Towards patient-specific multi-scale models and data integration for clinical stratification

Arnau Montagud
Computational Systems Biology of Cancer
U900 – Institut Curie
A bit about myself

• B.Sc. Biology
  – Uni València

• M.Sc. Cell Biology
  – Uni València

• Ph.D. in Applied Maths
  – Uni Polit València
  – Genome-scale metabolic model of hydrogen-producing cyanobacteria

• Moved on to cancer

• Arrived to Institut Curie in 2014

• Worked on
  – Data deconvolution, ICA
  – Logical models
  – Multi-scale models

• Involved in projects on
  – Breast cancer (INVADE)
  – Medulloblastoma (M5)
  – Prostate (PrECISE)
Logical models formalism

- The questions are **qualitative**
- The **data are discrete** (mutations, copy number, etc.)
- Expression data are **not absolute values**
- No information over **time**
- No details about the precise biochemical reactions
Translation of an influence network into Boolean logic

Regulatory graph

A = !B & C
B = A
C = input

Each variable can take two states: 0 or 1

Logical rules

Boolean logic:
Connectors: AND (&), OR (|), NOT (!), XOR (/)
Logic depends on incoming arrows

Solutions

Attractors are subgraphs of the state transition graph with no outgoing arrows.
stable states & cyclic attractors
Stable states = cell fates = phenotypes
**Boolean model results**

- **Stable states table**
  - Each stable state corresponds to a **biological situation/context**

- **Inputs** are set
- **Mechanisms** are known
- **Objectives** are stable states

- Once stable states are clear
  - Perturbing the mechanisms is highly informative
  - **Overexpression / Knock out**

---

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Outputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECMMicroenv</td>
<td>GF</td>
</tr>
<tr>
<td>DNA damage</td>
<td>TCFbeta</td>
</tr>
<tr>
<td>Metastasis</td>
<td>p21</td>
</tr>
<tr>
<td>Migration</td>
<td>CDH1</td>
</tr>
<tr>
<td>Invasion</td>
<td>CDH2</td>
</tr>
<tr>
<td>EMT</td>
<td>VIM</td>
</tr>
<tr>
<td>Apoptosis</td>
<td>TWIST1</td>
</tr>
<tr>
<td>CellCycleArrest</td>
<td>SNAI1</td>
</tr>
<tr>
<td>GF</td>
<td>SNAI2</td>
</tr>
<tr>
<td>TCFbeta</td>
<td>ZEB1</td>
</tr>
<tr>
<td>p21</td>
<td>ZEB2</td>
</tr>
<tr>
<td>CDH1</td>
<td>AKT1</td>
</tr>
<tr>
<td>CDH2</td>
<td>DKK1</td>
</tr>
<tr>
<td>VIM</td>
<td>CTNNB1</td>
</tr>
<tr>
<td>TWIST1</td>
<td>NICD</td>
</tr>
<tr>
<td>SNAI1</td>
<td>p63</td>
</tr>
<tr>
<td>SNAI2</td>
<td>p53</td>
</tr>
<tr>
<td>ZEB1</td>
<td>p73</td>
</tr>
<tr>
<td>ZEB2</td>
<td>miR200</td>
</tr>
<tr>
<td>AKT1</td>
<td>miR203</td>
</tr>
<tr>
<td>DKK1</td>
<td>miR34</td>
</tr>
<tr>
<td>CTNNB1</td>
<td>AKT2</td>
</tr>
<tr>
<td>NICD</td>
<td>ERK</td>
</tr>
<tr>
<td>p63</td>
<td>SMAD</td>
</tr>
</tbody>
</table>
Analysing mutants: GINsim

WT

TWIST1 KO
Limits of ODE and Boolean models

<table>
<thead>
<tr>
<th>ODE modelling:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• <em>Parameters</em>: chemical reaction <em>rate constants</em>, initial conditions.</td>
</tr>
<tr>
<td>• <em>Results</em>: quantitative concentrations of genes/proteins over <em>continuous time</em>.</td>
</tr>
<tr>
<td>Problems:</td>
</tr>
<tr>
<td>→ need to know <em>details</em> such as (relative) concentrations and speed of reactions</td>
</tr>
<tr>
<td>→ many <em>unknown</em> parameters</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Boolean modelling:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• <em>Parameters</em>: logical equations describing <em>activation of genes/proteins</em>.</td>
</tr>
<tr>
<td>• <em>Results</em>: activities (on/off) of genes/proteins.</td>
</tr>
<tr>
<td>Problems:</td>
</tr>
<tr>
<td>→ <em>coarse-grain description</em> of biological processes</td>
</tr>
<tr>
<td>→ <em>qualitative</em> results on final states only</td>
</tr>
</tbody>
</table>
Asymptotic solutions

