

NETWORK RECONSTRUCTION OF  
BIOCHEMICAL PATHWAYS  
USING DISCRETE DYNAMICAL SYSTEMS

Roxana Feier

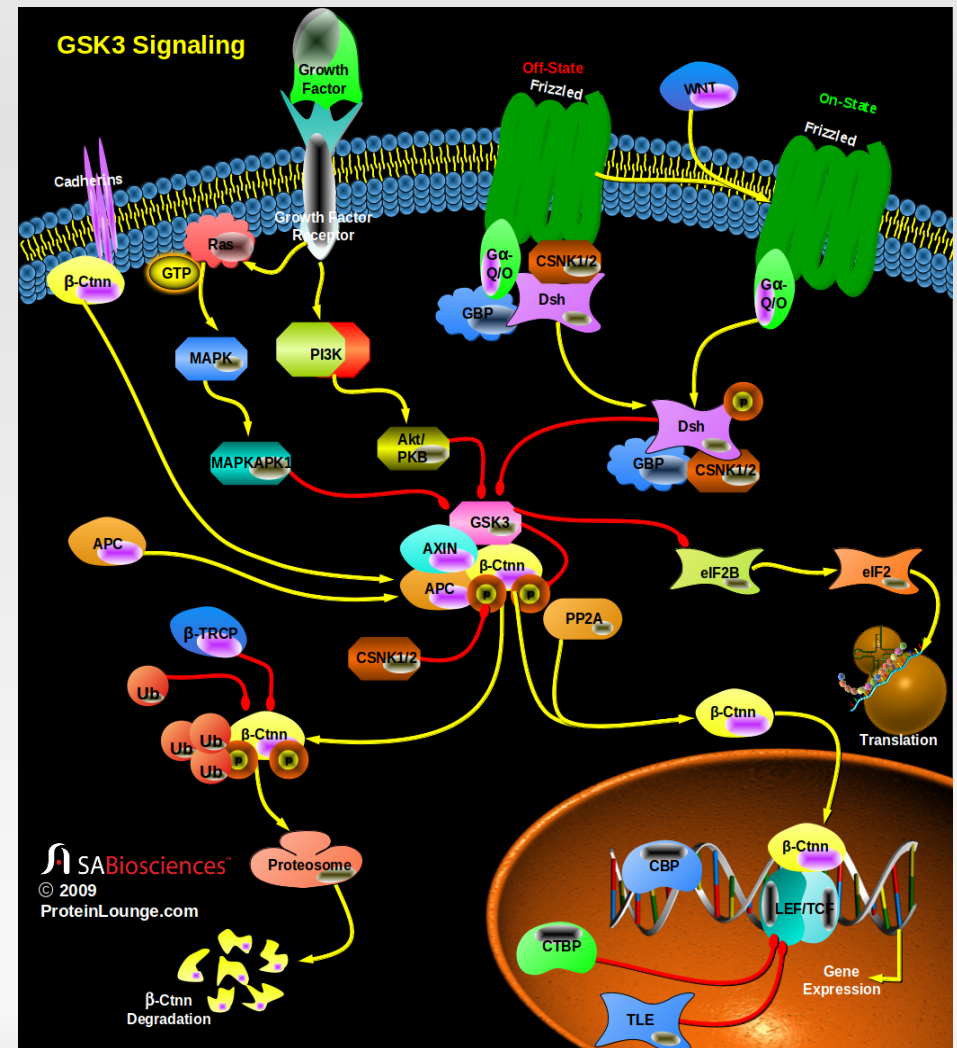
PRISE 2011

# Why mathematical models?

- **Golden age** for quantitative models:
  - DNA sequencing → large-scale measurements of chemical species (e.g. gene transcripts) within a pathway
- Such models are **predictive**:
  - Modify parameters to understand how *change* affects *function*
- **Feedback loop** between theory and experiment
  - Experimental data → model → *hypotheses* to be tested experimentally → better models → better experiments → ...

