



19th European Conference on Computational Biology

Planetary Health and Biodiversity

31st August 8th September 2020

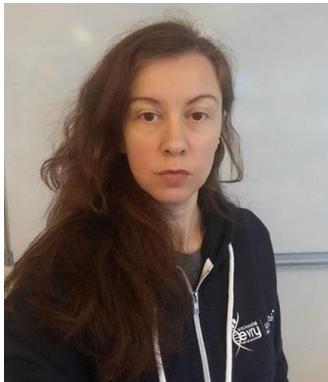
Building graphical and computational models in Systems Biology

NTB-T05

Computational modelling of
cellular processes: regulatory vs
metabolic systems

September 1st, 2020

Dr. Anna Niarakis – University of Evry, Paris-Saclay & INRIA Saclay



Associate Professor
GenHotel-UEVE, Genopole,
University of Paris-Saclay
Lifeware, INRIA, Saclay

@ anna.niaraki@univ-evry.fr
 Anna Niarakis



<https://www.genhotel.univ-evry.fr>



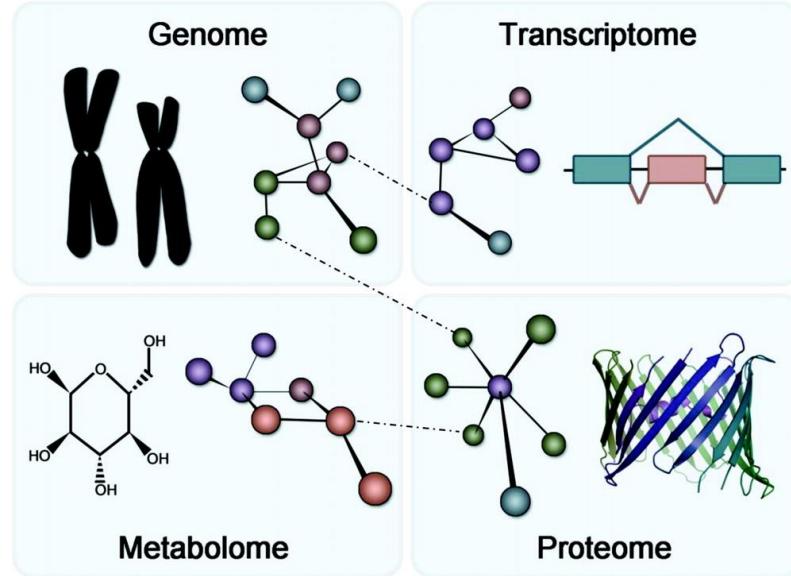
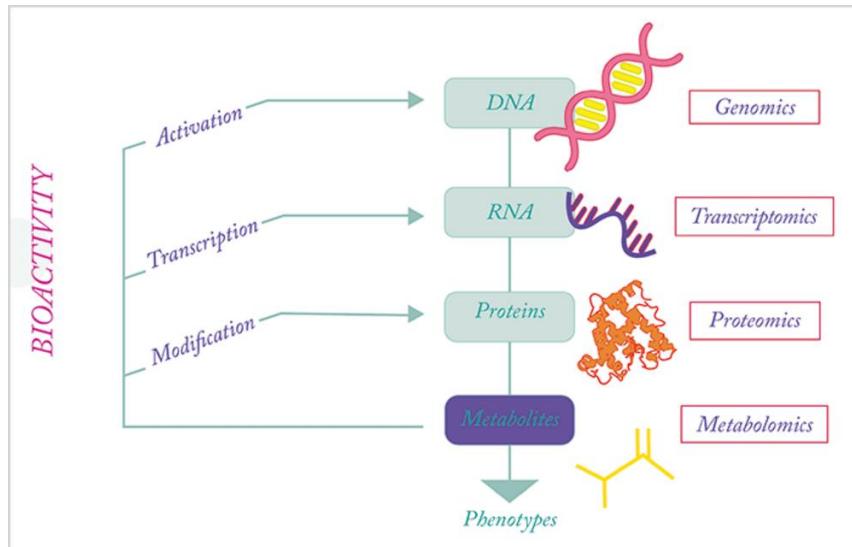
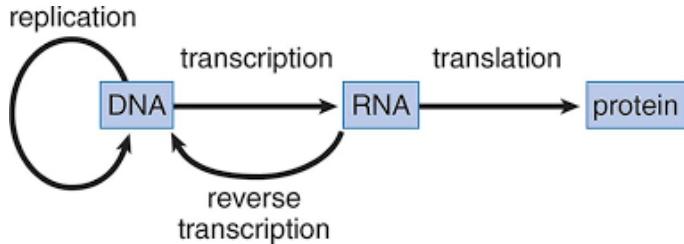
 informatics mathematics



Research Focus

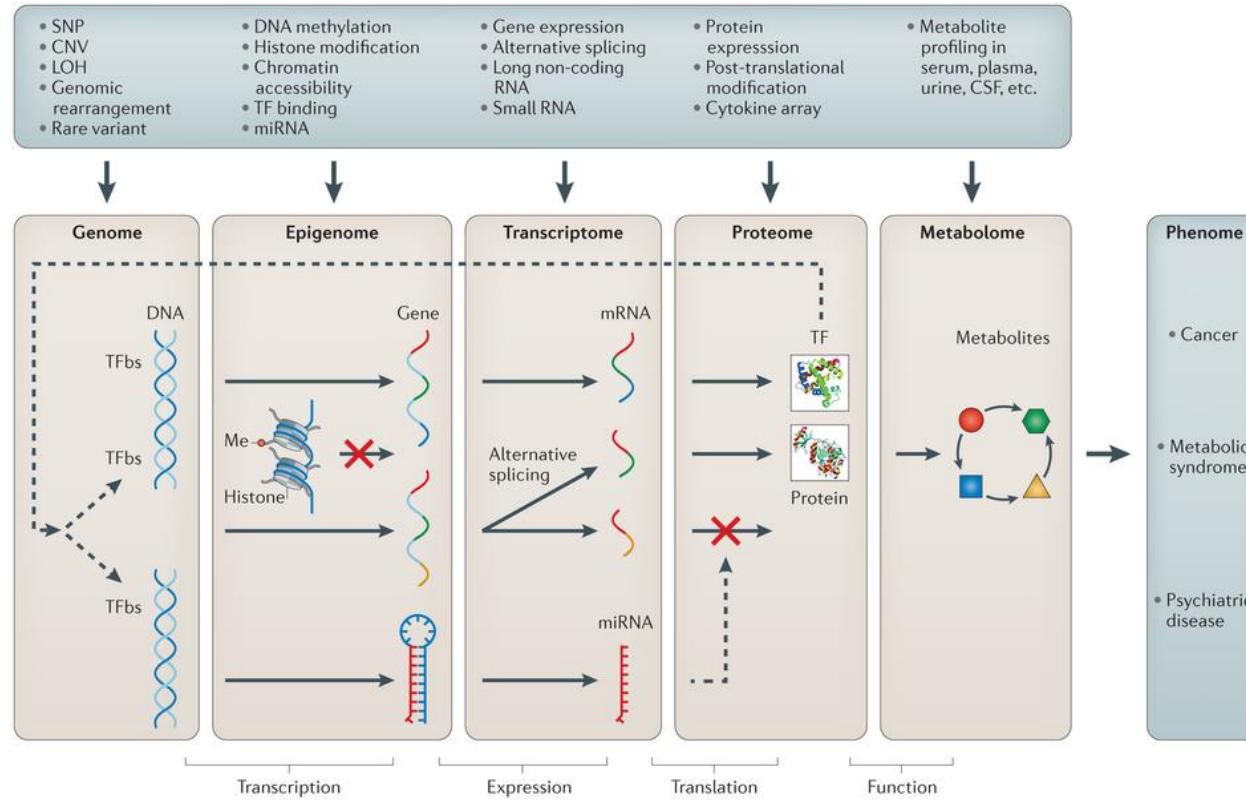
- Computational Systems Biology for complex human disease
- Disease maps construction and analysis
- Large scale Boolean modeling of signaling networks
- Automated inference of executable disease models
- Curation and Annotation of Logical Models in biology

Central dogma of molecular biology



(Figures adapted from Franklin and Vondriska, 2011) and <https://www.bioregulatory-systems-medicine.com/en/brsm-model/autoregulation-of-biological-networks>

Organisms: Complex systems



Goal of systems biology: put bits and pieces together

- How do the individual parts interact to yield system behavior?
- Biology has focused on figuring out the pieces.
- But what happens when you fit them together?

