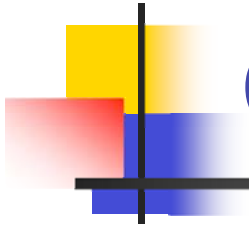
Genome-wide atlas of gene expression in the adult mouse brain

By Lein Ed S *et. al*

Presented by

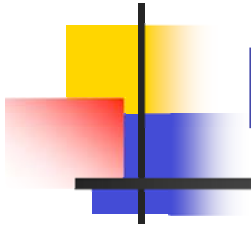
Shinsheng Yuan

STAT SINICA



Outline

- Motivation
- *in situ* hybridization
- Reproducibility
- Mapping ISH to Anatomic Data
- Global analysis
- Comments

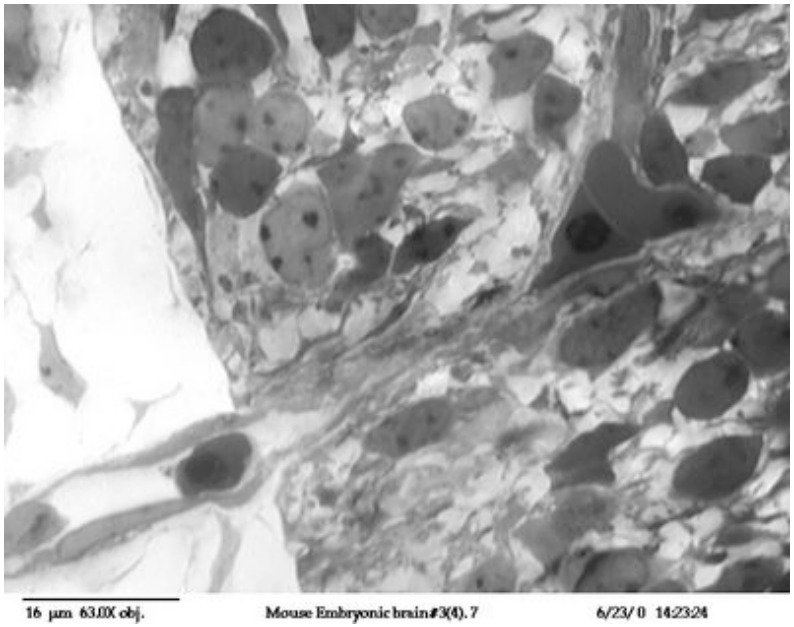


Motivation

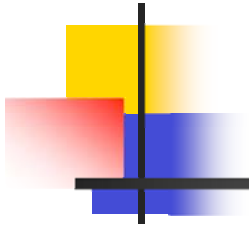
- The phenotypic properties of different neuronal and non-neuronal cells are the product of **unique combinations of expressed gene products**.
- Gene expression profiles => define the cell types. (Note: This has been done for cancer classification.)

Traditional Approach

- Histology



  - One gene at a time
  - Not systemically generated, analyzed, consolidated.

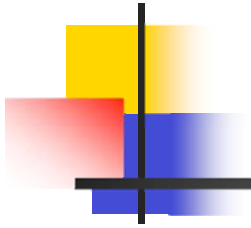


Measure Expression Levels

- DNA microarrays and SAGE
  - Usually applied to large brain region. (low resolution)
    - Have difficulty in differentiating the neuronal subtypes.
  - Single cell level microarray analysis (Kamme, F et. al. J. Neurosci (2003), Sugino, K. et. al. Nature Neurosci (2006))

Measure Expression Levels

- In this paper, *in situ* hybridization (ISH) technique is used.
  - Finer resolution (cellular but not single cell)
  - The data could be used to analyze the relationship among
    - Gene expression
    - Gene regulation
    - CNS function (spatial)
    - Cellular phenotype (spatial)



ISH vs microarray

- *in situ* hybridization vs microarray
  - *in situ* hybridization measures the expression and preserve the spatial information for a single gene.
  - microarray extracts all mRNAs out of tissues and measure thousands of genes at the same time. (What if we extract the “same” region of tissue only? Need registration!)
- For brain, it is important to keep the spatial information!

ISH

- Tissue preparation & Imaging process
  - Sectioning (Chop the tissues into slices ~25μm. Two directions are done (either Sagittal or Coronal)
  - Staining (Non-isotopic digoxigenine (DIG)-based in this paper.)
  - Washing
  - Imaging process

8 genes could be done on one mouse.

Each gene will have 20 Sections.

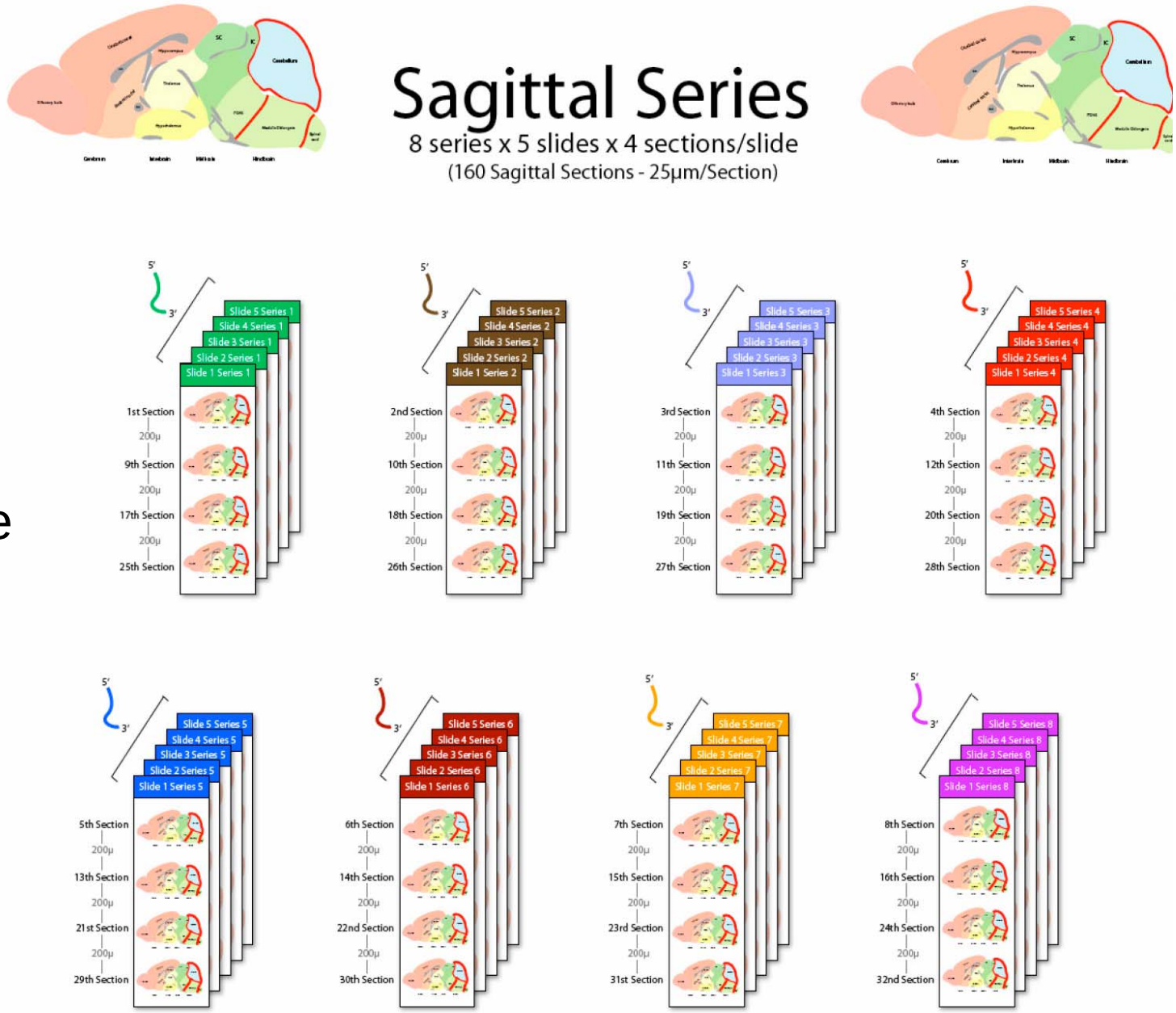


Figure 8 – *Standard series schema for a sagittally-sectioned brain*

6 genes could be done on one mouse brain.

Each gene has 56 sections.

2 of series are for Nissl staining.

