

Identification de régulateurs systémiques de l'horloge périphérique circadienne par apprentissage de modèles

Julien Martinelli



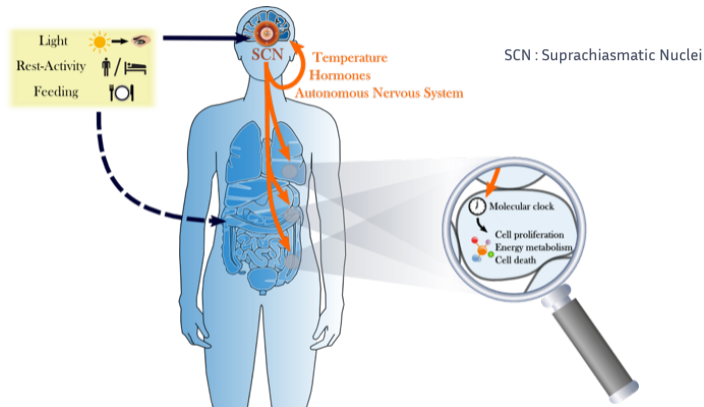
Inserm

La science pour la santé
From science to health



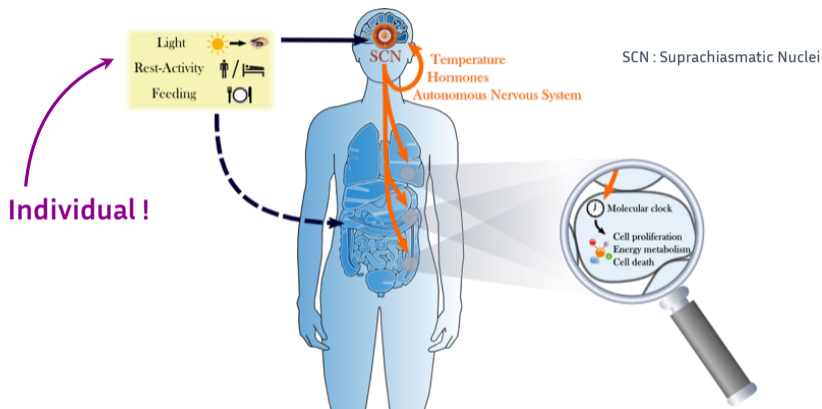
June 21st, 2021

The circadian timing system



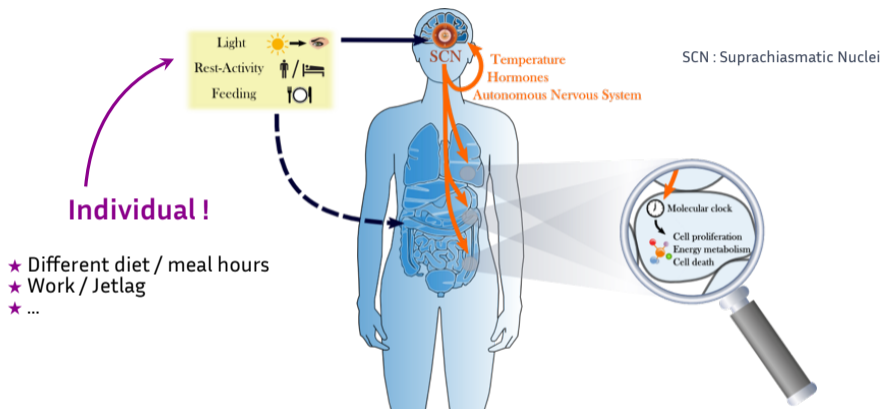
- A master clock acting as an autonomous $\approx 24\text{h}$ -oscillator synchronised by external cues
- This master clock **entrains** the peripheral clocks in the cells *via* physiological signals
- The peripheral clock induces oscillations in key intracellular processes

