

Protein bioinformatics: week 12

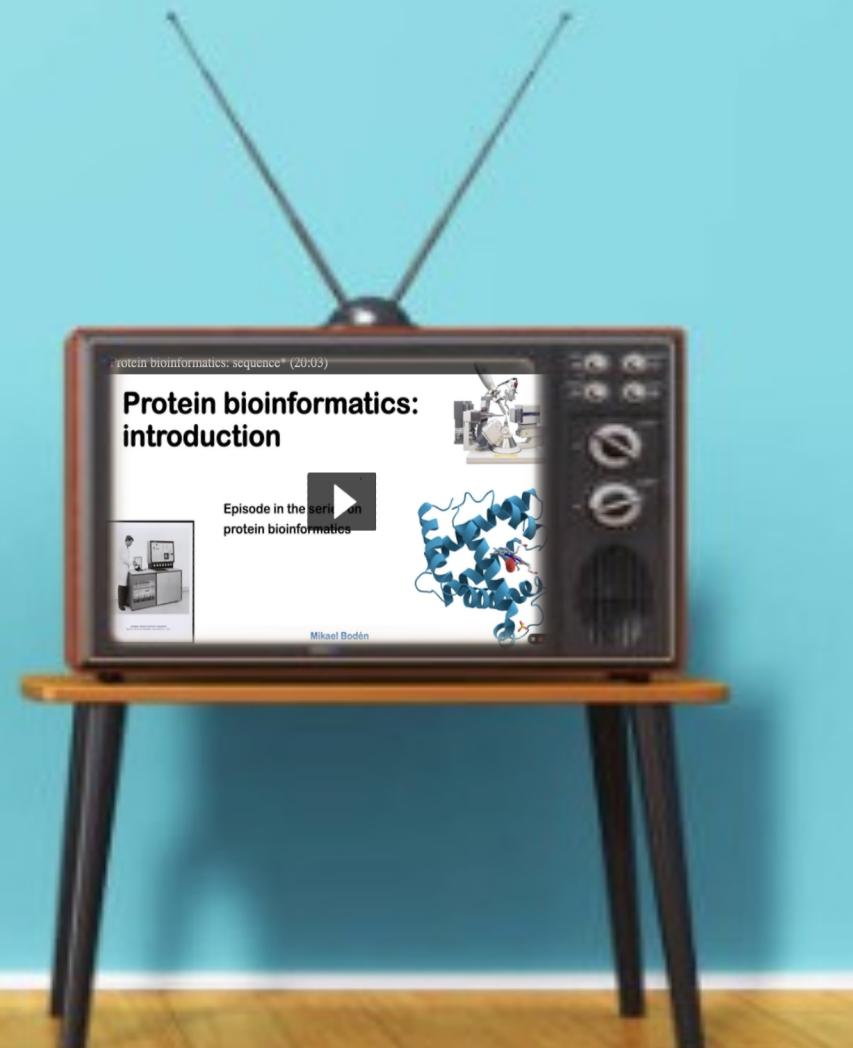
Watch the recordings

Introduction (3 parts)

- Protein sequence, structure, prediction and metrics

Systems (1 part)

- Protein function, localization, interaction, biological networks and data integration