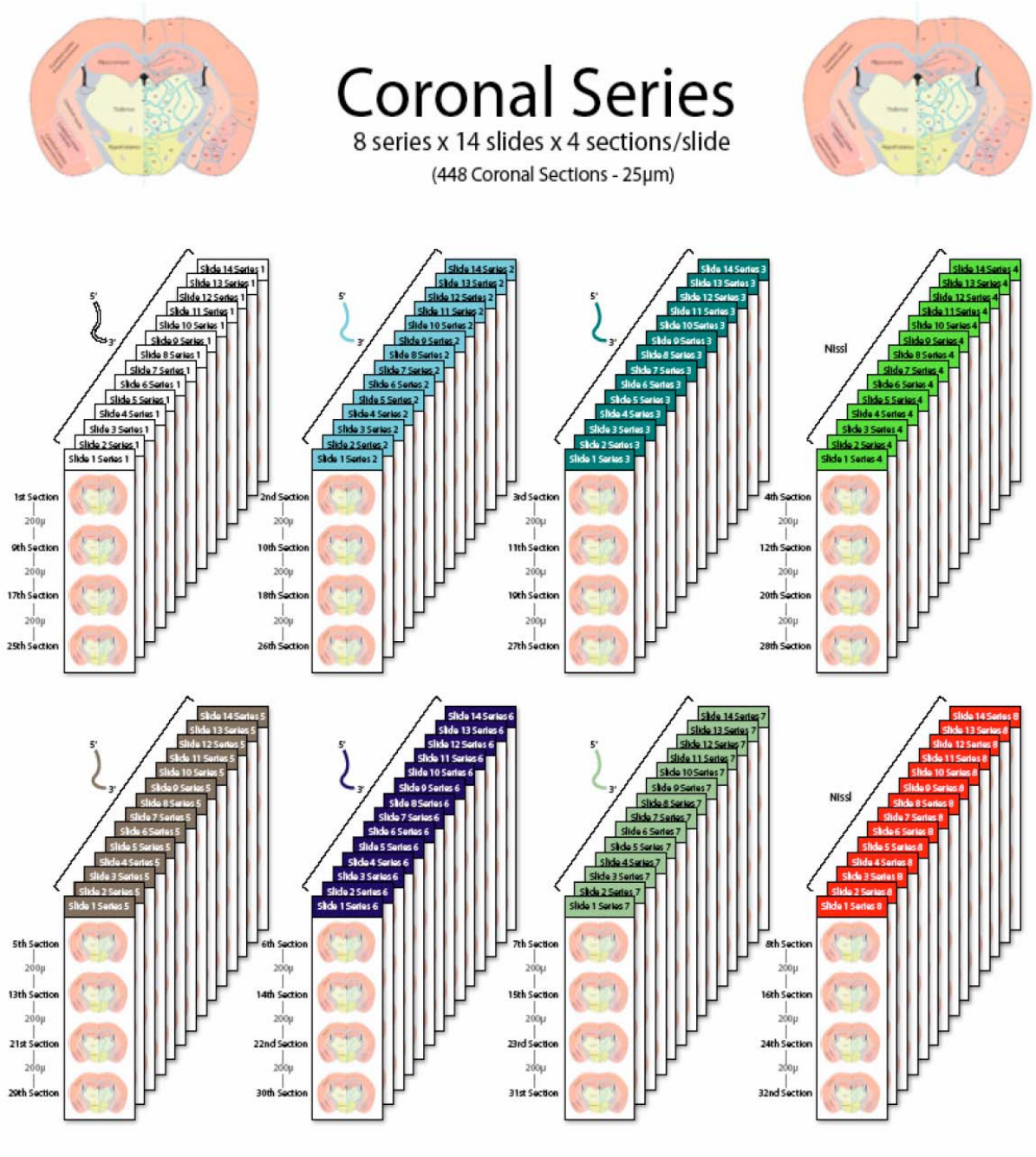
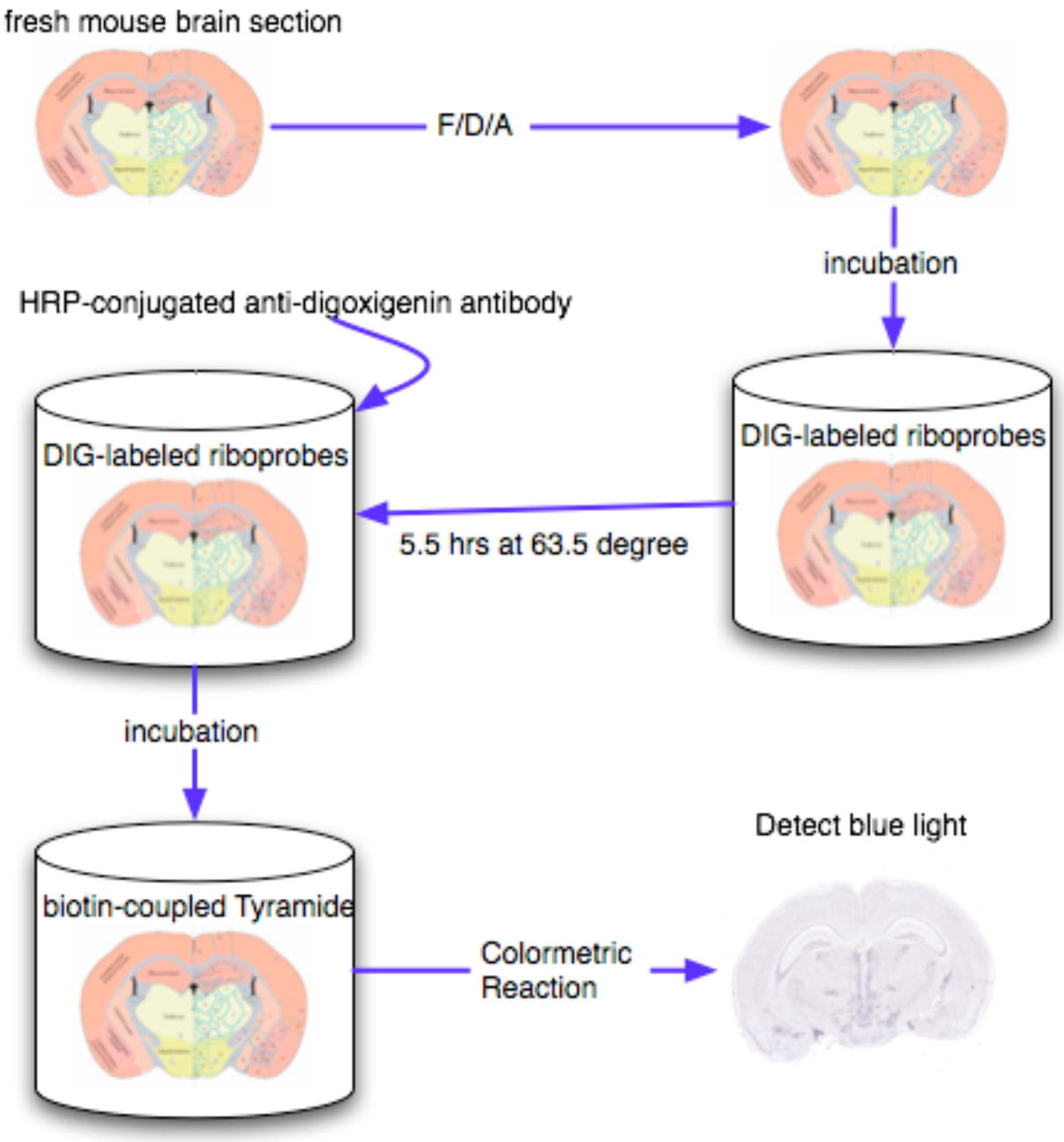
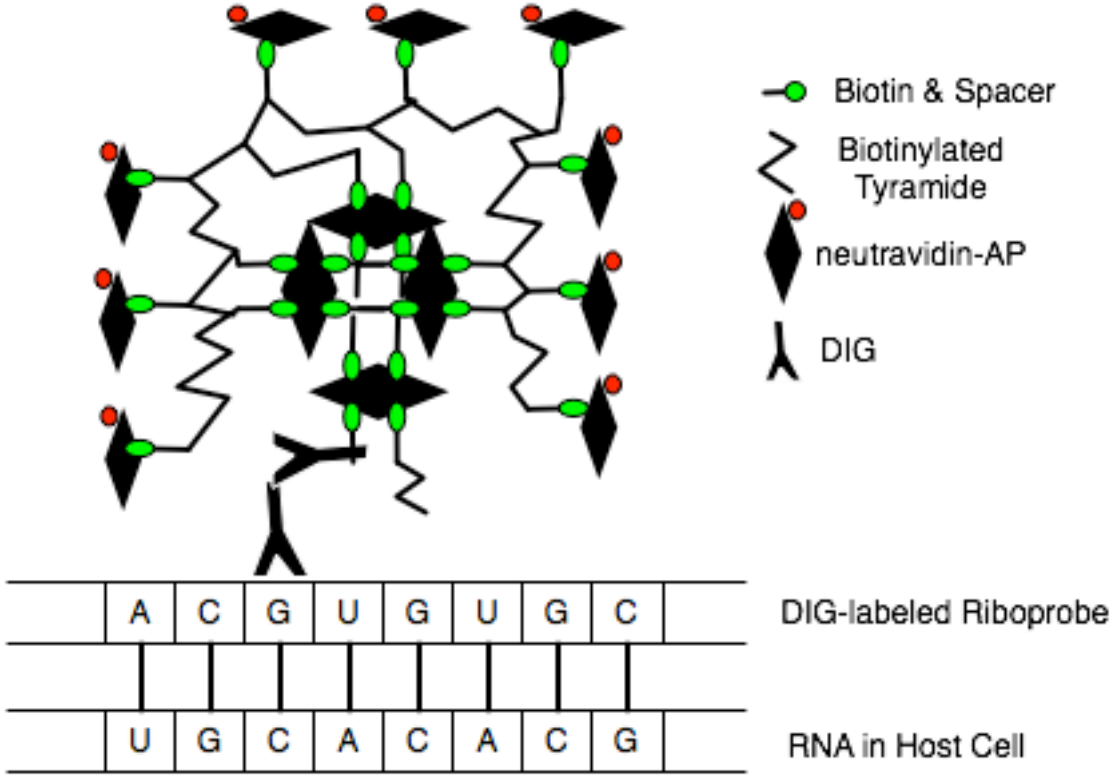


