Summary of MEA NFA data for Nov 12, 2020 Level 0 data Release

**Scripts**

See the scripts used in the folder “nfa-spike-list-to-mc0-scripts\_2020-11-12” or the commit to https://ncct-bitbucket.epa.gov/projects/MEA/repos/nfa-spike-list-to-mc0-r-scripts/browse on Nov 27, 2020 titled “nfa-spike-list-to-mc0-scripts\_2020-11-12.”

**The Data**

The level 0 data is saved in the folder “lvl0\_snapshots.” The level 0 data is the sum of the final output from each dataset. (Specifically, each *[dataset]\_longfile.RData* is taken from the *output* folder for each dataset, 'Brown2014', 'DNTGF2019', 'Frank2017', 'NTP91', 'OPP2015', 'PFAS2018', and 'ToxCast2016'. See the function “get\_latest\_dat” from the folder *supplemental\_scripts* for more details).

Before the data is saved, I check for:

* Flag any duplicated data
* Flag any NA concentrations
* Flag any NA SPIDs
* Set wllq to 0 for any apid where the median of controls on DIV 12 is < 10 spikes per min < 2 active electrodes
  + For this release, only 1 apid was removed: 20160921\_MW1160-23

For each snapshot, two RData files have been saved. The first file, “mea\_nfa\_lvl0\_[date]”, contains only the columns needed for the TCPL mc0. The second RData file, mea\_nfa\_lvl0\_extra\_cols\_[date]”, contains the following additional columns:

|  |  |
| --- | --- |
| **Column** | **Description** |
| wllq\_notes | Short explanation for any data point where wllq==0 |
| mea\_treatment\_name | The compound name as it was recorded in the source data |
| treatment | The updated mea\_treatment\_name. Should reflect the PREFERRED\_NAME listed in the spid map file used |
| dataset | Indicates the dataset from which the data was taken (e.g. ToxCast2016, NPT91, etc) |

Additionally, because the *rval* cannot be NA for TCPL, the wllq has been set to 0 wherever the rval is NA at DIV 12 only in the file *mea\_nfa\_lvl0\_2020-11-12.RData* (but not in *mea\_nfa\_lvl0\_extra\_cols\_2020-11-12.RData).* Here is a summary of the number of NA rval’s for each component:

alldat[grepl("DIV12",acsn) & wllq == 1, .(num\_pts\_not\_NA = sum(!is.na(rval)), num\_pts\_NA = sum(is.na(rval))), by = .(acsn)]

acsn num\_pts\_not\_NA num\_pts\_NA

1: CCTE\_Shafer\_MEA\_dev\_burst\_rate\_DIV12 11840 0

2: CCTE\_Shafer\_MEA\_dev\_interburst\_interval\_mean\_DIV12 10743 1097

3: CCTE\_Shafer\_MEA\_dev\_burst\_duration\_mean\_DIV12 10743 1097

4: CCTE\_Shafer\_MEA\_dev\_per\_burst\_interspike\_interval\_DIV12 10743 1097

5: CCTE\_Shafer\_MEA\_dev\_firing\_rate\_mean\_DIV12 11840 0

6: CCTE\_Shafer\_MEA\_dev\_mutual\_information\_norm\_DIV12 11840 0

7: CCTE\_Shafer\_MEA\_dev\_bursting\_electrodes\_number\_DIV12 11840 0

8: CCTE\_Shafer\_MEA\_dev\_active\_electrodes\_number\_DIV12 11840 0

9: CCTE\_Shafer\_MEA\_dev\_spike\_duration\_mean\_DIV12 10701 1139

10: CCTE\_Shafer\_MEA\_dev\_network\_spike\_duration\_std\_DIV12 10668 1172

11: CCTE\_Shafer\_MEA\_dev\_inter\_network\_spike\_interval\_mean\_DIV12 10668 1172

12: CCTE\_Shafer\_MEA\_dev\_per\_network\_spike\_spike\_number\_mean\_DIV12 10701 1139

