

# Pulling it all together

From protein sequences  
to PPI networks and beyond

Tamás Korcsmáros

**NetBio**  
Network Biology Group  
<http://netbiol.elte.hu>

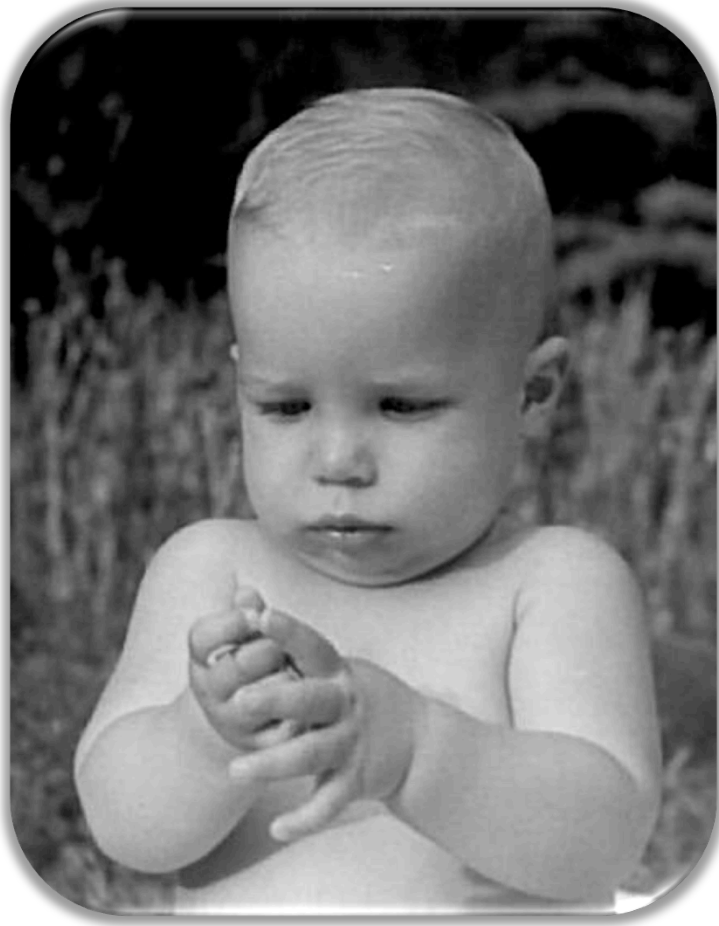


**TGAC**  
The Genome Analysis Centre™



**IFR** Institute of  
Food Research

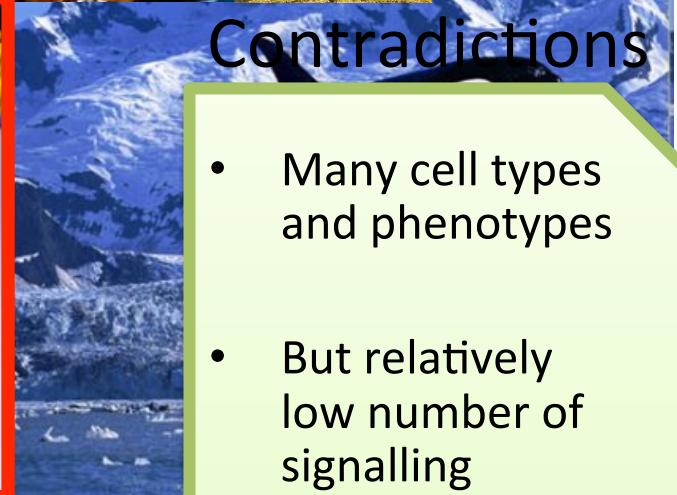
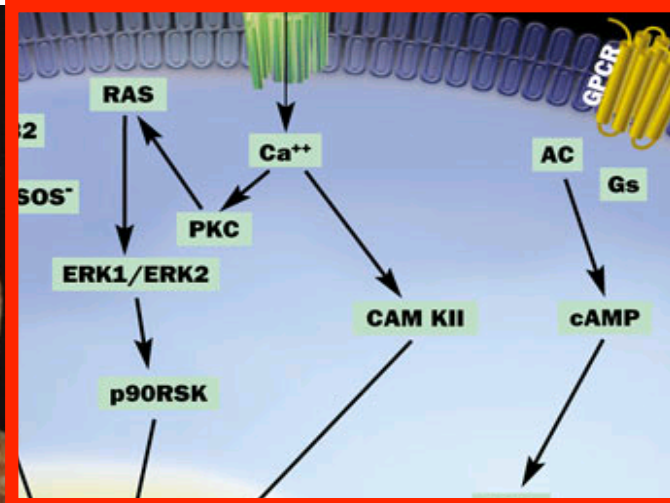
# Early research career



- Started as a high-school research student
- 4 years at a medical biochemistry lab
- MSc in biochemistry and molecular biology
- PhD on metazoan signalling networks

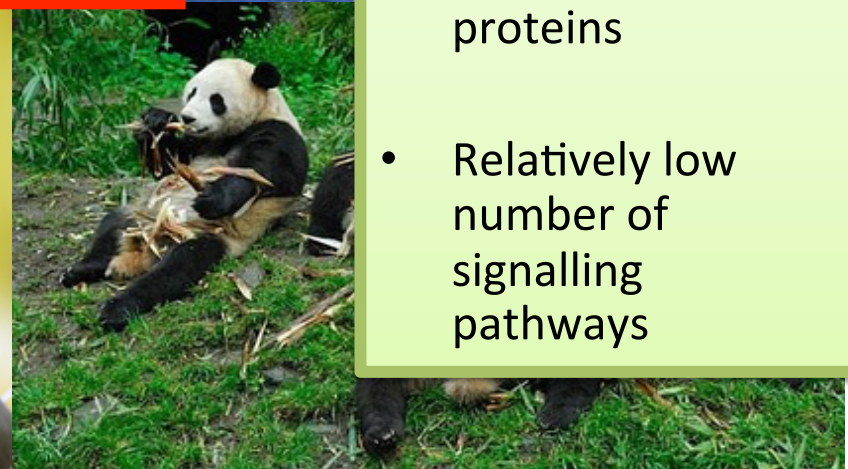


# Why I am interested in PPIs?



## Contradictions

- Many cell types and phenotypes
- But relatively low number of signalling proteins
- Relatively low number of signalling pathways



# Thematics

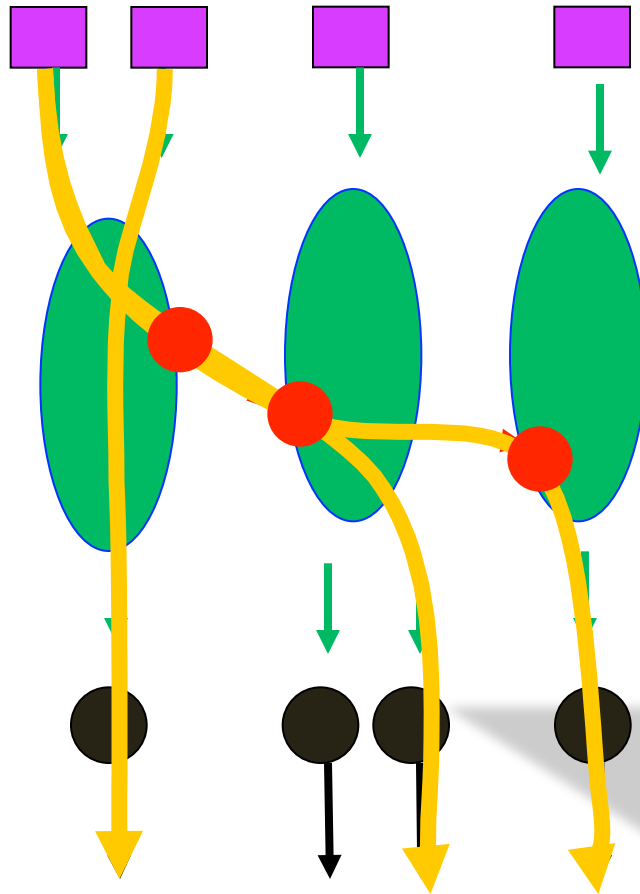
- **From pathways to networks**
  - Visualization challenges and solutions
- **Extending pathways**
  - Data integration
  - Power of prediction
- **Use cases for data integration**
  - NRF2ome
  - SignaLink 2

# Signalling pathways and networks

Input  
Ligands

Pathway  
mediators  
(**cross-talk  
proteins**)

Output  
Transcription  
factors



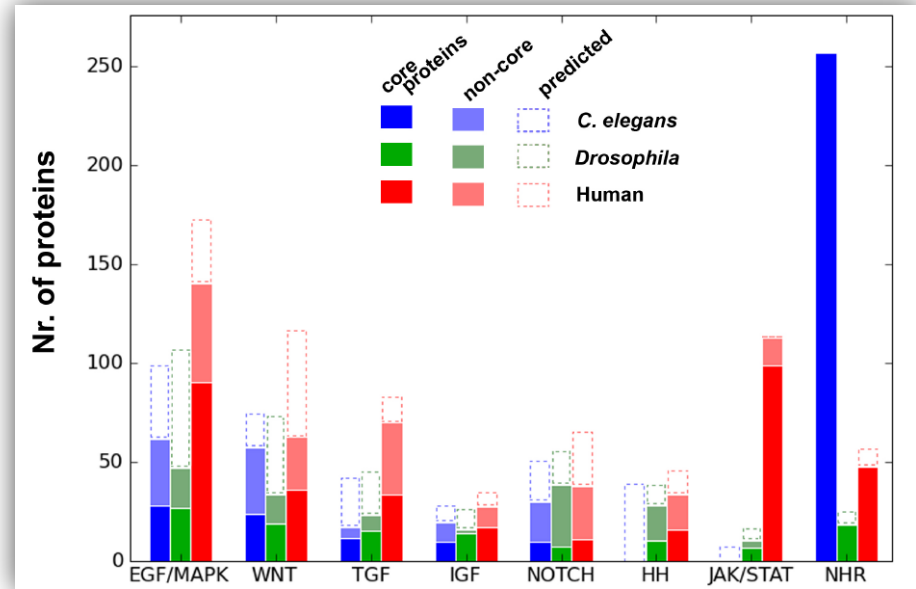
## Problems and challenges in signalling pathway databases

- Pathway definitions are not identical
- Curation rules are not uniform
- Not enough and identical details / sources
- Cross-talks and overlaps (multi-pathway proteins) cannot be examined (easily)

# Case study I. - the SignaLink database

<http://SignaLink.org>

- 3 metazoans
- 8 biochemically defined pathways
- Manually curated, uniform curation rules
- All interactions are directed and link to experimental research articles
- Contains multi-pathway proteins and cross-talks
- Versions: 2006, 2008, 2012\*



## Compared to 3 pathway databases

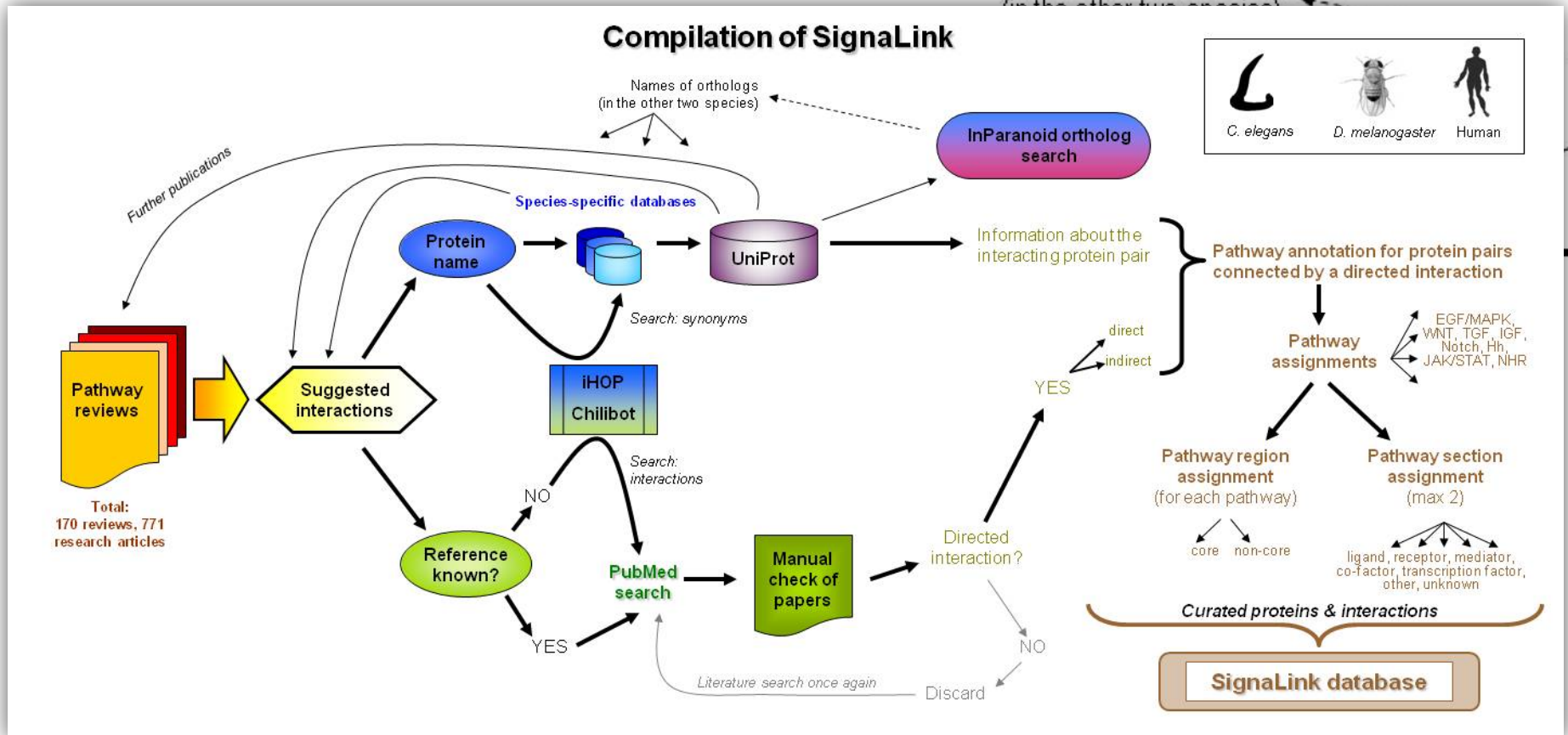
SignaLink contains the highest number of proteins, interactions, cross-talks and literature sources (between the examined pathways)

Allows the system-level examination of the signalling network



# Curation protocol of SignaLink

## Compilation of SignaLink



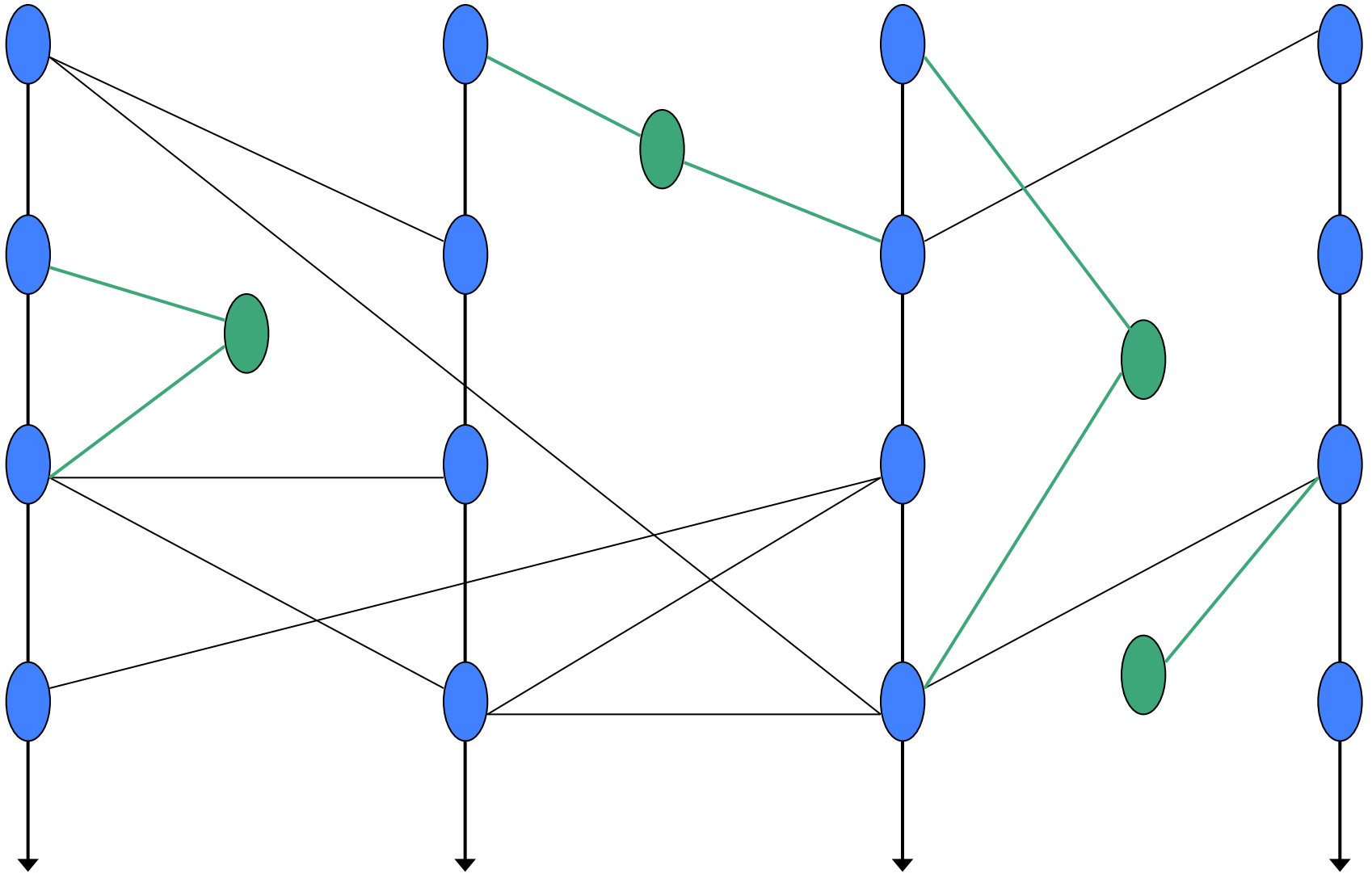
170 reviews, 771 research articles

Reference known?

