://www.go-fair.org/fair-principles) (Wilkinson et al., 2016), ensuring its long-term reuse. Following sections discuss the topics of graph database deployment and map FAIRification.

4 Disease map management in the Neo4j graph database

Maps offer visual representations primarily intended for human comprehension. For their exchange and edition, they are stored in machine-readable formats, e.g., CellDesigner SBML and SBGN-ML. While these formats are well suited for storing and exchanging graphical layout information, they do not offer a convenient solution for the network-based management (integration, querying, exploration) and analysis of collections of maps. The Neo4j graph database was proposed for an efficient management of large-scale highly-interconnected heterogeneous biological data (Lysenko et al., 2016), including biological pathway representations in systems biomedicine like Reactome (Fabregat et al., 2018), IntAct (Del Toro et al., 2022) and Recon2Neo4j (Thiele et al., 2013; Balaur et al., 2016). Collections of SBGN maps may also be stored and explored in a Neo4j database using the StonPy software (Rougny et al., 2023), which builds a Neo4j graph

from an SBGN-ML map or a CellDesigner map converted to SBGN-ML using cd2sbgnml (Balaur et al., 2020). Data model of StonPy is close to the visual representation of maps and that is easy to learn. In this model, SBGN nodes (e.g., proteins, states variables) are modelled using Neo4j nodes, while SBGN arcs (e.g., catalyses, production arcs) and relationships between concepts are modelled using Neo4j relationships (edges). Annotations are also stored and can be easily queried. Figure 5 shows an example of a map and an excerpt of its corresponding Neo4j graph as built by StonPy.

The StonPy workflow was used to perform a comparative analysis of integrated biological pathways in several major resources including the Atlas of Cancer Signalling Networks (ACSN), PANTHER, and Reactome (Mazein et al., 2021b; Rougny et al., 2021). The cd2sbgnml (Balaur et al., 2020) and StonPy tools were also used to develop the dedicated C19DM-Neo4j graph database component (https://c19dm-neo4j.covid.pages.uni.lu), which integrates the biological content of the COVID-19 Disease Map diagrams available in MINERVA.

5 FAIR principles for disease maps available in MINERVA

Disease Maps approaches evolve together with other systems biology communities such as COMBINE (Hucka et al., 2015), CoLoMoTo (Niarakis et al., 2021), SysMod (Dräger et al., 2021), SBGN (Le Novère et al., 2009), SBML (Keating et al., 2020) and FAIR (Wilkinson et al., 2016) to enhance the accessibility and reusability of the resources (Niarakis et al., 2022).

The FAIR (Findable, Accessible, Interoperable and Reusable) principles (https://www.go-fair.org/fair-principles) establish criteria to assess and improve the reuse of scientific resources including data, metadata and infrastructure.

