Supplemental text

Materials and Methods

General health check

Behavioral experiments were started one week after desflurane anesthesia. At the start of each test battery, each mouse was observed in a clean cage for a general health check (Supplemental table 1), followed by neurological screening. Posture was observed in the home cage (Crawley, 2007; Miyakawa et al., 2001; Takase et al., 2012).

Neurological screening test

Following the general health check, each mouse was challenged to evaluate their reflex responses (Crawley, 2007; Miyakawa et al., 2001; Takase et al., 2012). The neurological screening tests were designed to detect any gross abnormalities in physical function. The ear-twitch reflex was tested by approaching one ear from behind and touching the pinna, eliciting immediate movement of that ear. The eye-blink reflex was tested by approaching one eye with the covered end of a cotton-tipped applicator, eliciting immediate blinking by that eye. The

postural reflex was tested by shaking the cage, eliciting the extension of all four legs to maintain an upright, balanced position. The righting reflex was tested by turning the mouse over onto its back, eliciting an immediate turnover response to restore the upright posture of standing on all four paws. The whisker-touch reflex was tested by lightly touching one set of vibrissae, eliciting cessation of whisker movement and causing the mouse to turn its head ipsilaterally to the vibrissae that was touched.

Behavioral test batteries

Each behavioral test battery started when the mice were 8 weeks old (control group, n = 10 to 11/ a test battery; desflurane-treated group, n = 10/ a test battery) and ended by they were 9 weeks old. Our behavioral test batteries consisted of 7 test batteries as outlined in Supplementary Fig. 1, which minimize potential carry-over effects between behavioral tests (McIlwain *et al*., 2001; Paylor *et al*., 2006; Yonezaki *et al*., 2015). The experiments to examine the postanesthetic effects of desflurane were not performed all at once, since sufficient cage space is not available to breed large number of mice simultaneously. Therefore, a different control group was prepared

for the behavioral testing for desflurane-treated group. Behavioral testing, with the exception of the two bottle choice test, was performed between 9:00 AM and 5:00 PM. Before performing each test, the apparatus to be used was cleaned with 70% ethyl alcohol. The experimental rooms (Room A, 165 x 165 x 175 cm; Room B, 128 x 128 x 195 cm; Room C, 128 x 128 x 195 cm) were illuminated at 185 lx as measured at the floor of the room. The mice were habituated for 30 min before performing the behavioral tests in the room. The balance-beam test, attention tests, and contextual/cued fear conditioning test were performed in Room A. The anxiety tests, depression test, novel place/object recognition test, and social recognition test were performed in Room B. The sensory tests, motor tests (except the balance-beam test), and sociability tests were performed in Room C.

Sensory test battery

Visual placing test

Visual function was assessed using the visual placing test (Crawley, 2007; Heyser, 2003; Takase *et al*., 2012). This test involves holding the mouse by its tail approximately 30 cm above

a flat table surface. As the mouse is gradually lowered to the table, it extends its forepaws for a "soft landing". A blind mouse does not see the approaching surface and does not extend its forelimbs until its whiskers or nose touch the table. Extension of the forepaws was recorded as a yes or no response by the investigator.

Preyer reflex test

Auditory function was assessed by the Preyer reflex test (Crawley, 2007; Henry and Willott, 1972; Huang *et al*., 1995; Takase *et al*., 2012). The Preyer reflex is a flinch response to the sound of a loud hand clap. The reflex was recorded as a yes or no response by the investigator.

Von Frey hairs touch test

Tactile function was assessed using the Von Frey hairs touch test (Crawley, 2007; Fuchs *et al*., 1999; Pitcher *et al*., 1999). The mouse stood on an elevated platform with wide gauge wire mesh on the surface. The Von Frey hair (2.9 *N*) was inserted through the holes in the mesh from below to poke the undersurface of a hind paw. A normal response was defined as the mouse quickly flicking its paw away from the Von Frey hair. If the mouse showed the normal responses in two out of three consecutive trials, its tactile ability was recorded as normal.

