

Applications of Deep Learning for Single Cell Biology

Dr. Carlos Talavera-López Ph.D

Institute of AI for Health
Helmholtz Zentrum München
210708

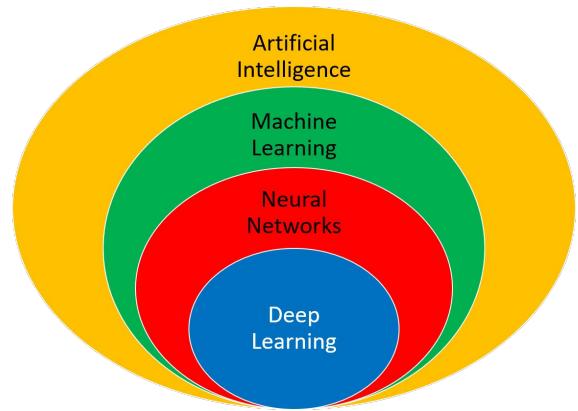
What is deep learning?



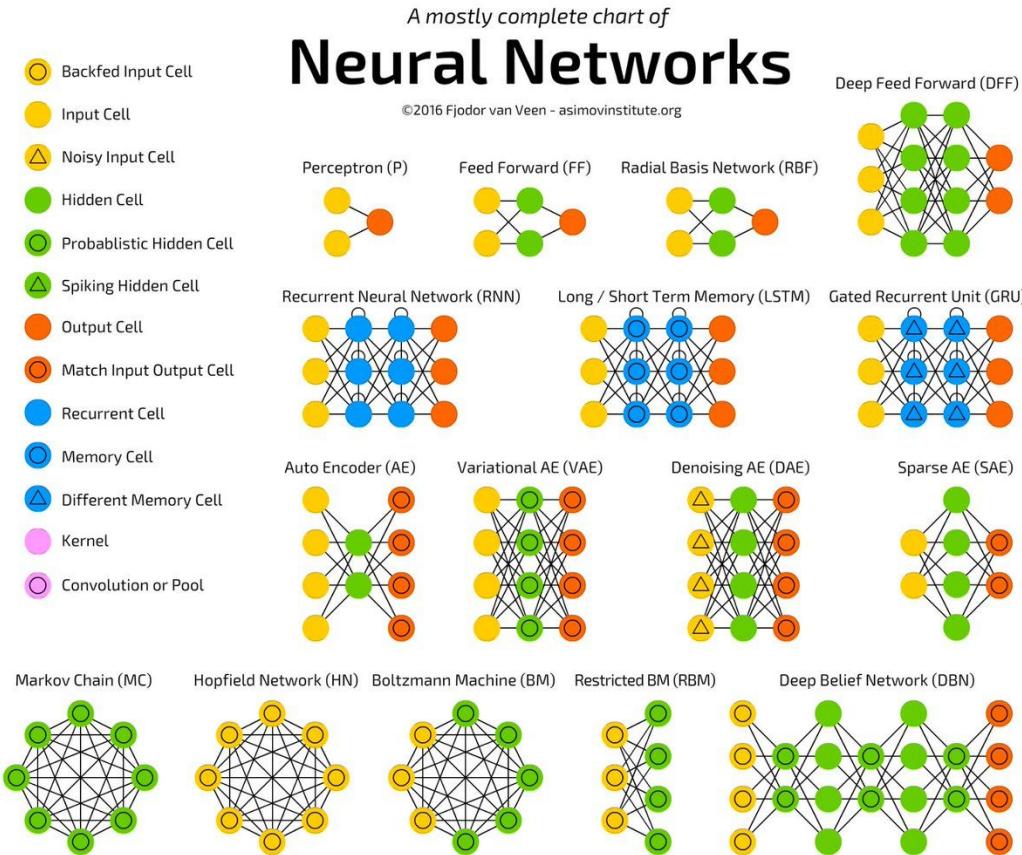
What is deep learning?

Deep Learning is a machine learning technique that constructs artificial neural networks to mimic the structure and function of the human brain.

