Notes on pre-processing MEA Acute TSCA2019 data

August 4, 2023

# Status of the TSCA2019 data pre-processing August 4, 2023

(see run\_me\_TSCA2019.Rmd)

Blue = complete

Black = to do

## Multi-concentration screen

* Steps 0 - 3 are complete
* Step 4
  + Check treatment labels - done
  + Check appropriate treatment labels for positive controls
    - There are 3 plates on which all 3 DMSO wells have rval’s that appear much closer to the Lysis wells than to other DMSO wells. (See plots in run\_me\_TSCA2019.Rmd under “LDH plates”). I haven’t seen a note in the lab notebook to indicate that anything is off here… but this does seem quite abnormal. Maybe the LDH readings were just a lot higher overall for these plates? Not sure if it’s worth investigating.
  + Spids – done
  + Wllt – done
  + Concentration checks
    - Preliminary checks – done
    - Compare conc's with expected conc based on stock conc & conc index– done
    - “Determine concentration for non-test wells”
      * Not done (I’m not sure what the actual concentration for PICRO and TTX is, but not sure it’s worth investigating)
    - Concentration units
      * Again, not sure about PICRO and TTX wells
  + Final well quality updates - done
  + Prepare LDH p wells – not done
    - Code in run\_me shows the approach taken in the past
    - However, recent conversations with Kathleen have brought up a concern. See notes in run\_me and “Selection of ‘p’ wells for LDH data” in *MEA Acute Pre-processing Notes.docx*
  + Check all acnm are registered in invitrodb – done
    - **(FYI 1 assay component, CCTE\_Shafer\_MEA\_acute\_per\_network\_burst\_interspike\_interval\_cv, has not been registered, but I’m not sure we really care to include it)**
  + Note the vehicle controls per plate – done
  + Data usability checks – not done
    - See notes in run\_me. I’ve included code to replicate what has been done in the past, but I’m not sure that is the best idea.
  + For repeated treatments, determine which culture(s) to keep
    - There are quite a few repeated treatments. Likely will want to keep all of them, but check with lab technicians.
  + Confirm every treatment has at least 4 conc's with wllq=1
    - Every treatment has at least 4 usable conc’s, unless changes made to wllq above
  + Check for the expected number of technical replicates for every treatment-conc-culture
  + Data summaries
    - Counts – done, but check again once finalized
    - Visualizations – check once data is finalized
* “Step 5” – see Data concerns/issues below

## Single-concentration screen

I started to pre-process the single-concentration data back in November 2020 (see folder *deprecated/single\_point\_screen*), but the methods have been updated so much that I would recommend starting over. Tim has sad that we will want to pipeline this data eventually, but it is a lower priority than the multi-concentration data.

See *L:\Lab\NHEERL\_MEA\Project TSCA 2019\Acute TSCA SPS*

Several treatments were only screened in single-concentration, but not the multi-concentration screen. See *TSCA Acutehitcall.xlsx* and *TSCA Hits.xlsx* and the One Note Lab notebook ([TSCA Acute Single Point Screen Summary](onenote://L:/Lab/NHEERL_MEA/Project%20TSCA%202019/TSCA%20Acute%20Notebook/TSCA%20Acute%202019.one#TSCA%20Acute%20Single%20Point%20Screen%20Summary&section-id={D33B9F6C-86D3-4451-99FE-808015B27399}&page-id={EF2743DF-F6A7-4923-BAA9-98687ED98029}&end)) for how Seline, Kathleen, and Theresa identified positives in the single-concentration screen.

There are 2 options for how the single-concentration data could be processed:

* Use the same endpoints and methods as are used for the multi-concentration data (i.e., from the Neural Statistics Compiler)
* Attempt to replicate the endpoints and methods used to generate the MEA acute sc data that has already been pipelined (i.e., the assay components CCTE\_Shafer\_MEA\_MFR, CCTE\_Shafer\_MEA\_LDH, CCTE\_Shafer\_MEA\_AB).
  + The data processed under these endpoints are annotated to this srcf: NHEERL\_MEA\_SS\_SOURCE\_RAW\_DATA\_150422.csv. I have found *a* copy of this srcf here: *L:\Lab\Toxcast\_Data\toxcast\_data\files\ccte\_shafer\nheerl\_mea\NHEERL\_MEA*
  + I’m not entirely sure if the MFR in the existing sc0 is the same as the Neural Statistics Compiler “Mean Firing Rate (Hz)” endpoint (or if it was calculated with the Neural Metrics Tool, DARB, or some other tool). So I don’t know if CCTE\_Shafer\_MEA\_MFR is a weighted or unweighted MFR, or if there are other differences.
  + Note that the LDH and AB values in the current sc0 are pre-normalized

(I’d probably go with option #1, since I think it’d be easier than trying to reproduce old methods, but that does mean that we’d have single-conc MEA acute data housed in 2 separate sets of endpoints)

See *L:\Lab\NHEERL\_MEA\Carpenter\_Amy\pre-process\_mea\_acute\_for\_tcpl\lvl0\_snapshots\Documentation\MEA Acute Update August 2022.docx* for additional investigations on the current sc0 MEA Acute data in invitroDB.

