THE UNIVERSITY OF ALABAMA AT BIRMINGHAM

Presenting Work in Progress Data

Chad Murchison, PhD

Department of Neurology, CNET

Department of Biostatistics

WIP Statistical Rigor

1. What - Hypothesis generation and experimental design
   Alignment of question and experiment
2. How - Selection and application of methods
   Are you using the right tool for the job
3. Why – Presentation, inference, and interpretation
   Is the result meaningfully conveyed

Key Points for Presenting Data

- Is there a research question and does it lead to a testable hypothesis
- Is the experimental and analytical design appropriate for addressing the research question
- Do the reported methods test the hypothesis while being sound
- Are the reported results (figures, tables) sufficiently and accurately presented while being understandable to the audience
- Is the information meaningfully conveyed or is it superfluous
- **Is the inference and interpretation in alignment with the design and the analysis**

Do I Understand the Research Question

- Should be clearly defined and understandable in moderately lay terms as an *experimental hypothesis* remember A -> B -> C
- Statistically, this leads to a testable *null hypothesis* and corresponding *alternative hypotheses* which should also be described in fairly lay terms

- Exploratory research questions exist even if *a priori* hypotheses do not; hypothesis testing (*how it happens*) versus generation (*if it happens*)
- Imagine you're writing an abstract or the last paragraph of the intro

Experimental Design

- What are the endpoints / outcomes / dependent variables?
- Primary independent variable
  - If comparing groups, what is the contrast e.g. control vs comparison
  - For relationships, what are the associate variables e.g. age
- Accounting for experimental bias
  - Definition of groups and inclusion/exclusion as needed
  - Biological vs technical replicates and power
  - Randomization, allocation methods, assessment control
- **A well-designed experiment should completely dictate the analytical plan and drive any statistical analysis**

Presenting Your Experimental Design

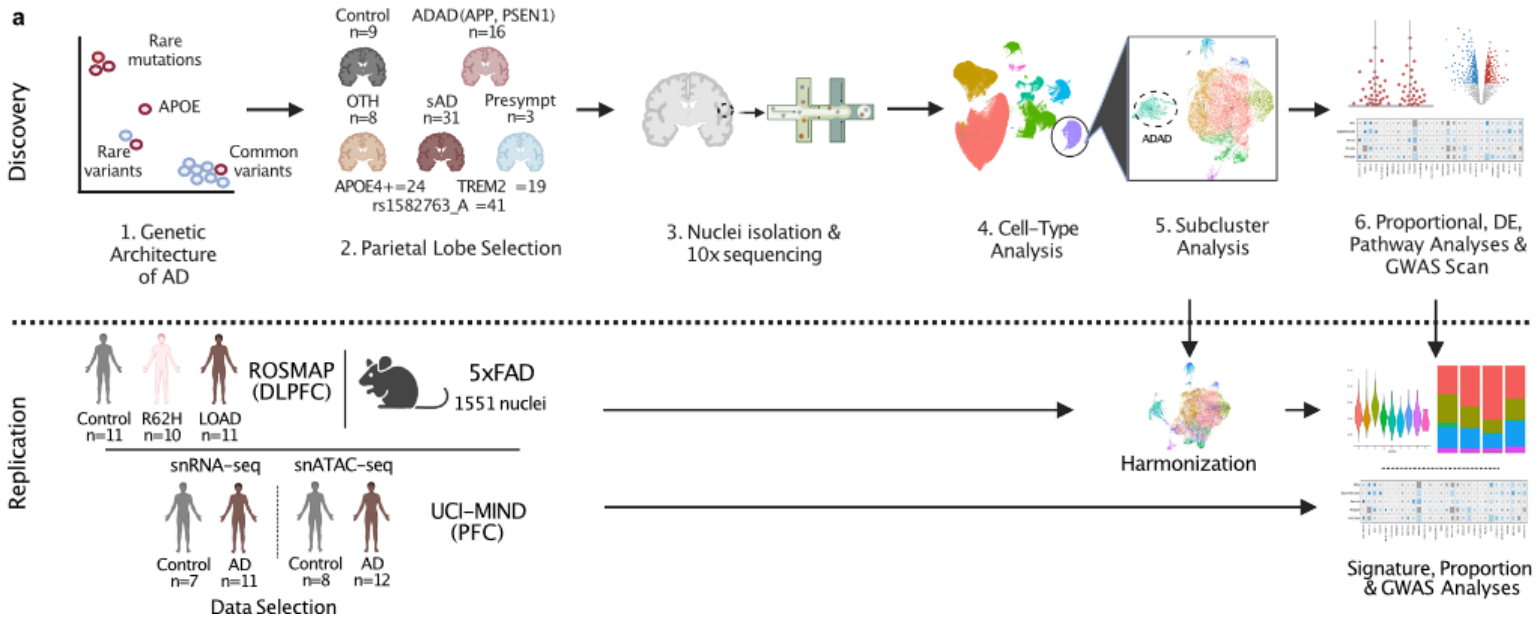
- Complex methods benefit from digestible design presentations

