COVID19 Disease Map, a computational knowledge repository of virus–host interaction mechanisms


1–82 The list of affiliations appears at the end of this article
*Corresponding author. Tel: +352 46 66 44 5604; E-mail: marek.ostaszewski@uni.lu
†FAIRDOMHub: https://fairdomhub.org/projects/190
© 2021 The Authors. Published under the terms of the CC BY 4.0 license
Abstract

We need to effectively combine the knowledge from surging literature with complex datasets to propose mechanistic models of SARS-CoV-2 infection, improving data interpretation and predicting key targets of intervention. Here, we describe a large-scale community effort to build an open, interoperable and computable repository of COVID-19 molecular mechanisms. The COVID-19 Disease Map (C19DMap) is a graphical, interactive representation of disease-relevant molecular mechanisms linking many knowledge sources. Notably, it is a computational resource for graph-based analyses and disease modelling. To this end, we established a framework of tools, platforms and guidelines necessary for a multifaceted community of biocurators, domain experts, bioinformaticians and computational biologists. The diagrams of the C19DMap, curated from the literature, are integrated with relevant interaction and text mining databases. We demonstrate the application of network analysis and modelling approaches by concrete examples to highlight new testable hypotheses. This framework helps to find signatures of SARS-CoV-2 predisposition, treatment response or prioritisation of drug candidates. Such an approach may help deal with new waves of COVID-19 or similar pandemics in the long-term perspective.

Keywords computable knowledge repository; large-scale biocuration; omics data analysis; open access community effort; systems biomedicine

Introduction

The coronavirus disease 2019 (COVID-19) pandemic due to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has already resulted in the infection of over 250 million people worldwide, of whom almost 5 million have died (https://covid19.who.int, accessed on 05.10.2021). This global challenge motivated researchers worldwide to an unprecedented effort towards understanding the pathology to treat and prevent it. To date, over 170 thousand articles have been published in relation to COVID-19 (PubMed query “covid-19[Title/Abstract] or sars-cov-2[Title/Abstract]”, accessed on 01.07.2021). The reported molecular pathobiology that links SARS-CoV-2 infection to the clinical manifestations and course of COVID-19 is complex and spans multiple biological pathways, cell types and organs (Gagliardi et al., 2020). Resources such as Protein Data Bank repository of viral protein structures (preprint: Lubin et al., 2020) or the IMEx coronavirus interactome (Perfetto et al., 2020) offer detailed information about particular viral proteins and their direct binding partners. However, the scope of this information is limited. To gain insight into the large network of molecular mechanisms, knowledge from the vast body of scientific literature and bioinformatic databases needs to be integrated using systems biology standards. A repository of such computable knowledge will support data analysis and predictive modelling.

With this goal in mind, we initiated a collaborative effort involving over 230 biocurators, domain experts, modellers and data analysts from 120 institutions in 30 countries to develop the COVID-19 Disease Map (C19DMap), an open access collection of curated computational diagrams and models of molecular mechanisms implicated in the disease (Ostaszewski et al., 2020). The C19DMap is a constantly evolving resource, refined and updated by ongoing biocuration, sharing and analysis efforts. Currently, it is a collection of 42 diagrams containing 1,836 interactions between 5,499 elements, supported by 617 publications and preprints. The summary of diagrams available in the C19DMap can be found online (https://covid.pages.uni.lu/map_contents) and in Table EV1.

In the article, we explain the effort of our multidisciplinary community to construct the interoperable content of the resource, involving biocurators, domain experts and data analysts. We introduce the scope of the C19DMap and the insight it brings into the crosstalk and regulation of COVID-19-related molecular mechanisms. Next, we outline analytical workflows that can be used on the contents of the map, including the initial outcomes of two case studies. We conclude with a discussion on the utility and perspectives of the C19DMap as a disease-relevant computational repository.

Results

An interoperable repository of comprehensive and computable diagrams

We constructed a comprehensive diagrammatic description of disease mechanisms in a way that is both human- and machine-readable, lowering communication barriers between experimental and computational biologists. To this end, we aligned the biocuration efforts of the Disease Maps Community (Mazein et al., 2018), Reactome (Jassal et al., 2020), and WikiPathways (Slenter et al., 2018) and developed guidelines for building and annotating these diagrams. In addition, we integrated relevant knowledge from public repositories (Licata et al., 2020; Perfetto et al., 2020; Rodchenkov et al., 2020; Türei et al., 2021) and text mining resources to update and refine the contents of the C19DMap based on other knowledge-building efforts. This work resulted in a series of pathway diagrams constructed de novo, describing key events in the COVID-19 infectious cycle and host response.

The C19DMap project involved three main groups of participants: the biocurators, the domain experts, and the analysts and modellers. Biocurators developed a collection of systems biology diagrams focused on the molecular mechanisms of SARS-CoV-2. Domain experts refined the contents of the diagrams using interactive visualisation and annotations. Analysts and modellers developed computational workflows to generate hypotheses and predictions about the mechanisms encoded in the diagrams. Figure 1 illustrates the ecosystem of the C19DMap Community, highlighting the roles of the participants, available format conversions, interoperable tools and downstream uses. The community members and their contributions are listed on FAIRDOMHub (Wolstencroft et al., 2017).

Creating and accessing the diagrams

The biocurators of the C19DMap diagrams followed the guidelines developed by the Community, WikiPathways (Slenter et al., 2018) and Reactome (Jassal et al., 2020) based on systems biology.
standards (Le Novère et al., 2009; Demir et al., 2010; Keating et al., 2020) and persistent identifiers (Wimalaratne et al., 2018). The diagrams are composed of biochemical reactions and interactions (altogether called interactions) between different molecular entities in various cellular compartments. As multiple teams worked on related topics, biocurators reviewed other diagrams, also across platforms (see also Materials and Methods). The diagrams are accessible online and can be explored using an intuitive user interface. Table 1 summarises information about the curated diagrams, and Table EV1 lists the diagrams and provides links to access them.

Enrichment using knowledge from databases and text mining

The knowledge of COVID-19 mechanisms is rapidly evolving, as shown by the growth of the COVID-19 Open Research Dataset (CORD-19), a source of manuscripts and metadata on COVID-19-related research (preprint: Lu Wang et al., 2020). CORD-19 currently contains almost 480,000 articles and preprints, over ten times more than when it was introduced more than a year ago (accessed on 05.10.2021). In such a quickly evolving environment, manual biocuration needs to be supported by automated procedures to identify and prioritise crucial articles, molecules and their interactions to be included in the C19DMap.

Potential knowledge sources for such assisted biocuration are interaction and pathway databases, especially those with dedicated COVID-19 content (Licata et al., 2020; Perfetto et al., 2020). Their structured and annotated information on protein interactions or causal relationships was generated using separate biocuration guidelines and formats. Nevertheless, their comparable identifiers and references to source publications make them plausible building blocks for constructing the C19DMap (see Materials and Methods).

Text mining approaches are another source of information that can direct the biocurators towards the most recent and relevant findings. They automatically extract and annotate biomolecule names and their interactions from abstracts, full-text documents or pathway figures (Bauch et al., 2020; Hanspers et al., 2020). Networks of molecule interactions constructed by text mining can carry substantially more noise than the contents of interaction databases but offer broader literature coverage.

Figure 1. Ecosystem of the COVID-19 Disease Map Community.

