Measuring Periodicity in Gene Expression with Persistence

I. Methods : Stability

II. Motivation : Somitogenesis

III. Methods : Measures

IV. Results : Rankings

I. Methods: Stability

I.1 FUNCTIONS

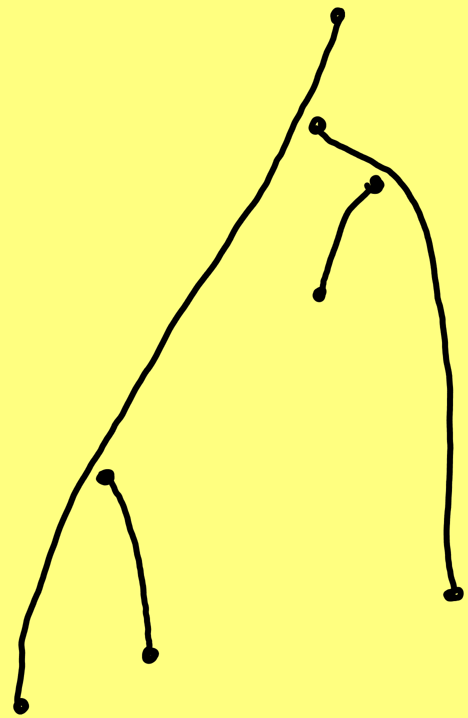
function

$f : S^1 \to \mathbb{R}$

I.1 Functions

function

$f : \mathbb{S}^1 \to \mathbb{R}$



merge tree

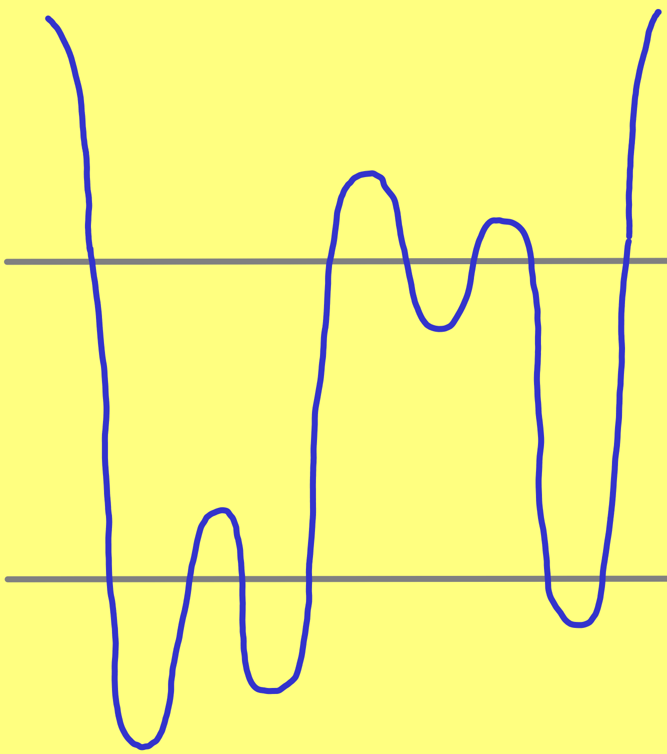
I.1 Functions

function

$f : \mathbb{S}^1 \to \mathbb{R}$

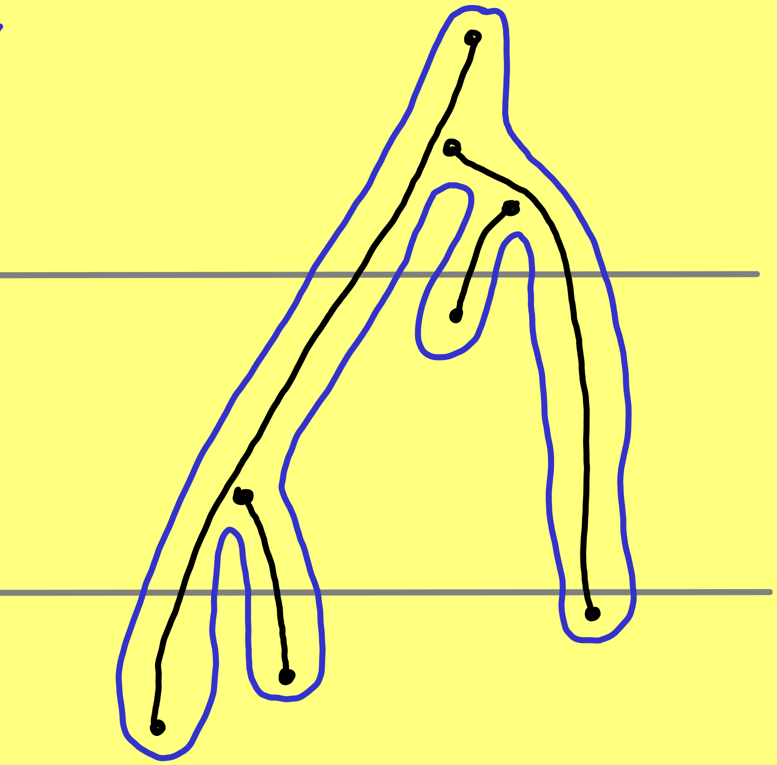
merge tree

I.1 Functions



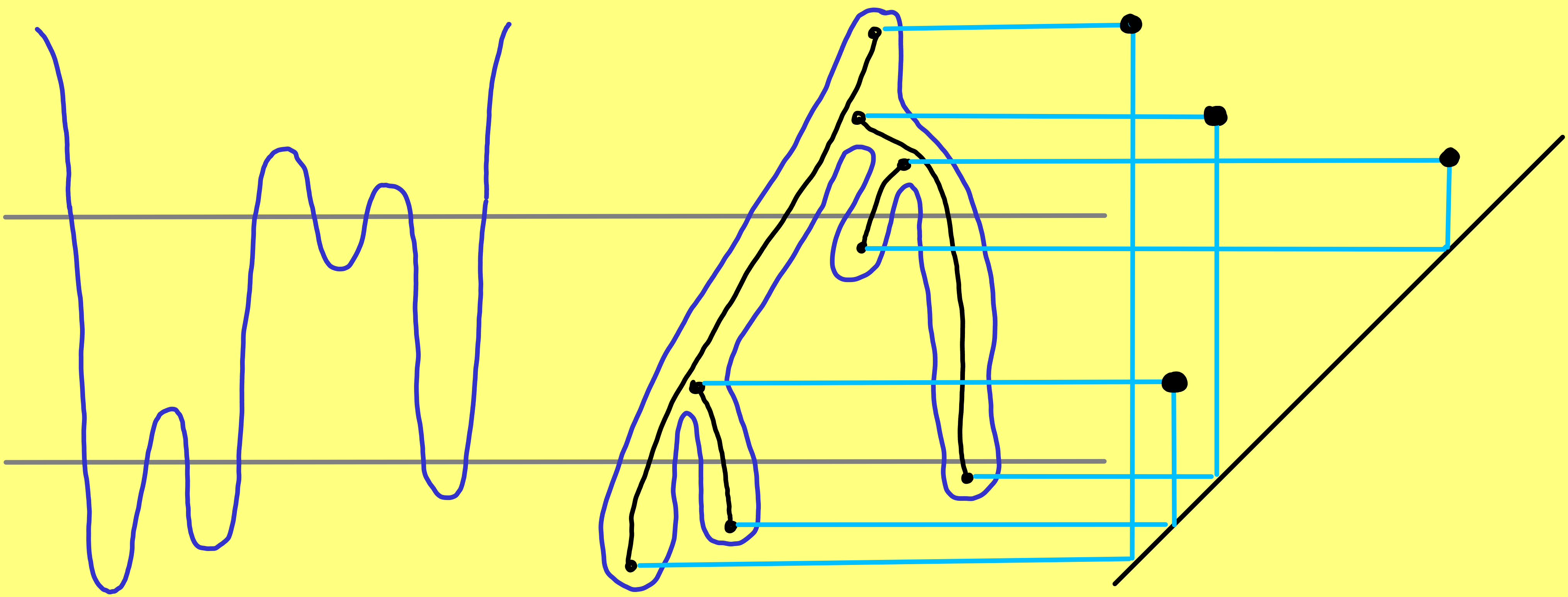
function

$f : \mathbb{S}^1 \to \mathbb{R}$



merge tree

I.1 Functions



function

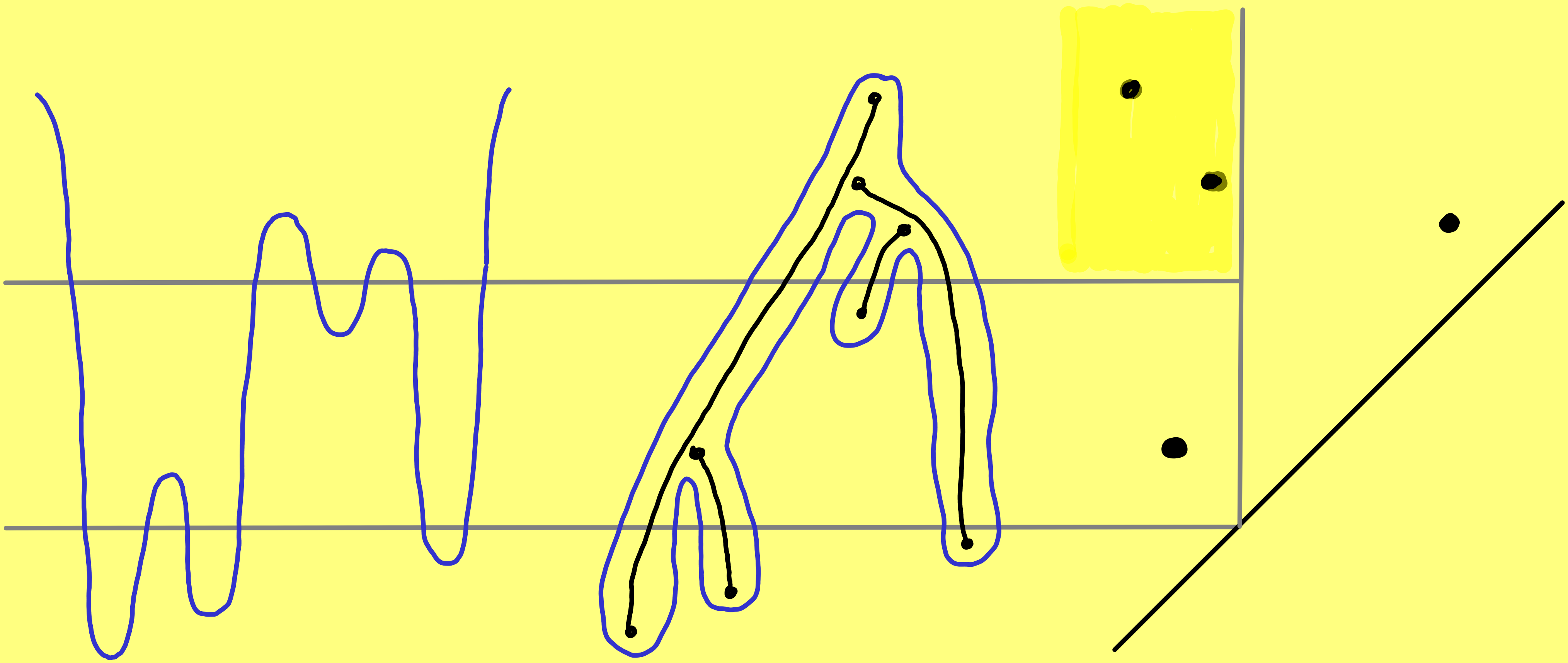
$f: \mathbb{S}^1 \to \mathbb{R}$

merge tree

persistence diagram

$\mathrm{Dgm}(f)$

I.1 Functions

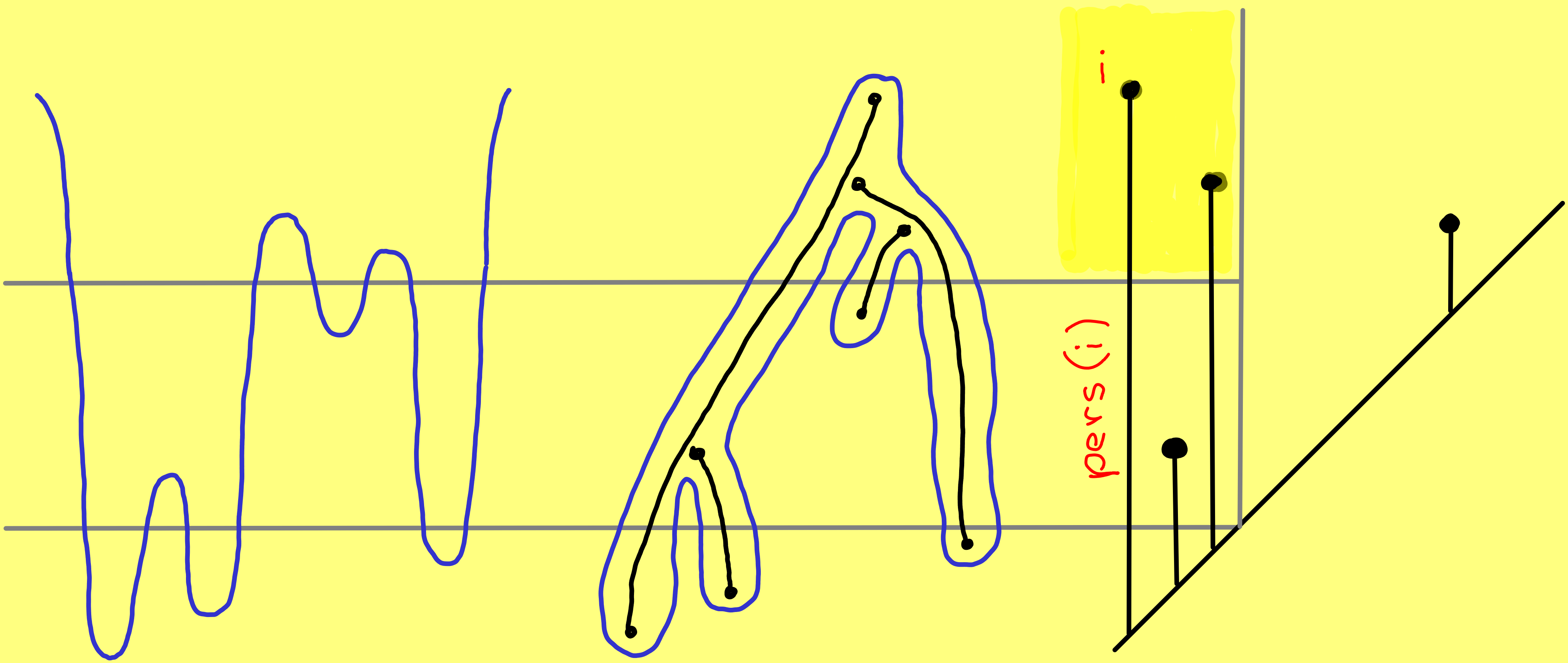


function $f: \mathbb{S}^1 \to \mathbb{R}$

merge tree

persistence diagram $\mathrm{Dgm}(f)$

I.1 FUNCTIONS



function

