Freie Universität Berlin

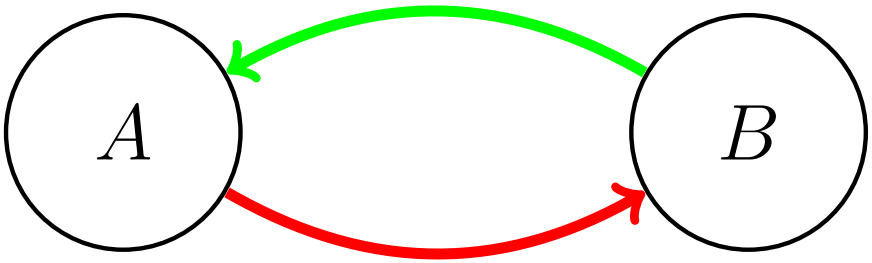
Analysing Cell Line Specific EGFR Signalling via Optimized Automata Based Model Checking

Adam Streck, Kirsten Thobe, Heike Siebert

18.09.2015

Introduction: Boolean Networks

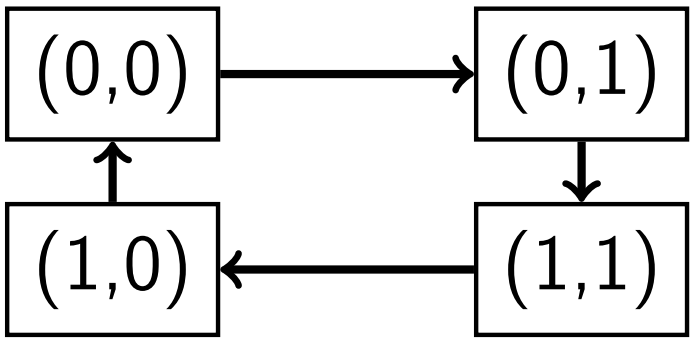
regulatory network:



regulatory functions:

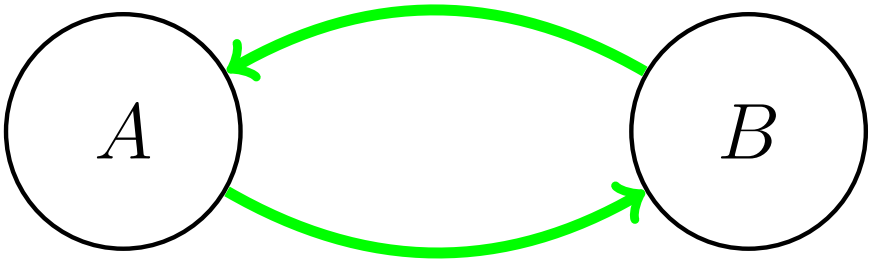
$$A' := B$$
$$B' := \neg A$$

transition system:



Introduction: Parameters

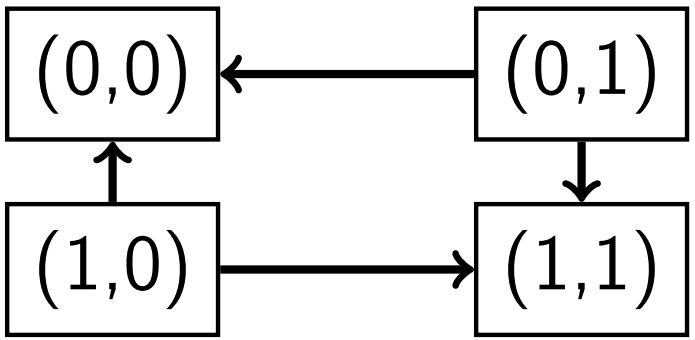
regulatory network:



regulatory functions:

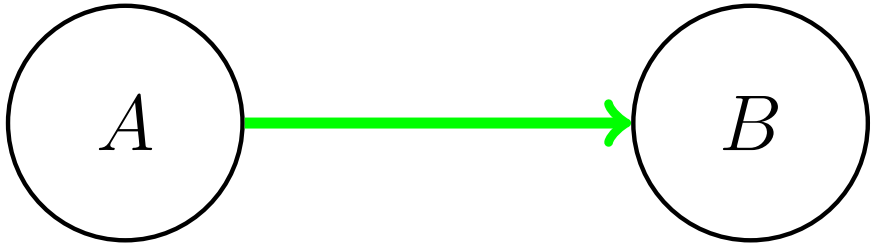
$$A' := B$$
$$B' := A$$

transition system:



