Second-generation high-throughput forward genetic screen in mice to isolate subtle behavioral mutants

Vivek Kumar, Kyungin Kim, Chryshanthi Joseph, Lisa C. Thomas, Heekyung Hong[1], and Joseph S. Takahashi[2]

Department of Neuroscience, Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, TX 75390-9111

Forward genetic screens have been highly successful in revealing roles of genes and pathways in complex biological events. Traditionally these screens have focused on isolating mutants with the greatest phenotypic deviance, with the hopes of discovering genes that are central to the biological event being investigated. Behavioral screens in mice typically use simple activity-based assays as endophenotypes for more complex emotional states of the animal. They generally set the selection threshold for a putative mutant at 3 SDs (z score of 3) from the average behavior of normal animals to minimize false-positive results. Behavioral screens using a high threshold for detection have generally had limited success, with high false-positive rates and subtle phenotypic differences that have made mapping and cloning difficult. In addition, targeted reverse genetic approaches have shown that when genes central to behaviors such as open field behavior, psychostimulant response, and learning and memory tasks are mutated, they produce subtle phenotypes that differ from wild-type animals by 1 to 2 SDs (z scores of 1 to 2). We have conducted a second-generation (G2) dominant N-ethyl-N-nitrosourea (ENU) screen especially designed to detect subtle behavioral mutants for open field activity and psychostimulant response behaviors. We successfully detect mutant lines with only 1 to 2 SD shifts in mean response compared with wild-type control animals and present a robust statistical and methodological framework for conducting such forward genetic screens. Using this methodology we have screened 229 ENU mutant lines and have identified 15 heritable mutant lines. We conclude that for screens in mice that use activity-based endophenotypic measurements for complex behavioral states, this G2 screening approach yields better results.

ethylnitrosourea mutagenesis | forward genetic screen | mouse behavior screen | open field behavior | psychostimulant response

Forward genetic screens are one of the cornerstones of modern biology, leading to the successful identification of numerous genes and pathways regulating complex biological phenomena. One hundred years ago, Thomas Hunt Morgan published two articles describing a series of overt eye and wing mutants in *Drosophila* created using mutagens (1, 2). He used these overt mutants as markers for genetic mapping, laying the foundations of modern genetics. Since then screens in model organisms have been successful in many phenotypic domains (3–7).

Forward genetic screens for complex behavior were pioneered by Seymour Benzer, who in a landmark study in 1967 used a countercurrent distribution procedure to isolate fly lines with phototactic defects (8). Although phototaxis had been previously used as an assay in *Drosophila*, Benzer had four key insights that allowed him to use it as an assay to isolate genetic mutants. First, he established a quick, simple, and high-throughput assay to screen thousands of flies in a single day. Second, rather than outbred fly stocks that were normally used (9), he used a genetically isogenic strain of flies to minimize phenotypic variance. Third, he screened through several isogenic strains before settling on one, Canton-S, which unlike the others has a very robust phototactic response. Fourth, to increase the probability of finding mutants, he used a potent mutagen, ethyl methanesulfonate, to induce unique mutations in the screening population. Benzer demonstrated that phototaxis, a complex sensorimotor behavioral response, could be dissected genetically by a seemingly simple assay (8, 10).

These same principles are used in modern day screening protocols for behavioral mutants; however, their application in mice has been more challenging. Many of the behavioral endophenotypes that are routinely measured are highly variable, either because of the inherent nature of these measurements or because the behavior itself is highly influenced by environmental and stochastic factors (11). This variability also exists in fly assays but can be overcome by increasing the screening population or screening multiple progeny from individual lines rather than relying on individual flies. Increasing the number of animals tested is often prohibitive with mice owing to cost, and instead the threshold for detection of mutants is raised to as high as 3 SDs to minimize false-positive results. In addition, locomotor output assays are often used as endophenotypes for complex emotionality, and these outputs can be affected by systems independent of the behavioral pathway being investigated, leading to identification of animals with pleiotropic phenotypes (12, 13). Moreover, mutations in genes known to regulate behaviors such as learning and memory, psychostimulant response, and open field behavior usually only produce a shift in the mean of 1 to 2 SDs (11, 13–15). Thus, when genetic screens are conducted to search for "strong" (greater than 3 SDs) mutants, the investigator is attempting to discover mutants that are more extreme than what would normally be found if a key gene, known to mediate the behavior, was deleted.

