



National Institutes
of Health

USC Stevens
Neuroimaging
and Informatics
Institute



Institute for
Systems Biology

Let the data lead to the discoveries

What can be done now,
that was not possible before?



BIG DATA *for* DISCOVERY SCIENCE



Arthur W Toga Principal Investigator, BDDS

USC Viterbi
School of Engineering
Information Sciences Institute

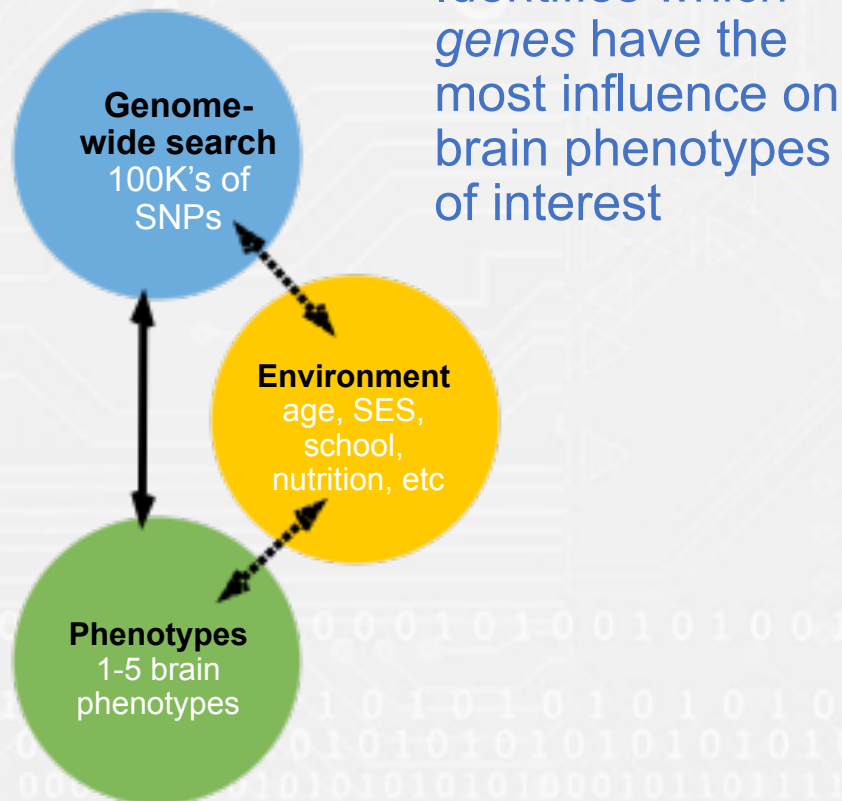
BDDS Platform – Multi-omic data

- Integrated multi-omic data for an end-to-end using standard format for data exchange
- Reusable processes, to clean, integrate, query multi-omic datasets
- Rapid, easy creation of cohorts based on phenotypes of interest
- Rapid, cost-effective, intuitive, reproducible analysis using on-demand cloud computing resources and on-premise HPC resources

