MECP2 Analysis

Andrew Willems and Tian Hong

5/23/2022

Objective

We are doing analysis on the new set of cohorts from the Krishnan lab for MECP2. First, we use intraclass correlation coefficient (ICC) values of the Cohort, Cell type, Cell number, and Image variables to determine if we need to build linear mixed effects models. After investigating if we need LMEs we then create heat maps of the MECP2 data. We compare the differences in means between the various conditions.

Step One

Load needed packages. `effectsize` is used to calculate the effect sizes of the differences in our various conditions/treatments. `ggforce` is used to make the sina plots. `ggpubr` is for the grouped plot support. `ggsignif` is used to add statistical results to ggplot plots. `gt` is used for making the nice tables. `ICC` is used to calculate the intraclass correlation coefficient to tell us if we can treat our predictors (variables) as independent or not. `magrittr` is a package that allows us to use pipes (%>%) in our code. `mclust` allows us to build Gaussian mixture models (GMMs) and calculate the mean intensity of the neurons of interest. `modelsummary` allows us to make our linear mixed effects (lme) model output more professional looking. `nlme` is the package that performs the linear mixed effects (lme) model fits. `rstatix` is used to do the pairwise t-tests and p-value correction. `tidyverse` is used for data manipulation. `webshot` is used to save our gt tables as .png files.

Step Two

Load the data and make separate data frames that are comprised of only 6 or 12 week data. The warning here is okay. When I make all columns numeric it introduces some NAs because not all columns have the same number of rows (some just have no data in that row and therefore they get an NA). I fill those NAs in with a 0.

Step Three: Building Gaussian Mixed Model (GMMs) for all of our data

Six week old data. We have a single mean intensity calculated for each sample

Twelve week old data. We have a single mean intensity calculated for each sample

Step Four: Now plotting and then saving all of the density plots for 6 week year old data

Plotting and saving for 12 week old data

Step Five: Adding the means of the GMM to our overall data frames

Cell_type	Cohort	Condition	Hemisphere	Image	Cell_number	Mean	Time
PNN-neg	#102319	NW	LH	1	1	1165.3707	6 wk
PNN-neg	#102319	NW	LH	1	2	835.6673	6 wk

PNN-neg	#102319	NW	LH	1	3	1130.7044	6 wk
PNN-neg	#102319	NW	LH	1	4	944.4801	6 wk
PNN-neg	#102319	NW	LH	2	1	804.8837	6 wk
PNN-neg	#102319	NW	LH	2	2	536.6386	6 wk