Exploratory behavior in an open field measures an animal's "emotionality" (16, 17). Ambulation, thigmotaxis, defecation, urination, sniffing, grooming, freezing, and many other measurements can be used as endophenotypes for an animal's mood or emotional state (18). Locomotor activity after administration of a psychoactive drug is a measure of an animal's sensitivity to the drug. Although not a test of addiction, the sensitivity and sensitization of locomotor response to drugs of abuse is correlated with an animal's propensity to become addicted to that drug (19, 20). Individual differences in response to drugs of abuse due to genetic factors have been well established in humans as well as mouse model systems (21). Selective breeding and quantitative trait loci (QTL) analysis have shown that both measurements routinely performed in the open field assay and psychostimulant response have high heritability, indicating that genetic factors make large contributions to the phenotype in mice and humans (21–24).

Here we conducted a screen specifically designed to discover mutants that are 1 to 2 SDs from the mean of normal animals. We analyzed open field assay (OFA) and psychostimulant response assay (PSY) in a second-generation (G2) screening strategy from

more than 2,300 mice from 230 families and analyzed more than 7,000 h of video to discover mutants in both these phenotypes.

Results

Screening Strategy. Typical behavioral screens in mice test generation 1 (G1) progeny for dominant and semidominant phenotypes or generation 3 (G3) progeny for recessive phenotypes using strict cutoffs such as 3 SDs (z score >3) to minimize the number of false-positive results (type I error) (25, 26). With the exception of circadian behavior, in large-scale recessive screens we have observed that the majority of behavioral mutants have subtle phenotypic effects (z score 1 to 2) (11, 27). To find dominant or semidominant mutants that are 1 to 2 SDs from the mean of a control screening population more efficiently, we used a G2 screening strategy. G1 mice were bred one more generation by back-crossing to C57BL/6J females to generate G2 mice that were screened (Fig. 1A). If there is a mutation affecting the phenotype of interest then 50% of the G2 mice should be carriers for this mutation on average. By screening 10 G2s from each G1 we are able to sample each mutant genome multiple times and screen individual families or lines rather than individual G1 mice. We hypothesized that any causative mutation would shift the mean of the entire G2 family in a statistically significant manner. To determine how many G2s should be tested for each G1 animal we carried out computational simulations to determine the optimal screening strategy. We simulated two populations with 1 SD difference in mean and randomly selected 5, 10, 15, or 20 individuals from these two distributions. The means from these two draws were compared in a t test to determine whether they were statistically significant in their difference. This was repeated 100,000 times, and the distribution of the P values was plotted (Fig. S1). We determined that screening 20 animals from each G1 would result in a P value of ≤0.05 more than 80% of the time. We choose to generate 10 G2 from each G1 and test each G2 twice, thus giving us 20 G2 tests for each G1 animal (Fig. 1A).

Behavioral Tests and Analysis Conducted. Two hundred twenty-nine mutagenized G1s were placed into mating to produce 2,306 G2s that were screened for OFA and PSY. On average, 10 G2s from each G1 were generated, and each G2 was tested twice over 2 wk for OFA and PSY. Because of the limited number of progeny being produced from each G1, males and females and the two tests were treated as independent. The 90-min test was split into two parts (Fig. 1B). The first 30 min (−30 to 0; Fig. 1B) consisted of the OFA, after which the animals were i.p. injected with cocaine (20 mg/kg; red arrow in Fig. 1B) and observed for a further 60 min (0–60; Fig. 1B) for PSY. This test was repeated 7 d after the first test. The testing chamber is a 55 cm × 55 cm × 36 cm well-lit matrix (700 ± 56 lx). For OFA the chamber is divided into a 5 × 5 grid consisting of 25 zones, and the time and distance an animal spent in the center, periphery, and corners was measured. We also measured the distance an animal traveled in the various zones during the test and crossing between the periphery and center. For PSY we analyzed the velocity of the animals before and after injection with cocaine. A detailed explanation of test measures is given in Fig. S2. All animals were tracked using video-based software (LimeLight; Actimetrics), and the videos are available upon request. Representative tracking data from eight animals are shown in Fig. 1C. Each row represents tracking data from one animal, and each frame shows 10 min of tracking data. During the open field part of the test, animals show exploratory behavior in the matrix, with more time spent in the corners and periphery than in the center. After injection with cocaine, the velocity of the animals increases significantly. During the postinjection phase [third, fourth, and fifth column (from left), Fig. 1C], the animal is running in the periphery with elevated velocity (Fig. 1B). As the drug effect wears off, animals return to preinjection activity level (Fig. 1B and last column, Fig. 1C).

