

Decoding regulation in ER+ breast cancer

A Data-Driven Dive into the ESR1-GATA3 Network



Michał Kloc

University Hospital Basel & University of Basel,
Swiss Institute of Bioinformatics,
June 19, 2025



University
of Basel



Swiss Institute of
Bioinformatics



Goals of this Presentation

Goals of this Presentation

- ① *Technical:* Introduce a method that combines mathematical modelling with neural networks (Neural ODEs, Universal ODEs...)

Goals of this Presentation

- ① *Technical:* Introduce a method that combines mathematical modelling with neural networks (Neural ODEs, Universal ODEs...)
- ② *Biological:* Apply the method to learn mechanistic model of mutual regulation of two transcription factors relevant in breast cancer

Goals of this Presentation

- ① *Technical:* Introduce a method that combines mathematical modelling with neural networks (Neural ODEs, Universal ODEs...)
- ② *Biological:* Apply the method to learn mechanistic model of mutual regulation of two transcription factors relevant in breast cancer
- ③ *Alert:* preliminary results!

ER+ Breast Cancer

- Breast Cancer: The most frequent cancer among women (1 in 8),

ER+ Breast Cancer

- Breast Cancer: The most frequent cancer among women (1 in 8),
- ER+ Breast Cancer: The most frequent subtype (70 - 80% of breast cancer patients), more common in post-menopausal women,

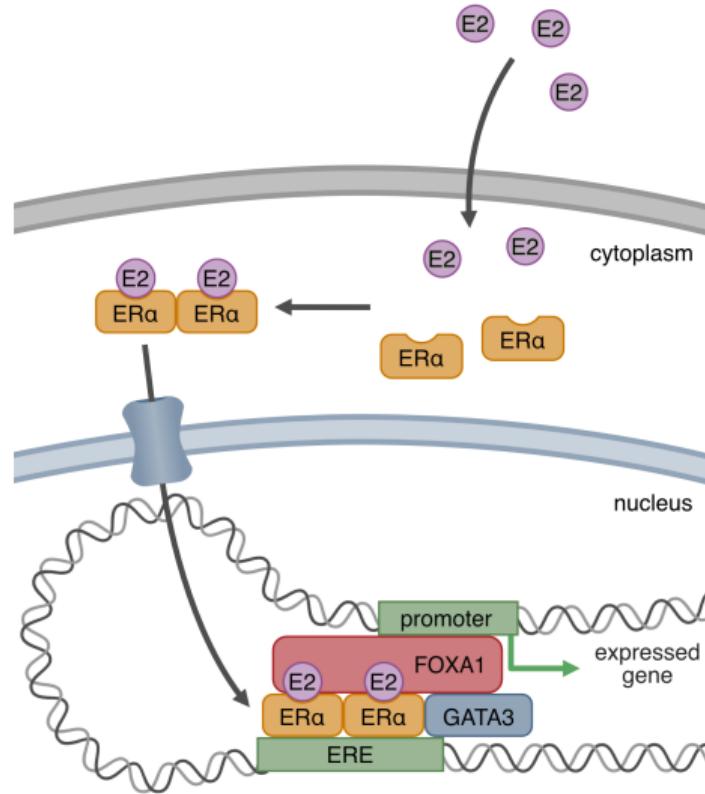
ER+ Breast Cancer

- Breast Cancer: The most frequent cancer among women (1 in 8),
- ER+ Breast Cancer: The most frequent subtype (70 - 80% of breast cancer patients), more common in post-menopausal women,
- Tumor proliferation conditioned on signaling via estrogen receptor (ER) that is ligand dependent (estradiol - E2),

ER+ Breast Cancer

- Breast Cancer: The most frequent cancer among women (1 in 8),
- ER+ Breast Cancer: The most frequent subtype (70 - 80% of breast cancer patients), more common in post-menopausal women,
- Tumor proliferation conditioned on signaling via estrogen receptor (ER) that is ligand dependent (estradiol - E2),
- *Endocrine therapy* as the backbone of treatment
 - *Tamoxifen* (Selective Estrogen Receptor Modulator) – Used in premenopausal women,
 - *Aromatase Inhibitors* (AIs) (e.g., Anastrozole, Letrozole, Exemestane) – Used in postmenopausal women,
 - *Fulvestrant* (Selective Estrogen Receptor Degrader) First-line treatment (in combination with targeted therapy) for metastatic or advanced ER+ breast cancer.

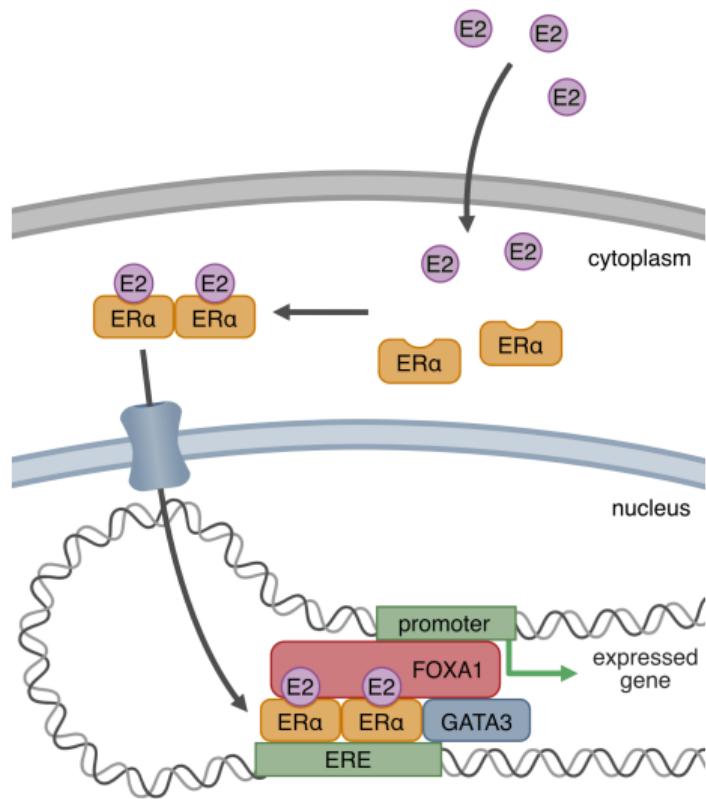
ER+ Breast Cancer



^aF. Yang et al., Mol. Cell 3 (2017)