The underlying MINERVA Platform provides a set of functionalities that addresses several elements of FAIR:
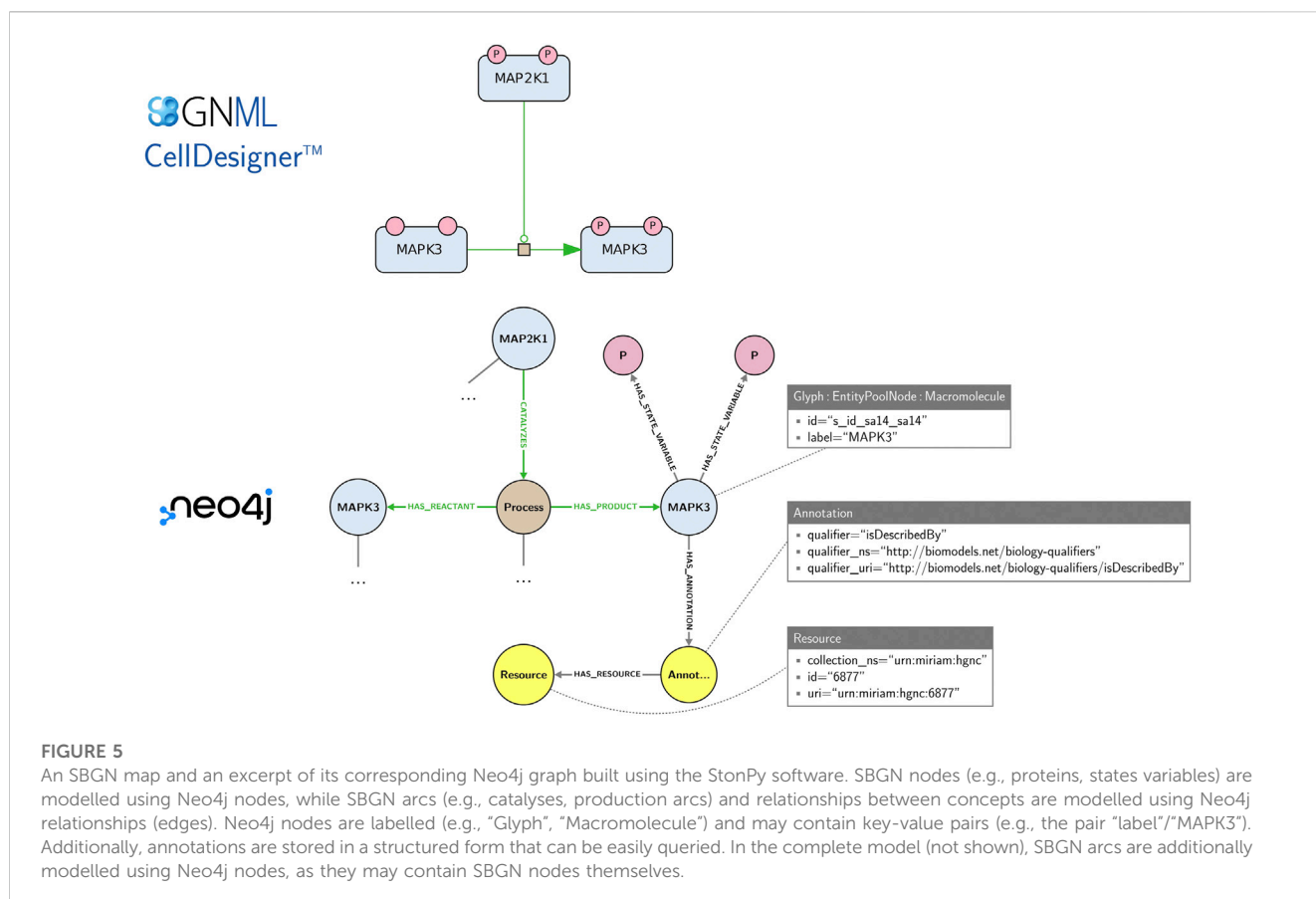
– unified API calls to access the standardised content of all diagrams on the platform;

– functionality for translating between systems biology standards;

– authentication and authorisation procedures for user management;

– explicitly mentioning the reuse licence (for example, the COVID-19 Disease Map diagrams are released in MINERVA under the CC-BY 4.0 licence).

More specifically, MINERVA conforms to FAIR expectations in the following ways:

Findability: The content of the disease maps diagrams in MINERVA is annotated with a range of information elements from initial publications, connected to well-established biomedical resources (e.g., HGNC, UniProt, ChEBI, ChEML, PubMed, etc.) using standard identifiers. The maps are openly available online (e.g., at https://covid19map.elixir-luxembourg.org) and indexed in all major search engines.

Accessibility: The content of biological maps in MINERVA is accessible easily online via standard web protocols. Moreover, MINERVA provides uniform API calls to access the content of the diagrams and can provide required authentication (credentials: user ID and password) and authorisation procedures where necessary.