Sixteen measures from this test were extracted and used in our analysis focusing on OFA distance traveled, OFA time in various zones, and PSY (Fig. S2). To visualize relationships between the measured traits, we constructed a correlogram (28). The correlogram used here iconically displays the correlation matrix, with
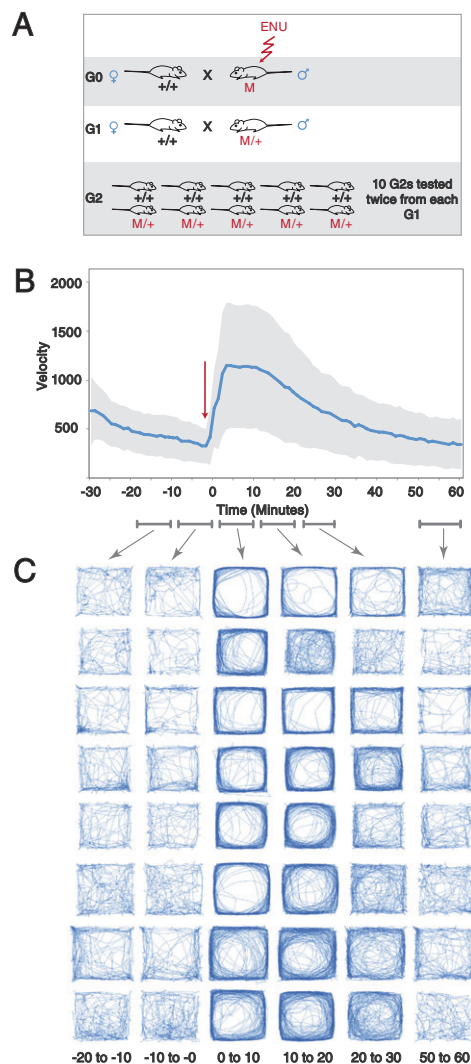
Fig. 1. Breeding scheme and behavioral sample data from screen. (A) G2 breeding scheme to generate the screening animals. Each mutagenized G0 male was bred to a wild-type female to produce the G1 population. The G1s were bred one more generation to produce 10 G2s from each G1. This G2 generation was then screened for phenotypic deviance. (B) Velocity data from 392 control WT001 animals. Velocity averaged for each minute for the 90 min of the behavior test. Injection with cocaine occurs at 0 min (red arrow). The 30 min before injection of cocaine is used for OFA, and the 60 min after injection is the PSY. The gray shaded region represents 1 SD range of the data. (C) Tracking data for eight representative mice from the screen. Each row represents data from one animal. Each frame is 10 min of tracking data. The third column (from left) represents tracking immediately after injection with cocaine.

