

Model learning to identify systemic regulators of the peripheral circadian clock

Julien Martinelli



Inserm

La science pour la santé
From science to health

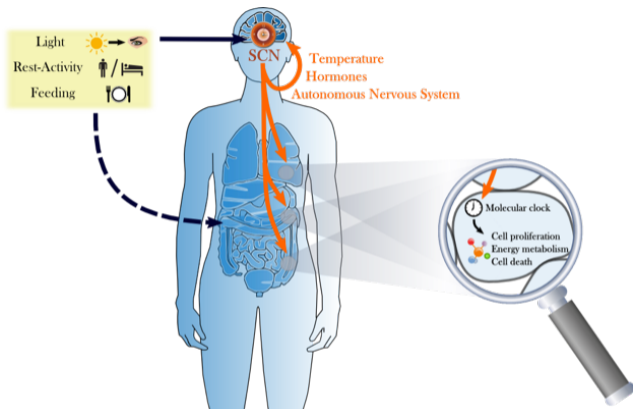


April 16th, 2021

Outline

- 1 Biological context and problem
- 2 Available data
- 3 Methods
- 4 Results

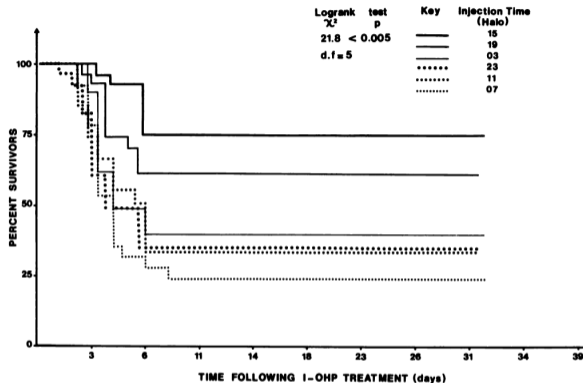
The circadian timing system



SCN : Suprachiasmatic
Nuclei

- 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

Chronotherapy



Oxaliplatin chronotoxicity in mice. Boughattas *et al.*, Cancer Research, 1989

Chronotherapy: optimal drug-time delivery based on the organisms circadian rhythms

Mouse: Chrono toxicity/efficiency for 40/28 drugs (Dallman *et al.*, Trends Mol Med., 2016)

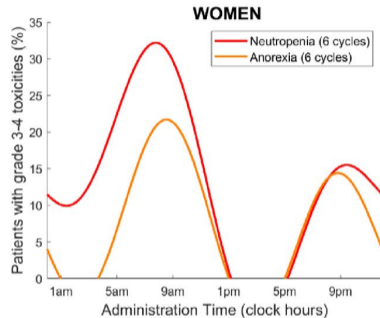
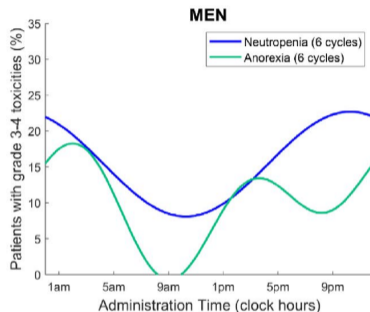
Inter-patient variability

Humans: 5-fold reduction of severe toxicities, doubled antitumoral response

Inter-patient variability

Humans: 5-fold reduction of severe toxicities, doubled antitumoral response

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

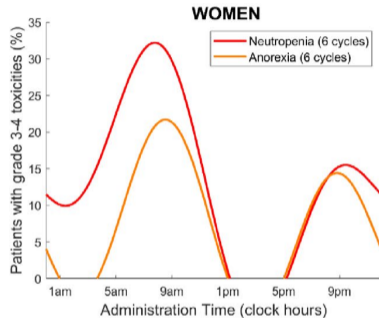
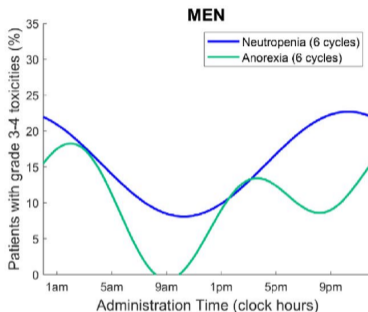


Innominato *et al.* Cancer Medicine, 2020.

Inter-patient variability

Humans: 5-fold reduction of severe toxicities, doubled antitumoral response

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

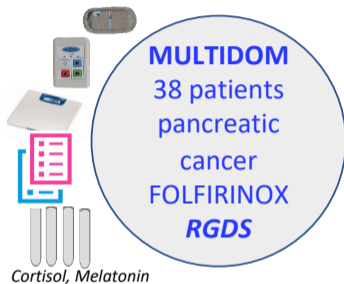


Innominato *et al.* Cancer Medicine, 2020.

Large inter-patient variability → Need for personalized optimal timing

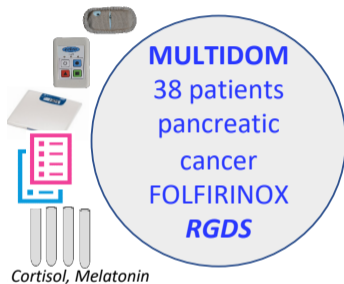
Collecting data at the patient level with eHealth platforms

Picado platform: remote data collection \Rightarrow **precision medicine**



Collecting data at the patient level with eHealth platforms

Picado platform: remote data collection \Rightarrow **precision medicine**



Statistical models untrainable \Rightarrow **Mechanistic models**

- Accounts for the lack of data
- Available data in mouse can be used for human: multi scale modelling

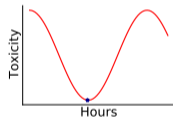
Global picture

Irinotecan Pharmacology Proteins

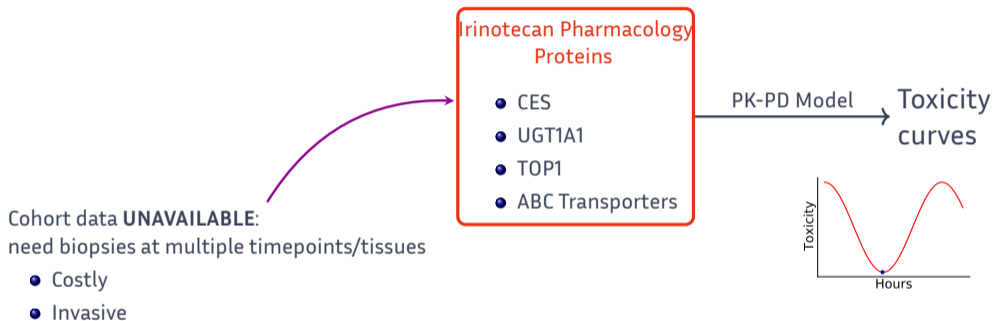
- CES
- UGT1A1
- TOP1
- ABC Transporters