The main groups of the C19DMap Community are biocurators, domain experts, and analysts and modellers; communicating to refine, interpret, and apply C19DMap diagrams. These diagrams are created and maintained by biocurators, following pathway database workflows or stand-alone diagram editors, and reviewed by domain experts. The content is shared via pathway databases or a GitLab repository; all can be enriched by integrated resources of text mining and interaction databases. The C19DMap diagrams are available in several layout-aware systems biology formats and integrated with external repositories, allowing a range of computational analyses, including network analysis and Boolean, kinetic or multiscale simulations.
Table 2 summarises open access interaction databases and text mining knowledge bases supporting the biocuration of the C19DMap. Molecular interactions from these sources have a broad coverage at the cost of depth of mechanistic representation. The biocurators used this content to build and update the map by manual exploration or by programmatic comparison. First, the biocurators visually explored the contents of such networks using available search interfaces to identify interactions of interesting molecules and encoded them in the diagrams. This task was supported by a dedicated visualisation tool COVIDminer (https://rupertoverall.net/covidminer). The biocurators also used assistant chatbots that respond to natural language queries and return

### Table 1. COVID-19 Disease Map contents.

<table>
<thead>
<tr>
<th>Source</th>
<th>Individual diagrams</th>
<th>Reactome</th>
<th>WikiPathways</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diagram contents</strong></td>
<td>21 diagrams</td>
<td>2 diagrams</td>
<td>19 diagrams</td>
</tr>
<tr>
<td></td>
<td>1,394 interactions</td>
<td>101 interactions</td>
<td>401 interactions</td>
</tr>
<tr>
<td></td>
<td>4,272 molecular entities</td>
<td>489 molecular entities</td>
<td>738 molecular entities</td>
</tr>
<tr>
<td></td>
<td>397 publications</td>
<td>227 publications</td>
<td>61 publications</td>
</tr>
<tr>
<td><strong>Access</strong></td>
<td>GitLab</td>
<td>SARS-CoV-1 and SARS-CoV-2 infections collection</td>
<td>COVID pathway collection</td>
</tr>
<tr>
<td></td>
<td>gitlab.lcsb.uni.lu/covid/models</td>
<td>reactome.org/PathwayBrowser/#/R-HSA-9679506</td>
<td>covid.wikipathways.org</td>
</tr>
<tr>
<td><strong>Exploration</strong></td>
<td>The MINERVA Platform (Gawron et al, 2016)</td>
<td>Native web interface</td>
<td>Native web interface</td>
</tr>
<tr>
<td><strong>Biocuration guidelines</strong></td>
<td>Community&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Platform-specific&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Platform-specific&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Diagram Editors</strong></td>
<td>CellDesigner (Matsuoka et al, 2014), Neirv&lt;sup&gt;d&lt;/sup&gt;, SBGN-ED (Czauderna et al, 2010), yEd+ySBGN&lt;sup&gt;e&lt;/sup&gt;</td>
<td>Reactome pathway editor&lt;sup&gt;f&lt;/sup&gt;</td>
<td>PathVisio (Kutmon et al, 2015)</td>
</tr>
<tr>
<td><strong>Formats</strong></td>
<td>CellDesigner SBML, SBGNML (Bergmann et al, 2020)</td>
<td>Internal, SBML and SBGNML compliant</td>
<td>GPML (Kutmon et al, 2015)</td>
</tr>
</tbody>
</table>

The table summarises biocuration resources and content of the C19DMap across three main parts of the repository. All diagrams are listed in Table EV1, and available online at: https://covid.pages.uni.lu/map_contents.

1. https://fairdomhub.org/documents/661
2. https://reactome.org/community/training
4. https://newteditor.org
5. https://github.com/sbgn/ySBGN

Table 2 summarises open access interaction databases and text mining knowledge bases supporting the biocuration of the C19DMap. Molecular interactions from these sources have a broad coverage at the cost of depth of mechanistic representation. The biocurators used this content to build and update the map by manual exploration or by programmatic comparison. First, the

### Table 2. Resources supporting biocuration of the COVID-19 Disease Map.

<table>
<thead>
<tr>
<th>Resource</th>
<th>Type</th>
<th>Manually curated</th>
<th>Directed</th>
<th>Layout</th>
<th>COVID-19 specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMEx Consortium (Orchard et al, 2012)</td>
<td>Interaction database</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes*</td>
</tr>
<tr>
<td>SIGNOR 2.0 (Licata et al, 2020)</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes*</td>
</tr>
<tr>
<td>OmniPath (Türei et al, 2016)</td>
<td>Pathway</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Elsevier Pathway Collection&lt;sup&gt;1&lt;/sup&gt;</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>INDRA (Cyori et al, 2017)</td>
<td>Text mining</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>BioKB&lt;sup&gt;3&lt;/sup&gt;</td>
<td></td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>AILANI COVID-19&lt;sup&gt;4&lt;/sup&gt;</td>
<td></td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>OpenNLP+GNormPlus&lt;sup&gt;5&lt;/sup&gt;</td>
<td></td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

They include (i) collections of COVID-19 interactions and pathways, (ii) interaction databases and (iii) text mining corpora.

1. https://www.ebi.ac.uk/intact/resources/datasets#coronavirus
2. https://signor.uniroma2.it/covid
3. https://pathwaystudio.com
4. http://dx.doi.org/10.17632/d5xwn2c8mw.1
5. https://emmaa.indra.bio/dashboard/covid19
6. https://biokb.lcsb.uni.lu
7. https://ailani.ai
meaningful answers extracted from text mining platforms. Second, we developed routines for programmatic queries of the interaction databases, providing automated and reproducible exploration of the selected databases. This was realised using data endpoints: Application programming interfaces (API) for INDRA, AlL AnI or Pathway Studio, and SPARQL for BioKB. This automated exploration retrieved functions, interactions, pathways or drugs associated with submitted queries, e.g. gene lists. This way, otherwise time-consuming tasks such as an assessment of completeness of a given diagram or search for new literature evidence were automated. Section “Exploration of the networked knowledge” describes an application of such automated queries in crosstalk analysis.

Interoperability of the diagrams and annotations
The biocuration of the C19DMap diagrams was distributed across multiple teams, using varying tools and associated systems biology representations. This required a common approach to annotations of diagram elements and their interactions. Additionally, to compare and combine the diagrams in the C19DMap, interoperability of layout-aware formats was needed.

The diagrams were encoded in three layout-aware formats for standardised representation of molecular interactions: SBML, SBGNML and GPML. All three formats, centred around molecular interaction repositories (Pillich et al, 2020; Hoksza et al, 2020) for harmonised visualisation of diagrams. Additionally, these diagrams were transformed into inputs of computational pipelines and data repositories, allowing network analysis, pathway modelling and interoperability with molecular interaction repositories (Pillich et al, 2017) (see Materials and Methods).

Structure and scope of the COVID-19 Disease Map
The C19DMap was built bottom-up, exploiting a rich bioinformatics framework discussed in Section “An interoperable repository of comprehensive and computable diagrams” of the Results, based on knowledge from existing studies of other coronaviruses (Fung & Liu, 2019) and contextualised with data emerging from studies of SARS-CoV-2 (Gordon et al, 2020). The contents of the C19DMap are available online, summarised in a constantly updated overview at https://covid.pages.uni.lu/map_contents (see also Table EV1). Currently, the C19DMap focuses on molecular processes involved in SARS-CoV-2 entry and replication and host-virus interactions (see Fig 2). Emerging scientific evidence of host susceptibility, immune

![Image of the COVID-19 Disease Map](https://covid.pages.uni.lu/map_contents)

**Figure 2.** The structure and content of the COVID-19 Disease Map.
An overview of the areas of focus of the C19DMap biocuration. Go to covid19map.elixir-luxembourg.org for an interactive version. Full list of diagrams and browsing instructions are available online at covid.pages.uni.lu/map_contents.
response, cell and organ specificity will be incorporated into the next versions in accordance with our curation roadmap (https://fairdomhub.org/documents/907).

While the interactions of SARS-CoV-2 with various host cell types are vital determinants of COVID-19 pathology (Hui et al., 2020; Mason, 2020; Ziegler et al., 2020), the current C19DMap represents an infection of a generic host cell. Several pathways included in the map are shared between different cell types; for example, the IFN-1 pathway is active in dendritic and lung epithelial cells and in alveolar macrophages (Hadjadj et al., 2020; Lee & Shin, 2020; Sa Ribero et al., 2020). Continued annotations of emerging expression datasets (Delorey et al., 2021) and other sources of information will allow the construction of cell-specific versions of the C19DMap to provide an integrated view of the effects of SARS-CoV-2 on the human organism. An example workflow to construct such a focused version of the map was proposed in Section “Case study: analysis of cell-specific mechanisms using single-cell expression data”.

SARS-CoV-2 infection and COVID-19 progression are sequential events that start with viral attachment and entry (Fig 3). These events involve various dynamic processes and different timescales that are not captured in static representations of pathways. The correlation of symptoms and potential drugs suggested to date helps downstream data exploration and drug target interpretation in the context of therapeutic interventions.