^bJ. Lee et al., Nucleic acids research 42 (2014)

ER+ Breast Cancer

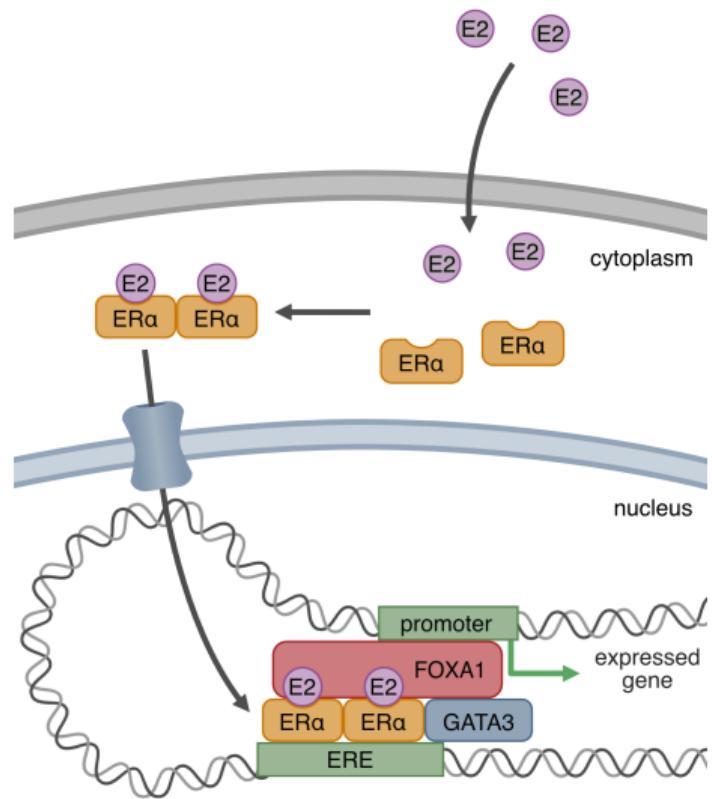


- annotation alert:
ESR1 (gene) → ER α (protein)

^aF. Yang et al., Mol. Cell 3 (2017)

^bJ. Lee et al., Nucleic acids research 42 (2014)

ER+ Breast Cancer

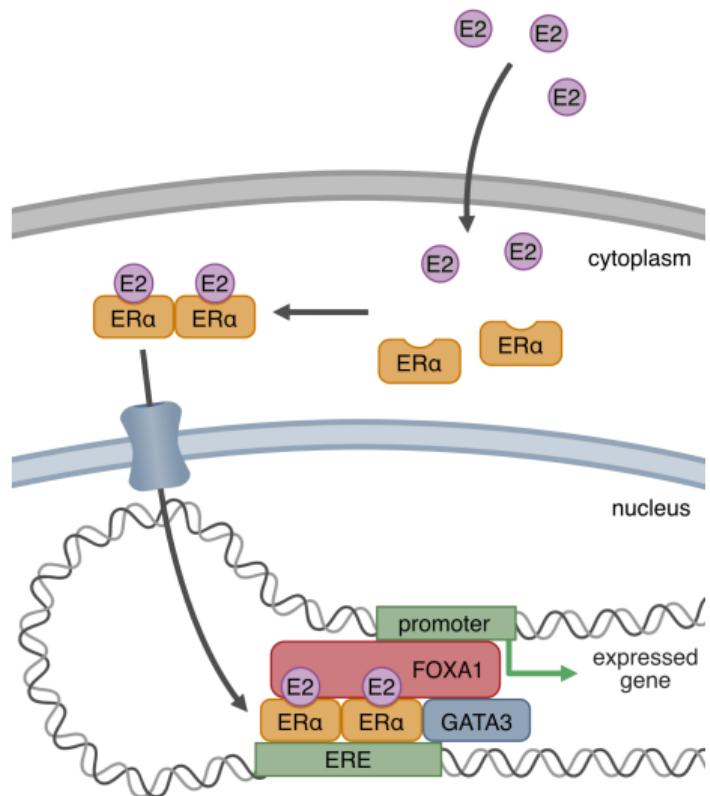


- *annotation alert:*
ESR1 (gene) → ER α (protein)
- transcriptional complex
ER α –GATA3–FOXA1^a

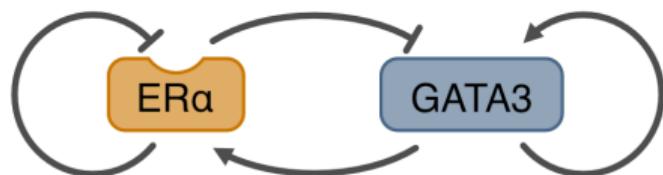
^aF. Yang et al., Mol. Cell 3 (2017)

J. Lee et al., Nucleic acids research 42 (2014)

ER+ Breast Cancer



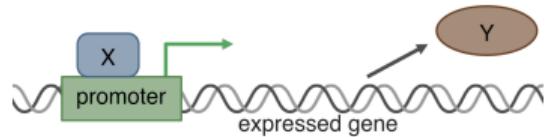
- *annotation alert:*
ESR1 (gene) → ER α (protein)
- transcriptional complex
ER α –GATA3–FOXA1^a
- Mutual regulation^b



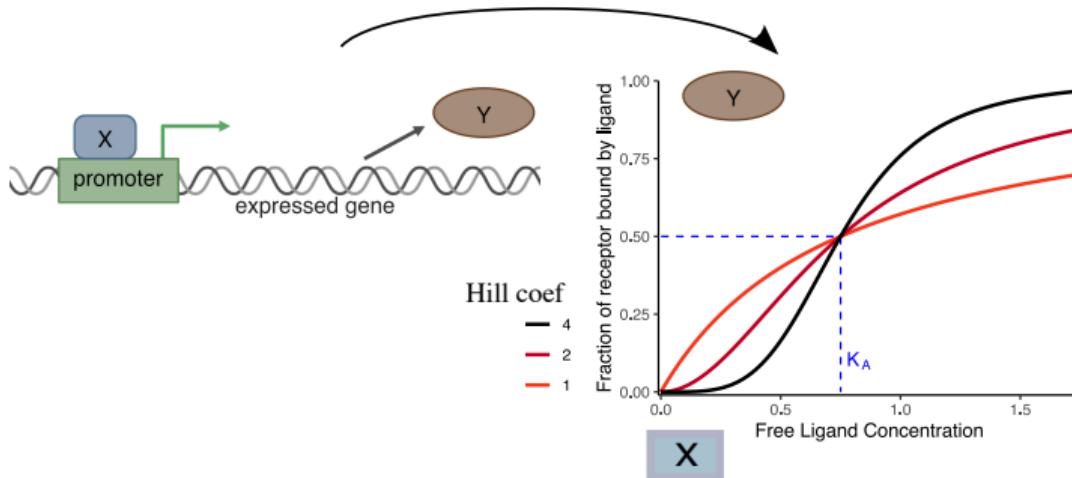
^aF. Yang et al., Mol. Cell 3 (2017)

