

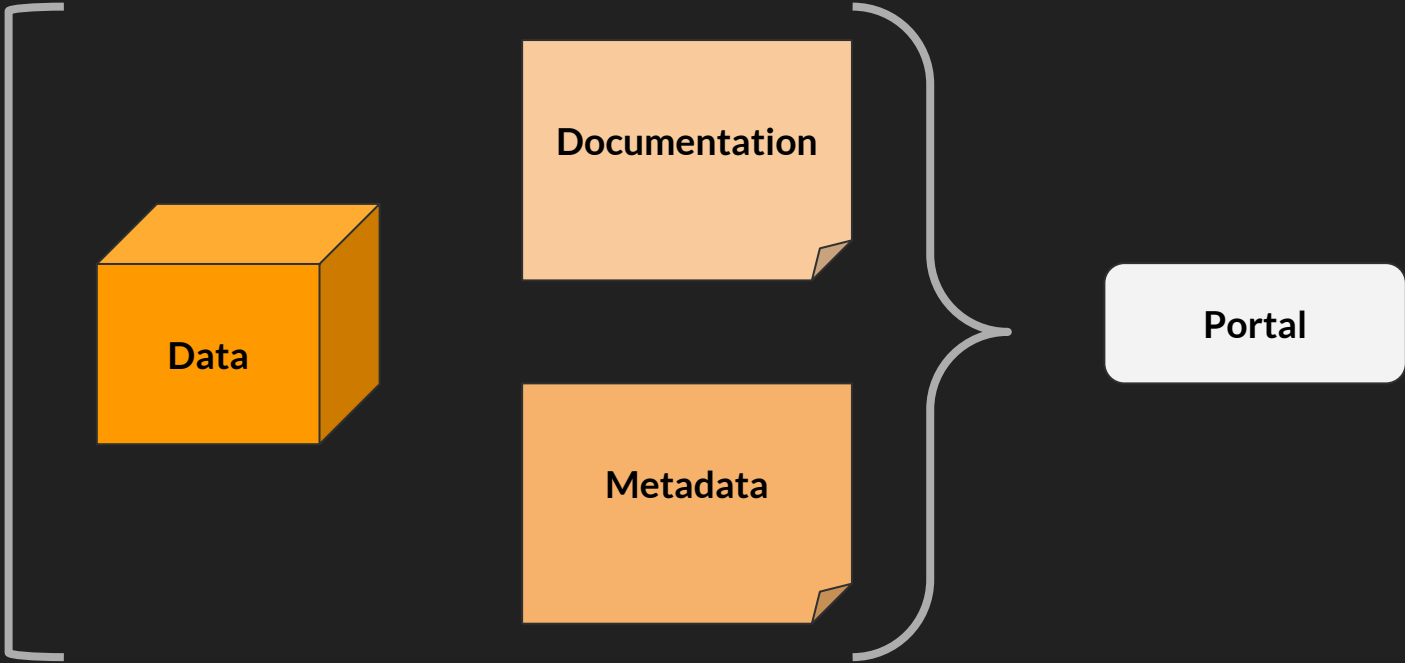
Study Content

Documentation

Metadata

Data

Data Ingress



Documentation

**Study
Description**

**Methods
Description
Assay 1**

**Methods
Description
Assay 2**

Documentation provides:

- A summary of the data source
- A summary of method(s) used for data generation

Metadata

Individual

Biospecimen

Assay(s)

Manifest

Structured Metadata is provided through 4 files:

- *Individual* – key variables that describe the individual the data comes from
- *Biospecimen* – key variables that describe the specimen the data has been generated from
- *Assay* – key variables that describe the assay used to generate data from the specimen (or an assessment done on an individual).
- *Manifest* – key variables that will be used as file annotations. Is used for upload of data and metadata files.

Metadata

Individual

Biospecimen

Assay(s)

Manifest

individualID	sex	ageDeath	diagnosis
participant1	male	78	Alzheimer Disease
participant2	female	82	control

Human cohort example

Metadata

Individual

Biospecimen

Assay(s)

Manifest

individualID	specimenID	organ	tissue
participant1	p1_spec1	brain	dorsolateral prefrontal cortex
participant2	p2_spec1	brain	anterior cingulate cortex

Brain tissue specimen example

Metadata

Individual

Biospecimen

Assay(s)

Manifest

specimenID	platform	RIN	libraryPrep
p1_spec1	HiSeq2500	9.2	polyAselection
p2_spec1	HiSeq2500	8.7	polyAselection

RNA sequencing example

Metadata

Individual

Biospecimen

Assay(s)

Manifest

path	parent	individualID	specimenID
~/p1.fastq	syn123	participant1	p1_spec1
~/p2.fastq	syn123	participant2	p2_spec1

The manifest provides metadata about the files and serves two purposes:

- Uploading the data in bulk via the Synapse client. Specifies the directory data will be uploaded
- Adding an initial set of annotations on the files when they are uploaded. The DCC will add annotations from the other metadata files. File annotations allow data queries in the portal

Metadata

Individual

Biospecimen


Assay(s)

Manifest

The first three metadata files can be joined by the individualID and specimenID columns to create the full metadata set.

individualID	specimenID	sex	ageDeath	...	organ	tissue	platform	RIN	...
participant1	p1_spec1	male	78	...	brain	dorsolateral prefrontal cortex	HiSeq2500	9.2	...
participant2	p2_spec1	female	82	...	brain	anterior cingulate cortex	HiSeq2500	8.7	...

