Big Data for Discovery Science (BDDS) knowledge discovery platform: Neuroimaging PheWAS

Kristi Clark[1], Clio Gonzalez-Zacarias[1], Gustavo Glusman[2], Surafael Yared[1], Sabir Saluja[1], Mike Darcy[3], Carl Kesselman[3], Arthur Toga[1]

[1]Laboratory of Neuro Imaging, USC Mark and Mary Stevens Neuroimaging and Informatics Institute, University of Southern California, Los Angeles, CA USA; [2]Institute for Systems Biology, Seattle, WA USA; [3]Information Sciences Institute, University of Southern California, Los Angeles, CA USA

ark:/88120/r87p49

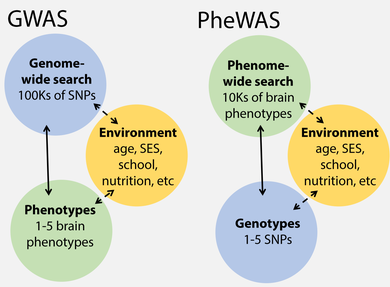
Abstract

The core of the BDDS knowledge discovery platform is the analysis and integration of large volumes of heterogeneous data to discover knowledge that is not discoverable using existing methodologies. In this example, we demonstrate how to use the BDDS knowledge discovery platform to perform a Neuroimaging PheWAS analysis. PheWAS stands for phenome-wide association study, which is used to identify which brain phenotypes (out of 10,000s) are influenced by a genotype of interest. This type of analysis yields unique insight into the mechanisms of how particular genotypes influence the brain.

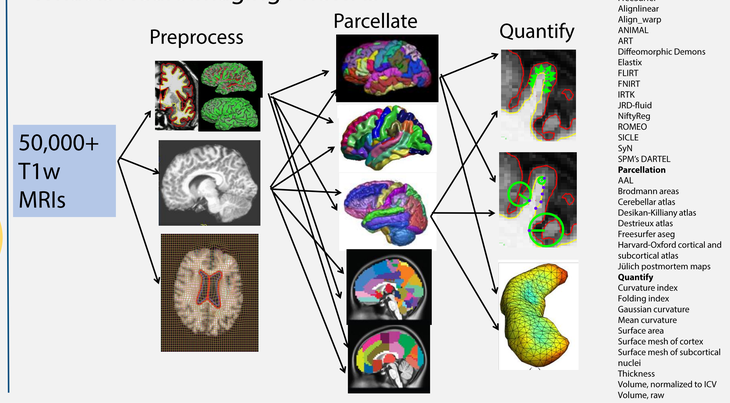
Introduction

What is PheWAS?

Phenome-wide study to discover gene-brain associations

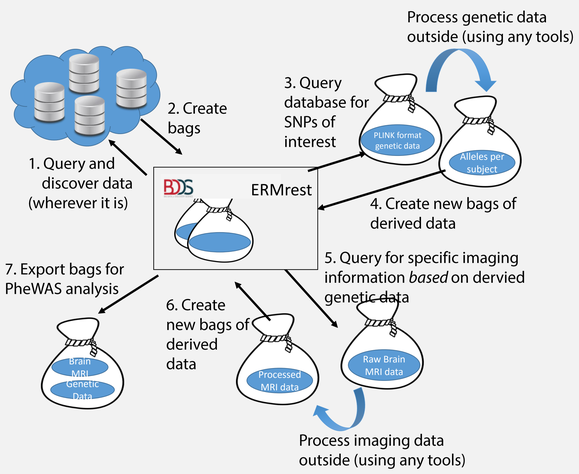


What is Neuroimaging PheWAS?



Materials & Methods

BDDS Dynamic Database Approach



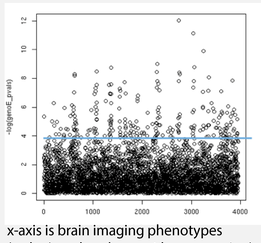
PheWAS Analysis Approach

1. Divide the subjects into groups based on genotype of interest, e.g. AG vs GG
2. For each brain imaging phenotype, e.g. average cortical thickness of precentral gyrus identified by the AAL atlas in data preprocessed with FreeSurfer, compute either a t-test (if 2 genotypes of interest) or ANOVA (if more than 2, e.g. AA, AG, or GG)
3. Plot –log (p-value) for all brain imaging phenotypes to generate a Manhattan plot
4. Identify a horizontal line representing Bonferroni multiple comparisons correction
5. Interpret most significant results
6. Combine or replicate data across databases

