Bioinformatics, 35(19), 2019, 3779–3785
doi: 10.1093/bioinformatics/btz132
Advance Access Publication Date: 21 February 2019
Original Paper

OXFORD

Data and text mining

CausalTAB: the PSI-MITAB 2.8 updated format for signalling data representation and dissemination

L. Perfetto[1,2], M. L. Acencio[3], G. Bradley[4], G. Cesareni[1,5], N. Del Toro[2], D. Fazekas[6,7], H. Hermjakob[2,8], T. Korcsmaros[7,9], M. Kuiper[10], A. Lægreid[3], P. Lo Surdo[1], R. C. Lovering[11], S. Orchard[2], P. Porras[2], P. D. Thomas[12], V. Touré[10], J. Zobolas [10] and L. Licata [1,*]

[1]Department of Biology, University of Rome Tor Vergata, Rome, Italy, [2]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridge, UK, [3]Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology (NTNU), Trondheim, Norway, [4]Computational Biology and Statistics, Target Sciences, GSK, UK, [5]IRCCS, Fondazione Santa Lucia, Rome, Italy, [6]Department of Genetics, Eötvös Loránd University, Budapest, Hungary, [7]Earlham Institute, Norwich, UK, [8]State Key Laboratory of Proteomics, Beijing Institute of Life Omics, Beijing Proteome Research Center, National Center for Protein Sciences, Beijing, China, [9]Quadram Institute, Norwich, UK, [10]Department of Biology, Norwegian University of Science and Technology (NTNU), Trondheim, Norway, [11]Department of Preclinical and Fundamental Science, Institute of Cardiovascular Science, University College London, UK and [12]Division of Bioinformatics, Department of Preventive Medicine, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA 90033, USA

*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on November 7, 2018; revised on February 1, 2019; editorial decision on February 17, 2019; accepted on February 19, 2019

Abstract

Motivation: Combining multiple layers of information underlying biological complexity into a structured framework represent a challenge in systems biology. A key task is the formalization of such information in models describing how biological entities interact to mediate the response to external and internal signals. Several databases with signalling information, focus on capturing, organizing and displaying signalling interactions by representing them as binary, causal relationships between biological entities. The curation efforts that build these individual databases demand a concerted effort to ensure interoperability among resources.

Results: Aware of the enormous benefits of standardization efforts in the molecular interaction research field, representatives of the signalling network community agreed to extend the PSI-MI controlled vocabulary to include additional terms representing aspects of causal interactions. Here, we present a common standard for the representation and dissemination of signalling information: the PSI Causal Interaction tabular format (CausalTAB) which is an extension of the existing PSI-MI tab-delimited format, now designated PSI-MITAB 2.8. We define the new term 'causal interaction', and related child terms, which are children of the PSI-MI 'molecular interaction' term. The new vocabulary terms in this extended PSI-MI format will enable systems biologists to model large-scale signalling networks more precisely and with higher coverage than before.

Availability and implementation: PSI-MITAB 2.8 format and the new reference implementation of PSICQUIC are available online (https://psicquic.github.io/ and https://psicquic.github.io/MITAB28Format.html).
Contact: luana.licata@uniroma2.it
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Cells are complex and dynamic systems responding to internal or environmental cues (Lodish *et al.*, 2000). Chemical, physical or mechanical stimuli are sensed by receptor proteins which trigger the propagation, amplification and modulation of the signal through a cascade of enzymatic reactions and physical interactions, culminating in the rewiring of the gene expression profile (Fig. 1A) (Lee and Yaffe, 2016). Collectively, the intricate interaction mesh underlying these processes is referred to as 'signal transduction'.

Given the importance that signal transduction has in determining cell phenotype, under either physiological or pathological conditions, obtaining a thorough understanding of the molecular mechanisms underlying the stimulus–phenotype relationships is one of the major goals of systems biology (Barabási and Oltvai, 2004).

To document and archive our growing understanding of signalling systems, a number of resources have undertaken the effort of retrieving from the scientific literature experimental observations supporting causal relationships between signalling proteins and to convert these into structured knowledge (Fig. 1B). These resources can be grouped according to the adopted data representation model in four main categories: activity flow, enzyme–substrate, indirect molecular interaction and process description (Türei *et al.*, 2016).