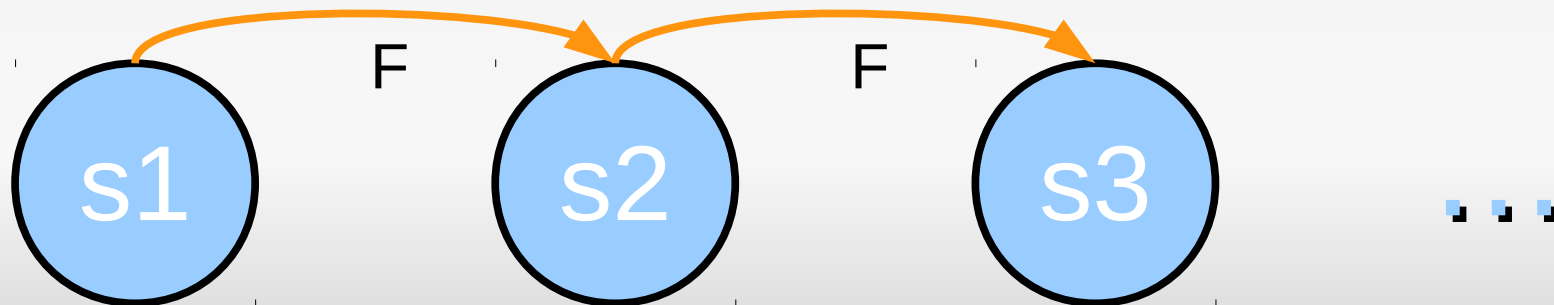
# The Wnt pathway

- Embryogenesis and adult tissue homeostasis
- Components mutated in ~90% of colorectal cancers
- Quantitative models → understand how **change** affects **function** → modulate these effects



# Interpolate data, recover dependencies

- N **species**  $\{x_1, \dots, x_N\}$
- Given: **time-series data**  $\{s_1, s_2, s_3, \dots\}$ 
  - $s_j \rightarrow$  vector of length  $N$  with concentrations at time  $j$
- Want:  $F = (f_1, \dots, f_N)$  with  $F(s_1) = s_2, \dots$ 
  - $f_j$  is the **transition function** for  $x_j$
- Species that appear in  $f_j \rightarrow$  connected to  $x_j$  in the reconstructed network



