

# Integrating literature-constrained and data-driven inference of signalling networks

**F Eduati**<sup>1,2</sup>, J De Las Rivas<sup>3</sup>, B Di Camillo<sup>1</sup>, G Toffolo<sup>1</sup>, J Saez-Rodriguez<sup>2</sup>




<sup>1</sup>Department of Information Engineering, University of Padova, Padova, Italy

<sup>2</sup>European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Cambridge, UK

<sup>3</sup>Bioinformatics & Functional Genomics Group, Cancer Research Center (CSIC/USAL), Salamanca, Spain

# Background

Available information about signalling networks:

- high-throughput methods  unsigned and undirected Protein-protein Interaction Networks (PINs)
- literature/databases  causal Prior Knowledge Networks (PKNs)
- perturbation experiments  cell-type and context specific data

# Modelling signalling networks

Two main approaches:

- training the PKN to the data

|     |  |
|-----|--|
| Adv | Efficient handling of large amount of data |
| Lim | Possible missing links                     |

- inference of Data-Driven Networks (DDNs) from data

|     |                                    |
|-----|------------------------------------|
| Adv | No prior knowledge required        |
| Lim | Nontrivial mapping of DDNs to PKNs |

# Aim

Integration of literature-constrained &  
data-driven methods  
to infer signalling networks

# Availability

R package **CNORfeeder**

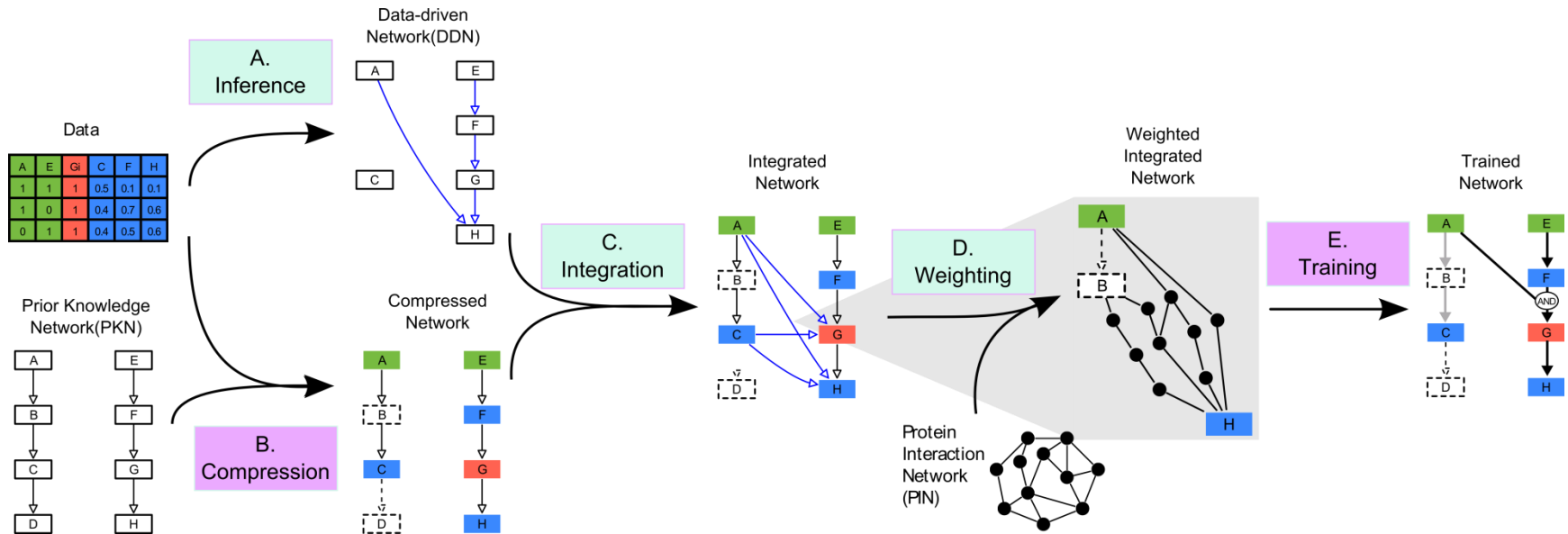
[designed to be integrated with CelNOptR (Saez-Rodriguez et al, 2009)]

available at [www.cellnopt.org](http://www.cellnopt.org)