Interoperability: The biological content in MINERVA follows XML-based systems biology standards such as CellDesigner SBML; translation

FIGURE 5
An SBGN map and an excerpt of its corresponding Neo4j graph built using the StonPy software. SBGN nodes (e.g., proteins, states variables) are modelled using Neo4j nodes, while SBGN arcs (e.g., catalyses, production arcs) and relationships between concepts are modelled using Neo4j relationships (edges). Neo4j nodes are labelled (e.g., "Glyph", "Macromolecule") and may contain key-value pairs (e.g., the pair "label"/"MAPK3"). Additionally, annotations are stored in a structured form that can be easily queried. In the complete model (not shown), SBGN arcs are additionally modelled using Neo4j nodes, as they may contain SBGN nodes themselves.

to other systems biology standard formats is also possible via MINERVA. Moreover, the content of maps is rich in metadata through annotations with information from initial publications connected to biomedical resources as recommended in Table 3, and there are ongoing efforts to connect the content to the Bricks Ontology (Rougny et al., 2021).

Reusability: The biological content of disease map diagrams in MINERVA is released under a specific data usage licence, e.g., under the CC-BY 4.0 usage licence for COVID-19 Disease Map (Ostaszewski et al., 2021). The biological diagrams are primarily edited using CellDesigner, thus complying with the incorporated rules for molecular mechanism representation, and are annotated with various types of information including references to publications (PMIDs) and standard identifiers for molecular entities (UniProt IDs for proteins, ChEBI IDs for metabolites, etc.).

By conforming to most aspects of the FAIR principles, the utility of disease maps available in MINERVA to the systems biomedicine community is greatly enhanced. While there is always room for improvement, and particularly in the domains of versioning and obsoleting policies, the application of FAIR principles throughout the development process of disease maps ensures the long-term relevance of the resource.

6 Conclusion

Knowledge of disease mechanisms is dispersed across many publications and domains. Systematic reconstructions of disease mechanisms with the help from disease domain experts allow building reference resources where fragmented information is integrated into disease maps as conceptual disease models. These maps are applied for interpreting newly generated omics data and for computational modelling in order to study diseases as complex systems. For this reason, disease maps need to be constructed in a responsible way, based on a formal, structured and detailed protocol enriched with illustrative examples. This will ensure high-quality and consistency, as well as enable efficient coordination in large-scale collaborations and help young researchers to be faster introduced to best practices, tools and methods.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

First draft: AM. Complete draft: AM, MO, MA, AR, IB and DW. Conceptualisation: AM, MO and MA. Project coordination: MO. Online materials: AM, MO and MA. Reviewing and editing: all authors. All authors contributed to the article and approved the submitted version.

Conflict of interest

Authors DR and AK were employed by company ITTM Information Technology for Translational Medicine.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aghakhani, S., Soliman, S., and Niarakis, A. (2022). Metabolic reprogramming in rheumatoid arthritis synovial fibroblasts: A Hybrid modeling approach. *PLoS Comput. Biol.* 18 (12), e1010408. doi:10.1371/journal.pcbi.1010408

Aghamiri, S. S., Singh, V., Naldi, A., Helikar, T., Soliman, S., and Niarakis, A. (2020). Automated inference of Boolean models from molecular interaction maps using CaSQ. *Bioinforma. Oxf. Engl.* 36, 4473–4482. doi:10.1093/bioinformatics/btaa484

Balaur, I., Mazein, A., Saqi, M., Lysenko, A., Rawlings, C. J., and Auffray, C. (2016). Recon2Neo4j: Applying graph database technologies for managing comprehensive genome-scale networks. *Bioinforma. Oxf. Engl.* 33, 1096–1098. doi:10.1093/bioinformatics/btw731

Balaur, I., Roy, L., Mazein, A., Karaca, S. G., Dogrusoz, U., Barillot, E., et al. (2020). cd2sbgnml: bidirectional conversion between CellDesigner and SBGN formats. *Bioinforma. Oxf. Engl.* 36, 4975. doi:10.1093/bioinformatics/btaa528

Balaur, I., Roy, L., Touré, V., Mazein, A., and Auffray, C. (2022). GraphML-SBGN bidirectional converter for metabolic networks. *J. Integr. Bioinforma.* 19, 20220030. doi:10.1515/jib-2022-0030

