LabAnimal

ARTICLES

https://doi.org/10.1038/s41684-019-0288-8

# Experiments done in Black-6 mice: what does it mean?

Johanna Åhlgren[1] and Vootele Voikar[1,2*]

**Low replicability of animal experiments is perceived as a major hurdle in the field of biomedicine. Attempts to enhance the replicability and to reduce the variability in basic research has led to the recommendation to use isogenic mice. The C57BL/6 strain has evolved as a gold standard strain for this purpose. However, C57BL/6 mice are maintained as substrains by multiple vendors. Evidence exists that the subtle differences between these mouse lines have not been systematically investigated and are often ignored. In the present study, we characterized the female mice of two closely related substrains (C57BL/6J and C57BL/6N) from three vendors in Europe (Charles River Laboratories, Envigo, Janvier Labs) in a battery of behavioral tests. Our data show and confirm substantial behavioral differences between the C57BL/6J and C57BL/6N mice. Importantly, the substrain differences were largely affected by the origin of the animals, as a significant effect of vendor or interaction between the substrain and vendor occurred in all tests. This work highlights the importance of adhering to precise international nomenclature in all publications reporting animal experiments. Moreover, the generalization of research findings from a single mouse substrain can be seriously limited due to genetic drift and environmental variables occurring at different vendors. However, heterogenization of samples, by including animals of different substrains, can enhance generalizability. These issues need to be seriously addressed to improve reproducibility, replicability, and the translational potential of the mouse models.**

Over the past 25 years, thousands of knockout mice have been developed worldwide, and the mouse is increasingly used as a model of choice for investigating the genetic basis of diseases and potential drug targets, with refined methodology and techniques[1]. Conventional knockout mice were created by gene targeting in embryonic stem cells from the 129-mouse strain, followed by backcrossing to the C57BL/6 strain. However, problems with mixed genetic background were soon identified[2,3], leading to the publication of recommendations for controlling the genetic background of mutant mice[4]. It is interesting to note that although the importance of nomenclature was emphasized, no particular attention was paid to the existence of different substrains of C57BL/6 mice at that time. The development of the C57BL/6 strain goes back to 1921 when the strain was created by Clarence Cook Little and initially maintained at the Jackson Laboratory (C57BL/6J). The substrain C57BL/6N was established in 1951 after transfer of the mice to the NIH. Mice from both parental colonies have then been moved to several large mouse vendors over the world (such as Charles River Laboratories, Envigo, formerly known as Harlan, Taconic, and Janvier Labs), where they are still maintained. The C57BL/6J (B6J) mouse was the first strain to have its genomic sequence published, and this strain is considered a gold standard in many research areas. However, to overcome the problems associated with a mixed genetic background and to facilitate the production of mutant mice, embryonic stem cell lines from the C57BL/6N (B6N) strain were established; the International Mouse Phenotyping Consortium is currently creating mutant mice for large-scale phenotyping in a C57BL/6N background[5–8].

It is well known that the phenotype of mutant mice can depend on their genetic background[9] and this holds true also for the C57BL/6 substrains[10]. Although genotypic and phenotypic differences between the substrains of C57BL/6 mice are well documented (Table 1)[11–13], the fact is that too many publications do not indicate the precise and accurate origin of the animals used[14]. Moreover, it may well be that the researchers are unaware or ignore this information. For instance, according to a recent survey carried out in Finnish research institutions, 39.5% of respondents were either not aware of genetic differences between these substrains or did not consider it important. Interestingly, among those who knew about these differences, still 26% of the respondents were not able to name the exact strain they were using. Among others, this factor can certainly be one of the major issues contributing to the current reproducibility crisis in basic research[15,16]. Moreover, concerns have been expressed that the low quality of basic and preclinical studies, not only in behavioral studies[17], may have a direct relationship with the failures in clinical trials[18,19].

For a long time, it has been suggested that the use of inbred (genetically homogeneous) strains increases the power of the study by reducing the variability between the subjects[20]. However, it is often overlooked that controlled genetic variation should be present in the test population. This can be achieved by using a battery of inbred strains in a factorial design in which both treatment and strain are varied simultaneously[21]. Moreover, in 1997, a report from the Banbury conference already recommended the back-crossing of mutant mice into at least two inbred strains to allow testing of the mutants in congenic lines but also in F1 hybrids derived from those[4]. However, the current dominating trend to keep the mice only in the C57BL/6 background is tremendously limiting the external validity and generalization of many findings[22,23]. Preference for using inbred strains, based on expected low inter-individual variability has been challenged by a recent report showing that trait variability is not larger in outbred stocks than it is in inbred strains and that therefore,outbred mice can be successfully employed to enhance reproducibility and replicability[24]. Similar concerns about low genetic and environmental diversity have been expressed for human genome-wide association studies[25,26].

[1]Laboratory Animal Center, Helsinki Institute of Life Science, University of Helsinki, Helsinki, Finland. [2]Neuroscience Center, Helsinki Institute of Life Science, University of Helsinki, Helsinki, Finland. *e-mail: vootele.voikar@helsinki.fi

**Table 1 | Substrains used for testing various behavioral phenotypes**

| Publication (Year) | Strains tested (Vendor) | Sex | Phenotype of B6J compared to B6N |
|---|---|---|---|
| Ashworth et al. (2015)[91] | •C57BL/6 J (JAX)<br>•C57BL/6N (JAX) | M and F | B6J females showed faster habituation in open field and were better on rotarod; B6J males were more active in open field and showed reduced freezing in novel context after fear conditioning |
| Bryant et al. (2008)[63] | •C57BL/6J (JAX)<br>•C57BL/6NCrl (CRL)<br>•C57BL/6Ntac (TAC)<br>•C57BL/6NHsd (Harlan) | M | B6J had enhanced motor coordination (rotarod), enhanced nociception (tail withdrawal and hot plate) and reduced contextual fear |
| Grottick et al. (2005)[79] | •C57BL/6J (JAX)<br>•C57BL/6NHsd (Harlan) | M | B6J showed enhanced startle and reduced PPI |
| Hager et al., (2014)[64] | •C57BL/6JIco (CRL)<br>•C57BL/6NCrl (CRL) | M | B6J showed reduced fear conditioning and enhanced activity during dark period |
| Kirkpatrick et al. (2017)[92] | •C57BL/6J (JAX)<br>•C57BL/6NJ (JAX) | ND | B6J did not show binge eating |
| Kumar et al. (2013)[65] | •C57BL/6J (JAX)<br>•C57BL/6N (NCI-Frederick) | ND | B6J had reduced acute and sensitized locomotor response to psychostimulants (cocaine, metamphetamine) |
| Labots et al. (2016)[77] | •C57BL/6JolaHsd (Harlan)<br>•C57BL/6NCrl (CRL) | M | B6J showed lower avoidance behavior (anxiety) |
| Matsuo et al. (2010)[47] | •C57BL/6J (JAX/CRL)<br>•C57BL/6NCrlCrlj (CRL)<br>•C57Bl/6CrSlc (Japan SLC) | ND | B6J showed enhanced motor coordination (rotarod), nociception (hot plate), increased open field activity, social interaction, reduced anxiety in elevated plus maze but not in light-dark box, enhanced acoustic startle and reduced PPI, but no difference in basal temperature, body weight |
| Mulligan et al. (2008)[66] | •C57BL/6J (JAX)<br>•C57BL/6NCrl (CRL) | M and F | B6J consumed more ethanol |
| Pinheiro et al. (2016)[78] | •C57BL/6JCrl (CRL)<br>•C57BL/6NCrl (CRL) | M | B6J showed enhanced dyadic social interaction |
| Radulovic et al. (1998)[67] | •C57BL/6J (Harlan)<br>•C57BL/6NCrl (CRL) | M | B6J showed reduced contextual and less generalized fear |
| Siegmund et al. (2005)[76] | •C57BL/6JCrl (CRL)<br>•C57BL/6JolaHsd (Harlan)<br>•C57BL/6NCrl (CRL) | M | B6J showed reduced contextual fear, faster extinction and reduced anxiety in light-dark box (latter also reduced in B6J compared with B6/JOlaHsd) |
| Simon et al. (2013)[55] | •C57BL/6J<br>•C57BL/6NTac | M and F | B6J showed enhanced startle and reduced PPI, enhanced motor coordination (rotarod), open field activity and anxiety-like behavior dependent on testing environment (either increased, decreased or no difference across 4 labs) |
| Stiedl et al. (1999)[68] | •C57BL/6JolaHsd (Harlan)<br>•C57BL/6NCrlBR (CRL) | M | B6J showed reduced contextual fear and faster extinction |
| Sturm et al. (2015)[69] | •C57BL/6J (CRL)<br>•C57BL/6NCrl (CRL) | M | B6J were less sensitive to chronic corticosterone treatment (reduced stress response) and more active in the open field and home cage |

The substrain names are presented as found in the main text of the publications. In most cases the substrains are from different vendors. CRL, Charles River; F, female; JAX, Jackson Laboratory; M, male; ND, no data; PPI, pre-pulse inhibition; TAC, Taconic

Several recommendations have been made for improving the design, analysis and reporting of preclinical studies involving animal models[27–32]. One suggested strategy for improving replicability has been to implement a rigorous standardization of experimental methods and conditions (in addition to standardized genetic backgrounds). However, the efficacy of environmental standardization has been extensively debated and questioned[33–35]. Indeed, rigorous standardization can lead to idiosyncratic and unreproducible findings and revised strategies are needed for experimental design[36–39]. Moreover, it is clear that environmental manipulations are an essential factor in disease modeling[40,41]. In addition, problematic issues with mouse (behavioral) phenotyping have been regularly highlighted in the headlines of major scientific journals[42–44]. Consequently, researchers need to improve the reproducibility and translatability of animal work, notably by adopting the current paradigm shift in the conduction and interpretation of animal experiments[45,46].

The substrains of C57BL/6 mice are genetically very close to each other, but mutations (few already identified and probably much more unknown) may lead to substantial phenotypic differences[47–49]. The retinal degeneration 8 (*rd8*) mutation, which makes the mice nearly blind by the age of 8 weeks, is a common feature for all B6N substrains[50]. C57BL/6JCrl and C57BL/6JRj (but not C57BL/6JRccHsd) mice carry a deletion in the *Nnt* gene encoding nicotinamide nucleotide transhydrogenase[51]. This mutation has been associated with impaired control of glucose homeostasis and reduced insulin secretion. We did not include in our panel the C57BL/6JOlaHsd strain, which is known to carry a deletion in the *Scna* gene[52]. These differences are caused by genetic drifts occurring in any independent mouse-breeding colony. While genetic drift can be controlled by careful colony management practices, it cannot be stopped completely[53,54]. On the other hand, despite many efforts for standardizing the operating procedures, several results of studies on the behavioral phenotyping of B6 substrains have revealed conflicting results between laboratories[55].