$f : \mathbb{S}^1 \to \mathbb{R}$

merge tree

persistence diagram

$\mathrm{Dgm}(f)$

I.2 Total Absolute Variation

$$V(f) = \int_0^{2\pi} |f'(t)|\, dt$$

I.2 Total Absolute Variation

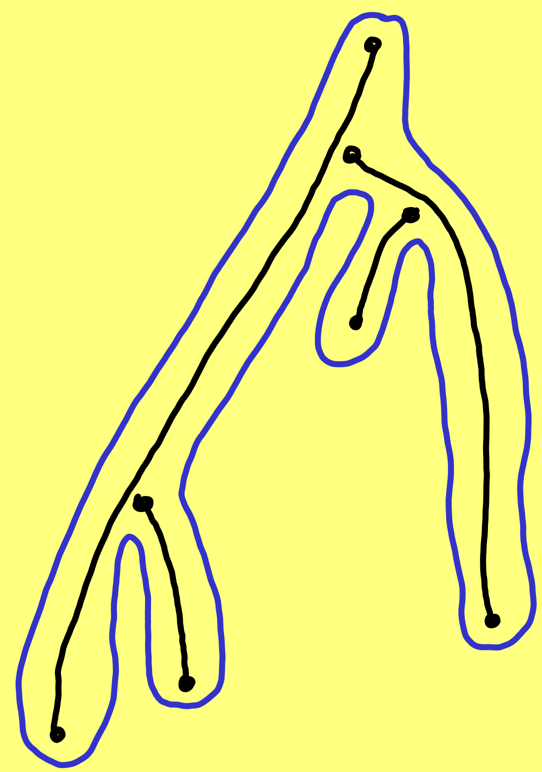
$$V(f) = \int_0^{2\pi} |f'(t)| \, dt$$

$$= 2 \sum_{i \in \mathrm{Dgm}(f)} \mathrm{pers}(i).$$

I.2 Total Absolute Variation

$$V(f) = \int_0^{2\pi} |f'(t)| \, dt$$

$$= 2 \sum_{i \in \mathrm{Dgm}(f)} \mathrm{pers}(i).$$

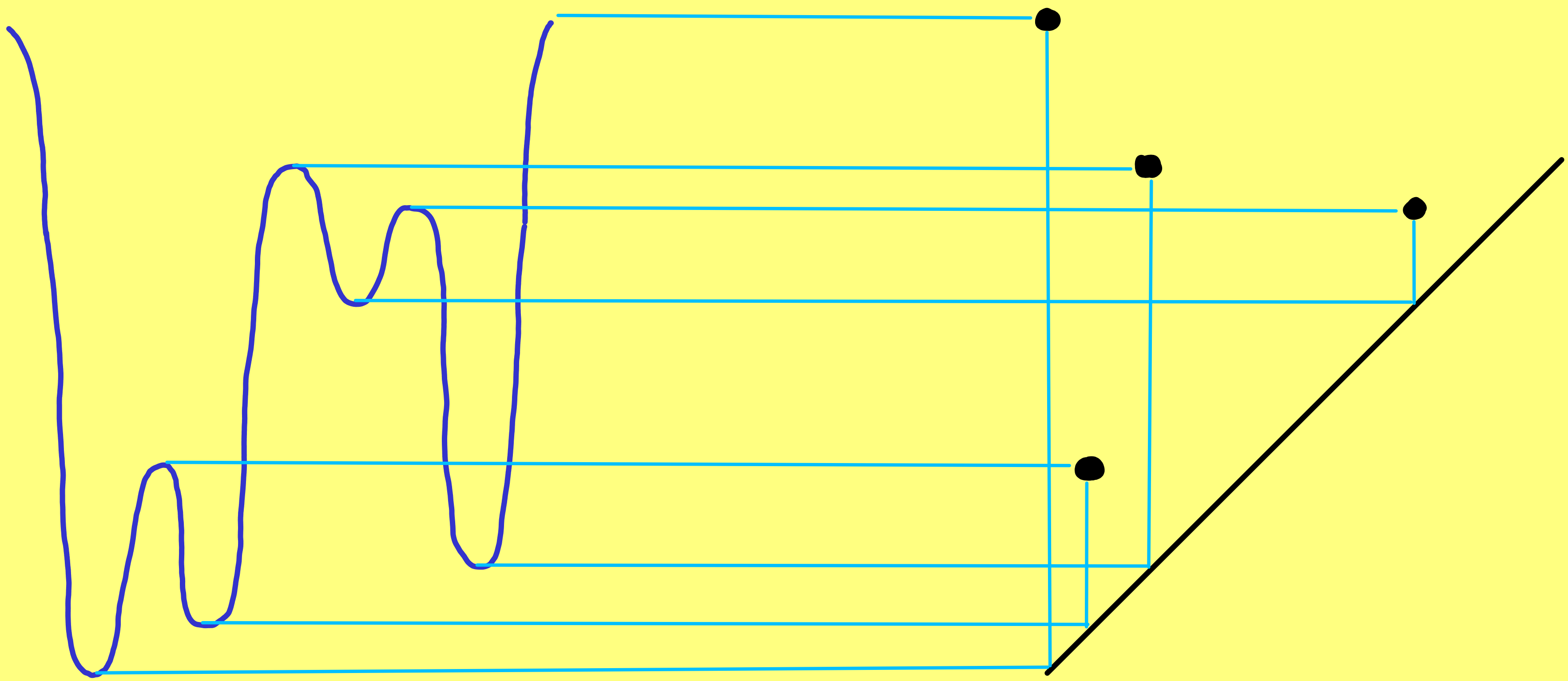


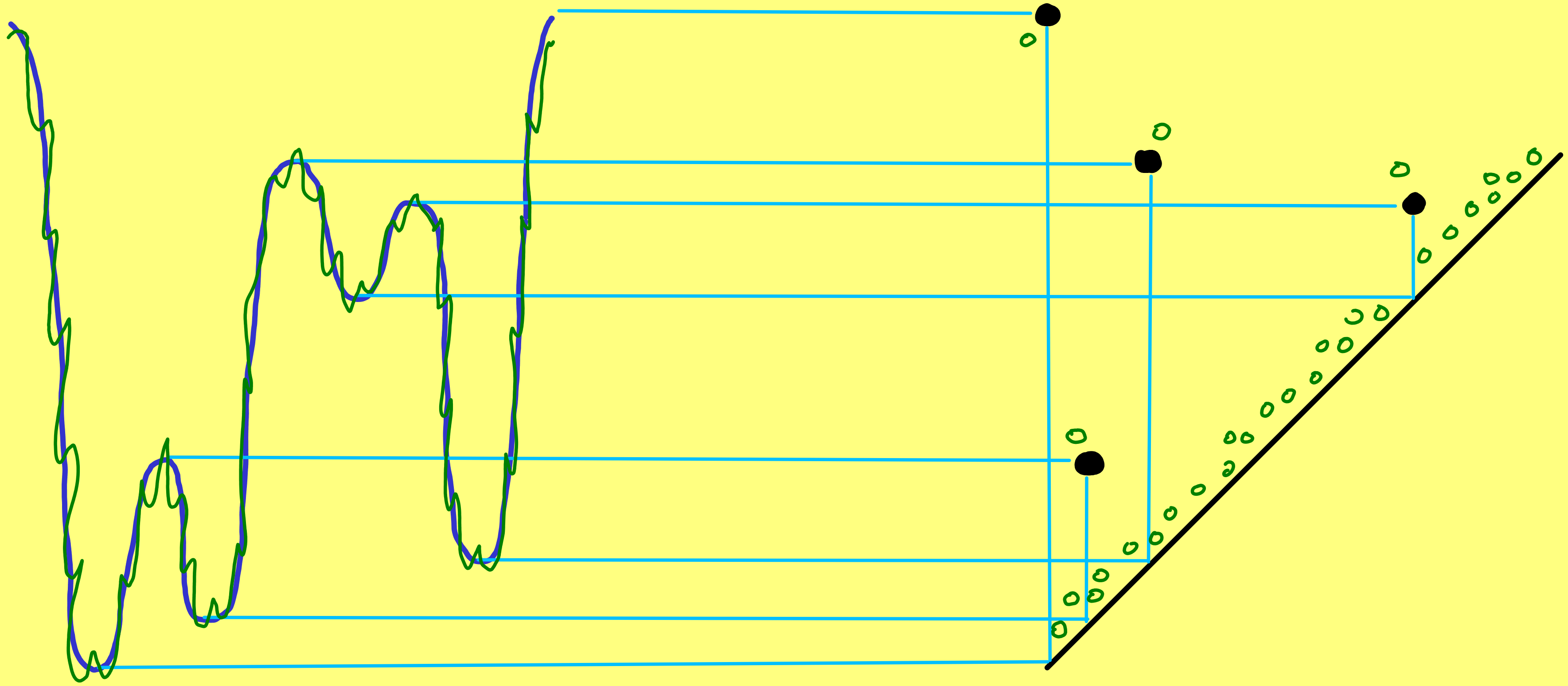
I.3 L_∞ - STABILITY

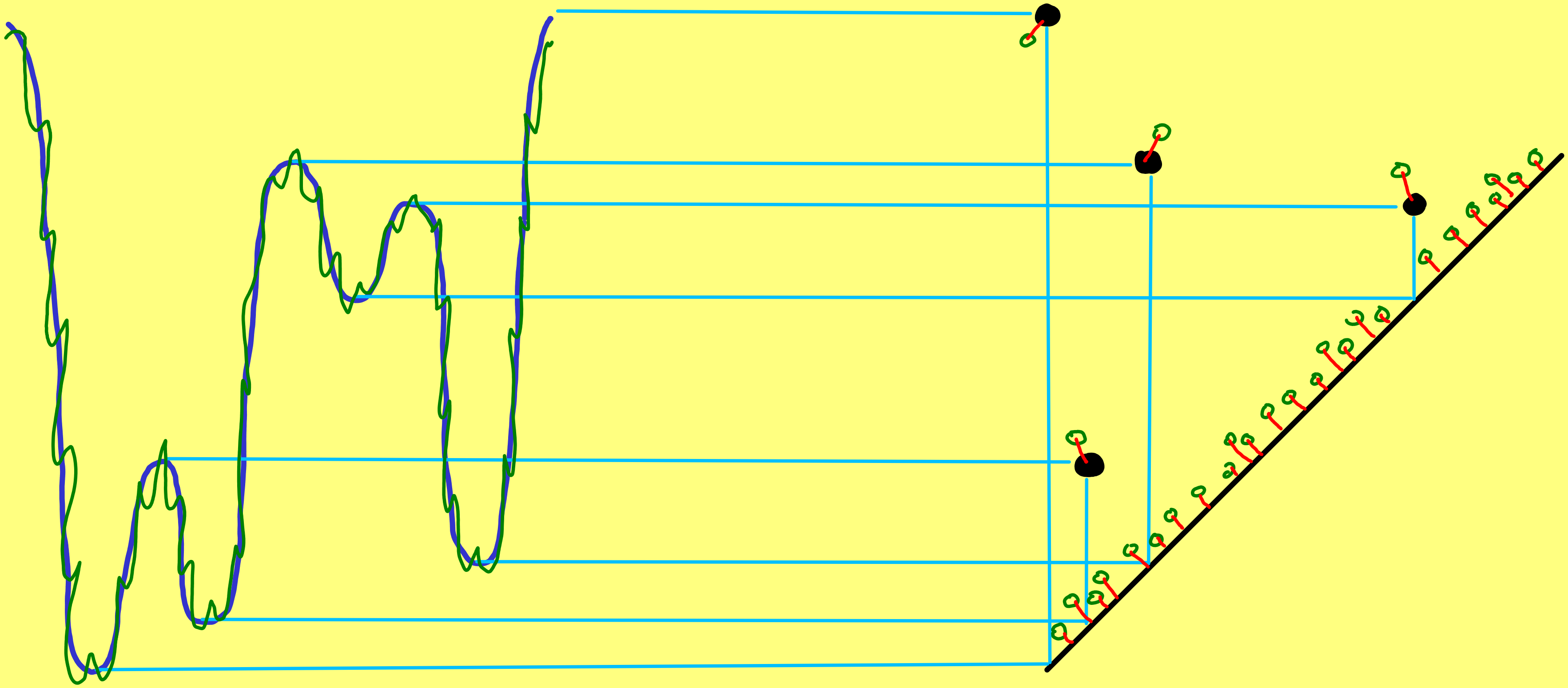
THM. (Cohen-Steiner, E, Harer 04)

For tame functions $f, g : \mathbb{X} \to \mathbb{R}$ the bottle-neck distance between their diagrams is bounded by the max-difference between the functions,

$$d_B(\mathrm{Dgm}(f), \mathrm{Dgm}(g)) \leq \|f - g\|_\infty .$$







$$d_B(\mathrm{Dgm}(f), \mathrm{Dgm}(g)) \leq \|f-g\|_\infty.$$

I.4 LIPSCHITZ

