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

Arnau Montagud

Computational Systems Biology of Cancer

U900, Institut Curie

institutCurie

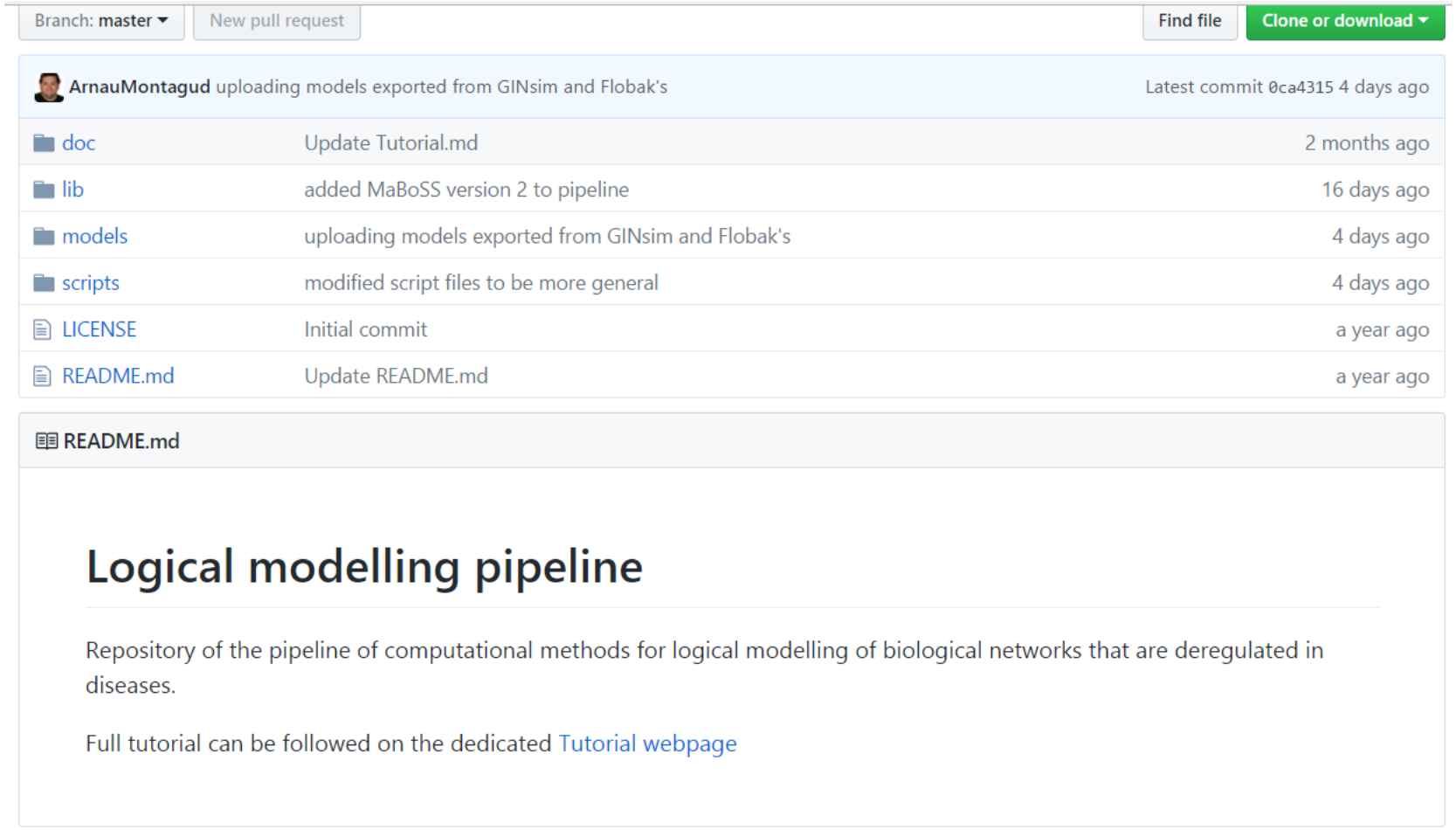
Instituts thématiques Inserm Institut national de la santé et de la recherche médicale

MINES ParisTech

PSL RESEARCH UNIVERSITY

Logical modelling pipeline

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



institutCurie

Instituts thématiques Inserm
Institut national de la santé et de la recherche médicale

MINES ParisTech

PSL RESEARCH UNIVERSITY

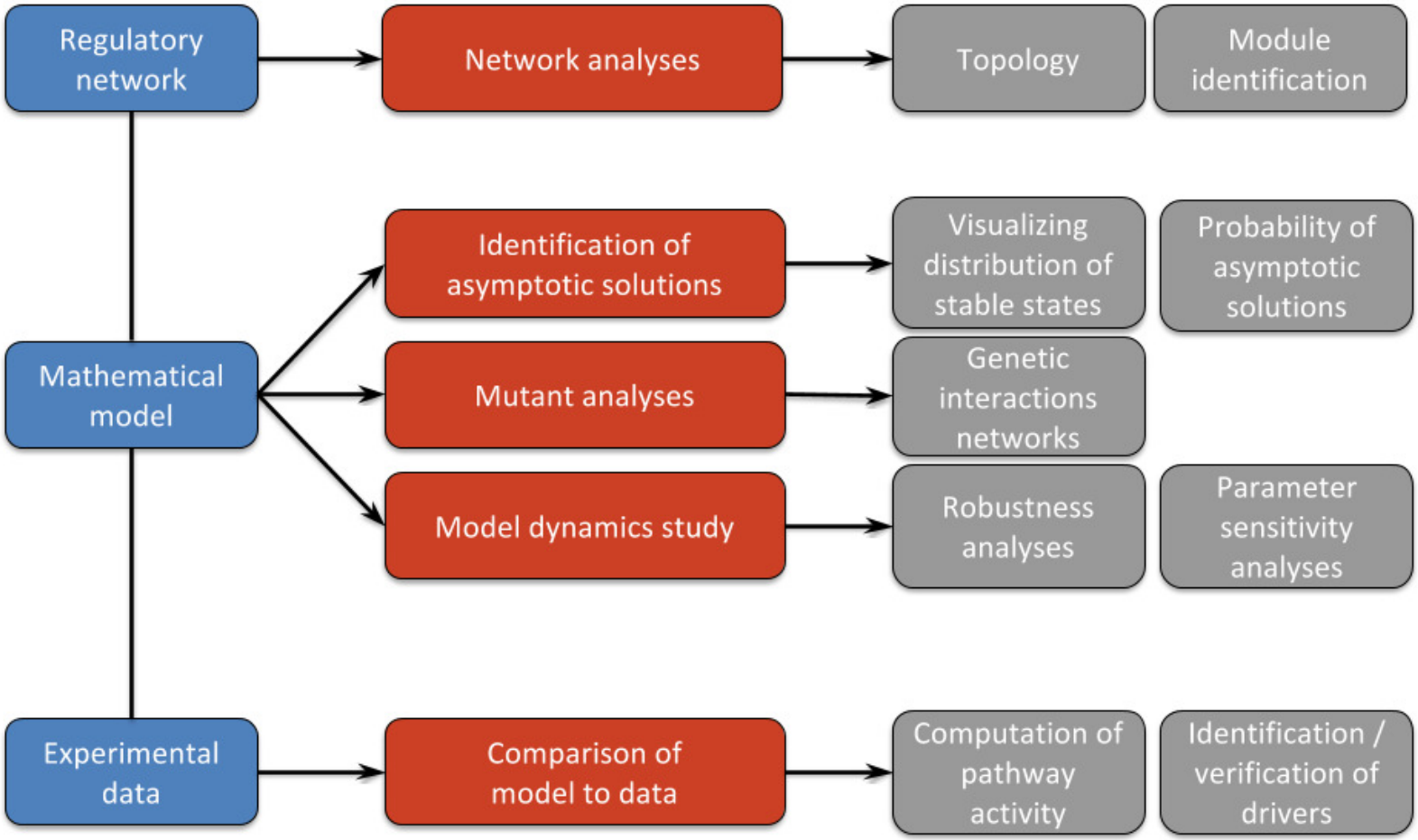
How to extract as much information as possible from a model?

A model is built to answer a particular question… but how much more can we get out of it?

3 types of approaches:

- analysis on the structure of the network
- analysis of the mathematical model
- link data with the network/model

Pipeline



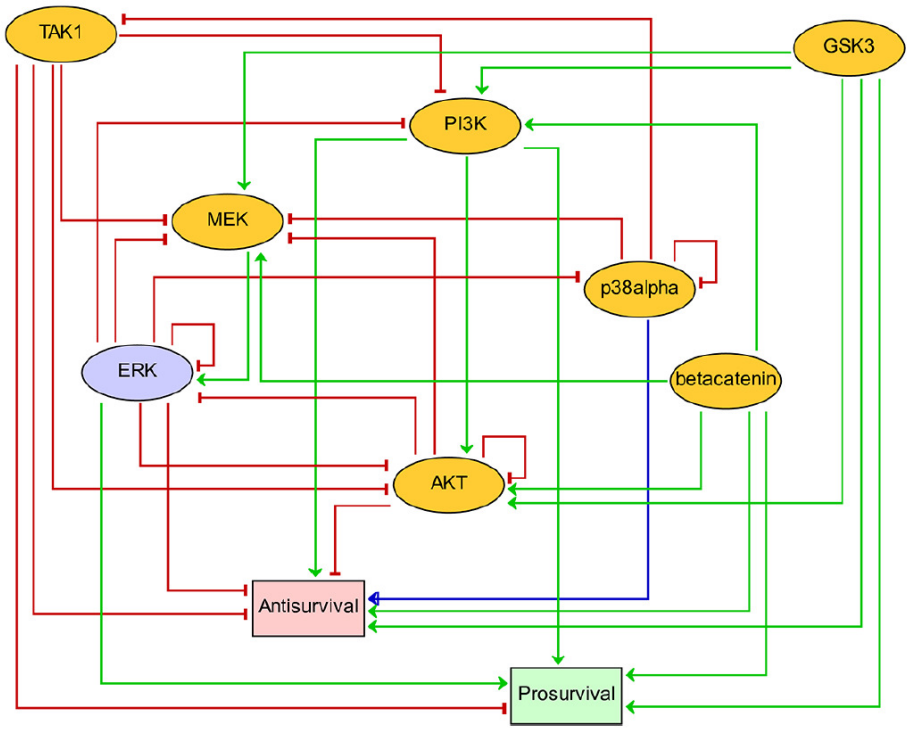
Example on a Boolean model

PLOS | COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Discovery of Drug Synergies in Gastric Cancer Cells Predicted by Logical Modeling

