

Sur l'apprentissage automatique de modèles mécanistes dynamiques à partir de données temporelles avec application aux chronothérapies personnalisées

Julien Martinelli

The logo for Inria, featuring the word "Inria" in a stylized, red, cursive font.

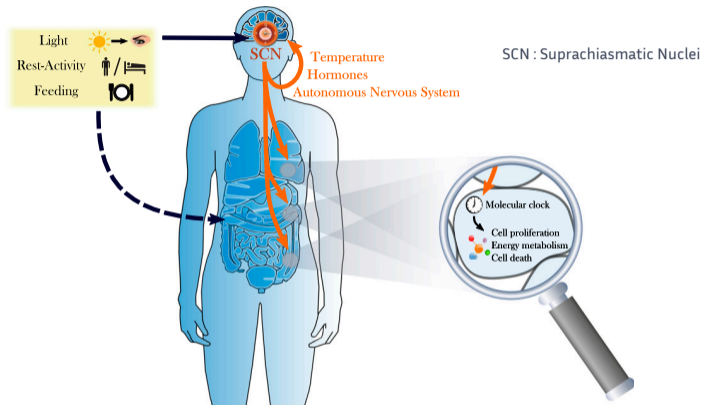
Inserm

La science pour la santé
From science to health



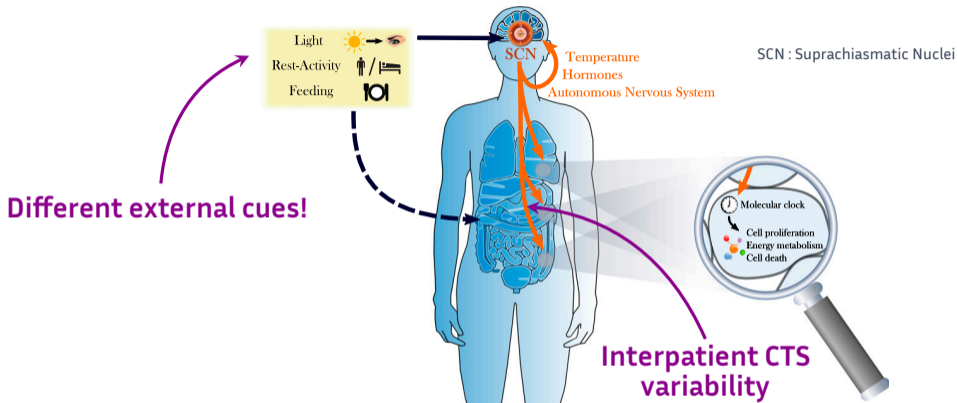
Vendredi 18 Février 2022

The circadian timing system (CTS)



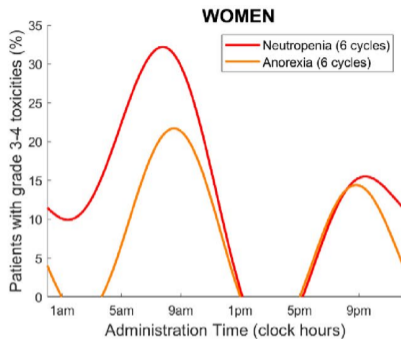
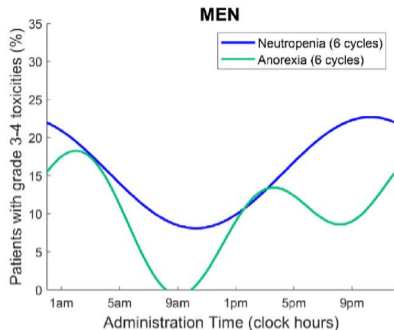
- A master clock acting as an autonomous $\approx 24\text{h}$ -oscillator synchronised by external cues
- This master clock **entrains** the peripheral clocks in the cells *via* physiological signals
- The peripheral clock induces oscillations in key intracellular processes

The circadian timing system (CTS)



- A master clock acting as an autonomous $\approx 24\text{h}$ -oscillator synchronised by external cues
- This master clock **entrains** the peripheral clocks in the cells *via* physiological signals
- The peripheral clock induces oscillations in key intracellular processes

Personalizing chronotherapies

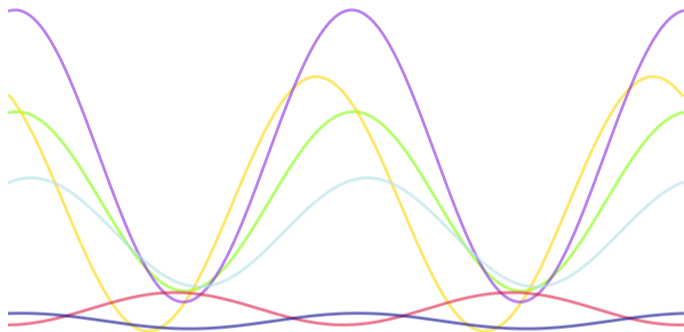


[Innominato *et al.*, *Cancer Medicine*, 2020]

- Multicentric study 193 patients - 67% men
- Metastatic colorectal cancer
- Irinotecan administrated at 6 different times

Large inter-patient variability → Need for personalized optimal timing

Personalizing chronotherapies

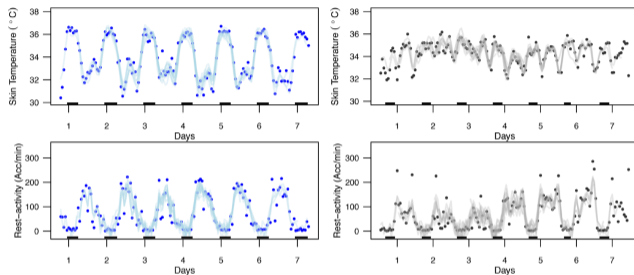
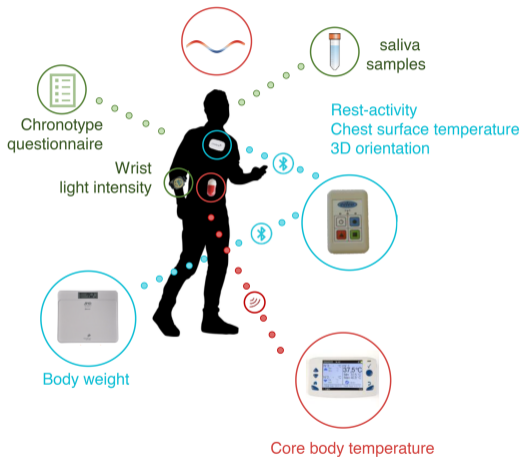


Necessary rhythms of pharmacological proteins required
[Dulong *et al.*, *Mol. Cancer Ther.*, 2015]

→ Data collection too **invasive** and **costly**

Personalizing chronotherapies

Nowadays, access to noninvasive measurements through **wearables data**



Woman, 71 years old (blue)
Man, 34 years old (grey)

Adapted from [Komarzynski *et al.*, *JCI insight*, 2019]

Can chronotherapies be personalized using noninvasive data measurements?

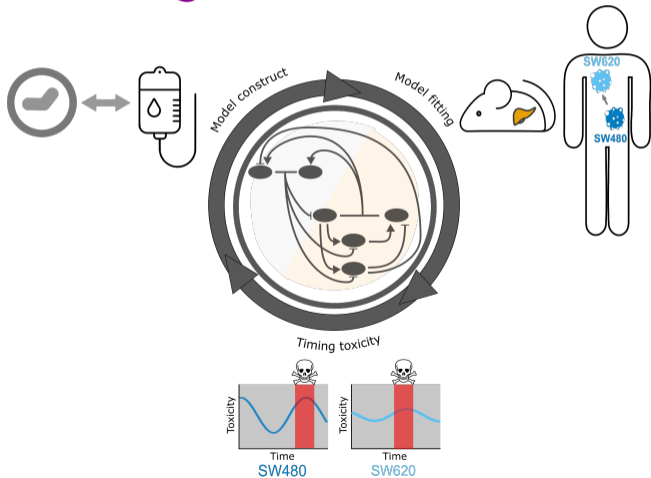
Can chronotherapies be personalized using noninvasive data measurements?

Model learning approaches can be designed to identify unknown interactions

Outline

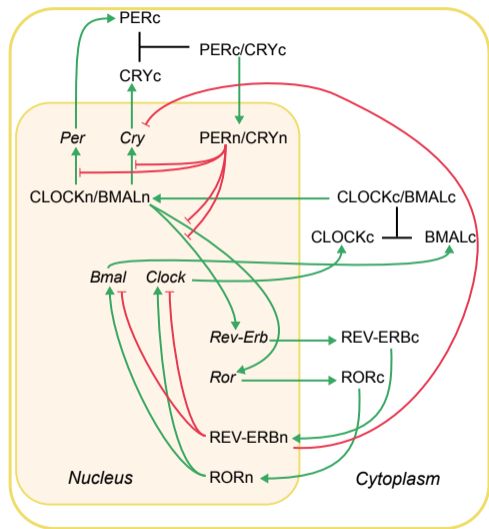
- 1 A mathematical model of the circadian clock and drug pharmacology to optimize irinotecan administration timing in colorectal cancer
- 2 Model learning to identify systemic regulators of the peripheral circadian clock
- 3 Reactmine: an algorithm for inferring biochemical reactions from time series data

Part 1 - A mathematical model of the circadian clock and drug pharmacology to optimize irinotecan administration timing in colorectal cancer



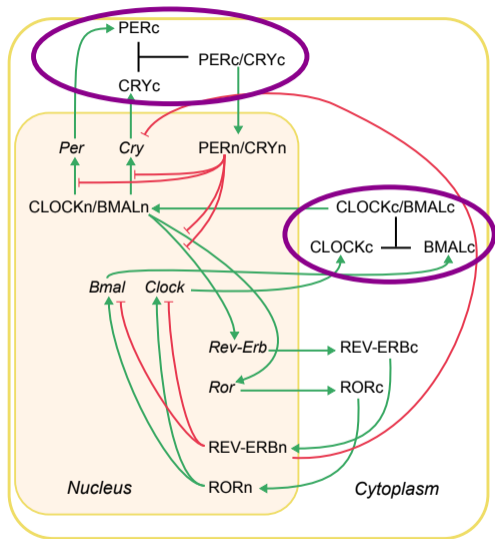
A novel mathematical model of the mammalian peripheral circadian clock

Ordinary Differential Equations (ODE) model
Based on [Relógio *et al.*, *PLoS CB*, 2011]



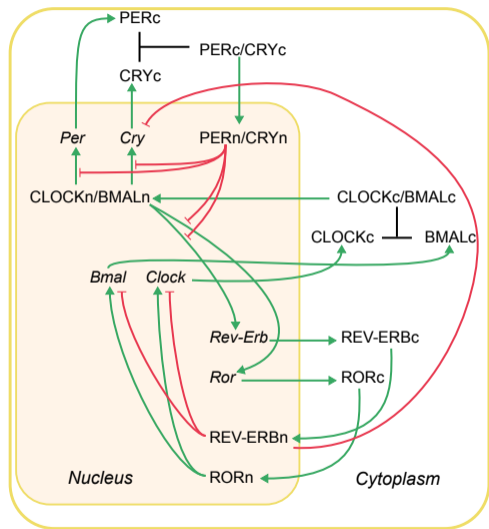
A novel mathematical model of the mammalian peripheral circadian clock