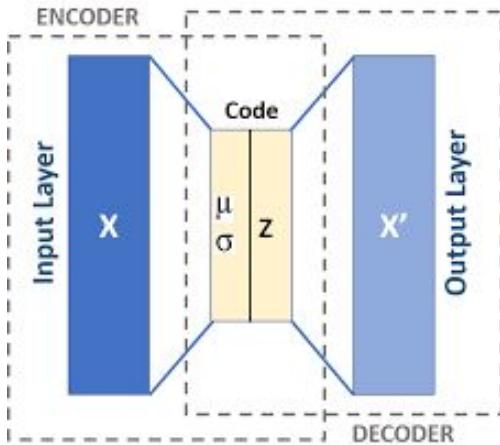
In practice, deep learning, also known as deep structured learning or hierarchical learning, uses a large number hidden layers -typically more than 6 but often much higher - of nonlinear processing to extract features from data and transform the data into different levels of abstraction (representations).



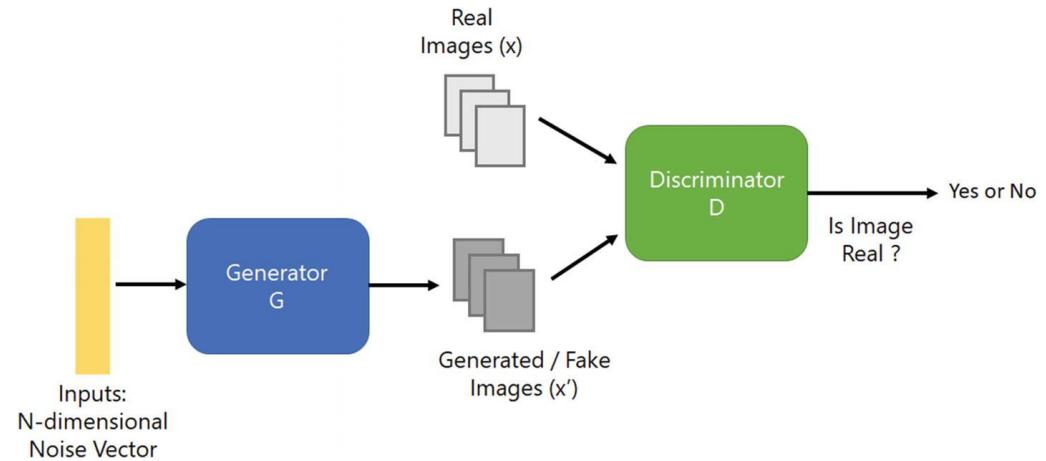
Types of deep learning: a bestiarium.



