NeuroImage 124 (2016) 1149–1154

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

The Pediatric Imaging, Neurocognition, and Genetics (PING) Data Repository

Terry L. Jernigan [a,b,c,*], Timothy T. Brown [d,e], Donald J. Hagler Jr. [d,f], Natacha Akshoomoff [a,c], Hauke Bartsch [d], Erik Newman [a,c], Wesley K. Thompson [c,g], Cinnamon S. Bloss [h], Sarah S. Murray [i], Nicholas Schork [j], David N. Kennedy [k], Joshua M. Kuperman [d,f], Connor McCabe [l], Yoonho Chung [m], Ondrej Libiger [h], Melanie Maddox [a], B.J. Casey [n], Linda Chang [o], Thomas M. Ernst [o], Jean A. Frazier [k], Jeffrey R. Gruen [p], Elizabeth R. Sowell [q], Tal Kenet [r], Walter E. Kaufmann [s], Stewart Mostofsky [t], David G. Amaral [u], Anders M. Dale [b,d,e,f], for the Pediatric Imaging, Neurocognition and Genetics Study

[a] Center for Human Development, University of California, San Diego, La Jolla, CA, USA
[b] Department of Cognitive Science, University of California, San Diego, La Jolla, CA, USA
[c] Department of Psychiatry, University of California, San Diego, La Jolla, CA, USA
[d] Multimodal Imaging Laboratory, University of California, San Diego, La Jolla, CA, USA
[e] Department of Neurosciences, University of California, San Diego, La Jolla, CA, USA
[f] Department of Radiology, University of California, San Diego, La Jolla, CA, USA
[g] Stein Institute for Research on Aging, University of California, San Diego, La Jolla, CA, USA
[h] The Qualcomm Institute, University of California, San Diego, La Jolla, CA, USA
[i] Department of Pathology, University of California, San Diego, La Jolla, CA, USA
[j] Human Biology, J. Craig Venter Institute, USA
[k] Department of Psychiatry, University of Massachusetts Medical School, Boston, MA, USA
[l] Department of Psychology, University of Washington, Seattle, WA, USA
[m] Department of Psychology, Yale University, New Haven, CT, USA
[n] Sackler Institute for Developmental Psychobiology, Weil Cornell Medical College, New York, NY, USA
[o] Department of Medicine, University of Hawaii, Queen's Medical Center, Honolulu, HI, USA
[p] Departments of Pediatrics and Genetics, Yale University, School of Medicine, New Haven, CT, USA
[q] Department of Pediatrics, University of Southern California, Children's Hospital Los Angeles, Los Angeles, CA, USA
[r] Department of Neurology, Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA
[s] Boston Children's Hospital, Boston, MA, USA
[t] Kennedy Krieger Institute, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[u] Department of Psychiatry and Behavioral Sciences, University of California-Davis, Davis, CA, USA

ARTICLE INFO

Article history:
Accepted 27 April 2015
Available online 1 May 2015

ABSTRACT

The main objective of the multi-site Pediatric Imaging, Neurocognition, and Genetics (PING) study was to create a large repository of standardized measurements of behavioral and imaging phenotypes accompanied by whole genome genotyping acquired from typically-developing children varying widely in age (3 to 20 years). This cross-sectional study produced sharable data from 1493 children, and these data have been described in several publications focusing on brain and cognitive development. Researchers may gain access to these data by applying for an account on the PING portal and filing a data use agreement. Here we describe the recruiting and screening of the children and give a brief overview of the assessments performed, the imaging methods applied, the genetic data produced, and the numbers of cases for whom different data types are available. We also cite sources of more detailed information about the methods and data. Finally we describe the procedures for accessing the data and for using the PING data exploration portal.

© 2015 Elsevier Inc. All rights reserved.