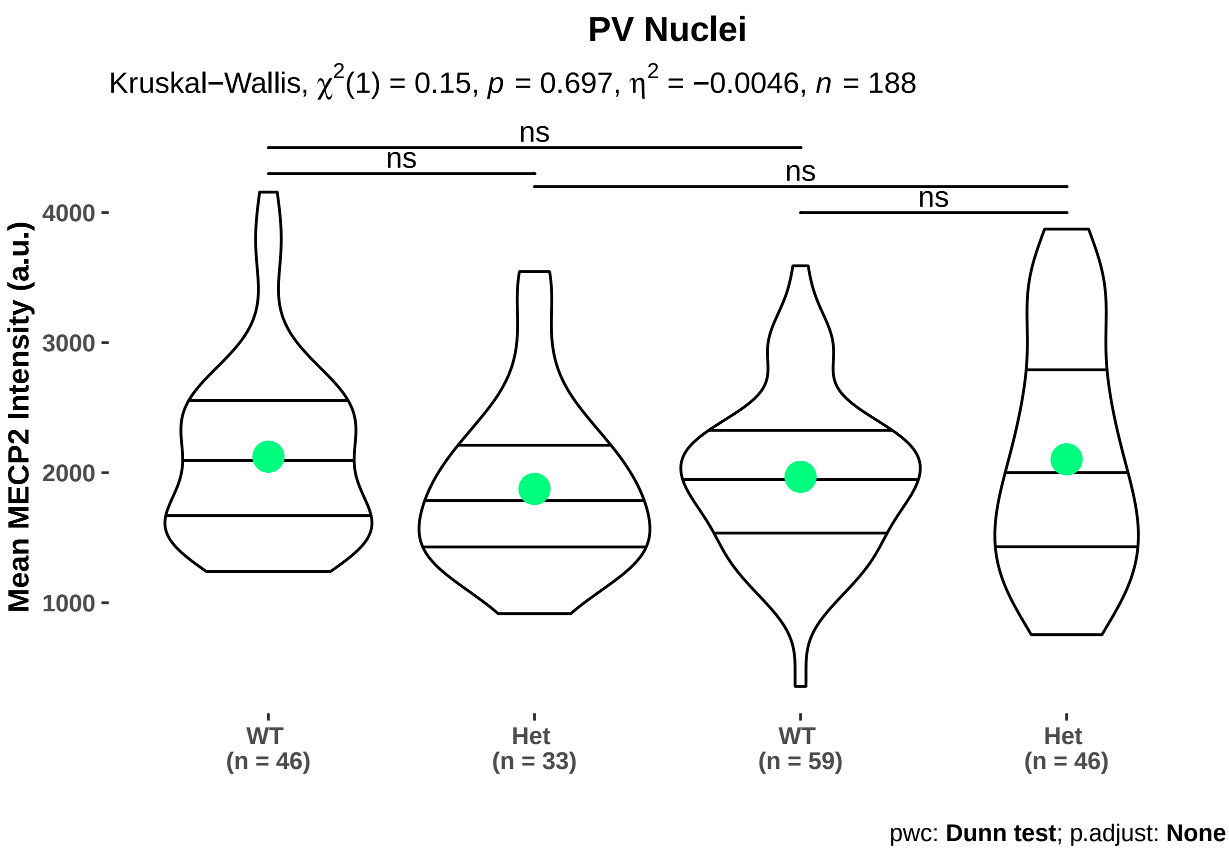
Cell_type	Cohort	Condition	Hemisphere	Image	Cell_number	Mean	Time
PNN-neg	#013118	NW	LH	1	2	816.7456	12 wk
PNN-neg	#013118	NW	LH	1	3	490.7493	12 wk
PNN-neg	#013118	NW	LH	1	4	601.9903	12 wk
PNN-neg	#013118	NW	LH	1	5	449.2910	12 wk
PNN-neg	#013118	NW	LH	2	1	479.0589	12 wk
PNN-neg	#013118	NW	LH	2	2	470.3309	12 wk

Now sub-setting our data frame to just NH and NW and then relabeling them as Het or WT. Finally, we factor them in the same order seen in the plot in the pre-print

Doing the same for 12 week old data

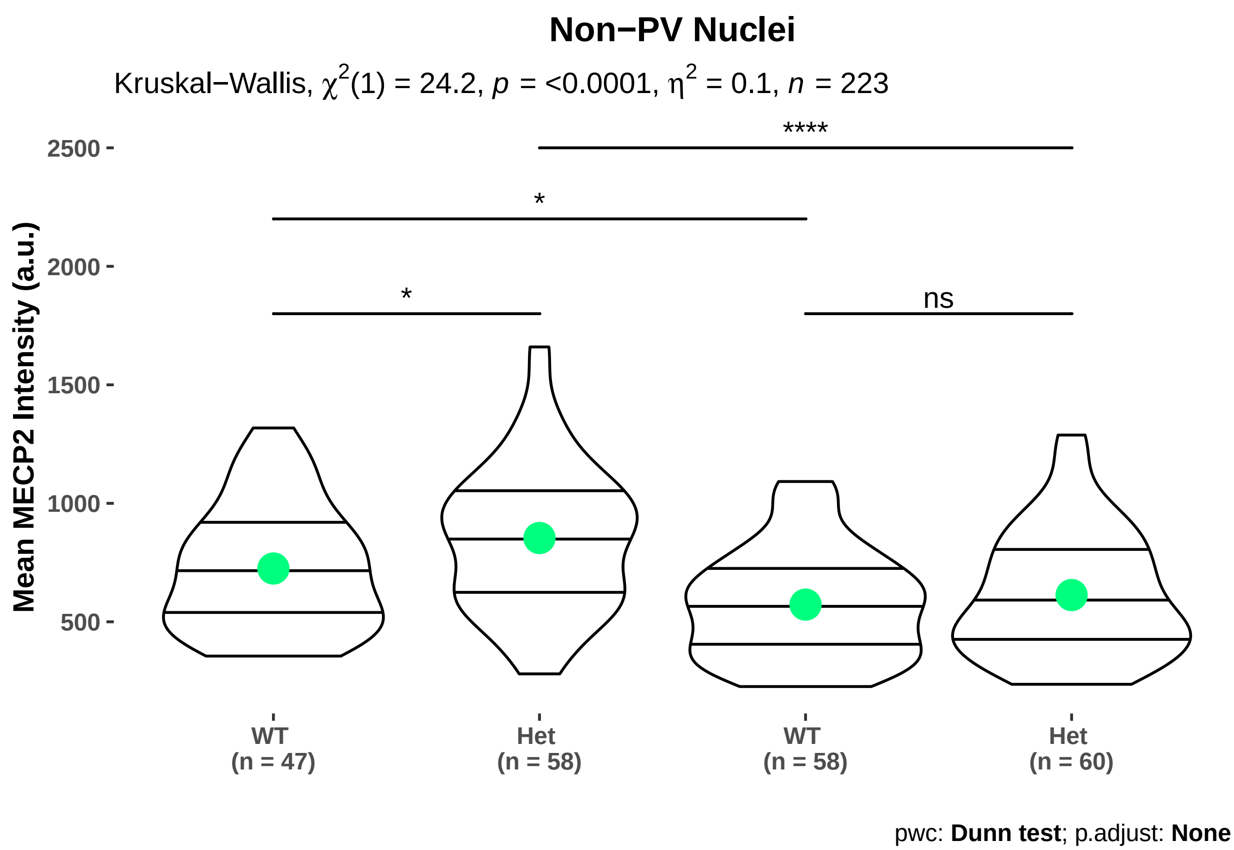
Now doing all the statistical analysis and plotting for the PV Nuclei (PNN-pos) containing samples

```
total_plot_pos
```