Among the resources representing signalling information as activity flows, databases such as non-metabolic KEGG (Kanehisa *et al.*, 2017), Reactome (Sidiropoulos *et al.*, 2017), SIGNOR (Perfetto *et al.*, 2016) and SignaLink (Csabai *et al.*, 2018) focus on the capture, organization and display of signalling relationships and represent them as binary, causal relationships between biological entities. Databases such as IntAct (Orchard *et al.*, 2014) focus on the capture of molecular physical interaction adding causality as an additional annotation of the interaction.

In addition, the Gene Ontology Consortium (GOC) (Ashburner *et al.*, 2000) has long been providing causal statements in the form of annotations to terms from the Gene Ontology (GO) branches *regulation of molecular function* (GO:0065009) and *regulation of biological process* (GO:0050789). More recently, the GOC has implemented an extension of GO annotations called GO-Causal Activity Modeling (GO-CAM; http://geneontology.org/cam, (The Gene Ontology Consortium, 2017). GO-CAM provides a 'grammar' for linking simple GO annotations into larger, semantically structured models (such as biological pathways) using relations from the Relations Ontology (RO) (Smith *et al.*, 2005).

The emerging picture is of a fragmented and sparse collection of annotated signalling relationships, which is structured according to different curation approaches. As no single database is comprehensive, users face the challenge of integrating different datasets to obtain maximum coverage. To increase interoperability, adequate standards, ontologies and controlled vocabularies (CVs) are required. The importance of producing standardized data for the life sciences has been documented many times. A particularly successful example of this is the work of the Molecular Interaction work group of the HUPO-Proteomics Standards Initiative (HUPO-PSI) (Deutsch *et al.*, 2017). Over the last 15 years, this group has developed and disseminated community standards, tools and CVs (Kerrien *et al.*, 2007; Sivade Dumousseau *et al.*, 2018a) and minimum information guidelines for authors (Orchard *et al.*, 2007). The ability to display molecular interactions in a single, unified PSI-MI XML format represented a milestone in the field of molecular interactions

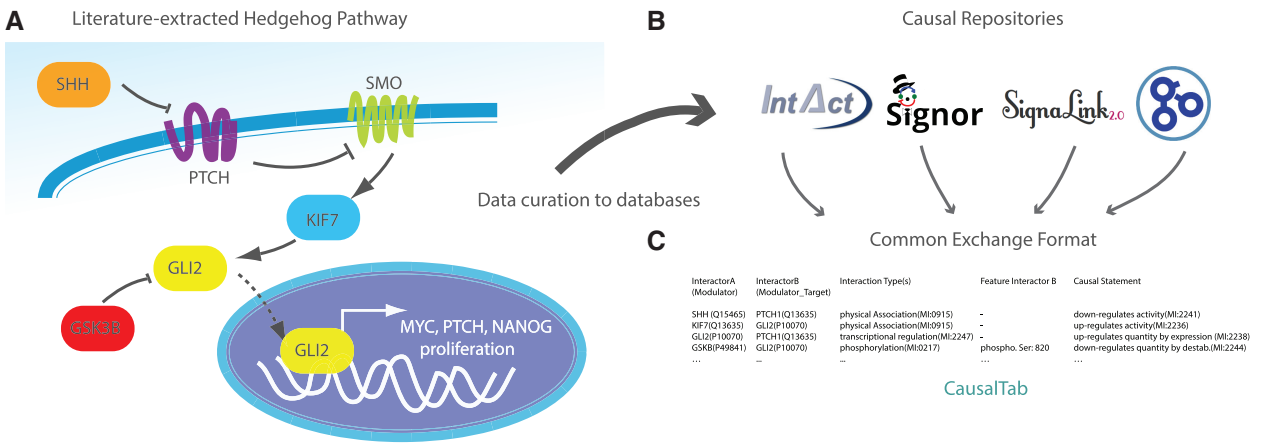

Fig. 1. Representation of causal interaction data in a standardized format. (A) The Hedgehog (HH) pathway as reported in the literature (Kaushal *et al.*, 2017; Yang *et al.*, 2010). In the absence of the HH ligands (e.g. Sonic Hedgehog, SHH), PTCH inhibits SMO signalling, while GLI2, the transcription factor responsible for HH signalling response, is maintained in a repressed state by GSK3B. GSK3B phosphorylates GLI2 and thus promotes its proteasomal degradation. In the presence of SHH ligand, PTCH is unable to inhibit SMO. Activated SMO positively regulates KIF7 which stabilizes GLI2. Activated GLI2 enhances transcription of the HH target genes (e.g. MYC, PTCH, NANOG), resulting in cell proliferation. (B) Data from the literature are independently curated in Causality-related resources. (C) The adoption of CausalTAB (PSI-MITAB 2.8), as a common exchange format allows users for the integration of causality data

