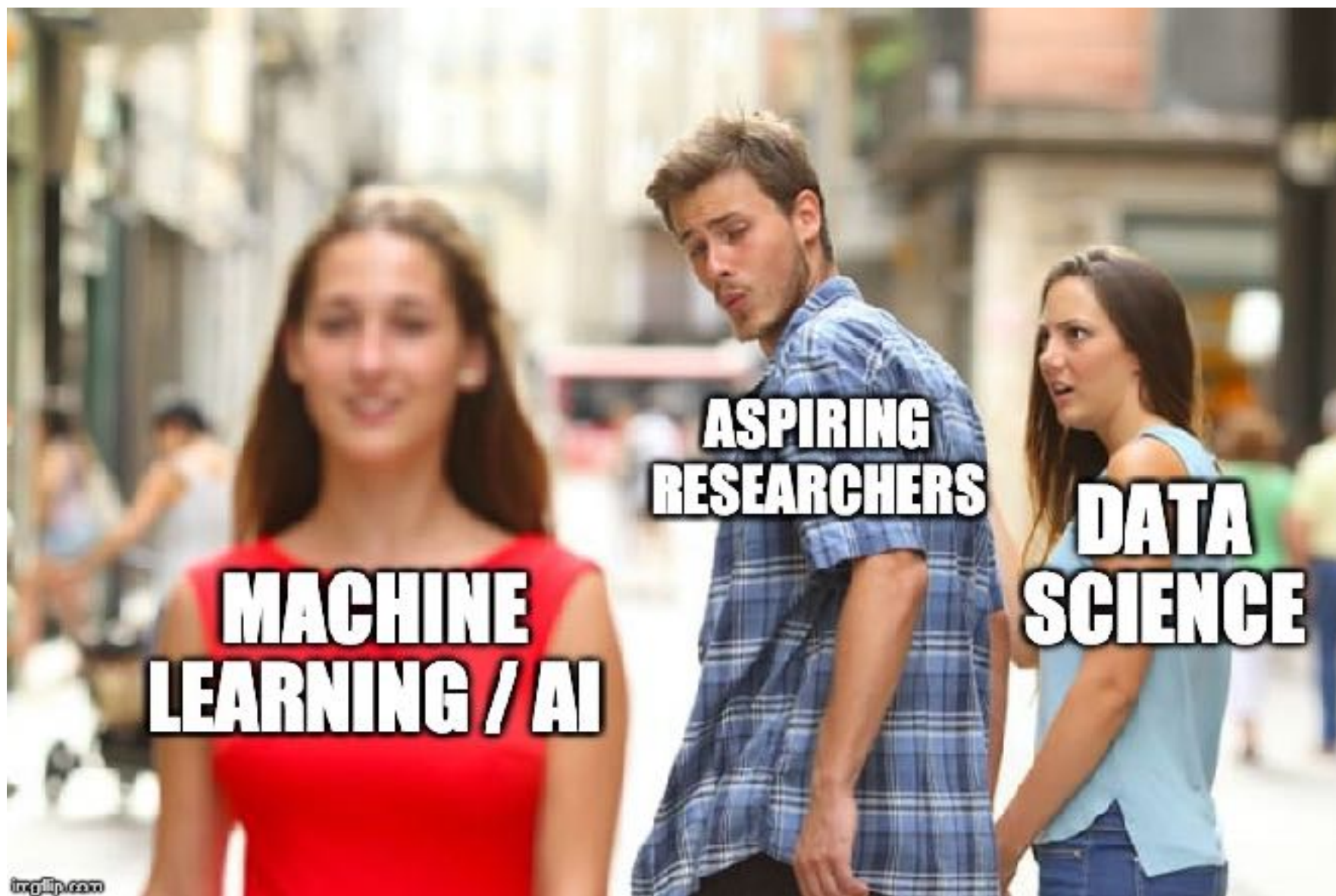


Come for the machine learning, stay for the data science

Shreejoy Tripathy
Krembil Centre for Neuroinformatics
Centre for Addiction and Mental Health
Department of Psychiatry, University of Toronto
website: triplab.org



