

# Operant behavior is reliably impaired in a mouse model of Angelman syndrome

**Caleigh D. Guynes**

Children's National Hospital

**Devante D. Kerr**

Children's National Hospital

**Zahraa S. Hotait**

Children's National Hospital

**Joseph K. Tanas**

Children's National Hospital

**Michael S. Sidorov**

[msidorov@childrensnational.org](mailto:msidorov@childrensnational.org)

Children's National Hospital

---

## Article

**Keywords:** Angelman syndrome, mouse behavior, neurodevelopmental disorders, Ube3a

**Posted Date:** March 20th, 2026

**DOI:** <https://doi.org/10.21203/rs.3.rs-9043922/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations:** No competing interests reported.

---

**Version of Record:** A version of this preprint was published at Scientific Reports on September 13th, 2026. See the published version at <https://doi.org/10.1038/s41598-026-71483-5>.

1 **Title:** Operant behavior is reliably impaired in a mouse model of Angelman syndrome

2

3 **Authors:** Caleigh D. Guoynes<sup>1&</sup>, Devante D. Kerr<sup>1&</sup>, Zahraa S. Hotait<sup>1</sup>, Joseph K. Tanas<sup>1</sup>, and Michael S.  
4 Sidorov<sup>1,2\*</sup>

5

6 **Affiliations:**

7 <sup>1</sup>Center for Neuroscience Research, Children's National Hospital, Washington, DC, USA

8 <sup>2</sup>Departments of Pediatrics and Pharmacology & Physiology, The George Washington University School of  
9 Medicine and Health Sciences, Washington, DC, USA

10

11 **&Indicates equal contributions** (CDG and DDK)

12

13 **\*Corresponding author:** Michael S. Sidorov (msidorov@childrensnational.org)

14

15

## ABSTRACT

Angelman syndrome (AS) is a neurodevelopmental disorder characterized by cognitive impairment, absent speech, seizures, motor and sleep impairments, and caused by loss of function of the maternal copy of the *UBE3A* gene. Modeling the AS phenotype in preclinical models across behavioral domains is critical for testing of new therapeutics. Our previous work reported behavioral phenotypes in male *Ube3a<sup>m-/p+</sup>* (AS model) mice on a cognitive-like operant extinction task. Here, we evaluated operant task performance in male and female *Ube3a<sup>m-/p+</sup>* mice across two cohorts, and correlated operant performance with performance on a battery of other widely used behavioral tests for *Ube3a* mutants. *Ube3a<sup>m-/p+</sup>* mice showed impairments in both the acquisition and extinction phases of operant testing that were not sex-specific. Inclusion of operant testing in the existing *Ube3a<sup>m-/p+</sup>* behavioral battery was feasible and improved clustering of genotypes in principal component space. Overall, operant acquisition and extinction testing is a reliable approach to quantify cognitive-like learning impairments in *Ube3a<sup>m-/p+</sup>* mice. Operant testing should be considered as part of a broad toolbox for evaluating the effectiveness of AS treatments preclinically.

**Keywords:** Angelman syndrome, mouse behavior, neurodevelopmental disorders, Ube3a

## INTRODUCTION

Angelman syndrome (AS) is a single-gene neurodevelopmental disorder marked by motor and cognitive impairment, developmental delay, absence of speech, seizures, sleep disturbances, and differences in social affective behavior<sup>1,2</sup>. Cognitive impairment in individuals with AS is pervasive, affecting quality of life for affected individuals and caregivers<sup>3-7</sup>. AS is caused by a loss of functional maternal *UBE3A* expression due to deletion, mutation, imprinting center defect, or paternal uniparental disomy<sup>8,9</sup>. *UBE3A* is paternally imprinted in neurons<sup>10-12</sup>, and approaches to re-express *UBE3A* via unsilencing the dormant paternal allele are now under evaluation in clinical trials<sup>13-15</sup>. As ongoing work identifies additional mechanism-based approaches to restore *UBE3A* function<sup>16-28</sup> and/or target downstream pathways<sup>29-38</sup>, preclinical testing of potential therapies remains critical.

Behavioral testing in *Ube3a*<sup>m-/p+</sup> mice<sup>39</sup> is a widely used approach to evaluate the preclinical effectiveness of potential AS therapies. This mouse model has strong construct validity, as *Ube3a* imprinting is conserved in mice<sup>11</sup>, and it has also shown good face validity across a number of behavioral domains<sup>40</sup>. For example, *Ube3a*<sup>m-/p+</sup> mice show impairments in motor function<sup>39,41-44</sup>, increased seizure susceptibility<sup>45-47</sup>, abnormal sleep and circadian rhythms<sup>48-51</sup>, and abnormal EEGs<sup>52-55</sup> that are generally consistent with patient populations. In addition, *Ube3a*<sup>m-/p+</sup> mice show impaired vocalizations in multiple behavioral contexts that are consistent with impaired speech in AS<sup>56,57</sup>. Recently, many groups have used a “gold standard” behavioral testing battery<sup>58</sup> to evaluate the effectiveness of multiple types of interventions<sup>21,25,28,38,59-62</sup>. The main advantage of this battery, which typically includes weight, rotarod, open field, marble burying, nest building, and forced swim, is that behavioral phenotypes are robust and consistent across cohorts of mice and different labs. However, a limit of using this battery alone is its overreliance on motor or motor-adjacent behaviors. More broadly, preclinical testing in AS has been limited by challenges in reliably evaluating cognitive or cognitive-like tests. Modeling cognitive impairment is inherently challenging in mice, and thus, new *Ube3a*<sup>m-/p+</sup> rat and pig models have been developed that are capable of more complex behavior<sup>63-65</sup>. However, evaluating cognitive-like behavior in mice remains important, as there are more advanced genetic tools in mice and many preclinical studies first screen drugs in mice.

Modeling cognitive impairment in *Ube3a*<sup>m-/p+</sup> mice can be approached in several ways. One approach is to use cognitive-like learning assays, such as novel object recognition, contextual fear conditioning, and the Morris water maze. *Ube3a*<sup>m-/p+</sup> mice generally show impairments on these tests<sup>44,66-71</sup>, though phenotypes are typically less reliable than the gold standard battery. However, these tasks are primarily dependent on the hippocampus and other subcortical brain structures<sup>72-74</sup>, which limits their interpretability as a broader assessment of cognitive impairment. An alternative approach for modeling cognitive impairment is to use operant learning tasks, such as operant acquisition and extinction and the five-choice serial reaction time test, that actively engage prefrontal circuits<sup>75-77</sup>. The prefrontal cortex (PFC) is central to cognition and critical for

high-level tasks in both humans and mice<sup>78,79</sup>, and *Ube3a*<sup>m-/p+</sup> mice show both electrophysiological impairments in PFC and behavioral impairments on operant tasks linked to PFC function<sup>80-83</sup>. Our prior work reported exaggerated extinction and a trend towards delayed acquisition of an operant learning task in adult male *Ube3a*<sup>m-/p+</sup> mice<sup>83</sup>. Here, we sought to evaluate the reliability and suitability of operant behavior for preclinical testing by assessing performance in male and female mice from multiple behavioral cohorts.

In addition, a secondary goal of this study was to evaluate the feasibility and value of adding operant testing to the existing *Ube3a*<sup>m-/p+</sup> behavioral battery. We previously demonstrated that using principal component analysis, performance on the gold standard *Ube3a*<sup>m-/p+</sup> behavioral battery can be summarized as a single “Angelman severity score” (PC1) that is sensitive to treatment<sup>84</sup>. Here, we performed operant testing and the behavioral battery concurrently, and asked whether inclusion of operant measures in multidimensional analysis improved our ability to distinguish *Ube3a* genotype based on behavior. This approach is conceptually and methodologically similar to our recent finding that including ultrasonic vocalization testing improves the breadth and effectiveness of the *Ube3a*<sup>m-/p+</sup> behavioral battery<sup>57</sup>.

## RESULTS

### *Operant acquisition and extinction are abnormal in Ube3a<sup>m-/p+</sup> mice*

We performed operant behavioral testing followed by a well-established behavior battery in adult male and female *Ube3a*<sup>m-/p+</sup> mice and littermate controls (**Fig. 1a**). Operant testing had three distinct stages (**Fig. 1b**): (1) magazine training, where food-deprived mice learned that pellet rewards were delivered via the food magazine; (2) acquisition, where a nose-poke into the light-cued aperture triggered reward delivery via the magazine; (3) extinction, where nose-pokes no longer triggered reward delivery. First, we analyzed the entire dataset together, which included two cohorts of mice, and males and females in equal ratios. There was no genotype difference in the number of pellets received during magazine training (**Fig. 1c**;  $t_{(54)} = 1.240$ ,  $p = 0.2202$ ), suggesting no gross difference in motivation. During acquisition, mice learned across days (**Fig. 1d**) and advanced to extinction after performing at criteria (>15 trials, >50% accuracy) for five consecutive days. *Ube3a*<sup>m-/p+</sup> mice took longer to reach operant acquisition learning criteria (**Fig. 1e**;  $t_{(54)} = 3.122$ ,  $p = 0.0030$ ). At criteria, *Ube3a* mice had fewer cued responses (**Fig. 1f**;  $t_{(54)} = 3.245$ ,  $p = 0.0020$ ) and fewer non-cued responses (**Fig. 1g**;  $t_{(54)} = 4.240$ ,  $p < 0.0001$ ), resulting in a higher accuracy (**Fig. 1h**;  $t_{(54)} = 3.380$ ,  $p = 0.0014$ ). Because response rates during acquisition were different by genotype, extinction was normalized to acquisition as in prior work<sup>83</sup>.

Extinction of cued responses (**Fig. 1i-j**; main effect of time:  $F_{(1.965, 106.1)} = 92.90$ ,  $p < 0.0001$ ) was exaggerated in *Ube3a*<sup>m-/p+</sup> mice (**Fig. i-j**; main effect of genotype:  $F_{(1, 54)} = 6.078$ ,  $p = 0.0169$ ; genotype  $\times$  time

interaction:  $F_{(2,108)} = 5.621$ ,  $p = 0.0048$ ). *Post hoc* tests revealed a significant group difference on extinction day 1 ( $p = 0.0048$ ). Non-cued responses during the extinction phase showed no genotype differences (Fig. 1k-l; main effect of genotype:  $F_{(1,54)} = 1.371$ ,  $p = 0.2469$ ; genotype  $\times$  time interaction:  $F_{(2,108)} = 0.4535$ ,  $p = 0.6366$ ). Accuracy during extinction also showed no genotype differences (main effect of genotype:  $F_{(1,54)} = 1.893$ ,  $p = 0.1745$ ; genotype  $\times$  time interaction:  $F_{(2,108)} = 1.187$ ,  $p = 0.3090$ ) (Fig. 1m-n). Extinction is plotted using raw data in Supplemental Figure 1. Overall, these results are consistent with our previous finding that *Ube3a*<sup>m/p+</sup> mice are more accurate at criteria during operant acquisition and that operant extinction is enhanced in *Ube3a*<sup>m/p+</sup> mice. However, in this cohort, we also found that *Ube3a*<sup>m/p+</sup> mice took longer to reach acquisition criteria.