Balci, H., and Dogrusoz, U. (2022). fCoSE: A Fast Compound graph layout algorithm with Constraint support. *IEEE Trans. Vis. Comput. Graph.* 28, 4582–4593. doi:10.1109/TVCG.2021.3095303

Balci, H., Siper, M. C., Saleh, N., Safarli, I., Roy, L., Kilicarslan, M., et al. (2020). Newt: A comprehensive web-based tool for viewing, constructing and analyzing biological maps. *Bioinformatics* 37, 1475–1477. doi:10.1093/bioinformatics/btaa850

Bento, A. P., Gaulton, A., Hersey, A., Bellis, L. J., Chambers, J., Davies, M., et al. (2014). The ChEMBL bioactivity database: An update. *Nucleic Acids Res.* 42, D1083–D1090. doi:10.1093/nar/gkt1031

Bergmann, F. T., Czauderna, T., Dogrusoz, U., Rougny, A., Dräger, A., Touré, V., et al. (2020). Systems biology graphical notation markup language (SBGNML) version 0.3. *J. Integr. Bioinforma.* 17, 20200016. doi:10.1515/jib-2020-0016

Bonnet, E., Viara, E., Kuperstein, I., Calzone, L., Cohen, D. P. A., Barillot, E., et al. (2015). NaviCell Web Service for network-based data visualization. *Nucleic Acids Res.* 43, W560–W565. doi:10.1093/nar/gkv450

Chanrion, M., Kuperstein, I., Barrière, C., El Marjou, F., Cohen, D., Vignjevic, D., et al. (2014). Concomitant Notch activation and p53 deletion trigger epithelial-to-mesenchymal transition and metastasis in mouse gut. *Nat. Commun.* 5, 5005. doi:10.1038/ncomms6005

Cooling, M. T., Nickerson, D. P., Nielsen, P. M. F., and Hunter, P. J. (2016). Modular modelling with Physiome standards. *J. Physiol.* 594, 6817–6831. doi:10.1113/JP272633

Czauderna, T., Klukas, C., and Schreiber, F. (2010). Editing, validating and translating of SBGN maps. *Bioinforma. Oxf. Engl.* 26, 2340–2341. doi:10.1093/bioinformatics/btq407

Del Toro, N., Shrivastava, A., Ragueneau, E., Meldal, B., Combe, C., Barrera, E., et al. (2022). The IntAct database: Efficient access to fine-grained molecular interaction data. *Nucleic Acids Res.* 50, D648–D653. doi:10.1093/nar/gkab1006

Dräger, A., Helikar, T., Barberis, M., Birtwistle, M., Calzone, L., Chaouiya, C., et al. (2021). SysMod: The ISCB community for data-driven computational modelling and multi-scale analysis of biological systems. *Bioinforma. Oxf. Engl.* 37, 3702–3706. doi:10.1093/bioinformatics/btab229

Fabregat, A., Korninger, F., Viteri, G., Sidiropoulos, K., Marin-Garcia, P., Ping, P., et al. (2018). Reactome graph database: Efficient access to complex pathway data. *PLoS Comput. Biol.* 14, e1005968. doi:10.1371/journal.pcbi.1005968

Ferguson, C., Araújo, D., Faulk, L., Gou, Y., Hamelers, A., Huang, Z., et al. (2021). Europe PMC in 2020. *Nucleic Acids Res.* 49, D1507–D1514. doi:10.1093/nar/gkaa994

Fujita, K. A., Ostaszewski, M., Matsuoka, Y., Ghosh, S., Glaab, E., Trefois, C., et al. (2013). Integrating pathways of Parkinson's disease in a molecular interaction map. *Mol. Neurobiol.* 49, 88–102. doi:10.1007/s12035-013-8489-4

Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., et al. (2012). ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 40, D1100–D1107. doi:10.1093/nar/gkr777

Gawron, P., Ostaszewski, M., Satagopam, V., Gebel, S., Mazein, A., Kuzma, M., et al. (2016). MINERVA-a platform for visualization and curation of molecular interaction networks. *NPJ Syst. Biol. Appl.* 2, 16020. doi:10.1038/npjsba.2016.20