# Pipeline

CNORfeeder

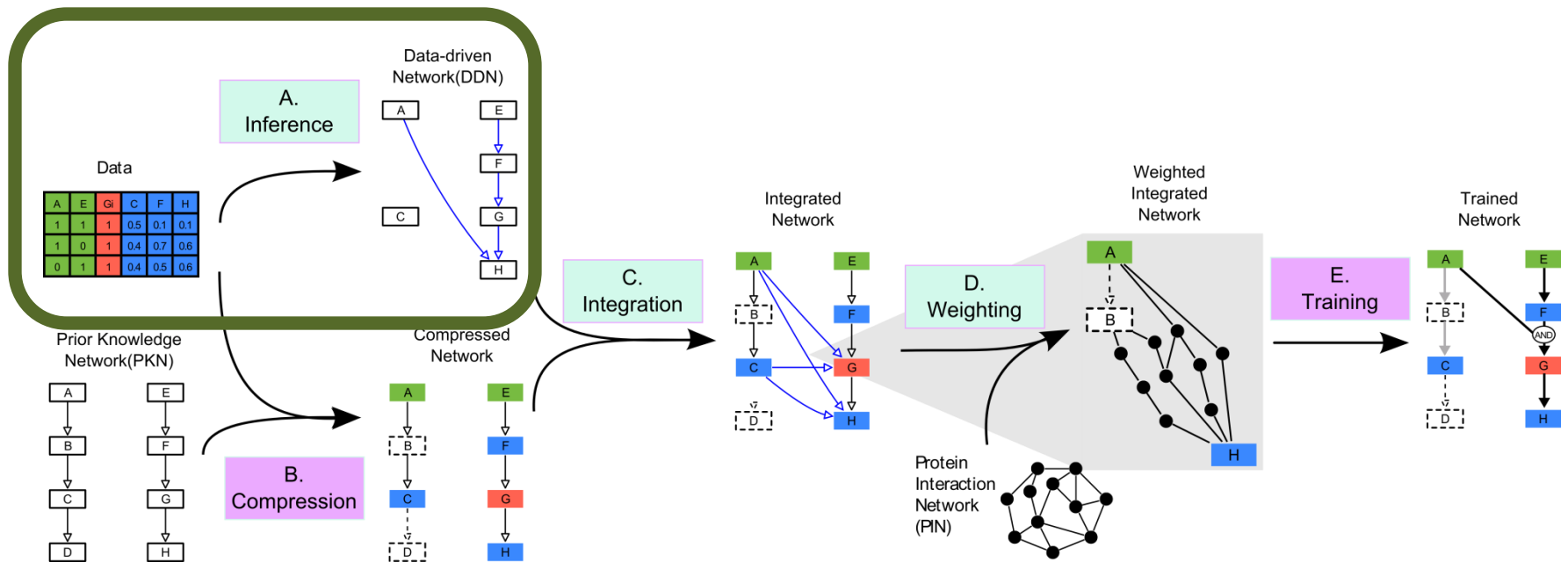
CellNOptR



# Pipeline

CNORfeeder

CellNOptR



stimulated

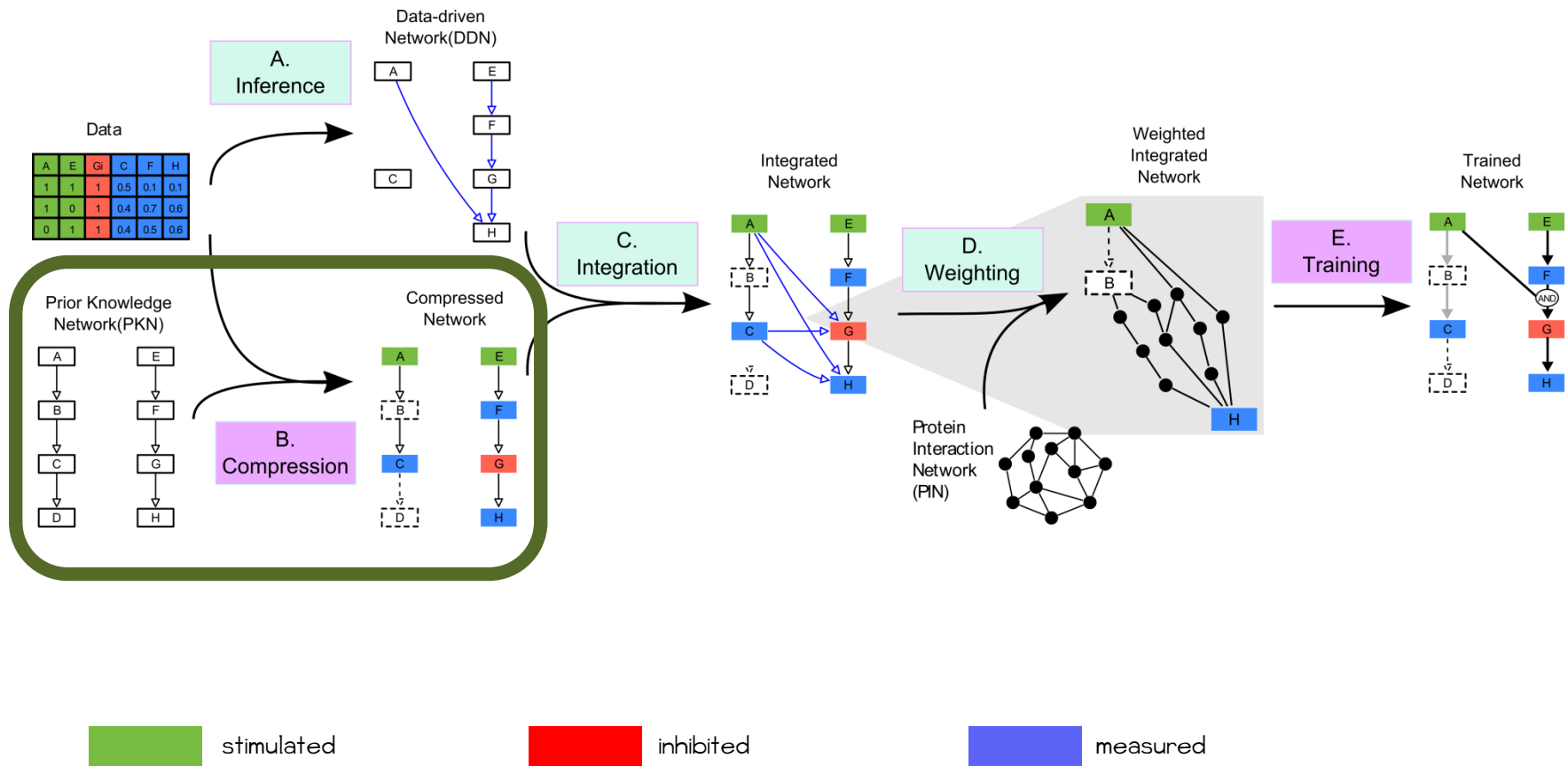
inhibited

measured

# Pipeline

CNORfeeder

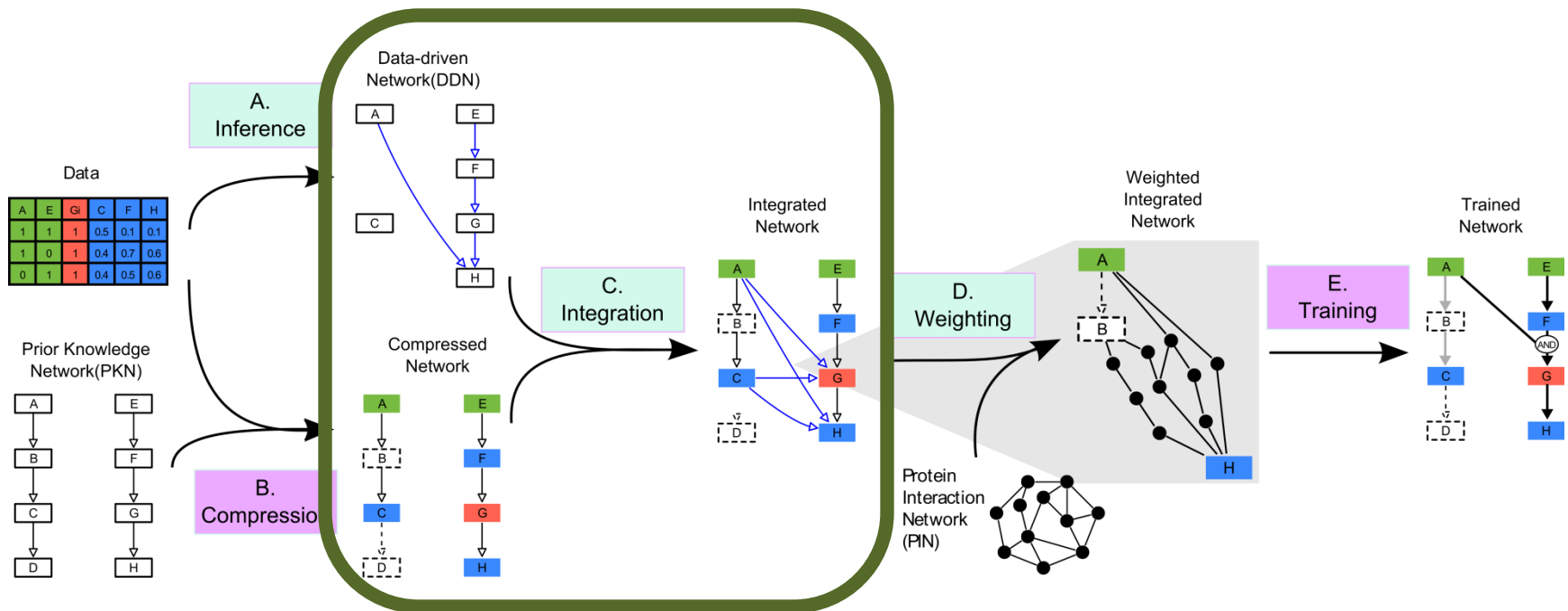
CellNOptR



# Pipeline

CNORfeeder

CellNOptR



stimulated

inhibited

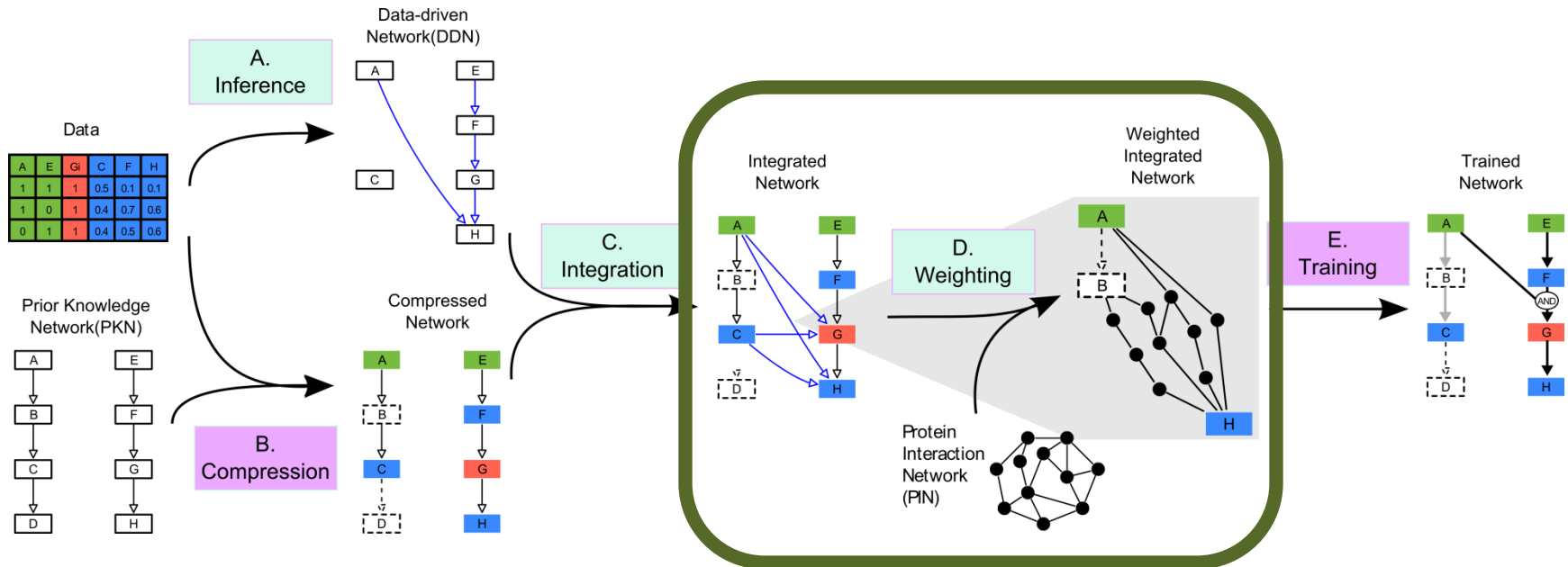
measured