Mikael Bodén

Biological networks

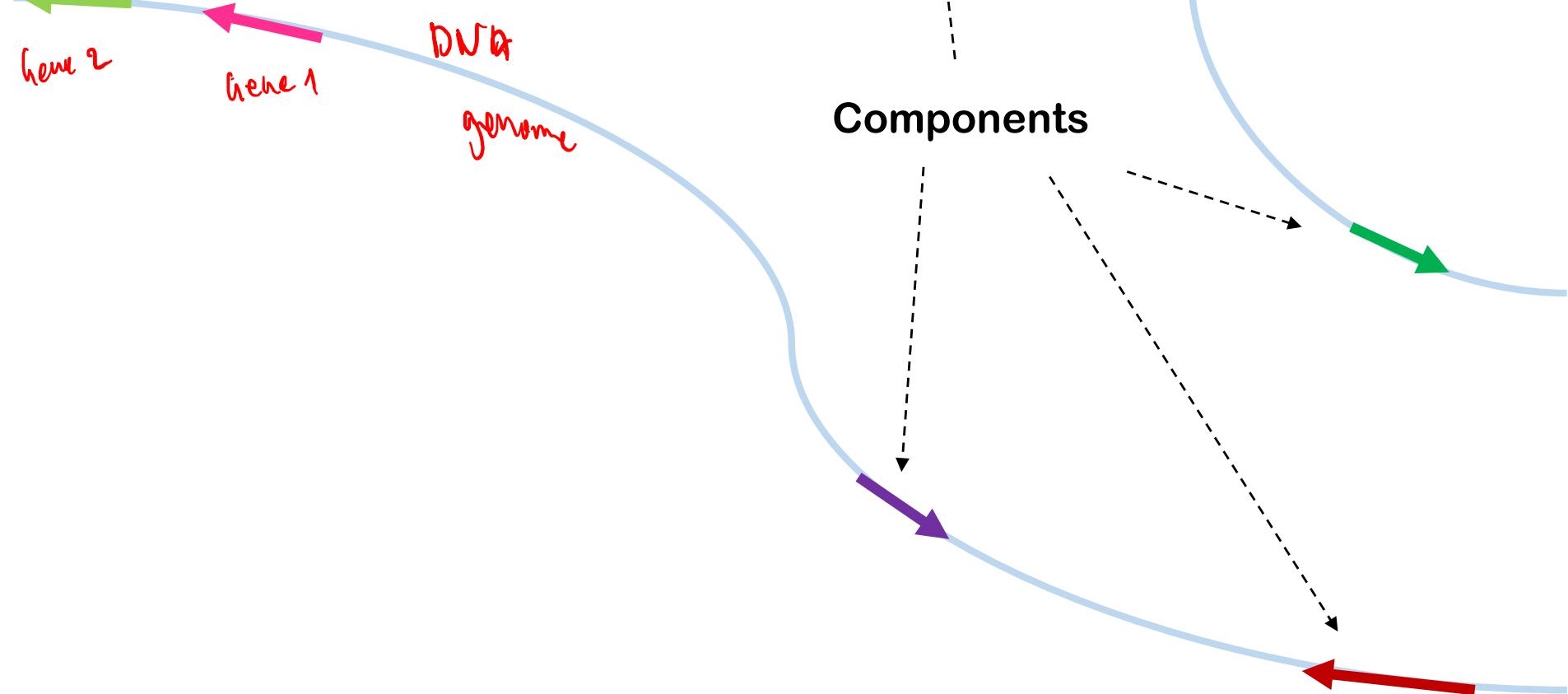
- Genome-wide assays
 - contain a catalog of “system components” (each data set)
 - indicate “relationships” (data sets jointly)
can be used to model “biological systems”
- Biological networks are used to represent
 - physical interactions and complexes
 - protein-protein, protein-DNA, protein-RNA, RNA-DNA, ...
 - permanent and transient interactions
 - cellular pathways (logical interactions)
 - metabolic/enzyme reaction
 - transcriptional/regulatory networks
- Biological networks can
 - integrate data types
 - Expression, co-expression
 - Subcellular localization
 - ...

What is a regulatory network?

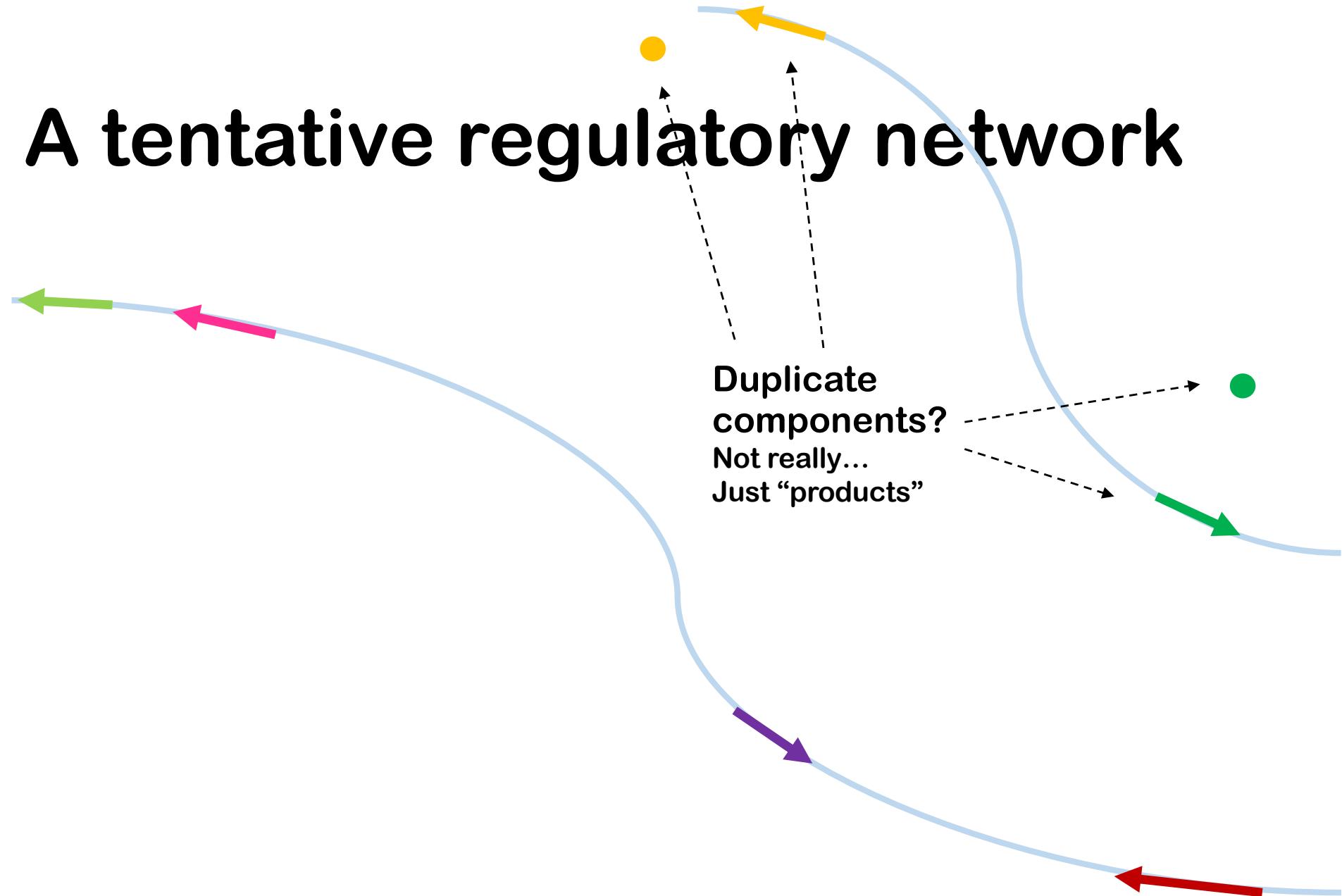
- A. Components are genes, they are linked if they share function
- B. Components are proteins, they are linked if they physically interact
- C. Components are RNA, they are linked if they bind to mRNA
- D. Components are genes, they are linked if the product of gene 1 binds and modulates gene 2's expression
- E. Components are proteins, they are linked if protein 1 is facebook friends with protein 2

