

# A Brief History of COMBINE

Chris J. Myers, Gary Bader, Padraig Gleeson, Martin Golebiewski,  
Michael Hucka, Nicolas Le Novère, David P. Nickerson, Falk Schreiber,  
and Dagmar Waltemath

COMBINE

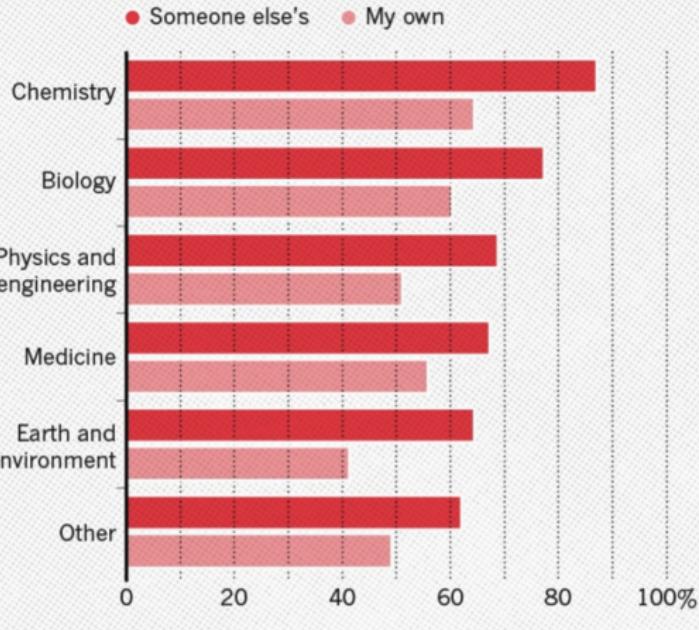
Winter Simulation 2017

December 5, 2017

# Reproducibility Crisis

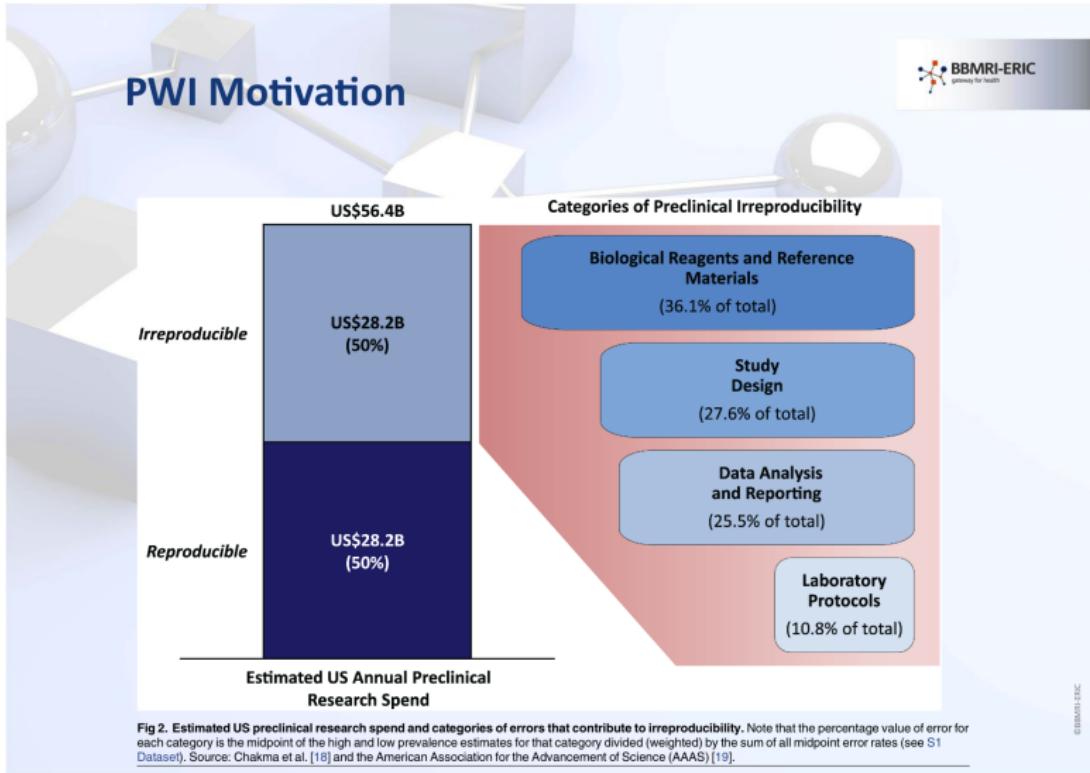
## HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.



(V. Simonyan, Center for Biologics Evaluation and Research FDA, USA)

# Reproducibility Crisis



# Reproducibility Crisis

The screenshot shows the header of the Science journal website. The main title "Science" is on the left, followed by navigation links: Home, News, Journals, Topics, and Careers. Below the header is a large banner featuring the logo of the Chinese University of Defense Technology (国防科技大学) and a green banner with the text "Welcome to join us >>". To the right of the banner, there are links for "Institution: UNIVE", "Log in", and "My acco".

SHARE

PERSPECTIVE



## Reproducible Research in Computational Science



Roger D. Peng

[+ See all authors and affiliations](#)



0

Science 02 Dec 2011;  
Vol. 334, Issue 6060, pp. 1226-1227  
DOI: 10.1126/science.1213847

Article

Figures & Data

Info & Metrics

eLetters

PDF

### Abstract

Computational science has led to exciting new developments, but the nature of the work has exposed limitations in our ability to evaluate published findings. Reproducibility has the potential to serve as a minimum standard for judging scientific claims when full independent replication of a study is not possible.

