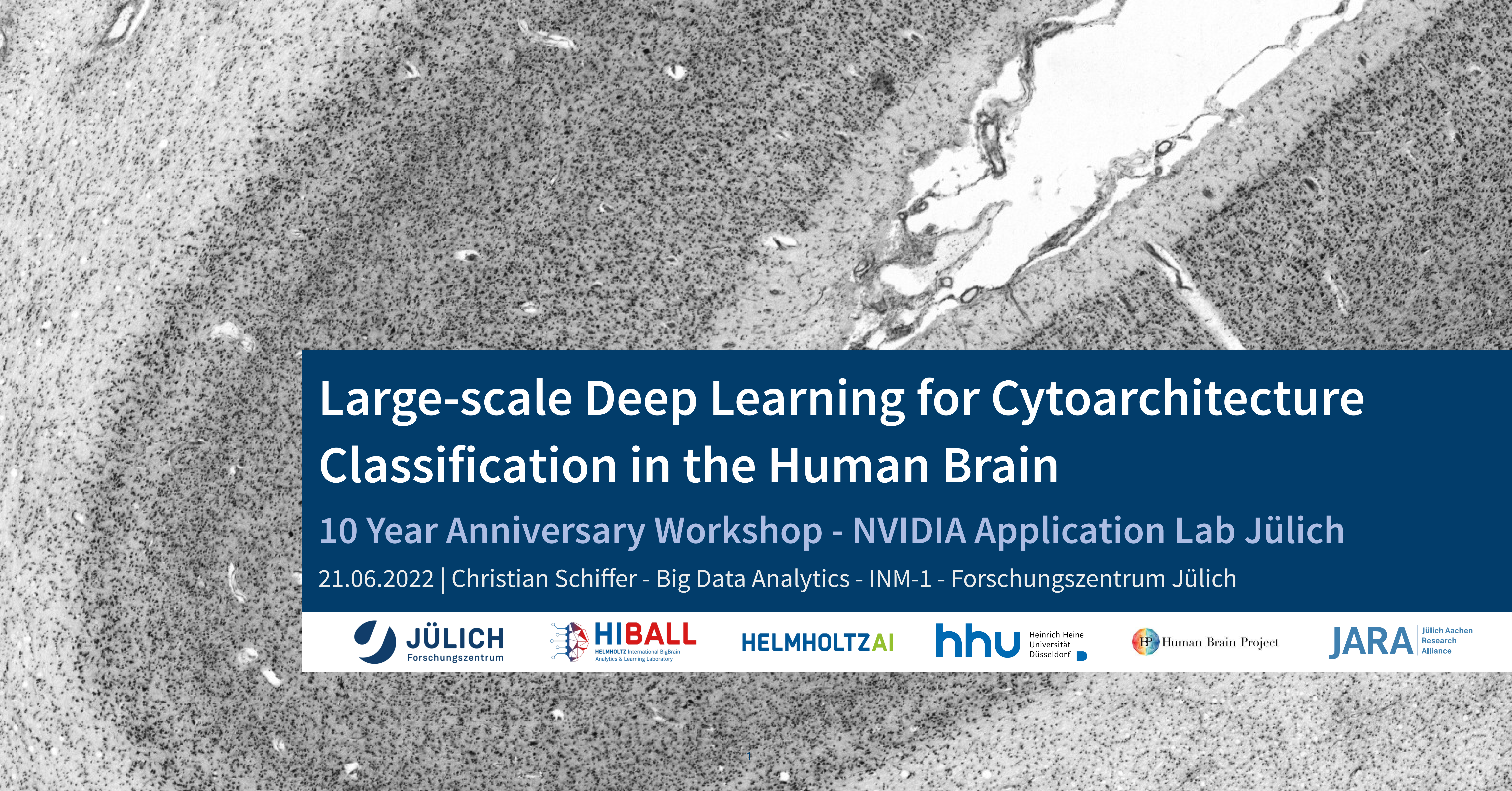


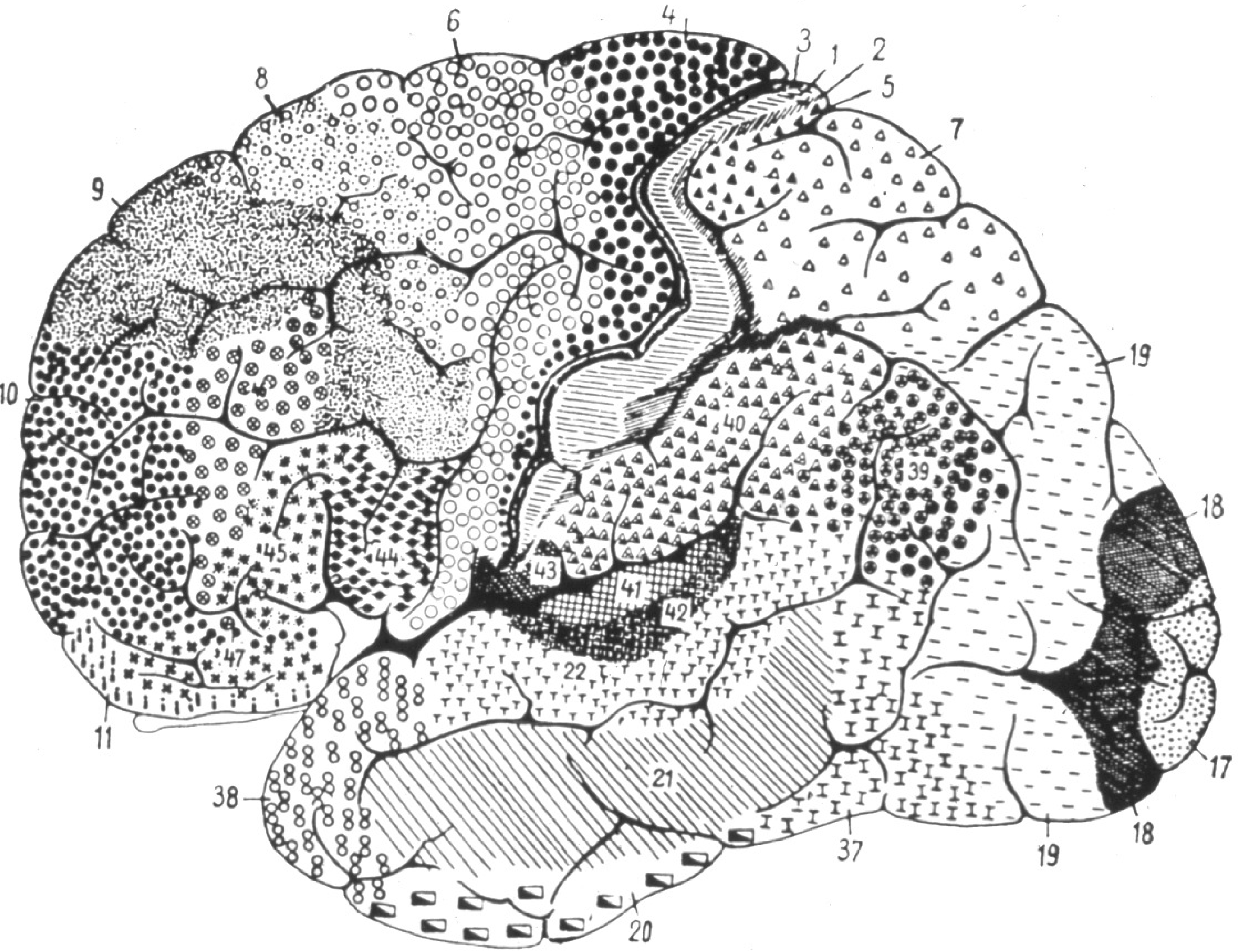
Large-scale Deep Learning for Cytoarchitecture Classification in the Human Brain