# Pipeline

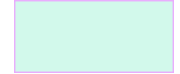
CNORfeeder

CellNOptR

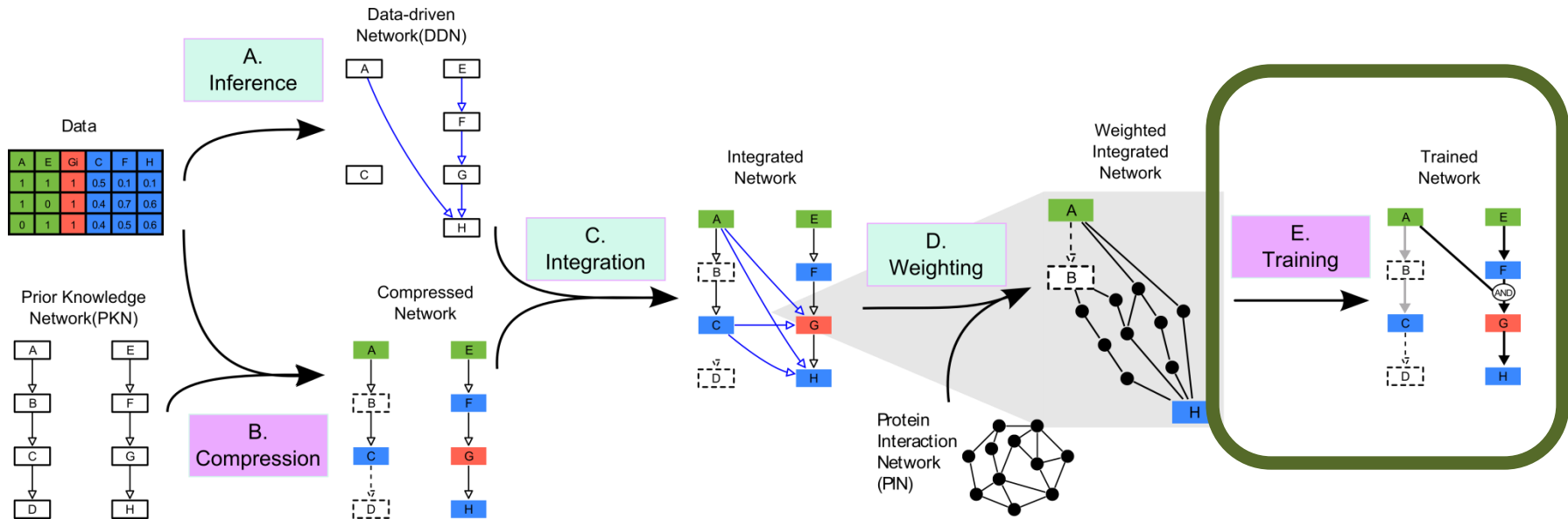


# Pipeline

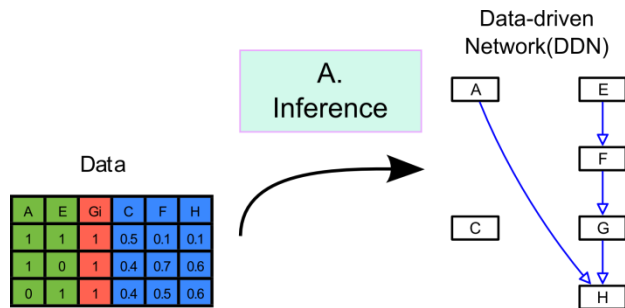
CNORfeeder



CellNOptR



# Inference

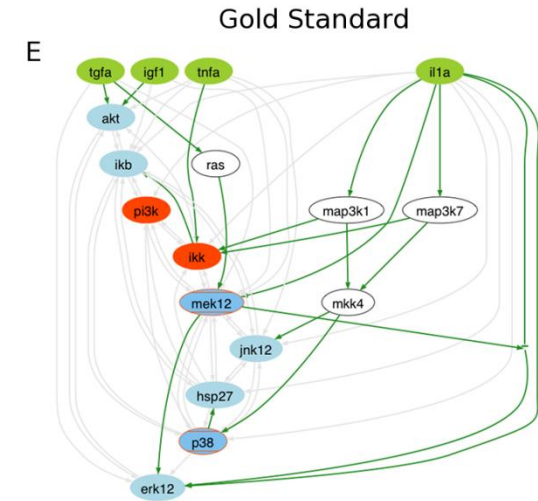
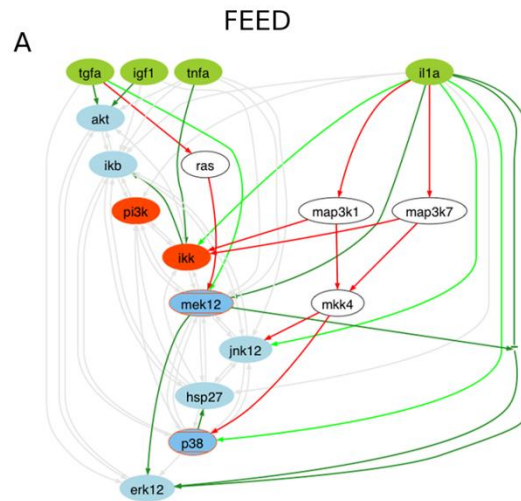


Data are used to **infer** a strictly **data-driven network (DDN)** using **reverse-engineering** methods.

