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. `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. `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`). When calculating the mean later those rows with `NAs` are not included in the calculation.

Step Three: Getting straight mean for all of our data

Six week old data

Twelve week old data

Step Four: Adding the means to our overall data frames

Filtering to just naive `Condition`

Relabeling `NW` and `NH` as `WT` and `Het` respectively

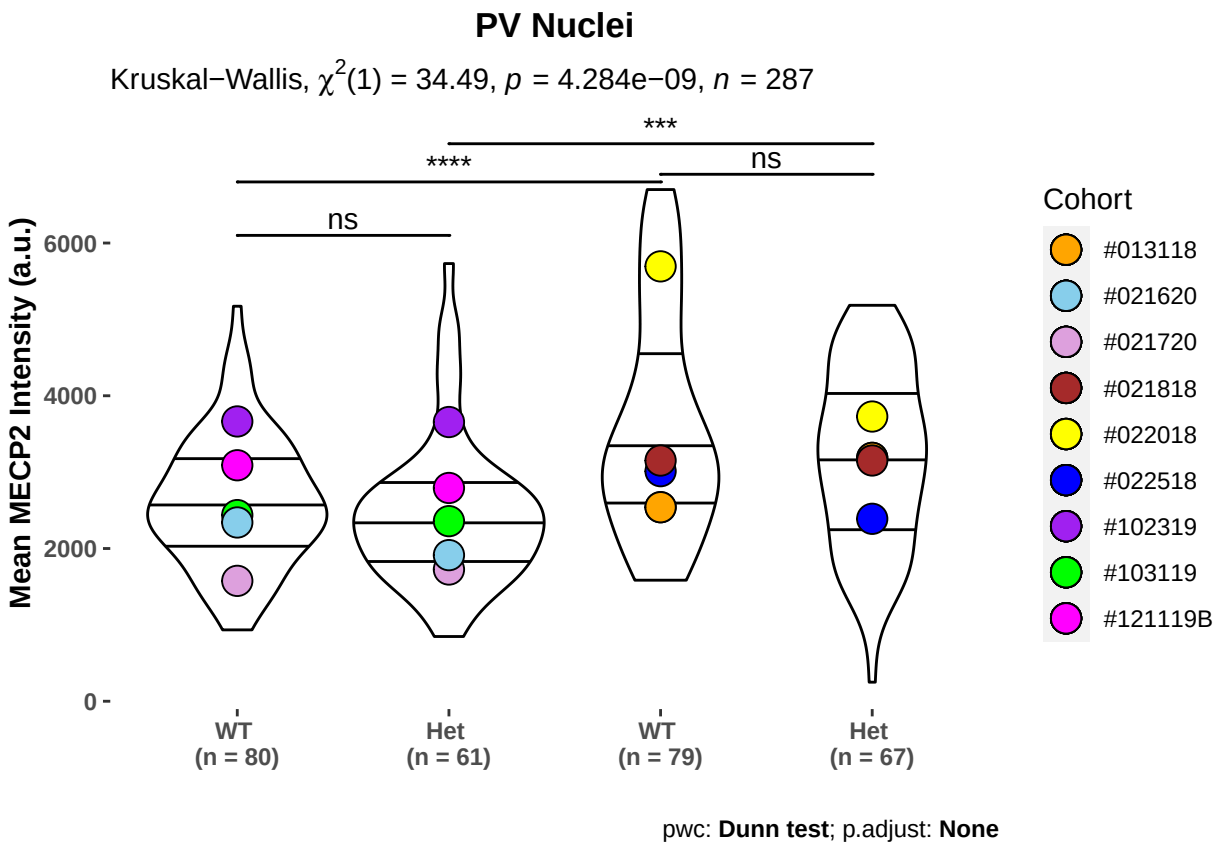
Now making the hemisphere all the same (`LH`) so that our analysis is correct for means