Åsmund Flobak[1]*, Anaïs Baudot[2], Elisabeth Remy[2], Liv Thommesen[1,3], Denis Thieffry[4,5,6], Martin Kuiper[7], Astrid Lægreid[1]*



Cell fate decision network in the AGS gastric cancer cell line, with 75 signalling and regulatory components

Reduced model has 10 nodes

Analyses by Pauline Traynard

What insights can we get from the mathematical model

Types of questions to be answered

- what are the solutions of the model that can be interpreted biologically?
- what are the important nodes of the network?
- how robust/sensitive is the model?
- what nodes could be altered (i.e. by mutations of genetic alterations) to account for a clinical output (e.g. stage of the tumor or metastasis) in a deregulation of a normal situation (e.g. tumorigenesis)?
- can we predict genetic interactions (epistasis, synthetic lethality) from the model?
- can we simplify/reduce the model to highlight the most important processes?

Asymptotic solutions

Stable state solutions, where the system can no longer evolve

Probabilities of reaching a state from an initial condition

Name	MEK	ERK	TAK1	p38alpha	AKT	PI3K	GSK3	betacatenin	Antisurvival	Prosurvival
	1	1	1						1	
	1	1	1		1	1		1		3
	1	1	1		1	1	1			1
	1	1	1		1	1	1	1		3

Name	MEK	ERK	TAK1	p38alpha	AKT	PI3K	GSK3	betacatenin	Antisurvival_b1	Antisurvival_b2	Antisurvival_b3	Prosurvival_b1	Prosurvival_b2	Prosurvival_b3
	1	1	1						1					
	1	1	1		1	1		1				1	1	1
	1	1	1		1	1	1					1		
	1	1	1		1	1	1	1				1	1	1

Method:

- continuous time Markov process / Gillespie algorithm on the transition state space
- a rate of change associated to each transition (separate rate up and rate down)

⇒ To each Boolean state, a probability is associated

Each stable state corresponds to a biological situation/context

http://www.ginsim.org

https://maboss.curie.fr

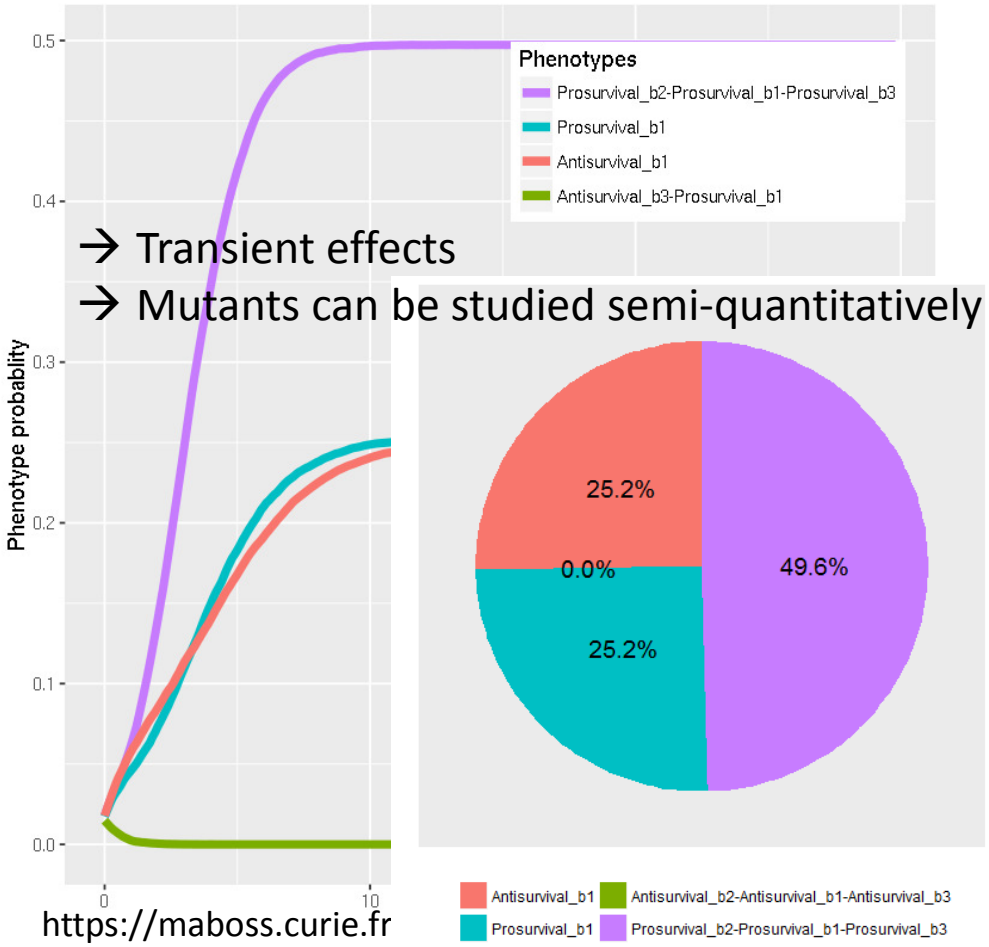
Asymptotic solutions

Stable state solutions, where the system can no longer evolve

Probabilities of reaching a state from an initial condition

Name	MEK	ERK	TAK1	p38alpha	AKT	PI3K	GSK3	betacatenin	Antisurvival	Prosurvival
	1	1	1						1	
	1	1	1		1	1		1		3
	1	1	1		1	1	1			1
	1	1	1		1	1	1	1		3

Name	MEK	ERK	TAK1	p38alpha	AKT	PI3K	GSK3	betacatenin	Antisurvival_b1	Antisurvival_b2	Antisurvival_b3	Prosurvival_b1	Prosurvival_b2	Prosurvival_b3
	1	1	1						1					
	1	1	1		1	1		1				1	1	1
	1	1	1		1	1	1					1		
	1	1	1		1	1	1	1				1	1	1



→ Transient effects

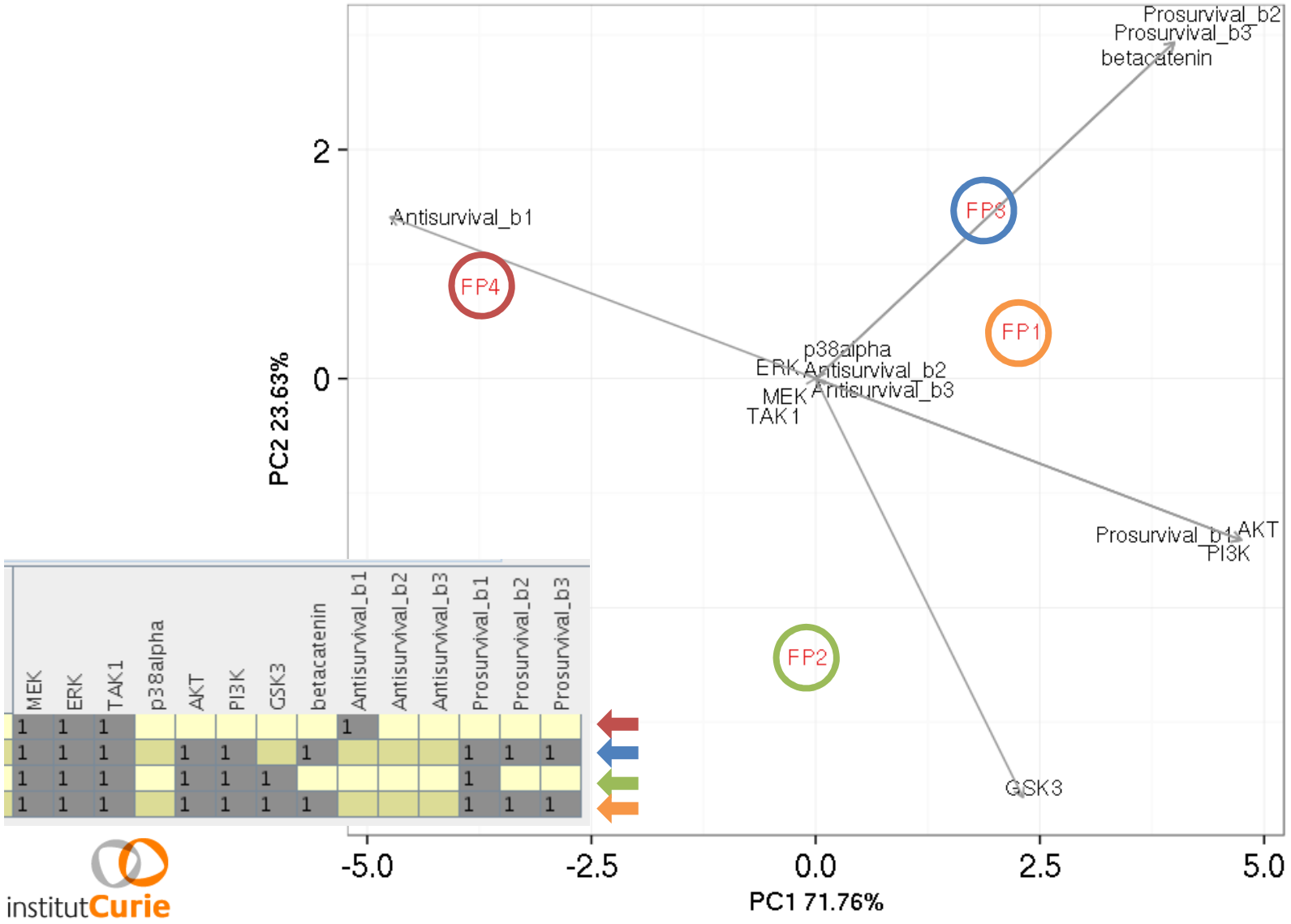
→ Mutants can be studied semi-quantitatively

Each stable state corresponds to a biological situation/context

http://www.ginsim.org

https://maboss.curie.fr

Can we classify the solutions of the Boolean model?