Introduction

Here we describe the data generated in the Pediatric Imaging, Neurocognition, and Genetics (PING) study, and the data repository that is now accessible to investigators through the PING Portal (http://pingstudy.ucsd.edu). The repository contains aggregated data collected

* Corresponding author at: Center for Human Development, University of California, San Diego, 9500 Gilman Drive, MC 0115, La Jolla, CA 92093, USA. Fax: +1 858 822 1602.
E-mail address: tjernigan@ucsd.edu (T.L. Jernigan).

http://dx.doi.org/10.1016/j.neuroimage.2015.04.057
1053-8119/© 2015 Elsevier Inc. All rights reserved.

at 10 sites in the United States. Multimodal neuroimaging data, developmental histories, behavioral and cognitive assessments, and/or genome-wide genotypes are now available for 1493 children and adolescents, aged 3 to 21 years. The goal of PING was to address the imbalance in existing imaging genomics data resources between those containing data collected from adult and elderly participants and the very limited data available from pediatric and adolescent participants. A wide pediatric age range for participants was included to allow investigators to search for interactions between age and genotype (i.e., to discover gene associations with developmental phenotypes). The PING infrastructure was designed specifically to address the challenges of large, multi-site studies involving multimodal imaging and assessment of behavioral phenotypes in a developmental context, and to facilitate the exploration as well as the dissemination of the sharable data. While PING was a cross-sectional study, all features of the PING infrastructure were designed to be extensible to longitudinal designs and the infrastructure currently supports ongoing longitudinal studies that have followed PING. Below we describe briefly the PING study and cohort; procedures for facilitating, standardizing, and optimizing data acquisition; procedures for processing the imaging and genomics data; and the infrastructure for sharing and exploration of accumulated data.

The PING cohort

Participants were recruited through local postings and outreach activities conducted in the greater metropolitan areas of Baltimore, Boston, Honolulu, Los Angeles, New Haven, New York, Sacramento, and San Diego. Children, adolescents, and young adults were screened as eligible for PING if they were between the ages of 3 and 20 years and fluent in English (some older 20-year-olds turned 21 by the completion of data collection). Exclusion criteria included: a) neurological disorders; b) history of head trauma; c) preterm birth (less than 36 weeks); d) diagnosis of an autism spectrum disorder, bipolar disorder, schizophrenia, or mental retardation; e) pregnancy; and f) daily illicit drug use by the mother for more than one trimester. Individuals with contraindications for MRI (such as dental braces, metallic or electronic implants, or claustrophobia) were also excluded from participating. Individuals with identified or suspected learning disability or ADHD were not excluded since these syndromes are fairly common in pediatric populations. Over 1700 participants were enrolled in PING at one of the 10 sites, however, only data from participants in whom acceptable data were obtained for at least two data types (i.e., demographic/developmental, behavioral, genomic, imaging) are included in the PING repository. This sample consists of 1493 participants (780 males) for whom acceptable imaging, genomics, and/or cognition assessments were obtained; acceptable imaging data were acquired for 1239 of these participants (645 males); acceptable cognition data for 1453 participants (752 males), and acceptable genotyping is available for 1391 (719 males). Similar proportions of males and females participated across the entire age range. Written parental informed consent was obtained for all PING subjects below the age of 18, and child assent was also obtained for all participants between the ages of 7 and 17. Written informed consent was obtained directly from all participants aged 18 years or older. For more information about the PING cohort, see Brown et al. (2012) and Akshoomoff et al. (2014).

Participant assessments

Neuromedical history

The PING Study Demographics and Child Health History Questionnaire was completed by parents or guardians of the minor participants, and participants aged 18 and over were given a self-report version of this questionnaire. The questionnaire assessed basic medical, developmental, and behavioral history, as well as family history of medical and neuropsychiatric disorders, including substance abuse. The measures from this questionnaire are identified in the PING Portal ontology and defined in the data dictionary with the prefix "FDH_".

NIH toolbox cognition battery measures

Cognitive assessments for the PING project were conducted using the NIH Toolbox Cognition Battery (NTCB). The NTCB was designed to tap key functions (executive function, attention, episodic memory, working memory, language, and processing speed) across the lifespan (ages 3 to 85 years). This computerized approach provides an economical method for assessing a wide range of cognitive abilities, which is appealing for large-scale studies. For pediatric studies, this also has the advantage of providing the same set of measures for use with young children, older children, and adolescents in an appealing format that provides automated data collection, storage, and scoring (Keator et al., 2008; McCarty et al., 2014). A detailed description of the NTCB results in the PING study is provided in Akshoomoff et al. (2014). More detailed information on the NIH Toolbox for Cognition is available at http://www.nihtoolbox.org/. The eight NTCB subtests for which data are available are: Dimensional Change Card Sort Test, Flanker Inhibitory Control and Attention Test, Picture Sequence Memory Test, Pattern Comparison Processing Speed Test, Oral Reading Recognition Test, List Sorting Working Memory Test, and Picture Vocabulary Test. The measures obtained with the NTCB are identified in the PING Portal ontology and data dictionary with the prefix "TBX_".