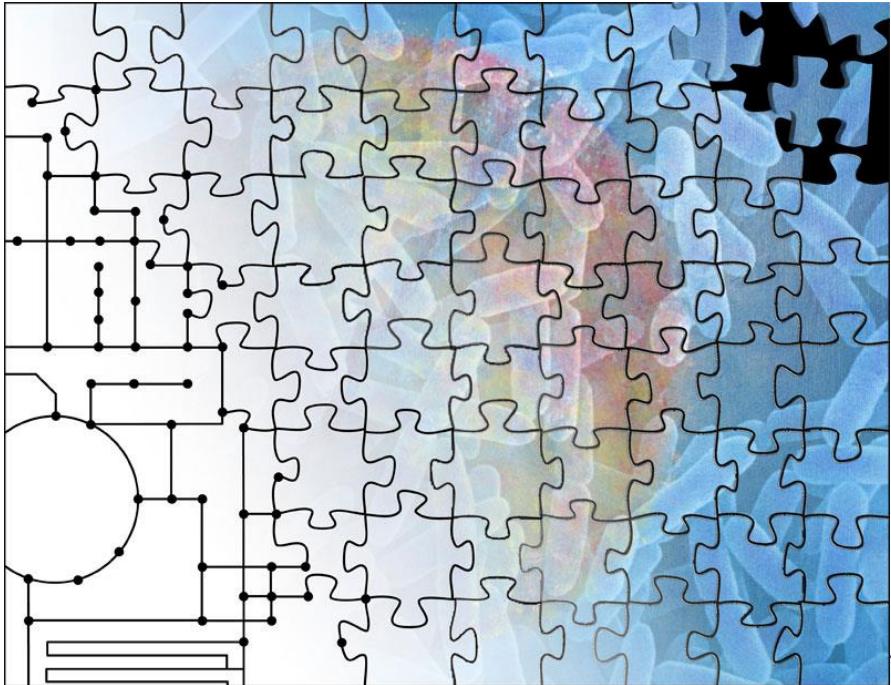
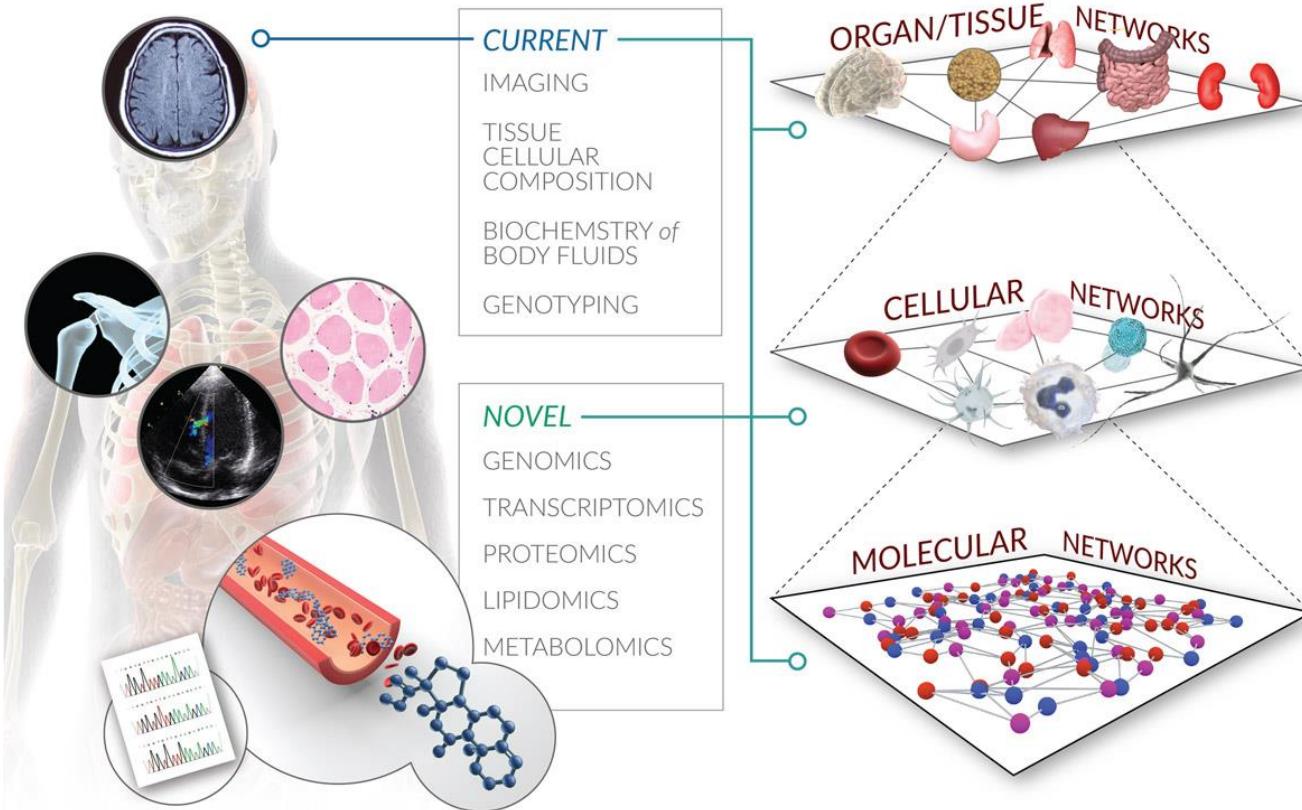


Image Creator: Joerg Buescher

ECCB 2020

Virtual!

DIAGNOSTIC APPROACHES



<https://www.bioregulatory-systems-medicine.com/en/brsm-model/autoregulation-of-biological-networks>

Full scale integration

- Key goal of systems biology: construct networks at different cellular levels to investigate cellular machinery.
- Currently no satisfactory method to construct an integrated cellular network.

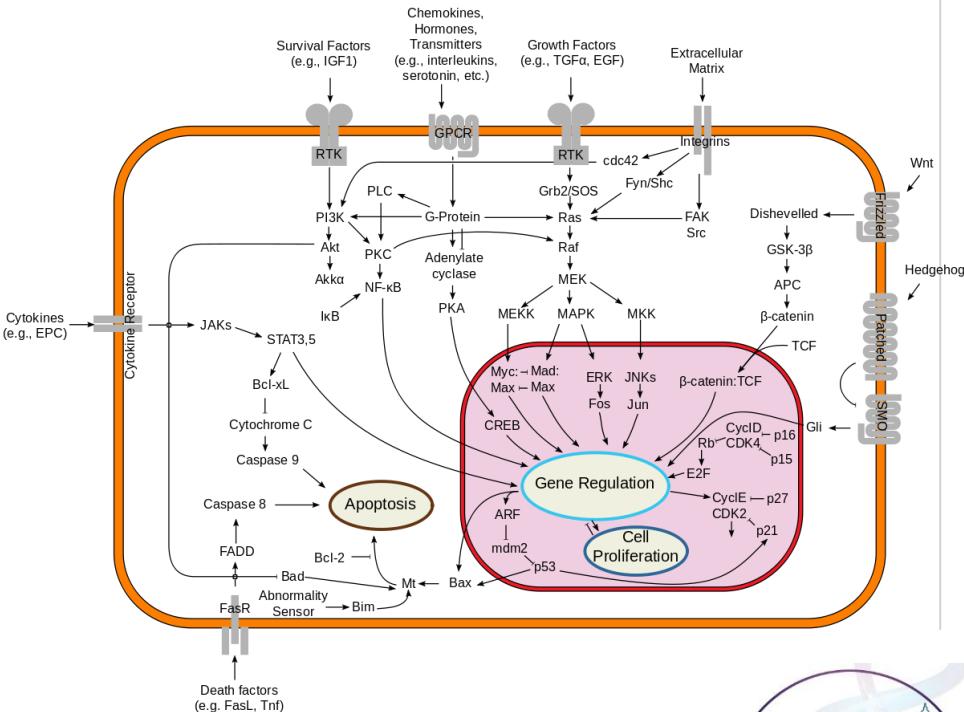
The challenge

- Interpret largescale data sets and extract true information to understand biological systems.
- Computational techniques, which can integrate and combine these large and heterogeneous data sets, will help gain more biological insights.



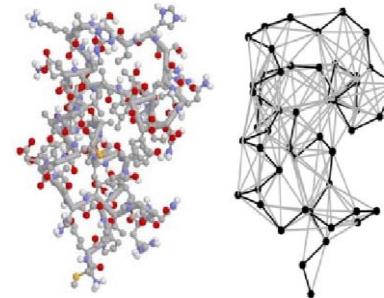
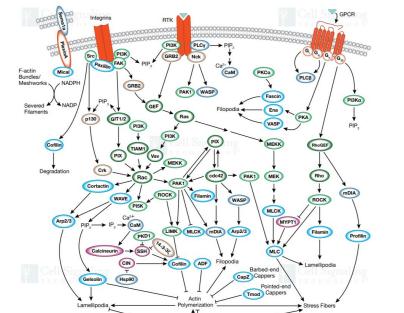
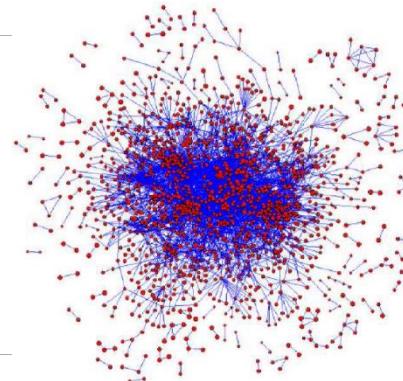
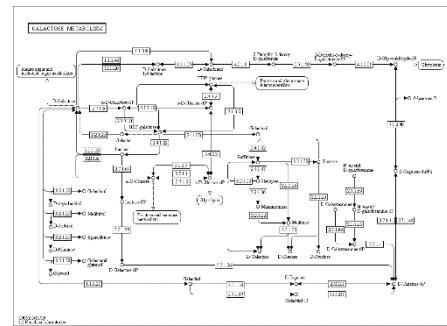
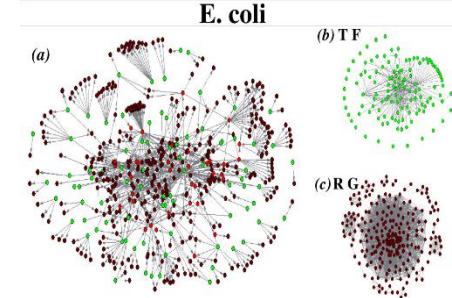
Viewing cells in terms of their underlying network

- Powerful concept, networks can be seen as graphs.
- Organizing biological information in the context of networks is fundamental to applying systems-level approaches to understanding biological function.



