Asbury *et al. Behavioral and Brain Functions* (2025) 21:1
https://doi.org/10.1186/s12993-024-00261-y

Behavioral and Brain Functions

RESEARCH Open Access

Host genetics maps to behaviour and brain structure in developmental mice

Sarah Asbury[1], Jonathan K. Y. Lai[1], Kelly C. Rilett[1], Zeeshan Haqqee[1], Benjamin C. Darwin[2], Jacob Ellegood[2], Jason P. Lerch[2,3] and Jane A. Foster[1,4,5*]

Abstract

Gene-environment interactions in the postnatal period have a long-term impact on neurodevelopment. To effectively assess neurodevelopment in the mouse, we developed a behavioural pipeline that incorporates several validated behavioural tests to measure translationally relevant milestones of behaviour in mice. The behavioral phenotype of 1060 wild type and genetically-modified mice was examined followed by structural brain imaging at 4 weeks of age. The influence of genetics, sex, and early life stress on behaviour and neuroanatomy was determined using traditional statistical and machine learning methods. Analytical results demonstrated that neuroanatomical diversity was primarily associated with genotype whereas behavioural phenotypic diversity was observed to be more susceptible to gene-environment variation. We describe a standardized mouse phenotyping pipeline, termed the Developmental Behavioural Milestones (DBM) Pipeline released alongside the 1000 Mouse Developmental Behavioural Milestones (1000 Mouse DBM) database to institute a novel framework for reproducible interventional neuroscience research.

Keywords Neurodevelopment, Structural MRI, Random forest, Machine learning, Early life stress

Background

Advances in translational neuroscience have expanded our understanding of gene-environment mechanisms that underlie neurodevelopment and behaviour. Moreover, employing clinically-relevant genetic and environmental strategies to perturb development in mice allows researchers to determine the influence and interaction of genetics, sex, and early life environmental insults on the neurodevelopmental trajectory. These relationships are biologically shared with neurodevelopment in children and can provide an understanding of the biological mechanisms that may contribute to neurodevelopmental disorders (NDD).

The current study investigated the influence of genetics, sex, and early life stress on neurodevelopmental trajectory of mouse behaviour and neuroanatomy. The current study utilized well-established developmental milestones and behavioural assays to map the trajectory of three commonly used inbred (Balb/C, C57Bl/6NCrl, FVB) and one commonly used outbred (CD1) strain of mice, as well as T cell deficient mice on a C57Bl/6 background, through knockout of the T cell receptor (TCR) β and δ chains (*TCRβ-/-δ-/-*) and *Fmr1* knock out (*FMR1-KO*) mice on a FVB background. Following behavioural testing, neuroanatomical differences were assessed by post-mortem brain volumetric analysis from MRI

*Correspondence:
Jane A. Foster
jane.foster@utsouthwestern.edu
[1] Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, ON, Canada
[2] Mouse Imaging Centre, The Hospital for Sick Children, Toronto, ON, Canada
[3] Wellcome Centre for Integrative Neuroimaging, FMRIB, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK
[4] Research Institute at St. Joe's Hamilton, Hamilton, ON, Canada
[5] Center for Depression Research and Clinical Care, Department of Psychiatry, O'Donnell Brain Institute, University of Texas Southwestern Medical Center, Dallas, TX, USA

at 4 weeks of age. The experimental design is shown in Fig. 1. To facilitate tracking of individual mice, mice were tattooed at postnatal day (P) 2. Stressors included immune challenge on P3 with lipopolysaccharide (LPS) and maternal separation overnight on P9. These early life stressors in rodents have been shown to have a long-term impact on behaviour and brain structure [1–3]. Milestones and behavioural tests were selected to model key milestones of development in children and particularly those domains that are impacted in NDD.

Mice were assessed using the 28-day Developmental Behavioural Milestones pipeline—which included righting reflex, ultrasonic vocalizations (USV), eye opening, open field (OF), social behaviour, and self-grooming (Fig. 1). Righting reflex is a motor milestone in mice that is typically achieved between P4 to P6 [4, 5]. The time it takes for a mouse pup to flip itself on to four paws from a supine position decreases daily during this period. In response to separation from the dam, pups emit ultrasonic vocalizations. USVs peak at the end of the first postnatal week of life and provide a measure of early-life social communication [6–8]. Eye opening is developmental milestone that occurs between P10 and P17 [4, 5, 9]. The open field test [10] measures exploratory behavior and locomotor activity [11–13]. OF was assessed at P17 using a modified protocol that included a smaller chamber and a shorter test time to accommodate the developmental stage. Sociability was measured using the 3-chambered test [12–14]. Self-grooming analysis is measured to evaluate repetitive behaviour [12, 13, 15]. Collectively, the DBM pipeline comprehensively phenotypes early-life behaviour and neurodevelopment metrics.

The influence of genetic background on neurodevelopment and behaviour in mice has long been recognized and extensively studied across many laboratory strains using numerous behavioural and neurodevelopmental tests. Activity differences have been described between mouse strains within and beyond the strains included in our experiments [16–20]. There are also recognized differences in sociability [12], early life social communication[7], grooming [19, 21, 22], and motor development [10, 19, 23] between strains. In accordance with inbred mouse strain background influencing behaviour, several studies have identified candidate genes or quantitative trait loci associated with phenotypic differences for mouse neurobehavioral traits including activity [24], rearing [25], grooming [25], early life social communication [26] and sociability [27]. These studies broadly demonstrate genetic determinants of normal developmental neurobehavioral phenotypes. A large-scale study of mouse behaviour and cognition demonstrated that single nucleotide polymorphisms also influence behavioural, anxiety, and cognitive outcomes in mice [28]. Such findings indicate genetics as a major source of variability in behavioural phenotypes.

Using both traditional statistical and machine learning methods, we identify neurodevelopmental, behavioural, and neuroanatomical differences between inbred mouse strains, and demonstrate similar neurodevelopmental trajectories exist between knockout mice and their wild type strains. These findings robustly characterize the early-life behavioural trajectories of common mouse laboratory strains and validate the importance of host genetics underlying behavioural and neuroanatomical differences. Furthermore, accurately summarizing established neurodevelopmental phenotypes previously reported between mouse strains demonstrates the external validity of the DBM Pipeline framework.

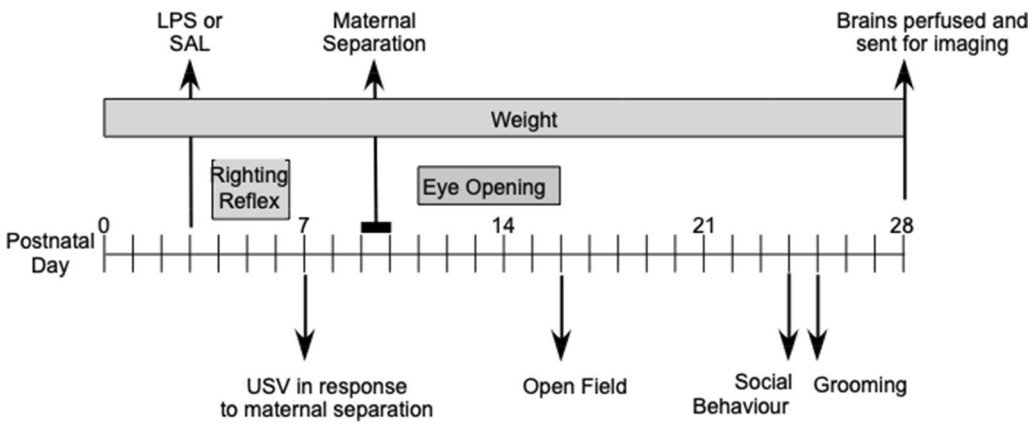


Fig. 1 Experimental design showing postnatal challenges, developmental outcomes and behavioural tests used in the first 4 weeks of postnatal life

Methods

Animals

Breeding pairs, purchased from Charles River, were bred in house and pregnant dams were ordered from Charles River. Lack of functional T-cells was due to genetic knockout of both β and δ chains of the T-cell receptor (*TCR β-/-δ-/-*) [29]. The mice were provided by Dr Andrew McPherson at McMaster University, and bred in house. Mice strains included FVB, C57Bl/6NCrl (B6), Balb/C, CD1, *FMR1*-KO (FVB.129P2.*Fmr1tm1Cgr*/J, stock #004624) and *TCRβ-/-δ-/-* on a C57Bl/6 background [30]. Mice were housed at the animal facility at St. Joseph's Healthcare with food and water available ad libitum, under a 12 h:12 h light dark cycle with lights on at 5 AM and lights off at 5 PM. Birth was set as postnatal day 0 (P0). On P2, litters were culled to 10 pups and pups were uniquely tattooed on their paw for identification. Pups were weaned on P21 and caged by sex with up to 4 littermates per cage. Metadata including breeding location, litter size, sex, and treatment for experimental mice is provided in Supplemental File 1. Weight of experimental mice is provided in Supplementary Fig. S1 and S2. All experimental procedures were approved by the Animal Research Ethics Board, McMaster University in accordance with the guidelines of the Canadian Council on Animal Care.

Experimental design

Our experimental design included developmental milestones, righting reflex, ultrasonic vocalizations (USVs), eye opening, open field, sociability and self-grooming. This analysis was completed over the first 28 days of life with postnatal challenges occurred on P3 (immune) and P9 (stress) (Fig. 1). Experimental mice (Table 1) included 1060 pups from 159 litters (29 FVB, 30 B6, 38 Balb/C, 28 CD1, 14 *FMR1*-KO, 20 *TCRβ-/-δ-/-*).

Postnatal challenges

On P3, pups were administered lipopolysaccharide (LPS) (0.1 mg/kg; *Escherichia coli* LPS; Sigma, St. Louis, MO) or saline i.p. at 50 µl/g. Injections were done at 4 PM. On P9, maternal separation (MS) or control [18] treatment was administered. For MS litters, pups were weighed at 4 PM, and then the dam was removed from the home cage at 5 PM. The home cage was then placed on a heating pad at 37 °C until 9 AM the next day, when the pups were weighed and the dam returned to the cage. Pups were again weighed at 4 PM. For control litters, pups were weighed at 4 PM on P9, 9 AM on P10 and 4 PM on P10; dams were removed briefly during weights and then returned to the home cage.

Table 1 Experimental mice included in behavioural analysis

Strain	Sex	P3 treatment	P9 treatment	N
B6	F	SAL	CON	23
			MS	25
		LPS	CON	24
			MS	22
	M	SAL	CON	25
			MS	23
		LPS	CON	24
			MS	23
TCR	F	SAL	CON	13
			MS	16
		LPS	CON	13
			MS	14
	M	SAL	CON	9
			MS	15
		LPS	CON	11
			MS	18
BalbC	F	SAL	CON	25
			MS	27
		LPS	CON	32
			MS	34
	M	SAL	CON	21
			MS	23
		LPS	CON	23
			MS	24
CD1	F	SAL	CON	31
			MS	32
		LPS	CON	29
			MS	28
	M	SAL	CON	36
			MS	28
		LPS	CON	36
			MS	30
FVB	F	SAL	CON	28
			MS	23
		LPS	CON	24
			MS	44
	M	SAL	CON	21
			MS	24
		LPS	CON	24
			MS	25
FMR1	F	SAL	CON	10
			MS	10
		LPS	CON	11
			MS	11
	M	SAL	CON	14
			MS	11
		LPS	CON	10
			MS	13

Righting reflex

On P4–P6, pups were tested for motor development by timing their ability to right themselves after being placed on their backs. Testing was done at 4 PM. A completed righting was defined by all four paws on the ground simultaneously. Time was kept with a stopwatch. The maximum score was set at 30 s at which time the pup was manually righted. The righting reflex score was calculated by taking the average righting reflex time from P4–P6. Lower righting reflex score therefore indicates earlier achievement of the righting reflex.

Eye opening

From P10 to P17, eye opening was scored daily: a score of 0, 1, or 2 was assigned per mouse reflecting the number of eyes open. Eye opening score was calculated by averaging the eye opening data for each mouse from P10 to P17. A higher eye opening score reflects earlier eye opening of one or both eyes.