The circadian timing system



- 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



- 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

Clock-induced oscillations in intracellular processes are individual

Repercussions e.g. cancer chronotherapy at the individual level

→ **Precision medicine, but with what data?**

Clock-induced oscillations in intracellular processes are individual

Repercussions e.g. cancer chronotherapy at the individual level

→ **Precision medicine, but with what data?**



Infer the links between measurable variables and the peripheral clock

Clock-induced oscillations in intracellular processes are individual

Repercussions e.g. cancer chronotherapy at the individual level

→ **Precision medicine, but with what data?**

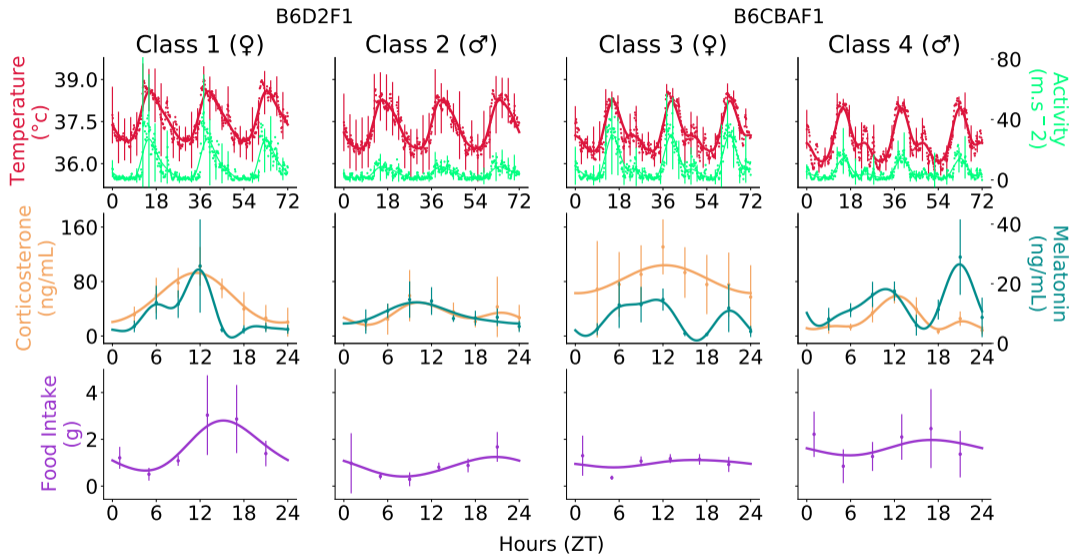