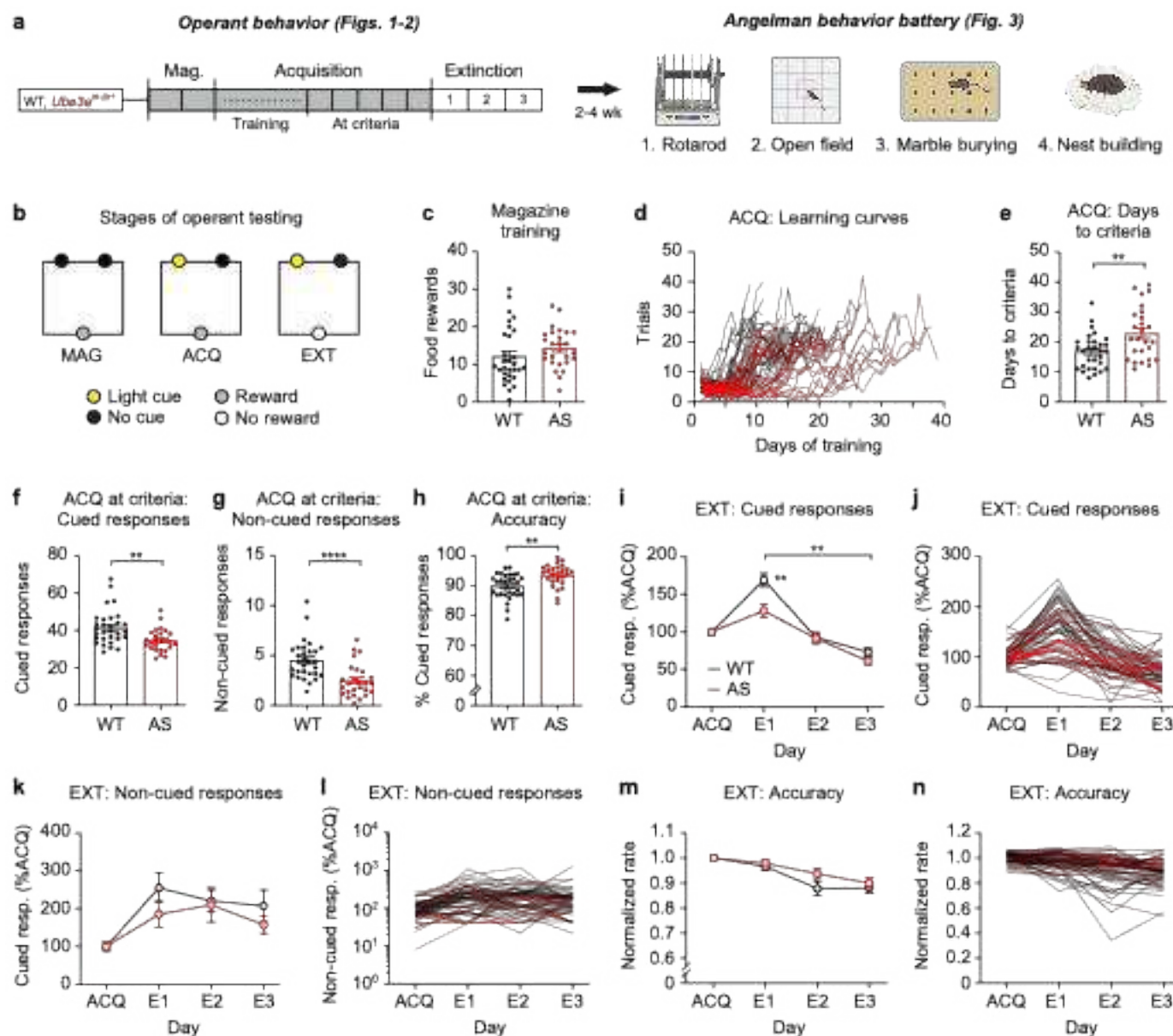


**Figure 1. Operant acquisition and extinction phenotypes in *Ube3a*<sup>m/p+</sup> mice.** (a) Experimental timeline for evaluating operant behavior and the Angelman behavior battery in the same animals. (b) Schematic of three stages of operant testing: magazine training (MAG), acquisition (ACQ), extinction (EXT). (c) No difference in

magazine training between WT and *Ube3a*<sup>m-/p+</sup> mice. (d) Learning curves during acquisition. (e) *Ube3a*<sup>m-/p+</sup> mice took more days to reach acquisition criteria. (f) *Ube3a*<sup>m-/p+</sup> mice had fewer cued responses at criteria. (g) *Ube3a*<sup>m-/p+</sup> mice had fewer non-cued responses at criteria. (h) *Ube3a*<sup>m-/p+</sup> mice were more accurate at criteria. (i) *Ube3a*<sup>m-/p+</sup> mice made fewer cued responses on extinction day 1. Bracket indicates day × genotype interaction; E1 asterisks represent *post hoc* test. (j) Cued responses during extinction in individual mice. (k) No genotype difference in non-cued responses during extinction. (l) Non-cued responses during extinction in individual mice. Four instances of zero non-cued responses are not plotted here because data are shown on a log scale, but were included in group analysis in K. (m) No genotype difference in accuracy during extinction. (n) Accuracy during extinction in individual mice. Data represent mean ± SEM; \*\*\*\**p* < 0.0001; \*\**p* < 0.01; black: WT (*n* = 30 mice), red: *Ube3a*<sup>m-/p+</sup> (AS; *n* = 26).

### *Operant behavioral phenotypes in Ube3a*<sup>m-/p+</sup> mice are generally consistent across experimental cohorts

Next, we asked whether operant *Ube3a*<sup>m-/p+</sup> behavioral phenotypes (days to criteria, accuracy at criteria, extinction) were consistent across experimental cohorts and sex. We compared behavior in Cohort 1 (WT: *n* = 13, *Ube3a*<sup>m-/p+</sup>: *n* = 11) and Cohort 2 (WT: *n* = 17, *Ube3a*<sup>m-/p+</sup>: *n* = 15), which were run using the same equipment by two separate experimenters. Cohort 2 used slightly smaller pellets for reward delivery (14 mg) than Cohort 1 (20 mg). For operant acquisition (days to criteria), there was a significant main effect of genotype (**Fig. 2a**;  $F_{(1, 52)} = 10.38$ , *p* = 0.0022) and no main effect of cohort ( $F_{(1, 52)} = 0.1209$ , *p* = 0.7295) or cohort × genotype interaction ( $F_{(1, 52)} = 1.168$ , *p* = 0.2848). *Post hoc* tests revealed a statistically significant effect in Cohort 1 (*p* = 0.0127) but not Cohort 2 (*p* = 0.2037). For accuracy at acquisition criteria, there was a significant main effect of genotype (**Fig. 2b**;  $F_{(1, 52)} = 12.48$ , *p* = 0.0009) and no main effect of cohort ( $F_{(1, 52)} = 2.654$ , *p* = 0.1094) or cohort × genotype interaction ( $F_{(1, 52)} = 0.8380$ , *p* = 0.3642). *Post hoc* tests revealed a statistically significant effect in Cohort 1 (*p* = 0.0098) but not Cohort 2 (*p* = 0.0988). For extinction of cued responses, there was a significant day × genotype interaction in both Cohort 1 (**Fig. 2c**; main effect of genotype:  $F_{(1, 22)} = 3.873$ , *p* = 0.0618; interaction:  $F_{(2, 44)} = 3.960$ , *p* = 0.0262; *post hoc*: E1 *p* = 0.0575, E2 *p* = 0.5218, E3 *p* = 0.9979) and Cohort 2 (**Fig. 2c**; main effect of genotype:  $F_{(1, 30)} = 2.603$ , *p* = 0.1171; interaction:  $F_{(2, 60)} = 3.554$ , *p* = 0.0348; *post hoc*: E1 *p* = 0.0962, E2 *p* = 0.9226, E3 *p* = 0.2119). Across cohorts, the cued extinction response was the most reliable. Although the days to criteria and accuracy at criteria during acquisition both had main effects of genotype, *post hoc* tests were only significant for Cohort 1.

### *Operant behavioral phenotypes are generally consistent in Ube3a*<sup>m-/p+</sup> males and females

We then compared behavior in males (WT: *n* = 16, *Ube3a*<sup>m-/p</sup>: *n* = 12) and females (WT: *n* = 14, *Ube3a*<sup>m-/p</sup>: *n* = 14) across the full dataset. For operant acquisition (days to criteria), there was a significant main effect of genotype (**Fig. 2d**;  $F_{(1, 52)} = 9.506$ , *p* = 0.0033) and a main effect of sex ( $F_{(1, 52)} = 8.174$ , *p* = 0.0061), with females taking longer to learn the task, but no genotype × sex interaction ( $F_{(1, 52)} = 0.3673$ , *p* = 0.5471).

Post hoc tests revealed a statistically significant effect in females ( $p = 0.0227$ ) but not males ( $p = 0.1670$ ). For accuracy at acquisition criteria, there was a significant main effect of genotype (**Fig. 2e**;  $F_{(1,52)} = 11.41$ ,  $p = 0.0014$ ) and no main effect of sex ( $F_{(1,52)} = 0.5311$ ,  $p = 0.4694$ ) or genotype  $\times$  sex interaction ( $F_{(1,52)} = 0.00076$ ,  $p = 0.9780$ ). Post hoc tests revealed statistically significant genotype differences in both males ( $p = 0.0401$ ) and females ( $p = 0.0415$ ). For extinction of cued responses, there was a significant day  $\times$  genotype interaction in males (**Fig. 2f**; main effect of genotype:  $F_{(1,26)} = 0.1342$ ,  $p = 0.7171$ ; interaction:  $F_{(2,52)} = 4.422$ ,  $p = 0.0168$ ; post hoc: E1  $p = 0.2167$ , E2  $p = 0.3421$ , E3  $p = 0.9985$ ). Females showed a significant main effect of genotype (Fig. 2F;  $F_{(1,26)} = 9.712$ ,  $p = 0.0044$ ; post hoc: E1  $p = 0.0205$ , E2  $p = 0.1705$ , E3  $p = 0.0560$ ) but no day  $\times$  genotype interaction ( $F_{(2,52)} = 1.103$ ,  $p = 0.3395$ ). Overall, the operant acquisition and extinction phenotypes were robust in both males and females; however, the increased days to acquisition were only significant for females.