Two bottle choice test

Gustatory function was assessed using the two bottle choice test (Crawley, 2007; Takase *et al*., 2012). The mouse was placed in a square aluminum cage (22 x 32 x 11 cm) and habituated for approximately 24 h. During this habituation period, the mouse was allowed to drink from two bottles containing water. After the habituation period, one of the two bottles was replaced with a bottle containing 5% sucrose and consumption monitored for 24 h as gram per gram of body weight. Standard pellet laboratory diet was provided *ad libitum* for the entire experimental period. The preference for sucrose was calculated using the following formula:

Preference (%) = sucrose consumption (g) / total consumption (g) x 100.

Olfactory habituation/dishabituation test

Olfaction was assessed by an olfactory habituation/dishabituation test (Silverman *et al*., 2010; Wrenn *et al*., 2003). The odorant stimuli were tap water, vanilla extract (Golden Kelly Pat. Flavor, Osaka, Japan) diluted 1:100 in tap water, and bitter almond extract (Golden Kelly Pat.

Flavor) diluted 1:100 in tap water. The odorant and dilution choices were based on data from pilot experiments in our laboratory. Experimenters presented stimuli by dipping a cotton-tipped applicator into the stimulus solution and then placing it through the wire lid of the cage. The applicator was stabilized 4.4 cm from the bottom of the cage. Each stimulus was presented for 3 min and then replaced by a new applicator three times in succession, for a total of nine presentations. The order of presentation was water (3x), vanilla (3x), and bitter almond (3x). An experimenter using a stopwatch recorded the cumulative time that the mouse spent sniffing the cotton-tipped applicator. Sniffing was defined as (a) tilting the head upward with the nose oriented toward and within 2 cm of the applicator, (b) rearing with the nose oriented toward and within 2 cm of the applicator, and (c) physically contacting the muzzle to the applicator if the mouth was closed. Occasional open-mouth contacts were considered to be chewing and not included in the cumulative sniff time.

Hot-plate test

The hot-plate test was performed by placing the mouse on a metal surface (19 cm in diameter, Muromachi Kikai, Tokyo, Japan) maintained at 54° ± 0.1°C. The hot plate was

surrounded by a transparent plastic barrier 20 cm in diameter and 25 cm in height. The latency to jumping off the plate or licking a hind paw was recorded. Sixty seconds was used as a cut-off time to protect the paw against injury (Blakeman et al., 2003; Crawley, 2007).

Motor test battery

Rotarod test

Motor coordination and balance were assessed using an accelerating rotarod (Muromachi Kikai) as described previously (Holmes *et al*., 2001; Paylor *et al*., 2006). Mice were placed on a cylinder that slowly accelerated from 4 to 40 rpm the latency to fall recorded for a maximum of 300 s. Each mouse performed three trials.

Balance-beam test

Motor coordination and balance were also assessed by measuring the ability of the mice to traverse a graded series of narrow beams to reach an enclosed safety platform (Carter *et al*., 1999). The beams consisted of long square or round strips of metal (1 m in length) with a cross-section of 5-, 12-, or 28-mm or diameter of 11, 17, or 28 mm, respectively. The beams were

placed horizontally 50 cm above the bench surface, with one end mounted on a narrow support and the other end attached to an enclosed box (20 cm square) into which the mouse could escape. Lights (1200 lx) were positioned above and to one side of the start of the beam. During training (1 day), mice were placed at the start of the 12-mm square beam and trained for 6 trials to traverse the beam to the enclosed box. Twenty four hours after the training, the mouse performed one trial on each of the square beams and each of the round beams, progressing from the narrowest to the widest. Mice were allowed up to 60 s to traverse each beam. The latency to traverse each beam and the number of times the hind feet slipped off were recorded for each trial.

Hanging wire test

Motor function was assessed using the wire hang test, which requires balance and grip strength (Crawley, 2007; Sango *et al*., 1999; Takase *et al*., 2012). A standard wire cage lid was used for this test. Masking tape placed around the perimeter of the lid prevented the mouse from walking off the edge. The test was performed by placing the mouse on the top of a wire cage lid. The investigator shook the lid lightly three times to cause the mouse to grip the wires and then turned the lid upside down. The upside-down lid was held approximately 40 cm above the cage

litter, high enough to prevent the mouse from easily climbing down but not high enough to cause harm in the event of a fall. Each mouse performed two trials. The investigator used a stopwatch to time the latency in falling off the wire lid with a maximum of 60 s.

