Dmitri L. Vyssotski

# Transgenerational epigenetic compensation of paternal drug treatment

## *SUPPORTING ONLINE MATERIAL*

## Table of contents

## MATERIALS AND METHODS

**Methods brief summary** [one page]

**Thyroxine experiments.** DBA/2J mice (P) were treated as neonates during the first 12 days (P0-P11) by subcutaneous injection of a daily dose of 2 µg L-thyroxine dissolved in 0.05 ml 0.9% NaCl made alkaline (pH 9.0) by adding a few drops of NaOH. Solution was prepared once 24 hr before the first administration (kept at +4°C). All pups in a given litter received the same treatment (between 17:00 and 18:00) and were kept in an original litter under their native DBA/2J mother (110-day-old at breeding). Control animals were left undisturbed. Reversed day-light cycle was used (8:00-20:00 – dark, 20:00-8:00 – light).

To have $F_1$, each DBA/2J male (P) at the age of 60 days was housed with 2 or 3 nulliparous 90-day-old naive DBA/2J females during 7 days. At birth pups were numbered and placed under primiparous NMRI foster-mothers to have 4 experimental and 4 control pups in each foster litter. To have $F_2$-incross, $F_1$ males at the age of 200 days were housed with $F_1$ females (2 females × 1 male, incross, but without inbreeding). To have $F_2$-outcross, $F_1$ males at the age of 230 days were housed with naive DBA/2J nulliparous 110-day-old females (2 females × 1 male). To have $F_3$, $F_2$-incross males at the age of 180 days were housed with $F_2$-incross females and $F_2$-outcross males at the age of 150 days were housed with $F_2$-outcross females (1 female × 1 male), simultaneously. NMRI foster-mothers were used in $F_1$, $F_2$ and $F_3$.

P and $F_1$ mice were tested in open-field test (30 min, 50 × 50 cm, 32 lx) and in the Morris water maze (see Methods in: Pleskacheva *et al.*, 2000). Mice starting from 40-60 days were housed individually.

P, $F_1$, $F_2$ and $F_3$ mice were tested in two-way avoidance task ("Mouse Shuttle Box", Campden Instruments Ltd., UK; Buselmaier *et al.*, 1981) at the age 90-155 days. Training: 5 days, 80 trials daily. The condition stimulus was light (5 sec), the negative reinforcement was foot-shock 0.15 mA (10 sec), which was supplied together with additional 10 sec of light, but both could be terminated by escaping to another compartment. This termination had a 0.8 sec delay – in order to have optimal DBA/2J training. Inter-trial interval: 5-15 sec. Averaged correct responses of 5 training days are shown in the figures.

For hippocampal mossy fiber (MF) morphometry and Timm staining P, $F_1$, $F_2$ and $F_3$ animals were perfused at the age of 190-605 days. The morphometric score for a given individual was taken as a ratio: (intra- and infrapyramidal MF)/(suprapyramidal MF).

**Morphine experiments.** Male Wistar rats, 42-day-old initially (P42; body weight 197 ± 20 g, mean ± SD), housed in groups 5-10 under normal day-light cycle, were injected intraperitoneally (i.p.) with morphine during 38 days. The first 7 days – twice daily (morning-evening, 8 hr between, mg/kg): 5-10, 15-15, 20-20, 25-30, 35-40, 45-50, 55-60 (10 mg/ml in 0.9% NaCl). Next day – 60 mg/kg in the morning and 6 hr later – injected i.p. with 2 mg/kg of naloxone (2 mg/ml) to induce early in life naloxone-precipitated morphine withdrawal. Next day – injected with morphine 60 mg/kg. The rest 29 days – injected with morphine 60 mg/kg twice daily Monday-Friday, and 60 mg/kg daily Saturday-Sunday. Control males – undisturbed.

During the last 5 days of morphine treatment P males were housed individually with drug-naive 75-day-old nulliparous Wistar females. To have $F_1$-2 ($F_1$, second brood), P males at the age of 175 days (*i.e.* 95 days of withdrawal) were housed individually with familiar females. To have $F_2$, $F_1$-2 males at the age of 85 days were bred individually with $F_1$-2 females (incross, but without inbreeding).

P, $F_1$, $F_2$ animals were tested in tail-withdrawal test at the age of 60-95 days. The distal part of the tail of a lightly restrained animal was dipped into circulating water thermostatically controlled at 56 ± 0.2°C. Latency to respond to the heat stimulus, by a vigorous flexion of the tail, was measured to the nearest 0.1 sec, cutoff latency – 15 sec. The test was done once before i.p. 10 mg/kg morphine injection (baseline latency) and 15, 30, 45, 60 and 90 min after. Baseline latency and 30-min latency divided by baseline are shown in the figures.

Opiate dependence was investigated in P, $F_1$, $F_2$ males at the age of 70-95 days. To have detectable morphine dependence in the offspring, $F_1$ and $F_2$ males (both experimental and control) were injected i.p. during 5.5 days (morning-evening, 12 hr interval, morphine, mg/kg): 10-10, 20-20, 30-30, 40-40, 50-50; next day – 60 mg/kg in the morning and 6 hr later – injected i.p. with 2 mg/kg of naloxone. Weight of each animal was measured to the nearest 1 g before naloxone administration and 24 hr later. Weight loss was taken as an indicator of opiate dependence.

The influence of 60 mg/kg morphine injection on locomotor activity was investigated in $F_1$ males 48 hours after above-mentioned naloxone administration (12-hr record: 3 hr before and 9 hr after injection).

Mann-Whitney U-test was used as a basic method for data analysis.

## 1. Thyroxine experiment

### 1.1. Animals and chemicals

For thyroxine study we have chosen inbred animals (DBA/2J mice) to check whether the post-meiotic germ cell selection, the selection which theoretically may take place in outbred stocks (*e.g.* Sprague-Dawley or Wistar rats used in previous studies), may be critical for observation of the discussed phenomena.

DBA/2J and NMRI mice were obtained from RCC Ltd (Itingen, Switzerland, www.rcc.ch). NMRI mice (outbred stock) were taken for fostering, because NMRI females (especially primiparous, who have fostered one native litter successfully) are known as very good foster mothers. All NMRI females, purchased for fostering of $F_1$ and $F_2$ generations, were primiparous. All purchased DBA/2J females were nulliparous (they had not any litter in advance). All mice starting from postnatal day P40-P60 were housed individually (one mouse per cage) – in the individual cages "Type 2" (polycarbonate, transparent; 267 × 207 × 140 mm). For other purposes (breeding, housing of a pregnant female, housing of female with pups) cage "Type 3" was used (polycarbonate, transparent; 425 × 266 × 155 mm). Reversed day-light cycle was used (8:00-20:00 – dark phase, 20:00-8:00 – light phase). The animals were tested during the dark phase. Animal room and experimental room temperature was kept about 21-22°C. In the animal room fluorescent (daylight) tube lamps provided intensity about 4-12 lx at the top of the cages during the light phase.

L-thyroxine was purchased from Fluka (Fluka, *BioChemika*, 89430, 1 g, L-Thyroxine Sodium salt, ~ 98% (HPLC, sum of enantiomers), Fluka Chemie GmbH, CH-9471 Buchs; Sigma-Aldrich, Pf, D-89552 Steinheim; Packed in Switzerland; www.sigmaaldrich.com/catalog/search/ProductDetail/FLUKA/89430).

#### 1.1.1. Primary breeding and postnatal treatment

DBA/2J animals were purchased from RCC at the age of 60 days (5 males and 15 females). Purchased males and females were born 2000-05-12 (year-month-day) in accordance with supplied record. They were bred at the age of 110 days (3 females × 1 male, in the cage "Type 3") to obtain P generation. No NMRI fostering was applied at this stage of experiment.

Hyperthyroidism was produced in newborn DBA/2J mice by subcutaneous injections of a daily dose of 2 µg L-thyroxine dissolved in 0.05 ml 0.9% NaCl, starting at postnatal day P0 and ending at day P11. The solution was made alkaline (pH 9.0) by adding a few drops of NaOH; prepared once 24 hr before the first administration, kept at temperature +4°C. Above-mentioned injections were done daily between 17:00 and 18:00. Control animals were left undisturbed. All pups in a given litter received the same treatment to prevent a competitive advantage of treated over untreated littermates or *vice versa*. There was no increased mortality rate, however. During treatment all pups were kept in original litters under their biological mothers. These animals, born about 2000-09-19 (year-month-day – hereinafter), were entitled as P. An additional (satellite) group of similarly treated P animals was born one month earlier – 2000-08-19.

#### 1.1.2. L-thyroxine solution preparation and injection

First of all, NaCl 0.9% solution with pH 9.0 was prepared, 50 ml. Under the temperature about 20°C 2 mg of L-thyroxine was dissolved in the 50 ml of above-mentioned NaCl solution with a help of magnetic stirring (30-min stirring) to have L-thyroxine concentration 0.04 mg/ml. Concentration 0.04 mg/ml was chosen to have 0.002 mg of L-thyroxine in 0.05 ml (one dose for injection). Prepared L-thyroxine solution was placed in the dark brown bottle and was kept at +4°C. L-thyroxine solution was prepared once for the whole 12-day treatment and this preparation was done exactly 24 hr before the first injection. To do the first (and each subsequent) injection some portion of solution was placed into small transparent bottle. The temperature of this small bottle was able to reach room temperature range (15-20°C) within 5-10 min. Then each experimental pup received 0.05 ml by subcutaneous injection. For this subcutaneous injection 0.05 ml

of solution was placed in 1 ml syringe with needle 0.5 × 12 mm. Pups were injected daily from day P0 until P11 (including both day of birth (P0) and day P11, *i.e*. 12 days in total) between 17:00 and 18:00.

Thus, during the first postnatal day (day P0) it was necessary to inject 0.05 ml into a newborn mouse pup (which had body weight about 1.25 g). Subcutaneous injection was done into the scruff of the neck, between ears. A pup was restrained (it was taken by the scruff of the neck, center scruff between ears) and was held horizontally, like an adult mouse. The needle was inserted (bevel up) over the dome of the scull at a 10° angle into the triangle of the skin between the ears, in rostro-caudal direction. The opening of the needle point was turned into "up" position to be visible. Both hands of an experimenter (with pup and with syringe) were kept on the table to diminish relative stochastic movements of pup and needle. Good light conditions were provided also. When the opening occurred under the skin, the solution was injected. Injected solution was visible like a relatively big bulb under the skin. The needle was removed without additional circular rotation to diminish possible tissue damage. Sometimes, especially in newborn pups, a small drop of solution went out from the mentioned above subcutaneous bulb. For each further day a new portion of solution was taken from above-mentioned dark brown bottle.

### 1.1.3. Breeding protocol

P-generation male adult DBA/2J mice (60 and 90 days of age /independent subgroups/) were housed with adult nulliparous DBA/2J females (90 days of age) for a period of 7 days. DBA/2J females were purchased from RCC in advance at the age of 60 days and they were living in groups in cages "Type 3" in our facilities during 1 month before the beginning of breeding period. During breeding period each male was housed with 2 or 3 females in the cage "Type 3" (425 × 266 × 155 mm). NMRI primiparous females were housed with NMRI males during the same period (2-3 females × 1 male) to be ready for fostering. At birth, DBA/2J pups were numbered and placed under NMRI foster-mothers by the following way. We were trying to form each foster litter using 2 experimental males, 2 experimental females, 2 control males and 2 control females (8 pups taken together). But in reality, due to natural temporal birth distribution (newborn pups foster litter formation took place daily at 5:00 p.m.), the amount of pups in a foster litter varied from 6 to 9 and each litter had at least 2 pups from the same group (experimental or control) and at least 2 pups of the same sex. If some original DBA/2J litter had only one alive pup at the day of birth (very rare event), this litter was excluded from litter size analysis, but above-mentioned pup was used in further experiment. Above-mentioned type of NMRI fostering and newborn pup numbering and redistribution into foster litters were used in all generations ($F_1$, $F_2$ and $F_3$). The animals were weaned at 32 days of age. Body weight and mortality rate were recorded daily from day P0 until P32. Above-mentioned weighing (to within 1%) was applied to all generations ($F_1$, $F_2$ and $F_3$). Obtained progeny, born about 2000-12-09, was entitled as "$F_1$, obtained from 2-month-old males", and "$F_1$, obtained from 3-month-old males" (both experimental groups had appropriate independent controls).

$F_1$ male adult mice (200 days of age), obtained from 2-month-old males, were housed with adult (200 days of age) $F_1$ females, obtained from 2-months old males and 3-months old males, for a period of 11 days. Each male was housed with 2 females; one of them was obtained from 2-months old male, and another one - from 3-months old male. Inbreeding was excluded (it means that despite $F_1$ males were bred with $F_1$ females, males and females were taken from different litters for each breeding cage – to exclude brother × sister mating; in reality all 3 animals in a breeding cage were taken from different litters to diminish the role of stochastic events). In control group all available females were used for breeding. In experimental group all females, obtained from 2-months old males, were used for breeding also. But we had more than enough females, obtained from 3-months old males, in the experimental group, and we took (selected) from them approximately ½ part – we took the animals, which were poor performers in two-way avoidance (the phenotype, more pronounced in all animals, obtained from 2-months old thyroxine-treated males). $F_1$ males, obtained from 3-months old males, were not used for breeding. $F_1$ males, used for breeding, were direct descendants of P males, tested in 2-way avoidance task using protocol with 0.8-sec delay (our basic protocol, see P in the Shuttle-box, Fig. S4; their $F_1$ male descendants consisted of 12 experimental and 14 control subjects). NMRI fostering (primiparous NMRI, purchased from RCC) and newborn pup numbering were used. Obtained progeny, born about 2001-07-22, was entitled as $F_2$-incross.

$F_1$ male adult mice (230 days of age), obtained from 2-month-old males, were housed with naïve nulliparous females (110 days of age, purchased from RCC at the age of 60 days) for a period of 11 days. Each male was housed with 2 females. All $F_1$ males, those were obtained from 2-month-old males and were used previously for breeding with $F_1$ females, were bred this time with naïve females. NMRI fostering (primiparous NMRI, purchased from RCC) and newborn pup numbering were used. Obtained progeny, born about 2001-08-21, was entitled as $F_2$-outcross.

Thus, in thyroxine study we had not only $F_2$ descendants, obtained by mating of $F_1$ males with $F_1$ females ($F_2$-incross), but also $F_2$ animals, obtained from $F_1$ males and naïve females ($F_2$-outcross). For $F_2$-outcross fostering we have chosen new primiparous NMRI (which were not used for $F_2$-incross fostering). Due to the existence of $F_2$-incross and $F_2$-outcross, $F_3$ descendents had to be obtained by incross of each independent subline ($F_3$-incross from $F_2$-incross or $F_3$-incross from $F_2$-outcross).

$F_2$-incross male adult mice (180 days of age) were housed with $F_2$-incross females (180 days of age) for a period of 11 days. $F_2$-outcross male adult mice (150 days of age) were housed with $F_2$-outcross females (150 days of age) for a period of 11 days. Each male was housed with one female. Thus, $F_2$-incross males were bred with $F_2$-incross females, $F_2$-outcross males – with $F_2$-outcross females, simultaneously. Inbreeding (brother × sister mating) was excluded. In the $F_2$ generation we had 8 groups using distinctions between control and experimental, males and females and incross and outcross. Inside each group of $F_2$ generation we selected for breeding equal numbers (8-9 mice) of the best and the worst performers in the 2-way avoidance task, considering mean performance value during 5-day training. Thus, each group of descendants was represented for breeding by two parts: good 2-way avoidance performers (8-9 mice) and poor 2-way avoidance performers (also 8-9 mice). Good performing males were bred with good performing females, whereas poor performing males were bred with poor performing females. In addition, the range inside formed subgroups was similar: the best performing male was bred with the best performing female, whereas good performing male was bred with good performing female; the worst performing male was bred with the worst performing female, whereas poor performing male was bred with poor performing female. In other words, special selection in accordance with 2-way avoidance performance took place: the best performers-males (25% of each group) were bred with the best performers-females (25% also), the poorest performers-males (25% of each group) were bred with the poorest performers-females (25% also), in range manner inside each mentioned above 25%. The same selection was done in control groups. Each above-mentioned 25% consisted of 8-9 mice. Thus, we had 8 groups of pregnant females: the 1-st distinction – "thyroxine" or "control", the 2-nd – "incross" or "outcross", the 3-d – "poor performers" or "good performers". NMRI fostering (nulliparous NMRI, purchased from RCC) and newborn pup's numbering were used (primiparous NMRI were not available at that period). Obtained progeny, born about 2002-02-08, was entitled as $F_3$ ($F_3$-incross from $F_2$-incross and $F_3$-incross from $F_2$-outcross).

### 1.2. Behavioural tests

Note that in the $F_1$ and $F_2$ generations all behavioural tests were done before animal breeding. $F_3$ was not bred at all. In the parental generation (P) thyroxine-treated males were tested after breeding period, whereas thyroxine-treated females were not bred.

#### 1.2.1. Emergence test

P-generation males at the age of 155 days and their sisters at the age of 158 days were tested in the Emergence test. $F_1$ males and females were tested in the Emergence test at the age of 50 days.

This test is called "Emergence", because in the frame of this test during 30-min video-tracking period an animal can hide in a small familiar "nest box" (and emerge from this nest box back).

About 3 days before Emergence test animals were distributed into individual home cages "Type 2" (polycarbonate, transparent; 267 × 207 × 140 mm). 24 hours before the test a small plastic non-transparent blue "nest box" was placed in each individual home cage (box 12 × 8 × 4 cm [L × W × H] with an opening about 8 × 4 cm; we use molecular biology ART tips boxes, "ART 10 Reach Molecular BioProducts",

external size 115 × 80 × 52 mm, internal size 115 × 80 × 35 mm, the opening 58 × 26 mm was hand made in the 80 × 52 mm wall, feels like polyethylene). For each particular animal just before the beginning of its testing its nest box was taken from the home cage and placed into 50 × 50 cm arena, surrounded by 37 cm walls, into one corner with distance of 5 cm from each wall, opening towards open space.

Mouse was introduced into above-mentioned arena with its own nest box installed (attached floor-to-floor by means of sticky tape loop) for a period of 30 min. Arena walls were made from non-reflective aluminium and were placed on a white non-slippery plastic surface to confine an open box 50 × 50 × 37 cm. Illumination was supplied by diffuse indirect room light (4 × 40 W halogen incandescent lamps; reflective rare part, the most popular E27 socket). These lamps provided intensity 30-33 lx in the middle of arena, 20-28 lx in the corners. Four units were operating concurrently and the paths of mice were recorded by means of a video camera suspended above four boxes. Animals were video-tracked at 4.2 Hz and 256 × 256 pixel spatial resolution using Noldus EthoVision 1.96 system (DOS-based program; Noldus Information Technology, Wageningen NL, www.noldus.com) and Windows 95 computer Pentium-133, 16 MB PC66 SDRAM. For each sample, the system recorded *xy* positions and object area. Raw data were then transferred to public domain software Wintrack 2.4, www.dpwolfer.ch/wintrack (Wolfer *et al*., 2001), for further analysis.

We have investigated all variables, typical for this test. For demonstration and discussion in this Supporting Online Material we have chosen the following variables (Wintrack variable names are given in square brackets):

1) Entries to box per 1 min - entries to box [ZIN] per 1 min total time;
2) Path traveled per 1 min visible (m) - path traveled [PTH], while active (V > 6 cm/s), per 1 min visible time;
3) Speed (cm/s) - average speed [SPD], while active (V > 6 cm/s) and visible (nest box is not transparent);
4) Time active (%) - % of time [ACT], while animal's speed was higher than 6 cm/s (time while an animal was visible was taken as 100%);
5) Time in box (%) - % of time [ZIT], while an animal was inside its nest box (total time was taken as 100%); chance level 3.70%.

### 1.2.2. Novelty test (including Open-field)

P-generation males at the age of 157 days and their sisters at the age of 160 days were tested in the Novelty test. $F_1$ males and females were tested in the Novelty test at the age of 52 days.

Total duration of Novelty test is 60 min for each animal.
The first 30 min of Novelty test can be used and analyzed as an Open-field test.

A mouse was introduced into square open-field arena 50 × 50 cm for a period of 30 min. Open-field walls were made from non-reflective aluminium (37 cm height) and were placed on a white non-slippery plastic surface to confine an open box 50 × 50 × 37 cm. Illumination was supplied by diffuse indirect room light (4 × 40 W halogen incandescent lamps) with intensity 30-33 lx in the middle of arena, 20-28 lx in the corners. Four units were operating concurrently and the paths of mice were recorded by means of a video camera suspended above four boxes. Animals were video-tracked at 4.2 Hz and 256 × 256 pixel spatial resolution using Noldus EthoVision 1.96 system (DOS-based program; Noldus Information Technology, Wageningen NL, www.noldus.com) and Windows 95 computer Pentium-133, 16 MB PC66 SDRAM. For each sample, the system recorded *xy* positions and object area. Raw data were then transferred to public domain software Wintrack 2.4, www.dpwolfer.ch/wintrack (Wolfer *et al*., 2001), for further analysis.

The number of entries to center zone per one minute was chosen for demonstration and discussion in this SOM. Center zone was determined as a square 40 × 40 cm minus circle d = 16 cm in the middle of arena. Thus, 50 × 50 cm arena was divided into three zones: center circle zone 8.04% (it will be "object zone" during the second 30 min of the test, when a new object will be introduced), center zone 55.96% and wall

zone 36.00%. Chosen indicator "number of entries to center zone" has better discriminative power than, for example, "number of entries to circle zone" or "number of entries to wall zone", however it is approximately equal (± 1 event) to the arithmetic sum of entries to the circle and to the wall zones.

After the first 30 min of test, while the mouse was still in the arena, into the centre of arena a small new object was placed for the next 30 min (we use 12 × 4 cm semi-transparent 50 ml Falcon tube, filled with water and placed vertically upside down; "CORNING 50 mL", d = 29 mm, total h = 116 mm, cover d = 35 mm, h = 12 mm). It was introduced by hand trying to avoid brisk movements. It was secured with special double-sided thick sticky tape, attached to the cover of Falcon tube in advance [Falcon tube was staying upside down on its cover]. For all animals new object was always introduced from the same side and with the same movements. Object zone was determined as a circle d = 16 cm in the middle of arena.

We have investigated all variables, typical for this test. For demonstration and discussion in this Supporting Online Material we have chosen the following variables:

1) Entries the object zone per 1 min - entries to the object zone [ZON] per 1 min total time (object zone was defined as a circle with d = 16 cm in the middle of arena);
2) Path traveled per 1 min visible (m) - path traveled [PTH], while active (V > 6 cm/s), per 1 min visible time [in this test visible time is equal to total time];
3) Speed (cm/s) - average speed [SPD], while active (V > 6 cm/s);
4) Time active (%) - % of time [ACT], while animal's speed was higher than 6 cm/s (total time was taken as 100%);
5) Time in the object zone (%) - % of time [ZOT], while an animal was inside object zone (total time was taken as 100%); chance level 8.04%.

### 1.2.3. Morris water maze

P-generation animals at the age of 175 days (females) and 180 days (males) and $F_1$ animals at the age of 105 days (males) and 120 days (females) were tested in the Morris water maze test.

The design of the apparatus followed the description given by Morris (Morris, 1984). It consisted of a white polypropylene circular pool of 150 cm diameter and 50 cm height, filled with water (15 cm deep, 24-26°C) made opaque by the addition of milk (1 liter, 4% fat). Distant visual cues for navigation were provided by numerous posters and black symbols on the walls. A wire-mesh platform (139 × 132 mm) painted white was placed 5 mm below the water surface, its center was 325 mm from the side of the pool. To avoid visual orientation prior to release, the mice were transferred from their cages to the pool in a white plastic cup from which they glided into the water, with the opening of the plastic cup toward the wall of the pool. Mice were released from eight symmetrically placed positions on the pool perimeter in a predetermined but not sequential order. They were left swimming either until they found a platform or until 120 s had elapsed. Finding the platform was defined as staying on the grid for at least 3 s. After staying on the platform for about 10 s, the mice were given the opportunity to climb onto a wire-mesh grid attached to a stick, which also was used to pick up swimming animals not having found the target. If mice did cross the platform without stopping (jumping immediately into the water), they were left swimming until they met the criteria above. After each trial, an animal was placed under infrared lamp for 3-7 min and allowed to warm up and dry off. Inter-trial times varied from 30 to 60 min (in our experiment it was slightly less than 60 min).

The entire procedure took 5 days. Each mouse did a total of 30 trials, six per day. In the morning (about 30 minutes before the beginning of each training session – training day) the animals were relocated from the animal facilities into the experimental room. They were removed from their individual home cages "Type 2" (polycarbonate, transparent; 267 × 207 × 140 mm) with saw dust bedding material and placed into individual cages "Type 1" (polycarbonate, transparent; 237 × 137 × 130 mm) with tissue paper. Three to five food pellets were placed into the cover of each cage "Type 1" to be available for a mouse, but bottle with water was not installed. Cages "Type 1" were placed in the vicinity of the water maze and the animals were able to see the same extra maze cues during inter-trail intervals as being in the water during training

trials. In the evening (after the end of each training session – training day) the animal were returned into their home cages "Type 2" and placed into the animal facilities. Illumination was supplied by diffuse indirect room light (4 × 40 W halogen incandescent lamps, socket E27). Computer and operator were in the same room, but remained hidden behind the wall of the pool. To monitor stress effects, the animals were weighed on the first and last day of the trials.

The position of the hidden platform remained fixed for the first 3 days (18 trials, acquisition phase). Afterwards, it was replaced in the opposite quadrant for 2 days (12 trials, reversal phase). The first trial of the reversal phase served as probe trial for spatial retention. Because spontaneous preference for one quadrant of the pool over others can never be excluded, four different versions of the protocol were used for place-navigation training, with goal positions and start points distributed over all four quadrants of the pool. The swim paths of the mice were recorded by means of a video camera suspended above the center of the pool. Animals were video-tracked at 4.2 Hz and 256 × 256 pixel spatial resolution using Noldus EthoVision 1.96 system (DOS-based program; Noldus Information Technology, Wageningen NL, www.noldus.com) and Windows 95 computer Pentium-133, i430TX, 16 MB PC66 SDRAM. For each sample, the system recorded *xy* positions and object area. Raw data were then transferred to public domain software Wintrack 2.4, www.dpwolfer.ch/wintrack (Wolfer *et al*., 2001), for further analysis.

We have investigated all variables, typical for this test (Wolfer *et al*., 1997). For demonstration and discussion in this Supporting Online Material we have chosen the following variables:

1) Escape latency (s) - escape time [TIM];
2) Swim path length (m) - path traveled [PTH], not counting noise while floating;
3) Swim speed (cm/s) - average speed [SPD], excluding periods of floating (speed slower than 6 cm/s, smoothing frame 1 s) and acute decelerations of 15 $cm/s^2$ or more. Trials 3 and 4 during the first training day had the best discriminative power concerning swim speed.

As a probe trial we have taken the first 60 s of trial 19 (the first trial of reversal phase). For the probe trial analysis we have chosen the following variables:

1) Target annulus crossings [XAT60S19] - crossings of target (trained) annulus (annulus was determined as a square 16 cm on side);
2) Adjacent annuli crossings / 2 [XAC60S19] - average of crossings over similar zones in adjacent quadrants. Crossings of adjacent annuli was chosen originally as a control, but in many cases it was found that well-trained animals (animals with good escape latency during several previous trials) have increased number of adjacent annuli crossings (as well as increased number of trained annulus crossings).

### 1.2.4. Two-way avoidance (Shuttle-box)

All mice (P, $F_1$, $F_2$ & $F_3$) were tested in the active avoidance task in an automated shuttle box ("Mouse Shuttle Box", Campden Instruments Ltd., UK, www.campdeninstruments.com ; this model is out of production now, however we can see its photograph in Buselmaier *et al*., 1981, p. 319). P males and females were tested at the age of 125 days, $F_1$ males – at the age of 135 days, $F_1$ females – 155 days. $F_2$ & $F_3$ males were tested at the age of 90 days, $F_2$ & $F_3$ females – at the age of 105 days.

In this behavioural task an animal has to learn to change compartment as a response to light presentation. Shuttle-box consists of two identical dark compartments. The light signal appears only in the compartment in which the animal is present. If the animal does move to the opposite dark compartment as a response to light presentation, the light switches off and a correct response is recorded for this trial. If the animal does not move to the opposite compartment during the first 5 sec of light presentation, it receives a negative reinforcement – the foot shock, which is applying together with the light and lasts up to 10 sec or until animal escape to the dark compartment. At the moment of escape the light and the foot shock switch off simultaneously. The next trial begins 5-15 sec later. An animal has to be exposed to 80 trails daily during 5 training days (*i.e*. 400 trials in total).

The apparatus consisted of a metallic chamber (length 270 mm, width 115 mm, height 130 mm) divided into two symmetric compartments (length 135 mm, width 115 mm, height 130 mm each) by the wall with 38 × 49 mm arch opening. An animal was introduced through Plexiglas front door (length 270 mm, height 130 mm), common for both compartments. An animal was able to move freely between two compartments through the arch opening in the separating wall. Two incandescent lamps (Type 687, 1 W, 24 V, 0.04 A) were incorporated into the two opposite walls of the chamber for light signal presentation. The floor of the chamber consisted of parallel bars of 2.4 mm diameter that were spaced 9 mm apart. This floor was used to deliver foot shocks generated by a shocker/scrambler unit that was able to adjust voltage automatically to maintain a constant current. Animal position was detected *via* mechanical movements of the floor. The floor was able to swing around horizontal axis passing through its center of gravity. Mouse movements from one compartment to another were able to induce small amplitude floor swinging (about 5-10 mm at the floor corner). Floor movements were detected with a help of hermetically sealed contact (Type MDRR-DT), attached to the corner of the chamber, and permanent magnet, attached to the corner of the swinging floor. Hermetically sealed contact was switching "ON" and "OFF" by above-mentioned magnet. Thus, the floor doubled as an animal position detector. The chamber was placed inside a sound proof ventilated enclosure. Four units (shuttle-boxes) were operating concurrently under control of IBM PC (processor IBM 6x86L PR200+, 64 MB PC66 SDRAM) running Windows 98 SE and custom developed software.

Each mouse was transferred directly from its home cage to the apparatus. At the beginning of the first session, mouse was allowed to habituate for 5 min. Then the animal performed 80 trials. Above-mentioned 5-min pre-session period was shortened to 2 min during the subsequent 4 days of training. Each trial begun with a 5 sec light stimulus (conditional stimulus – CS), delivered to the compartment where the animal happened to be at the particular moment. Then, with the light remaining on, a scrambled foot shock (unconditional stimulus – US) with 0.15 mA amplitude was applied for 10 sec (CS + US). The animal was able to terminate stimulus delivery at any point by moving to the opposite compartment. This light and current termination (or light termination only – in the case of correct response) had a 0.8 sec delay to make this task more difficult for DBA/2J. DBA/2J strain is known as a very good performer (Buselmaier *et al*., 1981). During each inter-trial interval, whose duration varied randomly between 5 and 15 sec, spontaneous crosses ("intercrosses") were recorded. The next trial has begun as soon as the inter-trial interval has elapsed. Pre-session activity, escape latencies and the number of escapes failed (negligible number in the case of DBA/2J) were recorded also. The number of correct responses in a training session served as an indicator of performance.

### 1.2.5. Establishment of spatial preference in the Intellicage™

$F_3$ females at the age of 150 days were tested for establishment of spatial preference in the Intellicage system (Newbehavior AG, Zurich, Switzerland; www.newbehavior.com ; see also Galsworthy *et al*., 2005).

The Intellicage is a group-housing cage that doubles a complete recording and testing apparatus. Animals can live in this system during many days like in standard home cages. The Intellicage records visits of individuals to "corners" by means of antennae which recognize the transponders located under the skin of the resident rodents. Each transponder has a unique code. The corners are effectively small (one-animal-only) operant chambers with access to water controlled by motorized doors. As such, the Intellicage provides resources for a variety of cognitive paradigms whilst offering an attractive alternative to repeatedly transferring animals into an alien arena for testing. Rather, the system allows animal exploration, learning and memory to be studied on an individual basis; yet within home cages, alongside peers, and with ample time for any paradigm.

The Intellicage (see Fig. S44a) is an apparatus design to fit inside a large standard rat cage ("Type 2000P", Tecniplast, Italy; www.tecniplast.it ; polycarbonate, transparent; 610 × 435 × 215 mm; this cage is bigger than "Type 4"). The apparatus itself provides four recording chambers that fit into the corners of the housing cage covering a right-angle triangular 150 × 150 × 210 mm area of floor space each. Access into the actual chambers is via an outer plastic ring (50 mm diameter) and then inner ring (30 mm diameter, 20 mm deep into outer ring). Such a width is designed to accommodate only one 10-40 g rodent. Furthermore

these rings contain circular antennae designed to register visits to the corner. The rodent entering this chamber encounters a choice between two 13 mm holes (one on the left, one on the right) which give access to water-bottled nipples. The holes can be closed by small motorized doors, thus barring access to either or both water bottles in each corner. Each hole-door combination was supplied with infrared nose-poke detector, and the door could be opened only when the mouse puts its nose into the vicinity of the door. In addition, each chamber had a small hole on its top through which a pulse of compressed air, controlled by the valve, could be delivered to the animal. This air pulse can be used as a negative reinforcement. Thus, if the animal comes into the chamber, its number can be read by the Intellicage and Intellicage can give the water to this subject by opening the door (only as a response to a nose-poke) or it can apply a negative reinforcement by the air pulse.

In addition to the Intellicage frame, each cage contained a small shelter in the center on which the animals could climb to reach the food (standard lab mouse chow, ad libitum). Each shelter consisted of 4 "Mouse house" units (right-angle triangular boxes made from translucent red plastic), attached to each other (Tecniplast, Italy, www.tecniplast.it ; see Fig. S44a).

A few days before the animals were introduced into the Intellicages, they were anaesthetized by inhalation of Methoxyflurane ("Metofane") vapour and subcutaneously injected with glass-covered microtransponders (11.5 mm length, 2.2 mm diameter; UKID System, Collison & Co., Riverside International Park, Caterall, Preston, UK), weighed and returned to their cages.

Our system consisted of two Intellicages which were served by Windows XP SP1a computer (PII-400, i440BX, 256 MB RAM, 18 GB SCSI HDD) running Newbehavior AG developed software (www.newbehavior.com). Thirty two female mice were injected with transponders and placed in these two Intellicages, 16 mice per cage, *i.e.* 8 experimental (4 "incross" and 4 "outcross") and 8 control (4 "incross" and 4 "outcross") were placed in each cage. Due to the injected transponder each animal had an individual number which can be read distantly in any Intellicage corner by means of built-in antenna.

The Intellicage has built-in capabilities for different protocols, but we used very simple one – two holes of any chamber led to identical bottles of water and an animal was able to use both of them without any difference. At the beginning of the experiment the animals were placed in the Intellicage system and they were able to investigate all corners, holes, nose-pokes and water delivery nipples without any negative reinforcement during more than 10 days (each animal was able to obtain water in any corner, though it was necessary to do a nose-poke to open the door). This period was used for adaptation of the animals to above-mentioned environment. During the next 20 days the training protocol for establishment of spatial preference was applied. Each animal was able to receive water in one corner and in the other 3 corners it received negative reinforcement (air pulse). Thus, there were "correct" corner visits, "opposite wrong", "right wrong" and "left wrong". Each corner was determined by the researcher as "correct" for 4 selected animals (1 experimental-incross, 1 experimental-outcross, 1 control-incross and 1 control-outcross) and as "wrong" for the others. In that way the corners were equally distributed between control and experimental animals and between their incross and outcross subgroups. The percent of correct corner visits was used as an indicator of performance, *i.e.*: "number of correct corner visits"/"number of all 4 corners visits" × 100. Daily averaged values are shown in the Fig. S44b. Chance level is obviously equal to 25%. Scores that are higher than 25% demonstrate the existence of spatial learning and spatial memory together with appropriate motivation.

### 1.3. Perfusion and Timm's staining for mossy fiber morphometry

P males and females were perfused at the age of 270 days.
$F_1$ males and females were perfused at the age of 250 days.
$F_2$-incross males were perfused at the age of 220 days, $F_2$-incross females – 225 days, $F_2$-outcross males – 190 days, $F_2$-outcross females – 195 days.
$F_3$ males were perfused at the age of 605 days, $F_3$ females – 555 days.

In the P generation we took for perfusion all animals those were tested in the two-way avoidance task using protocol with 0.8-sec delay (our basic protocol); P generation consisted of thyroxine-treated male parents and their female littermates (which were not used for breeding). Synchronous control was processed exactly by the same way as experimental animals in all generations.

In the $F_1$ generation we took for perfusion all descendants of above-mentioned males from P generation. Above-mentioned $F_1$ males were used as breeders to produce both $F_2$-incross and $F_2$-outcross generations.

In the $F_2$ generation we had 8 groups using distinctions between control and experimental, males and females and incross and outcross. Inside each group of $F_2$ generation we selected for perfusion equal numbers of the best and the worst performers in the 2-way avoidance task, considering mean performance value during 5-day training. Thus, each group of descendants was represented for morphological analysis by two parts: good 2-way avoidance performers and poor 2-way avoidance performers.

In the $F_3$ generation we used the same selection procedure as in the $F_2$. Above-mentioned selection was applied in the $F_2$ and $F_3$ to reveal possible correlations between 2-way avoidance performance and hippocampal intra- and infrapyramidal mossy fiber morphology.

### 1.3.1. Timm-staining

The mice were euthanized by an overdose of Nembutal injected intraperitoneally. They were then perfused transcardially with Ringer solution for 2 min (15 ml/min), followed by a 1% solution of sodium sulfide dissolved in 0.1 M phosphate buffer (pH 7.4) over 2 min (10 ml/min). This was followed by perfusion with a 3% solution of glutaraldehyde in 0.1 M phosphate buffer (pH 7.4), 4 min (20 ml/min). After removal, the brains were weighed to the nearest 2 mg, and kept in the glutaraldehyde fixative overnight. After 24-h storage in 20% sucrose solution in 0.1 M phosphate buffer, they were frozen and stored at -80°C for up to 6 months. The brains were sectioned horizontally in a cryostat (40 μm thickness), thaw-mounted in three parallel series, and developed for about 40-60 min in Timm's solution which consisted of 120 ml of a filtered 50% solution of acacia gum, 20 ml citrate buffer (containing 5.1 g citric acid + 4.7 g sodium citrate), 60 ml of 5% hydroquinone solution, and 0.5 ml of 17% $AgNO_3$, added immediately before development. Slides were developed at 30°C in the dark, allowing for complete visualization of the mossy fiber terminals and of the different strata of the region inferior, rinsed in tap water for 10 min, counterstained with methylene blue (two series) or neutral red (the rest one), dehydrated through alcohol, cleared in xylene and coverslipped with Merckoglas.

### 1.3.2. Morphometry

Quantitative analysis of the mossy fiber system was done on five horizontal sections spaced by 80 μm, plus a sixth one located 320 μm more ventrally. This level was included to assess better septotemporal gradients of the IIP-MF projection. Sampling started immediately below the most ventral extension of the septal pole of the hippocampus, the last section matching approximately the horizontal level of the anterior commissure. Both left and right hippocampi were analyzed. The levels for analysis were determined separately for each hemisphere in order to correct for tilted cutting planes.

Images of the hippocampi were digitized at the resolution 1280 × 1024 pixels, 8-bit gray, using a Videk Megaplus CCD camera mounted on Zeiss Axioplan. Images were analyzed using NIH Image 1.61 public-domain software on MAC OS 8.1 computer MAC PowerPC 7300/166, 192 MB EDO RAM, 2 GB SCSI HDD (above-mentioned PowerPC was upgraded during this study: 166 MHz processor was replaced by G3 processor [Sonnet Crescendo G3 300/512; 306 MHz, 512 KB cache], the second 4 GB SCSI HDD was added to have 6 GB HDD space in total and additional 128 MB EDO RAM were inserted to have 320 MB RAM in total). The mossy fiber distribution was then transformed into binary image. The gray level threshold for binary transformation was determined in four pairs of small circular masks along the suprapyramidal mossy fiber layer (SP-MF) and the hilar mossy fiber field (CA4-MF). For each pair of masks, one was placed within and the other outside the darkly stained mossy fiber zone. The average of the eight gray levels was then used to define the threshold level for obtaining the binary image of the mossy fiber distribution. Boundaries between the three mossy fiber fields (IIP-MF, SP-MF, and CA-MF) were

then drawn on the computer screen in order to measure the areas of each subfield. Note that these subfields refer only to the areas as covered by mossy fiber boutons. Thus, the hilar mossy fiber field referred to as CA4-MF does not include the full polymorphic layer of the dentate gyrus.

The morphometric scores for a given individual were than presented as a set of 12 local ratios - IIP-MF/SP-MF for each given section, separately for left and right parts. For data presentation and statistical analysis, above-mentioned ratios were averaged separately for left and right parts, and afterwards obtained values for left and right parts were averaged again.

### 1.4. EEG recording and analysis

In order to reveal deviance-related electrophysiological activity in $F_3$ incross and outrcross experimental and control mice we have chosen mismatch negativity (MMN) paradigm, adapted for mice, for EEG recording (Umbricht *et al.*, 2004, 2005).

Mismatch negativity (MMN) is an auditory event-related potential (ERP) that is generated when a stimulus violates the invariance or regularity of the recent auditory past. In the simplest paradigm this is the case when an infrequent stimulus (deviant) that differs in any physical characteristic such as frequency, duration, intensity or location is presented among repeatedly presented standard stimuli.

#### 1.4.1. Animals

DBA/2J males from $F_3$ generation, 32 mice in total, were used in this experiment: 19 experimental and 13 control males. Experimental group consisted of 10 “incross” and 9 “outcross” animals. Control group consisted of 5 “incross” and 8 “outcross” mice. Auditory event-related potentials in mismatch negativity paradigms were recorded in these mice at the age of 18 months.

#### 1.4.2. Animal preparation

Animals were anesthetized with ketamine-xylazine (87 mg/kg ketamine + 13 mg/kg xylazine, i.p.). Surgery was performed with a help of Stoelting Lab Standard™ Stereotaxis (www.stoeltingco.com) with World Precision Instruments Mouse Adaptor #502062 (www.wpiinc.com). Skin over the scull was removed, the scull surface was prepared using $H_2O_2$ (30%) and ethanol (91%). Five small burr holes of 0.8 mm diameter were drilled in the scull, two over the front, one at the back of the scull and one each over the estimated location of the auditory cortex (approximately 2.7 mm posterior to bregma, approximately 3.5 mm to the left/right of the midline). Five gold-plated screws were then carefully inserted. Prior to the operation miniature coaxial connectors (MK01/50G, www.distrelec.ch) had been coaxially soldered onto the head of these screws (screwdriver slot was made on the top of each connector before soldering procedure; each connector looked like very small gold-plated tube with internal diameter 0.5 mm and closed bottom; external diameter and length were 1.55 mm and 3.9 mm respectively; the bottom was soldered onto the screw head). Great care was taken not to injure the brain during the drilling or insertion of the screws. The screws were then fixated with dental cement (“Paladur”). Antiseptic “Merfen” (powder, produced by Novartis, 1 g contains 5 mg Chlorhexidingluconat, Benzoxoniumchlorid) was applied around dental cement immediately after operation. The animals were allowed to recover for 6-12 days from the operation before the first EEG recording (Fig. S46a-b). All event-related potentials shown in this Supporting Online Material (Figs. S45, S100-S108) were obtained from the electrode in the right auditory cortex – from the same hemisphere, in which the reference electrode was placed. Due to some reason the recording from the auditory cortex in the hemisphere of reference electrode produces slightly better MMN than the simultaneous recording from the auditory cortex in the opposite hemisphere. However event-related potentials obtained from both electrodes were used for statistical data analysis.

### 1.4.3. Auditory stimulation paradigms

To establish MMN-like activity we acquired ERPs with paradigms often used in human MMN studies. Mice were tested in 3 paradigms in which deviant stimuli differ in temporal position (Fig. S47a-c) and 4 paradigms in which deviants of different duration, different tone frequency and simultaneous difference in tone frequency and duration were presented (Fig. S47d-g). In addition, one paradigm with different inter-stimulus intervals (ISIs), used with C57BL/6J mice, is shown for explanation purposes (Fig. S45e-f, Fig. S46c-h). Note, that we discuss in our article only the last 4 paradigms (Fig. S47d-g) and in these 4 paradigms we focus our attention upon duration deviants only (the first string in the each paradigm in Fig. S47d-g). Standard and deviant tones were used in a mirror design (deviants in one trial served as standards in the next and vice-versa): accord 4 + 8 kHz *vs.* accord 3 + 6 kHz (frequency MMN), duration 50 ms *vs.* duration 150 ms (duration MMN) and mixed frequency-duration paradigm (simultaneous difference in frequency and duration, mixed MMN) were applied. In accordance with mirror design auditory stimuli were presented in 4 independent sets. In the first set the standard stimulus was accord 4 + 8 kHz, duration 50 ms, inter-stimulus interval (onset-to-onset) 500 ms; each $10^{th}$ stimulus was substituted for one of the 3 types of deviant stimuli: frequency deviant – accord 3 + 6 kHz, duration 50 ms; duration deviant – accord 4 + 8 kHz, duration 150 ms; mixed deviant – accord 3 + 6 kHz, duration 150 ms. The record duration was optimized to have 100 presentations of each type of deviant stimulus. In the second set standard stimulus was accord 4 + 8 kHz, duration 150 ms; in the third – accord 3 + 6 kHz, duration 50 ms; in the fourth – accord 3 + 6 kHz, duration 150 ms. Deviant stimuli in each set were chosen to be compatible with particular standard stimuli.

In the Fig. S47 the complete testing history is shown, because mid-latency electrophysiological events, including MMN, are in general history-dependent phenomena, which are somehow similar to animal behaviour. They are dependent not only on 1-2-second previous history, but on 2-3-hour previous history, at least. That is why the order of testing, the order of paradigms, is very important. We can not eliminate some testing episodes from the discussion just because they are not so interesting for us or just because they have not direct projections to our main conclusions.

Paradigms are shown in the order of their real presentation, with 3-5 min sound-free time intervals between consecutive paradigms (Fig. S46: paradigms c, d, e, f, g, h; Fig. S47: paradigms a, b, c, d, e, f, g). Scale bar 500 ms is valid for all shown paradigms. If a paradigm is shown in a few strings (Fig. S47), it means that the strings 1, 2, 3 or 1, 2, 3, 4 were run in the order 1, 2, 3, 1, 2, 3, 1, 2, 3, *etc*. or 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, *etc*. Note that paradigms shown in the Fig. S47 were applied to $F_3$ generation of thyroxine-treated mice, whereas paradigms in the Fig. S46 were used in C57BL/6J mice. Concerning C57BL/6J, the paradigms shown in the Fig. S46 were applied 6 months before paradigms from the Fig. S47.

### 1.4.4. Auditory stimuli

Auditory stimuli were generated with a RP1 system (Tucker-Davis Technologies, Alachua, FL, USA), amplified with PA5 amplifiers (Tucker-Davis Technologies, Alachua, FL, USA) and delivered through two electrostatic loudspeakers (Tucker-Davis Technologies, Alachua, FL, USA) mounted at an approximate height of 5 cm in the two short sides of the recording box. Stimuli were of an approximate SPL of 75 dB. Details of the stimuli are given in the Figs. S46 and S47.

### 1.4.5. EEG recording

EEGs were recorded in a plastic box (length 356 mm, width 183 mm, height 345 mm; internal size) placed into rectangular grounded copper box with a lid serving as a Faraday cage. An animal was insulated from the Faraday cage by the internal plastic walls and plastic floor, covered by tissue paper. Copper lid was slightly shifted for 1-2 cm to provide necessary ventilation and light (it was relatively dark in the chamber, but it was not absolute darkness). Miniature plugs were connected to the connectors mounted on the animal's head. For facilitation of connection procedure an animal was slightly anaesthetized by inhalation of Methoxyflurane ("Metofane") vapour. Two electrodes over the back and front of the scull served as ground electrodes, one electrode over the front of the scull (right) as reference electrode, one electrode over the left auditory cortex and one electrode of the right auditory cortex served as active electrodes (Fig.

S46a,b). The miniature plugs were connected to a swivel joint connector allowing the animal free range of movement. The EEG was amplified with a Siemens Mingograf 21 EEG amplifier (band pass filter 0.1-200 Hz, 50 Hz notch filter) and digitized with a Biopac M100 (Biopac, Goleta, CA, USA) system at a rate of 500 Hz. Today this information about Siemens Mingograf 21 can be interesting only as a historical fact, because very nice portable EEG recorder (Neurologger®) for mice is available now. It is so small that it can be hold on the mouse head (see www.vyssotski.ch/neurologger2.html and Vyssotski *et al*., 2006, 2009); for additional information contact info@evolocus.com. The raw EEG was continuously stored on a computer disk along with digital stimulus tags (Dell OptiPlex GX110 tower, PIII 733 MHz, i810E, 384 MB PC100 SDRAM, three IDE HDDs 40 + 60 + 120 GB, Windows 2000 Professional).

Data processing was performed off-line with the help of Neuroscan software Version 4.2 (Compumedics, ElPaso, TX, USA) using Dell Precision 650 workstation (single Xeon 2.8 GHz, 533 MHz FSB, 512 KB cache, E7505, 1 GB PC2100U ECC DDR, LSI SCSI, two 147 GB 10K RPM U320 HDDs) with Windows 2000 Professional. For the analysis of the auditory ERPs, epochs were constructed that consisted of a 100 ms pre-stimulus baseline and 800 ms post-stimulus interval (800 ms from the onset of stimulus). Epochs in which amplitudes exceeded ± 100 μV at the two active electrodes were excluded from further averaging. Data obtained in thyroxine study were not detrended. Following artifact rejection, epochs were averaged off-line for each animal and paradigm separately. The right recording channel was used for further analysis, because it provides slightly better MMN in comparison with the left channel, may be due to the presence of reference electrode in the right hemisphere.

### 1.4.6. EEG statistical analysis

Software filtering was applied with a help of Neuroscan 4.2: Bandpass, zero-phase shift filtering (no latency shifts) was performed with low pass filter setting of 15 Hz (24 dB) and a high pass filter of 0.1 Hz (24 dB). Data were not detrended. Filtered data are presented in the Fig. S45g-h. Data without above-mentioned software filtering are shown in the Fig. S45a-f. Data shown in the Fig. S45a-b (as well as in the corresponding filtered graphs, *e.g.* Fig. S45g) were obtained in two separate recording sessions with the same animals to have more smooth curves (for visual analysis only). The second recording session, applied 42 days after the first one, was not used for statistical data analysis. Time interval associated with statistical significance $P < 0.05$ (Fig. S45b) was identified using filtered data and t-test for independent samples. Experimental groups (Incross and Outcross) were taken as a single group and compared with Control. Significance $P < 0.05$ was observed in the time interval from 125 ms till 218 ms.

Discussed significance was revealed by the following way.

1) Epoch -100 800 ms, baseline correction and artifact rejection (described above) were applied to each record (4 records correspond to 4 paradigms for each animal);

2) All ERPs to standard stimuli were averaged (separately for each animal and separately for each paradigm) and all ERPs to duration deviant stimuli were averaged also (separately for each animal and separately for each paradigm);

3) Averaged standard ERP and averaged duration deviant ERP were filtered as described above (bandpass, zero-phase shift filtering with low pass filter setting of 15 Hz (24 dB) and a high pass filter of 0.1 Hz (24 dB)).

4) Averaged filtered standard ERP was subtracted from averaged filtered duration deviant ERP (separately for each animal and separately for each paradigm) to have averaged filtered "difference wave" (MMN);

5) Inside each paradigm averaged filtered difference waves of all experimental animals were averaged again and inside each paradigm waves of all control animals were averaged also to have "grand-average paradigm-1 experimental" (GAP-1E), "GAP-2E", "GAP-3E", "GAP-4E" and "grand-average paradigm-1 control" (GAP-1C), "GAP-2C", "GAP-3C", "GAP-4C".

6) To be sure that obtained difference was produced by all 4 paradigms each paradigm was taken as a unit of analysis. *I.e.* experimental group data were presented as 4 curves (paradigms) ($n$ = 4) and control group data were presented as 4 curves (paradigms) ($n$ = 4). The time interval with significant difference ($P < 0.05$) was identified using t-test for independent samples.

Above-mentioned analysis of mirror-designed paradigms was applied to reveal the difference between control and experimental groups that is specifically associated with deviance-related electrophysiological activity. The result of this analysis is shown in the Fig. S45b.

In addition to above-mentioned type of statistical analysis, several types of parametric and non-parametric statistical tests were applied to the same data set, where an individual animal was taken as a unit of analysis. These tests are: non-parametric Mann-Whitney U-test, parametric factorial ANOVA and parametric main effects ANOVA. All of them have brought qualitatively similar results, however the time window with statistical significance ($P < 0.05$) occurred to be slightly different. These results are described in the Notes of this SOM (Notes, section 1.5. “EEG recording in descendants of thyroxine-treated mice: statistical analysis”, pp. 56-59).

## 2. Morphine experiment

### 2.1. Animals and chemicals

Wistar rats were obtained from the National Research Center of Oncology (“Oncological Center”, Moscow, Russia). We received 30 males (20 for experimental group and 10 for control one) and 30 females (for breeding 1 female × 1 male). All rats in all morphine experiments were housed in groups 5-10 animals per cage – in the cages “Type 4” (polypropylene, non-transparent /“translucent or opaque”/; 595 × 380 × 200 mm). For other purposes (breeding, housing of a pregnant female, housing of female with pups) cage “Type 3” was used (polypropylene, non-transparent /“translucent or opaque”/; 425 × 266 × 155 mm). Normal day-light cycle was used. The animals were tested during the light phase.

Morphine sulfate and naloxone were generously provided by the National Research Center on Drug Abuse (“Narcological Center”, Moscow, Russia).

#### 2.1.1. Drug treatment

Male Wistar rats (prospective parents) used in our study were born about 1996-01-22. Starting from the age of 42 days (P42; body weight 197 ± 20 g, mean ± SD) 20 males (experimental group) were injected i.p. with morphine (dissolved in 0.9% NaCl, morphine concentration 10 mg/ml) during 38 days in accordance with the following protocol. During the first 7 days rats received morphine twice daily (morning-evening, 8 hr between; mg/kg): 5-10, 15-15, 20-20, 25-30, 35-40, 45-50, 55-60. Next day rats received 60 mg/kg of morphine in the morning and 6 hours later they were injected i.p. with naloxone (2 mg/kg, concentration 2 mg/ml) to induce early in life acute morphine withdrawal. Next day after naloxone injection rats received one morphine injection 60 mg/kg in the middle of a day (*i.e.* 24 hours after naloxone administration). During the rest 29 days of morphine treatment rats received 60 mg/kg of morphine twice daily (8 hr between) during Monday-Friday, and during Saturday-Sunday they received 60 mg/kg of morphine daily in the middle of a day. Morphine treatment has started 1996-03-04 and it has ended 1996-04-10. Control animals (10 males) were left undisturbed. All these animals were entitled as P.

### 2.1.2. Breeding protocol

During the last 6 days of morphine treatment P males (age P74 - P79, 1996-04-05 - 1996-04-10; *i.e.* days 33 - 38 of morphine treatment) were housed individually with drug-naive adult (75 days of age) nulliparous Wistar females (1 female × 1 male) in the cages "Type 3" (425 × 266 × 155 mm). At the end of morphine treatment (and breeding period) all males were removed from their females. This breeding has started 1996-04-05 and it has ended 1996-04-11. At birth the litters were left undisturbed (no cross-fostering and no litter-culling). Obtained progeny, born about 1996-04-28, was entitled as $F_1$-1 (the first generation, the first brood).

Sometimes in the text we use "$F_1$" instead of "$F_1$-1", but we never use "$F_1$" for generalized description of $F_1$-1 and $F_1$-2 or for something like that.

The next generation ($F_2$), suitable for investigation, had to be obtained from non-tested $F_1$. Thus, morphine-treated and control parents were bred second time to have $F_1$-second brood ($F_1$-2) generation.

P males at the age 175 days (95 days of morphine withdrawal) were housed (second time) individually with familiar females (175 days of age, primiparous; exactly the same mating pairs as earlier) for a period of 8 days. At the end of breeding period males were removed from females; at birth the litters were left undisturbed. Obtained progeny, born about 1996-08-08, was entitled as $F_1$-2 (the first generation, the second brood).

$F_1$ males at the age of 85 days were housed individually with $F_1$ females (85 days of age also) for a period of 8 days (one male with one female). It was incross, but without inbreeding (each mating pair consisted of experimental or control $F_1$ animals, but these animals were taken from different litters to avoid brother × sister mating). Practically all available $F_1$ males and females were used for breeding, without specific selection. At the end of breeding period males were removed from females; at birth the litters were left undisturbed. Obtained progeny, born about 1996-11-25, was entitled as $F_2$ (second generation; this $F_2$ might has been entitled as "$F_2$-incross", but we had not any $F_2$-outcross generation in morphine study, and, so, the "incross" extension was not introduced here).

## 2.2. Behavioural tests

### 2.2.1. Hole-board

$F_1$-1 males were tested in the hole-board test at the age of 39 days, $F_1$-1 females – 45 days, $F_1$-2 males – 54 days, $F_1$-2 females – 61 days.

Each $F_1$-1 animal was tested in this test twice: in the first 3-min session and 24 hours later in the second 3-min session. The apparatus consisted of a plywood open-field (380 × 390 mm) surrounded by vertical walls 270 mm height (one Plexiglas transparent wall and 3 plywood non-transparent). The floor (also plywood) was divided into 9 equal squares (130 × 130 mm each) by two pairs of X-crossed lines (visible for researcher). In addition, the floor had 9 holes (d = 30 mm), each of them made in the middle of each square. The floor had thickness about 3 mm. The device was equipped with the second floor placed 30 mm under the main one. Standard day-light illumination about 25-32 lx was used. The subjects were individually tested in 3 min sessions. Activity was recorded as the number of holes visits (holes investigation incidents), the number of center visits, the number of rearings, the number of grooming events and the number of lines crossed (locomotor activity). All these 5 indicators of behaviour were recorded by two observers in real time.

Each $F_1$-2 animal was tested in this test 3 times in 3 consecutive days: one 3-min session daily. In the $F_1$-2 animals only horizontal locomotor activity (the number of lines crossed) was recorded by one observer.

### 2.2.2. Passive avoidance (Step-down)

$F_1$-1 males were tested in step-down test at the age of 51 days, $F_1$-1 females – 62 days, $F_1$-2 males – 58 days, $F_1$-2 females – 67 days.

The apparatus consisted of an open squared arena (400 × 400 mm) surrounded by vertical Plexiglas walls 510 mm height. One front wall was transparent and the rest 3 – non-transparent. Wooden pine "cube" (parallelepiped – the platform 64 × 112 mm, 62 mm height) was placed in the middle of arena and attached to the brass rod (d = 4.5 mm approximately) floor. Standard day-light illumination about 25-32 lx was used. The negative reinforcement was a foot shock 0.35 mA, 50 Hz, supplied through 4 diodes (4-pole-system; see Bures *et al*., 1983). The time course of current application was controlled manually – short pulses of current were supplied regularly to keep animal in move. During training session the animal was placed on the platform accurately and step-down and step-up latencies were recorded. If the animal was sitting on the platform more than 180 sec, 180 sec was taken as step-down latency. But then in this case the animal was removed from the platform to the rod floor by hand and its step-up latency was measured. The animal was removed from the chamber when it was able to find the platform and it was sitting on it during 5 sec or when it was unable to find the platform during 60 sec. Each animal was trained during 3 days, one training session per day.

## 2.3. Drug-related tests

### 2.3.1. Tail withdrawal

Males P were tested in tail-withdrawal test at the age of 80 days (immediately after the end of breeding period, 1996-04-11), then P males were tested second time at the age of 11 months, $F_1$-1 males – at the age of 65 days, $F_1$-1 females – 70 days, $F_2$ males – 65 days, $F_2$ females – 60 days of age, $F_1$-2 males – 140 days, $F_1$-2 females were not tested.

All available P and $F_1$ animals were tested in this test, but in the $F_2$ generation the first born 120 males (80 experimental and 40 control) and 120 females (80 experimental and 40 control) were tested. The distal part of the tail (1/4-1/3 of its length) of a lightly restrained animal was dipped into circulating water thermostatically controlled at 56 ± 0.2°C. Latency to respond to the heat stimulus, by a vigorous flexion of the tail, was measured to the nearest 0.1 sec, cutoff latency – 15 sec. The test was done once before i.p. 10 mg/kg morphine injection (baseline latency) and 15, 30, 45, 60 and 90 min after. During this procedure an animal was packed into a peace of cotton towel (its head and tail remained visible) and was held by free hand. The latency of withdrawal was taken as an indicator of pain sensitivity.

### 2.3.2. Hot plate

The animals were tested in hot plate test practically simultaneously with tail-withdrawal procedure, using the same single morphine injection and time intervals (at each time point an animal was tested first in tail-withdrawal test and then (immediately after) – in hot-plate test). Hot-plate device consisted of metal 300 × 300 mm chrome-covered square plate, surrounded by vertical transparent Plexiglas walls 246 mm height. Immediately after each tail withdrawal measurement an animal was placed on the hot-plate (52 ± 0.4°C). The latency of paw licking or jumping was taken as an indicator of pain sensitivity. Cutoff latency was chosen 45 s for P generation and 30 s for $F_1$-1, $F_2$ and $F_1$-2. Results of this test are shown and discussed in SOM only, because these results occurred to be less impressive than tail-withdrawal data.

### 2.3.3. Dependence (5.5-day morphine treatment, naloxone-induced weight loss)

Physical dependence was investigated in P males at the age of 95 days, in $F_1$-1 males – at the age of 69 days, in $F_2$ males – 68 days, in $F_1$-2 males – 141 days. Age is given for the first day of 5.5-day morphine treatment.

$F_1$ and $F_2$ animals were injected i.p. with morphine twice daily during 5.5 days in the following way (morning-evening, 12 hr interval; mg/kg): 10-10, 20-20, 30-30, 40-40, 50-50, next day – 60 mg/kg of morphine in the morning and 6 hours later they were injected i.p. with naloxone 2 mg/kg (concentration 2 mg/ml). Weight of each animal was measured to within 1 g using Sartorius L610D at the following time points: i) before morphine treatment, ii) in the middle of the 3-d day, iii) before naloxone injection and iiii) 6, 12, 18, 24, 30, 36, 42 hours after naloxone injection. Naloxone-induced weight loss was taken as an indicator of physical morphine dependence.

Different protocol was used for parental generation. Experimental group of parental generation P was exposed to chronic 40-day morphine treatment, whereas control P group was left untreated. After the last 60 mg/kg morphine injection, 6 hours later, experimental group received naloxone i.p. 2 mg/kg injection (concentration 2 mg/ml). Weight of each animal was measured as described above.

### 2.3.4. The influence of morphine injection on locomotor activity

Locomotor activity was recorded in accordance with method developed by Thiel, Barnes & Mrosovsky in 1972 (Thiel *et al*., 1972), which was modified by us with respect to data acquisition hardware and software.

This test was applied to $F_1$-1 males at the age of 76 days (9 experimental and 6 control animals).

After 2 mg/kg naloxone injection, 45 hours later, each animal was placed into individual cage "Type 3" (polypropylene, non-transparent /"translucent or opaque"/; 425 × 266 × 155 mm). Small mechanical movements of the whole cage were measured with a help of piezoelectric sensor. The system consisted of 16 individual cages and all these cages were placed into special rack which allowed to measure small mechanical movements of each particular home cage. During the first 3 hours baseline locomotor activity was measured. Afterwards each animal was injected with morphine 60 mg/kg or saline, i.p., and its locomotor activity was recorded during additional 9 hours (it means that this 60 mg/kg morphine injection was done exactly 48 hours after 2 mg/kg naloxone administration; total duration of the record – 12 hours). In the graphs (Figs. S81-S82) we can see 1-min averaged values (morphine offspring – 9 rats, other groups – 6 rats per group). Individual locomotor activity during the first 3 hours (averaged value) was taken as a unit of measurement for each particular animal. For statistical analysis locomotor activity of each animal during 9 hours after injection (measured with respect to its activity before injection) was averaged and non-parametric Mann-Whitney U-test was applied.

Two additional independent control groups (6 and 6 rats) were tested at the age of 78 days with saline injection (instead of morphine). The first one consisted of control Exp.1 $F_1$-1 males, pretreated with morphine during 5.5 days and received naloxone 2 mg/kg injection, but, instead of morphine 60 mg/kg injection done 48 hr after naloxone administration, they have received saline injection during their locomotor activity recording 96 hr after naloxone administration. The second additional control group consisted of control Exp.1 $F_1$-1 males, which were not pretreated with morphine, never received naloxone (*i.e*. they are drug-naïve), they have received saline injection during their locomotor activity recording synchronously with above-mentioned the first additional control group. Both these controls are very important for interpretation of results of morphine 60 mg/kg injection.

Concerning technical details, in this system each cage was placed on three foam-rubber props. Each cage was equipped with 3 short (about 5 mm) spikes for improved mechanical contact with foam-rubber props to avoid horizontal slippage. These props were attached to the 180 × 335 mm base. Each prop had sickness 20 mm without load (*i.e*. without cage installed). The first prop had size 70 × 88 mm and it was placed under the right front corner of the cage (the right front corner of this prop was placed exactly over the right front corner of the base). The second prop 44 × 70 mm – under the left front corner (prop's left front corner – over the left front corner of the base). The third prop 70 × 88 mm – in the middle between two back corners of the cage (prop's back side - over the back side of the base). The longest side of the cage was parallel to

the longest side of the base and to the longest side of each prop. Water in the glass bottle (250 ml) with glass nipple was placed in the right front corner of the cage, on the grid lid, as usual. Food was provided as usual. Mechanical movements of the cage were measured at the top of the lid in the left-right direction. For this purpose metal plate 50 × 50 mm was attached to the top of the lid over the center of the cage. This metal plate was covered by glued tissue and needle of piezoelectric sensor had a contact with this surface. Piezoelectric sensor was attached to the rack by means of Y-shaped wire rocker arm which provided possibility of vertical movements without possibility of horizontal movements. All 3 ends of Y-shaped arm were placed in horizontal plane in the form of regular triangle with side about 130 mm. Two ends of this Y-shaped arm had point contacts with rack frame and the third one was holding piezoelectric sensor with its needle. The pressure of the needle on the above-mentioned tissue-covered plate was determined by the weight of piezoelectric sensor and the weight of the wire rocker arm. This construction was developed to provide good mechanical contact between cage and piezoelectric sensor and to avoid sensor damage in the case of high amplitude movements of the cage. The slippage between tissue-covered plate and needle was absent during small amplitude movements, but it was observed during high amplitude movements.

System had 16 independent cages "Type 3" and was served by DOS-computer Olivetti M240, XT 8086/87, 10 MHz, 640 KB RAM, 71.3 MB MFM HDD Micropolis 1335 (as a replacement of standard Olivetti 20 MB "Lexicon" HDD; both are old-fashioned MFM HDDs, *i.e.* non-IDE and non-SCSI). This computer was equipped with analog-digital converter board LA-70M3 (16-channel board for ISA-8, manufactured by ZAO "Rudnev-Sheliaev", www.rudshel.ru ). Very similar board, the next revision – LA-70M4, can be found here: www.rudshel.ru/show.php?dev=48 . Four 4-channel operant amplifiers OP482GP Analog Devices ($k$ = 100) were placed before LA-70M3 to avoid signal penetration from one channel to another. All 16 inputs of four OP482GP were connected to 16 piezoelectric sensors GZP-311 (*ГЗП-311*) by thin coaxial cables. Piezoelectric sensor GZP-311 was manufactured in Russia and can be seen here: http://oldradio.onego.ru/remast2.htm ). Data were digitized by means of LA-70M3 at the rate of 50 Hz per channel to avoid 50 Hz disturbance from power lines. 12-bit accuracy plus 4 bit for channel number (*i.e.* two bytes per reading) have produced 65.9 MB during 12 hr. Only raw data could be recorded to HDD using this system, because this XT 8086/87 (10 MHz) was not powerful enough to do any real-time data processing.

### 2.4. Synaptophysin brain distribution analysis

Synaptophysin (p38) is a synaptic vesicle protein thought to be associated with neurotransmitter release which has been extensively used to mark synaptic terminals and estimate synaptic density (Greengard *et al*., 1993; Belichenko *et al*., 1996, 2000).

Synaptophysin (p38) distribution was measured in $F_2$ males only. Four experimental males and four control males were taken for this analysis at the age of 75 days (age of perfusion). We have selected 8 rats (4 experimental and 4 control ones) that had shown practically normal phenotype in the drug-related tests, *i.e*. without detectable difference between selected experimental and control subjects (Fig. S60c-d). Each above-mentioned animal (experimental and control) has received two morphine injections (10 mg/kg, i.p.) 9 and 8 days before perfusion. During 5 days before perfusion animals were living in the individual cages "Type 3" (polypropylene, non-transparent) and received saline injections (0.25-0.65 ml per rat) twice daily (morning-evening). Next day (day 6 in cages Type 3) rats were taken for perfusion between 12:25 a.m. and 2:35 p.m.

In addition to these 8 rats, 8 other originally control males (not mentioned in our article, but shown in the Table S4) were injected i.p. with morphine during 6 days (morning-evening, 12 hr interval; mg/kg: 10-10, 20-20, 30-30, 40-40, 50-50, 60-60). Next day 4 of them were taken for synaptophysin distribution analysis (at the age of 75 days), whereas the rest 4 were taken for synaptophysin distribution analysis after 6 days of morphine withdrawal.

Synaptophysin level was investigated in somatomotor cortex, layer 3 (SMC-III) and layer 5 (SMC-V), somatosensory cortex, layer 3 (SSC-III) and layer 5 (SSC-V), nucleus accumbens (NA), hippocampus (HIPP), nucleus caudalis (NC) and ventrolateral thalamic nuclei (THAL).

Rats were deeply anaesthetized by nembutal 60 mg/kg i.p. injection and perfused with 4% paraformaldehyde in 0.1 M phosphate buffer (PB), pH 7.4. Their brains were removed and placed in cold (+4°C) PB until further processing (which took place 2 weeks later). 0.1-mm frontal sections were prepared with an Oxford vibratome and stored in PB at +4°C until immunofluorescence incubation. Immunofluorescence with synaptophysin (p38) antibody and quantitative analysis of p38 immunoreactivity (IR) and autofluorescent lipofuscin granules were performed as previously described (Belichenko *et al*., 1996). Briefly, sections were incubated for indirect immunofluorescence using rabbit anti-p38, (synaptophysin), raised in rabbits, dilution 1:2000. After preincubation in 5% non-fat milk in PB saline (PBS), the primary incubation was carried out overnight (+4°C) under free floating conditions. The sections were then rinsed in PBS (20 min, three changes) and incubated for 1 h at room temperature with biotinylated goat anti-rabbit IgG (1:200; Vector Labs, Burlingame, CA, USA) and rinsed again with PBS (20 min, three changes). After further incubation with fluorescein isothiocyanate (FITC)-conjugated streptavidin (1:200; Amersham, Little Chalfont, UK) for 1 h at room temperature and rinsing the sections were mounted on a microscope glass slide and coverslipped with glycerol in PB. All sections to be compared were prepared simultaneously under the same conditions. In control sections, where the primary antibody was omitted, immunofluorescence was not observed, but lipofuscin fluorescence was strongly present.

Sections were observed and scanned in a BioRad MRC-600 CLSM (Hercules, CA, USA) attached to a Nikon FXA fluorescence microscope, employing an argon/krypton mixed gas laser with exciting filters for FITC and lipofuscin (488 DF10) and for lipofuscin (568 DF10). The quantitative analysis of p38-IR and lipofuscin was carried out on single confocal images using Pixel Anatomy software (Belichenko *et al*., 1996). Three steps were carried out for every confocal image. As a first step, background fluorescence was estimated by analyzing the distribution of the pixel intensities in the image areas that did not contain any immunolabelled objects (the background threshold). This background was subtracted by setting the baseline of pixel intensities to the background value. In the second step, every image scanned for lipofuscin was subtracted in a pixel-by-pixel manner from the paired image of the same area scanned for p38-IR and lipofuscin. In the third step, an arbitrarily outlined polygon was chosen for quantification of p38-IR and lipofuscin in an image area without artifacts. In the polygon area chosen for analysis the relative area (%) of immunolabelled pixels with an intensity value above background was calculated. Data were statistically processed with StatView (SAS Institute, Cary, NC, USA). A non-parametric two-way analysis of variance (ANOVA) was used and $P < 0.05$ was considered significant.

**2a. Morphine experiment - 2**

**2a.1.1. Animals and chemicals**

Wistar rats were obtained from the National Research Center of Cardiology (“Cardiological Center”, Moscow, Russia) 1997-12-01 at the age of 51 days (born about 1997-10-11). We received 35 male rats: 12 male rats for chronically morphine treated experimental group, 12 – for experimental group injected with morphine before breeding only once (25 mg/kg) and 11 – for control group. Female rats (35 in total) were obtained from the same supplier for breeding with males (1 female × 1 male). Similar to the first morphine experiment, all rats were housed in groups 5-10 animals per cage – in the cages “Type 4” (polypropylene, non-transparent /“translucent or opaque”/; 595 × 380 × 200 mm). For other purposes (breeding, housing of a pregnant female, housing of female with pups) cage “Type 3” was used (polypropylene, non-transparent /“translucent or opaque”/; 425 × 266 × 155 mm). Normal day-light cycle was used. The animals were tested during the light phase.

Morphine sulfate and naloxone were generously provided by the National Research Center on Drug Abuse (“Narcological Center”, Moscow, Russia).

### 2a.1.2. Drug treatment

Male Wistar rats (prospective parents) used in this experiment were born about 1997-10-11. Starting from the age of 58 days (P58; 12 males, body weight 235 ± 27 g, mean ± SD) they were injected i.p. with morphine during 46 days in accordance with the following protocol. During the first 6 days rats received morphine twice daily (morning-evening, 12 hr interval; mg/kg): 10-10, 20-20, 30-30, 40-40, 50-50, 60-60. During the rest 40 days of morphine treatment rats received 60 mg/kg of morphine twice daily with 12 hr time intervals. Morphine treatment has started 1997-12-11 and it has ended 1998-01-26. An independent group of males at the age of 102 days (P102; 12 males, body weight 403 ± 58 g, mean ± SD) was injected i.p. with morphine (25 mg/kg) only once 1998-01-20 (24 hours before the beginning of the mating period). Control animals (11 males) were left undisturbed. All these animals were entitled as Exp.2 P.

### 2a.1.3. Breeding protocol

During the last 6 days of morphine treatment Exp.2 P males (age P102-P107, 1998-01-21 - 1998-01-26; *i.e.* days 42-47 of morphine treatment) were housed individually with drug-naive adult (about 90 days of age) nulliparous Wistar females (1 female × 1 male) in the cages "Type 3" (425 × 266 × 155 mm). Male rats from the independent group, which received 25 mg/kg morphine only once, 24 hours after morphine injection were placed to drug-naïve females (90 days of age) for a period of 6 days (1 female × 1 male). All 3 groups of males (2 experimental and 1 control) were bred with females synchronously at the age of P102-P107, 1998-01-21 - 1998-01-26. At the end of morphine treatment (and breeding period) all males were removed from their females. At birth, the litters were left undisturbed (no cross-fostering and no litter-culling). Obtained progeny, born about 1998-02-12, was entitled as Exp.2 $F_1$-1 (the second experiment, the first generation, the first brood).

Sometimes in the text we use "Exp.2 $F_1$" instead of "Exp.2 $F_1$-1", but we never use "Exp.2 $F_1$" for generalized description of Exp.2 $F_1$-1 and Exp.2 $F_1$-2.

Exp.2 P males at the age 150 days (43 days of morphine withdrawal) were housed (second time) individually with familiar females (about 140 days of age, primiparous; exactly the same mating pairs as earlier) for a period of 8 days (1998-03-10 - 1998-03-17). At the end of breeding period males were removed from females; at birth the litters were left undisturbed. Obtained progeny, born about 1998-04-04 (exactly: 1998-04-01 - 1998-04-08), was entitled as Exp.2 $F_1$-2 (the second experiment, the first generation, the second brood).

## 2a.2. Behavioural tests

### 2a.2.1. Fear conditioning (Freezing)

Exp.2 $F_1$-1 males were tested in fear conditioning test at the age of 51 days, Exp.2 $F_1$-1 females – at the age of 70 days. Above-mentioned ages are given for the first day of the test. Exp.2 $F_1$-2 animals were not tested in this test. For each animal the total duration of testing period in this test was 10 days and consisted of 5 experimental days: Day1, Day3, Day6, Day8 and Day10. Days in between (except Day5) were used for testing of another portion of animals using identical protocol. Five experimental days can be represented in the form of the following table:

| | | |
|---|---|---|
| Day1 | 3 min + 3 min + 3 min | Training-1 (spaced or massed) – Current 0.28 mA |
| Day3 | 6 min | Context test-1 |
| Day6 | 3 min + 3 min + 3 min | Training-2 (spaced or massed) – Current 0.48 mA |
| Day8 | 6 min | Context test-2 |
| Day10 | 6 min | Cue test (3 min before sound and 3 min during sound) |

Each group was subdivided into two equal parts – one part received spaced training, another one – massed training.

Two chambers were used in these experiments: a Biotest RK-5301 (made in USSR: "Биотест РК-5301") and a round chamber prepared in the laboratory. The Biotest RK-5301 was illuminated with a daylight lamp, mounted on top and separated from the chamber by semi-transparent white plastic. It was a rectangular chamber of 25 × 30 × 25 cm (L × W × H) with an electrically conducting floor made of metal rods 3 mm in diameter and located 10 mm apart; a loudspeaker was mounted on one of the chamber walls. This chamber had transparent Plexiglas front door, its other walls were metal and painted white. The round chamber was a Plexiglas cylinder of size (d = 20 cm, h = 32 cm), lying on its side.

Animals were transferred from the animal room to the experimental room and back using standard cage "Type 3" (polypropylene, non-transparent /"translucent or opaque"/; 425 × 266 × 155 mm) when animals were to be trained and tested to contexts, and in the same cage covered completely by dense tissue when animals were to be tested with the conditioned signal.

Table above shows the overall experimental scheme. The experiment started with spaced or massed training (two independent subgroups were formed from each experimental or control group). Testing to context was performed 48 h after training. The second training was applied 72 h after the first context test. During the second training the negative reinforcement was increased (current 0.48 mA instead of 0.28 mA), but all other training characteristics remained the same. Above-mentioned second training was necessary, because our animals, raised in the frame of our experiment, have demonstrated significantly less freezing, as a result of our standard training procedure, than animals, obtained being adults from external supplier (Immunological Center, Moscow) for our independent experiment (Vysotskii *et al*., 1999). The second testing to context was performed 48 h after the second training. Testing to the conditioned signal was performed 48 h after the second context test. Thus, each animal was trained twice and tested 3 times (2 context tests and one cue test, the last one with distinction 3 min before cue and 3 min during cue).

Training and testing to context were performed using the Biotest RK-5301 chamber; testing to the conditioned signal was performed using the round plastic chamber.

The first training consisted of three sessions. In the case of spaced training, the interval between sessions was 1 hour; in massed training the interval was 1 min. Rats were placed in the chamber for 3 min in each session. A sound of 2400 Hz (the conditioned signal), lasting 30 sec, was presented when the rats had completed 2 min in the chamber. Sound level was determined by Biotest RK-5301 at its maximum settings without attenuation. Sound level estimation is about 75 dB for training and about 70 dB for cue test. Rats were presented with an electric current (0.28 mA, 50 Hz) 2 sec before the sound ended; sound and shock were switched off together. Rats were left in the chamber for a further 30 sec after the sound and shock had ended. Between sessions each animal was transported from experimental room to animal room and back (distance about 10-15 m one way, these rooms were separated by 3 doors), in the cases of both spaced and massed training.

For testing to context, rats were placed in the chamber in which they had been trained, and were left there for 6 min.

The second training was similar to the first one, but current 0.48 mA was applied (instead of 0.28 mA).

The second context test was exactly the same as the first one.

For testing to the conditional signal, rats were placed in a new chamber (round), again for 6 min. The sound was presented for 3 min, starting at the end of the third minute. During cue test experimental room was dark and only special red lamp, which could provide relatively low light intensity, was used.

During testing the absolute durations (in sec) of freezing and grooming were measured. Behavioural recording was carried out visually, with immediate input of data into a computer via keyboard. MS-DOS-based program "timekbd2.exe" was used with computer AT 286/87. Key "`" was assigned to freezing, key

"1" – to grooming, for right hand usage. When particular key was pressed, freezing or grooming was recorded, and duration of pressed state was recorded as duration of particular behavioural episode. If both keys were pressed by chance simultaneously, freezing behaviour was recorded (freezing and grooming could not be recorded simultaneously for the same animal). Such usage of a standard PC keyboard is possible in DOS only, but not in Windows. In Windows the event duration can not be recorded through the key pressed state duration (two keys or key combinations are necessary in Windows – one for the beginning of the event, and the second one – for the end of this event). There are special keyboards and specially programmed micro-controllers, designed for this task and compatible with Windows, see, for example, www.xkeys.com .

The significance of differences between groups was assessed using Mann-Whitney U-test.

### 2a.2.2. Locomotor activity in a new cage

Exp.2 $F_1$-1 males were tested in this test at the age of 88 days, Exp.2 $F_1$-1 females – at the age of 86 days. Exp.2 $F_1$-2 animals were not tested.

In the animal room a rat was placed in a new fresh standard cage "Type 3" (polypropylene, non-transparent /"translucent or opaque"/; 425 × 266 × 155 mm) with standard bedding material, but without cover for 3 minutes. This cage Type 3 was placed on Coulbourn Instruments "Small animal movement monitor". "Movement monitor" itself was standing on a laboratory table in the animal room. This rather compact device looks like a platform with possibility of slight horizontal movements in two dimensions. These slight mechanical movements can be recorded. Cage with rat was placed directly on above-mentioned platform. The following device settings have been chosen: Session time – 180 s, Sensitivity – 100 g seconds per count, Movement separation – "g" (this is separation between "small movement" and "large movement", the whole scale is from "a" to "m", and "g" is exactly in the middle of this scale). During this test only "Small movement" and "Large movements" were recorded. These are mechanical movements of the cage on the platform, induced by animal locomotor activity inside this cage. "Large movements" occurred to be much more informative than "Small" ones. Mann-Whitney U-test was used for data analysis.

### 2a.2.3. Locomotor activity in Hole-board

Exp.2 $F_1$-1 males were tested in Hole-board test at the age of 90 days, Exp.2 $F_1$-1 females – at the age of 91 day. Exp.2 $F_1$-2 animals were not tested.

In the animal room a rat was placed in a Hole-board device. The apparatus consisted of a plywood open-field (380 × 390 mm) surrounded by vertical walls 270 mm height (one Plexiglas transparent wall and 3 plywood non-transparent). The floor (also plywood) was divided into 9 equal squares (130 × 130 mm each) by two pairs of X-crossed lines (visible for researcher). In addition, the floor had 9 holes (d = 30 mm), each of them made in the middle of each square. The floor had thickness about 3 mm. The device was equipped with the second floor placed 30 mm under the main one. Standard day-light illumination about 25-32 lx was used. The subjects were individually tested in 3 min sessions.

Above-mentioned Hole-board device as a whole unit was placed on Coulbourn Instruments "Small animal movement monitor". "Movement monitor" itself was standing on a floor in the animal room. This rather compact device looks like a platform with possibility of slight horizontal movements in two dimensions. These slight mechanical movements can be recorded. Hole-board device with rat was placed directly on above-mentioned platform. The following "Movement monitor" settings have been chosen: Session time – 180 s, Sensitivity – 100 g seconds per count, Movement separation – "g" (this is separation between "small movement" and "large movement", the whole scale is from "a" to "m", and "g" is exactly in the middle of this scale). During this test "Small movement" and "Large movements" were recorded. These are mechanical movements of the Hole-board device on the platform, induced by animal locomotor activity inside Hole-board device.

Number of holes visits (investigation incidents) and number of rearing events were recorded by an observer by means of direct visual observation. Mann-Whitney U-test was used for data analysis.

### 2a.3. Drug-related tests

#### 2a.3.1. Tail withdrawal

Exp.2 P males were tested in tail-withdrawal test at the age of 107 days (immediately after the end of breeding period, 1998-01-26) for measurement of basal pain sensitivity only. Progeny Exp.2 $F_1$-1 and Exp.2 $F_1$-2 was tested in accordance with our standard protocol with 10 mg/kg morphine i.p. injection; Exp.2 $F_1$-1 males were tested at the age of 105 days, Exp.2 $F_1$-1 females – 115 days, Exp.2 $F_1$-2 males – 65 days, Exp.2 $F_1$-2 females – 70 days of age. All available Exp.2 P, Exp.2 $F_1$-1 and Exp.2 $F_1$-2 animals were tested in this test. The distal part of the tail (1/4-1/3 of its length) of a lightly restrained animal was dipped into circulating water thermostatically controlled at 56 ± 0.2°C. Latency to respond to the heat stimulus, by a vigorous flexion of the tail, was measured to the nearest 0.1 sec, cutoff latency – 15 sec. The test was done once before i.p. 10 mg/kg morphine injection (baseline latency) and 15, 30, 45, 60 and 90 min after. During this procedure an animal was packed into a peace of cotton towel (its head and tail remained visible) and was held by free hand. The latency of withdrawal was taken as an indicator of pain sensitivity.

Note that in comparison with Exp.1, in the Exp.2 concerning the age of testing the mirror design has been chosen: Exp.1 $F_1$-1 was tested being young (P65-P70) and Exp.1 $F_1$-2 was tested being relatively old (P140), whereas Exp.2 $F_1$-1 was tested being relatively old (P105-P115) and Exp.2 $F_1$-2 was tested being young (P65-P70).

#### 2a.3.2. Hot plate

The animals were tested in hot plate test practically simultaneously with tail-withdrawal procedure, using the same single morphine injection and time intervals (at each time point an animal was tested first in tail-withdrawal test and then (immediately after) – in hot-plate test). Hot-plate device consisted of metal 300 × 300 mm square plate, covered by chrome and surrounded by vertical transparent Plexiglas walls 246 mm height. Immediately after each tail withdrawal measurement an animal was placed on the hot-plate (58 ± 0.4°C). The latency of paw licking or jumping was taken as an indicator of pain sensitivity. Cutoff latency was chosen 30 s for all generations, including Exp.2 P.

Note that in comparison with Exp.1 in the Exp.2 hot-plate temperature was increased from 52ºC up to 58ºC. It was done to make our data more compatible with previous publications, to decrease learning-related component of animal behaviour in this test, and to increase usually expected pain-sensitivity-related component of animal behaviour.

#### 2a.3.3. Dependence (5.5-day morphine treatment, naloxone-induced weight loss)

Physical dependence was investigated in Exp.2 P males at the age of 107 days, immediately after the end of morphine treatment and breeding period.

At P108 animals received 60 mg/kg of morphine in the morning and 6 hours later they were injected i.p. with naloxone 2 mg/kg (concentration 2 mg/ml). Weight of each animal was measured to within 1 g using Sartorius L610D at the following time points: i) before morphine treatment, ii) in the middle of the 3-d day, iii) before naloxone injection and iiii) 6, 12, 18, 24, 30, 36, 42 hours after naloxone injection. Naloxone-induced weight loss was taken as an indicator of physical morphine dependence.

Physical dependence was investigated in Exp.2 $F_1$-1 males at the age of 148 days, in Exp.2 $F_1$-2 males – at the age of 68 days.

Animals were injected i.p. with morphine twice daily during 5.5 days by the following way (morning-evening, 12 hr interval, mg/kg): 10-10, 20-20, 30-30, 40-40, 50-50, next day – 60 mg/kg of morphine in the morning and 6 hours later they were injected i.p. with naloxone 2 mg/kg (concentration 2 mg/ml). Weight of each animal was measured to within 1 g using Sartorius L610D at the following time points: i) before morphine treatment, ii) in the middle of the 3-d day, iii) before naloxone injection and iiii) 6, 12, 18, 24, 30, 36, 42 hours after naloxone injection. Naloxone-induced weight loss was taken as an indicator of physical morphine dependence.

**2a.3.4. The influence of morphine injection on locomotor activity**

Locomotor activity was recorded in accordance with method developed by Thiel, Barnes & Mrosovsky in 1972 (Thiel *et al*., 1972), which was modified by us with respect to data acquisition hardware and software. In comparison with the first experiment, further modifications of hardware and software were made.

This test was applied to Exp.2 $F_1$-1 males at the age of 155 days, to Exp.2 $F_1$-2 males – at the age of 75 days. Exp.2 females were not tested in this test.

After 2 mg/kg naloxone injection, 45 hours later, each animal was placed into individual cage "Type 3" (polypropylene, non-transparent; 425 × 266 × 155 mm). Small mechanical movements of the whole cage were measured with a help of piezoelectric sensor. The system consisted of 54 individual cages and all these cages were placed into special rack which allowed to measure small mechanical movements of each particular home cage. During the first 3 hours baseline locomotor activity was measured. Afterwards each animal was injected with morphine 60 mg/kg or saline, i.p., and its locomotor activity was recorded during additional 9 hours (it means that this 60 mg/kg morphine injection was done exactly 48 hours after 2 mg/kg naloxone administration; total duration of the record – 12 hours). In the graphs (Figs. S79-S80) we can see 1-min averaged values (morphine offspring – 9 rats, other groups – 6 rats per group). Individual locomotor activity during the first 3 hours (averaged value) was taken as a unit of measurement for each particular animal. For statistical analysis locomotor activity of each animal during 9 hours after injection (measured with respect to its activity before injection) was averaged and non-parametric Mann-Whitney U-test was applied.

Concerning technical details, in this system each cage was placed on three foam-rubber props. Each cage was equipped with 3 short (about 5 mm) spikes for improved mechanical contact with foam-rubber props to avoid horizontal slippage. These props were attached to the 180 × 335 mm base. Each prop had sickness 20 mm without load (*i.e*. without cage installed). The first prop had size 70 × 88 mm and it was placed under the right front corner of the cage (the right front corner of this prop was placed exactly over the right front corner of the base). The second prop 44 × 70 mm – under the left front corner (prop's left front corner – over the left front corner of the base). The third prop 70 × 88 mm – in the middle between two back corners of the cage (prop's back side - over the back side of the base). The longest side of the cage was parallel to the longest side of the base and to the longest side of each prop. Water in the glass bottle (250 ml) with glass nipple was placed in the right front corner of the cage, on the grid lid, as usual. Food was provided as usual. Mechanical movements of the cage were measured at the top of the lid in the left-right direction. For this purpose metal plate 50 × 50 mm was attached to the top of the lid over the center of the cage. This metal plate was covered by glued tissue and needle of piezoelectric sensor had a contact with this surface. Piezoelectric sensor was attached to the rack by means of Y-shaped wire rocker arm which provided possibility of vertical movements without possibility of horizontal movements. All 3 ends of Y-shaped arm were placed in horizontal plane in the form of regular triangle with side about 130 mm. Two ends of this Y-shaped arm had point contacts with rack frame and the third one was holding piezoelectric sensor with its needle. The pressure of the needle on the above-mentioned tissue-covered plate was determined by the weight of piezoelectric sensor and the weight of the wire rocker arm. This construction was developed to provide good mechanical contact between cage and piezoelectric sensor and to avoid sensor damage in the

case of high amplitude movements of the cage. The slippage between tissue-covered plate and needle was absent during small amplitude movements, but it was observed during high amplitude movements.

System had 54 independent cages "Type 3" and was served by DOS-running computer AMD 386 DX/ 387, 40 MHz, 8 MB RAM, IDE HDD. This computer was equipped with analog-digital converter board LA-70M3 (*ЛА-70М3*, 16-channel board for ISA-8, 13 kHz maximum sampling frequency [in total, not "per channel"], ZAO "Rudnev-Sheliaev", www.rudshel.ru). Very similar board, the next revision – LA-70M4 (*ЛА-70М4*), can be found here: www.rudshel.ru/show.php?dev=48 . This board was connected to multiplexer board LA-MUL-64 (*ЛА-МУЛ-64*, manufactured by ZAO "Rudnev-Sheliaev", www.rudshel.ru), and, thus, the number of channels was increased from 16 to 64. Multiplexer board LA-MUL-64 was connected to custom constructed board (consisted of two mechanically separate parallel parts due to mechanical reasons – each part had separate 32-channel connector for attachment to LA-MUL-64). This custom combination of two boards (both boards were in plane) was equipped with 16 (sixteen) 4-channel operant amplifiers OP482GP Analog Devices ($k$ = 100) to avoid signal penetration from one channel to another. 54 inputs of above-mentioned 16 operant amplifiers were connected to 54 piezoelectric sensors GZP-311 (*ГЗП-311*) by thin coaxial cables. Piezoelectric sensor GZP-311 was manufactured in Russia and can be seen here: http://oldradio.onego.ru/remast2.htm). The rest 10 channels were used the following way. Five were used for animal room monitoring: one – for humidity measurement, another one – for light intensity measurement, other three – for temperature measurement in 3 independent points (on the top of the rack, near the bottom of this rack, and in the middle). The rest 5 channels remained free (unused). Data were digitized by means of LA-70M3 at the rate of 200 Hz per channel (we had 59 active channels) to avoid 50 Hz disturbance from power lines. Each 4 measurements were averaged in real time as a matter of filtering of 50 Hz disturbance, and obtained averaged values were used as primary data. Primary data absolute values were averaged in 1-min intervals in real time and stored in HDD for further analysis.

### 2a.4. Measurement of brain tissue catecholamine levels by means of high performance liquid chromatography (HPLC) with electrochemical detection (ED)

HPLC-ED was used for investigation of Exp.2 $F_1$-1 males at the age of 84 days (6 males from each group).

Measurement of free catecholamines and their metabolites content in midbrain and hypothalamus of experimental animals was carried out by means of high performance liquid chromatography (HPLC) with electrochemical detection (ED) using chromatograph CC-4 (BAS, USA).

