

Selection for contextual fear conditioning affects anxiety-like behaviors and gene expression

C. A. Ponder*, C. L. Kliethermes[†], M. R. Drew[‡],
J. Muller[§], K. Das[¶], V. B. Risbrough^{**},
J. C. Crabbe^{††}, T. Conrad Gilliam^{‡‡} and
A. A. Palmer^{§§,1}

*Department of Genetics and Development, Columbia University, New York, NY, [†]Ernest Gallo Clinic and Research Center, University of California, San Francisco, Emeryville, CA, [‡]Department of Psychiatry, Center for Neurobiology and Behavior, Columbia University, New York, NY, [§]Department of Psychiatry, Columbia University and Department of Developmental Neuroscience New York State Psychiatric Institute, New York, NY, [¶]Gertrude H. Sergievsky Center, Columbia University, New York, NY, ^{**}Department of Psychiatry, University of California San Diego, La Jolla, CA, ^{††}Department of Behavioral Neuroscience, Oregon Health & Science University, Portland Alcohol Research Center, and Portland VA Medical Center, Portland, OR, ^{‡‡}Department of Human Genetics, University of Chicago, Chicago, IL, ^{§§}Departments of Human Genetics and Psychiatry, University of Chicago, Chicago, IL

¹Corresponding author: A. A. Palmer, Department of Human Genetics, University of Chicago, 920 E. 58th St. CLSC 507D, Chicago, IL 60637, USA. E-mail: aap@uchicago.edu

Conditioned fear and anxiety-like behaviors have many similarities at the neuroanatomical and pharmacological levels, but their genetic relationship is less well defined. We used short-term selection for contextual fear conditioning (FC) to produce outbred mouse lines with robust genetic differences in FC. The high and low selected lines showed differences in fear learning that were stable across various training parameters and were not secondary to differences in sensitivity to the unconditioned stimulus (foot shock). They also showed a divergence in fear potentiated startle, indicating that differences induced by selection generalized to another measure of fear learning. However, there were no differences in performance in a Pavlovian approach conditioning task or the Morris water maze, indicating no change in general learning ability. The high fear learning line showed greater anxiety-like behavior in the open field and zero maze, confirming a genetic relationship between FC and anxiety-like behavior. Gene expression analysis of the amygdala and hippocampus identified

genes that were differentially expressed between the two lines. Quantitative trait locus (QTL) analysis identified several chromosomal regions that may underlie the behavioral response to selection; cis-acting expression QTL were identified in some of these regions, possibly identifying genes that underlie these behavioral QTL. These studies support the validity of a broad genetic construct that includes both learned fear and anxiety and provides a basis for further studies aimed at gene identification.

Keywords: Anxiety, Behavioral genetics, Contextual fear conditioning, Correlated traits, Emotionality, Gene expression, Selective breeding

Received 5 October 2006, revised 3 January 2007, accepted 4 January 2007

Fear conditioning (FC) is a classic Pavlovian learning paradigm in which an aversive unconditioned stimulus (US) is paired with a previously neutral conditioned stimulus (CS) and recall of the fearful memory is measured by observation of freezing behavior, a species-specific response to fear. Fear learning is known to be heritable in mice (Caldarone *et al.* 1997; Wehner *et al.* 1997), and humans (Grillon *et al.* 1998; Hettema *et al.* 2003) and the neuroanatomy of fear learning is well established in rodents (LeDoux 2000) and is similar in humans (LaBar *et al.* 1995; Richardson *et al.* 2004). Individuals with an anxiety diagnosis display greater acquisition and slower extinction of learned fear behaviors (Lissek *et al.* 2005), and genetic studies of fear learning frequently cite the elucidation of genes involved in human anxiety disorders as a goal (Moldin 2000).

Common tests of anxiety-like behavior in rodents are highly sensitive to environmental factors (Francis *et al.* 1999; Wahlsten *et al.* 2006) and have no obvious parallels in humans. In contrast, fear learning can be readily examined in human subjects. While learned fear and anxiety have known neuroanatomical (Davis 1992) and pharmacological (Davis 1979a, 1979b; Risbrough *et al.* 2003; Santos *et al.* 2005) parallels, they have yet to be definitively genetically linked. In addition, it has not been established that genetic differences in fear learning are due to differences in the formation of learned fearful associations, vs. trivial differences related to other aspects of the paradigm (e.g. shock sensitivity). Thus it has not been clear whether genetic studies of fear learning will enhance our understanding of the genetic basis of anxiety disorders.

We used divergent short-term selective breeding to produce mouse lines with extremely high or low levels of FC.

Note: The first author was an Outstanding Young Investigator Awardee, 8th Annual Meeting of the International Behavioural and Neural Genetics Society, Vancouver, BC, Canada, May 19–22, 2006.

Selection acts on a genetically heterogeneous population by altering the frequency of alleles that influence the trait under selection. This allows for the examination of the effect of multiple trait-relevant alleles on putatively correlated phenotypes (Palmer & Phillips 2002). Short-term selection produces marked phenotypic differences in a few generations with minimal inbreeding (Belknap *et al.* 1997). In long term selection studies genetic drift may change the frequencies of nonselected alleles, confounding the interpretation of other traits that are correlated with the selected phenotype.

After the creation of the selected lines, we further explored the differences in FC by testing mice from the final selected generations with different numbers and intensities of shocks. We examined fear potentiated startle (FPS) to gauge fear learning in a freezing-independent manner. Acute sensitivity to foot shock was also evaluated. We examined the general learning ability of the mice to determine if the difference in learned fear resulted from a general learning difference between the lines. We then used the selected lines to examine the genetic relationship between fear learning and anxiety-like behavior. Finally, we used gene expression microarrays to identify genes that are differentially expressed between the hippocampus and the amygdala of the high and low selected lines; such genes might underlie the behavioral differences between these two lines.

Materials and methods

Environment and housing

All experiments were performed in accordance with the National Institutes of Health guidelines for the care and use of laboratory animals and were approved by the Columbia University and Portland VA Medical Center’s Institutional Animal Care and Use Committees. Mouse colony rooms were maintained on a 12/12 light/dark cycle with lights on at 07:00 hours. Two to five same-sex littermates were housed in clear plastic cages with standard corn-cob type bedding. All mice were maintained with food and water *ad libitum*, except during testing. Certain procedures were followed in all behavioral experiments: testing was conducted during the light phase between 09:00 and 17:00 and mice were brought into the testing room in their home cages and allowed to adapt for a minimum of 30 min before testing.

Fear conditioning

Fear conditioning (FC) procedures are based on Paylor *et al.* (1994). Fear conditioning (FC) chambers obtained from Med Associates (St. Albans, VT, USA) had inside dimensions of 29 cm × 19 cm × 25 cm with metal walls on each side, clear plastic front and back walls and ceilings, and stainless steel bars on the floor. A fluorescent light provided dim

illumination (~10 lux) and a fan provided a low level of masking background noise. Chambers were cleaned with 80% ethanol between animals. Behavior was recorded with digital video and analyzed with FreezeFrame software from Actimetrics (Evanston, IL, USA). Mice were between 7 and 14 weeks old on the first day of testing.

Testing for FC was conducted over three days (Fig. 1a). On test day 1 baseline activity was measured, beginning 30 seconds after the mice were placed into the test chambers, and terminated at 180

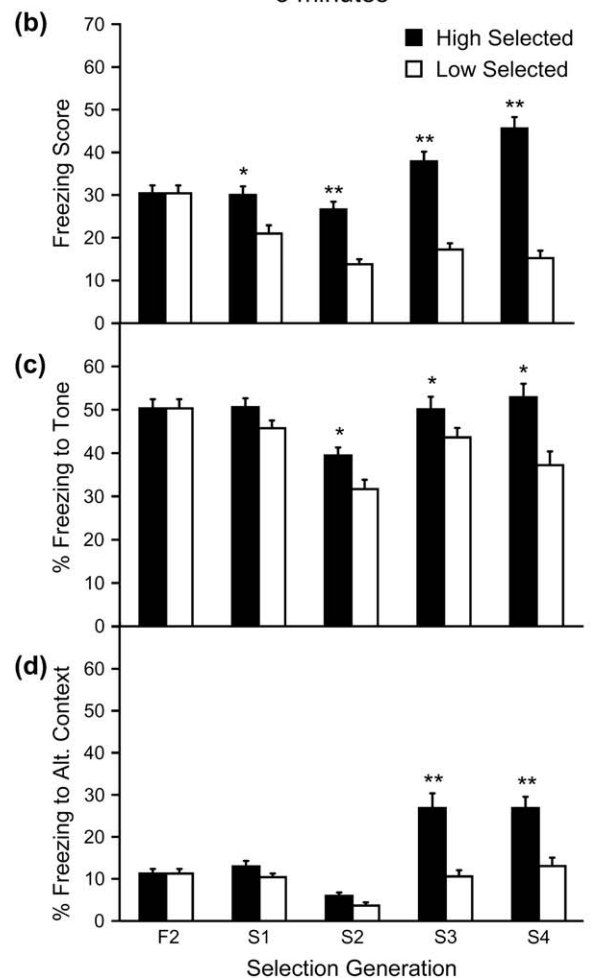
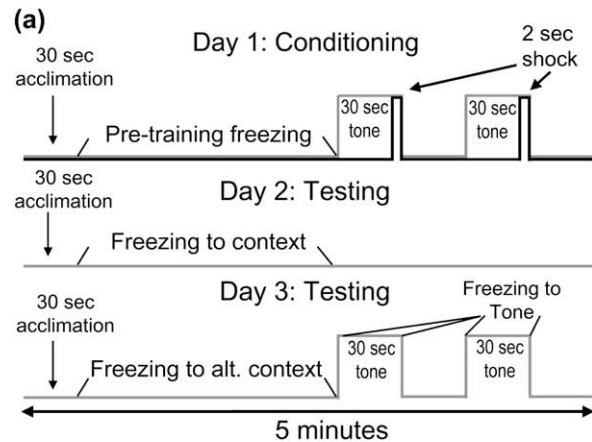


Figure 1: Response to bidirectional selection for freezing to a context previously paired with shock. (a) Diagram shows the FC paradigm used for selection. (b) ‘Freezing score’ of mice from each selection generation. (c) The response of the same mice to the tone presented on day 3. (d) The response of the same mice to the altered context presented on day 3 prior to presentation of the tone. *n* = 144 mice in the F₂, 96 mice per selected line in the S₁ through S₃, and 48 mice in the S₄. The F₂ generation was common to both selected lines, so bars are identical. **P* < 0.05, ***P* < 0.001 for comparisons between the high and low lines. All bars represent the mean ± standard error.

seconds. Mice were then exposed twice to the CS, which consisted of an 85 dB, 3 KHz tone that persisted for 30 seconds and coterminated with the US, which was a 2 second, 0.5 mA foot shock, with a 30-second intertrial interval (ITI) between pairings.