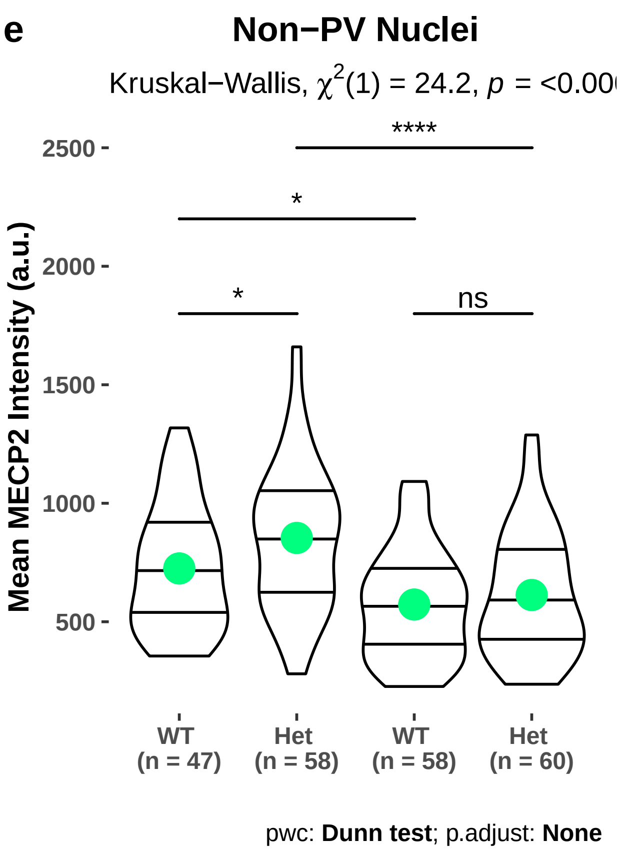


Now doing all the statistical analysis and plotting for the Non-PV Nuclei (PNN-neg) containing samples