13: CCTE\_Shafer\_MEA\_dev\_network\_spike\_number\_DIV12 11840 0

14: CCTE\_Shafer\_MEA\_dev\_network\_spike\_peak\_DIV12 10701 1139

15: CCTE\_Shafer\_MEA\_dev\_per\_network\_spike\_spike\_percent\_DIV12 11840 0

16: CCTE\_Shafer\_MEA\_dev\_per\_burst\_spike\_percent\_DIV12 11840 0

17: CCTE\_Shafer\_MEA\_dev\_correlation\_coefficient\_mean\_DIV12 11840 0

The new mc0 can replace the existing mc0. Here is a high-level summary of the datasets included in the latest mc0 release (the year corresponds to the approximate date tested):

       dataset number\_of\_spids

1:   Brown2014               8

2:   DNTGF2019              76

3:   Frank2017              86

4:       NTP91              50

5:     OPP2015              27

6:    PFAS2018              78

7: ToxCast2016             100

**New Assay Components and Registration**

The attached RData file contains 15 new assay components that will need to be registered. These 15 represent the component values at DIV 12, rather than the AUC collapsed over DIV. I set wllq to 0 wherever the DIV12 *rval* is NA. Could you register these new acnm’s and add the same methods as we are using for the rest of the CCTE\_Shafer\_MEA\_dev endpoints?

Also, can we change the aenm “CCTE\_Shafer\_MEA\_dev\_LDH\_up” to “CCTE\_Shafer\_MEA\_dev\_LDH\_dn”? This endpoint uses the same normalization methods as the rest of our “down” endpoints. It is a measurement of the total LDH from living cells, rather than extracellular LDH released from dying cells.

**SPID assignment**

The sample IDs were assigned as follows:

|  |  |  |
| --- | --- | --- |
| Dataset | Data Location | Spidmap files used |
| OPP2015 | L:\Lab\NHEERL\_MEA\PIP3 - Project\Data\Organophosphates | all spids taken from *EPA\_11118\_EPA-Mundy\_27FR\_100mM\_20150701\_cg.xlsx* |
| NTP91 | L:\Lab\NHEERL\_MEA\PIP3 - Project\Data\NTP Compound | all spids taken from *Copy of NTP91\_Compounds\_4NHEERL\_MEA\_dev\_cg.xlsx* |
| ToxCast2016 | L:\Lab\NHEERL\_MEA\PIP3 - Project\Data\ToxCast Compounds | all spids taken from '*EPA\_12088\_EPA-Shafer\_96misc\_75ul\_20160826\_key.xlsx*'  Exceptions: Valinomycin from NTP91, 3 compounds from *EPA\_11024\_TShafer\_384ph2\_75ul\_13May2015.xlsx* |
| DNTGF2019 | L:\Lab\NHEERL\_MEA\Project - DNT 2019 | *All Assays\_list\_toxcast\_OECD 20190524.xlsx*  Exceptions: Glufosinate Ammonium, L-Glufosinate Ammonium, 3 & 4 Glufosinate-P Technical: *glufosinates\_spidmap.xlsx* (these spids derived from *EPA\_ES202\_EPA-Shafer\_103\_20191218\_key.xlsx*)  Loperamide, Glyphosate: *Shafer\_sample\_info\_to\_register\_20201110\_afc.xlsx* |
| Brown2014 | L:\Lab\NHEERL\_MEA\PIP3 - Project\Data\Specific Aim 1 | Sodium Orthovanadate: *EPA\_ES203\_EPA-Shafer\_42\_20200110\_key.xlsx*  Loperamide (hydrochloride): *EPA\_ES202\_EPA-Shafer\_103\_20191218\_key.xlsx*  Acetaminophen, Bisindolylmaleimide 1, Domoic Acid, Mevastatin: *Shafer\_sample\_info\_to\_register\_20201110\_afc.xlsx* (note that spids for Mevastatin, Domoic Acid, and first dilution of Bis 1 were taken from Shafer\_42 list). |
| PFAS2018 | L:\Lab\NHEERL\_MEA\Project PFAS 2018 | *EPA\_9238\_EPA-Shafer\_75\_20180511.xlsx*  Exceptions: Loperamide (hydrochloride), Acetaminophen, Bisphenol A taken from *Shafer\_sample\_info\_to\_register\_20201110\_afc.xlsx* |
| Frank2017 | L:\Lab\NHEERL\_MEA\PIP3 - Project\Data\Specific Aim 2  L:\Lab\NHEERL\_MEA\PIP3 - Project\Data\Specific Aim 3  L:\Lab\NHEERL\_MEA\PIP3 - Project\Data\DNT Reference Compounds  L:\Lab\NHEERL\_MEA\PIP3 - Project\Data\DNT Reference Compounds\_2 | *EPA\_ES202\_EPA-Shafer\_103\_20191218\_key.xlsx* or *EPA\_ES202\_EPA-Shafer\_12*  Exceptions: TCEP, TPP *Copy of NTP91\_Compounds\_4NHEERL\_MEA\_dev\_cg.xlsx*  Acetaminophen, Glyphosate, Domoic Acid, Mevastatin, Bisindolylmaleimide 1: *Shafer\_sample\_info\_to\_register\_20201110\_afc.xlsx*  Sodium Orthovanadate: *EPA\_ES203\_EPA-Shafer\_42\_20200110\_key.xlsx* |

