CNR-INO
ISTITUTO NAZIONALE DI OTTICA
CONSIGLIO NAZIONALE DELLE RICERCHE

LENS
European Laboratory for Non-Linear Spectroscopy

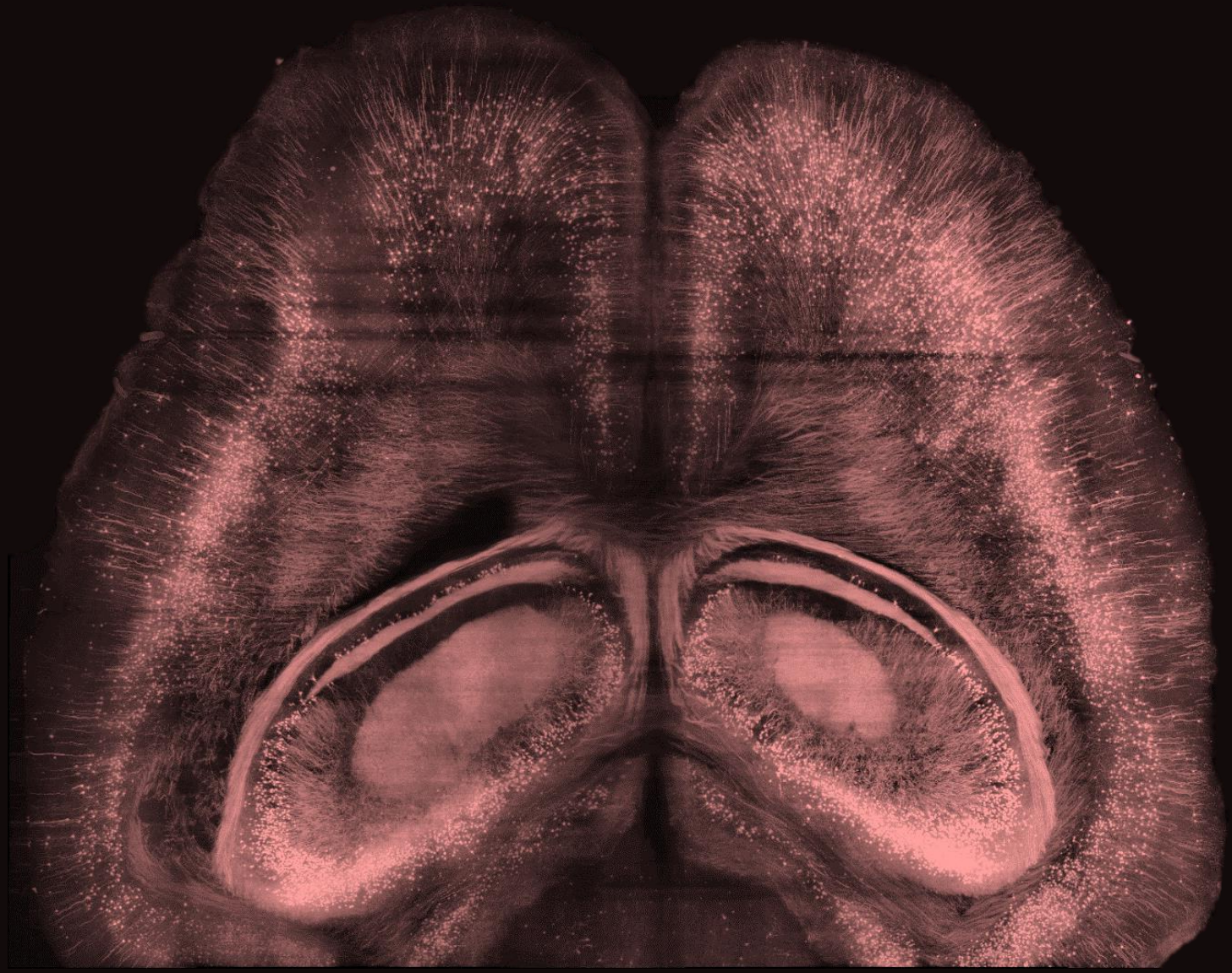


Image processing and management of large datasets in Light-Sheet Microscopy

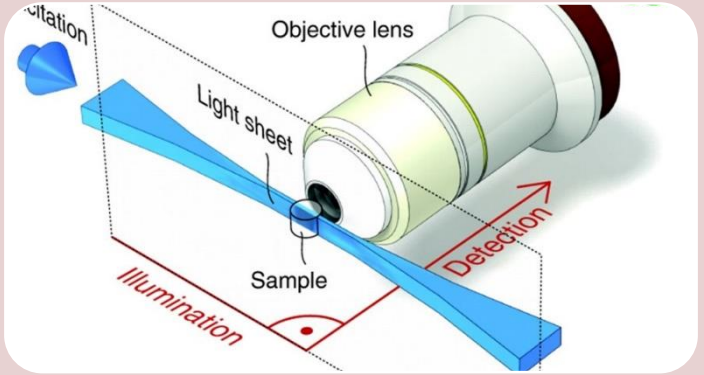
Giacomo Mazzamuto

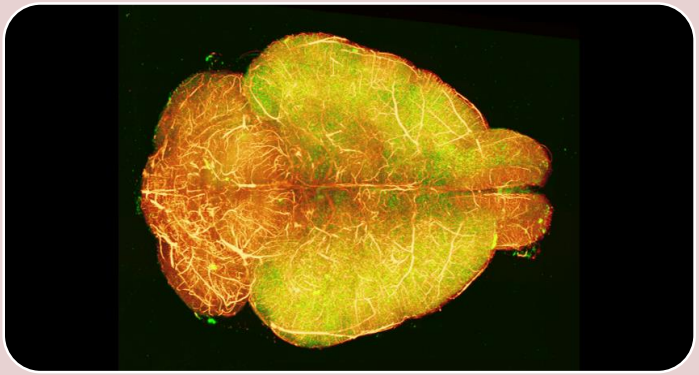
Laserlab-Europe / ELI/ CASUS Workshop

Better data for better science

29/10/2021

The three pillars of selective illumination microscopy





Optical clearing

Instruments

Light-Sheet
Two-Photon

Image processing

Microscopy Imaging: acquisition and processing (old days)

Sample

Image acquisition

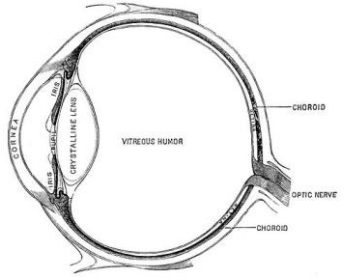
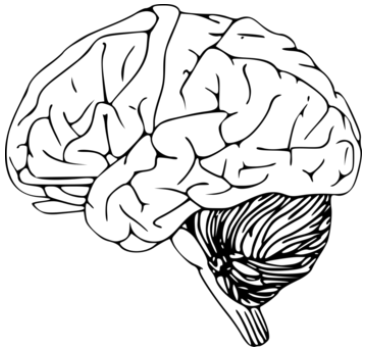
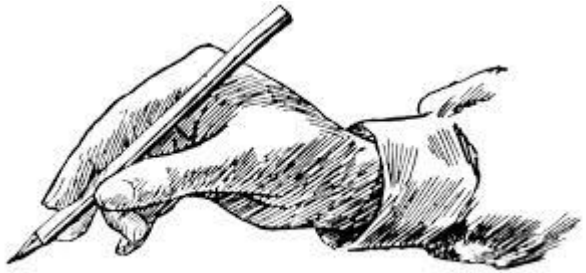
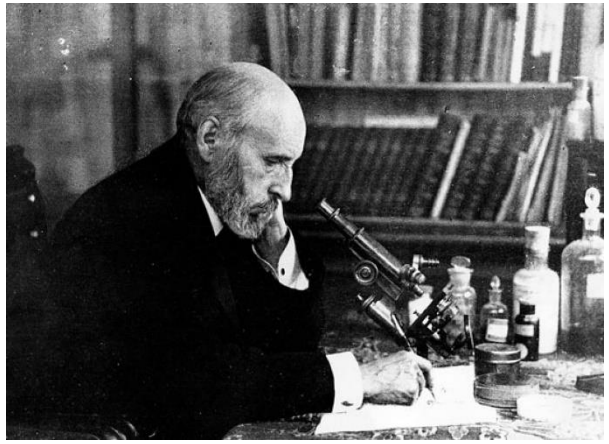


Image processing

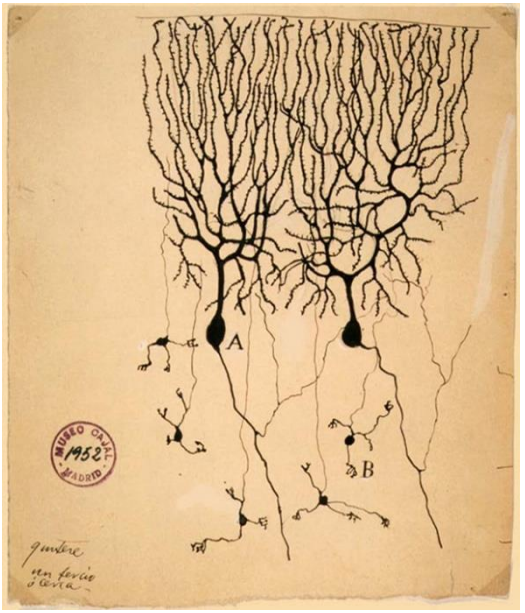


Visualization and storage





Santiago Ramón y Cajal at the microscope



Microscopy Imaging: acquisition and processing (today)

Sample

Image acquisition

Storage

Visualization

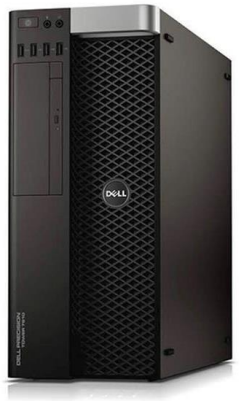


Image processing

Microscopy Imaging: acquisition and processing (tomorrow)

Sample

Image acquisition

Image processing

Visualization

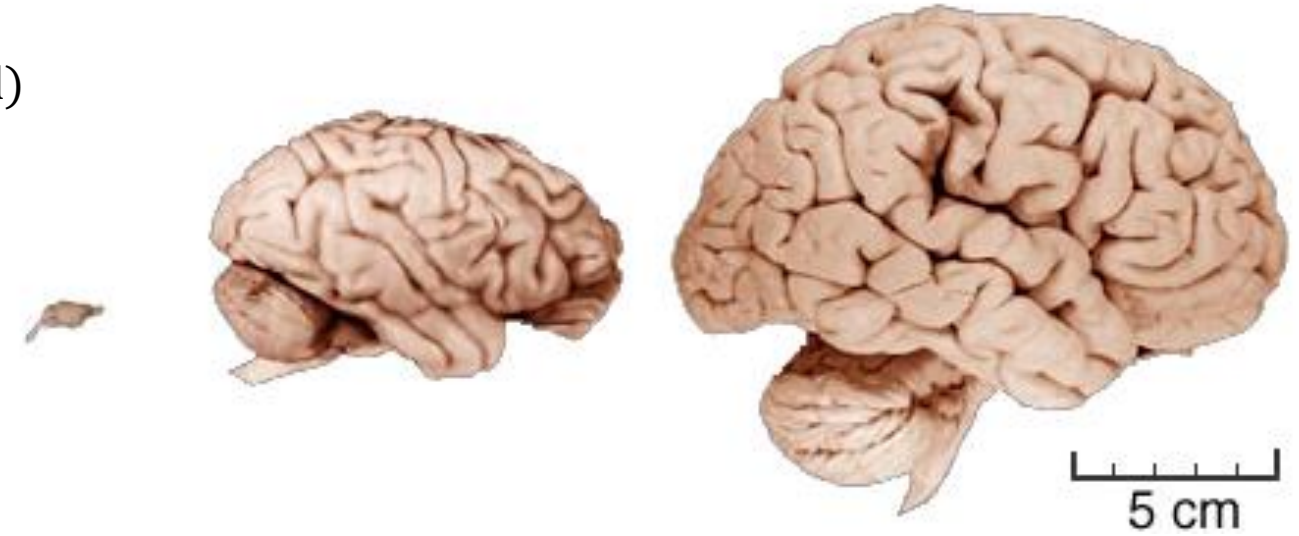
To be really scalable, image processing must be **as close as possible** to image acquisition

