

## BD2K CENTERS SIZE UP BRAIN DISEASE

About a year and a half ago, brain imaging researchers at the University of Southern California (USC) were shooting the breeze over salad and roast beef sandwiches when their lunch conversation took a turn. A skinny guy named **Arthur Toga, PhD**, confessed to his colleagues that he takes Lipitor—a common cholesterol-lowering drug prescribed to prevent heart attacks and strokes. Toga's total cholesterol had climbed above 200, prompting his cardiologist to recommend the statin therapy. "He said, 'this stuff should be in water like fluoride—there's no harm to it. Everybody should take it,'" Toga recalls.

Others at the lunch table weren't so sure: Because cholesterol is vital for brain health, they wondered if reducing cholesterol with statin therapies could lead to cognitive problems—perhaps even increase the likelihood of dementia.

Past studies looked for links between statins and Alzheimer's disease risk but were either too small, with sample sizes in the hundreds, or dealt with limited types of data. And their findings were mixed. "Nothing seemed definitive," says Toga, who runs USC's Laboratory of Neuro Imaging and leads BD2K's **Big Data for Discovery Science (BDDS)** Center there.

of the uncertainties of brain research will fade—helping the field more effectively diagnose and treat brain disease.

### More Data Yield More Definitive Results

Toga and USC colleague **Judy Pa, PhD**, decided to tackle the Alzheimer's question with a big-data approach. They put their computers to work combining through clinical and brain-imaging data from more than 2,100 participants enrolled in various studies at 40 research centers. The goal: look for relationships between statin use, brain structure and Alzheimer's disease status. Since the literature on statins and Alzheimer's is murky, says Pa, "we did not know where the results would take us."

The number crunching revealed a surprise: Statin use does appear to raise Alzheimer's risk but only in women. The findings have been submitted for publication.

This is just one example of the types of analyses made possible by the rise of big data. "If you don't have enough data, you can't possibly do something like this," Toga says. Traditionally, researchers start with a hypothesis and then go collect data to see if it supports the idea. But in the realm of big data, "we have the opportunity to not articulate a hypoth-

trends, relationships and other interesting features emerge. Even when researchers come in with certain ideas they hope to test, adds Pa, there are many more questions that can be asked of the data.

Recently BDDS researchers posed a particularly tough question: Using large quantities of complex, heterogeneous data from multiple centers and studies, can computers learn to identify which people have Parkinson's disease?

To find out, the team used data from the Parkinson's Progression Markers Initiative (PPMI). This \$60 million observational study launched in 2010 to find biomarkers for Parkinson's disease, which afflicts about 10 million people worldwide. PPMI has collected data and samples from nearly 1,000 participants—some with Parkinson's, some without—at 33 clinical sites in 11 countries.

The PPMI has gathered many kinds of data in vast quantities, including brain scans; medication histories; genotypes; and exam results reflecting answers to questions such as whether the person has cognitive issues, difficulty smelling, or the ability to pass a finger tapping test. All that information gets codified. Also, because people join the study at different stages of disease, the computer has to learn to assess differences in disease severity.

---

**With enough images and associated data, perhaps some of the uncertainties of brain research will fade—helping the field more effectively diagnose and treat brain disease.**

---

It's a common problem, as such analyses require huge numbers of brain images. Big data offers a solution: With enough images and associated data, perhaps some

of the uncertainties of brain research will fade—helping the field more effectively diagnose and treat brain disease. "Rather, we say to the data collection, tell me about yourself." Then they let machines sort through huge volumes of data and see what

To further complicate matters, the machines need to recognize different notations for the same information. "Somebody might code sex as

‘M’ or ‘F,’ ‘0’ or ‘1,’ ‘man’ or ‘woman,’ or ‘male’ or ‘female.’ A computer has no idea that those are all the same,” Toga says. “You have to teach it.”

The training seemed to work. Several machine-learning methods in the BDDS study—published August 2016 in *PLoS ONE*—correctly classified people as having Parkinson’s or not with greater than 95 percent accuracy, sensitivity and specificity. Previous studies using machine learning and data-mining methods to recognize Parkinson’s reported just 70 to 90 percent sensitivity.

The ultimate goal is to train computers to predict who’s on the verge of Parkinson’s in advance of symptoms, Toga

says, in order to be able to slow disease progression—similar to how doctors nowadays prescribe statins to people with high cholesterol hoping to prevent future heart disease.

