Pharmacology, Biochemistry and Behavior 178 (2019) 42–50

Contents lists available at ScienceDirect

Pharmacology, Biochemistry and Behavior

journal homepage: www.elsevier.com/locate/pharmbiochembeh

Dopamine transporter knockdown mice in the behavioral pattern monitor: A robust, reproducible model for mania-relevant behaviors

Molly A. Kwiatkowski[a], Gerhard Hellemann[b], Catherine A. Sugar[b,c], Zackary A. Cope[a], Arpi Minassian[a], William Perry[a], Mark A. Geyer[a,d], Jared W. Young[a,d,*]

[a] Department of Psychiatry, University of California San Diego, USA.
[b] Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, USA.
[c] Department of Biostatistics, University of California Los Angeles, USA.
[d] Research Service, VA San Diego Healthcare System, USA.

ARTICLE INFO

Keywords:
Reproducibility
Meta-analysis
Rodent
Hyperactivity
Exploration
Bipolar disorder

ABSTRACT

Efforts to replicate results from both basic and clinical models have highlighted problems with reproducibility in science. In psychiatry, reproducibility issues are compounded because the complex behavioral syndromes make many disorders challenging to model. We develop translatable tasks that quantitatively measure psychiatry-relevant behaviors across species. The behavioral pattern monitor (BPM) was designed to analyze exploratory behaviors, which are altered in patients with bipolar disorder (BD), especially during mania episodes. We have repeatedly assessed the behavioral effects of reduced dopamine transporter (DAT) expression in the BPM using a DAT knockdown (KD) mouse line (~10% normal expression). DAT KD mice exhibit a profile in the BPM consistent with acutely manic BD patients in the human version of the task—hyperactivity, increased exploratory behavior, and reduced spatial d (Perry et al., 2009). We collected data from multiple DAT KD BPM experiments in our laboratory to assess the reproducibility of behavioral outcomes across experiments. The four outcomes analyzed were: 1) transitions (amount of locomotor activity); 2) rearings (exploratory activity); 3) holepokes (exploratory activity); and 4) spatial d (geometrical pattern of locomotor activity). By comparing DAT KD mice to wildtype (WT) littermates in every experiment, we calculated effect sizes for each of the four outcomes and then calculated a mean effect size using a random effects model. DAT KD mice exhibited robust, reproducible changes in each of the four outcomes, including increased transitions, rearings, and holepokes, and reduced spatial d, vs. WT littermates. Our results demonstrate that the DAT KD mouse line in the BPM is a consistent, reproducible model of mania-relevant behaviors. More work must be done to assess reproducibility of behavioral outcomes across experiments in order to advance the field of psychiatry and develop more effective therapeutics for patients.

1. Introduction

The problem of reproducibility in science has recently been highlighted in editorials and surveys of scientists, with failures in large-scale efforts to replicate findings in the fields of cancer biology and psychology. In a survey of 1576 researchers conducted by *Nature* (2016), > 70% of researchers reported attempting and failing to reproduce another scientist's experiments (Baker, 2016). Pressure to publish and selective reporting of results were two factors that the majority of respondents (> 60%) identified as always or often contributing to reproducibility problems. Determining reproducibility rates in certain scientific fields is underway, and results so far underscore the need for improved reproducibility standards. For example, the Open Science Collaboration (OSC) attempted to replicate 98 studies published in three psychology journals (Open Science Collaboration, 2015). While 97% of the original 98 studies found significant effects, only 36% of the results were successfully replicated; furthermore, effect sizes were on average half those of the original studies. The OSC is now attempting to replicate widely cited studies in the cancer biology field. Out of five studies completed so far, two have been replicated, two were inconclusive due to technical issues, and one failed to replicate (Aird et al., 2017; Horrigan et al., 2017; Horrigan and Reproducibility Project: Cancer Biology, 2017; Kandela et al., 2017; Mantis et al., 2017). Confirmation of findings through replication is ongoing and will be refined over time, but overall, currently available data emphasize the need to evaluate and validate reproducibility in both clinical and

* Corresponding author at: Department of Psychiatry, UC San Diego School of Medicine, La Jolla, CA 92093-0804, USA.
E-mail address: jaredyoung@ucsd.edu (J.W. Young).

https://doi.org/10.1016/j.pbb.2017.12.007
Received 11 September 2017; Received in revised form 18 November 2017; Accepted 27 December 2017
Available online 28 December 2017
0091-3057/

basic models.

Most of the efforts addressing reproducibility have been aimed at assessing inter-laboratory reproducibility (i.e. conducting the same experiment across multiple different laboratories). Intra-laboratory reproducibility (i.e. repeating an experiment in the same laboratory multiple times) is also important to examine, particularly in the field of psychiatry where it has been problematic to generate reproducible basic models. These difficulties arise for two main reasons: 1) psychiatric disorders are comprised of diagnostic categories based on mainly subjective heterogeneous symptoms, and 2) the general knowledge of underlying etiology and pathophysiology of these disorders remains limited. Some intra-laboratory reproducibility studies have been conducted; e.g., isolation rearing-induced deficits in sensorimotor gating as measured by prepulse inhibition (PPI; Geyer et al., 1993), a paradigm commonly used to measure schizophrenia-relevant behavioral changes (Braff and Geyer, 1990; Geyer and Braff, 1987; Swerdlow et al., 2017), has been assessed. Largely reproducible isolation rearing-induced PPI deficits were reported, although only in 14 out of 18 cohorts (Cilia et al., 2005). Other laboratories have not reproduced isolation rearing-induced PPI deficits, however (Weiss et al., 1999). Methodological differences make it difficult to compare results across laboratories, and further complicate reproducibility issues in the field. Furthermore, many psychiatric disorders are complex and challenging to model, adding to reproducibility issues.

Bipolar disorder (BD) is a complex disease to model in rodents due to its genetic variability and heterogeneous symptoms (Seufuddin et al., 2013) that are often cyclical, including switching between extreme states (mania and depression). These alternating episodes of mania and depression in particular have been challenging to recreate in rodent models (Gould and Einat, 2007; Young and Dulcis, 2015). Modeling BD has focused on recreating specific mania-like symptoms in rodents, given that mania is the cardinal feature of BD, using pharmacological (Fries et al., 2015), environmental (Arent et al., 2015), and genetic methods (Prickaerts et al., 2006). We have reported that knockdown (KD) of the dopamine transporter (DAT) in mice (to 10% of normal levels) results in mania-like symptoms, including abnormal exploration (Perry et al., 2009; Young et al., 2010a) and increased risk-taking behavior (van Enkhuizen et al., 2015, 2014b; Young et al., 2011) seen in BD mania patients. The model was based on genetic linkage studies showing an association between DAT polymorphisms and BD (Greenwood et al., 2006, 2001) which may functionally reduce DAT expression, with confirmation of reduced DAT expression in postmortem frontal cortices of BD patients (Rao et al., 2012) and in unmedicated euthymic patients using PET (Anand et al., 2011). Hence, reduced DAT expression recreates several aspects of BD mania and has construct validity (Young et al., 2011).

In terms of reproducibility of this animal model of BD, we have repeatedly assessed the effect of reduced DAT expression on exploration in the behavioral pattern monitor (BPM). This exploratory chamber – originally developed for use in rats (Geyer et al., 1986) – is also available for human testing (Minassian et al., 2010), and has been used to show that BD mania patients exhibit abnormalities in three core exploratory domains: amount of motor activity, exploration of novel stimuli, and motor activity patterns (Henry et al., 2013; Minassian et al., 2011; Paulus and Geyer, 1993; Perry et al., 2009). These three aspects of exploratory behavior are similarly altered in DAT KD mice compared with respective controls in the BPM paradigm (Perry et al., 2009; Young et al., 2010a). Further validation is seen when BD mania patients treated with standard treatment (e.g. valproate, among others) exhibit diminished hyperactivity over time, while object interactions (hyperexploration) and motor activity pattern differences remain (Minassian et al., 2011). This effect is identical to that seen when DAT KD mice were treated chronically with valproate (van Enkhuizen et al., 2013), supporting the predictive validity of this model. Therefore, our previous research indicates that DAT KD mice in the BPM accurately model mania behavior of BD patients.

To our knowledge, we are the only laboratory conducting experiments in the BPM using the DAT KD mouse line, making our aggregate dataset collected across different cohorts unique. Given the cross-species relevance of the BPM and the reproducibility issues present in psychiatric research, we determined whether the mania-like behavioral profile of DAT KD mice in the BPM paradigm was reproducible across different experiments, as well as background strains, in our laboratory. Such reproducibility would increase the internal validity of using DAT KD mice as a model for mania-like behavior and contribute to current global efforts addressing issues of reproducibility in psychiatry research.