Figure 9 - *Standard series schema for a coronally-sectioned brain*

Staining Step

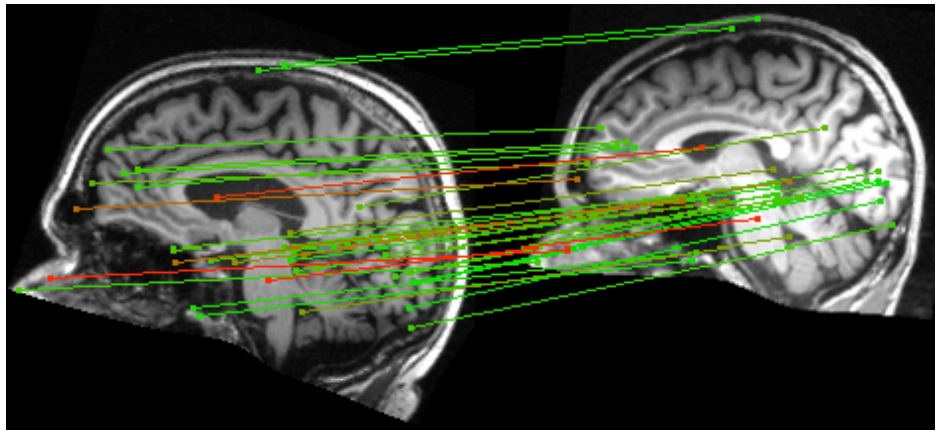


TSA (Tyramide Signal Amplification)



Data normalization on ISH data

- Background correction & Registration
  - Intensity normalization (Correct the background from negative control)
  - Registration (Map the image to the reference (concensus) atlas.)

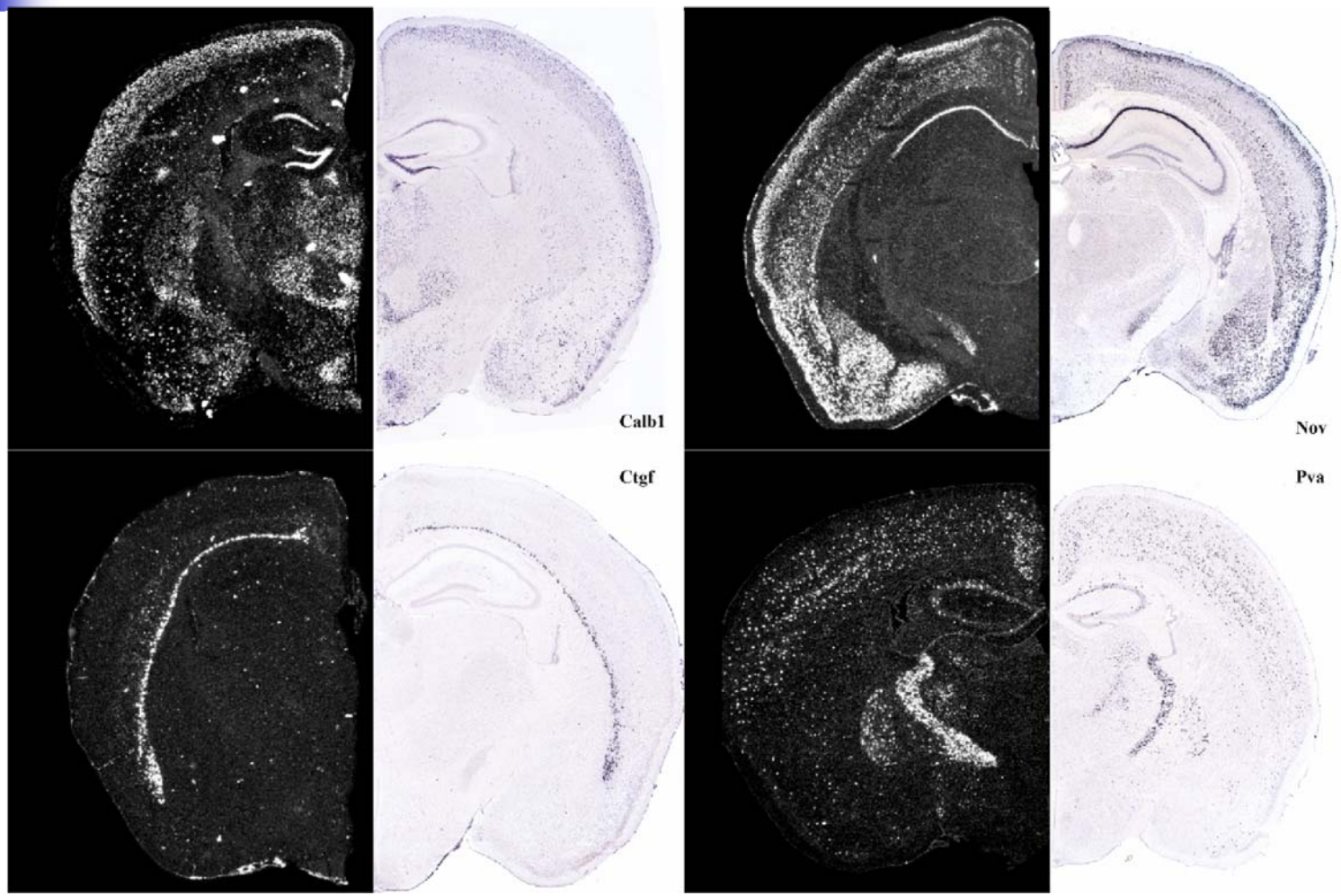


This is human.
Not mouse.
An example for registration.

More On ISH

- ISH only measures the expression for **single gene**.
- In order to measure multiple genes, a inbred mouse strain is used.
- Although different mice are used for different genes, the **expression** for **inbred mice under the same environmental conditions** are **reproducible**.
- An additional benefit from using inbred mice is that the brain image registration becomes **easier**.

Radioactive vs DIG-based ISH



Is ISH Reproducible?

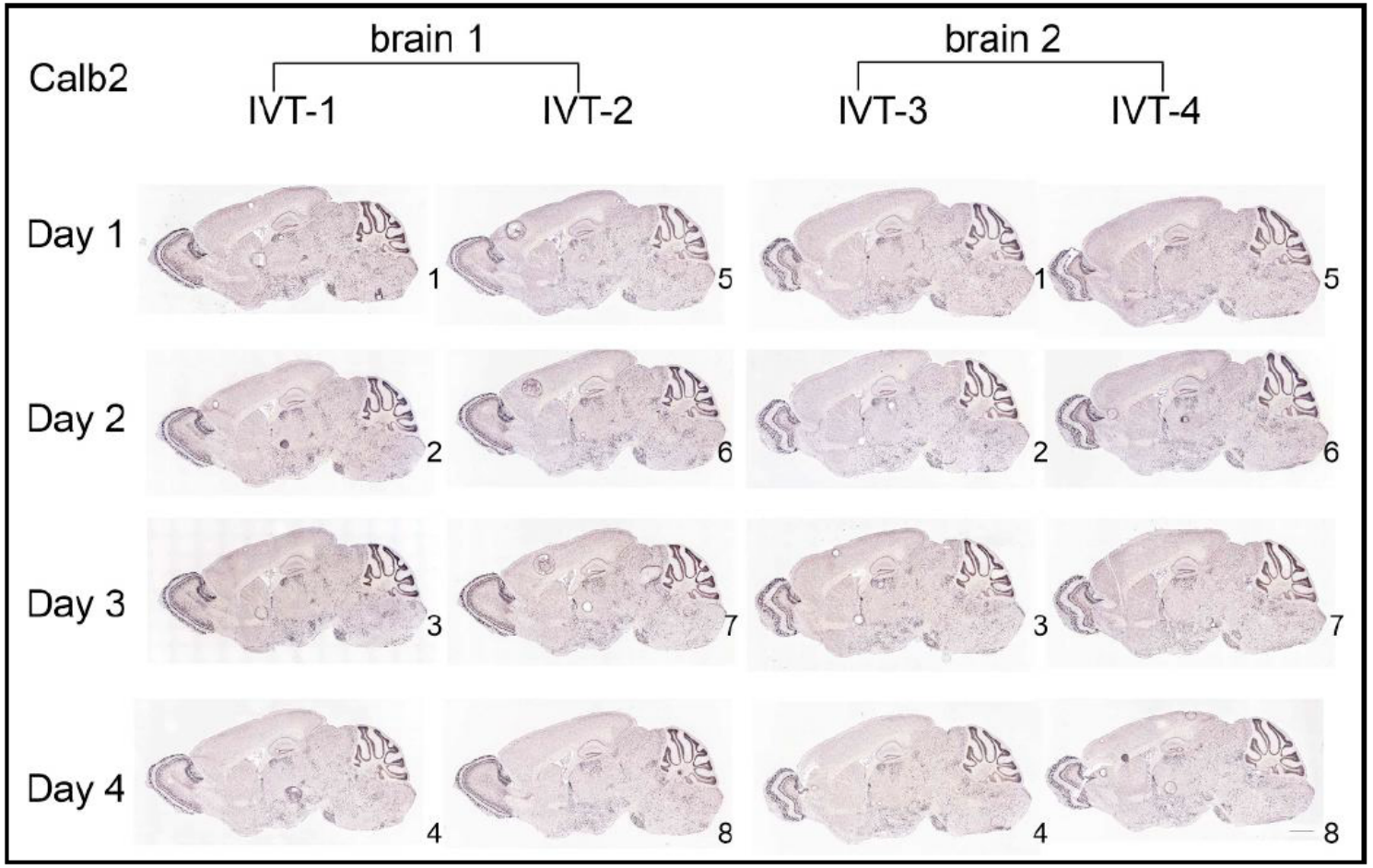
- Primary Source of variation comes from
  - Riboprobes
  - Day to day variability
  - Biological variability across brain sample (Not considered here since the inbred mouse is used.)