DEF. $f : \mathbb{S}^1 \to \mathbb{R}$ has Lipschitz constant C if $f(x) - f(y) \leq C\|x-y\|$.

I.4 LIPSCHITZ

DEF. $f : \mathbb{S}^1 \to \mathbb{R}$ has Lipschitz constant C if $f(x) - f(y) \le C \|x-y\|$.

$\Longrightarrow$ total absolute variation is $V(f) \le 2C\pi$.

I.5 L_p-STABILITY

THM. (Cohen-Steiner, E, Harer, Mileyko 07)

Let $\mathbb{X}$ be a compact metric space and $f, g : \mathbb{X} \to \mathbb{R}$ two Lipschitz functions. Then there exists a constant k that depends on $\mathbb{X}$ and the Lipschitz constants of f and g such that

$$\left| \sum_{i \in \mathrm{Dgm}(f)} \mathrm{pers}(i)^p - \sum_{j \in \mathrm{Dgm}(g)} \mathrm{pers}(j)^p \right| \leq \mathrm{Const} \cdot \| f - g \|_\infty$$

for every integer $p \geq k + 1$.

I.5 L_p-STABILITY

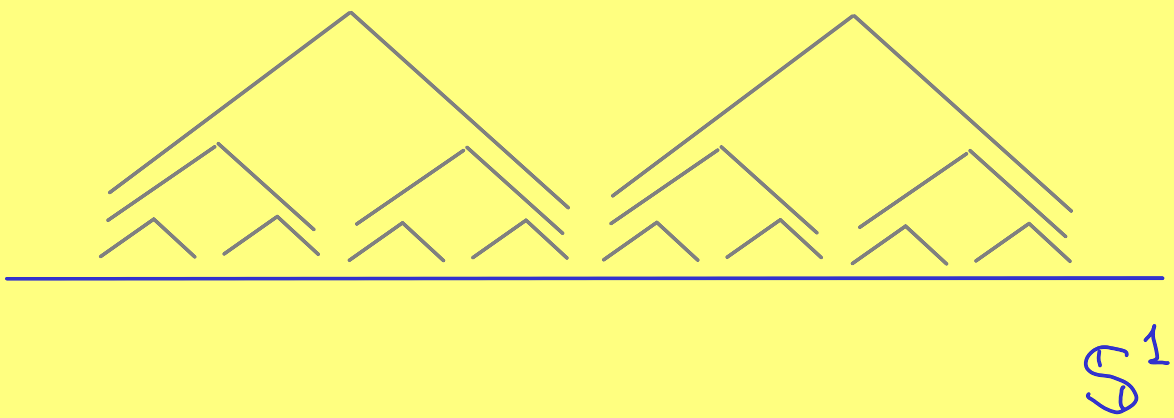
For $\mathbb{X} = \mathbb{S}^1$: $k = 1$.

Hence

$$\left| \sum_{i \in Dgm(f)} pers(i)^2 - \sum_{j \in Dgm(g)} pers(j)^2 \right| \leq Const \cdot \|f - g\|_\infty$$

WHY NOT $p = 1$?

WHY NOT $p = 1$?