Storage

Big data challenges in whole-brain imaging

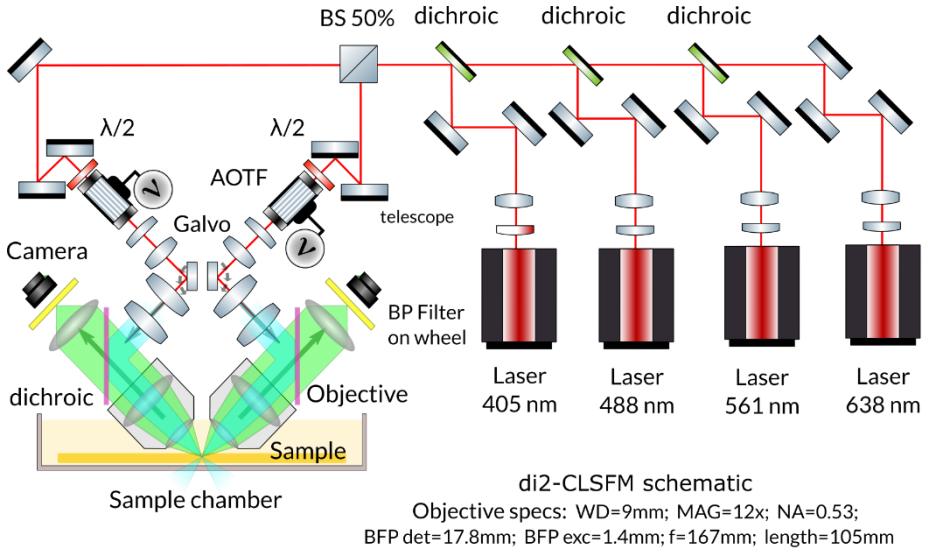
		Mouse	Monkey (Rhesus)	Human
Approx volume		1 cm^3	100 cm^3	1500 cm^3
Raw data size	1 channel, 10 µm res.	2 GB	200 GB	3 TB
	3 channels, 10 µm res.	6 GB	600 GB	9 TB
	1 channel, 1 µm res.	2 TB	200 TB	3000 TB
	3 channels, 1 µm res.	6 TB	600 TB	9000 TB

$$\frac{1\ \mathrm{cm}^3}{(10\ \mu\mathrm{m})^3} = 10^9\ \text{voxels} = 2\mathrm{GB}\ \ (@\ 16\text{bits/pixel})$$

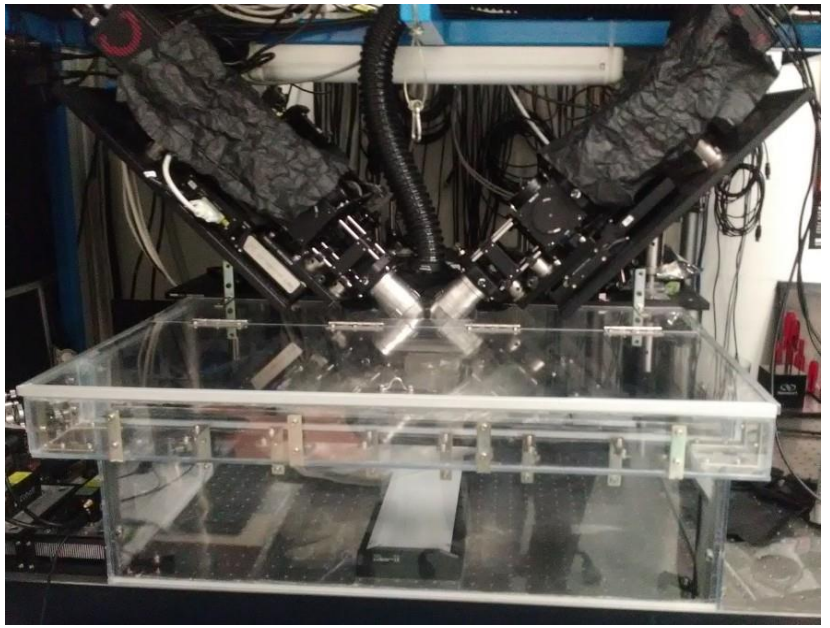


mouse monkey human

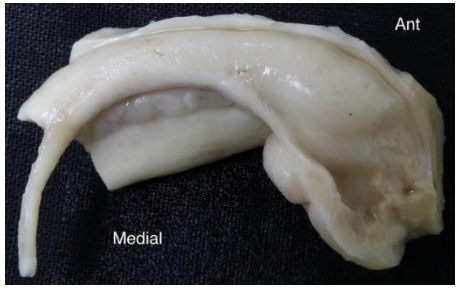
Dual view inverted Light Sheet Fluorescence Microscope



di2-CLSFM schematic
Objective specs: WD=9mm; MAG=12x; NA=0.53;
BFP det=17.8mm; BFP exc=1.4mm; f=167mm; length=105mm



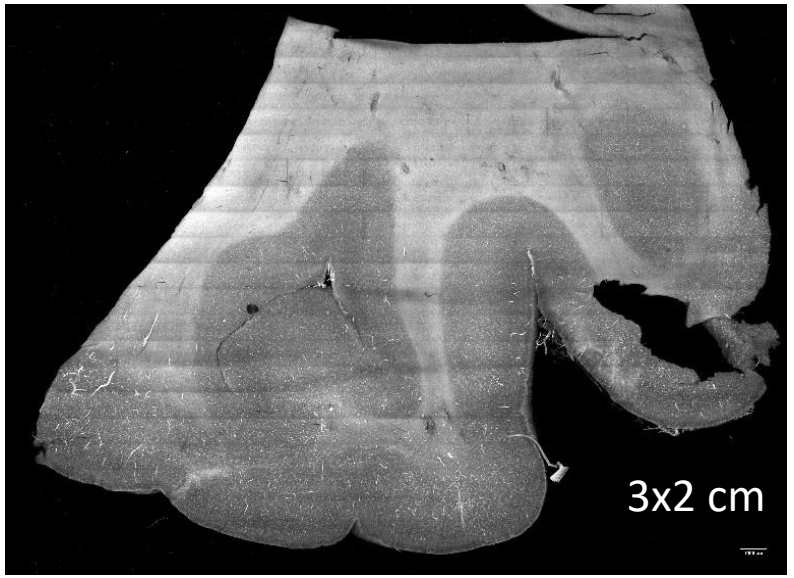
Human Brain Project

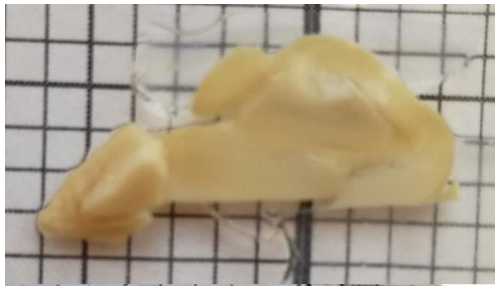


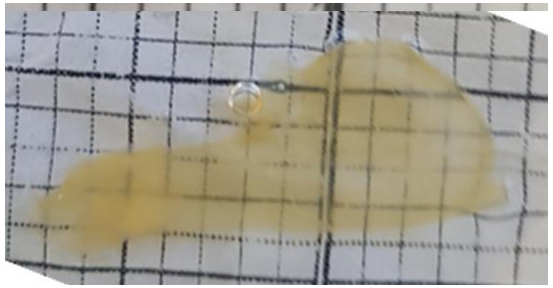
- dual-view inverted SPIM apparatus
  - Isotropic 1 um resolution
  - Up to 4 channels
  - Designed to image 1 mm thick sample
  - High data throughput (up to 1 GB/s, one channel)
  - Volumetric rate: 0.1 mm^3/s.

Hippocampus volume:
20 slices of 3cm x 1cm x 0.05 cm = $0.15cm^3$ = 0.15 x 10^{12} μm^3

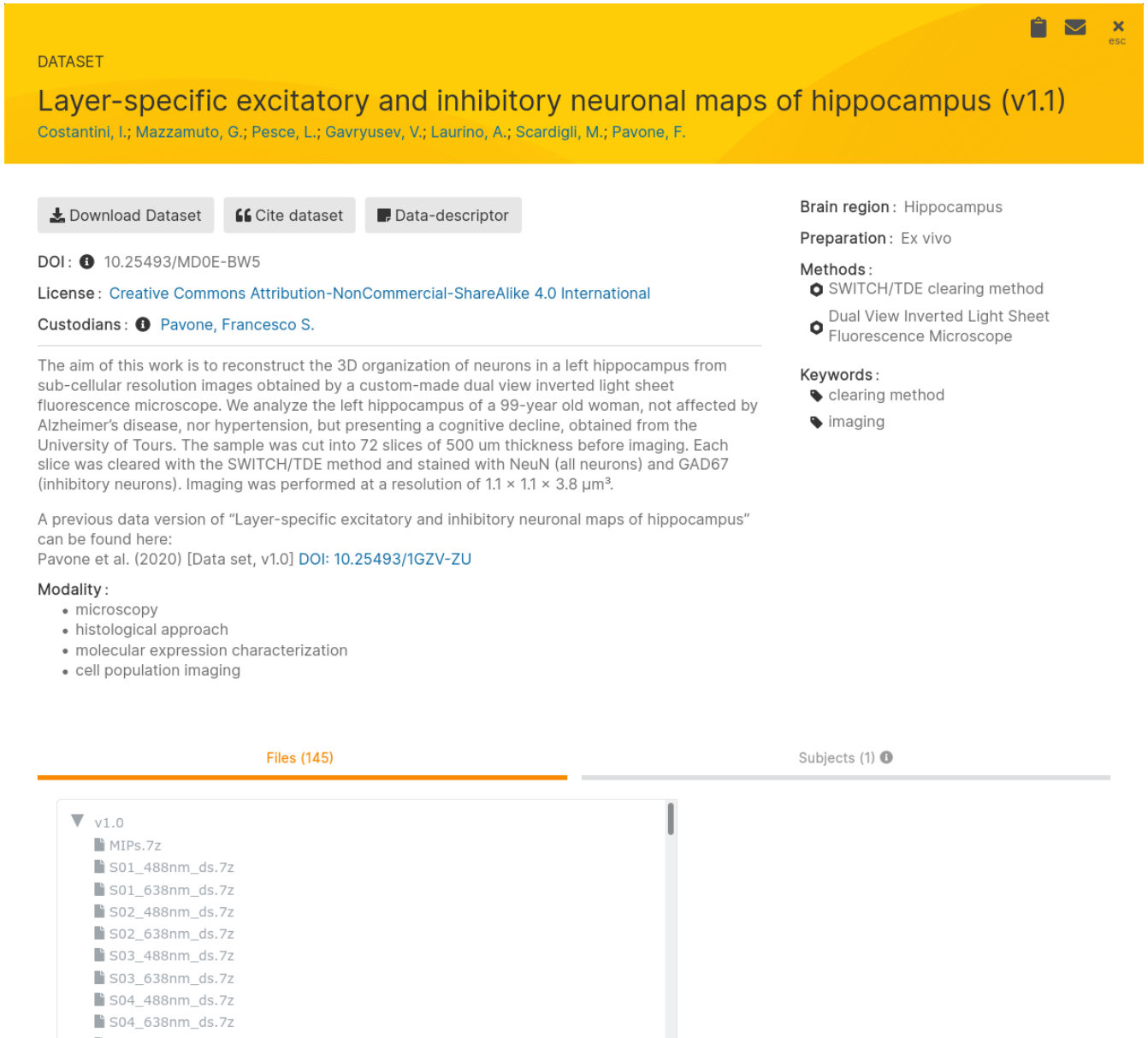
- Voxel size ≈ 1 μm^3 => 0.15 Tvoxels = 0.15 TB per channel = 0.6 TB per slice, Grand total: 12TB
- Data throughput: 24TB per sample (dual view)







