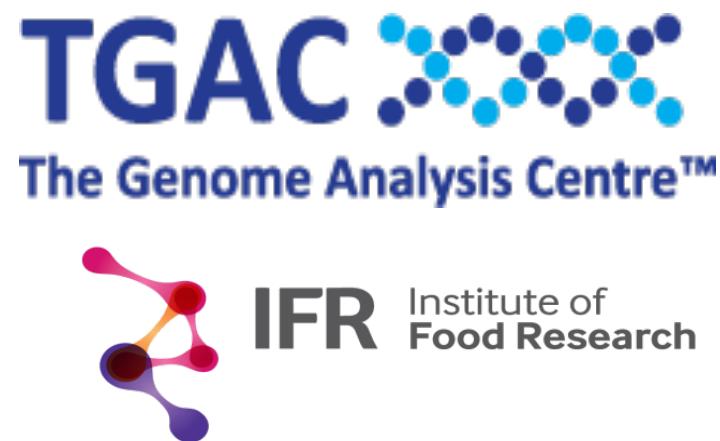
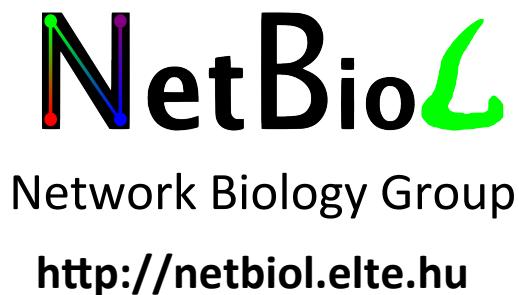


Pulling it all together

From protein sequences to PPI networks and beyond

Tamás Korcsmáros



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

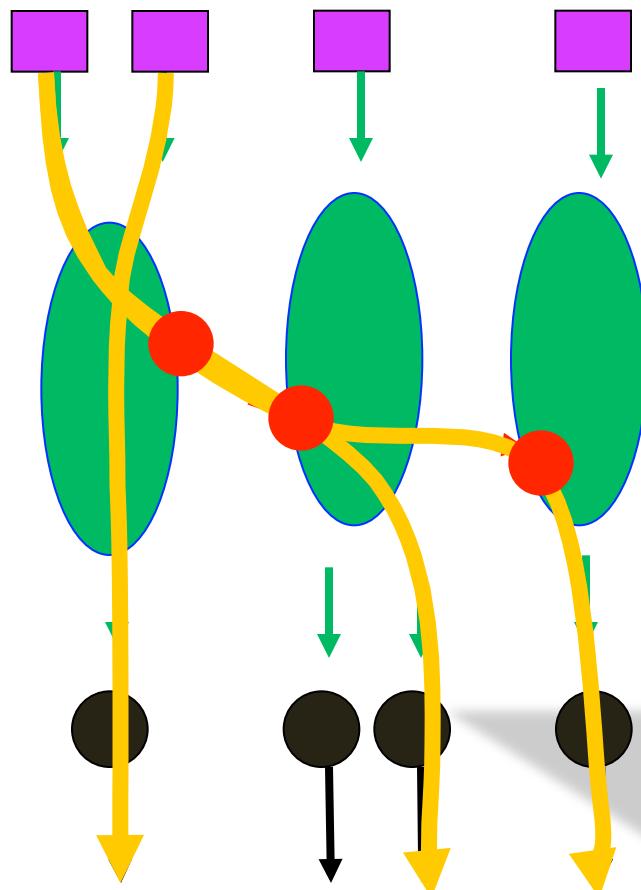
- **From pathways to networks**
 - Visualization challenges and solutions
- **Extending pathways**
 - Data integration
 - Power of prediction
- **Use cases for data integration**
 - NRF2ome
 - SignaLink 2

Signalling pathways and networks

Input
Ligands

Pathway
mediators
(cross-talk
proteins)

Output
Transcription
factors



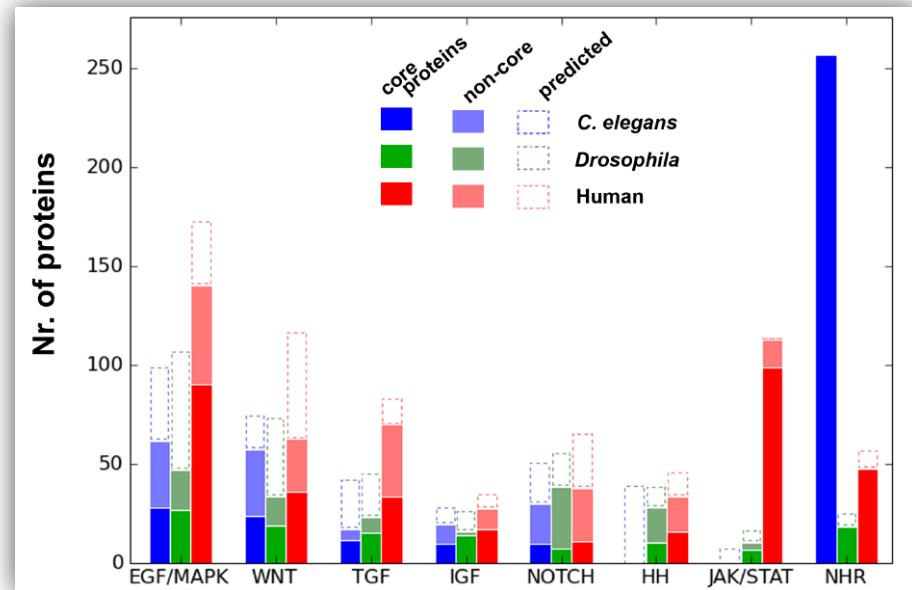
Problems and challenges in signalling pathway databases

- 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*



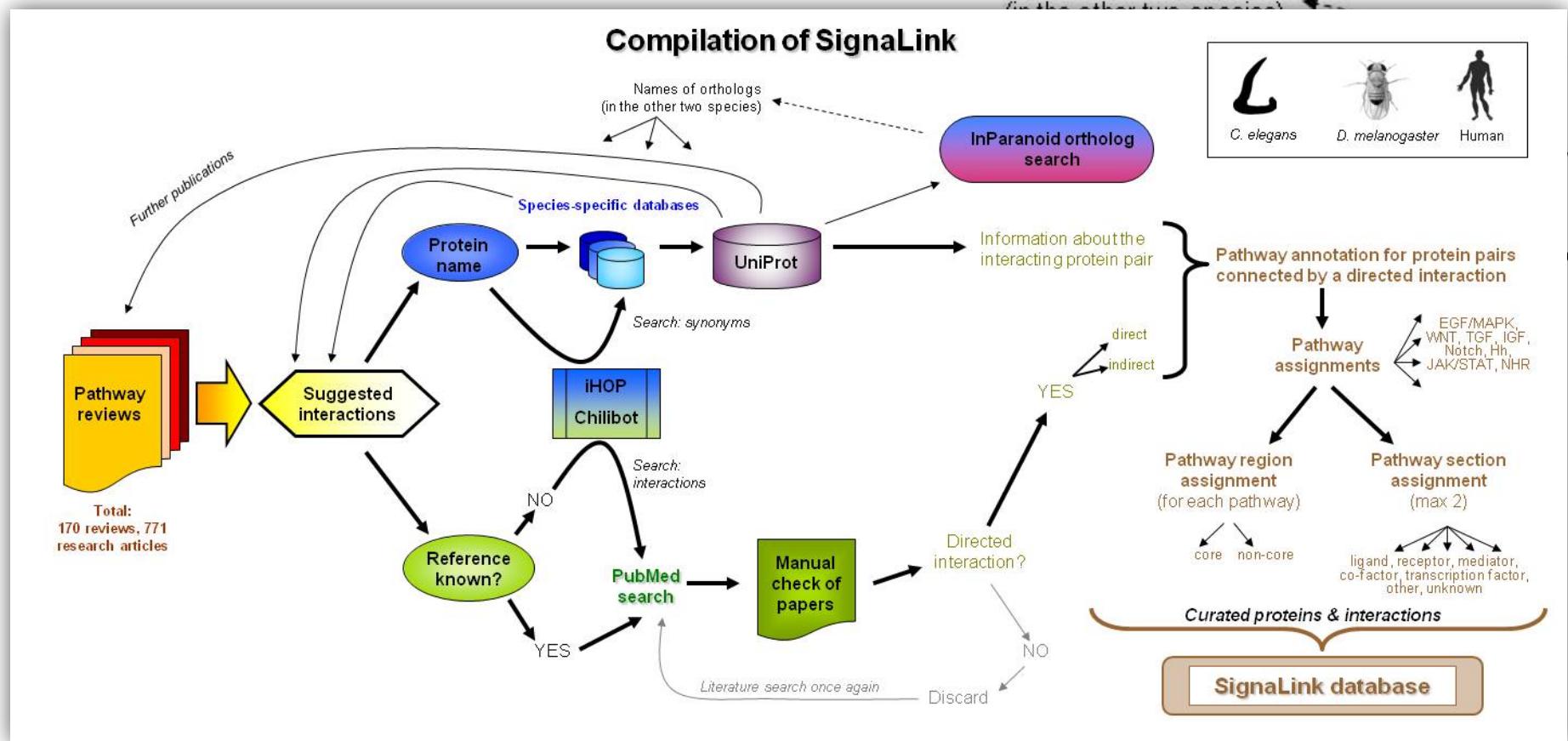
Compared to 3 pathway databases

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

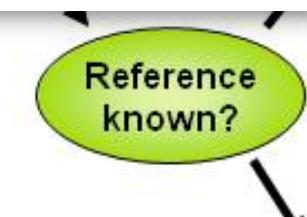
Allows the system-level examination of the signalling network