### Contents of the COVID-19 Disease Map

#### Virus replication cycle

The virus replication cycle includes binding of the spike surface glycoprotein (S) to angiotensin-converting enzyme 2 (ACE2) facilitated by TMPRSS2 (Hoffmann et al., 2020b; Letko et al., 2020) and
other receptors (preprint: Amraei et al., 2020; preprint: Gao et al., 2020). Viral entry occurs either by direct fusion of the virion with the cell membranes or by endocytosis (Hoffmann et al., 2020a; Xia et al., 2020) of the virion membrane and the subsequent injection of the nucleocapsid into the cytoplasm. Within the host cell, the C19DMap depicts how SARS-CoV-2 hijacks the rough endoplasmic reticulum (RER)-linked host translational machinery for its replication (Chen et al., 2010; Angelini et al., 2013; Nakagawa et al., 2016; V’kovski et al., 2019). The RER-attached translation machinery produces structural proteins, which, together with the newly generated viral RNA, are assembled into new virions and released to the extracellular space via smooth-walled vesicles (Nakagawa et al., 2016) or hijacked lysosomes (Ghosh et al., 2020).

These mechanisms are illustrated in the diagrams of the “Virus replication cycle” section in Table EV1: “Attachment and entry”, “Transcription, translation and replication” and “Assembly and release”.

**Viral subversion of host defence**

Endoplasmic reticulum (ER) stress results from the production of large amounts of viral proteins that create an overload of unfolded proteins (Krähling et al., 2009; DeDiego et al., 2011; Fukushima et al., 2012). The mechanisms of the unfolded protein response (UPR) include the mitigation of the misfolded protein load by reduced protein synthesis and increased protein degradation (Sureda et al., 2020) through the ubiquitin–proteasome system (UPS) and autophagy (Choi et al., 2018; Bello-Perez et al., 2020). SARS-CoV-2 may perturb the process of UPS-based protein degradation via the interaction of the viral Orf10 protein with the Cul2 ubiquitin ligase complex and its putative substrates (Gordon et al., 2020; Zhang et al., 2020). The involvement of SARS-CoV-2 in autophagy is less documented (Yang & Shen, 2020).

The increased burden of misfolded proteins due to viral replication and subversion of mitigation mechanisms may trigger programmed cell death (apoptosis). The C19DMap encodes major signalling pathways triggering this final form of cellular defence against viral replication (Diemer et al., 2010). Many viruses block or delay cell death by expressing anti-apoptotic proteins to maximise the production of viral progeny (Kanzawa et al., 2006; Liu et al., 2007) or induce it in selected cell types (Diemer et al., 2010; Chu et al., 2016; preprint: Chen et al., 2020b).

These mechanisms are illustrated in the diagrams of the “Viral subversion of host defence” section in Table EV1: “ER stress and unfolded protein response”, “Autophagy and protein degradation” and “Apoptosis”.

**Host integrative stress response**

Severe acute respiratory syndrome coronavirus 2 infection damages the epithelium and the pulmonary capillary vascular endothelium (Bao et al., 2020), impairing respiration and leading to acute respiratory distress syndrome (ARDS) in severe forms of COVID-19 (Huang et al., 2020). The release of pro-inflammatory cytokines and hyper-inflammation are known complications, causing further widespread damage (Chen et al., 2020a; Lucas et al., 2020). Coagulation disturbances and thrombosis are associated with severe cases, but specific mechanisms have not been described yet (Iba et al., 2020; Klok et al., 2020). Nevertheless, it was shown that SARS-CoV-2 disrupts the coagulation cascade and causes renin–angiotensin system (RAS) imbalance (Magro et al., 2020; Urwyler et al., 2020).

Angiotensin-converting enzyme 2, used by SARS-CoV-2 for host cell entry, is a regulator of RAS and is widely expressed in the affected organs. The diagrams in the repository describe how ACE2-converted angiotensins trigger the counter-regulatory arms of RAS and the downstream signalling via AGTR1, regulating the coagulation cascade (Ghebrawi et al., 2020; McFadyen et al., 2020).

These mechanisms are illustrated in the diagrams of the “Integrative stress response” section in Table EV1: “Renin–angiotensin system” and “Coagulopathy”.

**Host immune response**

The innate immune system detects specific pathogen-associated molecular patterns, through pattern recognition receptors (PRRs) that recognise viral RNA in the endosome during endocytosis or in the cytoplasm during virus replication. The PRRs activate associated transcription factors promoting the production of antiviral proteins such as interferon-alpha, interferon-beta and interferon-lambda (Takeuchi & Akira, 2010; Berthelot & Liote, 2020; Blanco-Melo et al., 2020; Hadjadji et al., 2020; Park & Iwasaki, 2020). SARS-CoV-2 impairs this mechanism (Chu et al., 2020), but the exact components are yet to be elucidated (Liao et al., 2005; Devaraj et al., 2007; Frie man et al., 2007; Li et al., 2016; Bastard et al., 2020). The C19DMap includes both the virus recognition process and the viral evasion mechanisms. It provides the connection between virus entry, its replication cycle, and the effector pathways of pro-inflammatory cytokines, especially of the interferon type 1 cascade (Wong et al., 2018; Mesey et al., 2019; Mantlo et al., 2020; Su & Jiang, 2020; Thoms et al., 2020; Ziegler et al., 2020).

Key metabolic pathways modulate the availability of nutrients and critical metabolites of the immune microenvironment (Rao et al., 2019). They are a target of infectious agents that reprogram host metabolism to create favourable conditions for their reproduction (Kedia-Meha & Finlay, 2019). The C19DMap encodes several immunometabolic pathways and provides detailed information about the way SARS-CoV-2 proteins interact with them. The metabolic pathways include haem catabolism (Batra et al., 2020) and its downstream target, the NLRP3 inflammasome (van den Berg & Te Velde, 2020), tryptophan-kynurenine metabolism governing the response to inflammatory cytokines (Murakami et al., 2013; preprint: Su et al., 2020), and nicotinamide and purine metabolism (Renz et al., 2020). Finally, we represent the pyrimidine synthesis pathway, tightly linked to purine metabolism, affecting viral DNA and RNA syntheses (Hayek et al., 2020; Xia et al., 2020).

These mechanisms are illustrated in the diagrams of the “Innate Immune Response” section in Table EV1: “PAMP signalling”, “Induction of interferons and the cytokine storm” and “Altered host metabolism”.

**Exploration of the networked knowledge**

The diagrams of the C19DMap were curated in a distributed manner across various platforms and tools. In order to coordinate such an effort and get a systematic overview of the contents of the map, we programatically analysed the content of the diagrams, benefiting from their standard encoding and annotation (see Materials and Methods). This allowed us to identify crosstalk and functional overlaps across pathways. Then, we linked the diagrams to interaction and text mining databases to fill the gaps in our understanding of COVID-19 mechanisms and generate new testable hypotheses.
Below, we discuss three specific examples of exploration of this networked knowledge (see Fig 4). Access to the complete content of the crosstalk diagrams can be found in Materials and Methods.

**Existing crosstalk between COVID-19 Disease Map diagrams**
First, the existing pathways crosstalk emerged by matching entities between the diagrams (Figs 4A, EV1 and EV2). For instance, they link different pathways involved in type I IFN (IFN-1) signalling. Responses to RNA viruses and pathogen-associated molecular patterns (PAMPs) share common pathways, involving RIG-I/Mda-5, TBK1/IKKE and TLR signalling, leading to the production of IFN-1s, especially IFN-beta (Häcker & Karin, 2006) and IFN-alpha (Mogensen, 2009). Downstream, IFN-1 activates Tyk2 and Jak1 protein tyrosine kinases, causing STAT1:STAT2:IRF9 (ISGF3) complex formation to promote the transcription of IFN-stimulated genes (ISGs). Importantly, TBK1 also phosphorylates IKBA, an NF-kB inhibitor, for proteasomal degradation in crosstalk with the UPS pathway, allowing free NF-kB and IRF3 to co-activate ISGs (Fang et al., 2017). Another TBK1 activator, STING, links IFN signalling with pyrimidine metabolism.