Ordinary Differential Equations (ODE) model
Based on [Relógio *et al.*, *PLoS CB*, 2011]



- Simplification of the PER/CRY loop
- Explicit modeling of CLOCK/BMAL dimerization

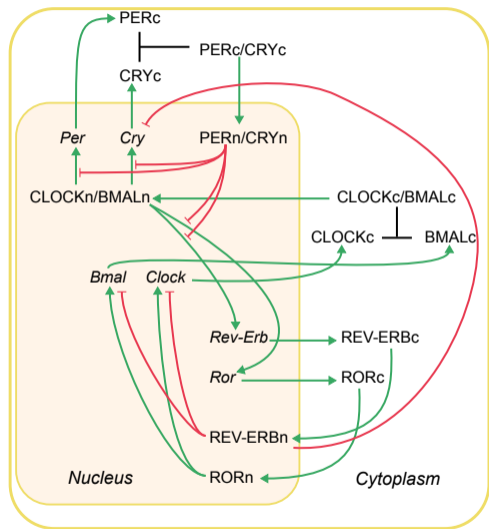
A novel mathematical model of the mammalian peripheral circadian clock



Ordinary Differential Equations (ODE) model
Based on [Relógio *et al.*, *PLoS CB*, 2011]

- 18 variables
- 58 parameters
- Integration of proteomics, genomics and sub cellular data
- **Absolutely quantitative:** mol/L output

A novel mathematical model of the mammalian peripheral circadian clock

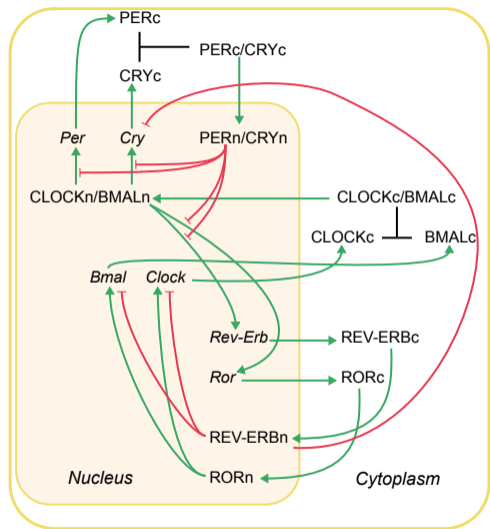


Ordinary Differential Equations (ODE) model
Based on [Relógio *et al.*, *PLoS CB*, 2011]

- 18 variables
- 58 parameters
- Integration of proteomics, genomics and sub cellular data
- **Absolutely quantitative**: mol/L output

$$\frac{dx}{dt} = V_{\max} \text{Transc}(M, \gamma) - \alpha x$$

A novel mathematical model of the mammalian peripheral circadian clock



Ordinary Differential Equations (ODE) model
Based on [Relógio *et al.*, *PLoS CB*, 2011]

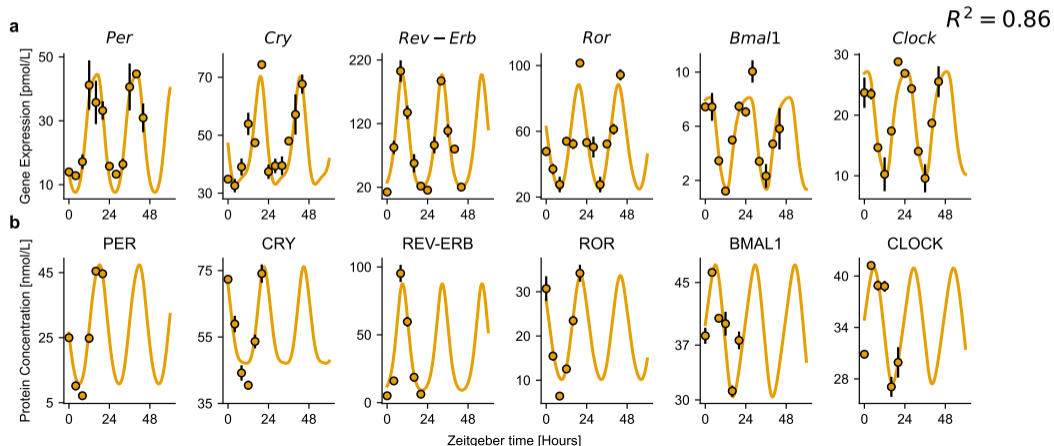
- 18 variables
- 58 parameters
- Integration of proteomics, genomics and sub cellular data
- **Absolutely quantitative:** mol/L output

$$\frac{dx}{dt} = V_{\max} \text{Transc}(M, \gamma) - \alpha x$$

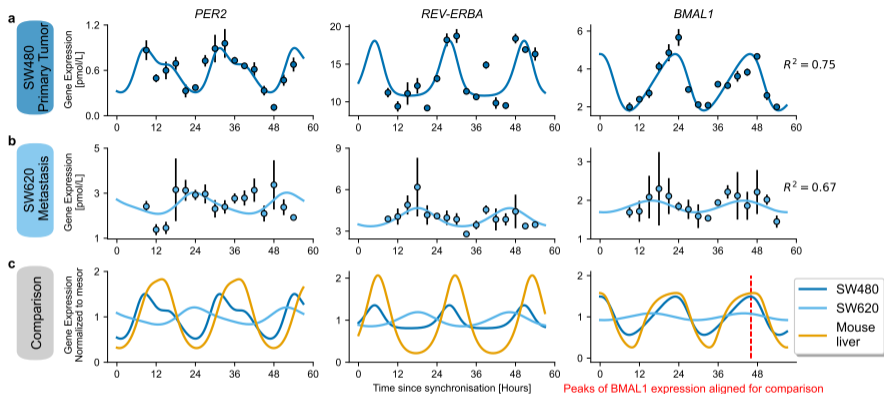
$$\text{Transc}_{Bmal} = \frac{1 + \gamma_1 \left(\frac{ROR}{\gamma_2}\right)^{\gamma_3}}{1 + \left(\frac{REV-ERB}{\gamma_4}\right)^{\gamma_5} + \left(\frac{ROR}{\gamma_2}\right)^{\gamma_3}}$$

Fit of the model on liver mouse data from [Narumi *et al.*, *PNAS*, 2016]

- Parameter fitting using CMA-ES [Hansen *et al.*, *Evolutionary Computation*, 2001].
- Fitness function: least squares
- Biological constraints added to the optimization



Application to human colon cancer cell lines SW480 and SW620

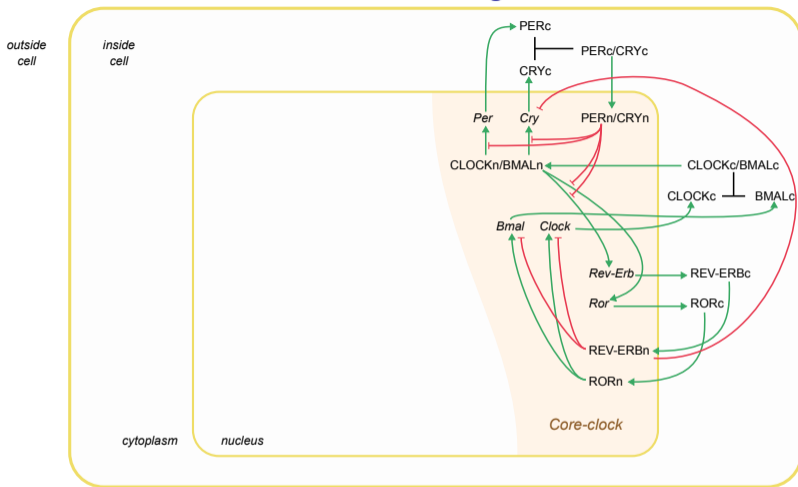


Cells synchronized by medium change

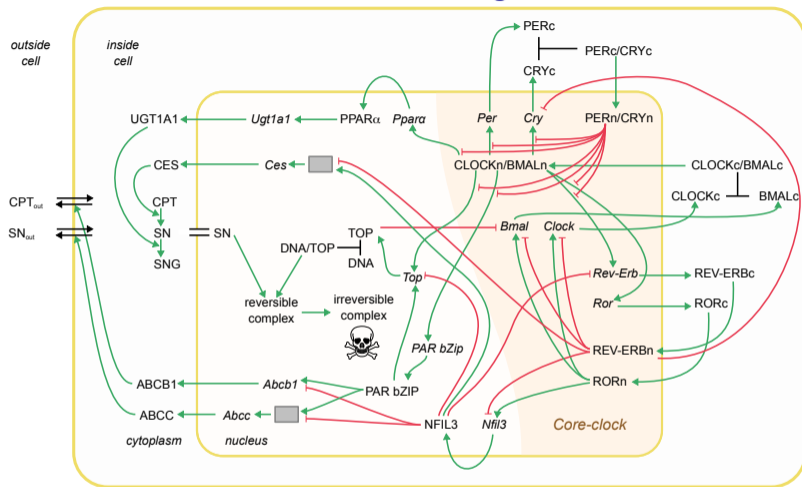
Model refitted to gene expression data from two different cell lines

- SW480 displays well-functioning clock
- SW620 oscillations shifted and dampened compared to liver and SW480

Extension to PK-PD network of anticancer drug irinotecan

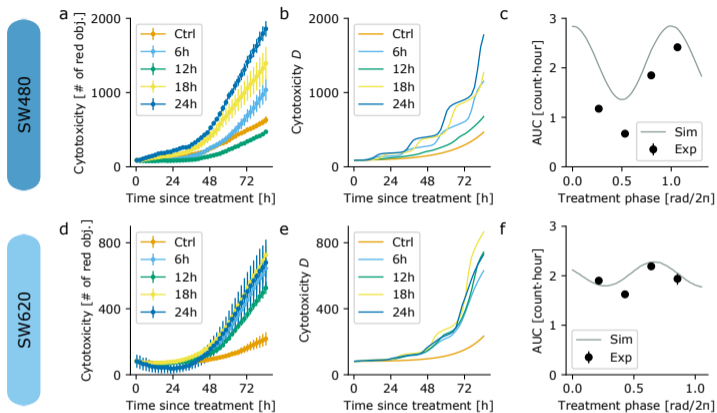


Extension to PK-PD network of anticancer drug irinotecan



- Clock model connected to Irinotecan PK-PD model from [Dulong *et al.*, *MCT*, 2015]
- Explicit modeling of the circadian clock effect instead of forcing functions.