# Problem ...

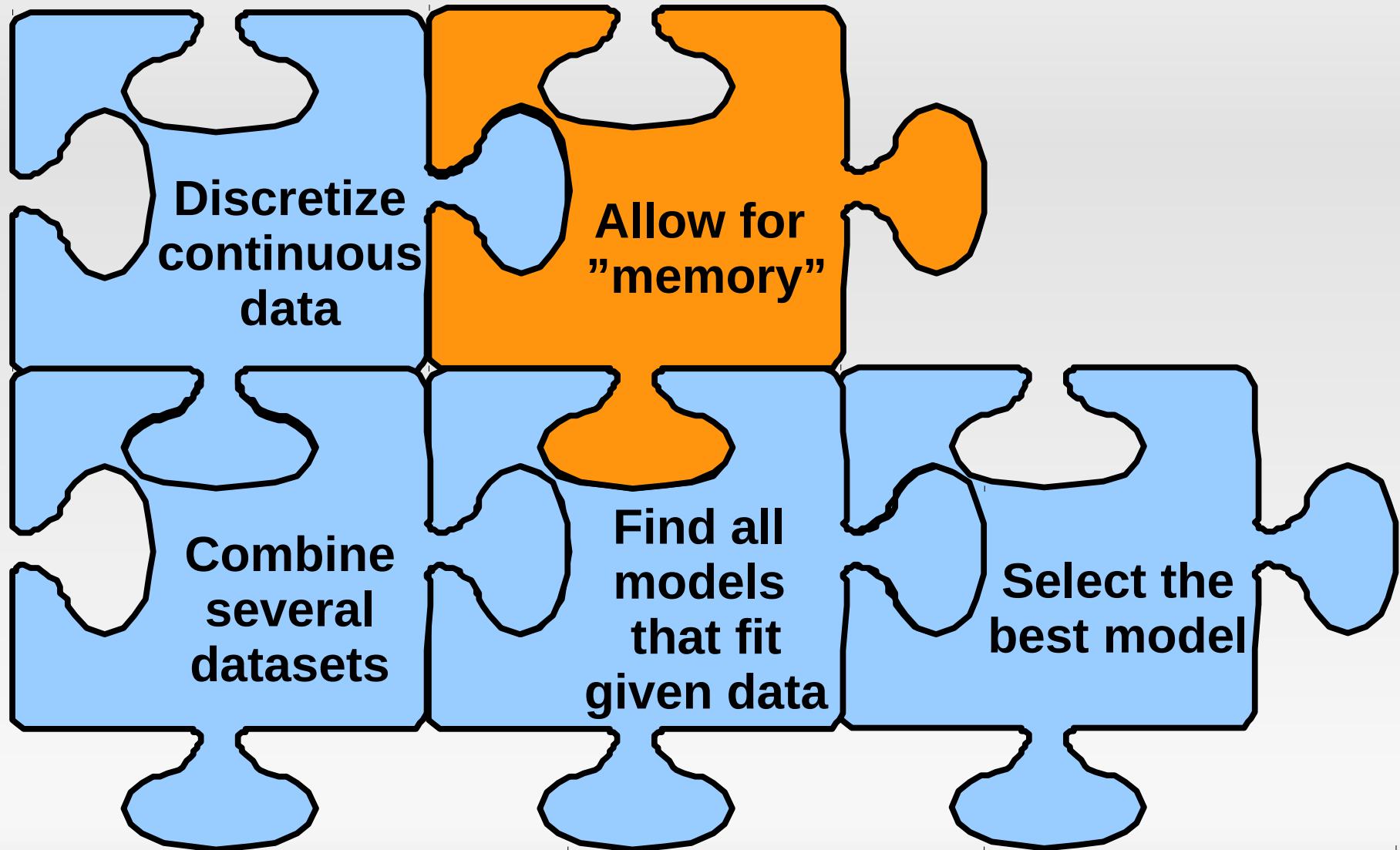
Many, many functions that fit  
a relatively small amount of data.

# Algebraic geometry to the rescue!

- Problem: many functions that interpolate a finite amount of data
- Solution: use **discrete** values for the concentrations
  - $k = \{0, 1, \dots, p - 1\}$
- Theorem: any function  $f : k^N \rightarrow k$  is a **polynomial**



# From the data to the network



# Step 1: Discretize the continuous data

- **Summary:** Pick a prime  $p$ , discrete values will be  $0, 1, \dots, p - 1$ .
- **Plus:** Keep essential information only: initial data is noisy.
- **Minus:** How to choose  $p$ ?

Time	x	y	z
1	0.9	0.92	0.89
2	0.47	0.27	0.26
3	0.24	0.39	0.42



$p = 3$

Time	x	y	z
1	2	2	2
2	1	0	0
3	0	1	1



# Step 2: Combine multiple datasets

- **Summary:** Use both **wildtype** and **knockout/knockdown** data.

2	2	2
1	0	0
Dataset #1		

F

- **Plus:** Better exploration of the space of all possible states.

- **Minus:** Can lead to **inconsistencies** upon discretization.

2	2	2
1	1	0
Dataset #2		

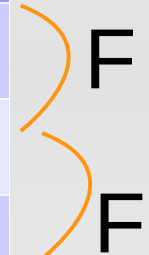
Same F??

*With memory → much fewer inconsistencies*

# Step 3: Find all models that fit the data

- **Summary:** Solutions of the form  $F + H$  where  $F$  fits the model and  $H$  vanishes on the given data (compare to differential equations).
- **Plus:** Finds all solutions!
- **Minus:** Finding all  $H$  that vanish on the data is computationally hard.