USV recordings

On P7, pups were consecutively maternally separated from the dam and littermates and placed in a custom-made sound-attenuating chamber. Testing took place during the first half of the active period, at least one hour after the active cycle began. Ultrasonic vocalizations were recorded for 3 min and then each pup was transferred to a separate holding cage. After all pups were tested, the pups were returned to the dam. Vocalizations were digitized using an Avisoft UltraSoundGate 116–200 recording device and USG CM116/CMPA microphone. The microphone was clamped to a retort stand and situated 17.5 cm above the center of the recording chamber. Calls were digitized in real-time and subsequently analyzed with Avisoft SAS Lab Pro.

Open field

At P17, pups were tested in the open field. Behavioural testing was conducted in a non-colony room after a 30 min habituation to the room. Testing took place in low light (approximately 200 Lux) during the first half of the active period. Behaviours were automatically recorded for 15 min using the Kinder Scientific Smart Rack System consisting of a 24 cm wide × 45 cm long × 24 cm high cage rack system, with 22 infrared beams (7 X & 15 Y) and a rearing option (22 additional beams). A Plexiglas® box was placed at one end of the chamber to reduce the testing chamber size to 24 × 23 cm. Data were collected using MotorMonitor® software (Kinder Scientific, Poway, CA). A maximum of 6 pups were tested at a time. After all pups had undergone testing, they were returned to the dam; maternal separation did not exceed 30 min.

Sociability

At P24, sociability was tested using a 3-chamber apparatus [14, 31]. Behavioural testing was conducted in a non-colony room after a 20 min habituation to the room. Testing took place during the first half of the active period. Mice were placed in the centre zone of the chamber with no access to the other chambers for 5 min. Subsequently, an age-, strain-, and sex- matched stranger mouse was placed in an inverted cup in one of the side chambers, the doors from the centre chamber to the outer chambers opened, and behaviour was recorded for 10 min. Live-tracking and automated videotape analysis was done using EthoVision® software.

Self-grooming

At P25, mice were observed for 10 min in a standard housing cage without bedding and scored for time spent grooming [13]. Mice were habituated to the testing cage for 10 min prior to grooming test. Grooming behaviours were manually scored using AnyMaze software. Grooming data was only included if inter-rater reliability was demonstrated ($\kappa > 0.05$).

Perfusions

Mice were perfused at P28 at St. Joseph's Healthcare in Hamilton, Ontario prior to being transferred to the Mouse Imaging Centre in Toronto for imaging and analysis. The perfusion protocol was as follows: Mice were anesthetized with ketamine/xylazine and intracardially perfused with 30 mL of 0.1 M PBS containing 10 U/mL heparin and 2 mM ProHance (Bracco Diagnostics Ltd. a Gadolinium contrast agent) followed by 30 mL of 4% paraformaldehyde (PFA) containing 2 mM ProHance [32]. Perfusions were performed with a minipump at a rate of approximately 1 mL/min. After perfusion, mice were decapitated and the skin, lower jaw, ears, and the cartilaginous nose tip were removed. The brain and remaining skull structures were incubated in 4% PFA + 2 mM ProHance overnight at 4 °C then transferred to 0.1 M PBS for at least 1 month prior to MRI scanning [33].

Behavioural analysis

Analysis of behavioural data to identify main effects of genotype, sex, and age was completed in SPSS Ver. 25 by multivariate general linear models followed by Bonferroni-corrected posthoc tests. Omnibus analysis was conducted in behavioural data prior to P9 (RR, USV), which has 2 treatment (P3) groups (LPS, SAL), and in behavioural data following P9 (eye opening, OF, sociability, self-grooming) which has 4 treatment (P3_P9) groups

(SALCON, SALMS, LPSCON, LPSMS). Based on main effects and interactions identified in the omnibus analyses, behavioural test-specific multivariate (genotype, sex, treatment—all test outcomes) and univariate analysis (genotype, sex—individual test outcomes) was conducted (SPSS Ver. 25) followed by pairwise comparisons using Bonferonni-corrected posthoc tests.

MRI analysis

All images were acquired using a 7.0 Tesla MRI Scanner (Agilent Inc., Palo Alto, CA). For neuroanatomical scans, a 40 cm inner bore diameter gradient was used with a maximum gradient strength of 30 G/cm. This was used in conjunction with a custom built solenoid array capable of scanning 16 brains at a time [34, 35]. To assess the volume differences throughout the brain, a T2-weighted 3D fast spin echo (FSE) sequence is used that is optimized for gray/white matter contrast. Parameters for the sequence: TR of 2000 ms, and TEs of 10 ms per echo for 6 echoes, with the centre of k-space being acquired on the 4th echo, TE_{eff} of 42 ms, two averages, field-of-view (FOV) of 14×28×25 mm and matrix size of 250×504×450 giving an image with 0.056 mm isotropic voxels. In the first phase encode direction consecutive lines of k-space were assigned to alternating echoes to move discontinuity related ghosting to the edges of the FOV [36]. This sequence involves oversampling k-space in the first phase encode direction by a factor of 2 to avoid ghosting in the final image. This gives a FOV of 28 mm but is subsequently cropped to 14 mm after reconstruction. Total imaging time was approximately 12 h [34].

To visualize and compare differences between mouse brains the images are registered together (both linear and nonlinear registration). All scans were then resampled with the appropriate transform and averaged to create a population atlas representing the average anatomy of the study sample. All registrations were performed with a combination of mni_autoreg tools [37] and advanced normalization tools (ANTs) [38, 39]. The result of the registration was to have all scans deformed into alignment with each other in an unbiased fashion. This allowed for the analysis of the deformations needed to take each individual mouse's anatomy into this final atlas space. Allowing modelling of the deformation fields as they relate to genotype [40, 41]. Significant volume changes were calculated by warping a pre-existing classified MRI atlas onto the population atlas, which allowed for the volume of 159 segmented structures encompassing cortical lobes, large white matter structures (i.e. corpus callosum), ventricles, cerebellum, brain stem, and olfactory bulbs to be assessed in all brains. This classified atlas incorporated the structures from three separate pre-existing atlases: (1) which delineated 62 different regions throughout the brain [42], (2) which further classified multiple different areas in the cerebellum [43], and (3) which divided the cortex into 64 different regions [44]. Moreover, these measurements were examined on a voxel-wise basis in order to localize the differences found within regions or across the brain. Regional and voxelwise differences were examined using two measures, absolute volume (mm^3) as well as relative volume (% total brain volume). Multiple comparisons in this study were controlled for using the False Discovery Rate [45].

Brain regions—as defined using the Dorr-Steadman-Ullmann-Richards-Qiu-Egan (40 micron, DSURQE) atlas and Allen Brain Institute hierarchy—were used for PCA, clustering and Random Forest analyses [46]. Hierarchical brain regions were pruned to level 8 for all regions for most regions. Hippocampus was instead pruned to level 7 and Midbrain was pruned to level 9. Only regions represented by terminal nodes were used for downstream analyses. Regions with absolute volume > 1 mm^3 were removed. The regions used in downstream analyses are available as a list in Supplemental File 2 or as a hierarchical visualization (Supplementary Fig. S3).

Integrative analysis of behavioural data

Principle component analysis

PCA and hierarchical clustering analyses were performed using *prcomp*, *hclust,* and *dist* in the stats R package [47]. Behavioural outcomes were Z-score scaled for PCA and hierarchical clustering. Behavioural outcomes include: average time for righting reflex (P4–P6); USV calls, duration, and intercall interval; average eye opening score; open field rearing; time spent in mouse, center, and empty chamber in three chamber test; and grooming frequency, duration, and latency. Z-score normalized MRI relative volumes from postnatal day 28 were used for neuroanatomical PCA. Brain regions used for analysis are listed in Supplementary file 3. Hierarchical clustering was performed on all neurodevelopmental and behavioural or MRI data using Euclidian distances and Ward.D2 clustering method. 3D PCA visualizations were generated using RGL R package [48] and *scatter3D* from car R package [49]. For behavioural network analysis, the optimal number of clusters (k) was determined using WSS and silhouette plots generated using Factoextra R package [50]. Clustering results were then visualized using *dendrogram* in base R, and separated into k = 2 groups using *cutree*. Cluster stability for behavioural and neurodevelopmental, neuroanatomical, or behavioural network clustering was determined using fpc package *clusterboot* [51]. Jaccard mean > 0.75 is considered a stable cluster, 0.6–0.75 is considered a pattern in the data, 0.5–0.6 is considered an unstable cluster and < 0.5 is considered a dissolved cluster.

Random Forest models

Random Forest models predicting genotype, sex, or combined treatment from behaviour and neurodevelopmental outcomes were generated using Z-score scaled behavioural and neurodevelopment outcomes as predictor variables. Sex, P3 and P9 treatment, or genotype was also included as predictor variables for the respective models. Models predicting genotype using relative neuroanatomical volume were trained using predictor variables listed in Supplementary File 1. Random Forest models were generated using the randomForest package in R [52]. Ten test and validation sets were generated via randomly sampled 80/20 splits. Test sets were further segmented via randomly sampled 80/20 split for fivefold cross-validation hyper parameter tuning of the number of trees (500–5000) and number of predictor variables available at each split (behaviour predictor variables mtry = 1:7 or neuroanatomical predictor variables mtry = 1:10). Model accuracy was reported as validation set results using Misclassification Error (ME) or Adjusted Rand Index (ARI) calculated using the e1071 package [53]. Predictor variable importance was measured via mean decreased Gini index for the best performing set. When the random forest model accurately predicts the response variable, high predictor variable importance rank implies an association between the predictor variable and response variable.

Results

Representative behavioral results for B6 mice

Representative graphs of the behavioural results are provided for male and female C57Bl/6 mice in Figs. 2 and 3. Righting reflex develops between P4 and P6 as shown in Fig. 2A and the average time to right is also provided. Behavioral read outs for all genotypes (n, mean, SE) are provided in supplemental file 3.

Behavioural analysis revealed primary effect of genotype

Genotype, sex, and treatment effects and interactions were investigated using multivariate and univariate ANOVAs for each neurodevelopmental and behavioural outcome. Neurodevelopmental and behavioural differences were primarily genotype effects. There were no independent sex effects for between-subject effects ANOVA and limited independent treatment effects (Table 2).

The average righting reflex neurodevelopmental metric measures the average righting reflex time between P4 and P6. Low average righting reflex values indicates faster development of the righting reflex neurodevelopmental milestone. Genotype was a main effect for average righting reflex ($p < 0.001$) (Table 2), and posthoc analysis indicates that FVB mice have earlier development of the righting reflex compared to other wildtype inbred genotypes (Table 3). *TCRβ-/-δ-/-* mice also have significantly higher average righting reflex time compared to wildtype B6 ($p < 0.001$), indicating delayed righting reflex development in T-cell deficient mice (Table 3).

USVs measure early life social communication. Genotype was a significant main effect for USV call number ($p < 0.001$) (Table 2). Most strains and knockout mice have significant pair-wise differences (Table 3). For wildtype mice, Balb/C mice have the lowest USV call number, while CD1 mice have the highest USV call number (Table 3). USV call duration was shortest in B6 mice and longest in FVB mice (Table 3). USV intercall intervals had a significant genotype effect ($p < 0.001$), driven by high intercall intervals of Balb/C mice—in addition to lowest number of USV calls compared to other mice (Table 3). *TCRβ-/-δ-/-* mice have a higher number of calls compared to wildtype B6 mice ($p < 0.001$), however T cell knockout did not affect USV duration (Table 3). Contrastingly, *FMR1-KO* mice have a lower number of USV calls compared to wildtype FVB mice ($p < 0.001$), with no differences in USV duration (Table 3).

Higher eye opening score indicates the eye opening milestone occurred earlier in development. Eye opening score was dependent on genotype ($p < 0.001$) (Table 2) and lowest in B6 and Balb/C mice (Table 3), indicating these strains attain eye-opening milestone later in development. In contrast, CD1 and FVB have the highest eye-opening scores (Table 3). There were no knockout effects in eye opening score (Table 3).

