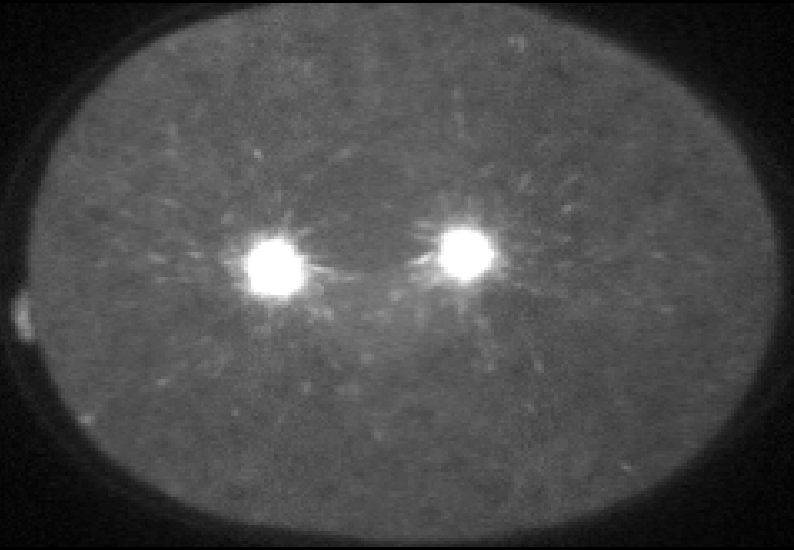


Zooming in On Life

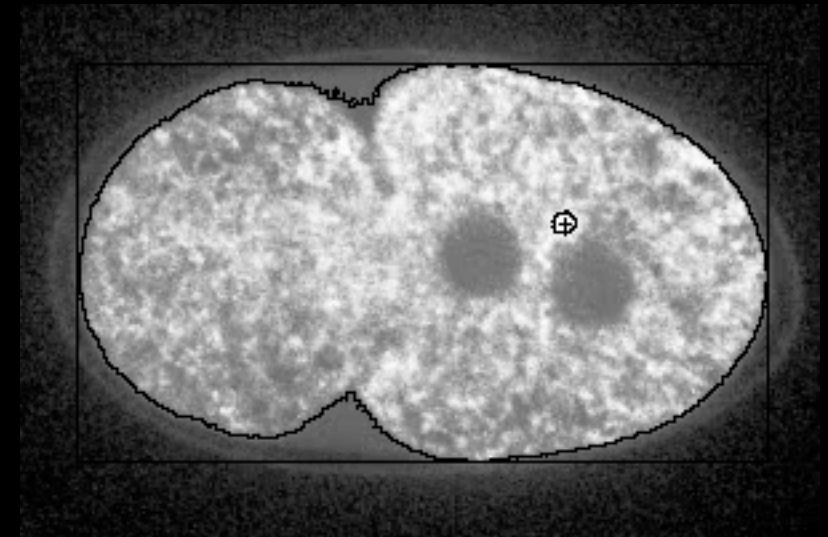
Gene Myers
Group Leader

HHMI Janelia Farm Research Campus

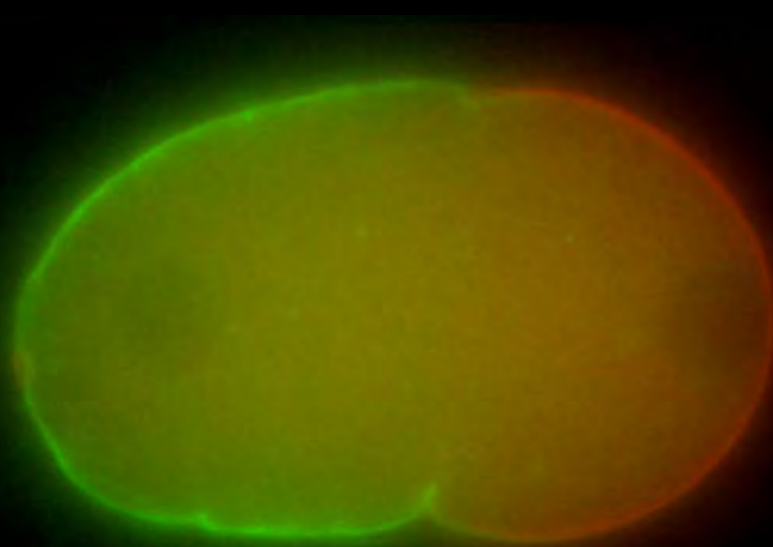
Yes, we can! (zoom in)



EB1-labelled
tubulin fibres.



Gamma-tubulin-labelled
centrosomes



Par2-Par6-labelled
membranes

Courtesy Tony Hyman Group
MPI, Dresden

Genesis: The Human Genome Project

1996 Weber & Myers propose a radical paired-end whole genome shotgun (PE-WGS) protocol for sequencing the human genome:

One 50,000,000-piece puzzle vs. 50,000 1,000-piece puzzles.

The Reaction



The critics said it wouldn't work: "The Swiss Cheese Genome"

Genesis: The Human Genome Project

1996 **Weber & Myers** propose a radical paired-end whole genome shotgun (PE-WGS) protocol for sequencing the human genome:

One 50,000,000-piece puzzle vs. 50,000 1,000-piece puzzles.

1998 **Celera** is formed to do so with **Craig Venter** as CEO

Celera's Sequencing Factory (1999)



Genesis: The Human Genome Project

1996 **Weber & Myers** propose a radical paired-end whole genome shotgun (PE-WGS) protocol for sequencing the human genome:

One 50,000,000-piece puzzle vs. 50,000 1,000-piece puzzles.