# Reproducibility Crisis

## Quantifying Reproducibility in Computational Biology: The Case of the Tuberculosis Drugome

Daniel Garijo, Sarah Kinnings, Li Xie, Lei Xie, Yinliang Zhang, Philip E. Bourne  , Yolanda Gil  

Published: November 27, 2013 • <https://doi.org/10.1371/journal.pone.0080278>

Article	Authors	Metrics	Comments	Related Content
				

### Abstract

Introduction

Methods and Analysis

Discussion

Conclusions

Supporting Information

Author Contributions

References

---

Reader Comments (0)

Media Coverage

Figures

### Abstract

How easy is it to reproduce the results found in a typical computational biology paper? Either through experience or intuition the reader will already know that the answer is with difficulty or not at all. In this paper we attempt to quantify this difficulty by reproducing a previously published paper for different classes of users (ranging from users with little expertise to domain experts) and suggest ways in which the situation might be improved. Quantification is achieved by estimating the time required to reproduce each of the steps in the method described in the original paper and make them part of an explicit workflow that reproduces the original results. Reproducing the method took several months of effort, and required using new versions and new software that posed challenges to reconstructing and validating the results. The quantification leads to "reproducibility maps" that reveal that novice researchers would only be able to reproduce a few of the steps in the method, and that only expert researchers with advance knowledge of the domain would be able to reproduce the method in its entirety. The workflow itself is published as an online resource together with supporting software and data. The paper concludes with a brief discussion of the complexities of requiring reproducibility in terms of cost versus benefit, and a desiderata with our observations and guidelines for improving reproducibility. This has implications not only in reproducing the work of others from published papers, but reproducing work from one's own laboratory.

# Reproducibility Crisis

*An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete ... set of instructions [and data] which generated the figures.*

– David Donoho, 1998

# Standards to the Rescue



(source <https://www.eaglegenomics.com/do-data-standards-really-matter/>)

# Word of Warning

## HOW STANDARDS PROLIFERATE:

(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)

SITUATION:  
THERE ARE  
14 COMPETING  
STANDARDS.

14?! RIDICULOUS!  
WE NEED TO DEVELOP  
ONE UNIVERSAL STANDARD  
THAT COVERS EVERYONE'S  
USE CASES.



YEAH!

SOON:

SITUATION:  
THERE ARE  
15 COMPETING  
STANDARDS.

(source [xkcd.com](http://xkcd.com))

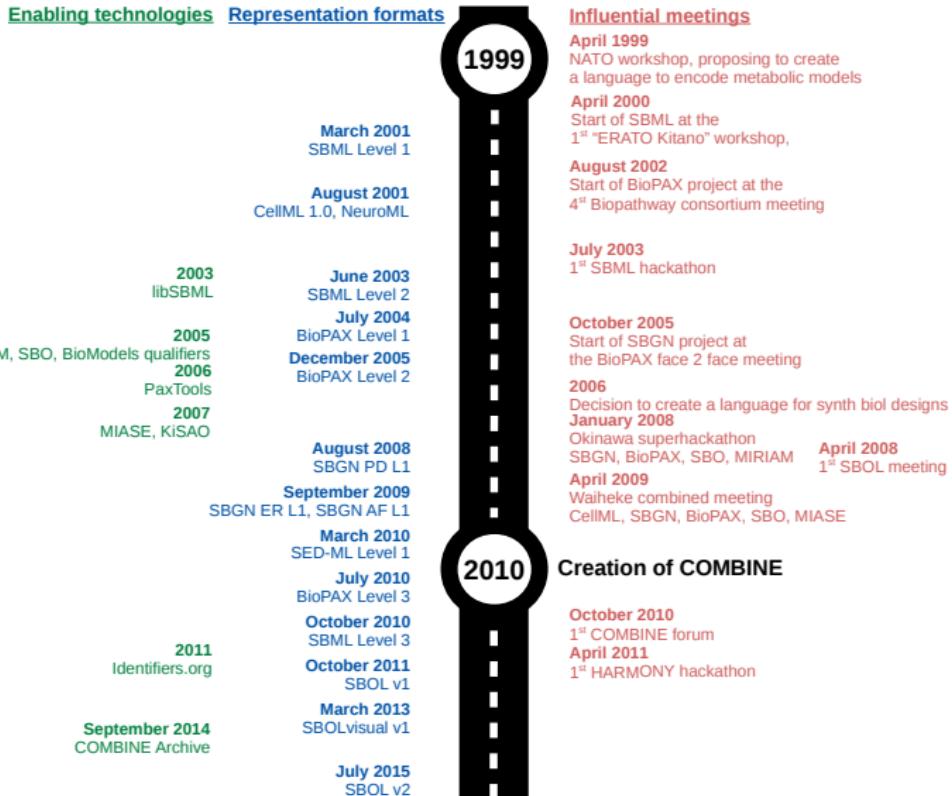
# Coordination of Standard Development in Systems/Synthetic Biology



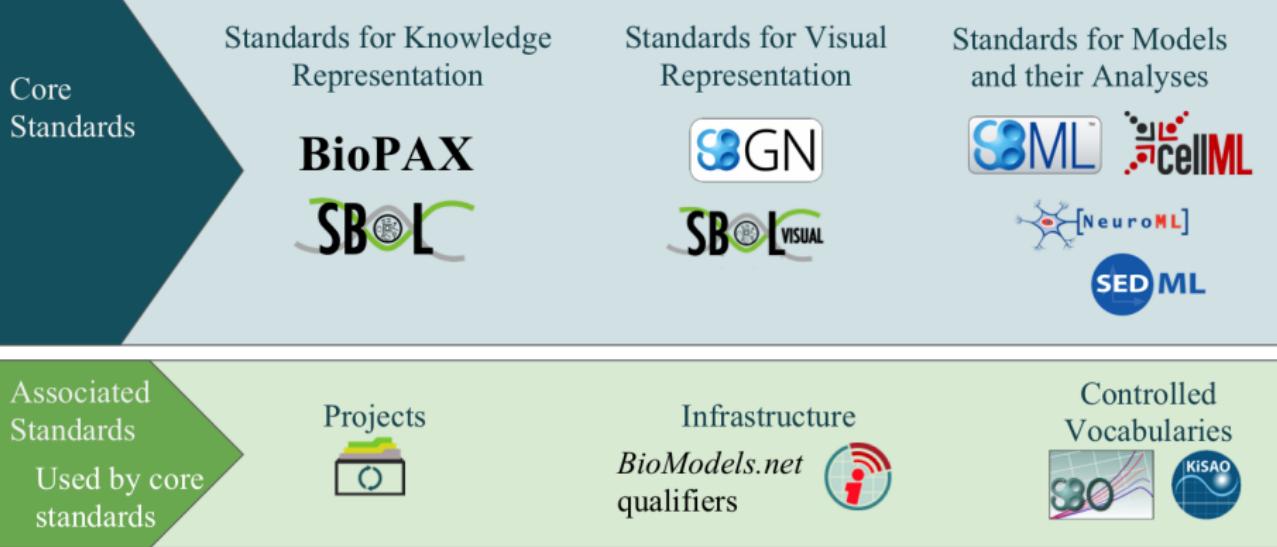
<http://co.mbine.org>

- Tasks and Actions:
  - Computational Modeling in Biology Network
  - Concerted meetings of standards: HARMONY & the COMBINE Forum
  - Training in application of standards (COMBINE tutorials)
  - Coordinate standards development
  - Develop common procedures & tools
  - Provide a recognized voice