Based on the available information and controversies, we set up a project for addressing the differences between the substrains of C57BL/6 mice from different sources. To that end, we compared the C57BL/6J and C57BL/6N female mice from three common vendors in Europe: Charles River Laboratories (CRL, Germany), Envigo (ENV, The Netherlands) and Janvier Labs (JAN, France). Only female mice were tested because we wanted to focus on strain differences. Female mice produce highly reliable data and are thus suitable for basic exploratory studies[56–58]. In addition, we wanted to avoid possible problems with escalating aggression in male mice, which is quite common in C57BL/6 strain, especially after transport and re-location of adolescent or adult animals[59]. The mice were tested in a battery of behavioral tests assessing exploratory and anxiety-like behavior, sociability, sensorimotor gating, fear conditioning, circadian activity. Similar batteries are commonly applied for the characterization of mutant mice[60–62].

## Results

Female mice of two C57BL/6 substrains from three different vendors—Charles River (CRL), Envigo (ENV) and Janvier (JAN)—arrived in our laboratory at the age of 7 weeks; they were allowed to adapt for 17–18 days before testing began at the age of 10 weeks (Fig. 1a). The experiment was carried out in three batches. At arrival, substantial differences between the transport boxes from different vendors were noted. There was abundant, though different nesting material available in the shipments by CRL and JAN, whereas no such enrichment was included in ENV's boxes (Fig. 1b).

The summary of three-way ANOVA results for the main parameters of all behavioral tests can be found in Table 2. The body weight of the mice was measured weekly and significant strain by vendor interaction was revealed: the B6N mice from ENV were much smaller compared to B6N from the other vendors. Moreover, the B6N from ENV weighed less than the B6J from ENV, whereas the B6N from CRL and JAN were heavier than B6J mice from the respective supplier (Fig. 1c).

The elevated plus-maze (EPM) and light-dark (LD) box are commonly used for measuring exploratory activity and anxiety-like behavior in mice. The significant main effects of the vendor established in these tests suggest that CRL mice displayed increased avoidance of exposed areas (Fig. 2a,b,d) as compared to the mice from other vendors. Moreover, the significant main effects of the strain for parameters measured in the LD box suggest that the B6N mice showed enhanced anxiety-like behavior (avoidance of brightly illuminated compartment) in comparison to the B6J mice. The locomotor activity (total distance travelled during the test) was not different between the strains in the EPM. However, activity was reduced in the B6N mice as compared to the B6J in the LD test, mainly due to the large difference between the ENV substrains (Fig. 2c). Next, we tested the spontaneous activity and exploration in the open field arena, where the B6N mice displayed significantly reduced activity as compared to the B6J mice (Fig. 2e). This difference was largest between the substrains from ENV. In addition, the number of rearings was significantly different between the substrains with B6J > B6N (Fig. 2f). Although the proportion of distance travelled in the center of the arena did not differ between the substrains and vendors, the time spent there was longer for the B6J mice as compared to B6N (Fig. 2g).

During the social approach test, the B6J mice from ENV were more active than the B6N from the same vendor as shown by the distance travelled (Fig. 3a). Overall, B6J mice spent more time in the interaction zone with unfamiliar mouse (Fig. 3b,c). However, this difference was more pronounced and significant between the substrains from CRL and JAN, whereas it was virtually absent in the mice from ENV. Acoustic startle reflex was elevated in the B6J mice from ENV and JAN, compared to the B6N strain from the respective vendor, whereas an opposite effect was found in the CRL mice (Fig. 3d). Pre-pulse inhibition was enhanced in the B6N mice from CRL and JAN compared with B6J mice from the respective vendor, while the substrains from ENV did not differ (Fig. 3e). The B6N mice from CRL and JAN reacted to the 0.6 mA foot-shock more vigorously than the B6J mice, as suggested by higher velocity during the administration of foot-shock (Fig. 3f). There was no difference between the groups in the freezing behavior at baseline, before conditioning. However, 24 h after conditioning, the B6N mice displayed enhanced contextual fear (freezing) as compared to the B6J mice. When placed in the novel context, not previously associated with delivery of foot-shock, the B6J mice displayed increased level of freezing compared with the B6N mice. Significant interaction between the strain and vendor in duration of freezing during the presentation of conditioned stimulus in the novel context indicated enhanced freezing in the B6N mice from ENV. An opposite effect (reduced freezing) was revealed in the B6N mice from CRL and JAN when compared to the respective groups of B6J mice (Fig. 3g).

Monitoring of circadian activity in single-housed mice over 7 days revealed a large difference between the B6N and B6J mice from ENV (Fig. 4a). Interestingly, the mice reacted differently to the individual housing: the body weight was increased in the B6N mice, whereas no change or even reduction was found in the B6J mice (Fig. 4b). However, there was no difference between the groups in the nest building abilities (average score 3 after the first night and 4 after the second night). The stress-induced hyperthermia (increased rectal temperature after two consecutive measurements) was stronger in the B6J mice (Fig. 4c).

## Discussion

In the present study, we examined the basic behavioral profile of C57BL/6J and C57BL/6N female mice, obtained from three different vendors (Charles River, Envigo, Janvier). Many of the previously known and published differences between the substrains of C57BL/6 mice were confirmed or expanded by the experiments presented here[47,55,63–69]. However, the effect of vendor has mostly been neglected in the previous studies. The general expectation seems to be that the differences between B6N and B6J substrains are universal and therefore, the substrains for comparison have often been ordered from different breeders (Table 1). Thus, to the best of our knowledge, this is the first systematic and simultaneous comparison of common B6N and B6J substrains from three different vendors, carried out in one laboratory environment. We found significant effects of strain, vendor or interaction between these factors in the majority of outcomes.

Transportation of the animals from vendors to the research institutions can be a significant stressor for the animals. There are certain rules for security and guaranteed well-being of the animals throughout the journey[70], which may take several days (in our case 48–72 h from door to door). Therefore, it was interesting to find that the transport boxes from Envigo did not contain any nesting material. The nest material has become a mandatory part for structuring the rodent cages and lack of nesting material may substantially change the physiology and behavior of the animals[71–73]. However, this particular vendor justifies the lack of nesting material in transport containers by the fact that for welfare reasons the animals need to be seen through a viewing window during the transportation (Envigo, personal communication). We also found that the body weight of the B6N mice from CRL and JAN was higher than that of B6J mice from the respective vendors, whereas a large opposite effect was found between the substrains ordered from ENV. These differences are in line with the information provided in technical sheets by the vendors. However, there seems to even be differences between mice from the same substrain and vendor but bred at different locations; for instance C57BL6/NHsd female mice are about 2–3 g smaller in the Netherlands than in the United States (https://www.envigo.com/products-services/research-models-services/

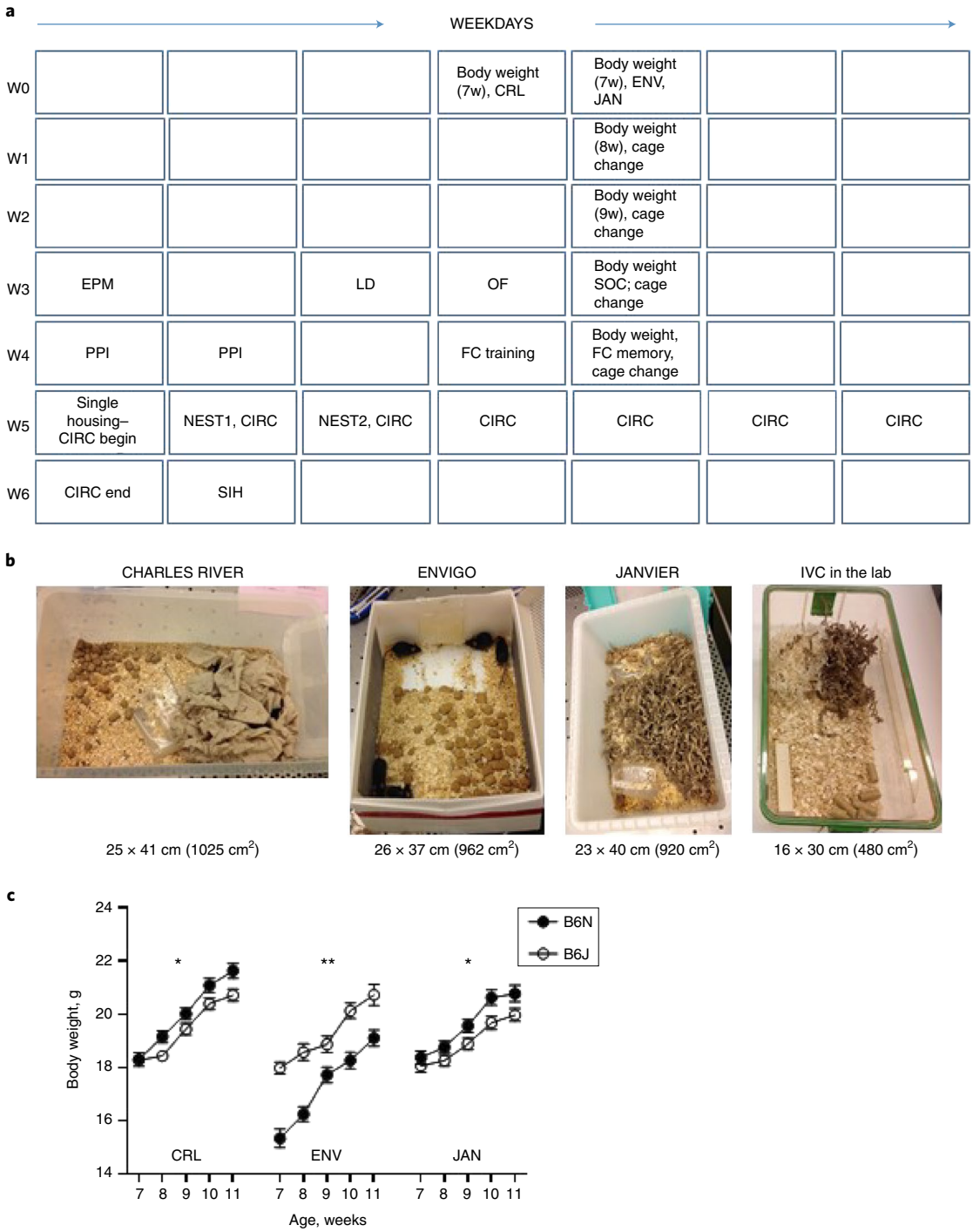


**Fig. 1 | Experimental design, transport boxes and mouse body weight. a**, Timeline of the experiment and behavioral testing. **b**, Characteristics of the transport boxes from the vendors and home cage in the destination. Animals were shipped in groups of six animals in respective boxes, and then randomly assigned to the individually ventilated cages in groups of three animals per cage. Notable differences were observed in the type and amount of nesting material provided by vendors. **c**, The body weight of the mice, measured during the course of the experiment. CIRC, circadian activity; CRL, Charles River; ENV, Envigo; EPM, elevated plus-maze; FC, fear conditioning; JAN, Janvier Lab; LD, light-dark box; NEST, nest building; OF, open field; PPI, pre-pulse inhibition; SIH, stress-induced hyperthermia; SOC, social approach. Filled circles, B6N; open circles, B6J. $^{*}P < 0.05$, $^{**}P < 0.01$ between the B6N and B6J mice from the same vendor.