Gene Ontology Consortium Douglass, E., Good, B. M., Unni, D. R., Harris, N. L., Mungall, C. J., et al. (2021). The gene ontology resource: Enriching a GOld mine. *Nucleic Acids Res.* 49, D325–D334. doi:10.1093/nar/gkaa1113

Gillespie, M., Jassal, B., Stephan, R., Milacic, M., Rothfels, K., Senff-Ribeiro, A., et al. (2022). The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* 50, D687–D692. doi:10.1093/nar/gkab1028

Grant, M. J., and Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Inf. Libr. J.* 26, 91–108. doi:10.1111/j.1471-1842.2009.00848.x

Hanspers, K., Kutmon, M., Coort, S. L., Digles, D., Dupuis, L. J., Ehrhart, F., et al. (2021). Ten simple rules for creating reusable pathway models for computational analysis and visualization. *PLoS Comput. Biol.* 17, e1009226. doi:10.1371/journal.pcbi.1009226

Hastings, J., Owen, G., Dekker, A., Ennis, M., Kale, N., Muthukrishnan, V., et al. (2016). ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res.* 44, D1214–D1219. doi:10.1093/nar/gkv1031

Hoksza, D., Gawron, P., Ostaszewski, M., Hausenauer, J., and Schneider, R. (2019a). Closing the gap between formats for storing layout information in systems biology. *Brief. Bioinform* 21, 1249–1260. doi:10.1093/bib/bbz067

Hoksza, D., Gawron, P., Ostaszewski, M., Smula, E., and Schneider, R. (2019b). MINERVA API and plugins: Opening molecular network analysis and visualization to the community. *Bioinforma. Oxf. Engl.* 35, 4496–4498. doi:10.1093/bioinformatics/btz286

Hucka, M., Nickerson, D. P., Bader, G. D., Bergmann, F. T., Cooper, J., Demir, E., et al. (2015). Promoting coordinated development of community-based information standards for modeling in biology: The COMBINE initiative. *Front. Bioeng. Biotechnol.* 3, 19. doi:10.3389/fbioe.2015.00019

Jdey, W., Thierry, S., Russo, C., Devun, F., Al Abo, M., Noguiez-Hellin, P., et al. (2016). Drug-driven synthetic lethality: Bypassing tumor cell Genetics with a combination of AsiDNA and PARP Inhibitors. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 23, 1001–1011. doi:10.1158/1078-0432.CCR-16-1193

Kanehisa, M., and Goto, S. (2000). Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi:10.1093/nar/28.1.27

Keating, S. M., Waltemath, D., König, M., Zhang, F., Dräger, A., Chaouiya, C., et al. (2020). SBML level 3: An extensible format for the exchange and reuse of biological models. *Mol. Syst. Biol.* 16, e9110. doi:10.15252/msb.20199110

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., et al. (2021). PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Res.* 49, D1388–D1395. doi:10.1093/nar/gkaa971

Kondratova, M., Sompairac, N., Barillot, E., Zinovyev, A., and Kuperstein, I. (2018). Signalling maps in cancer research: Construction and data analysis. *Database J. Biol. Databases Curation* 2018, bay036. doi:10.1093/database/bay036

Kuperstein, I., Cohen, D. P. A., Pook, S., Viara, E., Calzone, L., Barillot, E., et al. (2013). NaviCell: A web-based environment for navigation, curation and maintenance of large molecular interaction maps. *BMC Syst. Biol.* 7, 100. doi:10.1186/1752-0509-7-100

Kuperstein, I., Bonnet, E., Nguyen, H.-A., Cohen, D., Viara, E., Grieco, L., et al. (2015a). Atlas of cancer signalling network: A systems biology resource for integrative analysis of cancer data with Google maps. *Oncogenesis* 4, e160. doi:10.1038/oncsis.2015.19

Kuperstein, I., Grieco, L., Cohen, D. P. A., Thieffry, D., Zinovyev, A., and Barillot, E. (2015b). The shortest path is not the one you know: Application of biological network resources in precision oncology research. *Mutagenesis* 30, 191–204. doi:10.1093/mutage/geu078