Infer the links between measurable variables and the peripheral clock

Focus on mice: data available both at the systemic and cellular level

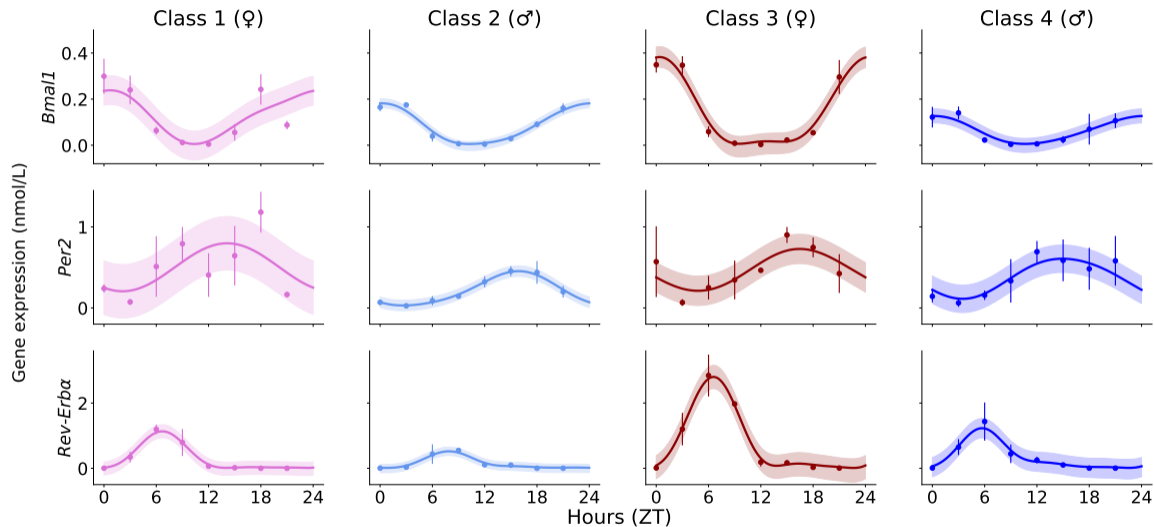


Mouse class systemic regulators data



Solid lines: gaussian process regression smoothing

Mouse class gene expression data



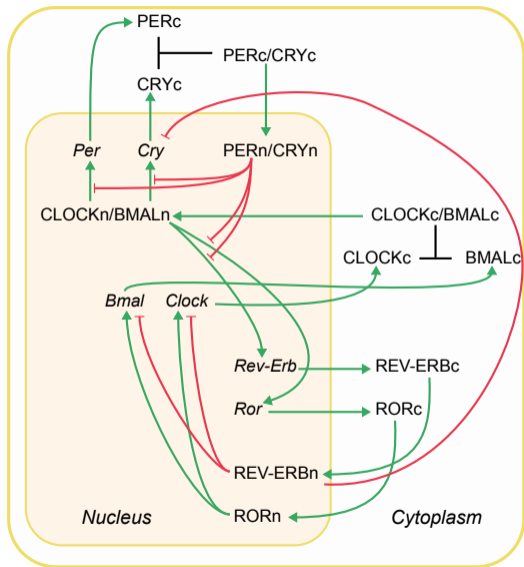
RT-qPCR acquired data. Solid lines and stds: gaussian process regression smoothing

A new model of the cellular circadian clock

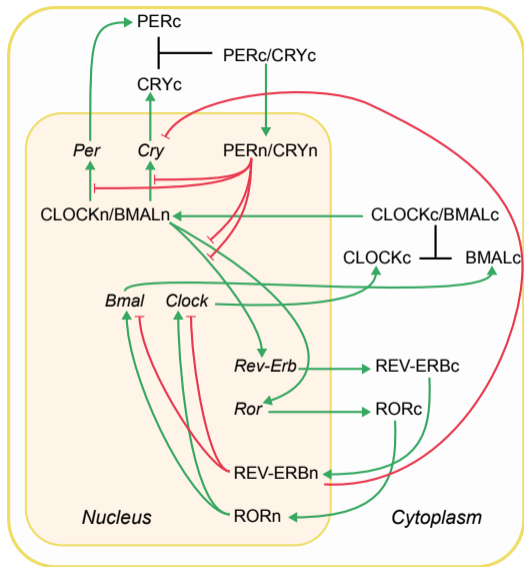
Ordinary differential equations

$$n_{vars} = 18$$

$$n_{params} = 58$$



A new model of the cellular circadian clock



Ordinary differential equations

$$n_{vars} = 18$$

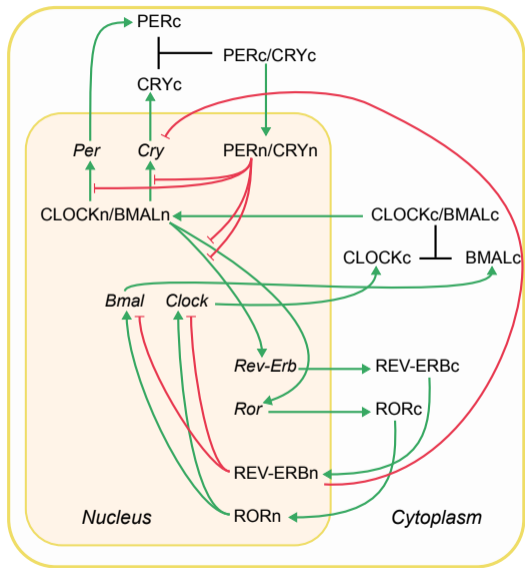
$$n_{params} = 58$$

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:

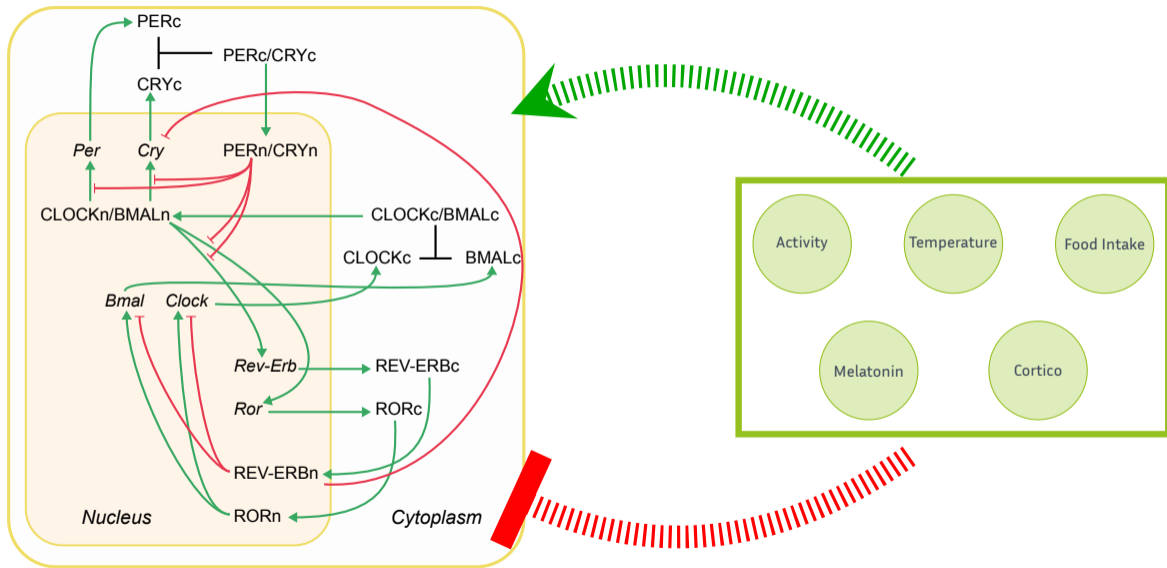
$$\frac{dx}{dt} = V_{max} \text{Transc}(M, \gamma) - \alpha x$$

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: Multiplicative control of systemic regulators z on gene transcription

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

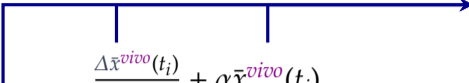
Incorporating systemic regulators action on gene expression

Hypothesis: Multiplicative control of systemic regulators z on gene transcription

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

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

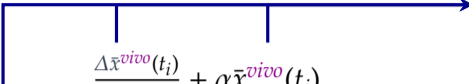
Systemic regulators identification as a regression problem



The diagram illustrates the relationship between mouse class data and a regression model. A horizontal arrow points from the text "Mouse class data $\bar{z} \quad \bar{x}$ " to the left. Three vertical lines descend from the arrow at different points. The first vertical line connects to the left side of the equation. The second vertical line connects to the numerator of the fraction in the equation. The third vertical line connects to the term $\alpha \bar{x}^{vivo}(t_i)$ in the numerator. This indicates that the regression model's input is the mouse class data, and its output is the systemic regulator identification, which is a function of the mouse class data.

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

Systemic regulators identification as a regression problem

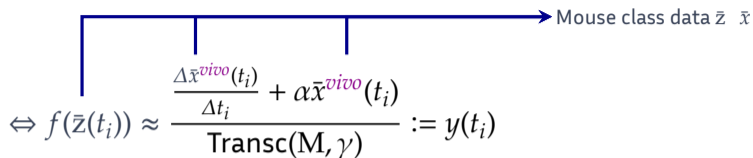


The diagram consists of a horizontal arrow pointing to the right, labeled "Mouse class data $\bar{z} \quad \bar{x}$ ". Three vertical lines descend from the arrow at different points. The first vertical line connects to the left side of the equation. The second vertical line connects to the term $\frac{\Delta \bar{x}^{vivo}(t_i)}{\Delta t_i}$. The third vertical line connects to the term $\alpha \bar{x}^{vivo}(t_i)$.