HBP data repository: EBRAINS



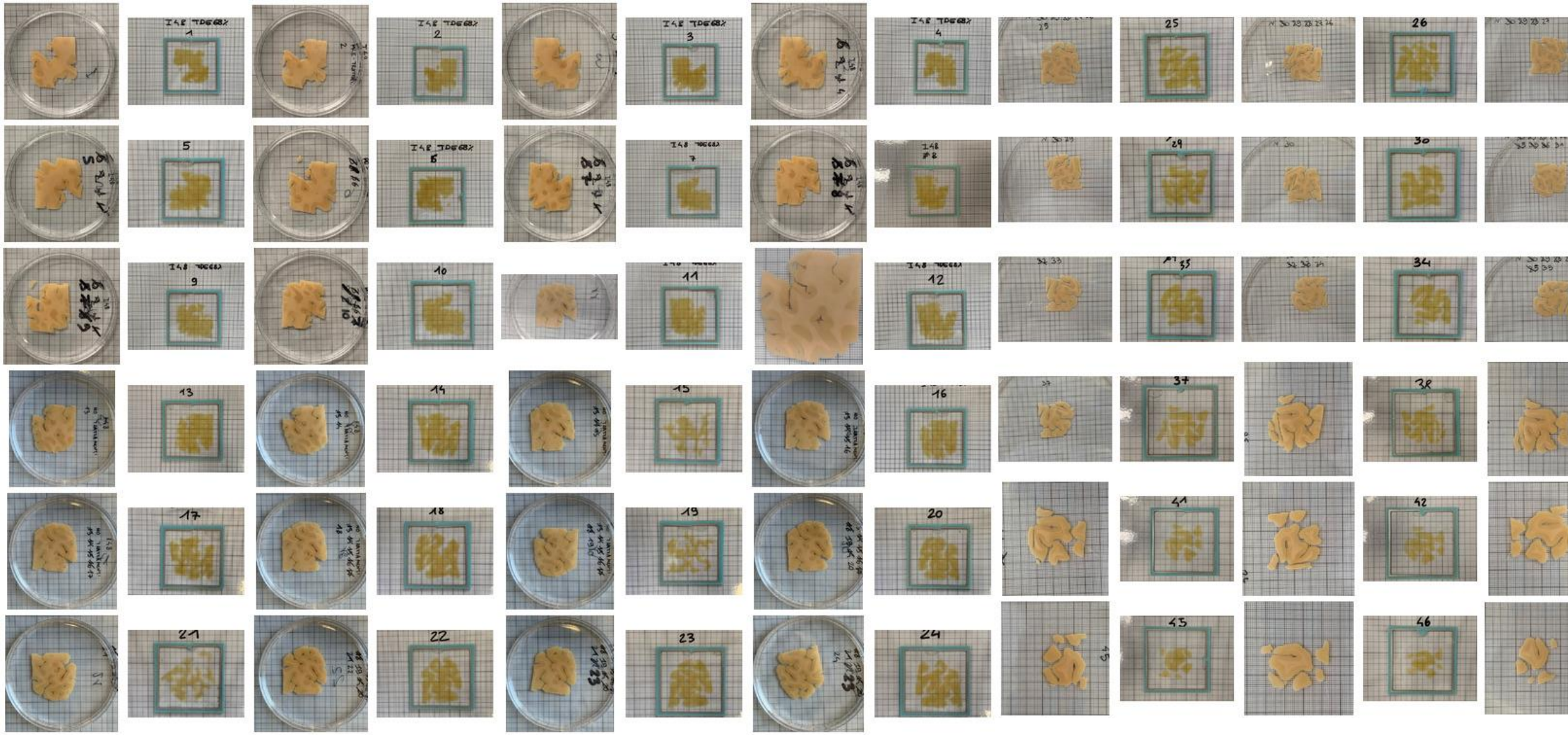
www.ebrains.eu

"EBRAINS is a new digital research infrastructure, created by the EU-funded Human Brain Project, that gathers an extensive range of data and tools for brain-related research."

Lower-resolution dataset: (4.4 x 4.4 x 3.3) μm^3
Total size: 116 GiB

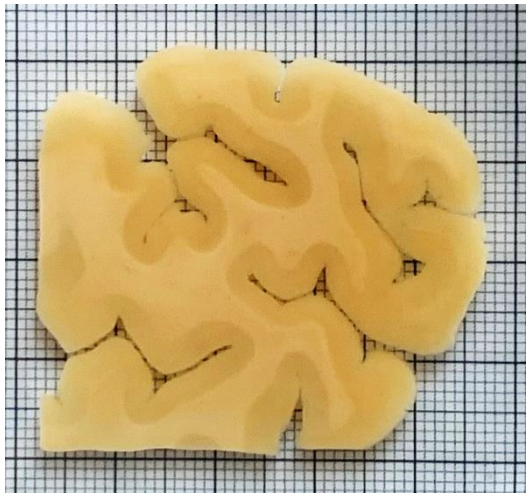
DOI: https://doi.org/10.25493/MD0E-BW5

A bigger challenge: human Broca's area (4x4x2cm^3, 48 slices)

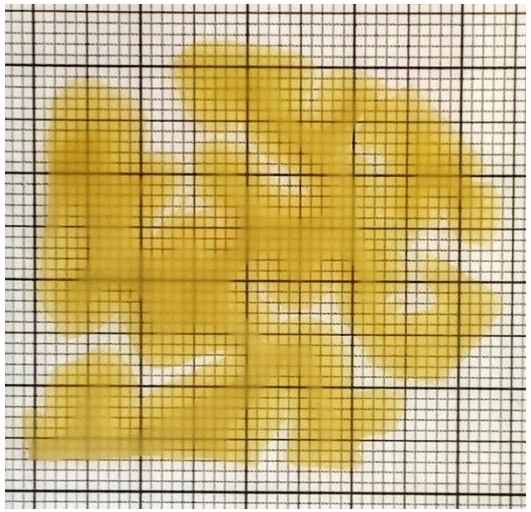


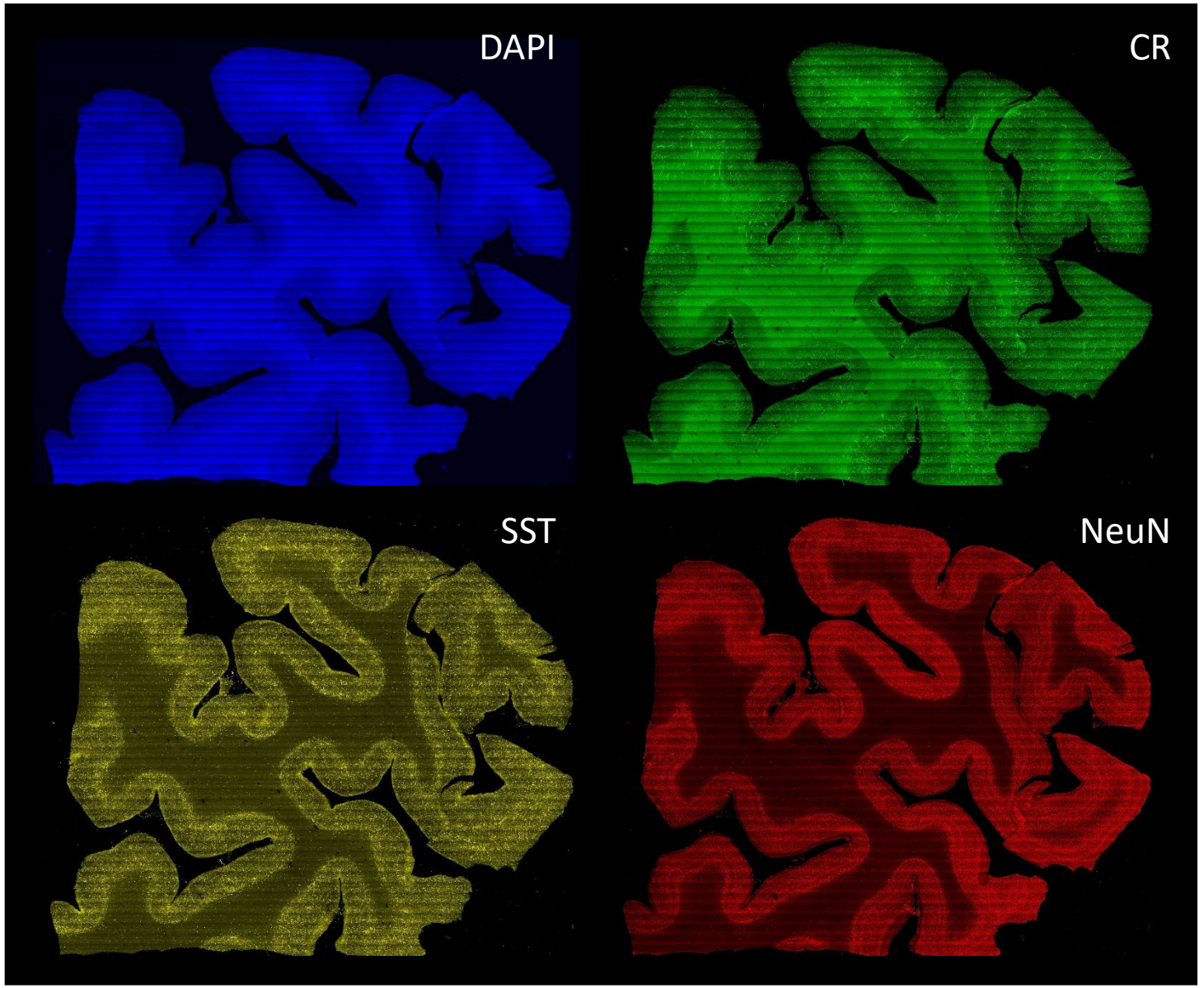
A bigger challenge: human Broca's area (4x4x2cm^3, 48 slices)