Table 1 | Demographic characteristics of samples

Samples	Control	ADAD	sAD	Presym	Other
Total	9	16	31	3	8
MS4A (AG %)*	55.6	46.7	45.2	33.3	12.5
TREM2@	-	-	15	-	4
PSEN1	-	13	-	-	-
APP	-	3	-	-	-
Braak Aβ (0/A,B/C)	2/7/0	0/0/16	0/0/31	0/0/3	2/5/1
Braak Tau (NA/I-III/IV-VI)	0/9/0	3/0/13	4/2/25	0/0/3	0/6/2
Sex (XY)%	33.3	56.3	45.2	33.3	50.0
AOD (mean, sd)y	90.1(9.6)	51.0(6.9)	81.5(6.4)	77.3(15.3)	88.8(6.1)
APOEε4+%$	11.1	25.0	54.8	33.3	12.5
PMI (mean, sd)h	10.9(5.5)	14.2(7.7)	11.9(6.3)	12.4(1.9)	11.3(9.1)

Other: (1:Dementia with Lewy bodies, 4:Argyrophilic grain disease, 1:Tramatic encephalopathy, 1:Neurofibrillary tangle-predominant AD, 1:Cerebrovascular disease).
ADAD autosomal dominant Alzheimer's disease, *sAD* sporadic Alzheimer's disease, *Presym* presymptomatic, *PMI* postmortem interval.
*MS4A is referring to SNP rs1582763 (**GG:25, AG:28, and AA:13**).
@Two African descent and one Asian descent (the p.H157Y is European descent).
$The total number of *APOE ε4+* were 24 (*APOE* genotypes: **23**:4, **24**:2, **33**:39, **34**:19, **44**:3).

Using demographics among disease types for clarity



Workflow indicating pipeline for both discovery of DEGs and transcript accessibility and how it relates to validation in additional datasets

- Help your very confused statistician out

Brase L et. al. *Nature Communications* 14, 2314 doi: 10.1038/s41467-023-37437-5.

Deciding on Your Analytical Methods

- Minimal assumptions
  - Normality – skewness, kurtosis, unimodal, ceiling / floor effects
  - Independence of units
  - Homogeneity of variance
- How were assumptions checked and what was done in response
  - Evaluation of residuals, ideally not just regurgitating Prism output
  - Any transformations on the outcome e.g. logarithmic transforms
  - Repeated measure and serial correlation adjustments
  - Non-Gaussian methods and models e.g. rank-sum instead of t-test
- Overly influential data and addressing outliers (especially for t-tests)

The Toolbox Problem

=

Statistical training for research frequently focuses on the individual tools and not on their contextual relation to the overall task

Easy Tasks Can Use Easy Tools

I want to compare multiple groups…

+

…that appear normally distributed…

->

…so, I'll use a one-way ANOVA

Statistics!

But When Things Get Complicated…

+

->

I have a mixture of cell cultures and animal models evaluating a collection of biomarkers…

…with some zero-inflated outcomes which sample within replicates over several time points…

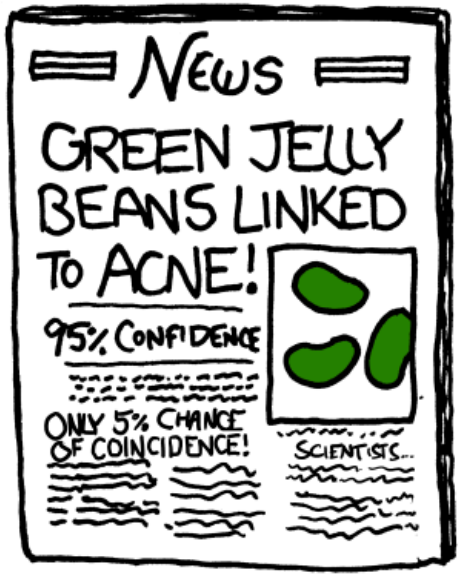
…but they're all just means so I'll use my one-way ANOVA a bunch of times for each comparison

...They May Not Go As Planned

Sta...statistics??

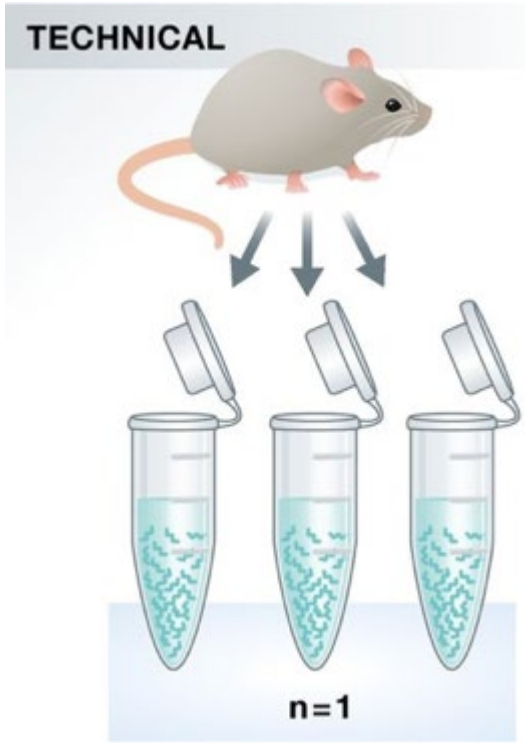
Some Statistical Red Flags

- P-hacking – either testing until something sticks OR restricting analysis set to force significance
- Not accounting for multiplicity
  1. Evaluation of multiple outcomes (Bonferroni, FDR)
  2. Multiple contrasts within a category (Tukey, Dunn)
- Post-hoc hypotheses – presenting generated hypotheses as *a priori* i.e. circular analysis
- Power and false discovery
  1. Conflating biological and technical replicates
  2. Taking absence of evidence as evidence of absence