Generative Deep Learning



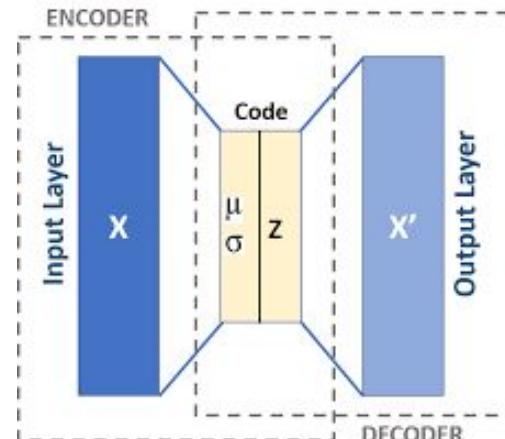
Variational Autoencoder



Generative Adversarial Networks

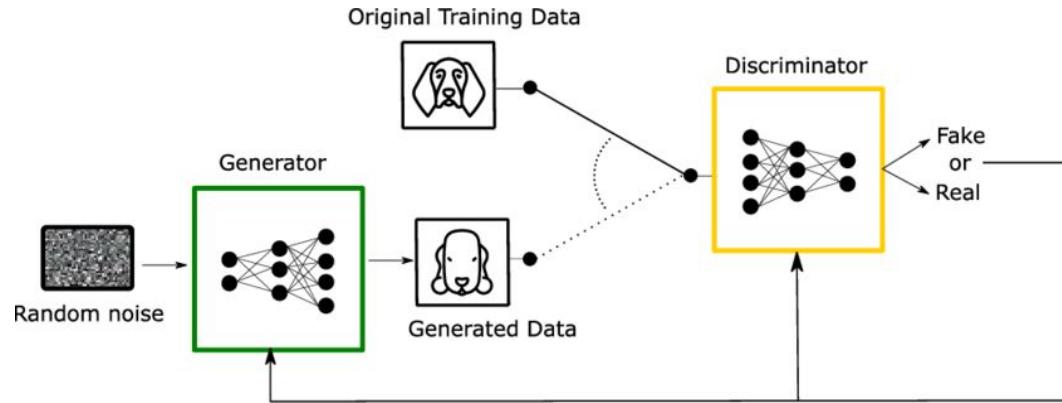
Generative Deep Learning: Autoencoders

The general idea of autoencoders is pretty simple and consists in setting an encoder and a decoder as neural networks and to learn the best encoding-decoding scheme using an iterative optimisation process (i.e: gradient descend).



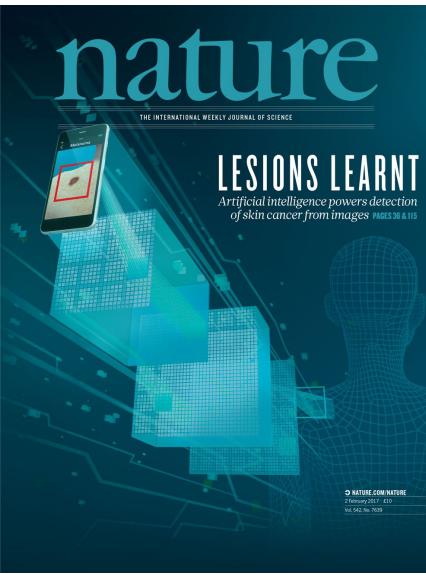
Variational Autoencoder

Generative Deep Learning: Generative Adversarial Networks

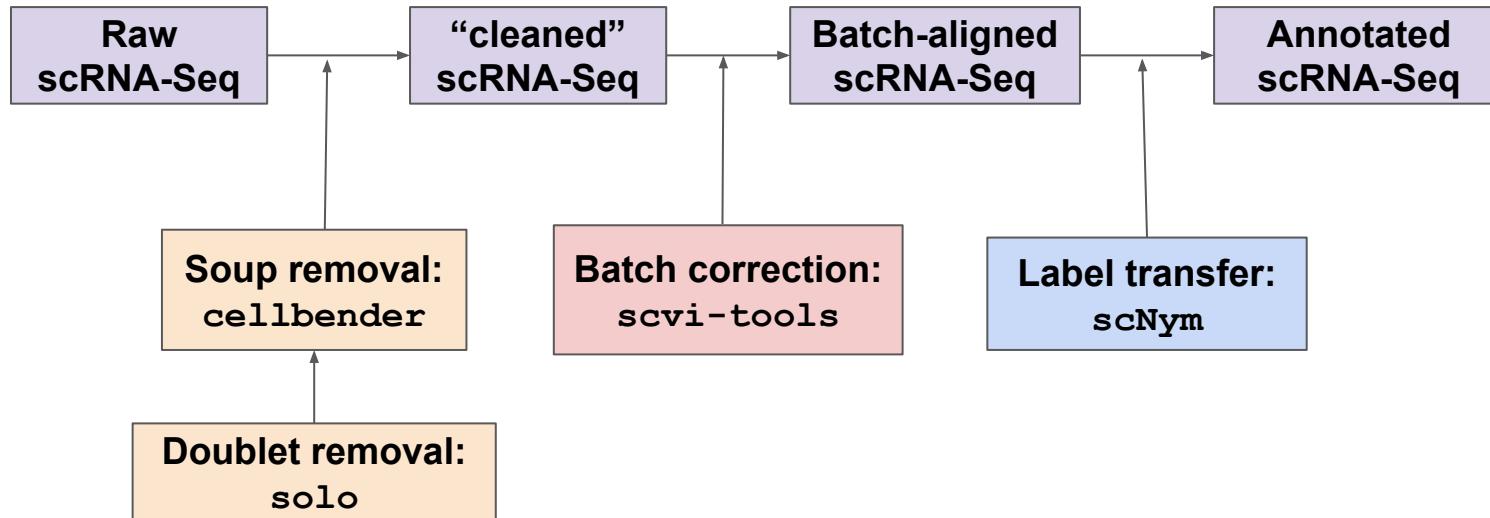


Generative Adversarial Networks are a type of deep learning architecture where two neural networks compete with each other, in a zero-sum game manner, to generate data that looks as similar to the input data as possible.