models/research-models/mice/inbred/c57bl-6-inbred-mice/c57bl-6nhsd/). Overall, such differences between strains and substrains, vendors and locations — that is genetic and environmental factors — can provide an interesting and reasonable resource for heterogenization of the population[74].

Testing exploratory activity of the mice by elevated plus-maze, light-dark box and open field revealed that the B6N mice displayed enhanced anxiety-like behavior as compared to the B6J. In addition, anxiety-like behavior was higher in the mice from CRL as compared to the other two vendors. Although widely used, the conventional

**Table 2 | Summary of statistics (three-way ANOVA) for the main outcome variables**

| Parameter | Batch | Strain | Vendor | Batch*Strain | Batch*Vendor | Strain*Vendor | Batch*Strain*Vendor |
|---|---|---|---|---|---|---|---|
| Body weight at 7 weeks | F(2,90) = 2.33, $P = 0.10$ | F(1,90) = 15.07, $^{a}P < 0.001$ | F(2,90) = 30.42, $^{a}P < 0.001$ | F(2,90) = 0.30, $P = 0.74$ | F(4,90) = 1.81, $P = 0.13$ | F(2,90) = 23.71, $^{a}P < 0.001$ | F(4,90) = 3.65, $^{a}P = 0.009$ |
| Body weight at 10 weeks | F(2,89) = 1.16, $P = 0.32$ | F(1,89) = 0.07, $P = 0.79$ | F(2,89) = 16.20, $^{a}P < 0.001$ | F(2,89) = 0.37, $P = 0.69$ | F(4,89) = 0.92, $P = 0.46$ | F(2,89) = 16.06, $^{a}P < 0.001$ | F(4,89) = 1.59, $P = 0.18$ |
| EPM: Total distance | F(2,90) = 1.82, $P = 0.17$ | F(1,90) = 2.27, $P = 0.14$ | F(2,90) = 0.93, $P = 0.40$ | F(2,90) = 3.34, $^{a}P = 0.04$ | F(4,90) = 1.74, $P = 0.15$ | F(2,90) = 0.45, $P = 0.63$ | F(4,90) = 1.62, $P = 0.18$ |
| EPM: Distance open (%) | F(2,90) = 0.09, $P = 0.91$ | F(1,90) = 0.39, $P = 0.53$ | F(2,90) = 7.12, $^{a}P = 0.001$ | F(2,90) = 4.18, $^{a}P = 0.02$ | F(4,90) = 2.11, $P = 0.09$ | F(2,90) = 0.10, $P = 0.91$ | F(4,90) = 4.08, $^{a}P = 0.004$ |
| EPM: Time open (%) | F(2,90) = 0.27, $P = 0.76$ | F(1,90) = 0.12, $P = 0.73$ | F(2,90) = 4.14, $^{a}P = 0.02$ | F(2,90) = 4.47, $^{a}P = 0.01$ | F(4,90) = 3.11, $^{a}P = 0.02$ | F(2,90) = 0.21, $P = 0.81$ | F(4,90) = 5.13, $^{a}P < 0.001$ |
| EPM: Time center (%) | F(2,90) = 3.14, $P = 0.05$ | F(1,90) = 10.34, $^{a}P = 0.002$ | F(2,90) = 2.34, $P = 0.11$ | F(2,90) = 2.24, $P = 0.11$ | F(4,90) = 1.99, $P = 0.10$ | F(2,90) = 5.88, $^{a}P = 0.004$ | F(4,90) = 0.87, $P = 0.49$ |
| LD: Latency to light | F(2,89) = 0.37, $P = 0.69$ | F(1,89 = 17.03, $^{a}P < 0.001$ | F(2,89) = 7.65, $^{a}P < 0.001$ | F(2,89) = 0.52, $P = 0.60$ | F(4,89) = 1.14, $P = 0.34$ | F(2,89) = 8.26, $^{a}P < 0.001$ | F(4,89) = 1.29, $P = 0.28$ |
| LD: Total distance | F(2,89) = 1.12, $P = 0.33$ | F(1,89) = 11.49, $^{a}P = 0.001$ | F(2,89) = 1.95, $P = 0.15$ | F(2,89) = 6.46, $^{a}P = 0.002$ | F(4,89) = 0.58, $P = 0.68$ | F(2,89) = 13.15, $^{a}P < 0.001$ | F(4,89) = 0.77, $P = 0.54$ |
| LD: Distance light (%) | F(2,89) = 0.14, $P = 0.87$ | F(1,89) = 18.59, $^{a}P < 0.001$ | F(2,89) = 3.30, $^{a}P = 0.04$ | F(2,89) = 2.07, $P = 0.13$ | F(4,89) = 1.07, $P = 0.38$ | F(2,89) = 0.79, $P = 0.46$ | F(4,89) = 3.03, $^{a}P = 0.02$ |
| LD: Time light (%) | F(2,89) = 0.19, $P = 0.82$ | F(1,89) = 12.12, $^{a}P < 0.001$ | F(2,89) = 3.09, $P = 0.05$ | F(2,89) = 2.01, $P = 0.14$ | F(4,89) = 0.59, $P = 0.67$ | F(2,89) = 0.18, $P = 0.83$ | F(4,89) = 2.13, $P = 0.08$ |
| LD: Total rearings | F(2,89) = 1.67, $P = 0.19$ | F(1,89) = 2.86, $P = 0.09$ | F(2,89) = 4.85, $^{a}P = 0.01$ | F(2,89) = 0.57, $P = 0.57$ | F(4,89) = 1.40, $P = 0.24$ | F(2,89) = 8.21, $^{a}P < 0.001$ | F(4,89) = 2.08, $P = 0.09$ |
| LD: Rearings light (%) | F(2,89) = 0.25, $P = 0.78$ | F(1,89) = 14.47, $^{a}P < 0.001$ | F(2,89) = 2.23, $P = 0.11$ | F(2,89) = 2.50, $P = 0.09$ | F(4,89) = 1.05, $P = 0.38$ | F(2,89) = 0.04, $P = 0.96$ | F(4,89) = 2.31, $P = 0.06$ |
| OF: Total distance | F(2,89) = 6.32, $^{a}P = 0.003$ | F(1,89) = 43.76, $^{a}P < 0.001$ | F(2,89) = 7.23, $^{a}P = 0.001$ | F(2,89) = 1.97, $P = 0.15$ | F(4,89) = 0.53, $P = 0.72$ | F(2,89) = 8.01, $^{a}P < 0.001$ | F(4,89) = 1.43, $P = 0.23$ |
| OF: Distance center (%) | F(2,89) = 0.76, $P = 0.47$ | F(1,89) = 0.04, $P = 0.85$ | F(2,89) = 0.63, $P = 0.53$ | F(2,89) = 0.95, $P = 0.39$ | F(4,89) = 0.36, $P = 0.84$ | F(2,89) = 1.12, $P = 0.31$ | F(4,89) = 1.05, $P = 0.38$ |
| OF: Time center (%) | F(2,89) = 0.73, $P = 0.48$ | F(1,89) = 12.86, $^{a}P < 0.001$ | F(2,89) = 2.28, $P = 0.11$ | F(2,89) = 0.02, $P = 0.98$ | F(4,98) = 0.54, $P = 0.71$ | F(2,89) = 2.74, $P = 0.07$ | F(4,89) = 1.14, $P = 0.34$ |
| OF: Rearings | F(2,89) = 4.76, $^{a}P = 0.01$ | F(1,89) = 6.72, $^{a}P = 0.01$ | F(2,89) = 5.72, $^{a}P = 0.005$ | F(2,89) = 0.99, P = 0.38 | F(4,98) = 1.58, $P = 0.19$ | F(2,89) = 7.46, $^{a}P = 0.001$ | F(4,89) = 1.57, $P = 0.19$ |
| OF: Rearings center (%) | F(2,89) = 0.70, $P = 0.50$ | F(1,89) = 0.45, $P = 0.50$ | F(2,89) = 0.73, $P = 0.49$ | F(2,89) = 0.65, $P = 0.53$ | F(4,89) = 0.89, $P = 0.48$ | F(2,89) = 3.62, $^{a}P = 0.03$ | F(4,89) = 0.64, $P = 0.64$ |
| SOC: Total distance | F(2,89) = 1.54, $P = 0.22$ | F(1,89) = 3.98, $^{a}P = 0.05$ | F(2,89) = 1.90, $P = 0.15$ | F(2,89) = 0.90, $P = 0.41$ | F(4,89) = 0.80, $P = 0.53$ | F(2,89) = 13.60, $^{a}P < 0.001$ | F(4,89) = 0.73, $P = 0.57$ |
| SOC: Time interaction zone | F(2,89) = 2.15, $P = 0.12$ | F(1,89) = 12.0, $^{a}P < 0.001$ | F(2,89) = 5.93, $^{a}P = 0.004$ | F(2,89) = 3.04, $^{a}P = 0.05$ | F(4,89) = 1.00, $P = 0.41$ | F(2,89) = 4.34, $^{a}P = 0.02$ | F(4,89) = 1.06, $P = 0.38$ |
| SOC: Social preference (%) | F(2,89) = 0.95, $P = 0.39$ | F(1,89) = 4.65, $^{a}P = 0.03$ | F(2,89) = 2.34, $P = 0.10$ | F(2,89) = 1.81, $P = 0.17$ | F(4,89) = 1.49, $P = 0.21$ | F(2,89) = 3.77, $^{a}P = 0.03$ | F(4,89) = 0.85, $P = 0.50$ |
| AS&PPI: Startle | F(2,89) = 1.04, $P = 0.36$ | F(1,89) = 3.33, $P = 0.07$ | F(2,89) = 2.68, $P = 0.07$ | F(2,89) = 0.43, $P = 0.65$ | F(4,89) = 1.58, $P = 0.19$ | F(2,89) = 8.83, $^{a}P < 0.001$ | F(4,89) = 3.08, $^{a}P = 0.02$ |
| AS&PPI: Mean PPI | F(2,89) = 1.30, $P = 0.28$ | F(1,89) = 14.41, $^{a}P < 0.001$ | F(2,89) = 0.26, $P = 0.78$ | F(2,89) = 0.82, $P = 0.45$ | F(4,89) = 0.99, $P = 0.42$ | F(2,89) = 2.95, $P = 0.06$ | F(4,89) = 1.00, $P = 0.41$ |
| FC: Shock reactivity | F(2,89) = 3.26, $^{a}P = 0.04$ | F(1,89) = 14.15, $^{a}P < 0.001$ | F(2,89) = 1.93, $P = 0.15$ | F(2,89) = 1.20, $P = 0.30$ | F(4,89) = 1.22, $P = 0.31$ | F(2,89) = 5.49, $^{a}P = 0.006$ | F(4,89) = 0.77, $P = 0.54$ |
| FC: Freezing baseline (%) | F(2,89) = 0.02, $P = 0.98$ | F(1,89) = 1.84, $P = 0.18$ | F(2,89) = 0.53, $P = 0.59$ | F(2,89) = 0.37, $P = 0.69$ | F(4,89) = 2.14, $P = 0.08$ | F(2,89) = 0.20, $P = 0.52$ | F(4,89) = 0.81, $P = 0.52$ |
| FC: Freezing context (%) | F(2,89) = 0.09, $P = 0.92$ | F(1,89) = 7.30, $^{a}P = 0.008$ | F(2,89) = 5.04, $^{a}P = 0.009$ | F(2,89) = 1.76, $P = 0.18$ | F(4,89) = 2.63, $^{a}P = 0.04$ | F(2,89) = 1.03, $P = 0.36$ | F(4,89) = 0.13, $P = 0.97$ |
| FC: Freezing novel (%) | F(2,89) = 0.89, $P = 0.41$ | F(1,89) = 7.55, $^{a}P = 0.007$ | F(2,89) = 0.12, $P = 0.89$ | F(2,89) = 4.87, $^{a}P = 0.01$ | F(4,89) = 2.73, $^{a}P = 0.03$ | F(2,89) = 2.70, $P = 0.07$ | F(4,89) = 1.03. $P = 0.40$ |
| FC: Freezing cue (%) | F(2,89) = 7.07, $^{a}P = 0.001$ | F(1,89) = 0.01, $P = 0.91$ | F(2,89) = 14.66, $^{a}P < 0.001$ | F(2,89) = 2.71, $P = 0.07$ | F(4,89) = 0.02, $P = 0.99$ | F(2,89) = 8.02, $^{a}P < 0.001$ | F(4,89) = 0.75, $P = 0.56$ |
| Nest score, day 2 | F(2,54) = 2.64, $P = 0.08$ | F(1,54) = 0.08, $P = 0.78$ | F(2,54) = 0.85, $P = 0.43$ | F(2,54) = 0.62, $P = 0.54$ | F(4,54) = 1.12, $P = 0.36$ | F(2,54) = 0.74, $P = 0.48$ | F(4,54) = 0.36, $P = 0.84$ |
| Mean activity, light period | F(2,54) = 0.97, $P = 0.39$ | F(1,54) = 0.45, $P = 0.50$ | F(2,54) = 0.02, $P = 0.98$ | F(2,54) = 0.35, $P = 0.71$ | F(4,54) = 0.17, $P = 0.95$ | F(2,54) = 1.06, $P = 0.35$ | F(4,54) = 0.78, $P = 0.55$ |
| Mean activity, dark period | F(2,54) = 2.33, $P = 0.11$ | F(1,54) = 3.54, $P = 0.07$ | F(2,54) = 2.61, $P = 0.08$ | F(2,54) = 0.06, $P = 0.94$ | F(4,54) = 0.55, $P = 0.70$ | F(2,54) = 2.45, $P = 0.10$ | F(4,54) = 1.92, $P = 0.12$ |
| Body weight change (%) | F(2,54) = 1.04, $P = 0.36$ | F(1,54) = 9.85, $^{a}P = 0.003$ | F(2,54) = 2.97, $P = 0.06$ | F(2,54) = 0.77, $P = 0.47$ | F(4,54) = 0.83, $P = 0.51$ | F(2,54) = 0.52, $P = 0.60$ | F(4,54) = 0.70, $P = 0.59$ |
| Basal temperature | F(2,54) = 5.76, $^{a}P = 0.005$ | F(1,54) = 1.89, $P = 0.17$ | F(2,54) = 1.70, $P = 0.19$ | F(2,54) = 0.84, $P = 0.44$ | F(4,54) = 1.20, $P = 0.32$ | F(2,54) = 8.26, $^{a}P < 0.001$ | F(4,54) = 1.19, $P = 0.33$ |
| SIH | F(2,54) = 1.57, $P = 0.22$ | F(1,54) = 6.86, $^{a}P = 0.01$ | F(2,54) = 0.87, $P = 0.42$ | F(2,54) = 1.04, $P = 0.36$ | F(4,54) = 0.80, $P = 0.53$ | F(2,54) = 0.27, $P = 0.77$ | F(4,54) = 2.17, $P = 0.09$ |