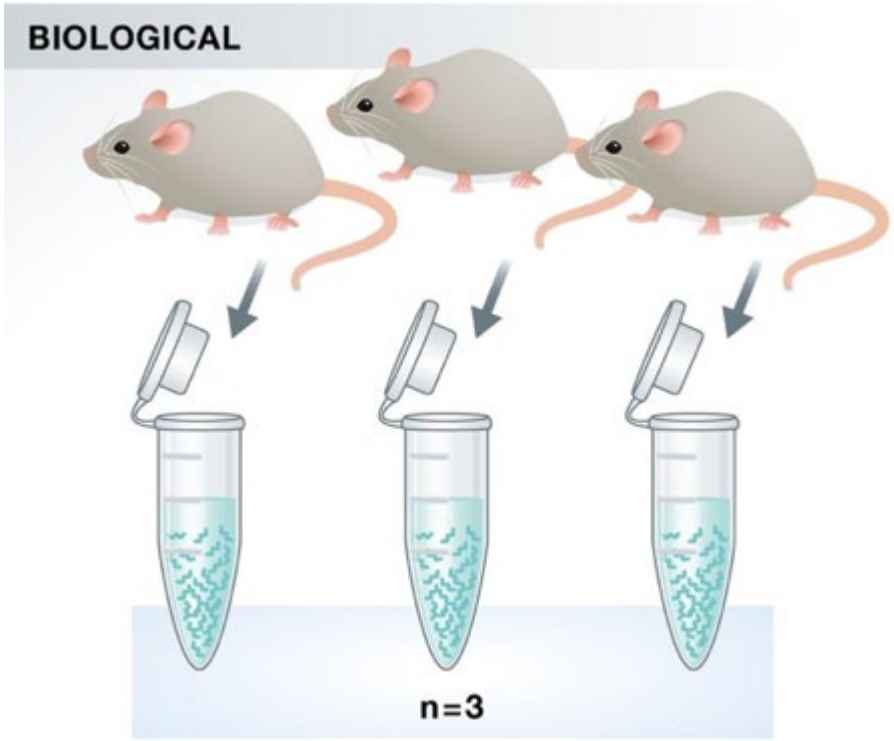


https://xkcd.com/882/

Biological vs Technical – Pseudoreplication



Technical Replicate:
Use of the same biological entity to repeat experimental steps to *accurately measure technical variation and assess experimental consistency*



Biological Replicate:
Use of different biological samples under the same set of research conditions to *measure inherent biological variation due to the experiment*

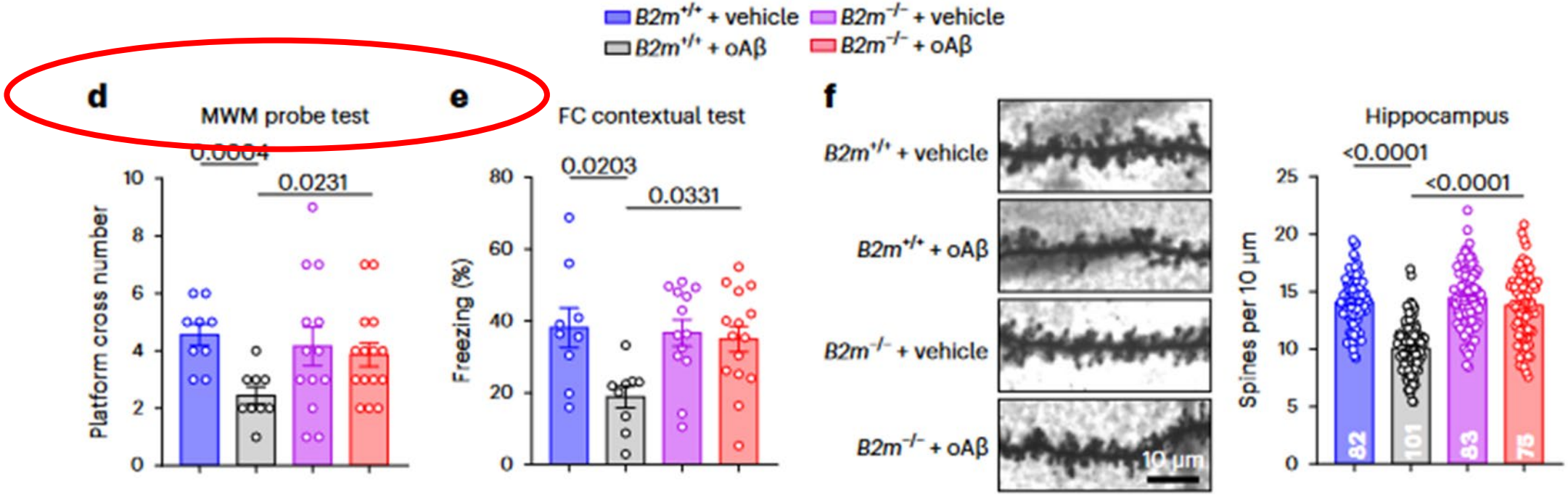
Klaus B. *EMBOJ* 34:2727-2730. doi: 10.15252/embj.201592958.