(Hermjakob *et al.*, 2004), and the use of common CVs has enabled the consistent annotation of the captured information. A tab-delimited format, MITAB, has proven more suitable for users requiring a simple, human-readable configuration (Kerrien *et al.*, 2007). Implementation of these standards by all leading molecular interaction databases has considerably contributed to data exchange, representation and comparison and encouraged the development of specific tools, such as PSICQUIC (Aranda *et al.*, 2011) and Cytoscape (Smoot *et al.*, 2011) to retrieve and visualize this information.

XML formats handling signalling information are already available (Hucka *et al.*, 2003). They support the representation of molecular interaction data using a reaction-based model that can store very detailed information about the physical and chemical properties of an interaction, such as equilibrium constant, reaction speed and reactant concentrations. Key reaction-based file formats are Systems Biology Markup Language (SBML) (Hucka *et al.*, 2003) and BioPax (Demir *et al.*, 2010), which are more suitable for storing chemical interactions rather than molecular interaction networks.

Curation into XML formats carries a large resource overhead, management requires programming skill and the files are not readily human-readable. Experience from the HUPO PSI-MI working group taught us that MITAB is much more popular than PSI-MI XML, albeit much less expressive. During an EMBL/EC-funded Causal Reasoning Workshop organized at EMBL-EBI in 2016, representatives of some of the main signalling repositories [SIGNOR, IntAct (Causal Reasoning), GO, SignaLink 2.0] recognized the necessity to move towards unification and standardization of signalling data and therefore to adopt common standards and CVs (Fig. 1B). At this meeting, it was agreed to adopt and expand the existing PSI-MI CV by adding new terms required to define different characteristics of causal interactions and to address the lack of a tab-delimited representation of causality by adopting and extending the CausalTAB format originally designed by the SIGNOR database curators.

Here, we report the first version of the new PSI-MI format and accompanying updates of the PSI-MI CV for the representation and dissemination of signalling information: the PSI Causal Interaction format TAB (CausalTAB or PSI-MITAB 2.8). This extension of the existing MITAB format was developed in compliance with the PSI framework and adopted by the MI workgroup of the HUPO-PSI. We have defined a new CV root term, 'causal interaction', and its related child terms, to annotate the different aspects of causal interactions.

2 Results

2.1 Causal interactions in the PSI-MI framework

CausalTAB is primarily inspired by the HUPO-PSI MITAB and has been designed to be a PSI MITAB compliant, Excel-compatible, tab-delimited format developed for users requiring only minimal information in a user-friendly structure. Since its development, the MITAB format has been extended to allow a more granular representation of the interaction data, with PSI-MITAB 2.6 and 2.7 versions now available (del-Toro *et al.*, 2013). The data are structured using PSI-MI CV terms that allows biocurators to capture molecular interaction data consistently and empowers users to perform data searches systematically.

The PSI-MI CV and MITAB were originally developed to capture a standardized representation of physical interactions; interaction directionality and the resulting effects (activation/inhibition for protein–protein interactions and up-/down-regulation for regulatory interactions) were not included. However, as a large number of physical interactions are known to be regulatory, there is a clear need for extending the PSI-MI standard to incorporate additional layers of abstraction. CausalTAB/PSI-MITAB 2.8 has therefore been designed to meet the needs of those members of the molecular interaction community who wish to extend their representation of molecular interaction data with directionality data.

2.2 Revision and extension of PSI-MI CV and MITAB file structure into CausalTAB standard

In a biological context, causal interactions are abstractions representing the regulatory effect that a regulator entity (a stimulus, a transcription factor, an enzyme, etc.) has on a target entity (a receptor, target gene, a substrate, etc.) (see also the example reported in Fig. 1). At their most basic level, causal interactions involve two partners, they have a direction (subject -> object) and a description of the regulatory effect (positive or negative). Chains of causal interactions underlie signal transductions, which modulate the cell response.

To enable the representation of causal relationships in a PSI-MI compatible format, we set out to scrutinize the data structure of resources annotating causal relationships to identify the need for extension of the PSI-MI CV. Given the binary, directed and effective nature of causal interactions, the process has been conceptually organized into four steps: (i) revision of terms to describe the directionality of the interaction; (ii) revision of terms to describe the causality of the interaction; (iii) revision of terms to describe the molecular mechanisms underlying the interaction and (iv) revision of terms to describe the entities.

