Automated pipeline for the inference of Boolean models from molecular interaction maps.

Anna Niarakis
MCF, UEVE- Paris-Saclay
anna.niaraki@univ-evry.fr
Describing mechanisms in an systematic fashion

- Molecular Interaction Maps:
  - High quality source of knowledge – template for data visualization (i.e. control versus treatment).
  - Can be seen and analyzed as a complex network (topology/structure).
  - Can serve as a scaffold for a mathematical model.
Molecular Interaction Maps:

- Increasing popularity

- Many systematic efforts (DiseaseMap Project, The Curie Atlas of Cancer Signalling Network, Cancer Cell Map Initiative etc.)

- Interdisciplinary teams (biologists, curators, clinicians, bioinformaticians etc)

- Standardization (SBGN, mEPN)
From static to dynamic:

Map (static)
- Template for data visualisation
- Focus on mechanistic details of interactions
- System biology graphical notation (SBGN)

Model (dynamic)
- System to simulate information flow
- Focus on flow of information and regulations
- Systems Biology Markup Language (SBML)

Objective

Focus

Format

Both processes are usually done separately

Shared information such as:
- Mode of influences
- Logic operations
- Network topology

Automated inference
The Molecular Interaction Map needs to be:

• **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.

(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)
CONSTRUCTION OF THE RA DISEASE MAP
Rheumatoid Arthritis (RA):

- Multifaceted autoimmune disease that causes chronic inflammation of the joints.
- **Etiology** of the disease remains unclear.
- Can also cause inflammation and injury in other organs in the body therefore considered as a **systemic disease**.

(Figure from McInnes & Schett, 2011)
Rheumatoid Arthritis (RA):

- RA greatly affects the synovial joints in the body:
- The immune system mistakenly attacks the synovial lining surrounding the joints leading to an inflammatory response.
- This response thickens the synovium by laying down fibroblasts and causes destruction of the cartilage and bone.
- The result of this process is severe deformation.

(Figure from Hawtree S, et al, 2013)
Molecular Interaction map of RA (Wu et al.2010):

- **28 studies high throughput**, including drug treatment experiments (2003-2009)
- **Large heterogeneity** of source studies (PBMC, SF, PMN, cartilage)
- **False positives** possible (due to RNA expression data)
- Lack of **experts validation**
- **Not very sophisticated layout**
- **Connectivity problems** (several nodes with very low degree)
- **Very basic annotation of the** CellDesigner file
- **No standards** (SBGN, MIRIAM annotations)
Current state of the RA map:

- **80 new mediators**, derived from literature published after 2010, using public databases and exhaustive manual curation (at least 2 articles for each molecule added, small scale experiments).

- All interactions and mediators are being reassessed using strict curation criteria (**20 mediators removed**).

- Detailed annotations including **PubMed identifiers**, **HUGO names** and **Cell/fluid types**, fully detailed MIRIAM annotation section.

- **Fully SBGN compliant**.

- **Quality control** of the integrated information and its representation is carried out by a collective effort of our collaborators (**clinicians, experts in RA and inflammation**) (restructure of the initial map), use of Ingenuity pathway analysis (IPA).
Updated RA map (CellDesigner file):

- ECM
- PLASMA
- MEMBRANE
- CYTOPLASM
- NUCLEUS
- SECRETED MOLECULES
- PHENOTYPES

>400 components
>150 scientific papers
Fully SBGN
Experts’ validation
Detailed annotations
Strict curation criteria
FROM A DISEASE NETWORK TO A BOOLEAN MODEL
Biological questions of more dynamic nature concerning RASFs

- Can we **induce apoptosis**? (either by forcing apoptosis pathway or by blocking cell survival pathways)

- Can we **block structural damage** by blocking intermediate components?

- Are they subjected to **negative feedback** control like macrophages?

- Do they **differentiate** depending on the initial stimuli?
Capture dynamic properties: Boolean model

- Simplest form of mathematical model governed by logic operators (AND, OR, NOT).
- Parameter free.
- Scalable, can range from 3 components up to more than 200 components.
- Suitable for modeling large signaling networks.
- *In silico* simulations, qualitative predictions

```
Boolean rules:
A* = NOT C
B* = A AND C
C* = NOT A
```
Different network representations:

(Figure from Le Novère, 2015)
How does it work exactly?

• **Process description** compressed to resemble more to an **activity flow** diagram.

