

# Higher-order interactions in single-cell expression data

What they are, and how to estimate them

---

Abel Jansma

 @Abelaer

16.03.2021

PhD:

@IGMM with Chris Ponting

@Particle Physics with Luigi del Debbio

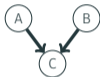
@Informatics with Ava Khamseh

# Spoiler Alert

Gene dynamics

Additive:  
 $C = A + B$

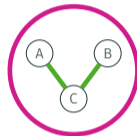
Causal graph



Correlation



Mutual Information

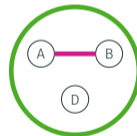
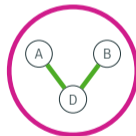
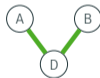
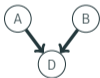


Interaction



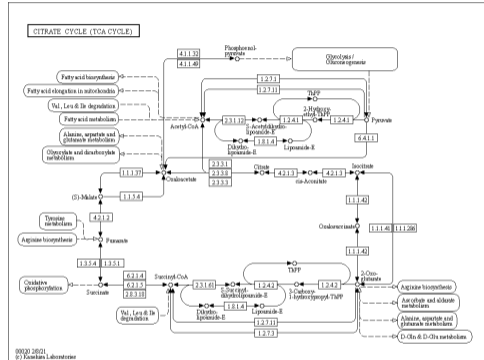
Multiplicative:

$D = A \times B$



# Biological Networks

- Molecular Cell Biology:
  - Pathways
  - Causal
  - Interactions
- Biologically meaningful
- Experiments & Experts
- Is it possible to learn these networks from observational data?



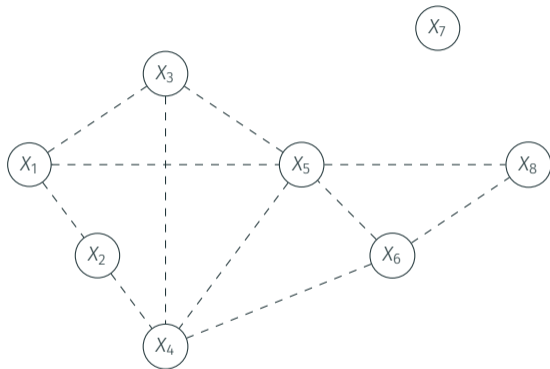
source: KEGG pathways



Source: UMAP documentation

# Reconstructing Genetic Networks

- Expression of genes  $X_i \in X$
- What does each  $X_i - - - - X_j$  represent?
- Existing methods:
  - Correlation
  - Mutual information
  - Causal graph
- non-parametric



# Non-parametric methods hide dynamics

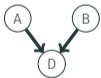
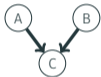
- Binary transcription factors  $A$  and  $B$ 
  - Affect  $C$  **independently**:  $C = A + B$
  - Affect  $D$  as **bound** complex:  $D = A \times B$

Gene dynamics

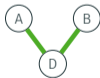
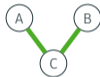
Additive:  
 $C = A + B$

Multiplicative:  
 $D = A \times B$

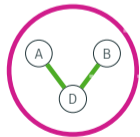
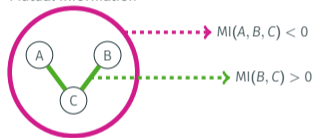
Causal graph



Correlation



Mutual Information



- **Do not differentiate between dynamics!**

## Misspecified Models introduce bias

- To get dynamics, you might try to fit an explicit model.
- Generate data as  $X_1, X_2 \sim \text{Unif}(-1, 1)$ ,

$$y = x_1 + x_2 + x_1x_2$$

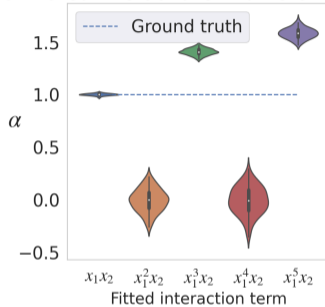
- Consider fitting a model of the form:

$$\hat{y} = x_1 + x_2 + \alpha x_1^n x_2$$

- Even worse:** If there is a hidden  $X_3$  such that

$$\begin{aligned}y &= x_1 + x_2 + \alpha x_1 x_2 + \beta x_1 x_2 x_3 \\ &= x_1 + x_2 + (\alpha + \beta x_3) x_1 x_2\end{aligned}$$

