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

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

Neuroscience Center[1] and Laboratory Animal Center[2], Helsinki Institute of Life Science, University of Helsinki, Finland

[#] Corresponding author – vootele.voikar@helsinki.fi

ABSTRACT

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 a recommendation to use isogenic mice. The C57BL/6 strain has evolved as a gold standard strain for this purpose. However, the C57BL/6 mice are maintained as sub-strains 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 sub-strains (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 sub-strain differences were largely affected by the origin of animals, as a significant effect of vendor or interaction between the sub-strain and vendor occurred in all tests. This work highlights the importance of adhering to precise international nomenclature in all publications reporting the animal experiments. Moreover, the generalization of research findings from a single mouse sub-strain can be seriously limited due to genetic drift and environmental variables at different vendors. However, generalizability can be enhanced by heterogenization of samples by including the animals of different sub-strains. These issues need to be seriously considered for improving reproducibility, replicability and translational potential of the mouse models.

This is a post-peer-review, pre-copyedit version of an article published in Lab Animal. The final authenticated version is available online at https://doi.org/10.1038/s41684-019-0288-8

INTRODUCTION

Over the last 25 years thousands of knockout mice have been developed worldwide and 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 the embryonic stem cells from the 129-mouse strain followed by backcrossing to the C57BL/6 strain. The problems with mixed genetic background were identified soon[2,3] and recommendations for controlling it were formulated[4]. It is interesting to note that although the importance of nomenclature was emphasized, no particular attention was paid to different sub-strains of C57BL/6 mice at that time. The development of this strain goes back to 1921 when the strain was created by Clarence Cook Little and initially maintained at the Jackson Laboratory (C57BL/6J). In 1951, the sub-strain C57BL/6N was established after transfer of the mice to the National Institute of Health (NIH). From both parental colonies, the B6J and B6N mice have been moved to and maintained by several large mouse vendors (e.g. Charles River Laboratories, Envigo (formerly Harlan), Taconic, Janvier Labs) over the world. The C57BL/6J (later B6J) mouse was the first strain which genomic sequence was published and this strain has been considered a gold standard for many research areas. However, to overcome the problems associated with a mixed genetic background and to facilitate the production of mutant mice, the embryonic stem cells from C57BL/6N (later B6N) strain were established. Thus, International Mouse Phenotyping Consortium is creating the mutant mice for large-scale phenotyping in C57BL/6N background[5-8].

It is well known that the phenotype of mutant mice can depend on the genetic background[9] and this, although often not thought about, holds true also for the C57BL/6 substrains[10]. Although the existence of genotypic and phenotypic differences between the sub-strains 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 name 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 sub-strains, or did not consider it important. Interestingly, among those who knew about these differences, still 26% were not able to name exact strain they were using. Among the other factors, this can certainly be one of the major issues contributing to the current "reproducibility crisis" in basic research[15,16]. Moreover, the concerns have been expressed that the low quality of basic and pre-clinical studies (not limited to behavioral studies)[17] may have a direct relation to the failures in clinical trials[18,19].

For a long time, it has been believed and suggested that application of inbred (genetically homogeneous) strains is increasing the power 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, already Banbury conference[4] recommended back-crossing of mutant mice into at least two inbred strains, which would allow testing of the mutants in congenic lines, but also in F1 hybrids derived from those. However, the current dominating trend to keep the mice only in the C57BL/6 background is tremendously limiting the external validity and generalization of findings[22,23]. Preference for using the inbred strains, based on expected low inter-individual variability has been challenged by a recent report showing that the trait variability is not larger in outbred stocks than it is in inbred strains and therefore, the outbred mice could be successfully employed for enhancing the reproducibility and replicability[24]. Similar concerns about low genetic and environmental diversity have been expressed regarding the human genome-wide association studies[25,26].

So far, 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 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, it has to be acknowledged that 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 part in disease modelling[40,41]. Nevertheless, the problematic issues with mouse (behavioral) phenotyping have been regularly highlighted in the headlines of major journals[42-44]. Beyond all other issues, the researchers need to adopt the shifts in paradigms used for assessing the conduction and interpretation of animal experiments[45,46].

The sub-strains of C57BL/6 mice are genetically very close to each other, but few identified (and probably much more unknown) mutations may lead to substantial phenotypic differences[47-49]. The common feature for all B6N sub-strains is rd8 (retinal degeneration 8) mutation, which makes them nearly blind by the age of 8 weeks[50]. The C57BL/6JCrl and C57BL/6JRj (but not C57BL/6JRccHsd) mice possess deletion of a gene encoding nicotinamide nucleotide transhydrogenase (Nnt)[51]. This mutation has been associated with impaired glucose homeostasis control and reduced insulin secretion. We did not include in our panel C57BL/6JOlaHsd strain, which is known for the deletion of alpha synuclein gene (Scna)[52]. These differences are caused by genetic drift 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 behavioral phenotyping of B6 sub-strains have revealed conflicting results between the laboratories[55].

Based on the available information and controversies, we set up a project for addressing the differences between the sub-strains 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 aimed at establishing the 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].

MATERIAL and METHODS

Ethics statement, animals and environmental conditions.

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).

Altogether 108 female mice were used for this study: C57BL/6JRccHsd and C57BL/6NHsd (Envigo, Horst, The Netherlands - ENV); C57BL/6JCrl (original Jackson strain, stock: 000664) and C57BL/6NCrl (Charles River Laboratories, Sulzfeld, Germany - CRL); C57BL/6JRj and C57BL/6NRj (Janvier Labs, Le Genest-Saint-Isle, France - JAN). 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 hours from Charles River and 72 hours from Envigo and Janvier. The second batch arrived two weeks and the third batch twelve weeks after the first one (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 – cage dimensions 391 x 199 x 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; Tecniplast, Italy). 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 (SPF) animal facility, in a large colony room together with hundreds of other mouse cages. Cage enrichment was provided by bedding (aspen chips 5 x 5 x 1 mm, 4HP, Tapvei, Estonia), nesting material (equal amount of aspen strips, PM90L, Tapvei, Estonia and Sizzle Nest (paper strands), Datesand Group, UK) and an aspen brick (100 x 20 x 20 mm, Tapvei). Food (Global Diet 2916C, pellet 12 mm, Envigo) and water (filtered and UV-irradiated) were available ad libitum. Room temperature was 22±2 °C and relative humidity 50±15 %. The lights were on between 6.00 and 18.00. The cages were cleaned once per week and animals were weighed before moving 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 the accident (escaped and injured when caught).

