

README file for MIDUS Gene Expression Project (P6)

October 2023

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?

- MR1_P6_RNAScores_N863_20210923.sav
 - A SPSS data file that contains CTRA and CDKN2A and SASP and DDR composites scores along with relevant sample quality metrics & RNA covariates;
 - Appropriate for most users;
 - Available through the MIDUS Colectica Portal (<http://midus.colectica.org/>).
- MR1_P6_RNAGenes_N863_20200605.sav
 - A SPSS data file that contains expression values for 51 individual CTRA indicator genes, along with relevant sample quality metrics & RNA covariates;
 - Note, this file does not contain individual indicator genes used to create the cell senescence composites scores. Those will be released separately at a later date;
 - Appropriate for use by investigators with experience in statistical genetics/genomics;
 - Information about accessing this datafile can be found here: [MIDUS Genomic Repository](#).
- MR1_P6_RNA_GeneExpressionLog2Matrix_N543_20231009.txt
 - A text data file that contains expression values for 60,675 distinct gene transcripts (genome-wide);
 - Rows denote genes; columns denote samples;
 - Data represent log2-transformed transcript abundance values (gene transcript counts per million total human transcriptome-mapped RNA sequencing reads), with values floored at 1 transcript per million (0 log2) to suppress spurious variability;
 - Appropriate for use by investigators with experience in statistical genetics/genomics;
 - Information about accessing this data file can be found here: [MIDUS Genomic Repository](#).

B. What is the structure of the MIDUS RNA Score Dataset?

The data file is a SPSS dataset comprised of CTRA composite scores along with relevant sample quality metrics and RNA covariates. The file contains 27 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:

MR1_P6_GeneExpression_Documentation_20231017: provides general information about the CTRA Scores 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 RNA Gene Score 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 RNA Score dataset to the other Refresher datasets.

Please report any errors or inconsistencies you find in the data or documentation to

midus_help@aging.wisc.edu