Note: multiple options may be true

A tentative regulatory network



A tentative regulatory network

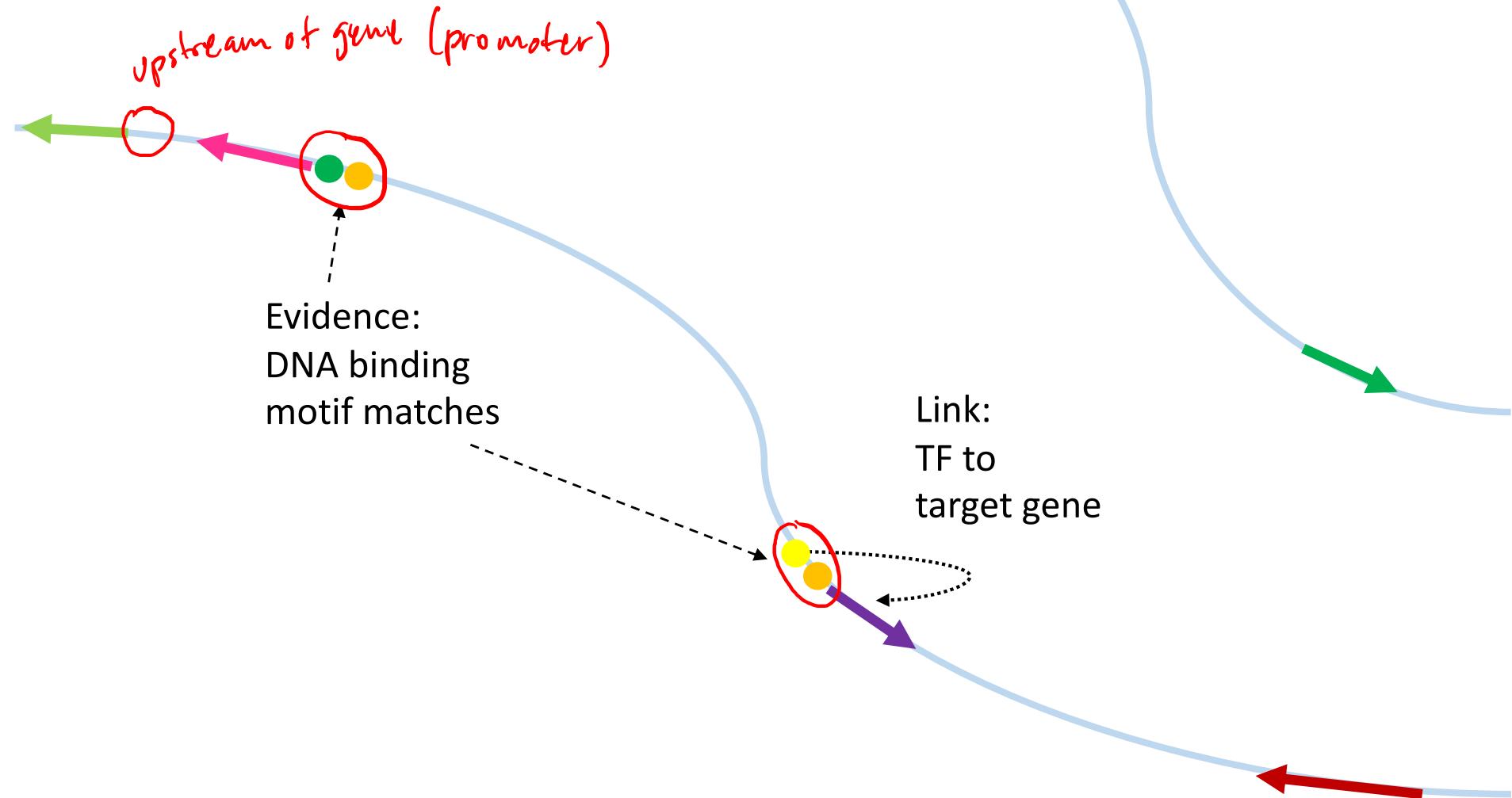


What evidence/data can we use to link the genes?