Figures and Tables

- Adequately labelled with descriptive titles and captions
- Tables should have precise values / significant digits on calculation
- Figures should have:
  - Clearly labelled axes with units
  - Sufficient breaks and ticks
  - Full ranges as needed; allowing for floor/ceiling values
  - Descriptive legends and labels
  - Definition of acronyms
- **There is no better abuse of statistics than a misleading figure**

Thinking at the Mouse Level

Fig. 3 | *B2m* deficiency abrogates oAβ-induced neurotoxicity in vivo.

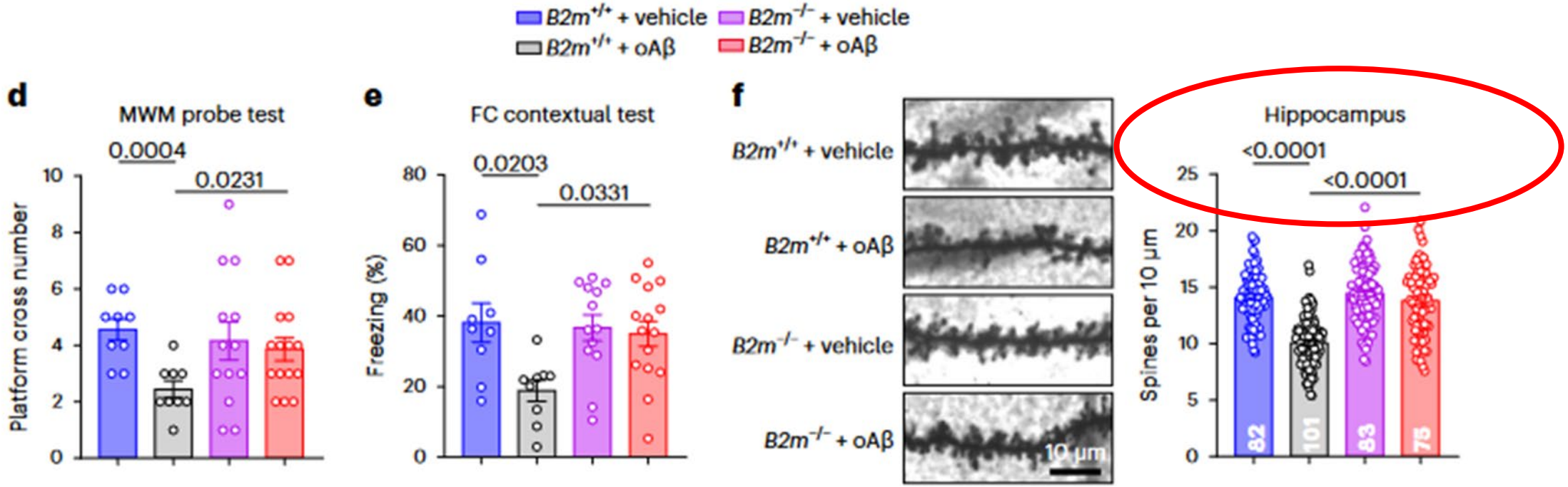


b–**e**, $B2m^{+/+}$ + vehicle, $n = 9$ mice; $B2m^{+/+}$ + oAβ, $n = 9$ mice; $B2m^{-/-}$ + vehicle, $n = 13$ mice; $B2m^{-/-}$ + oAβ mice, $n = 15$ mice.

Zhao Y et. al. *Nature Neuroscience* 26, 1170-1184. doi: 10.1038/s41593-023-01352-1.

But Analyzing at the Dendrite Level

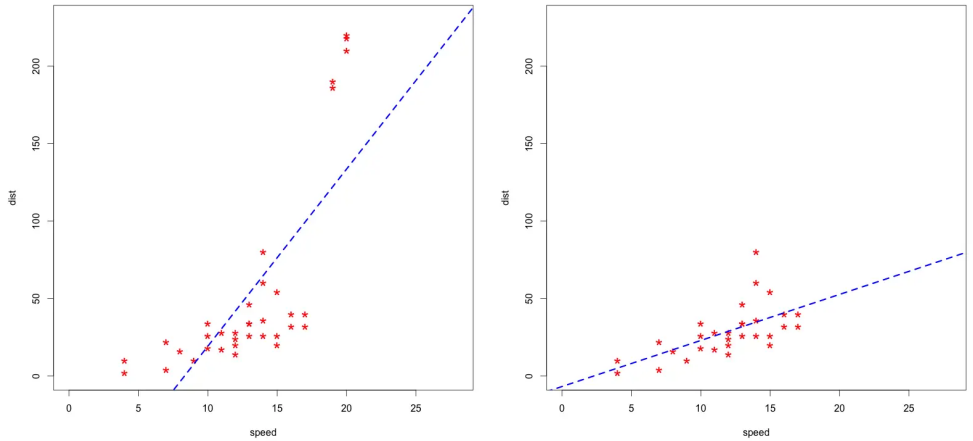
Fig. 3 | *B2m* deficiency abrogates oAβ-induced neurotoxicity in vivo.