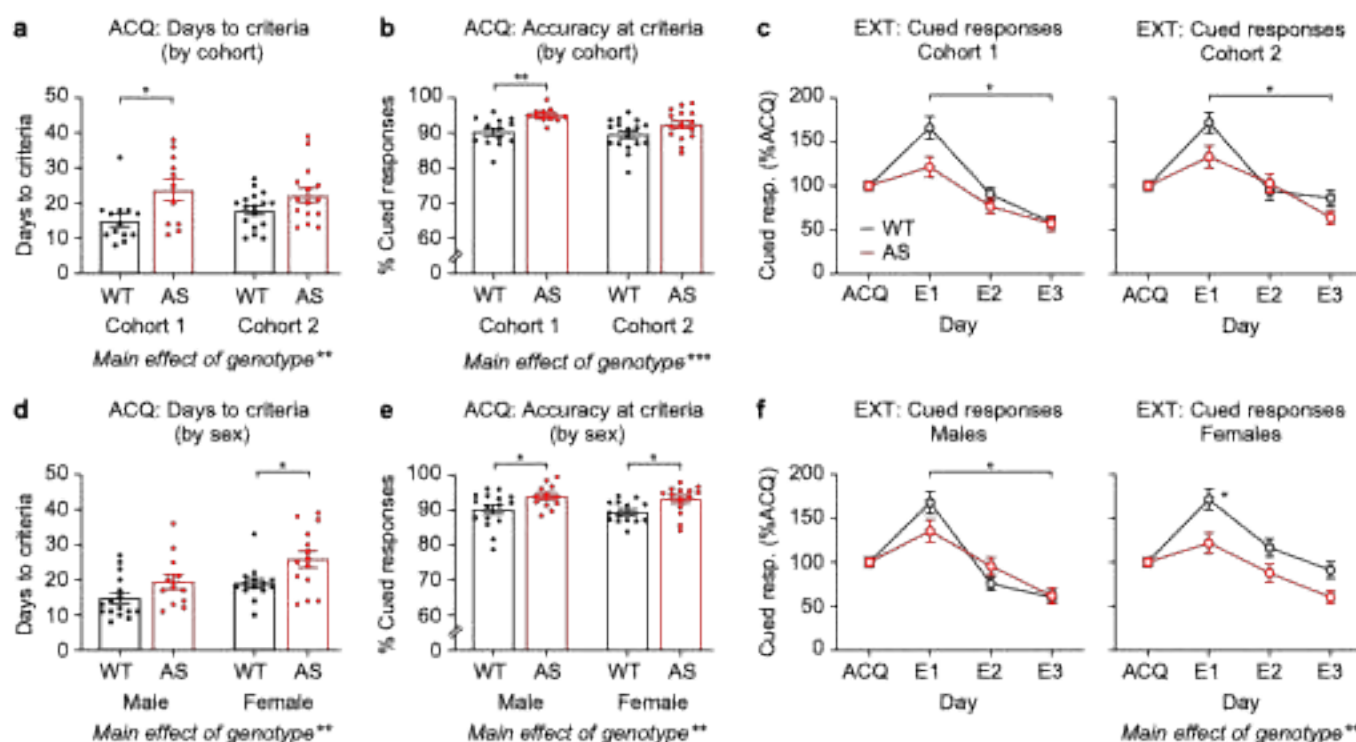


**Figure 2. Evaluation of operant phenotypes in *Ube3a<sup>m/p+</sup>* mice by cohort and sex.** (a) Acquisition days to criteria and (b) accuracy at criteria by cohort. (c) Left: cued responses during extinction in Cohort 1. Right: cued responses during extinction in Cohort 2. (d) Days to criteria and (e) accuracy at criteria by sex. (f) Left: cued responses during extinction in males. Right: cued responses during extinction in females. Data represent mean  $\pm$  SEM; \*\* $p < 0.01$ ; \* $p < 0.05$ ; black: WT, red: *Ube3a<sup>m/p+</sup>* (AS). Brackets in C and F represent day  $\times$  genotype interaction, other asterisks represent post hoc tests. For panels a-c, [Cohort 1: WT ( $n = 13$  mice), AS ( $n = 11$  mice); Cohort 2: WT ( $n = 17$  mice), AS ( $n = 15$  mice)]. For panels d-f, [Males: WT ( $n = 16$  mice), AS ( $n = 12$  mice); Females: WT ( $n = 14$  mice), AS ( $n = 14$  mice)]. When there is a main effect of genotype, it is indicated with text under the graph.

*Operant behavior complements and enhances the existing  $Ube3a^{m-/p+}$  behavioral battery*

Next, we sought to evaluate the added value of including operant behavioral testing into the existing motor-focused  $Ube3a^{m-/p+}$  behavioral battery. To do this, we performed the battery (weight, rotarod, open field, marble burying, nest building) in the same animals after operant testing concluded. We then correlated operant measures with overall performance on the  $Ube3a^{m-/p+}$  behavioral battery. As expected,  $Ube3a^{m-/p+}$  mice showed impaired performance on individual tests in this battery.  $Ube3a^{m-/p+}$  mice were heavier (**Fig. 3a**; main effect of genotype:  $F_{(1, 52)} = 40.69$ ,  $p < 0.0001$ ), showed impaired performance on the rotarod (**Fig. 3b**; main effect of genotype:  $F_{(1, 52)} = 42.33$ ,  $p < 0.0001$ ), marble burying (**Fig. 3c**; main effect of genotype:  $F_{(1, 52)} = 35.47$ ,  $p < 0.0001$ ), and nest building (**Fig. 3d**; main effect of genotype:  $F_{(1, 52)} = 54.50$ ,  $p < 0.0001$ ) tests. Surprisingly,  $Ube3a^{m-/p+}$  mice did not show a phenotype in open field testing (**Fig. 3e**; main effect of genotype:  $F_{(1, 52)} = 0.6670$ ,  $p = 0.4178$ , **Fig. 3f**; main effect of genotype:  $F_{(1, 52)} = 0.1391$ ,  $p = 0.7107$ ). All statistical testing from Figure 3 is reported in Supplementary File 1. These results are generally consistent with prior findings from our group and other groups<sup>58,84</sup>, except that mice that had previously performed operant testing did not show a genotype difference in distance traveled in subsequent open field testing.

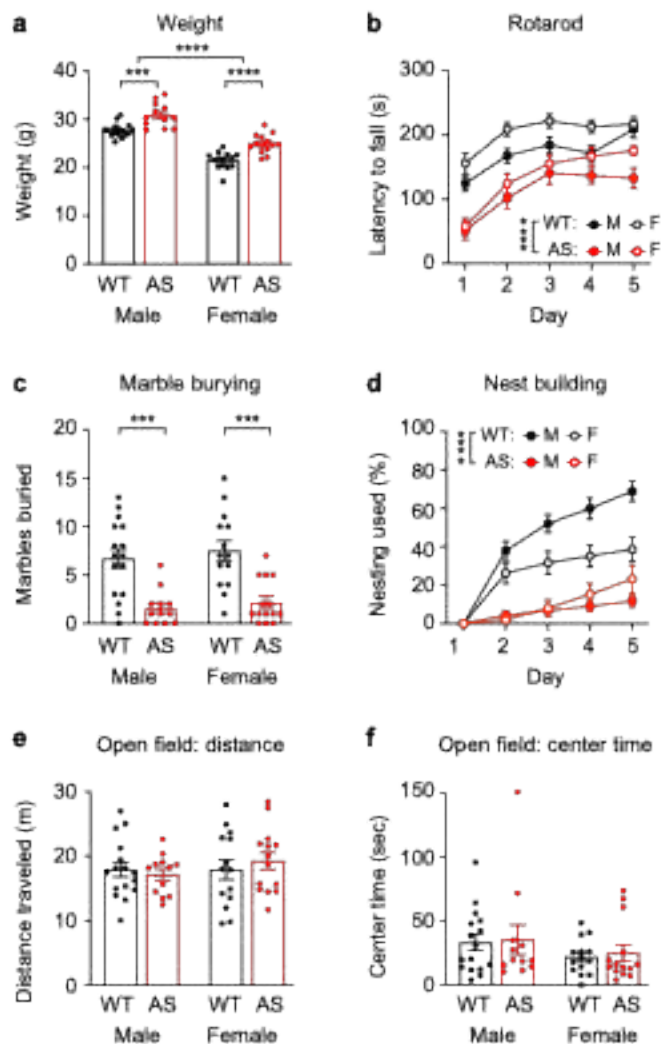
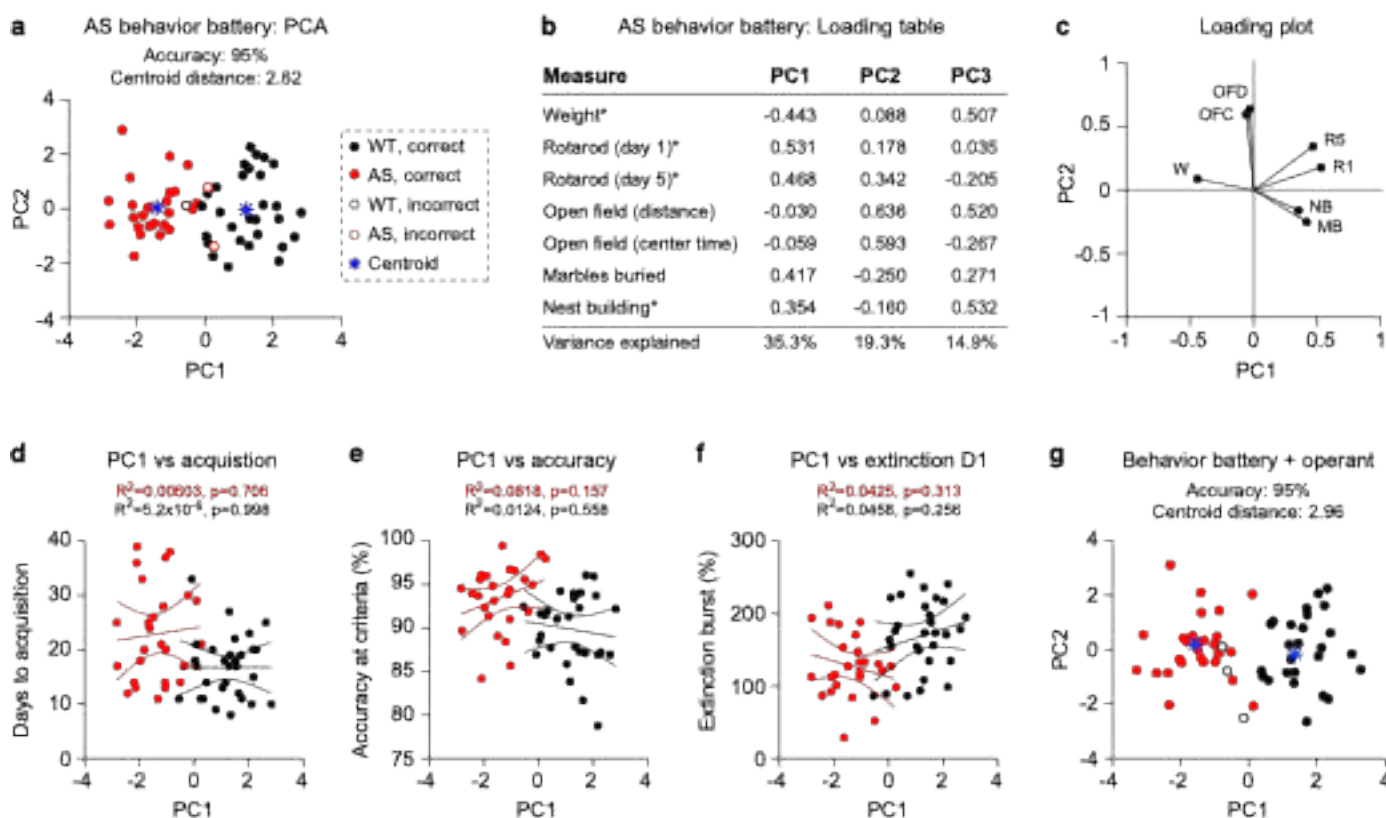