Balb/C mice have pronouncedly lower activity metrics in the open field test; both total distance and rearing scores were lower than other strains (Table 3). Rearing frequency was highest in FVB mice, followed closely by CD1 mice (Table 3). Notably, *FMR1-KO* mice had a higher rearing frequency compared to their background strain FVB ($p < 0.001$) (Table 3). Furthermore, rearing and open field total distance were lower in *TCRβ-/-δ-/-* mice compared to their wildtype B6 counterparts ($p = 0.027$, $p = 0.050$), indicating decreased activity in T-cell deficient mice (Table 3). Open field total distance was similar between the wildtype genotypes, except for Balb/C mice which had markedly lower activity (Table 3). Both open field total distance and rearing had a treatment effect ($p < 0.001$, $p = 0.005$) (Table 2) driven by reduced activity in maternally separated mice that were not LPS-treated compared to controls (Table 3).

Sociability, as measured by time spent in the mouse chamber, had a significant genotype effect ($p < 0.001$) and was lowest in Balb/C mice but similar between wildtype genotypes (Table 3). There were no significant differences in sociability for knockout mice (Table 3). LPS treated mice that were not maternally separated spent less time

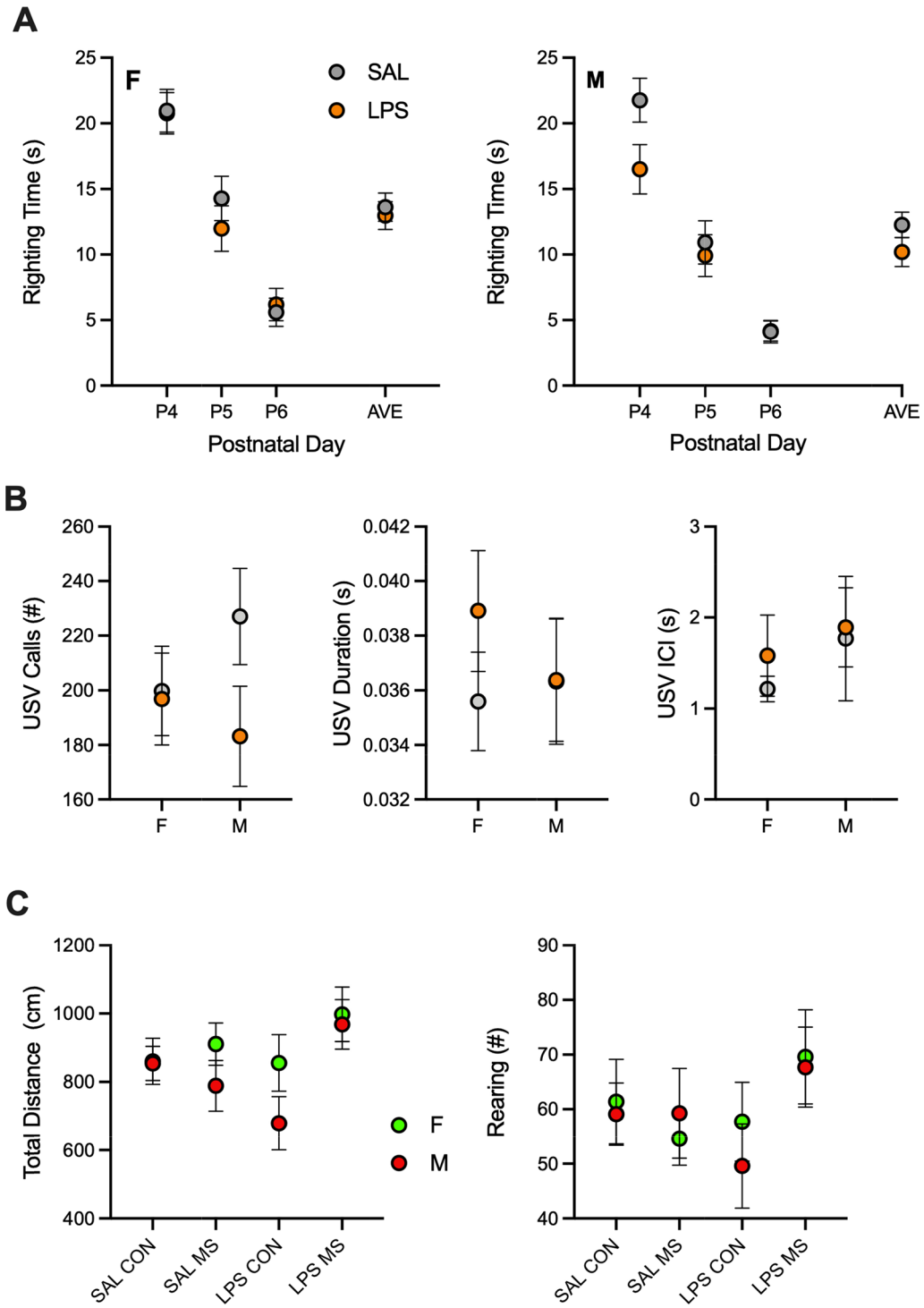


Fig. 2 Representative behavioural graphs for righting reflex on postnatal day (P) 4–6 and average righting reflex time (**A**), ultrasonic vocalizations (USV) (**B**), SAL treated mice are shown as grey and LPS treated mice as orange (**A**, **B**). Total distance travelled in the Open Field at P17 is shown for different treatment conditions at P3 and P9—SALCON, SALMS, LPSCON, LPSMS (**C**). Data is shown for female and male C57Bl/6 mice (mean ± S.E.)

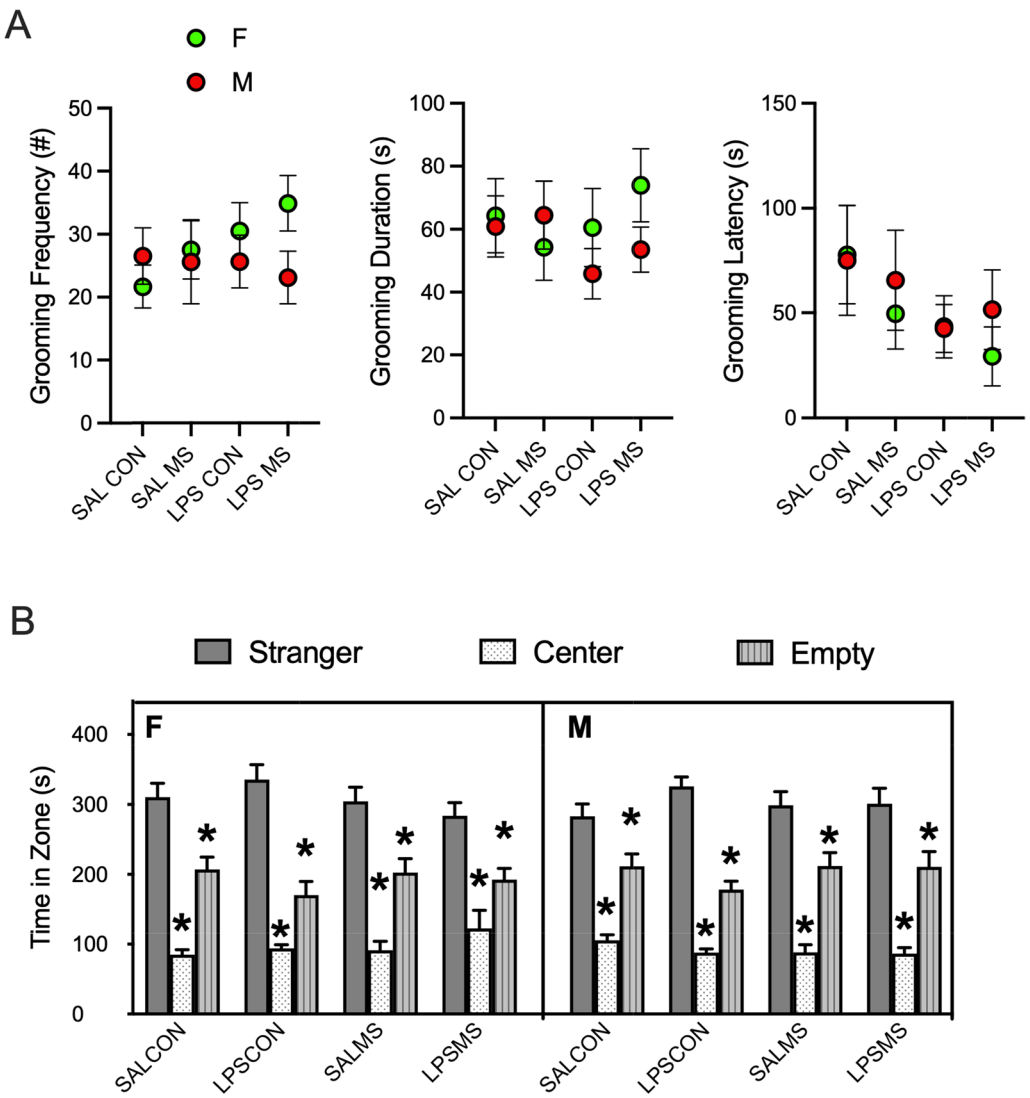


Fig. 3 Representative behavioural graphs for self grooming (**A**) and sociability (**B**). Data is shown for female and male C57Bl/6 mice (mean ± S.E.). Female (**B**-left panel) and male (**B**-right panel) mice showed typical sociability measured as social preference for chamber with stranger mouse. * $p < 0.05$, significantly different from stranger time

in the empty chamber according to posthoc analysis, however treatment effects were not detected for empty chamber time via between-subject ANOVA ($p = 0.119$).

Finally, grooming duration was highest in Balb/C mice, although CD1 mice also had high grooming duration relative to other strains (Table 3). While grooming frequency is highest in B6 mice, Balb/C also have higher grooming frequency compared to most other strains (Table 3). Balb/C mice have a grooming phenotype of longer frequent grooms, B6 mice have a phenotype of short and highly frequent grooms. For latency to first groom, CD1 have the highest metric compared to other wildtype mice (Table 3). The only knockout effect for grooming behaviours is decreased grooming duration in *TCRβ-/-δ-/-* compared to wildtype B6 ($p = 0.026$) (Table 3).

Unsupervised clustering of behaviors mapped mice to genotype

Inbred mice overall behavioural and neurodevelopmental phenotypes were visualized using PCA (Fig. 4A). Inbred mouse strains cluster based on neurodevelopmental and behavioural outcomes. We identified 3 clusters using unsupervised hierarchical clustering, whereby an unsupervised cluster predominantly maps to a single inbred strain (Fig. 4B). The structure of the clusters

Table 2 Multivariate ANOVA investigating association of genotype, sex, and treatment with behaviour

Test	Parameter	Effect	df	F	Sig
Righting reflex	Average righting reflex	Genotype	5	130.2	< 0.001
		Genotype * sex	5	3.8	0.002
USV	USV calls	Genotype	5	127.6	< 0.001
	USV duration	Genotype	5	115.3	< 0.001
		Genotype * treatment[1]	5	2.5	0.027
	USV ICI	Genotype	5	60.6	< 0.001
Eye opening	Eye opening Score	Genotype	5	30.0	< 0.001
Open field	OF rearing	Genotype	5	69.1	< 0.001
		Combined treatment	3	6.8	< 0.001
		Genotype * treatment	15	2.2	0.005
		Genotype * sex	5	3.8	0.002
	OF total distance	Genotype	5	86.5	< 0.001
		Combined treatment	3	4.5	0.004
		Genotype * treatment	15	2.8	< 0.001
		Genotype * sex	5	5.4	< 0.001
Sociabilty	Center chamber	Genotype	5	77.2	< 0.001
	Empty chamber	Genotype	5	3.9	0.002
	Mouse chamber	Genotype	5	31.9	< 0.001
		Sex * treatment	3	3.0	0.031
Grooming	Grooming duration	Genotype	5	43.3	< 0.001
	Grooming frequency	Genotype	5	29.9	< 0.001
		Genotype * sex	5	2.8	0.016
		Sex * treatment	3	3.5	0.015
	Grooming latency	Genotype	5	7.3	< 0.001
		Genotype * sex	5	3.3	0.006

[1] Treatment only refers to P3 treatment (LPS or SAL), as developmental milestone is assessed before P9 treatment (CON or MS)

demonstrates a robust relationship between genetic background and the overall behavioural and neurodevelopmental phenotype of mice.