Social–emotional and substance use assessments

In a subset of participants, a limited number of additional assessments of social and emotional functions and substance exposure were acquired through the PhenX Rising project. More information about the PhenX Toolkit and PhenX Rising is available at https://www.phenxtoolkit.org and about the data collected in PING in association with PhenX Rising in McCarty et al. (2014). PhenX assessments were obtained on only a subset of participants because PING was already underway when the PhenX Rising study began. Also, many of the assessments are age-specific. Variables associated with PhenX assessments are identified in the PING ontology and data dictionary with the prefix "PHX_".

Multimodal image acquisition and quality control

In PING, the imaging, quality control, and analysis protocols were developed specifically to meet the challenges associated with multisite, multimodal imaging of children. These procedures were also designed to ensure that acquisition and preprocessing methods were compatible with, and facilitated, the use of post-processing methods developed by leading imaging groups throughout the neuroimaging community as well as those applied by consortium investigators.

Image acquisition and preprocessing

The PING imaging protocol takes advantage of key technologies developed for the consortium and builds on earlier methods development performed as part of the Biomedical Informatics Research Network (BIRN Keator et al., 2008) and the Alzheimer's Disease Neuroimaging Initiative (ADNI Jack et al., 2008). Specifically, a standard PING scan session included: 1) a 3D T_1-weighted inversion prepared RF-spoiled gradient echo scan using prospective motion correction (PROMO), for cortical and subcortical segmentation; 2) a 3D T2-weighted variable flip angle fast spin echo scan, also using PROMO, for detection and quantification of white matter lesions and segmentation of CSF; 3) a high angular resolution diffusion imaging (HARDI) scan, with integrated B_0 distortion correction (DISCO), for segmentation of white matter tracts and measurement of diffusion parameters; and 4) a resting state blood oxygenation level-dependent (BOLD) fMRI scan, with integrated distortion

correction. Pulse sequence parameters used across (3 T) scanner manufacturers (GE, Siemens, and Phillips) and models were optimized for equivalence in contrast properties and consistency in image-derived quantitative measures.

Gradient nonlinearity correction (3D GradWarp)

Nonlinearity of the gradient fields used for spatial encoding in MRI is one of the most prominent sources of spatial distortion in MRI scans (Chang and Fitzpatrick, 1992; Jovicich et al., 2006). Through the involvement of PING neuroimaging investigators in the Biomedical Informatics Research Network (BIRN) and the Alzheimer's Disease Neuroimaging Initiative (ADNI), the group has led the development of a fully automated procedure to correct for gradient field nonlinearities using displacement maps computed based on scanner-specific specifications provided by the MRI scanner manufacturers. The correction software developed in the UCSD MultiModal Imaging Laboratory (MMIL) was adopted as part of the routine pre-processing routine for all scans acquired as part of ADNI, and has been shown to significantly improve the accuracy of longitudinal change estimates based on serial MRI scans (Holland and Dale, 2011).

Motion correction (PROMO)

An important recent advance in MRI acquisition technology is the development of real-time, or prospective, motion correction. The PROMO approach (White et al., 2010), first applied widely in PING, utilizes three orthogonal spiral navigators together with a recursive image-based estimation strategy based on the extended Kalman filter (EKF) for motion measurement. The spiral k-space trajectory allows image-domain reconstruction prior to motion estimation, which when combined with the flexible EKF framework, allows for efficient image-based tracking within an a priori region of interest. Significant reduction of motion-related image degradation in pediatric imaging is possible with this method (Brown et al., 2010; Kuperman et al., 2011).

EPI B_0 distortion correction (DISCO)

Single-shot echo planar imaging (EPI) is an efficient MRI acquisition scheme for producing fast, high-definition images for diffusion weighted imaging and fMRI. However, EPI suffers from severe spatial distortions and intensity variations due to susceptibility-induced B_0 field inhomogeneity. Anatomically accurate, undistorted images are essential for integrating the HARDI and fMRI images with anatomical (T_1-weighted) images in PING, i.e., for achieving accurate spatial registration of the information from different modalities. Since the B_0 distortion pattern depends on the exact position of the subject in the scanner, which may vary across scan sessions, correcting for such distortions is also essential for obtaining accurate estimates of change based on longitudinal MRI scans. Our group has developed a fast, robust, and accurate procedure for removing such spatial and intensity distortions from the EPI images obtained for HARDI and fMRI (Holland et al., 2010). The method involves acquisition of brief scans with opposite phase encoding polarities (resulting in opposite spatial distortion patterns) and subsequent alignment of the resulting images using a fast nonlinear registration procedure. The DISCO method, which requires minimal additional scan time, provides superior accuracy and better cross-modality registration relative to the more commonly used, and more time consuming, field mapping approach.