Checking to see if my means are the same as Logan's and Tian's

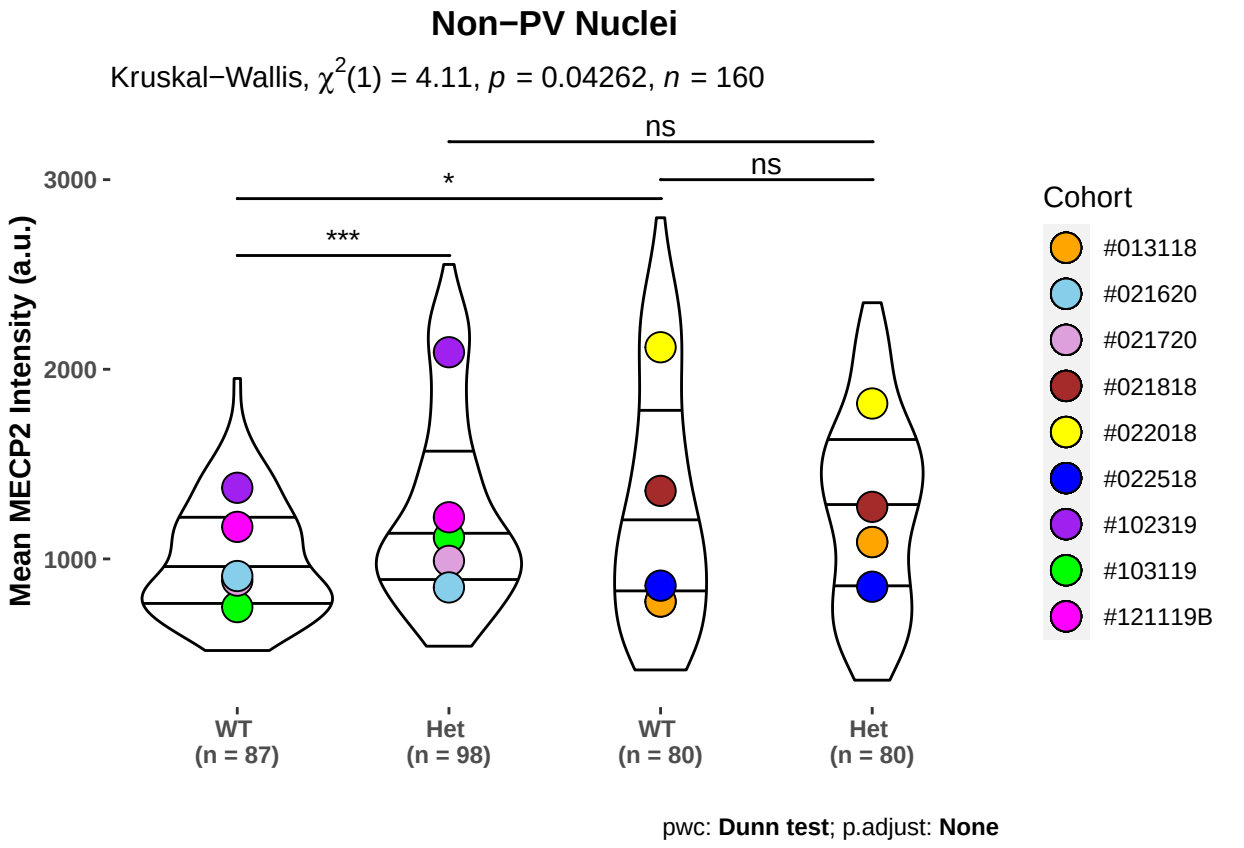
Filtering to just `PNN-pos` or `PNN-neg` cell types