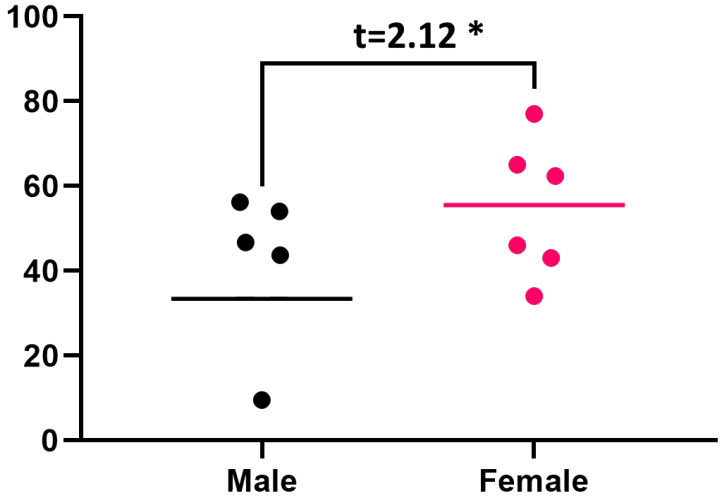
f, Golgi staining and quantification of dendritic spines in the hippocampi of oAβ-injected mice (the number of counted dendrites is indicated on the graphs, $n = 4$ mice per group).

Zhao Y et. al. *Nature Neuroscience* 26, 1170-1184. doi: 10.1038/s41593-023-01352-1.