Multimodal image analysis

Morphometric analysis of structural MRI data was performed using a specialized processing stream developed for PING that is based on FreeSurfer, with additional corrections and analyses developed at UCSD MMIL.

Structural MRI preprocessing

As described above, distortions caused by nonlinearity of the spatial encoding gradient fields were corrected with predefined, scanner specific, nonlinear transformations, provided by MRI scanner manufacturers (Jovicich et al., 2006). Non-uniformity of signal intensity was reduced using the nonparametric non-uniform intensity normalization (N3) method (Sled et al., 1998). Images were rigidly registered and resampled into alignment with an atlas brain with 1 mm isotropic voxels, facilitating standardized viewing and analysis of brain structure. If multiple, good quality scans were obtained for a participant, they were registered to each other and averaged.

Morphometric analysis

FreeSurfer encompasses tools for cortical surface reconstruction, subcortical segmentation, cortical parcellation, and estimation of various measures of brain morphometry using routinely acquired T_1-weighted MRI volumes (Dale and Sereno, 1993; Dale et al., 1999; Desikan et al., 2006; Destrieux et al., 2010; Fischl and Dale, 2000; Fischl et al., 1999a,b, 2001, 2002, 2004; Salat et al., 2009; Ségonne et al., 2004, 2007). Important extensions made at MMIL include maps of relative cortical surface area changes (Chen et al., 2012; Joyner et al., 2009) and genetically informed cortical parcellations (Chen et al., 2011, 2012, 2013). Cortical surface reconstruction involves skull-stripping (Ségonne et al., 2004), non-uniformity correction (Sled et al., 1998), white matter segmentation, initial mesh creation (Dale et al., 1999), correction of topological defects (Fischl et al., 2001; Ségonne et al., 2007), and generation of optimal white and pial surfaces (Dale and Sereno, 1993; Dale et al., 1999; Fischl and Dale, 2000). Subcortical structures were labeled using an automated, atlas-based, volumetric segmentation procedure (Fischl et al., 2002); volumes in mm^3 and average T_1-weighted intensity (T1w) were calculated for each structure. Labels for cortical gray matter and underlying white matter voxels were assigned based on surface-based nonlinear registration to atlas based on gyral and sulcal patterns (Fischl et al., 1999b) and Bayesian classification rules (Desikan et al., 2006; Destrieux et al., 2010; Fischl et al., 2004; Salat et al., 2009). White matter voxels adjacent to the cortical parcels were also labeled (Salat et al., 2009). Fuzzy-cluster parcellations based on genetic correlation of surface area were used to calculate weighted averages of cortical surface measures (Chen et al., 2012). Cortical thickness was calculated as the shortest distance between the white and pial surfaces (Fischl and Dale, 2000). Maps of relative cortical areal expansion were created by resampling individual subject surfaces to a standard tessellation, such that the area assigned to each mesh vertex reflects the degree of expansion or contraction relative to the atlas (Chen et al., 2012; Joyner et al., 2009). T1w was sampled to the cortical surface at a distance of ± 0.2 mm along the normal vector at each surface location, and T1w cortical contrast was calculated from gray and white matter values (Westlye et al., 2009). Average thickness, area, and T1w were calculated for each cortical parcel. Surface-based maps were sampled to the FreeSurfer atlas (Fischl et al., 1999a) and smoothed along the cortical surface (Hagler et al., 2006).

Diffusion MRI preprocessing

As described above, spatial and intensity distortions caused by B_0 field inhomogeneity were reduced using a robust and accurate procedure for reducing spatial and intensity distortions in EPI images (Holland et al., 2010) that relies on the reversing gradient method (Hagler et al., 2006; Jovicich et al., 2006). Eddy current distortions were corrected with a nonlinear estimation procedure that used the diffusion gradient orientations and amplitudes to predict the pattern of distortions across the entire set of diffusion weighted volumes (Hagler et al., 2009).

