Higher-order interactions in single-cell expression data

European Mathematical Genetics Meeting 2021

Abel Jansma, PhD candidate at the University of Edinburgh

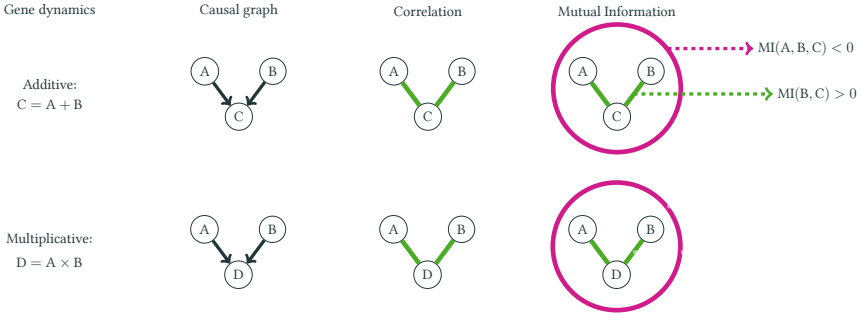
23.04.2021

@Abelaer

Supervised by:
Chris Ponting @ Institute of Genetics and Cancer
Luigi Del Debbio @ Higgs Centre for Theoretical Physics
Ava Khamseh @ School of Informatics

Non-parametric methods hide dynamics

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



- Do not differentiate between dynamics!

Model-free Interactions: Intuition

- Let's take a step back: What do we actually mean by interaction?
- The effect I_i of gene $X_i \in X$ on an outcome Y: 1-pt interaction

$$I_i = \frac{\partial Y}{\partial X_i}\Big|_{\underline{X}=0}$$

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

$$I_{ij} = \frac{\partial I_i}{\partial X_j}\Big|_{\underline{X}=0} = \frac{\partial^2 Y}{\partial X_j \partial X_i}\Big|_{\underline{X}=0}$$

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

$$I_{ijk} = \frac{\partial I_{ij}}{\partial X_k}\Big|_{\underline{X}=0} = \frac{\partial^3 Y}{\partial X_k \partial X_j \partial X_i}\Big|_{\underline{X}=0}$$

Model-free Interactions: Definition

- Interactions are derivatives of an outcome: $\mathrm{I}_{ij} = \frac{\partial^2 \mathrm{Y}}{\partial \mathrm{X}_j \partial \mathrm{X}_i}\Big|_{\underline{\mathrm{X}}=0}$
- Most general outcome: $\log \mathrm{p}(\mathrm{X})$
- For binary genes, we can calculate this [1]!

$$\mathrm{I}_{ij} = \frac{\partial^2 \log \mathrm{p}(\mathrm{X})}{\partial \mathrm{X}_j \partial \mathrm{X}_i}\Big|_{\underline{\mathrm{X}}=0} = \log \frac{\mathrm{p}\left(\mathrm{X}_i = 1, \mathrm{X}_j = 1 \mid \underline{\mathrm{X}} = 0\right) \mathrm{p}\left(\mathrm{X}_i = 0, \mathrm{X}_j = 0 \mid \underline{\mathrm{X}} = 0\right)}{\mathrm{p}\left(\mathrm{X}_i = 1, \mathrm{X}_j = 0 \mid \underline{\mathrm{X}} = 0\right) \mathrm{p}\left(\mathrm{X}_i = 0, \mathrm{X}_j = 1 \mid \underline{\mathrm{X}} = 0\right)}$$

- Symmetric: $\mathrm{I}_{ij} = \mathrm{I}_{ji}$.
- Conditionally independent genes do not interact: $\mathrm{X}_i \perp\!\!\!\perp \mathrm{X}_j \mid \underline{\mathrm{X}} \implies \mathrm{I}_{ij} = 0$.
- If $\underline{\mathrm{X}} = \emptyset$, 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 $\log \mathrm{p}(\mathrm{X})$.

[1] S. Beentjes & A. Khamseh, Phys. Rev. E 102, 053314

What are genetic higher-order interactions?

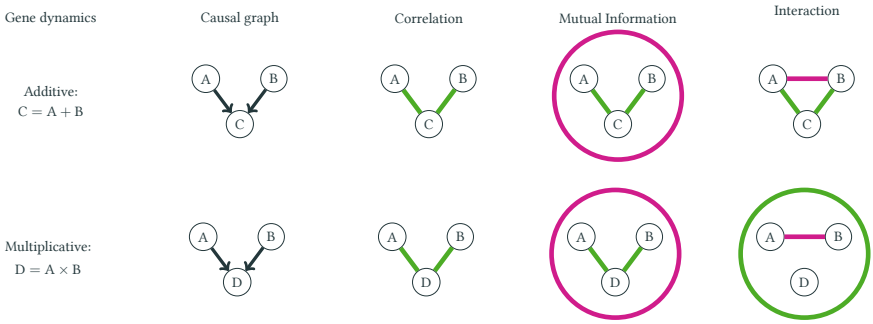
- 1-pt $I_i = \frac{\partial \log p(X)}{\partial X_i}\Big|_{\underline{X}=0}$
  - Innate tendency to be expressed.

