NeuroImage xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

NeuroImage

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

Northwestern University schizophrenia data sharing for SchizConnect: A longitudinal dataset for large-scale integration

Alex Kogan [a,*], Kathryn Alpert [a], Jose Luis Ambite [b,c,d], Daniel Marcus [e], Lei Wang [a,f]

[a] Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
[b] Information Sciences Institute, University of Southern California, Marina del Rey, CA, USA
[c] Digital Government Research Center, USA
[d] Department of Computer Science, University of Southern California, Los Angeles, CA, USA
[e] Neuroinformatics Research Group (NRG) at the Washington University School of Medicine, St. Louis, MO, USA
[f] Department of Radiology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

ARTICLE INFO

Article history:
Accepted 2 June 2015
Available online xxxx

Keywords:
Neuroinformatics
Schizophrenia data
Research datasets
Data mediation and integration
Data sharing
XNAT

ABSTRACT

In this paper, we describe an instance of the Northwestern University Schizophrenia Data and Software Tool (NUSDAST), a schizophrenia-related dataset hosted at XNAT Central, and the SchizConnect data portal used for accessing and sharing the dataset. NUSDAST was built and extended upon existing, standard schemas available for data sharing on XNAT Central (http://central.xnat.org/). With the creation of SchizConnect, we were able to link NUSDAST to other neuroimaging data sources and create a powerful, federated neuroimaging resource.

Introduction

Schizophrenia is a complex disease with heterogeneous clinical, behavioral, cognitive and genetic manifestations, and sharing of datasets is becoming essential in order to test hypotheses that can capture its variability and complexity (Poline et al., 2012). One example is the discovery of microRNA137 that succinctly illustrates the importance of data sharing: using computational biology techniques, Potkin et al. (2010) combined two previously published, separate datasets and discovered microRNA137 as a risk factor for schizophrenia. It should be noted that neither of the two distinct datasets had identified microRNA137. In a later confirmatory report on 51,695 individuals confirming microRNA137, the International Schizophrenia Consortium proclaimed that a new "cause" of schizophrenia had been found (Ripke et al., 2011).

In this paper, we describe the Northwestern University Schizophrenia Data (NUSDAST) (Wang et al., 2013) as part of SchizConnect, an NIH-funded neuroimaging resource for large-scale data sharing for schizophrenia research. With 451 subjects, the majority of whom have archived longitudinal data, NUSDAST is one of the largest single-site, single-platform neuroimaging datasets related to schizophrenia, making it a uniquely important resource to share with the research community. NUSDAST will benefit the neuroscience community in many ways. First, scientists will be able to use these data to generate or test new hypotheses related to abnormalities of brain structures and neural networks in individuals with schizophrenia. Second, scientists will be able to rapidly replicate findings produced using their own datasets. Third, the data could be used to test and validate new brain mapping tools.

What is available?

The data presented in NUSDAST were collected through the support of two NIH-funded grants on schizophrenia: (1) Neuromorphometry in Schizophrenia (R01-MH056584), and (2) Conte Center for the Neuroscience of Mental Disorders (P50 MH071616). Through these projects, our group has collected high-resolution structural MRI datasets from large cohorts of subjects using the same scanner platform and sequence protocols. We have also collected detailed clinical, cognitive and genetic information from these subjects.

Subjects

NUSDAST includes de-identified data from 451 individuals with schizophrenia, their non-psychotic siblings, comparison subjects and their siblings. Neuroimaging data exist for 368 individuals. Longitudinal neuroimaging data are also available on 171 individuals with schizophrenia (m/f = 114/57, age at baseline = 33.8 ± 12.5 years) and 170

* Corresponding author at: Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, 710 N. Lake Shore Dr, Abbott Hall 1322, Chicago, IL 60611.
E-mail address: a-kogan@northwestern.edu (A. Kogan).

http://dx.doi.org/10.1016/j.neuroimage.2015.06.030
1053-8119/

controls (m/f = 86/84, age at baseline = 31.4 ± 13.8 years). Within this group of subjects, 18 individuals with schizophrenia and 30 controls returned for a second follow-up (i.e., 3 time points). The average (SD) follow-up interval was 2.19 (0.82) years for individuals with schizophrenia and 2.28 (0.49) years for the controls. De-identification consisted of stripping HIPAA-mandated identifiable information in research data (such as name, initials, and phone numbers, etc.). Procedures to further anonymize imaging data such as defacing were not performed in order to share the same imaging data that we used in our publications so that others can replicate our findings using their own algorithms if they so desire. See Table 1 below for baseline information.

Clinical data includes information based on specific criteria for clinical stability (Rastogi-Cruz and Csernansky, 1997) and clinical rating scales such as the Scale for the Assessment of Positive Symptoms (Andreasen, 1984) and Scale for the Assessment of Negative Symptoms (Andreasen, 1983) (see Table 1 below for baseline information). Domains of psychopathology (i.e., psychotic symptoms, disorganized symptoms, and negative symptoms) (Andreasen et al., 1995) based on raw scales are also included. The reliability and practicality of using these scales in large populations of schizophrenic patients have been demonstrated by Andreasen, et al. (1995). Symptom assessments were performed by personnel specially trained for this purpose. Inter-rater reliability was monitored regularly for all rating scales and rater training sessions, including the conjoint assessment of difficult cases, were held weekly. In these sessions, a variety of patients were interviewed in a group. Two established raters reached a consensus of item scores after the interview was completed, and then this "gold standard" score was compared with the rest of the group. New raters were trained by first participating in a minimum of six of these sessions. They were allowed to participate in ratings only after they had demonstrated satisfactory agreement with trained personnel.