PK-PD Model

Toxicity
curves



Global picture



Global picture

Systemic Regulators

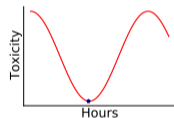
- Temperature
- Activity
- Melatonin
- Cortisol
- Food Intake

Irinotecan Pharmacology Proteins

- CES
- UGT1A1
- TOP1
- ABC Transporters

PK-PD Model

Toxicity curves



Global picture

Systemic Regulators

- Temperature
- Activity
- Melatonin
- Cortisol
- Food Intake

Systemic

- Wearables
- Saliva samples
- Daily questionnaires

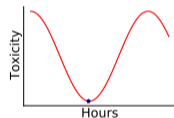


Irinotecan Pharmacology Proteins

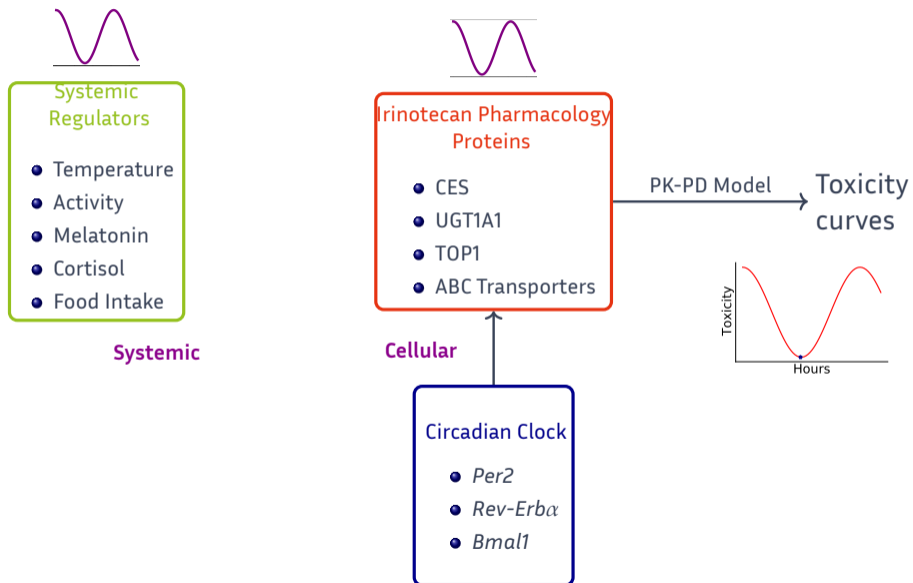
- CES
- UGT1A1
- TOP1
- ABC Transporters