Microstructural analysis

Diffusion parameters were computed for a set of major brain fiber tracts (Cann et al., 2002), as well as for other brain structures of interest. Conventional DTI methods were used to calculate measures related to microstructural tissue properties (Altshuler et al., 2010; Nelson et al., 2008; Xing et al., 2009), including the principal diffusion orientation, fractional anisotropy (FA), and mean, longitudinal, and transverse diffusivity (MD, LD, and TD). T_2-weighted intensity (T2w) was calculated from the b = 0 image (averaged if multiple b = 0 images). To remove arbitrary intensity variation across subjects due to scanner settings (i.e., gain), T2w images were normalized for each subject. A linear fit with zero intercept was calculated between MD and b = 0 intensity values using each voxel within a brain mask as a separate data point. The slope of the linear relationship was used to scale the T2w images.

AtlasTrack was used to automatically label long-range white matter tracts based on a probabilistic atlas of fiber tract locations and orientations (Hagler et al., 2009). The fiber atlas contains prior probabilities and orientation information for specific long-range projection fibers, including some additional fiber tracts not included in the original description, such as cortico-striate connections and inferior to superior frontal cortico-cortical connections.

Imaging variables associated with morphometry are identified in the PING ontology and data dictionary with the prefix "MRI_" followed by additional labels appropriate for specific measures (e.g., "cort_area", "cort_thick", "subcort_vol", etc.); and those associated with diffusion data with the prefix "DTI_" followed by additional labels appropriate for specific measures (e.g., "fiber", "aseg"). The "aseg" designation refers to regions of interest delineated in the volumetric analysis.

Imaging data quality control

Raw image quality control

Through the secure web-based application, individual sites uploaded DICOM images for each scan session. The data were automatically checked for completeness and protocol compliance, and images were reviewed for image quality by technicians trained by faculty. Specifically, images were inspected for motion artifacts, excessive distortion, operator error, or scanner malfunction. Quality ratings—good, average (usable), and bad (unacceptable)—were entered into the quality control utility within 24 h from time of upload to allow re-scanning of subjects when possible.

T_1-weighted images were examined slice-by-slice for evidence of excessive motion, such as stark ribbon or criss-cross artifacts within parenchyma and ghosting artifacts outside the head. Each volume was rated as either acceptable or recommended for rescan. Similarly, diffusion images were examined across all slices for signs of artifacts and poor image quality. Volumes with five or more slices showing significant slice-to-slice motion, motion artifacts, or whole-slice dropout were rejected (i.e., recommended for rescan). BOLD data were inspected for excessive subject movement and artifacts, and the mean frame-to-frame head motion was calculated.

Processed image quality control

Processed images from all modalities were also examined for all participants, including subcortical volumetric segmentations, cortical areal parcellations, and white and pial surface reconstructions. A series of QC movies were also produced using Matlab scripts for each subject that assisted in data examination. A movie showing coronal views in sequence was used to judge white matter texture consistency and possible temporal underestimation. A related horizontal sequence was used to check for temporal underestimation in other regions (i.e., superior). A movie showing sagittal views was used for examination and rating of pial and dural overestimation along parietal regions and for signs of excessive head motion. White matter tracts produced using AtlasTrack were inspected for contiguity and overall quality and rated as acceptable or not.

Processing and analysis of genetic information

Acquisition of samples and DNA extraction

Saliva collection in PING was performed using two different products from DNAgenoTek, Oragene•DISCOVER (OGR-500) and Oragene•DISCOVER (OGR-250). DNA extraction was carried out using respective protocols provided by Oragene, and DNA quantity was assessed using a Nanodrop fluorometer. DNA samples with at least 3ug total DNA were carried forward for further processing. Additional saliva samples were requested on specimens with less than 3 μg total DNA. Stock DNA was stored at −80 °C for long-term storage. Ultimately samples were processed for 1411 PING participants.