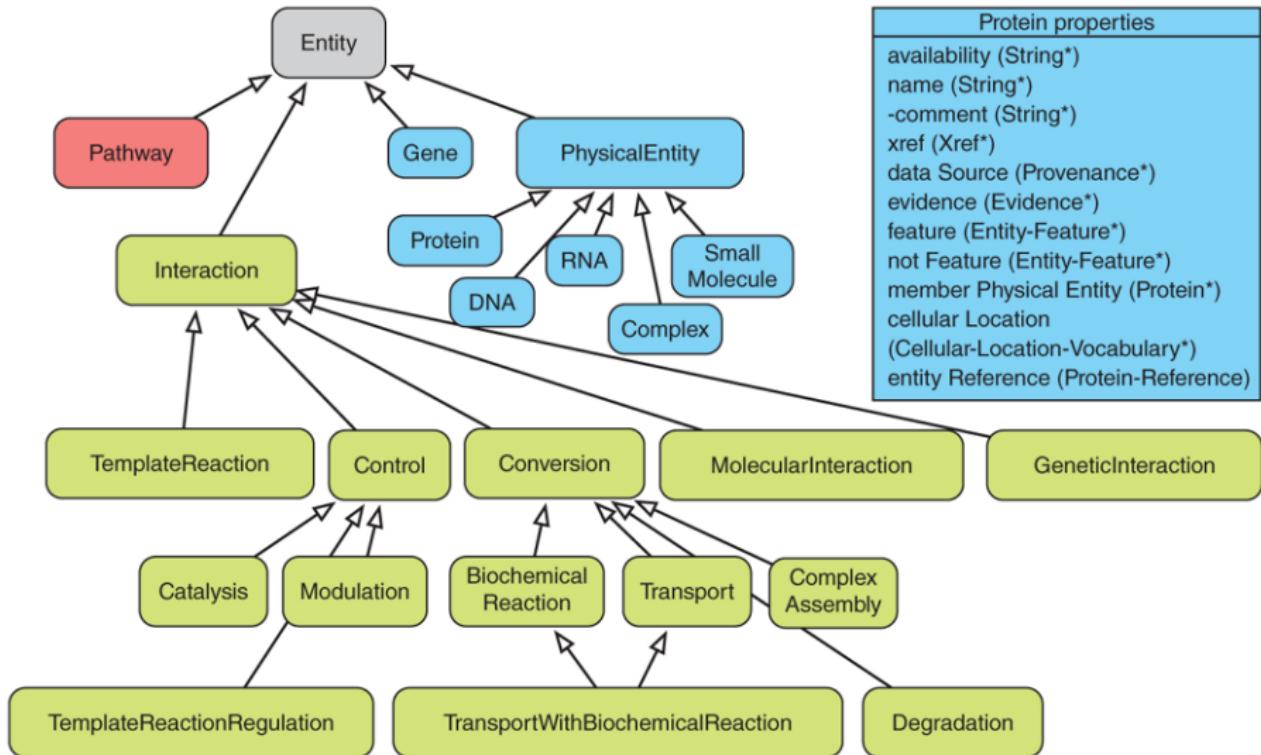
# COMBINE History



# COMBINE Overview



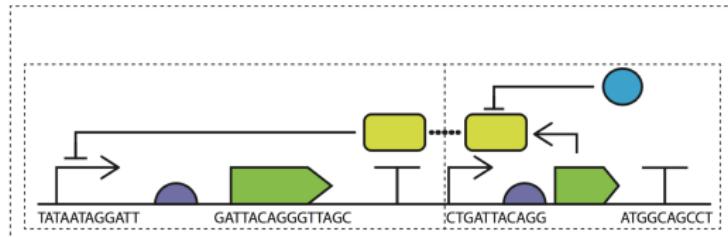
# BioPax: Biological Pathways



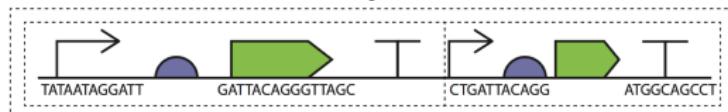
Demir et al., Nature Biotechnology (2010)

# Synthetic Biology Open Language (SBOL)

**SBOL**  
version 2



**SBOL**  
version 1



GenBank

Promoter RBS CDS Terminator Promoter RBS CDS Terminator  
TATAATAGGATTCCGAATGGATTACAGGGTTAGCAAATGGCAGCCTGATTACAGGGTTAGCAAATGGCAGCCT

FASTA

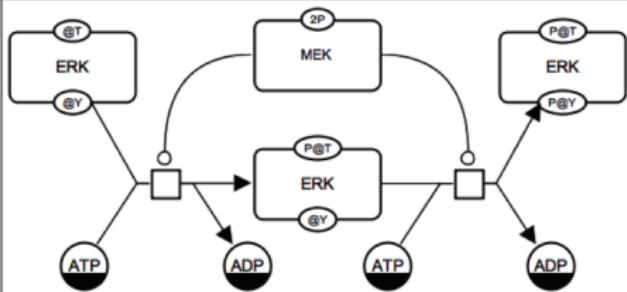
TATAATAGGATTCCGAATGGATTACAGGGTTAGCAAATGGCAGCCTGATTACAGGGTTAGCAAATGGCAGCCT

Galdzicki et al., Nature Biotechnology (2014)  
Roehner et al., ACS Synthetic Biology (2016)

# Systems Biology Graphical Notation (SBGN)

## Process Description

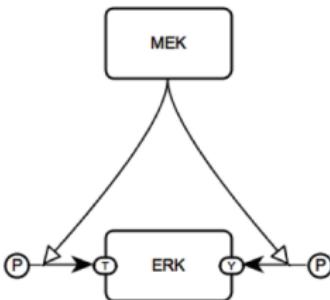
### maps



- ▶ Unambiguous
- ▶ Mechanistic
- ▶ Sequential
- ▶ Combinatorial explosion

## Entity Relationships

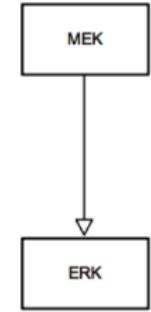
### maps



- ▶ Unambiguous
- ▶ Mechanistic
- ▶ Non-Sequential

## Activity Flow

### maps



- ▶ Ambiguous
- ▶ Conceptual
- ▶ Sequential

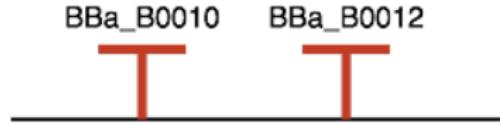
Le Novère et al., Nature Biotechnology (2009)

