MEA Acute pre-processing notes

Last updated August 4, 2023 by Amy Carpenter

# Things to keep in mind

## Culture vs experiment date in MEA Acute

* Culture.date = date the culture was prepared. Usually 2 groups (consisting of 3 plates each) are prepared on each culture.
* Experiment.date = date of recordings and treatment. This is almost always 13 or 15 days after the culture.date (one group is tested on DIV 13, the other on DIV 15). Presumably, the activity on DIV 13 and 15 is comparable.

# Getting set up

## (Recommended) To check with lab technicians before start pre-processing

* Make sure that the culture dates in folder names are present and correct (I know sometimes a culture date may be on one day, then it gets pushed back. So it’s helpful if the culture dates are updated in the folder names).
  + The 8-digit culture dates are pulled from the folder names
* Make sure that raw LDH files contains the word “LDH” in file name
  + The raw data will be read from these files directly. Any files in the “Culture Cytotoxicity” folder that do not include “LDH” in the name will be assumed to correspond to the Alamar Blue assay.

## Set up the project output folder

Make a copy of the folder ‘Template\_project’, and rename the folder and all sub-files with the project name in place of “Template.” Also make a copy of a recent run\_me script to use as a template (e.g., *mea-acute-neural-stats-to-mc0-scripts/run\_me\_TSCA2019.Rmd* and rename with the current project name.

## Prepare notebook section

See the OneNote notebook under L:\Lab\NHEERL\_MEA\CCTE\_Shafer pre-process for TCPL\MEA\_acute

Can create a copy of the “Template” to save notes related to the pre-processing

## Review lab notebook for well quality notes/data processing notes!

Check out the lab notebook and document any well quality issues! See below

## Defining the well quality

Three well quality columns are created within the run\_me and merged to make the final well quality determination:

* Wllq\_by\_recording – set by add\_wllq\_by\_recording()
* Wllq\_by\_well – determined by user entering information from lab notebook in the table “[project\_name]\_well\_quality\_table\_by\_well.csv”
* Wllq\_by\_trt – determined by user entering information from lab notebook in the table “[project\_name]\_well\_quality\_table\_by\_treatment\_cndx\_culture\_date.csv”

The wllq\_by\_well and wllq\_by\_trt are defined essentially the same as for the NFA. Refer to *documentation/guide\_to\_prepare\_mea\_nfa\_level0\_for\_tcpl.html* for more information on how to fill out these tables.

## Determine if any samples need to be registered

# Run the pre-processing script (run\_me)

## Step 0 - Gather and check files

The project.input.dir is searched for the 3 types of input data files. Manually modify these steps to get the desired input data files.

Full file paths to the selected files are saved in a text file ( “[project\_name]/[project\_name]\_[file type]\_files\_log.txt”). These lists of files will be read in steps 1 and 2.

## Step 1 – Get Neural Statistics Compiler Data

### Determining the run\_type

Each Neural statistics compiler file must be classified as either a “baseline” recording (before chemical treatment or “treated” recording (after chemical treatment). This is done 3 ways. None of these methods have been determined to be 100% reliable, so I am using all 3 methods and comparing the results.

### Analysis duration/start time:

* Analysis duration
  + This should be 2400 seconds (or 40 minutes). If the recording duration is significantly different than 2400 seconds, then you may want to assess whether this recording is still usable.
    - The definition of “significantly different than 2400 seconds” is up for interpretation. Previously, I have used a coarse filter of 500 seconds above or below 2400 to flag recordings.
    - Note that some endpoints are summed across the recording (e.g. number of spikes, number of bursts), so these endpoints are definitely affected by the analysis\_duration.
* Analysis start
  + I think this corresponds to the start time of the data relative for the beginning of the .RAW recording that is used to calculate the endpoint values with the Neural Statistics Compiler.
  + This field is taken from the neural statistics compiler header. If the analysis start is significantly different than 0, check if there were any abnormalities in the amount of rest time before the recording started (there should be 20 minutes usually). If so, flag the recording for follow-up analysis to determine if the data is still usable.

## Step 2 – Get cytotoxicity data

In the past, the cytotoxicity data was taken from the Calculations files on the sheet “LDH” and “CTB”. However, it was discovered in the TSCA2019 data set that sometimes the raw data may not be copied to the Calculations files for some projects. Since templates for the Calculations files are usually based on a copy from a previous culture, there were blank-corrected values in the file, but for the wrong plates (so there was no error and it would have been easy to miss this).

In order to get around this issue, we can read in the data for the raw cytotoxicity data files. Then we can compare the values from the raw cytotoxicity files to the Calculations files to get any plate-specific modifications. See steps outlined in the run\_me.

