ADNI Alzheimer's Disease Neuroimaging Initiative

Introduction and Procedures for Accessing Data from
Whole Genome Sequencing of ADNI Subjects
Posted September 18, 2013

ADNI WGS Overview

In July 2012 the Brin-Wojcicki Foundation and the Alzheimer's Association donated funds to support whole genome sequencing (WGS) of 818 ADNI participants (at the time, 128 with AD, 415 with MCI, 267 controls and 8 of uncertain diagnosis). Samples were sent to Illumina where non-CLIA WGS, as well as Illumina Omni 2.5M genome-wide association study (GWAS) single nucleotide polymorphism (SNP) arrays, were performed on each sample and completed in the spring of 2013. Basic quality control checks were then performed and others continue to be performed (see Quality Control section below). This document announces the availability of the data to qualified investigators and provides instructions on how to obtain it.

Informed Consent for Sharing WGS Data

ADNI has always pursued a philosophy of sharing data with investigators who abide by the ADNI Data Use Agreement. As part of this philosophy, very broad consent language for sharing ADNI data has been part of ADNI consent procedures from the start. This has worked well with over 4,100 investigators accessing the ADNI database, creating over 702 scientific reports that have been submitted to scientific journals with 418 manuscripts published or in-press papers to date. The data that has been accessed from ADNI includes diagnostic categories, longitudinal cognitive measures, structural, functional and molecular imaging, plasma and CSF biomarker data and genotype data including SNPs and CNVs from genome-wide arrays.

Since sharing of genome sequence data may raise additional privacy issues, we conducted an extensive review of all iterations of ADNI, ADNI-GO, and ADNI-2 consent documents provided by all participating sites. This consent document analysis characterized the information provided on each form regarding: what data would be shared, who it would be shared with, for what purposes would sharing be allowed in the future, and whether the consent document provided any additional constraints on data sharing. A database was created to summarize and capture relevant consent language in each document for future reference. Our review has led to the sequester of sequence data on 5 subjects from one site whose samples had been sent for WGS data sharing where we judged consent language to be more restrictive than appropriate for our intended sharing. Ultimately, these 5 subjects may be re-consented and their sequences data released, but until then their data will not be included in the ADNI WGS data release. All other consent documents that we reviewed were judged to be adequate and appropriate in their provision for sharing WGS data with members of the scientific community who agree to follow the ADNI data sharing guidelines.

Storage and Quality Control of WGS Data

ADNI WGS data has now been stored at three physical locations: LONI at the University of Southern California in Los Angeles, CA, the Broad Institute in Boston, MA and Indiana University in Indianapolis, IN. Another copy of the ADNI WGS data will soon be shared with the National Institute on Aging national genetics data repository (NIAGADS, see www.niagads.org). This is to provide backup for the data at several physical locations where investigators may work onsite with it, and the ability to combine these data with other sequencing data that are being collected as part of the national Alzheimer's Disease Sequencing Project. In all of these locations, investigators will be required to sign the ADNI Data Use Agreement before accessing the data.

The following QC steps have been conducted. File integrity was checked during upload, followed by a check of subject gender using current (Omni2.5M) GWAS data and self-reported gender information for 818 participants. Identity was cross-checked by comparing concordance (1) among a subset of fingerprint SNPs between prior genotype data and Omni2.5M on 816 of 818 individuals and (2) among all quality-controlled markers between Omni2.5M and VCF on all 818 individuals in the sequencing

project. These QC measures resulted in temporary sequester of 4 cases while anomalies are being investigated. When these are resolved, they will be added back into the pool of available data. Additional quality control measures assessed the entire sequencing dataset in relation to properties of naturally occurring human mutations, known biases in human polymorphism data, and distribution of allele frequencies and the proportion of variants seen in both chromosomal copies as opposed to variants seen in a single copy. Continued quality control, including experiments to compare these sequences with re-aligned and re-called sequences using Broad pipelines, are being conducted and will be shared here and with users when completed.

Application for ADNI Data

Procedures for applying for access to WGS data will be identical to that required for access to existing ADNI data. The intent of these policies is to facilitate data sharing with all interested and qualified investigators, to encourage academic productivity, and to provide a mechanism for tracking and archiving data requests, intended analyses and publications related to and resulting from ADNI WGS data. All users must formally agree to the ADNI data use agreement and follow ADNI publication procedures regardless of whether the WGS data is obtained through ADNI directly, through NIAGADS or through one of the centers where the data are physically stored. To view ADNI data request procedures and download the data use agreement, please visit: http://adni.loni.usc.edu/data-samples/access-data/.

Access to Genomic Data

With the nine sequestered cases as above, data will be available on 809 individuals. Three different types of files will be available for users: VCF files, BAM files and 2.5M array files. VCF files are now available for download from LONI through the ADNI data access website. Variant data in the BAM and VCF files were called using Illumina's proprietary CASAVA SNP caller (see http://res.illumina.com/documents/products/technotes/technote_snp_caller_sequencing.pdf). Some users may find this adequate to their needs, while others may wish to re-call variants or even realign the reads using other toolkits.

Persons wishing to obtain the BAM and other raw data files should note that the entire dataset will be between 160-200 terabytes and cannot be transmitted over the internet. Users wishing to have a copy of this dataset should send a hard drive array capable of storing 200 terabytes of data to Arthur Toga at Keck School of Medicine at USC, 2001 N. Soto Street, SSB1-102, Los Angeles, CA 90032, phone 323-442-7246). Requests should indicate the name and email address of the PI authorized to access ADNI data. Users are advised to check with LONI prior to making such arrangements to insure that the equipment will meet compatibility requirements. The entire dataset will be loaded onto this hard drive for you and sent back to you. Users must purchase or otherwise obtain this hard drive for themselves and cover all shipping. LONI will take all reasonable precautions, but cannot be responsible for damage to equipment sent in this manner.

For more information:

On overall process, data agreement and consent, contact
Robert C. Green, MD, MPH, Brigham and Women's Hospital and Harvard Medical School
Email: rcgreen@genetics.med.harvard.edu

On identity checks and SNP concordance among samples, contact
Andrew J. Saykin, PsyD, Indiana University School of Medicine
Email: asaykin@iupui.edu

On procedures for data download or hard drive shipping, contact
Arthur Toga, PhD, University of Southern California
Email: toga@loni.usc.edu