Time-dependent treatment of human cancer cell lines



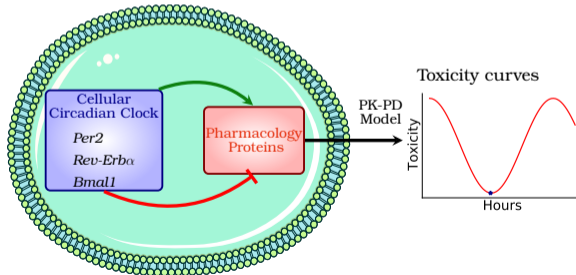
Time aligned to treatment onset $2\mu M$ irinotecan at 4 different circadian times

Dead cells count: cyanine (Incucyte Cytotox Dyes)

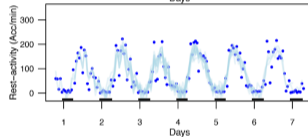
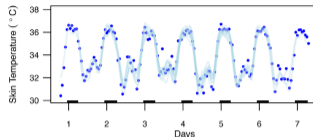
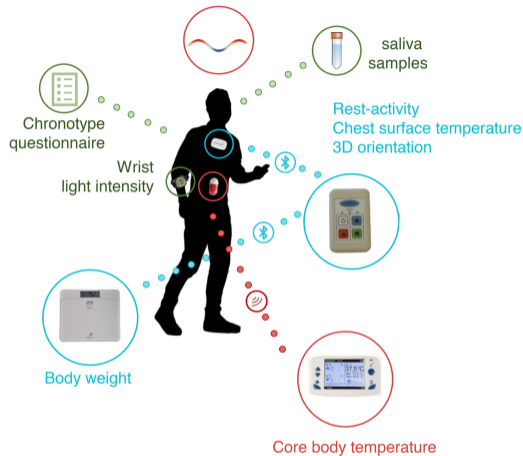
- Phase and amplitude well-captured for SW480, not mean.
- Low time dependency for SW620 toxicity profile

Conclusion - first part

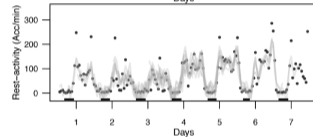
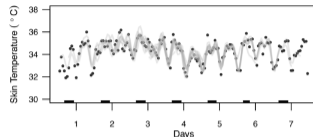
- A novel model of the mammalian circadian clock
 - ▶ Fitted with time-resolved gene expression and protein abundances
 - ▶ Absolutely quantitative
- Successfully connected to the irinotecan PK-PD network
- Enables **personalization** of irinotecan timing from clock and pharmacological mRNAs



Wearable technologies



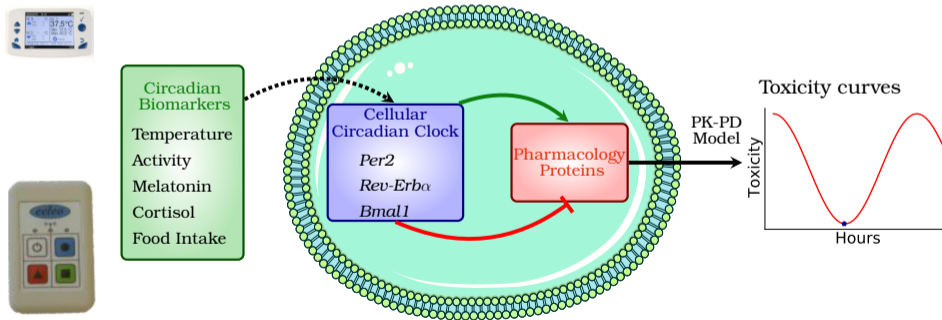
Woman (blue)



Man (grey)

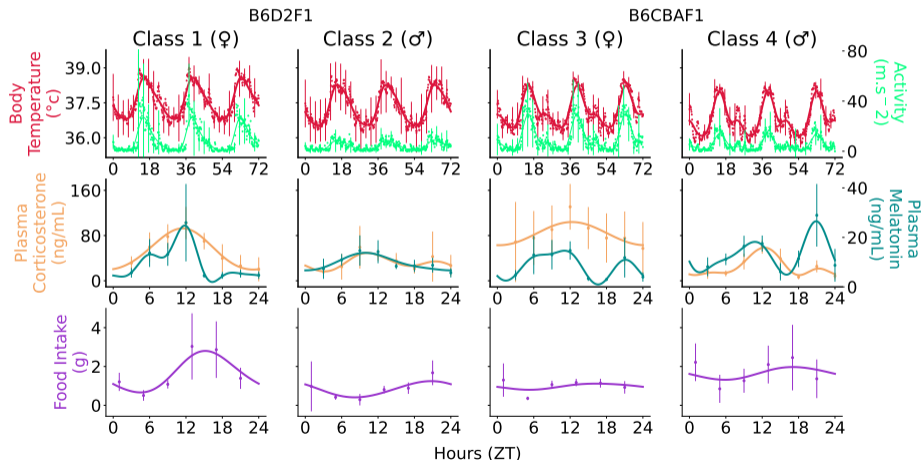
Adapted from [Komarzynski *et al.*, *JCI insight*, 2019]

Part 2 - Model learning to identify systemic regulators of the peripheral circadian clock



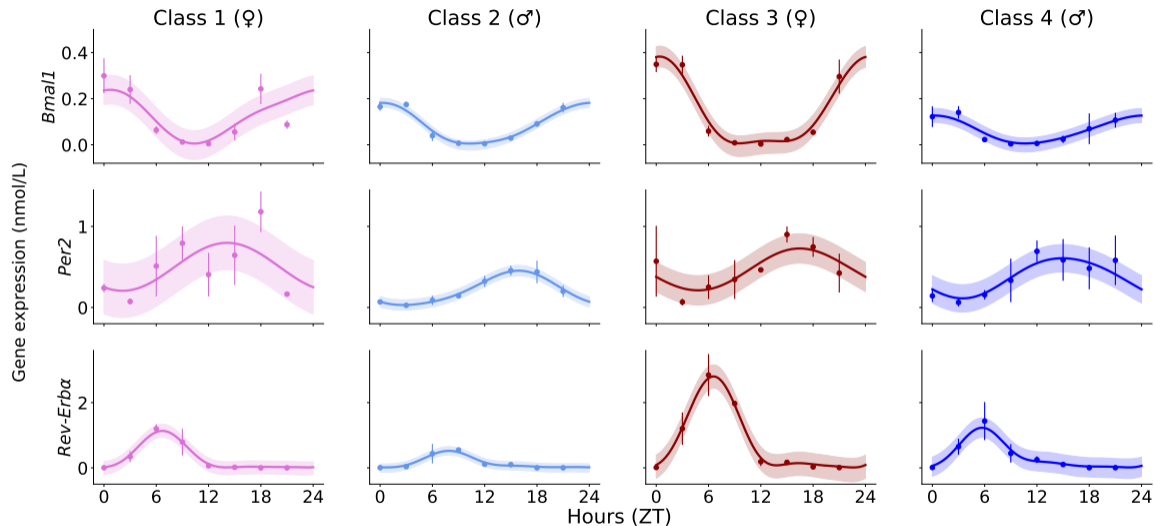
Mouse class systemic regulators data

- 4 mouse classes (2 sex / 2 strains)
- 12-hour light exposure followed by 12-hour darkness



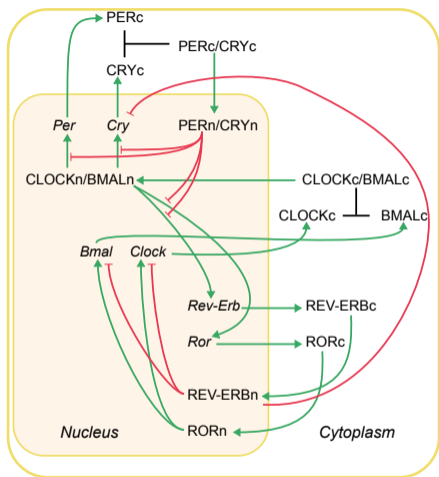
Solid lines: gaussian process regression smoothing.

Mouse class gene expression data (liver)

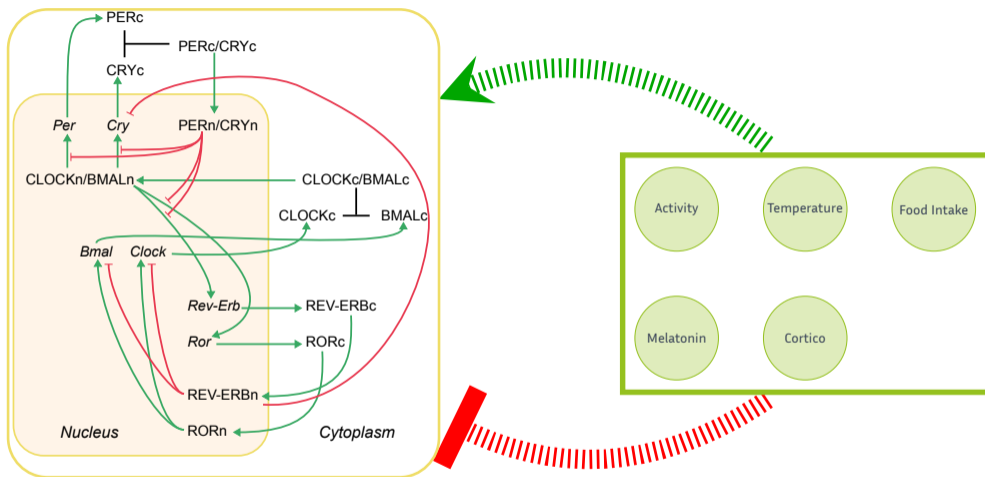


RT-qPCR data. Solid lines and standard deviations: gaussian process regression smoothing.

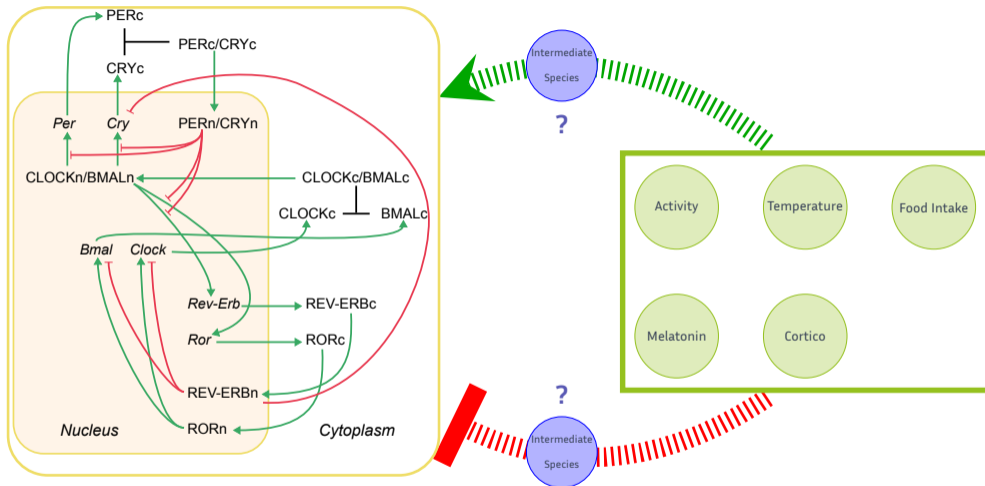
Identifying the action of systemic regulators on the peripheral circadian clock