Genome-wide genotyping

Genome-wide genotyping was performed on the extracted DNA using the Illumina Human660W-Quad BeadChip. The Illumina Human660W-Quad BeadChip (see www.illumina.com) contains more than 550,000 genetic markers (single nucleotide polymorphisms or SNPs and other variants) and is designed to measure most of the genetic variation present in the human genome (based on Hapmap release 21 reference data, see http://hapmap.ncbi.nlm.nih.gov/). The BeadChip measures variants on all autosomes (i.e., non-sex-chromosomes), the X and Y chromosomes, as well as mitochondrial DNA. The SNP call rate was >99% (i.e., >99% of the 539,865 SNPs were called). Acceptable genotyping data could be obtained by the Genomics Core for 1391 of the participant samples processed (out of the total 1396 received, or > 99.5%), including 727 males and 679 females.

Up to 1000 single nucleotide polymorphism values can be downloaded through the PING Portal Genetics Browser and the entire set of genotyping data is available in bulk.

Genetic ancestry assessment

In order to assess each participant's ancestry based on their genotype information, we constructed an ancestry-informative reference panel by bringing together genotype data from 2513 individuals of known ancestry from 63 populations around the world using several publicly available sources: 1) the Human Genome Diversity Project (HGDP) (Cann et al., 2002); 2) the Population Reference (POPRES) (Nelson et al., 2008); 3) the International HapMap 3 Consortium (HapMap3) (Altshuler et al., 2010); and 4) the University of Utah dataset (Xing et al., 2009). The reference panel was created in a stepwise fashion in order to ensure that the included individuals were not admixed among six major continental populations (African, Central Asian, East Asian, European, Native American, and Oceanic) and that each continental population was represented by a reasonably large number of diverse individuals originating in the relevant continent. The assembled reference panel contained genotype information at 16,433 strand-unambiguous SNPs. These markers exhibited low LD (r-squared less than 0.1 was observed between 99% of marker pairs), and allele frequency was higher than 1%.

To assess ancestry and admixture proportions in the PING participants, we used a supervised clustering approach implemented in the ADMIXTURE software (Alexander et al., 2009) and probabilistically assigned each participant to six clusters corresponding to the six major continental populations. The genotype profiles of the six populations were defined by the individuals who made up the reference panel. Although some individuals could be easily associated with a particular continental population, other individuals were clearly admixed. Such admixture could easily be associated with important phenotypic variation

and hence needs to be quantified (Goetz et al., 2014; Norden-Krichmar et al., 2014; Nievergelt et al., 2014; Libiger and Schork, 2012). We therefore determined the degree of ancestry of each participant, effectively quantifying the amount of their genome that is likely to be derived from each of the six populations. Variables associated with genetic ancestry factors are identified in the PING ontology and data dictionary with the prefix "GAF_".

Data sharing

Access to the PING Data Resource is available through an online web interface at http://pingstudy.ucsd.edu. Here, information about the study, the consortium, and the methods are available for browsing, and instructions are given for applying for approval to explore, download, or request bulk shipping of data (for a fee covering media and shipping charges). Full sharing of all data is not permitted by the IRB for some PING sites. This includes restrictions on the sharing of some raw image data on the NITRC site and restrictions on sharing of some genetics data. All sharable data are available through the PING Data Portal to any researcher who holds a position in a research institution and is at least at the postdoctoral level (upon assent to the PING Data Use Agreement and approval of a brief data use description). Students can gain access to data if sponsored by eligible supervising researchers who agree to supervise the students' compliance with the data use agreement. Raw image data for a subset of the participants is available through NITRC after an account is approved through the PING Portal.

Large data request downloads for PING are handled in two ways. First, raw image data in DICOM format is distributed using the dedicated image distribution platform on NITRC.org. The system used by NITRC (XNAT) is specifically built to host and download data across wide-area networks. Processed imaging and genetics data are also shared by PING using hard drives that can be ordered as PING-IN-A-BOX systems. The hard drive is shipped to customers (drive and shipping costs are billed to the recipient) and contains data in DICOM, Nifti, MGZ, and plink binary formats for easy integration into other existing post-processing pipelines.

Although PING data are publicly available, new data cannot be contributed into the PING data repository from outside sources. Given its strict standardization procedures for behavioral, imaging, and genomic data acquisition and processing, it is considered to be a completed resource.

The primary sponsors of the PING repository, NIDA and NICHD, as well as the NIH Office of the Director, have made a major, long term commitment to preservation of informative data repositories, particularly those with valuable genomic data, for future use. The contents of the repository may be transferred to one or more of the major NIH sponsored data repositories, such as NDAR or dbGaP, but it is likely that continuing support of the PING dataset will be provided. The Center for Human Development at UCSD is also committed to long-term maintenance of this resource.