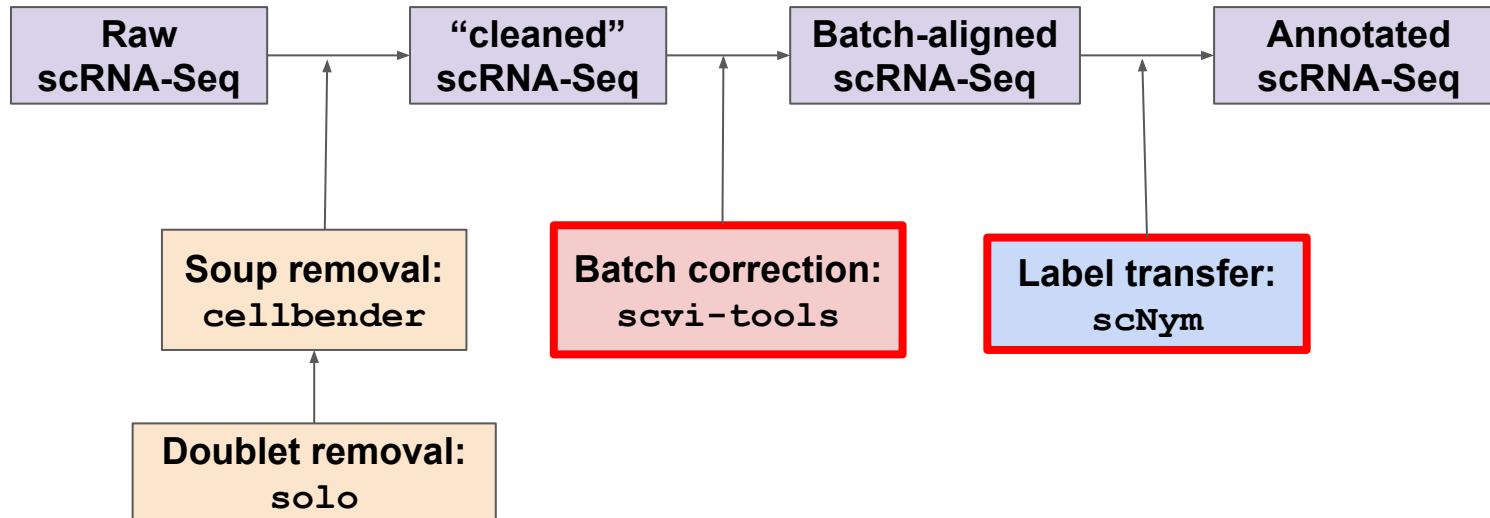
What are the applications of deep learning?



