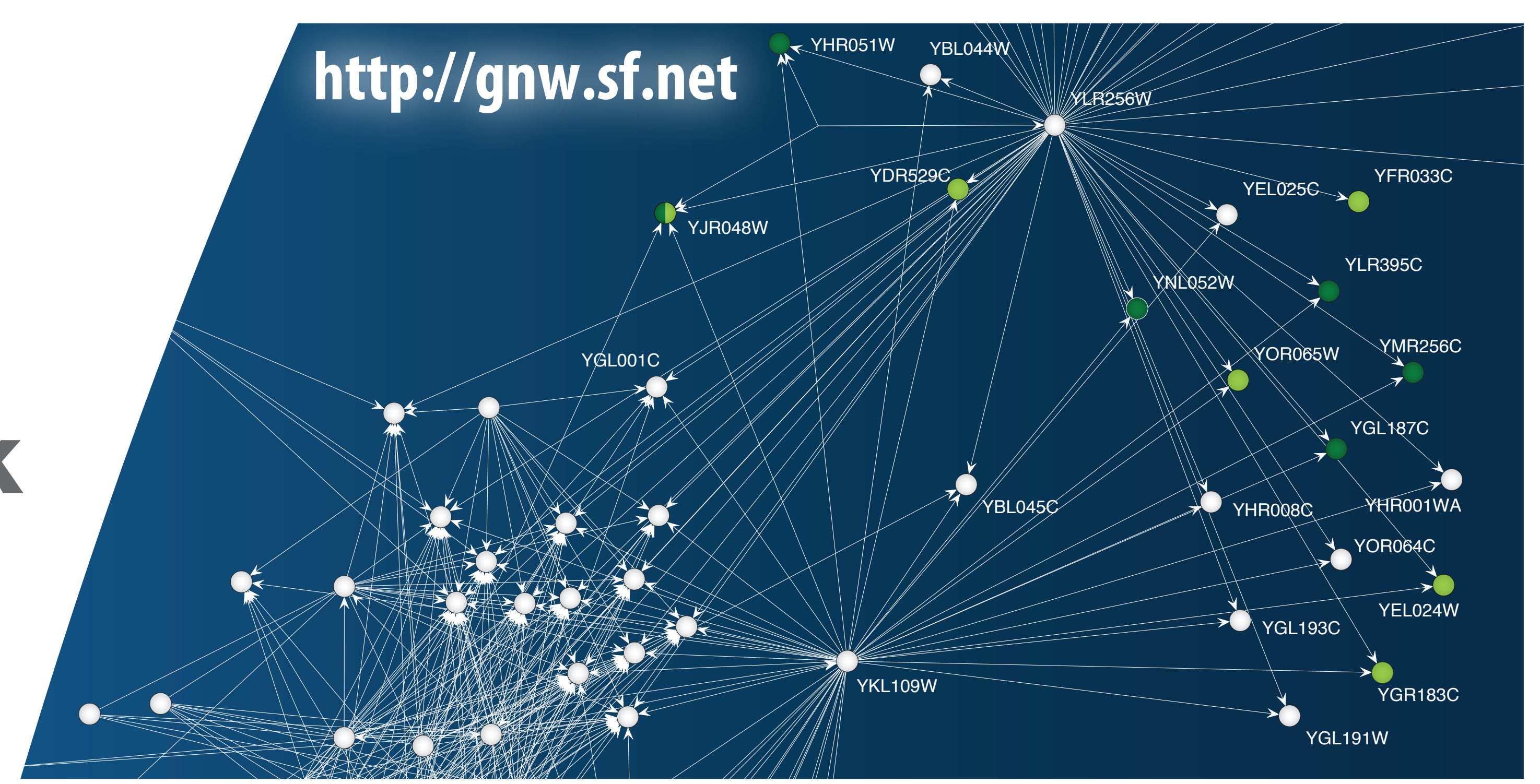


GeneNetWeaver: realistic benchmark generation and performance profiling of network inference methods

Thomas Schaffter^{1,*}, Daniel Marbach^{2,3,*}, Manolis Kellis^{2,3}, and Dario Floreano¹