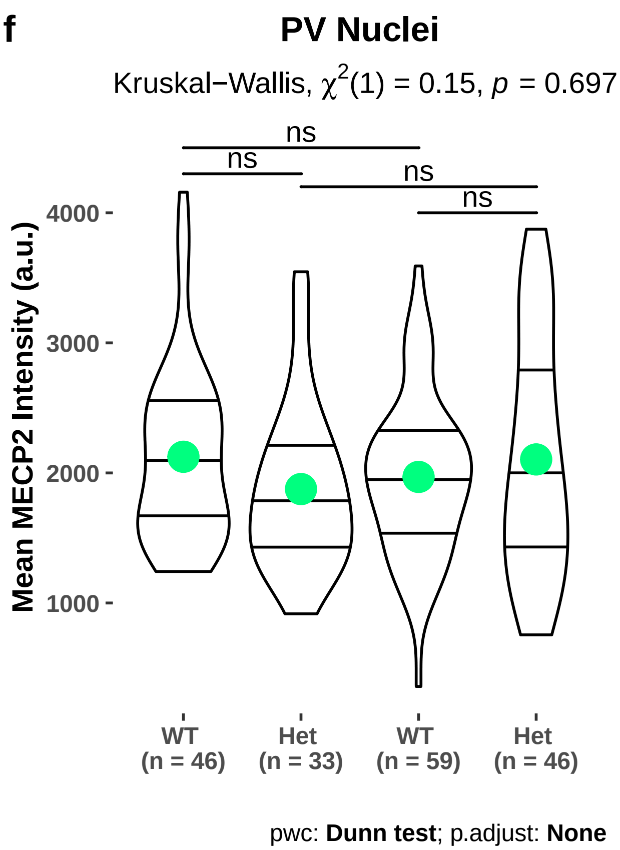
```
total_plot_neg
```


Non−PV Nuclei
Kruskal−Wallis, χ2(1) = 24.2, p = <0.0001, η2 = 0.1, n = 223
Mean MECP2 Intensity (a.u.)
2500
2000
1500
1000
500
****
*
*
ns
WT
(n = 47)
Het
(n = 58)
WT
(n = 58)
Het
(n = 60)
pwc: Dunn test; p.adjust: None

e


Non−PV Nuclei
Kruskal−Wallis, χ2(1) = 24.2, p = <0.00
Mean MECP2 Intensity (a.u.)
2500
2000
1500
1000
500
****
*
*
ns
WT
(n = 47)
Het
(n = 58)
WT
(n = 58)
Het
(n = 60)
pwc: Dunn test; p.adjust: None

f


PV Nuclei
Kruskal−Wallis, χ2(1) = 0.15, p = 0.697,
Mean MECP2 Intensity (a.u.)
4000
3000
2000
1000
ns
ns
ns
ns
WT
(n = 46)
Het
(n = 33)
WT
(n = 59)
Het
(n = 46)
pwc: Dunn test; p.adjust: None

Overall Non-PV Nuclei LME Model

	Model 1
Time12 wk	−200.977
	p = 0.217
SD (Observations)	212.118
Num.Obs.	222
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001	

Non-PV Nuclei 6 week WT v. 6 week Het LME Model

	Model 1
ConditionHet	118.020**
	p = 0.003
SD (Observations)	194.870
Num.Obs.	105
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001	

Now performing ICC analysis on the combinations we have previously tested to see if any of the variables have high levels of dependence

ICC for Non-PV MECP2 Data

Intraclass Correlation Coefficient (ICC) for Mean 6 and 12 week Non-PV MECP2 data.

Cohort	Cell number	Image
0.4977328	-0.004613448	-0.001720172

ICC for PV MECP2 Data

Intraclass Correlation Coefficient (ICC) for Mean 6 and 12 week PV MECP2 data.

Cohort	Cell number	Image
0.1235361	-0.01049272	-0.01039364

Building the lme for non-PV nuclei because of high ICC for `Cohort`

```
## Random effect variances not available. Returned R2 does not account for random effects.
```

Doing 6 week WT vs. Het lme model

```
## Random effect variances not available. Returned R2 does not account for random effects.
```

Doing 12 week WT vs. Het lme model

```
## Random effect variances not available. Returned R2 does not account for random effects.
```