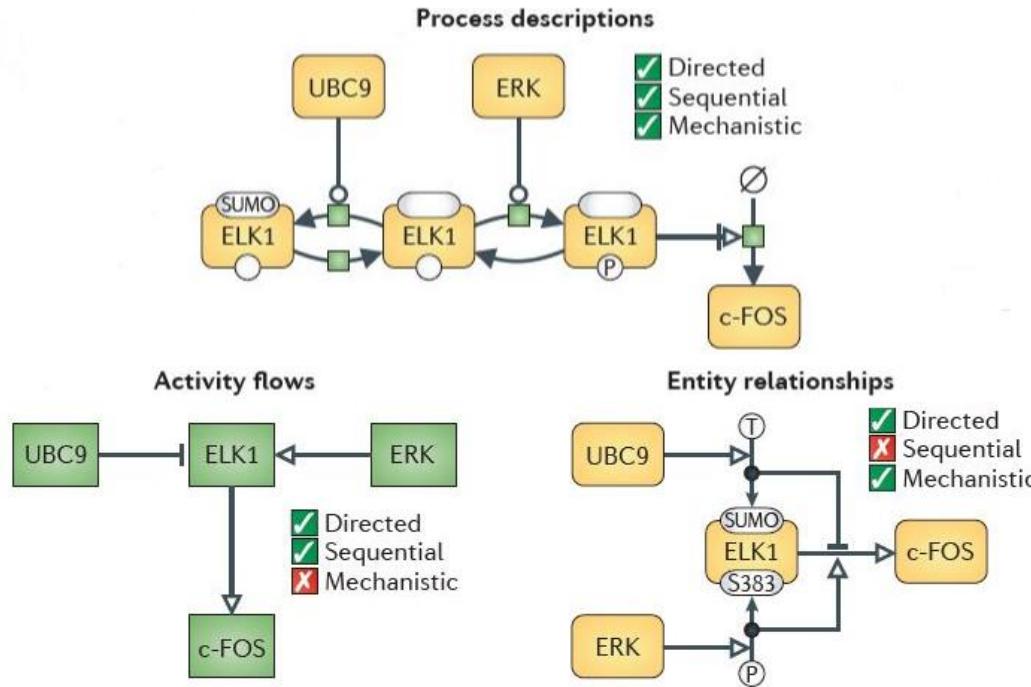
Intra-cellular networks

- Transcriptional regulation networks
- Protein structure networks
- Metabolic networks
- Protein-protein interaction (PPI) networks
- Cell signaling networks



Systems Biology Graphical Notation (SBGN) standard languages

Describing mechanisms in a systematic fashion

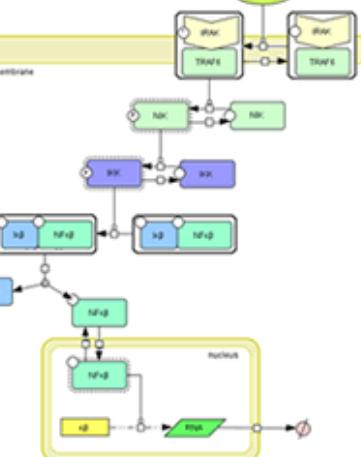
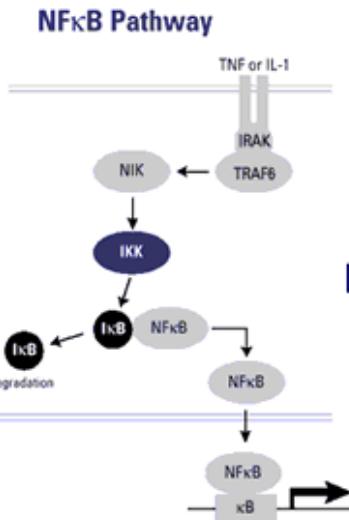
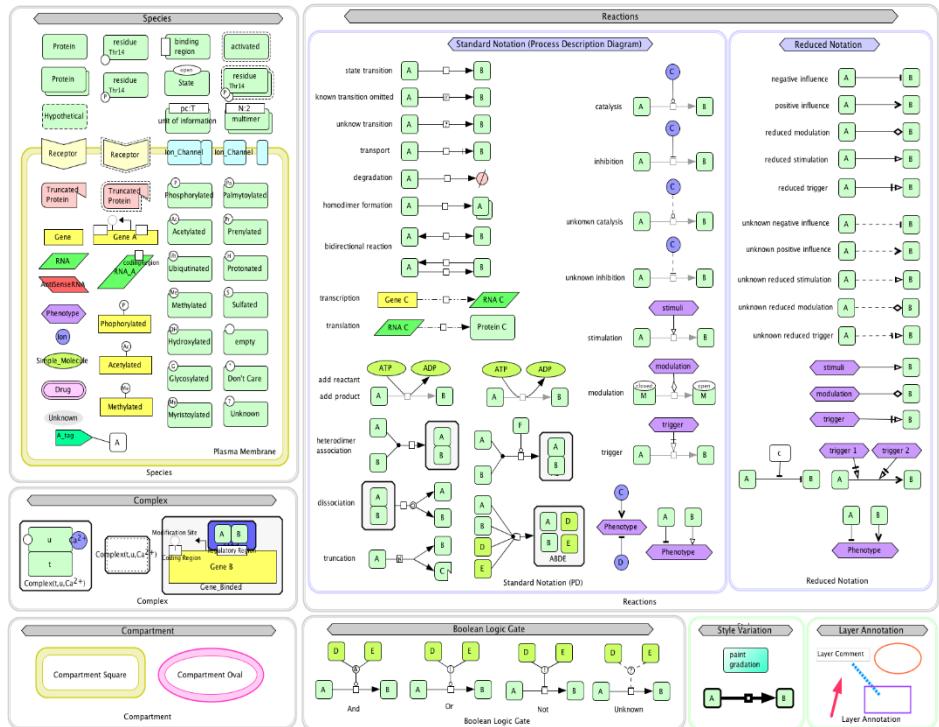


(Le Novere, 2015, Mazein et al., 2018, Vogt et al., 2013)

Process description diagrams of biological mechanisms

- Pioneering works of Prof. Hiroaki Kitano.
- Process description representation of signalling networks.
- First comprehensive disease-relevant extensive reconstructions of signalling pathways.
- Cancer Signalling Atlas – Curie Institute -
<https://acsn.curie.fr/ACSN2/ACSN2.html>
- Disease Maps project <https://disease-maps.org/>
- REACTOME – PD like diagrams <https://reactome.org/>

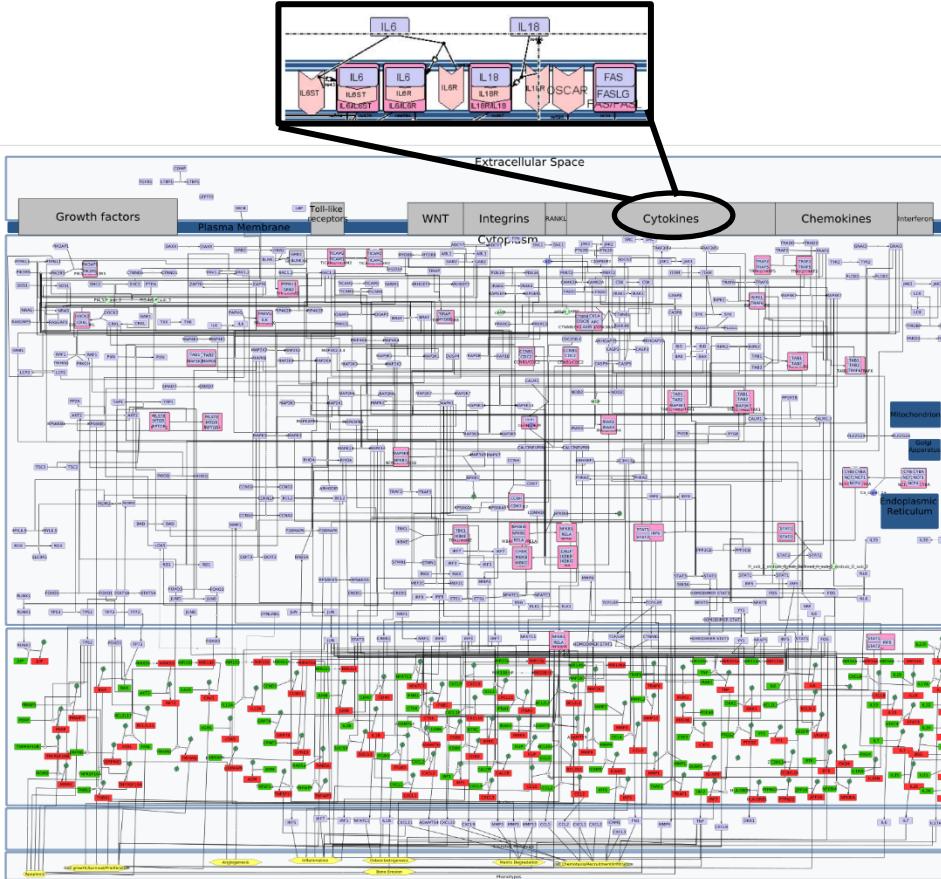
Building molecular maps using CellDesigner