Anxiety test battery

Elevated plus-maze test

The elevated plus-maze test was conducted as described previously (Holmes *et al*., 2002a,b; Holmes *et al*., 2003). The apparatus consisted of two open arms (29.7 x 5.4 cm) and two closed arms (30 x 6 x 15 cm) extending from a common central platform (6 x 6 cm). A small raised lip (0.3 cm) around the perimeter of the open arms prevented the mouse from falling. The apparatus was constructed from polypropylene, with gray floor and gray walls, and elevated 40 cm above the floor. Mice were placed individually on the center square facing an open arm and allowed to freely explore the apparatus under overhead fluorescent lighting (200 lx) for 5 min. Time spent in the open arms and open and closed arm entry (all four paws in an arm) were scored by a highly trained observer using behavioral scoring software (ANY-maze, Stoelting, IL, USA).

Open field test

The open field test was conducted as described previously (Paylor *et al*., 1998). Each subject was placed in the center of a clear Plexiglas (50 x 50 x 40 cm) chamber in standard room-lighting conditions (70 lx). Activity in the open field was quantitated by TimeOFCR4 (O'hara, Tokyo, Japan). Horizontal activity in the center (30 x 30 cm) or peripheral area, total distance (cm), and time in the center or periphery were recorded in seconds. The center distance was also divided by the total distance to obtain a center distance/total distance ratio, which can be used as an index of anxiety. Data were collected during 10 min test sessions.

Light-dark exploration test

The light-dark exploration test was performed as described previously (Holmes *et al*., 2002a,b; Holmes *et al*., 2003; Mathis *et al*., 1994). The apparatus consisted of a polypropylene cage (44 x 21 x 21 cm, Muromachi Kikai) separated into two compartments by a partition with a small aperture (12 x 5 cm) at floor level. The larger compartment (28 cm long) was open-topped, transparent, and brightly illuminated by white light from a 40 W desk lamp (1000 lx). The

smaller compartment (14 cm long) was close-topped and painted black. Mice were placed individually in the center of the light compartment, facing away from the partition, and allowed to freely explore the apparatus for 10 min. The number of light-dark transitions between the two compartments and the total time spent in the dark compartment were automatically recorded by ANY-maze.

Depression test battery

Porsolt forced swim test

Antidepressant activity was assessed by the Porsolt forced swim test. For swim sessions, mice were placed in individual glass cylinders (24.5 cm tall, 19 cm in diameter) filled with water (23° to 25°C water) to a depth of 15 cm. The depth was deep enough so that mice could not support themselves by placing their paws on the base of the cylinder. The procedure was similar to that of Porsolt *et al*. (1977). Behavioral scoring employed a standard 6-min test duration (Borsini and Meli, 1988). The water was changed between subjects. All test sessions were recorded, and the duration of immobility during the last 4 min of the test period and the latency

to first immobility during the 6 min test period automatically quantified by ANY-maze. A mouse was judged to be immobile when making only those movements necessary to keep its head above water.

Tail suspension test

Antidepressant activity was also assessed by the tail suspension test as described previously (Cryan and Mombereau, 2004; Holmes *et al*., 2002a,b). Mice were securely fastened to a flat metallic surface by the tip of the tail (2 to 3 cm) using medical adhesive tape and suspended 30 cm above the ground in a 40 cm^3 white Plexiglas box that isolated the mouse from visual distractions but permitted observation of behavior from above. The latency to first immobility, defined as the absence of limb movement, and the time of immobility was sampled by ANY-maze.