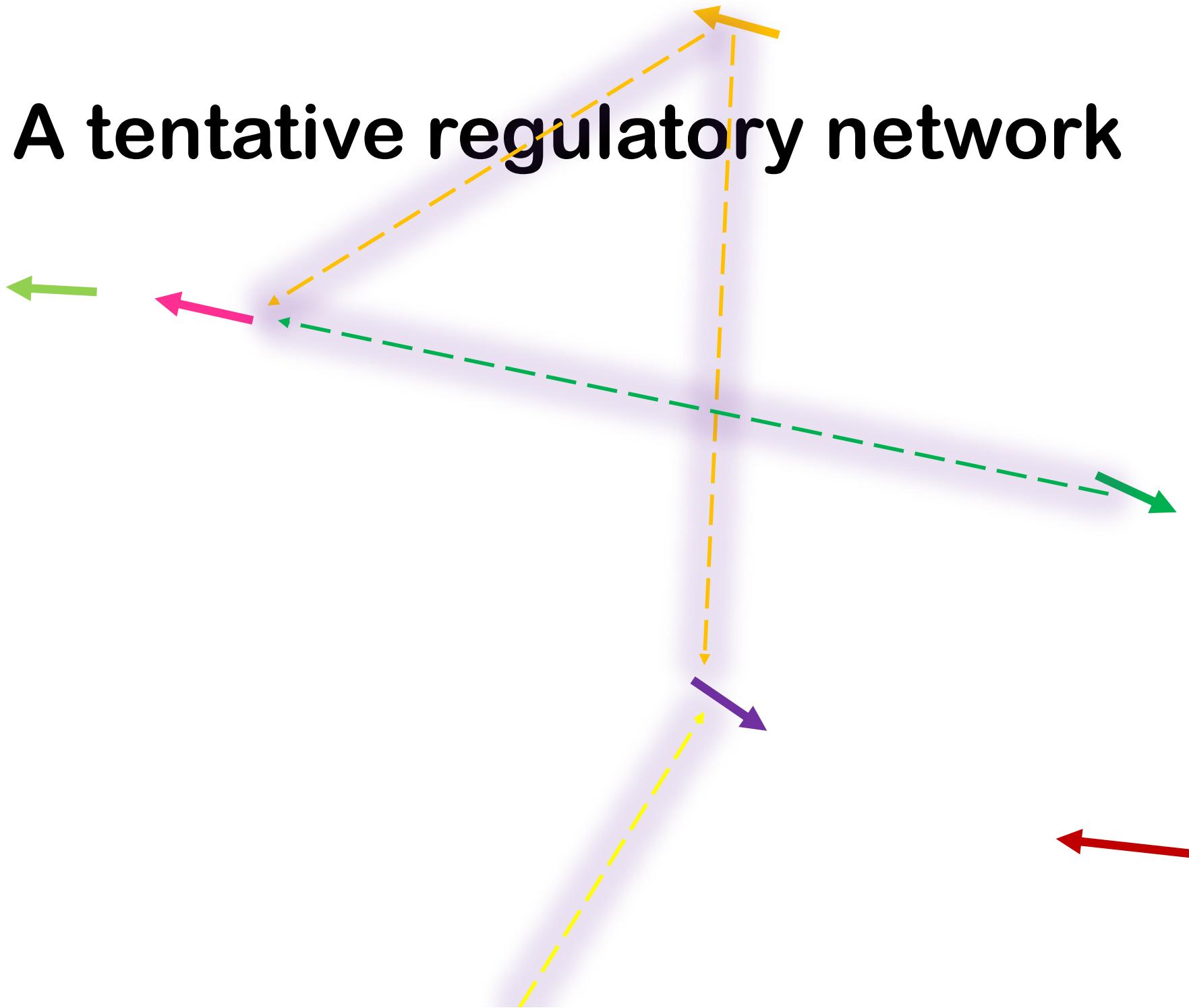
- A. Co-expression of gene
- B. Protein-protein interaction (pairs of gene products)
- C. Co-localization of proteins (pairs of gene products)
- D. Search signalling pathway database for source protein to target gene association
- E. Transcription factor DNA binding motif at target gene promoter

