DOE and NIH Partnerships
Cancer and Brain

Rick Stevens

Argonne National Laboratory

The University of Chicago

Crescat scientia; vita excolatur

Acknowledgements

Material in this talk is from Argonne, Berkeley, Oak Ridge, Livermore and Los Alamos, NCI, UCSF, Harvard and University of Chicago

Special Thanks to

Bobby Kasthuri (Argonne and Uchicago)
Jeff Lichtman (Harvard)
HansPeter Pfister (Harvard)
Peter Denes (Berkeley)
Gina Tourassi (Oak Ridge)
Fred Streitz (Livermore)
Frank Alexander (Los Alamos)
Eric Stahlberg (NCI)
Peter Littlewood (Argonne and Uchicago)

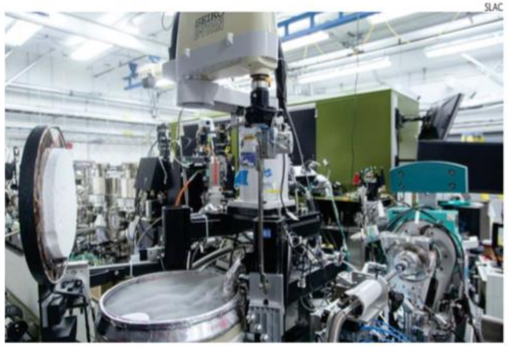
Cancer, Brain and Supercomputing

Three White House Initiatives

- **National Strategic Computing**
- **Precision Medicine**
- **BRAIN**

Cancer, brain research, and supercomputing

By contributing to health research, the Department of Energy could transform its approach to designing the next generation of high-performance computers.



SLAC

A HIGHLY AUTOMATED, ROBOTIC X-RAY CRYSTALLOGRAPHY SYSTEM at SLAC's Linac Coherent Light Source x-ray laser. The metal drum at the lower left contains liquid nitrogen for cooling crystallized samples. This setup was used to explore the molecular machinery involved in brain signaling in atomic-scale detail.

When computers capable of working at the exascale level (10^{18} floating-point calculations per second) come on line, they will be brought to bear on figuring out how another, quite different computer, the human brain, works. With that goal in mind, Energy secretary Ernest Moniz and National Institutes of Health director Francis Collins are exploring how to bring the Department of Energy, which houses the nation's leading supercomputers, into the presidential initiative known as BRAIN (Brain Research through Advancing Innovative Neurotechnologies; see PHYSICS TODAY, December 2013, page 20).

The brain is just one area of biomedical research that could benefit from the computational and physical sciences expertise at DOE and its national laboratories. In December Moniz asked his Secretary of Energy Advisory Board (SEAB) to look for ways to increase DOE's contribution to biomedical sciences. A SEAB task force, cochaired by former NIH and National Cancer Institute (NCI) director Harold Varmus and former DOE undersecretary Steven Koonin, will report to him in September.

The BRAIN Initiative will require advances across several scientific fields. "We need better ways of detecting and recording neural signals," says Roderic Pettigrew, director of NIH's National Institute of Biomedical Imaging and Bioengineering. "Then we need analytical tools to interpret those signals. We need ways of deciphering meaningful signals from noise, an area DOE scientists are accustomed to dealing with."

Another area of focus is the modeling of what goes on in the brain, resolved in three dimensions and in time. "People often don't think of the time domain of medical data," notes Pettigrew, the designated liaison to DOE. "But life is temporal, and biological dimensions change in the time domain. Proteins fold and unfold, protein receptors go from inactive to active state."

In October representatives from the two agencies held a jointly sponsored BRAIN workshop at Argonne National Laboratory that coincided with a major neuroscience conference in nearby Chicago. Reports from those discussions were delivered to Moniz and Collins but haven't been made public.

"There is a lot of opportunity and a lot of need in the neuroscience community to benefit from the tools and the organization of the labs to do this kind of big project," Moniz told reporters in November, days before issuing his charge to SEAB.

Dimitri Kusnezov, chief scientist for DOE's National Nuclear Security Administration, is involved in discussions with NIH. "The question we're asking ourselves is, Are there real wins in pushing diagnostics—for example, in a multimode analysis—or is the community geared to move forward at the same pace anyway?" he says. "Can we accelerate things in a significant way or not? We don't have the answer yet."

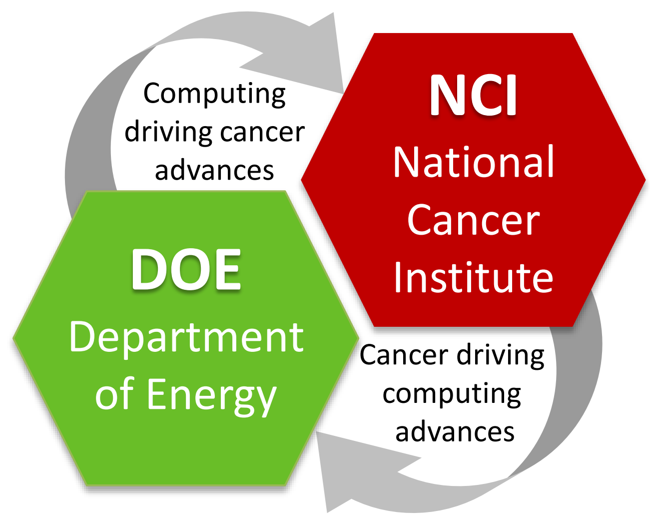
Biomedical research has long benefited from DOE assets. Life-sciences researchers represent the single largest sector of users (about 40%) at the DOE national laboratories' x-ray light sources, half of whom are supported by NIH. And initial genome-sequencing work at Los Alamos National Laboratory begat the NIH-led Human Genome Project.

The nanoscale-science research centers operated by the national labs and other groups have been developing sensors that can read nanoparticles. "It's conceivable that nanoparticles with certain characteristics can be embedded in a living system like a brain," says Steve Binkley, associate director for advanced scientific computing research in DOE's Office of Science. "And one could then also conceive of reading the signals coming out of them. The holy grail is to get real-time mapping of signals that exist in neurons as a function of time to certain stimuli," he says. Such mapping has been done with mice, but scientists used invasive probes not suitable for research on humans.

Imaging is another DOE strength that will be useful to BRAIN, Binkley says. The labs have expertise using UV, x rays, IR, coherent light sources, and lasers for imaging. "It's often not obvious at the outset how one puts all those things together to image a certain type of thing.

Joint Design of Advanced Computing Solutions for Cancer

DOE-NCI partnership to advance cancer research and high performance computing in the U.S.



December 11, 2015

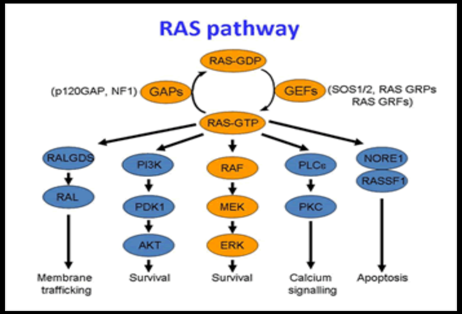
Presented to:
Secretary Moniz and Director Lowy

U.S. DEPARTMENT OF ENERGY | NIH NATIONAL CANCER INSTITUTE

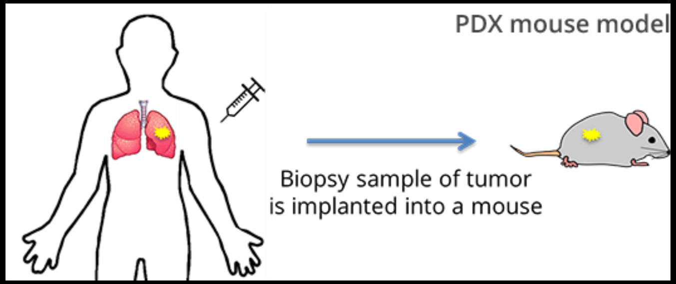
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. LLNL-PRES-678642-DRAFT.

The NCI-DOE partnership will extend the frontiers of precision oncology (Three Pilots)

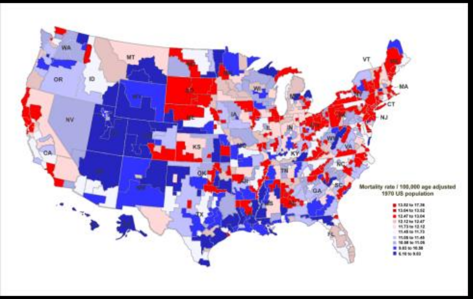
- **Cancer Biology**
  - **Molecular Scale Modeling of RAS Pathways**
  - Unsupervised Learning and Mechanistic models
  - Mechanism understanding and Drug Targets



- **Pre-clinical Models**
  - **Cellular Scale PDX and Cell Lines**
  - ML, Experimental Design, Hybrid Models
  - Prediction of Drug Response

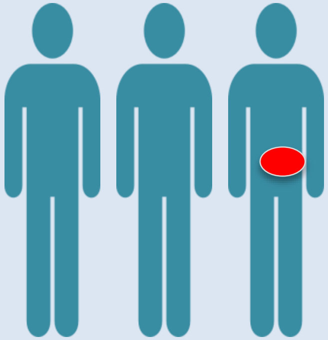


- **Cancer Surveillance**
  - **Population Scale Analysis**
  - Natural Langue and Machine Learning
  - Agent Based Modeling of Cancer PateintTrajectories

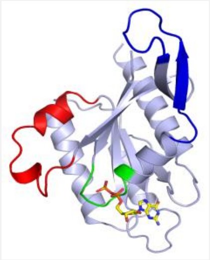


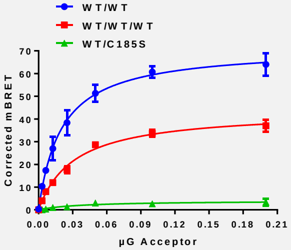
Developing new therapeutic approaches to target RAS-driven cancer

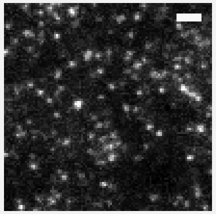
30% of cancers have mutated RAS
~1M deaths/year



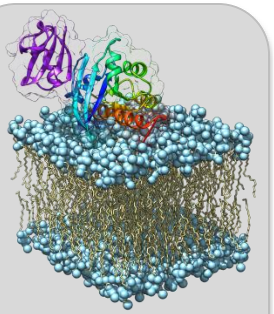
Current therapies ineffective against RAS-driven cancer



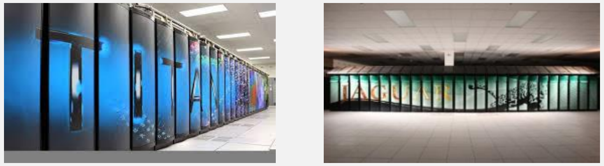




Molecular Dynamics
Simulation
Modeling

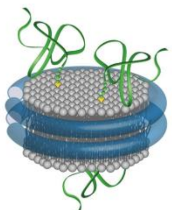


RAS biology
ID targets
New inhibitors

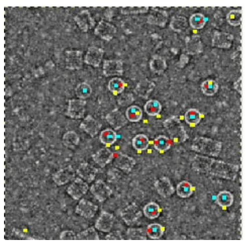


Pilot 2: RAS proteins in membranes

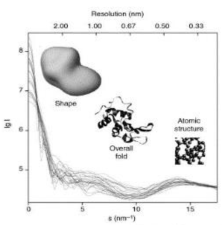
RAS activation experiments at NCI/FNL



Experiments on nanodisc



CryoEM imaging



X-ray/neutron scattering

Multi-modal experimental data, image reconstruction, analytics

Protein structure databases

New adaptive sampling molecular dynamics simulation codes

Adaptive time stepping

Coarse-grain MD

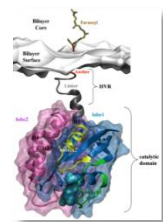
Classical MD

Quantum MD

Adaptive spatial resolution

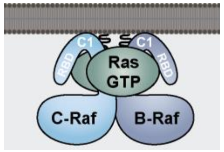
High-fidelity subgrid modeling

Predictive simulation and analysis of RAS activation



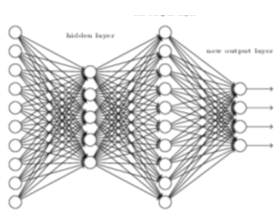
Granular RAS membrane interaction simulations

Atomic resolution sim of RAS-RAF interaction

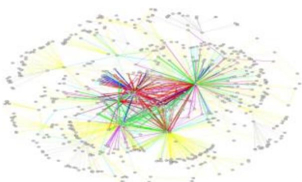


Inhibitor target discovery

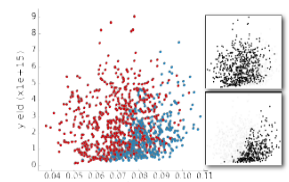
Machine learning guided dynamic validation