PubMed search

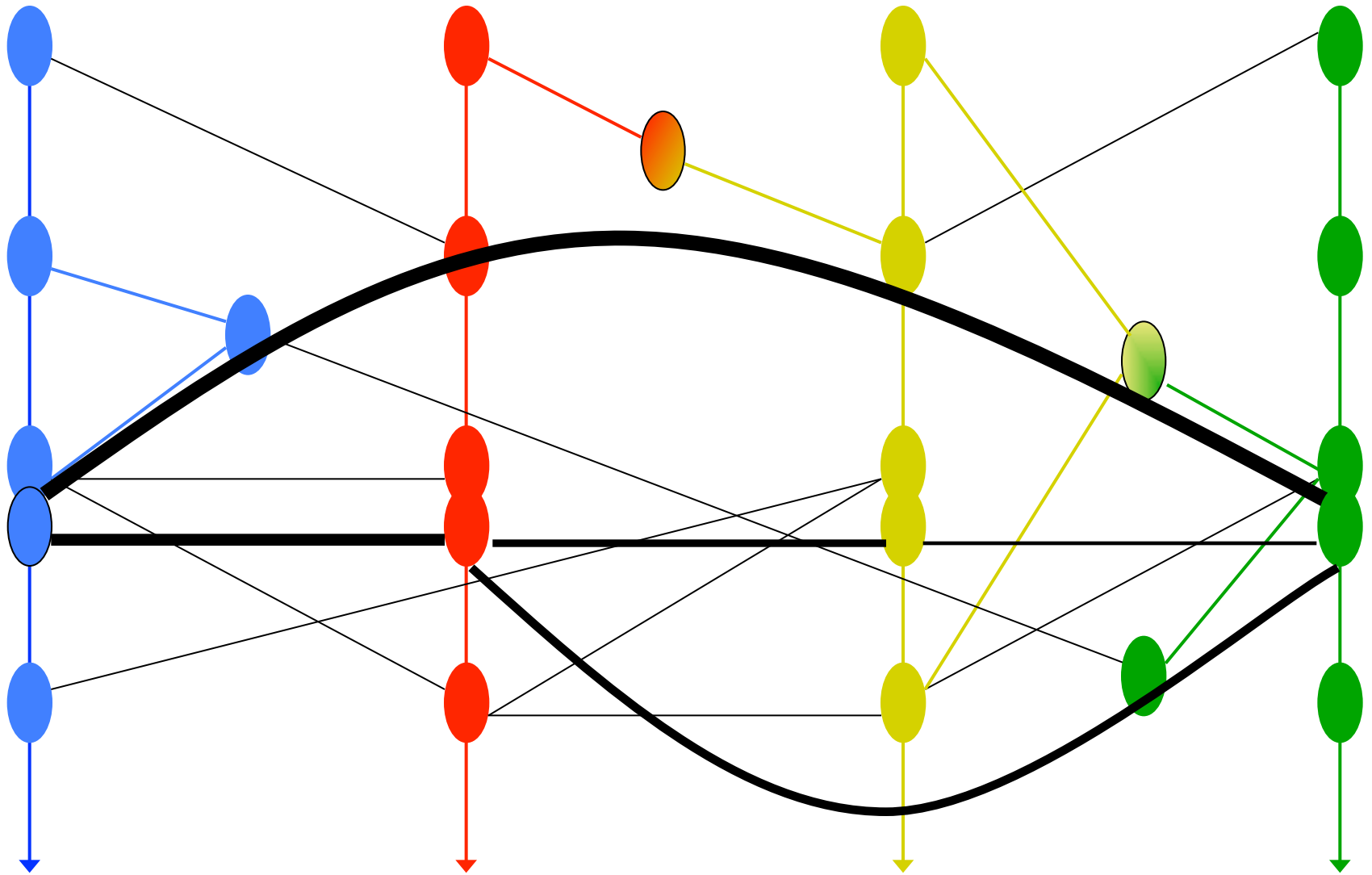
Manual check of papers

# Network of signalling pathways



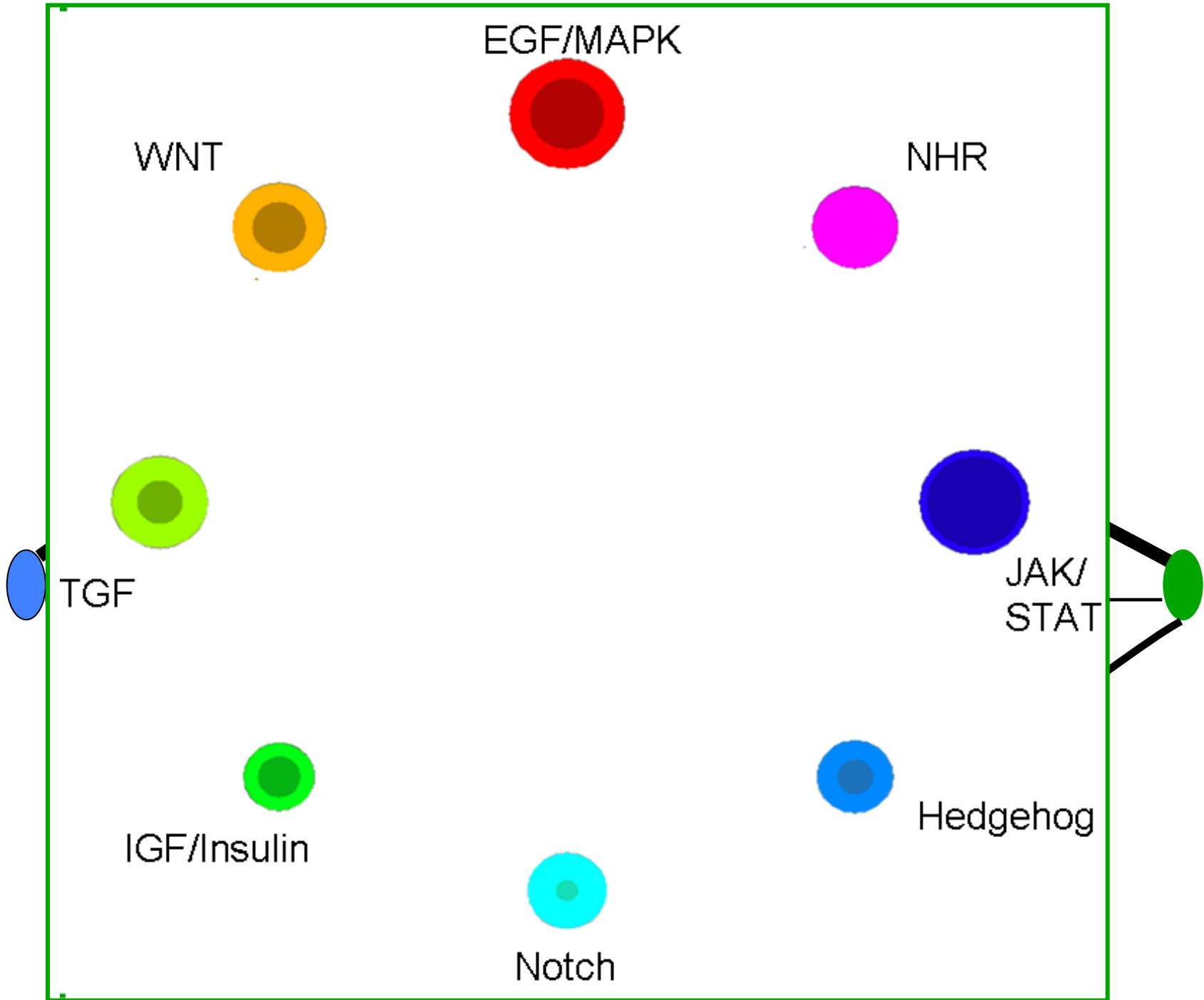
How can we visualize cross-talks?

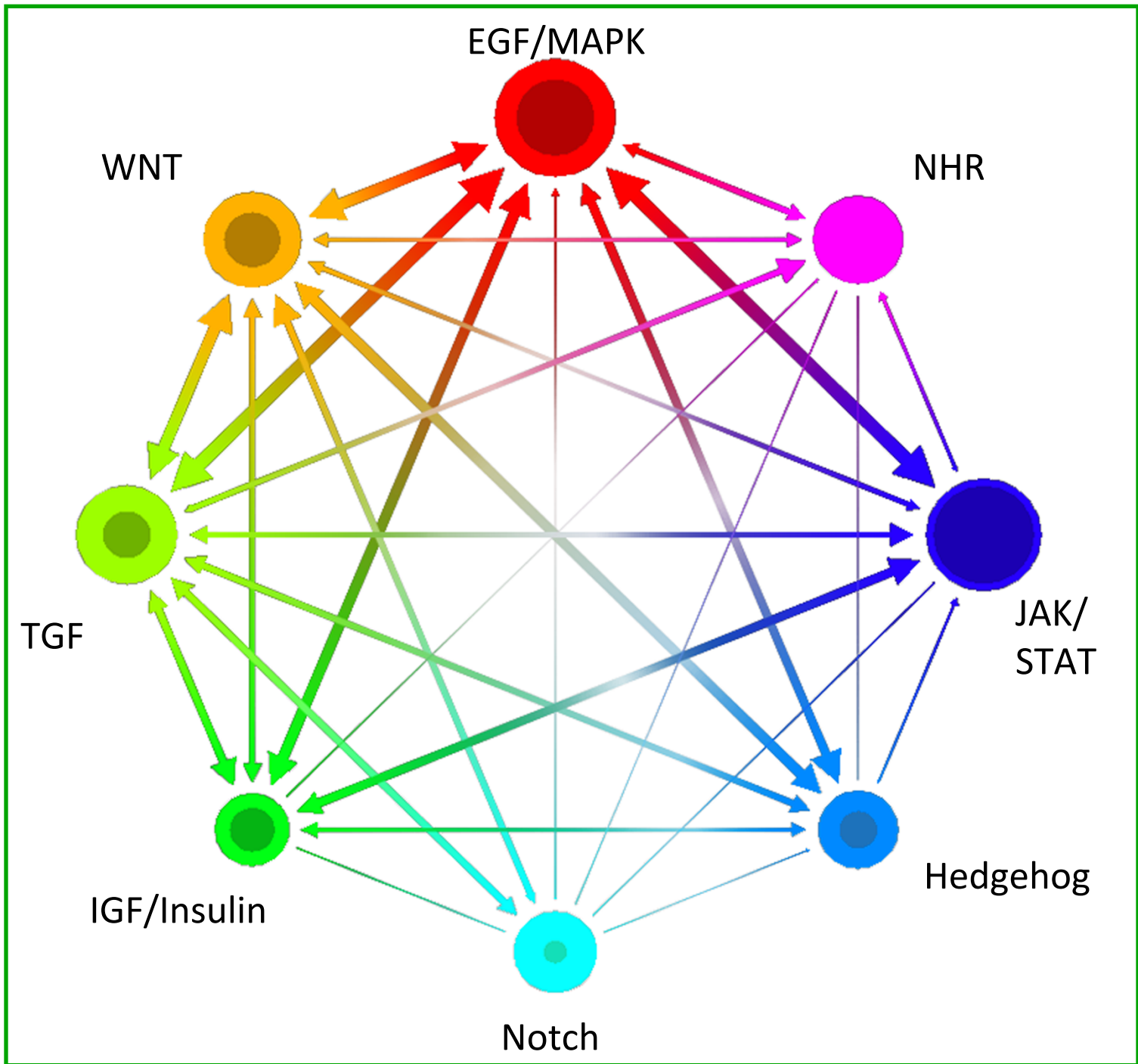
# Network of signalling pathways



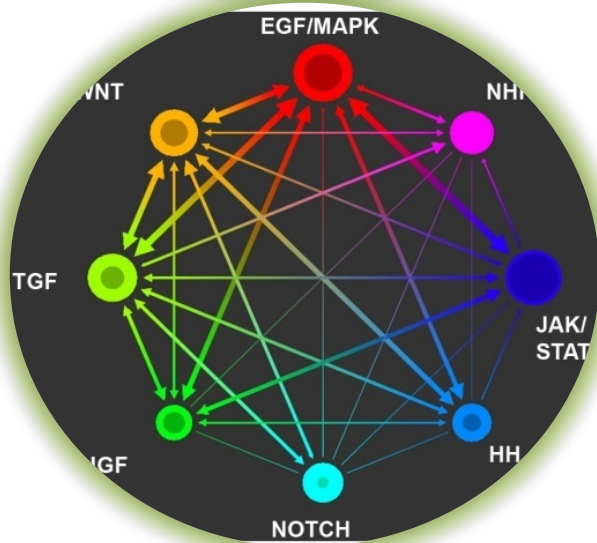
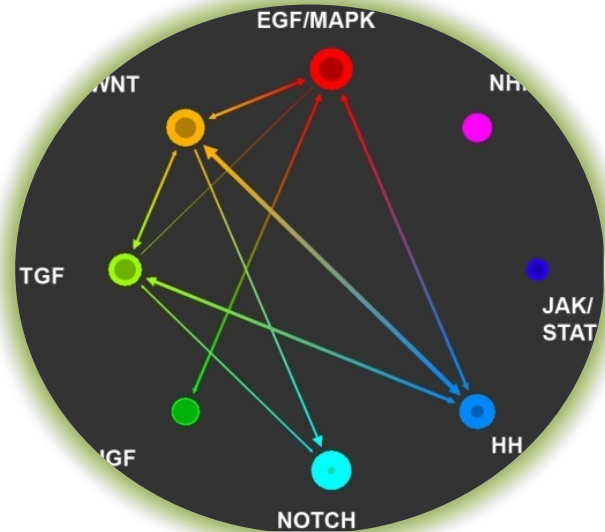
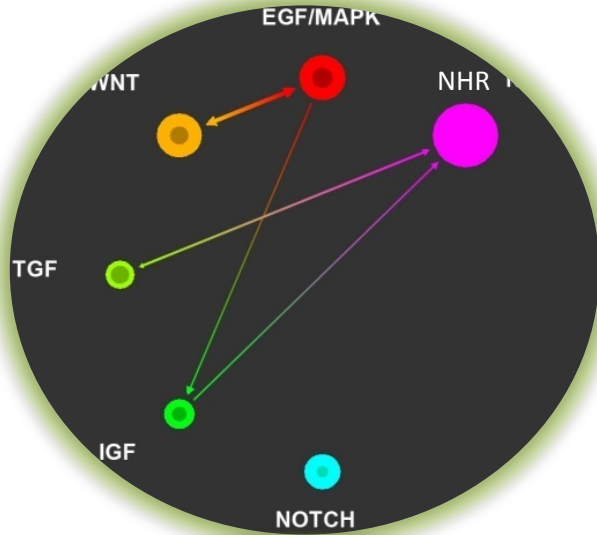
**1) Color-coded protein-protein interaction network**











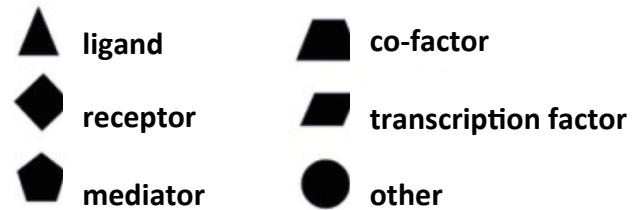
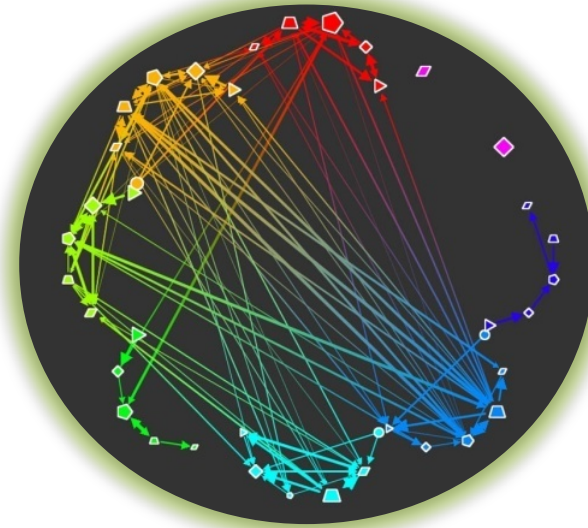


# Cross-talk networks in 3 metazoans



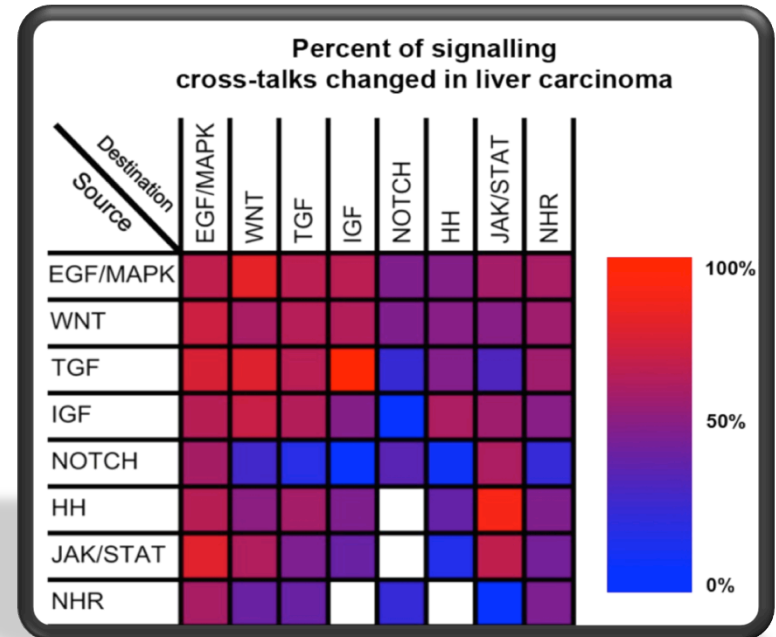
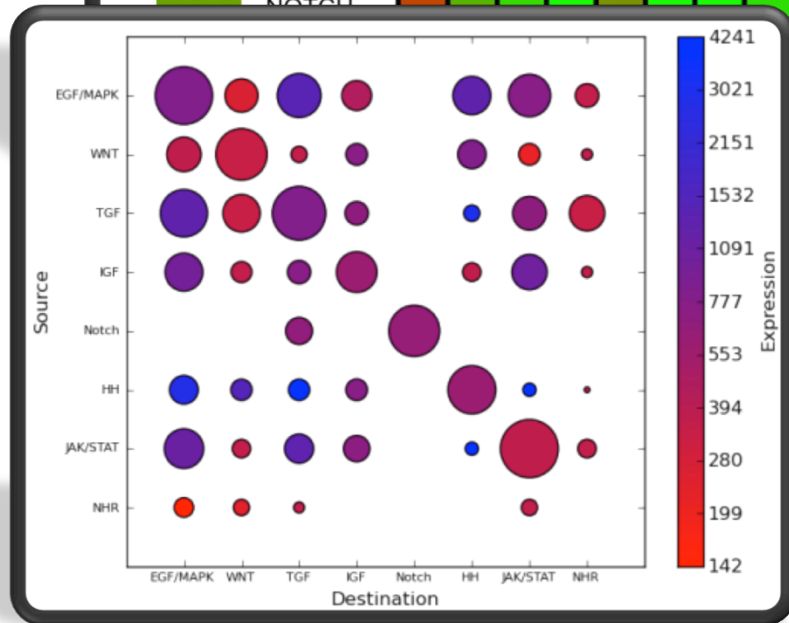
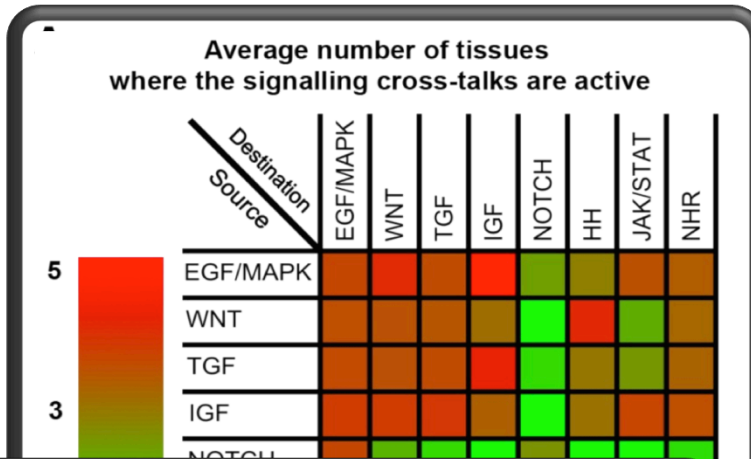
Nr. of proteins			
	50	150	250
Nr. of cross-talks			
	1	10	100

