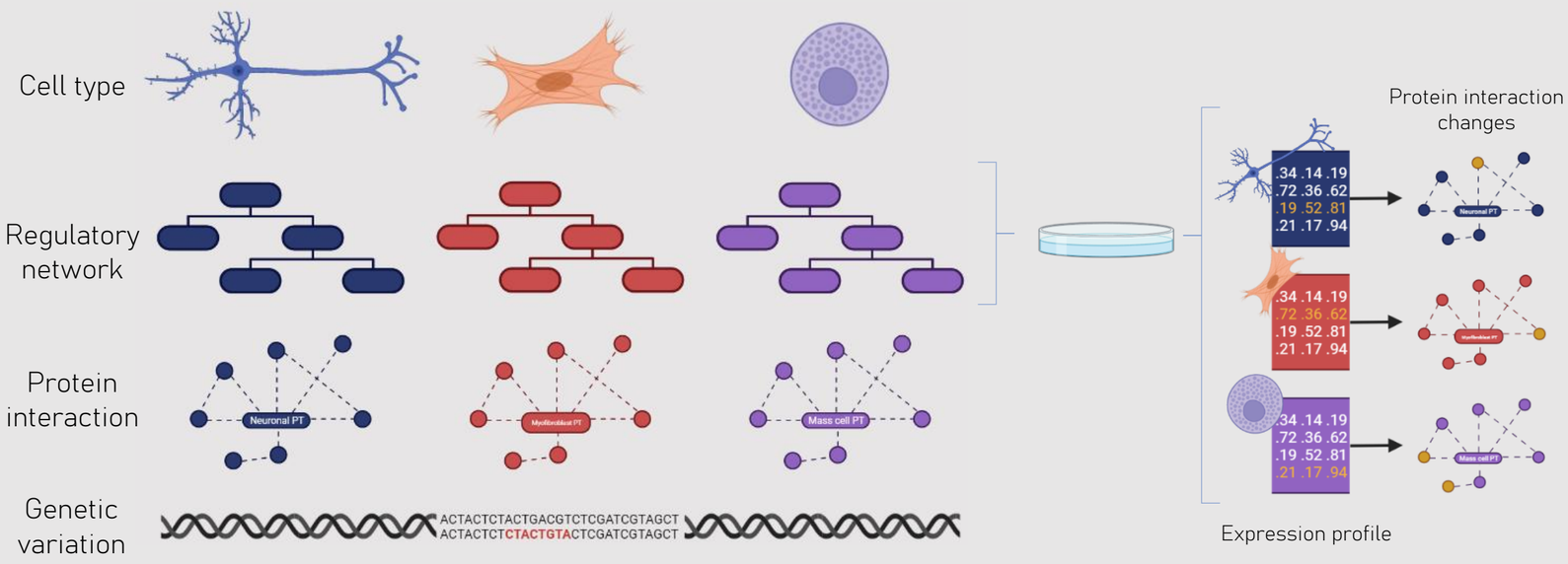
Using Single Cell Gene Expression Data to Study the Biology of Cell Types

Emiliano Navarro Garre || Current Topics in Bioinformatics

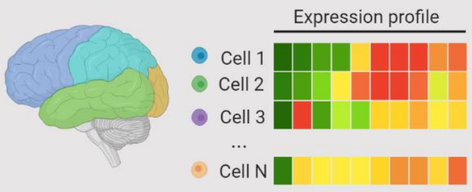
Where do we start?



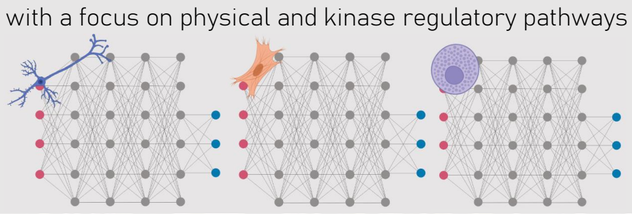
- Different **cell types are given by regulatory networks and protein interactions.**
- **Single cell techniques** allow us to sequence different cell types from a single heterogeneous sample in order to know **which genes are expressed** by these cells **under *in vivo* conditions**.
- **Protein interaction networks** can **change** between different cell types and **are described** in the **expression profiles** for each cell type.
- **Genetic variations** can **cause** different **effects** on **cell expression profiles** and **protein interaction networks**, depending on the cell type.