In the current analysis, we collated the results of studies performed in our laboratory (from 2004 to 2015) on DAT KD vs. WT mice using the BPM. Cohen's *d* effect sizes were calculated for four core dimensions of the mania-like behavioral profile. These effect sizes were compared using a meta-analysis across studies and between behavioral dimensions to assess the reproducibility of the DAT KD mouse line in the BPM as a model for BD mania.

2. Methods

2.1. Study identification

To examine the reproducibility of the DAT KD mouse model of mania in our BPM paradigm, we identified studies conducted in our laboratory between 2004 and 2015 using DAT KD vs. WT mice in the BPM (see Section 2.2 for task description). The DAT KD mouse line has been maintained on a 129/SvJ (129) and C57BL/6J (C57) genetic background over this period. Overall analyses included DAT KD and WT littermates from both genetic backgrounds; separate sub-analyses were conducted on each strain to examine inter-background differences in behavioral outcomes. We included both published and unpublished results in this analysis; findings that have been previously published have been cited and identified accordingly. Data from drug treatment groups in three published studies were also collected in order to assess effects on BPM outcomes in this model.

2.2. Behavioral pattern monitor (BPM)

Locomotor and exploratory behavior was analyzed in BPM chambers (San Diego Instruments, San Diego, CA) as described previously by our group (e.g. Risbrough et al., 2006; van Enkhuizen et al., 2013). Briefly, each Plexiglas chamber (30.5 × 61 × 38 cm) contains eight wall holes (1.25 cm diameter, 1.9 cm above floor) and three floor holes (Fig. 1). Each hole contains an infrared beam to detect holepoking. A grid of 12 × 24 infrared photobeams located 1 cm above the floor records mouse location every 0.1 s., allowing calculation of transitions (i.e. locomotor activity) from one of nine defined regions to another. A second set of 16 infrared photobeams located 2.5 cm above the floor keeps track of number of mouse rearings. Each session lasts 45–180 min, and mice are free to explore the BPM during this time. Each chamber is enclosed such that external light and noise is minimized, and sessions are performed with an internal white light (350 lx in the center, 92 lx in the four chamber corners).

The four BPM outcome measures assessed in the current meta-analysis were: 1) number of transitions (locomotor activity); 2) number of rearings (exploratory behavior); 3) number of holepokes (exploratory behavior); and 4) spatial d (locomotor pattern). Spatial d (ranging from 1 to 2) quantifies the dimensionality of the pattern of locomotor activity, where d = 1 describes a straight-line path and d = 2 describes small, circumscribed movements (Paulus and Geyer, 1991).

2.3. Effect size calculation

Sample sizes, means, and associated standard deviations for each BPM outcome were collected from previous analyses in order to

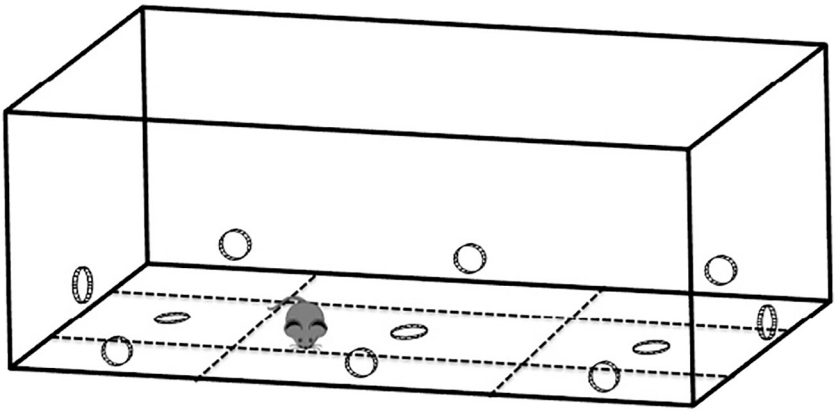

Fig. 1. Schematic of the mouse behavioral pattern monitor (BPM). The BPM is a Plexiglas chamber (30.5 × 61 × 38 cm) containing 8 wall holes (1.25 cm diameter, 1.9 cm above floor) and 3 floor holes. Each hole contains an infrared beam to detect holepoking. A 12 × 24 grid of infrared photobeams located 1 cm above the floor records mouse location every 0.1 s allowing for calculation of number of transitions from 1 of 9 defined regions to another (dashed lines illustrate the 9 regions). The mouse location using this grid is also used to quantify locomotor patterns such as spatial d. Another grid of 16 infrared photobeams located 2.5 cm above the floor tracks the number of rearings. Mice are free to explore the chamber during each test session, which can last 45–180 min. Each chamber is enclosed to minimize external light and noise, and sessions are performed with an internal white light (350 lx in the center, 92 lx in the 4 chamber corners).

calculate Cohen's *d* effect sizes in DAT KD vs. WT mice. Cohen's *d* was computed using the following equation:

$$d = \frac{\text{mean}_{DATKD} - \text{mean}_{WT}}{\sqrt{\frac{(N_{WT}-1)SD^2_{WT} + (N_{DATKD}-1)SD^2_{DATKD}}{N_{WT}+N_{DATKD}-2}}}$$

Positive *d* values reflected a larger mean in the DAT KD vs. WT group, while negative *d* values reflected the opposite. An effect size was calculated for each of the four BPM outcome measures per experiment (transitions, rearing, holepokes, and spatial d). In experiments where treatment dosing occurred, effect sizes for drug data were calculated by comparing DAT KD drug data to WT drug data (Figs. 2–5, purple data points). Comparing drug effect sizes to effect sizes calculated from DAT KD vs. WT vehicle treatment data enabled drug effect assessment on BPM outcomes in DAT KD mice.

2.4. Statistical methods for meta-analyses

Effect sizes obtained from individual experiments were used to calculate a mean effect size for each BPM outcome measure. First, each effect size was weighted by inverse variance to correct for bias due to differences in sample size; this process gave greater weight to effect sizes that were more reliably estimated. After weighting each effect size by inverse variance, a Q statistic was calculated using a meta-analysis macro in SPSS 24.0 (IBM Corp., Armonk, NY) to assess for homogeneity of effect sizes within each BPM outcome measure (meta-analysis macros for SPSS were retrieved June 7, 2017, from http://mason.gmu.edu/~dwilsonb/ma.html). To aid interpretation of the degree of heterogeneity, an I^2 statistic was calculated from the Q statistic and associated df: $I^2 = \frac{Q-df}{Q} * 100$. I^2 values of 25, 50, and 75 represent low, medium, and high heterogeneity, respectively (Higgins and Thompson, 2002). Negative I^2 values were rounded up to zero.

The same meta-analysis macro (Wilson, DB 2006) was used to calculate mean effect sizes for each BPM outcome using a random effects model. The random effects model assumes there are factors changing across experiments (e.g. experimenter experience, animal age, time of day during testing) that might influence observed effect sizes, making it the most appropriate model for the current analyses. Drug data were not included in the meta-analyses. Forest plots were generated using GraphPad Prism 7 (GraphPad Software, La Jolla, CA).

2.5. Statistical comparison of DAT KD effect sizes in C57 vs. 129 genetic backgrounds

To assess whether DAT KD mutant behavioral profiles differed from WT mice across the two genetic backgrounds studied, DAT KD vs. WT littermate effect sizes were calculated separately for each genetic background (C57 and 129). For each BPM outcome measure, the difference in effect size between C57 and 129 mice was calculated as a z-