• **Unregulated components** = **inputs**.

• **Logical operators** assigned based on the network’s topology and the semantics already present in the map (activation/inhibition, complex formation etc).

• **Annotated references** and **layout** retained in the output file.
Executable Disease Map

**Cell Collective modelling platform**

**CellDesigner**

**CaSQ**

**Disease Network:**
- Global
- Heterogeneous
- Too big and complex to tune and execute

*View of the boolean network generated from the Disease Map, in Cell Collective*
Construction of RASF specific model:

• Focusing on 4 functional outcome of RASFs.

• “Module” is defined by pathway(s) that lead(s) to certain expression.
Simulation: understanding dynamic properties

Boolean modules

Merged RASF model

Easy to observe input-output relation and tune systems

Smaller size, less complexity, easier to check input output relationships (one output per module)

To observe global response

252 components
390 interactions (high overlap between modules)

Easier fine-tuning
Simulations on the AKT pathway:

- Results show that both the module and the whole model were able to **recapitulate** known regulation of AKT.

<table>
<thead>
<tr>
<th>Function</th>
<th>Components</th>
<th>Biological function</th>
<th>Module simulation</th>
<th>Model simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apoptosis</td>
<td>BAX</td>
<td>inactive</td>
<td>inactive</td>
<td>inactive</td>
</tr>
<tr>
<td></td>
<td>MDM2</td>
<td>active</td>
<td>active</td>
<td>active</td>
</tr>
<tr>
<td></td>
<td>FOXO1</td>
<td>inactive</td>
<td>inactive</td>
<td>inactive</td>
</tr>
<tr>
<td></td>
<td>BCL2L11</td>
<td>inactive</td>
<td>inactive</td>
<td>inactive</td>
</tr>
<tr>
<td>Proinflammatory response</td>
<td>NF-kB</td>
<td>active</td>
<td>N/A</td>
<td>active</td>
</tr>
<tr>
<td></td>
<td>IL-6</td>
<td>active</td>
<td>N/A</td>
<td>active</td>
</tr>
<tr>
<td></td>
<td>TNF</td>
<td>active</td>
<td>N/A</td>
<td>active</td>
</tr>
<tr>
<td>Chemokine secretion</td>
<td>IL-8</td>
<td>active</td>
<td>N/A</td>
<td>active</td>
</tr>
<tr>
<td></td>
<td>CXCL 1-3</td>
<td>active</td>
<td>N/A</td>
<td>active</td>
</tr>
<tr>
<td>Bone &amp; matrix degradation</td>
<td>MMP3</td>
<td>active</td>
<td>N/A</td>
<td>active</td>
</tr>
<tr>
<td></td>
<td>RANKL</td>
<td>active</td>
<td>N/A</td>
<td>active</td>
</tr>
<tr>
<td>Cell cycle control</td>
<td>CDKN1B</td>
<td>inactive</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Glycolysis</td>
<td>GSK3</td>
<td>inactive</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Difficulties:

• Issues of interoperability between different tools (CellDesigner, Cell Collective, GINsim).
• Hard to calculate attractors.
• 4 modules still large and complex.
• Computationally costly.
Conclusions:

• **Successful translation** of a molecular map to an executable model.

• Pipeline **applicable to other maps** available.

• **Flexibility:** direct translation for cell specific maps, modular approach for network subtraction.

• The resulting model **retains annotations, references and hierarchical layout** of the original map, facilitating reusability and simulations.
To do list:

- **Further simulations and testing is needed**, both on the sub-modules and the merged RASF model in order to have a fully functional model.

- **Model reductions will be required** to solve the computational explosion for in depth analysis of dynamical properties (finding attractors).

- **Tool development is needed** in order to bypass technical issues and computational cost.

- Application of **control theory to Boolean networks** (currently not adapted to large scale) based on topological features (like betweenness centrality).
Dream team:

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

Saran Pankaew,
University of Evry, Paris-Saclay
M2SSB Student

Anna Niarakis, PhD
University of Evry, Paris-Saclay
Associate Professor, Department of Biology

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

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

Alexander Mazein, PhD
European Institute for Systems Biology and Medicine, France
Senior Researcher

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

Tomas Helikar, PhD
University of Nebraska-Lincoln
Assistant professor

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

In loving memory

Gilles Chiocchia, PhD
University of Versailles-Saint-Quentin en-Yvelines
Professor - Hospital Practitioner