Each term is assigned to an identifier (as MI:2233) and annotated with a description and a reference. We first defined the new CV term 'causal interaction' as a new branch term in the PSI-MI CV (Fig. 2 and Supplementary Table S1), and subsequently proceeded with the creation of related child terms. Each new term is linked to a definition and a reference and systematically integrated into the PSI-MI CV. The 'causal interaction' term is defined as 'Binary causative relationships between biological entities'. The granularity of the new CV terms allows the precise description of causal interactions using the PSI-MI schema (Fig. 2).

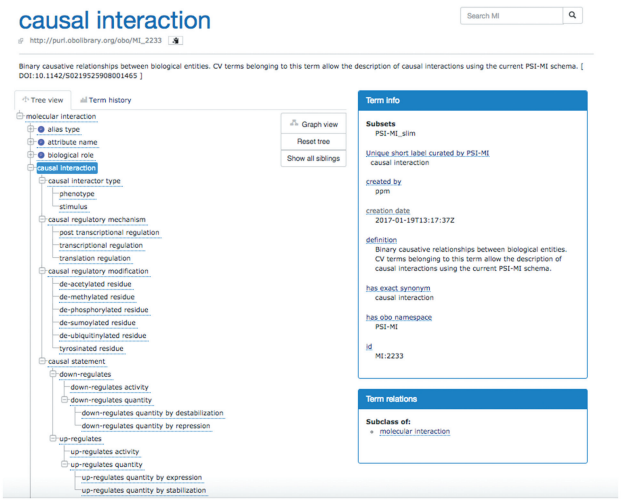

Fig. 2. The hierarchical structure of the new PSI-MI CVs terms. Causal Interaction class and child terms as shown in the Ontology Lookup Service

2.3 New entities added to PSI-MI CV

2.3.1 Entity types

Entities to be represented can encompass different kinds of molecules, such as proteins, DNA, RNA, chemicals, drugs and also non-molecule entities, such as stimuli, phenotypes and biological processes which are fundamental to understanding and representing signalling events. The PSI-MI CV supports the representation of molecular entities (see MI:0313 and children), but not of non-molecules interactors. To fill this gap, we created a 'causal interactor type' term (MI:2259) and related children ('phenotype' and 'stimulus') (Fig. 2 and Supplementary Table S1).

2.3.2 Biological role (directionality)

Causal interactions are directional and therefore, by definition, asymmetric, as the interacting pair of entities have two distinguishable roles: one entity constitutes the modulator, while the other one is the target of the modulation. Some terms in the PSI-MI ontology already describe such asymmetry, for instance terms like 'enzyme' (MI:0501) or 'enzyme target' (MI:0502). To generalize this concept, we defined two new 'biological role' terms: 'regulator' (MI:2274) and 'regulator target' (MI:2275) (Supplementary Table S1).

2.3.3 Causality (effect)

A fundamental piece of information required for causality is the definition of the effect (positive or negative) that the regulator entity has on the regulated entity. Such information was completely missing in the PSI-MI CV. To address this issue, we created the 'causal statement' term. Under the causal statement parent term, it is now possible to find all the necessary definitions to explain the biological effect that (the function of) an entity has on (the function of) another entity and therefore provide more detail on how a modulator entity acts on a modulated entity. A modulator entity can act by up- or down-regulating (the function of) another entity. The new terms also enable the users to distinguish between the types of regulation and to specify whether the regulation of the modulator entity acts on the activity or the quantity (by controlling the expression or the stability) of the modulated entity (Fig. 2 and Supplementary Table S1).

2.3.4 Interaction type (mechanism)

The major goal of defining terms for the CausalTAB is to enable representation of interactions occurring in signalling events. Many resources, such as SIGNOR, IntAct and GO, capture data on molecular events underlying causal interactions. In a biological context, signalling events are mostly cascades of physical interactions, enzymatic reactions and post-translational modifications.

The PSI-MI ontology already contained all the interaction type terms necessary to represent 'direct causal interaction' such as molecular interactions and enzymatic reactions between the partners that is the regulator entity being immediately upstream of the target one (see branch 'Interaction type' in the PSI-MI CV).