II. Motivation: Somitogenesis

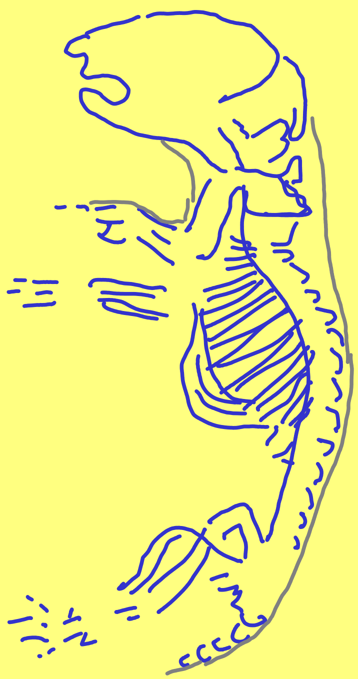
work by Olivier Pourquié

Mary-Lee Dequéant

II.1 SEGMENTATION of vertebrate body plan

II.1 SEGMENTATION of vertebrate body plan

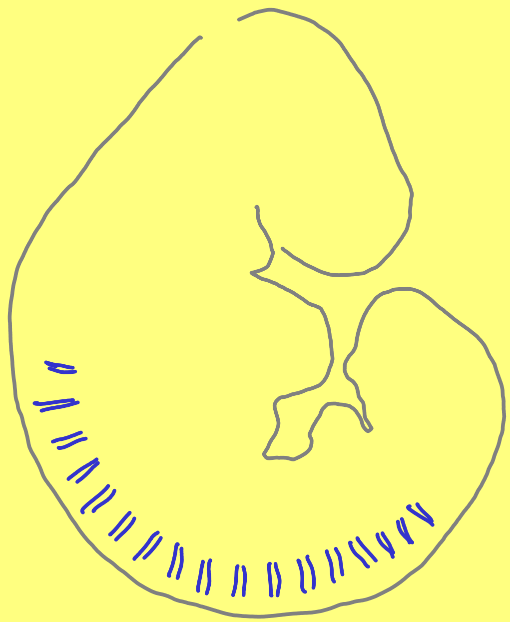
adult mouse



II.1 SEGMENTATION of vertebrate body plan

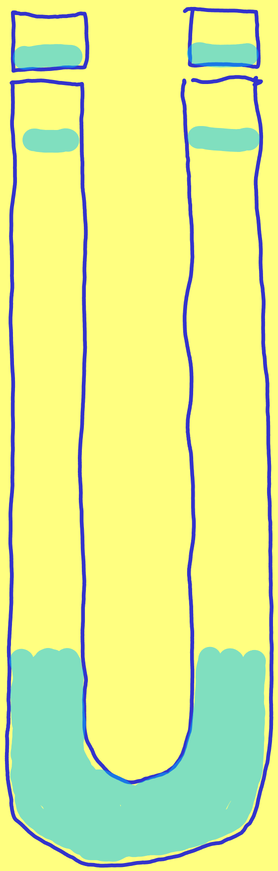
adult mouse

mouse embryo

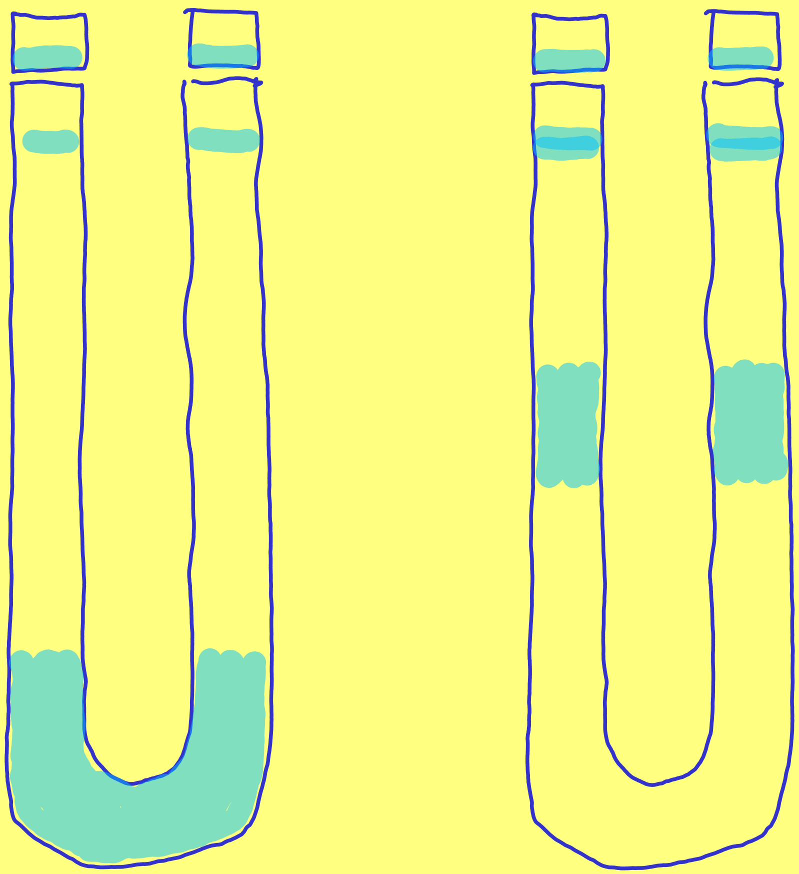


II.2 SOMITOGENESIS . . . a periodic event

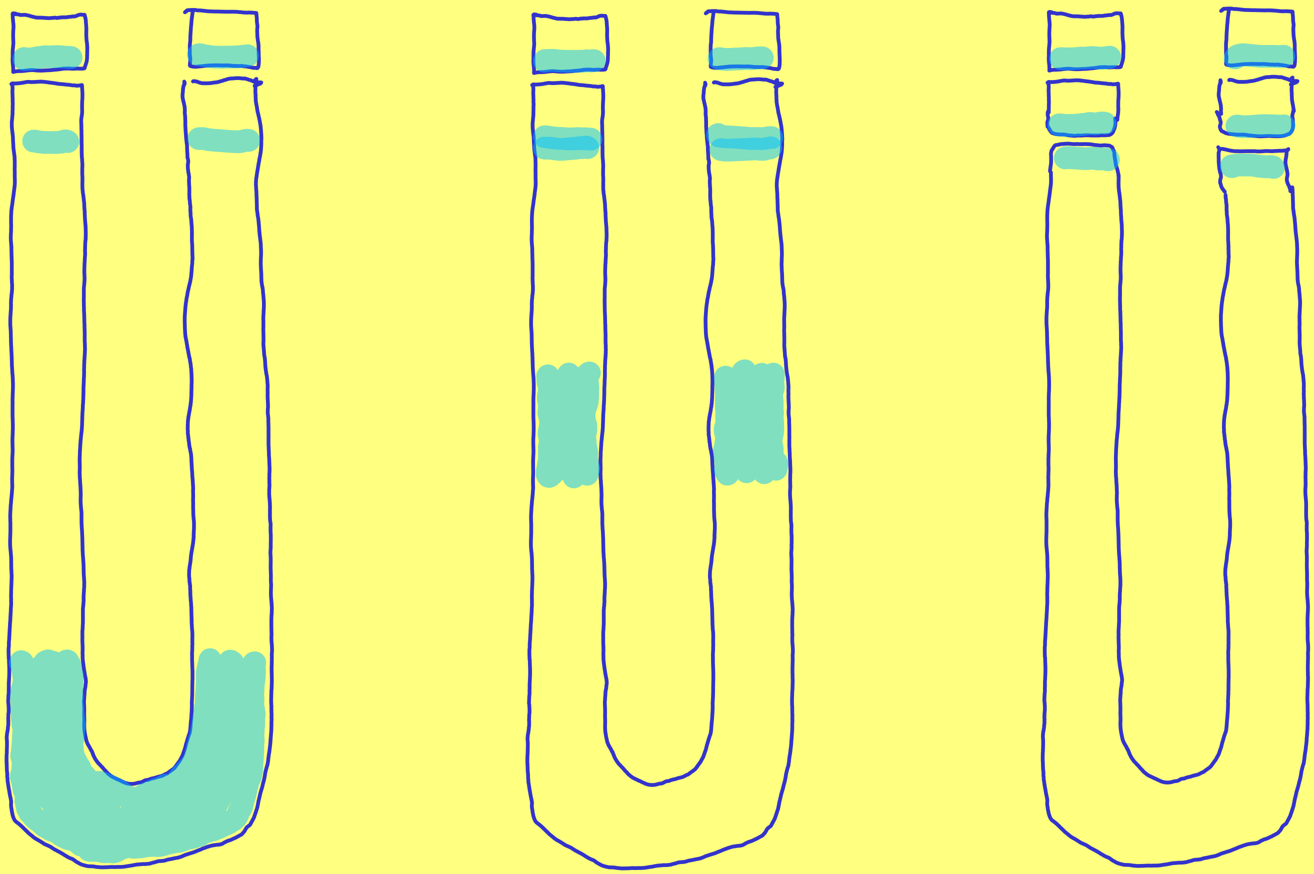
II.2 SOMITOGENESIS ... a periodic event



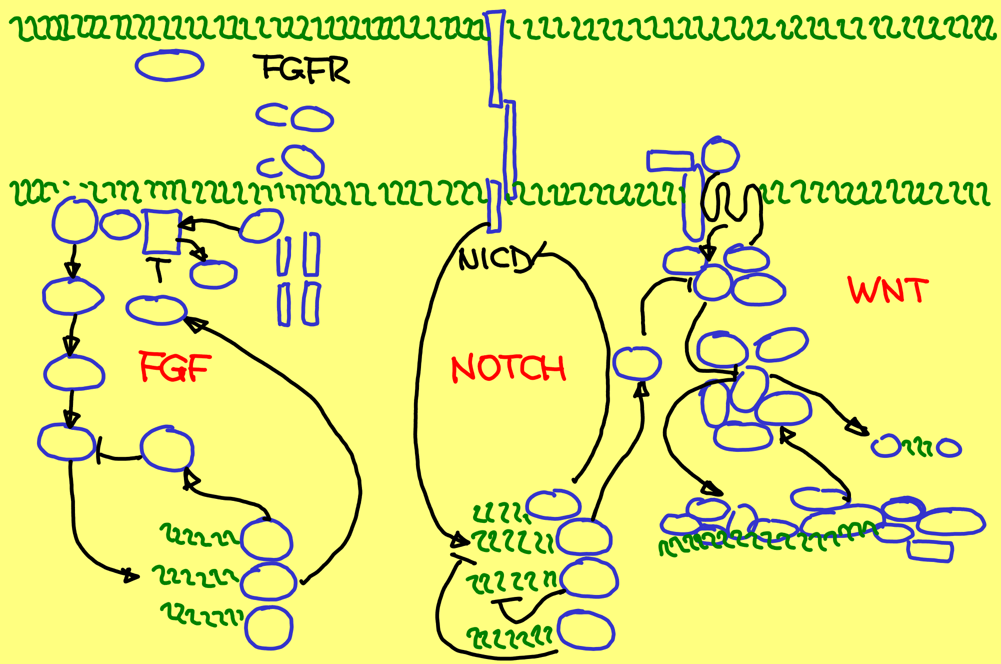
II.2 SOMITOGENESIS ... a periodic event



II.2 SOMITOGENESIS ... a periodic event



