



UNIVERSITÀ
DI TORINO

di.unito.it
DIPARTIMENTO
DI INFORMATICA



From Data to Models: Analyzing and Integrating Biological Data into Mechanistic Models

Dora Tortarolo, Simone Pernice, Francesca Cordero

Course of Doctoral school

PhD in Complex Systems for Quantitative Biomedicine

When?

| Sunday | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday |
|--------|--------|---------|-----------|----------|---|----------|
| 29 | 30 | 1 | 2 | 3 | • 09:00 - 13:00 Seminari  | 4 |
| 6 | 7 | 8 | 9 | 10 | • 09:00 - 13:00 Sala conferenze, terzo piano  | 11 |
| 13 | 14 | 15 | 16 | 17 | 18 | 19 |

DAY 2: Outline

From Case Study to Data Analysis

- From Data to Models: the importance of model parameterization using real (raw) experimental data.
- Real-World Case Study: understanding the problem context.
- Hands-on Data Analysis using ORCA.
- Real-World Case Study: model definition.

Materials: https://drive.google.com/drive/folders/1mGFzbS1nflQ80TwQxC0FIkY0Mz6y_j7i?usp=drive_link



<https://github.com/qBioTurin/From-Data-to-Models.git>

Brief Recap: different modeling approaches

Statistical / Data-Driven Models

These models aim to uncover patterns, correlations, and predictions from data. They do **not explicitly model biological mechanisms**, and instead rely on data structure.

Exploratory: Identify structures or patterns in data without prediction (PCA, clustering)

Inferential: Use statistical inference to test hypotheses or estimate parameters (GLMs, ANOVA)

Machine Learning: Learn patterns from unlabeled/labeled data for classification/regression(SVM, RF, DL)

Mechanistic / Knowledge-Driven Models

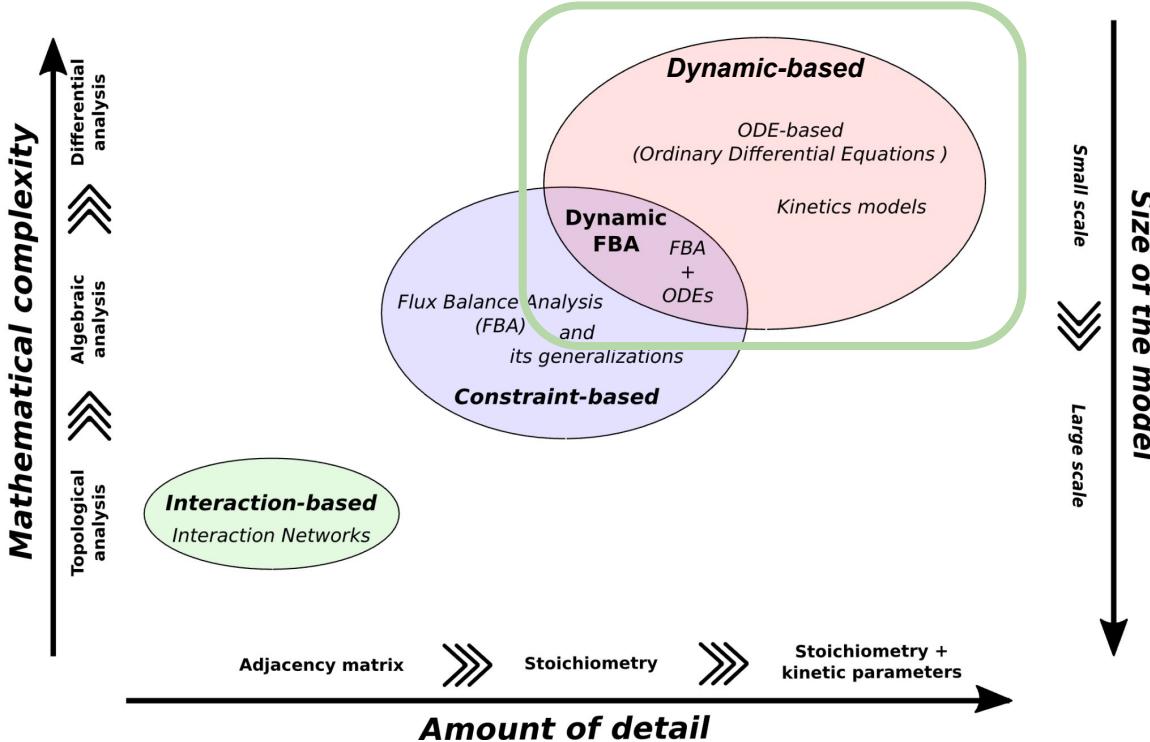
These models are grounded in **explicit biological rules or hypotheses**, often formulated with equations or logic. They aim to explain causal behavior, not just correlations.

Interaction-based: network reconstruction processes that yield representations capturing structural information only (protein-protein interaction networks).

Mechanism-based:

- **Constraint-based (no dynamics):** commonly exploited to study systems under specific constraints (considered "mechanistic-lite" without dynamics)
- **Dynamical (with dynamics):** Models that represent explicit mechanistic hypotheses, including temporal dynamics.

Brief Recap: mechanistic modeling approaches



Brief Recap: Petri Net: graphical formalism to draw models

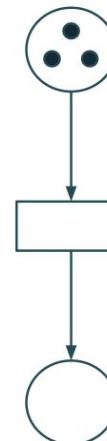
Petri Net (PN) is graphical formalism (bipartite graph), conveniently used for the analysis of complex models. It allows us to derive qualitative and quantitative properties of the system.

Places (circles): Represent *states* or *conditions* (e.g., a molecule present).

Transitions (rectangles): Represent *events* or *actions* that change the state (e.g., reactions).

Tokens (dots inside places): Indicate the *current state* (how many molecule are in a place).

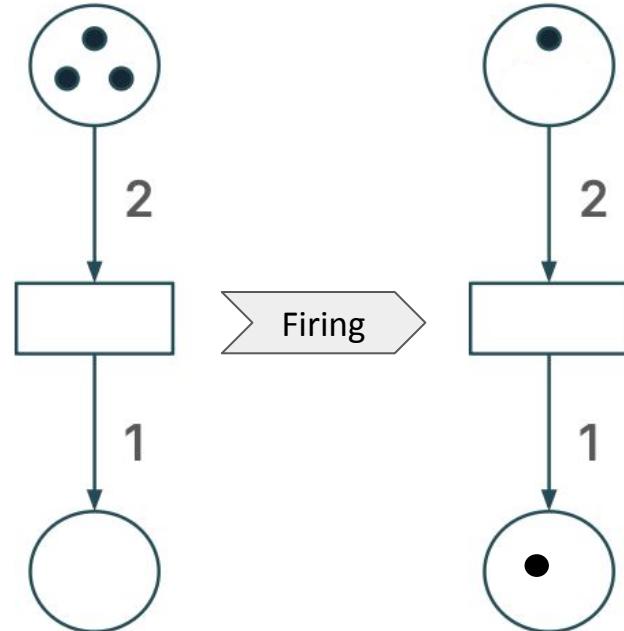
Arcs (arrows): Show how places and transitions are connected (direction matters).



Brief Recap: Petri Net: graphical formalism to draw models

Arcs Multiplicity:

- Arcs can have multiplicities, indicating the number of tokens transferred between places and transitions.



Transitions Firing:

- A transition 'fires' when the input places contain enough tokens based on the arc multiplicities.
- Firing consumes tokens from input places and produces tokens in output place

Firing Rate:

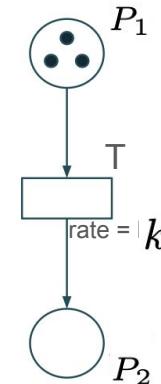
- Determines how quickly a transition converts input tokens into output tokens

Brief Recap: transition velocities

The transition firings velocity is derived from the **propensity function!**

Mass Action (The Default)

- **Mass Action kinetics** assumes that transition velocities are **proportional** to the number of **available tokens** from the **input places**.
- This is a **simplifying assumption**, often valid when:
 - All molecules are well-mixed.
 - Transitions occur freely without spatial/structural constraints.



$$a_T(x) = x_{P_1} * k$$

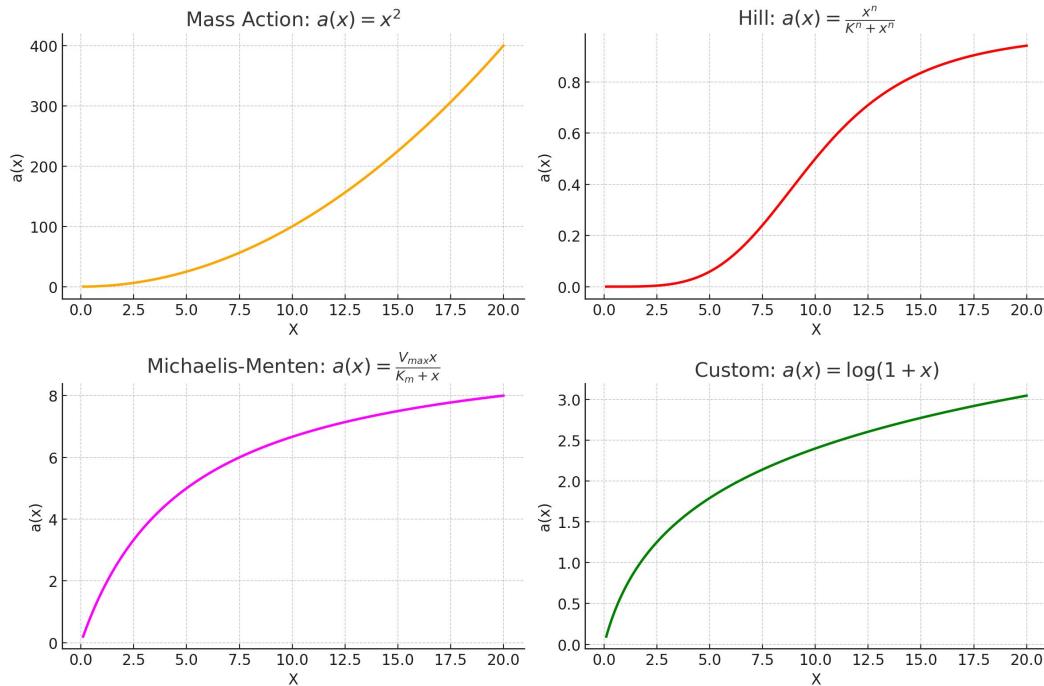
$$x = [x_{P_1}, x_{P_2}] \longrightarrow \text{The number of tokens in each place}$$

Brief Recap: transition velocities

The transition firings velocity is derived from the **propensity function!**

Many biological systems **do not follow mass action**:

- **Enzyme reactions** reach saturation → **Michaelis-Menten**
- **Gene regulation** often shows switch-like behavior → **Hill kinetics**
- **Feedback or inhibition** may lead to nonlinear or **custom rates**



Translating biological knowledge into a computational model

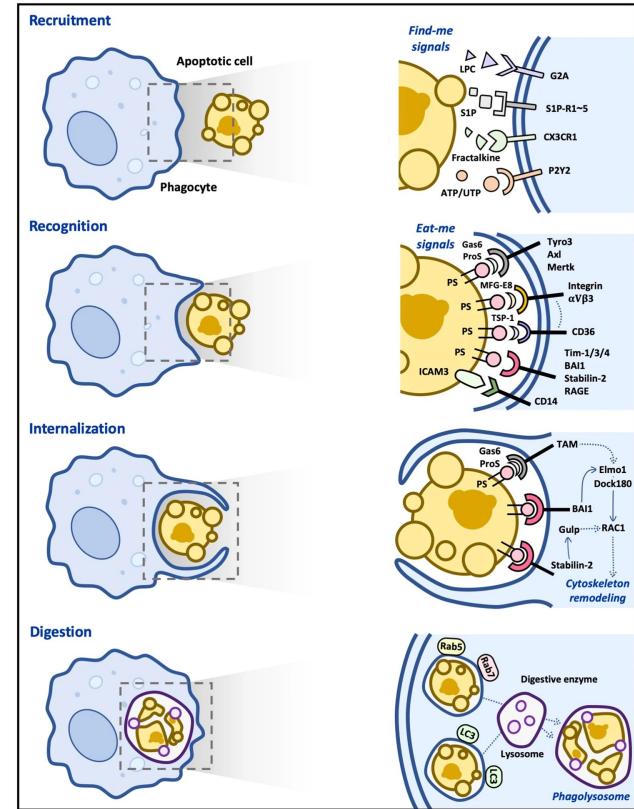
Model construction:

- translating biological graphical models into formal models
- multi-scale complexity: different levels of complexity and how they might interact across scales and levels of organization

Data integration: models are dependent on quantitative data (e.g. lab experiments, omics, imaging)

data is:

- required for parameter estimation, validation and prediction
- sparse and noisy, with limited time points
- can come from heterogeneous sources (in-house and from the literature) → can you compare it?
- can come from different types of experiments of different types of molecules → can you merge different types measurement units?
- is often a relative abundance data, but models require absolute values
- some data is unavailable and must be estimated (kinetic rates, binding constants)

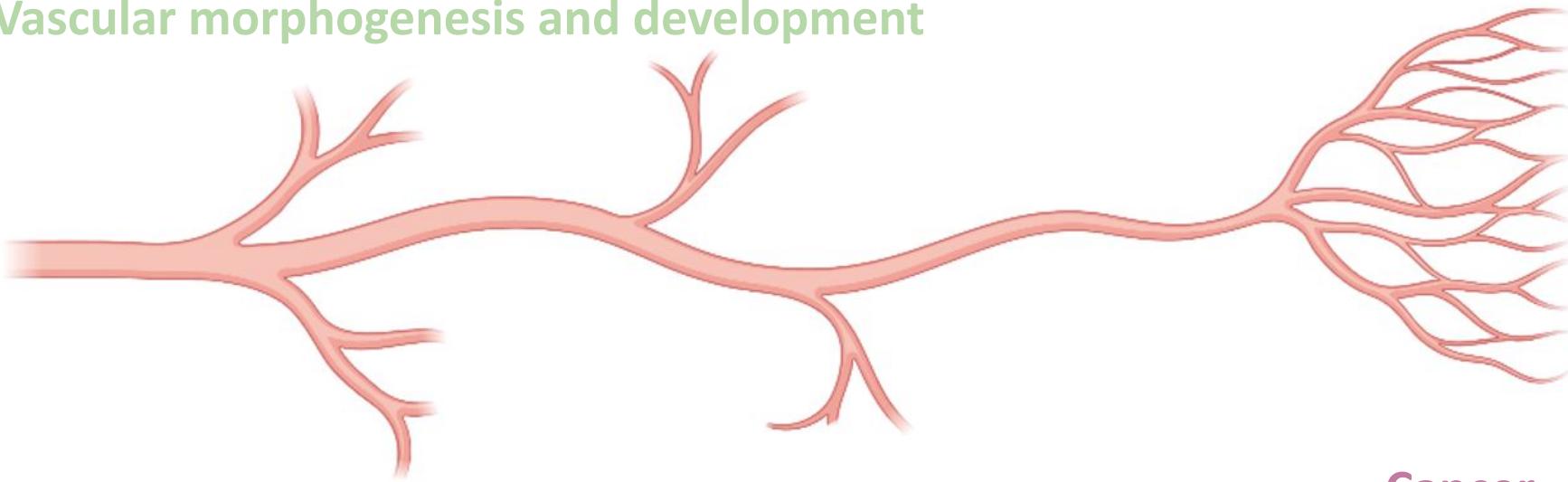


Byeongjin Moon et al., Experimental & Molecular Medicine, 2023

Real world case study: endothelial cell polarity

Endothelial cell polarity is **crucial**
for physiological and pathological processes

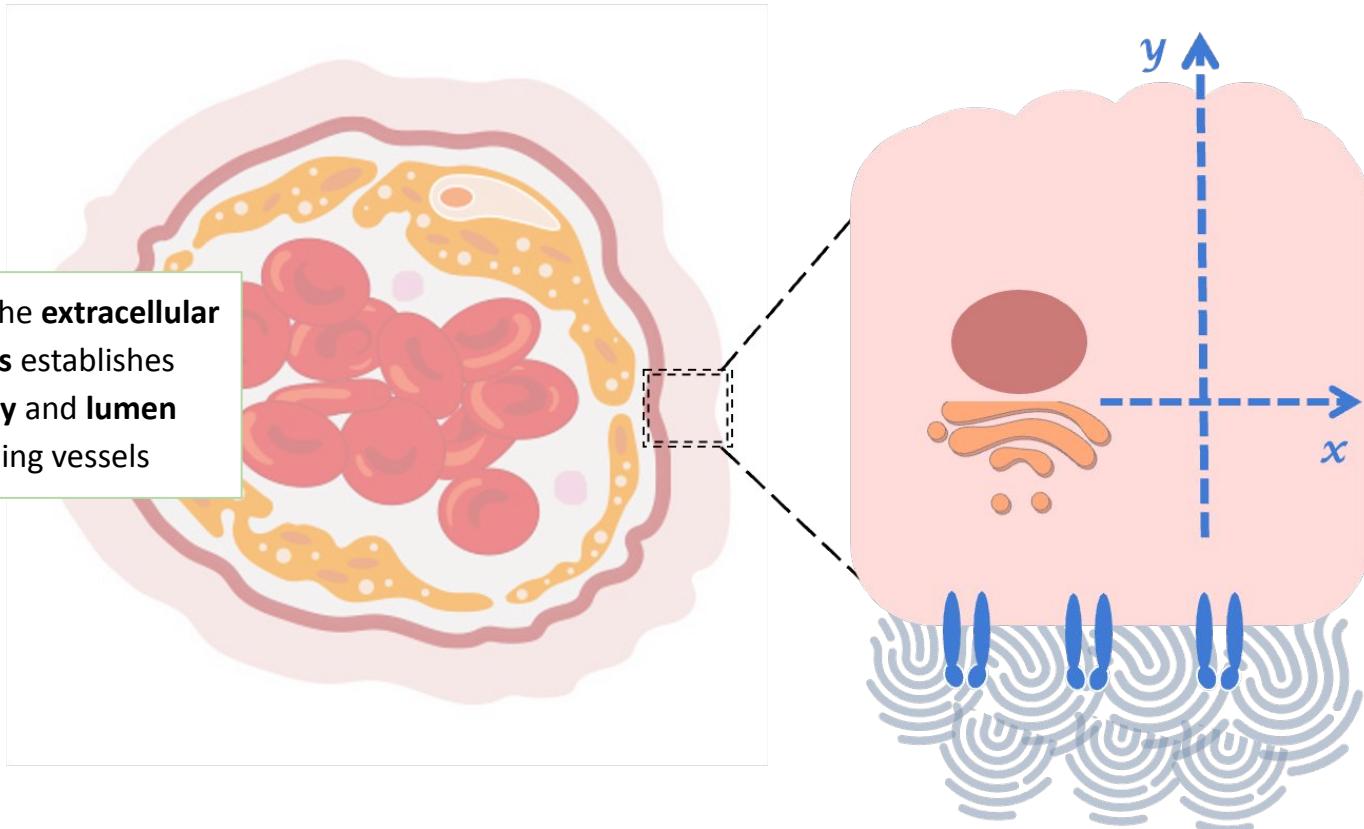
Vascular morphogenesis and development



Cancer

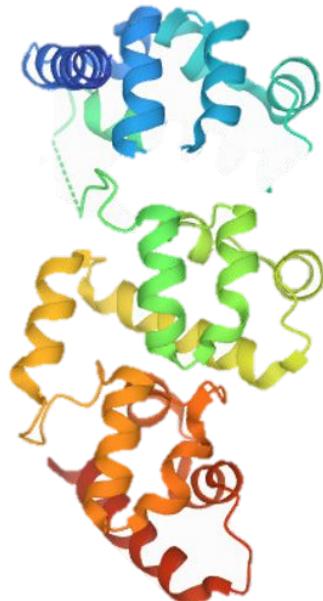
Real world case study: endothelial cell polarity

The interaction between the **extracellular matrix** and **$\beta 1$ integrins** establishes **endothelial cell polarity** and **lumen formation** in developing vessels

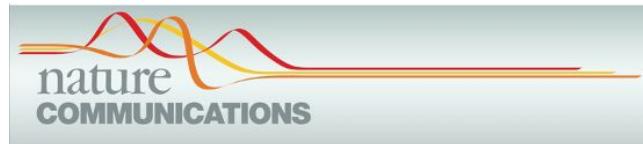


Real world case study: endothelial cell polarity

PTPRF-interacting protein $\alpha 1$ (PPFIA1) is a key mediator of EC polarity



Crystal structure of
PPFIA1 SAM domain



ARTICLE

Received 23 Aug 2015 | Accepted 13 Oct 2016 | Published 23 Nov 2016

DOI: [10.1038/ncomms13546](https://doi.org/10.1038/ncomms13546)

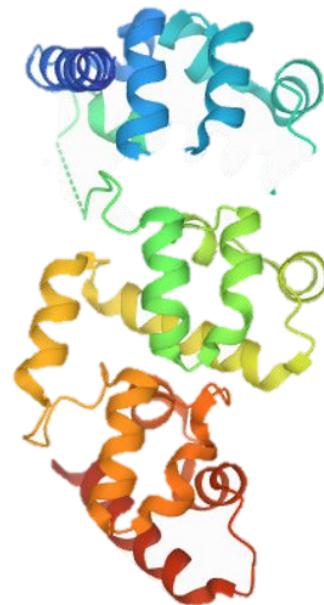
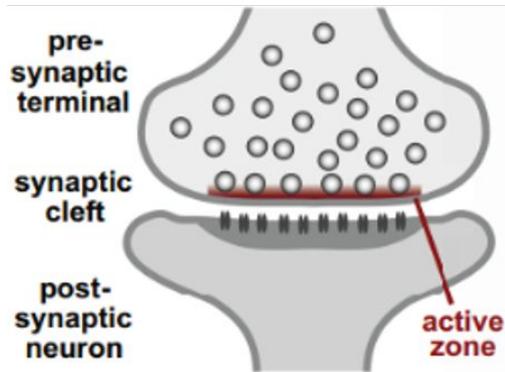
OPEN

PPFIA1 drives active $\alpha 5\beta 1$ integrin recycling and controls fibronectin fibrillogenesis and vascular morphogenesis

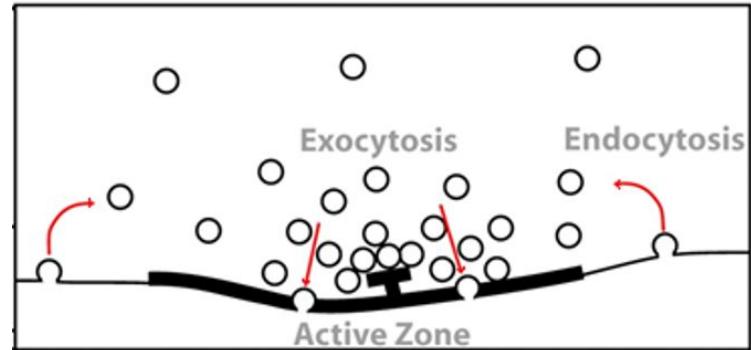
Giulia Mana^{1,2,*}, Fabiana Clapero^{1,2,*}, Emiliano Panieri³, Valentina Panero³, Ralph T. Böttcher⁴, Hui-Yuan Tseng⁴, Federico Saltarin^{1,2}, Elena Astanina^{1,5}, Katarzyna I. Wolanska⁶, Mark R. Morgan⁶, Martin J. Humphries⁷, Massimo M. Santoro^{3,8}, Guido Serini^{1,2,**} & Donatella Valdembri^{1,2,**}