the magnitude and direction of correlation indicated by color intensity (Fig. 2A, *Lower*) and bivariate concentration ellipse containing 68% of the data points (1 SD ellipse) with a LOWESS fit through the data (Fig. 2A, *Upper*). As expected, many of the measured behaviors are correlated with each other. For instance, all three of the psychostimulant response measures are correlated, as are the OFA distance and time measures (Fig. 2A). Given the correlative nature of the data, we wanted to determine the independent factors and extract a combined behavioral score for each of these factors. One exploratory data analysis tool that can be used to understand multidimensional data is principal components analysis (PCA). PCA uses linear weighted composites of measured traits (the behavioral responses) to yield the primary components or factors that can account for the total variance. Although not confirmatory like structural equation modeling,
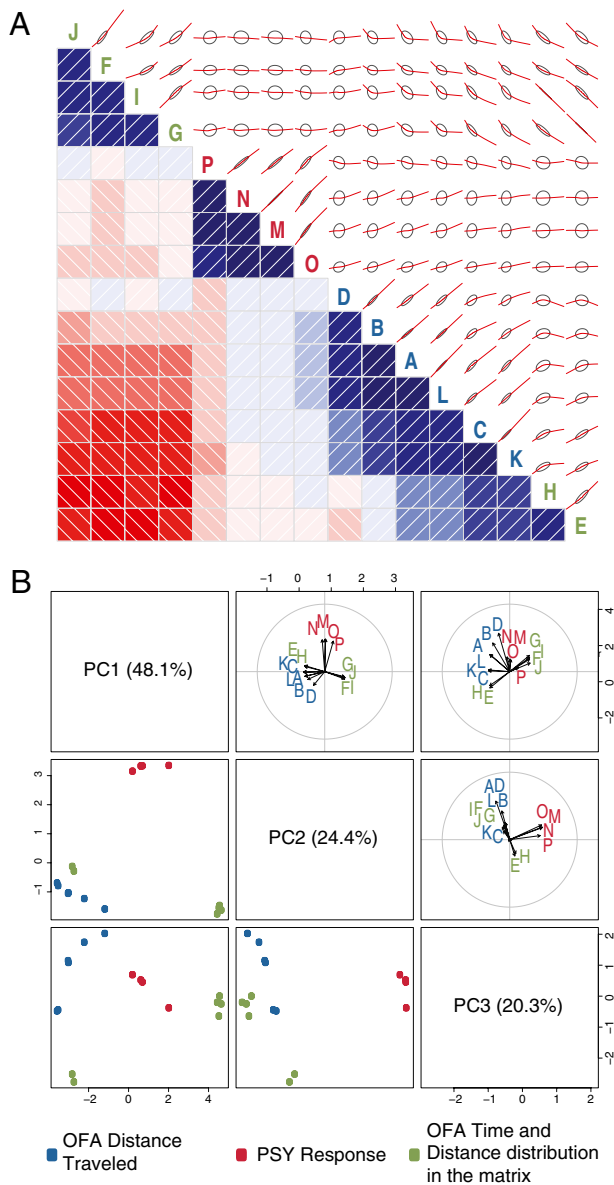
Fig. 2. Measured behavioral traits fall into distinct but correlated independent groups. (*A*) Correlogram ordered by the first two principal components. In lower half, blue and red indicate positive and negative correlation, respectively, and increasing shade of each color indicates increasing correlation coefficient. Upper half represents a schematic scatterplot in which the 4,600 individual data points have been omitted for clarity. Ellipse is a 1-SD circle that covers 68% of the value around the mean fit of the distribution. Red curve is a LOWESS smoothened curve to the scatterplot data. (*B*) PCA indicates that the first three factors account for 92.8% of the phenotypic variance. Biplot vector analysis indicates clustering of psychostimulant response, OFA distance traveled, and OFA time and distance distribution in the matrix. The vector biplot is shown on the top of the scatterplot matrix. The bottom of the scatterplot shows the results from the vector biplot scaled. The phenotypes are color coded, with OFA distance traveled measures in blue, OFA time and distance distribution in the matrix in green, and PSY in red. The individual phenotypes have been coded as described in Fig. S2. A, total distance; B, peripheral distance; C, center distance; D, corner distance; E, percent distance center; F, percent distance corner; G, percent distance periphery; H, percent time center; I, percent time periphery; J, percent time corner; K, crossing; L, baseline; M, 30-min average; N, sum 30-min average; O, sum 60-min average; P, net response.