*These authors contributed equally to this work

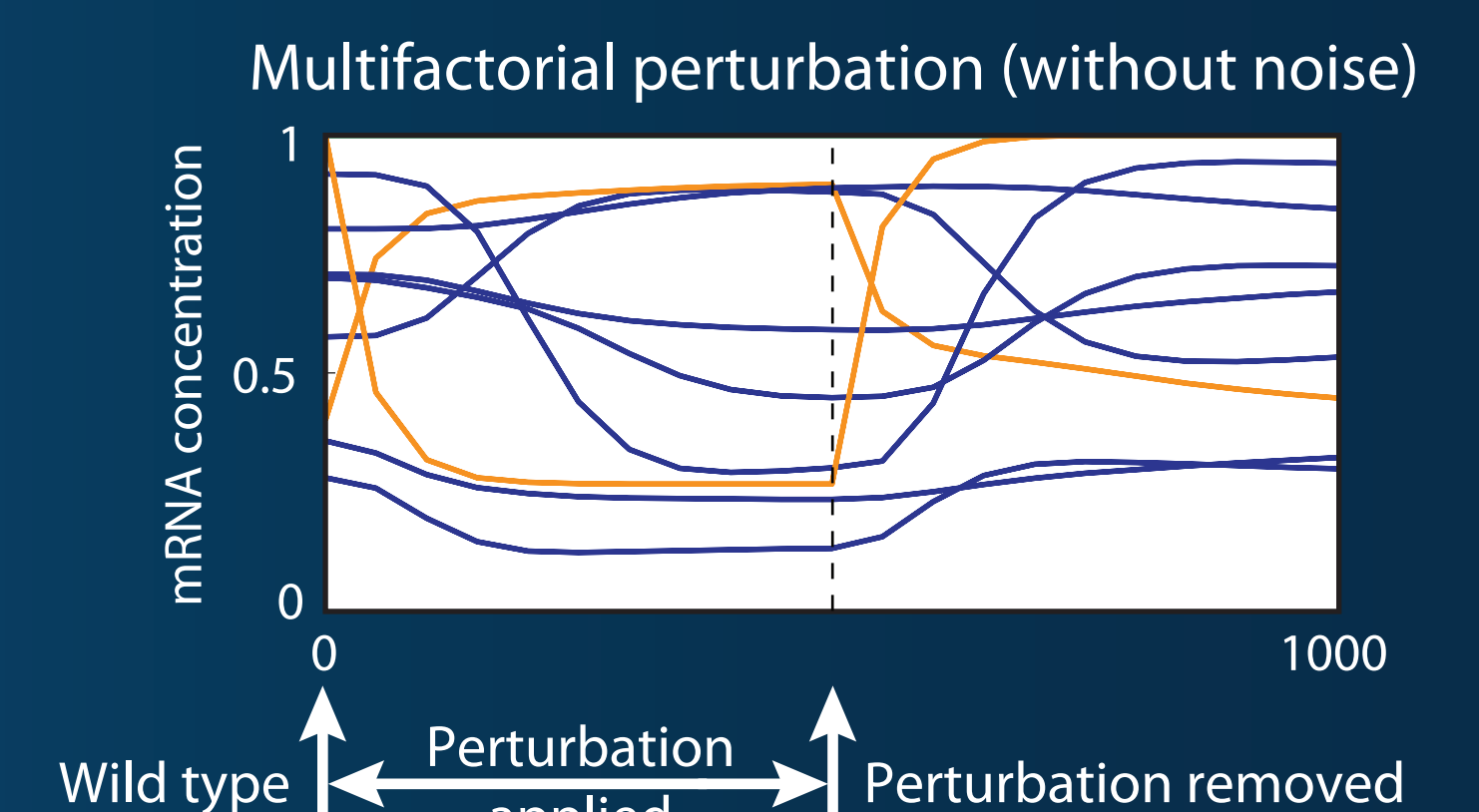
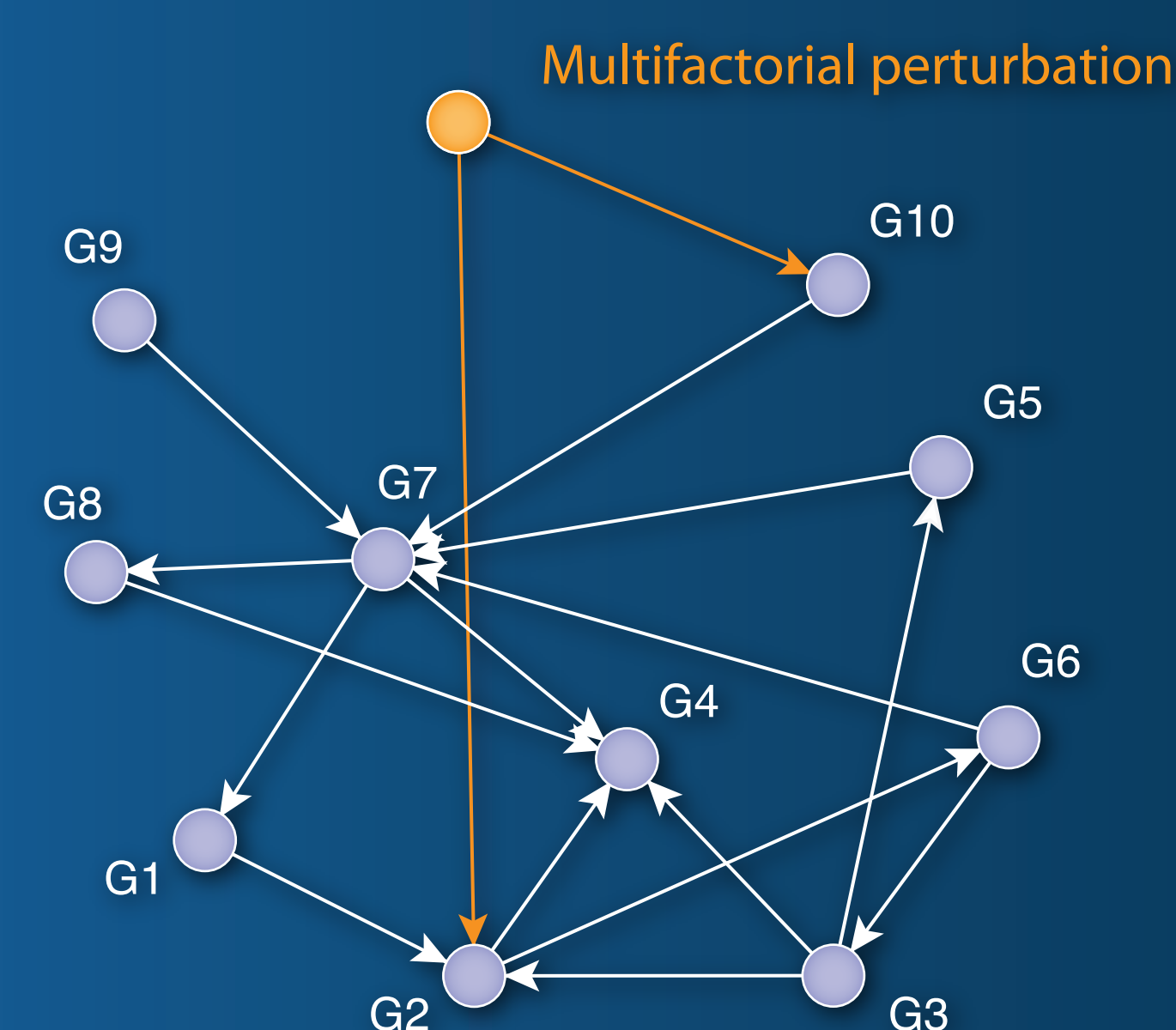
