README file for MIDUS 3 Gene Expression Project (P6)
February 2024

The purpose of this memo is to provide basic information about the Gene Expression Project (P6) data and documentation for the MIDUS 3 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 (https://www.icpsr.umich.edu/web/ICPSR/series/203)
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?

Three data files containing gene expression data have been generated. The first file contains gene expression scores and is available through the MIDUS Portal:

- *M3_P6_RNA_SCORE_N747_20240216.sav*
  - An SPSS data file that contains CTRA and CDKN2A and SASP and DDR composite scores along with relevant sample quality metrics & RNA covariates. This dataset is available through the MIDUS Colectica Portal (http://midus.colectica.org/).

- *M3_P6_RNA_Gene_n747_20240214.sav*
  - An SPSS data file contains expression values for 51 individual CTRA indicator genes along with relevant sample quality metrics & RNA covariates. Due to the sensitive nature of the data, this dataset is not available through the MIDUS Colectica Portal. Information about accessing this datafile can be found here: MIDUS Genomic Repository.

- *M3_P6_RNA_GeneExpressionLog2Matrix_N672_20240222.txt*
  - A text data file that contains expression values or 60,675 distinct gene transcripts (complete genome-wide transcriptome data). Due to the sensitive nature of the data, this dataset is not available through the MIDUS Colectica Portal. Information about accessing this datafile can be found here: MIDUS Genomic Repository.

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

The data file is an SPSS dataset comprised of CTRA composite scores along with relevant sample quality metrics and RNA covariates. The file contains 28 variables for 747 cases (MIDUS Core sample and Milwaukee African American sample) that completed the MIDUS 3 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 3 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:

M3_P6_GeneExpression_Documentation_20240219: 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 and MIDUS 3 Survey (Project 1) data (https://midus-study.github.io/public-documentation/M3P1/Documentation/MR_M3_NamingAndCodingConventions_20220531.pdf) and describes conventions for naming and coding variables. It is included with MR1 and M3 survey project documentation in the portal or at ICPSR.

C. Other important information.

ID system. The M3 RNA Score dataset uses the M2ID variable to identify individual respondents. This variable can be used to merge any MIDUS datasets derived from the original longitudinal sample.

```
************************************************************************
         Please report any errors or inconsistencies you find in the data or documentation to
                    MIDUS HelpDesk (http://midus.wisc.edu/helpdesk.php).
************************************************************************
```