4 methods included so far:

1. FEED (Eduati et al., 2010)
2. ARACNe ('minet R package' (Mayer et al, 2008))
3. CLR
4. Bayesian networks ('catnet R package')

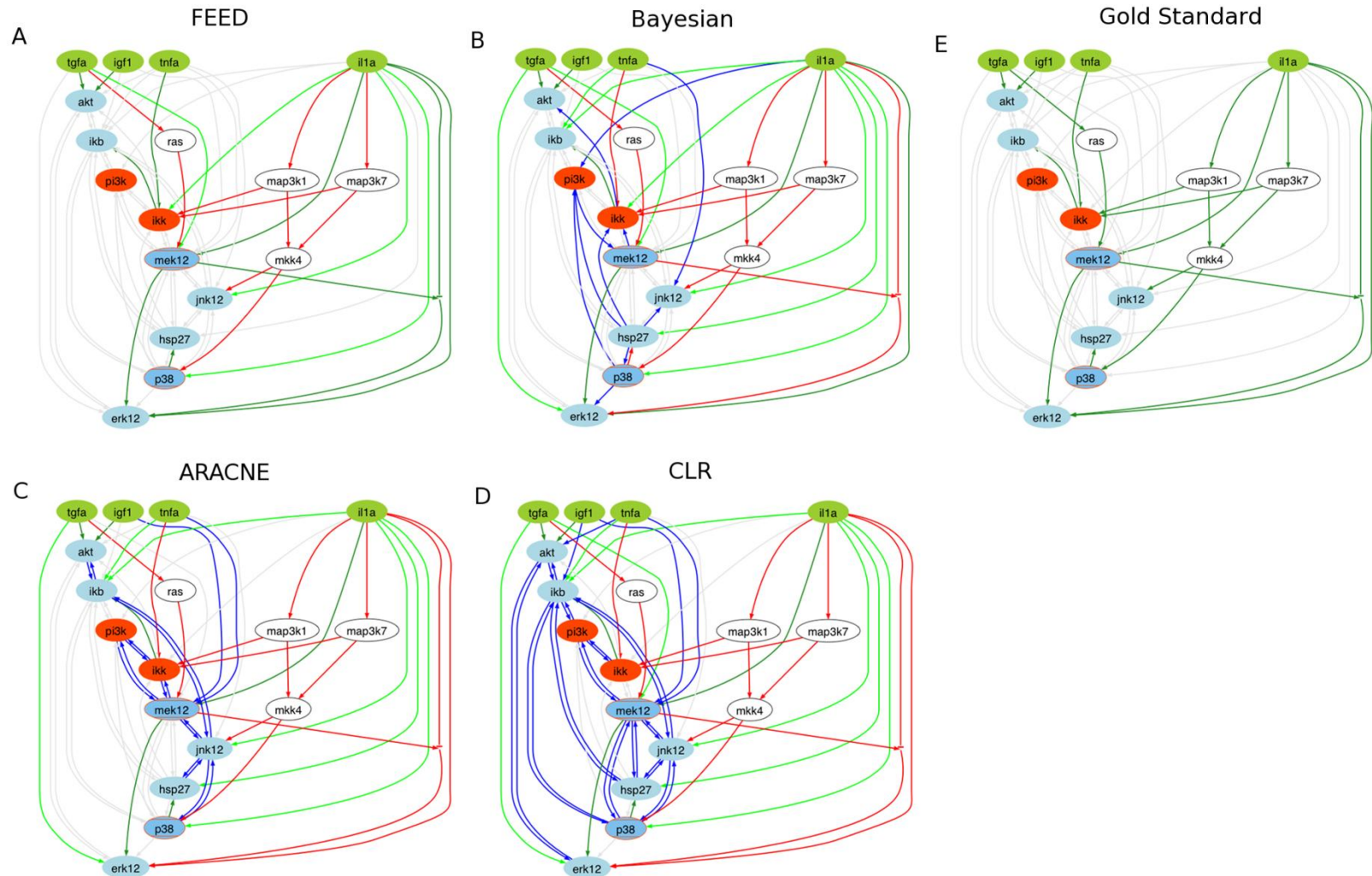
# Inference of benchmark model



Comparison of the inferred networks (DDNs) with the Gold Standard:

→ in both networks    
 → path in the Gold Standard    
 → in Gold Standard not in DDN    
 → in DDN not in Gold Standard

# Inference of benchmark model

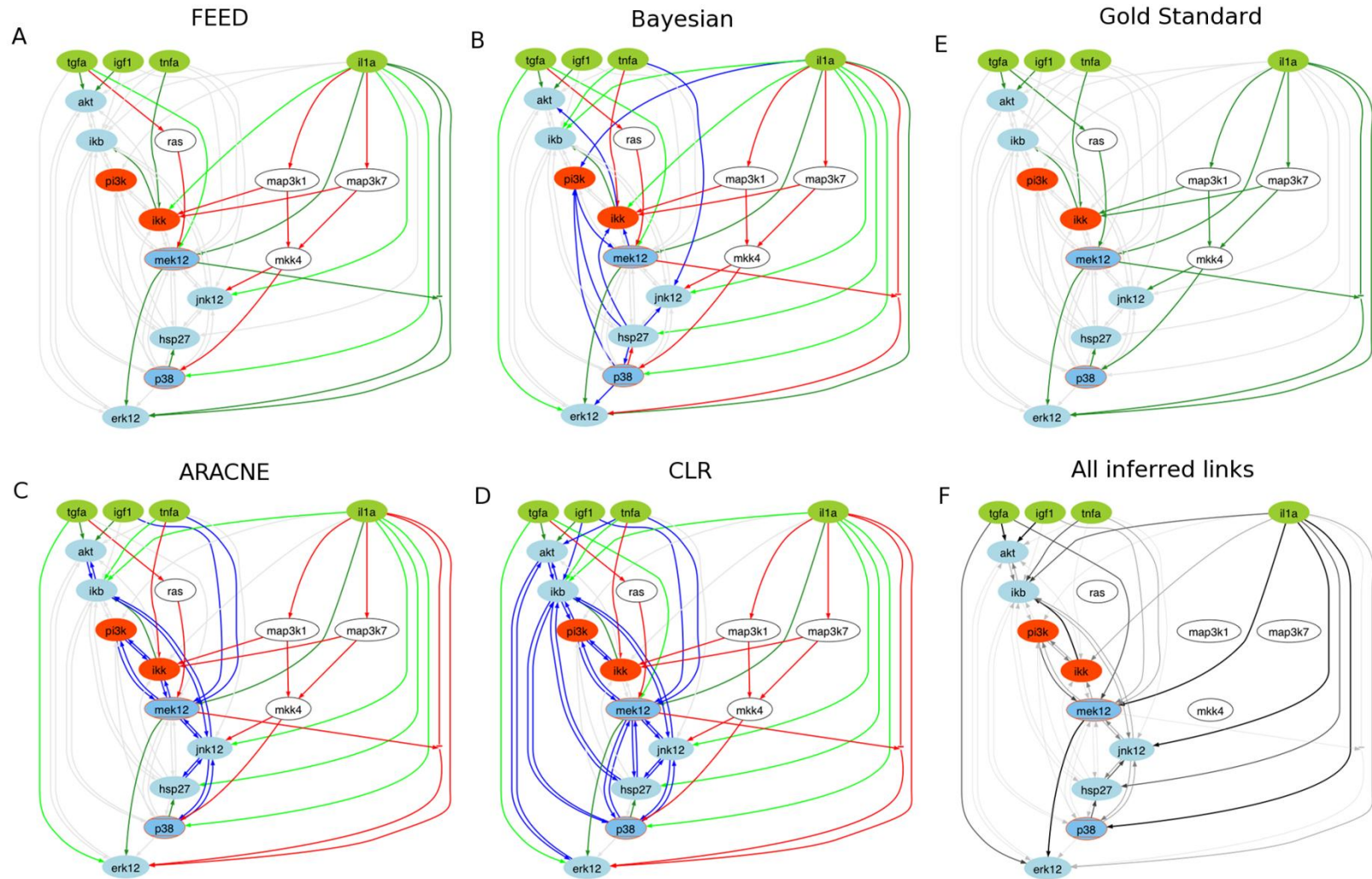


Comparison of the inferred networks (DDNs) with the Gold Standard:

→ in both networks    → path in the Gold Standard    → in Gold Standard not in DDN    → in DDN not in Gold Standard



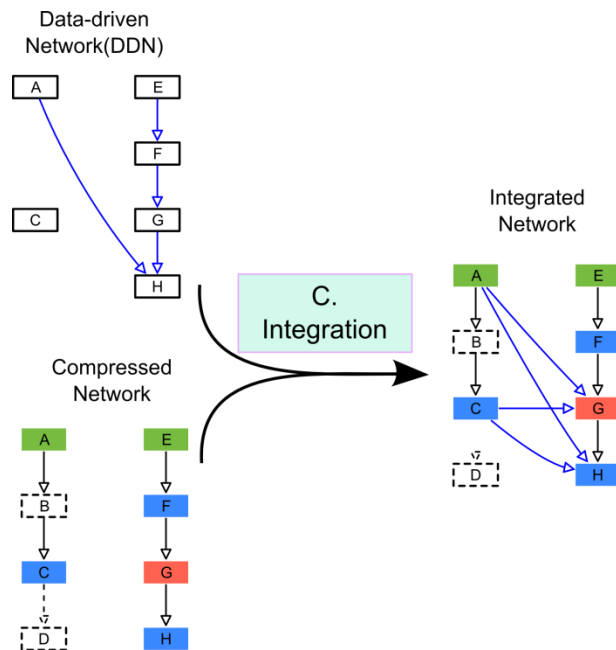
# Inference of benchmark model



Comparison of the inferred networks (DDNs) with the Gold Standard:

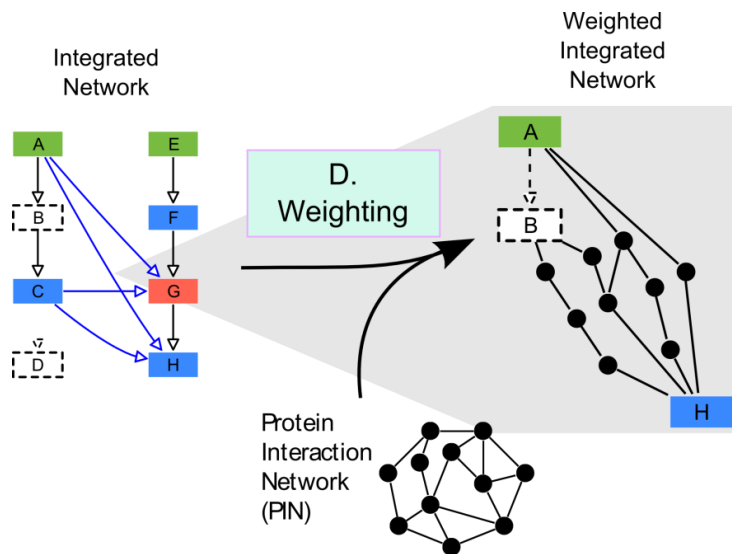
→ in both networks   
 → path in the Gold Standard   
 → in Gold Standard not in DDN   
 → in DDN not in Gold Standard

# Integration



- compressed network is integrated with the DDN  
(blue: links from the DDN,  
black: links from the PKN)
- one link in the DDN can correspond to multiple links in the PKN,  
(e.g. for the link from A to H)

# Weighting (and training)



Objective  
function

$$\theta(P) = \theta_f(P) + \alpha \cdot \theta_s(P)$$

P: candidate  
solution

fit

size

$$\frac{1}{N} \sum_{n=1}^N (data_n - pred_n)^2 = MSE$$

$$\theta_{pkn}(P) + \beta \cdot \theta_{add}(P)$$

PINs weighting \*

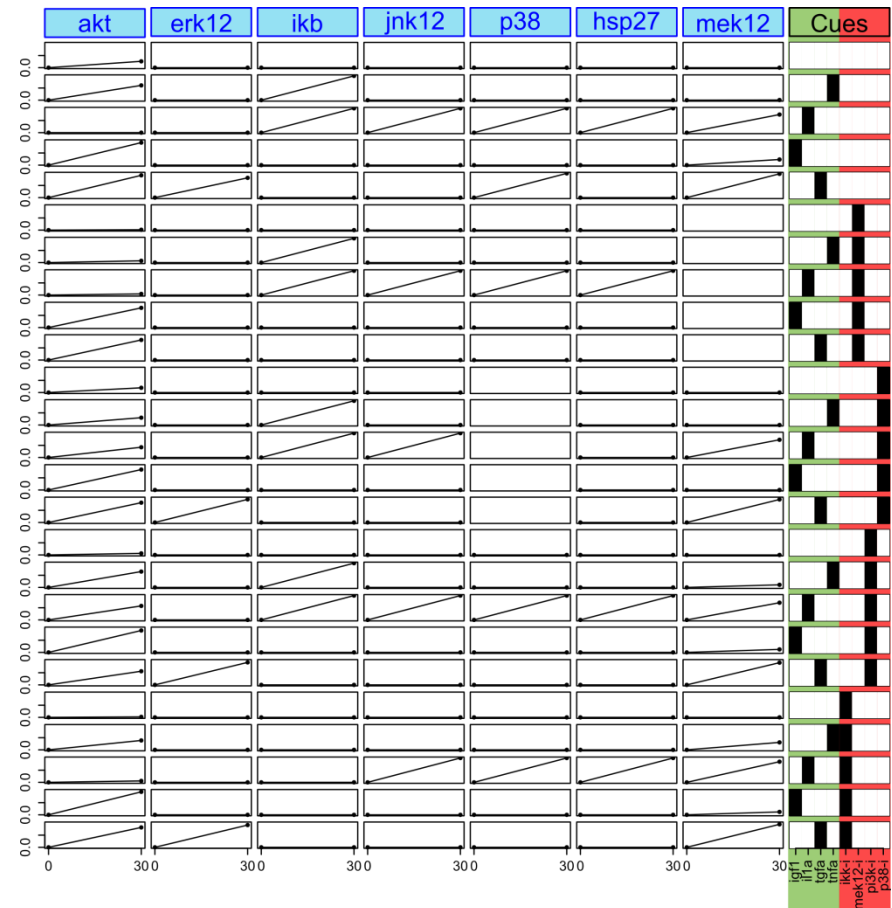
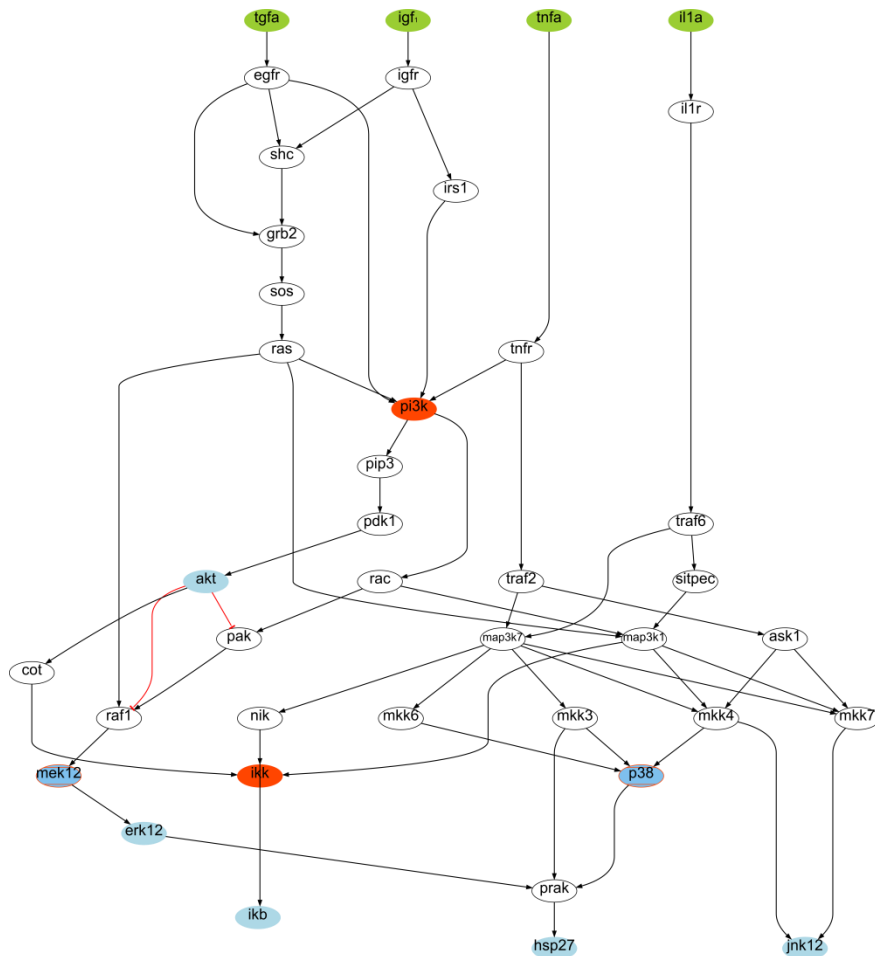
\* shorter path in the PIN