Fixed tissue



Cleared tissue





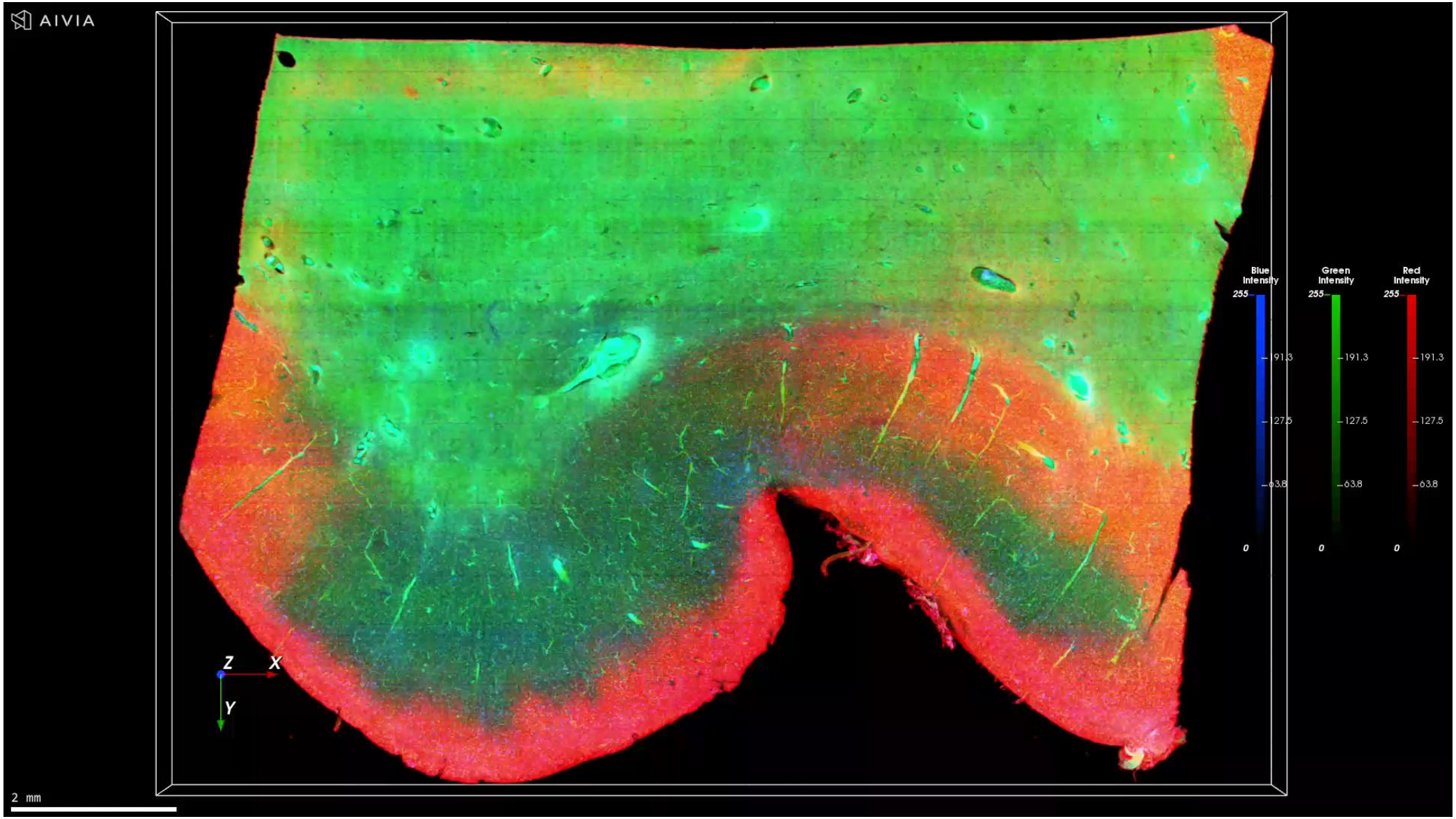
We successfully applied the SHORT clearing method to a whole Broca's Area sliced in 48 slabs of 400µm of thickness.

Brain Initiative Cell Census Network (BICCN)

Supported by The General Hospital Corporation Center of the National Institutes of Health, award number 1U01MH117023-01

NIH

3D slice


AIVIA
Blue Intensity
255
191.3
127.5
63.8
0
Green Intensity
255
191.3
127.5
63.8
0
Red Intensity
255
191.3
127.5
63.8
0
Z
X
Y
2 mm

Some numbers

August 2021

4 channels, 48 slices

Number of files	8602
Raw data	507 TiB
Lossy Compression, JPEG2000, original resolution (0.5x0.5x3.5µm)	25 TiB
Resliced (lower res, 3.5µm isotropic)	10.4 TiB
Resliced (lossless compression, 3.5µm, delivered)	4.2 TiB

BRAIN initiative data repository: DANDI

Human brain cell census for BA 44/45

ID: 000026 DRAFT | Contact **Morgan, Leah** | File Count **18530** | File Size **6.2 TB**

Created **June 24, 2020** | Last update **October 19, 2021**

Mazzamuto, Giacomo; Costantini, Irene; Gavryusev, Vladislav; Castelli, Filippo Maria; Pesce, Luca; Scardigli, Marina; Pavone, Francesco Saverio; Roffilli, Matteo; Silvestri, Ludovico; Hof, Patrick R.; Boas, David A.; Fischl, Bruce; Morgan, Leah; Yang, Jiarui; Chang, Shuaibin; Laffey, Jessie; Magnain, Caroline; Varadarajan, Divya; Wang, Hui; Frost, Robert; Kouwe, Andre van der; Player, Allison Stevens; Atzeni, Alessia; Gonzalez, Juan Eugenio Iglesias; Balbastre, Yael; Vera, Matthew; Cordero, Devani; Nestor, Kimberly; Ammon, William; Nolan, Jackson; Mora, Jocelyn; Pallares, Erendira Garcia; Augustinack, Jean; Diamond, Bram; Fogarty, Morgan; Boyd, Emma; Varghese, Merina; Dalca, Adrian V.; Edlow, Brian; Frosche, Matthew; Chen, I-Chun Anderson; Wicinski, Bridget

Magnetic resonance imaging (MRI) is used to establish a macroscopic reference coordinate system of laminar and cytoarchitectural boundaries. Cell counting is obtained with both traditional immunohistochemistry, to provide a stereological gold standard, and with a custom-made inverted confocal light-sheet fluorescence microscope (LSM) for 3D imaging at cellular resolution. Finally, polarization-sensitive optical coherence tomography (PSOCT) enables registration of the distorted histological cell typing obtained with LSM to the MRI-based atlas coordinate system. [- see less]

Keywords: multi-modal imaging | MRI | OCT | SPIM | human cortex | Broca's area | Motor cortex | Stereology

Licenses: spdx:CC-BY-4.0

Subject matter of the dataset

- Broca's Area
- Motor Cortex

Access information

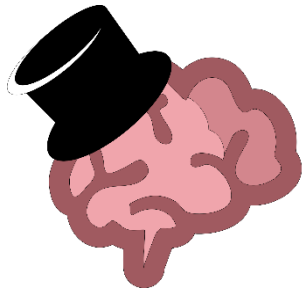
- dandi:OpenAccess

Assets Summary

DANDI: **D**istributed **A**rchives for **N**europhysiology **D**ata **I**ntegration

www.dandiarchive.org

"The BRAIN Initiative archive for publishing and sharing neurophysiology data including electrophysiology, optophysiology, and behavioral time-series, and images from immunostaining experiments."

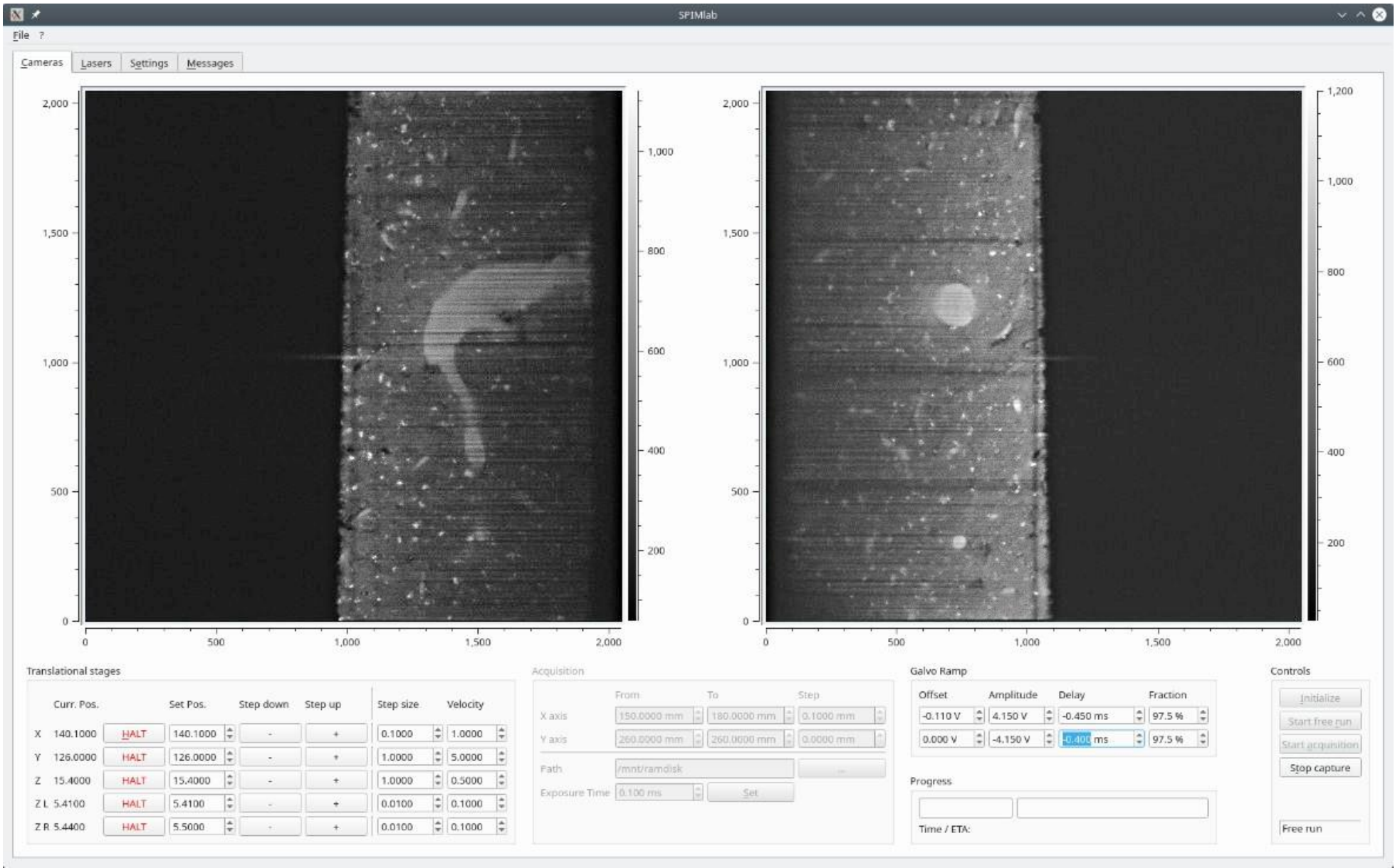


DANDI

```
pip-install dandi
```

BIDS extension proposal 31 (BEP31)
(The Brain Imaging Data Structure)

Acquisition and control software (SPIMlab + QtLab)



- developed in C++ using Qt
- 7000 SLOC (Single Lines of Code)
- multi threaded
- data rate 1 GB/s with two cameras
- flexible and modular architecture