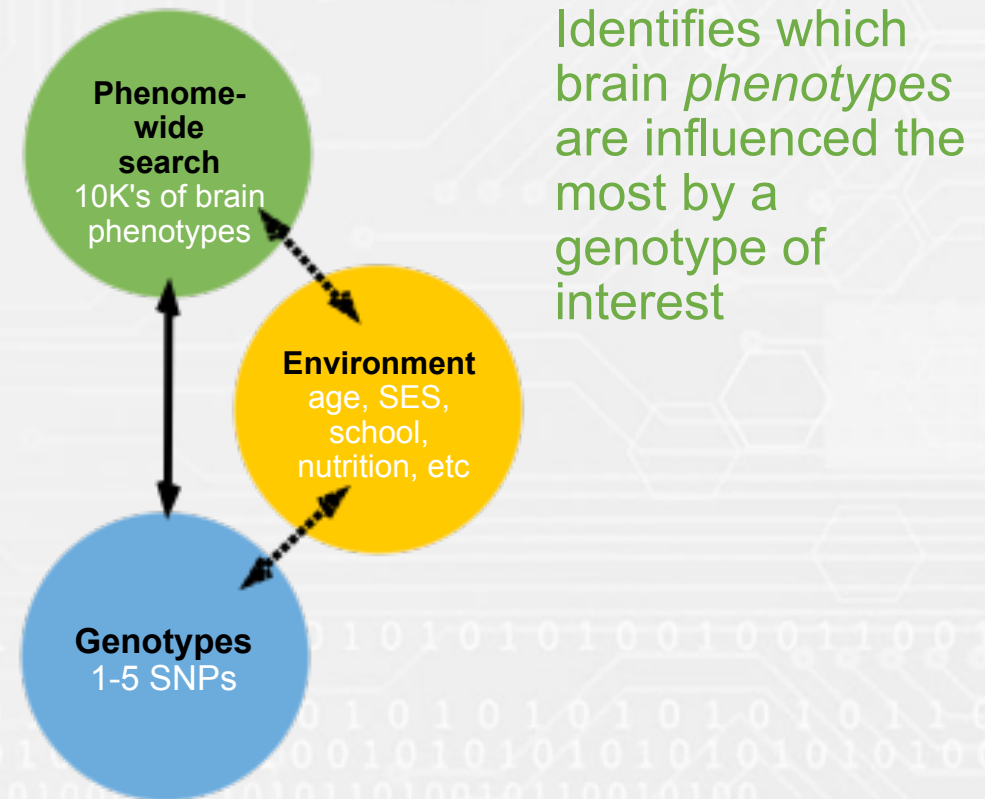
What is PheWAS?

- Phenome-wide study to discover gene-brain associations

GWAS



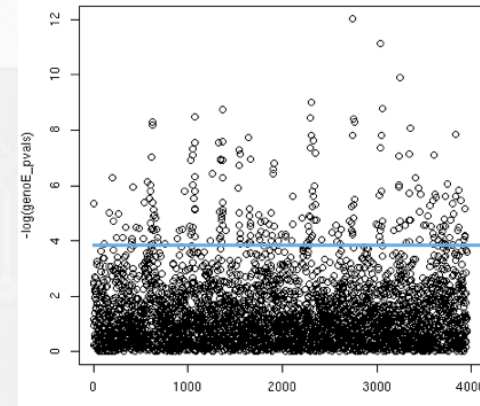
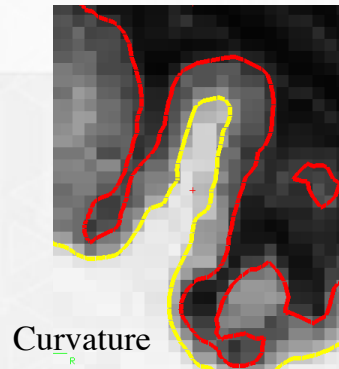
PheWAS



BDDS Platform: Neuroimaging PheWAS

Neureglin-1 (rs35753505)