The origin of *Shafer\_sample\_info\_to\_register\_20201110\_afc.xlsx*:

(Most of the files referenced below are in the folder: *C:\Users\ACARPE01\OneDrive - Environmental Protection Agency (EPA)\Profile\Documents\MEA\_NFA Workflow Development*)

On Oct 19, 2020, I found 10 SPIDs that I was currently re-using in multiple datasets. I wanted to make sure that the same sample was indeed used in each culture (See *Notes\_Oct19\_2020.docx*). I matched by Lot/Supplier where possible, but I found that in several cases the Lot/Supplier information in the lab notebook did not match the Lot/Supplier in the spid map file (or was missing) (see *MEA NFA Spid Notes Oct 23 2020.docx*). We had to register these compounds that did not have an appropriate SPID. In the process, we learned from Chris Grulke that we can register separate samples for distinct dilutions of the same compound. For some compounds, we had to use our best judgement to guess the dilution dates. See *Sample IDs\spid\_registration\Verification of stock concentrations 2020-11-09.docx* for notes on determining the correct solvent and *Notes\_to\_Tim\_Oct21\_2020.docx* and *Notes\_Oct26\_2020.docx* for details on determining the correct dilution dates, solvent, and other investigations. Here are a few specific notes I am not sure I documented elsewhere:

* For sodium orthovanadate – we made a new sample of this every time we dosed (i.e., since we dosed 3 times throughout the experiment, we would technically have 3 separate SPIDs per plate). Since we can’t honestly separate by dilution for this compound, we decided to leave all samples under the same SPID.
* For Bis 1, I only found 1 dilution sheet. Kathleen found a record indicating that multiple dilutions had been made, but no dates were including. We (Tim, Kathleen, and I) used our best judgement to determine which samples should be distinct based on the 1 dilution sheet available, the rate at which a sample Bis 1 would have been used up in each experiment, and the fact that a sample of Bis 1 is only good for up to 6 months.

**Concentration-correction**

I have corrected the concentrations using the ‘stck’ in the "sample" table on invitrodb for each SPID. (I did this because some of the concentrations were already corrected in the source data, but some were not. I thought that would be easier to sort out on my end than on your end).

**Well Quality**

I have not yet developed standardized thresholds for acceptability (e.g. based on the mean firing rate or number of active electrodes in control wells). However, there is 1 plate that is a clear outlier (20160921\_MW1160-23, see *plateSN\_controls\_grouped\_by\_culture\_DIV12.pdf* in the folder *overall\_analysis*). The control wells did not develop – the median mean firing rate is 5.7 spikes per minute at DIV 12. For comparison, the next lowest plate median of controls is 35 spikes per minute. Usually we like to see controls above 50 spikes per minute at DIV 12. I have set wllq to 0 for this apid. All other well quality assignments are based on the discretion of the lab technicians.