conventional diagram

CellDesigner's diagram

The RA map



Extracellular space

Plasma membrane

Cytoplasm

Nucleus

Secreted Components

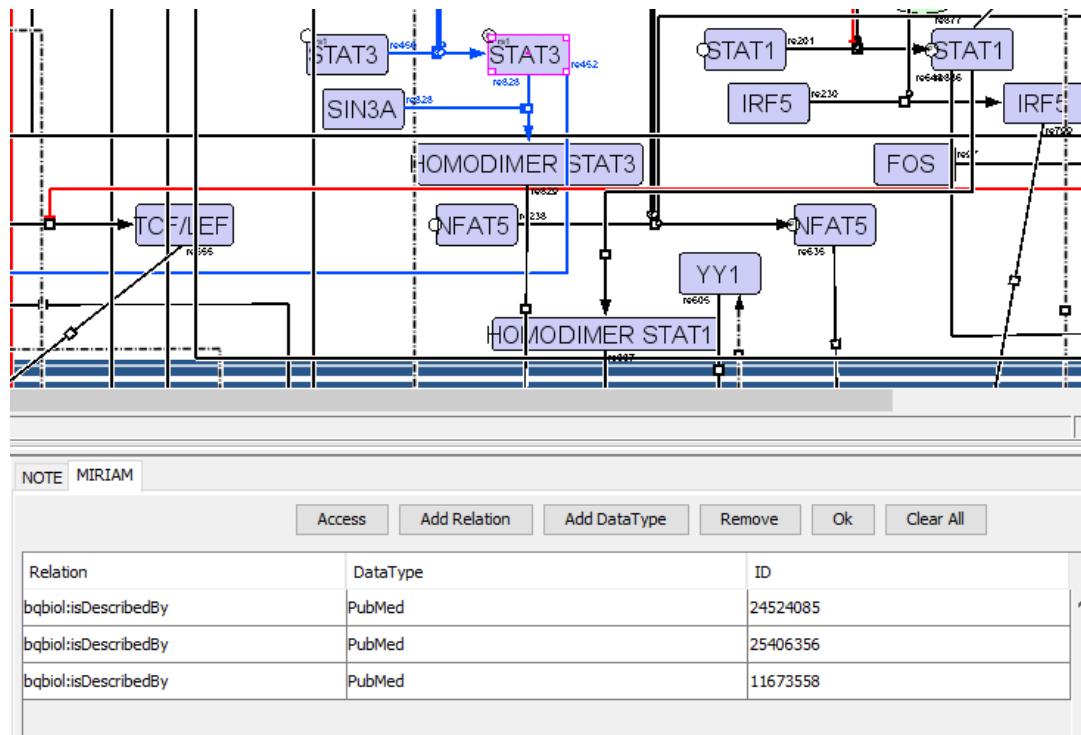
Phenotypes

- SBGN compliant
- 506 species (303 proteins, 61 molecular complexes, 106 genes, 106 RNA), 449 reactions
- 353 scientific publications used
- Expert validated
- Detailed annotations



Building molecular maps using CellDesigner

MIRIAM: Minimal Information Required In the Annotation of Models



- Facilitating interoperability and model reusability
 - Annotations retrieved in the resulting model

Interactive knowledge base for RA using MINERVA –(Molecular Interaction NEtwoRk VisuAlization) platform



SEARCH OVERLAYS INFO

GENERIC DRUG CHEMICAL MIRNA

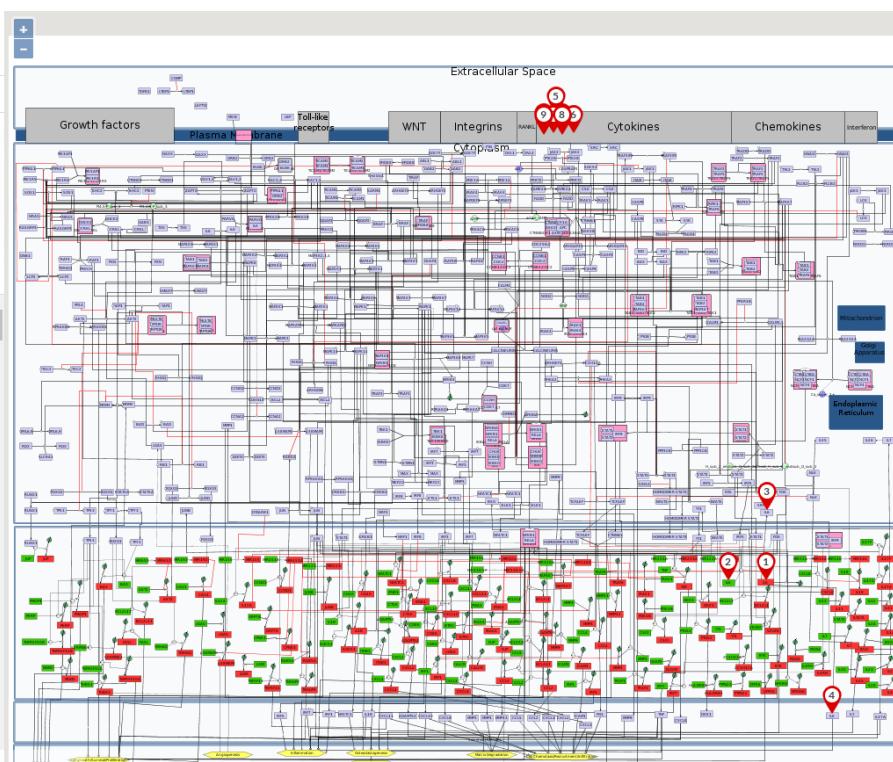
SEARCH: IL6

il6

RNA: IL6

Compartment: Nucleus
Full name: interleukin 6
Symbol: IL6
Former symbols: IFNB2
Synonyms: BSF2, HGF, HSF, IL-6
Annotations:
Source: Annotated by curator

[1] PUBMED(10688908)
[2] PUBMED(17652167)
[3] PUBMED(2462501)
[4] PUBMED(18281366)
[5] PUBMED(18454843)
[6] PUBMED(24524083)
[7] PUBMED(22870451)
[8] PUBMED(8484679)
[9] PUBMED(12905466)
[10] PUBMED(28494214)
[11] PUBMED(18205922)



Protein: IL6

Interacting drugs

Interacting chemicals

Interacting Micro RNAs

AM1 FN1

AM2 CAM1 ITGAV TGFB1 TGFAL

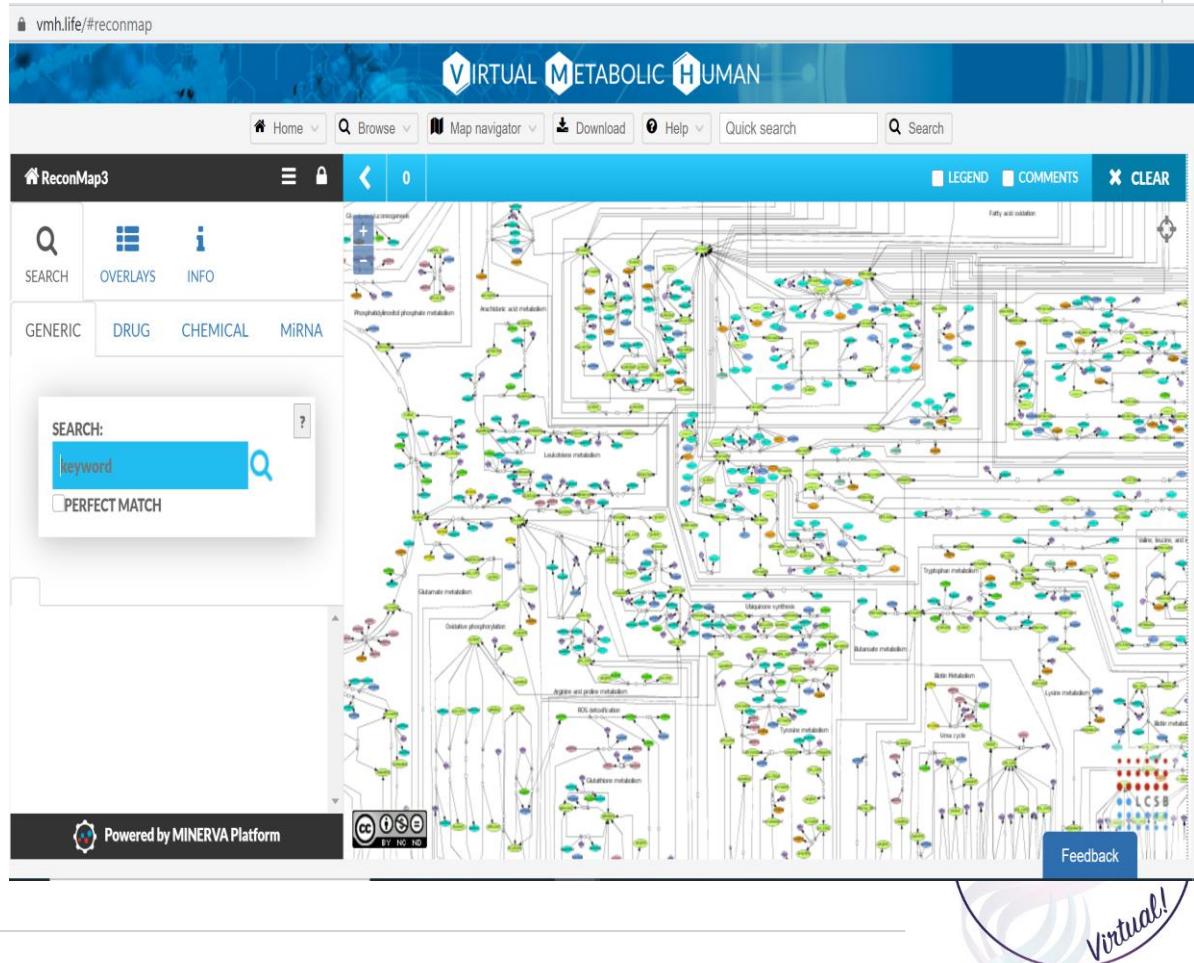
FYN DVL1 JAK2 PTK2B SRC TRAF3IP2

RANK RANKL L6ST L6R IL18R OSCAR FAS IL17A IL18 IL17



The ReconMap

- ReconMap content obtained from the Virtual Metabolic Human database (VMH, <http://vmh.uni.lu>).
- Recon-derived simulation results can be visualized on ReconMap using a new extension to the COBRA Toolbox (Schellenberger et al., 2011).
- User can perform a simulation, e.g. Flux Balance Analysis, using the COBRA toolbox function 'optimizeCBmodel', then call the function 'buildFluxDistLayout' to write the input file for a context-specific ReconMap Overlay.



Prerequisites for high-quality biochemical interaction maps:

(Inspired by Systems Biology, ed. Nielsen and Hofmann, Chapter 8, Wiley -VCH, 2016 and Community-driven roadmap for integrated disease maps, Ostaszewski et al., 2018)

- **Accurate** – correctly represents our empirical knowledge.
 - **Reusable** – well annotated and referenced.
 - **Comprehensive** – accounts for all known reactions within the selected scope.
 - **Machine readable** – can be processed and analyzed using computers.
-
- **Executable** – corresponds to a computational model that can be simulated.
 - **Functional** – can explain the known system-level behavior of the biological network.