1998 **Celera** is formed to do so with **Craig Venter** as CEO

2001 Celera “ties” the public project on the **human genome**, but also sequences the **fruitfly**, **mouse**, and **mosquito** by the end of 2001.

Now all genomes are sequenced with PE-WGS

Computer Science served a critical enabling role:

- novel protocol
- a solution to a daunting computation

Next-Gen Sequencing in 2009



1 of these
=
Celera



Outpacing Moore's
Law

New algorithmic
challenges

Computations
severely strain
existing hardware

What's Next

- Next-next gen sequencers

Your genome for < \$1,000

- “Genotype-phenotype correlation”

Your genome's implication for your health and treatment

- “Sequencing everything”

Answers key questions in diversity, evolution, cancer, pathogenicity, etc.

- Synthetic biology ...

All of these involve vast amounts of data.
Computation and analysis are the bottlenecks!!!

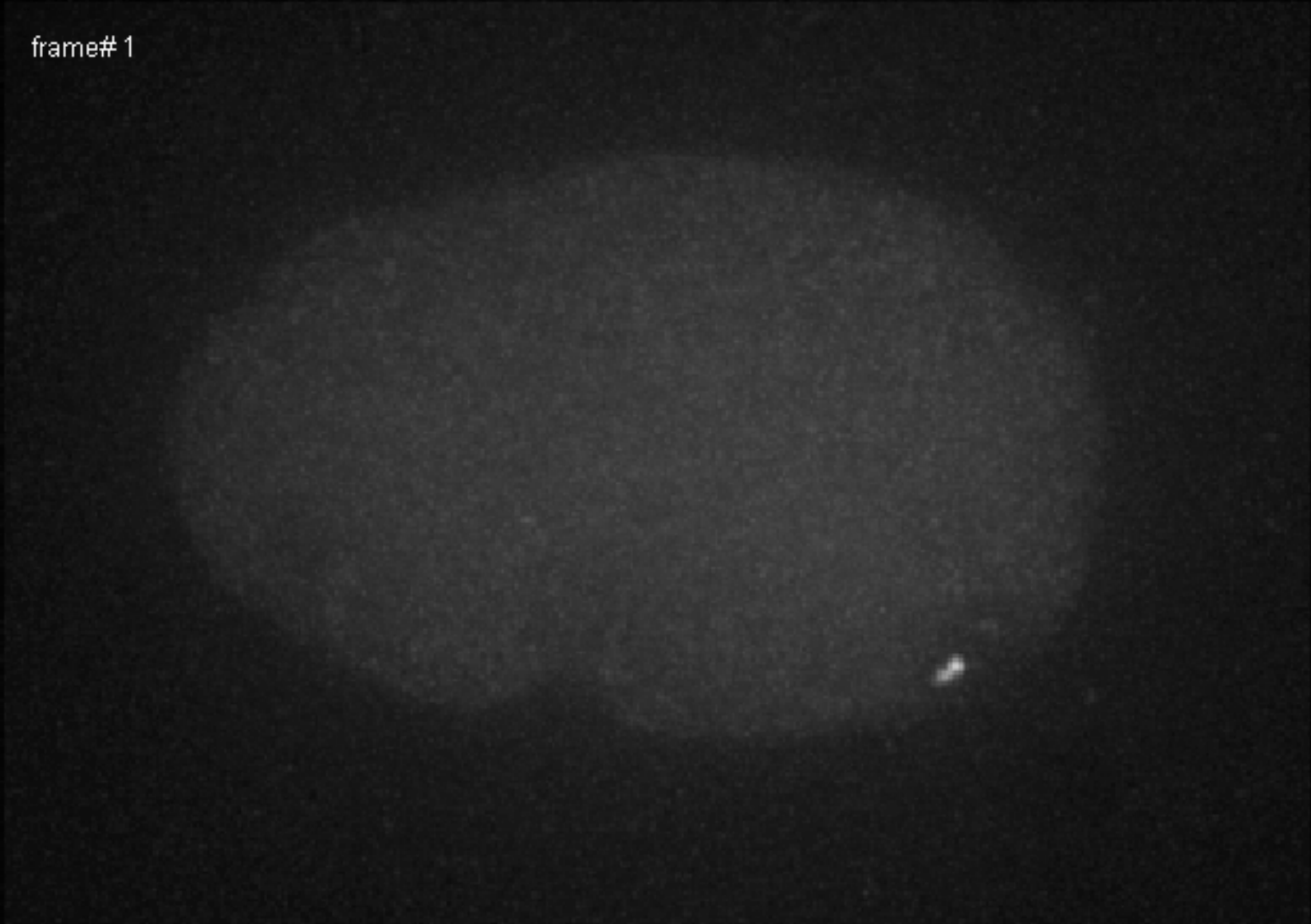
My view

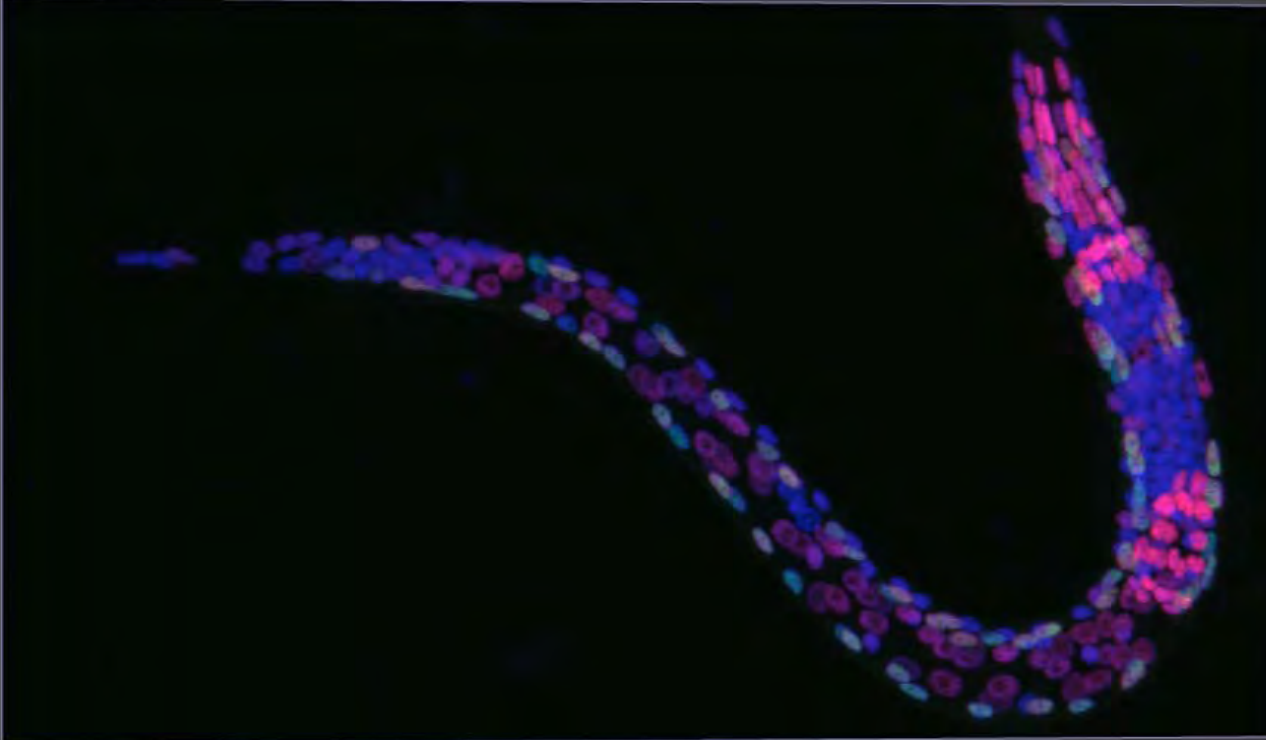
- In my view, the most significant outcome of the genome project(s) is that it permits us to do recombinant genetics on an entire genome.
- That is, it enables “systems genetics”
 - O’Shea: Protein-GFP of all genes in yeast to get localization (*Nature* 425 (2003), 737-41.)
 - Hyman: RNAi knockdown of all genes in worm to understand mitosis. (*Nature* 434 (2005), 462-9.)
 - Rubin: Promotor-GFP fusions in fly to get fine-grained control of functionally-related neurons.
- Interpreting the, typically visual, results of these system-wide studies computationally is unavoidable.

“BioImaging Informatics”