Stable state solutions, where the system can no longer evolve

Probabilities of reaching a state from an initial condition

Method:
- continuous time Markov process / Gillespie algorithm
- a rate of change associated to each transition (separate rate up and rate down)

⇒ To each Boolean state, a probability is associated

http://www.ginsim.org
https://maboss.curie.fr
Analysing stable states: MaBoSS

- Continuous time Markov process on the Boolean transition state space
  - Each Boolean state has an associated probability
  - Rate of change associated to each transition
    - rate up and rate down
  - Stochasticity, time, probabilities, ...

- Perturbations can be studied in a probabilistic manner
  - Transient effects, such as Knock downs
  - Dosage experiments

- Software developed by Stoll *et al*, *BMC Systems Biology* 2012
  - DOI: 10.1186/1752-0509-6-116
  - MaBoSS 2.0: DOI: 10.1093/bioinformatics/btx123

Stoll et al. (2012) *BMC Syst Biol*
Gillespie algorithm allows of stochastic simulations

**Conditions**

- **Nutrients**: 50%
- **Androgen**: 50%
- **GFs**: 50%
- **Carcinogen**: 10%
- **Hypoxia**: 10%

**Signalling pathways**

- **Apoptosis**: 10%
- **Proliferation**: 23%
- **Quiescence**: 67%
- **EMT**: 0%
- **Metastasis**: 0%
- **Angiogenesis**: 10%
- **DNA repair**: 5%
- **Migration**: 0%

**Probabilities based on 1000 trajectories**

- Mutations (ex: MEK==0): 75%

**State probabilities**

- 0.00 to 1.00
- Time: 0 to 15
The solutions of the Boolean model can be interpreted biologically

- **Time variations**
  - Similar to robustness analysis of dynamic systems
  - Transient effects can be seen

- **Population heterogeneity**
  - Mutants can be studied semi-quantitatively
The model confirms the appearance of metastasis in the Notch++/p53-- double mutant


*Chanrion et al. Nature Communications 2014
Logical modelling pipeline

• Extensive and comprehensive studies can be done with a validated model and tools on the field

• Experimentally friendly hypotheses can come out of these studies

• We aim to:
  – include other tools in the pipeline to get more insight from a model
  – be included in existing frameworks
  – show examples using other models in SBML format
  – provide a repository with all scripts

• [https://github.com/sysbio-curie/Logical_modelling_pipeline](https://github.com/sysbio-curie/Logical_modelling_pipeline)

Briefings in Bioinformatics, 2017, 1–12
doi: 10.1093/bib/bbx163
Paper

Conceputal and computational framework for logical modelling of biological networks deregulated in diseases

Arnau Montagud, Pauline Traylor, Loredana Martignetti, Eric Bonnet, Emmanuel Barillot, Andrei Zinovyev and Laurence Calzone
Instantiation of patient-specific logical models with multi-omics data allows clinical stratification of patients
Modelling patient-specific Boolean networks

Phenotype probabilities can be studied as time trajectories or populations’ dynamics.

Growth factors, nutrients and androgen

Same with hypoxia
Selecting instantiation recipe for the patients’ profiles

• Choosing the proper **data type** in the proper **model variable**

• Modelling framework:
  - Node states = mutants
  - **Initial conditions** = growth media conditions or experimental setup
  - Transitions rates = gene’s ability to activate or deactivate

• Data types:
  - CNA
  - Mutations
  - **Expression**: RNA and/or proteins

![Boolean state transition graph]

Stoll et al. (2012)
*BMC Syst Biol*
Selecting instantiation recipe: METABRIC

- Different data in different variables yields different phenotypes’ distributions
- Which one do we use and why?