Behavioral tests

For all conventional tests (carried out during the light phase, between 9.00 and 16.00) the animals were moved from colony room to the testing rooms in the same animal facility at least 30 min before beginning of the experiment. 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)[63]*.* The maze consisted of two open arms (30 x 5 cm) and two enclosed arms (30 x 5 cm, inner diameter) connected by central platform (5 x 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- 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 minutes. 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 in the end of the trial.

Light-dark exploration (LD)[64]*.* LD-test was done 48 h after the EPM. The test was carried out in the square open field arena (30 x 30 x 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 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. Animal was placed in the dark compartment and allowed to explore the arena for 10 minutes. 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 in the end of trial.

Open field (OF). OF-test was performed 24 h after the LD-test. The same arena and monitoring system as for LD test was used, but without dark insert, illumination of the arena was ~150 lx. Animals were released in the corner of the arena and monitored for 30 minutes. 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 – 3-compartment test[65] and social interaction arena[66]. Large cage (dimensions 48 x 37.5 x 21 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- 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 minutes. 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 – ratio between 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 4.5 cm, length 8 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 minutes. 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) – the white noise burst 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 1st and 12th 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 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), 0.6 mA, 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 minutes.

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 mm x 207 mm x 140 mm) with bedding (aspen chips, Tapvei) and nesting material (see the next chapter). 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 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 square, ~2.5 g) of compressed cotton fiber (Nestlets, Ancare, Bellmore, NY) was added into the cage. 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)[67]. Assessment was repeated 24 hours later.

Stress-induced hyperthermia (SIH). This test was carried out after 7 days of activity recording in singly housed animals[68]. 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. Ten minutes later, 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, repeated measurements) were added where appropriate. The significance threshold was set at 0.05 and the results of analysis are presented in the 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 drawing the figures. The data on 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.

RESULTS

The female mice of two C57BL/6 sub-strains from three vendors 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 Charles River and Janvier, whereas no such enrichment was included in Envigo's boxes (Fig.1B).

The summary of three-way ANOVA results for the main parameters of all behavioral tests can be found in the 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 as compared to the B6J from the respective supplier (Fig.1C).

The elevated plus-maze and light-dark box are commonly used for measuring exploratory activity and anxiety-like behavior in the 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 other vendors. Moreover, the significant main effects of the strain for parameters measured in the light-dark 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 (Fig.2C, mainly due to the large difference between the ENV sub-strains).

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 sub-strains from ENV. In addition, the number of rearings was significantly different between the sub-strains (B6J>B6N; Fig.2F). Although the proportion of distance travelled in the center of the arena did not differ between the sub-strains 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 as shown by the distance travelled (Fig.3 A). The B6J mice spent significantly more time in the interaction zone with unfamiliar mouse (Fig.3B,C). However, this difference was more pronounced between the sub-strains from CRL and JAN and 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, 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 while the sub-strains from ENV did not differ (Fig.2E).

The B6N mice 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 hours 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 than 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 substantially enhanced freezing in the B6N mice from ENV. The 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 significantly 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 sub-strains of C57BL/6 mice were confirmed or expanded by the experiments presented here[48,55,69-75]. 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 sub-strains are universal and therefore, the sub-strains 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 sub-strains 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[76], which may take several days (in our case 48-72 hours 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[77-79]. However, this particular vendor is justifying the lack of nesting material in transport containers by the fact that for welfare reasons the animals need to be seen through viewing window during the transportation (personal communication). We also found that the body weight of the B6N mice from CRL and JAN was higher than in the B6J mice from the respective vendors, whereas a large opposite effect was found between the sub-strains ordered from ENV. These differences are in line with the information provided in technical sheets by the vendors. However, there seem to be differences even between the breeding nodes of the same vendor, for instance the C57BL6/NHsd female mice are about 2-3 grams smaller in the Netherlands than in the United States (https://www.envigo.com/products-services/research-models-services/models/research-models/mice/inbred/c57bl-6-inbred-mice/c57bl-6nhsd/). Overall, such differences between the strains and sub-strains, vendors and nodes (i.e. genetic and environmental factors) can provide an interesting and reasonable resource for heterogenization of the population[80].

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 tests for unconditioned anxiety have often produced contradictory findings, dependent on the laboratory environment[33,81]. Nevertheless, our finding of reduced anxiety and higher activity shown by the B6J mice is in line with several other reports[48,75,82,83]. 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 the others[48,84]. However, it has to be noted that in our experiment the difference in social interaction was robust between the sub-strains from the CRL and JAN, but not from ENV.

Augmented startle reflex and reduced pre-pulse inhibition in the B6J mice has been shown previously[48,55,85]. 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 sub-strains. Moreover, pre-pulse inhibition was reduced in the B6J-CRL and B6J-JAN mice, but no difference was found between the sub-strains from ENV. Fear (freezing) in the environment (context) associated with the foot-shock has been consistently shown to be reduced in the B6J mice as compared to the B6N[69-72,82], and this was the case also in our study. However, this difference was limited to the sub-strains 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. The others have shown that home cage activity during the dark period is lower in the B6N mice as compared to the B6J[72,75]. In our study, this difference was seen only between the sub-strains ordered from ENV. Social isolation or separation of group-housed mice can be a stressful experience, especially for female mice[86]. The acute effect of the isolation stress may be seen in the changes of body weight[87]. In our study, the 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 sub-strains. The B6N mice have been shown to be more vulnerable to the chronic treatment with corticosterone which is used as a model of stress[75]. 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 less stressful experience for the less social B6N mice. Thus, different sub-strains 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 sub-strains under different conditions imposing stress on animals.