Identifying the action of systemic regulators on the peripheral circadian clock



Identifying the action of systemic regulators on the peripheral circadian clock



Systemic regulators: z

Intermediate species \rightarrow Integral regulators $\int z$
(obtained after data normalization)

Incorporating systemic regulators action on gene expression

Hypothesis 1: Multiplicative control of systemic regulators z on gene transcription

$$\frac{dx^{vivo}}{dt} = f_{\text{Transc}}(z) V_{\max} \text{Transc}(M, \gamma) - \alpha x^{vivo}$$

Incorporating systemic regulators action on gene expression

Hypothesis 1: Multiplicative control of systemic regulators z on gene transcription

$$\frac{dx^{vivo}}{dt} = f_{\text{Transc}}(z) V_{\max} \text{Transc}(M, \gamma) - \alpha x^{vivo}$$
$$\implies f_{\text{Transc}}(z) \propto \frac{\frac{dx^{vivo}}{dt} + \alpha x^{vivo}}{\text{Transc}(M, \gamma)}$$

Incorporating systemic regulators action on gene expression

Hypothesis 1: Multiplicative control of systemic regulators z on gene transcription

$$\frac{dx^{vivo}}{dt} = f_{\text{Transc}}(z) V_{\max} \text{Transc}(M, \gamma) - \alpha x^{vivo}$$
$$\implies f_{\text{Transc}}(z) \propto \frac{\frac{dx^{vivo}}{dt} + \alpha x^{vivo}}{\text{Transc}(M, \gamma)}$$

Hypothesis 2: Multiplicative control of systemic regulators z on gene mRNA degradation

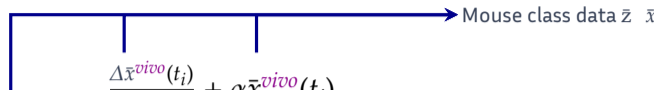
$$\frac{dx^{vivo}}{dt} = V_{\max} \text{Transc}(M, \gamma) - f_{\text{Deg}}(z) \alpha x^{vivo}$$
$$\implies f_{\text{Deg}}(z) \propto \frac{V_{\max} \text{Transc}(M, \gamma) - \frac{dx^{vivo}}{dt}}{x^{vivo}}$$

Data for $x = Bmal1, Per2$ and $Rev-Erb\alpha$ mRNAs

Systemic regulators identification as a regression problem

$$\Leftrightarrow f_{\text{Transc}}(\bar{z}(t_i)) \approx \frac{\frac{\Delta \bar{x}^{\text{vivo}}(t_i)}{\Delta t_i} + \alpha \bar{x}^{\text{vivo}}(t_i)}{\text{Transc}(\mathbf{M}, \gamma)} := y(t_i)$$

Systemic regulators identification as a regression problem



The diagram consists of a horizontal arrow pointing to the right, labeled "Mouse class data $\bar{z} \quad \bar{x}$ ". From the left end of this arrow, a vertical line goes down, then a horizontal line goes to the right, and finally a vertical line goes down to the top of a fraction. This horizontal line is positioned above the fraction's numerator. The fraction is $\frac{\Delta \bar{x}^{vivo}(t_i) + \alpha \bar{x}^{vivo}(t_i)}{\Delta t_i}$. Below the fraction is the text "Transc(M, γ)". To the left of the fraction is the expression $\Leftrightarrow f_{\text{Transc}}(\bar{z}(t_i)) \approx$. To the right of the fraction is the text $:= y(t_i)$.

$$\Leftrightarrow f_{\text{Transc}}(\bar{z}(t_i)) \approx \frac{\Delta \bar{x}^{vivo}(t_i) + \alpha \bar{x}^{vivo}(t_i)}{\Delta t_i} := y(t_i)$$

Learn f_{Transc} using the samples $\left\{ \left(\bar{z}(t_i), y(t_i) \right), i = \{1, \dots, N-1\} \right\}$

Systemic regulators identification as a regression problem

$$\Leftrightarrow f_{\text{Transc}}(\bar{z}(t_i)) \approx \frac{\frac{\Delta \bar{x}^{\text{vivo}}(t_i)}{\Delta t_i} + \alpha \bar{x}^{\text{vivo}}(t_i)}{\text{Transc}(\mathbf{M}, \gamma)} := y(t_i)$$

Learn f_{Transc} using the samples $\{(\bar{z}(t_i), y(t_i)) , i = \{1, \dots, N - 1\}\}$

 Systemic Regulators

Systemic regulators identification as a regression problem

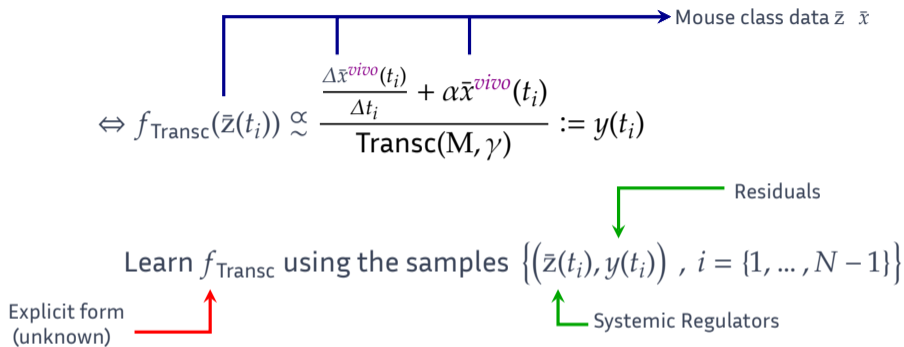
$$\Leftrightarrow f_{\text{Transc}}(\bar{z}(t_i)) \approx \frac{\frac{\Delta \bar{x}^{\text{vivo}}(t_i)}{\Delta t_i} + \alpha \bar{x}^{\text{vivo}}(t_i)}{\text{Transc}(\mathbf{M}, \gamma)} := y(t_i)$$

Learn f_{Transc} using the samples $\left\{ \left(\bar{z}(t_i), y(t_i) \right), i = \{1, \dots, N-1\} \right\}$

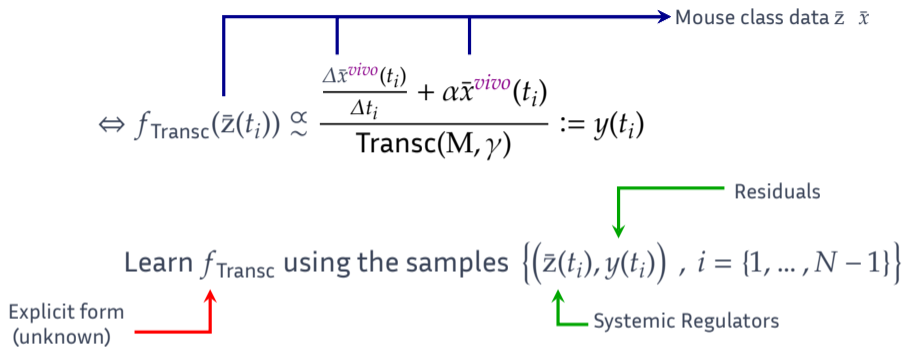
Residuals

Systemic Regulators

Systemic regulators identification as a regression problem



Systemic regulators identification as a regression problem

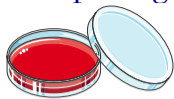


Learning f_{Transc} usually boils down to solve

$$\underset{\hat{f} \in \mathcal{F}}{\text{argmin}} \sum_{i=1}^{N-1} (y(t_i) - \hat{f}(\bar{z}(t_i)))^2$$

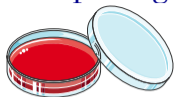
For this study, \mathcal{F} will be the space of linear functions.

Computing y : acquisition of clock parameters and protein levels *in vitro*

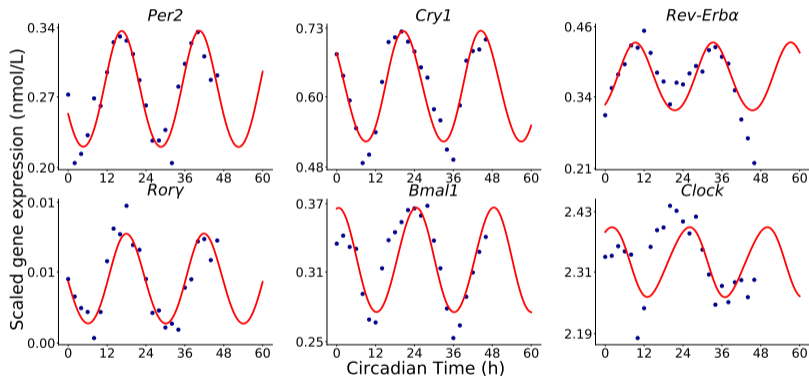


In vitro setting $\implies f_{\text{Transc}}(z)$ constant \implies Fit model on hepatocytes data

Computing y : acquisition of clock parameters and protein levels *in vitro*



In vitro setting $\Rightarrow f_{\text{Transc}}(z)$ constant \Rightarrow Fit model on hepatocytes data



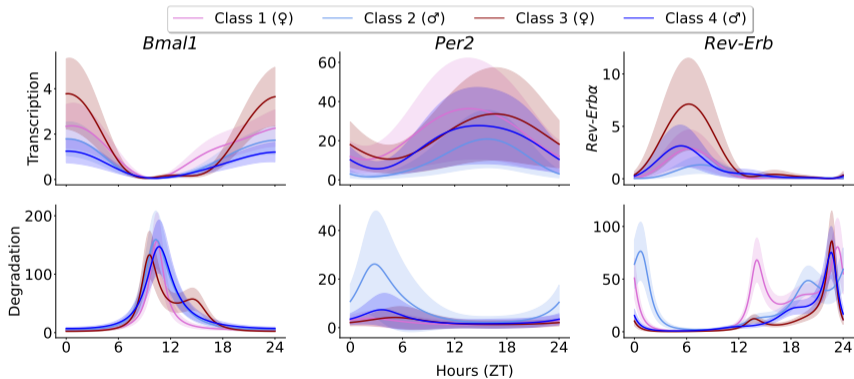
[Atwood *et al.*, PNAS, 2011]

$\rightarrow \alpha, \gamma$ and $M(t)$ estimates obtained $\Rightarrow y(t) = \frac{\frac{\Delta \bar{x}^{\text{vivo}}(t_i)}{\Delta t_i} + \alpha \bar{x}^{\text{vivo}}(t_i)}{\text{Transc}(M, \gamma)}$ can be computed.