Curation protocol of SignaLink

Compilation of SignaLink

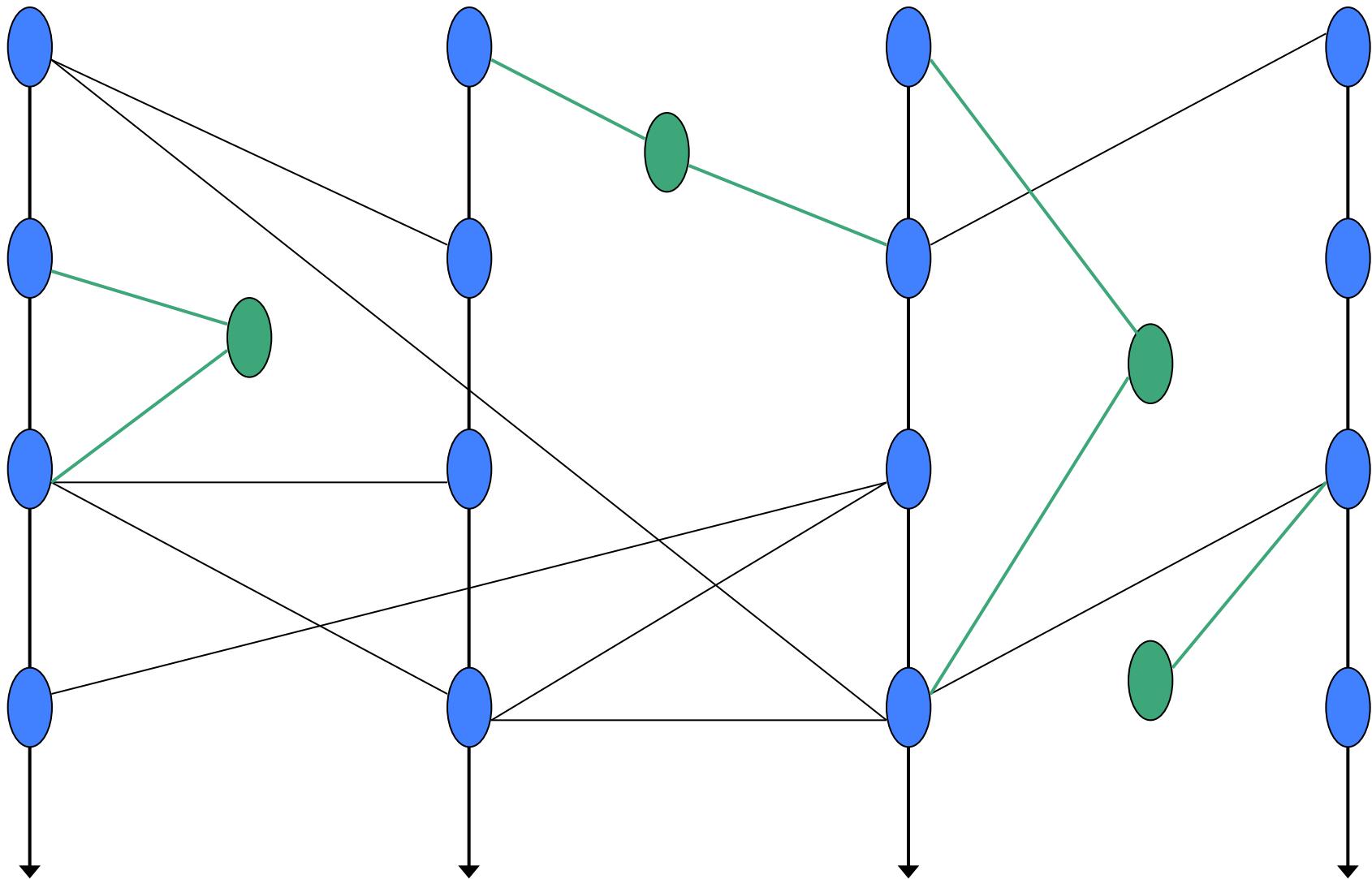


170 reviews, 771 research articles



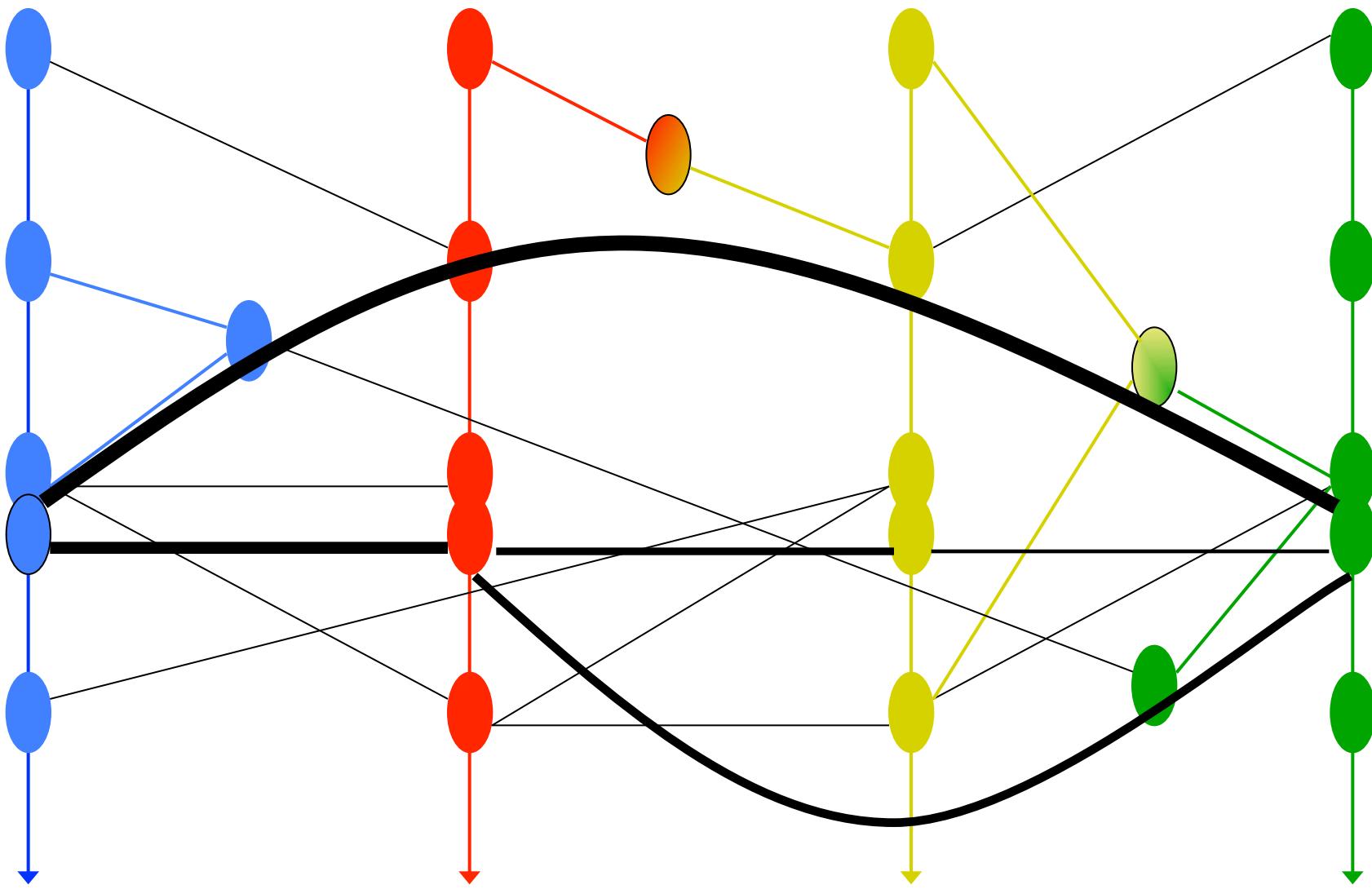
in

Network of signalling pathways

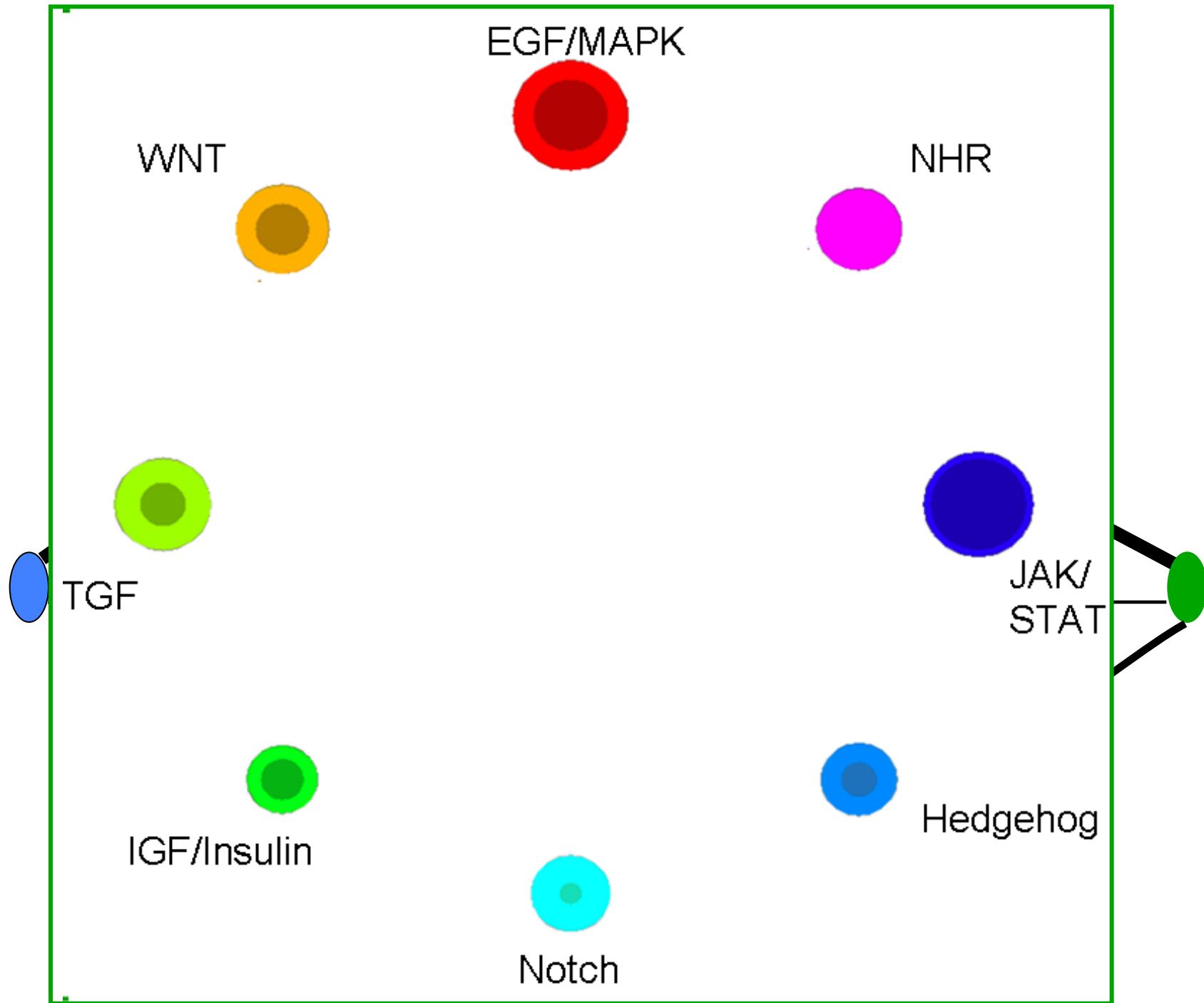


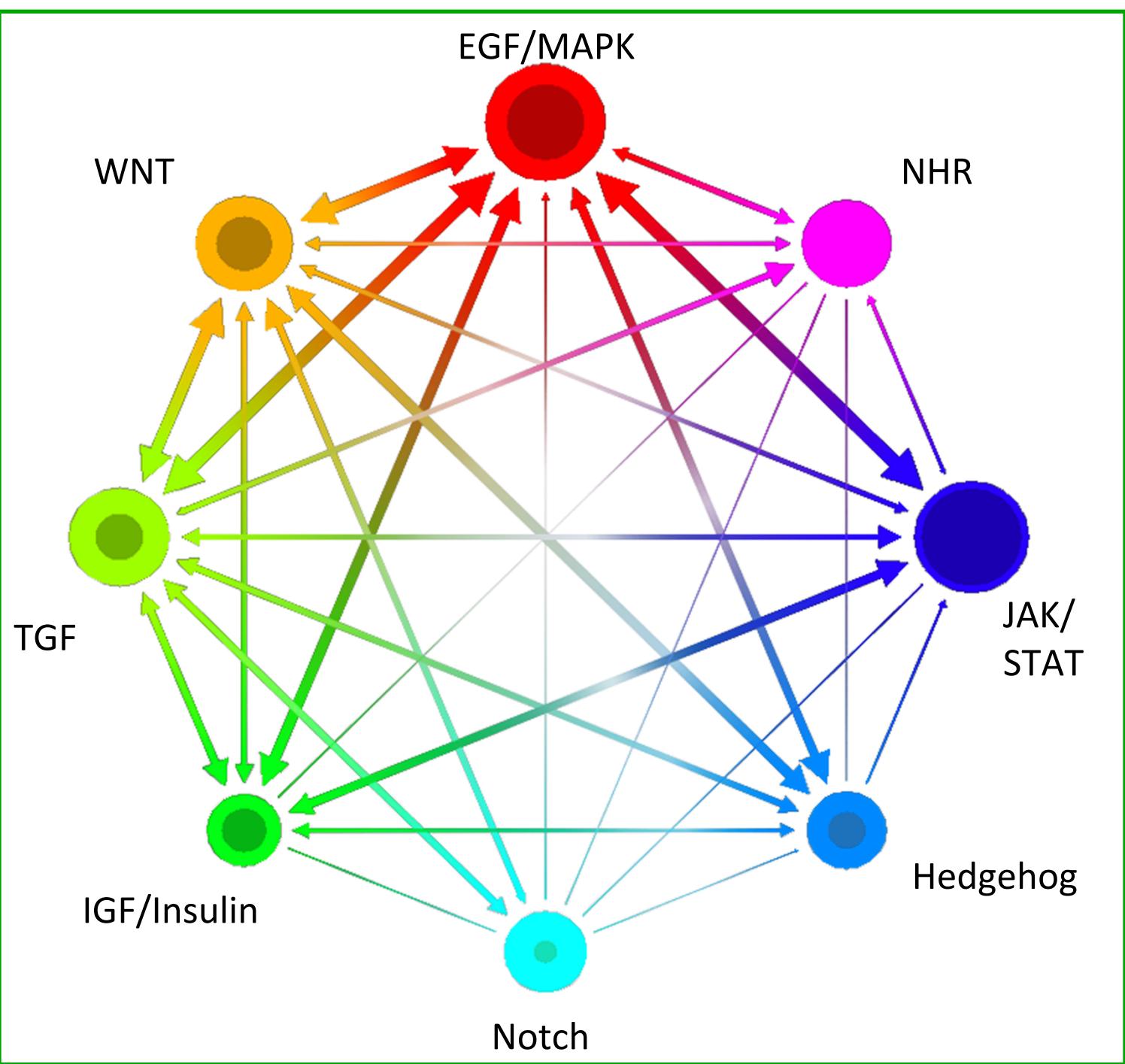
How can we visualize cross-talks?

Network of signalling pathways

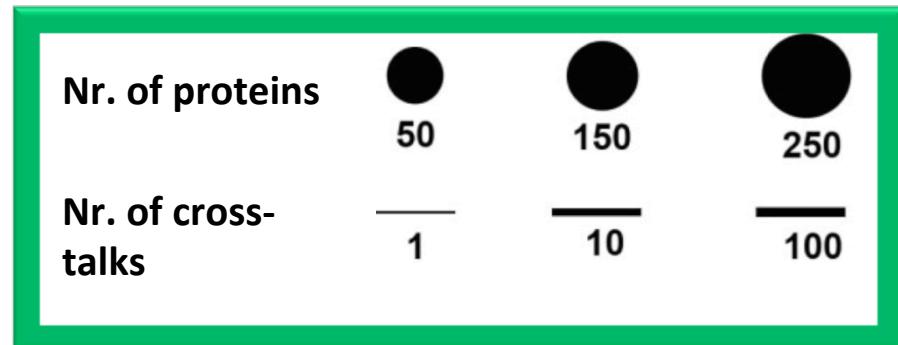
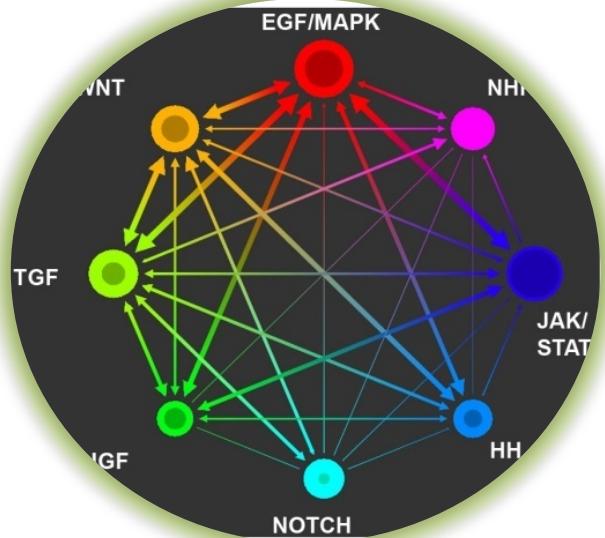
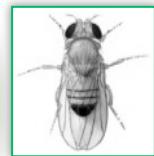
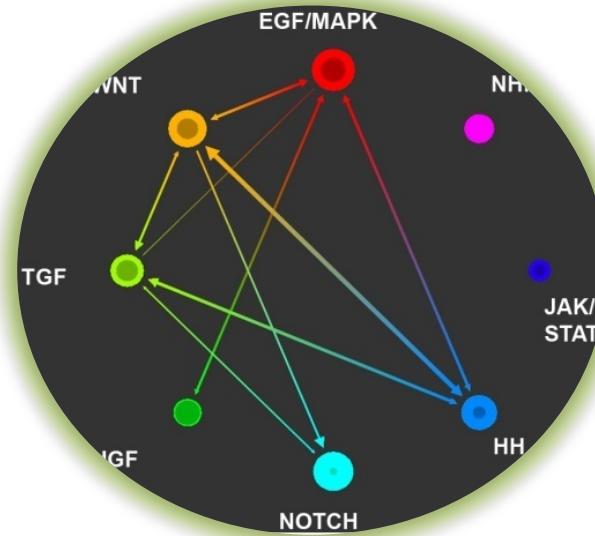
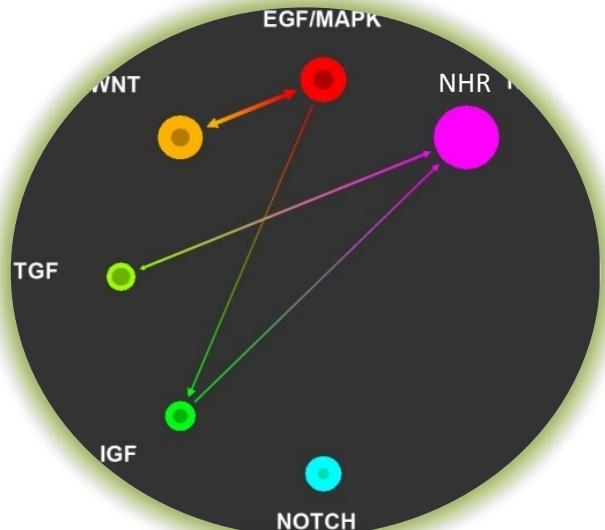


1) Color-coded protein-protein interaction network

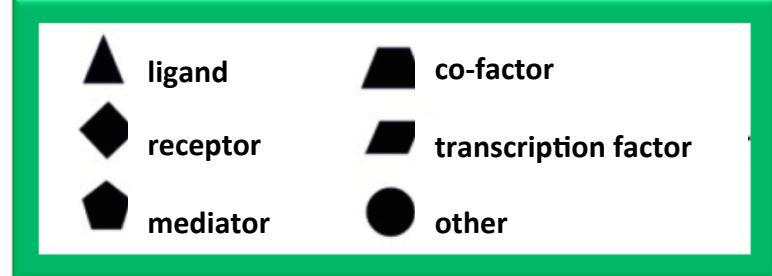
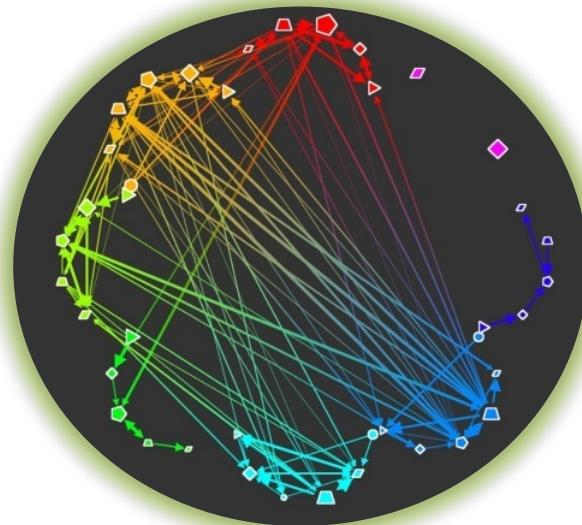
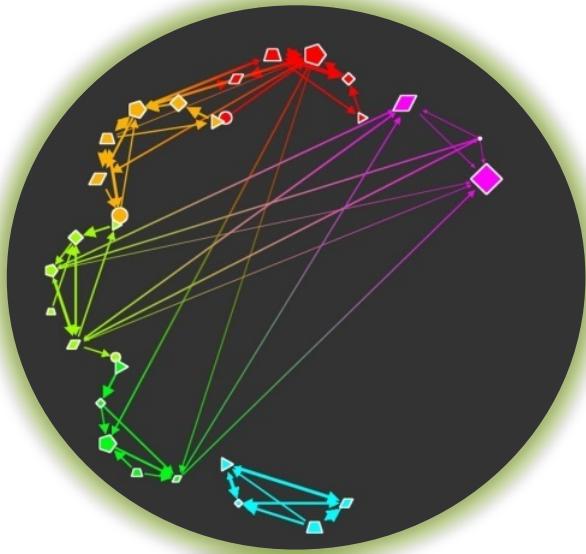




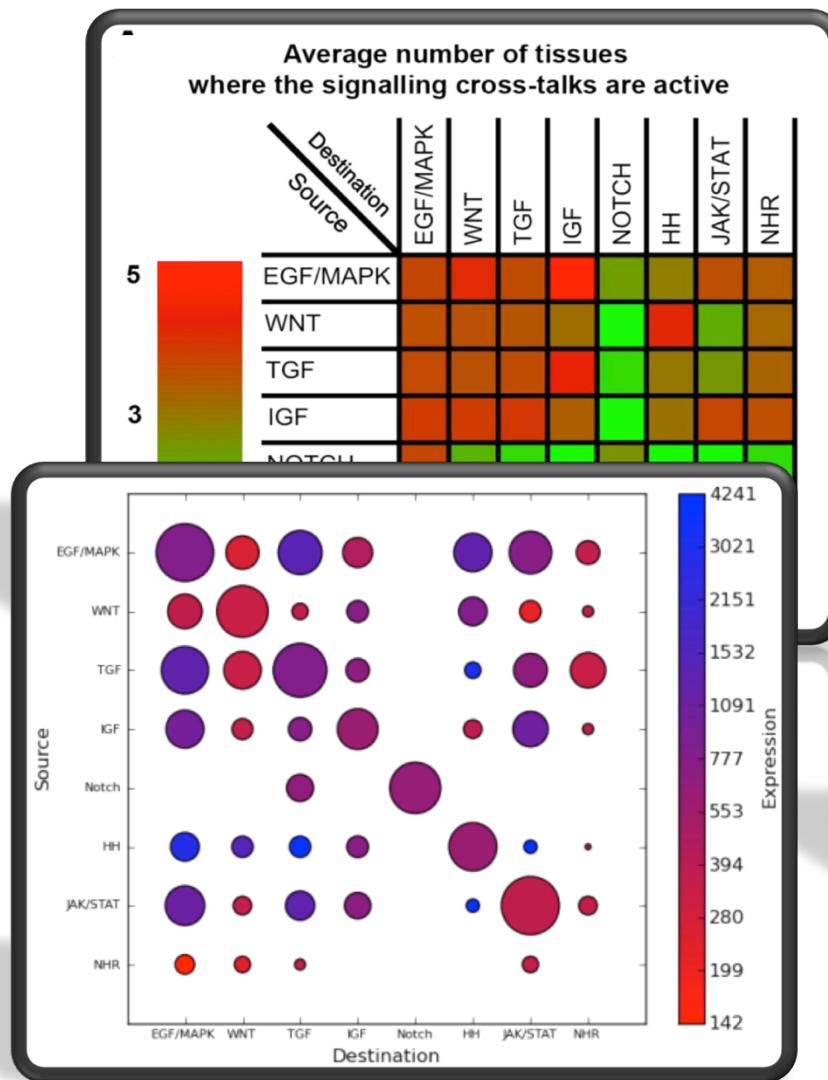
Cross-talk networks in 3 metazoans



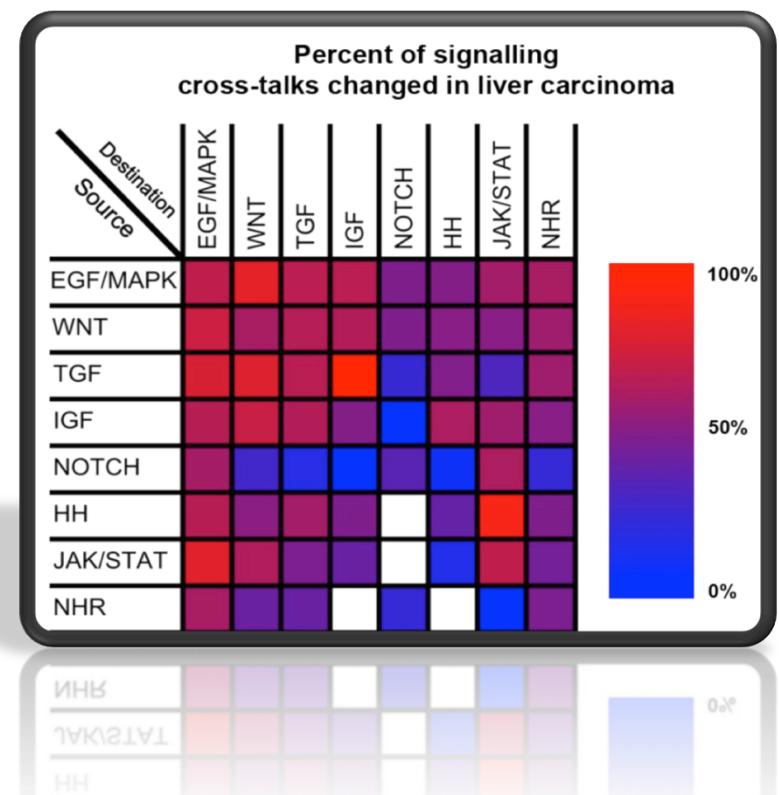
Detailed cross-talk networks in 3 metazoans