Sociability test battery

Social interaction test

Social behavior was examined using the conventional social interaction test (Crawley, 2007; Silverman *et al*., 2010; Yonezaki *et al*., 2015). Each subject was placed in the center of a cubical box (30 x 36 x 17 cm) in standard room-lighting conditions. A naive same-sex mouse was then introduced into the cubical box and allowed to explore freely for 5 min. The interaction frequency and total duration of the interaction were recorded. Social interaction comprised sniffing, grooming, exploring, following, kicking, mounting, jumping on, wrestling, and other forms of physical contact.

Tube test

The tube test assay was adapted from Lindzey *et al*. (1961). The test employed a transparent Plexiglas tube 30 cm in length with a 3 cm inside diameter, which is sufficient to permit one adult mouse to pass through without reversing direction. For training, each mouse was released at alternating ends of the tube to run through the tube, sometimes with the help of a plastic stick pushing at its back. Each animal was given eight training trials per day on two successive days. On the third day, the mice were given the test trial. During the test trial, two mice were released simultaneously into opposite ends, and care was taken to ensure that they met in the middle of

the tube. The mouse that first retreated from the tube within 2 min was designated the “loser” of that trial. In rare cases when no mouse retreated within 2 min, the tests were repeated. Between each trial, the tube was cleaned with 70% ethanol. Between the control and desflurane-treated groups, paired encounters were staged based on body weight.

Attention test battery

Prepulse inhibition test

The prepulse inhibition test was performed as described by Miyakawa *et al*. (2003) using a startle reflex measurement system (O’hara). A test session began by placing a mouse in a Plexiglas cylinder, where it was left undisturbed for 5 min. The duration of white noise used as the startle stimulus was 40 ms for all trial types. The startle response was recorded for 160 ms (measuring the response every 1 ms), starting with the onset of the prepulse stimulus. The background noise level in each chamber was 70 dB. The peak startle amplitude recorded during the 160-ms sampling window was used as the dependent variable. A test session consisted of six trial types (i.e., two types for startle stimulus-only trials, and four types for prepulse inhibition trials). The intensity of the startle stimulus was 110 or 120 dB. The prepulse sound was presented

100 ms before the startle stimulus, and its intensity was 74 or 78 dB. Four combinations of prepulse and startle stimuli were employed (74-110, 78-110, 74-120, and 78-120). Six blocks of the six trial types (four trial types with the combinations of prepulse and startle stimulus and two startle stimulus-only trials) were presented in pseudorandom order, such that each trial type was presented once within a block. The average intertrial interval was 15 s (range, 10 to 20 s). The percentage of prepulse inhibition was calculated as follows: ((startle response for pulse alone - startle response for pulse with prepulse) / startle response for pulse alone) x 100.

Latent inhibition test

This experiment was performed similar to that reported by Miyakawa *et al*. (2003). On the first day, each mouse was placed in a conditioning chamber (Muromachi Kikai). The mice were divided into two groups: pre-exposed (P) and non-pre-exposed (NP). The P group was exposed to 40 white noise (68 dB, 5-s duration, 25-s interstimulus interval), and the NP group was exposed to no stimulus during an equivalent period. Immediately after the pre-exposure period, white noise -shock pairs consisting of a 5-s white noise co-terminating with a 2-s foot shock at 0.4 mA were delivered to both groups with a 25-s interstimulus interval. Afterward, mice

remained in the chamber for 25 s before being returned to the home cage. On day 2, the mice were placed in the conditioning chamber for 5 min to measure freezing to the context. On day 3, the mice were put in a white Plexiglas chamber scented with vanilla essence, and after 180 s, a 180-s white noise was delivered to measure cued freezing.

Learning test battery

Novel place/object recognition test

The novel place/object recognition test was performed using a cubical box (30 x 36 x 17 cm) made of clear Plexiglas (Ng *et al*., 2009). All behavioral events were scored and analyzed by a highly trained observer. The test consisted of three sessions with intertrial intervals of 2 min, during which mice were returned to their home cages. During the habituation session, four different plastic objects were presented in the cubical box. Exploration of the four different plastic objects in the cubical box was measured every 5 min for 15 min under dim lighting (4 lx). In the place recognition session, the four objects initially placed in a square arrangement were reconfigured into a polygon-shaped pattern by moving two objects (displaced objects, DOs). The

remaining two objects were left at the same location (non-displaced objects, NDOs). The time spent exploring the DOs and NDOs was recorded for 5 min and expressed as a percentage of the total time for object investigation. In the object recognition session, one of the NDOs was replaced with a new object (NO) at the same location, and the two DOs were removed. The time examining the NO or familiar NDO (FO) was recorded for 5 min and expressed as a percentage of the total time for object investigation.