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

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

Systemic regulators identification as a regression problem



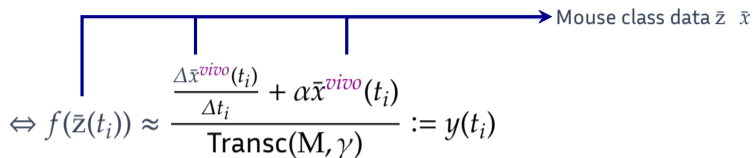
A diagram illustrating the regression problem. A horizontal blue arrow points from the left towards the text "Mouse class data $\bar{z} \quad \bar{x}$ ". Three vertical blue lines descend from the arrow at different positions. The first vertical line is connected to the left side of the equation. The second vertical line is connected to the term $\frac{\Delta \bar{x}^{vivo}(t_i)}{\Delta t_i}$. The third vertical line is connected to the term $\alpha \bar{x}^{vivo}(t_i)$.

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

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

 Systemic Regulators

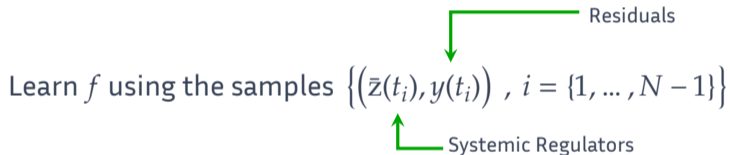
Systemic regulators identification as a regression problem



A blue line starts from the right, labeled "Mouse class data $\bar{z} \quad \bar{x}$ ", and branches into three vertical lines that point to the terms $\bar{z}(t_i)$, $\bar{x}^{vivo}(t_i)$, and $\text{Transc}(\mathbf{M}, \gamma)$ in the equation below.

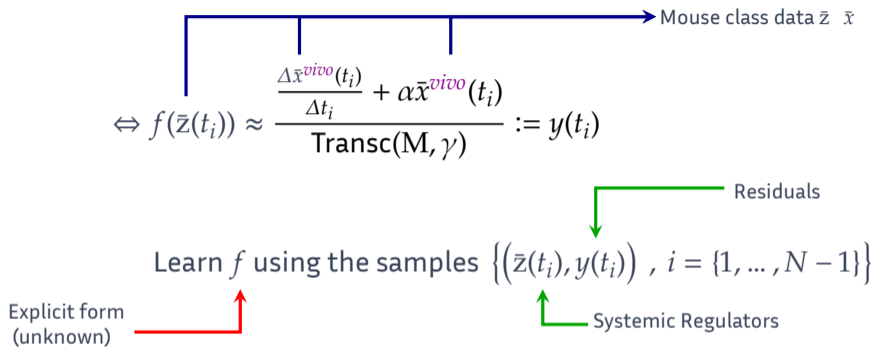
$$\Leftrightarrow f(\bar{z}(t_i)) \approx \frac{\frac{\Delta \bar{x}^{vivo}(t_i)}{\Delta t_i} + \alpha \bar{x}^{vivo}(t_i)}{\text{Transc}(\mathbf{M}, \gamma)} := y(t_i)$$

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

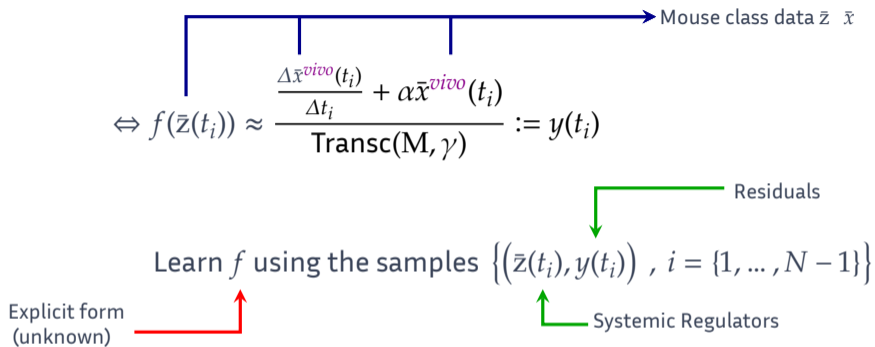


Two green arrows point towards the sample set. One arrow from the top right, labeled "Residuals", points down to the sample set. Another arrow from the bottom right, labeled "Systemic Regulators", points up to the sample set.