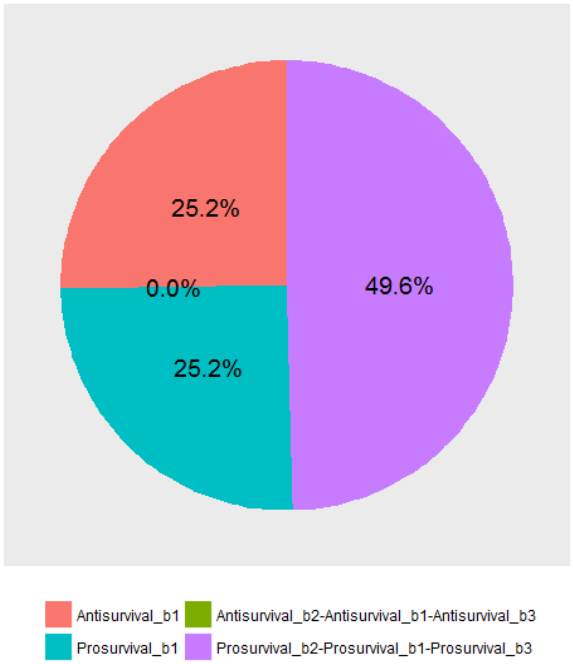


institutCurie

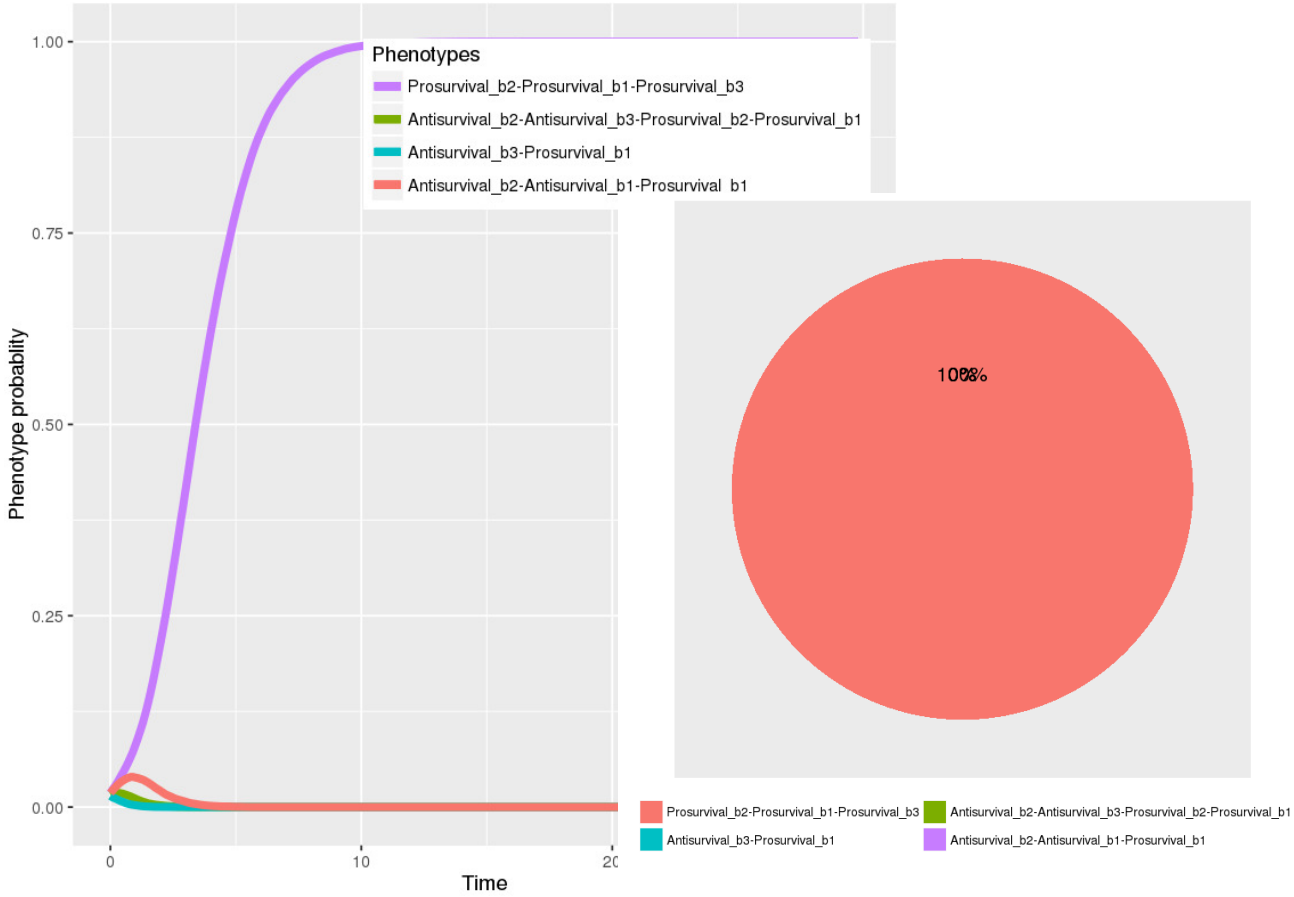
PSL RESEARCH UNIVERSITY

Mutants in MaBoSS

Wild type



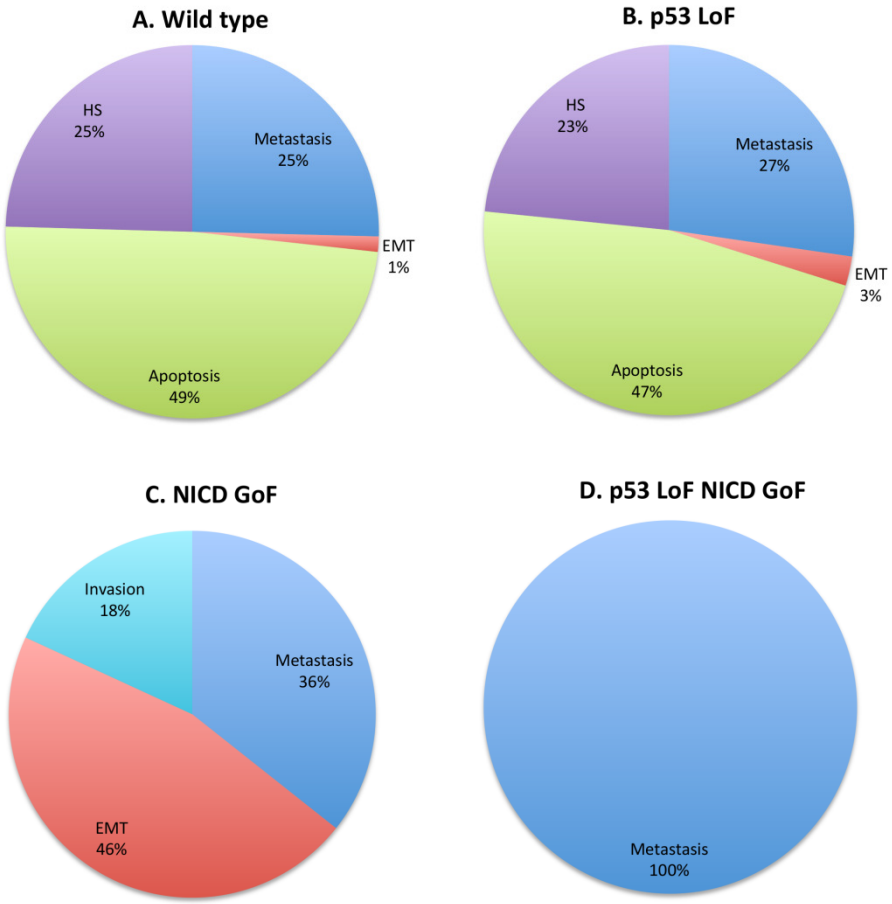
Mutant



betacatenin=1 and *GSK3*=0: Prosurvival stable state is selected

Mutants in MaBoSS

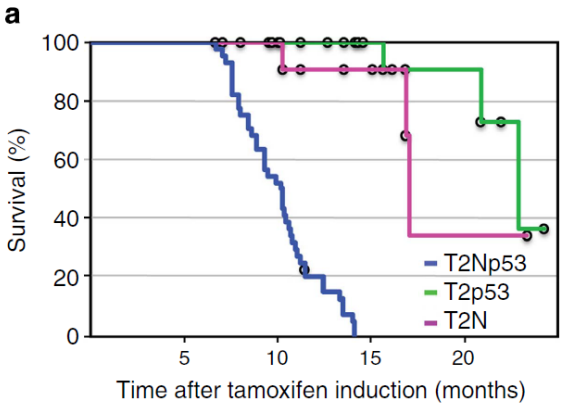
Model prediction

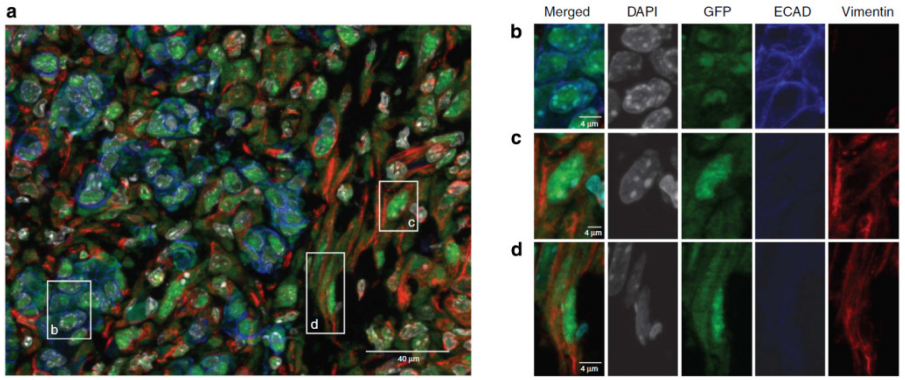


Cohen et al. (2015) PLoS Comp Biol

Mouse experiment

NICD++/p53–

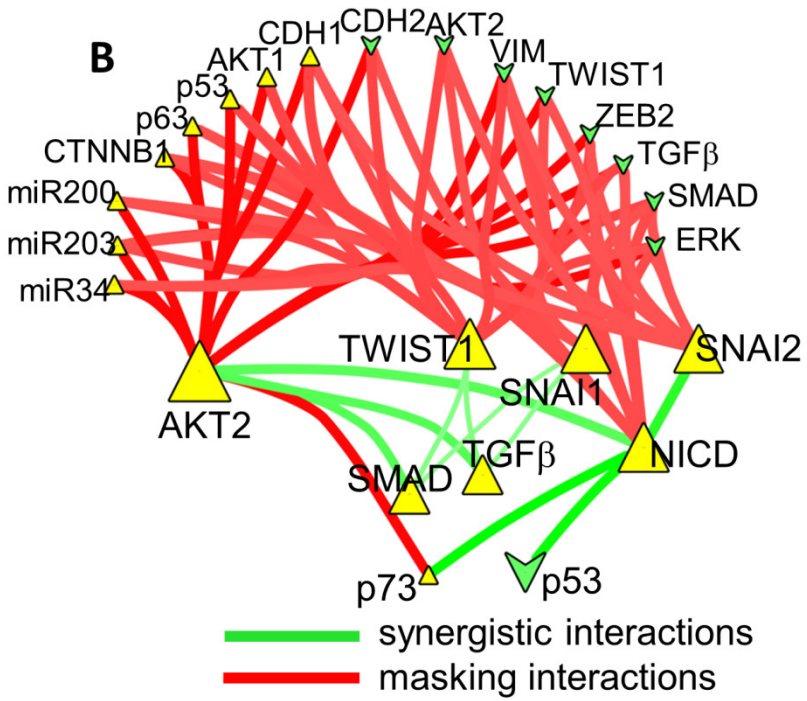




Chanrion et al. (2014) Nat Comm

The model confirms the appearance of metastasis in the Notch++/p53-- double mutant

We can predict genetic interactions



Calzone et al. (2015) Integr. Biol.

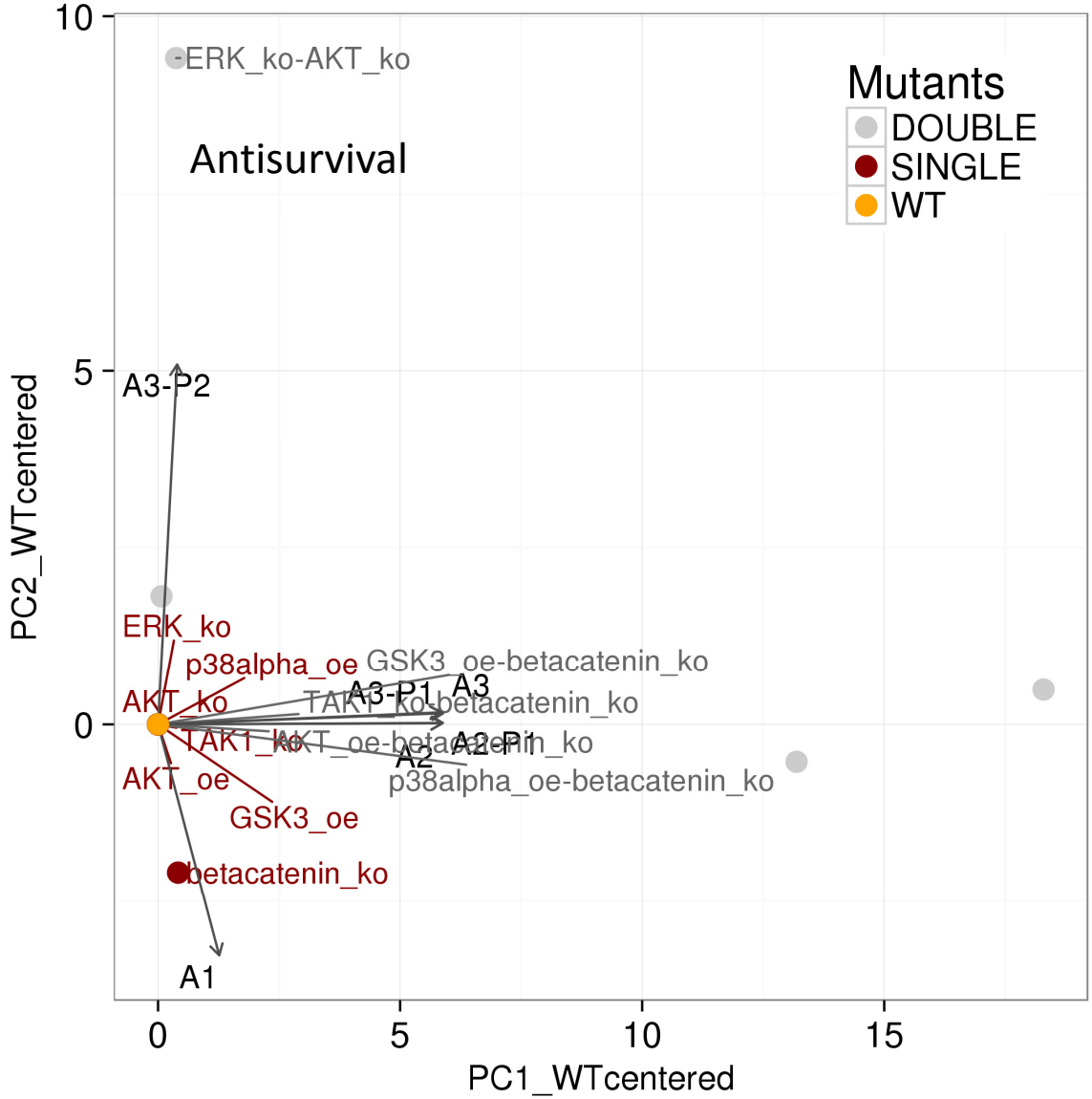
1. we generate all single and double mutants
2. we simulate MaBoSS to associate to each mutant a probability of phenotype (e.g. Metastasis)