“Why build models?”

Jay Bailey

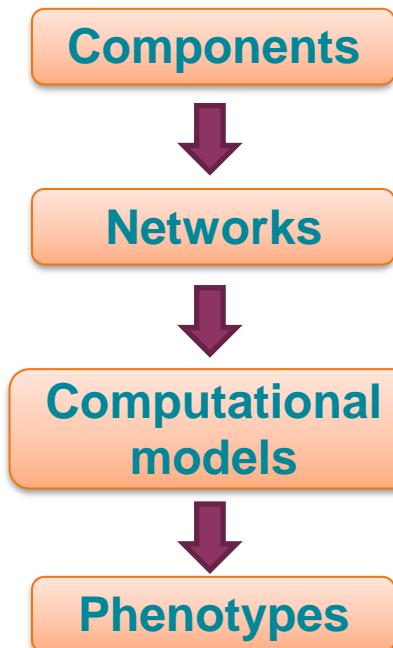
- To organize disparate information into a coherent whole.
- To think (and calculate) logically about what components and interactions are important in a complex system.
- To discover new strategies.
- To make important corrections to the conventional wisdom.
- To understand the essential qualitative features.



Computational models

- Models are used to represent actual **quantitative/qualitative** relations between the molecules in the system.
- **Abstract representations** of biological processes
- Have an inherent **execution scheme** attached to the model.

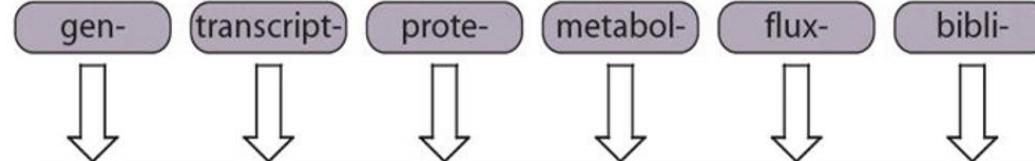
Central dogma of computational systems biology



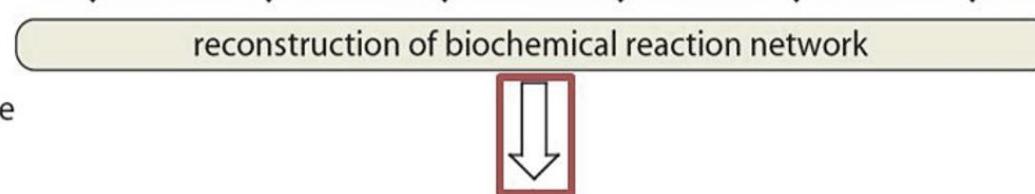
The Systems Biology paradigm:

Components -> networks -> computational models -> phenotypes

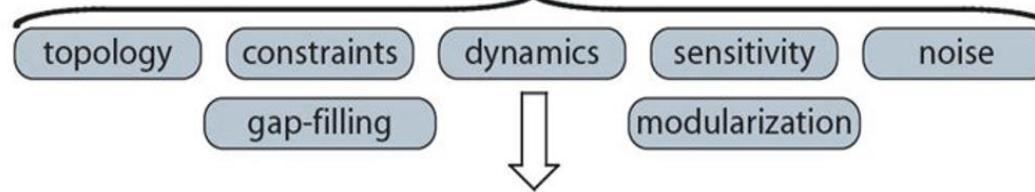
1. Database:
- Plurality of -omics



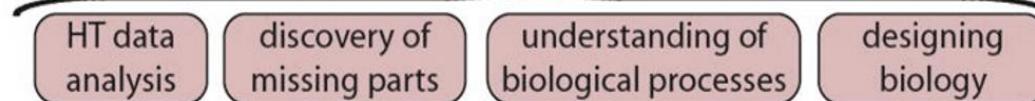
2. Knowledge Base:
- One set of reactions
encoded by a genome



3. *In silico* modeling:
- Query Tools

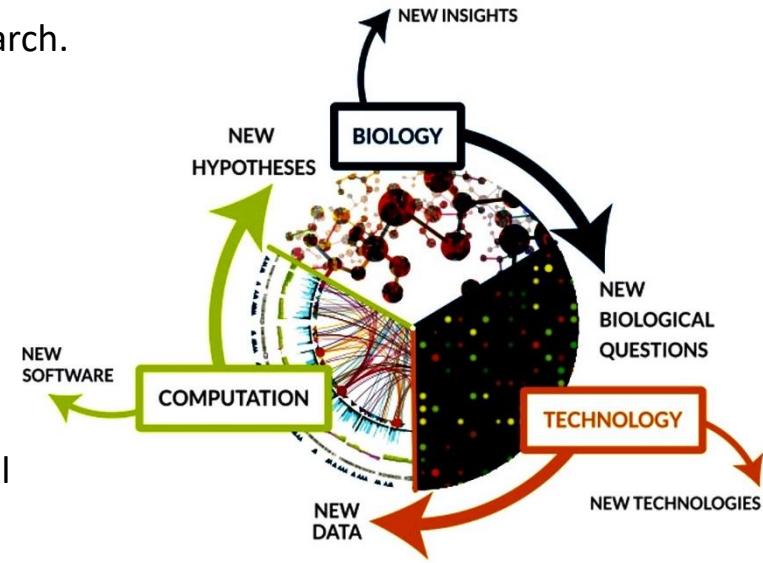


4. Validation, Discovery,
and Use



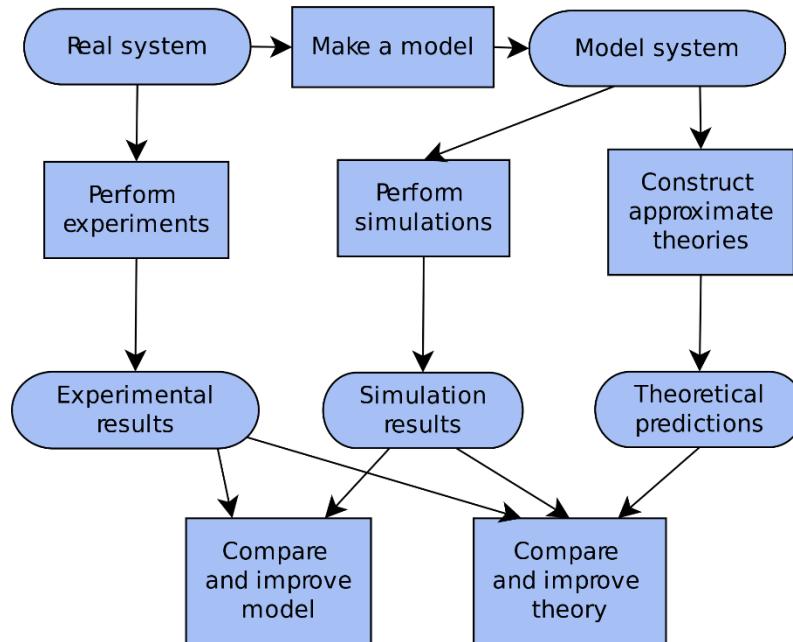
Model building: a step by step process

- Generation of the structural model based on a literature search.
- Compilation of a calibration dataset using experimentally validated biological knowledge.
- Model fitting to the experimental data.
- *In silico* simulations to generate predictions.
- Validation of predictions through experimental testing.
- Model refinement by feeding newly generated experimental data back to the model.
- Generation of novel hypotheses.
- Reiteration.



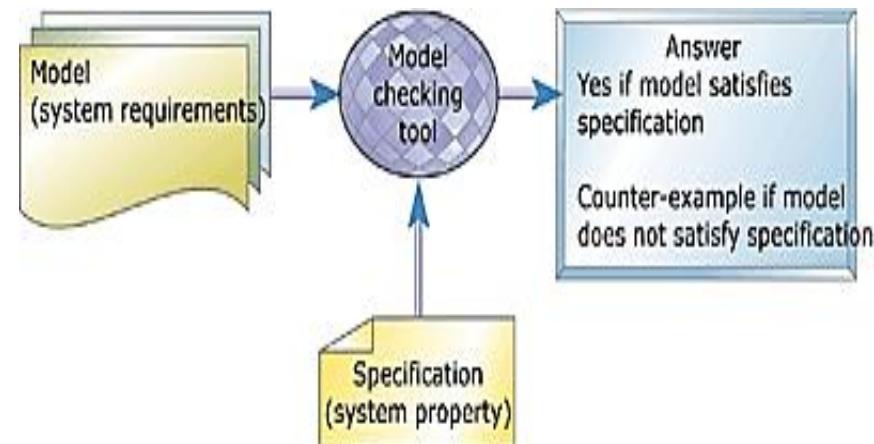
***“In theory, there is no difference between theory and practice.
But in practice, there is.”***