10 Year Anniversary Workshop - NVIDIA Application Lab Jülich

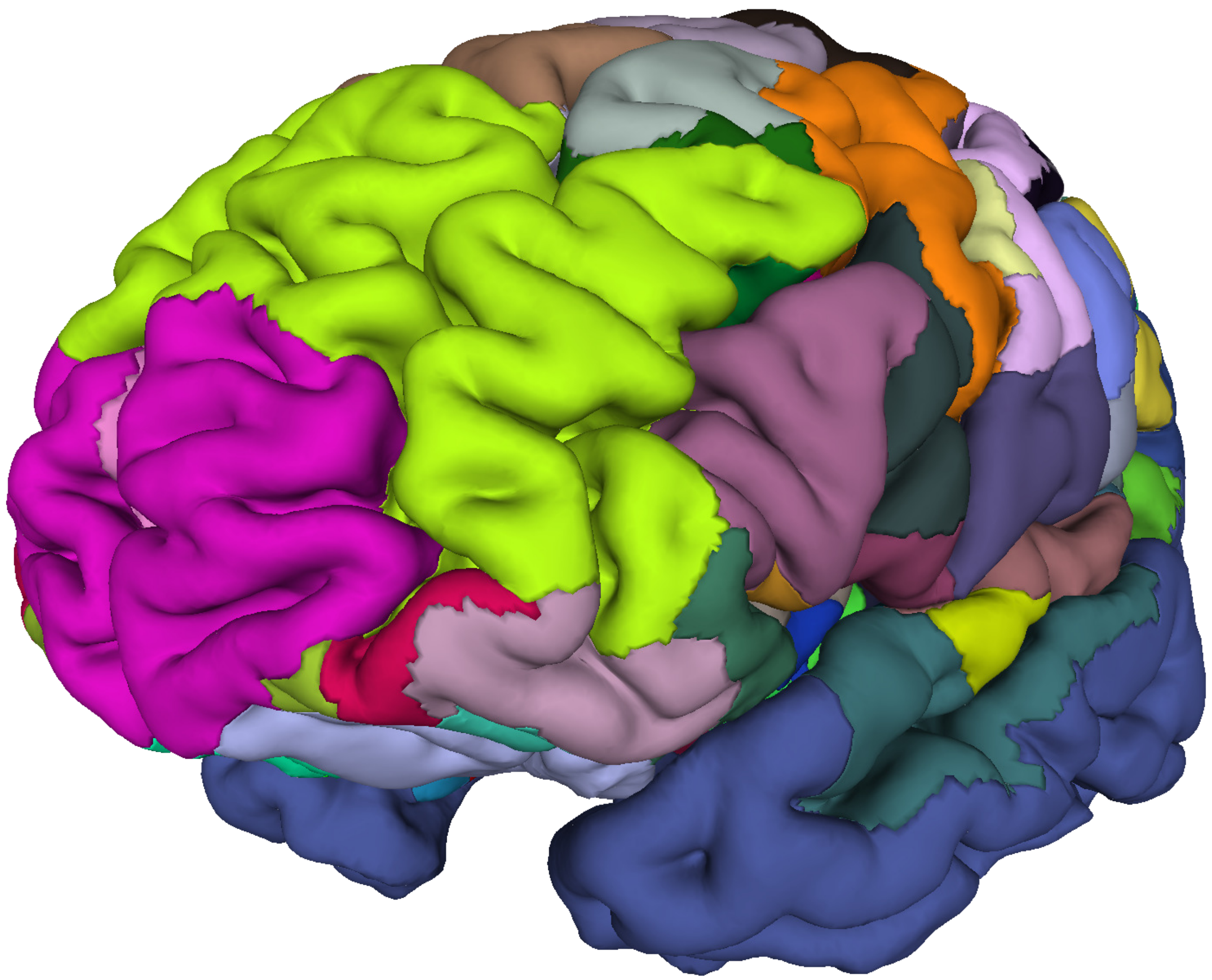
21.06.2022 | Christian Schiffer - Big Data Analytics - INM-1 - Forschungszentrum Jülich

JÜLICH Forschungszentrum | HIBALL HELMHOLTZ International BigBrain Analytics & Learning Laboratory | HELMHOLTZAI | hhu Heinrich Heine Universität Düsseldorf | Human Brain Project | JARA Jülich Aachen Research Alliance