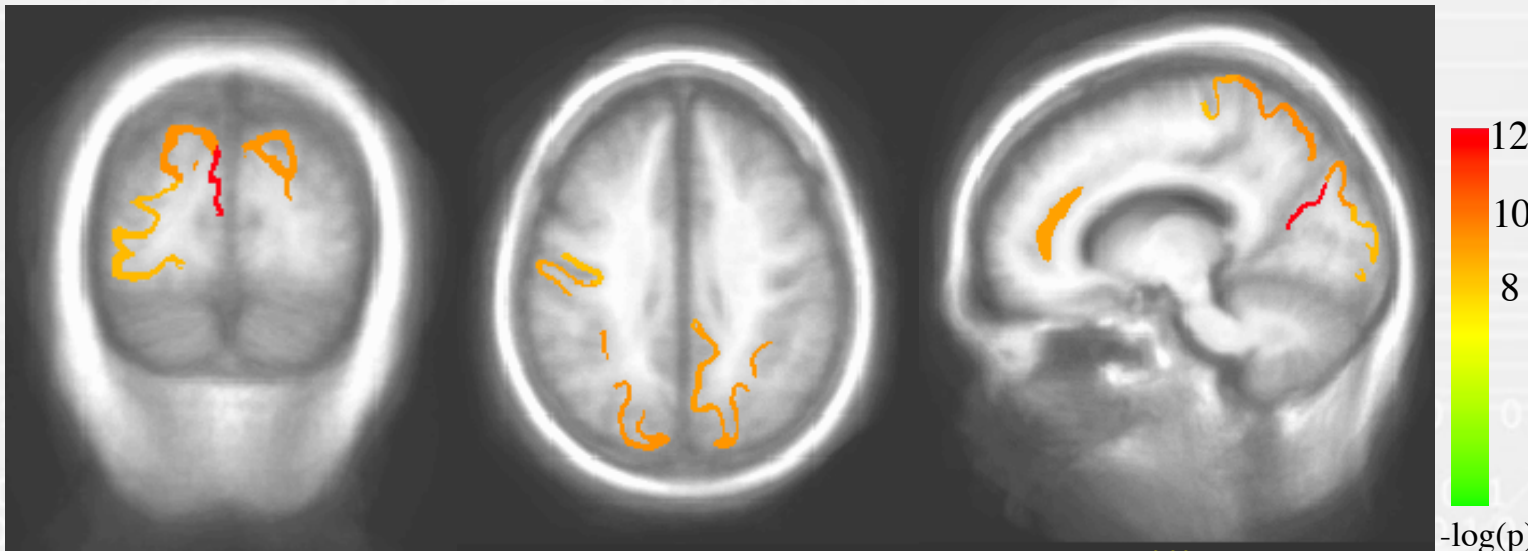
- Mediates cell signaling
- Plays a role in receptor binding and growth factor activity
- Associated with *sensory* neuron development
- Associated with schizophrenia



Preprocess
 FreeSurfer
 Alignlinear
 Align_warp
 ANIMAL
 ART
 Diffeomorphic Demons
 Elastix
 FLIRT
 FNIRT
 IRTK
 JRD-fluid
 NiftyReg
 ROMEO
 SICLE
 SyN
 SPM's DARTEL

Parcellation
 AAL
 Brodmann areas
 Cerebellar atlas
 Desikan-Killiany atlas
 Destrieux atlas
 Freesurfer aseg
 Harvard-Oxford cortical and subcortical atlas
 Jülich postmortem maps

Quantify
 Curvature index
 Folding index
 Gaussian curvature
 Mean curvature
 Surface area
 Surface mesh of cortex
 Surface mesh of subcortical nuclei
 Thickness
 Volume, normalized to ICV
 Volume, raw



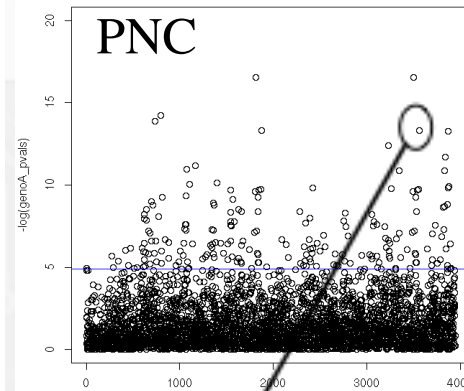
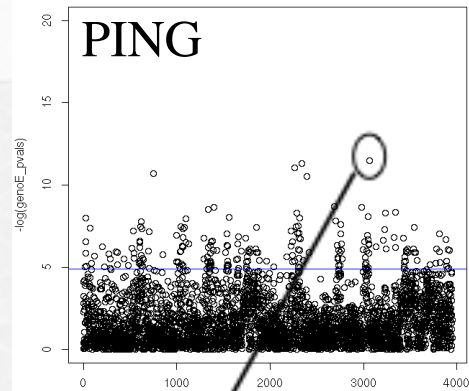
3 trillion variables

~50,000 subjects * ~75 databases * 80,000+ imaging-derived metrics * 10 modalities

BDDS Platform: Neuroimaging PheWAS

BDNF (rs6265)

- Promotes survival of neurons
- Supports synaptic plasticity
- When deleted, causes weight gain and intellectual disability
- Associated with BMI