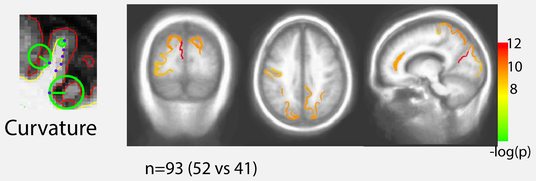
Results

BDDS Dynamic Database example #1: SNP rs35753505 (NRG1; neureglin-1)

Neureglin-1 (rs35753505)
- Mediates cell signaling
- Plays a role in receptor binding and growth factor activity
- Associated with *sensory* neuron development
- Associated with schizophrenia



x-axis is brain imaging phenotypes (order/number does not have meaning)



n=93 (52 vs 41)

BDDS Dynamic Database example #2: SNP rs6265 (BDNF; brain-derived neurotrophic factor)

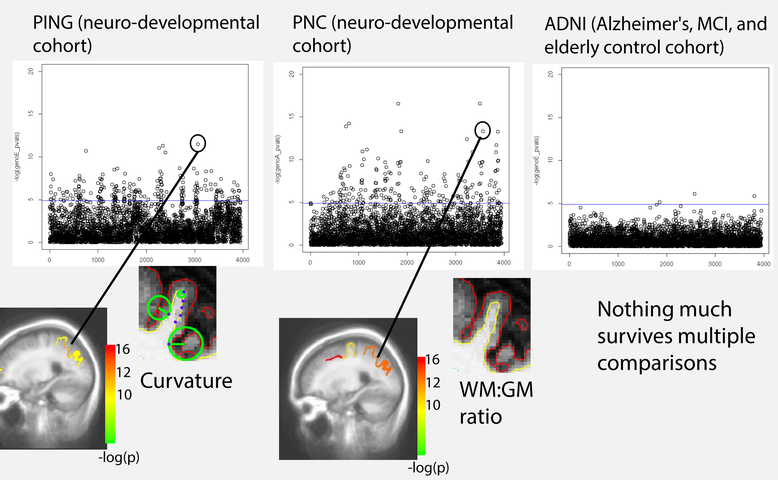
BDNF (rs6265)
- Promotes survival of neurons
- Supports synaptic plasticity
- When deleted, causes weight gain and intellectual disability
- Associated with BMI

Subjects

PING: 736 (227 vs 509)
ages 3-21 years

PNC: 971 (84 vs 887)
ages 8-21 years

ADNI: 324 (241 vs 83)
ages 55-96 years
*Note: results only included one time point per subject



Discussion

Applying the BDDS Neuroimaging PheWAS approach to a neurodevelopmental cohort (PNC) investigating SNP rs35753505, which corresponds to the 5'-flanking region of the neureglin-1 gene, explains a significant amount of variance in the curvature of visual and sensory areas, particularly in the right hemisphere. This is in concordance with known literature from other fields that demonstrate that NRG1 is associated with sensory neuron development. It yields new insight that: a) curvature in visual and sensory areas are related to each other, b) cortical curvature in the right hemisphere visual and sensory areas are under common genetic control, and c) NRG1 plays a role in determining the biological factors that underlie cortical curvature. In future studies, this finding may yield important insight into why rs35753505 is a risk factor for schizophrenia.

Using the same approach to study a second SNP, rs6265 (known as Val66Met in the BDNF gene), we replicated results across two neurodevelopmental databases showing that BDNF explains a signficant amount of variance in the morphometry of the superior parietal regions (and other regions not shown here). In the PING database, this variance was explained for the curvature of these regions, while in the PNC database, this variance was explained in the white matter to gray matter ratio. No significant findings were identified in the ADNI database, possibly because this database is an older and sicker population. It is possible that age and disease effects mask the influence of BDNF in this population.

BDDS Platform creates a *dynamic* database for knowledge discovery. Not only dynamic in terms of the raw data, but also in terms of incorporating derived, processed data. Neuroimaging PheWAS offers unique insight into the mechanisms by which specific SNPs of interest influence the brain.

Big Data for Discovery Science (BDDS): http://bd2k.ini.usc.edu
NIH U54 EB020406