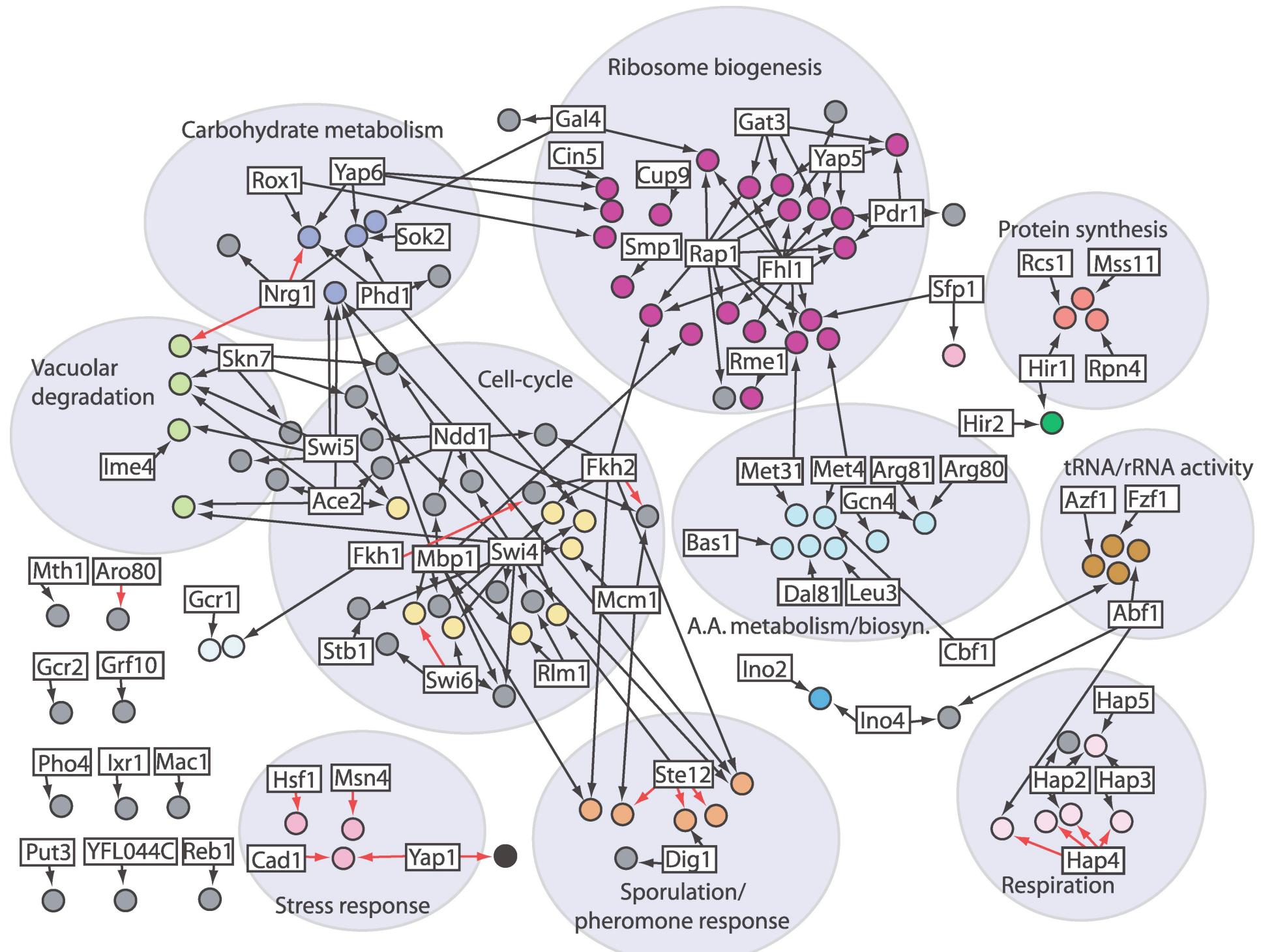
Note: multiple options may be true

A tentative regulatory network



A tentative regulatory network



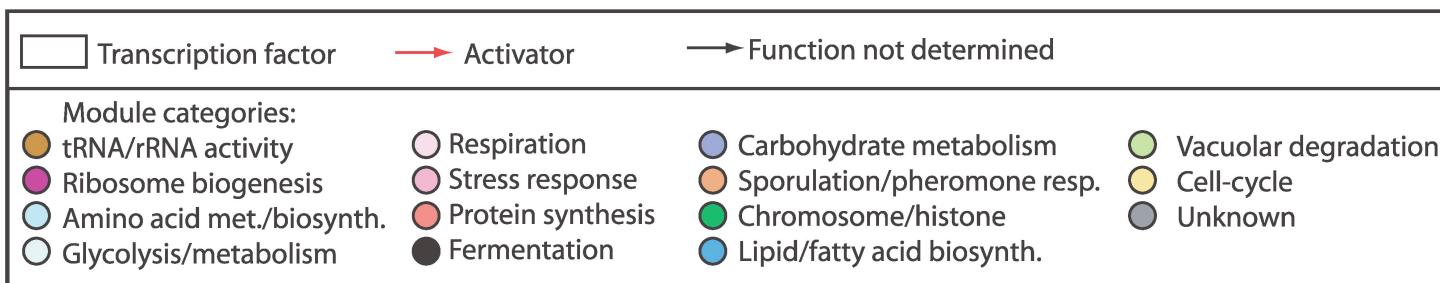
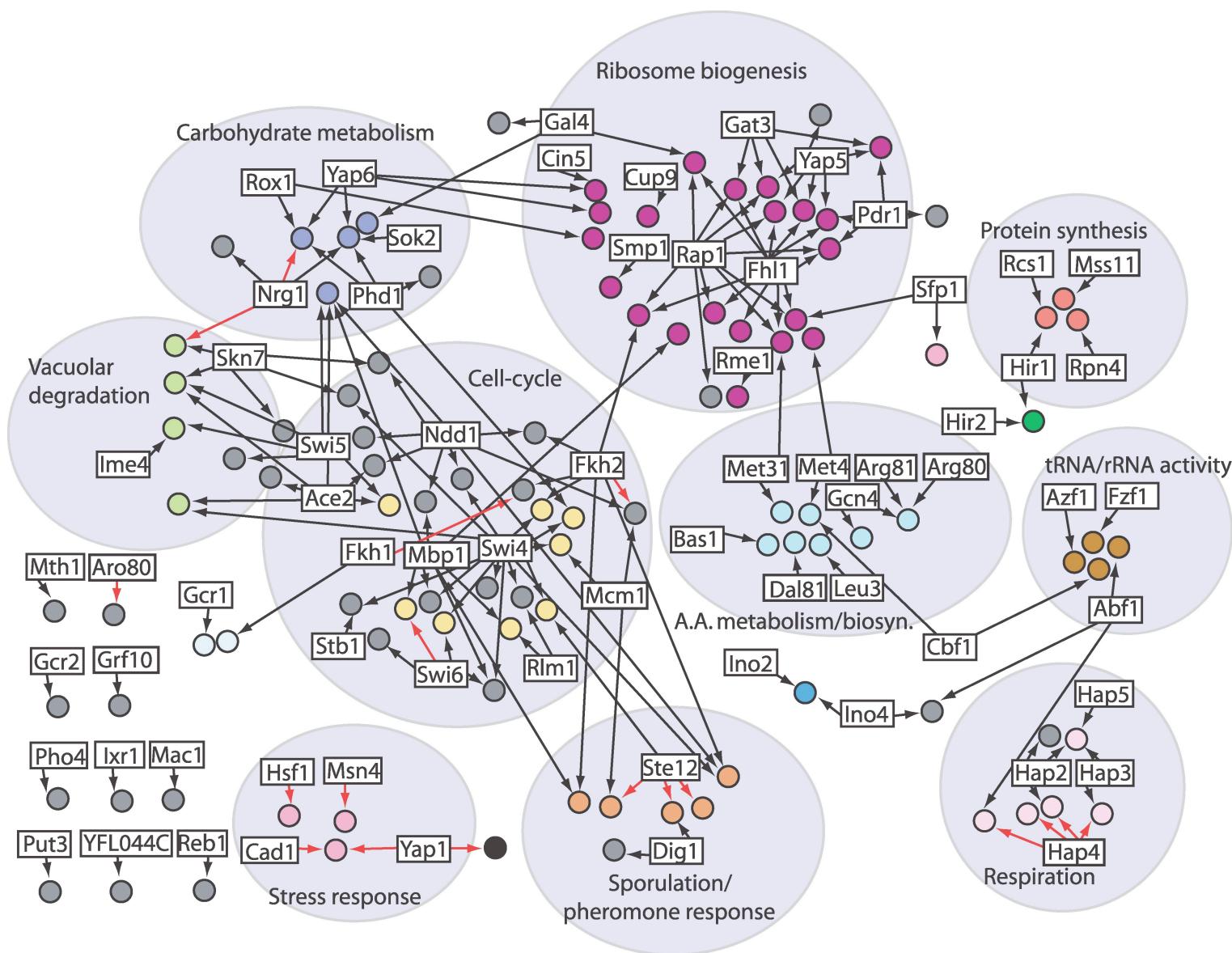


Expression @ condition

	x_1	x_2	...	x_m
G_1	✓	✓	✗	✗