Unsupervised deep feature learning



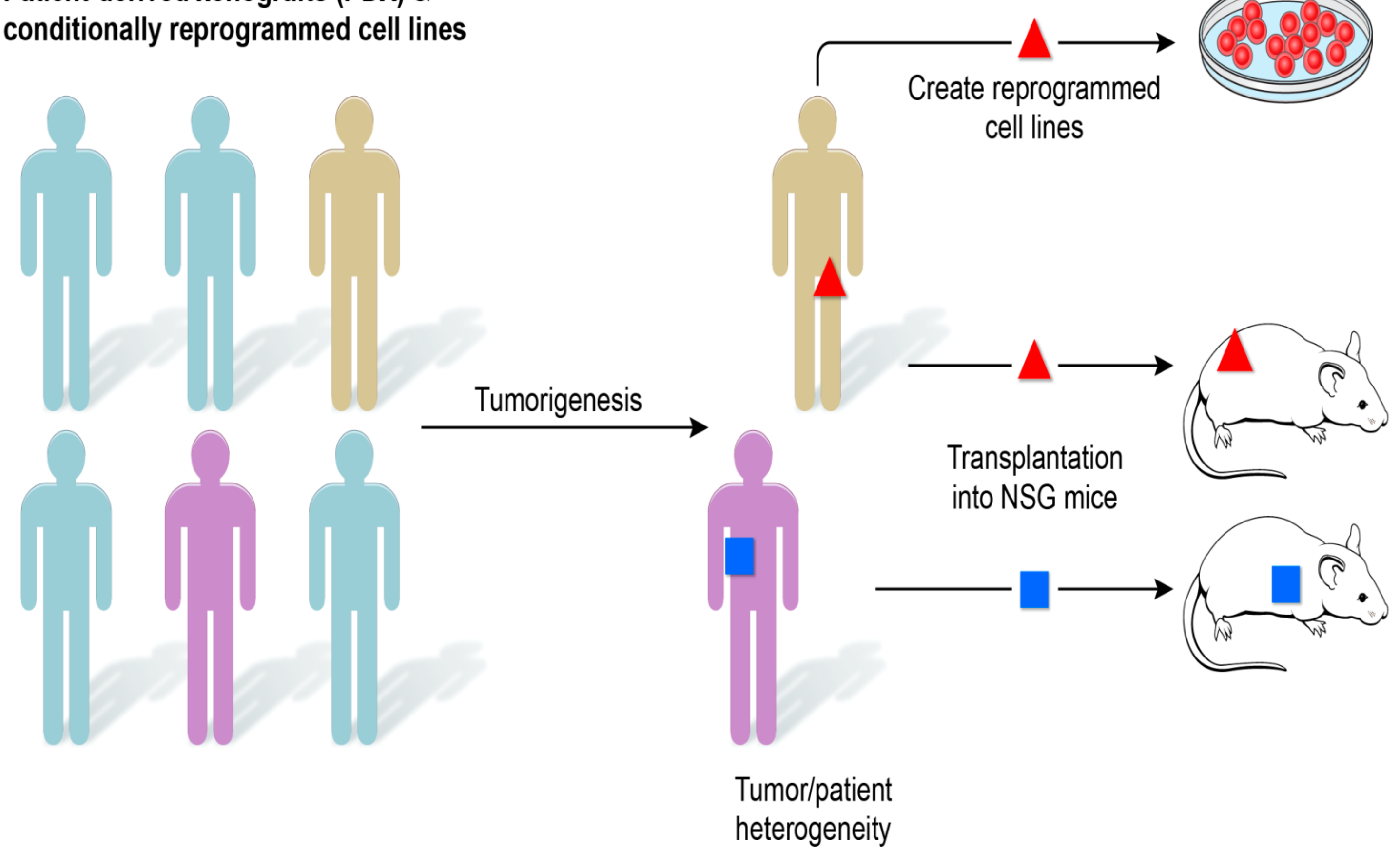
Mechanistic network models



Uncertainty quantification

Patient Derived Xenograft Models

Patient-derived xenografts (PDX) & conditionally reprogrammed cell lines

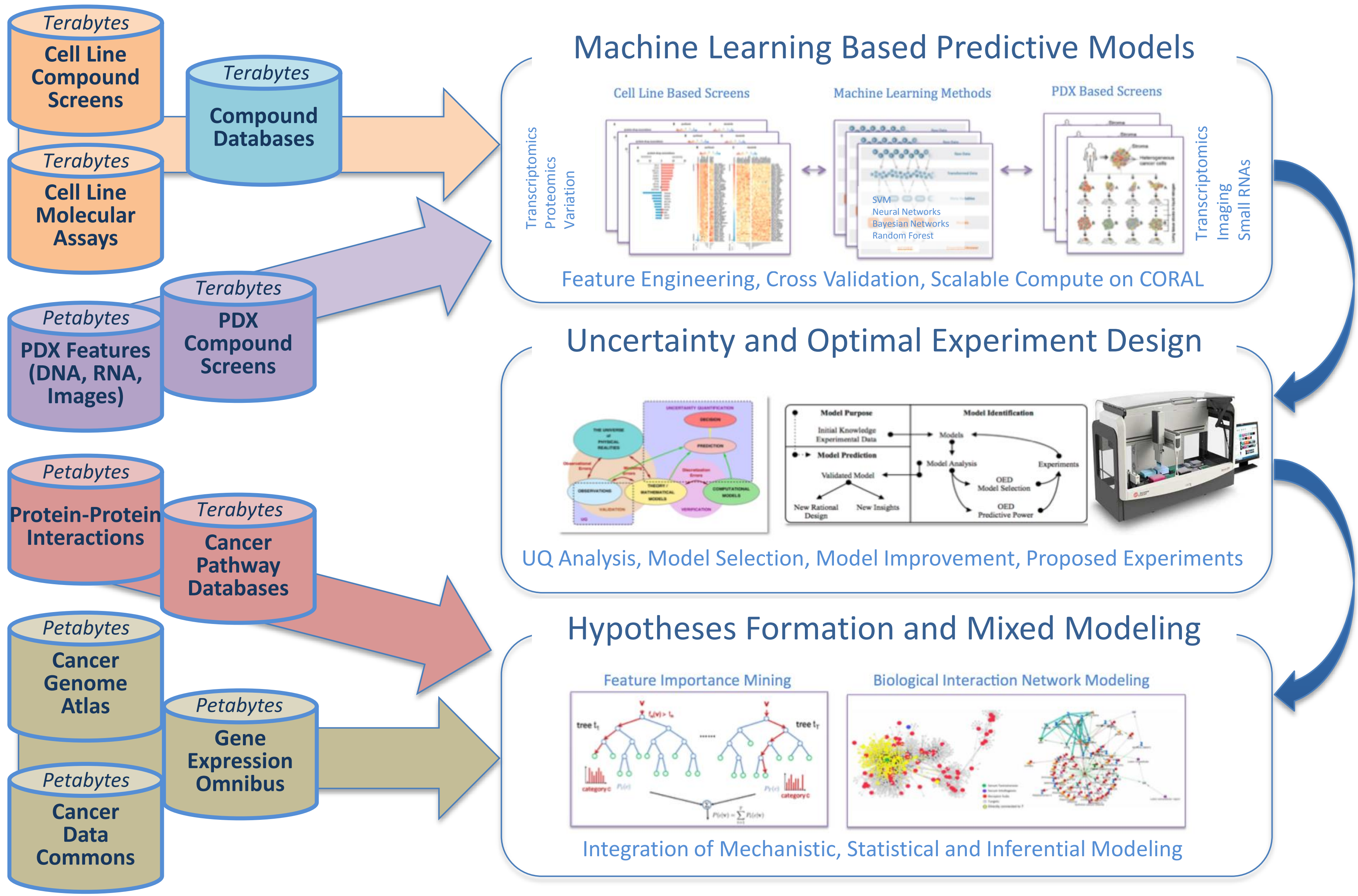


Molecularly characterize, treat/screen mice bearing transplants & cells with relevant drugs.

"Pre-clinical clinical trials"

Nature Rev. Clin. Oncol. 11: 649-662, 2014.

Pilot 1: Predictive Models for Pre-Clinical Screening

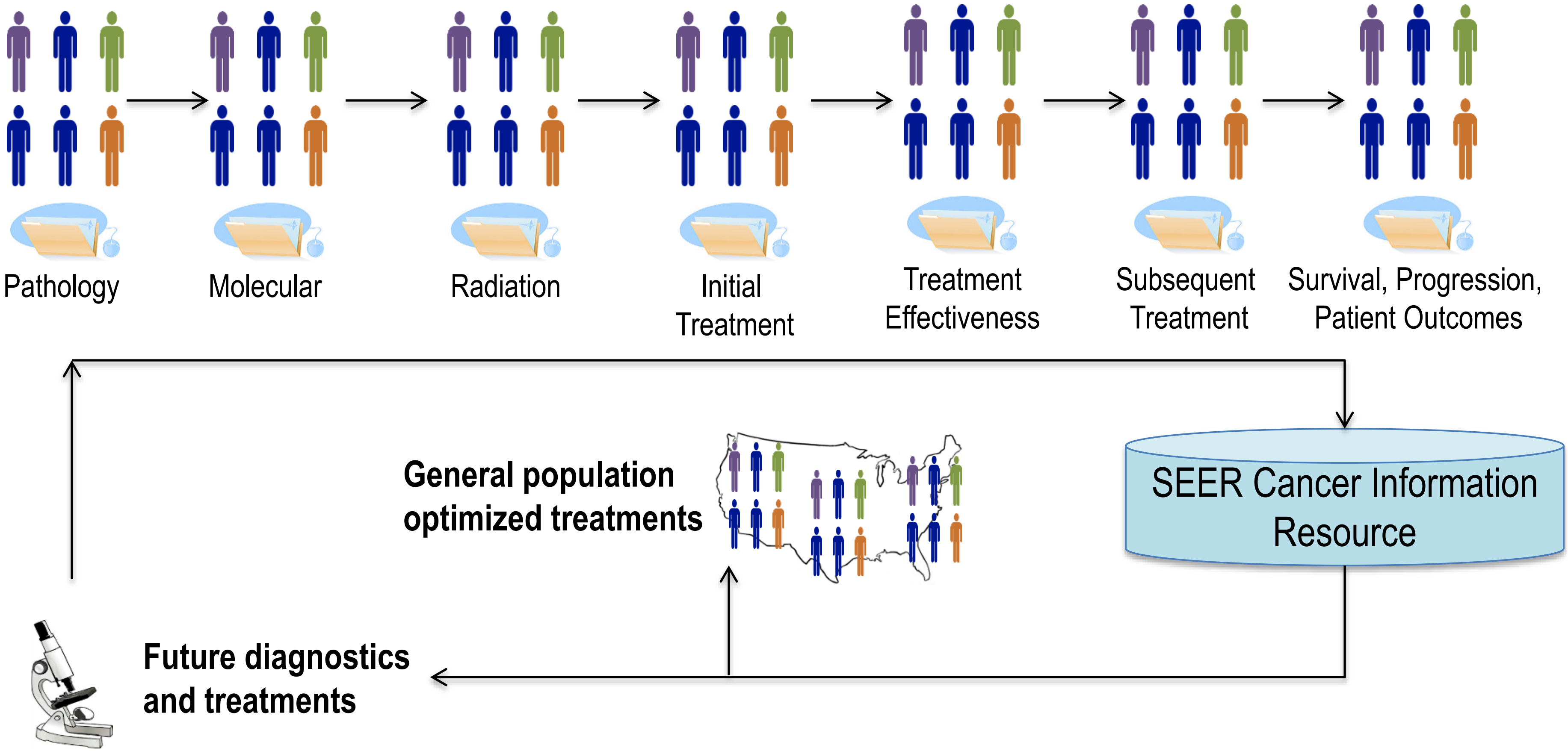


Aims for Pre-Clinical Screening Pilot

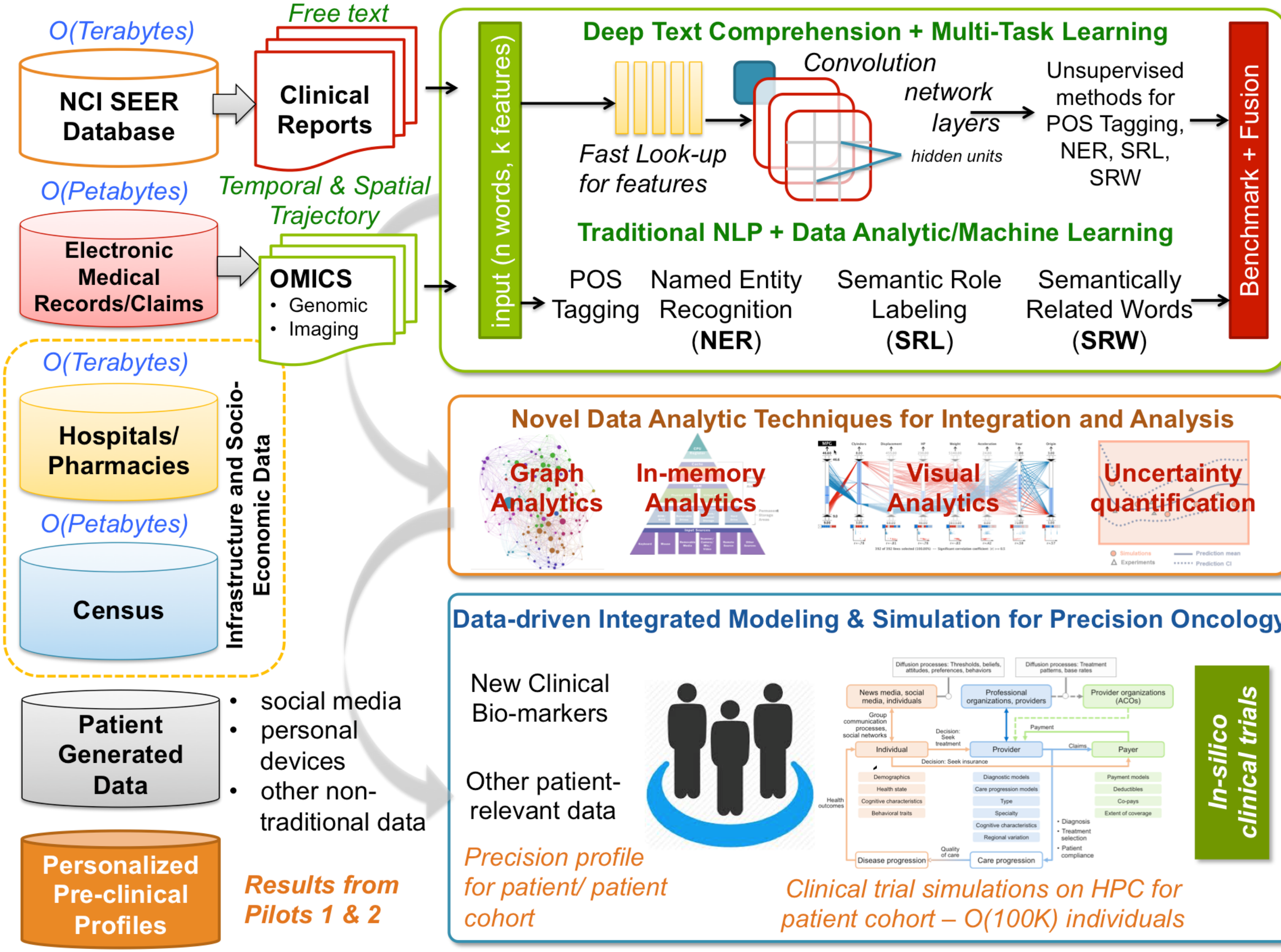
- Reliable machine learning based predictive models of drug response that enable the projection of screening results from and between cell-lines and PDX models
- Uncertainty quantification and optimal experimental design to assert quantitative limits on predictions and to recommend experiments that will improve predictions
- Improved modeling paradigms that support the graded introduction of mechanistic models into the machine learning framework and to rigorously assess the potential modeling improvements obtained thereof

Cancer Patient Surveillance and Information Integration

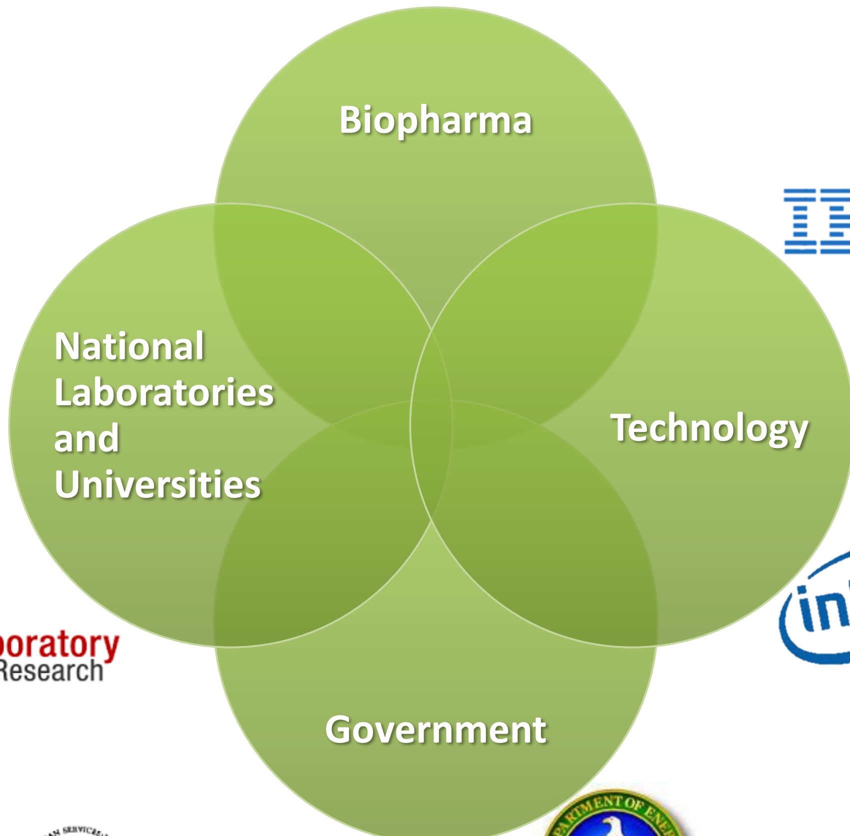
Cancer patient demographic and clinical outcomes data



Pilot 3: Population Information Integration, Analysis and Modeling



Emerging NSCi Public Private Partnership for Computing Precision Medicine



Frederick National Laboratory for Cancer Research

National Laboratories and Universities
ANL
ORNL
LLNL
LANL
FNLCR
Harvard University
University of Chicago