Manfred Eigen, 1967 Nobel Laureate in Chemistry



Model checking techniques - Comparing mechanistic models to specifications

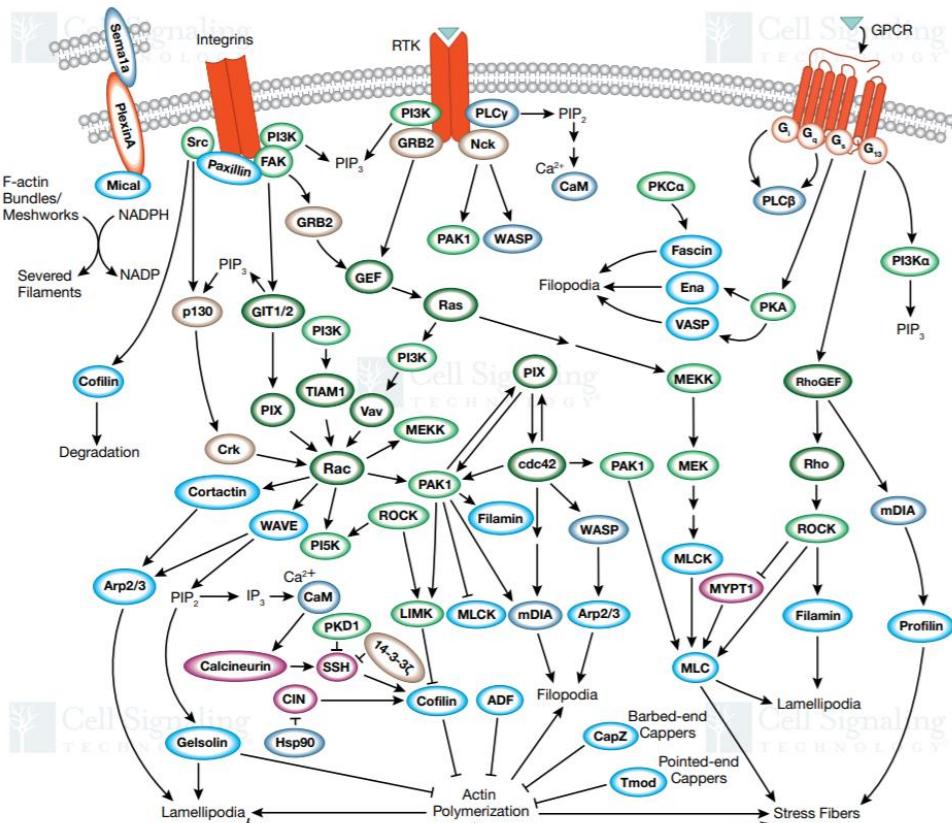
- A formal model of the system under study is constructed.
- Experimental evidence is formalized as specifications (observations).
- Model checking is used to ensure that the model reproduces the experimental observations.
- **Mismatch with experimental observations:** model should be refined by additional information.
- **Match with experimental observations:** could lead to further querying and testing of the model to suggest further experimental studies.



(Figure from <https://www.embedded.com/an-introduction-to-model-checking/>)

Reachability analysis

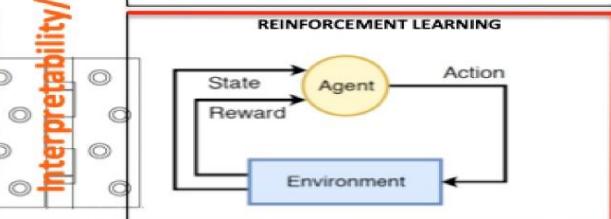
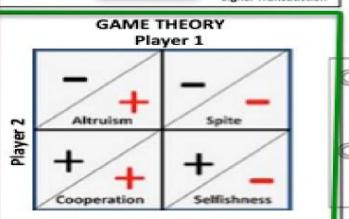
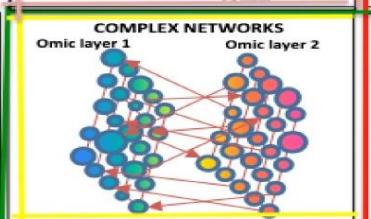
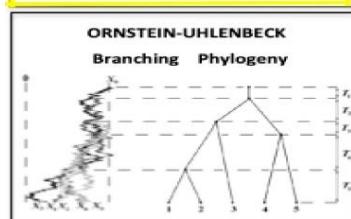
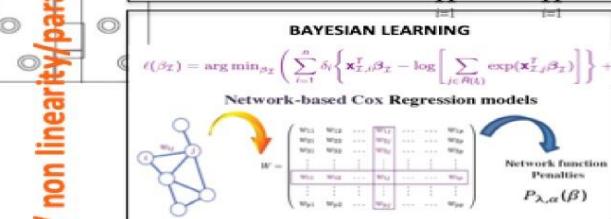
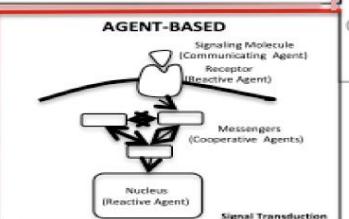
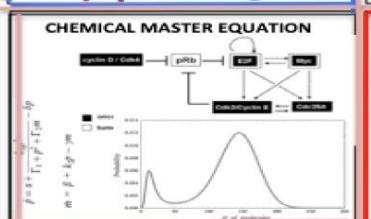
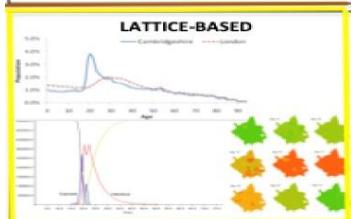
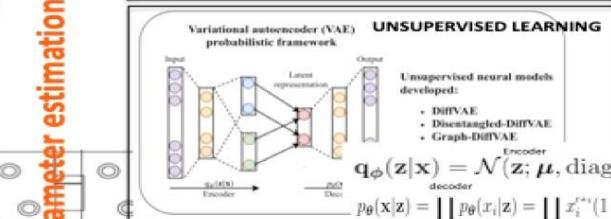
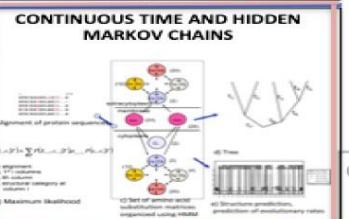
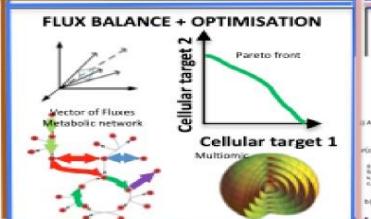
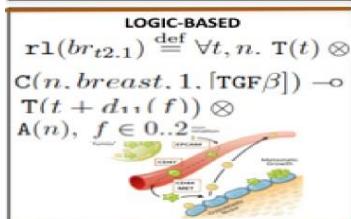
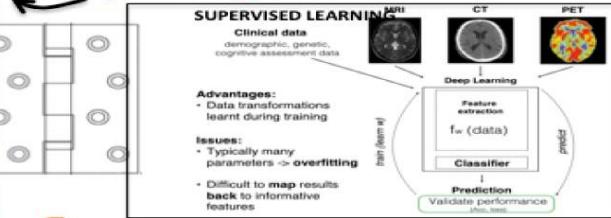
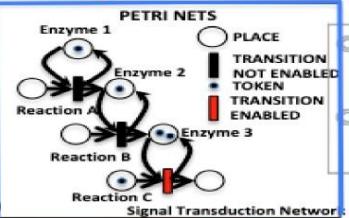
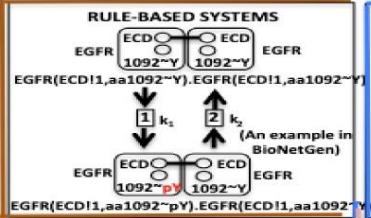
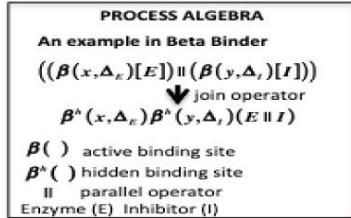
- Extra-cellular molecules trigger a response inside the cell by initiating a signal at special membrane receptors.
- Signal is transmitted to reporters/targets through various chains of interactions among proteins.
- Understanding whether such a signal **can reach from membrane receptors to reporters** is essential in studying the cell response to extra-cellular events.