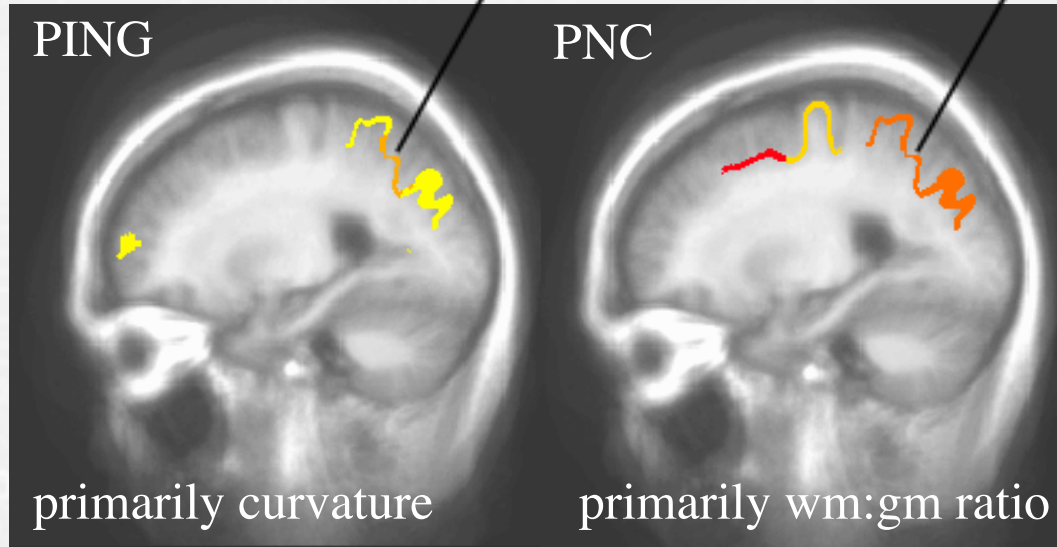
Findings in two neurodevelopmental cohorts

PING

n=736, ages 3-21

PNC

n=971, ages 8-21



Preprocess

FreeSurfer
Alignlinear
Align_warp
ANIMAL
ART
Diffeomorphic Demons
Elastix
FLIRT
FNIRT
IRTK
JRD-fluid
NiftyReg
ROMEO
SICLE
SyN
SPM's DARTEL

Parcellation

AAL
Brodmann areas
Cerebellar atlas
Desikan-Killiany atlas
Destrieux atlas
Freesurfer aseg
Harvard-Oxford cortical and subcortical atlas
Jülich postmortem maps