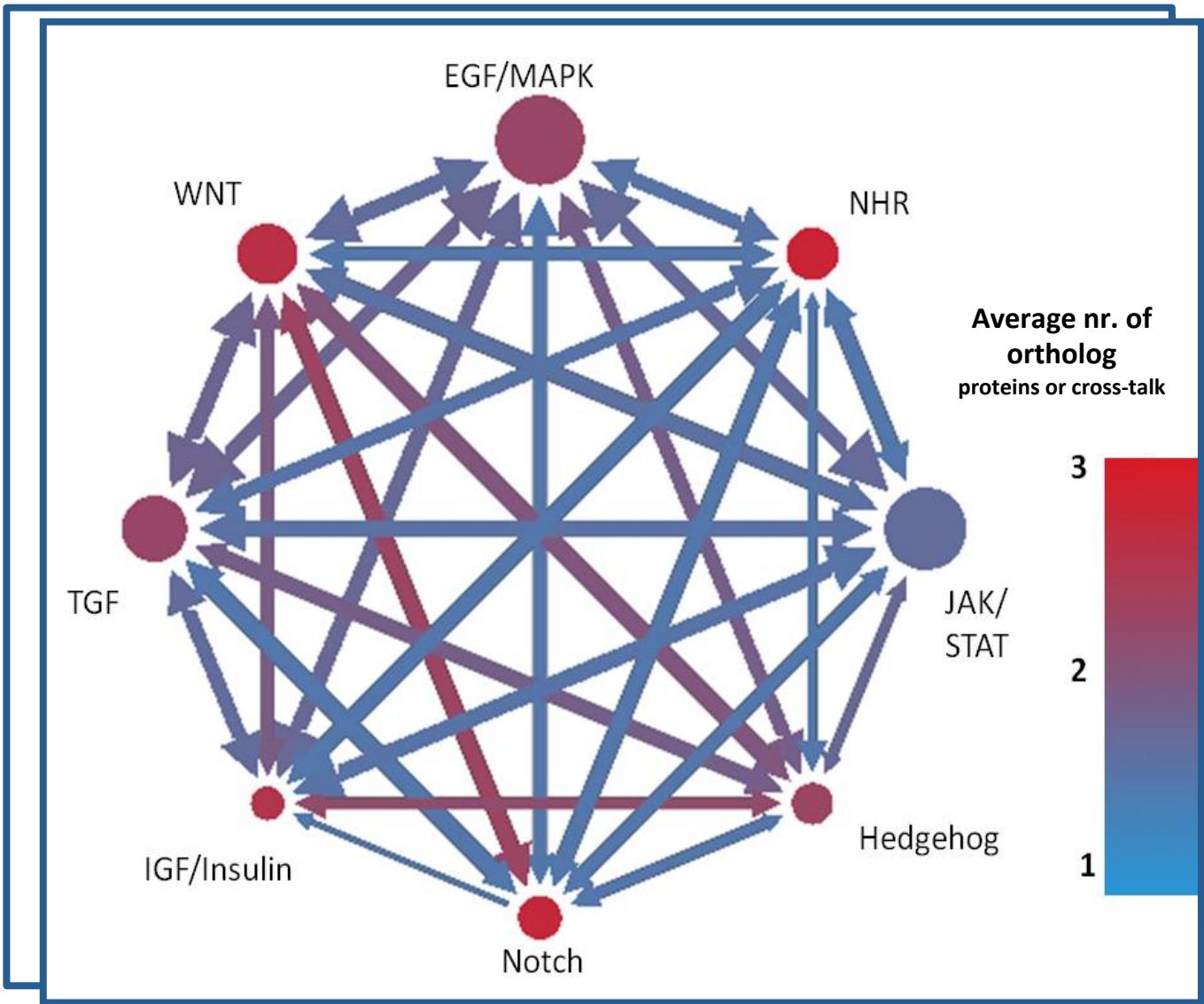
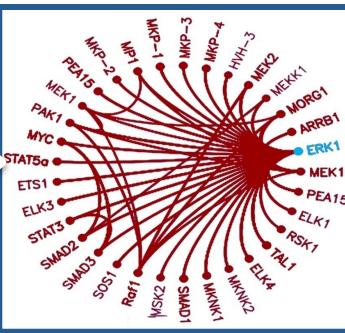
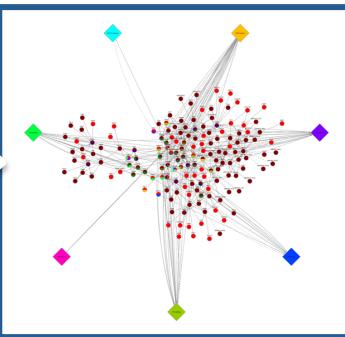
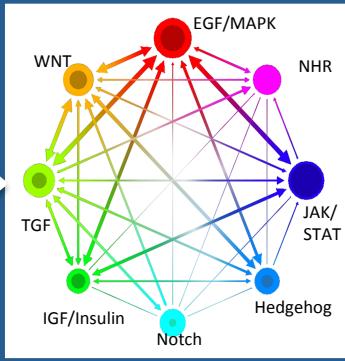
Visualizing cross-talk expressions



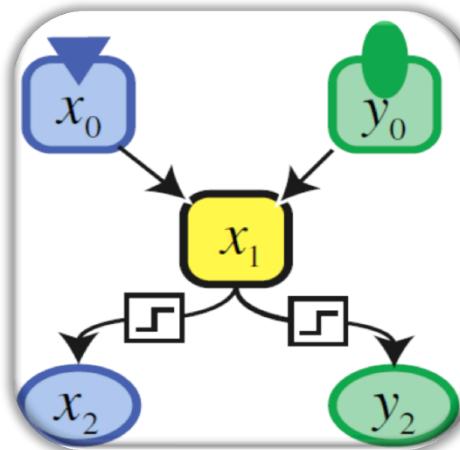
Identification of tissue-, disease-, and cell-type specific cross-talks



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

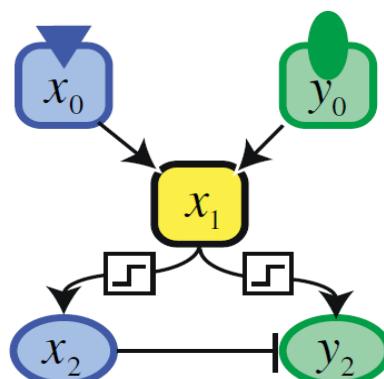
Combinatorial problems

- Protein isoforms
- Alternative splicing
- Post-translational sites

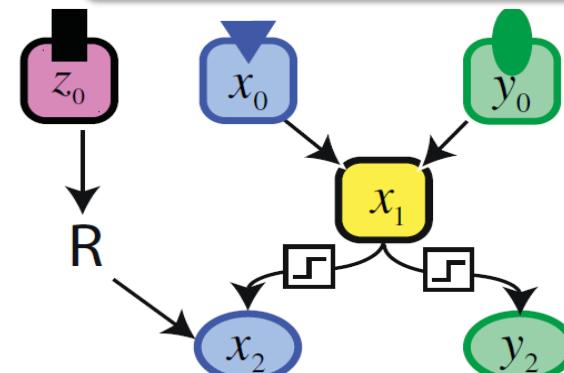
How cross-talks and signalling flow are regulated?

Insulating mechanisms to regulate signalling flow

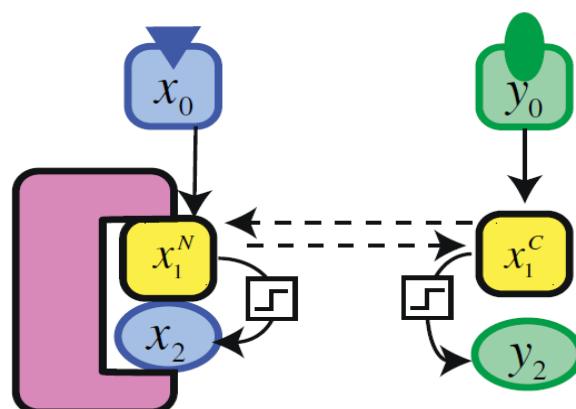
Cross-pathway inhibition



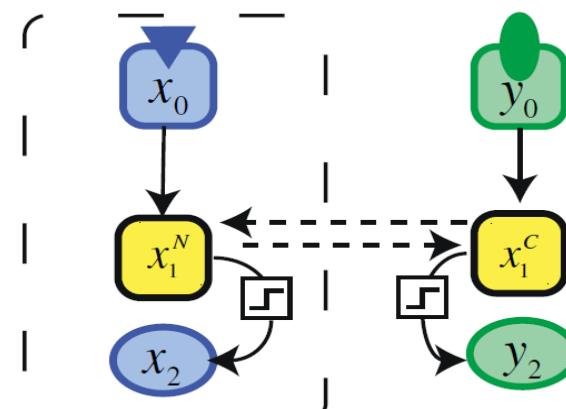
Combination of signalling



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)

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

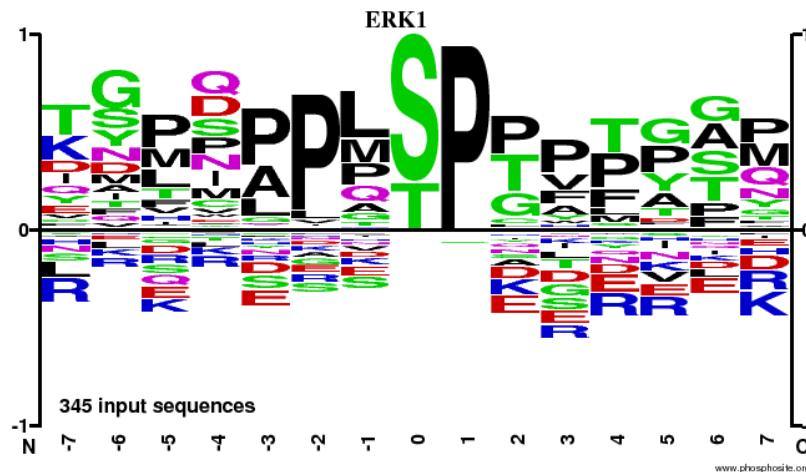


Specificity?



Active site

Target motifs



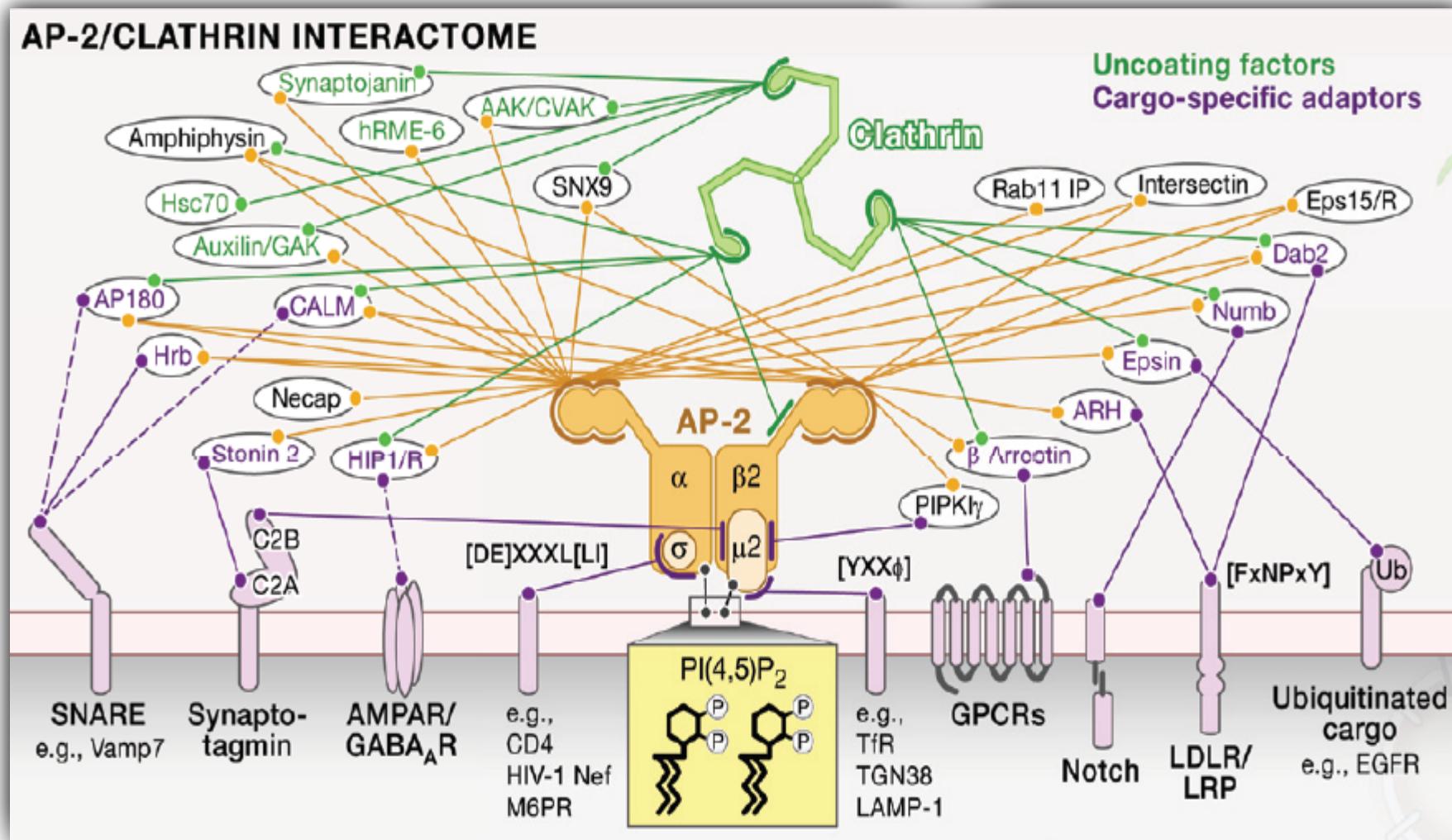
Docking motifs

Docking interactions

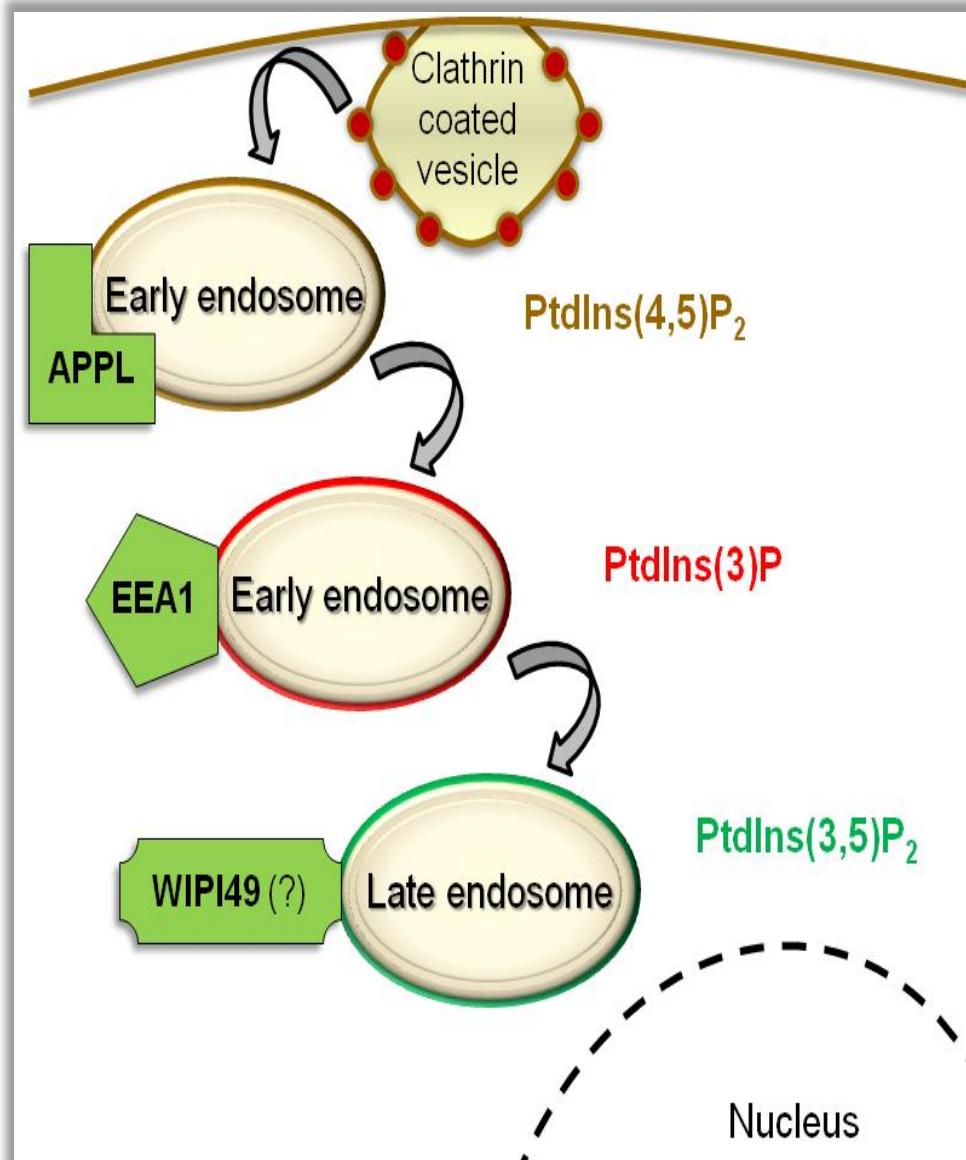
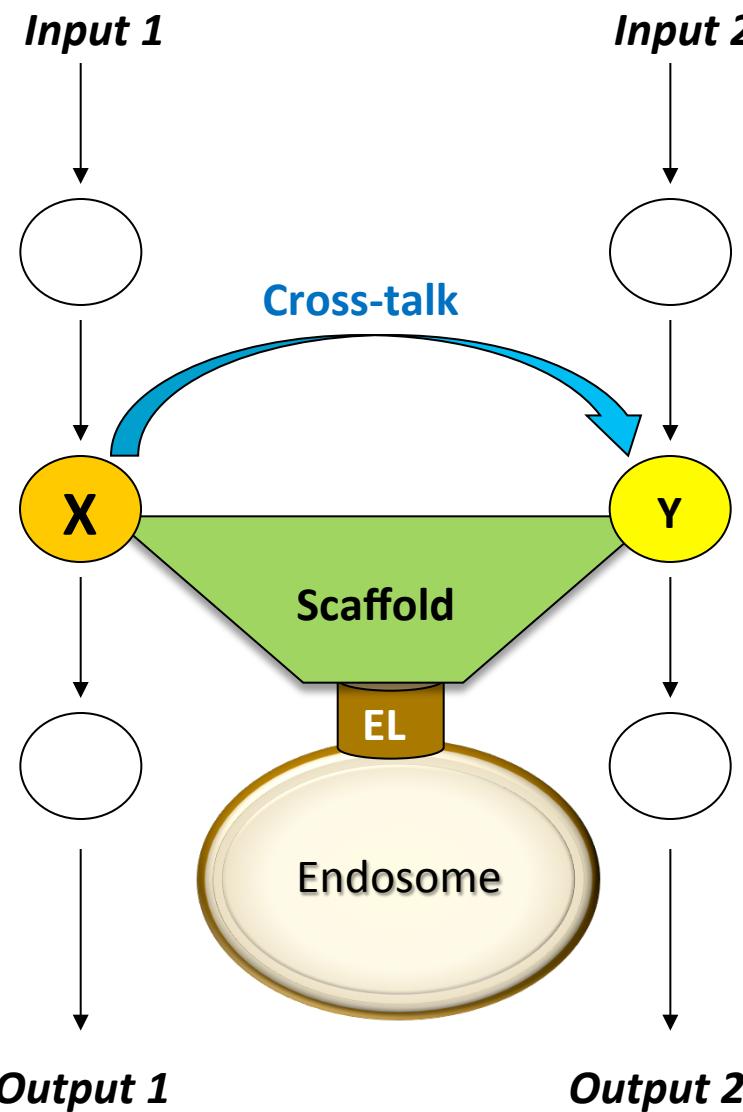
Sequence	Start	End	Subsequence
NCF1_HUMAN	363	368	TQRSKPQPAVPPRPSADLIL
POLG_HCVJA	2323	2328	LPSTKAFPIPPPFRKRTVVL
FAK1_MOUSE	750	755	SGGSDEAFPKPSRPGYPSPR
DYN1_HUMAN	833	838	FGGPPOVPSRPNDAPPGVPS
P85A_HUMAN	305	313	RQAPAPLPEKPKPTTVANN
P85A_HUMAN	305	310	WNERQFAPALPPKPKPTTV
RPGF1_HUMAN	284	289	VVDNSPPPALPPKKRQSAPS
PTN22_MOUSE	614	619	RTDDEIFPPLPERTPESFIV
NEF_HV1BR	72	77	EVGFPTVQVPLRPMTYKAA
PAK1_RAT	13	18	LDVQDKPAPPMRNTSTMIG
SOS1_HUMAN	1152	1157	DEVFPVPFPVPRRFESAPA