# SBOL Visual (SBOLv)

BBa\_F2620

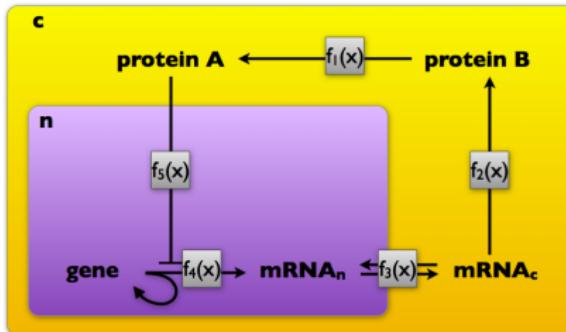


BBa\_B0015



Quinn et al., PLoS Biology (2015)

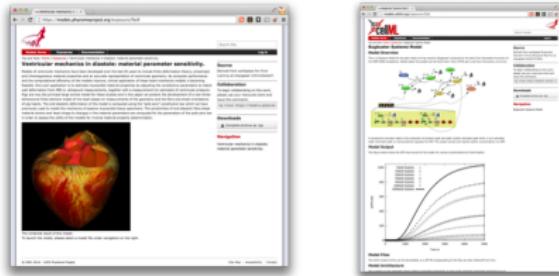
# Systems Biology Markup Language (SBML)



Hucka et al., Bioinformatics (2003)

- Models are composed of a number of species (i.e. proteins, genes, chemical compounds, etc.) and reactions that transform these species.
- Also include parameters, functions, units, initial assignments, constraints, rules for continuous dynamics, & events for discontinuous state changes.
- Biological entities and context are defined mainly via (RDF) annotations.
- Models can be archived in the BioModels database and/or JWS Online.
- Supported by more than 280 tools, enabling researchers to create, annotate, simulate, store, exchange, and visualize models.

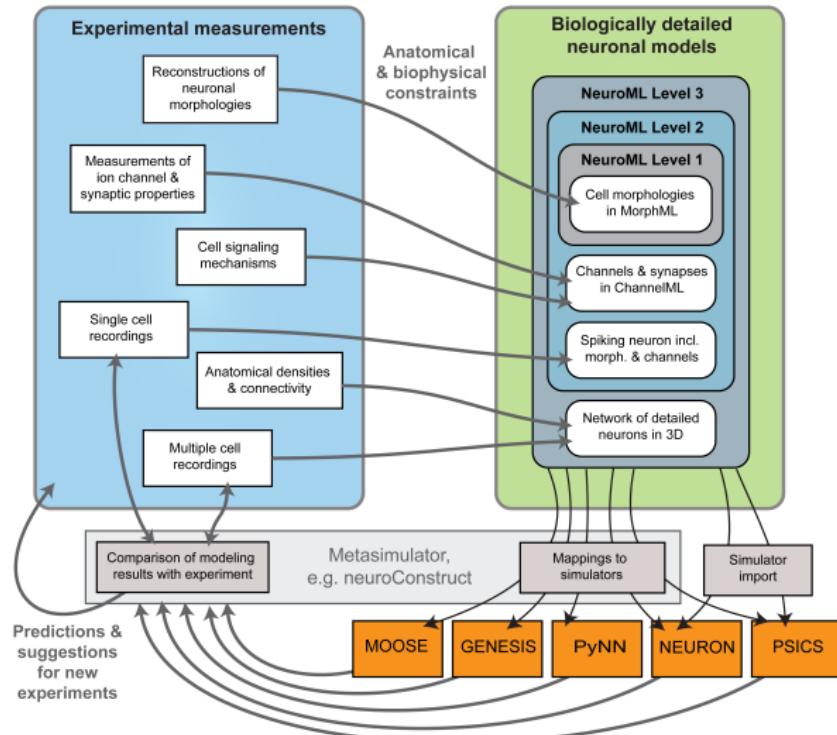
# CellML



Cuellar et al., Simulation (2003)

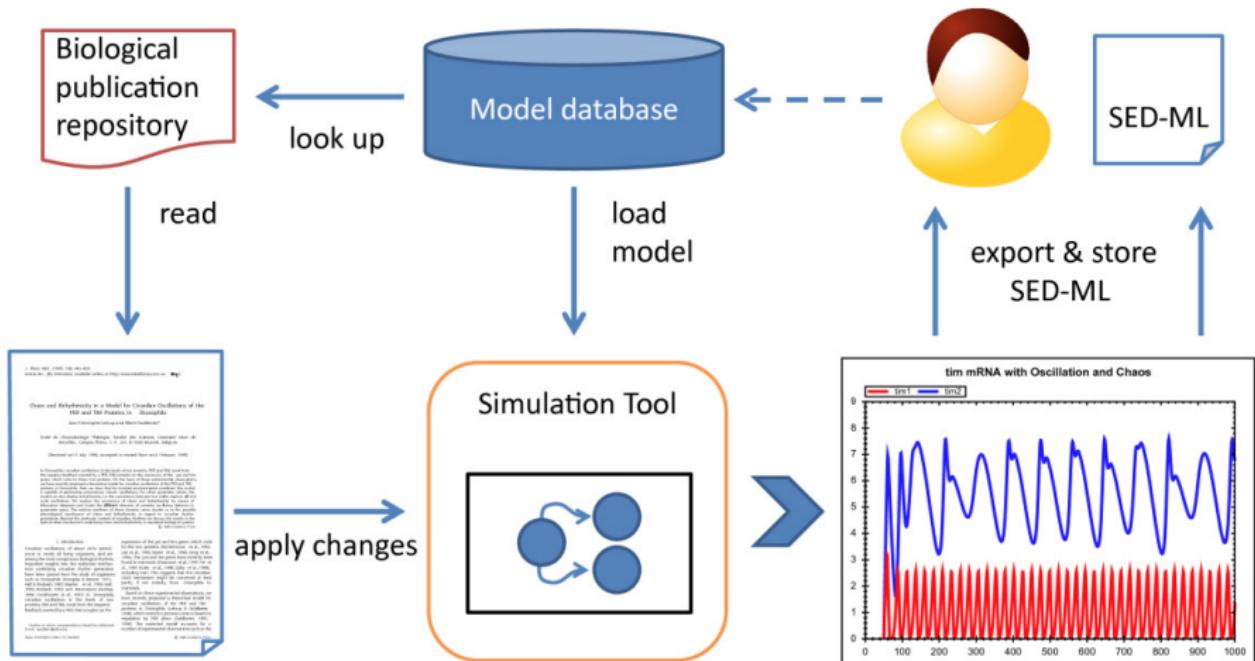
- Modular framework for encoding mathematical models:
  - Encodes models consisting of differential algebraic equations w/MathML.
  - Math is the primary data, biological context, provenance, etc., provided through annotations using RDF.
- All quantities require physical units ensuring unambiguous conversion.
- Can define hierarchies of modules to enable mathematical abstraction.
- External models can be imported enabling reuse.
- Free and open repository (<https://models.physiomeproject.org/>) supporting versioned model reuse, archiving, and collaboration.