Introduction: Functional edges

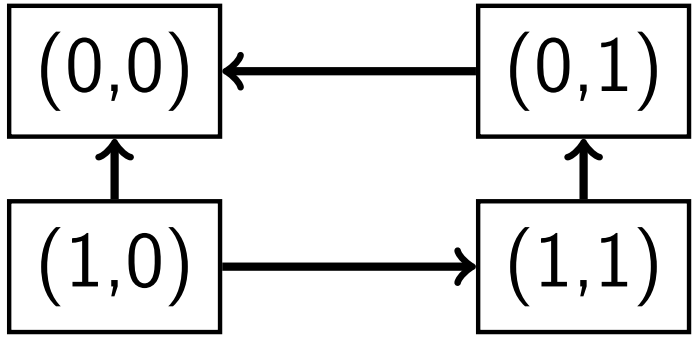
regulatory network:



regulatory functions:

$$A' := 0$$
$$B' := A$$

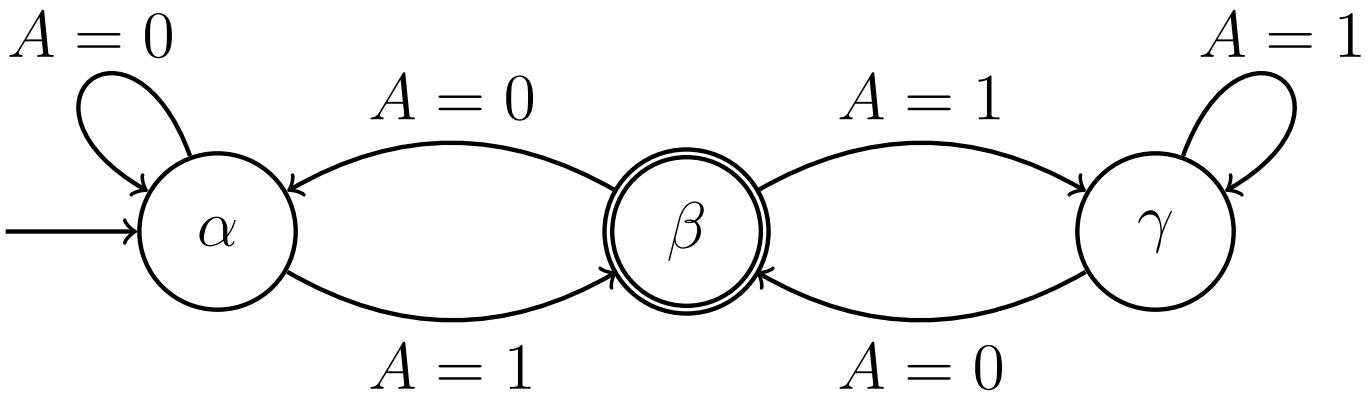
transition system:



Introduction: LTL Model Checking

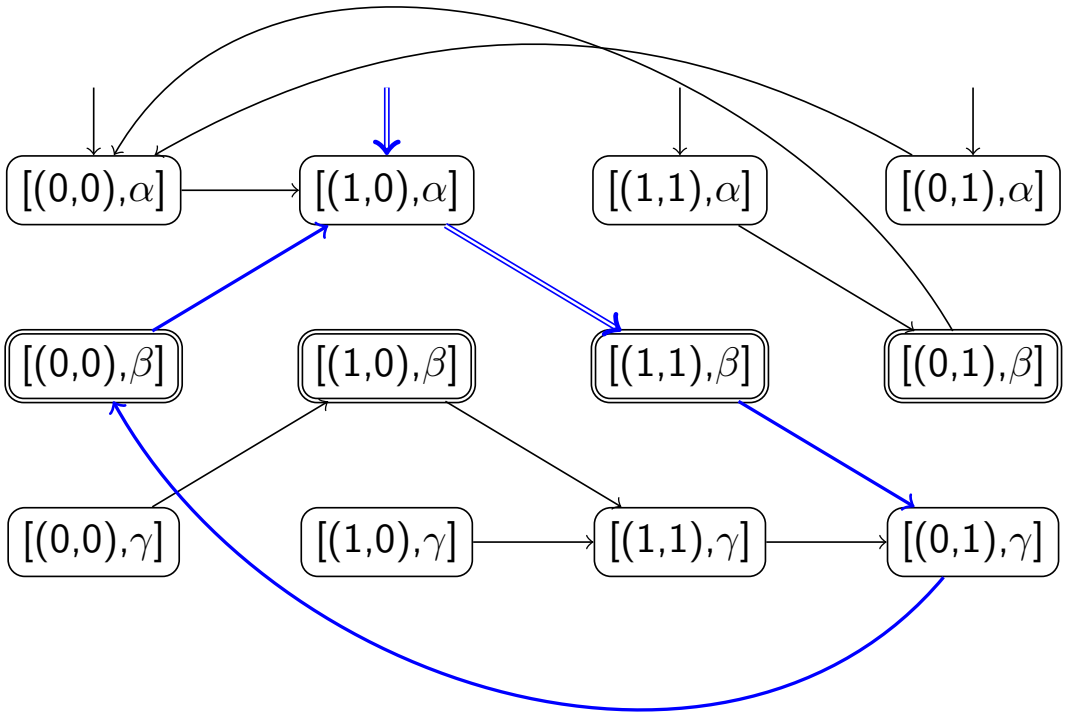
Linear temporal logic (LTL) as a basis for property specification.