Building a Human Brain Atlas for Cytoarchitecture



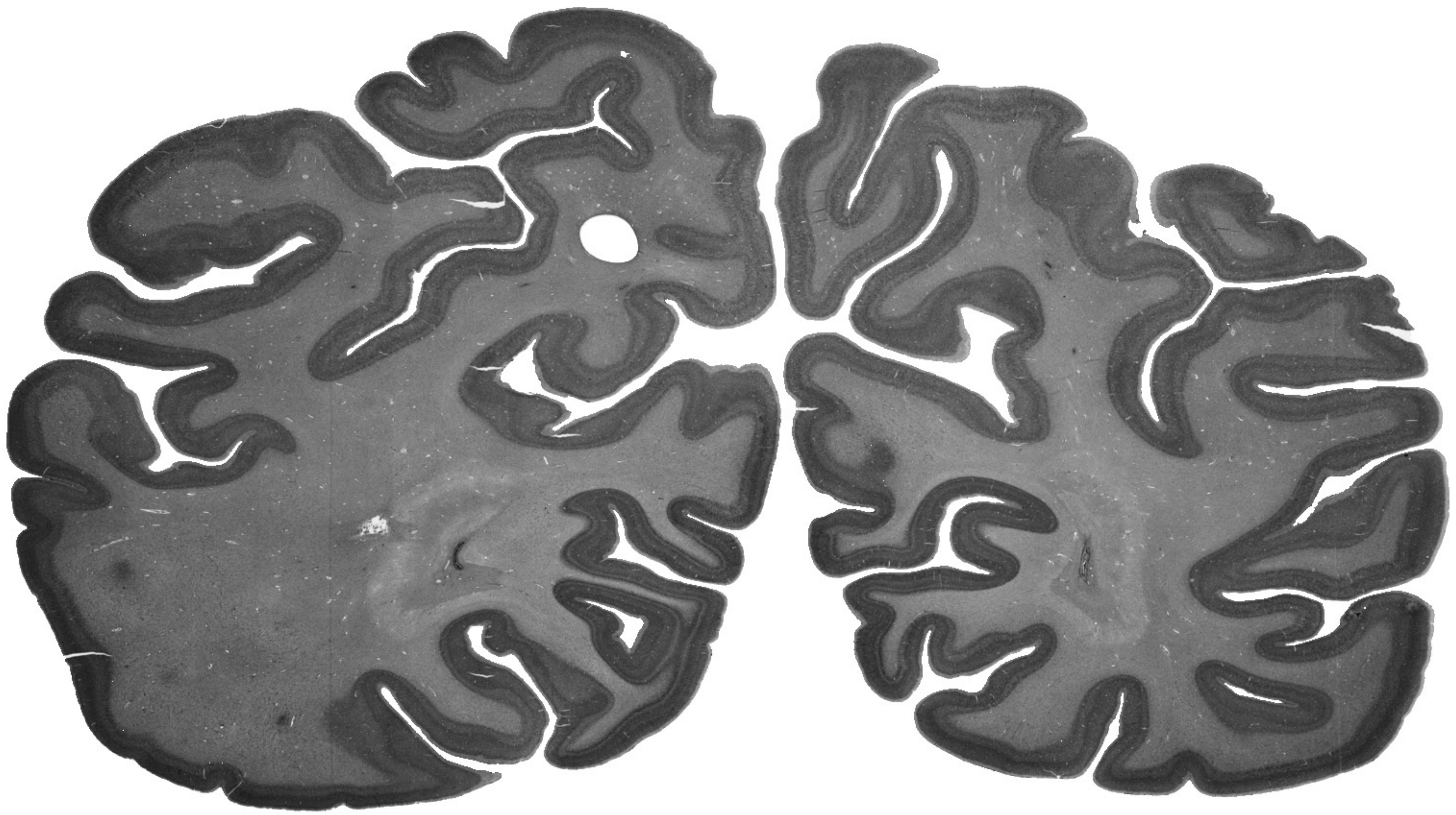
Brodmann, 1909



Amunts et al., 2020

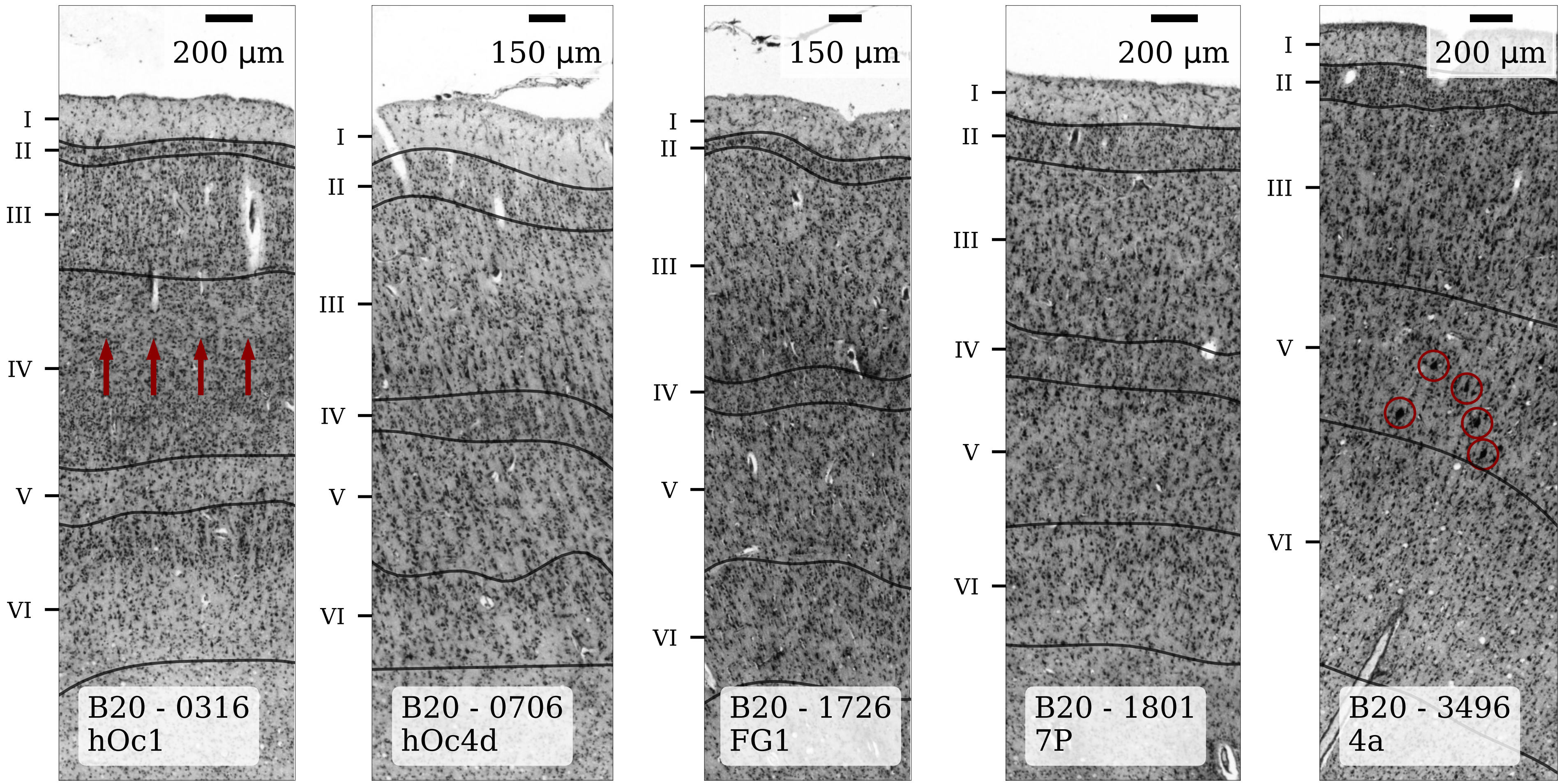
Histological Human Brain Sections

- **Postmortem** human brains
- Fixate and **cut** into histological sections
  - 6000-8000 sections per brain
  - Thickness: $20\mu m$
- Stain for **cell bodies**
- Microscopic **imaging** at $1\mu m$ pixel resolution
- **Cerebral cortex:** Outer layer of the cerebrum