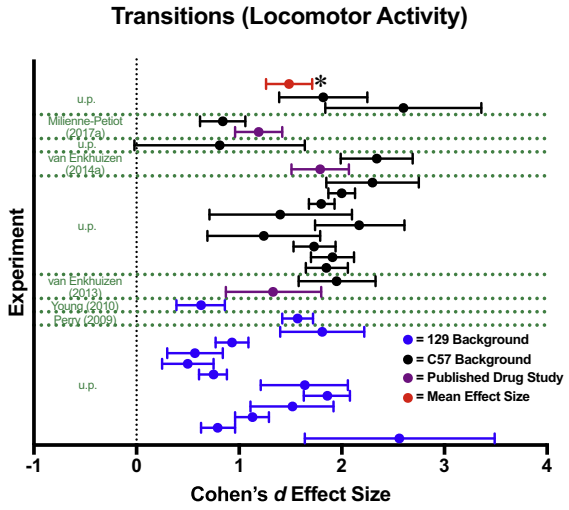

Fig. 2. DAT KD mice exhibit a significant, reproducible increase in locomotor activity, as measured by transitions in the BPM, compared with WT littermates across experiments. Each row represents a different experiment, and each point is the calculated effect size for that experiment. Blue points: data from DAT KD mice on a 129/SvJ (129) background; black points: data from DAT KD mice on a C57BL/6J (C57) background. The dashed black vertical line denotes a Cohen's *d* effect size of zero. All data points are located to the right of the dashed line, reflecting an increase in transitions in DAT KD mice vs. WT littermates for each experiment. Homogeneity of effect sizes was tested using a Q statistic prior to mean effect size calculation. $Q(27) = 71.0$ ($p < 0.0001$; $I^2 = 62.0$), indicating heterogeneity of effect sizes in the overall effect size pool. Using a random effects model, the calculated mean effect size was 1.49 (95% CI = 1.26, 1.71, Z = 13.0, $^*p < 0.0001$; red data point), reflecting a substantial, reproducible increase in transitions in DAT KD mice vs. WT littermates across all experiments. When analyzed separately by background strain, the effect sizes from mice on a C57 background were homogeneous ($Q(14) = 18.0$, $p > 0.05$; $I^2 = 22.2$), and resulted in a mean effect size of 1.80 (95% CI = 1.57, 2.02, Z = 15.6, $p < 0.0001$). Effect sizes from mice on a 129 background remained heterogeneous ($Q(12) = 27.0$, $p < 0.01$; $I^2 = 55.6$), and a mean effect size of 1.15 was calculated (95% CI = 0.87, 1.44, Z = 7.9, $p < 0.0001$). Purple data points represent effect sizes calculated from published studies that included a drug treatment group (DAT KD on drug vs. WT littermates on drug). DAT KD mice treated with 15 mg/kg valproate chow for 28 days exhibited a reduction in transitions vs. DAT KD mice treated with normal chow (black (control) and purple (drug) data points labeled "van Enkhuizen et al. (2013)"). A reduction in transitions was also observed in DAT KD mice treated with 30 mg/kg of a tyrosine hydroxylase inhibitor alpha-methyl-p-tyrosine (AMPT) for 4 days vs. DAT KD mice treated with vehicle (black (control) and purple (drug) data points labeled "van Enkhuizen et al. (2014a)"). In contrast, treatment of DAT KD mice with 0.3 mg/kg of the $D_{2/3}$ antagonist brexpiprazole resulted in an increase in transitions vs. DAT KD mice treated with vehicle (black (control) and purple (drug) data points labeled "Milienne-Petiot et al. (2017a)"). Data points are organized in ascending (bottom to top) chronological order based on date of data acquisition. Published data are denoted by: first author last name (year of publication); "u.p." = unpublished data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

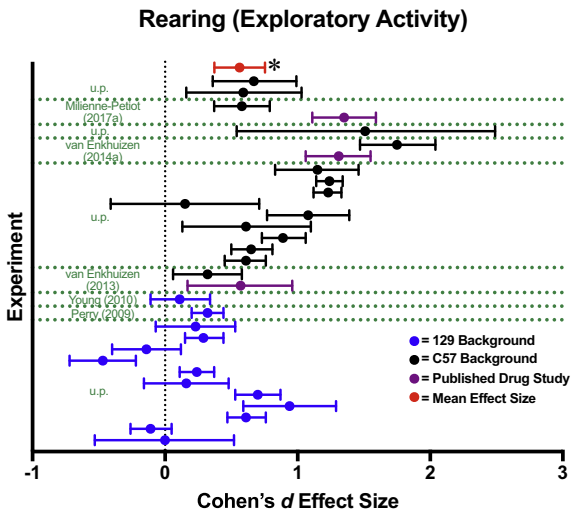

Fig. 3. DAT KD mice exhibit a significant, reproducible increase in exploratory behavior, as measured by number of rearings in the BPM, compared with WT littermates across experiments. Each row represents a different experiment, and each point is the calculated effect size for that experiment. Blue points: data from DAT KD mice on a 129/SvJ (129) background; black points: data from DAT KD mice on a C57BL/6J (C57) background. The dashed black vertical line denotes a Cohen's *d* effect size of zero. All black points are located to the right of the dashed line, whereas 3 blue points are to the left of the line. Homogeneity of effect sizes was tested using a Q statistic prior to mean effect size calculation. Q(27) = 62.0 ($p < 0.01$; $I^2 = 56.5$), indicating heterogeneity of effect sizes in the overall effect size pool. Using a random effects model, the calculated mean effect size was 0.56 (95% CI = 0.37, 0.75, Z = 5.8, $^*p < 0.0001$; red data point), reflecting a moderate, reproducible increase in rearing in DAT KD mice vs. WT littermates across experiments. When analyzed separately by background strain, the effect sizes from mice on a C57 background were homogeneous (Q(14) = 19.0, $p > 0.05$; $I^2 = 26.3$) and a mean effect size of 0.89 (95% CI = 0.69, 1.10, Z = 8.5, $p < 0.0001$) was calculated. Effect sizes from mice on a 129 background were also homogenous (Q(12) = 14.2, $p > 0.05$; $I^2 = 15.5$), and a mean effect size of 0.24 was calculated (95% CI = 0.05, 0.43, Z = 2.5, $p < 0.05$). Purple data points represent effect sizes calculated from published studies that included a drug treatment group (DAT KD on drug vs. WT littermates on drug). DAT KD mice treated with 15 mg/kg valproate chow for 28 days exhibited an increase in number of rearings vs. DAT KD mice treated with normal chow (black (control) and purple (drug) data points labeled "van Enkhuizen et al. (2013)"). A reduction in number of rearings was observed in DAT KD mice treated with 30 mg/kg of a tyrosine hydroxylase inhibitor alpha-methyl-p-tyrosine (AMPT) for 4 days vs. DAT KD mice treated with vehicle (black (control) and purple (drug) data points labeled "van Enkhuizen et al. (2014a)"). Treatment of DAT KD mice with 0.3 mg/kg of the $D_{2/3}$ antagonist brexpiprazole resulted in a substantial increase in number of rearings vs. DAT KD mice treated with vehicle (black (control) and purple (drug) data points labeled "Milienne-Petiot et al. (2017a)"). Data points are organized in ascending (bottom to top) chronological order based on date of data acquisition. Published data are denoted by: first author last name (year of publication); "u.p." = unpublished data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

score: $z = {}^{-}|d_{C57} - d_{129}| / \sqrt{(SE^2_{C57} + SE^2_{129})}$.

A two-tailed *p*-value was determined from the z-score using a standard normal distribution (van de Lagemaat et al., 2017).

3. Results

For each BPM outcome measure, 28 effect sizes were obtained for the differences between DAT KD and WT mice. Initially, DAT KD mice were maintained on a 129 background. Subsequently, DAT KD mice were crossed over to the C57 background strain. Data for both 129 and C57 background strains were obtained and utilized in the mean effect size analyses. For holepokes, only 26 effect sizes were obtained given that two experiments did not record any holepokes during the session. All mean effect sizes were calculated using a random effects model.

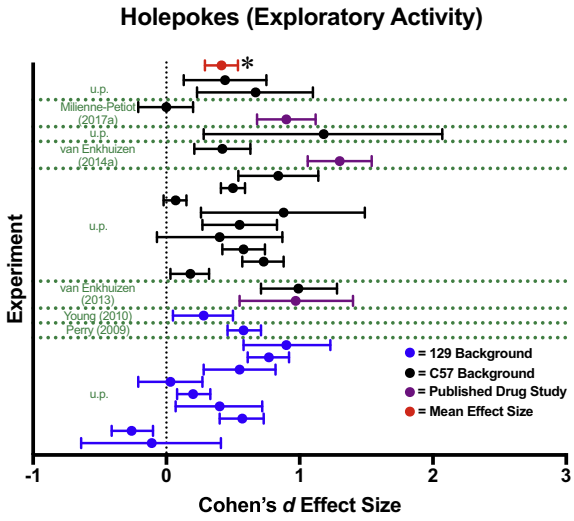

Fig. 4. DAT KD mice exhibit a significant, reproducible increase in exploratory behavior, as measured by number of holepokes in the BPM, compared with WT littermates across experiments. Each row represents a different experiment, and each point is the calculated effect size for that experiment. Blue points: data from DAT KD mice on a 129/SvJ (129) background; black points: data from DAT KD mice on a C57BL/6J (C57) background. The dashed black vertical line denotes a Cohen's *d* effect size of zero. Homogeneity of effect sizes was tested using a Q statistic prior to mean effect size calculation. Q(25) = 25.4 ($p > 0.05$; $I^2 = 1.6$), and the calculated mean effect size was 0.41 (95% CI = 0.29, 0.54, Z = 6.5, $^*p < 0.0001$; red data point), reflecting a moderate, reproducible increase in holepokes in DAT KD mice vs. WT littermates across experiments. Purple data points represent effect sizes calculated from published studies that included a drug treatment group (DAT KD on drug vs. WT littermates on drug). DAT KD mice treated with 15 mg/kg valproate chow for 28 days exhibited no change in number of holepokes vs. DAT KD mice treated with normal chow (black (control) and purple (drug) data points labeled "van Enkhuizen et al. (2013)"). A substantial increase in holepokes was observed in DAT KD mice treated with 30 mg/kg of a tyrosine hydroxylase inhibitor alpha-methyl-p-tyrosine (AMPT) for 4 days vs. DAT KD mice treated with vehicle (black (control) and purple (drug) data points labeled "van Enkhuizen et al. (2014a)"). Similarly, treatment of DAT KD mice with 0.3 mg/kg of the $D_{2/3}$ antagonist brexpiprazole resulted in a sizeable increase in holepokes vs. DAT KD mice treated with vehicle (black (control) and purple (drug) data points labeled "Milienne-Petiot et al. (2017a)"). Data points are organized in ascending (bottom to top) chronological order based on date of data acquisition. Published data are denoted by: first author last name (year of publication); "u.p." = unpublished data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.1. BPM outcome measure: Transitions

Testing for homogeneity of effect sizes for transitions resulted in Q (27) = 71.0 ($p < 0.0001$; $I^2 = 62.0$). The resulting mean effect size was 1.49 (95% CI = 1.26, 1.71; Z = 13.0; $p < 0.0001$; Fig. 2), indicating that DAT KD mice exhibited a large, reproducible increase in number of transitions in the BPM vs. WT littermates. This hyperactivity of DAT KD mice compared to WT littermates is consistent with modeling BD mania.