The LTL formula: $\Phi \equiv G(F(A = 1) \wedge F(A = 0))$
translates into the Büchi automaton (BA):



Introduction: Automaton Based MC

The TS and the BA are combined into the *synchronous product*:



In blue is a *witness* of satisfaction of Φ.

Methods: Motivation

- The MC procedure is extremely resource-heavy:

Methods: Motivation

- The MC procedure is extremely resource-heavy:
  - The size of a transition system is *exponential* w.r.t. the number of components.

Methods: Motivation

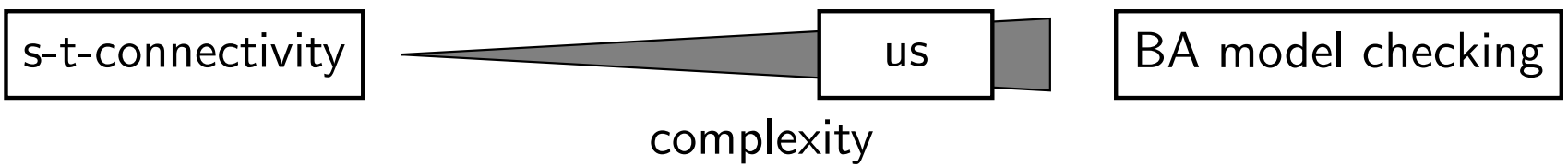
- The MC procedure is extremely resource-heavy:
  - The size of a transition system is *exponential* w.r.t. the number of components.
  - The automaton is *exponential* w.r.t. the size of the formula.

Methods: Motivation

- The MC procedure is extremely resource-heavy:
  - The size of a transition system is *exponential* w.r.t. the number of components.
  - The automaton is *exponential* w.r.t. the size of the formula.
  - Under parameter uncertainty, the number of possible regulatory functions is *more than double exponential* w.r.t. the number of components.

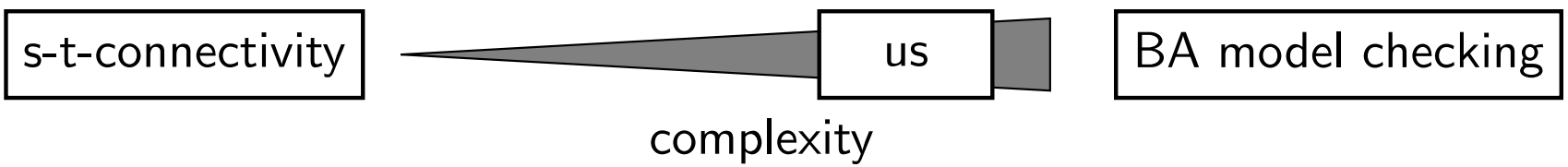
Methods: Motivation

- The MC procedure is extremely resource-heavy:
  - The size of a transition system is *exponential* w.r.t. the number of components.
  - The automaton is *exponential* w.r.t. the size of the formula.
  - Under parameter uncertainty, the number of possible regulatory functions is *more than double exponential* w.r.t. the number of components.
- Aim: Utilize the structure of the problem to simplify the method:



Methods: Motivation

- The MC procedure is extremely resource-heavy:
  - The size of a transition system is *exponential* w.r.t. the number of components.
  - The automaton is *exponential* w.r.t. the size of the formula.
  - Under parameter uncertainty, the number of possible regulatory functions is *more than double exponential* w.r.t. the number of components.
- Aim: Utilize the structure of the problem to simplify the method:



- A custom model checker necessary.

Methods (1/3): Experimental Set-up

- Simple idea: If a component is stimulated or inhibited from the beginning, remove it from the transition system.

Methods (1/3): Experimental Set-up

- Simple idea: If a component is stimulated or inhibited from the beginning, remove it from the transition system.
- Values of fixed components are kept as constants in a lookup table.

Methods (1/3): Experimental Set-up

- Simple idea: If a component is stimulated or inhibited from the beginning, remove it from the transition system.
- Values of fixed components are kept as constants in a lookup table.
- Reduction on the number of components.

Methods (2/3): Measurement Series

Freie Universität Berlin

- Most of the properties are coming from the measurements (microarray, immunoblotting,...).