Cytoarchitecture

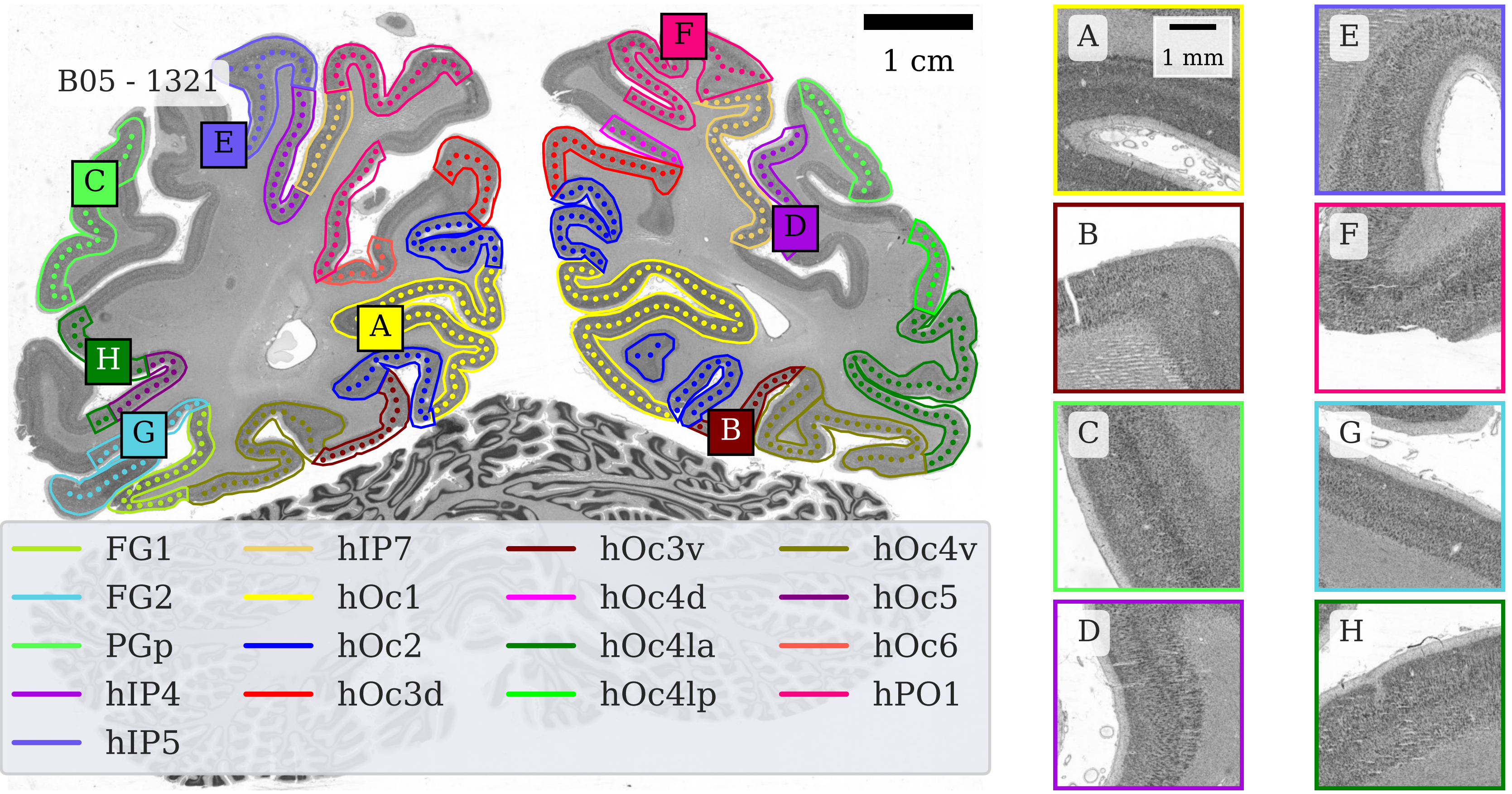
- **Cytoarchitecture:** Distribution, shape, and type of neuronal cells
- Organization into **cortical layers**
- Regional differences define **cortical areas**
- Indicators for **connectivity** and **function**



JÜLICH Forschungszentrum

Cytoarchitectonic Brain Mapping

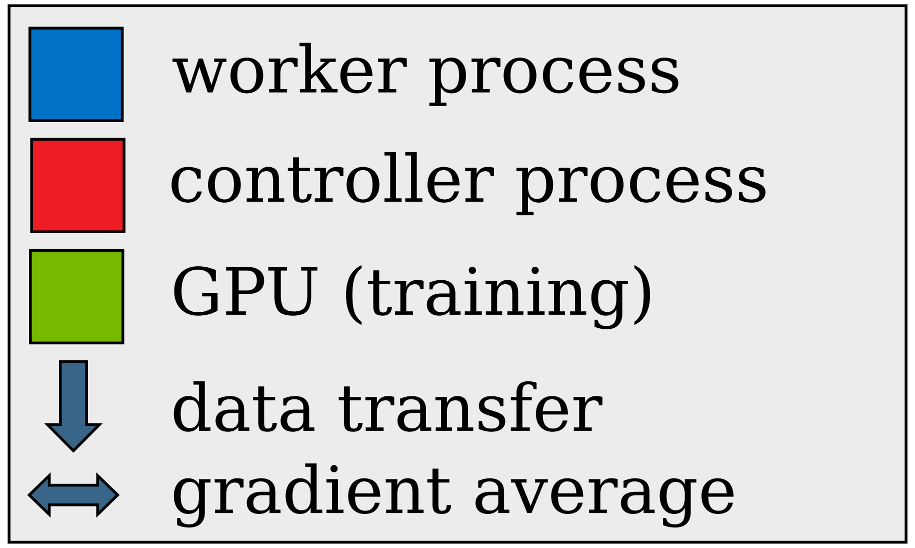
- **Brain mapping:** Identify cytoarchitectonic areas
- **Gold standard method:** Schleicher et al., 1999
  - Statistical image analysis
  - Reproducible and observer-independent
  - **Time intensive:** ≥ 30-60 min/section/area
- **Goal:** Automated cytoarchitectonic mapping to enable **large-scale cytoarchitecture analysis**
- Train **deep neural networks** to predict areas from images



Distributed Deep Learning on HPC

- **Dataset size**
  - **Large images:** ~80,000×100,000 px (> 8 GB)
  - **Many images:** 6000-8000 images per brain
  - **Large patches:** 2048×2048 px/patch (4mm^2@2µm/px)
- **Technical challenges**
  - **I/O:** Random access to patches → **flash-based storage**
  - **Preprocessing:** Augmenting large image patches → **CPUs**
  - **Training:** Data parallel deep learning → **GPUs**