### Cracking the Brain’s Structural and Genetic Code

In the field of neuroimaging, researchers are studying brain scans to identify structural features that associate with neurological and psychiatric disorders. The **Enhancing Neuroimaging Genetics through Meta-analysis (ENIGMA)**

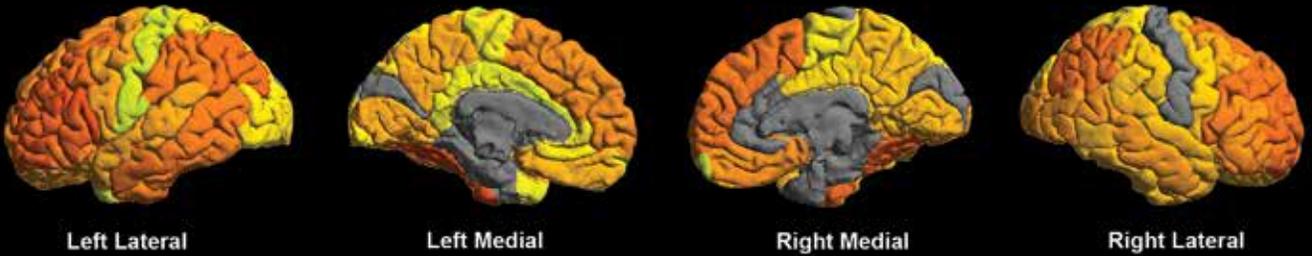
## RELEVANT NIH INSTITUTES:

**NIMH, NINDS, NHLBI, NIA, NIBIB, NIDA, NHGRI, NICHD, and NIAAA**

Consortium goes further and tries to look for the genetic underpinnings of these phenotypes and diseases. Because gene effects tend to be subtle, teasing them out requires huge datasets amassed and analyzed with a global team-science approach. Since its launch in 2009, ENIGMA has rallied more than 800 scientists at 340 institutions in 35 countries to crack the genetic code underlying 18 brain diseases.

## ENIGMA Bipolar Disorder and Major Depressive Disorder Cortical Thickness Findings

1,837 bipolar patients and 2,582 healthy controls



1,902 major depression patients and 7,658 healthy controls



The ENIGMA consortium has published the world’s largest neuroimaging studies of bipolar disorder and major depression. In separate studies, ENIGMA researchers observed significant cortical thinning associated with both illnesses. Bipolar disorder (top) was associated with widespread thinning of frontal, temporal, and parietal regions, whereas major depression (bottom) was associated with thinning

in paralimbic regions. Scientists are now working to quantify the similarities and differences between these psychiatric illnesses thanks to harmonized brain measures derived by the ENIGMA consortium. Image courtesy of Christopher Ching, Paul Thompson, and the ENIGMA Bipolar Disorder and Major Depressive Disorder Working Groups.

Poring through magnetic resonance imaging (MRI) scans, along with corresponding clinical and genetic information from pooled datasets, ENIGMA researchers can analyze cohorts 10 to 30 times larger than a typical neuroimaging study.

In some cases, ENIGMA researchers have found that boosting sample size for MRI-based data is enough to gain insight—even without considering genetics. For example, in a January 2017 study in the *American Journal of Psychiatry*, an international team found that children and adults with obsessive-compulsive disorder (OCD) have distinct patterns of subcortical abnormalities. Whereas smaller brain imaging studies in OCD produced mixed results, the conclusions were clear when the ENIGMA team

from 7,957 healthy people and 2,148 depressed patients at 20 sites around the world. Compared to controls, adults with depression—but not children—had thinner cortical gray matter in the orbitofrontal cortex, anterior and posterior cingulate, insula and temporal lobes. Depressed adolescents had different brain abnormalities—namely, lower surface area in frontal regions as well as primary and higher-order visual, somatosensory and motor regions. The large sample size allowed the researchers to distinguish effects in children versus adults, suggesting that depression correlates with brain structure distinctly during different stages of life.

Several recent ENIGMA papers focus more squarely on identifying gene variants that underlie fundamental

“Rather than just download the data... you have a community to help you really dig into a question,” says **Paul Thompson, PhD**, professor of neurology at USC and principal investigator for the ENIGMA BD2K Center.

He compares the situation to wanting to become better at chess. “Let’s say someone says, ‘I really want to be a world-class chess player. I’ve bought all the pieces.’ In fact, my home is full of chess pieces,” Thompson says. But to improve at chess, “I would say they need to be with people who are really active and playing a lot of chess. Really, what’s going to take the science to the next level is working with a large team of experts. The data is a requirement but not the clincher.”