On day 2, the testing environment was identical to the first day; however neither tones nor shocks were presented. Freezing in response to the test chamber (context) was measured beginning 30 seconds after the start of the test and ending at 180 seconds. Selective breeding was based on a 'freezing score' which was the percentage of time spent freezing on day 2 minus the percentage of time spent freezing on day 1.

On day 3, the context was altered in several ways: a different experimenter wore a different style of gloves, the transfer cages had no bedding, the metal shock grid was covered with a white plastic floor, a bent white plastic wall was inserted into the test chamber, a yellow light filter was placed over the chamber lights, chambers were cleaned with 0.1% acetic acid solution, and the vent fan was partially obstructed to change the background noise. The CS was again presented twice, but without a foot shock. The 'freezing to tone' score is an average of the percentage of time spent freezing during the two 30-second CS presentations. Freezing to the altered context was defined as freezing that occurred between 30 and 180 seconds on day 3. Data from day 3 were *not* used to determine which animals would be selected as breeders.

Selective breeding

To create the selected lines, an F₁ cross between C57BL/6J and DBA/2J (B6D2F₁) mice was obtained from The Jackson Laboratory (Bar Harbor, ME, USA). The B6D2F₁ mice were intercrossed at Columbia University to obtain an F₂ population. 144 F₂ mice (72 of each sex) were phenotyped for FC. From this population the 12 male and 12 female mice with the highest contextual freezing score were selected to breed the high selected line (avoiding brother–sister breeder pairs). Three alternate breeder pairs were created from the mice with the next highest freezing scores, their offspring were used for the subsequent line only when primary breeder pairs failed to produce offspring. From the high freezing breeder pairs, 96 offspring from 12 families (as close to four male and four female animals per family as possible) constituted the first selected generation (denoted S₁). These mice were tested and bred in the same manner, until the S₄ litter had been produced. The procedure to produce the low selected line was identical, except that mice with the lowest freezing scores were bred in each generation. The line names reflect the number of generations removed from the F₂ population, and their freezing behavior (L for low, H for high), so that the offspring of the fourth selected breeders for the high freezing line were called S₄H. Forty-eight mice per line from the S₄ generation were phenotyped for FC. From the S₄ mice, the six male and six female mice with the highest or lowest freezing scores were used to produce a small S₅ generation. These S₅ mice were used for experiments that could not be completed with available S₄ mice. Phenotypes were measured in behaviorally naïve mice, with a few exceptions, as noted below. The lines were terminated after the S₅ generation because beyond this point inbreeding would confound interpretation of genetic correlations. The inbreeding coefficient for the fourth selected generation is 0.075 (Falconer & Mackay 1996). The decision to halt the selection after four generations (other than the creation of a small S₅ population) was based on the calculations of Belknap *et al.* (1997).

Effect of number of CS/US pairings on freezing behavior

The effect that the number of CS–US pairings had on the freezing exhibited on days 2 and 3 was tested by using different numbers of pairings on the training day in behaviorally naïve male and female mice from the S₄ generation (*n* = 40 per line) that were between 7 and 10 weeks old. All sessions were identical to those described for the selection procedure except that the length of the session was shortened or extended as needed to accommodate the different number of CS/US pairings. The 0 CS–US group was presented with two 30-second tones but received no shocks.

Effects of shock magnitude on learning

To test the effect of the shock magnitude (in milliamps) on the freezing behavior exhibited on days 2 and 3, we varied the intensity of the shocks presented on day 1 in behaviorally naïve male and female mice from the S₄ generation (*n* = 48 per line) that were between 7 and 11 weeks old. All sessions were identical to those described for the selection procedure except that the shock intensity on day 1 was 0, 0.1, 0.3, 0.5, 0.7 or 0.9 mA (*n* = 8 per line per intensity).

Extinction of contextual freezing

Extinction of contextual freezing behavior was tested in behaviorally naïve male and female mice from the S₅ generation (*n* = 24 per line) when they were between 12 and 14 weeks old. Testing on days 1 and 2 was similar to previous experiments (Fig. 1a) except that a 'no-shock' control group was also used. The 'no-shock' control group received a training session that was identical to day 1 except it did not include any shocks. The testing sessions used for days 3–9 were identical to day 2. Corrected scores for each day were obtained by subtracting the baseline freezing on day one from time spent freezing on each subsequent testing day.

Fear potentiated startle

Fear potentiated startle (FPS) was tested in behaviorally naïve male and female mice (*n* = 40 from S₄L; *n* = 36 from S₄H) when they were between 7 and 12 weeks old. Twelve mice from each line were randomly assigned to each of 3 treatment groups with the exception that 16 mice from the low selected line were assigned to the UC group. The protocol and apparatus are similar to those that have been described previously (Risbrough *et al.* 2003). On day 1 mice were randomly presented with one of three training sessions. The paired training session (P) consisted of 20 light + shock pairings (30-second chamber light coterminating with 0.5 seconds 0.14 mA shock). The uncorrelated protocol (UC) consisted of 10 light + shock pairings, 10 shock alone trials (0.5-second 0.14 mA shock) and 10 light alone trials (30-second chamber light, no shock). The no shock training session (NS) consisted of 20 light alone trials – no shocks were presented to mice in this treatment group. On day 2 all mice were presented with the same set of stimuli, which consisted of 12 light plus startle trials and 12 trials that consisted of that startle stimuli alone, in the absence of the light cue (Risbrough *et al.* 2003). The maximum response for each trial in the 65 milliseconds following the onset of the startle stimulus was used for all calculations. The dependent measure was %FPS, which was calculated as follows: %FPS = [(startle response on cue trials) – (startle response on no-cue trials)] / (startle response on cue trials) × 100%.

Shock sensitivity

The sensitivity to foot shock was assessed using the same apparatus used to measure FPS. Behaviorally naïve male and female mice from the S₄ generation (*n* = 20 per line) that were between 7 and 9 weeks old were used for this study. After a 5-min adaptation period, five startle trials were presented, separated by an average of 15 seconds, with a pseudorandom ITI. Each startle trial consisted of a 40-millisecond burst of 100-dB white noise. Next, a total of 25 shocks were presented in 5 blocks that consisted of one shock of each of the five intensities (0.1, 0.3, 0.5, 0.7 and 0.9 mA; all shocks were 500 milliseconds in duration) in a pseudorandom order. These trials were separated by a pseudorandom ITI that averaged 90 seconds. After the last shock, a final block of five startle trials were presented, after which the session ended. The startle response was measured in the same manner as in FPS except that the response window was 500 milliseconds.

Pavlovian approach conditioning

Pavlovian Approach Conditioning was tested in behaviorally naïve female mice from the S₅ generation (*n* = 10 per line) when they were between 13 and 19 weeks old. Mice were housed singly and placed

on a restricted diet of Bio-Serv (Frenchtown, NY) complete purified diet pellets. Mice were provided a titrated ration of food once per day, to maintain mice at between 85% and 90% of their baseline weight (ration size averaged 1.5 g). *Ad libitum* water was available throughout the experiment.

The subjects were tested in modular conditioning chambers (Coulbourn Instruments, Allentown, PA; model H10-11R-TC) with internal dimensions of 25.5 (deep) × 28.5 (high) × 28.5 cm (wide) located inside a sound-attenuating isolation cubicle (Coulbourn Instruments, Model H10-24) with a white noise generator (60 dB) in the testing room. A feeder trough (dimensions 2.5 × 2.5 cm) was centered in the right wall panel 1.5 cm above the floor. Reward pellets (20 mg, Research Diets, New Brunswick, NJ, USA, Product # PJA-0020) could be dropped into the feeder by an automatic pellet dispenser located outside each chamber. Directly inside the feeder trough was a yellow feeder light, the illumination of which served as the CS. Infrared photocell detectors inside the feeder trough were used to detect mouse head pokes (the conditioned response).

Prior to the start of conditioning, mice received two 15-min habituation sessions, conducted on consecutive days. Two days after the final habituation training session conditioning sessions were conducted once daily, 5 days a week, for a total of 10 sessions. Each of 15 trials consisted of the illumination of the feeder light (the CS) followed by the delivery of a reward pellet 8 seconds later, with a 72-second ITI between trials. Head entries into the feeder were recorded throughout each session. To assess conditioned approach to the feeder light CS, we compared the number of head entries during the first 8 seconds of each CS presentation (prior to pellet delivery) to the number of head entries during the final 8 seconds of each ITI. These values were compared using elevation scores, which were computed by subtracting ITI head entries from CS head entries.

Morris water maze

Two weeks after the completion of the Pavlovian approach conditioning test, the same mice were used to examine behavior in the Morris water maze. Thus, at the beginning of testing mice were 18–24 weeks old. The maze consisted of a white plastic tank (1.7-m internal diameter, 56 cm deep) filled to a depth of 51 cm, and water temperature was maintained at 21°C at all times. A circular, clear Plexiglas platform (14.6 cm diameter) was submerged 0.5 below the surface of the water in the middle of one quadrant of the maze. Trials were administered at the rate of 3 trials per day with an ITI of approximately 30-min. On visible trials, the platform location was identified by a cylindrical, multicolored plastic tube (12 cm high × 1 cm diameter) fixed atop the submerged platform. The maximum trial duration was 2 min. If a mouse did not find the platform by that time, it was either led there or placed there by the experimenter.

One day of pretraining preceded behavioral training and behavior was not recorded. Following pretraining, there were 3 days of visible platform training, then 5 days of hidden platform training (Meshi *et al.* 2006). Data from visible and hidden training were collected using the Chromotrack system (San Diego Instruments). For both visible and hidden training, the water was made opaque with white, nontoxic tempera paint so that the computerized tracking system could easily identify the location of the mice (Pearl Art # 2910). Data were then analyzed offline using in-house software to calculate the following measures: total distance, latency to platform, swim speed, percent time in each quadrant, and (on probe trials) crossing of the platform location.