PCA can yield important clues about the latent structure of the data. In addition, the factor loadings can be used to calculate a combined factor score that can be used as a calculated behavioral score for screening. Before conducting PCA, all phenotypes were confirmed to be normally distributed (Fig. S3 *A and B*). We carried out PCA on the 16 behavioral scores and 4,600 tests. In this dataset three major components or eigenvalues explained more than 92% of the phenotypic variance. Analysis of the PCA loadings indicated that PC1, which accounts for 48% of the variance, loads highly for OFA measurements. PSY measurements load highly on PC2, which accounts for 24.4% of the variance, and a subset of OFA measurements load highly on PC3, which account for 20% of the variance. To further interpret the biological basis of these three components and to visualize their relationship, we constructed a biplot scatterplot matrix (Fig. 2*B*). Biplots are helpful in revealing multicollinearity and clustering of variables used in PCA. Mathematically they represent a graphic display of matrix multiplication that can be used to represent any two-way table (29). The biplot scatterplot matrix of the first three principal components and an analysis of the component loadings show that OFA distance traveled, OFA time spent in various areas of the matrix, and psychostimulant response form three distinct clusters (Fig. 2*B*). Vector biplot analysis (Fig. 2*B*, *Upper*) as well as scaled biplot analysis (Fig. 2*B*, *Lower*) show that PSY measurements cluster together (Fig. 2*B*, red points). The OFA distance traveled measurements cluster with crossing from periphery to center measurement (Fig. 2*B*, blue points), and OFA time spent and distance traveled cluster together (Fig. 2*B*, green points). The only exception seems to be OFA percent time and distance in center, which cluster with OFA distance measurements. PCA analysis indicates three independent factors: PC1, OFA activity as measured by the distance an animal travels (OFA distance); PC2, response to cocaine (PSY response); and PC3, OFA activity as measured by where an animal spends its time, particularly in the periphery and corners of the matrix (OFA time and distance distribution). Thus, how much an animal travels, where it travels, and how much it responds to cocaine seem to be independent. Ambulation in the open field has been interpreted as a measure of fearfulness or anxiety, as well as an animal's tendency for exploratory behavior (30). Using factor loadings from the first three principal components, three composite behavioral scores were calculated for each animal tested, and the scores were treated as a calculated behavioral trait (PC1, PC2, and PC3).

Next we were interested in finding G2 families that have multiple individuals that are behaviorally different in the measured phenotypes compared with control animals. To visualize behavioral data for each phenotype, the data of each family for all 19 phenotypic scores were plotted (Fig. 3 and Fig. S4 *A–D*). Fig. 3 shows four representative phenotypes from the 19 phenotypes used in analysis. Fig. S4 *A–D* contains a detailed phenotype of each animal separated by family and by phenotype. Each row represents one family, with the mean of their family displayed as a point, and the line represents ± 1 SD range. For visual aid, the red lines represent mean and 1 SD range of the WT001 control family (*n* = 392). WT001, the first family in each graph, is the unmutagenized control group. Progeny from WT001, unmutagenized C57BL/6J matings, were produced throughout the screen and tested alongside mutagenized screening population for proper control. This family serves as the group with which all of the other animals are compared. Most of the mean values for each family that were screened fall within 1 SD of the control population; however, even cursory examination reveals families that are deviants in certain phenotypes. For instance, ENU042 has high corner time (Fig. 3*C*), low OFA total distance (Fig. 3*B*), and high PC1 score (Fig. 3*D*), although it is normal for PSY (Fig. 3*A*). ENU100 has low psychostimulant response (Fig. 3*A*) but seems normal in all other phenotypes (Fig. 3 *B–D*). ENU172 seems to be affected in all four phenotypes shown in Fig. 3. Even though visual inspection and shift in mean of family can imply that a line is interesting, we wanted to use significance in statistical tests as strict criteria to determine whether a family is mutant.