# Data concerns/issues

## See Known issues/FYIs with MEA acute in general

See *MEA Acute Pre-Processing Notes.docx* for general concerns about the MEA Acute assay (but not specific to the TSCA2019 data)

## Well quality issues noted in lab notebook to follow up

Several well quality notes from reading the lab notebook were entered into this table:

*TSCA2019\tables\_manually\_reviewed\TSCA2019\_culture\_folders\_to\_review\_2023-03-02\_afc.xlsx*

See rows where ‘need lab followup’ == 1 or ‘need analysis followup´== 1. Examples of the kind of issues to investigate:

* Starting in Group 11 (Culture date 20210224, experiment date 20210311), they started performing a half media change on DIV 14 (presumably to feed the cells before the experiments on DIV 15). Is the data from previous experiments performed on DIV 15 without the half media change (within TSCA2019 and previous MEA Acute data) sufficiently comparable to data from experiments on and after Group 11?
* Usability of plate for which the recording after treatment was accidentally started 34 minutes after the treatment was added (instead of the usual 20 minutes)
* Differences between the LDH plates ran on the same experiment date (e.g. new vial of LDH control on each plate). Since the apid is currently defined as the experiment.date, are these differences okay or would they interfere with optimal normalization? (maybe this is a reason to change the apid to experiment.date\_plate.id?)
* ~3 compounds were noted to be blue in color. Tim noted that this might cause interference with the LDH/AB assays. Is this data still usable?

Feel free to address or ignore these notes as you see fit 😊

## Temperature Inconsistency

The temperature set point was changed for some experiments in TSCA2019:

* Cultures from 20201104 – 20210429 and 20210811: **37 C**
* Cultures from 20210505 – 20210714: **35 C**

I think that this change was related to concerns about the interior wells being too warm for some Maestros… talk to Kathleen/Theresa for more info and whether the data is usable.

## Variable of % DMSO in wllt n and t wells

Within the TSCA2019 project, different concentration of DMSO were used at the highest concentration versus other concentrations tested:

Top concentration tested:

Dosing plate: 4.5uL of chemical (in DMSO). Add 295.5 uL of Media -> 4.5/300 = 1.5% DMSO

MEA plate: 1:10 dilution -> 0.15% DMSO

Other conc’s + DMSO-only wells

Dosing plate: 5uL of chemical (in DMSO) from dilution plate. Add 495 uL of Media -> 5/500 = 1% DMSO

MEA plate: 1:10 dilution -> 0.1% DMSO

I think this was done due to practical issues of getting a top concentration of 30uM from a stock concentration of 20mM. However, given that addition of DMSO can affect the activity, I am concerned that the increased amount of DMSO in the top concentration may confound the results.

Additionally, for 1 plate from G16 (MW75-8105), a dosing error occurred for the top concentration tested such that the final concentration of DMSO in the MEA plate was 0.225%. The affected wells are noted in TSCA2019\_well\_quality\_table\_by\_well.csv, but currently I left wllq = 1.

The course of action should probably be considered within the context of the entire MEA acute data set. The concentration of DMSO has probably not been consistent throughout all MEA Acute data that has already been pipelined to date (both within and across assays). Note that the concentration of DMSO reported in the mc0 files to date is not 100% reliable (e.g., for APCRA2019, I labelled all DMSO wells as 0.15%, but I have other notes indicating that the concentration of DMSO may have varied throughout these experiments).

So the question is - Is the data from the top concentrations tested usable, or are any effects confounded by the higher concentration of DMSO? Or does it not matter given the variability of DMSO used throughout all MEA Acute data?

So I’m not sure what to do, just wanted to make you aware.

## Split recordings to patch

For 2 plates, the treated recordings were unintentionally interrupted:

L:\Lab\NHEERL\_MEA\Project TSCA 2019\Acute TSCA Conc Response\20201125 Culture G2\Neural Statistic Compiler

* AC\_20201125\_MW71-7113\_13\_00(001)(000).csv – records first ~900 seconds
* AC\_20201125\_MW71-7113\_13\_00(002)(001).csv – records ~ 1500 seconds
* AC\_20201125\_MW71-7113\_13\_treated.csv – attempted manual splicing of some parameter values from above 2 files

L:\Lab\NHEERL\_MEA\Project TSCA 2019\Acute TSCA Conc Response\20210512 Culture G24 (200H02 may need repeat)\Neural Statistic Compiler

* AC\_20210512\_MW75-8213\_13\_35(001)(000).csv – IDK what this file is. It’s only 3 KB.
* AC\_20210512\_MW75-8213\_13\_35(002)(000).csv – recording for 420 seconds

In the first case, Seline attempted to splice together some of the parameter values from the 2 recordings (e.g., theoretically the number of spikes could be simply added together). However, spot-checking these values doesn’t quite line up (e.g., the number of spikes from the first and second recording in well A1 doesn’t quite add up to the value in the file that Seline made).