Anxiety-like behavioral measures

The elevated zero maze, open field, and light/dark box all examined anxiety-like behaviors and were all performed at the Portland VA Medical Center in Portland, Oregon. Behaviorally naive female S4 mice ($n = 36$ per line) were shipped from New York to Portland and acclimated for 3 weeks prior to behavioral testing. Mice were between 14 and 18 weeks old when tested. The Portland VA Medical Center uses similar mouse housing systems to those used at Columbia University (clear plastic shoebox cages with corn-cob bedding).

To increase statistical power we tested additional mice naive to a particular test, but previously tested in one of the three tests. The second test in the new apparatus was conducted a minimum of 6 days after the first test. The mice that were first tested in the elevated zero maze were retested in the open field; those that were tested in the light/dark box were retested in the elevated zero maze; and the group of mice tested first in the open field were retested in the light/dark box. This resulted in a final sample size of 24 female mice/line/test (12 experimentally naive and 12 previously tested within each line).

Elevated zero maze

The elevated zero maze used 48 S₄ female mice ($n = 24$ per line) that were between 14 and 16 weeks old. It consists of a black acrylic ring (45 cm external diameter) elevated to a height of 46 cm above the floor and 5.5 cm from the inner to the outer edge. The ring was divided into quadrants comprised of two closed and two open sections, the closed sections had walls that were 11 cm tall and made of clear acrylic, while the open sections were surrounded by a short (3 mm) lip designed to prevent mice from falling. The lighting during the test was 15 lux. The procedure for testing mice on the elevated zero maze has been described previously (Kliethermes *et al.* 2004). The variables recorded during the 5-min test include the time spent on the open sections, the number of open section entrances, the latency to the first open section entrance, and the total number of line crossings (defined as open plus closed section entrances, plus crossings over either of two additional lines spaced evenly within each arm).

Open field

The open field used 48 S₄ female mice ($n = 24$ per line) that were between 14 and 18 weeks old. The procedure for the open field has been previously described (Crabbe *et al.* 1982, 1987). The open field consisted of a large, circular arena (60.96 cm diameter) enclosed within high walls (42.55-cm height) under bright (350 lux) illumination. Results from the open field were analyzed by a repeated measures analysis of variance (ANOVA) with line as the between subjects factor and time bin as the within subjects factor. The mouse was placed into the center of the arena for a 15-min test and the data were recorded from videotape in 5-min bins. The variables recorded included the time spent in the center (defined as any time the mouse was not within 8 cm of the wall), the total number of line crossings, and the number of fecal boli.

Light/dark box

The light/dark box used 48 S₄ female mice ($n = 24$ per line) that were between 14 and 16 weeks old. The apparatus used was similar to that proposed by Crawley & Goodwin (1980), and consisted of a light compartment (28 × 28 cm) made from clear acrylic and an enclosed, dark compartment (28 × 19 cm) made from black acrylic. A small door allowed access to both compartments, and the light intensity in the light compartment during the test was approximately 250 lux. Procedures for its use were recently described in Kliethermes *et al.* (2004). The variables scored during the 10 min test included the number of transitions between the compartments, the time spent in the light compartment, the initial latency to exit the light compartment, the latency to re-enter the light compartment, and the number of line crossings in the light compartment.

Gene expression measurement by Affymetrix microarray

We measured gene expression in 72 experimentally naive male mice from the S₄ generation that were between 7 and 12 weeks old. We focused our study on the amygdala and the hippocampus, both of which are known to be essential to FC (LeDoux 2000). Although the

mPFC is known to be important for extinction of FC; we did not examine gene expression of this region. We collected a total of 72 samples, 36 per selected line, with 18 representing the hippocampus and the other 18 representing the amygdala. Different mice were used to obtain amygdala and hippocampal sections because the dissection procedures precluded obtaining both samples from the same mouse. Mice were sacrificed by decapitation and the amygdala or hippocampus was rapidly dissected out and frozen in liquid nitrogen. For the amygdala dissection, the brain was placed in a brain block and two razor blades were used to cut a 2-mm coronal section. This section was then placed under a dissecting scope and the amygdala dissected bilaterally by making three cuts as described in (Shumyatsky *et al.* 2002). For the hippocampus, the cerebellum was removed, then the brain bisected along the sagittal plane. The hippocampus was then removed by blunt dissection using a spatula underneath the ventral hippocampus as described in (Verbitsky *et al.* 2004) and total RNA was isolated using the Qiagen RNeasy Mini Kit.

For each microarray, total RNA from three male siblings was pooled and hybridized to a single Affymetrix MOE 430 v2 array, thus six arrays were used per brain region per line. The concentration of the purified total RNA from each individual was quantified using a spectrophotometer and combined such that each individual's sample was represented equally in each pool. Total RNA samples were sent to the NINDS/NIMH Microarray Consortium (<http://arrayconsortium.tgen.org>) where they were processed, labeled and hybridized to the arrays using standard Affymetrix protocols.

Array data were normalized in Bioconductor 1.8 using the GCRMA algorithm (Gentleman *et al.* 2004; Irizarry *et al.* 2003). Normalized expression levels were examined using a two-way ANOVA for the factors brain region and line. We then used the program QValue (<http://faculty.washington.edu/~jstorey/qvalue>) to calculate q -values (Storey & Tibshirani 2003) and used a $q = 0.05$ threshold for significance. Present, marginal or absent (P/M/A) calls were then calculated in R using the 'mas5calls'. The fold change for each significant transcript was calculated for each brain region to identify probe sets with large expression differences. Map locations of candidate probe sets were verified by using the BLAT alignment tool (<http://genome.ucsc.edu/>) to compare the probe sequences to the mouse genome (Build 36).

Quantitative PCR

Quantitative PCR (qPCR) using Sybr Green was used to obtain relative quantification of transcript abundance as described previously (Palmer *et al.* 2005). Primers were designed to amplify multiple exons across at least one intron to avoid amplifying genomic DNA that might have contaminated the cDNA sample. Expression was examined in 46 unpooled hippocampal samples. These samples included all of the S_4 subjects from the microarray study (18 per line) plus an additional 10 subjects that were not included in the microarray studies ($n = 20$ for S_4H and 26 for S_4L). A t -test for gene expression between the high and low lines was used to confirm differential expression.

Genotyping of DNA from qPCR samples

DNA was isolated from spleens of the 46 animals used in qPCR experiments ($n = 20$ for S_4H and 26 for S_4L) using a standard salting out protocol. Single nucleotide polymorphisms (SNPs) that distinguish between the B6 and D2 strains were selected from Petkov *et al.* (2004) based on proximity to significantly differentially expressed genes as determined by the microarray studies. These SNPs were genotyped by kbiosciences (Hoddesdon, UK). Each qPCR result was compared with the genotype of a nearby SNP (B6, Het or D2) in order to test the hypothesis that the expression difference was due to a heritable factor that is in close apposition to the location of the gene itself; such differences are termed *cis*-acting expression quantitative trait locus (*cis*-eQTL). We used a one-way ANOVA to examine the effect of SNP genotype (B6, Het, D2) on expression of the nearby target gene. In all cases the ANOVA reflected a significant effect of genotype on gene expression. We next sought to determine whether the inheritance pattern of the eQTL was dominant or additive. A *cis*-eQTL

was designated as additive if expression in the heterozygote group was significantly different from each homozygote; *cis*-eQTLs were said to be dominant if the heterozygote was statistically equivalent to one homozygote and different from the other.

Genotyping and QTL analysis of markers from S_3 generation

DNA was collected from the subset of S_3 mice that were selected to breed the S_4 generation (alternate breeders were also included; total $n = 30$ for S_3L and $n = 29$ for S_3H). We examined 305 informative SNP markers in these animals spanning all chromosomes. Single nucleotide polymorphisms (SNPs) were genotyped by the Harvard Partners Center for Genetics and Genomics. SNPs had an average distance of about 5 cM. We searched for quantitative trait loci by determining the likelihood that the difference in frequency of each SNP in the S_3L and S_3H mice was due to genetic drift using the method of Belknap (1997).

Results

Creation of high and low FC selected lines

Bidirectional selection for freezing to context resulted in a highly significant divergence in freezing scores between the high and low lines (Fig. 1b). A three-way ANOVA with line, generation and sex as factors showed no significant effect of or interactions with sex. A two-way ANOVA of line and generation identified a significant interaction between the two ($F_{[3,663]} = 10.0$; $P < 0.001$). The difference between the lines was significant in the first generation ($F_{[1,198]} = 10.6$; $P < 0.0013$) and behavior progressively diverged in each subsequent generation. To determine the significance of the response to selection, separate one-way ANOVAs were performed within each line, including data from the F_2 population. A significant effect of generation (F_{2-S_4}) was found in the high line ($F_{[4,474]} = 8.89$; $P < 0.001$) and the low line ($F_{[4,475]} = 14.11$; $P < 0.001$).

Selection for freezing to context also altered freezing in response to the CS (tone) that had been paired with the shock (Fig. 1c). The freezing to tone was analyzed with a three-way ANOVA for sex, generation (S_1 to S_4) and line. Females had significantly higher freezing than males ($F_{[1,651]} = 7.85$; $P < 0.0052$), but there was no interaction between sex and either line or generation. A two-way ANOVA for generation and line showed significantly more freezing in the high line, ($F_{[1,659]} = 43.1$; $P < 0.001$), but there was no significant interaction between line and generation. A one-way ANOVA of generation (F_{2-S_4}) in each line was significant in both the high line ($F_{[4,465]} = 6.53$; $P < 0.001$), and the low line ($F_{[4,475]} = 15.21$; $P < 0.001$).

Selection was also associated with a difference in the time spent freezing in the altered context (Fig. 1d). No main effects of sex or interaction with generation or line were found for altered context, so further analyses are collapsed on sex. A two-way ANOVA for the factors generation and line on freezing in the altered context before the tone revealed a significant interaction between the two ($F_{[3,659]} = 12.2$; $P < 0.001$). Simple main effects analysis showed that the S_3H and S_4H spent significantly more time freezing in the altered context than the S_3L and S_4L , respectively. The

increase in freezing before the tone in the altered context may represent an increase in the nonassociative learning component in the high line (Kamprath & Wotjak 2004).