Experiments

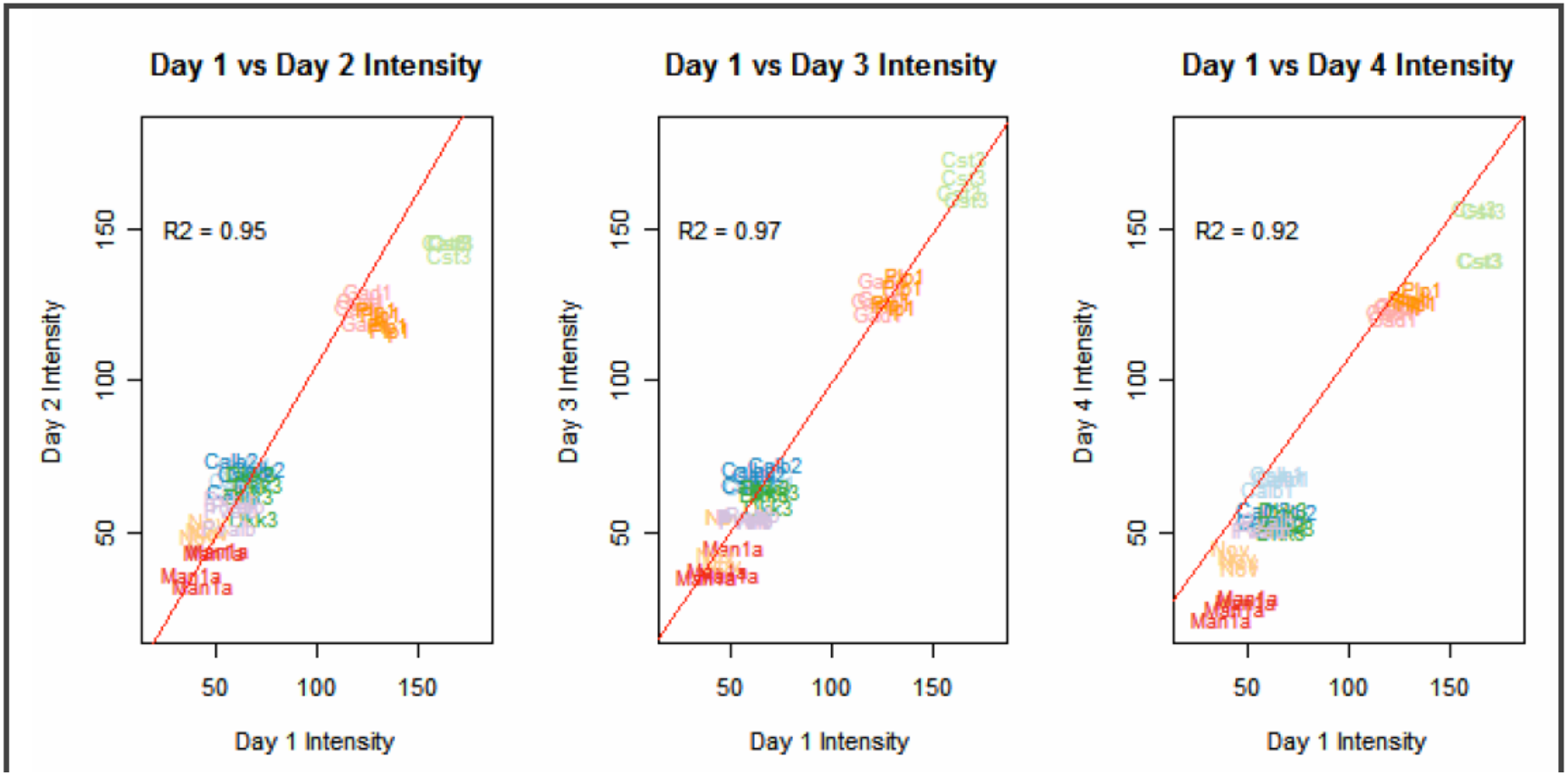
- Four riboprobes were independently generated by IVT (In vitro translation).
- 9 genes are included.

ISH Run	brain 1		brain 2	
Day 1	**IVT-1** *Series 1*	**IVT-2** *Series 5*	**IVT-3** *Series 1*	**IVT-4** *Series 5*
Day 2	**IVT-1** *Series 2*	**IVT-2** *Series 6*	**IVT-3** *Series 2*	**IVT-4** *Series 6*
Day 3	**IVT-1** *Series 3*	**IVT-2** *Series 7*	**IVT-3** *Series 3*	**IVT-4** *Series 7*
Day 4	**IVT-1** *Series 4*	**IVT-2** *Series 8*	**IVT-3** *Series 4*	**IVT-4** *Series 8*

