The Molecular Language of the Body:

Translating knowledge of cell-to-cell communication molecules from immunology to neuroscience with RNAseq data

Lucia Guerri, Miranda Darby, Jingwen Gu, Amanda Bell, Van Truong, Saba Nafees

1st Women-led Biodata Science NCBI Hackathon (2019)

Team Lead

Lucia Guerri → mom of scientific idea

The Professionals

Miranda Darby → Supervisor and Review Committee

Amanda Bell → Dataset magician

Saba Nafees → Machine Learning magician & writer

Jingwen Gu → Stats magician & Sysadmin

Van Truong → Drawing magician

WHY?

- The study of cell-to-cell communication molecules has been particularly strong in immunology, while remaining largely understudied in brain.
- The same pairs of communication molecules are employed by several tissues throughout the body.

- By translating knowledge of cell-to-cell communication molecules (and their conserved intracellular signaling) from immunology to neuroscience with RNAseq data, we can fasten hypothesis generation in neuroscience.

GOALS:

Hackathon:

Day 1:

- ✓ **Brainstorm** about best **strategies** to address the scientific question (“Translate knowledge of cell-to-cell communication molecules from immunology to neuroscience”)
- ✓ **Breakthrough** theoretical project into concrete technical pipeline
- ✓ Identify best RNAseq **datasets** (and specie: Hs vs Mm)
- ✓ Define **pipeline** to generate a (non-thorough) **database** of **“PAIRS of cell-to-cell communication molecules from immunology”**

Day 2:

- Generate a (non-thorough) **database** of **“PAIRS of cell-to-cell communication molecules from immunology”**
- **Brainstorm** about
  - **Scoring system** for supervised pipeline
  - **Machine-learning** for partially-unsupervised tool (adapt WGCNA?)

Day 3:

- Generate working prototype of **supervised pipeline**

Stretch:

- Generate working prototype of **machine-learning partially-unsupervised tool**

HOW?

- Use a RNAseq input file for i une cell expression data
- Filter GO terms for m oteins only
- Pair the molecu wn relationships
- Verify outp
  - High sco known pairs of cell-to-cell communication molecules
- Our tool will identify brain cell molecule pairs with high interaction score

OLD SLIDE

BRAINSTORM TOPICS:

- Binary Y/N for gene expression: For first pipeline only. Discuss cutoff (avoid complex normalization if possible) across da
- scRNAseq as input: use expr clusters vs. single cell data. If clusters, then use mean of top quartile/ value for each gene, to address technical loss of data. If single cell, th to address the cell pairs, and how to process that potentially heavy data
- Machine learning: discover new potential pairs of molecules of cell-to cell-communication, instead of being limited by current knowledge to guide the scores. Discuss normalization strategy and challenges to exploit the actual level of expression, instead of binary Y/N expression.
- Soluble signals: how to deal with soluble signals and their receptors (membrane-bound and cytoplasmic) in this pipeline.
- Mouse vs human data: less “noise” vs higher N (=power)

OLD SLIDE