However, the information reported in the literature about the molecular mechanism through which a molecule or an environmental condition influences the status of a downstream entity can be limited. SIGNOR and SignaLink also annotate causal interactions where the regulator is not immediately upstream of the target. For example, the information that DNA damage can induce intermolecular autophosphorylation of the protein ATM, through the up-regulation of its kinase activity (Bakkenist and Kastan, 2003).

To enable the representation of this data and to ensure the distinction between 'direct causal interactions' and 'indirect causal interactions', we created a new interaction type: 'Functional association' defined as

> Binary relationship between biological entities when one of them modulates the other in terms of function, expression, degradation or stability of the other and the relationship between the partners cannot be ascertained as direct, so intermediate steps are implicitly present. This relation specifically does not imply a physical interaction between the entities involved (Supplementary Table S1).

To add further information on 'Functional associations', we created the new 'causal regulatory mechanism', defined as:

> Type of relationship between entities involved in a causal interaction. This term is to be used only to describe the effect of a modulator entity A on a modulated entity B when A is not immediately upstream of B (Fig. 2 and Supplementary Table S1).

To conclude, the result of this revision-integration effort is the creation of 29 new terms (Supplementary Table S1), which allow the description of signalling relationships reported in all the repositories mentioned above, using the PSI-MI schema. The full set of terms is also available at https://www.ebi.ac.uk/ols/ontologies/mi/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FMI_2233.

2.4 The new PSI-MITAB 2.8 version or CausalTAB

The new format derives from the PSI-MI standard, enhanced to capture causal interactions among biological entities. PSI-MITAB 2.7 has been extended by adding four more columns allowing the description of signalling data and, therefore, updated to the new PSI-MITAB 2.8 version or CausalTAB.

CausalTAB maintains the same structure as previous versions of HUPO-PSI MITAB: it is a tab-delimited file where each row corresponds to an interaction. The file is organized in two main parts: one part reports annotations on the interactors while the second part reports annotations on the occurring interaction. The extended PSI-MITAB format now has 46 columns instead of 42 (Fig. 3 and Supplementary Table S2). This ensures a compatible merge of non-directional binary interactions with both directional and causality data, when available.

The PSI-MITAB 2.7 already contained many columns required to define the elements of a causal interaction, for example: 'Biological role(s) interactor A or B', 'Feature(s) interactor A or B', 'Interactor type'.

In particular, the 'Interaction type' column will contain all the terms that describe the type of relationship between entities involved in a causal interaction (Fig. 3 and Supplementary Table S2). Terms, such as 'phosphorylation reaction' (MI:0217) or 'physical association' (MI:0915) or 'ubiquitination reaction' (MI:0220), are already present in the PSI-MI CV, under the 'Interaction type' CV term and depict all the mechanism underlying a 'direct causal interaction'. 'Indirect causal Interactions' are identifiable by the new interaction type: 'Functional association'.

The four new columns inserted allow a full export of causal data (Fig. 3) and refer to the new CV terms described in the previous paragraph. We, here, briefly discuss the new columns:

The 'Causal Statement' column is designed to report the effect of modulator entity A on a modulated entity B. It might contain any child term of 'causal statement', including new terms such as 'up-regulates activity' (MI:2236) that define 'the effect of a modulator entity A on a modulated entity B that increases the frequency, rate or extent of the molecular function of B, an elemental biological