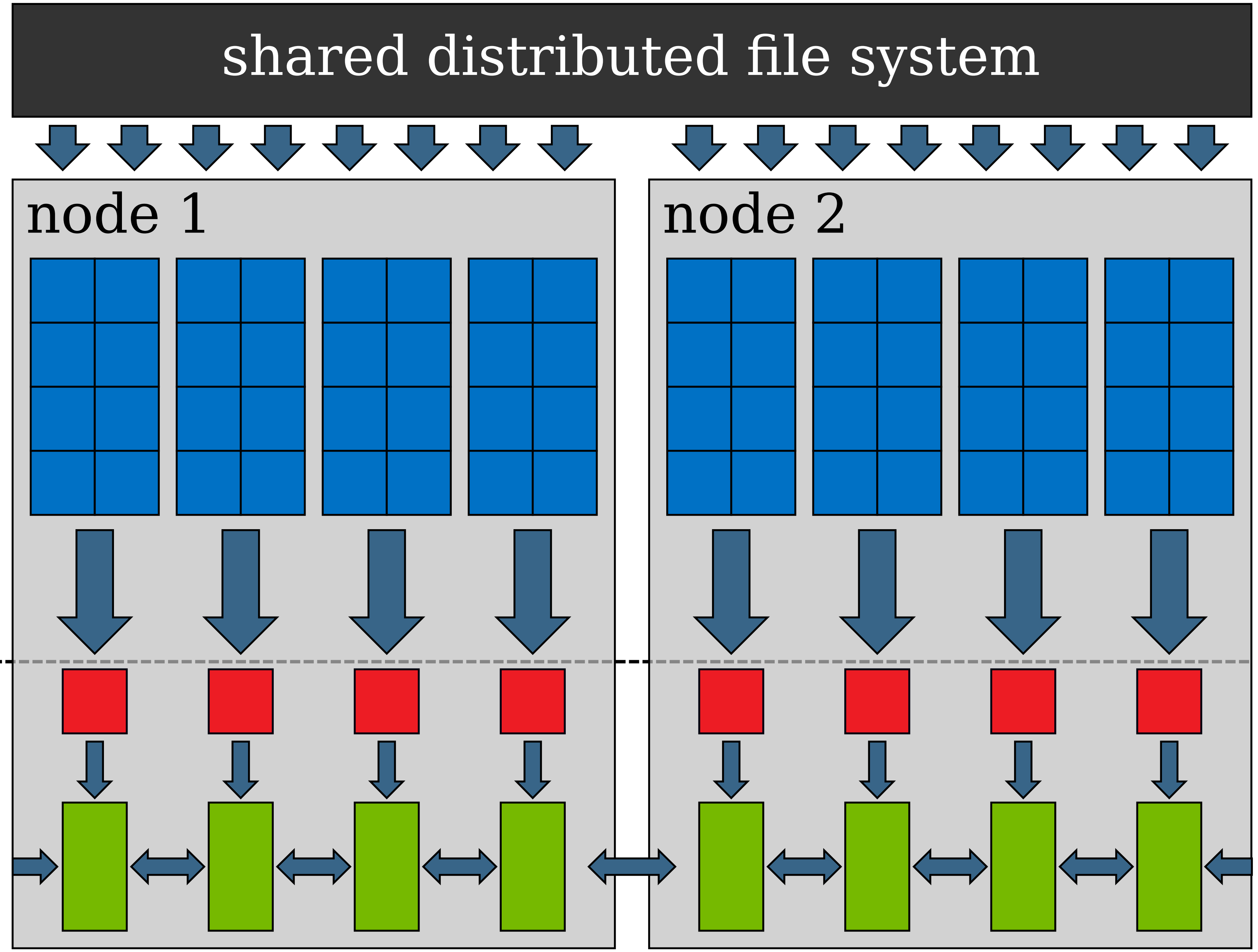
Distributed Deep Learning on HPC



data stage
- read data
- data augmentation
- transfer to masters

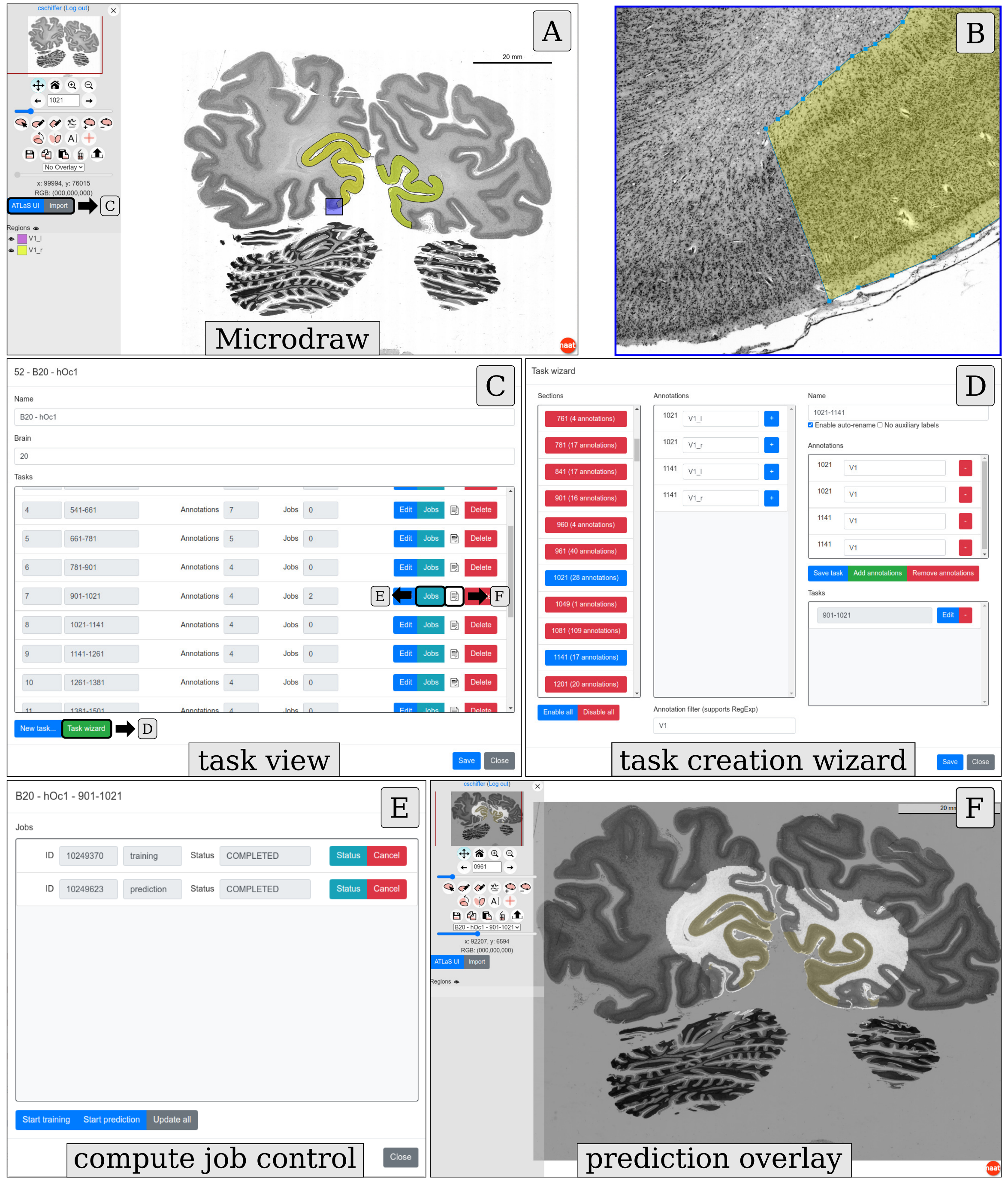
---

training stage
- receive data
- transfer to GPU
- parameter update



Application 1: Supporting Cytoarchitectonic Mapping with Deep Learning

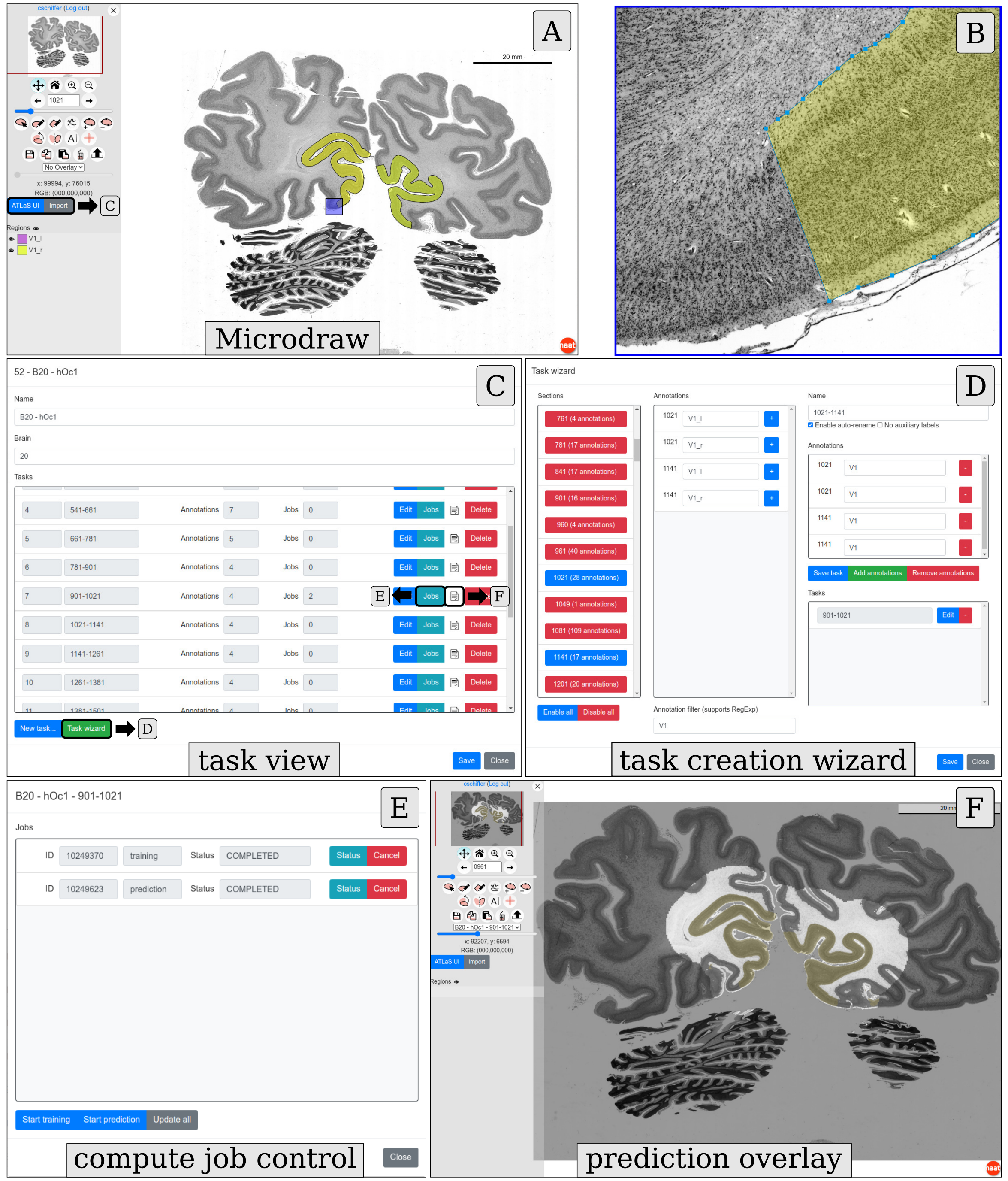
- **Goal:** Interactive workflow to support brain mapping
- **Idea:** Train specialized models using few annotations
  - Provide annotation on **every n-th** brain section
  - Train model on pairs of adjacent annotated sections