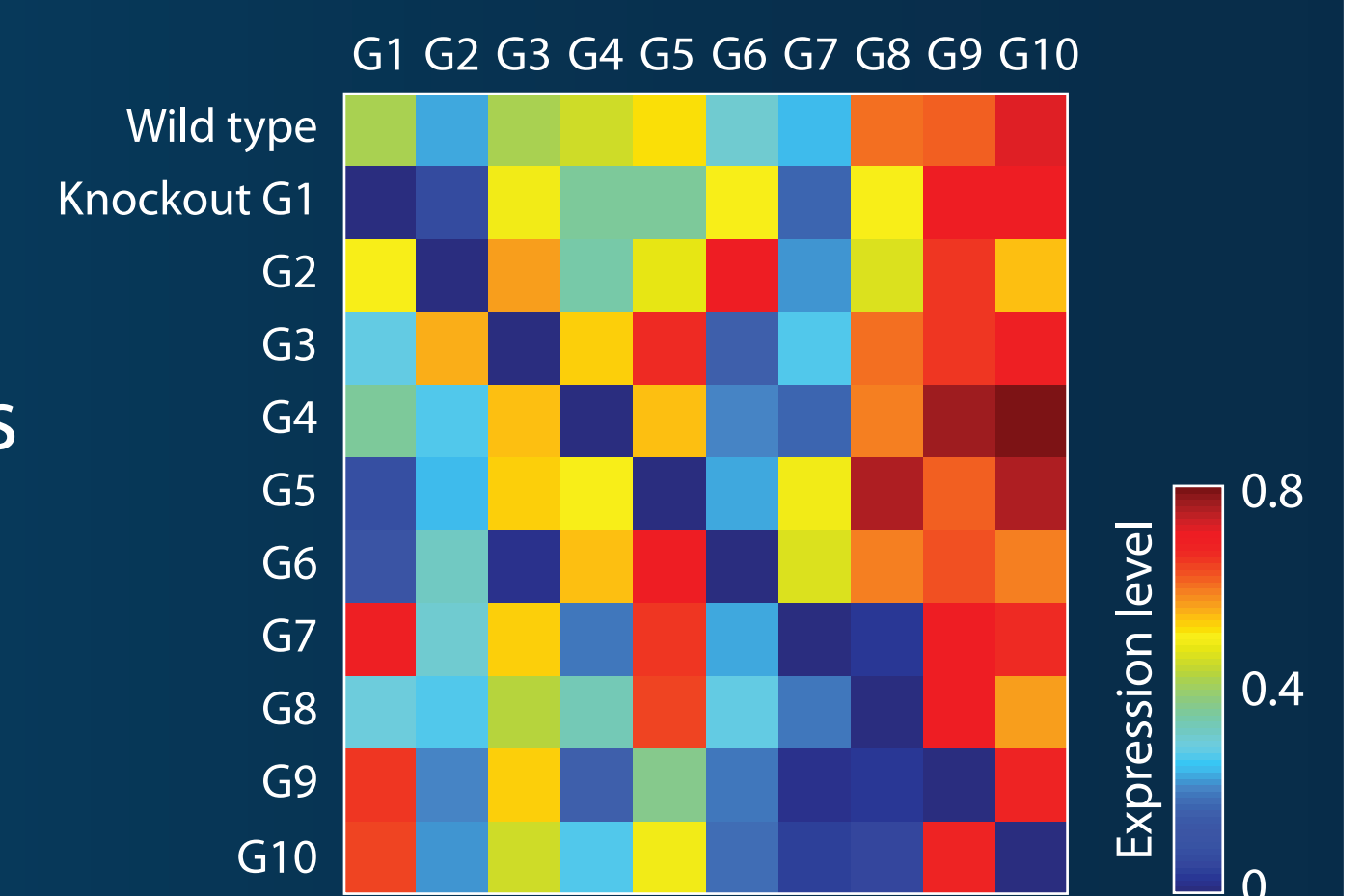
¹Laboratory of Intelligent Systems, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland; ²MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, Massachusetts, USA; ³Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.