PK-PD Model

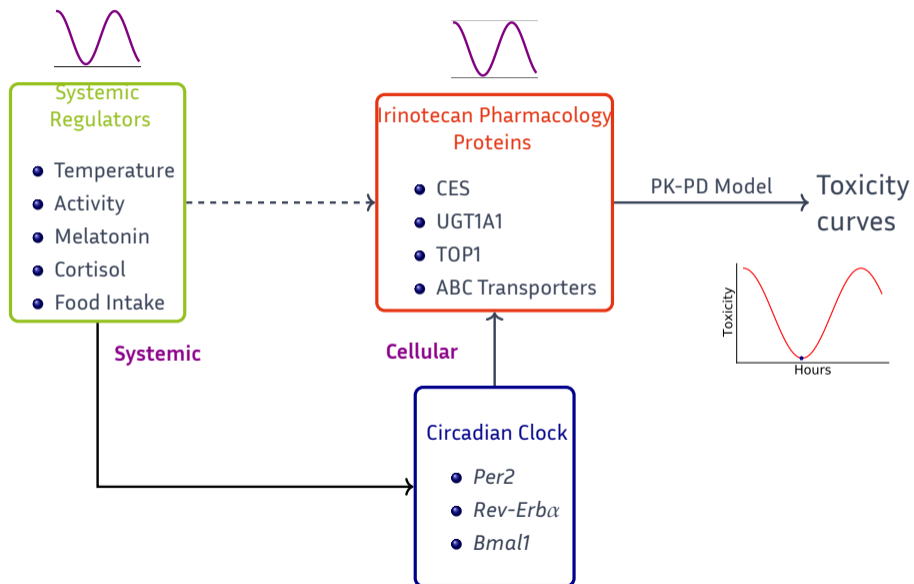
Toxicity curves



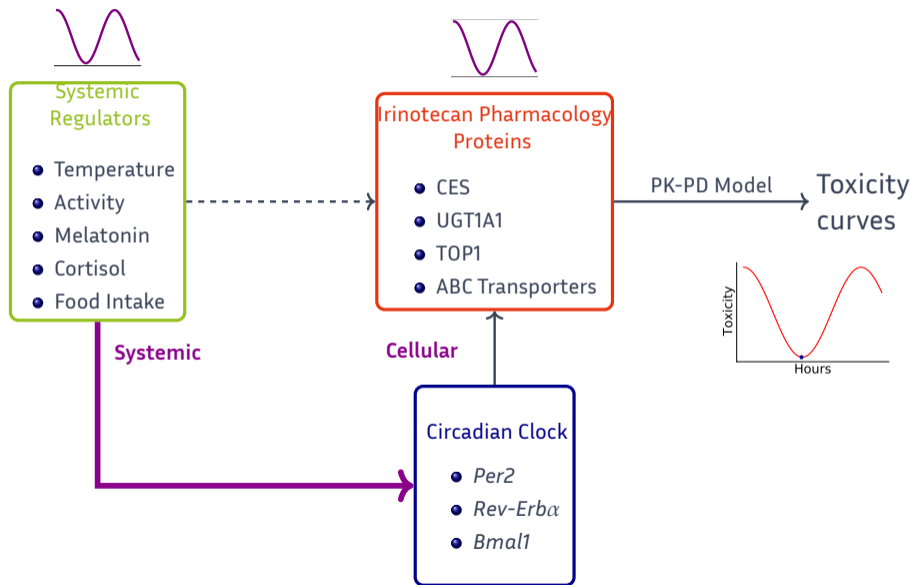
Global picture



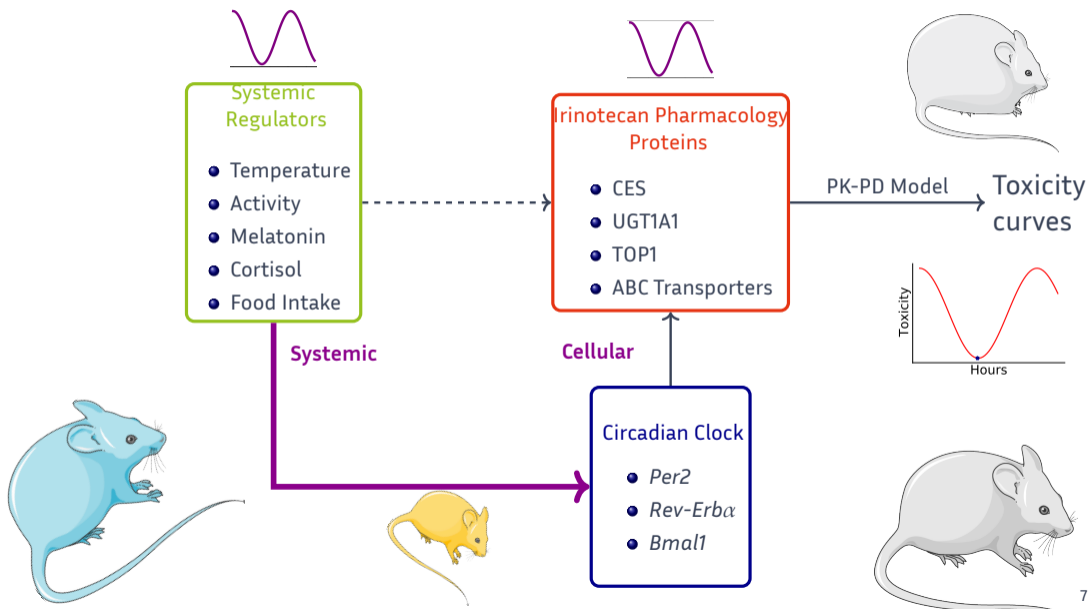
Global picture



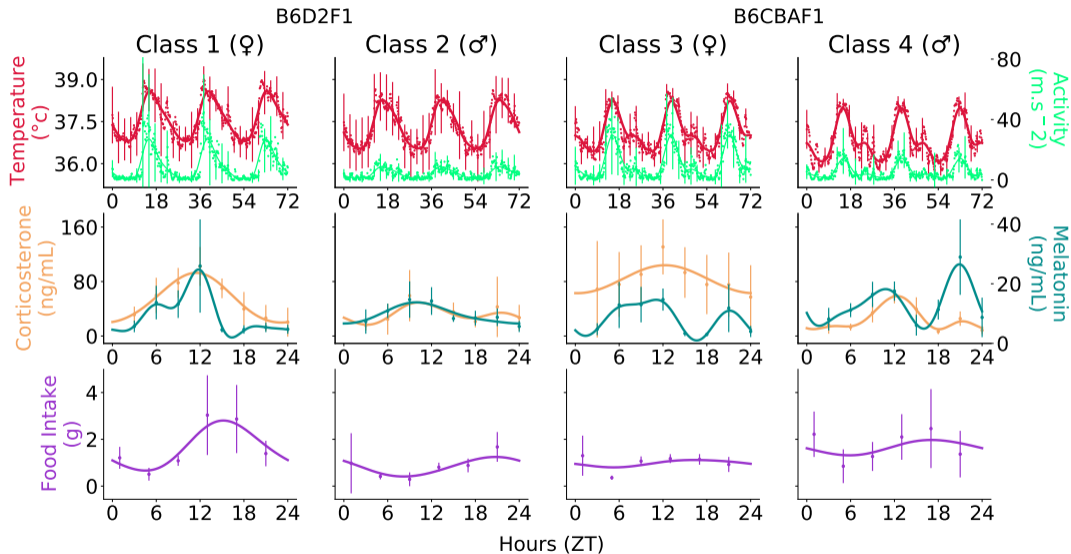
Global picture



Global picture

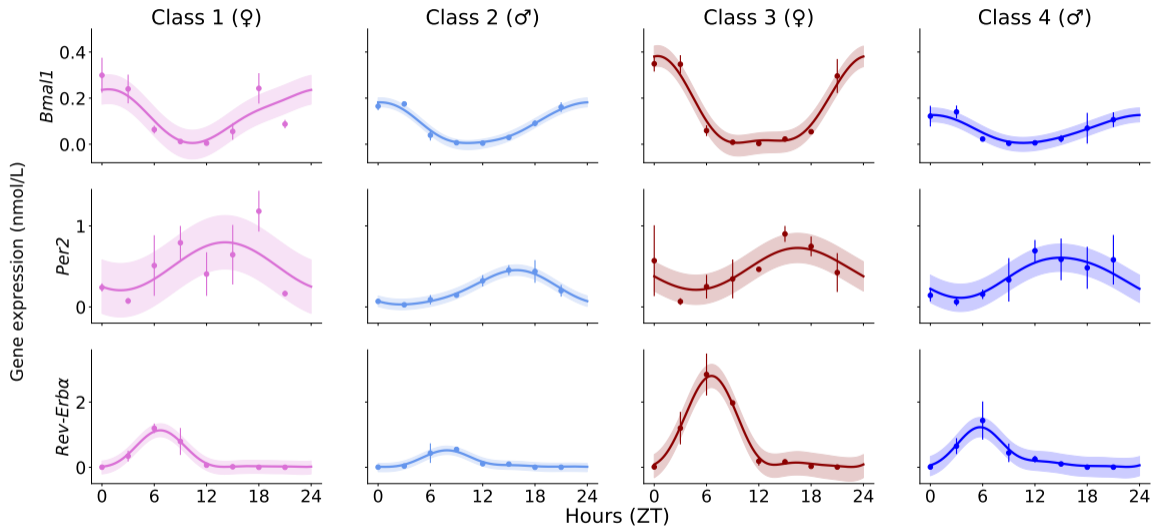


Mouse class systemic regulators data



Gaussian process regression smoothing

Mouse class gene expression data



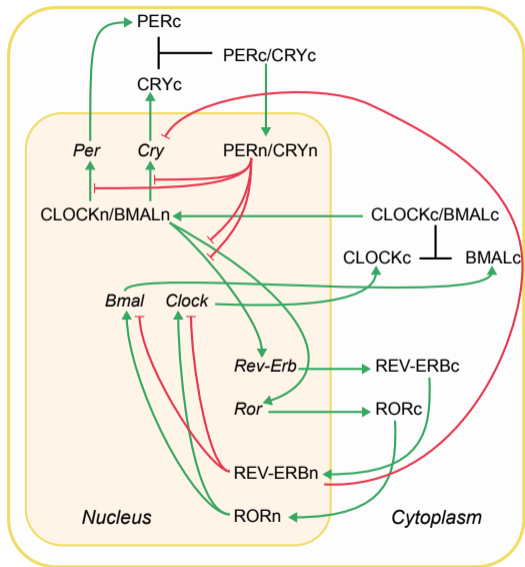
RT-qPCR acquired data. Gaussian process regression smoothing

A new model of the cellular circadian clock

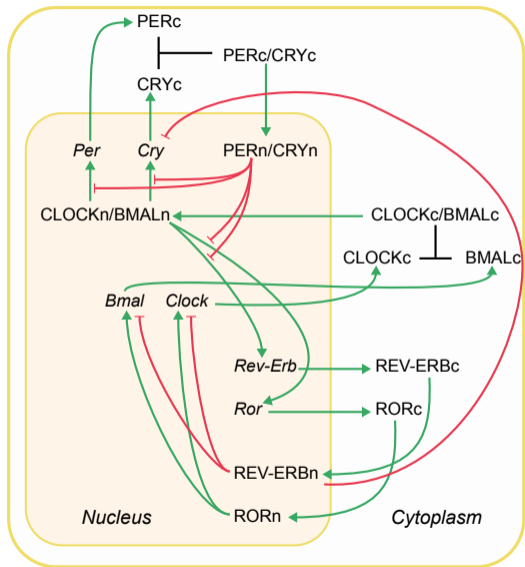