Real world case study: endothelial cell polarity

PPFIA1 docks neurotransmitter-containing vesicles in synapses

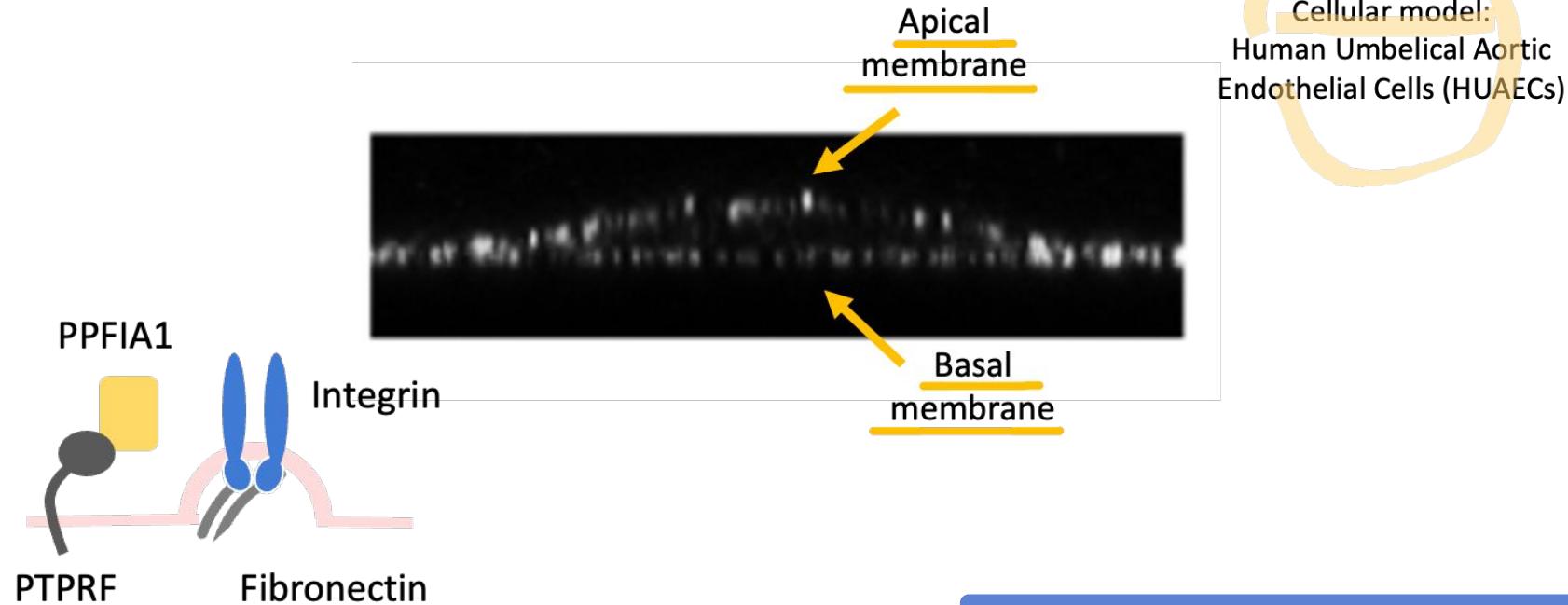


Crystal structure of
PPFIA1 SAM domain



Real world case study: endothelial cell polarity

PPFIA1 docks vesicles containing cellular adhesion receptors on the basolateral membrane of ECs



Polarized cell



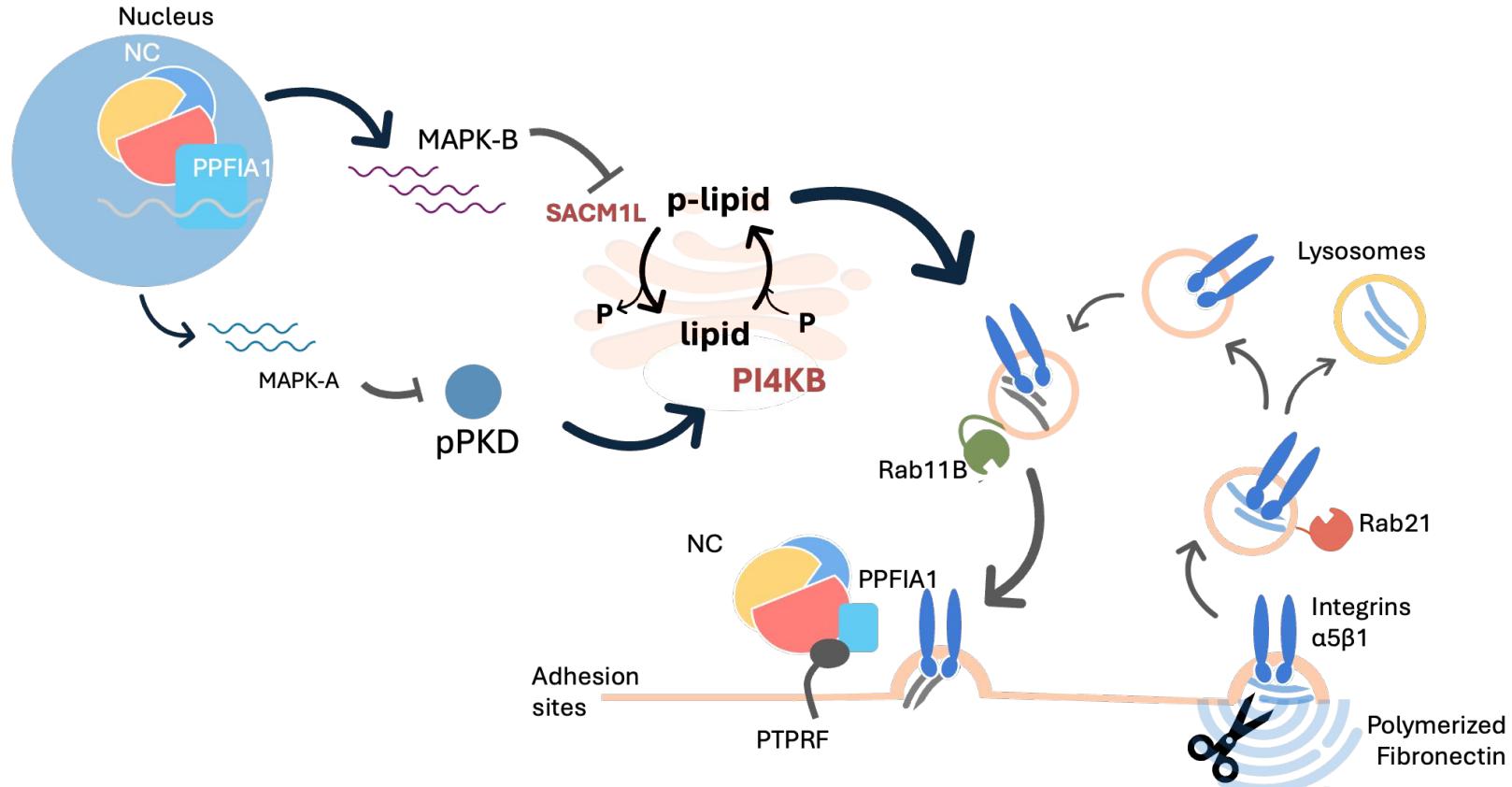
Integrin distributed on the membrane in CONTROL cells

NON - polarized cell



Integrin distributed on the membrane
in PPFA1 KNOCK-OUT cells

PPFIA1-dependent pathway overview



Creating the computational model

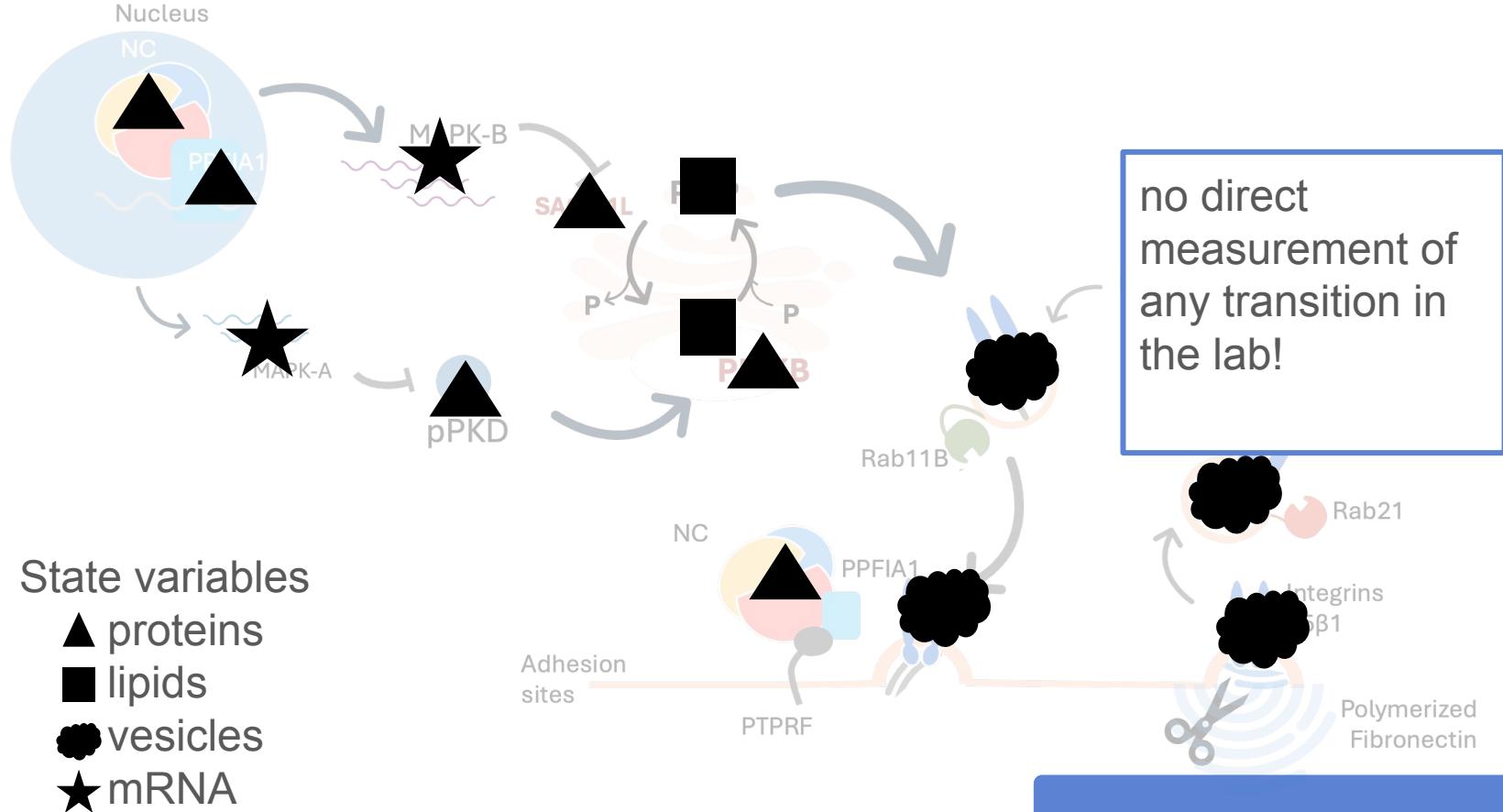
Model construction:

- translating biological graphical models into formal models
- multi-scale complexity: different levels of complexity and how they might interact across scales and levels of organization

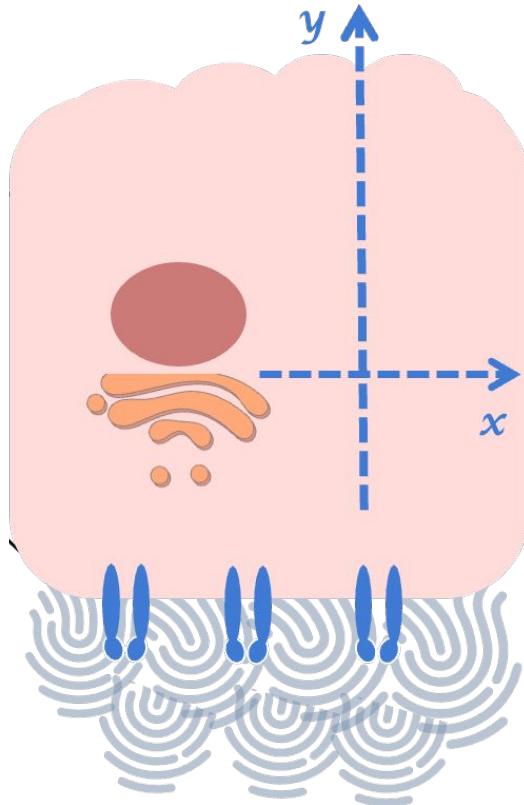
Data integration: models are dependent on quantitative data (e.g. lab experiments, omics, imaging) data is:

- required for parameter estimation, validation and prediction
- sparse and noisy, with limited time points
- can come from heterogeneous sources (in-house and from the literature) → can you compare it?
- can come from different types of experiments of different types of molecules → can you merge different types measurement units?
- is often a relative abundance data, but models require absolute values
- some data is unavailable and must be estimated (kinetic rates, binding constants)

PPFIA1-dependent pathway overview



What are we modeling and why?