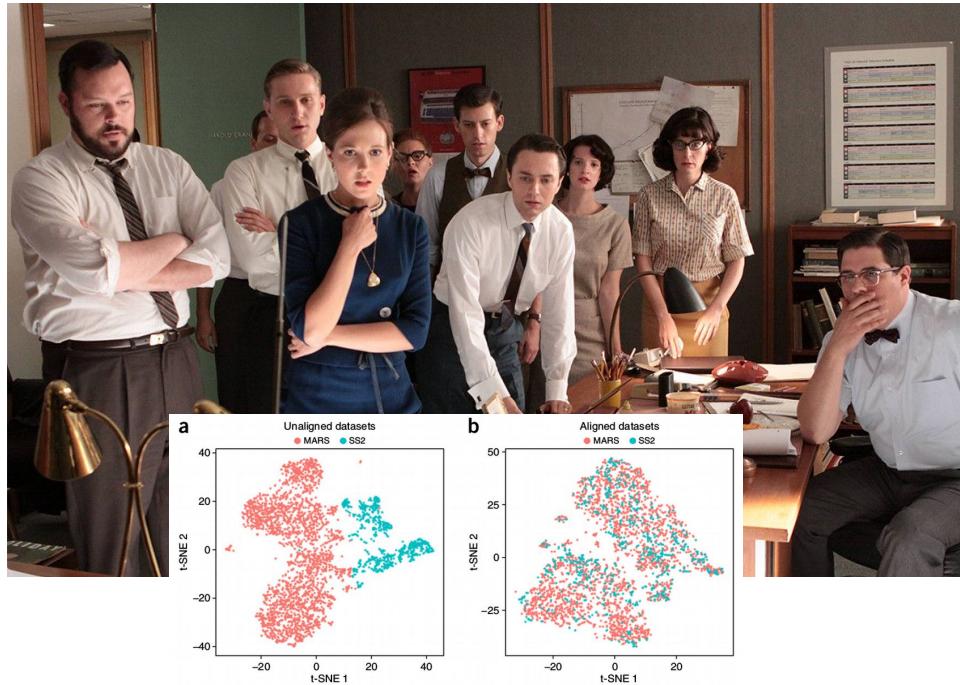
A DL-based single cell data analysis workflow



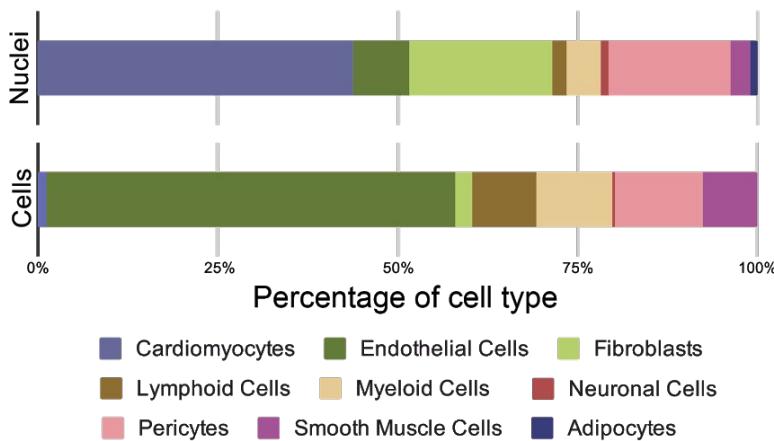
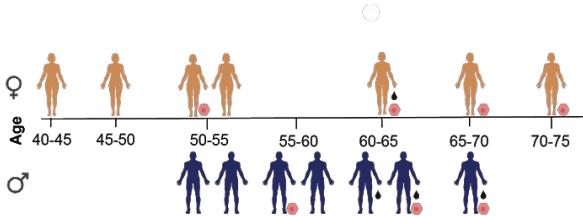
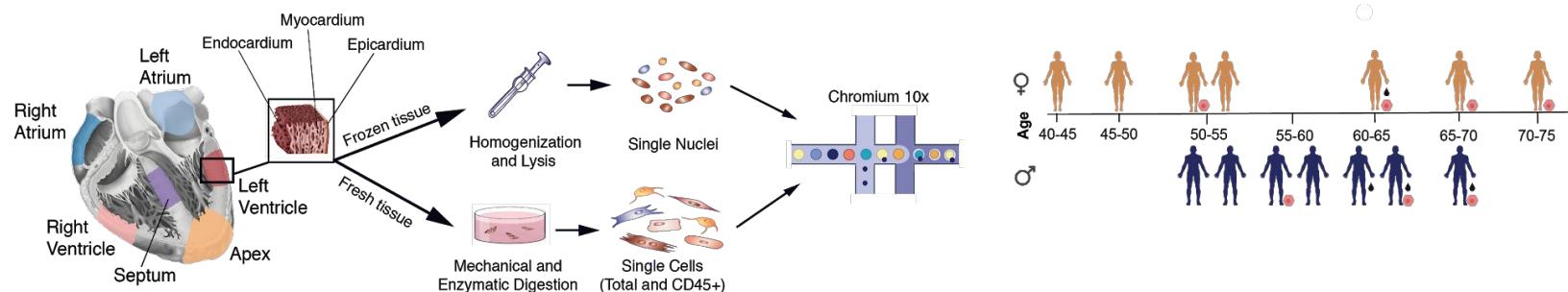
A DL-based single cell data analysis workflow



