README file for MIDUS Gene Expression Project (P6)
January 2021

The purpose of this memo is to provide basic information about the Gene Expression Project (P6) data and documentation for the MIDUS Refresher that are publicly available via the MIDUS Colectica Portal. Note, these data are not currently available through ICPSR.

Data and documentation for other MIDUS projects are available through:

1. The public archive at ICPSR (http://www.icpsr.umich.edu/icpsrweb/NACDA/studies/36532/version/2)
2. The MIDUS Colectica Portal (http://midus.colectica.org/) which houses interactive codebooks for all the publicly available MIDUS projects. The Portal also includes search and explore functions, links to documentation, and a custom download function.

A. What data files are available?

Two data files containing gene expression data have been generated. The first file contains data on expression of genes involved in the Conserved Transcriptional Response to Adversity (CTRA) and biological aging (the senescence marker $p16^{INK4a}$/*CDKN2A* and composite scores capturing the senescence-associated secretory phenotype/SASP and DNA damage response/DDR) and is available through the Portal as an SPSS dataset:

MR_P6_RNA_SCORE_N863_MRID_20210106.sav

The second file is also an SPSS dataset and contains expression values for 51 individual CTRA indicator genes along with relevant sample quality metrics & RNA covariates. Due to the sensitive nature of this data, the datafile is not available through the Portal. More information about accessing this datafile can be found here:
(http://midus.wisc.edu/midus_genetic_data.php/)

B. What is the structure of the MIDUS RNAScore Dataset?

The RNAScore datafile is a SPSS (ver. 23) dataset comprised of CTRA and CDKN2A and SASP and DDR composite scores along with relevant sample quality metrics and RNA covariates. The file contains 26 variables for 863 cases (MainRDD and Milwaukee African American) that completed the MIDUS Refresher Biomarker project (P4).

The dataset has been cleaned, meaning that value ranges and skip patterns have been checked. Variables have been named according to the MIDUS Refresher naming conventions. All variables include labels to aid interpretation. Value labels have been applied where appropriate and discrete missing values have been defined.

The following documents provide additional information about the data:

MR_P6_Gene Expression Documentation_V2_20210125: provides general information about the RNAScores dataset along with technical and other details about how the scores

are generated. It also includes information about related gene transcript variables.

MR_M3_Naming and Coding Conventions: is posted with the MIDUS Refresher Survey (Project 1) data and describes conventions for naming and coding variables. It is included with the MR Project 1 documentation in the Portal or at ICPSR: (http://www.icpsr.umich.edu/icpsrweb/ICPSR/series/203/studies/36532?archive=ICPSR&sortBy=7

C. Other important information.

ID system. The MIDUS RNAScore dataset only contains Refresher cases thus it uses the 5-digit respondent identification variable called MRID. This ID system is used throughout the publicly available MIDUS Refresher datasets and can be used to link the MIDUS RNAScore dataset to the other Refresher datasets.

****

Please report any errors or inconsistencies you find in the data or documentation to glove@wisc.edu or bradler@wisc.edu

****