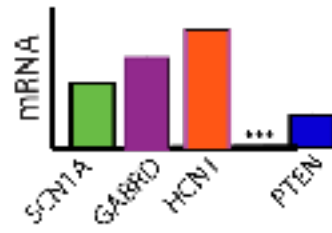
Laboratory for Computational Genomics

Bridging scales of investigation in neuroscience and psychiatry

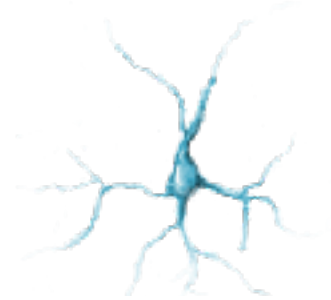
Genetics



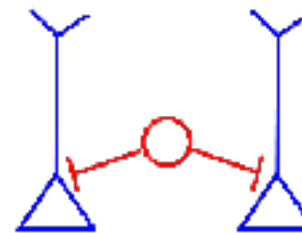
Gene Expression



Cellular Physiology



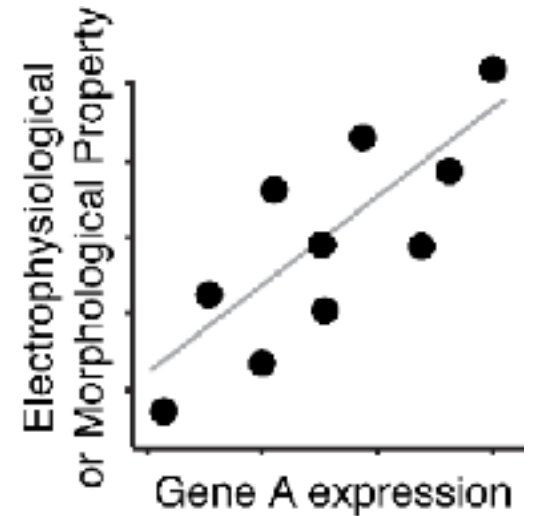
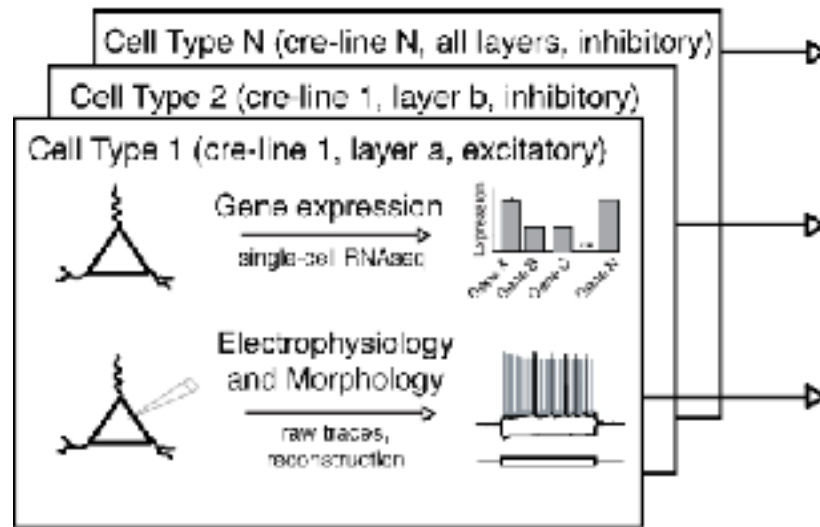
Microcircuit