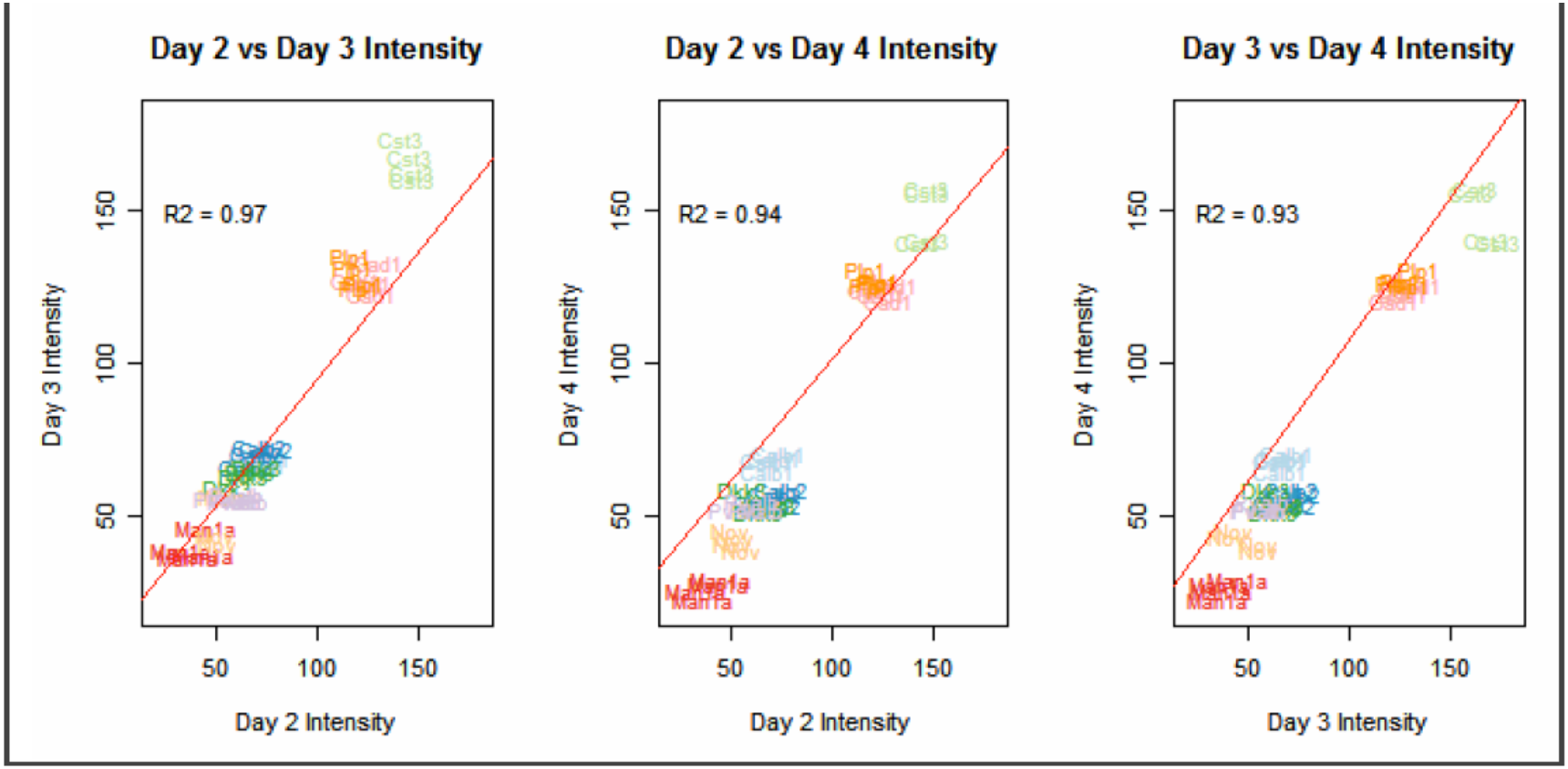
Images for Calb2



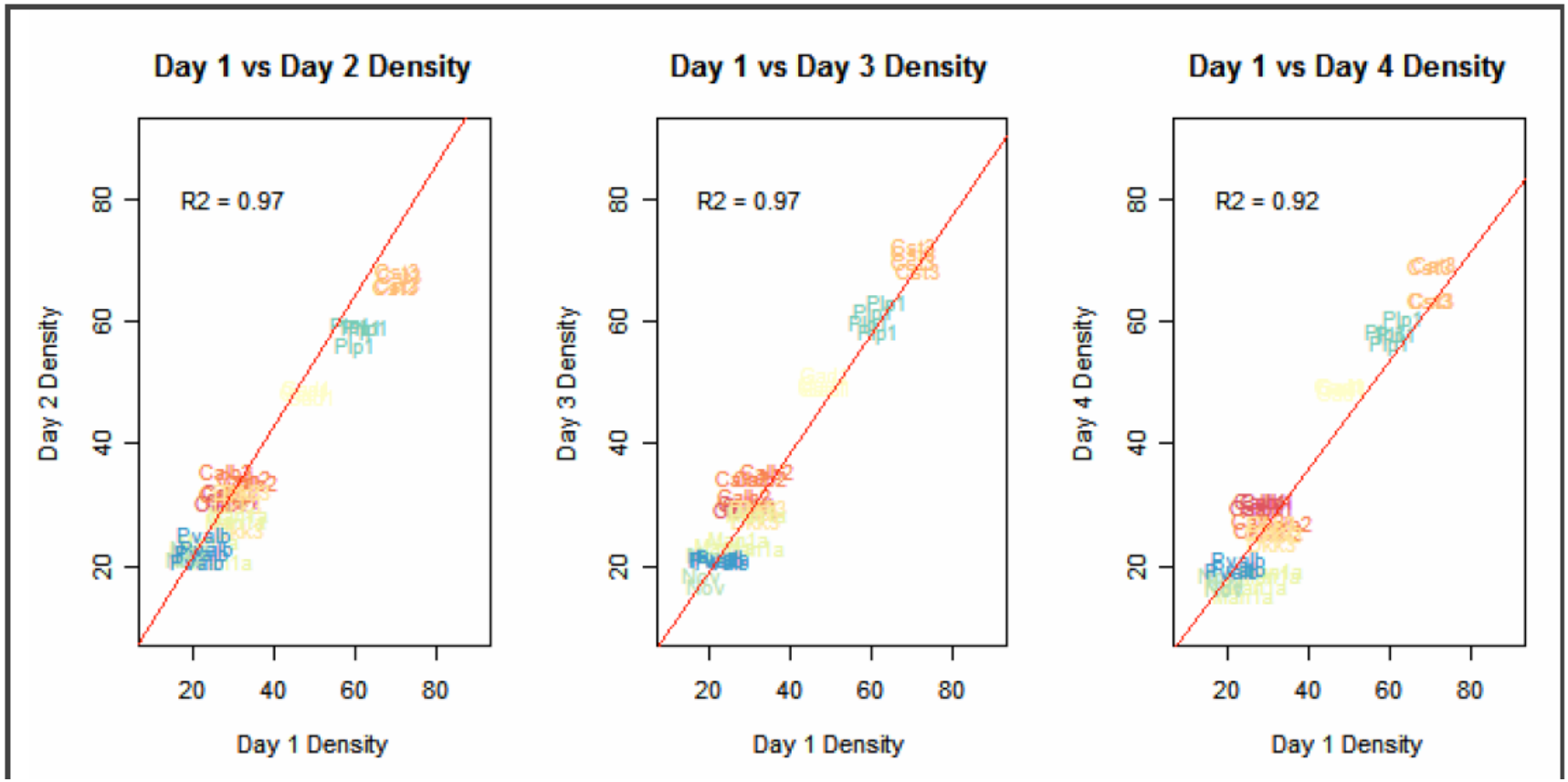
Comparison: Intensity



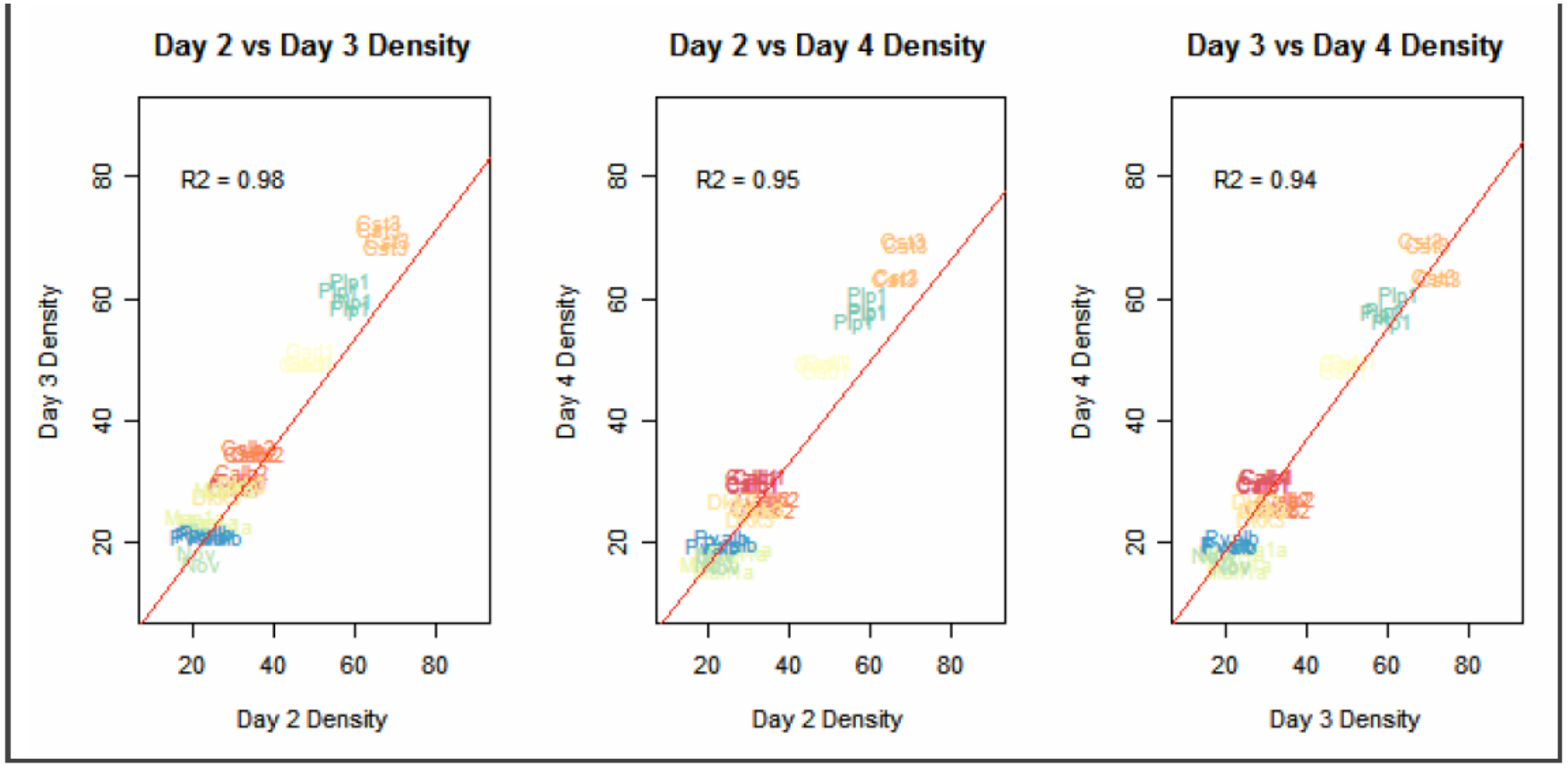
Comparison: Intensity



Comparison: Density

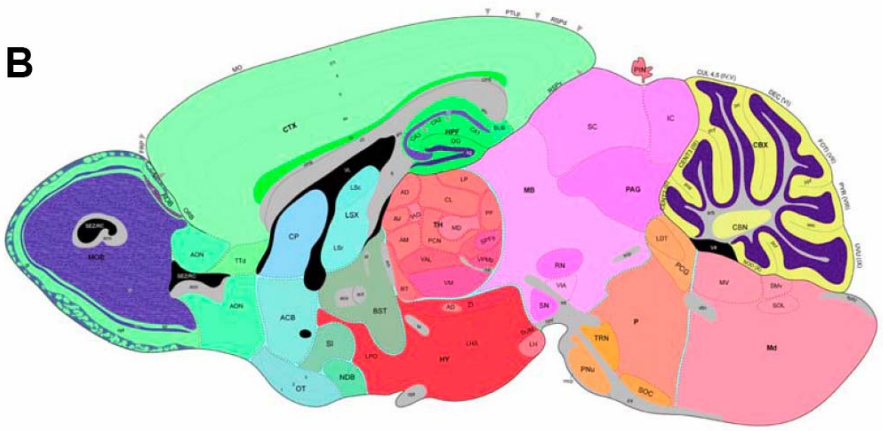


Comparison: Density



Mapping ISH to Anatomic Data

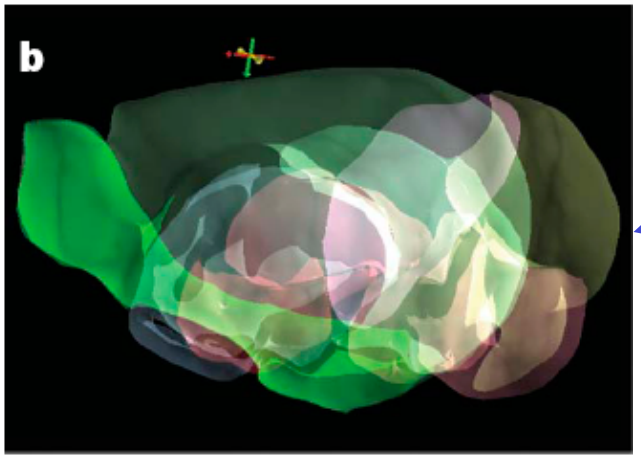
- A registration problem.
- Even with inbred mice, the variation between individual brains with the same age, sex, weight, etc, is still significant.