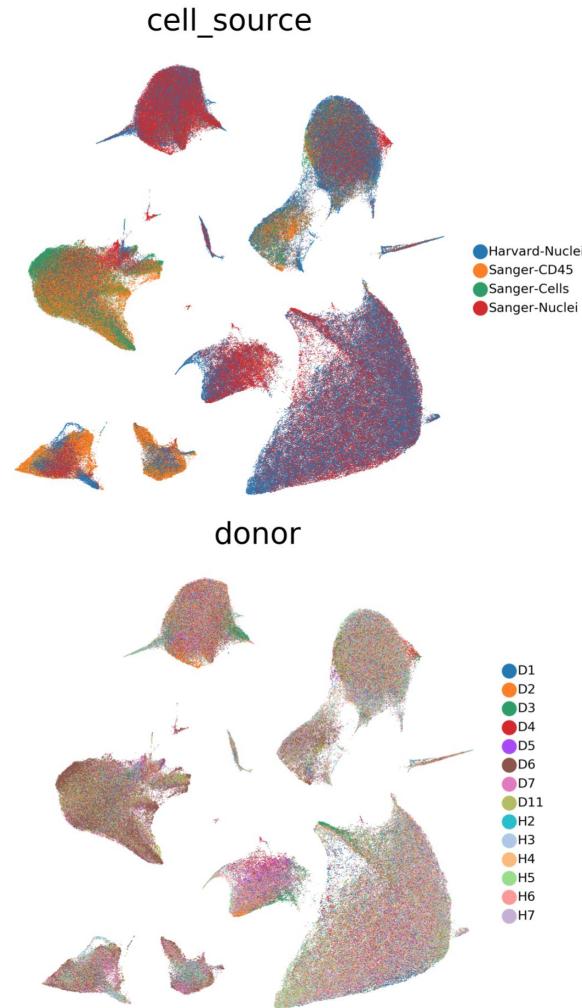
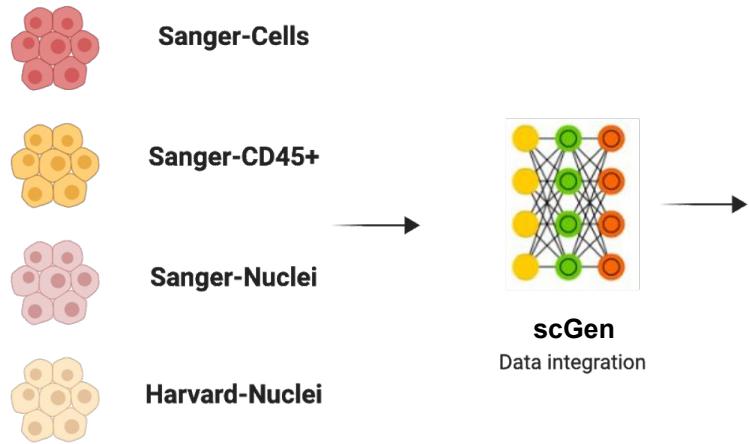
Batch Correction