In general, the phenotypic differences between the inbred strains have been shown to be stable and robust[88]. However, for obtaining such replicability (external validity), certain quality of study design and conduct (internal validity) is needed[89,90]. The phenotypic differences between the genetically close sub-strains 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[91-93]. Moreover, further confounds can be caused by the local breeding schemes at the research institutions[94]. Nevertheless, genetically defined strains are and continue to be instrumental for elucidating the genetic basis of disease[95] (however, see[24]). Our data suggest that more emphasis and attention must be paid on the precise and accurate nomenclature when publishing the research findings and designing the 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[36,80,96]. Therefore, deliberate variation of the mouse strains and sub-strains can be recommended as another way for improving the study design and addressing the issues of poor replicability.

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 valuable comments and discussions during preparation of this manuscript.

Table 1. Sub-strains used for testing the various behavioral phenotypes. The sub-strain names are presented as found in the main text of the publications, vendors added in parenthesis. Note that in most cases the sub-strains have been from different vendors. (M – males, F – females)

PUBLICATION	STRAINS TESTED	SEX	Phenotype of B6J compared to B6N
Ashworth et al., 2015[97]	C57BL/6J (Jax) C57BL/6N (Jax)	M & F	B6J females faster habituation in open field, better on rotarod; B6J males more active in open field, reduced freezing in novel context after fear conditioning
Bryant et al., 2008[69]	C57BL/6J (Jax); C57BL/6NCrl; C57BL/6NTac; C57BL/6NHsd	M	B6J enhanced motor coordination (rotarod), enhanced nociception (tail withdrawal, hot plate), reduced contextual fear
Grottick et al., 2005[85]	C57BL/6J (Jax); C57BL/6NHsd	M	B6J enhanced startle and reduced PPI
Hager et al., 2014[72]	C57BL/6JIco (Crl); C57BL/6NCrl	M	B6J reduced fear conditioning, enhanced activity during dark period
Kirkpatrick et al., 2017[98]	C57BL/6J (Jax); C57BL/6NJ	?	B6J not showing binge eating
Kumar et al., 2013[73]	C57BL/6J (Jax); C57BL/6N (NCI-Frederick)	?	B6J reduced acute and sensitized locomotor response to psychostimulants (cocaine, metamphetamine)
Labots et al., 2016[83]	C57BL/6JOlaHsd; C57BL/6NCrl	M	B6J show lower avoidance behavior (anxiety) compared to B6N
Matsuo et al., 2010[48]	C57BL/6J (Jax); C57BL/6NCrlCrlj; C57Bl/6CrSlc (Japan SLC)	?	B6J 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, no difference in basal temperature, body weight
Mulligan et al., 2008[74]	C57BL/6J (Jax); C57BL/6NCrl	M & F	B6J consume more ethanol than B6N
Pinheiro et al., 2016[84]	C57BL/6JCrl C57BL/6NCrl	M	B6J enhanced dyadic social interaction
Radulovic et al., 1998[70]	C57BL/6J (Harlan); C57BL/6NCrl	M	B6J reduced contextual and less generalized fear
Siegmund et al., 2005[82]	C57BL/6JCrl; C57BL/6JOlaHsd; C57BL/6NCrl	M	B6J reduced contextual fear, faster extinction, reduced anxiety in light-dark box (latter also reduced when compared to B6/JOlaHsd)
Simon et al., 2013[55]	C57BL/6J (?); C57BL/6NTac	M & F	B6J 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[71]	C57BL/6JOlaHsd; C57BL/6NCrlBR	M	B6J reduced contextual fear and faster extinction
Sturm et al., 2015[75]	C57BL/6J (Crl); C57BL/6NCrl	M	B6J less sensitive to chronic corticosterone treatment (reduced stress response); B6J more active in the open field and home cage

Table 2. Summary of statistics (three-way ANOVA) for the main outcome variables. The significant results are highlighted.

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, p<0.001	F(2,90)=30.42, p<0.001	F(2,90)=0.30, p=0.74	F(4,90)=1.81, p=0.13	F(2,90)=23.71, p<0.001	F(4,90)=3.65, 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, p<0.001	F(2,89)=0.37, p=0.69	F(4,89)=0.92, p=0.46	F(2,89)=16.06, 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, 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, p=0.001	F(2,90)=4.18, p=0.02	F(4,90)=2.11, p=0.09	F(2,90)=0.10, p=0.91	F(4,90)=4.08, 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, p=0.02	F(2,90)=4.47, p=0.01	F(4,90)=3.11, p=0.02	F(2,90)=0.21, p=0.81	F(4,90)=5.13, p<0.001
EPM: Time Center%	F(2,90)=3.14, p=0.05	F(1,90)=10.34, 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, 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, p<0.001	F(2,89)=7.65, p<0.001	F(2,89)=0.52, p=0.60	F(4,89=1.14, p=0.34	F(2,89)=8.26, 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, p=0.001	F(2,89)=1.95, p=0.15	F(2,89)=6.46, p=0.002	F(4,89)=0.58, p=0.68	F(2,89)=13.15, 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, p<0.001	F(2,89)=3.30, 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, p=0.02
LD: Time Light%	F(2,89)=0.19, p=0.82	F(1,89)=12.12, 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, p=0.01	F(2,89)=0.57, p=0.57	F(4,89)=1.40, p=0.24	F(2,89)=8.21, 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, 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, p=0.003	F(1,89)=43.76, p<0.001	F(2,89)=7.23, p=0.001	F(2,89)=1.97, p=0.15	F(4,89)=0.53, p=0.72	F(2,89)=8.01, 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, 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, p=0.01	F(1,89)=6.72, p=0.01	F(2,89)=5.72, p=0.005	F(2,89)=0.99, p=0.38	F(4,98)=1.58, p=0.19	F(2,89)=7.46, 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, 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, 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, p<0.001	F(4,89)=0.73, p=0.57
SOC: Time InteractionZone	F(2,89)=2.15, p=0.12	F(1,89)=12.0, p<0.001	F(2,89)=5.93, p=0.004	F(2,89)=3.04, p=0.05	F(4,89)=1.00, p=0.41	F(2,89)=4.34, 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, 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, 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, p<0.001	F(4,89)=3.08, p=0.02
AS&PPI: Mean PPI	F(2,89)=1.30, p=0.28	F(1,89)=14.41, 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, p=0.04	F(1,89)=14.15, 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, 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, p=0.008	F(2,89)=5.04, p=0.009	F(2,89)=1.76, p=0.18	F(4,89)=2.63, 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, p=0.007	F(2,89)=0.12, p=0.89	F(2,89)=4.87, p=0.01	F(4,89)=2.73, 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, p=0.001	F(1,89)=0.01, p=0.91	F(2,89)=14.66, p<0.001	F(2,89)=2.71, p=0.07	F(4,89)=0.02, p=0.99	F(2,89)=8.02, 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, 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, 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, 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, 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