Government
Department of Energy
National Cancer Institute

Biopharma
Glaxo Smith-Kline

Technology
Intel
IBM

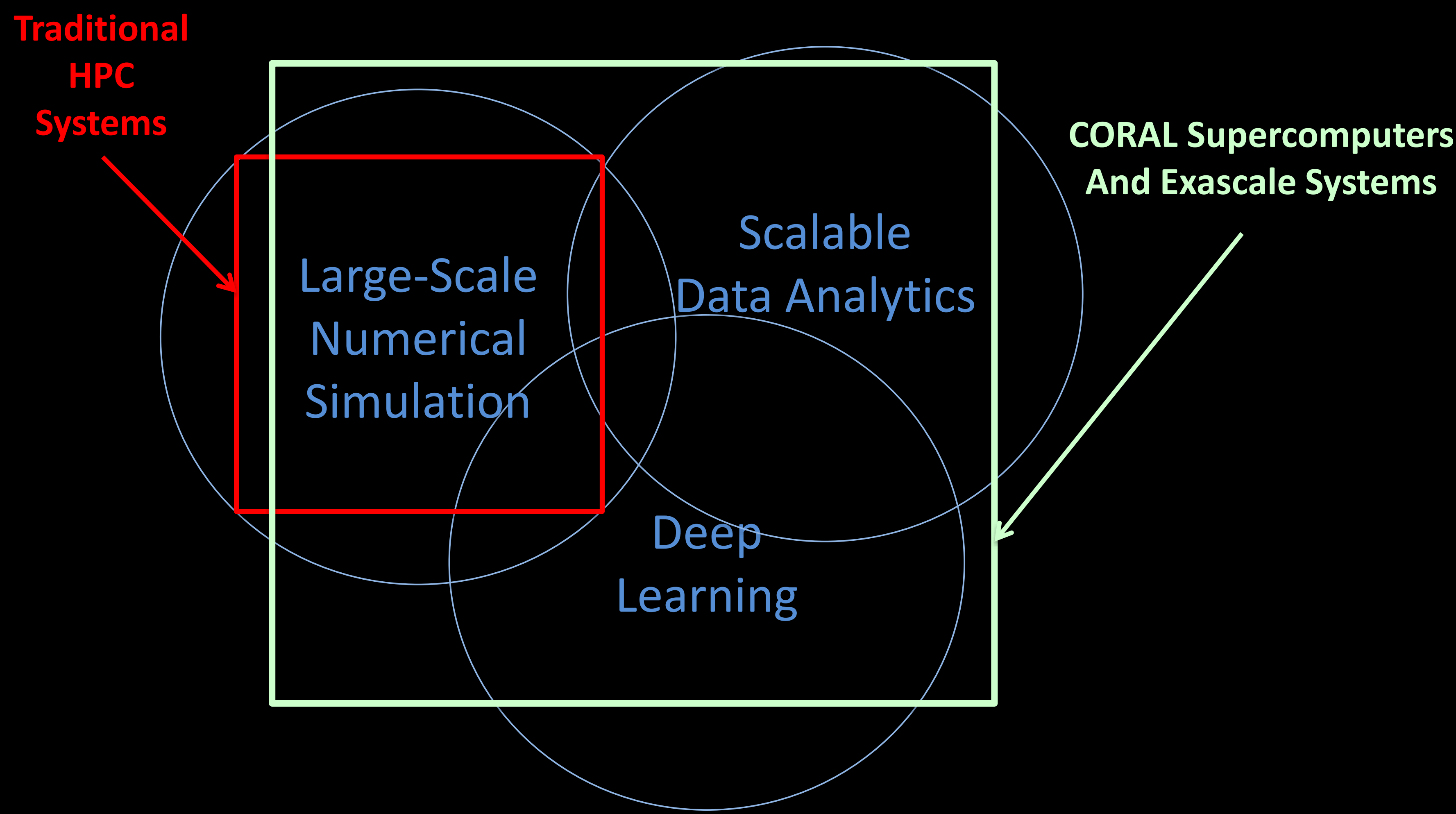
The NCI-DOE partnership will extend the frontiers of DOE computing capabilities

- **In simulation**
  - Atomic-resolution MD simulations of critical protein complex interactions that will require exaflops of floating point performance
  - New integrations of QM and multi-timescale methods that enable high-accuracy interactions over extended time windows
  - Integration of data-driven modeling and analytics at scale for rapid-cycle new intervention development and testing *in silico*
  - Extended theory and tools for UQ in multiple spatial and temporal scales
- **In data analytics**
  - Learning dynamic patterns from molecular to population scale data sets on CORAL-class architectures
  - Integrated machine-learning and simulation systems that bring together mechanistic and probabilistic models
- **In new computing architectures**
  - Codesign of architectures integrating learning systems and simulation in new memory-intensive hierarchies
  - Growth of new computing ecosystems bringing together leadership-class HPC and cloud-based data systems
  - Integration of beyond Von Neumann architectures into mission workflows

The NCI-DOE partnership will extend the frontiers of precision oncology (Three Pilots)

- **In understanding cancer biology**
  - Deepen awareness of disease initiation in key RAS-related cancers
  - Improve understanding of critical cancer pathways
  - Develop new molecular models to probe and explain complexities of cancer
  - Develop predictive models to identify novel targets and substances
- **In pre-clinical models**
  - Develop technologies to bridge insight between cell line and PDX models
  - Accelerate identification and evaluation of new promising cancer drugs
  - Prepare foundations to expand breadth of treatments and conditions for cancer precision medicine
- **By expanding the population's role in future advances**
  - Increase comprehensiveness and efficiency of critical information within cancer registries
  - Identify new biomarkers impacting patient outcomes
  - Develop capabilities to identify optimal care pathways for cancer patients
  - Develop data-driven predictive models of patient health trajectories

Integration of Simulation, Data Analytics and Machine Learning



Deep Learning and Drug Screening @Johnson and Johnson

Jörg K. Wegner and Hugo Ceulemans, et. al. (NIPS2014)

“**Deep learning outperformed all other methods with respect to the area under ROC (auc 0.83) curve and was significantly better than all commercial products**. Deep learning surpassed the threshold to make virtual compound screening possible and has the potential to become a standard tool in industrial drug design.”

Hybrid Models in Cancer

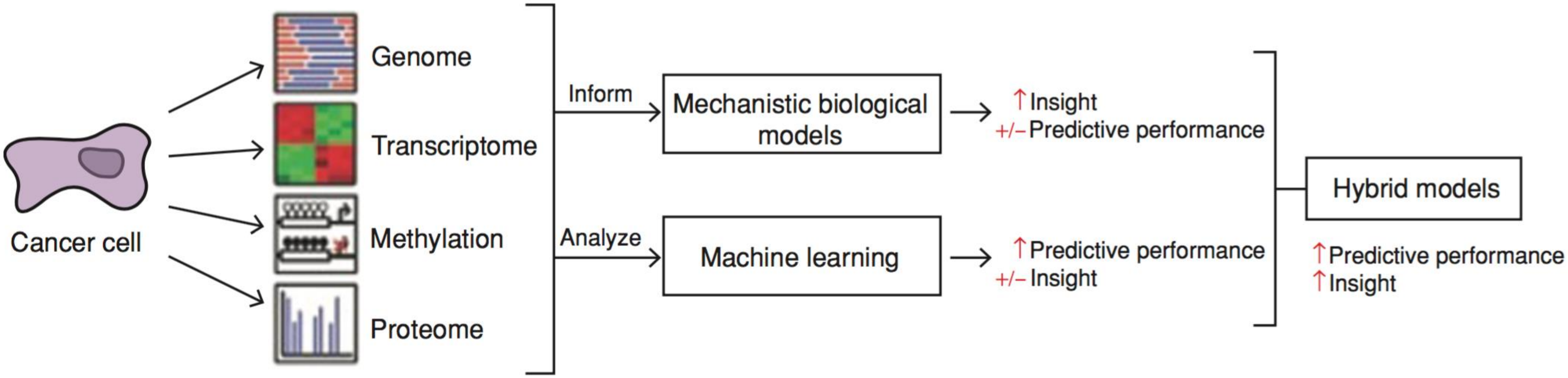


Figure 1. In two DREAM challenges, high throughput data characterizing cancer cells are used to build predictive models. Mechanistic models provide insight into the underlying biology, but do not take full advantage of the information within the data to achieve high performance. Machine learning methods are associative and extract maximum predictive value from the data, but do not always provide insight about mechanism. The future may bring hybrid models that combine the best of both approaches.

Predicting Cancer Drug Response: Advancing the DREAM

Russ B. Altman

Summary: The DREAM challenge is a community effort to assess current capabilities in systems biology. Two recent challenges focus on cancer cell drug sensitivity and drug synergism, and highlight strengths and weaknesses of current approaches. *Cancer Discov; 5(3); 237–8. ©2015 AACR.*

the WHITE HOUSE

BRAIN INITIATIVE

BRAIN RESEARCH THROUGH ADVANCING INNOVATIVE NEUROTECHNOLOGIES

The Big Picture Goal

- **The challenge is to map the circuits of the brain, measure the fluctuating patterns of electrical and chemical activity flowing within those circuits, and understand how their interplay creates our unique cognitive and behavioral capabilities.**
- We should pursue this goal simultaneously in humans and in simpler nervous systems in which we can learn important lessons far more quickly. But our ultimate goal is to understand our own brains.

DENTATE GYRUS

POSSIBLE LONG-TERM OUTCOMES

The BRAIN Initiative has the potential to do for neuroscience what the Human Genome Project did for genomics by supporting the development and application of innovative technologies that can create a dynamic understanding of brain function. It aims to help researchers uncover the mysteries of brain disorders, such as Alzheimer's and Parkinson's diseases, depression, Post-Traumatic Stress Disorder (PTSD), and traumatic brain injury (TBI).

The Human Genome Project demonstrates the potential impact that ambitious research programs like the BRAIN initiative can have.

From 1988-2003, the Federal Government invested $3.8 billion in the Human Genome Project, which has since generated an economic output of $796 billion —a return of $141 for every $1 invested.

Goals of the BRAIN 2025

- Discovering diversity: cell types
- Maps at multiple scales: connectome
- Brain in action: dynamic activity
- Demonstrating causality: link to behavior
- Identifying fundamental principles
- Advancing human neuroscience
- BRAIN to brain: integration and translation

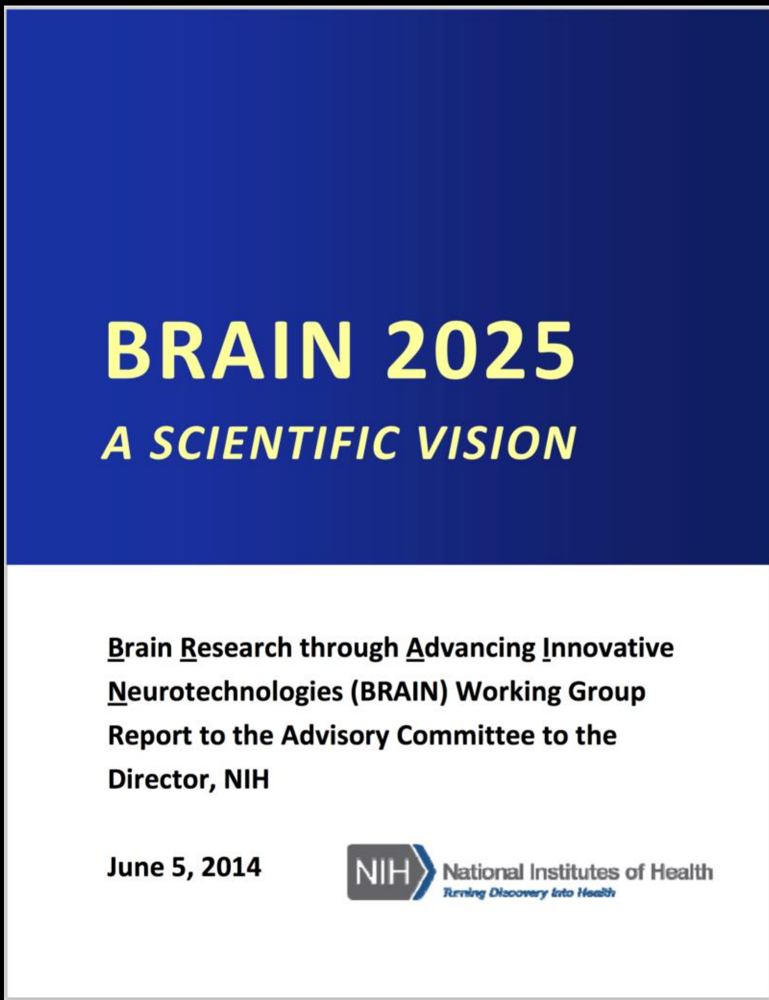
NIH Blueprint for Brain

started in 2004.. GC launched in 2009 (15 institutes)

Blueprint Grand Challenges

- The **Human Connectome Project** is an effort to map the connections of the healthy brain. It is expected to help answer questions about how genes influence brain connectivity, and how this in turn relates to mood, personality and behavior. The investigators will collect brain imaging data plus genetic and behavioral data from 1,200 adults. They are working to optimize brain imaging techniques to see the brain's wiring in unprecedented detail. Building on the success of the Connectome Project, in 2014 the Blueprint authorized funds to expand the age range of normal subjects to include both young people and older adults.
- The **Grand Challenge on Chronic Neuropathic Pain** supports research to understand the changes in the nervous system that cause acute, temporary pain to become chronic. The initiative has supported multi-investigator projects to partner researchers in the pain field with researchers in the neuroplasticity field.
- The **Blueprint Neurotherapeutics Network** is helping small labs develop new drugs for nervous system disorders. The Network provides research funding, plus access to millions of dollars worth of services and expertise to assist in every step of the drug development process, from laboratory studies to preparation for clinical trials. Project teams across the U.S. have received funding to pursue drugs for conditions from vision loss to neurodegenerative disease to depression.

Overall Planning Document
(15 academic authors, NIH, NSF, DARPA, FDA)



BRAIN 2025

A SCIENTIFIC VISION

Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Working Group Report to the Advisory Committee to the Director, NIH

June 5, 2014

NIH National Institutes of Health
Turning Discovery Into Health

- Vision and Philosophy
- Priority Research Areas
- Implementation goals, deliverables, timelines and Costs
- 6 workshops
- ~100 Participants

- computer scientists?
- Mathematicians?

146 pages

CS mentioned 5 times

Math mentioned 4 times

Proposed BRAIN Initiative 12 Year Budget

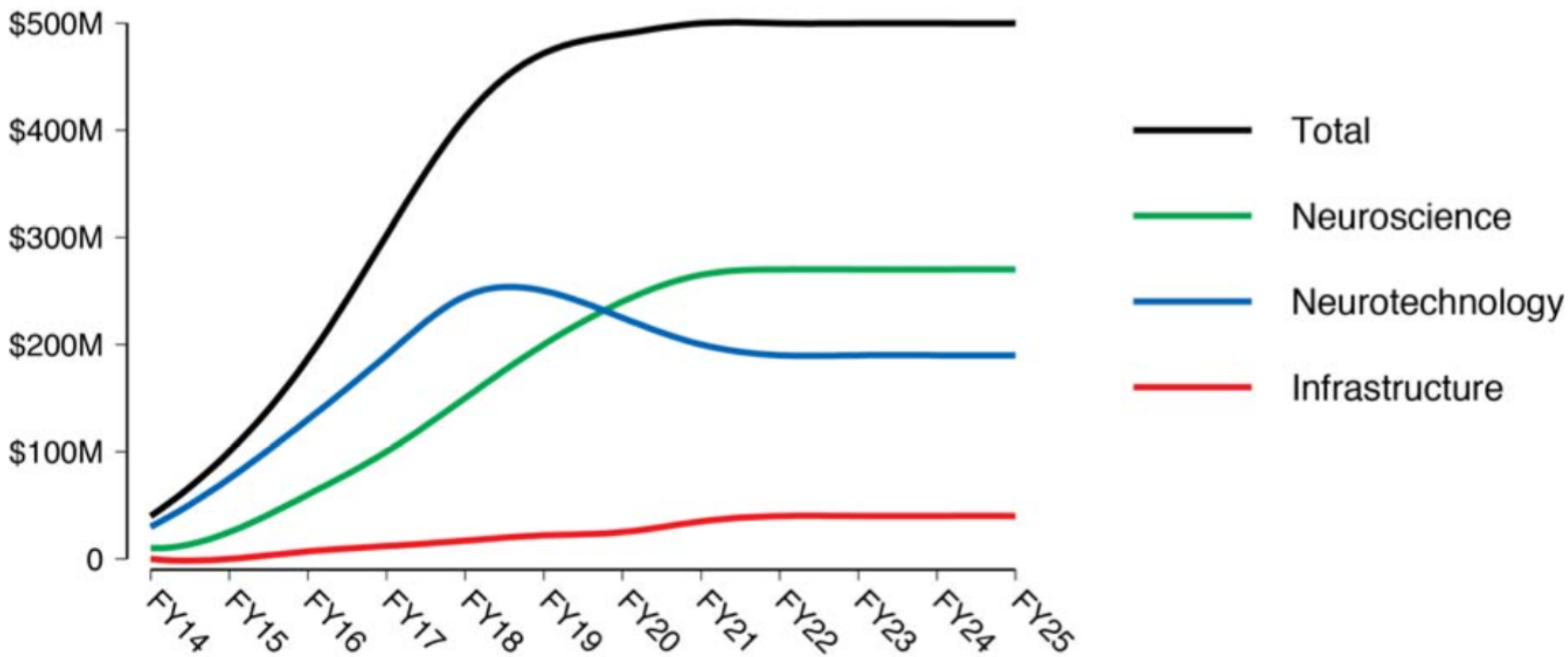
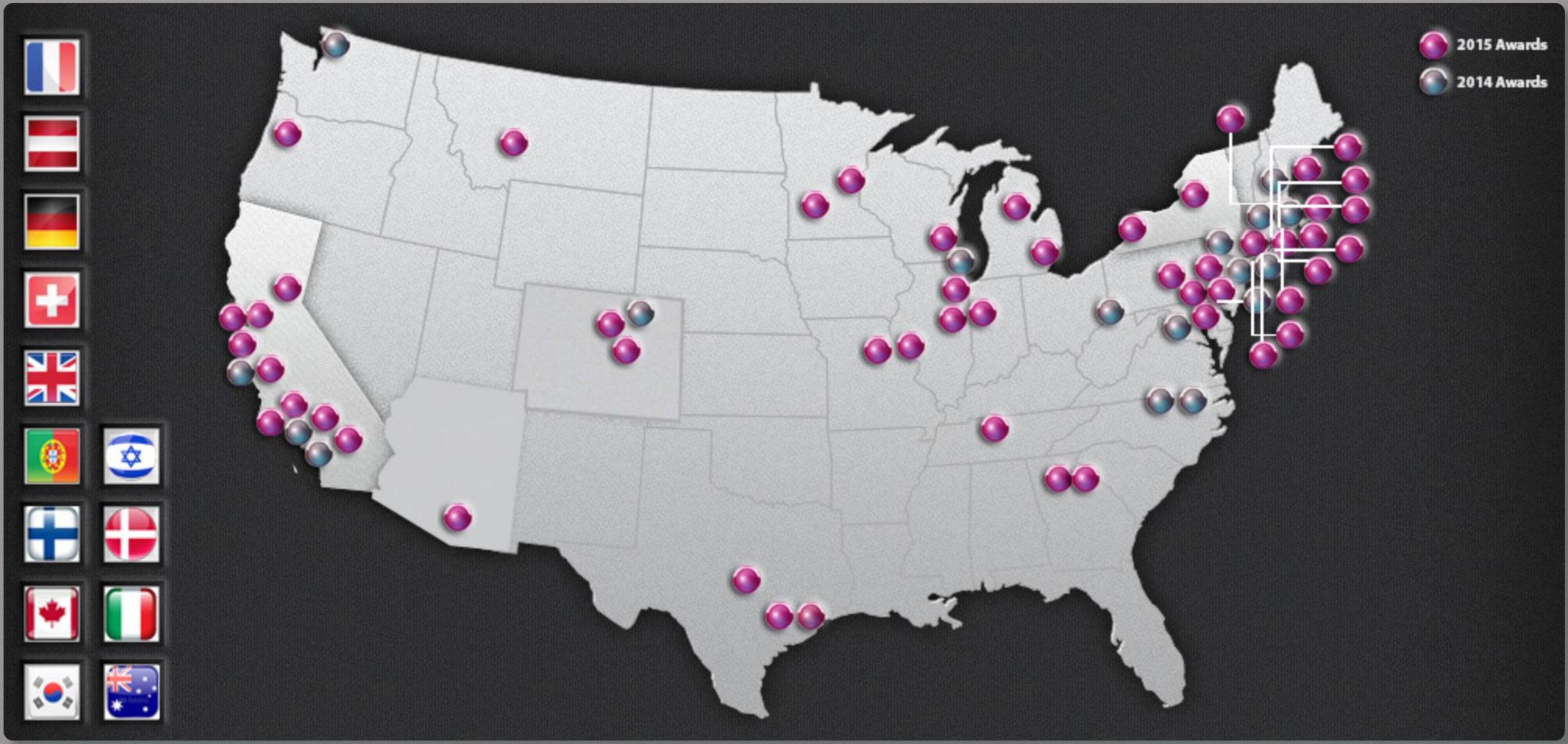