# NeuroML: Computational Neuroscience Models



Gleeson et al., PLoS Computational Biology (2010)

# Simulation Experiment Description Markup Lang. (SED-ML)



Waltemath et al., BMC Systems Biology (2011)

# COMBINE Archive

**COMBINE Archive format =**  
single file that supports exchange  
of all information necessary for any  
modeling and simulation  
experiment

- Not SBML-specific at all
- Not programming-language specific
- Not domain specific

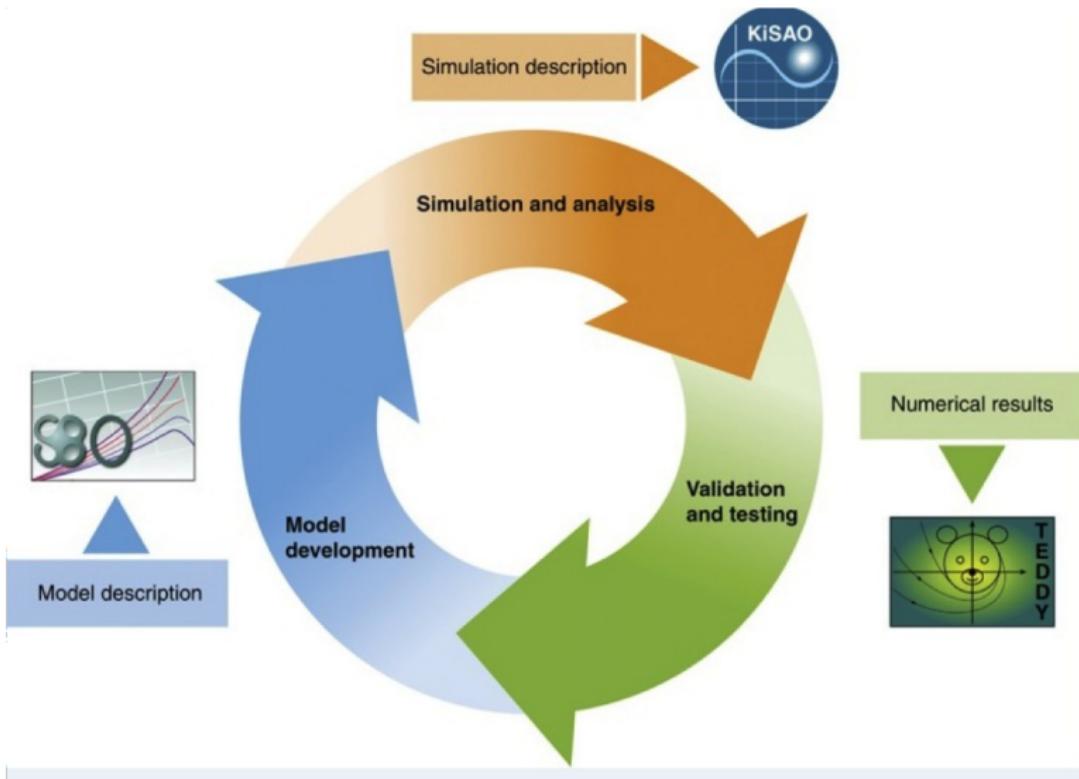


**OMEX** = file format for COMBINE Archive

- ZIP file containing manifest file (in XML form) + other files
- Use of ZIP leverages many existing programming libraries