Generate synthetic expression datasets

» Dynamical models are simulated to reproduce various biological experiments. Simulations can be done either **deterministically or stochastically (internal noise)**

» Types of experiments in GNW
Wild type, knockouts, knockdowns, multifactorial perturbations, time series



Abstract

Numerous methods have been developed for inference of gene regulatory networks from expression data, however, their strengths and weaknesses remain poorly understood. Accurate and systematic evaluation of these methods is hampered by the difficulty of constructing adequate benchmarks and the lack of tools for a differentiated analysis of network predictions on such benchmarks.

We present the new version (3.0) of **GeneNetWeaver (GNW)**, an open-source tool for **in silico benchmark generation** and **performance profiling of network inference methods**. GNW can be launched directly from any web browser thanks to the deployment technology Java Web Start.

GeneNetWeaver and the DREAM Initiative

We are using GNW to provide an **annual network inference challenge** for the DREAM project [2]. In the past three editions, **a total of 91 teams submitted about 800 network predictions** to evaluate the performance of their methods on GNW-generated benchmarks.

Generate network dynamical model

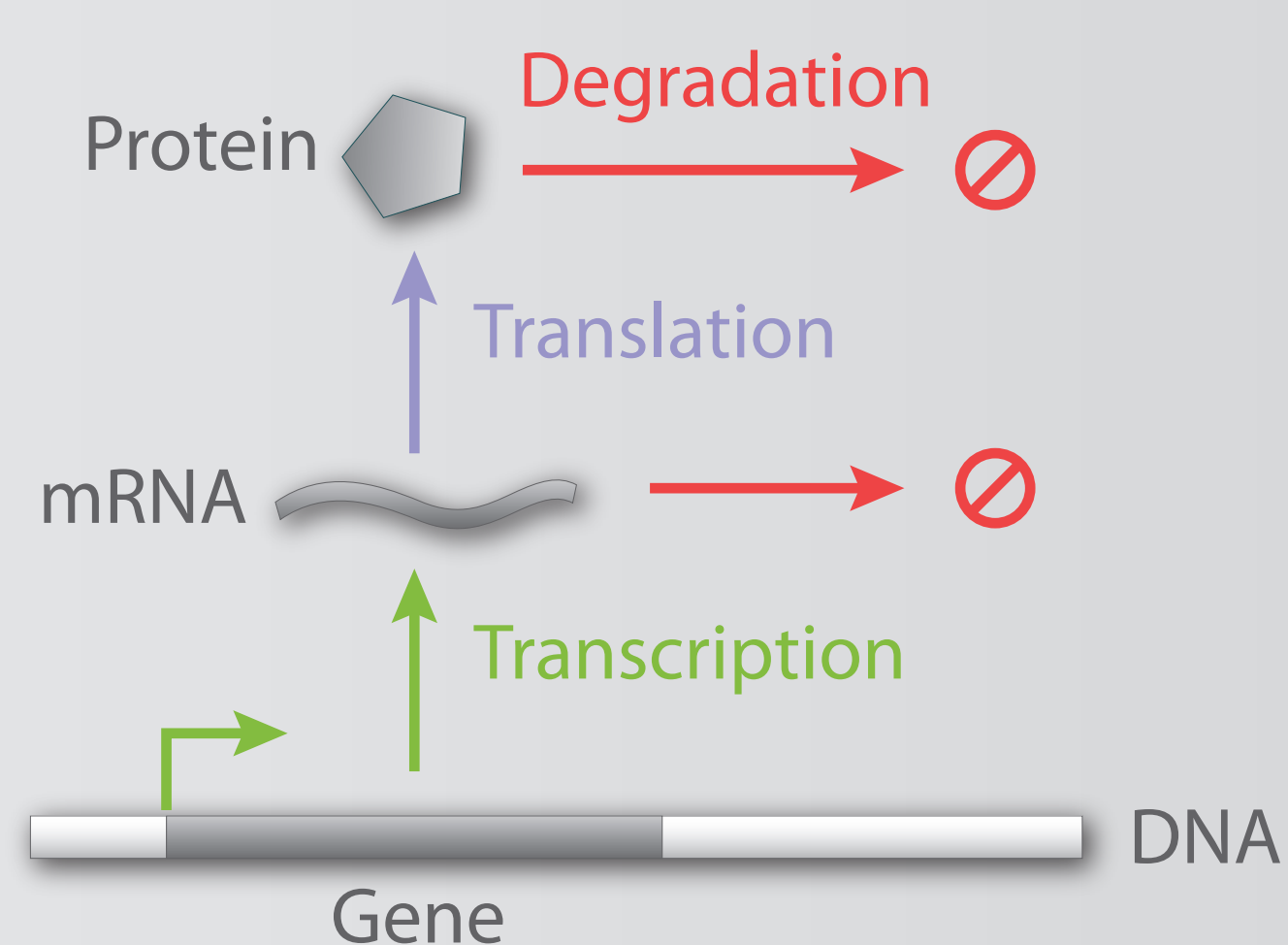
- » Realistic network structures are generated by extracting modules of size N from known biological interaction networks [1]
- » Transcriptional regulatory networks of *E. coli* and *S. cerevisiae* are used for generating the DREAM *in silico* challenges
- » The networks extracted are then endowed with dynamics using a standard approach based on thermodynamics
- » Transcription / translation and
- » Both independent and synergistic interactions are modeled