SARS-CoV-2 M protein affects these IFN responses by inhibiting the RIG-I-MAVS:TRAF3 complex and TBK1, preventing IRF3 phosphorylation, nuclear translocation and activation (Zheng et al., 2020). In severe COVID-19 cases, elevated NF-kB activation
associated with impaired IFN-1 (Hadjadj et al., 2020) may be a host attempt to compensate for the lack of IFN-1 activation (Rubio et al., 2013), leading to NF-κB hyperactivation and release of pro-inflammatory cytokines. Also, SARS-CoV-1 viral papain-like proteases, contained within the nsp3 and nsp16 proteins, inhibit STING and its downstream IFN secretion (Chen et al., 2019). Perturbations in these pathways may impair the IFN response against SARS-CoV-2 and explain persistent blood viral load and an exacerbated inflammatory response in COVID-19 patients (Hadjadj et al., 2020).

**New crosstalk from interaction and text mining datasets**

New relationships emerging from associated interaction and text mining databases (see Section “Exploration of the networked knowledge” of the Results) suggested new pathway crosstalk (see Figs 4B and EV3). One of these was the interplay between ER stress and the immune pathways, as PPP1R15A regulates the expression of TNF and the translational inhibition of both IFN-1 and IL-6 (Smith, 2018). This finding coincided with the proposed interaction of pathways responsible for protein degradation and viral detection, as SQSTM1, an autophagy receptor and NFKB1 regulator, controls the activity of cGAS, a double-stranded DNA detector (Seo et al., 2014). The suggested NFE2L2-HMOX1 interaction is supported by modulation of the NFE2L2-HMOX1 axis was already proposed as a mechanism to counteract viral entry mechanisms (Hassan et al., 2020) and potential interactions with viral entry mechanisms (Hassan et al., 2020). Interestingly, the modulation of the NFE2L2-HMOX1 axis was already proposed as a therapeutic measure for inflammatory diseases (Attucks et al., 2014), making it an appealing extension of the C19DMap.

**Computational analysis and modelling for hypothesis generation**

The standardised representation and programmatic access to the contents of the C19DMap support reproducible analytical and modelling workflows. Here, we discuss the range of possible computational approaches that can be used to explore the map, its transformed contents, and the associated datasets.

### Table 3. Examples of computational workflows using the COVID-19 Disease Map

<table>
<thead>
<tr>
<th>Workflow</th>
<th>COVID-19 Disease Map contents</th>
<th>User input</th>
<th>Tools</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data interpretation</td>
<td>Online diagrams</td>
<td>Transcriptomics, Proteomics, Metabolomics</td>
<td>The MINERVA Platform (Gawron et al., 2016) PathVisio (Kutmon et al., 2015) Reactome (Jassal et al., 2020)</td>
<td>Visualisation of SARS-CoV-2 mechanisms contextualised to data&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Diagnostics in SIF format (via CasQ)</td>
<td>Transcriptomics</td>
<td>DoRoTHEA (Garcia-Alonso et al., 2019)</td>
<td>Contextualised evaluation of transcription factors activity under SARS-CoV-2 infection&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Diagnostics in SIF format (via CasQ)</td>
<td>Interactome data Proteomics</td>
<td>Network clustering (Messina et al., 2020)</td>
<td>Identification of new SARS-CoV-2-relevant interactions</td>
</tr>
<tr>
<td>Mechanistic modelling</td>
<td>Diagrams in SIF format (via CasQ)</td>
<td>TranscriptomicsMetabolomics</td>
<td>HiPATHIA (Salavert et al., 2016) CARNIVAL (Liu et al., 2019)</td>
<td>Endpoint prediction&lt;sup&gt;c&lt;/sup&gt; Drug target effect prediction</td>
</tr>
<tr>
<td>Discrete modelling</td>
<td>Diagrams in SBML qual format (via CasQ)</td>
<td>Perturbation hypothesis (loss/gain of function)</td>
<td>Cell Collective (Helikar et al., 2012) GINSim (Naldi et al., 2018a) BoolNet</td>
<td>Perturbation outcomes&lt;sup&gt;d&lt;/sup&gt; - Real-time simulation - Attractor analysis</td>
</tr>
<tr>
<td>Stochastic &amp; Multiscale modelling</td>
<td>Diagrams in SBML qual format (via CasQ)</td>
<td>Perturbation hypothesis (loss/gain of function)</td>
<td>PhysiBoSS (Letort et al., 2019) MaBoSS (Stoll et al., 2017) + PhysiCell (GhaffariZadeh et al., 2018)</td>
<td>Perturbation outcomes&lt;sup&gt;d&lt;/sup&gt; - Real-time simulation - Stochastic multiscale modelling</td>
</tr>
</tbody>
</table>

Each workflow relies on the input from the C19DMap, either a direct diagram or its transformed contents, available in the GitLab repository. The workflow users may supply omics datasets to interprete them in the context of the map or test their hypotheses about how disease models will behave under specific perturbations.

<sup>a</sup>Example: see Results, Case study—analysis of cell-specific mechanisms using single-cell expression data.
<sup>b</sup>Example: see Results, Case study—RNA-Seq-based analysis of transcription factor activity.
<sup>c</sup>Example: see Results, Case study—RNA-Seq-based analysis of pathway signalling.
<sup>d</sup>https://colomoto.github.io/colomoto-docker/
Data interpretation and network analysis

The projection of omics data onto the C19DMap broadens and deepens our understanding of disease-specific mechanisms, in contrast to classical pathway enrichment analyses, which often produce lists of generic biological mechanisms. Visualisation of omics datasets on the map diagrams creates overlays, allowing interpretation of specific conditions, such as disease severity or cell types (Satagopam et al., 2016).

Datasets projected on the C19DMap can create signatures of molecular regulation determined by the expression levels of the corresponding molecules. Together, multiple omics readouts and multiple measurements can increase the robustness of such signatures (De Meulder et al., 2018). This interpretation can be extended using available SARS-CoV-2-related omics and interaction datasets (Bouhaddou et al., 2020) to infer which transcription factors, their target genes and signalling pathways are affected upon infection (Dugourd & Saez-Rodriguez, 2019). Combining regulatory interactions of the C19DMap with such data collections extends the scope of the analysis and may suggest new mechanisms to include in the map.

Besides the visual exploration of omics datasets, the network structure of the C19DMap allows extended network analysis of viral-human protein–protein interactions (PPIs) (Gordon et al., 2020). It can be expanded by merging virus–host with human PPIs and proteomics data to discover clusters of interactions indicating human biological processes affected by the virus (Messina et al., 2020). These clusters can be interpreted by visualising them on the C19DMap diagrams to reveal additional pathways or interactions to add to the map.

Mechanistic and dynamic computational modelling

Diagrams from the C19DMap can be coupled with omics datasets to estimate their functional profiles and predict the effect of interventions, e.g. effects of drugs on their targets (Salavert et al., 2016). However, such an approach has a substantial computational complexity, limiting the size of the input diagrams. Large-scale mechanistic pathway modelling can address this challenge but requires transformation of diagrams into causal networks, which, combined with transcriptomics, (phospho-)proteomics or metabolomics data (Dugourd et al., 2021), contextualise the networks and hypotheses about intervention outcomes. Both approaches provide a set of coherent causal links connecting upstream drivers such as stimulations or pathogenic mutations to downstream changes in diagram endpoints or transcription factor activities.

Dynamic modelling allows analysis of changes of molecular networks in time to understand their complexity under disease-related perturbations (Naldi et al., 2018b). C19DMap diagrams, translated to SBML qual using CaSQ (see Materials and Methods), can be used in discrete modelling, using modelling software that supports SBML qual file import. Notably, multiscale processes involved in viral infection, from molecular interactions to multicellular behaviour, can be simulated using a dedicated computational architecture. In such a multiscale setup, single-cell models run in parallel to capture the behaviour of heterogeneous cell populations and their intercellular communications at different time scales, e.g. diffusion, cell mechanics, cell cycle, or signal transduction (Osborne et al., 2017; preprint: Wang et al., 2020). Implementing detailed COVID-19 signalling models in the PhysiBoS framework (Letort et al., 2019) may help better understand complex dynamics of interactions between immune system components and the host cell.