Catecholamines and their metabolites were separated on a reverse-phase (RF) column Nucleosil-$C_{18}$ or Absorbosphere-$C_{18}$, 5 µm, 150 mm × 4.6 mm (Alltech, USA) using guard column Absorbosphere-$C_{18}$, 5 µm, 10 mm × 4.6 mm (Alltech, USA) and they were detected electrochemically by means of amperometric detector LC-4B with electrode MF1000 or MF1046 (BAS, USA).

The rats were decapitated, their midbrain and hypothalamus were removed quickly in the cold and frozen at -70 ºC, stored at -70 ºC up to the moment of catecholamine measurement. Homogenization was done in cold (t = 0 - 4 ºC) 0.1 H $HClO_4$ with an addition of internal standard 3,4-dihydroxybenzylamine (DHBA) 100 ng/ml. Homogenization was carried out in glass homogenizer with Teflon pestle, placed in the ice bath; 5000 rpm, 30 sec.

Tissue homogenate was centrifuged at 10000g, 20 min (t = 0 - 4 ºC). The supernatant was filtered by means of centrifugation through kapron filters 0.2 µm, d = 8.5 mm. A sample of 20 µl was taken from the resulting filtrate and introduced into the chromatographic system.

a). For separation of catecholamines – noradrenaline (NA), adrenaline (A), dopamine (DA), the composition of the mobile phase was as follows: 70 mM Na-phosphate buffer ($Na_2HPO_4$), with 33 mM citric acid, 1 mM $Na_4EDTA$, 0.5 mM 1-heptane sulfonic acid sodium salt, 8% methanol (v/v), pH = 5.2 (Krstulovic, 1981).

The separation was carried out using a flow rate 0.6 mL/min. The electrochemical detector was configured with a glassy carbon working electrode (MF1000 or MF1046) held at a potential of +0.600 V (versus Ag/AgCl reference, MW2021, BAS).

b). For separation of catecholamine metabolites – 3-methoxy-4-hydroxyphenylglycol (MOPEG), normetanephrine (NMN), 3,4-dihydroxyphenylacetic acid (DOPAC), homovanillic acid (HVA), the composition of the mobile phase was as follows: 0.1 M Na-phosphate buffer (pH = 3.2), with 0.15 mM EDTA, 0.3 mM 1-heptane sulfonic acid sodium salt, 18% methanol (v/v).

The separation was carried out using a flow rate 0.8 mL/min. The electrochemical detector was configured with a glassy carbon working electrode held at a potential of +0.850 V (*vs*. Ag/AgCl).

Data analysis was carried out using integrator CJ4100 (LDC/Milton Roy), or computer Hyundai HCM1420 (AT 286/287) with chromatographic program – 712 (Gilson, France).

The system for HPLC consisted of the following components also: manometric module – 802C, pump – 302, chromatograph-computer interface system – 506B (Gilson, France).

For reference, standard solutions containing mixtures of catecholamines or their metabolites were prepared such that the concentration of each component was 50 ng/mL in 0.1 H chloric acid.

The following reagents were used in the study: noradrenaline, adrenaline, dopamine, 3,4-dihydroxybenzylamine, homovanillic acid, vanillic acid, 3,4-dihydroxyphenylacetic acid, normetanephrine, 3-methoxy-4-hydroxyphenylglycol (Sigma, USA); citric acid, 1-heptane sulfonic acid sodium salt, $Na_2HPO_4$, $NaH_2PO_4$, EDTA, $Na_2EDTA$, $Na_4EDTA$ (Serva, Federal Republic of Germany); methanol (Chemapol, Czechoslovakia); $HClO_4$ (DIA-M, USSR).

In the experiments we used distilled water passed through the system of filters Milli-Q, Reagent Grade Water System (Millipore, France). The mobile phase was filtered before usage through nylon membranes 0.2 μm, 47 mm (Alltech).

The catecholamine data were expressed as absolute values of tissue concentrations of individual compounds (ng per g of wet tissue) and compared between groups using Mann-Whitney U-test (6 rats in each group, *i.e.* n = 6).

## 3. Cage enrichment experiment

### 3.1. Animals and enrichment

During postnatal days P21-P63 female mice (C57BL/6J, DBA/2J and their $F_1$ hybrid B6D2F1) were housed in the cages "Type II L" (365 × 207 × 140 mm) /Tecniplast, Italy/ (standard housing conditions) or "Type IV" (595 × 380 × 200 mm) (enriched); 4 mice of the same strain per cage. Different toys were placed in the enriched cage and these toys were renewed twice weekly (see sections "Animals" and "Housing conditions" below).

#### 3.1.1. Animals

We used females of the two inbred strains C57BL/6J and DBA/2J and their $F_1$-hybrid B6D2F1 from Taconic M&B A/S, Ry, Denmark. Freshly weaned females (C57BL/6J, DBA/2J & B6D2F1) were ordered. Animals arrived at the age 21 day having the following body weights: C57BL/6J – 9.71 ± 1.65 g, DBA/2J – 9.33 ± 2.16 g, B6D2F1 – 9.96 ± 1.76 g (mean ± SD). These weights more or less correspond to the age P21-P22. Upon arrival (on Tuesday), animals were weighed and ear-marked and assigned in groups of 4 of

the same strain to either standard or enriched housing. Finally there were 6 groups of mice (72 animals in each, 432 in total). Experiments were organized in 3 independent laboratories in 3 independent replicates in each, as it was described previously by Wolfer *et al.* (2004).

### 3.1.2. Housing conditions

Mice were housed under standard and enriched conditions for six weeks (P21-P63) in temperature (21±1°C) and humidity (50±5%) controlled conventional colony rooms under reversed 12-12 h light-dark cycle (lights on at 19:00 h) with water and standard rodent pellets ad libitum. Standard housed mice were kept in "Eurostandard Type II L" cages (365 × 207 × 140 mm; polycarbonate, transparent; "L" means "long"; these cages are also known as "Type 2a") with sawdust as bedding. Enriched housed mice were kept in "Eurostandard Type IV" cages (595 × 380 × 200 mm; polycarbonate, transparent; known also as "Type 4") with sawdust as bedding and a "Mouse House" (Tecniplast, Indulab, Gams, Switzerland) as shelter. In addition, twice a week (Tuesdays and Fridays), one enrichment item (autoclaved) was added to the enriched cages. Enrichments added on Tuesdays (when also new cages with fresh sawdust were provided to all mice) remained in the cage for one week until the next cage change (soft enrichments). Enrichments added on Fridays remained in the cage until the end of the housing period (hard enrichments). Soft enrichments included a soft paper tissue (wk 1), a coarse paper tissue (wk 2), a handful of straw (wk 3), a handful of shredded paper in stripes (wk 4), a handful of pieces of bark (wk 5), and a handful of rodent pellets that were hidden in the sawdust (wk 6). Hard enrichments included a wooden tunnel (25 cm long, inner diameter: 4 cm) with several holes (wk 1), a trapeze (12 cm long, diameter: 1 cm) hung from the cage lid (wk 2), three wooden branches (ca. 30 cm long, wk 3), a cardboard roll (15 cm long, diameter: 4 cm, wk 4), and a cardboard house "Shepherd shack" (Shepherd Speciality Papers, Indulab, Gams, Switzerland, wk 5). Thus, enrichment was a combination of more space, additional resources, increased environmental complexity, and novelty (novel items and environmental change). On the last Friday (wk 6), mice from enriched cages were placed in standard cages (Type 2a) until testing started on the following Monday.

## 3.2. Behavioural Testing

Mice were subjected to 4 standard behavioural tests (all in the same order): day 1 – Elevated O-Maze Test; day 3 – Open-Field Test; day 4 – Novel Object Test; and days 8-12 – spatial navigation in the Morris Water Maze. All tests were run during the dark phase of the cycle (07:00-19:00 h). Test rooms were indirectly illuminated by 4 40W bulbs adjusted to yield 32 lx in the centre of the test arena. Animals were video-tracked in all tests using the Noldus EthoVision 3.00 system (Noldus Information Technology, Wageningen NL, www.noldus.com ) which recorded centre point position and subject area at 4.2 Hz. Additional behaviours could be monitored using the built-in keyboard event recorder. Combined data were transferred to public domain software Wintrack 2.4 (www.dpwolfer.ch/wintrack ; Wolfer *et al.*, 2001) for analysis.

### 3.2.1. Elevated O-Maze Test

The elevated plus maze test is the most frequently used test to study anxiety-related behaviours in pharmacology and neuroscience (Belzung & Griebel, 2001). The elevated O-Maze is a modification of the elevated plus maze that has the advantage that it lacks the ambiguous central area of the elevated plus maze (Crawley, 2000).

*Apparatus and procedure.* A 5.5 cm wide annular runway made of grey plastic with an outer diameter of 46 cm was placed 40 cm above the floor. Two opposing 90° sectors were protected by 16 cm height inner and outer walls made of grey polyvinyl-chloride (closed sectors). The remaining two 90° sectors were without walls (open sectors). Animals were released in one of the closed sectors and observed for 5 min.

*Variables.* Sector entries were defined as the animal entering the respective sector with all four paws. As an anxiety related measure we have chosen time spent on open sectors (%) [ITXE].

### 3.2.2. Open-Field Test

The Open-Field Test is clearly the most frequently used of all behavioural tests in pharmacology and neuroscience. Despite the simplicity of the apparatus, however, open field behaviour is complex. Consequently, it has been used to study a variety of behavioural traits, including general motor function, exploratory activity and anxiety-related behaviours (Crawley, 2000; Prut & Belzung, 2003).

*Apparatus and procedure.* Four quadratic arenas (50 × 50 cm, 37 cm height) made of non-reflective white plastic were concurrently used. Mice were placed in the arena for 30 min.

*Variables*. To assess changes over time related to habituation we calculated changes in the length of path travelled between the first and the last 10 min [TPMX_H].

### 3.2.3. Novel Object Test

The Novel Object Test is not a very frequently used behavioural test. However, in combination with an open field test, it serves to discriminate between approach and avoidance tendencies towards novel stimuli (*e.g.* Dulawa *et al*., 1999).

*Apparatus and procedure*. 24 h after the Open-Field test, the animals were re-exposed for 15 min to the same arena. Then, a semi-transparent 50 ml Falcon tube (height 12 cm, diameter 4 cm) was placed vertically in the centre of the arena and the behaviour of the mice monitored for another 15 min.

*Variables*. An object zone was defined such that the mouse was detected inside the zone by the video-tracking system whenever it was touching the object with at least its nose. Object exploration was estimated by calculating the difference in the amount of small movements (Mohajeri *et al*., 2004) inside the object zone between time period with object (the second 15 min) and time period without object (the first 15 min). It is so called horizontal object exploration [DNSEQUNO].

### 3.2.4. Place Navigation in the Water Maze

The water maze has become the most frequently used tool in the study of learning and memory in mice (D'Hooge & De Deyn, 2001).

*Apparatus and procedure*. A round swim tank made of poly-propylene with a diameter of 150 cm was filled with water (temperature 24-26°C, depth 15 cm) that was made opaque by adding 1 litre of milk. A quadratic goal platform (14 × 14 cm) was hidden at a constant location 0.5 cm below the water surface. Its center was always 325 mm from the side of the pool. Each experimental and control group was subdivided into four subsets, each with a different platform location (a different target quadrant). The mice performed 16 training trials (4 per day, max. duration 90 s) from varying (pseudo random) starting positions, with an inter-trial interval of 30 s which they spent on the goal platform (massed training). To minimize handling, they were transferred to the pool using a white plastic cup and allowed to climb onto a wire mesh grid for retrieval. On day 5, the mice performed a 60 s probe test without the goal platform.

*Variables*. From training trials, we calculated average escape latency [TIM01X16] as a measure of overall escape performance. In addition, we calculated average swim speed [SPD01X16] and average swim path length [PTH01X16]. From probe trials, we calculated two measures of spatial selectivity, crossings of target (trained) annulus [XAT17] (annulus was determined as a square 16 cm on side) and average of crossings over similar zones in adjacent quadrants [XAC17] (average of adjacent annuli crossings).

### 3.2.5. Sound frequency and sound duration discrimination (Go/NoGo)

Sound discrimination was investigated in 48 mice at the age of 7 months (8 mice per group, all females; the same 6 groups as in the previous tests: C57BL/6J standard & enriched, DBA/2J standard & enriched,

B6D2F1 standard & enriched). Sound frequency discrimination (and, later, sound duration discrimination) was investigated in Go/NoGo paradigm. "Mouse Shuttle Box" (Campden Instruments Ltd., UK) was used (Buselmaier *et al*., 1981). It consisted of a metallic chamber (270 × 115 × 130 mm) with two identical compartments (135 × 115 × 130 mm each; L × W × H), supplied with grid floor. Compartments were separated by the wall with 38 × 49 mm arch opening and were illuminated by 1 W bulb per compartment. Animals at the age of 7 months were trained during 7 days (40 "Go" and 40 "NoGo" trails daily) to discriminate between pairs of sound. The sound pressure level (SPL) was 75 dB in the centre of experimental compartment. In the Go/NoGo sound **frequency** discrimination task "Go" signal consisted of two sounds: 50 ms 2.5 kHz and 50 ms 10 kHz, which were separated by 200 ms of silence. "NoGo" signal consisted of two identical 50 ms 5 kHz sounds separated by 200 ms of silence. Each "Go" trial consisted of 5 "Go" signal presentations with inter-signal interval 1 s (onset-to-onset). But if the animal did not move to the opposite compartment, it received additional "Go" signal presentations (maximum 5), paired with negative reinforcement – with electric current, 200 ms, 0.20 mA (the onset of 200 ms current coincided with the onset of the second sound in the sound pair). Inter-trial time interval was varying by chance in the range 5-15 s. Each "NoGo" trial consisted of 5 "NoGo" cue presentations. If the animal was moving to the opposite compartment during these 5 sec, it received negative reinforcement – current 200 ms, 0.20 mA, once. At the moment of current application, "NoGo" sound presentation was terminated even if the animal was not exposed to the whole 5 "NoGo". The order of "Go" and "NoGo" trials was pseudo-stochastic (Lipp & Van der Loos, 1991), but fixed for all animals and all training days. After 7 days of task-free period the animals were tested in Go/NoGo sound **duration** discrimination task during 7 days. "NoGo" signal was taken from sound frequency discrimination task. "Go" signal consisted of two sounds: 50 ms 5 kHz and 150 ms 5 kHz, separated by 200 ms of silence. An animal should be able to discriminate the duration of the second sounds – 150 ms in "Go" and 50 ms in "NoGo". All files for sound discrimination were prepared using Sonic Foundry Sound Forge, Version 5.0b (Build 162) (www.sonicfoundry.com) and were recorded at 44100 Hz sample rate, 16-bit bit depth, stereo (two identical channels). "Fade In" (5 ms) and "Fade Out" (5 ms) were applied at the beginning and at the end of each sound. Discrimination D was calculated as D [%] = ("Correct Go" - "Mistaken Go")/40 × 100. Mean values of 7 training days are shown in the figures.

### 3.2.6. Visual-tactile and olfactory discrimination (Hole-board)

Visual-tactile and olfactory discrimination was investigated in 48 mice at the age of 11 months (8 mice per group, all females; these mice were used for Go/NoGo sound discrimination task 4 month earlier). Each animal was tested during 5 days, 6 min daily. During day 1 and day 2 it was tested in classic hole-board, during day 3 – in visual-tactile discrimination, during days 4 and 5 – in olfactory discrimination. The test system consisted of square 40 × 40 cm 16-hole hole-board with 32 cm walls and nosepoke detectors under the floor (4 infrared beams, 4 channels per box). The signals from infrared detectors were transmitted to IBM PC (Pentium-233MMX, i430TX, 64 MB PC66 SDRAM, Windows 98 SE). Hole-board had grey polyvinyl chloride (PVC) floor with 16 round holes D = 25 mm, spaced in configuration 4 × 4 with 100 mm step between consecutive rows or columns. The floor itself had thickness 4 mm and cylindrical space D = 72 mm, h = 21 mm under each hole. Plastic Petri dish of above-mentioned diameter was placed under each hole. Illumination was 25 lx in the center of arena. The following three indicators of behaviour were registered with a help of IBM PC during 6-min session: a) total number of nosepokes; b) total duration of all nosepoke activity (total exploration time); c) mean duration of a nosepoke (c = b/a). During day 3 (visual-tactile discrimination) each second row of holes was replaced with beech plywood floor, natural colour, thickness 4 mm, with 6-point star holes, each side of equilateral triangle 30 mm (Fig. S96d). Above-mentioned indicators of behaviour were recorded and analyzed for each row separately. Discrimination D was calculated as D [%] = ("New" - "Old")/("New" + "Old") × 100, were "New" and "Old" – total exploration time of new and old holes. During day 4 (olfactory discrimination) only PVC floor with round holes was used, but under the ½ part of the floor in the each hole a portion of Mint was added. *I.e.* "Mint Tea" sold under "Migros" brand (Switzerland) was placed into a Petri dish (contents of one tea bag was placed into a Petri dish, bag itself was discarded). In accordance with manufacturer's description each above-mentioned portion consisted of 1.4 g of dried powder of Mint (*Mentha piperita*, that is hybrid [*M. aquatica* × *M. spicata*]). During an experimental day *Mentha piperita* in Petri dishes was replaced each 4 hours to keep it fresh (each tea bag of Migros "Mint Tea" is individually factory sealed). During day 5 the protocol of olfactory discrimination was applied the second time (*i.e.* day 5 is a

replication of day 4). Olfactory discrimination D was calculated as D [%] = ("No-odor" - "Odor")/("No-odor" + "Odor") × 100, were "No-odor" and "Odor" – total exploration time of holes without and with Mint odor. Olfactory discrimination during day 4 (the first day of olfactory discrimination) is shown in the Fig. S96f.

### 3.3. EEG (auditory event-related potentials) recording

EEG recording was done in 48 mice at the age of 7 months (8 mice per group, all females). These mice were never used for any sound-discrimination in any experiment. Their EEG was recorded exactly at the same age as the age of Go/NoGo training of independent subset of mice (in sound frequency discrimination task). EEG-related procedures were described in details previously in the section "Thyroxine Experiment". However there are the following differences in data analysis and EEG recording: 1) data obtained in the cage enrichment study were detrended; 2) three preliminary paradigms (Fig. S47a-c) were omitted and each animal was exposed to 4 main MMN paradigms only (Fig. S47d-g). The last difference is important, because particular electrophysiological processes are shown to be history-dependent. Briefly, recording electrode was placed 2.7 mm posterior to bregma, 3.5 mm to the right of the midline, reference – on the same hemisphere near the right olfactory bulb. Auditory stimuli SPL 75 dB were presented in 4 independent sets. In the first set the standard stimulus was accord 4 + 8 kHz, duration 50 ms, inter-stimulus interval (onset-to-onset) 500 ms; each $10^{th}$ stimulus was substituted for one of the 3 types of deviant stimuli: frequency deviant – accord 3 + 6 kHz, duration 50 ms; duration deviant – accord 4 + 8 kHz, duration 150 ms; mixed deviant – accord 3 + 6 kHz, duration 150 ms. The record duration was optimized to have 100 presentations of each type of deviant stimulus. In the second set standard stimulus was accord 4 + 8 kHz, duration 150 ms; in the third – accord 3 + 6 kHz, duration 50 ms; in the fourth – accord 3 + 6 kHz, duration 150 ms. Deviant stimuli in each set were chosen to be compatible with particular standard stimuli. Both N1 (Figs. S100-S101) and MMN (Figs. S102-S103) were taken from mentioned above paradigm (Fig. S47d-g).

### 3.4. Survival rate recording during aging

Survival rate recording during aging was not a planned part of the project with enrichment of housing conditions. In other words, we did not expect any differences here. At the age of 11 month all mice (8 mice in a group, all females) in all groups (6 groups) were looking healthy (100% survivors), however at the age of 22 months we have found that one group (enriched DBA/2J) has lost the majority of animals and this result occurred to be statistically significant (*Chi*-square, $P < 0.045$; Fig. S99a). During particular time interval (age 7-22 months) all animals from this experiment were housed in the individual cages "Type 2" (267 × 207 × 140 mm; polycarbonate, transparent) with free access to water and food.

### 3.5. Cage enrichment experiment statistical analysis

Data were analyzed using Mann-Whitney U-test. This non-parametric test can be applied to bimodal and multimodal distributions (contrary to parametric methods, which can be used if and only if the data meet assumptions of normality and homogeneity of variances). The possibility to analyze data with bimodal and multimodal distributions allows to pool appropriate groups from different laboratories and different replicates (especially if the number of animals in each lab and replicate in each control and experimental group is the same). Using Mann-Whitney U-test all standard C57BL/6J were compared with all enriched C57BL/6J, all standard DBA/2J – with all enriched DBA/2J, all standard B6D2F1 – with all enriched B6D2F1. 2-way factorial ANOVA model with between subject factors housing condition (standard versus enriched housing) and strain (DBA/2J, C57BL/6J, B6D2F1) was applied for illustrative purposes only, being one of the most commonly used statistical methods.

## 4. *Gdi1* knockout mice experiment

### 4.1. Animals

#### 4.1.1. *Gdi1* gene known functions

*GDI1* gene (D'Adamo *et al*., 1998, 2002) encodes one of the proteins regulating the small GTPases of the Rab family, a group of small Ras-like GTPases, involved in vesicle fusion in the exocytic and endocytic pathways (Novick & Zerial, 1997). The Rab GTPases cycle between an active GTP-bound and an inactive GDP-bound state through the action of regulatory proteins. Among the regulatory proteins, GDIs are required to retriev the GDP-bound form from the membrane and to maintain a pool of soluble Rab-GDP (Wu *et al*., 1996; Schalk *et al*., 1996). Two different GDIs have been described in mammals (Erdman & Maltese, 2001; Bachner *et al*., 1995; Nishimura *et al*., 1994). They interact with >40 Rab proteins and participate in fusion of different cellular membranes.

Since αGDI encoded by GDI1 is the most abundant form in brain (Bachner *et al*., 1995), it was suggested that the main consequence of mutations in GDI1 could be a modification of the pool of Rab3a, the most abundant of the Rab proteins in brain (Fischer von Mollard *et al*., 1991; Johnston *et al*., 1991), leading eventually to alterations of synaptic vesicle exocytosis, similar to those described in the mouse knockout (KO) for *Rab3A* (Lonart *et al*., 1998; Geppert *et al*., 1994, 1997). *GDI1*, however, may also be relevant through its general role in the endo/exocytic pathways. The finding that in the mouse, Gdi1 is expressed early in development and is upregulated at early stages of brain differentiation, suggested that it may also play a role in neuron migration and/or differentiation (D'Adamo *et al*., 1998).

#### 4.1.2. Generation of mice carrying a null *Gdi1* gene

To generate *Gdi1*-null mice, D'Adamo *et al*. (2002) constructed a targeting vector in which exons 1-4 of the mouse *Gdi1* gene were substituted for the *LacZ* gene under the *Gdi1* promoter and a neomycin-resistance gene cassette transcribed from the *PGK1* promoter. The homologous recombinant clones having the highest percentage of normal caryotypes (93%) were injected into blastocysts. Chimeras were obtained from both ES cell clones and were crossed to C57BL/6J and B6D2F1 (C57BL/6J × DBA/2J) females.

The heterozygote females were crossed to WT males of the corresponding genetic background. Wild-type (WT) and KO males from the crosses ($F_2$) were analyzed, as well as WT and KO mice born from $F_2$ heterozygote females crossed to WT males of the same generation and of the same genetic background ($F_3$). Backcrosses into the C57BL/6J background were done, and the seventh generation (N7) was used in all our tests. For all animals, the genotype was determined by PCR, using the primers G1/G2 and Z1/Z2 and by Southern blot. RT-PCR of RNA extracted from KO mice showed that the *Gdi1* mRNA was not produced. Western blot analysis of total protein extracts from brain confirmed the result, showing that αGdi was absent from brains of KO mice.

The *Gdi1*-null mice were viable and fertile: the mutant allele was transmitted in the expected Mendelian segregation ratio of an X-chromosome gene. Both male and female KOs appeared normal and healthy, indicating that also in the mouse the *Gdi1* gene is not essential for life.

Additional details concerning isolation of the mouse *Gdi1* gene, gene targeting and generation of *Gdi1* KO mice can be found in the article of D'Adamo *et al*. (2002), p. 2577.

All *Gdi1*-null mice and control mice used in described below study were born about 2002-07-14. They are known as "Gdi17" (because they are 7$^{th}$ generation backcross of *Gdi1* into C57BL/6J background). Only males were tested.

### 4.2. Behavioural testing of *Gdi1* mice

#### 4.2.1. Sound frequency discrimination (Go/NoGo) in *Gdi1* mice

Sound frequency discrimination was investigated in 27 male mice at the age of 3 months 2002-10-14 in Go/NoGo paradigm (14 mutants and 13 controls, all males). "Mouse Shuttle Box" (Campden Instruments Ltd., UK) was used. It consisted of a metallic chamber (270 × 115 × 130 mm) with two identical compartments (135 × 115 × 130 mm; L × W × H), supplied with grid floor. Compartments were separated by the wall with 38 × 49 mm arch opening and were illuminated by 1 W bulb per compartment. Animals were trained during 7 days (40 "Go" and 40 "NoGo" trails daily) to discriminate between pairs of sound, SPL 75 dB. Signal "Go" consisted of two sounds: 50 ms 2.5 kHz and 50 ms 10 kHz, which were separated by 200 ms of silence. "NoGo" signal consisted of two identical 50 ms 5 kHz sounds separated by 200 ms of silence. "Fade In" (5 ms) and "Fade Out" (5 ms) were applied at the beginning and at the end of each sound. Each "Go" trial consisted of 5 "Go" signal presentations with inter-signal interval 1 s (onset-to-onset). But if the animal did not move to the opposite compartment, it received additional "Go" signal presentations (maximum 5), paired with negative reinforcement – with electric current, 200 ms, 0.20 mA (the onset of 200 ms current coincided with the onset of the second sound in the sound pair). Inter-trial time interval was varying by chance in the range 5-15 s. Each "NoGo" trial consisted of 5 "NoGo" cue presentations. If the animal was moving to the opposite compartment during these 5 sec, it received negative reinforcement – current 200 ms, 0.20 mA, once. At the moment of current application, "NoGo" sound presentation was terminated even if the animal was not exposed to the whole 5 "NoGo". The order of "Go" and "NoGo" trials was pseudo-stochastic (Lipp & Van der Loos, 1991), but fixed for all animals and all training days. Numbers of correct and wrong responses for each training day are shown in the Fig. S104e-f. Discrimination D was calculated as D [%] = ("Correct Go" - "Mistaken Go")/40 × 100. Averaged discrimination values of 7 training days are shown in the Fig. S104g.

### 4.3. EEG (auditory event-related potentials) recording in *Gdi1* mice

Different EEG recordings were done in 16 mice at the age of 7, 8, 11 and 13 months (8 mutants and 8 controls, all males; these mice were used previously [4 months before the beginning of EEG recording] in sound frequency discrimination task). EEG recording in *Gdi1*-deficient mice has started exactly at the same age (7 months) as the age of EEG recording in standard and enriched C57BL/6J, DBA/2J and B6D2F1 mice in the experiment with enrichment of housing conditions. EEG-related procedures were described in details previously in the Thyroxine Experiment section of Materials and Methods. Data obtained in the *Gdi1*-deficient mice were detrended.

Animals were operated at the age of 7 months several days before the first EEG recording. Briefly, recording electrode was placed 2.7 mm posterior to bregma, 3.5 mm to the right of the midline, reference – on the same hemisphere near the right olfactory bulb (Fig. S46a-b).

At the age about 7 months *Gdi1* animals (8 experimental and 8 control) were exposed to their first EEG recording paradigm with the following parameters.

| Frequency (kHz) | Stimulus Duration (ms) | Sound Level SPL (dB) | ISI (ms) | Record Duration (min) | # of stimuli |
|---|---|---|---|---|---|
| 10 | 100 | 75 | 200 | 5 | 1500 |
| 10 | 100 | 75 | 300 | 5 | 1000 |
| 10 | 100 | 75 | 500 | 4 | 480 |
| 10 | 100 | 75 | 900 | 4 | 240 |
| 10 | 100 | 75 | 1700 | 5 | 180 |
| 10 | 100 | 75 | 3300 | 10 | 180 |
| 10 | 100 | 75 | 6500 | 20 | 180 |

The results of this record are shown in the Figs. S45e-f, S104a-d.

Event-related potentials obtained during the last two records (with 3300 and 6500 ms ISIs) were practically identical and due to this reason in the figures ERPs with 6500 ms ISI are not shown. Inter-stimulus intervals (ISIs) are given here as onset-to-onset. See Fig. S46c-h for representation of above-mentioned paradigm in graphic form (record with 6500 ms ISI is omitted in this figure also).

Immediately after above-mentioned paradigm, at the same age of 7 months, *Gdi1* animals (8 experimental and 8 control) were exposed to the second paradigm. The second paradigm differs in stimulus frequency only (5 kHz instead of 10 kHz; see table below, the first column).

| Frequency (kHz) | Stimulus Duration (ms) | Sound Level SPL (dB) | ISI (ms) | Record Duration (min) | # of stimuli |
|---|---|---|---|---|---|
| 5 | 100 | 75 | 200 | 5 | 1500 |
| 5 | 100 | 75 | 300 | 5 | 1000 |
| 5 | 100 | 75 | 500 | 4 | 480 |
| 5 | 100 | 75 | 900 | 4 | 240 |
| 5 | 100 | 75 | 1700 | 5 | 180 |
| 5 | 100 | 75 | 3300 | 10 | 180 |
| 5 | 100 | 75 | 6500 | 20 | 180 |

The results of above-mentioned record are shown in the Fig. S105a-d.

Immediately after this paradigm at the age of 7 months *Gdi1* animals (8 experimental and 8 control) were exposed to the following paradigm with different sound level stimuli.

| Frequency (kHz) | Stimulus Duration (ms) | Sound Level SPL (dB) | ISI (ms) | Record Duration (min) | # of stimuli |
|---|---|---|---|---|---|
| 10 | 100 | 56.94 | 1000 | 5 | 300 |
| 10 | 100 | 66.48 | 1000 | 5 | 300 |
| 10 | 100 | 70.92 | 1000 | 5 | 300 |
| 10 | 100 | 73.84 | 1000 | 5 | 300 |
| 10 | 100 | 76.02 | 1000 | 5 | 300 |

The results of above-mentioned record are shown in the Figs. S106a-b, S107a.

Immediately after above-mentioned paradigm at the age of 7 months *Gdi1* animals (8 experimental and 8 control) were exposed to similar paradigm, but with different sound frequency (5 kHz instead of 10 kHz; see table below).

| Frequency (kHz) | Stimulus Duration (ms) | Sound Level SPL (dB) | ISI (ms) | Record Duration (min) | # of stimuli |
|---|---|---|---|---|---|
| 5 | 100 | 56.94 | 1000 | 5 | 300 |
| 5 | 100 | 66.48 | 1000 | 5 | 300 |
| 5 | 100 | 70.92 | 1000 | 5 | 300 |
| 5 | 100 | 73.84 | 1000 | 5 | 300 |
| 5 | 100 | 76.02 | 1000 | 5 | 300 |

The results of above-mentioned record are shown in the Figs. S106c-d, S107b.

At the age of 8 months, *i.e.* one month later, *Gdi1* animals (8 experimental and 8 control) were exposed to logically similar paradigms, but the stimulus duration was chosen 25 ms instead of 100 ms. Frequency 10 kHz for different ISIs paradigm and 5 kHz for different SPLs one. See the following two tables below.

| Frequency (kHz) | Stimulus Duration (ms) | Sound Level SPL (dB) | ISI (ms) | Record Duration (min) | # of stimuli |
|---|---|---|---|---|---|
| 10 | 25 | 75 | 200 | 5 | 1500 |
| 10 | 25 | 75 | 300 | 5 | 1000 |
| 10 | 25 | 75 | 500 | 4 | 480 |
| 10 | 25 | 75 | 900 | 4 | 240 |
| 10 | 25 | 75 | 1700 | 5 | 180 |
| 10 | 25 | 75 | 3300 | 10 | 180 |
| 10 | 25 | 75 | 6500 | 20 | 180 |

The results of above-mentioned record are shown in the Fig. S105e-h.

| Frequency (kHz) | Stimulus Duration (ms) | Sound Level SPL (dB) | ISI (ms) | Record Duration (min) | # of stimuli |
|---|---|---|---|---|---|
| 5 | 25 | 56.94 | 1000 | 5 | 300 |
| 5 | 25 | 66.48 | 1000 | 5 | 300 |
| 5 | 25 | 70.92 | 1000 | 5 | 300 |
| 5 | 25 | 73.84 | 1000 | 5 | 300 |
| 5 | 25 | 76.02 | 1000 | 5 | 300 |

The results of above-mentioned record are shown in the Figs. S106e-f, S107c.

At the age of 11 months, *i.e.* 3 months later, *Gdi1* animals (7 experimental and 4 control) were exposed to two paradigms with very long (200 ms) and more usual (100 ms) stimuli. The first record was done with 200 ms stimulus duration, the second one with 100 ms – immediately after the end of the first one.

200 ms paradigm
Frequency 10 kHz, Stimulus duration 200 ms, Sound pressure level 75 dB, Inter-stimulus interval 900 ms (onset-to-onset), Record duration 66.6 min, # of stimuli 4440.
The results of this paradigm with 200 ms stimulus duration are shown in the Fig. S108a.

100 ms paradigm
Frequency 10 kHz, Stimulus duration 100 ms, Sound pressure level 75 dB, Inter-stimulus interval 900 ms (onset-to-onset), Record duration 66.6 min, # of stimuli 4440.
The results of this paradigm with 100 ms stimulus duration are shown in the Fig. S108b.

At the age of 13 months, *i.e.* 2 months later, *Gdi1* animals (4 experimental and 2 control) were tested in mismatch negativity (MMN) paradigm shown in the Fig. S47d-g.

Auditory stimuli with sound pressure level (SPL) 75 dB were presented in 4 independent sets. In the first set the standard stimulus was accord 4 + 8 kHz, duration 50 ms, inter-stimulus interval (onset-to-onset) 500 ms; each $10^{th}$ stimulus was substituted for one of the 3 types of deviant stimuli: frequency deviant – accord 3 + 6 kHz, duration 50 ms; duration deviant – accord 4 + 8 kHz, duration 150 ms; mixed deviant – accord 3 + 6 kHz, duration 150 ms. The record duration was optimized to have 100 presentations of each type of deviant stimulus. In the second set standard stimulus was accord 4 + 8 kHz, duration 150 ms; in the third – accord 3 + 6 kHz, duration 50 ms; in the fourth – accord 3 + 6 kHz, duration 150 ms. Deviant stimuli in each set were chosen to be compatible with particular standard stimuli. Duration MMN is shown in the Fig. S45d.

**General statistical analysis**

Mann-Whitney U-test was taken as a basic statistical method for our study. The reasons for this choice are explained in this SOM (p. 45). *Chi*-square was used for mortality data analysis. ANOVA and MANOVA – in some special cases (in each such a case an appropriate remark was placed in the particular graph or in the text). Statistica 6.0, Statistica 5.5 and previous versions Statistica 5.1 and Statistica 4.5 (StatSoft Inc., Tulsa, OK, USA, www.statsoft.com) were used for all statistical calculations on different computers (including Alpha – the first relatively popular 64-bit processor, 233 MHz, "Multia" VX42B-F2 from Digital Equipment Corporation (DEC), 88 MB FPM ECC RAM [four 72-pin SIMMs, 12 + 12 + 32 + 32 MB], with Windows NT 4.0 SP6a and FX!32 installed). FX!32 probably was not used by Statistica 5.5, because Windows NT has build-in support for 16-bit mode and Statistica 5.5 can be used in both 16-bit and 32-bit modes, *i.e.* 16-bit mode was used with Alpha CPU and Windows NT. Up to 95% of all graphs and related data analysis were done in my projects using Statistica 5.5. This version works great with Windows 2000 Professional, Windows 98 SE, Windows 95 and even Windows NT 4.0 for Alpha, but it can not be used with Windows XP Professional 32-bit (with any available Service Pack and in any "compatibility mode" – it produces error message after opening of several graphs: "Application: STA_BAS, The Win 16 Subsystem has insufficient resources to continue running. Click on OK, close your applications, and restart your machine."). Statistica 5.5 can not be used also with Windows Server 2003 Standard Edition 32-bit (with any available Service Pack and in any "compatibility mode" – it produces error message immediately after attempt to start: "ntvdm.exe – System Error, NTVDM encountered a hard error."). Compatibility of Statistica 5.5 with more recent Windows versions, including 64-bit ones, was not checked by me. Statistica 5.5 is not compatible with Linux openSUSE 10.3 32-bit (it just can not be opened), and probably it is not compatible with other Linux versions also. The next version of Statistica itself, Statistica 6.0, is not compatible with Statistica 5.5: 1) all graph files *.stg are not compatible – Statistica 5.5 graphs are looking like garbage in Statistica 6.0 and most of information is missing; 2) data files *.sta are compatible in one direction only – data files from Statistica 5.5 can be used in Statistica 6.0, but after being saved in Statistica 6.0 format they can not be used back in Statistica 5.5 (there is an option "do not convert" in Statistica 6.0, but if conversion is done due to some mistake, there is no way back to Statistica 5.5 format). StatView (SAS Institute Inc., Cary, NC, USA, www.statview.com) was used for synaptophysin distribution analysis due to local traditions.

We can mention here a list of computers used for data analysis, but it would be rather useless, even if it will be limited to Intel chips and chipsets: i486DX-33 (with ISA), i486DX4-100 (with VLB), P-166 (i430HX), P-233MMX (i430TX), P2-233 (i440LX with Adaptec SCSI), P2-300 (i440BX), P3-700 (i810), P3S-1.4 (i815E), P4-2.4 (E7205), P4-2.53 (i850E), P4-2.66 (i845E), Xeon-2.8 (1) (E7505 with LSI U320 SCSI), P4-3.06 (i850E), P4-3.4E (i875P with LSI MegaRAID 320-1 SCSI and Adaptec ICH5-R SATA RAID 0).

## SOM Text

## DISCUSSION

### 1. Historical remarks

Here we would like to mention a few relatively old experimental facts concerning observation of inversed phenotype in the progeny of drug-treated animals and plants.

In plants (*Linum usitatissimum*): "Phosphorus without nitrogen produces large parent plants but induces small plants in the next generation; nitrogen without phosphorus produces small parent plants but induces large plants in the next generation" (Durrant, 1962, p. 56).

In insects (*Pieris brassicae*): LSD treatment produces decreased percentage of diapausing animals in parent generation but entails increased percentage of diapausing animals in the untreated or LSD-treated progeny, obtained from LSD-treated males and intact females (Vuillaume & Berkaloff, 1974).

In mammals (Sprague-Dawley rats): neonatal thyroxine treatment of male rats accelerates their eye opening (Bakke *et al*., 1975), but decelerates eye opening in their $F_1$ untreated offspring (Bakke *et al*., 1976). "It is of possible significance that some of the changes in the untreated progeny tend to be the opposite of those observed in the neo-T4 fathers themselves" (Bakke *et al.*, 1975, pp. 106-107) (fathers were treated as neonates by subcutaneous injection of L-thyroxine during the first 7 or 10 days of postnatal life).

Morphine treatment of adult male Sprague-Dawley rats not only produces drug tolerance in these treated animals, but entails enhanced sensitivity to morphine-induced analgesia in their $F_1$ untreated male offspring (Eriksson *et al*., 1989; Cicero *et al*., 1995). Male parents received morphine together with liquid food during 8 days with 5-day withdrawal before breeding (Eriksson *et al*., 1989) or were injected with single large dose of morphine (25 mg/kg) 24 hours before mating (Cicero *et al*., 1995). In both above-mentioned studies $F_1$ adult males have shown significantly enhanced sensitivity to the antinociceptive effect of the first morphine injection.

We would like to represent the research of transgenerational epigenetic effects together with phenotypic inversion as a new branch of science. It is new, in fact. However nothing was extracted from vacuum.

Henri Bergson (Nobel Prize Winner, 1927) has written in 1907:

"We should propose, then, to introduce a distinction between the hereditability of *deviation* and that of *character.* An individual which acquires a new character thereby *deviates* from the form it previously had, which form the germs, or oftener the half-germs, it contains would have reproduced in their development. If this modification does not involve the production of substances capable of changing the germ-plasm, or does not so affect nutrition as to deprive the germ-plasm of certain of its elements, it will have no effect on the offspring of the individual. This is probably the case as a rule. If, on the contrary, it has some effect, this is likely to be due to a chemical change which it has induced in the germ-plasm. This chemical change might, by exception, bring about the original modification again in the organism which the germ is about to develop, but there are as many and more chances that it will do something else. In this latter case, the generated organism will perhaps deviate from the normal type *as much as* the generating organism, but it will do so *differently*. It will have inherited deviation and not character. In general, therefore, the habits formed by an individual have probably no echo in its offspring; and when they have, the modification in the descendants may have no visible likeness to the original one. Such, at least, is the hypothesis which seems to us most likely" (Bergson, 1998, p. 83).

Trofim D. Lysenko has added in 1948:

"Numerous facts go to show that changes in various sections of the body of a vegetable or animal organism are not fixed by the reproductive cells with the same frequency or to the same extent" (Lysenko, 1949, p. 37).

## 2. Perspectives and further experiments

Questions about molecular mechanisms remain the most attractive, as usual. However, pharmacological treatment, as morphine or thyroxine, is not the most direct way to go into the field of molecular mechanisms.

Really new possibilities will be opened starting from using "knockout" and transgenic animals (see, for example, Allen *et al*., 1990) for investigation of transgenerational epigenetic compensation.

The insertion of transgene sometimes can produce very strong developmental abnormalities and induce complex adaptive response. The duration of this adaptive response in terms of further generations was never checked carefully. There are several published reports (Crabbe *et al*., 1999; Phillips *et al*., 1999) and a lot of unofficial information about situations when in transgenic animal previously detected phenotype disappears in a few subsequent generations, in spite of undisrupted transgene. Epigenetic mechanisms, involved into transgenerational compensatory processes in transgenic and "knockout" mice, still have to be discovered.

To investigate above-mentioned question systematically it is necessary to compare the phenotypes of the descendants, which have officially the same genotype (for example, -/-), but were obtained from homozygous (-/-) or heterozygous (+/-) *male* parents (female parents should have identical genotype, in our example, +/-) (Vyssotski, 2004, pp. 284-291). A lot of mutants with strong phenotype can be used for these purposes (however chosen homozygous males still should be able to breed). We will see that the phenotype of a descendant depends not only on its own genotype, but also on paternal one, which can be heterozygous or homozygous. And *it will be a new branch of experimentation in which transgenic animals will be used as an experimental model for investigation of epigenetic compensation*.

We will take here as an example a simple constitutive knockout.

It is necessary to apply 2 types of breeding:

1) (+/-) females × (+/-) males (classic breeding paradigm, officially recommended for knockout animals);
2) (+/-) females × (-/-) males (also well-known paradigm, but it is not in use very often due to the absence of (+/+) animals as littermates).

Finally it will be possible to compare the following 5 groups:

1) (-/-) mice, obtained from (+/-) females and (+/-) males;
2) (+/-) mice, obtained from (+/-) females and (+/-) males;
3) (+/+) mice, obtained from (+/-) females and (+/-) males;
4) (-/-) mice, obtained from (+/-) females and (-/-) males;
5) (+/-) mice, obtained from (+/-) females and (-/-) males.

Usually, it is commonly accepted to discuss the differences between groups 1 (-/-) and 3 (+/+), whereas for investigation of transgenerational epigenetic compensation it is necessary to focus our attention upon the differences between groups 1 (-/-) and 4 (-/-), and, in addition, between groups 2 (+/-) and 5 (+/-).

Nevertheless further experiments are not limited to transgenic animals. It is possible to apply classical methods also.

For example, we can look at the brain lesion technique. It is possible to make cytotoxic or electrolytic hippocampal lesions on the left or right side of male mouse or rat hippocampus and investigate possible asymmetry in behaviour and brain morphology in its progeny (Vyssotski, 2004, pp. 291-293). For behavioural phenotyping it is reasonable to choose Open-field and Morris water maze, in which one can see intrinsic animal's preference to clockwise or counterclockwise movements. For brain morphology investigation it is very convenient to take morphometry of hippocampal intra- and infrapyramidal mossy fibers (the same measurements, that were used in our study; they can reveal morphological asymmetry).

Using 5 groups of males (left side lesion, right side lesion, sham lesion, both sides lesions and naïve) it is possible to reveal in the progeny the proportion of asymmetric deviations with respect to general (symmetric) deviations.

With paternal brain lesion (as with paternal drug treatment) all time intervals should be important (time interval between lesion and breeding, animal age at the moment of lesion, *etc*.). The consequences of neonatal and adult paternal brain lesions should be investigated separately. Possible differences between different lesion methods (cytotoxic, electrolytic and cold-induced lesions /*i.e.* freezing lesions/) are the questions for further discussion, because in general the most precise cytotoxic lesions can be associated with possible direct influence of cytotoxic agent on germ cells.

Paternal drug treatment, brain lesion and the appearance of transgene in the homozygous state are quite strong artificial influences, and some of them are shown to produce a set of deviations in the further generations. But concerns have been raised that this phenomenon might take place only under artificial influences, but not under natural ones.

There is at least one possibility to resolve this question. Laboratory mice, including inbred ones, were kept under standard laboratory housing conditions, which were really highly restrictive in comparison with natural environment, during more than 100 years (DBA mice are known since 1904). It can be assumed that a lot of natural features of these animals were masked by epigenetic mechanisms to make them compatible with laboratory conditions. And its phenotype looks rather stable.

But may be their phenotype and the phenotype of their untreated offspring can be converted towards wild type again if we will place young male inbred laboratory mice into enriched housing conditions until breeding maturity (Vyssotski, 2004, p. 293). We can apply the same enrichment of housing conditions that was used by us (described in this SOM) and by Wolfer *et al*. (2004). What consequences will it bring for the progeny? It is known that in these males themselves the phenotype will be shifted towards the phenotype of wild animals (enhanced dominance behavior and aggression). Using the term "wild animals" here we are keeping in mind not wild-type laboratory mice, but real wild-caught mice, which can be trapped in the mixed deciduous and evergreen forest (see, for example, Galsworthy *et al*., 2005).

The same prediction can be checked with respect to the offspring of "enriched" male mice (inbred laboratory mice, which were housed under enriched conditions during their childhood). Descendants, being housed under standard laboratory conditions, should share some characters with wild animals. We can suppose that "wild-mouse" characteristics can be unmasked in the progeny of inbred laboratory mice if their male parents were exposed to enriched housing conditions during their own maturation (for example, during postnatal days P21-P62). This hypothesis still has to be verified.

All above-mentioned experiments are supposed to be applied to inbred laboratory animals. Independent branch of experiments should be formed with hybrid animals. $F_1$ hybrid animals can have a set of transgenerational epigenetic mechanisms, activated by genetic novelty, which can not be expected in the inbred laboratory animals. We suppose that during maturation of a hybrid animal a lot of epigenetic modifications can occur, and some of them can be heritable. They can be revealed by means of comparison of descendants, obtained from the same father at different ages. To have above-mentioned comparison, the same hybrid male should be bred at different ages with standard-aged females (for example, B6D2F1 hybrid male mouse at the ages 2, 3 and 6 months can be bred with 3-month old DBA/2J inbred strain females). Broods from 2, 3 and 6-month old hybrid fathers should be compared with their synchronous controls (broods from 2, 3 and 6-month old inbred fathers). Of course, the progeny from hybrid males will be heterogeneous and sufficient numbers of descendants are necessary for solid conclusion.

Additional and the most important comparison can be organized if we will have one group of hybrid fathers housed under standard laboratory conditions and another one housed under enriched conditions. If these two groups of males will be bred with females at the ages of 2, 3 and 6 months, it will be possible to reveal the proportion of heritable epigenetic changes, induced by genetic novelty *per se* (in the progeny of standard hybrid fathers) and proportion of heritable epigenetic changes, induced by both genetic novelty and enrichment of housing conditions (in the progeny of enriched hybrid fathers). Of course, simultaneously inbred strain males taken from standard and enriched housing conditions should be bred with standard females as a control.

And it will be the first experiment in which the role of internal and external changes/influences in the real evolutionary process will be compared synchronously and under controlled experimental conditions.

# NOTES

In this section we would like to provide information about progeny development (birthweight, postnatal body weight (P00-P30), litter size, neonatal mortality) and some information about temporal distribution of the processes (learning curves, analgesic effect curves, naloxone-induced weight loss curves), which were shown in the article as mean or selected single-point values.

In addition, some methodical details, which can be classified as unusual, are explained here (in reality they were taken from our pilot experiments).

## GENERAL OBSERVATIONS

### 1. Comparison of progeny and parents – comparison through their synchronous controls

Due to well-known reasons progeny and parents have to be tested in different calendar periods (seasons or years). There is no question if the control group results of consecutive generations occurred to be the same or about the same. But as a rule the difference between parental control group and control group of the next generation is quite visible. Sometimes the difference between the same age control groups in the different generations (tested in different seasons) is about the range of the difference between experimental and control groups (tested synchronously). On the other hand, the difference during synchronous testing of the animals with slightly different age is often relatively small. Usually it is difficult or just impossible to reveal the primary reason of discussed above seasonal difference in control groups. That is why it is extremely important to have representative control groups in each particular generation. All comparisons should be done only with respect to control group tested synchronously. This order was kept throughout our study. All comparisons of drug-treated males with their untreated descendants were done here through comparisons with their synchronous controls. There is also additional limitation concerning impossibility of synchronous testing of male and female progeny. In the majority of behavioural tests (*e.g.* water-maze, two-way avoidance) the maximum number of animals in one batch is limited to 48 subjects. Only this number can pass through 5-day testing procedure synchronously. That is why male and female descendants have to be tested in different weeks (usually males go first). Due to this reason synchronous comparison is possible only between two groups (for example, between experimental males and control males). In addition, some behavioural tests produce normal distribution, whereas others do not. Thus, we have chosen Mann-Whitney U-test (a simple non-parametric test) as a basic universal statistical method for our study. For visual data analysis, in addition to usual graphs, it is useful to represent data in the form of "% change over control (Δ, %)". It can be calculated in accordance with the following formula: (("experimental group mean" - "control group mean") / "control group mean") × 100. Experimental group dispersion (SE or SD) can be shown in these graphs also. If the obtained dispersion in the experimental and control group is about the same, such figures provide quite representative information.

### 2. The importance of animal age for drug treatment, breeding and testing

The age of progeny testing is very important parameter. The testing of $F_1$ offspring at the age of 70 days gives usually more impressive consequences of paternal drug exposure than the testing of $F_1$ independent group at the age of 150 days (compare our morphine Exp.1 and Exp.2). May be some part of abnormalities can be attenuated or compensated during progeny ontogenetic development.

The time interval between the end of drug treatment and breeding period is also important. Breeding during the last 5 days of morphine treatment has produced the results shown in our article, whereas breeding after

45 or 95 days of morphine withdrawal has produced just some slight differences, mainly without statistical significance (see our morphine Exp.1 and Exp.2). May be it happens due to some reversion of epigenetic heredity to wild type during withdrawal period. It is also possible, that some additional shift in dominance (presumably epigenetic, as in genetic imprinting) can mask the pre-existing heritable epigenetic change ($F_1$ descendants are obviously heterozygous due to wild-type mothers; epigenetically modified alleles come from the fathers). Whatever it is, it does not reverse the situation to the wild type completely, because significant changes can be found afterwards in the second generation ($F_2$) – in the animals, obtained from this first generation ($F_1$) (see our article). After neonatal thyroxine treatment, thyroxine-treated males produce progeny with more pronounced changes being bred at the age of 2 months than being bred at the age of 3 months (in both cases males were bred with 3-month-old females; Figs. S6, S12-S13).

The age of male parents at the beginning of drug treatment is also important. At the age of 60 days single morphine injection 25 mg/kg done 24 hours before mating in the independent experiment (Cicero *et al.*, 1995) has produced more severe abnormalities in the progeny than similar treatment at the age of 100 days in our second experiment (see our morphine Exp.2). Younger males are better for treatment. Despite this idea sounds reasonable, it will be better to check it by means of synchronous breeding in one laboratory. In general, the influence on younger males produces better (more impressive) results in their $F_1$ progeny.

### 3. Gender-linked differences in the expression of behavioural abnormalities in the progeny

A lot of gender-linked morphological deviations and gender-linked drug-related abnormalities were found in our research in the $F_1$ progeny (thyroxine study, Fig. S4; morphine study, Fig. S51). In addition to these results, in the morphine experiment $F_1$ males, but not females, have shown impaired performance in the passive avoidance task (decreased step-down latencies; $P < 0.0026$ – males, $P < 0.13$ – females; Fig. S66c). On the other hand, in the same $F_1$ progeny only females, but not males, have shown some deviations in the hole-board test (supernormal number of hole visits during the second day; $P < 0.66$ – males, $P < 0.0052$ – females; Fig. S66b). In the frame of thyroxine study very impressive gender-linked behavioural abnormalities were observed only in the second generation ($F_2$ mice, impaired active avoidance performance; $P < 0.39$ – males, $P < 0.00023$ – females; Fig. S4). However equal behavioural abnormalities were found in males and females in the previous generation ($F_1$ mice, impaired active avoidance performance; $P < 0.015$ – males, $P < 0.010$ – females; Fig. S4). Once again we can see that the results of transgenerational epigenetic inheritance are extremely gender-dependent. Male and female progeny have different changes. And it looks reasonable if we will take into account recently reported observation of tissue-specific expression and regulation of sexually dimorphic genes in mice (Yang *et al.*, 2006) or previously published data about gender-related differences in the antinociceptive properties of morphine (Cicero *et al.*, 1996) and gender-linked differences in the expression of physical dependence in the rat (Cicero *et al.*, 2002). See (Vige *et al.*, 2008) for general review.

### 4. Paternal drug treatment: increased *vs.* normal mortality in the progeny

Strong deviations in the phenotype of the progeny of drug treated fathers are not necessarily associated with increased neonatal (*i.e.* P0-P21) mortality. In the thyroxine experiment strongly decreased birthweight and deviations in the adult behaviour and brain morphology took place together with normal neonatal mortality in all tested generations ($F_1$-$F_3$). For example, in the $F_1$ generation mortality was 6.8% in the experimental group and 8.5% in the control one, $P < 0.59$ – there is no reason to expect any influence of paternal thyroxine treatment on neonatal mortality in the offspring. The same result, obtained under similar conditions, was reported earlier in rats (Bakke *et al.*, 1975, 1976).

In the morphine experiment we had highly increased neonatal mortality in the $F_1$ progeny (Fig. S67a-b). It is interesting to mention that this increased mortality was observed simultaneously with significantly increased (elevated) birthweight (Fig. S66a,g). Increased neonatal mortality took place in the $F_1$ progeny, obtained from morphine-treated males bred with drug-naïve females during the last 5 days of morphine treatment (mortality 42% *vs.* control 15%, $P < 0.0001$). The same result was observed in the $F_1$ progeny, obtained from the same male rats bred with the same females after 95 days of morphine withdrawal

(mortality 50% *vs.* control 29%, $P < 0.0046$). The deaths occurred in the time window P6-P14 with maximum at P8-P10 in both mentioned above situations (Fig. S67a-b). This temporal death distribution precisely coincides with previously published result, obtained in the $F_1$ progeny of methadone or morphine treated male rats (Smith & Joffe, 1975; Soyka & Joffe, 1980).

In our morphine experiment in the $F_1$ experimental progeny the litter size was normal at birth, but occurred to be significantly decreased at P21 due to above-mentioned neonatal mortality (Fig. S67c). Conclusion: neonatal mortality in the offspring, obtained from drug-treated fathers, can be normal or highly increased, and this trait is absolutely independent from birthweight, which can be decreased simultaneously with normal mortality (as we can see in the thyroxine experiment) or can be even elevated together with increased neonatal mortality (as in the morphine experiment).

### 5. Paternal drug treatment: increased *vs.* decreased birthweight in the progeny

It is generally accepted that decreased birthweight is a common consequence of paternal drug treatment (Friedler, 1974, 1996). But in our morphine experiment the birthweight of $F_1$ progeny, obtain from morphine-treated fathers, was significantly increased ($P < 0.015$; group size: 146 – exp., 67 – contr.; Fig. S66a). It was observed in spite of absolutely normal litter size at birth ($P < 0.73$, Fig. S67c). And this increased in the $F_1$ birthweight was replicated in the independent experiment ($P < 0.0081$; group size: 49 – exp., 79 – contr.; Fig. S66g). It has happened together with replication of the increased in the $F_1$-males naloxone-precipitated morphine withdrawal ($P < 0.025$; group size: 14 – exp., 24 – contr.; Fig. S66f,h). Interestingly, that the bias in the direction of increased birthweight has been reported in the $F_1$ rat pups, obtained from the 60-day-old fathers after single 25 mg/kg morphine injection (Cicero *et al*., 1995), but it was not declared as significant. We had similar experimental group of $F_1$, obtained from the 100-day-old fathers after single 25 mg/kg morphine injection (made 24 hours before breeding period), and their birthweight was also increased ($P < 0.033$; group size: 39 – exp., 66 – contr.; Fig. S70b). Conclusion: the birthweight of the offspring, obtained from drug-treated fathers, being in general unpredictable (unknown in advance), may be decreased (as in the thyroxine experiment) or increased (as in several morphine experiments).

### 6. $F_2$ generation: probability of Mendelian distribution

The question about Mendelian distribution in accordance with single-locus hypothesis arises automatically when we are going to discuss $F_2$ data. If we suppose that this model can be applied to the current morphine and thyroxine data, we have to expect that all $F_1$ animals are heterozygous due to drug-naïve mothers and drug-treated fathers. In the morphine experiment we have $F_2$-incross generation and due to this reason we should have about ¼ homozygous, ½ heterozygous and ¼ wild-type animals in the $F_2$ generation. In the thyroxine experiment we have both $F_2$-incross and $F_2$-outcross. In the $F_2$-incross we should have the same situation as in the morphine experiment: ¼ homozygous, ½ heterozygous and ¼ wild-type. $F_2$-outcross was obtained from wild-type females and heterozygous $F_1$ males and we have to expect in their $F_2$ offspring ½ heterozygous and ½ wild-type animals.

In theory we should be able to observe two features in $F_2$ generation: a) bimodal or multimodal distribution; b) increased variability. Real experimental data do not provide any support for bimodal or multimodal distribution. We can see in the Figs. S20, S21a-h, S60-S65 that distribution in all experimental and control groups is unimodal. It is practically normal distribution in the thyroxine experiment. In the morphine experiment original analgesic effect data are not distributed normally, but they can be normalized in the control and experimental groups simultaneously by the logarithmic transformation: $y = \ln x$. On the other hand, we can see remarkable increase in variability in all experimental groups, obtained from $F_2$ incross breeding. It is quite impressive in the morphine experiment (Figs. S61c, S62c), especially in males (Fig. S61c), and rather moderate in the thyroxine one (Fig. S21a). For investigation of this distribution we have taken the traits with significant changes and relatively large numbers of animals in the particular groups (for the last reason physical opiate dependence and mossy fiber morphology data are not included into this analysis). So, we have analyzed the following traits: analgesic effect during the first and the second

morphine administration in the morphine experiment and birthweight and 2-way active avoidance performance (shuttle-box) in the thyroxine experiment. In the $F_2$-outcross in the experimental groups the variability is about the same as in control ones (Fig. S21e-h). One exception exists for shuttle-box results in females, in which variability has some trend to be increased in the experimental animals (Fig. S21g).

In addition we would like to mention one interesting fact in the $F_2$ progeny in the morphine experiment: in the $F_2$ we can see increased analgesic effect in the experimental animals after the first morphine administration ($P_1 < 0.022$ – males, $P_1 < 0.00012$ – females) and absolutely normal analgesic effect after the second one, made 24 hr later (Figs. S61c-d, S62c-d). So, with respect to this trait only the first morphine injection, but not the second one, can reveal difference between control and experimental groups – it can induce enhanced analgesic effect in the experimental offspring. Similar result was obtained in the previous generation ($F_1$) in males (Fig. S61a-b; $P_1 < 0.00024$, $P_2 < 0.11$). In the experimental $F_1$ females the second morphine injection has induce even slightly smaller, but not significantly smaller, analgesic effect than in control group (Fig. S62a-b; $P_2 < 0.088$). It seems that enhanced analgesic effect in the progeny is not a primary heritable change, but it is rather flexible trait, which was formed during progeny ontogenesis as a result of secondary adaptation (adaptation under rather strong pressure of some other change, that is really heritable).

In general, the distribution in the experimental groups of $F_2$ progeny in the morphine and thyroxine experiments is definitely unimodal and due to this reason it does not support simple model of Mendelian distribution with one main locus, but it can not rule it out due to increased variability in the most of experimental groups and relatively high general variability.

## 7. Several independent loci are supposed to be involved into epigenetic inheritance

In our article we can see an important statement (statement III): "Heritable epigenetic changes are distributed in a few independent loci and disappear gradually and independently of one another during 2-3 untreated generations". This statement has originated from two observations: 1) the asynchronous disappearance of modified traits in $F_2$-$F_3$ generations; 2) the absence of individual correlations of modified traits in each experimental group of $F_2$-$F_3$ generations. We believe that in our experiments above-mentioned several loci are present in the chromosomal genome. In the experiment with vinclozolin treatment (Anway *et al*., 2005) about 25 different PCR products were identified that had altered DNA methylation patterns. This estimation is compatible with our data. In our experiments we suppose now typical number of primary heritable epigenetic changes in the range of 3-12, but for different treatments this range should be wider, from 0 till unknown number. Zero means that in many experimental setups we expect no heritable epigenetic changes at all. It is possible to assume that up to 25 DNA sequences can have altered methylation as a result of paternal thyroxine or morphine treatment. However it is also possible (and we have to take it into account) that modified loci can be present in the RNA-mediated heredity, which was discovered recently in *Arabidopsis* (Lolle *et al*., 2005) and mouse (Rassoulzadegan *et al*., 2006).

We have started the first experimental project (historically it was morphine project) keeping in our mind very simple hypothesis: a) there is only one primary heritable epigenetic change in a single locus; b) all observed phenotypic changes are direct or indirect consequences of primary heritable epigenetic change. Morphine project results occurred to be rather confusing for us. We were able to explain the $F_1$-generation results using our simple hypothesis, but we were unable to do this with the second generation ($F_2$).

And thyroxine project has led us to the final conclusion. Now we are absolutely sure that our primary very simple hypothesis about a single locus is just incorrect (with respect to both morphine and thyroxine experiments). It is not compatible with observed reality. There are several independent heritable factors that are simultaneously involved into discussed phenomena. We suppose that they are distributed in several independent loci. Thus, there are several independent heritable epigenetic changes after paternal morphine treatment and there are several independent heritable epigenetic changes after paternal thyroxine treatment.

## 8. Selection of $F_1$ animals for further breeding: criteria

And here is the last methodical remark. In all transgenerational studies where $F_2$ is supposed to be tested, the question about selection of $F_1$ generation animals for breeding appears important. The number of available animals in the $F_1$ usually is about 2-4 times higher than reasonable number of breeders (due to external limitations, at least). Selection "by chance" or "without health problems" seems unacceptable. Even if one male and one female are taken from each litter for further breeding, it should be absolutely clear, how these two were selected. Any criterion will be better than "chance". Even the most arbitrary and confusing criterion being applied to both control and experimental groups will produce better outcome than "chance" selection of animals those "look good". In general, this problem is not specific for transgenerational epigenetic experiments, because in a lot of other studies with development and behaviour, morphology and electrophysiology, the subset of animals for morphology and electrophysiology is significantly smaller than the number of animals involved into behavioural and developmental testing ("screening"). And some selection is unavoidable. Selection of the "middle" in unidimensional or multidimensional space of behavioural and/or developmental parameters or selection of "extremes" in some unidimensional space (*e.g*. selection of the best and the worst performers in particular behavioural task in both control and experimental groups) are just some possible examples. Of course, for different experimental models criteria of selection can be different. There are rare examples when all available animals were used for further breeding, but in all other cases the criterion of breeder selection at the level of each generation should be clearly described. If breeder selection criteria were different for males and females (for example, all available females were used, but one half of males was selected), it should be described also. If selection was different in control and experimental groups (for example, due to different numbers of available animals) it is not good, but should be mentioned.

## 9. Reproducibility: transgenerational experiments *vs*. single-generation ones

Experiments with transgenerational epigenetic inheritance are about fourfold less reproducible than typical single-generation experiment (*e.g*. with similar pharmacological treatment). We can see this in the independent sets of experiments with prenatal vinclozolin treatment, neonatal L-thyroxine treatment and young adult morphine treatment. All 3 above-mentioned sets of experiments have about the same level of reproducibility and it is not very high. Poor reproducibility in multi-generational experiments is not a result of poor experimentation, technical problems or inappropriate statistical analysis. The observed level of reproducibility is an essential feature of particular biological system. This is a part of natural biological reality. The outcome of a transgenerational epigenetic experiment is dependent on many factors those are not so important for a single-generation experiment. As an example we can consider very sharp age-dependency or very sharp dependency on drug administration pattern in the transgenerational experiments with paternal drug treatment. Sometimes in a transgenerational experiment it is possible to select for publication and discussion some traits with relatively high reproducibility, with reproducibility which is about the same as reproducibility in a typical single-generation experiment. However above-mentioned selection can mask real situation and, in addition, in the case of 2-3 experiments the results of this selection will be definitely stochastic and not enough reliable. Artificial selection of traits with high reproducibility or long-lasting ones for publication and discussion is not a good practice in this field. Many old expectations occurred to be wrong here. And expectation of "normal" reproducibility is one of them.

"Fourfold less reproducible" estimation is a bit subjective up to now, because it was obtained by means of observation and analysis of available experiments which were not organized to solve the reproducibility question directly. To have real quantitative data concerning reproducibility it will be better to conduct 3 independent replications in 3 independent laboratories to have 9 experiments in total with specially equalized experimental setup. Similar approach was used, for example, in the experiment with enrichment of housing conditions (Wolfer *et al*., 2004). Single-generation reproducibility can be taken from comparison of drug-treated male parents with untreated controls. Transgenerational reproducibility can be taken from $F_1$ generation (further generations are not realistic for above-mentioned setup).

# PROJECT-SPECIFIC DETAILS

## 1. THYROXINE EXPERIMENTS

### 1.1. Neonatal thyroxine treatment: decreased *vs.* increased mossy fiber fields

This section provides insight into early effects of thyroxine and suggests that different heritable epigenetic changes in our experiment were formed during different periods of paternal ontogenesis and some of them were formed probably after the end of treatment period (P11), but not during thyroxine treatment itself (P0-P11). Some of the previously obtained results, discussed below, look contradictive at the phenomenological level. However they can be combined into unified picture if we will take into account two additional variables – efficacy and time period of thyroxine treatment.

In our experiment neonatal thyroxine treatment has led to decreased intra- and infrapyramidal mossy fiber fields measured in adult animals, whereas in a set of other experiments neonatal thyroxine treatment has produced increased mossy fiber fields in rats (Lipp *et al*., 1984) and mice (Lipp *et al*., 1988). In rats increased mossy fiber projections can be observed not only after neonatal thyroxine treatment, but also after exposition to ethanol in utero (West *et al*., 1981, p. 958; Lipp *et al*., 1988, p. 1917). And what is more, adult rats have increased mossy fibers after being given hippocampal lesions shortly after birth (Laurberg & Zimmer, 1980; West *et al*., 1981, p. 958). It is difficult to compare effective doses of thyroxine in mice and rats. Therefore we shall discuss here mice experiments only. In our experiment with neonatal thyroxine treatment DBA/2J mice had practically zero postnatal mortality in experimental and control groups. In comparable experiments, in which increase of mossy fiber projections was achieved, postnatal mortality in the experimental group was "fairly high" – up to 50% (Schwegler *et al*., 1991, p. 2103). The protocols of treatment with respect to doses were formally identical – 0.002 mg daily dissolved in 0.05 ml (per one pup), from day P0 until P11. But there is some slight difference in the protocols – we used 0.9% NaCl solution with pH 9.0 as a vehicle, whereas in Schwegler's *et al*. experiments Sorensen's buffer with pH 9.0 was used as a vehicle solution. Two types of appropriate control injections were applied in the last mentioned experiment in similar intervals and volumes: Sorensen's vehicle solution and physiological saline. However it can not rule out that simultaneous administration of Sorensen's buffer and thyroxine produces some synergetic effect. There is also rather speculative explanation based on supposed initial differences in the efficacy of L-thyroxine sodium salt, obtained from different suppliers: Fluka (see Materials and Methods) and Sigma. However we believe that sodium levothyroxine from Fluka should be about the same as from Sigma in accordance with available specifications (shown below).

89430 L-Thyroxine sodium salt pentahydrate
Fluka *BioChemika*, ≥97.0% (sum of enantiomers, HPLC)

| | |
|---|---|
| Synonym | 3-[4-(4-Hydroxy-3,5-diiodophenoxy)-3,5-diiodophenyl]-L-alanine sodium salt<br>Sodium levothyroxine<br>$T_4$ |
| Molecular Formula | $HOC_6H_2I_2OC_6H_2I_2CH_2CH(NH_2)COONa \cdot 5H_2O$ |
| Molecular Weight | 888.93 |
| Properties | |
| product line | *BioChemika* |
| assay | ≥97.0% (sum of enantiomers, HPLC) |
| optical activity | [α]20/D +16±1°, c = 2% in ethanol/1 M HCl (4:1) |
| total impurities | ~10% water |
| mp | 207-210 (dec.)(lit.) |
| storage temp. | 2-8°C |

T2501 L-Thyroxine sodium salt pentahydrate
Sigma ≥98% (HPLC), powder

| | |
|---|---|
| Synonym | 3-[4-(4-Hydroxy-3,5-diiodophenoxy)-3,5-diiodophenyl]-L-alanine sodium salt<br>Sodium levothyroxine<br>$T_4$ |
| Molecular Formula | $HOC_6H_2I_2OC_6H_2I_2CH_2CH(NH_2)COONa \cdot 5H_2O$ |
| Molecular Weight | 888.93 |
| Properties | |
| assay | ≥98% (HPLC) |
| form | powder |
| mp | 207-210 (dec.)(lit.) |
| storage temp. | −20°C |

From other experiments it is known that C57BL/6J mice, which can not survive during mentioned above neonatal thyroxine treatment (Lipp *et al*., 1988, p. 1915), can be injected with 0.02 mg (*i.e*. 10 times higher) daily from day P5 until P8 and remain alive (Seyfried *et al*., 1979, p. 598). It means that the most important period with respect to mortality due to thyroxine treatment took place between P0 and P4. It is known that DBA/2J mice have smaller mossy fiber projections in comparison with C57BL/6J (Lipp *et al*., 1989, p. 850). It is also known that endogenous concentration of thyroxine in DBA/2J is 2 times higher than in C57BL/6J during P9-P16 (period of maximum of endogenous concentration) (Seyfried *et al*., 1979, p. 599). In accordance with a straightforward logic one can assume that neonatal thyroxine treatment should decrease mossy fiber fields in adult animals (and this effect was really found in our experiment). However it is also known that knockout animals with the absence of thyroid hormone receptor (Thrb -/-) have 2-2.5 times elevated levels of thyroxine (T4), triiodothyronine (T3) and thyroid-stimulating hormone (TSH), measured at the age of 10 weeks (Forrest *et al*., 1996, p. 3010). It means that this functional system has effective and really active feedback.

When 0.9% NaCl solution with pH 9.0 was used as a vehicle (*e.g*. Lipp *et al*., 1988), we assume that L-thyroxine solution was prepared just before the first administration (as a rule this time interval is not mentioned in the articles).

When in our pilot study we have applied time interval about 5-15 min between L-thyroxine solution preparation (in 0.9% NaCl with pH 9.0) and the first injection to P0 pups, we had 100% mortality at day P3 (Table S1). However when this time interval was increased up to 24 hr (exactly) and solution was stored at +4°C, no neonatal mortality was observed (Table S1). This protocol was used throughout our study. It has led to significantly decreased body weight starting from day P9 (Fig. S3a) and decreased adult brain weight and body weight in both males and females (Fig. S2; see legend). Difference between control and experimental groups in body weight was visible starting from day P6 (Fig. S2) and, thus, the first effects of thyroxine administration take place during P0-P5.

Contrary to experiments with strong thyroxine treatment, in our study we observed decreased intra- and infrapyramidal mossy fiber fields in thyroxine-treated mice (Fig. S4). Thus, “strong” neonatal thyroxine treatment leads to increased IIP-MF projections, whereas “weak” one entails decreased IIP-MF projections in the same inbred DBA/2J strain.

Does “strong” or “weak” treatment mimic natural function of thyroxine in the organism? Strain-related differences in IIP-MF projections and endogenous thyroxine levels have shown that strain with the largest IIP-MF projections (C57BL/6J) has the lowest endogenous concentration of L-thyroxine (in comparison with DBA/2J). In C57BL/6J L-thyroxine peak in plasma appears later than in DBA/2J with maximum at P16 and this peak is about 2 times lower than in DBA/2J (Seyfried *et al*., 1979, 1984). *I.e*. natural endogenous concentration of L-thyroxine in C57BL/6J is significantly lower than in DBA/2J. In soft terms, strain-related differences in endogenous thyroxine level do not provide support for hypothesis “more

thyroxine – larger IIP-MF projections", because they show the opposite: "more thyroxine – smaller IIP-MF projections".

Two explanations of this situation are possible: 1) experiments with strong neonatal thyroxine treatment do not mimic natural function of thyroxine, because thyroxine at these doses has mainly disruptive (suppressive) effect on particular functional system; 2) DBA/2J mice have genetically fixed very low efficacy of thyroxine reception and high thyroxine level is some secondary effect (also genetically fixed). Like in knockout mice with the absence of thyroid hormone receptor (Forrest *et al*., 1996) those have 2-2.5 times elevated levels of T3, T4, and TSH at the age of 10 weeks. Phenotype of DBA/2J mice after strong neonatal thyroxine treatment supports the idea #1 (about disruptive effect). On the other hand, idea #2 (about relatively weak reception of T4 in DBA/2J) is supported by the following observations. It is known that C57BL/6J mice can not survive at all during mentioned above 2 μg daily neonatal thyroxine treatment during P0-P11, which produces mortality up to 50% in DBA/2J (Lipp *et al*., 1988, p. 1915). The first days of treatment (days P0-P4) are important here, because C57BL/6J can be injected with 0.02 mg daily (*i.e*. 10 times higher) from day P5 until P8 and remain alive (Seyfried *et al*., 1979, p. 598). Despite some adjustment of the actual thyroxine level towards the existing efficacy of its reception can take place in DBA/2J, it is unlikely that this adjustment can provide about 4-fold difference in the efficacy of thyroxine reception (4-fold decreased efficacy of thyroxine reception in DBA/2J mice in comparison with C57BL/6J), the difference which is necessary for compatibility with the idea "more thyroxine – larger IIP-MF projections". Contrary to this, it seems that natural relationship between thyroxine level and IIP-MF projections should be described as "more thyroxine – smaller IIP-MF projections" and the effect of strong neonatal thyroxine treatment has more similarities with the action of endocrine disruptors (Anway *et al*., 2005) than with natural thyroxine function in healthy organism.

In general it seems that in mentioned above previous experiments with DBA/2J mice (Lipp *et al*., 1988; Schwegler *et al*., 1991) effective concentration of injected thyroxine during P0-P4 occurred to be out of physiological range. We can say so looking at 50% mortality (Schwegler *et al*., 1991, p. 2103) and at the absence of dose-dependency (Lipp *et al*., 1988, p. 1911). In our experiment effective concentration occurred to be much smaller *de facto* and did not induce any mortality.

We can give the following mechanistic explanation of above-mentioned situation. Suppose that in Schwegler's experiment the first injection of thyroxine (day P0) has disrupted the mechanisms of thyroxine reception. One half of all animals died. Further additional injections of thyroxine and also endogenous thyroxine were ineffective and had zero effect. That is why it was impossible to see any further dose dependency. Zero effect of thyroxine has led to the situation when DBA/2J mice with initially small mossy fiber projections and originally high endogenous concentration of thyroxine have evolved towards C57BL/6J phenotype (animals with large mossy fiber projections and low endogenous level of thyroxine). Increased mossy fiber fields were found in the Schwegler's experiment as a consequence of dramatically decreased (disrupted) physiological effect of thyroxine (disrupted by the first injection during day P0). In our experiment the injected dose occurred to be somehow smaller and linear relationship (negative correlation) between thyroxine level and adult mossy fiber fields was found, as in the inbred mouse strains (DBA/2J and C57BL/6J).

If we suppose that in our experiment thyroxine receptive pathway was partially disrupter at the end of thyroxine treatment and this disruption occurred somehow heritable, we should also assume that it will lead to increased mossy fiber fields in the $F_1$-$F_3$ progeny (IIP-MF should be larger due to decreased reception of endogenous thyroxine level). The same conclusion appears if we have in our experiment the suppression of endogenous thyroxine production, but not the suppression of thyroxine receptive pathway. Nevertheless all these ideas (*i.e*. ideas of 5 lines above) <u>are incorrect</u>, because in our experiment *de facto* decreased mossy fiber projections were observed in the $F_1$-$F_3$ generations and also in the parental generation.

If we suppose that the heritable epigenetic changes were formed not during thyroxine treatment (P0-P11), but during later developmental stages – as a compensation of some primary thyroxine disruptive effect, we have to assume that mossy fiber fields should be diminished both in the parents and in the progeny. Exactly this result was obtained in our experiment.

That is why we believe that in mammals, if we have some treatment during an early developmental stage (which sometimes can be mainly mosaic and relatively independent from far-ranging feedbacks), the heritable epigenetic changes will be formed during later developmental stage with a help of regulatory feedbacks. This is very important, because it shows that the results of direct disruption of an early ontogenetic stage can not contain any heritable epigenetic component (at least in mammals), but heritable epigenetic components can be formed during later regulatory stage. In other words, the results of direct disruption are not heritable, but further compensation can produce some heritable epigenetic components.

An alternative explanation consists of the idea that heritable epigenetic change was formed very fast during paternal thyroxine treatment, – during days P0-P4, and it was not disruptive, but had more similarities with priming (I mean priming in birds, during their early ontogenesis), when very early exposure to weak stimulus enhances further reception of particular agent. Not tolerance, but enhanced sensitivity is developed as a result of priming. This enhanced sensitivity to thyroxine should lead to decreased IIP-MF projections in both treated parents and descendants.

However we should keep in our mind that not all changes were similar in thyroxine-treated males and their progeny. Practically all behavioural measures were opposite in P and $F_1$-$F_2$ animals. It means that despite above-mentioned ambiguity, we can be pretty sure that some heritable epigenetic changes were formed after the end of neonatal thyroxine treatment. And here there is no contradiction with previous explanation, because there are several (not one) heritable epigenetic changes and one of them can be formed at the beginning of neonatal thyroxine treatment (let say during P0-P4), whereas other one can be formed after the end of thyroxine treatment (after P11).

In our experiment effective thyroxine concentration did not induce any mortality and it is the reason to provide detailed protocol of our thyroxine treatment in Materials and Methods. This protocol has led to decreased mossy fiber projections in the treated mice. The progeny of treated male mice have shown decreased mossy fiber projections also.

### 1.2. Two-way avoidance: protocol optimization for DBA/2J strain and learning curves

In our thyroxine study for all generations (including parental one) we have chosen modified protocol of two-way avoidance with 0.8 sec delay of light or light-current termination (see Materials and Methods). Originally we were trying to apply standard 5-day protocol without any delay (Lipp *et al*., 1988). But it was found immediately that DBA/2J mice can learn this task in two days and during the last three days they perform as 100% correct responses without any individual variability (Fig. S14c-d). It is not good, because under such conditions the test has luck of sensitivity due to general pinning. This result coincides with previously published data (Buselmaier *et al*., 1981, p. 321), obtained with the same model of shuttle-boxes ("Mouse Shuttle Box", Campden Instruments Ltd., UK). We did pilot experiment with naïve female mice and different delays of light and light-current termination (0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.6, 0.8, 1.0 sec; see Fig. S22a). Finally we have chosen optimal 0.8 s delay for P, $F_1$, $F_2$ and $F_3$ generations.

Two-way avoidance learning curves are not just nice-looking pictures (Figs. S8-S11), supporting important statements based on 5-day averaged values. Learning curves reveal mechanisms of phenotype normalization in the $F_3$ generation (Fig. S11a,c) and provide information which can not be taken from 5-day averaged values. Learning curves show that normalization of phenotype in the $F_3$-incross was achieved due to improvement of performance during the first 3 training days, whereas impaired performance during the last 2 training days was uncorrected in both $F_3$-incross (Fig. S11a) and $F_3$-outcross (Fig. S11c). These effects were observed only in males (compare Fig. S11a,c with Fig. S11b,d), but they show that some part of epigenetic heredity was transmitted through females, because only difference in $F_1$ generation females (new naïve or experimental $F_1$) can be a source of observed difference in $F_3$ males. Observed difference in learning curves demonstrates that epigenetic transgenerational compensation and final phenotype normalization not necessarily go through reversion to previous naïve state, but, instead, they lead to some third state, which just reminds previous normal state in general, but not in details.

### 1.3. Mossy fiber fields and two-way avoidance: negative *vs.* positive correlation

In previous studies (Lipp *et al*., 1984, 1988) with neonatal thyroxine treatment of rats and mice it was shown negative correlation between mossy fiber fields and two-way avoidance (both individual and group correlations). In the mentioned above articles thyroxine treatment has led to increased mossy fiber fields and decreased 2-way avoidance performance. In our experiment neonatal thyroxine treatment also induced negative correlation between mossy fiber fields and two-way avoidance in parental generation (both individual correlations and group correlations), but the direction of observed changes was opposite (decreased mossy fiber fields and increased 2-way avoidance performance). The most unusual results were obtained in the progeny ($F_1$, $F_2$ and $F_3$). First of all, in the progeny any individual correlation between mentioned traits inside each particular group was absolutely absent. Second, as we can see in the Fig. S4 and Fig. S5, any offspring group with decreased mossy fiber projections has decreased two-way avoidance performance. Thus, in the progeny there is positive group correlation between mossy fiber projections and two-way avoidance performance. This observation is valid for all offspring generations ($F_1$, $F_2$ and $F_3$), without any exception. It means that causal relationships between discussed traits are more complex than it was previously assumed. And developmental dissociation of these traits is possible.

Correlation analysis of two-way avoidance performance and hippocampal mossy fiber fields (in $F_1$-$F_3$ progeny, Figs. S17-S19) shows the absence of statistically significant correlations: even when these both traits are significantly changed with respect to control, individual correlations between them can not be observed and, thus, these heritable modifications are independent. Remember (see Methods, p. 12 of this SOM) that in P and $F_1$ generations all available animals were taken for perfusion and mossy fiber morphometry, but in $F_2$ and $F_3$ generations we have selected for perfusion extreme performers in each subgroup (the best and the worst) and due to this reason we have artificially organized bimodal distribution in the Figs. S17c-d, S18c-f, S19a-d. This artificial bimodal distribution should not be a source of confusion and it does not compromise the main conclusion about the absence of individual correlations between significantly modified traits in the $F_2$-$F_3$ generations.

When all available animals were taken for correlation analysis, as it was done with birthweight and two-way avoidance performance in $F_2$ generation, the absence of individual correlations was obvious also (Fig. S21a-h). However this should not lead to wrong generalization that all correlations were not detected due to some “noise” or other technical problems, because significant correlation was detected in the same $F_2$ generation between the same two-way avoidance performance and body weight at the age of 21 day (P21), specifically in females (Fig. S20d,f). Gender-specific correlation is not unusual for particular case, because both traits were changed in $F_2$ in experimental females only, whereas experimental $F_2$ males were normal with respect to both two-way avoidance and mossy fibers (Fig. S4). Why body weight at P21 but not birthweight (P0) has produced significant individual correlation – we do not know. In all control groups and in the experimental males such correlation was absent (at both P0 and P21).

### 1.4. EEG recording in mice: the origin of mismatch negativity (MMN)

Mismatch negativity (MMN) is an auditory event-related potential (ERP) that is generated when a stimulus violates the invariance or regularity of the recent auditory past. In the simplest paradigm this is the case when an infrequent stimulus (deviant) that differs in any physical characteristic such as frequency, duration, intensity or location is presented among repeatedly presented standard stimuli. MMN is usually expressed as a negative deflection in a particular time window. We use terms “event-related potential (ERP)” and “evoked potential” as convertible terms.

In the Fig. S45a we can see that event-related potentials (ERPs) to standard stimuli are practically absent in all DBA/2J mice of particular age (18 months). It could be expected, because age-related hearing loss is a very well known phenomenon for DBA/2J strain (Ralls, 1967; Erway *et al*., 1993, 1996; Willot *et al*., 1995; Willot & Erway, 1998). On the other hand, incross and outcross $F_3$ males have shown detectable ERPs to duration deviant stimuli, *i.e.* pure deviant-related activity (Fig. S45b). Detectable stimulus-related activity was also observed in all DBA/2J mice in the paradigm with very short inter-stimulus intervals (ISIs) (Fig.

S45c). The existence of detectable short-ISI ERPs is not so surprising, because the facilitation of auditory ERP at very short inter-stimulus intervals was observed previously in humans (Budd & Michie, 1994; Loveless *et al*., 1989). However it is really surprising that particular mice have not normal auditory event-related potentials, but have pure deviant-related activity. It's a confusing result, because it is in contradiction with two known models of mismatch negativity: "lateral inhibition" (May *et al*., 1999) and "stimulus-specific adaptation" (Ulanovsky *et al*., 2003; Jaaskelainen *et al*., 2004). However we will not discuss these models in details here. Instead, we will look at the other experimental results, which will help us to resolve the question.

To resolve this question we are going to discuss the results, obtained with the same paradigms in other mice (C57BL/6J and younger DBA/2J). Naïve C57BL/6J male mice have shown classical event-related potential, induced by standard stimulus (Fig. S45d, green), slightly different duration deviant ERP (Fig. S45d, red) and expected "Difference wave" = "Deviant ERP" minus "Standard ERP" (Fig. S45d, blue). Note that observed negative deflection time window (Fig. S45d) exactly coincides with time window of statistical significance between experimental and control groups, 125-218 ms (Fig. S45b).

We have seen stimulus-induced activity in DBA/2J mice in paradigm with very short inter-stimulus intervals (Fig. S45c). This is a reason to look at the ERPs in C57BL/6J in paradigms with different inter-stimulus intervals (Fig. S45e-f). Paradigms with these ISIs are shown schematically in the Fig. S46. Long ISI paradigms (900-3300 ms) produce classical N1 (sharp negative peak), as it has been expected. Note gradual increase of N1 amplitude with increase of ISI. Note that in the short-ISI paradigms (200-300 ms) instead of sharp N1 we can see diffused delayed stimulus-induced, but not precisely time-locked activity (negative wave). When ISI becomes longer (500 ms) it is possible to distinguish both mentioned above diffused activity and appearance of N1, which can be classified as a classical N1 later (when ISI will be 900-3300 ms). Here we can see the conversion of diffused activity into time-locked N1 during the increase of inter-stimulus interval. Note that the paradigms with the shortest ISI were applied to each animal before paradigms with the longest ISI. Mentioned here diffused activity will be classified later as a physical basis of deviance-related activity, observed during mismatch-negativity paradigm. Note that this diffused activity can be observed along, *i.e.* without good (sharp) N1.

Interesting remark concerning MMN in DBA/2J mice: discussed above pure deviant-related activity in DBA/2J mice is reproducible at the group level, if we have 10-20 mice in a group; however probably due to the essential stochastic nature of this phenomenon, our methods do not give good reproducibility at the level of individual animals.

Above-mentioned MMN-like diffused activity can not be helpful for organization of operant behaviour, at least in the Go/NoGo shuttle-box-based sound duration discrimination task (discrimination 50 *vs.* 150 ms). DBA/2J of particular age can not learn this task, whereas C57BL/6J and $F_1$ hybrids B6D2F1 (obtained from C57BL/6J females and DBA/2J females) can learn this task successfully (Fig. S88c, S91c-d). Both C57BL/6J and B6D2F1 have classical ERP under particular conditions (Fig. S100a,c, S101a,e). Thus, classical ERP, but not stimulus-induced diffused activity, is necessary for operant behaviour (we did not check non-operant paradigms yet).

From the single-neuron recordings in auditory cortex in the rat it is known that diffused activity can be formed by neurons, which properties do not exactly match the properties of the stimulus (DeWeese *et al.,* 2003). These neurons usually show delayed and very approximate time-locked responses with relatively low probability. Note that both neurons with high and low probability of responses under particular conditions will give only one or zero response per neuron as a result of particular stimulus application (DeWeese *et al.,* 2003).

Above-mentioned data look reasonable if we assume the following. Under condition of repeated stimuli a neuron in the primary auditory cortex produce a response if and only if this response will entail detectable result (*i.e.* it will be effective in terms of participation in the evoked potential formation). The efficacy of response can be detected by this neuron with a help of feedback (*i.e*. neuron analyzes its own response together with associated consequences). We can not specify the nature of this feedback now, but we can obtain an estimation of temporal features of this feedback from our data (see the next paragraph below). It

the response occurred to be inefficient under particular conditions, next time under the same conditions the response will be suppressed by this neuron – the neuron's activity will be self-suppressed. It the neuron has induced distinctive event-related potential, next time neuron's activity will not be suppressed or self-suppressed. If particular conditions occurred to be locally new, a few neurons will produce a response, even if it will be ineffective and even if in the future it will be self-suppressed during the next presentation of the same stimulus. Each neuron prefers to take part in the processes in which its participation makes a difference, *i.e*. in which this neuron has some portion of control over its local environment. And *vice versa*, a neuron does not like to show any stimulus-related activity, if this activity is ineffective.

The facilitation of ERP in very short ISI paradigm is achieved due to artificial substitution of natural feedback by the next stimulus presentation. Under discussed above conditions (stimulus duration 100 ms) artificial substitution occurs to be the most effective with ISI of 300 ms. Under ISIs 200 and 500 ms the diffused event-related activity is significantly smaller (Fig. S45f). Above-mentioned observation provides insight that relevant feedback with information about the efficacy of neuron's activity is expected by the neuron about 300 ms from the onset of 100 ms sound.

Note that all MMN and deviant-related activity which were observed in our experiments in mice, were observed with duration deviants, but not with other 2 types of deviants (frequency deviants and mixed frequency-duration deviants; see, for example, Fig. S102b-c, S103c-f). Frequency deviants do not induce significant MMN-like activity in mice, the result which was reported previously (Ehlers & Somes, 2002; Siegel *et al*., 2003). Mixed deviants with simultaneous difference in frequency and duration do not induce statistically significant MMN-like activity (at least, this activity is shorter and smaller than in the case of duration deviants; our mice data; see, for example, Figs. S102c, S103e-f). The absence of frequency MMN in mice supposes independent processing of different frequencies in these animals.

Reported in our article duration deviant MMN was observed in the descendants of thyroxine-treated males, whereas in the control group MMN was significantly smaller ($P < 0.05$, time window 125-218 ms). Here we show that relatively good duration deviant MMN is not a unique feature, associated exclusively with paternal thyroxine treatment. Qualitatively and quantitatively similar MMN was observed in the normal DBA/2J mice (age 7 months, Fig. S45h), which were just younger than our experimental and control animals (age 18 months, Fig. S45g). Looking at this trait we can see that thyroxine treatment does not induce something qualitatively new, but, instead, the descendants of thyroxine-treated males demonstrate attenuated or delayed aging. Keeping in our mind all necessary precautions (*e.g*., the following generalization can be true only with respect to one particular trait), we can conclude that the progeny of thyroxine-treated males at the age of 18 months looks younger than their age-matched control.

It is interesting that there is a set of observations where neonatal thyroxine treatment accelerates development, but slows down the development of untreated descendants. Previously it was shown that neonatal thyroxine treatment of male rats accelerates their eye opening (Bakke *et al.,* 1975), but decelerates eye opening in their $F_1$ untreated offspring (Bakke *et al.,* 1976). Using auditory ERPs and duration deviant mismatch negativity paradigm in the $F_3$ descendants of thyroxine-treated males we can see the deceleration of aging process once again. And probably once again we can see the inversion of the acquired paternal phenotype in the untreated progeny.

### 1.5. EEG recording in descendants of thyroxine-treated mice: statistical analysis

Duration MMN in individual mice, $F_3$ male descendants of thyroxine-treated males, was represented in a table in the form of "duration deviant auditory event-related potential (ERP) minus standard stimulus auditory ERP" (*i.e.* mismatch negativity – MMN), microvolt (μV). The table represents EEG-related data of 32 mice and it has size 461 variables × 256 cases. It has 256 cases for 32 mice, because it contains averaged ERPs of individual animals in 4 separate records (shown in the Fig. S47d,e,f,g) and each animal has 2 channels (one from the right and one from the left auditory cortex). The table has 461 variables, because 451 variables represent epoch 900 ms with 2 ms step (100 ms before stimulus onset and 800 ms after) and additional 10 variables are used for animal grouping (they are discussed below). For each individual animal we have in the table a long set of values which are "averaged duration deviant stimulus

ERP" minus "averaged standard stimulus ERP". To obtain "averaged standard stimulus ERP" all standard stimuli from particular record (for example, from paradigm 1 [or 2, or 3, or 4]) were averaged.