# Detailed cross-talk networks in 3 metazoans



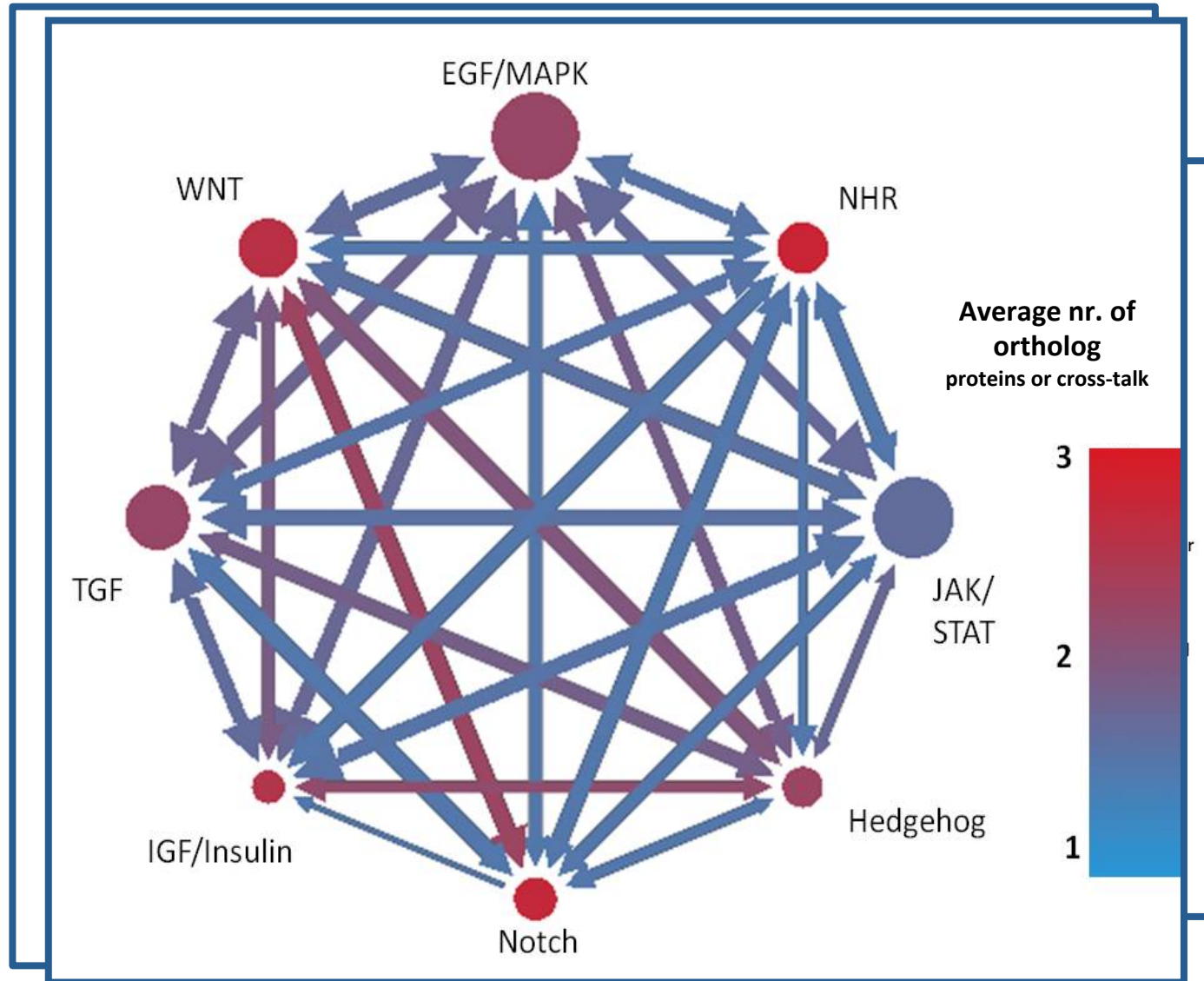
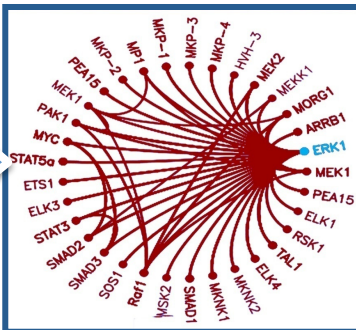
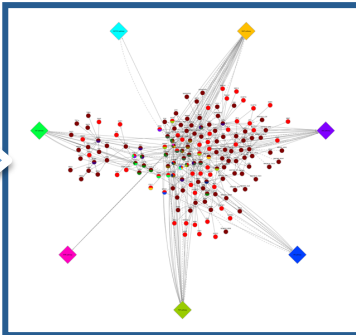
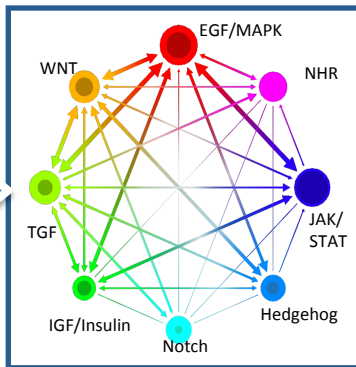
# Visualizing cross-talk expressions

Identification of tissue-, disease-, and cell-type specific cross-talks

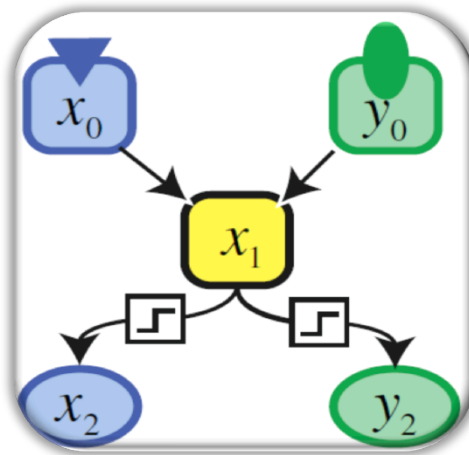




# Further cross-talk analyses and visualizations



# Multi-pathway proteins and the direction of signalling flow



## A solution for the contradictions?

- Many cell types and phenotypes
- But relatively low number of signalling proteins
- Relatively low number of signalling pathways

## Combinatorial problems

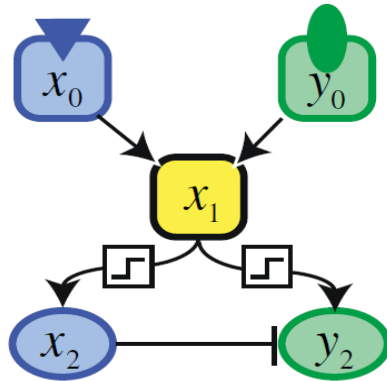
- Protein isoforms
- Alternative splicing
- Post-translational sites

How cross-talks and signalling flow are regulated?

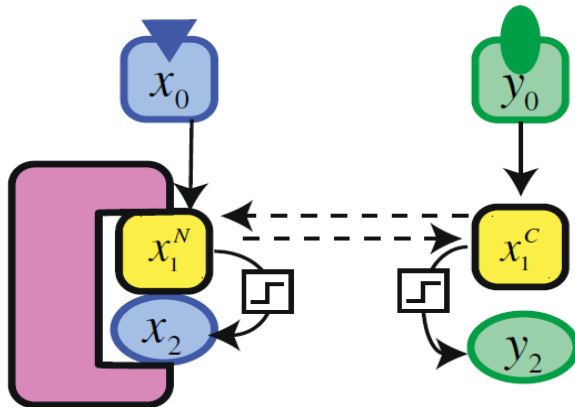
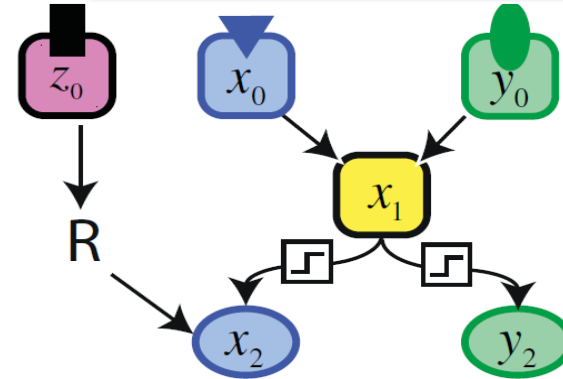


# Insulating mechanisms to regulate signalling flow

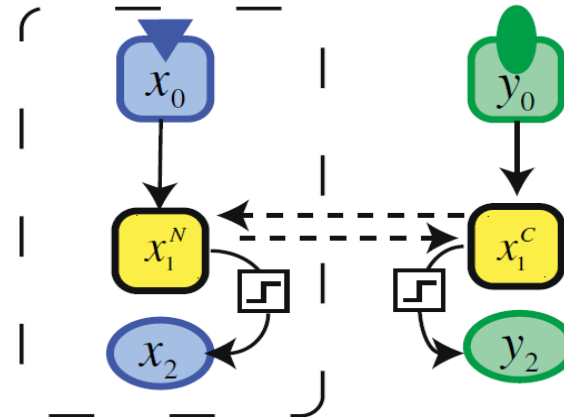
Cross-pathway inhibition



Combination of signalling



Scaffolds



Compartmentalization

Precise regulation of  
signalling pathways is important.

**Extending signalling pathways**

# Regulation of signalling duration

## Kinases, phosphatases, ubiquitin-ligases, peptidases, etc.

- Reversible or irreversible modulation of specific proteins
- Priming, activation, temporal de-activation, cleavage, destruction, etc.

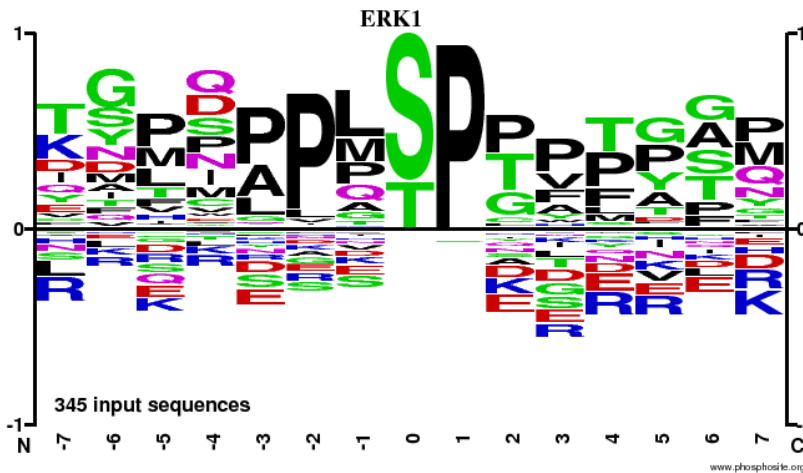
## Resources of post-translational modifications (PTMs)

- NetworkKIN (<http://networkin.info>)
- Phosphosite (<http://phosphosite.org>)
- dbPTM (<http://dbptm.mbc.nctu.edu.tw>)
- ELM server (<http://elm.eu.org>)
- ...



## Specificity?

### Target motifs



### Docking motifs

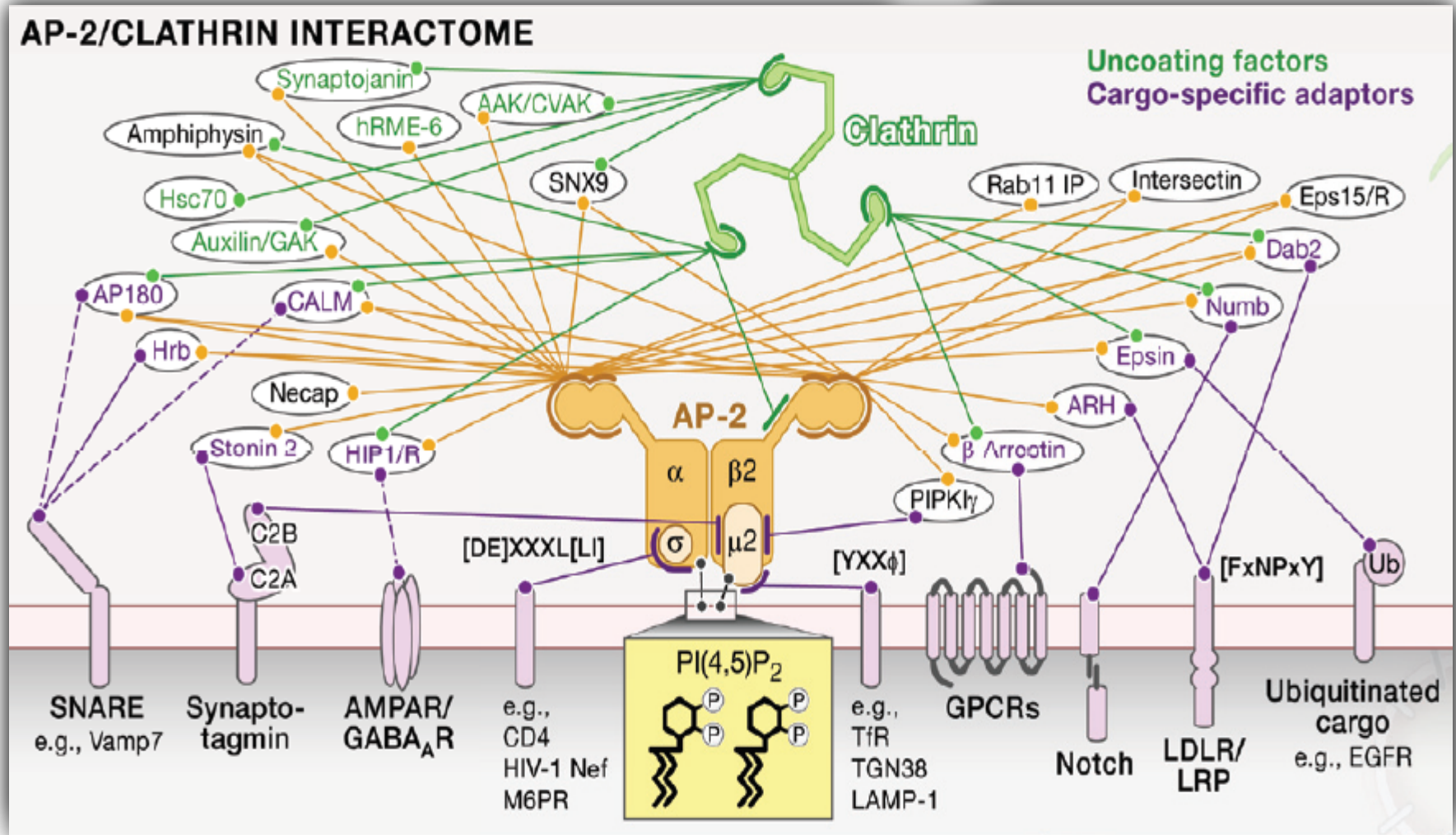
Sequence	Start	End	Subsequence
NCF1_HUMAN	363	368	TQRSKPQ <b>PAVFP</b> PSADLIL
POLG_HCVJA	2323	2328	LPSTKAP <b>PIPP</b> PRKRRTVVL
FAK1_MOUSE	750	755	SGGSDEA <b>PKP</b> SRPGYPSPR
DYN1_HUMAN	833	838	FGFEPQV <b>ESR</b> ENAPPGVFS
P85A_HUMAN	305	313	RQEA <b>P</b> PKPKPTTVANN
P85A_HUMAN	305	310	WNERQPA <b>PALFP</b> PKPKPTTV
RPGF1_HUMAN	284	289	VVDNS <b>FPALFP</b> PKRQSAPS
PTN22_MOUSE	614	619	RTDDEI <b>PPLP</b> ERTFESFIV
NEF_HV1BR	72	77	EVGF <b>PVTPQVPL</b> RPMTYKAA
PAK1_RAT	13	18	LDVQDK <b>PAPF</b> PMRNTSTMIG
SOS1_HUMAN	1152	1157	DEVF <b>VFPFVPPFR</b> RFESAPA

P...P.[KR]

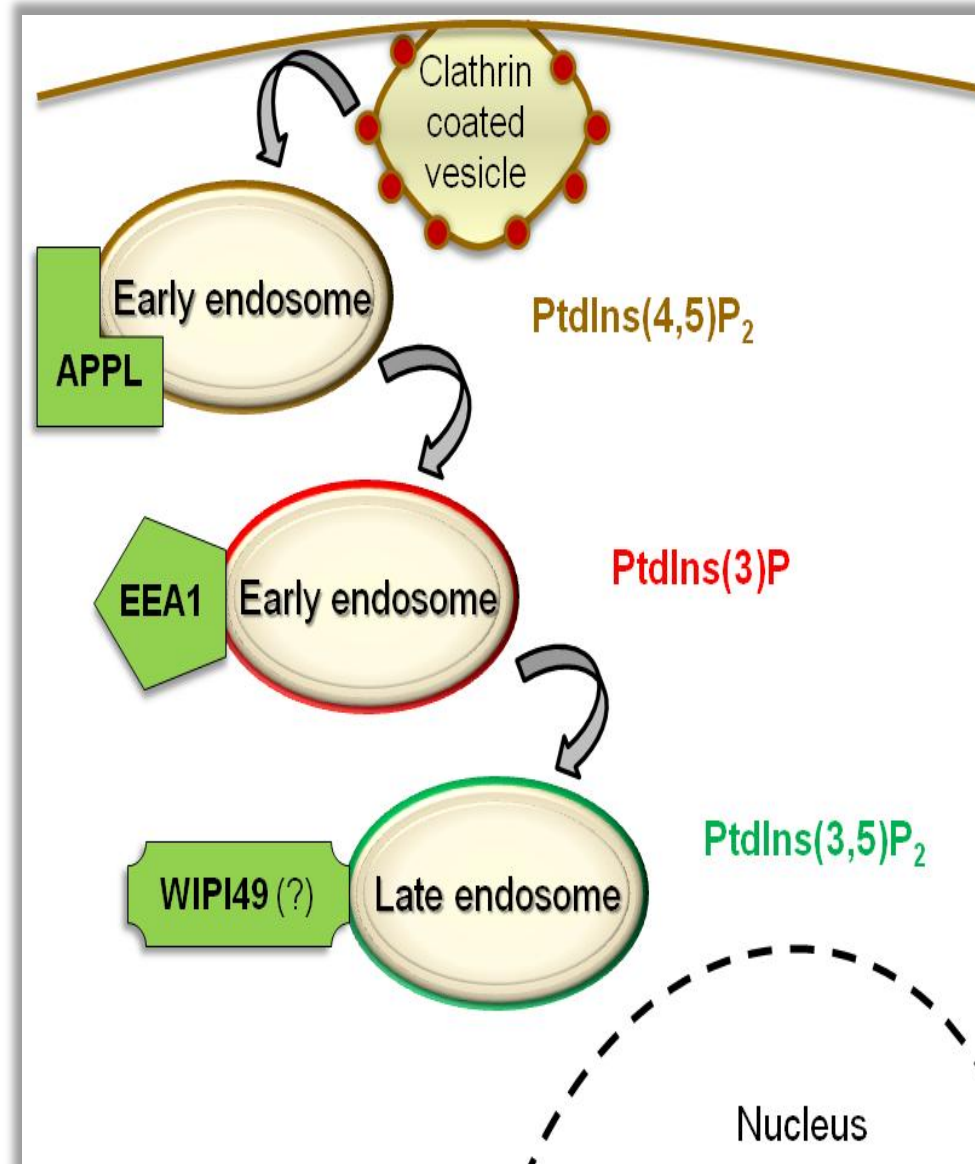
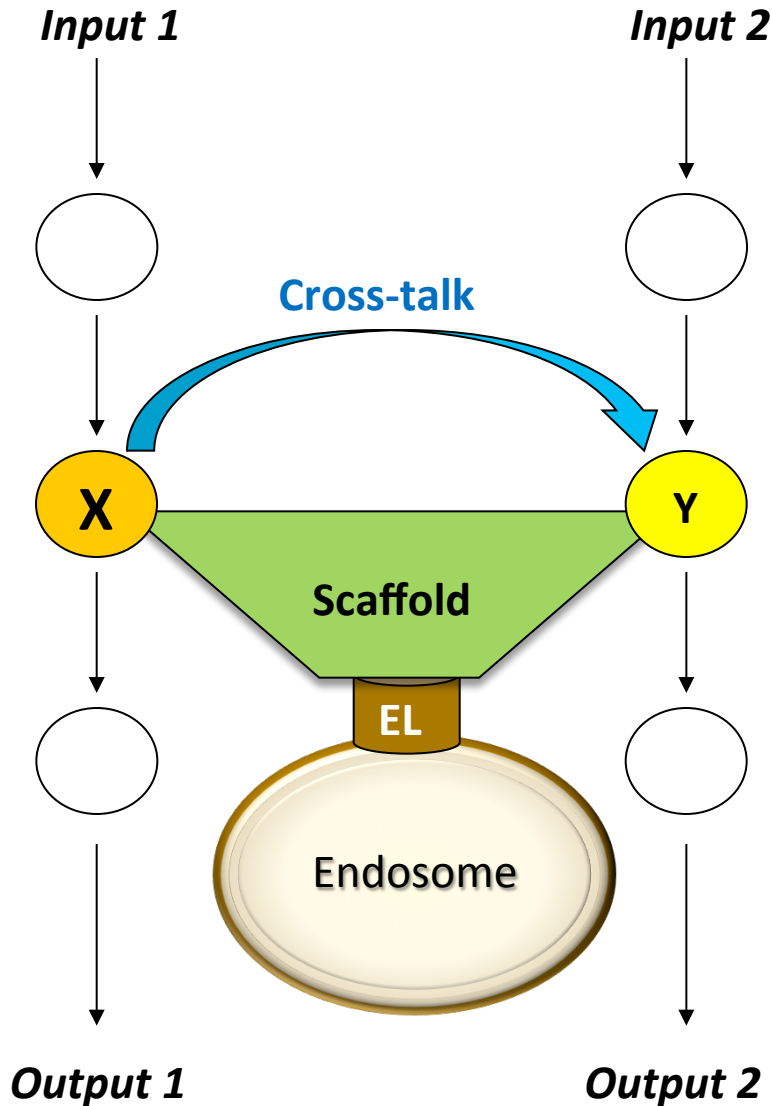
Docking interactions

# Spatial regulation of signalling by endocytosis

Specific down-regulation, recycling, destruction or signal modulation by general and cargo-specific factors