To address the heterogeneity seen in effect sizes for transitions, a sub-analysis was performed to determine whether background strain contributed to the observed heterogeneity. DAT KD vs. WT mice on a C57 background were analyzed for homogeneity of effect sizes. Q(14) = 18.0 ($p > 0.05$; $I^2 = 22.2$), and the mean effect size was 1.80 (95% CI = 1.57, 2.02; Z = 15.6; $p < 0.0001$). In contrast, analyses using mice on a 129 background resulted in Q(12) = 27.0 ($p < 0.01$; $I^2 = 55.6$) and a mean effect size of 1.15 (95% CI = 0.87, 1.44; Z = 7.9; $p < 0.0001$). Taken together, effect sizes observed from DAT KD vs. WT littermates on a 129 background were the main driver of heterogeneity in the overall transitions analysis, as effect sizes from DAT KD vs. WT mice on a C57 background displayed a low level of heterogeneity.

Another sub-analysis was performed to address the remaining

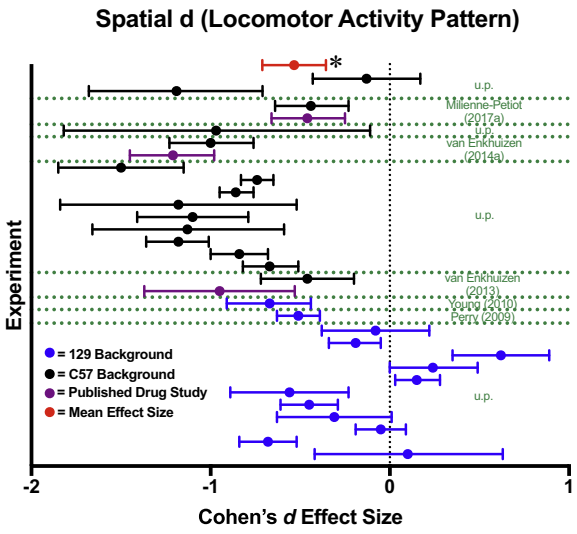

Fig. 5. DAT KD mice exhibit a significant, reproducible reduction in spatial d in the BPM compared with WT littermates across experiments. Each row represents a different experiment, and each point is the calculated effect size for that experiment. Blue points: data from DAT KD mice on a 129/SvJ (129) background; black points: data from DAT KD mice on a C57BL/6J (C57) background. The dashed black vertical line denotes a Cohen's *d* effect size of zero. All black points are located to the left of the dashed line, whereas 4 blue points are to the right of the line. Homogeneity of effect sizes was tested using a Q statistic prior to mean effect size calculation. $Q(27) = 54.4$ ($p < 0.01$; $I^2 = 50.4$), indicating heterogeneity of effect sizes in the overall effect size pool. Using a random effects model, the calculated mean effect size was -0.53 (95% CI = -0.71, -0.36, $Z = -5.9$, $^*p < 0.0001$; red data point), reflecting a moderate, reproducible decrease in spatial d in DAT KD mice vs. WT littermates across experiments. When analyzed separately by background strain, the effect sizes from mice on a C57 background were homogeneous ($Q(14) = 12.1$, $p > 0.05$, $I^2 = 0.0$), and resulted in a mean effect size of -0.83 (95% CI = -1.00, -0.66, $Z = -9.7$, $p < 0.0001$). Effect sizes from mice on a 129 background were also homogenous ($Q(12) = 16.9$, $p > 0.05$, $I^2 = 29.0$), and a mean effect size of -0.21 was calculated (95% CI = -0.41, 0.00, $Z = -1.9$, $p > 0.05$). Purple data points represent effect sizes calculated from published studies that included a drug treatment group (DAT KD on drug vs. WT littermates on drug). DAT KD mice treated with 15 mg/kg valproate chow for 28 days exhibited a reduction in spatial d vs. DAT KD mice treated with normal chow (black (control) and purple (drug) data points labeled "van Enkhuizen et al. (2013)"). Similarly, a reduction in spatial d was also observed in DAT KD mice treated with 30 mg/kg of a tyrosine hydroxylase inhibitor alpha-methyl-p-tyrosine (AMPT) for 4 days vs. DAT KD mice treated with vehicle (black (control) and purple (drug) data points labeled "van Enkhuizen et al. (2014a)"). Treatment of DAT KD mice with 0.3 mg/kg of the $D_{2/3}$ antagonist brexpiprazole resulted in no change in spatial d vs. DAT KD mice treated with vehicle (black (control) and purple (drug) data points labeled "Milienne-Petiot et al. (2017a)"). Data points are organized in ascending (bottom to top) chronological order based on date of data acquisition. Published data are denoted by: first author last name (year of publication); "u.p." = unpublished data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

heterogeneity of effect sizes seen in DAT KD mice on a 129 background. The overall mean effect size analysis included cohorts with repeated exposure to the BPM paradigm, and it was hypothesized that these exposures might introduce variability to the analysis. Experiments on DAT KD mice (129 background) with repeated exposure to the BPM were excluded in this sub-analysis to test this hypothesis. This "first exposure" analysis yielded $Q(5) = 13.3$ ($p < 0.05$; $I^2 = 62.4$), indicating that heterogeneity of effect sizes still persisted in this subgroup. This analysis was not conducted in mice on a C57 background given the consistency in effect sizes observed in the strain-specific analysis.

Beyond homogeneity across studies, the mean effect size of DAT KD vs. WT mice on a C57 background was larger than the 129 background, indicative of a larger difference in transitions between DAT KD and WT mice on a C57 vs. 129 background. This difference in effect size was confirmed by computing a z-score and deriving a two-tailed *p*-value from a standard normal distribution, resulting in a z-score of -3.50 ($p < 0.001$ at $\alpha = 0.05$). Therefore, compared to mice on a 129 background, DAT KD mice on a C57 background provided a significantly more robust model of hyperactivity in the BPM.

3.2. BPM outcome measure: Rearing

The test for homogeneity of effect sizes of DAT KD vs. WT mice for rearing resulted in $Q(27) = 62.0$ ($p < 0.01$; $I^2 = 56.5$) and a mean effect size of 0.56 was subsequently calculated (95% CI = 0.37, 0.75; $Z = 5.8$; $p < 0.0001$; Fig. 3). The mean effect size for rearing was smaller in magnitude compared to the mean effect size for transitions, reflecting a moderate, reproducible increase in exploratory activity for DAT KD vs. WT littermates in the BPM across experiments.

Given the heterogeneity of effect sizes for rearing, follow-up sub-analyses were performed to determine sources of heterogeneity. Consistent with transitions (above), heterogeneity analyses of C57 and 129 background strains were conducted separately. For the C57 background, $Q(14) = 19.0$ ($p > 0.05$; $I^2 = 26.3$) and the mean effect size was 0.89 (95% CI = 0.69, 1.10; $Z = 8.5$; $p < 0.0001$). For mice on a 129 background, $Q(12) = 14.2$ ($p > 0.05$; $I^2 = 15.5$) and the mean effect size was 0.24 (95% CI = 0.05, 0.43; $Z = 2.5$; $p < 0.05$). Therefore, both background strains exhibited internal homogeneity of effect sizes, indicating that the combination of background strains in the overall analysis contributed to the observed heterogeneity.

Similar to transitions, DAT KD mice on a C57 background exhibited a more robust increase in number of rearings vs. WT littermates compared to DAT KD mice on a 129 background (mean effect size of 0.89 for C57 mice vs. 0.24 for 129 mice). The computed z-score for the C57 and 129 background effect sizes was -4.54 ($p < 0.00001$). Again, the DAT KD mice on a C57 background were a significantly more robust model for observing rearing differences vs. WT littermates in the BPM compared to mice on the 129 background.