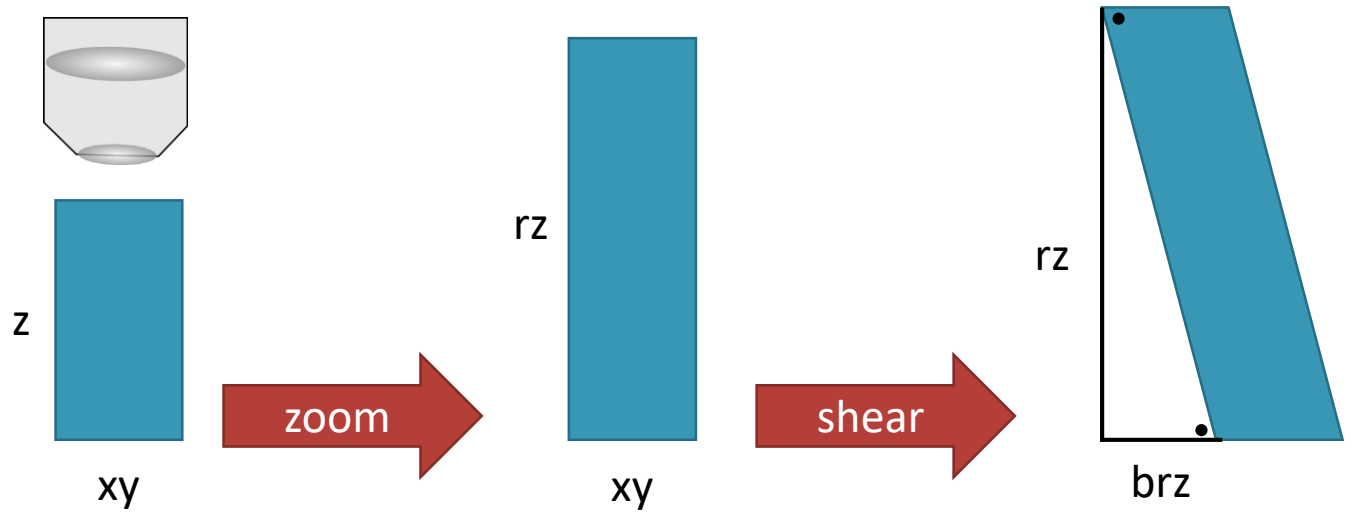
GitHub

C++

Qt

Image reslicing

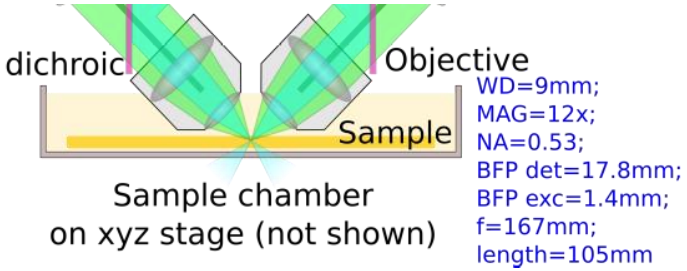
Objective reference frame



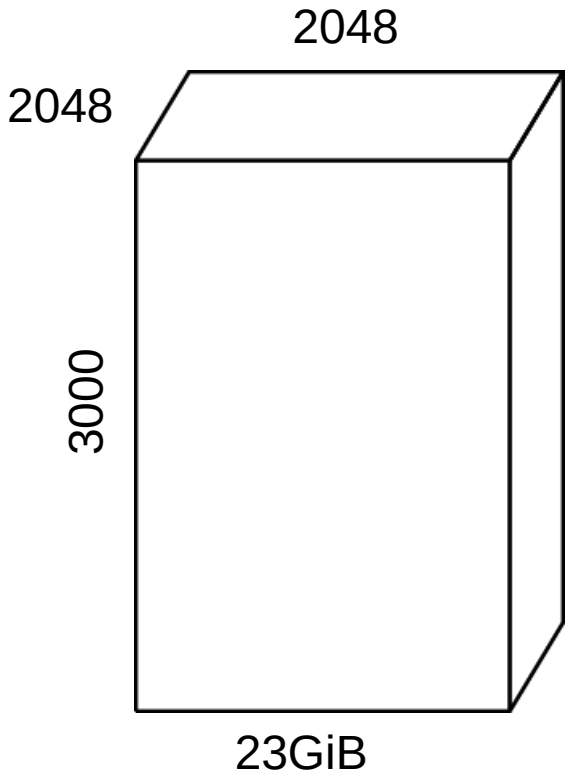
Affine transformation using zooms, rotations and shears to go from the objective's reference frame to the lab's reference frame.

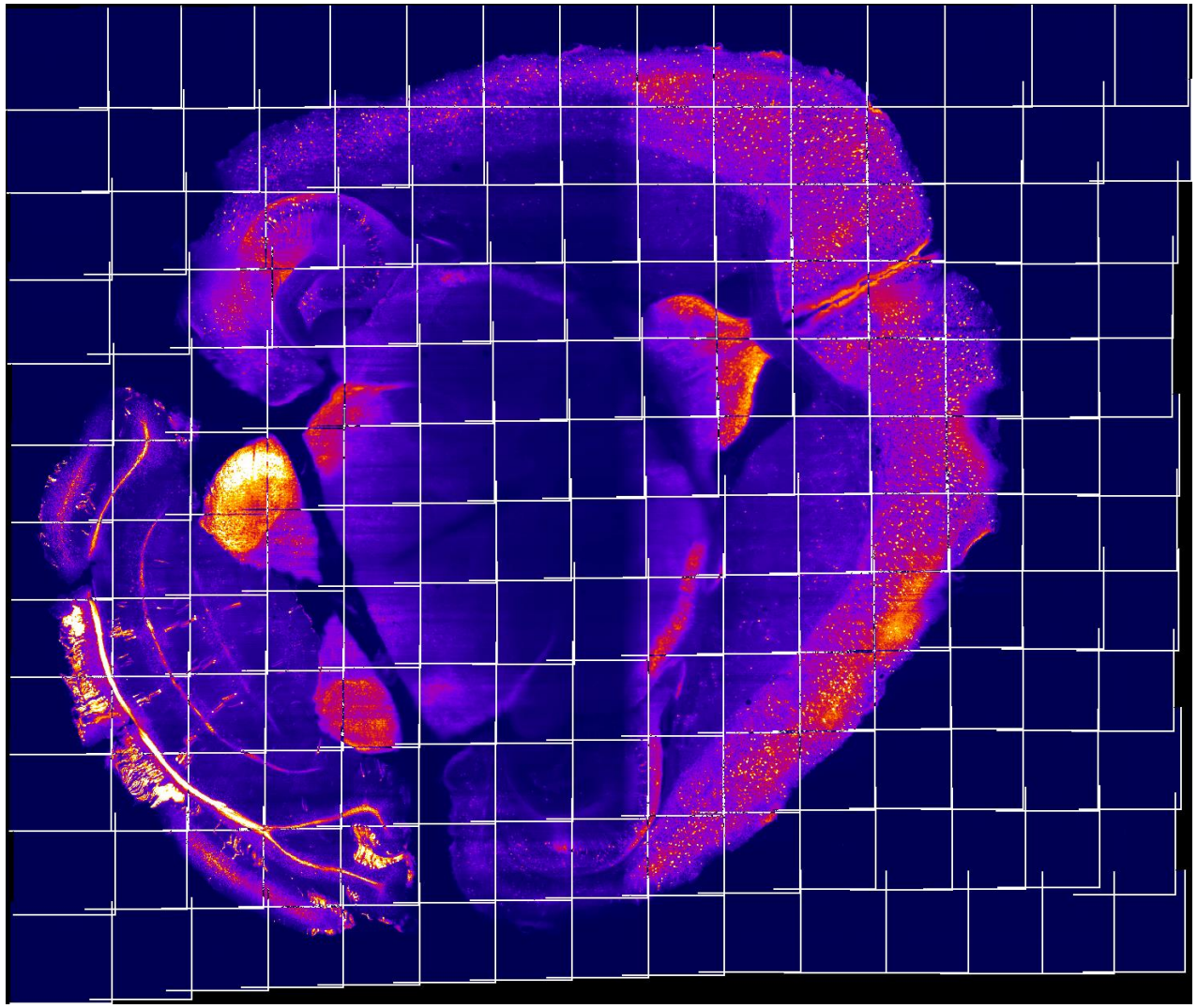
$$A = T\ R\ S\ Z$$

Lab reference frame



Volumetric stitching





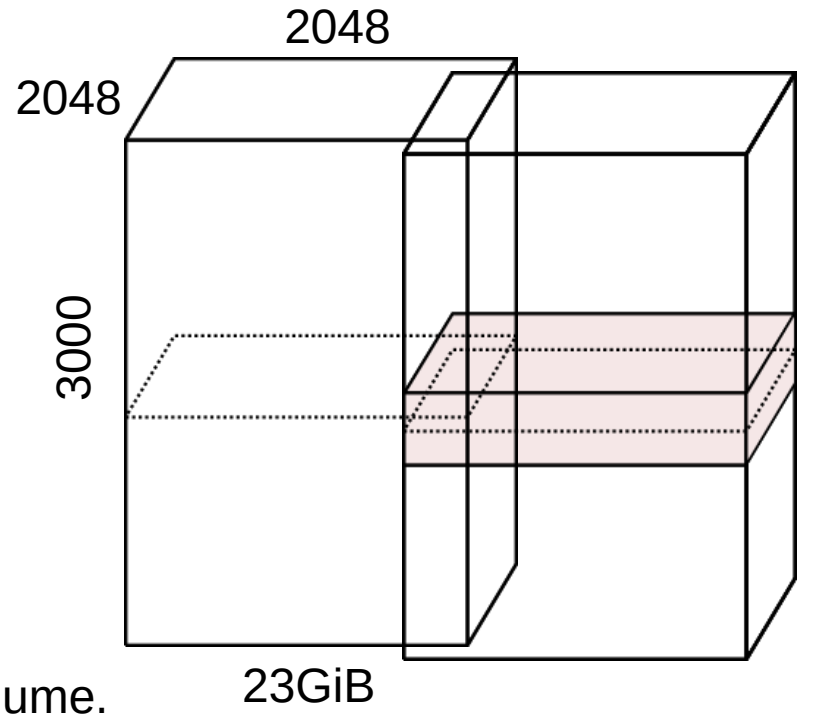
- Sample size (half mouse brain): 0.6 x 1.2 x 1.2 cm^3 (ZYX)
- Pixel size: 2 x 0.65 x 0.65 μm^3 (ZYX), 16 bit
- Mosaic: 15 x 12 stacks, 23 GiB each
- Whole dataset: 4.2 TiB