3. we associate to double mutants, a type of genetic interactions depending on the computed epistasis value

masking interaction: the double mutant has no advantage over one of the single mutants
synergistic interaction: the double mutant is increasing or decreasing the probability of single mutants

$$\varepsilon_\phi(A,B) = f_\phi^{AB} - \psi(f_\phi^A, f_\phi^B)$$

$$\begin{aligned}
\psi^{ADD}(x,y) &= x+y && (additive)\\
\psi^{LOG}(x,y) &= log_2((2^x-1)(2^y-1)+1) && (log)\\
\psi^{MLT}(x,y) &= xy && (multiplicative)\\
\psi^{MIN}(x,y) &= \min(x,y) && (min)
\end{aligned}$$

Predicting genetic interactions

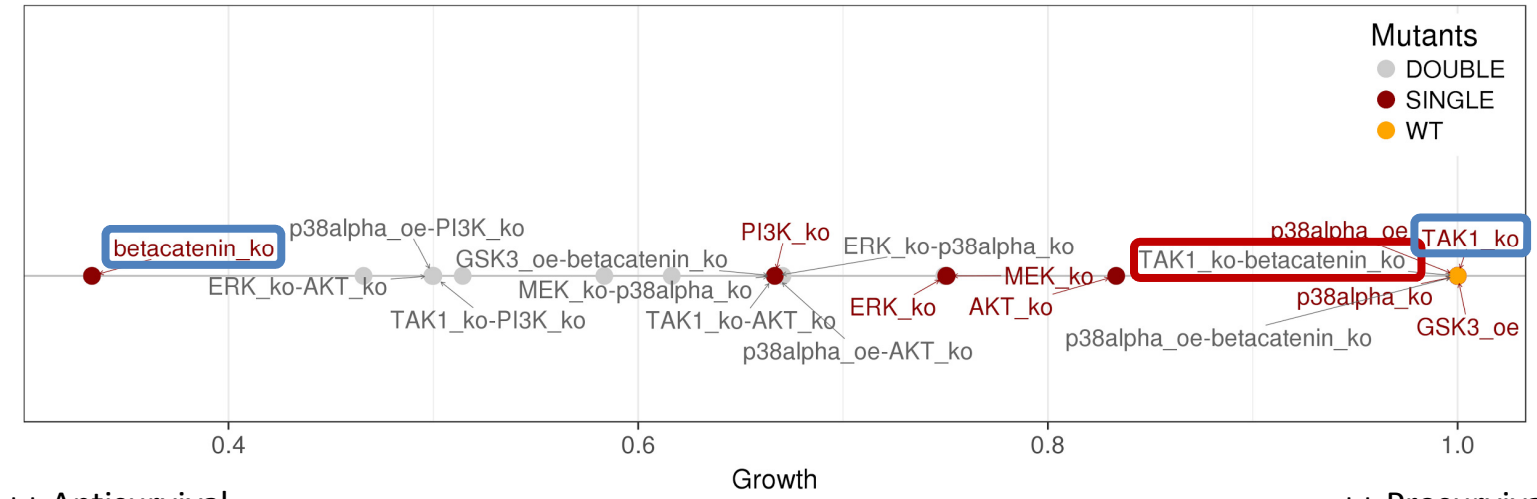


- PCA on MaBoSS output
  - WT at the centre
  - Selected phenotypes as variables
  - Mutants projected on these phenotypes
- Only looking at
  - Prosurvival
  - Antisurvival