^bJ. Lee et al., Nucleic acids research 42 (2014)

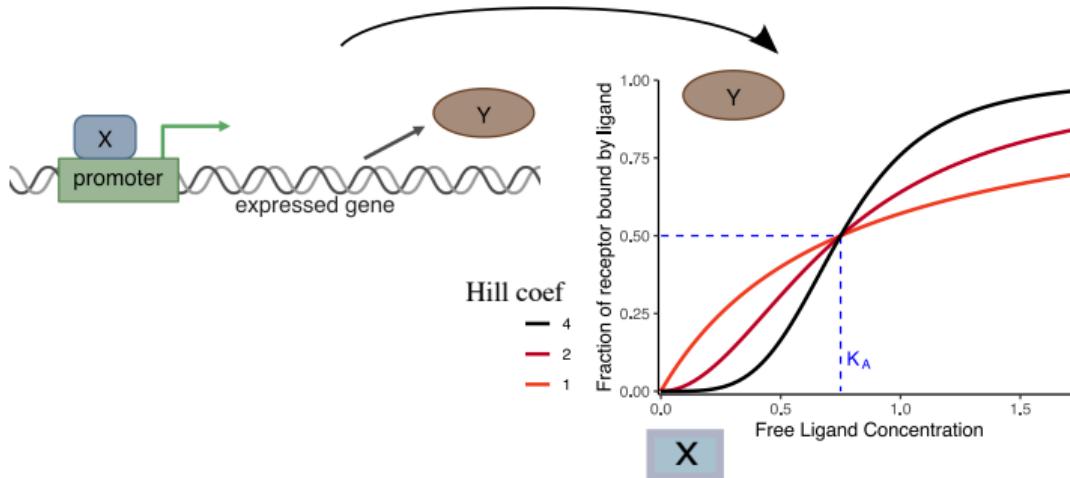
Michaelis-Menten and Hill Kinetics



Michaelis-Menten and Hill Kinetics



Michaelis-Menten and Hill Kinetics



Hill equation

$$H(x, K, n) = \frac{x^n}{K^n + x^n} = \frac{\left(\frac{x}{K}\right)^n}{1 + \left(\frac{x}{K}\right)^n}$$

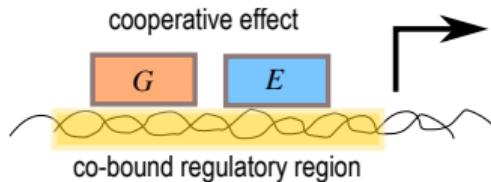
- x : concentration of the transcription factor
- K : half-activation concentration of x
- n : Hill coefficient: level of cooperativity (monomer, dimer...)

Mechanistic Model

Ordinary differential equations

$$\frac{dE}{dt} = \beta_E (1 - f_{EE} H(E, K_{EE}, n_{EE})) (1 + f_{GE} H(G, K_{GE}, n_{GE})) - \gamma_E E,$$
$$\frac{dG}{dt} = \beta_G (1 - f_{EG} H(E, K_{EG}, n_{EG})) (1 + f_{GG} H(G, K_{GG}, n_{GG})) - \gamma_G G.$$

Multiplicatively coupled feedback loops ¹



¹Tiwari and Igoshin, Phys. Biol. **9** (2012)

Mechanistic Model

Ordinary differential equations

$$\frac{dE}{dt} = \beta_E (1 - f_{EE} H(E, K_{EE}, n_{EE})) (1 + f_{GE} H(G, K_{GE}, n_{GE})) - \gamma_E E,$$

$$\frac{dG}{dt} = \beta_G (1 - f_{EG} H(E, K_{EG}, n_{EG})) (1 + f_{GG} H(G, K_{GG}, n_{GG})) - \gamma_G G.$$

- (E, G) : state vector, ESR1 and GATA3 concentrations
- γ_E, γ_G : protein decay rates
- β_E, β_G : baseline transcription rates
- $f_{T_1 T_2}$: effect size (fold change) of the target T_2 synthesis from T_1 binding on the regulatory region,
+: activator, -: repressor

Mechanistic Model

Ordinary differential equations

$$\frac{dE}{dt} = \beta_E (1 - f_{EE} H(E, K_{EE}, n_{EE})) (1 + f_{GE} H(G, K_{GE}, n_{GE})) - \gamma_E E,$$

$$\frac{dG}{dt} = \beta_G (1 - f_{EG} H(E, K_{EG}, n_{EG})) (1 + f_{GG} H(G, K_{GG}, n_{GG})) - \gamma_G G.$$

- (E, G) : state vector, ESR1 and GATA3 concentrations
- γ_E, γ_G : protein decay rates
- β_E, β_G : baseline transcription rates
- $f_{T_1 T_2}$: effect size (fold change) of the target T_2 synthesis from T_1 binding on the regulatory region,
+: activator, -: repressor

Mechanistic Model

Ordinary differential equations

$$\frac{dE}{dt} = \beta_E (1 - f_{EE} H(E, K_{EE}, n_{EE})) (1 + f_{GE} H(G, K_{GE}, n_{GE})) - \gamma_E E,$$

$$\frac{dG}{dt} = \beta_G (1 - f_{EG} H(E, K_{EG}, n_{EG})) (1 + f_{GG} H(G, K_{GG}, n_{GG})) - \gamma_G G.$$

- (E, G) : state vector, ESR1 and GATA3 concentrations
- γ_E, γ_G : protein decay rates
- β_E, β_G : baseline transcription rates
- $f_{T_1 T_2}$: effect size (fold change) of the target T_2 synthesis from T_1 binding on the regulatory region,
+: activator, -: repressor