3.3. BPM outcome measure: Holepokes

For the overall holepokes effect size analysis, $Q(25) = 25.4$ ($p > 0.05$; $I^2 = 1.6$), and meta-analysis resulted in a mean effect size of 0.41 (95% CI = 0.29, 0.54; $Z = 6.5$; $p < 0.0001$; Fig. 4). This value reflected a moderate, reproducible increase in holepokes in DAT KD vs. WT littermates, corroborating the increase in exploratory activity seen with rearing. Sub-analysis by background strain confirmed homogeneity of effect sizes in both C57 and 129 background strain groups (C57: $Q(14) = 12.6$, $p > 0.05$, $I^2 = 0.0$; mean effect size = 0.45, 95% CI = 0.28, 0.61, $Z = 5.4$, $p < 0.0001$; 129: $Q(10) = 12.4$, $p > 0.05$, $I^2 = 19.4$; mean effect size = 0.37, 95% CI = 0.15, 0.58, $Z = 3.4$, $p < 0.001$).

The calculated z-score between the C57 and 129 sub-groups was -0.58 ($p > 0.05$), indicating that the increased effect size magnitude in the C57 sub-group was not significantly different from the effect size observed in the 129 sub-group. Therefore, for exploratory activity measured by holepokes in the BPM, genetic background strain did not influence behavioral profiles.

3.4. BPM outcome measure: Spatial d

Spatial d, a calculated measure of the geometrical pattern of locomotor activity in the BPM chamber, was the fourth BPM outcome measure analyzed. Spatial d values range between 1 and 2, with a value closer to 1 indicating a straight-line path and values closer to 2 indicating local, circumscribed movements (closer to a filled plane). For the overall analysis, $Q(27) = 54.4$ ($p < 0.01$; $I^2 = 50.4$), and the mean effect size was -0.53 (95% CI = -0.71, -0.36; $Z = -5.9$; $p < 0.0001$; Fig. 5), reflecting a moderate, reproducible reduction of spatial d in DAT KD vs. WT littermates across BPM experiments.

Sub-analyses revealed that effect sizes in both the C57 and 129 backgrounds were homogenous (C57: $Q(14) = 12.1$, $p > 0.05$,

I^2 = 0.0; 129: Q(12) = 16.9, $p > 0.05$, I^2 = 29.0). The mean effect size for DAT KD mice on a C57 background was −0.83 (95% CI = −1.00, −0.66, Z = −9.7, $p < 0.0001$), whereas the mean effect size for DAT KD mice on a 129 background was −0.21 (95% CI = −0.41, 0.00, Z = −1.9, $p > 0.05$). The *p*-value for the mean effect size in the 129 sub-analysis failed to reach significance, indicating that DAT KD mice on a 129 background do not exhibit a significant reduction in spatial d in the BPM compared to WT littermates. Accordingly, DAT KD mice on a C57 background exhibited a significantly larger mean effect size magnitude vs. WT littermates compared to DAT KD vs. WT mice on a 129 background (z-score = −4.52, $p < 0.00001$).

3.5. Drug effects on BPM outcomes in DAT KD mice

For three published studies, data from drug treatment groups were collected in order to compute effect sizes comparing DAT KD mice vs. WT littermates treated with various drugs. Drug effect sizes were qualitatively compared to effect sizes generated from vehicle treatment groups in the same study (DAT KD vs. WT littermates treated with vehicle) to determine how behavioral outcomes in DAT KD mice were affected by treatment. All drug treatments were conducted with DAT KD and WT littermate mice on a C57 background.

When DAT KD mice were treated over 28 days with 15 mg/kg valproate chow, they exhibited a reduction in number of transitions and spatial d, an increase in rearing, and no change in holepokes, vs. DAT KD mice treated with normal chow (Figs. 2–5, data points labeled "van Enkhuizen et al., 2013"). When mice were treated with a tyrosine hydroxylase inhibitor alpha-methyl-p-tyrosine (AMPT; reduces the synthesis of dopamine), DAT KD mice exhibited reduced transitions, rearing, and spatial d, and increased holepokes (Figs. 2–5, data points labeled "van Enkhuizen et al., 2014a"). Finally, treatment with the $D_{2/3}$ antagonist brexpiprazole increased the number of transitions, rearings, and holepokes, with no change in spatial d, in DAT KD mice (Figs. 2–5, data points labeled "Milienne-Petiot et al., 2017a").

4. Discussion

Across cohorts and experiments, DAT KD mice exhibited a reproducible increase in number of transitions, rearing, and holepokes, plus reduced spatial d, vs. WT littermates in the BPM (see Fig. 6 for a summary of overall mean effect sizes calculated for the four BPM outcome measures discussed). The reproducibility of this model was observed irrespective of background strain as DAT KD vs. WT littermate mice exhibited this profile on both C57 and 129 background strains (with the exception of spatial d—DAT KD mice on a 129 background did not exhibit a significant reduction in spatial d vs. WT littermates). However, DAT KD mice on a C57 background were a more robust model for these outcomes compared to mice on a 129 background for three out of the four BPM outcomes (transitions, rearing, and spatial d). The reproducible changes in the exploratory profile of these mice in the BPM, in addition to their face, construct, and predictive validity, support the use of this model for BD mania research. Drug-induced changes in DAT KD BPM locomotor activity, exploratory activity, and spatial d support use of the DAT KD mouse line as a translatable model of bipolar mania in humans. Furthermore, drug-induced changes in DAT KD BPM profiles can be used to screen the effects of novel therapeutics on various behavioral aspects of mania.

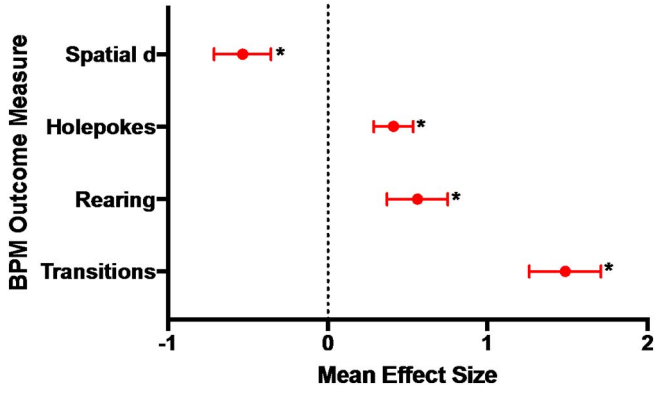

Fig. 6. Summary of mean effect sizes generated for each BPM outcome measure. The dashed black vertical line denotes a mean effect size of zero. All 4 BPM outcomes were significantly altered in DAT KD mice vs. WT littermates, with increases in transitions, rearing, and holepokes, and a reduction in spatial d. Transitions: random effects model mean effect size = 1.49 (95% CI = 1.26, 1.71, Z = 13.0). Rearing: random effects model mean effect size = 0.56 (95% CI = 0.37, 0.75, Z = 5.8). Holepokes: random effects model mean effect size = 0.41 (95% CI = 0.29, 0.54, Z = 6.5). Spatial d: random effects model mean effect size = −0.53 (95% CI = −0.71, −0.36, Z = −5.9). *$p < 0.0001$.

A dysregulated dopamine system has been implicated in multiple psychiatric disorders, including schizophrenia, attention-deficit hyperactivity disorder, bipolar disorder, and autism spectrum disorders (Dichter et al., 2012). It is, therefore, essential to fully understand the contributions of altered dopamine homeostasis on these disorders. The DAT KD mouse line was created for this purpose, in addition to circumventing the growth retardation phenotype seen with dopamine transporter knockout lines (Zhuang et al., 2001). Numerous studies have been conducted with the DAT KD mouse line since its creation, yielding valuable insights into the influences of altered dopamine signaling on psychiatry-relevant behavioral profiles (Berridge et al., 2005; Cagniard et al., 2006; Milienne-Petiot et al., 2017b; Peciña et al., 2003; Perry et al., 2009; Tilley et al., 2007; van Enkhuizen et al., 2014a, 2014b; Young et al., 2011, 2010a; Zhuang et al., 2001). This work is the first to demonstrate reproducible findings—across > 20 cohorts of studies spanning > 13 years of research—that these mice exhibit reliable BD mania-relevant behavior in the BPM (BD mania patients exhibit increased locomotor activity, more object interactions (exploratory activity), and reduced spatial d (Perry et al., 2009) when compared with healthy participants in the BPM). Therefore, the DAT KD model in our BPM paradigm is reliable, reproducible, and also translatable to human populations.