A cell atlas of the adult human heart



Batch alignment using a VAE



Label transfer



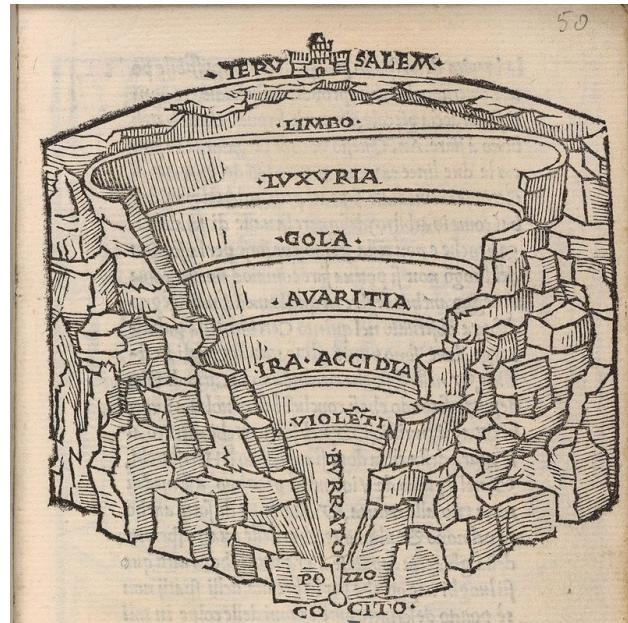
Image source: AMC, Creative Commons

Cell type/state annotation

Once you have your manifold and you are happy with it, the next step is to figure out what is in there in terms of cellular populations.

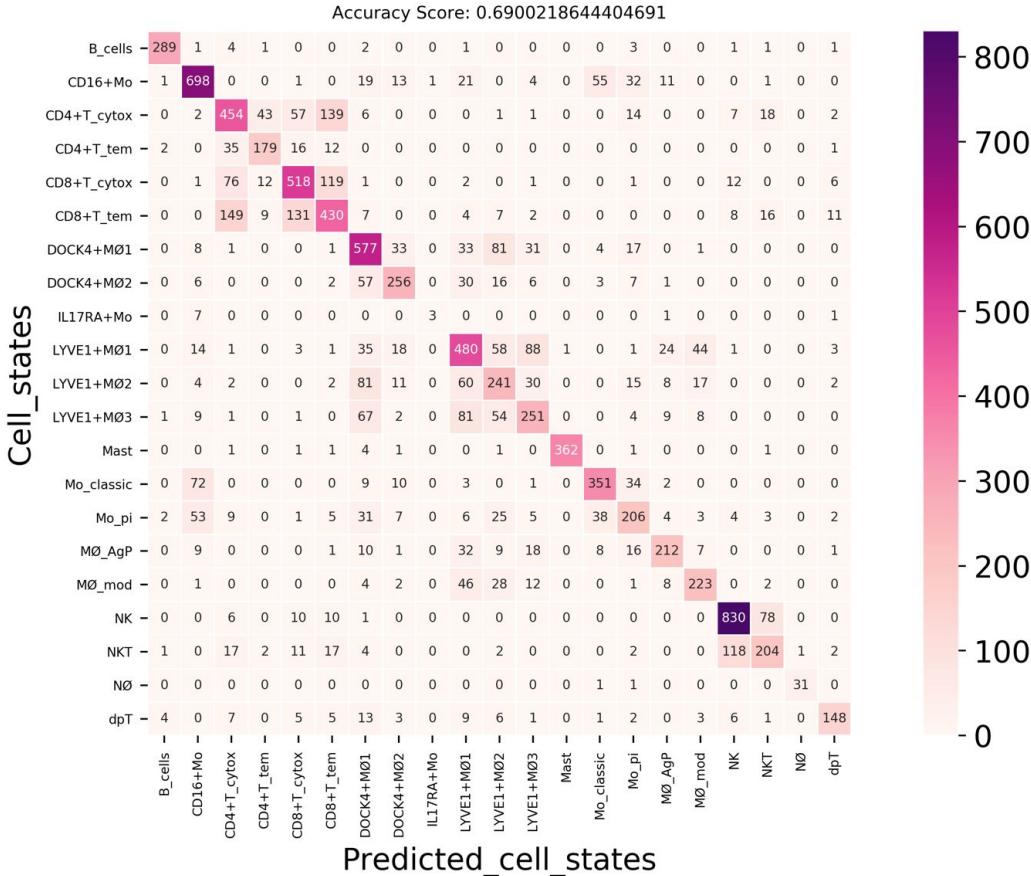
The “traditional” way to do this is by clustering the manifold of cells, perform DGE and manually annotate each cell state.