$$\sum_{i=1}^A v_i \left( \frac{1}{\omega_i} + 1 \right) P_i$$

more plausible link integrated in the PKN



## Application: growth and inflammatory signalling



stimulated

 inhibited

 measured

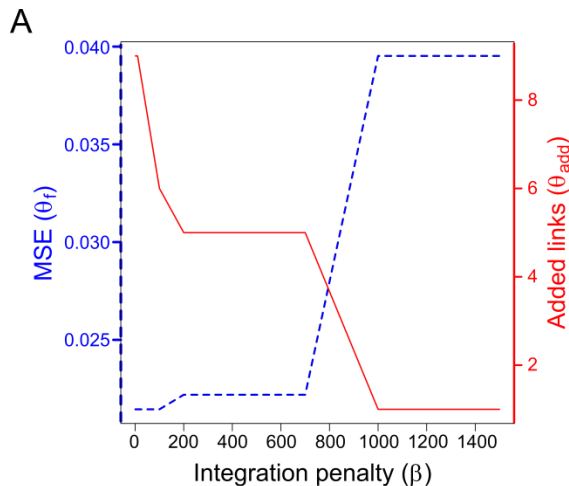
# Results: tuning integration penalty

Settings:

- Inference method: FEED
- $\alpha = 0.001$

$$\theta(P) = \theta_f(P) + \alpha \cdot [\theta_{pkn}(P) + \beta \cdot \theta_{add}(P)]$$

Integration penalty



B

|                  | 1 | 100 | 500 |
|------------------|---|-----|-----|
| igf1=akt         | x | x   | x   |
| tnfa=ikk         | x | x   | x   |
| il1a+mek12=erk12 | x | x   | x   |
| il1a=mek12       | x | x   | x   |
| tgfa=akt         | x | x   |     |
| tgfa=mek12       | x |     |     |
| il1a=jnk12       | x |     |     |

- low value of  $\beta$

Best fit

Many links

- high value of  $\beta$

Less links

Bad fit

- reduced number of links selected/prioritized when using PIN to weight links (dark green).

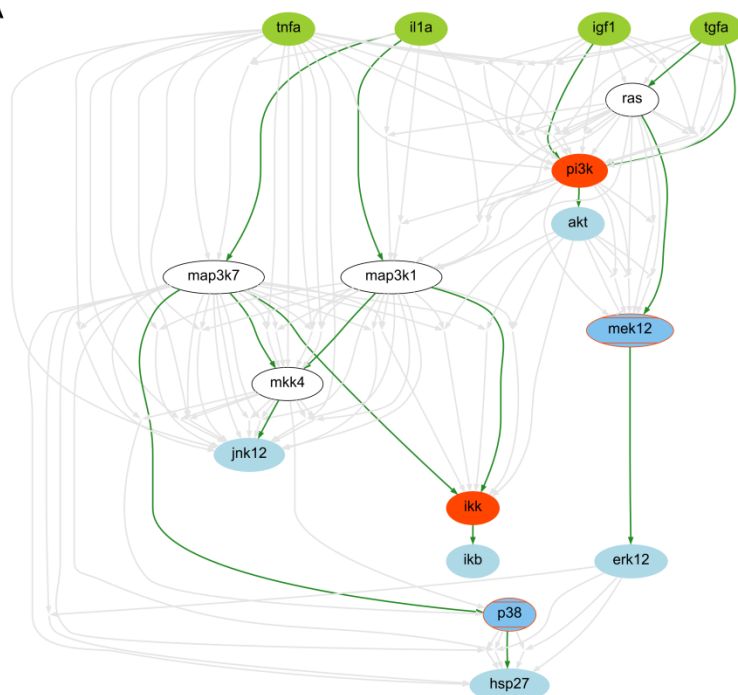
# Results: integrated network and fit

Settings:

- Inference method: FEED
- $\alpha = 0.001$
- $\beta = 700$

## A Compressed network

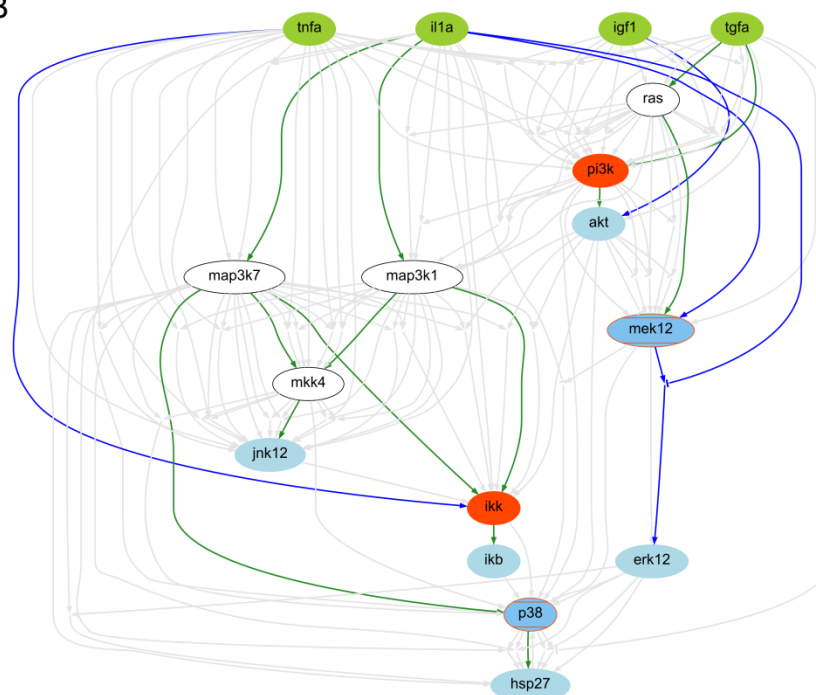
A



|     | akt   | erk12 | ikb   | jk12  | p38   | hsp27 | mek12 | TOTAL        |
|-----|-------|-------|-------|-------|-------|-------|-------|--------------|
| MSE | 0.119 | 0.002 | 0.159 | 0.000 | 0.050 | 0.000 | 0.125 | <b>0.064</b> |