Given the vast number of studies utilizing genetically engineered mutant mouse lines to address research questions, surprisingly little work has been done to assess the reproducibility of experimental findings across mouse background strains and cohorts. Recently, van de Lagemaat et al. (2017) assessed the similarity of mutant behavioral phenotypes across different genetic background strains using three mutations crossed onto two strains. They analyzed the behavioral impact of each genetic mutation on both background strains using 16 behavioral variables, calculating mutant vs. WT effect sizes for each variable. Importantly, it was observed that 85% of the mutant phenotypes exhibited similar effect sizes across C57 and 129S5 strains. The current research adds to these findings that behavioral consistency of mutant strains can be observed across multiple backgrounds.

Some minor differences between robustness of signal and background strain were observed in the current dataset, however. For example, DAT KD mice on a 129 background did not exhibit a significant reduction in spatial d in the BPM vs. WT littermates. Accordingly, DAT KD mice on a C57 background were a significantly more robust model for spatial d in the BPM, as well as for transitions and rearing outcomes. Importantly, however, while effect sizes were larger in magnitude for C57 mice for transitions and rearing outcomes, the direction of change was the same across strains (increased number of transitions and rearings). Effect sizes for holepokes did not differ significantly between the two background strains. This similarity was reflected by the fact that the overall pool of effect sizes used to compute the holepoke mean effect size was homogeneous. Hence, despite minor differences, the effect of the mutation drove the same changes in behavior in the same direction of effect.

Similar changes in effect sizes were observed in the original study

using 129 background DAT KD mice (Zhuang et al., 2001). Although not tested in the BPM, overall activity and rearing was measured using an open-field test. In terms of activity, the KD vs. WT effect size was 1.12, almost identical to the 1.15 effect size in our BPM 129 background strain studies. These data support the inter-laboratory reproducibility in addition to the intra-laboratory reproducibility described above. Although other activity studies in DAT KD mice have been conducted and report similar directionality of effect (hyperactivity in DAT KD vs. WT littermates), insufficient data were presented to calculate effect sizes for comparison (Cagniard et al., 2014; Tilley et al., 2007). Increased rearing in DAT KD vs. WT littermates was also observed (Zhuang et al., 2001), with a larger effect size (1.07) compared with our reports (0.24). Although not similar in magnitude, they are consistent in direction. Magnitude differences were likely due to inter-laboratory variations (e.g. BPM which includes holepokes vs. open-field test without, testing in the dark vs. light phase, etc.). The inter-laboratory consistency supports the strong intra-laboratory reproducibility described here.

In our intra-laboratory analysis, heterogeneity of effect sizes observed with transitions, rearing, and spatial d prompted sub-analyses by strain for these three BPM outcomes. When considered separately, homogeneity of effect sizes was observed within each background strain with the exception of transitions, as while the C57 sub-analysis was homogeneous, the 129 sub-analysis remained heterogeneous. We hypothesized that DAT KD mice on a 129 background might be more sensitive to subsequent exposure to the BPM chamber, and performed an additional sub-analysis that included only first-time exposures to the BPM. Effect sizes for this sub-analysis remained heterogeneous, suggesting that factors other than repeated exposure drove heterogeneity of transitions in the 129 background strain in the BPM. DAT KD mice on a 129 background may be more susceptible to subtle changes in experimental conditions, such as light levels, noise levels, or room temperature, thereby driving effect size variation. Alternatively, the 129 background strain in particular contains numerous sub-strains that cross four distinct genetic lineages (Kiselycznyk and Holmes, 2011). The sub-strain used for the generation of the DAT KD line, 129/SvJ, was created by combining the 129/Sv strain with another unknown strain (Threadgill et al., 1997), perhaps introducing intra-genetic background variability and driving the heterogeneous effect sizes seen with transitions in the current analysis. The consistency of effect sizes of DAT KD vs. WT mice on the C57 background supports its continued use in research on BD mania-related behaviors and their treatment (Milienne-Petiot et al., 2017b; van Enkhuizen et al., 2014a, 2014b, 2013; Young et al., 2011).

Utilizing the C57 background strain of mutation, we have conducted several drug treatment studies. Chronic valproate exposure resulted in a reduction in locomotor activity, an increase or no change in exploratory behavior, and a worsening of locomotor patterns (van Enkhuizen et al., 2013). Chronic valproate (and other) treatment-induced reduction in locomotor effect sizes was also observed in mania patients repeatedly tested over time in the human BPM, where no change on specific exploration or locomotor pattern effect sizes was observed (Minassian et al., 2011). These data highlight the need for therapeutics that address multiple aberrant behaviors in mania, rather than drugs restricted to certain behavioral dimensions (e.g. hyperactivity). Furthermore, given that valproate and lithium exhibit equivalent response rates when used to treat mania episodes in patients with bipolar disorder (Yildiz et al., 2011), the inclusion of this valproate study further demonstrates the validity of the DAT KD mouse line in modeling mania-like behaviors. Future studies should test the response to other anti-mania treatments (e.g. lithium) in order to further assess the validity of this model.

With administration of AMPT, which depleted tyrosine hydroxylase levels, improvements in 3 out of 4 BPM outcomes were observed (excluding holepokes). These improvements in the DAT KD BPM profile lend support to the hypothesis that reduced DAT expression, and resultant hyperdopaminergia, drove the aberrant behavioral profile observed at baseline in DAT KD mice (van Enkhuizen et al., 2014a). Finally, brexpiprazole treatment worsened or exerted no effect on all 4 BPM outcomes in DAT KD mice (Milienne-Petiot et al., 2017a), potentially indicating that this drug did not specifically target the mechanism by which DAT KD mice exhibit mania-like symptoms. Overall, these data demonstrate that testing the DAT KD mouse line in the BPM provides a strong model for psychiatric therapeutic development and testing.

When developing animal models of human disorders, it is important to also assess response to negative control treatments (Young et al., 2010c). Previous studies have reported an amphetamine-induced reduction in locomotor activity in DAT KD mice (Zhuang et al., 2001), which is seemingly contradictory to its use as a model of mania-like behavior given that amphetamine can induce mania episodes in patients and is used for modeling mania in normal rodents (Arban et al., 2005; Cosgrove et al., 2016; Valvassori et al., 2008). This reduction in locomotor activity, however, could reflect a hypersensitivity to amphetamine (as was observed with another DAT inhibitor, GBR 12909; Young et al., 2010b). Amphetamine at 1 mg/kg induced hyperactivity in WT mice, yet did not affect baseline hyperactivity in DAT KD mice (Zhuang et al., 2001); WT mice continued to exhibit a dose-dependent increase in locomotor activity at 3 mg/kg, whereas DAT KD mice exhibited a reduction in locomotion activity at this dose. A bimodal effect of amphetamine on normal rodent locomotor behaviors was previously reported, where low doses (< 2 mg/kg) induced heightened locomotor activity and higher doses reduced locomotor activity via increased stereotypic behaviors (Minassian et al., 2016; Yates et al., 2007). Hence, it is likely that DAT KD mice exhibit a leftward shift in this bimodal pattern, whereby lower amphetamine doses that induce hyperactivity in normal rodents now functionally act like higher doses to suppress locomotor activity – consistent with hypersensitivity of DAT KD mice to GBR12909. Therefore, the results reported by Zhuang et al. (2001) may reflect a hypersensitivity to amphetamine, consistent with the hypersensitivity to amphetamine seen in patients with bipolar disorder (Wingo and Ghaemi, 2008), and requires lower doses to be assessed for confirmation.

In conclusion, the DAT KD mouse line in the BPM paradigm is a robust, reproducible model for quantitatively analyzing mania-like behaviors, as well as for testing potential therapeutics on such behaviors. A general effort must be made to characterize and understand the consistency of behavioral profiles across experiments, as well as to measure the impact of genetic background strains on those profiles (Bespalov et al., 2016). Particularly in the field of psychiatry, where disorders mainly manifest as alterations in behavior, ensuring reproducibility in behavioral outcomes from animal models is critical to advancing the field and developing more effective treatments to improve patient care.

References

Aird, F., Kandela, I., Mantis, C., Reproducibility Project: Cancer Biology, 2017. Replication study: BET bromodomain inhibition as a therapeutic strategy to target c-Myc. elife 6. http://dx.doi.org/10.7554/eLife.21253.

Anand, A., Barkay, G., Dzemidzic, M., Albrecht, D., Karne, H., Zheng, Q.-H., Hutchins, G.D., Normandin, M.D., Yoder, K.K., 2011. Striatal dopamine transporter availability in unmedicated bipolar disorder. Bipolar Disord. 13, 406–413. http://dx.doi.org/10.1111/j.1399-5618.2011.00936.x.

Arban, R., Maraia, G., Brackenborough, K., Winyard, L., Wilson, A., Gerrard, P., Large, C., 2005. Evaluation of the effects of lamotrigine, valproate and carbamazepine in a rodent model of mania. Behav. Brain Res. 158, 123–132. http://dx.doi.org/10.1016/j.bbr.2004.08.015.