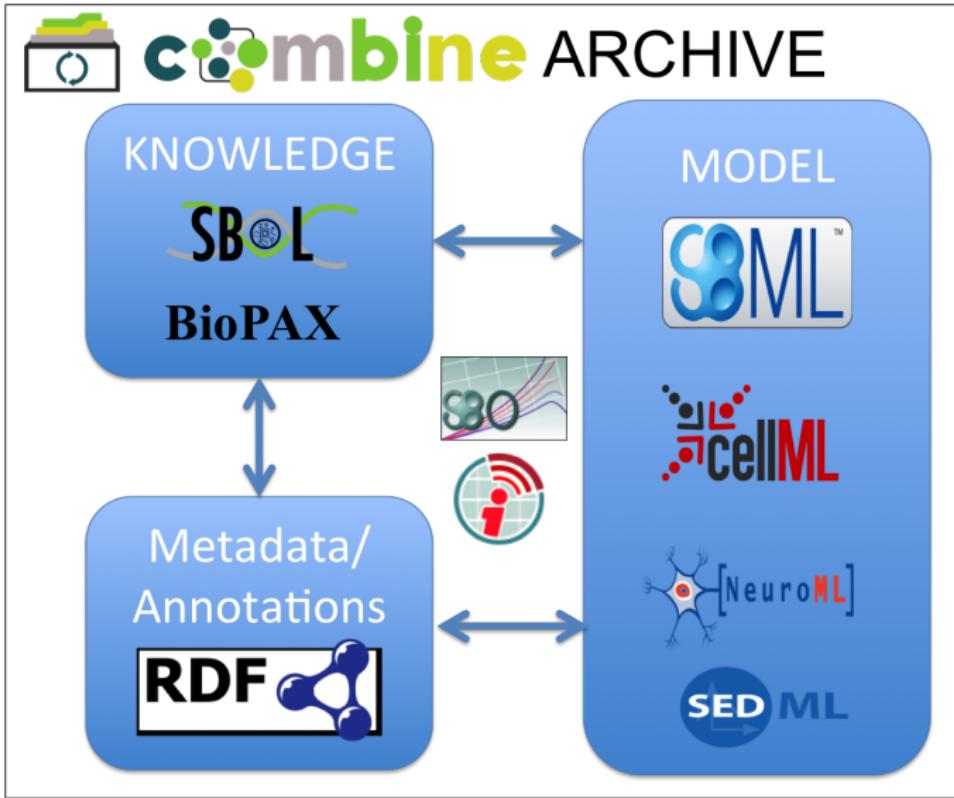
<http://combine.org/documents/archive>

# Ontologies and Controlled Vocabulary

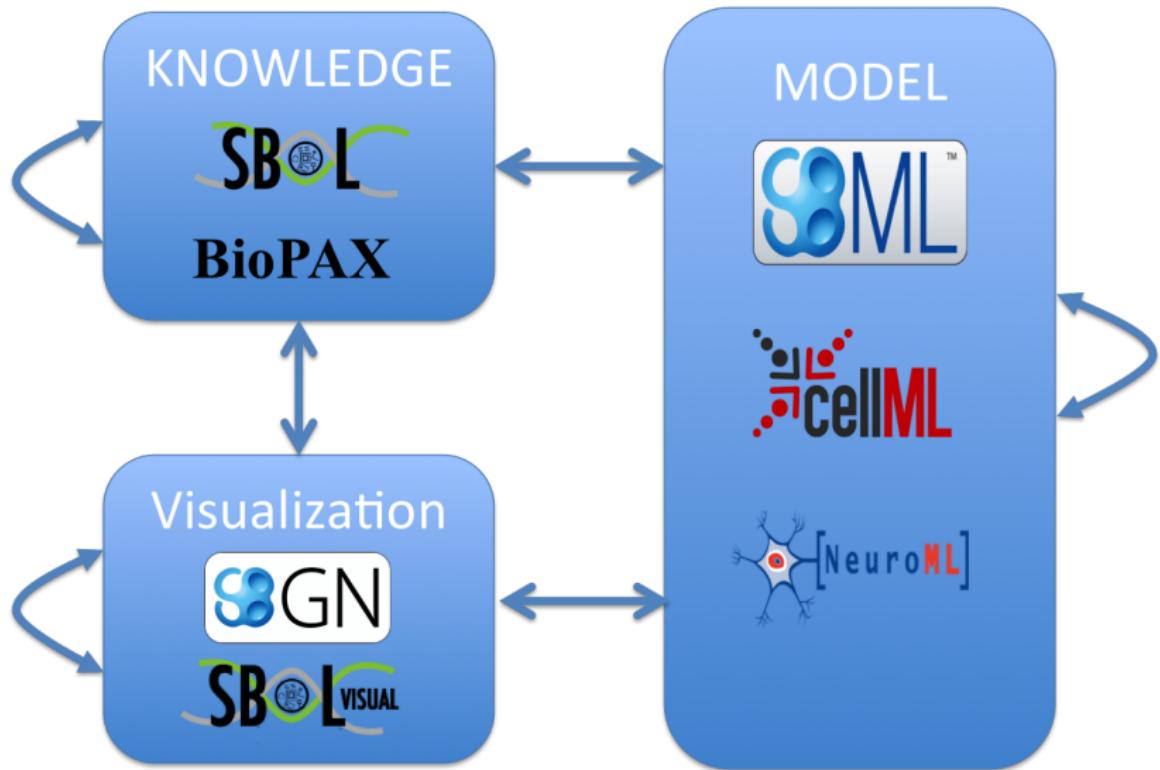


Courtot, et al., Mol Syst Biol. (2011)