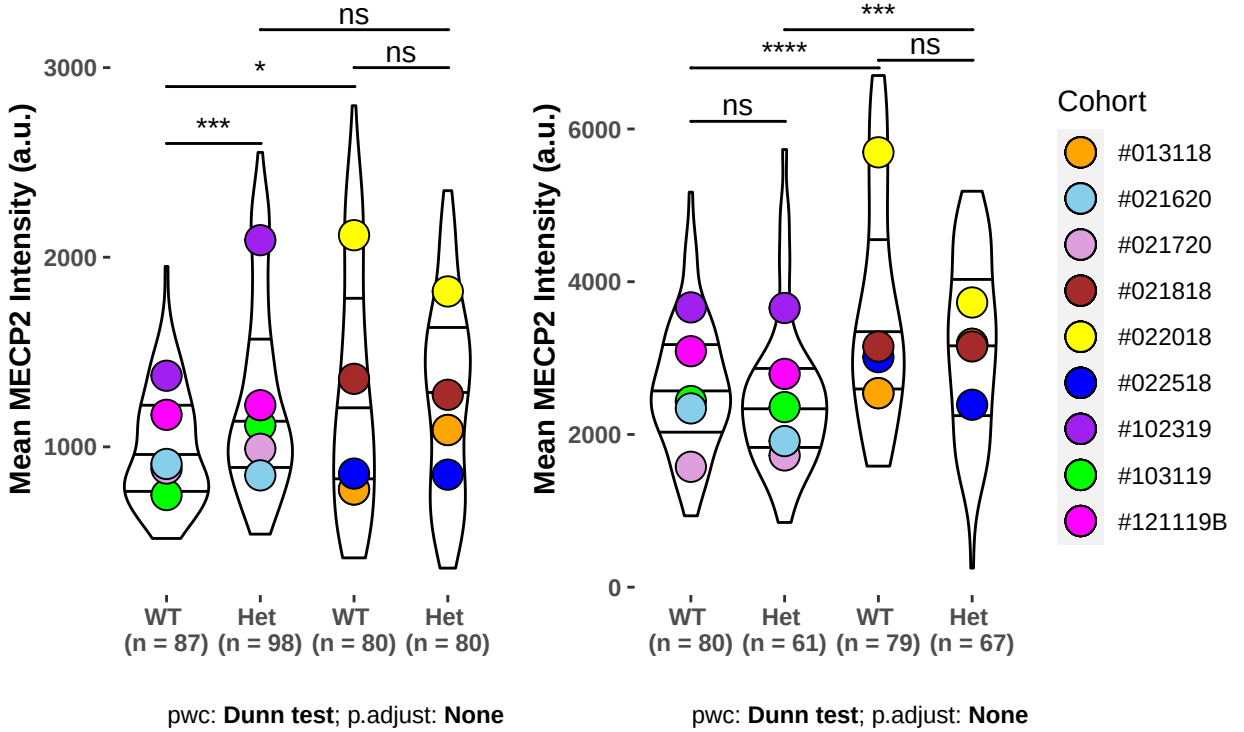
The table that contains the means that are equivalent to Logan's

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



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


Non−PV Nuclei
Kruskal−Wallis, $\chi^2(1) = 4.11$, $p = 0.04262$, $n = 160$
ns
ns
*
***
Mean MECP2 Intensity (a.u.)
3000
2000
1000
WT
(n = 87)
Het
(n = 98)
WT
(n = 80)
Het
(n = 80)
Cohort
#013118
#021620
#021720
#021818
#022018
#022518
#102319
#103119
#121119B
pwc: Dunn test; p.adjust: None


e
Non−PV Nuclei
Kruskal−Wallis, $\chi^2(1) = 4.11$, p
ns
ns
*
***
Mean MECP2 Intensity (a.u.)
3000
2000
1000
WT
(n = 87)
Het
(n = 98)
WT
(n = 80)
Het
(n = 80)
pwc: Dunn test; p.adjust: None
f
PV Nuclei
Kruskal−Wallis, $\chi^2(1) = 34.49$, $p = 4.284e{-}09$, n
***
****
ns
ns
Mean MECP2 Intensity (a.u.)
6000
4000
2000
0
WT
(n = 80)
Het
(n = 61)
WT
(n = 79)
Het
(n = 67)
Cohort
#013118
#021620
#021720
#021818
#022018
#022518
#102319
#103119
#121119B
pwc: Dunn test; p.adjust: None

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
0.6177129	-0.0113295	-0.005600858