Objectives

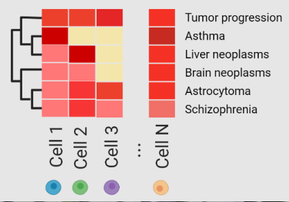
1. Create expression profiles for each human tissue



2. Describe cell type-specific interaction and/or regulatory networks with a focus on physical and kinase regulatory pathways



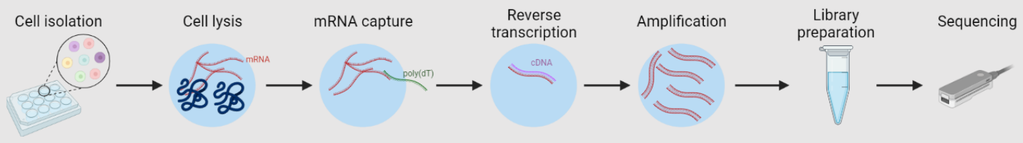
3. Explore associations of cell types with human diseases



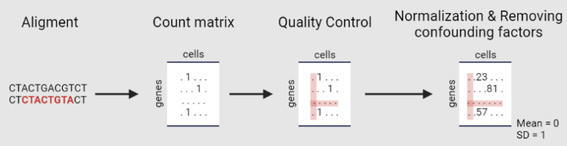
Methods

- **Workflow for single-cell RNA-sequencing experiment**

1. Experimental method



2. Data processing



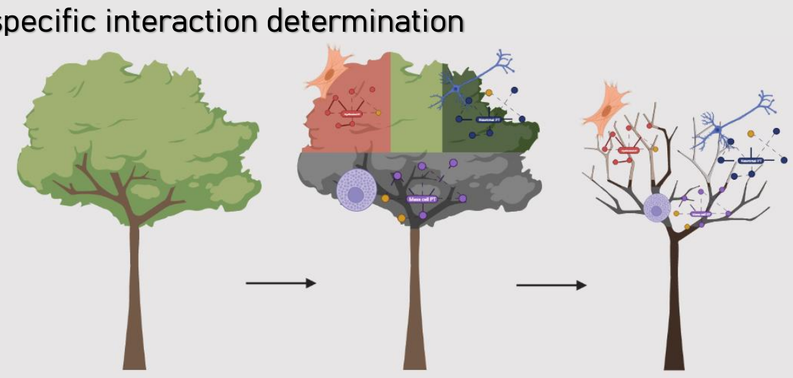
3. Data analysis

- The highly variable genes indicate the specificity. These would be the genes that are responsible for the differences between each cell type.
- The Principal Component Analysis (PCA) help us to see which combination of genes best explain the variability in expression data.
- The clustering consists in group cells with a similar expression pattern together.

Once we have classified the different cell types, we carry out the analyses shown above to determine which gene clusters are enriched, construct the protein networks...

- This analysis will be carried out with different tissue cells or *in vitro* organoids, such as cerebral organoids.

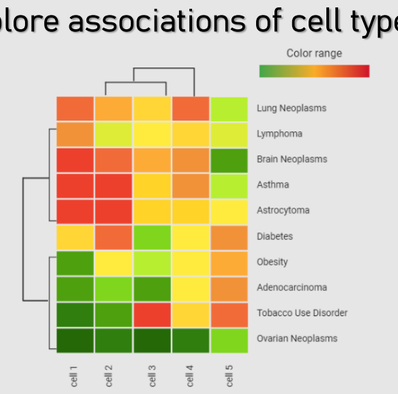
- **Cell type-specific interaction determination**



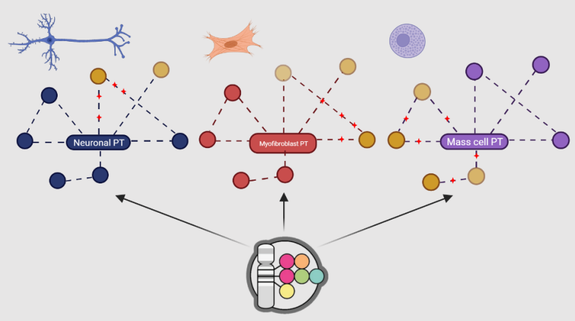
Observe which gene modules are expressed in each cell type from scRNA-seq data and compare with the theoretical network, pruning the network and weighting for each of the interactions according to their scores.

It will be necessary perturbation experiments to validate PPI networks.

- **Explore associations of cell types with human diseases**



Different cell types in different tissues (e.g. cerebral organoids) can show association with different diseases at the same time.



Observe which GWAS hits are in the cell types studied (e.g. Alzheimer's hits) and observe neighboring genes that may be causative of the disease.

Conclusions

- All this implies **a shift in cell sorting**, from being merely morphological by microscopy to **sorting cells by its expression data**.
- The **data analysis** allow us to **predict the effect of any changes** in the **interaction network**, the consequences at the cellular level, at the tissue level or at the disease level.
- Similarly annotated clusters exhibit higher correlation of the expression profiles for the differentially expressed genes between datasets.
- We can create the **characteristic metabolism** of each cell type to **understand** the **signalling cascade** of each one and what happens when there is an alteration.
- We can create a "**computational human**" that allows us to **model different behaviors** of different interaction networks **under different disturbances**.

Source: Rosana C. Garrido: "*¡Esto no me lo habían dicho! Lo que no está escrito sobre la cultura científica*".

All the figures are self-made by BioRender

GitHub