Ordinary differential equations

$$n_{vars} = 18$$

$$n_{params} = 58$$



A new model of the cellular circadian clock



Ordinary differential equations

$$n_{vars} = 18$$

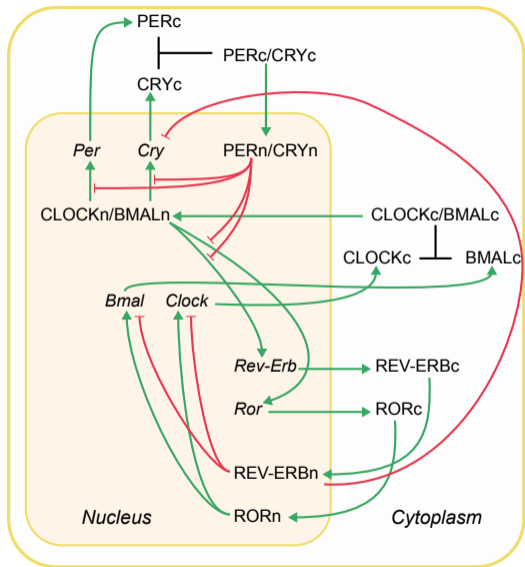
$$n_{params} = 58$$

Dynamics of gene expression:

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

Modulators

A new model of the cellular circadian clock



Ordinary differential equations

$$n_{vars} = 18$$

$$n_{params} = 58$$

Dynamics of gene expression:

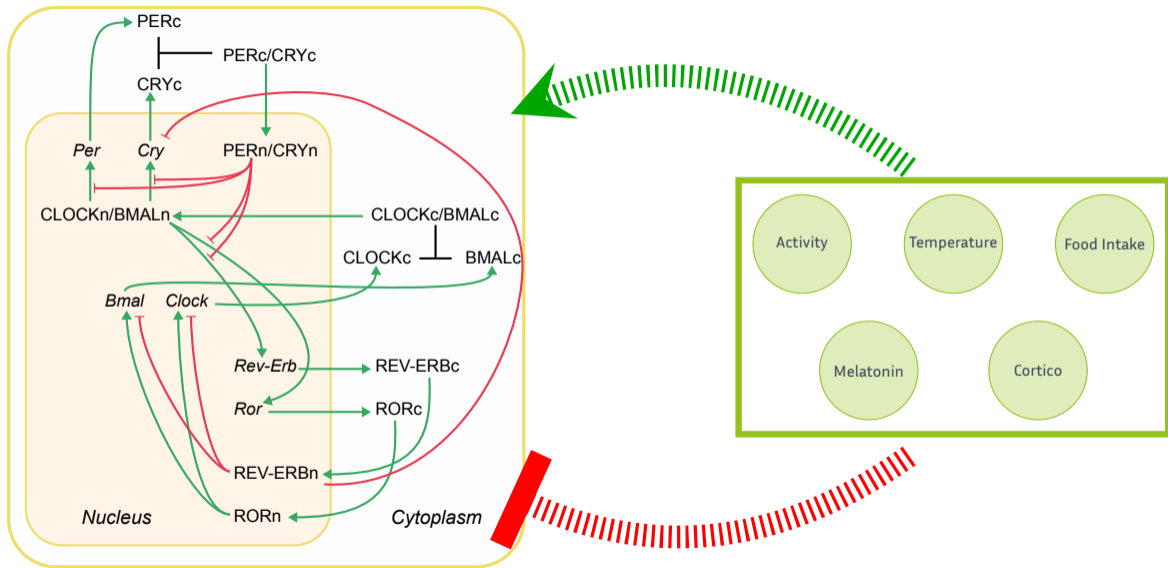
$$\frac{dx}{dt} = V_{max} \text{Transc}(M, \gamma) - \alpha x$$