- 2-pt $I_{ij} = \frac{\partial I_i}{\partial X_j}\Big|_{\underline{X}=0}$
  - How X_j changes X_i's tendency to be expressed
  - Isolated gene regulation

- 3-pt $I_{ijk} = \frac{\partial I_{ij}}{\partial X_k}\Big|_{\underline{X}=0}$
  - 'Combinatorial' gene regulation
  - Expression 'epistasis' (Saswati Saha this morning!)

Model-free Interactions: in practice

- The TF-models from before:



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

Interactions are something new

Correlation is not Causation

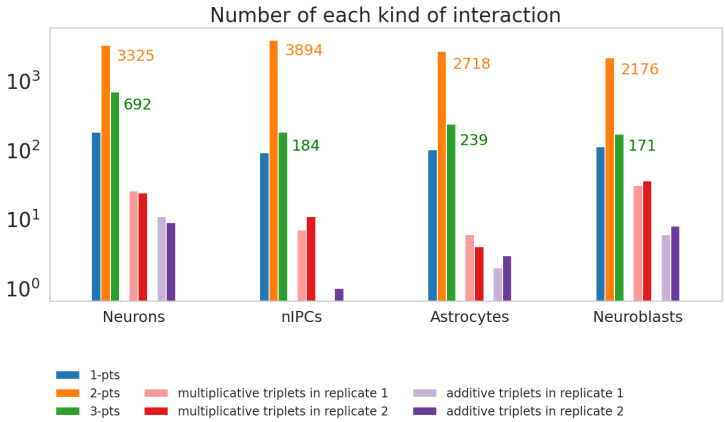
is not

is not

Interaction

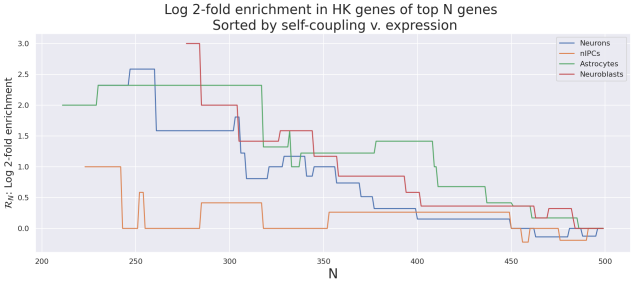
Results: we find hundreds of interactions

- 10X 1.3M E18.5 mouse brain cells
- 10k cells of 4 types each, 500 genes
- 1-pt: Effects
- 2-pt: Interaction
- 3-pt: Higher-order
- Multiplicative: Binding TFs
- Additive: Independent TFs



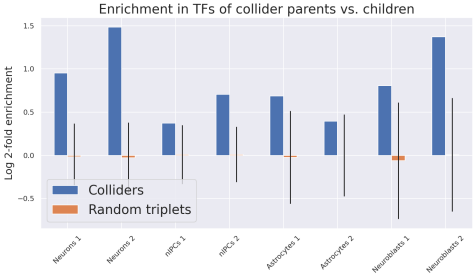
Results: 1-point interactions

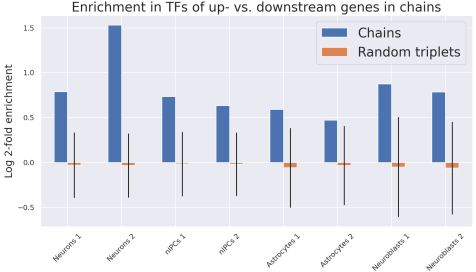
Q: Are 1-point interactions an indication of housekeeping genes?

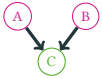


- Yes! Even more so than expression.

Results: Connected triplets





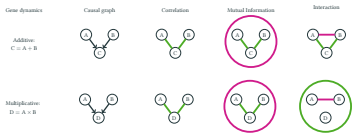


Parents are more likely to be transcription factors than children.

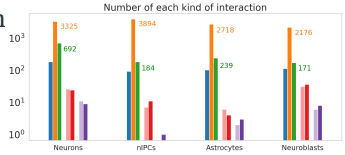
Upstream genes are more likely to be transcription factors than downstream genes.

Summary

- Higher-order interactions reveal hidden dynamics



- We find hundreds of higher order interactions in the mouse brain
- They differ across cell types



- They contain biological information about the proteins

A look ahead

- How to validate 2-point interactions?
- Can we integrate our predictions with CHIP-seq data?
- Differential expression $\rightarrow$ 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
- Sjoerd Beentjes