**Experimental Design**

For each plate, separate out the wild type and mutant wells. Then, we will try and figure out whether or not certain sets of features, DIVs and electrode level data lead to clustering of certain wells in PC space.

For example, let there be four plates of 3 wells each. The orange colors are all mutants, with different shades for wells from different plates. The blue are all wild type, with different shades for different plates.

Then, each well will be represented by one data point in PC space. This data point will be determined by doing PCA on a matrix where each row corresponds to a well and each column corresponds to the variables being considered. The variables being considered are ‘features’ that are written out explicitly below. We will do different permutations of features used, in addition to electrode level data and multiple data points for DIVs to try and find a combination of factors where WT wells from different plates do not cluster separately, while mutant and wild type wells do cluster.

Example:

Plates

PC Space (Example on left, all sets cluster by plate and mutant v wild type. Example on right, controls do not cluster by plate. Looking to find sets of features for example on right, since this would suggest underlying difference in activity that is primarily due to mutation and not dominated by well to well and animal to animal variability)

***Pseudo Code***

Experimental Data Legend

Features – nAE (no active electrodes), mFR (mean firing rate), mISI (mean Inter-Spike interval), sdISI (st dev of Inter-Spike interval), mDB (mean duration of bursts), bpm (burst per minute), cvIBI (coefficient of variation for inter burst interval), mFB (mean frequency in burst), sdFB (st dev frequency in burst), mSPB (mean spikes per burst), mIBI (mean inter burst interval), pm (peak mean), mNBT (mean network burst time), SPB (spikes per burst), pSIB (% spikes in bursts), number of bursts, si (have written as si per active electrode, not sure what this feature is, can’t read my writing)

Plates – plates listed as Pi where a indicates which plate. E.g For 4 plates, Pa, Pb, Pc and Pd. If I write for all Pi, indicates I want to do a for loop that goes through all four plates.

Wells – Let’s say plate Pa has 20 wells, 10 wild type and 10 mutant. Matrix formulations I will use will have Paw be contained within Plate A and include the data points for the 10 wild type wells. Pam refers to the mutant wells in Plate A.

Electrodes – Number of Electrodes = 16

DIV – DIV of initial analysis will refer specifically to some combination of DIVs 25,27 and 29 for initial analysis.

***Code*** *-* ***Well level analysis with DIVs***

%% This symbol refers to comments within pseudo code

MI = Mutant index = 0

WTI = Wild Type index = 0

FM = Feature Matrix = 3 \* 15 where each row corresponds to DIVs (25,27,29) and each column represents one of the 15 features.

For i = 1:# of Plates

For j = 1:# of wells

If nAE passes filter continue, else break

If well is mutant

MI = MI + 1

Pim(MI) = FM of current well in loop

Else

WTI=WTI+1

Piw(WTI) = FM of current well in loop

%% Paw should be of size x\*y\*z where x = number of wild type wells in Plate A, y = 3 and z = 15, which is number of features excluding nAE, which is just being used for prefiltering.

Fts = Features for analysis. Using order above, setting Fts = [2,3,4] means you want to do PCA using mFR, mISI and sdISI.

PCAAni = Matrix that will be used for PCA, where i is just an index. In the below example, a separate PCAAn matrix is used for mutant and for wild type for each plate used in analysis. For 4 plates with both mutant and wild type, there will be 8 of them.

For m = all matrices

%% (if 4 wells with mutant and wild type, all matrices will be a set of 8 matrices, where each matrix is x\*y\*z as described above. E.g Paw,Pbw,Pcw,Pdw,Pam,Pbm,Pcm,Pdm)

Concatenated = reshape m to size x \* yz

%% E.g for descriptions above, y\*z is 3 DIVs \* 15 features = 45 columns. Now m is a matrix that corresponds to either the wild type or mutant wells for a given plate of size # wells by 45. The concatenation should be done so that first 15 columns are DIV25, second 15 DIV 27 and the last 15 columns DIV 29)

AM = analysis matrix = portion of concatenated matrix that only uses the features established by instantiating Fts = Concatenated[:,(Fts,Fts\*2,Fts\*3)

%% Using example above with three features, AM should be size x\*y where x = # mutant (or wild type) wells and y = 9, because three features for three days each.

PCAAni = AM

%% At this point we should have 8 PCAAni, each are of size # wells by three times number of futures considered.

DoPCA and Graph

%% don’t know R well enough to pseudo code graphing, but basically, do PCA of all combinations of the 8 PCAAni. So, take two of those matrices, stack them on top of each other, run PCA and graph. See whether, for this set of features, the graph of PC1 to PC2 appears to cluster based on mutant, wild type, plate or other. The functionality of this set of code should allow us to put in numbers for the indices of features at the top and then run code and we should get 28 graphs (8 choose 2, that’s number of combos of two that can be done from 8 things).

***Code*** *-* ***Well level analysis with Electrodes***

FM = Feature Matrix = For all features with electrode level analysis, extract that data per electrode.

Let’s say 8 features have electrode level data, FM should be 16 \* 8.

Continue with code from above, just concatenated matrix should be 16 times 8 columns per row, if all features are used, instead of 3\*15. If r can do PCA in 3 dimensions, not a problem, but otherwise, need to think about the smartest way to flatten data so each row really describes a well as best as possible.