Predicting genetic interactions

- We performed a manual merging of single phenotypes into a phenotype Growth that corresponds to the difference of
  - "Prosurvival -- Antisurvival"
  - normalized between 0 and 1

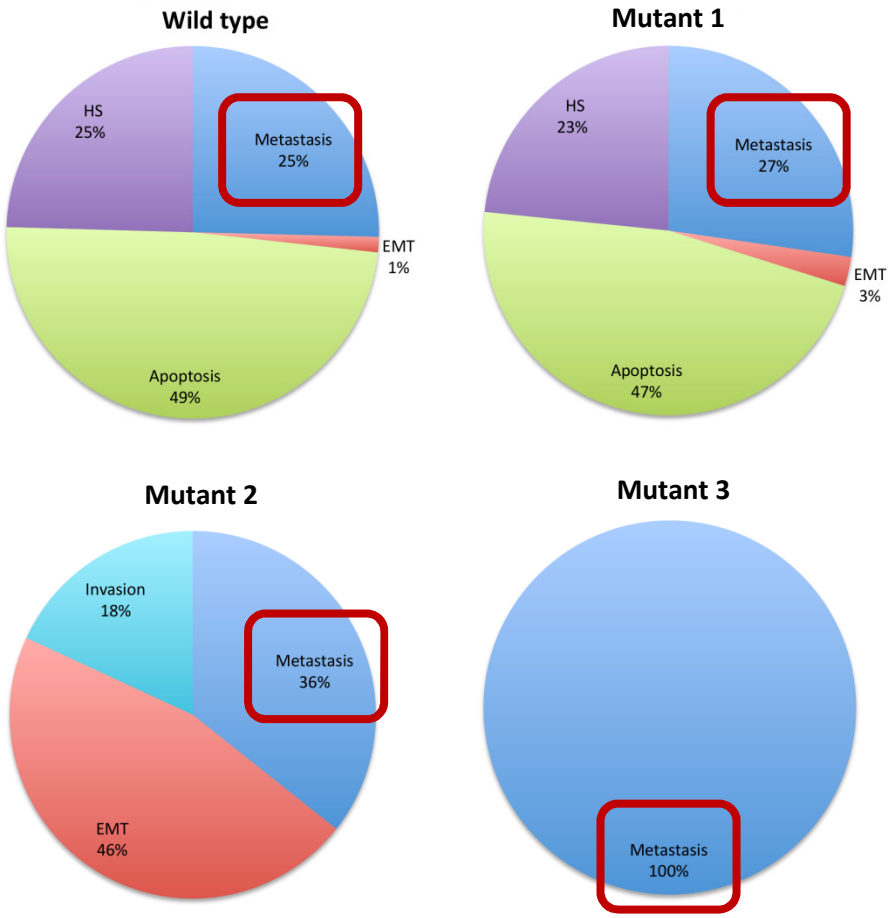
- PCA values on MaBoSS output
  - WT-normalized
  - Growth pseudo-phenotype
  - Mutants projected on this phenotype



++ Antisurvival

++ Prosurvival

Robustness analysis of genetic interactions with respect to the phenotype probability



- Ratio mutant / WT on Growth
  - Prosurvival - Antisurvival
  - Mutants and WT have different probabilities for this phenotype
  - WT bin in red

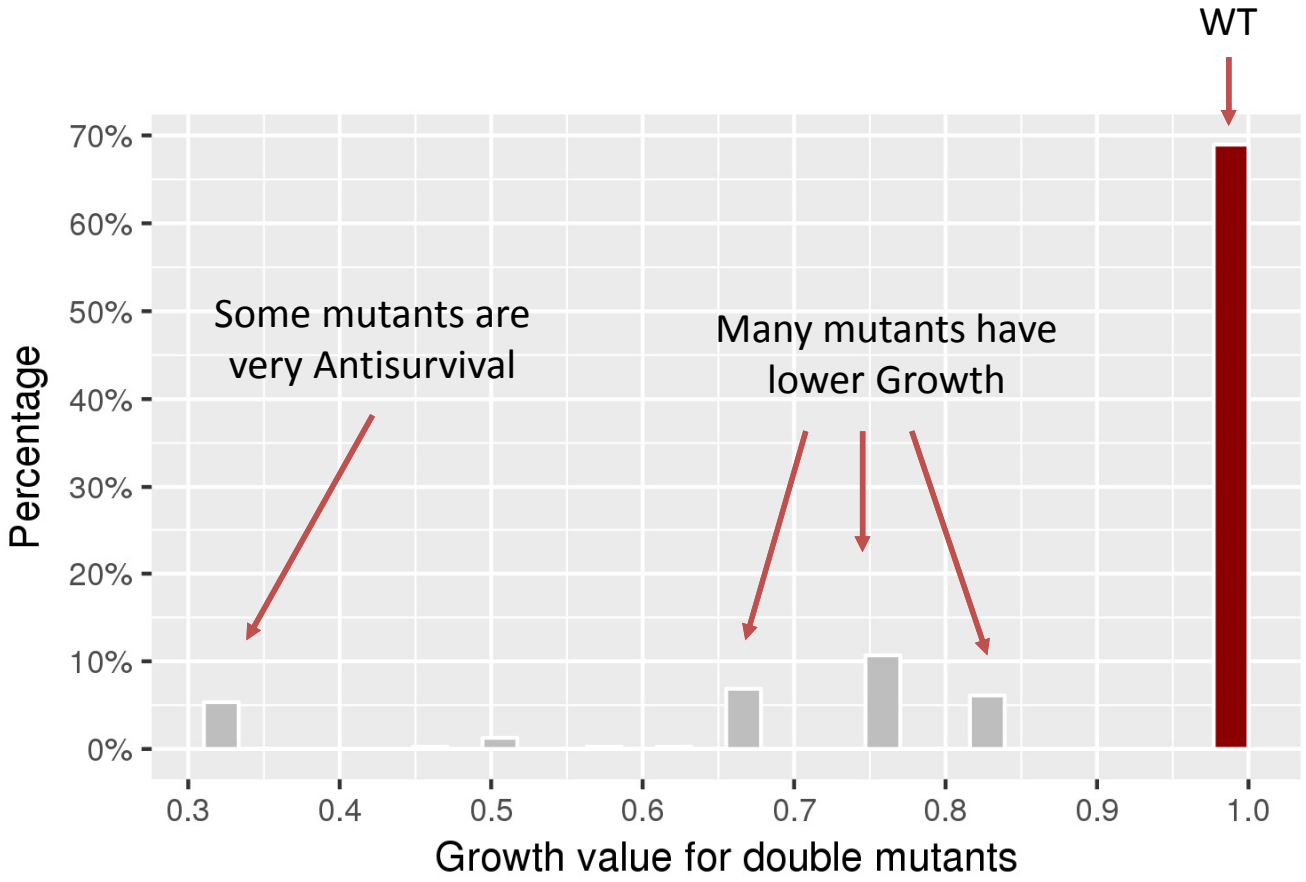
institutCurie

Instituts thématiques Inserm
Institut national de la santé et de la recherche médicale

MINES ParisTech

PSL RESEARCH UNIVERSITY

Robustness analysis of genetic interactions with respect to the phenotype probability



- Ratio mutant / WT on Growth
  - Prosurvival - Antisurvival
  - Mutants and WT have different probabilities for this phenotype
  - WT bin in red

institutCurie

Instituts thématiques Inserm
Institut national de la santé et de la recherche médicale

MINES ParisTech

PSL RESEARCH UNIVERSITY

Robustness of the model

- Can we confirm that the proposed model is robust with respect to small changes?
- Is there one model or a family of models that could be equivalent?
- Can we identify the “weak spots” of the model?

