R Code READ ME for Haggard et al. "Using a high-throughput steroidogenesis assay to predict effects on estrogen and androgen production or overall perturbation of the steroid biosynthesis pathway"

**Disclaimer:** The United States Environmental Protection Agency (EPA) GitHub project code is provided on an "as is" basis and the user assumes responsibility for its use. EPA has relinquished control of the information and no longer has responsibility to protect the integrity, confidentiality, or availability of the information. Any reference to specific commercial products, processes, or services by service mark, trademark, manufacturer, or otherwise, does not constitute or imply their endorsement, recommendation or favoring by EPA. The EPA seal and logo shall not be used in any manner to imply endorsement of any commercial product or activity by EPA or the United States Government.

The R code used for all data analysis and figures follows an R project template, and the “R Project.Rproj” file should be used, as this makes every operation and file load to the relative path of the R Project directory.

“ANOVA” contains the output for all of the OECD ANOVA analysis, including figures and supplemental files/figures.

“Mahalanobis Distance” contains all of the output for the Mahalanobis Distance analysis, including figures and supplemental files/figures.

“Master Data Table” contains the txt file of the processed and MTT-filtered H295R assay data.

“Misc” contains miscellaneous output from some of the R scripts.

“RData” contains all of the R output and data files that are used and presented in the manuscript. Note that all of the R scripts will reference specific time-stamped RData files when they are loaded and will need to be adjusted accordingly when re-running all of the analysis.

“Scripts” contains all of the R scripts needed to run the analysis and can be opened in a manual R session or via the R Studio GUI when running from the “R Project.Rproj”. Note: Each script will output both RData files and txt files as time-stamped output, see the individual scripts to reference where these output are placed in the R Project environment.

The following R scripts should be run first in order to generate the figures and supplemental tables presented in the manuscript (Note: the *“H295R\_master\_table\_2017-08-08.RData”* file is needed for these scripts to work):

* *“H295R\_manova\_to\_estimate\_residual\_matrix.R”* – this script generates the matrix of residuals of the H295R data by block using a multivariate linear model in order to calculate the Mahalanobis distance. This script should be run first.
* *“mahalanobis\_distance\_calculation\_and\_Supp9.R”* – this scripts uses the output from the *“H295R\_manova\_to\_estimate\_residual\_matrix.R”* to calculate the Mahalanobis distance, maxmMd, and critical limits. This scripts should be run second.
* *“OECD\_GLOBAL\_anova\_script\_and\_Supp4\_Supp5\_Supp6.R”* – performs the ANOVA analysis and all related supplemental tables. This code should be run third.
* All remaining R scripts can then be run, in any order, as the above three scripts generates all of the necessary RData and txt files (alongside the *“H295R\_master\_table\_2017-08-08.RData”* file) needed for subsequent analyses and figure generation.