AS&PPI, acoustic startle and pre-pulse inhibition; EPM, elevated plus-maze; FC, fear conditioning; LD, light-dark; OF, open field; SIH, stress-induced hyperthermia; SOC, social approach. [a]significant results.

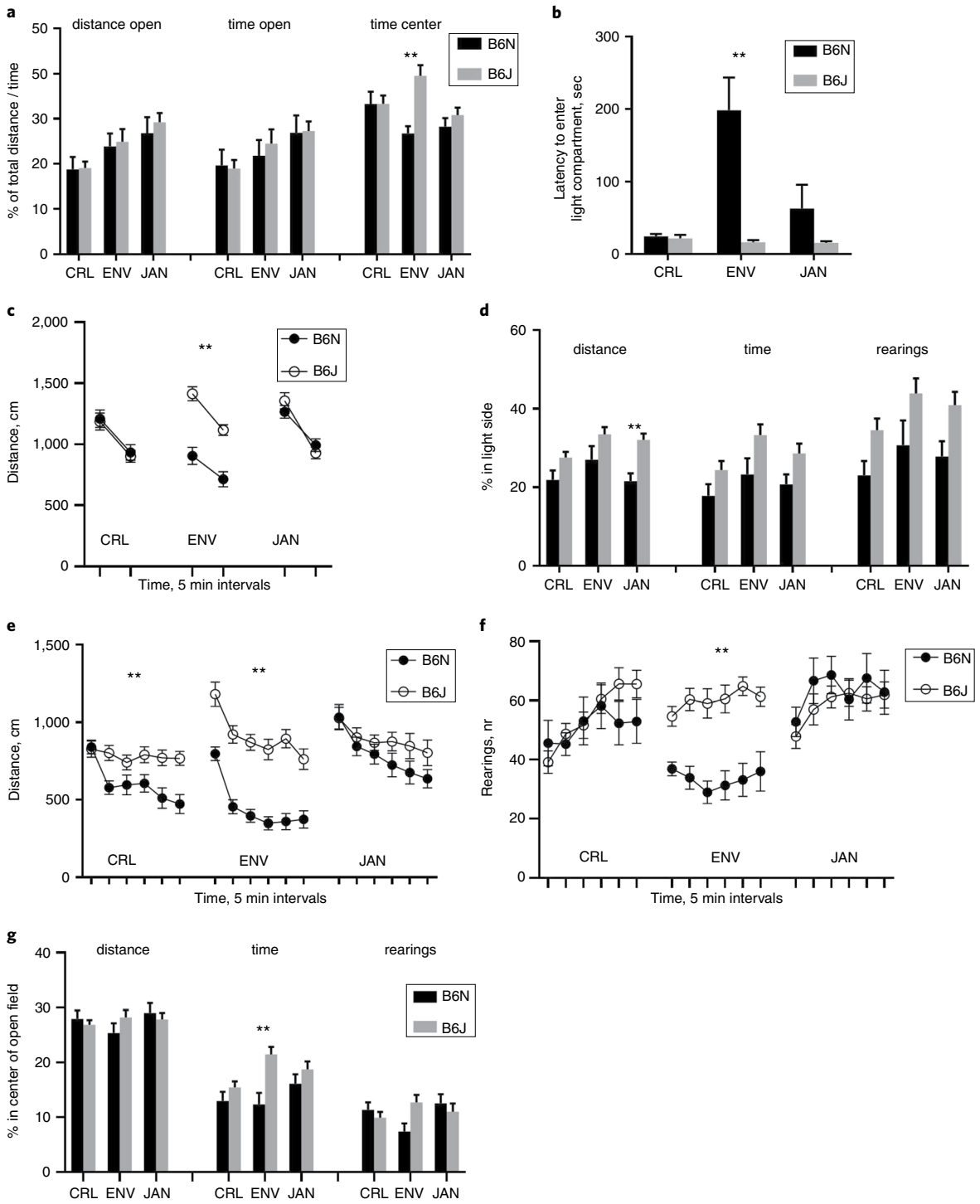


**Fig. 2 | Elevated plus-maze, light-dark box and open field. a**, Proportion of time and distance in open arms and central platform of the elevated plus-maze. **b**, Latency to enter the light compartment in light-dark test. **c**, Activity (distance travelled) during 10 min of testing in the light-dark box. **d**, Proportion of time, distance and exploratory rearings in the light compartment of the light-dark box. **e**, Distance travelled during 30 min in the open field arena. **f**, Number of rearings during 30 min in the open field arena. **g**, The proportion of activity (distance, time and rearings) in the center area of the open field. Filled circles, black bars B6N; open circles, grey bars B6J. $^{*}P < 0.05$, $^{**}P < 0.01$ between the B6N and B6J mice from the same vendor.

tests for unconditioned anxiety have often produced contradictory findings, which are dependent on the laboratory environment[33,75]. Nevertheless, our finding of reduced anxiety and higher activity shown by the B6J mice is in line with several other reports[47,69,76,77]. In addition to the enhanced anxiety-like behavioral profile, the B6N mice showed less interest towards a novel mouse (social approach) and similar data have been previously published by others groups[47,78]. However, it has to be noted that in our experiment the difference in social interaction was robust between the substrains from the CRL and JAN, but not from ENV.