Variables:

"ANIMAL" – individual number of each animal with extension indicating record number and type of applied filtering (this information normally should not be used for further statistical analysis)

"CHANNEL" – channel number; can be 1 (right channel) or 2 (left channel), signals from right and left primary auditory cortex respectively

"T_C" – "thyroxine" or "control"; can be 1 (descendants of thyroxine-treated males) or 2 (descendants of control males)

"I_O" – from "incross" or "outcross"; can be 1 (descendants of $F_2$-incross) or 2 (descendants of $F_2$-outcross)

"PARADIGM" – paradigm number; can be 1, 2, 3 or 4 corresponding to paradigms d, e, f and g shown in the Fig. S47; d & f with long duration deviant, e & g with short duration deviant

"TI_TO_C" – "thyroxine-incross" or "thyroxine-outcross" or "control"; can be 1 (descendants of thyroxine-treated males, incross), 2 (descendants of thyroxine-treated males, outcross) or 3 (descendants of control males, incross and outcross together)

"TITOCICO" – "thyroxine-incross" or "thyroxine-outcross" or "control-incross" or "control-outcross"; can be 1 (descendants of thyroxine-treated males, incross), 2 (descendants of thyroxine-treated males, outcross), 3 (descendants of control males, incross) or 4 (descendants of control males, outcross)

"DAYTEST" – day of test; can be 1, 2, 3, 4, 5, 6, 7 or 8; all 32 animals were operated during 3 consecutive days and 3 days later they were recorded during 8 consecutive days (4 mice daily; mice were recorded in the same order as they were operated); "DAYTEST" variable approximately represents time interval between operation and recording

"TIMETEST" – time of test; can be 1, 2, 3 or 4 corresponding to morning (8:00-11:00 a.m.), middle of a day (11:00 a.m.-2:00 p.m.), evening (3:00-6:00 p.m.) and late evening (6:00-9:00 p.m.); remember that in the animal room reversed daylight cycle was used: light (8:00 p.m. – 8:00 a.m.) and dark (8:00 a.m. – 8:00 p.m.), but absolute (normal human) time is shown here and in the Methods

"BP_GP" – obtained from $F_2$ "bad performers" or "good performers" in the shuttle-box test; can be 1 (obtained from bad performers) or 2 (from good performers); note that in the shuttle-box test the light was used as a conditional stimulus (not sound)

"_100" – "0" – the difference between duration deviant ERP and standard stimulus ERP, electric potentials from primary auditory cortex, measured in microvolt; 50 variables which represent time interval 100 ms before sound presentation (with 2 ms step); from "-100 ms" till "0 ms"; this time interval serves as a baseline

"0" – "`800" – the difference between duration deviant ERP and standard stimulus ERP, electric potentials from primary auditory cortex, measured in microvolt; 401 variables which represent time interval 800 ms during and after sound presentation (with 2 ms step); from "0 ms" till "800 ms"; remember that sound stimulus duration was 150 ms or 50 ms and due to this reason the duration deviant novelty can be detected starting from the variable "`50"

The table was obtained from Neuroscan 4.2 *.avg files. These *.avg files were imported into EEGLAB 5.03 in MATLAB 7.0.4 using command "loadavg". Compensating multiplier $k = 6$ was applied to all data to eliminate software error in the EEGLAB 5.03 import (the origin of this error is not clear, however our compensating multiplier was verified using Siemens Mingograf 21 calibration with 100 μV peak-to-peak standard pulse and the exact value of this multiplier is 5.92417). From MATLAB data were imported into Statistica 5.5. Finally data were analyzed in Statistica 6.0.

In the Fig. S45b a mean group value in a "paradigm" was taken as a unit of analysis (there are 4 paradigms with duration deviants, see Fig. S47d-g, thus, n = 4 for each "group" – experimental group 4 mean values were compared with control group 4 mean values). Using this method and t-test for independent samples we have compared united experimental group (incross + outcross) with united control one (incross + outcross) and we have received significance ($P < 0.05$) from 125 ms till 218 ms (Fig. S45b). However, usually an individual animal or an individual record of an individual animal is taken as a unit of analysis. In this section we use an individual record of an individual animal as a unit of analysis for processing of the same data set (discussed table).

**Non-parametric Mann-Whitney U-test**

All experimental *vs.* all control
$P < 0.05$: time window 152-208 ms
The best significance: 176 ms, $P < 0.011$ ($n_1$=152, $n_2$=104)

Experimental incross *vs.* all control
N.S. ($n_1$=80, $n_2$=104)

Experimental outcross *vs.* all control
$P < 0.05$: time window 152-212 ms
The best significance: 172 ms, $P < 0.0087$ ($n_1$=72, $n_2$=104)

Experimental incross *vs.* control incross
$P < 0.05$: time window 110-142 ms
The best significance: 130 ms, $P < 0.010$ ($n_1$=80, $n_2$=40)

Experimental outcross *vs.* control outcross
$P < 0.05$: time window 162-186 ms
The best significance: 172 ms, $P < 0.016$ ($n_1$=72, $n_2$=64)

**Parametric factorial ANOVA**
Factors:
1. Channel (1 or 2)
2. Thyroxin or Control
3. Incross or Outcross
4. Paradigm (1, 2, 3, 4)
Factor 2 (Thyroxine *vs.* Control), $P < 0.05$: time window 136-212 ms
The best significance: 164 ms, $P < 0.010961$

**Parametric main effects ANOVA**
Factors:
1. Channel (1 or 2)
2. Thyroxin or Control
3. Incross or Outcross
4. Paradigm (1, 2, 3, 4)
Factor 2 (Thyroxine *vs.* Control), $P < 0.05$: time window 140-214 ms
The best significance: 166 ms, $P < 0.008091$

We can see that above-mentioned standard statistical tests have brought approximately the same result as our specific statistical test, optimized for particular task. Specific test has revealed difference with $P < 0.05$ in the time window 125-218 ms. This time interval is slightly wider than, for example, time interval given by factorial ANOVA (136-212 ms).

However this table is discussed here not only for support of statistical result, presented in the article. It shows several factors which can be classified as “technical” or “methodical”, but they are also important for observation of good duration mismatch negativity (MMN). We do not provide detailed statistical analysis of these factors here, because it is space consuming. However for primary impression we will show here One-way ANOVA result for each factor. In addition, we will show parametric factorial ANOVA result with two factors (the first factor will be Thyroxine-Control in all cases). We will choose here time point 170 ms (one of the best).

Hereinafter:
"*P*" (without any number in round brackets and without round brackets at all) – statistical significance given by One-way ANOVA (t-test for independent samples) for particular factor (*e.g.* "Channel").
"*P*(1)" – parametric factorial ANOVA result for factor "Thyroxine-Control" ("*P*(1)" is always for factor "Thyroxine-Control"; this rule works up to the end of this section).
"*P*(2)" – parametric factorial ANOVA result for particular factor (*e.g.* "Channel").
"*P*(1,2)" – parametric factorial ANOVA result for interaction between factor "Thyroxine-Control" and particular factor (*e.g.* "Channel").
Time point 170 ms is chosen for all comparisons.

Factor Channel. Channel 1 (placed in the right hemisphere – hemisphere of reference electrode) produces slightly better duration MMN than Channel 2. $P < 0.61$, $P(1) < 0.012$, $P(2) < 0.67$, $P(1,2) < 0.68$

Factor T_C (Thyroxine-Control). $P < 0.012$ (One-way ANOVA)

Factor I_O (Incross-Outcross). $P < 0.41$, $P(1) < 0.0087$, $P(2) < 0.24$, $P(1,2) < 0.96$

Factor Paradigm. Paradigms 1 & 3 with longer duration deviant produce slightly better MMN than paradigms 2 & 4 with shorter duration deviant. $P < 0.62$, $P(1) < 0.013$, $P(2) < 0.61$, $P(1,2) < 0.94$

Factor DAYTEST (time interval between surgery and recording, here it is 6-12 days). Animals were tested one by one during 8 days and slight relative decrease of duration MMN was observed in the animals tested during the last 2 or 3 days of 8-day period. When this recording session was replicated with the same animals 42 days later (data not shown), duration MMN occurred to be decreased in comparison with the first recording session; about 50% decrease (approximately). Statistical result for the first recording session: $P < 0.031$, $P(1) < 0.0078$, $P(2) < 0.0011$, $P(1,2) < 0.0015$.

Factor TIMETEST (time of recording with respect to daylight cycle) can have some influence on duration MMN. Duration MMN is slightly better when animals are active (during beginning of dark phase and during last hours of dark phase). $P < 0.41$, $P(1) < 0.022$, $P(2) < 0.36$, $P(1,2) < 0.0014$

Factor BP_GP (from Bad Performers – from Good Performers). Here we take into account 2-way avoidance performance of their parents (male good performers were bred with female good performers, male bad performers – with female bad performers, inside each group), light as a conditional stimulus. Descendants of good shuttle-box performers have slightly better duration MMN than descendants of bad performers. $P < 0.12$, $P(1) < 0.012$, $P(2) < 0.12$, $P(1,2) < 0.99$ The role of parental 2-way avoidance performance looks not very impressive here, however if we take this factor into consideration, it improves general situation significantly (see below).

**Parametric factorial ANOVA**
Factors:
1. Thyroxin or Control
2. Incross or Outcross
3. From Bad performers or Good performers
4. Paradigm (1, 2, 3, 4)
Factor 1 (Thyroxine *vs.* Control), $P < 0.05$: time window 130-220 ms
The best significance: 174 ms, $P < 0.001989$
Interaction: $P(1,2,3) < 0.001161$, 174 ms

Here we have investigated duration MMN (EEG) in $F_3$ descendants of thyroxine-treated males. We can see here highly significant effect of ancestor thyroxine treatment ($P < 0.0020$) and highly significant interaction ($P < 0.0012$) between ancestor treatment, breeding paradigm and parental shuttle-box performance.

## 2. MORPHINE EXPERIMENTS

### 2.1. Paternal morphine treatment and early in life naloxone-precipitated morphine withdrawal

In the morphine study during paternal morphine treatment after the first 7 days of chronic morphine treatment we applied naloxone-precipitated morphine withdrawal and 24 hours later morphine treatment was prolonged to have 38 days of morphine treatment in total (see Materials and Methods). It was done to achieve 2 independent goals. First of all, this protocol of morphine treatment makes possible the direct comparison of opiate dependence development in fathers and offspring (asynchronous, but in comparable age; data not shown). Second, and the most important, the naloxone-precipitated morphine withdrawal enhances the subjective importance of morphine treatment. This event occurs in relatively young animals (in our experiment in the 49-day-old). All results of morphine experiments, described in our article, were obtained using paternal morphine treatment with early in life naloxone-precipitated withdrawal. Afterwards, an additional independent experiment [exp.2] was done (see this SOM), in which male parents were treated with morphine without any withdrawal – regularly with 12 hr intervals during 46 days (see Materials and Methods). Some parameters of $F_1$ offspring occurred to be similar in the two above-mentioned experiments (see, for example, birthweight: Fig. S66a,g, $P < 0.015$ – paternal opiate withdrawal, $P < 0.0081$ – no withdrawal). However some other very important parameters, *e.g.* opiate dependence after standard morphine treatment, occurred to be more pronounced in the experiment that was done with paternal early in life opiate withdrawal (Fig. S66f,h, $P < 0.00066$ – paternal opiate withdrawal, $P < 0.025$ – no withdrawal). Despite these experiments were not synchronous (thus, other factors are probably involved also), mentioned above observation can be taken as a primary guideline.

### 2.2. Analgesic effect: tail-withdrawal test *vs.* hot-plate test

In the morphine study we used tail-withdrawal test to demonstrate analgesic effect of standard dose of morphine (10 mg/kg), whereas in a lot of other studies we can find hot-plate test, used in the similar situations. It is well-known fact that different mechanisms are involved in tail-withdrawal and hot-plate reactions. Tail withdrawal response depends mainly on peripheral mechanisms, whereas into hot-plate response both central and peripheral processes are involved. In all our experiments we did both tail-withdrawal and hot-plate tests simultaneously (see Materials and Methods), but the difference between experimental and control groups in all our progeny ($F_1$ & $F_2$) was greater in the tail-withdrawal test. We do not know exactly the reason of this difference between two tests. May be, there is a methodical reason – for example, the parameters of hot-plate test might have been optimized not as good as in tail-withdrawal test. But may be there is a natural reason – may be the effect of paternal treatment is not so visible in progeny in hot-plate test due to some behavioural compensation, which probably can not play such a role in the tail-withdrawal test. Anyway we have chosen tail-withdrawal test, because it occurred to be much more demonstrative.

There is at least one previously published study where similar difference between tail-withdrawal and hot-plate tests was demonstrated (Kirby *et al*., 1982). This study was undertaken to compare the morphine-induced analgesic response in adult offspring of rats which had been injected during the last half of gestation on schedules known to produce fetal tolerance (5 mg/kg morphine at 6 hour intervals) versus a schedule known not to produce fetal tolerance (10 mg/kg at 12 hour intervals). At 30 days postnatally the offspring of animals injected on these 2 schedules show no changes in their responsiveness to the analgesic effect of morphine as determined in the hot-plate test. However adult offspring of mothers injected with 20 mg/kg/day of morphine in four divided doses on days 12-20 of gestation (E12-E20) have an enhanced analgesic response to morphine in the tail-flick test (here we consider tail-withdrawal and tail-flick tests as physiologically similar). In contrast, offspring of mothers injected during the same period of gestation with 20 mg/kg/day of morphine in two divided doses respond to the analgesic effect of morphine in the same manner as the offspring of saline-treated mothers. These results show that the schedule for prenatal morphine administration can play a role in the behavioural effects of morphine in adulthood.

Note also, that prenatally morphine-treated animals have shown enhanced sensitivity to morphine-induced analgesia (Kirby *et al*., 1982), similar to $F_1$ progeny of morphine-treated males in our own or previously published experiments (Eriksson *et al*., 1989; Cicero *et al*., 1995). Prenatal morphine treatment enhances sensitivity to morphine-induced analgesia in adult animals (Kirby *et al*., 1982; Castellano & Amassari-Teule, 1984; Eriksson & Ronnback, 1989; Gagin *et al*., 1996). However in some other experiments the decreased morphine-induced analgesia was reported at the age of 3, 5 and 11 weeks in prenatally (E5-E12) morphine-treated rats (O'Callaghan & Holtzman, 1976). The discrepancy could be attributed to differences in the time of drug administration, as well as the time of analgesia testing.

In the study with neonatal morphine treatment (Arjune & Bodnar, 1989), neonatal morphine injections (20 μg, days 1-7) has altered analgesia on the tail-flick and jump tests induced by morphine in adult male and female rats (2.5 and 5.0 mg/kg, s.c.). Neonatal morphine treatment significantly increased the magnitude of morphine analgesia on both tests in females, and significantly decreased the magnitude of morphine analgesia on both tests in males, thereby acting to vitiate the observed gender differences in morphine analgesia (Arjune & Bodnar, 1989).

When the mothers were treated with morphine before pregnancy (during 5 days) and then they were bred with drug-naive fathers (5-day withdrawal before breeding), their male and female offspring had standard predictable change – decreased sensitivity to morphine-induced analgesia (tolerance) (Friedler, 1974).

Thus, paternal or maternal morphine treatment before mating, morphine treatment during different periods of pregnancy and neonatal morphine treatment may have different and sometimes opposite, sometimes gender-dependent effects on morphine-induced analgesia in young adult animals.

### 2.3. Tail-withdrawal test: analgesic curves

In the Fig. S66d and Fig. S66e one can see nice analgesic curves, obtained in the tail-withdrawal test, with the following time points: before morphine injection and 15, 30, 45, 60 and 90 min after. Two narrow black bars on the left side of each picture show basal pain sensitivity – the latency of tail-withdrawal in seconds before morphine administration. In this particular case we can see $F_1$ males (the same animals as in Fig. 5, A and B), during the first morphine administration and during the second one, done 24 hours later. Basal pain sensitivity is fairly stable as well as analgesic effect in the control animals. However the analgesic effect in the experimental group during the second day is significantly smaller. Thus, this enhanced sensitivity to morphine analgesic effect can not be classified as a stable trait. Enhanced analgesic effect of the first morphine administration is very impressive in the male progeny of morphine-treated males. However analgesic effect of the second injection is dramatically smaller. We can describe this phenomenon in terms of tolerance development: the rate of tolerance development is higher in the progeny of morphine-treated males, despite their higher initial sensitivity to morphine-induced analgesia.

Extremely fast disappearance of enhanced sensitivity to morphine-induced analgesia after the first injection of morphine was observed in tail-withdrawal test in $F_1$ males as highly significant result (Fig. S54a,c), in $F_1$ females – as non-significant trend (Fig. S55a,c), in $F_2$ males and $F_2$ females – as highly significant result (Figs. S57a,c and S58a,c, respectively). Note that control animals in all above-mentioned tests have demonstrated extremely stable analgesic effect (no difference between the 1$^{st}$ and the 2$^{nd}$ morphine injection). Very high flexibility of particular trait in experimental animals shows that its primary appearance is a result of secondary adaptation.

From the methodical viewpoint it is very important that animals were not used in any pain sensitivity test or in any other test with strong negative reinforcement before above-mentioned test with morphine. As we will see below, such tests can push initially abnormal phenotype towards normalization.

In the hot-plate test we can see that there is a great difference between the 1$^{st}$ and the 2$^{nd}$ day in all groups, including control ones (Figs. S54b,d, S55b,d, S56b,d, S57b,d, S58b,d). Looking at such learning curves we can say that it is not an analgesic test *per se*, but rather a behavioural test for investigation of learning and memory (and learning and memory were not impaired here). May be one of the reasons of this result is

relatively low temperature of hot-plate (52ºC). In our independent experiment (Exp.2) we used higher temperature of hot-plate (58ºC) and behaviour was changed dramatically: instead of paw leaking, animals were jumping out of the chamber with relatively low latency (about 2.0 s for control and 3.4 s for experimental male rats).

In this Exp.2 $F_1$ rats were tested for basal pain sensitivity two times: at the age of 2.5 months and 3.5 months. We had 3 groups of animals: 1) $F_1$ progeny of chronically morphine-treated males and drug-naïve females; 2) $F_1$ progeny of males after acute (25 mg/kg) morphine injection and drug-naïve females; $F_1$ progeny of drug-naïve animals. No differences were observed in tail-withdrawal test both at the age of 2.5 months and at the age of 3.5 months (Fig. S83a). Basal tail-withdrawal latencies were about 2.5 s in males and about 2.0 s in females at both ages. But in hot-plate test increased latency was observed in both experimental groups (in males) at the age 2.5 months, but not at the age of 3.5 months (Fig. S83b). There were no differences in females. Note that normalization of basal pain sensitivity took place here between age 2.5 months and 3.5 months without any morphine treatment. Only two factors contribute here: natural aging and pain sensitivity testing procedure. We can not discriminate between these two factors now and should consider them together. And what is more, at the age of 3.5 months these animals did not show enhanced sensitivity to morphine-induced analgesia in both tail-withdrawal and hot-plate tests (Fig. S72a,c). It means that not only morphine injection can lead to normalization of analgesia-related phenotype in $F_1$ males, but natural aging together with pain sensitivity testing procedure can lead to the same result. May be aging and pain sensitivity testing do not do it so fast as morphine injection, but final result (normal phenotype) is the same.

### 2.4. Naloxone-precipitated morphine withdrawal: weight loss *vs.* behavioural signs

For investigation of physical drug dependence we have chosen the weight loss during 24 hours after naloxone injection. However in a lot of other studies for the same purpose a battery of behavioural measurements was used (for example, in the open-field, about 30 min after naloxone injection). In both setups experimental animals were treated with morphine or other opiate drug during several days before above-mentioned naloxone administration. At the beginning of our study we were trying to apply one known behavioural battery. But, to tell the truth, during this observation it was possible to see some degree of suffering in the experimental animals. This physiological state can be expressed in different individuals in variable behavioural patterns. On the other hand, the weight loss during 24 hours (but not during 30 min) after naloxone injection is expressed in very homogenous manner, in a manner which produces relatively high significance level. Time point 24 hr is the real optimum for this test, as we can see in the Fig. S66f,h. It has better discriminative power than other time points (6, 12, 18 and 30, 36, 42 hours). Our observation coincides with conclusions, published previously (Cicero *et al*., 2002, p. 693).

### 2.5. Naloxone-induced weight loss and animal age

Opiate dependence data, shown in the Fig. S66h, were taken from an independent experiment (Exp.2). It was not included into the article: there is $F_1$ progeny, obtained from morphine-treated or control males and drug-naive females. This experiment (Exp.2) was absolutely independent from the main one (Exp.1) and original parents were obtained even from different supplier (see Materials and Methods). As we can see, the results are qualitatively the same. However we should note here that the absolute values of naloxone-induced weight loss are different. And the most impressive difference between these two experiments can be seen in the absolute values of weight change during 5.5-day morphine treatment. Morphine treatment, applied before naloxone administration, is an essential part of this test. In fact, obtained results are extremely age-dependent. In the younger animals (age 70 days, Fig S66f) the absolute values of naloxone-induced weight loss are significantly smaller than in the older animals (age 150 days, Fig. S66h). And what is more, younger animals have acquired weight during 5.5-day morphine treatment (about 18-24 g, Fig. S66f). Older animals have lost their weight during the same treatment period (about 8-18 g, Fig. S66h). Thus, weight loss during 5.5-day morphine treatment and further weight loss during 24 hours after naloxone administration can serve as very useful indicators of opiate-related phenotype. However these

indicators are strongly age-dependent (only groups with identical age can be compared or one can apply special non-linear data correction procedure).

### 2.6. At least two epigenetically modified loci are involved in morphine effects in the $F_2$ progeny

Enhanced analgesic effect of morphine during the first morphine injection looks like dominant trait in $F_1$ males ($P < 0.00024$) and recessive trait in $F_1$ females ($P < 0.44$). In the second generation ($F_2$) the observed dominance is reversed: enhanced analgesic effect of morphine during the first morphine injection is recessive in $F_2$ males ($P < 0.022$) and it is dominant in $F_2$ females ($P < 0.00012$). These significance levels are obtained using non-parametric Mann-Whitney U-test and these results are shown in the Figs. S61-S62 (abscissa).

Enhanced analgesic effect of morphine during the second morphine injection is not statistically significant (0.11, 0.98; 0.088, 0.36), but animal distribution is remarkable here. In the Figs. S61-S62 I have drawn provisional lines of demarcation between normal and abnormal phenotypes, looking at both $F_1$ and $F_2$ generations simultaneously. In the first generation ($F_1$) during the second morphine injection the enhanced analgesic effect is practically absent in males (Fig. S61a, ordinate). In females during the second morphine injection this trait is absolutely absent in both $F_1$ and $F_2$ (Fig. S62a,c, ordinate).

However in the $F_2$ males we can see the following.

1. During the first morphine injection the enhanced analgesic effect exists in 5/16 of all experimental $F_2$ males (approximately in ¼ part).
2. During the second morphine injection the enhanced analgesic effect exists in 1/8 part of all experimental $F_2$ males.
3. Above-mentioned 1/8 part that has shown increased analgesic effect during the second morphine injection is clearly divided into two equal subgroups: a) 1/16 part of all experimental $F_2$ males has shown increased analgesic effect during both the first and the second morphine injections; b) 1/16 part of all experimental $F_2$ males has shown enhanced analgesic effect only during the second morphine injection, but not during the first one.

The appearance of 1/16 in $F_2$ is a sign of involvement of two independent loci.

In the experimental setup with drug-treated P males and drug-naïve P females all $F_1$ animals are supposed to be heterozygous, because they have received epigenetically modified allele from male and epigenetically normal allele from female.

During the analysis of epigenetically modified traits we will use the following convention. The lowercase letter designates the normal allele and the uppercase letter designates the epigenetically modified allele.

We do not use lowercase and uppercase letters for distinction of dominant and recessive traits, because dominance/recessiveness is very complex in the field of transgenerational epigenetic inheritance (it is not something like a fixed feature here).

For two loci we will have the following Punnett square (see the next page).

$F_2$ generation

| | | Gamete $F_1$ ♂ | | | |
|---|---|---|---|---|---|
| | | AB | Ab | aB | ab |
| Gamete $F_1$ ♀ | AB | AABB | AABb | AaBB | AaBb |
| | Ab | AABb | AAbb | AaBb | Aabb |
| | aB | AaBB | AaBb | aaBB | aaBb |
| | ab | AaBb | Aabb | aaBb | aabb |

Allele A in homozygous state AA entails enhanced analgesic effect during the first morphine injection in the $F_2$ males and this result is independent from another allele (b or B) sitting in another locus. These male animals have epigenotypes AABB, AABb, AABb, AAbb, and taken together they are ¼ part of all experimental $F_2$ males.

Alleles A and B in homozygous state AABB entail enhanced analgesic effect of morphine during the first and during the second morphine injections in the $F_2$ males. These AABB animals are 1/16 of all experimental $F_2$ males. Total number of tested experimental $F_2$ males is 80 and 1/16 is 5 rats, see Fig. S61c.

Male $F_2$ animals with wild-type allele a in homozygous state aa and with epigenetically modified allele B in homozygous state BB have normal analgesic effect during the first morphine injection and they have enhanced analgesic effect during the second morphine injection. These aaBB animals are 1/16 of all experimental $F_2$ males.

Thus, during the second morphine injection only $F_2$ males with epigenotypes AABB and aaBB have shown enhanced analgesic effect. These AABB and aaBB animals taken together are 1/8 of all experimental $F_2$ males. To have this effect these animals should have epigenetically modified allele B in homozygous state BB and they should have the first allele a or A in homozygous state aa or AA.

If the first allele is in the heterozygous state Aa, the presence of allele B in homozygous state BB is not sufficient to induce enhanced analgesic effect during the second morphine injection.

It means that heterozygous state aA provides some dimension of plasticity and due to this plasticity the effect of allele B in homozygous state BB on male phenotype can be decreased up to non-significant level.

There are also other factors and other dimensions of plasticity those are involved into morphine effects, because $F_2$ females do not show any effects of $F_2$ males ($F_2$ females are 100% normal during the second morphine injection, see Fig. S62c).

It means that generally greater ontogenetic plasticity of females provides additional possibilities for developmental compensation of phenotypic effects of epigenetically modified alleles.

We know that above-mentioned analysis can not be considered as a proof of something, but it provides important clues and shows a variety of factors which can form transgenerational epigenetic effects in different experimental situations.

## 3. CAGE ENRICHMENT EXPERIMENTS

### 3.1. Enriched living conditions are more beneficial for hybrids than for others

Our experiment with enriched housing conditions and hybrid and inbred mice has discovered new phenomenon of general interest, which is far beyond laboratory experimentation. These results can be extrapolated to human population without significant limitations. It was shown that enrichment of living conditions is more beneficial for development of superior behavioural phenotype in hybrids than in usual inbred animals. Previously it was very well known that hybrids can demonstrate better learning abilities than inbred strains (so called "hybrid vigour" or "heterosis"). However it was absolutely unclear how improvement of living conditions can change or modify this situation. In our study it was shown that enrichment of living conditions can improve behavioural phenotypes in both inbred and hybrid animals, but improvement in the case of hybrids is dramatically higher. As a response to enriched living conditions hybrids can develop superior learning abilities in comparison with inbred animals. This observation is obviously pleasant for a lot of multinational countries, including US, at least at the philosophical level. Animal studies are commonly used to find out some conclusions applicable to humans (however some reasonable precautions are always necessary).

In our study it was shown that ontogenetic development of hybrid vigour not only can be modulated by external conditions, but already superior behavioural properties of hybrids can be enhanced even further if these hybrids were living in the enriched environment during their ontogenesis (between weaning and puberty, postnatal days P21-P60 in mice). Simultaneously, the same environment does not induce equivalent improvement in inbred animals (it induces some improvement in inbred animals, but it is statistically less significant than in hybrids). Enhanced behavioural hybrid vigour can be detected in the "enriched" hybrids many-many months after the end of enrichment period (probably up to the end of their life). The importance of environment for behavioural hybrid vigour formation was not shown previously. Previously it was assumed that hybrid vigour (heterosis) is something like mechanical consequence of maternal and paternal genotypes. Honestly, above-mentioned simplified notion was not supported by all research groups, especially in plant science. For example, Trofim D. Lysenko discussed the development of hybrid vigour and dominance in 1935 in the following terms. "We maintain that in all cases when a hybrid plant is given really different conditions of existence for its development, this cause corresponding changes in dominance: *the dominant character will be the one that has more favourable conditions for adapting itself to its development*. We repeat that dominance is not predetermined in one particular direction in the zygote. Combining both parental hereditary foundations, the zygote contains the *potentiality* of developing both allelomorphs. The question of dominance is decided by the adaptedness of either allelomorph to develop under the given conditions of existence. Consequently, what will be dominant under one set of conditions will be recessive under another" (Lysenko, 1954, p. 83).

Views of T.D. Lysenko concerning development of hybrid vigour in plants look very reasonable today, as well as his thoughts about transgenerational effects (he did not use modern term "epigenetic"). However due to known political and other reasons (Soyfer, 1989, 2001) it is difficult to say whether his ideas have an impact upon development of modern science.

It was demonstrated that heterosis is an environmentally modified quantitative phenotype, but the positive role of environment is still underestimated.

Modern explanations of heterosis can be found in the following reviews: (Hochholdinger & Hoecker, 2007; Lippman & Zamir, 2006; Birchler *et al*., 2003). Most of them are focused on plants. Despite very

interesting discoveries made in plants concerning epigenetic inheritance and transgenerational effects (*e.g*. Durrant, 1962; Molinier *et al*., 2006) we shall focus our attention upon even more complex phenomenon: animal behaviour and ontogenesis of nervous system.

### 3.2. The enhancement of hybrid vigour by enrichment reveals creative neuroontogenesis

In our experiment with hybrid mice and enrichment of housing conditions we have shown that postnatal neuroontogenesis can create beneficial behavioural phenotypes on the basis of unexpected heritable changes (Fig. S88). Cage enrichment during days P21-P60 enhances hybrid vigour in B6D2F1 mice [$F_1$ obtained from C57BL/6J females and DBA/2J males]. Enhanced hybrid vigour can be observed during the rest of their life in a variety of operant behavioural tasks.

Thus, the phenotypes of progeny from drug-treated males are formed by cooperation of transgenerational epigenetic inheritance and creative neuroontogenesis.

Creative neuroontogenesis observed in $F_1$ generation can facilitate selective attenuation of some heritable epigenetic changes and it can promote selective reinforcement of some other heritable epigenetic changes. The consequences of this process can be seen in $F_2$ and further generations.

Water-maze test in the study of Wolfer *et al.* (2004) has revealed environmental induction of hybrid vigour (Fig. S88a). Escape latency is the most operant indicator of performance in the Water-maze (the presence in water is a negative reinforcement for mouse). Enriched $F_1$ hybrids are absolutely the best performers in Water-maze. The effect of enrichment is significant only in the hybrid mice (average escape latency of 16 training trials was taken as an indicator of performance): C57BL/6J, $P < 0.80$; DBA/2J, $P < 0.44$; B6D2F1, $P < 0.0016$ (Mann-Whitney U-test). In the Water-maze enrichment did not change group ranks, because standard B6D2F1 occurred to be the second overall performers. In this test enrichment just elevates pre-existing hybrid vigour.

In $F_1$ hybrids Water-maze general performance was increased by two ways: due to slight decrease of average swim path length ($P < 0.046$) and due to increase of average swim speed ($P < 0.000074$) (Fig. S89a-b). Swim speed was increased in the inbred C57BL/6J mice also ($P < 0.0099$), but, unfortunately for C57BL/6J, their swim path length occurred to be slightly increased and finally they were unable to show any improvement of escape latency (Fig. S89a-b; Fig. S88a).

Probe trial (without real target platform) also has revealed some deviation in behaviour of enriched hybrids. Surprisingly, there were no differences in the number of target (trained) annulus crossings (Fig. S89c). However enriched $F_1$ hybrids have shown increased number of adjacent annuli crossings ($P < 0.0025$; Fig. S89d). Looking at the overall escape latencies during acquisition trials we can say that adjacent annuli crossing occurred to be more efficient behaviour than behaviour of control and inbred groups *de facto*.

At the age of 7 months we took mice (C57BL/6, DBA/2 and their $F_1$ hybrids B6D2F1; standard and enriched; 96 in total) from Wolfer *et al.* (2004) study for sound discrimination tasks and auditory event-related potential (ERP) recording. We tested 48 mice (8 per group) in visual-tactile and olfactory discrimination tasks (hole-board-based), sound-frequency and sound-duration discrimination Go/NoGo operant tasks (shuttle-box-based). Independent subset of 48 mice was used for EEG recording. Auditory event-related potentials (ERPs) were recorded in awake, freely moving animals.

Auditory discrimination was performed in the shuttle-box Go/NoGo paradigm, which is an operant task with strong negative reinforcement. Animals were trained during 7 days (40 “Go” and 40 “NoGo” trails daily) to discriminate between sounds of different frequencies (2.5 and 10 kHz *vs*. 5 kHz). Difference between number of correct responses (correct Go) and wrong responses (mistaken Go) are shown in Fig. S88b. Highly significant strain effect was found ($P < 10^{-7}$, ANOVA): C57 mice and $F_1$ hybrids occurred to be good performers, whereas DBA were unable to learn this task (due to known age-related hearing loss). Significant enrichment × strain interaction ($P < 0.045$, ANOVA) was found in this sound **frequency**

discrimination task: enrichment improves performance in $F_1$ hybrids only ($P < 0.027$, Mann-Whitney U-test). After 7 days of task-free period the animals were trained in Go/NoGo sound **duration** discrimination task to discriminate between sounds of different duration (150 *vs*. 50 ms). Significant strain effect ($P < 10^{-6}$, ANOVA), enrichment effect ($P < 0.0014$) and strain × enrichment interaction ($P < 0.0041$) were found (Fig. S88c). Very impressive effect of enrichment has been seen in the $F_1$ hybrids ($P < 0.0011$, Mann-Whitney U-test). Standard $F_1$ hybrids occupied intermediate position between C57 and DBA, whereas enriched $F_1$ hybrids have shown absolutely the best performance. It is remarkable that environmental induction of hybrid vigour is more significant in more difficult task.

Surprisingly, in the sound frequency and sound duration discrimination Go/NoGo tasks enrichment did not increase the number of correct responses ("Correct Go"; Fig. S90b,d). However enriched $F_1$ hybrids have shown significantly decreased number of wrong responses ("Mistaken Go"; Fig. S91a,c). It was observed both in sound frequency ($P < 0.023$) and sound duration ($P < 0.030$) discrimination tasks.

Taken all mentioned above results together we can conclude that cage enrichment uncovers dormant hybrid vigour both in Go/NoGo and Water-maze tasks. Both auditory discrimination and spatial orientation can be improved by cage enrichment specifically in hybrid mice, but not in the inbred strains. One of the parental strains (DBA/2J) has age-related hearing loss, but our conclusion is not limited to the situation of functional disability in one of the parents. Parental strains DBA/2J and C57BL/6J have not detectable physiological problems in Water-maze, in which environmental induction of hybrid vigour in B6D2F1 was shown also.

Auditory event-related potentials of standard and enriched C57BL/6J, DBA/2J and B6D2F1 mice are shown in Fig. S100. Enrichment induced qualitatively similar changes in the shape of event event-related potential in C57 and $F_1$ mice: the local maximum in the 75-100 ms time window occurred to be increased, whereas the next local maximum in 150-200 ms time window was decreased (Fig. S100a,c). The strongest behavioral effect of enrichment seen in $F_1$ can not be explained by the difference in N1 amplitude (negative peak around 35-50 ms; Fig. S100c), but it is associated with some events in the 75-200 ms time window, because impressive enrichment-induced changes in the shape of auditory event-related potential in $F_1$ were found here.

The differences in ERP shape in "standard" and "enriched" mice (Fig. S100a,c) are qualitatively similar to the differences between "standard ERP" and "duration deviant ERP" in male C57BL/6J mice in duration deviant paradigm (Fig. S45d). It means that "enriched" mice have a tendency to process standard stimulus as a duration deviant stimulus in "standard" mice is processed. In addition, "enriched" mice have a tendency to process duration deviant stimulus as more distinctive duration deviant stimulus (Fig. S102a).

During EEG recording session of each animal 4 subsequent paradigms were applied (Fig. S47d-g). Detailed analysis of auditory ERPs has revealed several interesting facts. **First**, there is systematic decrease in N1 amplitude from Paradigm 1 to Paradigm 4 (see especially Fig. S101e-f). It means that event-related potentials are in general history-dependent phenomena, similar to animal behaviour, observed in different operant and exploratory paradigms (note that in this experiment all animals used for EEG recording were not exposed to any special sounds during their previous life). **Second**, in the standard $F_1$ hybrids there is dramatic change in ERP shape in 80-180 ms time window during paradigms 1-4 (Fig. S101e), whereas ERP shape of the enriched $F_1$ hybrids looks absolutely stable (Fig. S101f). Taken into account better performance of enriched hybrids in all operant paradigms, we have to conclude that despite standard hybrids still have, perhaps, some plasticity, enriched hybrids were able to develop on the basis of this plasticity efficient local circuitries in advance, during enrichment period. **Third**, in duration MMN paradigm standard $F_1$ hybrid mice have shown strong symmetric positive and negative deflections in paradigms with long and short deviants (Fig. S103a), whereas enriched $F_1$ hybrids did not show such physical-features-dependent activity, but, instead, they have shown better novelty-dependent activity (duration MMN and increased amplitudes of N1 and P2 in frequency MMN paradigm and mixed duration-frequency paradigm; Fig. S103b,d,f,). It seems that enriched $F_1$ hybrids have local brain circuitries which allow them to suppress inefficient activities both at electrophysiological (ERP) and behavioural levels (see Go/NoGo wrong responses, Water-maze path length), whereas useful activities can be selectively reinforced.

### 3.3. Reconsolidation of previously disrupted cascades can not explain hybrid vigour

These phenomena can not be explained by reconsolidation of previously disrupted cascades (mechanisms), *i.e.* by hypothesis used usually for explanation of the origin of hybrid vigour. All necessary features were not written in the heredity (in genetic or epigenetic heredity) in preexisting form, but, instead, they were developed by means of creative ontogenetic process on the basis of a variety of unexpected heritable changes (new changes, originated from crossing of very different inbred maternal and paternal strains).

The most important conclusion from the experiments with hybrid mice is that unexpected heritable changes (presumably both genetic and epigenetic) lead to development of useful phenotype by means of ontogenetic processes that are trying to build helpful brain circuitries using unexpected heritable changes as new resources. These ontogenetic processes seem to be entirely creative. These processes do not necessarily require the existence of long feedbacks from, for example, real animal behaviour. They can probably act at the level of local brain circuitries, being looking at their local efficiency. These local brain circuitries can be used later for organization of animal behavior.

We are using the word "creative" here not as a metaphor. We use it as an exact term, the term that was introduced by Henri Bergson in 1907 in his book "*Creative Evolution*" (see ref.: Bergson, 1998).

The next question is whether these creative ontogenetic processes can induce new heritable epigenetic changes or modify heritable epigenetic changes, which were initially induced, for example, by paternal drug treatment. Significant difference between $F_1$ and $F_2$, observed in our experiments with morphine and thyroxine, suggests that above-mentioned expectation is quite realistic. Of course, this hypothesis needs further validations.

Creative neuroontogenesis, observed in hybrid animals, leads to additional question about genome (and epigenome) activity in the germ cells of hybrid animals. This question is independent from any hypothesis of brain-germ line interaction and can be formulated by the following way: if there is a heterozygous animal, will it produce different descendants being at the age of 2, 3 or 6 months? (The numbers are taken for mice). If there are creative germ line processes that optimize epigenetic heredity, the progeny from 6-month old hybrid animal should statistically differ from the progeny from 2-month old hybrid one.

To check this possibility it is possible to take mice C57BL/6J, DBA/2J and their hybrid B6D2F1 and cross male animals at the age of 2, 3 and 6 months with 3-month old females of C57BL/6J and DBA/2J inbred strain (C57BL/6J females are better for fostering, but DBA/2J females are better for detection of epigenetic effects in the progeny; other option is to use DBA/2J females as biological mothers, but then females from outbred stock (*e.g.* NMRI) should be used as foster mothers). Statistically sufficient number of descendants should be tested in a set of exploratory and operant tasks at the age, standard for particular tasks (Place navigation in the water maze, Go/NoGo sound discrimination task, other operant tasks). Operant tasks are important here, because only operant tasks can reveal positive or negative changes by objective way. Exploratory tasks can reveal important changes also, but interpretation of all these changes remains entirely subjective. As an example we can take superior performance of hybrid and wild animals in all operant tasks and confusing results of the same animals in different exploratory tests (wild mice sometimes just "do not work" in standard laboratory exploratory tests).

### 3.4. Hybrid fathers have a set of epigenetic processes, activated by genetic novelty

The comparison of the progeny from inbred and hybrid fathers is important here, because hybrid fathers can have a set of epigenetic processes, activated by genetic novelty. These processes can lead to optimization of epigenetic heredity and they can be absolutely absent in inbred animals. Transgenerational effects of above-mentioned optimization processes should be a target for further studies. Note that these processes theoretically can go locally (only inside germ cells) and can be ruled by local efficiency. An interaction of above-mentioned germ-cell processes with the body or brain can exist in real nature, but it should be treated as an independent hypothesis.

The existence of above-mentioned interaction between germ cells and brain or body in the hybrid animals can be investigated at the experimental level. For this purpose it is necessary to compare described above experiment with similar experiment where experimental male parents (both hybrid and inbred groups) will be housed under enriched conditions. If the results of experiments with standard and enriched housing conditions will be about the same, we should not assume the existence of significant interactions between germ cells and brain or body under particular experimental conditions. Otherwise, these interactions should be taken into account. On the basis of out experiments with morphine and thyroxine it is possible to predict that some kind of indirect interactions between germ cells, brain and body does exist in nature.

### 3.5. Exploratory tasks can not demonstrate hybrid vigour, but can reveal useful behavioural patterns

At the age of 2-3 months all standard and enriched mice were tested in the set of classic exploratory tasks: a) Elevated 0-maze; b) Open-field; c) Object exploration test (Fig. S96). All of them have revealed highly significant statistical differences between standard and enriched mice in both inbred strains (C57BL/6J and DBA/2J) and hybrids (B6D2F1), but nothing that can be described as hybrid vigour.

At the age 11 months (*i.e*. 9 months later) we have tested the subset of 48 mice in the independent exploratory task (Hole-board).

During the last 25 years Hole-board task has evolved from classic version with manual recording of mouse behaviour towards modern computer-controlled system with infrared beams under the floor. Modern system can register duration of exploratory episodes to within 10 ms (using the same classic geometry with 16 holes (d = 25 mm) in the 40 × 40 cm arena, surrounded by vertical walls). In the classic Hole-board the single specific indicator of behavior was the number of nosepokes (for example, during 6-min session). In the modern setup it was discovered that two other indicators have better discriminative power: 1) "total exploration time" (this time is about 40-70 s for 6-min session) and, especially, 2) mean "nosepoke duration" (typical nosepoke duration is about 0.5-1.2 s). Contrary to straightforward logic, mean nosepoke duration can provide higher statistical significance than the number of nosepokes or total exploration time.

Video tracking is also useful for hole-board task. It can not be used instead of nosepoke registration, but systems with simultaneous video tracking and infrared beams under the floor are available now. The best ones can video track 4 animals in 4 hole-board units with concurrent registration of all nosepoke activity (for additional information see www.evolocus.com ). In our own study, discussed in this Supporting Online Material, we did not use video tracking during hole-board test (but infrared beams under the floor were used). Our system consisted of one hole-board unit.

In addition to above-mentioned technical progress, there is significant progress in hole-board protocols. Today after application of classic hole-board protocol during the first two days (for investigation of habituation) it is reasonable to apply al least one day of visual-tactile discrimination and at least one day of olfactory discrimination. During visual-tactile discrimination test at least one hole of hole-board is substituted for hole with different shape [for example, circular hole is substituted for star-shaped hole], and/or the floor material, corresponding to at least one above-mentioned hole, is substituted for material which differ in color and/or surface texture; above-mentioned material can also have entirely different origin and/or production history [as, for example, plywood in comparison with polyvinyl chloride]. See Fig. S96d as an example. During olfactory discrimination test the unusual odour is placed under at least one hole of hole-board [it can be, for example, Mint odour]. Usually during olfactory discrimination test the classic floor with 16 circular holes is used. There are publications with 4-hole hole-board for olfactory discrimination, but the decrease in the number of holes leads to significant decrease in discriminative power (*i.e*. efficiency) of hole-board test, at least in accordance with our observations. During both visual-tactile and olfactory discrimination we used equal numbers of new and old holes (*i.e*. 8 circular holes *vs*. 8 star-shaped holes and 8 holes without odour *vs*. 8 holes with Mint odour). Minimal modern protocol consists of 4 days: days 1 & 2 – classic hole-board, day 3 – visual-tactile discrimination, day 4 – olfactory discrimination. In our study we did olfactory discrimination twice: during days 4 & 5.

During the first day (classic hole-board) animals of all genotypes did not show any differences between standard and enriched subgroups in all 3 variables (number of nosepokes, total exploration time and nosepoke duration, see Fig. S97a-c). However during the second day (classic hole-board) enriched hybrids have shown increased total exploration time (Fig. S97b) and increased nosepoke duration (Fig. S97c). Note that in this test the enrichment has pushed the phenotype of B6D2F1 from C57BL/6J level towards DBA/2J level (Fig. S96f, S97b-c). On the other hand, we remember that in the Go/NoGo task (Fig. S91a) the same enrichment has pushed "wrong responses" of the same B6D2F1 from DBA/2J level towards C57BL/6J level. Thus, in terms of dominance, it is impossible to say as a general conclusion that, for example, in hybrid B6D2F1 mice the enrichment converts normally dominant C57BL/6J-like phenotype to DBA/2J-like phenotype, because for some other traits the effect of enrichment is opposite and it converts normally dominant DBA/2J-like phenotype to C57BL/6J-like phenotype.

In addition it is interesting to mention that the second day of classic hole-board has better discriminative abilities than the first one, at least in females. Better significant differences were obtained between important groups during the second day in female mice (Fig. S97b-c) (see our cage enrichment experiment) and in female rats (Fig. S66b) (see our morphine experiment). An additional day of testing can reveal important differences between groups not only in the exploratory tasks like hole-board, but in the passive avoidance tasks like step-down (Figs. S67d-e, S66c) (in this case the differences were better in males; Wistar rats, morphine experiment; days 1 and 2 are not shown in the Fig. S66c). Step-down is a commonly used one-trial learning technique (the first day – training, the 2-d day – testing). An additional day of step-down training (also one-trial) can really improve discriminative power of this method. But, of course, this protocol has known limitations concerning drug-discovery studies (because it consists of 3 days: the first day – training, the 2-d day – training-testing trial and the 3-d day – testing). However for descendants of drug-treated animals and for knockouts an additional day of testing in hole-board or step-down can be very useful.

During the third day (visual-tactile discrimination) the effect of enrichment was revealed not only in hybrid B6D2F1, but in inbred DBA/2J also (Fig. S97d,f). Surprisingly, first of all, that statistical difference was observed not in the new star-shaped hole exploration (green bars), but in the old circular hole exploration (white bars). The second surprise is that in the DBA/2J the enrichment has led to decreased nosepoke number (Fig. S97d), but increased nosepoke duration (Fig. S97f), whereas total exploration time occurred unchanged (Fig. S97e). Visual-tactile discrimination *per se* was not modified by enrichment up to significant level (Fig. S96e).

During the fourth day (the first day of olfactory discrimination) visual-tactile discrimination *per se* was significantly modified (improved) in hybrid mice only (Fig. S96f). No significant enrichment-induced changes were observed in the other variables (Fig. S98a-c). It is interesting that all DBA/2J mice strongly avoid new mint odour, whereas all C57BL/6J do not show any discrimination at all (Fig. S98a-c).

During the fifth day (the second day of olfactory discrimination) the effect of enrichment was registered in hybrid B6D2F1 mice as increased exploration time of holes without odour (Fig. S98e) and as increased nosepoke duration (but, once again, nosepoke duration was increased only with respect to holes without odour; Fig. S98f).

Looking at the strain-related differences with respect to enrichment effect in the hole-board task we can conclude that: 1) the strongest effect of enrichment was observed in hybrid B6D2F1 mice, 2) very mild effect of enrichment was observed in inbred DBA/2J mice and 3) the absence of enrichment effect was observed in inbred C57BL/6J mice.

Looking at the same data from methodical viewpoint we can say that: 1) the strongest differences were observed in nosepoke duration, 2) rather mild differences were observed in total exploration time and 3) the practical absence of differences was observed in nosepoke number.

To draw the conclusion for transgenerational epigenetic inheritance from the experiment with enrichment of housing conditions we have to look at the time frame of the whole experiment with enrichment of housing conditions (Fig. S99b). We see that relatively short period of enrichment during P21-P63 has led to

significant behavioural differences registered many months later, at the age of 7 and 11 months. In DBA/2J strain the enrichment effect was significant up to the end of their life in the form of decreased survival rate (registered by us at the age of 22 months, Fig. S99a). DBA/2J strain was chosen by us as a basic strain for investigation of transgenerational epigenetic effects of paternal thyroxine treatment. Looking backward we can speculate that this choice was probably better than more common choice of relatively "stable" C57BL/6J strain.

## 4. *GDI1*-DEFICIENT MICE EXPERIMENTS

In our electrophysiological experiments with EEG and auditory event-related potentials recording we have found that local neural circuits can improve organism's profit by selective participation in the processes where their participation makes a difference. Neurons can self-regulate their own activity looking at its efficiency with a help of local feedback. It was shown in the experiments with descendants of thyroxine-treated mice, with inbred and hybrid mice after enriched housing conditions and with mutant *Gdi1*-deficient mice [mice with slowed neurotransmitter endo- and exocytosis].

### 4.1. *Gdi1*-deficient mice – good tool for investigation of general neuronal mechanisms

*Gdi1*-knockout mice are animals with slowed neurotransmitter endo- and exocytosis (see Materials and Methods for references). Usually researchers are investigating knockout mice to demonstrate some important differences between knockouts and controls. Here we would like to use knockout mice for some other purpose: we would like to show that regularities in our particular *Gdi1* knockouts are qualitatively similar to control mice, but quantitatively they are expressed in more demonstrative manner. In other words, some general features are more pronounced in particular knockout mice. Thus, this knockout should be used for investigation of particular general mechanisms. We have investigated several very different mutant mice (Umbricht *et al*., 2004, 2005; Bickel *et al*., 2007, 2008). And *Gdi1* knockout was chosen by us due to its fitness to our practical purpose – for demonstration of the principles of neuronal activity and the origin of mismatch negativity (MMN). Of course, there are several statistically significant differences between *Gdi1*-deficient mice and controls, but it is another story. Here would like to focus our attention upon general mechanisms, which occurred to be more pronounced and easy for investigation in *Gdi1*.

### 4.2. Neurons are trying to produce activity that brings detectable electrophysiological results

In the Fig. S45e,f we have seen that paradigms with very short inter-stimulus intervals can produce diffused and very approximately time-locked activity instead of classic auditory even-related potentials. In the Fig. S104c,d we can see that this phenomenon is extremely strong in Gdi1-deficient mice (look at the red curve), but rather moderate in controls under particular stimulation conditions (10 kHz stimulus frequency, 100 ms stimulus duration). Despite this impressive abnormality seen in *Gdi1*-deficient mice with very short (300 ms) inter-stimulus intervals, *Gdi1*-deficient mice demonstrate very slightly decreased correct responses in operant sound frequency discrimination Go/NoGo task when inter-stimulus intervals are rather long (1000 ms) – see Fig. S104e-g. It is very important to mention that when we use other stimulation conditions, which seems to be not optimal for induction of classic evoked potential (5 kHz instead of 10 kHz) or just different (25 ms instead of 100 ms), in *Gdi1*-deficient mice N1 becomes smaller, and, simultaneously, the very approximately time-locked activity occurs to be more pronounced in control mice – see Fig. S105c-d, g-h (look at the green curve). Thus, *Gdi1*-deficient and control mice can show qualitatively the same delayed, diffused and very approximately time-locked electrophysiological activity, but optimal conditions for its observation are slightly different for these two groups.

We focus our attention upon above-mentioned slightly delayed, diffused and very approximately time-locked activity, because we would like to compare two types of neuronal activity: 1) classic evoked potential, which has P1, N1, P2 and is very-well time-locked with respect to stimulus; 2) delayed, diffused

and very approximately time-locked with respect to stimulus neural activity. Classic evoked potential exists when particular stimulus exactly fit to the features of particular neurons, whereas diffused potential exists when there is no exact coincidence between stimulus properties and features of available neurons. Usually, if we have a lot of identical repeated stimuli, diffused activity occurs to be completely suppressed, and we can see only activity of neurons with classic evoked potentials. However if we add several deviant stimuli in a row of identical ones, these deviant stimuli will induce both classic and diffused activity (in different neurons, of course). Here we have the roots of mismatch negativity. And it is not so important for the functionality of mismatch negativity whether neurons with classic and diffused activity will be literally in the same brain region or in slightly different regions. It is reasonable to expect that spatial position of averaged centre of diffused and classic activity will be slightly shifted with respect to each other. And, of course, if we change parameters of stimulation and parameters of deviant stimuli, some neurons with previously classic response can produce diffused response, and some neurons with previously diffused response can produce classic response.

### 4.3. Neurons can suppress their own activity, if it is not efficient enough

Usually, if we have a lot of identical stimuli, without any deviants, we can see only classic responses and diffused activity is completely self-suppressed. It is very important that this activity is self-suppressed and it is not suppressed by some external force, for example by means of concurrence (competition) with activity of neurons with classic responses (extreme example: neurons with classic responses suppress diffused activity of neurons with diffused responses). In the animals with age-dependent hearing loss (DBA/2J) it is possible to see duration MMN when any responses to standard stimuli are absent. Duration MMN can be seen as clear diffused activity (Fig. S45g-h). In this experimental situation there is no classic evoked potential and there is no neurons which produce classic evoked potentials, but all diffused activity induced by standard stimulus occurs to be suppressed. But deviants with different duration still can evoke diffused activity (activity which can be seen as MMN). The absence of diffused activity induced by standard stimuli can be explained only by self-suppression of this activity. This self-suppression can take place if and only if neurons have feedback from the results of their activity. If there are several repeated stimuli and particular neuron takes part in formation of classic evoked potential, it is not suppresses its own activity. Whereas if it can not participate in formation of classic evoked potential or classic evoked potential is not formed at all, this particular neuron suppresses its own activity, induced by particular stimuli. However if really new stimulus comes, this neuron can try to produce activity which can be effective (it will help to form classic evoked potential), but there is statistically more probable outcome – its activity will be ineffective and it will help to form diffused activity only.

Look at the *Gdi1* and control mice once again. *Gdi1* mice can show impressive diffused activity with delayed maximum (Fig. S104c-d). Control mice can show some diffused activity also, but its maximum is slightly delayed (Figs. S104c-d, S105c-d, g-h) and in general this diffused activity is much smaller. When we apply stimuli with different sound level, *Gdi1* mice demonstrate smaller evoked potentials, especially in situation with low sound frequency and low sound pressure level (Figs. S106c-d, e-f). In situation with optimal sound frequency and relatively high sound level there is no difference between *Gdi1* and control in P1, N1 and P2 (Fig. S104a-b). It is interesting that if stimuli have normal duration (100 ms), normal inter-stimulus interval (900 ms), normal frequency (10 kHz) and normal sound level (75 dB) there is no impressive difference between *Gdi1* and control (Fig. S108b), but if stimulus becomes twice longer (200 ms instead of 100 ms), very impressive difference appears in the time window 130-228 ms (Fig. S108a). In both cases the time is counted from the onset of sound.

Special remark concerning statistical methods is needed here. Mann-Whitney U-test was finally used for all *Gdi1*-deficient mice ERP data analysis, as it is mentioned in the graph legends. However, an important preliminary ERP data processing was used. For example, four independent records, each with significantly different duration, were used for each animal in the last experiment (Fig. S108). Also, EEG responses from both right and left hemispheres were utilized in our analysis, whereas only signal from the right hemisphere is shown in the figures. And, of course, filtered data were used for statistical analysis (filter settings see in this SOM, p. 19), whereas in the figures the original data are shown. In fact, shown significance levels can not be achieved by straightforward application of common statistical methods to the raw data.

### 4.4. Neurons are using feedback from their own activity for estimation of its efficiency

Previously we have concluded that under conditions of repeated stimuli all diffused activity should be self-suppressed with a help of local feedbacks. However in the experimental situation with very short inter-stimulus intervals (200-300 ms) we can see strong diffused activity in *Gdi1* and some diffused activity in control mice. Obviously, these stimulation conditions are not evolutionary optimal, but what are the real mechanisms of appearance of diffused activity here? One of the possible explanations is a false feedback. Neurons receive the second (n+1) stimulus, when they are expecting to receive feedback about the effect of their activity concerning the first (n) stimulus. Thus, due to repeated appearance of false feedback these neurons classify this situation as situation with normal classic evoked potential. It seems that *Gdi1* have slowed neurotransmitter exocytosis and slowed neurotransmitter endocytosis, and, probably, their window for feedback acceptance is wider than in normal mice. That is the reason for more impressive diffused activity in *Gdi1* mice. On the basis of this hypothesis, the expected time of feedback appearance is about 300 ms for *Gdi1* and about 200 ms for normal C57BL/6J. Alternative explanation is that these evoked potentials are in fact real classic evoked potentials, but they are cut from their tails. This explanation works badly for control mice and it does not work at all for *Gdi1*, because it does not explain the increased duration of negative peak (N1 equivalent) and the shift of its local extremum towards late period.

### 4.5. Local neural circuits can be self-evolving units, useful for behaviour of the whole organism

Thus, neurons of primary auditory cortex in mice can demonstrate the following features:

1. Neurons would like to be involved into processes where their participation makes a difference
2. Neurons can suppress their own activity, if it is not efficient enough
3. Neurons are using feedback from their own activity for estimation of its efficiency
4. Local neural circuits can be self-evolving units, useful for behaviour of the whole organism

In general, local neural circuits can improve organism's profit by selective participation in the processes where their participation makes a difference. Neurons can self-regulate their own activity looking at its efficiency with a help of local feedback.

### 4.6. Local neural circuits are self-evolving units during early ontogenesis

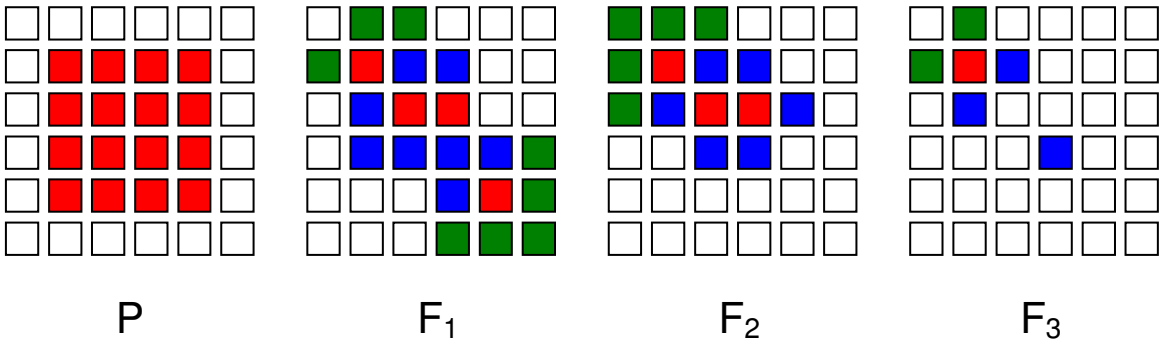
The most important part of application of above-mentioned statement is that it works not only during adult life (as it was investigated in our experiments with *Gdi1* and other mice), but during early ontogenesis also. Neural circuits during early ontogenesis can improve prospective profit of an adult organism by selective participation in the developmental processes where their participation makes a difference. Their activity during above-mentioned developmental processes is entirely creative and its local features are determined in part by current situation and not only by heritable factors (genetic and epigenetic). During this early ontogenesis neurons can self-regulate not only their own activity, but their spatial configuration also, looking at the integral efficiency of particular activity and configuration with a help of local feedback.

## GENERAL CONCLUSION

All above-mentioned materials lead us to the final conclusion: paternal drug treatment entails selective transgenerational transmission of several (not all!) epigenetic modifications (not all genes those had changed expression during drug treatment period will have associated heritable epigenetic changes as a consequence). Above-mentioned several heritable epigenetic changes induce during early ontogenesis of descendant creative response (response to particular inherited novelty). Creative ontogenetic response entails additional detectable phenotypic results. The results of this response are modified further by self-controlled neural activity which is trying to extract as much benefits as possible from particular developmental situation. And it is very important that some results of above-mentioned processes can modify heritable epigenetic changes further to induce significant differences between $F_1$ and $F_2$ generations. Qualitatively similar, but quantitatively different, processes take place during ontogenesis of $F_2$ generation and they lead to significant differences between $F_2$ and $F_3$. And it is still an opened question how these epigenetic changes (*i.e*. epigenetic evolution) are going to be transformed into classic genetic changes (*i.e*. classic genetic evolution). All above-mentioned phenomena are important for evolution of central nervous system – neuroevolution.

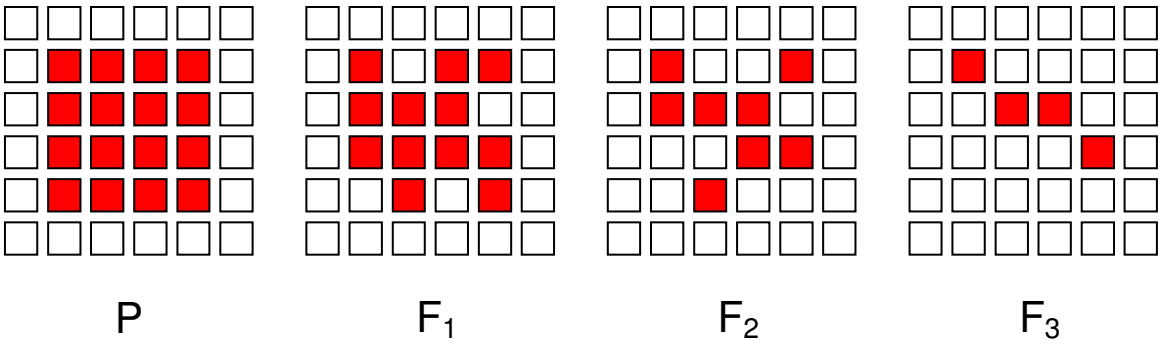
***P.S.***: The last question is about the existence of more efficient mechanisms of transition from epigenetic heredity (in modern sense, *e.g*. Graff & Mansuy, 2008) to genetic one. We assume the existence of more efficient processes than "genetic assimilation", proposed by Conrad H. Waddington (Waddington, 1942, 1953a-b; see also Pigliucci & Murren, 2003). We are not going to provide here complete historical review of all possible interactions between epigenetic and genetic heredity, but it should be mentioned that even the theory of "germinal selection" (*On Germinal Selection as a Source of Definite Variation*), proposed by August Weismann (Weismann, 1902), now can be easily reformulated (refreshed) by means of several relatively simple substitutions. For example, "determinant selection" [on the basis of their "strength"] can be replaced with "inheritance of epigenetic changes" [developed on the basis of functional activity of particular genes and their regulatory sites], because both processes provide the same final quantitative effect which can be expressed in modern terms as a change in gene expression or a change in the efficacy of particular regulatory site. From the purely mechanistic viewpoint "germinal selection" has a direct descendant in the form of post-meiotic (pre-zygotic) selection (Vyssotski, 2004, pp. 22-25, 503, 531).

**a** Obtained results (scheme)




- Intact phenotypic traits
- Drug-induced paternal phenotypic traits and their partial inheritance
- Opposite phenotypic changes in the progeny
- Qualitatively new changes in the progeny

**b** Classical view on the transgenerational action of "endocrine disruptors" and epigenetic inheritance




- Intact phenotypic traits
- Drug-induced paternal phenotypic traits and their partial inheritance

**Figure S1 | Summary of the main result**. **a**, Figure summarizing the main result of this article. **b**, Previous expectations. Each small square represents schematically one phenotypic trait (quantitative trait). Each cluster of 36 small squares represents the whole body of an organism or its chosen functional system.

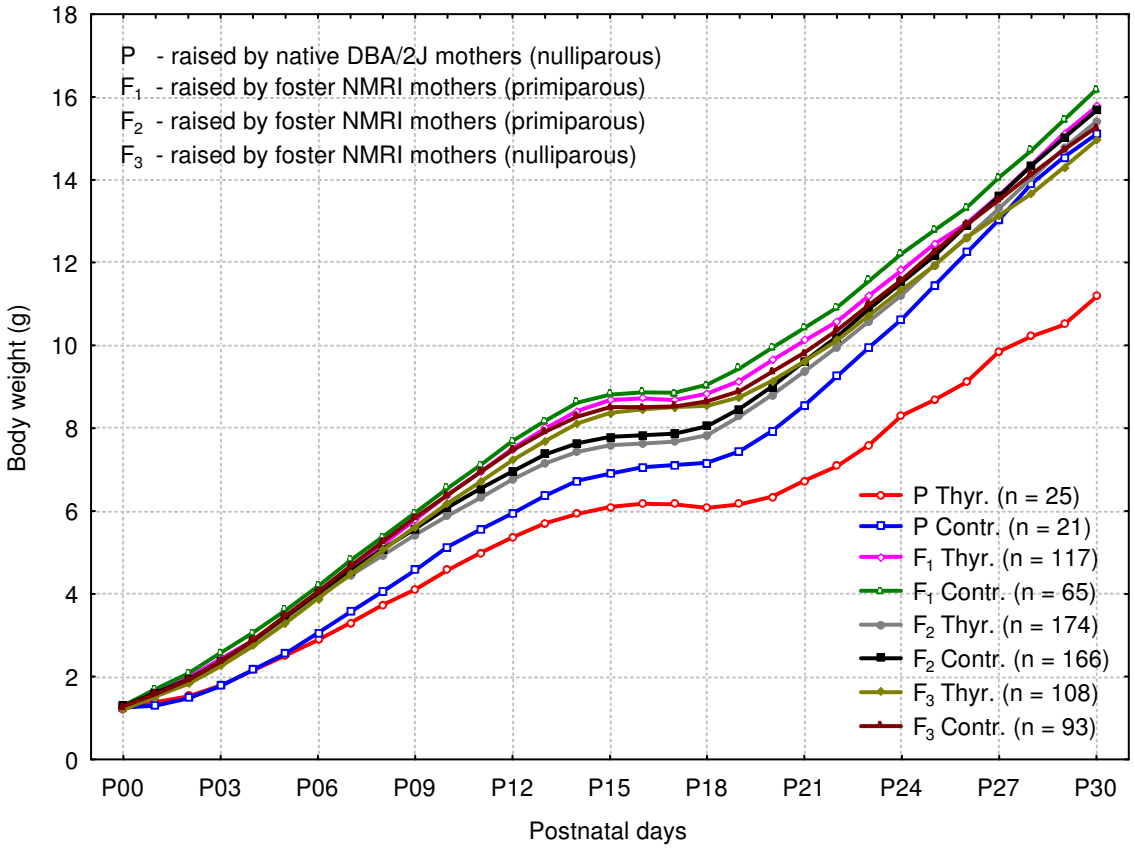



**Figure S2 | Postnatal body weight in thyroxine-treated mice and $F_1$-$F_3$ offspring.** DBA/2J mice. Postnatal P00-P30 body weight (g) in the P, $F_1$, $F_2$ and $F_3$ generations (mean). Note that all groups, except thyroxine-treated, have normal body weight at P27-P30, despite some early differences induced by the presence or absence of NMRI mice fostering (see control groups) or by ancestor's thyroxine treatment. The differences between primiparous ($F_1$-$F_2$) and nulliparous ($F_3$) NMRI fostering occurred to be smaller than the differences between DBA/2J (P) and NMRI ($F_1$-$F_3$) fostering. Some delay in the weight gain during P15-P18 seen in all groups can be explained by switching from mother milk to solid food consumption. It can be seen in other mouse strains also, *e.g.* in C57BL/6J. Note that the differences between experimental and control groups (except P generation) are usually smaller than the differences between control $F_1$, $F_2$ and $F_3$, tested in different calendar periods (seasons). Decreased body weight persisted in the treated animals (P) throughout their life (with high significance in both males ($P < 0.0020$) and females ($P < 0.030$); Mann-Whitney U-test). Brain weight was also decreased in these males ($P < 0.00023$) and females ($P < 0.014$) at the age of 270 days (mean ± SD):

| P generation | *n* | Body weight (g) | Brain weight (mg) |
|---|---|---|---|
| Thyr. males | 9 | 26.13 ± 1.40 | 435 ± 15 |
| Contr. males | 11 | 30.72 ± 2.77 | 476 ± 12 |
| Thyr. females | 9 | 21.99 ± 0.65 | 436 ± 20 |
| Contr. females | 8 | 23.24 ± 1.81 | 472 ± 28 |

No thyroxine-related differences in brain weight were observed in $F_1$, $F_2$ and $F_3$.

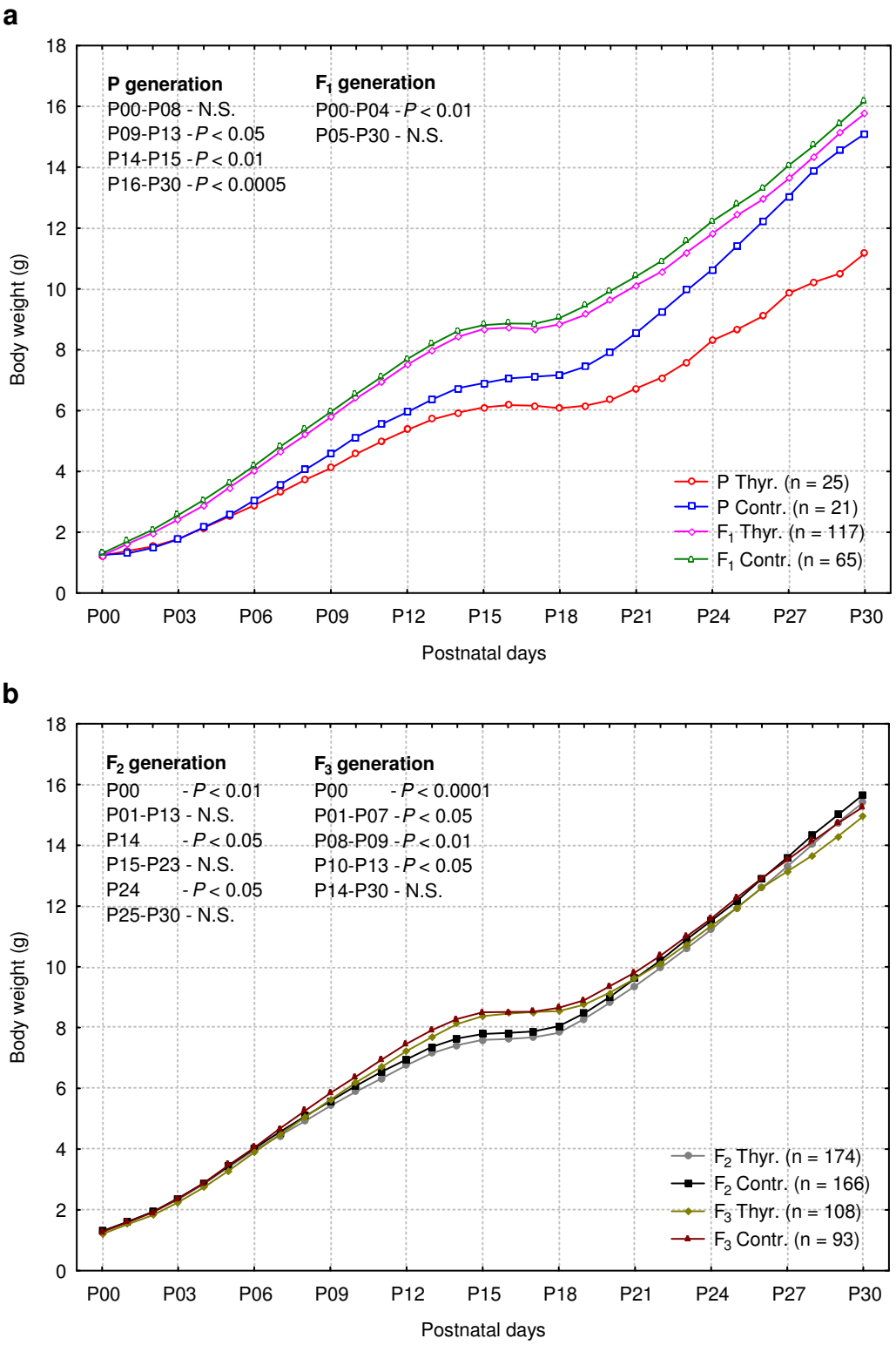



**Figure S3 | Statistics of postnatal body weight in thyroxine-treated mice and $F_1$-$F_3$.** **a**, P generation and $F_1$ generation. **b**, $F_2$ and $F_3$ generations. Note decreased weight in the thyroxine-treated group after P09 and decreased weight in the progeny during P00-P04.

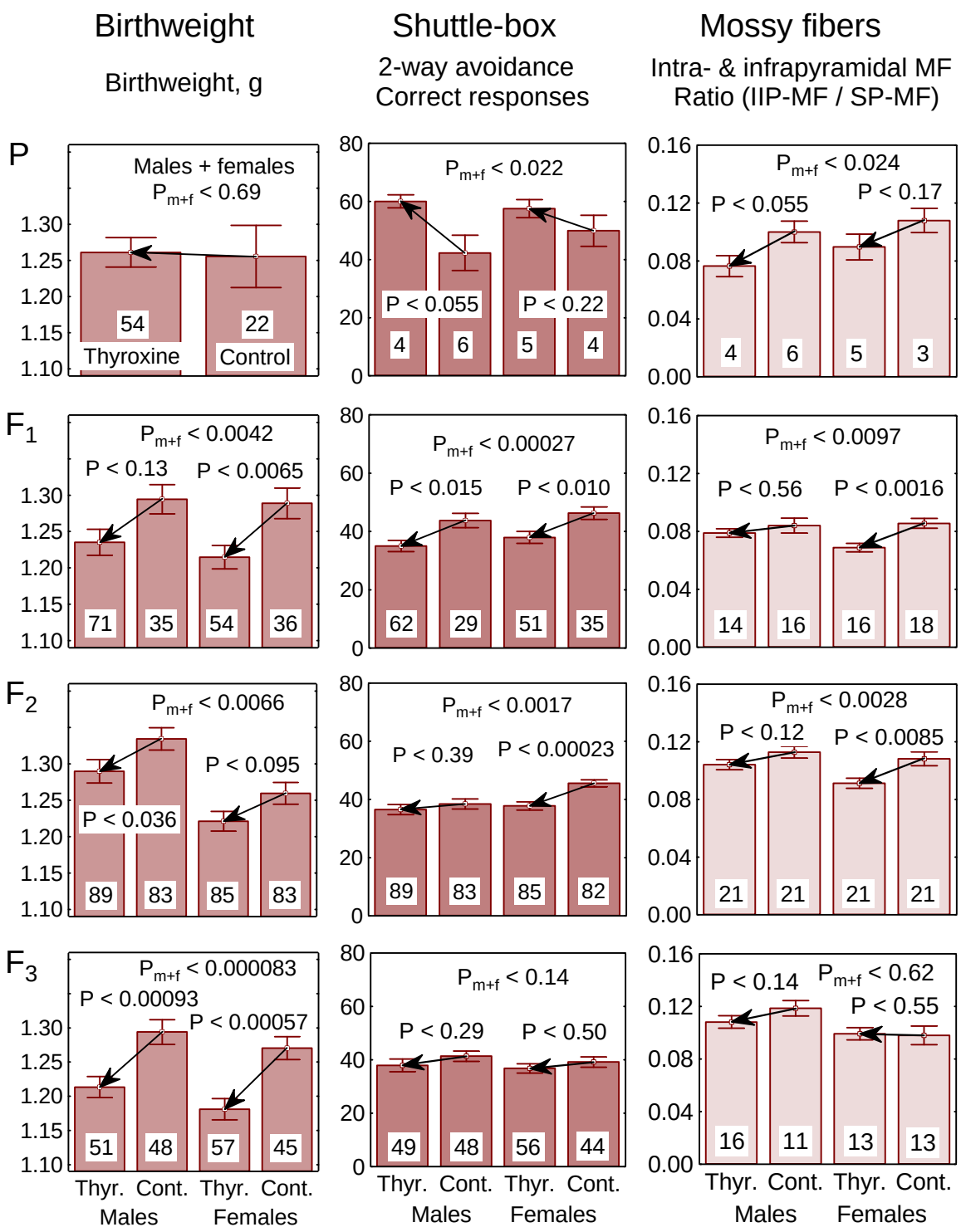



**Figure S4 | Phenotype of thyroxine-treated mice and $F_1$-$F_3$ descendants.** Neonatally thyroxine-treated mice and untreated descendants of thyroxine-treated males (mean ± SE). Birthweight. Two-way avoidance, averaged correct responses of 5-day training, 80 trials daily. Hippocampal intra- and infrapyramidal mossy fibers, ratio of intra- and infrapyramidal mossy fiber fields (IIP-MF) to suprapyramidal mossy fiber fields (SP-MF). Group size is shown as a number in a particular bar. Arrows indicate the direction of the changes, induced by paternal thyroxine treatment.

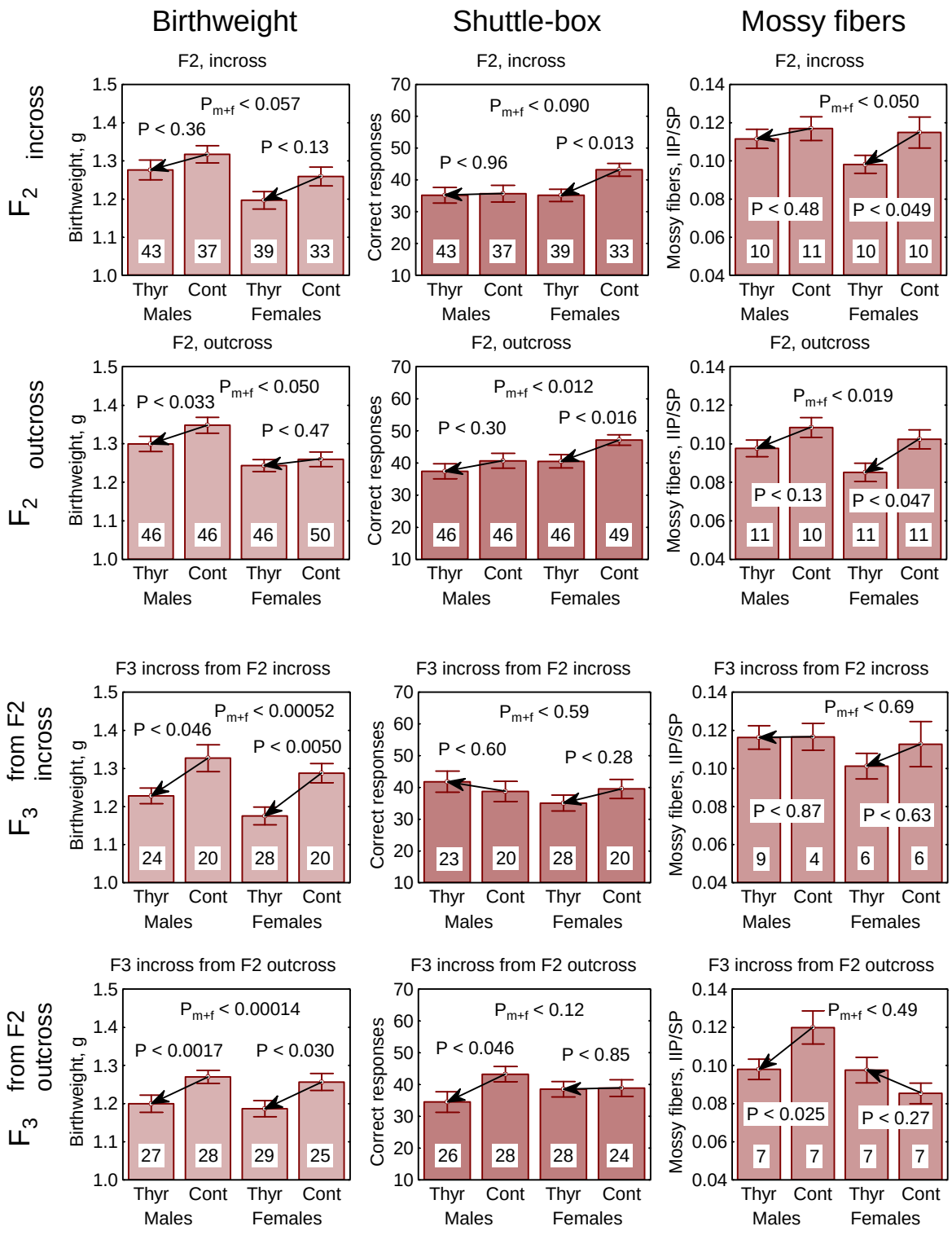



**Figure S5 | Phenotype of incross and outcross $F_2$-$F_3$ offspring.** Untreated offspring of thyroxine-treated male mice (mean ± SE). Birthweight. Two-way avoidance. Hippocampal intra- and infrapyramidal mossy fibers. Note similar changes in $F_2$ incross and $F_2$ outcross and significant difference between $F_3$ from $F_2$-incross and $F_3$ from $F_2$-outcross (changes are significantly greater in $F_3$ males from $F_2$-outcross). See Tables S2 and S3 for statistical details.

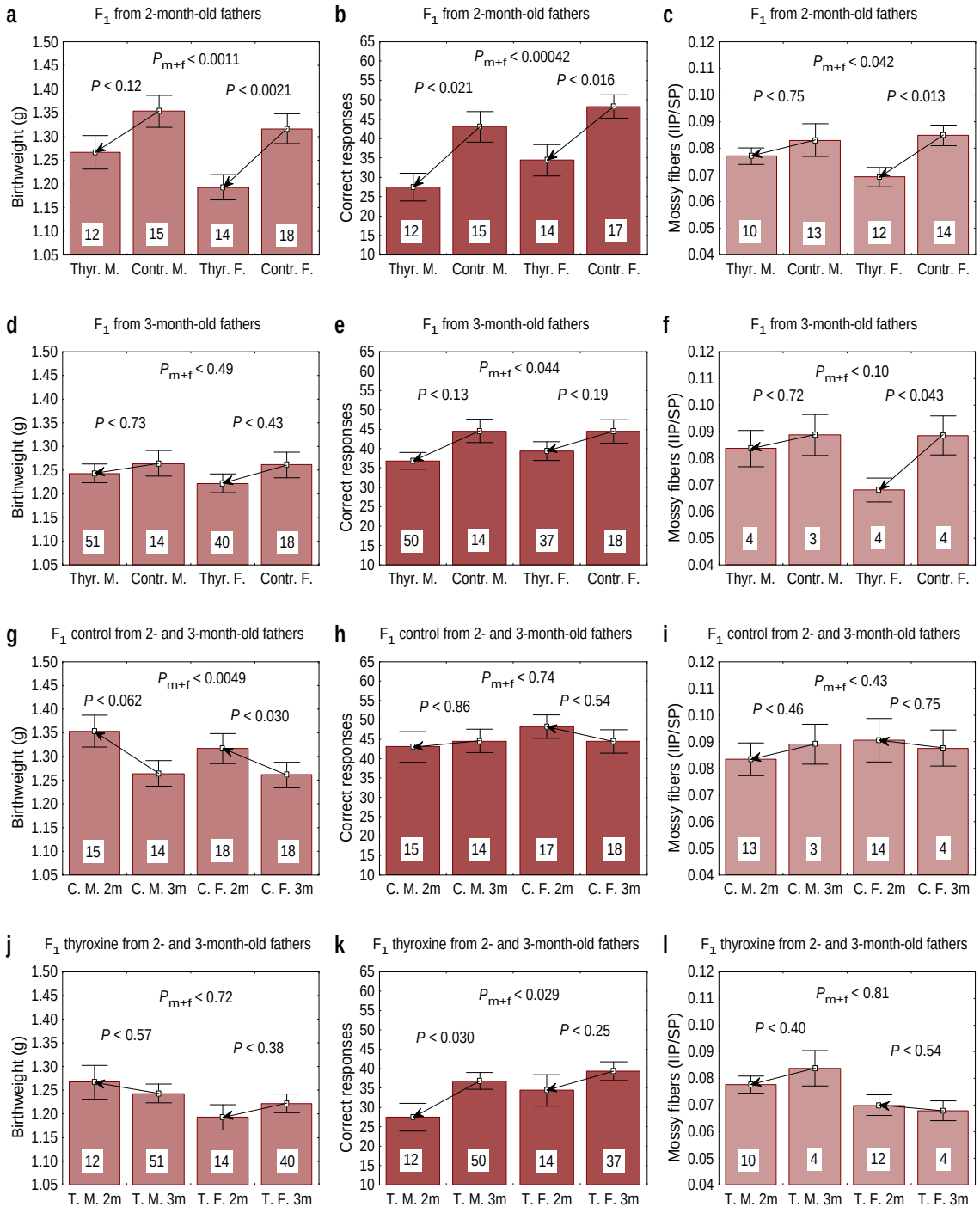



**Figure S6 | The importance of paternal age for thyroxine effects in the $F_1$ progeny.** Asynchronously born and thyroxine-treated, but synchronously bred, 2- and 3-month-old males have produced $F_1$ with 3-month-old females. Decreased birthweight and impaired 2-way avoidance performance were more pronounced in the progeny of 2-month-old males (**a-b**, **d-e**). Hippocampal mossy fibers were decreased similarly (**c**, **f**). Unexpectedly, control from 2-month-old had significantly increased birthweight in comparison with control from 3-month-old (**g**). Mean ± SE. Mann-Whitney U-test.

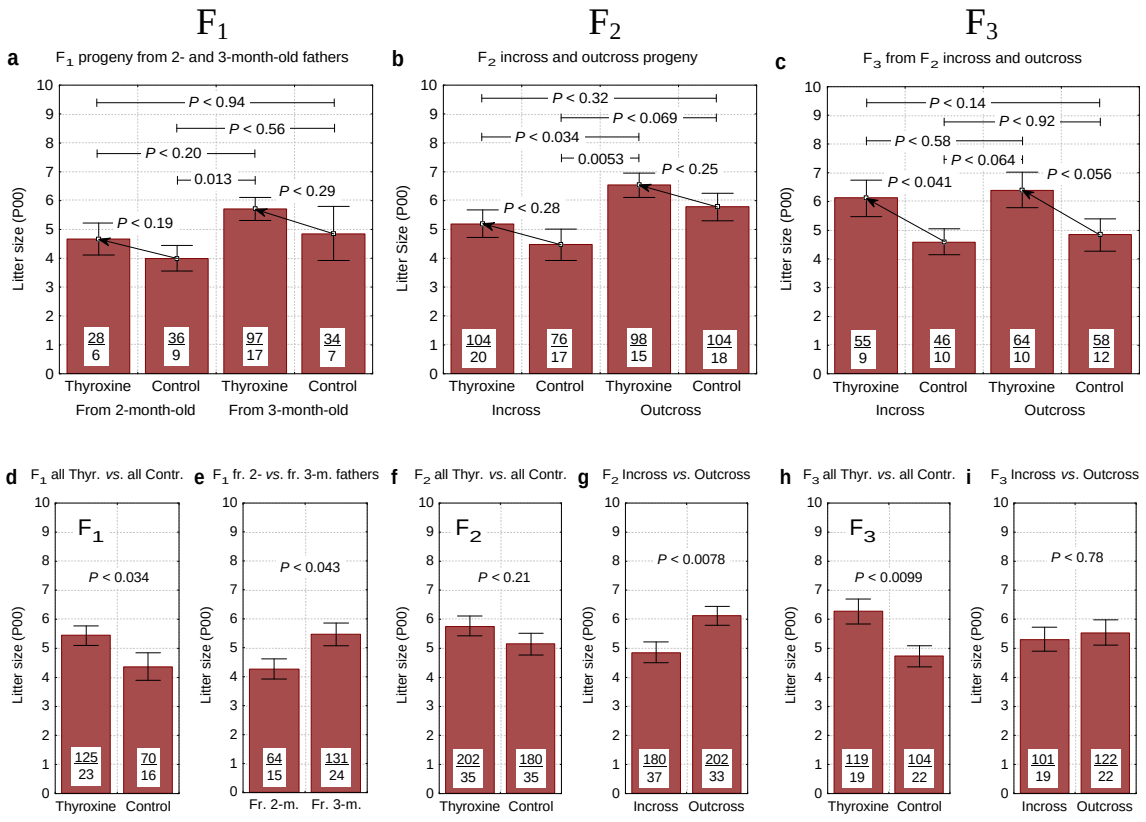



**Figure S7 | Litter size in the $F_1$-$F_3$ progeny of thyroxine-treated males.** $F_1$ was obtained from 2- or 3-month-old males. $F_2$ & $F_3$ are descendants of 2-month-old P males. The number of pups and the number of litters are shown as an indicated division on each bar. Observed birthweight, increased in the progeny of 2-month-old males and decreased in the progeny of thyroxine-treated males, can be explained by the differences in litter size (**a, d-e**). Direct comparison of subgroups does not provide official significance (**a**), but pooling of these subgroups shows both effects of neonatal paternal thyroxine treatment (**d**) as well as of paternal age (**e**). In the thyroxine project the most long-lasting change (in terms of consecutive generations) and the most significant change in the $F_3$ generation in general occurred to be the decreased birthweight (Fig. S4). However this decreased birthweight is not a consequence of some pathological process. It is a consequence of slightly increased litter size, observed in the $F_1$-$F_3$ progeny of thyroxine-treated males. Increased litter size in the progeny of thyroxine-treated males becomes highly significant in the $F_3$ only (**h**), simultaneously in both Incross and Outcross sublines (**c**). May be the increased litter size is a non-specific compensation of an unidentified developmental problem. As we can see from this figure (**g**) and table below, the ages of both parents are important for litter size. Younger females (110-day-old) have bigger litters (see $F_2$-outcross) than older ones (200-day-old; see $F_2$-incross). Mean ± SE. Mann-Whitney U-test.

| | | $F_1$(2m) | $F_1$(3m) | $F_2$(In) | $F_2$(Out) | $F_3$(In) | $F_3$(Out) |
|---|---|---|---|---|---|---|---|
| Parental age at | *Female* | 90 | 90 | 200 | 110 | 180 | 150 |
| breeding (days) | *Male* | 60 | 90 | 200 | 230 | 180 | 150 |

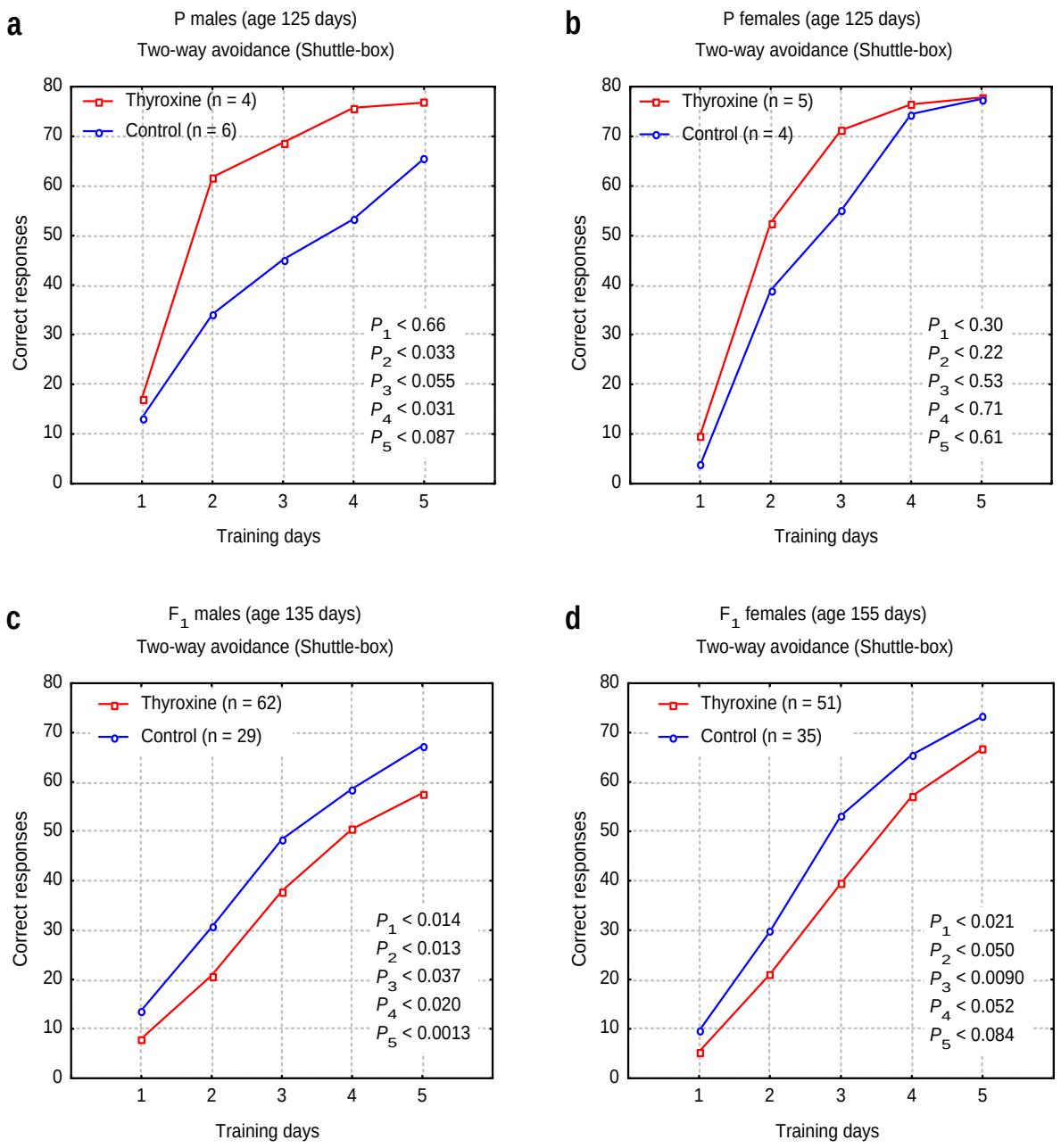



**Figure S8 | Two-way avoidance in thyroxine-treated mice and untreated $F_1$ progeny.** Two-way avoidance correct responses in neonatally L-thyroxine treated DBA/2J males (**a**) and females (**b**) and $F_1$ males (**c**) and females (**d**), obtained from thyroxine-treated males and drug-naïve females. Note that control group performance is identical in P and $F_1$. Simultaneously, thyroxine-treated animals themselves have improved performance (**a-b**), whereas progeny of thyroxine-treated males and drug-naïve females have shown decreased performance during all 5 training days (**c-d**). Mean number of correct responses is shown for each day, Mann-Whitney U-test. SE or SD is omitted for clarity. For these animals Mean ± SE is shown in the Fig. S22b-c. Each animal was exposed to 80 training trials daily and 80 correct responses daily is theoretical maximum.