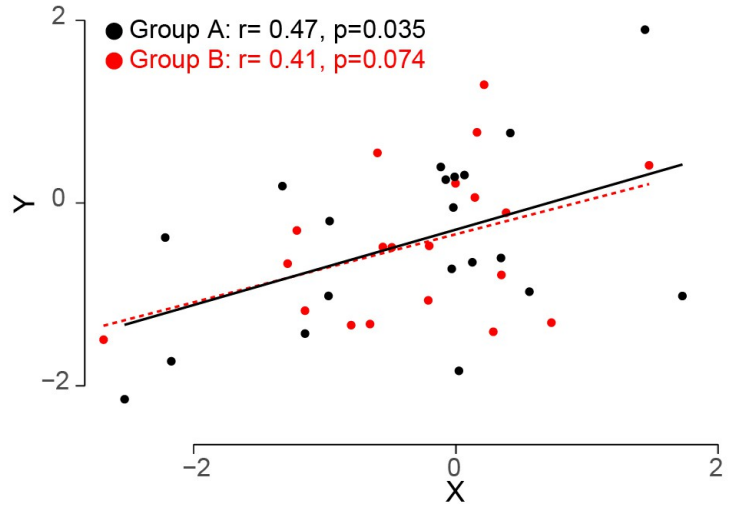
Figures as an Evaluation Tool



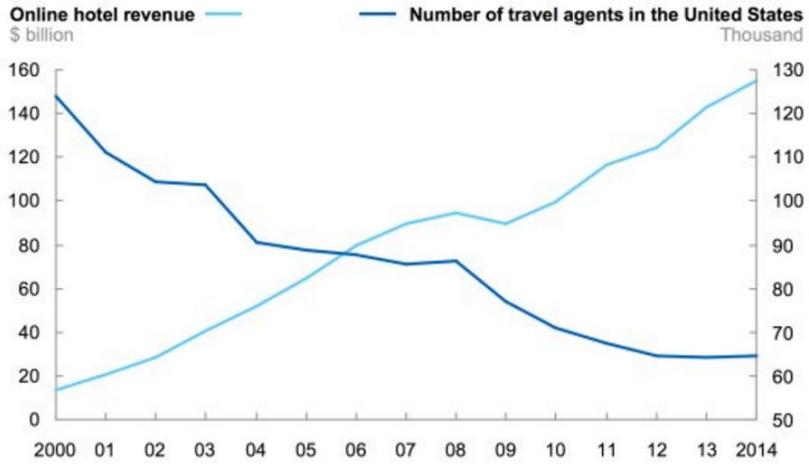
Influential points and spuriousness



Model assumptions or leverage



Indirect group comparisons

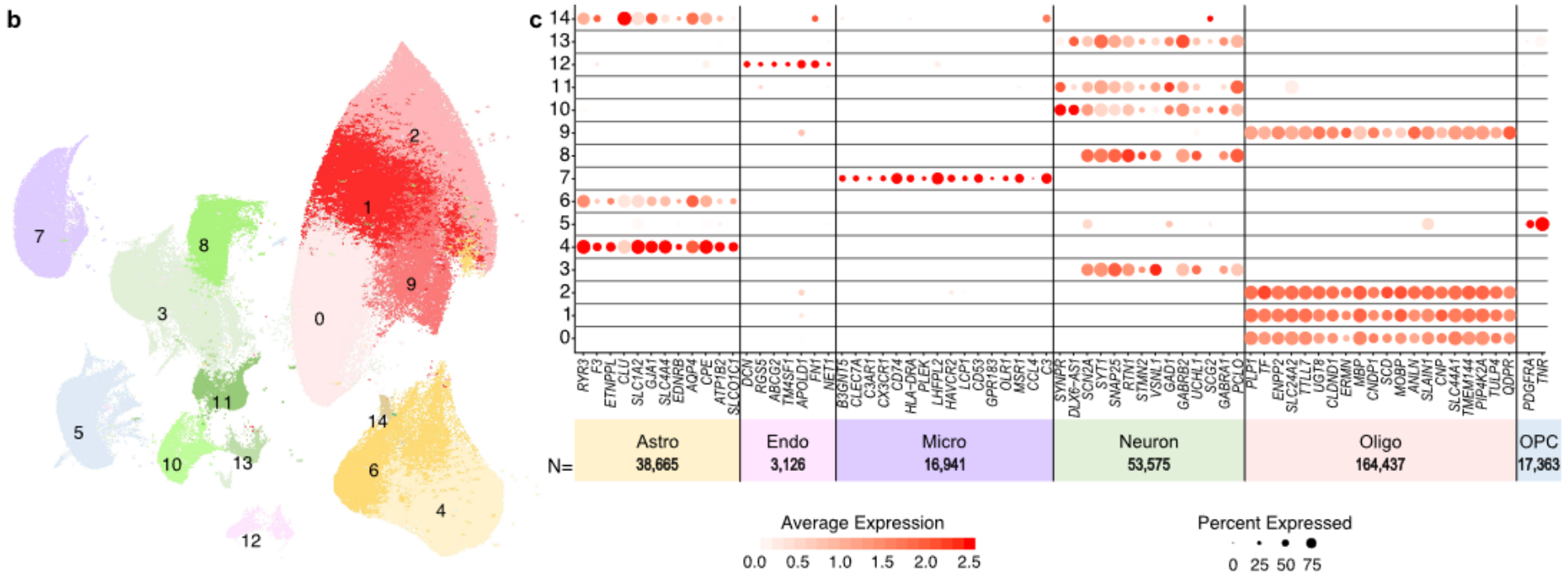


Deceptive axes (USA Today plots)

https://doi.org/10.7554/eLife.48175 / https://www.storytellingwithdata.com / https://online.stat.psu.edu / http://rstatistis.co

Extracting Meaning From a Plot

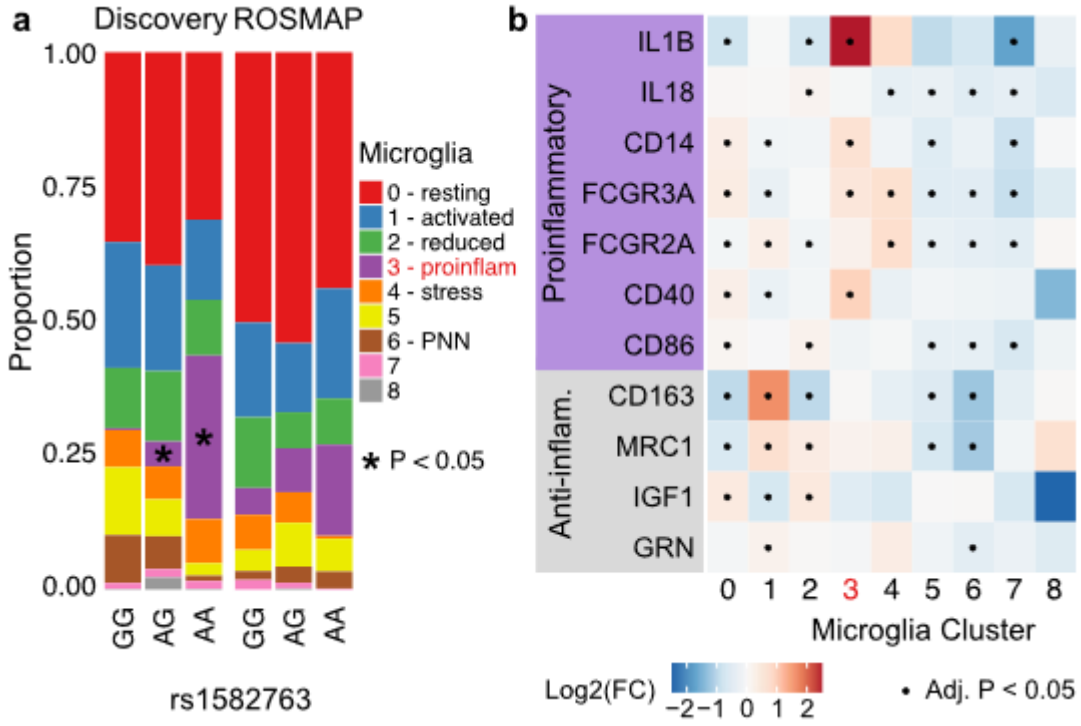
- The best plots are companion pieces to other data presentations
- Every aspect of your figures should convey meaning



Brase L et. al. *Nature Communications* 14, 2314 doi: 10.1038/s41467-023-37437-5.