Estimated interactions in different models



## Model-free Interactions: Intuition

- Let's take a step back: What do we actually **mean** by **interaction**?
- The **effect**  $l_i$  of gene  $X_i \in X$  on an outcome  $Y$ :

$$l_i = \left. \frac{\partial Y}{\partial X_i} \right|_{\underline{X}=0}$$

- Two genes  $X_i$  and  $X_j$  **interact** when expression of  $X_j$  changes the effect of  $X_i$  on  $Y$ :

$$l_{ij} = \left. \frac{\partial l_i}{\partial X_j} \right|_{\underline{X}=0} = \left. \frac{\partial^2 Y}{\partial X_j \partial X_i} \right|_{\underline{X}=0}$$

- A third gene  $X_k$  can modulate this interaction, which we call a **3-point interaction**:

$$l_{ijk} = \left. \frac{\partial l_{ij}}{\partial X_k} \right|_{\underline{X}=0} = \left. \frac{\partial^3 Y}{\partial X_k \partial X_j \partial X_i} \right|_{\underline{X}=0}$$



## Model-free Interactions: Definition

- Interactions are derivatives of an outcome:  $l_{ij} = \frac{\partial^2 Y}{\partial X_j \partial X_i} \Big|_{\underline{X}=0}$
- Most general outcome:  $\log p(X)$
- For binary genes, we can calculate this!

$$l_{ij} = \frac{\partial^2 \log p(X)}{\partial X_j \partial X_i} \Big|_{\underline{X}=0} = \log \frac{p(X_i = 1, X_j = 1 | \underline{X} = 0) p(X_i = 0, X_j = 0 | \underline{X} = 0)}{p(X_i = 1, X_j = 0 | \underline{X} = 0) p(X_i = 0, X_j = 1 | \underline{X} = 0)}$$

- **Symmetric**:  $l_{ij} = l_{ji}$ .
- Conditionally independent genes do not interact:  $X_i \perp\!\!\!\perp X_j | \underline{X} \implies l_{ij} = 0$ .
- If  $\underline{X} = \emptyset$ , **log-odds ratio**.
- **Model-independent** - can be directly estimated from expression data.
- It can be generalised to an  **$n$ -point interaction** by taking  $n$  derivatives of  $\log p(X)$ .

# What are higher-order interactions?

- 1-pt  $l_i = \frac{\partial \log p(X)}{\partial X_i} \Big|_{\underline{X}=0}$ 
  - **Innate** tendency to be expressed.
- 2-pt  $l_{ij} = \frac{\partial l_i}{\partial X_j} \Big|_{\underline{X}=0}$ 
  - How  $X_j$  changes  $X_i$ 's tendency to be expressed
  - Vanilla **gene regulation**
- 3-pt  $l_{ijk} = \frac{\partial l_{ij}}{\partial X_k} \Big|_{\underline{X}=0}$ 
  - '**Combinatorial**' gene regulation
  - Expression 'epistasis'

# Model-free Interactions: in practice

- The TF-models from before:

Gene dynamics

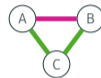
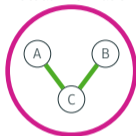
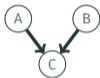
Causal graph

Correlation

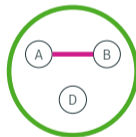
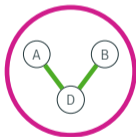
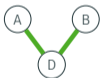
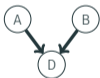
Mutual Information

Interaction

Additive:  
 $C = A + B$



Multiplicative:  
 $D = A \times B$



- A 2-point interaction for independent transcription factors.
- A 3-point interaction for bound complex of transcription factors.