Figure caption. Proposed 12-year budget for the BRAIN Initiative. Collaborative technology development is emphasized through FY2019, while discovery-driven science receives priority beginning in FY2020. 'Infrastructure' is for facilities and capabilities that will benefit researchers across the entire nation, with emphasis on data sharing resources, training in the use of new technologies and quantitative methods, and possible regional instrumentation centers during the last half of the BRAIN Initiative.

BRAIN initiative Awards



BRAIN Initiative Major Areas

- Cell Types
- Circuit Diagrams
- Monitor Neural Activity
- Interventional Tools
- Theory and Data Analysis Tools
- Human Neuroscience
- Integrated Approaches

FY2016 Investments

FY2015 ~$200M

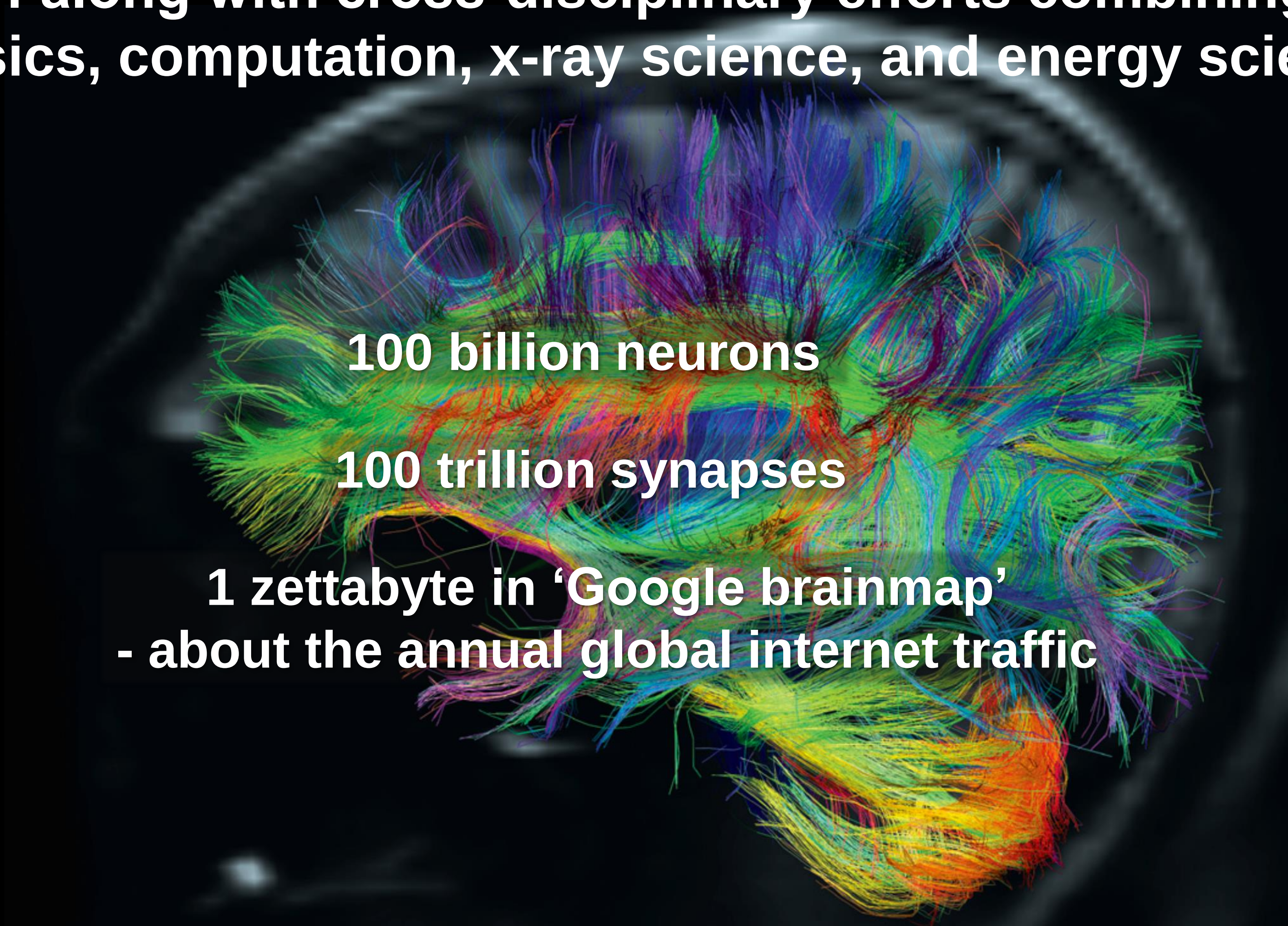
- NIH $135M
- DARPA $95M
- NSF $72M
- IARPA $XM
- FDA $XM

Building off of $100 million in commitments announced last year at NIH, NSF and DARPA, the BRAIN Initiative is growing to five participating federal agencies with the addition of FDA and IARPA.

DOE has a proposed FY17 role for BES, BER and ASSCR

The challenge of understanding the brain requires extraordinary advances in neuroscience…

… along with cross-disciplinary efforts combining physics, computation, x-ray science, and energy science

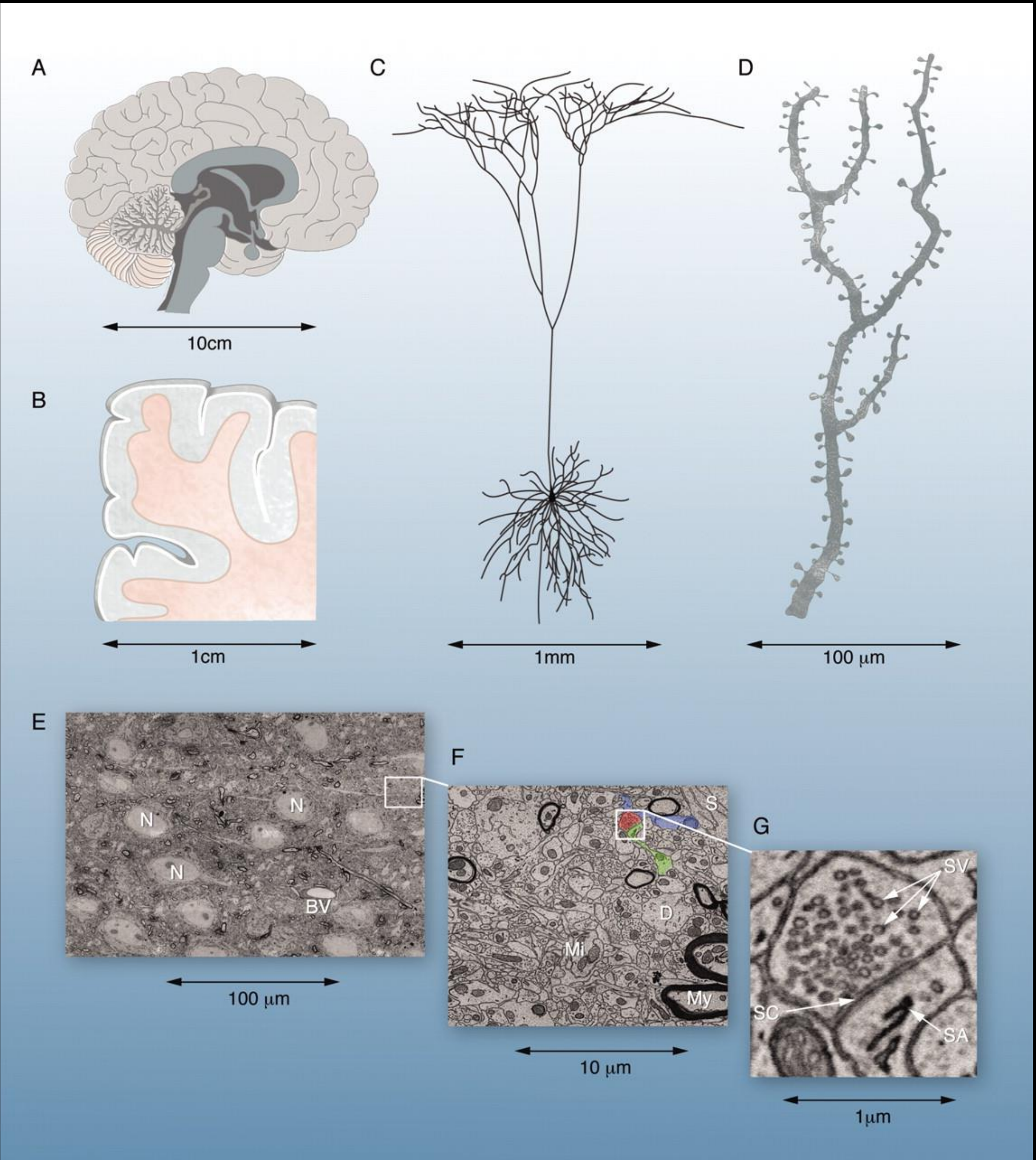


100 billion neurons

100 trillion synapses

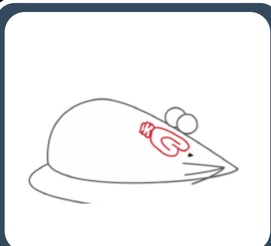
1 zettabyte in ‘Google brainmap’
- about the annual global internet traffic