We are modeling a single endothelial cell as PPFIA1 is progressively removed by siRNA.

We want to:

- explore the dynamics of the system
- identify gaps in our knowledge
- test hypotheses *in silico* first and validate *in vitro*



GreatORCA combines:

- data analysis of most common lab experiments
- data integration into computational models

ORCA purposes



Open Science

Sharing of research methods, outputs and data freely to enhance collaboration and knowledge dissemination



Reproducibility

Scientific results can be consistently replicated by independent researchers using the same methodology and data



User-friendly

Approachable by scientists without advanced R language knowledge



Transparency

Clear and open communication of all aspects of the research process, including methodologies, data, and analyses



Co-Distribution Report



Standardization

Establishing uniform protocols and guidelines to ensure consistency, accuracy, and comparability of results across studies

How to install ORCA

<https://github.com/qBioTurin/ORCA>

How to Install

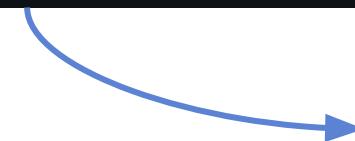
To install ORCA, you can use devtools:

```
install.packages("devtools")
devtools::install_github("qBioTurin/ORCA", ref="main", dependencies=TRUE)
```

How to Run

To run the Shiny application:

```
ORCA::ORCA.run()
```

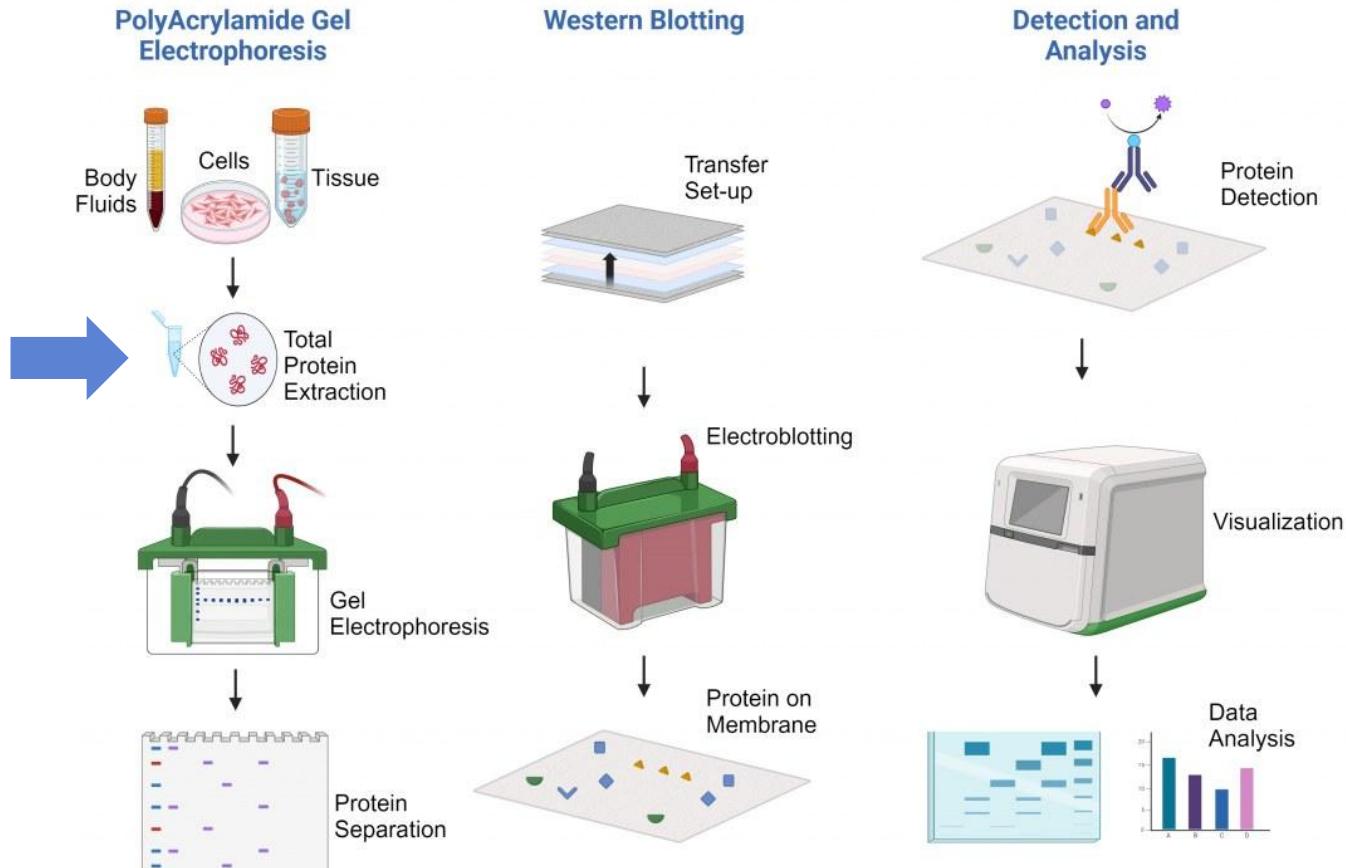


ORCA functionalities

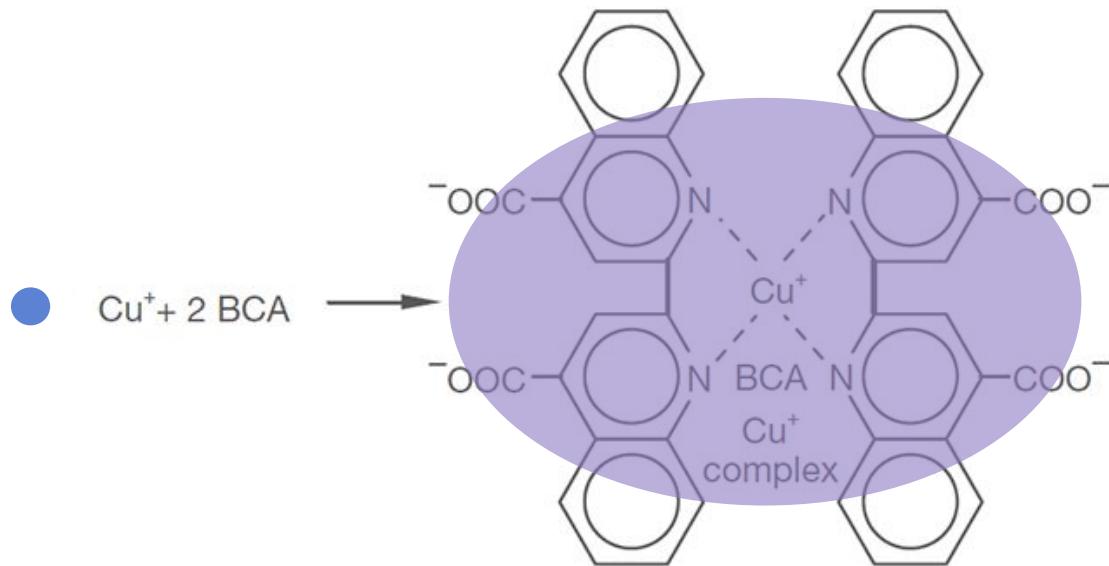
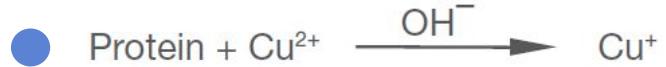


- ❖ Analysis of wet lab experiments
 - ➡ • Western Blot (**WB**),
 - ➡ • Reverse Transcription-quantitative PCR (**RT-qPCR**),
 - Enzyme-Linked ImmunoSorbent Assay (**ELISA**),
 - Endocytosis
 - Cytotoxicity
 - Immunofluorescence
 - Flow Cytometry analysis experiments (**FACS**)
 - ... and growing
- ➡ ❖ Support for statistical analysis
- ➡ ❖ Integration of wet lab experiment results in a computational model

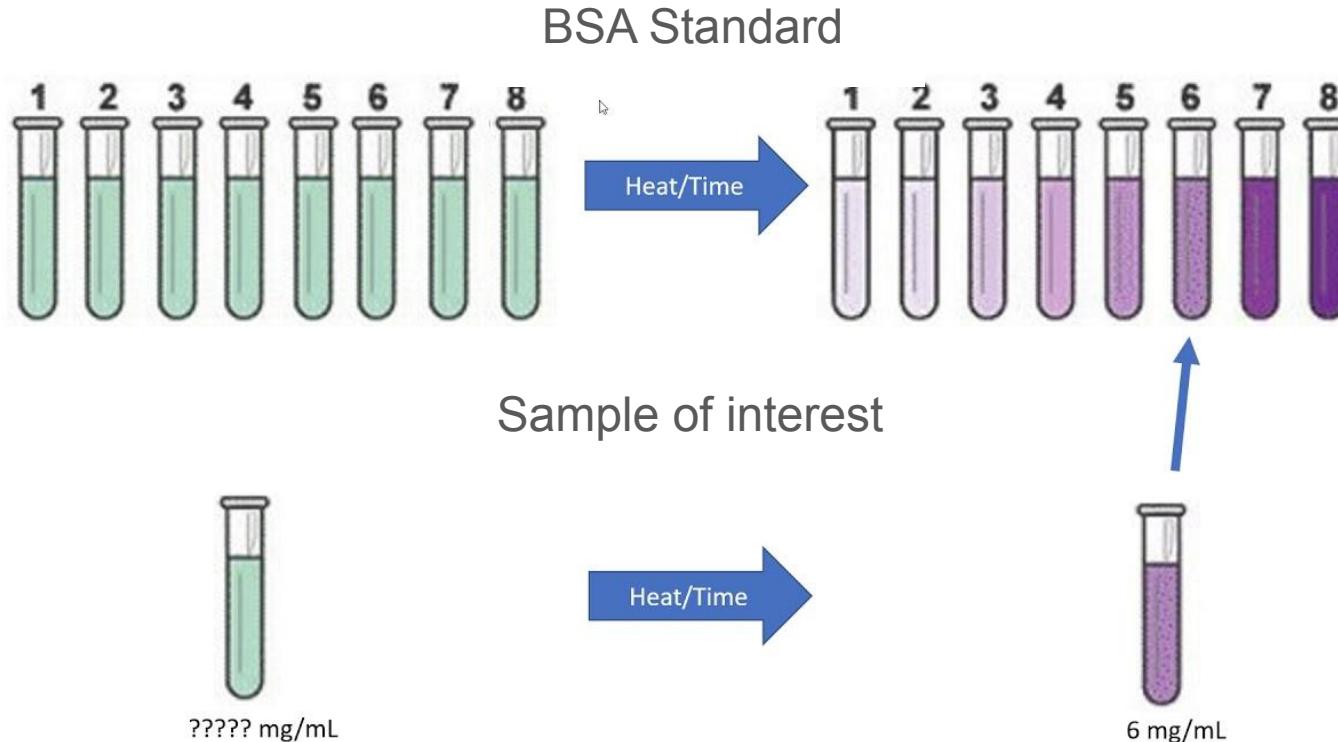
Western blot analysis: detect and quantify *proteins* in a sample