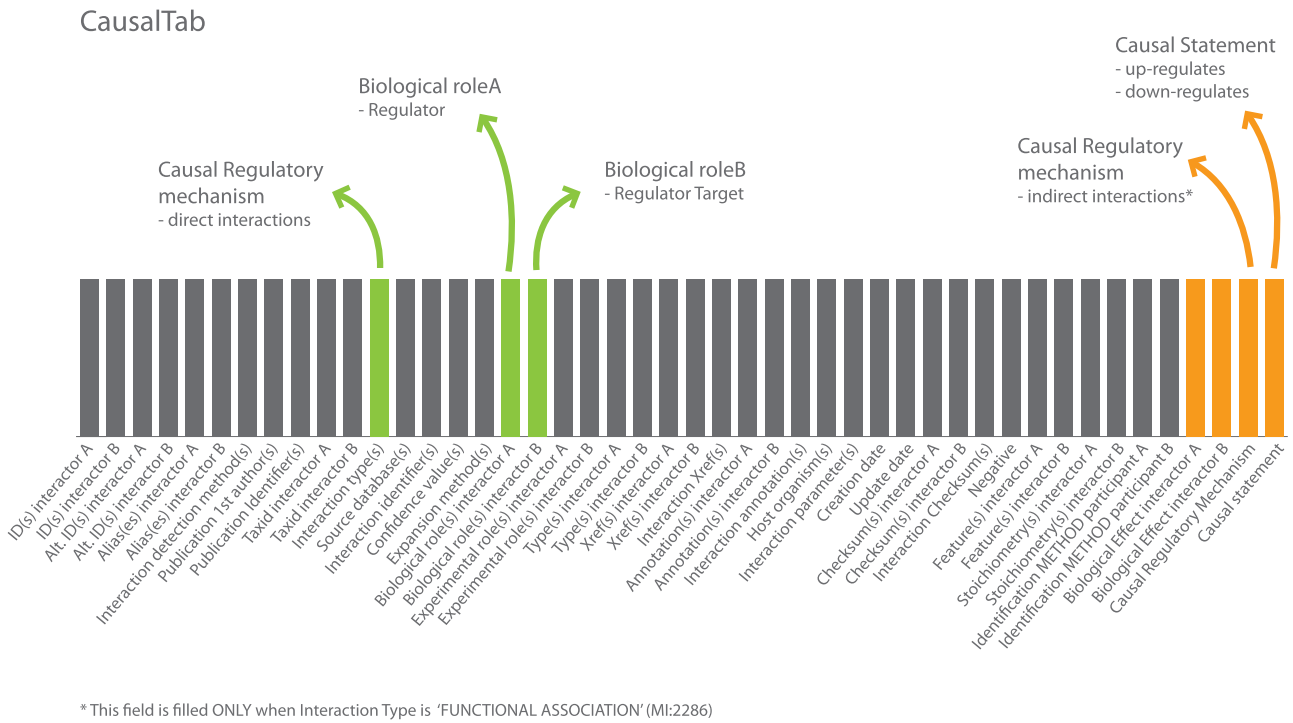

Fig. 3. A schematic representation of the CausalTAB. The last four columns represent the new columns introduced in the PSI-MI TAB 2.8. The 'Causal regulatory mechanism', 'Biological Role A' and 'Biological Role B' columns indicated columns of the PSI-MI TAB necessary to describe a causal interaction

activity occurring at the molecular level, such as catalysis or binding' (Figs 1 and 3).

The 'Biological Effect Interactor A' column contains the GO term associated with the Molecular Function of interactor A that is responsible for its regulatory activity. For example, this column will contain the 'kinase activity' GO term, for a kinase phosphorylating its substrate; or the 'RNA polymerase II transcription factor activity, sequence-specific DNA binding' GO term for a transcription factor binding the promoter sequence of its target gene.

Similarly, the 'Biological Effect Interactor B' column contains the GO term associated with the Molecular Function of interactor B that is modulated by the entity A.

The 'Causal Regulatory Mechanism' contains terms that describe indirect causal interactions, such as 'post transcriptional regulation', 'transcriptional regulation' and 'translation regulation', where the effect of entity A is not necessary immediately upstream the entity B. These terms have always to be associated with the term 'Functional Association' at the 'Interaction Type' level (Supplementary Table S2).

The new PSI-MITAB 2.8, under the name CausalTAB, is already downloadable from the download page the SIGNOR database (https://signor.uniroma2.it/downloads.php) (see also Supplementary Table S2), and will also become a download option through a future version of the PSICQUIC web service (Aranda *et al.*, 2011) (see Section 2.5).

2.5 Tools

Over the years, several tools have been developed and maintained to empower the PSI-MI format usage. These tools allow, for example, graphical network representation, data format conversions, and use of the schema validation, and are all accessible through the PSI-MI web pages (http://www.psidev.info/groups/molecular-interactions). The new PSI-MITAB 2.8 format is now compatible with most of these tools. The PSI Common QUery InterfaCe (PSICQUIC) web service (Aranda *et al.*, 2011; del-Toro *et al.*, 2013), an application that allows the retrieval of PSI-MI standardized interaction data, has recently been updated to be compatible with the new PSI-MITAB 2.8 format. The Molecular Interactions Query Language (MIQL), a common way to access the data in PSICQUIC and perform search queries has also been updated to the version 2.8 to allow querying the four new columns that describe causality. Information details about the PSI-MITAB 2.8 format, the MIQL 2.8 extension, installation instructions and usage examples of clients that query the PSICQUIC service in many programming languages, as well as the new reference implementation of PSICQUIC that supports the CausalTAB, can be found on the online PSICQUIC documentation Github site (https://psicquic.github.io/).