G_2	✗	✗	✓	✗
G_3	✓	✓	✗	✗
G_4				
...			...	
G_n				

Co-expression

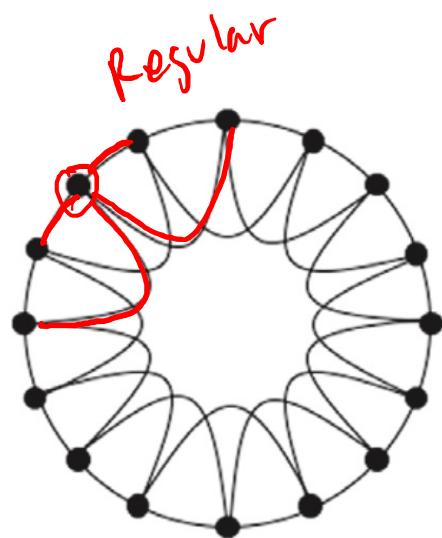
	G_1	G_2	...	G_n
G_1	l			
G_2	-l			
G_3	l			
G_4				
...			...	
G_n				



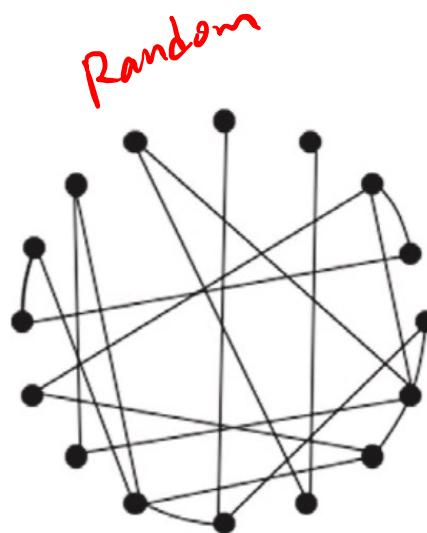
Yeast
TF/DNA
binding
motifs
combined
with
expression
data