Arent, C.O., Valvassori, S.S., Steckert, A.V., Resende, W.R., Dal-Pont, G.C., Lopes-Borges, J., Amboni, R.T., Bianchini, G., Quevedo, J., 2015. The effects of n-acetylcysteine and/or deferoxamine on manic-like behavior and brain oxidative damage in mice submitted to the paradoxal sleep deprivation model of mania. J. Psychiatr. Res. 65, 71–79. http://dx.doi.org/10.1016/j.jpsychires.2015.04.011.

Baker, M., 2016. 1,500 scientists lift the lid on reproducibility. Nature 533, 452–454. http://dx.doi.org/10.1038/533452a.

Berridge, K.C., Aldridge, J.W., Houchard, K.R., Zhuang, X., 2005. Sequential super-stereotypy of an instinctive fixed action pattern in hyper-dopaminergic mutant mice: a

model of obsessive compulsive disorder and Tourette's. BMC Biol. 3, 4. http://dx.doi.org/10.1186/1741-7007-3-4.
Bespalov, A., Steckler, T., Altevogt, B., Koustova, E., Skolnick, P., Deaver, D., Millan, M.J., Bastlund, J.F., Doller, D., Witkin, J., Moser, P., O'Donnell, P., Ebert, U., Geyer, M.A., Prinssen, E., Ballard, T., Macleod, M., 2016. Failed trials for central nervous system disorders do not necessarily invalidate preclinical models and drug targets. Nat. Rev. Drug Discov. 15, 516. http://dx.doi.org/10.1038/nrd.2016.88.
Braff, D.L., Geyer, M.A., 1990. Sensorimotor gating and schizophrenia. Human and animal model studies. Arch. Gen. Psychiatry 47, 181–188.
Cagniard, B., Balsam, P.D., Brunner, D., Zhuang, X., 2006. Mice with chronically elevated dopamine exhibit enhanced motivation, but not learning, for a food reward. Neuropsychopharmacology 31, 1362–1370. http://dx.doi.org/10.1038/sj.npp.1300966.
Cagniard, B., Sotnikova, T.D., Gainetdinov, R.R., Zhuang, X., 2014. The dopamine transporter expression level differentially affects responses to cocaine and amphetamine. J. Neurogenet. 28, 112–121. http://dx.doi.org/10.3109/01677063.2014.908191.
Cilia, J., Hatcher, P.D., Reavill, C., Jones, D.N.C., 2005. Long-term evaluation of isolation-rearing induced prepulse inhibition deficits in rats: an update. Psychopharmacology 180, 57–62. http://dx.doi.org/10.1007/s00213-004-2139-5.
Cosgrove, V.E., Kelsoe, J.R., Suppes, T., 2016. Toward a valid animal model of bipolar disorder: how the research domain criteria help bridge the clinical-basic science divide. Biol. Psychiatry 79, 62–70. http://dx.doi.org/10.1016/j.biopsych.2015.09.002.
Dichter, G.S., Damiano, C.A., Allen, J.A., 2012. Reward circuitry dysfunction in psychiatric and neurodevelopmental disorders and genetic syndromes: animal models and clinical findings. J. Neurodev. Disord. 4, 19. http://dx.doi.org/10.1186/1866-1955-4-19.
van Enkhuizen, J., Geyer, M.A., Kooistra, K., Young, J.W., 2013. Chronic valproate attenuates some, but not all, facets of mania-like behaviour in mice. Int. J. Neuropsychopharmacol. 16, 1021–1031. http://dx.doi.org/10.1017/S1461145712001198.
van Enkhuizen, J., Geyer, M.A., Halberstadt, A.L., Zhuang, X., Young, J.W., 2014a. Dopamine depletion attenuates some behavioral abnormalities in a hyperdopaminergic mouse model of bipolar disorder. J. Affect. Disord. 155, 247–254. http://dx.doi.org/10.1016/j.jad.2013.08.041.
van Enkhuizen, J., Henry, B.L., Minassian, A., Perry, W., Milienne-Petiot, M., Higa, K.K., Geyer, M.A., Young, J.W., 2014b. Reduced dopamine transporter functioning induces high-reward risk-preference consistent with bipolar disorder. Neuropsychopharmacology 39, 3112–3122. http://dx.doi.org/10.1038/npp.2014.170.
van Enkhuizen, J., Geyer, M.A., Minassian, A., Perry, W., Henry, B.L., Young, J.W., 2015. Investigating the underlying mechanisms of aberrant behaviors in bipolar disorder from patients to models. Neurosci. Biobehav. Rev. 58, 4–18. http://dx.doi.org/10.1016/j.neubiorev.2015.08.008.
Fries, G.R., Valvassori, S.S., Bock, H., Stertz, L., Magalhães, P.V. da S., Mariot, E., Varela, R.B., Kauer-Sant'Anna, M., Quevedo, J., Kapczinski, F., Saraiva-Pereira, M.L., 2015. Memory and brain-derived neurotrophic factor after subchronic or chronic amphetamine treatment in an animal model of mania. J. Psychiatr. Res. 68, 329–336. http://dx.doi.org/10.1016/j.jpsychires.2015.05.006.
Geyer, M.A., Braff, D.L., 1987. Startle habituation and sensorimotor gating in schizophrenia and related animal models. Schizophr. Bull. 13, 643–668.
Geyer, M.A., Russo, P.V., Masten, V.L., 1986. Multivariate assessment of locomotor behavior: pharmacological and behavioral analyses. Pharmacol. Biochem. Behav. 25, 277–288.
Geyer, M.A., Wilkinson, L.S., Humby, T., Robbins, T.W., 1993. Isolation rearing of rats produces a deficit in prepulse inhibition of acoustic startle similar to that in schizophrenia. Biol. Psychiatry 34, 361–372.
Gould, T.D., Einat, H., 2007. Animal models of bipolar disorder and mood stabilizer efficacy: a critical need for improvement. Neurosci. Biobehav. Rev. 31, 825–831. http://dx.doi.org/10.1016/j.neubiorev.2007.05.007.
Greenwood, T.A., Alexander, M., Keck, P.E., McElroy, S., Sadovnick, A.D., Remick, R.A., Kelsoe, J.R., 2001. Evidence for linkage disequilibrium between the dopamine transporter and bipolar disorder. Am. J. Med. Genet. 105, 145–151.
Greenwood, T.A., Schork, N.J., Eskin, E., Kelsoe, J.R., 2006. Identification of additional variants within the human dopamine transporter gene provides further evidence for an association with bipolar disorder in two independent samples. Mol. Psychiatry 11 (125–33), 115. http://dx.doi.org/10.1038/sj.mp.4001764.
Henry, B.L., Minassian, A., Patt, V.M., Hua, J., Young, J.W., Geyer, M.A., Perry, W., 2013. Inhibitory deficits in euthymic bipolar disorder patients assessed in the human behavioral pattern monitor. J. Affect. Disord. 150, 948–954. http://dx.doi.org/10.1016/j.jad.2013.05.020.
Higgins, J.P.T., Thompson, S.G., 2002. Quantifying heterogeneity in a meta-analysis. Stat. Med. 21, 1539–1558. http://dx.doi.org/10.1002/sim.1186.
Horrigan, S.K., Reproducibility Project: Cancer Biology, 2017. Replication study: the CD47-signal regulatory protein alpha (SIRPa) interaction is a therapeutic target for human solid tumors. elife 6. http://dx.doi.org/10.7554/eLife.18173.
Horrigan, S.K., Courville, P., Sampey, D., Zhou, F., Cai, S., Reproducibility Project: Cancer Biology, 2017. Replication study: melanoma genome sequencing reveals frequent PREX2 mutations. elife 6. http://dx.doi.org/10.7554/eLife.21634.
Kandela, I., Aird, F., Reproducibility Project: Cancer Biology, 2017. Replication study: discovery and preclinical validation of drug indications using compendia of public gene expression data. elife 6. http://dx.doi.org/10.7554/eLife.17044.
Kiselycznyk, C., Holmes, A., 2011. All (C57BL/6) mice are not created equal. Front. Neurosci. 5, 10. http://dx.doi.org/10.3389/fnins.2011.00010.
van de Lagemaat, L.N., Stanford, L.E., Pettit, C.M., Strathdee, D.J., Strathdee, K.E., Elsegood, K.A., Fricker, D.G., Croning, M.D.R., Komiyama, N.H., Grant, S.G.N., 2017. Standardized experiments in mutant mice reveal behavioural similarity on 129S5 and C57BL/6J backgrounds. Genes Brain Behav. 16, 409–418. http://dx.doi.org/10.1111/gbb.12364.
Mantis, C., Kandela, I., Aird, F., Reproducibility Project: Cancer Biology, 2017. Replication study: coadministration of a tumor-penetrating peptide enhances the efficacy of cancer drugs. elife 6. http://dx.doi.org/10.7554/eLife.17584.
Milienne-Petiot, M., Geyer, M.A., Arnt, J., Young, J.W., 2017a. Brexpiprazole reduces hyperactivity, impulsivity, and risk-preference behavior in mice with dopamine transporter knockdown—a model of mania. Psychopharmacology 234, 1017–1028. http://dx.doi.org/10.1007/s00213-017-4543-7.
Milienne-Petiot, M., Kesby, J.P., Graves, M., van Enkhuizen, J., Semenova, S., Minassian, A., Markou, A., Geyer, M.A., Young, J.W., 2017b. The effects of reduced dopamine transporter function and chronic lithium on motivation, probabilistic learning, and neurochemistry in mice: modeling bipolar mania. Neuropharmacology 113, 260–270. http://dx.doi.org/10.1016/j.neuropharm.2016.07.030.
Minassian, A., Henry, B.L., Geyer, M.A., Paulus, M.P., Young, J.W., Perry, W., 2010. The quantitative assessment of motor activity in mania and schizophrenia. J. Affect. Disord. 120, 200–206. http://dx.doi.org/10.1016/j.jad.2009.04.018.
Minassian, A., Henry, B.L., Young, J.W., Masten, V., Geyer, M.A., Perry, W., 2011. Repeated assessment of exploration and novelty seeking in the human behavioral pattern monitor in bipolar disorder patients and healthy individuals. PLoS One 6, e24185. http://dx.doi.org/10.1371/journal.pone.0024185.
Minassian, A., Young, J.W., Cope, Z.A., Henry, B.L., Geyer, M.A., Perry, W., 2016. Amphetamine increases activity but not exploration in humans and mice. Psychopharmacology 233, 225–233. http://dx.doi.org/10.1007/s00213-015-4098-4.
Open Science Collaboration, 2015. Estimating the reproducibility of psychological science. Science 349 (80) (aac4716-aac4716). https://doi.org/10.1126/science.aac4716.
Paulus, M.P., Geyer, M.A., 1991. A temporal and spatial scaling hypothesis for the behavioral effects of psychostimulants. Psychopharmacology 104, 6–16.
Paulus, M.P., Geyer, M.A., 1993. Three independent factors characterize spontaneous rat motor activity. Behav. Brain Res. 53, 11–20.
Peciña, S., Cagniard, B., Berridge, K.C., Aldridge, J.W., Zhuang, X., 2003. Hyperdopaminergic mutant mice have higher "wanting" but not "liking" for sweet rewards. J. Neurosci. 23, 9395–9402.
Perry, W., Minassian, A., Paulus, M.P., Young, J.W., Kincaid, M.J., Ferguson, E.J., Henry, B.L., Zhuang, X., Masten, V.L., Sharp, R.F., Geyer, M.A., 2009. A reverse-translational study of dysfunctional exploration in psychiatric disorders: from mice to men. Arch. Gen. Psychiatry 66, 1072–1080. http://dx.doi.org/10.1001/archgenpsychiatry.2009.58.
Prickaerts, J., Moechars, D., Cryns, K., Lenaerts, I., van Craenendonck, H., Goris, I., Daneels, G., Bouwknecht, J.A., Steckler, T., 2006. Transgenic mice overexpressing glycogen synthase kinase 3beta: a putative model of hyperactivity and mania. J. Neurosci. 26, 9022–9029. http://dx.doi.org/10.1523/JNEUROSCI.5216-05.2006.
Rao, J.S., Kellom, M., Reese, E.A., Rapoport, S.I., Kim, H.-W., 2012. Dysregulated glutamate and dopamine transporters in postmortem frontal cortex from bipolar and schizophrenic patients. J. Affect. Disord. 136, 63–71. http://dx.doi.org/10.1016/j.jad.2011.08.017.
Risbrough, V.B., Masten, V.L., Caldwell, S., Paulus, M.P., Low, M.J., Geyer, M.A., 2006. Differential contributions of dopamine D1, D2, and D3 receptors to MDMA-induced effects on locomotor behavior patterns in mice. Neuropsychopharmacology 31, 2349–2358. http://dx.doi.org/10.1038/sj.npp.1301161.
Seifuddin, F., Pirooznia, M., Judy, J.T., Goes, F.S., Potash, J.B., Zandi, P.P., 2013. Systematic review of genome-wide gene expression studies of bipolar disorder. BMC Psychiatry 13, 213. http://dx.doi.org/10.1186/1471-244X-13-213.
Swerdlow, N.R., Light, G.A., Thomas, M.L., Sprock, J., Calkins, M.E., Green, M.F., Greenwood, T.A., Gur, R.E., Gur, R.C., Lazzeroni, L.C., Nuechterlein, K.H., Radant, A.D., Seidman, L.J., Siever, L.J., Silverman, J.M., Stone, W.S., Sugar, C.A., Tsuang, D.W., Tsuang, M.T., Turetsky, B.I., Braff, D.L., 2017. Deficient prepulse inhibition in schizophrenia in a multi-site cohort: internal replication and extension. Schizophr. Res. http://dx.doi.org/10.1016/j.schres.2017.05.013.
Threadgill, D.W., Yee, D., Matin, A., Nadeau, J.H., Magnuson, T., 1997. Genealogy of the 129 inbred strains: 129/SvJ is a contaminated inbred strain. Mamm. Genome 8, 390–393.
Tilley, M.R., Cagniard, B., Zhuang, X., Han, D.D., Tiao, N., Gu, H.H., 2007. Cocaine reward and locomotion stimulation in mice with reduced dopamine transporter expression. BMC Neurosci. 8, 42. http://dx.doi.org/10.1186/1471-2202-8-42.
Valvassori, S.S., Petronilho, F.C., Réus, G.Z., Steckert, A.V., Oliveira, V.B.M., Boeck, C.R., Kapczinski, F., Dal-Pizzol, F., Quevedo, J., 2008. Effect of N-acetylcysteine and/or deferoxamine on oxidative stress and hyperactivity in an animal model of mania. Prog. Neuro-Psychopharmacol. Biol. Psychiatry 32, 1064–1068. http://dx.doi.org/10.1016/j.pnpbp.2008.02.012.
Weiss, I.C., Feldon, J., Domeney, A.M., 1999. Isolation rearing-induced disruption of prepulse inhibition: further evidence for fragility of the response. Behav. Pharmacol. 10, 139–149.
Wingo, A.P., Ghaemi, S.N., 2008. Frequency of stimulant treatment and of stimulant-associated mania/hypomania in bipolar disorder patients. Psychopharmacol. Bull. 41, 37–47.
Yates, J.W., Meij, J.T.A., Sullivan, J.R., Richtand, N.M., Yu, L., 2007. Bimodal effect of amphetamine on motor behaviors in C57BL/6 mice. Neurosci. Lett. 427, 66–70. http://dx.doi.org/10.1016/j.neulet.2007.09.011.
Yildiz, A., Vieta, E., Leucht, S., Baldessarini, R.J., 2011. Efficacy of antimanic treatments: meta-analysis of randomized, controlled trials. Neuropsychopharmacology 36, 375–389. http://dx.doi.org/10.1038/npp.2010.192.
Young, J.W., Dulcis, D., 2015. Investigating the mechanism(s) underlying switching between states in bipolar disorder. Eur. J. Pharmacol. 759, 151–162. http://dx.doi.org/