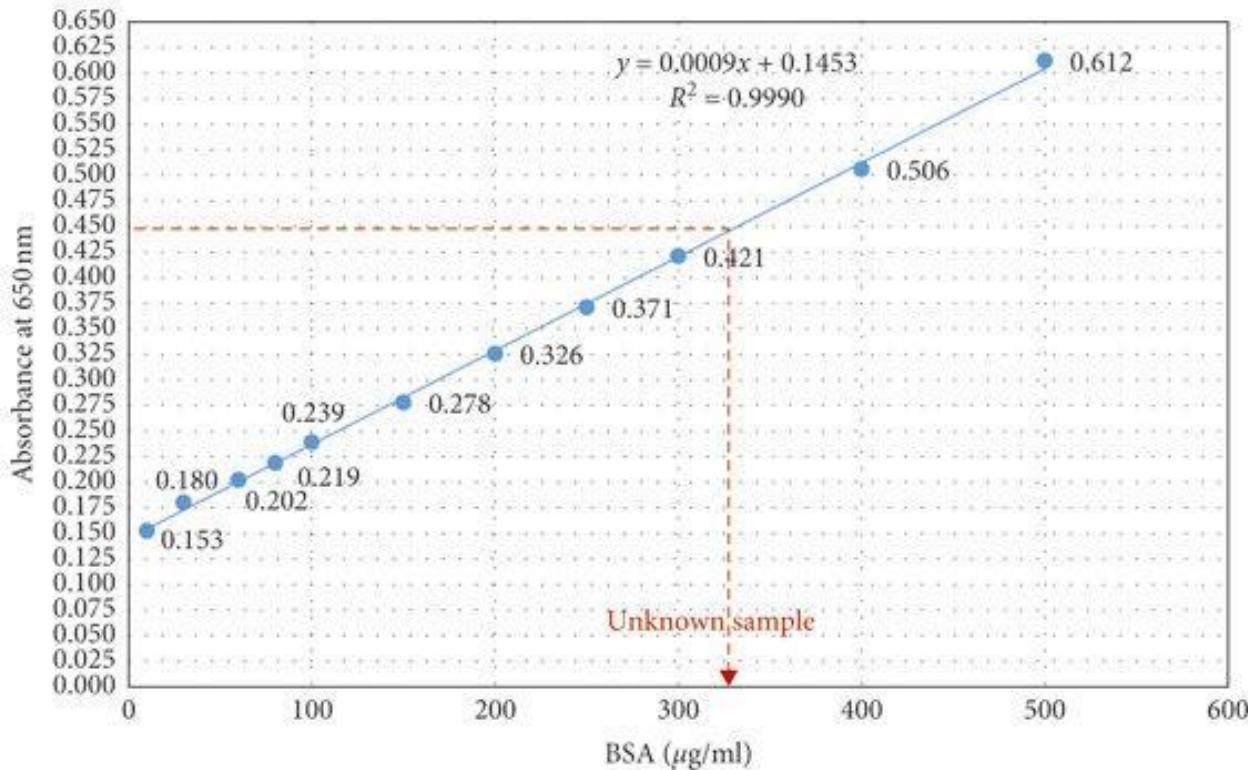
Bicinchoninic acid technique for protein quantification



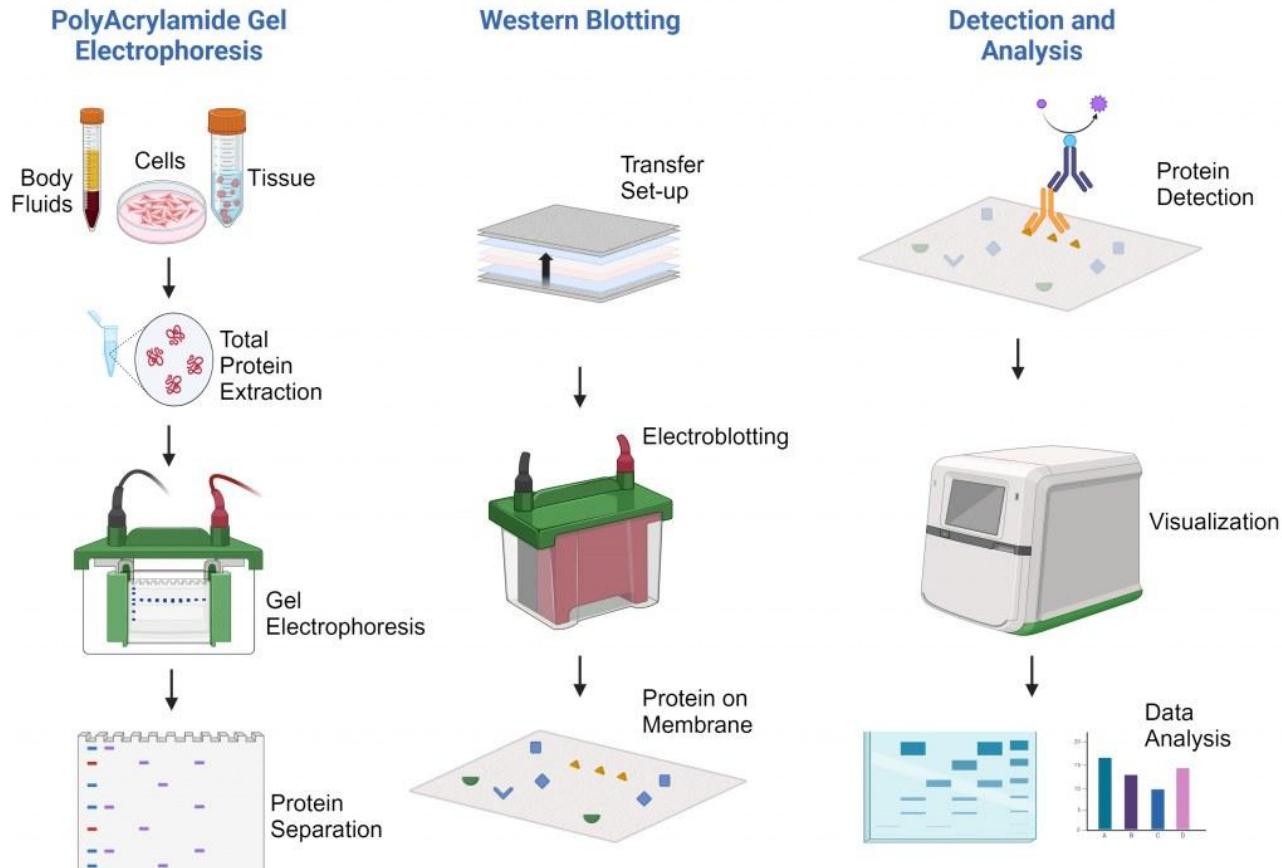
Bicinchoninic acid technique for protein quantification



Bicinchoninic acid technique for protein quantification

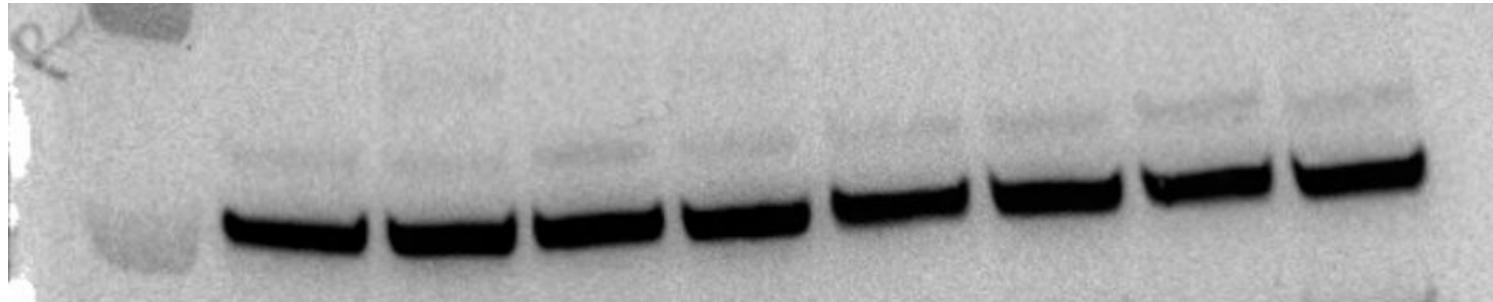


Western blot analysis: detect and quantify *proteins* in a sample



Western blot analysis: detect and quantify *proteins* in a sample

WB membrane imaging:

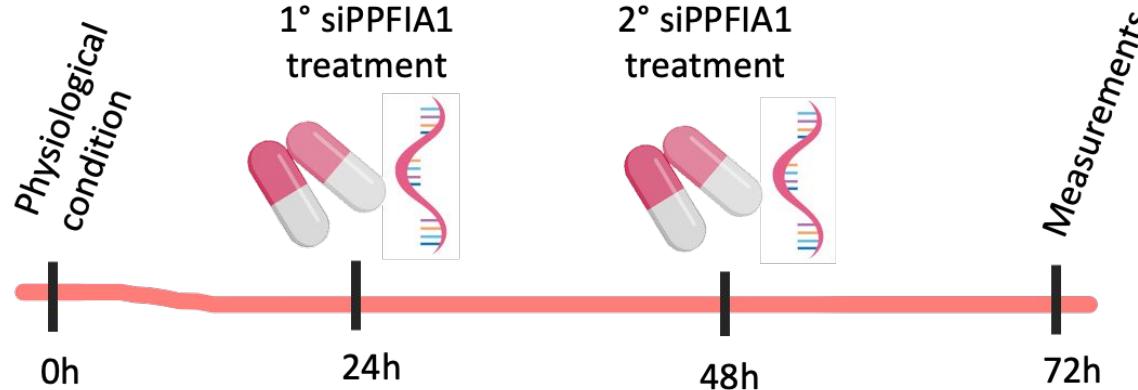


Densitometric analysis:

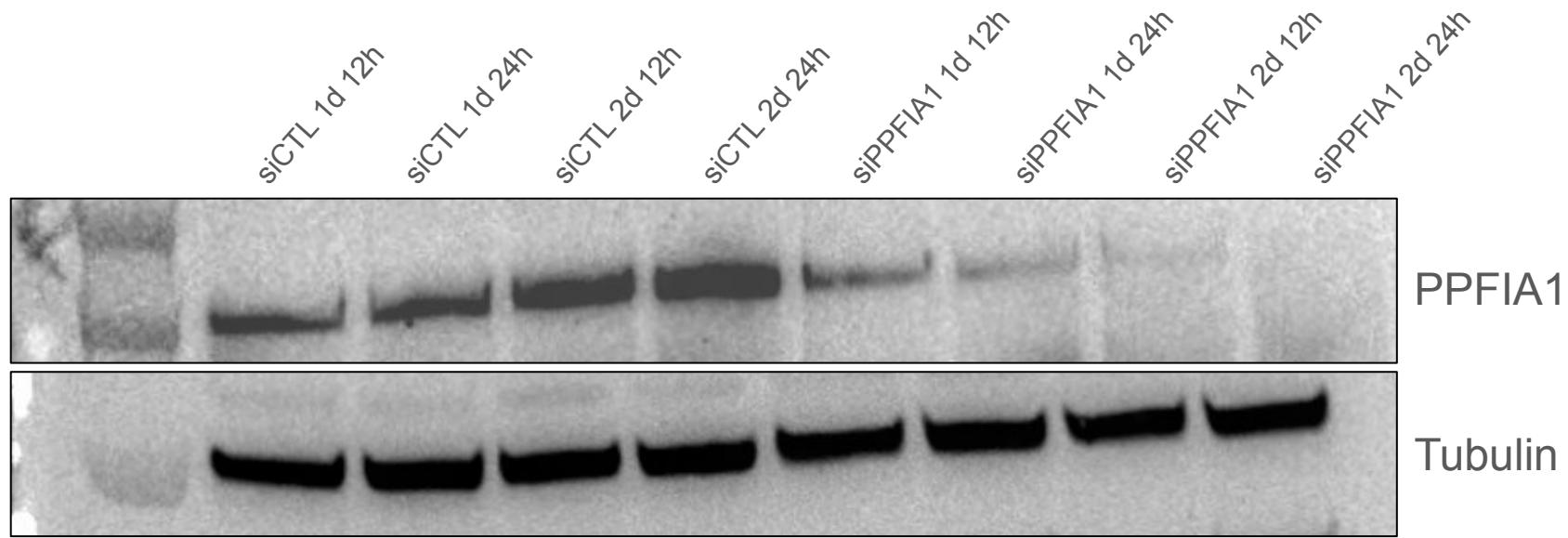
quantifies the intensity of protein bands on the blot,
providing a relative measure of protein abundance across different samples

What are we modeling and why?