Modulators

$$\text{Transc}_{Bmal1} = \frac{1 + \gamma_1 \left(\frac{ROR}{\gamma_2} \right)^{\gamma_3}}{1 + \left(\frac{REV-ERB}{\gamma_4} \right)^{\gamma_5} + \left(\frac{ROR}{\gamma_2} \right)^{\gamma_3}}$$

Hill-like kinetics

A new model of the cellular circadian clock



Incorporating systemic regulators action on gene expression

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

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

Incorporating systemic regulators action on gene expression

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

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

Incorporating systemic regulators action on gene expression

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

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

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

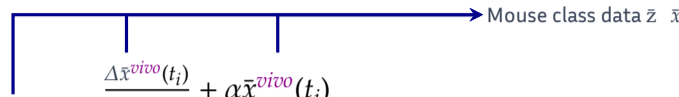
$$\frac{dx^{vivo}}{dt} = V_{\max}\text{Transc}(M, \gamma) - f(z)\alpha x^{vivo}$$
$$\Leftrightarrow f(z) = \frac{V_{\max}\text{Transc}(M, \gamma) - \frac{dx^{vivo}}{dt}}{x^{vivo}}$$

Data for $x = Bmal1, Per2$ and $Rev-Erba$

Systemic regulators identification as a regression problem

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

Systemic regulators identification as a regression problem




A diagram consisting of a horizontal arrow pointing to the right, labeled "Mouse class data $\bar{z} \quad \bar{x}$ ". Three vertical lines descend from the arrow to the top of a fraction in the equation below. A vertical line also descends from the left side of the fraction to the left side of the equation.

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

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

Systemic regulators identification as a regression problem



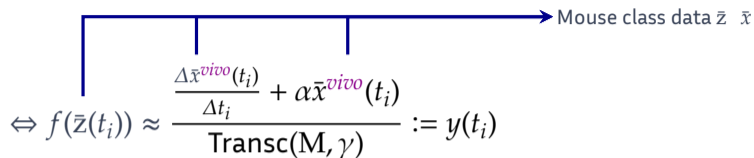
A diagram consisting of a horizontal blue arrow pointing to the right, labeled "Mouse class data $\bar{z} \quad \bar{x}$ ". Three vertical blue lines descend from the arrow to the top of a fraction in the equation below. The equation is:

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

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

 Systemic Regulators

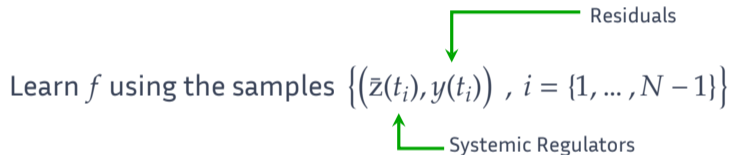
Systemic regulators identification as a regression problem



A blue arrow points from the text "Mouse class data \bar{z}, \bar{x} " to the regression equation below. Three vertical lines connect the top of the equation to the arrow's path, indicating that the equation is derived from the data.

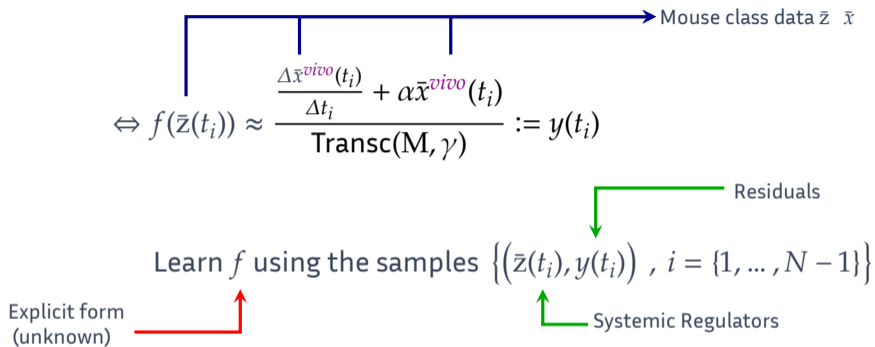
$$\Leftrightarrow f(\bar{z}(t_i)) \approx \frac{\frac{\Delta \bar{x}^{vivo}(t_i)}{\Delta t_i} + \alpha \bar{x}^{vivo}(t_i)}{\text{Transc}(\mathbf{M}, \gamma)} := y(t_i)$$

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

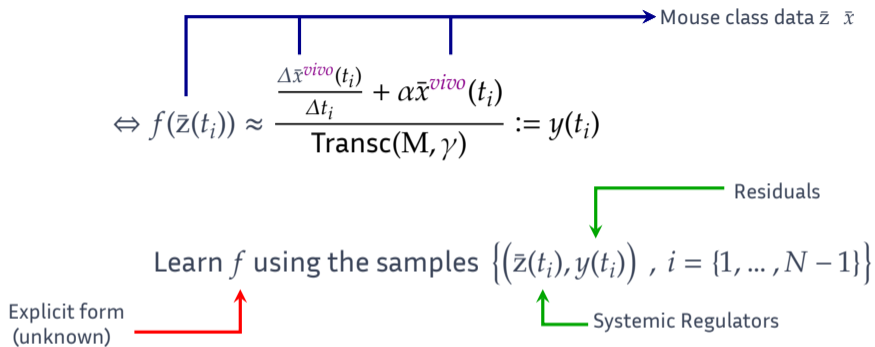


A green arrow labeled "Residuals" points from the top of the sample set to the right. A green arrow labeled "Systemic Regulators" points from the bottom of the sample set to the left.