Moreover, common methods for interpreting omics data such as GO-based functional classification or Gene Set Enrichment Analyses (Subramanian *et al.*, 2005) are sufficient for deciphering general areas of biology that are altered in response to a stimulus, but cannot model the signalling response or uncover the mechanisms of action. This requires causal network analysis methods, and while suitable algorithms are freely available (Bradley and Barrett, 2017), publicly available causal interaction data needs to increase to encourage wider uptake and application for these more recently developed methods.

3 Discussion

There is a strong drive from the scientific community towards the adoption of the FAIR (Findable, Accessible, Interoperable, Reusable) Data Principles that provide data resources, tools and vocabularies with guidelines that promote data availability and reusability by other users (Wilkinson *et al.*, 2016).

Our work has been inspired by the FAIR principles and stimulated by the need to make biological signalling data from disparate

resources compatible with each other and consequently interoperable and available to the scientific community.

The signalling community has discussed, revised and formalized the new standards during the Causal Reasoning Workshop in 2016, the Seattle (2015), Ghent (2016), Beijing (2017) and Heidelberg (2018) HUPO-PSI meetings (http://www.psidev.info/events), the Malta and Lisbon COST Action 15205 GREEKC (http://greekc.org/) workshops and in Rome during an ELIXIR Italy funded Curation Workshop on Molecular and Causal Interactions. The proposed format represents the consensus view of the signalling community and it has also been discussed with potential user groups, for example members of the EMBL-EBI Industry program at workshops in Boston and San Diego, USA.

Moreover, during the curation workshop, we have developed new curation rules to make the use of CausalTAB more efficient and useful. Those rules have been accepted by the IMEx Consortium (Orchard *et al.*, 2012) and SignaLink curators and will be adopted for the annotation of signalling data.

In the future, we aim to define the fundamental and mandatory information to capture in a causal interaction (e.g. specific information for causal interaction and entity objects, recommendation on ontology to use), by building a guideline called the 'Minimum Information about a Causal Statement', to unify the representation of causal interactions. Moreover, we will map our new terms to the RO terms that also contain some terms suitable for the description of causal relations and their directionality.

In parallel, we plan to implement more informative formats to represent causality. The release of a new version of PSI-MI XML, able to capture every aspect of causality is under discussion within the HUPO-PSI community. Also, we intend to extend the current version of MI-JSON (a JSON-compatible version of PSI-MITAB 2.7 built in the JAMI library) (Sivade Dumousseau *et al.*, 2018b) to include the new CausalTAB/PSI-MITAB 2.8.

In terms of curation approach, the resources will continue their independent, but coordinated annotation of causal interactions. SIGNOR, SignaLink and the GO-CAM initiative are ongoing projects focusing on the capture of signalling interaction data. Introduction of causality statements into IMEx resources (especially IntAct and MINT databases) will be first limited to newly created entries, although identification and re-curation of existing entries, to add causality, remains a long-term goal, dependent on resources availability.

To conclude, thanks to this community effort, we are now able to capture causal interaction data in a structured format, compatible with PSI-MI data and tools. The development of new PSI-MI CV terms specific to describing and defining signalling events has allowed the update of the HUPO-PSI MITAB to the new 2.8 version that contains the appropriate descriptions for causal interaction. Causal data can now be annotated in a structured format and exchanged and analyzed by the user. Moreover, thanks to its structure, the PSI-MI CV can be modified and updated according to community needs.

Acknowledgements

We thank Birgit Meldal for critically reading and reviewing the manuscript.

Funding

This work was supported by ELIXIR-IIB, the Italian Node of the European ELIXIR infrastructure [to L.L.]; Italian Association for Cancer Research, triennial fellowship [Starwood Hotels & Resorts ID. 18137 to L.P.]; Gene Regulation Ensemble Effort for the Knowledge Commons [CA15205 to M.K., J.Z.]; the DEPTH Project of the European Research Council [Grant Agreement 322749 to G.C., L.P.]; AIRC [Project IG 2017 ID. 20322 to G.C., L.P.]; the Research Council of Norway [Project 247727 to A.L., M.A.]; the Norwegian University of Science and Technology's Strategic Research Area 'NTNU Health' [to V.T.]; the ERACoSysMed grant COLOSYS, RCL by the British Heart Foundation [RG/13/5/30112 to V.T., J.Z.]; the National Institute for Health Research University College London Hospitals Biomedical Research Centre [to V.T., J.Z.]; EMBL core funding and Open Targets [Grant Agreement OTAR-044 to H.H., P.P., S.O., N.D., L.P.]; fellowship in computational biology at Earlham Institute (Norwich, UK) in partnership with the Quadrams Institute (Norwich, UK) and strategically supported by Biological Sciences Research Council [BB/J004529/1, BB/P016774/1 to T.K.].