Doing 6 week WT vs. 12 week WT lme model

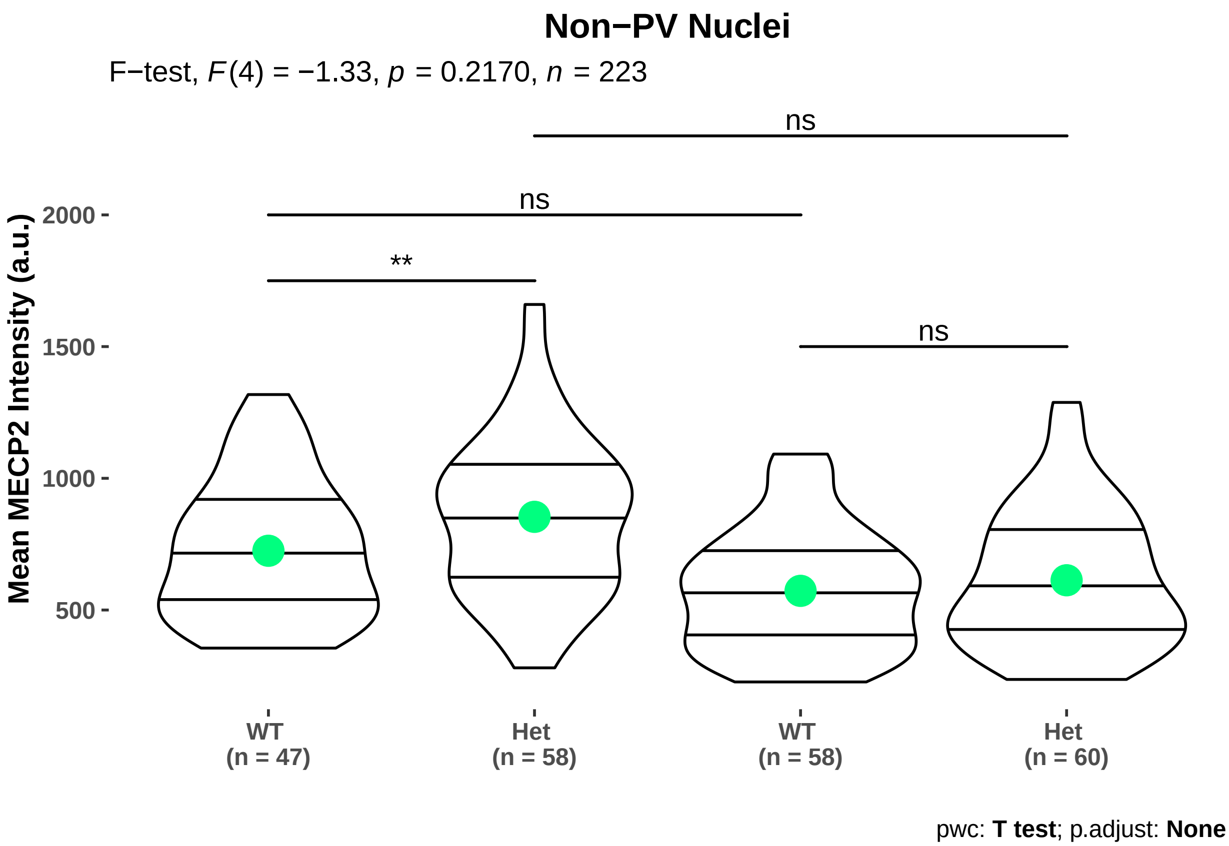
```
## Random effect variances not available. Returned R2 does not account for random effects.
```

Doing 6 week Het vs. 12 week Het lme model

```
## Random effect variances not available. Returned R2 does not account for random effects.
```

Non-PV Nuclei 12 week WT v. 12 week Het LME Model	
	Model 1
ConditionHet	35.918
	p = 0.381
SD (Observations)	218.650
Num.Obs.	117
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001	

Non-PV Nuclei 6 week v. 12 week WT LME Model	
	Model 1
Time12 wk	−161.190
	p = 0.296
SD (Observations)	185.782
Num.Obs.	104
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001	



Non-PV Nuclei 6 week v. 12 week Het LME Model	
	Model 1
Time12 wk	−235.897
	p = 0.182
SD (Observations)	223.115
Num.Obs.	118
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001	

Getting an lme of PV nuclei even though their ICC is low (0.12) and plotting to see if anything changes