Intracellular Processes: Centrosome Formation

frame# 1

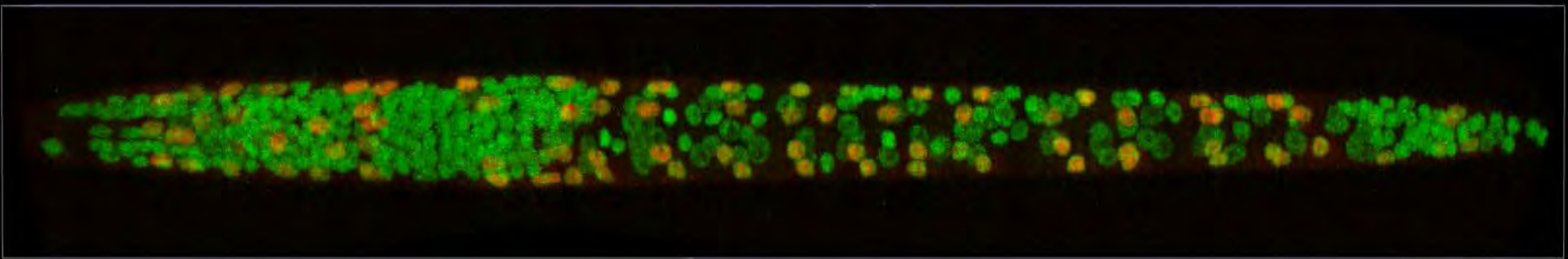




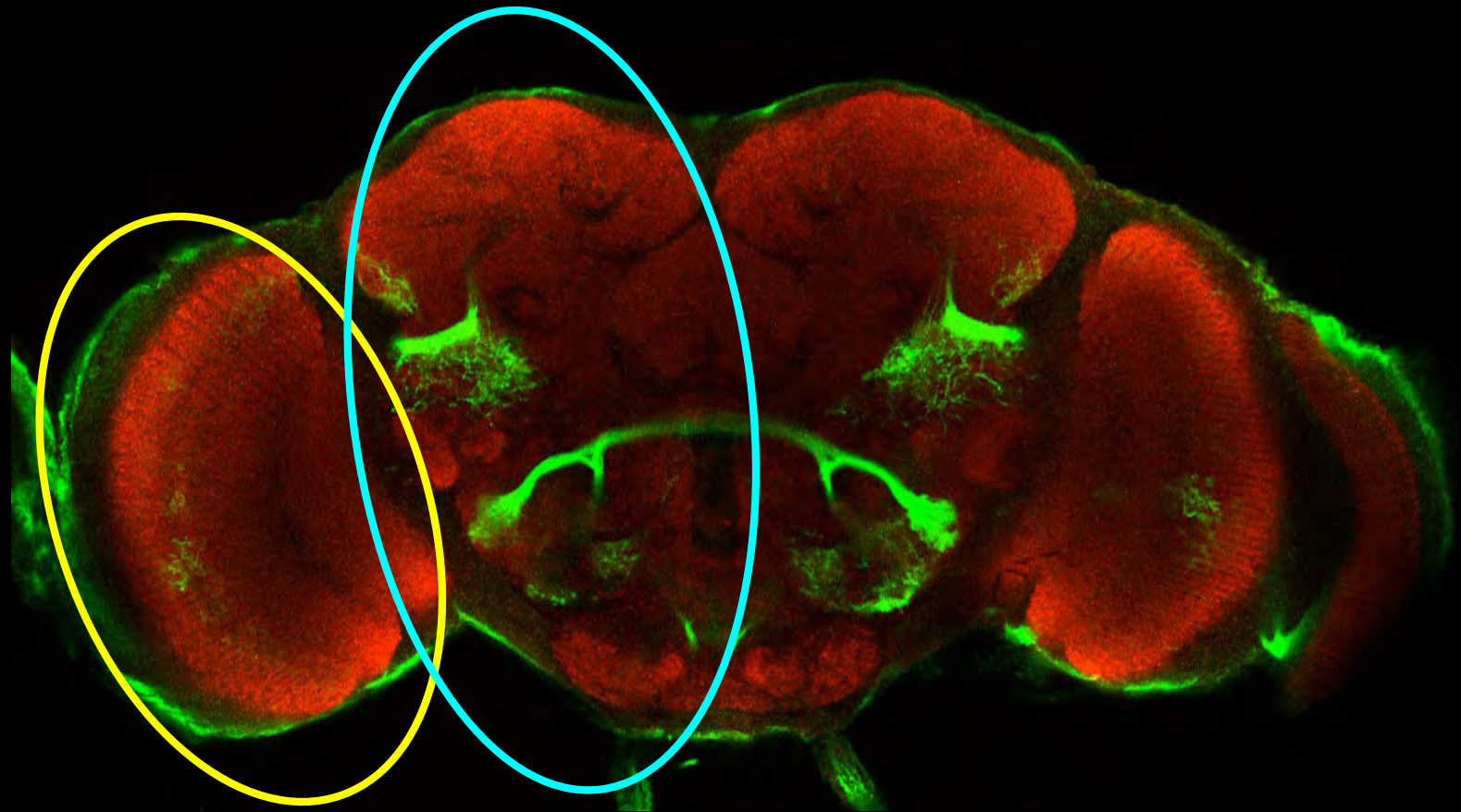
- L1, 558 nuclei
- Leica confocal
63x/1.40 oil lens
- **DAPI**: all
- **GFP**: 82 muscle cells
(fiducial)
- **mCherry**: subject