Final Product – Documentation

 Study Details

 Study Data

STUDY DESCRIPTION

The Mount Sinai Brain Bank (MSBB) study

Brain specimens were obtained from the [Mount Sinai/JJ Peters VA Medical Center Brain Bank \(MSBB\)](#) which holds over 1,700 samples. This cohort was assembled after applying stringent inclusion/exclusion criteria and represents the full spectrum of disease severity. Neuropathological assessments are performed according to the Consortium to Establish a Registry for [Alzheimer's Disease \(CERAD\) protocol](#) and include assessment by hematoxylin and eosin, modified Bielschowski, modified thioflavin S, and anti- β amyloid (4G8), anti-tau (AD2) and anti-ubiquitin (Daka Corp.). Each case is assigned a [Braak AD-staging score](#) for progression of neurofibrillary neuropathology. Quantitative data regarding the density of neuritic plaques in the middle frontal gyrus, orbital frontal cortex, superior temporal gyrus, inferior parietal cortex and calcarine cortex are also collected as [described](#). Clinical dementia rating scale (CDR) and mini-mental state examination (MMSE) severity tests are conducted for assessment of dementia and cognitive status. Final diagnoses and CDR scores are conferred by consensus. Based on [CDR classification](#), subjects are grouped as no cognitive deficits (CDR = 0), questionable dementia (CDR = 0.5), mild dementia (CDR = 1.0), moderate dementia (CDR = 2.0), and severe to terminal dementia (CDR = 3.0–5.0). Covariates including demographic and neuropathological data were collected on the samples used for this project including postmortem interval, race, age of death, clinical dementia rating, clinical neuropathology diagnosis, CERAD, Braak, sex, and a series of neuropathological variables. See the Mount Sinai cohort of [large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease](#) for a detailed description of the study and the data.

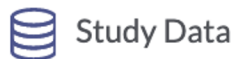
METHODS: GENOMIC VARIANTS (WHOLE EXOME SEQUENCING)

Library preparation : Genomic DNA samples are sheared into small DNA fragments and libraries were prepared with Illumina compatible adapters and indices. Biotinylated cRNA baits were incubated with the library for 16 hours and then targeted regions were selected using magnetic streptavidin beads. Targeted regions were amplified, producing an exome enriched library.

Sequencing : The sequence ready libraries were loaded onto Illumina HiSeq 2500 System with 125 bp paired-end sequencing on V4 flow cell. One trio (3 samples) was pooled per sequencing lane, aiming for 80X mean coverage per sample.

Data processing : The raw sequence reads were aligned to human genome hg19 with the BWA aligner. Then the sequence variants were called using the DNaseq Variant Analysis workflow of GTAK Best Practices version 3. Following the QC described BELOW, samples with QC actions "Remap" or "Exclude", and plink genotypes from individuals with missing rate > 0.5, SNPs with MAF < 0.01, missing genotyping rate > 0.5, or HWE test P value < 0.001 were removed.

Final Product – Data and Metadata



Study Data (6,411)

Data Type ⌵ ↕ ✕ Assay ⌵ ↕ ✕ Org



■ (40.5%) geneExpression
■ (29.9%) proteomics
■ (19.9%) chromatinActivity



■ (40.5%) rnaSeq
■ (25%) TMT quantitation
■ (19.9%) ATACSeq

Showing 6411 results

Access	Id	Study	Data Type	Assay	Organ	Tissue	Specie
View Terms	processing_code_alignment_star.lsf	MSBB	geneExpression	rnaSeq	-	-	Human
View Terms	processing_code_quantitation_feat...	MSBB	geneExpression	rnaSeq	-	-	Human
View Terms	BM_22_213.accepted_hits.sort.coo...	MSBB	geneExpression	rnaSeq	brain	superior temporal gyrus	Human
View Terms	BM_22_213.unmapped.fastq.gz	MSBB	geneExpression	rnaSeq	brain	superior temporal gyrus	Human
View Terms	BM_22_222.accepted_hits.sort.coo...	MSBB	geneExpression	rnaSeq	brain	superior temporal gyrus	Human
View Terms	BM_22_222.unmapped.fastq.gz	MSBB	geneExpression	rnaSeq	brain	superior temporal gyrus	Human
View Terms	BM_22_229.accepted_hits.sort.coo...	MSBB	geneExpression	rnaSeq	brain	superior temporal gyrus	Human
View Terms	BM_22_229.unmapped.fastq.gz	MSBB	geneExpression	rnaSeq	brain	superior temporal gyrus	Human
View Terms	BM_22_230.accepted_hits.sort.coo...	MSBB	geneExpression	rnaSeq	brain	superior temporal gyrus	Human

Metadata Files (9)

Showing 9 results via:

Access	Id	Metadata Type
View Terms	MSBB_individual_metadata.csv	individual
View Terms	MSBB_biospecimen_metadata.csv	biospecimen
View Terms	MSBB_assay_TMT_metadata.csv	assay
View Terms	MSBB_assay_ATACseq_metadata.csv	assay
View Terms	MSBB_assay_MethylationArray.csv	assay
View Terms	MSBB_assay_proteomics_metadata.csv	assay
View Terms	MSBB_assay_wholeGenomeSeq_metadata.csv	assay
View Terms	MSBB_assay_rnaSeq_metadata.csv	assay
View Terms	MSBB_assay_exomeSeq_metadata.csv	assay