Volumetric stitching: ZetaStitcher

ZetaStitcher

GitHub

https://github.com/lens-biophotonics/ZetaStitcher



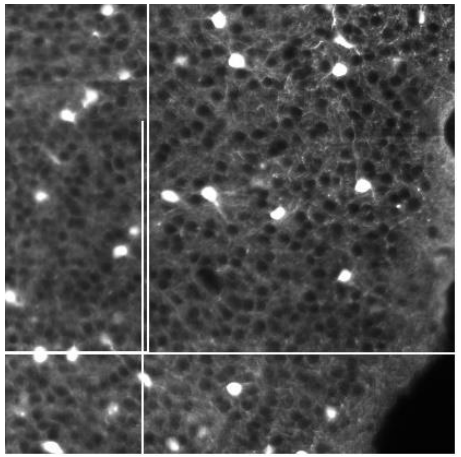
- Developed entirely in Python
- Open source (GPLv3)
- High throughput

python™

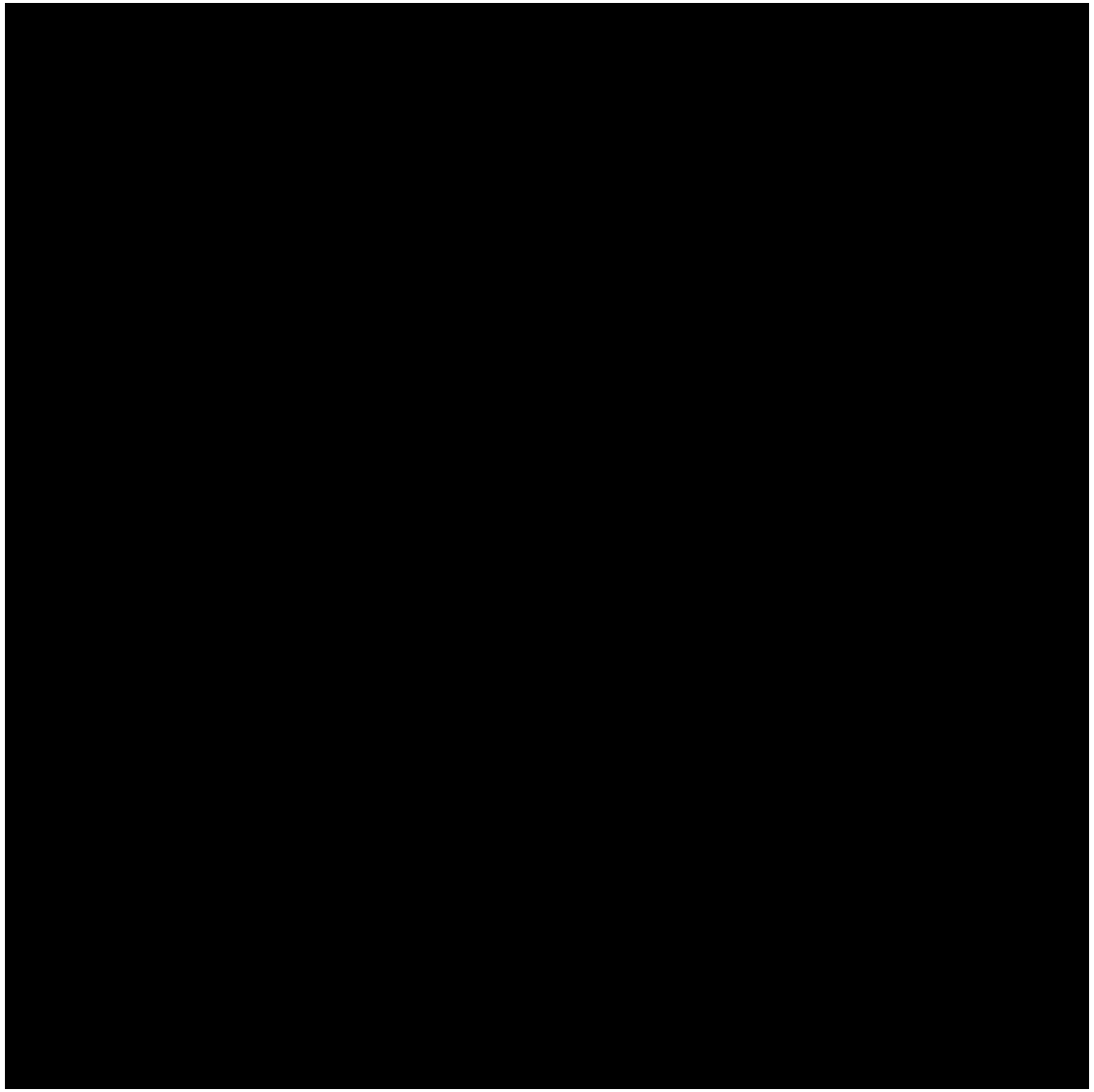
A powerful and simple Python API to query arbitrary regions within the fused volume.

```
>>> from zetastitcher import VirtualFusedVolume
>>> vfv = VirtualFusedVolume('stitch.yml')
>>> vfv.shape
(2985, 18924, 23486)

a = vfv[2000:2500, 12000:13000, 15500:16500]
```



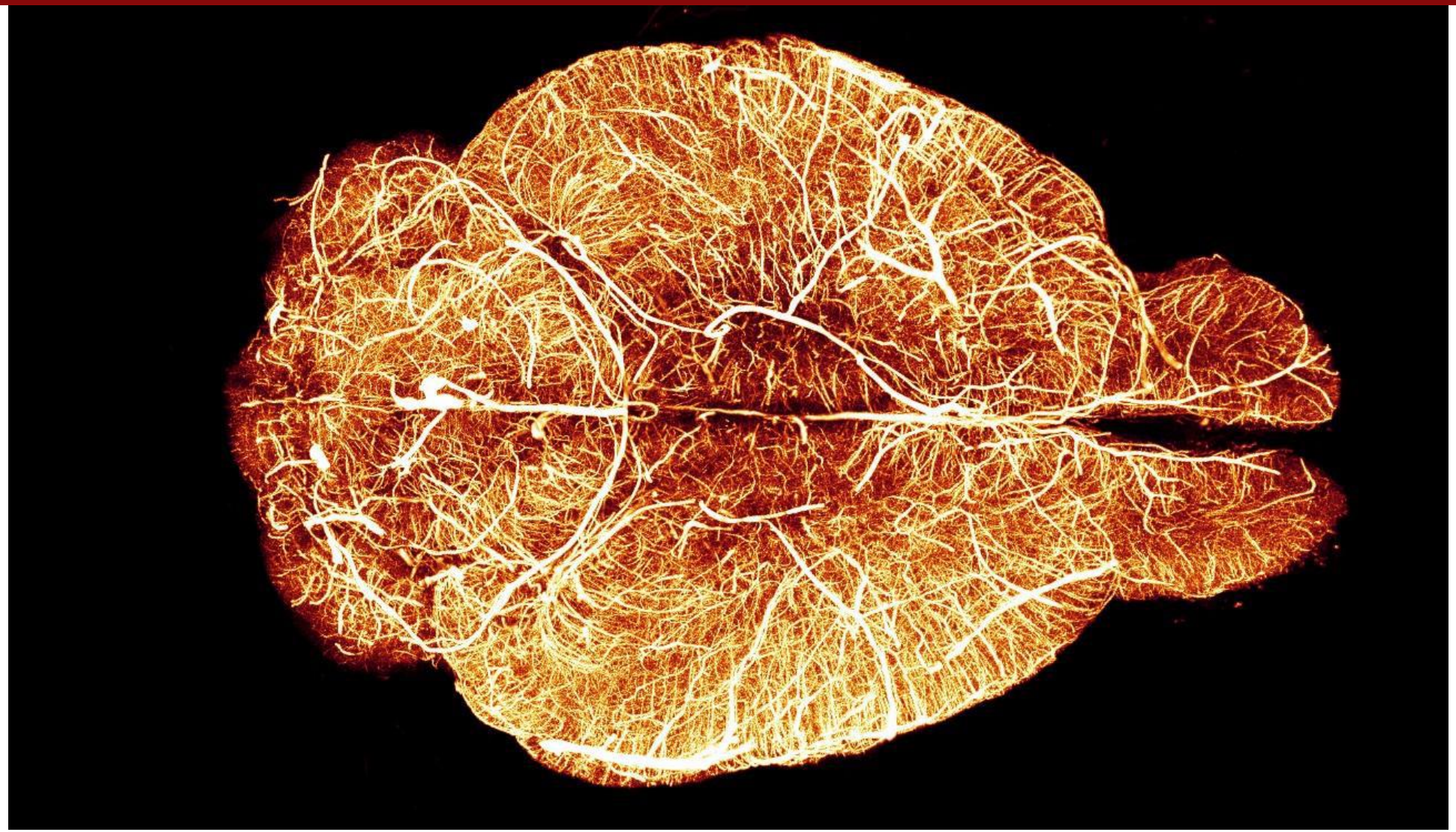
Fused volume (half mouse brain)



Downscaled volume:
px size: 10 x 10.4 x 10.4 μm^3 (ZYX)

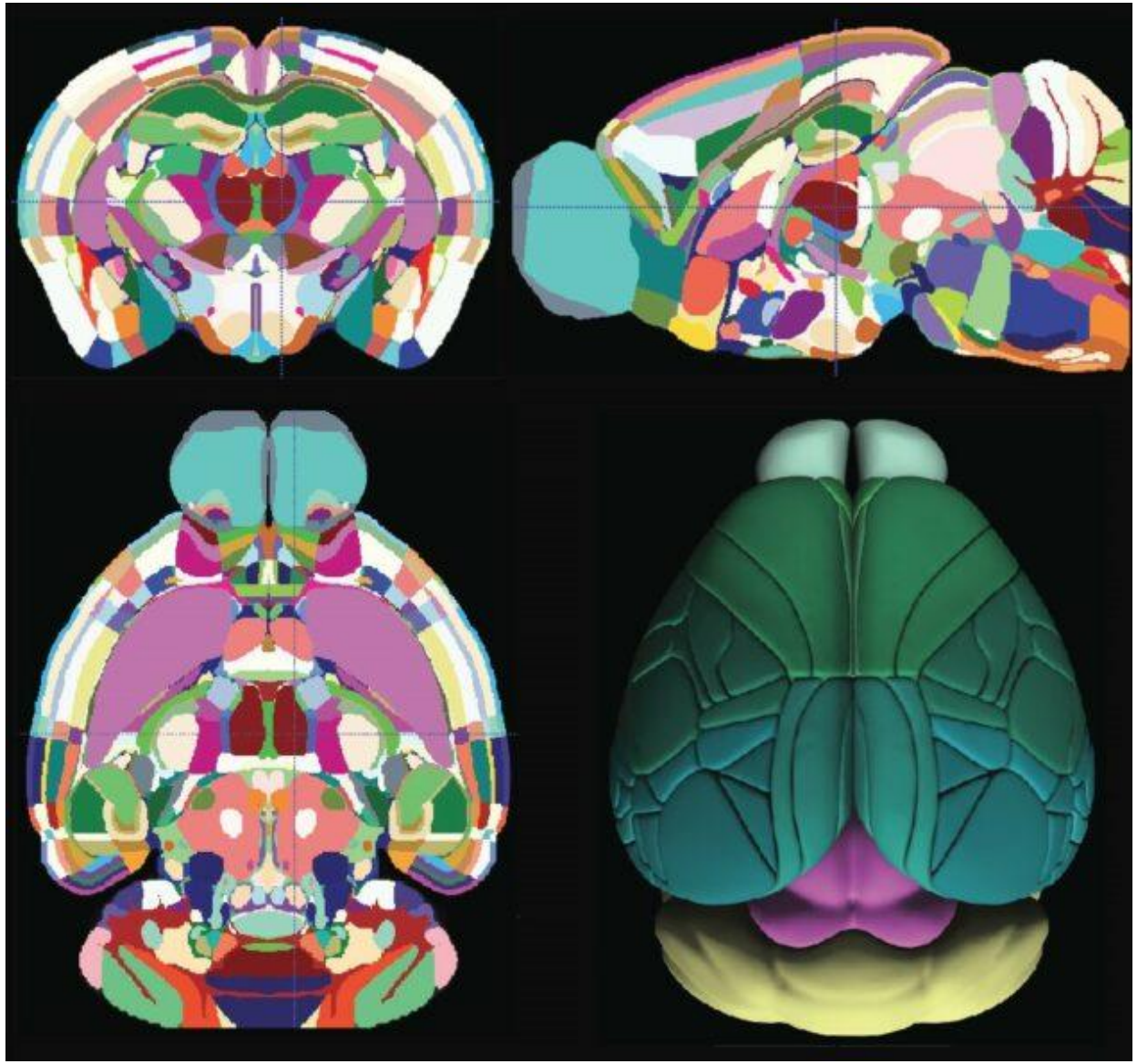
Original volume:
2985 × 18924 × 23486 px (Z × Y × X) ≈ 1.3 10^{12} voxels