**Figure 3. *Ube3a*<sup>m-/p+</sup> mice show behavioral impairments on a widely used behavioral battery.** (a) *Ube3a*<sup>m-/p+</sup> mice were heavier than WT littermates. Top bracket indicates main effect of sex; other asterisks represent *post hoc* tests. (b) *Ube3a*<sup>m-/p+</sup> mice had a decreased latency to fall in rotarod performance across days. Asterisks represent main effect of genotype. (c) *Ube3a*<sup>m-/p+</sup> mice buried fewer marbles during the marble burying task. Asterisks indicate *post hoc* tests. (d) *Ube3a*<sup>m-/p+</sup> mice used less nesting material across days. Asterisks indicate main effect of genotype. (e) No difference in open field distance between genotypes. (e) No difference in open field center time between genotypes. Data represent mean ± SEM; \*\*\*\**p* < 0.0001; \*\*\**p* < 0.001; \*\**p* < 0.01; \**p* < 0.05; black: WT, red: *Ube3a*<sup>m-/p+</sup> (AS).

We next performed principal component analysis on the standard battery and assessed correlations with operant performance. Multidimensional analysis of standard behavioral battery assessments (weight, rotarod day 1, rotarod day 5, open field distance, open field center time, marble burying, and nest building) predicted *Ube3a* genotype with 95% accuracy (Fig. 4a; 53/56 mice). The distance between centroids representing each genotype was 2.62 units in 2PC space. All behavioral assessments with significant group differences loaded strongly onto PC1 (Fig. 4b-c), confirming our prior finding that PC1 can be used a

213 composite Angelman behavioral severity score in mice<sup>84</sup>. We next asked whether PC1 correlates with each  
 214 operant behavior phenotype (days to acquisition, accuracy at acquisition, day 1 extinction performance) within  
 215 animals. No operant measures were significantly correlated with PC1. Days to acquisition showed no  
 216 relationship with PC1 in either AS ( $R^2 = 0.0003$ ,  $p = 0.798$ ) or WT ( $R^2 = 5.2 \times 10^{-6}$ ,  $p = 0.998$ ) mice (**Fig. 4d**).  
 217 Similarly, accuracy at criteria was not correlated with PC1 in either AS ( $R^2 = 0.0818$ ,  $p = 0.157$ ) or WT ( $R^2 =$   
 218  $0.0124$ ,  $p = 0.558$ ) mice (**Fig. 4e**). Extinction day 1 performance also showed no significant correlation with  
 219 PC1 in either AS ( $R^2 = 0.0425$ ,  $p = 0.313$ ) or WT ( $R^2 = 0.0458$ ,  $p = 0.256$ ) mice (**Fig. 4f**). These findings  
 220 suggest that operant testing captures behavioral variance that is unique from the standard battery. To test this  
 221 hypothesis, we asked whether including operant assessments in multidimensional analysis would improve its  
 222 effectiveness. Improved effectiveness could be evaluated either by an increased accuracy or by WT and AS  
 223 clusters moving farther apart in PC space. Inclusion of operant measures in multidimensional analysis resulted  
 224 in a consistent high accuracy (95%), while increasing the centroid distance between WT and AS clusters from  
 225 2.62 to 2.96 units in PC space (**Fig. 4g**). These results suggest that operant extinction measures provide  
 226 complementary information to the existing behavioral battery and could enhance the sensitivity of  
 227 multidimensional phenotyping approaches for detecting treatment effects in *Ube3a*<sup>m-/p+</sup> mice.



228  
 229 **Figure 4. Operant behaviors are independent of AS battery behaviors and modestly enhance genotype**  
 230 **discrimination.** (a) Multidimensional analysis using data from the *Ube3a*<sup>m-/p+</sup> behavior battery demonstrates  
 231 accurate clustering of animals using two principal components. (b) Loading table and (c) loading plot for PCA  
 232 in panel a. (d) Days to acquisition does not correlate with PC1. (e) Accuracy at criteria does not correlate with  
 233 PC1. (f) The extinction burst response on day 1 does not correlate with PC1. Lines show linear fits with 95%

confidence intervals. (g) Adding operant extinction behaviors to PCA analysis increases centroid distance between WT and *Ube3a*<sup>m-/p+</sup> clusters. Centroids and centroid distance are calculated based on actual genotypes, not k-means clusters.

### Sample size guidance for future operant studies

To guide the design of future operant studies in *Ube3a*<sup>m-/p+</sup> mice, we performed power analyses to determine the sample size required to detect statistically significant effects ( $\alpha = 0.05$ ) for different levels of power ( $1-\beta = 0.8, 0.9$ , and  $0.95$ ) on operant behavioral measures. For each sample size ( $n = 4-30$ ), we ran 10,000 iterations of a bootstrap analysis to determine the minimum number of mice per group. At 80% power, sample sizes ranged from  $n = 20-24$  per measure in operant behavior, and from  $n = 7-12$  for measures in the standard behavior battery (excluding open field, as we did not observe a phenotype). At 90% power, sample sizes increased to  $n = 26-30$  per measure for operant behavior and  $n = 9-15$  for measures in the standard behavior battery. At 95% power, sample sizes were greater than 30 per group for operant behaviors but ranged from  $n = 11-18$  for measures in the standard battery (**Table 1**).

| Behavioral measure   | Achieved effect size<br>(Cohen's d) | Sample size per<br>group<br>( $1-\beta = 0.8$ ) | Sample size per<br>group<br>( $1-\beta = 0.9$ ) | Sample size per<br>group<br>( $1-\beta = 0.95$ ) |
|----------------------|-------------------------------------|-------------------------------------------------|-------------------------------------------------|--------------------------------------------------|
| Days to acquisition  | 0.84                                | 24                                              | 30                                              | >30                                              |
| Accuracy at criteria | 0.89                                | 20                                              | 26                                              | >30                                              |
| Extinction day 1     | 0.91                                | 21                                              | 28                                              | >30                                              |
| Rotarod day 1        | 1.64                                | 7                                               | 9                                               | 11                                               |
| Rotarod day 5        | 1.22                                | 12                                              | 15                                              | 18                                               |
| Open field distance  | 0.07                                | >30                                             | >30                                             | >30                                              |
| Marbles buried       | 1.62                                | 7                                               | 9                                               | 11                                               |
| Nest building        | 1.52                                | 9                                               | 11                                              | 14                                               |

**Table 1. Bootstrap analysis to guide sample size for operant behavioral testing in *Ube3a* mutants.**

Sample sizes (“n”) required to detect statistically significant effects ( $\alpha = 0.05$ ) for different levels of power ( $1-\beta = 0.8, 1-\beta = 0.9$ , and  $1-\beta = 0.95$ ). Behaviors from operant learning and extinction are above the dashed line, and behaviors from the gold standard behavior battery are below the dashed line.

### DISCUSSION

Overall, our results confirm that operant learning is impaired in *Ube3a*<sup>m-/p+</sup> mice and demonstrate that operant phenotypes are not sex-dependent. Relative to WT littermates, *Ube3a*<sup>m-/p+</sup> mice took longer to learn the

operant acquisition task (**Fig. 1e**), had higher accuracy at criteria (**Fig. 1h**), and showed an exaggerated extinction of cued responses on extinction day 1 (**Fig. 1i**). Together, these phenotypes suggest impaired cognitive function during operant learning. Within animals, operant performance did not correlate with overall performance on the gold-standard behavior battery (**Figs. 4d-f**), suggesting that operant testing may capture a different behavioral domain from existing motor-like tasks in *Ube3a*<sup>m-/p+</sup> mice. Consistent with this idea, including operant measures increased the centroid distance between WT and *Ube3a*<sup>m-/p+</sup> mice in principal component space (**Fig. 4g**), suggesting that these tasks further distinguished the *Ube3a*<sup>m-/p+</sup> phenotype. These data demonstrate that, in addition to being a reliable test, operant extinction captures distinct cognitive-like impairments in *Ube3a*<sup>m-/p+</sup> mice.

This study builds on recent operant behavioral findings in *Ube3a*<sup>m-/p+</sup> mice. We replicated our previous findings that *Ube3a*<sup>m-/p+</sup> male mice have enhanced operant extinction and greater accuracy at criteria<sup>83</sup>, and showed that with sufficient sample sizes, we can also detect significant differences in days to acquisition. While previous studies used only male mice, our current study included both male and female mice. We found that no operant phenotypes were sex dependent, which was valuable to rule out because previous studies in *Ube3a*<sup>m-/p+</sup> mice found sex differences in complex behaviors<sup>85,86</sup>. Accuracy at criteria and extinction phenotypes were present independently in males and females, but days to acquisition was statistically significant only in female *Ube3a*<sup>m-/p+</sup> mice, suggesting that the effect size may be greater in females. Similarly, when we broke operant phenotypes down by each of our two cohorts, not all phenotypes were present individually in each cohort. Only exaggerated extinction on day 1 was independently significant in each cohort (**Fig. 2c**). Acquisition accuracy and days to criteria were independently significant in cohort 1, but not cohort 2 (**Figs. 2a-b**). Breaking down our results by cohort and sex revealed that, to detect significant effects, much larger sample sizes (n=20+) are needed than for behaviors in the gold-standard behavior battery (**Table 1**).