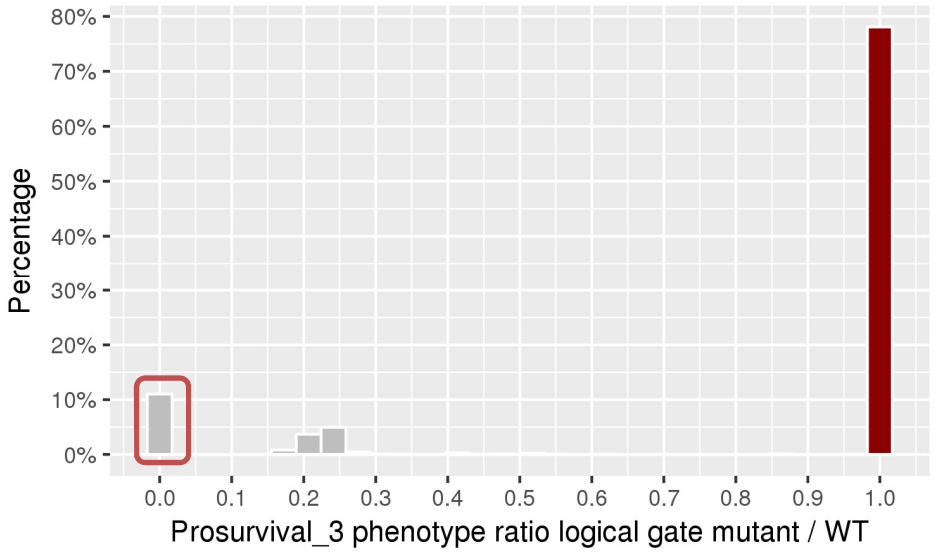
Three tests were performed:

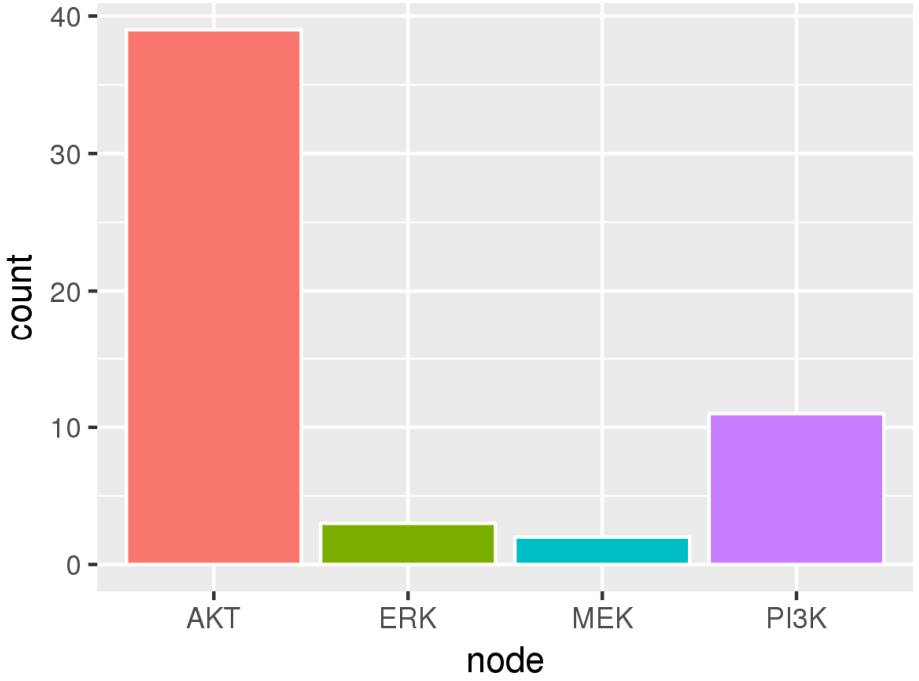
- One operator in all rules was changed
- Two operators in one rule were changed
- One operator in two rules was changed

Question: how do these changes affect the probability to reach a phenotype?

Robustness analysis of logical gates with respect to the phenotype probability

- Identify nodes whose logical rules have a drastic effects on the model properties
- The rules of some genes need to be carefully studied: AKT and PI3K in particular





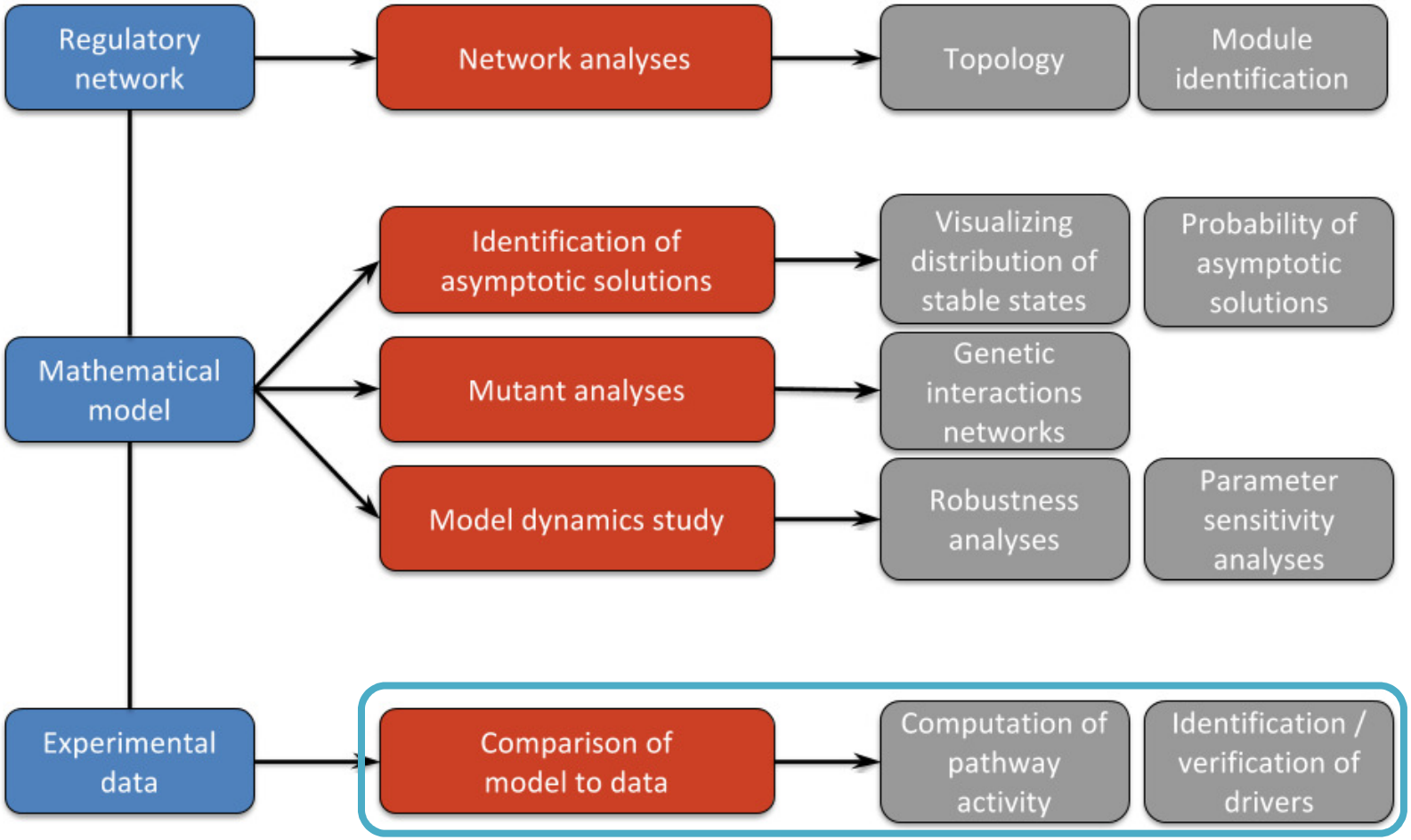
institutCurie

Instituts thématiques Inserm
Institut national de la santé et de la recherche médicale

MINES ParisTech

PSL RESEARCH UNIVERSITY

Pipeline



Logical modelling pipeline

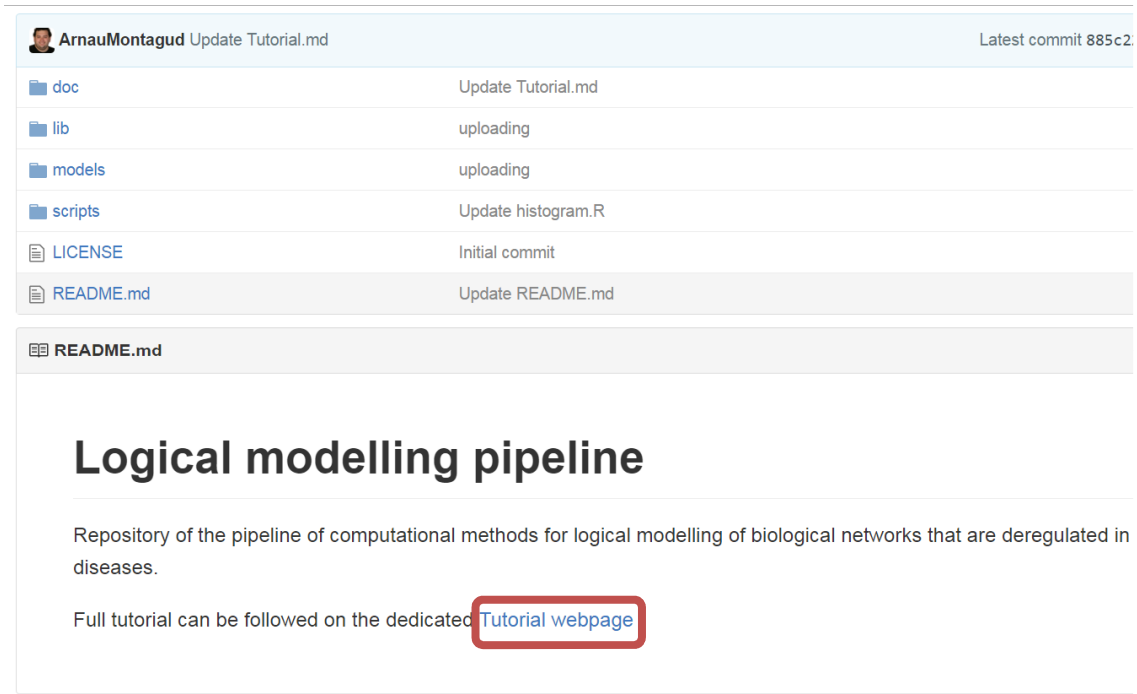
Acknowledgments

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

Laurence Calzone

Pauline Traynard

Eric Bonnet	LemonTree
Andrei Zinovyev	Robustness, epistasis
Loredana Martignetti	ROMA
Gautier Stoll	MaBoSS



institutCurie

Data to Model

- Types of questions to be answered
  - can we confirm that the genes included in the model are reasonable with respect to datasets?
  - can the model stratify patients based on the stable state solutions?
    - More aggressive tumours are associated to proliferative stable states
  - can we identify over/under activated pathways when comparing two conditions?