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

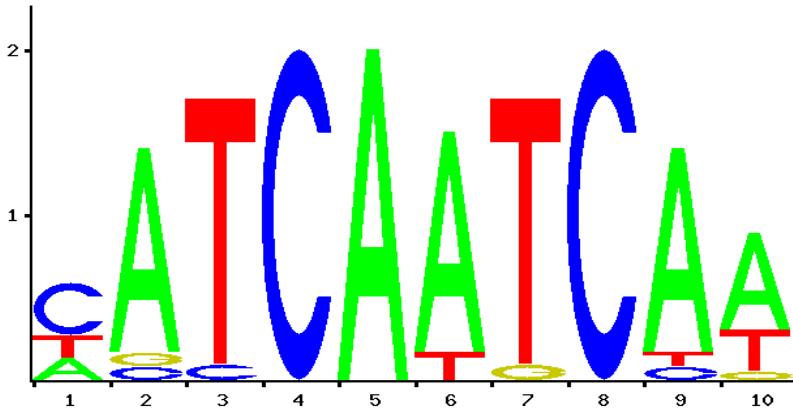
- Activation or inhibition of specific gene expression by transcription factor – transcription factor binding site connection
- JASPAR (<http://jaspar.genereg.net>)
- TFe (<http://cisreg.ca/cgi-bin/tfe/home.pl>)
- HTRIdb (<http://www.lbpc.ibb.unesp.br/htri>)
- PAZAR (<http://pazar.info>)

Post-transcriptional regulation

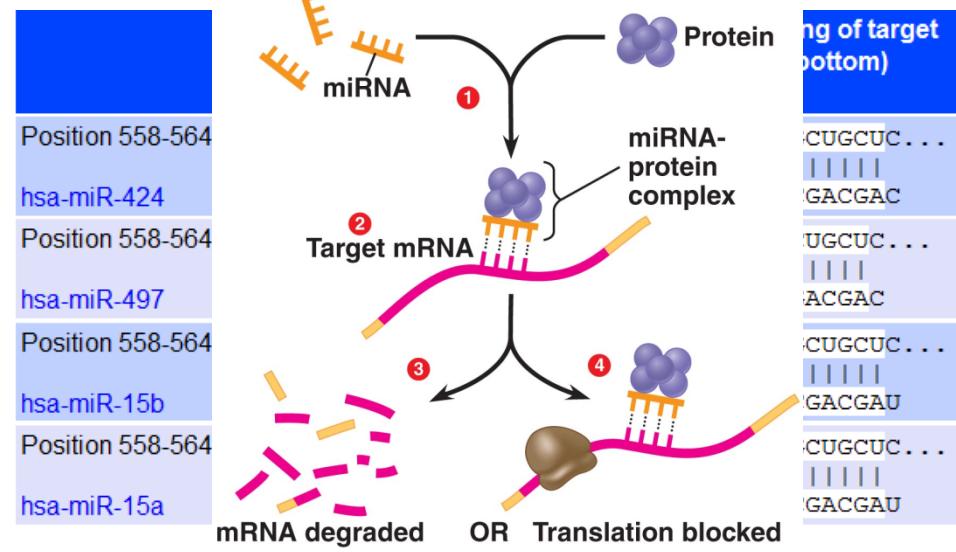
- Destruction or inhibition of specific mRNAs by miRNAs
- miRBase (<http://mirbase.org>)
- miRGen (<http://diana.cslab.ece.ntua.gr/mirgen>)
- miRecords (<http://mirecords.umn.edu/miRecords>)
- TarBase (<http://www.microrna.gr/tarbase>)

Specificity?

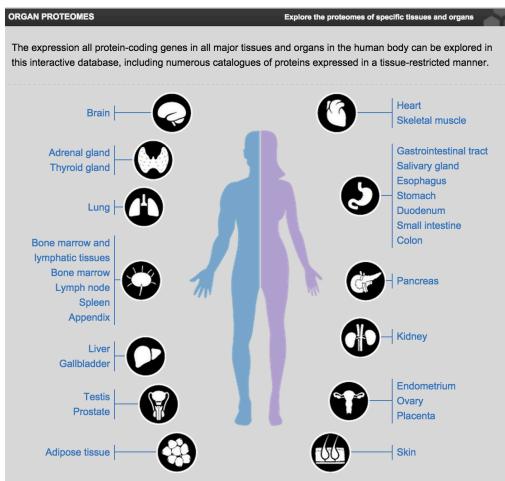
TF-TFBS target motifs



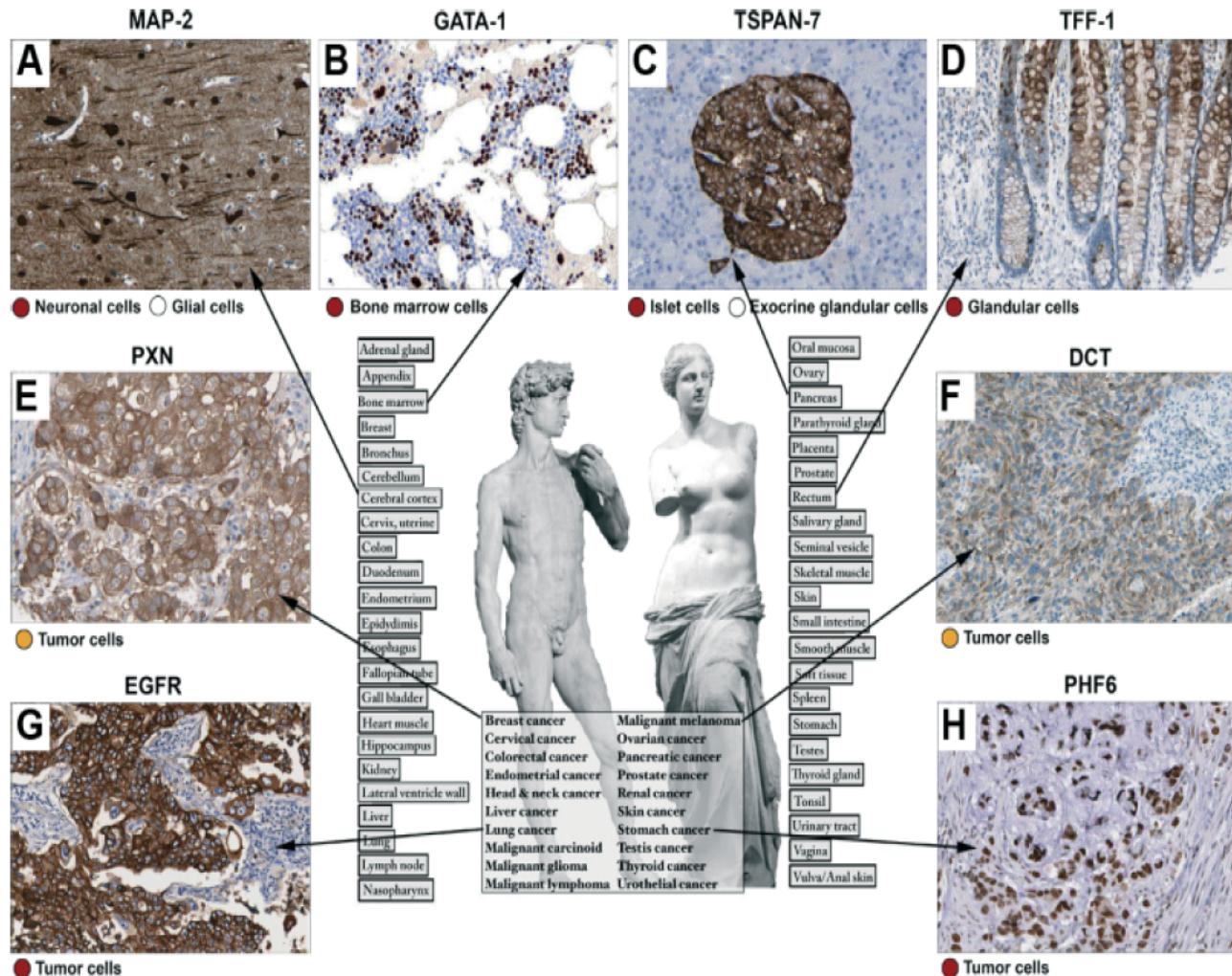
mRNA-miRNA binding



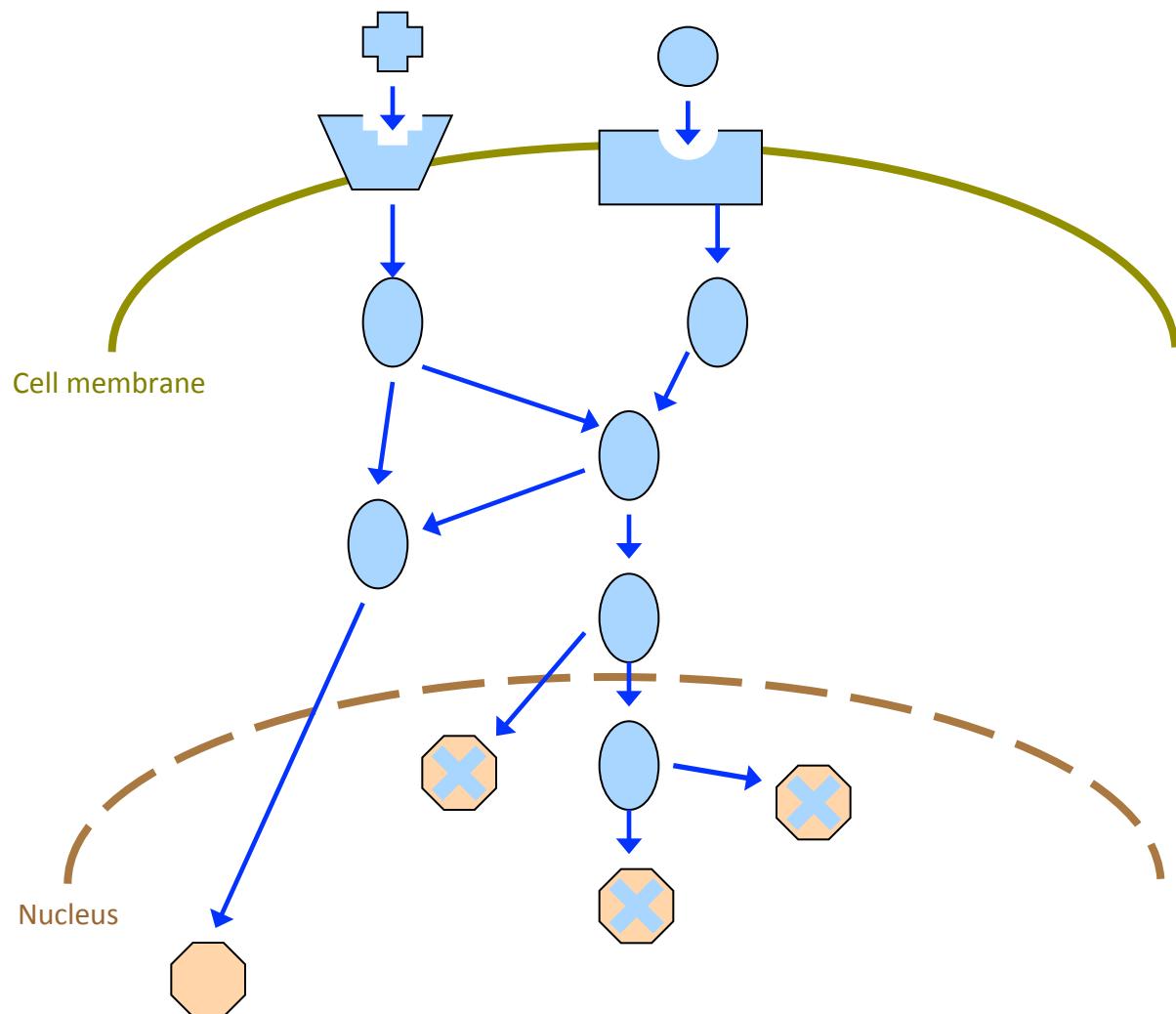
An important issue: tissue specificity



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



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



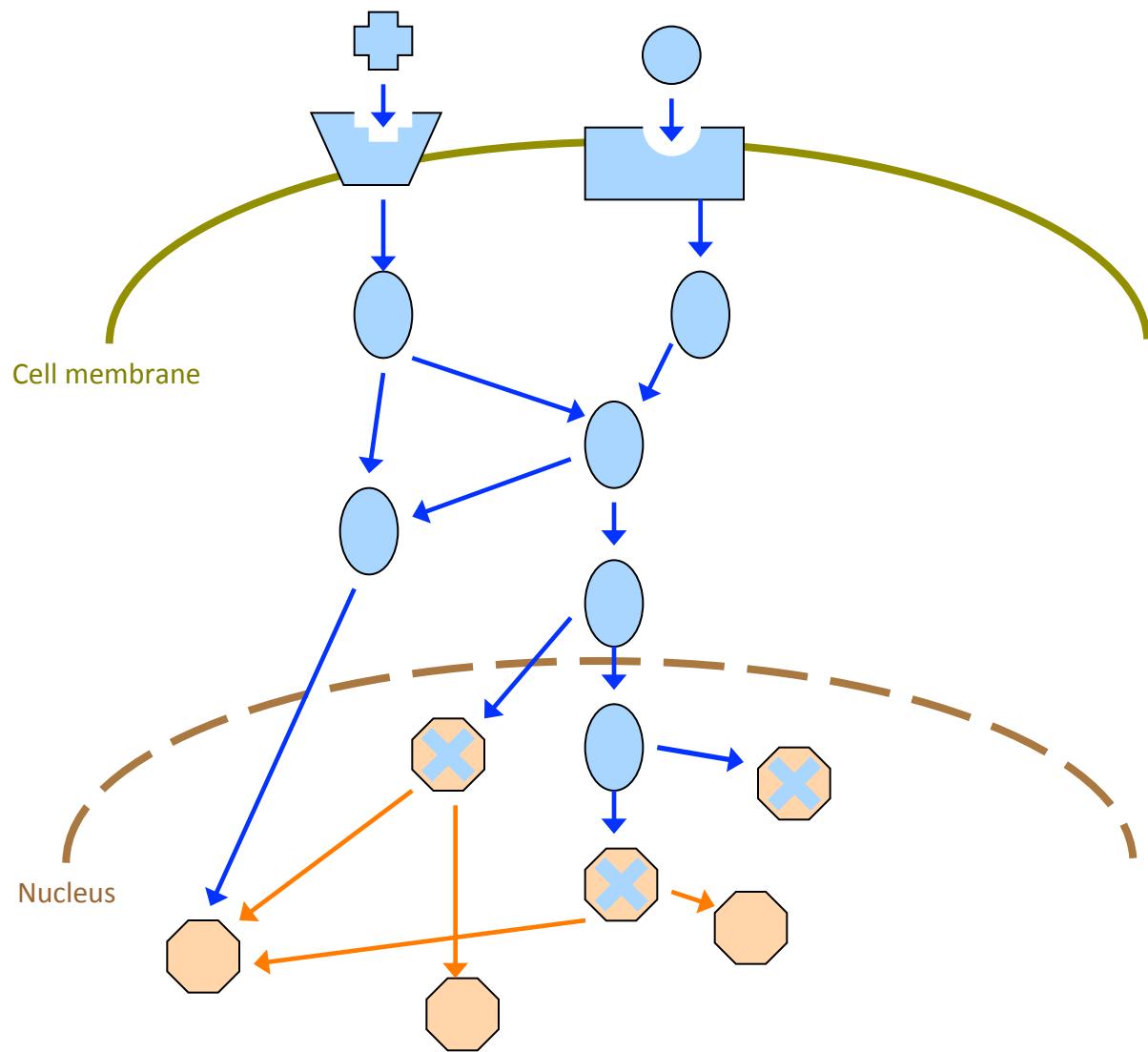
Legends:

- Signalling elements
- Transcription factors

Types of networks:



Network of
signalling
pathways

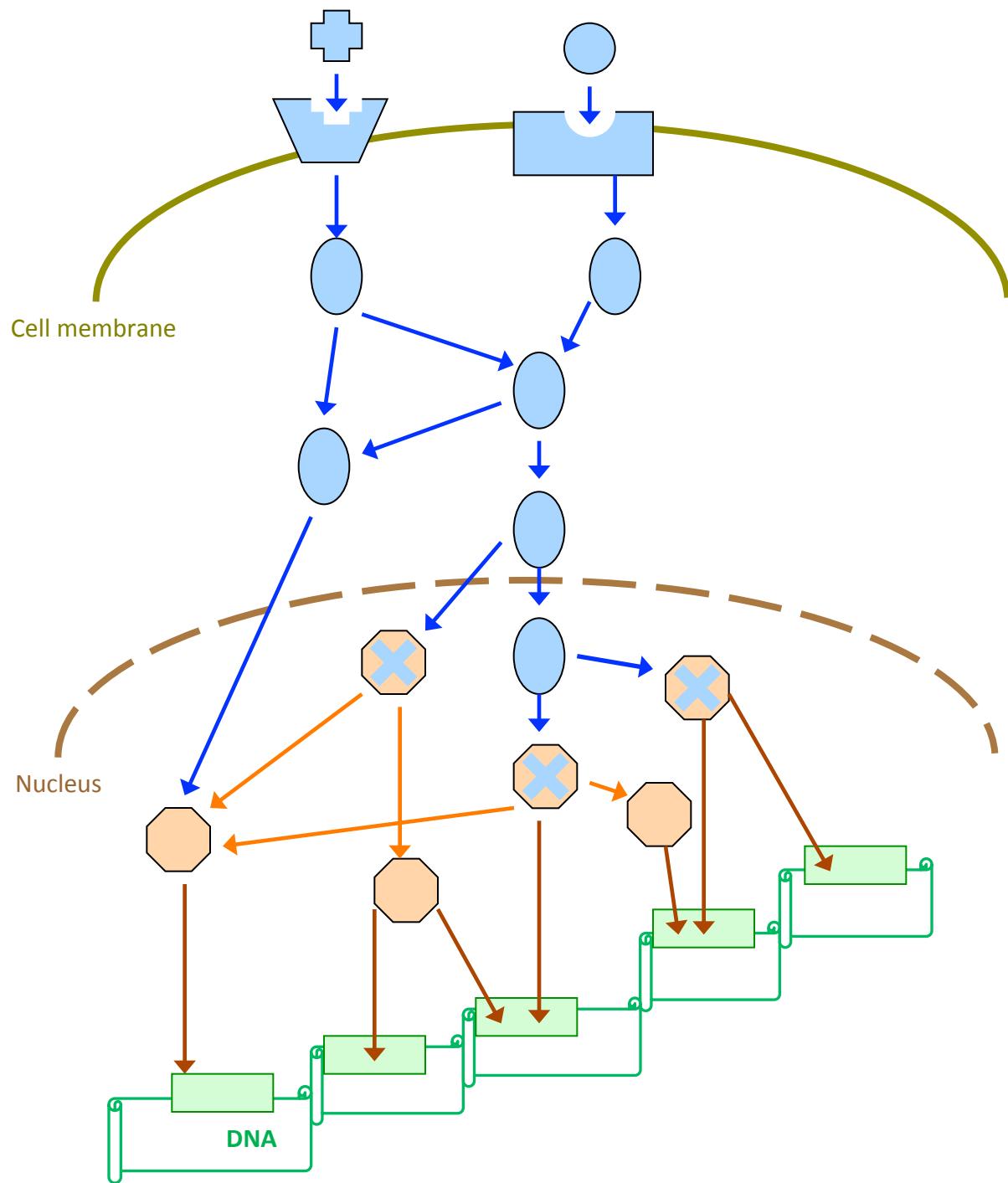


Legends:

- Signalling elements
- Transcription factors

Types of networks:

- Network of signalling pathways
- Network of transcription factors

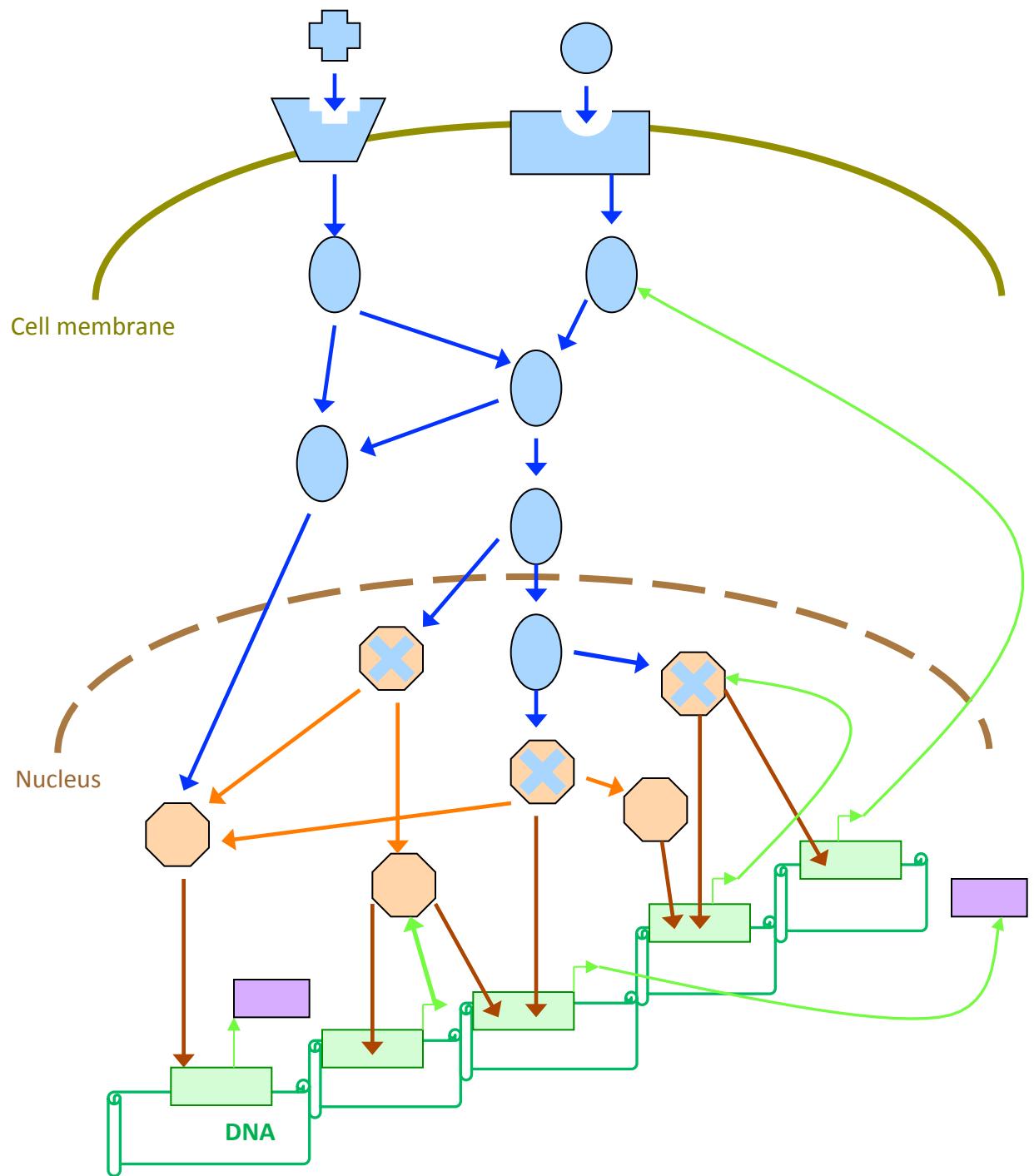


Legends:

- [Blue square] Signalling elements
- [Orange square] Transcription factors
- [Green rectangle] Promoters

Types of networks:

- [Blue square → Blue square] Network of signalling pathways
- [Orange square → Orange square] Network of transcription factors
- [Orange square → Green rectangle] Network of transcription factors' binding sites

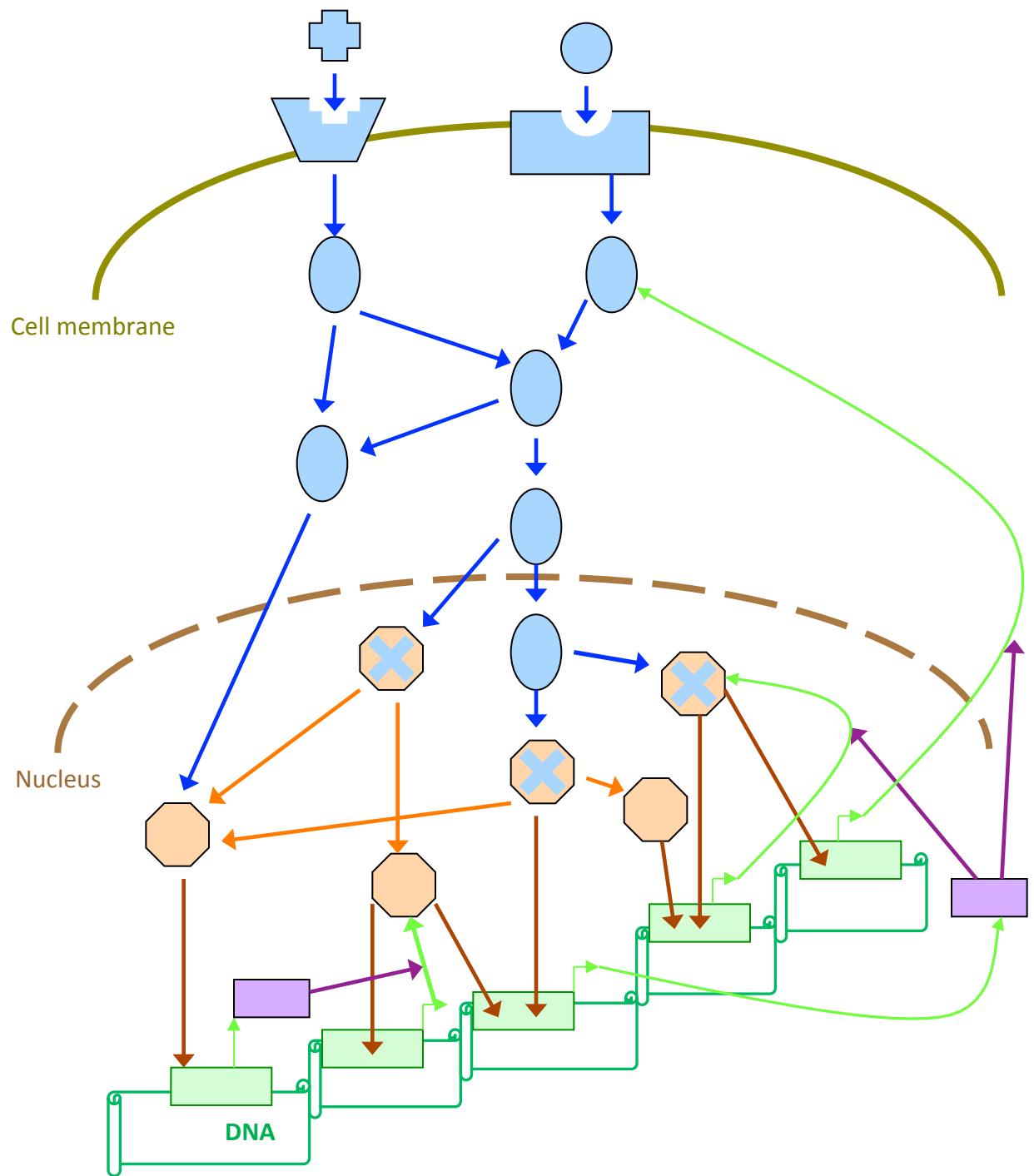


Legends:

- [Blue rectangle] Signalling elements
- [Orange rectangle] Transcription factors
- [Green rectangle] Promoters
- [Purple rectangle] miRNAs

Types of networks:

- [Blue rectangle] → [Blue rectangle] Network of signalling pathways
- [Orange rectangle] → [Orange rectangle] Network of transcription factors
- [Orange rectangle] → [Green rectangle] Network of transcription factors' binding sites
- [Green rectangle] → [Blue rectangle / Orange rectangle / Purple rectangle] Network of promoters and their transcripts



Legends:

- [Blue rectangle] Signalling elements
- [Orange rectangle] Transcription factors
- [Green rectangle] Promoters
- [Purple rectangle] miRNAs

Types of networks:

- [Blue rectangle] → [Blue rectangle] Network of signalling pathways
- [Orange rectangle] → [Orange rectangle] Network of transcription factors
- [Orange rectangle] → [Green rectangle] Network of transcription factors' binding sites
- [Green rectangle] → [Blue rectangle] Network of promoters and their transcripts
- [Purple rectangle] → [Orange rectangle] Network of miRNAs and transcripts

- Reactome
- SignaLink
- ConsensusPathDB
- ...

• Fantom4

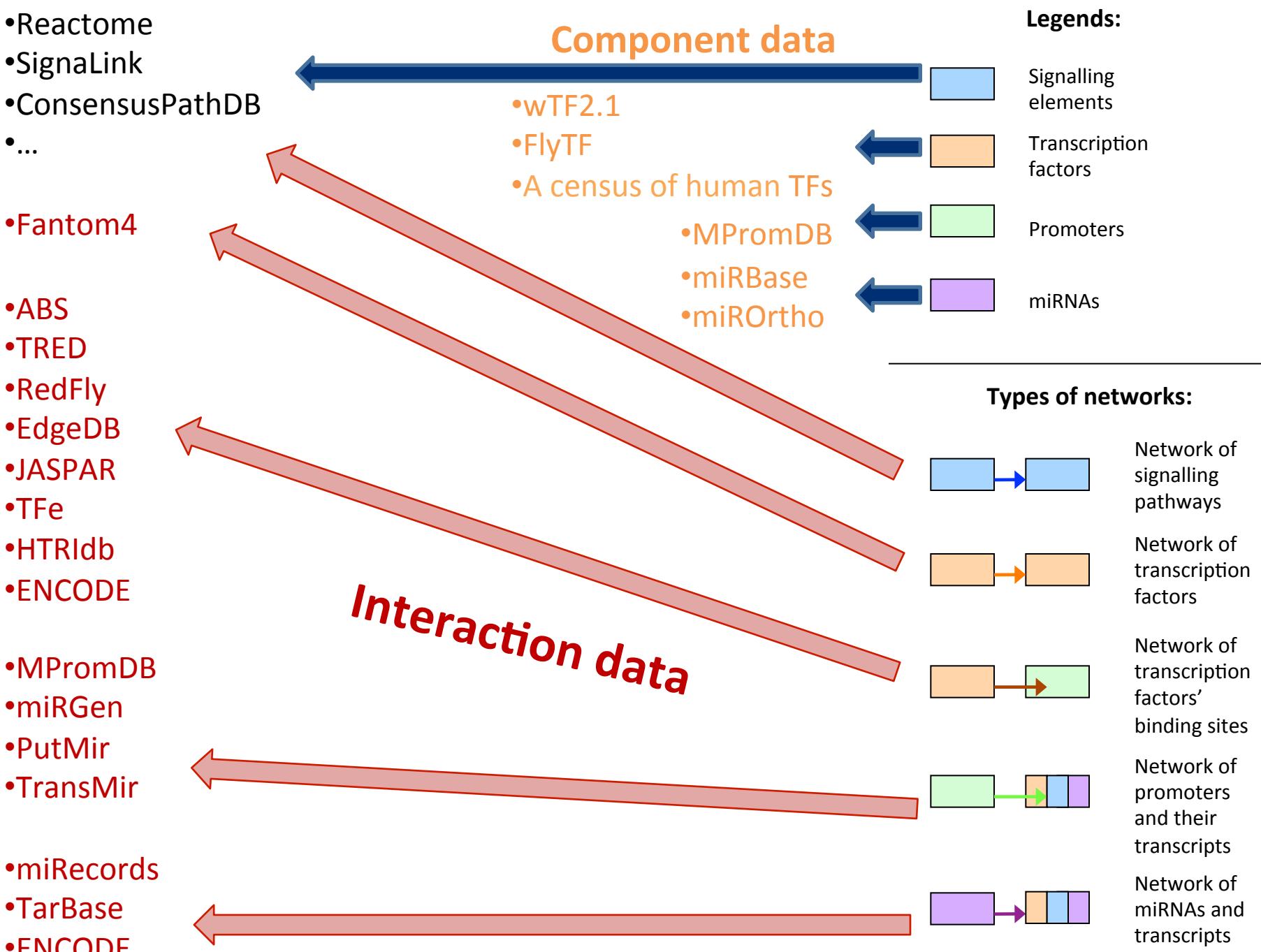
- ABS
- TRED
- RedFly
- EdgeDB
- JASPAR
- TFe
- HTREdb
- ENCODE

- MPromDB
- miRGen
- PutMir
- TransMir

- miRecords
- TarBase
- ENCODE

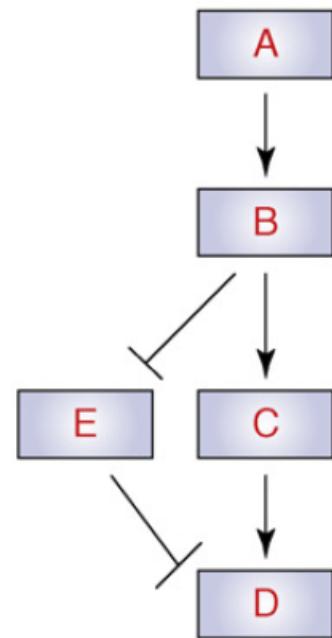
Component data

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

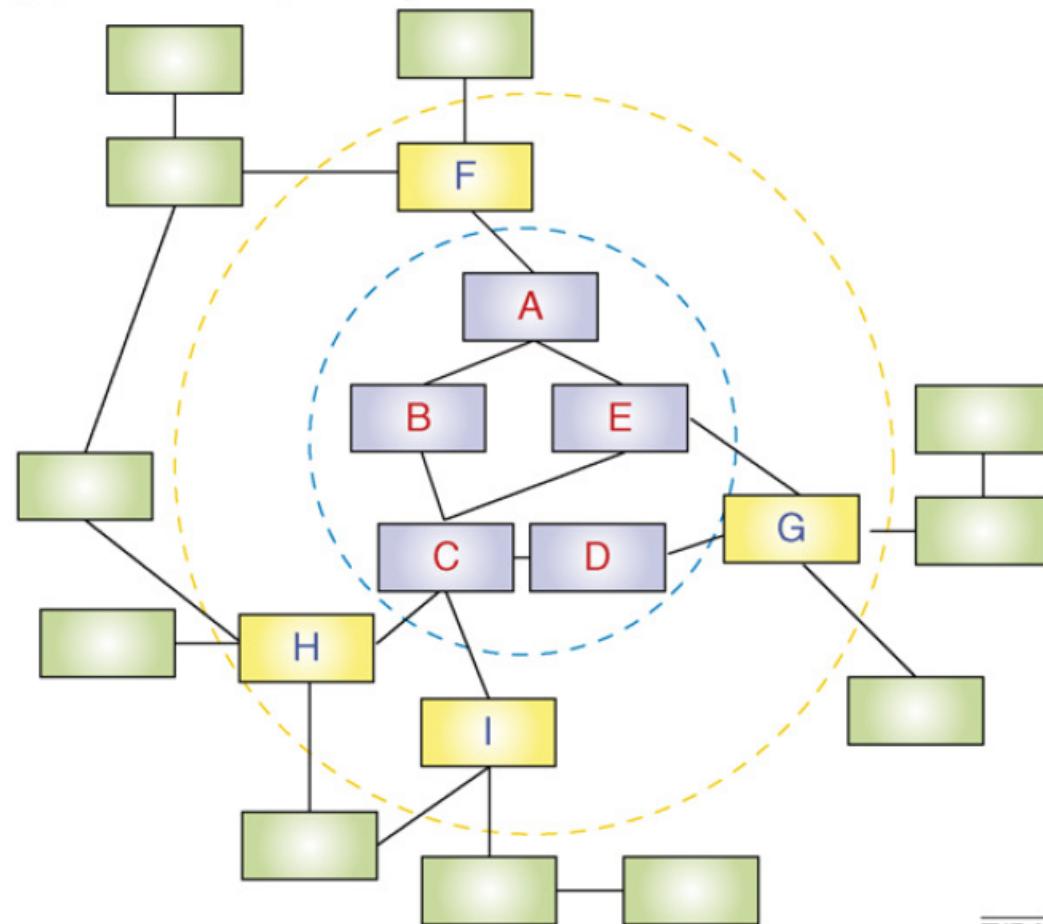


Embedding pathways into interaction networks

(a) Classical pathway



(b) Embedded pathway



TiBS

Where can I find PPIs to
connect with my pathway?