Le Novère, N., Hucka, M., Mi, H., Moodie, S., Schreiber, F., Sorokin, A., et al. (2009). The systems biology graphical notation. *Nat. Biotechnol.* 27, 735–741. doi:10.1038/nbt.1558

Le Novère, N. (2015). Quantitative and logic modelling of molecular and gene networks. *Nat. Rev. Genet.* 16, 146–158. doi:10.1038/nrg3885

Licata, L., Lo Surdo, P., Iannuccelli, M., Palma, A., Micarelli, E., Perfetto, L., et al. (2020). SIGNOR 2.0, the SIGnaling network open resource 2.0: 2019 update. *Nucleic Acids Res.* 48, D504–D510. doi:10.1093/nar/gkz949

Lysenko, A., Roznovăț, I. A., Saqi, M., Mazein, A., Rawlings, C. J., and Auffray, C. (2016). Representing and querying disease networks using graph databases. *BioData Min.* 9, 23. doi:10.1186/s13040-016-0102-8

Matsuoka, Y., Matsumae, H., Katoh, M., Eisfeld, A. J., Neumann, G., Hase, T., et al. (2013). A comprehensive map of the influenza A virus replication cycle. *BMC Syst. Biol.* 7, 97. doi:10.1186/1752-0509-7-97

Mazein, A., Ostaszewski, M., Kuperstein, I., Watterson, S., Le Novère, N., Lefaudeux, D., et al. (2018). Systems medicine disease maps: Community-driven comprehensive representation of disease mechanisms. *Npj Syst. Biol. Appl.* 4, 21. doi:10.1038/s41540-018-0059-y

Mazein, A., Ivanova, O., Balaur, I., Ostaszewski, M., Berzhitskaya, V., Serebriyskaya, T., et al. (2021a). AsthmaMap: An interactive knowledge repository for mechanisms of asthma. *J. Allergy Clin. Immunol.* 147, 853–856. doi:10.1016/j.jaci.2020.11.032

Mazein, A., Rougny, A., Karr, J. R., Saez-Rodriguez, J., Ostaszewski, M., and Schneider, R. (2021b). Reusability and composability in process description maps: RAS-RAF-MEK-ERK signalling. *Brief. Bioinform.* 22, bbab103. doi:10.1093/bib/bbab103

Meldal, B. H. M., Perfetto, L., Combe, C., Lubiana, T., Ferreira Cavalcante, J. V., Bye-A-Jee, H., et al. (2022). Complex portal 2022: New curation frontiers. *Nucleic Acids Res.* 50, D578–D586. doi:10.1093/nar/gkab991

Mi, H., Schreiber, F., Moodie, S., Czauderna, T., Demir, E., Haw, R., et al. (2015). Systems biology graphical notation: Activity Flow language Level 1 version 1.2. *J. Integr. Bioinforma.* 12, 340–381. doi:10.1515/jib-2015-265

Mi, H., Muruganujan, A., Ebert, D., Huang, X., and Thomas, P. D. (2019). PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* 47, D419–D426. doi:10.1093/nar/gky1038

Miagoux, Q., Singh, V., de Mézquita, D., Chaudru, V., Elati, M., Petit-Teixeira, E., et al. (2021). Inference of an integrative, executable network for rheumatoid arthritis combining data-driven machine learning approaches and a state-of-the-art mechanistic disease map. *J. Pers. Med.* 11, 785. doi:10.3390/jpm11080785

Mizuno, S., Iijima, R., Ogishima, S., Kikuchi, M., Matsuoka, Y., Ghosh, S., et al. (2012). AlzPathway: A comprehensive map of signaling pathways of Alzheimer's disease. *BMC Syst. Biol.* 6, 52. doi:10.1186/1752-0509-6-52

Monraz Gomez, L. C., Kondratova, M., Ravel, J.-M., Barillot, E., Zinovyev, A., and Kuperstein, I. (2019). Application of Atlas of cancer signalling network in preclinical studies. *Brief. Bioinform.* 20, 701–716. doi:10.1093/bib/bby031