II.3 THE CLOCK



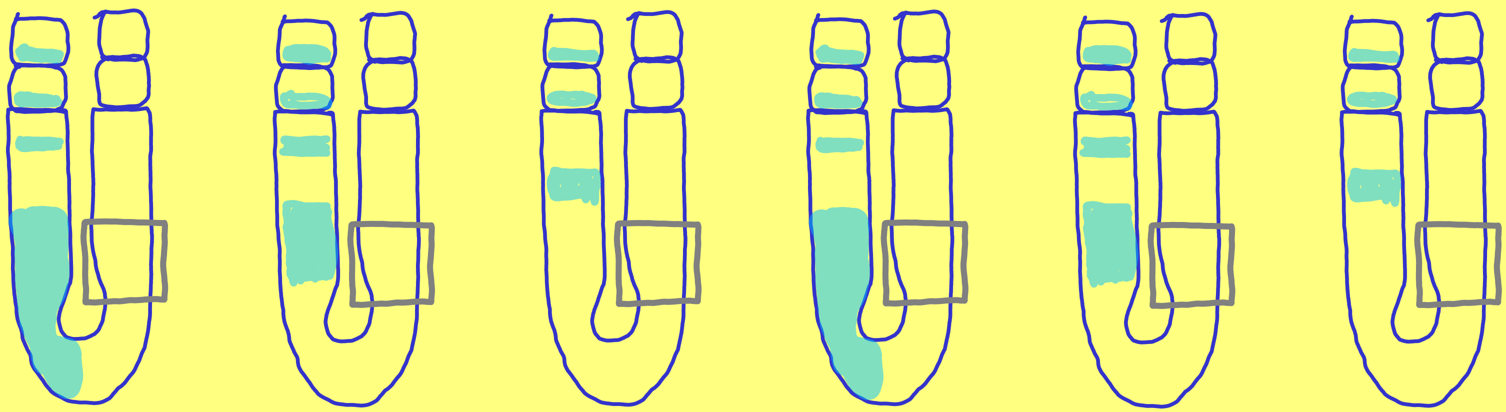
proposed model

II.4 Cyclic Genes

... time series microarray analysis

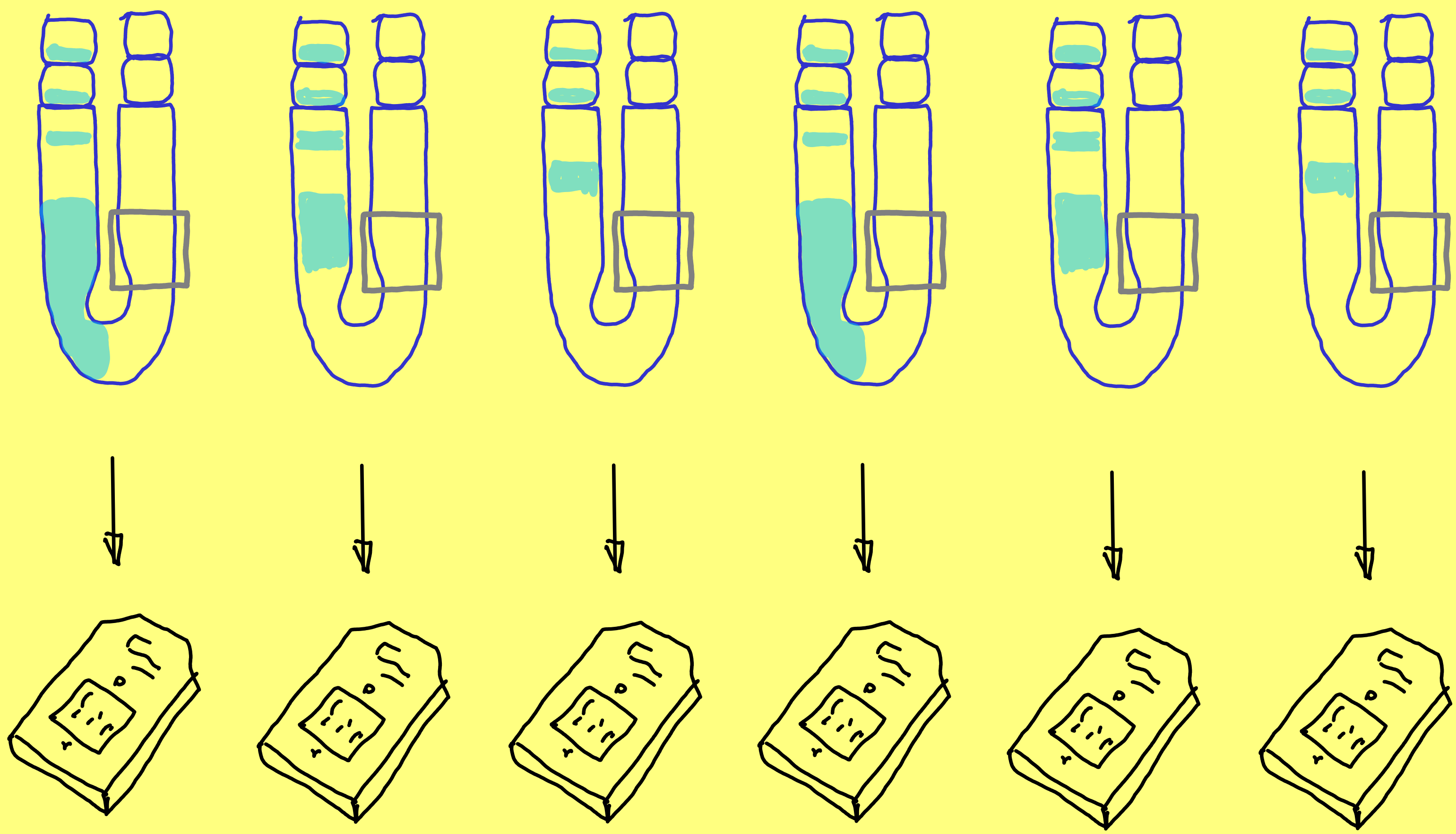
II.4 Cyclic Genes

... time series microarray analysis



II.4 Cyclic Genes

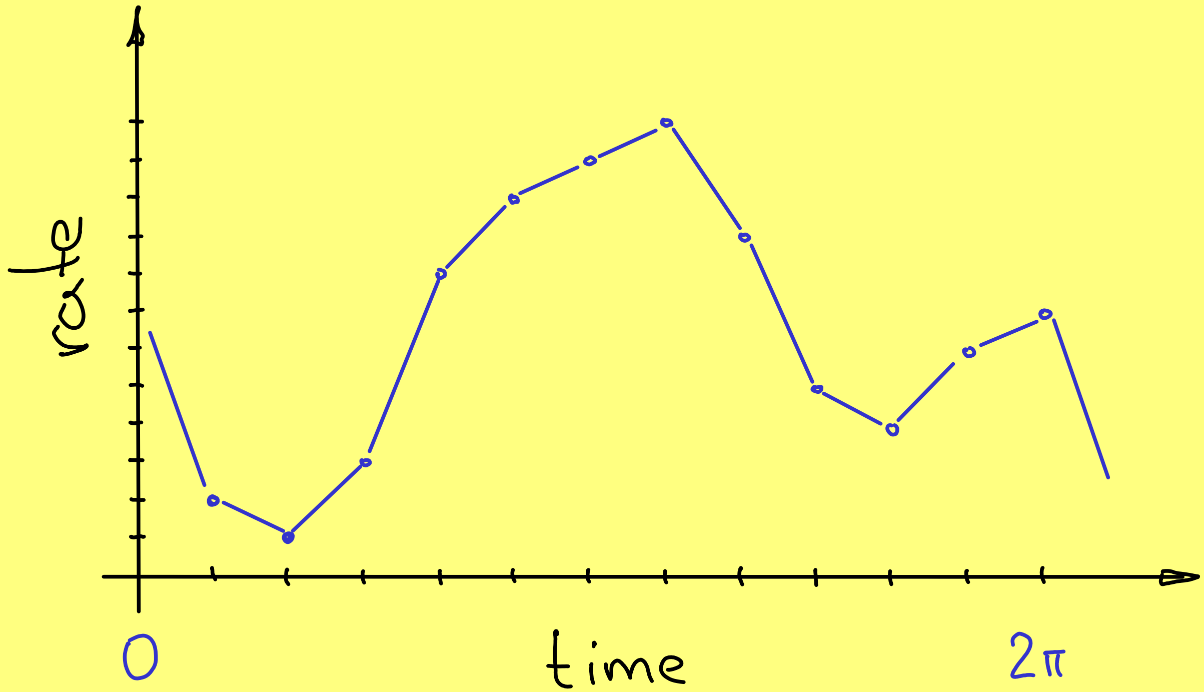
... time series microarray analysis



III. Methods: Measures

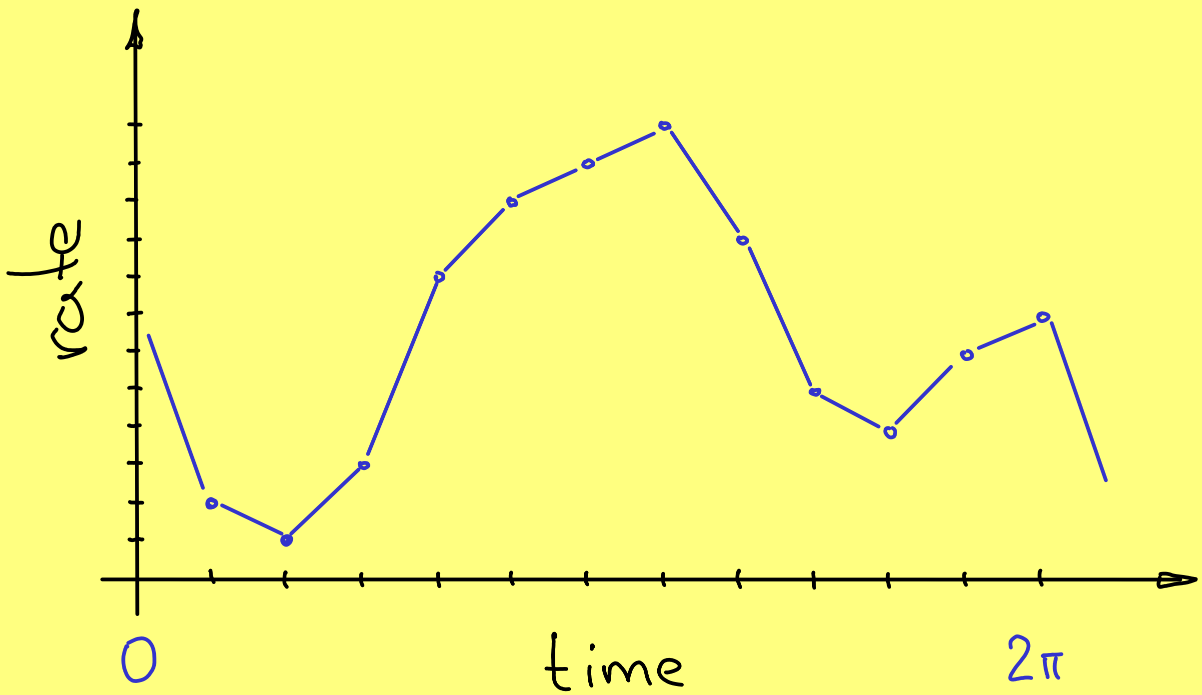
III.1 Data Normalization

pts. in time, expression rates $\rightarrow$ ranks



III.1 Data Normalization

pts. in time, expression rates $\longrightarrow$ ranks



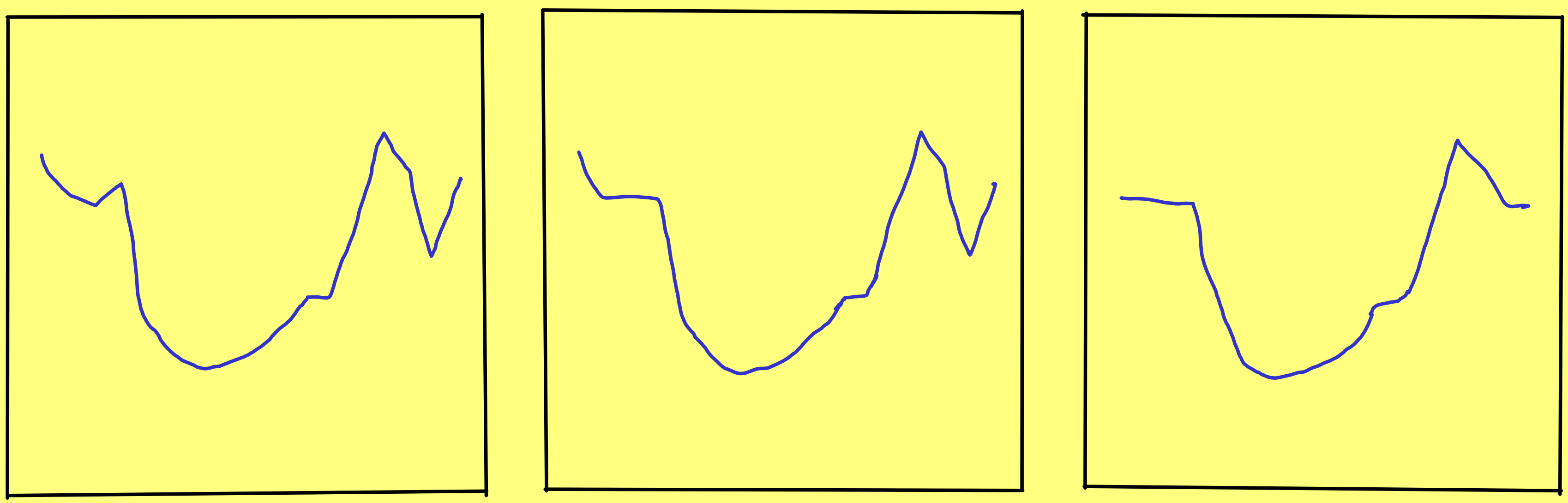
normalized to $\operatorname{amp}(f) = \max_x f(x) - \min_x f(x) = 1$