Mechanistic Model

Ordinary differential equations

$$\frac{dE}{dt} = \beta_E (1 - f_{EE} H(E, K_{EE}, n_{EE})) (1 + f_{GE} H(G, K_{GE}, n_{GE})) - \gamma_E E,$$

$$\frac{dG}{dt} = \beta_G (1 - f_{EG} H(E, K_{EG}, n_{EG})) (1 + f_{GG} H(G, K_{GG}, n_{GG})) - \gamma_G G.$$

- (E, G) : state vector, ESR1 and GATA3 concentrations
- γ_E, γ_G : protein decay rates
- β_E, β_G : baseline transcription rates
- $f_{T_1 T_2}$: effect size (fold change) of the target T_2 synthesis from T_1 binding on the regulatory region,
+: activator, -: repressor

Mechanistic Model

Ordinary differential equations

$$\frac{dE}{dt} = \beta_E (1 - f_{EE} H(E, K_{EE}, n_{EE})) (1 + f_{GE} H(G, K_{GE}, n_{GE})) - \gamma_E E,$$

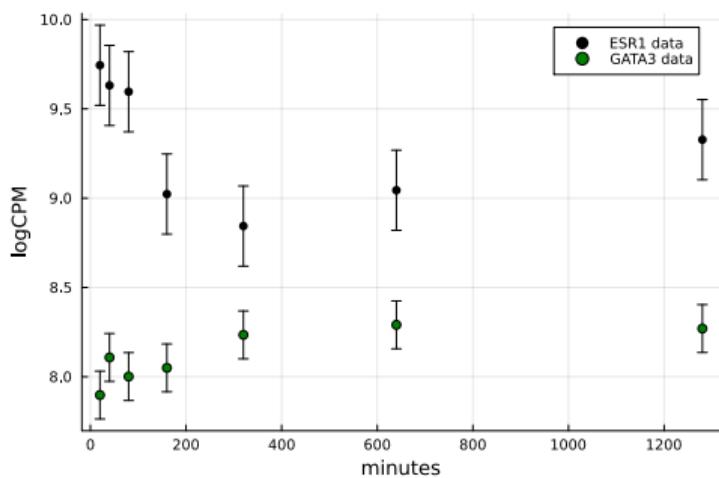
$$\frac{dG}{dt} = \beta_G (1 - f_{EG} H(E, K_{EG}, n_{EG})) (1 + f_{GG} H(G, K_{GG}, n_{GG})) - \gamma_G G.$$

- fit the model with phenomenological parameters¹ (20 hours, 14 Western blot measurements)
Example: $n_{GG} \approx 10$, $n_{EE} \approx 1.3$
- *Fixed Biological Parameters: Can the Model Still Fit?*

¹J. Lee et al., Nucleic acids research 42 (2014)

Solving the Inverse Problem

RNA-seq data²: MCF7 upon E2 stimulation



- Fix known biological coefficients

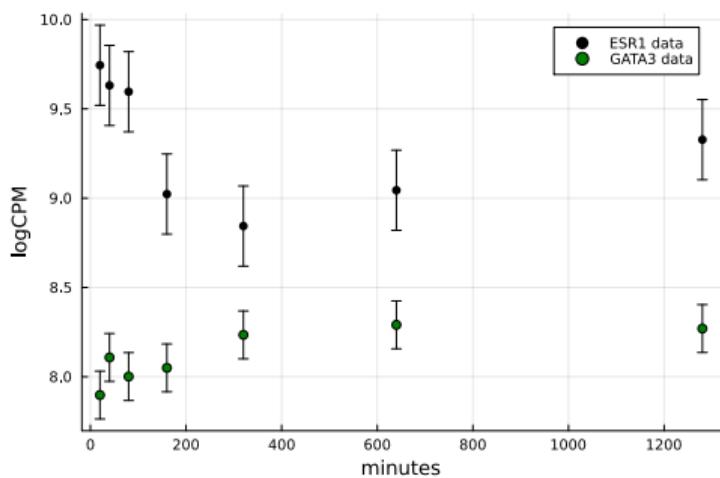
$$n_{EE} = n_{EG} = 2, \quad n_{GE} = n_{GG} = 1,$$
$$\gamma_G = 1h, \gamma_E = 3.6h,$$

- Gradient-based optimization over N measured datapoints with objective (loss function) $\mathcal{L} = c_1\mathcal{L}_1 + c_2\mathcal{L}_2$
- \mathcal{L}_1 : Weighted Squared Error,
 \mathcal{L}_2 : Restriction on parameters

²GSE62789, Honkela et al. (2015)

Solving the Inverse Problem

RNA-seq data²: MCF7 upon E2 stimulation



- Fix known biological coefficients

$$n_{EE} = n_{EG} = 2, \quad n_{GE} = n_{GG} = 1,$$
$$\gamma_G = 1h, \gamma_E = 3.6h,$$