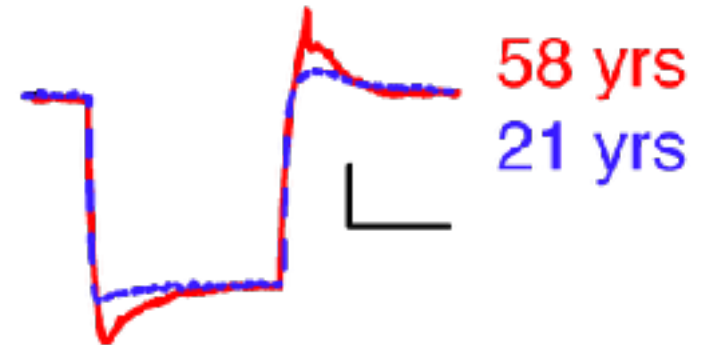
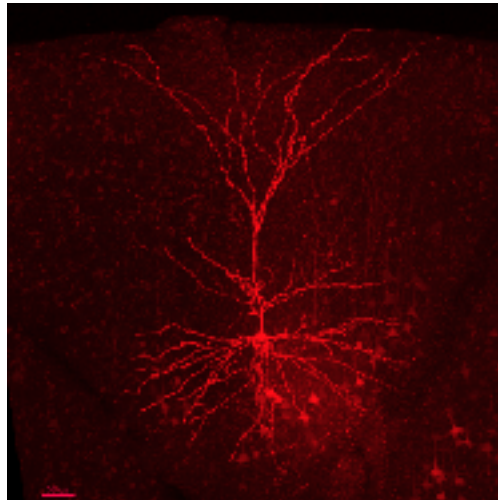
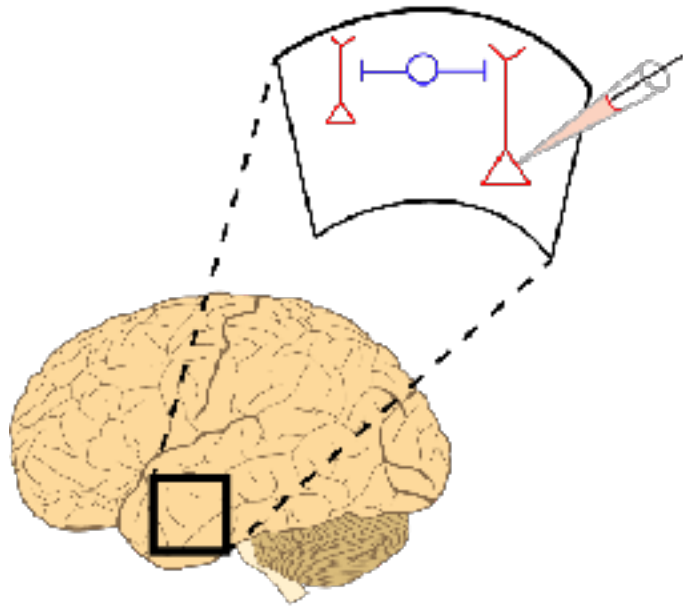
Traits & Behaviors