Quantify

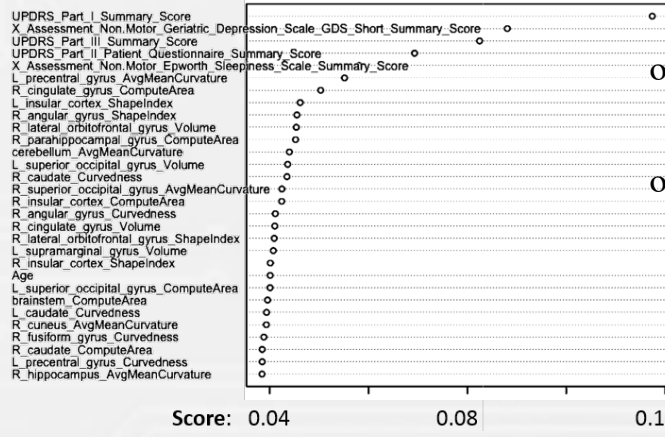
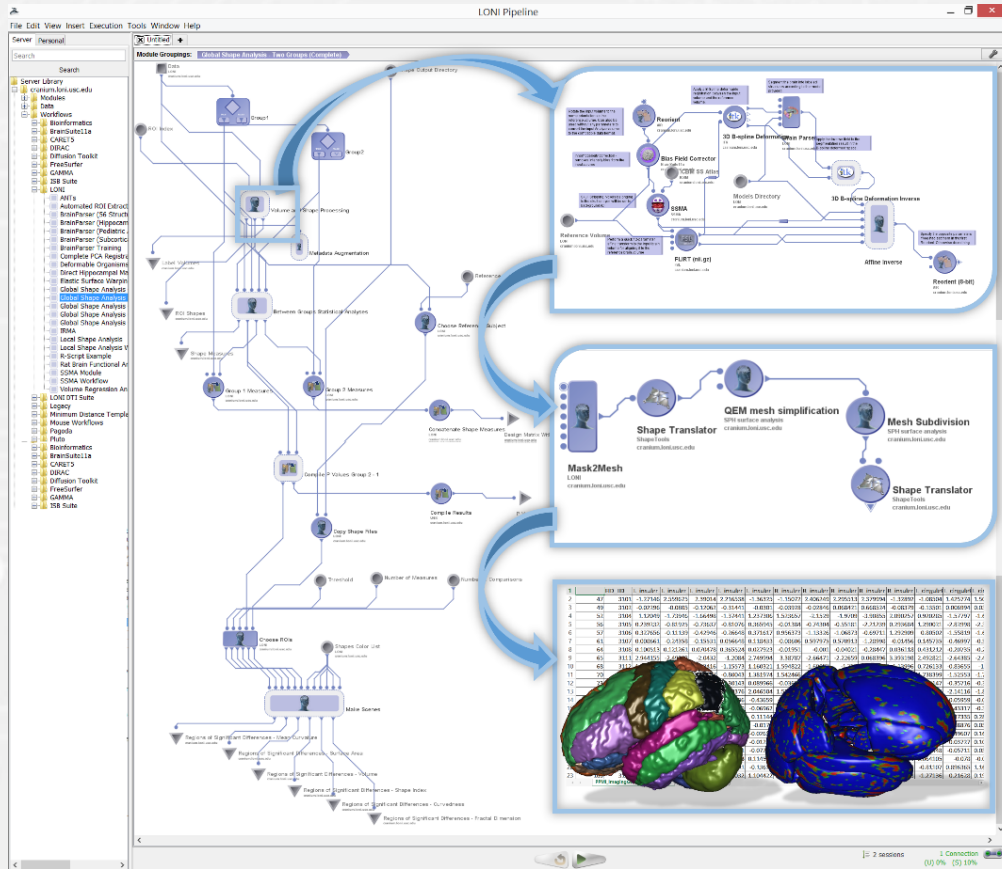
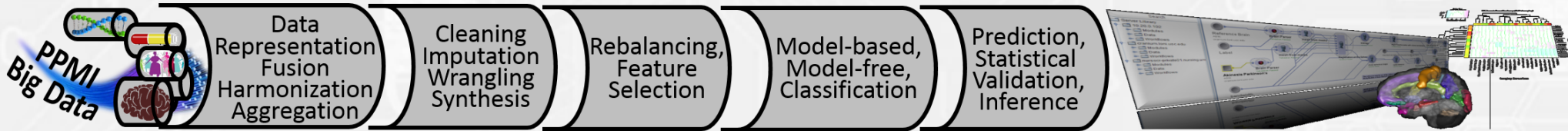
Curvature index
Folding index
Gaussian curvature
Mean curvature
Surface area
Surface mesh of cortex
Surface mesh of subcortical nuclei
Thickness
Volume, normalized to ICV
Volume, raw

3 trillion variables

~50,000 subjects * ~75 databases * 80,000+ imaging-derived metrics * 10 modalities

Predictive Analytics using Large, Complex, Incongruent, Heterogeneous Multi-source & Incomplete Observations

• A Big Data Study of Parkinson's Disease



Varplot illustrating:
the critical predictive data elements (Y-axis) and their impact scores (X-axis)
AdaBoost classifier for Controls vs. Patients prediction

ML classifier	accuracy	sensitivity	specificity	positive predictive value	negative predictive value	log odds ratio (LO)
AdaBoos t	0.996324	0.994141	0.998264	0.9980392	0.9948097	11.4888
SVM	0.985294	0.994140	0.977431	0.9750958	0.9946996	8.9021

BDDS Platform Demos

www.bd2k.org

- Minimal Viable Information Identifier
 - Identifying Research Data Objects <http://minid.bd2k.org/>
 - Consortium Activity with bioCADDIE, CEDAR & HeartBD2K
- Amyloid Burden – PD & AD
 - Integrated Exchange of Multi-omic data
- PheWAS
 - Gene-Brain Associations
- Predictive Big Data Analysis – PD
 - Defining Data Characteristics