mRNA concentration x_i

$$\frac{dx_i}{dt} = m_i \cdot f_i(\mathbf{y}) - \delta_i \cdot x_i$$

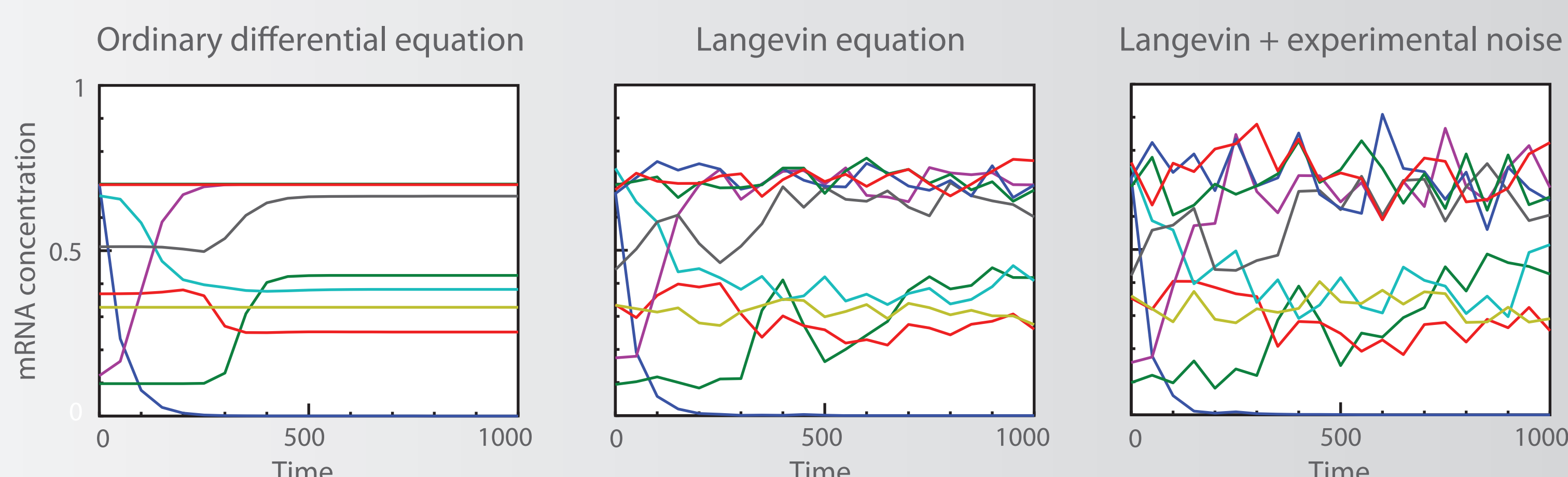
Protein concentrations y_i

$$\frac{dy_i}{dt} = r_i \cdot x_i - \lambda_i \cdot y_i$$



Models of noise

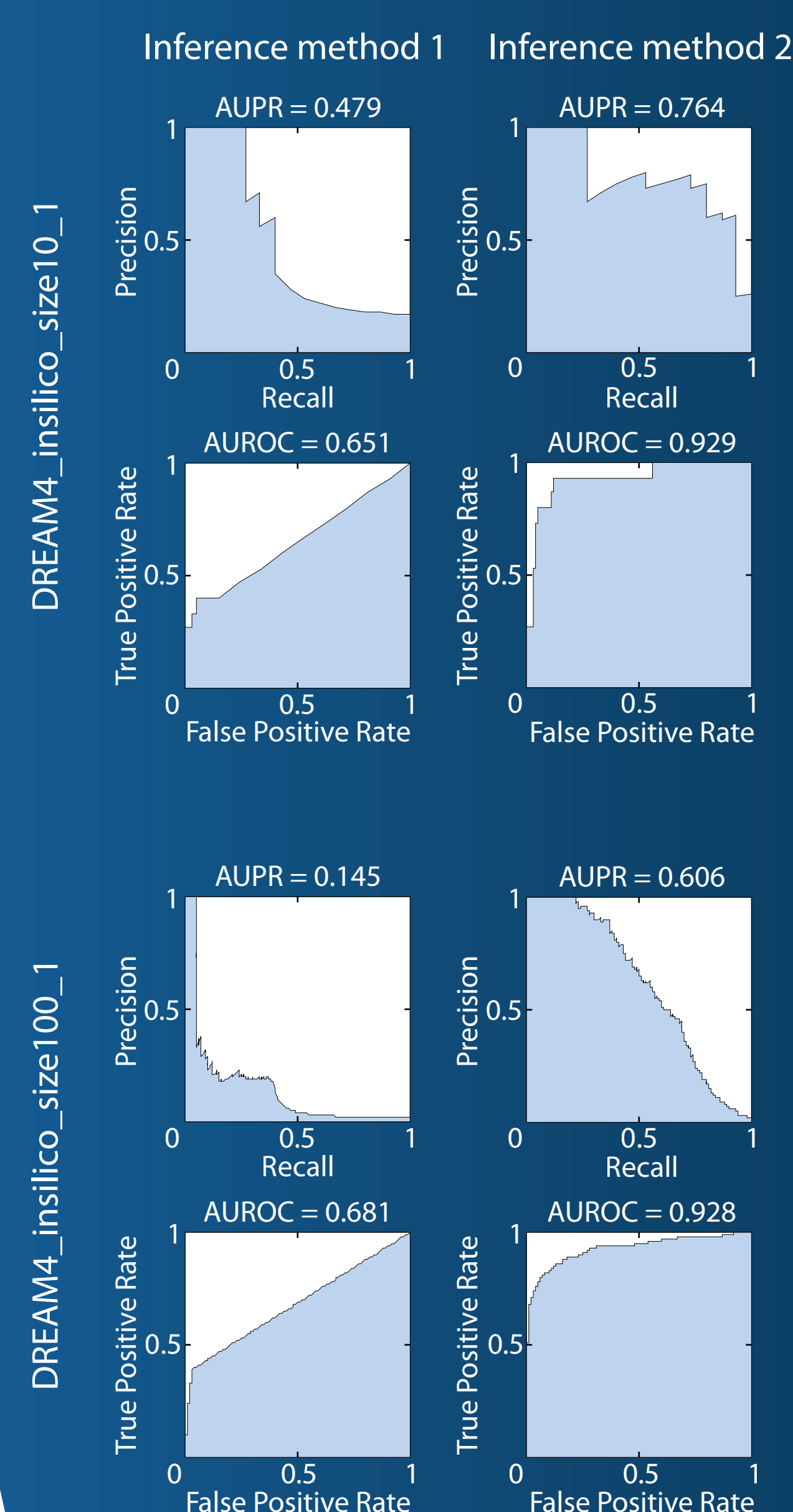
- » Internal noise in the dynamics of the networks
Chemical Langevin equation (Gillespie 2000, *J Chem Phys*)
- » Experimental noise (measurement error)
Model of noise in microarrays (Tu et al. 2002, *PNAS*)
- » Example: comparison of noise-free and noisy data