Scales in the brain

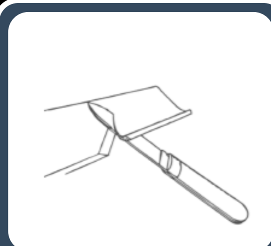


Lichtman and Denk 2011

Connectomics Workflow

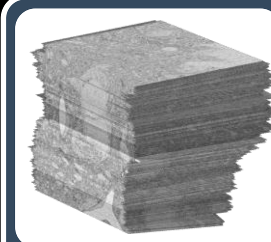


Tissue
Preparatio
n



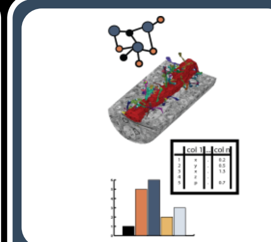
Sectioning
& Wafer
Prep

Image
Acquisition



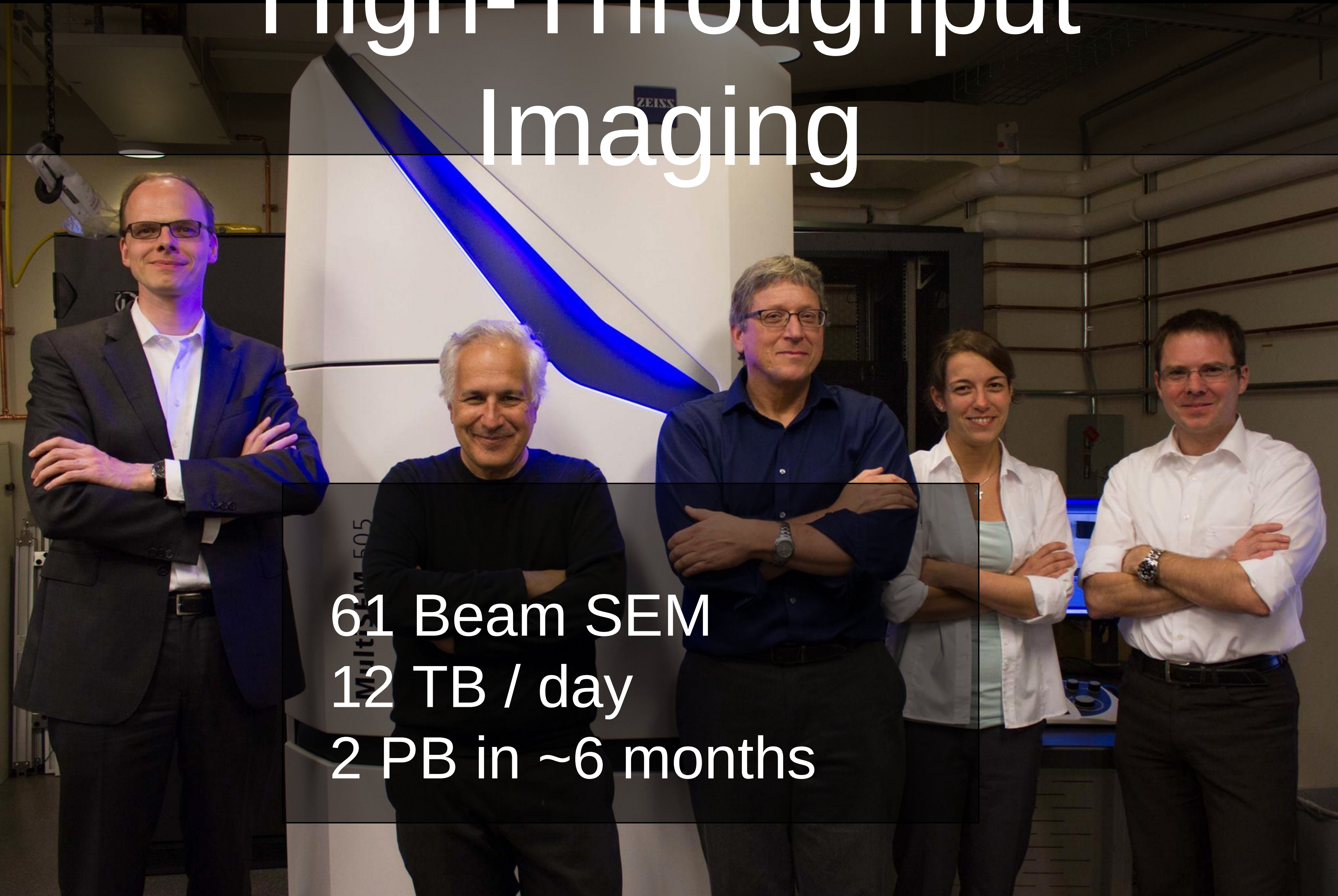
Registratio
n

Segmentati
on &
Synapse
Detection

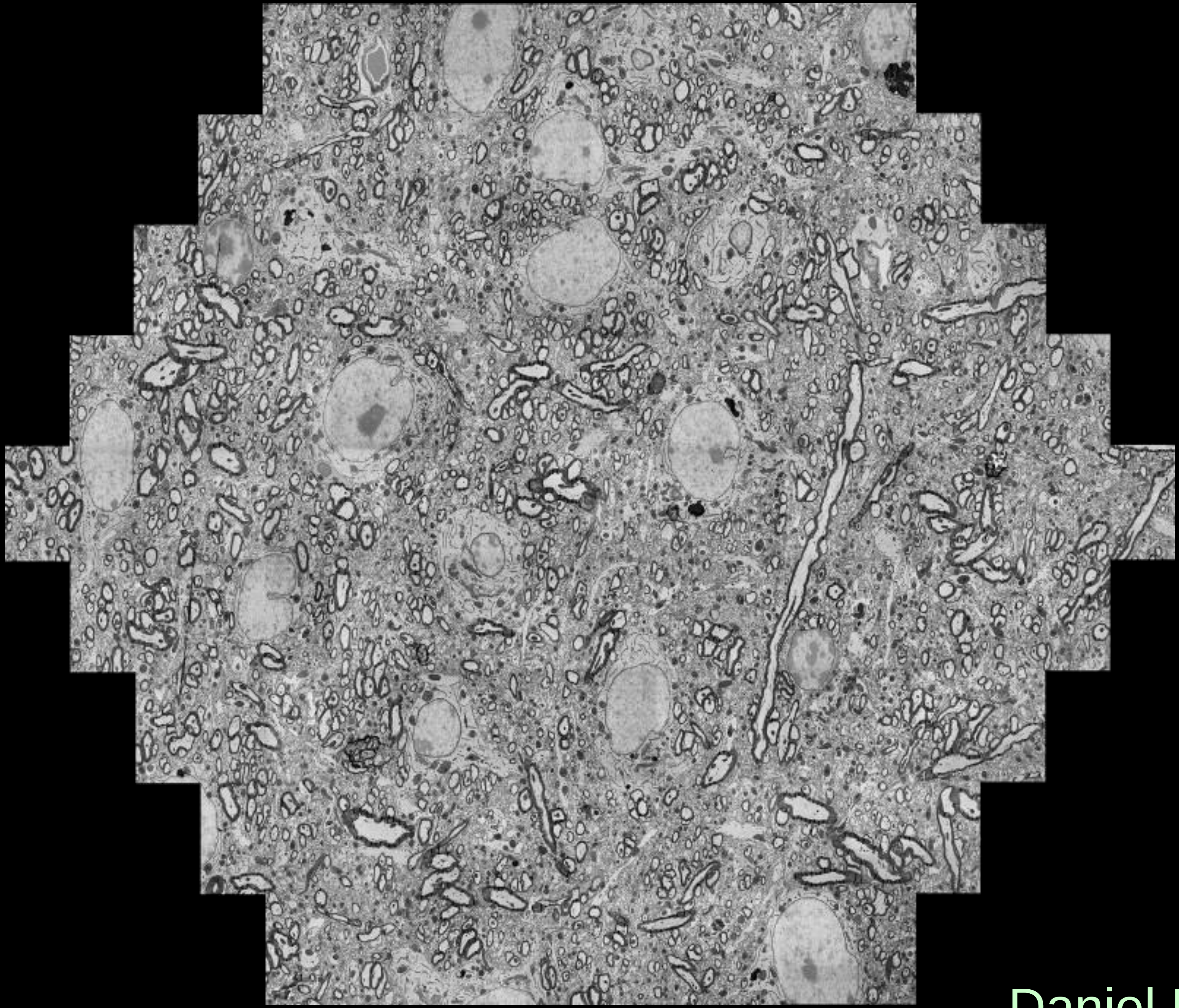


Visualizati
on &
Analysis

High-Throughput Imaging

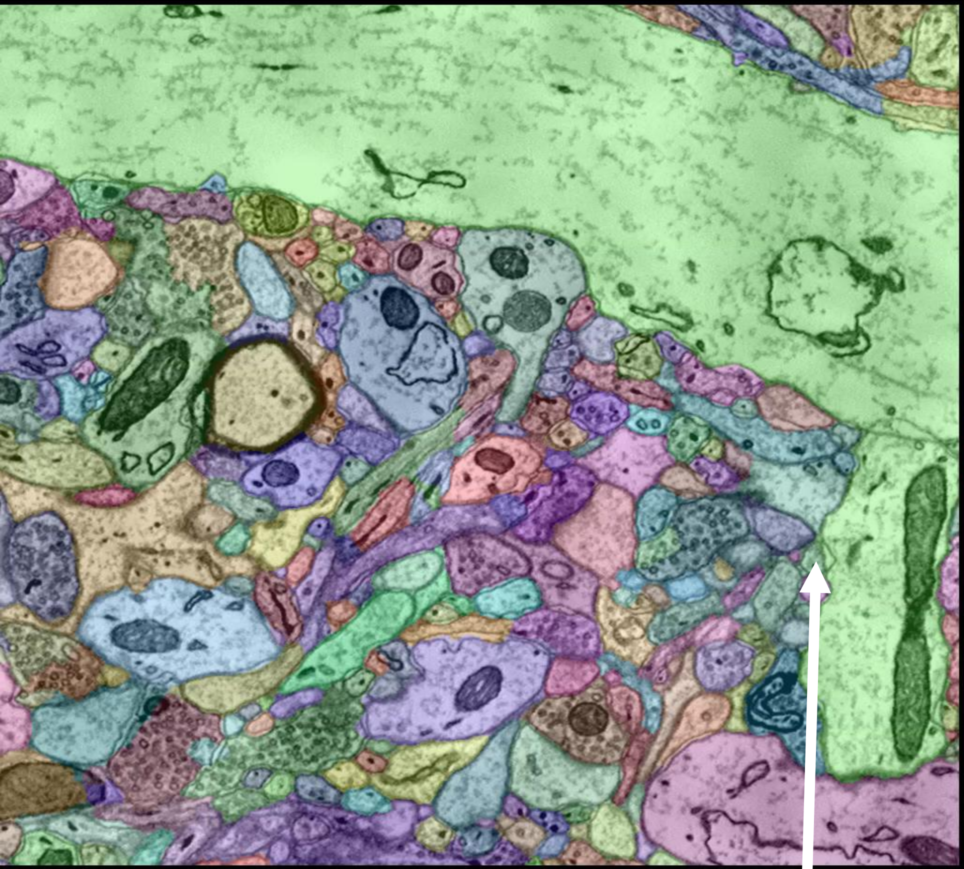


61 Beam SEM
12 TB / day
2 PB in ~6 months

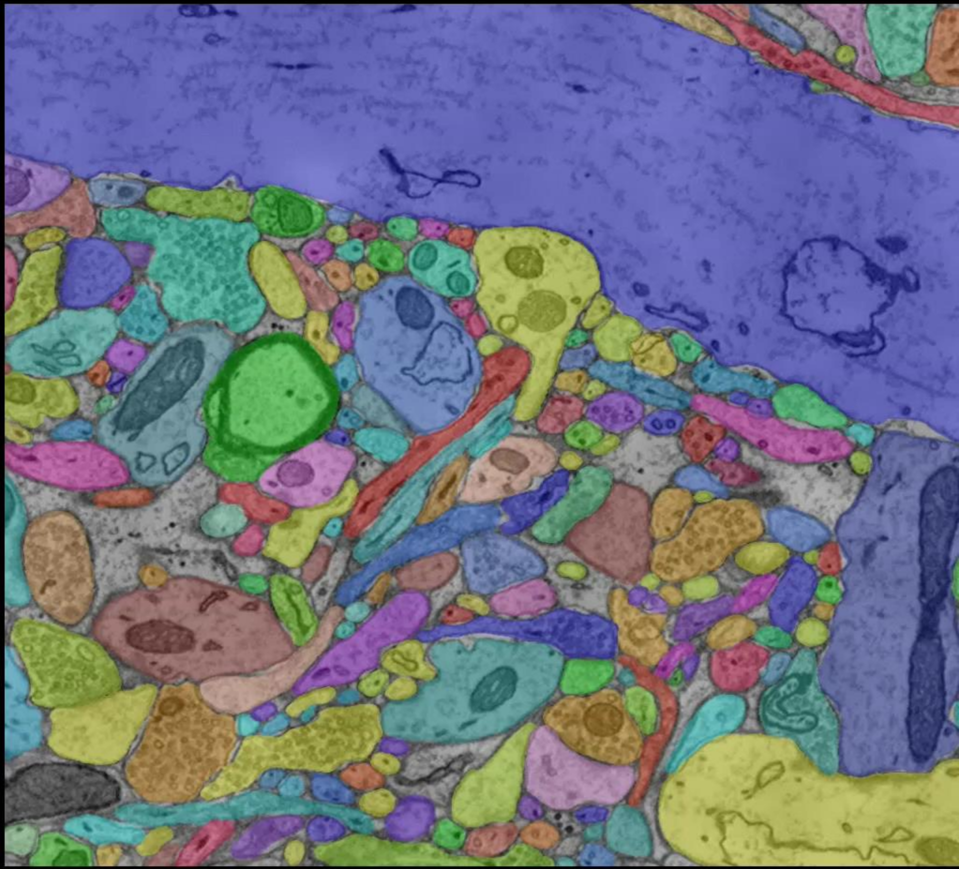


Daniel Haehn

Fully automatic (Rhoana)

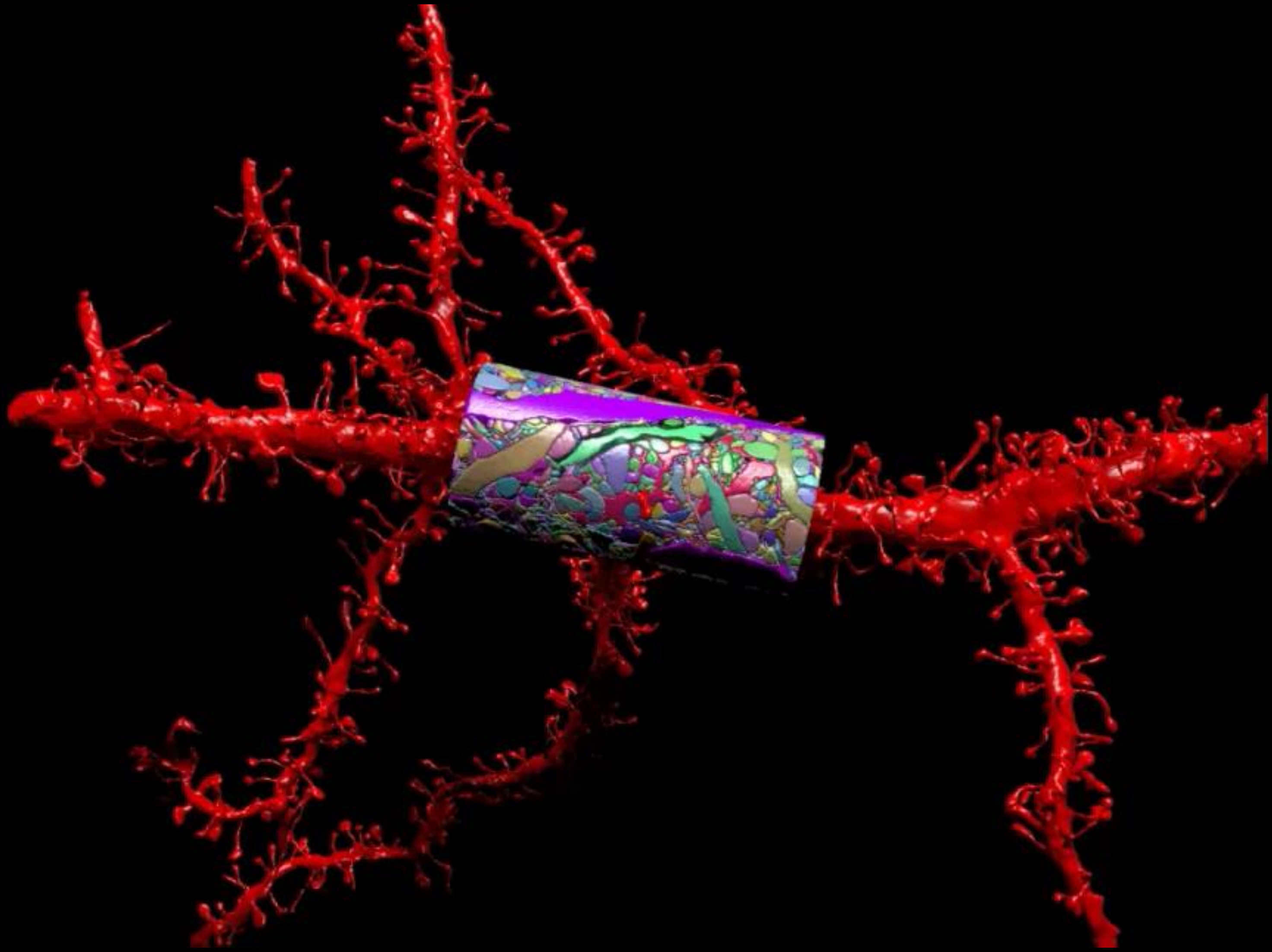


Hand segmentation (VAST)

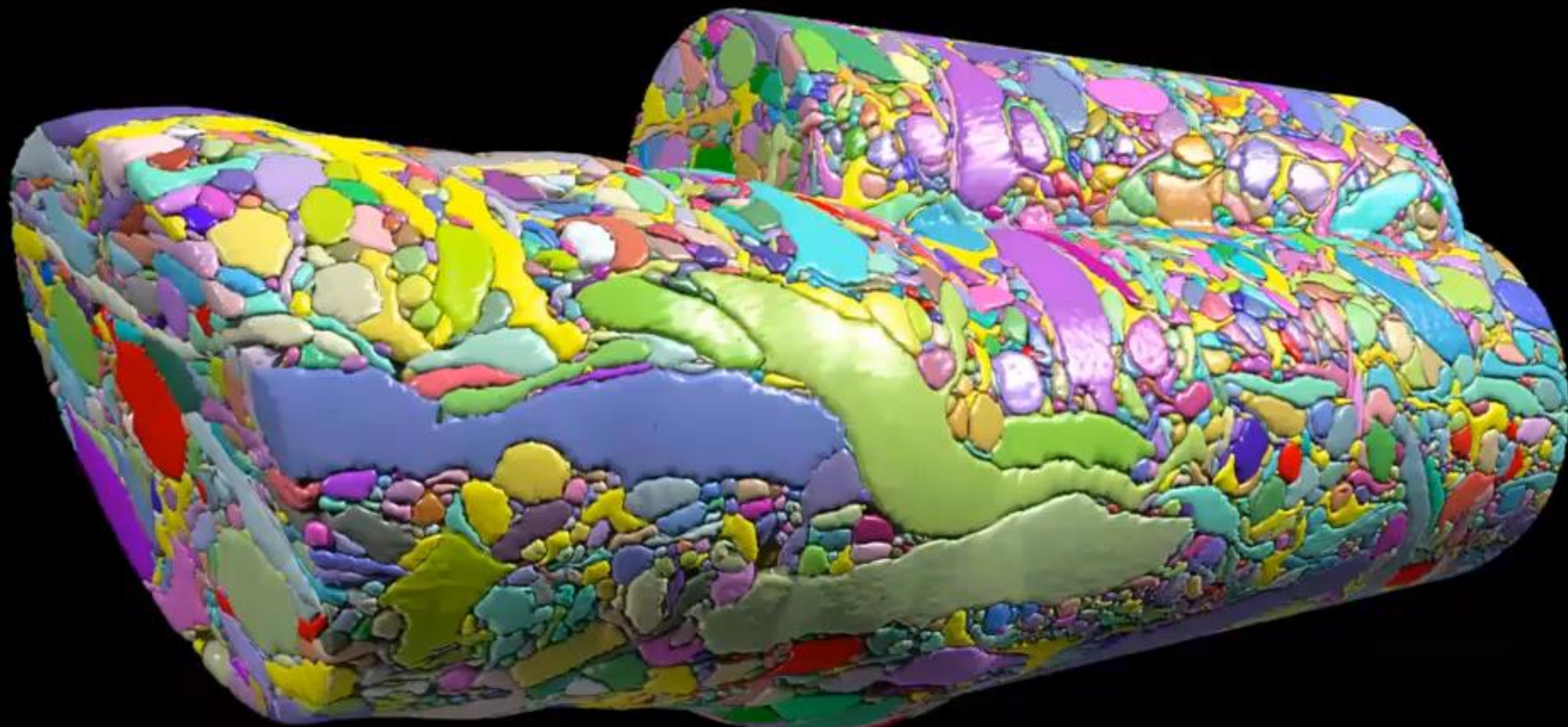


Kasthuri et al., Cell 2015
Knowles-Barley et al.