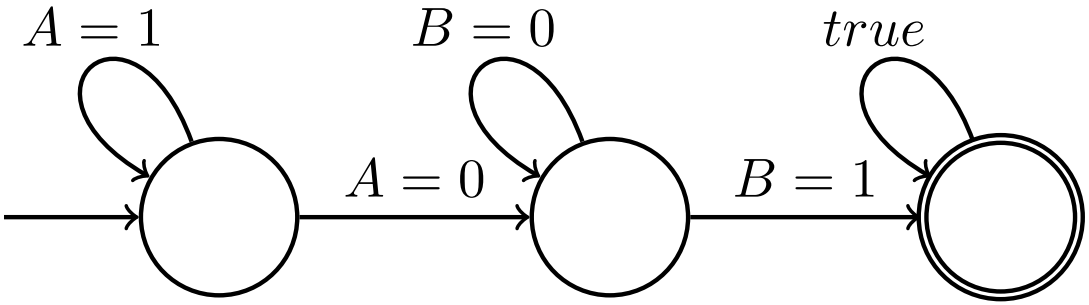
Methods (2/3): Measurement Series

Freie Universität Berlin

- Most of the properties are coming from the measurements (microarray, immunoblotting,...).
- Consider a sequence: First $A = 0$ and then $B = 1$.

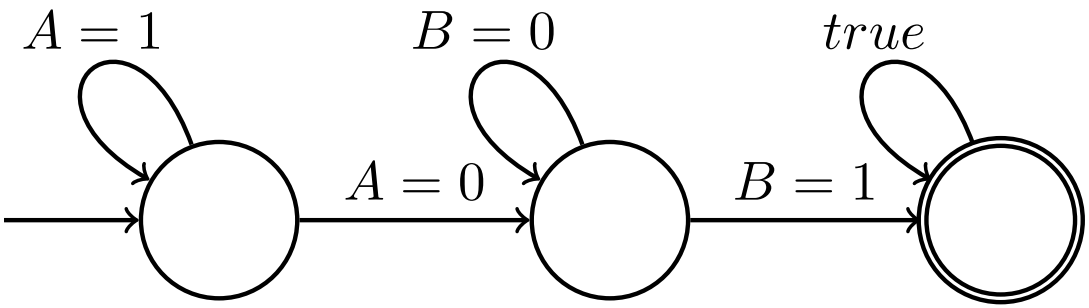
Methods (2/3): Measurement Series

- Most of the properties are coming from the measurements (microarray, immunoblotting,...).
- Consider a sequence: First $A = 0$ and then $B = 1$.
- Measurement series are guaranteed to be linear in size:



Methods (2/3): Measurement Series

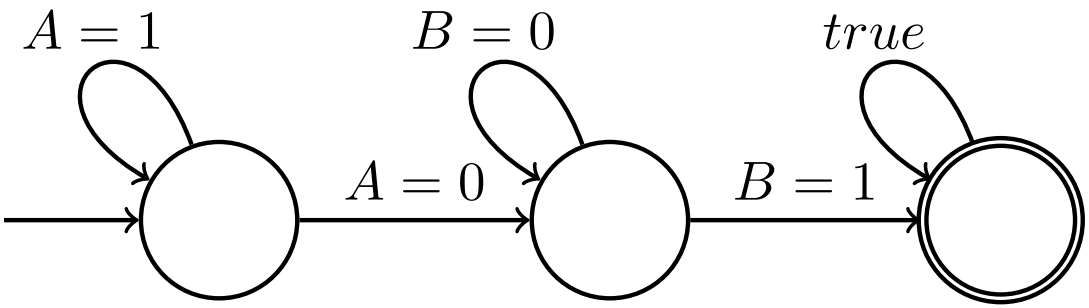
- Most of the properties are coming from the measurements (microarray, immunoblotting,...).
- Consider a sequence: First $A = 0$ and then $B = 1$.
- Measurement series are guaranteed to be linear in size:



- We can put first and last measurement into predicates over the transition system $\Rightarrow$ size is $1 \Rightarrow$ problem reduced to *s-t-connectivity*.

Methods (2/3): Measurement Series

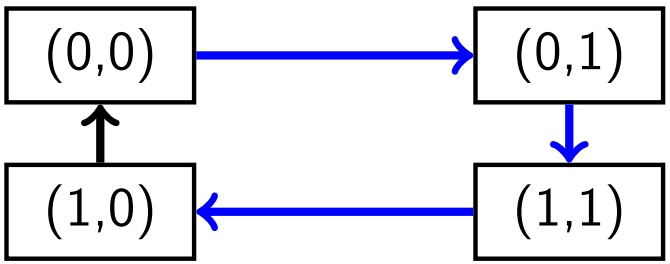
- Most of the properties are coming from the measurements (microarray, immunoblotting,...).
- Consider a sequence: First $A = 0$ and then $B = 1$.
- Measurement series are guaranteed to be linear in size:



- We can put first and last measurement into predicates over the transition system $\Rightarrow$ size is $1 \Rightarrow$ problem reduced to *s-t-connectivity*.
- Also applicable to cycles and stability requirements.