Effect of number of CS/US pairings on freezing behavior

Mice were trained with between 0 and 8 CS/US pairings to investigate the possibility that the low line would freeze as much as the high line if presented with additional training trials (Fig. 2a). A three-way ANOVA for line, number of shocks and sex showed no main effect of or interaction with sex. A two-way ANOVA for line and number of shocks showed a significant interaction between the two ($F_{(4,70)} = 9.07$; $P < 0.001$). The high line froze more than the low line for each treatment except the 0 shock condition, for which the two lines showed comparable (low) freezing scores (Fig. 2a). Even eight shocks did not produce freezing scores in the low line that were comparable to the high line after a single shock.

Effects of shock intensity on learning

Mice were trained with between 0 and 0.9 mA shock intensity to investigate the possibility that the low line would freeze as much as the high line given a more intense shock. The high line showed much greater freezing than the low line at all shock intensities above 0.1 (Fig. 2b). An initial three-way ANOVA found a main effect of sex ($F_{(1,72)} = 8.06$; $P < 0.006$), indicating that females froze more than males, but no interaction involving either line or shock, therefore, subsequent analyses did not include sex as a factor. A two-way ANOVA for the factors line and shock intensity showed a significant interaction between the two ($F_{(15,84)} = 7.14$; $P < 0.001$). *Post hoc* comparisons showed the two lines were not significantly different in the 0 or 0.1 mA shock intensity, but were significantly different at all higher shock intensities.

Extinction of contextual freezing

We investigated the extinction of freezing behavior over successive presentations of the context in the absence of shock. The high selected line showed higher freezing after training and took more days to show extinction of freezing behavior, as compared to the low line (Fig. 2c). An initial four-way ANOVA identified a significant interaction between line, treatment and day ($F_{(7,273)} = 3.1$; $P < 0.005$), as well as an interaction between line, sex and treatment ($F_{(1,39)} = 6.1$; $P < 0.02$). To investigate the source of the line by treatment by day interaction, we examined the effect of treatment and day on the two lines separately. For the low freezing line the two treatment groups were statistically different on day 2 of testing, but were similar on all subsequent days. In the high freezing line, the two groups were statistically different on days 2 through 8 of testing. While the number of testing sessions required for extinction differed between the two lines, it was unclear whether this was due to a difference in the rate of extinction or a difference in the magnitude of the initial response.

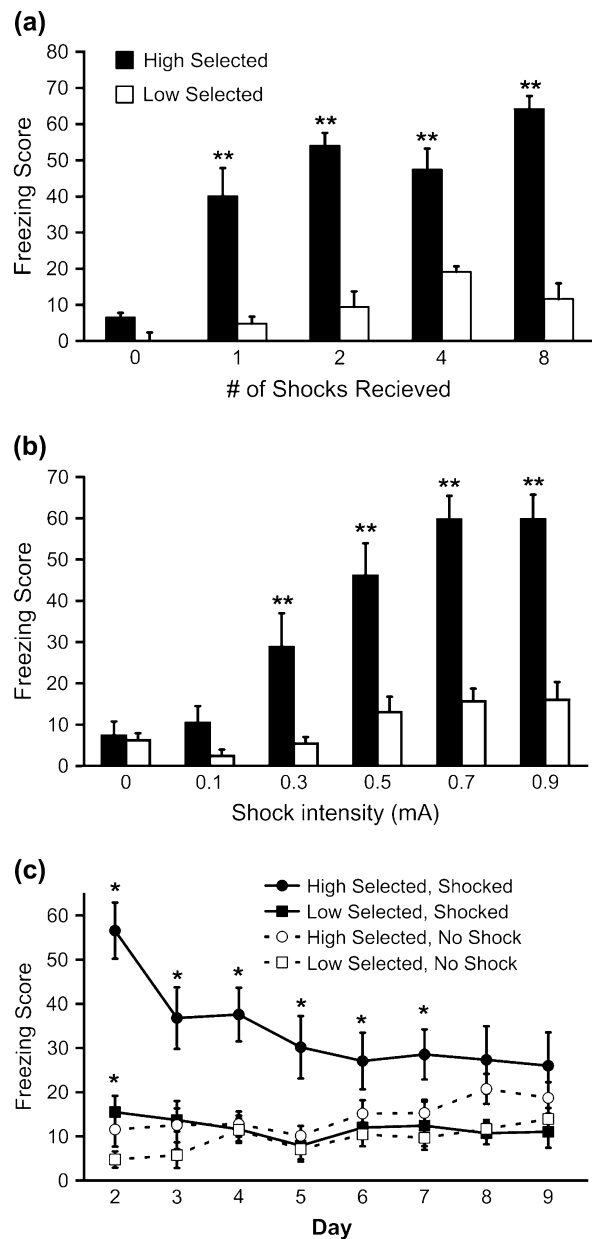


Figure 2: Effects of experimental parameters on freezing score in S4 high and low selected lines. (a) Freezing score for mice trained with different numbers of CS/UC pairings on day 1. (b) Freezing score for mice trained with different shock intensities on day 1. (c) Extinction of freezing to context over the course of 9 days of testing. Each point represents the mean \pm standard error of 12 mice, except for the points corresponding to the high selected, shocked group, which consisted of 11 mice. * $P < 0.05$, ** $P < 0.001$ for comparisons between the high and low lines in panels A and B, or for comparisons between the shocked and un-shocked groups within each selected line for the indicated day in panel C.

Fear potentiated startle

Fear potentiated startle (FPS) measures FC by assessing the increase in the startle response in the presence of a CS after it has been paired with the US, relative to the startle response in the absence of that CS. Fear potentiated startle (FPS) was exhibited by the high line in the paired treatment group (P), but was not observed in the P groups from the low line or in the NS or the UC negative control groups from either line (Fig. 3). A two-way ANOVA for the factors training (P, UC, NS) and line showed a significant interaction between the two ($F_{12,701} = 3.1$; $P < 0.05$). We used simple main effects analysis to identify the source of this interaction. For the high selected line, the P group was significantly different from the UC and NS groups ($P < 0.01$), however there were no differences between the treatment groups for the low selected line. The %FPS of the P group was significantly greater in the high selected line as compared to the low selected line ($P < 0.01$). These results demonstrate a difference in fear learning between the high and the low lines that does not depend on freezing behavior and is therefore not confounded by differences in locomotor activity.

Shock sensitivity

Selection for high or low fear learning could have changed the frequency of alleles that influence sensitivity to shock. We measured the acute startle response at each of five shock intensities in both high and low selected mice (Fig. 4). There was no evidence of a change in response to a shock of a given intensity over the course of the session (no interaction or main effect of shock number when it was treated as a repeated measure). Therefore, averages of the responses to the 5 shocks at each intensity were used for analysis. A

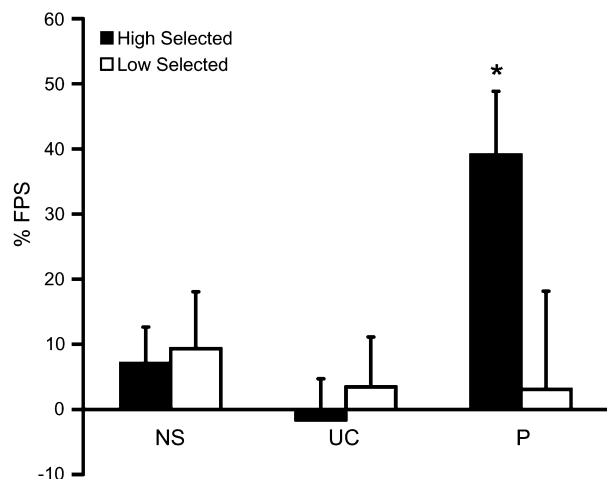


Figure 3: Fear potentiated startle (FPS) in S_4 high and low selected lines. Each bar represents the behavior of 12 mice with the exception of the low selected UC group, which represents 16 mice. All values are means \pm standard error. * $P < 0.01$ for the comparison of the high selected P group to the other two high selected groups and also for the comparison of the high selected P group to the low selected P group.

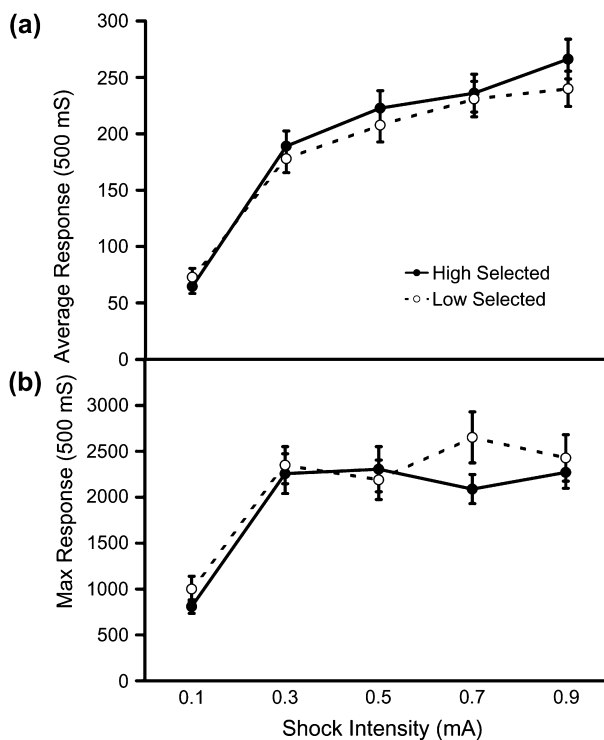


Figure 4: Acute shock sensitivity in S_4 high and low selected lines. (a) Maximum startle response over a 500-millisecond period immediately after the onset of the shock. (b) Average response over a 500-millisecond period. Each point represents the average response of 20 mice to five shocks at each of the indicated intensities. All values are means \pm standard error.

three-way ANOVA identified significant main effects of sex for both the maximum response to the shock ($F_{1,361} = 5.0$; $P < 0.05$) and average response to the shock ($F_{1,361} = 9.4$; $P < 0.005$) but no significant interactions involving sex or line. A two-way ANOVA for line and shock intensity found no main effect of line, and no interaction between line and shock intensity for either the max or the average measures in either sex. There was a clear effect of shock intensity for both the max ($F_{4,1521} = 56.3$; $P < 0.001$) and average measures ($F_{4,1521} = 248$; $P < 0.001$), indicating greater response in both lines to increasing shock intensities. There was also no effect of line on the response to the 0.5 mA shock intensity, which we tested alone *a priori*, since that was the intensity used in selection (Max: $F_{1,381} = 0.12$; Ave: $F_{1,381} = 0.49$; both $P > 0.5$). Therefore, we concluded that selection had not altered acute sensitivity to foot shock.