Systemic regulators identification as a regression problem



Systemic regulators identification as a regression problem

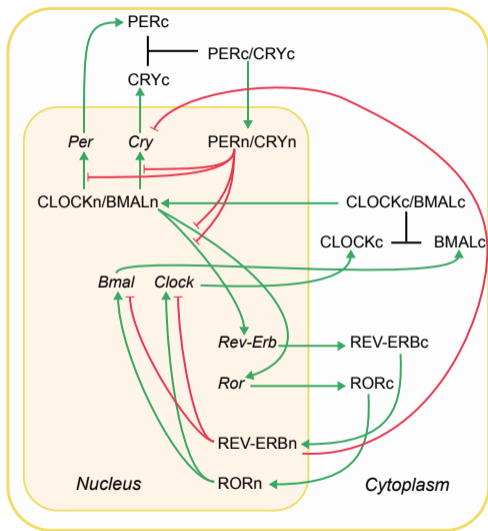


Learning f usually boils down to solve

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

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

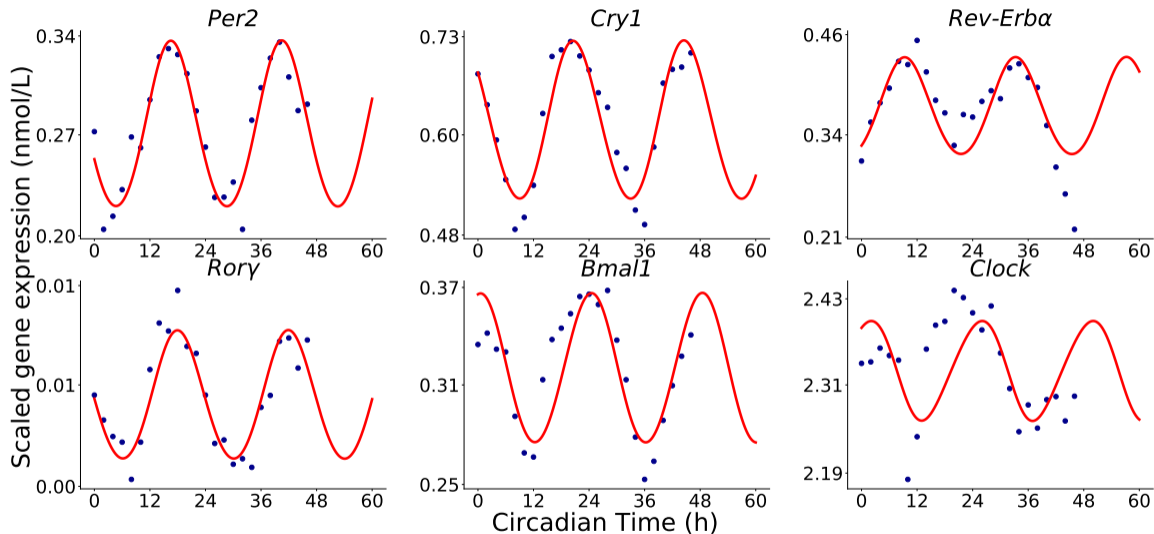
Computing residuals γ : acquisition of clock parameters and protein levels



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

- *In vitro* setting $\implies f(z)$ constant
- Fit model on *in vitro* hepatocytes data (Atwood *et al.*, PNAS, 2011)

Clock model fit on *in vitro* hepatocytes data



$\Rightarrow \alpha, \gamma$ and $M(t)$ estimates obtained (fit performed with *CMA-ES*)

Multiple trajectories for stronger inference results

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

- α, γ and M are educated guesses...
- ...But are just estimates from an *in vitro* dataset

Multiple trajectories for stronger inference results

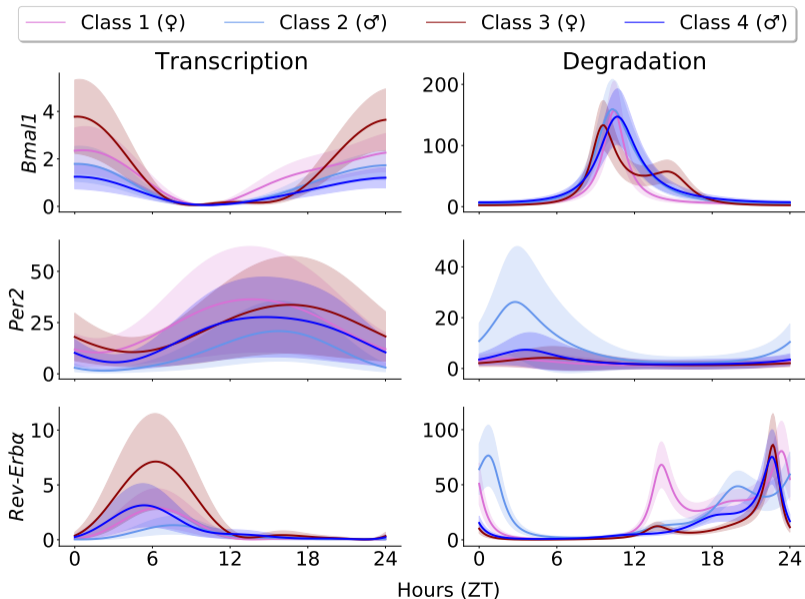
$$f(\bar{z}(t_i)) \approx \frac{\frac{\Delta \bar{x}^{vivo}(t_i)}{\Delta t_i} + \alpha \bar{x}^{vivo}(t_i)}{\text{Transc}(M, \gamma)} := y(t_i) \quad (*)$$

- α, γ and M are educated guesses...
- ...But are just estimates from an *in vitro* dataset

Solution:

- 1 Perturbed clock model parameter vectors are sampled
- 2 $n = 2000$ new residual trajectories y are generated from the perturbed clocks

Residual trajectories y



Linear regression

For each residual y , a linear model $\sum_j \beta_j z_j$ is fitted

- The **active regulators** of the fitted model should be the same classwise.
- **Different weights** β for a regulator from one class to another are allowed

Linear regression

For each residual y , a linear model $\sum_j \beta_j z_j$ is fitted

- The **active regulators** of the fitted model should be the same classwise.
- **Different weights** β for a regulator from one class to another are allowed

0.8 Food Intake (Class 1)
+ 0.3 Temperature

Linear regression

For each residual y , a linear model $\sum_j \beta_j z_j$ is fitted

- The **active regulators** of the fitted model should be the same classwise.
- **Different weights** β for a regulator from one class to another are allowed

0.8 Food Intake (Class 1)
+ 0.3 Temperature

0.7 Food Intake (Class 2)
+ **0.5 Activity**

Linear regression

For each residual y , a linear model $\sum_j \beta_j z_j$ is fitted

- The **active regulators** of the fitted model should be the same classwise.
- **Different weights** β for a regulator from one class to another are allowed