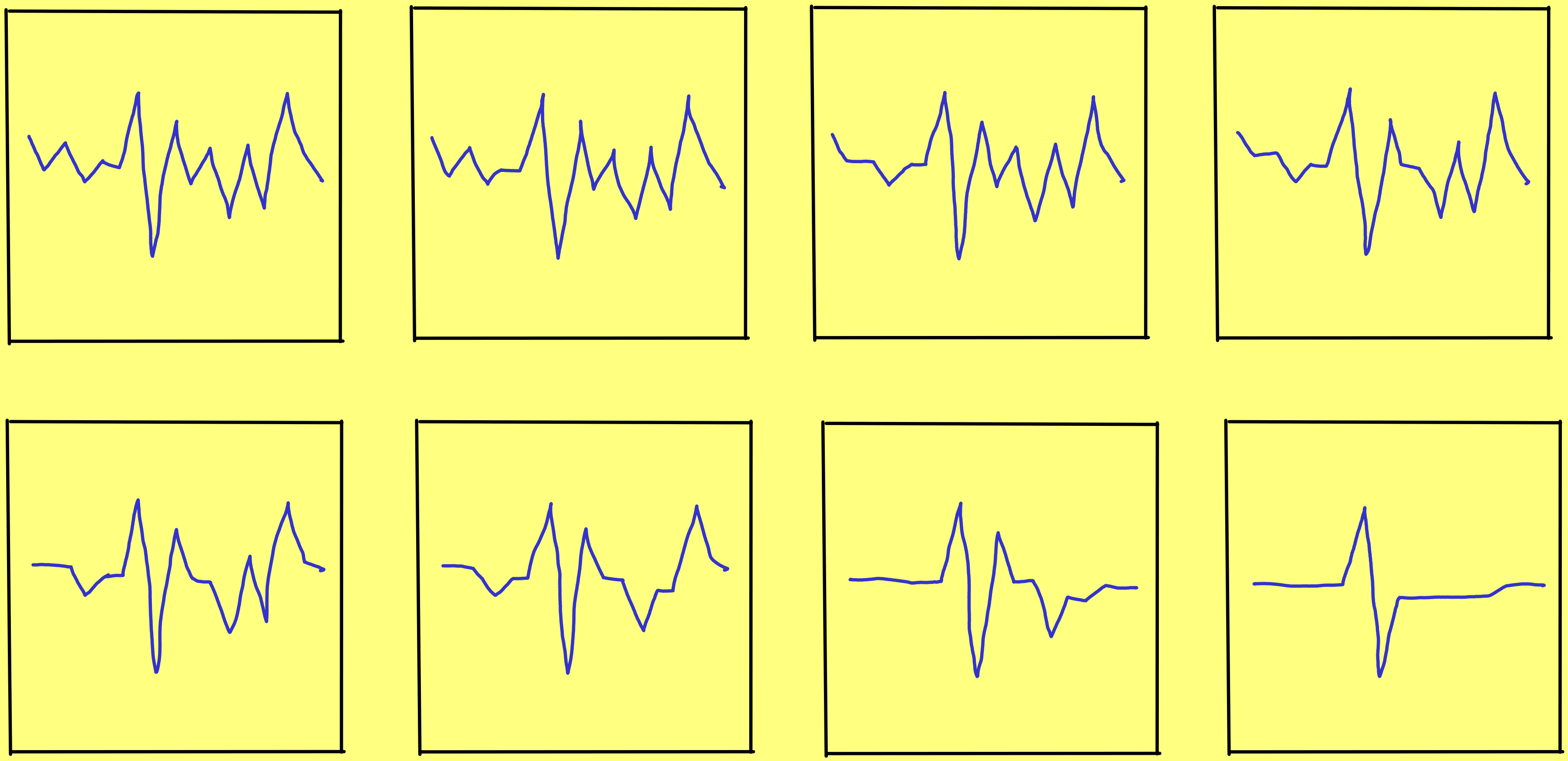
III.2 Simplification

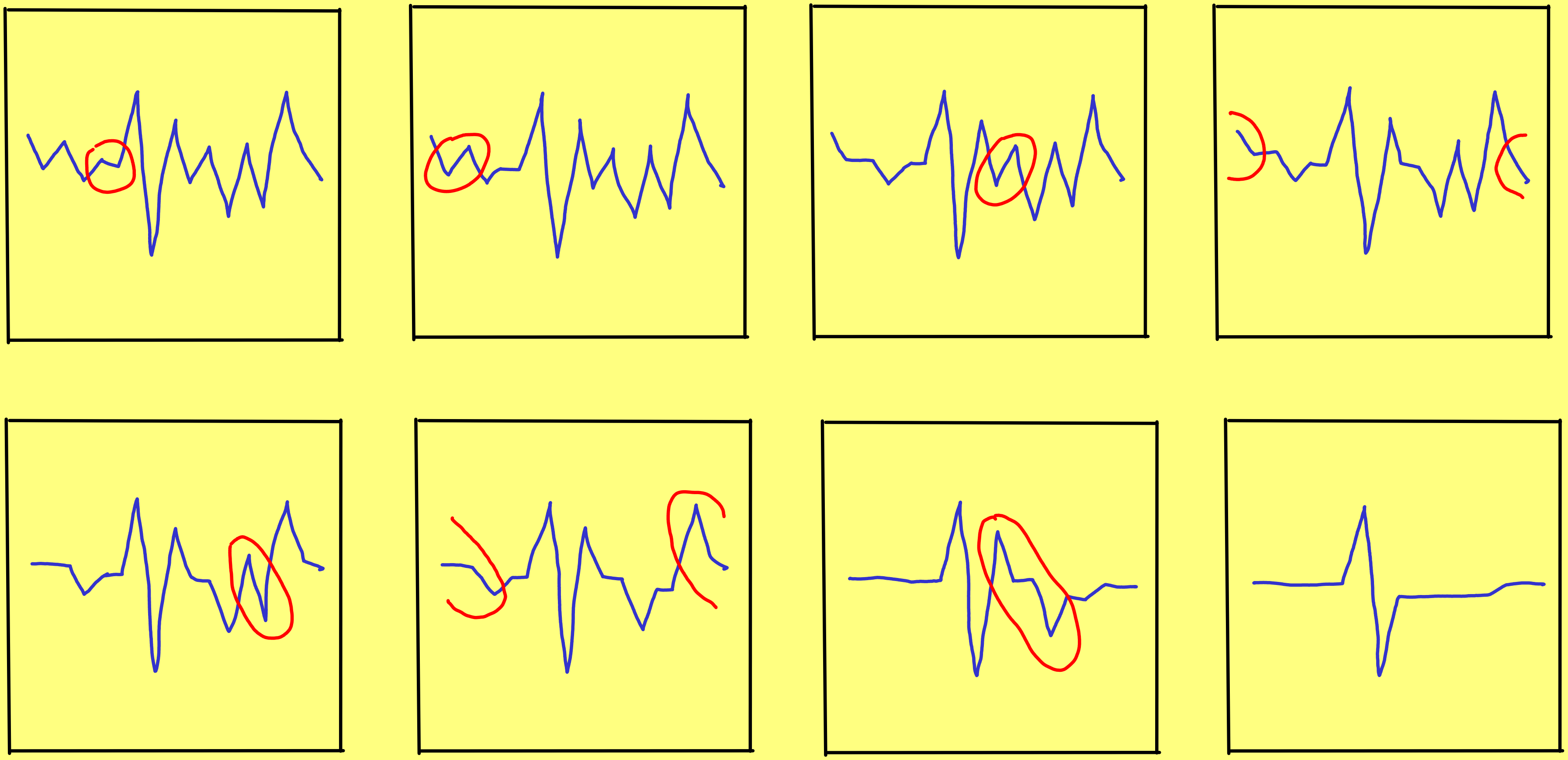
Def. (E, Morozov, Pascucci 06) An ε-simplification of f is a function $f_\varepsilon : \mathbb{X} \to \mathbb{R}$ with $\|f - f_\varepsilon\|_\infty \leq \varepsilon$ whose diagram $\mathrm{Dgm}(f_\varepsilon)$ is same as $\mathrm{Dgm}(f)$ without points of persistence at most ε.

III.2 SIMPLIFICATION

DEF. (E, Morozov, Pascucci 06) An ε-simplification of f is a function $f_\varepsilon : \mathbb{X} \to \mathbb{R}$ with $\| f - f_\varepsilon \|_\infty \leq \varepsilon$ whose diagram $\mathrm{Dgm}(f_\varepsilon)$ is same as $\mathrm{Dgm}(f)$ without points of persistence at most ε.







III.3 Periodicity Measures

$f : S^1 \to \mathbb{R}$

$M_0(f) = \frac{1}{2} \#$critical points

III.3 Periodicity Measures

$f : \mathbb{S}^1 \to \mathbb{R}$

$$M_0(f) = \tfrac{1}{2}\,\#\text{critical points}$$

$$M_1(f) = \int_{\varepsilon=0}^{1} M_0(f_\varepsilon)\,d\varepsilon = \sum_i \mathrm{pers}(i)$$

III.3 Periodicity Measures

$f: \mathbb{S}^1 \to \mathbb{R}$

$$M_0(f) = \frac{1}{2} \# \text{critical points}$$

$$M_1(f) = \int_{\varepsilon=0}^{1} M_0(f_\varepsilon)\, d\varepsilon = \sum_i \text{pers}(i)$$

$$M_2(f) = \int_{\varepsilon=0}^{1} M_1(f_\varepsilon)\, d\varepsilon = \sum_i \text{pers}(i)^2$$

etc.

III.3 Periodicity Measures

$$f : \mathbb{S}^1 \to \mathbb{R}$$

not stable

$$M_0(f) = \tfrac{1}{2} \,\#\text{critical points}$$

$$M_1(f) = \int_{\varepsilon=0}^{1} M_0(f_\varepsilon)\, d\varepsilon = \sum_i \mathrm{pers}(i)$$

stable

$$M_2(f) = \int_{\varepsilon=0}^{1} M_1(f_\varepsilon)\, d\varepsilon = \sum_i \mathrm{pers}(i)^2$$

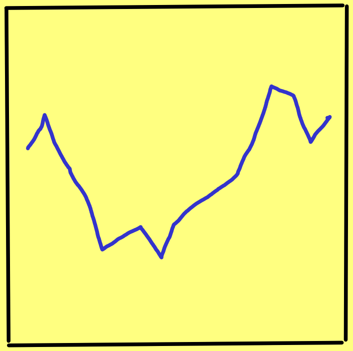
etc.

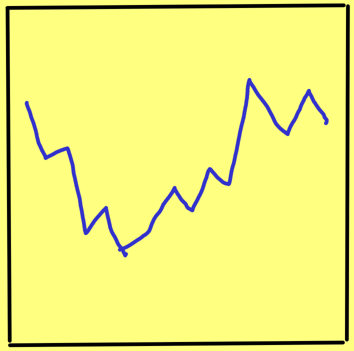
IV. Results: Rankings

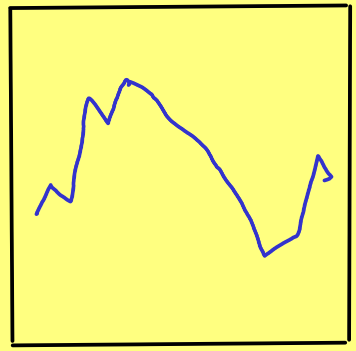
IV.1 VERIFIED - 30

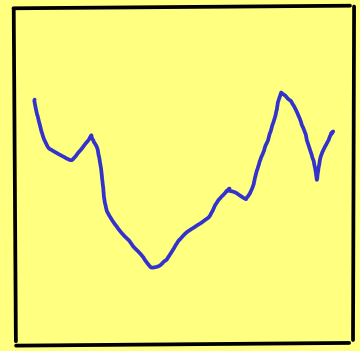
(biologically confirmed to participate in the periodic process of somite development)

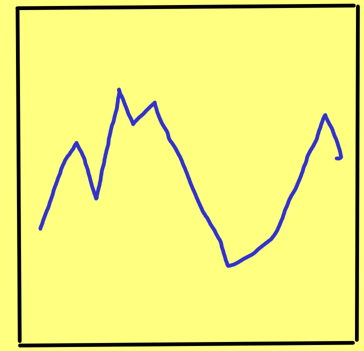
Dkk1	Tnfrsf19	Hes1	Axin2	Hspg2
Myc	Hes5	Dact1	Sp5	Efna1
Bcl2l11	a-Tnfrsf19	Lfng	Spry2	Klf10
s-Dsp	Hey1	Plxdc2	Nudt13	Bcl9l
Id1	Has2	5-Nrarp	Dsp	Phlda1
Arfl4	Nkd1	6-Nrarp	x-Cyr61	a-Cyr61