Whole mouse brain tomography with LSFM: vasculature



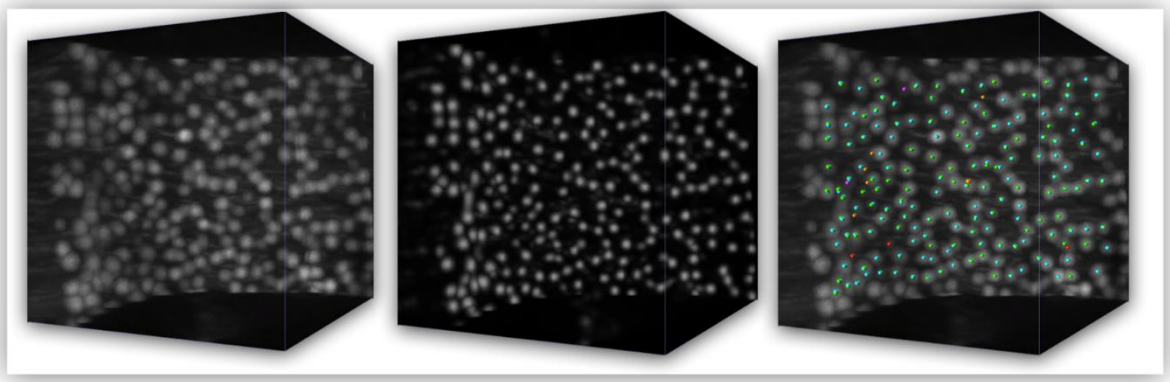
Di Giovanna, Antonino Paolo, et al. "Whole-brain vasculature reconstruction at the single capillary level." Scientific reports 8.1 (2018): 1-11.

Alignment to the Allen Mouse Brain Reference Atlas



Atlas warping using ANTs on downsampled version of the dataset (Avants et al. IEEE transactions on medical imaging, http://stnava.github.io/)

Neuron localization in the whole mouse brain

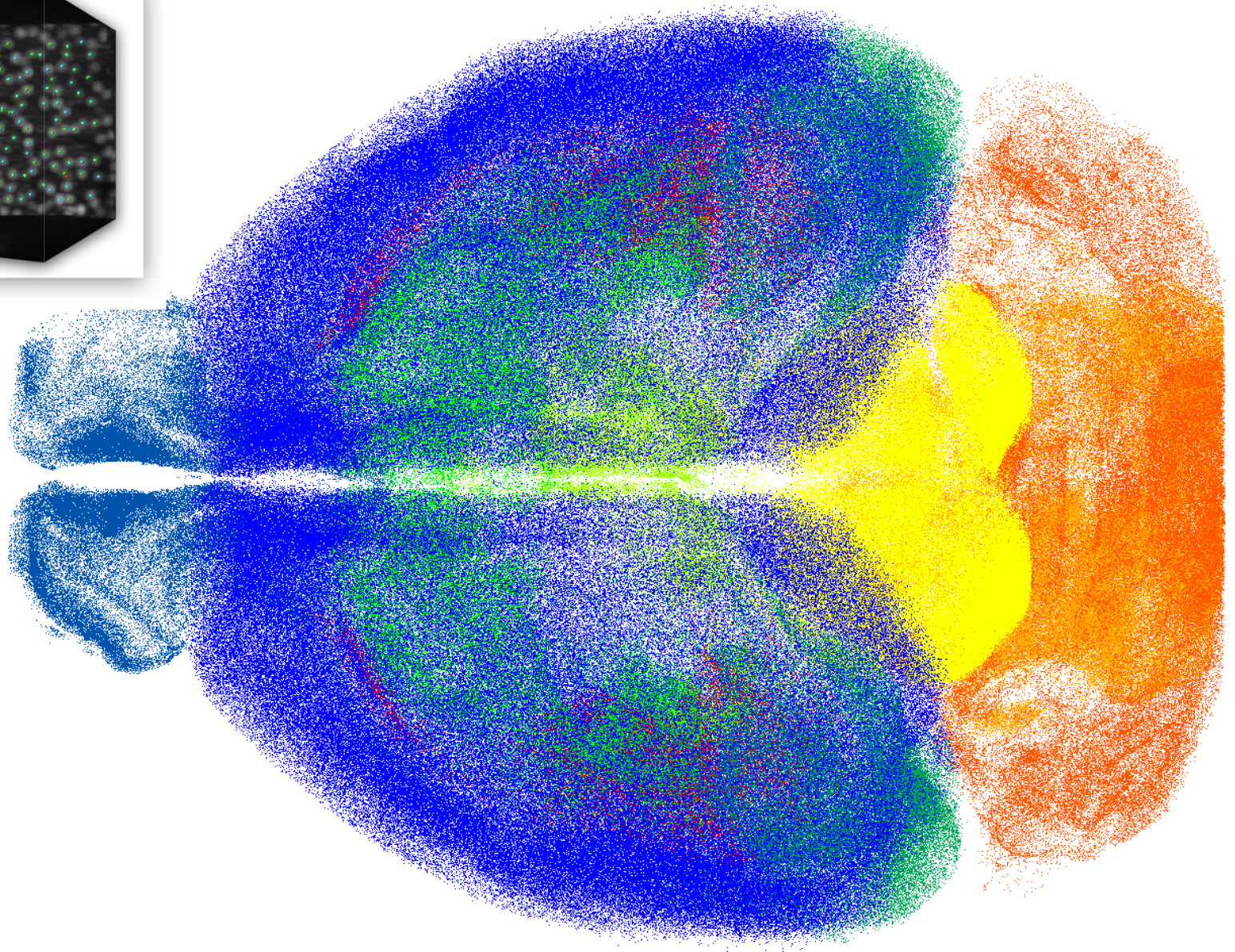


Semantic deconvolution

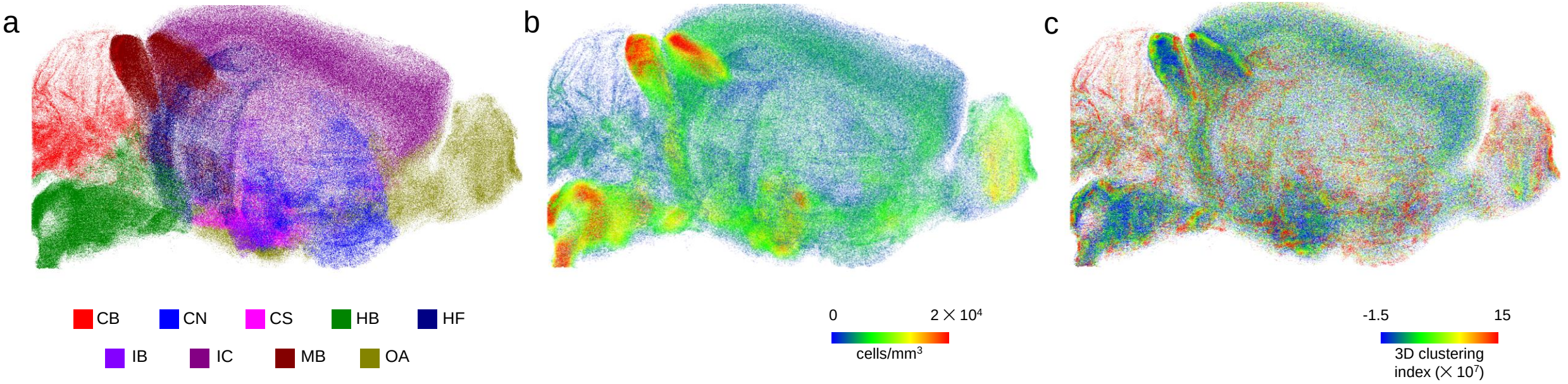
SST-positive cells

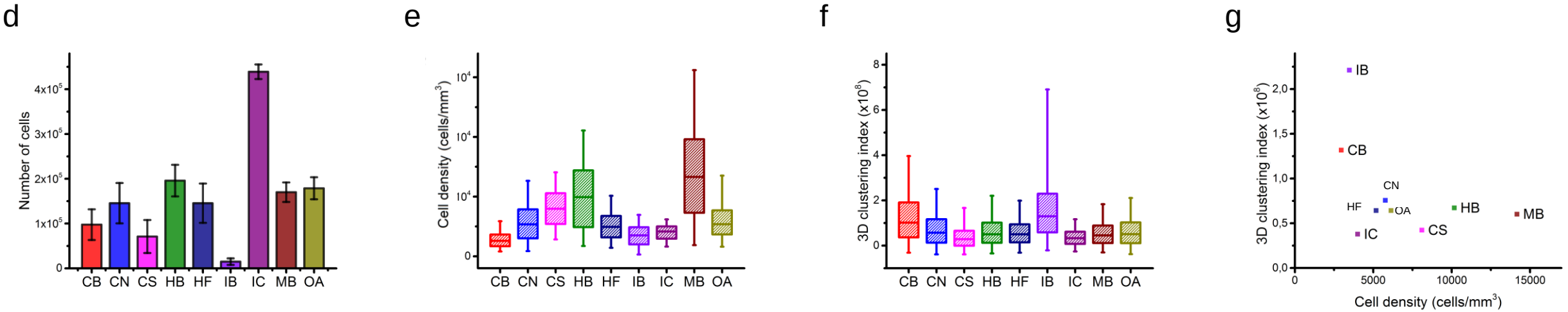
About 1.5×10^6 neurons

University of Florence, Department of Information Engineering (DINFO)
Paolo Frasconi