Data to Model

- Tools
  - LemonTree (inference of modules of co-regulated genes and their regulatory programs from data)
  - R (to compute distance from data to model)
  - ROMA (module activity)

Interpreting data with the network

- Tool: ROMA (Representation Of Module Activity)
- Command line tool

The main idea behind ROMA is:

- to define a metagene that captures the largest amount of variance
- this variance is interpreted as a result of the variability in the pathway biological activity
- to explore the activity of sets of genes (modules) rather than individual genes across samples explained by the genes in the module

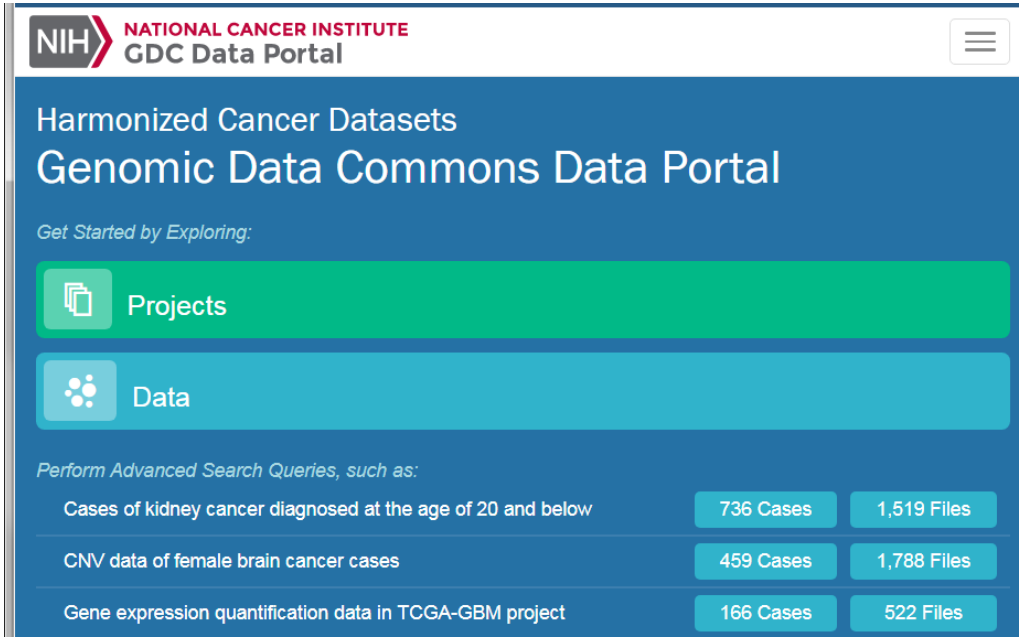
A *module* is a list of target genes of a TF, list of genes composing a process, etc.

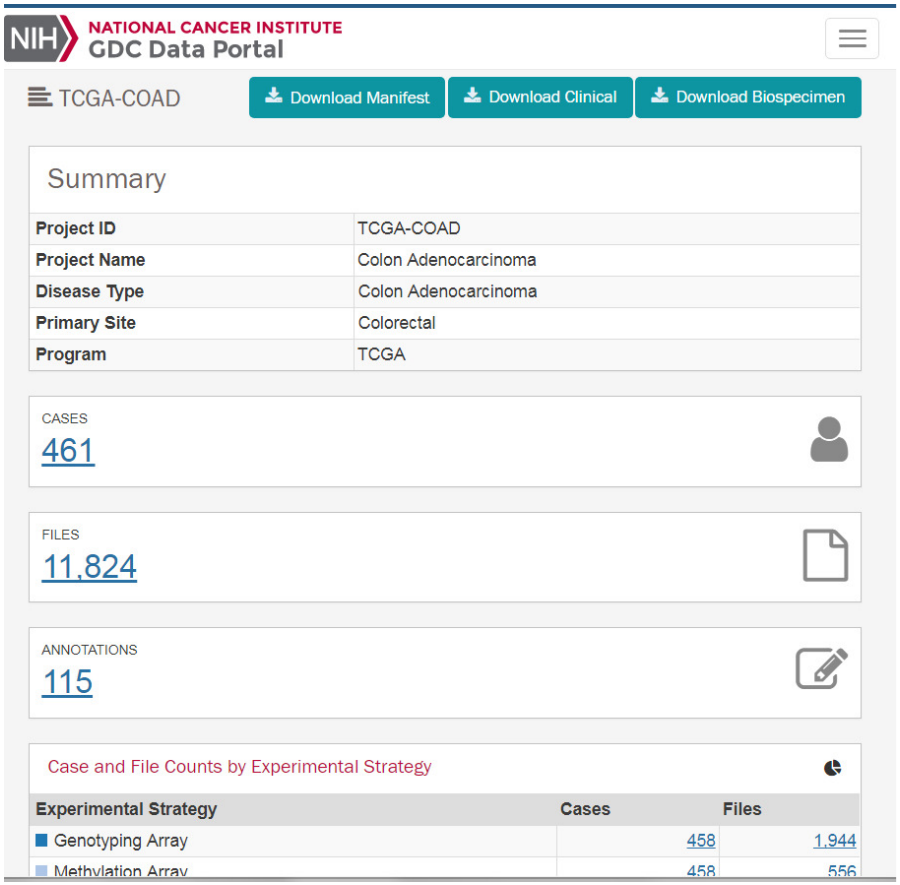
Example of response to cetuximab (EGFR inhibitor) for 8 colon cancer patients

- 4 responders and 4 non responders
- GSE56386 (no paper associated to the data)

Data: Transcriptomics data of colon tumour biopsies

- Colon tumours on TCGA
- 17 metastatic and 88 non-metastatic patients





institutCurie

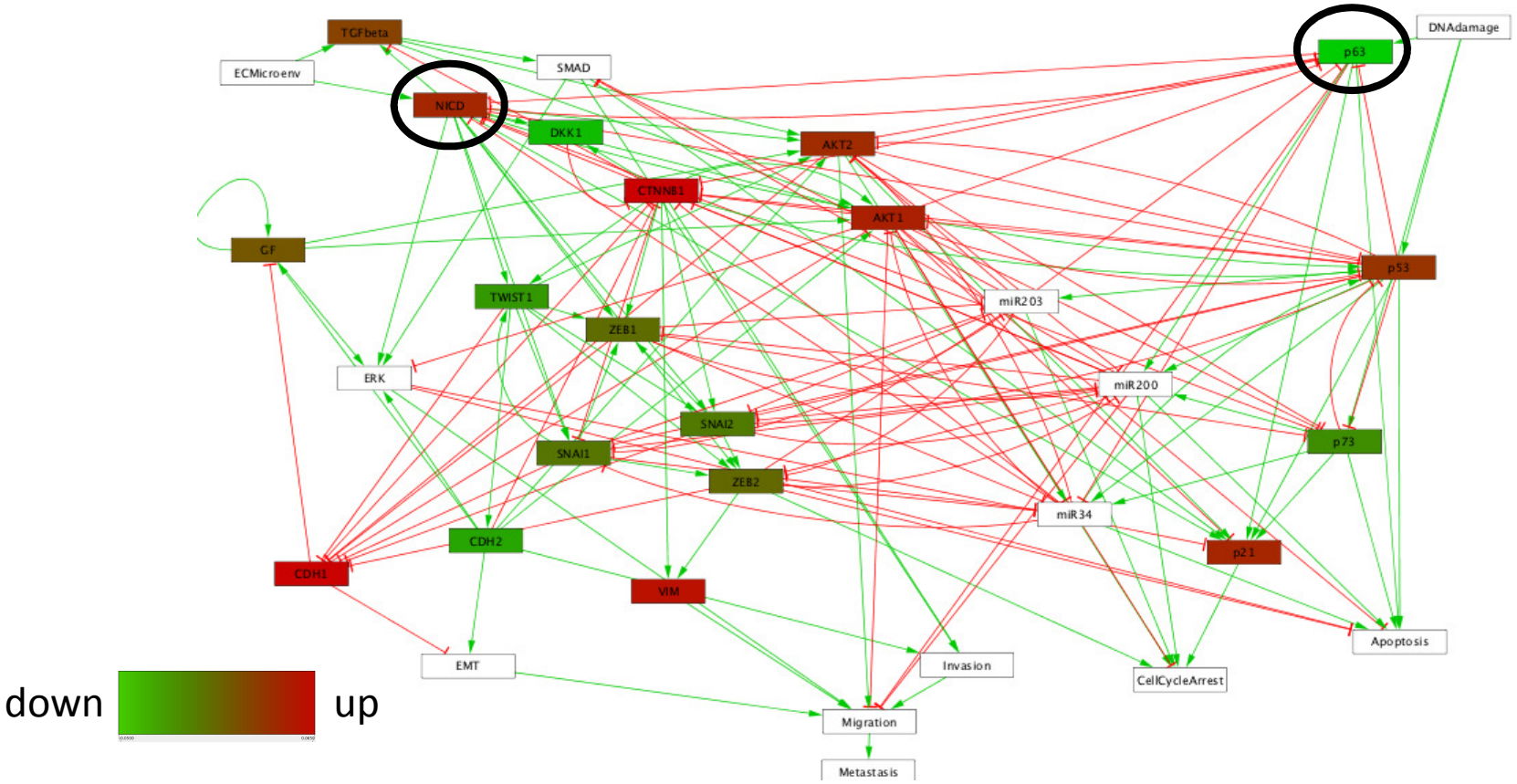
Instituts thématiques Inserm
Institut national de la santé et de la recherche médicale