Social recognition test

Recognition memory about other individuals was assessed by a social recognition test (Crawley, 2007; Silverman *et al*., 2010; Yonezaki *et al*., 2015). For training, mice were placed in a cubical box (30 x 36 x 17 cm) with a naïve mouse for 10 min. One hour later, the mouse was returned to the cubical box and exposed to the same mouse and a novel mouse for 10 min. The time spent exploring each mouse was recorded during both the training and test phases. The percentage of novelty preference was calculated follows: (novel individual exploration duration) / (novel individual exploration duration + familiar individual exploration duration) x 100.

Cued/contextual fear conditioning test

This cued/contextual fear conditioning test was performed as described by Miyakawa *et al*. (2003). On the first day, each mouse was placed in a conditioning chamber (Muromachi Kikai). The mice were habituated to the chamber for 1200 s. Immediately after exposure to the chamber, white noise-shock pairs consisting of a 5-s white noise co-terminating with a 2-s foot shock at 0.4 mA were delivered with a 25-s interstimulus interval. Afterward, mice remained in the chamber for 25 s before being returned to their home cage. On day 2, the mice were placed back in the conditioning chamber for 5 min to measure freezing to the context. On day 3, the mice were put in a white Plexiglas chamber scented with vanilla essence, and after 180 s, a 180-s white noise was delivered to measure cued freezing.

References

Borsini F, Meli A (1988). Is the forced swimming test a suitable model for revealing antidepressant activity? *Psychopharmacology (Berl)* **94**: 147-60.

Carter RJ, Lione LA, Humby T, Mangiarini L, Mahal A, Bates GP, et al. (1999). Characterization of progressive motor deficits in mice transgenic for the human Huntington's disease mutation. *J Neurosci* **19**: 3248-57.

Crawley JN (2007). *What's Wrong with My Mouse? Behavioral Phenotyping of Transgenic and Knockout Mice.* New York: Wiley-Liss.

Cryan JF, Mombereau C (2004). In search of a depressed mouse: utility of models for studying depression-related behavior in genetically modified mice. *Mol Psychiatry* **9**: 326-57.

Fuchs PN, Roza C, Sora I, Uhl G, Raja SN (1999). Characterization of mechanical withdrawal responses and effects of mu-, delta- and kappa-opioid agonists in normal and mu-opioid receptor knockout mice. *Brain Res* **821**: 480-6.

Henry KR, Willott JF (1972). Unilateral inhibition of audiogenic seizures and Preyer reflexes. *Nature* **240**: 481-2.

Heyser CJ (2003). Assessment of developmental milestones in rodents. In: *Current Protocols in Neuroscience*. New York: Wiley-Liss; pp 8.18.1-15.

Holmes A, Hollon TR, Gleason TC, Liu Z, Dreiling J, Sibley DR, et al. (2001). Behavioral characterization of dopamine D5 receptor null mutant mice. *Behav Neurosci* **115**: 1129-44.

Holmes A, Kinney JW, Wrenn CC, Li Q, Yang RJ, Ma L, et al. (2003). Galanin GAL-R1 receptor null mutant mice display increased anxiety-like behavior specific to the elevated plus maze. *Neuropsychopharmacology* **28**: 1031-44.

Holmes A, Wrenn CC, Harris AP, Thayer KE, Crawley JN (2002a) Behavioral profiles of inbred strains on novel olfactory, spatial and emotional tests for reference memory in mice. *Genes Brain Behav* **1**: 55-69.

Holmes A, Yang RJ, Crawley JN (2002b) Evaluation of an anxiety-related phenotype in galanin overexpressing transgenic mice. *J Mol Neurosci* **18**: 151-65.