ICC for PV MECP2 Data

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

Cohort	Cell number	Image
0.5192296	-0.01426585	-0.005552031

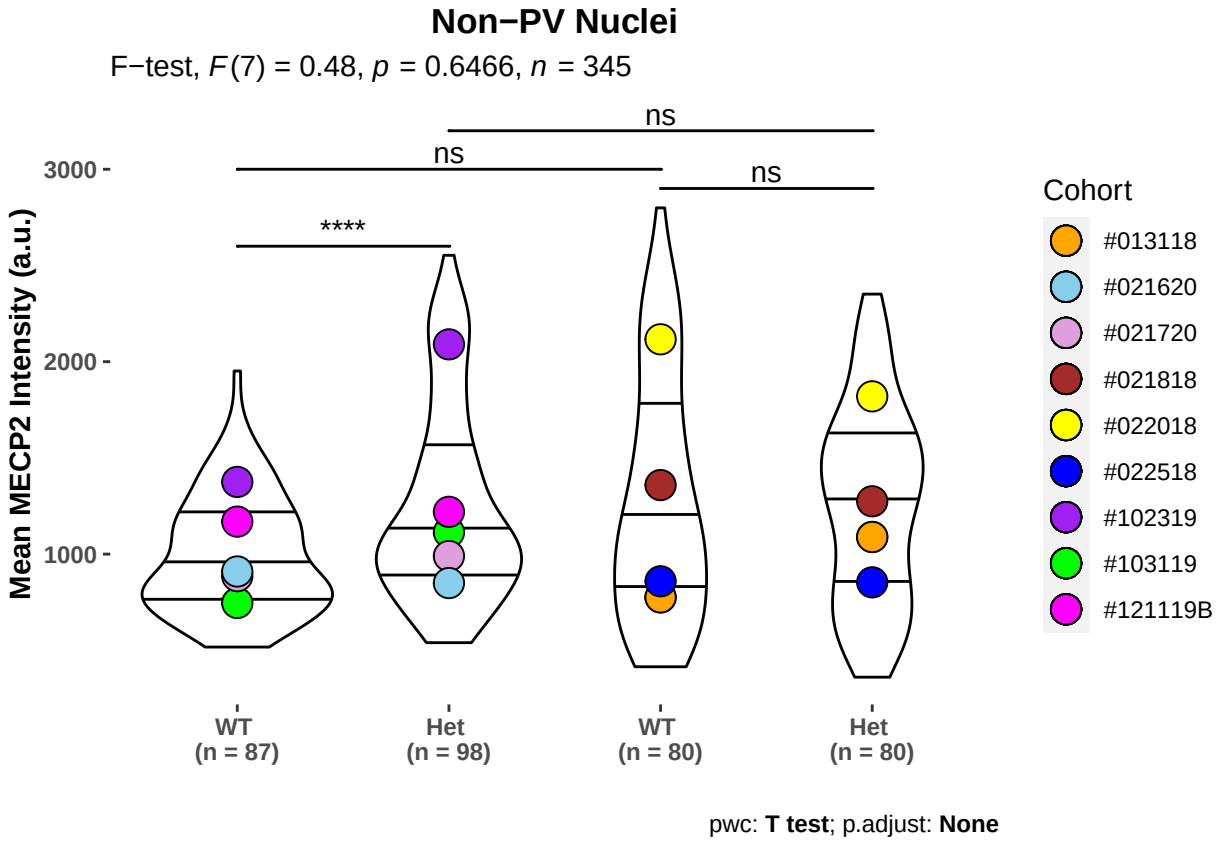
Building the lme for non-PV nuclei because of high ICC for `Cohort`

Doing 6 week WT vs. Het lme model

Doing 12 week WT vs. Het lme model

Doing 6 week WT vs. 12 week WT lme model

Doing 6 week Het vs. 12 week Het lme model



Getting an lme of PV nuclei because of high ICC (~0.51) compared to before the ICC was only 0.14