Niarakis, A., Kuiper, M., Ostaszewski, M., Malik Sheriff, R. S., Casals-Casas, C., Thieffry, D., et al. (2021). Setting the basis of best practices and standards for curation and annotation of logical models in biology-highlights of the [BC]2 2019 CoLoMoTo/SysMod Workshop. *Brief. Bioinform* 22, 1848–1859. doi:10.1093/bib/bbaa046

Niarakis, A., Waltemath, D., Glazier, J., Schreiber, F., Keating, S. M., Nickerson, D., et al. (2022). Addressing barriers in comprehensiveness, accessibility, reusability, interoperability and reproducibility of computational models in systems biology. *Brief. Bioinform.* 23, bbac212. doi:10.1093/bib/bbac212

Noronha, A., Daníelsdóttir, A. D., Gawron, P., Jóhannsson, F., Jónsdóttir, S., Jarlsson, S., et al. (2017). ReconMap: An interactive visualization of human metabolism. *Bioinforma. Oxf. Engl.* 33, 605–607. doi:10.1093/bioinformatics/btw667

Ogishima, S., Mizuno, S., Kikuchi, M., Miyashita, A., Kuwano, R., Tanaka, H., et al. (2016). AlzPathway, an updated map of curated signaling pathways: Towards deciphering Alzheimer's disease pathogenesis. *Methods Mol. Biol. Clifton N. J.* 1303, 423–432. doi:10.1007/978-1-4939-2627-5_25

Ostaszewski, M., Gebel, S., Kuperstein, I., Mazein, A., Zinovyev, A., Dogrusoz, U., et al. (2018). Community-driven roadmap for integrated disease maps. *Brief. Bioinform.* 20, 659–670. doi:10.1093/bib/bby024

Ostaszewski, M., Niarakis, A., Mazein, A., Kuperstein, I., Phair, R., Orta-Resendiz, A., et al. (2021). COVID19 Disease Map, a computational knowledge repository of virus-host interaction mechanisms. *Mol. Syst. Biol.* 17, e10387. doi:10.15252/msb.202110387

Parton, A., McGilligan, V., Chemaly, M., O'Kane, M., and Watterson, S. (2019). New models of atherosclerosis and multi-drug therapeutic interventions. *Bioinforma. Oxf. Engl.* 35, 2449–2457. doi:10.1093/bioinformatics/bty980

Pereira, C., Mazein, A., Farinha, C. M., Gray, M. A., Kunzelmann, K., Ostaszewski, M., et al. (2021). CyFi-MAP: An interactive pathway-based resource for cystic fibrosis. *Sci. Rep.* 11, 22223. doi:10.1038/s41598-021-01618-3

Pratt, D., Chen, J., Welker, D., Rivas, R., Pillich, R., Rynkov, V., et al. (2015). NDEx, the network data exchange. *Cell Syst.* 1, 302–305. doi:10.1016/j.cels.2015.10.001

Ravel, J.-M., Monraz Gomez, L. C., Sompairac, N., Calzone, L., Zhivotovsky, B., Kroemer, G., et al. (2020). Comprehensive map of the regulated cell death signaling network: A powerful analytical tool for studying diseases. *Cancers* 12, 990. doi:10.3390/cancers12040990

Rougny, A., Touré, V., Moodie, S., Balaur, I., Czauderna, T., Borlinghaus, H., et al. (2019). Systems biology graphical notation: Process Description language Level 1 version 2.0. *J. Integr. Bioinforma.* 16, 20190022. doi:10.1515/jib-2019-0022

Rougny, A., Touré, V., Albanese, J., Waltemath, D., Shirshov, D., Sorokin, A., et al. (2021). SBGN Bricks Ontology as a tool to describe recurring concepts in molecular networks. *Brief. Bioinform* 22, bbab049. doi:10.1093/bib/bbab049