This process is time consuming and requires some level of expertise (and bias) to accurately identify known populations and potentially new ones.



Manual cell annotation

Cardiac immune model confusion matrix



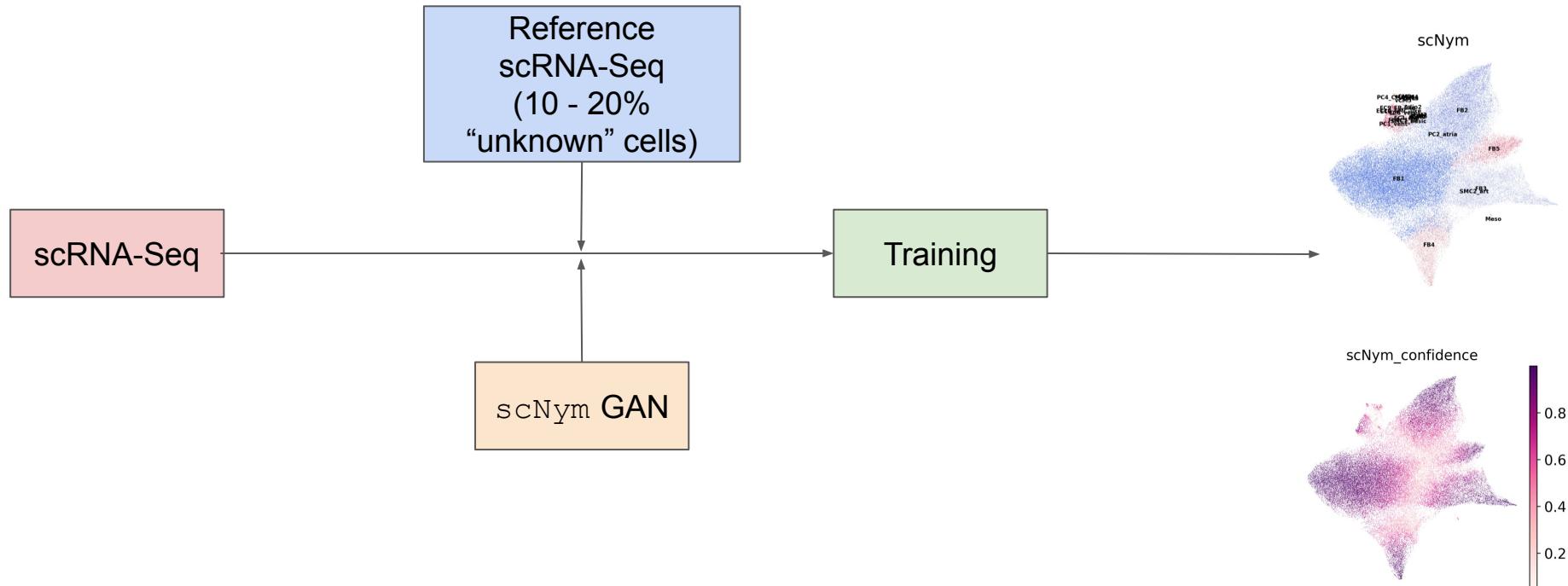
Automated cell annotation

You can also leverage on all the expertise that has gone into annotating all the scRNA-Seq data out there and use them as reference for automated cell annotation.

This will not solve 100% of your annotation problem, but it will definitely speed things up by identifying known populations and novel ones.



Automated cell annotation with scNym



The possibilities are endless!



Conclusions:

- Deep learning has recently started being applied to single cell biology.
- Deep learning plays well with the large matrix-based data.
- Each deep learning architecture has multiple potential applications, like transformers.
- Just because it is new and shiny doesn't mean it is better.

Materials for practical session

GitHub repo: https://github.com/cartal/lisbon_single_cell_2021

Batch correction notebook:

https://colab.research.google.com/drive/1xBd2hVI6o0oAm-Qrf_E9XcnAYPDrqPG7?usp=sharing

Label transfer notebook:

<https://colab.research.google.com/drive/1BbOa-u8Jt9IHjGUTz2NIFKOj8UXUbEEq?usp=sharing>

Thank you!