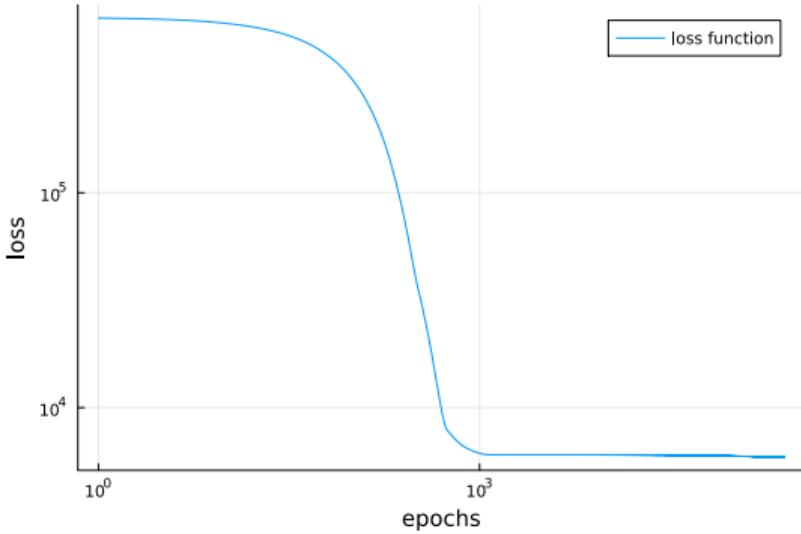
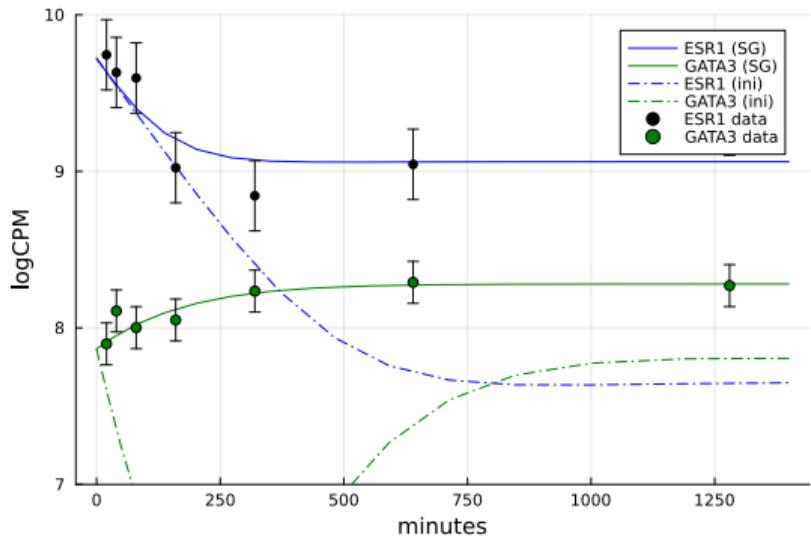
- Gradient-based optimization over N measured datapoints with objective (loss function) $\mathcal{L} = c_1 \mathcal{L}_1 + c_2 \mathcal{L}_2$
- \mathcal{L}_1 : Weighted Squared Error,
 \mathcal{L}_2 : Restriction on parameters

$$\mathcal{L}_1 = \sum_{i=1}^N \left[\frac{y_i - \hat{y}_i}{\sigma} \right]^2, \quad \mathcal{L}_2 = \sum_{i \in \{EE, EG, GG, GE\}} |K_i - K_i^{ref}|^2,$$

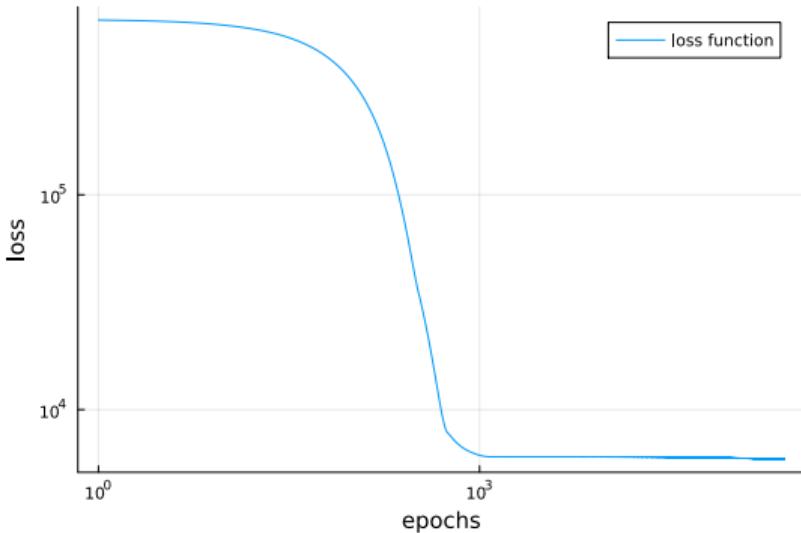
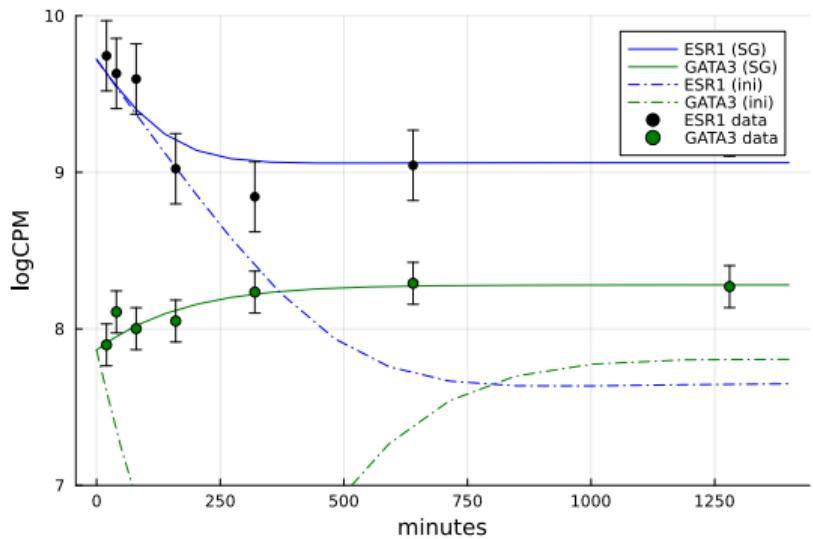
K_i^{ref} : average expressions of E and G

²GSE62789, Honkela et al. (2015)