Rougny, A., Balaur, I., Luna, A., and Mazein, A. (2023). StonPy: A tool to parse and query collections of SBGN maps in a graph database. *Bioinforma. Oxf. Engl.* 39, btad100. doi:10.1093/bioinformatics/btad100

Satagopam, V., Gu, W., Eifes, S., Gawron, P., Ostaszewski, M., Gebel, S., et al. (2016). Integration and visualization of translational medicine data for better understanding of human diseases. *Big Data* 4, 97–108. doi:10.1089/big.2015.0057

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi:10.1101/gr.1239303

Siebenhaller, M., Nielsen, S. S., McGee, F., Balaur, I., Auffray, C., and Mazein, A. (2018). Human-like layout algorithms for signalling hypergraphs: Outlining requirements. *Brief. Bioinform.* 21, 62–72. doi:10.1093/bib/bby099

Singh, V., Kalliolias, G. D., Ostaszewski, M., Veyssiere, M., Pilalis, E., Gawron, P., et al. (2020). RA-Map: Building a state-of-the-art interactive knowledge base for rheumatoid arthritis. *Database J. Biol. Databases Curation* 2020, baaa017. doi:10.1093/database/baaa017

Sud, M., Fahy, E., Cotter, D., Brown, A., Dennis, E. A., Glass, C. K., et al. (2007). Lmsd: LIPID MAPS structure database. *Nucleic Acids Res.* 35, D527–D532. doi:10.1093/nar/gkl838

Tang, Y. A., Pichler, K., Füllgrabe, A., Lomax, J., Malone, J., Munoz-Torres, M. C., et al. (2019). Ten quick tips for biocuration. *PLOS Comput. Biol.* 15, e1006906. doi:10.1371/journal.pcbi.1006906

Thiele, I., and Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* 5, 93–121. doi:10.1038/nprot.2009.203

Thiele, I., Swainston, N., Fleming, R. M. T., Hoppe, A., Sahoo, S., Aurich, M. K., et al. (2013). A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.* 31, 419–425. doi:10.1038/nbt.2488

Touré, V., Flobak, Å., Niarakis, A., Vercruysse, S., and Kuiper, M. (2021a). The status of causality in biological databases: Data resources and data retrieval possibilities to support logical modeling. *Brief. Bioinform.* 22, bbaa390. doi:10.1093/bib/bbaa390

Touré, V., Vercruysse, S., Acencio, M. L., Lovering, R. C., Orchard, S., Bradley, G., et al. (2021b). The minimum information about a molecular interaction CAusal STatement (MI2CAST). *Bioinforma. Oxf. Engl.* 36, 5712–5718. doi:10.1093/bioinformatics/btaa622

Türei, D., Korcsmáros, T., and Saez-Rodriguez, J. (2016). OmniPath: Guidelines and gateway for literature-curated signaling pathway resources. *Nat. Methods* 13, 966–967. doi:10.1038/nmeth.4077

Varusai, T. M., Jupe, S., Sevilla, C., Matthews, L., Gillespie, M., Stein, L., et al. (2020). Using Reactome to build an autophagy mechanism knowledgebase. *Autophagy* 0, 1543–1554. doi:10.1080/15548627.2020.1761659

Vogt, T., Czauderna, T., and Schreiber, F. (2013). Translation of SBGN maps: Process description to activity Flow. *BMC Syst. Biol.* 7, 115. doi:10.1186/1752-0509-7-115

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3, 160018. doi:10.1038/sdata.2016.18

Wishart, D. S., Tzur, D., Knox, C., Eisner, R., Guo, A. C., Young, N., et al. (2007). Hmdb: The human Metabolome database. *Nucleic Acids Res.* 35, D521–D526. doi:10.1093/nar/gkl923

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018). DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* 46, D1074–D1082. doi:10.1093/nar/gkx1037

Zerrouk, N., Aghakhani, S., Singh, V., Augé, F., and Niarakis, A. (2022). A mechanistic cellular Atlas of the rheumatic Joint. *Front. Syst. Biol.* 2. doi:10.3389/fsysb.2022.925791