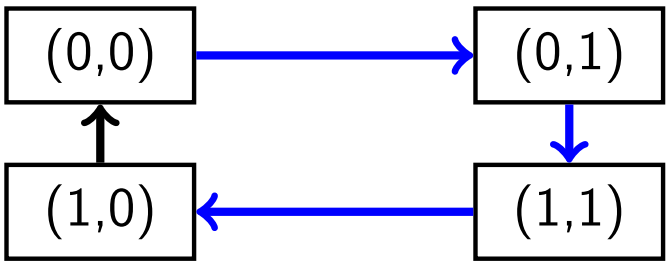
Methods (3/3): One-sided Monotonicity

- A requirement that a component behaves monotonously between two measurements. E.g. B is not monotonous between $(0, 0)$ and $(1, 0)$:



Methods (3/3): One-sided Monotonicity

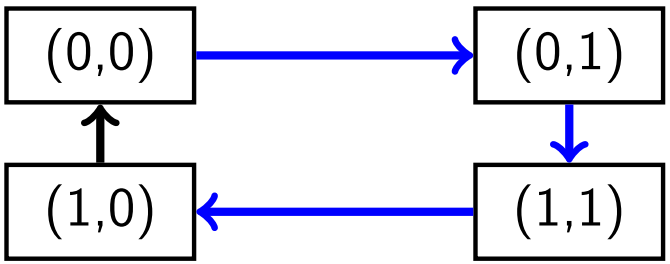
- A requirement that a component behaves monotonously between two measurements. E.g. B is not monotonous between $(0, 0)$ and $(1, 0)$:



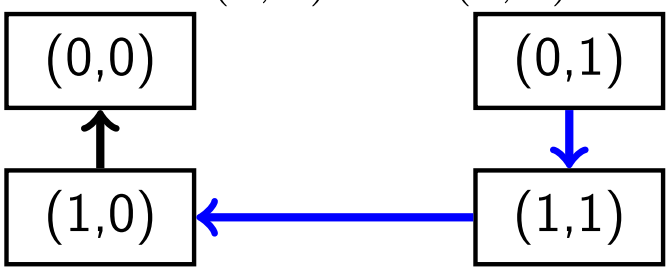
- Hard to LTL encode.

Methods (3/3): One-sided Monotonicity

- A requirement that a component behaves monotonously between two measurements. E.g. B is not monotonous between $(0, 0)$ and $(1, 0)$:

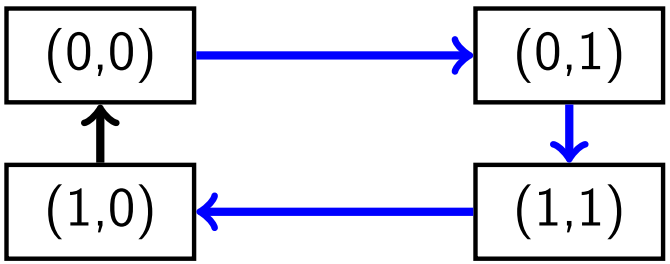


- Hard to LTL encode.
- One-sided monotonicity easy on the transition system—just remove transitions in the wrong direction. E.g. B is monotonously decreasing between $(0, 0)$ and $(1, 0)$:

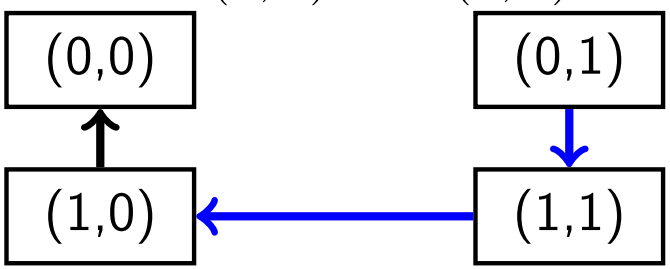


Methods (3/3): One-sided Monotonicity

- A requirement that a component behaves monotonously between two measurements. E.g. B is not monotonous between $(0, 0)$ and $(1, 0)$:



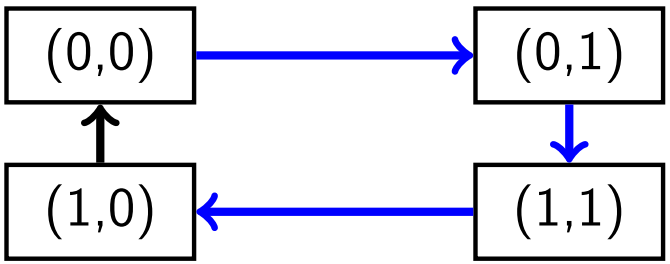
- Hard to LTL encode.
- One-sided monotonicity easy on the transition system—just remove transitions in the wrong direction. E.g. B is monotonously decreasing between $(0, 0)$ and $(1, 0)$:



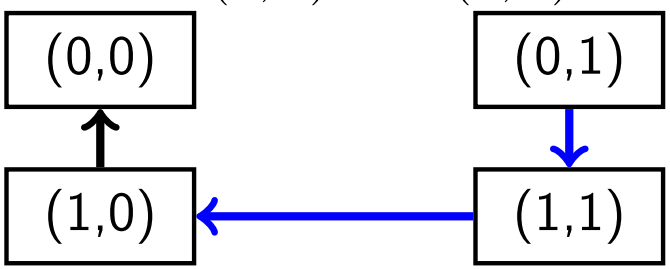
- One-sided usually sufficient.