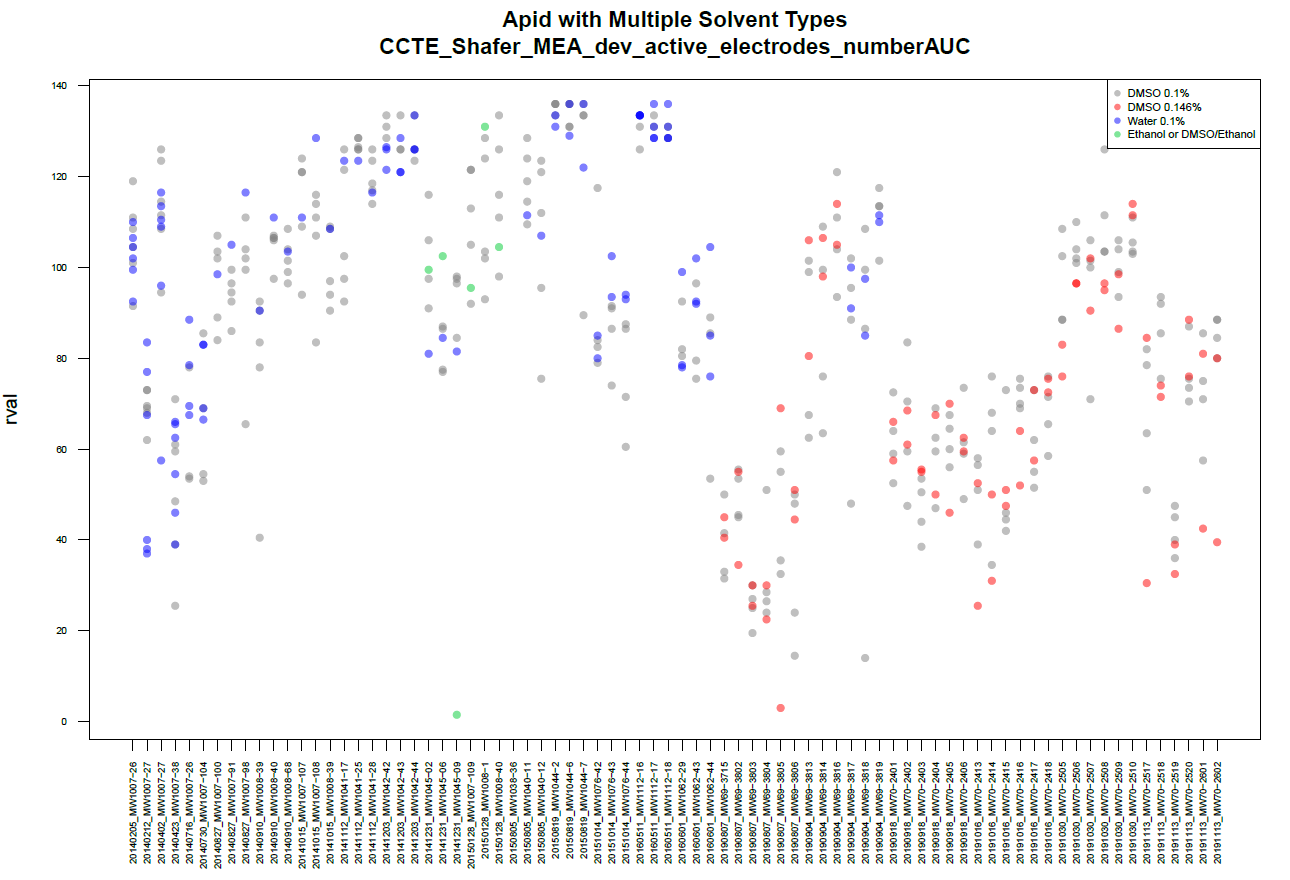
**Vehicle Solvents**

Here are the compounds and concentrations for all wells where wllt == “n”

|  |  |  |
| --- | --- | --- |
| Treatment | Conc (by volume) | Number of wells |
| DMSO | 0.1% | 1334 |
| DMSO | 0.146% | 134 |
| Water | 0.1% | 66 |
| Ethanol | 0.1% | 3 |
| 1:1 DMSO:Ethanol | 0.1% | 3 |

I entered the conc’s in the data as 0.001, 0.00146, etc.

