The Aging, Dementia and Traumatic Brain Injury Study is a detailed neuropathologic, molecular and transcriptomic characterization of brains of control and TBI exposure cases from a unique aged population-based cohort from the Adult Changes in Thought (ACT) study. This study was developed by a consortium consisting of the University of Washington, Kaiser Permanente Washington Health Research Institute, and the Allen Institute for Brain Science, and was supported by the Paul G. Allen Family Foundation. This freely available resource (http://aging.brain-map.org/) presents a systematic and extensive data set of study participant metadata, quantitative histology and protein measurements of neuropathology, and RNA sequencing (RNA-seq) analysis of hippocampus and neocortex. In total, 377 samples collected from cortical grey (parietal and temporal) and white matter (parietal) and hippocampus from a total of 107 brains are presented in this resource.

In this NIAGADS data set we present the raw data used to derive the RNA sequencing data presented as part of the Aging, Dementia and Traumatic Brain Injury Study. For each sample, eight files are presented:

- *[Code]_SAMPLE_[Code]_R[1/2]_001.fastq.gz* = zipped fastq files for all reads (i.e., the fastq file coming directly from the sequencer after demultiplexing).
- *SAMPLE.aligned_transcriptome.sorted.bam* = bam file for alignment to transcriptome
- *SAMPLE.aligned_transcriptome.sorted.bam.bai* = index file
- *SAMPLE.aligned_non_transcriptome.sorted.bam* = bam file for reads not in above file that align to genome
- *SAMPLE.aligned_non_transcriptome.sorted.bam* = index
- *SAMPLE.unaligned_R[1/2].fastq.gz* = zipped fastq files for all reads not mapping to genome or transcriptome.

In addition, an excel file (*ConversionCode_AllenAgingDementiaTBI.csv*) is included, which makes it possible to match the data across all the file types, and with the data on the website.

Specific methodological details are available on the "Documentation" tab at http://aging.brain-map.org/. In short, collection of tissue samples from temporal and parietal neocortex, parietal white matter, and hippocampus was done by manual macrodissection. Tissue was immediately transferred to prepared tubes where RNA was isolated using the RNeasy Lipid Tissue Mini Kit (Qiagen #78404) as per manufacturer's instructions. RNA was then quantified on a Nanodrop 8000 spectrophotometer (Thermo Scientific, Wilmington, DE) and normalized to 5ng/µl before RNA QC was performed using a Bioanalyzer (Agilent Technologies) and RNA Integrity Number (RIN) was recorded. Total RNA (250ng) was used as input into the Illumina TruSeq Stranded Total RNA Sample Prep Kit (RS-122-2203), which uses random hexamer first strand cDNA synthesis and includes rRNA depletion (Ribo-Zero Gold rRNA depletion kit) and fragmentation. At the time of project inception, this sequencing strategy provided the most reliable option for quantification of transcriptomic reads from tissue of widely varying quality, allowing the broadest inclusion of donors from the ACT cohort. External RNA Controls Consortium (ERCCs) at a 1:10,000 dilution were spiked into each sample. RNA sequencing was done on Illumina HighSeq 2500 using v4 chemistry, producing a minimum of 30

million 50bp paired-end clusters per sample. Expression Analysis, Inc. (Morrisville, NC) performed both the TruSeq Stranded Sample Prep as well as the Illumina sequencing.

Raw read (fastq) files were aligned to the GRCh38.p2 human genome (current as of 01/15/2016). Illumina sequencing adapters were then clipped using the fastqMCF program, and then mapped to the transcriptome using RNA-Seq by Expectation-Maximization (RSEM) using default settings except for two mismatch parameters: bowtie-e (set to 500) and bowtie-m (set to 100). RSEM aligns reads to known isoforms and then calculates gene expression as the sum of isoform expression for a given gene, assigning ambiguous reads to multiple isoforms using a maximum likelihood statistical model. Reads that did not map to the transcriptome were then aligned to the hg38 genome sequence using Bowtie with default settings, after which remaining unmapped reads were mapped to ERCCs. Reads mapping to transcriptome and genome were saved as bam files and remaining unaligned reads were stored in zipped fastq files, as indicated above.