MRI data

All MR scans were collected using the same 1.5 T Vision scanner platform (Siemens Medical Systems) at each time point. The Vision scanner had actively shielded gradients and echo-planar capability with very high gradient linearity (<0.4% over a 22-cm diameter spherical volume compared to 2%–5% over 22-cm for our other scanners), which yielded anatomical images with virtually no distortion (<0.4% voxel displacement), critical to analyses of neuroanatomical structures. Using the same scanner provided stable longitudinal MR data throughout the entire period of data collection from 1998 to 2006.

Acquisition of all scans was performed at the Mallinckrodt Institute of Radiology at Washington University School of Medicine, where scanner stability (e.g., frequency, receiver gain, transmitter voltage, SNR) and artifacts were regularly monitored. Phantoms of known size were scanned to confirm image dimensions. Further tests and adjustments (shims, gradient calibrations, EPI switch delays, etc.) were made as needed. During each scan session, a small standardization object (i.e., vitamin E gelcap) was placed on the left side of the forehead for each subject to clearly indicate laterality in the scans. Each scan session included a high-resolution T1-weighted turbo-FLASH scan (Venkatesan and Haacke, 1997), multiple (2–4) MPRAGE scans, and MPRAGE average. Source MR scan data were in Siemens MAGNETOM VISION IMA format and subsequently converted into Analyze format using in-house software. Since Analyze format images may cause confusion with regard to laterality, even though the abovementioned vitamin E gelcap information may help verify laterality, all Analyze format images are being converted into NIFTI format and uploaded. The multiple MPRAGE images for each subject are aligned with the first image and averaged to create a low-noise image volume (Buckner et al., 2004. See Table 2 for detailed scan protocol parameters.

Neuroimaging metadata

In our template-based brain mapping applications, we have focused on a network of structures previously implicated in the pathophysiology of schizophrenia (Weinberger et al., 1992; Csernansky and Bardgett, 1998; Goldman-Rakic, 1999). This network included regions with the prefrontal cortex (e.g., middle frontal gyrus, Brodman area 46) (John et al., 2006; Harms et al., 2010), the cingulate gyrus (Qiu et al., 2007; Wang et al., 2007a), the hippocampus (Wang et al., 2001; Csernansky et al., 2002), the parahippocampal gyrus (Karnik-Henry et al., 2012), as well as the thalamus (Csernansky et al., 2004a; Harms et al., 2007; Smith et al., 2011) and the basal ganglia (Mamah et al., 2007; Wang et al., 2008), which directly or indirectly link these structures via cortical–subcortical connections. We have constructed manual segmentation datasets for all these structures, which can be used for the validation of new computational methods. In addition, we have also used FreeSurfer (Desikan et al., 2006) to generate cortical surface parcellations and measures of cortical regional volume, thickness and surface area (Cobia et al., 2011).

Template data

The templates for the hippocampus and amygdala were generated using a T1-weighted MR scan collected in a healthy subject (Wang et al., 2008). The templates for the thalamus and basal ganglia (caudate nucleus, putamen, nucleus accumbens and globus pallidus) were generated using a seven-time averaged T1-weighted MR scan collected in another healthy subject (Wang et al., 2007b). These segmentations were manually performed using Analyze software in these scans by consensus of experts using atlas guidelines (Duvernoy, 1988, 1991; Mai et al., 1997). Surfaces (.byu format) of each structure were generated using the marching cubes algorithm (Lorensen and Cline, 1987; Claudio and Roberto, 1994). The left and right surfaces have corresponding nodes so that analyses of shape asymmetry can be performed. These templates and subject-level landmark and surface data (below) have been shared here for the purpose of replication and facilitating potential, further modeling work (Haller et al., 1997; Csernansky et al., 2004a; Wang et al., 2007b.

Landmark and surface data

Mapping of the template MR scan occurred in a two-step process. First, it was coarsely aligned to each target scan using landmarks, and then the diffeomorphic map was applied. Surfaces for subcortical structures in the target scans were generated by carrying the template surfaces through these maps (Joshi et al., 1997; Csernansky et al., 2004b).

To facilitate our template-based mapping, global and local (i.e., structure-dependent) neuroanatomical landmarks were placed

Table 1
Subject characteristics at baseline.

	Schizophrenia Subjects	Control Subjects	Schizophrenia Siblings	Control Siblings
N	171	170	44	66
Age at baseline (years)	33.8(12.5 [17–63])	31.4(13.8 [13–67])	N/A	N/A
Gender (male/female)	114/57	86/84	21/23	16/50
Race (Caucasian/African-American/other)	90/78/3	61/105/2	17/27/0	16/50/0
Global SAPS Score	11.1 (12.7 [0–81])	0.06(0.3 [0–4])	0.5 (1.3 [0–7])	N/A
Global SANS Score	9.6 (10.7 [0–62])	0.04(1.7 [0–19])	2.3 (4.8 [0–38])	N/A

Please cite this article as: Kogan, A., et al., Northwestern University schizophrenia data sharing for SchizConnect: A longitudinal dataset for large-scale integration, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.06.030