Most of our control wells contain DMSO at 0.1% by volume, but some apid have a few solvent control wells with Water, Ethanol, or DMSO at 0.15% (78 out of the 252 apid have multiple solvent types). We have decided to pool all control wells together from each apid regardless of the vehicle solvent. It seems that the plate-to-plate variability is far greater than the variability between solvent type within a plate. See *mea\_nfa\_controls\_solvent\_comparison.pdf* and *apid\_with\_multiple\_solvent\_types\_AUC.pdf pdf* in the folder *overall\_analysis.* Below is a graph of the AUC number of active electrodes for the 78 plates with multiple solvent types.



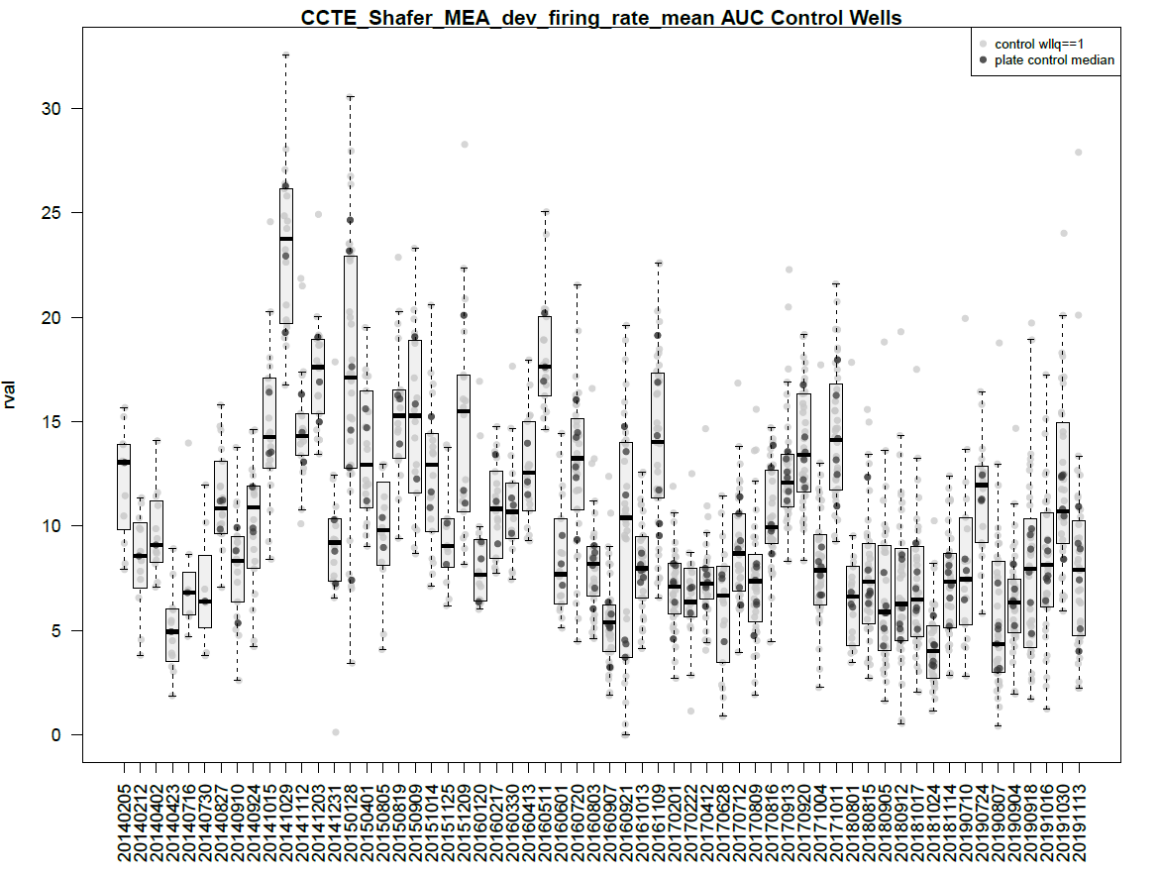
**Normalization methods**

Apid by plate or by culture

We have decided to let the ‘apid’ correspond to each 48-well plate (rather than the culture date, as we are doing for MEA\_acute). Even though the medians from each plate generally cluster by culture, there are a handful of plates that are noticeably different than the culture median. See *plateSN\_controls\_grouped\_by\_culture\_AUC/\_DIV12.pdf* and *plateSN\_controls\_boxplots\_by\_culture\_AUC/\_DIV12.pdf pdf* in the folder *overall\_analysis.* In the first set of files, the control wells from each culture date are shown in a column. The control values from each plate are plotted in 1 vertical line within each culture date group. The second set of files shows boxplots of the control wells from each culture date overlaid with the individual plate median of controls.