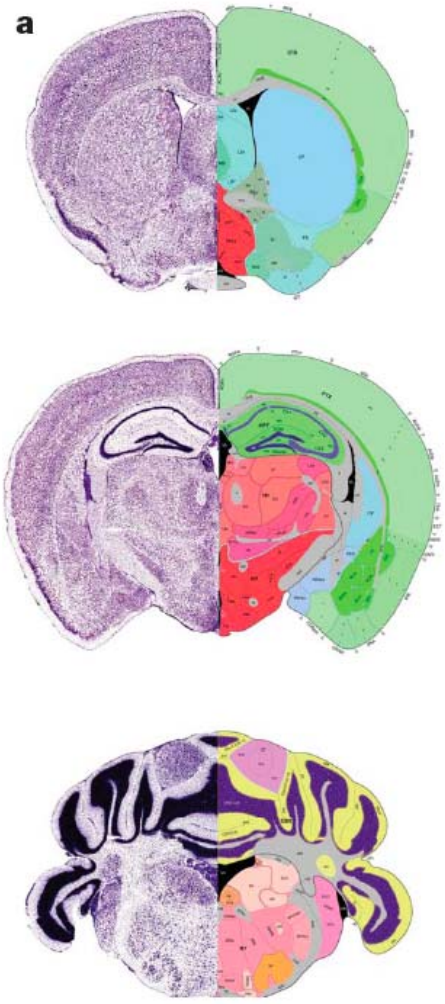


Allen Reference Atlas

Allen Reference Atlas

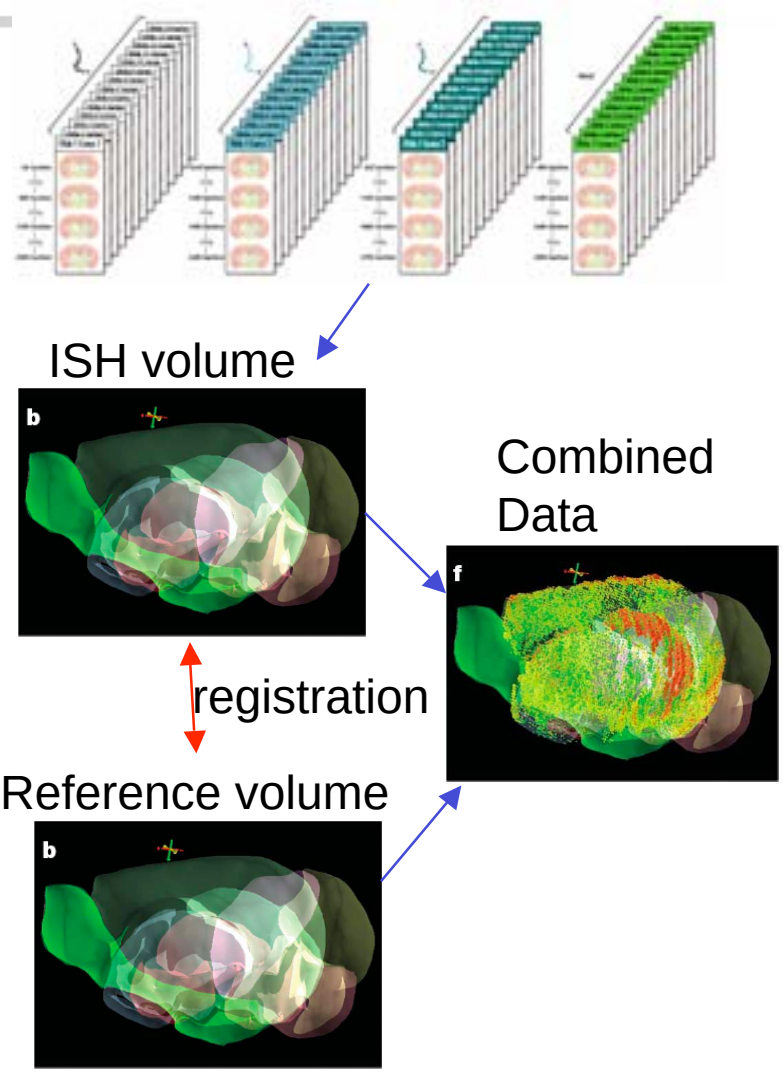
- Render the reference atlas into 3D reference volume.





Finding the Nearest Annotated Section

- For each mouse brain, 448 sections for coronal direction, 160 sections for saggital direction.
- 3D registration
  - The 3D ISH volume is reconstructed based on ISH images.
  - The registration algorithm will be map the 3D reference volume to 3D ISH volume.



Global Analysis of gene expression

- Expressed versus non-expressed genes.
- Enriched expression in major cell types
- Regionally enriched gene expression
- Cluster of correlated gene expression

Expressed versus non-expressed genes.

- **Expression Level** L, I is the average intensity, a_g is the expressing area, a_{max} is the maximum expressing area. [0, 255]

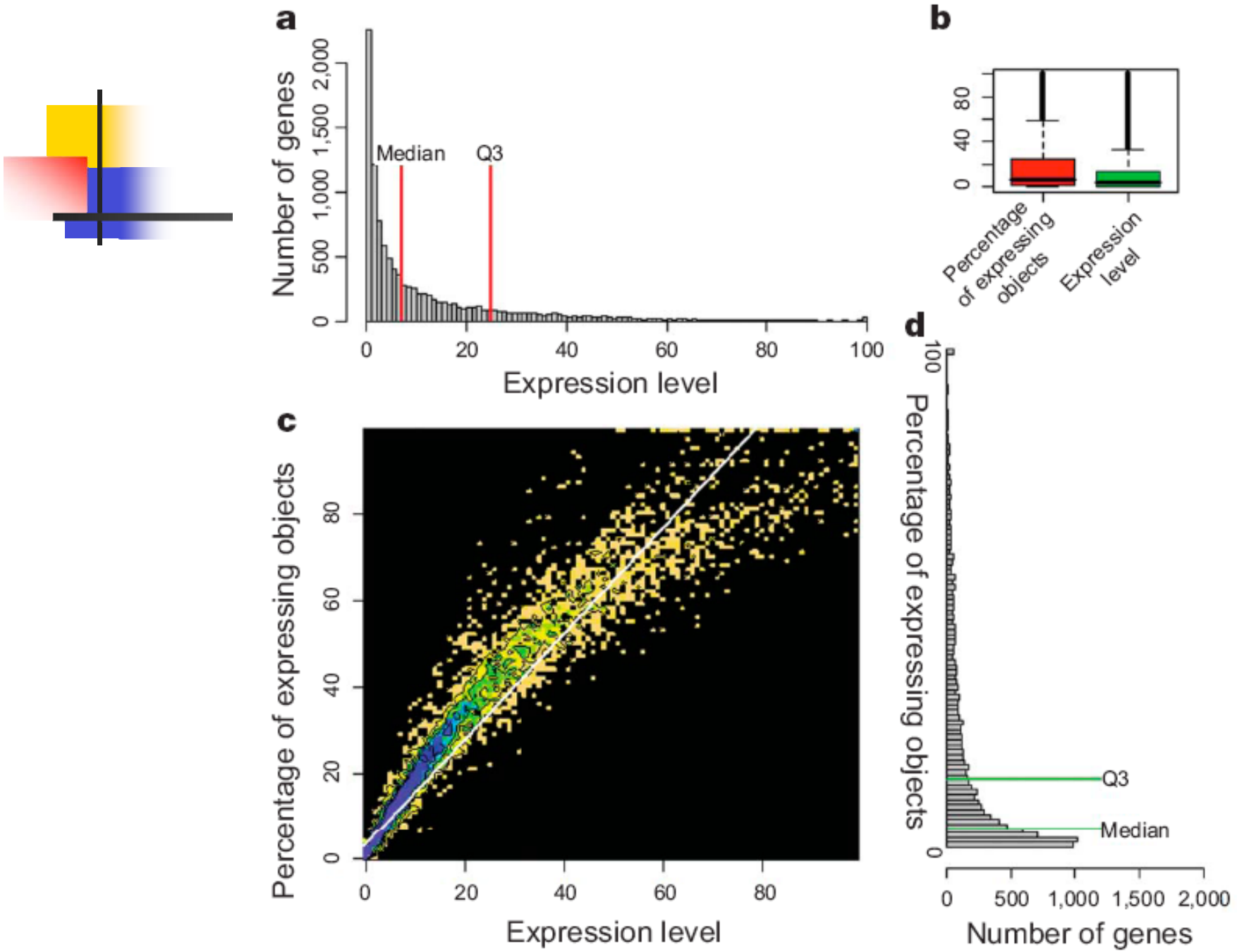
$$L = \frac{a_g}{a_{\max}} \bar{I}$$

- **Expression Density** D, n_g is the number of expressing cells, n_{max} is the maximum number of expressing cells derived from a set of ubiquitos genes. [0, 100]