0.8 Food Intake (Class 1)
+ 0.3 Temperature

0.7 Food Intake (Class 2)
+ **0.5 Temperature**

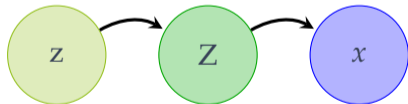
Linear regression

For each residual y , a linear model $\sum_j \beta_j z_j$ is fitted

- The **active regulators** of the fitted model should be the same classwise.
- **Different weights** β for a regulator from one class to another are allowed

Need to account for the delay introduced by moving in different compartments

\Rightarrow *Integral* regulators $Z_j(t) = \int_0^t z_j(s) ds$ are added: $z \leftarrow (z, Z)$



A regulator z_j and its integral Z_j are never found together in a model for all j

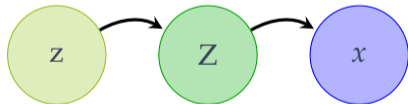
Linear regression

For each residual y , a linear model $\sum_j \beta_j z_j$ is fitted

- The **active regulators** of the fitted model should be the same classwise.
- **Different weights** β for a regulator from one class to another are allowed

Need to account for the delay introduced by moving in different compartments

\Rightarrow *Integral* regulators $Z_j(t) = \int_0^t z_j(s) ds$ are added: $z \leftarrow (z, Z)$



A regulator z_j and its integral Z_j are never found together in a model for all j

0.8 Food Intake (Class 1)
+ 0.9 \int Food Intake

0.7 Food Intake (Class 2)
+ 0.6 \int Food Intake

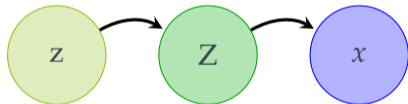
Linear regression

For each residual y , a linear model $\sum_j \beta_j z_j$ is fitted

- The **active regulators** of the fitted model should be the same classwise.
- **Different weights** β for a regulator from one class to another are allowed

Need to account for the delay introduced by moving in different compartments

\Rightarrow *Integral* regulators $Z_j(t) = \int_0^t z_j(s) ds$ are added: $z \leftarrow (z, Z)$

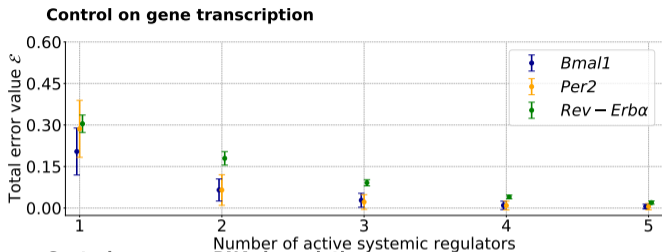


A regulator z_j and its integral Z_j are never found together in a model for all j

0.8 Food Intake (Class 1)
+ 0.4 \int Melatonin

0.7 Food Intake (Class 2)
+ 0.2 \int Melatonin

Total error as a function of the number of involved regulators



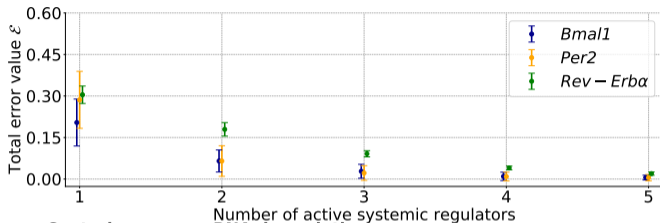
$$\mathcal{E}(y, \bar{z}) := \frac{1}{4n} \sum_{c=1}^4 \sum_{k=1}^n \min_{\beta_k^{(c)}} \ell(y_k^{(c)}, \bar{z}^{(c)}, \beta_k^{(c)})$$

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

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

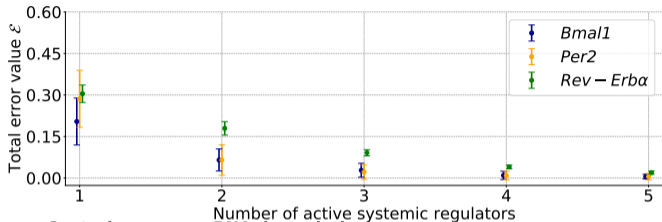
$$\mathcal{E}(y, \bar{z}) := \frac{1}{4n} \sum_{c=1}^4 \sum_{k=1}^n \min_{\beta_k^{(c)}} \ell(y_k^{(c)}, \bar{z}^{(c)}, \beta_k^{(c)})$$

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

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

Total error as a function of the number of involved regulators

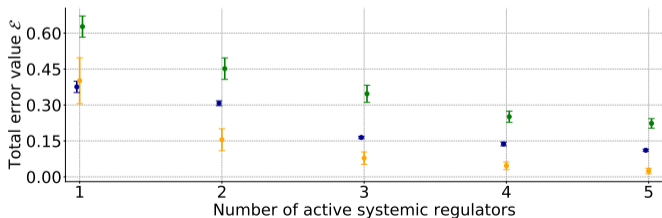
Control on gene transcription



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

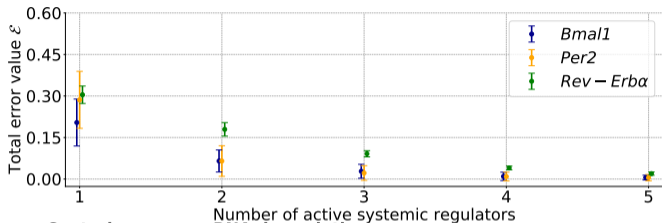
⇒ No linear control of regulators on *Rev-Erbα* transcription