FIGURE LEGENDS

Figure 1. 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. Symbols: B6N – filled circles; B6J – open circles; *, ** - $p<0.05$, $p<0.01$ respectively, between the B6N and B6J mice from the same vendor.

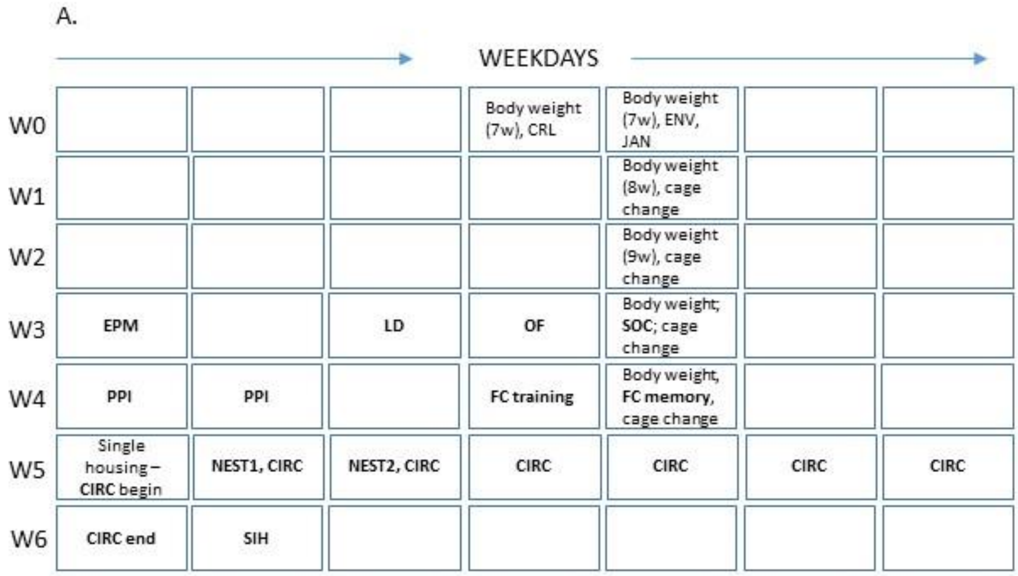
Figure 2. Elevated plus-maze, light-dark box and open field. A. Proportion of time and distance in open arms and central platform of the elevated plus-maze. B. Latency to enter the light compartment in light-dark test. C. Activity (distance travelled) during 10 min of testing in the light-dark box. D. 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, rearings) in the center area of the open field. Symbols: B6N – filled circles, black bars; B6J – open circles, grey bars; *, ** - $p<0.05$, $p<0.01$ respectively, between the B6N and B6J mice from the same vendor.

Figure 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 seconds 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). Symbols: B6N – filled circles, black bars; B6J – open circles, grey bars; *, ** - $p<0.05$, $p<0.01$ respectively, between the B6N and B6J mice from the same vendor.

Figure 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. Symbols: B6N – filled circles, black bars; B6J – open circles, grey bars; *, ** - $p<0.05$, $p<0.01$ respectively, between the B6N and B6J mice from the same vendor.

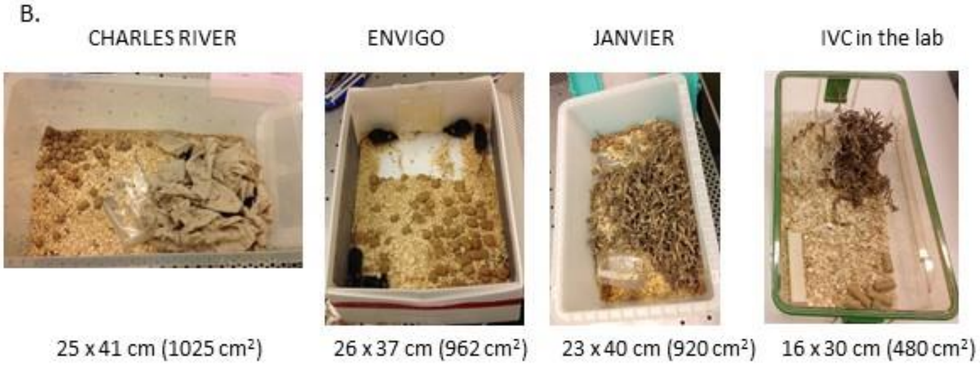
Figure 1

A.



WEEKDAYS

W0				Body weight (7w), CRL	Body weight (7w), ENV, JAN		
W1					Body weight (8w), cage change		
W2					Body weight (9w), cage change		
W3	EPM		LD	OF	Body weight; SOC; cage change		
W4	PPI	PPI		FC training	Body weight, FC memory, cage change		
W5	Single housing – CIRC begin	NEST1, CIRC	NEST2, CIRC	CIRC	CIRC	CIRC	CIRC
W6	CIRC end	SIH					