When *FMR1-KO* mice are included for unsupervised hierarchical clustering, all *FMR1-KO* mice are members of the behavioural and neurodevelopmental cluster associated with their background strain, FVB (Supplementary Fig. S4). *FMR1-KO* mice do not cluster with the Balb/C or B6 dominant clusters. While differences exist between FVB and *FMR1-KO* mice in specific USV and OF metrics, their overall behavioural and neurodevelopmental phenotype retains higher similarity to FVB mice than other inbred strains in our study. However, the inclusion of *FMR1-KO* mice in the PCA plot decreases the effectiveness of hierarchical clustering discernment of FVB and B6 mice when there are three unsupervised clusters. Although the percent composition of each B6 and FVB predominant clusters remains similar in both analyses, 36 B6 mice shift to the FVB cluster when *FMR1-KO* mice are included. Changes in the B6 dominant cluster are an artefact of the sensitivity of hierarchical clustering to separating the overlap between the B6 and FVB behavioural phenotypes, whereby FVB mice with a more distant phenotype from the average mouse in the FVB clusters are assigned to the B6 cluster.

Similarly, when *TCRβ-/-δ-/-* mice are included in the PCA inbred plots, their behavioural and neurodevelopmental phenotype overlaps with their background strain B6 (Supplementary Fig. S4). They cluster primarily with B6 mice, with few *TCRβ-/-δ-/-* mice assigned to the FVB predominant cluster. There are more behavioural and neurodevelopmental differences of T-cell knockout mice compared to their background strain than *FMR1-KO* and FVB mice. Differences between B6 and *TCRβ-/-δ-/-* neurodevelopmental and behavioural metrics include: righting reflex, USV, OF, and grooming. Despite this, *TCRβ-/-δ-/-* mice still predominantly cluster with their background strain, albeit with lower cluster purity compared to the *FMR1*-KO and FVB mice cluster.

The addition of the outbred heterozygous strain CD1 disrupts the hierarchical clustering such that the B6 and FVB clusters are dissolved (Supplementary Fig. S5). The clustering pattern of CD1 mice suggests their behavioural and neurodevelopmental phenotype varies on a

Table 3 Behaviour univariate ANOVA with Bonferroni-corrected posthoc

Test	Parameter	Effect	df	F	Sig	Variable 1	Variable2	Abs Mean Difference	Mean Difference (I-J)	Std. Error	Sig
Righting reflex	Average righting reflex	Genotype	5	130.2	< 0.001	B6	TCR	9.29	− 9.29	0.68	< 0.001
						B6	FVB	5.95	5.95	0.56	< 0.001
						BalbC	FVB	5.45	5.45	0.56	< 0.001
						CD1	FVB	4.53	4.53	0.54	< 0.001
USV	USV calls	Genotype	5	127.6	< 0.001	BalbC	CD1	273.6	− 273.6	11.3	< 0.001
						BalbC	FVB	154.2	− 154.2	11.3	< 0.001
						B6	BalbC	147.6	147.6	11.7	< 0.001
						B6	CD1	126.1	− 126.1	11.2	< 0.001
						CD1	FVB	119.5	119.5	10.8	< 0.001
						B6	TCR	92.2	− 92.2	13.7	< 0.001
	USV duration	Genotype	5	115.3	< 0.001	B6	FVB	0.0337	− 0.0337	0.0019	< 0.001
						B6	CD1	0.0215	− 0.0215	0.0019	< 0.001
						BalbC	FVB	0.0179	− 0.0179	0.0019	< 0.001
						B6	BalbC	0.0157	− 0.0157	0.0020	< 0.001
						CD1	FVB	0.0122	− 0.0122	0.0019	< 0.001
						FMR1	FVB	0.0104	0.0104	0.0024	< 0.001
						BalbC	CD1	0.0058	− 0.0058	0.0019	0.045
	USV ICI	Genotype	5	60.6	< 0.001	BalbC	CD1	4.99	4.99	0.33	< 0.001
						BalbC	FVB	4.46	4.46	0.33	< 0.001
						B6	BalbC	4.13	− 4.13	0.34	< 0.001
Eye opening	Eye opening score	Genotype	5	30.0	< 0.001	B6	CD1	0.197	− 0.197	0.02	< 0.001
						BalbC	CD1	0.160	− 0.160	0.02	< 0.001
						B6	FVB	0.158	− 0.158	0.02	< 0.001
						BalbC	FVB	0.121	− 0.121	0.02	< 0.001
Open field	OF rearing	Genotype	5	69.1	< 0.001	BalbC	FVB	86.5	− 86.5	4.97	< 0.001
						BalbC	CD1	71.6	− 71.6	5.12	< 0.001
						B6	FVB	43.8	− 43.8	5.31	< 0.001
						B6	BalbC	42.7	42.7	5.49	< 0.001
						B6	CD1	28.9	− 28.9	5.45	< 0.001
						FMR1	FVB	28.7	− 28.7	6.15	< 0.001
						B6	TCR	19.9	19.9	6.35	0.027
						CD1	FVB	14.9	− 14.9	4.93	0.039
		Treatment	3	6.8	< 0.001	SAL CON	SAL MS	16.9	16.9	4.64	0.002
	OF total distance	Genotype	5	86.5	< 0.001	BalbC	CD1	696.3	− 696.3	38.25	< 0.001
						BalbC	FVB	663.0	− 663.0	37.13	< 0.001
						B6	BalbC	594.9	594.9	41.01	< 0.001
						B6	TCR	139.5	139.5	47.38	0.050
		Treatment	3	4.5	0.004	SAL CON	SAL MS	100.5	100.5	34.66	0.023
Sociabilty	Mouse chamber	Genotype	5	31.9	< 0.001	B6	BalbC	119.49	119.49	12.90	< 0.001
						BalbC	FVB	113.08	− 113.08	11.67	< 0.001
						BalbC	CD1	107.87	− 107.87	12.03	< 0.001
	Center chamber	Genotype	5	77.2	< 0.001	BalbC	CD1	160.70	160.70	9.64	< 0.001
						B6	BalbC	144.99	− 144.99	10.34	< 0.001
						BalbC	FVB	141.46	141.46	9.36	< 0.001
	Empty chamber	Genotype	5	3.9	0.002	BalbC	CD1	53.33	− 53.33	11.56	< 0.001
		Treatment	3	1.957	0.119	LPS CON	SAL CON	31.84	− 31.84	10.31	0.012

Table 3 (continued)

Test	Parameter	Effect	df	F	Sig	Variable 1	Variable2	Abs Mean Difference	Mean Difference (I-J)	Std. Error	Sig
Grooming	Grooming duration	Genotype	5	43.3	< 0.001	BalbC	FVB	55.95	55.95	4.86	< 0.001
						B6	BalbC	52.88	− 52.88	5.37	< 0.001
						BalbC	CD1	41.17	41.17	5.01	< 0.001
						B6	TCR	19.52	19.52	6.20	0.026
						CD1	FVB	14.78	14.78	4.82	0.033
	Grooming frequency	Genotype	5	29.9	< 0.001	B6	FVB	14.5	14.5	1.4	< 0.001
						B6	CD1	14.4	14.4	1.5	< 0.001
						BalbC	FVB	8.1	8.1	1.4	< 0.001
						BalbC	CD1	8.0	8.0	1.4	< 0.001
						B6	BalbC	6.4	6.4	1.5	< 0.001
	Grooming latency	Genotype	5	7.3	< 0.001	BalbC	CD1	55.17	− 55.17	9.27	< 0.001
						B6	CD1	45.26	− 45.26	9.87	< 0.001
						CD1	FVB	38.79	38.79	8.92	< 0.001

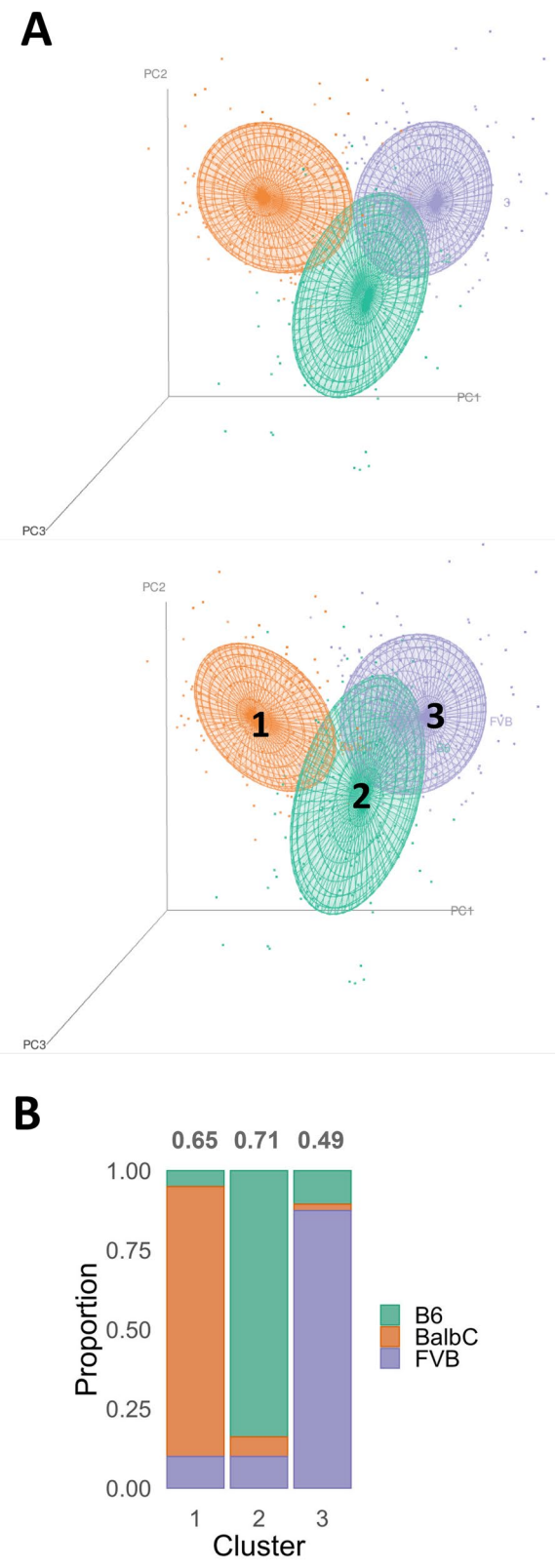
phenotypic spectrum that intersects with B6 and FVB mice, creating overlap that dissolves the mice into a single cluster. The remaining portion of the B6 predominant cluster begins to cluster with a portion of the Balb/C predominant cluster (Supplementary Fig. S5). These clustering patterns reflect that outbred heterozygous CD1 mice have heterogenous behavioural and neurodevelopmental phenotypes that span B6 and FVB trajectories.

The influence of sex and treatment on behavioural and neurodevelopmental clustering of mice was also assessed. There was no appreciable cluster separation of any sex-treatment subgroup (Supplementary Fig. S6).

Genotype was predicted using behavioral and neurodevelopmental milestons

Accurate prediction of genotype from behavioural and neurodevelopmental milestones further validates distinct phenotypes between genotypes and inbred strains of mice in our study. Random Forest models were trained to predict mouse genotype using sex, treatment, and behavioural and neurodevelopmental outcomes. The average genotype prediction accuracy of our model was 73% across 10 training/test/validation sets (Misclassification Error (ME) = 0.27 ± 0.02, Adjusted Rand Index (ARI) = 0.49 ± 0.03). 73% accuracy of machine learning models for predicting genotype from behaviour reinforces that the structure of behavioural and neurodevelopmental data differs between mouse genotypes and strains—indicative of unique phenotypes. A breakdown of prediction accuracy for each genotype is reported as a proportionate classification matrix (Fig. 5). As demonstrated in PCA and hierarchical clustering analysis, inbred strains can be accurately distinguished from one another for most observations (Figs. 3, 5A). However, behavioural and neurodevelopmental phenotypic overlap often occurs between the outbred CD1 mice, or knockout mice and their background strains (Fig. 5A, Fig S2). CD1 mice are most frequently misclassified as FVB or B6 mice, akin to the results when CD1 were included for hierarchical clustering (Fig. 5A). FVB misclassifications are predominantly misclassified as their associated genetic knockout—*FMR1-KO*—in 20% of cases, or as outbred CD1 strain in 14% of cases. B6 genotype misclassifications are predominantly CD1 or *TCRβ-/-δ-/-* genotypes. In accordance with observed overlap between knockout mice and their background strain in hierarchical clustering and PCA visualization method, *TCRβ-/-δ-/-* and *FMR1*-KO misclassification is almost solely attributable to misclassification as their wildtype background strains. Balb/C mice do not demonstrate appreciable misclassification as any other genotype, reaffirming their distinct behavioural and neurodevelopmental phenotype in our study (Fig. 5A).