Time	x	y	z
1	2	2	2
2	1	0	0
3	0	1	1



For example:

$$F(x, y, z) = (y^2, y + 1, 2y - z + 1)$$

Check:

$$F(2, 2, 2) = (1, 0, 0)$$

$$F(1, 0, 0) = (0, 1, 1)$$

# Step 4: Select the best model

- **Summary:** Pick a model which is minimal in some sense.
- **Plus:** Can incorporate **prior knowledge** about the network.
- **Minus:** What does "minimal" mean?

Time	x	y	z
1	2	2	2
2	1	0	0
3	0	1	1

} F  
}

$$F(x, y, z) = (y^2, y + 1, 2y - z + 1)$$

But  $y - z = 0$  on the dataset.

Better:

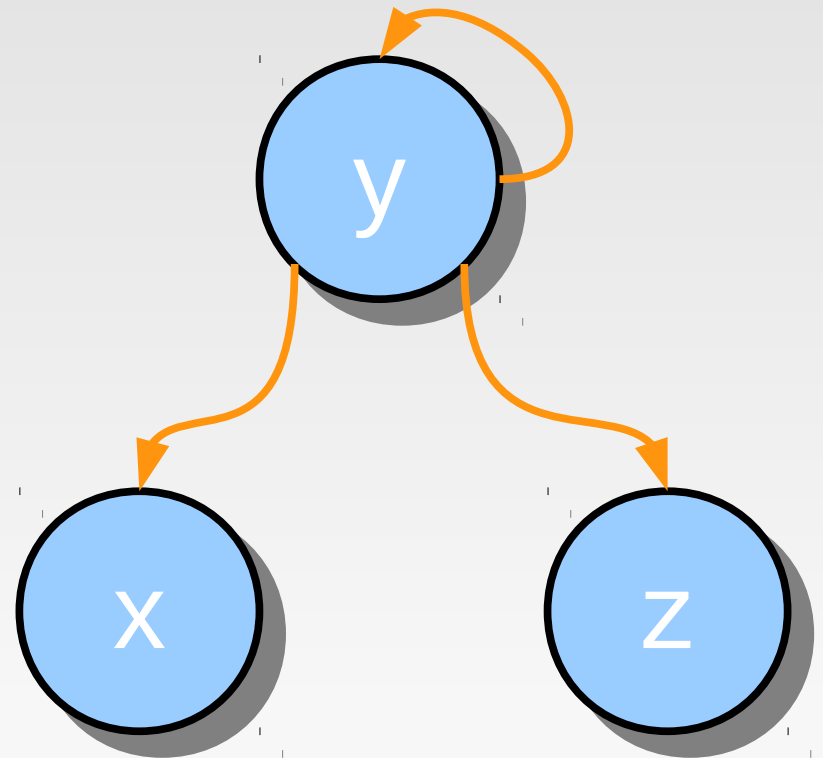
$$F(x, y, z) = (y^2, y + 1, y + 1)$$

# Final step: reconstruct the network

- The network is a **directed graph**
- The **nodes** are the **species**
- If the function for  $x$  coordinate depends on  $y$ , draw an **edge** from  $y$  to  $x$ .