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: mecp2_6_12_pos
##       AIC     BIC   logLik
##   2936.48 2949.34 -1464.24
##
## Random effects:
##  Formula: ~1 | Cohort
##         (Intercept) Residual
## StdDev:    226.9015 674.9409
##
## Fixed effects:  Mean ~ Time
##                  Value Std.Error  DF   t-value p-value
## (Intercept) 2035.6019  152.4186 178 13.355341  0.0000
## Time12 wk    -16.5627  212.0883   4 -0.078094  0.9415
##  Correlation:
##            (Intr)
## Time12 wk -0.719
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -2.5092504 -0.6378502 -0.1994135  0.5998646  2.7687115
##
## Number of Observations: 184
## Number of Groups: 6
```

Doing 6 week WT vs. Het lme model

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: six_wk_comp_df
##        AIC      BIC    logLik
##   1243.615 1253.093 -617.8076
##
## Random effects:
##  Formula: ~1 | Cohort
##         (Intercept) Residual
## StdDev:    270.1284 581.3881
##
## Fixed effects:  Mean ~ Condition
##                   Value Std.Error DF   t-value p-value
## (Intercept)  2146.3578  180.4454 75 11.894779   0.000
## ConditionHet -259.7963  134.3924 75 -1.933118   0.057
##  Correlation:
##              (Intr)
## ConditionHet -0.312
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -2.1642464 -0.6203736 -0.1458772  0.6283684  2.9647792
##
## Number of Observations: 79
## Number of Groups: 3
```

Doing 12 week WT vs. Het lme model

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: twelve_week_comp
##        AIC      BIC    logLik
##   1692.963 1703.579 -842.4815
##
## Random effects:
##  Formula: ~1 | Cohort
##         (Intercept) Residual
## StdDev:    180.0981 726.5549
##
## Fixed effects:  Mean ~ Condition
##                  Value Std.Error  DF  t-value p-value
## (Intercept)  1972.0631  141.9317 101 13.89445  0.0000
## ConditionHet  110.8158  144.6399 101  0.76615  0.4454
##  Correlation:
##              (Intr)
## ConditionHet -0.445
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -2.2562926 -0.6649898 -0.1281592  0.6001132  2.4305366
##
## Number of Observations: 105
## Number of Groups: 3
```

Doing 6 week WT vs. 12 week WT lme model

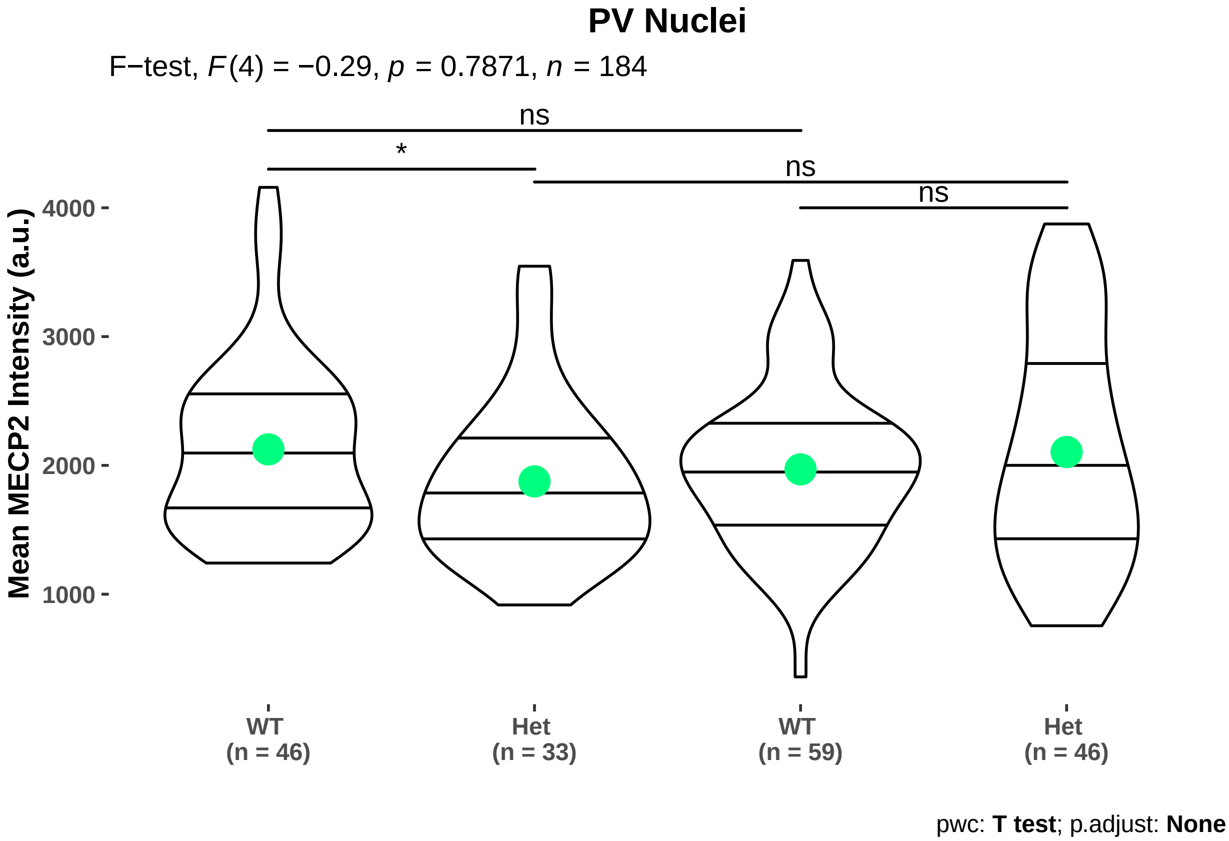
```
## Linear mixed-effects model fit by maximum likelihood
##   Data: wt_v_wt_comp
##        AIC    BIC    logLik
##   1651.185 1661.8 -821.5923
##
## Random effects:
##  Formula: ~1 | Cohort
##         (Intercept) Residual
## StdDev:    259.1322 579.9995
##
## Fixed effects:  Mean ~ Time
##                 Value Std.Error DF   t-value p-value
## (Intercept) 2151.4607  174.2789 99 12.344930  0.0000
## Time12 wk   -178.7261  242.9116  4 -0.735766  0.5027
##  Correlation:
##           (Intr)
## Time12 wk -0.717
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -2.45808538 -0.59875010 -0.08961367  0.45481706  2.94225852
##
## Number of Observations: 105
## Number of Groups: 6
```

Doing 6 week Het vs. 12 week Het lme model

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: het_v_het_comp
```

```
##        AIC      BIC    logLik
##   1280.189 1289.667 -636.0946
##
## Random effects:
##  Formula: ~1 | Cohort
##         (Intercept) Residual
## StdDev:    314.1724 724.8964
##
## Fixed effects:  Mean ~ Time
##                 Value Std.Error DF  t-value p-value
## (Intercept) 1879.5604  223.9510 73 8.392730  0.0000
## Time12 wk    171.0378  309.6096  4 0.552431  0.6101
##  Correlation:
##            (Intr)
## Time12 wk -0.723
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -1.9786892 -0.6411193 -0.1143075  0.5990460  2.1070045
##
## Number of Observations: 79
## Number of Groups: 6
```