w. Stuart Kim
Stanford

Cellular Collections: Worm Atlas



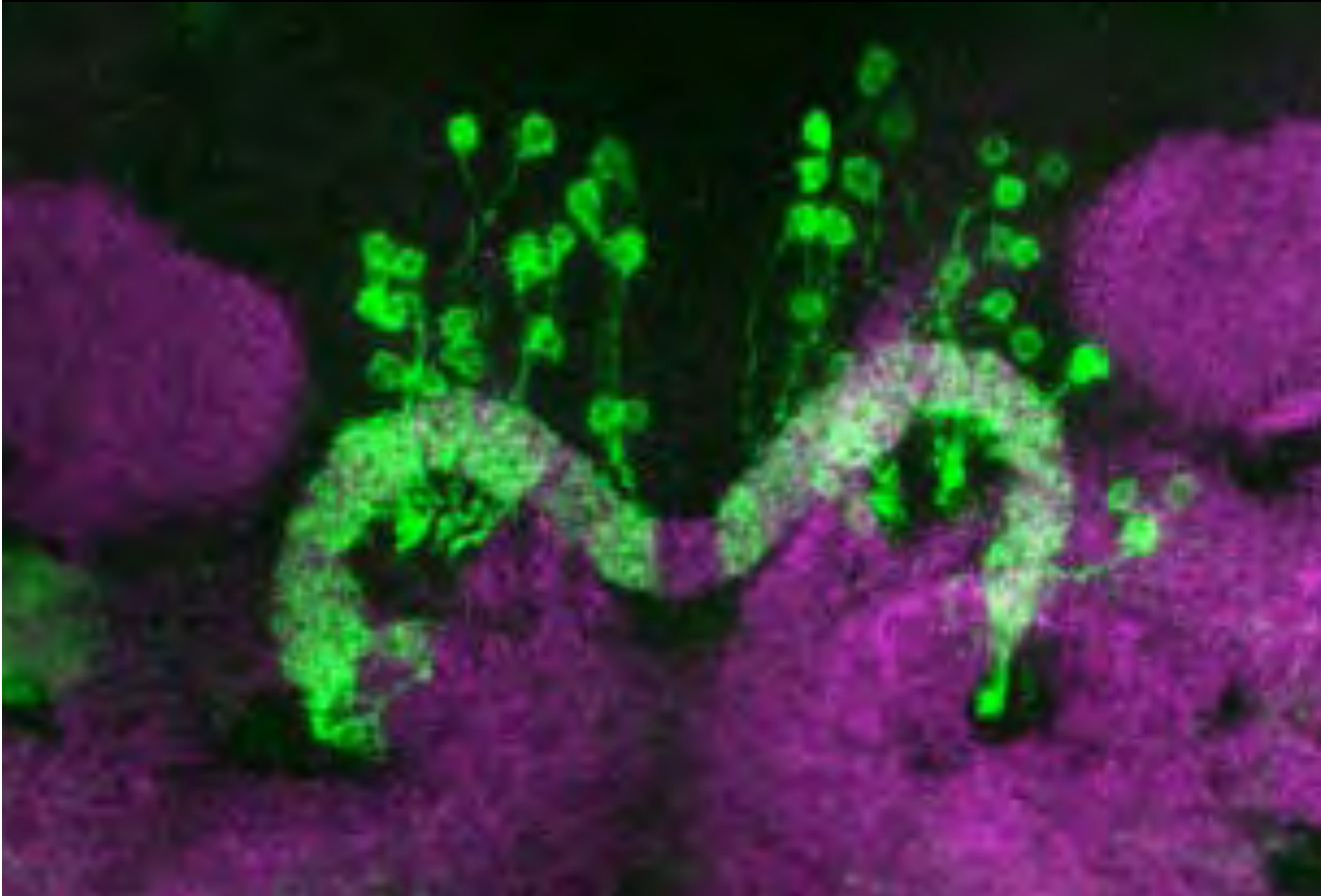
Cellular Collections: A Fly's Brain



Optic Lobe
30K neurons

Core Hemisphere
20K neurons

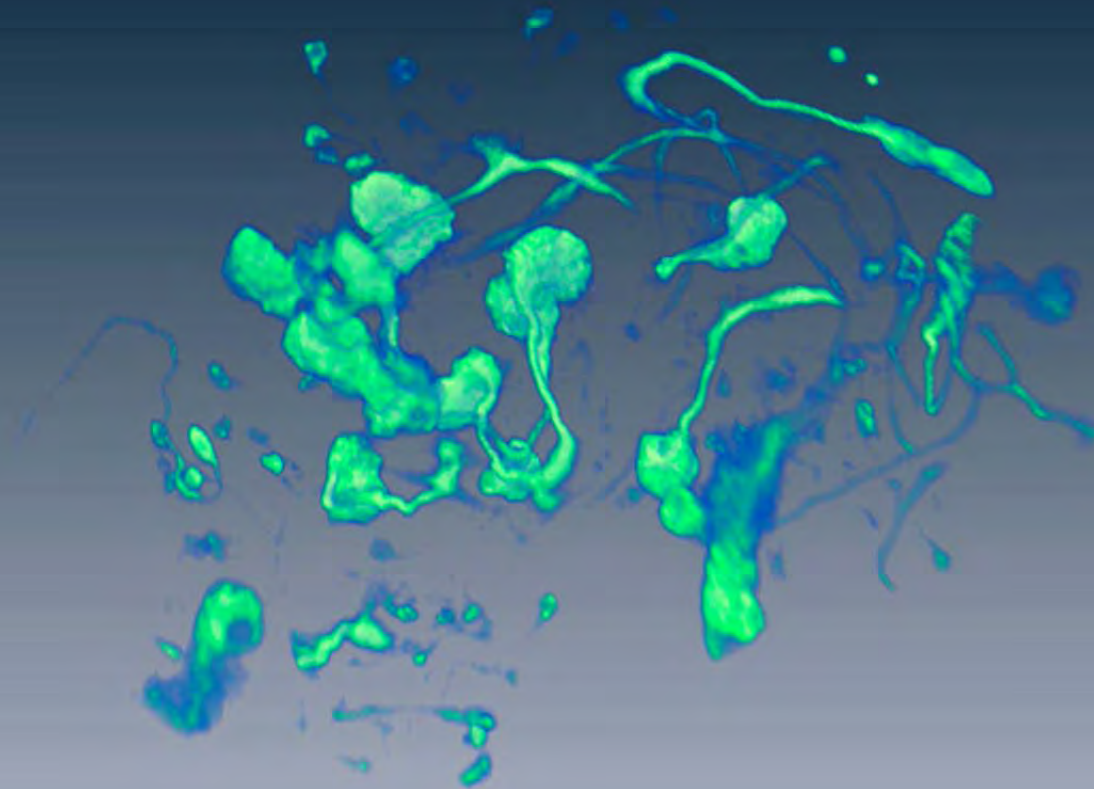
Fly's Brain (continued)