Machine learning on massive datasets to bridge genes to neuron activity

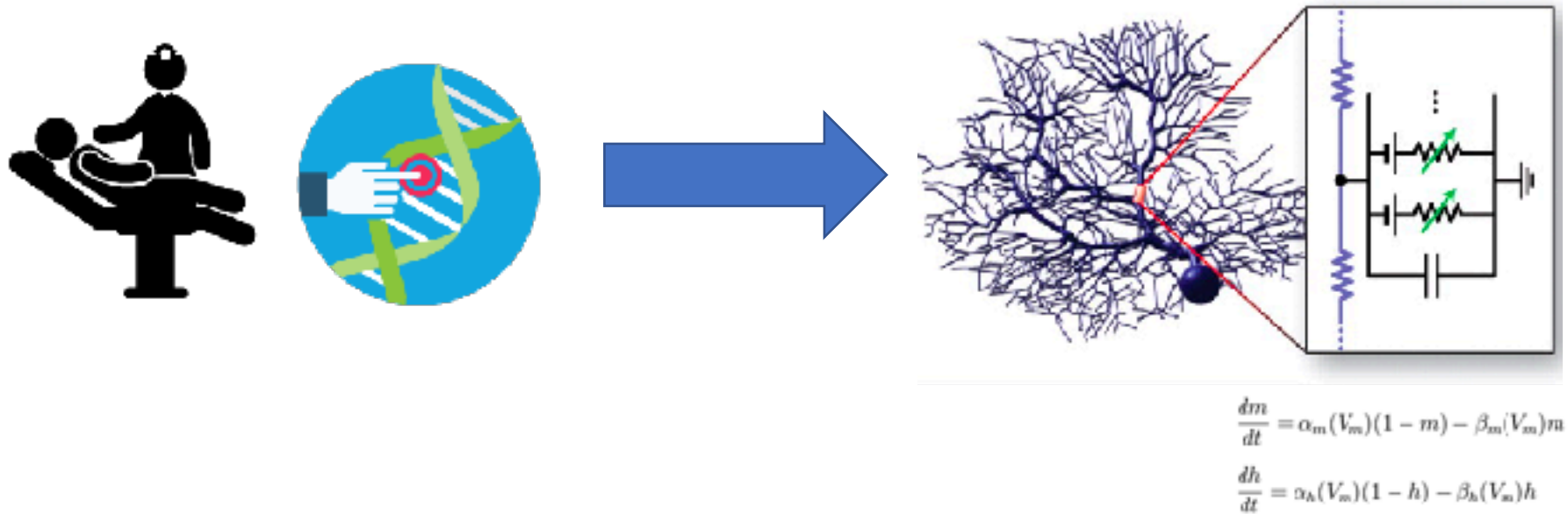