Correlation **is not** Causation

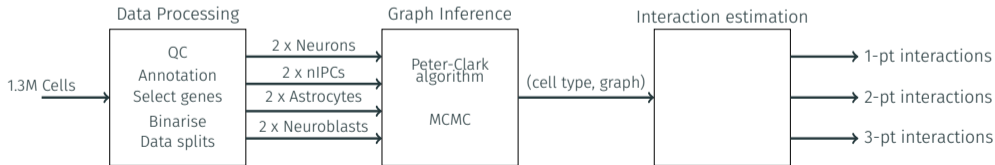
*is not*

*is not*

Interaction

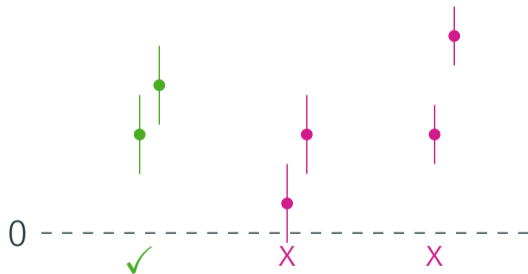
# Pipeline

- scRNA-seq: 1.3M embryonic (E18.5) mouse brain cells
- 4 cell types:
  - Inhibitory neurons, olfactory bulb
  - Neuronal intermediate progenitor cells (nIPCs)
  - Astrocyte-like, dorsal midbrain
  - Neuroblasts, olfactory bulb
- Two replicates of 10k cells.



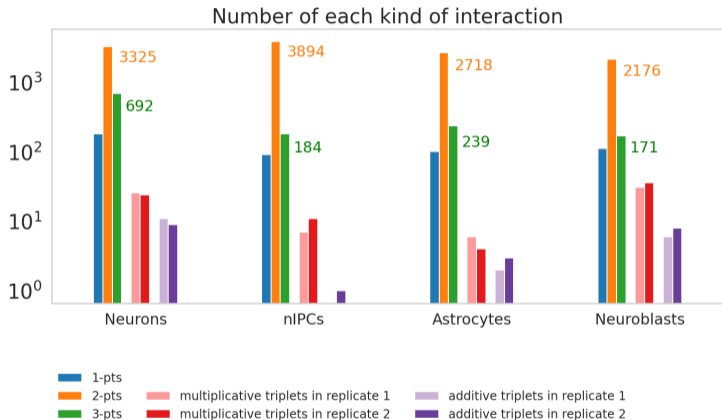
# Significance

- When do we call a result significant?
- Bootstrap a 95% confidence interval
- Interaction should:
  - Be significant in both replicates
  - Have overlapping 95% confidence intervals



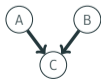
# Results: we find hundreds of interactions

- **1-pt:** Effects
- **2-pt:** Interaction
- **3-pt:** Higher-order
- **Multiplicative:**  
Binding TFs
- **Additive:**  
Independent TFs



# Biological validation

- Problem: Gold standards are pairwise.
- **Biological questions**
- **1-pts:**
  - Among all genes that are expressed highly, are the ones with a strong 1-point interactions more likely to be housekeeping genes than similarly expressed genes with a weak 1-point?
- **2-pts:**
  - Are pairs that interact enriched in protein-protein interactions as compared to pairs that correlate?
- **3-pts:**

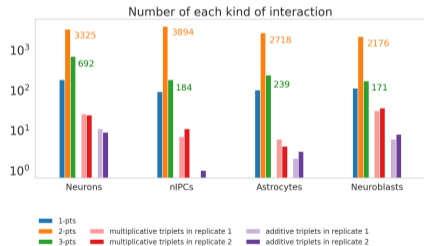


- In a multiplicative triplet, are the parents on the graph more likely to have a protein-protein interaction than the parents in an additive triplet?
- Are the upstream genes more likely to be transcription factors than the downstream genes?



# Summary

- Higher-order interactions reveal hidden dynamics
- No model bias
- We find hundreds of higher order interactions in the mouse brain
- They differ across cell types



- How do these interactions differ between cell types?
- Differential expression → differential regulation.
- Can we predict novel interactions?

# Collaborators

Supervisors:

Chris Ponting, Luigi Del Debbio, Ava Khamseh

Ponting group:



Not pictured:

- Joshua Dibble
- Louise Docherty
- Catherine Heath
- Jenna Stephen
- Kelsey Tetley-Campbell