Augmented startle reflex and reduced pre-pulse inhibition in the B6J mice compared with B6N mice has been shown previously[47,55,79].

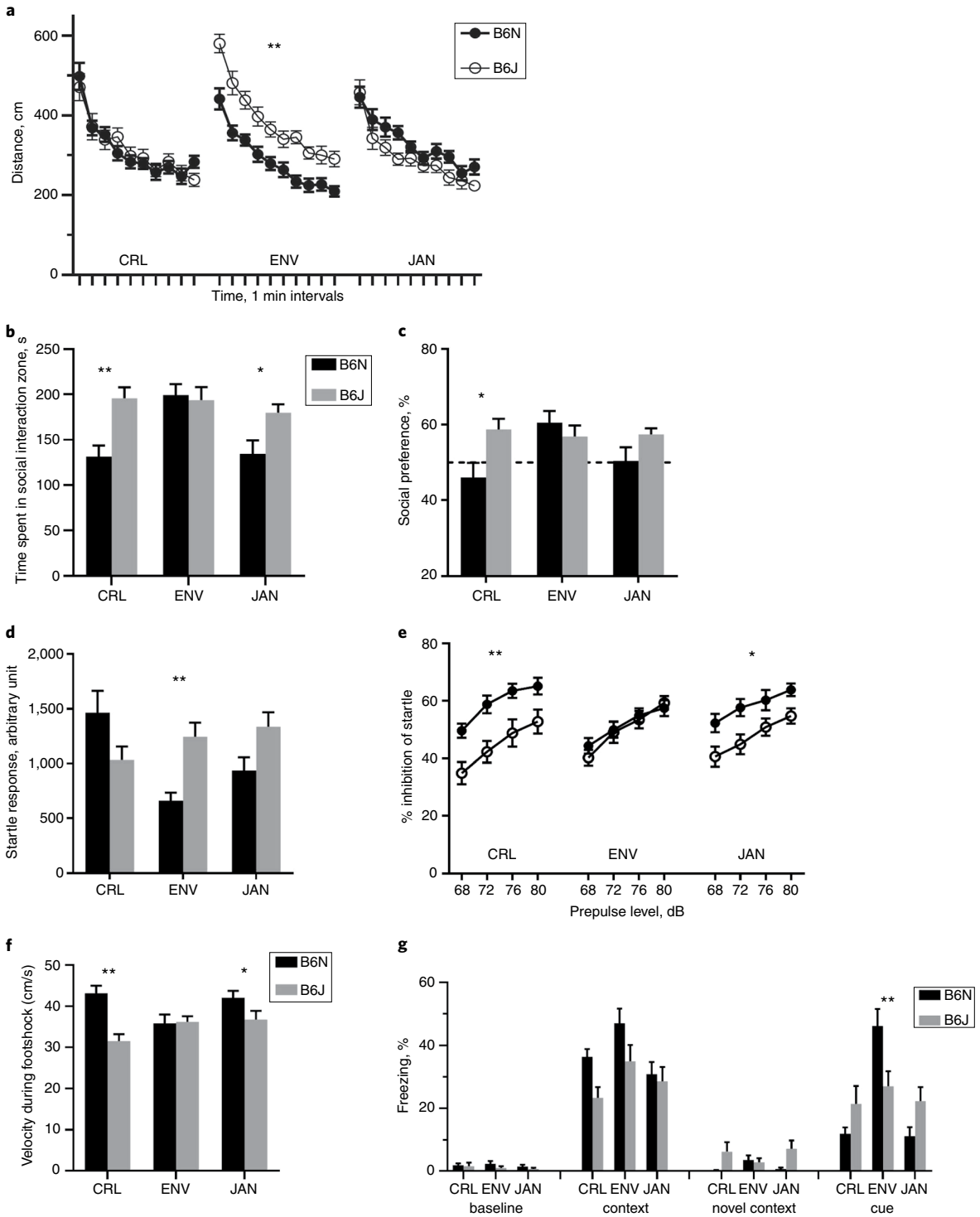


**Fig. 3 | Social approach, acoustic startle and pre-pulse inhibition, fear conditioning. a**, Distance travelled during 10 min test of social approach. **b**, Time spent in the social interaction zone during the social approach test. **c**, Preference for the cylinder with social stimulus, calculated as proportion of total time spent in exploring two cylinders. **d**, Magnitude of the startle response to 120 dB acoustic stimulus (40 ms white noise). **e**, Pre-pulse inhibition of the acoustic startle response at increasing pre-pulse intensities. **f**, Reaction to foot-shock in fear conditioning experiment, expressed as a mean velocity during two foot-shock applications (2 s each). **g**, Percentage of freezing time in different phases of fear conditioning experiment: baseline (2 min before first application of conditioned stimulus); context (3 min) ; novel context (2 min in the altered arena); cue (2 min of tone presentation in novel context). Filled circles, black bars B6N; open circles, grey bars, B6J. $^{*}P < 0.05$, $^{**}P < 0.01$ between the B6N and B6J mice from the same vendor.

In our panel, the startle was enhanced in the B6J mice from ENV and JAN, but reduced in the B6J from CRL when compared to the respective B6N substrains. Moreover, pre-pulse inhibition was reduced in the B6J-CRL and B6J-JAN mice, but no difference was found between the substrains from ENV. Fear (freezing) in the environment (context) associated with the foot-shock has been

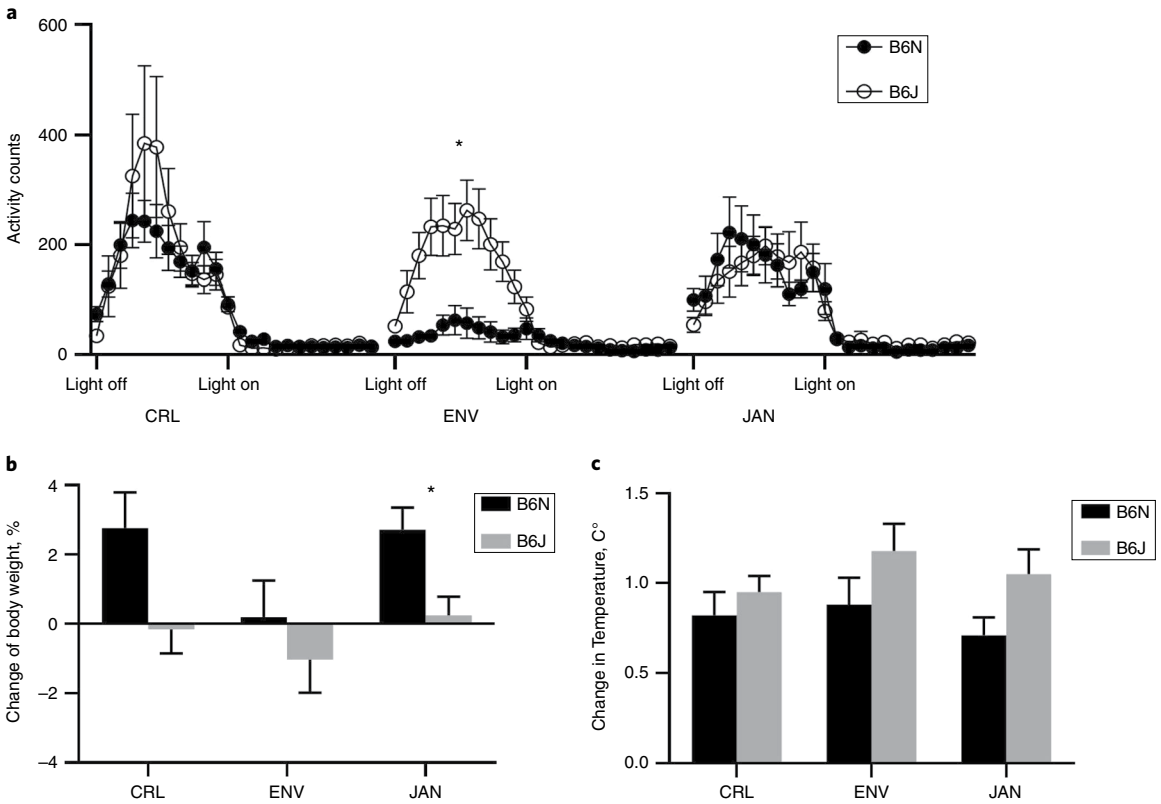


**Fig. 4 | Recording of circadian activity and stress-induced hyperthermia in individually caged mice. a**, Circadian activity (average counts per hour). **b**, The changes in body weight, shown as a difference in percentage between the end and start of single housing for measuring circadian activity. **c**, Stress-induced hyperthermia – difference between two consecutive (interval 10 min) measurement of rectal temperature. Filled circles, black bars, B6N; open circles, grey bars B6J. $^{*}P < 0.05$, $^{**}P < 0.01$ between the B6N and B6J mice from the same vendor.

consistently shown to be reduced in the B6J mice as compared to the B6N[63,64,67,68,76], and this was the case also in our study. However, this difference was limited to the substrains from CRL and ENV, and not detected in mice from JAN.

For measuring the circadian activity, nest building and stress-induced hyperthermia, the mice were housed individually for 8 days. Other groups have shown that home cage activity during the dark period is lower in B6N mice as compared to B6J[64,69]. In our study, this difference was seen only between the substrains ordered from ENV. Social isolation or separation of group-housed mice can be a stressful experience, especially for female mice[80]. The acute effect of the isolation stress may be seen in the changes of body weight[81]. In our study, weight gain was detected in the B6N mice, whereas no change or even decrease was found in the B6J mice. This finding suggests that the metabolic response and coping in stressful situations may be different between the C57BL/6 substrains. The B6N mice have been shown to be more vulnerable to the chronic treatment with corticosterone which is used as a model of stress[69]. They also displayed higher anxiety-like behavior and reduced social interaction. In contrast, the stress-induced hyperthermia was increased in the B6J mice. Based on these findings, it could be speculated that individual housing is a less stressful experience for the less social B6N mice. Thus, different substrains of C57BL/6 mice could be useful for elucidating the quantitative trait loci involved in the stress-related behavior. At the same time, further studies are warranted for characterization of these substrains under different conditions imposing stress on animals.