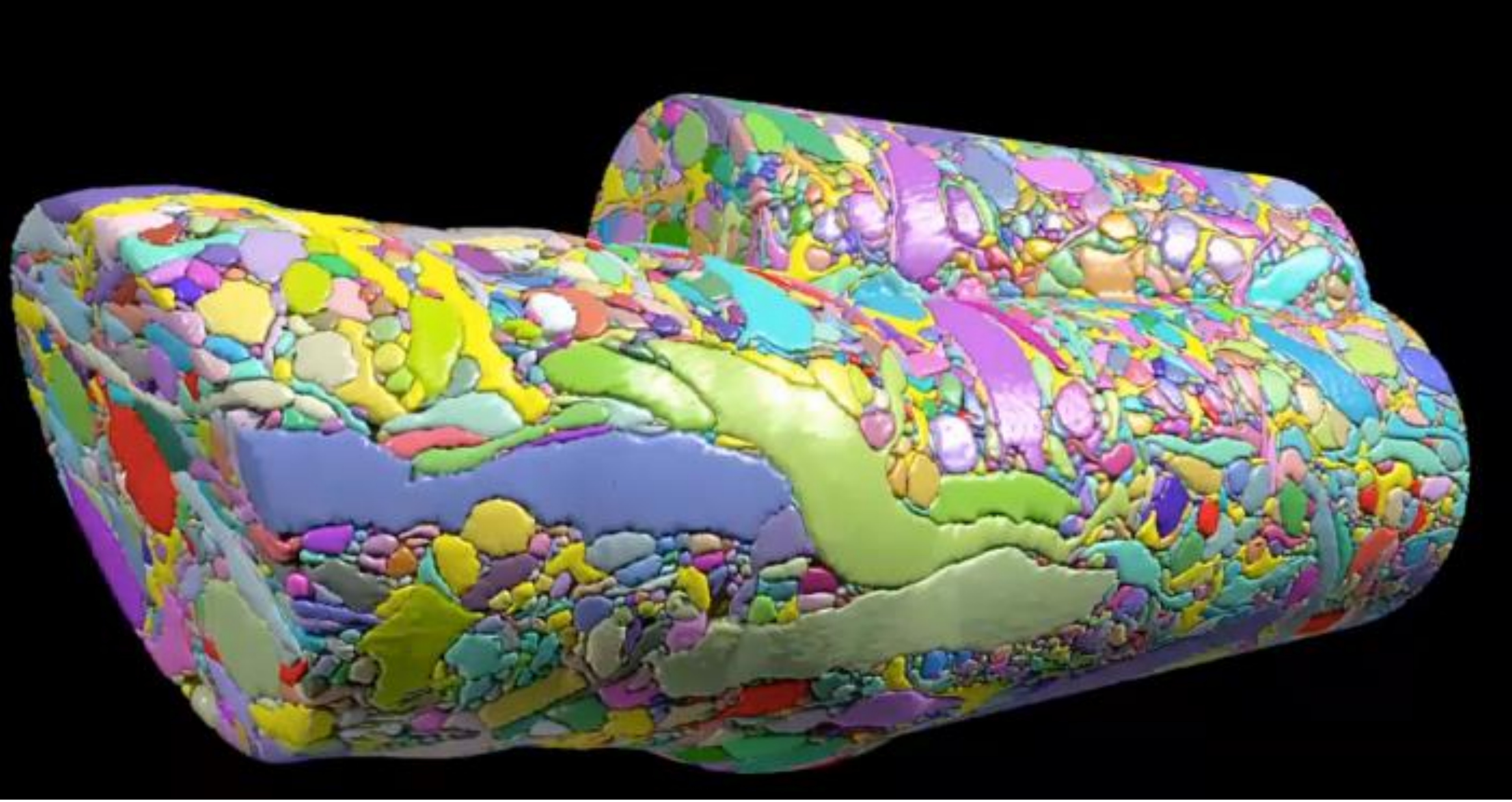
Progress on the Connectome



Bobby Kasthuri, et. al. Argonne, Uchicago and Harvard



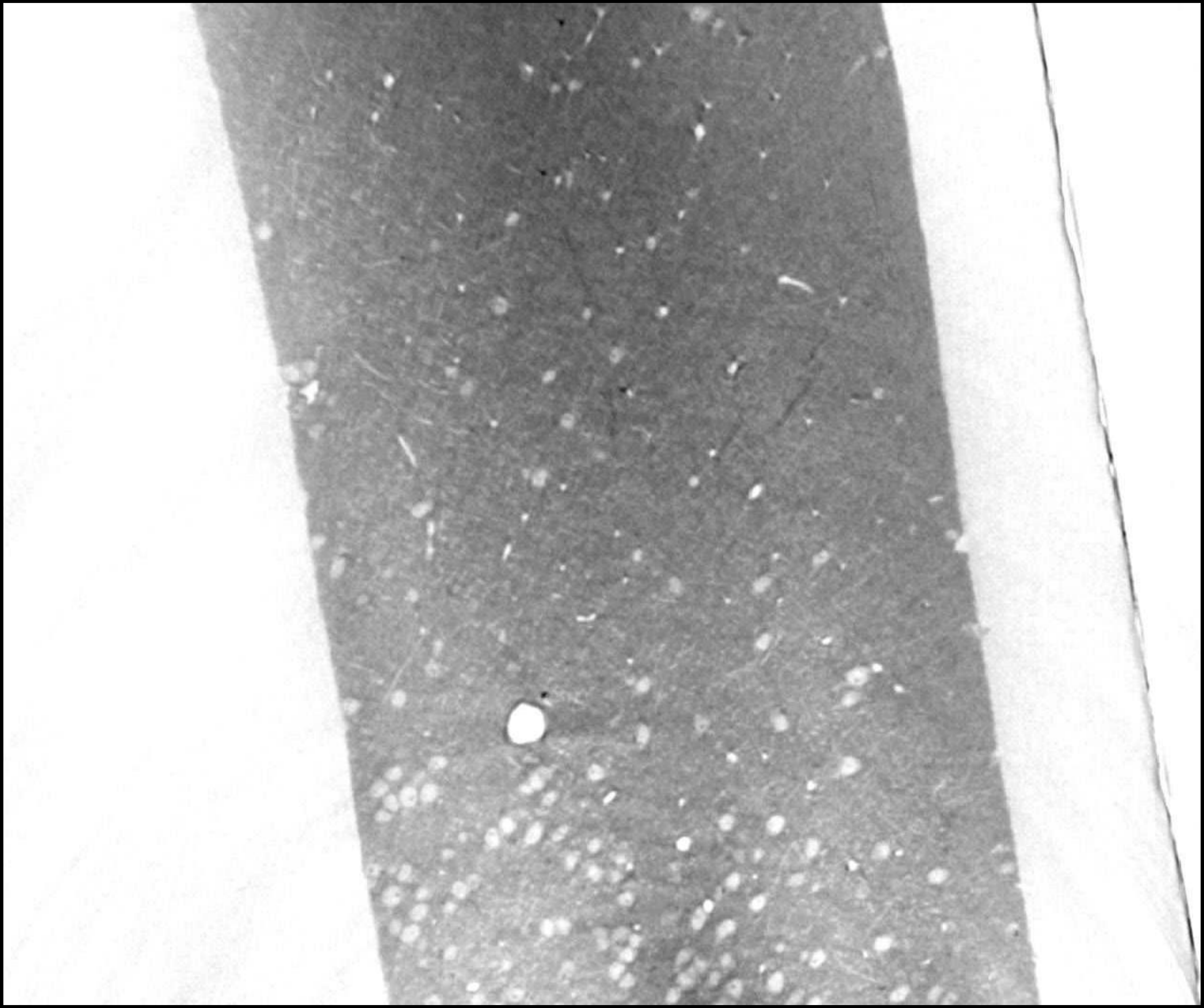
Kasthuri et al., Cell 2015



2 person-years
1500 μm^3
1/666,666th of 1 mm^3

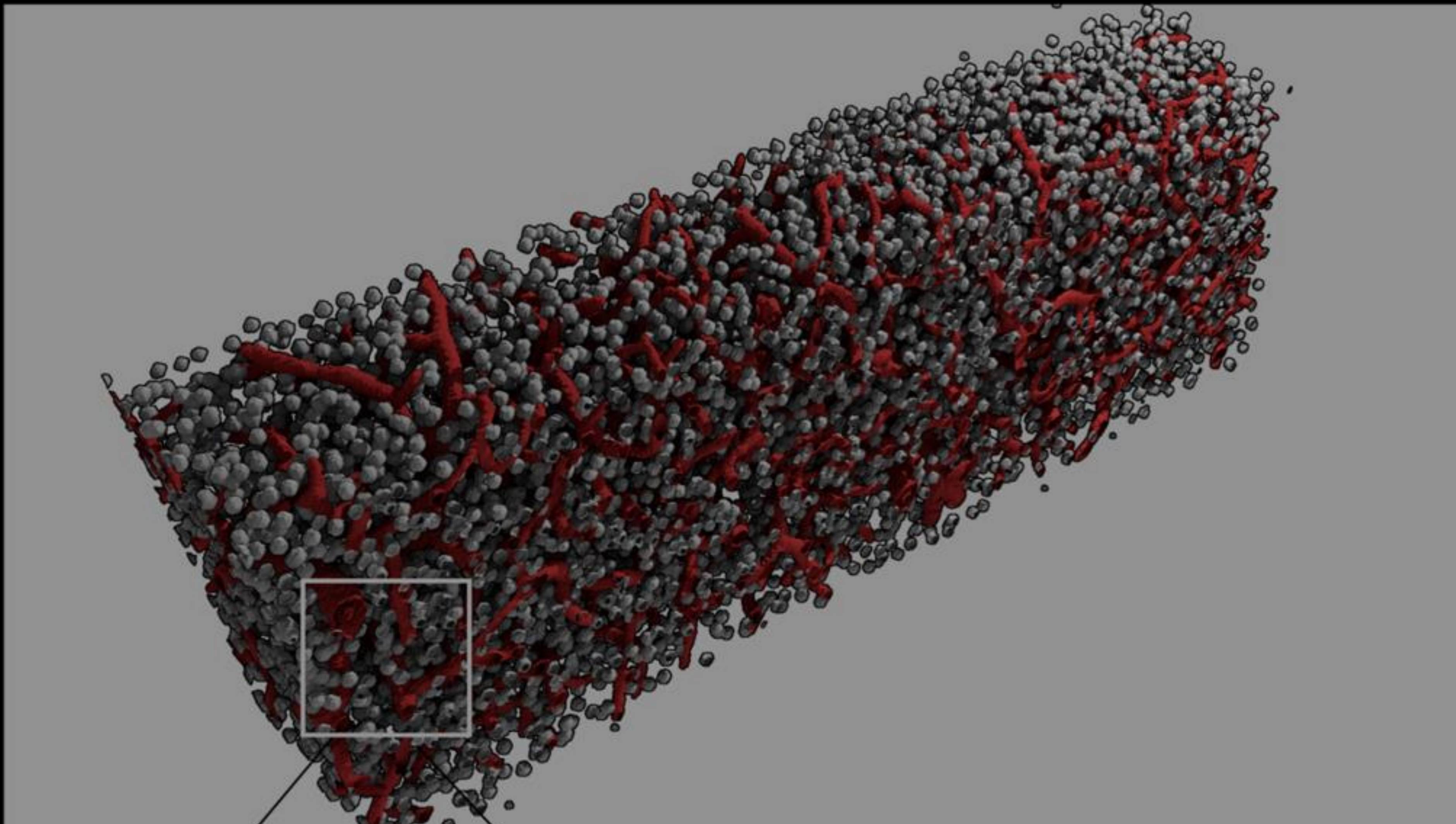
Kasthuri et al., Cell 2015

APS and X-Rays for Connectome

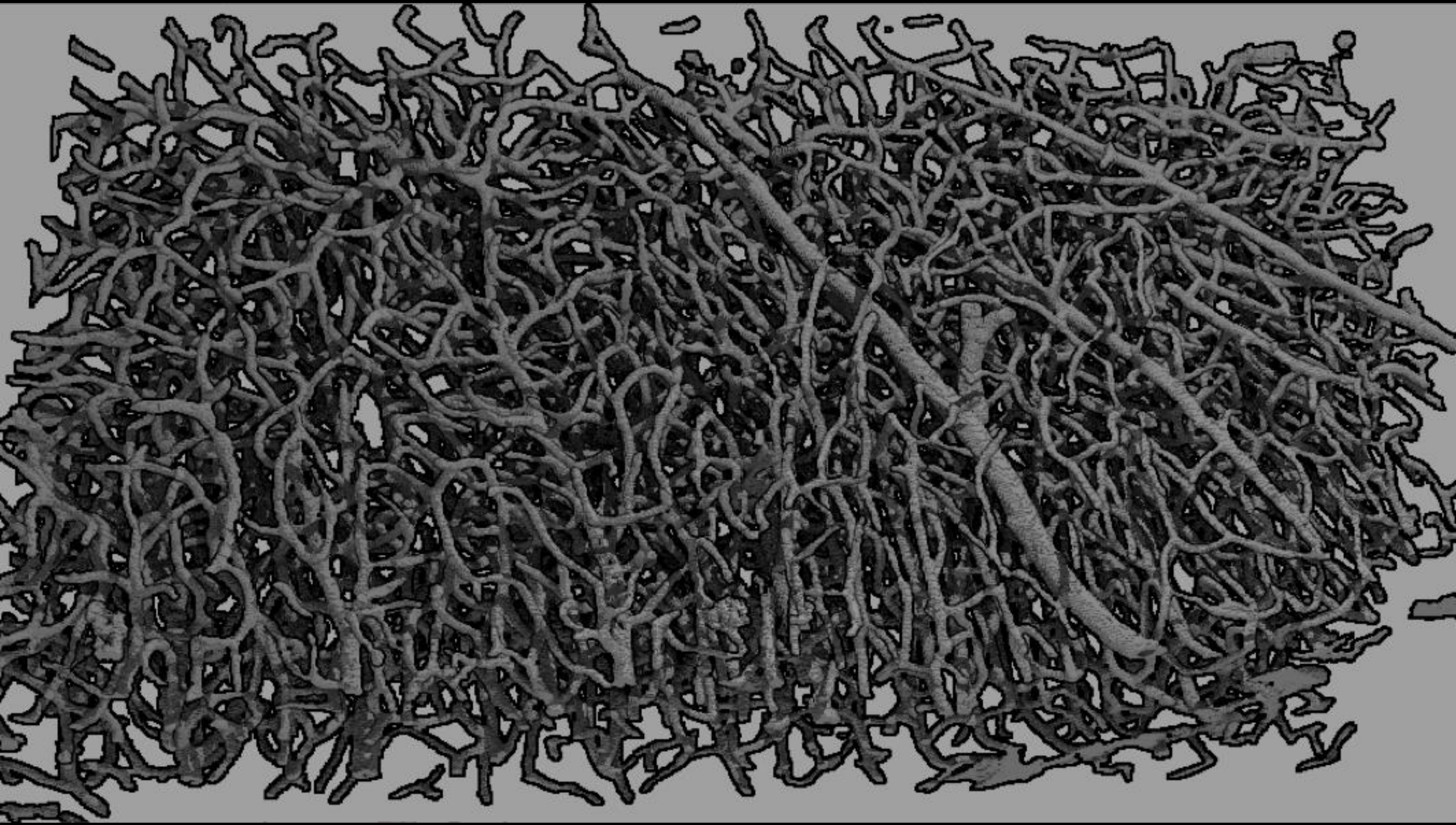


Kasthuri et. al. Argonne

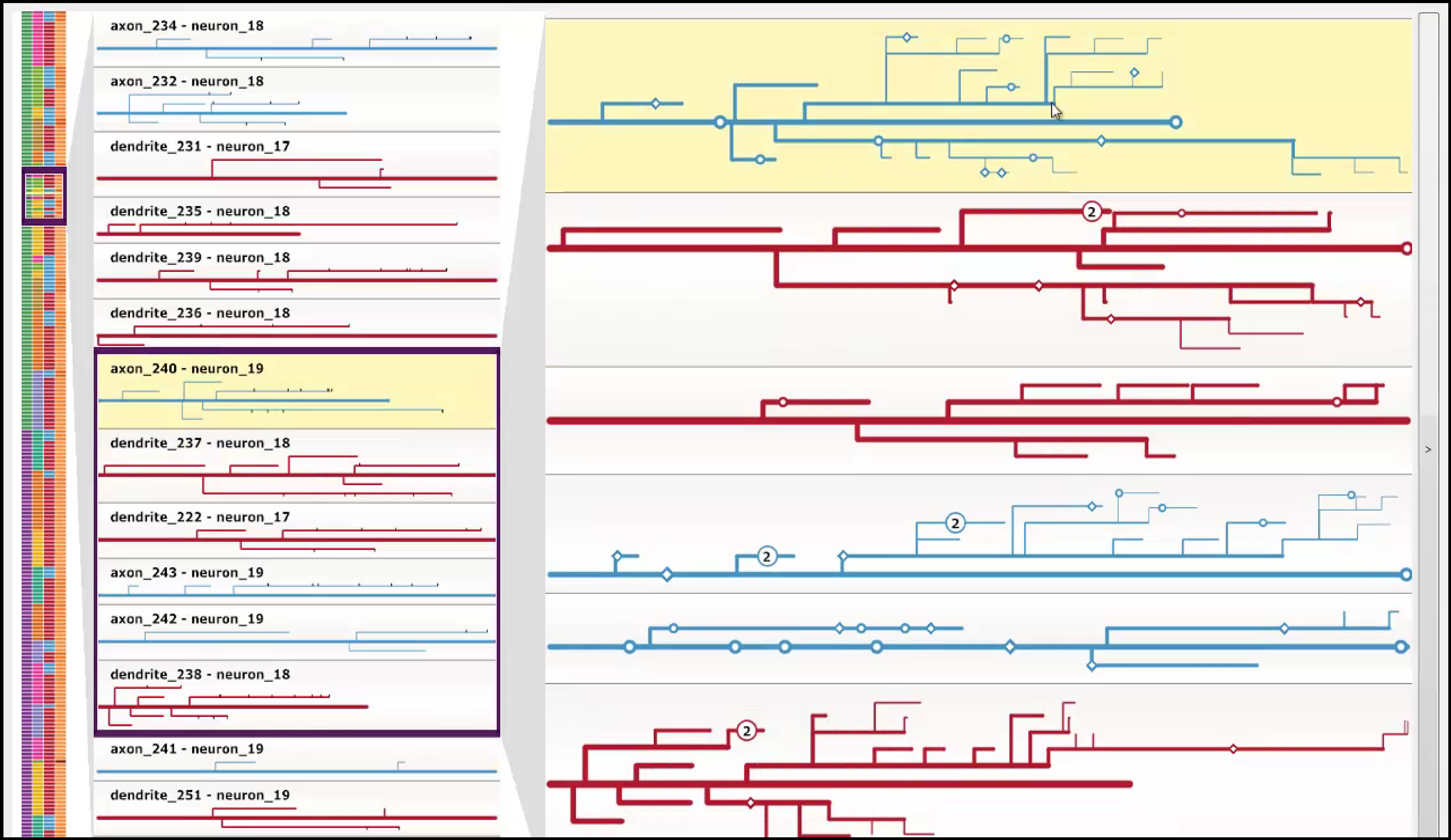
In-situ Reconstruction via X-ray Tomography



In-situ Reconstruction via X-ray Tomography

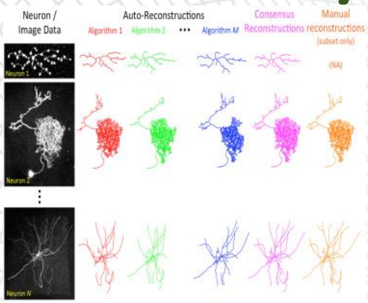


NeuroLines –
Neuronal Connectivity Analysis



Al-Awami et al., TVCG 2014

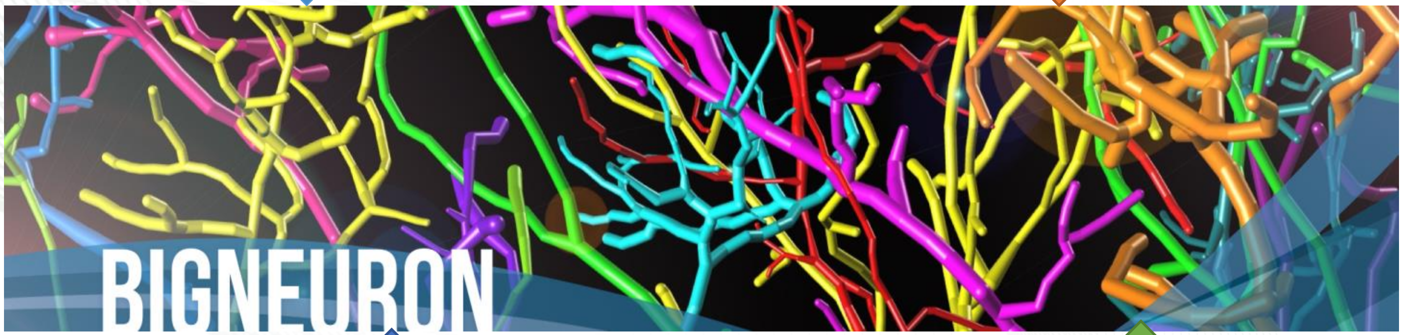
BIGNEURON: Automated Image Reconstruction for Large-scale Phenotyping of Neuron Morphologies



Example snapshot of automated neuronal reconstruction algorithms

Algorithms (M > 20)
Algorithm porting and data analysis hackathons (4)

Neuron Images (N > 30K) Annotation workshop to establish "gold standard" manual reconstruction



Supercomputing (4 centers) Bench testing of algorithms across all images

Data hosting (1-3 mirror sites)

Nov 2015: Hackathon at EVEREST visualization center

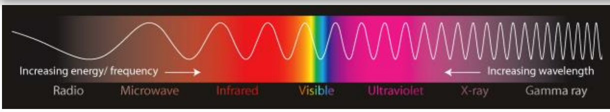