COMPUTATIONAL (BIO) MODELING METHODOLOGIES → LEARNING METHODOLOGIES

Hypothesis-driven

(Mostly) Data-driven



Interpretability/ non linearity/parameter estimation

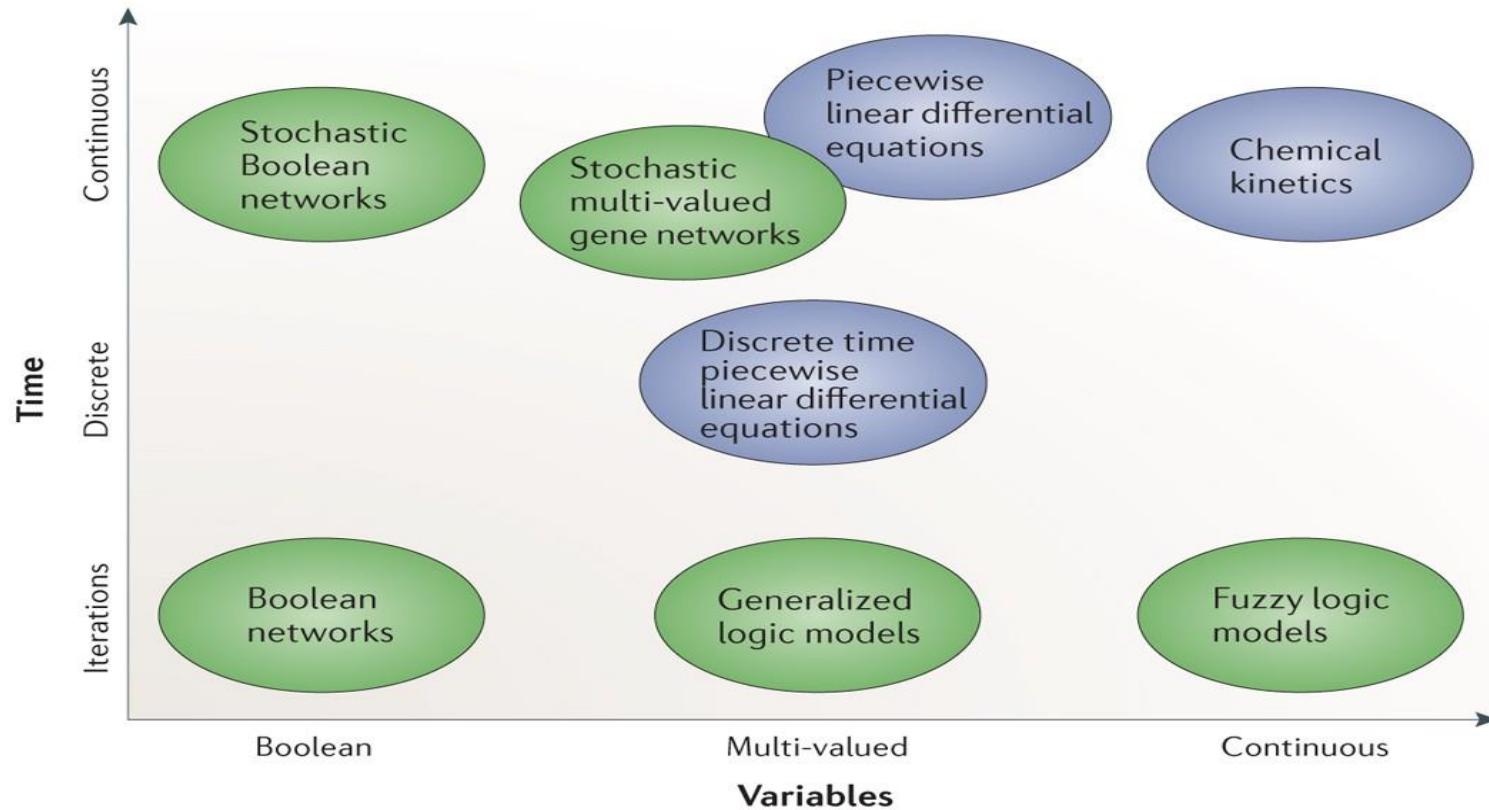
Quantitative versus qualitative models

Table 1

Comparison matrix of quantitative and qualitative models

	Quantitative model	Logic model
Suitable for	Time series	Phenotypes
Time representation	Linear representation	Abstract iterations
Variables	Quantitative	Qualitative
Mechanism representation	Yes	No
What can we do?	Compute concentrations and durations; evaluate the effect of parameter values	Compute state transitions and attractors (steady-states and cyclic attractors)
Data necessary to build the model	Molecular species, genes, interactions, biochemical processes	Activities, defined phenotypes, rules linking those
Data to parameterize and validate the model	Amount of molecular species, timecourses, quantitative phenotype	Perturbations of activities such as RNA interference, inhibitors, qualitative phenotypes
Advantages	Quantitative, precise; direct comparison with quantitative measurements; large existing toolkit	Easy to build; easy to compose; easy simulation of perturbations
Weaknesses	Requires quantitative knowledge of initial conditions and kinetics	Cannot provide quantitative predictions; difficult to choose between alternative behaviours

Granularity of time representation and variable values for various modelling approaches.



Many different modelling approaches – one common language?

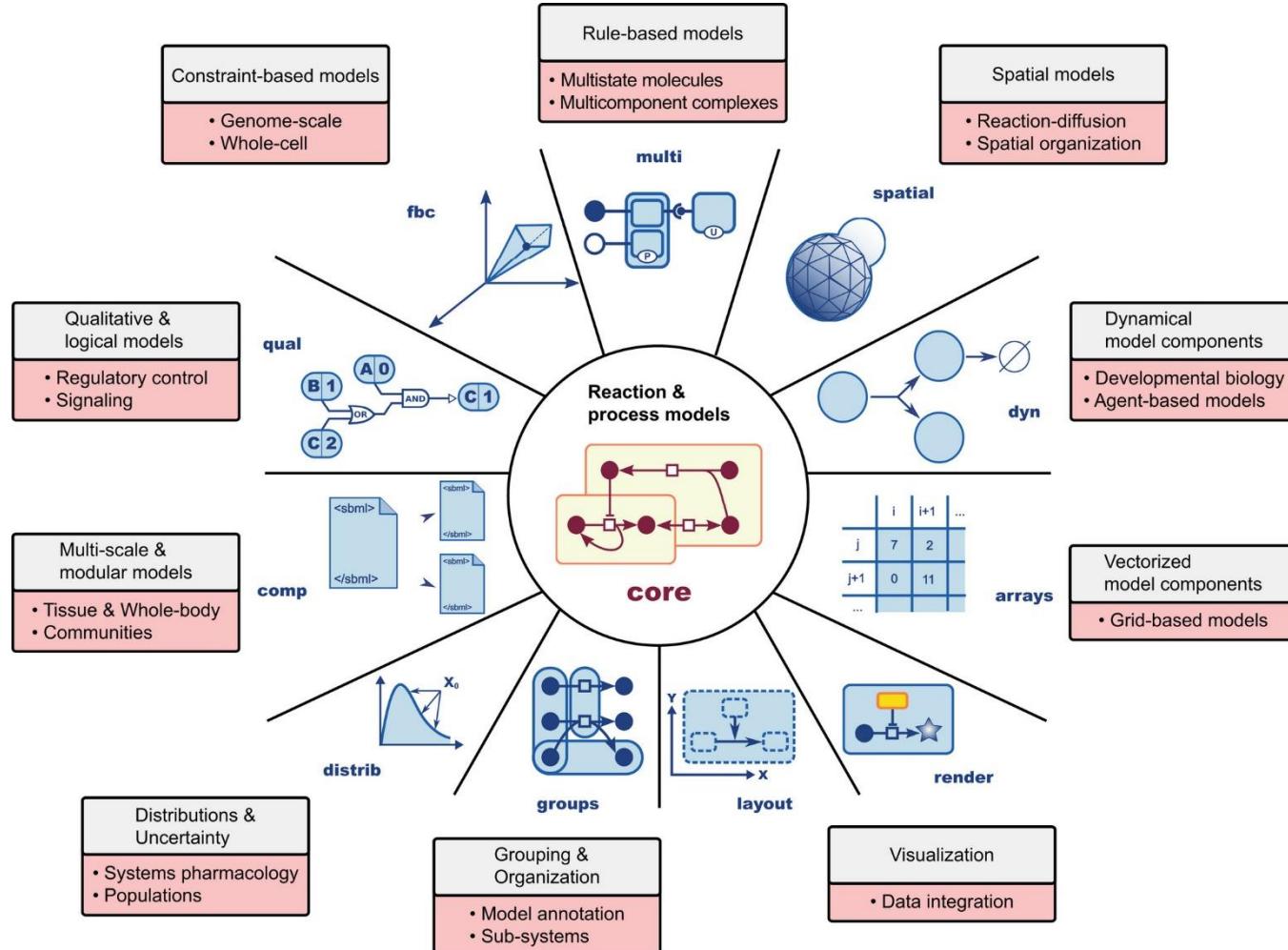


- A format to encode mathematical models that is used in systems biology.
- Initially focused on non-spatial, reaction-based biochemical models.
- Packages covering different modelling approaches (qual).
- Supported by software libraries in different programming languages .
- Can be imported or exported by a variety of modelling and simulation tools.
- Does not store experimental data, or simulation descriptions.
- Based on xml format, intended to be machine and not human-readable.



020

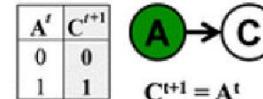
Virtual!



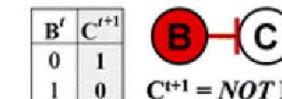
Logic models

- Form of mathematical model governed by logic operators (AND, OR, NOT).
- Parameter free.
- Suitable for modeling gene regulatory networks.
- In silico* simulations, qualitative predictions.
- Each node in a logic model has a corresponding logic function that controls its regulation each time the model is updated.
- Two updating schemes: synchronous and asynchronous.

A Logic functions with one molecular regulator



$$C^{t+1} = A^t$$



$$C^{t+1} = \text{NOT } B^t$$

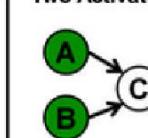
Truth table

Truth table

Logic functions with two molecular regulators

Non-specific Interaction Network

B Two Activators



AND C is only ON in one condition

A^t	B^t	C^{t+1}
0	0	0
1	0	0
0	1	0
1	1	1

The presence of A and the presence of B activates C

OR

C is only OFF in one condition

A^t	B^t	C^{t+1}
0	0	0
1	0	1
0	1	1
1	1	1

Either the presence of A or the presence of B activates C.

C One Activator and One Inhibitor



$$C^{t+1} = A^t \text{ AND NOT } B^t$$

A^t	B^t	C^{t+1}
0	0	0
1	0	1
0	1	0
1	1	0

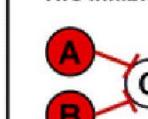
The presence of A and the absence of B activates C.

$$C^{t+1} = A^t \text{ OR NOT } B^t$$

A^t	B^t	C^{t+1}
0	0	1
1	0	1
0	1	0
1	1	1

Either the presence of A or the absence of B activates C.