Control on gene mRNA degradation



Total error as a function of the number of involved regulators

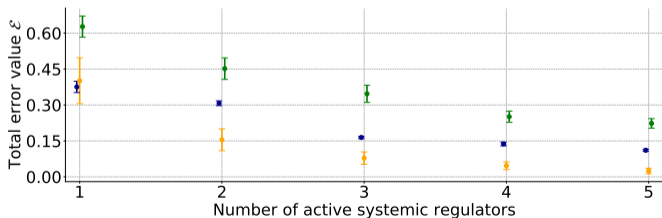
Control on gene transcription



- *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

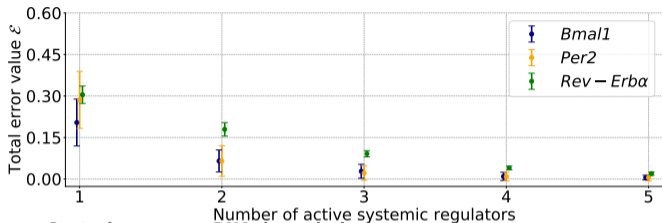


No model with < 3 terms

⇒ No linear control of regulators on gene mRNA degradation

Total error as a function of the number of involved regulators

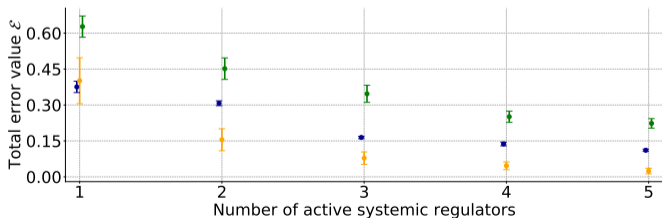
Control on gene transcription



- *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

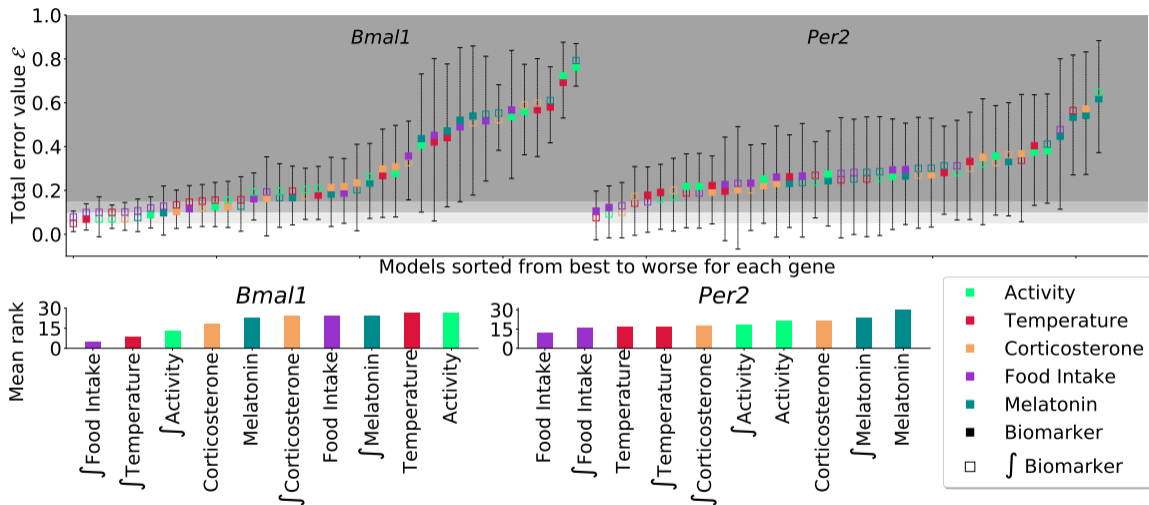


No model with < 3 terms

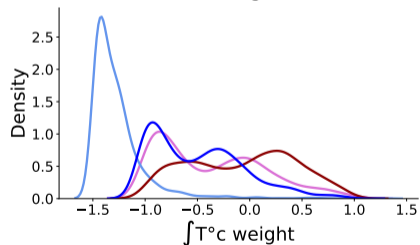
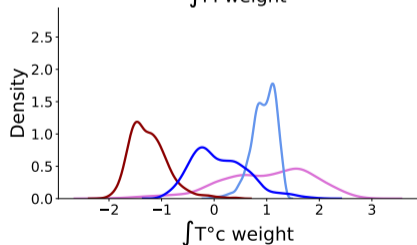
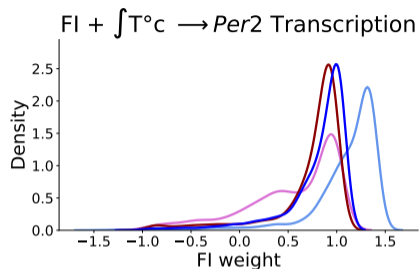
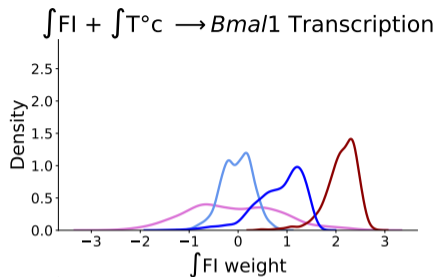
⇒ No linear control of regulators on gene mRNA degradation

Focus on 2-term models for Transcription: 40 models

2-term models ranking



Classwise weights analysis for best 2-term models



— Class 1 (♀) — Class 2 (♂) — Class 3 (♀) — Class 4 (♂)

Conclusion & Perspectives

Under all hypotheses:

- Food Intake / T^oc main actors for transcription control: consistent with literature
- Linear control of studied systemic regulators on gene mRNA degradation unlikely

Conclusion & Perspectives

Under all hypotheses:

- Food Intake / T^oc main actors for transcription control: consistent with literature
- Linear control of studied systemic regulators on gene mRNA degradation unlikely

Model learning approach:

- Integration of data at systemic and cellular level
- Knowledge encompassed in model, mechanistic predictions on unknown parts
- Handle large number of variables within the sparse multi-task regression framework

Conclusion & Perspectives

Under all hypotheses:

- Food Intake / $T^{\circ}C$ main actors for transcription control: consistent with literature
- Linear control of studied systemic regulators on gene mRNA degradation unlikely

Model learning approach:

- Integration of data at systemic and cellular level
- Knowledge encompassed in model, mechanistic predictions on unknown parts
- Handle large number of variables within the sparse multi-task regression framework

What's next:

- Integration of best regulator models back in the ODEs
- Validation on human data

Want to know more? Paper accepted at Bioinformatics (ECCB21 Proceedings)!



Julien Martinelli, Sandrine Dulong, Xiao-Mei Li, Michèle Teboul, Sylvain Soliman, Francis Lévi, François Fages, and Annabelle Ballesta. *Model learning to identify systemic regulators of the peripheral circadian clock*. working paper or preprint. Mar. 2021. url: <https://hal.inria.fr/hal-03183579>.