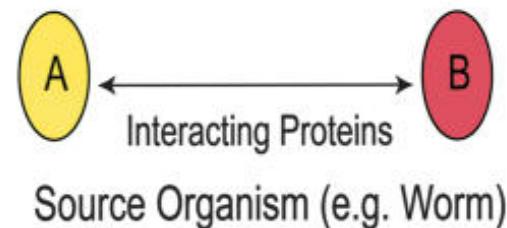
The power of prediction

Applying sequence and structural information
to predict novel functions / connections

Protein-protein interactions

Predicted interactions

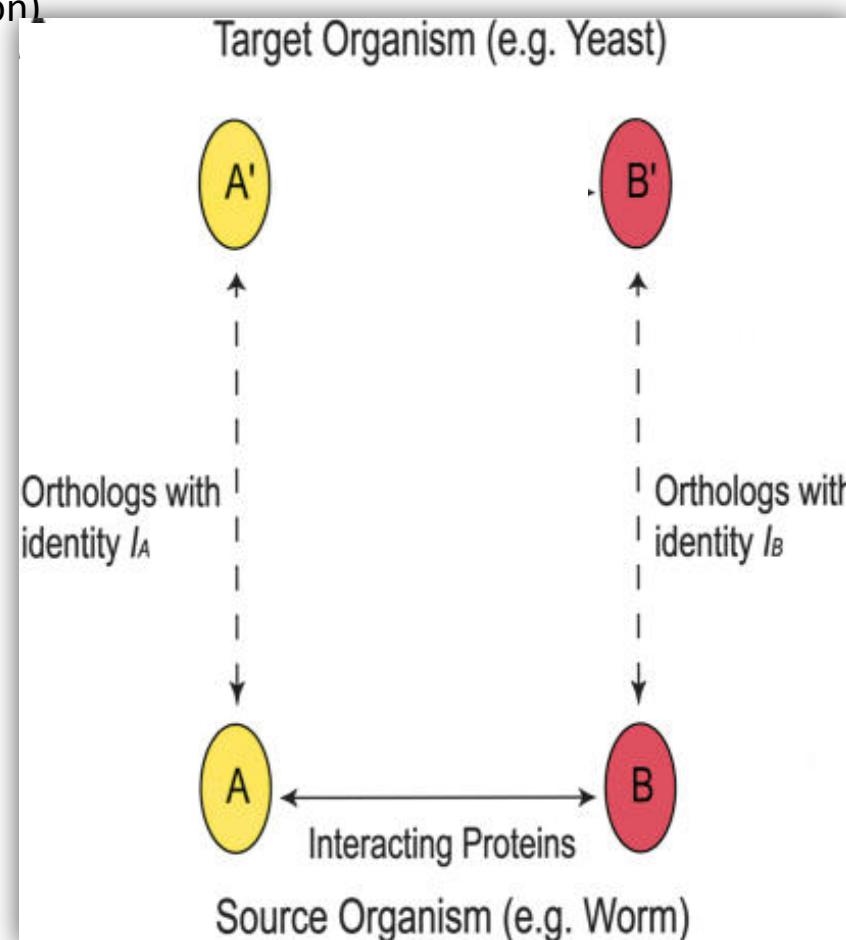
- Homology/orthology-based (interologs)
- Domain-motifs based (directed)
- Domain-domain based (interaction & direction)



Protein-protein interactions

Predicted interactions

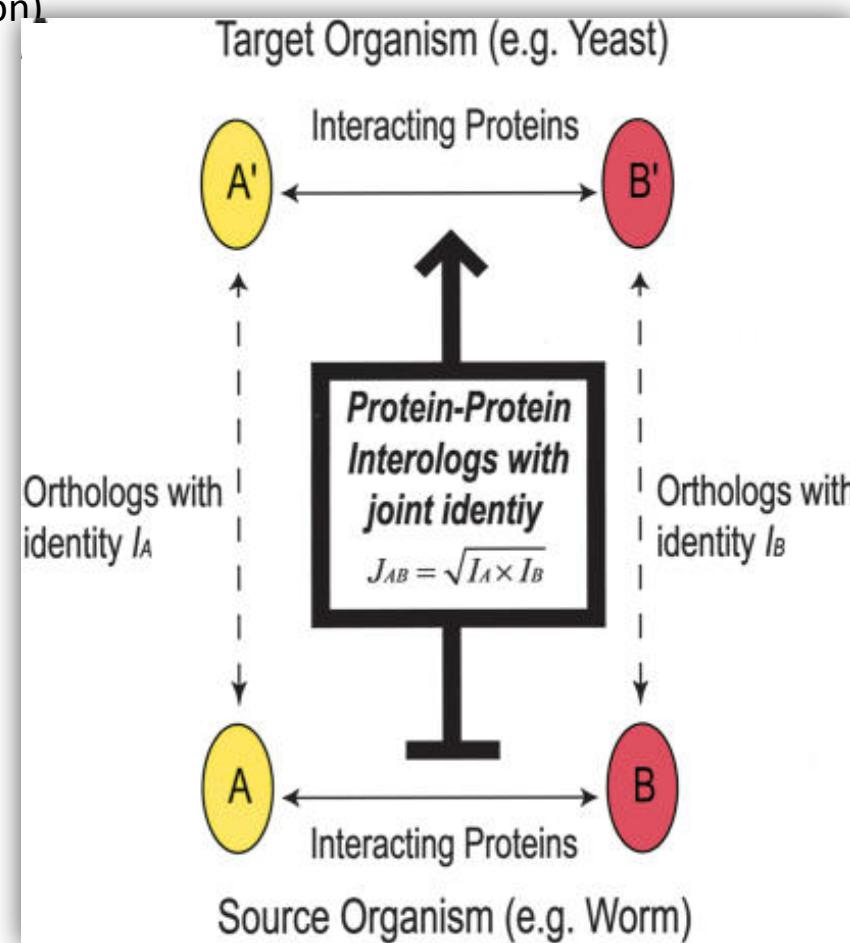
- Homology/orthology-based (interologs)
- Domain-motifs based (directed)
- Domain-domain based (interaction & direction)



Protein-protein interactions

Predicted interactions

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



Protein-protein interactions

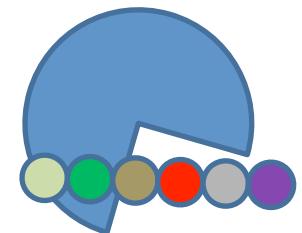
Predicted interactions

- 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



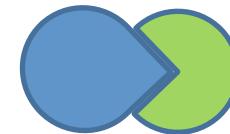
**Predicted PPI based on
domain-motif interaction**

Protein-protein interactions

Predicted interactions

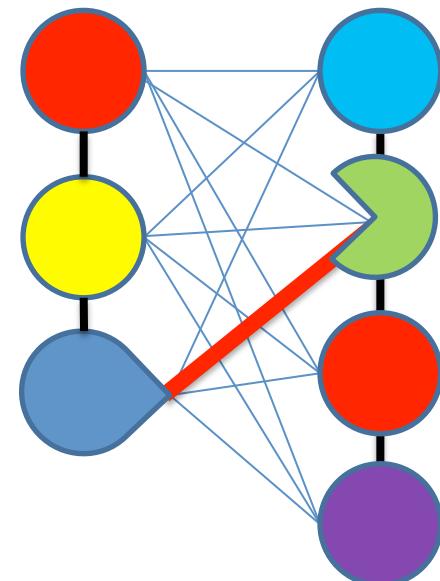
- Homology/orthology-based (interologs)
- Domain-motifs based (directed)
- **Domain-domain based (interaction & direction)**

Domain-domain
interaction data
(eg., DOMINE)



Protein-domain
composition
data (eg., PFAM)

Possible
domain
pairs

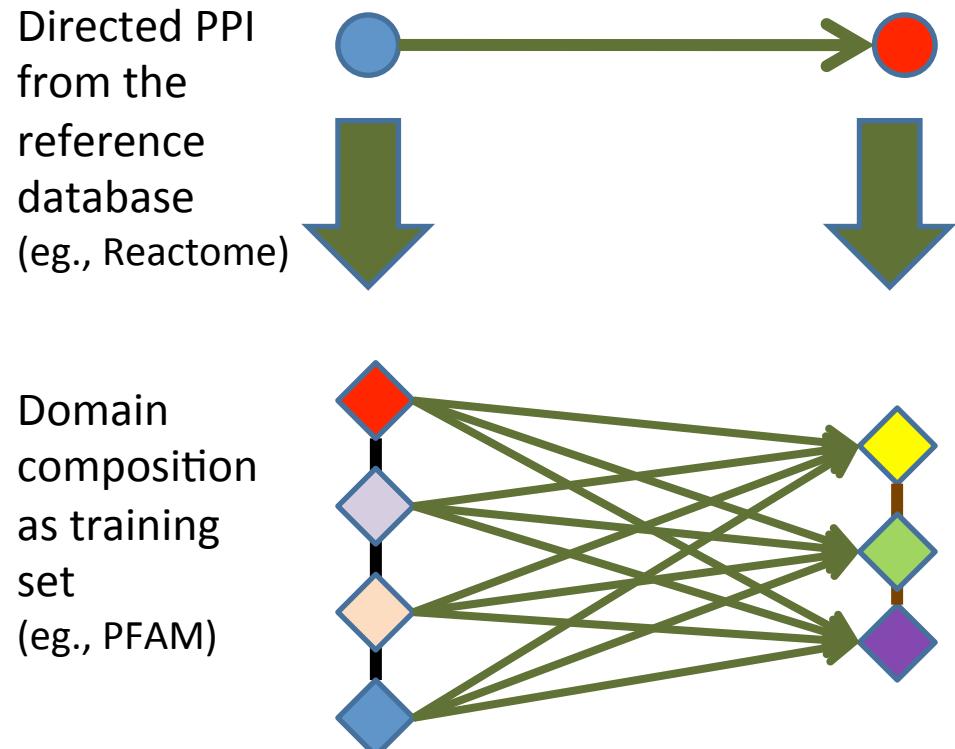


**Predicted PPI based on
domain-domain interaction**

Protein-protein interactions

Predicted interactions

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



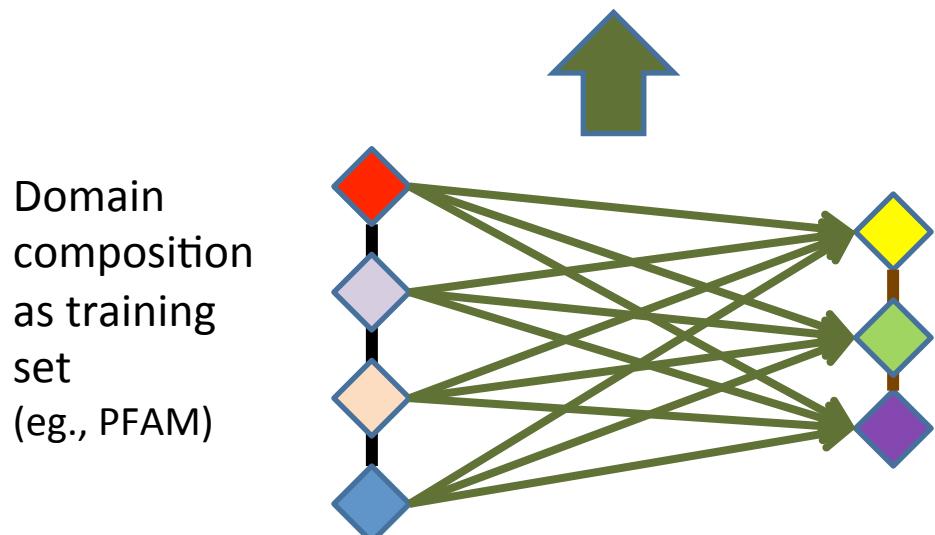
Liu *et al.*, MCP (2009) and
Rhodes *et al.*, Nature Biotechnology (2005)

Protein-protein interactions

Predicted interactions

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

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



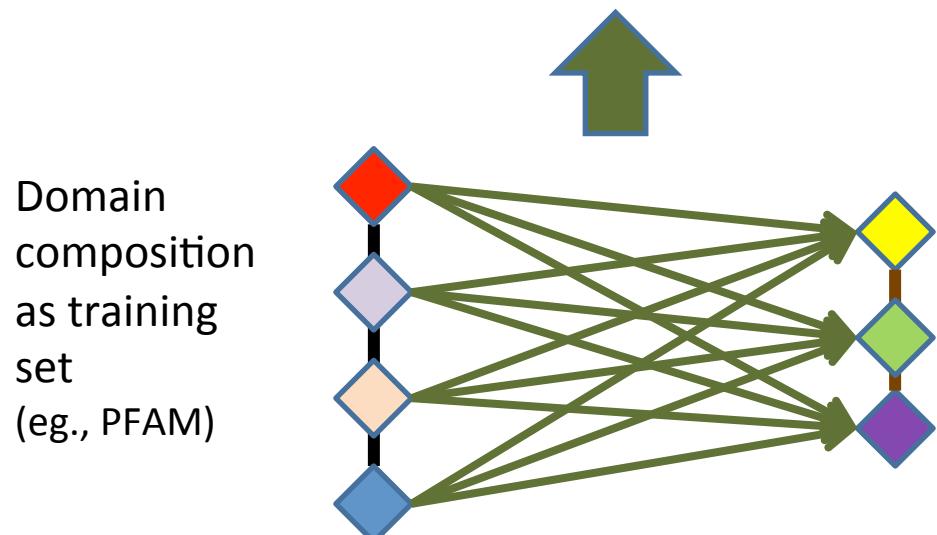
Liu *et al.*, MCP (2009) and
Rhodes *et al.*, Nature Biotechnology (2005)

Protein-protein interactions

Predicted interactions

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

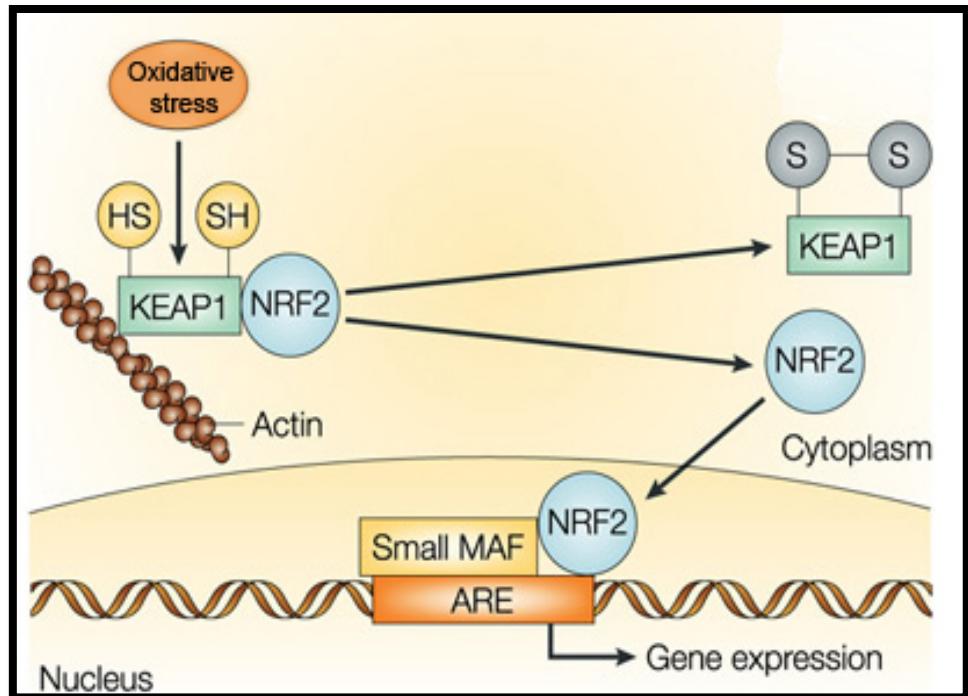
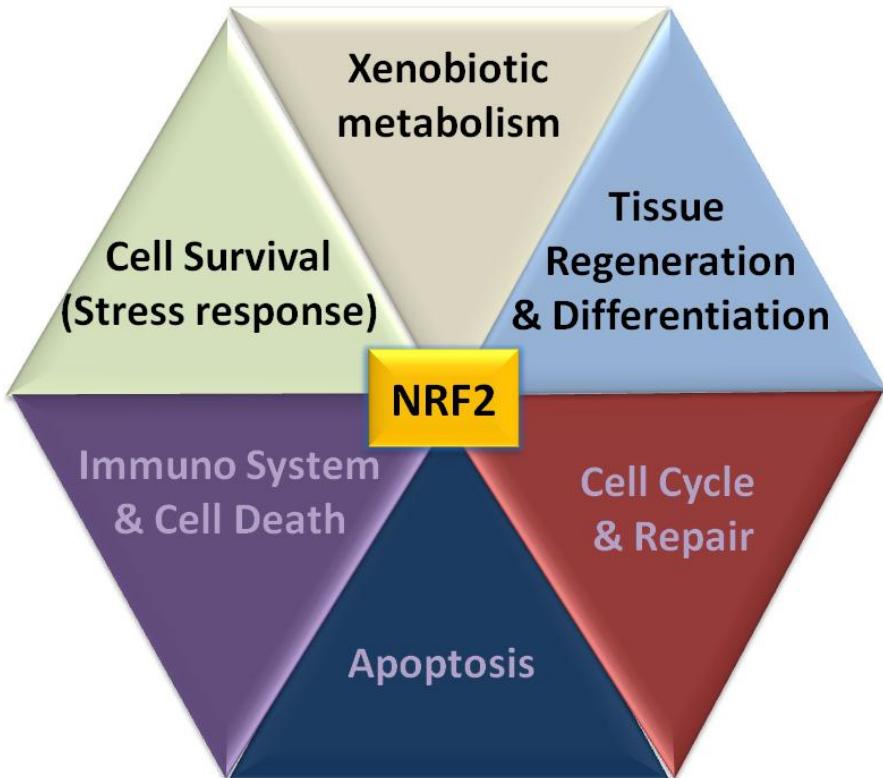
$$F(d_{mn}) = \frac{\Pr(d_m \rightarrow d_n) - \Pr(d_n \rightarrow d_m)}{\Pr(d_m) \times \Pr(d_n)}$$



Liu *et al.*, MCP (2009) and
Rhodes *et al.*, Nature Biotechnology (2005)