Pavlovian approach conditioning

In order to assess differential associative conditioning between the selected lines, we tested mice in a Pavlovian approach conditioning paradigm in which mice learned the association between a feeder light (CS) and food reward (US). We measured the conditioned response (CR), which was head-poking toward the feeder during the CS. All mice

acquired the anticipatory head-poking during the CS ($F_{[1,9]} = 8.4$; $P < 0.001$ for overall trial effect) and there was no difference in the acquisition curves ($F_{[1,9]} = 0.67$; $P > 0.5$) as measured by the interaction of line and day (Fig. 5a).

We characterized learning speed in the two groups with a method described by Gallistel *et al.* (2004). This procedure recursively examines the data for each subject and detects the points at which responding in the CS exceeds each subject's baseline level of responding to estimate the number of pairings between the CS and the US necessary for consistent responding. Although there was considerable individual variability in the number of pairings to acquisition (range from 29 to 255 pairings), there was no significant difference in acquisition rate ($F_{[1,9]} = 0.13$; $P > 0.5$) with the average number of pairings to acquisition being 141 ± 29 for the high line and 134 ± 27 for the low line (Fig. 5b).

Morris water maze

The Morris water maze was used to evaluate the selected lines for differences in general learning ability. The visible platform and hidden platform data were analyzed separately and the three trials of each training day were averaged, except for day 5 when the average was taken across the two training trials only (because the third trial was a probe trial). Separate two-way ANOVAs for visible and hidden data, using line (high and low) and day (1, 2, 3 for visible; 1, 2, 3, 4, 5 for hidden) as factors showed total distance per trial decreased as a function of day (Fig. 5c). In visible training there was a significant effect of day, ($F_{[2,36]} = 49.07$, $P < 0.001$), but not of line ($F_{[1,18]} = 0.77$; $P > 0.5$) or of the interaction ($F_{[2,36]} = 1.74$, $P > 0.19$). Hidden training exhibited a similar pattern, with a significant effect of day ($F_{[4,72]} = 15.75$, $P < 0.001$), but not of line ($F_{[1,18]} = 1.25$, $P > 0.25$) or the interaction ($F_{[4,72]} = 1.21$, $P > 0.25$).

Swim speed increased slightly over days in visible training (data not shown). There was a significant effect of day ($F_{[2,36]} = 5.32$, $P < 0.009$) but not of line ($F_{[1,18]} = 0.41$) or their interaction ($F_{[2,36]} = 1.49$, $P > 0.2$). In hidden platform training, there were no significant effects of day ($F_{[2,36]} = 2.44$, $P = 0.054$), line ($F_{[1,18]} = 0.87$; $P > 0.5$) or their interaction ($F_{[2,36]} = 0.13$; $P > 0.5$).

The two dependent measures for the probe trial data, proportion of time in target quadrant (Fig. 5d) and crossings of platform location (Fig. 5e) were analyzed together using multivariate ANOVA. Neither measure differed between lines ($F_{[2,17]} = 0.61$; $P > 0.5$). To confirm that each mouse line was performing above chance level, one-sample *t*-tests were used to assess whether the proportion of time in the target quadrant exceeded 0.25. Both lines exceeded the chance level of performance, ($t_{[9]} > 7.0$; $P < 0.001$).

Elevated zero maze

In order to assess anxiety-like behavior in the selected lines we examined behavior on the elevated zero maze. A two-sample *t*-test to compare the two lines indicated that the low line spent more time in the open sections of the elevated zero maze (Fig. 6c; $t_{[46]} = 2.44$; $P < 0.05$), and made more open

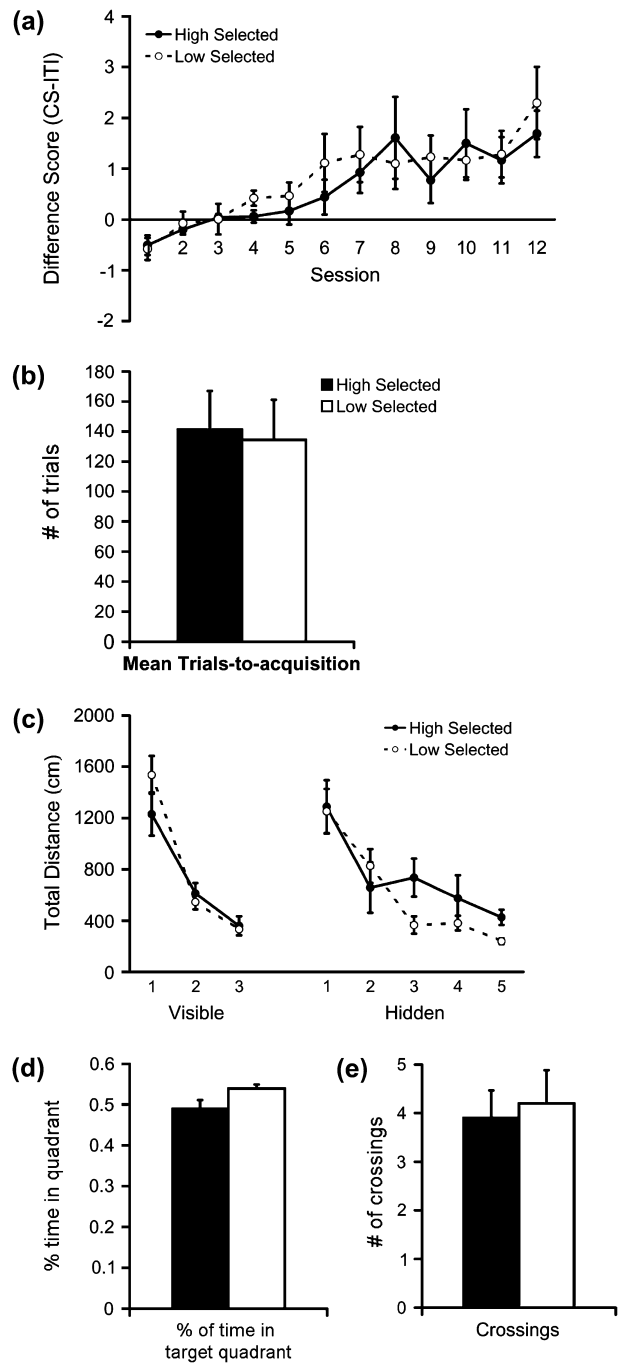


Figure 5: Evaluation of two tests of learning in the S_5 high and low selected lines. (a) Pavlovian approach conditioning: the number of nose pokes made during the 8-second CS presentation minus the number of nose pokes made in the 8 seconds just prior to the presentation of the CS. (b) Mean trials-to-acquisition in Morris Water Maze. (c) Total distance traveled in the target quadrant on the probe trial (day 9) when the platform was removed. (d) Number of times the mice crossed the target location where the platform had previously been located. Each test used 10 mice per selected line. All values are means \pm standard error.

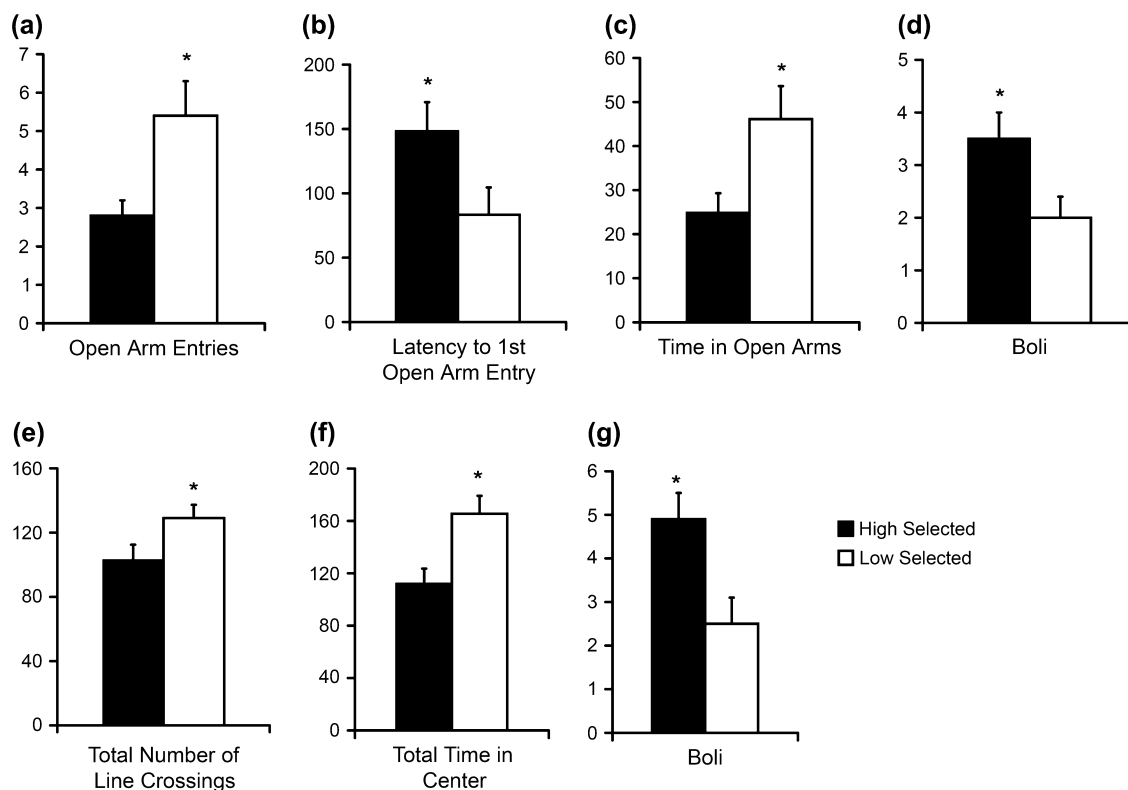


Figure 6: Anxiety-like behavior in S_4 high and low selected lines in two anxiety tests. *Elevated zero maze (a–d).* (a) Number of open arm entries in the zero maze. (b) Time until first entry into the open arm of the zero maze. (c) Time spent in the open arms of the zero maze over the 5 min test session. (d) Number of fecal boli left in the zero maze at the end of testing. *Open field test (e–g).* (e) Number of line crossings observed during the test session. (f) Total time spent in the center of the open field. (g) Number of fecal boli left in the open field at the end of the 15 min test. For all figures each bar represents the behavior of 24 mice. All values are mean \pm standard error. * $P < 0.05$ for comparisons between the high and low lines.

section entrances (Fig. 6a; $t_{461} = 2.69$; $P < 0.05$) as compared to the high line. The lines did not differ significantly in the total number of line crossings, although the low line showed a trend toward higher numbers of crossings ($t_{461} = 1.84$; $P = 0.07$). The low line also took less time to make the first entry into the open sections of the maze (Fig. 6b; $t_{461} = 2.10$; $P < 0.05$), and produced fewer numbers of fecal boli over the course of the test (Fig. 6d; $t_{451} = 2.32$; $P < 0.05$).