Systemic regulators identification as a regression problem



Systemic regulators identification as a regression problem

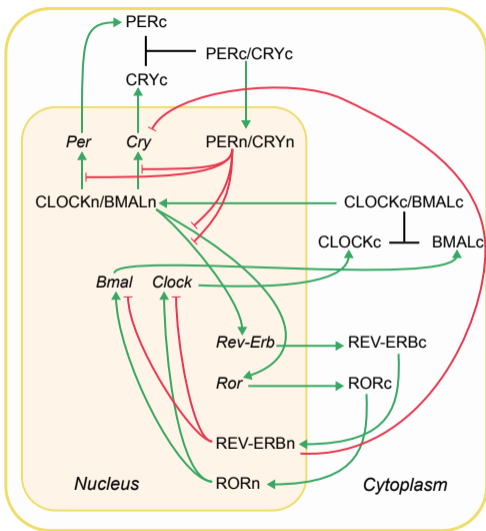


Learning f usually boils down to solve

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

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

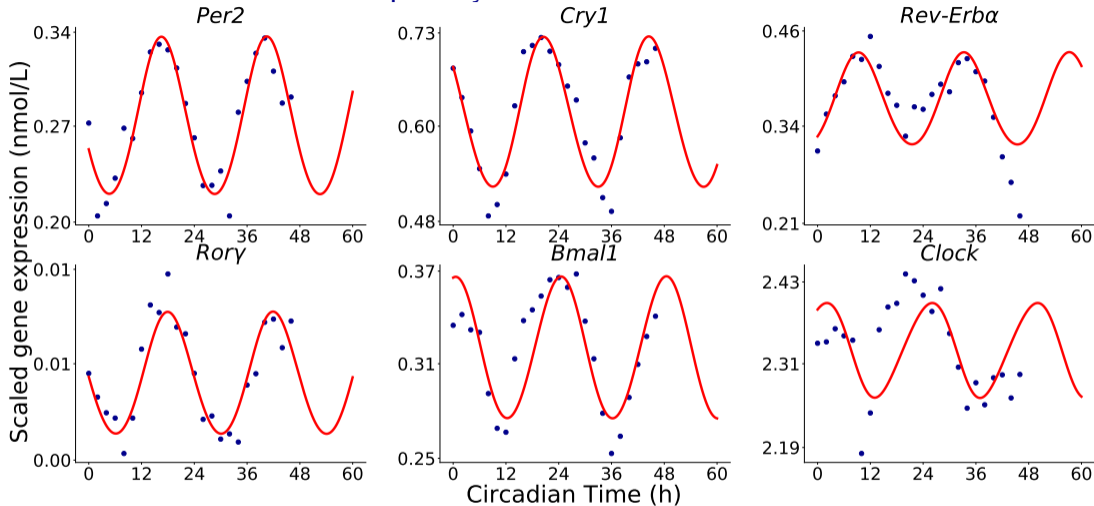
Computing residuals γ : acquisition of clock parameters and protein levels