Two use cases
for data integration

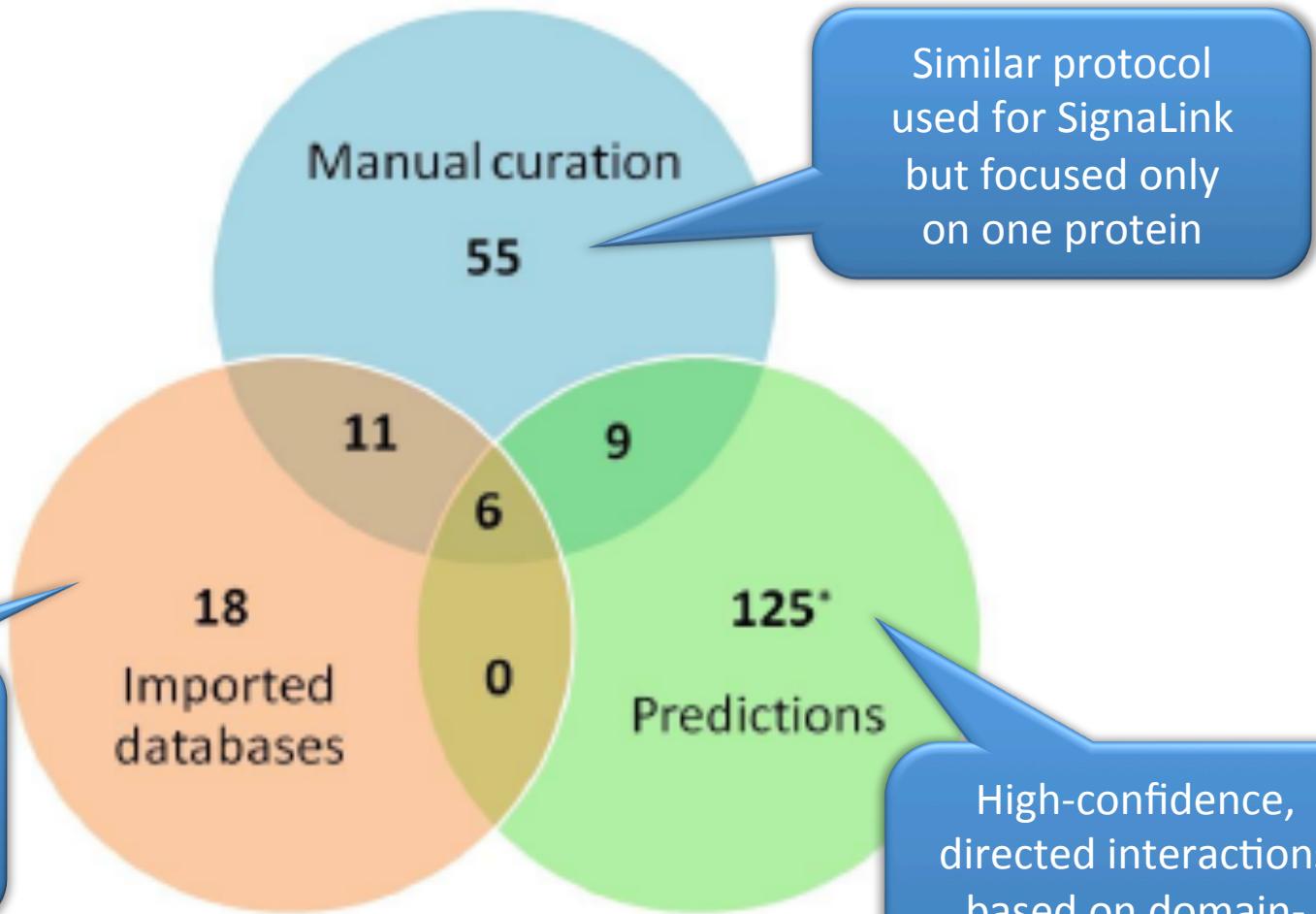
NRF2 interactome & regulome



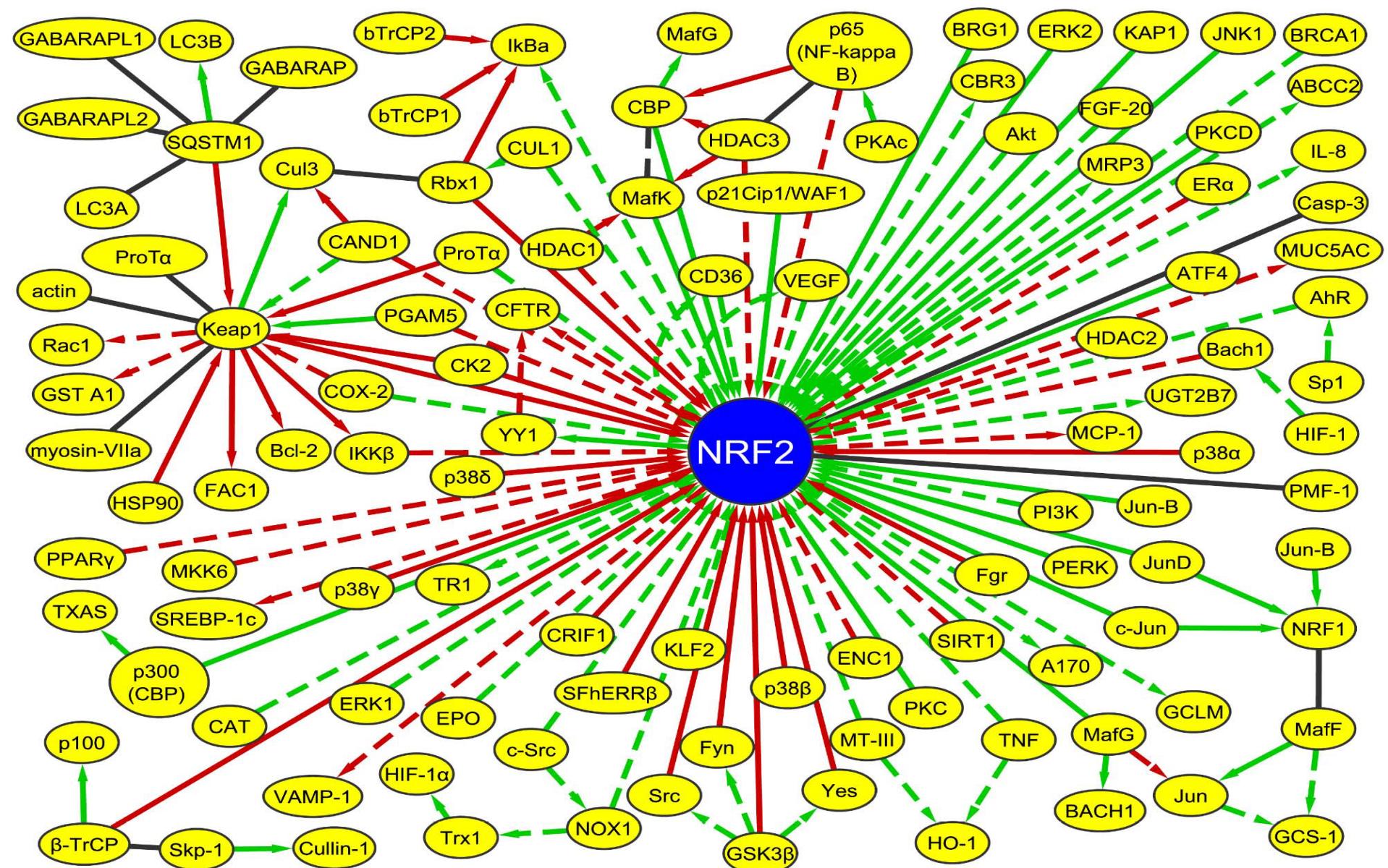
- Cancer
- Ageing
- Inflammation
- Diseases with oxidative stress

Reconstructing the NRF2 interactome

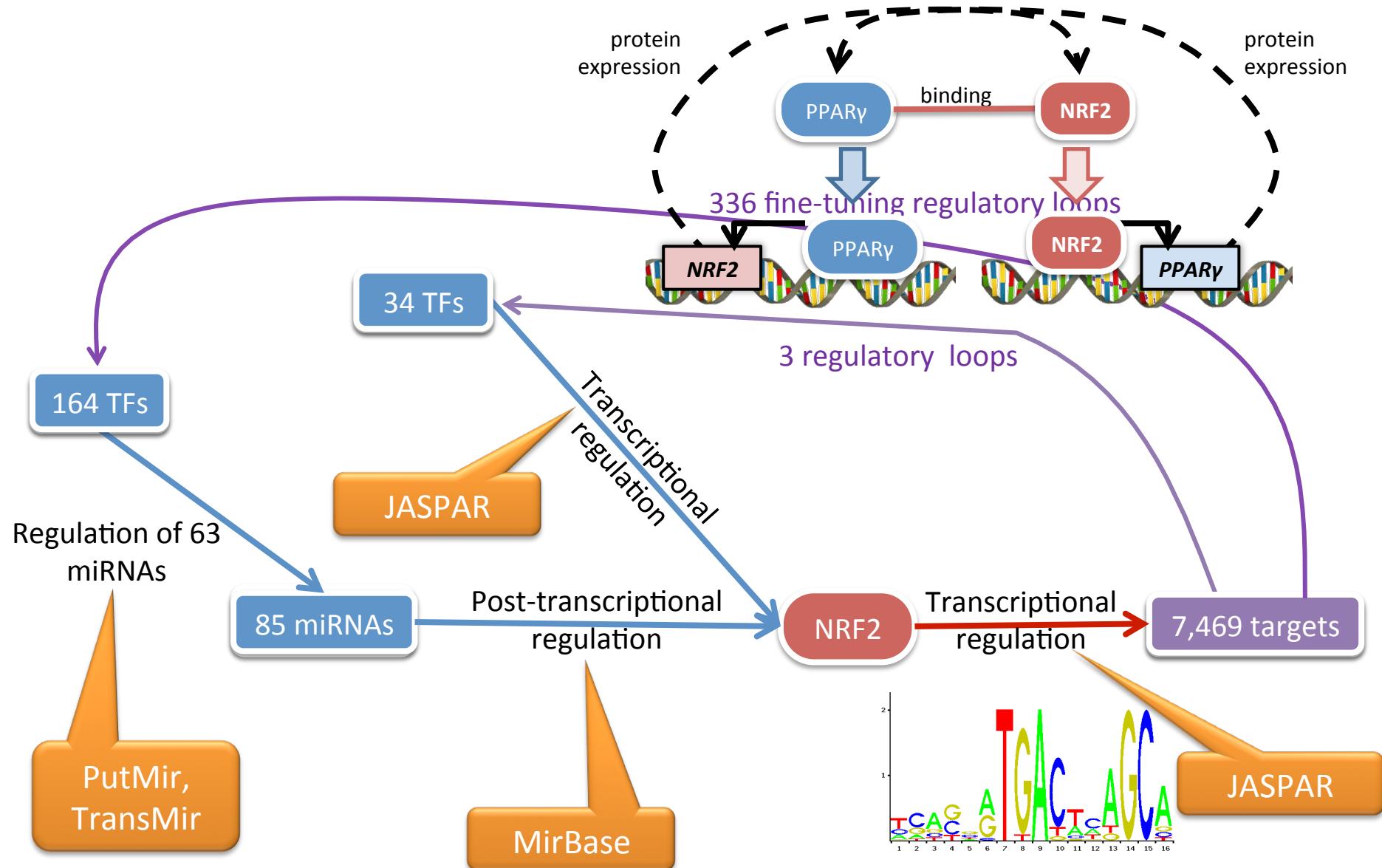
Distribution of the NRF2 interactors by sources



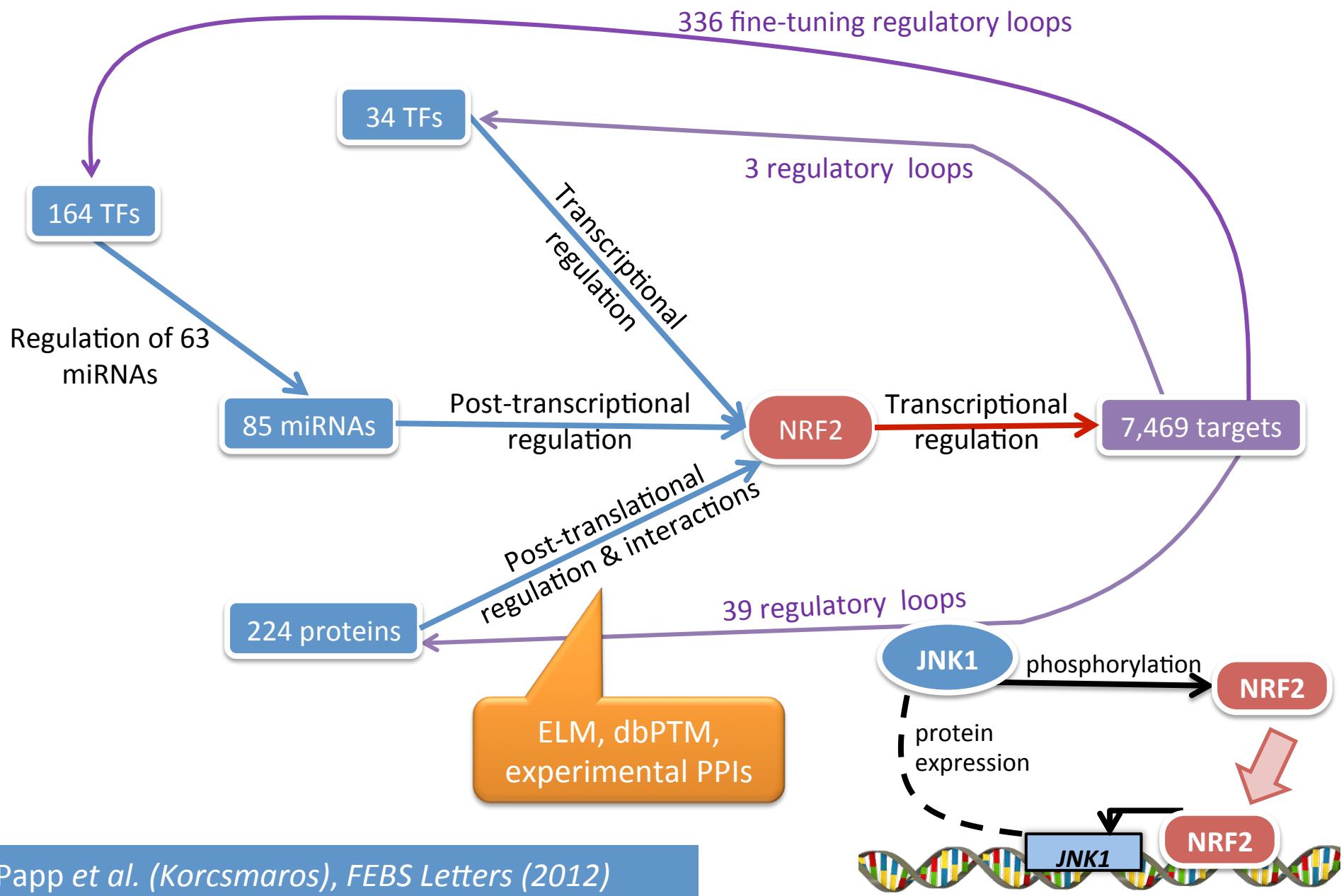
Reconstructing the NRF2 interactome



Reconstructing the NRF2 regulome



Reconstructing an integrated NRF2 network



- Reactome
- SignaLink
- ConsensusPathDB
- ...

• Fantom4

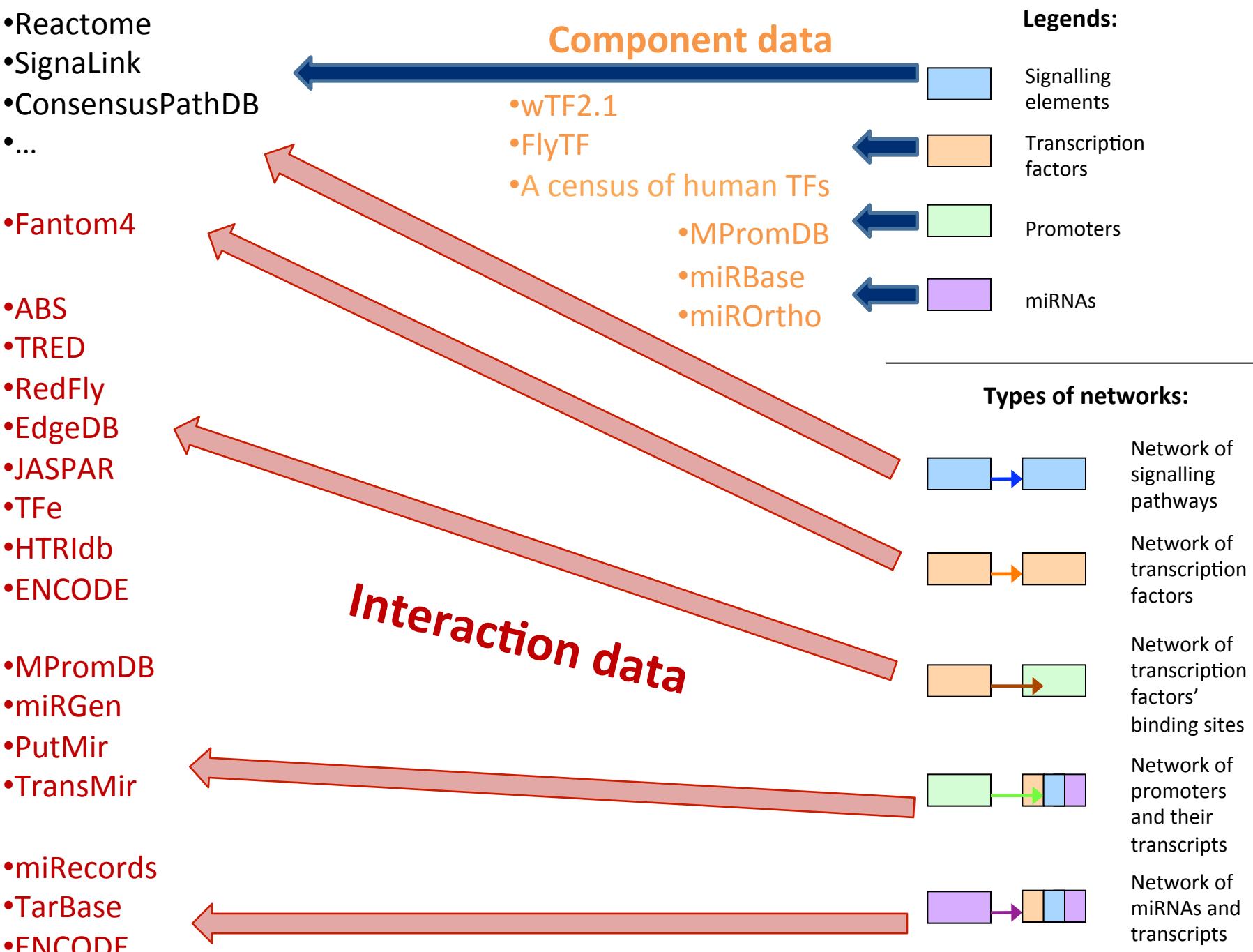
- ABS
- TRED
- RedFly
- EdgeDB
- JASPAR
- TFe
- HTREdb
- ENCODE