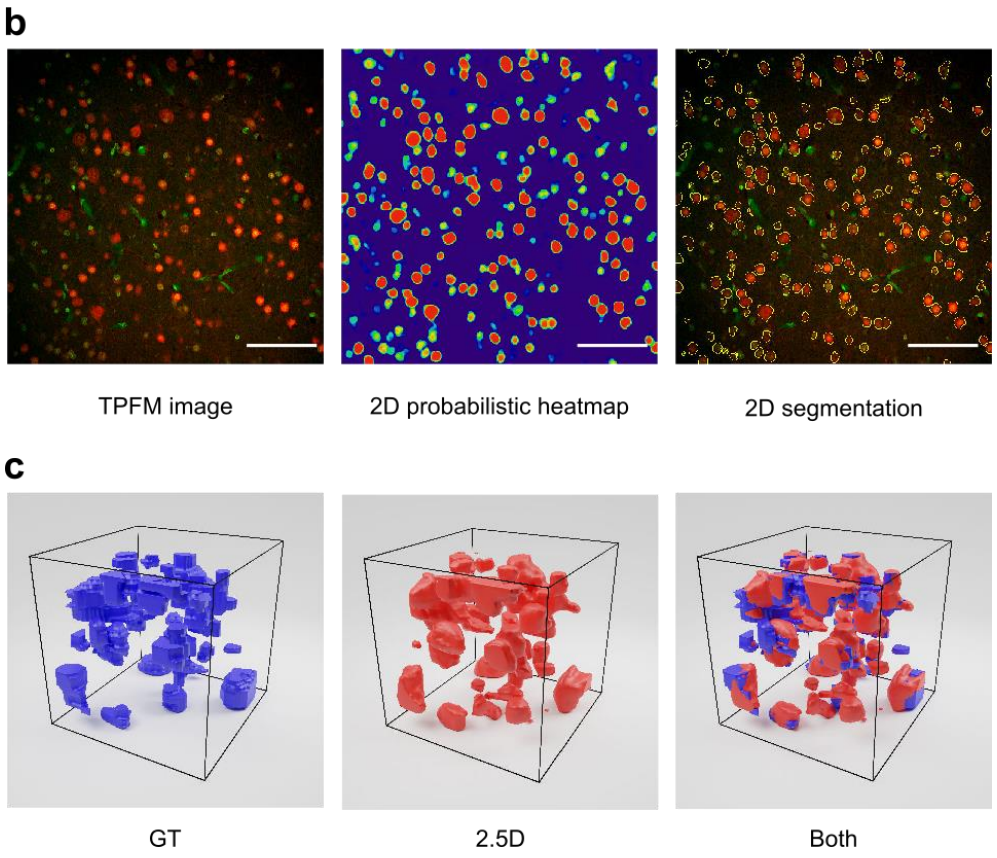
Quantitative data analysis

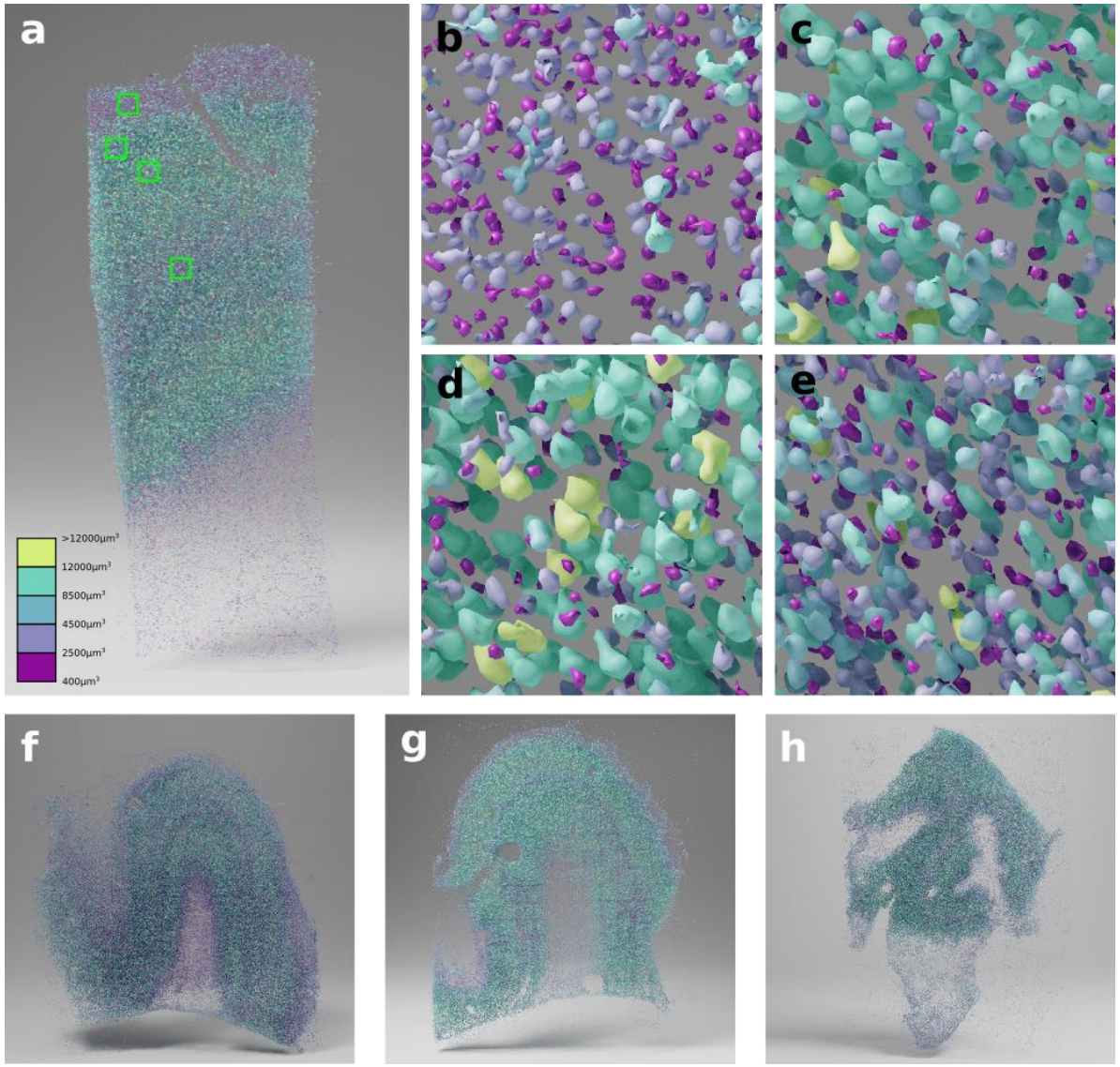




Silvestri, L. *et al. Nat Methods* **18,** 953–958 (2021).

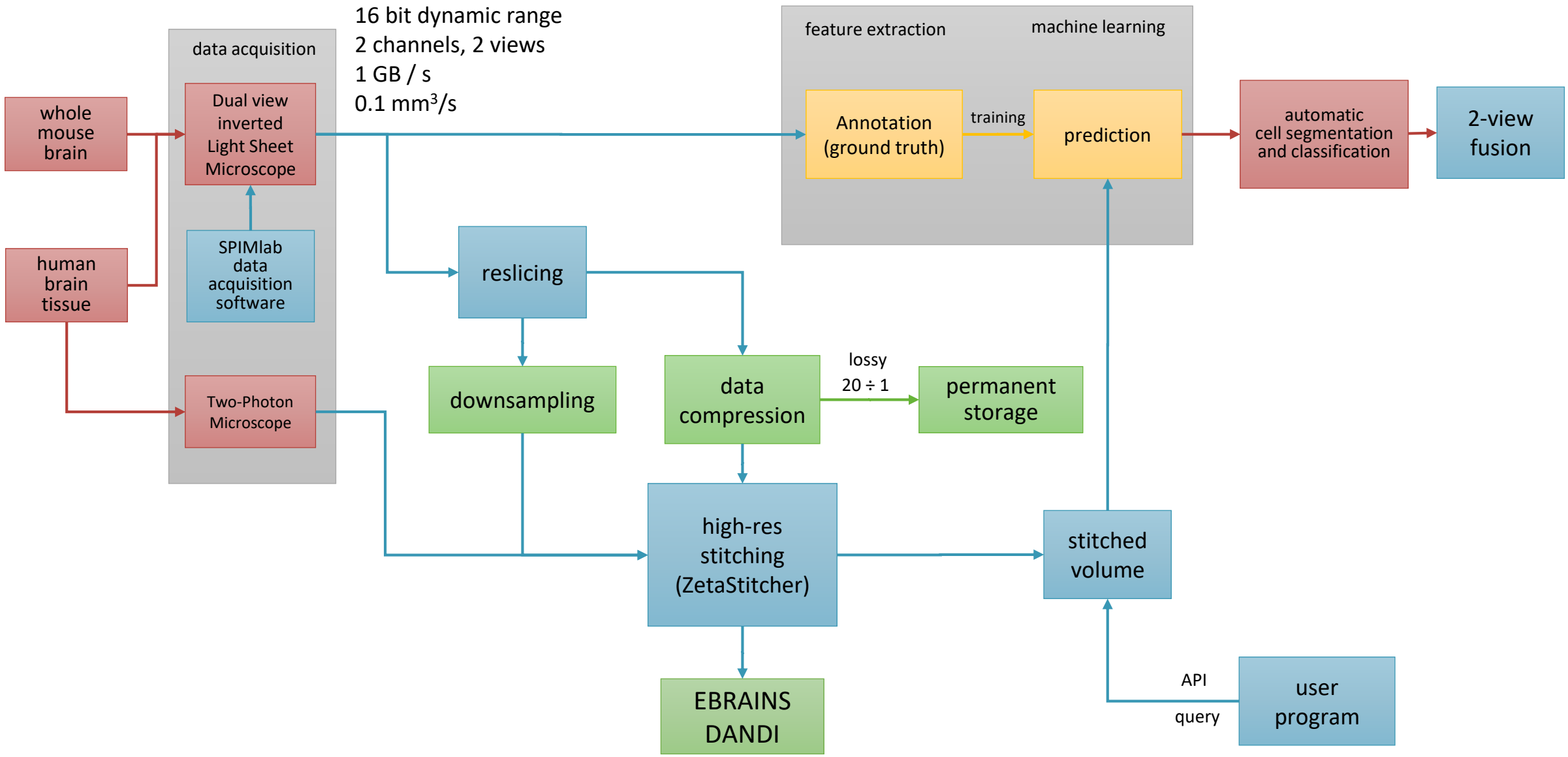
Digital histology





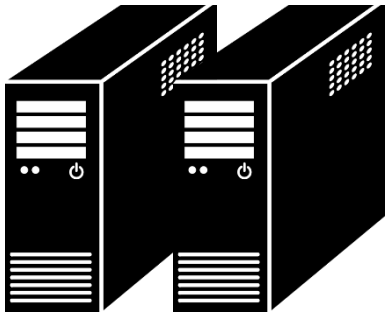
I. Costantini, G. Mazzamuto et al. **Biomedical Optics Express,** Vol. 12, Issue 6, pp. 3684-3699 (2021)

Data management workflow



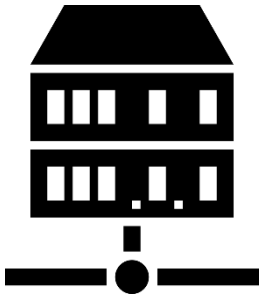
Computing cluster

1 GPU server with:
1x GeForce 3090 24GB
256GB RAM
AMD Ryzen Threadripper 32 core



2 GPU servers each with:
4x GeForce 2080Ti 12GB
192GB RAM
Intel Xeon 24 core

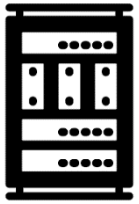
1 GPU server with:
1x Pascal P100 12GB
128GB RAM
Intel Xeon 16 core



NAS storage
500 TB

Central manager node
- HTCondor
- Docker registry
- LDAP

10 Gbit/s



CINECA (Bologna)

Internet

1 Gbit/s

Acknowledgements

European Laboratory for Non-linear Spectroscopy (LENS)
Ludovico Silvestri, Irene Costantini, Vladislav Gavryusev, Giuseppe Sancataldo, Marina Scardigli, Luca Pesce, Filippo Maria Castelli, Francesco S. Pavone

University of Florence, Department of Information Engineering (DINFO)
Francesco Orsini, Paolo Frasconi

Bioretics Srl, Cesena
Mattia Neri, Matteo Roffilli


Thank you!
www.ino.it
bio.lens.unifi.it
github.com/lens-biophotonics