Courtesy
Arnim Jenett,
Gerry Rubin

Want the fine-grained arbor to arbor co-occupancy

A Fly Neural Lineage

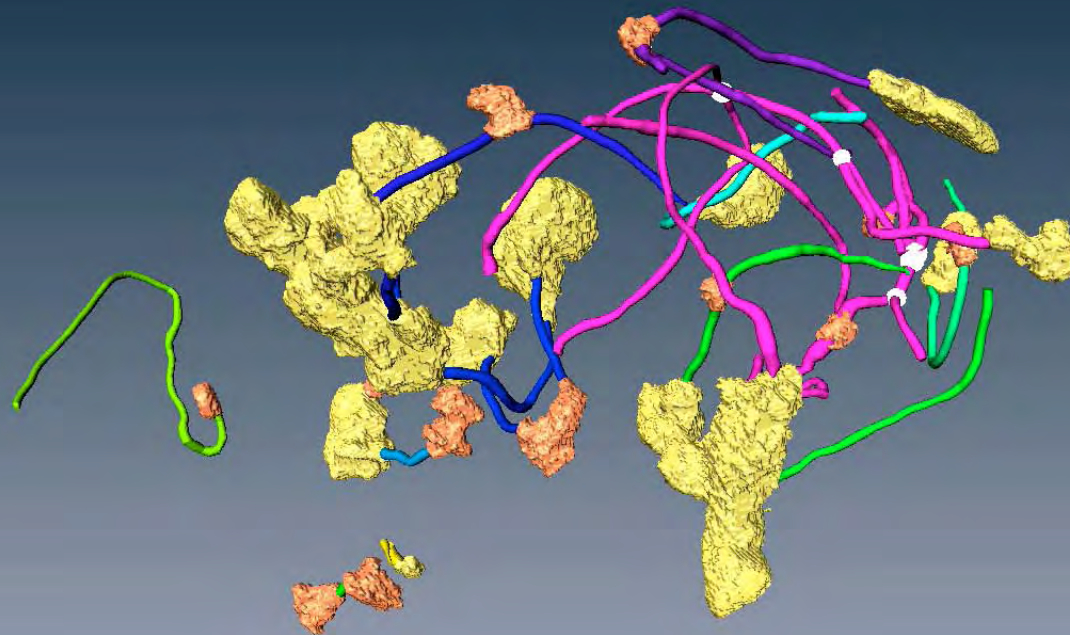