Huang JM, Money MK, Berlin CI, Keats BJ (1995). Auditory phenotyping of heterozygous sound-responsive (+/dn) and deafness (dn/dn) mice. *Hear Res* **88**: 61-4.

Lindzey G, Winston H, Manosevitz M (1961). Social dominance in inbred mouse strains. *Nature* **191**: 474-6.

Mathis C, Paul SM, Crawley JN (1994). Characterization of benzodiazepine-sensitive behaviors in the A/J and C57BL/6J inbred strains of mice. *Behav Genet* **24**: 171-80.

McIlwain KL, Merriweather MY, Yuva-Paylor LA, Paylor R (2001). The use of behavioral test batteries: effects of training history. *Physiol Behav* **73**: 705-17.

Miyakawa T, Leiter LM, Gerber DJ, Gainetdinov RR, Sotnikova TD, Zeng H, et al. (2003). Conditional calcineurin knockout mice exhibit multiple abnormal behaviors related to schizophrenia. *Proc Natl Acad Sci USA* **100**: 8987-92.

Miyakawa T, Yared E, Pak JH, Huang FL, Huang KP, Crawley JN (2001). Neurogranin null mutant mice display performance deficits on spatial learning tasks with anxiety related components. *Hippocampus* **11**: 763-75.

Ng D, Pitcher GM, Szilard RK, Sertié A, Kanisek M, Clapcote SJ, et al. (2009). Neto1 is a novel CUB-domain NMDA receptor-interacting protein required for synaptic plasticity and learning. *PLoS Biol* **7**: e41.

1 Paylor R, Nguyen M, Crawley JN, Patrick J, Beaudet A, Orr-Urtreger A (1998). Alpha7 nicotinic
2 receptor subunits are not necessary for hippocampal-dependent learning or sensorimotor
3 gating: a behavioral characterization of Acra7-deficient mice. *Learn Mem* **5**: 302-16.
4 Paylor R, Spencer CM, Yuva-Paylor LA, Pieke-Dahl S (2006) The use of behavioral test
5 batteries, II: effect of test interval. *Physiol Behav* **87**: 95-102.
6 Pitcher GM, Ritchie J, Henry JL (1999). Paw withdrawal threshold in the von Frey hair test is
7 influenced by the surface on which the rat stands. *J Neurosci Methods* **87**: 185-93.
8 Porsolt RD, Le Pichon M, Jalfre M (1977). Depression: a new animal model sensitive to
9 antidepressant treatments. *Nature* **266**: 730-2.
10 Sango K, McDonald MP, Crawley JN, Mack ML, Tifft CJ, Skop E,et al. (1996). Mice lacking
11 both subunits of lysosomal beta-hexosaminidase display gangliosidosis and
12 mucopolysaccharidosis. *Nat Genet* **14**: 348-52.
13 Silverman JL, Yang M, Lord C, Crawley JN (2010). Behavioural phenotyping assays for mouse
14 models of autism. *Nat Rev Neurosci* **11**: 490-502.

Takase K, Yamamoto Y, Yagami T (2012). Maternal deprivation in the middle of a stress hyporesponsive period decreases hippocampal calcineurin expression and causes abnormal social and cognitive behaviours in adult male Wistar rats: relevance to negative symptoms of schizophrenia. *Behav Brain Res* **232**: 306-15.

Wrenn CC, Harris AP, Saavedra MC, Crawley JN (2003). Social transmission of food preference in mice: methodology and application to galanin-overexpressing transgenic mice. *Behav Neurosci* **117**: 21-31.

Yonezaki K, Uchimoto K, Miyazaki T, Asakura A, Kobayashi A, Takase K, et al. (2015). Postanesthetic effects of isoflurane on behavioral phenotypes of adult male C57BL/6J mice. *PLoS One* **10**: e0122118.

Supplementary Table 1. The items in the general health check

Bald patch

Body size

Body weight

Crustiness around the nostrils/ eyes

Ear pinna/ Footpad color

Fur

Gait

Lesions on the feet/ tail

Posture

Scabs on the tail, rump, back

Tumor

Whisker