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

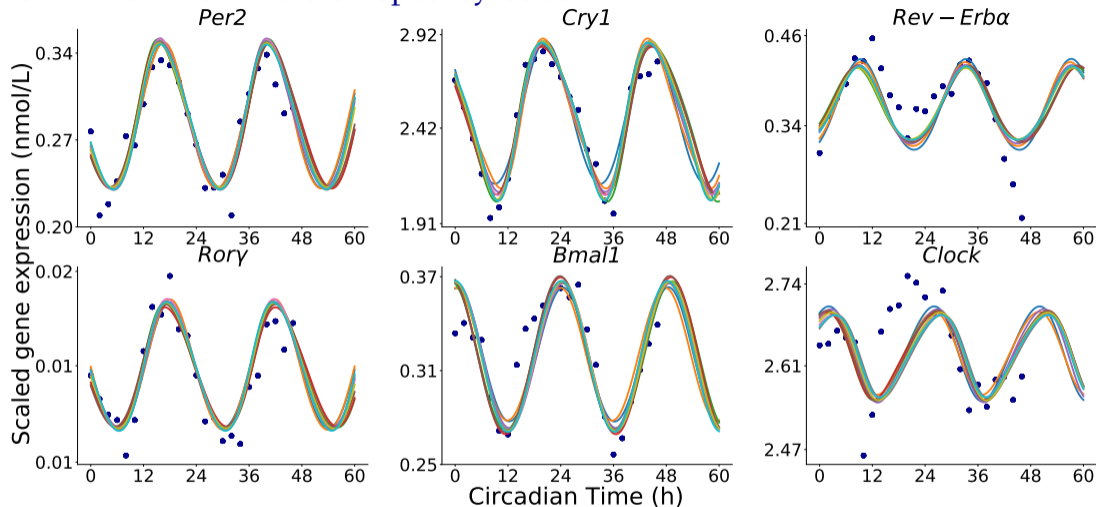
- *In vitro* setting $\implies f(z)$ constant
- Fit model on *in vitro* hepatocytes data (Atwood *et al.*, PNAS, 2011)

Clock model fit on *in vitro* hepatocytes data



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

Clock model fit on *in vitro* hepatocytes data



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

Multiple α, γ and $M(t) \Rightarrow$ multiple residual trajectories $y(t)$ for each gene / mouse class.

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

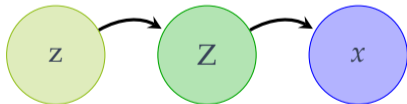
Linear regression

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

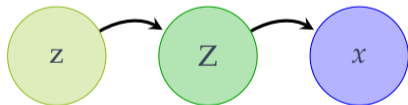
Linear regression

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

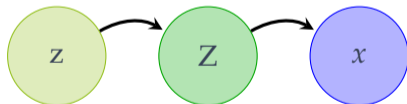
Linear regression

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: $z \leftarrow (z, Z)$

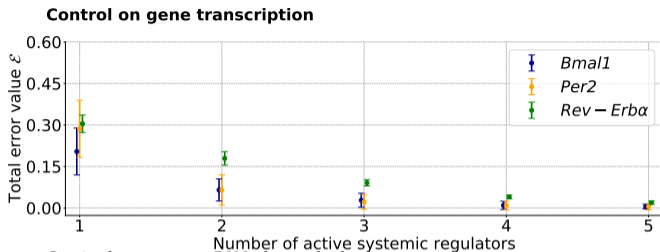


A regulator z_j and its integral Z_j are never found together in a model for all j

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

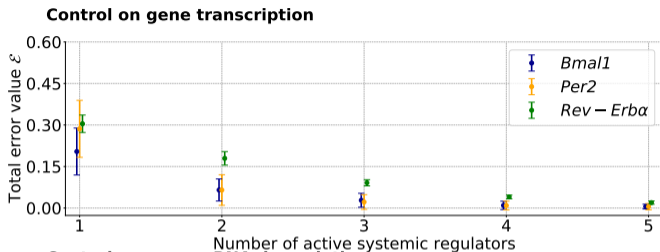


$$\mathcal{E}(y, \bar{z}) := \frac{1}{4n} \sum_{c=1}^4 \sum_{k=1}^n \min_{\beta_k^{(c)}} \ell(y_k^{(c)}, \bar{z}^{(c)}, \beta_k^{(c)})$$