Case study: analysis of cell-specific mechanisms using single-cell expression data

To investigate cell-specific mechanisms of COVID-19, we projected single-cell expression data onto the C19DMap. To this end, we calculated differentially expressed genes (DEGs) for two datasets relevant to the disease. The first dataset describes non-infected bronchial secretory cells (Luikassen et al., 2020; Data ref: Luikassen et al., 2020), where we selected DEGs from three different subtypes of secretory cells dubbed (i) secretory1, (ii) secretory2 and (iii) secretory3 (transient) cells. The second dataset describes SARS-CoV-2-infected intestinal organoids (Triana et al., 2021; Data ref: Triana et al., 2021), where we selected DEGs from (iv) infected and (v) bystander immature enterocytes from the intestinal organoids infected with SARS-CoV-2. DEGs (i), (ii) and (iii) can serve as an illustration of pathway activity across normal lung cells, while datasets (iv) and (v) demonstrate a comparison of molecular activity in cells of infected intestinal tissue. These selected datasets are available as overlays in the C19DMap and can be interactively explored, showing cell-type-specific dysregulation of particular diagrams (see Materials and Methods).

Visual exploration of the differential expression profiles in the C19DMap revealed that transient secretory cells specifically express molecules associated with the virus replication cycle (TMPRSS2). This suggests that these cells are more susceptible to viral entry than the other types of bronchial secretory cells. Also, the interferon 1 signalling pathway was up-regulated in both secretory1 and transient secretory cells. However, transient secretory cells showed up-regulation of elements up- and downstream of the pathway (IFNAR1-JAK1, and ISG15 or OAS1); in secretory1 cells, the up-regulated proteins were downstream (transcription factor AP-1). In the intestinal organoid dataset, the comparison of infected and bystander immature enterocytes confirmed the downregulation of the ACE2 receptor reported by the original article (Triana et al., 2021; Data ref: Triana et al., 2021), as visualised in the virus replication cycle diagram. In addition, exploration of other affected pathways may suggest the context of this observation – for instance, the C19DMap demonstrated the differential activity of the pyrimidine deprivation pathway, which could suggest a reduction of transcriptional activity as a host response to the viral infection. Enrichment analysis of diagrams indicated that mitochondrial dysfunction, apoptosis, and inflammasome activation were dysregulated in infected enterocytes. The enrichment analysis of the cell-type-specific overlays was obtained by the GSEA plugin of the C19DMap. These results can be replicated and examined directly by the users via the visual interface of the C19DMap (see https://covid.pages.uni.lu/minerva-guide/ and Materials and Methods).

Case study: RNA-Seq-based analysis of transcription factor activity

As discussed above, the diagrams of the C19DMap can be coupled with omics datasets. Here, we highlight how the map systematically
reveals the transcription factors (TFs) related to SARS-CoV-2 infection. To do so, we conducted differential expression analysis between SARS-CoV-2 infected Calu-3 human lung adenocarcinoma cell line and controls. Results were used to estimate TF activity deregulation upon viral infection. We mapped the outcomes of the TF activities to pathway diagrams of the C19DMap (see Materials and Methods).

The results for the interferon type I signalling diagram are shown in Fig 5. This pathway included some of the most active TFs after SARS-CoV-2 infection, such as STAT1, STAT2, IRF9 and NFKB1. These are well-known components of cytokine signalling and antiviral responses (Cheon et al, 2013; Fink & Grandvaux, 2013). Interestingly, these TFs were located downstream of various viral proteins (E, S, Nsp1, Orf7a and Orf3a) and members of the MAPK pathway (MAPK8, MAPK14 and MAP3K7). SARS-CoV-2 infection is known to promote MAPK activation, which mediates the cellular response to pathogenic infection and promotes the production of pro-inflammatory cytokines (Bouhaddou et al, 2020). Overall, these results highlighted that the molecular mechanisms of the response of the human cells to SARS-CoV-2 infection can be investigated by combining omics datasets with the diagrams of the C19DMap.

**Case study: RNA-Seq-based analysis of pathway signalling**

The diagrams of the C19DMap allow for a complex analysis of how the infection may affect signalling sequences in encoded pathways based on available omics data. To demonstrate this approach, we applied a mechanistic modelling algorithm that estimates the functional profiles of signalling circuits in the context of omics datasets. We used expression profiles from nasopharyngeal swabs of COVID-19 patients and controls (Lieberman et al, 2020; Data ref: Lieberman et al, 2020) to calculate the differential expression profiles and derive the pathway signalling activities (see Materials and Methods).

To illustrate this approach, we focused on the results of the analysis of the apoptosis pathway, also shown in Fig 6 and Table EV2. We observed an overall downregulation of both the CASP3 and CASP7 subpathways and an inhibition of the circuit ending in effector protein CASP3, possibly due to the downregulation of AKT1 and BAD and the downstream inhibition of BAX. Although the BAX downstream genes were up-regulated, the signal arriving at them was diminished by the effect of the previous nodes. Although CASP8 was up-regulated, the cumulative effect of the individual node activities resulted in the inhibition of CASP7. Indeed, inflammatory response via CASP8 has been described as a result of SARS-CoV-2

**Figure 5. Interferon type I signalling pathway diagram of the COVID-19 Disease Map integrated with TF activity derived from transcriptomics data after SARS-CoV-2 infection.**

A zoom was applied in the area containing the most active TFs (red nodes) after infection. Node shapes: host genes (rectangles), host molecular complex (octagons), viral proteins (V shape), drugs (diamonds) and phenotypes (triangles).
infection (Li et al., 2020), and the role of caspase-induced apoptosis has been established, together with the ripoptosome/caspase-8 complex, as a pro-inflammatory checkpoint (Chauhan et al., 2018), which may be triggering up-regulation of such processes in other pathways. Overall, our findings recapitulate reported outcomes and provide explanations of the effects of interactions on pathway elements.

Discussion

Our knowledge of COVID-19 molecular mechanisms is growing at a great speed, fuelled by global research efforts to investigate the pathophysiology of SARS-CoV-2 infection. Keeping an overview of all the findings, many of which focus on individual molecules, is a great challenge just one year after the start of the pandemic. The C19DMap aggregates this knowledge into molecular interaction diagrams, making it available for visual exploration by life science and clinical researchers and analysis by computational biologists.

The map complements and interfaces with other COVID-19 resources such as interaction databases (Licata et al., 2020; Perfetto et al., 2020), protein-centric resources (preprint: Lubin et al., 2020) and relevant omics data repositories (Delorey et al., 2021) by providing a context to particular pieces of information and helping with data interpretation. The diagrams of the C19DMap describe molecular mechanisms of COVID-19, grounded in the relevant published SARS-CoV-2 research, completed where necessary by mechanisms discovered in related beta-coronaviruses.

We developed the contents of the C19DMap de novo in an unprecedented, community-driven effort involving independent biocurators, as well as WikiPathway and Reactome biocurators. Over forty diagrams with molecular resolution have been constructed since March 2020, shared across three platforms. In this work, we combined and harmonised expertise in biocuration across multiple teams, formulated clear guidelines and cross-reviewed the outcomes of our work with domain experts. Although the approach of community curation was applied in the past (Slayden et al., 2013; Naithani et al., 2019), we are not aware of any curation effort on a similar scale for a single human disease to date.

In this work, we established a computational framework accompanying the biocuration process, integrating interaction databases and text mining solutions to accelerate diagram building. This allowed us not only to enrich particular diagrams but also to explore crosstalk between them and prioritize key novel regulators of the encoded pathways. Thanks to the interoperability of different systems biology formats, we performed this analysis for diagrams constructed in different biocuration environments, extending current advances in pathway interoperability (Bohler et al., 2016).
Moreover, by developing reproducible analysis pipelines for the contents of the C19DMap, we promoted early harmonisation of formats, support of standards and transparency in all steps. Preliminary results of such efforts are illustrated in the case studies above. Notably, the biocurators and domain experts participated in the analysis helped to evaluate the outcomes and corrected the curated content if necessary. This way, we improve the quality of the analysis and increase the reliability of the models used to generate testable predictions.

The C19DMap is an open access repository of diagrams and reproducible workflows for content conversion and analysis. We followed FAIR principles in making our content and code available to the entire research community (Wilkinson et al., 2016). Importantly, FAIRDOMHub is an essential platform for disseminating all information about the project and linking contributors to their contributions. The C19DMap Community is open and expanding as a result of this approach.