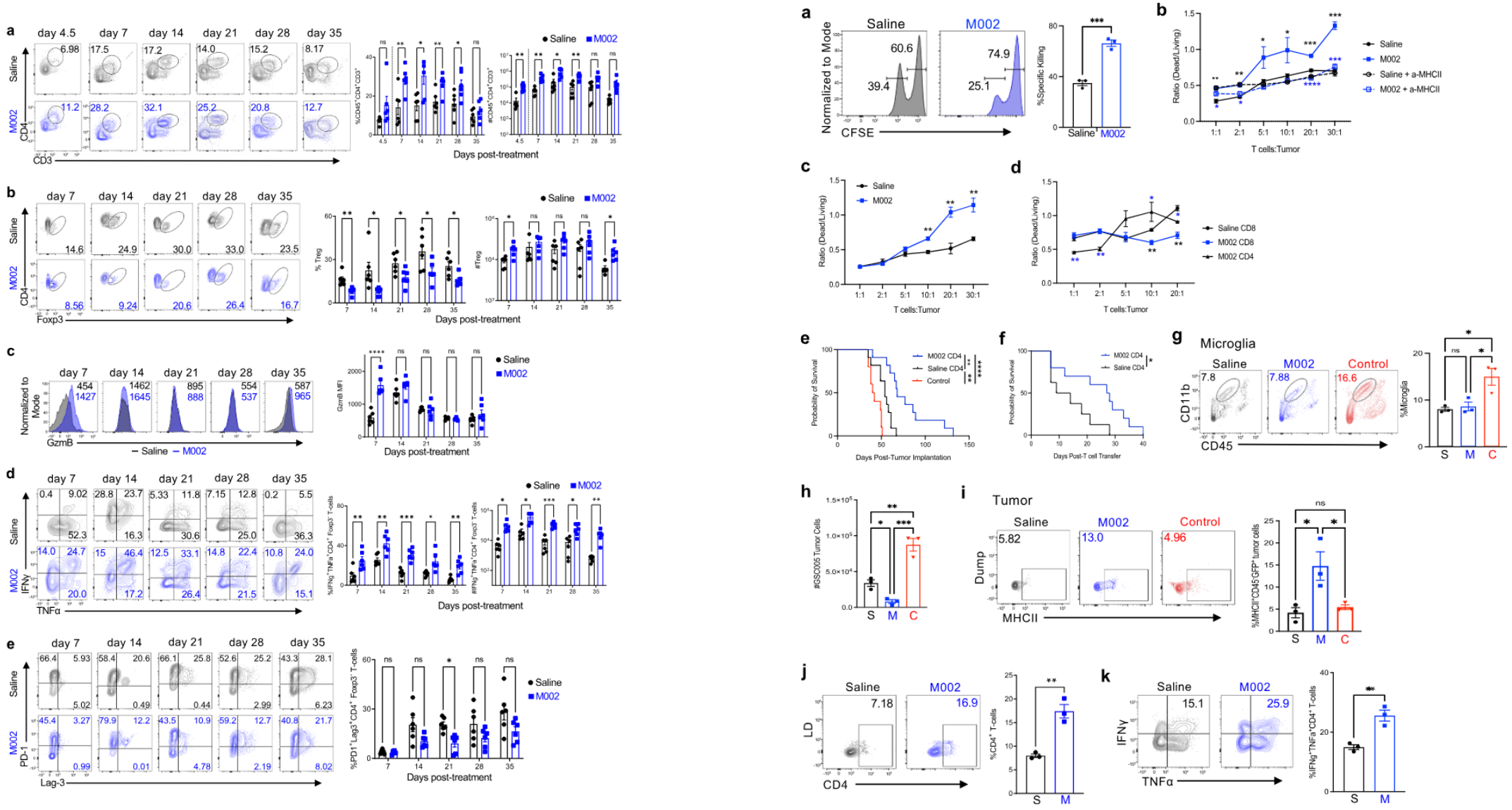
Combining Information for that Meaning

- Heat maps are often relegated to expression or abundance
- Can easily be applied to any scale measure e.g. p-values
- Naturally lends itself to overlays similar to a Dot Plot
- Again, the best plots should inform one another



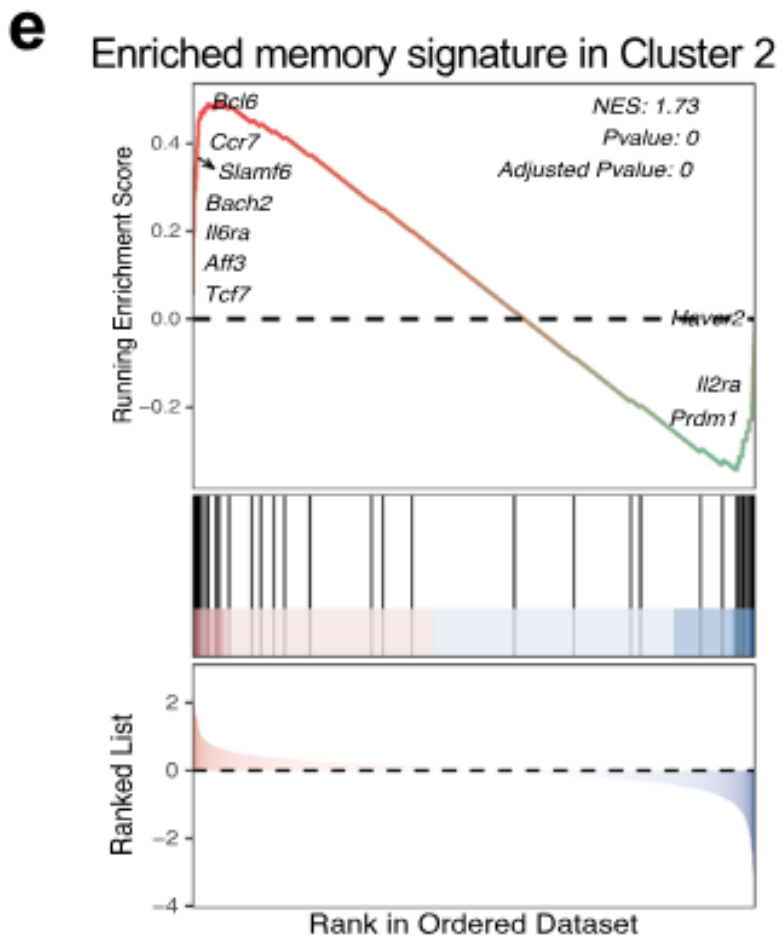
Brase L et. al. *Nature Communications* 14, 2314 doi: 10.1038/s41467-023-37437-5.