Conflict of Interest: none declared.

References

Aranda,B. *et al.* (2011) PSICQUIC and PSISCORE: accessing and scoring molecular interactions. *Nat. Methods*, **8**, 528–529.

Ashburner,M. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.

Bakkenist,C.J. and Kastan,M.B. (2003) DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature*, **421**, 499–506.

Barabási,A.L. and Oltvai,Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.*, **5**, 101–113.

Bradley,G. and Barrett,S.J. (2017) CausalR: extracting mechanistic sense from genome scale data. *Bioinformatics*, **33**, 3670–3672.

Csabai,L. *et al.* (2018) SignaLink: multilayered regulatory networks. *Methods Mol. Biol.*, **1819**, 53–73.

del-Toro,N. *et al.* (2013) A new reference implementation of the PSICQUIC web service. *Nucleic Acids Res.*, **41**, W601–606.

Demir,E. *et al.* (2010) The BioPAX community standard for pathway data sharing. *Nat. Biotechnol.*, **28**, 935–942.

Deutsch,E.W. *et al.* (2017) Proteomics standards initiative: fifteen years of progress and future work. *J. Proteome Res.*, **16**, 4288–4298.

Hermjakob,H. *et al.* (2004) The HUPO PSI's molecular interaction format—a community standard for the representation of protein interaction data. *Nat. Biotechnol.*, **22**, 177–183.

Hucka,M. *et al.* (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524–531.

Kanehisa,M. *et al.* (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.*, **45**, D353–361.

Kaushal,J.B. *et al.* (2017) The regulation of Hh/Gli1 signaling cascade involves Gsk3β-mediated mechanism in estrogen-derived endometrial hyperplasia. *Sci. Rep.*, **7**, 6557.

Kerrien,S. *et al.* (2007) Broadening the horizon—level 2.5 of the HUPO-PSI format for molecular interactions. *BMC Biol.*, **5**, 44.

Lee,M.J. and Yaffe,M.B. (2016) Protein regulation in signal transduction. *Cold Spring Harb. Perspect. Biol.*, **8**, 1–19.

Lodish,H. *et al.* (2000) *Molecular Cell Biology*. 4th edn. W. H. Freeman, New York.

Orchard,S. *et al.* (2014) The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.*, **42**, D358–363.

Orchard,S. *et al.* (2012) Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nat. Methods*, **9**, 345–350.

Orchard,S. *et al.* (2007) The minimum information required for reporting a molecular interaction experiment (MIMIx). *Nat. Biotechnol.*, **25**, 894–898.

Perfetto,L. *et al.* (2016) SIGNOR: a database of causal relationships between biological entities. *Nucleic Acids Res.*, **44**, D548–554.

Sidiropoulos,K. *et al.* (2017) Reactome enhanced pathway visualization. *Bioinformatics*, **33**, 3461–3467.

Sivade Dumousseau,M. *et al.* (2018a) Encompassing new use cases—level 3.0 of the HUPO-PSI format for molecular interactions. *BMC Bioinform.*, **19**, 134.

Sivade Dumousseau,M. *et al*. (2018b) JAMI: a Java library for molecular interactions and data interoperability. *BMC Bioinform*., **19**, 133.

Smith,B. *et al*. (2005) Relations in biomedical ontologies. *Genome Biol*., **6**, R46.

Smoot,M.E. *et al*. (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, **27**, 431–432.

Subramanian,A. *et al*. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA*, **102**, 15545–15550.

The Gene Ontology Consortium. (2017) Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res*. **45**, D331–D338.

Türei,D. *et al*. (2016) OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat. Methods*, **13**, 966–967.

Wilkinson,M.D. *et al*. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data*, **3**, 160018.

Yang,L. *et al*. (2010) Activation of the hedgehog-signaling pathway in human cancer and the clinical implications. *Oncogene*, **29**, 469–481.