These plots show that there is generally good agreement among the control well values within each culture. We are leaning towards normalizing by the median on each individual plate because there are a few cultures where the individual plate medians vary significantly from the culture median. For example, looking at the AUC MFR:

* 20150128 – very big range of medians from all 6 plates.
* 20151209 (1 plate median is significantly above the IQR)
* 20160921 – just a really big range for all 6 plates
* 20161109 – another big range
* Plus a few other instances where the plate median falls outside of the overall culture IQR



I think that the only reason that we might want to normalize by culture instead of by plate would be to handle any “really bad” plates (e.g. from 20150128, 20160907, or 20160921). However, I think that if there really are any “really bad” plates, then perhaps the question at hand is not how to normalize them, but if we should include those plates at all. Since we have no real reason or need to normalize by culture, and there is some plate-to-plate variability, we are going to continue normalizing by plate.

Bval method

Currently, we normalize to the median of solvent control wells (bval.apid.nwlls.med). However, for the MEA\_acute assay, we normalize to the median of solvent controls and the two lowest concentrations tested (bval.apid.nwllslowconc.med). I have not yet investigated whether we might want to change the level 3 method for MEA\_dev. However, I think that it is reasonable that we stick with normalizing by solvent control wells only because:

* There is not a noticeable “DMSO effect” in the MEA\_dev assay as there is in the MEA\_acute assay
* I think it is very possible that we might see a treatment-related effect at the two-lowest concentrations tested for some MEA\_dev apid, since the treatment is on the plate for much longer than in the MEA\_acute assay

**Ideas for ongoing updates**

Overactive electrodes highlighting issues with the parameter definitions

Looking at the values in control wells over culture time, it is clear that several endpoints such as the burst rate, number of network spikes, and mutual information are consistently lower in more recent cultures than in earlier cultures. On the other hand, endpoints such as the mean burst duration and network spike duration and generally a bit higher for more recent cultures. I have a theory that this difference is due to the presence of “overactive electrodes” in the more recent cultures.

Rationale:

For network spikes, we have the requirement that the beginning and end of a network spike are sandwiched by a 0.05s time bin with 0 spikes from any electrodes. So, if just 1 electrode is firing constantly, there will be very few time bins with no spikes. Thus, what the human observer might call separate network spikes might get grouped into a single network spike. This would reduce the network spike number, but increase the network spike duration, as observed. Note that, while you would think that the mean number of spikes in network spikes would also increase, this endpoint actually only counts the number of spikes that occur at the 0.05s time bin at the “peak” of the network spike, not throughout the entire network spike.

Similarly, for bursts, if an electrode is just firing constantly, the algorithm will count really long but few bursts. In that way, there might be more activity, but fewer bursts counted.

Regardless, normalization will smooth out a lot of these differences.

If there are more “overactive electrodes” no than in the past… I wonder why that is?

Other ideas

* Change how the number of spikes in a network spike are counted
* Change how a network spike is calculated (do we have to require a time bin at 0?). Possibly make this more of a sliding-scale/adaptable to the conditions in the well, particularly to have a more robust measure that can adapt to the changes in activity patterns we are seeing over culture time (i.e. changes in network spike activity – with current parameters, looks like a lot fewer ns.n… but are the ns.n’s just different now?)
* Change bursting parameter so that we don’t have the inconsistency with min.IBI and max.isi
* Change the inter-network spike interval: right now, this is essentially just (time of last ns – time of first ns)/ns.N. Since the numerator will always be close to 900 seconds… this endpoint is currently not telling us much that the ns.N isn’t already communicating. Perhaps we could make this endpoint actually correspond to the mean time in between network spikes – so I would average the time from the end of one network spike to the beginning of the next (probably using the half-max metric to define the beginning and end).
* Reconsider endpoints that can be NA for the AUC calculation (does it make sense to set all NA’s to 0?)