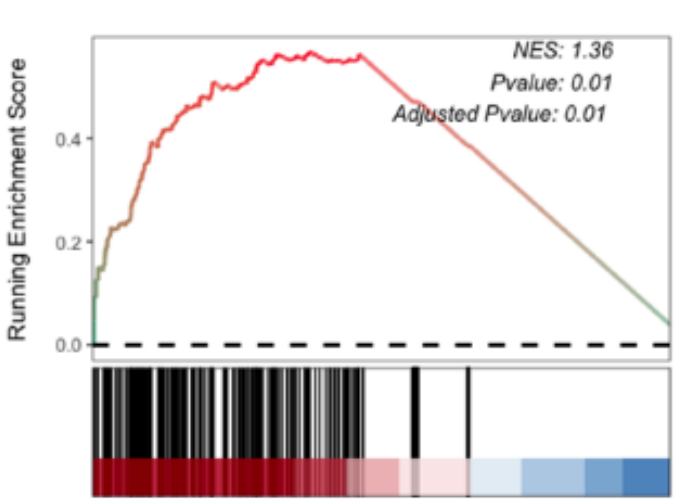
Be Wary of Figure Overload



Grimes J et. al. *Nature Communications* 16, 1095. doi: 10.1038/s41467-024-55455-9.

And Make Sure You Can Explain Them





Enriched memory signature among genes associated with interactions of microglia, myeloid and tumor cells with CD4+ T-cell

e, k GSEA test statistic of enrichment score (ES, one-sided) was used to calculate the normalized enrichment score (NES) using a one-sided permutation test with FDR adjustments for multiple comparisons.

Grimes J et. al. *Nature Communications* 16, 1095. doi: 10.1038/s41467-024-55455-9.

Results and Conclusion – A Disconnect

PD Risk Variant-Based PRS Is Associated with Increased Risk for LID

PRS analyses aggregating PD-associated variants showed that higher values of PRS were associated with a very mild increase in LID risk (OR = 1.02; 95% CI, 1.002–1.035; $P = 0.0298$) (Fig. 3B). When dividing the PRS in quartiles, logistic regression showed a significant association between the fourth quartile and LID, with a greater risk compared to the analyses using PRS as a continuous variable ($OR_{fourth\_quartile} = 1.27$; 95% CI, 1.03–1.56; $P = 0.0210$) (Fig. 3A, Supplementary Table S8). Cox regression did not show any significant associations between PRS and time to development of LID (Supplementary Fig. S5A,B, Supplementary Table S9). The PD PRS logistic regression was significant for a moderate heterogeneity ($I^2 = 43.90\%$, $P = 0.0449$) and repeating the meta-analysis using a random-effect model, which accounts for heterogeneity, the results did not show statistically significant associations (OR = 1.02, $P = 0.2038$). PD PRS Cox regression did not show heterogeneity (I2 = 0%, $P = 0.6236$).

The significant association between the two PRS analyses suggests that aggregating multiple common variants that might have a scarce effect on LID individually could contribute to uncovering the overall genetic impact on LID. In particular, the association between the PRS including PD risk variants suggests that patients with a stronger genetic risk profile for PD are also more at risk for LID, a factor to consider for patient counselling and potential clinical trials, although the magnitude of the increased risk was small.

It's not what you think but how you think

Sosero Y et. al. *Movement Disorders* 2024 39, 1773-1783. doi: 10.1002/mds.29960.

Final Thoughts

- Be sure to present your research question clearly
- Help your audience understand complex experimental designs
- At least one person is going to be looking at your analysis methods
- Your figures are powerful; make sure they carry meaning
- Don’t oversell, interpretation and inference matter
- **Statistical significance and contextual importance are both key**

Resources

- NIH - https://grants.nih.gov/policy/reproducibility/index.htm
- NINDS - https://www.ninds.nih.gov/funding/preparing-your-application/preparing-research-plan/rigorous-study-design-and-transparent-reporting
- SAMPL Guidelines - https://www.equator-network.org/wp-content/uploads/2013/03/SAMPL-Guidelines-3-13-13.pdf
- ARRIVE Checklist - https://arriveguidelines.org/sites/arrive/files/documents/Author%20Checklist%20-%20Full.pdf
- Journals and Organizations
  - STAR Methods for Cell Press - https://www.cell.com/pb-assets/journals/research/cell/methods/Methods_Guide_general-1678470557763.pdf
  - PLoS - https://journals.plos.org/plosone/s/submission-guidelines
    https://plos.org/resource/how-to-report-statistics/
  - Center for Open Science - https://www.cos.io/initiatives/top-guidelines
  - Zenodo - https://about.zenodo.org/policies/