## B Integrated network

B



|     | akt   | erk12 | ikb   | jk12  | p38   | hsp27 | mek12 | TOTAL        |
|-----|-------|-------|-------|-------|-------|-------|-------|--------------|
| MSE | 0.087 | 0.002 | 0.000 | 0.000 | 0.050 | 0.000 | 0.023 | <b>0.022</b> |

→ form PKN

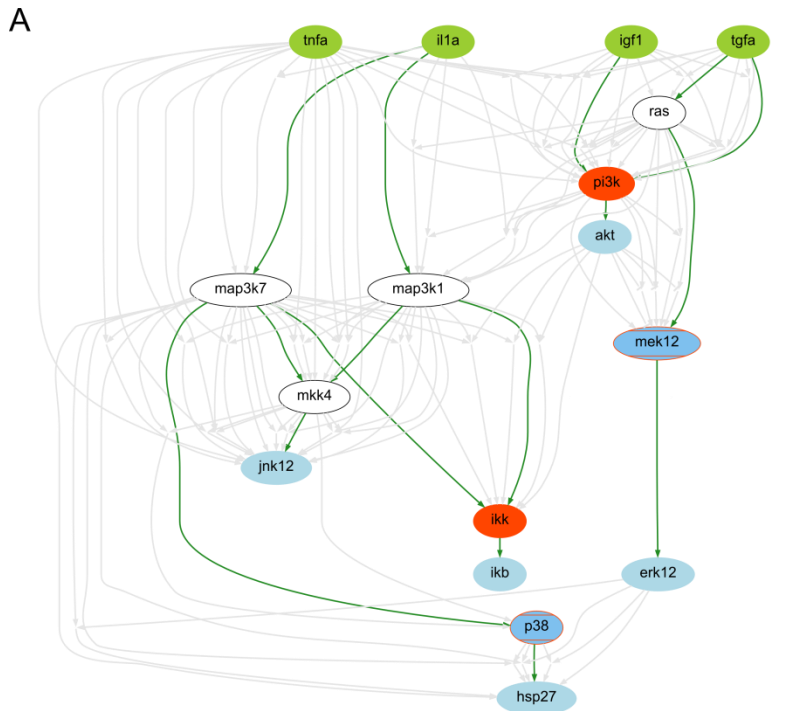
→ added using CNORfeeder

# Results: integrated network and fit

Settings:

- Inference method: FEED
- $\alpha = 0.001$
- $\beta = 700$

A Compressed network

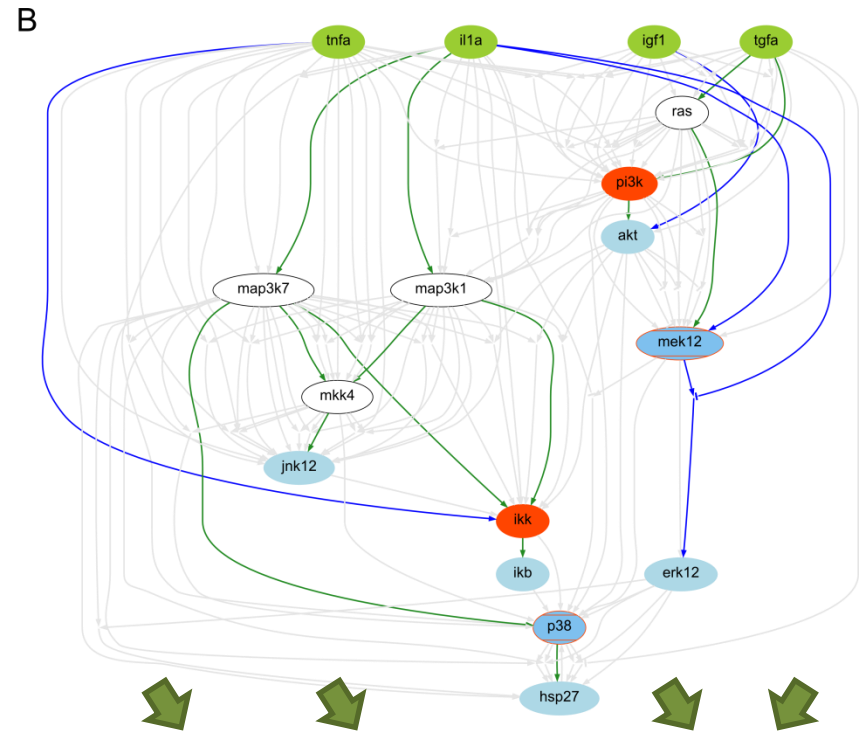


|     | akt   | erk12 | ikb   | jk12  | p38   | hsp27 | mek12 | TOTAL        |
|-----|-------|-------|-------|-------|-------|-------|-------|--------------|
| MSE | 0.119 | 0.002 | 0.159 | 0.000 | 0.050 | 0.000 | 0.125 | <b>0.064</b> |

→ form PKN

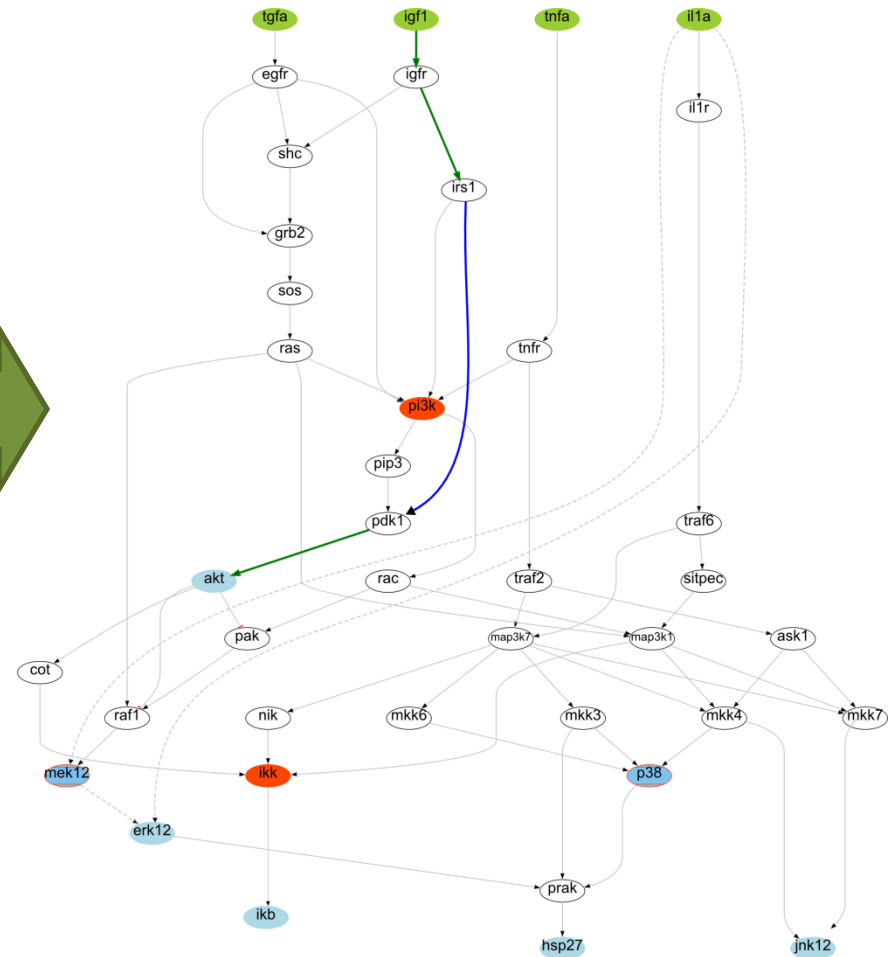
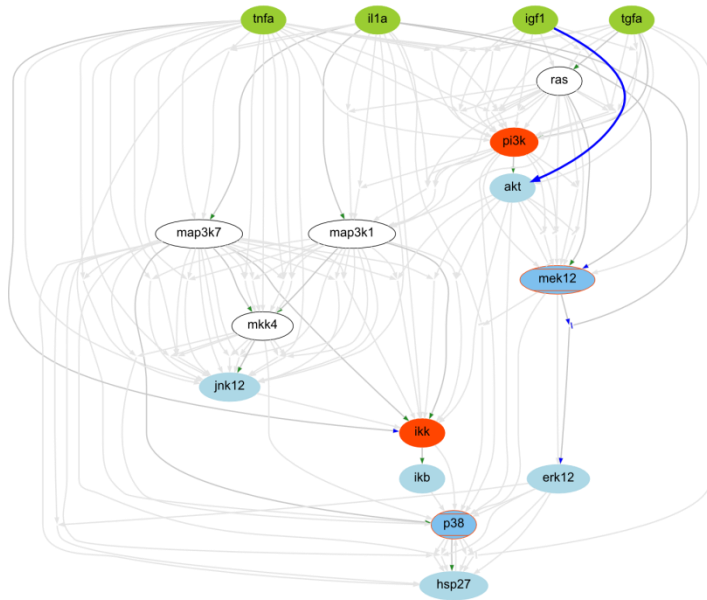
→ added using CNORfeeder