Residual trajectories $y(t)$

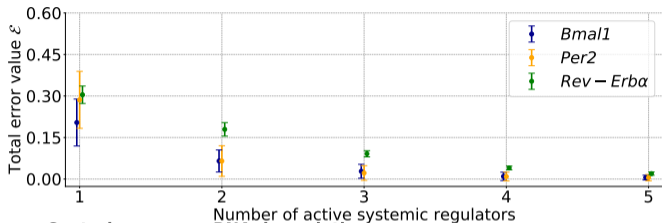
Hypothesis: “*in vivo* clock \approx *in vitro* clock + systemic control + perturbation noise”

→ Perturbations of parameter values to obtain multiple residual trajectories



Total error as a function of the number of involved regulators

Control on gene transcription



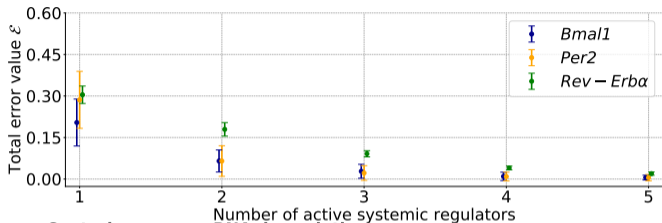
$$\ell(y_k^{(c)}, \bar{z}^{(c)}, \beta_k^{(c)}) := \frac{1}{N-1} \sum_{i=1}^{N-1} \left(y_k^{(c)}(t_i) - \sum_j \beta_{k,j}^{(c)} \bar{z}_j^{(c)}(t_i) \right)^2$$

$$\mathcal{E}(y, \bar{z}) := \frac{1}{n_{class} n_{traj}} \sum_{c=1}^{n_{class}} \sum_{k=1}^{n_{traj}} \min_{\beta_k^{(c)}} \ell(y_k^{(c)}, \bar{z}^{(c)}, \beta_k^{(c)})$$

Input/output normalized $\implies \mathcal{E}$ is an average % of unexplained variance

Total error as a function of the number of involved regulators

Control on gene transcription



- *Bmal1* / *Per2* residuals well fitted with 2-term models, not *Rev-Erbα*
- F-test for nested models concludes on 2-terms

⇒ No linear control of regulators on *Rev-Erbα* transcription

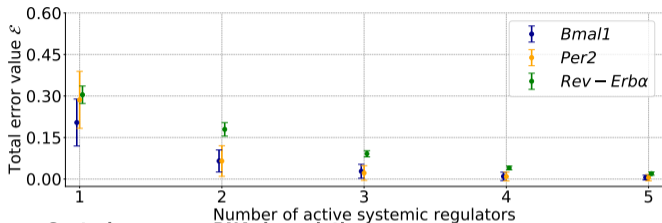
$$\ell(y_k^{(c)}, \bar{z}^{(c)}, \beta_k^{(c)}) := \frac{1}{N-1} \sum_{i=1}^{N-1} \left(y_k^{(c)}(t_i) - \sum_j \beta_{k,j}^{(c)} \bar{z}_j^{(c)}(t_i) \right)^2$$

$$\mathcal{E}(y, \bar{z}) := \frac{1}{n_{class} n_{traj}} \sum_{c=1}^{n_{class}} \sum_{k=1}^{n_{traj}} \min_{\beta_k^{(c)}} \ell(y_k^{(c)}, \bar{z}^{(c)}, \beta_k^{(c)})$$

Input/output normalized ⇒ \mathcal{E} is an average % of unexplained variance

Total error as a function of the number of involved regulators

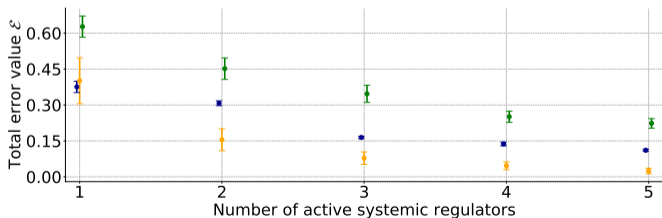
Control on gene transcription



- *Bmal1* / *Per2* residuals well fitted with 2-term models, not *Rev-Erbα*
- F-test for nested models concludes on 2-terms

⇒ No linear control of regulators on *Rev-Erbα* transcription

Control on gene mRNA degradation

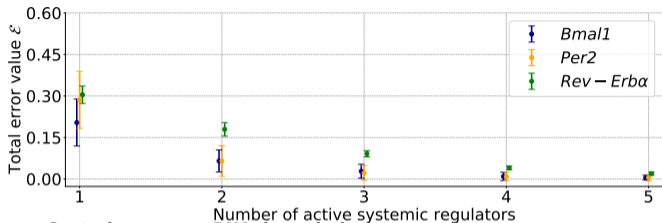


No model with < 3 terms

⇒ No linear control of regulators on gene mRNA degradation

Total error as a function of the number of involved regulators

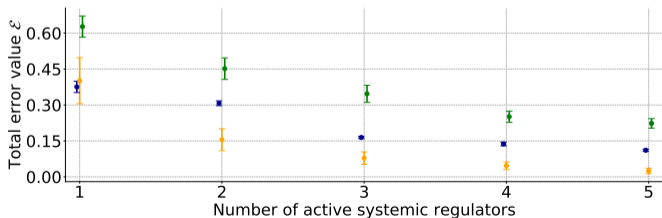
Control on gene transcription



- *Bmal1* / *Per2* residuals well fitted with 2-term models, not *Rev-Erbα*
- F-test for nested models concludes on 2-terms

⇒ No linear control of regulators on *Rev-Erbα* transcription

Control on gene mRNA degradation

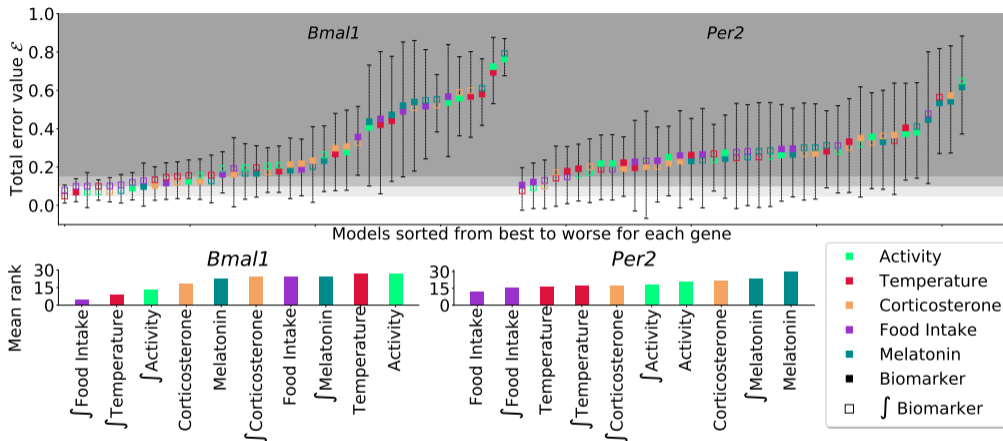


No model with < 3 terms

⇒ No linear control of regulators on gene mRNA degradation

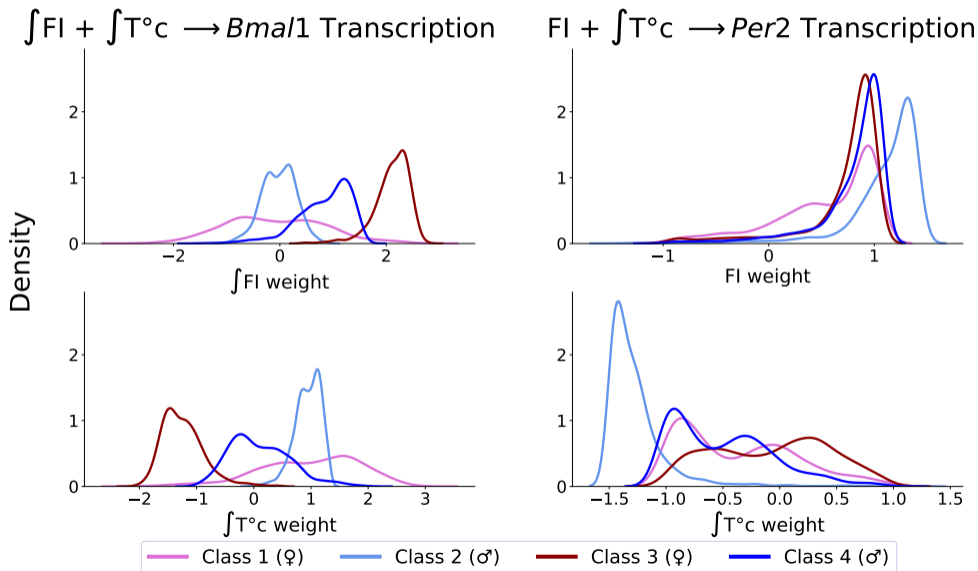
Focus on 2-term models for Transcription: 40 models

2-term models ranking



- Food Intake and Temperature stand out as best models key components.
- Melatonin included as negative control: validation of the approach.

Classwise weights analysis for best 2-term models



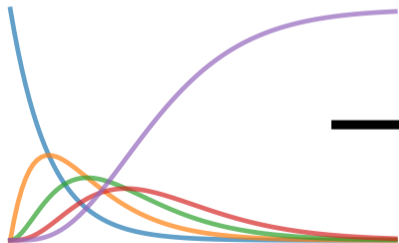
Conclusion - second part

Biological insights and perspectives:

- No realistic control for all 3 genes mRNA degradation & *Rev-Erb α* transcription
 - ▶ Control might not be linear
 - ▶ Non measured systemic variables may be responsible
- Systemic regulators may act on a part of the clock where data is not available
- **Food Intake** and **Temperature** main actors for *Bmal1* and *Per2* transcription
- Statistically significant differences of regulator weights on the basis of genetic background and sex: need for patient stratification

Part 3 - Reactmine: an algorithm for inferring biochemical reactions from time series data

Input: single time series data $Y = (y_{l,i})_{\substack{1 \leq l \leq n \\ 1 \leq i \leq m}}$



Output:
Chemical Reaction Network

Hidden CRN	Learned CRN
$A \xrightarrow{1} B$	$A \xrightarrow{0.999} B$
$B \xrightarrow{1} C$	$B \xrightarrow{1.001} C$
$C \xrightarrow{1} D$	$C \xrightarrow{1.002} D$
$D \xrightarrow{1} E$	$D \xrightarrow{0.999} E$



Framework

Reaction: (R, P, f) with R (resp. P) set of reactants (resp. products) and f rate function.

Chemical Reaction Network (CRN): Finite set of reactions

Framework

Reaction: (R, P, f) with R (resp. P) set of reactants (resp. products) and f rate function.

Chemical Reaction Network (CRN): Finite set of reactions

- 0/1 Stoichiometry
- Elementary reactions: at most two reactants
- At most 1 catalyst

Framework

Reaction: (R, P, f) with R (resp. P) set of reactants (resp. products) and f rate function.