In practice, there are several advantages and disadvantages to operant testing when considering this approach for preclinical testing in *Ube3a* mutants. A major advantage of operant testing is the ability to reliably test cognitive-like behavior. The commonly used behavior battery for *Ube3a*<sup>m-/p+</sup> mice<sup>58</sup> emphasizes motor tasks, and our results suggest that operant testing captures a different aspect of the AS phenotype. We demonstrate that operant testing can feasibly precede the gold-standard behavior battery and that adding a cognitive-like task to this battery can help distinguish the *Ube3a*<sup>m-/p+</sup> phenotype in principal component space. However, operant testing has several key limitations that should guide usage. First, the long test duration (often 4+ weeks of daily testing) provides a practical barrier. While operant learning/extinction is less time-consuming than other prefrontally encoded tasks, such as the five-choice serial reaction time test (~3 months)<sup>58</sup>, it is more time-consuming than hippocampally encoded cognitive-like tasks, such as novel object recognition (~4 days)<sup>87</sup>. Another disadvantage is the large sample size required to detect effects. We report that the sample size needed to find a statistically significant effect ( $\alpha = 0.05$ ) and ( $1-\beta = 0.8$ ) was  $n = 24$  mice for days to acquisition,  $n = 19$  for accuracy at criteria, and  $n = 21$  for extinction. Thus, our two cohorts combined have sufficient power to detect an effect (WT:  $n = 30$ ; *Ube3a*<sup>m-/p+</sup>:  $n = 26$ ), but each cohort individually was underpowered (Cohort 1: WT:  $n = 13$ ; *Ube3a*<sup>m-/p+</sup>:  $n = 11$ ; Cohort 2: WT:  $n = 17$ ; *Ube3a*<sup>m-/p+</sup>:  $n = 15$ ). The

sample size needed to detect a significant effect is much higher than that required for many of the tests in the gold-standard behavior battery, making it more costly and time-consuming to run. Overall, it may be best practice to run the gold-standard behavior battery first and in a separate cohort, and then only run operant behavior if cognitive performance is an important secondary measure.

Limited face validity is another major limitation of operant extinction testing in *Ube3a<sup>m-/p+</sup>* mice. While impairments in task acquisition can be easily interpreted, it is less clear how greater accuracy at task acquisition and enhanced extinction in mice translate directly to cognitive dysfunction in people with AS. An operant behavioral task in mice with potentially greater translatability is the five-choice serial reaction time test (5CSRTT). We previously demonstrated that *Ube3a<sup>m-/p+</sup>* mice had increased omissions during 5CSRTT training, suggesting that attentional behavior is impaired<sup>82</sup>. However, 5CSRTT is much longer and more labor-intensive than operant extinction, and is more susceptible to the confounds of motor dysfunction<sup>82</sup>. The trade-offs between face validity and task complexity and duration remain a challenge for studying cognition in mice. While operant behavioral testing in mice has limited face validity in patient populations, it does provide a useful readout of prefrontal circuit function<sup>82,83,88-90</sup> and a reliable assessment of cognitive-like performance.

A surprising finding from this study was that *Ube3a<sup>m-/p+</sup>* mice did not show hypoactivity during open field testing (**Fig. 3e**). Hypoactivity in the open field is typically one of the more reliable phenotypes in *Ube3a* mutants<sup>54,58,86,91</sup>, and our group has previously reported open field hypoactivity using the same arena<sup>84</sup> and protocol. We hypothesize that the lack of an open field phenotype could be driven by weeks of prior experience in a similar environment – the operant chamber – during prior operant testing. Both the operant testing box and the open field testing box are large, soundproof chambers with square behavior arenas. Prior work has demonstrated that total distance traveled in the open field decreases with experience<sup>92,93</sup>. We hypothesize that experimental mice habituated to the operant testing environment, resulting in decreased distance traveled in a similar open field arena. Indeed, the lack of open field phenotype was driven mainly by decreased distance traveled in WT mice (~20 meters per session in this study vs. ~40 meters in prior work)<sup>84</sup>. This finding has practical implications for running the behavior battery and suggests that best practice may be to run the battery separately from operant extinction if the open field phenotype is important to measure.

Overall, we demonstrate that operant behavioral phenotypes are consistent in *Ube3a<sup>m-/p+</sup>* males and females, and include both differences in acquisition and extinction. Operant testing is feasible in conjunction with other behavioral tests, but it may be best to run operant extinction and the behavior battery in separate cohorts. Operant behavior may be a strong readout for measuring cognitive performance and prefrontal dysfunction, and could be a valuable tool for mechanistic studies and screening therapeutic drugs in AS. This test is likely not suitable to be added to every behavior battery because it is labor-intensive and does not have strong face validity, unlike other behavioral tests such as ultrasonic vocalizations<sup>91</sup> that are also reliable and add an additional behavioral domain but are much easier to implement and have stronger translatability.

## METHODS

### *Animals*

All experimental protocols were approved by the Institutional Animal Care and Use Committee (IACUC) of Children's National Hospital, and were performed in accordance with all relevant guidelines and regulations. All experiments in this study complied with the ARRIVE guidelines<sup>94</sup>. All datasets used *Ube3a*<sup>m-/p+</sup> mice and WT littermate controls (*Ube3a*<sup>m+/p+</sup>) on a congenic C57BL6/J background purchased from The Jackson Laboratory (strain #016590), Bar Harbor, Maine, USA, and bred in a colony in the Children's National Hospital research core animal facility. Experimenters were blind to genotype during all testing and analysis. Experimental WT and *Ube3a*<sup>m-/p+</sup> littermates were generated by crossing female *Ube3a*<sup>m+/p-</sup> mice and male WT mice. Two cohorts of mice were used in this study. Cohort 1 included 24 total adult mice aged ~P60-155 (WT male: *n* = 8, *Ube3a*<sup>m-/p+</sup> male: *n* = 6, WT female: *n* = 5, *Ube3a*<sup>m-/p+</sup> female: *n* = 5). Cohort 2 included 30 total adult mice aged ~P83-114 (WT male: *n* = 8, *Ube3a*<sup>m-/p+</sup> male: *n* = 6, WT female: *n* = 9, *Ube3a*<sup>m-/p+</sup> female: *n* = 9). We did not exclude any mice from analysis. Ages of mice reported are on day 1 of operant testing.

### *Behavioral testing*

*Operant acquisition and extinction.* We used the same testing protocol as was used in our previous study<sup>83</sup>. Mice were food-restricted for the duration of the experiments (2 h of unrestricted feeding after testing each day). We performed daily testing during the light phase at the same time each day. Testing was performed using modular operant conditioning chambers (MED Associates, ENV-307W). These chambers had five available nose-poke apertures on one wall, a food delivery magazine on the opposite wall, and a house light that illuminated the chamber. To provide operant cues, stimulus lights inside the nose-poke apertures were individually controlled. We covered three nose-poke apertures (holes 1, 3, and 5), leaving two open for testing (**Fig. 1b**). There were three phases of operant testing: magazine training (MAG), operant acquisition (ACQ), and operant extinction (EXT). On the day prior to operant testing, mice were habituated to the chambers with the house light on and no reward delivery for 30 minutes. Next, we trained mice (two 10 min sessions on consecutive days) to receive a food reward (Cohort 1: 20 mg dustless precision pellets, Bio-Serv; Cohort 2: 14 mg dustless precision pellets, Bio-Serv) each time they nose-poked into the food magazine [MAG training]. During operant acquisition (15 min sessions), we illuminated one aperture at a time (the "cued" aperture), and mice received a reward only after a nose-poke into the cued aperture. When a mouse did a nose-poke into the non-illuminated aperture, this was considered a "non-cued" response. For each mouse, the light cue was presented in the same aperture during every trial across days, but cue location (left or right aperture) was randomized across mice. A "trial" is defined as the receipt of a food reward following one or more cued responses, and each trial was self-initiated by the mice, so the number of trials per session was open-ended. To encourage mice to self-initiate more trials, we "primed" mice on by placing one food pellet inside both the cued and non-cued apertures. Priming was performed on the first two days of acquisition and on subsequent sessions following a session with <5 trials. During acquisition, pellet rewards were delivered on

a fixed FR1 schedule for the first 10 rewards, then on a variable VR2 schedule for the following rewards. Mice needed to complete >15 trials within a session, and >75% cued response rate [cued responses/(cued + non-cued responses)], over five consecutive days to meet criteria for successful acquisition. Immediately following acquisition, mice underwent three days of extinction training, where the light cue was presented without reward. For analysis of extinction data, we normalized cued and non-cued responses to the group average of the last five days of acquisition. After all mice in a cage completed their last day of operant extinction, mice were returned to normal *ad libitum* food access for the remainder of the study.

*Ube3a<sup>m-p+</sup> behavior battery.* Mice were given 2-4 weeks to recover from operant behavioral testing and food restriction prior to starting the behavior battery. The behavior battery consisted of accelerating rotarod (five consecutive days), open field (one day), marble burying (one day), and nest building (five consecutive days). There was two to three days between the end of one test and the start of the next test. Weight was recorded on the first day of the rotarod prior to behavior testing.

*Rotarod.* Mice were positioned on a rotating rod that increased speed from 4 to 40 rpm over a five-minute duration with an acceleration of 7.2 rpm<sup>2</sup> (Ugo Basile model #47600). Testing concluded when the animal dropped from the apparatus, completed three successive wrapping rotations, or when the five-minute time limit was reached. Daily testing consisted of two trials separated by a one-hour inter-trial period, with results averaged for each day. The protocol spanned five consecutive days.

*Open field test.* Mice were introduced into a 42 cm square arena (AccuScan Instruments, Inc., Columbus, OH) and permitted to explore freely during a single 10-minute session. The center zone was 21 cm x 21 cm. Open field activity was monitored using an automatic tracking system (Omnitech Electronics, Inc. SuperFlex Open Field System) employing photocell emitters and receptors that create an x-y grid of infrared beams. Total locomotor distance and duration in the center were measured via infrared beam interruption data. Testing occurred over a single day.

*Marble burying.* Mice were placed individually in a 16 × 8 in cage with ~4 cm of bedding (Bed-o'Cobs 1/4" bedding) and 20 black glass marbles arranged in a 5 × 4 array for a single 30 min trial. A marble was considered buried if it was >50% covered with bedding at the end of the trial. Testing occurred over a single day. Immediately after marble burying, mice were singly housed for nest building habituation.

*Nest building.* Immediately after marble burying, mice were then singly-housed and habituated to new nesting material (Bio-Rad 7.5 × 10 cm extra thick block filter paper; 11 ± 1 g) for 3 days prior to testing. During testing, new nesting material was introduced on day 1, and unused material was weighed daily across five days.

#### *Multidimensional analysis of mouse behavior*

We used the MATLAB graphical user interface PUMBAA ([github.com/sidorovlab/PUMBAA](https://github.com/sidorovlab/PUMBAA)) to perform multidimensional analysis of behavior. Multidimensional analysis using PUMBAA consists of a series of steps: data selection, standardization, principal component analysis (PCA), k-means clustering, and validation, as described in our prior work. Here, we performed multidimensional analysis twice—once with the behavior

battery measures alone (**Fig. 4a**), and a second time with measures from operant behavior also included (**Fig. 4g**). On the first run, our goal was to determine how accurately PCA would predict genotype if only given data from the *Ube3a*<sup>m-/p+</sup> behavior battery. On the second run, we included operant behaviors to determine if these additional behavioral points improved clustering accuracy or the distance between group centroids in PC space. *Data selection*: For the *Ube3a*<sup>m-/p+</sup> behavior battery, seven total measures were included: weight, rotarod day 1, rotarod day 5, open field distance traveled, open field center time, marbles buried, and nest building at day 5 (**Fig. 4a**). For the *Ube3a*<sup>m-/p+</sup> behavior battery plus operant extinction (**Fig. 4g**), 10 total measures were included: *Ube3a*<sup>m-/p+</sup> behavior battery (seven measures as above), days to acquisition, accuracy at criteria, and extinction day 1. *Standardization*: All measures were standardized using a z-score ( $z = (\text{data point} - \text{group means}) / \text{standard deviation}$ ) to account for different units across tests. In prior work, we standardized male and female data separately for tests where a sex difference was observed, using the same method here. *Principal component analysis*: We performed PCA using the `pca()` function in MATLAB and calculated the amount of variance explained by each PC using a Scree plot and the loading distribution of principal components using the coefficient outputs from PCA (**Fig. 4b-c**). *k-means clustering*: k-means clustering was performed in principal component space using the `kmeans()` function in MATLAB with  $k = 2$  clusters. *Validation*: We compared the actual genotypes of the mice to their assigned cluster and calculated the percentage of correct assignments. Centroids of genotype-based groups were calculated by taking the mean of all x and y positions for each mouse in principal component space. Here, centroids were calculated based on the actual genotype, rather than the k-means cluster. The distance between WT and *Ube3a*<sup>m-/p+</sup> centroids was defined as the absolute value of the square root of  $(x_1 - x_2)^2 + (y_1 - y_2)^2$ , where x and y represent the coordinates of centroids defined by genotype in 2PC space.