Solving the Inverse Problem

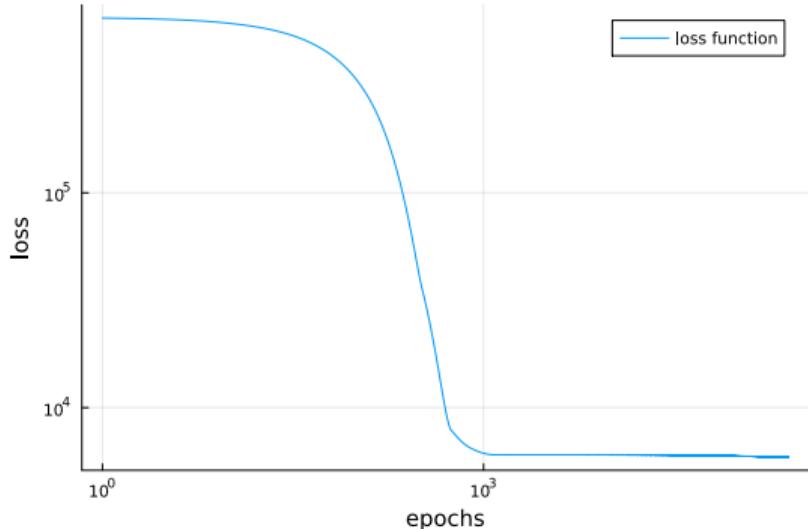
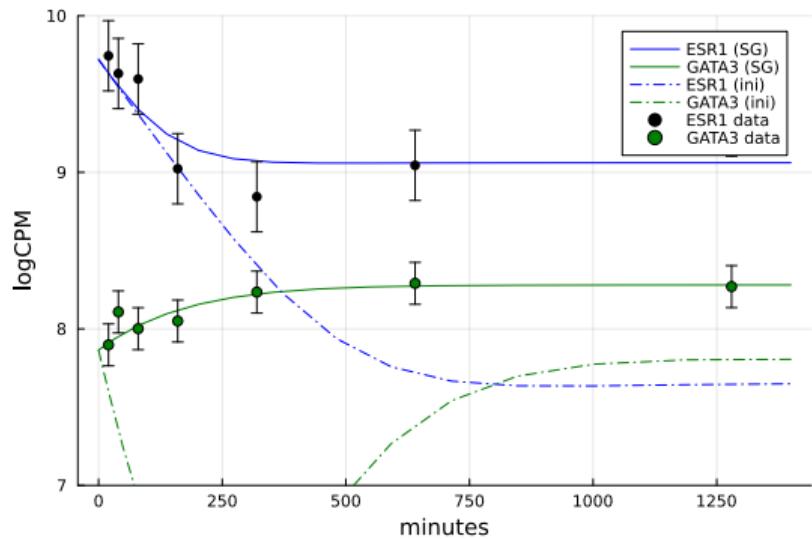


Solving the Inverse Problem



Not bad (the model fits the average behavior)

Solving the Inverse Problem



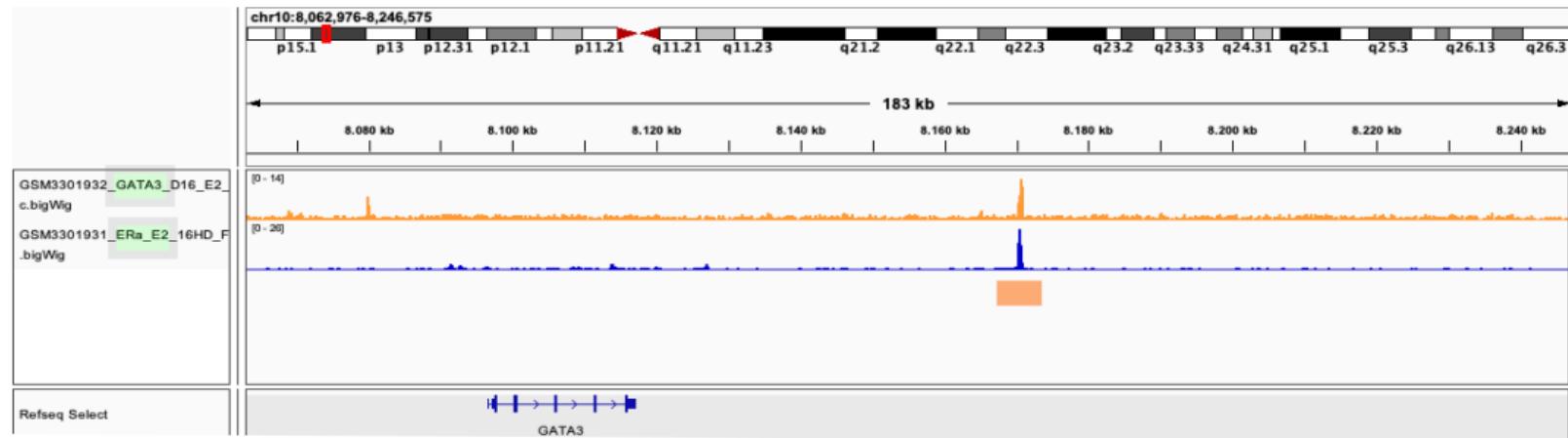
Not bad (the model fits the average behavior)
but also not good! (the model does not capture the detailed features of the dynamics)

Are We Missing Something in the Model?

³GSE117492, Nair et. al, Nat Struct Mol Biol **26** (2019)

Are We Missing Something in the Model?

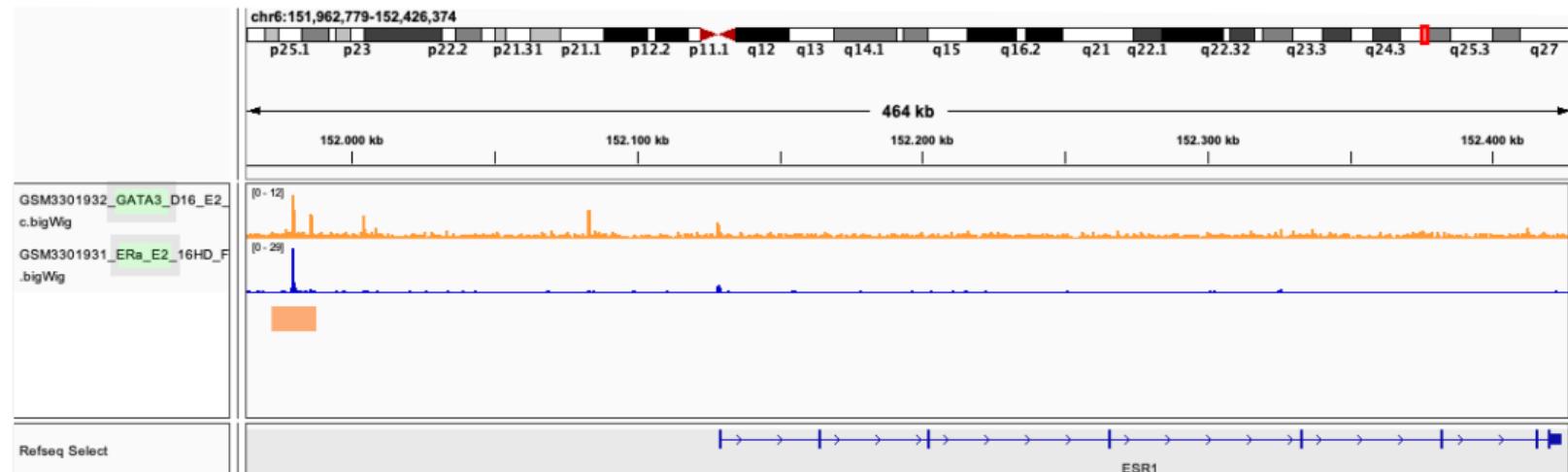
GATA3 regulatory region, ChIPseq data³



³GSE117492, Nair et. al, Nat Struct Mol Biol 26 (2019)

Are We Missing Something in the Model?