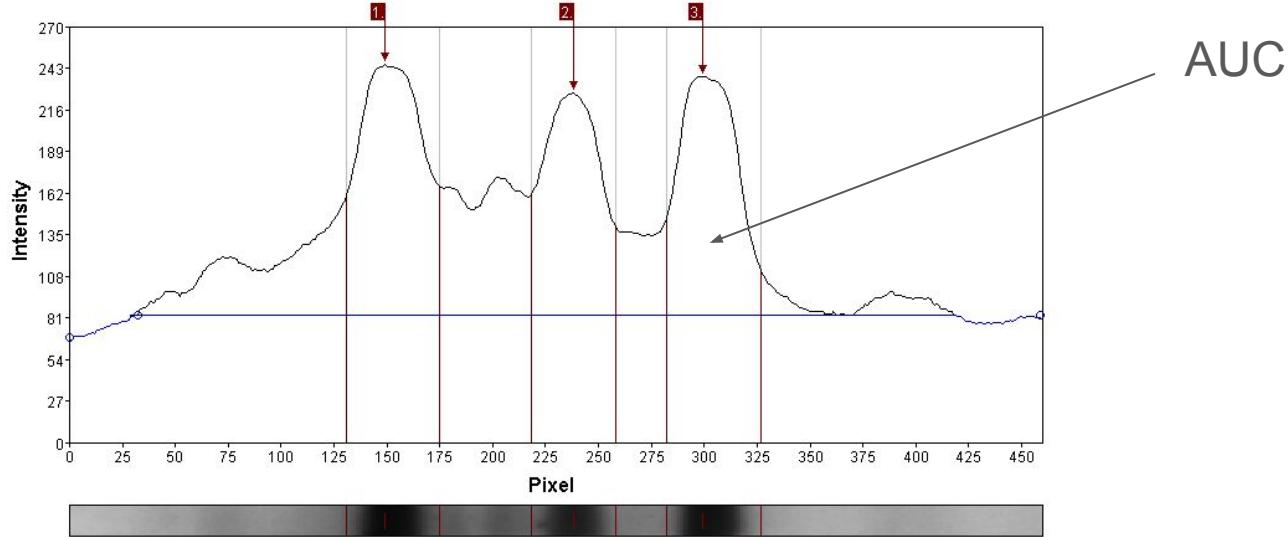
We are modeling a single endothelial cell as
PPFIA1 is progressively removed by siRNA.



Western blot analysis: quantifying PPFIA1

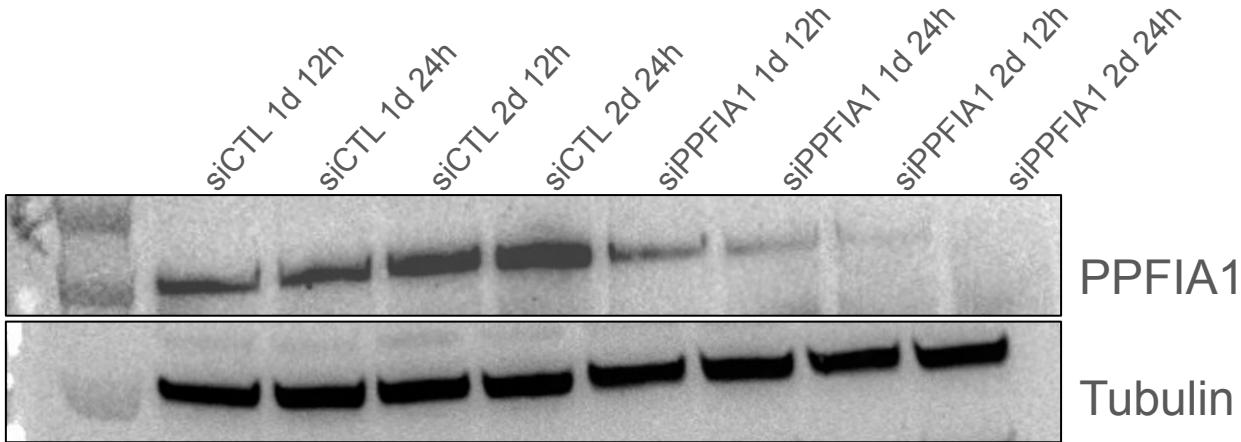


Western blot analysis: quantifying PPFA1



Each pixel in these digital images is assigned an intensity value that is related to the number of photons detected by the corresponding pixel in the sensor until it reaches saturation

Western blot analysis: quantifying PPFIA1



1. Relative density: in each blot, ORCA normalize the samples from the protein of interest WB on its control sample
1. Adjusted relative density: the AUC from the protein of interest is normalized on the AUC of the housekeeping/normalizer

WB analysis in ORCA



WB module in ORCA:

- quantification of sample protein content using BCA
- densitometric analysis of WB membrane image

Let's use ORCA to do a WB analysis

1. *Install and run ORCA:*

<https://github.com/qBioTurin/ORCA>

1. *Download data:*

https://drive.google.com/drive/folders/1mGFzbS1nflQ80TwQxC0FIkY0Mz6y_j7i?usp=drive_link

Let's use ORCA to do statistical analysis on WB experiments

GreatORCA: a strategy for data harmonization in comp. models

Mass spectrometry (iBAQ)

Madugudungu et al., *Proteomics*, 2019

$$iBAQ = \frac{\sum \text{peak intensity}}{\# \text{ theoretical peptides}}$$



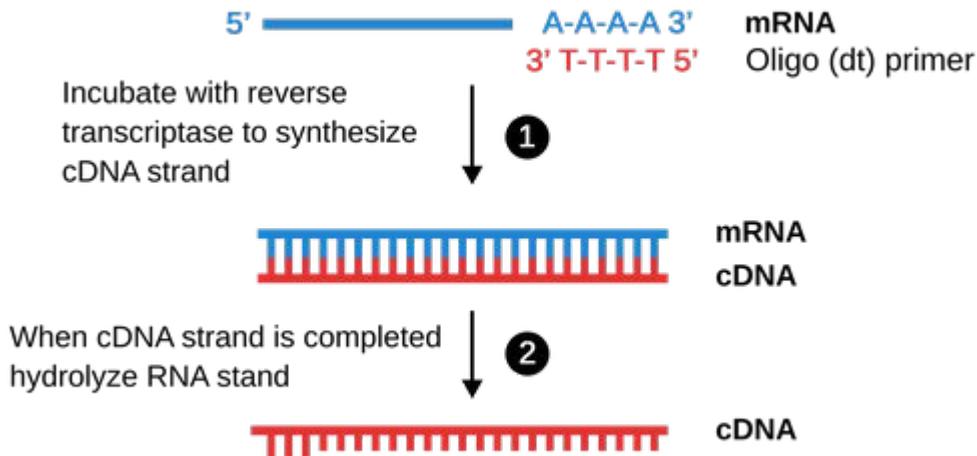
from relative abundance to
absolute abundance

Download data: https://drive.google.com/drive/folders/1mGFzbS1nflQ80TwQxC0FIkY0Mz6y_j7i?usp=drive_link

RT-qPCR analysis: detect and quantify target mRNA in a sample

Two steps:

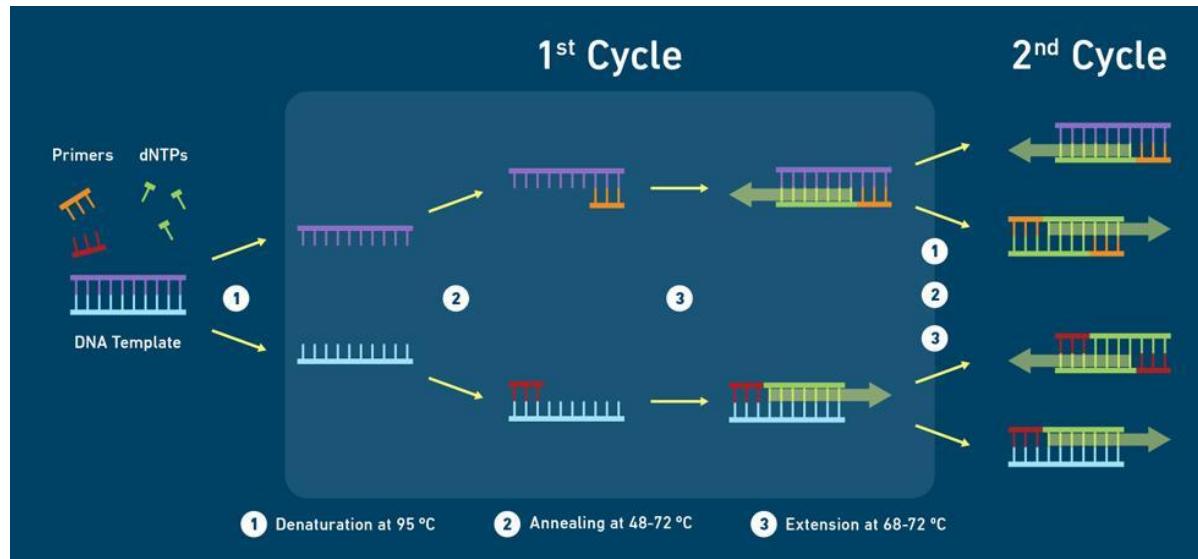
1. Reverse transcription: total RNA or mRNA is transcribed into complementary DNA (cDNA)



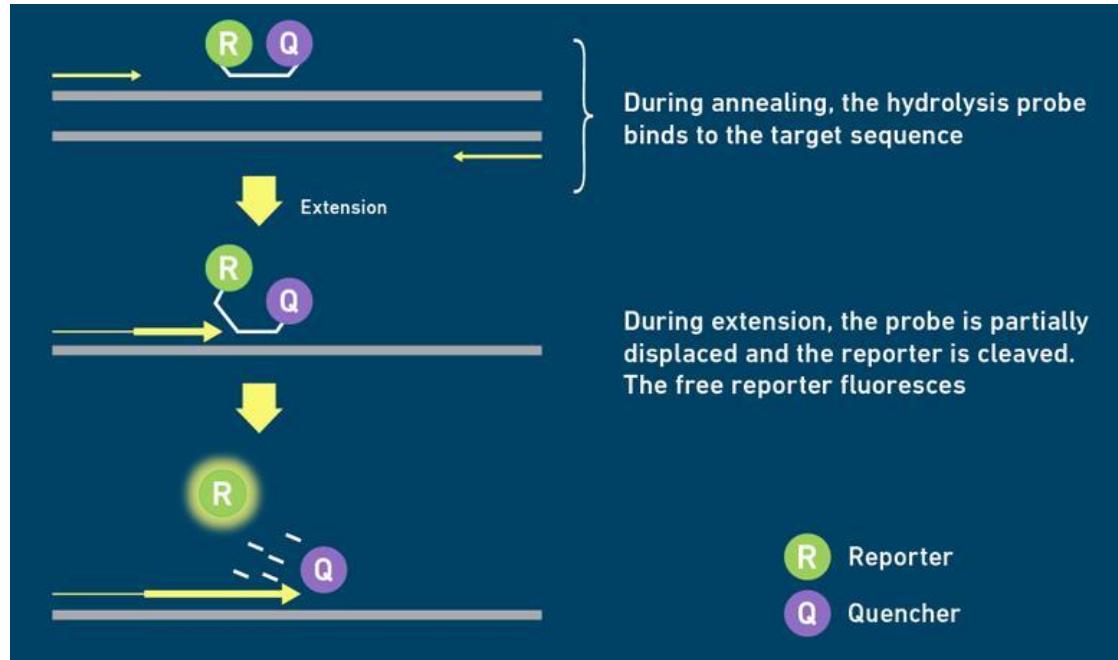
RT-qPCR analysis: detect and quantify target mRNA in a sample

Two steps:

2. qPCR: the cDNA is used as the template for the amplification of target mRNA, allowing mRNA quantification