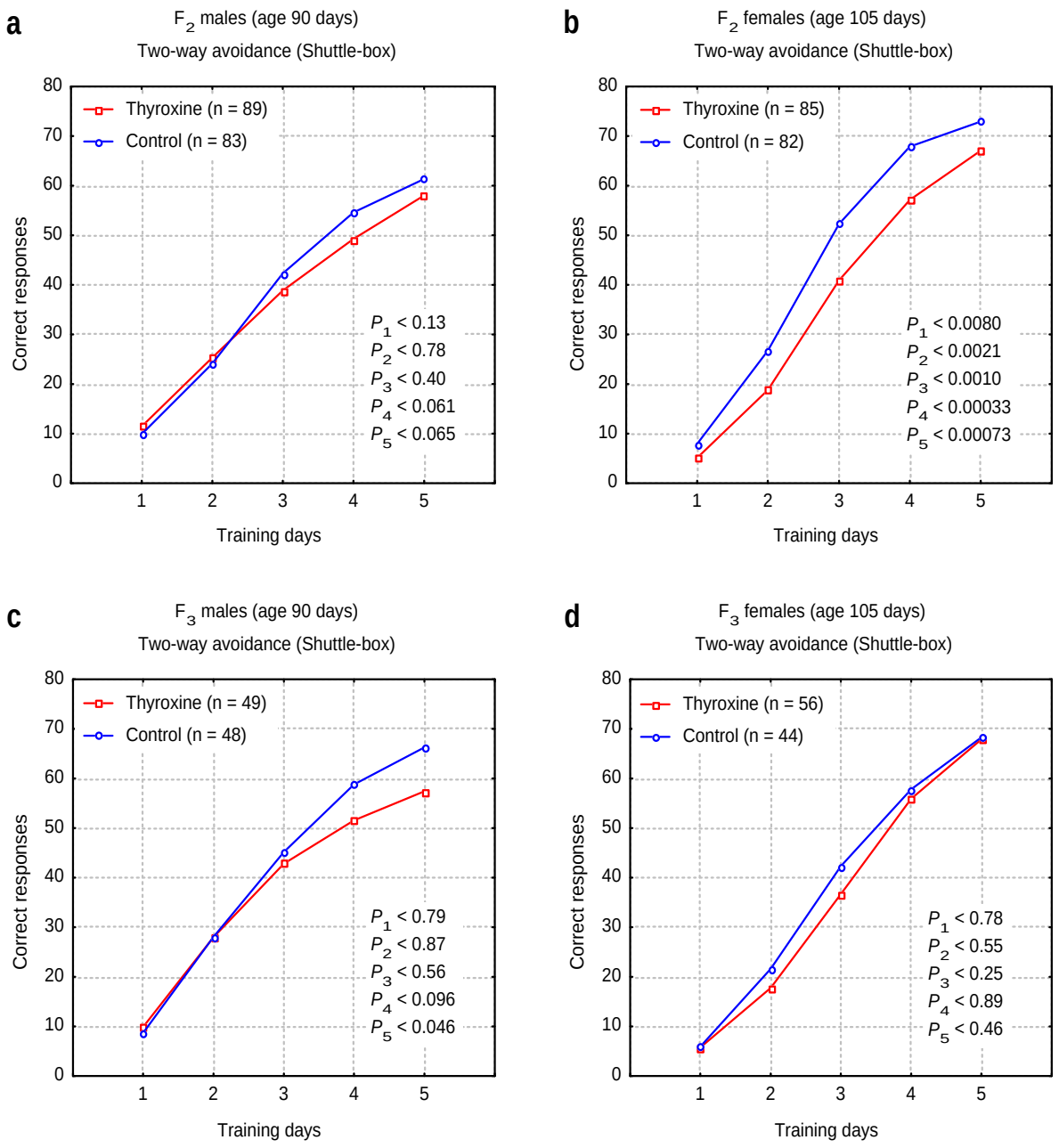



**Figure S9 | Two-way avoidance in the $F_2$ and $F_3$ progeny of thyroxine-treated males.** Two-way avoidance correct responses in the $F_2$ (**a-b**) and $F_3$ (**c-d**) descendants of neonatally L-thyroxine-treated males and drug-naïve females. **a, c**, Males. **b, d**, Females. Incross and outcross animals are shown together here. Note slight trend towards impairment of late training stages in the $F_2$ males (**a**) and highly significant impairment of all training stages in the $F_2$ females (**b**). In the $F_3$ generation the late training stages occurred to be impaired significantly in males (**c**), but differences in females have disappeared completely (**d**). $F_3$ male phenotype (**c**) looks confusing here, but it can be explained after separate analysis of incross and outcross subgroups (see Fig. S11a,c).

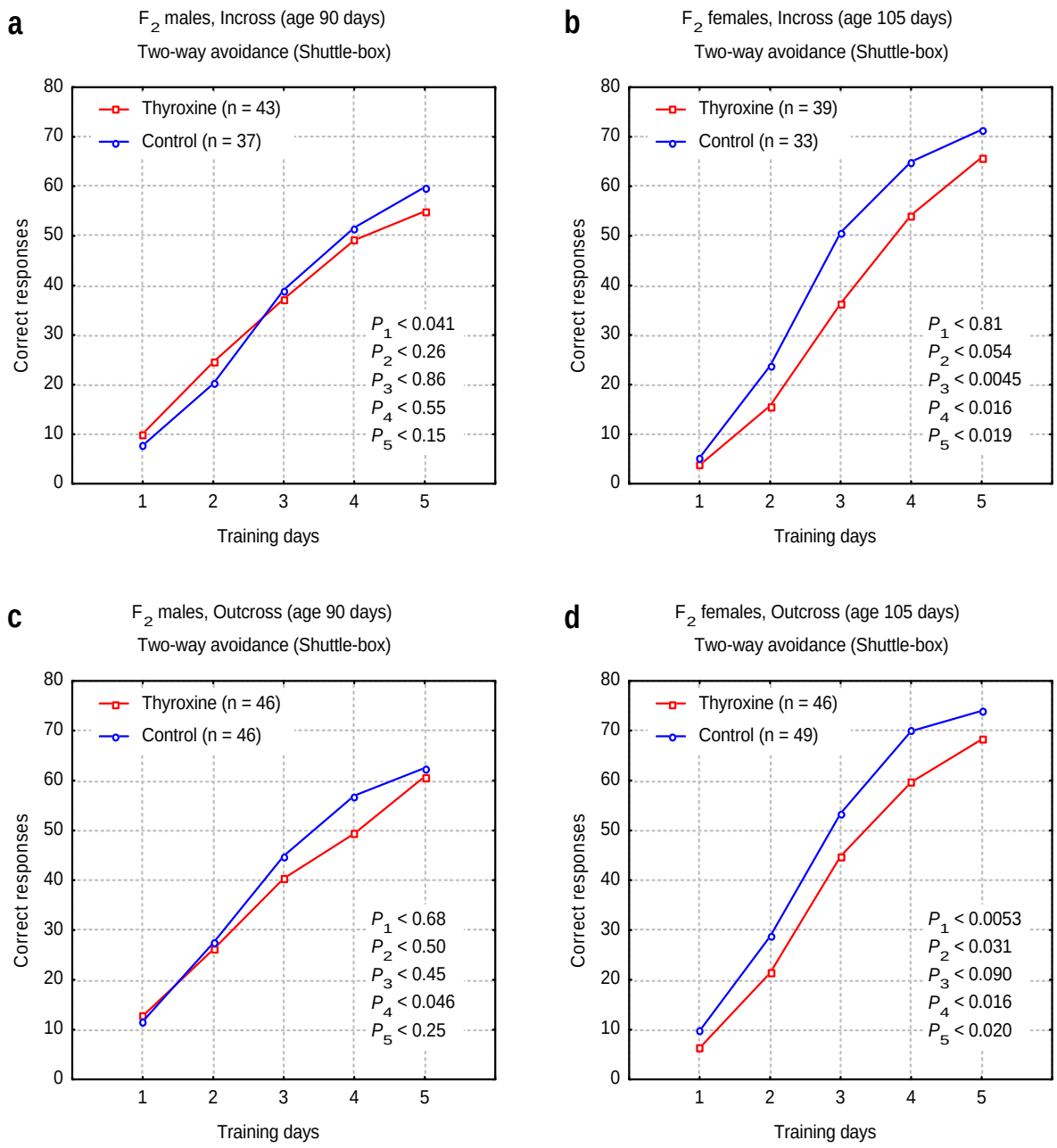



**Figure S10 | Two-way avoidance in $F_2$ incross and outcross mice.** Two-way avoidance correct responses in the $F_2$ incross males (**a**) and females (b) and $F_2$ outcross males (**c**) and females (**d**), descendants of neonatally L-thyroxine-treated males and drug-naïve females. Note that there is not much difference in $F_2$ males, in incross (**a**) and outcross (**c**), just some slight improvement at the 1st training day in incross (**a**), and slight impairment at the 4th training day in outcross (**c**). However in $F_2$ females the impairment of performance is highly significant in both incross (**b**) and outcross (**d**), approximately to the same extent.

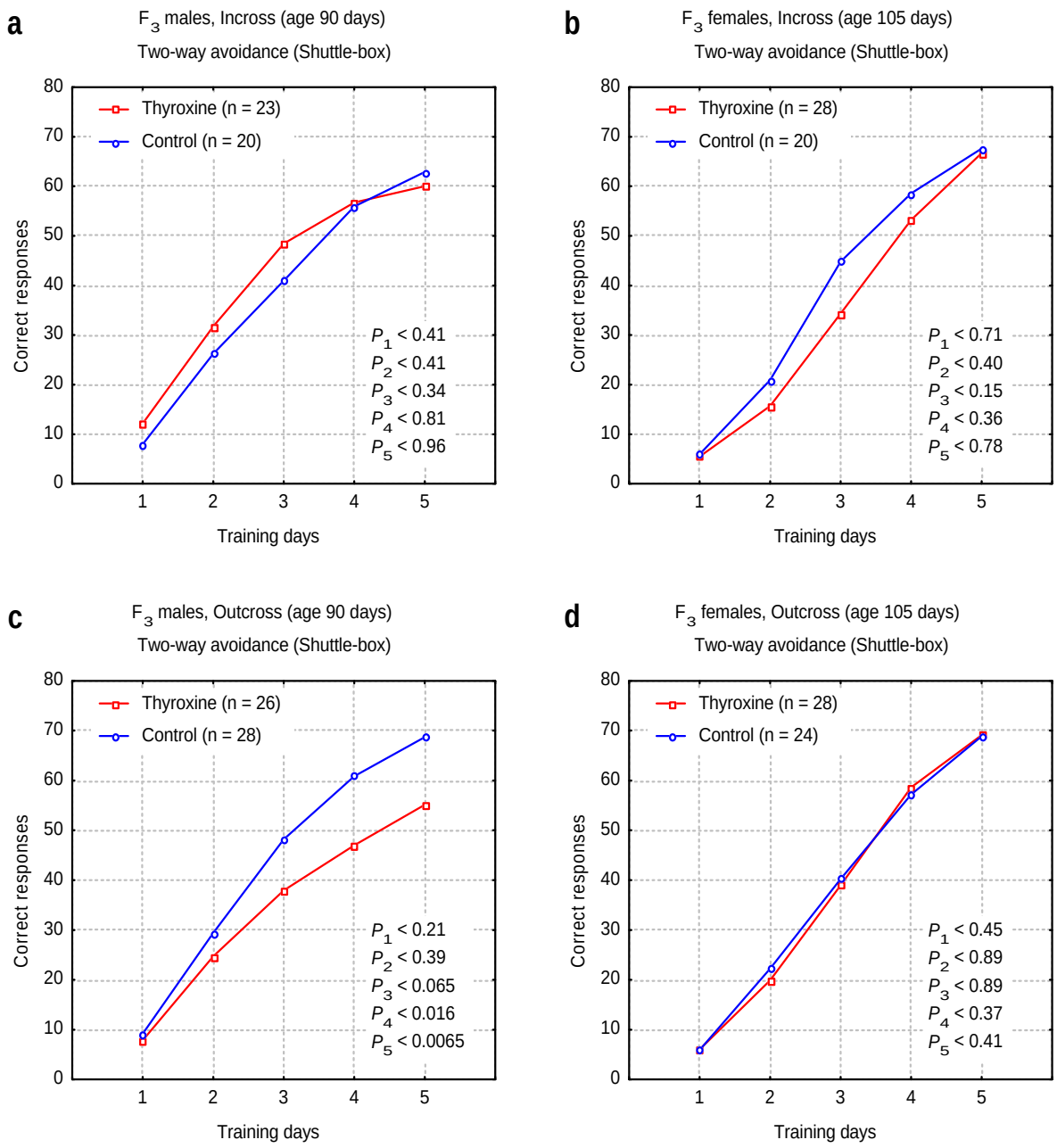



**Figure S11 | Two-way avoidance in $F_3$ incross and outcross mice.** "$F_3$-incross" was obtained from incross breeding of $F_2$-incross females with $F_2$-incross males, but our "$F_3$-outcross" is not really "outcross", because it was obtained from incross breeding of $F_2$-outcross females with $F_2$-outcross males and, thus, $F_3$-outcross mice are just descendants of $F_2$-outcross parents. Note that $F_3$-outcross males have significantly impaired late learning stages (**c**). However $F_3$-incross males are even more interesting: they have the same shape of learning curve, but this curve is shifted up as a whole (**a**), due to some compensation which is not a reversion to wild type.

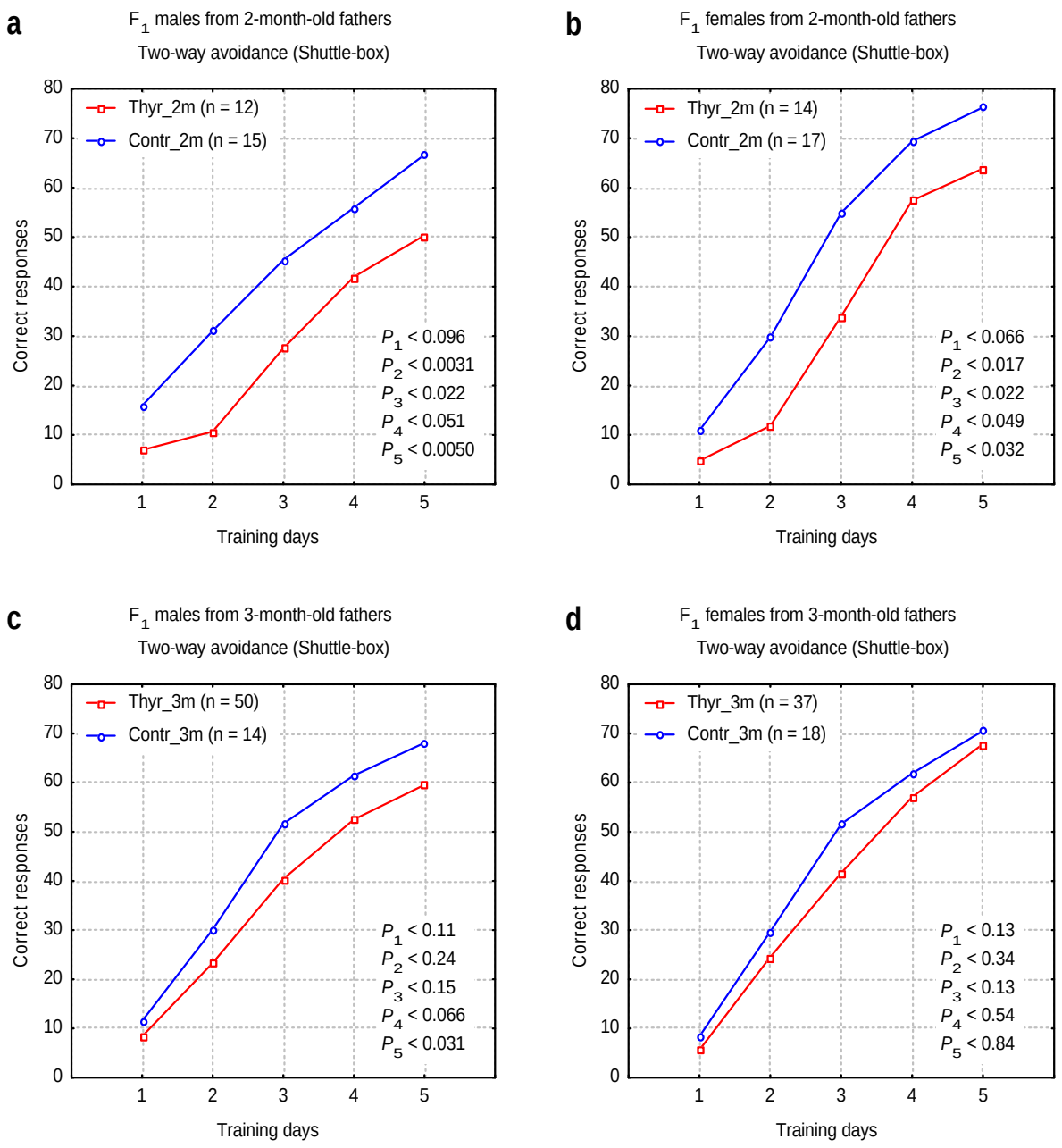



**Figure S12 | Two-way avoidance in $F_1$ mice from 2- and 3-month-old thyroxine-treated males.** All $F_1$ descendants shown here were born and tested synchronously. The single difference is that male parents of $F_1$ in (**c-d**) were born (and experimental group was thyroxine-treated) one month earlier that in (**a-b**). Note highly significant effect of thyroxine in the progeny of 2-month-old fathers (**a-b**) and only slight tendency in the progeny of 3-month-old ones (**c-d**).

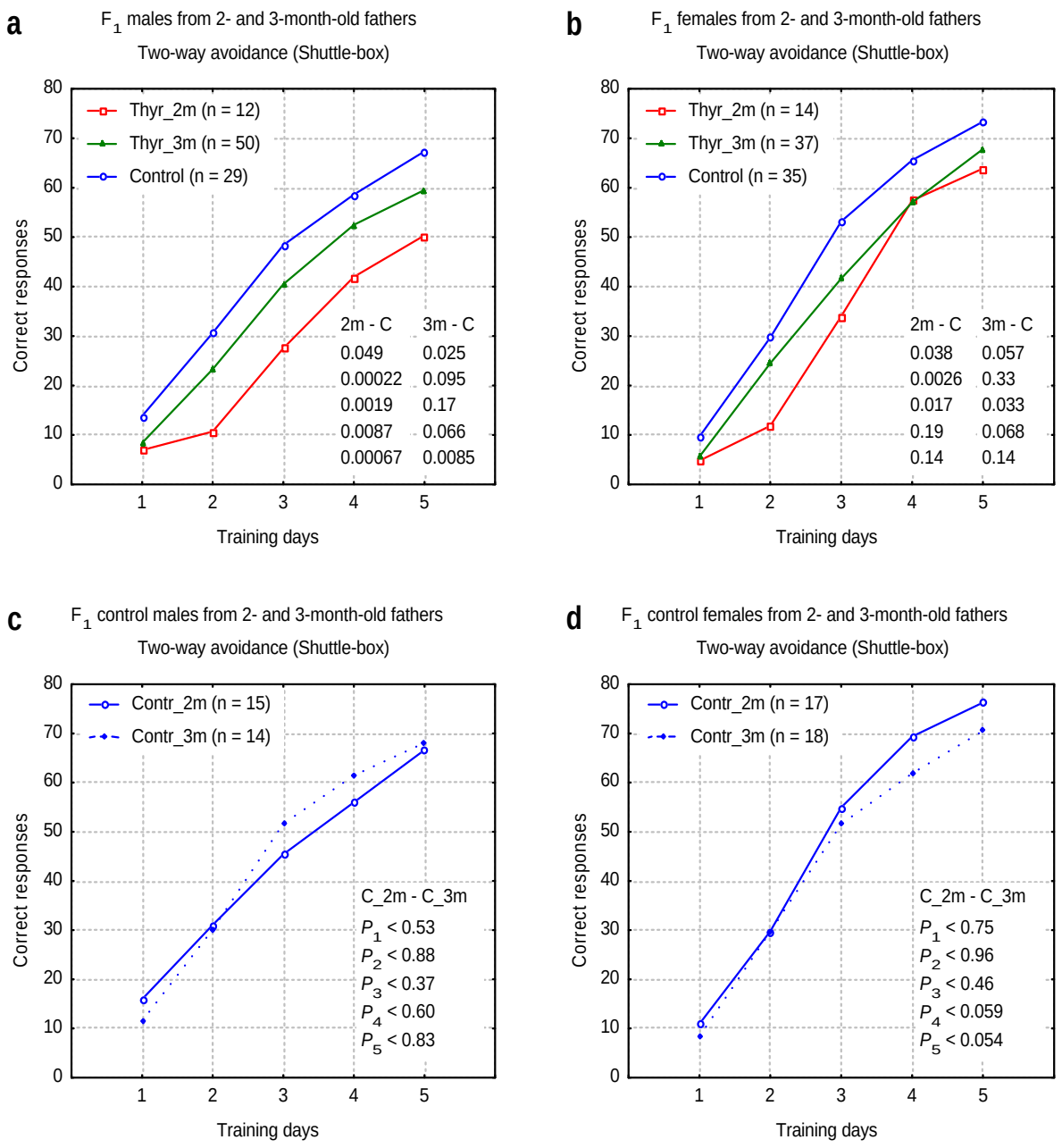



**Figure S13 | Two-way avoidance in $F_1$ mice relatively to united control from 2- and 3-month-old.** $F_1$ progeny, obtained from 2- and 3-month-old neonatally thyroxine-treated male parents, in comparison with united control group, consisting of descendants of control 2-month-old males and control 3-month-old males (**a-b**). Note that all thyroxine effects are about 3-4 times more significant in the progeny of 2-month-old males. Simultaneously, control groups from 2- and 3-month-old males do not differ from each other in this task (**c-d**).

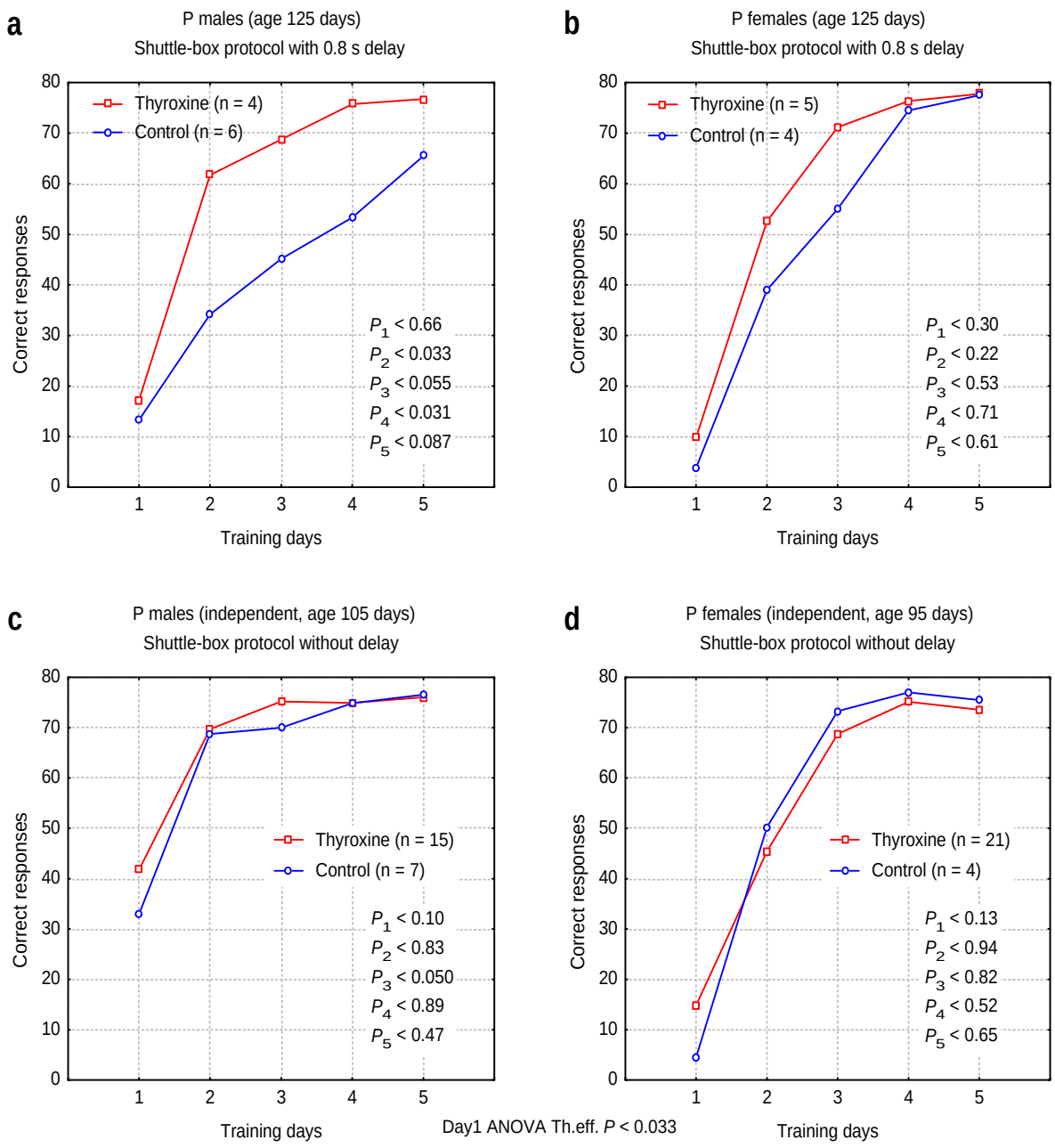



**Figure S14 | Two-way avoidance in neonatally thyroxine-treated mice (different protocols).** In all our experiments we used rather unusual protocol for 2-way avoidance task with 0.8 s delay of light and light + current cut off (**a-b**). We have chosen it on the basis of pilot experiment with different cut off latencies (Fig. S22a). However before this some part of thyroxine-treated animals, namely those were bred at the age of 3 months (males only were bred), were tested in usual 2-way avoidance task without any delay (**c-d**). Enhanced performance of thyroxine-treated males and females was detectable only during the first training day (**c-d**), $P < 0.033$ (ANOVA, thyroxine effect). Then, practical maximum was achieved by all males at day 2 (**c**), and by all females – at day 3 (**d**). Obviously, good protocol optimization is very-very important here.

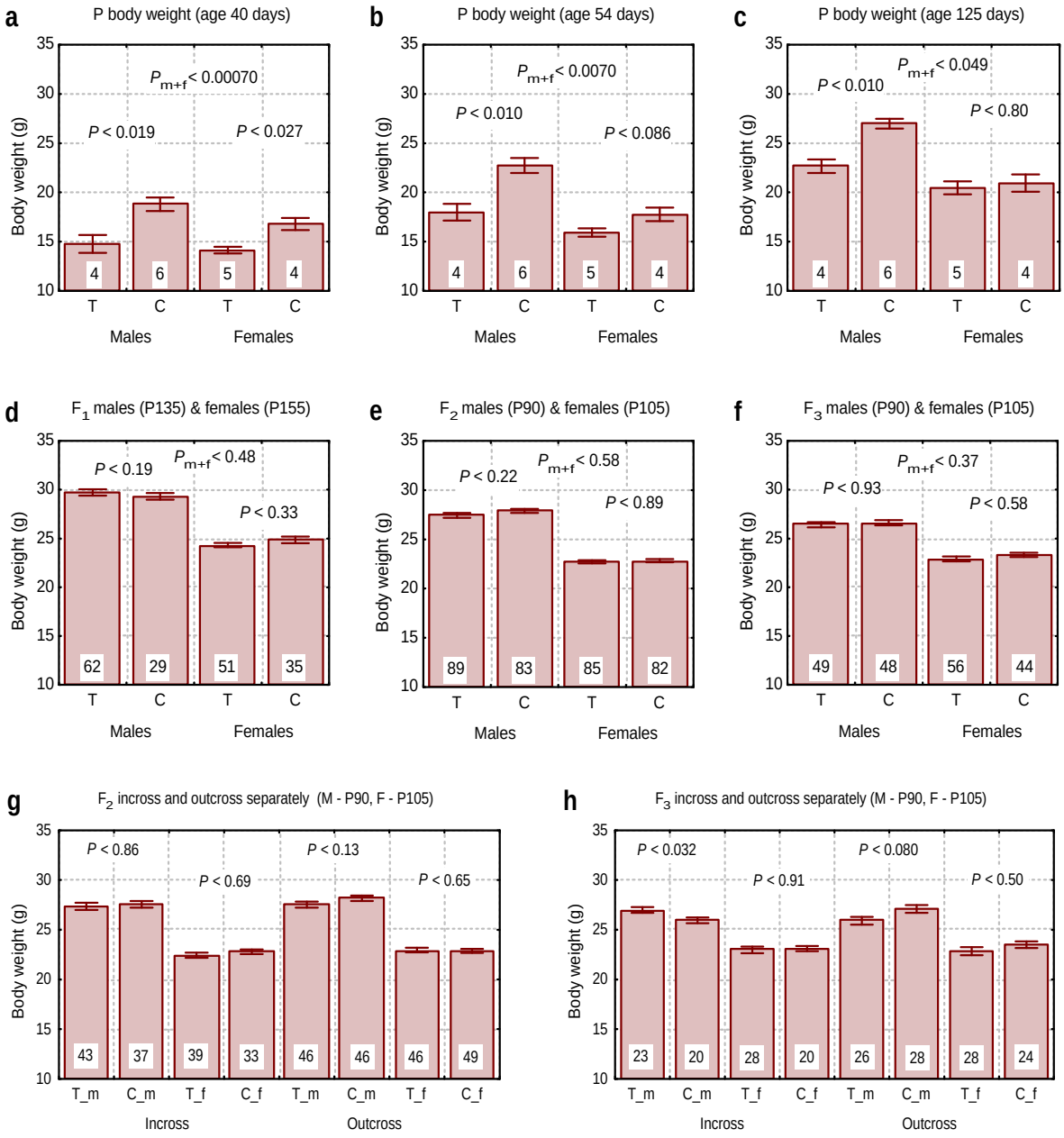



**Figure S15 | Body weight before two-way avoidance test in P and $F_1$-$F_3$ mice.** Mean ± SE. Body weight of neonatally L-thyroxine-treated male and female P mice at the age of 40 days (**a**) and 54 days (**b**) and just before behavioural testing in two-way avoidance task (**c**). Body weight of $F_1$-$F_3$ descendants of neonatally L-thyroxine-treated males and drug-naïve females measured just before behavioural testing in two-way avoidance task (**d-h**). Note that body weight of P generation was significantly decreased (**a-c**) and theoretically it can induce improved 2-way avoidance performance. However in the $F_1$ and $F_2$ generations body weight was absolutely normal (**d-e**) and, thus, it can not explain decreased 2-way avoidance performance here. Interesting that $F_3$-incross and $F_3$-outcross males (**h**) have body weight changes in the opposite direction ($P < 0.032$ and $P < 0.080$, respectively), which can be linked to unusual compensation in the $F_3$-incross males observed in 2-way avoidance task.

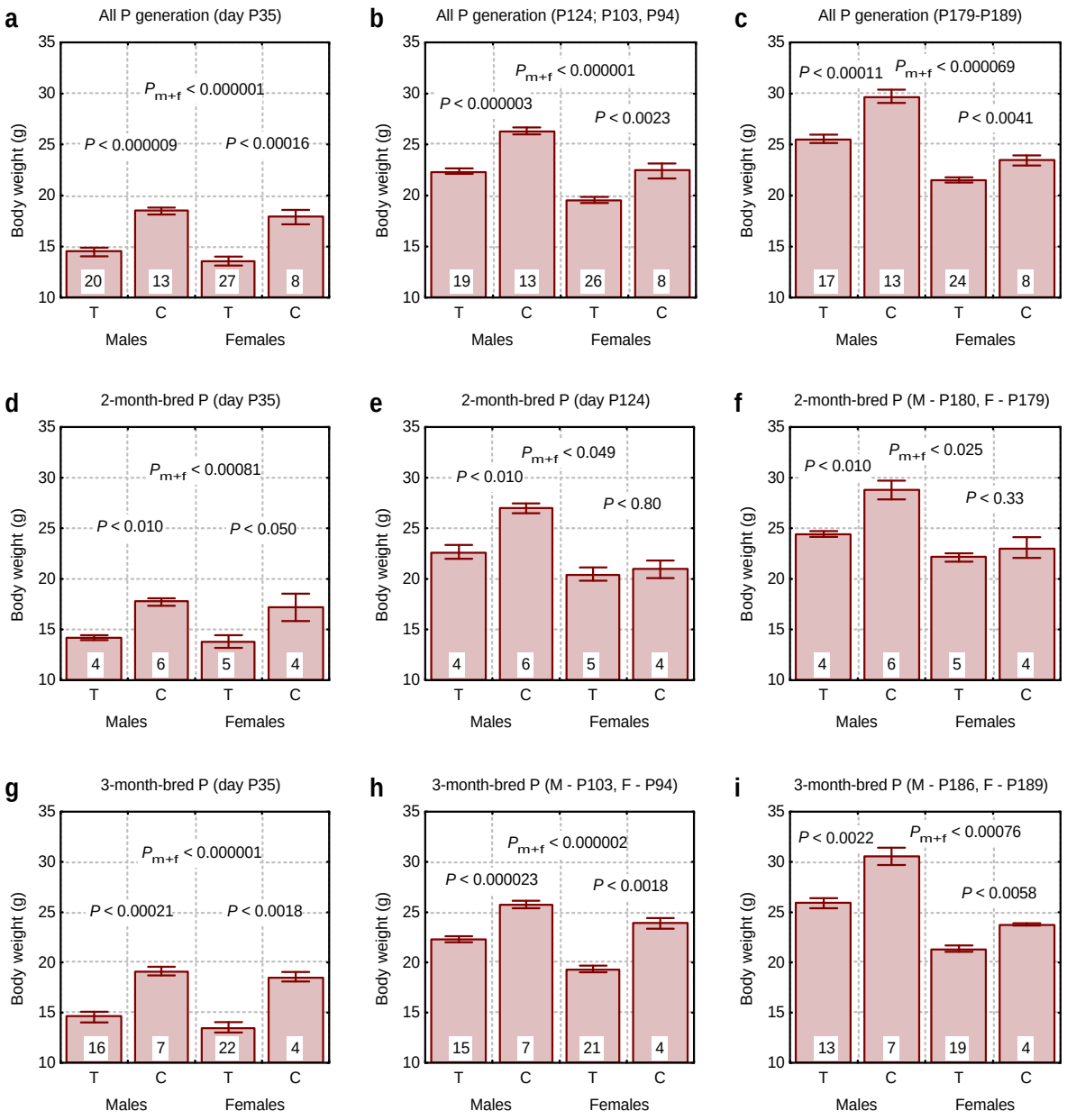



**Figure S16 | Body weight in thyroxine-treated P mice bred at the age of 2 or 3 months.** Body weight at the age of 35 days (**a**, **d**, **g**), at the age of 94-124 days just before 2-way avoidance test (**b**, **e**, **h**), and at the age 179-189 days just around Morris water maze test (**c**, **f**, **i**). **a-c**, All P generation male and female mice. **d-f**, Male mice, those were bred at the age of 2 months, and their female litter-mates. **g-i**, Male mice, those were bred at the age of 3 months, and their female litter-mates. All P females were not used for breeding. Note significantly decreased body weight in all thyroxine-treated animals at all time points (35 days, 3 and 6 months). Note also that there is no difference between 2-month-bred and 3-month-bred males in body weight. It means that their L-thyroxine treatment procedures were approximately similar. Mean ± SE. Mann-Whitney U-test.

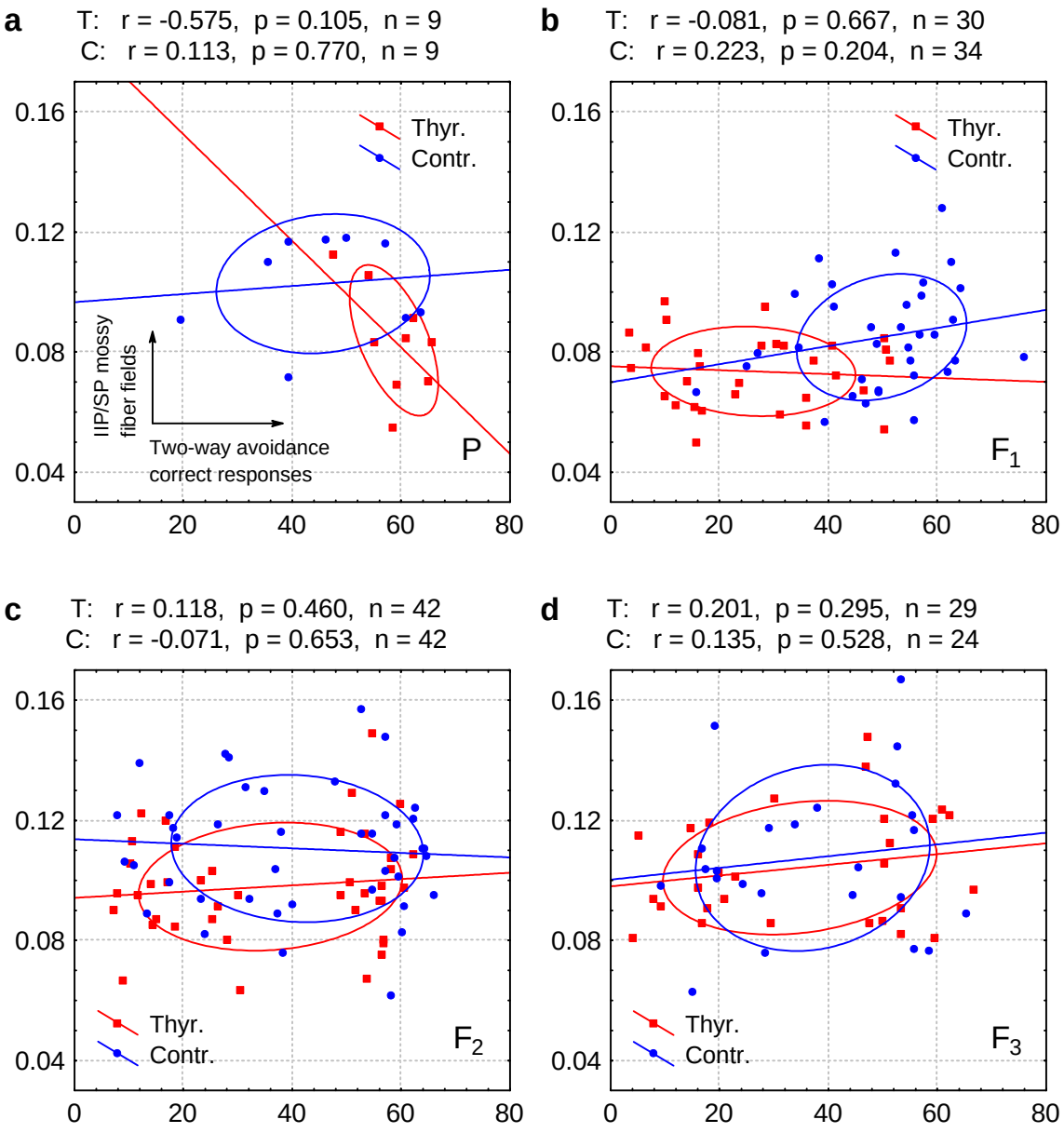



**Figure S17 | Correlations in the thyroxine-treated mice and their $F_1$-$F_3$ progeny.** Correlation between two-way avoidance performance (number of correct responses, averaged value of 5 training days, 80 trials daily) and ratio of intra- and infrapyramidal to suprapyramidal hippocampal mossy fiber fields (IIP/SP mossy fibers). **a**, P generation. **b**, $F_1$. **c**, $F_2$. **d**, $F_3$. Males and females are pooled here. Red – experimental group. Blue – control one. Note the existence of relatively high negative correlation in the thyroxine-treated animals only (**a**, red). In the progeny of thyroxine-treated males (**b-d**) any correlation between particular traits is absolutely absent, in spite of significantly changed mean values (see significance levels in the Fig. S4). Separate analysis of males and females (Fig. S18) and separate analysis of incross and outcross subgroups in the $F_2$ and $F_3$ generations (Fig. S19) revealed the absence of individual correlations between above-mentioned traits in all experimental groups.

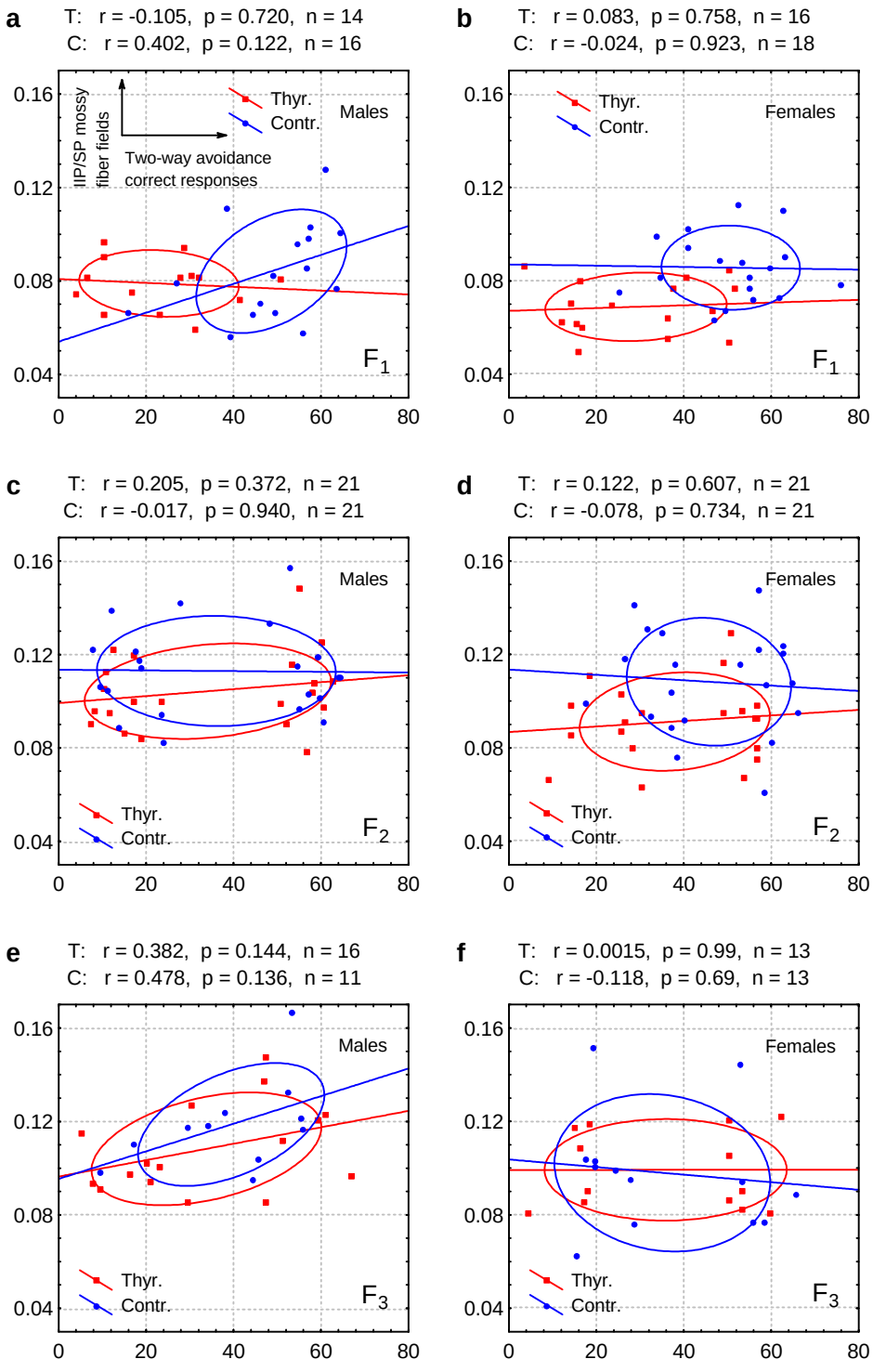



**Figure S18 | Correlations in the $F_1$-$F_3$ male and female descendants of thyroxine-treated males.** Correlation between two-way avoidance performance (number of correct responses, averaged value of 5 training days, 80 trials daily) and ratio of intra- and infrapyramidal to suprapyramidal hippocampal mossy fiber fields (IIP/SP mossy fibers). **a-b**, $F_1$ generation. **c-d**, $F_2$ generation, **e-f**, $F_3$ generation. **a**, **c**, **e**, Males. **b**, **d**, **f**, Females.

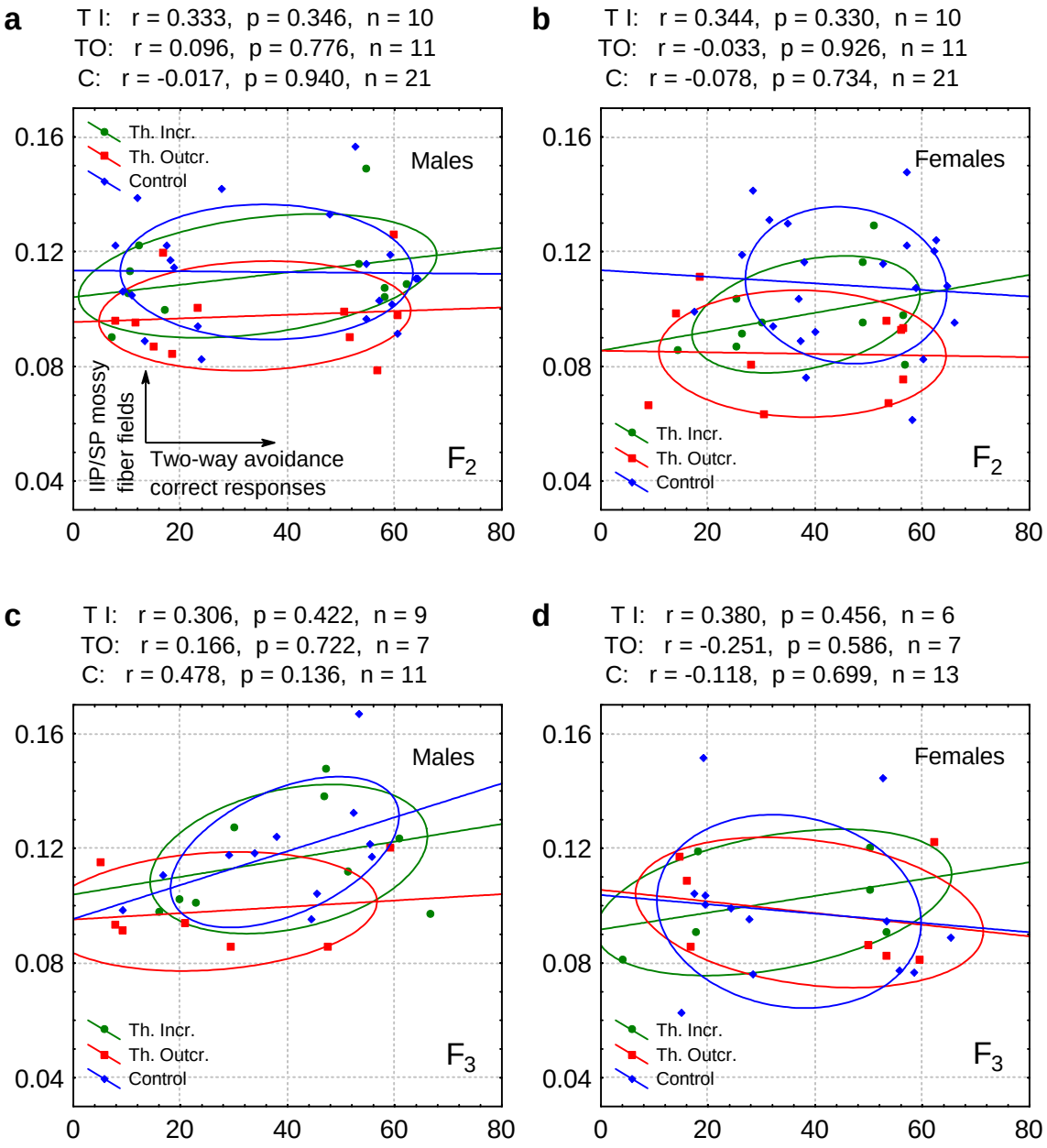



**Figure S19 | Correlations in the $F_2$-$F_3$ incross and outcross descendants of thyroxine-treated males.** Correlation between two-way avoidance performance (number of correct responses, averaged value of 5 training days, 80 trials daily) and ratio of intra- and infrapyramidal to suprapyramidal hippocampal mossy fiber fields (IIP/SP mossy fibers). **a-b**, $F_2$ generation. **c-d**, $F_3$ generation. **a**, **c**, Males. **b**, **d**, Females. Green – descendants of thyroxine-treated males, obtained after incross breeding. Red – descendants of thyroxine-treated males, obtained after outcross breeding. Blue – descendants of control males, incross and outcross control groups are pooled. Note that the outcross differs from the control group to the greater extent than incross does, especially in the $F_2$ females (**b**) and $F_3$ males (**c**).

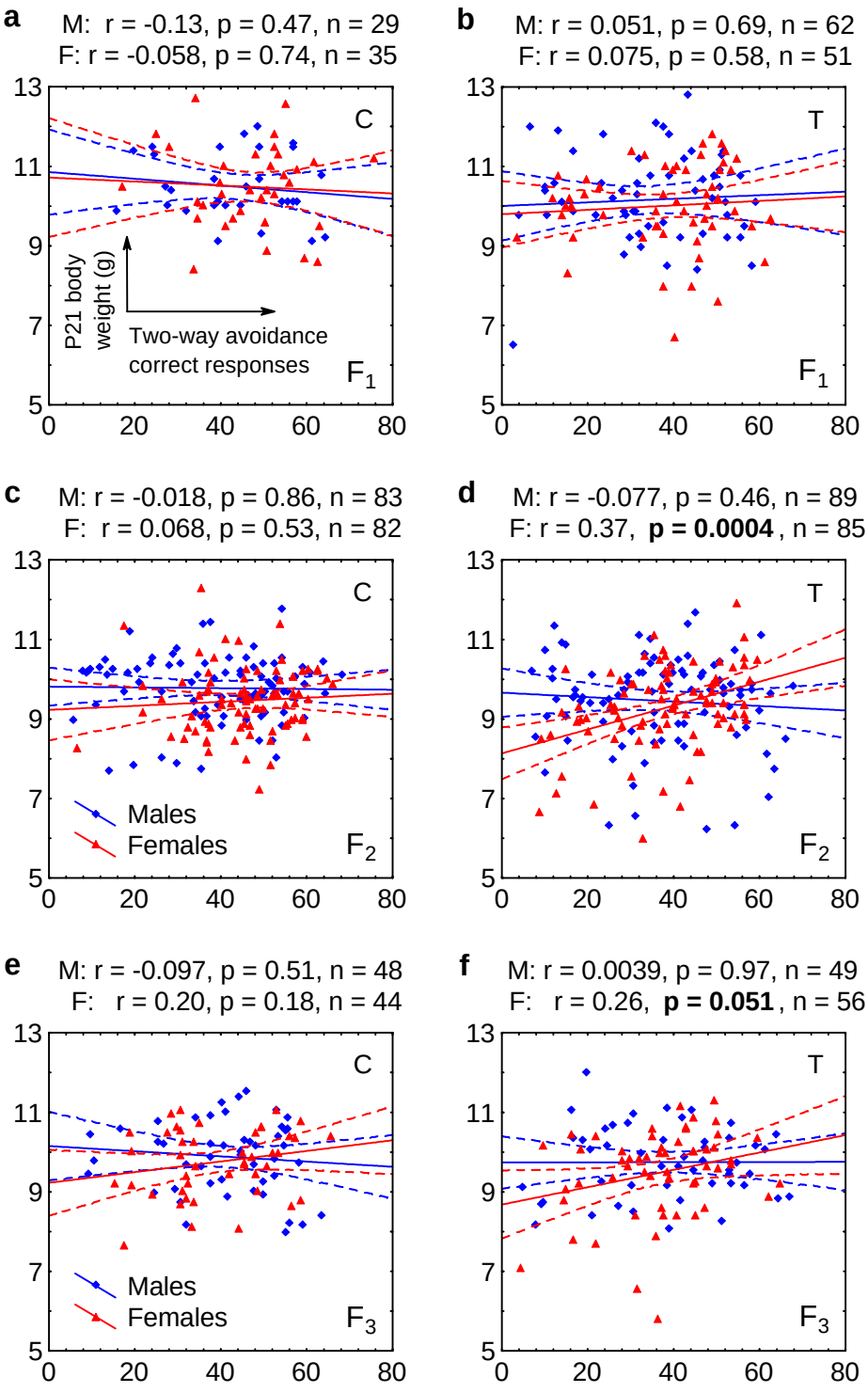



**Figure S20 | Gender-dependent correlations in the $F_1$-$F_3$ offspring of thyroxine-treated males.** Correlation of adult two-way avoidance performance and body weight at the age of 21 day (P21) in the $F_1$-$F_3$ offspring of thyroxine-treated male mice. $F_1$ (**a**, **b**), $F_2$ (**c**, **d**) and $F_3$ (**e**, **f**) descendants of thyroxine-treated males and drug-naïve females. **a**, **c**, **e**, Control. **b**, **d**, **f**, Experimental. Blue – males, red – females. Note significant positive correlation in the experimental $F_2$-$F_3$ females only (**d**, **f**).

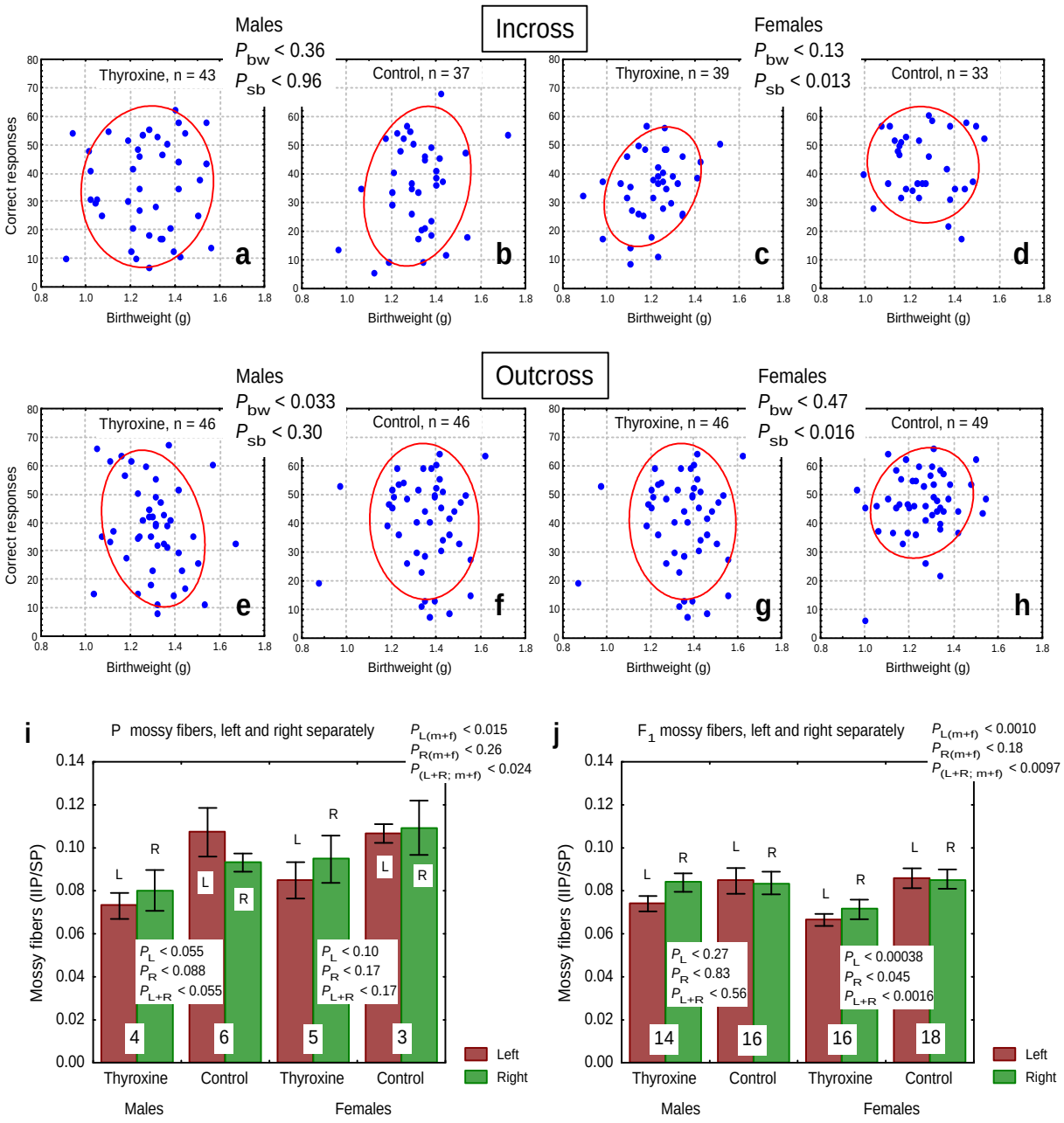



**Figure S21 | Correlation "birthweight – 2-way avoidance" in $F_2$ mice and P-$F_1$ MF asymmetry. a-h**, Correlation between birthweight and adult shuttle-box performance in the $F_2$-incross (**a-d**) and $F_2$-outcross (**e-h**) males (**a-b, e-f**) and females (**c-d, g-h**). $P_{bw}$ and $P_{sb}$ are significance levels in accordance with Mann-Whitney U-test for birthweight and two-way avoidance ("shuttle-box"), respectively. Note that decreased two-way avoidance performance of $F_2$ incross and outcross females (**c-d, g-h**) is not associated with any difference in their birthweight. **i-j**, Hippocampal mossy fiber projections, ratio of intra- and infrapyramidal mossy fibers to suprapyramidal mossy fibers in the left and right hemispheres in P and $F_1$ mice. Thyroxine effects are more pronounced in the left hemisphere, at least in P and $F_1$ generations. Similar asymmetry was not observed in the $F_2$ and $F_3$ (data not shown). Mean ± SE. Mann-Whitney U-test.

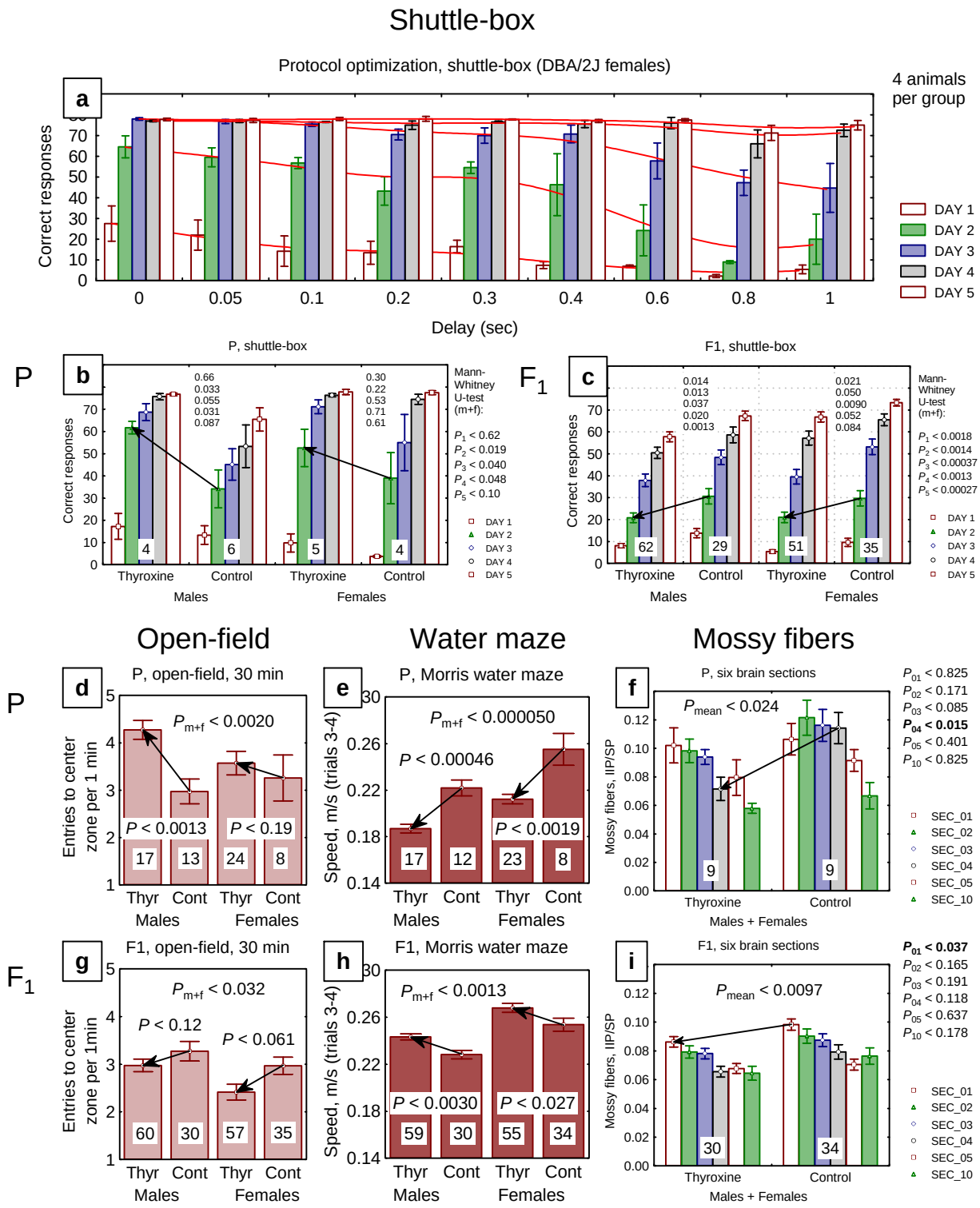



**Figure S22 | Comparison of thyroxine-treated mice and $F_1$ descendants. a**, Two-way avoidance protocol optimization using naïve DBA/2J females (mean ± SE), different delays of light-current termination. Delay 0.8 sec was chosen. **b-c**, Two-way avoidance, P and $F_1$. Note identical learning curves in control groups and opposite changes in experimental parents and progeny. **d**, **g**, **e**, **h**, Open field and water maze. Note similarity in control groups and opposite changes in the experimental ones. **f**, **i**, Hippocampal mossy fibers. Note similar changes in the experimental groups.

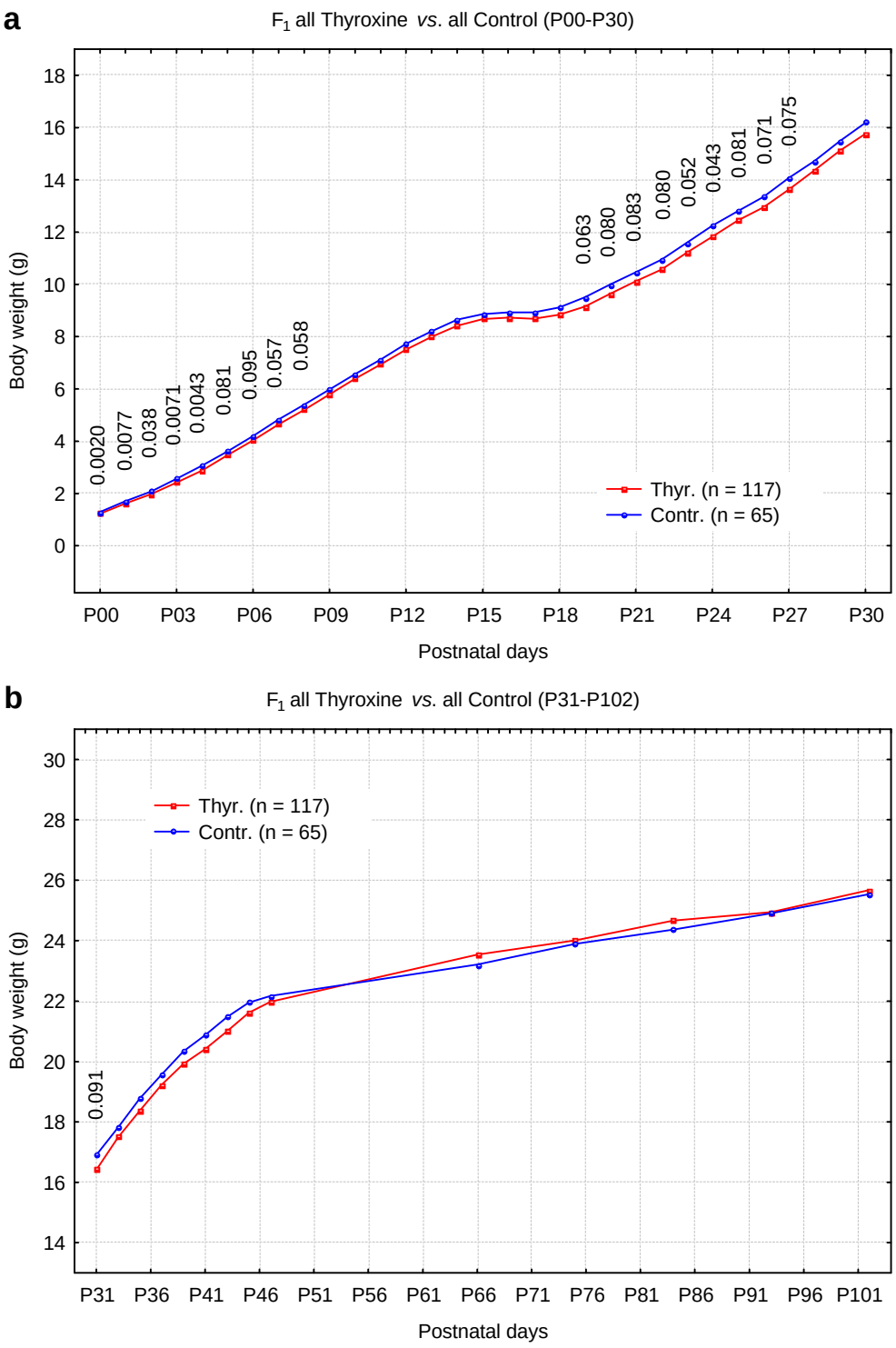



**Figure S23 | Body weight in the $F_1$ offspring of neonatally thyroxine-treated males.** Males and females are pooled here. Descendants of 2- and 3-month-old males are pooled also. Note significantly decreased body weight in the experimental animals during postnatal days P00-P04. Remember that at P00 all pups were placed under NMRI foster mothers (each NMRI foster mother took about 4 experimental and 4 control pups).

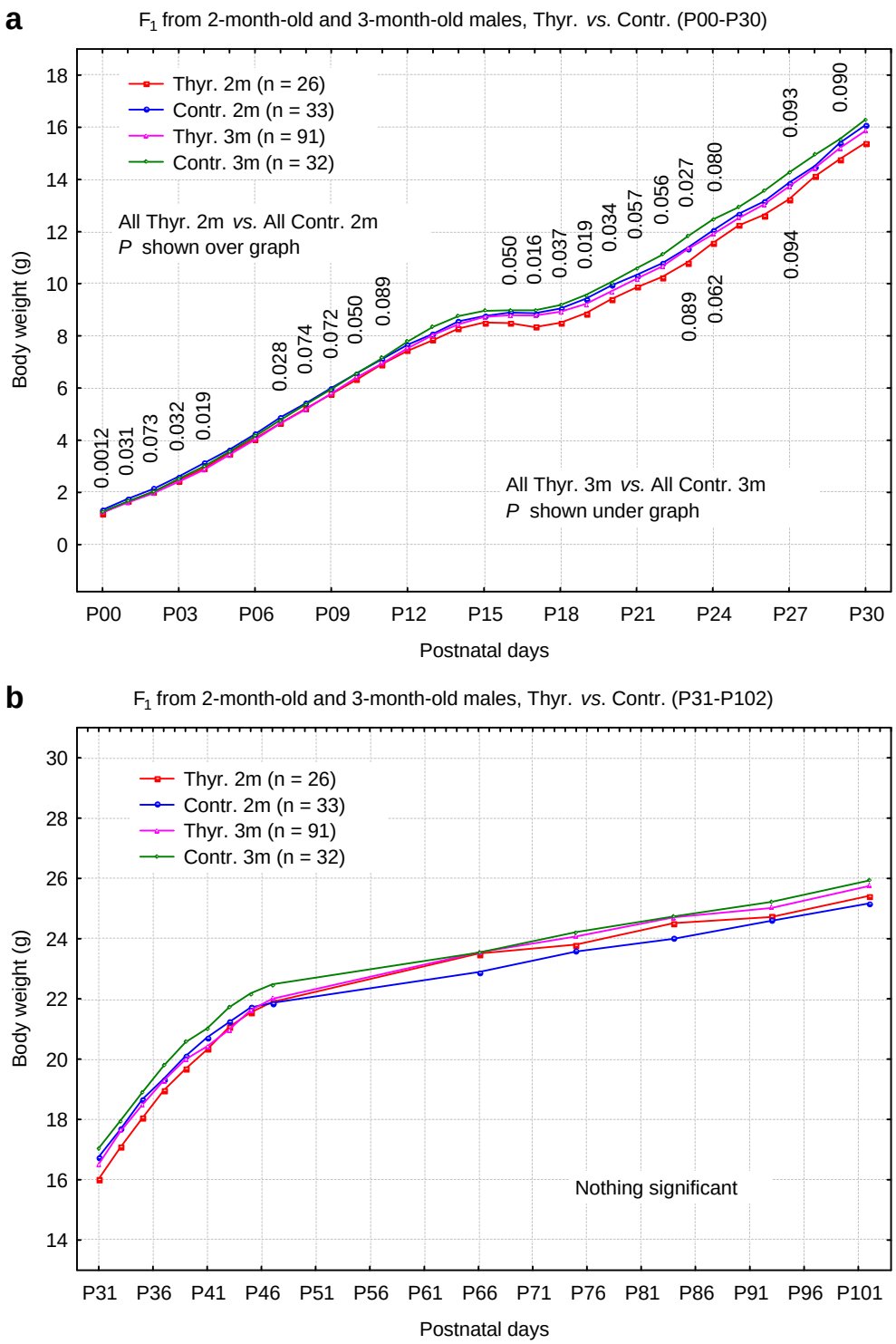



**Figure S24 | Body weight in the $F_1$ offspring of 2- and 3-month-old thyroxine-treated males.** Descendants of 2- and 3-month-old males are shown separately here. Note that all effects of paternal neonatal thyroxine treatment are greater in the progeny of 2-month-old males than of 3-month-old ones. Mann-Whitney U-test. Significance levels with $P < 0.1$ are shown. $P < 0.05$ is considered significant.

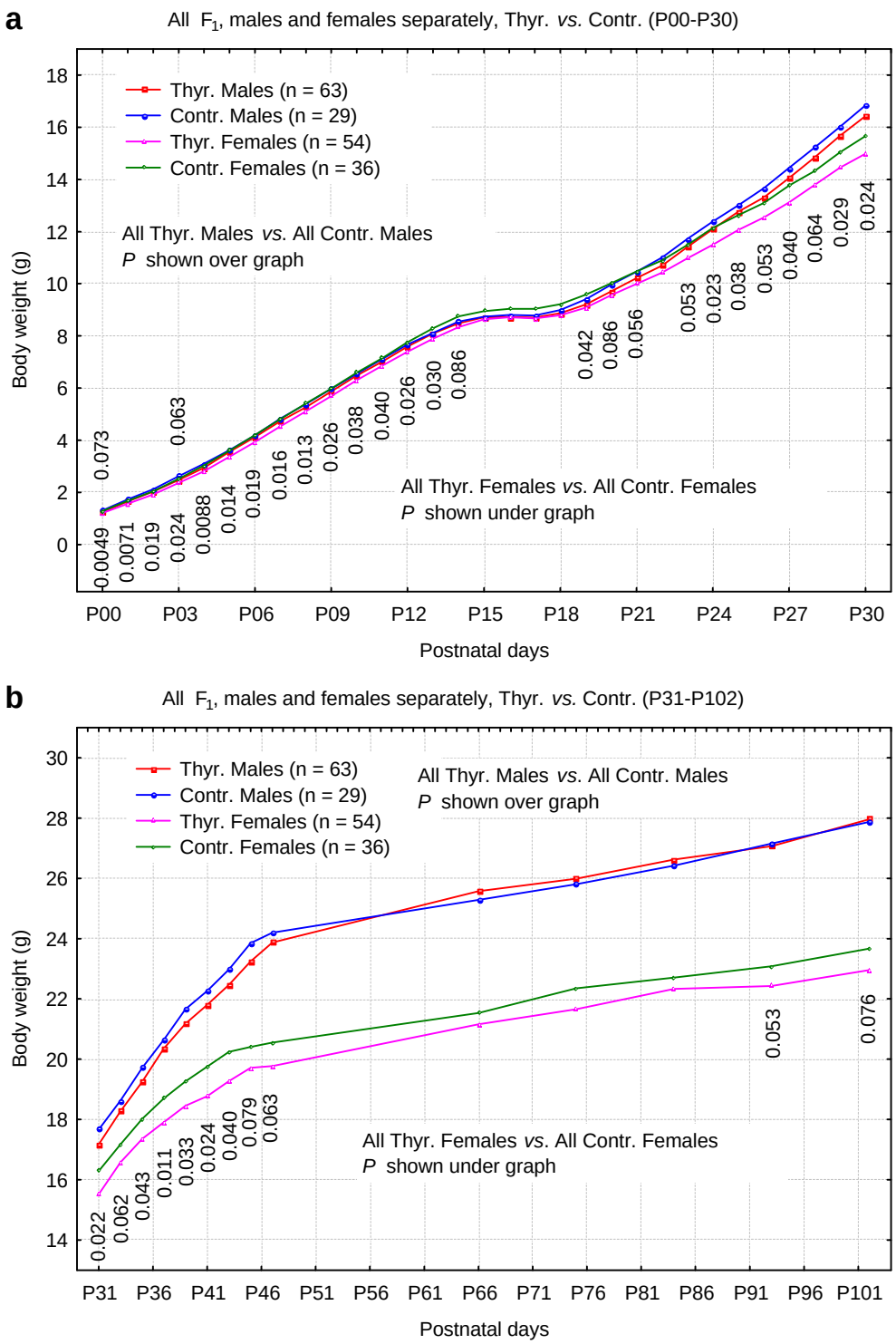



**Figure S25 | Body weight in male and female $F_1$ offspring of neonatally thyroxine-treated males.** Descendants of 2- and 3-month-old males are pooled here. Males and females are shown separately. Note that all body-weight-related thyroxine effects are achieved due to females.

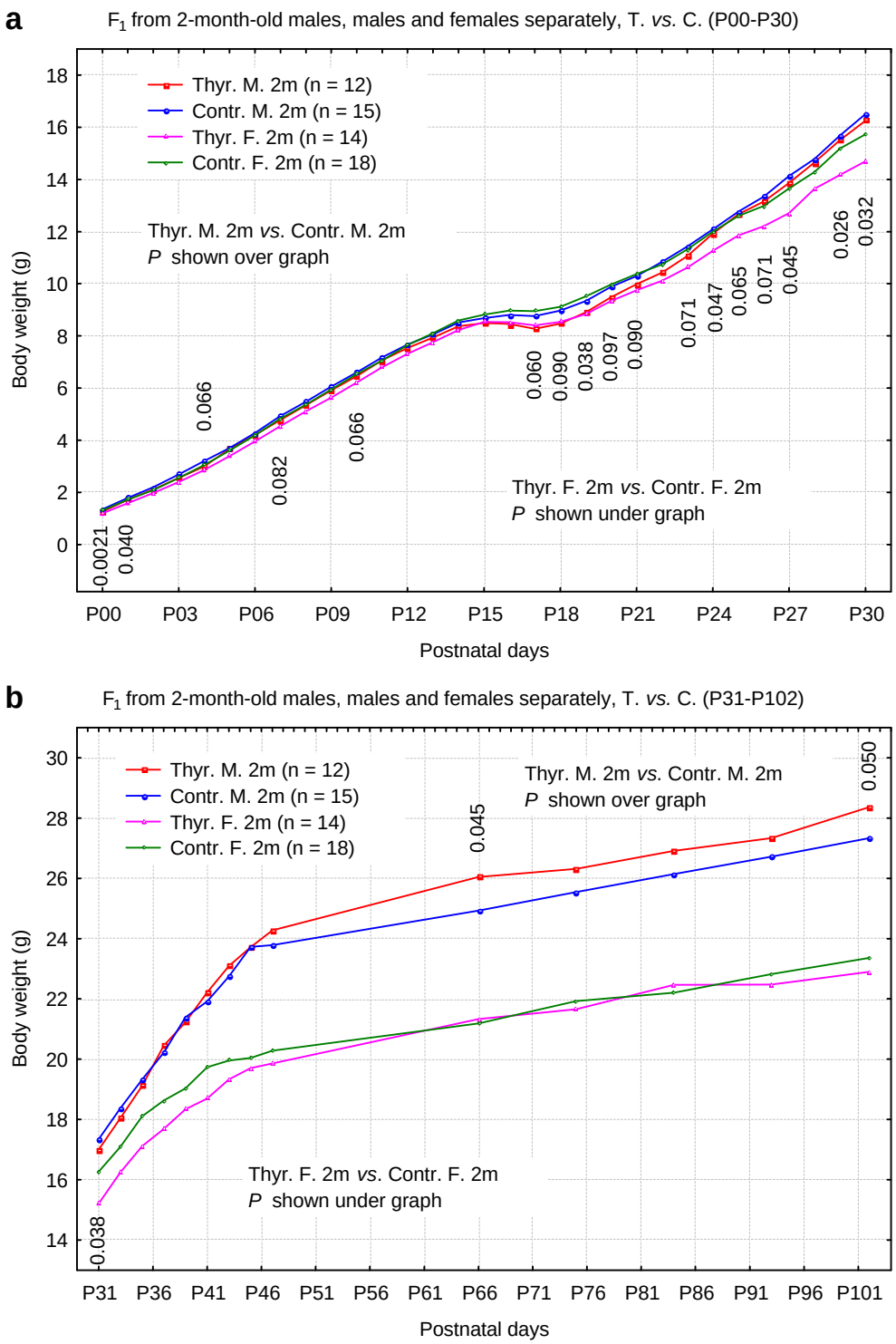



**Figure S26 | Body weight in male and female $F_1$ offspring of 2-month-old thyroxine-treated males.** Note once again that decreased body weight is highly significant in females only (**a**), whereas in males it is slightly increased in adults only (not really very significant, see **b**).

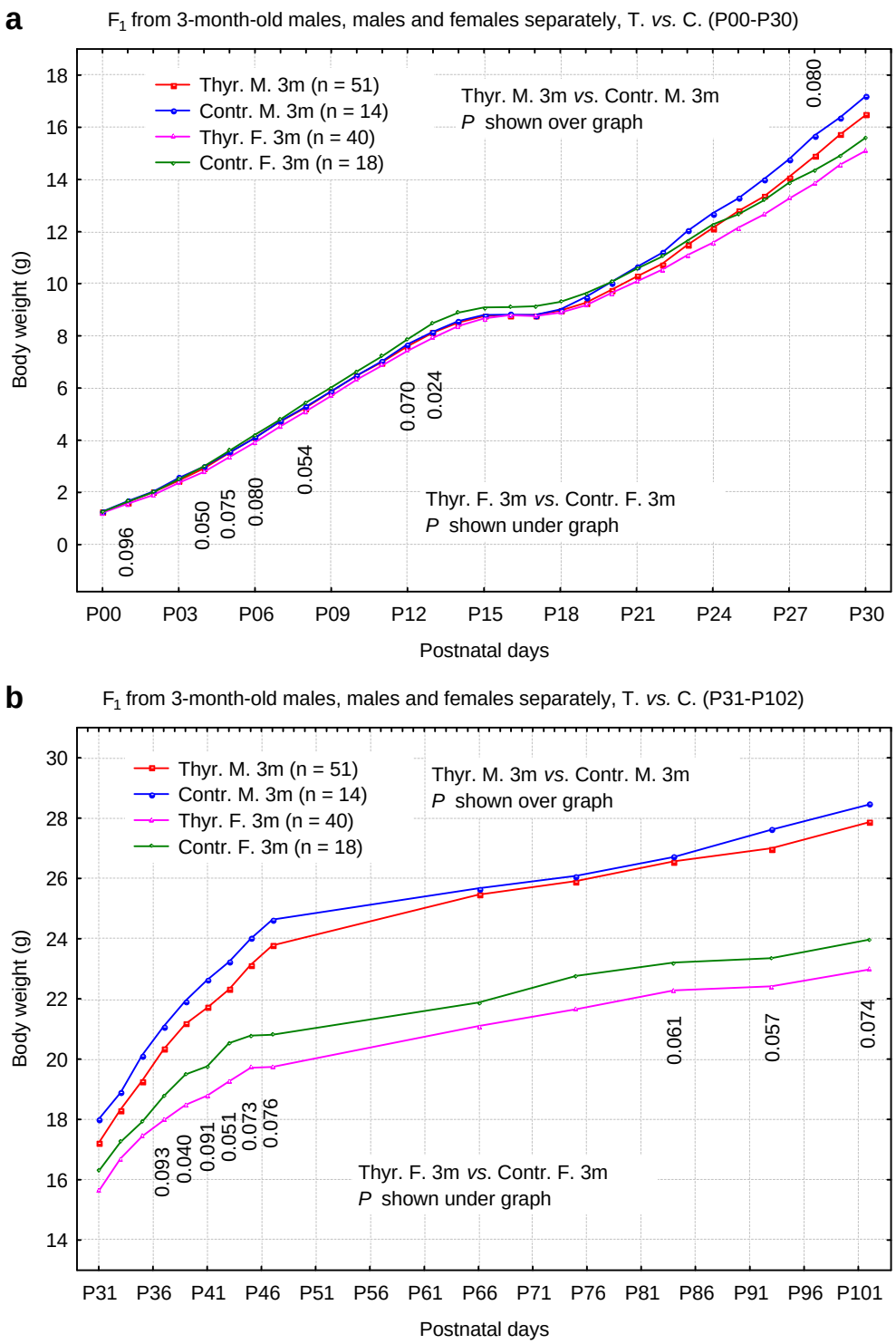



**Figure S27 | Body weight in male and female $F_1$ offspring of 3-month-old thyroxine-treated males.** Note slight decrease in body weight of females and nothing in males. Breeding of 2-month-old males (previous figure) produces definitely better results concerning thyroxine effects.

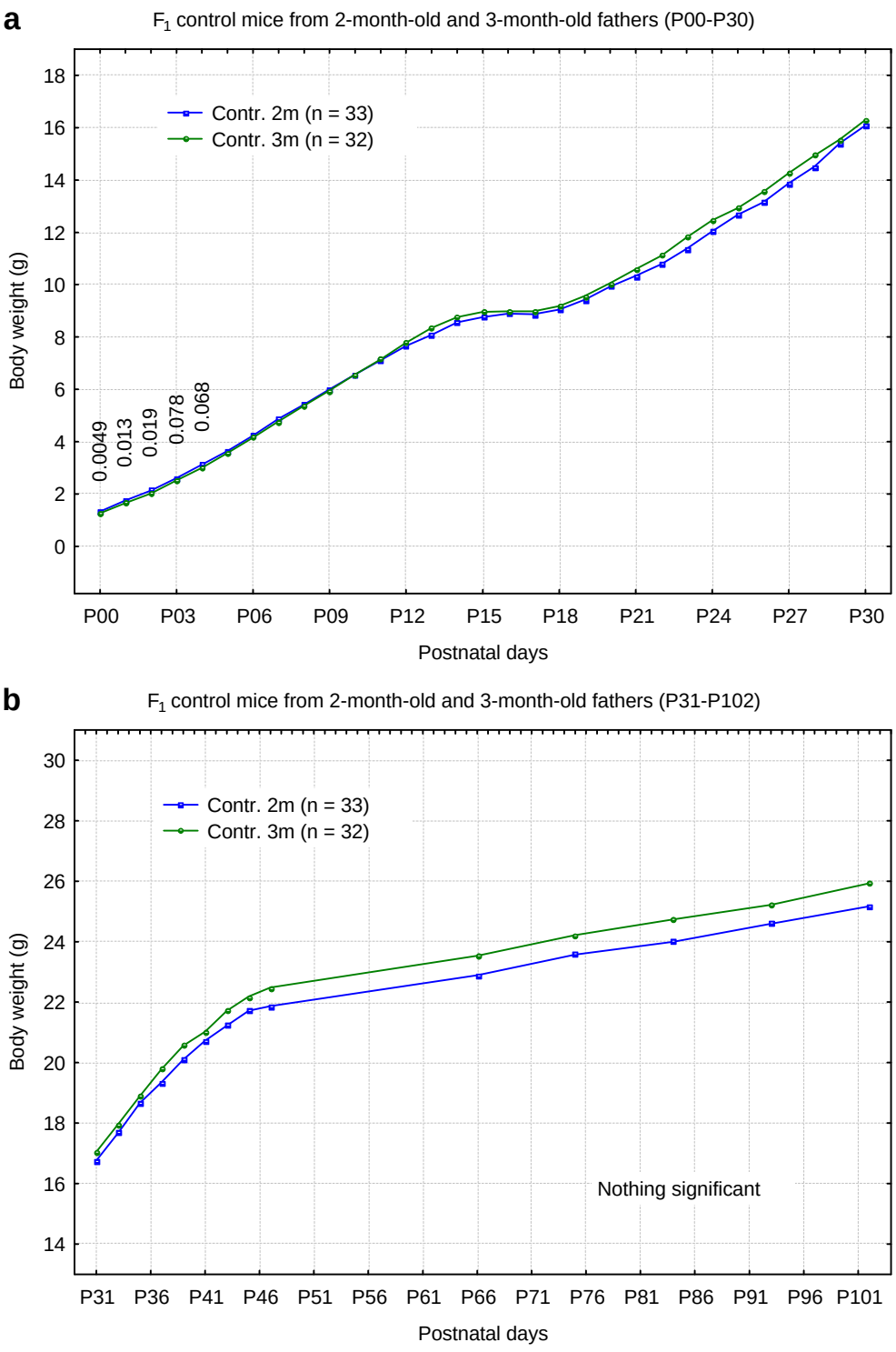



**Figure S28 | Body weight in control $F_1$ progeny of 2- and 3-month-old fathers.** Mothers of $F_1$ generation were always 3-month-old at breeding. Male and female progeny are pooled here. Note that descendants of 2-month-old males have increased birthweight and their weight remains increased during the first 3 days of postnatal life.

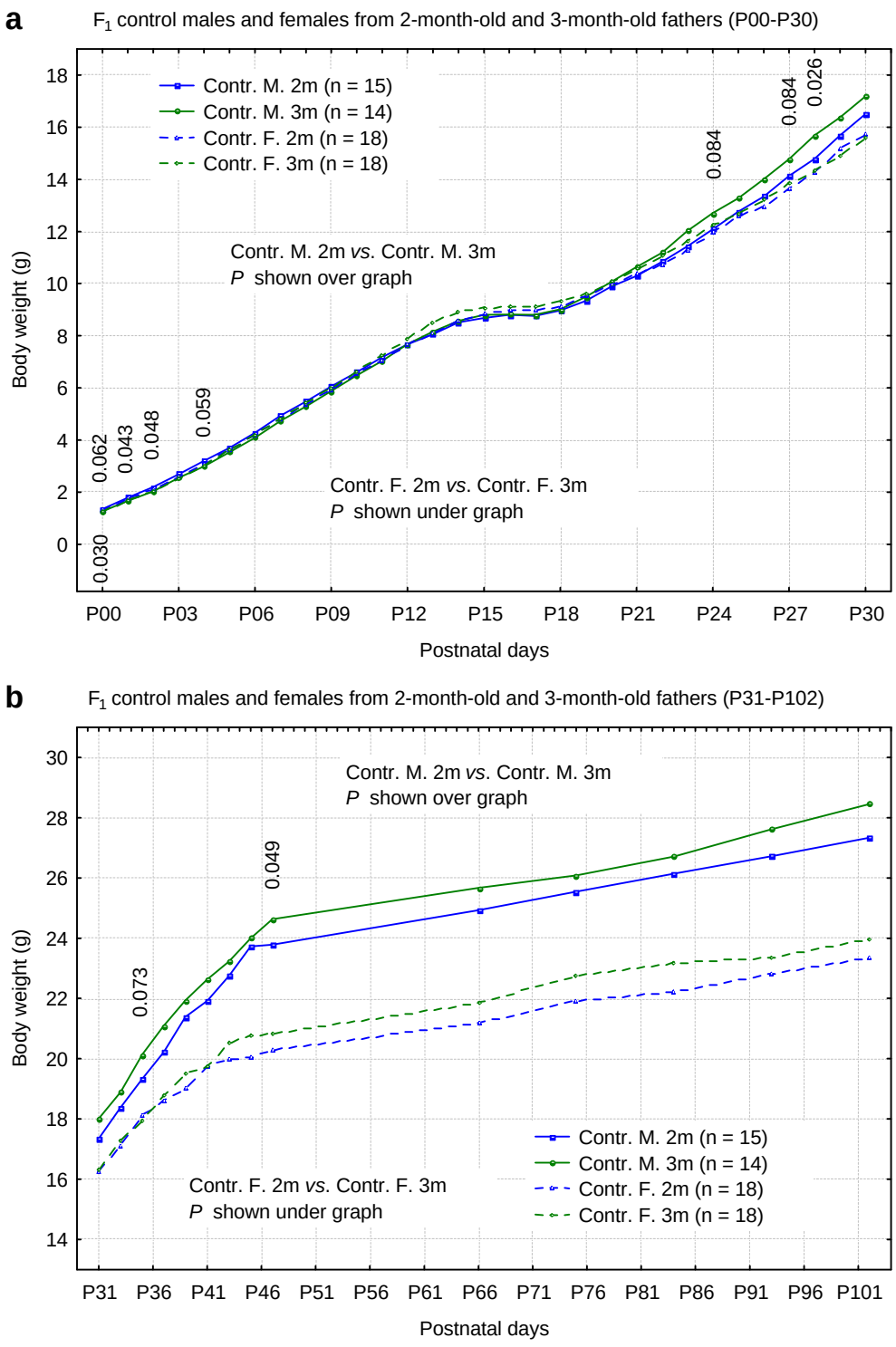



**Figure S29 | Body weight in $F_1$ control males and females obtained from 2- and 3-month-old fathers.** Note that the age of drug-naïve male parent has approximately the same impact upon body weight of male and female descendants. The role of litter size is important here (see Fig. S7e).

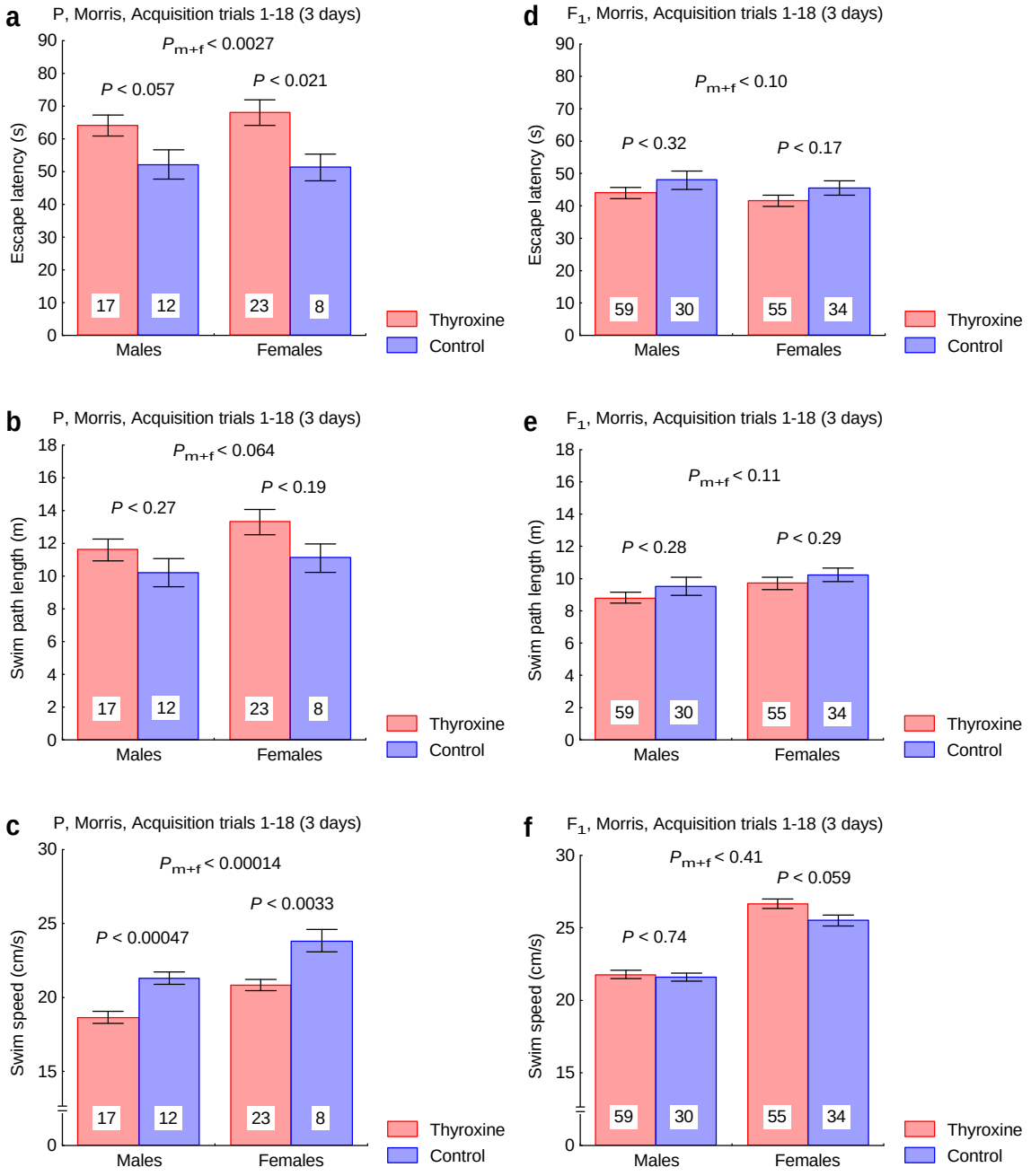



**Figure S30 | Thyroxine-treated mice and their $F_1$ progeny in the Morris water maze (acquisition).** The first 3 training days, acquisition phase, trials 1-18. Escape latency (**a**, **d**), swim path length (**b**, **e**) and swim speed (**c**, **f**). Each bar represents averaged value of 18 training trials. These 18 trials took place during 3 consecutive days, 6 trials daily. During each day 6 trials were separated by 60-min inter-trial intervals (spaced training). Note that despite impressive effects in thyroxine-treated mice (**a-c**), all differences in descendents (**d-f**) are rather mild during this acquisition phase (shown time-averaged values are not statistically significant). Nevertheless all of them are in the opposite direction in comparison with P generation.