Extracted Model

Yellow:
soma

Brown:
arbor

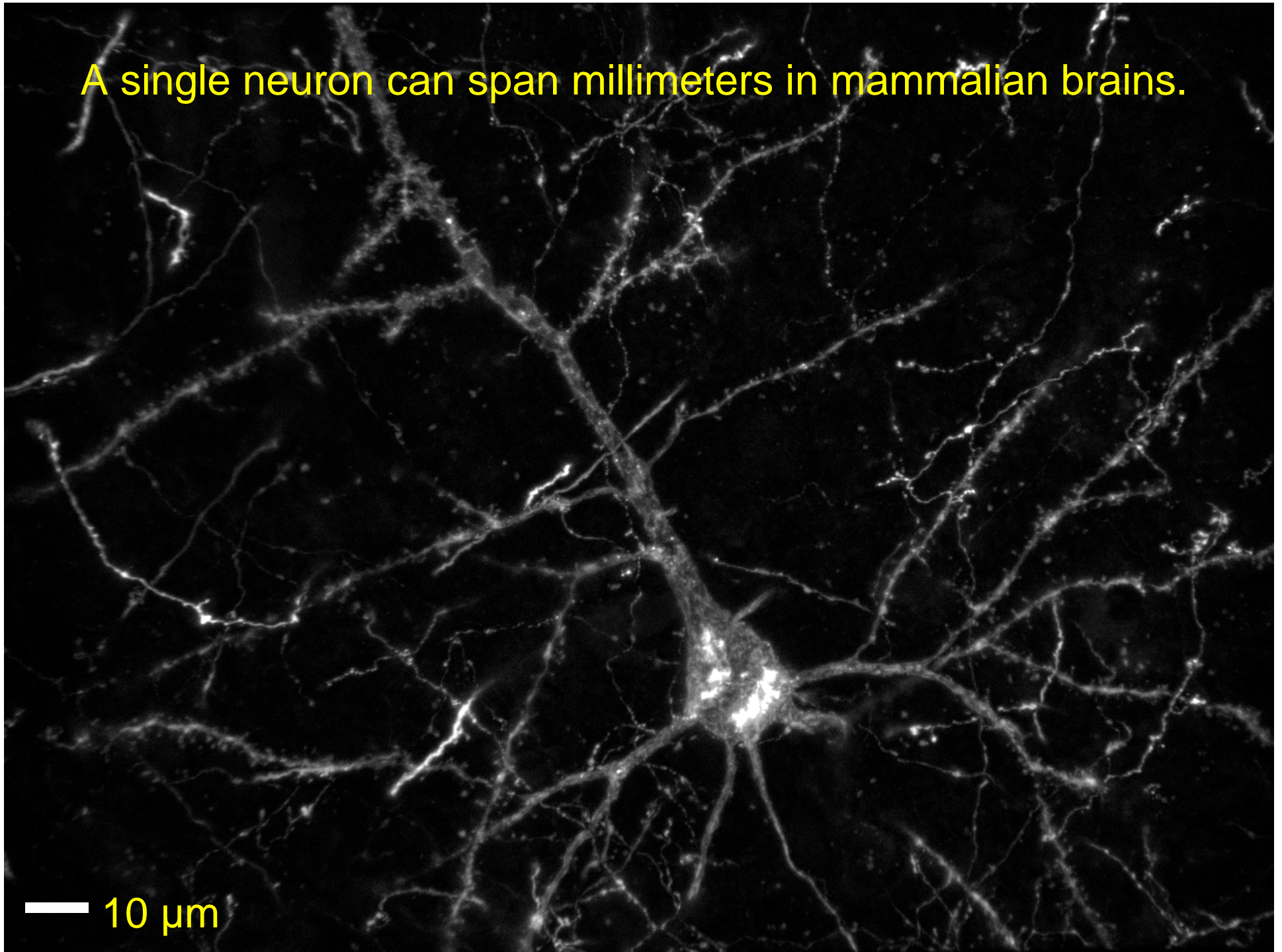
White:
branch

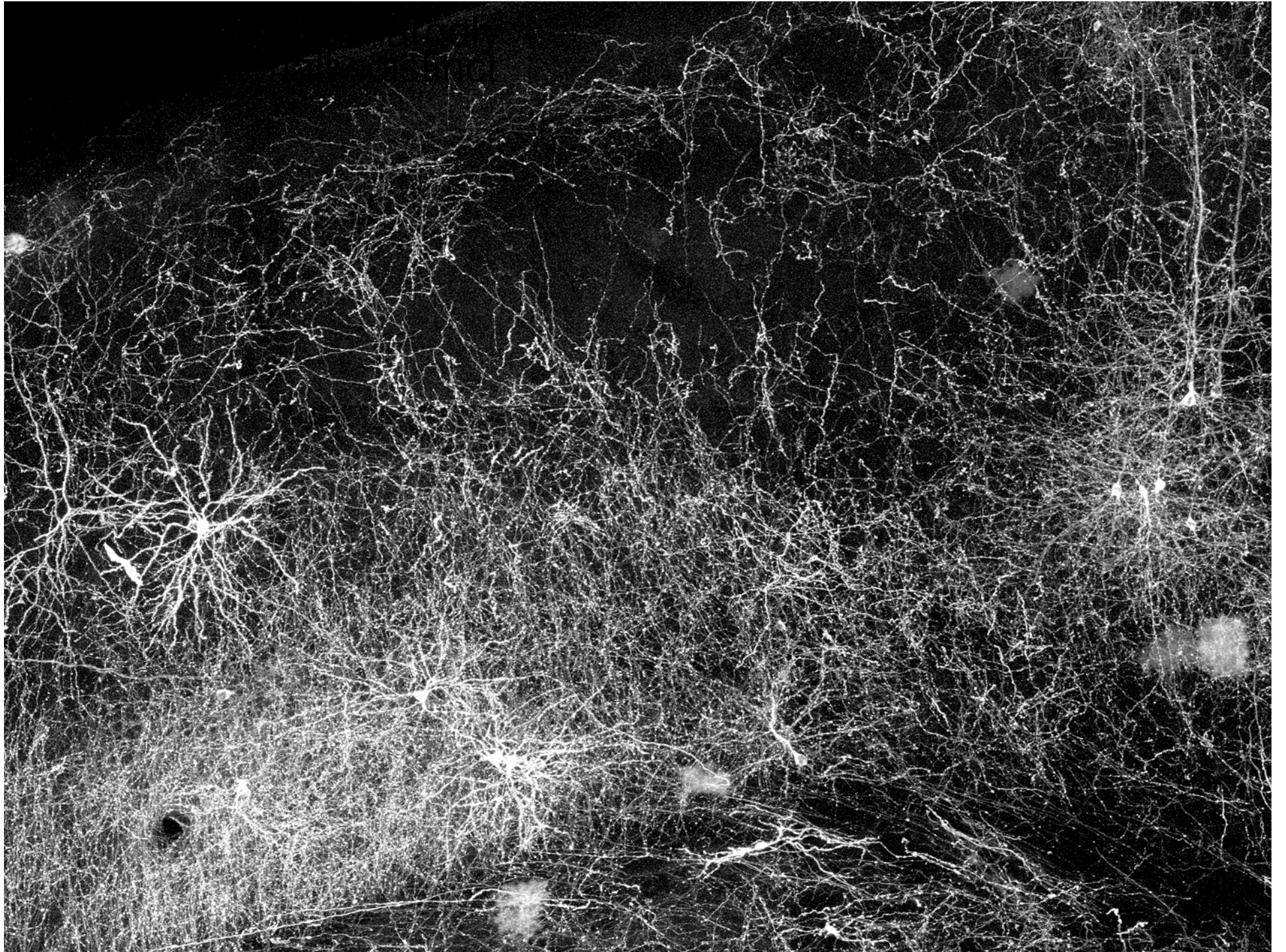