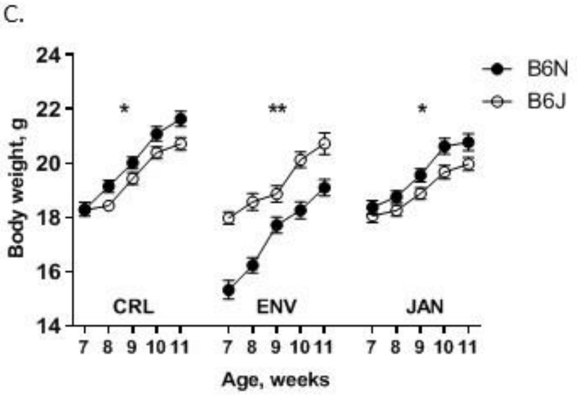


Figure 2

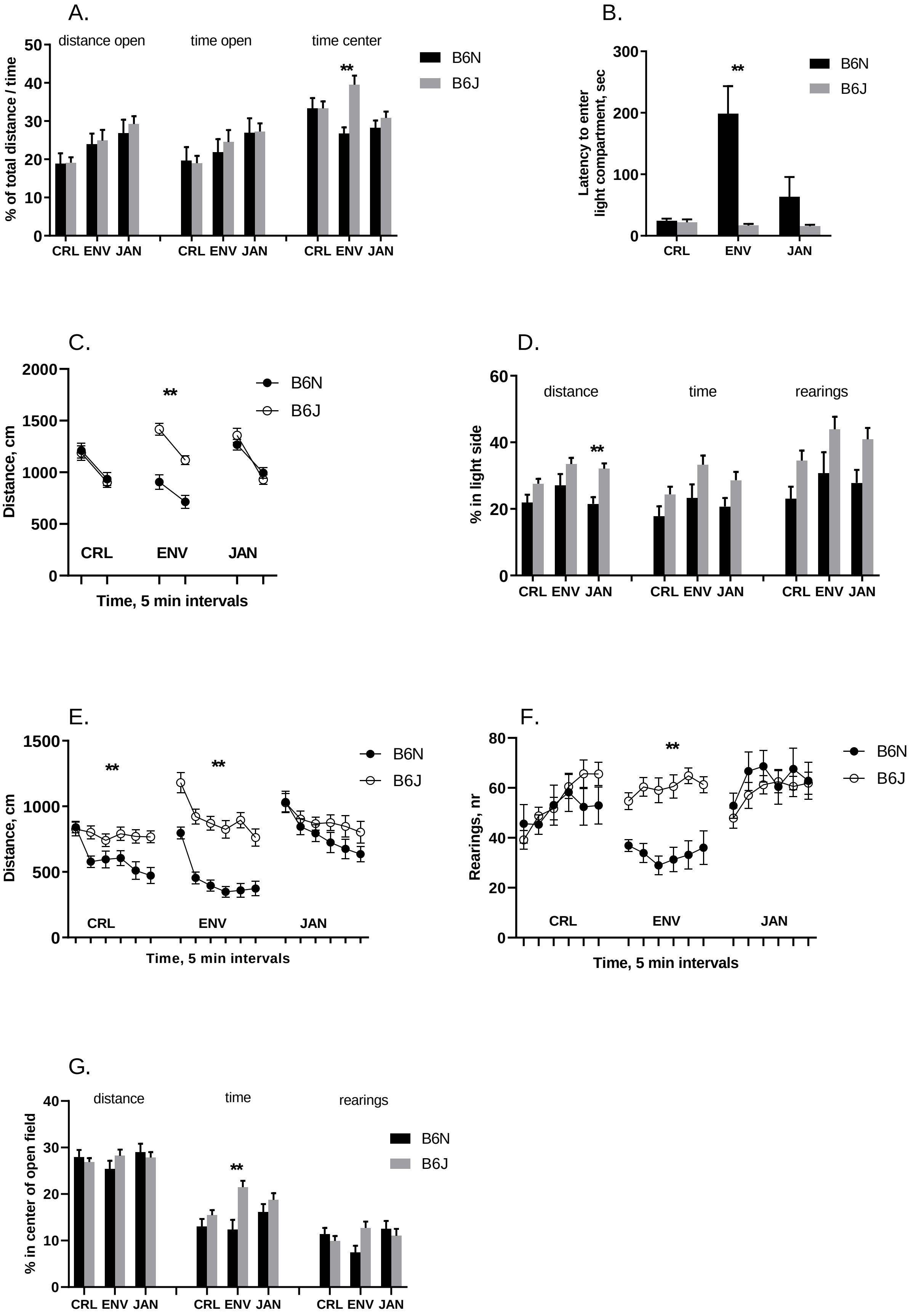

A.
distance open
time open
time center
% of total distance / time
0
10
20
30
40
50
**
CRL ENV JAN
CRL ENV JAN
CRL ENV JAN
B6N
B6J
B.
Latency to enter light compartment, sec
0
100
200
300
**
CRL
ENV
JAN
B6N
B6J
C.
Distance, cm
0
500
1000
1500
2000
**
B6N
B6J
CRL
ENV
JAN
Time, 5 min intervals
D.
% in light side
0
20
40
60
distance
time
rearings
**
CRL ENV JAN
CRL ENV JAN
CRL ENV JAN
E.
Distance, cm
0
500
1000
1500
**
**
B6N
B6J
CRL
ENV
JAN
Time, 5 min intervals
F.
Rearings, nr
0
20
40
60
80
**
B6N
B6J
CRL
ENV
JAN
Time, 5 min intervals
G.
% in center of open field
0
10
20
30
40
distance
time
rearings
**
B6N
B6J
CRL ENV JAN
CRL ENV JAN
CRL ENV JAN