# An opinion: Cross-talk endosomes





# Regulating the expression of signalling components

## Transcriptional regulation

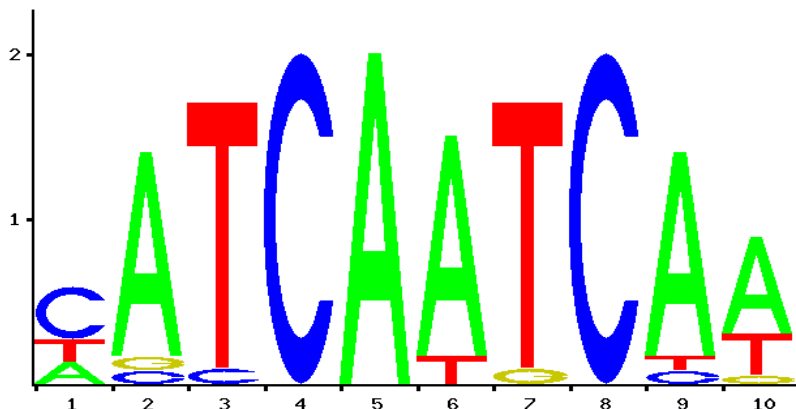
- Activation or inhibition of specific gene expression by transcription factor – transcription factor binding site connection
- JASPAR (<http://jaspar.genereg.net>)
- TFe (<http://cisreg.ca/cgi-bin/tfe/home.pl>)
- HTRIdb (<http://www.lbbc.ibb.unesp.br/htri>)
- PAZAR (<http://pazar.info>)

## Post-transcriptional regulation

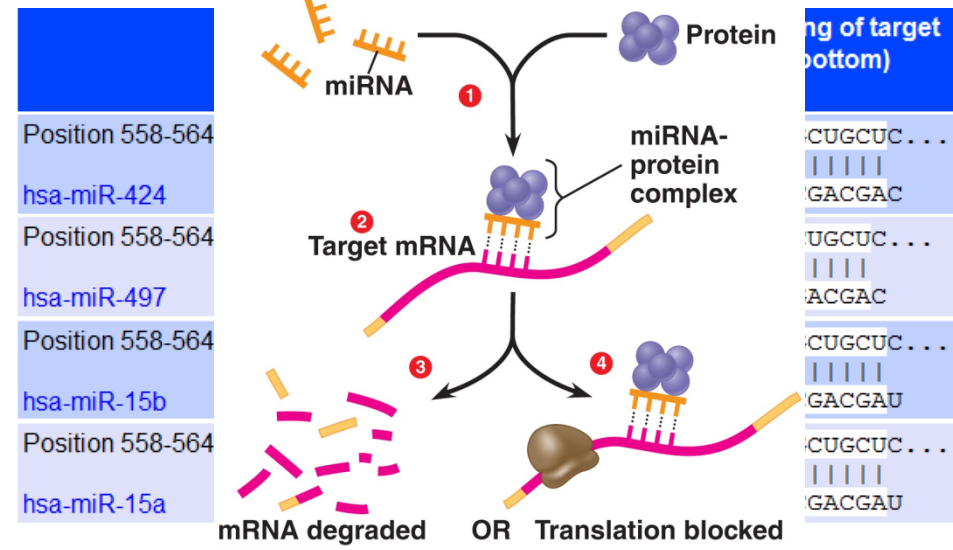
- Destruction or inhibition of specific mRNAs by miRNAs
- miRBase (<http://mirbase.org>)
- miRGen (<http://diana.cslab.ece.ntua.gr/mirgen>)
- miRecords (<http://mirecords.umn.edu/miRecords>)
- TarBase (<http://www.microna.gr/tarbase>)

## Specificity?

### TF-TFBS target motifs

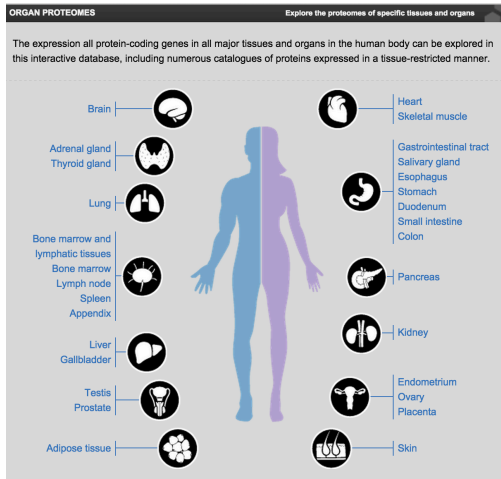


### mRNA-miRNA binding

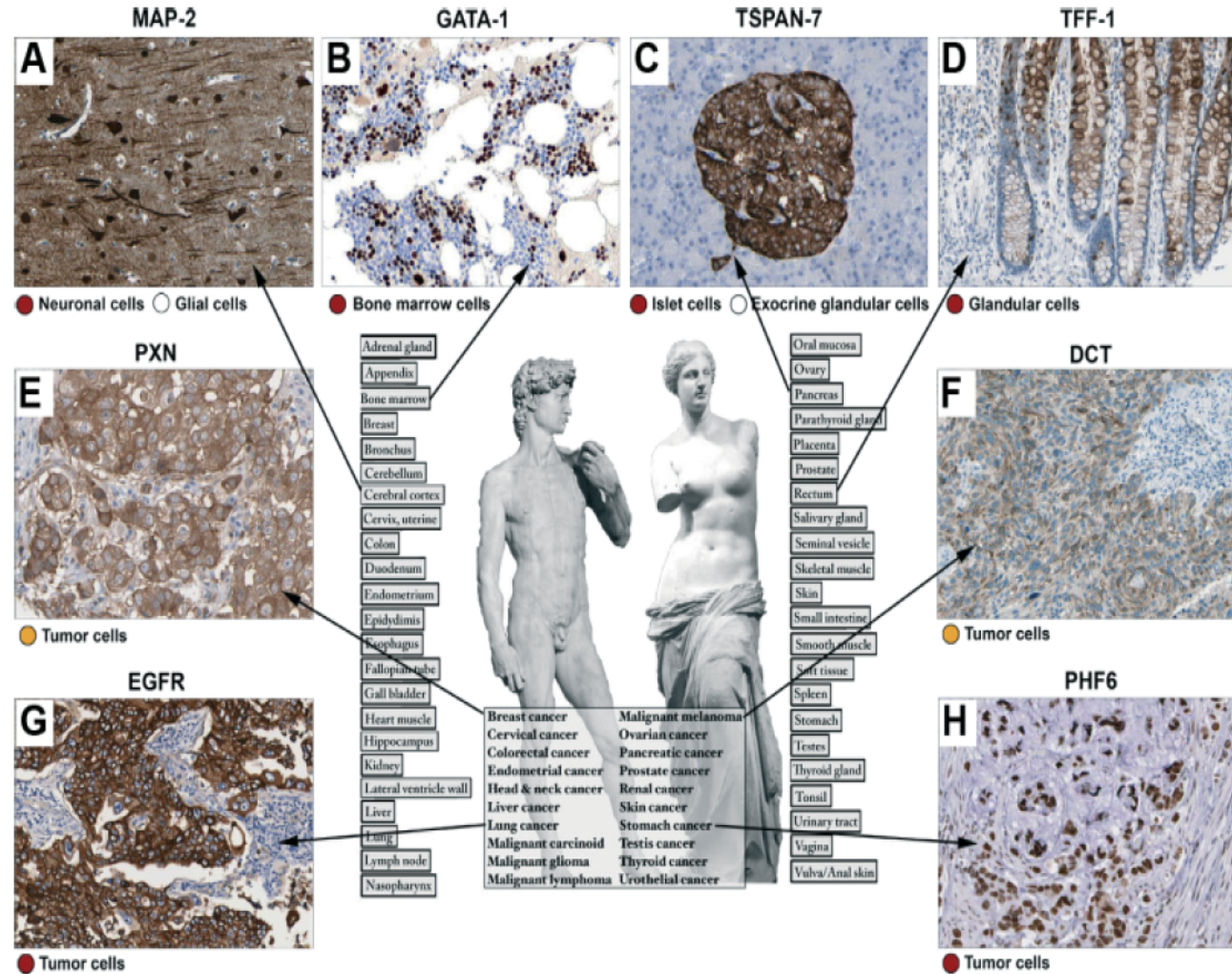


# An important issue: tissue specificity

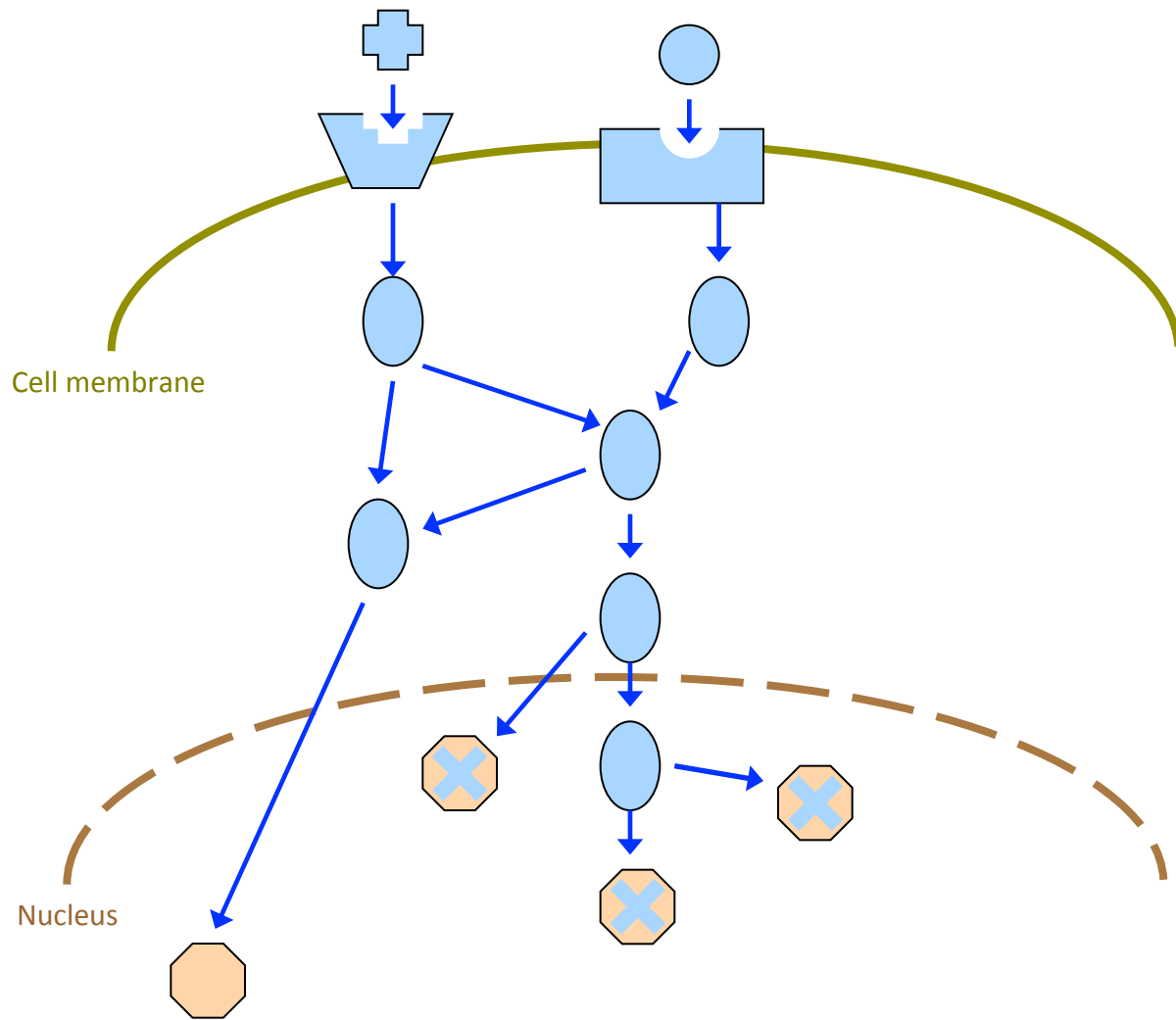
Human Protein Atlas <http://proteatlas.org>



- High-resolution images
- Spatial distribution of proteins
- 44 different normal human tissues and 20 different cancer types
- 46 different human cell lines







Cell membrane

Nucleus

**Legends:**

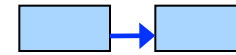


Signalling elements

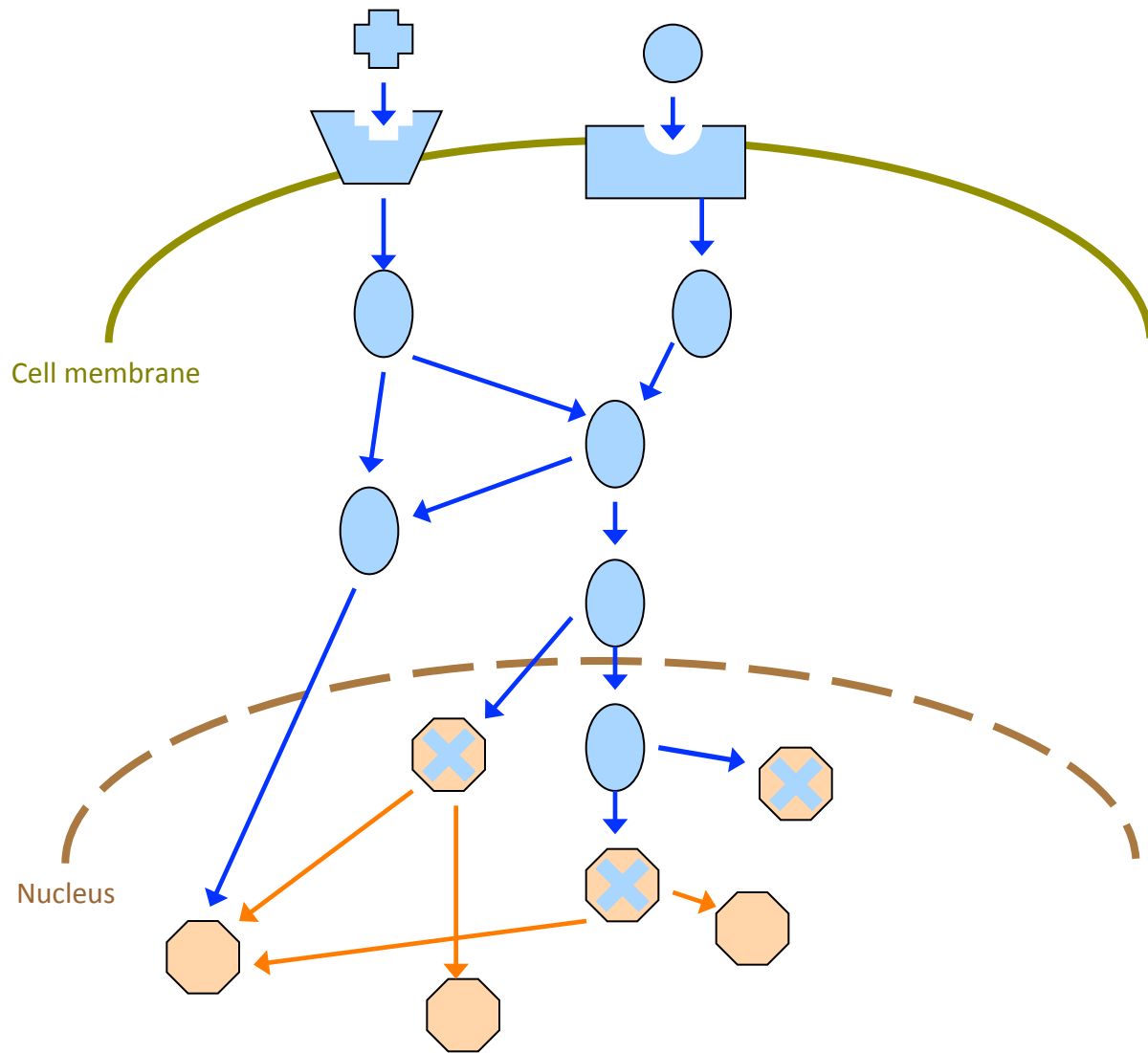


Transcription factors

**Types of networks:**



Network of signalling pathways



### Legends:

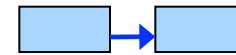


Signalling elements

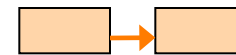


Transcription factors

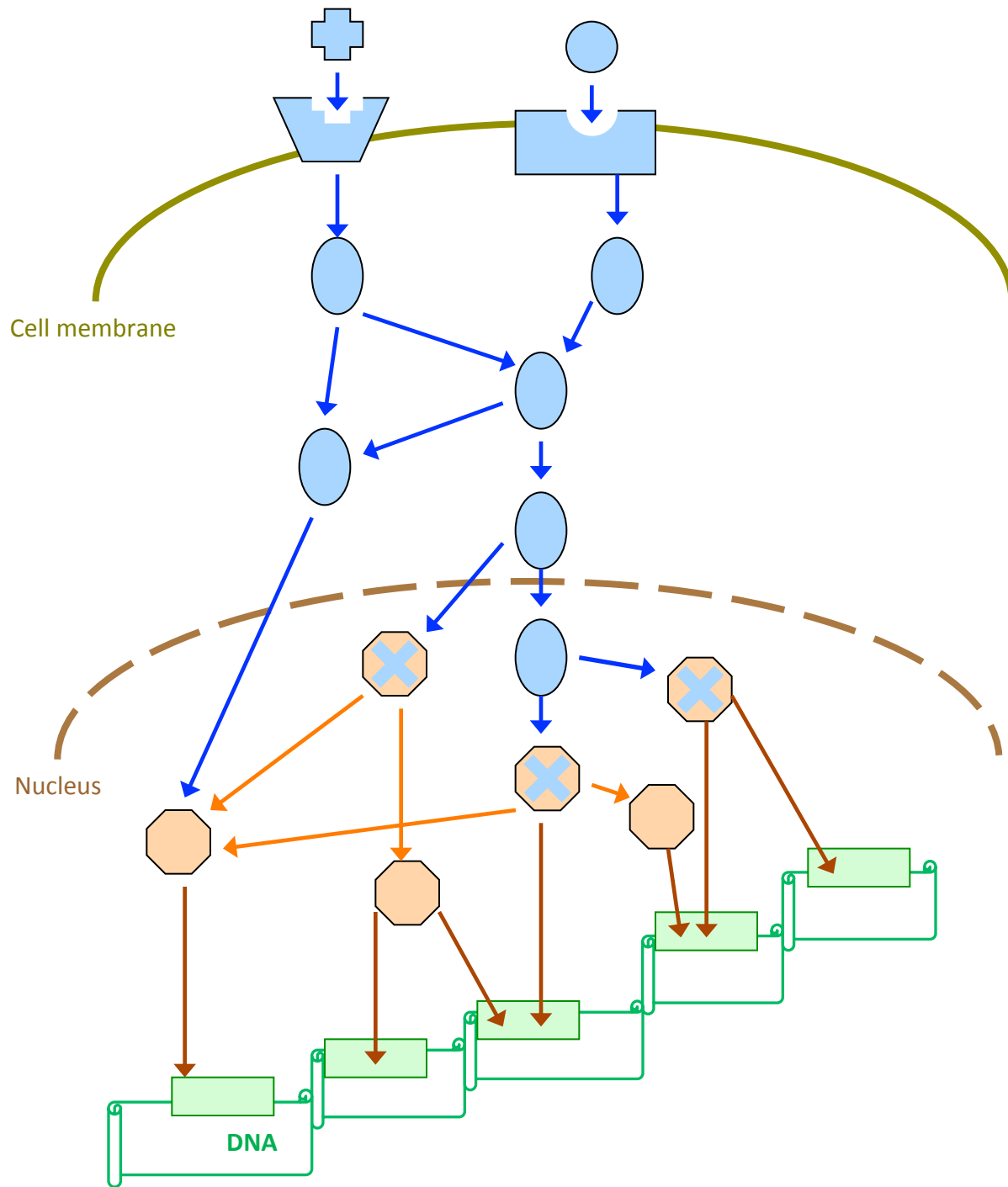
### Types of networks:



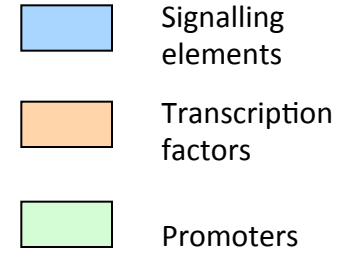
Network of signalling pathways



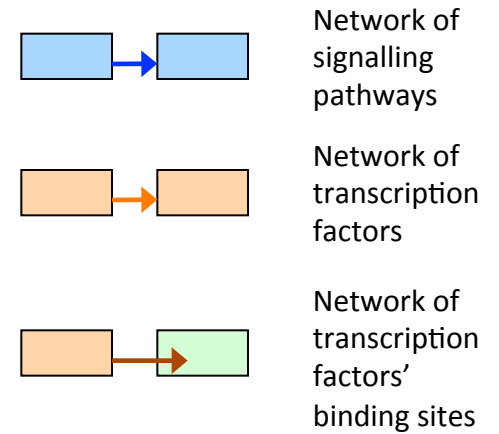
Network of transcription factors

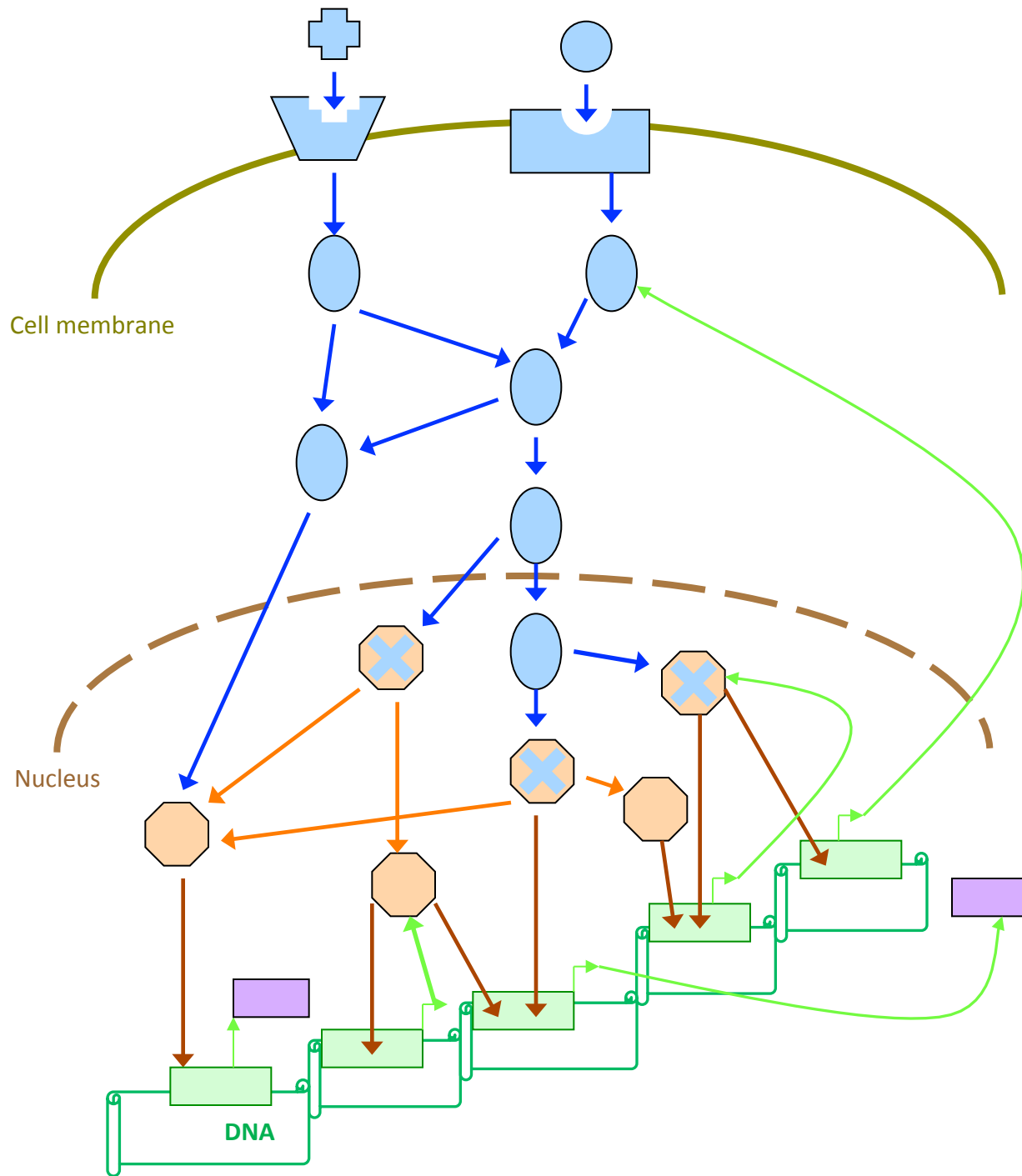


### Legends:

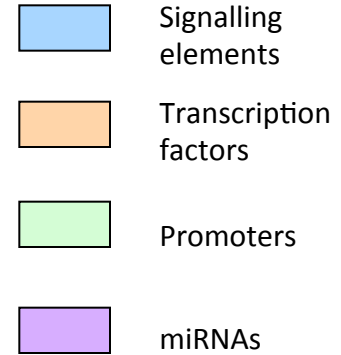


### Types of networks:

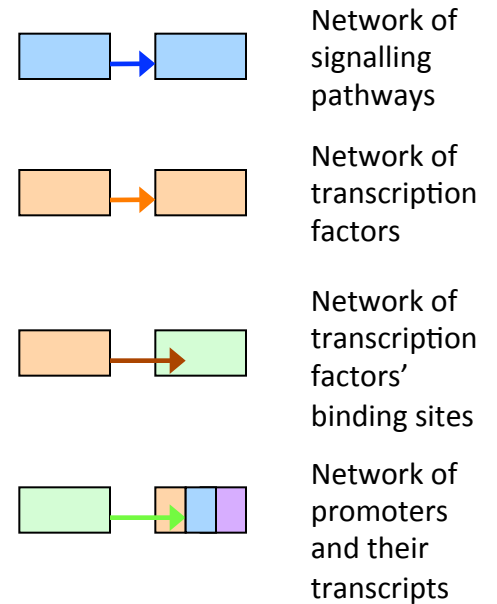


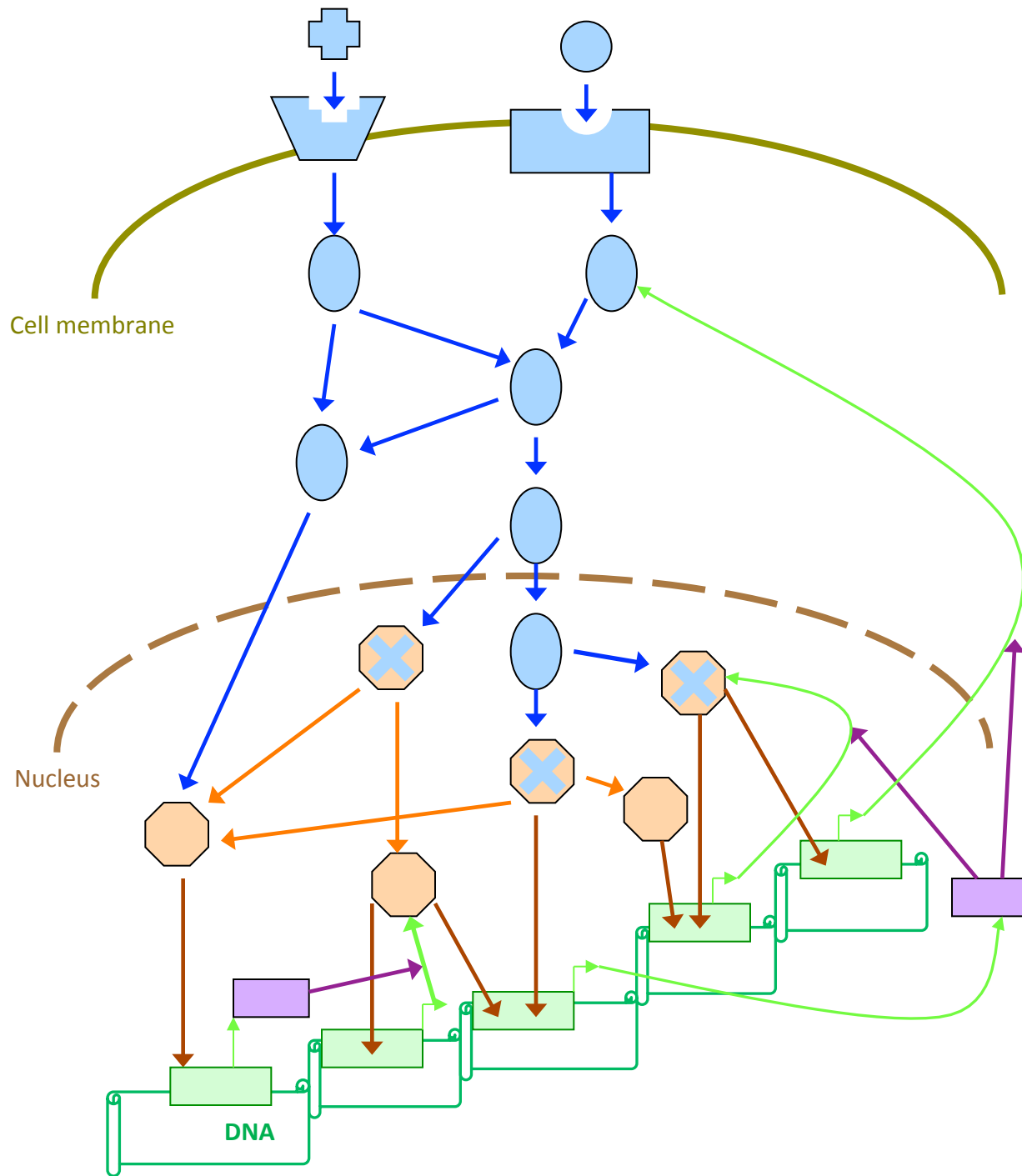


### Legends:







### Types of networks:

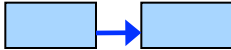
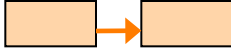







### Legends:

-  Signalling elements
-  Transcription factors
-  Promoters
-  miRNAs

### Types of networks:

-  Network of signalling pathways
-  Network of transcription factors
-  Network of transcription factors' binding sites
-  Network of promoters and their transcripts
-  Network of miRNAs and transcripts

- Reactome
- Signalink
- ConsensusPathDB
- ...

•Fantom4

- ABS
- TRED

- RedFly
- EdgeDB
- JASPAR
- TFe
- HTRIdb
- ENCODE

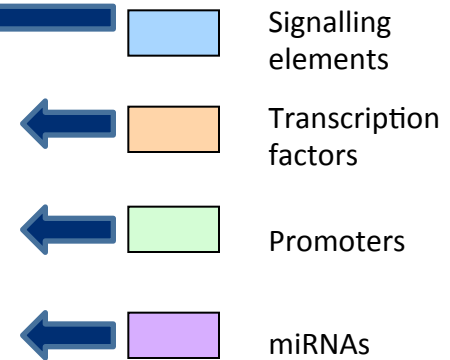
- MPromDB
- miRGen
- PutMir
- TransMir

- miRecords
- TarBase
- ENCODE

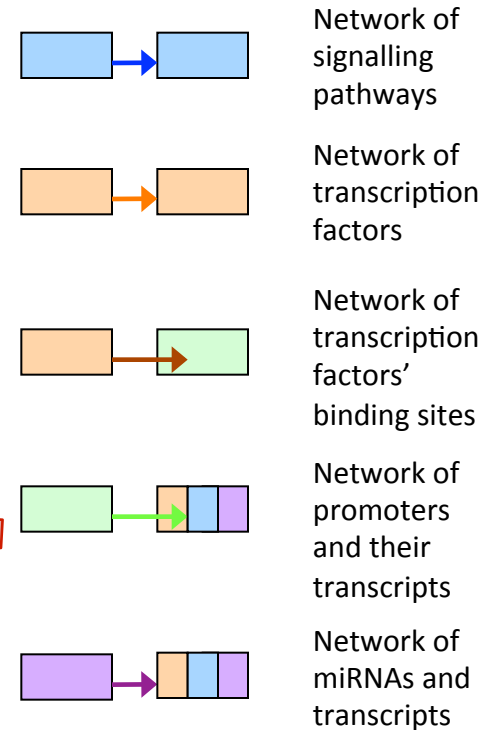
## Component data

- wTF2.1
- FlyTF
- A census of human TFs
- MPromDB
- miRBase
- miROrtho

## Legends:



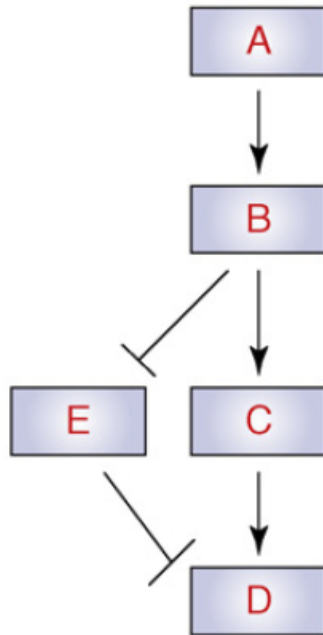
## Types of networks:



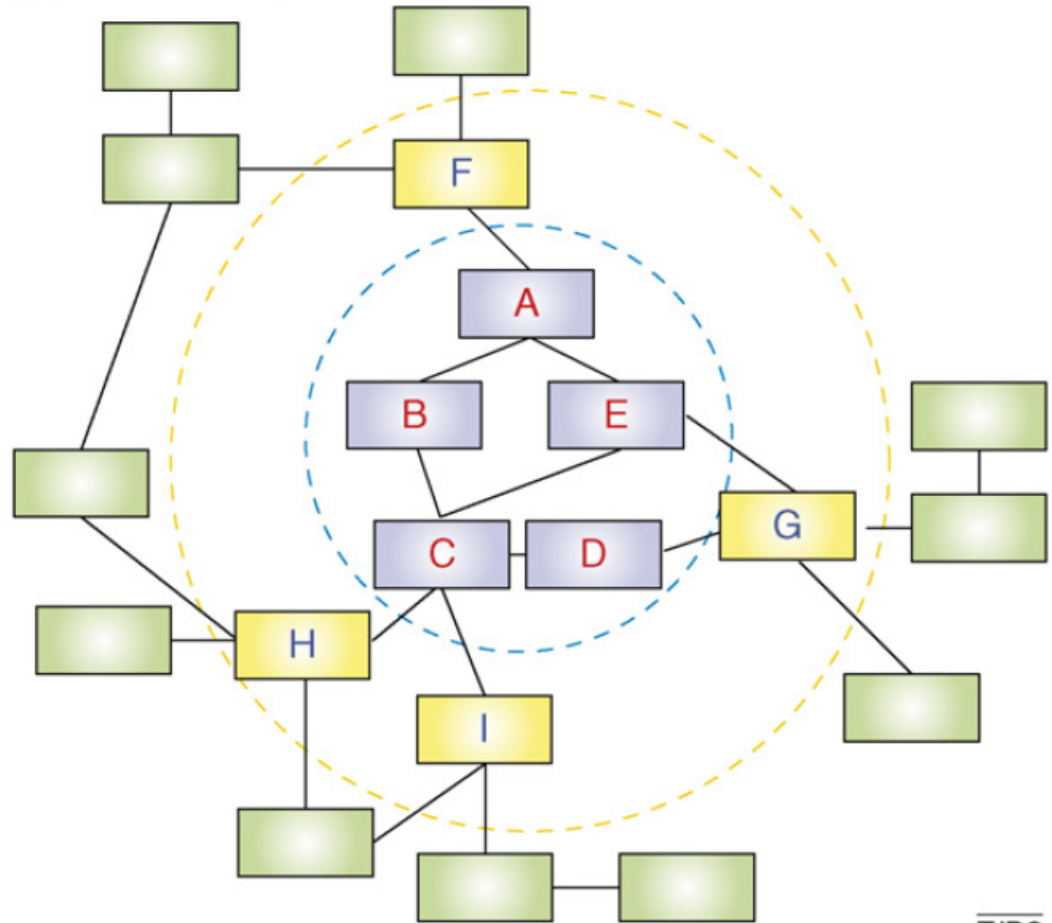
## Interaction data

# Embedding pathways into interaction networks

(a) Classical pathway



(b) Embedded pathway



TiBS



Where can I find PPIs to  
connect with my pathway?

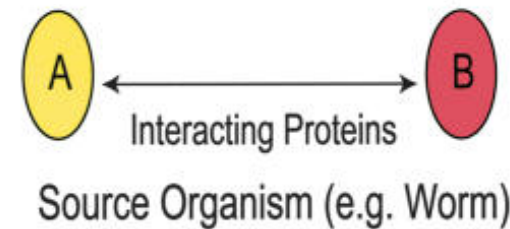
# **The power of prediction**

Applying sequence and structural information  
to predict novel functions / connections

# Protein-protein interactions

## Predicted interactions

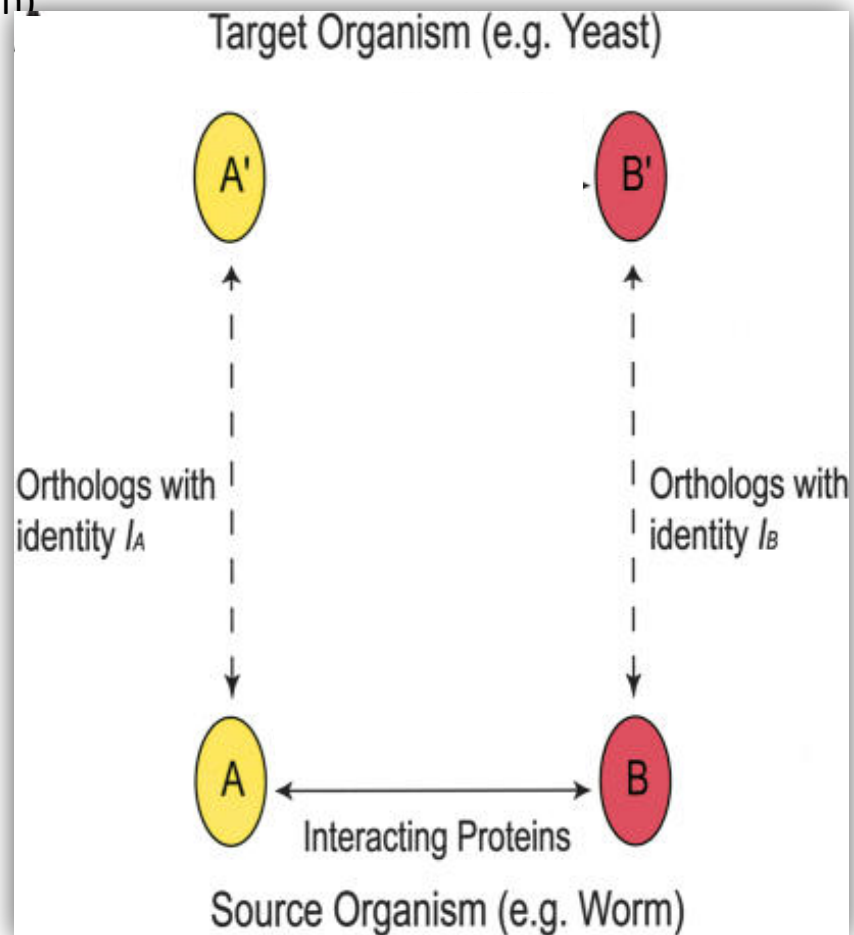
- **Homology/orthology-based (interologs)**
- Domain-motifs based (directed)
- Domain-domain based (interaction & direction)



# Protein-protein interactions

## Predicted interactions

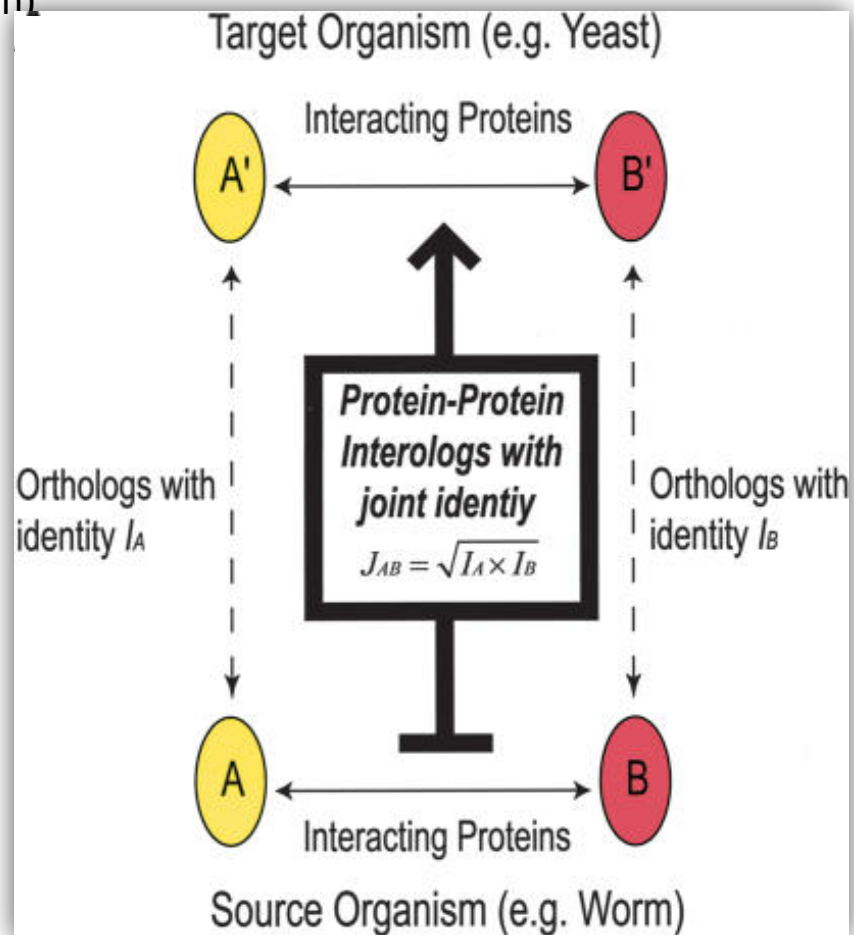
- **Homology/orthology-based (interologs)**
- Domain-motifs based (directed)
- Domain-domain based (interaction & direction)