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

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

Total error as a function of the number of involved regulators



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

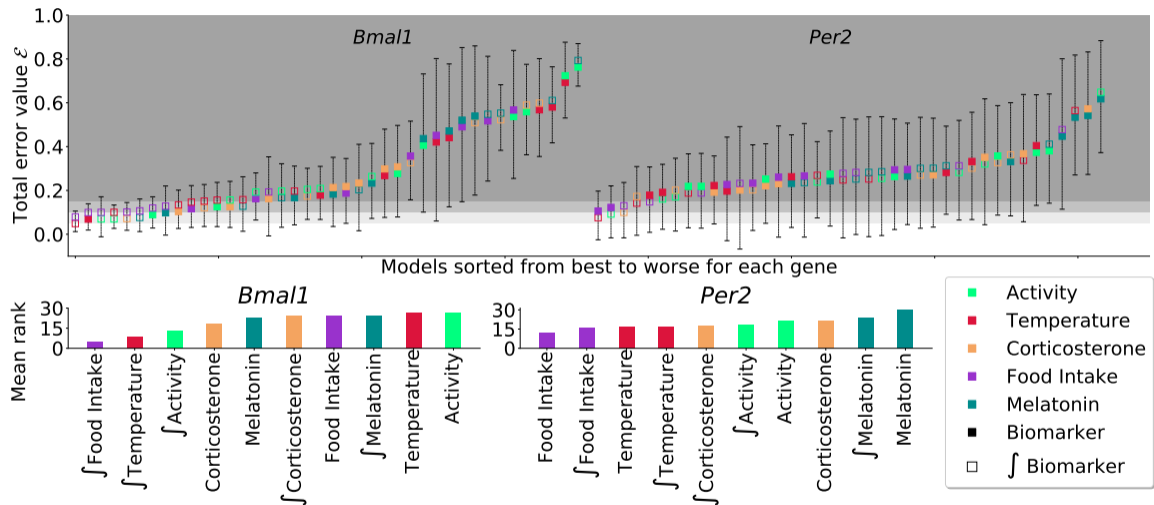
⇒ No linear control of regulators on *Rev-Erbα* transcription

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

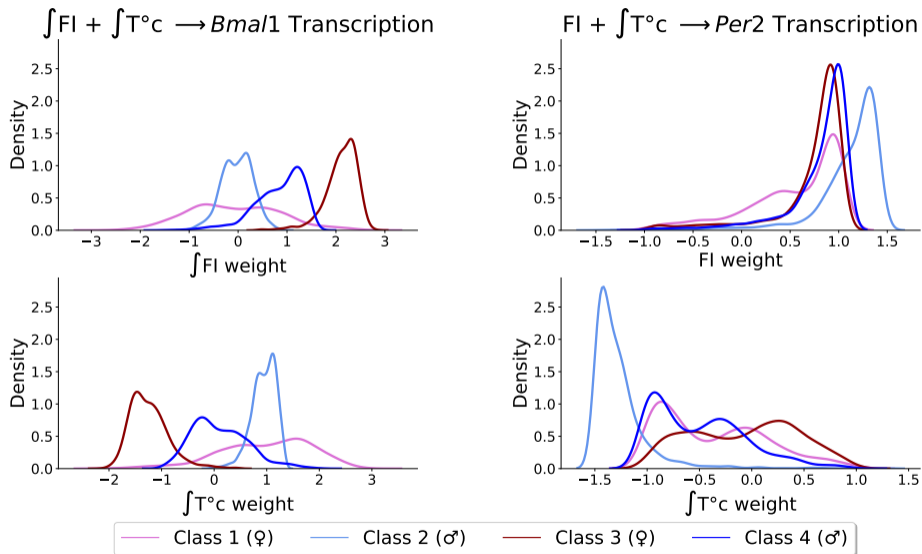
$$\ell(y_k^{(c)}, \bar{z}^{(c)}, \beta_k^{(c)}) := \frac{1}{N-1} \sum_{i=1}^{N-1} \left(y_k^{(c)}(t_i) - \sum_j \beta_{k,j}^{(c)} \bar{z}_j^{(c)}(t_i) \right)^2$$

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

2-term models ranking



Classwise weights analysis for best 2-term models



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

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

Conclusion & Perspectives

Under all hypotheses:

- Food Intake / $T^{\circ}\text{C}$ 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 to appear in *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>.