Open field

To further assess anxiety-like behavior in the selected lines we also examined behavior on the open field. The number of line crossings was significantly different (Fig. 6e; $F_{1, 41} = 8.22$; $P < 0.01$), with the low line making more line crossings than the high line. The low line spent more time in the center of the open field than did the high line (Fig. 6f; $F_{2, 821} = 7.86$; $P < 0.01$) and produced fewer fecal boli over the course of the test (Fig. 6g; $t_{461} = 2.42$; $P < 0.01$). A main effect of time bin on activity counts was also observed ($F_{2, 821} = 4.48$; $P < 0.05$), with increasing activity counts seen later in the 15 min session. No time bin by line interaction was apparent for

activity counts ($F_{2, 821} = 0.68$; $P > 0.25$), indicating that the change in activity seen over the course of the session was similar in both lines (data not shown).

Light/dark box

As a third means of assessing anxiety-like behavior in the selected lines we examined behavior on the light/dark box. Two-sample t -tests of all measures in the light/dark box indicated that there were no significant differences between the lines (all $P > 0.1$). The low line and the high line did not differ for the number of boli in the apparatus (Low: 0.5 ± 0.2 ; High: 0.3 ± 0.2), the latency to exit the light side of the box (Low: 13.9 ± 1.8 seconds; High: 15.4 ± 2.4 seconds), the number of light/dark transitions (Low: 27.7 ± 2.6 ; High: 25.5 ± 2.5), the latency to re-enter the light side (Low: 77.4 ± 23.4 seconds; High: 115.1 ± 18.1 seconds), the total number of lines crossed in the light side (Low: 24 ± 2.8 ; High: 18.6 ± 2.4), or total time spent in the light side of the box (Low: 130.1 ± 11.2 seconds; High: 134.3 ± 11.4 seconds). Based on these data we conclude that there was no difference between the lines for this test.

Gene expression measurement by Affymetrix microarray

The results of a 2-way ANOVA of microarray data for the factors brain region and line identified 183 probe sets that showed a main effect of line and that had q -values < 0.05 . Of the 183 probe sets identified using the AVOVA, 82 had a fold change of 1.5 or greater in at least one brain region (Supplementary Table 1). The list contained 8 genes represented by 2 probe sets, so 74 unique genes were identified by this analysis. We selected 6 genes for further examination with qPCR based on possible support for their roles in fear learning and anxiety in the published literature, statistical significance, magnitude of fold change, agreement between multiple probe sets and the location of the probe sets within the target gene (favoring exons, avoiding the 3' untranslated regions). These genes were found to be differentially expressed between the high and low hippocampal samples using qPCR, supporting the results of the microarray study (Table 1).

Genotyping of DNA from qPCR samples

Each of the six genes measured by qPCR was found to be significantly differently expressed by a t -test between line (high vs. low) and also in a one-way ANOVA for genotype at an adjacent SNP (B6, Het, D2; Table 1). The analysis of expression based on nearby SNPs supported the presence of *cis*-eQTL for all six genes. *Post hoc* comparisons were used to determine whether these *cis*-eQTLs showed a dominant/recessive or additive pattern of inheritance. All *cis*-eQTL were confirmed by expression data from the hippocampus of BXD RI lines available on line at www.genenetwork.org.

Genotyping and QTL analysis of markers from S₃ generation

The most significant P -values resulting from the analysis of the 305 SNPs genotyped in the S₃ breeders using the methods of Belknap (1997) are shown in Table 2. We found evidence for QTL in the selected lines relatively close to each

of the genes confirmed to be differentially expressed between the selected lines, suggesting a possible relationship between the differences in gene expression (*cis*-eQTL) and these QTL.

Discussion

Short-term, bidirectional selection for contextual FC produced mouse lines with extremely divergent fear learning. These lines also differed in their response to tone (CS) and altered context, consistent with previous reports (Bolivar *et al.* 2001; Fernández-Teruel *et al.* 2002; Radcliffe *et al.* 2000). Though short-term selection has previously been used to produce lines of mice with differential FC, Radcliffe *et al.* (2000), did not evaluate any additional phenotypes in those lines. Thus, it was not clear whether genetic differences in FC reflected a difference in the capacity to form fearful associations, or whether the difference in FC was due to trivial factors (e.g. shock sensitivity). Previous genetic studies of FC have not fully explored these possibilities (Caldarone *et al.* 1997; Fernández-Teruel *et al.* 2002; Talbot *et al.* 2003; Valentinuzzi *et al.* 1998; Wehner *et al.* 1997), nor have they adequately examined the putative genetic correlation between fear learning and anxiety-like behaviors.

Using the fourth selected generation, we sought to determine whether the lines differed in their maximal capacity for fear learning, or whether the low learning mice simply required additional training. Higher shock intensities or additional shocks produced greater freezing in both selected lines, as reported previously in both rats (Baldi *et al.* 2004) and mice (Laxmi *et al.* 2003); however, even at the highest intensities or number of shocks the low line never froze as much as the high line. Had the line difference been due to a difference in shock sensitivity, increasing shock intensity would have produced a rightward shift in the shock-response curve of the low line; instead, we observed a difference in maximum freezing.

The high line exhibited slower extinction of the freezing response, recapitulating the differences in acquisition and

Table 1: qPCR results

Abbreviation	Chr	Mb	qPCR fold change	Line with higher expression	SNP genotype with higher expression	Inheritance pattern
Cdh7	1	110	6.2	Low**	B6**	B6 Dominant
Mcm6	1	128.2	2.5	Low***	B6***	Additive
CAP1	4	121.9	2.3	Low*	D2*	B6 Dominant
CaMK2n1	4	137.3	2.0	High*	B6**	B6 Dominant
Gnb1	4	154	2.0	Low*	D2***	Additive
Supt16h	14	47.3	3.6	Low***	D2***	Additive

The abbreviation and location of each gene is listed. The fold change (by qPCR) and the significance of the t -test for the effect of selected line (high or low) on gene expression is given. The nearby SNP genotype (B6 of D2) that was associated with higher expression is listed along with the significance of the one-way ANOVA for genotype. The inheritance pattern (dominant or additive) as determined by comparison of the heterozygote group to the two homozygotes is listed.

* $P < 0.05$, ** $P < 0.001$, *** $P < 0.00001$

Table 2: QTL analysis

SNP	Chr	Mb	cM	P-value
rs3022830	1	116	63	0.0001
rs4224940	4	146	72	0.0097
rs4225248	5	65	41	0.0007
rs13480650	10	74	40	0.0022
rs13481924	13	84	45	0.0005
rs3090594	14	49	22	0.0004
rs4232247	19	47	44	0.0056

The RefSNP accession ID (rs) for each SNP is listed, along with the chromosome and location (in Mb and cM), and the significance of the difference in allele frequency in S_3 breeders as determined by the method of Belknap (1997).

extinction of fear learning observed in anxious human populations (Lissek *et al.* 2005).

We evaluated whether selection had altered fear learning generally, or whether it merely altered the propensity to freeze in response to a fear eliciting stimulus. Fear potentiated startle (FPS) measures the increase of the startle response in the presence of a fearful cue and is therefore not confounded by differences in freezing or locomotor behaviors. Only the high freezing line exhibited FPS, indicating the genetic changes underlying selection altered the central formation of fearful associations, rather than simply the propensity to freeze. These are the first data that we are aware of to suggest that these two tests of fear learning have a common genetic basis. This genetic association is in contrast to the known neuroanatomical dissociation between freezing to context and FPS (McNish *et al.* 1997). Interestingly, FPS is diminished by pretreatment with anxiolytic drugs (Falls *et al.* 1997; McCaughan *et al.* 2000), suggesting that it is sensitive to pharmacology manipulations that alter anxiety-like behavior in both rodents and humans (Risbrough *et al.* 2003; Santos *et al.* 2005; Walker & Davis 1997).

One possible explanation for the effect of selection on freezing is that selection caused hyperactivity in the low line and hypoactivity in the high line, leading to an apparent difference in freezing that is secondary to competing differences in locomotor drive. We selected for a difference between freezing on days 1 and 2 specifically to avoid inadvertently selecting for differences in baseline locomotor activity. Nevertheless, we observed a small but significant difference in the number of line crossings in the open field test (Fig. 6e). FPS measures fear learning without depending on freezing behavior. Therefore the difference in FPS suggests that selection did not act exclusively by changing activity levels. In addition, we observed no difference in freezing on day 2 when shocks were not administered on day 1 (Fig. 2a–c), showing that activity differences do not account for the difference in freezing.

Selection could have changed the frequency of alleles that influence the sensitivity to foot shock. Such a difference would explain the observed response to selection, but would not reflect a true difference in fear learning. We measured the acute response to five trials at each of five different shock

intensities in 40 mice (1000 total measurements) but were unable to identify any differences. Previous studies have used similar approaches to identify differences in shock sensitivity due to both environmental (Kosten *et al.* 2005) and genetic (Shumyatsky *et al.* 2002) factors.

We assessed behavior in an appetitive learning task and in the Morris water maze to determine whether differences in fear learning generalized to other tests of learning. We found no differences in these tests, indicating that selection had altered fear learning, but not these other forms of learning. Similarly, Owens *et al.*, (1997) found no correlation between FC and the Morris water maze in a panel of inbred strains. Furthermore, in a B6D2F₂ cross, quantitative trait loci for spatial learning and an avoidance conditioning task mapped to different chromosomes (Steinberger *et al.* 2003). Our data indicate that alleles that effect fear learning do not overlap with alleles for general learning ability.