ESR1 regulatory region, ChIPseq data⁴



⁴GSE117492, Nair et. al, Nat Struct Mol Biol 26 (2019)

Learn Missing Dynamics from the Data

*"A model is worth a thousand datasets"*⁵ ⁶

⁵Rackauckas et. al, arXiv:2001.04385v4 (2020)

⁶Chen, et al, Advances in neural information processing systems **31** (2018)

Learn Missing Dynamics from the Data

"A model is worth a thousand datasets"⁵ ⁶

Neural Ordinary differential equations (NODEs)

$$\frac{dE}{dt} = \beta_E \left[(1 - f_{EE} H(E, K_{EE}, n_{EE})) (1 + f_{GE} H(G, K_{GE}, n_{GE})) + \textcolor{blue}{NN_\Theta(E, G)} \right] - \gamma_E E,$$

$$\frac{dG}{dt} = \beta_G (1 - f_{EG} H(E, K_{EG}, n_{EG})) (1 + f_{GG} H(G, K_{GG}, n_{GG})) - \gamma_G G.$$

⁵Rackauckas et. al, arXiv:2001.04385v4 (2020)

⁶Chen, et al, Advances in neural information processing systems 31 (2018)

Learn Missing Dynamics from the Data

"A model is worth a thousand datasets"^{5 6}

Neural Ordinary differential equations (NODEs)

$$\frac{dE}{dt} = \beta_E [(1 - f_{EE} H(E, K_{EE}, n_{EE})) (1 + f_{GE} H(G, K_{GE}, n_{GE})) + \textcolor{blue}{NN}_{\Theta}(E, G)] - \gamma_E E,$$
$$\frac{dG}{dt} = \beta_G (1 - f_{EG} H(E, K_{EG}, n_{EG})) (1 + f_{GG} H(G, K_{GG}, n_{GG})) - \gamma_G G.$$

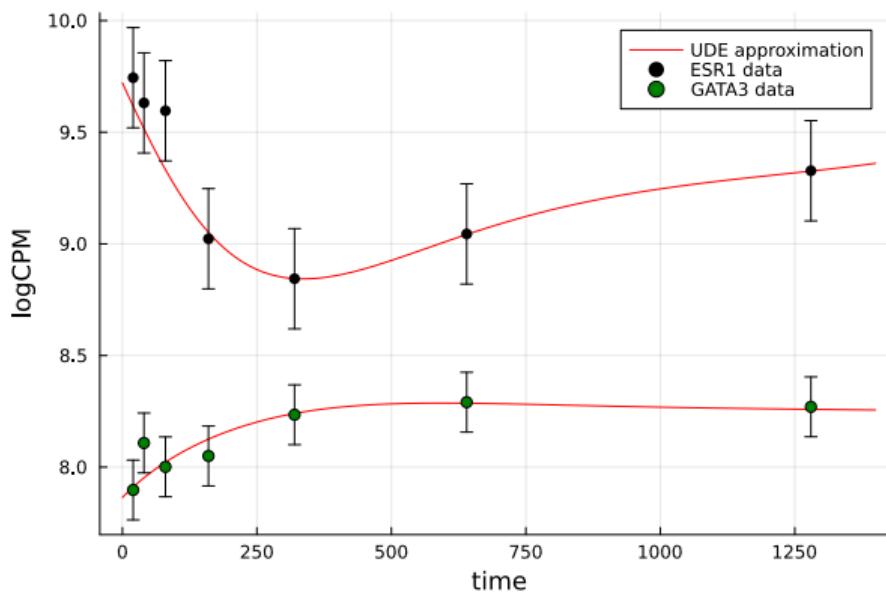
- *Universal differential equations (UDEs):* NN learns correction to our mechanistic model to approximate the data.

```
#####
# prepare the neural network with small weights
# Construct a neural network using Lux
nn = Lux.Chain(
    Lux.Dense(2 => 8, tanh; init_weight=truncated_normal(; std=1e-3)),
    Lux.Dense(8 => 8, tanh; init_weight=truncated_normal(; std=1e-3)),
    Lux.Dense(8 => 8, tanh; init_weight=truncated_normal(; std=1e-3)),
    Lux.Dense(8 => 1;  init_weight=truncated_normal(; std=1e-3))
)
```

⁵Rackauckas et. al, arXiv:2001.04385v4 (2020)

⁶Chen, et al, Advances in neural information processing systems 31 (2018)

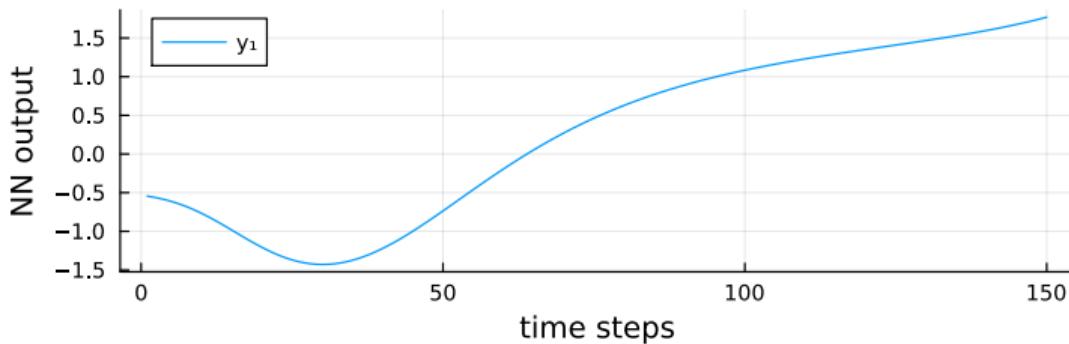
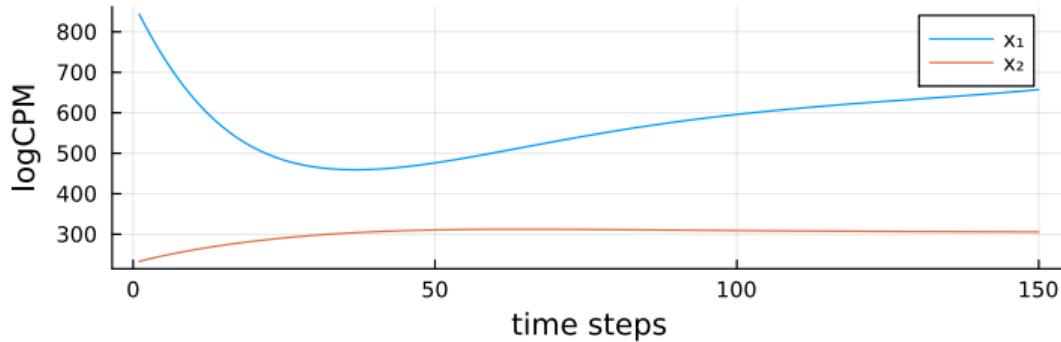
Learn Missing Dynamics from the Data