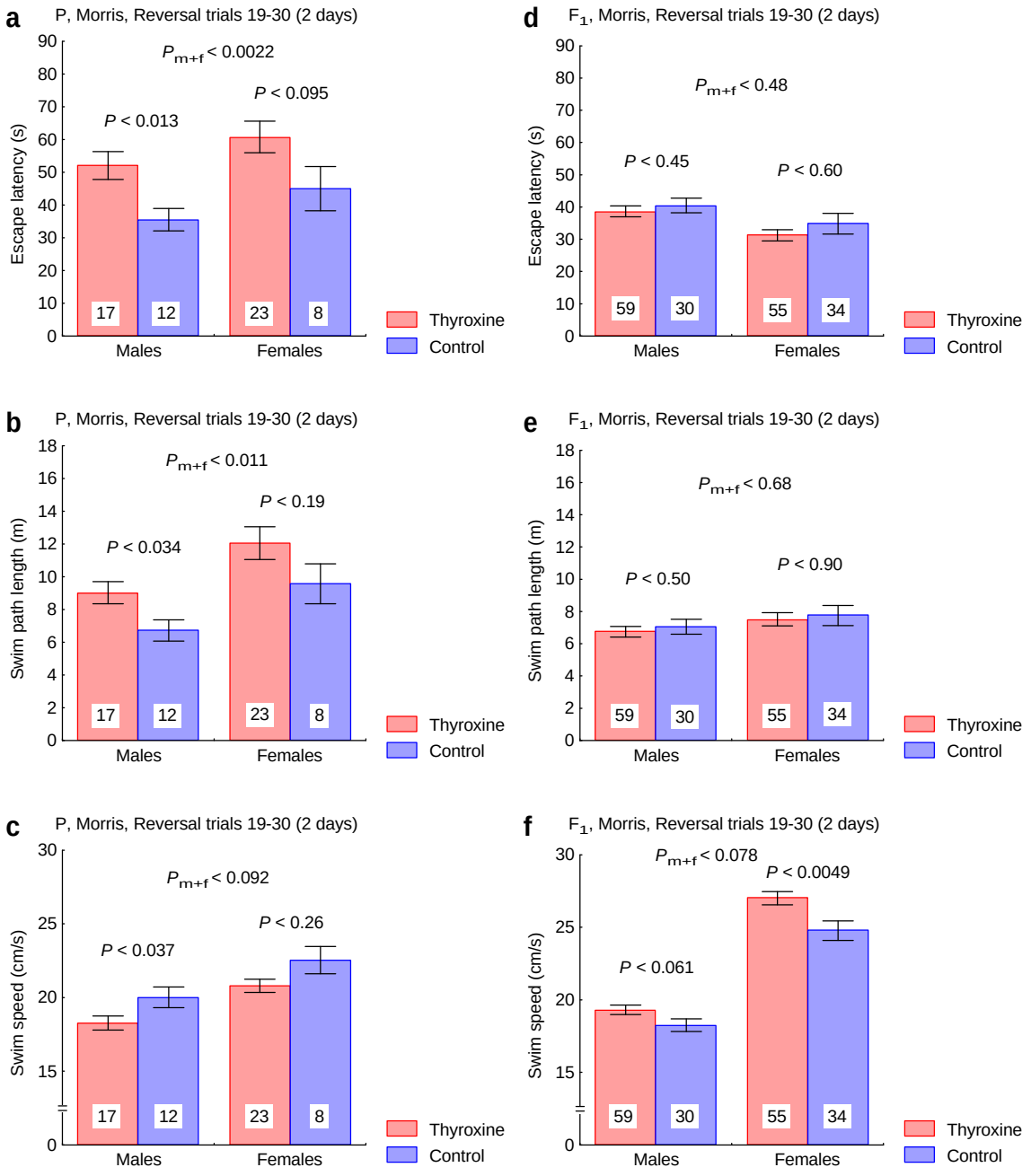



**Figure S31 | Thyroxine-treated mice and their $F_1$ progeny in the Morris water maze (reversal).** The last 2 training days (day 4 and day 5), reversal phase, trials 19-30. During these trials the platform was placed into the opposite quadrant. Escape latency (**a**, **d**), swim path length (**b**, **e**) and swim speed (**c**, **f**). Each bar represents averaged value of 12 reversal trials (two days; 6 trials daily, separated by 60-min inter-trial intervals). Note that female descendants have shown increased swim speed (**f**) during this reversal phase. However it is not a clear phenotypic inversion, because thyroxine-treated mice (P) have decreased swim speed mainly during acquisition phase (see **c** in the previous figure), but not reversal phase (see **c** here).

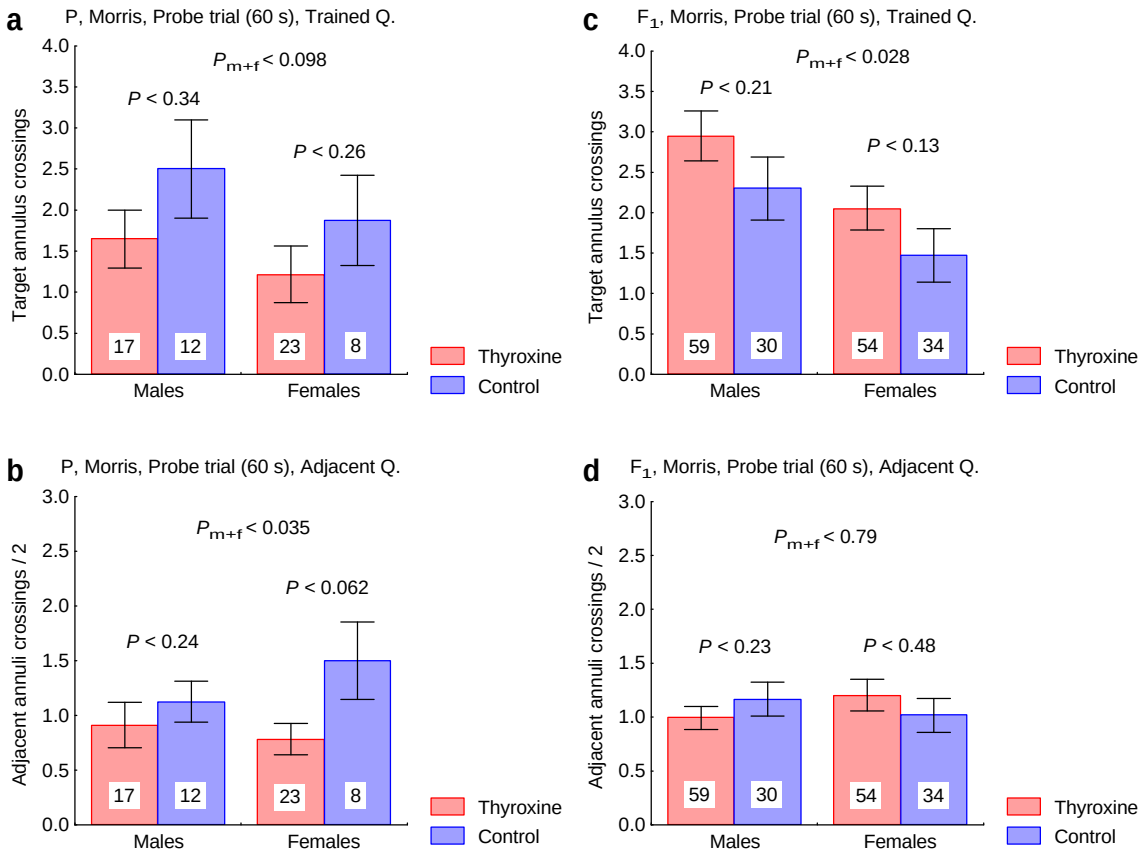



**Figure S32 | Thyroxine-treated mice and their $F_1$ progeny in the Morris water maze (probe trial).** As a probe trial we have taken the first 60 sec of the first reversal trial (trial 19; maximum trial duration can be up to 120 sec). We calculated two measures of spatial selectivity: (**a**, **c**) crossings of target (trained) annulus (annulus was determined as a square 16 cm on side) and (**b**, **d**) average of crossings over similar zones in adjacent quadrants (average of adjacent annuli crossings). Note that thyroxine-treated mice have shown slightly decreased number of target annulus crossings ($P < 0.098$) and decreased number of adjacent annuli crossings ($P < 0.035$) - both changes are indicators of decreased performance in this task. $F_1$ progeny has slightly increased number of target annulus crossings ($P < 0.028$) and unchanged number of adjacent annuli crossings ($P < 0.79$). All these changes have relatively low statistical significance (despite sufficient numbers of animals in the $F_1$ generation). However all above-mentioned changes are signs of phenotypic inversion (parental drug treatment entails opposite changes in the untreated progeny). The analysis of temporal dynamics given in the next 6 figures will discover specific time periods in the training protocol, associated with relatively high statistical significance.

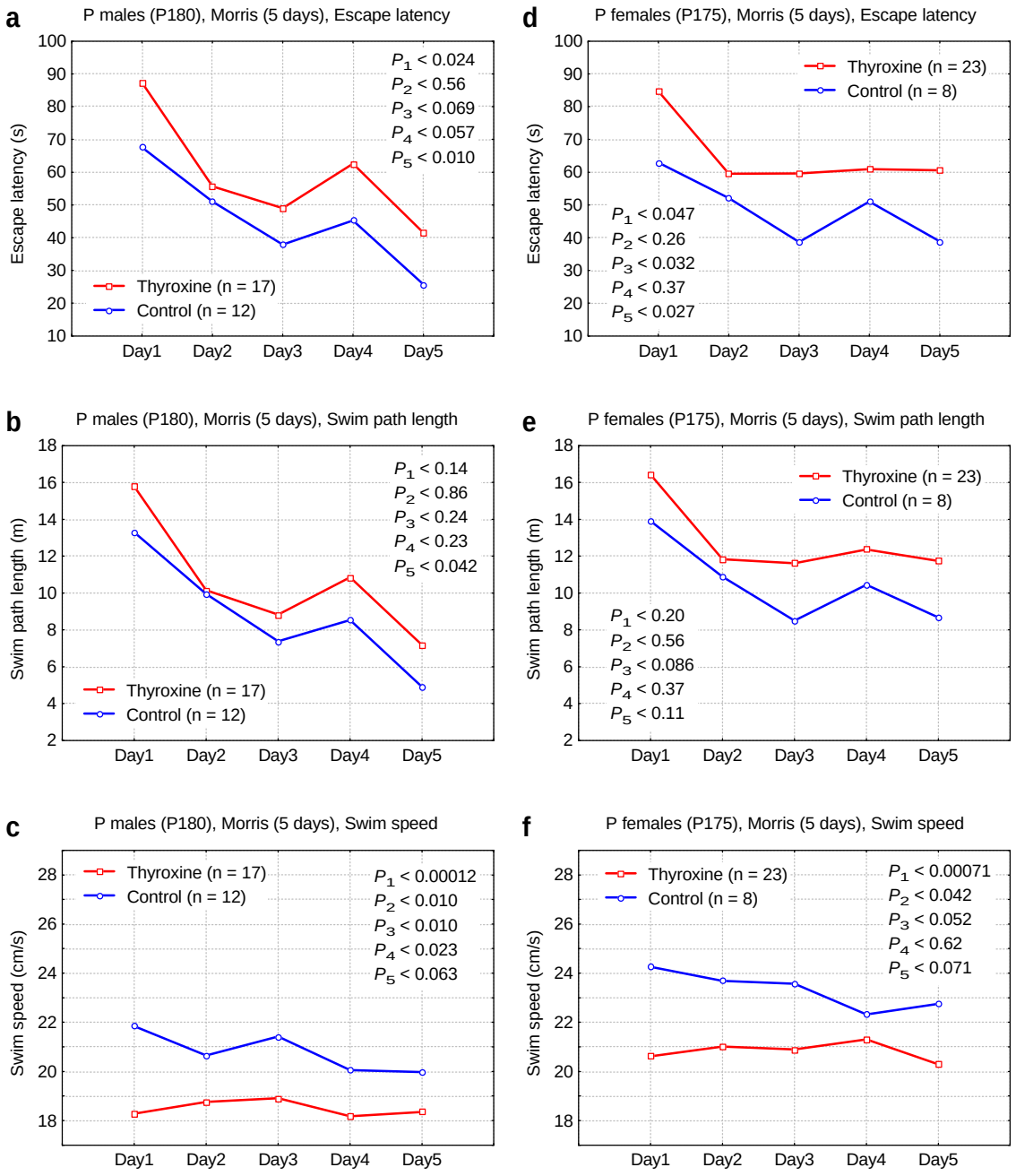



**Figure S33 | Thyroxine-treated males and females in the MWM, 5 days separately.** Escape latency (**a**, **d**), swim path length (**b**, **e**) and swim speed (**c**, **f**) of thyroxine-treated males (**a-c**) and females (**d-f**) in the Morris water maze. In our experiment P males were used for breeding, P females - not. Each time point represents averaged value of 6 consecutive daily training trials. These 6 trials were separated by 60-min inter-trial intervals (spaced training). During the last 2 days (day 4 and day 5) the platform was moved into the opposite quadrant (reversal phase of training). Note increased escape latency during the first day in both males (**a**) and females (**d**) and very significantly decreased swim speed during the same day in both males (**c**) and females (**f**).

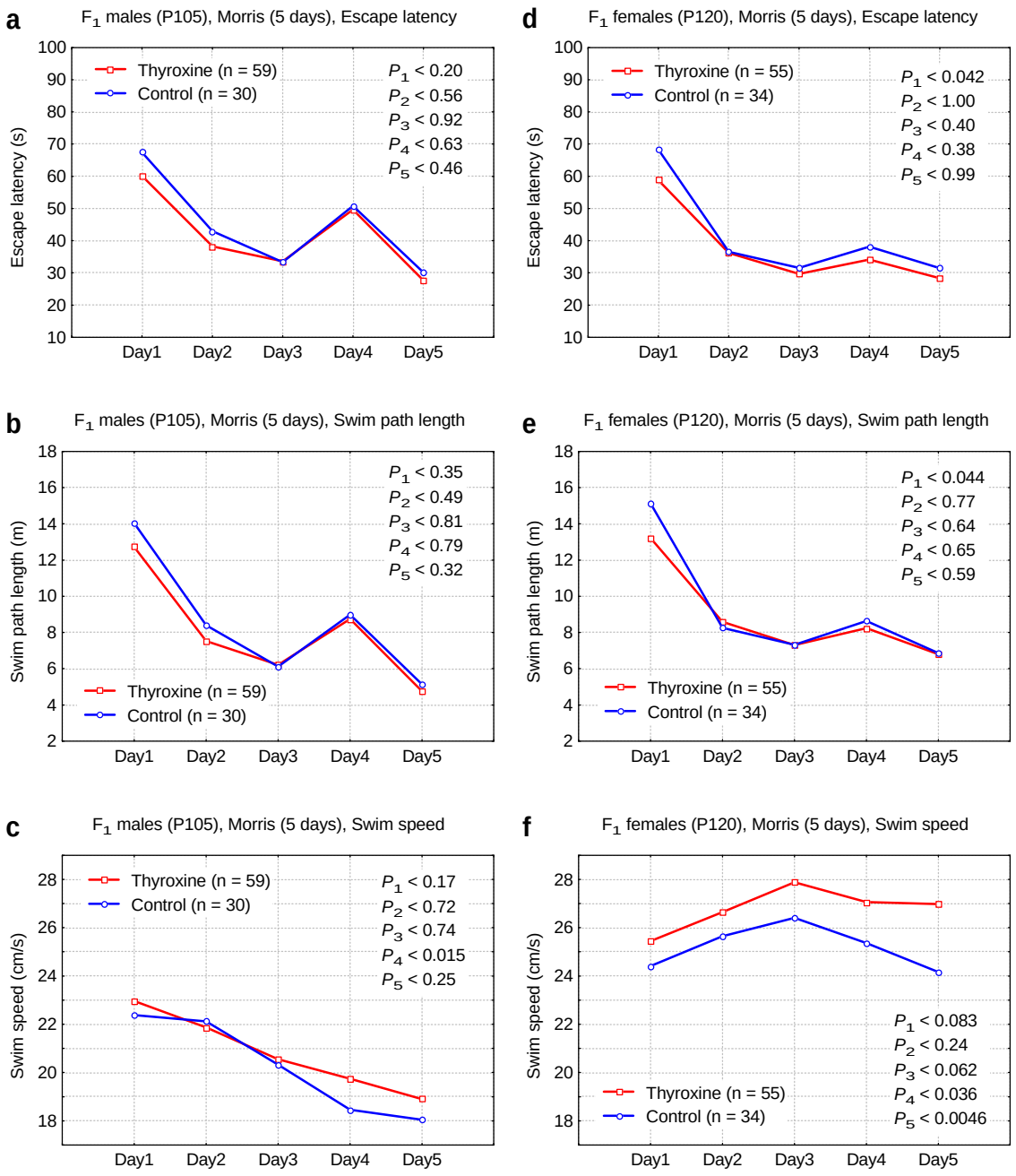



**Figure S34 | $F_1$ progeny of thyroxine-treated males in the MWM, 5 days separately.** Escape latency (**a**, **d**), swim path length (**b**, **e**) and swim speed (**c**, **f**) of $F_1$ male (**a-c**) and female (**d-f**) progeny, obtained from thyroxine-treated males and drug-naïve females. Note that, in comparison with thyroxine-treated mice *per se* (previous graphs), $F_1$ generation has all changes in the opposite direction, mainly with border-line significance. However, the days with the highest statistical significance are not the same for P and $F_1$. For example, concerning swim speed, it was the first day in P males and females (acquisition phase), but day 4 in $F_1$ males and days 4 and 5 in $F_1$ females (reversal phase). Pure phenotypic inversion can be seen if we look at the same data with higher temporal resolution (2-trial resolution, see the next 4 figures).

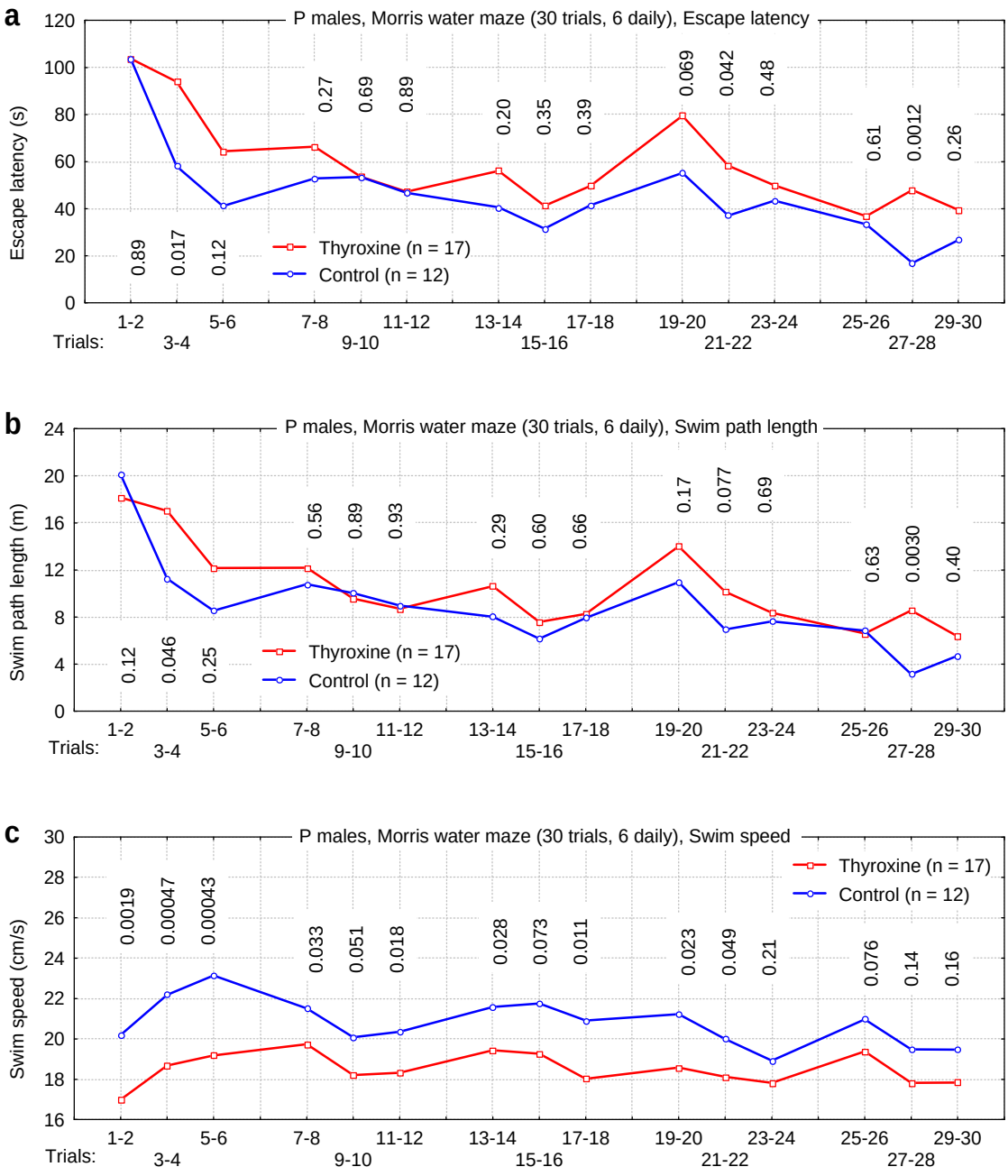



**Figure S35 | Thyroxine-treated P males in the Morris water maze, 30 trials in pairs.** Escape latency (**a**), swim path length (**b**) and swim speed (**c**) of thyroxine-treated male mice (P generation) in the Morris water maze. Each time point represents averaged value of 2 consecutive trials (thus, 6 daily trials conform to 3 time points). Note that trials 3-4 look the most interesting in terms of statistical significance (all three shown variables are changed significantly here).

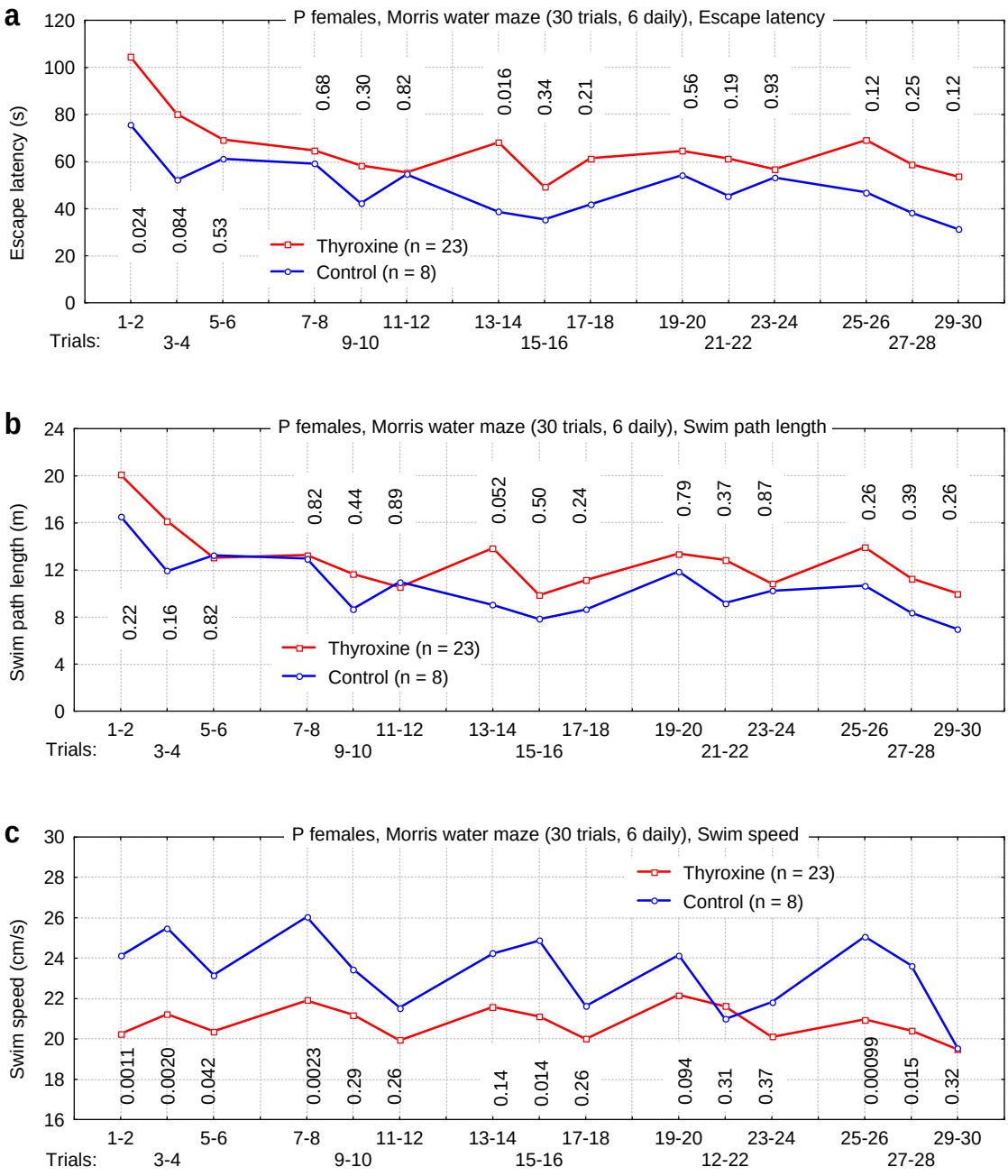



**Figure S36 | Thyroxine-treated P females in the Morris water maze, 30 trials in pairs.** Escape latency (**a**), swim path length (**b**) and swim speed (**c**) of thyroxine-treated female mice (known as P generation) in the Morris water maze. Each time point represents averaged value of 2 consecutive trials (thus, 6 daily trials conform to 3 time points). Note that, similar to thyroxine-treated males (previous graphs), trials 3-4 look rather interesting.

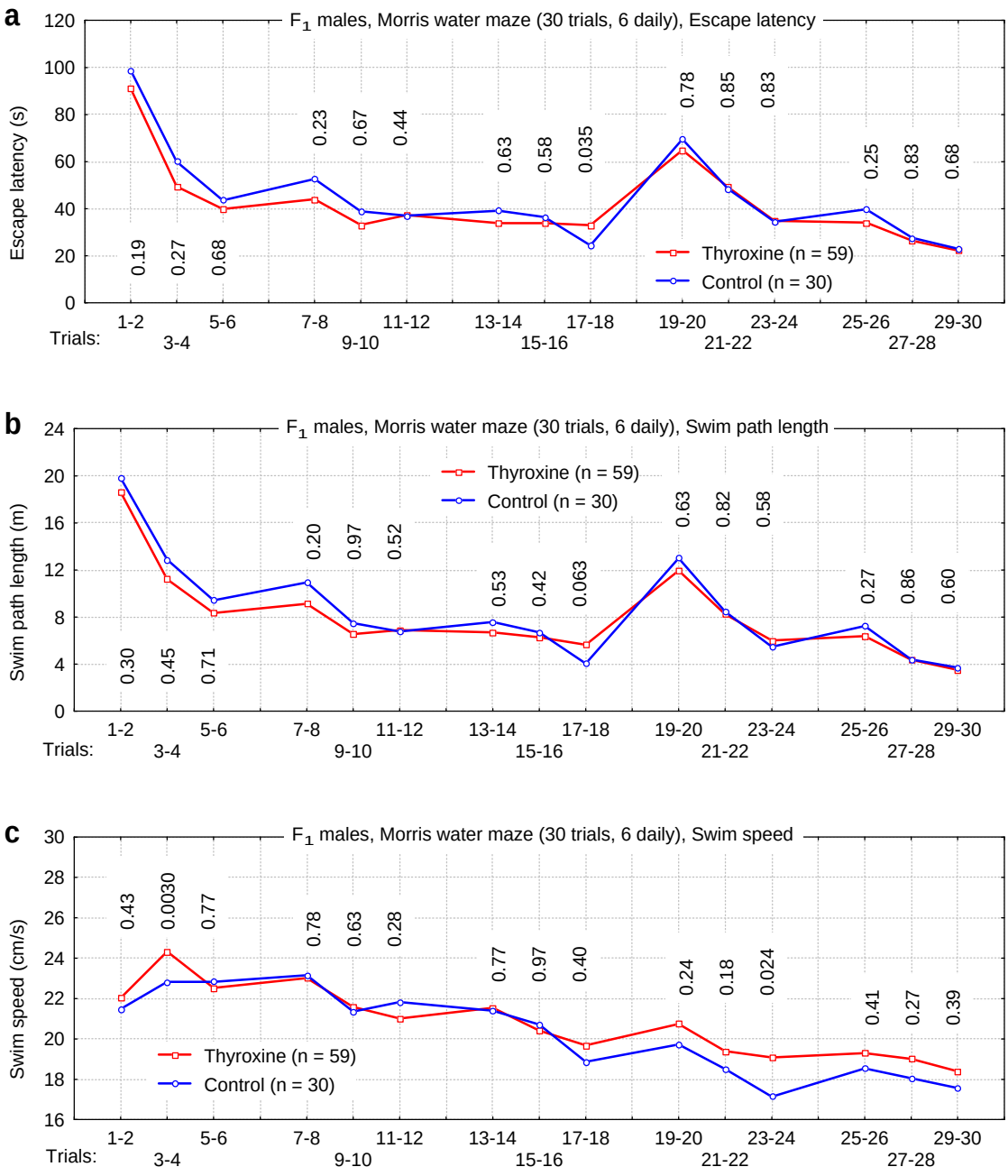



**Figure S37 | $F_1$ males, progeny of thyroxine-treated males, MWM 30 trials in pairs.** Escape latency (**a**), swim path length (**b**) and swim speed (**c**) of the $F_1$ male progeny of thyroxine-treated males and drug-naïve females. Note that all changes in this drug-naïve $F_1$ progeny are less wide spread than similar changes in their thyroxine-treated male parents. In $F_1$ many variables in many time points show exact coincidence between experimental and control groups (see **a**-**b**). However in some specific time points the changes in descendents are clearly reversed in comparison with parents (**c**, trials 3-4).

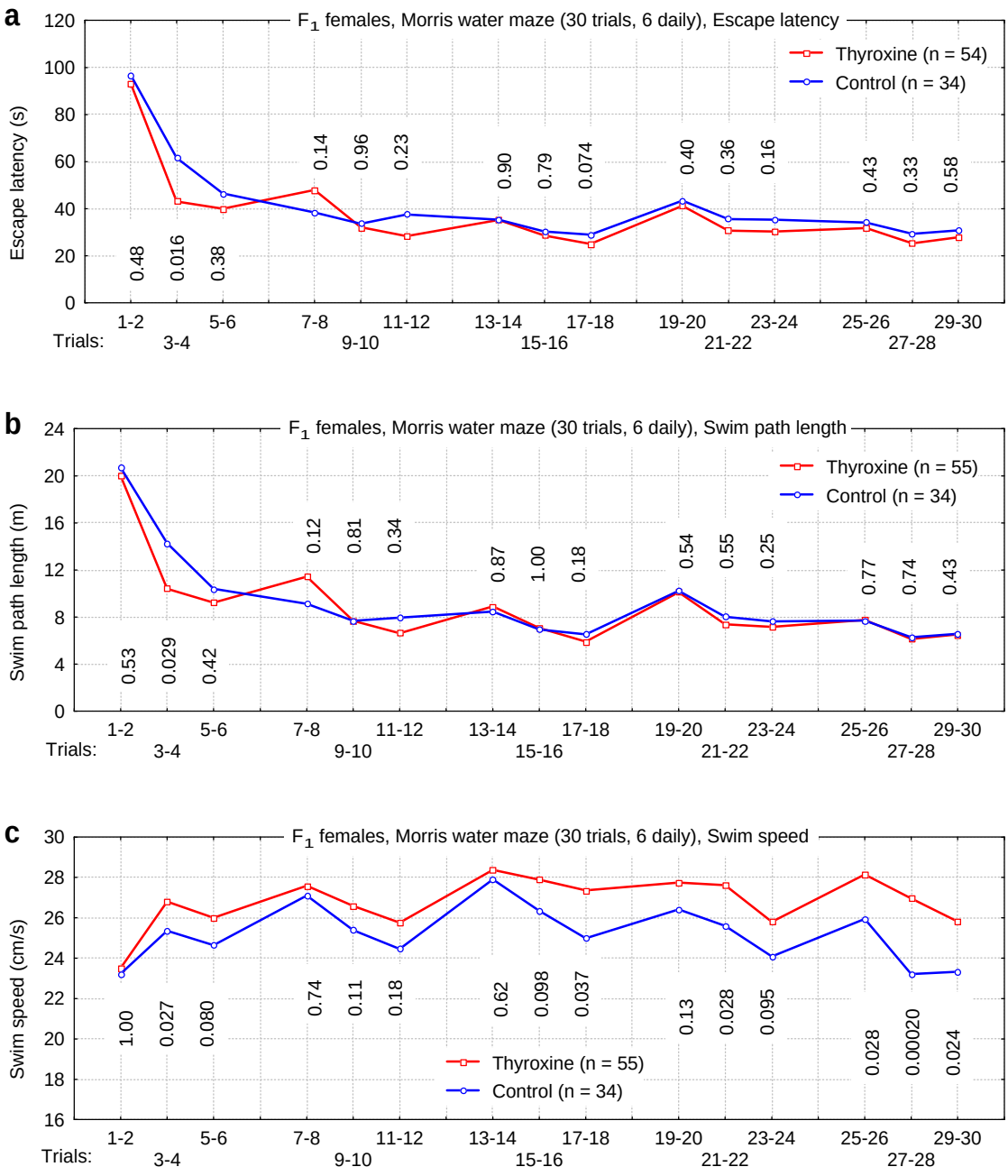



**Figure S38 | $F_1$ females, progeny of thyroxine-treated males, MWM 30 trials in pairs.** Escape latency (**a**), swim path length (**b**) and swim speed (**c**) of the $F_1$ female progeny in the Morris water maze. Note that, similar to $F_1$ males, all changes in $F_1$ females are less wide spread than in thyroxine-treated animals (*e.g.* thyroxine-treated females). Escape latency (**a**) and swim path length (**b**) during reversal phase (trials 19-30) show exact coincidence between groups. However swim speed (**c**) differs significantly during this period. During trials 3-4 all shown variables are changed significantly and these changes show clear phenotypic inversion. Note also that all females use different learning strategy in comparison with males: females use high swim speed and it goes up, whereas male speed goes down and they use spatial orientation more (see probe trial).

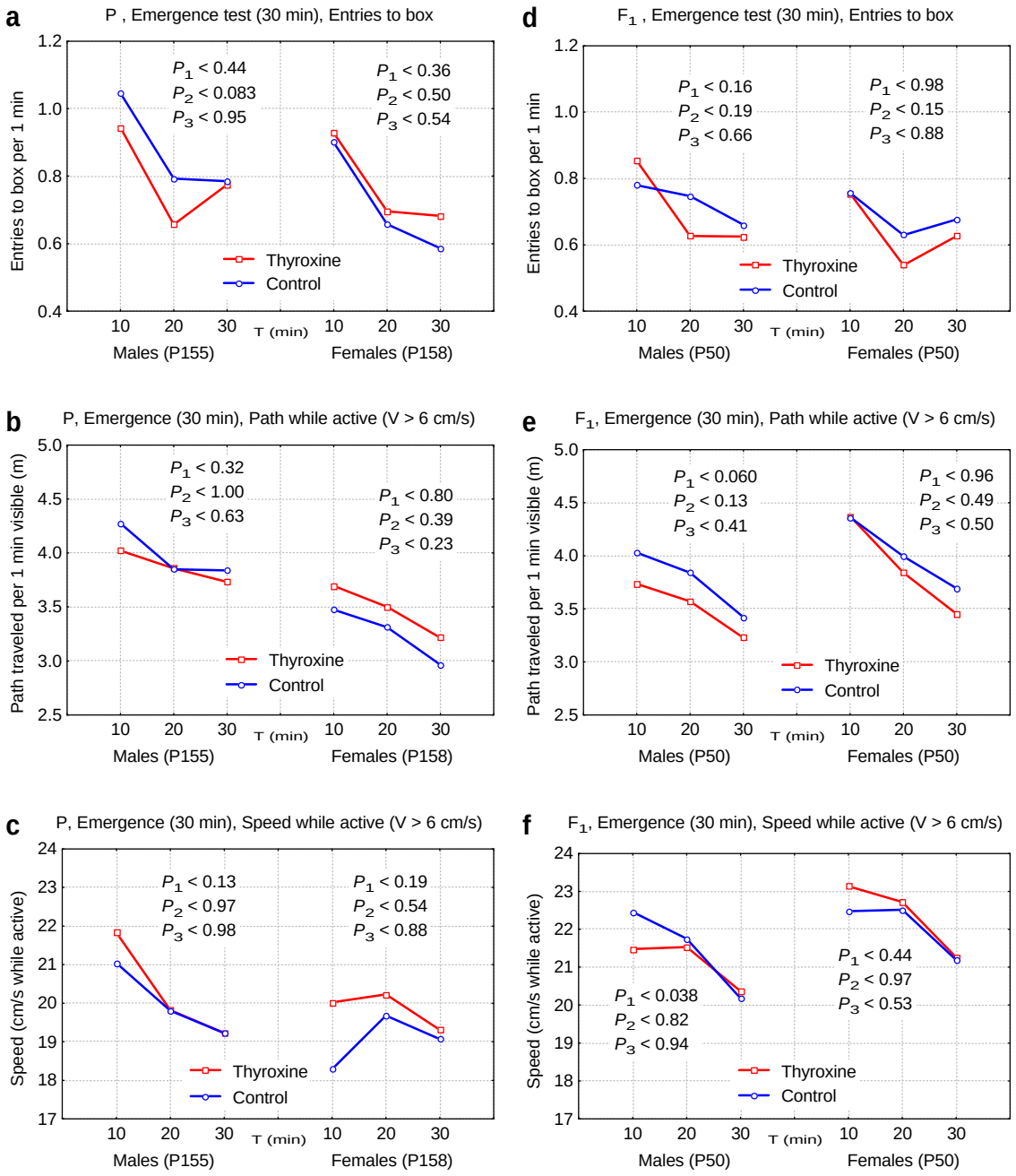



**Figure S39 | Thyroxine-treated mice and their $F_1$ progeny in the Emergence test.** About 3 days before Emergence test animals were distributed into individual home cages (Type 2). 24 hours before the test a small plastic box was placed in each individual home cage (box 12 × 8 × 4 cm with an opening 8 × 4 cm; we use molecular biology ART tips boxes). During the test this box was taken from the home cage and placed into 50 × 50 cm arena, surrounded by 37 cm walls, into one corner with distance of 5 cm from each wall, opening towards open space. As we can see in **a-f**, neither thyroxine-treated mice, nor their offspring have shown any differences with respect to their control groups, except very slight decrease in speed of $F_1$ males during the first 10 min (**f**). Mean. Mann-Whitney U-test. P50, P155, P158 - age of mice (postnatal days).

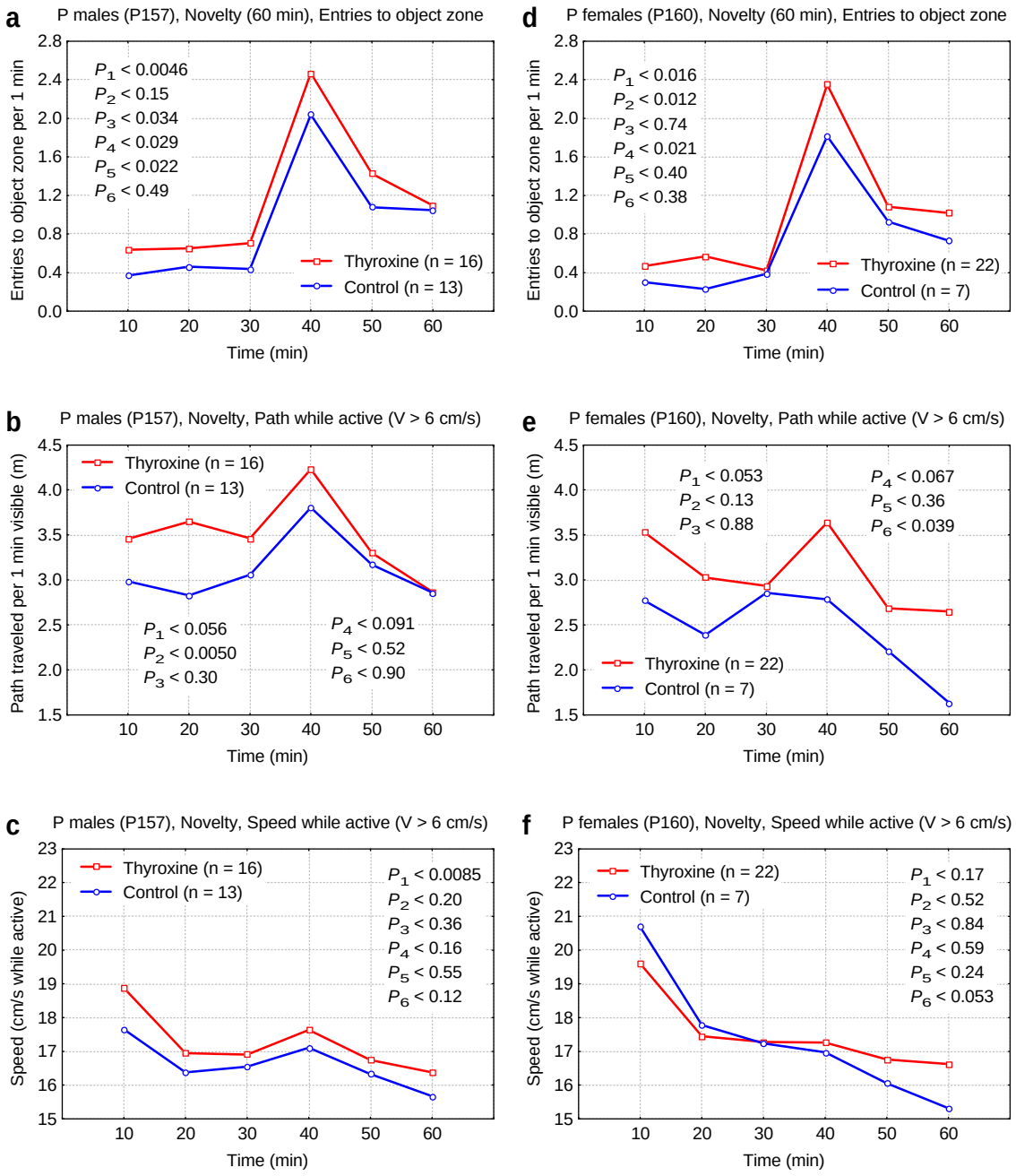



**Figure S40 | Thyroxine-treated male and female mice in the Novelty test.** 24 hours after the Emergence test (see previous figure) each mouse was introduced into the same arena, but without any box (empty arena) for 60 min. The first 30 min of this period can be analyzed as Open-field test. After the first 30 min into the middle of arena a small new object was placed for the next 30 min (we use 12 × 4 cm semi-transparent 50 ml Falcon tube, placed vertically). Object zone was determined as a circle d = 16 cm in the middle of arena. Thyroxine-treated male parents and their thyroxine-treated sisters (not used for breeding) are shown in this figure. Note that, contrary to Emergence test, there are some differences here. Animal numbers (n = …) are the same for Novelty and Emergence tests.

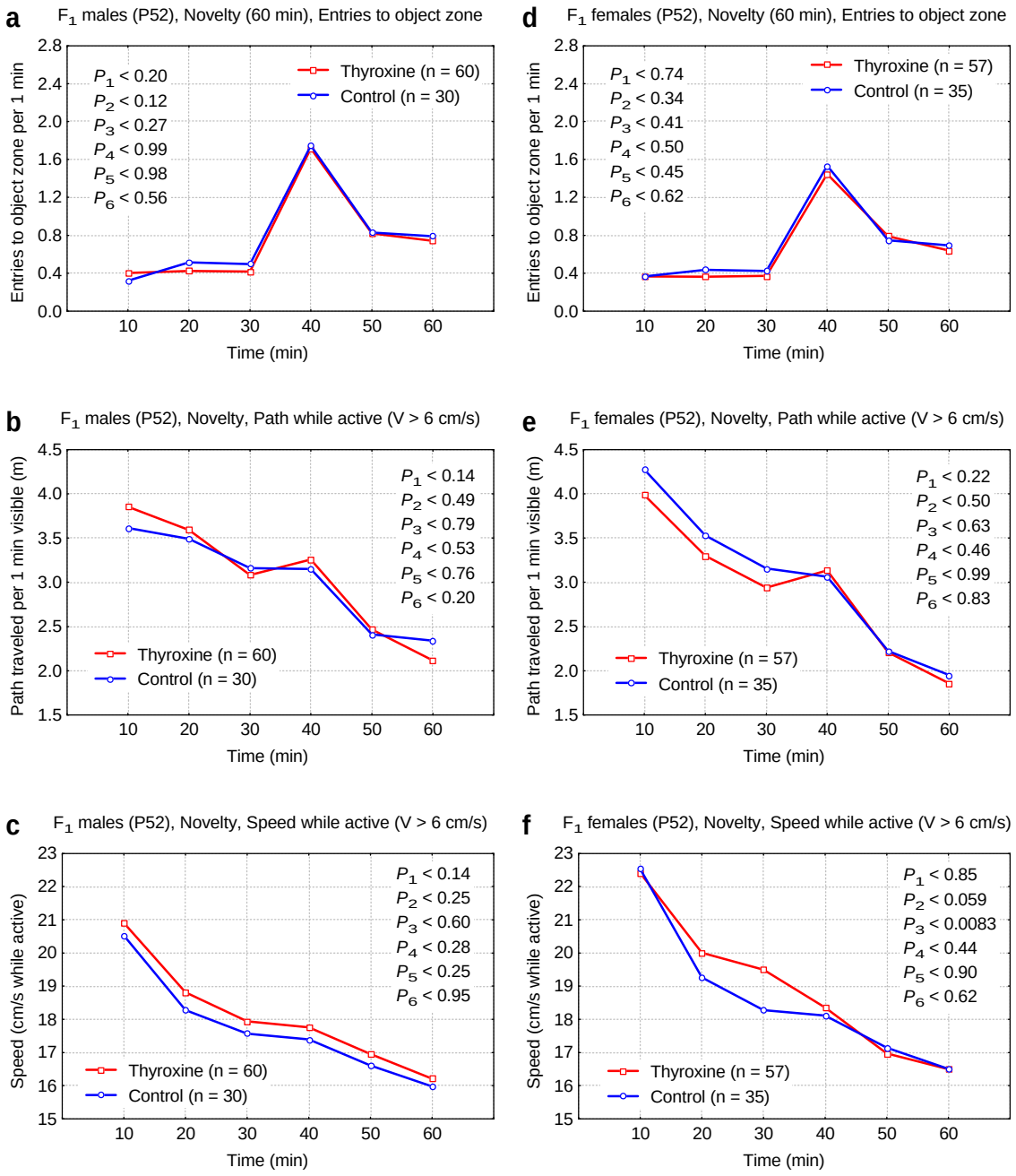



**Figure S41 | $F_1$ male and female progeny of thyroxine-treated males in the Novelty test.** Object zone was determined as a circle with d = 16 cm in the middle of arena. Animals were considered "active" when their speed was higher than 6 cm/s. $F_1$ generation, obtained from thyroxine-treated males and drug-naïve females, is shown here. Note that experimental and control animals demonstrate practically identical behaviour in this test (see especially **a**, **d**). Some difference exists only in females (**f**): they have decreased habituation [decreased speed decline] in empty arena (increased speed *per se* with respect to control group during the second and the third 10-min intervals). This increased speed may have the same origin as increased swim speed observed in the same mice in Morris water maze afterwards.

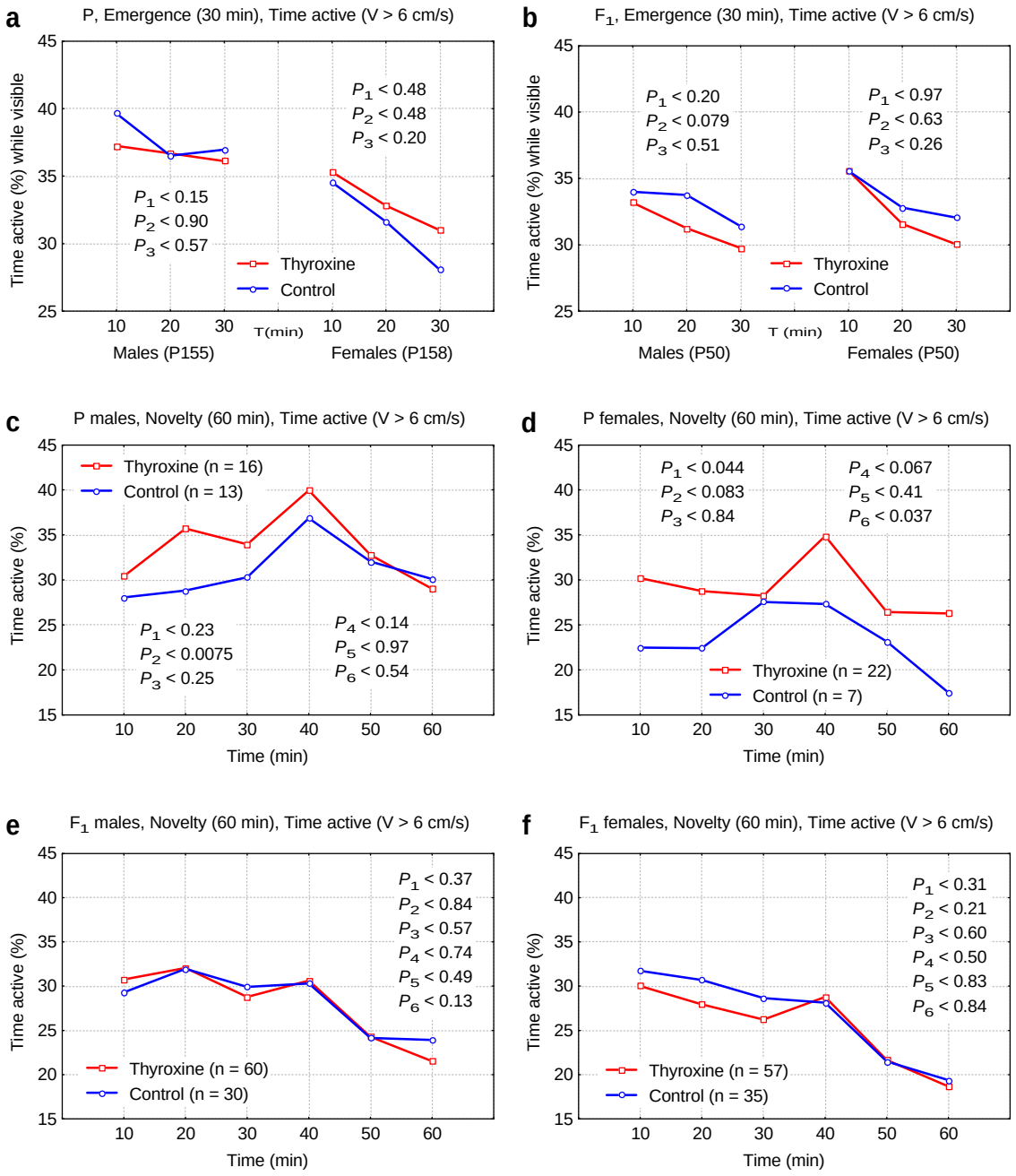



**Figure S42 | Time active (V > 6 cm/s) in Emergence and Novelty tests (P and $F_1$).** Behavioural continuum was arbitrary divided by us into active (V > 6 cm/s) and passive behavioural episodes and due to this reason we decided to show % of time which was spent by an animal being in active state in both Emergence (**a-b**) and Novelty (**c-f**) tests. This active time is about 30-35% in both Emergence and Novelty tests. There is no difference between all groups in the Emergence test (**a-b**). In the Novelty test neonatally thyroxine-treated mice are slightly more active (**c-d**) and there is no difference in progeny (**e-f**). We can see that Time active (%) correlates best of all with Path traveled while active per 1 min visible (**b**, **e** in three previous figures).

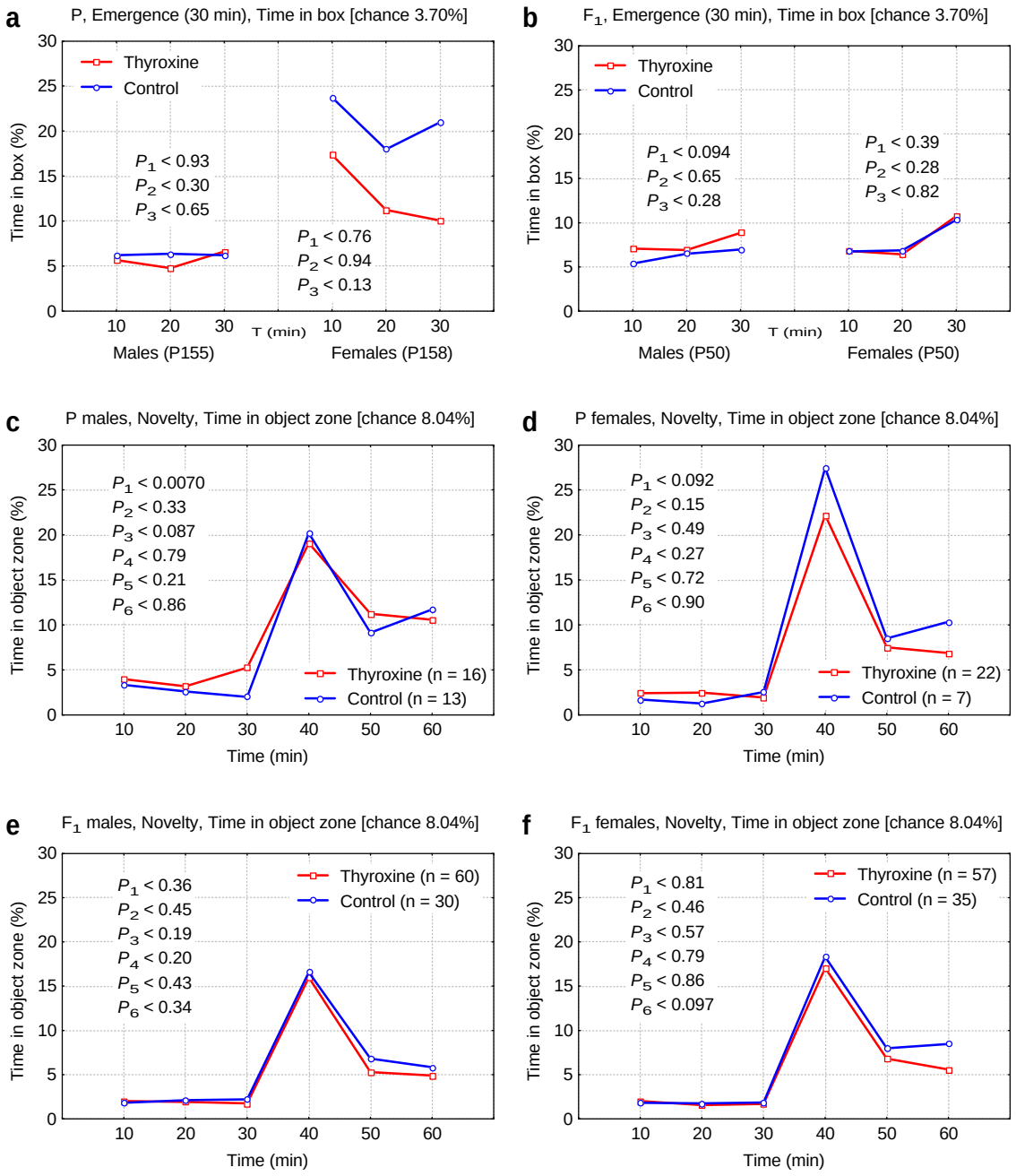



**Figure S43 | Time in box in Emergence test and time in object zone in Novelty test (P and $F_1$).** Note that both experimental and control mice have spent the same time inside the box in Emergence test. In Novelty test the appearance of a new object in the middle of arena increases time spent in object zone, but this increase is identical for all groups. In $F_1$ generation the tests were applied in the following order: Emergence & Novelty, Morris water maze, and Two-way avoidance. We discuss these tests in the opposite order, because we would like to show the most significant results first (Two-way avoidance), then - less significant (Morris water maze), and finally - non-significant (Emergence & Novelty tests). In the Novelty test there is some decrease in habituation in $F_1$ females in the Open-field only (the first 30 min of Novelty test).

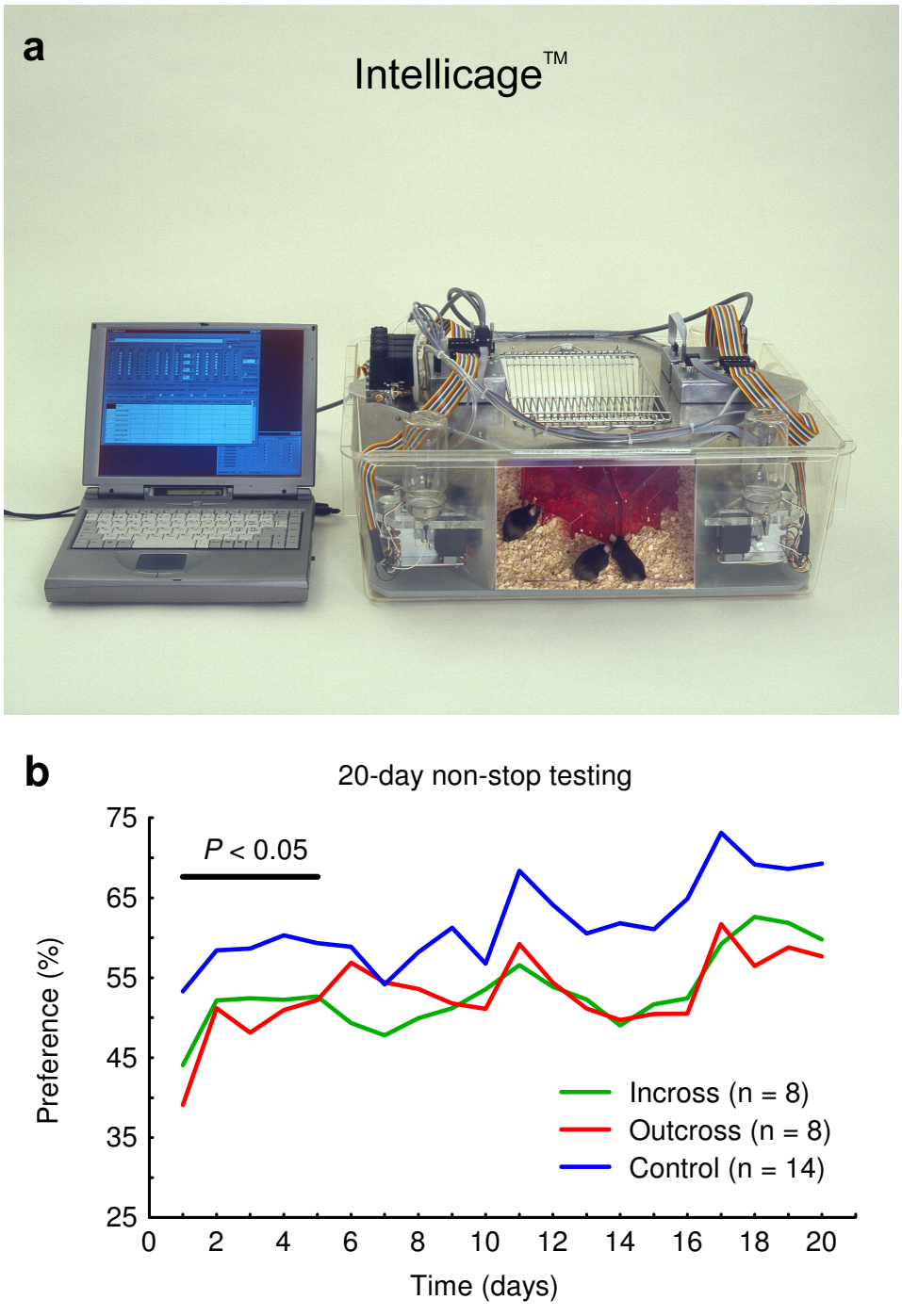



**Figure S44 | Spatial learning of $F_3$ female descendants of thyroxine-treated male mice. a**, Transponder-tagged animals were living in the Intellicage™ in balanced mixed groups (16 mice per cage, 4 mice from each group) during 20 days. Each corner was equipped with antenna (transponder reader). Water reinforcement was provided for each animal in one corner and in the others the animal received air puff weak negative reinforcement. **b**, Corner preference (percent of visits) over 25% chance level indicates spatial learning. Bar shows time period with $P < 0.05$ "Incross + Outcross" *vs*. Control. Mean. Note that spatial learning is slightly impaired in both Incross and Outcross groups.

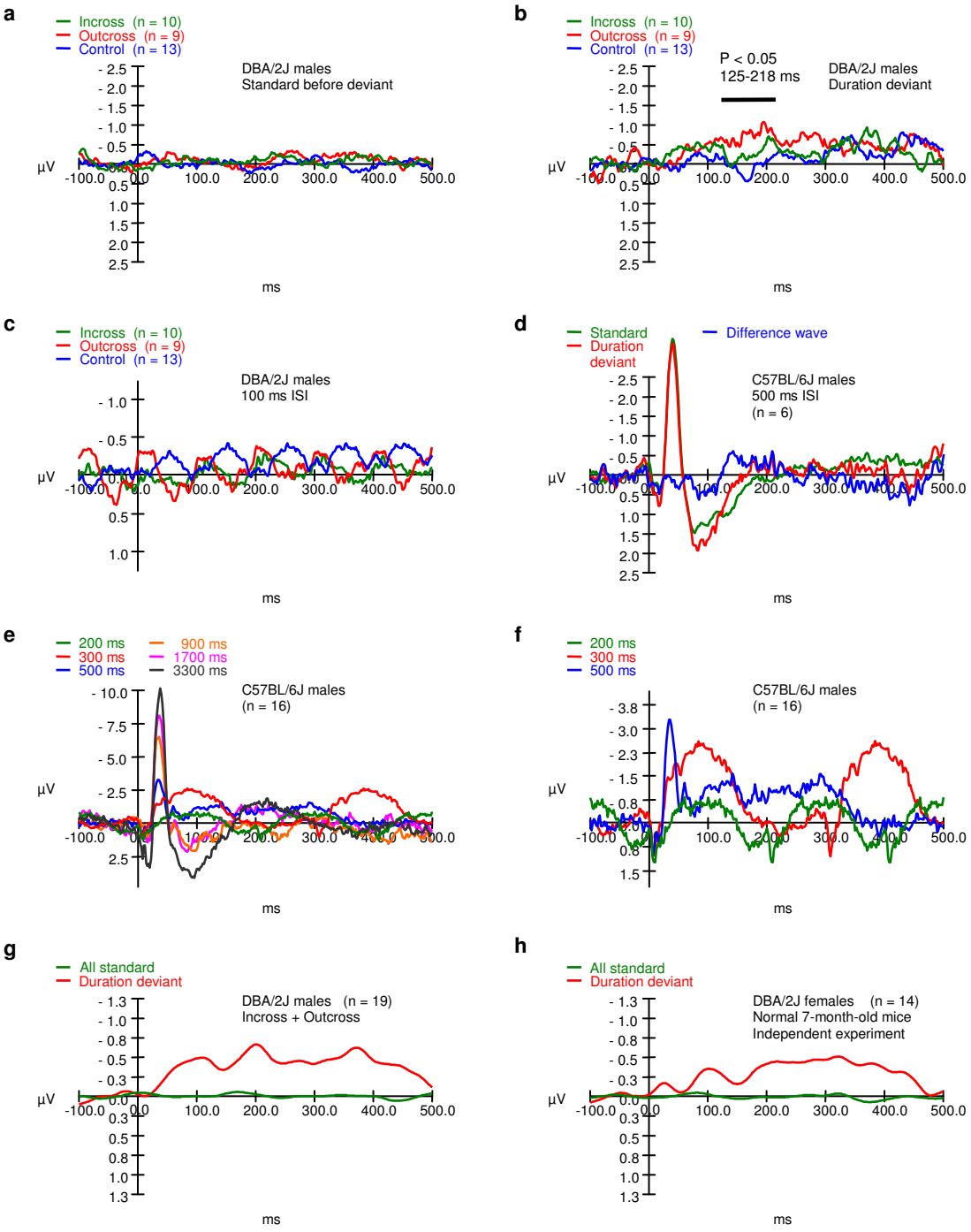



**Figure S45 | Auditory evoked potentials in the $F_3$ male descendants of thyroxine-treated males.** Standard before deviant (**a**) and duration deviant (**b**) evoked potentials in $F_3$ males (grand-average of paradigms d-g from Fig. S47). **c**, Evoked potentials (standard) in $F_3$ males (paradigm b from Fig. S47). **d-f,** Evoked potentials in C57BL/6J males from the experiment with *Gdi1*-mutant mice: **d**, paradigms d-g from Fig. S47; **e**, paradigms c-h from Fig. S46; **f**, paradigms c-e from Fig. S46. **g-h**, Standard and duration deviant evoked potentials in the experimental $F_3$ males (DBA/2J, age 18 months) and in the DBA/2J mice from the experiment with cage enrichment (age 7 months), paradigms d-g, Fig. S47.

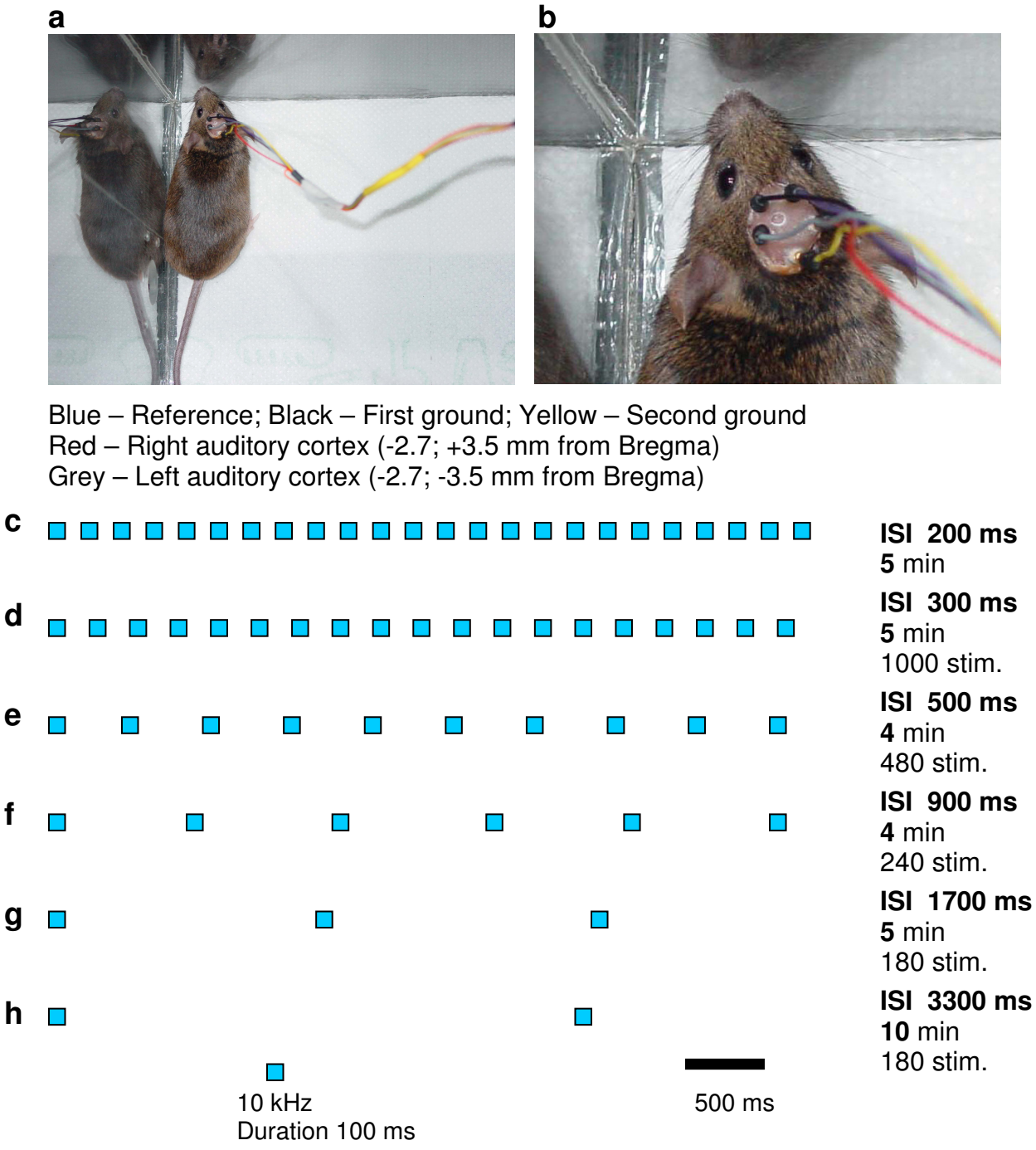



**Figure S46 | Electrode coordinates and paradigms with different inter-stimulus intervals.** **a-b**, EEG electrode coordinates (129sv mouse is shown as an example). **c-h**, schematic representation of paradigms with different inter-stimulus intervals (onset-to-onset) for investigation of auditory evoked potentials from primary auditory cortex in C57BL/6J male mice. C57BL/6J mice are chosen for comparison as the most standard inbred mouse strain used in biomedical research. The order from **c** to **h** coincides with the real order of an animal exposure to these paradigms (with 3-5 min sound-free time intervals between consecutive paradigms).

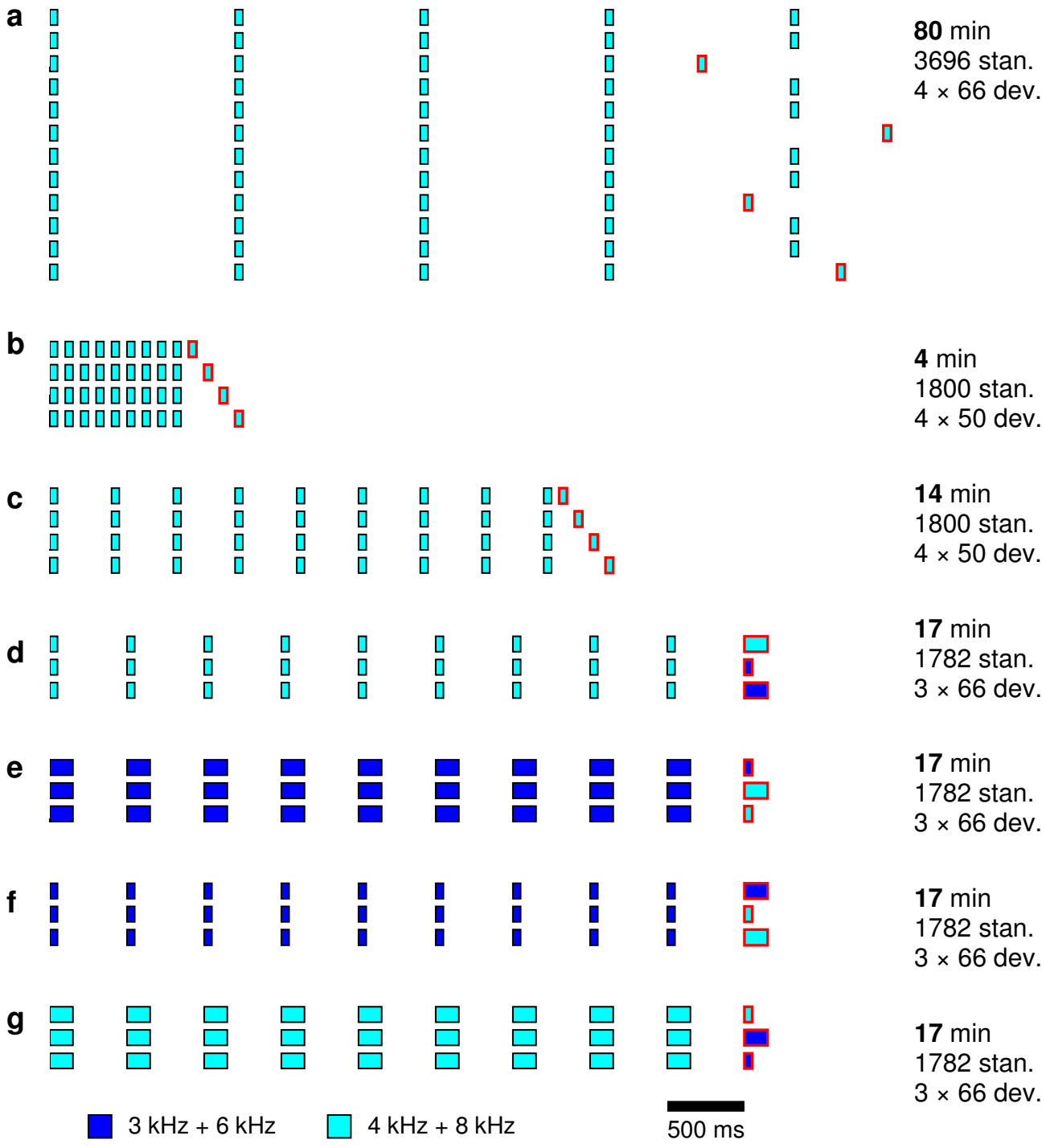



**Figure S47 | Mismatch negativity paradigms for $F_3$ male descendants of thyroxine-treated males.** Schematic representation of paradigms for investigation of mismatch negativity in $F_3$ male mice. Stimulus duration 50 or 150 ms. Red frame – deviants. The order from **a** to **g** coincides with the real order of an animal exposure to these auditory paradigms (with 3-5 min sound-free time intervals between consecutive paradigms, *i.e.* between **a**, **b**, **c**, **d**, **e**, **f** and **g**). Data from paradigms **d-g** are shown in the Fig. S45a-b.

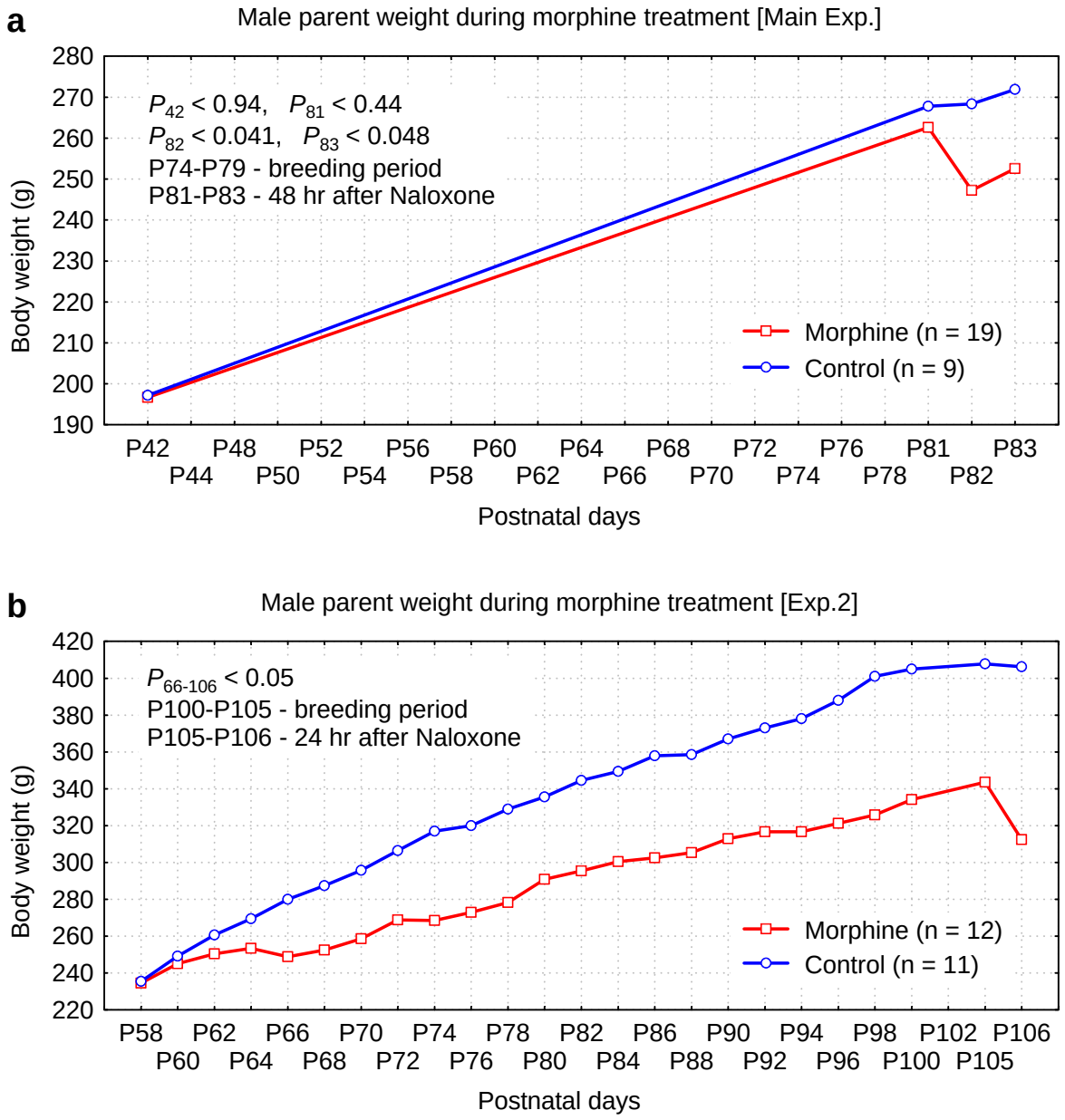



**Figure S48 | Male parent body weight during chronic morphine treatment.** Wistar rats. **a**, Main experiment male parent body weight during chronic morphine treatment (P42-P81) and after naloxone (opiate receptor antagonist) administration (P81), *i.e.* during naloxone-precipitated morphine withdrawal. **b**, Satellite experiment [Exp.2] male parent body weight during chronic morphine treatment (P58-P105) and after naloxone administration (P105). Main and satellite experiments were similar with respect to regular maximum dose (60 mg/kg twice daily), but differed in initial animal age (P42 *vs.* P58) and regularity of morphine treatment (episodes of morphine withdrawal were used in the main experiment, including naloxone-induced morphine withdrawal after the first 7 days of morphine treatment, see Methods, whereas in the satellite experiment animals received morphine with 12-hr intervals). Morphine treatment started at P42 (**a**) produced less impressive differences in body weight than treatment started at P58 (**b**), but all differences in the $F_1$ progeny occured to be greater in the case of early morphine treatment (**a**).

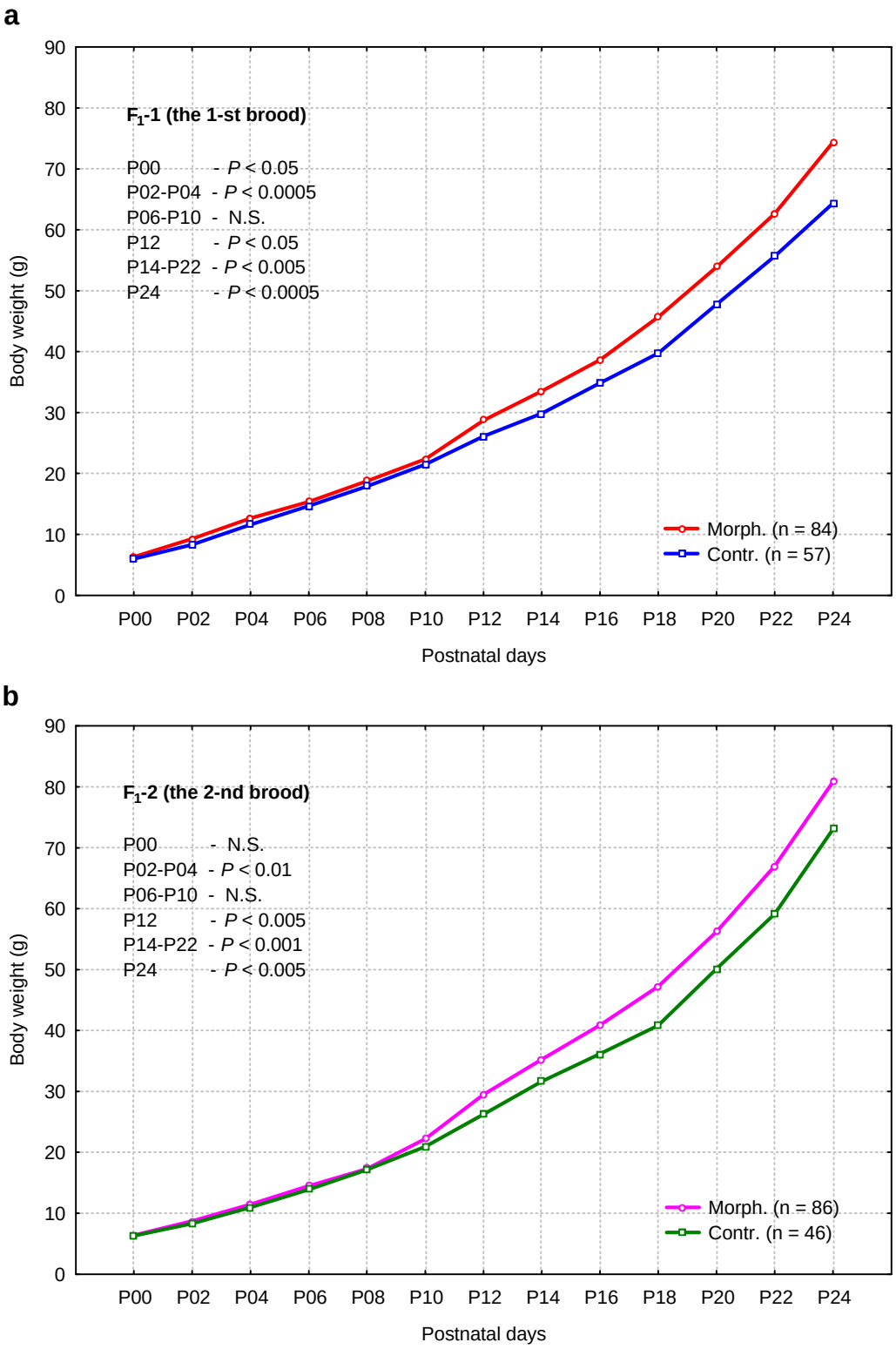



**Figure S49 | Effect of paternal morphine treatment on body weight in $F_1$ progeny.** $F_1$-1 and $F_1$-2 Wistar rats. Mean. **a**, $F_1$ progeny, the first brood. **b**, $F_1$ progeny, the second brood (parents of $F_2$ generation). Note similarity between the first and the second brood.

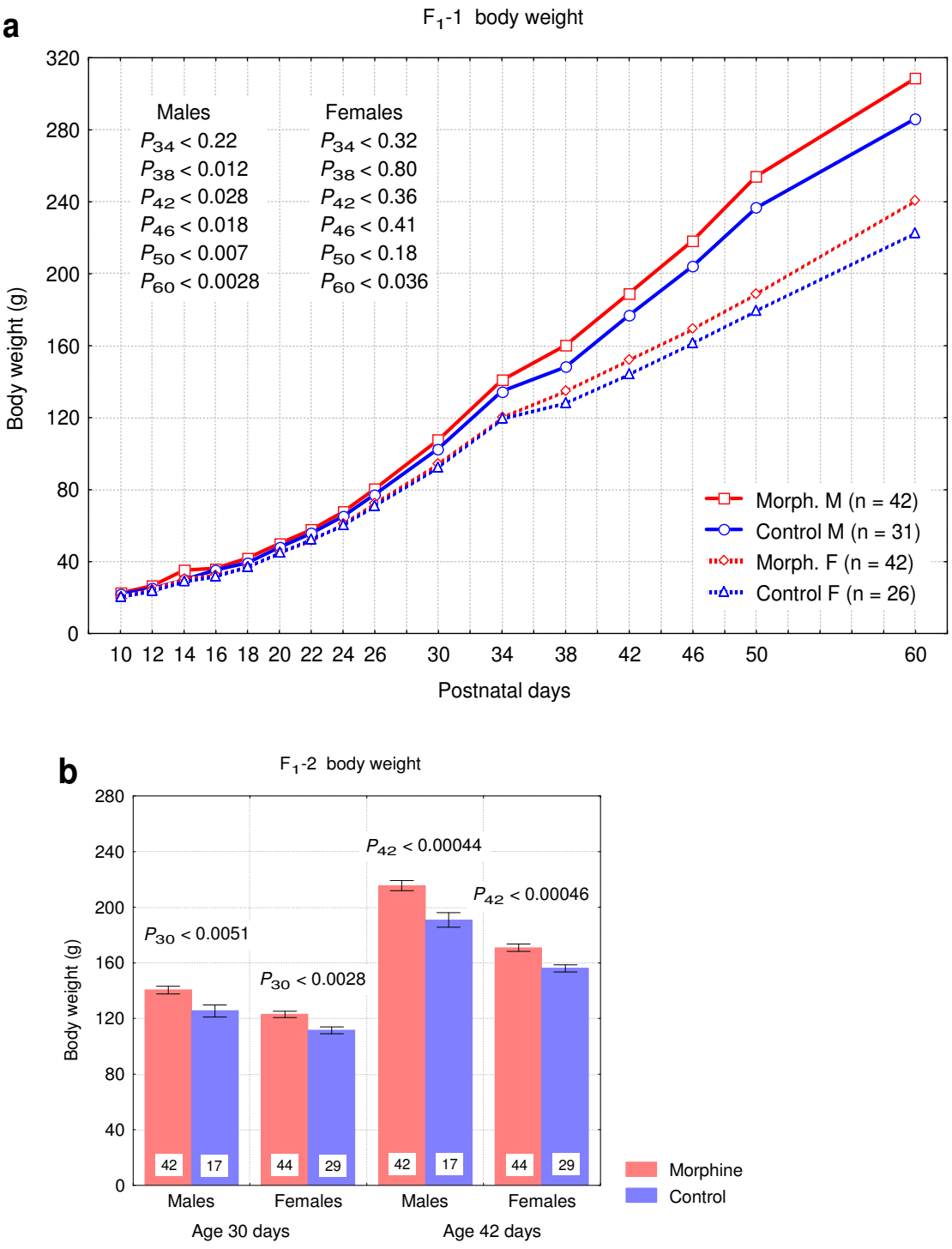



**Figure S50 | Body weight of $F_1$-1 and $F_1$-2 progeny, males and females separately.** This figure represents the same progeny of morphine-treated male rats as Fig. S49. We can see here that the observed increase in body weight (relatively to control) was achieved due to both males and females. In the $F_1$-1 the impact of males on body weight increase was slightly higher than females (**a**). In the $F_1$-2 the roles of males and females in body weight increase were practically equal (**b**). Mann-Whitney U-test. Mean ± SE.

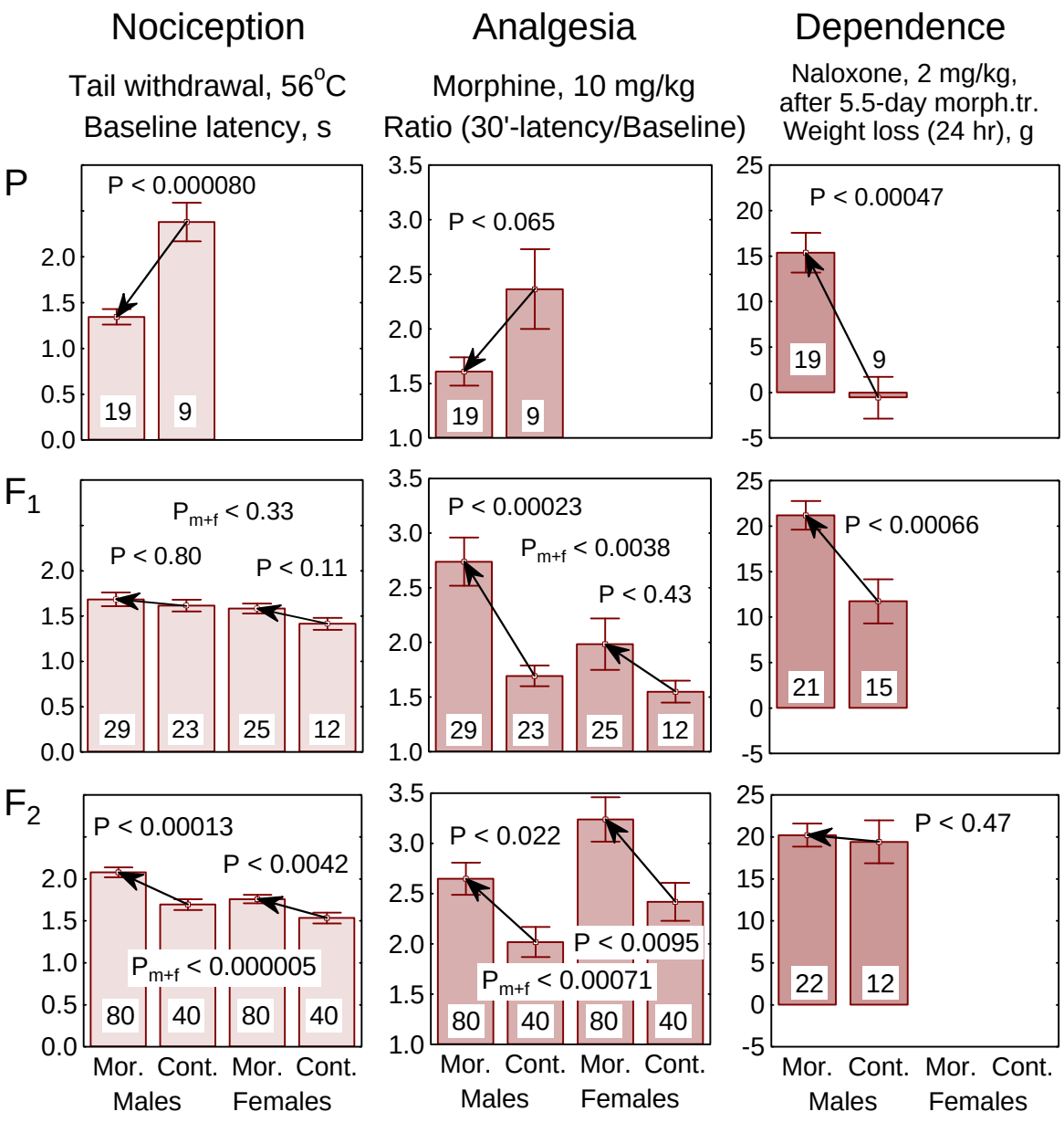



**Figure S51 | Phenotype of morphine-treated male rats and $F_1$-$F_2$ progeny.** Chronically morphine-treated male rats and their untreated progeny (mean ± SE). Pain sensitivity, baseline latency. Morphine analgesia, ratio of tail withdrawal latency, measured 30 min after 10 mg/kg morphine administration, to baseline latency. Naloxone-precipitated weight loss after 5.5-day morphine treatment (in the $F_1$ and $F_2$ offspring) or after 40-day treatment (in the experimental fathers).

