The Informatics Backbone of the Brain Genomics Superstruct Project
Open Data Release

T.M. O’Keefe1,2, V.I. Petrov1,3, R.L. Buckner1,2,3,4, A.J. Holmes2, G.R. Fariello1,2,3
Neuroinformatics Research Group, Harvard University 1, Center for Brain Science, Harvard University 2, Athinoula A. Martinos Center for Biomedical Imaging 3, HHMI 4

Introduction

Large scale imaging data sets are necessary to address questions regarding the relationship between brain and behavior. Generating, storing and analyzing the required data are a daunting enterprise for many independent research groups. In 2007, the Open Access Series of Imaging Studies (OASIS) sought to remove these obstacles by developing a distribution model for free and reusable magnetic resonance imaging data sets (Marcus et al., 2007a, 2009). The community has benefited from these and other open data initiatives including the 1000 Functional Connectomes Project (Biswal et al., 2010) and the ongoing NIH Human Connectome Project (Van Essen et al., 2012). The availability of open data creates opportunities for researchers to contribute scientifically while spending less time and resources gathering independent, and often redundant, data. In the spirit of these initiatives, the Brain Genomics Superstruct Project Open Data Release presented here reflects the public release and informatics behind a uniform, high-quality collection of neuroimaging, cognitive, behavioral and derived data for 1500 human participants. These data sets will be available from a hosted or downloadable installation of the Extensible Neuroimaging Archive Toolkit (XNAT, Marcus et al., 2007b). Each data set will contain T1-weighted and bandwidth-matched T2-weighted structural data, low-resolution DTI, resting state BOLD acquisitions and, for a subset of subjects, DSI data amenable to tractography. Demographic, cognitive (e.g., Matrix Reasoning, Dot Comparisons, personality (e.g., STAI, NEO-FFI) and lifestyle assessments will also be provided. Each data set will be accompanied by behavioral, personality and cognitive scoring, a fully-automated quality assessment of functional acquisitions, manual quality assessments of anatomical acquisitions, pre-computed analysis of intrinsic connectivity (Van Dijk et al., 2010) and morphometrics (Fischl et al., 2000, 2004).

Data Features

MRI

• CAP (15 minutes) and CAP-Extended (34 minutes, Above)
• Matched Siemens 3T Tim Trio® scanners (VB15 upgraded to VB17)

Demographics and Lifestyle

• Annett Hand Preference Questionnaire
• Race, ethnicity, socioeconomic status, education, occupation, health, psychiatric screening, medication, family history, alcohol and tobacco use

Personality and Behavior

• State-Trait Anxiety (STAI)
• Neuroticism Extraversion Openness to Experience, Five Factor Inventory (NEO-FFI)
• Behavioral Inhibition, Behavioral Activation, and Affective Responses to Impending Reward and Punishment (BIS/BAS)
• Imaginal Processes Inventory Frequency Scale
• Domain-Specific Risk Attitude (DOSPERT)
• Barratt Impulsiveness Scale
• Profile of Mood States 1 (POMS)
• Shipley Institute of Living Scale (SILLS) vocabulary for VIQ
• Temperament and Character Inventory (TCI-9), subset

Cognition

• WASI Matrix Reasoning for PIQ
• Mental Rotation
• Dot Comparison
• Visual Paired Associates
• Flanker Reversed
• Reading the Mind in the Eyes Task (RMET)
• Emotional Faces (Buckner et al., 2011, O’Keefe et al., 2011)

Preparing for Launch

• The initial GSP Open Data Release is targeted for Fall of 2012.
• Efforts are underway to release pipeline source code and APs for public consumption. In addition, virtualization of the GSP processing environment will be provided in OVF format and for use within commercial cloud computing platforms such as Amazon EC2, Google Compute Engine and Microsoft Azure.
• A GSP-specific registration template for FreeSurfer is expected to decrease transform uncertainties as well as address individual cases where reconstructions fail (currently 0.3% of all cases). Results from this analysis will be included in the GSP Open Data Release.
• Following migration to XNAT 1.6, all pipeline content (e.g., schemas, views) will be submitted to the XNAT Marketplace.

Automated Pipeline: Morphometrics

Automated Pipeline: BOLD Quality Assessment

Automated Pipeline: Intrinsic Functional Connectivity

Behavior, personality and cognitive assessments are automatically scored. Example related measures are illustrated including NEO-FFI neuroticism vs. STAI trait anxiety (N=2007, r=0.78) and POMS total mood disturbance vs. STAI trait anxiety (N=1940, r=0.70).

Automated Pipeline: Cognitive Scoring

Automated Pipeline: Behavior, Personality and Cognitive Scoring

References