We examined the effect of selection for FC on anxiety-like behaviors. We found differences between the lines in the zero maze and open field but not the light dark box test, suggesting that these three tests are influenced by at least partially nonoverlapping alleles. This finding is not without precedent, and it was because we were aware that anxiety was not a unitary genetic construct that we chose to use a battery of tests. The observed differences in anxiety-like behavior strongly support the proposed genetic relationship between fear learning and anxiety-like behavior (Lissek *et al.* 2005). Emotionality is a construct that reflects the emotional reactivity of an animal to its environment (Flint *et al.* 1995; Hall 1934). Studies of emotionality have used selective breeding in rodents for anxiety-like behaviors, including defecation (Broadhurst 1975), open field activity (DeFries *et al.* 1978; Ramos *et al.* 2003), plus maze activity (Kromer *et al.* 2005; Liebsch *et al.* 1998), juvenile ultrasonic vocalizations (Dichter *et al.* 1996) and conditioned avoidance (Bignami 1965; DeFries *et al.* 1978; Liebsch *et al.* 1998). Using an F₂ cross of the inbred progeny from Bignami's selection, Fernández-Teruel *et al.* (2002) use factor analysis to identified one factor for learned fear and a second that included learned fear and anxiety-like behavior, which they termed emotional reactivity. Our results support the possibility of a genetically determined emotionality factor including both learned fear and anxiety. Our results are not confounded by possible environmental correlations, as is the case for correlations observed in an F₂ population (Palmer & Phillips 2002).

Prior studies have identified effects of maternal behavior (Cierpial *et al.* 1990; Francis *et al.* 1999) and ultrasonic vocalizations by pups (Dichter *et al.* 1996) on later life anxiety-like behavior. The response to selective breeding for apomorphine sensitivity was associated with differences in maternal behavior (Ellenbroek *et al.* 2000). Based on these results, we examined multiple measures of maternal behavioral (e.g. licking behavior) in our selected lines during the first two weeks of postnatal life (data not shown), but despite careful examination of extensive observational data, we were unable to identify any differences. Because there was no suggestion of a difference in maternal behavior, we did not attempt cross fostering experiments, which would have been required to establish the effect of maternal behavior on the observed phenotypic differences.

Previous studies using selection for anxiety-like behaviors have employed selection over many (15+) generations, which produces higher levels of inbreeding, resulting in changes in the frequencies of both trait-relevant and irrelevant alleles. While longer term selection produces a sustainable resource, the unintentional inbreeding can give rise to genetic differences that are unrelated to the selection phenotype. In order to avoid this problem, the creation of replicated selected lines has been suggested (Crabbe 1989) and employed (DeFries *et al.* 1978). Short-term selection studies such as ours are designed to minimize inbreeding by using large numbers of families over a very small number of generations.

Short-term selected lines are especially appropriate for gene expression studies (Palmer *et al.* 2005). We identified numerous gene expression differences between naïve mice from the fourth selected generation. Of six genes confirmed by qPCR, two are part of genetic pathways known to affect behavior. We identified differential expression of Calcium-Calmodulin-dependent protein kinase II inhibitor (Camk2n). Anxiety-like behavior induced by microinjection of CRF receptor agonists into rat brain has been shown to be blocked by CaMKII inhibitors (Rainnie *et al.* 2004). We also identified differential expression of guanine nucleotide binding protein, beta 1 (GNB1), which is a member of the 'Ras protein signal transduction' gene ontology (GO) category. We identified this GO category as overrepresented using ermineJ software (<http://www.bioinformatics.ubc.ca/ermineJ/>). RAS-GRF knockout mice have reduced contextual and cued FC (Brambilla *et al.* 1997). Of the other genes found to be differentially expressed, Cyclase associated protein homolog 1 (CAP1) has been identified to be differentially expressed in both a rat model of schizophrenia (Wong *et al.* 2005) and ethanol sensitive mice (MacLaren & Sikela 2005), indicating it may be important for a variety of behavioral phenotypes. Cadherin 7 (Cdh7) was represented twice in the top 20 statistically different transcripts and has been implicated in nervous system development. Finally, two of our confirmed differences are genes that have direct roles in DNA replication. Minichromosome maintenance deficient 6 (MCM6) affects DNA replication initiation, and is represented twice in our twenty most significantly differentially expressed genes, with a very high fold change in the hippocampus. Suppressor of Ty 16 homolog (Supt16h) affects translation elongation and is represented twice in the 10 most significantly differentially expressed transcripts. It is possible expression of MCM6 and Supt16h have a role in hippocampal neurogenesis; our data are the first that we are aware of to implicate these genes in emotional learning. Genotyping adjacent SNPs allowed us to associate higher expression of each gene with a particular allele (Table 1). Quantitative trait locus (QTL) for the behavioral differences between these strains were identified that are close to these six genes (Table 2).

We have interpreted the present data in terms of pleiotropic effects of some alleles on fear learning and anxiety. However it is possible that close genetic linkage of different alleles may be responsible for the observed correlations (Villafuerte & Burmeister 2003). Talbot *et al.* (2003) showed evidence that linkage, not pleiotropy, caused a genetic correlation between fear and anxiety for an allele

on chromosome 1. While the allele studied by Talbot *et al.* (2003) is not polymorphic between our strains, Talbot's data illustrates that genetic correlations do not always reflect pleiotropy.

The present data suggest that selection altered the frequency of alleles that influence multiple measures of learned fear and anxiety-like behavior, and that these differences are not secondary to trivial explanations such as differences in general learning ability, activity, shock sensitivity or the propensity to freeze in response to a fear eliciting stimulus. Therefore we conclude that efforts to identify the underlying genetic variants are warranted in the search for the genetic basis of emotionality.

References

- Baldi, E., Lorenzini, C.A. & Bucherelli, C. (2004) Footshock intensity and generalization in contextual and auditory-cued fear conditioning in the rat. *Neurobiol Learn Mem* **81**, 162–166.
- Belknap, J.K., Richards, S.P., O'Toole, L.A., Helms, M.L. & Phillips, T.J. (1997) Short-term selective breeding as a tool for QTL mapping: ethanol preference drinking in mice. *Behav Genet* **27**, 55–66.
- Bignami, G. (1965) Selection for high rates and low rates of avoidance conditioning in the rat. *Anim Behav* **13**, 221–227.
- Bolivar, V.J., Pooler, O. & Flaherty, L. (2001) Inbred strain variation in contextual and cued fear conditioning behavior. *Mamm Genome* **12**, 651–656.
- Broadhurst, P.L. (1975) The Maudsley reactive and nonreactive strains of rats: a survey. *Behav Genet* **5**, 299–319.
- Caldarone, B., Saavedra, C., Tartaglia, K., Wehner, J.M., Dudek, B.C. & Flaherty, L. (1997) Quantitative trait loci analysis affecting contextual conditioning in mice [see comments]. *Nat Genet* **17**, 335–337.
- Cierpial, M.A., Murphy, C.A. & McCarty, R. (1990) Maternal behavior of spontaneously hypertensive and Wistar-Kyoto normotensive rats: effects of reciprocal cross-fostering of litters. *Behav Neural Biol* **54**, 90–96.
- Crabbe, J.C. (1989) Genetic animal models in the study of alcoholism. *Alcohol Clin Exp Res* **13**, 120–127.
- Crabbe, J.C., Jr., Johnson, N.A., Gray, D.K., Kosobud, A. & Young, E.R. (1982) Biphasic effects of ethanol on open-field activity: sensitivity and tolerance in C57BL/6N and DBA/2N mice. *J Comp Physiol Psychol* **96**, 440–451.
- Crabbe, J.C., Young, E.R., Deutsch, C.M., Tam, B.R. & Kosobud, A. (1987) Mice genetically selected for differences in open-field activity after ethanol. *Pharmacol Biochem Behav* **27**, 577–581.
- Crawley, J. & Goodwin, F.K. (1980) Preliminary report of a simple animal behavior model for the anxiolytic effects of benzodiazepines. *Pharmacol Biochem Behav* **13**, 167–170.
- Davis, M. (1979a) Diazepam and flurazepam: effects on conditioned fear as measured with the potentiated startle paradigm. *Psychopharmacology (Berl)* **62**, 1–7.
- Davis, M. (1979b) Morphine and naloxone: effects on conditioned fear as measured with the potentiated startle paradigm. *Eur J Pharmacol* **54**, 341–347.
- Davis, M. (1992) The role of the amygdala in fear-potentiated startle: implications for animal models of anxiety. *Trends Pharmacol Sci* **13**, 35–41.
- DeFries, J.C., Gervais, M.C. & Thomas, E.A. (1978) Response to 30 generations of selection for open-field activity in laboratory mice. *Behav Genet* **8**, 3–13.
- Dichter, G.S., Brunelli, S.A. & Hofer, M.A. (1996) Elevated plus-maze behavior in adult offspring of selectively bred rats. *Physiol Behav* **60**, 299–304.
- Ellenbroek, B.A., Sluyter, F. & Cools, A.R. (2000) The role of genetic and early environmental factors in determining apomorphine susceptibility. *Psychopharmacology (Berl)* **148**, 124–131.
- Falconer, D.S. & Mackay, T.F.C. (1996) *Introduction to Quantitative Genetics*, 4th Edn. Longman, Essex, UK.