Patients’ phenotypes distribution

Scores for the original model without instantiation
Selecting instantiation recipe: METABRIC

- We use correlation as
  - a means to validate results and
  - select a good combination of data in the model variables that tallies clinical data

- **METABRIC:**
  - NPI and MKI67 RNA and Proliferation score

- **TCGA:**
  - Protein or RNA signature and Proliferation score
Selecting instantiation recipe: TCGA Breast

- We use correlation as
  - a means to validate results and
  - select a good combination of data in the model variables that tallies clinical data

- METABRIC:
  - NPI and MKI67 RNA and Proliferation score

- TCGA:
  - Protein or RNA signature and Proliferation score
Patients’ phenotypes correlate with clinical data: METABRIC models vs PAM50

RNAseq data related to model’s genes

Model simulations

PAM50 Subtypes - Model-related RNA
With principal components PC1 and PC2 as X and Y axis

PAM50 Subtypes - PCA with Simulation Outputs from Case #5
With principal components PC1 and PC2 as X and Y axis

Subtypes
- Basal
- Claudin-low
- Her2
- LumA
- LumB
- Normal

Density levels
- 0.005
- 0.010
- 0.015
- 0.020
- 0.025
- 0.030
- 0.035
Patients’ phenotypes correlate with clinical data: METABRIC models vs survival data.
Instantiation of logical models

- Flexible pipeline of methods
  - Available at GitHub

- Generates data-tailored models
  - Cell lines
  - Patients

- Correlates with clinical data
  - Highly dependent on available clinical data

https://github.com/sysbio-curie/Instantiation_logical_models

Instantiation of Patient-Specific Logical Models With Multi-Omics Data Allows Clinical Stratification of Patients

Jonas Béal¹, Arnaud Montagud¹, Pauline Traynard¹, Emmanuel BARILLOT¹, Laurence Calzone¹

¹Institut Curie, France

Submitted to Journal:
Frontiers in Physiology
Specialty Section:
Systems Biology
PhysiBoSS: a multi-scale agent based modelling framework integrating physical dimension and cell signalling
# Genotype-to-phenotype modelling

- **Hormones**
- **ECM density**
- **Mutations**
- **Cell-cell adhesion**
- **Cell-ECM adhesion**

**Multiscale modelling**
- Gene mutations
- Signalling pathways
- Cell - environment
- ECM modification

---

<table>
<thead>
<tr>
<th>Multicellular migration</th>
<th>Cell-cell junctions</th>
<th>Tumor type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amoeboid</td>
<td>-</td>
<td>Leukemia, lymphoma cell subsets (all tumors)</td>
</tr>
<tr>
<td>Mesenchymal</td>
<td>-</td>
<td>Stromal tumors, epithelial tumors after EMT</td>
</tr>
<tr>
<td>Amoeboid (multicellular)</td>
<td>?</td>
<td>All tumors developing amoeboid single-cell dissemination</td>
</tr>
<tr>
<td>Mesenchymal (multicellular)</td>
<td>(+)</td>
<td>Tumors with mesenchymal invasion; fibroblasts leading tumor cells</td>
</tr>
<tr>
<td>Cluster</td>
<td>++</td>
<td>Moderately differentiated epithelial tumors</td>
</tr>
<tr>
<td>Solid strand</td>
<td>++</td>
<td>Moderately differentiated epithelial tumors with subregions after EMT; basal and squamous cell carcinoma</td>
</tr>
<tr>
<td>Strand (with lumen)</td>
<td>++</td>
<td>Differentiated epithelial tumors; vascular neoplasia</td>
</tr>
<tr>
<td>Strand (protrusive)</td>
<td>++</td>
<td>Moderately differentiated epithelial tumors lacking EMT</td>
</tr>
<tr>
<td>Outward pushing tumor</td>
<td>++</td>
<td>All solid tumors</td>
</tr>
</tbody>
</table>

From Friedl and Alexander, Cell, 2011
PhysiBoss: multiscale modelling framework

- Developed jointly with Gaelle Letort
- Multi-scale modelling of cancerous phenotypes

- Merges agent-based modelling and Boolean modelling
  - PhysiCell
    - Paul Macklin's Lab, Indiana University
    - MaBoSS
- Each agent has a Boolean model