Predictor variable importance and probability density function plots of behavioural and neurodevelopmental outcomes for each genotype reveals behavioural tests and neurodevelopmental milestones that are important for genotype discrimination. USV duration was the most important variable for genotype prediction across all genotypes (Fig. 5C) and had substantial probability density



◀ **Fig. 4** Principal component analysis of all neurodevelopment and behavioral metrics in inbred mice. Variance explained: PC1 = 25.4%, PC2 = 14.1% PC3 = 11.8%. **A** Top: PCA points coloured by genotype. Balb/C = Orange, FVB = Purple, B6 = Green. Bottom: PCA points coloured by 3 clusters generated by unsupervised hierarchical clustering (k = 3). Orange = Cluster 1 (Jaccard Index = 0.65), Green = Cluster 2 (Jaccard Index = 0.71), Purple = Cluster 3 (Jaccard Index = 0.49). **B** Behavioural and neurodevelopmental hierarchical cluster membership by mouse strain. Cluster stability is annotated above the bar graph for each cluster. Cluster stability was measured by the average Jaccard similarity index with the clusters of 1000 bootstrapped samples

function separation between most genotypes (Fig. 5B). Specifically, B6 background, FVB background and Balb/C or CD1 background demonstrate visual separation of USV duration density functions (Fig. 5B). Other important predictor variables include: average righting reflex, OF total distance, USV calls, and OF rearing (Fig. 5C). Visual separation of probability density functions for at least one genotype exists for most of these predictor variables (Fig. 5B).

Sex and treatment were also included as predictor variables for genotype prediction in our Random Forest models, however these variables were not important predictor variables for classification of mouse genotype (Fig. 5C). This supports our ANOVA and posthoc results, which did not demonstrate any independent treatment or sex effect, and minimal treatment interactions. In our dataset, sex and LPS and/or MS treatment have limited influence on early life behavioural and neurodevelopmental trajectories. To further validate the absence of sex and treatment differences in our dataset, random forest models were trained to predict sex or treatment from behavioural and neurodevelopmental outcomes, treatment or sex, and genotype data. Accuracy for sex prediction was approximately 55%, akin to random assignment of two labels (ME = 0.45 ± 0.04, ARI = 0.01 ± 0.02) (Supplementary Fig. S7). Similarly, accuracy for treatment prediction was 26%, akin to random assignment of four labels (ME = 0.74 ± 0.02, ARI = 0.001 ± 0.014) (Supplementary Fig. S8).

Genotype was predicted from neuroanatomical volumes

Mouse neuroanatomical volumes were highly associated with genotype in our study. PCA visualization of relative brain volumes from 69 regions in wildtype inbred mice suggests clustering by genotype (Fig. 6A). These visually observed clusters were validated using unsupervised hierarchical clustering (Fig. 6B). Three highly stable clusters each dominated by a single genotype emerged (Fig. 6C). Balb/C mice have a distinct neuroanatomical phenotype that forms a pure Balb/C unsupervised cluster

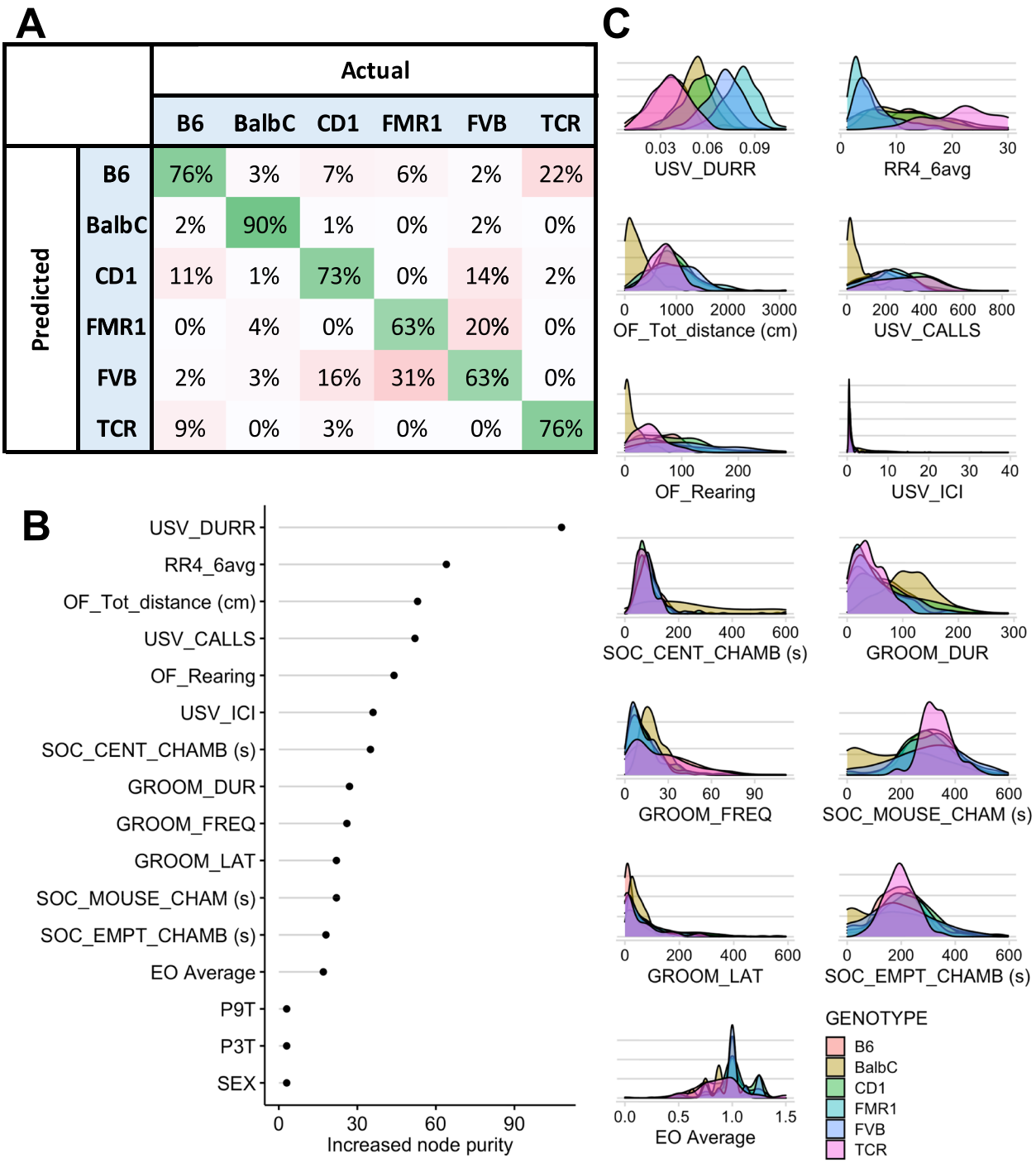
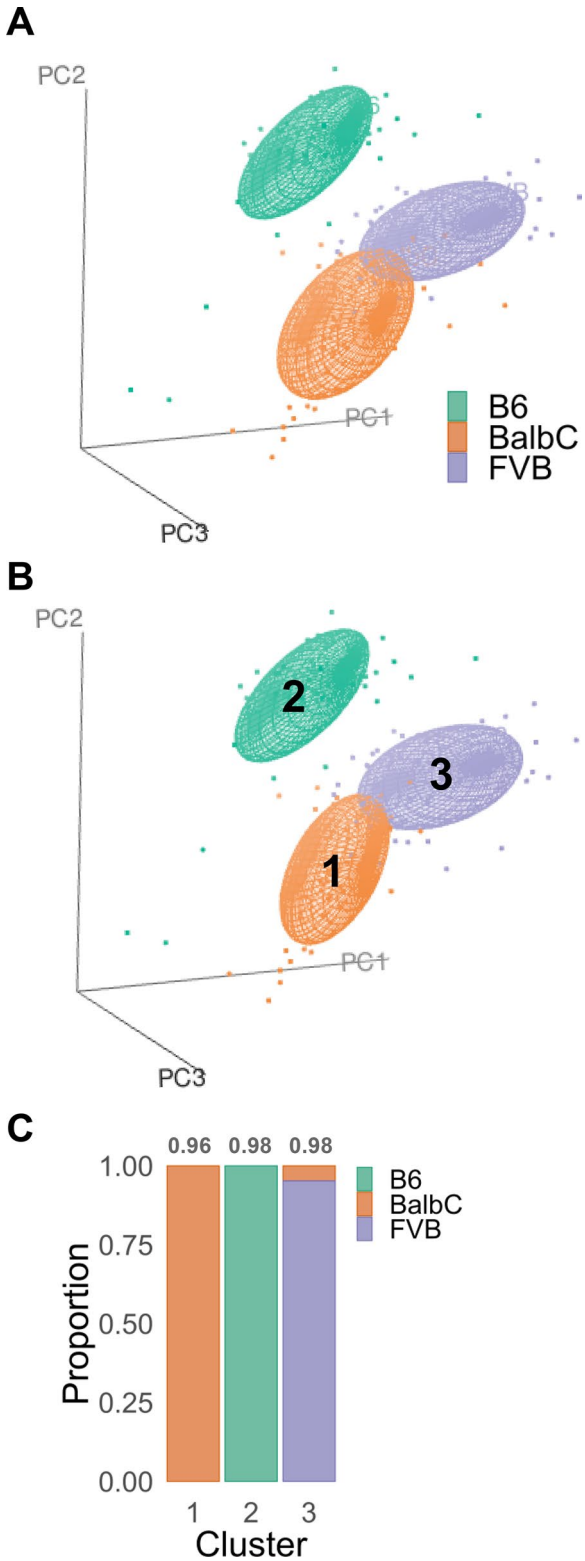


Fig. 5 Random Forest machine learning models accurately predict mouse genotype using behavioural and neurodevelopmental outcomes. Behavioural outcomes, neurodevelopmental outcomes, sex, and treatment (P3T and P9T) were input as predictor variables. **A** Confusion matrix represented as proportion of prediction from the total observations across 10 validation sets for each genotype. **B** Variable importance of predictor variables for genotype as measured by increased Gini index when included in the model. **C** Density plots of behavioural and neurodevelopmental outcomes for each mouse genotype P3T—postnatal day 3 treatment, P9T—postnatal day 9 treatment, EO—eye opening, SOC_EMP_CHAMB—time in empty chamber, SOC_CENT_CHAMB—time in the center chamber, SOC_MOUSE_CHAM—time in mouse chamber, DUR—duration, FREQ—frequency, LAT—latency, USV—ultrasonic vocalization, ICI—intercall interval, OF—open field, Tot—total, RR4—righting reflex postnatal day 4, DURR—duration



◀ **Fig. 6** Principal component analysis of MRI relative volumes in inbred mice. Brain regions included in analysis are available in Supplementary File 3 and Supplementary Fig. S3. Variance explained: PC1 = 23.2%, PC2 = 15.9% PC3 = 11.3%. **A** PCA, observations coloured by Genotype. **B** PCA points coloured by 3 clusters generated by unsupervised hierarchical clustering. Orange = Cluster 1, Green = Cluster 2, Purple = Cluster 3. **C** Bar graph describing the proportion of observations belonging to each Genotype in Clusters I, 2, and 3

(Fig. 6C). Furthermore, Balb/C mice do not contaminate the B6 nor FVB neuroanatomical volume clusters. There is some overlap between the B6 and FVB clusters, with some B6 genotype mice being misclassified as FVB mice based on their relative neuroanatomical volumes. These results demonstrates that the fingerprint of mouse neuroanatomical volumes across is highly distinct for inbred mice, but more similar between B6 and FVB mice than Balb/C mice. High genotype purity of clusters suggest that relative neuroanatomical volume is highly influenced by genotype.