Dkk1

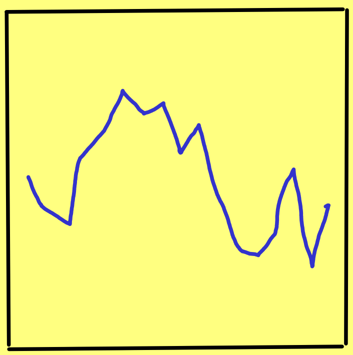

Tnfrsf19

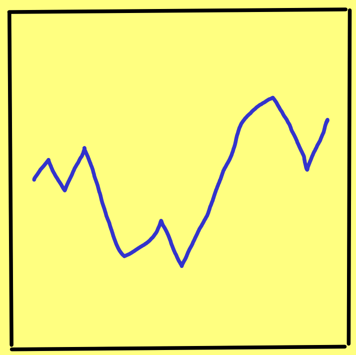

Hes1

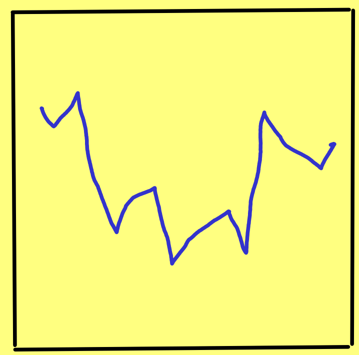

Axin2

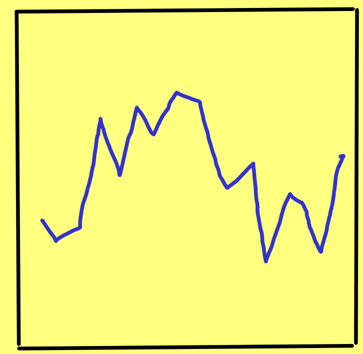

Hspg2


Myc

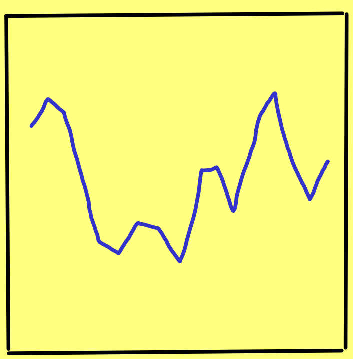

Hes5

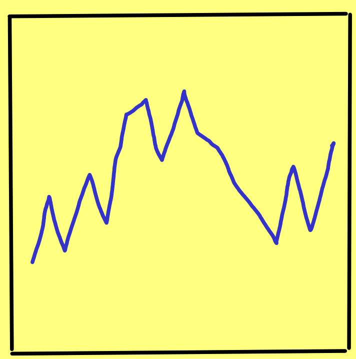

Dact1

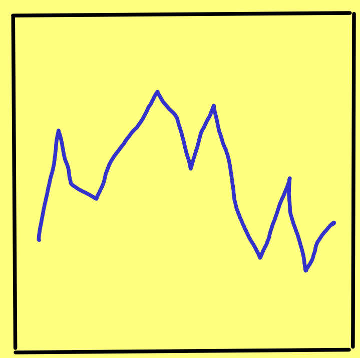

Sp5

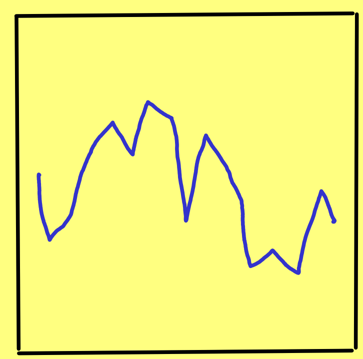

Efna1

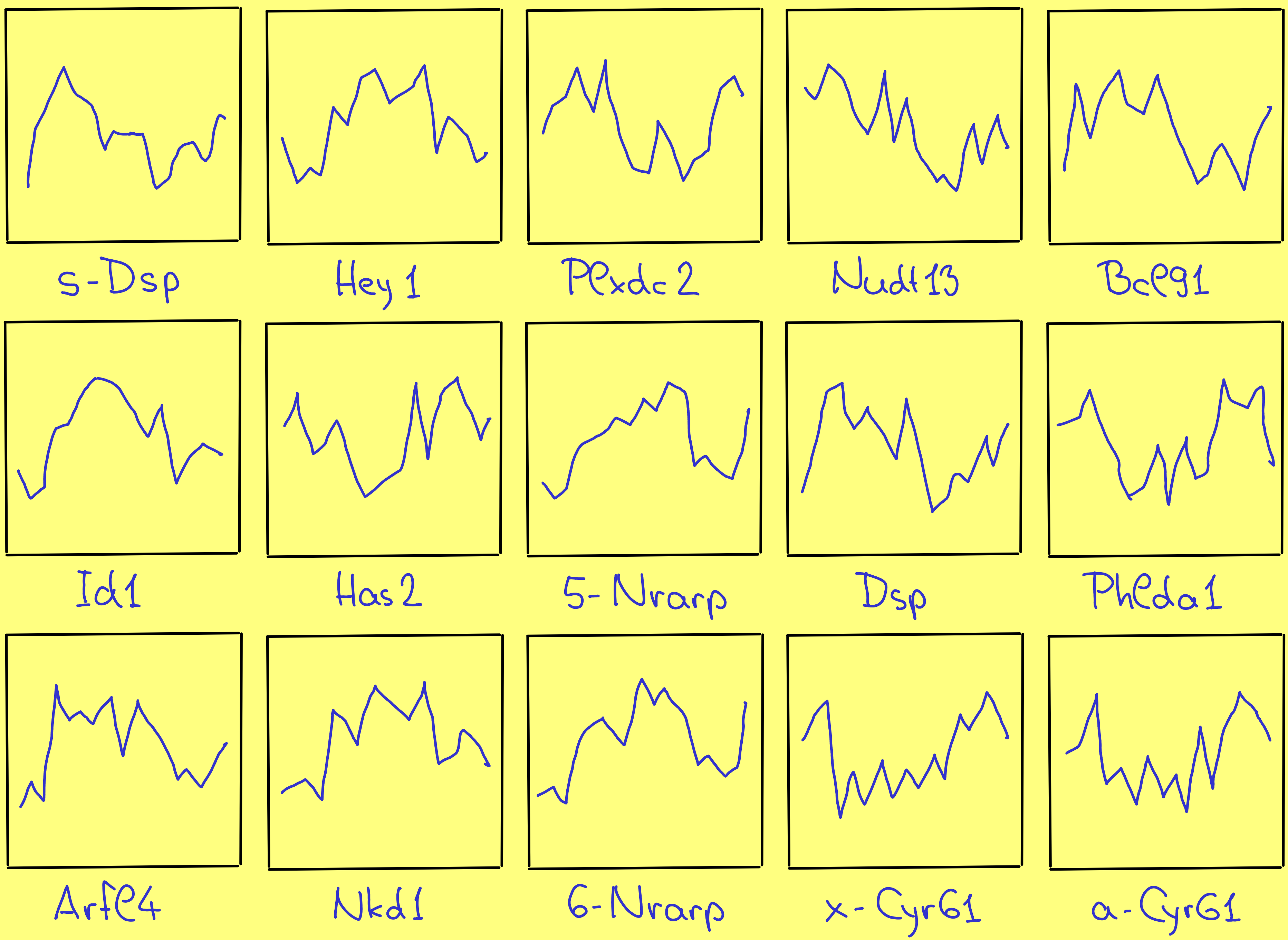

Bcl2l11


a-Tfnrsf19


Lnfg


Spry2


Klf10

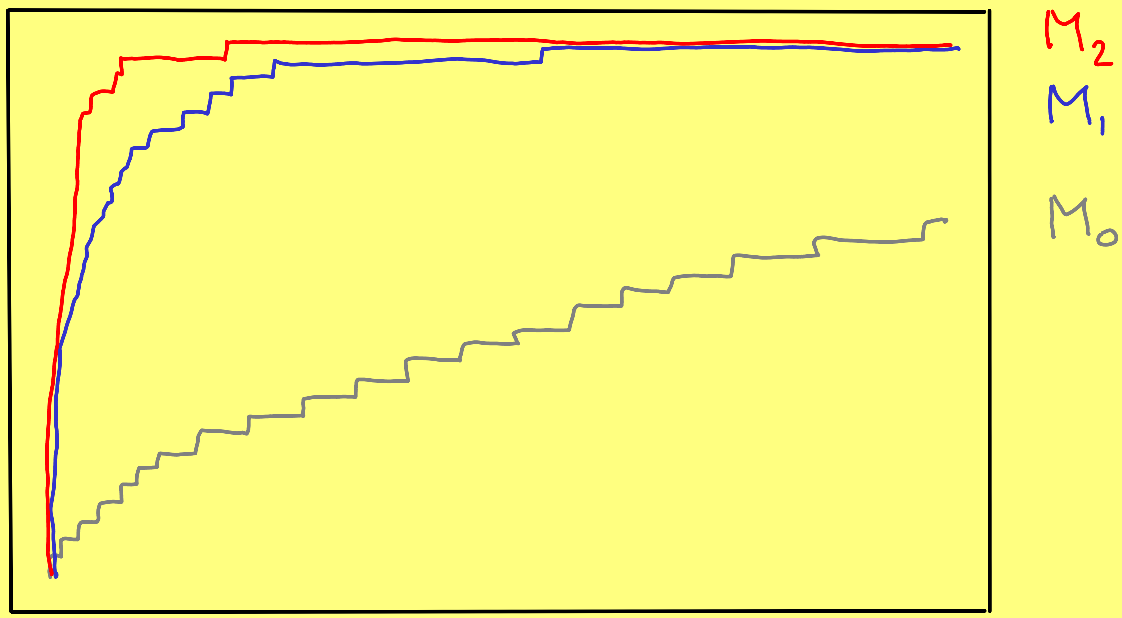

s-Dsp
Hey 1
Plxdc 2
Nudt13
Bcl9l
Id1
Has 2
5-Nrarp
Dsp
Phlda1
Arl4
Nkd1
6-Nrarp
x-Cyr61
a-Cyr61

IV.2 Internal Comparison

	M_1	M_2	M_3	M_4	avg
Dkk1	1	1	1	1	1
Tnfrsf19	12	2	2	2	5
Hey1	23	12	9	9	13
Nudt13	882	170	104	83	310
Arfl4	302	86	74	68	133
...	...	...	...	...	...
avg	250	109	104	117	145

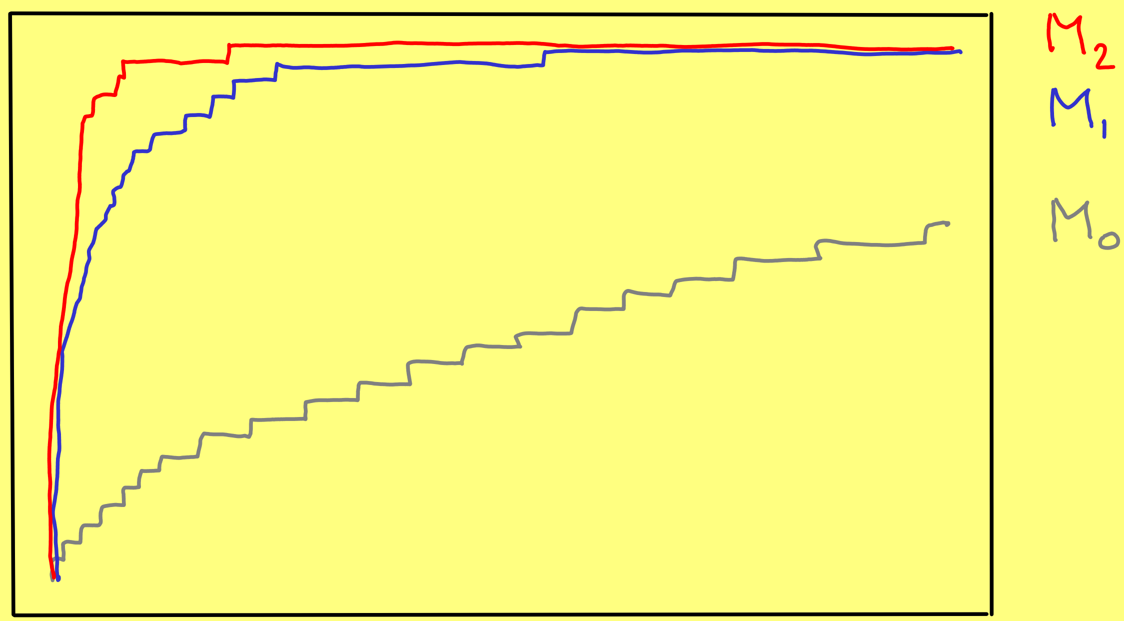
IV.2 Internal Comparison

ROC Curves



IV.2 Internal Comparison

ROC Curves



	M_0	M_1	M_2	M_3	M_4	max
Area	4.42	9.75	10.18	10.19	10.15	10.50

IV.2 Internal Comparison

Promising Candidates

	M_1	M_2	M_3	M_4	L
Tnfrsf9	5	4	4	6	8
Ptpn11	8	6	6	7	22
Mtm1	9	7	5	3	15
Stom	7	8	12	16	299
Star	6	9	11	13	39
...	...	...	...	...	...

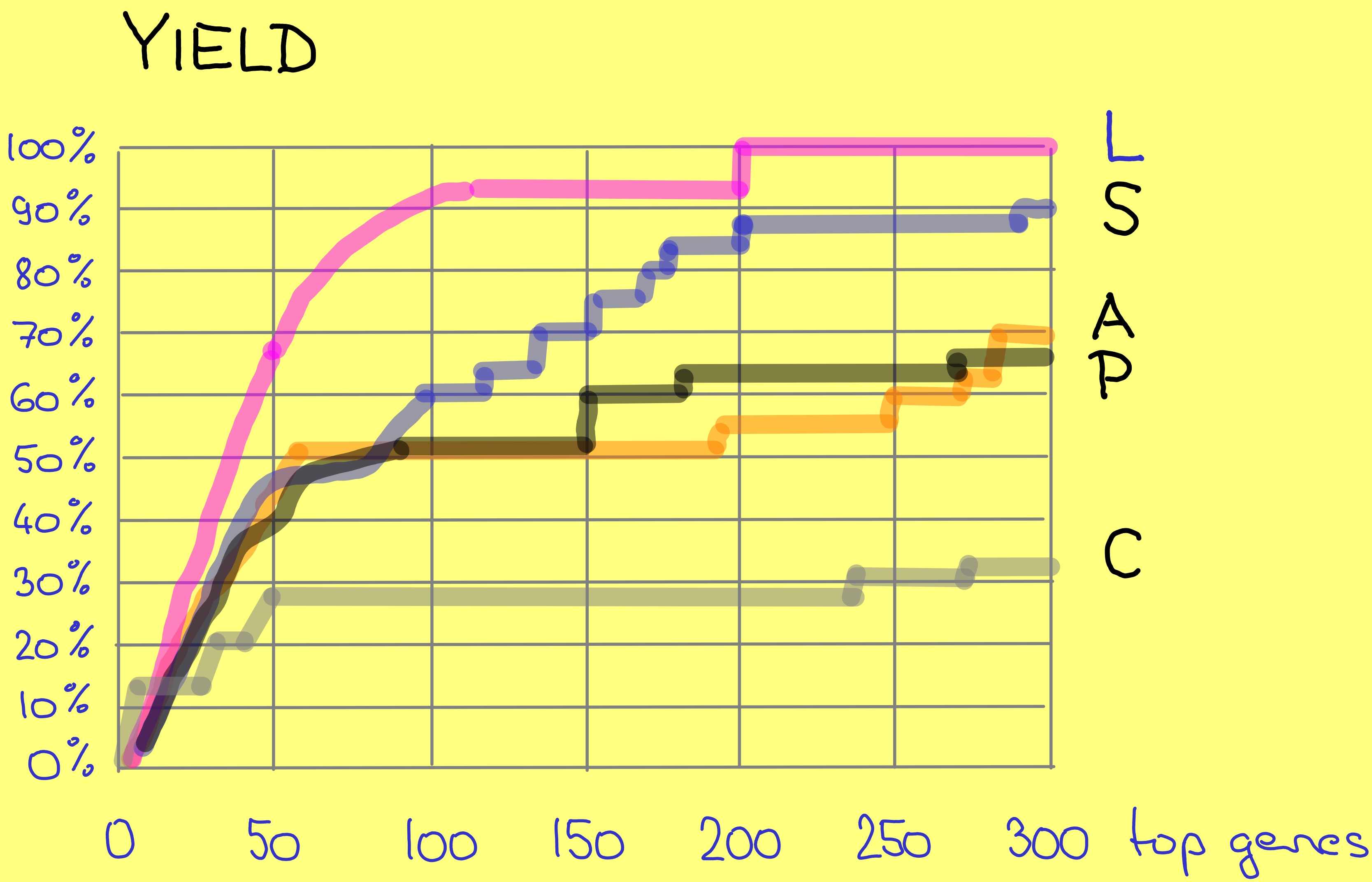
IV.3 External Comparison

	1 + 4 Methods	
L	Lomb-Scargle	Stowers
P	Phase Consistency	Dallas
A	Address Reduction	Marie Curie
C	Cyclohedron Test	Berkeley
S	Stable Persistence	Duke