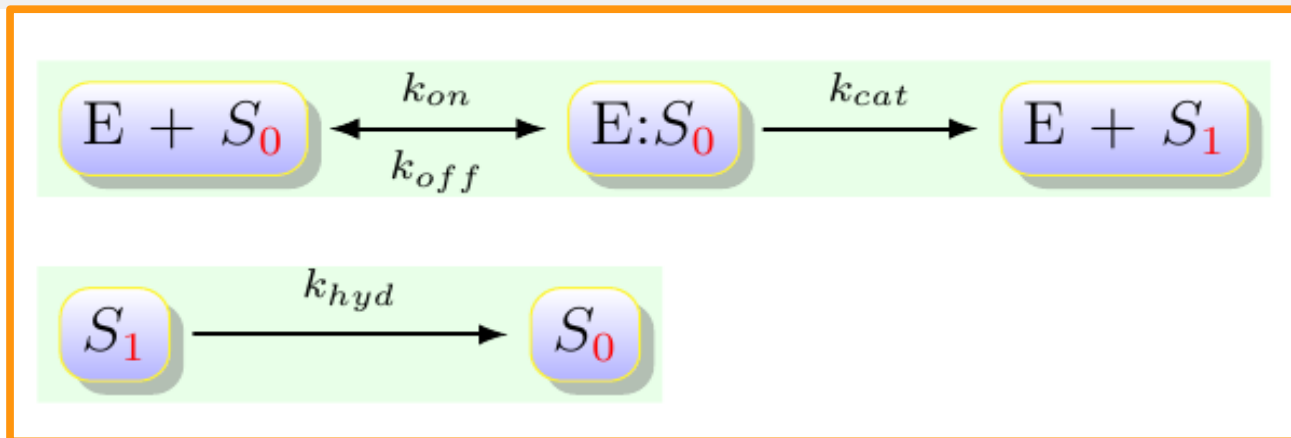
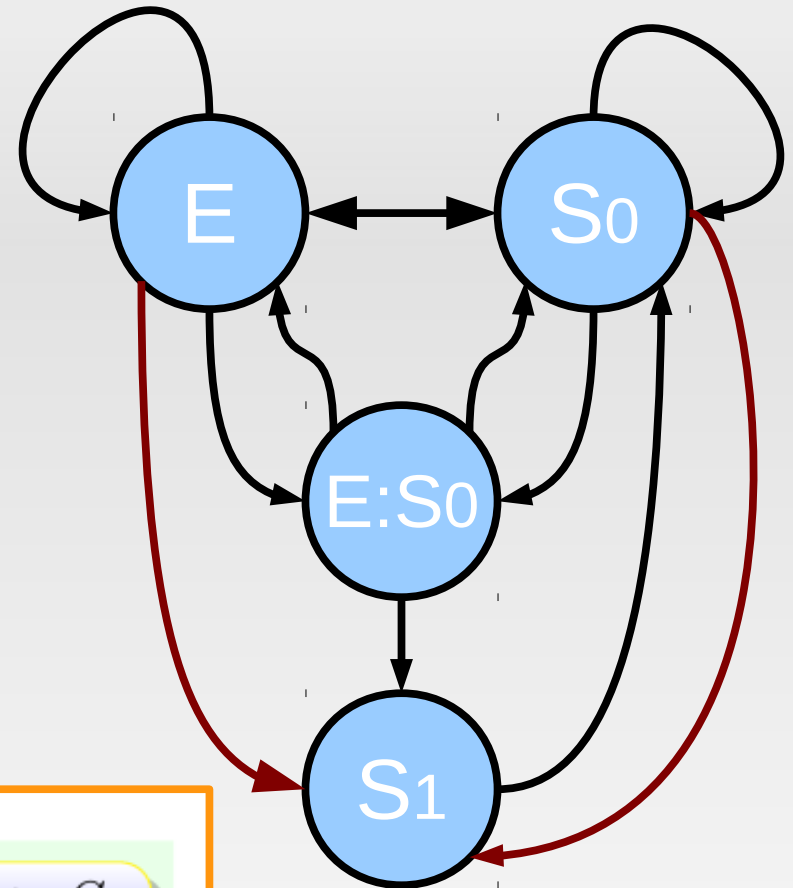
- Our toy example:

$$F(x, y, z) = (y^2, y + 1, y + 1)$$



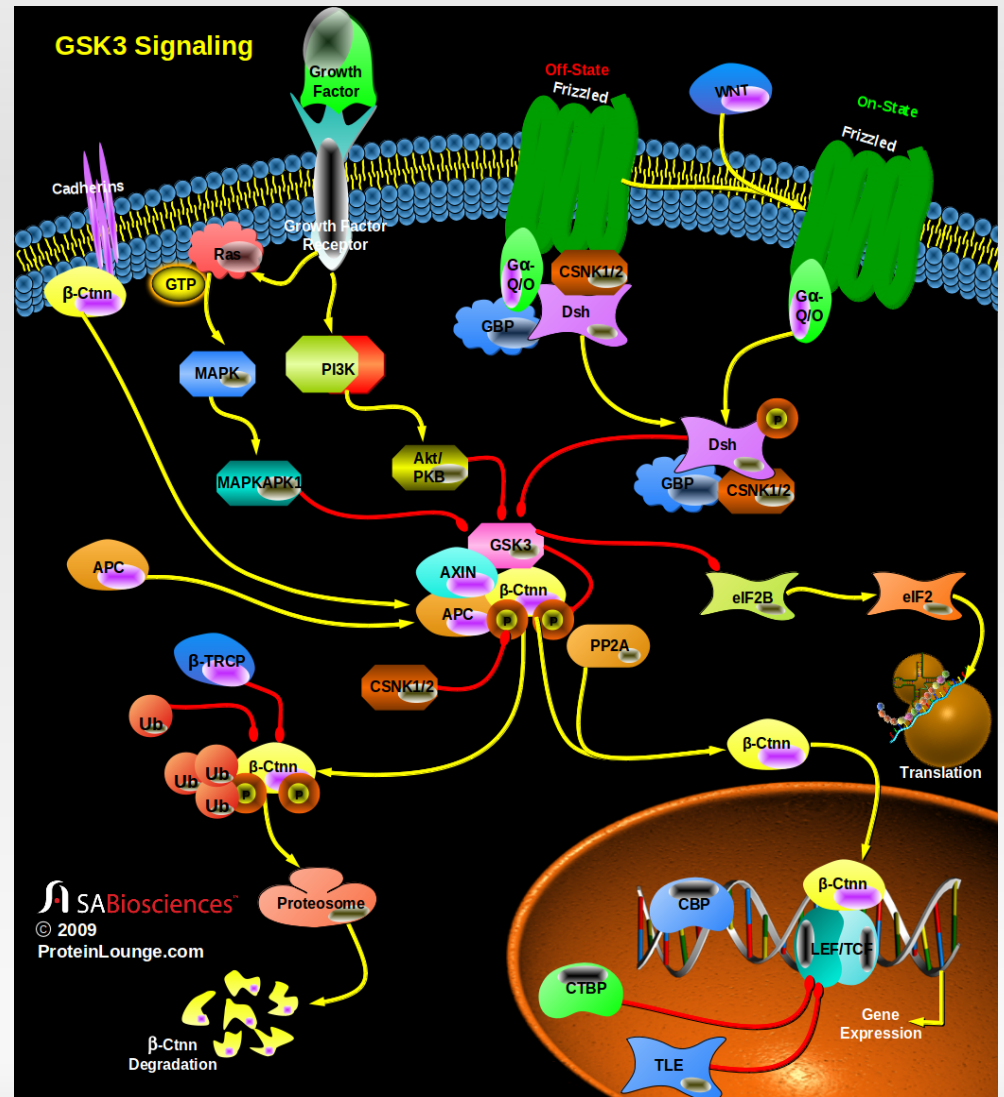
# Example 1: One-site phosphorylation

- No false negatives
- Two false positives
  - Due to indirect influence?
- Ways to detect this