- Integrated data analysis and visualization for petabyte scale datasets on TITAN/EVEREST
  - Hosting about 30,000 neurons + O(200 TB) reconstruction data
- Enable standardization of neuronal reconstruction algorithms at scale
- Interactive portal to visualize and download datasets for users

Contacts: Arvind Ramanathan (ORNL) & Hanchuan Peng (Allen Institute)
ramanathana@ornl.gov & hanchuan.peng@gmail.com

OAK RIDGE National Laboratory

https://www.ornl.gov/news/digitizing-neurons

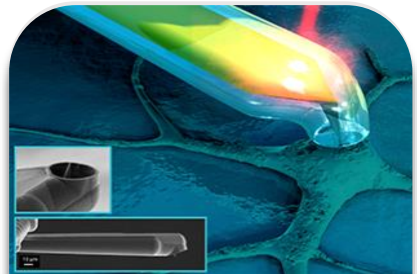
Multi-modal tools for interrogating **brain function**



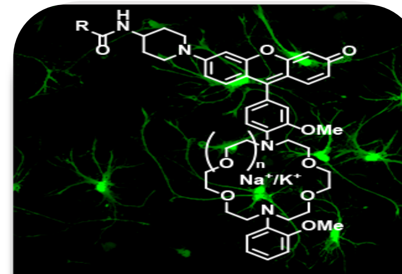
Optimize signals for specific questions

Next-generation optical technologies

UpConverting Nanoparticles

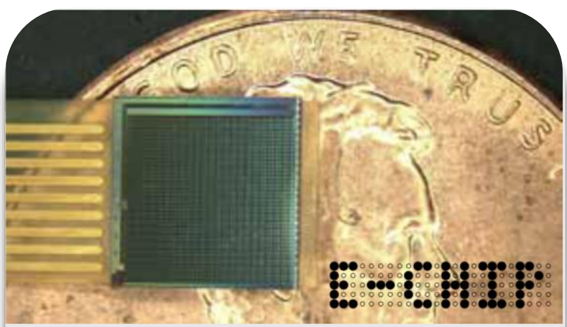


Opto-acoustic Approaches



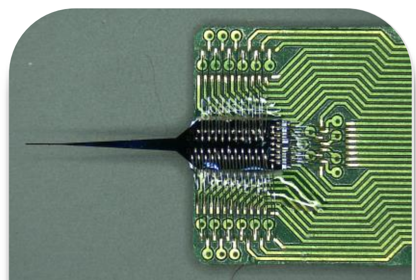
Fast Ionic Sensors

High-channel count electrical recordings

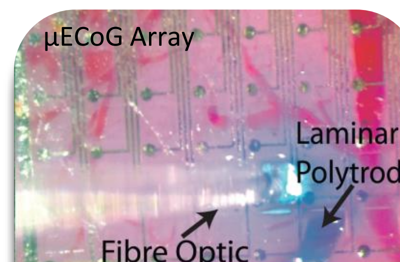


High Channel-Count Electrocorticography

High-throughput Data acquisition and FPGA

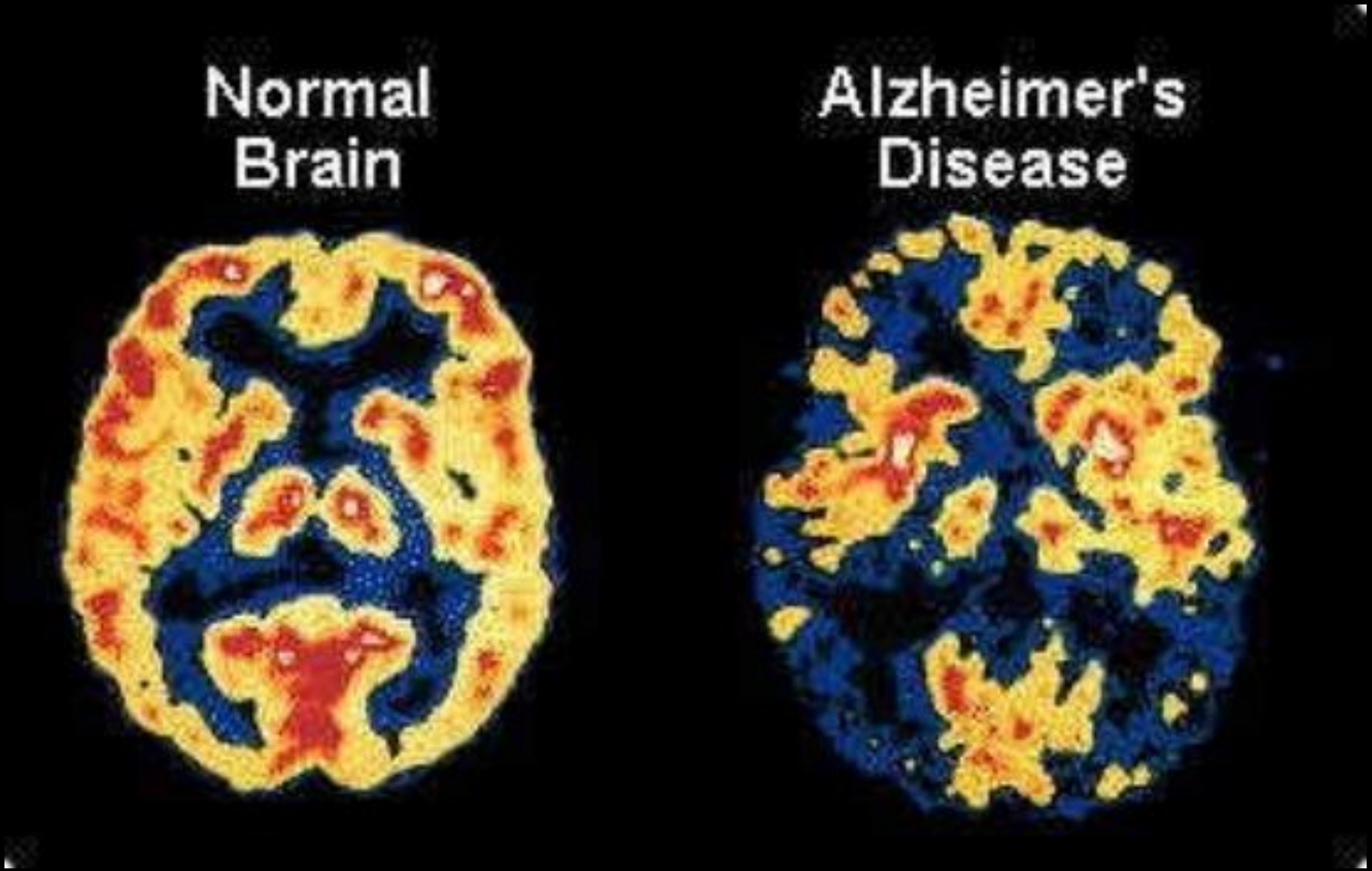


Novel Probes (e-beam litho)



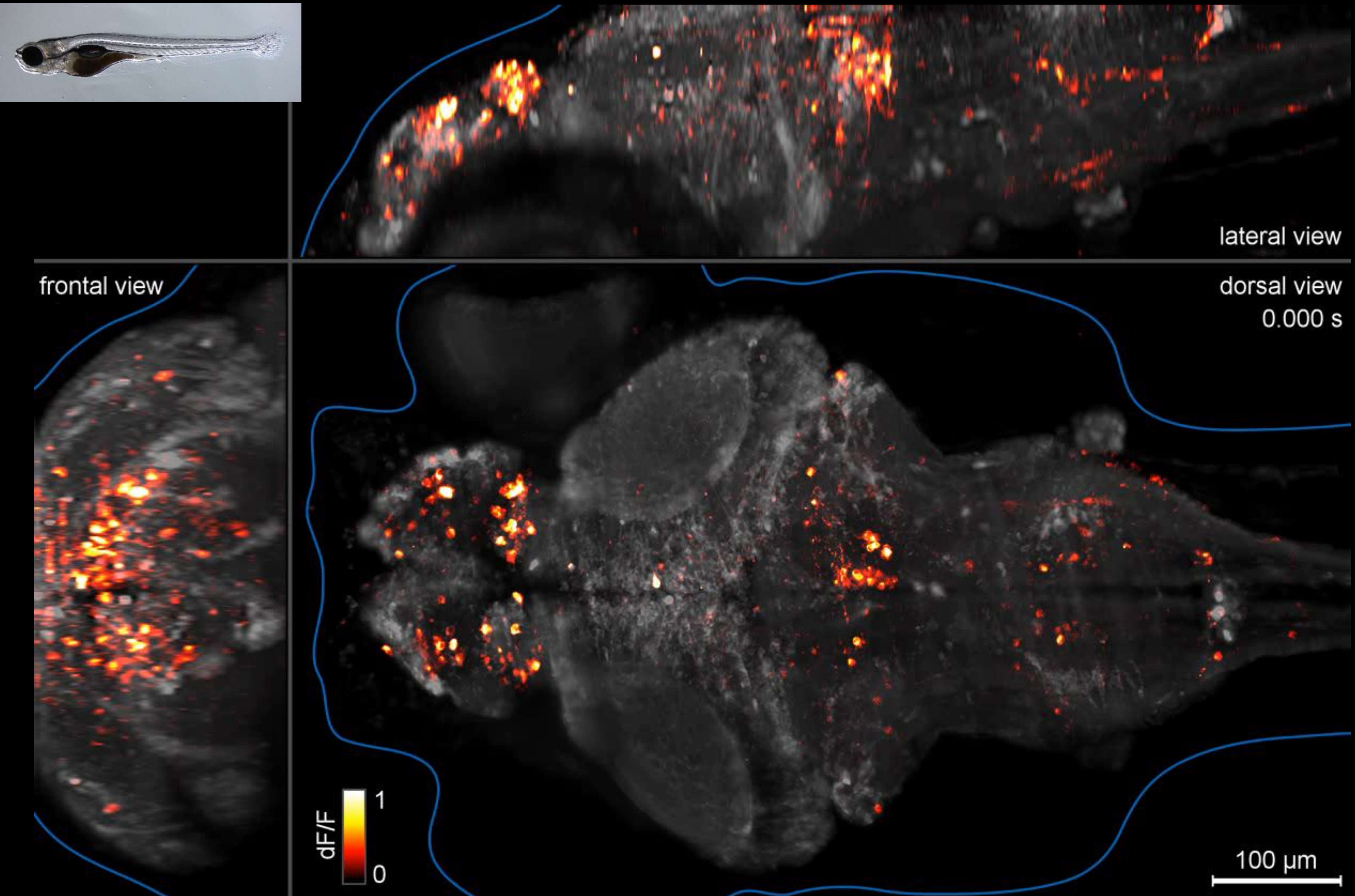
Neural Engineering Lab

Critical disease processes take place at scales that simply cannot be seen, even by today's best tools