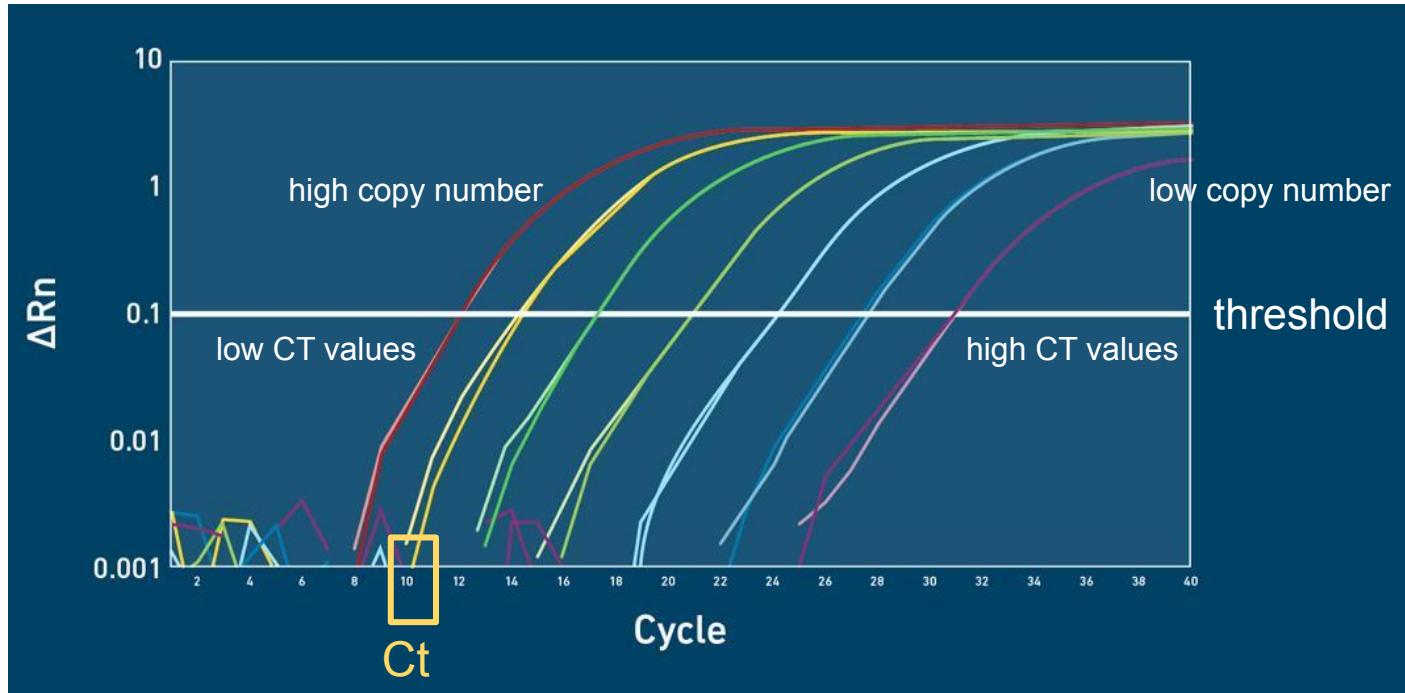
RT-qPCR analysis: detect and quantify target mRNA in a sample



Quantification:

fluorescent signal is detected by the fluorometer and it is proportional to the number of ssDNA fragments being amplified that are complementary to the probe.

RT-qPCR analysis: detect and quantify target mRNA in a sample



Ct or Cq (threshold cycle): the cycle number at which the fluorescence generated by a well crosses the threshold when sufficient amplicons have accumulated

RT-qPCR analysis: the output

| Sample Name | Detector Name | Reporter | Task | Ct |
|-------------|---------------|----------|---------|---------------|
| siCTRL | GAPDH | FAM | Unknown | 20.092 |
| siCTRL | GAPDH | FAM | Unknown | 20.080 |
| siCTRL | GAPDH | FAM | Unknown | 20.198 |
| siPPFIA1 | GAPDH | FAM | Unknown | 19.998 |
| siPPFIA1 | GAPDH | FAM | Unknown | 20.128 |
| siPPFIA1 | GAPDH | FAM | Unknown | 19.975 |
| siCTRL | MAPK-A | FAM | Unknown | 27.853 |
| siCTRL | MAPK-A | FAM | Unknown | 27.937 |
| siCTRL | MAPK-A | FAM | Unknown | 27.966 |
| siPPFIA1 | MAPK-A | FAM | Unknown | 27.601 |
| siPPFIA1 | MAPK-A | FAM | Unknown | 27.608 |
| siPPFIA1 | MAPK-A | FAM | Unknown | 27.689 |
| siCTRL | MAPK-B | FAM | Unknown | 25.358 |
| siCTRL | MAPK-B | FAM | Unknown | 25.327 |
| siCTRL | MAPK-B | FAM | Unknown | 25.463 |
| siPPFIA1 | MAPK-B | FAM | Unknown | 26.143 |
| siPPFIA1 | MAPK-B | FAM | Unknown | 26.171 |
| siPPFIA1 | MAPK-B | FAM | Unknown | 26.211 |
| siCTRL | TBP | FAM | Unknown | 25.832 |
| siCTRL | TBP | FAM | Unknown | 25.799 |
| siCTRL | TBP | FAM | Unknown | 25.946 |
| siPPFIA1 | TBP | FAM | Unknown | 26.235 |
| siPPFIA1 | TBP | FAM | Unknown | 26.381 |
| siPPFIA1 | TBP | FAM | Unknown | 26.464 |

Housekeeping genes:
GAPDH and *TBP*

Target genes:
MAPK-A and *MAPK-B*

RT-qPCR analysis: the output

| Sample Name | Detector Name | Reporter | Task | Ct |
|-------------|---------------|----------|---------|--------|
| siCTRL | GAPDH | FAM | Unknown | 20.092 |
| siCTRL | GAPDH | FAM | Unknown | 20.080 |
| siCTRL | GAPDH | FAM | Unknown | 20.198 |
| siPPFIA1 | GAPDH | FAM | Unknown | 19.998 |
| siPPFIA1 | GAPDH | FAM | Unknown | 20.128 |
| siPPFIA1 | GAPDH | FAM | Unknown | 19.975 |
| siCTRL | MAPK-A | FAM | Unknown | 27.853 |
| siCTRL | MAPK-A | FAM | Unknown | 27.937 |
| siCTRL | MAPK-A | FAM | Unknown | 27.966 |
| siPPFIA1 | MAPK-A | FAM | Unknown | 27.601 |
| siPPFIA1 | MAPK-A | FAM | Unknown | 27.608 |
| siPPFIA1 | MAPK-A | FAM | Unknown | 27.689 |
| siCTRL | MAPK-B | FAM | Unknown | 25.358 |
| siCTRL | MAPK-B | FAM | Unknown | 25.327 |
| siCTRL | MAPK-B | FAM | Unknown | 25.463 |
| siPPFIA1 | MAPK-B | FAM | Unknown | 26.143 |
| siPPFIA1 | MAPK-B | FAM | Unknown | 26.171 |
| siPPFIA1 | MAPK-B | FAM | Unknown | 26.211 |
| siCTRL | TBP | FAM | Unknown | 25.832 |
| siCTRL | TBP | FAM | Unknown | 25.799 |
| siCTRL | TBP | FAM | Unknown | 25.946 |
| siPPFIA1 | TBP | FAM | Unknown | 26.235 |
| siPPFIA1 | TBP | FAM | Unknown | 26.381 |
| siPPFIA1 | TBP | FAM | Unknown | 26.464 |

$$\Delta CT_{\text{gene}} = \text{avg}(CT_{\text{gene}}) - \text{avg}(CT_{\text{housekeeping}})$$

$$\Delta CT_{\text{baseline}} = \text{avg}(CT_{\text{baseline}}) - \text{avg}(CT_{\text{housekeeping}})$$

$$\Delta\Delta CT = \Delta CT_{\text{gene}} - \Delta CT_{\text{baseline}}$$

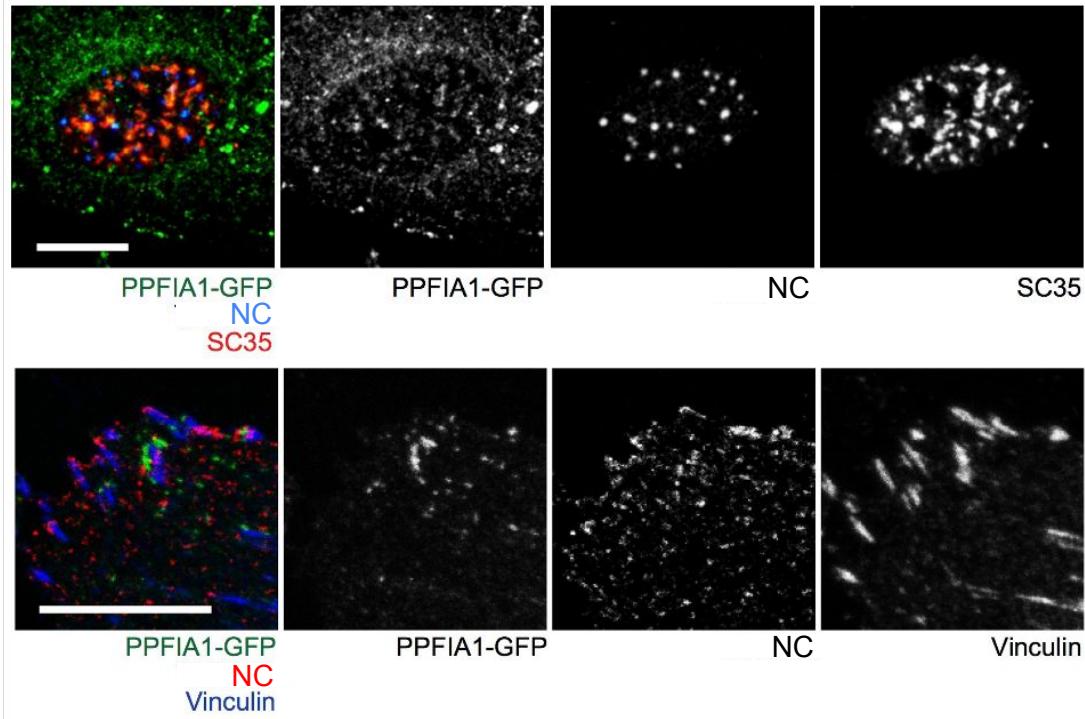
$$Q = 2^{-\Delta\Delta CT}$$

Let's use ORCA to do a RT-qPCR analysis

Download data: https://drive.google.com/drive/folders/1mGFzbS1nflQ80TwQxC0FIkY0Mz6y_i7i?usp=drive_link

Given this data, how would you design your petri net?