Unlike typical genetic screens, which use cutoffs or simply select the most extreme animals in a distribution, we used statistical analysis to determine which G2 family differed significantly from
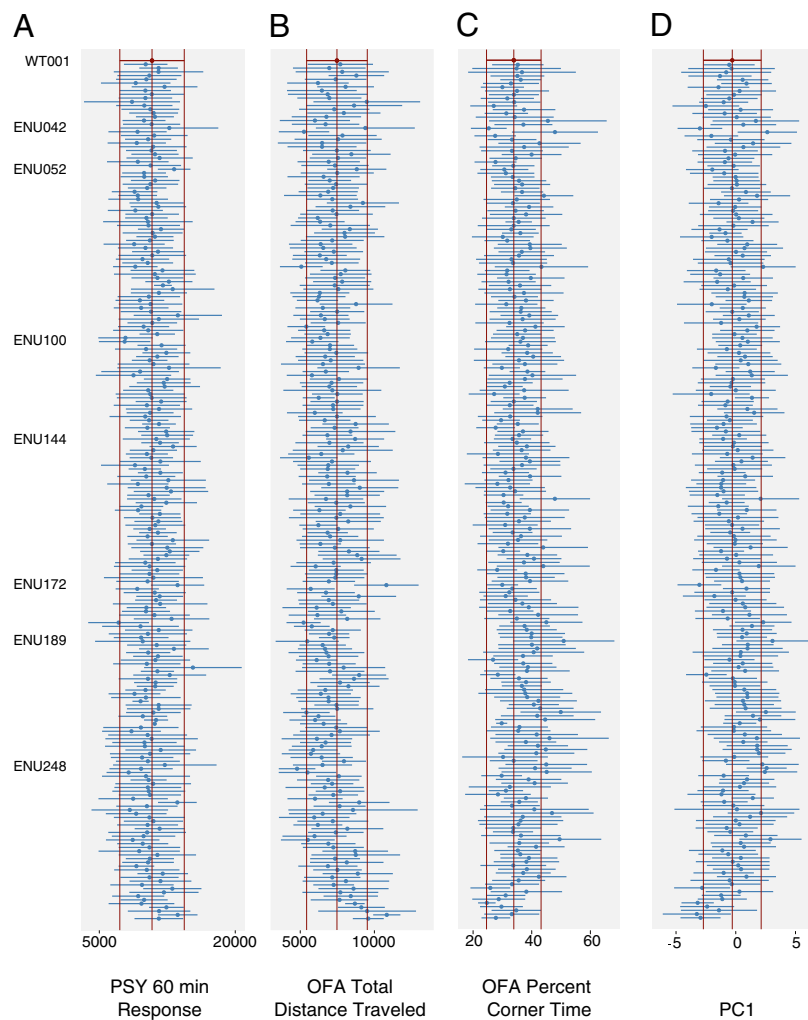
Fig. 3. Representative phenotype plots for all 230 families that were screened. Four of the 19 phenotypes are shown in this figure; a complete dataset for all of the phenotypes is in Fig. S4. Each bar represents mean (dot) ± 1 SD (bar) for each family. The control family, WT001, is shown in red at the beginning of each panel, and the mean ± 1 SD range is drawn throughout the graph for comparison. Each family that is screened is in sequential order. Mutant families that are mentioned in the text are labeled on the left. Data for each animal for each family is shown in Fig. S4 *A–D*.

the control population (designated WT001). In statistical hypothesis testing, when a large number of hypotheses (*m* hypotheses) are tested simultaneously and the final inferences need to be valid across all experimental variables that are being tested, the probability of committing type I error increases as *m* increases dramatically. In this screen we have 230 groups, and even at a very conservative $\alpha = 0.001$, the probability of committing at least one type I error is 28.12% $(1 - [1 - \alpha]^m)$. To control for multiple testing we used Dunnett's test with step-down procedure using the multcomp package in R (31–33). Step-down procedures in multiple testing offer a more powerful approach than single-step procedures and come in two flavors, step-up and step-down (34, 35). Step-down procedures, used in the current analysis, test ordered hypothesis and reject null hypothesis until the first nonrejection is reached using the closure principle to increase power of the statistical test (36). Because any hypothesis rejected by the single-step procedure will also be rejected by the stepwise method but not vice versa, the multistep procedures are considered more powerful (37).