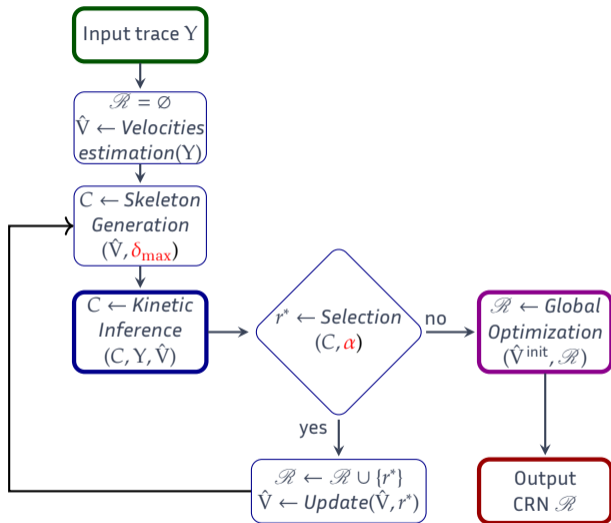
Chemical Reaction Network (CRN): Finite set of reactions

- 0/1 Stoichiometry
- Elementary reactions: at most two reactants
- At most 1 catalyst

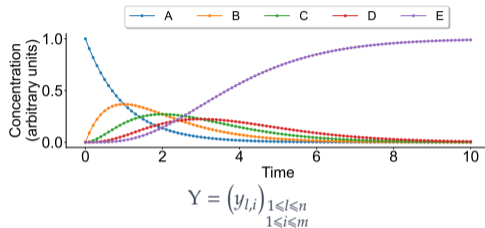
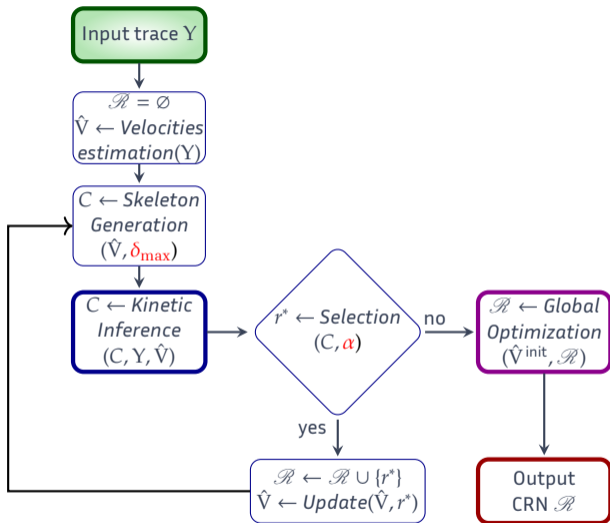
Learning protocol:

- ▶ Learn a CRN involving only observed species
- ▶ Based on a single trace (no combinatorics of initial states and Knockouts)

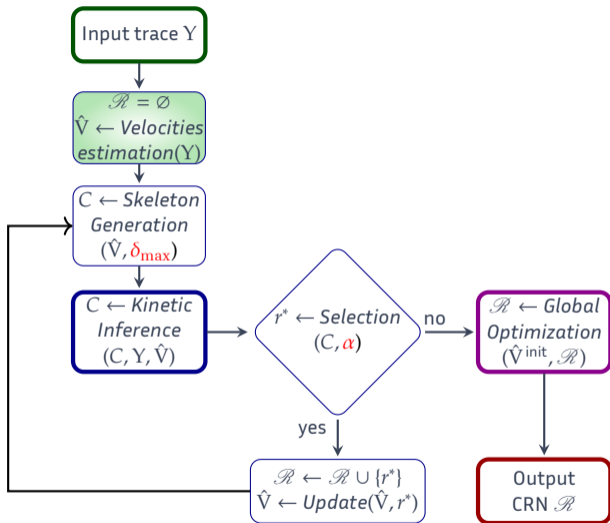
Reactmine flowchart



Reactmine flowchart

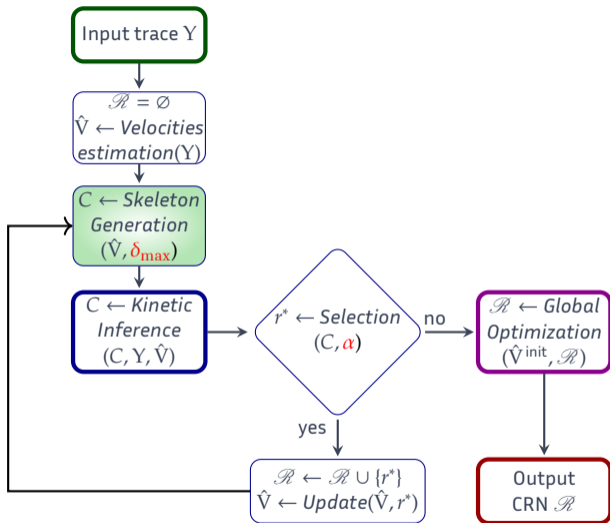


Reactmine flowchart

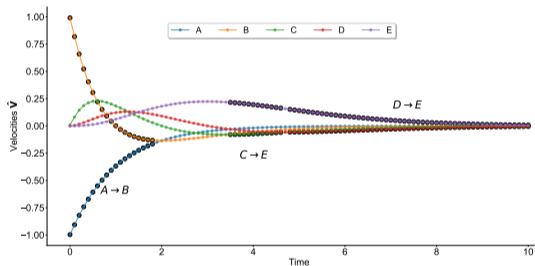


$$\text{Velocities } \hat{V} = (\hat{v}_{l,i})_{\substack{1 \leq l \leq n \\ 1 \leq i \leq m}}$$
$$\hat{v}_{l,i} = \frac{y_{l+1,i} - y_{l,i}}{t_{l+1} - t_l}$$

Reactmine flowchart

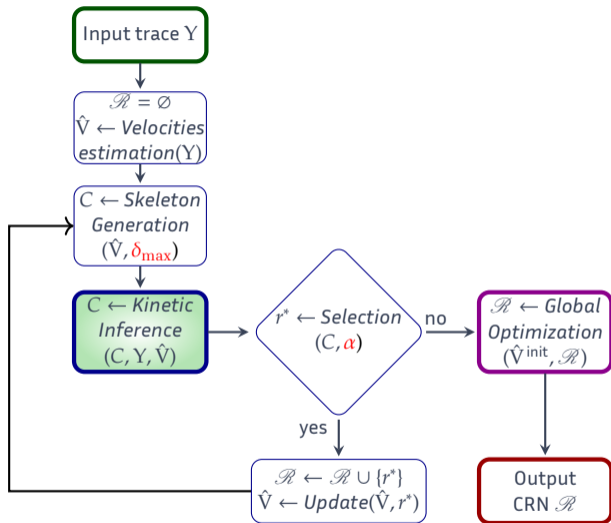


Each reaction skeleton $r = (R, P)$ is inferred based on time points $t_l(r)$ where it is **preponderant**: support set $\mathcal{T}(r)$



Reactants and products belonging to a skeleton have similar absolute variations up to δ_{\max}

Reactmine flowchart



For each reaction skeleton $r = (R, P)$
associate kinetic rate

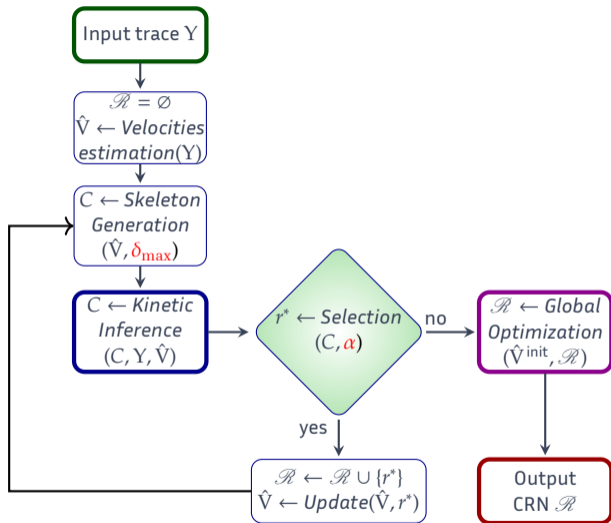
$$\forall j \in R \cup P, \forall l \in \{1, \dots, n\}, |v_{l,j}| = k \prod_{u \in R} y_{l,u}$$

Estimate k reliably on the support set $\mathcal{F}(r)$

$$\hat{k} = \frac{1}{\#\mathcal{F}(r)} \sum_{l \in \mathcal{F}(r)} \frac{|\hat{v}_{l,j}|}{\prod_{u \in R} y_{l,u}}$$

Coefficient of variation (CV) $\rho = \frac{\sigma}{|\hat{k}|}$

Reactmine flowchart

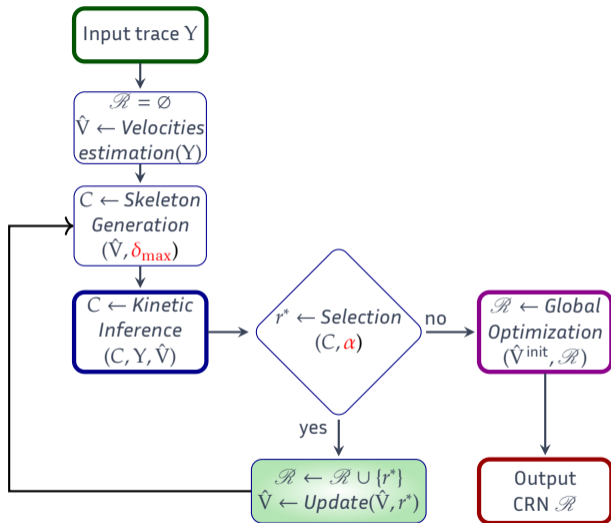


Select reaction minimizing CV

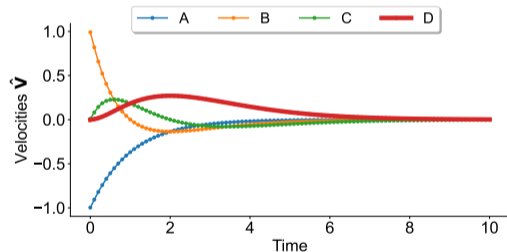
$$r^* = \operatorname{argmin}_r \rho_r$$

Accept r^* if $\rho_{r^*} < \alpha$

Reactmine flowchart



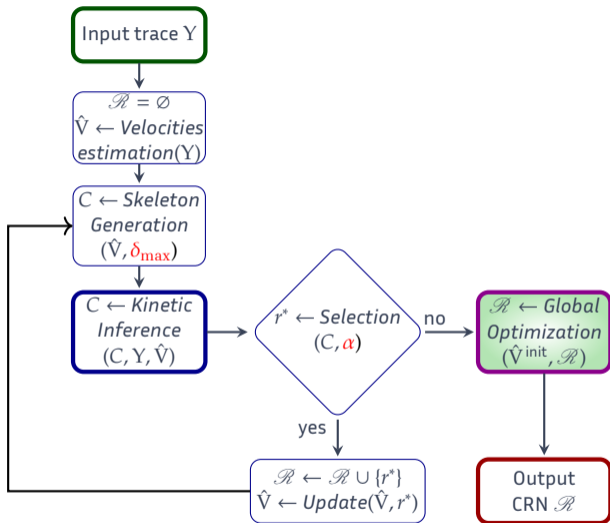
Remove the effect of accepted reaction on the velocities



$$\hat{V} \leftarrow \hat{V} - \begin{pmatrix} f(Y_{1,\bullet}) \\ \vdots \\ f(Y_{n,\bullet}) \end{pmatrix} s^T$$

effect of the reaction \uparrow \uparrow stoichiometry vector
 $Y_{i,\bullet}$: species concentration vector at time t_i