Investigating the neuronal basis of human-to-human differences



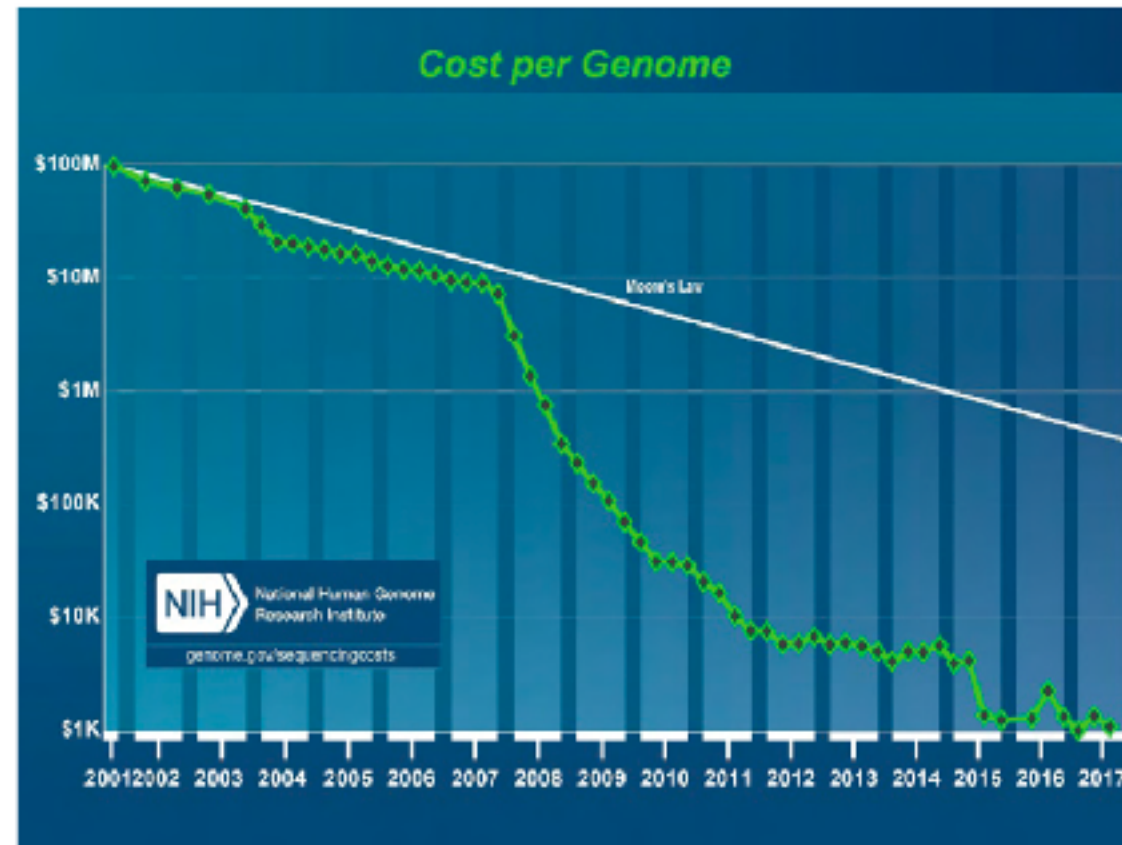
(with Krembil Brain Inst.
at Toronto Western)