The relationship between genotype and relative brain region volume is further supported by high accuracy of machine learning models predicting genotype from neuroanatomical volumes. Random Forest models trained using relative brain volume were 98% accurate at predicting all genotypes (ME = 0.02 ± 0.01, ARI = 0.96 ± 0.03). Misclassification was highest in FVB mice (ME = 0.05), which were most frequently misclassified as outbred CD1 mice (Fig. 7A). Inbred mice B6 and FVB and their knockout strains had very low misclassification error rates between ranging 1–2% (Fig. 6A), unlike the relatively higher 9–31% misclassification error when predicting wildtype strains from behavioural outcomes (Fig. 5A). The top 20 most important brain regions for genotype discrimination were visualized as density plots (Fig. 7B, C). The most important brain region for genotype discrimination was Olfactory Areas (Other)—of which *TCRβ-/-δ-/-* and B6 mice have distinct distributions from the other genotypes. The caudoputamen is the next most important brain region due to high relative caudoputamen volume in Balb/C mice compared to other genotypes (Fig. 7C). Overall, high genotype prediction accuracy of Random Forest models trained using relative brain volume indicates that genotype is an important factor related to individual brain region volumes.

Random forest models were also trained to predict sex and treatment from relative neuroanatomical volumes. Random forest models predicting sex had poor prediction accuracy, only marginally better than random assignment (ME = 0.33 ± 0.04, ARI = 0.12 ± 0.06) (Supplementary Fig. S9). The Bed Nuclei of the stria terminalis was the most important predictor variable for sex.

A

		Actual					
		B6	**BalbC**	**CD1**	**FMR1**	**FVB**	**TCR**
Predicted	**B6**	99%	0%	0%	0%	0%	0%
	BalbC	0%	99%	0%	0%	0%	0%
	CD1	0%	0%	97%	0%	2%	0%
	FMR1	0%	0%	1%	97%	1%	0%
	FVB	0%	1%	1%	3%	96%	0%
	TCR	1%	0%	1%	0%	0%	100%

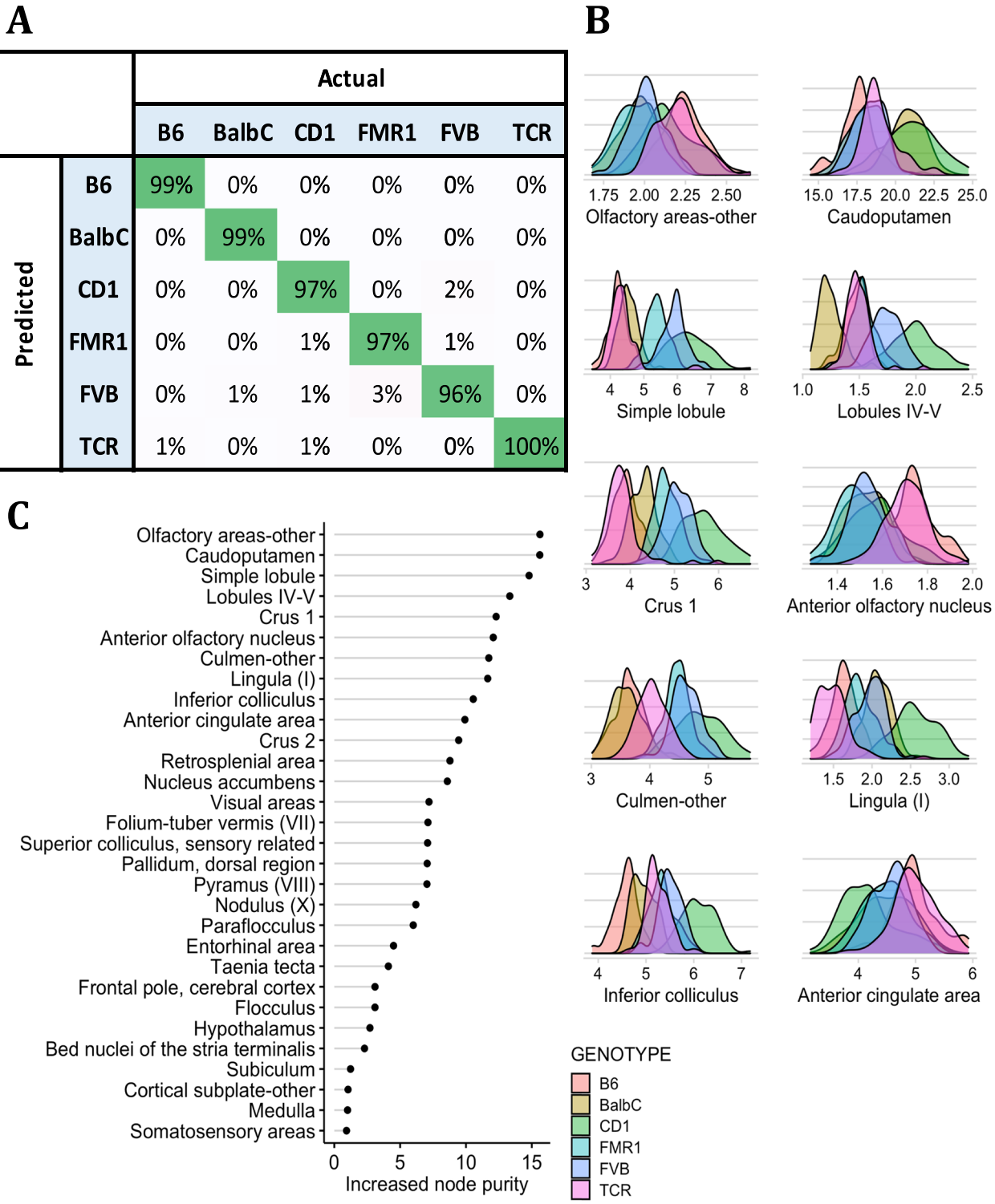


Fig. 7 Random Forest machine learning models accurately predict mouse genotype from relative neuroanatomical volumes. The brain regions used as predictor variables are listed in Supplemental File 3. **A** Confusion matrix represented as proportion of genotype predictions from the total number of each genotype across 10 validation sets. **B** Variable importance of predictor variables for genotype as measured by increased Gini index when included in the model. The top 20 ranked, random 5, and lowest 5 ranked predictor variables were plotted. **C** Density plots of brain regions in the top 20 of predictor variable importance; grouped by mouse genotype

Random forest models were unable to predict treatment from relative neuroanatomical volume; prediction was akin to random assignment (ME = 0.76 ± 0.04, ARI = 0.02 ± 0.03) (Supplementary Fig. S10).

Hierarchical clustering of behavioural data

Hierarchical clustering was used to investigate relationships between behavioural and neurodevelopmental metrics. Clustering was performed using behavioural and milestone outcomes from mice spanning all genotypes, treatments and both sexes. Two stable clusters of neurodevelopmental milestones and behaviours emerged (Fig. 8). Development metric for open field total distance was removed because of the high correlation and outcome redundancy with open field rearing. However, when open field was included in the analysis, the composition and stability the two clusters generated via hierarchical clusters were not different (Supplementary Table S1, S2). Similar behavioural networks with two highly stable clusters were identified when using inbred mice only (Supplementary Table S1, S2). Using 2 clusters (k = 2) was determined empirically using a Silhouette plot and the elbow method interpreted form a Within-Sum-Squares (Supplementary Fig. S11). The dendogram in Fig. 8 demonstrates the two stable clusters of behaviours and neurodevelopmental milestones. In one cluster the metrics related to grooming, righting reflex, sociability center time (movement) and intercall interval of USVs are linked. The other cluster links open field activity, eye opening, USV number and duration, and sociability. These associated networks of behaviours may indicate the neurocircuitry programming behaviours in each cluster are developmentally linked, and that mice may follow a neurodevelopmental trajectory that can be broadly characterized using two behavioural networks.

Discussion

We have developed the DBM Pipeline behavioural framework, a 28-day pipeline that tracks the neurodevelopmental and behavioural trajectory of mice. Using the DBM Pipeline, we assessed the influence of genetics, early-life challenge, and sex on the behavioural and neurodevelopmental trajectory for over 1060 mice. To our knowledge, our study is the first big data mouse

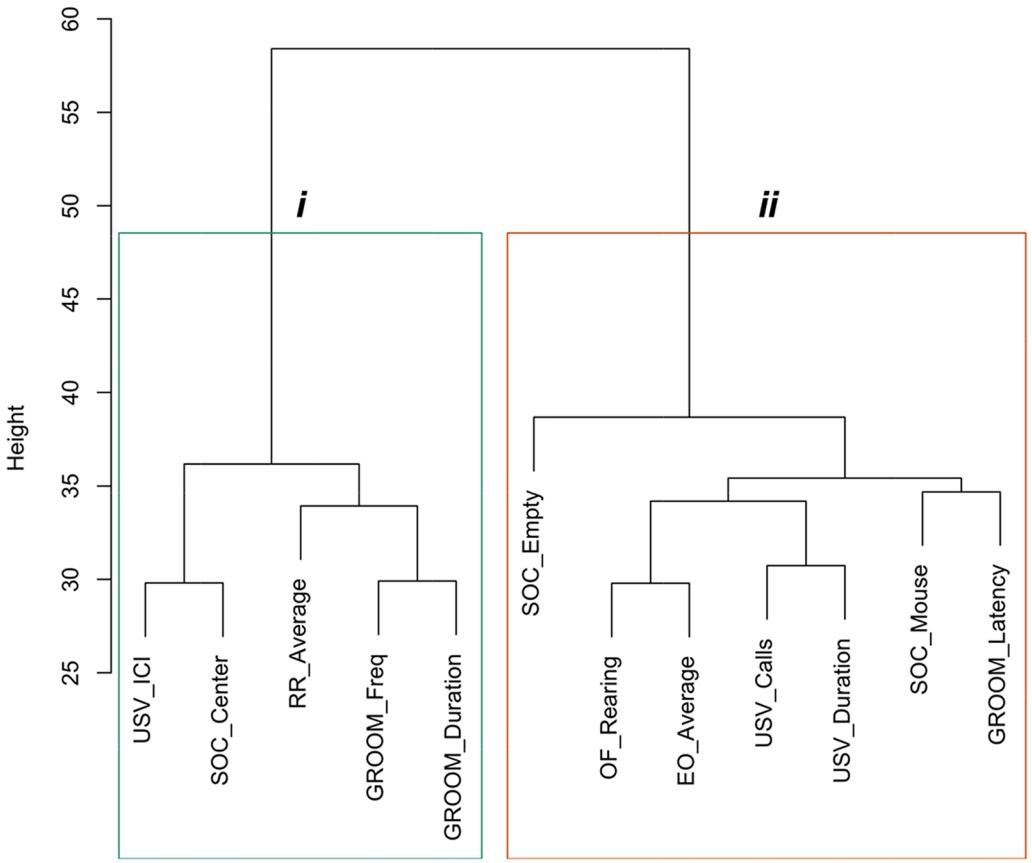


Fig. 8 Neurodevelopment and behavioral metrics from 871 clustered by Euclidian distances. Ward.D2 hierarchical clustering method was used to generate 2 clusters. Clusters were bootstrapped to validate cluster stability (Jaccard Mean = 0.972 (Cluster i), 0.986 (Cluster ii))

neurodevelopment project using clinically relevant developmental milestones and behavioural tests. While other behavioural neuroscience studies have used large sample sizes, the sample sizes are often diluted across experimental groups, preventing the use of statistical methods requiring large sample sizes for accurate analyses, such as hierarchical clustering and machine learning class predictions.