PV Nuclei lme plot



Now doing ICC for just the non-pv het samples between 6 and 12 weeks to see if the ICC is large for this specific comparison or not

ICC for Non-PV Het Only MECP2 Data

Intraclass Correlation Coefficient (ICC) for Mean 6 and 12 week Non-PV Het Only MECP2 data.

Cohort	Cell number	Image
0.5143417	-0.03070989	0.003773949

Given the change from very statistically significant to non-significance between the 6 week and 12 week `Het` groups I wanted to see how many correlated neurons equaled one uncorrelated neuron and how many more neurons we would need to see a difference between these groups. I took the total number of neurons from the MECP2 negative group and divided it by the number of cohorts (because `cohort` is the variable with a high ICC). From this I got the average cluster size `M`. From there I calculated the Design Effect (`deff`). This tells us how many dependent neurons equal one uncorrelated neuron. From this we can get the effective sample size (`neff`) which tells us the equivalent number of cohorts if there was no correlation/clustering.

M	Design effect	Effective size
37.17	19.5	11.44

Our results show that about 11.44 cohorts is what we would need to get a sample size that would be equivalent to a sample size that had no correlation/clustering. This is about 1.91 times as many cohorts. (e.g. 12 needed instead of the 6 currently done)

Checking if WT only has high ICC for Keerthi

ICC for Non-PV WT Only MECP2 Data

Intraclass Correlation Coefficient (ICC) for Mean 6 and 12 week Non-PV WT Only MECP2 data.

Cohort	Cell number	Image
0.5157874	0.003545972	-0.01952475