PING data exploration portal

An intelligent data exploration tool is available to facilitate the application of advanced statistical models to PING data, and to enable region of interest- and vertex-wise mapping of effects and 3D visualization of the results onto the cortical surface (Bartsch et al., 2014). In contrast to many data sharing tools, the Portal also integrates appropriate statistical modeling capabilities, with structured descriptors for all of the PING measurements and an intuitive user interface to control diverse quality control and analysis workflows.

The Portal supports online collaborative exploration of interrelationships among measurements obtained from structural, behavioral, and genetic analyses. Using this web application, a user can select a specific variable of interest (e.g., cortical thickness) from the data dictionary and, with a single click of a button fit a statistical model (e.g., generalized additive model; GAM) with one or more independent variables and covariates (e.g., age, sex, scanner, genetic ancestry factor), and plot the resulting model fits along with the individual data points. For example, using the Portal one can interactively produce a scatter plot of each participant's total cortical surface area as a function of age, color-coded by sex, and controlling for other factors such as socioeconomic status (SES) and genetically derived ethnic ancestry. This statistical model can then be applied to every cortical vertex to produce a map of this relationship between age and surface area with covariates of interest. Main effects and interactions can be modeled for summary variables, averages, or by vertex, and displayed interactively, in real-time, using a WebGL-based application with control over color mapping, orientation of brain hemispheres, and corrections for multiple comparisons by controlling the false discovery rate (FDR) (Genovese et al., 2002). The same modeling functions can be applied to region of interest analyses either specified by the user or using built-in cortical parcellations from Freesurfer.

The model, or query, can then be stored by the user and shared with other users. The ease of interaction permits deeper understanding of the complex relationships within the dataset and facilitates the discovery of hidden structure in the data. In PING, for example, the Portal has been an effective way to visualize scanner effects and the results of different methods for modeling them. The Portal supports an online chat feature that is used to send model descriptions to other people visiting the page, as well as a user feedback forum monitored by the developer. Users can construct and download datasets, statistical reports, results tables, model specifications, and figures for off-line analysis, archiving, or publication.

Acknowledgments

The PING Project was supported by the National Institute on Drug Abuse and the Eunice Kennedy Shriver National Institute of Child Health and Human Development with the following awards: RC2DA029475 and R01 HD061414.

References

Akshoomoff, N., et al., 2014. The NIH toolbox cognition battery: results from a large normative developmental sample (PING). Neuropsychology 28, 1–10.
Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664.
Altshuler, D.M., et al., 2010. Integrating common and rare genetic variation in diverse human populations. Nature 467, 52–58.
Bartsch, H., Thompson, W.K., Jernigan, T.L., Dale, A.M., 2014. A web-portal for interactive data exploration, visualization, and hypothesis testing. Front. Neuroinformatics 8, 25.
Brown, T.T., et al., 2010. Prospective motion correction of high-resolution magnetic resonance imaging data in children. Neuroimage 53, 139–145.
Brown, T.T., et al., 2012. Neuroanatomical assessment of biological maturity. Curr. Biol. 22, 1693–1698 (doi:S0960-9822(12)00793-2 [pii] http://dx.doi.org/10.1016/j.cub.2012.07.002).
Cann, H.M., et al., 2002. A human genome diversity cell line panel. Science 296, 261–262.
Chang, H., Fitzpatrick, J.M., 1992. A technique for accurate magnetic resonance imaging in the presence of field inhomogeneities. IEEE Trans. Med. Imaging 11, 319–329.
Chen, C.H., et al., 2011. Genetic influences on cortical regionalization in the human brain. Neuron 72, 537–544.
Chen, C.H., et al., 2012. Hierarchical Genetic Organization of Human Cortical Surface Area. Science 335 (80), 1634–1636.
Chen, C.H., et al., 2013. Genetic topography of brain morphology. Proc. Natl. Acad. Sci. U. S. A. 110, 17089–17094.
Dale, A.M., Sereno, M.I., 1993. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. J. Cogn. Neurosci. 5, 162–176.
Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9, 179–194.
Desikan, R.S., et al., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31, 968–980.
Destrieux, C., Fischl, B., Dale, A., Halgren, E., 2010. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage 53, 1–15.
Fischl, B., Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc. Natl. Acad. Sci. U. S. A. 97, 11050–11055.
Fischl, B., Sereno, M.I., Tootell, R.B., Dale, A.M., 1999a. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum. Brain Mapp. 8, 272–284.