Learn Missing Dynamics from the Data

What does the
NN actually do?

$x_1 := \text{ESR1}$,
 $x_2 := \text{GATA3}$



Sparse symbolic regression

- We can be happy with the NN correction and test the model for predictions, but extrapolation using the NN might still be limited (many parameters)

Sparse symbolic regression

- We can be happy with the NN correction and test the model for predictions, but extrapolation using the NN might still be limited (many parameters)
- Idea: try to approximate the NN as a combination of rational functions (implicit SINDy)⁷ \rightsquigarrow simplify the predictor to "expected dynamics"

⁷J. Mangan et al., IEEE Transactions on Molecular, Biological, and Multi-Scale Communications 2 (2017)

Sparse symbolic regression

- We can be happy with the NN correction and test the model for predictions, but extrapolation using the NN might still be limited (many parameters)
- Idea: try to approximate the NN as a combination of rational functions (implicit SINDy)⁷ \rightsquigarrow simplify the predictor to "expected dynamics"

$$NN(E, G) \approx \frac{f_N(E, G)}{f_D(E, G)},$$

$f(E, G)$ expand in terms E^2, G, E^2G, G^2 ,
apply sparse regularization to the regression

⁷J. Mangan et al., IEEE Transactions on Molecular, Biological, and Multi-Scale Communications 2 (2017)

Recovered dynamics

$$\begin{aligned}\frac{dE}{dt} &= \beta_E [(1 - f_{EE} H(E, K_{EE}, n_{EE})) (1 + f_{GE} H(G, K_{GE}, n_{GE})) + NN(E, G)] - \gamma_E E, \\ \frac{dG}{dt} &= \beta_G (1 - f_{EG} H(E, K_{EG}, n_{EG})) (1 + f_{GG} H(G, K_{GG}, n_{GG})) - \gamma_G G.\end{aligned}$$

Polynomial approximation

$$NN(E, G) \approx \frac{p_1 E^2 + p_2 G + p_3 G^2}{-p_4 G - p_5 G^2}$$

with parameters p_i from the symbolic regression.

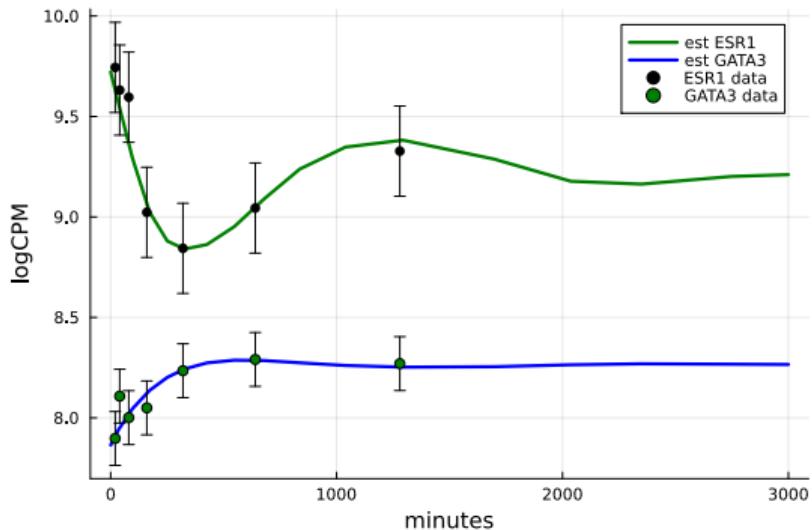
Recovered dynamics

$$\frac{dE}{dt} = \beta_E [(1 - f_{EE} H(E, K_{EE}, n_{EE})) (1 + f_{GE} H(G, K_{GE}, n_{GE})) + NN(E, G)] - \gamma_E E,$$
$$\frac{dG}{dt} = \beta_G (1 - f_{EG} H(E, K_{EG}, n_{EG})) (1 + f_{GG} H(G, K_{GG}, n_{GG})) - \gamma_G G.$$

Polynomial approximation

$$NN(E, G) \approx \frac{p_1 E^2 + p_2 G + p_3 G^2}{-p_4 G - p_5 G^2}$$

with parameters p_i from the symbolic regression.



What next

- ① Now: "Proof of concept" and application of UDEs in biological networks,

What next

- ① Now: "Proof of concept" and application of UDEs in biological networks,
- ② Ongoing: generating better dataset (qPCR on ER/GATA3 in MCF7 24h, evenly distributed)

What next

- ① Now: "Proof of concept" and application of UDEs in biological networks,
- ② Ongoing: generating better dataset (qPCR on ER/GATA3 in MCF7 24h, evenly distributed)
- ③ Later time points (around 48h) for validation

What next

- ① Now: "Proof of concept" and application of UDEs in biological networks,
- ② Ongoing: generating better dataset (qPCR on ER/GATA3 in MCF7 24h, evenly distributed)
- ③ Later time points (around 48h) for validation
- ④ Can we simulate perturbation with Fulvestrant (ESR1 degrader) via it's effect on γ_E ?

Acknowledgment



Momo Bentires-Alj team

- Charly Jehanno, Martino Bernasconi

Bioinformatics Core Facility, DBM

- Florian Geier, Anastasiya Börsch, Robert Ivánek, Igor Červenka

Julia lab, MIT

- Frank Schäfer
(now Axiomatic AI)