# Protein-protein interactions

## Predicted interactions

- **Homology/orthology-based (interologs)**
- Domain-motifs based (directed)
- Domain-domain based (interaction & direction)



# Protein-protein interactions

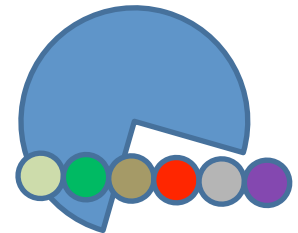
## Predicted interactions

- Homology/orthology-based (interologs)
- **Domain-motifs based (directed)**
- Domain-domain based (interaction & direction)

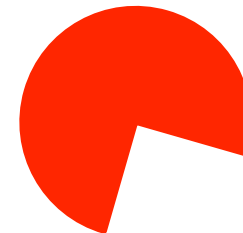
Protein sequence of  
a protein of interest  
(eg., from UniProt)



Domain-motif  
database  
(eg., ELM server)



Enzymatic domain  
capable to target  
the protein



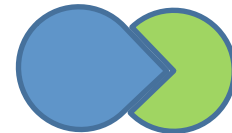
**Predicted PPI based on  
domain-motif interaction**

# Protein-protein interactions

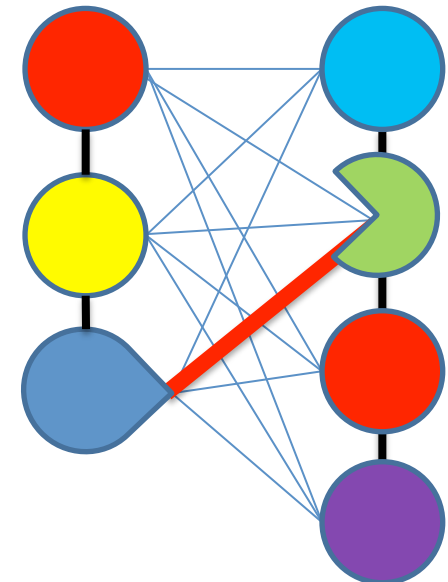
## Predicted interactions

- Homology/orthology-based (interologs)
- Domain-motifs based (directed)
- **Domain-domain based (interaction & direction)**

Domain-domain  
interaction data  
(eg., DOMINE)



Protein-domain  
composition  
data (eg., PFAM)



Possible  
domain  
pairs

**Predicted PPI based on  
domain-domain interaction**

# Protein-protein interactions

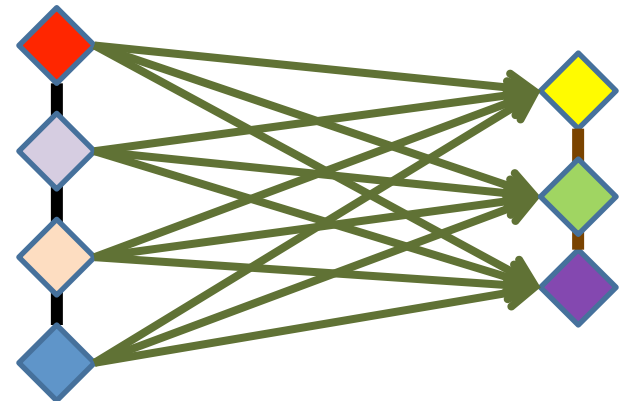
## Predicted interactions

- Homology/orthology-based (interologs)
- Domain-motifs based (directed)
- **Domain-domain based** (interaction & direction)

Directed PPI  
from the  
reference  
database  
(eg., Reactome)



Domain  
composition  
as training  
set  
(eg., PFAM)





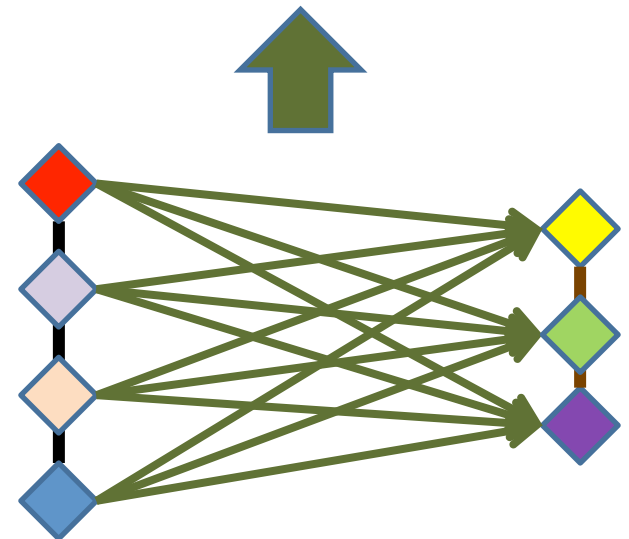
# Protein-protein interactions

## Predicted interactions

- Homology/orthology-based (interologs)
- Domain-motifs based (directed)
- **Domain-domain based** (interaction & direction)

$$F(\text{blue} \rightarrow \text{yellow}) = \frac{\Pr(\text{blue} \rightarrow \text{yellow}) - \Pr(\text{yellow} \rightarrow \text{blue})}{\Pr(\text{blue}) \times \Pr(\text{yellow})}$$

Domain  
composition  
as training  
set  
(eg., PFAM)



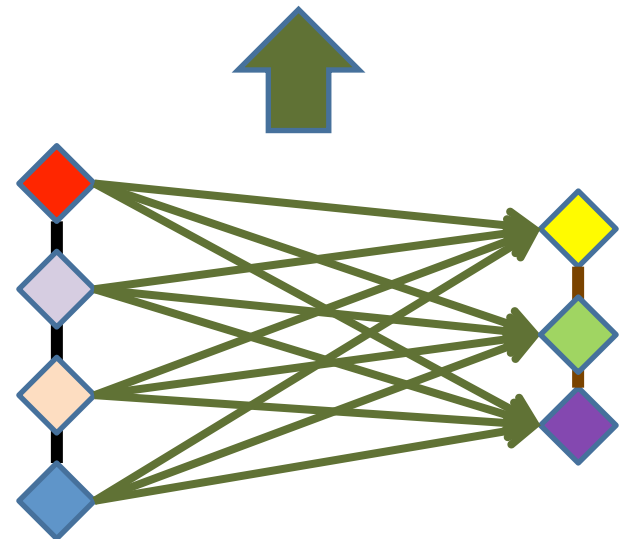
# Protein-protein interactions

## Predicted interactions

- Homology/orthology-based (interologs)
- Domain-motifs based (directed)
- **Domain-domain based** (interaction & direction)

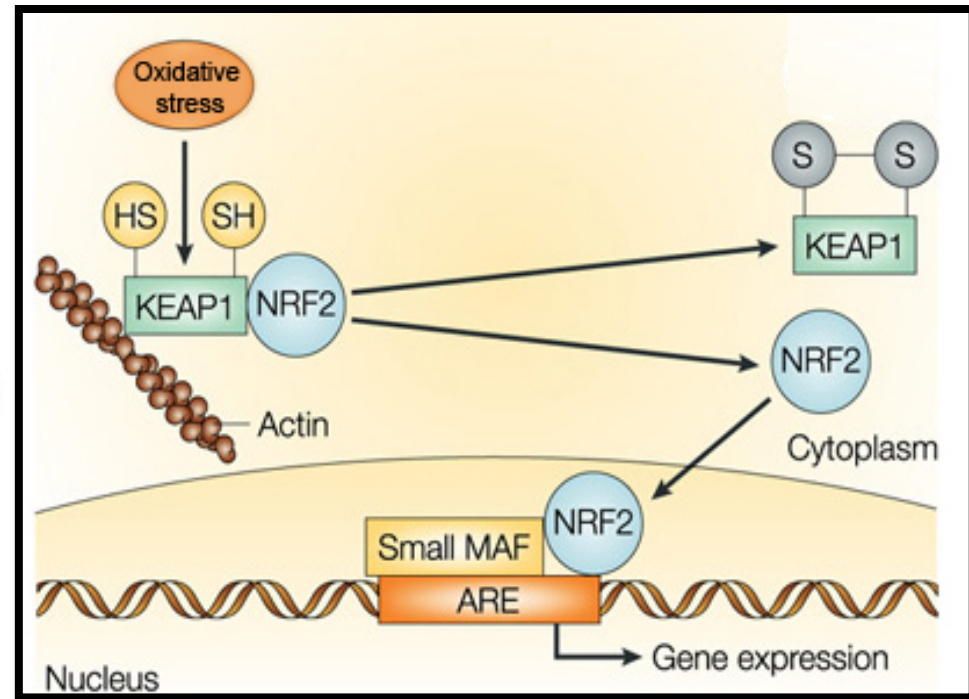
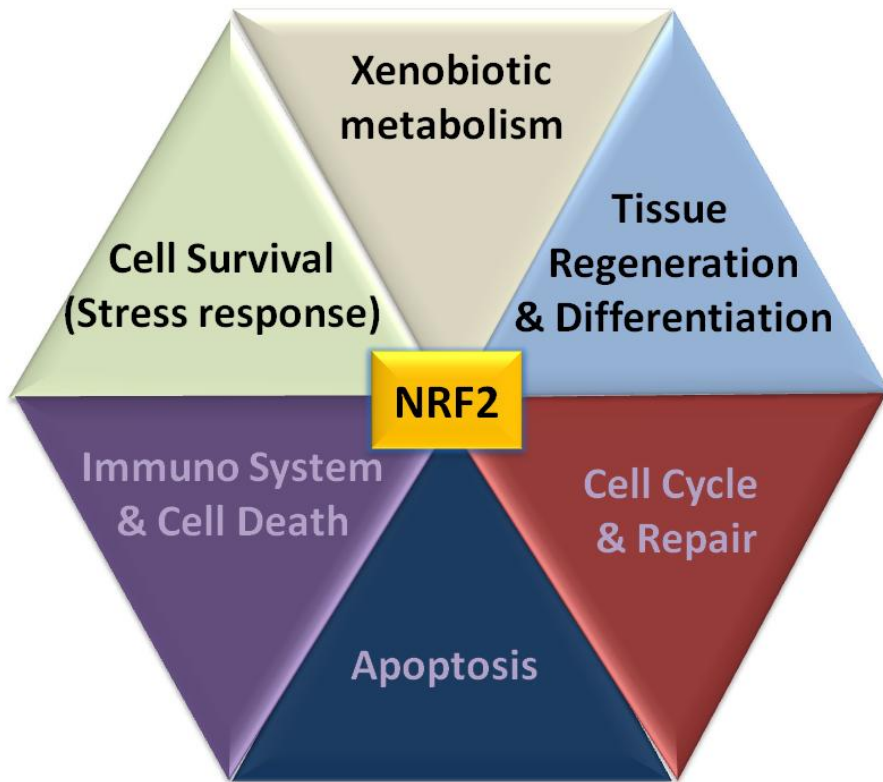
$$F(d_{mn}) = \frac{\Pr(d_m \rightarrow d_n) - \Pr(d_n \rightarrow d_m)}{\Pr(d_m) \times \Pr(d_n)}$$

Domain  
composition  
as training  
set  
(eg., PFAM)



# Two use cases for data integration

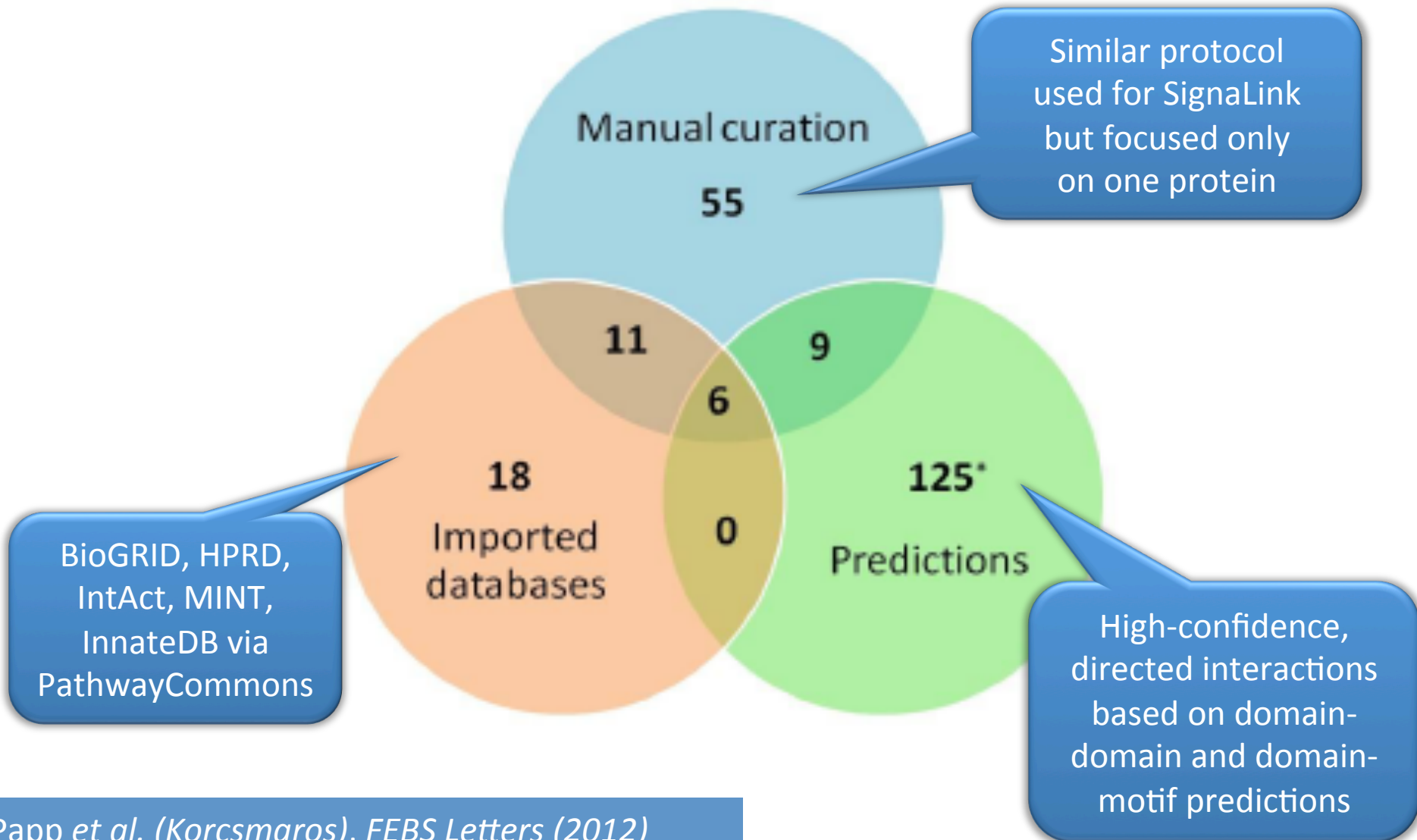
# NRF2 interactome & regulome



- Cancer
- Ageing
- Inflammation
- Diseases with oxidative stress

# Reconstructing the NRF2 interactome

## Distribution of the NRF2 interactors by sources

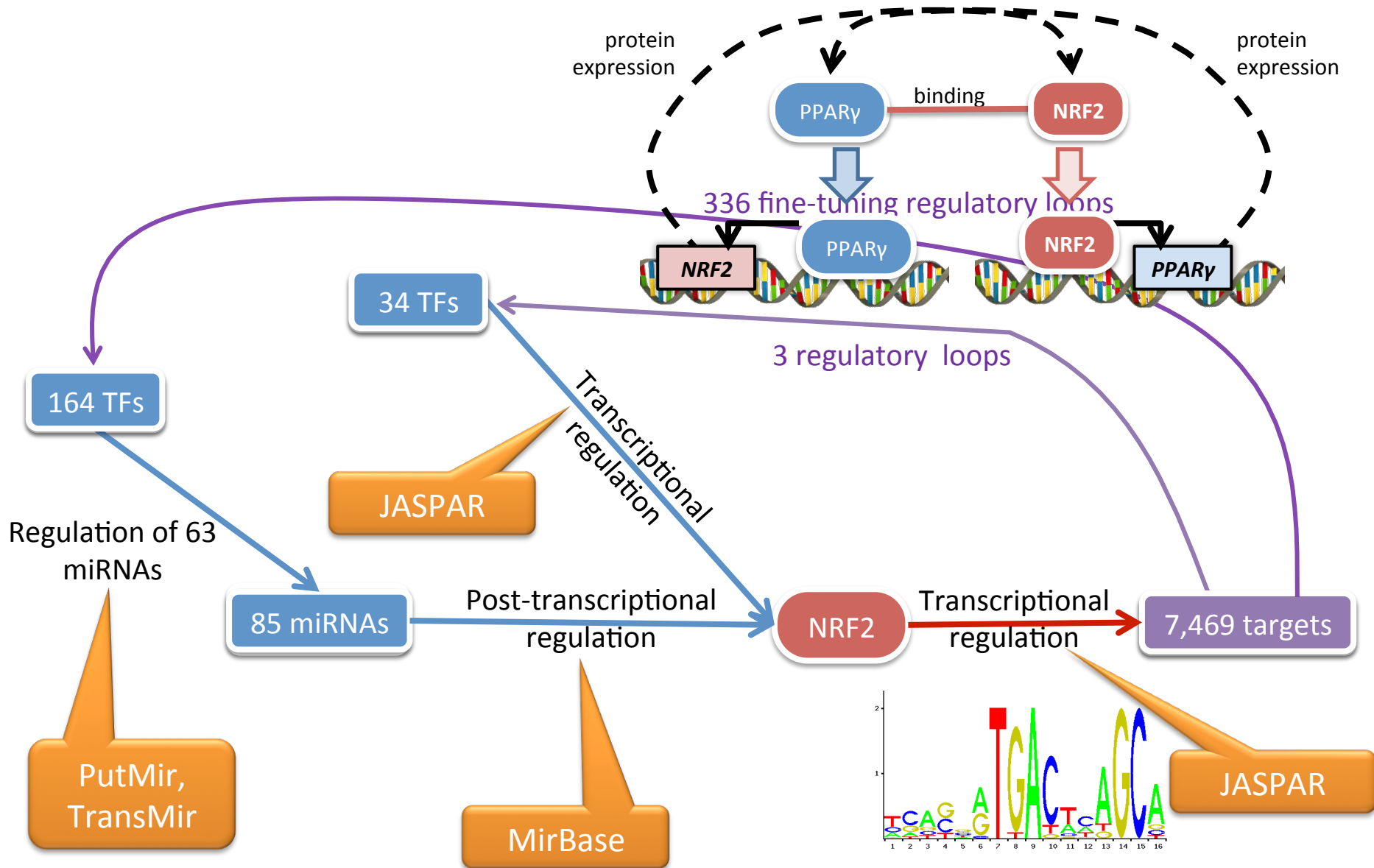




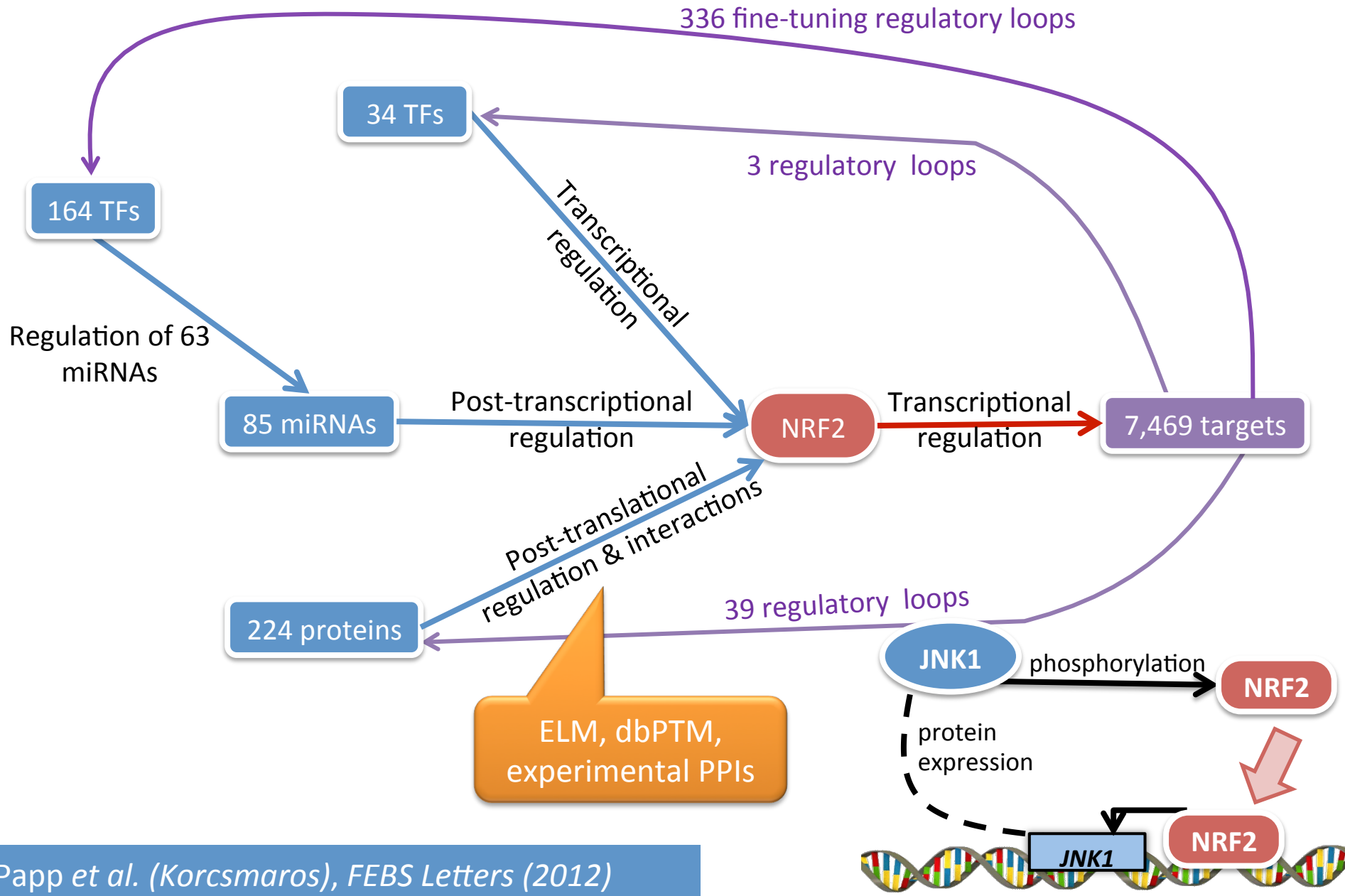




# Reconstructing the NRF2 regulome



# Reconstructing an integrated NRF2 network



- Reactome
- Signalink
- ConsensusPathDB
- ...