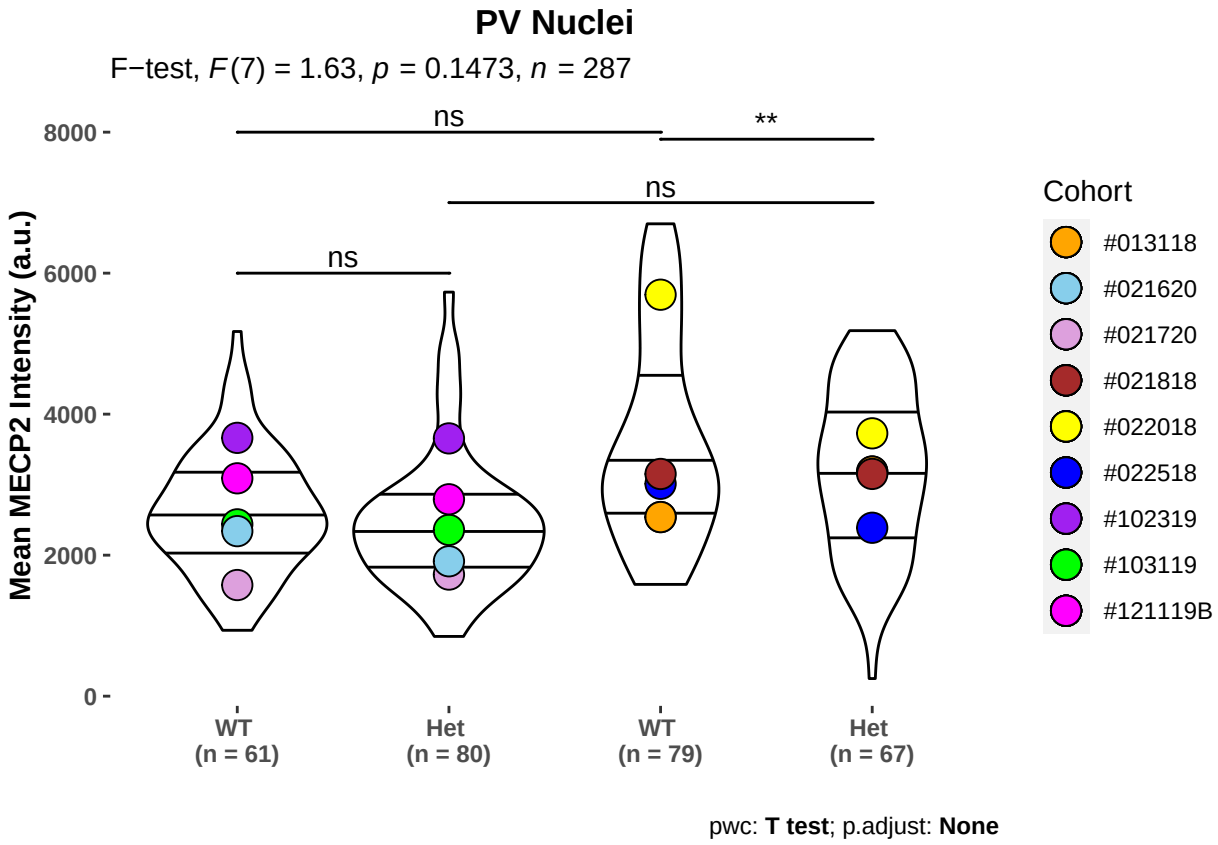
Doing 6 week WT vs. Het lme model

Doing 12 week WT vs. Het lme model

Doing 6 week WT vs. 12 week WT lme model

Doing 6 week Het vs. 12 week Het lme model

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
0.6573317	-0.02370183	-0.009198375

Our results show that about 11.58 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 1.93 times as many cohorts. (e.g. 12 needed instead of the 6 currently done). Given this we recommend an additional `n` of 6 mouse cohorts for analysis.

Non-PV Het only recommendation

Checking if WT only non-PV 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
0.7312525	-0.02050287	-0.01184338

Non-PV WT only recommendation

Figure 9c comparison. We are comparing the data we have to that in the pre-print to ensure consistency

PNN Figure 9 Plot

Overall Conclusion

We see a definite change in the results with these additional cohorts. We see a large rise in the ICC of the PV data. When using an LME for PV data we get results that are much closer to previous results. For Non-PV data we do not see previous results whether we use a vanilla model or an LME. These results could be problematic as a key result hinges on the Non-PV het v. het comparison. When looking at the number of WT vs. Het samples for Non-PV we see that they are roughly symmetrical (178 Het and 167 WT). Examining the different cohorts for Non-PV shows that they are evenly distributed (range of 32-40 with the majority being 40). In addition, we have 9 total cohorts now and they are unevenly distributed with 5 associated with 6 week data and only 4 associated with 12 week data.