Methods (3/3): One-sided Monotonicity

- A requirement that a component behaves monotonously between two measurements. E.g. B is not monotonous between $(0, 0)$ and $(1, 0)$:



- Hard to LTL encode.
- One-sided monotonicity easy on the transition system—just remove transitions in the wrong direction. E.g. B is monotonously decreasing between $(0, 0)$ and $(1, 0)$:



- One-sided usually sufficient.
- In a similar manner we can enforce component stability.

Case Study: Colorectal Cancer Study

- Epidermal growth factor receptor pathway.

Case Study: Colorectal Cancer Study

- Epidermal growth factor receptor pathway.
- Human colorectal cancer data, 5 cell-lines.

Case Study: Colorectal Cancer Study

Freie Universität Berlin

- Epidermal growth factor receptor pathway.
- Human colorectal cancer data, 5 cell-lines.
- Data from "Network quantification of EGFR signaling unveils potential for targeted combination therapy", Klinger et al., Molecular Systems Biology 9:673, 2013

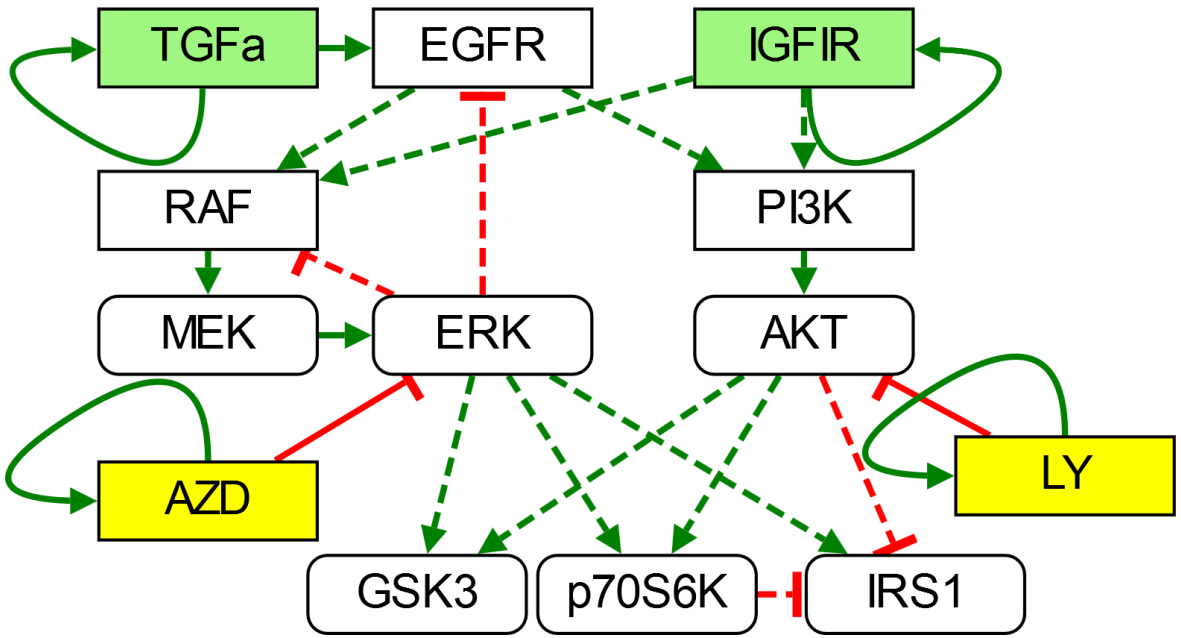
Case Study: Colorectal Cancer Study

Freie Universität Berlin

- Epidermal growth factor receptor pathway.
- Human colorectal cancer data, 5 cell-lines.
- Data from "Network quantification of EGFR signaling unveils potential for targeted combination therapy", Klinger et al., Molecular Systems Biology 9:673, 2013
- Thanks to Torsten Gross and Nils Blütken for consultations.

Case Study: EGFR Pathway

Research question: Which regulations are disrupted in the individual cell lines?
259200 ways how to parametrize the network.



green edge = activation; red edge = inhibition;
full edge = mandatory regulation; dashed edge = optional edge;
lime component = stimulus; yellow component = inhibitor

Case Study: Data Interpretation

Freie Universität Berlin

- Alltogether 134 measurement series:

Case Study: Data Interpretation

Freie Universität Berlin

- Alltogether 134 measurement series:
  - 5 cell lines.

Case Study: Data Interpretation

Freie Universität Berlin

- Alltogether 134 measurement series:
  - 5 cell lines.
  - 8 experimental set-ups—some were problematic to discretize in some cell lines, not included.