## Step 3 – Compile all well-level data

* Load dat.neural.stats from Step 1 (this contains the data from both treated and baseline recordings in long format)
* Collapse the dat.neural.stats by well by calculating the percent change in activity from baseline to treated recordings in each well.
* Combine the neural statistics compiler data with the cytotoxicity data

## Step 4 – Data cleaning & checks

See notes in run\_me.

Things to check in data visualizations:

* MEA points should be more or less centered around 0
* there should be points filling up the whole range from -100 to 300
* Expect to see at least some points around -100 at the higher conc's
* outliers - do we see any/several extreme outliers for some of the more common endpoints? Does this pose a concern?

# Preparing to tcplWriteLvl0()

What I did in the past to prepare data for tcplWriteLvl0() (i.e., for the mc0 shared on 2020-07-29):

(see *mea-acute-neural-stats-to-mc0-scripts/deprecated\_create\_lvl0\_snapshots.R*)

* Got the most recent “dat4” from the “output” folder of all projects that were ready to be pipelined and combined into 1 table
* Set wllq to 0 for DMSO wells identified as outliers based on the entire data to pipeline
* Saved 2 tables under *lvl0\_snapshots*
  + Dat4 – the results above with all columns
  + Mea\_acute\_lvl0 – same as dat4 but restricted to columns needed for the mc0

A simpler approach we could use going forward:

* Within in each project folder, the final output of the *run\_me\_[project\_name].Rmd* should be ready to add to the existing mc0 in invitroDB (e.g., *TSCA2019/output/ TSCA2019\_MEA\_Acute\_for\_tcpl\_lvl0.RData*)
  + (If we want to remove DMSO outliers based on the entire MEA acute mc0, then an extra step may be needed here)
* When combining the new level 0 data with the existing mc0, we may want to check that none of the data in the new mc0 has already been pipelined (e.g., check for shared apid or srcf’s). (This could happen if chemicals from multiple projects were tested on the same plates, and so copies of the experimental data for those plates are present in multiple project folders on the L drive).
* The folder *lvl0\_snapshots* may not be needed anymore, other than optionally to house documentation of updates to the mc0/sc0 in invitroDB.

# Known issues/FYIs with MEA acute in general

## Definition of apid as the experiment.date (probably should be plate.id!!)

For all MEA acute mc data that has been pipelined to date, the “apid” is defined as the experiment date rather than the plate ID. Thus, all 3 plates tested on the same experiment date have the same apid.

I think I made this choice (back in ~2020) because there are only 3 DMSO control wells per plate. By pooling 3 plates together, we might get a more reasonable bval. HOWEVER, there might be important plate-to-plate variability that is missed with this definition of the apid. **\*\*This is especially an issue for the cytotoxicity assays, where the exact amount of time that a reagent is present may vary from plate to plate and thus may be important for normalization (per Kathleen).**

Furthermore, the bval is currently defined as the median of DMSO + cndx 1 and 2 on each apid (meaning that 9 DMSO wells + 12\*3 test wells = 45 wells are used to calculate the bval for each experiment date!). So even if the apid is set to the plate.id, there would still be 15 wells to used to calculate the bval on each plate (which seems like more than enough). Of course, there may be concerns about whether we might see effects at cndx 1 and 2… but what can you do.

So I would probably change the apid to experiment.date\_plate.id. BUT, to be consistent, we’d have to do that for all MEA acute data that has been processed to date, which may not be desirable.

Gah… I’m sorry!

The apid is currently defined in *fileToLongdat.R*

## Extrapolation of treatment & conc labels for MEA parameters from cytotoxicity data

Currently, the scripts are set to extract the meta data for the LDH and AB data only. Then, in level 4, the meta data for the LDH and AB assays is extrapolated to the data rows for the other parameters by well.

This should work UNLESS there is a discrepancy between the actual treatment in an MEA plate and an LDH or AB plate. (This could happen if there was a mix-up when transferring media from the MEA plate to the LDH or AB plate). Thus, it would be better to get the meta data for the MEA plates from a separate source from the meta data for the LDH and AB plates. For example, the meta data could be taken from the Calculations files, sheet “Plate X”, from the “Dosing Plate” sheet, or a sheet similar to the “Log” sheets in the NFA could be created. In any case, communicate with the lab technicians so that they know which sheets the scripts read from, and which sheets should be updated if they need to make an update to the treatments and/or concentrations.

## Selection of ‘p’ wells for LDH data – half vs full lysis wells

In TCPL, the LDH blank-corrected values are normalized to both the median of neutral controls and the positive controls with the resp.pc method. The “pval” is calculated as the median of the positive control wells on each apid. Positive control wells are meant to indicate the amount of LDH released from cells that are completely lysed.