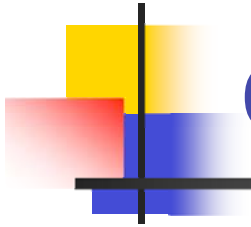
$$D = \frac{n_g}{n_{\max}}$$

Expressed genes

- 80% of genes display some cellular expression above background in the brain.
- 70.5% of genes are expressed in less than 20% of total cells.


a
Number of genes
0
500
1,000
1,500
2,000
Median
Q3
0
20
40
60
80
100
Expression level
b
0
40
80
Percentage of expressing objects
Expression level
c
Percentage of expressing objects
0
20
40
60
80
0
20
40
60
80
Expression level
d
Percentage of expressing objects
0
100
Q3
Median
0
500
1,000
1,500
2,000
Number of genes

Enriched expression in major cell types



- **Expression Pattern**
  - A binary classification represented by the labels, clustered and not clustered.
  - Not clustered => uniformly distributed throughout the anatomic structure.
  - Clustered => has some regional aspect in the anatomic structure.
- The result is a binary vector over all anatomic structures.

Finding enriched genes

- Seeding with known structure-specific genes.
  - Oligodendrocyte (Mbp, Mobp, Cnp1)
  - Choroid-plexus (Col8a2, Lbp, Msx1)
- Find the genes with similar expression patterns.

Result

- All well-established markers for different cell types were identified.
- GO enrichment analysis show different aspects.
  - Oligodendrocyte-enriched genes => myelin production.

Regionally enriched gene expression

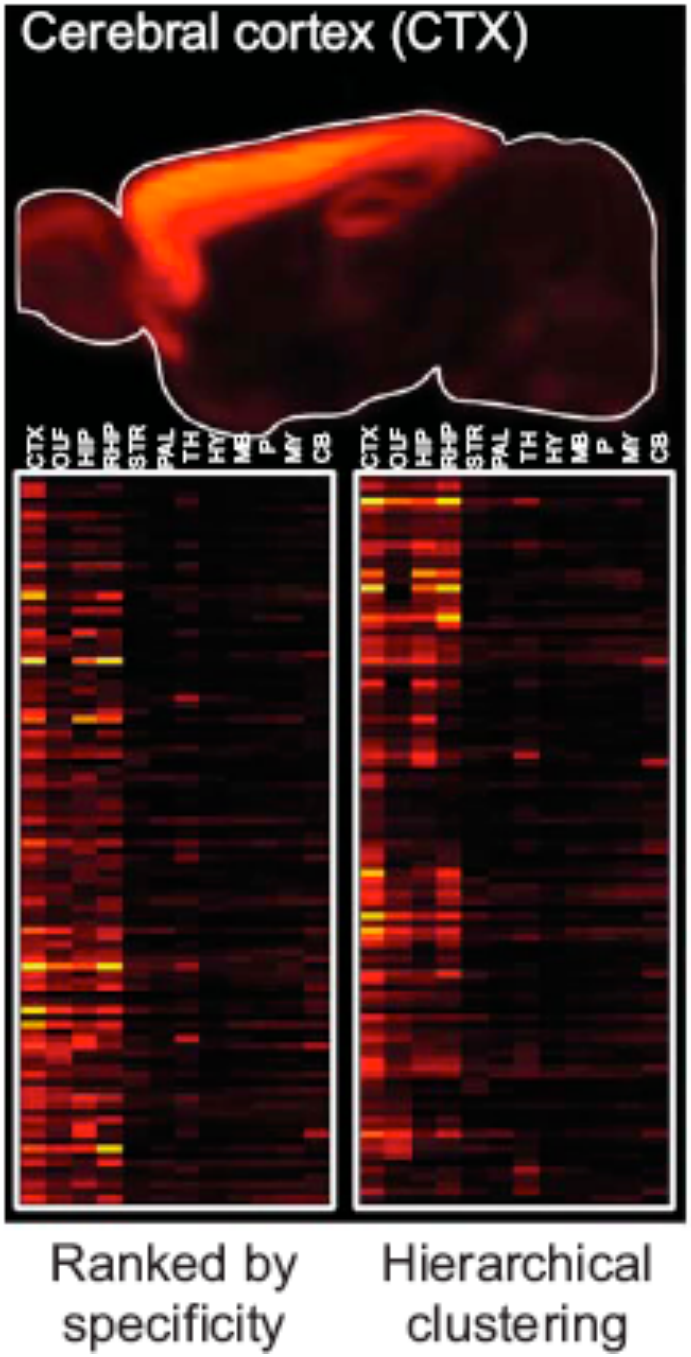
- Genes with regionalized expression patterns provides potential substrates for functional differences between brain regions.

Finding regional-specific genes.

- For each gene, the area occupied by signal is A and the area in the region occupied by the signal is B. B/A is the percentage of signal in the region.
- For 12 major brain regions, 100 top genes are reported for each region.

Result

Top 100 genes



Similar regions are clustered together.

Cluster of correlated gene expression

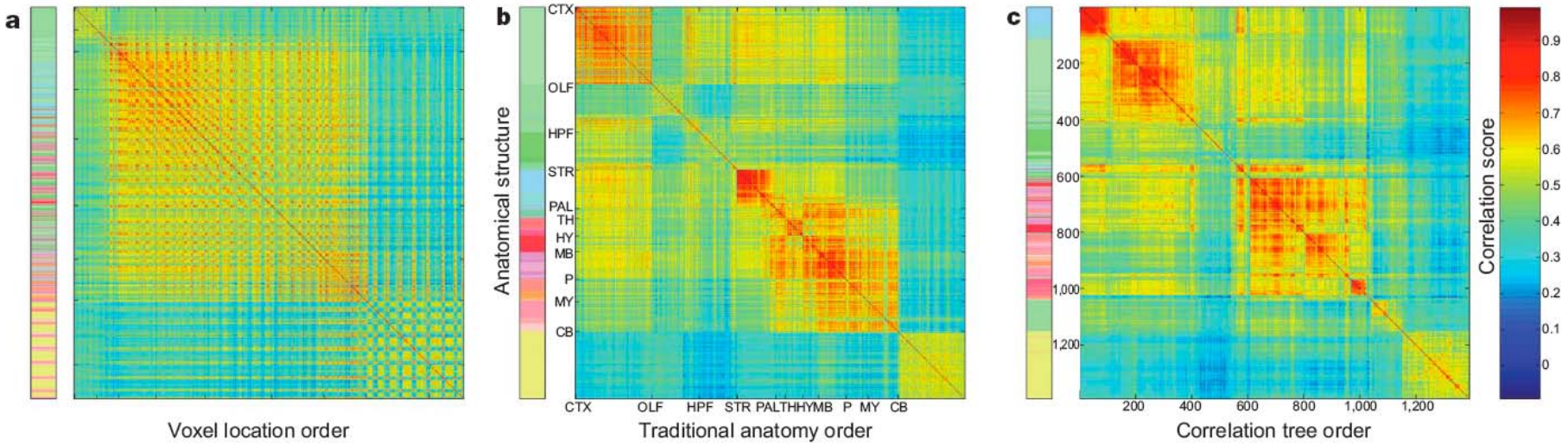
- Classical definition of brain regions
  - Overall Morphology
  - Cellular Cytoarchitecture
  - Ontological Development
  - Functional Connectivity

Use gene expression for finding brain regions

- Genes with regionalized expression (not the brain region) are used.
  - Initially, 5195 genes are selected. The expressing fraction is between [0.2, 0.14]
- The non-expressers and genes expressed widely are excluded.
- Grid level correlations are computed.
  - The voxel size is 300 μm^2. Among 7000 voxels, 1500 representative voxels are selected.

