Let the data lead to the discoveries

What can be done now, that was not possible before?
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
- Genomewide search
  - 100K’s of SNPs
- Environment
  - age, SES, school, nutrition, etc
- Phenotypes
  - 1-5 brain phenotypes
- Identifies which genes have the most influence on brain phenotypes of interest

PheWAS
- Phenome-wide search
  - 10K’s of brain phenotypes
- Environment
  - age, SES, school, nutrition, etc
- Genotypes
  - 1-5 SNPs
- Identifies which brain phenotypes are influenced the most by a genotype of interest
**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

Preprocess
- FreeSurfer
- Alignlinear
- Align_warp
- ANIMAL
- ART
- Diffeomorphic Demons
- Elastix
- FLIRT
- FNIRT
- IRTK
- JRD-fluid
- NiityReg
- 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

Findings in two neurodevelopmental cohorts

- **PING**
  - n=736, ages 3-21

- **PNC**
  - n=971, ages 8-21

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

Preprocess
- FreeSurfer
- Alignlinear
- Align_warp
- ANIMAL
- ART
- Diffeomorphic Demons
- Elastix
- FLIRT
- FNIRT
- IRTK
- JRD-fluid
- NiityReg
- 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

Findings in two neurodevelopmental cohorts

- **PING**
  - n=736, ages 3-21

- **PNC**
  - n=971, ages 8-21

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

<table>
<thead>
<tr>
<th>ML classifier</th>
<th>accuracy</th>
<th>sensitivity</th>
<th>specificity</th>
<th>positive predictive value</th>
<th>negative predictive value</th>
<th>log odds ratio (LOD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AdaBoost</td>
<td>0.99632</td>
<td>0.994141</td>
<td>0.998264</td>
<td>0.9980392</td>
<td>0.9948097</td>
<td>11.4882</td>
</tr>
<tr>
<td>SVM</td>
<td>0.98529</td>
<td>0.994140</td>
<td>0.977431</td>
<td>0.9750958</td>
<td>0.9946996</td>
<td>8.9021</td>
</tr>
</tbody>
</table>
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