MINES ParisTech

PSL RESEARCH UNIVERSITY

Gene level

Mean value expression of genes mapped on the network:
17 metastatic and 88 non-metastatic patients



⇒ The figure is very similar for both metastastic and non-metastatic patients

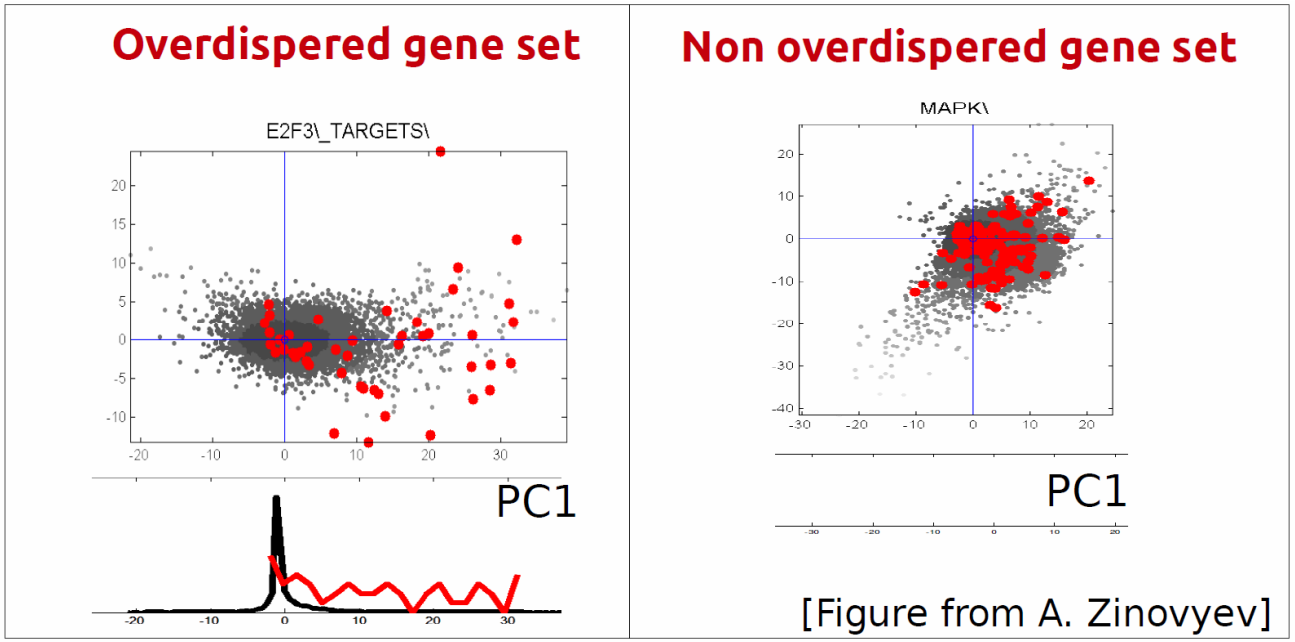
⇒ No obvious differences at the transcriptomics level for Notch and p53

ROMA

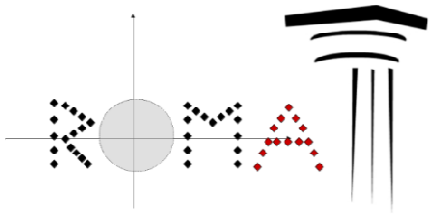
Martignetti et al, Front Genet. 2016
https://github.com/sysbio-curie/Roma

- ROMA: Representation Of Module Activity

- The main idea behind ROMA is:
  - to define a metagene that captures the largest amount of variance
  - to explore the activity of sets of genes (modules) rather than individual genes across samples explained by the genes in the module



[Figure from A. Zinovyev]



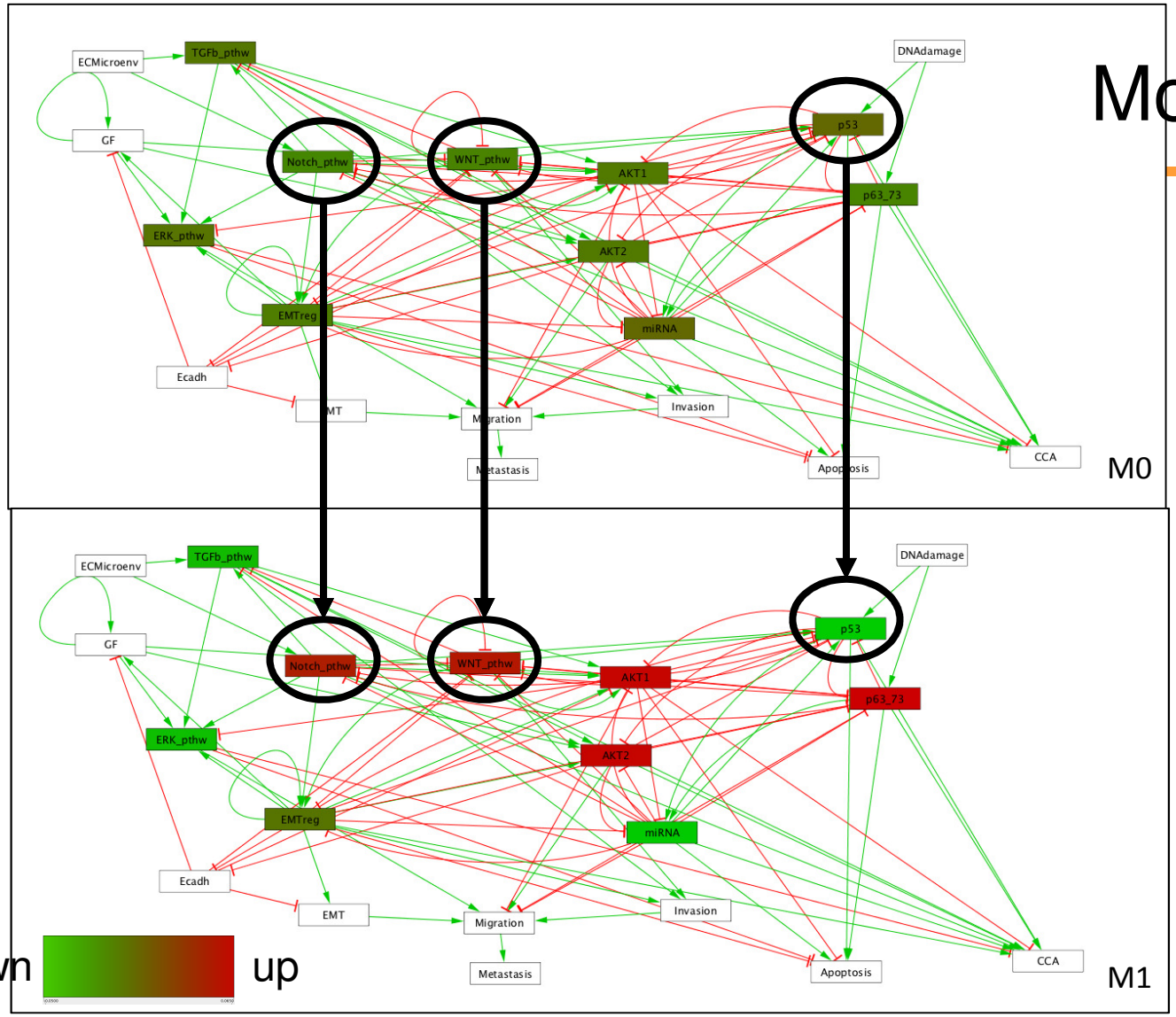
ROMA

Martignetti et al, Front Genet. 2016
https://github.com/sysbio-curie/Roma

- Gene set: set of genes with a functional relationship
  - ACSN signalling pathways
  - KEGG metabolic pathways
  - Can have weights and sign
- The data is not analysed per gene but per gene-set
- In this case, gene-set is a module and its genes

- KEGG_CITRATE_CYCLE_TCA_CYCLE: IDH3B, DLST, PCK2, CS, PDHB, PCK1, PDHA1, LOC642502, PDHA2, LOC283398, FH, SDHD, OGDH, SDHB, IDH3A, SDHC, IDH2, IDH1, ACO1, ACLY, MDH2, DLD, MDH1, DLAT, OGDHL, PC, SDHA, SUCLG1, SUCLA2, SUCLG2, IDH3G, ACO2
- G3-Kinases: CSNK2A1[18.09], CDK1[11.76], PRKDC[9.95], GSK3B[9.50], AURKA[6.33], ADRBK1[4.52], HIPK2[4.52], MAPK3[4.52], MAPK1[3.61], AKT1[2.71], CLK1[2.71], ATM[2.26], TGFBR2[2.26], TTK[2.26], CDK4[1.8], CSNK2A2[1.8], PRKCA[1.8], ATR[1.35], CDK2[1.35], CDK5[1.35], DMPK[1.35], EIF2AK2[1.35], GSK3A[1.35]

Module level



Colon tumour data

Modules are the result of the model reduction

Activity of each module = sum of the expression of genes

What about EMT?

- EMT transient
- only a small proportion of cells go through EMT

⇒ Search for time series of EMT induction