Notably, the biocurators and domain experts participated in the curation process. Such an approach may help to deal with the constantly growing C19DMap content and share knowledge. By aligning our efforts, we strive to provide COVID-19-specific pathway models, synchronise content with similar resources and encourage discussion and feedback at every stage of the curation process. Such an approach may help to deal with new waves of COVID-19 or similar pandemics in the long-term perspective.

The project aims to provide the tools to deepen our understanding of the mechanisms driving the infection and help boost drug development supported by testable suggestions. It offers insights into the dynamic nature of the disease at the molecular level and its propagation at the systemic level. Thus, it provides a platform for a precise formulation of models, accurate data interpretation, the potential for disease mitigation and drug repurposing. In the longer run, the constantly growing C19DMap content will be used to facilitate the finding of robust signatures related to SARS-CoV-2 infection predisposition, disease evolution or response to various treatments, along with the prioritisation of new potential drug targets or drug candidates.

This approach to an emerging worldwide pandemic leveraged the capacity and expertise of an entire swath of the bioinformatics community, bringing them together to improve the way we build and share knowledge. By aligning our efforts, we strive to provide COVID-19-specific pathway models, synchronise content with similar resources and encourage discussion and feedback at every stage of the curation process. Such an approach may help to deal with new waves of COVID-19 or similar pandemics in the long-term perspective.

### Materials and Methods

#### Reagents and Tools table

<table>
<thead>
<tr>
<th>Reagent/Resource</th>
<th>Reference or Source</th>
<th>Identifier or Catalog Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Software</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CellDesigner v4.4.2</td>
<td><a href="http://www.celldesigner.org">http://www.celldesigner.org</a> (Matsuoka et al., 2014)</td>
<td></td>
</tr>
<tr>
<td>Newt v3.0</td>
<td><a href="https://newteditor.org">https://newteditor.org</a></td>
<td></td>
</tr>
<tr>
<td>ySBGN</td>
<td><a href="https://github.com/sbgn/ySBGN">https://github.com/sbgn/ySBGN</a></td>
<td></td>
</tr>
<tr>
<td>The MINERVA Platform v15.1.4</td>
<td><a href="https://minerva-web.lcsb.uni.lu">https://minerva-web.lcsb.uni.lu</a> (Gawron et al., 2016)</td>
<td></td>
</tr>
<tr>
<td>Reactome</td>
<td><a href="https://reactome.org">https://reactome.org</a> (Jassal et al., 2020)</td>
<td></td>
</tr>
<tr>
<td>WikiPathways</td>
<td><a href="https://www.wikipathways.org">https://www.wikipathways.org</a> (Slenter et al., 2018)</td>
<td></td>
</tr>
<tr>
<td>PathVisio v5.3.0</td>
<td><a href="https://pathvisio.github.io">https://pathvisio.github.io</a> (Kutmon et al., 2015)</td>
<td></td>
</tr>
<tr>
<td>AILANI COVID-19</td>
<td><a href="https://ai.lani.ai">https://ai.lani.ai</a></td>
<td></td>
</tr>
<tr>
<td>BioKB</td>
<td><a href="https://biokb.lcsb.uni.lu/topic/DOID:0080599">https://biokb.lcsb.uni.lu/topic/DOID:0080599</a></td>
<td></td>
</tr>
<tr>
<td>COVIDminer</td>
<td><a href="https://rupertoeverall.net/covidminer">https://rupertoeverall.net/covidminer</a></td>
<td></td>
</tr>
<tr>
<td>rWikipedia v1.12</td>
<td>10.18129/89.bioc.rWikiPathways</td>
<td></td>
</tr>
<tr>
<td>OmniPathR</td>
<td><a href="https://github.com/saezlab/OmnipathR">https://github.com/saezlab/OmnipathR</a></td>
<td></td>
</tr>
<tr>
<td>cd2sbgnl</td>
<td><a href="https://github.com/sbgn/cd2sbgnl">https://github.com/sbgn/cd2sbgnl</a> (Balaur et al, 2020)</td>
<td></td>
</tr>
<tr>
<td>rnef2sbgn</td>
<td><a href="https://github.com/goilovatenkop/rnef2sbgn">https://github.com/goilovatenkop/rnef2sbgn</a></td>
<td></td>
</tr>
<tr>
<td>Seurat v4.0</td>
<td><a href="https://satijalab.org/seurat/">https://satijalab.org/seurat/</a> (Hao et al, 2021)</td>
<td></td>
</tr>
<tr>
<td>DESeq2</td>
<td>10.18129/89.bioc.DESeq2 (Love et al, 2014)</td>
<td></td>
</tr>
<tr>
<td>Viper v1.26.0</td>
<td>10.18129/89.bioc.viper (Alvarez et al, 2016)</td>
<td></td>
</tr>
<tr>
<td>DoRothEA v1.4.1</td>
<td>10.18129/89.bioc.dorothea (Garcia-Alonso et al, 2019)</td>
<td></td>
</tr>
<tr>
<td>CaSQ v0.9.11</td>
<td><a href="https://ai.lani.ai">Aghamiri et al</a> (2020)</td>
<td></td>
</tr>
</tbody>
</table>
Reagents and Tools table (continued)

<table>
<thead>
<tr>
<th>Reagent/Resource</th>
<th>Reference or Source</th>
<th>Identifier or Catalog Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>CoV-HiPathia</td>
<td>Rian et al (2021)</td>
<td></td>
</tr>
<tr>
<td>Datasets</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIGNOR 2.0 COVID-19 dataset</td>
<td>Licata et al (2020)</td>
<td></td>
</tr>
<tr>
<td>OmniPath</td>
<td>Türei et al (2021)</td>
<td></td>
</tr>
<tr>
<td>RNA-Seq transcriptomic single-cell profiles</td>
<td><a href="https://eils-lung.cells.ucsc.edu">https://eils-lung.cells.ucsc.edu</a></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-infected intestinal organoids</td>
<td>Triana et al (2021)</td>
<td>GSE156760</td>
</tr>
<tr>
<td>SARS-CoV-2-infected Calu-3 cells</td>
<td>Blanco-Melo et al (2020)</td>
<td>GSE147507</td>
</tr>
<tr>
<td>SARS-CoV-2 nasopharyngeal swabs</td>
<td>Lieberman et al (2020)</td>
<td>GSE152075</td>
</tr>
</tbody>
</table>

Methods and Protocols

**Biocuration platforms**

Individual diagrams were encoded in systems biology layout-aware formats (see below) by biocurators using CellDesigner (Matsuoka et al., 2014), Newt (https://newteditor.org), SBGN-ED (Czauderna et al., 2010) and ySBGN (https://github.com/sbgn/ySBGN). This community-based curation was coordinated by sharing curation topics, e.g. relevant pathways or particular SARS-CoV-2 proteins across the community to cover the available literature and identify synergies. Curation guidelines (https://fairdomhub.org/documents/661) were established to ensure proper representation and annotation of the key features of the diagrams. Curation guidelines for logical models (Niarakis et al., 2020) were followed. Regular technical reviews of the diagrams were performed following a previously established checklist to harmonise their content. The diagrams are stored and versioned in a GitLab repository (https://gitlab.lesb.uni.lu/covid/models). Individual diagrams are visualised in the MINERVA Platform (Gawron et al., 2016). The entry-level view is based on Fig. 2.

Reactome (Jassal et al., 2020) biocuration efforts initially focused on SARS-CoV-1 and its proteins, and their functions are extensively documented in the experimental literature. Reactome curators were assigned a subpathway from the viral life cycle, a host pathway or potential therapeutics. Curators were supported by an editorial manager and a dedicated SARS literature triage process. The resulting set of pathways for SARS-CoV-1 provided the basis for computational inference of the corresponding SARS-CoV-2 pathways based on structural and functional homologies between the two viruses. The computationally inferred SARS-CoV-2 infection pathway events and entities were then reviewed and manually curated using published SARS-CoV-2 experimental data. Reactome diagrams are available via a dedicated pathway collection (https://reactome.org/PathwayBrowser/#/R-HSA-9679506).

