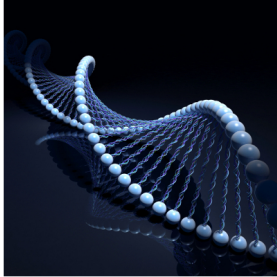


## LGBIO2010: Inference of gene regulatory networks

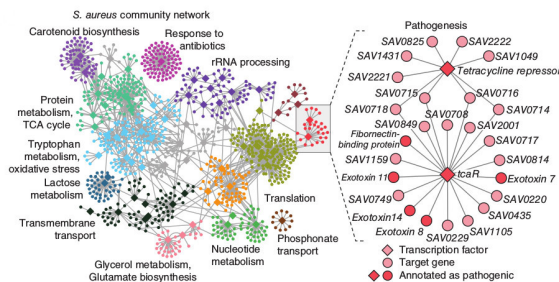
Pierre Dupont



UCL – ICTEAM

## Gene Regulatory Network (GRN)

- The set of all regulatory transcription interactions in a cell
- Typically represented as a graph with edges connecting
  - ▶ regulator = **transcriptor factor** (TF) (or the gene coding for it)
  - ▶ regulated gene = **target gene** (TG)



- TFs are usually **known**, TGs are at least partly **unknown**
  - otherwise consider each gene as potential "TF" or TG

*Illustration from [Marbach et al., 12]*

## Module-based versus direct inference

## Module-based inference

- 1 search for **clusters (= modules)** of genes that exhibit a similar expression behavior
- 2 infer the transcriptional program of **each module**
  - ▶ better at finding interactions between regulators and targets with less similar expression profiles

## Direct inference

- look for the regulators of **each gene**
  - ▶ better at finding the target(s) of regulators dedicated to one or a few genes

## Expression-based versus integrative inference

## Expression-based methods

- take **gene expression data** as sole input
- useful for organisms for which there is little regulatory information available

## Integrative methods

- complement gene expression with **additional information**: motif data or protein-DNA interaction data
- more likely to predict true positive interactions
- the added value depends on the quality of the additional information

## Global versus query-driven inference

### Global inference

- provide a **general view** of the active GRN
- identify **large pathways** that consist of many genes

### Query-driven inference

- search for genes that are co-expressed, in a condition-dependent way, with a predefined set of **query genes**
- find **local regulatory** interactions

## Supervised versus unsupervised inference

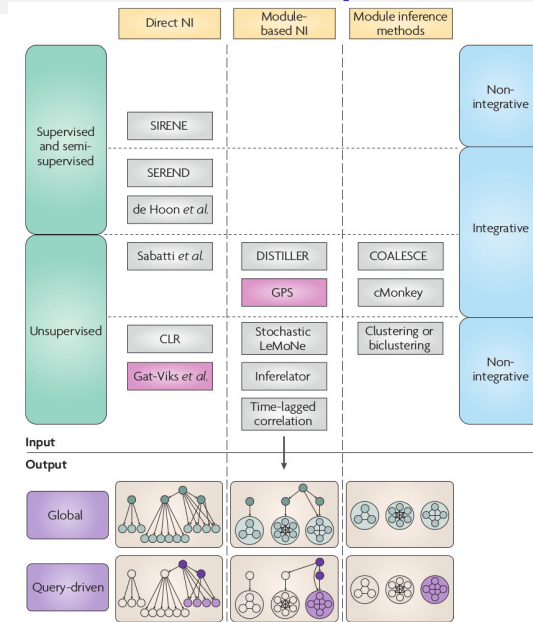
### Supervised inference

- treat network inference as **multiple classification problems**
  - which target genes are regulated by a **given transcription factor**
- require a **training set** of known interactions between TFs and TGs and expression data
- each learned classifier is then used to predict whether further genes should be considered as target of a given TF

### Unsupervised inference

No such training set is available

## A taxonomy of GRN inference [De Smet and Marchal, 10]







## LASSO-based inference [Haury et al., 12]

	gene 1	gene 2	...	gene p
sample 1	$x_{1,1}$	$x_{1,2}$	...	$x_{1,p}$
...	...	...	...	...
sample n	$x_{n,1}$	$x_{n,2}$	...	$x_{n,p}$

- each gene  $G_i$  is separately regressed on the **TFs**  
(a known subset of columns)
- linear regression with L1-regularization
 
$$\hat{\mathbf{w}} = \operatorname{argmin}_{\mathbf{w}} \|\mathbf{x}_{i*} - \mathbf{X}_{TF} \mathbf{w}\|_2^2 + \lambda \|\mathbf{w}\|_1$$
- for a specific  $\hat{\mathbf{w}}$ ,  $\hat{\mathbf{w}}_j \neq 0$  means dependence  $TF_j \rightarrow G_i$
- direct, non-integrative, global and unsupervised inference
  - Expression data + list of known TFs is the only required information

## Further Reading

-  [De Smet, R. and Marchal, K. \(2010\)](#)  
*Advantages and limitations of current network inference methods*  
[Nature Reviews](#), Vol. 8, pp. 717-729.
-  [Haury, A.-C., Mordelet, F., Vera-Licona, P. and Vert, J.P. \(2012\)](#)  
*TIGRESS: Trustful Inference of Gene Regulation using Stability Selection*,  
[BMC Systems Biology](#), Vol. 6 (145).
-  [Marbach, D., Costello, J.C., Küffner, R. Vega, N.M., Prill, R.J., Camacho, D., Allison, K.R., The DREAM5 Consortium, Kellis, M., Collins, J.J., and Stolovitzky, G. \(2012\).](#)  
*Wisdom of crowds for robust gene network inference*,  
[Nature Methods](#), Vol. 9 (8), pp. 796-804.
-  [Ruyssinck, J., Huynh-Thu, V.A., Geurts, P., Dhaene, T., Demeester, P. and Saeys Y. \(2014\)](#)  
*NIMEFI: Gene Regulatory Network Inference using Multiple Ensemble Feature Importance Algorithms*  
[PLOS One](#), Vol. 9 (3), e92709.