- Modular
- Function specific structure
- Robust and redundant
- Predictive

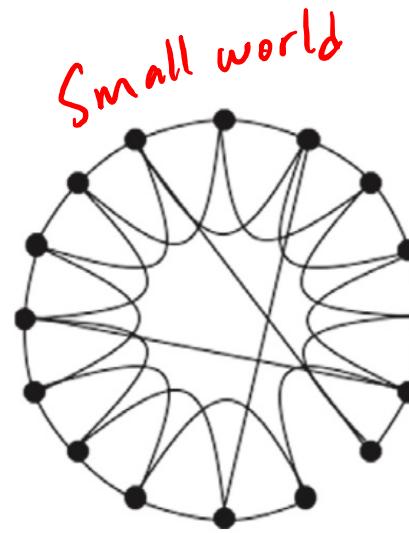
Reproduced from Bar-Joseph et al., *Nat Biotechnol* 21:1337-1342 (2003)



A



B



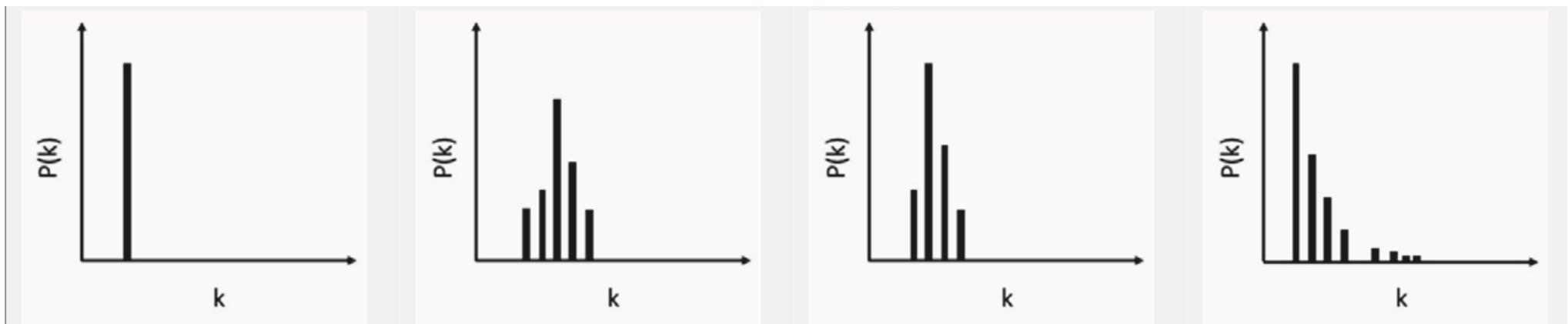
C



D

Which network has a topology of a typical biological network?

What is the degree distribution of a typical biological network?



A

B

C

D

Questions so far?

About biological networks and systems

Final exam 2018: Protein bioinformatics



The Chou-Fasman propensity (P) values for each amino acid to form an α -helix or a β -strand, respectively is shown in the table (right). A larger value denotes higher propensity.

Based on your understanding of the formation of protein secondary structure and the propensity table, predict the secondary structure class (**α -helix**, **β -strand** or **coil**) at the highlighted position of the following amino acid sequences. You do not need to use Chou-Fasman's algorithm, but similar principles should apply in determining your answers.

(a) His – Lys – Glu – Ile – Cys – Leu – Pro
– Ile – Val – Phe – Lys – Asp

...

(d) Justify the predictions for (a)-(c) by explaining what the table is based on and the strategy with which predictions were made. Calculations are not required.

Amino acid	α -helix		β -strand	
	Designation	P	Designation	P
Ala	F	1.42	b	0.83
Cys	I	0.70	f	1.19
Asp	I	1.01	B	0.54
Glu	F	1.51	B	0.37
Phe	f	1.13	f	1.38
Gly	B	0.61	b	0.75
His	f	1.00	f	0.87
Ile	f	1.08	F	1.60
Lys	f	1.16	b	0.74
Leu	F	1.21	f	1.30
Met	F	1.45	f	1.05
Asn	b	0.67	b	0.89
Pro	B	0.57	B	0.55
Gln	f	1.11	f	1.10
Arg	I	0.98	I	0.93
Ser	I	0.77	b	0.75
Thr	I	0.83	f	1.19
Val	f	1.06	F	1.70
Trp	f	1.08	f	1.37
Tyr	B	0.69	F	1.4

(a)

His – Lys – Glu – Ile – Cys – Leu – Pro – Ile – Val – Phe – Lys – Asp

α
 β

0.70 1.21 0.57 1.08 1.06 1.13
1.19 1.30 0.55 1.60 1.70 1.38

Amino acid	α -helix		β -strand	
	Designation	P	Designation	P
Ala	F	1.42	b	0.83
Cys	I	0.70	f	1.19
Asp	I	1.01	B	0.54
Glu	F	1.51	B	0.37
Phe	f	1.13	f	1.38
Gly	B	0.61	b	0.75
His	f	1.00	f	0.87
Ile	f	1.08	F	1.60
Lys	f	1.16	b	0.74

Leu	F	1.21	f	1.30
Met	F	1.45	f	1.05
Asn	b	0.67	b	0.89
Pro	B	0.57	B	0.55
Gln	f	1.11	f	1.10
Arg	I	0.98	I	0.93
Ser	I	0.77	b	0.75
Thr	I	0.83	f	1.19
Val	f	1.06	F	1.70
Trp	f	1.08	f	1.37
Tyr	B	0.69	F	1.4