There are logistical challenges to curating a large-scale behavioural mouse dataset. Any big data mouse pipeline is vulnerable to issues of reproducibility and test validity. Each mouse in the data collection pipeline must undergo the same experimental tests at specific time points, as behavioural test results in mice are impacted by developmental age and the order of behavioural tests. Furthermore, errors arise from collecting data over an extended period, therefore robust quality control protocols ensuring data collected from daily assessments and tests have been accurately transferred to the database, and automated behavioural test equipment is used whenever possible. The 1000 Mouse DBM database addressed these challenges by using the standardized DBM pipeline and integrating robust quality control steps at every experimental stage—ensuring high quality mouse phenotyping data.

Our findings demonstrate the significant influence of background genetics on behaviour and neurodevelopmental trajectory. The main effects on neurodevelopmental milestones and behaviour tests identified in the between-subjects' effects were predominantly genotype differences. Posthoc analysis further validated broad behavioural strains differences, and select differences between KO mice and wild type counterparts. Inbred mouse strain behaviours and developmental milestones differences culminate into to a neurodevelopmental phenotype unique to each strain, which can be identified using supervised and unsupervised machine learning methods. Hierarchical Clustering validates visually observed inbred strains clusters. Random Forest models trained using neurodevelopmental and behavioural data accurately predict genotype, which further validates distinct neurodevelopmental and behavioural phenotypes. Machine learning model misclassifications typically occurred between knockout mice and their background strains, or inbred mice and outbred CD1 mice, demonstrating the phenotypic overlap between these genotypes. It is likely genetic diversity between individual outbred CD1 mice contributes to the heterogeneous CD1 neurodevelopmental phenotype, which varied between the FVB phenotype and a subset of the B6 neurodevelopmental phenotype in our study. While the relationship between behaviour and genetics in adult mice has already been robustly established, our study demonstrates how genetic background, and consequently behavioural outcomes, define the range of overall early-life behavioural and neurodevelopmental trajectory. Indeed, each mouse genotype could be broadly classified by a unique behavioural and neurodevelopmental phenotype. Notably, we observed no sex across any behavioural assays in our study in our study and limited treatment differences exclusive to sociability and open field tests. However, our study was limited to early-life behaviours and neuroanatomy observed up until post-natal day 28. Treatment and especially sex differences may be more profound after adolesence. Further, our study focuses on behaviours and milestones related to neurodevelopment. Our pipeline does not explore other behaviours related to cognitive function such as memory, learning, and executive function, which may have revealed additional treatment or sex differences. Knockout of key neurodevelopment-associated genes produce diverse behavioural and neurodevelopmental phenotypes dependent on the background strain [7, 54–59]. Indeed, hierarchical clustering predominantly grouped knockout mice with their background strain in the PCA analysis. Random Forest models trained to predict genotype further validated this relationship, as knockout mice were primarily misclassified as their own background strain. Although knockout mice differ from their background strain in select behavioural tests and neurodevelopmental milestones, multi-dimensional analyses using machine learning demonstrate that the phenotype of knockout mice broadly intersects with mice in their background strain. Our study validates the relevance of the background strain on behavioural phenotype in knockout mice models, as the background strain is the primary determinant of overall neurodevelopmental phenotype. Our results further reveal that the heterogeneous outbred CD1 strain exhibits a neurodevelopmental phenotype intermediate to common inbred mouse strains, potentially better reflecting the phenotypic variation found in human populations. Outbred strains may preferable to study the effect of genetic knockouts on neurodevelopment when clinical translation of results is important.

We also demonstrate robust neuroanatomical phenotypes delineated by mouse strain and knockout genotype. Random Forest models trained using relative brain region volume predict mouse strain and knockout genotype with 98% accuracy, suggesting that neuroanatomical volumes were largely driven by genotype. Additionally, both knockout strains can be differentiated from their background strain, indicating the FMR1 gene (*FMR1-KO* model) and T cells (*TCRβ-/-δ-/-* model) influence neuroanatomical volumes in adolescence. Significant differences between mouse strains' neuroanatomical volume and morphological variance have been previously

documented in several studies [60–62]. Additionally, quantitative trait loci that influence neuroanatomical volume have been identified, indicating the role of genetics [63, 64]. We build upon existing literature by using machine learning models to demonstrate that neuroanatomical volume differences between mouse genotypes are profound enough for near-perfect supervised and unsupervised classification.

Conversely, there is high purity when classifying early-life neuroanatomical volumes of knockout mice (*FMR1-KO* and *TCRβ-/-δ-/-* model) and their background strains (FVB and B6). These results suggest a disconnect between neuroanatomical differences and behavioural differences that may be driven by background genetics. Broadly, neuroanatomical diversity may be primarily dictated by genotype whereas behavioural phenotypic diversity may be more susceptible to gene-environment interactions. The observed genotypic stability of neuroanatomical phenotypes indicates genetics as a potential confounding variable when measuring neuroanatomical differences between populations.

Finally, we provide evidence for two highly stable networks of connected early-life behaviours and neurodevelopmental milestones. The existence of these networks implies that early developmental milestones may accurately predict behavioural phenotypes later in life related to behavioural activation and inhibition. We identified two networks of behavioural outcomes positively correlated to neurodevelopmental milestones using hierarchical clustering of behavioural parameter principal components. Cluster 1 can be summarized as behavioural inhibition. Righting reflex is the developmental milestones associated with Cluster 1, and behaviours include: grooming frequency and duration, 3-chambered sociability center time, and USV inter-call interval. The behaviours in the Cluster 1 are repetitive or represent low engagement for early life and adolescent social parameters. The cluster 1 milestone—righting reflex—also represents behavioural inhibition, as higher average righting reflex score indicates delayed motor development. Late onset of motor development in mice may be predictive of an inhibited behavioural phenotype later in life. Cluster 2 can be summarized as behavioural activation and is associated with the eye-opening milestone. The behaviours of Cluster 2 include: USV Calls and duration, 3-Chambered Sociability Mouse or Empty chamber time, Open-field parameter, and latency to groom. The behavioural readouts are proxies for high locomotor and social activity throughout early and adolescent life. Mice who achieve the eye-opening score earlier in development may be more likely to have an activated behavioural phenotype in adolescence. Indeed, early life behavioural milestones achievement have clinical associations with behavioural trajectory, in particular, for early detection of neurodevelopmental disorders in clinical populations [65–68]. The discovery of milestone-associated behavioural networks in our study indicates the external validity of our behavioural pipeline for modelling neurodevelopmental trajectories.

The 1000 Mouse DBM database containing the behavioural and neuroanatomical data from the 1060 mice phenotyped using the DBM Pipeline framework has been released as a publicly available data resource. It will also serve as both a data access portal for researchers to test or validate novel brain-behaviour hypotheses in the existing database. Furthermore, publicly releasing both a standardized animal experiment behavioural pipeline and a reference behavioural and neuroanatomical database will allow researchers to compare their own data for common mouse strains and estimate behavioural laboratory effects. Most importantly, researchers can leverage the 1000 Mouse DBM database's extensive sample size and reproducibility to design and conduct their own neuroanatomical and behavioral studies using the DBM pipeline. The 1000 Mouse database serves as a valuable reference point, enabling scientists to compare and validate their findings against a comprehensive reference dataset. Overall the integration of the DBM Pipeline and 1000 Mouse DBM database creates a framework for interventional neurodevelopmental pipeline and a reference database to support subsequent discoveries.

Conclusions

In summary, we developed the Developmental Behavioural Milestones pipeline, which was used to demonstrate the influence of genetic background on neurodevelopmental trajectory and neuroanatomical phenotype, as well as the relationship between neurodevelopmental milestones and behavioural outcomes. Our pipeline contextualizes the effect of knockouts targeting neurodevelopmental on behavioural outcomes that differ by strain. By applying hierarchical clustering to neurodevelopmental data of knockout mice and their background strain, the broad similarity of their neurodevelopmental trajectory can be assessed. Detection of robust genotype effects for behaviours and neuroanatomy, as well as early neurodevelopmental milestones being predictive of adolescent behavioural trajectory, indicates the external validity of the DBM framework. We have concurrently released the behavioural, neurodevelopmental and neuroanatomical data of the 1060 mice used in this study as the 1000 Mouse DBM database. Ultimately, the DBM pipeline and 1000 Mouse DBM database can be reproduced in other experimental designs to model how genotype, environmental insults, or drug interventions influence clinically relevant neurodevelopmental and

behavioural outcomes throughout early-life and adolescent development.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12993-024-00261-y.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.

Acknowledgements
Not applicable.

Author contributions
Conceptualization, JAF, JPL, JE. Methodology, JAF, JL, JE, SA, BCD. Behavioral Experiments, Analyses, Visualization, JAF, JKYL, KCR, ZH, SA. Imaging experiments, Analyses, Visualzation, JKYL, KCR, BCD, JE, JPL. Writing—Original Draft, SA, JE, JAF. Writing—Review and Editing, SA, JKYL, KCR, ZH, BCD, JE, JPL, JAF. Funding Acquisition, JAF, JPL.

Funding
This work was supported by an operating grant from the Ontario Brain Institute (OBIPOND JAF and JL).

Availability of data and materials
The 1000 Mouse DBM database containing the behavioural and neuroanatomical data from the 1060 mice phenotyped using the DBM Pipeline framework will be released as a publicly available data resource (https://www.braincode.ca/content/open-data-releases).

Declarations

Ethics approval and consent to participate
All experimental procedures were approved by the Animal Research Ethics Board, McMaster University in accordance with the guidelines of the Canadian Council on Animal Care.

Consent for publication
Not applicable.

Competing interests
JAF serves on the Scientific Advisory Board for MRM Health NL and has received consulting/speaker fees from Alphasights, Novozymes, Klaire Labs, Takeda Canada, Rothman, Benson, Hedges Inc, and WebMD. All other authors have no conflicts to report.

Received: 4 July 2024 Accepted: 4 December 2024
Published online: 02 January 2025