## Statistics

Statistical analysis was performed using GraphPad Prism 9 and MATLAB R2023a (Mathworks). We used unpaired t-tests for the majority of comparisons between WT and *Ube3a*<sup>m-/p+</sup> mice when assessing single behavioral timepoints (**Figs. 1c, 1e-h**). When assessing behavioral outcome measures that occurred over days, we used a two-way RM ANOVA for genotype and day (**Figs. 1i, 1k, 1m, 2c, 2f, S1a, S1c, S1e**). When assessing the effects of cohort or sex by genotype, we used a two-way ANOVA (**Figs. 2a-b, 2d-e, 3a, 3c, 3e-f**). When assessing the effects of sex by genotype by days, we used a three-way RM ANOVA (**Figs. 3b, 3d**). To determine sample sizes needed to achieve a statistically significant result, we performed power analyses of data presented in Figs. 1e, 1h, and 1i (days to acquisition criteria, accuracy at criteria, and E1 extinction). Here, we used MATLAB to conduct a bootstrap simulation using 10,000 iterations per each “n” and set significance to 95% ( $\alpha = 0.05$ ) and power to 80% ( $1 - \beta = 0.8$ ). For all figures, data are represented as mean  $\pm$  SEM, were two-tailed, and \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , and \*\*\*\* $p < 0.0001$ .

**Acknowledgments:** Work was supported by NIH R00EY028964 to MSS, NIH T32NS115656 to CDG, a grant from the Angelman Syndrome Foundation to MSS, and the District of Columbia Intellectual and Developmental

443 Disabilities Research Center (DC-IDRC) Award P50HD105328 by NICHD. We thank Dr. Li Wang (Children's  
444 National) for assistance with open field testing.

445

446 **Author contributions:** Conceptualization: MSS. Methodology: CDG, DDK, and MSS. Investigation: CDG,  
447 DDK, ZSH, and JKT. Formal Analysis: CDG, DDK, ZSH, and MSS. Writing – Original Draft: CDG and MSS.  
448 Writing – Review & Editing: CDG, DDK, ZSH, JKT, and MSS. Supervision: MSS. Funding Acquisition: MSS  
449 and CDG.

450

451 **Competing interests:** The authors declare no competing interests.

452

453 **Data availability:** All data generated or analyzed during this study are included in the published article (and its  
454 Supplementary Information files).

## FIGURE LEGENDS

**Figure 1. Operant acquisition and extinction phenotypes in *Ube3a*<sup>m-/p+</sup> mice.** (a) Experimental timeline for evaluating operant behavior and the Angelman behavior battery in the same animals. (b) Schematic of three stages of operant testing: magazine training (MAG), acquisition (ACQ), extinction (EXT). (c) No difference in magazine training between WT and *Ube3a*<sup>m-/p+</sup> mice. (d) Learning curves during acquisition. (e) *Ube3a*<sup>m-/p+</sup> mice took more days to reach acquisition criteria. (f) *Ube3a*<sup>m-/p+</sup> mice had fewer cued responses at criteria. (g) *Ube3a*<sup>m-/p+</sup> mice had fewer non-cued responses at criteria. (h) *Ube3a*<sup>m-/p+</sup> mice were more accurate at criteria. (i) *Ube3a*<sup>m-/p+</sup> mice made fewer cued responses on extinction day 1. Bracket indicates day × genotype interaction; E1 asterisks represent *post hoc* test. (j) Cued responses during extinction in individual mice. (k) No genotype difference in non-cued responses during extinction. (l) Non-cued responses during extinction in individual mice. Four instances of zero non-cued responses are not plotted here because data are shown on a log scale, but were included in group analysis in K. (m) No genotype difference in accuracy during extinction. (n) Accuracy during extinction in individual mice. Data represent mean ± SEM; \*\*\*\**p* < 0.0001; \*\**p* < 0.01; black: WT (*n* = 30 mice), red: *Ube3a*<sup>m-/p+</sup> (AS; *n* = 26).

**Figure 2. Evaluation of operant phenotypes in *Ube3a*<sup>m-/p+</sup> mice by cohort and sex.** (a) Acquisition days to criteria and (b) accuracy at criteria by cohort. (c) Left: cued responses during extinction in Cohort 1. Right: cued responses during extinction in Cohort 2. (d) Days to criteria and (e) accuracy at criteria by sex. (f) Left: cued responses during extinction in males. Right: cued responses during extinction in females. Data represent mean ± SEM; \*\**p* < 0.01; \**p* < 0.05; black: WT, red: *Ube3a*<sup>m-/p+</sup> (AS). Brackets in C and F represent day × genotype interaction, other asterisks represent *post hoc* tests. For panels a-c, [Cohort 1: WT (*n* = 13 mice), AS (*n* = 11 mice); Cohort 2: WT (*n* = 17 mice), AS (*n* = 15 mice)]. For panels d-f, [Males: WT (*n* = 16 mice), AS (*n* = 12 mice); Females: WT (*n* = 14 mice), AS (*n* = 14 mice)]. When there is a main effect of genotype, it is indicated with text under the graph.

**Figure 3. *Ube3a*<sup>m-/p+</sup> mice show behavioral impairments on a widely used behavioral battery.** (a) *Ube3a*<sup>m-/p+</sup> mice were heavier than WT littermates. Top bracket indicates main effect of sex; other asterisks represent *post hoc* tests. (b) *Ube3a*<sup>m-/p+</sup> mice had a decreased latency to fall in rotarod performance across days. Asterisks represent main effect of genotype. (c) *Ube3a*<sup>m-/p+</sup> mice buried fewer marbles during the marble burying task. Asterisks indicate *post hoc* tests. (d) *Ube3a*<sup>m-/p+</sup> mice used less nesting material across days. Asterisks indicate main effect of genotype. (e) No difference in open field distance between genotypes. (e) No difference in open field center time between genotypes. Data represent mean ± SEM; \*\*\*\**p* < 0.0001; \*\*\**p* < 0.001; \*\**p* < 0.01; \**p* < 0.05; black: WT, red: *Ube3a*<sup>m-/p+</sup> (AS).

**Figure 4. Operant behaviors are independent of AS battery behaviors and modestly enhance genotype discrimination.** (a) Multidimensional analysis using data from the *Ube3a*<sup>m-/p+</sup> behavior battery demonstrates

accurate clustering of animals using two principal components. (b) Loading table and (c) loading plot for PCA in panel a. (d) Days to acquisition does not correlate with PC1. (e) Accuracy at criteria does not correlate with PC1. (f) The extinction burst response on day 1 does not correlate with PC1. Lines show linear fits with 95% confidence intervals. (g) Adding operant extinction behaviors to PCA analysis increases centroid distance between WT and *Ube3a<sup>m-/p+</sup>* clusters. Centroids and centroid distance are calculated based on actual genotypes, not k-means clusters.

**Supplementary Figure S1. Raw data during extinction.** Data are associated with extinction reported as normalized to baseline in Figure 1. Black: WT, red: *Ube3a<sup>m-/p+</sup>* (AS). (a) Cued responses during acquisition and extinction. *Ube3a<sup>m-/p+</sup>* mice had fewer cued responses during the five days at criteria in the acquisition phase, represented as an average of five days in “ACQ” (as reported in Fig. 1g). During the extinction phase, analysis of raw responses revealed a main effect of genotype ( $F_{(1,54)} = 25.00$ ,  $p < 0.0001$ ) and a day  $\times$  genotype interaction ( $F_{(2,108)} = 9.292$ ,  $p = 0.0002$ ). Bracket represents interaction and asterisks represent *post hoc* tests (E1  $p < 0.0001$ , E2  $p = 0.1781$ , E3  $p = 0.0117$ ). (b) Cued responses during extinction in individual mice. (c) Non-cued responses during acquisition and extinction. *Ube3a<sup>m-/p+</sup>* had fewer non-cued responses during the five days at criteria in the acquisition phase, represented as an average of five days in “ACQ” (as reported in Fig. 1g). During the extinction phase, analysis of raw responses revealed a main effect of genotype ( $F_{(1,58)} = 20.66$ ,  $p < 0.0001$ ) but no day  $\times$  genotype interaction ( $F_{(2,116)} = 0.3888$ ,  $p = 0.6788$ ). Asterisks represent *post hoc* tests (E1  $p = 0.0021$ , E2  $p = 0.0091$ , E3  $p = 0.0316$ ). (d) Non-cued responses during extinction in individual mice. (e) Accuracy [cued responses / (cued responses + non-cued responses)] during acquisition and extinction. *Ube3a<sup>m-/p+</sup>* mice had higher accuracy during the five days at criteria in the acquisition phase, represented as an average of five days in “ACQ” (as reported in Fig. 1h). During the extinction phase, analysis of accuracy revealed a main effect of genotype ( $F_{(1,54)} = 9.046$ ,  $p = 0.0040$ ) but no day  $\times$  genotype interaction ( $F_{(2,108)} = 1.128$ ,  $p = 0.3276$ ). Asterisks represent *post hoc* tests (E1  $p = 0.0356$ , E2  $p = 0.0247$ , E3  $p = 0.1677$ ). (f) Accuracy during extinction in individual mice. Data represent mean  $\pm$  SEM; \*\*\*\* $p < 0.0001$ ; \*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ .

**Supplementary File S1. Statistics for all figures.** Data represent mean  $\pm$  SEM; \*\*\*\* $p < 0.0001$ ; \*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ .

## REFERENCES

- 1 Bird, L. M. Angelman syndrome: review of clinical and molecular aspects. *The application of clinical genetics*, 93-104 (2014).
- 2 Thibert, R. L., Larson, A. M., Hsieh, D. T., Raby, A. R. & Thiele, E. A. Neurologic manifestations of Angelman syndrome. *Pediatr Neurol* **48**, 271-279 (2013).  
<https://doi.org/10.1016/j.pediatrneurol.2012.09.015>
- 3 Sadhwani, A. *et al.* Developmental skills of individuals with Angelman syndrome assessed using the Bayley-III. *Journal of autism and developmental disorders* **53**, 720-737 (2023).
- 4 Gwaltney, A. *et al.* Adaptive skills of individuals with angelman syndrome assessed using the Vineland Adaptive Behavior Scales. *Journal of Autism and Developmental Disorders* **54**, 3863-3887 (2024).
- 5 Keute, M. *et al.* Angelman syndrome genotypes manifest varying degrees of clinical severity and developmental impairment. *Molecular Psychiatry* **26**, 3625-3633 (2021).
- 6 Peters, S. U. *et al.* Cognitive and adaptive behavior profiles of children with Angelman syndrome. *American Journal of Medical Genetics Part A* **128**, 110-113 (2004).
- 7 Willgoss, T. *et al.* Measuring what matters to individuals with Angelman syndrome and their families: Development of a patient-centered disease concept model. *Child Psychiatry & Human Development* **52**, 654-668 (2021). <https://doi.org/https://doi.org/10.1007/s10578-020-01051-z>
- 8 Kishino, T., Lalande, M. & Wagstaff, J. UBE3A/E6-AP mutations cause Angelman syndrome. *Nat Genet* **15**, 70-73 (1997). <https://doi.org/10.1038/ng0197-70>
- 9 Clayton-Smith, J. & Laan, L. Angelman syndrome: a review of the clinical and genetic aspects. *J Med Genet* **40**, 87-95 (2003). <https://doi.org/10.1136/jmg.40.2.87>
- 10 Yamasaki, K. *et al.* Neurons but not glial cells show reciprocal imprinting of sense and antisense transcripts of Ube3a. *Human molecular genetics* **12**, 837-847 (2003).
- 11 LaSalle, J. M., Reiter, L. T. & Chamberlain, S. J. Epigenetic regulation of UBE3A and roles in human neurodevelopmental disorders. *Epigenomics* **7**, 1213-1228 (2015).
- 12 Hsiao, J. S. *et al.* A bipartite boundary element restricts UBE3A imprinting to mature neurons. *Proceedings of the National Academy of Sciences* **116**, 2181-2186 (2019).
- 13 Elgersma, Y. & Sonzogni, M. UBE3A reinstatement as a disease-modifying therapy for Angelman syndrome. *Developmental Medicine & Child Neurology* **63**, 802-807 (2021).
- 14 Markati, T., Duis, J. & Servais, L. Therapies in preclinical and clinical development for Angelman syndrome. *Expert Opinion on Investigational Drugs* **30**, 709-720 (2021).
- 15 Hipp, J. F. *et al.* The UBE3A-ATS antisense oligonucleotide rugonersen in children with Angelman syndrome: a phase 1 trial. *Nat Med* (2025). <https://doi.org/10.1038/s41591-025-03784-7>
- 16 Huang, H.-S. *et al.* Topoisomerase inhibitors unsilence the dormant allele of Ube3a in neurons. *Nature* **481**, 185-189 (2012).
- 17 Lee, H. M. *et al.* Characterization and structure-activity relationships of indenoisoquinoline-derived topoisomerase I inhibitors in unsilencing the dormant Ube3a gene associated with Angelman syndrome. *Mol Autism* **9**, 45 (2018). <https://doi.org/10.1186/s13229-018-0228-2>
- 18 Vihma, H. *et al.* Ube3a unsilencer for the potential treatment of Angelman syndrome. *Nature Communications* **15**, 5558 (2024).
- 19 Meng, L. *et al.* Truncation of Ube3a-ATS unsilences paternal Ube3a and ameliorates behavioral defects in the Angelman syndrome mouse model. *PLoS genetics* **9**, e1004039 (2013).
- 20 Meng, L. *et al.* Towards a therapy for Angelman syndrome by targeting a long non-coding RNA. *Nature* **518**, 409-412 (2015).
- 21 Milazzo, C. *et al.* Antisense oligonucleotide treatment rescues UBE3A expression and multiple phenotypes of an Angelman syndrome mouse model. *JCI insight* **6** (2021).
- 22 Dindot, S. V. *et al.* An ASO therapy for Angelman syndrome that targets an evolutionarily conserved region at the start of the UBE3A-AS transcript. *Science Translational Medicine* **15**, eabf4077 (2023).
- 23 Clarke, M. T. *et al.* Prenatal delivery of a therapeutic antisense oligonucleotide achieves broad biodistribution in the brain and ameliorates Angelman syndrome phenotype in mice. *Molecular Therapy* **32**, 935-951 (2024).
- 24 Schmid, R. S. *et al.* CRISPR/Cas9 directed to the Ube3a antisense transcript improves Angelman syndrome phenotype in mice. *The Journal of Clinical Investigation* **131** (2021).

Wolter, J. M. *et al.* Cas9 gene therapy for Angelman syndrome traps Ube3a-ATS long non-coding RNA. *Nature* **587**, 281-284 (2020).

Li, J. *et al.* A high-fidelity RNA-targeting Cas13 restores paternal Ube3a expression and improves motor functions in Angelman syndrome mice. *Molecular Therapy* **31**, 2286-2295 (2023).

Daily, J. L. *et al.* Adeno-associated virus-mediated rescue of the cognitive defects in a mouse model for Angelman syndrome. *PLoS One* **6**, e27221 (2011). <https://doi.org/10.1371/journal.pone.0027221>

Judson, M. C. *et al.* Dual-isoform hUBE3A gene transfer improves behavioral and seizure outcomes in Angelman syndrome model mice. *JCI insight* **6**, e144712 (2021).

Sell, G. L. *et al.* Deleting a UBE3A substrate rescues impaired hippocampal physiology and learning in Angelman syndrome mice. *Sci Rep* **11**, 19414 (2021). <https://doi.org/10.1038/s41598-021-97898-w>

Sun, A. X. *et al.* Potassium channel dysfunction in human neuronal models of Angelman syndrome. *Science* **366**, 1486-1492 (2019). <https://doi.org/10.1126/science.aav5386>

Jamal, I. *et al.* Rescue of altered HDAC activity recovers behavioural abnormalities in a mouse model of Angelman syndrome. *Neurobiol Dis* **105**, 99-108 (2017). <https://doi.org/10.1016/j.nbd.2017.05.010>

Hu, J. H. *et al.* Activity-dependent degradation of Kv4.2 contributes to synaptic plasticity and behavior in Angelman syndrome model mice. *Cell Rep* **44**, 115583 (2025). <https://doi.org/10.1016/j.celrep.2025.115583>

Sun, J. *et al.* mTORC1-S6K1 inhibition or mTORC2 activation improves hippocampal synaptic plasticity and learning in Angelman syndrome mice. *Cell Mol Life Sci* **73**, 4303-4314 (2016). <https://doi.org/10.1007/s00018-016-2269-z>

Bi, X., Sun, J., Ji, A. X. & Baudry, M. Potential therapeutic approaches for Angelman syndrome. *Expert Opin Ther Targets* **20**, 601-613 (2016). <https://doi.org/10.1517/14728222.2016.1115837>

Cao, C. *et al.* Impairment of TrkB-PSD-95 signaling in Angelman syndrome. *PLoS Biol* **11**, e1001478 (2013). <https://doi.org/10.1371/journal.pbio.1001478>

Egawa, K. *et al.* Decreased tonic inhibition in cerebellar granule cells causes motor dysfunction in a mouse model of Angelman syndrome. *Sci Transl Med* **4**, 163ra157 (2012). <https://doi.org/10.1126/scitranslmed.3004655>

Pandya, N. J. *et al.* Secreted retrovirus-like GAG-domain-containing protein PEG10 is regulated by UBE3A and is involved in Angelman syndrome pathophysiology. *Cell Rep Med* **2**, 100360 (2021). <https://doi.org/10.1016/j.xcrm.2021.100360>

Cruz, E. *et al.* CIM6P/IGF-2 receptor ligands reverse deficits in Angelman syndrome model mice. *Autism Research* **14**, 29-45 (2021).

Jiang, Y.-h. *et al.* Mutation of the Angelman ubiquitin ligase in mice causes increased cytoplasmic p53 and deficits of contextual learning and long-term potentiation. *Neuron* **21**, 799-811 (1998).

Rotaru, D. C., Mientjes, E. J. & Elgersma, Y. Angelman syndrome: from mouse models to therapy. *Neuroscience* **445**, 172-189 (2020).

Bruinsma, C. F. *et al.* Dissociation of locomotor and cerebellar deficits in a murine Angelman syndrome model. *J Clin Invest* **125**, 4305-4315 (2015). <https://doi.org/10.1172/JCI83541>

Heck, D. H., Zhao, Y., Roy, S., LeDoux, M. S. & Reiter, L. T. Analysis of cerebellar function in Ube3a-deficient mice reveals novel genotype-specific behaviors. *Hum Mol Genet* **17**, 2181-2189 (2008). <https://doi.org/10.1093/hmg/ddn117>

Allensworth, M., Saha, A., Reiter, L. T. & Heck, D. H. Normal social seeking behavior, hypoactivity and reduced exploratory range in a mouse model of Angelman syndrome. *BMC Genet* **12**, 7 (2011). <https://doi.org/10.1186/1471-2156-12-7>

Huang, H.-S. *et al.* Behavioral deficits in an Angelman syndrome model: effects of genetic background and age. *Behavioural brain research* **243**, 79-90 (2013).

Judson, M. C. *et al.* GABAergic Neuron-Specific Loss of Ube3a Causes Angelman Syndrome-Like EEG Abnormalities and Enhances Seizure Susceptibility. *Neuron* **90**, 56-69 (2016). <https://doi.org/10.1016/j.neuron.2016.02.040>

Gu, B. *et al.* Cannabidiol attenuates seizures and EEG abnormalities in Angelman syndrome model mice. *J Clin Invest* **129**, 5462-5467 (2019). <https://doi.org/10.1172/JCI130419>

Mandel-Brehm, C., Salogiannis, J., Dhamne, S. C., Rotenberg, A. & Greenberg, M. E. Seizure-like activity in a juvenile Angelman syndrome mouse model is attenuated by reducing Arc expression. *Proc Natl Acad Sci U S A* **112**, 5129-5134 (2015). <https://doi.org/10.1073/pnas.1504809112>

Shi, S. Q., Bichell, T. J., Ihrle, R. A. & Johnson, C. H. Ube3a imprinting impairs circadian robustness in Angelman syndrome models. *Curr Biol* **25**, 537-545 (2015). <https://doi.org/10.1016/j.cub.2014.12.047>

Ehlen, J. C. *et al.* Maternal Ube3a Loss Disrupts Sleep Homeostasis But Leaves Circadian Rhythmicity Largely Intact. *J Neurosci* **35**, 13587-13598 (2015). <https://doi.org/10.1523/JNEUROSCI.2194-15.2015>

Copping, N. A. & Silverman, J. L. Abnormal electrophysiological phenotypes and sleep deficits in a mouse model of Angelman Syndrome. *Mol Autism* **12**, 9 (2021). <https://doi.org/10.1186/s13229-021-00416-y>

Colas, D., Wagstaff, J., Fort, P., Salvert, D. & Sarda, N. Sleep disturbances in Ube3a maternal-deficient mice modeling Angelman syndrome. *Neurobiol Dis* **20**, 471-478 (2005). <https://doi.org/10.1016/j.nbd.2005.04.003>

Sidorov, M. S. *et al.* Delta rhythmicity is a reliable EEG biomarker in Angelman syndrome: a parallel mouse and human analysis. *J Neurodev Disord* **9**, 17 (2017). <https://doi.org/10.1186/s11689-017-9195-8>

Spencer, E. R. *et al.* Longitudinal EEG model detects antisense oligonucleotide treatment effect and increased UBE3A in Angelman syndrome. *Brain Commun* **4**, fcac106 (2022). <https://doi.org/10.1093/braincomms/fcac106>

Born, H. A. *et al.* Strain-dependence of the Angelman Syndrome phenotypes in Ube3a maternal deficiency mice. *Scientific reports* **7**, 8451 (2017).