Dunnett's test using the step-down method resulted in a *P* value for each family for each measured trait. Forty-two and 23 lines had at least one phenotype that met the criteria for significance at $\alpha = 0.05$ and 0.01, respectively. We initially plotted the $-\log P$ value for each family in each trait, as shown in Fig. S5, and noticed that certain G2 families showed significant differences in multiple phenotypes. This was expected because many of the measures are correlated. To visualize whether certain families have similar significant differences, the families were clustered according to the *P* values obtained from the Dunnett's test for multiple comparisons. Agglomerative hierarchical clustering places multidimensional

observations into groups or clusters according to the dissimilarity metric between observations. Cluster analysis based on *P* values revealed six classes of mutants that are significantly different from control animals (Fig. 4). These consist of various combinations of the 19 measured traits, and to our surprise these clusters or classes of mutants roughly fall into combinations of the three major principle components.

Each cluster represents a group of G2 families that have similar statistically significant differences in their respective phenotypes. For instance, cluster VI consists of lines that are affected in PSY measures only, whereas cluster V consists of lines that are affected in PSY as well as OFA distance measures. Biologically cluster V families are hypoactive or hyperactive animals in OFA and also have significantly different response to cocaine, and cluster VI are families that are normal in OFA but have a specific phenotype in cocaine response. Cluster I consists of lines that have a phenotype in the percent time and distance in corners only. This seems to be a highly specific phenotype that is a subset of PC1. Cluster II and IV are the two largest clusters and include all OFA phenotypes. These lines have phenotypic differences in percent time, percent distance, and total distance traveled; however, their psychostimulant response is not affected. Cluster III consists of families that are affected in OFA distance traveled measurements only. These animals are hyper- or hypoactive, but the relative distribution of where in the matrix they spend their time is not different from control animals. Clusters I, III, and VI are specific to OFA time, distance, and PSY respectively. Clusters II, IV, and V are combinations of
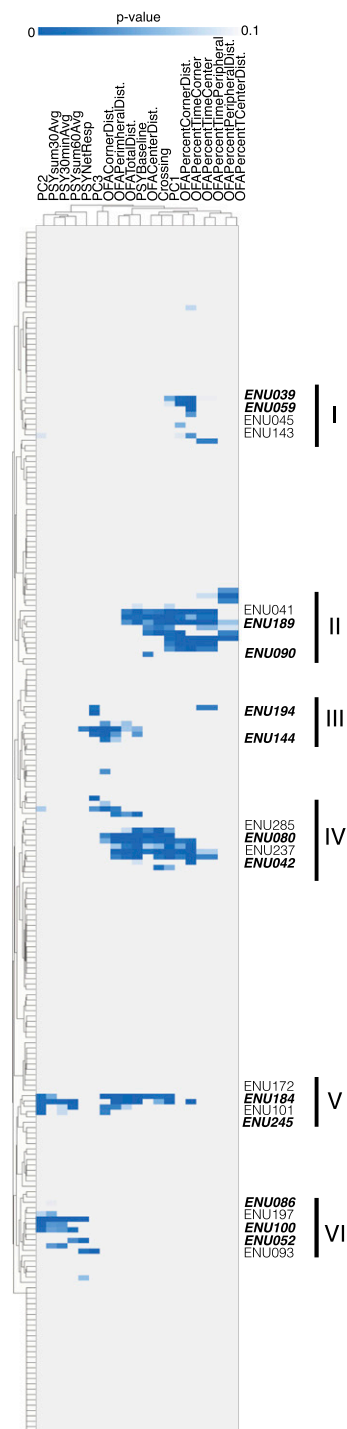
Fig. 4. Hierarchical clustering of phenotypes and families by *P* value reveals six distinct classes of mutants. Each G2 family was clustered according to *P* values obtained from Dunnett's test. The *y* axis represents each of 229 families tested, and the *x* axis represents each of the 19 phenotypes measured. Both the phenotypes and families are clustered. Another representation of these data without clustering in which the *P* value for each line is plotted by phenotype is shown in Fig. S5. The lines that were placed into mating for heritability testing are indicated on the right. Lines in bold and italic passed heritability testing, and lines in plain text did not pass heritability testing. Two heritable lines, WT248 and WT054, are not shown marked in the list of heritable lines in this figure.

these three specific clusters. Thus, clustering based on *P* value has revealed distinct classes of mutants.