- Falls, W.A., Carlson, S., Turner, J.G. & Willott, J.F. (1997) Fear-potentiated startle in two strains of inbred mice. *Behav Neurosci* **111**, 855–861.
- Fernández-Teruel, A., Escorihuela, R.M., Gray, J.A., Aguilar, R., Gil, L., Giménez-Llort, L., Tobeña, A., Bhomra, A., Nicod, A., Mott, R., Driscoll, P., Dawson, G.R. & Flint, J. (2002) A quantitative trait locus influencing anxiety in the laboratory rat. *Genome Res* **12**, 618–626.
- Flint, J., Corley, R., DeFries, J.C., Fulker, D.W., Gray, J.A., Miller, S. & Collins, A.C. (1995) A simple genetic basis for a complex psychological trait in laboratory mice. *Science* **269**, 1432–1435.
- Francis, D., Diorio, J., Liu, D. & Meaney, M.J. (1999) Nongenomic transmission across generations of maternal behavior and stress responses in the rat. *Science* **286**, 1155–1158.
- Gallistel, C.R., Fairhurst, S. & Balsam, P. (2004) The learning curve: implications of a quantitative analysis. *Proc Natl Acad Sci USA* **101**, 13124–13131.
- Gentleman, R.C., Carey, V.J., Bates, D.M., et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80.
- Grillon, C., Dierker, L. & Merikangas, K.R. (1998) Fear-potentiated startle in adolescent offspring of parents with anxiety disorders. *Biol Psychiatry* **44**, 990–997.
- Hall, C.S. (1934) Emotional Behavior in the Rat I. Defecation and urination as measures of individual differences in emotionality. *J Comp Psychol* **18**, 385–403.
- Hettema, J.M., Annas, P., Neale, M.C., Kendler, K.S. & Fredrikson, M. (2003) A twin study of the genetics of fear conditioning. *Arch Gen Psychiatry* **60**, 702–708.
- Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U. & Speed, T.P. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264.
- Kamprath, K. & Wotjak, C.T. (2004) Nonassociative learning processes determine expression and extinction of conditioned fear in mice. *Learn Mem* **11**, 770–786.
- Kliethermes, C.L., Cronise, K. & Crabbe, J.C. (2004) Anxiety-like behavior in mice in two apparatuses during withdrawal from chronic ethanol vapor inhalation. *Alcohol Clin Exp Res* **28**, 1012–1019.
- Kosten, T.A., Miserendino, M.J., Bombace, J.C., Lee, H.J. & Kim, J.J. (2005) Sex-selective effects of neonatal isolation on fear conditioning and foot shock sensitivity. *Behav Brain Res* **157**, 235–244.
- Kromer, S.A., Kessler, M.S., Milfay, D., Birg, I.N., Bunk, M., Czibere, L., Panhuysen, M., Putz, B., Deussing, J.M., Holsboer, F., Landgraf, R. & Turck, C.W. (2005) Identification of glyoxalase-I as a protein marker in a mouse model of extremes in trait anxiety. *J Neurosci* **25**, 4375–4384.
- LaBar, K.S., LeDoux, J.E., Spencer, D.D. & Phelps, E.A. (1995) Impaired fear conditioning following unilateral temporal lobectomy in humans. *J Neurosci* **15**, 6846–6855.
- Laxmi, T.R., Stork, O. & Pape, H.C. (2003) Generalisation of conditioned fear and its behavioural expression in mice [erratum appears in *Behav Brain Res* 2994 May 5;1411(2):337]. *Behav Brain Res* **145**, 89–98.
- LeDoux, J.E. (2000) Emotion circuits in the brain. *Annu Rev Neurosci* **23**, 155–184.
- Liebsch, G., Montkowski, A., Holsboer, F. & Landgraf, R. (1998) Behavioural profiles of two Wistar rat lines selectively bred for high or low anxiety-related behaviour. *Behav Brain Res* **94**, 301–310.
- Lissek, S., Powers, A.S., McClure, E.B., Phelps, E.A., Woldehawariat, G., Grillon, C. & Pine, D.S. (2005) Classical fear conditioning in the anxiety disorders: a meta-analysis. *Behav Res Ther* **43**, 1391–1424.
- McCaughran, J.A., Jr., Bell, J., 3rd & Hitzemann, R.J. (2000) Fear-potentiated startle response in mice: genetic analysis of the C57BL/6J and DBA/2J intercross. *Pharmacol Biochem Behav* **65**, 301–312.
- McNish, K.A., Gewirtz, J.C. & Davis, M. (1997) Evidence of contextual fear after lesions of the hippocampus: a disruption of freezing but not fear-potentiated startle. *J Neurosci* **17**, 9353–9360.
- Meshi, D., Drew, M.R., Saxe, M., Ansorge, M.S., David, D., Santarelli, L., Malapani, C., Moore, H. & Hen, R. (2006) Hippocampal neurogenesis is not required for behavioral effects of environmental enrichment. *Nat Neurosci* **9**, 729–731.
- Moldin, S.O. (2000) Neurobiology of anxiety and fear: challenges for genomic science of the new millennium. *Biol Psychiatry* **48**, 1144–1146.
- Owen, E.H., Logue, S.F., Rasmussen, D.L. & Wehner, J.M. (1997) Assessment of learning by the Morris water task and fear conditioning in inbred mouse strains and F1 hybrids: implications of genetic background for single gene mutations and quantitative trait loci analyses. *Neuroscience* **80**, 1087–1099.
- Palmer, A.A. & Phillips, T.J. (2002) Effect of forward and reverse selection for ethanol-induced locomotor response on other measures of ethanol sensitivity. *Alcohol Clin Exp Res* **26**, 1322–1329.
- Palmer, A.A., Verbitsky, M., Suresh, R., Kamens, H.M., Reed, C.L., Li, N., Burkhart-Kasch, S., McKinnon, C.S., Belknap, J.K., Gilliam, T.C. & Phillips, T.J. (2005) Gene expression differences in mice divergently selected for methamphetamine sensitivity. *Mamm Genome* **16**, 291–305.
- Paylor, R., Tracy, R., Wehner, J. & Rudy, J.W. (1994) DBA/2 and C57BL/6 mice differ in contextual fear but not auditory fear conditioning. *Behav Neurosci* **108**, 810–817.
- Petkov, P.M., Ding, Y., Cassell, M.A., Zhang, W., Wagner, G., Sargent, E.E., Asquith, S., Crew, V., Johnson, K.A., Robinson, P., Scott, V.E. & Wiles, M.V. (2004) An efficient SNP system for mouse genome scanning and elucidating strain relationships. *Genome Res* **14**, 1806–1811.
- Radcliffe, R.A., Lowe, M.V. & Wehner, J.M. (2000) Confirmation of contextual fear conditioning QTLs by short-term selection. *Behav Genet* **30**, 183–191.
- Ramos, A., Correia, E.C., Izidio, G.S. & Bruske, G.R. (2003) Genetic selection of two new rat lines displaying different levels of anxiety-related behaviors. *Behav Genet* **33**, 657–668.
- Richardson, M.P., Strange, B.A. & Dolan, R.J. (2004) Encoding of emotional memories depends on amygdala and hippocampus and their interactions. *Nature Neuroscience* **7**, 278–285.
- Risbrough, V.B., Brodtkin, J.D. & Geyer, M.A. (2003) GABA-A and 5-HT1A receptor agonists block expression of fear-potentiated startle in mice. *Neuropsychopharmacology* **28**, 654–663.
- Santos, J.M., Gãrgaro, A.C., Oliveira, A.R., Masson, S. & Brandão, M.L. (2005) Pharmacological dissociation of moderate and high contextual fear as assessed by freezing behavior and fear-potentiated startle. *Eur Neuropsychopharmacol* **15**, 239–246.
- Shumyatsky, G.P., Tsvetkov, E., Malleret, G., Vronskaya, S., Hatton, M., Hampton, L., Batten, J.F., Dulac, C., Kandel, E.R. & Bolshakov, V.Y. (2002) Identification of a signaling network in lateral nucleus of amygdala important for inhibiting memory specifically related to learned fear. *Cell* **111**, 905–918.
- Steinberger, D., Reynolds, D.S., Ferris, P., Lincoln, R., Datta, S., Stanley, J., Paterson, A., Dawson, G.R. & Flint, J. (2003) Genetic mapping of variation in spatial learning in the mouse. *J Neurosci* **23**, 2426–2433.
- Storey, J.D. & Tibshirani, R. (2003) Statistical methods for identifying differentially expressed genes in DNA microarrays. *Methods Mol Biol* **224**, 149–157.
- Talbot, C.J., Radcliffe, R.A., Fullerton, J., Hitzemann, R., Wehner, J.M. & Flint, J. (2003) Fine scale mapping of a genetic locus for conditioned fear. *Mamm Genome* **14**, 223–230.
- Valentinuzzi, V.S., Kolker, D.E., Vitaterna, M.H., Shimomura, K., Whiteley, A., Low-Zeddies, S., Turek, F.W., Ferrari, E.A., Paylor, R. & Takahashi, J.S. (1998) Automated measurement of mouse freezing behavior and its use for quantitative trait locus analysis of contextual fear conditioning in (BALB/cJ x C57BL/6J)F2 mice. *Learn Mem* **5**, 391–403.
- Verbitsky, M., Yonan, A.L., Malleret, G., Kandel, E.R., Gilliam, T.C. & Pavlidis, P. (2004) Altered hippocampal transcript profile accompanies an age-related spatial memory deficit in mice. *Learn Mem* **11**, 253–260.
- Villafuerte, S. & Burmeister, M. (2003) Untangling genetic networks of panic, phobia, fear and anxiety. *Genome Biol* **4**, 224.
- Wahlsten, D., Bachmanov, A., Finn, D.A. & Crabbe, J.C. (2006) Stability of inbred mouse strain differences in behavior and brain size between laboratories and across decades. *Proc Natl Acad Sci USA* **103**, 16364–16369.
- Walker, D.L. & Davis, M. (1997) Double dissociation between the involvement of the bed nucleus of the stria terminalis and the central nucleus of the amygdala in startle increases produced by conditioned versus unconditioned fear. *J Neurosci* **17**, 9375–9383.
- Wehner, J.M., Radcliffe, R.A., Rosmann, S.T., Christensen, S.C., Rasmussen, D.L., Fulker, D.W. & Wiles, M. (1997) Quantitative trait

locus analysis of contextual fear conditioning in mice [see comment]. *Nat Genet* **17**, 331–334.

Acknowledgments

This work was supported by MH70933, AA10270, AA12714, a NARSAD young investigator award and a grant from the Department of Veterans Affairs.

Supplementary material

The following supplementary material is available for this article:

Table S1: Transcripts that had a statistically significant main effect of selected line (high or low) with a q -value < 0.05 . The gene name and abbreviation and the Affymetrix probeset ID are listed, along with the chromosome and location (in Mb) and the position of the probeset as determined by using the BLAT alignment tool (<http://genome.ucsc.edu/>) to compare the probeset to the mouse genome (Build 36). The q -values (<http://faculty.washington.edu/~jstorey/qvalue/>) were derived from the P -value distribution of a two-way ANOVA for

Correlation of contextual fear, anxiety and gene expression

the factors line and region. The fold change for just the amygdala or just the hippocampal samples are shown; white cells are high $>$ low and shaded cells are low $>$ high. The fold change between the brain regions (with both amygdala and hippocampal samples included) is also shown; white cells are amygdala $>$ hippocampus and shaded cells are hippocampus $>$ amygdala. Finally, the presence or absence of the gene in each group (e.g. 'A L' = amygdala, low selected) as determined by the 'MAS5Calls' command in Bioconductor is shown, 'P' denotes a present call in at least 4 of the 6 microarrays, 'A' denotes an absent call in 4 or more of the 6 microarrays.

This material is available as part of the online article from: <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1601-4183X.2007.00306.x>

(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.