•Fantom4

- ABS
- TRED

- RedFly
- EdgeDB
- JASPAR
- TFe
- HTRIdb
- ENCODE

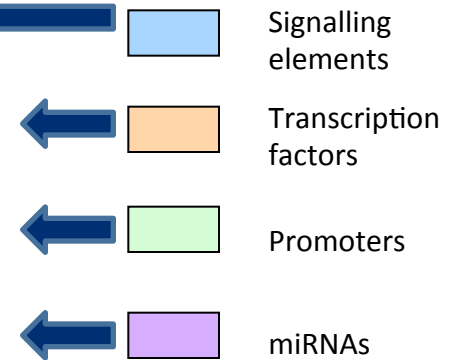
- MPromDB
- miRGen
- PutMir
- TransMir

- miRecords
- TarBase
- ENCODE

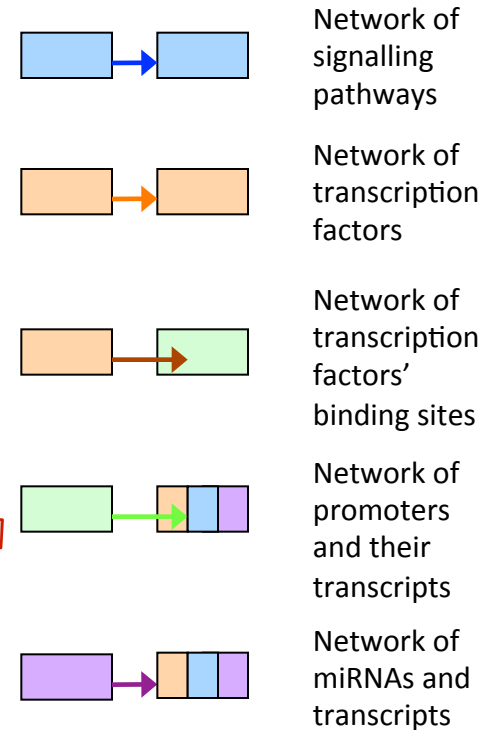
## Component data

- wTF2.1
- FlyTF
- A census of human TFs
- MPromDB
- miRBase
- miROrtho

## Legends:



## Types of networks:

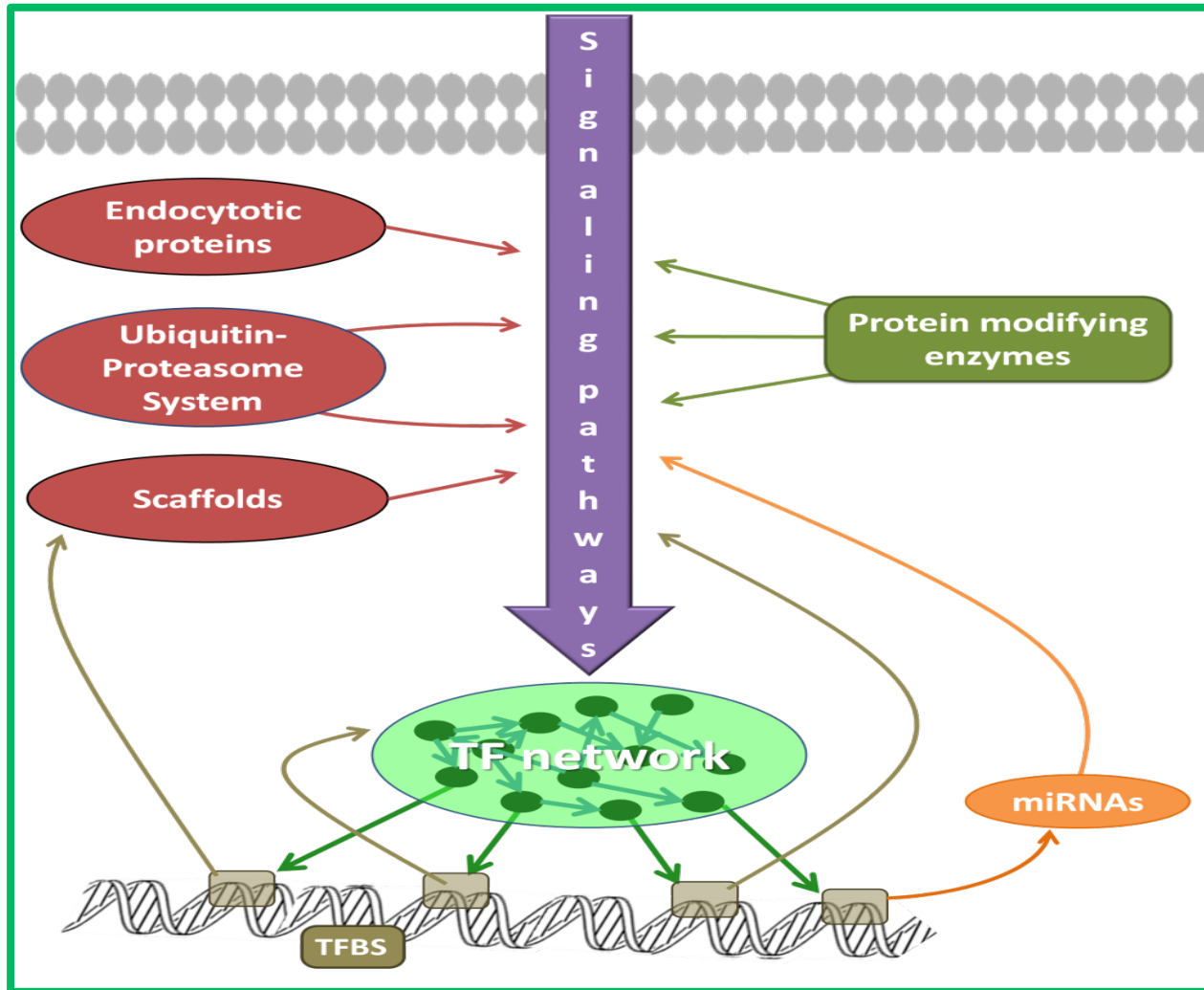


## Interaction data

# Signalink 2.0

Fazekas *et al* BMC Syst Biol, 2013

A signalling pathway resource with a multi-layered regulatory network



# SignaLink 2.0

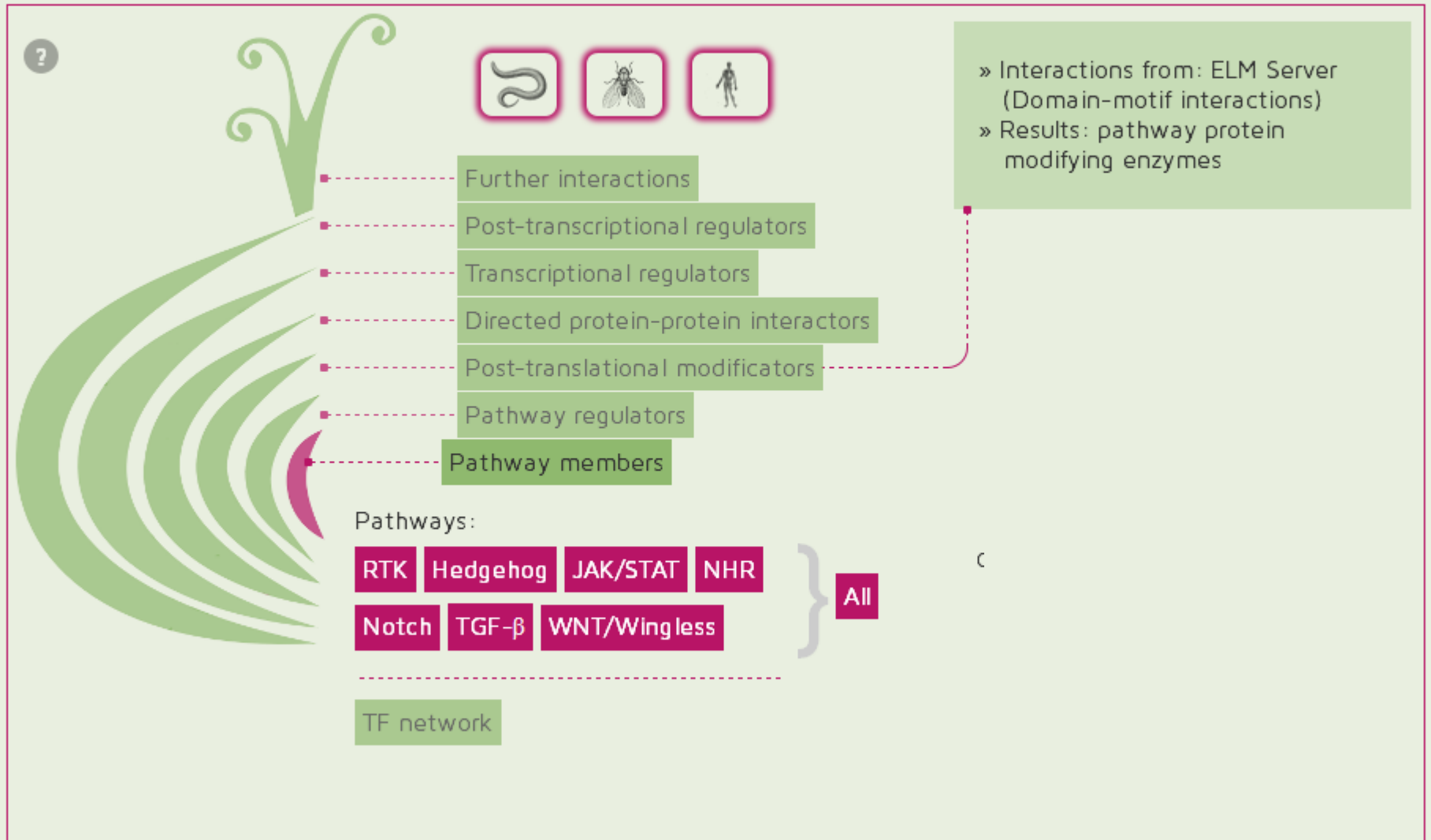
search

download  
tools

publications  
people

contact  
faq

A signaling pathway resource with multi-layered regulatory networks



# SignaLink 2.0

search

download  
tools

publications  
people

contact  
faq

A signaling pathway resource with multi-layered regulatory networks

?

Further interactions

Post-transcriptional regulators

Transcriptional regulators

Directed protein-protein interactors

Post-translational modifiers

Pathway regulators

Pathway members

Pathways:

RTK Hedgehog JAK/STAT NHR } All

Notch TGF- $\beta$  WNT/Wingless

TF network

Range from pathway specific TF:

1 2 3 all

» Interactions from literature curation

» Results: endocytotic and scaffold proteins



# SignaLink 2.0

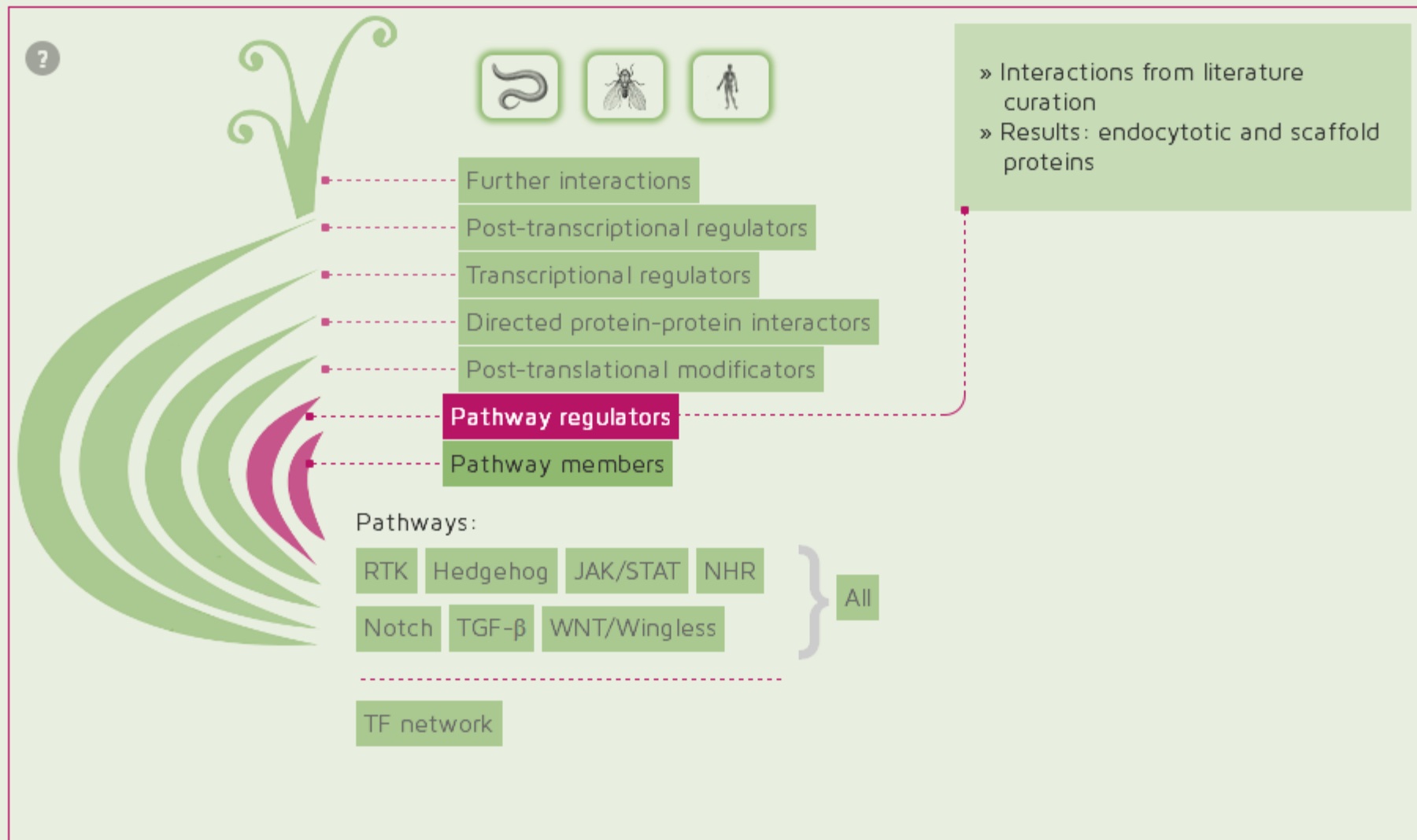
search

download  
tools

publications  
people

contact  
faq

A signaling pathway resource with multi-layered regulatory networks



# SignaLink 2.0

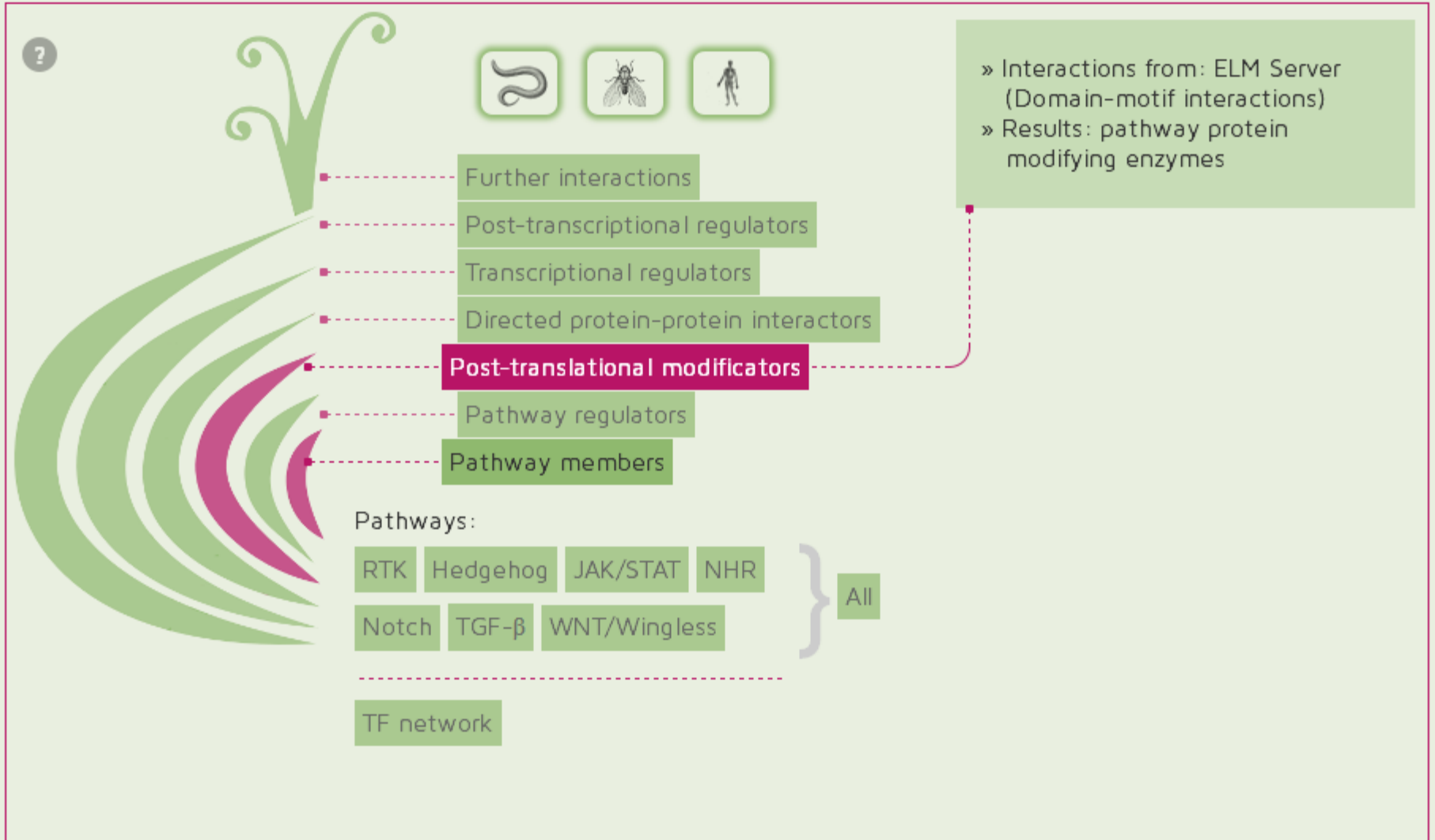
search

download  
tools

publications  
people

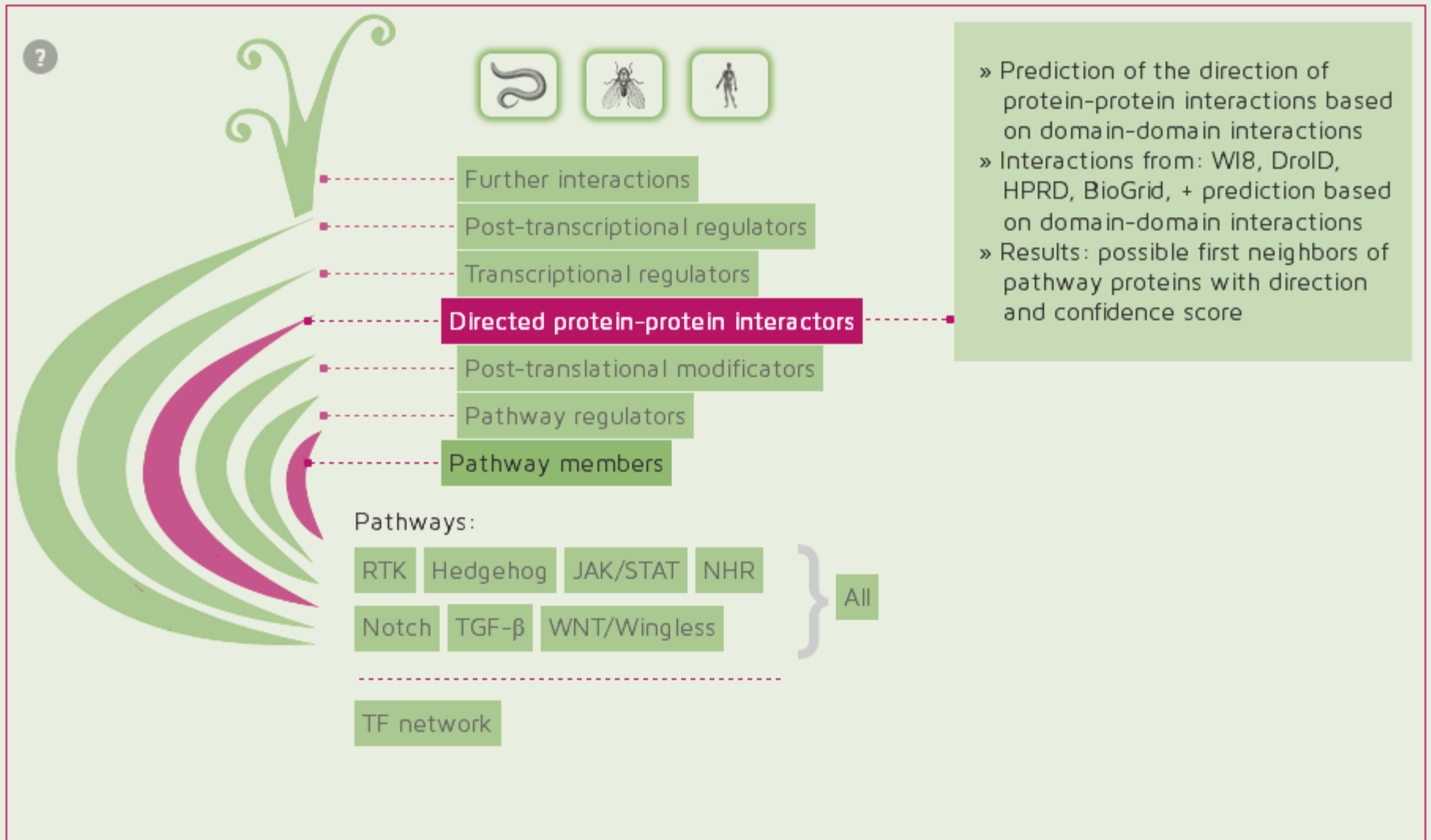
contact  
faq

A signaling pathway resource with multi-layered regulatory networks



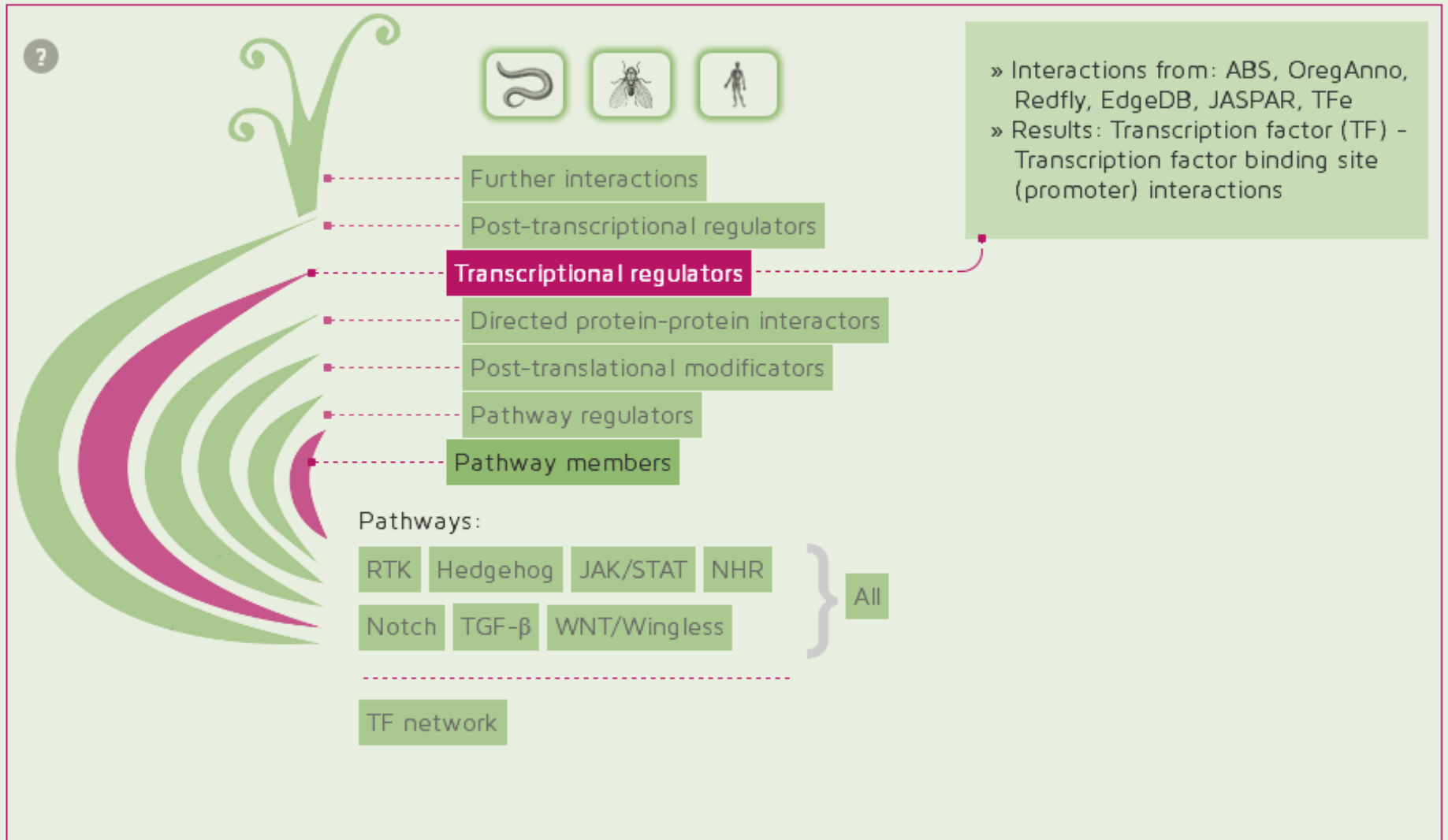
# SignaLink 2.0

search download tools publications people contact faq  
A signaling pathway resource with multi-layered regulatory networks



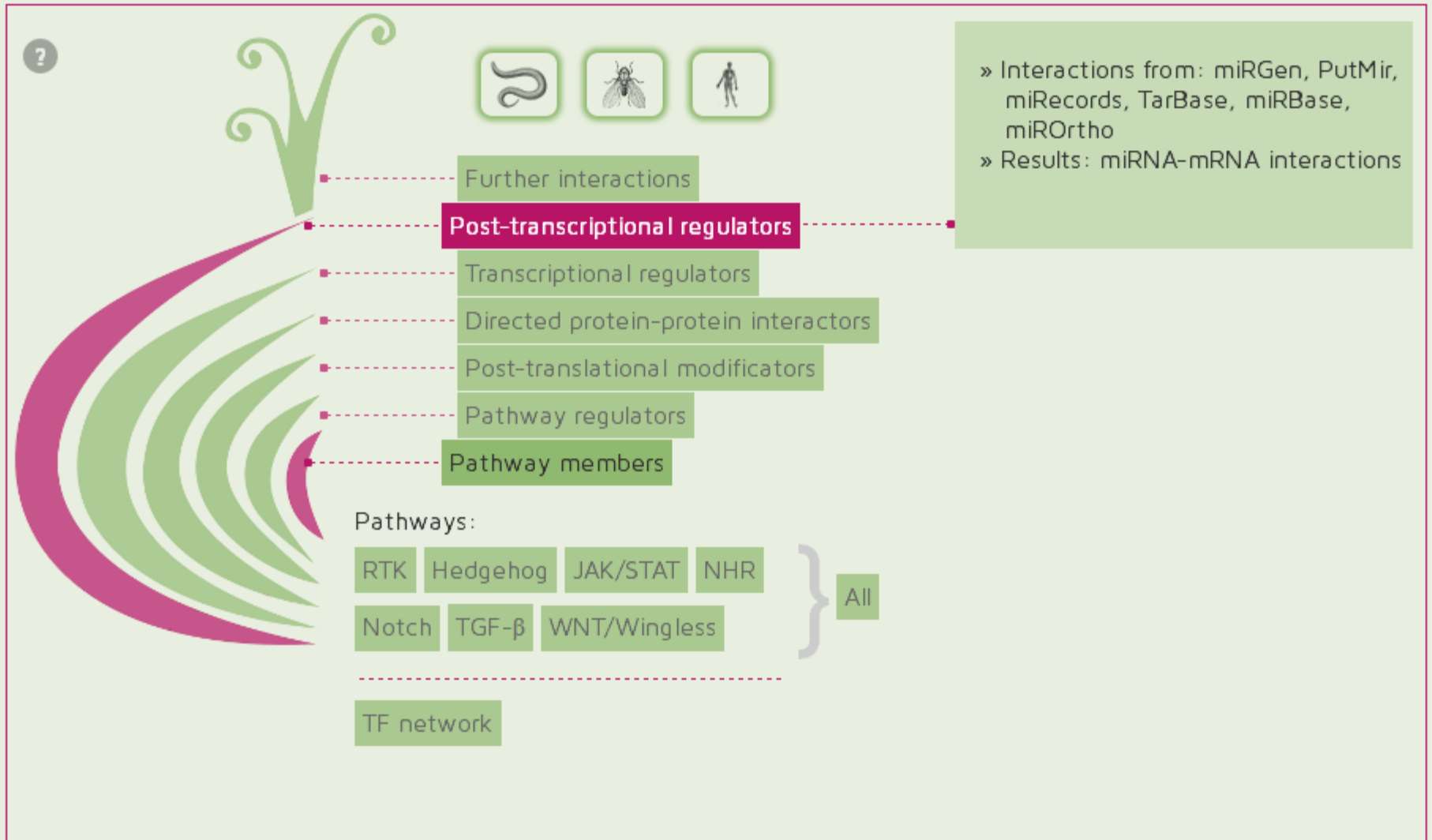
# SignaLink 2.0

search download tools publications people contact faq  
A signaling pathway resource with multi-layered regulatory networks



# SignaLink 2.0

search download tools publications people contact faq  
A signaling pathway resource with multi-layered regulatory networks





# SignaLink 2.0

search

download

tools

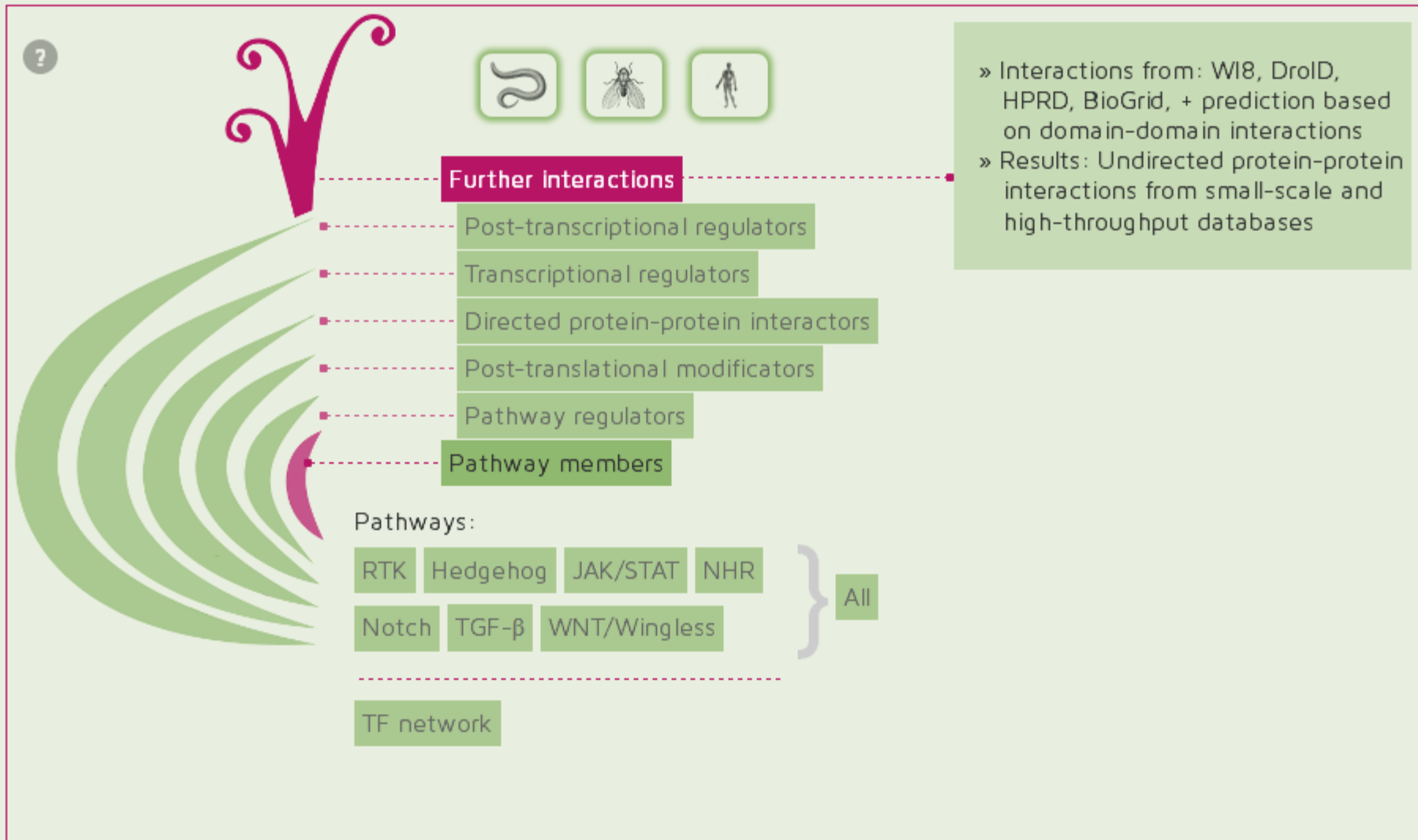
publications

people

contact

faq

A signaling pathway resource with multi-layered regulatory networks



# SignaLink 2.0

search

download

tools

publications

people

contact

faq

A signaling pathway resource with multi-layered regulatory networks

**Further Interactions**

- Post-transcriptional regulators
- Transcriptional regulators
- Directed protein-protein interactors
- Post-translational modifiers
- Pathway regulators
- Pathway members

**Pathways:**

- RTK
- Hedgehog
- JAK/STAT
- NHR
- Notch
- TGF- $\beta$
- WNT/Wingless

TF network

Output format:

- csv
- biopax (level 3)
- psimi tab
- psimi xml
- sbml
- cytoscape

Compress:  gzip  zip  none

**Proceed to download**

OR

Complete database

» Interactions from: WI8, Droid, HPRD, BioGrid, + prediction based on domain-domain interactions  
» Results: Undirected protein-protein interactions from small-scale and high-throughput databases

?

Further interactions

Post-transcriptional regulators

Transcriptional regulators

Directed protein-protein interactors

Post-translational modifiers

Pathway regulators

Pathway members

Pathways:

RTK Hedgehog JAK/STAT NHR

Notch TGF- $\beta$  WNT/Wingless

All

Linkouts from UniProt, Wormbase, Flybase

**SignaLink 3 – coming soon**

- Zebrafish
- Extended coverage
- More signs for PPIs

Output format:

CSV biopax (level 3) psimi tab

psimi xml sbml cytoscape

Compress: » gzip » zip » none

Proceed to download

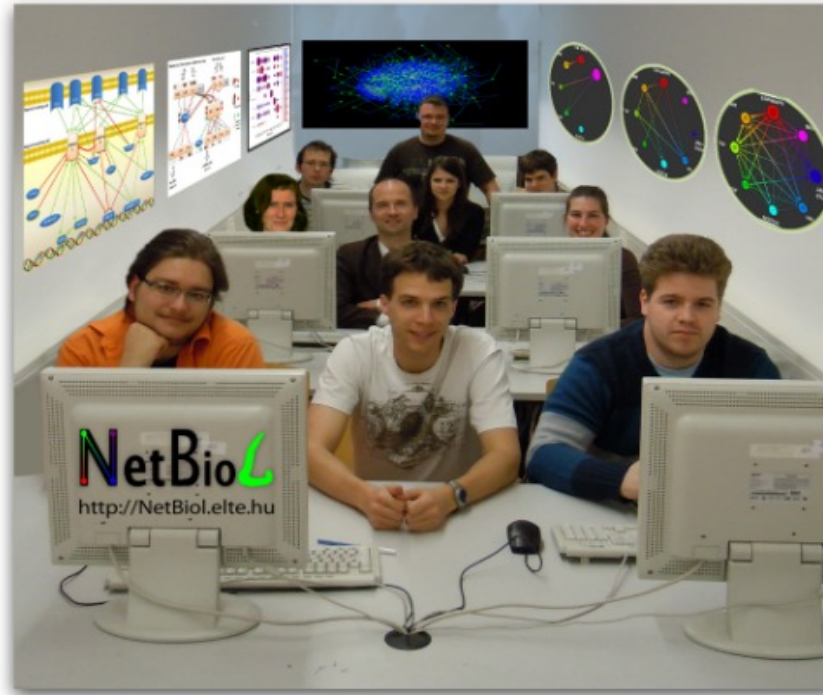
OR

Complete database

# Discussion



# Thank you for your attention!



[Tamas.Korcsmaros@tgac.ac.uk](mailto:Tamas.Korcsmaros@tgac.ac.uk)

# Signalling Networks: From data to modelling

Monday 25 to Friday 29 January 2016

The Genome Analysis Centre, Norwich, UK

**Introduction and hands-on training about pathway resources, tools and modelling approaches from expert researchers**

## Course Faculty

**Laurence Calzone**

Institut Curie (France)

**Tamas Korcsmaros**

TGAC (UK)

**Pablo Porras**

EMBL-EBI

**Julio Saez-Rodriguez**

JRC for Computational Biomedicine (Germany)

**Jean-Marc Schwartz**

Univ. of Manchester (UK)

**Denes Turei**

EMBL-EBI

## You will learn about

Reliable signalling databases

Cytoscape

Network reconstruction

Pathway visualisation

Model building

Logic modelling

Apply a model for your own work

<http://tinyurl.com/signet16>

**Registration closes 30 October 2015**

No programming skills and modelling background are required.