16 days
PhysiBoss: multiscale modelling framework

Cells can:
- Grow
- Divide
- Attach
- Move
- Degrade ECM

Legend:
- Green: growth
- Blue: quiescent
- Red: apoptosis
- Grey: ECM proxy

Time = 45 hours
PhysiBoss: multiscale modelling framework

Cells can:
• Grow
• Divide
• Attach
• Move
• Degrade ECM

Legend:
- Green: growth
- Blue: quiescent
- Red: apoptosis
- Grey: ECM proxy

Time = 45 hours
PhysiBoss: multiscale modelling framework

Cells can:
- Grow
- Divide
- Attach
- Move
- Degrade ECM

Legend:
- Blue: cells centre
- Grey: ECM proxy

Time = 45 hours
PhysiBoss: different cell strains or environment

- **Different cell strains** can be used
  - Different **biology**
    - Boolean network
    - MaBoSS parameters
  - Different **physical properties**
    - Cell-cell adhesion
    - Cell-matrix adhesion
- **Dynamical environment** can be used
  - Pulses of TNF

Pulses frequency

- one pulse = 0.5 ng/mL
  - 10 min

**Every 150 min**

**Every 300 min**

**Every 600 min**
PhysiBoss: multiscale modelling framework

- Open source code for multiscale modelling
  - Available at bioRxiv and GitHub

- Uses agent-based modelling for physical phenomena
  - Forked from PhysiCell
    - Paul Macklin's Lab, Indiana University

- Uses Boolean modelling for biological phenomena
  - Adapted from MaBoSS

https://github.com/sysbio-curie/PhysiBoSS
ICA uncovers clinical traits that cause breast cancer stratification
Identify independent sources that cause data variability

Assumption of statistical independence of factor activities: Independent Component Analysis (ICA)

Factor 1

Factor 2

...  Factor m

Gene 1

Gene 2

Gene 3

...  Gene n

\[ m \ll n \]

(Cocktail party problem)

(Zinovyev et al, BBRC, 2013)

(Zinovyev et al, BBRC, 2013)

(Zinovyev et al, BBRC, 2013)

(Cocktail party problem)

(Alex Katz, The Cocktail Party (1965))

(Alex Katz, The Cocktail Party (1965))

(Alex Katz, The Cocktail Party (1965))
Identify independent sources that cause data variability

- ICA is a statistical method for
  - transforming an *observed multidimensional* random vector into
  - *components* that are statistically as *independent* from each other as possible

- When do we generally need ICA?
  - when data is *not* Gaussian
  - when *raw data* appear to be very *noisy*
  - when a sensor records *several source signals simultaneously*

---

Zinovyev et al, BBRC, 2013
Independent Component Analysis on RNAseq data

• Unique set of early invasion breast cancer data from Vincent-Salomon lab, Institut Curie
• 56 samples of diverse
  – histology
    • 19 in situ
    • 21 invasive
    • 16 micro-invasive
  – molecular subtypes
    • 23 LumA
    • 14 LumB
    • 11 Her2
    • 3 LumB/Her2
    • 5 TN
Invasion

Molecular subtypes

Immuno

Basal subtype

Invasive Subtype

ICA components on RNAseq
Invasive components

Invasion by proliferation

Invasion by ECM modification
Tumour invasion is diverse

• Tumour is a mix of cells with potentially different
  – Genome
  – Transcriptome
  – Neighbours
  – Local microenvironment
    • Density
    • Architecture

• Diversity of invasion modes
  – Tissue-specific
  – Epithelial to mesenchymal transition (EMT)

Video from P. Chavrier, Institut Curie
Breast invasion model
Outward pushing tumour

- **PhysiBoSS**: Multi-scale modelling of
  - Physics-oriented cellular population
  - Biology-oriented intracellular cell fates

- Stable state
  - Attached cells proliferating

- Colour
  - (Up) Red is ECM node status
  - (Down) Red is ECM signal
Multi-scale modelling of invasive phenotypes: on going work

- Use patient-specific models
- Obtain patient-specific phenotypes
- Compare these phenotypes with visual resources from patients
Acknowledgments

Computational Systems Biology of Cancer group

Emmanuel Barillot
Jonas Béal
Laurence Calzone
Luca Albergante
Laura Cantini
Urszula Czerwinska
Mihaly Koltaï
Maria Kondratova
Gaelle Letort
Pauline Traynard
Inna Kuperstein
Christine Lonjou
Loredana Martignetti
Cristóbal Monraz
Andrei Zinovyev

Eric Bonnet
Gautier Stoll
David Cohen
Paul Deveau

Collaborators
Paul Macklin (Indiana University)
Denis Thieffry (ENS)
Luca Grieco (University College London)
Sylvie Robine (Institut Curie, UMR 144)

arnau.montagud@gmail.com