Case Study: Data Interpretation

Freie Universität Berlin

- Alltogether 134 measurement series:
  - 5 cell lines.
  - 8 experimental set-ups—some were problematic to discretize in some cell lines, not included.
  - 4 different levels of strictness: no requirement; monotone; stabilizing; monotone and stabilizing.

Case Study: Data Interpretation

Freie Universität Berlin

- Alltogether 134 measurement series:
  - 5 cell lines.
  - 8 experimental set-ups—some were problematic to discretize in some cell lines, not included.
  - 4 different levels of strictness: no requirement; monotone; stabilizing; monotone and stabilizing.
- Each measurement series has 2 steps.

Case Study: Data Interpretation

Freie Universität Berlin

- Alltogether 134 measurement series:
  - 5 cell lines.
  - 8 experimental set-ups—some were problematic to discretize in some cell lines, not included.
  - 4 different levels of strictness: no requirement; monotone; stabilizing; monotone and stabilizing.
- Each measurement series has 2 steps.
- When a value of a component seemed not to change, the component is required to be stable.

Case Study: Basic Results

Freie Universität Berlin

- For every cell line there are parameters such that all measurement series are fit, no matter the strictness.

Case Study: Basic Results

Freie Universität Berlin

- For every cell line there are parameters such that all measurement series are fit, no matter the strictness.
- About 10 hours on a desktop, less than 10MB of RAM used.

Case Study: Present Regulations

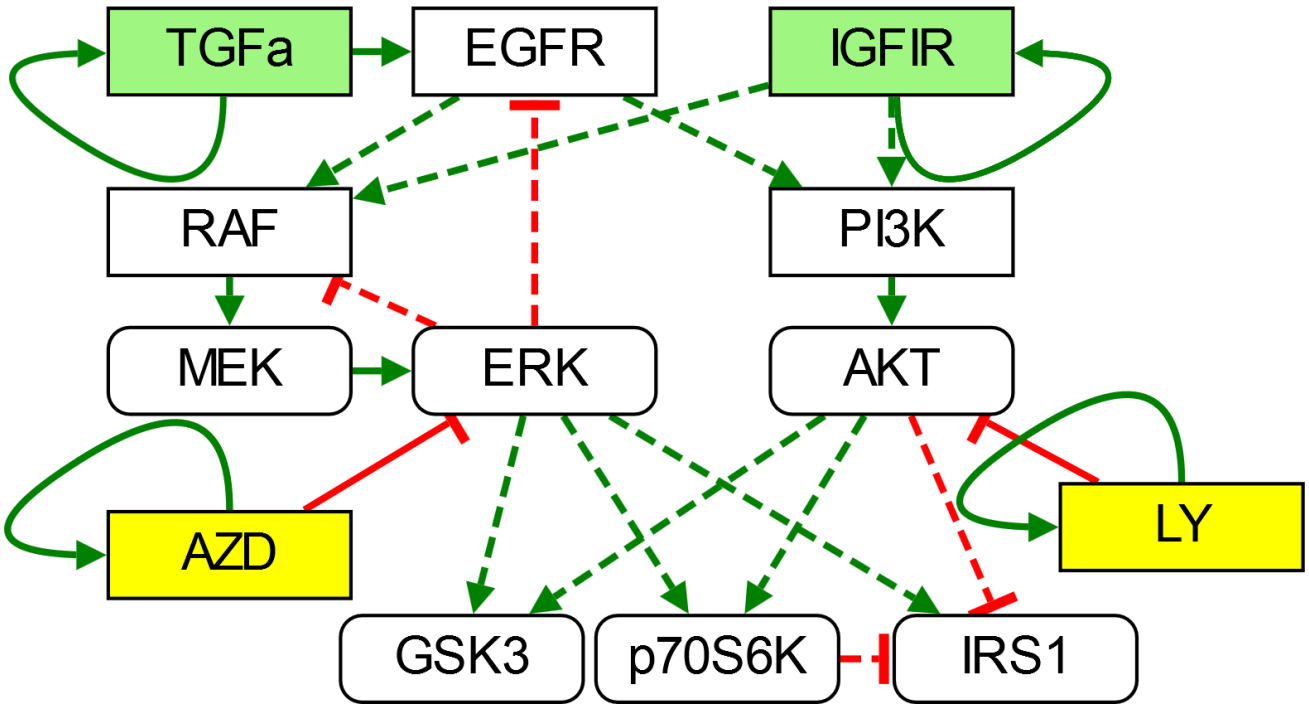
Main focus of Klinger et al.—present regulators under the steady state.