PET scan showing activity in the brain of an Alzheimer's patient, National Institute on Aging

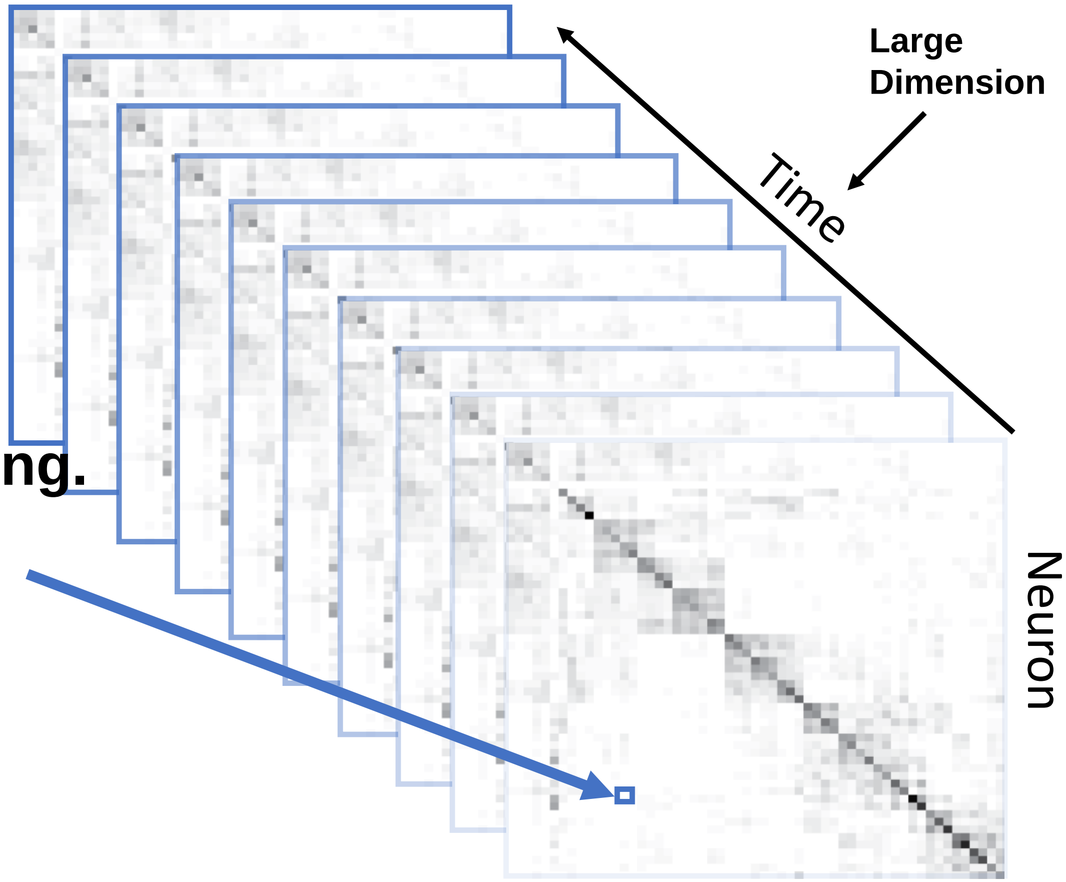
In vivo: Optogenetics allows visualisation of neural activity

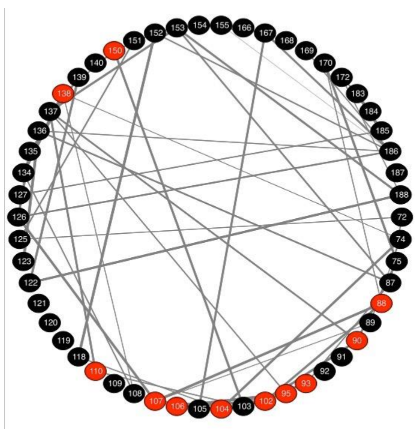


The Functional Connectome:
structure of brain functions

A weighted, direct graph describing the dynamic, casual interactions amongst neurons in the functioning brain.

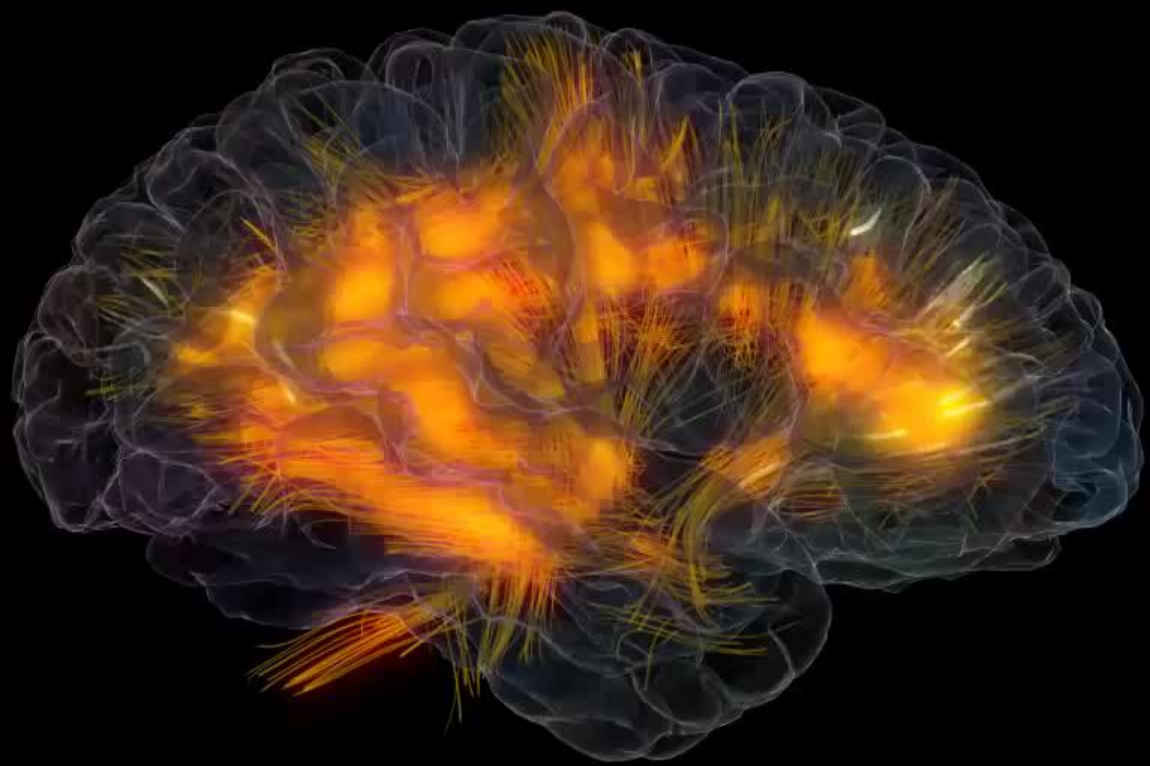
(e.g.) Each edge is **estimated** from data using **machine learning.**





Functional network from human electrophysiology derived using LBNL developed algorithms

GlassBrain — Visualization

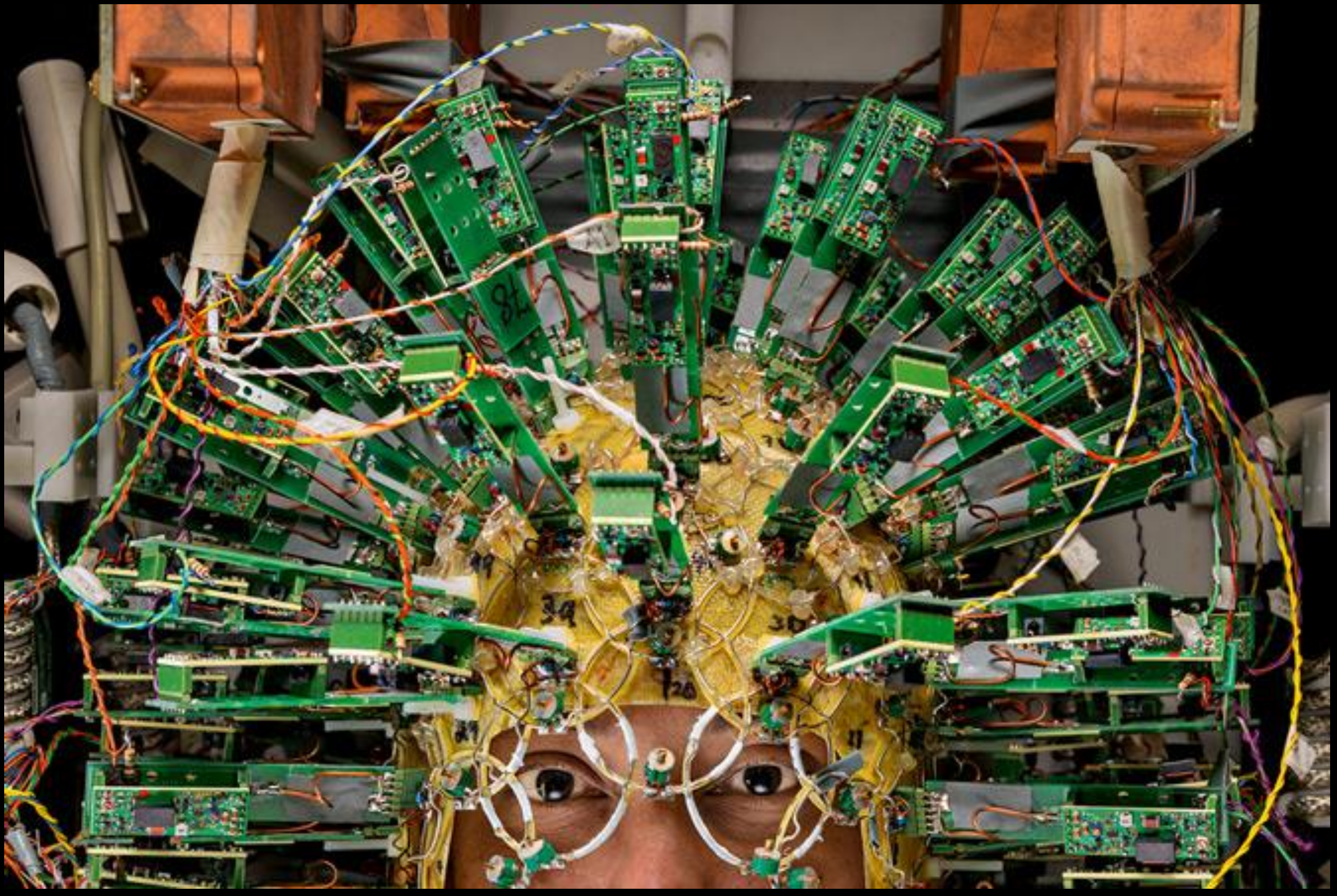


EEG powered by BCILAB | SIFT

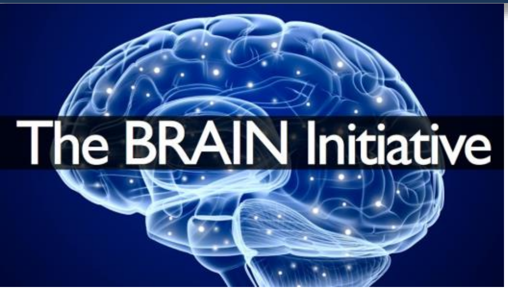
EEG (color indicates "power" and frequency, MRI (Diffusion Tensor Imaging) for structure

Adam Gazzeley Lab - UCSF

Brain-mapping tools



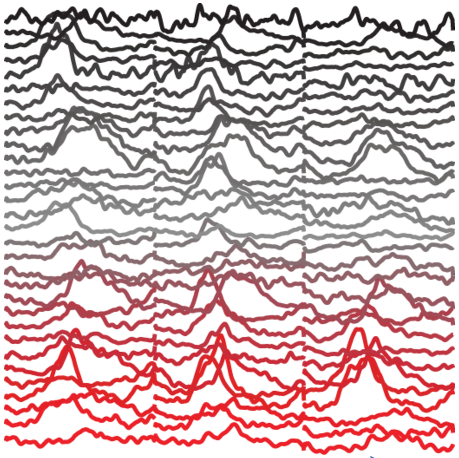
ASCR can uniquely contribute to BRAIN



ASCR can play a unique role in BRAIN computing through advances in applied mathematics and computer science together with HPC facilities.

Function

dynamic data

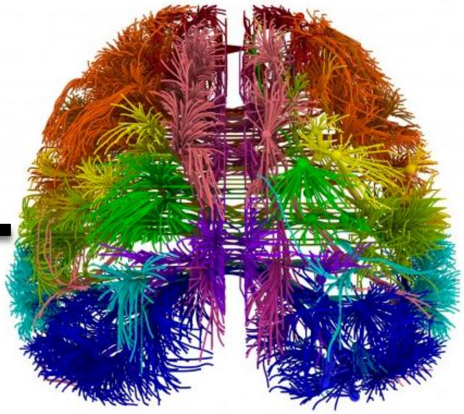


Theory & Models

abstractions

Structure

static data



Generation and analysis of raw data

Linking structure to function is a 'grand challenge' in general biology and materials

Top Level Computing Opportunities

- Large-scale Data Analysis
  - Reverse engineering (e.g. BRAIN, microbiome)
  - Searching, diagnostics and sensors (e.g. id, amr)
  - Data integration and bioinformatics (e.g. amr, Cancer)
- Large-scale Predictive (statistical) Modeling
  - Predictive Tools (PMI, cancer, public health, amr)
  - Hypothesis Formation (e.g. id, amr, cancer)
- Large-scale Explainatory (mechanistic) Modeling
  - Molecular Interactions (molecules, pathways)
  - Physiological Modeling (cell, organ, organism)
  - Cellular Populations (brain, id, cancer, evo-devo, etc.)

Labs are Particularly Good at

- Building flexible teams that cross disciplines and cross laboratories
- Sustained technology development needed to reach a goal and involves partners (vendors, uni, labs, etc.)
- Large-scale project management
- Production quality software development and software engineering (code teams)
- Building user communities around new scientific capabilities (facilities, and online services)
- Integrating across multiple domains and facilities

DOE NIH Interaction Models

- NIH can and does tap into labs via University Grants and Contracts (extramural)
  - Lab PIs with Joint appointments
- NIH supported PIs are users of lab user facilities
  - Light sources, EMSL, JGI, LCFs, Nanoscience, etc.
  - Software and materials sharing
- Direct funding/hosting arrangements
  - Structural Biology at some labs (APS, …)
- Agency-to-Agency arrangements
  - DOE-NCI Pilots?, BRAIN?, ….

What Might be Needed for the Future

- A flexible framework for larger scale partnerships that tap into the capabilities and culture of the labs in a more direct fashion
  - Partnership projects (Intramural/Laboratory)
    - DOE/NCI/Moonshot, Informatics, Computing?
  - NIH centric facility hosted at Labs
    - National Brain Observatory (e.g. connectome facility)
    - Hosted computers/data infras (e.g. “X Commons”)
  - NIH projects as “plug-ins”
    - Co-Design participant for Exascale Apps
    - Intramural computing partnership with ASCR facilities

Computing is a Great Integrator

- To truly understand something means we can build models that predict future states or outcomes and give us insights or explainations of the behavior of the system
- To do this often requires integration of knowledge from many sources into a coherent computable representation that can be used to test hypotheses and conjectures
- In this sense computing/modeling collaborations often play the role of grand scientific integrators this is true in DOE mission space and also is often true in NIH mission space
- Computing collaborations with DOE could improve how NIH integrates science across domains, institutes and projects

Acknowledgements

Many thanks to DOE, NSF, NIH, DOD, ANL, UC, Moore Foundation, Sloan Foundation, Apple, Microsoft, Cray, Intel and IBM for supporting my research group over the years