References

1. Foster JA, MacQueen G: Neurobiological factors linking personality traits and major depression. Can J Psychiatry 2008, 53(1):6–13.
2. Belay H, Burton CL, Lovic V, Meaney MJ, Sokolowski M, Fleming AS: Early adversity and serotonin transporter genotype interact with hippocampal glucocorticoid receptor mRNA expression, corticosterone, and behavior in adult male rats. Behav Neurosci 2011, 125(2):150–160.
3. Bilbo SD: Early-life infection is a vulnerability factor for aging-related glial alterations and cognitive decline. Neurobiol Learn Mem 2010, 94(1):57–64.
4. Fox WM. Reflex-ontogeny and behavioural development of the mouse. Anim Behav. 1965;13(2):234–41.
5. Hill JM, Lim MA, Stone MM. Developmental milestones in the newborn mouse. In: Gozes I, editor. Neuropeptide techniques. Totowa, NJ: Humana Press; 2008. p. 131–49. https://doi.org/10.1007/978-1-60327-099-1_10.
6. Ehret G. Infant rodent ultrasounds—a gate to the understanding of sound communication. Behav Genet. 2005;35(1):19–29.
7. Lai JK, Sobala-Drozdowski M, Zhou L, Doering LC, Faure PA, Foster JA. Temporal and spectral differences in the ultrasonic vocalizations of fragile X knock out mice during postnatal development. Behav Brain Res. 2014;259:119–30.
8. Scattoni ML, McFarlane HG, Zhodzishsky V, Caldwell HK, Young WS, Ricceri L, Crawley JN. Reduced ultrasonic vocalizations in vasopressin 1b knockout mice. Behav Brain Res. 2008;187(2):371–8.
9. Scudder CL, Karczmar AG, Lockett L. Behavioural developmental studies on four genera and several strains of mice. Anim Behav. 1967;15(2):353–63.
10. Roth A, Kyzar EJ, Cachat J, Stewart AM, Green J, Gaikwad S, O'Leary TP, Tabakoff B, Brown RE, Kalueff AV. Potential translational targets revealed by linking mouse grooming behavioral phenotypes to gene expression using public databases. Prog Neuropsychopharmacol Biol Psychiatry. 2013;40:312–25.
11. Brooks SP, Dunnett SB. Tests to assess motor phenotype in mice: a user's guide. Nat Rev Neurosci. 2009;10(7):519–29.
12. Moy SS, Nadler JJ, Young NB, Perez A, Holloway LP, Barbaro RP, Barbaro JR, Wilson LM, Threadgill DW, Lauder JM, et al. Mouse behavioral tasks relevant to autism: phenotypes of 10 inbred strains. Behav Brain Res. 2007;176(1):4–20.
13. Silverman JL, Yang M, Lord C, Crawley JN. Behavioural phenotyping assays for mouse models of autism. Nat Rev Neurosci. 2010;11(7):490–502.
14. Moy SS, Nadler JJ, Perez A, Barbaro RP, Johns JM, Magnuson TR, Piven J, Crawley JN. Sociability and preference for social novelty in five inbred strains: an approach to assess autistic-like behavior in mice. Genes Brain Behav. 2004;3(5):287–302.
15. Kalueff AV, Stewart AM, Song C, Berridge KC, Graybiel AM, Fentress JC. Neurobiology of rodent self-grooming and its value for translational neuroscience. Nat Rev Neurosci. 2016;17(1):45–59.
16. Bolivar VJ, Caldarone BJ, Reilly AA, Flaherty L. Habituation of activity in an open field: a survey of inbred strains and F1 hybrids. Behav Genet. 2000;30(4):285–93.
17. Bothe GW, Bolivar VJ, Vedder MJ, Geistfeld JG. Behavioral differences among fourteen inbred mouse strains commonly used as disease models. Comp Med. 2005;55(4):326–34.
18. Loos M, Koopmans B, Aarts E, Maroteaux G, van der Sluis S, Neuro BMPC, Verhage M, Smit AB. Sheltering behavior and locomotor activity in 11 genetically diverse common inbred mouse strains using home-cage monitoring. PLoS ONE. 2014;9(9): e108563.
19. Molenhuis RT, de Visser L, Bruining H, Kas MJ. Enhancing the value of psychiatric mouse models; differential expression of developmental behavioral and cognitive profiles in four inbred strains of mice. Eur Neuropsychopharmacol. 2014;24(6):945–54.
20. Tucci V, Lad HV, Parker A, Polley S, Brown SD, Nolan PM. Gene-environment interactions differentially affect mouse strain behavioral parameters. Mamm Genome. 2006;17(11):1113–20.
21. Kalueff AV, Tuohimaa P. Contrasting grooming phenotypes in C57Bl/6 and 129S1/SvImJ mice. Brain Res. 2004;1028(1):75–82.
22. Kalueff AV, Tuohimaa P. Contrasting grooming phenotypes in three mouse strains markedly different in anxiety and activity (129S1, BALB/c and NMRI). Behav Brain Res. 2005;160(1):1–10.
23. Dierssen M, Fotaki V, Martinez de Lagran M, Gratacos M, Arbones M, Fillat C, Estivill X. Neurobehavioral development of two mouse lines commonly used in transgenic studies. Pharmacol Biochem Behav. 2002;73(1):19–25.
24. Kelly MA, Low MJ, Phillips TJ, Wakeland EK, Yanagisawa M. The mapping of quantitative trait loci underlying strain differences in locomotor activity between 129S6 and C57BL/6J mice. Mamm Genome. 2003;14(10):692–702.
25. Delprato A, Algeo MP, Bonheur B, Bubier JA, Lu L, Williams RW, Chesler EJ, Crusio WE. QTL and systems genetics analysis of mouse grooming and behavioral responses to novelty in an open field. Genes Brain Behav. 2017;16(8):790–9.

26. Ashbrook DG, Roy S, Clifford BG, Riede T, Scattoni ML, Heck DH, Lu L, Williams RW. Born to cry: a genetic dissection of infant vocalization. Front Behav Neurosci. 2018;12:250.
27. Takahashi A, Tomihara K, Shiroishi T, Koide T. Genetic mapping of social interaction behavior in B6/MSM consomic mouse strains. Behav Genet. 2010;40(3):366–76.
28. Janecka M, Marzi SJ, Parsons MJ, Liu L, Paya-Cano JL, Smith RG, Fernandes C, Schalkwyk LC. Genetic polymorphisms and their association with brain and behavioural measures in heterogeneous stock mice. Sci Rep. 2017;7:41204.
29. Mombaerts P, Mizoguchi E, Ljunggren HG, Iacomini J, Ishikawa H, Wang L, Grusby MJ, Glimcher LH, Winn HJ, Bhan AK, et al. Peripheral lymphoid development and function in TCR mutant mice. Int Immunol. 1994;6(7):1061–70.
30. Rilett KC, Friedel M, Ellegood J, MacKenzie RN, Lerch JP, Foster JA. Loss of T cells influences sex differences in behavior and brain structure. Brain Behav Immun. 2015;46:249–60.
31. Nadler JJ, Moy SS, Dold G, Trang D, Simmons N, Perez A, Young NB, Barbaro RP, Piven J, Magnuson TR, et al. Automated apparatus for quantitation of social approach behaviors in mice. Genes Brain Behav. 2004;3(5):303–14.
32. Cahill LS, Laliberte CL, Ellegood J, Spring S, Gleave JA, Eede MC, Lerch JP, Henkelman RM. Preparation of fixed mouse brains for MRI. Neuroimage. 2012;60(2):933–9.
33. de Guzman AE, Wong MD, Gleave JA, Nieman BJ. Variations in post-perfusion immersion fixation and storage alter MRI measurements of mouse brain morphometry. Neuroimage. 2016;142:687–95.
34. Lerch JP, Sled JG, Henkelman RM. MRI phenotyping of genetically altered mice. Methods Mol Biol (Clifton, NJ). 2011;711:349–61.
35. Dazai J, Spring S, Cahill LS, Henkelman RM. Multiple-mouse neuroanatomical magnetic resonance imaging. J Vis Exp JoVE. 2011. https://doi.org/10.3791/2497-v.
36. Thomas DL, De Vita E, Roberts S, Turner R, Yousry TA, Ordidge RJ. High-resolution fast spin echo imaging of the human brain at 4.7 T: implementation and sequence characteristics. Magn Reson Med. 2004;51(6):1254–64.
37. Collins DL, Neelin P, Peters TM, Evans AC. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J Comput Assist Tomogr. 1994;18(2):192–205.
38. Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med Image Anal. 2008;12(1):26–41.
39. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage. 2011;54(3):2033–44.
40. Lerch JP, Carroll JB, Spring S, Bertram LN, Schwab C, Hayden MR, Henkelman RM. Automated deformation analysis in the YAC128 Huntington disease mouse model. Neuroimage. 2008;39(1):32–9.
41. Nieman BJ, Lerch JP, Bock NA, Chen XJ, Sled JG, Henkelman RM. Mouse behavioral mutants have neuroimaging abnormalities. Hum Brain Mapp. 2007;28(6):567–75.
42. Dorr AE, Lerch JP, Spring S, Kabani N, Henkelman RM. High resolution three-dimensional brain atlas using an average magnetic resonance image of 40 adult C57Bl/6J mice. Neuroimage. 2008;42(1):60–9.
43. Steadman PE, Ellegood J, Szulc KU, Turnbull DH, Joyner AL, Henkelman RM, Lerch JP. Genetic effects on cerebellar structure across mouse models of autism using a magnetic resonance imaging atlas. Autism Res. 2013. https://doi.org/10.1002/aur.1344.
44. Ullmann JF, Watson C, Janke AL, Kurniawan ND, Reutens DC. A segmentation protocol and MRI atlas of the C57BL/6J mouse neocortex. Neuroimage. 2013;78:196–203.
45. Genovese CR, Lazar NA, Nichols T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage. 2002;15(4):870–8.
46. Lerch JP, van der Kouwe AJ, Raznahan A, Paus T, Johansen-Berg H, Miller KL, Smith SM, Fischl B, Sotiropoulos SN. Studying neuroanatomy using MRI. Nat Neurosci. 2017;20(3):314–26.
47. R Core Team: R: a language and environment for statistical computing. In. Vienna, Austria: R Foundation for Statistical Computing; 2021.
48. Adler DM, D.: rgl: 3D visualization using OpenGL. In., 0.100.54 edn; 2020.
49. Fox JWS. An R companion to applied regression. 3rd ed. Thousand Oaks, CA: Sage Publications; 2019.
50. Kassambara AM, F.; : factoextra: extract and visualize the results of multivariate data analyses. In., 1.0.7 edn; 2020.
51. Henning C: fpc: Flexible procedures for clustering. In., 2.2–7 edn; 2020.
52. Liaw AW, M.: Classification and regression by randomForest. In: R News. vol. 2(3). Wien, Austria: Institut für Statistik und Wahrscheinlichkeitstheorie; 2002: 4.
53. Meyer DD, E.; Hornik, K.; Weingessel, A.; Leisch, F.: e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071). In.; 2020.
54. Born HA, Dao AT, Levine AT, Lee WL, Mehta NM, Mehra S, Weeber EJ, Anderson AE. Strain-dependence of the Angelman Syndrome phenotypes in Ube3a maternal deficiency mice. Sci Rep. 2017;7(1):8451.
55. Huang HS, Burns AJ, Nonneman RJ, Baker LK, Riddick NV, Nikolova VD, Riday TT, Yashiro K, Philpot BD, Moy SS. Behavioral deficits in an Angelman syndrome model: effects of genetic background and age. Behav Brain Res. 2013;243:79–90.
56. Lai JK, Lerch JP, Doering LC, Foster JA, Ellegood J. Regional brain volumes changes in adult male FMR1-KO mouse on the FVB strain. Neuroscience. 2016;318:12–21.
57. Pietropaolo S, Guilleminot A, Martin B, D'Amato FR, Crusio WE. Genetic-background modulation of core and variable autistic-like symptoms in Fmr1 knock-out mice. PLoS ONE. 2011;6(2): e17073.
58. Samaco RC, McGraw CM, Ward CS, Sun Y, Neul JL, Zoghbi HY. Female Mecp2(+/-) mice display robust behavioral deficits on two different genetic backgrounds providing a framework for pre-clinical studies. Hum Mol Genet. 2013;22(1):96–109.
59. Spencer CM, Alekseyenko O, Hamilton SM, Thomas AM, Serysheva E, Yuva-Paylor LA, Paylor R. Modifying behavioral phenotypes in Fmr1KO mice: genetic background differences reveal autistic-like responses. Autism Res. 2011;4(1):40–56.
60. Chen XJ, Kovacevic N, Lobaugh NJ, Sled JG, Henkelman RM, Henderson JT. Neuroanatomical differences between mouse strains as shown by high-resolution 3D MRI. Neuroimage. 2006;29(1):99–105.
61. Lin HY, Ni HC, Lai MC, Tseng WI, Gau SS. Regional brain volume differences between males with and without autism spectrum disorder are highly age-dependent. Mol Autism. 2015;6:29.
62. Scholz J, LaLiberte C, van Eede M, Lerch JP, Henkelman M. Variability of brain anatomy for three common mouse strains. Neuroimage. 2016;142:656–62.
63. Rosen GD, Williams RW. Complex trait analysis of the mouse striatum: independent QTLs modulate volume and neuron number. BMC Neurosci. 2001;2:5.
64. Zygourakis CC, Rosen GD. Quantitative trait loci modulate ventricular size in the mouse brain. J Comp Neurol. 2003;461(3):362–9.
65. Havdahl A, Farmer C, Schjolberg S, Oyen AS, Suren P, Reichborn-Kjennerud T, Magnus P, Bresnahan M, Hornig M, Susser E, et al. Age of walking and intellectual ability in autism spectrum disorder and other neurodevelopmental disorders: a population-based study. J Child Psychol Psychiatry. 2021;62(9):1070–8.
66. Johnson MH, Gliga T, Jones E, Charman T. Annual research review: Infant development, autism, and ADHD–early pathways to emerging disorders. J Child Psychol Psychiatry. 2015;56(3):228–47.
67. Provost B, Lopez BR, Heimerl S. A comparison of motor delays in young children: autism spectrum disorder, developmental delay, and developmental concerns. J Autism Dev Disord. 2007;37(2):321–8.
68. Gurevitz M, Geva R, Varon M, Leitner Y. Early markers in infants and toddlers for development of ADHD. J Atten Disord. 2014;18(1):14–22.

Publisher's Note