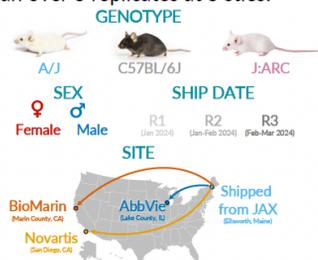


INTRODUCTION

- The **replication crisis** challenges all aspects of preclinical research.
- Factors that complicate cross-site replication are **interlaboratory variation and noise**, perhaps from short-duration studies.
- Digital *in vivo* measures show great promise in enhancing study replicability, potentially through **long-duration studies**.
- Variation in digital *in vivo* studies must be measured and mastered.

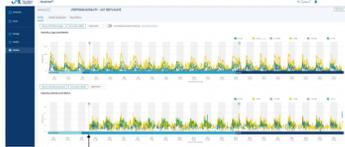
METHODS

- Study design based on a classic study interlaboratory variation study (Crabbe et al., 1999), was run over 3 replicates at 3 sites.



- 9-week replicability study: 21 days per cage, 3 mice per cage, collected **73,504 hours** of individual mouse behavior.
- Advanced machine learning from a beta version of the **JAX Envision™** platform used to generate the outcome measure of cage activity (average speed in cm/s per mouse).

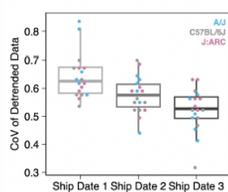
EXPERIMENT TIMELINE (in Envision™ App)



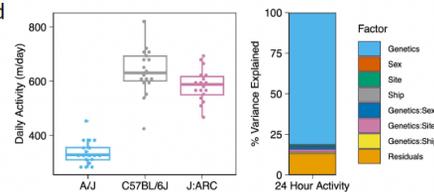
- Analysis used linear modeling:

$$\text{activity} = \beta_0 + \beta_1 \cdot \text{site} + \beta_2 \cdot \text{ship} + \beta_3 \cdot \text{genotype} + \beta_4 \cdot \text{sex} + \beta_5 \cdot (\text{genotype} \times \text{sex}) + \beta_6 \cdot (\text{site} \times \text{genotype}) + \beta_7 \cdot (\text{ship} \times \text{genotype}) + \epsilon$$
- Hourly aggregated data down-sampled at different durations.

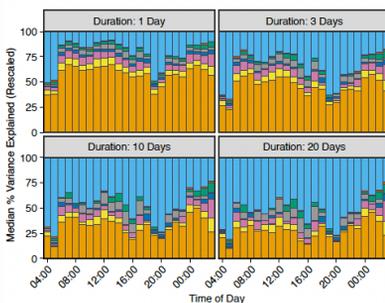
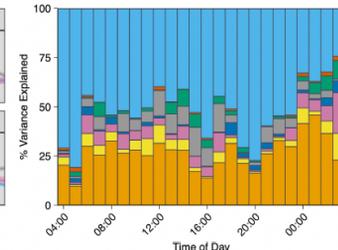
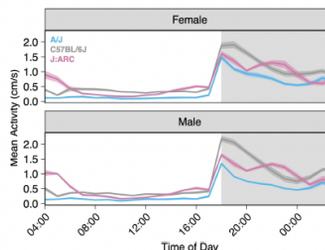
RESULTS



- Statistically significant differences were observed in hour-to-hour variation due to ship date (**left**). Earlier ship dates had higher coefficients of variation.
- Despite this, long-duration digital phenotyping elicits a highly replicable genotype effect in averaged 24-hour aggregates (**right**).

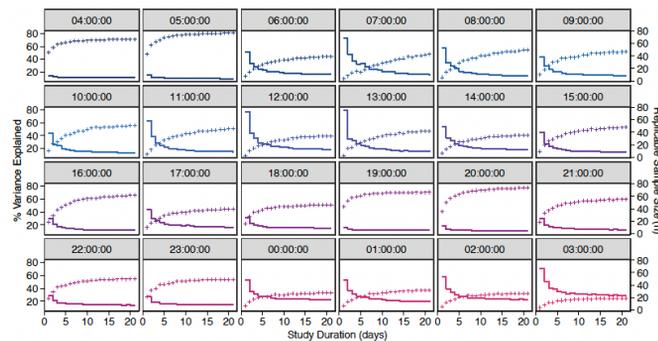


- Hour-to-hour aggregates showed genotype effects varying over the day (**near right**, mean \pm SEM).
- Resulting dataset demonstrates strong genotype effects during most hours.
- Residuals and technical terms were high during typical working hours (**far right**, 08:00-16:00).



- Increasing study duration strongly increased effect size within the first 10 days of the study regardless of the hour of the day sampled (**right**, + signs).
- This resulted in **decreases in sample size required for replicable results** for each hour (**right**, stair step lines).
- Replicable sample size was lowest during the dark cycle.

- Down-sampled dataset indicated that **noise decreased with longer duration** (**left**, see residuals in orange).
- Increasing duration resulted in greater replication of statistically significant effects across sites (**right**, dark gray = replicated findings across three sites).



CONCLUSIONS

- Digital *in vivo* studies **enhance experimental replicability**.
- In contrast to the usual short duration of behavioral experiments, **long-duration sampling in the home cage increases replicability**.
- The **time of day** is an important factor determining experimental replication.
- By sampling animals for longer, it is possible to **decrease the sample size** needed for a replicable experiment.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the important contributions of a multidisciplinary team of members, including Brian Baridon, Brian Berridge, Kristin Blanchette, Christina Boykin, Joseph Brewster, Michael Ellis, Kristin Evans, Ellie Karlsson, Vivek Kumar, Jared Ostmeier, Nicole Peltier, and Vivek Philip.

The Digital In Vivo Alliance (DIVA) is a collaboration of pharmaceutical industry and academic scientists with a shared interest in the validation and application of AI-enabled *in vivo* digital measures. For more information, visit DIVA.bio.