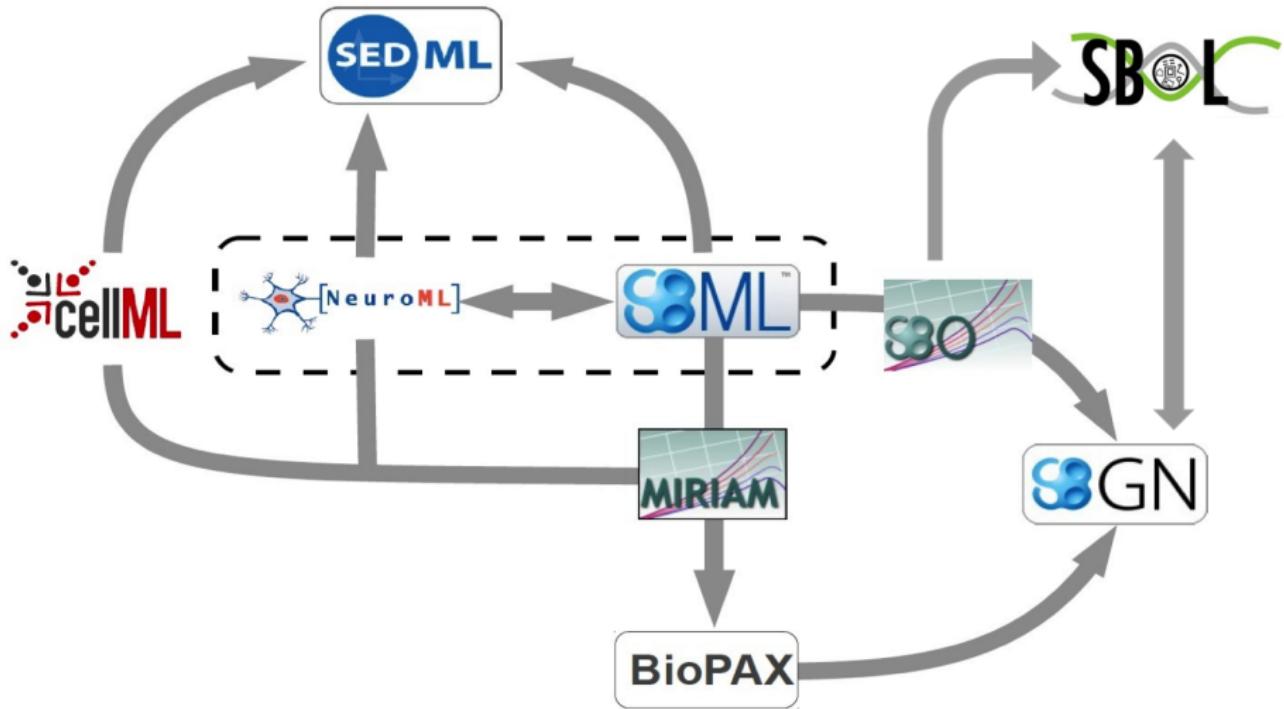
# Unified Metadata and Annotations



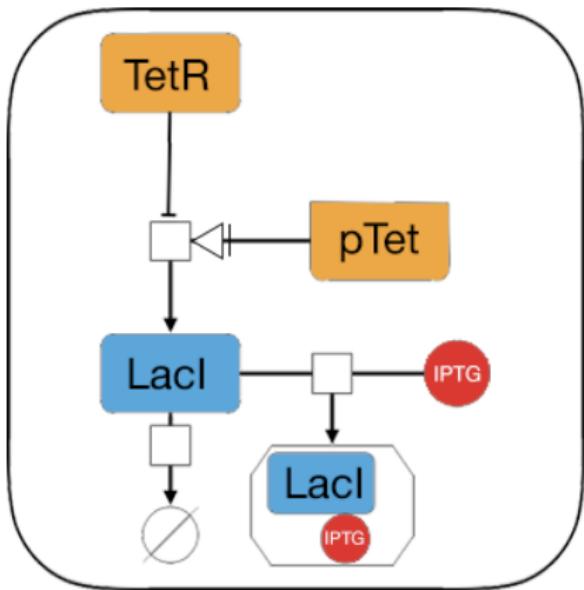
# Data Standard Conversions



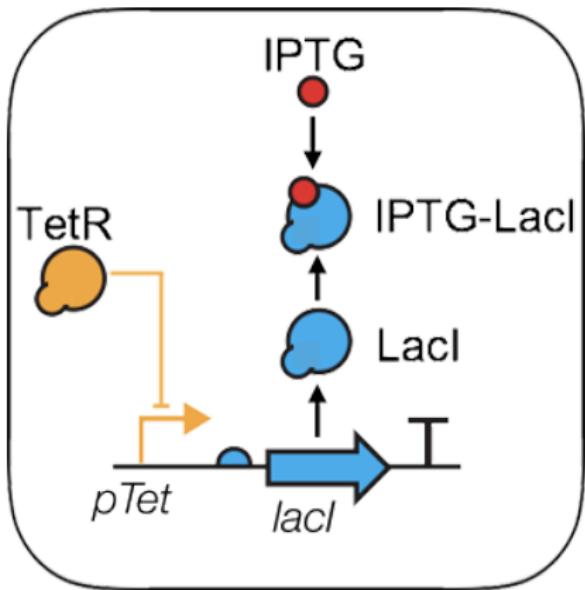
# Interfacing and Interoperability of Standards



# Unified Visualizations (SBGN/SBOLv)

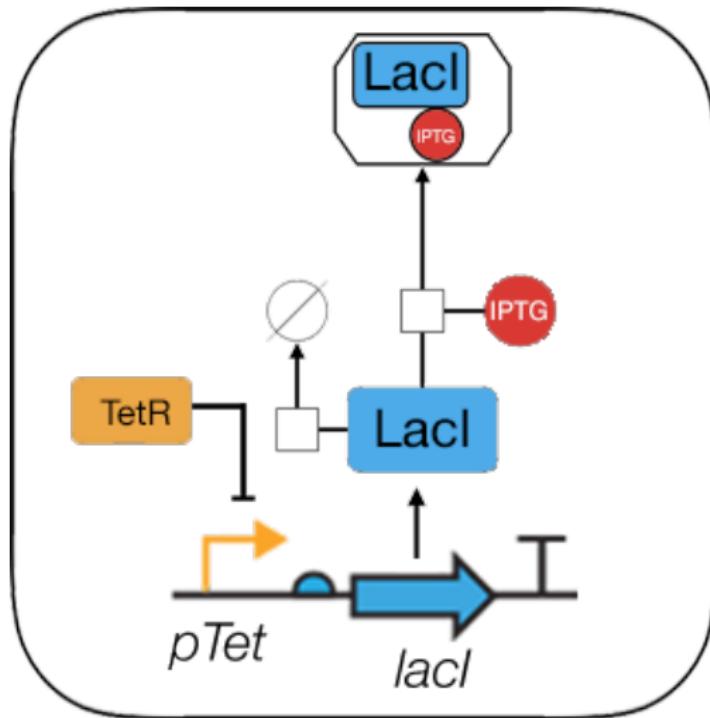


SBGN



SBOLv

# Unified Visualizations (SBGN/SBOLv)



Unified SBGN/SBOLv

# Repositories



# Pathway Commons

Access and discover data integrated from public pathway and interactions databases.

Pathway Commons, a web resource for biological pathway data.

[Data](#) [Tools](#) [FAQ](#) [Contact](#)

ps

**Search**

Search the entire collection of pathways  
names or gene IDs (e.g. 'glycolysis', 'TP53')

**PCViz**

Get details about genes and their interactions  
Gene IDs (e.g. 'MDK2 TP53')

<http://www.pathwaycommons.org>

<http://biomodels.net>



# SynBioHub

SynBioHub is a design repository for people designing biological constructs. It enables DNA and protein designs to be uploaded, then provides a shareable link to allow others to view them. SynBioHub also facilitates searching for information about existing useful parts and designs by combining data from a variety of sources.