The WikiPathways (Slenter et al., 2018) diagrams were constructed using PathVisio (Kutmon et al., 2015), with annotation of pathway elements from the integrated BridgeDb identifier mapping framework (van Iersel et al., 2010). All pathways are stored in GPML format (Kutmon et al., 2015). The WikiPathways diagrams are available via a dedicated pathway portal, grouping pathway models specific to SARS-CoV-2, other coronaviruses and general cellular processes relevant to the virus–host interactions (https://www.wikipathways.org/index.php/Portal:COVID-19).

**Layout-aware systems biology formats**

The diagrams are available in SBML format (Keating et al., 2020), allowing computational modelling of biological processes. SBML stores visual information about encoded elements and reactions using render (Bergmann et al., 2018) and layout (Gauges et al., 2015) packages. An early version of SBML adapted by CellDesigner allows storing layout and rendering information. Systems Biology Graphical Notation (SBGN) format is a graphical standard for visual encodings of molecular entities and their interactions, implemented using SBGNML (Bergmann et al., 2020) for encoding the layout of SBGN maps and their annotations. Finally, GPML (Kutmon et al., 2015) is a structured XML format for computable representation of biological knowledge used by the WikiPathways platform.

Interactions and interacting entities are annotated following a uniform, persistent identification scheme, using either MIRIAM Registry or Identifiers.org (Juty et al., 2012) and the guidelines for annotations of computational models. Viral protein interactions are explicitly annotated with their taxonomy identifiers to highlight findings from strains other than SARS-CoV-2. Stable protein complexes from SARS-CoV-2 and SARS are annotated using the Complex Portal.

**Interaction databases**

The biocuration process was supported by interaction and pathway databases storing structured, annotated and curated information about COVID-19 virus–host interactions. The IMEx Consortium (Meldal et al., 2019) dataset (Perfetto et al., 2020) contains curated Coronaviridae-related interaction data from reviewed manuscripts and preprints, resulting in a dataset of roughly 7,300 interactions extracted from over 250 publications, including data from SARS-CoV-2, SARS, CoV, and other strains of Coronaviridae. The dataset is updated with every release of IMEx data and is open access (https://www.ebi.ac.uk/intact/resources/datasets#coronavirus). The SIGNOR 2.0 (Licata et al., 2020) dataset contains manually annotated and validated signalling interactions related to the host–virus interaction, including cellular pathways modulated during SARS-CoV-2 infection. The dataset was constructed from the literature on causal interactions between SARS-CoV-2, SARS-CoV-1, MERS proteins and the human host and is openly available (https://signor.
uniroma2.it/covid/). The Elsevier Pathway Collection (Daraselia et al., 2004; Nesterova et al., 2020) COVID-19 dataset comprises manually reconstructed and annotated pathway diagrams. Statements about molecular interactions are extracted into a knowledge graph by a dedicated text mining technology adapted for extracting facts about viral proteins and viruses from the literature. These interactions were filtered for experimental evidence, used for pathway reconstruction and made openly available (http://dx.doi.org/10.17632/d55xn2c8mw.1). Information from OmniPath (Türe et al., 2021) on existing interactions gathered from pathway and interaction databases was used in a programmatic way to suggest cell-specific interactions and cell–cell interactions specific to immune reactions.

**Text and figure mining**

Text mining was performed on the CORD-19: COVID-19 Open Research Dataset (preprint: Lu Wang et al., 2020). INDRA (Gyori et al., 2017), AILANI COVID-19 (https://ailani.ai) and BioKB processed CORD-19 dataset (https://biokb.lcb.uni.lu/topic/DOID:0080599), with their results available programmatically via REST API and SPARQL interfaces. An OpenNLP-based (https://opennlp.apache.org/) text mining workflow using GNormPlus (Wei et al., 2015) was applied to the CORD-19 dataset and the collection of MEDLINE abstracts associated with the genes in the SARS-CoV-2 PPI network (Gordon et al., 2020) using the Entrez GeneRIFs, https://www.ncbi.nlm.nih.gov/gene/about-generif. (https://gitlab.lcb.uni.lu/covid/models/-/tree/master/Resources/Text%20mining). Also, we used data from 221 CORD-19 dataset figures using a dedicated Figure Mining Workflow (Hanspers et al., 2020), with results available at https://gladstone-bioinformatics.shinyapps.io/shiny-covid pathways. Results of text mining were accessed by the curators in the form of molecular interactions with references to the articles and to sentences from which these interactions were derived. We systematically aligned the C19DMAP with assembled INDRA Statements, both to enrich and to extend the map (see “Crosstalk analysis” below). The content of AILANI and AILANI COVID-19 was accessible via interfaces that allow users to provide natural language queries, such as “What are COVID-19 risk factors?” or “What are the interactors of ACE2?”, facilitating extracting knowledge from the results of text mining workflows. The results of the INDRA workflow were visualised using the COVIDminer project (https://rupertogether.net/covidminer). Each extracted statement describes a directed interaction between two gene products, small molecules or biological processes. The causal network representing the COVIDminer database is browsable through a web interface. The results of the OpenNLP-based text mining workflow were imported into a BioKc biocuration platform for structured processing and SBML export.

**Crosstalk analysis**

Crosstalk analysis was performed for the list of C19DMAP diagrams (Table EV1). The code is available at: https://gitlab.lcb.uni.lu/covid/models/-/tree/master/Resources/Crosstalks. Individual diagrams were accessed via the API of the MINERVA Platform, WikiPathway diagrams via the rWikipathways package (https://github.com/wikipathways/rWikiPathways) and Reactome diagrams via the Reactome API. Text mining interactions are from the INDRA EMMAA Collection (https://emmaa.indra.bio/dashboard/covid19), dataset timestamp: 2020-12-01-21-05-54. Verified molecular interactions for quality control of the text mining data were obtained from OmniPath using the OmnipathR package (https://github.com/semzlab/OmnipathR). We filtered text mining interactions of the EMMAA dataset for “belief” of 0.8 or higher and retained those matching the direction and interacting molecules to the OmniPath dataset. We call this filtered group of interactions “EMMAA-OP interactions”.

Crosstalk between C19DMAP diagrams was calculated based on the HGNC identifiers of their elements. For simplification, all elements of the same diagram were considered to be interacting with each other. Three types of networks were constructed: existing crosstalk, new crosstalk and new regulators. Diagram groups followed the scheme in the list of C19DMAP diagrams (Table EV1). The networks were visualised using Cytoscape (Shannon et al., 2003). The colour code is common for the networks: light green for nodes representing a diagram or a diagram group, light blue for nodes having one or two neighbours, yellow for nodes having three or four neighbours and red for nodes with five or more neighbours. Diagram nodes have prefixes indicating their provenance. Diagram groups have no prefixes, as they combine diagrams across platforms. Existing crosstalk between diagrams, or groups of diagrams, was calculated by identifying shared HGNC identifiers linking diagrams or groups of diagrams. To calculate new crosstalk between diagrams, we merged the EMMAA-OP interactions with the network of existing crosstalk and kept only those new interactions that link at least two upstream and two downstream diagrams or diagram groups. To calculate new upstream regulators of existing diagrams, we merged the EMMAA-OP interactions with the network of existing crosstalk. We kept interactions with source elements, not within existing diagrams, and target elements in at least one existing diagram or diagram group.

**Diagram interoperability and translation for computational modelling**

Bidirectional translation of curated diagrams between CellDesigner and SBGNML formats is supported by the MINERVA Conversion API https://minerva.pages.uni.lu/doc/api/15.1/converter/ (Hoksza et al., 2020), and cd2sbgnml converter https://github.com/sbgn/cd2sbgnml (Balaur et al., 2020). The MINERVA Conversion API supports bidirectional translation between CellDesigner, SBML and GPML. Unidirectional translation from Reactome format to GPML is supported by the Reactome-to-WikiPathways converter (Bohler et al., 2016). Diagrams in the RNEF format of Elsevier Pathway Studio were translated to SBGNML using a dedicated rnef2sbgn software (https://github.com/golovatenkop/rnef2sbgn).