Table for illustration (SW408 cell line):

Target	Must	Klinger et al.	May	Match
EGFR	TGFa	TGFa, ERK	TGFa	no
RAF	∅	EGFR, IGFIR, ERK	EGFR, IGFIR	no
PI3K	EGFR, IGFIR	EGFR, IGFIR	EGFR, IGFIR	total
GSK3	∅	AKT	ERK, AKT	yes
p70S6K	ERK, AKT	ERK, AKT	ERK, AKT	total
IRS1	ERK	p70S6K	ERK, AKT, p70S6K	no

A negative feedback from ERK to RAF and EGFR seems to be a problem.

Case Study: Present Regulators 2



The negative feedback from RAF is a problem.

Edge semantics not included in Klinger et al., possible inconsistency?

Case Study: Cell Line Difference

Research question: what are the differences between the cell lines.

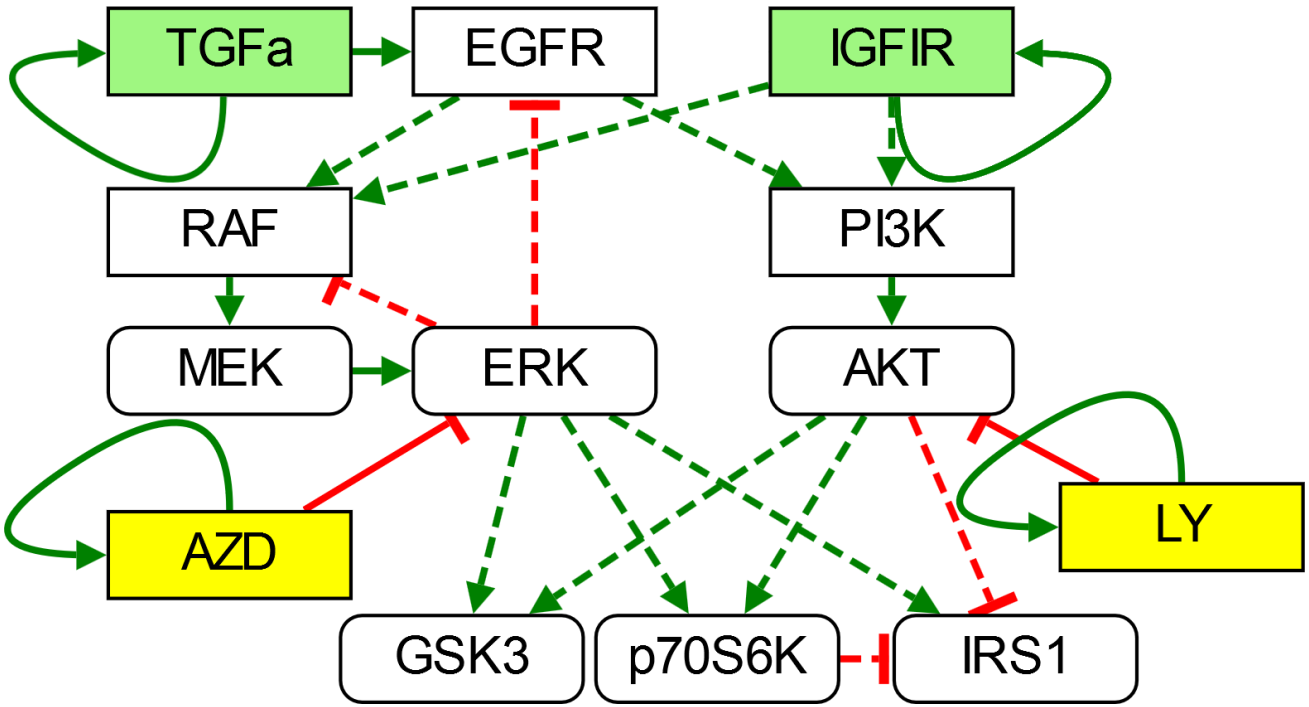
Table for illustration (comparison of cell lines LIM1515 and HCT116):

Target	LIM1215-HCT116
EGFR	differences in frequency
RAF	LIM1215 allows for 15 functions, HCT116 only for $y = 1$
PI3K	strong increase in $y = 1$
GSK3	strong increase in $y = 1$
p70S6K	$y = 1$ appears
IRS1	no difference

LIM1215 suggests that the PI3K cascade stops to respond to the upstream regulation, being constantly active.

HCT116 suggests the same for the RAF cascade, with lower impact on the output components.

Case Study: Cell Line Difference 2



LIM1215 seems to favour a disruption of the PI3K cascade.

HCT116 requires a disruption of the RAF cascade.

Conclusion

- We are usually asking much simpler questions about biological systems than about electrical systems.

Conclusion

- We are usually asking much simpler questions about biological systems than about electrical systems.
- An intervention in the MC process can move the boundary of what is practically computable.

Conclusion

- We are usually asking much simpler questions about biological systems than about electrical systems.
- An intervention in the MC process can move the boundary of what is practically computable.
- Toolset TREMPPI used for analysis. Publicly available (alpha).

Questions?

Thank you for your attention.