Researchers can propose new stud-

---

**“Really, what’s going to take the science to the next level is working with a large team of experts,” Thompson says. “The data is a requirement but not the clincher.”**

---

pooled 35 sets of structural brain scans from 1,759 healthy controls and 1,830 OCD patients—about a sixth of whom were under age 18. Compared with healthy peers, children with OCD had a larger thalamus, a brain area important for sleep, consciousness and higher-order brain processing. However, in adults with OCD, greater volumes were measured in other brain regions—namely, the hippocampus and the pallidum, an area important for motivating rewards and incentives. These results are in line with the developmental nature of OCD and suggest that further research on neuroplasticity—the brain’s ability to reorganize and form new neural connections throughout life—could be useful.

Combining datasets, as well as separating children and adult subgroups, also proved important in a May 2016 *Molecular Psychiatry* study that revealed cortical differences in people with depression. The analysis pooled MRI scans

brain features and specific diseases. An international team undertook a massive study of more than 32,000 adults at 52 sites. In a paper published October 2016 in *Nature Neuroscience*, the researchers reported identifying seven genes that not only regulate brain volume, memory and reasoning but also seem to influence Parkinson’s disease risk. And in study of people with schizophrenia, ENIGMA scientists found that certain measures of volume and thickness in affected brain regions correlate with gene variants known to confer disease risk. They also found that schizophrenia shares some of these neurogenetic signatures with other psychiatric disorders. These findings appeared October 2016 in *Molecular Psychiatry*.

And it’s not just about pooling data. Each ENIGMA analysis gets vetted by one of 30 working groups—teams of neuroscientists, imagers, geneticists, methods developers and others devoted to a specific disease or subfield of study.

ies by joining the monthly phone calls held by each working group. The calls update members on the group’s ongoing projects and offer a chance for people with new ideas to thrash them out.

### A Networked Brain: Discovering Causal Relationships

Beyond structure and genetics, the brain can also be viewed as a network.

Another BD2K Center—the **Center for Causal Modeling & Discovery of Biomedical Knowledge from Big Data (CCD)**—renders big data as networks. And it connects the network’s nodes, or variables, not with mere lines but arrows. “Our business is computer algorithms that will find causal relations from measured data,” says **Clark Glymour, PhD**, a professor of philosophy at Carnegie Mellon University (CMU) who leads the CCD group focused on the brain.

The basic algorithm was developed

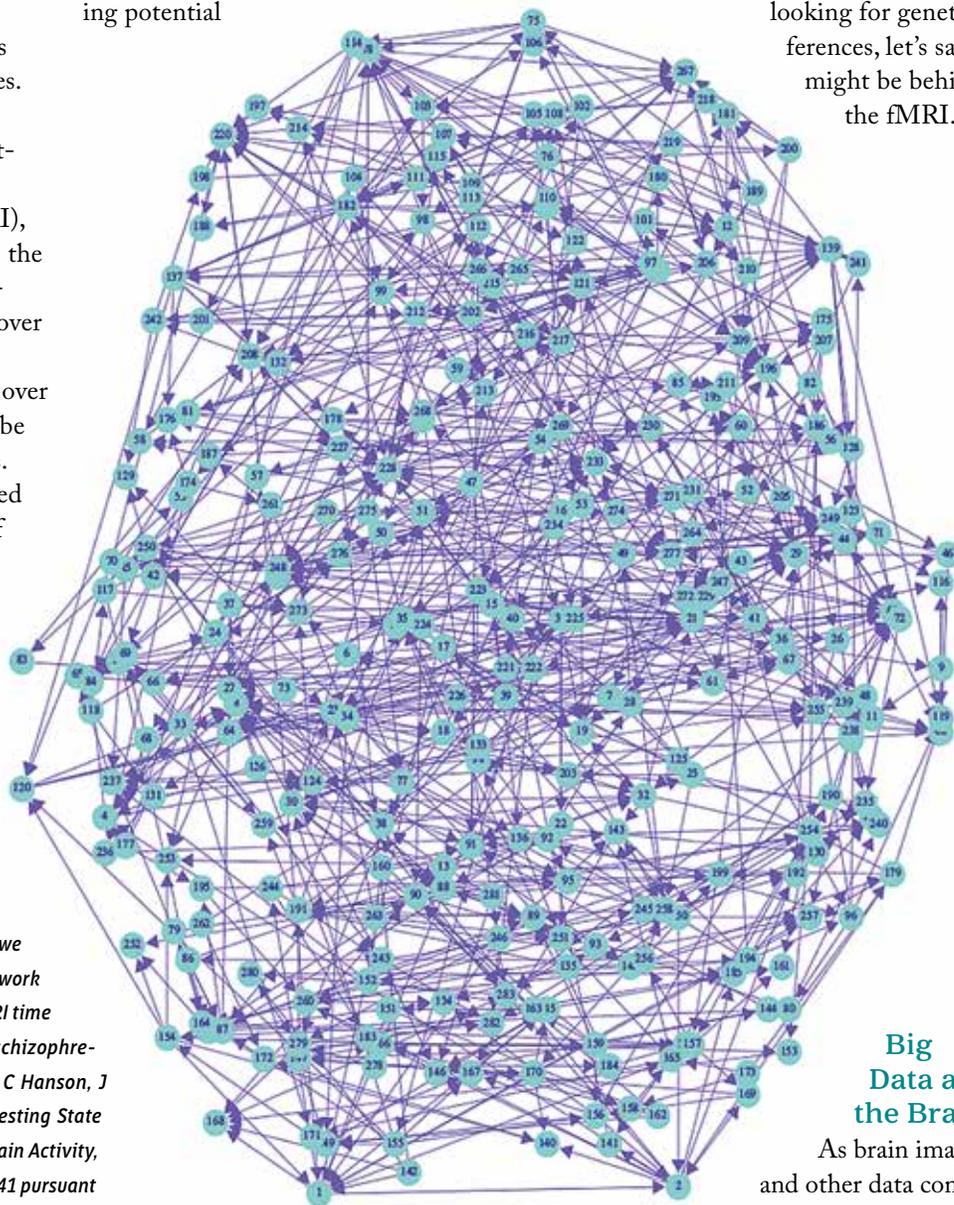
by a CMU graduate student. It could handle 15 to 20 variables—features that take different numerical values over time. About five years ago, with careful programming Glymour’s team got it to run on a few hundred variables. And with further improvements last year the algorithm, called Fast Greedy Equivalence Search (FGES), now runs in about 12 hours on a million variables.