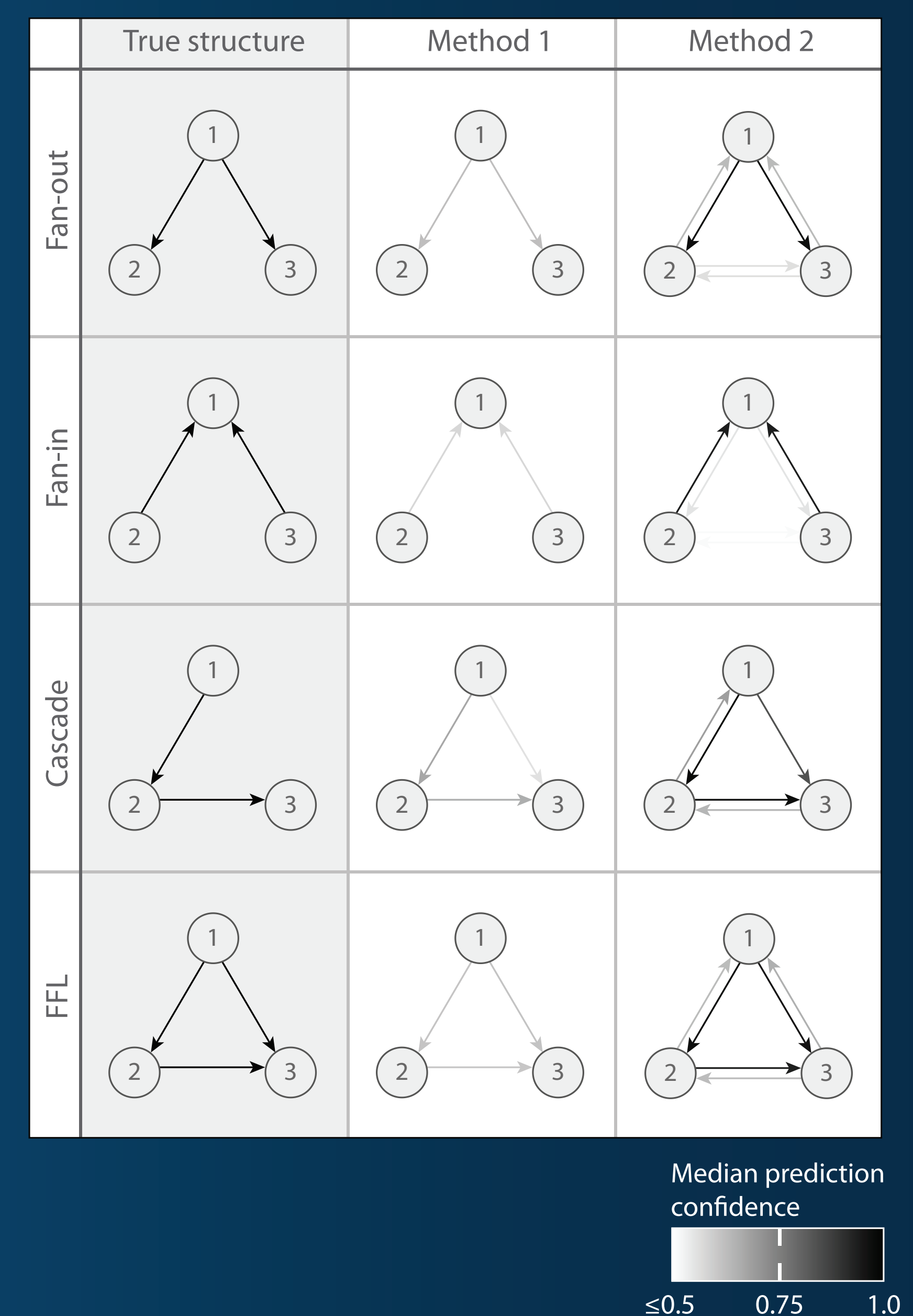
Performance profiling of network inference methods

- » GNW allows systematic evaluation of the predictions from different inference methods on *in silico* networks
- » **Comprehensive reports in PDF format are automatically generated**, including standard metrics used to assess the accuracy of network inference methods such as precision-recall and receiver operating characteristic (ROC) curves
- » The reports include **network motif analysis** [2], where the performance of inference methods is profiled on local connectivity patterns

Standard metrics (PR and ROC curves)



Motif prediction confidence



Selected publications

[1] Marbach D, Schaffter T, Mattiussi C, and Floreano D (2009) Generating realistic in silico gene networks for performance assessment of reverse engineering method. *J Comput Biol*, 16(2):229-239

[2] Marbach D, Prill RJ, Schaffter T, Mattiussi C, Floreano D, and Stolovitzky G (2010) Revealing strengths and weaknesses of methods for gene network inference. *PNAS*, 107(14):6286-6291