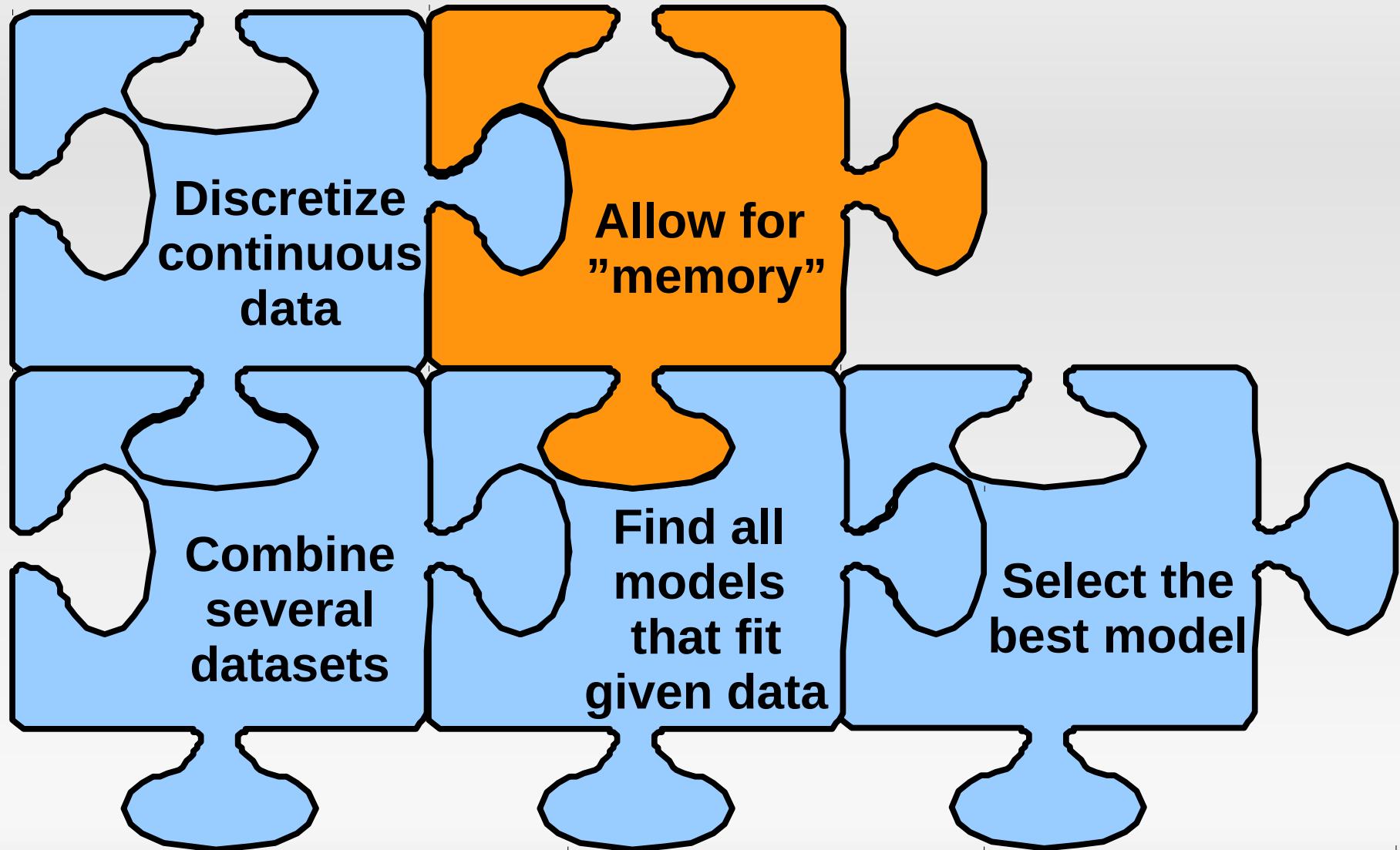


# Example 2: the Wnt pathway

- 17 species, data simulated from ODE model
- **Preliminary** results:
  - 160 out of  $17^2 = 279$  interactions detected
  - Many false positives
  - Few false negatives



# Summary



# What next?

- How to choose **p**?
- What makes good **data**?
- How does **noise** affect the reconstruction?
- How do **coefficients** in the transition functions correlate with **strengths** of interactions between species?



# Thank you!

- Prof. Jeremy Gunawardena
- Dr. Bobby Karp
- Bianca Dumitrascu & Stefano de Pretis
- HHMI-IUF, FAS Center for Systems Biology
- Greg Llacer & PRISE
- This audience!