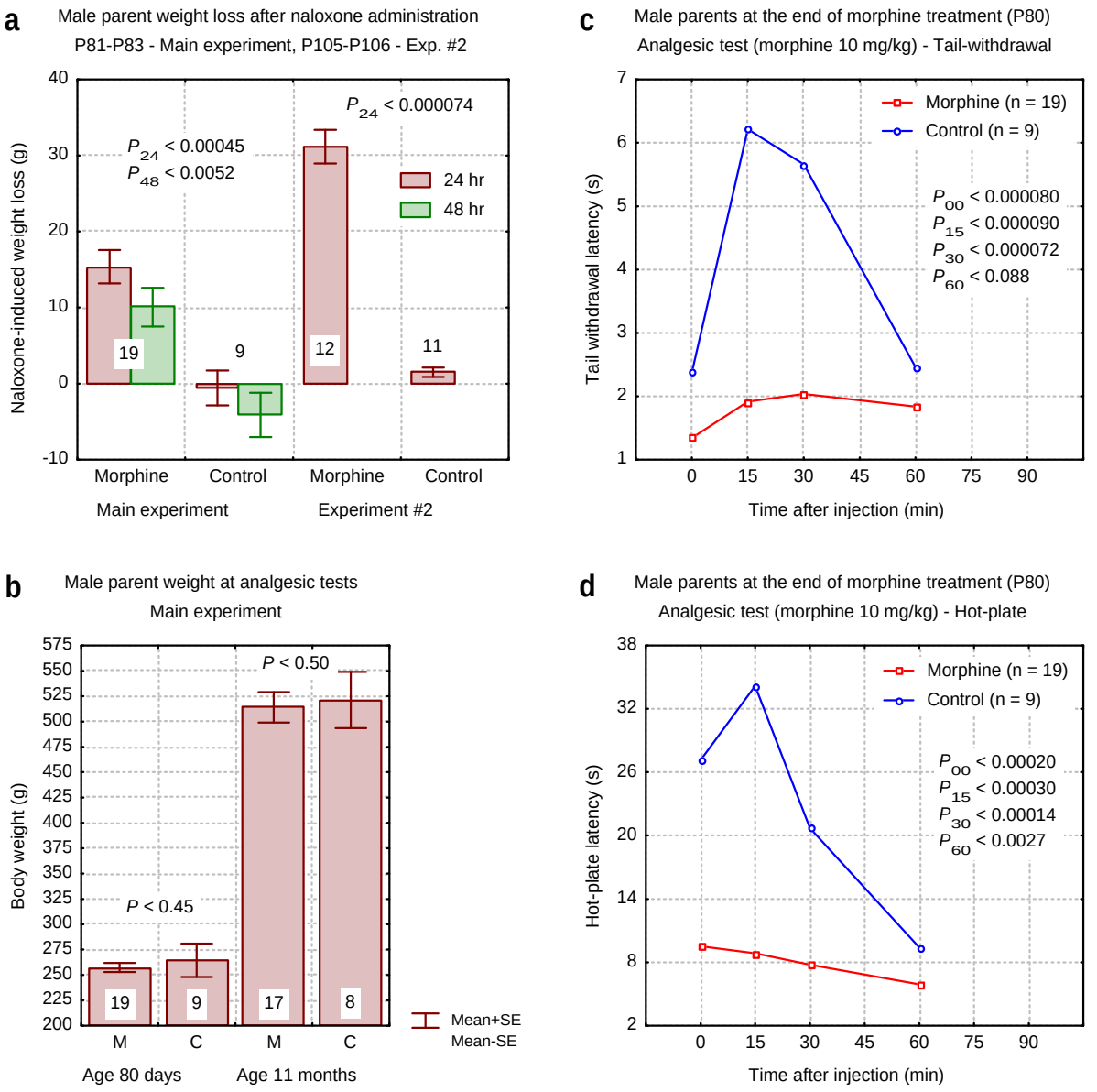



**Figure S52 | Phenotype of male parents after chronic morphine treatment.** **a**, Naloxone-induced weight loss in morphine-treated male parents in the Main experiment (during 24 and 48 hr) and Satellite experiment (#2, during 24 hr) after the end of chronic morphine treatment. Note that experimental, but not control, rats have lost body weight after naloxone injection. **b**, Main experiment male parent body weight during analgesic tests (these tests were applied twice: at the age of 80 days and at 11 months). Note that experimental and control males had identical body weight. Tail-withdrawal (**c**) and hot-plate (**d**) latencies immediately after chronic morphine treatment (age 80 days): before morphine injection (time 0') and after morphine 10 mg/kg i.p. administration (time 15', 30', 60') Note that 10 mg/kg did not induce any analgesic effect in the chronically treated animals. Note also that basal latencies (time 0') are decreased in the experimental animals in both tail-withdrawal and hot-plate tests.

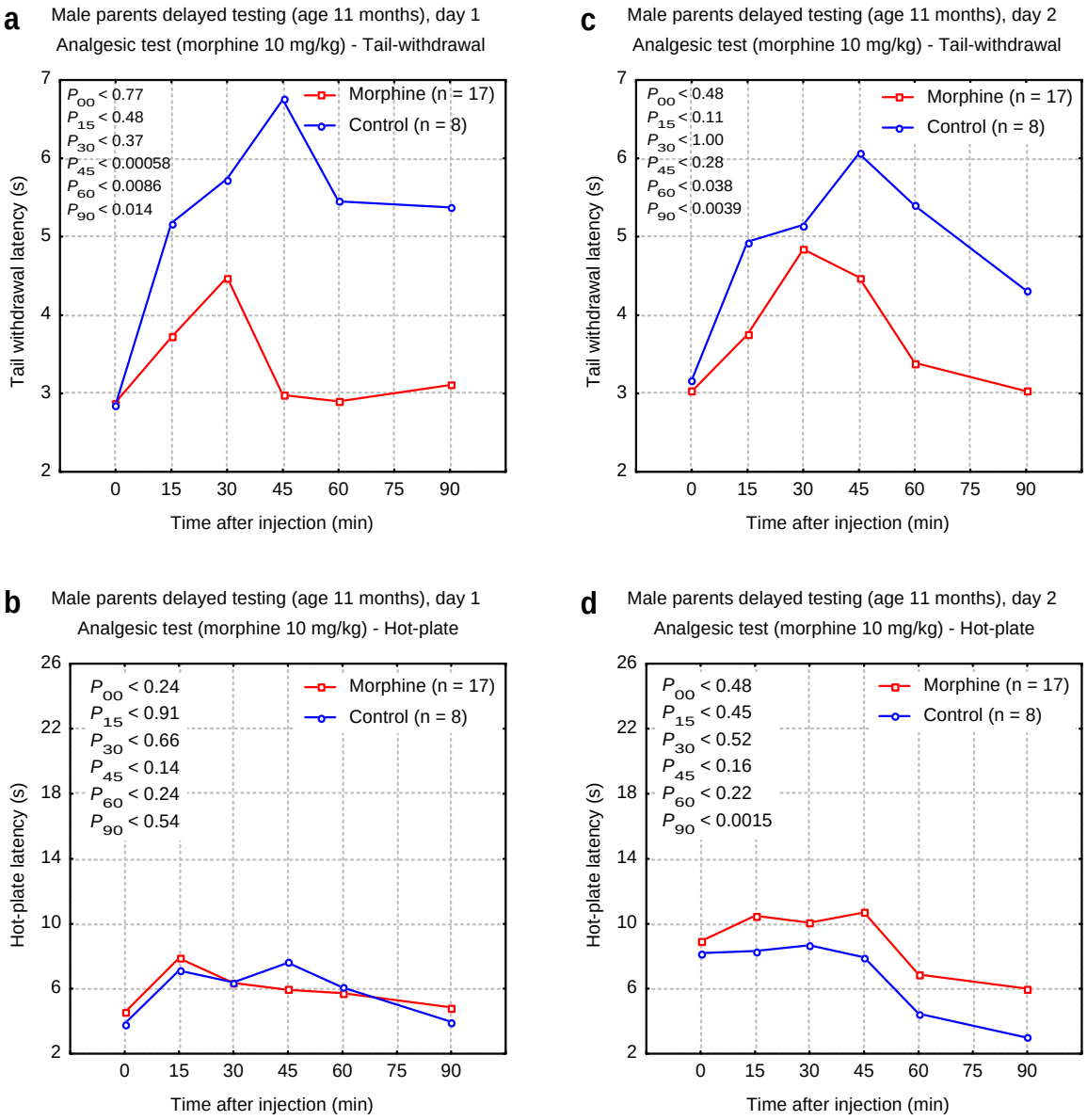



**Figure S53 | Male parents, tail-withdrawal and hot-plate tests at the age of 11 months.** Tail-withdrawal (**a, c**) and hot-plate (**b, d**) tests in chronically morphine-treated male parents about 8 months after the end of chronic morphine treatment. Here analgesic effect of 10 mg/kg morphine injection was measured twice with 24 hr time interval, see day 1 (**a-b**) and day 2 (**c-d**). Note that basal pain sensitivity is absolutely normal (**a-d**, time 0') and morphine does induce detectable analgesic effect in the experimental animals (**a-b**, time 15', 30'), but this analgesic effect is still diminished at the time points 45', 60', 90' at day 1 (**a**) and time 60', 90' at day 2 (**b**). In the hot-plate test some significance was observed only at day 2, time 90', and here due to some reason analgesic effect in the experimental rats was higher than in controls (**d**).

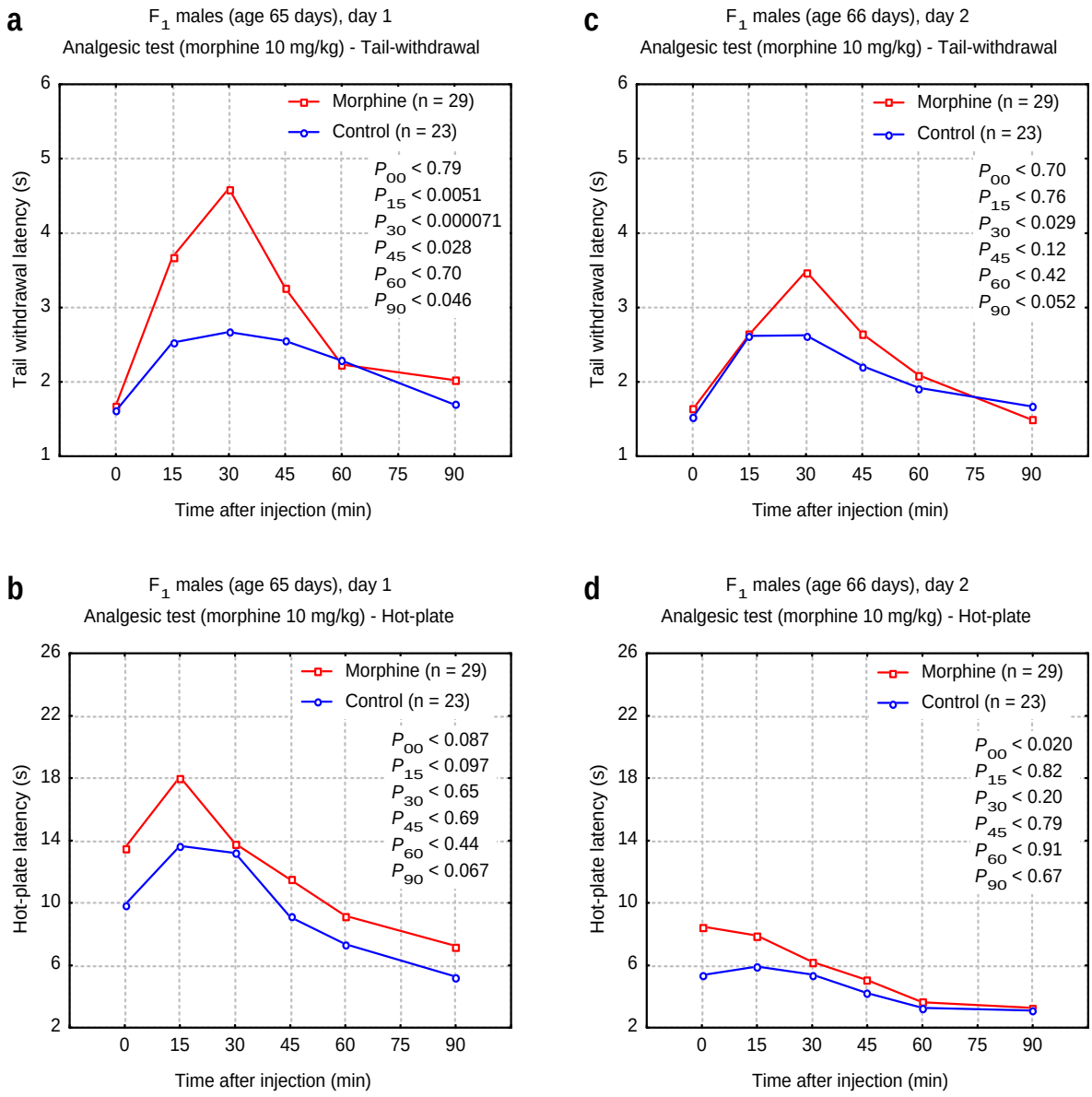



**Figure S54 | $F_1$ males, tail-withdrawal and hot-plate tests, day 1 and day 2.** Latency before morphine injection (time 0') and latencies after morphine 10 mg/kg i.p. administration (time 15', 30', 45', 60', 90') in the $F_1$ **male** descendants of chronically morphine-treated males and drug-naïve females in tail-withdrawal (**a, c**) and hot-plate (**b, d**) tests in the case of the 1st morphine administration (**a-b**) and the 2nd one done 24 hr later (**c-d**). Note normal basal pain sensitivity (**a, c**, time 0') and enhanced analgesic effect of morphine during the 1st day in tail-withdrawal test (**a**). In control animals in tail-withdrawal test analgesic effect is practically the same during day 1 (**a**, blue) and day 2 (**c**, blue), whereas in the experimental rats it is dramatically decreased in day 2 (**c**, red) in comparison with day 1 (**a**, red), it is decreased up to control group level (**a, c**). In hot-plate test we see nice learning effect in all rats – compare day 1 (**b**) and day 2 (**d**).

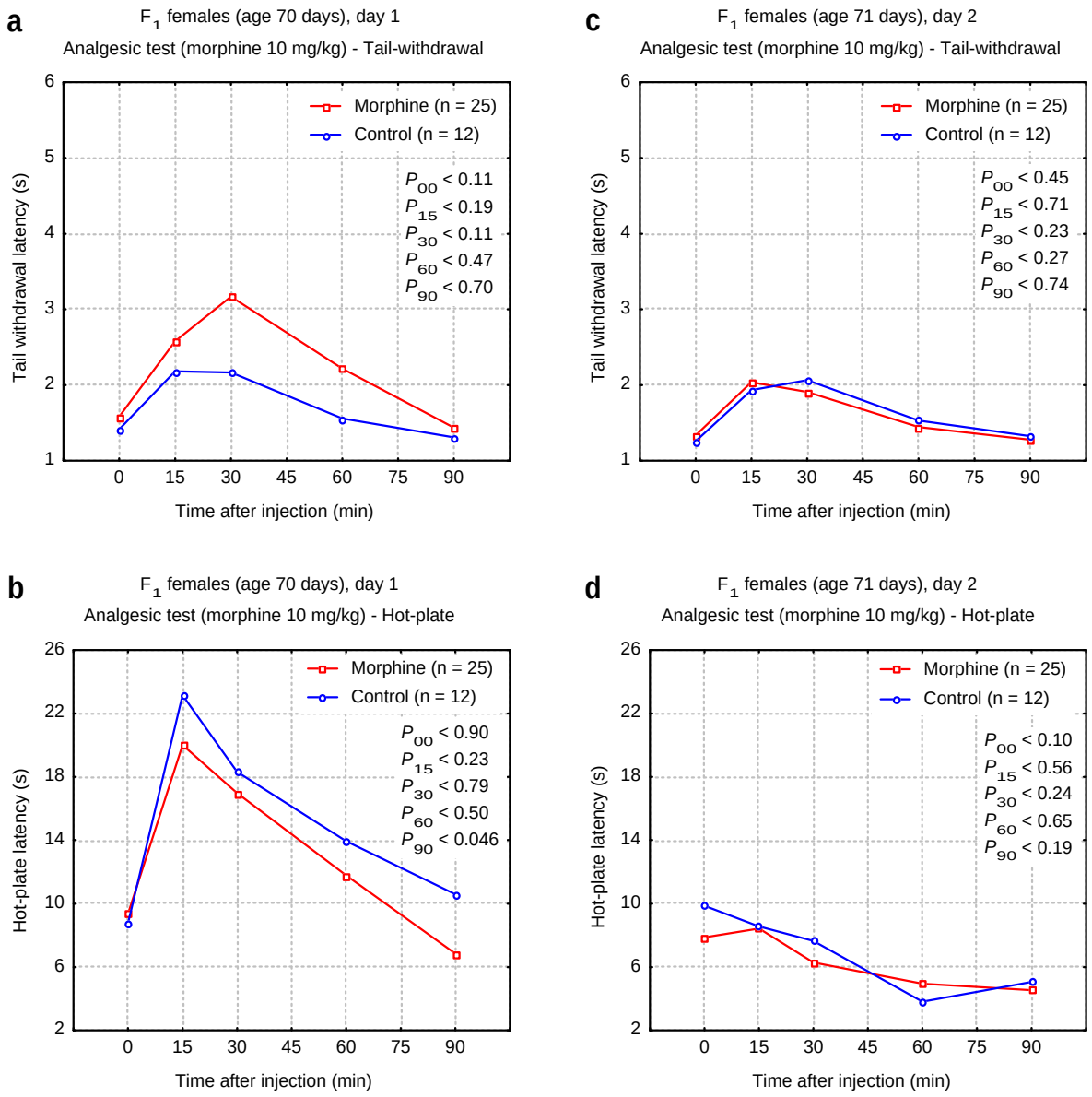



**Figure S55 | $F_1$ females, tail-withdrawal and hot-plate tests, day 1 and day 2.** Latency before morphine injection (time 0') and latencies after morphine 10 mg/kg i.p. administration (time 15', 30', 60', 90') in the $F_1$ **female** descendants of chronically morphine-treated males and drug-naïve females in tail-withdrawal (**a, c**) and hot-plate (**b, d**) tests in the case of the 1st morphine administration (**a-b**) and the 2nd one done 24 hr later (**c-d**). Note normal basal pain sensitivity (**a, c**, time 0') and very slightly enhanced analgesic effect during day 1 (**a**) – it is non-significant and looks like analgesic effect in males during day 2. In females during day 2 analgesic effect is absolutely normal (**c**). Thus, great gender-related differences can be seen in the $F_1$ experimental animals concerning sensitivity to morphine-induced analgesia in tail-withdrawal test. In hot-plate test we see in all females nice learning curves, as in males – compare day 1 (**b**) and 2 (**d**).

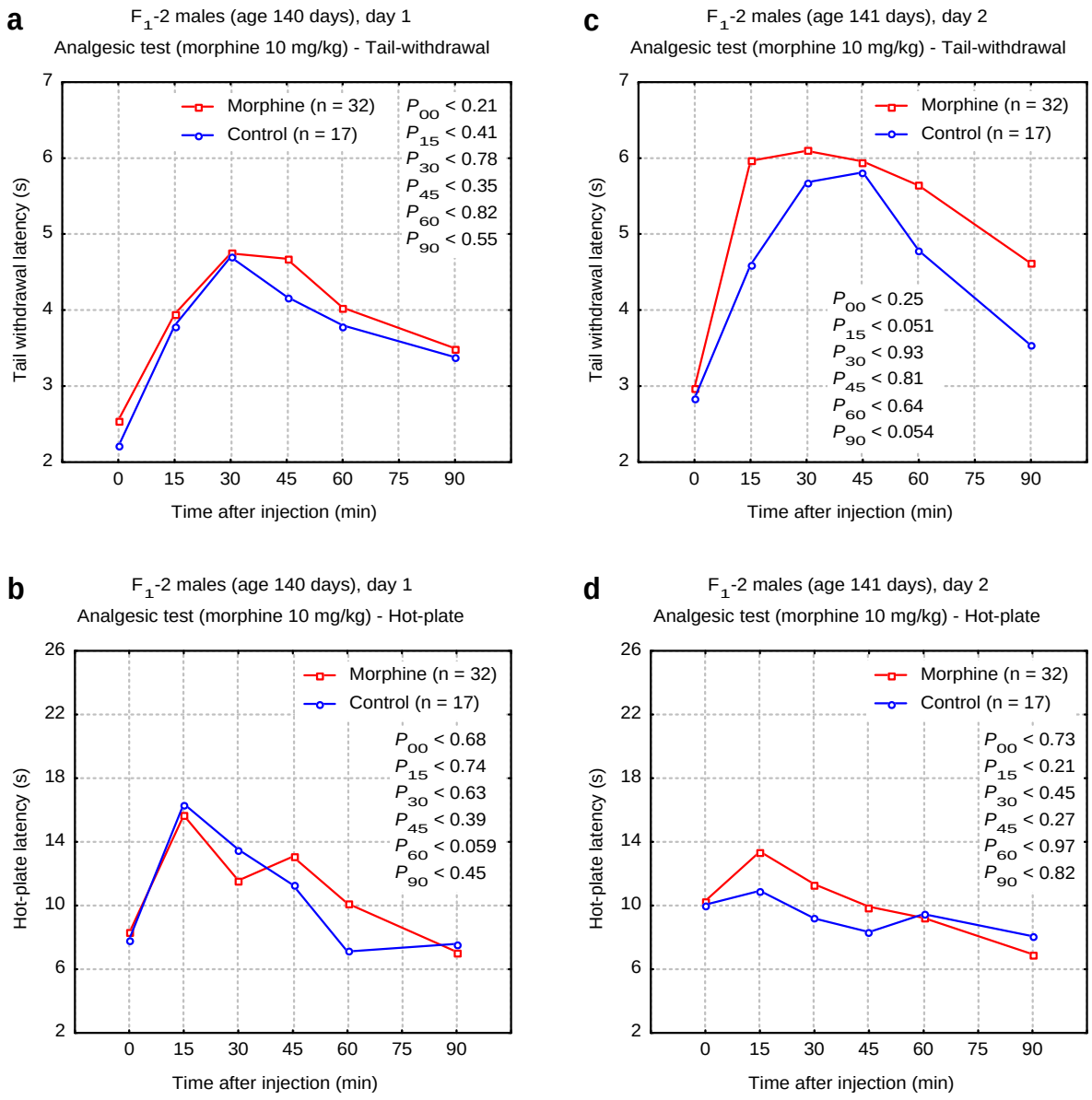



**Figure S56 | $F_1$-2 males, tail-withdrawal and hot-plate tests, day 1 and day 2.** Latency before morphine injection (time 0') and latencies after morphine 10 mg/kg i.p. administration (time 15', 30', 45', 60', 90') in the $F_1$-2 (**second brood**) **male** descendants of chronically morphine-treated males and drug-naïve females in tail-withdrawal (**a, c**) and hot-plate (**b, d**) tests in the case of the 1st morphine administration (**a-b**) and the 2nd one done 24 hr later (**c-d**). These $F_1$-2 males are male parents of $F_2$ generation and they were tested here many days after breeding period, at the age of 138 days. We can not see here any difference between experimental and control groups. The origin of difference between day 1 (**a**) and day 2 (**c**) in tail-withdrawal test in all latencies, including basal, is unclear and it looks for me as decreased water temperature during day 2, however I have not appropriate remarks in my records and temperature is described as 56ºC during both day 1 and day 2.

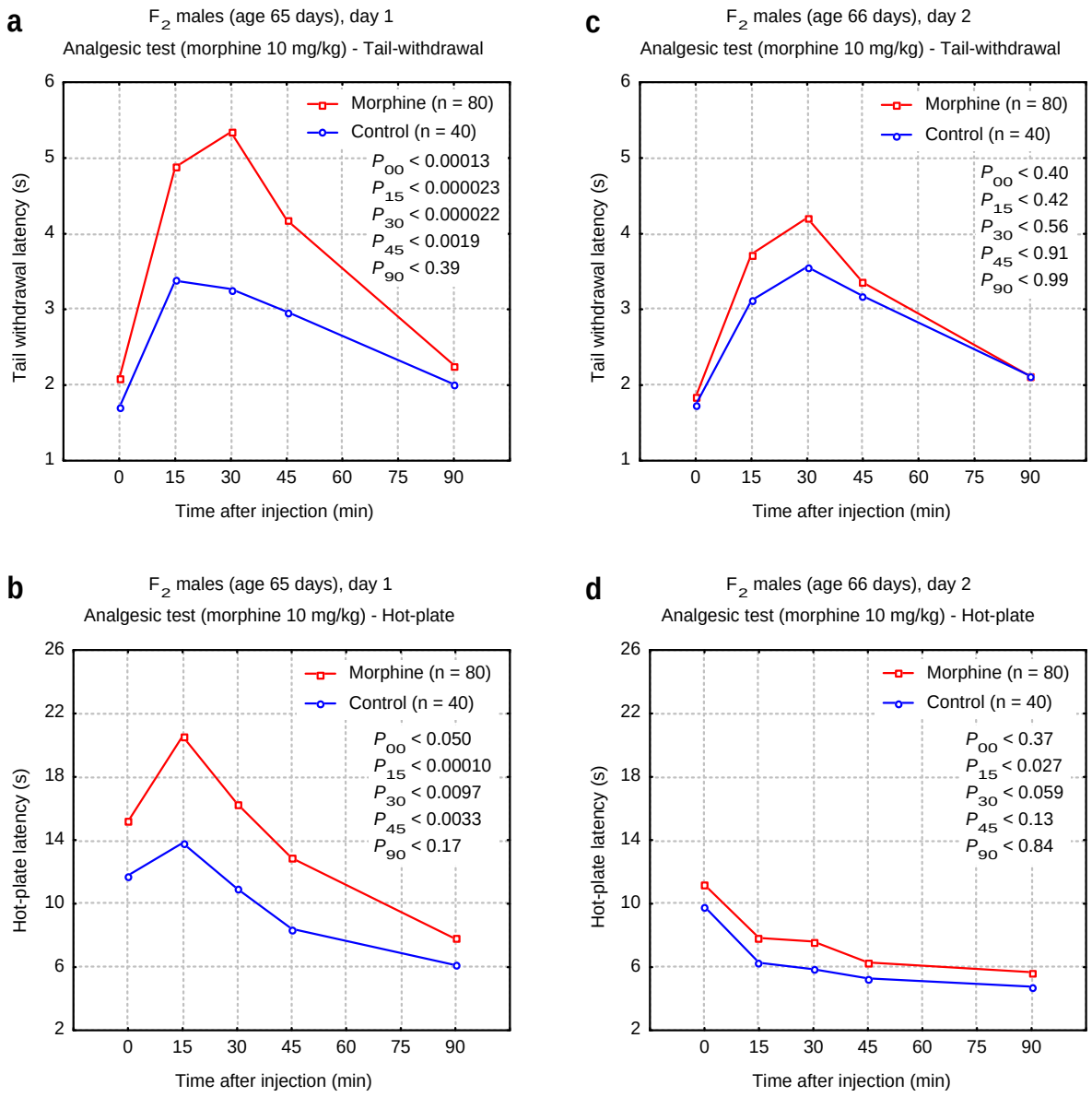



**Figure S57 | $F_2$ males, tail-withdrawal and hot-plate tests, day 1 and day 2.** Latency before morphine injection (time 0') and latencies after morphine 10 mg/kg i.p. administration (time 15', 30', 45', 90') in the $F_2$ **male** descendants of chronically morphine-treated males and drug-naïve females in tail-withdrawal (**a, c**) and hot-plate (**b, d**) tests in the case of the 1st morphine administration (**a-b**) and the 2nd one done 24 hr later (**c-d**). Note dramatically enhanced analgesic effect during the 1st day in both tail-withdrawal (**a**) and hot-plate (**c**) tests and practically no differences during the 2nd day (**c, d**). Note changed basal pain sensitivity during day 1 only (**a-b**). Exactly like in the $F_1$ males, in the $F_2$ males enhanced sensitivity to morphine-induced analgesia disappears after the 1st morphine injection during the next 24 hours. In the $F_2$ generation it is true with respect to both tail-withdrawal and hot-plate tests.

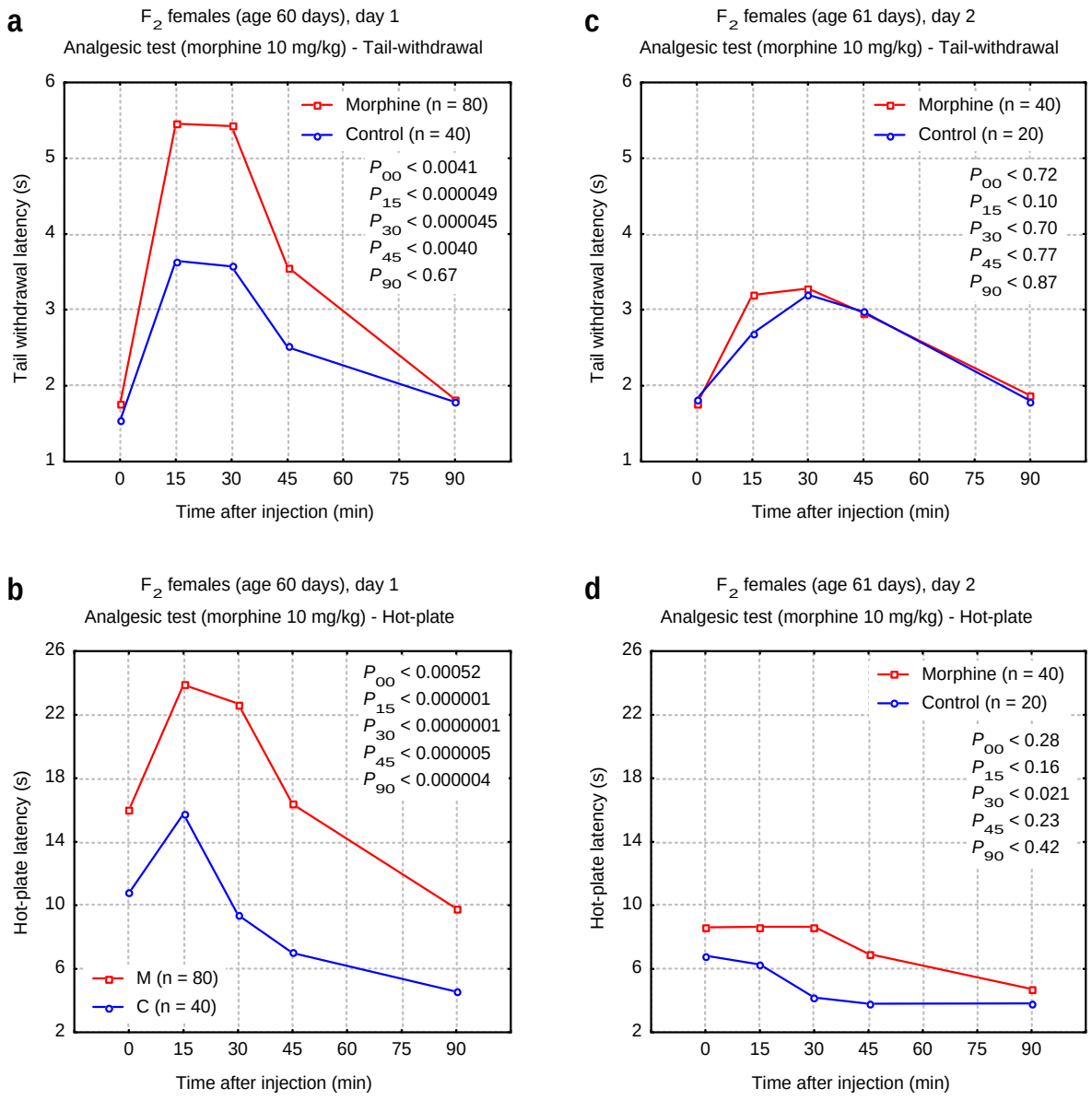



**Figure S58 | $F_2$ females, tail-withdrawal and hot-plate tests, day 1 and day 2.** Latency before morphine injection (time 0') and latencies after morphine 10 mg/kg i.p. administration (time 15', 30', 45', 90') in the $F_2$ **female** descendants of chronically morphine-treated males and drug-naïve females in tail-withdrawal (**a, c**) and hot-plate (**b, d**) tests in the case of the 1st morphine administration (**a-b**) and the 2nd one done 24 hr later (**c-d**). Note, like in $F_2$ males, dramatically enhanced analgesic effect in both tail-withdrawal (**a**) and hot-plate (**b**) tests and, similar to $F_2$ males, complete normalization of sensitivity to morphine induced analgesia detected at the 2nd day (**c**) in tail-withdrawal test. Hot-plate test works here more like learning paradigm which involves long-term memory formation (**b, d**). Note also differences in basal pain sensitivity during the 1st, but not during the 2nd day (time 0').

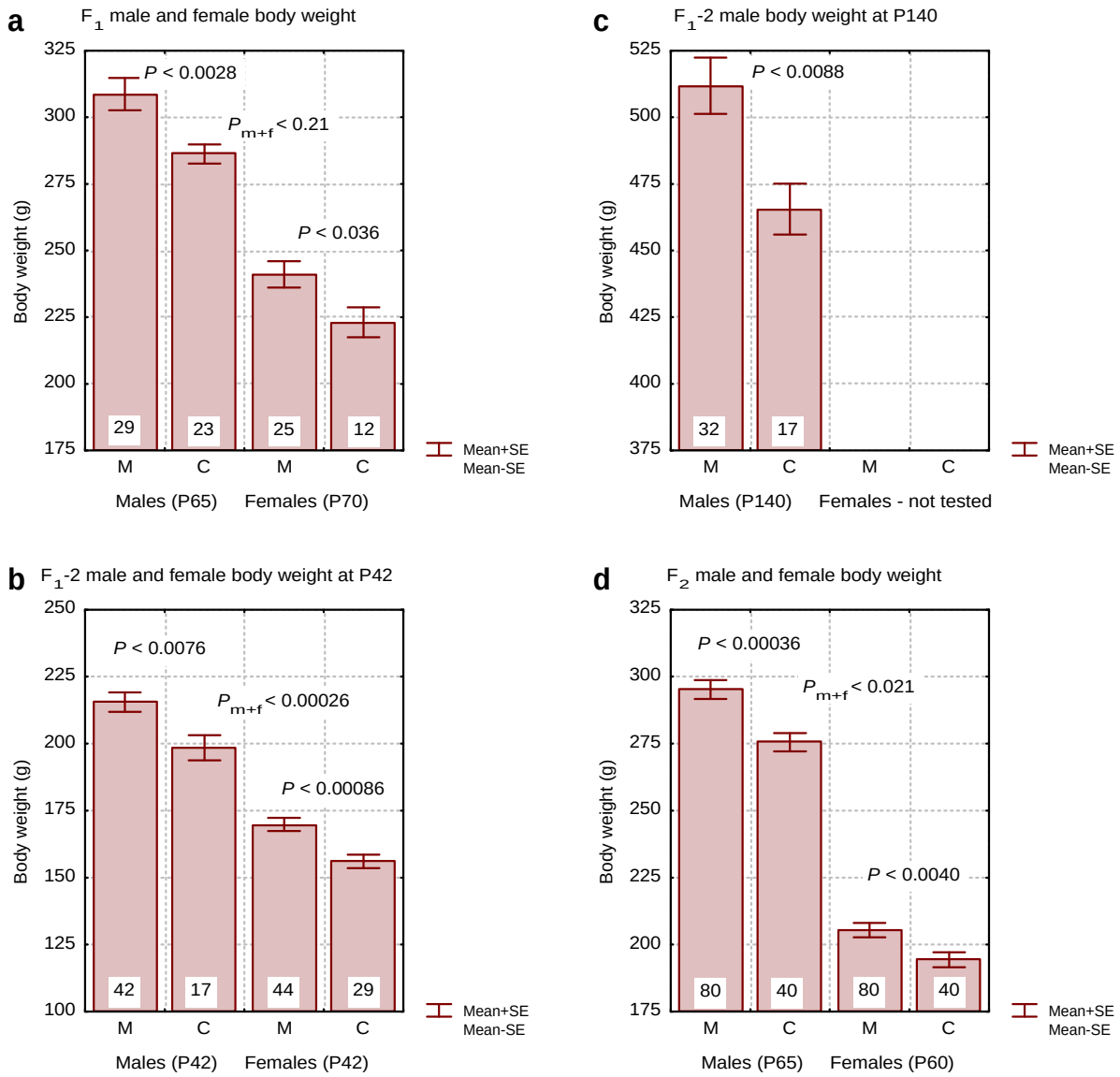



**Figure S59 | $F_1$ and $F_2$ progeny body weight during tail-withdrawal and hot-plate tests.** Body weight of descendants of chronically morphine-treated males and drug-naïve females: $F_1$ (**a**), $F_1$-2 (**b-c**) and $F_2$ (**d**). M – experimental group ("morphine"), C – control. In (**a**) and (**c-d**) body weight was measured just before the first analgesic test. However $F_1$-2 females were not tested and we show their weight at the age of 42 days (**b**). Note that in the $F_1$, $F_1$-2 and $F_2$ experimental animals have increased body weight with respect to control (both males and females). However it can not be the main reason of enhanced analgesic effect of morphine in these animals, because in the $F_1$-2 males (**c**) body weight was also increased, but analgesic effect was absolutely normal.

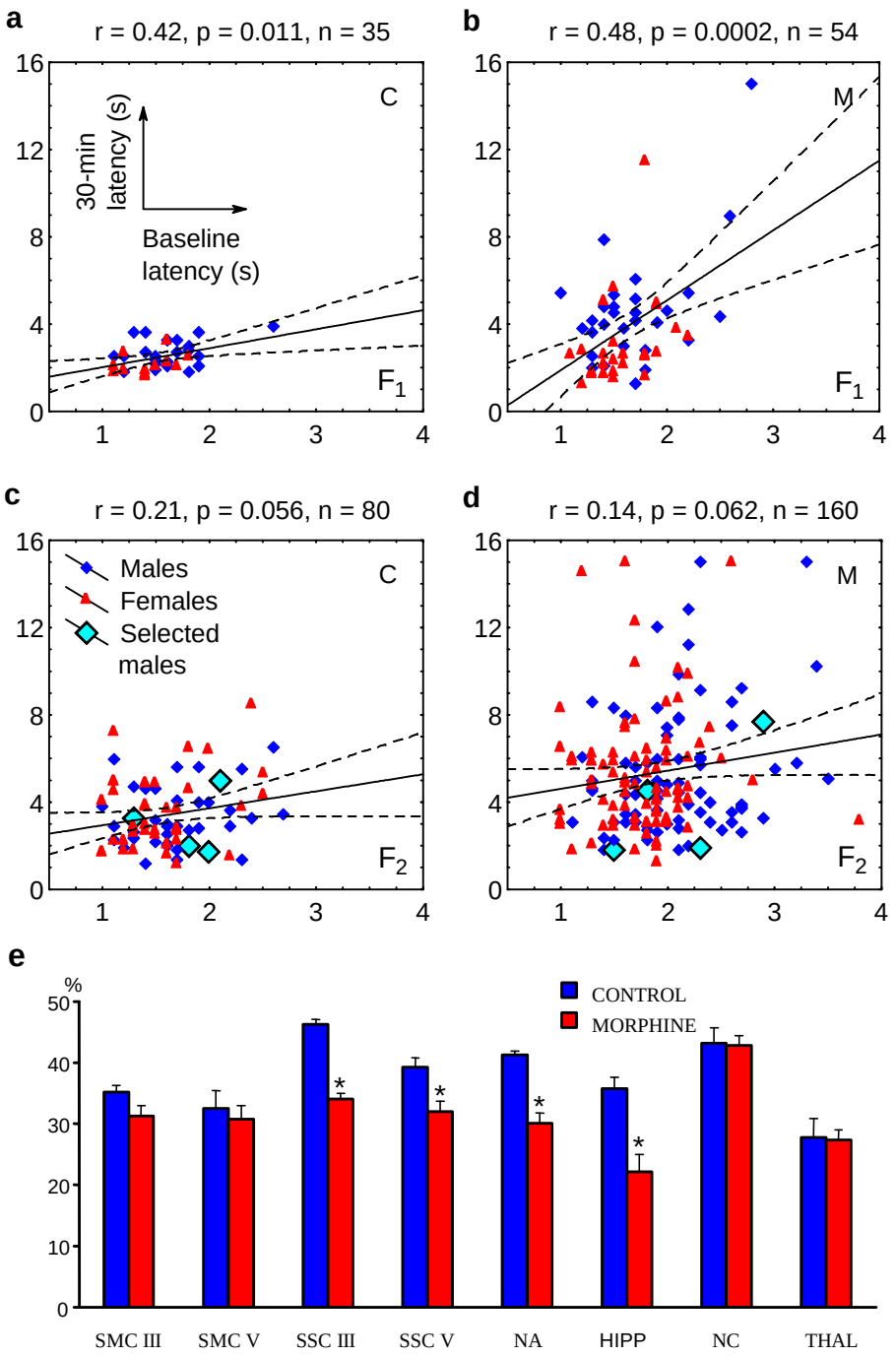



**Figure S60 | Correlations in the $F_1$-$F_2$ descendants of morphine-treated male rats.** **a-d**, Correlations between pain sensitivity and morphine analgesia. **a**, **c**, Control. **b**, **d**, Experimental group. **e**, Synaptophysin (p38) distribution in the selected $F_2$ males (fluorescent area, %). SMC III, somatomotor cortex, layer 3; SMC V, somatomotor cortex, layer 5; SSC III, somatosensory cortex, layer 3; SSC V, somatosensory cortex, layer 5; NA, nucleus accumbens; HIPP, hippocampus; NC, nucleus caudatus; THAL, ventrolateral thalamic nuclei. Mean + SE. Asterisk, $P < 0.05$.

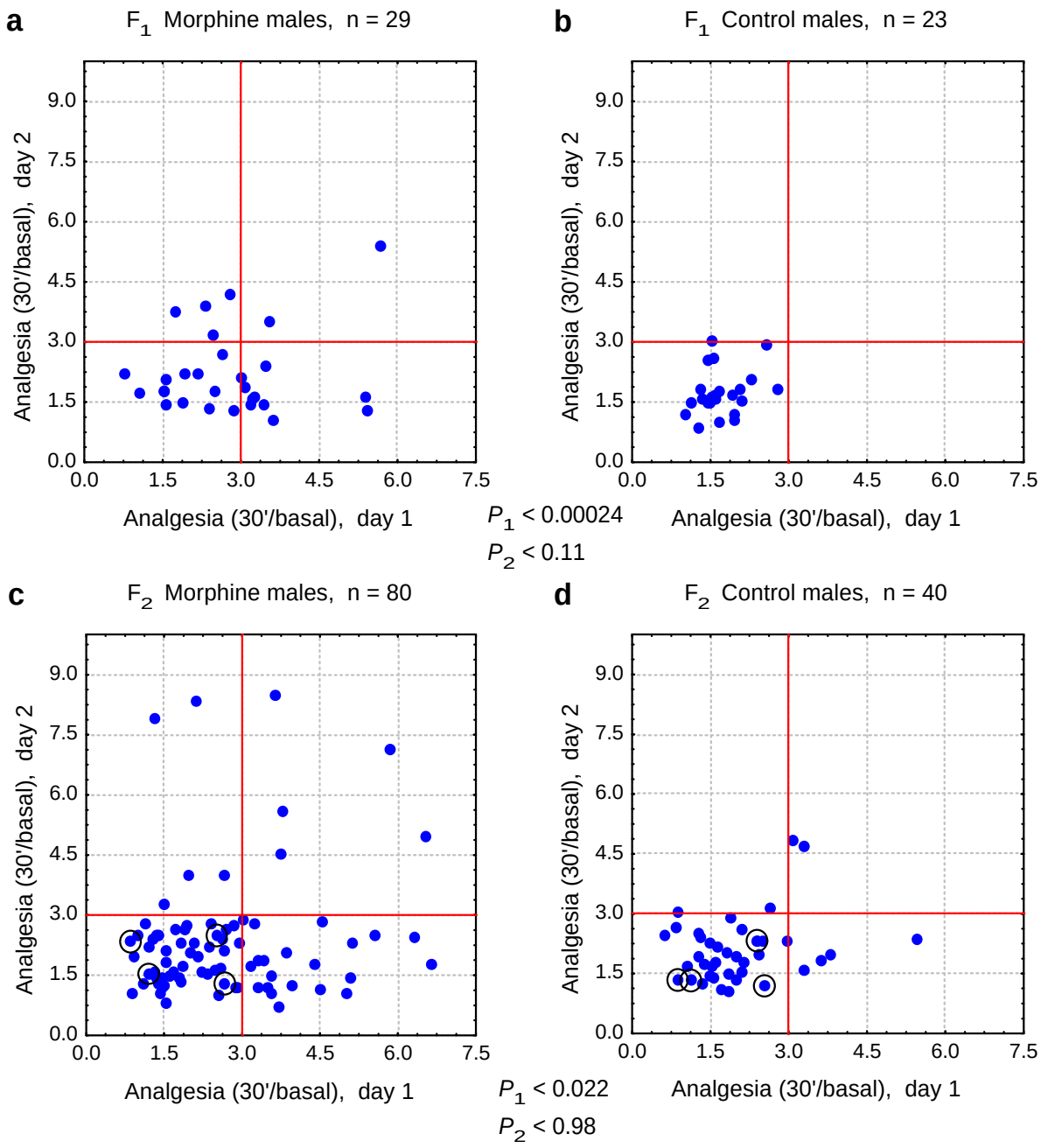



**Figure S61 | Correlation between 1st and 2nd day analgesic effect in $F_1$ and $F_2$ males.** Tail-withdrawal test, **male** descendants of chronically morphine-treated males and drug-naïve females. Ratio of tail-withdrawal latency 30 min after morphine 10 mg/kg i.p. administration to tail-withdrawal latency measured before morphine injection for day 1 (abscissa) and day 2 (ordinate). **a**, Experimental $F_1$ males. **b**, Control $F_1$ males. **c**, Experimental $F_2$ males. **d**, Control $F_2$ males. Red lines are provisional lines of demarcation between normal and abnormal phenotypes. $P_1$ and $P_2$ are significance levels in accordance with Mann-Whitney U-test for day 1 and day 2, respectively. Four dots in small circles (**c-d**) show rats selected for synaptophysin distribution analysis in the brain (Fig. S60e) – these rats had normal phenotype in tail-withdrawal test.

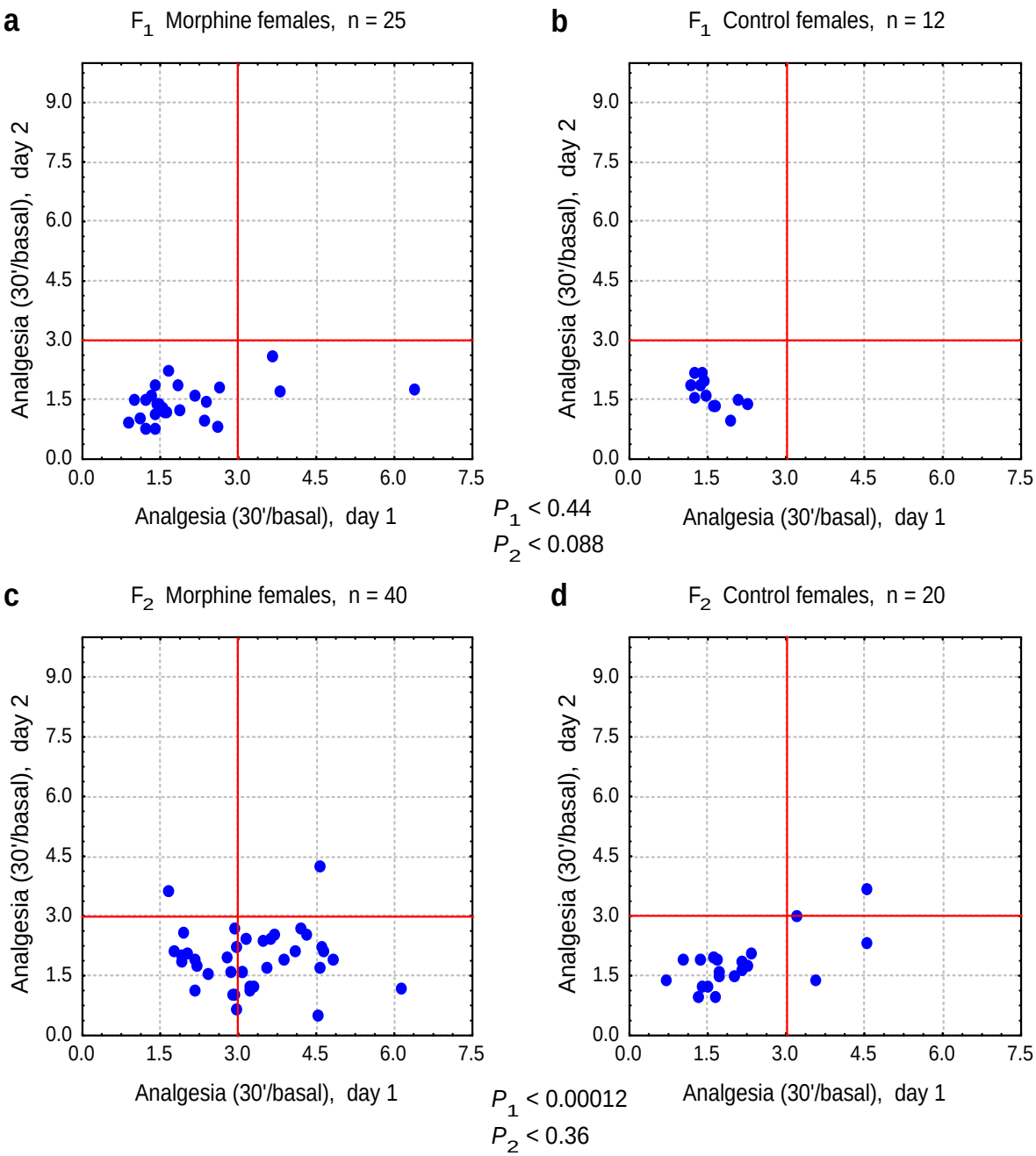



**Figure S62 | Correlation between 1st and 2nd day analgesic effect in $F_1$ and $F_2$ females.** Tail-withdrawal test, **female** descendants of chronically morphine-treated males and drug-naïve females. Ratio of tail-withdrawal latency 30 min after morphine 10 mg/kg i.p. administration to tail-withdrawal latency measured before morphine injection for day 1 (abscissa) and day 2 (ordinate). **a**, Experimental $F_1$ females. **b**, Control $F_1$ females. **c**, Experimental $F_2$ females. **d**, Control $F_2$ females. Red lines are provisional lines of demarcation between normal and abnormal phenotypes. $P_1$ and $P_2$ are significance levels in accordance with Mann-Whitney U-test for day 1 and day 2, respectively. Note that females have shown enhanced analgesic effect only during the first day and only in the $F_2$ generation (**c**, abscissa).

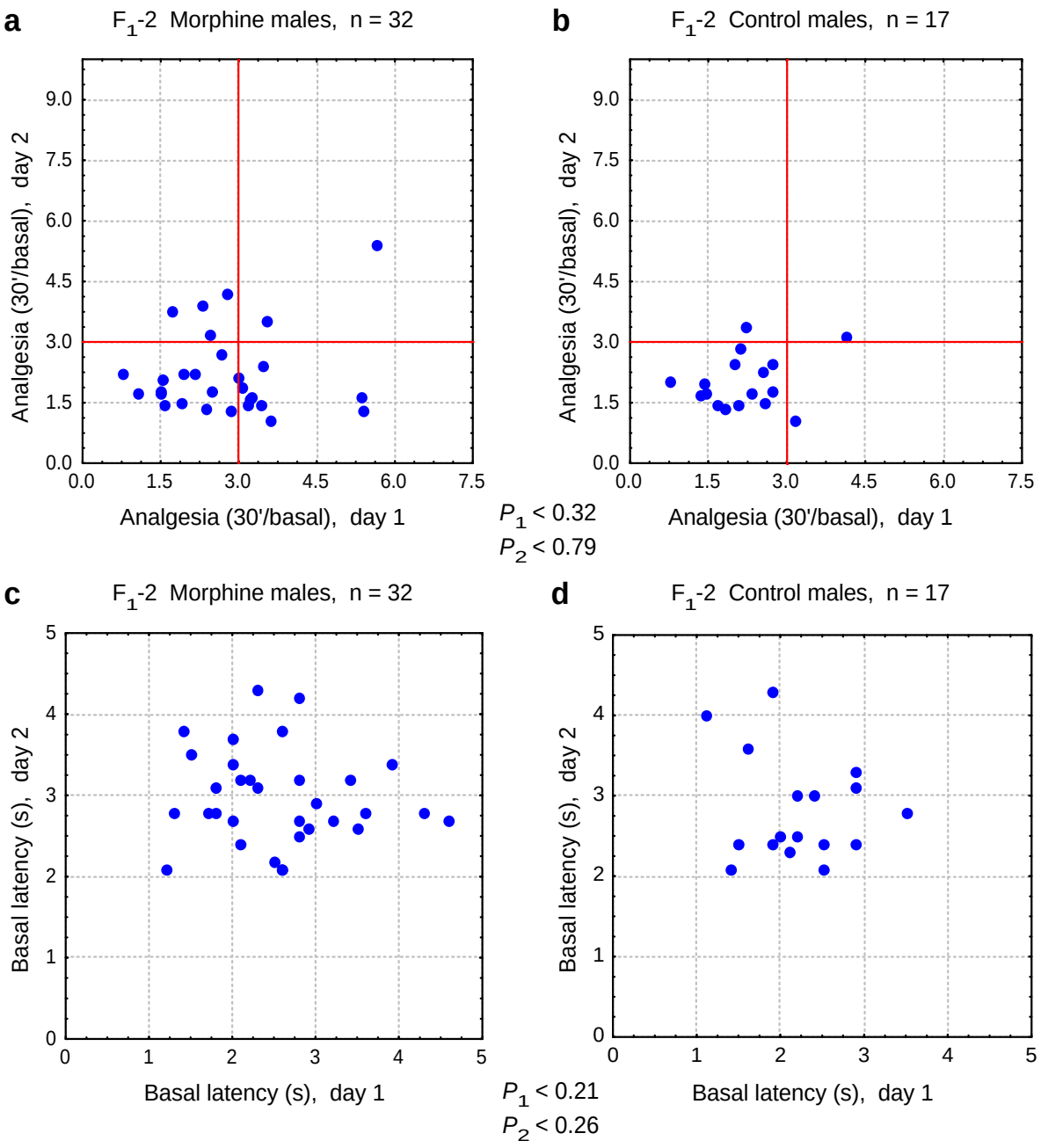



**Figure S63 | Correlation between 1st and 2nd day analgesia, nociception in $F_1$-2 males.** Tail-withdrawal test, $F_1$-2 (**second brood**) **male** descendants of chronically morphine-treated males and drug-naïve females. **a-b**, Ratio of tail-withdrawal latency 30 min after morphine 10 mg/kg i.p. administration to tail-withdrawal latency before morphine administration for day 1 (abscissa) and day 2 (ordinate). **c-d**, Tail-withdrawal latency measured before morphine administration (*i.e.* basal latency, nociception) for day 1 (abscissa) and day 2 (ordinate). **a**, **c**, Experimental $F_1$-2 males. **b**, **d**, Control $F_1$-2 males. Note that phenotype of $F_1$-2 males with respect to nociception and morphine analgesia looks normal at the particular age (138 days). Their direct descendants ($F_2$) have shown decreased nociception and enhanced morphine analgesia (at the age of 60-65 days).

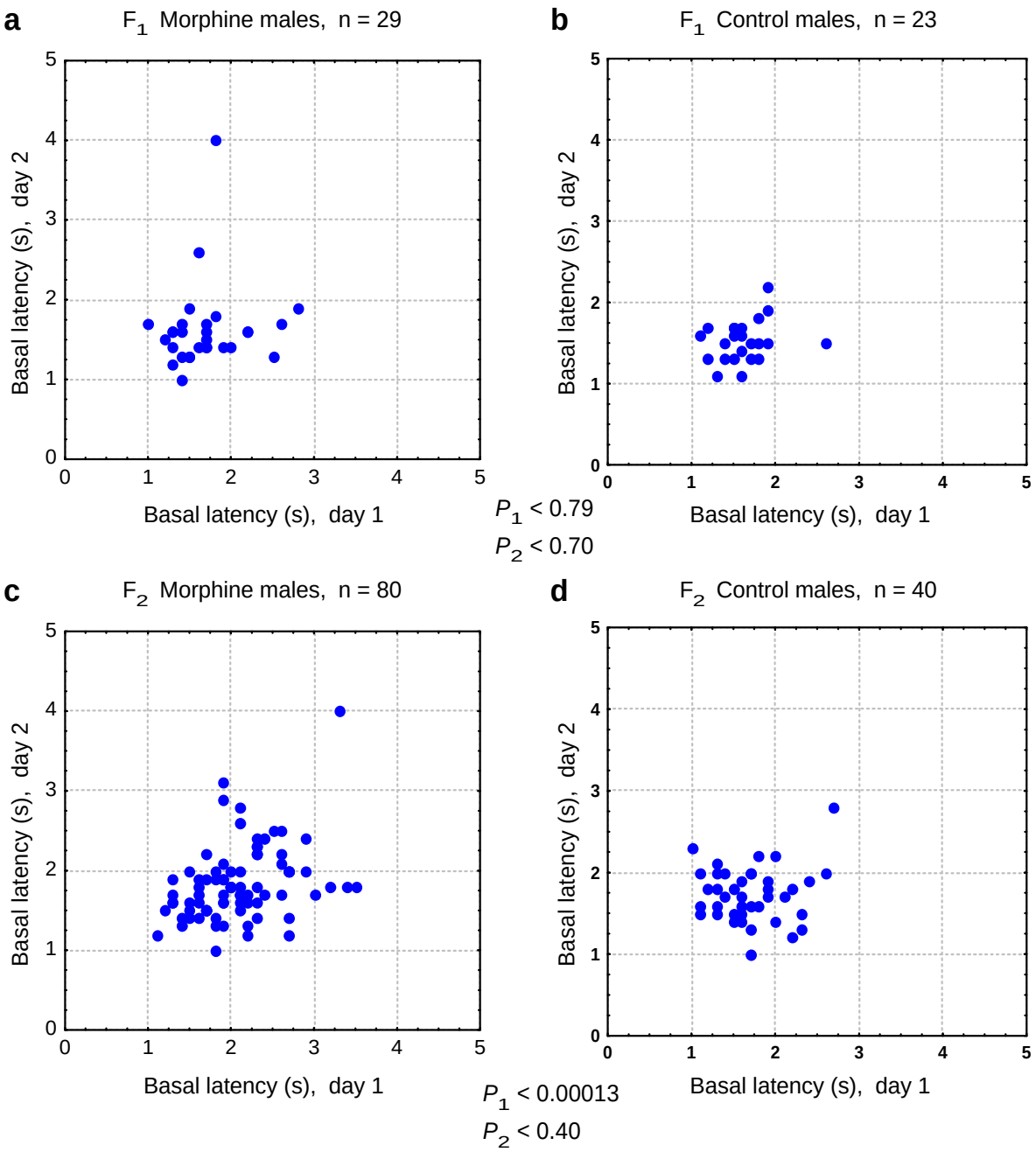



**Figure S64 | Correlation between 1st and 2nd day nociception in $F_1$ and $F_2$ males.** Tail-withdrawal test, **male** descendants of chronically morphine-treated males and drug-naïve females. Tail-withdrawal latency measured before morphine administration (*i.e.* basal latency, nociception) for day 1 (abscissa) and day 2 (ordinate). **a**, Experimental $F_1$ males. **b**, Control $F_1$ males. **c**, Experimental $F_2$ males. **d**, Control $F_2$ males. $P_1$ and $P_2$ are significance levels in accordance with Mann-Whitney U-test for day 1 and day 2, respectively. Note that basal pain sensitivity is absolutely normal in $F_1$ males during both day 1 (**a**) and day 2 (**b**). And only in the $F_2$ generation, only during the 1st day, basal tail-withdrawal latency was significantly increased (**c**, abscissa).

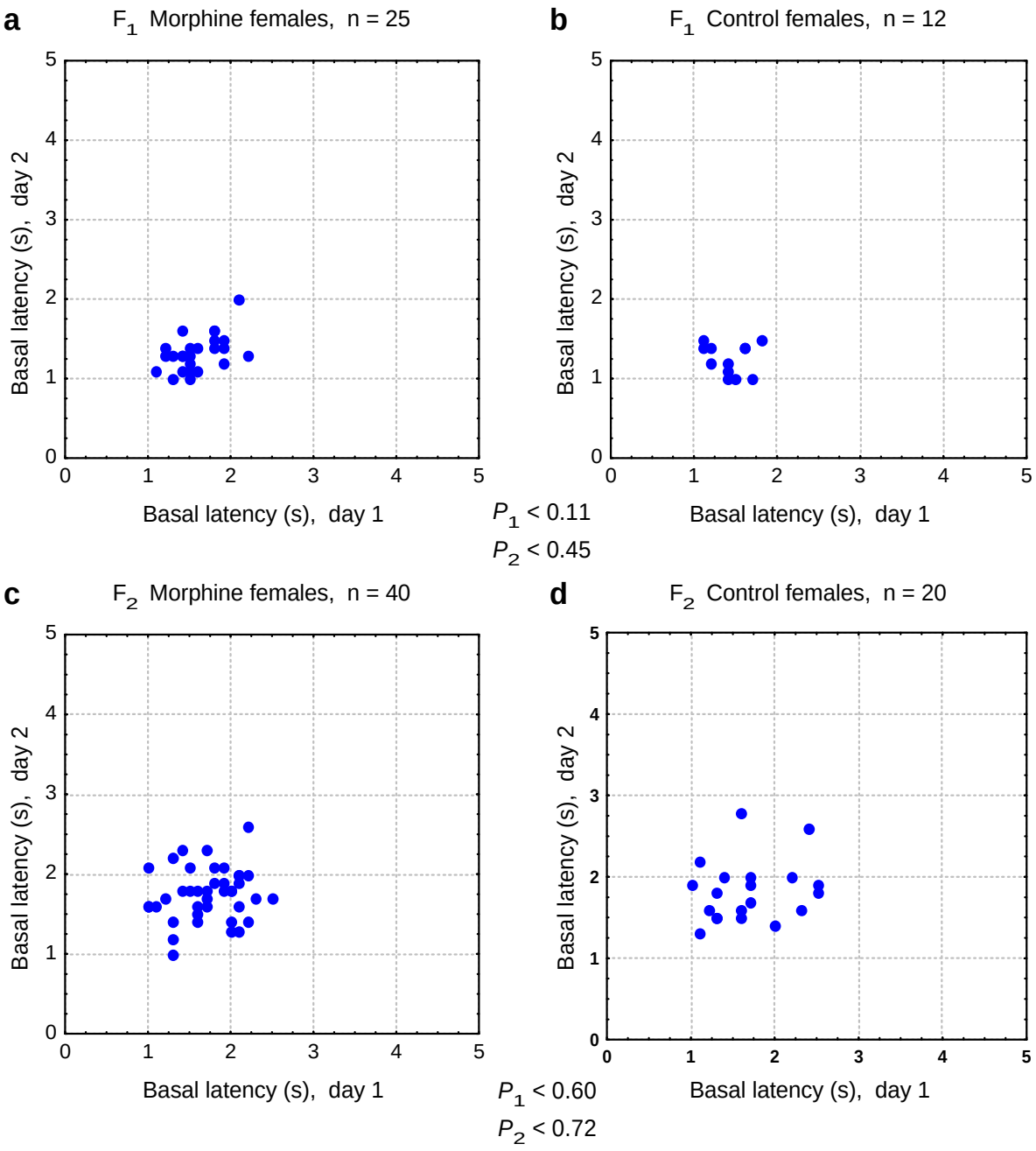



**Figure S65 | Correlation between 1st and 2nd day nociception in $F_1$ and $F_2$ females.** Tail-withdrawal test, **female** descendants of chronically morphine-treated males and drug-naïve females. Tail-withdrawal latency measured before morphine administration (*i.e.* basal latency, nociception) for day 1 (abscissa) and day 2 (ordinate). **a**, Experimental $F_1$ females. **b**, Control $F_1$ females. **c**, Experimental $F_2$ females. **d**, Control $F_2$ females. Note that, like in $F_1$ males, basal pain sensitivity is normal in $F_1$ females (**a-b**). In the $F_2$ generation ½ of all females were tested during the 2nd day and this ½ part did not show significant difference in nociception during the 1st day ($P_1 < 0.60$, **c-d**). However all $F_2$ females tested at day 1 have shown increased basal tail-withdrawal latency with significance $P < 0.0041$ (Fig. S58a), similar to all $F_2$ males ($P < 0.00013$; Fig. S57a).

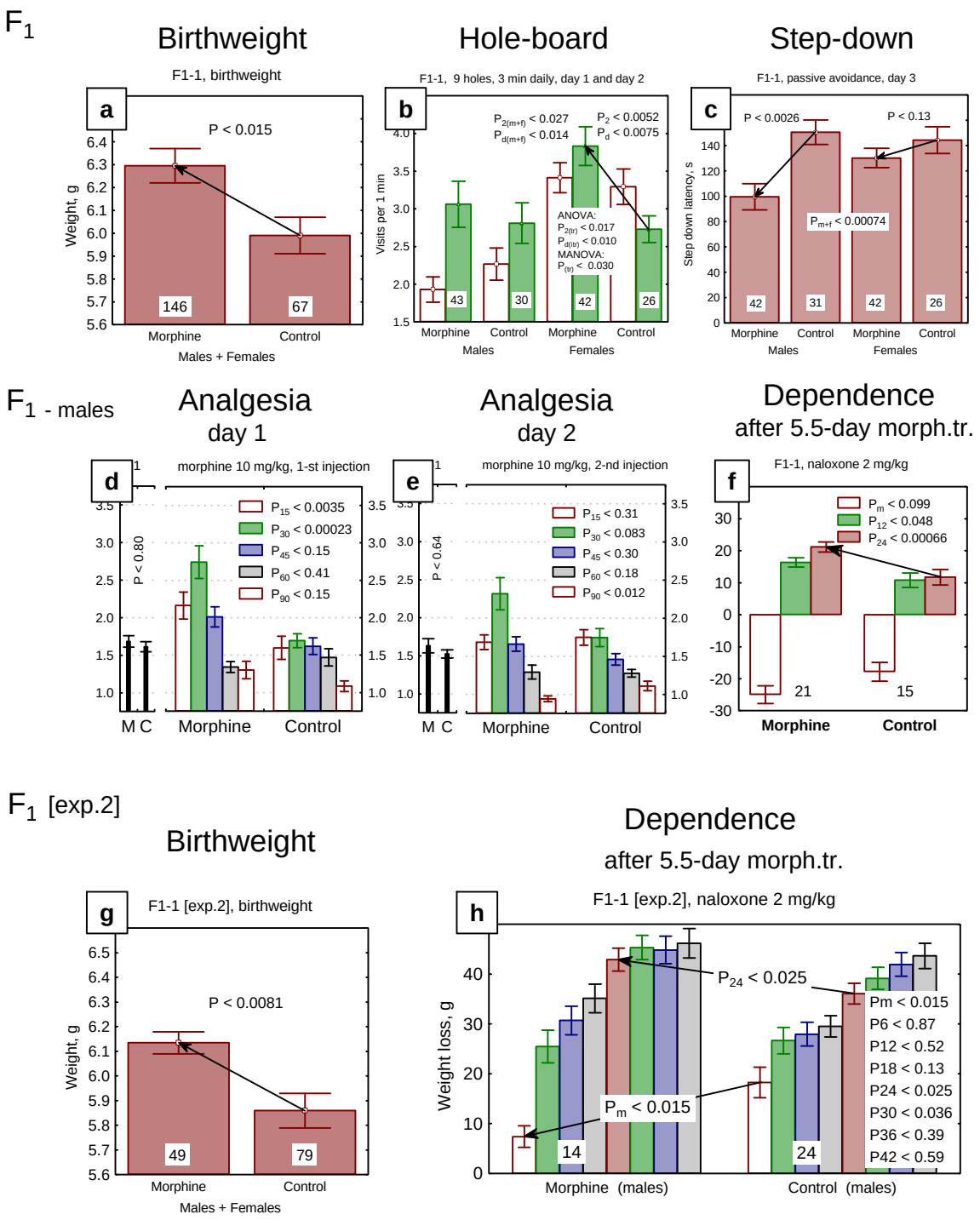



**Figure S66 | Birthweight and behaviour of $F_1$ progeny of morphine-treated male rats.** **a**, Birthweight. **b**, Hole-board. $P_2$ and $P_d$ - second day and "1$^{st}$ day *minus* 2$^{nd}$ day" (females). $P_{2(tr)}$ and $P_{d(tr)}$ - ANOVA, effect of paternal treatment. **c**, Step-down. **d**, **e**, Tail-withdrawal (males), initial pain sensitivity (black narrow bars, sec) and analgesia (wide bars, ratio). Note enhanced analgesic effect in the experimental group during the first but not during the second day. **f**, Weight loss in males during 5.5-day morphine treatment (white bars) and 12 and 24 hours after naloxone administration (coloured bars). **g**, **h**, The second (independent) experiment. Note similarity in results (**a**, **g**, **f**, **h**). Mean ± SE.

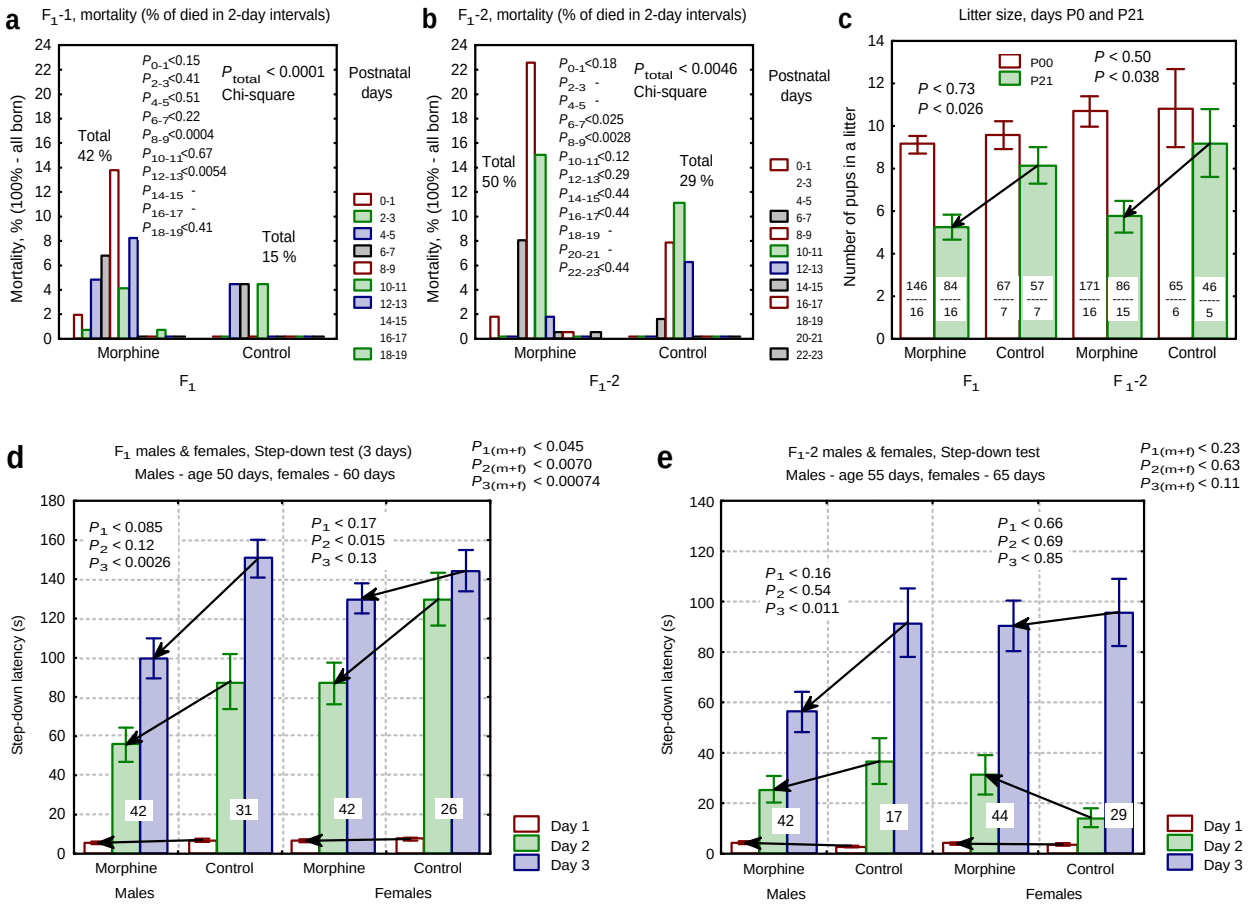



**Figure S67 | Neonatal mortality and step-down performance in the $F_1$ and $F_1$-2 rats.** **a-b**, Neonatal mortality in the $F_1$ and $F_1$-2 generations, obtained from morphine-treated males and drug-naïve females ($F_1$-2 – the second brood, they are male and female parents of $F_2$, see Fig. 1B). **c**, Litter size at birth (P0, white bars) and postnatal day 21 (P21, coloured bars). **d-e**, Step-down 3-day learning in $F_1$ and $F_1$-2 animals (step-down latencies). Mean ± SE. Note that $F_1$ and $F_1$-2 are rather similar concerning neonatal mortality and litter size (**a-c**). $F_1$ and $F_1$-2 males have shown similar decrease in step-down latencies at the 3-d day of test, however in general step-down performance was impaired in the $F_1$-2 animals to the less extent (**d-e**). Better performance of all $F_1$ animals in comparison with $F_1$-2 can be explained by better handling of $F_1$ generation (handling in terms of human-related neonatal experience). $F_1$ rats have received a lot of neonatal handling. Neonatal handling of $F_1$-2 rats was limited to their weighing (each 2$^{nd}$ day).

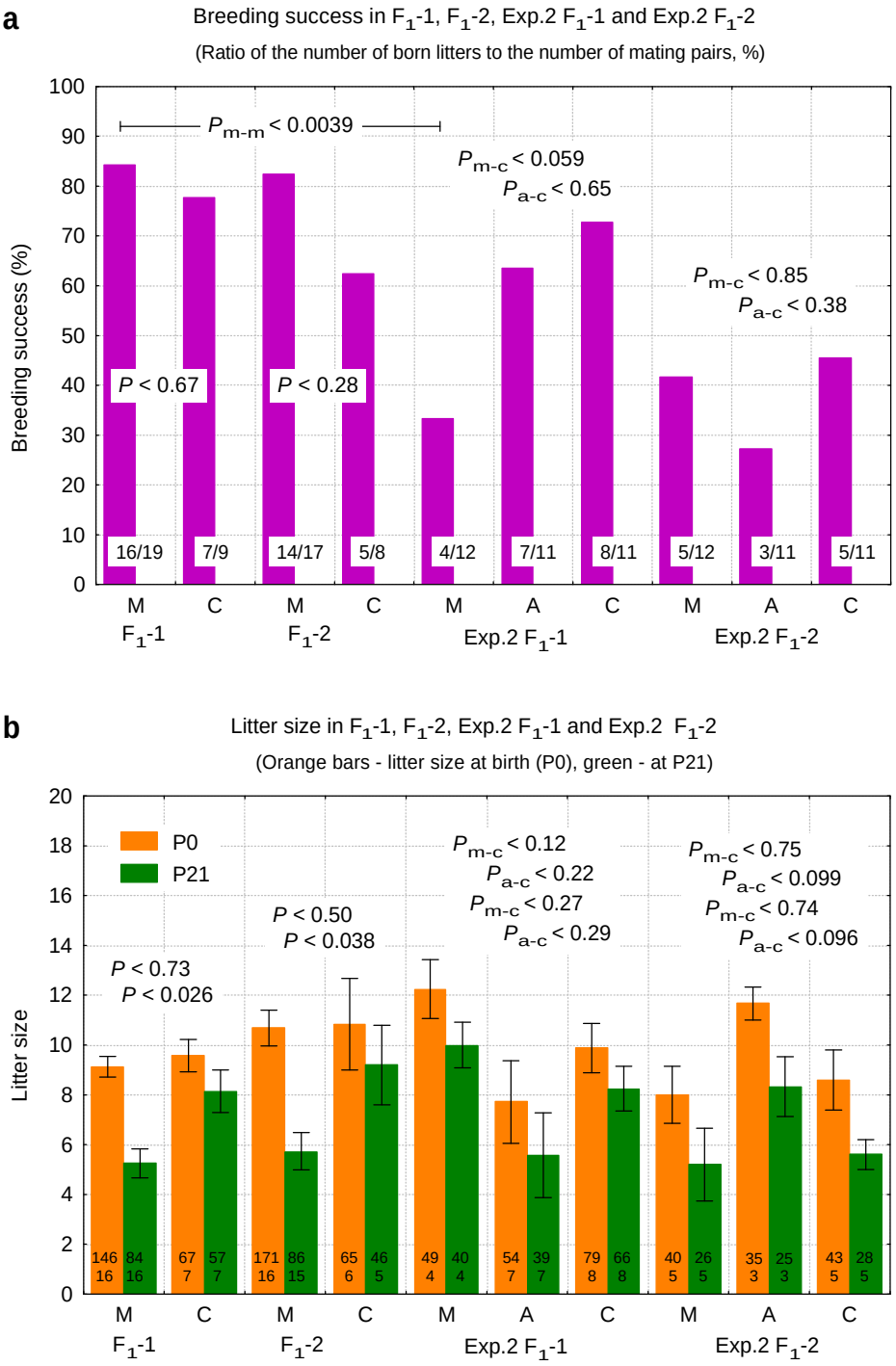



**Figure S68 | Breeding success and litter size, progeny of morphine-treated rats.** **a**, Note that breeding success of chronically morphine-treated males in Exp.2 was significantly lower than in Exp.1 ($P < 0.0039$, *Chi*-square). The reason is absolutely regular morphine treatment with 12h intervals in Exp.2 and remarkable episodes of morphine withdrawal in Exp.1 (see Methods). We can see that already at breeding these treatment protocols have brought distinguishable results. **b**, Litter size. Mean ± SE.

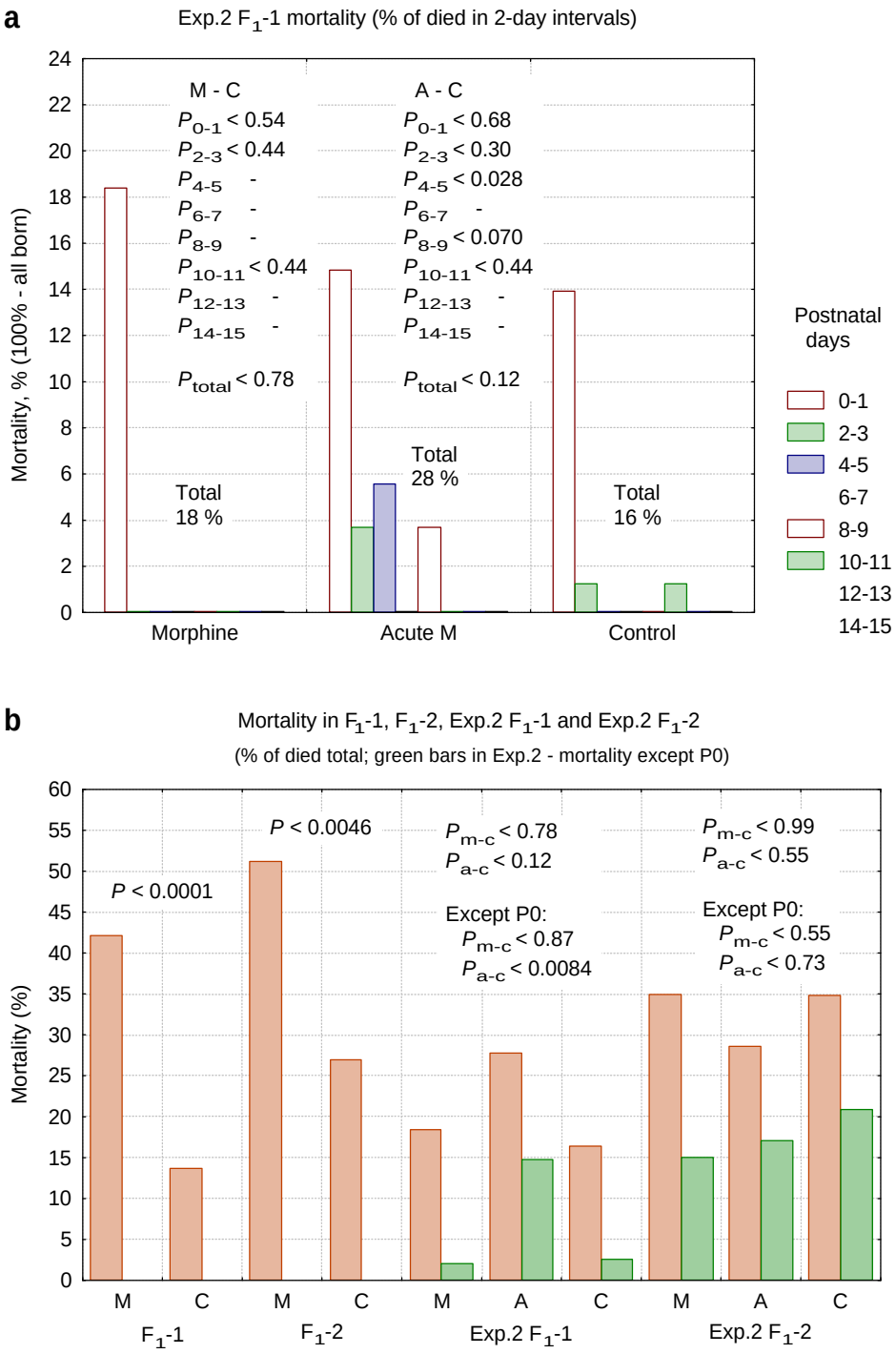



**Figure S69 | Neonatal mortality in Exp.2 $F_1$-1 and mortality comparison chart.** **a**, Exp.2 $F_1$-1 progeny of chronically and acute morphine-treated males, postnatal mortality. **b**, Comparison chart of mortality in $F_1$ offspring in Exp.1 and Exp.2. Note increased mortality in Exp.1, where male parents were chronically morphine treated with episodes of morphine withdrawal, and normal mortality in Exp.2. In the Exp.2 $F_1$-1 from acute morphine-treated males slight mortality increase can be seen (green), excluding P0.

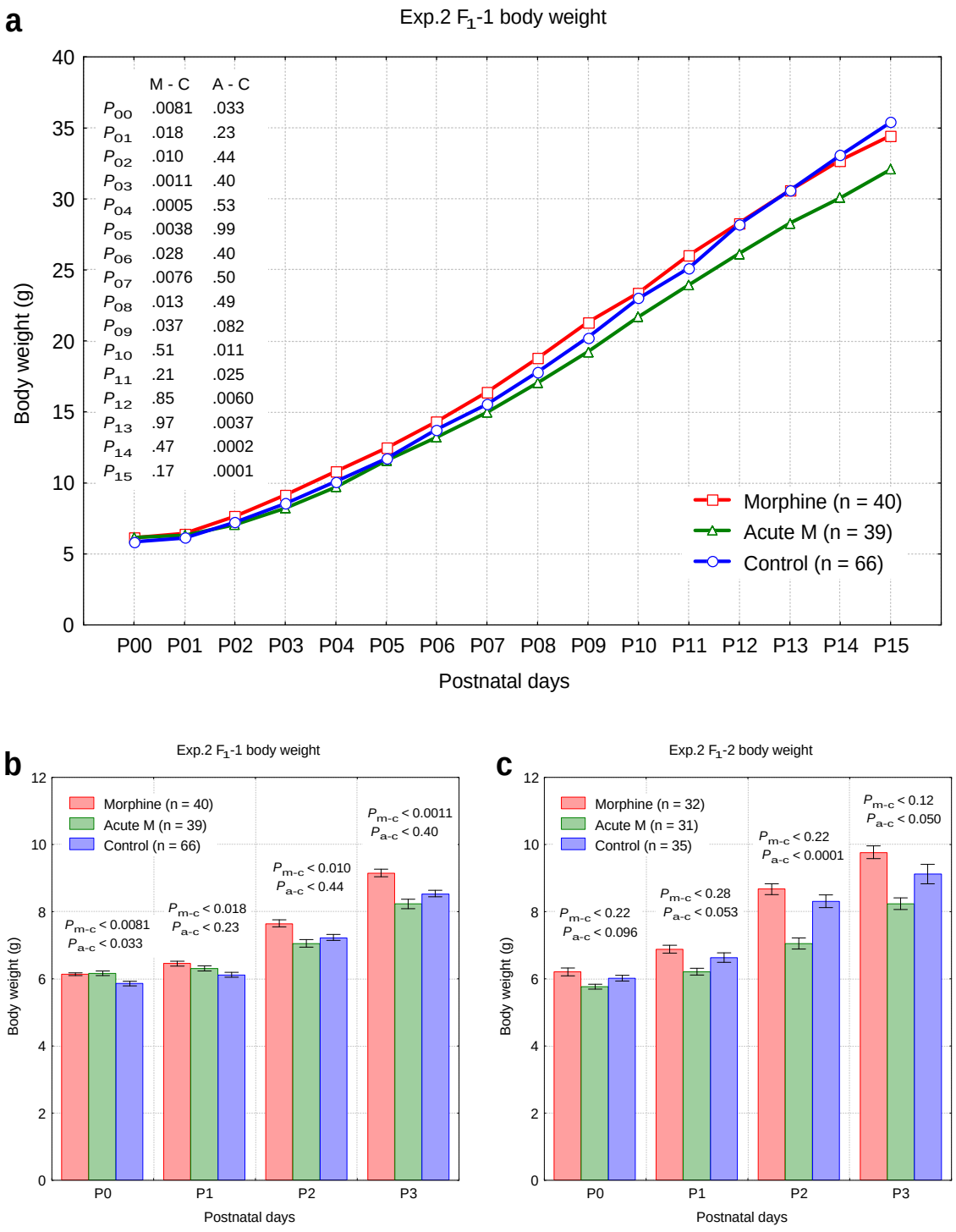



**Figure S70 | Neonatal body weight in Exp.2 $F_1$-1 and Exp.2 $F_1$-2 descendants.** Progeny of chronically and acute morphine-treated males in Exp.2. **a-b**, Note increased body weight at P00-P09 in descendants of chronically morphine-treated males and increased weight at birth only (P00) in descendants of acute treated males, however followed by decreased weight starting from P10. **c**, The second brood of Exp.2 – note some decrease in acute group only (green). Mann-Whitney U-test. Mean ± SE.

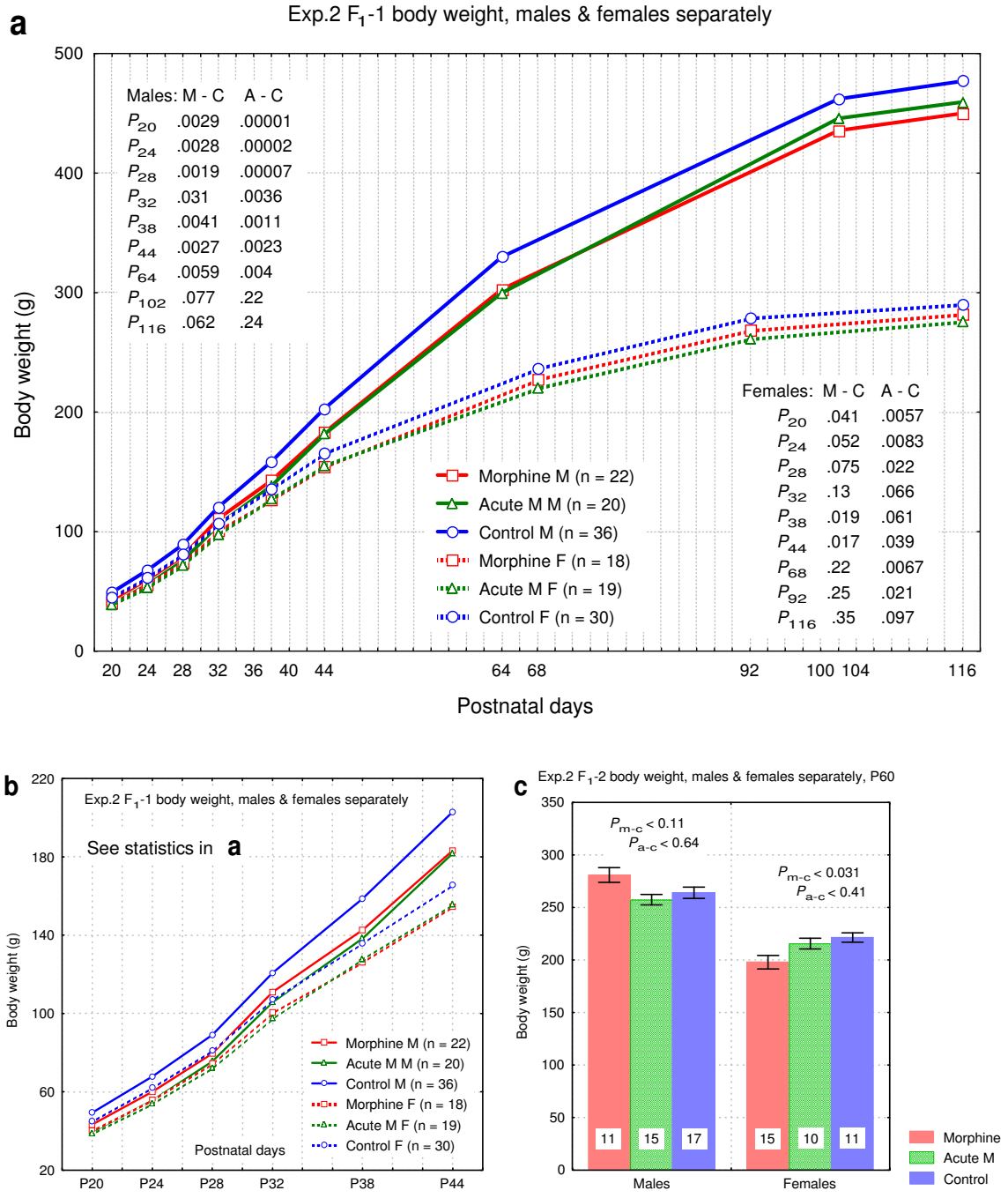



**Figure S71 | Exp.2 $F_1$-1 and Exp.2 $F_1$-2 body weight, males and females separately.** **a-b**, Note that body weight of both males and females in progeny of both chronically and acute morphine-treated males is decreased in general in Exp.2 $F_1$-1. Contrary to this observation, body weight of the same rats was increased at birth and increased body weight was evident in the offspring of chronically treated males from P00 till P09 (Fig. S70). Thus, Exp.1 and Exp.2 share increased neonatal body weight, but starting from P12-P20 body weight remains increased in Exp.1 (Fig. S49a) and becomes decreased in Exp.2 (**a**). **c**, Exp.2 $F_1$-2 – decreased weight in females, contrary to Exp.1 $F_1$-2 females.

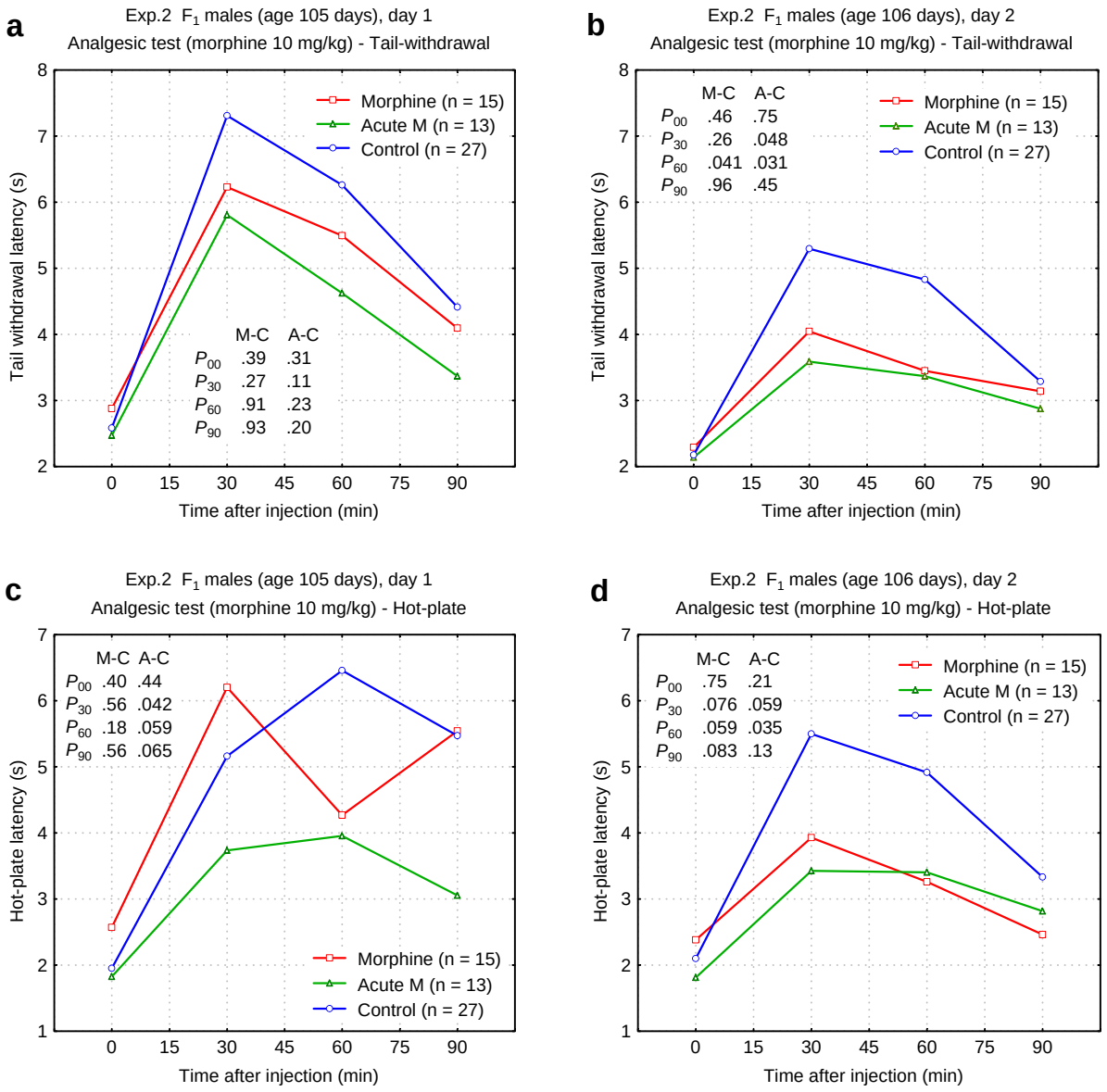



**Figure S72 | Exp.2 $F_1$ males, tail-withdrawal and hot-plate tests, day 1 and day 2.** In the second experiment (Exp.2) hot-plate temperature was increased from previous 52°C to 58°C. Due to this reason we have not previously observed learning curves here. Instead, typical analgesic curves were obtaimed in hot-plate, similar to tail-withdrawal test. Contrary to Exp.1 $F_1$-1 males, Exp.2 $F_1$-1 males have shown no differences during day 1 (**a**). But during day 2 both experimantal groups have shown decreased analgesic effect in comparison with relatively stable control (**b**). In hot-plate test the results were similar (**d**). Note that this decreased analgesic effect of morphine is more pronounced in the progeny of acute morphine-treated males, than in the offspring of chronically morphine-treated ones (**a-d**). We see here that in Exp.1 we had increased analgesic effect during day 1 and normal during day 2, whereas during Exp.2 we have normal analgesic effect during day 1 and decreased during day 2. Experimantal animals have different initial sensitivity to morphine-induced analgesia in these experiments. But increased rate of tolerance development remains the same in both Exp.1 $F_1$-1 males and Exp.2 $F_1$-1 males.