B Integrated network



|     | akt   | erk12 | ikb   | jk12  | p38   | hsp27 | mek12 | TOTAL        |
|-----|-------|-------|-------|-------|-------|-------|-------|--------------|
| MSE | 0.087 | 0.002 | 0.000 | 0.000 | 0.050 | 0.000 | 0.023 | <b>0.022</b> |

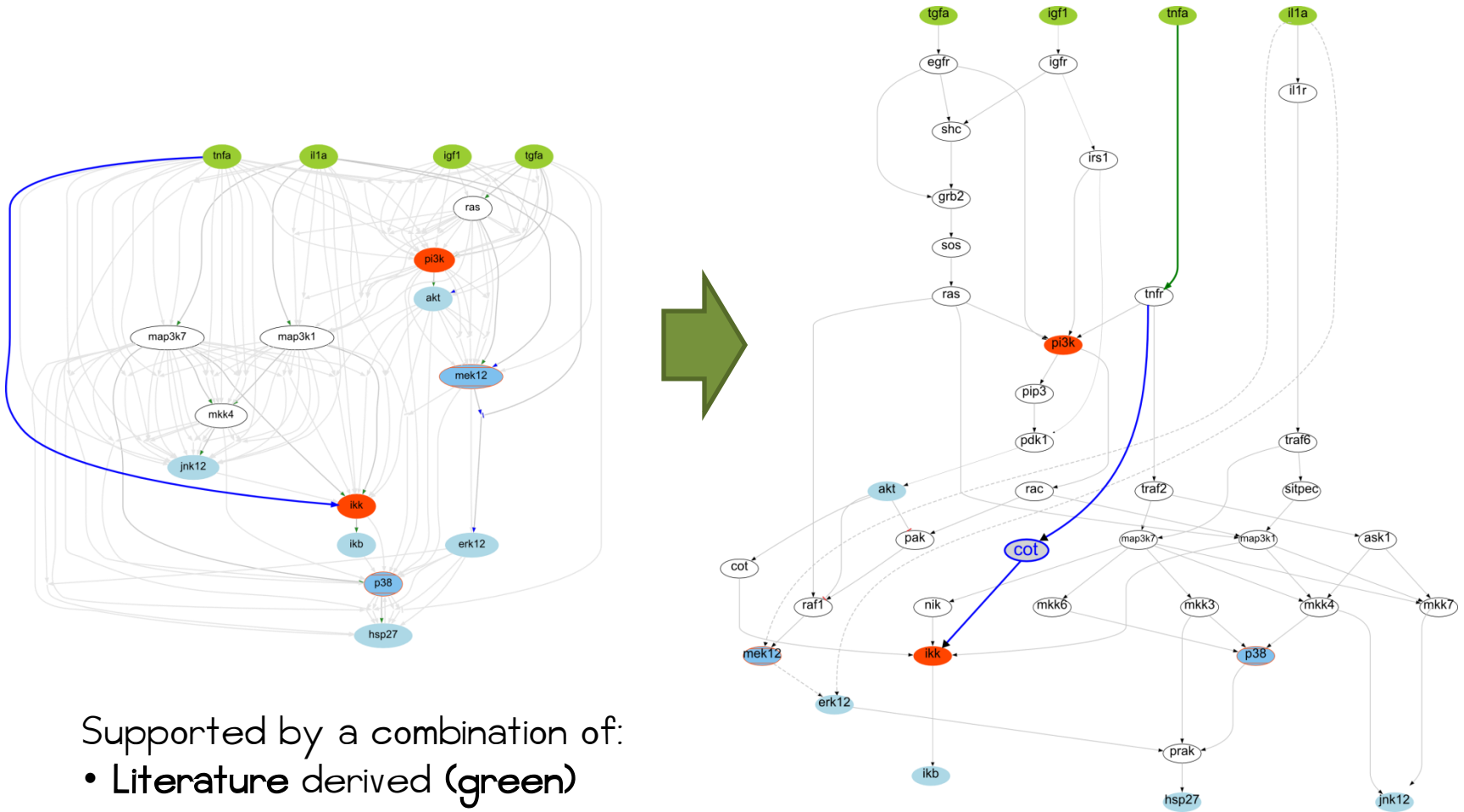
# Results: added link (example 1)



Supported by a combination of:

- Literature derived (green)
- Protein-protein interaction (blue)

# Results: added link (example 2)



Supported by a combination of:

- Literature derived (green)
- Protein-protein interaction (blue)

# Conclusions

CNORfeeder allows to:

- obtain a logic model that better describes data (with the minimum number of link);
- identify possible missing links in the PKN (e.g. incomplete biological Knowledge);
- use PINs as complementary information to suggest and support new interactions.

## REFERENCE PAPER

EDUATI F, DE LAS RIVAS J, DI CAMILLO B, TOFFOLO G, SAEZ-RODRIGUEZ J.

Integrating literature-constrained and data-driven inference of signalling networks. Bioinformatics, 2012

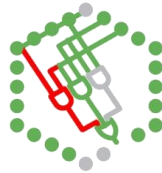
# Open points for discussion

1. Inclusion of other reverse-engineering methods able to use prior knowledge also in the inference step (e.g. Bayesian).
2. Assessment of reverse-engineering with large number of benchmark network and more realistic simulation of data.
3. Shortest paths were used as a metric to discriminate between links in the PPI but different methods could be applied.



# Acknowledgements

EMBL-EBI



- **Julio Saez-Rodriguez**
- Thomas Cokelaer
- Aidan MacNamara
- Camille Terfve



DEPARTMENT OF  
INFORMATION  
ENGINEERING  
UNIVERSITY OF PADOVA



- **Gianna Toffolo**
- Barbara Di Camillo



Cancer  
Research  
Center,  
Salamanca

- **Javier De Las Rivas**