Figure 3

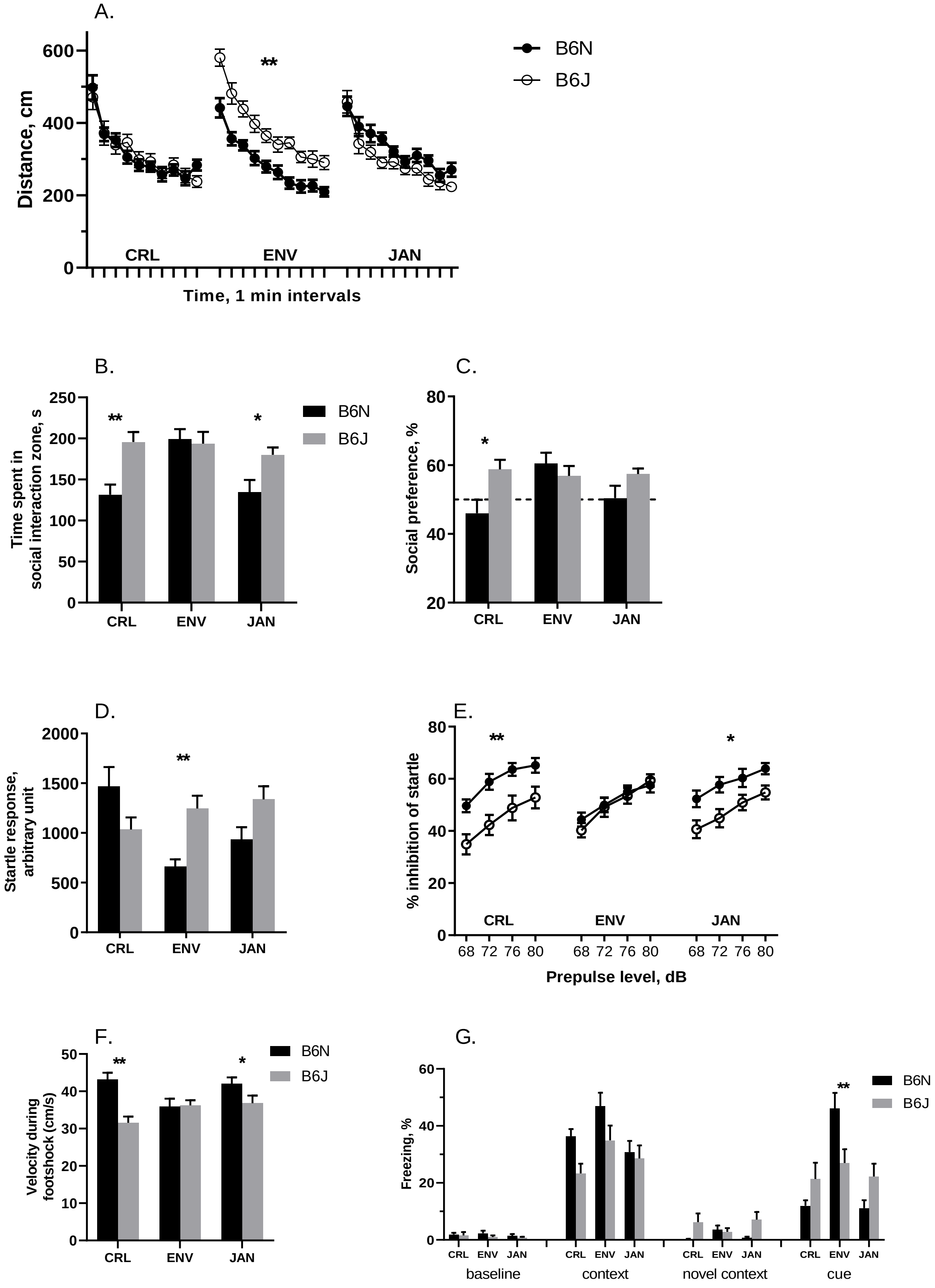
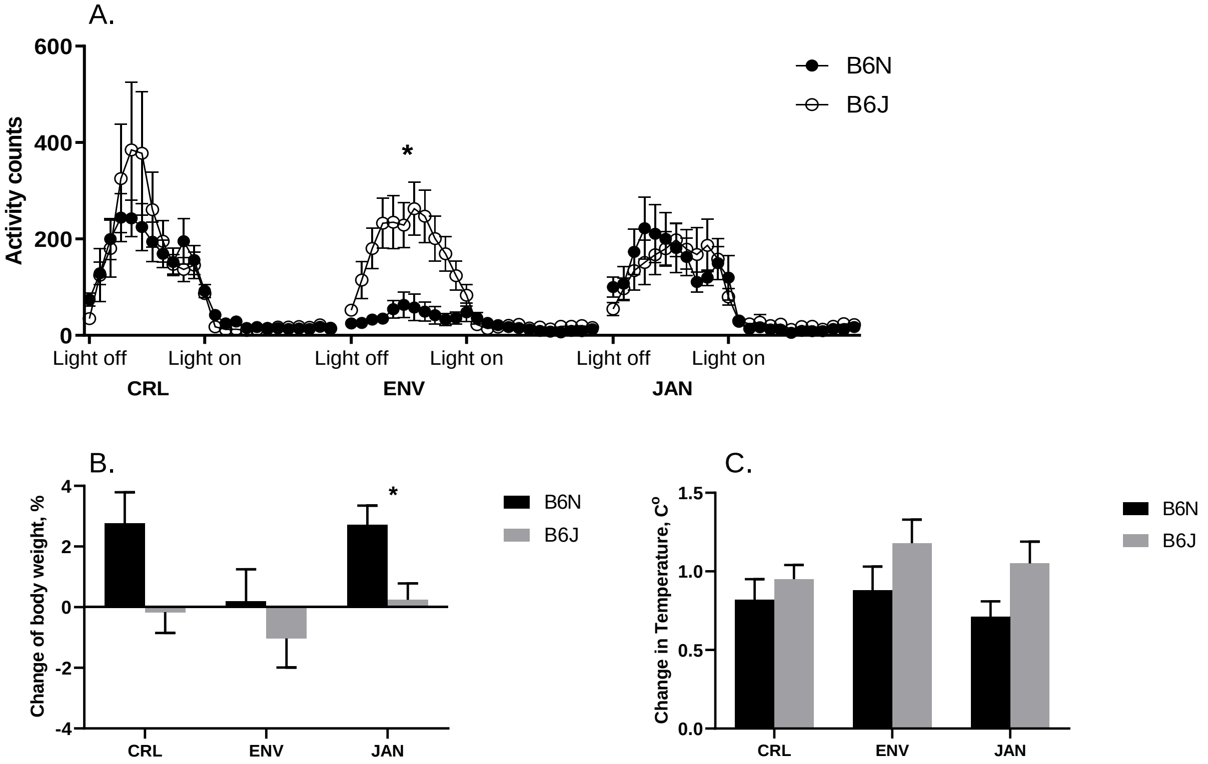