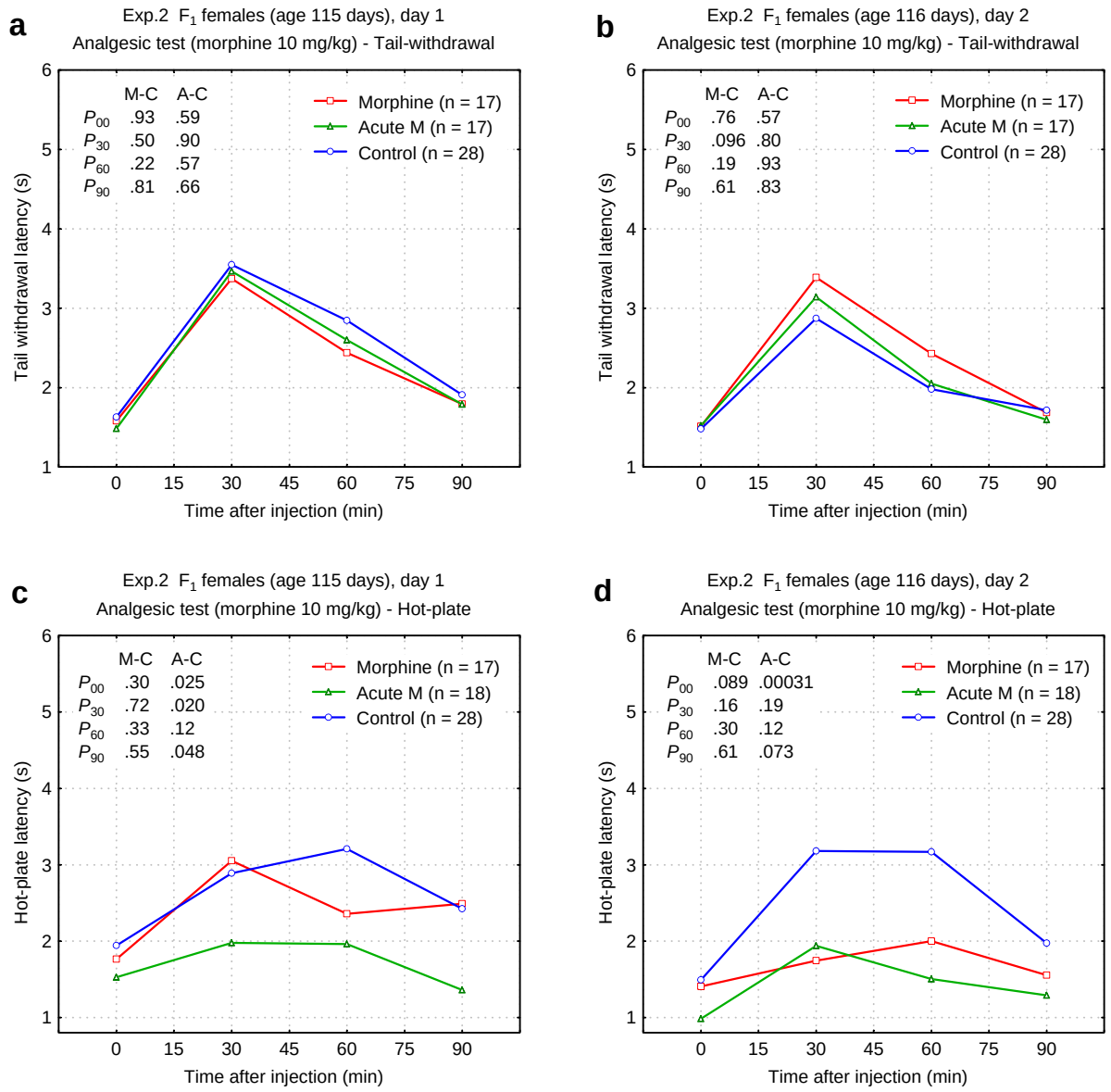



**Figure S73 | Exp.2 $F_1$ females, tail-withdrawal and hot-plate tests, day 1 and day 2.** Females Exp.2 $F_1$-1 have shown normal curves in tail-withdrawal test during both day 1 & 2 (similar to Exp.1 $F_1$-1 females). However in hot-plate test Exp.2 $F_1$-1 female progeny of acute morphine-treated males have shown decreased basal latency (before morphine administration) during both day 1 and 2. Two findings are interesting here. First, decreased basal hot-plate latency is a developing feature, because it is more statistically significant during day 2 ($P < 0.00031$) than day 1 ($P < 0.025$), and day 1 is more significant than previous testing 20 days earlier ($P < 0.67$, Fig. S83b). In males hot-plate latency is also developing feature, and it is developing in the same direction, but from another starting point: from increased latency at the beginning to the normal at the end (Fig. S83b). Second, some abnormalities in the progeny of acute morphine-treated males can be greater than in offspring of chronically morphine-treated males. We have seen it already in male descendants (see previous Figure) and in females we can see the same regularity.

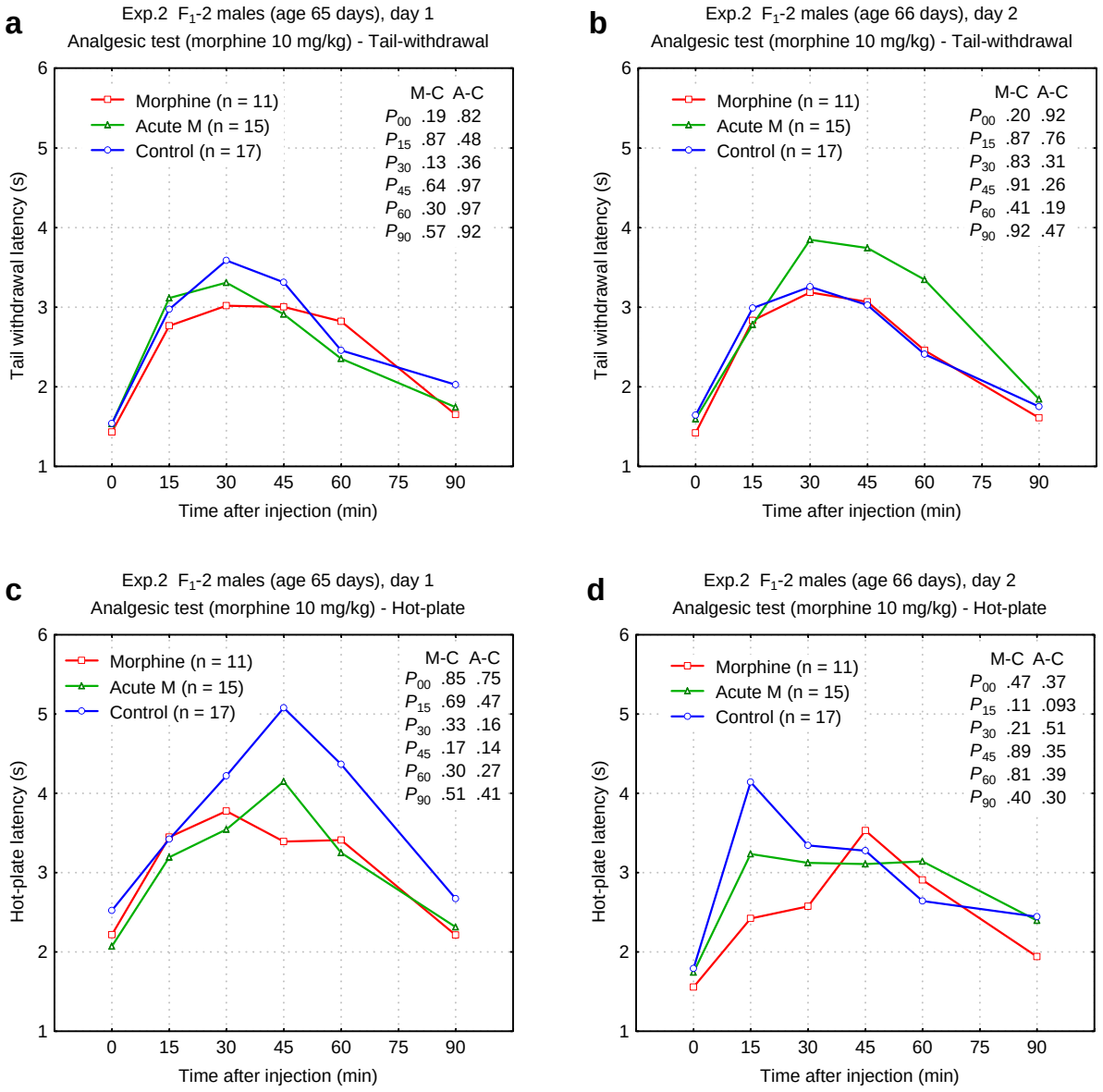



**Figure S74 | Exp.2 $F_1$-2 males, tail-withdrawal and hot-plate tests, day 1 and day 2.** The second brood Exp.2 males (Exp.2 $F_1$-2 males) did not show any differences between groups in both tail-withdrawal (**a-b**) and hot-plate (**c-d**) tests. In this aspect they are about the same as the second brood males of the first experiment (Exp.1 $F_1$-2 males). However it should be mentioned that in the second experiment the second brood males were tested being young (65 days old), and the first brood males - being relatively old (105 days). Whereas in the first experiment ages were reversed: the first brood was tested being young (65 days), and the second brood - being relatively old (138 days). And in the last case only males were tested. Thus, we can see that independently from the age of testing (P65 or P105) the second brood males do not show much differences in the analgesic tests (both descendants of chronic and acute morphine-treated males).

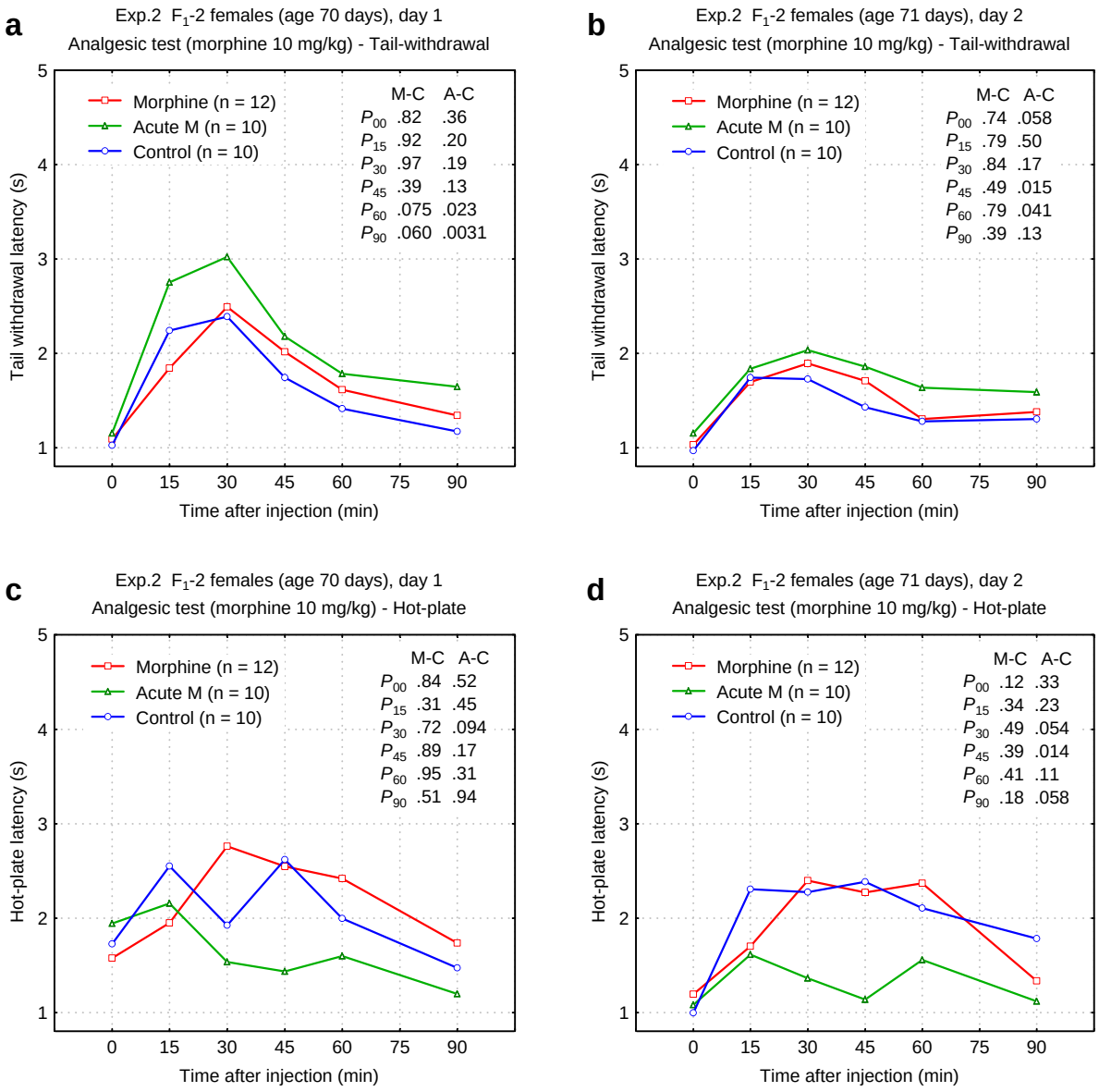



**Figure S75 | Exp.2 $F_1$-2 females, tail-withdrawal and hot-plate tests, day 1 and day 2.** Female progeny of chronically morphine-treated males Exp.2 $F_1$-2 did not differ from control in tail-withdrawal and hot-plate tests (**a-d**), see M-C. However female offspring of acute morphine-treated males Exp.2 $F_1$-2 have shown interesting pattern (**a-d**), see A-C. In tail-withdrawal test at late time points (at 45 min (day 2, $P < 0.015$) and afterwards) they have demonstrated increased analgesic effect (**a-b**). Contrary to tail-withdrawal test, in hot-plate test they have shown decreased analgesic effect, especially during day 2 (**d**), including above-mentioned time point 45 min ($P < 0.014$). Thus, in the second brood female progeny of acute morphine-treated males the enhanced analgesic effect can be observed in tail-withdrawal test and, simultaneously in the same animals, the decreased analgesic effect can be observed in hot-plate test.

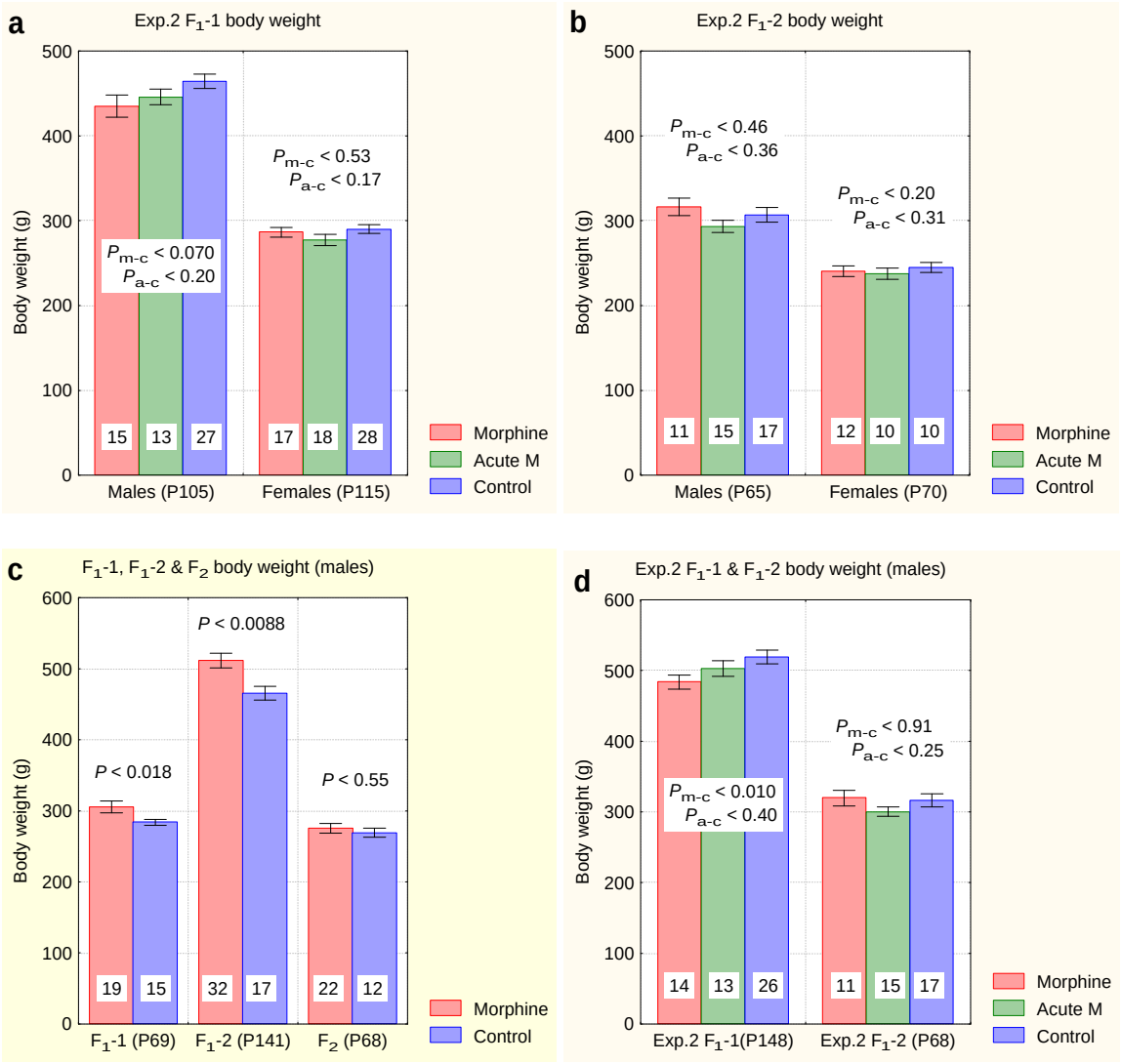



**Figure S76 | Exp.1 and Exp.2 body weight before morphine-related tests.** Body weight just before tail-withdrawal and hot-plate tests (**a-b**) and just before 5.5-day morphine treatment for investigation of naloxone-precipitated morphine withdrawal (**c-d**). **a**, Exp.2 $F_1$-1 males and females. **b**, Exp.2 $F_1$-2 males and females. **c**, Exp.1 $F_1$-1, $F_1$-2 & $F_2$ males (females were not used in naloxone test). **d**, Exp.2 $F_1$-1 & $F_1$-2 males (females were not tested). Age is provided in parentheses. Mann-Whitney U-test. Mean ± SE. Analgesic test results can not be explained by differences in body weight just because all rats have practically normal body weight. Naloxone test results, in particular increased naloxone-induced weight loss in Exp.1 $F_1$-1 and Exp.2 $F_1$-1, can not be explained by body weight difference also, because increased naloxone-induced weight loss can coexist with both increased body weight in Exp.1 $F_1$-1 (**c**) and decreased body weight in Exp.2 $F_1$-1 (**d**). In addition, increased body weight in Exp.1 $F_1$-2 (**c**) can coexist with normal naloxone-induced weight loss, and normal body weight in Exp.2 $F_1$-2 (**d**) can coexist with normal naloxone-induced weight loss (see Fig. S77 for naloxone test results).

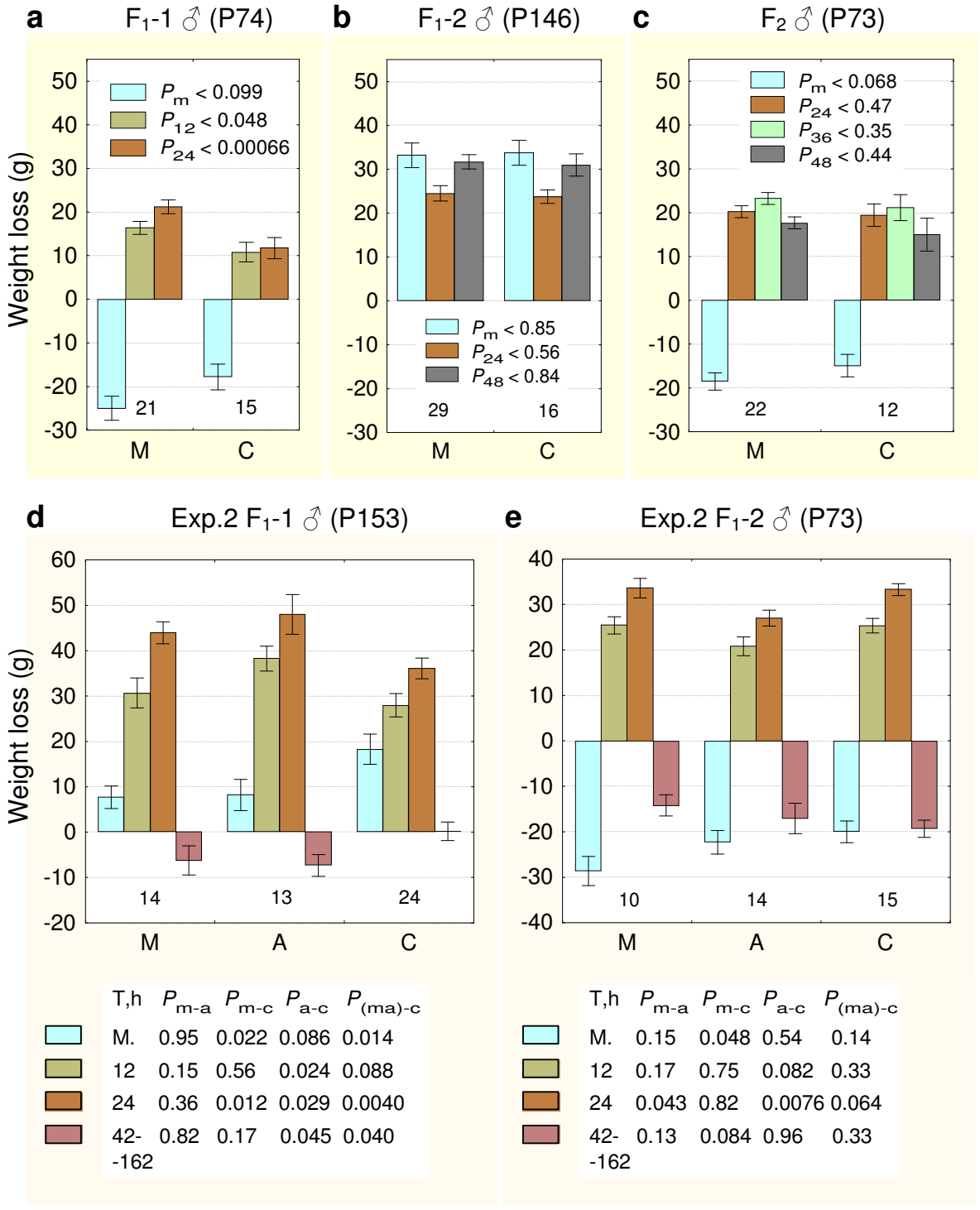



**Figure S77 | Naloxone-precipitated morphine withdrawal in Exp.1 and Exp.2.** Weight loss during 5.5-day morphine treatment (blue), during 12 hours after naloxone 2 mg/kg administration (green), during 24 hours (brown), during 36 hours (light green), during 48 hours (grey) and weight loss in time interval starting from 42 hours after naloxone and ending at 162 hours after naloxone administration (red). Postnatal day of naloxone administration is given in parentheses. Mann-Whitney U-test. Mean ± SE.

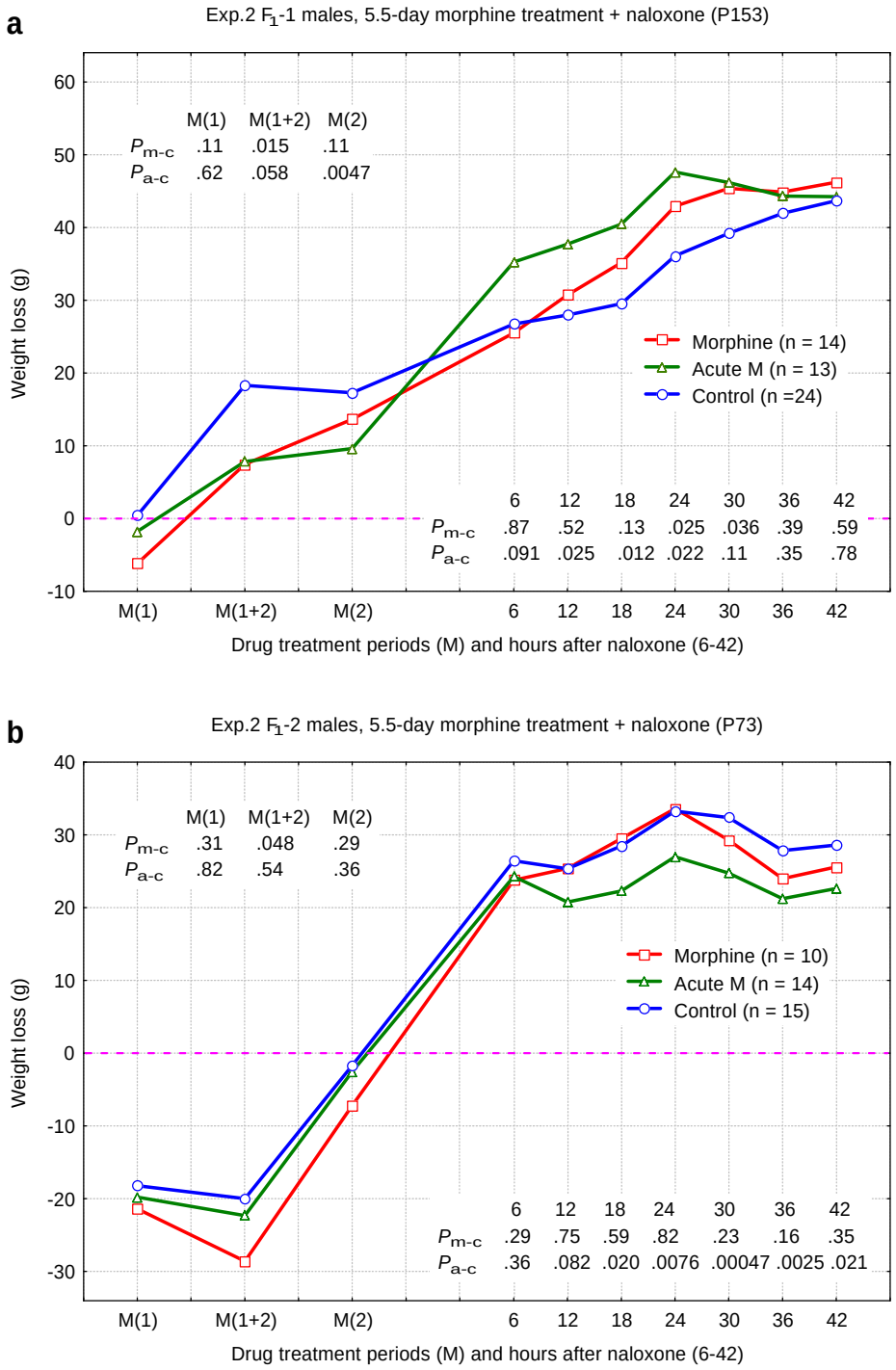



**Figure S78 | Weight loss during morphine treatment and withdrawal in Exp.2.** M(1), Weight loss during the first 2.5 days of morphine treatment. M(1+2), Weight loss during the whole 5.5-day morphine treatment. M(2), Weight loss during the second part, during the last 3 days of morphine treatment. 6-…-42, Weight loss during 6-…-42 hours after naloxone administration. Age of naloxone administration is shown in parentheses: (P153) & (P73). **a**, Exp.2 $F_1$-1 males. **b**, Exp.2 $F_1$-2 males. Mann-Whitney U-test. Mean.

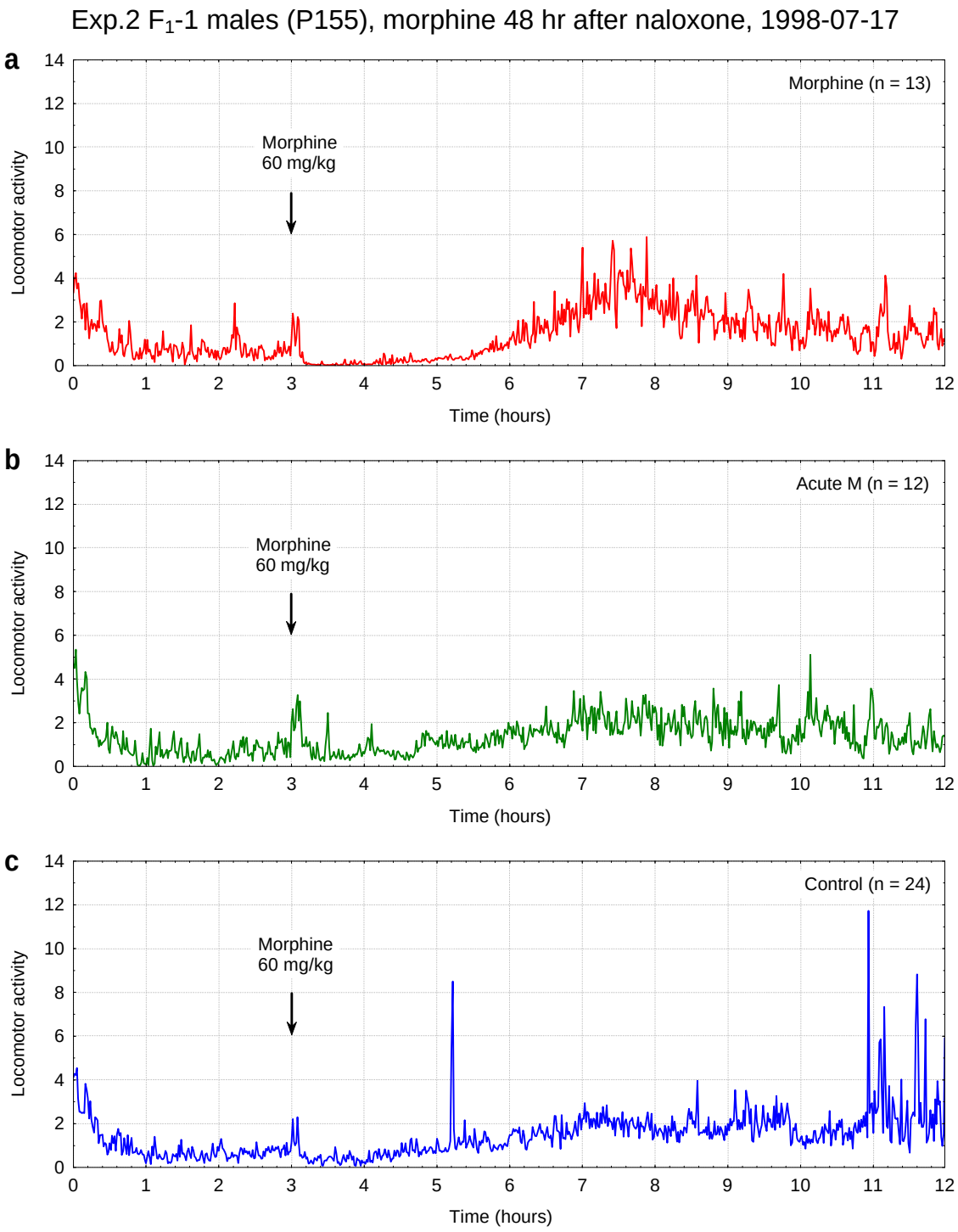



**Figure S79 | Locomotor activity in Exp.2 $F_1$-1 – effect of morphine injection.** Note suppressing effect of morphine injection on locomotor activity in progeny of chronically treated males only (see the first two hours after morphine administration). At time "0" each rat was placed in a new unfamiliar individual cage – that's why we can see increased locomotor activity during the first 30 minutes. All rats were morphine-treated during 5.5 days and received naloxone 2 mg/kg 48 hr before shown 60 mg/kg injection.

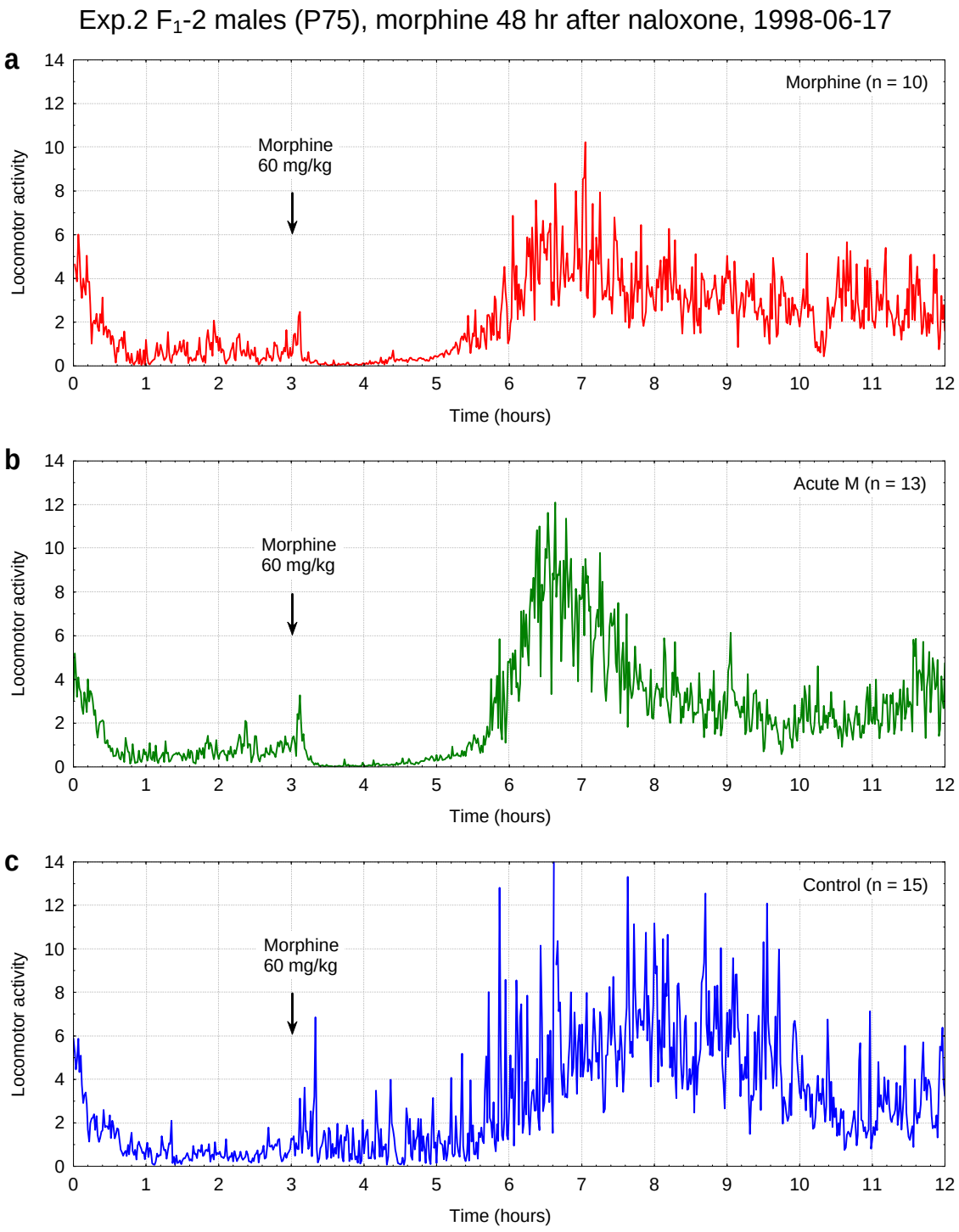



**Figure S80 | Locomotor activity in Exp.2 $F_1$-2 – effect of morphine injection.** Note suppressing effect of morphine administration on locomotor activity in offspring of chronically and acute morphine-treated males. In the offspring of acute morphine-treated males we can see here relatively rare case, when an effect is absent in the first brood (Fig. S79b), but present in the second one (Fig. S80b). It means that this effect was formed in epigenetic heredity significantly after acute morphine treatment (25 mg/kg *i.p.* injection).

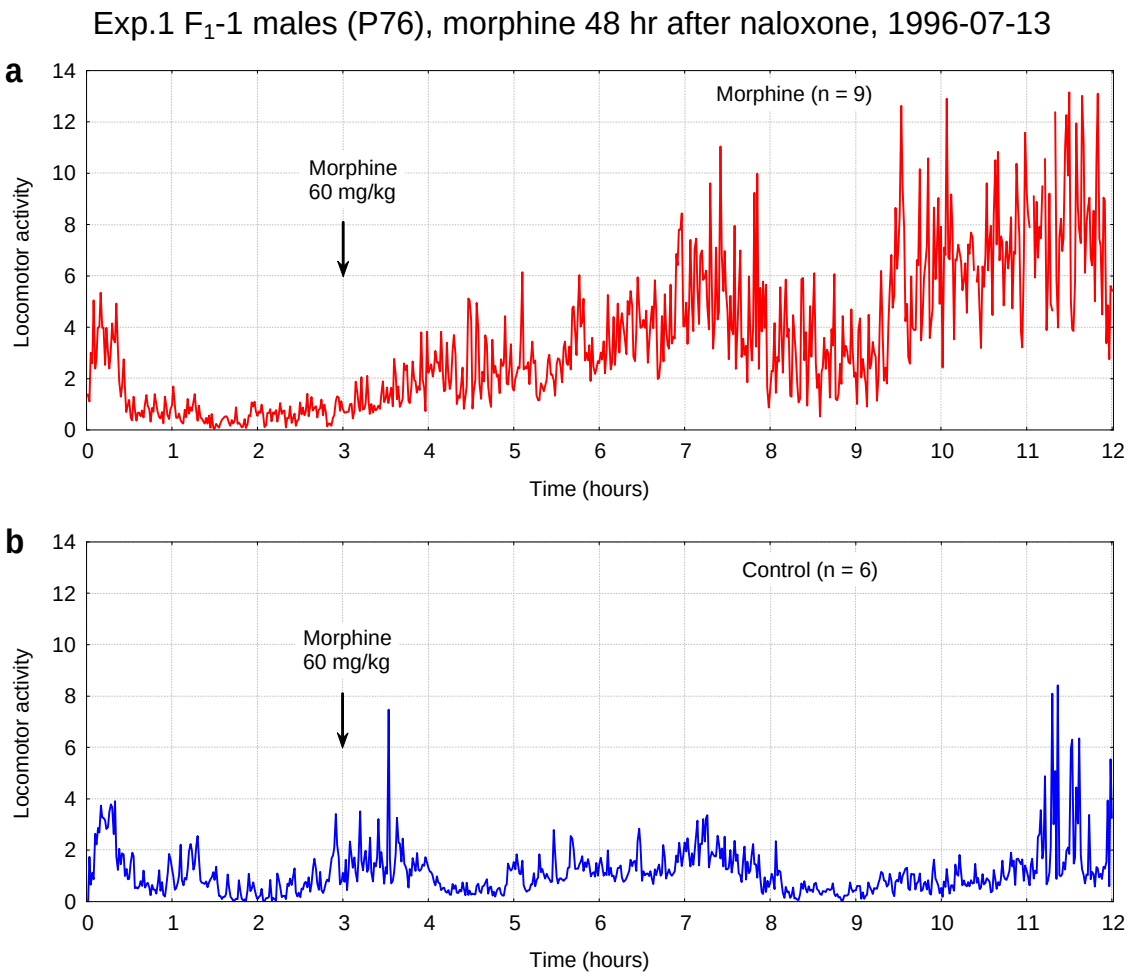



**Figure S81 | Locomotor activity in Exp.1 $F_1$-1 – effect of morphine injection.** The first experiment. Note that morphine injection has evoked increased locomotor activity in experimental rats (**a**) in comparison with synchronous control (**b**). If we compare Exp.1 with Exp.2, we can see that different protocols of paternal drug treatment can lead to different results in the progeny. Short-term suppressing effect of morphine injection on locomotor activity was observed in the progeny after regular chronic paternal morphine treatment (Exp.2, Fig. S80a). Relatively long-lasting enhancing effect of morphine injection on locomotor activity was observed in the offspring after chronic paternal morphine treatment with episodes of morphine withdrawal (Exp.1, Fig. S81a). Note also, that if we look at controls (Figs. S79c, S80c, S81b), we can see that Figs. S79c and S81b are practically identical, whereas in the Fig. S80c locomotor activity after morphine injection is about threefold higher. It can not be explained by animal age, because ages in Exp.2 $F_1$-2 and Exp.1 $F_1$-1 are identical (P75 and P76, respectively). However we have discovered that identical results in control groups were obtained when animals were tested exactly during the same calendar period: July 17 – Exp.2, July 13 – Exp.1. Despite different age (P155 – Exp.2, P76 – Exp.1) control rats have shown identical locomotor activity here. On the other hand, rats with similar age (P75), but tested one month earlier (not in the middle of summer, but in its beginning – June 17), have shown increased locomotor response.

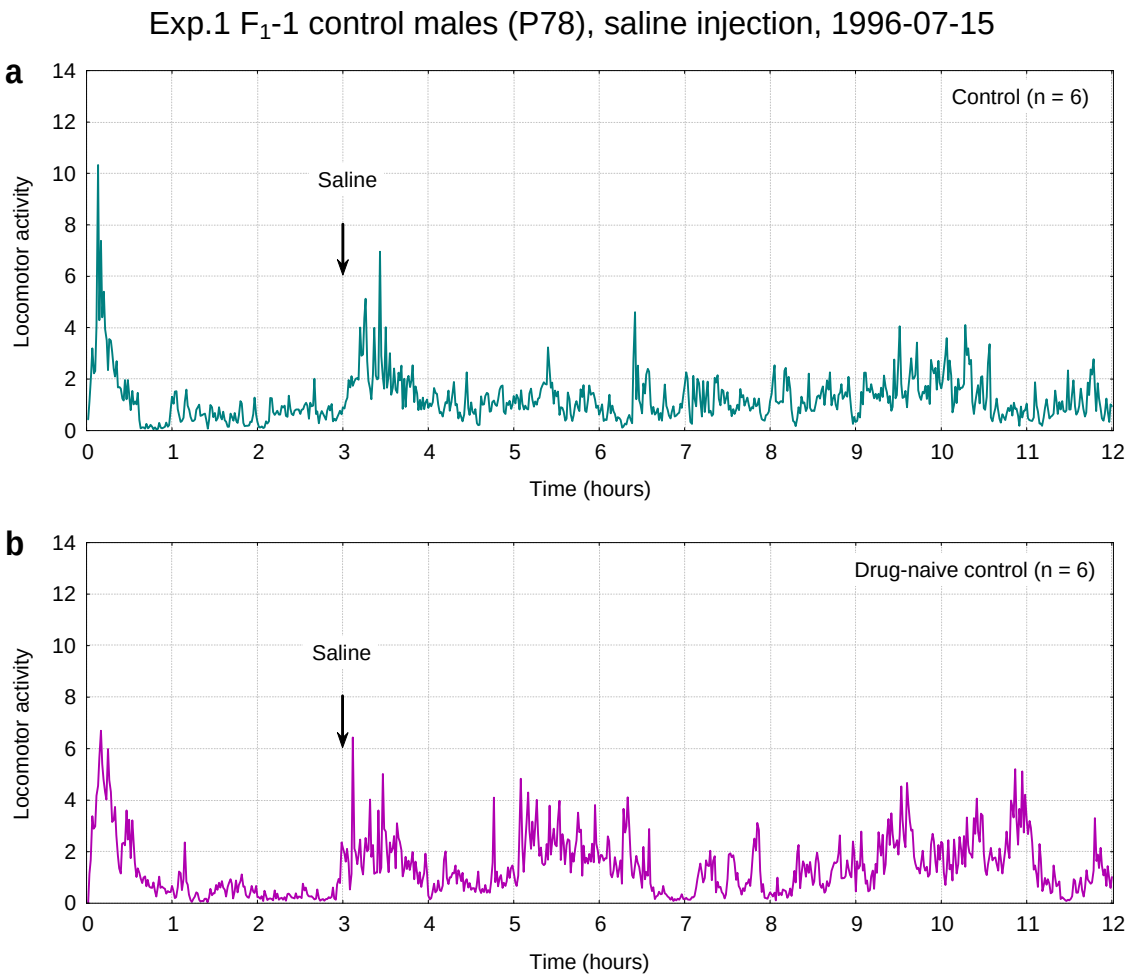



**Figure S82 | Locomotor activity in Exp.1 $F_1$-1 control – effect of saline injection.** The first experiment. **a**, Control Exp.1 $F_1$-1 males, pretreated with morphine during 5.5 days and received naloxone 2 mg/kg injection, but, instead of morphine 60 mg/kg injection done 48 hr after naloxone administration, they have received saline injection during their locomotor activity recording 96 hr after naloxone administration. **b**, Control Exp.1 $F_1$-1 males, which were not pretreated with morphine, never received naloxone (*i.e.* they are drug-naïve), they have received saline injection during their locomotor activity recording synchronously with above-mentioned 5.5-day morphine-treated and naloxone-treated group. Both these controls are very important for interpretation of results of morphine 60 mg/kg injection. They show that in Exp.1 $F_1$-1 males locomotor activity was enhanced by morphine 60 mg/kg injection in the experimental group (Fig. S81a) and locomotor activity was not suppressed by morphine 60 mg/kg injection in the control group (Fig. S81b). They also show that general locomotor activity was not impaired by previous 5.5-day morphine treatment and naloxone administration, at least in control animals (compare Fig. S82a with Fig. S82b). Take into account some remark that saline was applied (due to pure technical reasons) 96 hr after naloxone administration, whereas morphine 60 mg/kg injection was done exactly 48 hr after naloxone administration in all experiments and in all groups, except above-mentioned two groups which received saline.

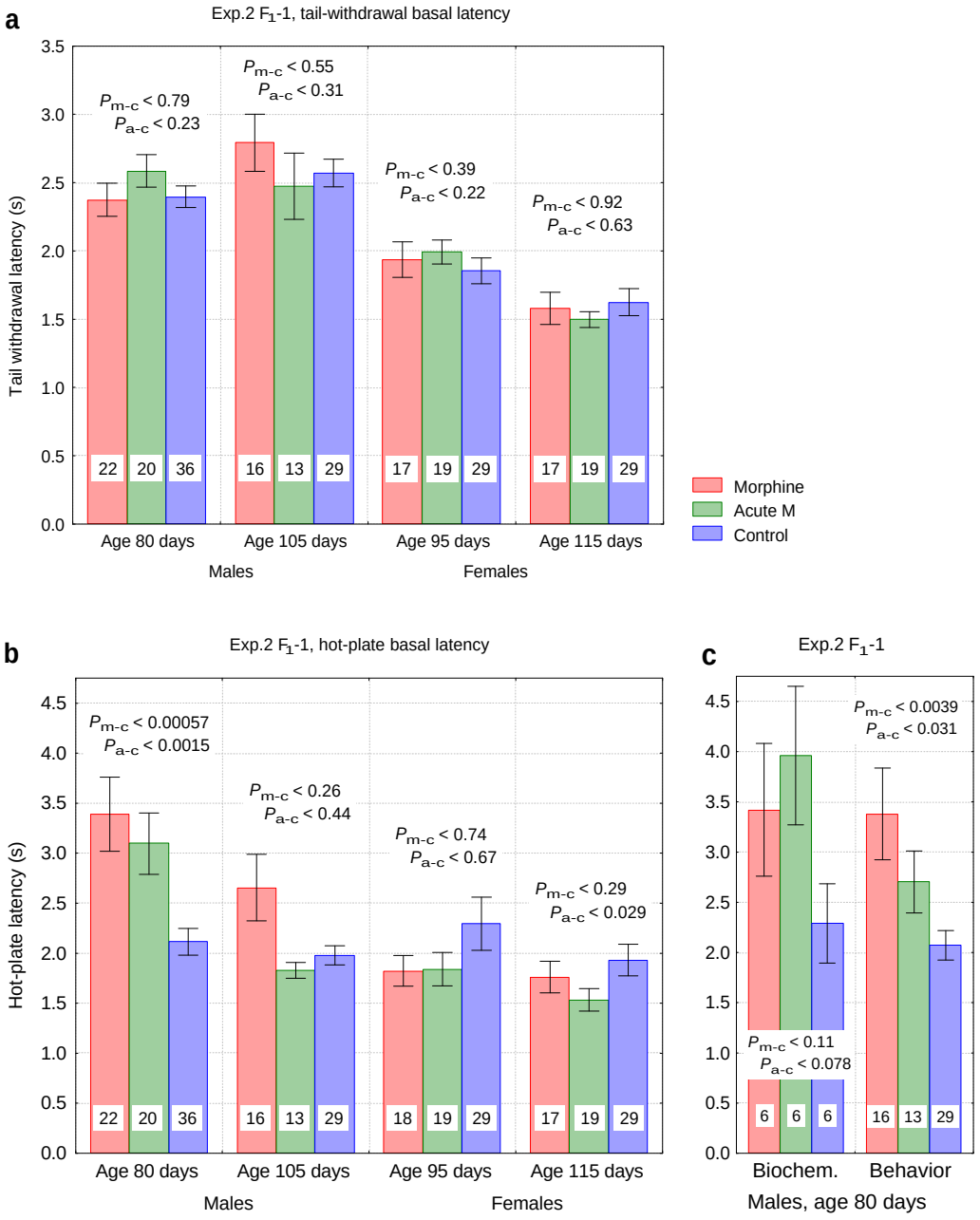



**Figure S83 | Basal pain sensitivity at 2.5 and 3.5 months in Exp.2 $F_1$-1 rats.** Tail-withdrawal (**a**) and hot-plate (**b-c**) latencies in Exp.2 $F_1$-1 descendants of morphine-treated males. Note increased basal latency in hot-plate (**b**), but not tail-withdrawal (**a**), test at the age of 80 days in males. Increased latency was observed in male progeny of both chronically (red) and acute (green) morphine-treated males. Due to aging and/or experience this deviation disappears (see age 105 days), disappears faster in the progeny of acute-treated males. In females, normal at P95 (**b**), the decrease of latency goes so far that the progeny of acute-treated males has decreased latency at P115 ($P < 0.029$). Six males from each group were taken before P105 for measurement of catecholamine brain tissue levels (Fig. S84) and they are shown separately here (**c**, Biochem.). Mean ± SE.

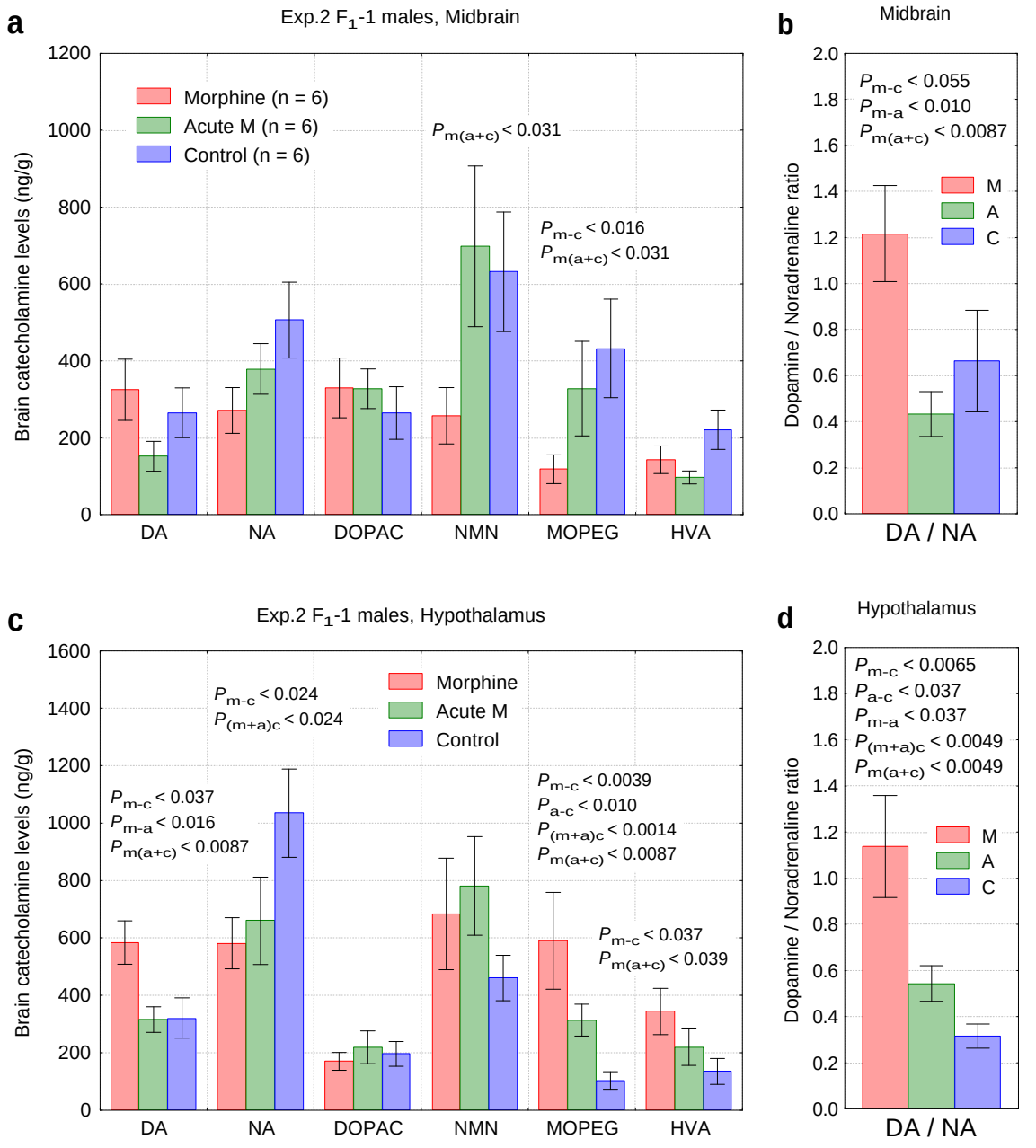



**Figure S84 | Brain tissue catecholamine levels in Exp.2 $F_1$-1 males.** Progeny of chronically morphine-treated males (red), progeny of males after single (acute) 25 mg/kg morphine injection (green) and progeny of control males (blue) at the age of 84 days. Catecholamines and their metabolites in midbrain (**a-b**) and hypothalamus (**c-d**). Brain homogenate levels (nanograms per gram wet tissue) of dopamine (DA), noradrenaline (NA), 3,4-dihydroxyphenylacetic acid (DOPAC), normetanephrine (NMN), 3-methoxy-4-hydroxyphenylglycol (MOPEG), homovanillic acid (HVA) measured by means of high-performance liquid chromatography (HPLC) with electrochemical detection (ED). All significant differences ($P < 0.05$) are shown. Mann-Whitney U-test. Six rats in each group. Mean ± SE.

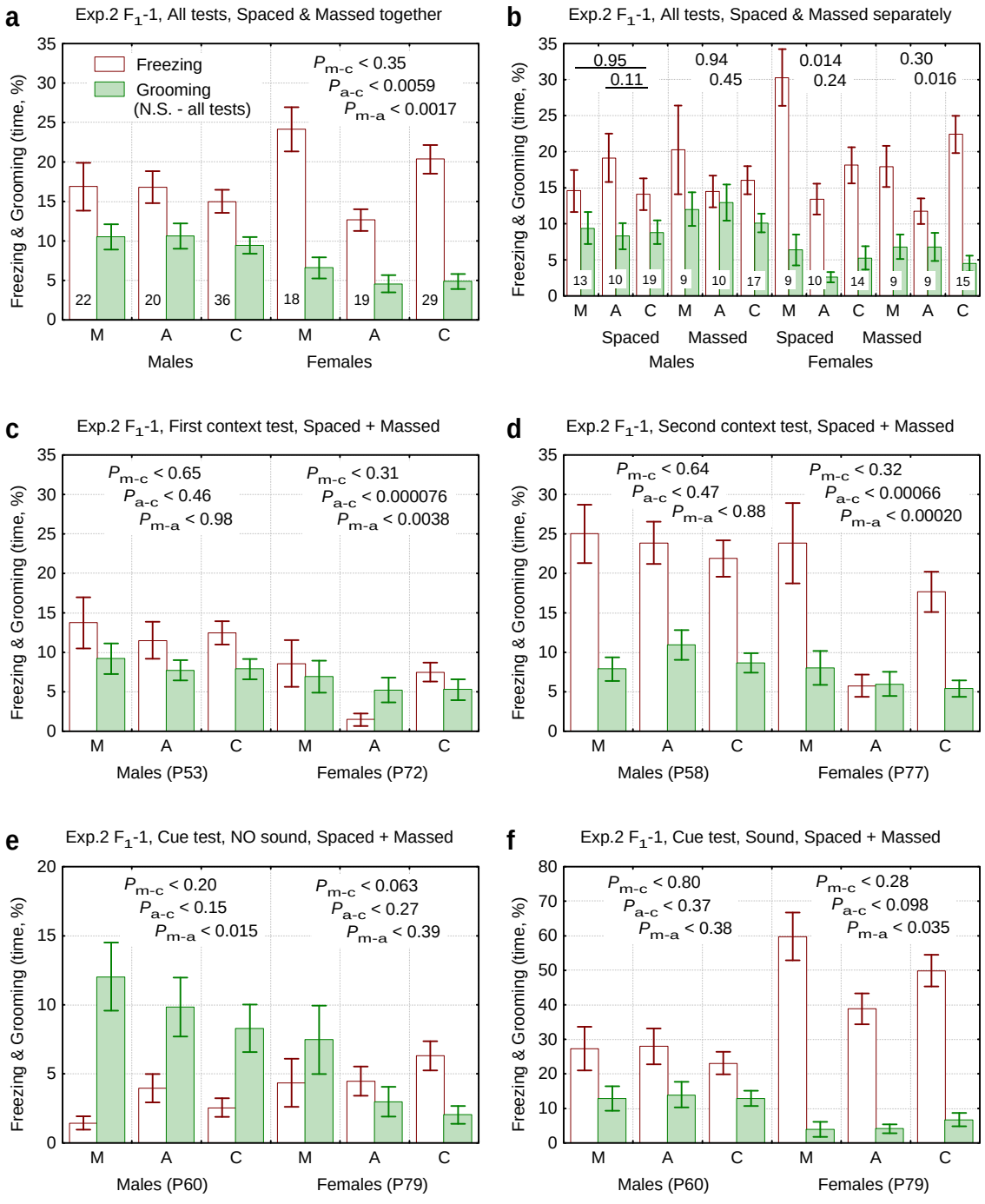



**Figure S85 | Fear conditioning in Exp.2 $F_1$-1 rats, spaced or massed training.** Freezing (white columns) and grooming (shaded columns) as a proportion of the total recording time, %. M – progeny of chronically morphine-treated males, A – progeny of acute morphine-treated males, C – control. Age is given in parentheses. **a**, All tests **c-f** pooled. **b**, The same as **a**, but spaced-trained and massed-trained groups are shown separately. **c**, The first context test, after 0.28 mA training. **d**, The second context test, after 0.48 mA training. **e**, Cue test before sound. **f**, Cue test during sound. Significance is given for freezing only. Grooming is not significant. Mann-Whitney U-test. Mean ± SE.

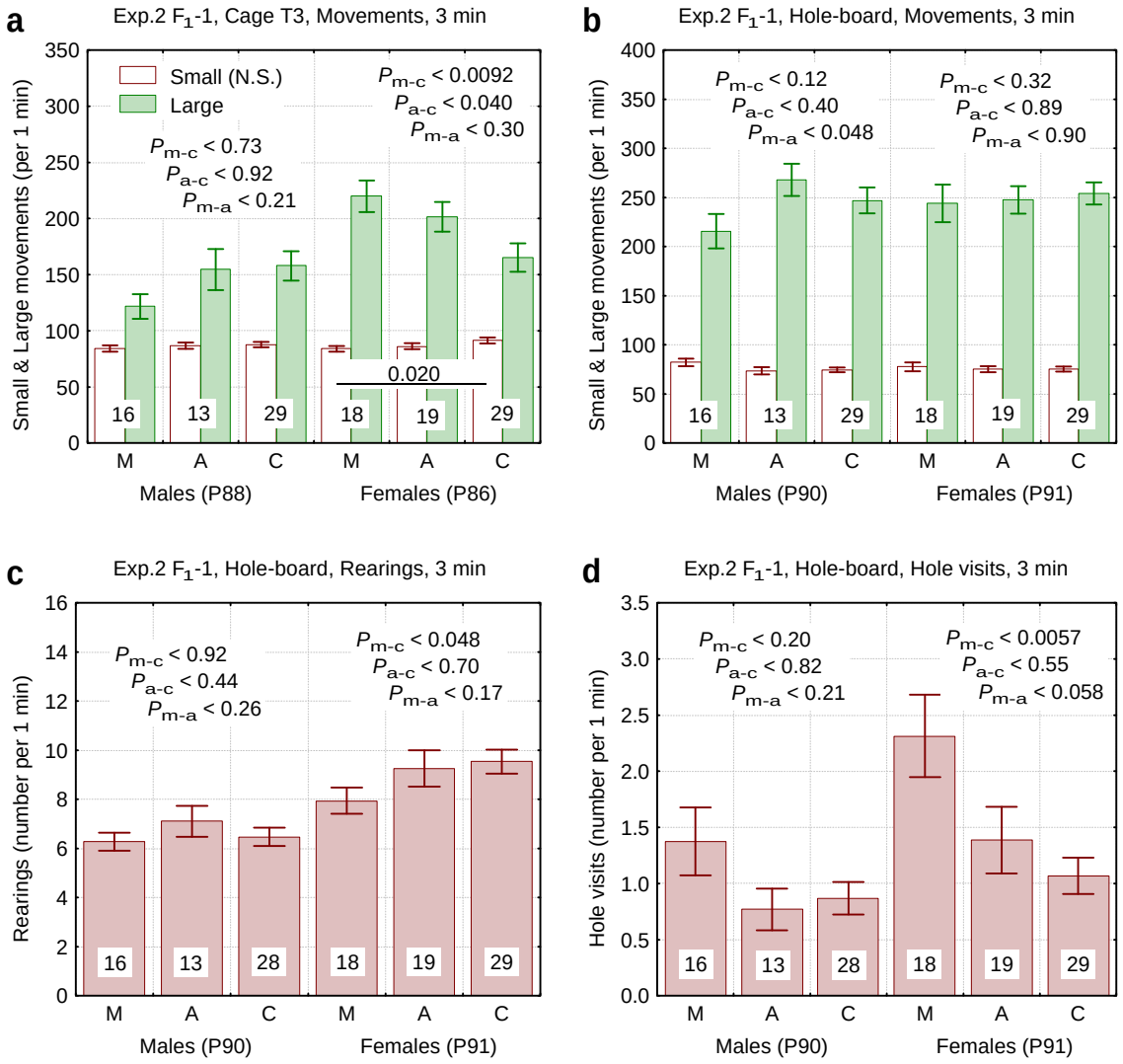



**Figure S86 | New cage T3 exploration and hole-board test in Exp.2 $F_1$-1 rats. a-b**, Small (white columns) and large (shaded columns) movements in a new cage Type 3 (**a**) and in Hole-board apparatus (**b**). In **a-b** statistical difference is given for large movements only. There were no differences in small movements, except one $P < 0.020$, shown in **a**. **c**, Rearings during hole-board test. **d**, Hole visits in hole-board test. Note that in the previous test (Fear conditioning) all animals have shown more or less normal behaviour, except females, obtained from acute morphine-treated males. The last ones have shown decreased freezing response, highly significant in both context tests. One can assume that it might be just due to their general increase in locomotor activity. However their behaviour in a new cage T3 and in hole-board shows that it is not the case, because in a new cage (**a**) and hole-board (**b-d**) another group of descendants, - females from chronically morphine-treated males, differs from control to a greater extent. It means that different protocols of paternal morphine treatment have targeted different aspects of behaviour in their untreated female offspring. Mann-Whitney U-test. Mean ± SE.

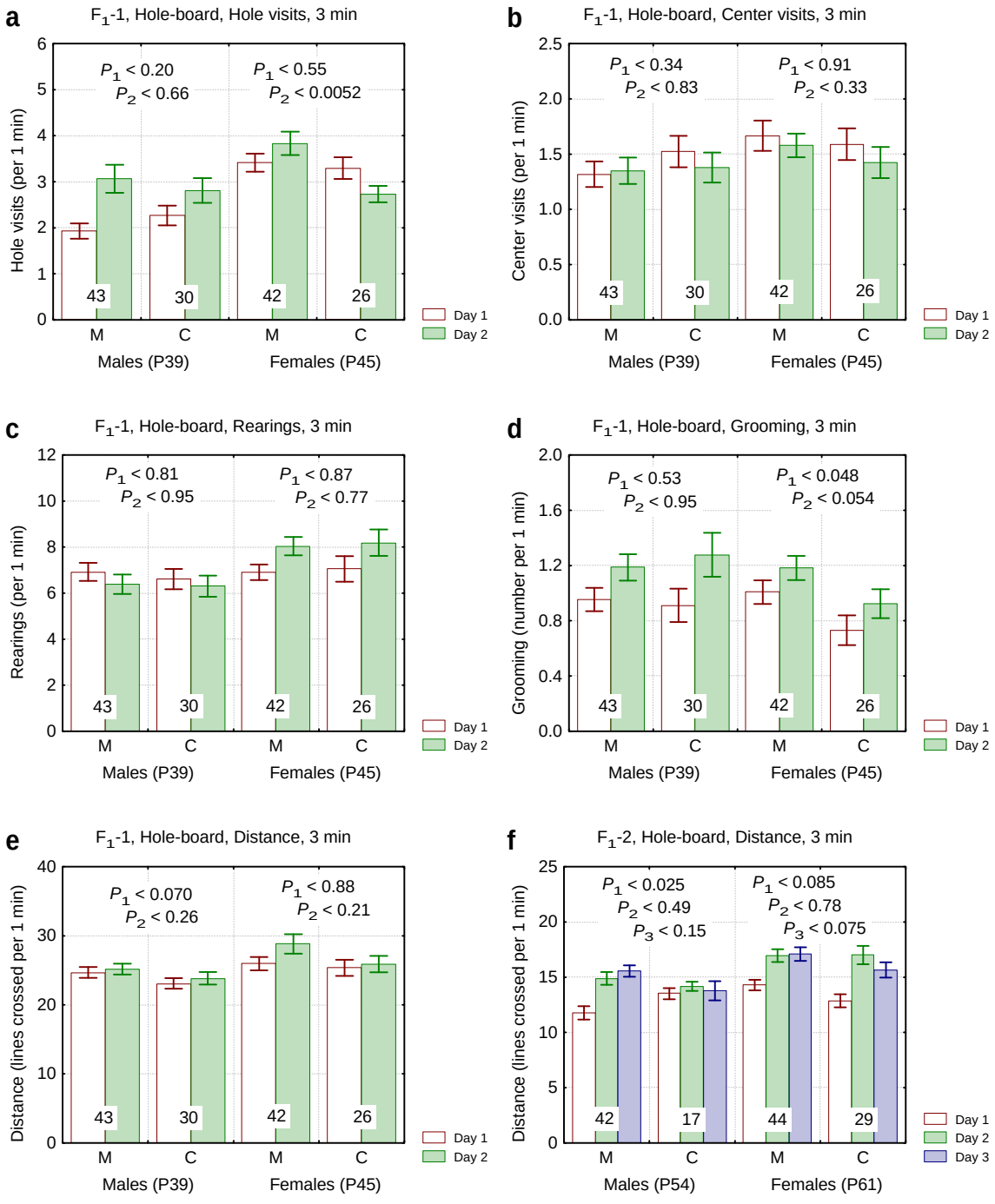



**Figure S87 | Hole-board test in Exp.1 $F_1$-1 and Exp.1 $F_1$-2 rats. a-e**, Hole-board test in $F_1$-1, day 1 (white columns) and day 2 (shaded columns). **a**, Hole visits. **b**, Center visits. **c**, Rearings. **d**, Grooming events. **e**, Lines crossed (distance). **f**, Hole-board test in $F_1$-2, day 1 (white columns), day 2 (green columns) and day 3 (blue columns). M - progeny of chronically morphine-treated males, C - control. Age is shown in parentheses. Note that shown behaviour is rather normal here, except increased hole visits in $F_1$-1 females during day 2 (**a**). Mann-Whitney U-test. Mean ± SE.

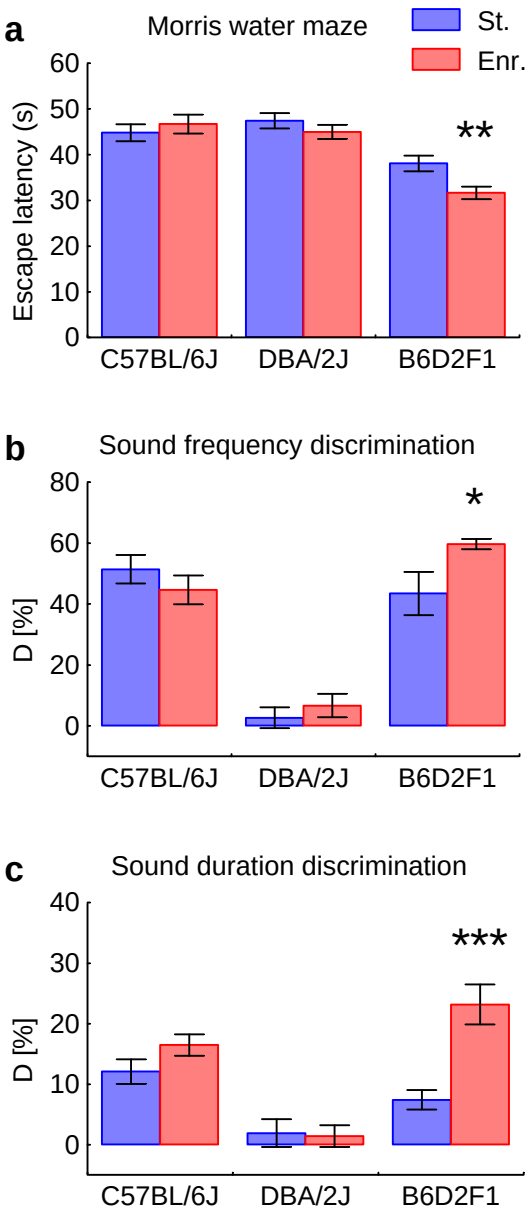



**Figure S88 | Cage enrichment enhances hybrid vigour in $F_1$ mice (B6D2F1).** **a**, Morris water maze. **b**, Sound frequency discrimination. **c**, Sound duration discrimination. Standard (blue) and enriched (red) C57BL/6J, DBA/2J and B6D2F1 female mice. Note remarkable environmental induction of hybrid vigour in all these operant tasks. Each group consists of 72 mice in **a** and 8 mice in **b-c**. Mean ± SE. Asterisk, $P < 0.05$; double asterisk, $P < 0.01$; triple asterisk, $P < 0.001$.

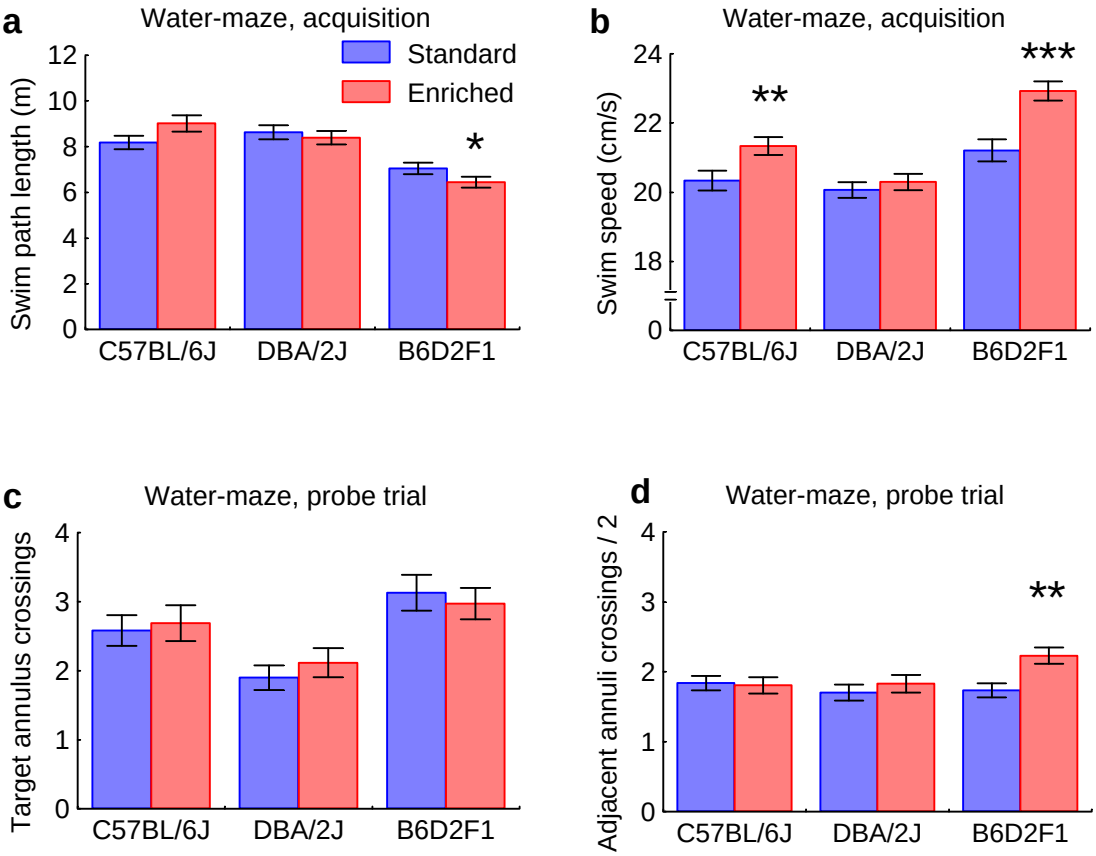



**Figure S89 | Standard and enriched C57BL/6J, DBA/2J and B6D2F1 mice in Morris water maze. a**, Swim path length. **b**, Swim speed. **c**, Probe trial target annulus crossings. **d**, Probe trial adjacent annuli crossings. Swim path and speed are averaged values of 16 training trials. Annuli crossings were measured during probe trial. Note that swim speed (**b**) is increased in both enriched C57BL/6J and B6D2F1. However only in the enriched B6D2F1 we can see decreased path length (**a**) and improved overall performance - decreased escape latency (Fig. S88a). Note that efficient behavior of B6D2F1 can be seen in the probe trial also - increased number of adjacent annuli crossings (**d**). Mean ± SE. Asterisk, $P < 0.05$; double asterisk, $P < 0.01$; triple asterisk, $P < 0.001$.

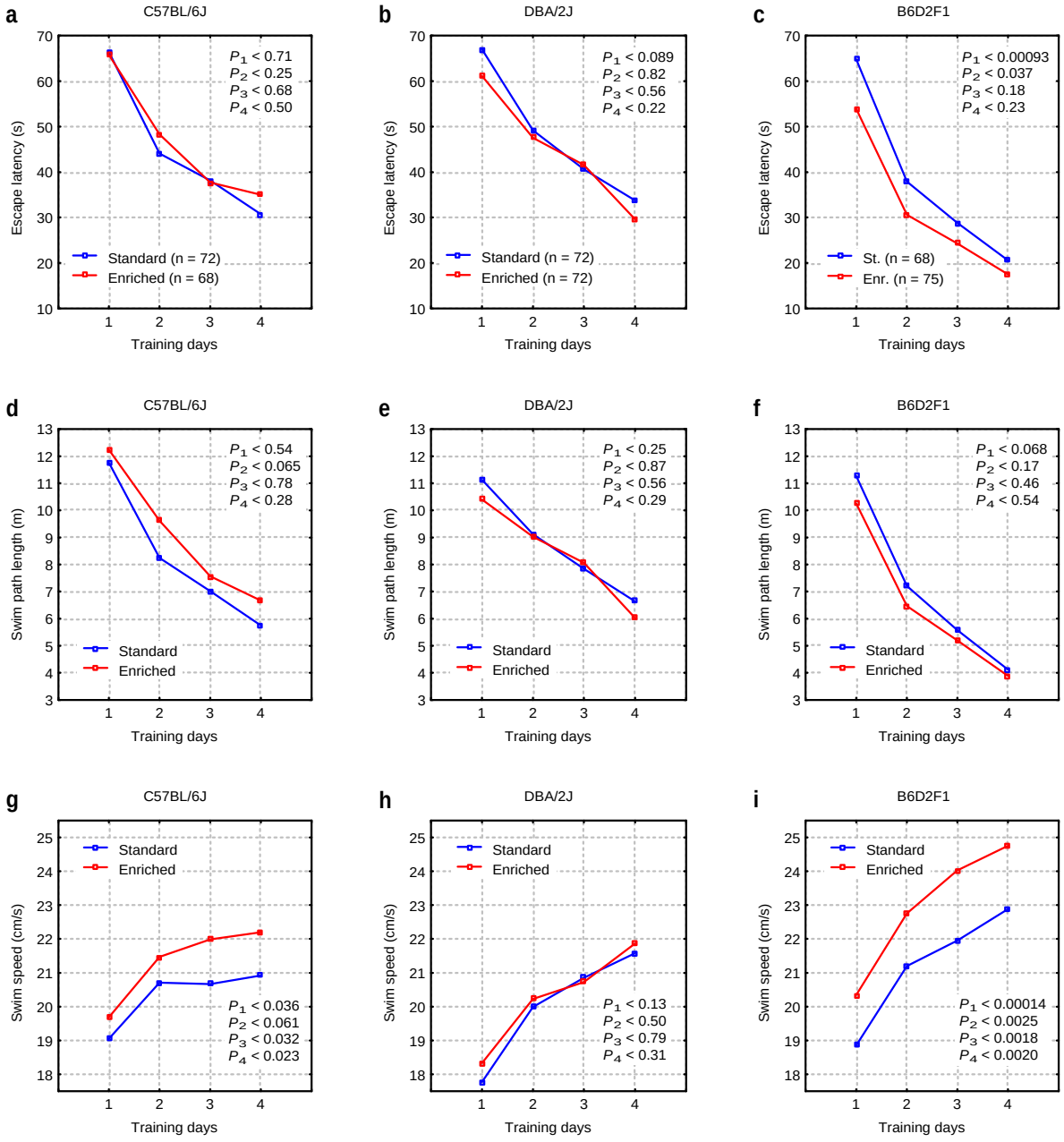



**Figure S90 | Learning curves of standard and enriched mice in the Morris water maze.** Escape latency (**a-c**), swim path length (**d-f**) and swim speed (**g-i**) of standard and enriched C57BL/6J, DBA/2J and B6D2F1 female mice at the age of 73 days. Each time point represents averaged value of 4 consecutive daily training trials. These 4 trials were separated by 30-sec inter-trial intervals (massed training). Note that all 3 variables (latency, path length and speed) produced nice learning curves. However very significant environmental induction of hybrid vigour was observed only in two of them: escape latency (**c**) and swim speed (**i**). Note that enriched hybrids are the best learners with respect to swimming *per se* (**i**).

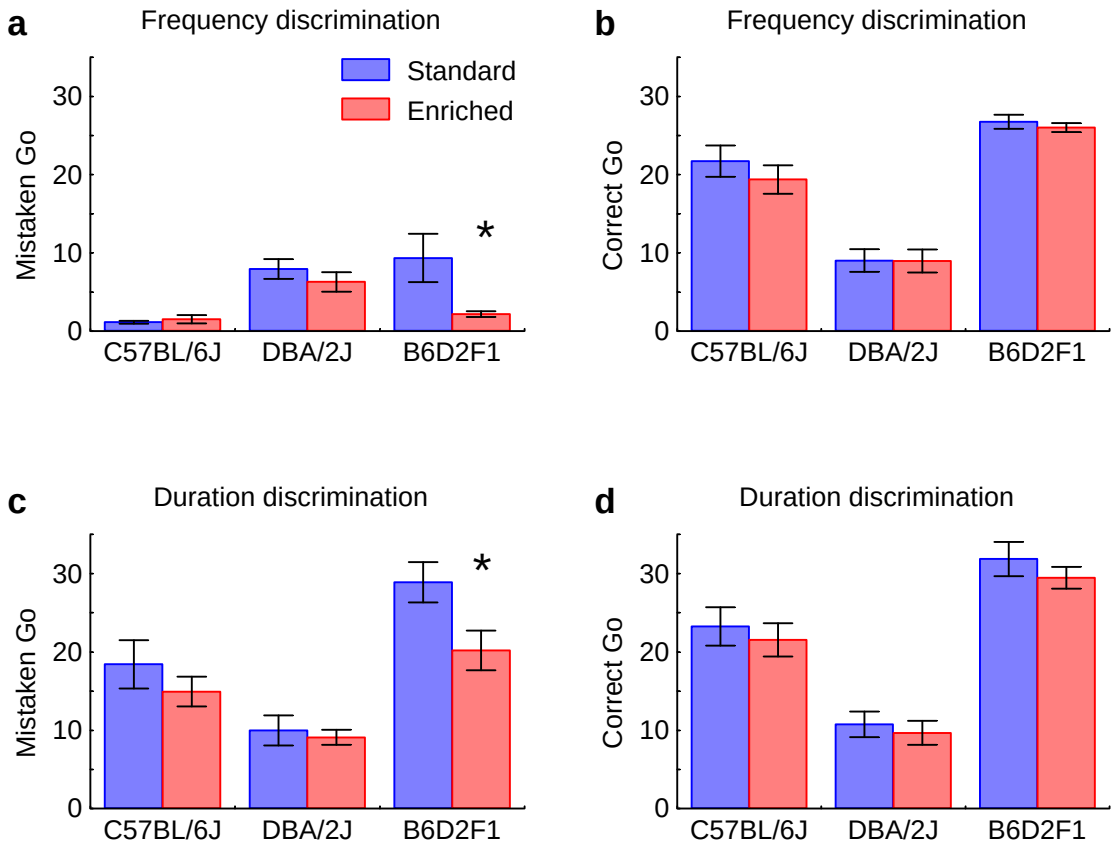



**Figure S91 | Standard and enriched mice in Go/NoGo sound discrimination tasks.** **a-b**, Sound frequency discrimination. **c-d**, Sound duration discrimination. **a**, **c**, Mistaken Go responses. **d**, **b**, Correct Go responses. Averaged values of 7-day training (40 "Go" and 40 "NoGo" trials daily). Note very good sound **frequency** discrimination (**a**, **b**) and relatively poor sound **duration** discrimination (**c**, **d**). However in both tasks enriched $F_1$-hybrids have shown improved overall performance due to decreased number of wrong responses ("Mistaken Go"; **a**, **c**). Mean ± SE. Asterisk, $P < 0.05$.

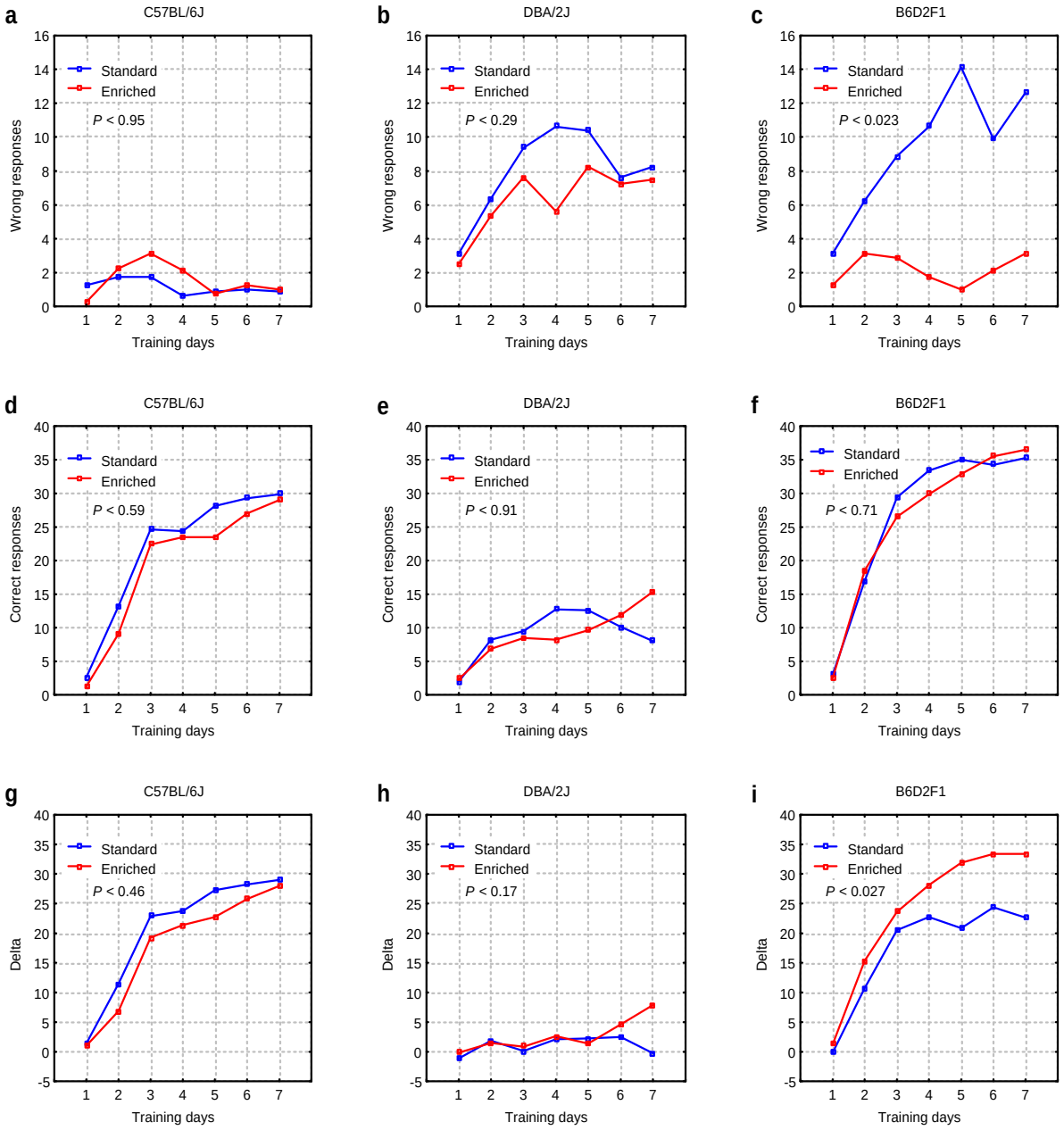



**Figure S92 | Learning curves in Go/NoGo sound frequency discrimination task.** Standard and enriched mice wrong responses (**a-c**), correct responses (**d-f**) and their delta (correct *minus* wrong, **g-i**) in Go/NoGo sound **frequency** discrimination task. C57BL/6J, DBA/2J and B6D2F1 female mice were tested at the age of 7 months after standard or enriched housing conditions during P21-P62. Shown significance is given by Mann-Whitney U-test applied to averaged values of 7-day training (40 Go and 40 NoGo trials daily). Each group consists of 8 animals. Note nice learning curves in C57BL/6J and B6D2F1 mice: sound frequency discrimination is a relatively simple task. The environmental enrichment during P21-P62 has led to enhanced behavioural hybrid vigour in B6D2F1 females at the age of 7 months (**i**). Superior performance of enriched B6D2F1 was achieved due to decreased number of wrong responses (**c**).

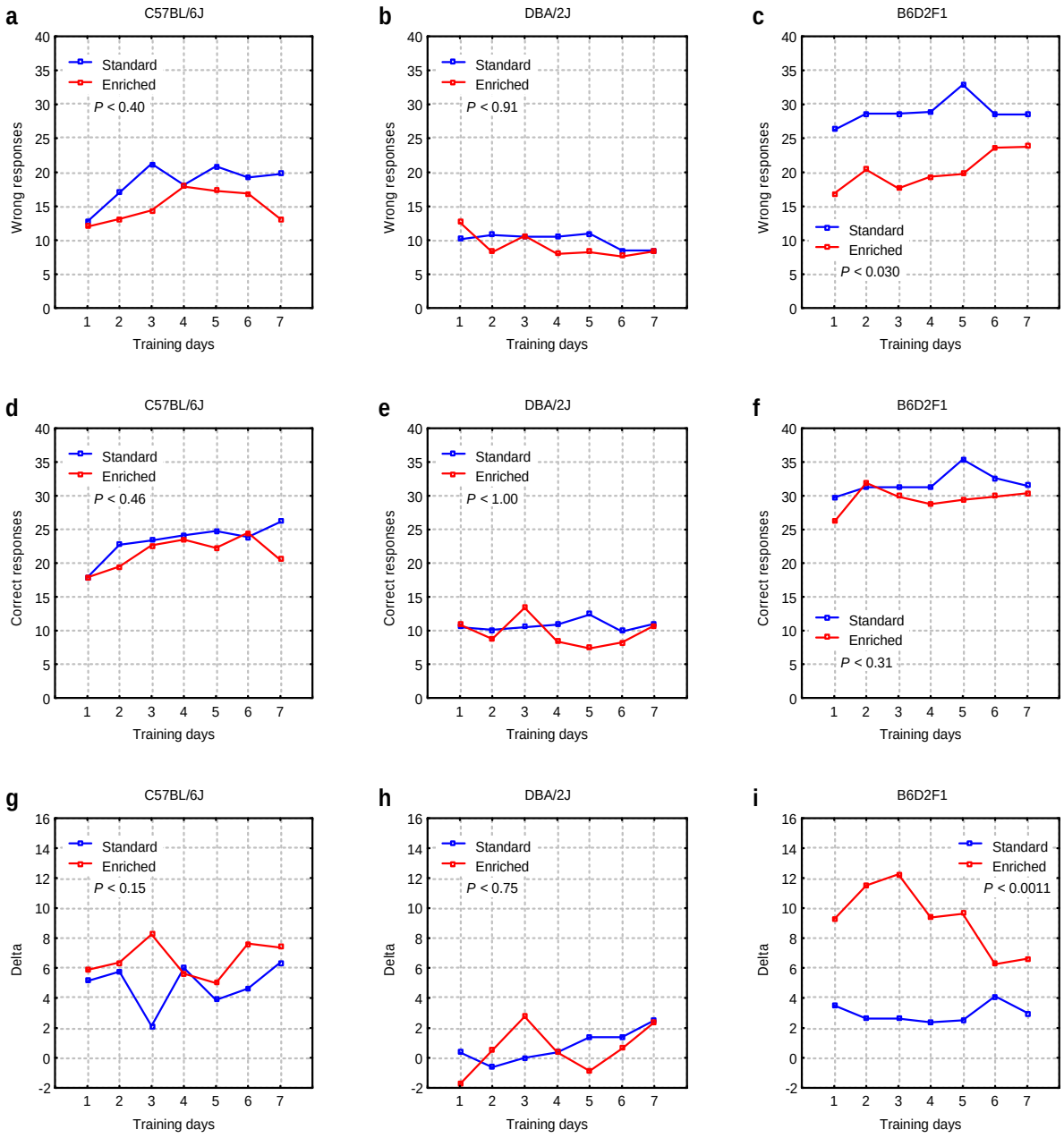



**Figure S93 | Learning curves in Go/NoGo sound duration discrimination task.** Standard and enriched mice wrong responses (**a-c**), correct responses (**d-f**) and their delta (**g-i**) in sound **duration** discrimination task. C57BL/6J, DBA/2J and B6D2F1 female mice at 7.5 months. This task was applied after sound frequency discrimination task to the same animals, after 7-day task-free period. Note that sound duration discrimination task is very difficult – no learning curve and even some decrease in performance from day 3 till day 7 (**i**). However it is the task that has revealed the environmental induction of hybrid vigour with very high (for particular group size, n = 8) statistical significance ($P < 0.0011$). In theory the highest significance for Mann-Whitney U-test and this group size is $P < 0.00078$. Significance $P < 0.0011$ is achieved when only one experimental animal has lover performance than the best performing animal from control group.