D Two Inhibitors



$$C^{t+1} = \text{NOT } A^t \text{ AND NOT } B^t$$

A^t	B^t	C^{t+1}
0	0	1
1	0	0
0	1	0
1	1	0

The absence of A and the absence of B activates C.

$$C^{t+1} = \text{NOT } A^t \text{ OR NOT } B^t$$

A^t	B^t	C^{t+1}
0	0	1
1	0	1
0	1	1
1	1	0

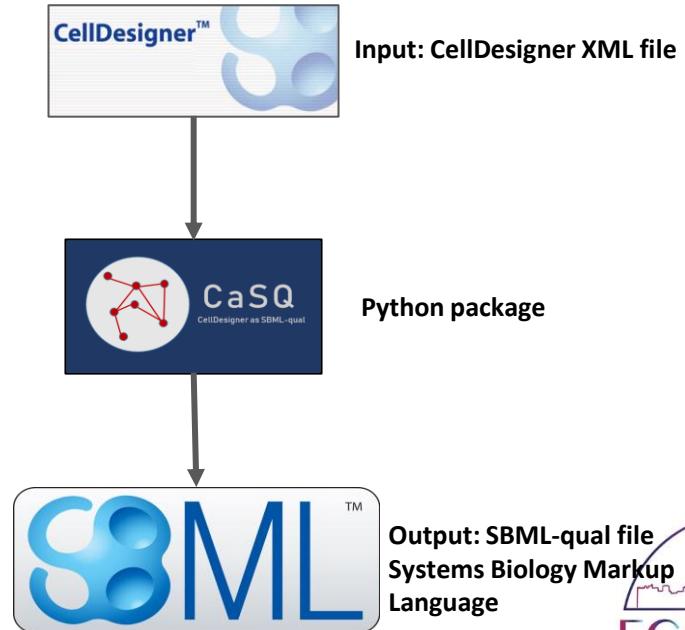
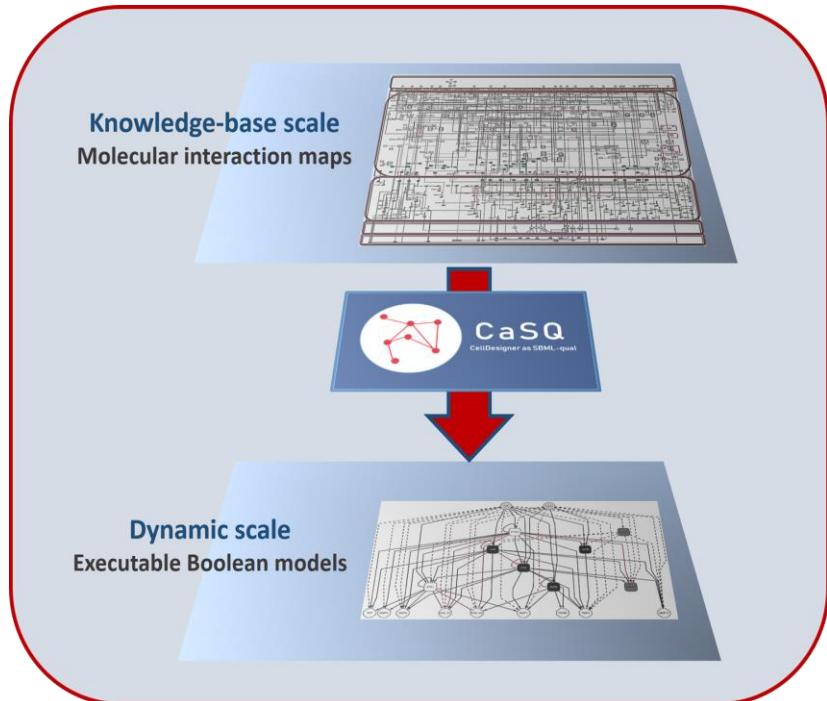
Either the absence of A or the absence of B activates C.

Attractors' search: an important aspect of dynamical analysis

- Dynamical analysis can reveal **attractors** (complex or simple) that could correspond to **fixed points** (steady states) or **oscillatory behaviours** (periodic states).
- Attractors represent a stable behaviour of a system, as reflected by a fixed trajectory in the space of all possible states of the system.

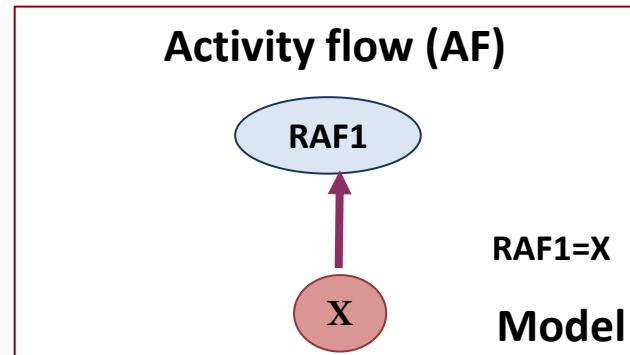
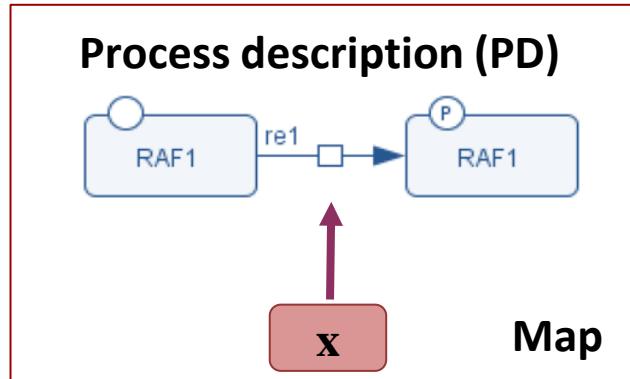


Executable Boolean models from molecular interaction maps



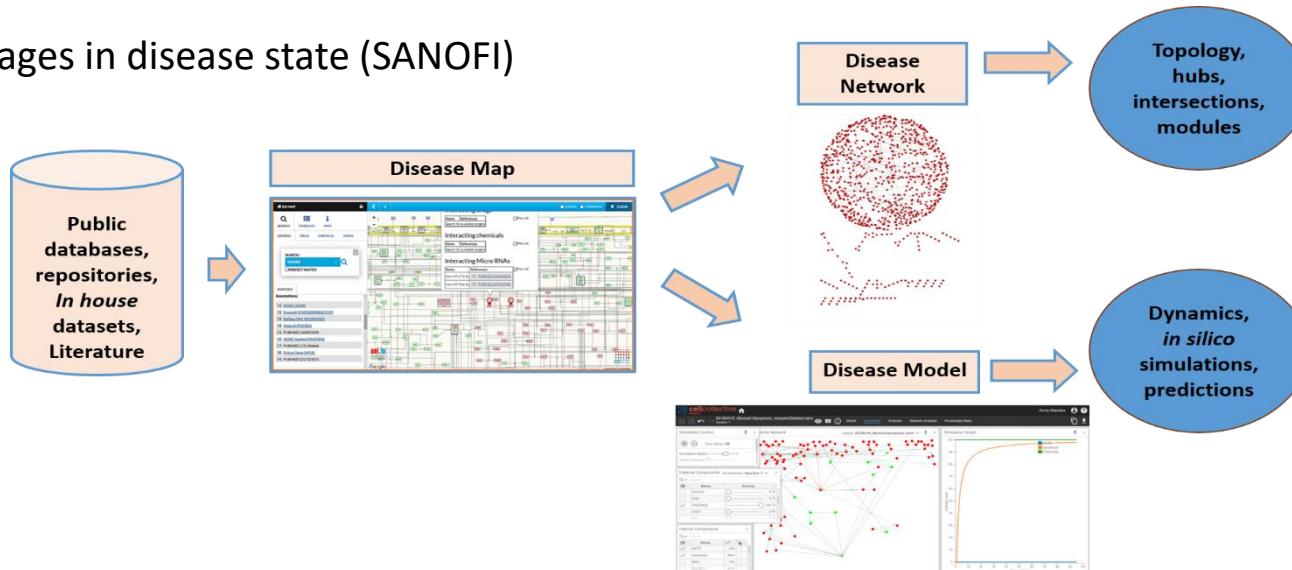
CaSQ (CellDesigner as SBML-qual)

- CaSQ: a tool that bridges static and dynamic representations of biological networks.
- **Logical rules inference based on topology and semantics** already encoded in the map.
- **Use of standards (SBGN and SBML)** facilitates interoperability and minimizes assumptions and interpretations.
- Models retain **all annotations and references** of the original map, so they can be reusable by others.
- **SBML qual: standard format**, model can be analyzed with all tools that can import the format (**CoLoMoTo notebook**).



Map to Model projects that use the CaSQ framework

- RA FLS - UPSaclay/ UEVE/Genopole
- ImmunAID - Immunome project consortium for AutoInflammatory Disorders Cordis Europa (SAID)
- Covid-19 Disease Map (Disease Map consortium)
- Macrophages in disease state (SANOFI)



Highlights

- Capture knowledge with minimal assumptions and interpretations (SBGN language).
- Support documentation and references of the integrated data sources.
- Provide commonly recognized unique identifiers for components for further annotation (MIRIAM annotations).
- Translation from static to dynamic must be scalable to support comprehensive network reconstruction, without compromising the accuracy or mechanistic resolution of the network.
- Strict syntax to make the language machine readable, and precise semantics that define how the network corresponds to an executable model.
- Use of standard formats and languages (SBML) to facilitate interoperability and reusability.

Acknowledgements:



Sylvain Soliman, PhD
INRIA Saclay-Île-de-France
CR



Vidisha Singh, MSc
University of Evry, Paris-Saclay
PhD Student



Sara Sadat Aghamiri, Msc
University of Evry, Paris-Saclay
INSERM Engineer



Yacine Hadjar,
SANOFI, Paris
M2 Student



Tomas Helikar, PhD
University of Nebraska-Lincoln
Assistant professor



Denis Thieffry, PhD
Computational Systems Biology Group
IBENS, Paris
Professor



Aurelien Naldi, PhD
INRIA Saclay-Île-de-France
Researcher



Laurence Calzone, PhD
Institut Curie, Paris
Research Engineer



Georges D. Kalliolias, MD, PhD
Hospital for Special Surgery, New York City
Weill Cornell Medical College, New York City
Assistant Professor, Assistant Scientist



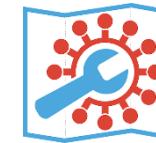
Vassili Soumelis, MD, PhD
Inserm U976 "Human Immunology,
Physiopathology and Immunotherapy"
Saint Louis Hospital, Paris
Professor - Hospital Practitioner



Zsolt Bocskei, PhD
SANOFI, Paris
Group leader



Marek Ostaszewski, PhD
University of Luxembourg, Centre for
Systems Biomedicine, Luxembourg
Researcher





THANK YOU
FOR
YOUR
ATTENTION
ANY QUESTIONS?