In general, the phenotypic differences between inbred strains have been shown to be stable and robust[82]. However, for obtaining such replicability (external validity), certain quality of study design and conduct (internal validity) is needed[83,84]. The phenotypic differences between the genetically close substrains of the C57BL/6 mice, coupled with the effect of vendor highlight the possible problems associated with choosing the background for genetically modified mice, interpretation of the results and reproducibility of the findings. Obviously, these cautions are not limited only to the C57BL/6 mice[85–87]. Moreover, further confounds can be caused by the local breeding schemes at the research institutions[88]. Nevertheless, genetically defined strains are and continue to be instrumental for elucidating the genetic basis of disease[89], although a recent report show that the trait variability is not larger in outbred stocks than it is in inbred strains[24]. Our data suggest that more emphasis and attention must be paid on the precise and accurate nomenclature when publishing research findings and designing future experiments. Heterogenization of the study samples, multi-laboratory experiments and refined statistical models have been proposed to be effective means for improving the reproducibility[36,74,90]. Therefore, deliberate variation of the mouse strains and substrains can be recommended as another way for improving the study design and addressing the issues of poor replicability.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41684-019-0288-8.

Received: 16 October 2018; Accepted: 19 March 2019;
Published online: 22 April 2019

## References


1. Collins, F. S., Rossant, J. & Wurst, W. A mouse for all reasons. *Cell* **128**, 9–13 (2007).
2. Gerlai, R. Gene-targeting studies of mammalian behavior: is it the mutation or the background genotype? *Trends Neurosci.* **19**, 177–181 (1996).
3. Crawley, J. N. et al. Behavioral phenotypes of inbred mouse strains: implications and recommendations for molecular studies. *Psychopharmacology (Berl)* **132**, 107–124 (1997).
4. Silva, A. J. et al. Mutant mice and neuroscience: recommendations concerning genetic background. Banbury Conference on genetic background in mice. *Neuron* **19**, 755–759 (1997).
5. Ayadi, A. et al. Mouse large-scale phenotyping initiatives: overview of the European Mouse Disease Clinic (EUMODIC) and of the Wellcome Trust Sanger Institute Mouse Genetics Project. *Mamm. Genome* **23**, 600–610 (2012).
6. Brown, S. D. & Moore, M. W. The international mouse phenotyping consortium: past and future perspectives on mouse phenotyping. *Mamm Genome* **23**, 632–640 (2012).
7. Pettitt, S. J. et al. Agouti C57BL/6N embryonic stem cells for mouse genetic resources. *Nat. Methods* **6**, 493–495 (2009).
8. Bradley, A. et al. The mammalian gene function resource: the International Knockout Mouse Consortium. *Mamm. Genome* **23**, 580–586 (2012).
9. Doetschman, T. Influence of genetic background on genetically engineered mouse phenotypes. *Methods Mol. Biol.* **530**, 423–433 (2009).
10. Bourdi, M., Davies, J. S. & Pohl, L. R. Mispairing C57BL/6 substrains of genetically engineered mice and wild-type controls can lead to confounding results as it did in studies of JNK2 in acetaminophen and concanavalin A liver injury. *Chemical research in toxicology* **24**, 794–796 (2011).
11. Wotjak, C. T. C57BLack/BOX? The importance of exact mouse strain nomenclature. *Trends Genet.* **19**, 183–184 (2003).
12. Kiselycznyk, C. & Holmes, A. All (C57BL/6) mice are not created equal. *Front. Neurosci* **5**, 10 (2011).
13. Bryant, C. D. The blessings and curses of C57BL/6 substrains in mouse genetic studies. *Ann. N Y Acad. Sci.* **1245**, 31–33 (2011).
14. Fontaine, D. A. & Davis, D. B. Attention to background strain is essential for metabolic research: C57BL/6 and the international knockout mouse consortium. *Diabetes* **65**, 25–33 (2016).
15. Kafkafi, N. et al. Reproducibility and replicability of rodent phenotyping in preclinical studies. *Neurosci. Biobehav. Rev.* **87**, 218–232 (2018).
16. Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **533**, 452–454 (2016).
17. Bespalov, A. & Steckler, T. Lacking quality in research: Is behavioral neuroscience affected more than other areas of biomedical science? *J. Neurosci. Methods* **300**, 4–9 (2018).
18. Perrin, S. Preclinical research: Make mouse studies work. *Nature* **507**, 423–425 (2014).
19. Begley, C. G. & Ellis, L. M. Drug development: Raise standards for preclinical cancer research. *Nature* **483**, 531–533 (2012).
20. Festing, M. F. Warning: the use of heterogeneous mice may seriously damage your research. *Neurobiol. Aging* **20**, 237–244 (1999). discussion 245–236.
21. Festing, M. F. Inbred strains should replace outbred stocks in toxicology, safety testing, and drug development. *Toxicologic pathology* **38**, 681–690 (2010).
22. Rivera, J. & Tessarollo, L. Genetic background and the dilemma of translating mouse studies to humans. *Immunity* **28**, 1–4 (2008).
23. Sittig, L. J. et al. Genetic Background Limits Generalizability of Genotype-Phenotype Relationships. *Neuron* **91**, 1253–1259 (2016).
24. Tuttle, A. H., Philip, V. M., Chesler, E. J. & Mogil, J. S. Comparing phenotypic variation between inbred and outbred mice. *Nat. Methods* **15**, 994–996 (2018).
25. Mills, M. C. & Rahal, C. A scientometric review of genome-wide association studies. *Commun. Biol* **2**, 9 (2019).
26. Haga, S. B. Impact of limited population diversity of genome-wide association studies. *Genet. Med* **12**, 81–84 (2010).
27. Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M. & Altman, D. G. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol.* **8**, e1000412 (2010).
28. Steward, O. & Balice-Gordon, R. Rigor or mortis: best practices for preclinical research in neuroscience. *Neuron* **84**, 572–581 (2014).
29. Mogil, J. S. & Macleod, M. R. No publication without confirmation. *Nature* **542**, 409–411 (2017).
30. Smith, A. J., Clutton, R. E., Lilley, E., Hansen, K. E. A. & Brattelid, T. PREPARE: guidelines for planning animal research and testing. *Lab Anim* **52**, 135–141 (2017).
31. Ioannidis, J. P. A. The proposal to lower P value thresholds to .005. *Jama* **319**, 1429–1430 (2018).
32. Steckler, T. et al. The preclinical data forum network: A new ECNP initiative to improve data quality and robustness for (preclinical) neuroscience. *Eur. Neuropsychopharmacol.* **25**, 1803–1807 (2015).
33. Crabbe, J. C., Wahlsten, D. & Dudek, B. C. Genetics of mouse behavior: interactions with laboratory environment. *Science* **284**, 1670–1672 (1999).
34. Wurbel, H. Behaviour and the standardization fallacy. *Nat. Genet.* **26**, 263 (2000).
35. Richter, S. H., Garner, J. P. & Wurbel, H. Environmental standardization: cure or cause of poor reproducibility in animal experiments? *Nat. Methods* **6**, 257–261 (2009).
36. Voelkl, B., Vogt, L., Sena, E. S. & Wurbel, H. Reproducibility of preclinical animal research improves with heterogeneity of study samples. *PLoS Biol.* **16**, e2003693 (2018).
37. Voelkl, B. & Wurbel, H. Reproducibility crisis: Are we ignoring reaction norms? *Trends Pharmacol. Sci.* **37**, 509–510 (2016).
38. Bailoo, J. D., Reichlin, T. S. & Wurbel, H. Refinement of experimental design and conduct in laboratory animal research. *ILAR J.* **55**, 383–391 (2014).
39. Karp, N. A. Reproducible preclinical research-Is embracing variability the answer? *PLoS Biol.* **16**, e2005413 (2018).
40. Ehrenreich, H. The impact of environment on abnormal behavior and mental disease: To alleviate the prevalence of mental disorders, we need to phenotype the environment for risk factors. *EMBO Rep* **18**, 661–665 (2017).
41. Sundberg, J. P. & Schofield, P. N. Living inside the box: environmental effects on mouse models of human disease. *Dis. Model Mech.* **11**, 1–7 (2018).
42. Editorial. Considerations for experimental design of behavioral studies using model organisms. *J.Neurosci* **39**, 1–2 (2019).
43. Editorial. Building a better mouse test. *Nat. Methods* **8**, 697–697 (2011).
44. Editorial. Troublesome variability in mouse studies. *Nat. Neurosci.* **12**, 1075 (2009).
45. Garner, J. P. The significance of meaning: why do over 90% of behavioral neuroscience results fail to translate to humans, and what can we do to fix it? *ILAR J.* **55**, 438–456 (2014).
46. Garner, J. P., Gaskill, B. N., Weber, E. M., Ahloy-Dallaire, J. & Pritchett-Corning, K. R. Introducing Therioepistemology: the study of how knowledge is gained from animal research. *Lab Anim (NY)* **46**, 103–113 (2017).
47. Matsuo, N. et al. Behavioral profiles of three C57BL/6 substrains. *Front. Behav. Neurosci.* **4**, 29 (2010).
48. Mekada, K. et al. Genetic differences among C57BL/6 substrains. *Exp. Anim.* **58**, 141–149 (2009).
49. Zurita, E. et al. Genetic polymorphisms among C57BL/6 mouse inbred strains. *Transgenic Res.* **20**, 481–489 (2011).
50. Mattapallil, M. J. et al. The Rd8 mutation of the Crb1 gene is present in vendor lines of C57BL/6N mice and embryonic stem cells, and confounds ocular induced mutant phenotypes. *Invest. Ophthalmol Vis Sci* **53**, 2921–2927 (2012).
51. Huang, T. T. et al. Genetic modifiers of the phenotype of mice deficient in mitochondrial superoxide dismutase. *Hum. Mol. Genet.* **15**, 1187–1194 (2006).
52. Specht, C. G. & Schoepfer, R. Deletion of the alpha-synuclein locus in a subpopulation of C57BL/6J inbred mice. *BMC Neurosci.* **2**, 11 (2001).
53. Zeldovich, L. Genetic drift: the ghost in the genome. *Lab Anim (NY)* **46**, 255–257 (2017).
54. Taft, R. A., Davisson, M. & Wiles, M. V. Know thy mouse. *Trends Genet.* **22**, 649–653 (2006).
55. Simon, M. M. et al. A comparative phenotypic and genomic analysis of C57BL/6J and C57BL/6N mouse strains. *Genome Biol.* **14**, R82 (2013).
56. Fritz, A. K., Amrein, I. & Wolfer, D. P. Similar reliability and equivalent performance of female and male mice in the open field and water-maze place navigation task. *Am. J. Med. Genet. C Semin. Med. Genet* **175**, 380–391 (2017).
57. Prendergast, B. J., Onishi, K. G. & Zucker, I. Female mice liberated for inclusion in neuroscience and biomedical research. *Neurosci. Biobehav. Rev.* **40**, 1–5 (2014).
58. Deacon, R. M. Housing, husbandry and handling of rodents for behavioral experiments. *Nat. Protoc.* **1**, 936–946 (2006).
59. Kappel, S., Hawkins, P. & Mendl, M. T. To group or not to group? Good practice for housing male laboratory mice. *Animals* **7**, pii: E88 (2017).
60. Paylor, R., Spencer, C. M., Yuva-Paylor, L. A. & Pieke-Dahl, S. The use of behavioral test batteries, II: Effect of test interval. *Physiol. Behav.* **87**, 95–102 (2006).
61. McIlwain, K. L., Merriweather, M. Y., Yuva-Paylor, L. A. & Paylor, R. The use of behavioral test batteries: Effects of training history. *Physiol. Behav.* **73**, 705–717 (2001).
62. Crawley, J. N. & Paylor, R. A proposed test battery and constellations of specific behavioral paradigms to investigate the behavioral phenotypes of transgenic and knockout mice. *Horm. Behav.* **31**, 197–211 (1997).
63. Bryant, C. D. et al. Behavioral differences among C57BL/6 substrains: implications for transgenic and knockout studies. *J. Neurogenet.* **22**, 315–331 (2008).
64. Hager, T. et al. Display of individuality in avoidance behavior and risk assessment of inbred mice. *Front. Behav. Neurosci.* **8**, 314 (2014).
65. Kumar, V. et al. C57BL/6N mutation in Cytoplasmic FMRP interacting protein 2 regulates cocaine response. *Science* **342**, 1508–1512 (2013).