  - Apply model to **fill the gaps** between annotations
- **Web interface** for visualization, annotation, configuration



JÜLICH Forschungszentrum

Application 1: Computational requirements

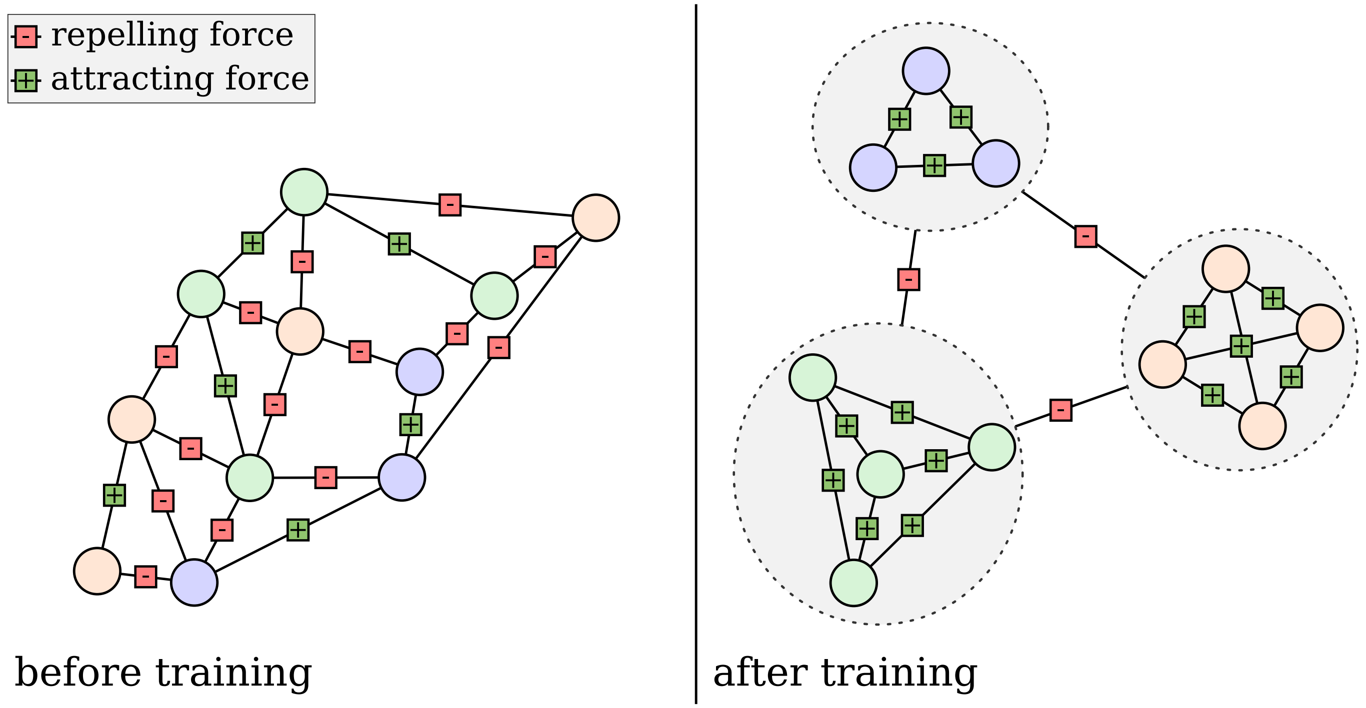
- Users define and submit **training** and **prediction** jobs
- Training and prediction on **JURECA-DC**, each job using...
  - Four **A100** GPUs (4 × 40GB)
  - **64 MPI ranks**, four threads per rank (256 total)
- Number of models depends on area size (≤ 20)
- **Runtime:** 10-15min → **Interactive use**



