TraIN:

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 → Mentor and Review Committee

Amanda Bell → Dataset magician

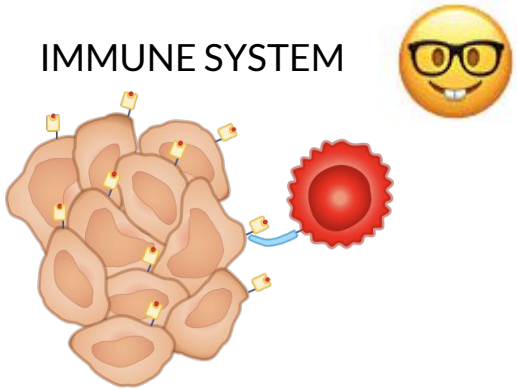
Saba Nafees → Mathematics magician & Writer

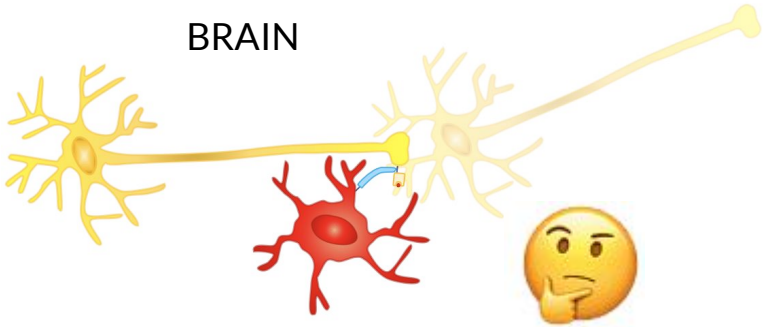
Jingwen Gu → Stats magician & Sysadmin

Van Truong → Data & Scientific Illustrator magician

WHY?

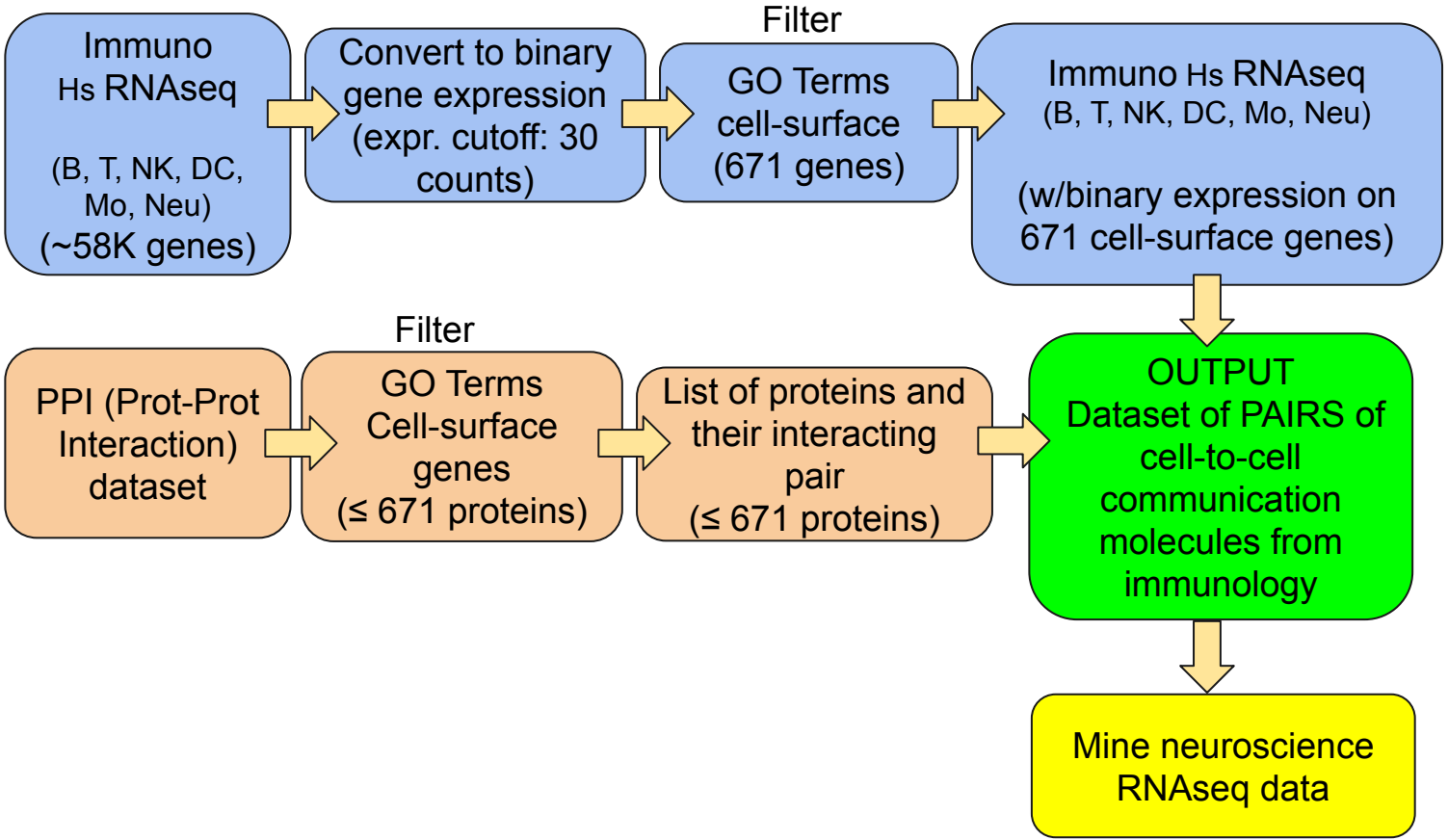
- The study of cell-to-cell communication molecules has been particularly strong in immunology, while remaining largely understudied in brain.
- The same pairs of communication molecules are employed by several tissues throughout the body.





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

WORKFLOW:


Immuno
Hs RNAseq
(B, T, NK, DC, Mo, Neu)
(~58K genes)
Convert to binary gene expression (expr. cutoff: 30 counts)
Filter
GO Terms cell-surface (671 genes)
Immuno Hs RNAseq
(B, T, NK, DC, Mo, Neu)
(w/binary expression on 671 cell-surface genes)
PPI (Prot-Prot Interaction) dataset
Filter
GO Terms Cell-surface genes (≤ 671 proteins)
List of proteins and their interacting pair (≤ 671 proteins)
OUTPUT
Dataset of PAIRS of cell-to-cell communication molecules from immunology
Mine neuroscience RNAseq data

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”)
- ✓ **Break down** the theoretical project into concrete technical pipeline
- ✓ Identify best RNAseq human and murine **datasets**
- ✓ Define **pipeline** to generate a **database** of **“PAIRS of cell-to-cell communication molecules from immunology”**

Day 2:

- 😱😀👍 Generate a **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**