66. Mulligan, M. K. et al. Alcohol trait and transcriptional genomic analysis of C57BL/6 substrains. *Genes Brain Behav.* **7**, 677–689 (2008).
67. Radulovic, J., Kammermeier, J. & Spiess, J. Generalization of fear responses in C57BL/6N mice subjected to one-trial foreground contextual fear conditioning. *Behav. Brain Res.* **95**, 179–189 (1998).
68. Stiedl, O. et al. Strain and substrain differences in context- and tone-dependent fear conditioning of inbred mice. *Behav. Brain Res.* **104**, 1–12 (1999).
69. Sturm, M., Becker, A., Schroeder, A., Bilkei-Gorzo, A. & Zimmer, A. Effect of chronic corticosterone application on depression-like behavior in C57BL/6N and C57BL/6J mice. *Genes Brain Behav* **14**, 292–300 (2015).
70. Swallow, J. et al. Guidance on the transport of laboratory animals. *Lab Anim.* **39**, 1–39 (2005).
71. Baumans, V. & Van Loo, P. L. How to improve housing conditions of laboratory animals: The possibilities of environmental refinement. *Vet. J.* **195**, 24–32 (2013).
72. Kulesskaya, N., Rauvala, H. & Voikar, V. Evaluation of social and physical enrichment in modulation of behavioural phenotype in C57BL/6J female mice. *PLoS ONE* **6**, e24755 (2011).
73. Gaskill, B. N. et al. Impact of nesting material on mouse body temperature and physiology. *Physiol Behav* **110-111C**, 87–95 (2013).
74. Richter, S. H. et al. Effect of population heterogenization on the reproducibility of mouse behavior: a multi-laboratory study. *PLoS ONE* **6**, e16461 (2011).
75. Harro, J. Animals, anxiety, and anxiety disorders: How to measure anxiety in rodents and why. *Behav. Brain Res.* **352**, 81–93 (2018).
76. Siegmund, A., Langnaese, K. & Wotjak, C. T. Differences in extinction of conditioned fear in C57BL/6 substrains are unrelated to expression of alpha-synuclein. *Behav. Brain Res.* **157**, 291–298 (2005).
77. Labots, M., Zheng, X., Moattari, G., Ohl, F. & van Lith, H. A. Effects of light regime and substrain on behavioral profiles of male C57BL/6 mice in three tests of unconditioned anxiety. *J. Neurogenet.* **30**, 306–315 (2016).
78. Pinheiro, B. S. et al. Dyadic social interaction of C57BL/6 mice versus interaction with a toy mouse: conditioned place preference/aversion, substrain differences, and no development of a hierarchy. *Behav. Pharmacol.* **27**, 279–288 (2016).
79. Grottick, A. J. et al. Neurotransmission- and cellular stress-related gene expression associated with prepulse inhibition in mice. *Brain Res. Mol. Brain Res.* **139**, 153–162 (2005).
80. Martin, A. L. & Brown, R. E. The lonely mouse: Verification of a separation-induced model of depression in female mice. *Behav. Brain Res.* **207**, 196–207 (2010).
81. Schipper, L., Harvey, L., van der Beek, E. M. & van Dijk, G. Home alone: a systematic review and meta-analysis on the effects of individual housing on body weight, food intake and visceral fat mass in rodents. *Obesity Reviews* **19**, 614–637 (2018).
82. Wahlsten, D., Bachmanov, A., Finn, D. A. & Crabbe, J. C. Stability of inbred mouse strain differences in behavior and brain size between laboratories and across decades. *Proc. Natl. Acad. Sci. U S A* **103**, 16364–16369 (2006).
83. Gulinello, M. et al. Rigor and reproducibility in rodent behavioral research. *Neurobiol. Learn. Mem.* in press, https://doi.org/10.1016/j.nlm.2018.01.001 (2018).
84. Tanila, H. Testing cognitive functions in rodent disease models: Present pitfalls and future perspectives. *Behav. Brain Res.* **352**, 23–27 (2018).
85. Boleij, H., Salomons, A. R., van Sprundel, M., Arndt, S. S. & Ohl, F. Not all mice are equal: welfare implications of behavioural habituation profiles in four 129 mouse substrains. *PLoS ONE* **7**, e42544 (2012).
86. Cook, M. N., Bolivar, V. J., McFadyen, M. P. & Flaherty, L. Behavioral differences among 129 substrains: implications for knockout and transgenic mice. *Behav. Neurosci.* **116**, 600–611 (2002).
87. Sittig, L. J. et al. Phenotypic instability between the near isogenic substrains BALB/cJ and BALB/cBy. *J. Mamm. Genome* **25**, 564–572 (2014).
88. Olfe, J., Domanska, G., Schuett, C. & Kiank, C. Different stress-related phenotypes of BALB/c mice from in-house or vendor: alterations of the sympathetic and HPA axis responsiveness. *BMC Physiol.* **10**, 2 (2010).
89. Meehan, T. F. et al. Disease model discovery from 3,328 gene knockouts by The international mouse phenotyping consortium. *Nat. Genet.* **49**, 1231–1238 (2017).
90. Kafkafi, N. et al. Addressing reproducibility in single-laboratory phenotyping experiments. *Nat. Methods* **14**, 462–464 (2017).
91. Ashworth, A. et al. Comparison of neurological function in males and females from two substrains of C57BL/6 Mice. *Toxics* **3**, 1–17 (2015).
92. Kirkpatrick, S. L. et al. Cytoplasmic FMR1-interacting protein 2 is a major genetic factor underlying binge eating. *Biol. Psychiatry* **81**, 757–769 (2017).

## Acknowledgements

This study was supported by Jane and Aatos Erkko foundation (VV), Biocenter Finland and Helsinki Institute of Life Science. Professors Iiris Hovatta and Jaan-Olle Andressoo are greatly acknowledged for their valuable comments and discussions during the preparation of this manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Reprints and permissions information** is available at www.nature.com/reprints.

**Correspondence and requests for materials** should be addressed to V.V.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Methods

**Ethical statement.** The animal experiments were performed according to the EU legislation harmonized with Finnish legislation and have been approved by the National Animal Experiment Board of Finland (ESAVI/10165/04.10.07/2016).

**Animals.** Altogether 108 female mice from 6 strains were used for this study: C57BL/6JRccHsd and C57BL/6NHsd (Envigo, Horst, The Netherlands); C57BL/6JCrl (original Jackson strain, stock: 000664) and C57BL/6NCrl (Charles River Laboratories, Sulzfeld, Germany); C57BL/6JRj and C57BL/6NRj (Janvier Labs, Le Genest-Saint-Isle, France). The number of animals was determined by power analysis with medium effect size and power of 0.80 (G*Power version 3.1.9.2). The mice were ordered and tested in three batches (6 mice per strain in one batch, 18 mice per strain in total). The duration of transportation was 48 h from Charles River and 72 h from Envigo and Janvier. The second and third batches arrived two weeks and twelve weeks respectively after the first one (the batches arrived in January, February and April of 2018, respectively). All mice were shipped at the age of 7 weeks. After arrival, the mice were housed in groups of three in the individually ventilated cage system (Mouse IVC Green Line, Tecniplast, Italy; cage dimensions 391 × 199 × 160 mm, floor area 501 $cm^2$; air inlet and outlet valves located in the cage lid, on top of the cage; the rate of air change was set at 75 times per hour with air speed at animal level max. 0.05 m/s; half of the cage covered by a wire bar food hopper). On the next day after arrival, the mice were marked by ear punching and the first body weights were recorded. The animals were maintained in the specific pathogen free animal facility, in a large colony room together with hundreds of other mouse cages. Cage enrichment was provided by bedding (aspen chips 5 × 5 × 1 mm, 4HP, Tapvei, Estonia), nesting material (equal amount of aspen strips, PM90L, Tapvei, Estonia and Sizzle Nest, Datesand Group, UK) and an aspen brick (100 × 20 × 20 mm, Tapvei). Food (Global Diet 2916 C, pellet 12 mm, Envigo) and water (filtered and UV-irradiated) were available ad libitum. Room temperature was 22 °C ± 2 °C and relative humidity 50% ± 15%. The light cycle was 12:12 light:dark (lights were on between 6.00 and 18.00). The cages were cleaned once per week and animals were weighed before being moved to the new cage. The mice were checked for microphthalmia, fur and whiskers (barbering) without any notable findings. Behavioral testing was started when the animals were 10 weeks old (after 17–18 days of adaptation). One mouse (C57BL/6NCrl from the second batch) was discarded after the first test day (elevated plus-maze) due to an accident (the mouse escaped and was injured).