Fischl, B., Sereno, M.I., Dale, A.M., 1999b. Cortical surface-based analysis II: inflation, flattening, a surface-based coordinate system. Neuroimage 9, 195–207.
Fischl, B., Liu, A., Dale, A.M., 2001. Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. IEEE Trans. Med. Imaging 20, 70–80.
Fischl, B., et al., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neurotechnique 33, 341–355.
Fischl, B., et al., 2004. Sequence-independent segmentation of magnetic resonance images. Neuroimage 23 (Suppl. 1), S69–S84.
Genovese, C.R., Lazar, N.A., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage 15, 870–878.
Goetz, L.H., Uribe-Bruce, L., Quarless, D., Libiger, O., Schork, N.J., 2014. Admixture and clinical phenotypic variation. Hum. Hered. 77, 73–86.
Hagler Jr., D.J., Saygin, A.P., Sereno, M.I., 2006. Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data. Neuroimage 33, 1093–1103.
Hagler Jr., D.J., et al., 2009. Automated white-matter tractography using a probabilistic diffusion tensor atlas: application to temporal lobe epilepsy. Hum. Brain Mapp. 30, 1535–1547.
Holland, D., Dale, A.M., 2011. Nonlinear registration of longitudinal images and measurement of change in regions of interest. Med. Image Anal. 15, 489–497.
Holland, D., Kuperman, J.M., Dale, A.M., 2010. Efficient correction of inhomogeneous static magnetic field-induced distortion in Echo Planar Imaging. Neuroimage 50, 175–183.
Jack Jr., C.R., et al., 2008. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J. Magn. Reson. Imaging 27, 685–691.
Jovicich, J., et al., 2006. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. Neuroimage 30, 436–443.
Joyner, A.H., et al., 2009. A common MECP2 haplotype associates with reduced cortical surface area in humans in two independent populations. Proc. Natl. Acad. Sci. U. S. A. 106, 15483–15488.
Keator, D.B., et al., 2008. A national human neuroimaging collaboratory enabled by the Biomedical Informatics Research Network (BIRN). IEEE Trans. Inf. Technol. Biomed. 12, 162–172.
Kuperman, J.M., et al., 2011. Prospective motion correction improves diagnostic utility of pediatric MRI scans. Pediatr. Radiol. 41, 1578–1582.
Libiger, O., Schork, N.J., 2012. A method for inferring an individual's genetic ancestry and degree of admixture associated with six major continental populations. Front. Genet. 3, 322.
McCarty, C.A., et al., 2014. PhenX RISING: real world implementation and sharing of PhenX measures. BMC Med. Genomics 7, 16.
Nelson, M.R., et al., 2008. The population reference sample, POPRES: a resource for population, disease, and pharmacological genetics research. Am. J. Hum. Genet. 83, 347–358.
Nievergelt, C.M., et al., 2014. Chip-based direct genotyping of coding variants in genome wide association studies: utility, issues and prospects. Gene 540, 104–109.
Norden-Krichmar, T.M., 2014. Correlation analysis of genetic admixture and social identification with body mass index in a Native American community. Am. J. Hum. Biol. 26, 347–360.
Salat, D.H., et al., 2009. Regional white matter volume differences in nondemented aging and Alzheimer's disease. Neuroimage 44, 1247–1258.
Ségonne, F., et al., 2004. A hybrid approach to the skull stripping problem in MRI. Neuroimage 22, 1060–1075.
Ségonne, F., Pacheco, J., Fischl, B., 2007. Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. IEEE Trans. Med. Imaging 26, 518–529.
Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A non-parametric method for automatic correction of intensity non-uniformity in MRI data. IEEE Trans. Med. Imaging 17, 87–97.
Westlye, L.T., et al., 2009. Increased sensitivity to effects of normal aging and Alzheimer's disease on cortical thickness by adjustment for local variability in gray/white contrast: a multi-sample MRI study. Neuroimage 47, 1545–1557.
White, N., et al., 2010. PROMO: real-time prospective motion correction in MRI using image-based tracking. Magn. Reson. Med. 63, 91–105.
Xing, J., et al., 2009. Mobile elements create structural variation: analysis of a complete human genome. Genome Res. 19, 1516–1526.