1. PPFIA1 interacts with NC in the nucleus

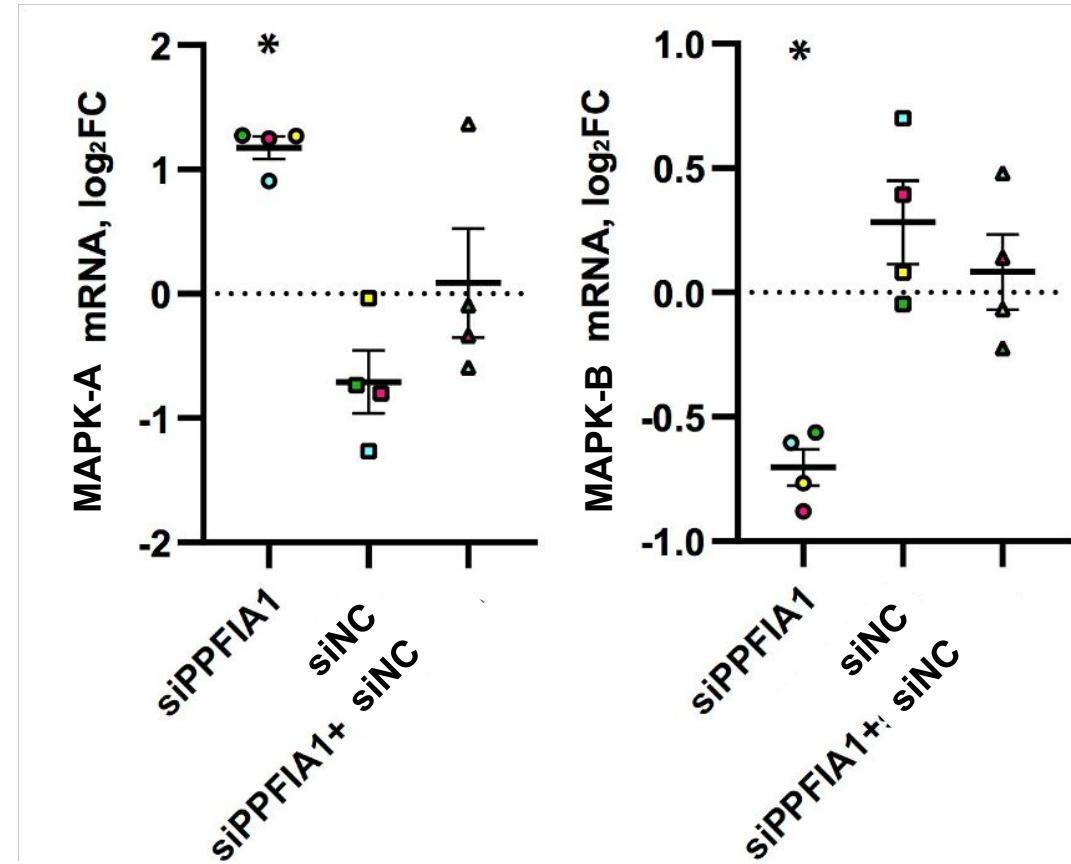


2. PPFIA1 interacts with NC at adhesion sites

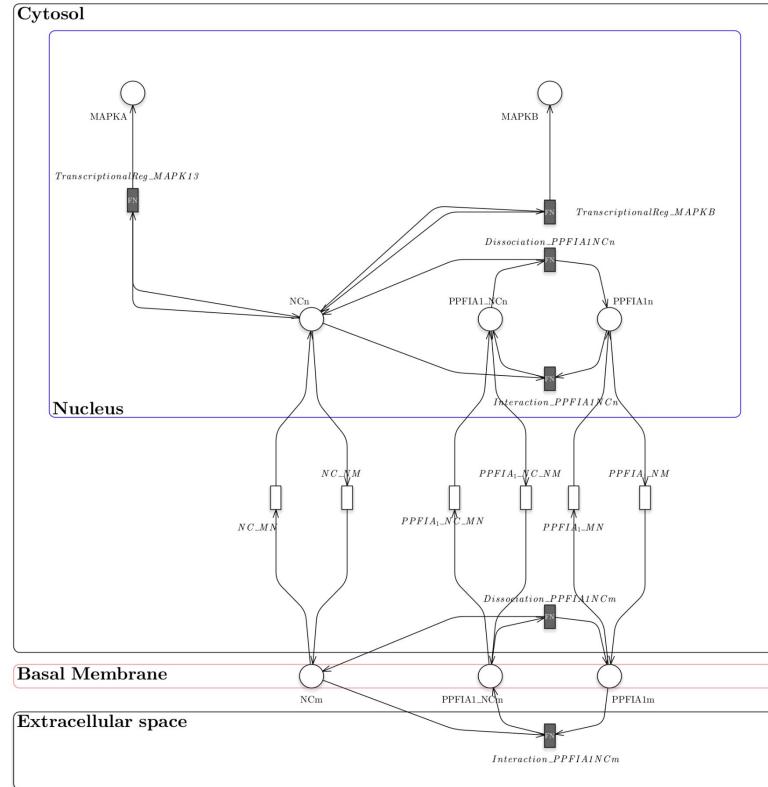
Given this data, how would you design your petri net?

3. PPFIA1 regulates MAPKA and MAPKB transcription *via* NC

Experimental replicates
#1 #2 #3 #4



Given this data, how would you design your petri net?



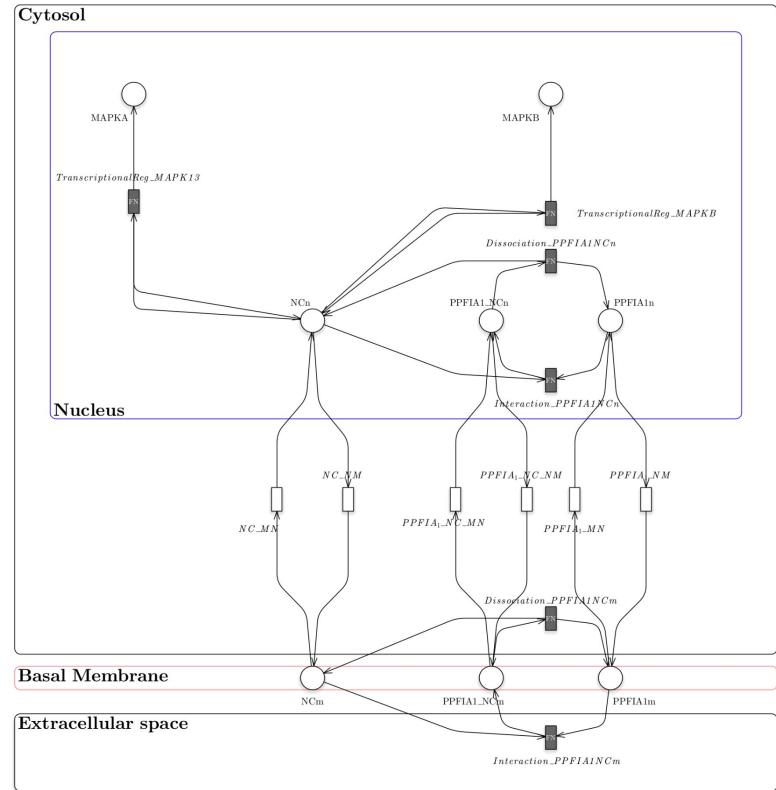
Let's parametrize our model

Model marking:

1. Add tokens to PPFIA1
2. Add tokens to NC
3. Add tokens to MAPKA and MAPKB

Transition rates:

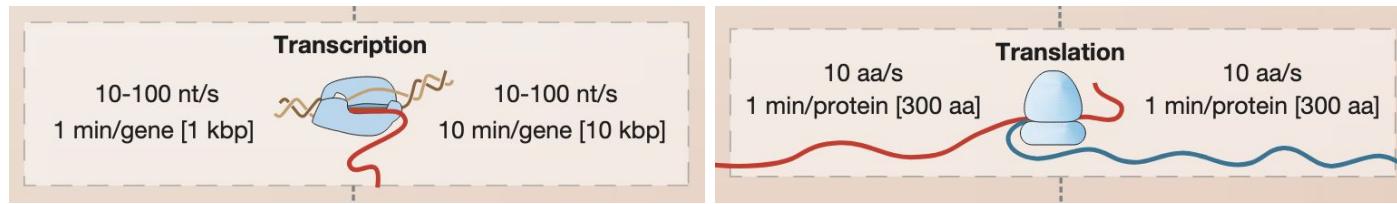
1. Removing PPFIA1 at set times
2. Interaction and movement rates: to be calibrated
3. Transcription rates



Parametrizing transcription reaction

MAPK-A and MAPK-B transcriptional regulation reactions have to take into account two aspects:

1. temporal frame in which the biological processes of transcription and translation occur → “rate_expression”



[https://www.cell.com/cell/pdf/S0092-8674\(16\)30208-2.pdf?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0092867416302082%3Fshowall%3Dtrue](https://www.cell.com/cell/pdf/S0092-8674(16)30208-2.pdf?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0092867416302082%3Fshowall%3Dtrue)

MAPK-A: 6316 nt and 365 AA ~450 seconds → rate_expression_A = 1/450s

MAPK-B: 4222 nt and 360AA ~325 seconds → rate_expression_B = 1/325s

2. modulation of gene transcription by NC activity → “cost_modulation” → estimated

Parametrizing transcription reaction

MAPK-A and MAPK-B transcriptional regulation reactions have to take into account two aspects:

1. temporal frame in which the biological processes of transcription and translation occur → “rate_expression”
2. modulation of gene transcription by NC activity → “cost_modulation” → estimated

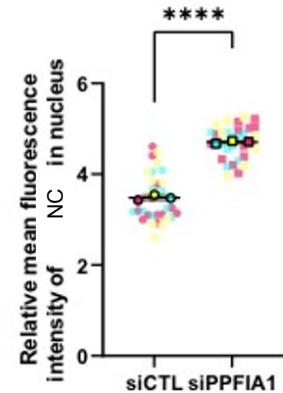
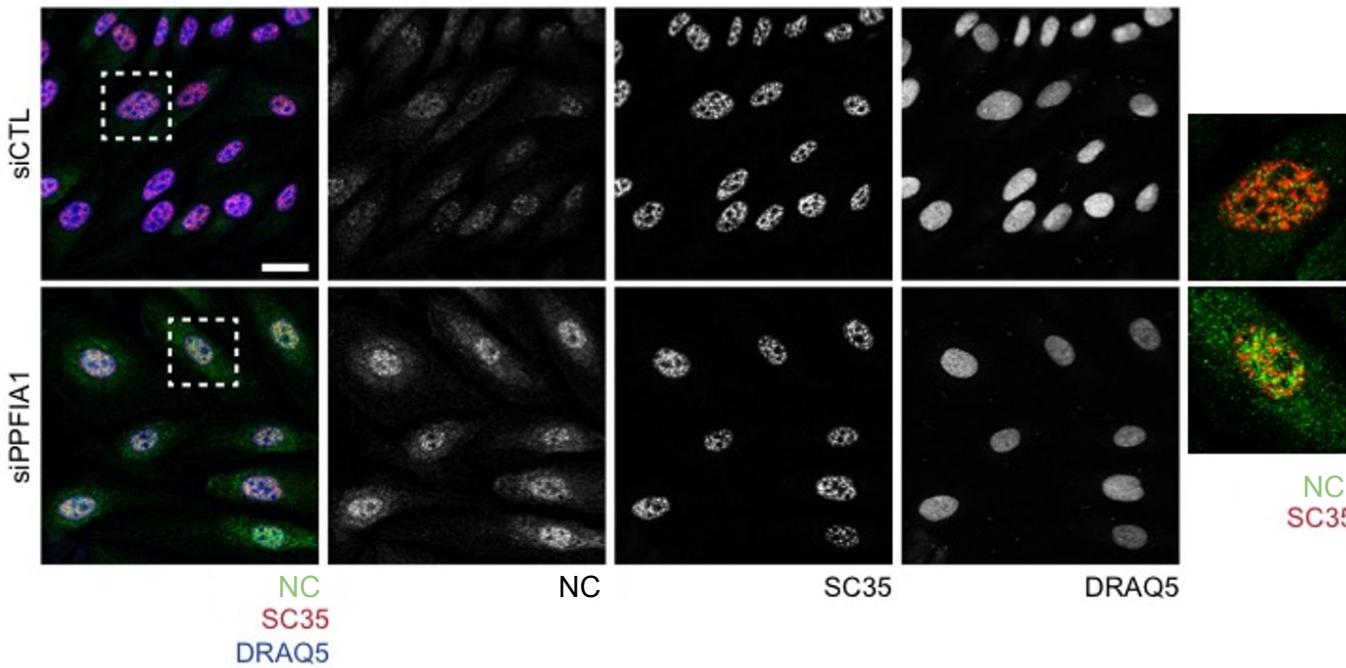
$$f(\text{TranscriptionalReg_MAPK_A}) = \text{rate_expression_A} + x_{\text{NC}} * \text{cost_modulation_13};$$

$$f(\text{TranscriptionalReg_MAPK14}) = \text{rate_expression_14} - x_{\text{NC}} * \text{cost_modulation_14},$$

$$\text{cost_modulation} \geq \text{rate_expression_B}/\max(x_{\text{NC}}).$$

Given this data, how would you design your petri net?

Hypothesis confirmation in the lab:
PPFIA1 physically and functionally inhibits ATAC complex



Given this data, how would you design your petri net?

Hypothesis confirmation in the lab:
PPFIA1 physically and functionally inhibits ATAC complex