Background on positive controls (to my understanding):

The amount of LDH released by the lysed cells in a Media-only well is far outside the optimal reading window for the equipment that reads the optical density. Thus, 2 times the value in "half lysis" wells are often preferred to be used as the positive control values rather than the readings from the “full lysis” wells. (I think the “half lysis” is a dilution of the contents of a “full lysis” well, but check with Kathleen).

What I've done for previous data sets:

* For plates that have at least one 1/2 Lysis well with wllq=1,
  + multiply the 1/2 Lysis rval's by 2 and set wllt = "p"
  + set wllt for Full Lysis wells to "x" (We don't want these to be part of the "p" wells on these plates)
* For plates that do not have at least one 1/2 Lysis well with wllq=1,
  + Label full Lysis wells as wllt = "p"

HOWEVER, based on more recent conversations with Kathleen, she actually prefers to consider each plate individually to determine whether the Full or Half lysis wells should be used. So, we may want to have some conversations about how to determine which wells should be used as the positive controls for the LDH assay, perhaps based on standardized cutoffs or by expert-judgment.

## No system for keeping track of dose units

Usually, the dose units are in uM, but sometimes they are in mg/mL. Currently, there is no standardized system to communicate the dose units for the MEA Acute. Be sure to note any of these occurrences and make sure the tested\_conc\_unit in ChemTrack for the given sample agrees with the units in the mc0 (I think that’s the solution?) Or try to standardize all to uM?

## Problematic dataset-wide control wells wllq thresholds

(copying notes from run\_me\_TSCA2019.Rmd)

In previous project, a final well quality check was performed on the DMSO control wells to remove outliers in the MFR % change values (as opposed to the automated wllq updates based on the activity level in the baseline recording only).

What has been done for the all projects that have been pipelined to-date ('APCRA2019','DNT2019','GF2019','ToxCast2016'):

* Calculate the mean and SD where wllt == 'n' for ALL MEA acute data pre-processed to date
* Set wllq to 0 for any wells where the CCTE\_Shafer\_MEA\_acute\_firing\_rate\_mean is less than 2 SDs below the mean or more then 2 SDs above the mean for all parameters

I think that this approach was developed mostly to justify removing several low-activity control wells from the APCRA2019/DNT2019 projects. The idea was that as new data is added, the mean and SD of wllt == 'n' wells would be recalculated and the well quality would be reset for all existing MEA Acute data.

HOWEVER, this approach may not be preferable going forward because:

* If we have to update the mean and SD based on all wllt == 'n' wells for ALL MEA acute data every time a new project is added, then the well quality for some wells that have already been pipelined might change.
* I'm not sure that it makes sense to set the wllq to 0 for the LDH and AB parameters based on the MFR, particularly where the MFR is 2SD above the mean.

Some alternative options going forward:

* We could use the data quality threshold from the NFA (exclude plates where plate-wise median of controls on DIV 12 is < 10 spikes per min or < 2 active electrodes).
* We could set the standardized cutoffs based on only the MEA Acute data that has already been pipelined, and not update the cutoffs as more data is added.

## Unreliable documentation of DMSO conc in mc0

I’m not confident that I entered the correct concentrations for DMSO in the data that has been pipelined to date. I have some old notes that indicate that the concentration of DMSO in the APCRA project alternated between experiments (even though in the mc0 I entered 0.015 for all). I’m not sure if the actual concentrations of DMSO used were recorded in the lab notebook or elsewhere. So overall, I wouldn’t rely on the DMSO concentrations in the mc0 without further verification.

## Source files

Similar situation as for the NFA:

* The input source files
  + Background: All experimental and meta data sources files used to pre-process data for the ToxCast Pipeline will ultimately be saved in Clowder (I think). With the current pre-processing approach, the file paths to the source files used are recorded in the files\_log.txt. However, these source files could be modified after the pre-processing is completed. Therefore, the source files should be copied to a stable location before running the pre-processing scripts. However, this seems like it could result in a lot of data copying, especially if the files will need to be copied again to the ToxCast\_data drive. Going forward, work with the ToxCast data team to determine the best path, and check for any modifications to the source files that occurred after pre-processing before sharing with the ToxCast team.
* The 'srcf' columns:
  + Multiple source files may be used to generate the numeric and meta data for a single data row (e.g., the raw Neural Statistics Compiler file and the meta data in the Calculations file)
  + Thus, a solution is to package the source files into .zip files (e.g., 1 for each plate for the MEA endpoints?). These .zip files would become the new 'srcf'. This solution was proposed by Jason Brown to address a similar issue for the HCI data.