**Behavioral tests.** For all conventional tests (carried out during the light phase, between 9.00 and 16.00), the animals were moved from the colony room to the testing rooms in the same animal facility at least 30 min before the beginning of the experiment. The testing order of the cages and animals was counterbalanced and randomized for each experiment and the experimenter was blinded regarding the genotypes. Behavioral testing of individual batches was carried out in the following order (Fig. 1a): day 1, elevated plus-maze (9.00–15.00); day 3, light-dark box (9.00–11.00); day 4, open field (9.00–13.00); day 5, sociability (9.00–15.00); days 8–9, pre-pulse inhibition of acoustic startle (9.00–16.00); day 11, fear conditioning (training, 9.00–11.00); day 12, fear conditioning (contextual memory 9.00–11.00; cued memory 11.00–13.00); day 15, individual housing and start of recording circadian activity; days 16–17, assessment of nest building (at 9.00); day 23, end of recording the circadian activity; day 24, stress-induced hyperthermia (9.00–10.00).

*Elevated plus-maze (EPM).* The maze[93] consisted of two open arms (30 × 5 cm) and two enclosed arms (30 × 5 cm, inner diameter) connected by central platform (5 × 5 cm), elevated to 40 cm above the floor. The floor of each arm was light grey and the closed arms had transparent (15 cm high) side walls and end walls. The illumination level in all arms was ~150 lx. The mouse was placed in the center of the maze facing one of the enclosed arms and observed for 5 min. The latency to the first open arm entry, number of open and closed arm entries (four paw criterion), distance travelled and the time spent in different zones of the maze were measured (tracking by Ethovision XT 10.0, Wagenigen, The Netherlands). The number of fecal boli was counted at the end of the trial.

*Light-dark exploration (LD).* LD-test [94] was done 48 h after the EPM. The test was carried out in the square open field arena (30 × 30 × 20 cm, Med Associates, St. Albans, VT) equipped with infrared light sensors detecting horizontal and vertical activity. The dark insert (non-transparent for visible light) was used to divide the arena into two halves, an opening (a door with a width of 5.5 cm and height of 7.0 cm) in the wall of the insert allowed animal's free movement from one compartment to another. Illumination in the center of the light compartment was ~550 lx. The animal was placed in the dark compartment and allowed to explore the arena for 10 min. Latency to enter the light side, distance travelled, number of rearings, and time spent in different compartments were recorded by the program (Activity Monitor, version 5.8). The number of fecal boli was counted at the end of trial.

*Open field (OF).* OF-test was performed 24 h after the LD-test. The same arena and monitoring system used for LD test were used for OF-test, but without dark insert, illumination of the arena was ~150 lx. Animals were released in the corner of the arena and monitored for 30 min. For analysis, the arena was divided into center and periphery (peripheral zone defined as a 6 cm wide corridor along the wall).

*Social approach (SOC).* SOC-test was done 24 h after the OF. The equipment and method for testing sociability was a combination based on two approaches, the 3-compartment test[95] and social interaction arena[96]. The large cage (dimensions 48.0 × 37.5 × 21.0 cm) contained two transparent and perforated cylinders (diameter 9 cm, height 15 cm) which were fixed at the center of the opposite short walls (distance between the cylinders 30 cm). One of the cylinders (position counterbalanced between subjects) contained a stimulus mouse (unfamiliar age-matched and sex-matched NMRI mouse (Envigo), kept in groups and previously adapted to confinement in the cylinder) whereas another was empty. The test was performed under reduced light conditions (~30 lx). The test mouse was released in the center of the arena and movement of the mouse was recorded by Ethovision XT 10.0 during 10 min. Total distance travelled and the time spent in exploring the cylinders were measured (the interaction zone was defined as a 5 cm corridor around the cylinders; the ratio between the time exploring the social vs empty cylinder was calculated as an index of social preference).

*Pre-pulse inhibition of acoustic startle reflex (PPI).* The animal was enclosed in a transparent plastic tube (inner diameter of 4.5 cm and length of 8.0 cm). The tube was placed and fixed on the piezoelectric platform, inside a sound-attenuating startle chamber (Med Associates, St. Albans, VT) with a background white noise of 65 dB and left undisturbed for 5 min. Testing was performed in 12 blocks of 5 trials and five trial types were applied. One trial type was a 40-ms, 120-dB white noise acoustic startle stimulus (SS) presented alone. In the remaining four trial types, the startle stimulus was preceded by a 20-ms acoustic pre-pulse stimulus (PPS) with the white noise bursting at the level of 68, 72, 76 or 80 dB. The delay between the onset of PPS and SS was 100 ms; for controlling the baseline movement there was a null-period of 200 ms included before presentation of acoustic stimuli. The first and twelfth block consisted of SS-only trials. In remaining blocks, the SS and PPS + SS trials were presented in pseudorandomized order, such that each trial type was presented once within a block of 5 trials. The inter-trial interval ranged between 10 s and 20 s. The startle response was recorded for 65 ms starting with the onset of the startle stimulus. The maximum startle amplitude recorded during the 65-ms sampling window was used as the dependent variable. The startle response was averaged over 10 trials from blocks 2–11 for each trial type. The pre-pulse inhibition for each PPS was calculated by using the following formula: 100-[(startle response on PPS + SS trials/startle response on SS trials)×100].

*Fear conditioning (FC).* The experiments were carried out employing a computer-controlled fear conditioning system (TSE, Bad Homburg, Germany). Training was performed in a transparent acrylic arena (23 × 23 × 35 cm) within a constantly illuminated (~100 lx) conditioning chamber. A loudspeaker provided a constant, white background noise (68 dB) for 120 s, followed by a 10 kHz tone — conditioned stimulus (CS), 76 dB, pulsed at 5 Hz —for 30 s. The tone was terminated by a foot-shock — unconditioned stimulus (US) of 0.6 mA for 2 s, constant current — delivered through a stainless steel floor grid (rod diameter 4 mm, distance 10 mm). Two CS–US pairings were separated by a 30 s pause, and the trial ended 30 s after the second foot-shock. Contextual memory was tested 24 h after the training. The animals were returned to the conditioning arena and the total time of freezing (defined as an absence of any movements for more than 3 s) was measured by infrared light beams (scanned continuously with a frequency of 10 Hz) during 3 min. Memory for the CS (tone) was tested 2 h later in a novel context. The new context was an acrylic box of similar size with black opaque walls and a smooth floor. A layer of wood chips (standard bedding material) under the floor provided a novel odor to the chamber. After 120 s of free exploration in the novel context, the CS was applied during next 120 s and freezing was measured as above.

*Circadian activity of single housed mice.* The InfraMot system (TSE, Germany) was used for recording the activity of single-housed animals by heat sensor. The system consisted of 24 units. Therefore, two animals from each original home cage were randomly assigned for testing the circadian activity. The mice were housed in Type II open cages (267 × 207 × 140 mm) with bedding (aspen chips, Tapvei) and nesting material (see the next paragraph). The sensor assembly was mounted on top of a cage lid. Food and water were available ad libitum. The recording continued for 7 days.

*Nest building.* Nest building was assessed after the first and second night of accommodation in single cages of the InfraMot system. One hour before the dark phase, one piece (5 $cm^2$, ~2.5 g) of compressed cotton fiber (Nestlets, Ancare, Bellmore, NY) was added into the cage. The next morning (~16 h later), the nests were assessed by visual inspection on a rating scale of 1–5 (1 = Nestlets > 90% intact, 2 = Nestlets 50–90% intact, 3 = Nestlets mostly shredded but no identifiable nest site, 4 = identifiable but flat nest, 5 = crater-shaped nest)[97]. Assessment was repeated 24 h later.

*Stress-induced hyperthermia (SIH).* This test was carried out after 7 days of activity recording in singly housed animals[98]. Briefly, a mouse was removed from the cage

and rectal temperature was measured. Then the body weight was measured and an animal was immediately returned to the same cage. After 10 min, the measurement of rectal temperature was repeated. Difference between these two measurements was defined as a stress-induced hyperthermia.

**Statistics.** Data were analyzed by using a three-way ANOVA model with batch (1,2,3), strain (B6N, B6J) and vendor (CRL, ENV, JAN) as between-subject factors. Within-subject factors (time and repeated measurements) were added when appropriate. The significance threshold was set at 0.05 and the results of the analysis are presented in Table 2. Newman-Keuls post-hoc comparisons were applied for further analysis if significant main effects or interactions were revealed. Software packages GraphPad Prism for Windows (v. 7.03) and STATISTICA (v. 12, StatSoft, Inc.) were used for the statistical analysis and for drawing the figures. The data on the figures are shown as mean values with error bars for standard error of mean. The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

## References


93. Lister, R. G. The use of a plus-maze to measure anxiety in the mouse. *Psychopharmacology* **92**, 180–185 (1987).
94. Crawley, J. N. & Goodwin, F. K. Preliminary report of a simple animal behavior model for the anxiolytic effects of benzodiazepines. *Pharmacol. Biochem. Behav.* **13**, 167–170 (1980).
95. Moy, S. S. et al. Sociability and preference for social novelty in five inbred strains: an approach to assess autistic-like behavior in mice. *Genes Brain Behav.* **3**, 287–302 (2004).
96. Golden, S. A., Covington, H. E. 3rd, Berton, O. & Russo, S. J. A standardized protocol for repeated social defeat stress in mice. *Nat. Protoc.* **6**, 1183–1191 (2011).
97. Deacon, R. M. Assessing nest building in mice. *Nat. Protoc.* **1**, 1117–1119 (2006).
98. Van der Heyden, J. A., Zethof, T. J. & Olivier, B. Stress-induced hyperthermia in singly housed mice. *Physiol. Behav* **62**, 463–470 (1997).