A.
Distance, cm
0
200
400
600
**
CRL
ENV
JAN
Time, 1 min intervals
B6N
B6J
B.
Time spent in social interaction zone, s
0
50
100
150
200
250
**
*
CRL
ENV
JAN
B6N
B6J
C.
Social preference, %
20
40
60
80
*
CRL
ENV
JAN
D.
Startle response, arbitrary unit
0
500
1000
1500
2000
**
CRL
ENV
JAN
E.
% inhibition of startle
0
20
40
60
80
**
*
CRL
ENV
JAN
68 72 76 80
68 72 76 80
68 72 76 80
Prepulse level, dB
F.
Velocity during footshock (cm/s)
0
10
20
30
40
50
**
*
CRL
ENV
JAN
B6N
B6J
G.
Freezing, %
0
20
40
60
**
CRL ENV JAN
CRL ENV JAN
CRL ENV JAN
CRL ENV JAN
baseline
context
novel context
cue
B6N
B6J

Figure 4


A.
600
400
200
0
Activity counts
*
B6N
B6J
Light off
Light on
Light off
Light on
Light off
Light on
CRL
ENV
JAN
B.
4
2
0
-2
-4
Change of body weight, %
*
CRL
ENV
JAN
B6N
B6J
C.
1.5
1.0
0.5
0.0
Change in Temperature, C°
CRL
ENV
JAN
B6N
B6J

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, doi:10.1007/s00335-012-9418-y (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, doi:10.1007/s00335-012-9427-x (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, doi:10.1007/s00335-012-9422-2 (2012).

9 Doetschman, T. Influence of genetic background on genetically engineered mouse phenotypes. *Methods Mol Biol* **530**, 423-433, doi:10.1007/978-1-59745-471-1_23 [doi] (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, doi:10.1021/tx200143x (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, doi:10.3389/fnins.2011.00010 (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, doi:10.1111/j.1749-6632.2011.06325.x (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, doi:10.2337/db15-0982 (2016).

15 Kafkafi, N. *et al.* Reproducibility and replicability of rodent phenotyping in preclinical studies. *Neurosci Biobehav Rev* **87**, 218-232, doi:10.1016/j.neubiorev.2018.01.003 (2018).

16 Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **533**, 452-454, doi:10.1038/533452a (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, doi:10.1016/j.jneumeth.2017.10.018 (2018).

18 Perrin, S. Preclinical research: Make mouse studies work. *Nature* **507**, 423-425, doi:10.1038/507423a (2014).

19 Begley, C. G. & Ellis, L. M. Drug development: Raise standards for preclinical cancer research. *Nature* **483**, 531-533, doi:10.1038/483531a (2012).

20 Festing, M. F. Warning: the use of heterogeneous mice may seriously damage your research. *Neurobiol Aging* **20**, 237-244; discussion 245-236 (1999).

21 Festing, M. F. Inbred strains should replace outbred stocks in toxicology, safety testing, and drug development. *Toxicologic pathology* **38**, 681-690, doi:10.1177/0192623310373776 (2010).

22 Rivera, J. & Tessarollo, L. Genetic background and the dilemma of translating mouse studies to humans. *Immunity* **28**, 1-4, doi:10.1016/j.immuni.2007.12.008 (2008).

23 Sittig, L. J. *et al.* Genetic Background Limits Generalizability of Genotype-Phenotype Relationships. *Neuron* **91**, 1253-1259, doi:10.1016/j.neuron.2016.08.013 (2016).

24 Tuttle, A. H., Philip, V. M., Chesler, E. J. & Mogil, J. S. Comparing phenotypic variation between inbred and outbred mice. *Nature Methods* **15**, 994-996, doi:10.1038/s41592-018-0224-7 (2018).

25 Mills, M. C. & Rahal, C. A scientometric review of genome-wide association studies. *Commun Biol* **2**, 9, doi:10.1038/s42003-018-0261-x (2019).

26 Haga, S. B. Impact of limited population diversity of genome-wide association studies. *Genetics in medicine : official journal of the American College of Medical Genetics* **12**, 81-84, doi:10.1097/GIM.0b013e3181ca2bbf (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, doi:10.1371/journal.pbio.1000412 [doi] (2010).

28 Steward, O. & Balice-Gordon, R. Rigor or mortis: best practices for preclinical research in neuroscience. *Neuron* **84**, 572-581, doi:10.1016/j.neuron.2014.10.042 (2014).

29 Mogil, J. S. & Macleod, M. R. No publication without confirmation. *Nature* **542**, 409-411, doi:10.1038/542409a (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*, 23677217724823, doi:10.1177/0023677217724823 (2017).

31 Ioannidis, J. P. A. The Proposal to Lower P Value Thresholds to .005. *Jama* **319**, 1429-1430, doi:10.1001/jama.2018.1536 (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, doi:10.1016/j.euroneuro.2015.05.011 (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, doi:10.1371/journal.pbio.2003693 (2018).

37 Voelkl, B. & Wurbel, H. Reproducibility Crisis: Are We Ignoring Reaction Norms? *Trends Pharmacol Sci* **37**, 509-510, doi:10.1016/j.tips.2016.05.003 (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, doi:10.1093/ilar/ilu037 (2014).

39 Karp, N. A. Reproducible preclinical research-Is embracing variability the answer? *PLoS Biol* **16**, e2005413, doi:10.1371/journal.pbio.2005413 (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, doi:10.15252/embr.201744197 (2017).

41 Sundberg, J. P. & Schofield, P. N. Living inside the box: environmental effects on mouse models of human disease. *Dis Model Mech* **11**, doi:10.1242/dmm.035360 (2018).

42 Editorial. Considerations for Experimental Design of Behavioral Studies Using Model Organisms. *The Journal of Neuroscience* **39**, 1, doi:10.1523/JNEUROSCI.2794-18.2018 (2019).

43 Editorial. Building a better mouse test. *Nat Meth* **8**, 697-697 (2011).

44 Editorial. Troublesome variability in mouse studies. *Nat Neurosci* **12**, 1075, doi:nn0909-1075 [pii]

10.1038/nn0909-1075 [doi] (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, doi:10.1093/ilar/ilu047 (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 animal* **46**, 103-113, doi:10.1038/laban.1224 (2017).