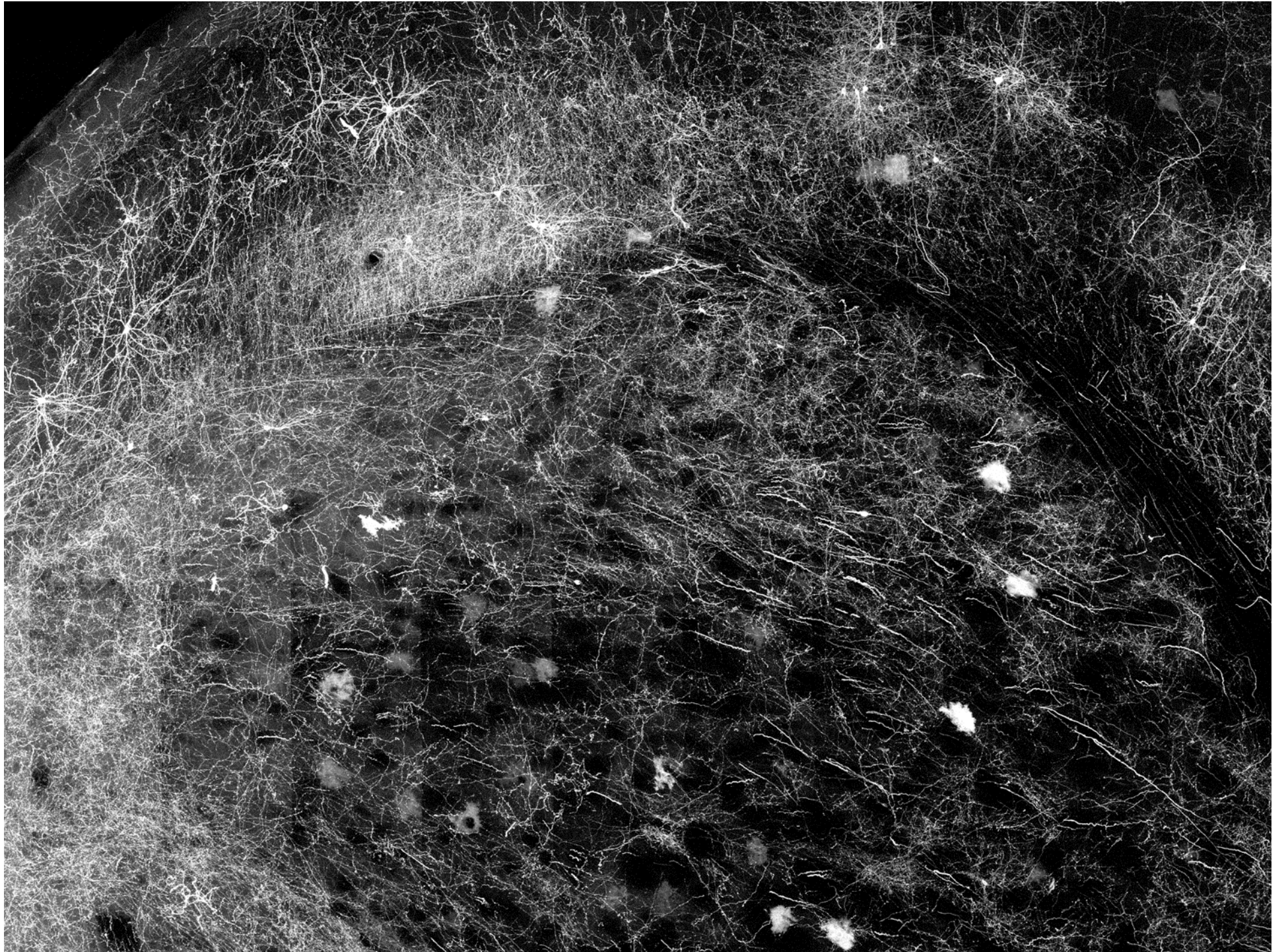
Hope to Build An Atlas of Every Neuron In A Fly's Brain



A single neuron can span millimeters in mammalian brains.