The C19DMAP diagrams (Table EV1) in CellDesigner format were translated using CaSQ (Aghamiri et al., 2020) into executable Boolean networks. Conversion rules and logical formulae were inferred according to the topology and the annotations of the diagrams. SBML-qual files (Chaouiya et al., 2013) generated with CaSQ (Aghamiri et al., 2020) retained their references, annotations and layout of the original CellDesigner file. They can be used for *in silico* simulations and analysis with Cell Collective (Helikar et al., 2012), GINsim (Naldi et al., 2018a) or MaBoSS (Stoll et al., 2017). CaSQ was adapted to produce SIF files necessary for HiPATHIA (Hidalgo et al., 2017) and CARNIVAL (Liu et al., 2019) pipelines. The C19DMAP GitLab repository (https://gitlab.lcb.uni.lu/covid/models) was configured to translate stable versions of diagrams into SBML qual and SIF files. The diagrams were translated to XGMML using Cytoscape and GINSim.
Calculation and visualisation of single-cell RNA-Seq expression profiles

RNA-Seq transcriptomic single-cell profiles were calculated for (i) non-infected airway cells (Lukassen et al, 2020; Data ref: Lukassen et al, 2020) and (ii) SARS-CoV-2-infected intestinal organoids (Triana et al, 2021; Data ref: Triana et al, 2021). The Seurat package (Hao et al, 2021) was used to calculate cell-specific transcriptional profiles. For dataset (i), differential expression was calculated using every cell type against remaining cell types and applying the FindAllMarkers function of the Seurat package with min pc 0.25 and log fold change threshold 0.25. For dataset (ii), the cells were classified into bystander or infected based on the absence or presence of SARS-CoV-2 mRNA measured by scRNAseq (Triana et al, 2021; Data ref: Triana et al, 2021). Differential expression was calculated by contrasting the mock organoids with the bystander or infected cells after 12 h or 24 h of treatment. Expression profiles for the following cell types and conditions were selected for visualisation and enrichment analysis: for dataset (i), three types of secretory cells; and for dataset (ii), infected and bystander immature enterocytes 24 h post-infection versus mock. The datasets were selected to recapitulate the findings in the original papers and demonstrate the capability of the C19DMap for cell-specific data interpretation. Selected differentially expressed genes (DEGs) were prepared for visualisation in the MINERVA Platform as follows. Differential expression values were normalised to [−1,1] range by dividing by three and setting the outliers to their respective border values. Expression values and their corresponding HGNC symbols were used to create visual overlays in the C19DMap in the MINERVA Platform (https://covid19map.elixir-luxembourg.org/minerva/index.xhtml?id=covid19_map_17Jun21, “Overlays” tab).

On-the-fly exploration and enrichment analyses using the GSEA plugin (Hoksza et al, 2019) are described in a dedicated guide (https://covid.pages.uni.lu/minerva-guide/). Complete expression analysis and transformation scripts are available in RMarkdown files at https://gitlab.lcsb.uni.lu/covid/models/-/tree/master/Reso urces/Omics%20analysis.

RNA-Seq-based analysis of transcription factor activity

RNA-Seq transcriptomic profiles of SARS-CoV-2 infection come from SARS-CoV-2-infected Calu-3 cells measured 24 h after infection, Gene Expression Omnibus reference GSE147507 (Blanco-Melo et al, 2020; Data ref: Blanco-Melo et al, 2020). Differential expression analysis of the transcript abundances between conditions was performed with DESeq2 (Love et al, 2014). The resulting t-values from the differential expression analysis were used to estimate the effect of SARS-CoV-2 at the transcription factor (TF) activity level. This analysis was performed using the software Viper (Alvarez et al, 2016) algorithm coupled with TF–target interactions from DoRothEA (Garcia-Alonso et al, 2019). DoRothEA TF–target interactions have a confidence level based on the reliability of their source, which ranges from A (most reliable) to E (least reliable). Here, interactions with confidence levels A, B and C were selected. Activities of TFs having at least five different targets were computed. The TFs’ normalised enrichment score from the Viper output was mapped on the “Interferon type I signalling pathway diagram” (https://fairdomhub.org/models/713) of the C19DMap using the SIF files generated by CaSQ. The resulting network was visualised using Cytoscape (Shannon et al, 2003). Notebooks to reproduce the results of this case study are available at https://github.com/saezlab/Covid19.

RNA-seq-based analysis of pathway signalling

The CoV-HiPathia (Rian et al, 2021) web tool was used to calculate the level of activity of the subpathways of the apoptosis diagram (https://fairdomhub.org/models/712) from the C19DMap. RNA-Seq transcriptomic profiles come from a public dataset of nasopharyngeal swabs from 430 individuals with SARS-CoV-2 and 54 negative controls, Gene Expression Omnibus reference GSE152075 (Lieberman et al, 2020; Data ref: Lieberman et al, 2020). RNA-Seq gene expression data with the trimmed mean of M-values (TMM) normalisation (Robinson et al, 2010) were rescaled to the range [0;1] for the calculation of the signal and normalised using quantile normalisation (Bolstad et al, 2003). Normalised gene expression values and the experimental design (case/control sample names files) were uploaded to CoV-Hipathia to calculate the level of activation of the signalling in the selected diagram. A case/control contrast with a Wilcoxon test was used to assess differences in signalling activity between the two conditions. To reproduce the results, files with normalised gene expression data and the experimental design can be generated using the code https://gitlab.lcsb.uni.lu/covid/models/-/tree/master/Resources/Hipathia/data_preprocessing. These files can then be used in CoV-HiPathia at http://hiopathia.babelomics.org/covid19/ under “Differential signalling” tab. Diagrams from the C19DMap can be selected in the “Pathway source” section, under “Disease Maps Community curated pathways”.

Data availability

COVID-19 Disease Map diagrams are available via:
- the GitLab repository (https://gitlab.lcsb.uni.lu/covid/models).
- Reactome collection (https://reactome.org/PathwayBrowser/#/R-HSA-9679506).

Workflows, executable models and network models are available via:
- the GitLab repository (https://gitlab.lcsb.uni.lu/covid/models).

Expanded View for this article is available online.

Acknowledgements

We would like to thank Andjela Tatarovic, architect, and Gina Crovetto, a researcher in the field of cancer, for their help with the design of the top-level view diagrams. We would like to acknowledge the Responsible and Reproducible Research (R3) team of the Luxembourg Centre for Systems Biomedicine for supporting the project and providing necessary communication and data sharing resources. The work presented in this paper was carried out using the ELIXIR Luxembourg tools and services. This study was supported by the Luxembourg National Research Fund (FNR) COVID-19 Fast-Track grant programme, grant COVID-19-2020-1/14715687/CovScreen (E. Glaab); European Commission, INFORE grant H2020-ICT-825070 (A. Montagud, M. Ponce de Leon, M. Vazques and A. Valencia); European Commission, PerMedCoE grant H2020-ICT-951773 (A. Montagud, M. Ponce de Leon, M. Vazques and A. Valencia); the Federal Ministry of Education and Research (BMBF, Germany) and the Baden-Württemberg Ministry of Science, the Excellence Strategy of the German Federal and State
Conflict of interest

A. Niaraks collaborates with SANOFI-AVENTIS R&D via a public–private partnership grant (CIFRE contract, n° 2020/0766). D. Maier and A. Bauch are employed at Biomax Informatics AG and will be affected by any effect of this publication on the commercial version of the AlLAnI software. J.A. Bachman and B. Gyori received consulting fees from Two Six Labs, LLC. T. Helikar has served as a shareholder and/or has consulted for Discovery Collective, Inc. R. Balling and R. Schneider are founders and shareholders of MEGENO S.A. and ITTM S.A. J. Saez-Rodriguez receives funding from GSK and Sanofi and consults for Travere Therapeutics. The remaining authors have declared that they have no Conflict of interest.

References


Chen X, Yang X, Zheng Y, Yang Y, Xing Y, Chen Z (2014) SARS coronavirus papain-like protease inhibits the type I interferon signaling pathway through interaction with the STING-TRAF3-TBK1 complex. Protein Cell 5: 369–381


Fink K, Grandvaux N (2013) STAT2 and IRF9: Beyond ISGF3. JAK-STAT 2: e27521


© 2021 The Authors

Molecular Systems Biology 17: e10387 | 2021 of 22


**License:** This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.