- MPromDB
- miRGen
- PutMir
- TransMir

- miRecords
- TarBase
- ENCODE

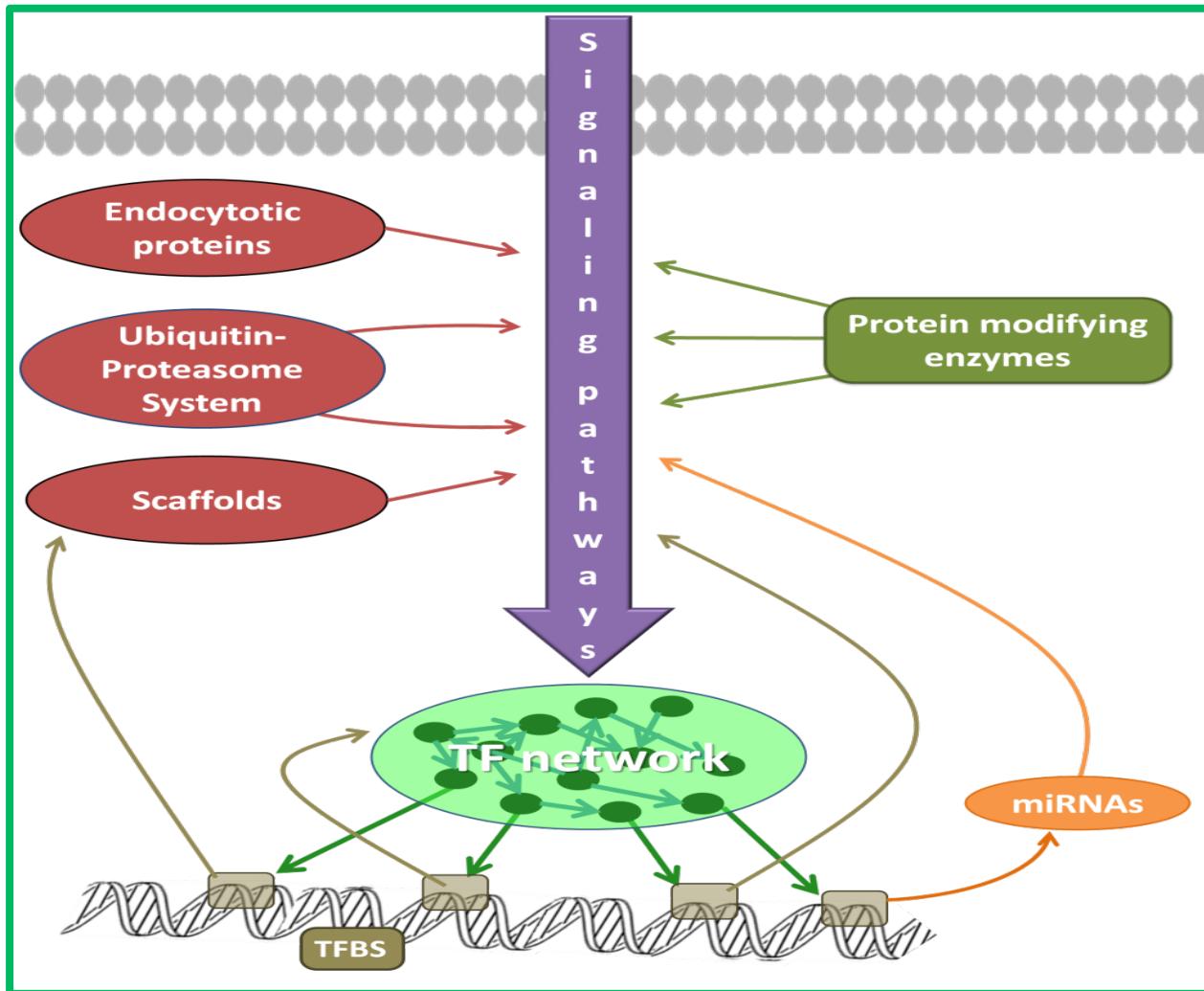
Component data

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



SignaLink 2.0

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



- 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- β WNT/Wingless

All

TF network

- » Interactions from: ELM Server (Domain-motif interactions)
- » Results: pathway protein modifying enzymes

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- β WNT/Wingless
- All

TF network

Range from pathway specific TF:

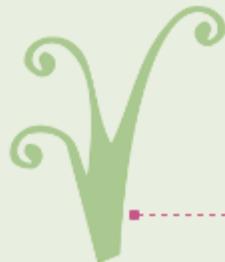
1 2 3 all

- » 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



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

Pathways:



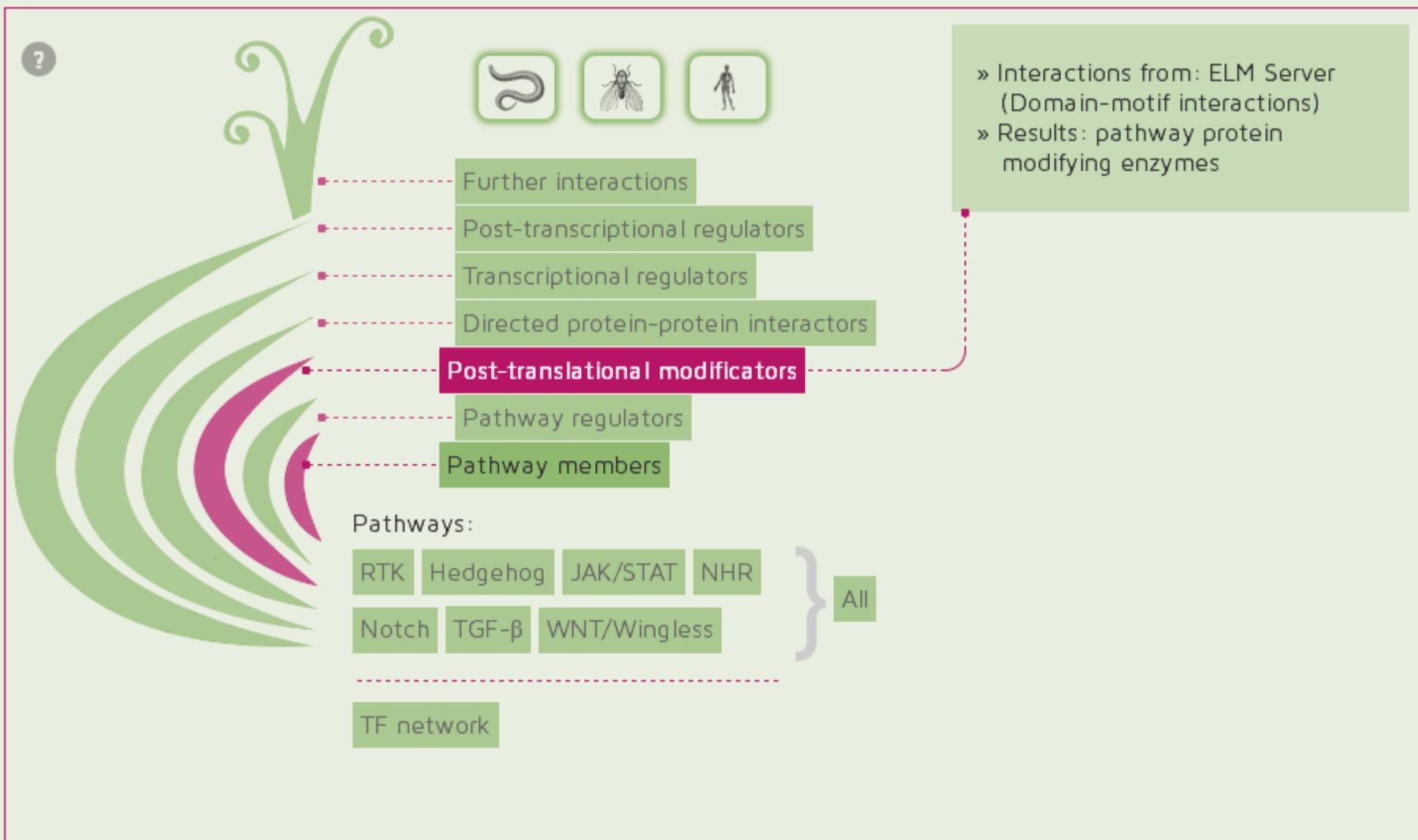
TF network

- » 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



?

Further interactions

Post-transcriptional regulators

Transcriptional regulators

Directed protein-protein interactors

Post-translational modifiers

Pathway regulators

Pathway members

Pathways:

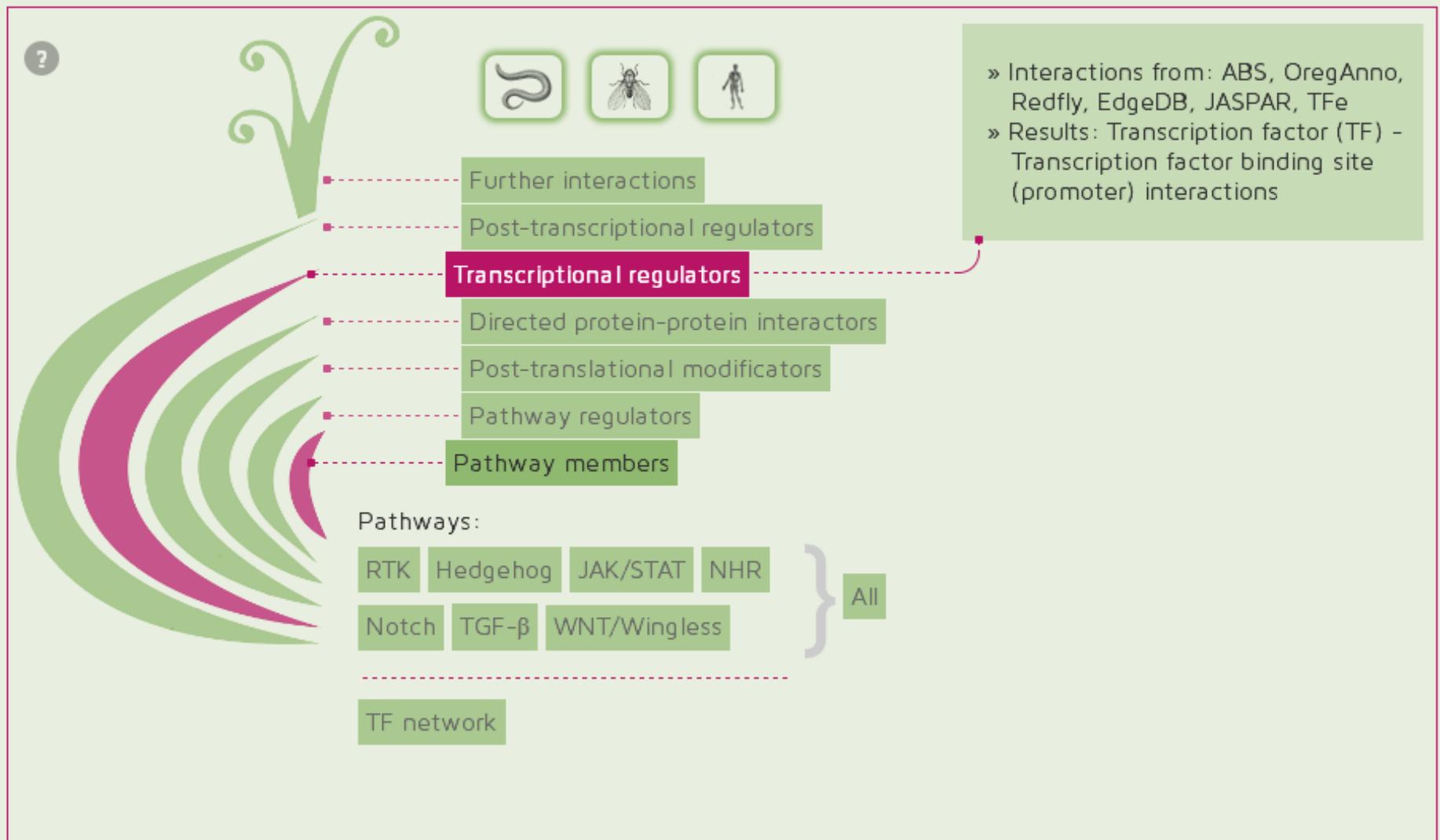
RTK Hedgehog JAK/STAT NHR

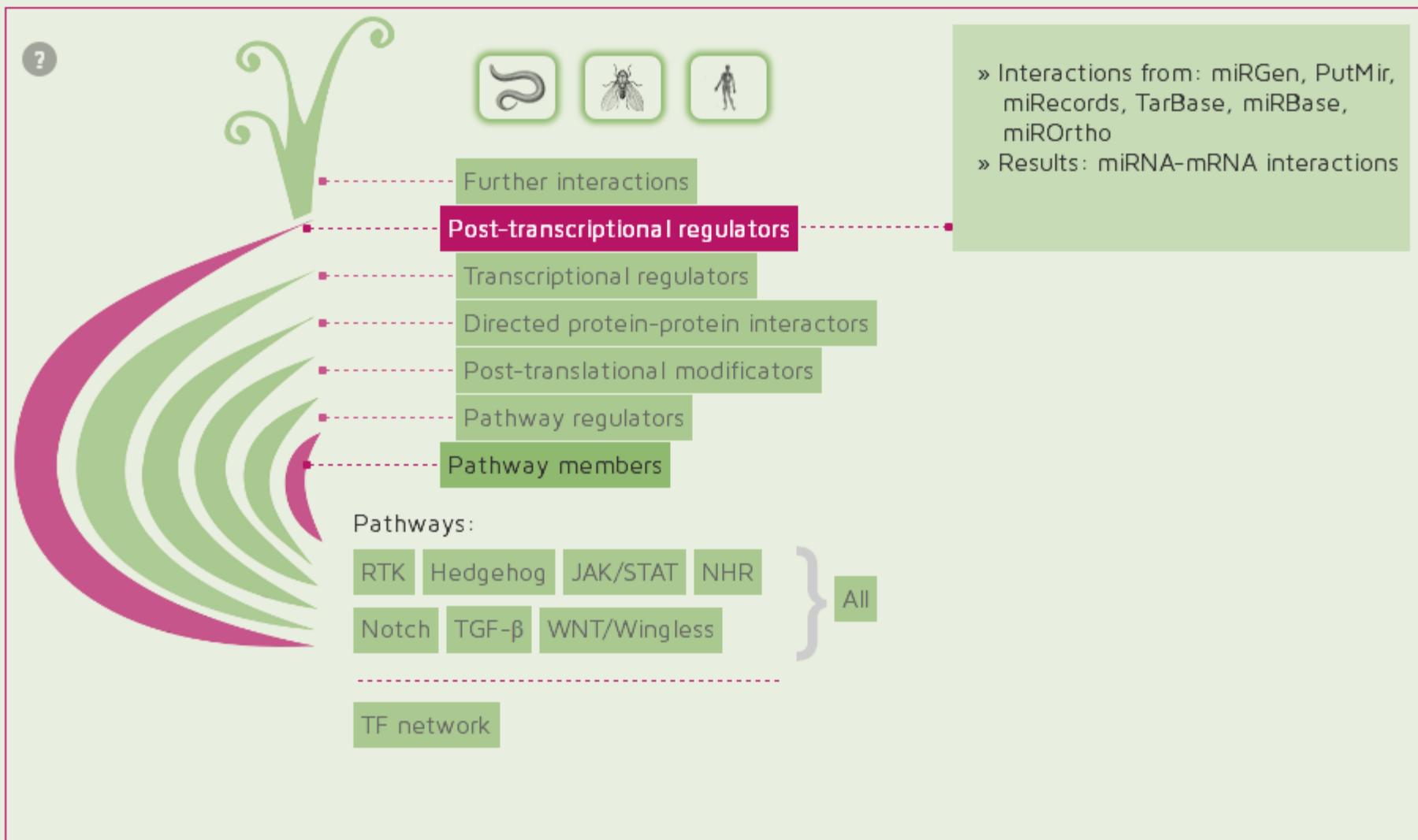
Notch TGF- β WNT/Wingless

All

TF network

- » Prediction of the direction of protein-protein interactions based on domain-domain interactions
- » Interactions from: WI8, DroID, HPRD, BioGrid, + prediction based on domain-domain interactions
- » Results: possible first neighbors of pathway proteins with direction and confidence score





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- β	WNT/Wingless	All

TF network

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

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- β WNT/Wingless

All

TF network

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

Output format:

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

Compress: » gzip » zip » none

Proceed to download

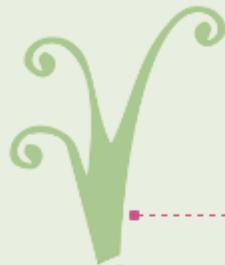
OR

Complete database

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- β 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

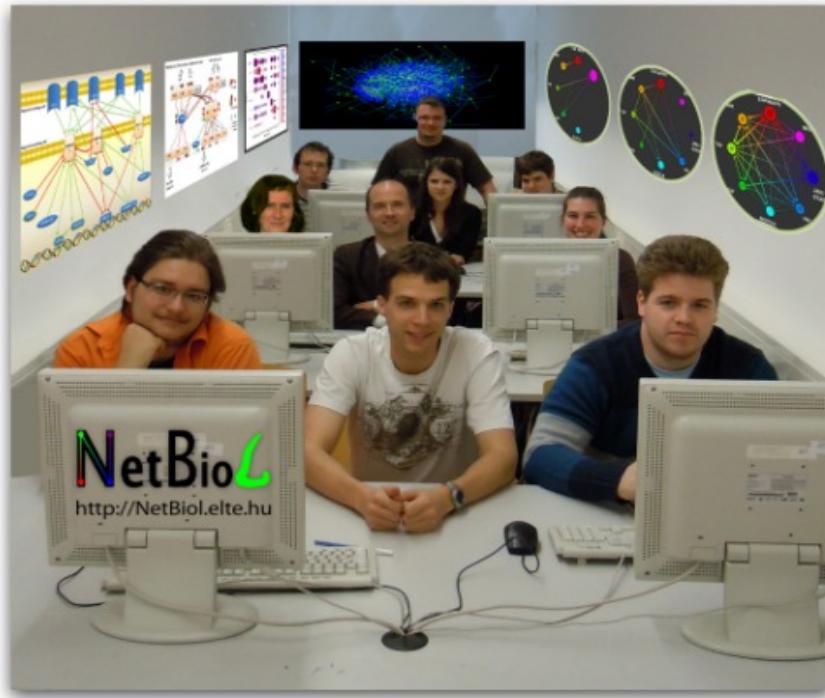
OR

Complete database

Discussion



Thank you for your attention!



Tamas.Korcsmaros@tgac.ac.uk

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.