Using patient genetics for personalized brain cell and circuit simulations



Automation and technology advances are driving down costs to generate data

Figure 1: Human Genome Sequencing – Historical Cost

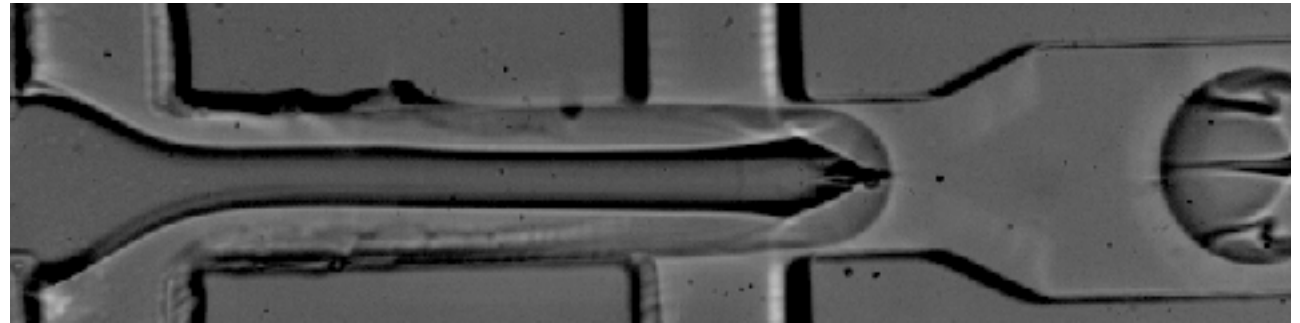


Source: Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP).¹

Automation and technology advances are driving down costs to generate data



Microfluidics device that enables massive-scale single cell genomics



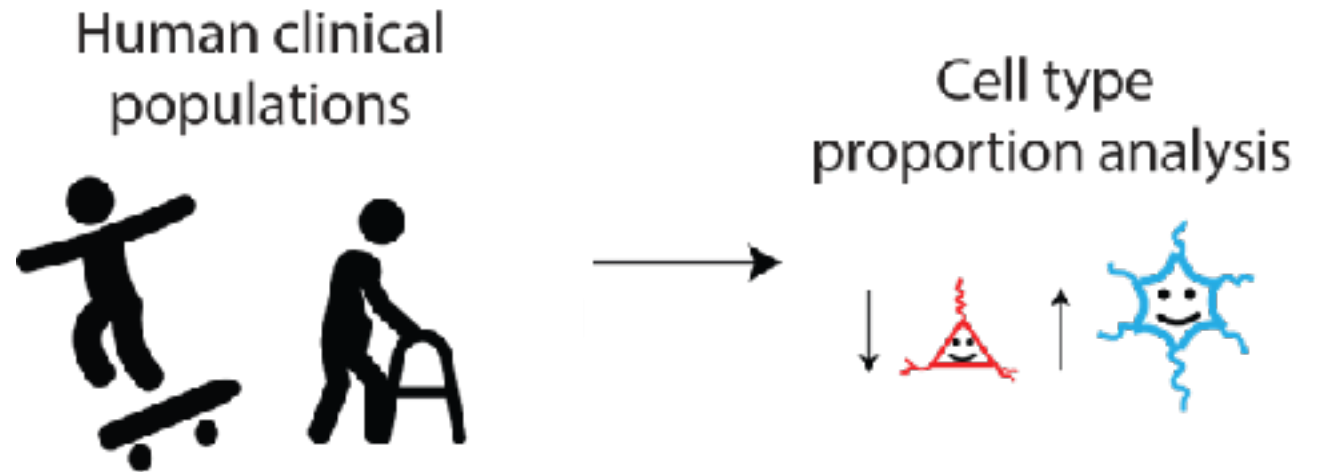


Teaching the next generation of biological data scientists



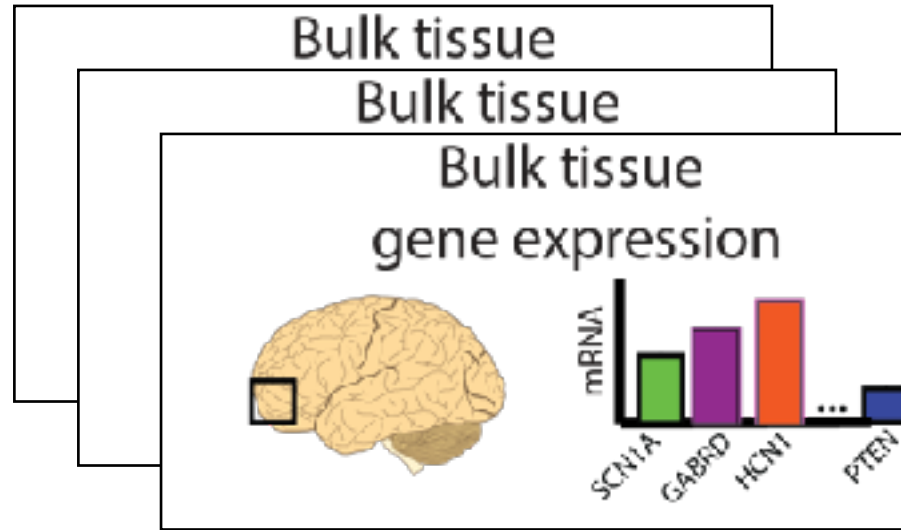
What's needed for a good data science project

- A question
- *What types of brain cells are increasing or decreasing in number as humans age?*