IV.3 External Comparison

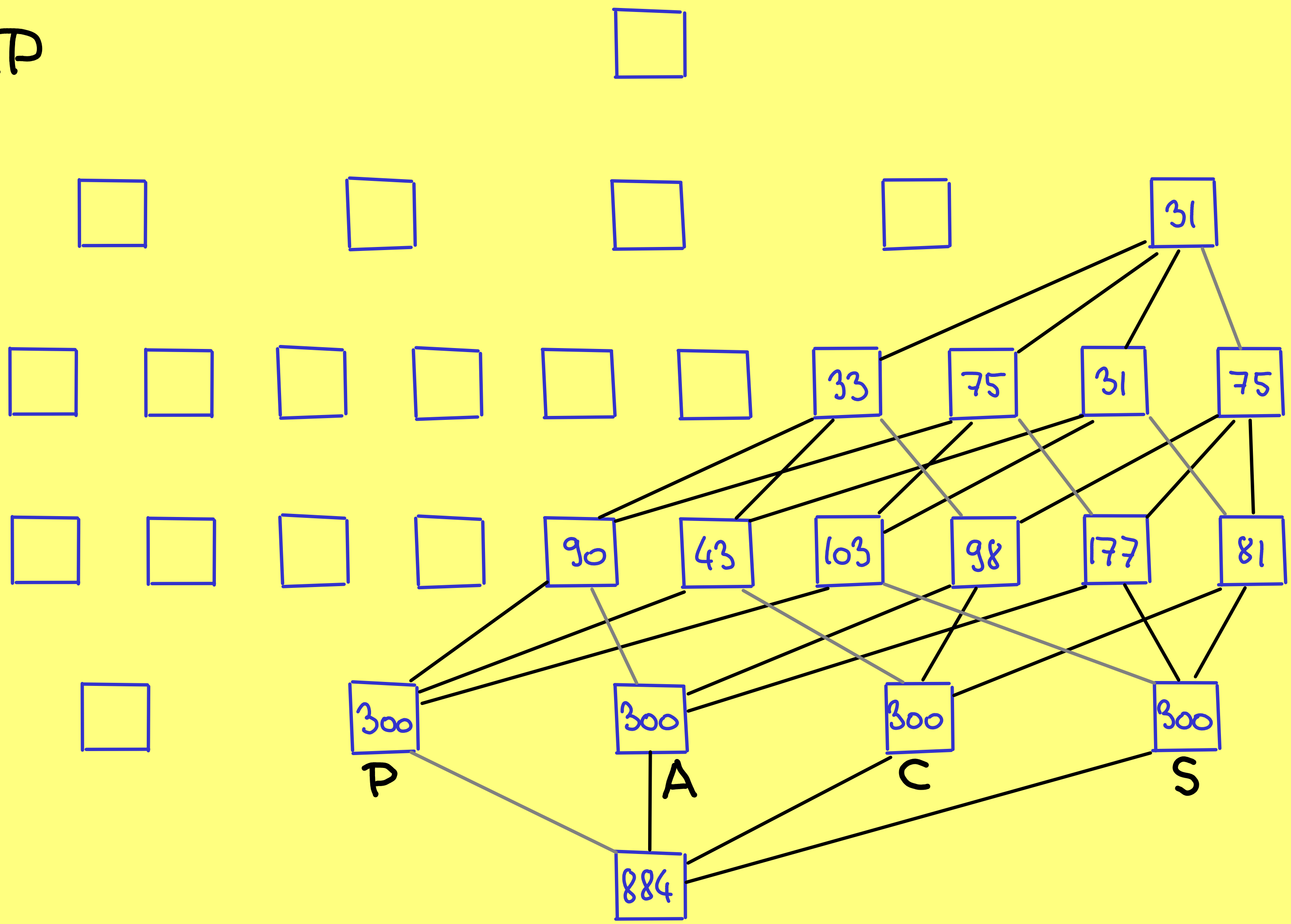
	1 + 4 Methods	
L	Glynn, Hattem, Mishegian	Stowers
P	Kudlicki, Rowicka	Dallas
A	Ahnert, Fink	Marie Curie
C	Morton, Pachter, Shiu, Sturmfels	Berkeley
S	Edelsbrunner, Mileyko	Duke

IV.3 External Comparison



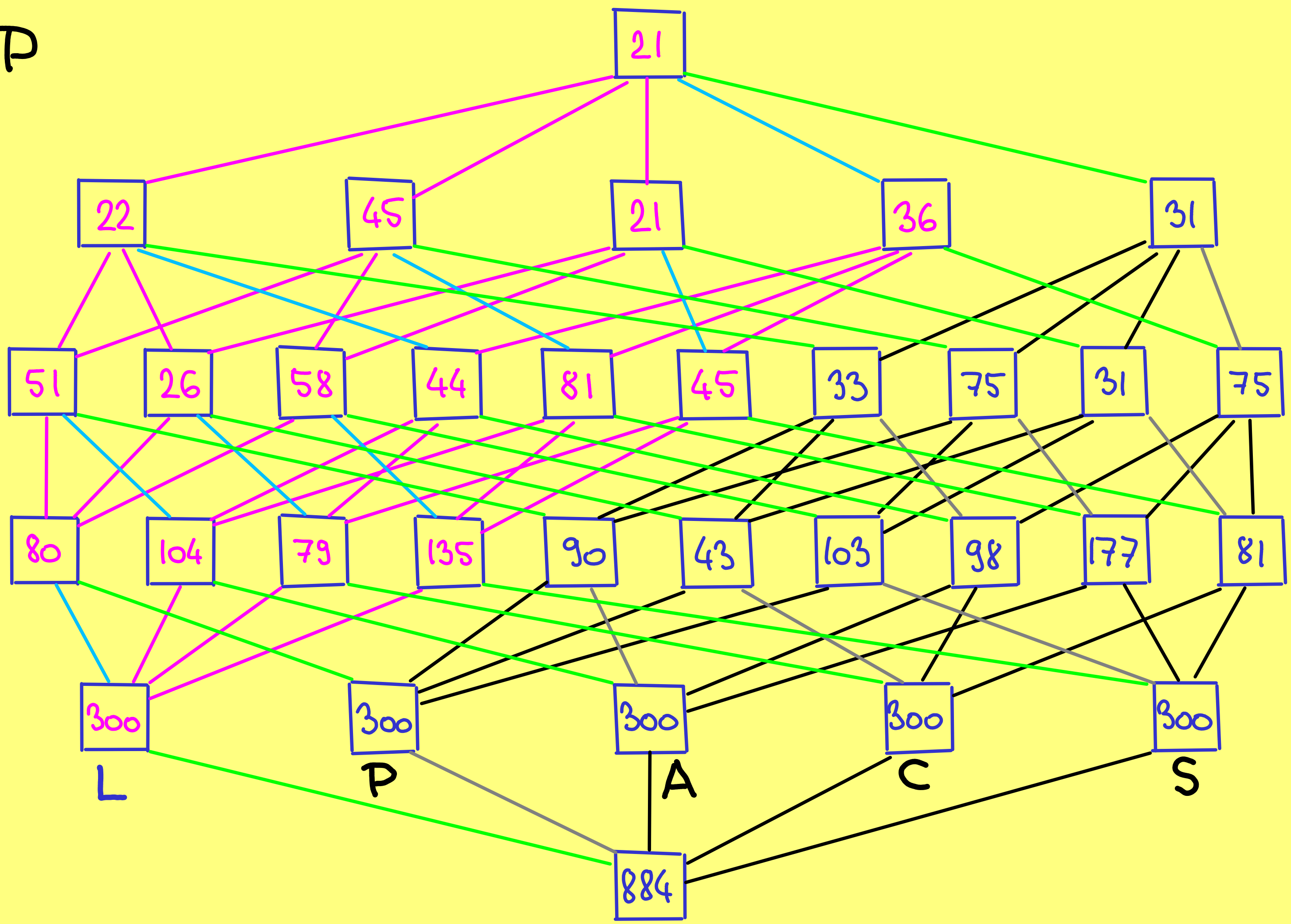
IV.3 External Comparison

Overlap



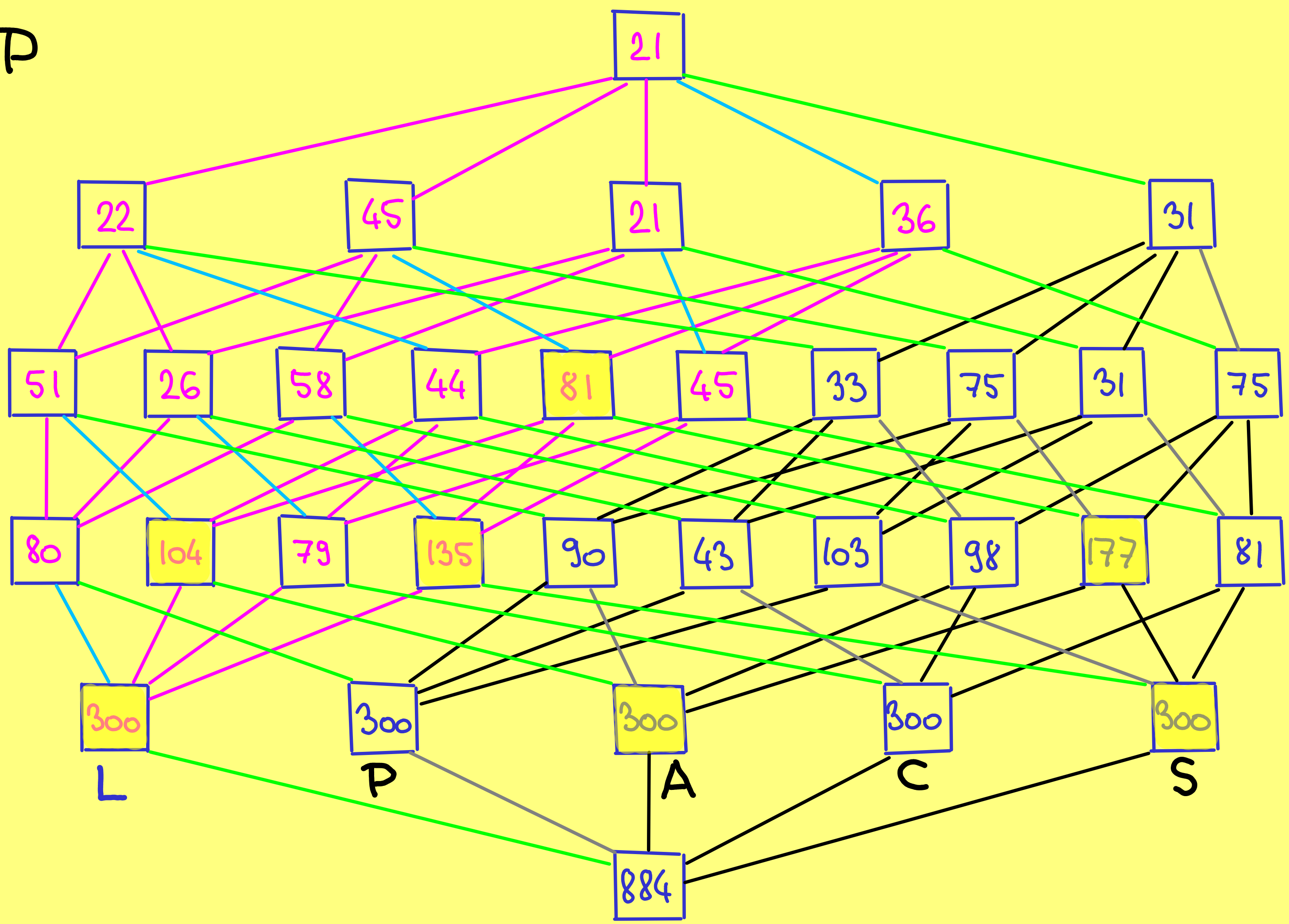
IV.3 External Comparison

Overlap



IV.3 External Comparison

Overlap



IV.3 · External Comparison

WNT Cluster

		L	P	A	C	S
Wnt cyclic genes	validated					
	candidates					

THANK YOU