10.1016/j.ejphar.2015.03.019.
Young, J.W., Goey, A.K.L., Minassian, A., Perry, W., Paulus, M.P., Geyer, M.A., 2010a. The mania-like exploratory profile in genetic dopamine transporter mouse models is diminished in a familiar environment and reinstated by subthreshold psychostimulant administration. Pharmacol. Biochem. Behav. 96, 7–15. http://dx.doi.org/10.1016/j.pbb.2010.03.014.
Young, J.W., Goey, A.K.L., Minassian, A., Perry, W., Paulus, M.P., Geyer, M.A., 2010b. GBR 12909 administration as a mouse model of bipolar disorder mania: mimicking quantitative assessment of manic behavior. Psychopharmacology 208, 443–454. http://dx.doi.org/10.1007/s00213-009-1744-8.
Young, J.W., Zhou, X., Geyer, M.A., 2010c. Animal models of schizophrenia. Curr. Top. Behav. Neurosci. 4, 391–433.
Young, J.W., van Enkhuizen, J., Winstanley, C.A., Geyer, M.A., 2011. Increased risk-taking behavior in dopamine transporter knockdown mice: further support for a mouse model of mania. J. Psychopharmacol. 25, 934–943. http://dx.doi.org/10.1177/0269881111400646.
Zhuang, X., Oosting, R.S., Jones, S.R., Gainetdinov, R.R., Miller, G.W., Caron, M.G., Hen, R., 2001. Hyperactivity and impaired response habituation in hyperdopaminergic mice. Proc. Natl. Acad. Sci. 98, 1982–1987. http://dx.doi.org/10.1073/pnas.98.4.1982.