What's needed for a good data science project

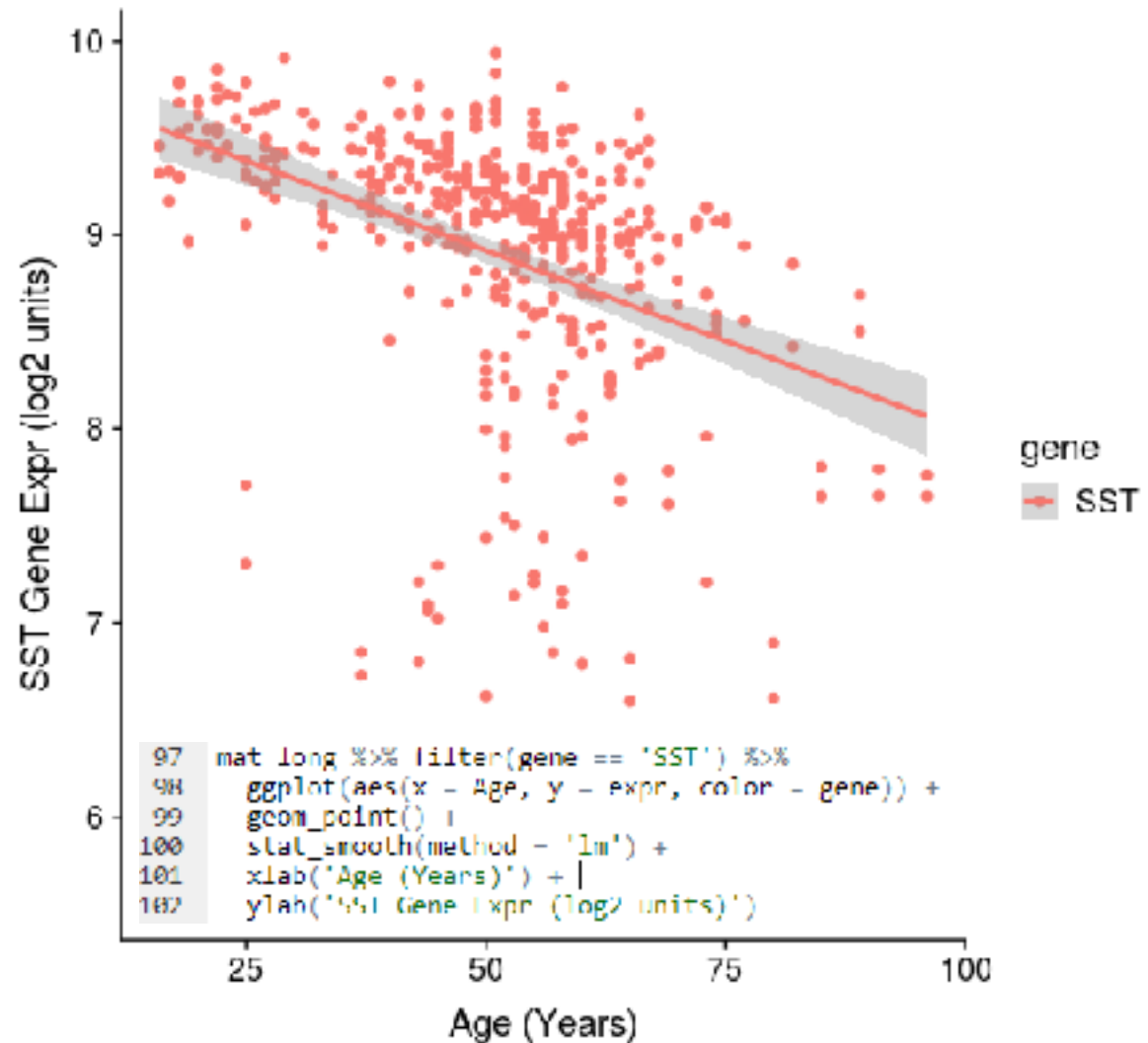
- A question
- A dataset
 - Bulk brain gene expression sampled from 100s of post-mortem human brains at various ages



	Sample 1	Sample 2	...	Sample N
Gene 1				
Gene 2				
...				
Gene 20K				

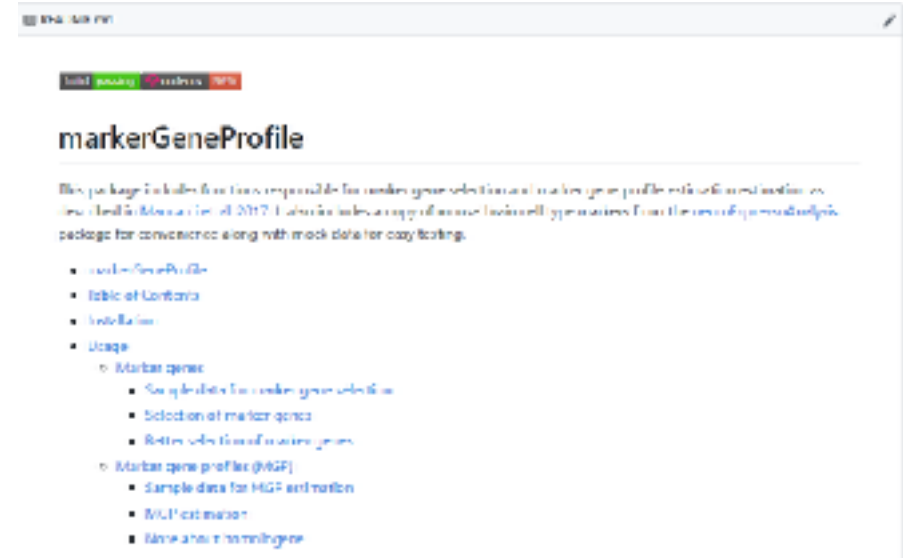
What's needed for a good data science project