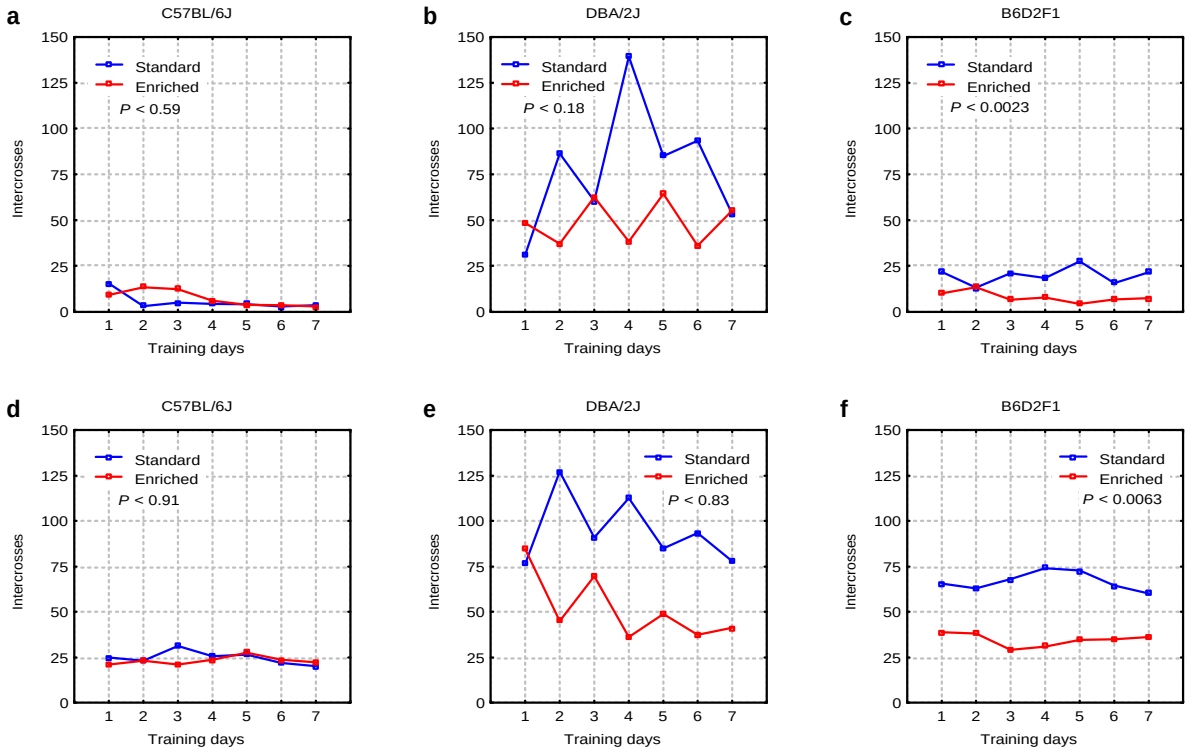



**Figure S94 | Standard and enriched mice intercrosses in Go/NoGo tasks.** Spontaneous intercrosses of standard and enriched C57BL/6J, DBA/2J and B6D2F1 mice in sound frequency (**a-c**) and sound duration (**d-f**) discrimination tasks. An intercross is an event when an animal changes compartment between presentations of stimuli, *i.e.* when neither "Go" nor "NoGo" stimulus is presented. It is entirely spontaneous behaviour which is not linked to any positive or negative reinforcement in the frame of this learning paradigm. DBA/2J mice show very variable intercrosses (**b, e**), because this strain is very jumpy and jumping on the mechanically swinging floor can lead to false-positive intercrosses (one jump – two intercrosses). Note that standard, but not enriched, B6D2F1 mice have rather high number of spontaneous intercrosses in both sound frequency (**c**) and sound duration (**f**) discrimination tasks. In anthropomorphic terms standard B6D2F1 mice are more "nervous" than enriched ones.

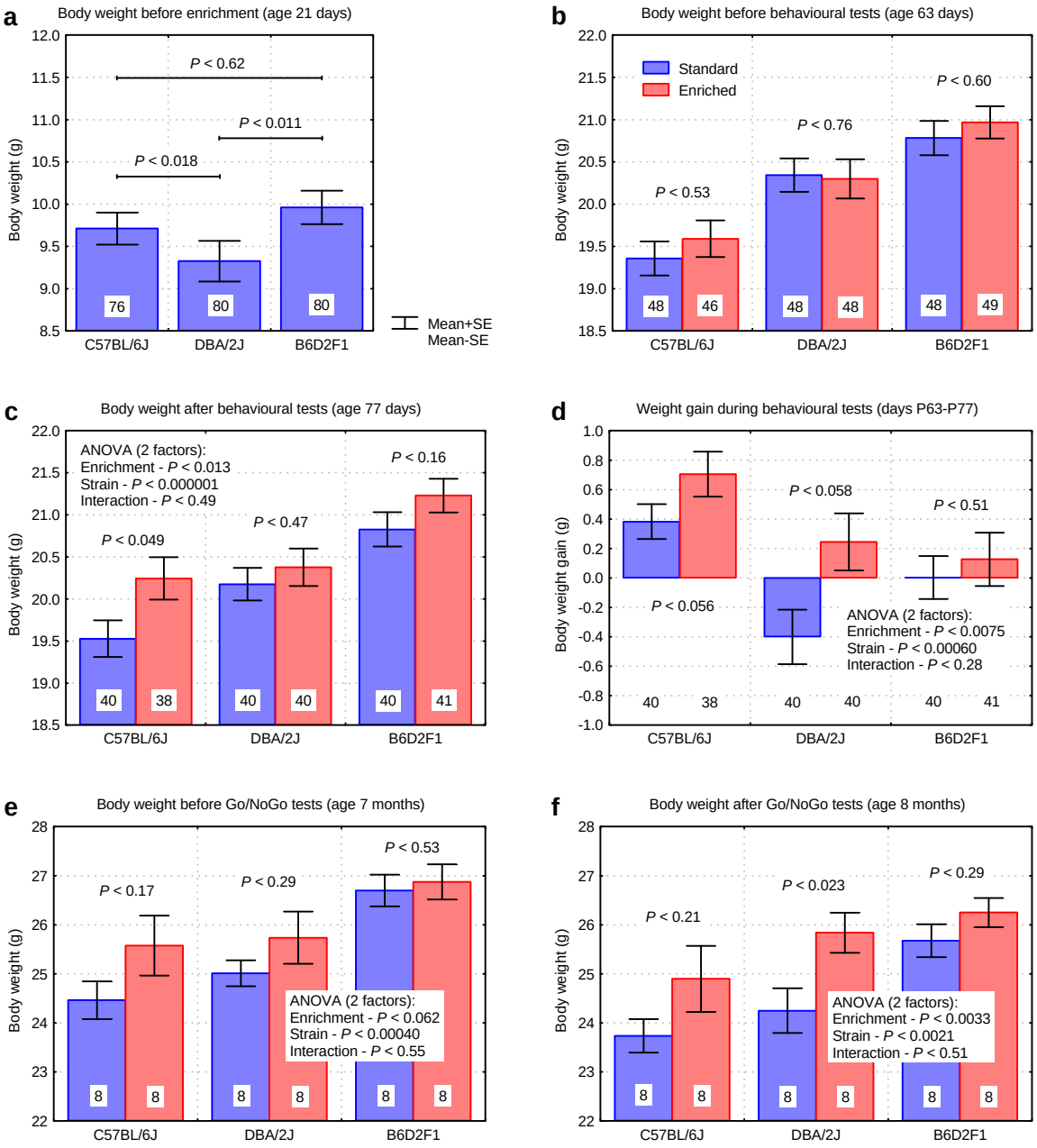



**Figure S95 | Standard and enriched female mice body weight before and after tests.** **a**, Body weight of inbred C57BL/6J, DBA/2J and hybrid B6D2F1 female mice before cage enrichment. **b**, Weight immediately after enrichment and before behavioural tests. **c**, Weight after behavioural tests (O-maze, Open-field, Object exploration, Morris water maze). **d**, Weight gain during above-mentioned tests. **e**, Weight at the age of 7 months, before Go/NoGo tests. **f**, Weight after Go/NoGo sound frequency discrimination and sound duration discrimination tasks. Cage enrichment *per se* during P21-P63 does not change body weight (**b**). But after behavioural testing in Morris water maze (**c**) and Go/NoGo sound discrimination tasks (**f**) body weight of enriched animals occured to be slightly increased in comparison with standard ones. Note that cage enrichment does not enhance hybrid vigour concerning body weight. The enhancement of hybrid vigour by cage enrichment is specific for behaviour and it can be observed in operant behavioural tests.

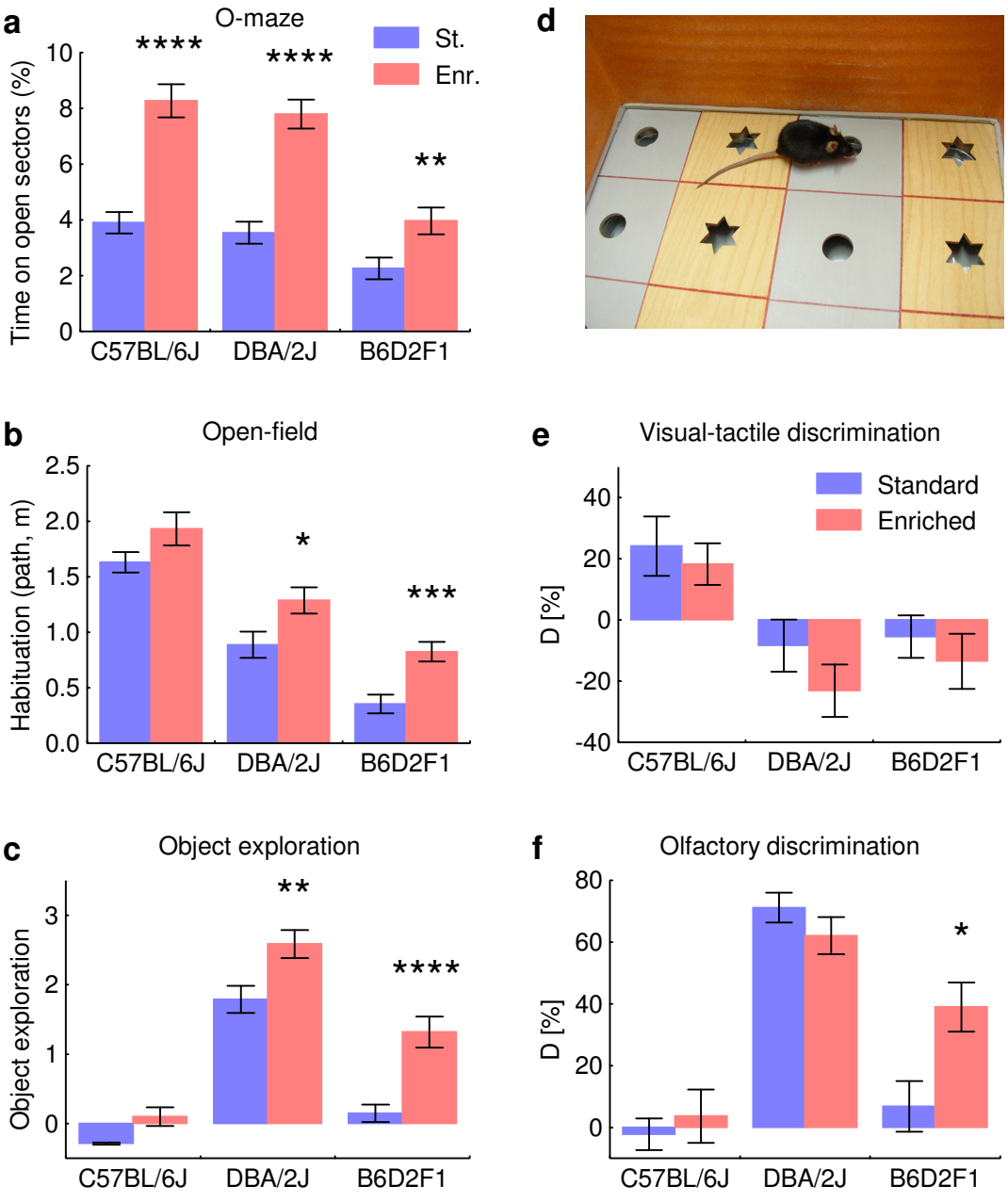



**Figure S96 | Behaviour of standard and enriched mice in the exploratory tasks.** **a**, O-maze. **b**, Open-field. **c**, Object exploration. **d**, Hole-board device for visual-tactile discrimination. **e**, Visual-tactile discrimination. **f**, Olfactory discrimination. The difference in total time of exploration was taken as an indicator of discrimination (**e**-**f**). Standard (blue) and enriched (red) C57BL/6J, DBA/2J and B6D2F1 female mice. Note remarkable effect of enrichment in **a**-**c** and **f** and strong strain-related differences in all tests, but nothing that can be classified as hybrid vigour. Each group consists of 72 mice in **a**-**c** and 8 mice in **e**-**f**. Mean ± SE. Asterisk, $P < 0.05$; double asterisk, $P < 0.01$; triple asterisk, $P < 0.001$; quadruple asterisk, $P < 0.00001$.

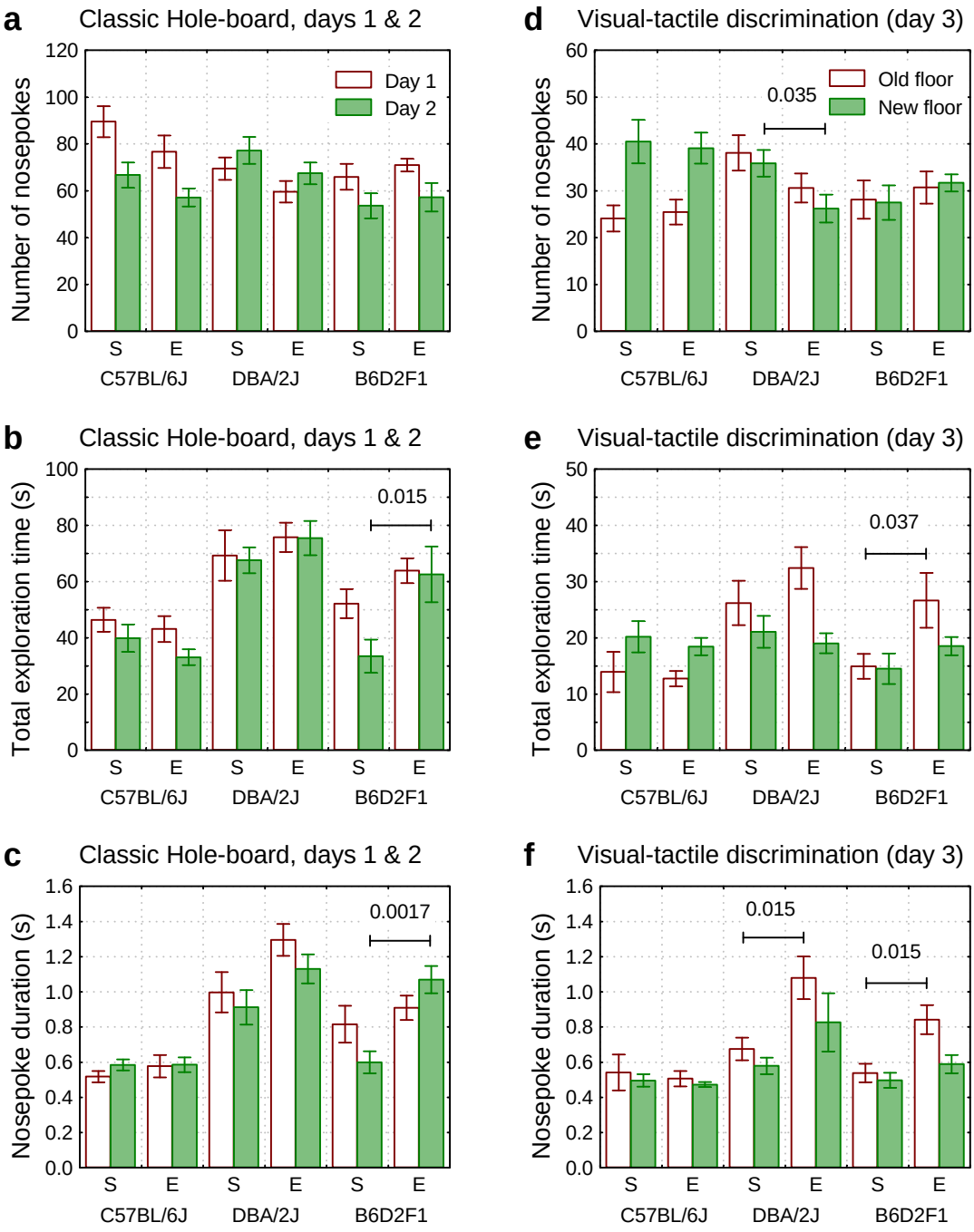



**Figure S97 | Visual-tactile discrimination of standard and enriched mice in the Hole-board task. a-c**, Classic Hole-board, 6 min daily, days 1 & 2 (16 holes, 40x40 cm). **d-f**, Visual-tactile discrimination (see Fig. S96d for details). Number of nosepokes (**a**, **d**), total time of exploration (**b**, **e**) and averaged nose-poke duration (**c**, **f**). S, standard; E, enriched. Note the absence of enrichment effect in the number of nosepokes (**a**), and remarkable increase of nosepoke duration in enriched hybrid B6D2F1 mice during the second day (**c**, green). Note the absence of enrichment effect in the exploration of new floor (**e-f**, green), but significant increase in nosepoke duration in the old floor in enriched DBA/2J and B6D2F1 (**f**, white). Mean ± SE.

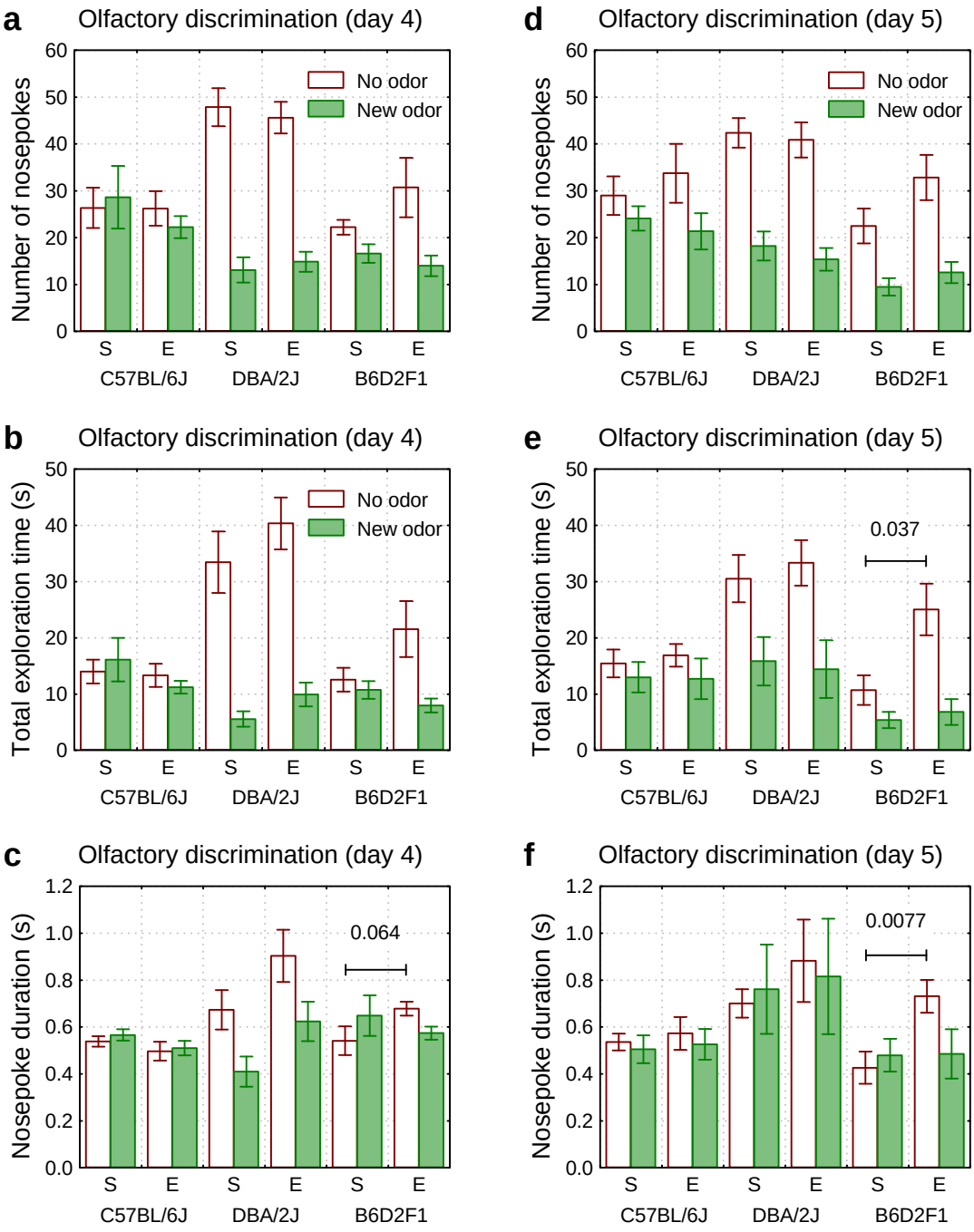



**Figure S98 | Olfactory discrimination of standard and enriched mice in the Hole-board task.** **a-f**, Olfactory discrimination in Hole-board, the first day (**a-c**), and its replication (**d-f**). 16-hole Hole-board, 40 x 40 cm, 6 min daily, under 1/2 of the floor the mint odor was placed (see Methods). Note strong avoidance of new mint odor in all DBA/2J mice and absolute absence of discrimination in all C57BL/6J. Once again, the enrichment effect was observed in the nosepoke duration in hybrid mice, but only concerning the holes without new odor and mainly during the second day of olfactory discrimination (**f**, white). Nevertheless in the Fig. S96f the first day of olfactory discrimination is shown (just for simplicity of interpretation). Mean ± SE.

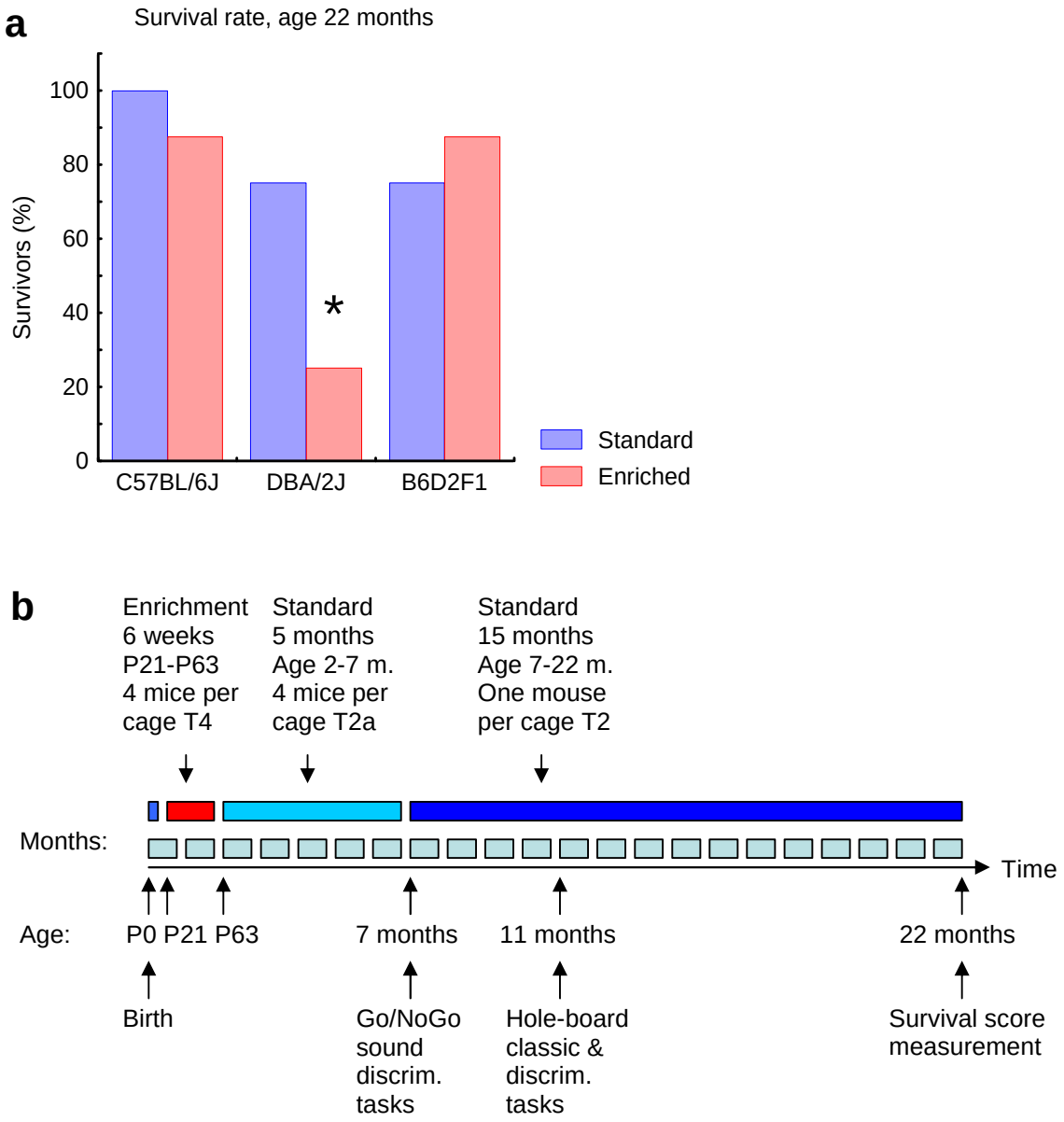



**Figure S99 | Survival rate of standard and enriched mice in usual laboratory conditions. a**, Survivors (%). Asterisk, $P < 0.05$, *chi*-square. **b**, Time frame of the whole experiment with enrichment of housing conditions. Note that enrichment took place only up to the end of the second month, whereas statistically significant effects of enrichment were registered at the ages of 7 and 11 months (see Figures S88b-c, S91-S94 and S96f, S97-S98). And what is more, not all effects of enrichment occurred to be beneficial: in DBA/2J mice the effect of enrichment in combination with further standard housing has led to decreased life span (**a**), probably due to some kind of contradiction between initial enriched developmental conditions and further standard ones.

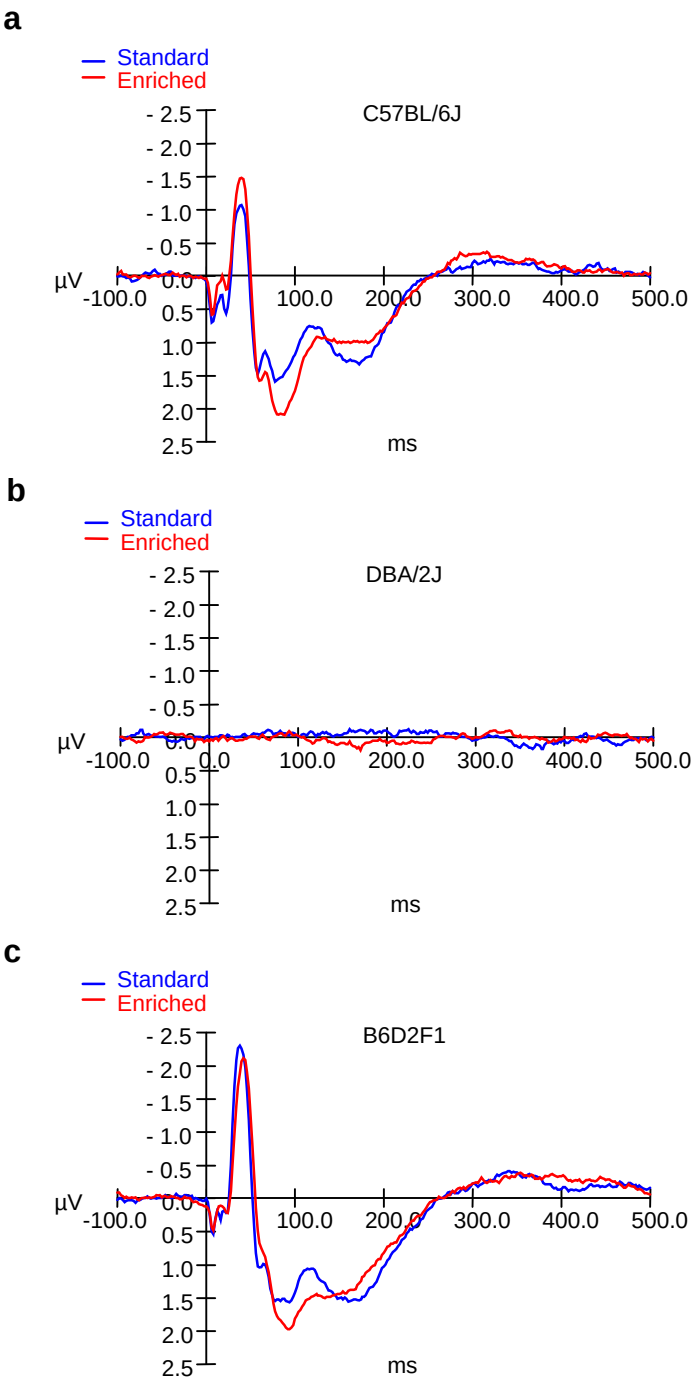



**Figure S100 | Auditory evoked potentials in standard and enriched mice.** Age 7 months. Grand averaged auditory evoked potential waveforms from primary auditory cortex in awake, freely moving female mice after standard (blue) and enriched (red) housing. **a**, C57BL/6J. **b**, DBA/2J. **c**, B6D2F1. Note that, by convention, negative voltage is plotted upwards. Note qualitatively similar changes, induced by enrichment of housing conditions, in the shape of evoked potential in C57BL/6J and B6D2F1. Each curve was formed by 8 mice. 9600 stimuli were averaged for each animal.

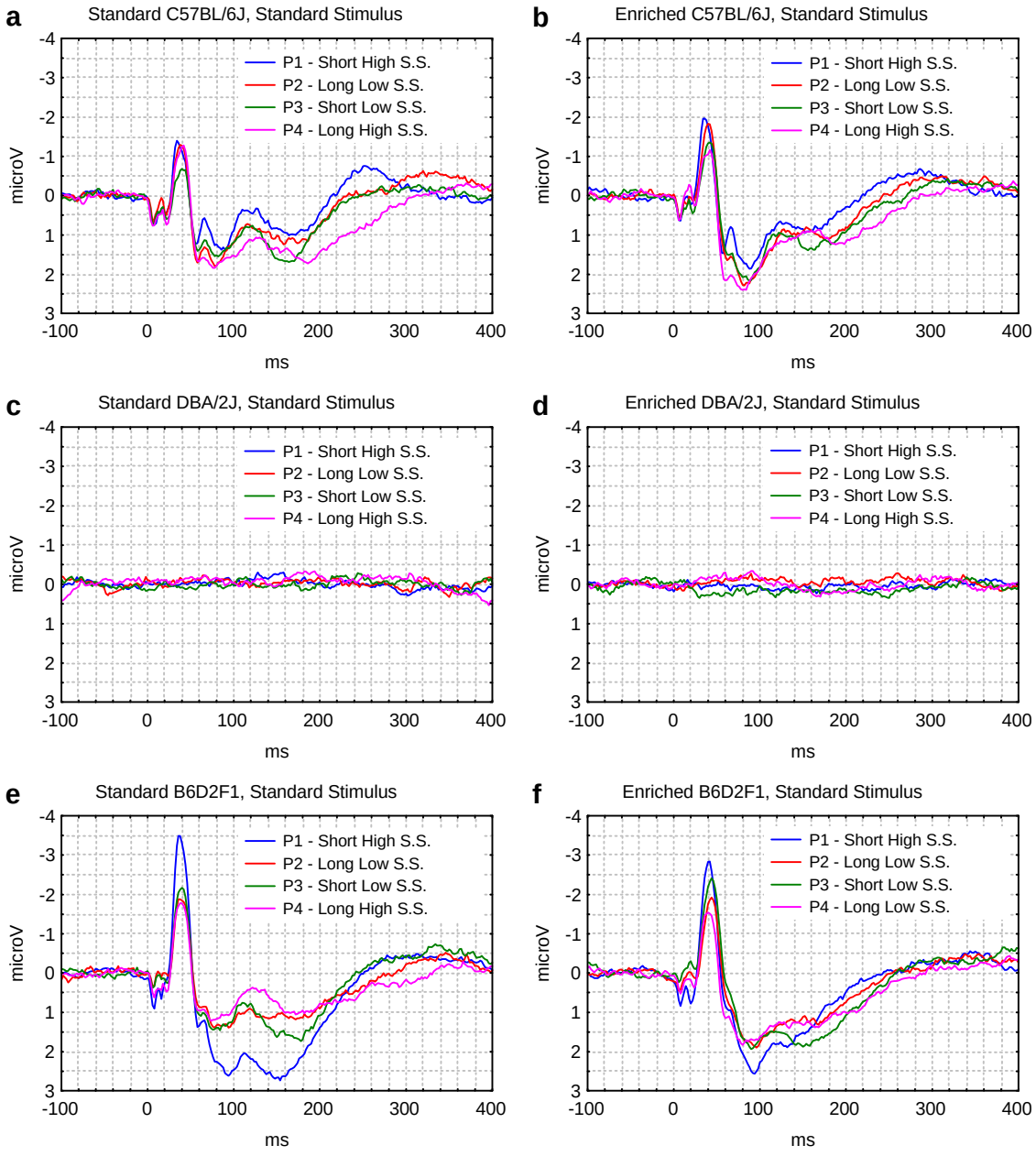



**Figure S101 | Detailed analysis of auditory evoked potentials in standard and enriched mice. a-b**, C57BL/6J. **c-d**, DBA/2J. **e-f**, B6D2F1. **a, c, e,** Standard housing. **b, d, f,** Enriched housing. Paradigms 1 (blue), 2 (orange), 3 (green) and 4 (pink) are shown schematically in the Fig. S47d,e,f,g. Note systematic decrease in N1 amplitude from Paradigm 1 to Paradigm 4. This decrease demonstrates the importance of previous history for particular evoked potential shape. Note dramatic changes in the B6D2F1 standard (**e**) and remarkable stability in the B6D2F1 enriched (**f**).

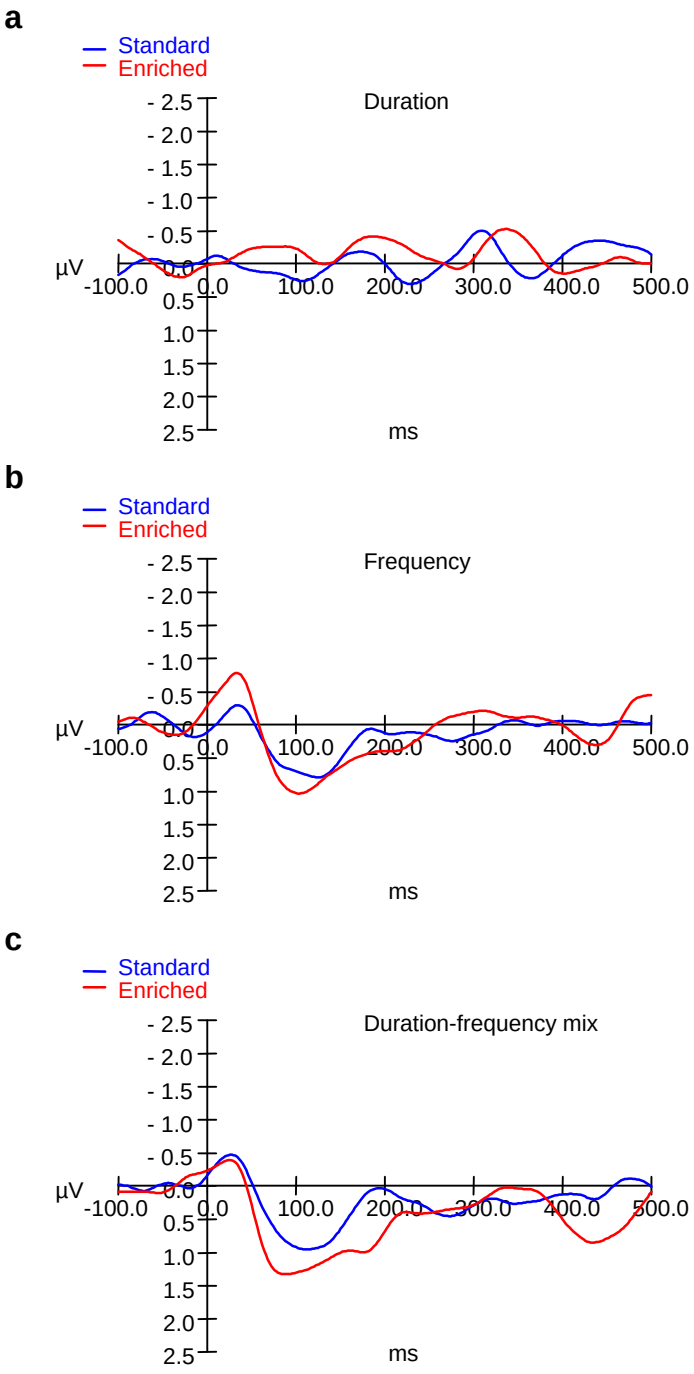



**Figure S102 | Mismatch negativity in standard and enriched $F_1$ hybrid mice (B6D2F1).** Grand averaged difference waves ("deviant stimulus evoked potential" minus "standard stimulus evoked potential") after standard (blue) and enriched (red) housing. **a**, Duration deviant paradigm. **b**, Frequency deviant paradigm. **c**, Mixed (duration and frequency) deviant paradigm. Note increased amplitude of deviant-related activity in the enriched animals in all paradigms. Negative deflection in 140-200 ms time window (**a**) can be classified as mismatch negativity. Each curve was formed by 8 mice. Each animal was exposed to 400 deviant stimuli of each type.

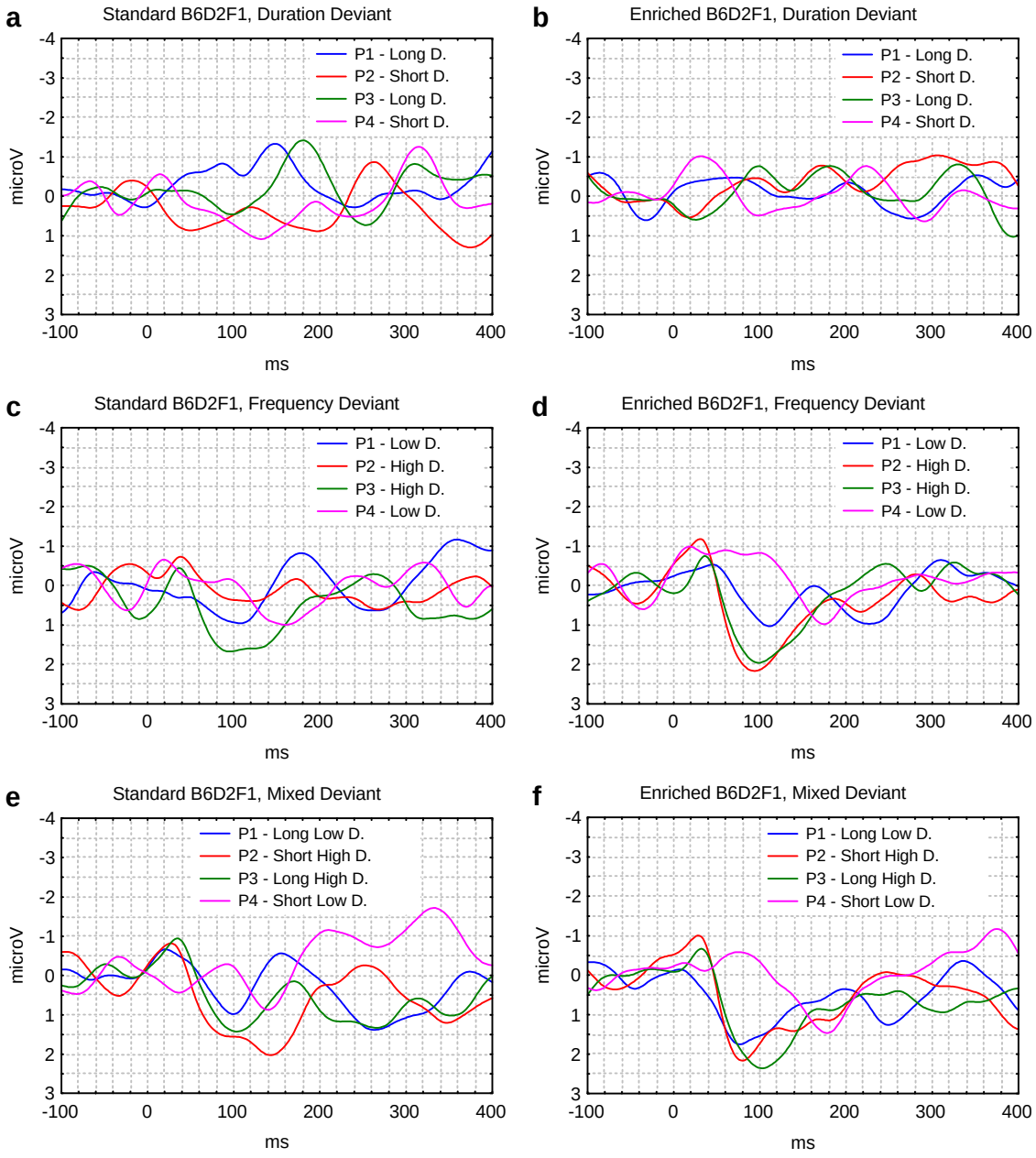



**Figure S103 | Detailed analysis of mismatch negativity in standard and enriched B6D2F1 mice. a-b**, Duration mismatch negativity. **c-d**, Frequency mismatch negativity. **e-f**, Mixed (duration and frequency) mismatch negativity. **a, c, e,** Standard housing. **b, d, f,** Enriched one. Note mirror-shaped curves of short- and long-deviant mismatch negativity in standard B6D2F1 (**a**), *i.e.* negative deflection in long-deviant paradigms and symmetric positive deflection in the short-deviant ones. Grand average of all these paradigms does not reveal novelty detection in standard B6D2F1, but we can see this detection in the enriched ones (Fig. S102a).

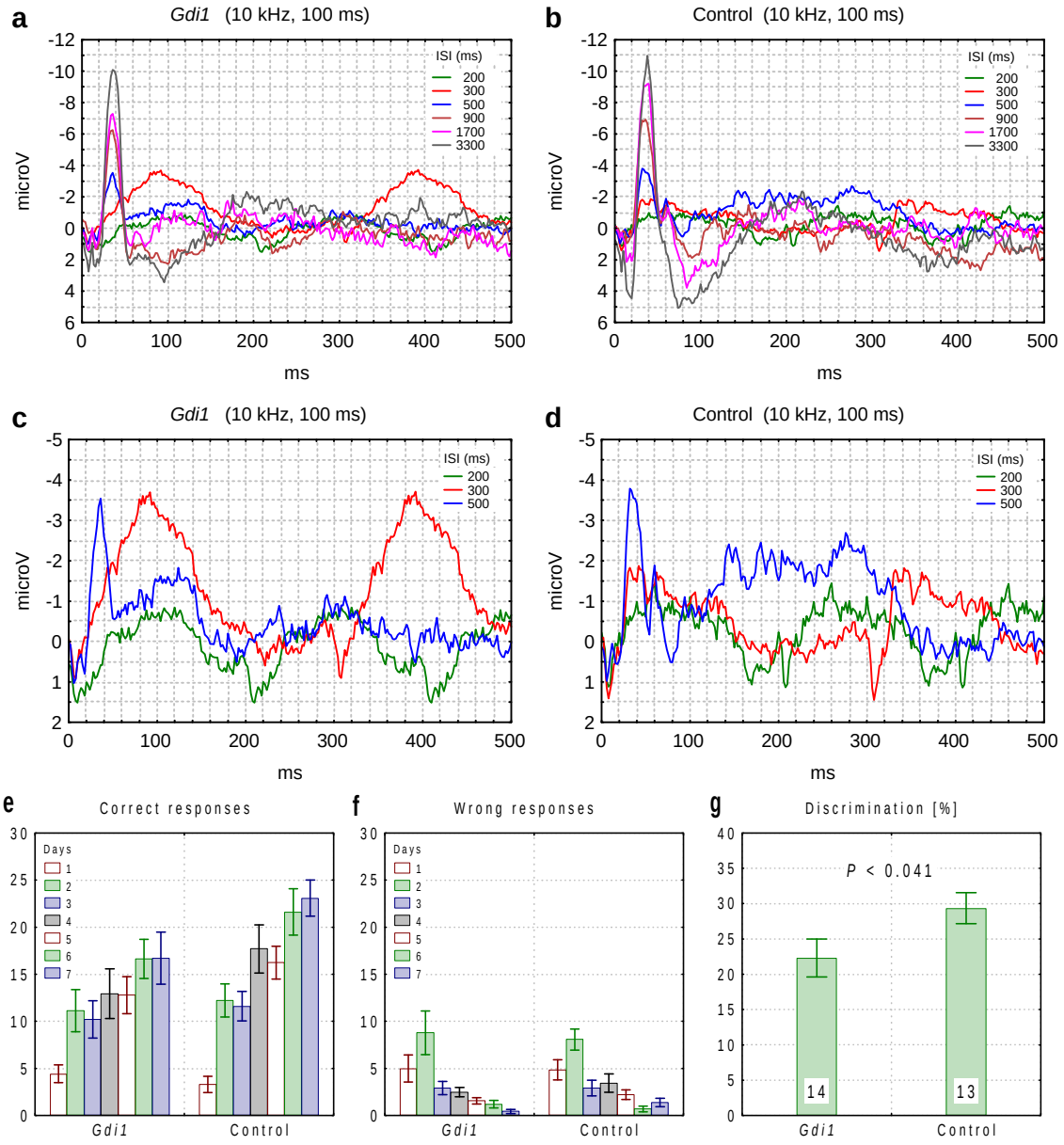



**Figure S104 | Auditory evoked potentials and frequency discrimination in *Gdi1*-deficient male mice.** **a-d**, Auditory evoked potentials in freely moving *Gdi1*-mutant male mice (age 7 months). Frequency 10 kHz, stimulus duration 100 ms. Paradigms with different inter-stimulus intervals (ISI; onset-to-onset), Fig. S46c-h. Background C57BL/6J, 8 mice per group. Note enhanced and diffused evoked potential in the *Gdi1*-knockout mice with ISI 300 ms (**c**, red; *Gdi1 vs.* Control, $P < 0.01$ in 94-108 ms, $P < 0.05$ in 78-124 ms). Note enhanced diffused part of evoked potential in control mice with ISI 500 ms (**d**, blue; *Gdi1 vs.* Control $P < 0.05$ in 160-230 ms). See qualitatively similar diffused evoked potentials in mutant and control mice with ISI 200 ms (**c**-**d**, green; *Gdi1 vs.* Control, N.S.). **e**-**g**, *Gdi1* male mice in Go/NoGo sound frequency discrimination task (before EEG recording; age 3 months); 2.5 & 10 kHz *vs.* 5 & 5 kHz (see Methods). Note slight decrease in the averaged performance of *Gdi1* mice (**g**). Mean ± SE.

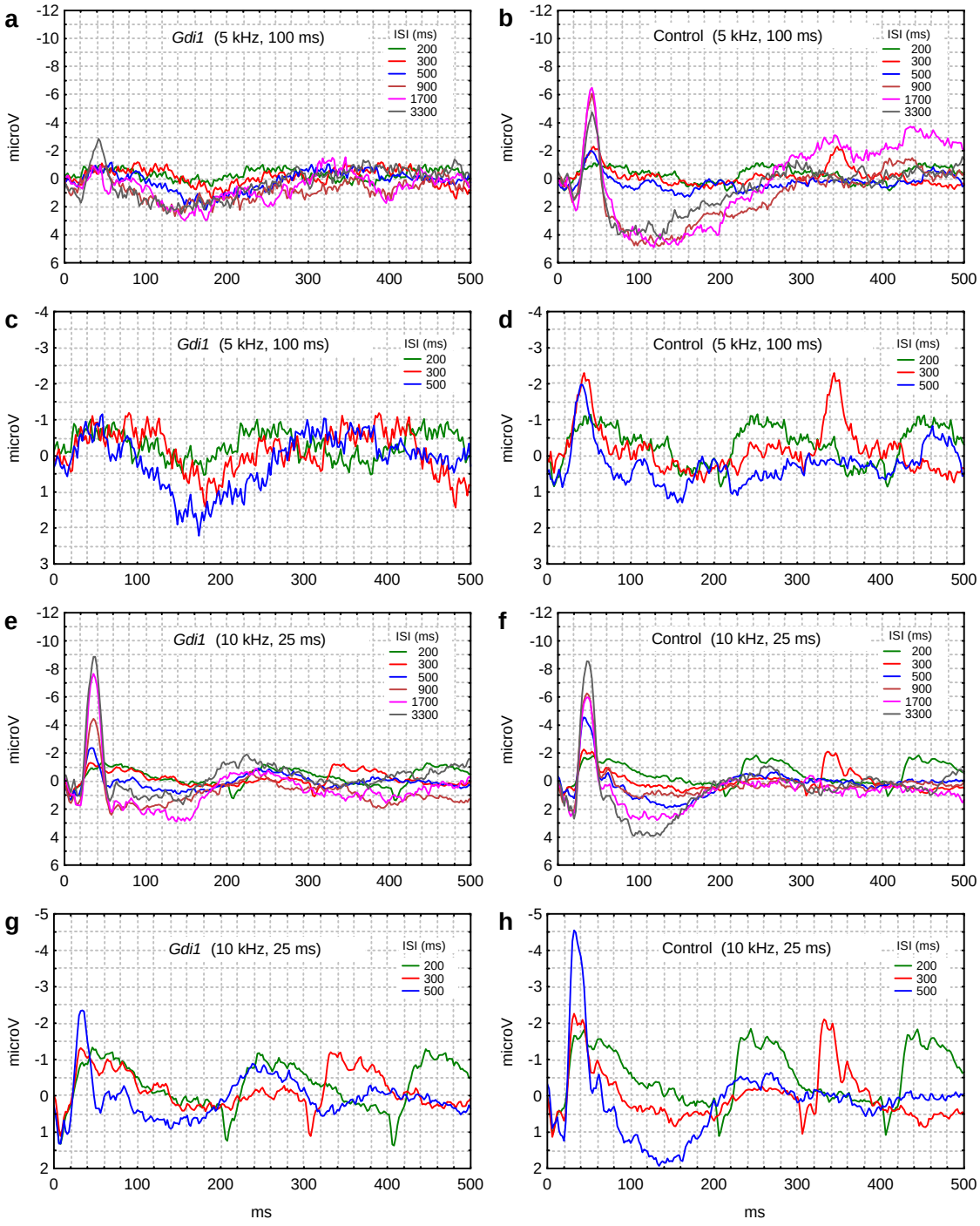



**Figure S105 | Auditory evoked potentials in *Gdi1*-deficient mice induced by non-optimal stimuli. a-d**, Low frequency (5 kHz) stimuli. **e-h**, Short duration (25 ms) stimuli. The other parameters of applied paradigms with different inter-stimulus intervals (ISI) are shown in the Fig. S46c-h. **a**, **c**, **e**, **g**, *Gdi1*. **b**, **d**, **f**, **h**, Control. There are 8 mice in a group. Note that differences between mutant and control mice are only quantitative, but not qualitative. *I.e.* using different parameters of stimulation one can obtain the same evoked potential shape in mutants and controls.

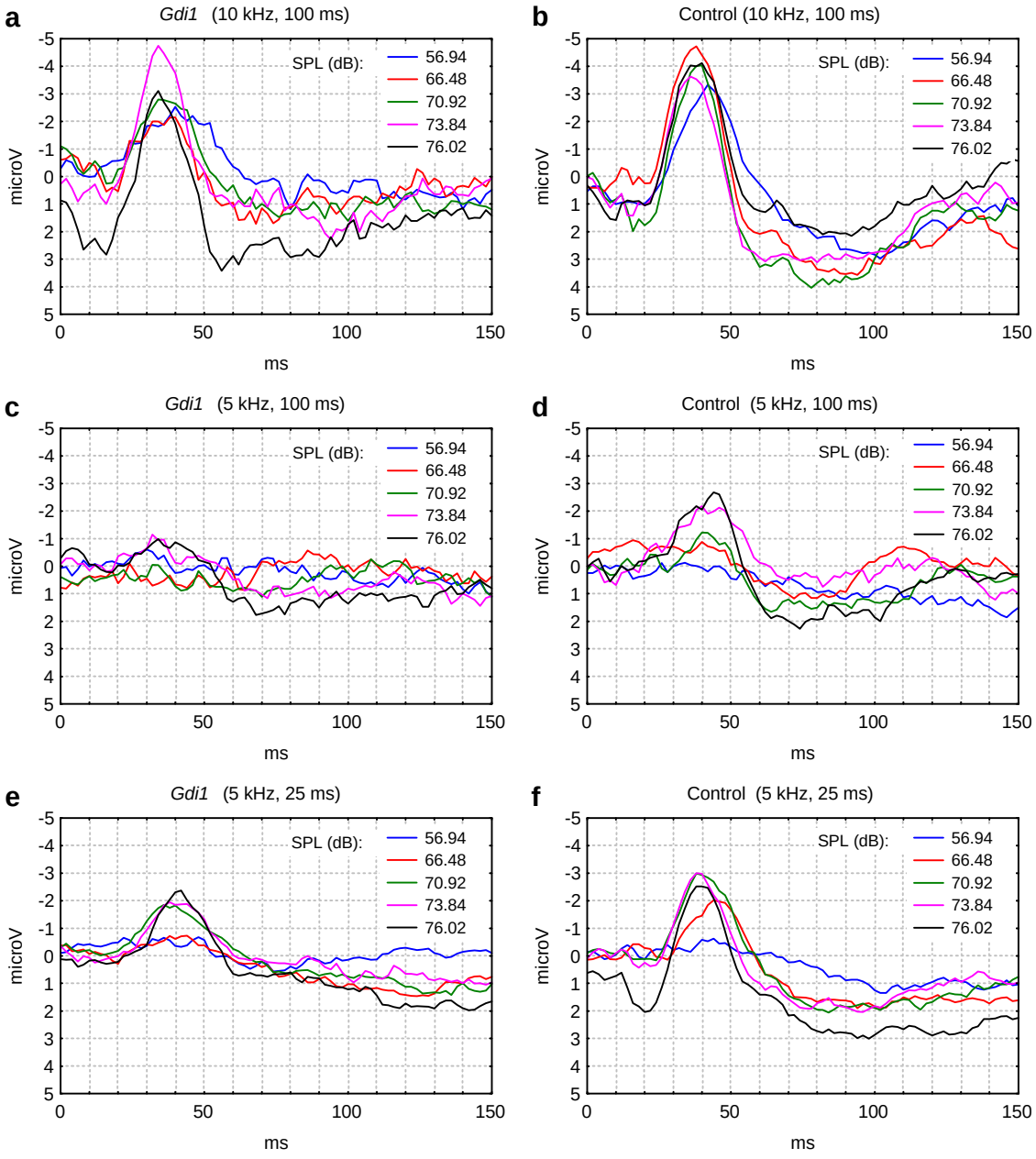



**Figure S106 | Auditory evoked potentials in *Gdi1* mice induced by different sound level stimuli.** Inter-stimulus interval (ISI) was 1000 ms (onset-to-onset). SPL (dB) is a sound pressure level; reference effective pressure (*re*) is 20 µPa, as usual. **a-b**, Frequency 10 kHz, duration 100 ms. **c-d**, Frequency 5 kHz, duration 100 ms **e-f**, Frequency 5 kHz, duration 25 ms. During each paradigm an animal was exposed to 300 stimuli. **a**, **c**, **e**, *Gdi1*. **b**, **d**, **f**, Control. 8 mice in a group. Note that evoked potentials induced by optimal frequency (10 kHz) are SPL-independent in the 56-76 dB range (**a-b**), whereas non-optimal frequency (5 kHz) induces evoked potentials, which are dependent not only on SPL, but on stimulus duration also (**c-f**). Short stimulus (25 ms; **e-f**) sometimes can produce better and less noisy evoked potential, than long one (100 ms, **c-d**).

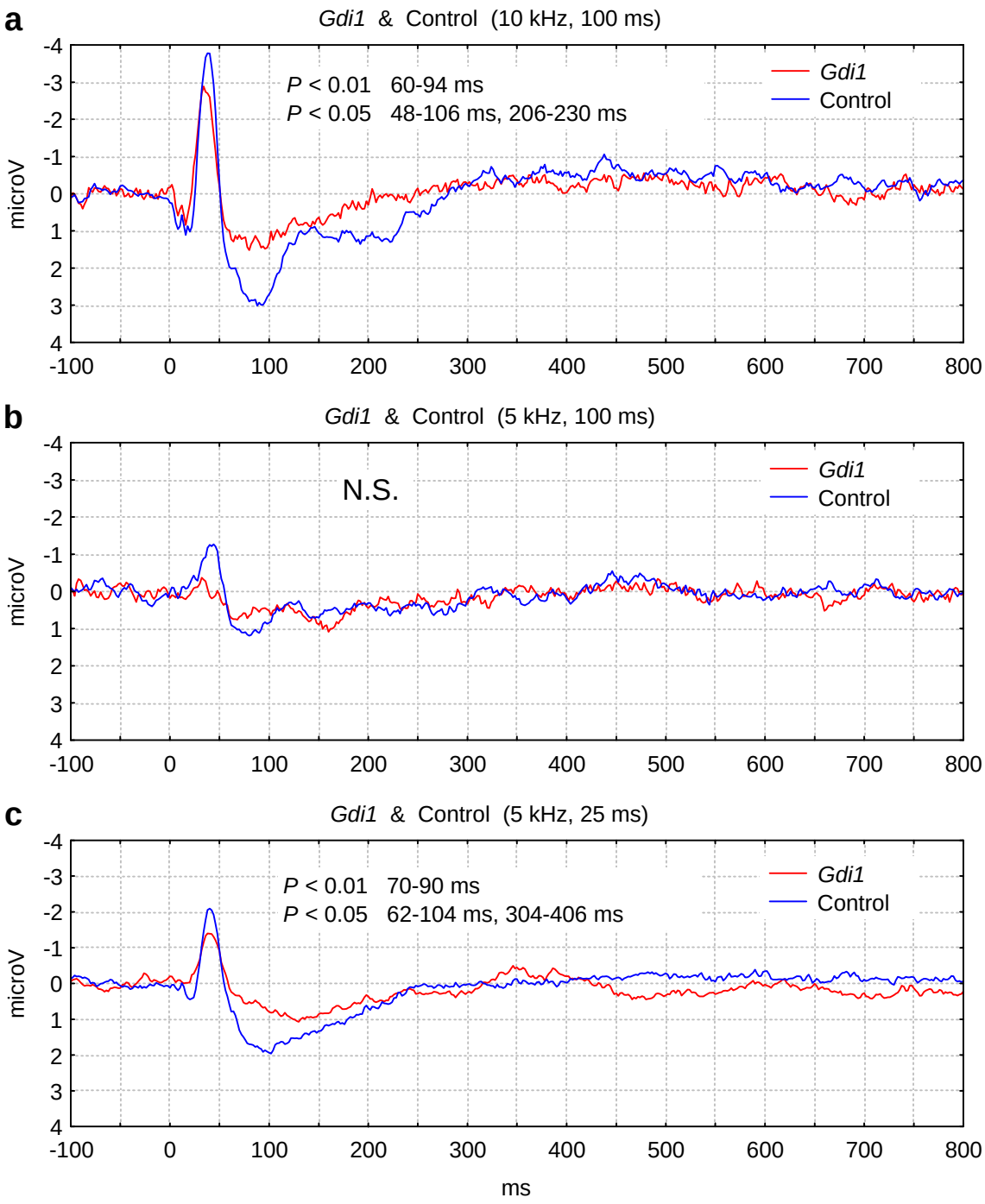



**Figure S107 | Auditory evoked potentials in *Gdi1* mice, averaged curves of all sound level stimuli**. Each curve represents mean value of 5 curves obtained in different sound pressure level (SPL) paradigms shown in the Fig. S106. SPL 56-76 dB. **a**, Frequency 10 kHz, duration 100 ms. **b**, Frequency 5 kHz, duration 100 ms. **c**, Frequency 5 kHz, duration 25 ms. Inter-stimulus interval 1000 ms. There are 8 mice in a group. 1500 stimuli were averaged for each curve (5 × 300). Note decreased positive deflection in *Gdi1* in 70-90 ms time window. Mann-Whitney U-test.

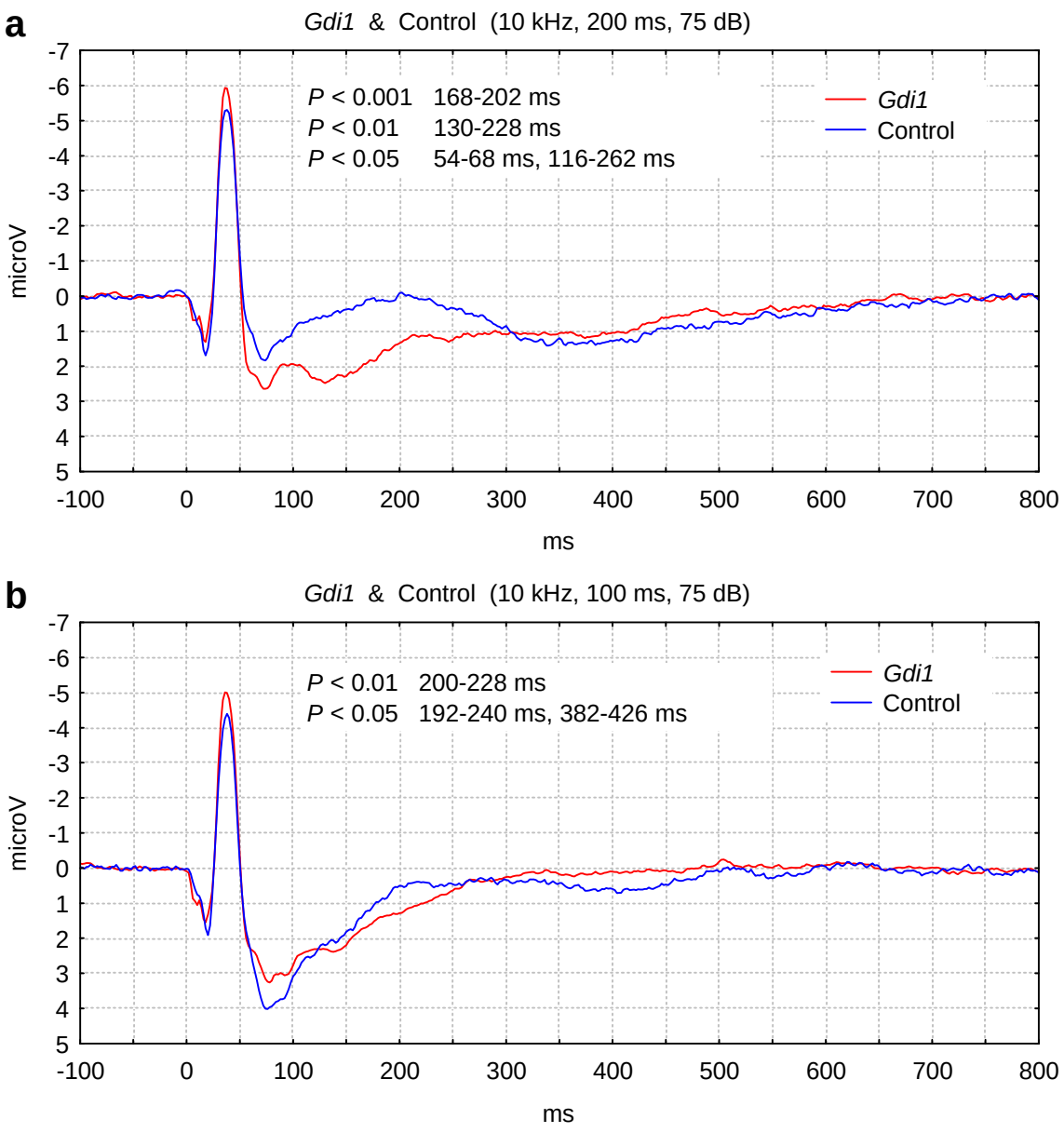



**Figure S108 | Auditory evoked potentials in *Gdi1*-deficient mice induced by relatively long stimuli. a**, Long duration (200 ms) stimuli. **b**, Usual (100 ms) stimuli. Frequency 10 kHz, sound pressure level (SPL) 75 dB. Inter-stimulus interval (ISI) was 900 ms (onset-to-onset). 4440 stimuli were averaged for each curve. There are 7 mice in the experimental group and 4 mice in the control one. Note that, contrary to experiment with different (56-76 dB) sound level stimuli (Fig. S107a), in the paradigm with 200 ms stimuli *Gdi1* mice have increased positive deflection (**a**), however it belongs to relatively late time period (130-228 ms). In the paradigm with 100 ms stimuli (**b**) above-mentioned period is shorter (200-228 ms), whereas in the earlier period (60-94 ms) *Gdi1* mice have decreased positive deflection (**b**), similar to paradigm with different sound level stimuli (Fig. S107a). Mann-Whitney U-test.

## Tables S1 to S4

**Table S1**
**The importance of time interval between L-thyroxine solution preparation and injection**

| The 1-st L-thyroxine injection done **10 min** after preparation, pilot exp. | | | | The 1-st L-thyroxine injection done **24 hr** after preparation, main exp. | | | | Control - untreated | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| P00 | P01 | P02 | P03 | P00 | P01 | P02 | P03 | P00 | P01 | P02 | P03 |
| 1.48 | 1.37 | 1.26 | | 1.38 | 1.59 | 1.84 | 2.08 | 1.43 | 1.49 | 1.73 | 2.02 |
| 1.44 | 1.22 | | | 1.37 | 1.53 | 1.71 | 1.99 | 1.38 | 1.48 | 1.66 | 1.98 |
| 1.44 | 1.21 | | | 1.31 | 1.47 | 1.69 | 1.96 | 1.38 | 1.40 | 1.64 | 1.97 |
| 1.30 | 1.13 | | | 1.26 | 1.46 | 1.67 | 1.96 | 1.14 | 1.19 | 1.44 | 1.75 |
| 1.01 | | | | 1.22 | 1.41 | 1.64 | 1.89 | | | | |
| | | | | 1.16 | 1.40 | 1.61 | 1.84 | | | | |
| | | | | 1.13 | 1.36 | 1.55 | 1.80 | | | | |

P00-P03 – the first 4 postnatal days. P generation pup body weight (g) and neonatal mortality in the pilot and main experiments. In the main experiment prepared L-thyroxine solution was stored at +4 °C during 24 hours before the first (day P00) administration (during 48 hr before the second one, during 72 hr before the 3-d one, *etc*). Note that the L-thyroxine injection done 10 min after solution preparation (at P00, pilot exp.) has lead to 100% mortality (at P03). Similar deaths were not observed if above-mentioned time interval was 24 hours. Three different litters are shown (the last one – control) with 5, 7 and 4 pups respectively.

**Table S2**
**Phenotype of $F_3$ descendants of thyroxine-treated male mice**
**Mann-Whitney U-test**

| | | p-level | Valid N Thyroxine | Valid N Control | MEAN Thyroxine | MEAN Control | SE Thyroxine | SE Control |
|---|---|---|---|---|---|---|---|---|
| $F_3$ males | Birthweight | **0.0462** | 24 | 20 | 1.228 | 1.327 | 0.021 | 0.035 |
| from | Shuttle-box | 0.6006 | 23 | 20 | 41.82 | 38.79 | 3.31 | 3.21 |
| $F_2$-incross | IIP/SP MF | 0.8773 | 9 | 4 | 0.1162 | 0.1165 | 0.0062 | 0.0071 |
| $F_3$ males | Birthweight | **0.0017** | 27 | 28 | 1.201 | 1.271 | 0.023 | 0.017 |
| from | Shuttle-box | **0.0464** | 26 | 28 | 34.53 | 43.27 | 3.23 | 2.41 |
| $F_2$-outcross | IIP/SP MF | **0.0253** | 7 | 7 | 0.0981 | 0.1198 | 0.0052 | 0.0087 |
| $F_3$ females | Birthweight | **0.0051** | 28 | 20 | 1.175 | 1.287 | 0.023 | 0.025 |
| from | Shuttle-box | 0.2814 | 28 | 20 | 35.09 | 39.58 | 2.52 | 2.99 |
| $F_2$-incross | IIP/SP MF | 0.6309 | 6 | 6 | 0.1011 | 0.1127 | 0.0067 | 0.0118 |
| $F_3$ females | Birthweight | **0.0306** | 29 | 25 | 1.186 | 1.256 | 0.019 | 0.023 |
| from | Shuttle-box | 0.8543 | 28 | 24 | 38.53 | 38.91 | 2.44 | 2.62 |
| $F_2$-outcross | IIP/SP MF | 0.2774 | 7 | 7 | 0.0976 | 0.0854 | 0.0066 | 0.0054 |

Birthweight – body weight (g)
Shuttle-box – number of correct responses (maximum 80), mean value of 5-day two-way avoidance training
IIP/SP MF – ratio of intra- and infrapyramidal to suprapyramidal hippocampal mossy fiber area, horizontal brain sections

**Table S3**

**Hippocampal mossy fibers in the $F_3$ descendants of thyroxine-treated mice, ANOVA**

IIP/SP mossy fiber ANOVA: Summary of all Effects
1 - thyroxin *vs.* control, 2 - incross *vs.* outcross, 3 - males *vs.* females

| Factors | Df Effect | MS Effect | df Error | MS Error | F | p-level |
|---|---|---|---|---|---|---|
| 1 | 1 | 0.000368 | 45 | 0.000365 | 1.00814 | 0.32071 |
| 2 | 1 | 0.001658 | 45 | 0.000365 | 4.54856 | **0.03843** |
| 3 | 1 | 0.002287 | 45 | 0.000365 | 6.27164 | **0.01596** |
| 1, 2 | 1 | 3.98E-06 | 45 | 0.000365 | 0.01091 | 0.91725 |
| 1, 3 | 1 | 0.000414 | 45 | 0.000365 | 1.13426 | 0.29255 |
| 2, 3 | 1 | 0.000200 | 45 | 0.000365 | 0.54790 | 0.46301 |
| 1, 2, 3 | 1 | 0.001627 | 45 | 0.000365 | 4.46205 | **0.04023** |

**Table S4**

**Synaptophysin brain distribution in the $F_2$ male descendants of morphine-treated male rats and in their drug-naive and morphine-treated controls**

| Origin | Control $F_2$ | Experimental $F_2$ | Control | Control |
|---|---|---|---|---|
| Treatment | Drug-naive | Drug-naive | After 6-day morphine treatment | After 6-day morphine treatment and 6-day morphine withdrawal |
| Valid N | 4 | 4 | 4 | 4 |
| SMC III | 35.2 ± 1.1 | 31.3 ± 1.7 | | |
| SMC V | 32.5 ± 2.9 | 30.8 ± 2.2 | | |
| SSC III | 46.3 ± 0.8 | **34.1 ± 0.9*** | | |
| SSC V | 39.3 ± 1.5 | **32.0 ± 1.7*** | | |
| NA | 41.3 ± 0.6 | **30.1 ± 1.7*** | 37.5 ± 3.1 | **32.9 ± 1.4*** ($P < 0.011$) |
| HIPP | 35.8 ± 1.8 | **22.1 ± 2.9*** | 30.3 ± 2.7 | **27.0 ± 2.6*** ($P < 0.018$) |
| NC | 43.2 ± 2.5 | 42.8 ± 1.6 | 44.6 ± 1.8 | 40.5 ± 1.9 |
| THAL | 27.8 ± 3.1 | 27.4 ± 1.6 | | |

Fluorescent area, %, Mean ± SE. Synaptophysin (p38) is represented as relative area (in %) of pixels occupied by synaptophysin with intensity value above a background threshold. Control animals shown in the last two columns were treated with morphine using 6-day protocol (morning-evening, 12 hr interval; mg/kg): 10-10, 20-20, 30-30, 40-40, 50-50, 60-60.

SMC III, somatomotor cortex, layer 3; SMC V, somatomotor cortex, layer 5; SSC III, somatosensory cortex, layer 3; SSC V, somatosensory cortex, layer 5; NA, nucleus accumbens; HIPP, hippocampus; NC, nucleus caudatus; THAL, ventrolateral thalamic nuclei. Asterisk, $P < 0.05$; all statistical comparisons with "Control $F_2$, Drug-naive" (the first column).

## References


Allen, N.D., Norris, M.L. & Surani, M.A. Epigenetic control of transgene expression and imprinting by genotype-specific modifiers. *Cell* **61**, 853-861 (1990).

Anway, M.D., Cupp, A.S., Uzumcu, M. & Skinner, M.K. Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science* **308**, 1466-1469 (2005).

Arjune, D. & Bodnar, R.J. Post-natal morphine differentially affects opiate and stress analgesia in adult rats. *Psychopharmacol*. (Berl). **98**, 512-517 (1989).

Bachner, D., Sedlacek, Z., Korn, B., Hameister, H. & Poustka, A. Expression patterns of two human genes coding for different rab GDP-dissociation inhibitors (GDIs), extremely conserved proteins involved in cellular transport. *Hum. Mol. Genet*. **4**, 701-708 (1995).

Bakke, J.L., Lawrence, N.L., Bennett, J. & Robinson, S. in *Perinatal thyroid physiology and disease* (eds Fisher, D.A. & Burrow, G.H.) 79-116 (Raven Press, New York, 1975).

Bakke, J.L., Lawrence, N.L., Robinson, S. & Bennett, J. Observations on the untreated progeny of hypothyroid male rats. *Metabolism* **25**, 437-444 (1976).

Belichenko, P.V., Fedorov, A.A. & Dahlstrom, A.B. Quantitative analysis of immunofluorescence and lipofuscin distribution in human cortical areas by dual-channel confocal laser scanning microscopy. *J. Neurosci. Meth.* **69**, 155-161 (1996).

Belichenko, P.V., Brown, D., Jeffrey, M. & Fraser, J.R. Dendritic and synaptic alterations of hippocampal pyramidal neurons in scrapie-infected mice. *Neuropath. Appl. Neurobiol*. **26**, 143-149 (2000).

Belzung, C. & Griebel, G. Measuring normal and pathological anxiety-like behaviour in mice: a review. *Behav. Brain Res.* **125**, 141-149 (2001).

Bergson, H. *Creative Evolution* (Dover Publications, Inc., Mineola, New York, 1998).

Bickel, S., Lipp, H.-P. & Umbricht, D. Impaired attentional modulation of auditory evoked potentials in N-methyl-D-aspartate NR1 hypomorphic mice. *Genes Brain Behav.* **6**, 558-568 (2007).

Bickel, S., Lipp, H.-P. & Umbricht, D. Early auditory sensory processing deficits in mouse mutants with reduced NMDA receptor function. *Neuropsychopharmacol.* **33**, 1680-1689 (2008).

Birchler, J.A., Auger, D.L. & Riddle, N.C. In search of the molecular basis of heterosis. *Plant Cell* **15**, 2236-2239 (2003).

Budd, T.W. & Michie, P.T. Facilitation of the N1 peak of the auditory ERP at short stimulus intervals. *Neuroreport* **5**, 2513-2516 (1994).

Bures, J., Buresova, O. & Huston, J.P. *Techniques and Basic Experiments for the Study of Brain and Behavior* (Elsevier Science Publishers B.V., Amsterdam, 1983).

Buselmaier, W., Vierling, Th., Balzereit, W. & Schwegler, H. Genetic analysis of avoidance learning by means of different psychological testing systems with inbred mice as model organisms. *Psychol. Res.* **43**, 317-333 (1981).

Castellano, C. & Ammassari-Teule, M. Prenatal exposure to morphine in mice: enhanced responsiveness to morphine and stress. *Pharmacol. Biochem. Behav*. **21**, 103-108 (1984).

Cicero, T.J., Nock, B., O'Connor, L., Adams, M.L. & Meyer, E.R. Adverse effects of paternal opiate exposure on offspring development and sensitivity to morphine-induced analgesia. *J. Pharmacol. Exp. Ther*. **273**, 386-392 (1995).

Cicero, T.J., Nock, B & Meyer E.E. Gender-related differences in the antinociceptive properties of morphine. *J. Pharmacol. Exp. Ther.* **279**, 767-773 (1996).

Cicero, T.J., Nock, B & Meyer E.E. Gender-linked differences in the expression of physical dependence in the rat. *Pharmacol. Biochem. Behav.* **72**, 691-697 (2002).

Crabbe, J.C., Wahlsten, D. & Dudek, B.C. Genetics of mouse behaviour: interactions with laboratory environment. *Science* **284**, 1670-1672 (1999).

Crawley, J.N. *What's wrong with my mouse? Behavioural phenotyping of transgenic and knockout mice* (Wiley-Liss, New York, 2000).

Greengard, P., Valtorta, F., Czernik, A.J. & Benfenati, F. Synaptic vesicle phosphoproteins and regulation of synaptic function. *Science* **259**, 780-785 (1993).

D'Adamo, P., Menegon, A., Lo Nigro, C., Grasso, M., Gulisano, M., Tamanini, F., Bienvenu, T., Gedeon, A.K., Oostra, B., Wu, S.K., Tandon, A., Valtorta, F., Balch, W.E., Chelly, J. & Toniolo, D. Mutation in GDI1 are responsible for X-linked non-specific mental retardation. *Nat. Genet*. **19**, 134-139 (1998).

D'Adamo, P., Welzl, H., Papadimitriou, S., Raffaele di Barletta, M., Tiveron, C., Tatangelo, L., Pozzi, L., Chapman, P.F., Knevett, S.G., Ramsay, M.F., Valtorta, F., Leoni, C., Menegon, A., Wolfer, D.P., Lipp, H.-P. & Toniolo, D. Deletion of the mental retardation gene Gdi1 impairs associative memory and alters social behavior in mice. *Hum. Mol. Genet.* **11**, 2567-2580 (2002).

D'Hooge, R. & De Deyn, P.P. Applications of the Morris water maze in the study of learning and memory. *Brain Res. Brain Res. Rev*. **36**, 60-90 (2001).

DeWeese, M.R., Wehr, M. & Zador, A.M. Binary spiking in auditory cortex. *J. Neurosci.* **23**, 7940-7949 (2003).

Dulawa, S.C., Grandy, D.K., Low, M.J., Paulus, M.P. & Geyer, M.A.. Dopamine D4 receptor-knock-out mice exhibit reduced exploration of novel stimuli. *J. Neurosci*. **19**, 9550-9556 (1999).

Durrant, A. The environmental induction of heritable change in Linum. *Heredity* **17**, 27-61 (1962).

Ehlers, C.L. & Somes, C. Long-latency event-related potentials in mice: effects of stimulus characteristics and strain. *Brain Res.* **957**, 117-128 (2002).

Erdman, R.A. & Maltese, W.A. Different Rab GTPases associate preferentially with a or ß GDP-dissociation inhibitors. *Biochem. Biophys. Res. Commun*. **282**, 4-9 (2001).

Eriksson, P.S., Ronnback, L. & Hansson, E. Do persistent morphine effects involve interactions with the genome? *Drug Alcohol Depend*. **24**, 39-43 (1989).

Eriksson, P.S. & Ronnback, L. Effects of prenatal morphine treatment of rats on mortality, bodyweight and analgesic response in the offspring. *Drug Alcohol Depend*. **24**, 187-194 (1989).

Erway, L.C., Willott, J.F., Archer, J.R. & Harrison, D.E. Genetics of age-related hearing loss in mice: I. Inbred and F1 hybrid strains. *Hear. Res*. **65**, 125-132 (1993).

Erway, L.C., Shiau, Y.W., Davis, R.R. & Krieg E.F. Genetics of age-related hearing loss in mice. III. Susceptibility of inbred and F1 hybrid strains to noise-induced hearing loss. *Hear. Res*. **93**, 181-187 (1996).

Fischer von Mollard, G., Sudhof, T.C. & Jahn, R. A small GTP-binding protein dissociates from synaptic vesicles during exocytosis. *Nature* **349**, 79-81 (1991).

Forrest, D., Hanebuth, E., Smeyne, R.J., Everds, N., Stewart, C.L., Wehner, J.M. & Curran, T. Recessive resistance to thyroid hormone in mice lacking thyroid hormone receptor β: evidence for tissue-specific modulation of receptor function. *EMBO J.* **15** (12), 3006-3015 (1996).

Friedler, G. in *Perinatal Pharmacology: Problems and Priorities* (eds Dancis, J. & Hwang, J.C.) 207-216 (Raven Press, New York, 1974).

Friedler, G. Paternal exposures: impact on reproductive and developmental outcome. An overview. *Pharmacol. Biochem. Behav.* **55**, 691-700 (1996).

Gagin, R., Cohen, E. & Shavit, Y. Prenatal exposure to morphine alters analgesic responses and preference for sweet solutions in adult rats. *Pharmacol. Biochem. Behav*. **55**, 629-634 (1996).

Galsworthy, M.J., Amrein, I., Kuptsov, P.A., Poletaeva, I.I., Zinn, P., Rau, A., Vyssotski, A. & Lipp, H.-P. A comparison of wild-caught wood mice and bank voles in the Intellicage: assessing exploration, daily activity patterns and place learning paradigms. *Behav. Brain Res*. **157**, 211-217 (2005).

Geppert, M., Bolshakov, V.Y., Siegelbaum, S.A., Takei, K., De Camilli, P., Hammer, R.E. & Sudhof, T.C. The role of Rab3a in neurotransmitter release. *Nature* **369**, 493-497 (1994).

Geppert, M., Goda, Y., Stevens, C.F. & Sudhof, T.C. The small GTP-binding protein Rab3A regulates a late step in synaptic vesicle fusion. *Nature* **387**, 810-814 (1997).

Graff, J. & Mansuy, I.M. Epigenetic codes in cognition and behaviour. *Behav. Brain Res*. **192**, 70-87 (2008).

Hochholdinger, F. & Hoecker, N. Towards the molecular basis of heterosis. *Trends Plant Sci*. **12**, 427-432 (2007).

Jaaskelainen, I.P., Ahveninen, J., Bonmassar, G., Dale, A.M., Ilmoniemi, R.J., Levanen, S., Lin, F.-H., May, P., Melcher, J., Stufflebeam, S., Tiitinen, H. & Belliveau, J.W. Human posterior auditory cortex gates novel sounds to consciousness. *Proc. Natl. Acad. Sci. USA* **101**, 6809-6814 (2004).

Johnston, P.A., Archer, B.T. 3rd, Robinson, K., Mignery, G.A., Jahn, R. & Sudhof, T.C. rab3A attachment to the synaptic vesicle membrane mediated by a conserved polyisoprenylated carboxy-terminal sequence. *Neuron* **7**, 101-109 (1991).

Kirby, M.L., DeRossett, S.E. & Holtzman, S.G. Enhanced analgesic response to morphine in adult rats exposed to morphine prenatally. *Pharmacol. Biochem. Behav*. **17**, 1161-1164 (1982).

Krstulovic, A.M. in *Advances in Chromatography* (ed Zlatkis, A.) 561-575 (Houston, 1981).

Laurberg, S. & Zimmer, J. Lesion-induced rerouting of hippocampal mossy fibers in developing but not in adult rats. *J. Comp. Neurol.* **190** (4), 627-650 (1980).

Lipp, H.-P., Schwegler, H. & Driscoll, P. Postnatal modification of hippocampal circuitry alters avoidance learning in adult rats. *Science* **225**, 80-82 (1984).

Lipp, H.-P., Schwegler, H., Heimrich, B. & Driscoll, P. Infrapyramidal mossy fibers and two-way avoidance learning: Developmental modification of hippocampal circuity and adult behavior of rats and mice. *J. Neurosci.* **8** (6), 1905-1921 (1988).

Lipp, H.-P., Schwegler, H., Crusio, W.E., Wolfer, D.P., Heimrich, B., Driscoll, P. & Leisinger-Trigona, M.-C. Using genetically-defined rodent strains for the identification of hippocampal traits relevant for two-way avoidance learning: a non-invasive approach. *Experientia* **45**, 845-859 (1989).

Lipp, H.-P. & Van der Loos, H. A computer-controlled Y-maze for the analysis of vibrissotactile discrimination learning in mice. *Behav. Brain Res*. **45**, 135-145 (1991).

Lippman, Z.B. & Zamir, D. Heterosis: revisiting the magic. *Trends Genet*. **23**, 60-66 (2006).

Lolle, S.J., Victor, J.L., Young J.M. & Pruitt, R.E. Genome-wide non-mendelian inheritance of extra-genomic information in *Arabidopsis*. *Nature* **434**, 505-509 (2005).

Lonart, G., Janz, R., Johnson, K.M. & Sudhof, T.C. Mechanism of action of rab3A in mossy fiber LTP. *Neuron* **21**, 1141-1150 (1998).

Loveless, N., Hari, R., Hämäläinen, M. & Tiihonen, J. Evoked responses of human auditory cortex may be enhanced by preceding stimuli. *Electroencephalog. Clin. Neurophysiol*. **74**, 217-227 (1989).

Lysenko, T.D. in *The Situation in Biological Science* (Proceedings of the Lenin Academy of Agricultural Sciences of the USSR, Session: July 31 – August 7, 1948, Verbatim Report) 11-50 (Foreign Languages Publishing House, Moscow, 1949).

Lysenko, T.D. *Agrobiology. Essays on Problems of Genetics, Plant Breeding and Seed Growing* (Foreign Languages Publishing House, Moscow, 1954).

May, P., Tiitinen, H., Ilmoniemi, R.J., Nyman, G., Taylor, J.G. & Naatanen, R. Frequency change detection in human auditory cortex. *J. Comp. Neurosci.* **6**, 99-120 (1999).

Mohajeri, M.H., Madani, R, Saini, K, Lipp, H.-P., Nitsch, R.M. & Wolfer, D.P. The impact of genetic background on neurodegeneration and behavior in seizured mice. *Genes Brain Behav*. **3**, 228-239 (2004)

Molinier, J., Ries, G., Zipfel, C. & Hohn, B. Transgenerational memory of stress in plants. *Nature* **442**, 1046-1049 (2006).

Morris, R.G.M. Developments of a water-maze procedure for studying spatial learning in the rat. *J. Neurosci. Methods* **11**, 47-60 (1984).

Nishimura, N., Nakamura, H., Takai, Y. & Sano, K. Molecular cloning and characterization of two rab GDI species from rat brain: brain-specific and ubiquitous types. *J. Biol. Chem*. **269**, 14191-14198 (1994).

Novick, P. & Zerial, M. The diversity of Rab proteins in vesicle transport. *Curr. Opin. Cell. Biol*. **9**, 496-504 (1997).

O'Callaghan, J.P. & Holtzman, S.G. Prenatal administration of morphine to the rat: tolerance to the analgesic effect of morphine in the offspring. *J. Pharmacol. Exp. Ther*. **197**, 533-544 (1976).

Phillips, T.J., Hen, R. & Crabbe, J.C. Complications associated with genetic background effects in research using knockout mice. *Psychopharmacol.* **147**, 5-7 (1999).

Pigliucci, M. & Murren, C.J. Perspective: Genetic assimilation and a possible evolutionary paradox: can macroevolution sometimes be so fast as to pass by? *Evolution* **57**, 1455-1464 (2003).

Pleskacheva, M.G., Wolfer, D.P., Kupriyanova, I.F., Nikolenko, D.L., Scheffrahn, H., Dell'Omo, G. & Lipp, H.-P. Hippocampal mossy fibers and swimming navigation learning in two vole species occupying different habitats. *Hippocampus* **10**, 17-30 (2000).

Prut, L. & Belzung, C. The open field as a paradigm to measure the effects of drugs on anxiety-like behaviors: a review. *Eur. J. Pharmacol*. **463**, 3-33 (2003).

Ralls, K. Auditory sensitivity in mice, Peromyscus and Mus musculus. *Anim. Behav.* **15**, 123-128 (1967).

Rassoulzadegan, M., Grandjean, V., Gounon, P., Vincent, S., Gillot, I. & Cuzin F. RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* **441**, 469-474 (2006).

Schalk, I., Zeng, K., Wu, S.K., Stura, E.A., Matteson, J., Huang, M., Tandon, A., Wilson, I.A. & Balch, W.E. Structure and mutational analysis of Rab GDP-dissociation inhibitor. *Nature* **381**, 42-48 (1996).

Schneider, S., Kaufmann, W., Buesen, R. & van Ravenzwaay, B. Vinclozolin – The lack of a transgenerational effect after oral maternal exposure during organogenesis. *Reprod. Toxicol*. **25**, 352-360 (2008).

Schwegler, H., Crusio, W.E., Lipp, H.-P., Brust, I. & Mueller, G.G. Early postnatal hyperthyroidism alters hippocampal circuity and improves radial-maze learning in adult mice. *J. Neurosci.* **11** (7), 2102-2106 (1991).

Seyfried, T.N., Glaser, G.H. & Yu, R.K. Thyroid hormone influence on the susceptibility of mice to audiogenic seizures. *Science* **205**, 598-600 (1979).

Seyfried, T.N., Glaser, G.H. & Yu, R.K. Genetic analysis of serum thyroxine content and audiogenic seizures in recombinant inbred and congenic strains of mice. *Exp. Neurol*. **83**, 423-428 (1984).

Siegel, S.J., Connoly, P., Liang, Y., Lenox, R.H., Cur, R.E., Bilker, W.B., Kanes, S.J. & Turetsky, B.I. Effects of strain, novelty, and NMDA blockade on auditory-evoked potentials in mice. *Neuropsychopharmacol.* **28**, 675-682 (2003).

Smith, D.J. & Joffe J.M. Increased neonatal mortality in offspring of male rats treated with methadone or morphine before mating. *Nature* **253**, 202-203 (1975).

Soyfer, V.N. New light on the Lysenko era. *Nature* **339**, 415-420 (1989).

Soyfer, V.N. The consequences of political dictatorship for Russian science. *Nat. Rev. Genet*. **2**, 723-729 (2001).

Soyka, L.F. & Joffe, J.M. Male mediated drug effects on offspring. *Prog. Clin. Biol. Res*. **36**, 49-66 (1980).

Thiel, V., Barnes, D.S. & Mrosovsky, N. A simple method for recording activity patterns of small animals. *Physiol. Behav*. **8**, 549-551 (1972).

Ulanovsky, N., Las, L. & Nelken, I. Processing of low-probability sounds by cortical neurons. *Nat. Neurosci*. **6**, 391-398 (2003).

Umbricht, D., Vyssotski, D., Latanov, A., Nitsch, R., Brambilla, R., D'Adamo, P. & Lipp, H.-P. Midlatency auditory event-related potentials in mice: comparison to midlatency auditory ERPs in humans. *Brain Res.* **1019**, 189-200 (2004).

Umbricht, D., Vyssotski, D., Latanov, A., Nitsch, R. & Lipp, H.-P. Deviance-related electrophysiological activity in mice: is there mismatch negativity in mice? *Clin. Neurophysiol.* **116**, 353-363 (2005).

Vige, A., Gallou-Kabani, C. & Junien, C. Sexual dimorphism in non-mendelian inheritance. *Pediatr. Res*. **63**, 340-347 (2008).

Vuillaume, M. & Berkaloff, A. LSD treatment of Pieris brassicae and consequences on the progeny. *Nature* **251**, 314-315 (1974).

Vysotskii, A.L., Vysotskii, D.L., Gudasheva, T.A., Ostrovskaya, R.U. & Anokhin, K.V. Modulation of long-term memory by delayed administration of the Amide of L-pyroglutamyl-D-alanine, a nootropic agent, in massed and spaced training in rats. *Neurosci. Behav. Physiol.* **29**, 137-142 (1999).

Vyssotski, D.L. *Elements of Biological Concepts. The Theory of Biological Concept Construction with Applications and Examples* (Nauka, Novosibirsk, 2004) [in Russian], ISBN 5-02-032061-7. www.evolocus.com/publications/vyssotski2004.pdf

Vyssotski, A.L., Serkov, A.N., Itskov, P.M., Dell'Omo, G., Latanov, A.V., Wolfer, D.P. & Lipp, H.-P. Miniature neurologgers for flying pigeons: multichannel EEG and action and field potentials in combination with GPS recording. *J. Neurophysiol*. **95**, 1263-1273 (2006).

Vyssotski, A.L., Dell'Omo, G., Dell'Ariccia, G., Abramchuk, A.N., Serkov, A.N., Latanov, A.V., Loizzo, A., Wolfer, D.P. & Lipp, H.-P. EEG responses to visual landmarks in flying pigeons. *Curr. Biol.* **19** (14), 1159-1166 (2009).

Waddington, C.H. Canalization of development and the inheritance of acquired characters. *Nature* **150**, 563-565 (1942).

Waddington, C.H. Genetic assimilation of an acquired character. *Evolution* **7**, 118-126 (1953a).

Waddington, C.H. The "Baldwin effect", "genetic assimilation" and "homeostasis". *Evolution* **7**, 386-387 (1953b).

Weismann, A. *On Germinal Selection as a Source of Definite Variation* (Open Court, Chicago, 1902).

West, J.R., Hodges, C.A. & Black, A.C. Jr. Prenatal exposure to ethanol alters the organization of hippocampal mossy fibers in rats. *Science* **211** (4485), 957-959 (1981).

Willott, J.F., Erway, L.C., Archer, J.R. & Harrison, D.E. Genetics of age-related hearing loss in mice. II. Strain differences and effects of caloric restriction on cochlear pathology and evoked response thresholds. *Hear. Res*. **88**, 143-155 (1995).

Willott, J.F. & Erway, L.C. Genetics of age-related hearing loss in mice. IV. Cochlear pathology and hearing loss in 25 BXD recombinant inbred mouse strains. *Hear. Res*. **119**, 27-36 (1998).

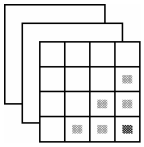
Wolfer, D.P., Müller, U., Stagliar, M. & Lipp, H.-P. Assessing the effects of the 129/Sv genetic background on swimming navigation learning in transgenic mutants: a study using mice with a modified ß-amyloid precursor protein gene. *Brain Research* **771**, 1-13 (1997).

Wolfer, D.P., Madani, R., Valenti, P. & Lipp, H.-P. Extended analysis of path data from mutant mice using the public domain software Wintrack. *Physiol. Behav.* **73**, 745-753 (2001).

Wolfer, D.P., Litvin, 0., Morf, S., Nitsch, R.M., Lipp, H.-P. & Würbel, H. Laboratory animal welfare: Cage enrichment and mouse behaviour. *Nature* **432** (7019), 821-822 (2004).

Wu, S.K., Zeng, K., Wilson, I.A. & Balch, W.E. Structural insights into the function of the Rab GDI superfamily. *Trends Biochem. Sci*. **21**, 472-476 (1996).

Yang, X., Schadt, E.E., Wang, S., Wang, H., Arnold, A.P., Ingram-Drake, L., Drake, T.A. & Lusis A.J. Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Research* **16**, 995-1004 (2006).



Evolocus

*Equipment for Biomedical Research*
*Animal Behavior & Electrophysiology*

Dmitri Vyssotski
Evolocus LLC
177 White Plains Road, 42B
Tarrytown, NY 10591, USA
Phone: +1 (914) 329-0035
Fax: +1 (914) 631-3421
E-mail: vyssotski@evolocus.com

www.evolocus.com