Reactmine flowchart

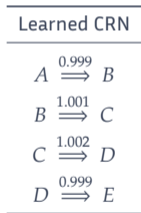
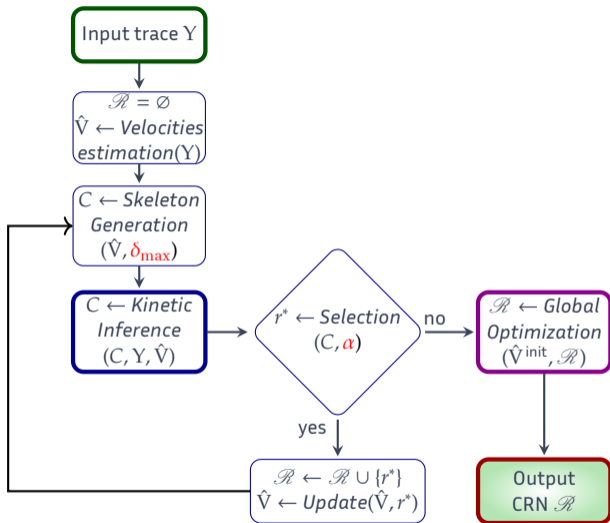


Joint optimization
of kinetic parameters
over whole trace

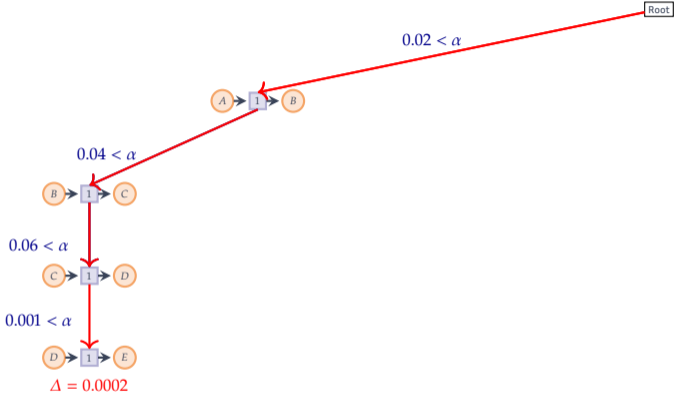
$$k = \operatorname{argmin}_{k \in \mathbb{R}_+^p} \|\hat{V}^{\text{init}} - F(Y, k)S\|_2^2$$

= Δ whole trace CRN transition discrepancy

Reactmine flowchart

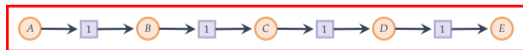
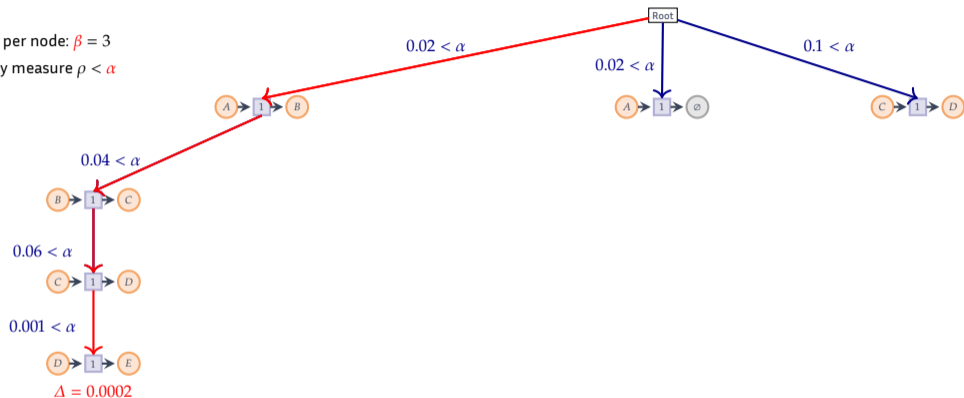


Reactmine extended to a search tree algorithm



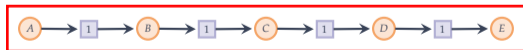
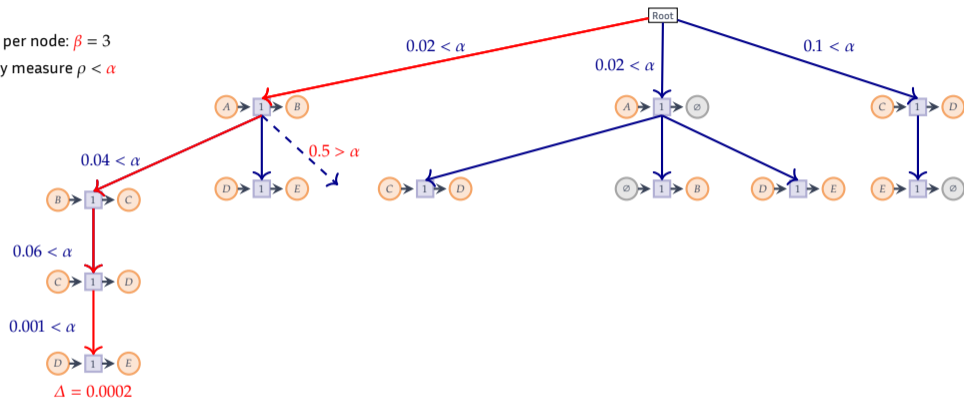
Reactmine extended to a search tree algorithm

number of candidates per node: $\beta = 3$
node accepted if quality measure $\rho < \alpha$



Reactmine extended to a search tree algorithm

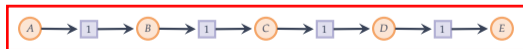
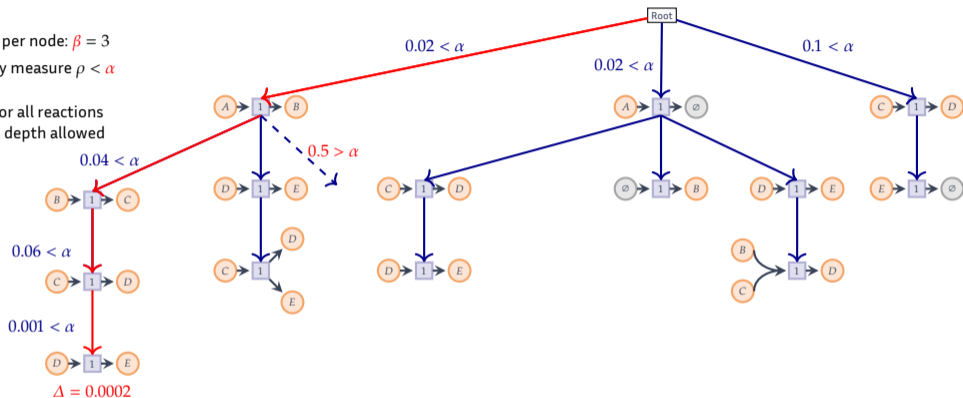
number of candidates per node: $\beta = 3$
node accepted if quality measure $\rho < \alpha$



Reactmine extended to a search tree algorithm

number of candidates per node: $\beta = 3$
node accepted if quality measure $\rho < \alpha$

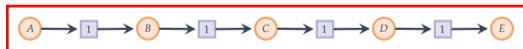
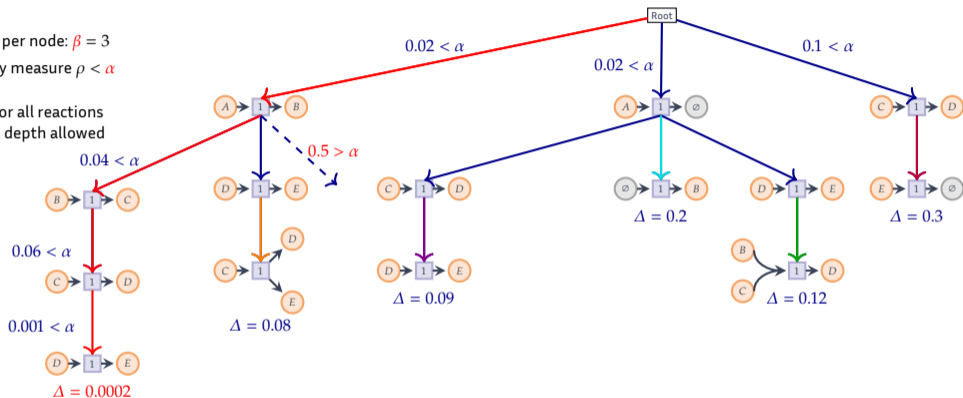
Termination: $\rho > \alpha$ for all reactions
or reached γ maximal depth allowed



Reactmine extended to a search tree algorithm

number of candidates per node: $\beta = 3$
node accepted if quality measure $\rho < \alpha$

Termination: $\rho > \alpha$ for all reactions
or reached γ maximal depth allowed

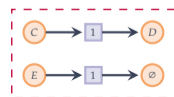
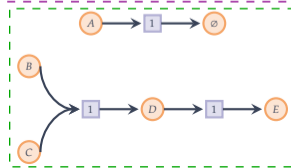
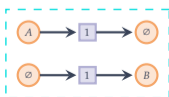
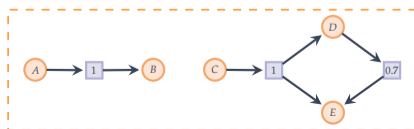
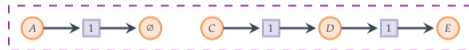
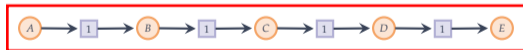
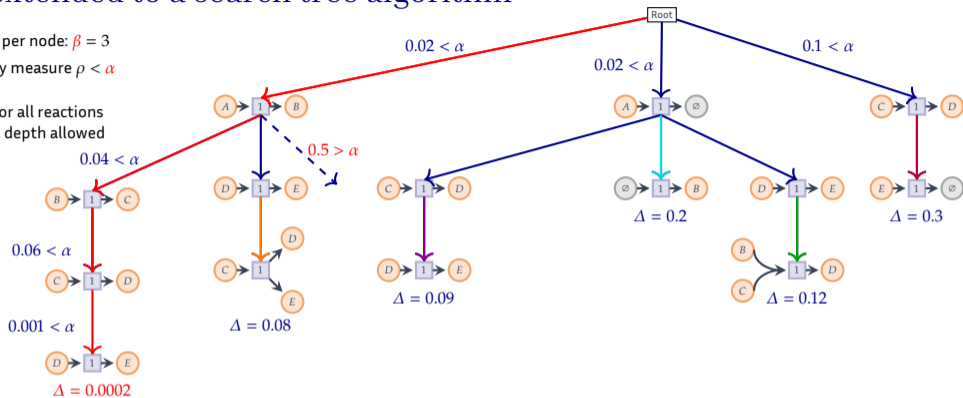


Reactmine extended to a search tree algorithm

number of candidates per node: $\beta = 3$

node accepted if quality measure $\rho < \alpha$

Termination: $\rho > \alpha$ for all reactions
or reached γ maximal depth allowed