- A question
- A dataset
- A data science toolkit
 - Importing data into R
 - readr, readxl
 - Combining data with metadata
 - dplyr, tidyr
 - Visualizing data using plots and summary stats
 - ggplot2



What's needed for a good data science project

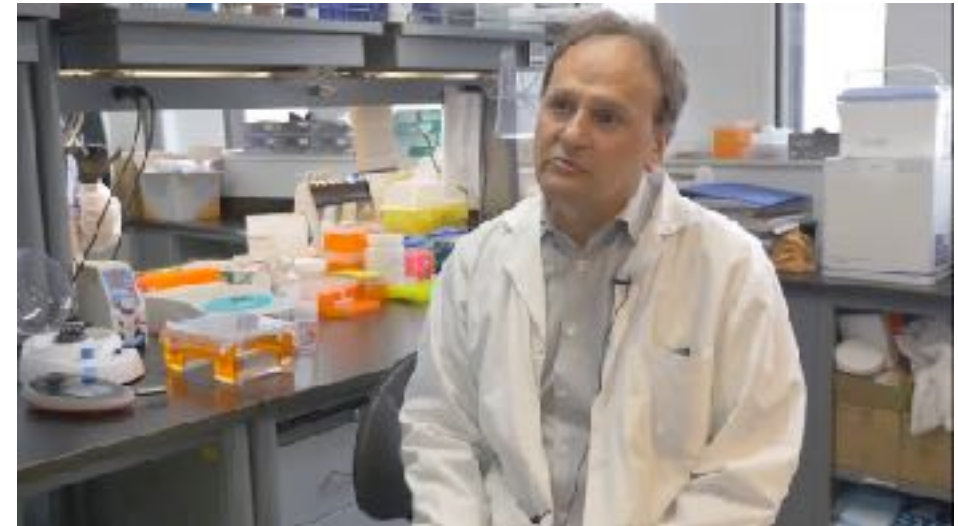
- A question
- A dataset
- A data science toolkit
- The *science* of data science
 - Knowing when to develop your own analysis tools vs reusing someone else's
 - incl. machine learning / AI
 - Was my initial question even the right one?
 - How do I know if my answer is right? Might there be other explanations that I need to rule out?
 - Iterate



```
estimations = mgpEstimate(exprData=mgp_LesnickParkinsonExp,  
  genes=mouseMarkerGenesM1EDrain,  
  geneColName="Gene.Symbol",  
  outlierSampleRemove=F, # should outlier samples removed. This is done using boxplot stats  
  geneTransform =function(x){homologene::mouse2human(x);humanGene}, # this is the default o  
  groups=mgp_LesnickParkinsonMetaDisease, #if there are experimental groups provide them  
  reekConsensus = FALSE, # ensures gene rotations are positive in both of the groups  
  removalMinority = TRUE) # removes genes if they are the minority in terms of rotation sign
```


What's needed for a good data science project

- A question
- A dataset
- A data science toolkit
- The *science* of data science
- Communication



Dr. Etienne Sibille, CAMH

What's needed for a good data science project

- # What's needed for a good data science project
- A question
 - A dataset
 - A data science toolkit
 - The *science* of data science
 - Communication



The goal of the Krembil Centre is to use the power of big data, artificial intelligence, and brain modelling to treat mental illness.

Director,
Sean Hill



**Computational
Genomics**
Shreejoy Tripathy
(USA)



**Brain Cells &
Circuits**
Etay Hay
(Israel)



**Whole Brain
Dynamics**
John Griffiths,
(UK)



**Integrative
Neuroinformatics**
Leon French,
(Canada)



**Cognitive
Networks**
Andreea
Diaconescu,
(Romania)



**Whole Person
and Populations**
Daniel Felskey
(Canada)



Acknowledgements

- Tripathy Lab @ UofT and CAMH Email: shreejoy.tripathy@utoronto.ca
• Now hiring for Post-Docs and graduate students!
Website: triplab.org
Twitter: @neuronJoy
- Krembil Centre for Neuroinformatics @ CAMH
- Data Providers