(a)

His – Lys – Glu – Ile – Cys – Leu – **Pro** – Ile – Val – Phe – Lys – Asp
 1.00 1.16 1.51 1.08 0.70 1.21 0.57 1.08 1.06 1.13 1.16 1.01
 0.87 0.74 0.37 1.60 1.19 1.30 0.55 1.60 1.70 1.38 0.74 0.54

Amino acid	α -helix		β -strand	
	Designation	P	Designation	P
Ala	F	1.42	b	0.83
Cys	I	0.70	f	1.19
Asp	I	1.01	B	0.54
Glu	F	1.51	B	0.37
Phe	f	1.13	f	1.38
Gly	B	0.61	b	0.75
His	f	1.00	f	0.87
Ile	f	1.08	F	1.60
Lys	f	1.16	b	0.74

Leu	F	1.21	f	1.30
Met	F	1.45	f	1.05
Asn	b	0.67	b	0.89
Pro	B	0.57	B	0.55
Gln	f	1.11	f	1.10
Arg	I	0.98	I	0.93
Ser	I	0.77	b	0.75
Thr	I	0.83	f	1.19
Val	f	1.06	F	1.70
Trp	f	1.08	f	1.37
Tyr	B	0.69	F	1.4

(b)

Arg – Pro – Met – Ala – Lys – Thr – Gln – Ala – Phe – Cys – Gly

0.98 0.57 1.45 1.42 1.16 0.83 1.11 1.42 1.13 0.70 0.61

0.93 0.55 1.05 0.83 0.74 1.19 1.10 0.83 1.38 1.19 0.75

Amino acid	α-helix		β-strand	
	Designation	P	Designation	P
Ala	F	1.42	b	0.83
Cys	I	0.70	f	1.19
Asp	I	1.01	B	0.54
Glu	F	1.51	B	0.37
Phe	f	1.13	f	1.38
Gly	B	0.61	b	0.75
His	f	1.00	f	0.87
Ile	f	1.08	F	1.60
Lys	f	1.16	b	0.74

Leu	F	1.21	f	1.30
Met	F	1.45	f	1.05
Asn	b	0.67	b	0.89
Pro	B	0.57	B	0.55
Gln	f	1.11	f	1.10
Arg	I	0.98	I	0.93
Ser	I	0.77	b	0.75
Thr	I	0.83	f	1.19
Val	f	1.06	F	1.70
Trp	f	1.08	f	1.37
Tyr	B	0.69	F	1.4

(c)

Pro – Gly – Cys – His – Pro – Ser – Tyr – Ala

0.57 0.61 0.70 1.00 0.57 0.77 0.69 1.42

0.55 0.75 1.19 0.87 0.55 0.75 1.40 0.83

Amino acid	α-helix		β-strand	
	Designation	P	Designation	P
Ala	F	1.42	b	0.83
Cys	I	0.70	f	1.19
Asp	I	1.01	B	0.54
Glu	F	1.51	B	0.37
Phe	f	1.13	f	1.38
Gly	B	0.61	b	0.75
His	f	1.00	f	0.87
Ile	f	1.08	F	1.60
Lys	f	1.16	b	0.74

Leu	F	1.21	f	1.30
Met	F	1.45	f	1.05
Asn	b	0.67	b	0.89
Pro	B	0.57	B	0.55
Gln	f	1.11	f	1.10
Arg	I	0.98	I	0.93
Ser	I	0.77	b	0.75
Thr	I	0.83	f	1.19
Val	f	1.06	F	1.70
Trp	f	1.08	f	1.37
Tyr	B	0.69	F	1.4

The correct answers were (a) beta, (b) alpha, (c) coil

What *sensitivity* to the class alpha helix do I get by predicting *alpha helix* for all three questions?

- (a) observing not α predicting α FP
- (b) observing α predicting α TP
- (c) observing not α predicting α FP

- Sensitivity
$$\frac{tp}{tp + fn}$$
- Specificity
$$\frac{tn}{tn + fp}$$

↓
Two decimal places multiplied by 100

e.g. 100 50 33 0

correct

The correct answers were (a) *beta*, (b) *alpha*, (c) *coil*

What *specificity* to the class *alpha helix* do I get by predicting *alpha helix* for all three questions?

(a) observing not α
predicting α

FP

(b) observing α
predicting α

TP

(c) observing not α
predicting α

FP

• Sensitivity

$$\frac{tp}{tp + fn}$$

• Specificity

$$\frac{tn}{tn + fp} \quad \frac{0}{0+2}$$

multiply by 100

e.g.

100 50 33 0

The correct answers were (a) *beta*, (b) *alpha*, (c) *coil*

What *accuracy* (Q_3) do I get by predicting *alpha helix* for all three questions?

	Helix prediction	Strand prediction	Coil prediction
Helix observation	1	0	0
Strand observation	1	0	0
Coil observation	1	0	0

$$Q_k = \frac{\sum_{j=1}^k tp(j)}{\sum_{j=1}^k tp(j) + fn(j)} \cdot 100$$

0.33 $\frac{1}{3}$

Questions so far?

About prediction, protein sequence and structure