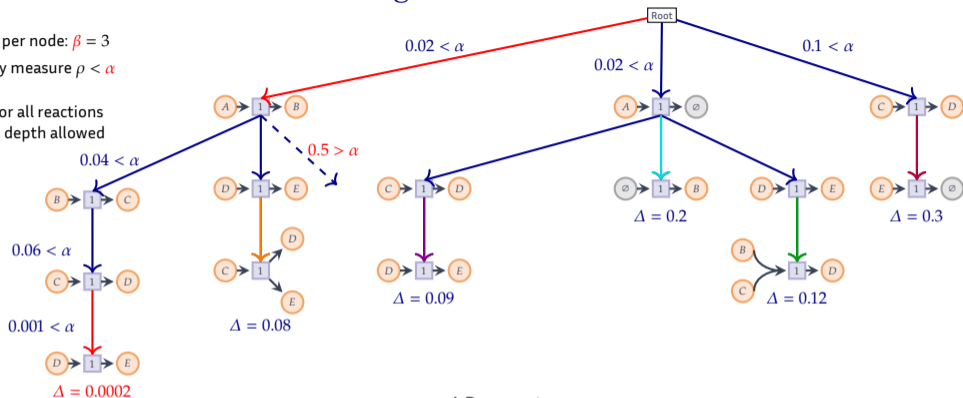


Reactmine extended to a search tree algorithm

number of candidates per node: $\beta = 3$

node accepted if quality measure $\rho < \alpha$

Termination: $\rho > \alpha$ for all reactions
or reached γ maximal depth allowed

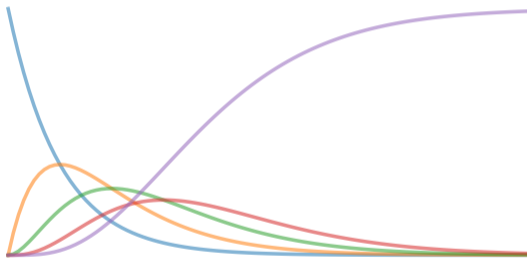


4 Parameters:

- δ_{\max} Species variations similarity threshold
- α CV threshold
- γ CRN size limit
- β Number of reaction candidates per node

Evaluation on Toy CRNs

- δ_{\max} species variations similarity threshold fixed to 3
- CRN size limit γ fixed to 6



Chain

Hidden CRN	Learned CRN
$A \xrightarrow{1} B$	$A \xrightarrow{0.999} B$
$B \xrightarrow{1} C$	$B \xrightarrow{1.001} C$
$C \xrightarrow{1} D$	$C \xrightarrow{1.002} D$
$D \xrightarrow{1} E$	$D \xrightarrow{0.999} E$

Evaluation on Toy CRNs

- δ_{\max} species variations similarity threshold fixed to 3
- CRN size limit γ fixed to 6

To be compared with SINDy
[Brunton *et al.*, *PNAS*, 2016]
which finds a weight matrix

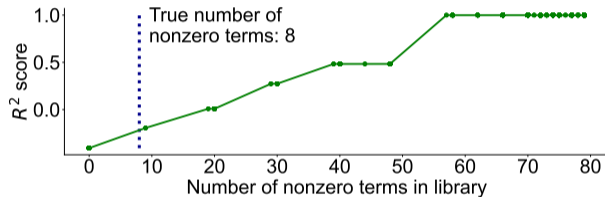
$$\mathcal{E} = \operatorname{argmin}_{\mathcal{E} \in \mathbb{R}^{p \times m}} \|\hat{\mathbf{V}}^{\text{init}} - \Theta(\mathbf{Y})\mathcal{E}\|_2^2 + \lambda \|\mathcal{E}\|_1$$

Library of functions

e.g. $y \mapsto y_1$
 $y \mapsto y_1 y_2$

Weights

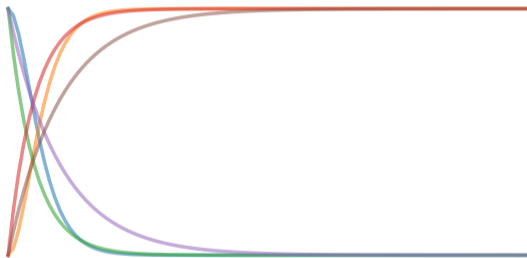
Sparsity enforcing
regularization



$$\begin{cases} \frac{dA}{dt} = -k_1 A \\ \frac{dB}{dt} = k_1 A - k_2 B \\ \frac{dC}{dt} = k_2 B - k_3 C \\ \frac{dD}{dt} = k_3 C - k_4 D \\ \frac{dE}{dt} = k_4 D \end{cases}$$

Evaluation on Toy CRNs

- δ_{\max} species variations similarity threshold fixed to 3
- CRN size limit γ fixed to 6

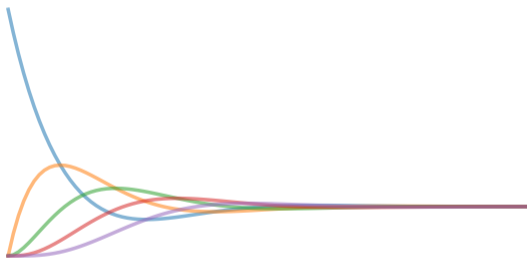


Parallel

Hidden CRN	Learned CRN
$A + D \xrightarrow{1} B + D$	$A + D \xrightarrow{2.938} B + D$
$C \xrightarrow{2} D$	$C \xrightarrow{2.001} D$
$E \xrightarrow{1} F$	$E \xrightarrow{1.0007} F$

Evaluation on Toy CRNs

- δ_{\max} species variations similarity threshold fixed to 3
- CRN size limit γ fixed to 6

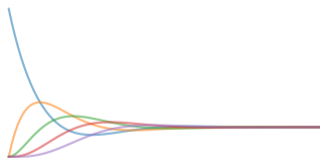


Loop

Hidden CRN	Learned CRN
$A \xrightarrow{1} B$	$A \xRightarrow{1.0103} B$
$B \xrightarrow{1} C$	$B \xRightarrow{1.009} C$
$C \xrightarrow{1} D$	$C \xRightarrow{1.009} D$
$D \xrightarrow{1} E$	$D \xRightarrow{1.009} E$
$E \xrightarrow{1} A$	$E \xRightarrow{1.01} A$

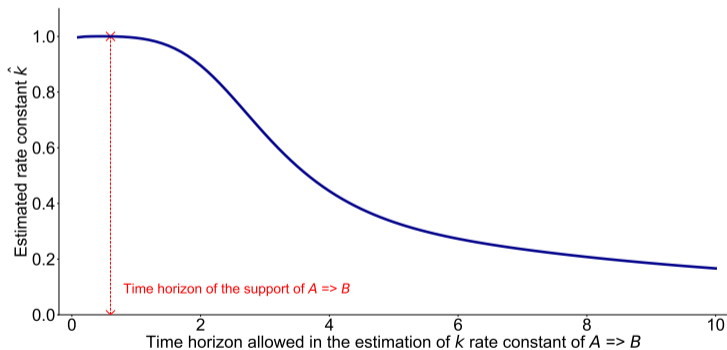
Results on the Loop CRN

Hidden CRN	Learned CRN
$A \xrightarrow{1} B$	$A \xrightarrow{1.0103} B$
$B \xrightarrow{1} C$	$B \xrightarrow{1.009} C$
$C \xrightarrow{1} D$	$C \xrightarrow{1.009} D$
$D \xrightarrow{1} E$	$D \xrightarrow{1.009} E$
$E \xrightarrow{1} A$	$E \xrightarrow{1.01} A$



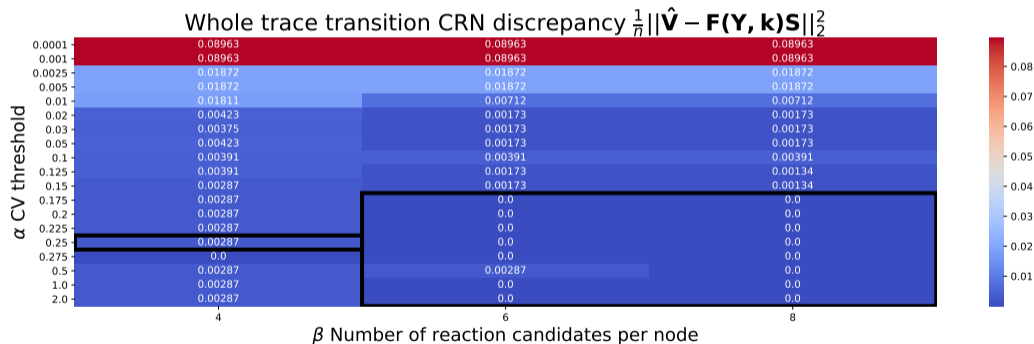
Recipe for success:

- $A \implies B$ witnessed with enough candidates allowed
- Kinetic inference based on support rather than whole trace



Reactmine parameter sensibility on the Loop CRN

- Loop CRN recovered for specific parameter values
- But consistent results for numerous (α, β) parameter combinations



Parameter can be selected by minimizing whole trace CRN transition discrepancy

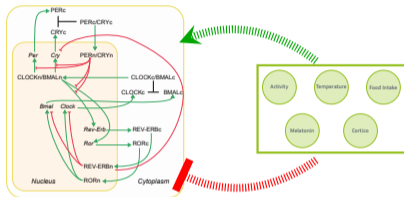
Conclusion - third part

- A method to **sequentially** infer biochemical reactions.
 - ▶ **Parsimony** of the inferred network integrated by construction.
- Philosophy: “**mining**” **reactions** at specific time points where they are preponderant.
 - ▶ More reliable estimation of reaction kinetics based on support
 - ▶ **Explainability** of the method through the support set of inferred reaction
- Extension from greedy to research algorithm allowed reconstruction of a cyclic CRN.
- Application to biological models with multiple time scales in preparation.

Summary of methodological and biomedical contributions

Summary of methodological and biomedical contributions

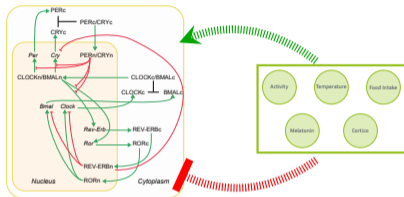
New strategies for mechanistic model learning from time series data



- **Prior knowledge** taken into account under the form of known reactions
- Provide **quantitative** information about inferred interactions

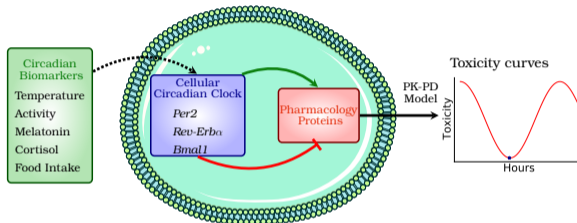
Summary of methodological and biomedical contributions

New strategies for mechanistic model learning from time series data



- **Prior knowledge** taken into account under the form of known reactions
- Provide **quantitative** information about inferred interactions

Towards personalizing chronotherapies through wearable sensors



- Novel **absolutely quantitative** clock model usable for other circadian studies
- Computation of time-dependent toxicity curves from set of mRNAs
- Integrating wearables data to investigate impact of genetic/sex differences on optimal timing