JÜLICH Forschungszentrum

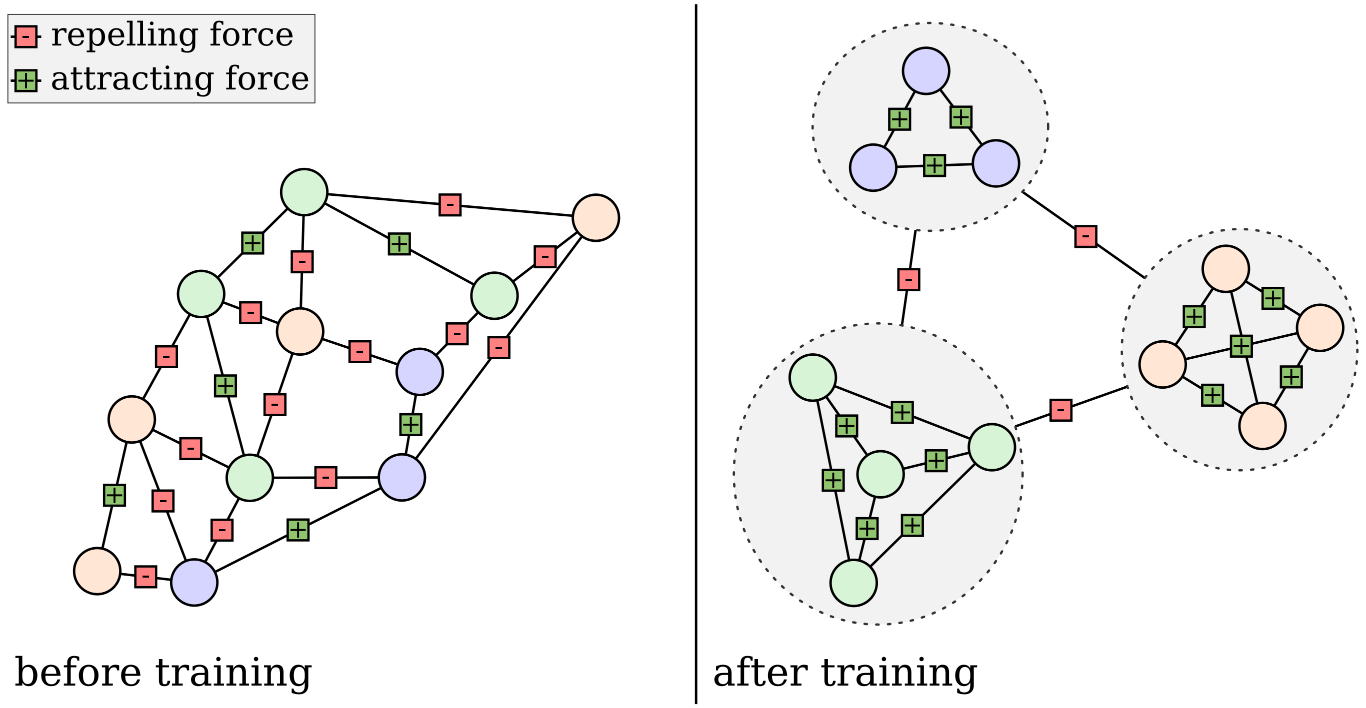
Application 2: Contrastive Cytoarchitectonic Feature Learning at Large Scale

- **Goal: General model** for cytoarchitecture classification
- **Approach:** Contrastive learning
  - Learn features by **comparison**
  - Make features of **similar images** similar
  - Make features of **dissimilar images** *dissimilar*
- **Similarity** based on **labels** or **probabilities**
- Learned features enable **classification** and **clustering**



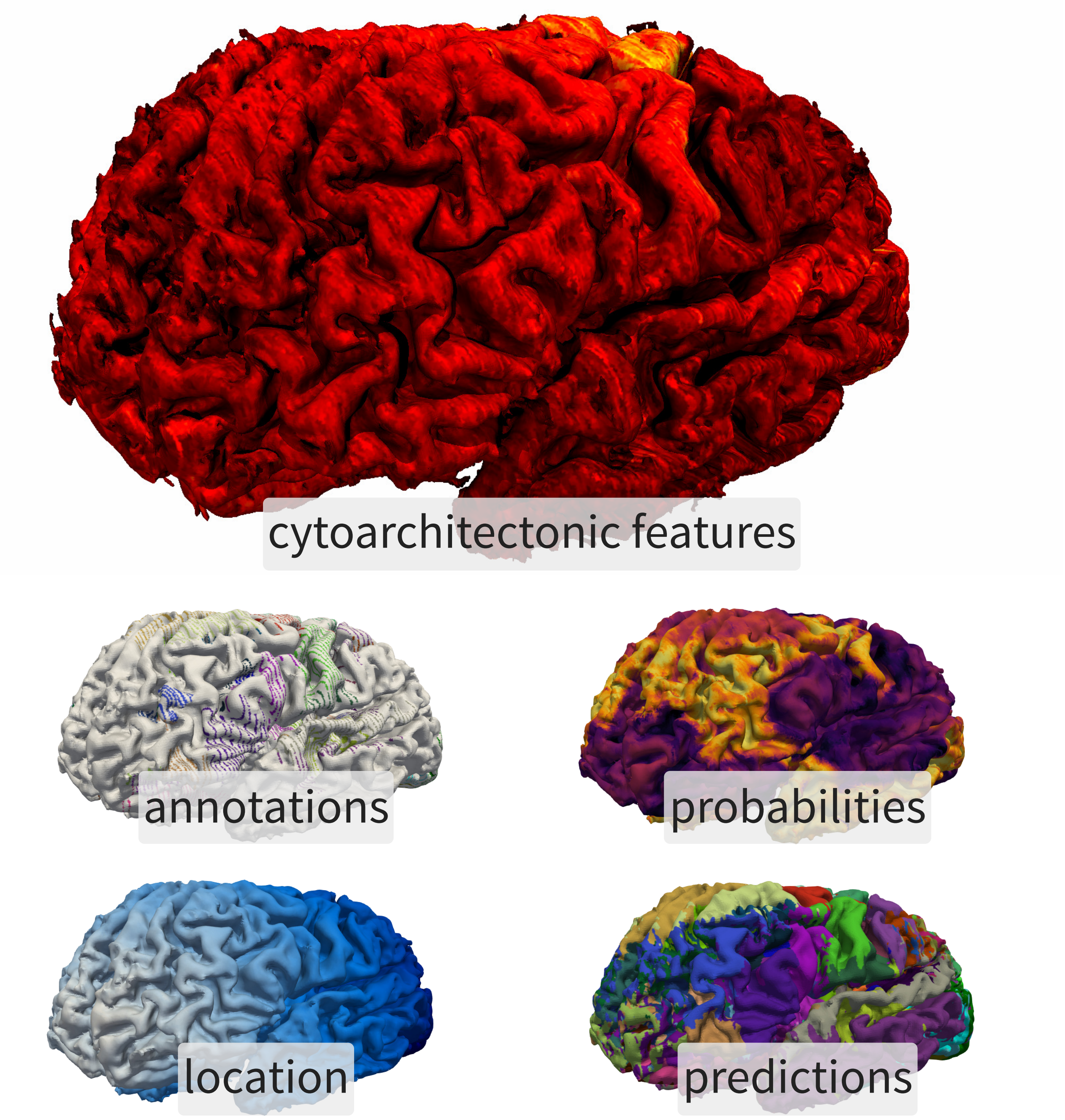
Application 2: Computational requirements