Montgomery, D. P., Shide, J. J., Didouchevski, A. M., Dickinson, A. H. & Sidorov, M. S. Periodic and aperiodic contributions to EEG delta power are translatable and complementary Angelman syndrome biomarkers. *bioRxiv* (2025). <https://doi.org/https://doi.org/10.1101/2025.09.08.674941>

Perrino, P. A., Chamberlain, S. J., Eigsti, I. M. & Fitch, R. H. Communication-related assessments in an Angelman syndrome mouse model. *Brain and Behavior* **11**, e01937 (2021).

Guoynes, C. D., Pavalko, G. & Sidorov, M. S. Courtship and distress ultrasonic vocalizations are altered in a mouse model of Angelman syndrome. *J Neurodev Disord* **17**, 59 (2025). <https://doi.org/10.1186/s11689-025-09648-y>

Sonzogni, M. *et al.* A behavioral test battery for mouse models of Angelman syndrome: a powerful tool for testing drugs and novel Ube3a mutants. *Molecular autism* **9**, 1-19 (2018).

Qu, S. *et al.* Dietary Intake of Octanoic Acid Restores UBE3A Expression and Improves the Behavioral Phenotypes in a Mouse Model of Angelman Syndrome. *The FASEB Journal* **39**, e70559 (2025).

Silva-Santos, S. *et al.* Ube3a reinstatement identifies distinct developmental windows in a murine Angelman syndrome model. *J Clin Invest* **125**, 2069-2076 (2015). <https://doi.org/10.1172/JCI80554>

Sonzogni, M., Zhai, P., Mientjes, E. J., van Woerden, G. M. & Elgersma, Y. Assessing the requirements of prenatal UBE3A expression for rescue of behavioral phenotypes in a mouse model for Angelman syndrome. *Mol Autism* **11**, 70 (2020). <https://doi.org/10.1186/s13229-020-00376-9>

Sonzogni, M. *et al.* Delayed loss of UBE3A reduces the expression of Angelman syndrome-associated phenotypes. *Mol Autism* **10**, 23 (2019). <https://doi.org/10.1186/s13229-019-0277-1>

Dodge, A. *et al.* Generation of a novel rat model of Angelman syndrome with a complete Ube3a gene deletion. *Autism Research* **13**, 397-409 (2020).

Berg, E. *et al.* Translational outcomes in a full gene deletion of ubiquitin protein ligase E3A rat model of Angelman syndrome. *Translational psychiatry* **10**, 39 (2020).

Myers, L. S. *et al.* A preclinical pig model of Angelman syndrome mirrors the early developmental trajectory of the human condition. *Proceedings of the National Academy of Sciences* **122**, e2505152122 (2025).

Moreira-de-Sá, A. *et al.* Adenosine A2A receptors format long-term depression and memory strategies in a mouse model of Angelman syndrome. *Neurobiology of Disease* **146**, 105137 (2020).

Schultz, M. N. & Crawley, J. N. Evaluation of a TrkB agonist on spatial and motor learning in the Ube3a mouse model of Angelman syndrome. *Learning & Memory* **27**, 346-354 (2020).

Guzzetti, S. *et al.* Taurine administration recovers motor and learning deficits in an Angelman syndrome mouse model. *International journal of molecular sciences* **19**, 1088 (2018).

Berg, E. L. *et al.* Insulin-like growth factor-2 does not improve behavioral deficits in mouse and rat models of Angelman syndrome. *Molecular Autism* **12**, 1-16 (2021).

Liu, Y. *et al.* Enhancement of synaptic plasticity and reversal of impairments in motor and cognitive functions in a mouse model of Angelman Syndrome by a small neurogenic molecule, NSI-189. *Neuropharmacology* **144**, 337-344 (2019).

687 71 Kaphzan, H. *et al.* Reversal of impaired hippocampal long-term potentiation and contextual fear  
688 memory deficits in Angelman syndrome model mice by ErbB inhibitors. *Biological psychiatry* **72**, 182-  
689 190 (2012).

690 72 Vorhees, C. V. & Williams, M. T. Morris water maze: procedures for assessing spatial and related forms  
691 of learning and memory. *Nature protocols* **1**, 848-858 (2006).

692 73 Phillips, R. & LeDoux, J. E. Differential contribution of amygdala and hippocampus to cued and  
693 contextual fear conditioning. *Behavioral neuroscience* **106**, 274 (1992).

694 74 Broadbent, N. J., Gaskin, S., Squire, L. R. & Clark, R. E. Object recognition memory and the rodent  
695 hippocampus. *Learning & memory* **17**, 5-11 (2010).

696 75 Koike, H. *et al.* Chemogenetic inactivation of dorsal anterior cingulate cortex neurons disrupts  
697 attentional behavior in mouse. *Neuropsychopharmacology* **41**, 1014-1023 (2016).

698 76 Norman, K. J. *et al.* Chemogenetic suppression of anterior cingulate cortical neurons projecting to the  
699 visual cortex disrupts attentional behavior in mice. *Neuropsychopharmacology Reports* **41**, 207-214  
700 (2021).

701 77 Green, J. T. & Bouton, M. E. New functions of the rodent prelimbic and infralimbic cortex in instrumental  
702 behavior. *Neurobiology of learning and memory* **185**, 107533 (2021).

703 78 Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* **24**,  
704 167-202 (2001). <https://doi.org/10.1146/annurev.neuro.24.1.167>

705 79 Le Merre, P., Ahrlund-Richter, S. & Carlen, M. The mouse prefrontal cortex: Unity in diversity. *Neuron*  
706 **109**, 1925-1944 (2021). <https://doi.org/10.1016/j.neuron.2021.03.035>

707 80 Rotaru, D. C., van Woerden, G. M., Wallaard, I. & Elgersma, Y. Adult Ube3a Gene Reinstatement  
708 Restores the Electrophysiological Deficits of Prefrontal Cortex Layer 5 Neurons in a Mouse Model of  
709 Angelman Syndrome. *J Neurosci* **38**, 8011-8030 (2018). [https://doi.org/10.1523/JNEUROSCI.0083-](https://doi.org/10.1523/JNEUROSCI.0083-18.2018)  
710 [18.2018](https://doi.org/10.1523/JNEUROSCI.0083-18.2018)

711 81 Sidorov, M. S. *et al.* Visual Sequences Drive Experience-Dependent Plasticity in Mouse Anterior  
712 Cingulate Cortex. *Cell Rep* **32**, 108152 (2020). <https://doi.org/10.1016/j.celrep.2020.108152>

713 82 Negron-Moreno, P. N., Diep, D. T., Guoyne, C. D. & Sidorov, M. S. Dissociating motor impairment  
714 from five-choice serial reaction time task performance in a mouse model of Angelman syndrome. *Front*  
715 *Behav Neurosci* **16**, 968159 (2022). <https://doi.org/10.3389/fnbeh.2022.968159>

716 83 Sidorov, M. S. *et al.* Enhanced operant extinction and prefrontal excitability in a mouse model of  
717 Angelman syndrome. *Journal of Neuroscience* **38**, 2671-2682 (2018).

718 84 Tanas, J. K. *et al.* Multidimensional analysis of behavior predicts genotype with high accuracy in a  
719 mouse model of Angelman syndrome. *Translational psychiatry* **12**, 426 (2022).

720 85 Koyavski, L. *et al.* Sex-dependent sensory phenotypes and related transcriptomic expression profiles  
721 are differentially affected by Angelman syndrome. *Molecular Neurobiology* **56**, 5998-6016 (2019).

722 86 Cosgrove, J. A., Kelly, L. K., Kiffmeyer, E. A. & Kloth, A. D. Sex-dependent influence of postweaning  
723 environmental enrichment in Angelman syndrome model mice. *Brain and Behavior* **12**, e2468 (2022).

724 87 Jamal, I. *et al.* Environmental enrichment improves behavioral abnormalities in a mouse model of  
725 Angelman syndrome. *Molecular neurobiology* **54**, 5319-5326 (2017).

726 88 Finger, E. C., Mitchell, D. G., Jones, M. & Blair, R. J. R. Dissociable roles of medial orbitofrontal cortex  
727 in human operant extinction learning. *Neuroimage* **43**, 748-755 (2008).

728 89 Alderman, N. Central executive deficit and response to operant conditioning methods.  
729 *Neuropsychological Rehabilitation* **6**, 161-186 (1996).

730 90 Krueger, D. D., Osterweil, E. K., Chen, S. P., Tye, L. D. & Bear, M. F. Cognitive dysfunction and  
731 prefrontal synaptic abnormalities in a mouse model of fragile X syndrome. *Proceedings of the National*  
732 *Academy of Sciences* **108**, 2587-2592 (2011).

733 91 Guoyne, C. D., Pavalko, G. & Sidorov, M. S. Courtship and distress ultrasonic vocalizations are  
734 disrupted in a mouse model of Angelman syndrome. *Research Square*, rs. 3. rs-5953744 (2025).

735 92 Bolivar, V. J., Caldarone, B. J., Reilly, A. A. & Flaherty, L. Habituation of activity in an open field: a  
736 survey of inbred strains and F1 hybrids. *Behavior genetics* **30**, 285-293 (2000).

737 93 Rudeck, J., Vogl, S., Banneke, S., Schönfelder, G. & Lewejohann, L. Repeatability analysis improves  
738 the reliability of behavioral data. *PloS one* **15**, e0230900 (2020).

739 94 Kilkenny, C., Browne, W., Cuthill, I. C., Emerson, M. & Altman, D. G. Animal research: reporting in vivo  
740 experiments: the ARRIVE guidelines. *British journal of pharmacology* **160**, 1577 (2010).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigure1.pdf](#)
- [SupplementaryFile1.xlsx](#)
- [SupplementaryFile2.xlsx](#)