**Search for useful parts and designs**



[Browse Public Designs](#)

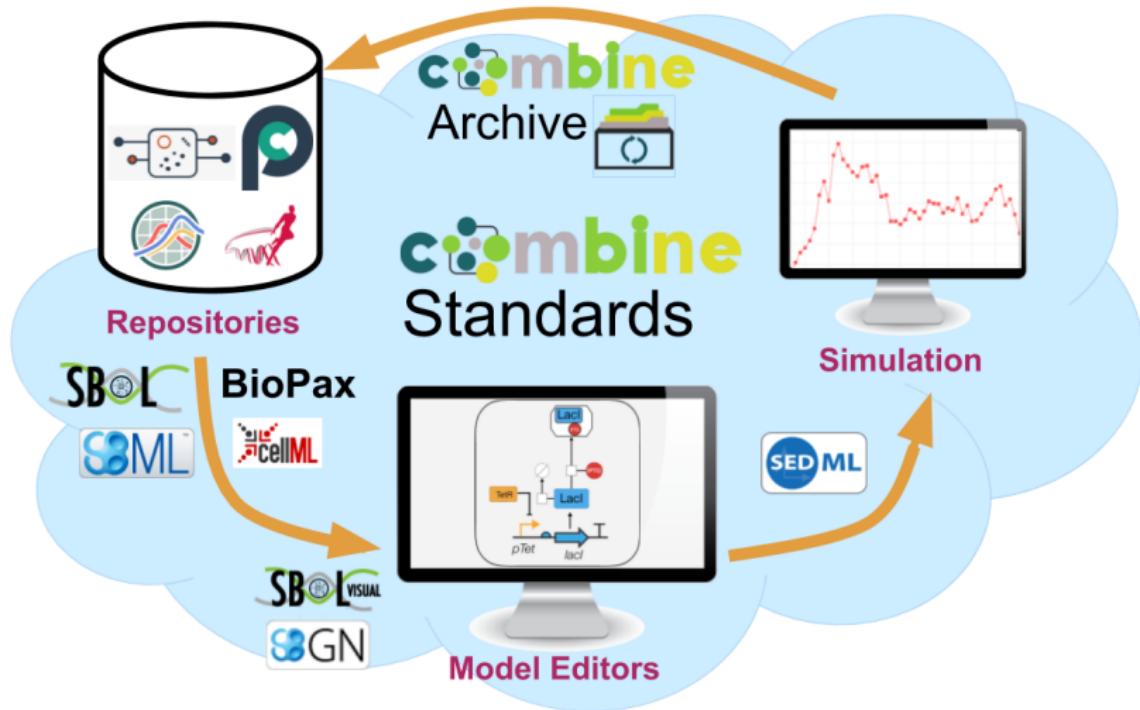
 **Upload your design for safekeeping**  
[Submit a Design](#)

 **Share designs for publication or collaboration**  
[Manage Submissions](#)

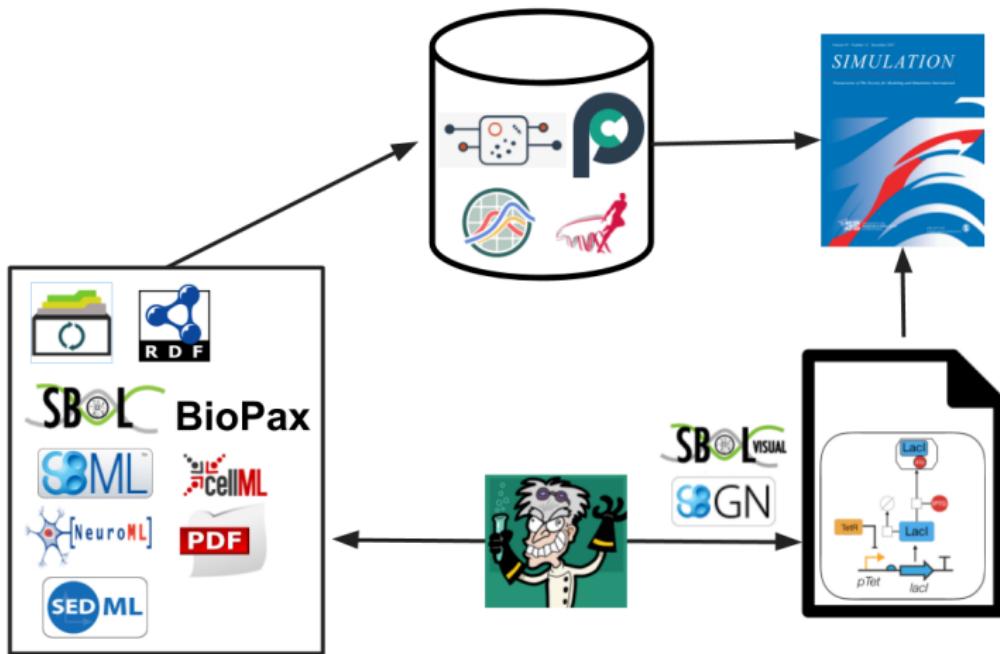
<https://synbiohub.org>

<https://models.physiomeproject.org>

# Standard Enabled System/Synthetic Biology Workflows



# Journal Workflow for Reproducibility



# Conclusion

*While reproducibility remains a challenge in computational modeling in biology, the COMBINE community and the standards being developed within this community have the potential to make this a fully reproducible scientific endeavor.*

# COMBINE Coordination Board



**Gary D. Bader**  
(University of  
Toronto, Canada)  
BioPAX Delegate



**Mike Hucka**  
(Caltech  
Pasadena, CA)  
SBML Delegate



**Chris Myers**  
(University of  
Utah, USA)  
SBOL Delegate



**David Nickerson**  
(University of  
Auckland, NZ)  
CellML Delegate



**Falk Schreiber**  
(Monash University  
Melbourne, AUS)  
SBGN Delegate



**Dagmar Waltemath**  
(University of  
Rostock, Germany)  
SED-ML Delegate



**Padraig Gleeson**  
(University College  
London, UK)  
NeuroML Delegate



**Martin  
Golebiewski**  
(HITS gGmbH,  
Germany)



**Nicolas Le Novère**  
(Babraham  
Institute, UK)

# COMBINE Coordination Board



**Gary D. Bader**  
(University of  
Toronto, Canada)  
BioPAX Delegate



**Sarah Keating**  
(EMBL-EBI,  
Hinxton, UK)  
SBML Delegate



**Chris Myers**  
(University of  
Utah, USA)  
SBOL Delegate



**David Nickerson**  
(University of  
Auckland, NZ)  
CellML Delegate



**Falk Schreiber**  
(Monash University  
Melbourne, AUS)  
SBGN Delegate



**Dagmar Waltemath**  
(University of  
Rostock, Germany)  
SED-ML Delegate



**Padraig Gleeson**  
(University College  
London, UK)  
NeuroML Delegate



**Martin  
Golebiewski**  
(HITS gGmbH,  
Germany)



**Nicolas Le Novère**  
(Babraham  
Institute, UK)

# Invitation

- You are invited to join the COMBINE community.
- Contact the COMBINE Coordinators or standard editors to join the appropriate mailing lists.
- Attend our upcoming events:



HARMONY 2018  
University of Oxford  
June 18-22



COMBINE Forum 2018  
Boston University  
October 8-12