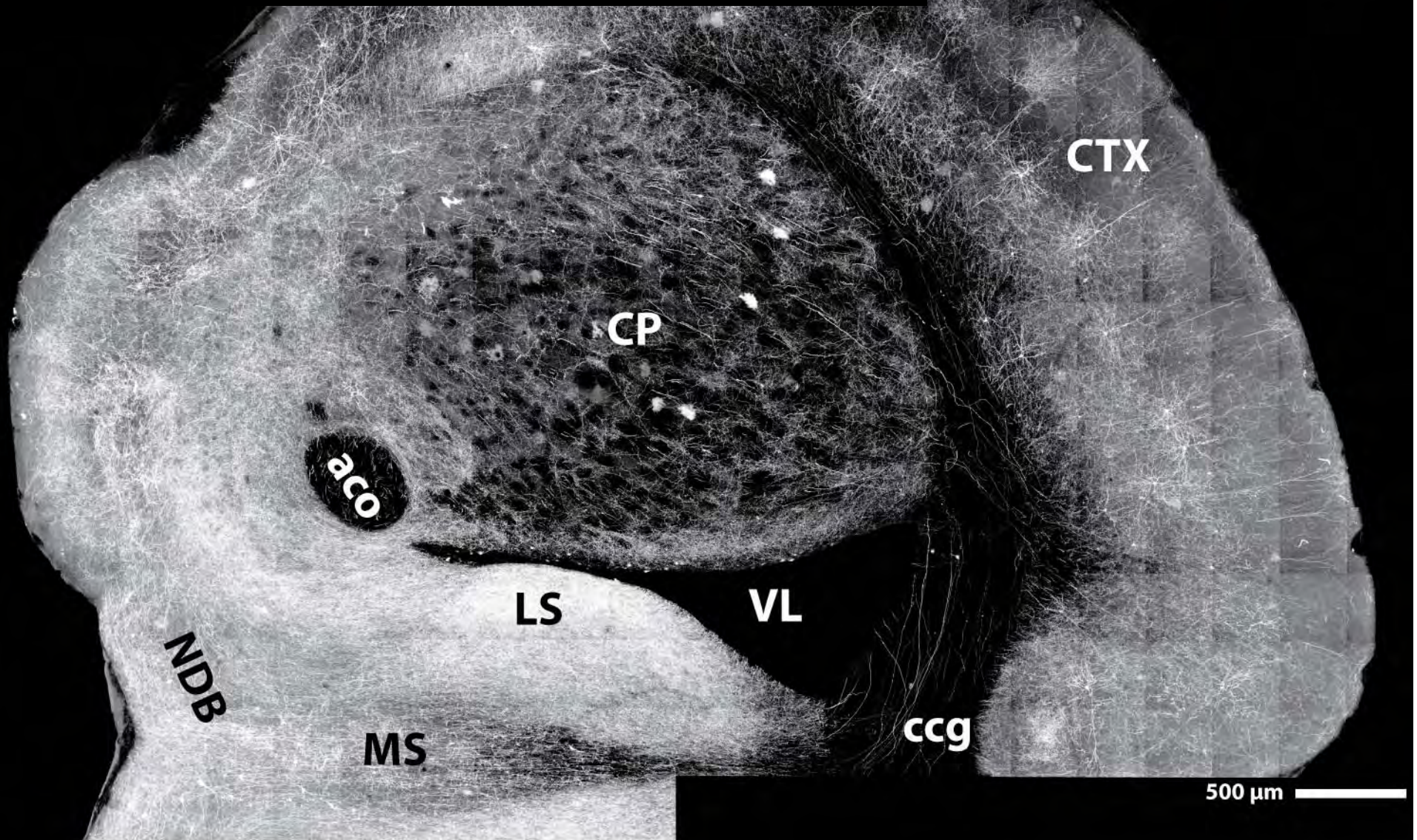






We are building a microscope that will
image an entire mouse brain in 1 week

4.2 trillion voxels!



A Mouse Brain Atlas?



**100K
neuron
tracings
per brain!**

In Conclusion

- Genomics is driving Molecular Biology into a big data science. Even microscopes are becoming high-throughput instruments.
- We can map genomes, can we now map
 - cells at the “particle” level?
 - organs/organisms as collections of (developing) cells?
- Computation is the bottleneck in every project
 - Massive amounts of data
 - Must perform at human-or-better levels else curation becomes the bottleneck (e.g. EM reconstruction)

Acknowledgements

Myers Group

Fuhui Long (*C. elegans*)

Ting Zhao (neuron tracing, stitching)

Jun Xie (neuron tracing)

Nathan Clack (mouse brain)

Peng Group (fly brain registration)

Svoboda Group (mouse samples)

Danny Connors (whisker data)

Stuart Kim Lab (*C. elegans*)

Stanford

Xiao Liu

Kendall Wu

Tony Hyman Lab (Mitosis)

MPG-CBG Dresden

Karsten Hogue

Martin Stracko