As one test case, CCD decided to apply the FGES algorithm to the resting state brain. Much of their work uses data from functional MRI (fMRI), which approximates neural activity as the amount of energy consumed by thousands of tiny subregions of the brain over time. Generally fMRI produces a full image of the brain every few seconds over a 15 to 20 minute period. And it can be easily performed on many individuals. Highly scalable algorithms are required to make sense of the large quantity of data produced by a typical fMRI-

*CCD researchers used functional MRI (fMRI) data to build causal networks that depict different patterns of brain connectivity in normal versus autistic and schizophrenic individuals. They found that the networks of neuro-atypical individuals exhibited more fractionated and less controllable dynamics in the resting state brain. They even showed that disease severity was revealed by the global topology of the network itself. Here we see the graphical structure of the pattern of network connectivity in the resting state brain (from fMRI time series) aggregated over 10 individuals with schizophrenia. Reprinted from SJ Hanson, D Mastrovito, C Hanson, J Ramsey, C Glymour, Scale-Free Exponents of Resting State are Biomarkers of Neuro-Typical and Atypical Brain Activity, biorxiv.org, doi: <https://doi.org/10.1101/068841> pursuant to Creative Commons license CC-BY-NC-ND 4.0.*

based research study. “What the fMRI work does is give us a really, really hard case for making the best algorithms we can,” says Glymour. In a recent study posted to *bioRxiv* in August 2016, CCD researchers analyzed resting-state fMRI data from one healthy adult, 60 people

with autistic spectrum disorder, and 60 with schizophrenia. Applying the FGES algorithm to the reams of scan data produced causal networks that depict different patterns of brain connectivity in normal versus neuro-atypical individuals. But beyond providing potential



## Big Data and the Brain

As brain images and other data continue to accumulate, the tools developed by the BD2K

There are very likely to be multiple causes and types,” says **Greg Cooper, MD, PhD**, professor of biomedical informatics at the University of Pittsburgh and director of CCD. And if robust patterns were to emerge within the autism group, adds Glymour, “you could start looking for genetic differences, let’s say, that might be behind the fMRI.”

diagnostic information—for example, being able to distinguish healthy individuals from people with autism—the connectivity patterns could help researchers sort autism cases into different subgroups. “These conditions are almost certainly not single monolithic diseases.

Centers are setting the standard for high quality computational neuroscience. Using big data to discover how the brain works in health and disease is becoming routine, as researchers address questions raised over roast beef sandwiches from the comfort of a single workstation. □