- **Challenge:** Large batch size for comparison
- Training on **JURECA-DC**
- **Contrastive training configuration**
  - **64 A100 GPUs** (16 nodes)
  - **1024 MPI ranks**, four threads per rank (4096 total)
  - 16 images per GPU (total GPU memory: 2.5 TB)
  - **Total data read:** ≥ 155 TB
  - **Runtime:** ≥ 6h
- Methods using **more data** in development



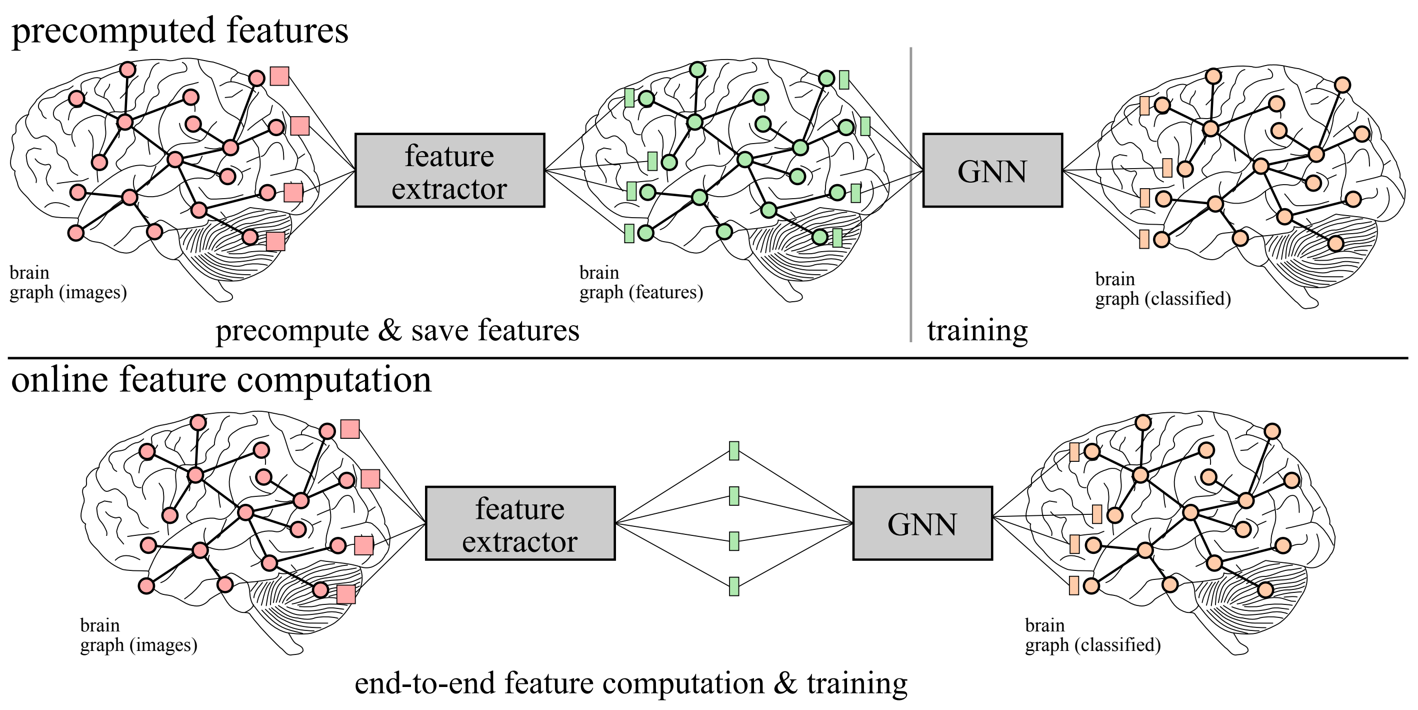
Application 3: Graph Neural Networks for Cytoarchitecture Classification

- **Previously:** Classify **individual** images
  - **Ill-defined:** Classification often requires context
  - Directly providing context (e.g., 3D) is **infeasible**
- **Idea:** Model brain as a **graph**
  - Coarse brain reconstruction to obtain a **mesh/graph**
  - Assign **image features to graph nodes**
  - Apply **graph neural networks (GNNs)** to classify nodes
- Improves performance by combining high-resolution **image features** with **context encoded in the graphs**



Application 3: Computational requirements

- Currently: **Pre-computed features**
- Training on **JURECA-DC**
- **Graph neural network training configuration**
  - **8 A100 GPUs** (2 nodes)
  - **128 MPI ranks,** four threads per rank (256 total)
  - **Runtime:** 20 - 120 min
  - Pre-computed attributed graphs: **~60 GB**
- **End-to-end feature and graph learning** in development



Future work

- Advanced **feature learning methods**
  - Use **non-annotated data** (self-supervised learning)
  - Compute requirements grow **linearly** with data
- **End-to-end feature and graph learning**
  - **End-to-end learning**
  - Enable **data augmentation** for robustness
  - Potentially combination with **contrastive learning**
  - **Challenge:** I/O and compute requirements **grow exponentially** with model depth

Christian Schiffer
Mail: c.schiffer@fz-juelich.de
Phone: +49 2461 61-3678

Team *Big Data Analytics*
Institute of Neuroscience and Medicine (INM-1)
Forschungszentrum Jülich

JÜLICH Forschungszentrum
HIBALL HELMHOLTZ International BigBrain Analytics & Learning Laboratory
HELMHOLTZ AI
hhu Heinrich Heine Universität Düsseldorf
Human Brain Project
JARA Jülich Aachen Research Alliance