In the second case, the lab notebook indicates that the recording was interrupted in the last 7 minutes of recording. So theoretically, there might be an additional file containing the parameter values from the first 33 minutes.

Options that I see:

* Just not include any data from these either of plates. (This is the current approach in the run\_me\_TSACA2019.Rmd)
* Attempt to get the parameter values
  + For MW75-8213, ask Theresa/Kathleen if the recording for the first 33 minutes is available on a DROBO/elsewhere
  + Create a new “spliced” files with R to calculate the parameters that are easy to combine

## Maestro, Firmware, and AxIS versions Inconsistencies

Different Maestro’s, firmware versions, and AxIS versions were used to do the recordings for the MEA Acute assay (even within the TSCA2019 project). Where different Maestros were used (e.g., Maestro vs Maestro Pro), other settings are also different (e.g., the Voltage scale, Digital Low/High Pass Filters, etc). How much does this matter? I am wondering whether the different machines/software/firmware produce sufficiently similar results to be considered the same assay.

The maestro type, firmware, axis version, and many other settings for the TSCA2019 data were pulled from the Neural Statistics Compiler file headers and saved in the table *TSCA2019/output/ neural\_stats\_settings\_table.RData*. A similar table could be created for the other MEA Acute data processed to date (see code in run\_me\_TSCA2019.Rmd)

Within TSCA2019:

* 6 neural statistics files came from Maestro, the other 212 came from Maestro Pro
* ~3/4 of the recordings used MaestroPro Firmware 2.0.4.21, the other 1/4 used 1.5.2.4
* 4 different AxIS versions were used

## “Baseline Decline”

Kathleen and Theresa have been troubleshooting a decline in baseline activity and/or DMSO-effect in the MEA Acute assay for a while. Notably, Kathleen and Theresa did an experiment to look at the change in activity over time even without the addition of DMSO and saw some decline in activity.

See here for some of my past investigations on this issue:

* Notes on pre-processing data from a study specifically conducted to investigate the decline in baseline: L:\Lab\NHEERL\_MEA\CCTE\_Shafer pre-process for TCPL\MEA\_acute\MEA Acute Pre-process for TCPL\Baseline Decline ([Baseline Decline](onenote://L:/Lab/NHEERL_MEA/CCTE_Shafer%20pre-process%20for%20TCPL/MEA_acute/MEA%20Acute%20Pre-process%20for%20TCPL/Baseline%20Decline.one#section-id={06904D31-379F-4184-82F9-717C9D6A0075}&end))
* Analyses with the “Baseline Decline” data: *L:\Lab\NHEERL\_MEA\Carpenter\_Amy\pre-process\_mea\_acute\_for\_tcpl\investigations\baseline\_decline*

I’m not sure if the TSCA2019 data poses any more of a concern on this issue than past projects. Below is a figure of the changes in mean firing rate in control wells (see TSCA2019\figs, created with the script investigations\_2023-07-31.Rmd). It looks like the percent change in mean firing rate is more variable with the TSCA2019 than past projects.

For the purposes of pre-processing, I think the question is – Is the data still worth releasing despite possible declines in baseline?

Chart, scatter chart

Description automatically generated

## Problem in right bottom wells

Note from OneNote lab notebook (see page “Quality Control Notes”, written by Seline)

“Right bottom wells started having low number of active electrode since the beginning of 2021

* Checked plate lot (used different lot of plates, no change)
* Tried reading the plate with old Maestro, plate activity looked fine
* Maybe some problem with POGO pins
* Sent back Maestro Pro to Axion in end of July to upgrade for Impedance reading and replace POGO pins, checked by Axion and sent back to EPA, same problem
* Purchased new Maestro Pro, plate activity look fine on new machine”

In Groups 32-35 (20210630 – 20210811), no chemicals were tested in the wells in the bottom right. However, that there are still 26 groups from earlier in 2021 in which the bottom right wells were used.

So, is the data from the affected plates / machine still usable? Can we just use the standard thresholds of activity for the baseline recording (i.e., >= 10 active electrodes and MFR >= 0.6377603 Hz)?

## Plates proposed by lab technicians to discard

In the final page of the TSCA 2019 lab notebook (see L:\Lab:\NHEERL\_MEA:\Project TSCA 2019), Seline listed 19 plates to exclude from the analysis due to low controls quality. These 19 plates are currently noted in *TSCA2019\_well\_quality\_table\_by\_well.csv* with a wllq\_note, but I left the wllq = 1.

I’m sure Seline really did observe low activity on these plates, but I have misgivings about throwing out these plates entirely because:

* I doubt that the thresholds of “low controls quality” were applied to all plates in the project consistently.
* 12 treatments will not have any data with wllq == 1 if these plates are excluded.

Some individual points from these 19 plates already have wllq == 0 based on standardized cutoffs for the MFR or AE in the baseline recording, but not all points on these plates.

So, some options might include:

* Just take Seline’s advice and set the wllq to 0 on the plates listed
* Make up standards thresholds for acceptability at the plate level (e.g., based on MFR or AE)
* Not worry about plate-wise thresholds of acceptability