47 Mekada, K. *et al.* Genetic differences among C57BL/6 substrains. *Exp Anim* **58**, 141-149, doi:JST.JSTAGE/expanim/58.141 [pii] (2009).

48 Matsuo, N. *et al.* Behavioral profiles of three C57BL/6 substrains. *Front Behav Neurosci* **4**, 29, doi:10.3389.fnbeh.2010.00029 (2010).

49 Zurita, E. *et al.* Genetic polymorphisms among C57BL/6 mouse inbred strains. *Transgenic Res* **20**, 481-489, doi:10.1007/s11248-010-9403-8 [doi] (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. *Investigative ophthalmology & visual science* **53**, 2921-2927, doi:10.1167/iovs.12-9662 (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, doi:10.1093/hmg/ddl034 (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 animal* **46**, 255-257, doi:10.1038/laban.1275 (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, doi:10.1186/gb-2013-14-7-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, doi:10.1002/ajmg.c.31565 (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, doi:10.1016/j.neubiorev.2014.01.001 (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 : an open access journal from MDPI* **7**, doi:10.3390/ani7120088 (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 Lister, R. G. The use of a plus-maze to measure anxiety in the mouse. *Psychopharmacology (Berl)* **92**, 180-185 (1987).

64 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).

65 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).

66 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, doi:10.1038/nprot.2011.361 (2011).

67 Deacon, R. M. Assessing nest building in mice. *Nat Protoc* **1**, 1117-1119 (2006).

68 Van der Heyden, J. A., Zethof, T. J. & Olivier, B. Stress-induced hyperthermia in singly housed mice. *Physiol Behav* **62**, 463-470, doi:S0031-9384(97)00157-1 [pii] (1997).

69 Bryant, C. D. *et al.* Behavioral differences among C57BL/6 substrains: implications for transgenic and knockout studies. *J Neurogenet* **22**, 315-331, doi:906582175 [pii]

10.1080/01677060802357388 [doi] (2008).
70 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).
71 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).
72 Hager, T. *et al.* Display of individuality in avoidance behavior and risk assessment of inbred mice. *Front Behav Neurosci* **8**, 314, doi:10.3389/fnbeh.2014.00314 (2014).
73 Kumar, V. *et al.* C57BL/6N mutation in Cytoplasmic FMRP interacting protein 2 regulates cocaine response. *Science* **342**, 1508-1512, doi:10.1126/science.1245503 (2013).
74 Mulligan, M. K. *et al.* Alcohol trait and transcriptional genomic analysis of C57BL/6 substrains. *Genes Brain Behav* **7**, 677-689, doi:10.1111/j.1601-183X.2008.00405.x (2008).
75 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, doi:10.1111/gbb.12208 (2015).
76 Swallow, J. *et al.* Guidance on the transport of laboratory animals. *Lab Anim* **39**, 1-39, doi:10.1258/0023677052886493 (2005).
77 Baumans, V. & Van Loo, P. L. How to improve housing conditions of laboratory animals: The possibilities of environmental refinement. *Vet J* **195**, 24-32, doi:10.1016/j.tvjl.2012.09.023 (2013).
78 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, doi:10.1371/journal.pone.0024755 [doi]

PONE-D-11-07441 [pii] (2011).
79 Gaskill, B. N. *et al.* Impact of nesting material on mouse body temperature and physiology. *Physiol Behav* **110-111C**, 87-95, doi:10.1016/j.physbeh.2012.12.018 (2013).
80 Richter, S. H. *et al.* Effect of population heterogenization on the reproducibility of mouse behavior: a multi-laboratory study. *PLoS ONE* **6**, e16461, doi:10.1371/journal.pone.0016461 [doi] (2011).
81 Harro, J. Animals, anxiety, and anxiety disorders: How to measure anxiety in rodents and why. *Behav Brain Res* **352**, 81-93, doi:10.1016/j.bbr.2017.10.016 (2018).
82 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).
83 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*, 1-10, doi:10.1080/01677063.2016.1249868 (2016).
84 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, doi:10.1097/fbp.0000000000000223 (2016).
85 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, doi:10.1016/j.molbrainres.2005.05.020 (2005).
86 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).
87 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, doi:10.1111/obr.12663 (2018).
88 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).

89 Gulinello, M. *et al.* Rigor and Reproducibility in Rodent Behavioral Research. *Neurobiol Learn Mem* **in press**, doi:10.1016/j.nlm.2018.01.001 (2018).
90 Tanila, H. Testing cognitive functions in rodent disease models: Present pitfalls and future perspectives. *Behav Brain Res* **352**, 23-27, doi:10.1016/j.bbr.2017.05.040 (2018).
91 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, doi:10.1371/journal.pone.0042544 [doi]

PONE-D-12-18941 [pii] (2012).
92 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).
93 Sittig, L. J. *et al.* Phenotypic instability between the near isogenic substrains BALB/cJ and BALB/cByJ. *Mamm Genome* **25**, 564-572, doi:10.1007/s00335-014-9531-1 (2014).
94 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, doi:1472-6793-10-2 [pii]

10.1186/1472-6793-10-2 [doi] (2010).
95 Meehan, T. F. *et al.* Disease model discovery from 3,328 gene knockouts by The International Mouse Phenotyping Consortium. *Nat Genet*, doi:10.1038/ng.3901 (2017).
96 Kafkafi, N. *et al.* Addressing reproducibility in single-laboratory phenotyping experiments. *Nat Methods* **14**, 462-464, doi:10.1038/nmeth.4259 (2017).
97 Ashworth, A. *et al.* Comparison of Neurological Function in Males and Females from Two Substrains of C57BL/6 Mice. *Toxics* **3**, 1-17, doi:10.3390/toxics3010001 (2015).
98 Kirkpatrick, S. L. *et al.* Cytoplasmic FMR1-Interacting Protein 2 Is a Major Genetic Factor Underlying Binge Eating. *Biol Psychiatry* **81**, 757-769, doi:10.1016/j.biopsych.2016.10.021 (2017).