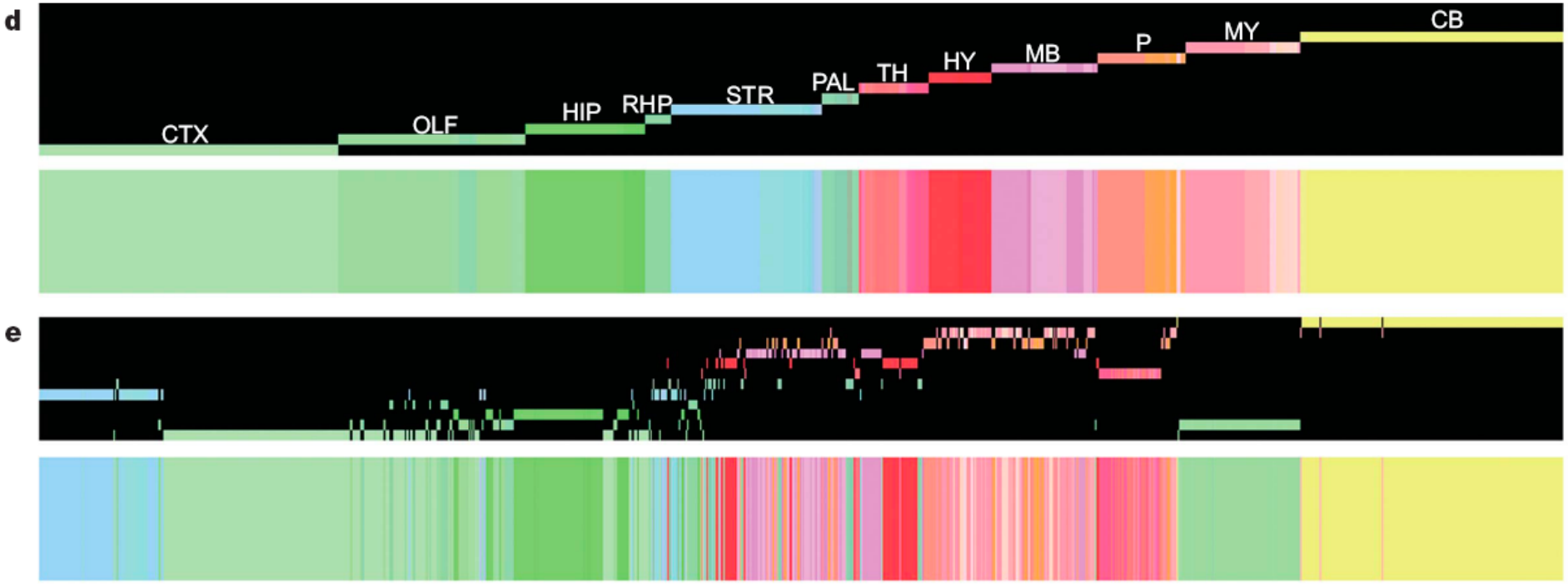
Result

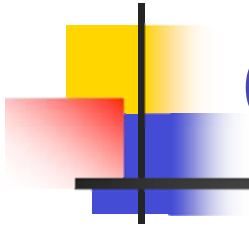
3D to 1D by RGB coding



Within each region, the correlation tends to be higher.

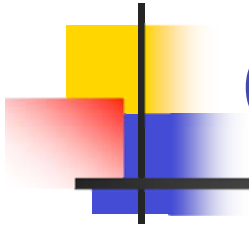
Result





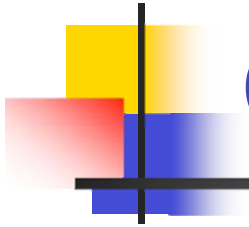
Comments

- The data produced in this paper is huge!
- The expression “profile” is in fact an 3 dimension matrix.
- The gene expression could be used to differentiate the cell types and possibly discover the sub-cell types. Those genes will be the potential marker for Histology staining.



Comments

- The inbred mouse strain plays important roles in this data. Without it, the gene expression measured on different mice would be subject to the genetic variation.



Comments

- Is this really better than microarray approach?

ABA Data Production Process

- 1000 slides/4000 brain sections daily

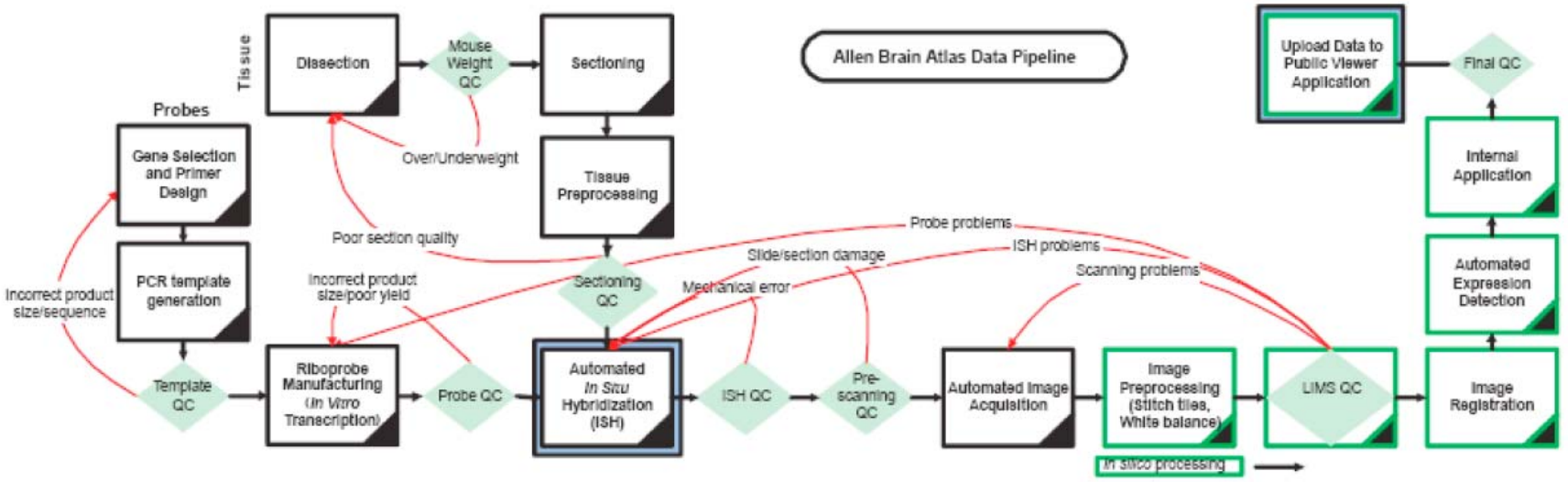
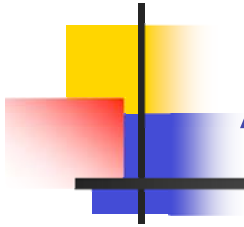


Figure 1 - *ABA production process*



ABA Process

- In their data, 20000 genes are measured. 20000x(5 or 14) slides.
- That would take around 1 year to finish all slides.

Riboprobe

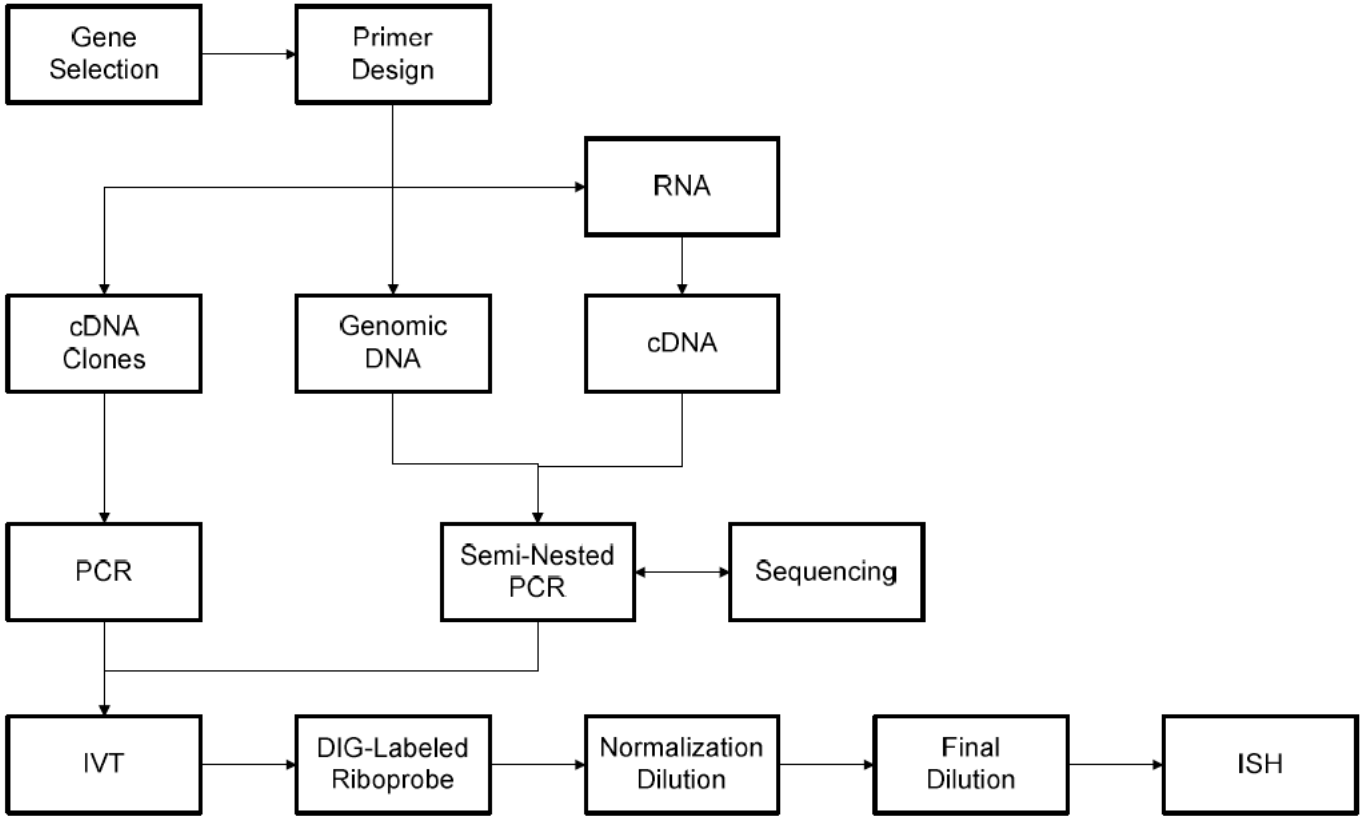


Figure 2 - *ABA Probe Production Workflow*