Mutant Lines Are Heritable. In traditional genetic screens after a putative mutant is identified, the animal is bred and the genetic transmission of the mutant trait is confirmed in successive generations. This confirmation of "vertical transmission" is critical for identification of a true genetic mutant. By the nature of the screen we conducted, multiple individuals in a G2 family must be affected sufficiently to change the mean response in a statistically significant manner. A single extreme animal should not change the mean in a statistically significant manner. However, it is still possible that environmental factors could affect multiple individuals from the same line and change the overall phenotype of the line. Therefore, we wanted to confirm that lines displayed the mutant phenotype in multiple generations and are indeed heritable. We choose 27 lines and intercrossed the extreme animals together and analyzed the progeny for vertical transmission of the affected phenotype in the next generation. We specifically chose traits such as percent time in corner and periphery because these are traits for which standard screening cutoffs have been difficult to apply owing to high variability in measurement. The ENU042 line was originally identified as a family with low overall activity and high time spent in corners. This family is representative of the phenotype seen in cluster IV that is significantly different in OFA distance traveled and percent time spent in various areas of the matrix but not in PSY and constitutes the second largest phenotypic class cluster. When G2 animals high in percent time spent in corners were intercrossed, the next generation of animals (G3) also displayed significantly higher time in corners and low total distance traveled (Fig. 5 *A* and *B*, respectively). This line, which displays low exploratory behavior with high levels of thigmotaxis and low ambulation, seems to be a high-anxiety line.

ENU052 was chosen for heritability testing because it belonged to cluster VI, which has phenotype specific for cocaine response only. This cluster was not affected in any of the OFA traits. When G2 animals from this family with high response to cocaine were intercrossed to produce G3 progeny, the next generation had a significantly higher response to cocaine, like the G2 generation (Fig. 5*C*). In both G2 and G3 the response to cocaine is ≈1 SD higher than in the control group. Because this line has normal baseline activity and OFA phenotypes, we would argue that it does not have anxiety or exploratory behavior abnormalities but is affected in only cocaine response.

We also chose animals from cluster III, which is specific for OFA distance measurements but not in OFA time or PSY measurements. This cluster can be considered as mutants that are normal for anxiety-like behaviors but have high or low exploratory behavior as measured by ambulation. ENU144, which belongs to this cluster, is significantly higher than the control group in the OFA total distance traveled measure in two consecutive generations (Fig. 5*D*). The total distance traveled during the OFA is 1 SD higher than in the control group in both G2 and G3 for ENU144. Thus, this line has a mutation that specifically affects components of PC1, exploratory behavior. ENU189 was chosen from cluster II, which constitutes the largest phenotypic class of mutants. This class has significant differences in OFA distance and time measurements but is phenotypically normal for psychostimulant measures. Animals from ENU189 G2 and G3 had higher time spent in periphery as well as decreased total distance traveled (Fig. 5 *E* and *F*, respectively). This line is similar to ENU042 in having low exploratory activity and high anxiety-like behavior.

Overall in our screen we placed progeny from 27 G2 families for heritability testing and confirmed inheritance in 15 lines. Putative mutants from all six phenotypic clusters were placed into mating. Data for heritable lines not shown in Fig. 5 are shown in Fig. S6*A*, and data for lines that were not heritable are shown in Fig. S6*B*. We were able to discover at least one heritable mutant from each cluster, even when there were very few individuals in the cluster. For instance, heritable lines ENU059 and ENU039 from cluster I had high percent time in corners compared with control animals in both G2 and G3 (Fig. S6*A*). This was the smallest cluster, with only seven mutant lines. We discovered two lines with high response to cocaine, ENU052 (Fig. 5*D*) and ENU086 (Fig. S6*A*), and one line with low response to cocaine, ENU100 (Fig. S6*A*). Lines that were
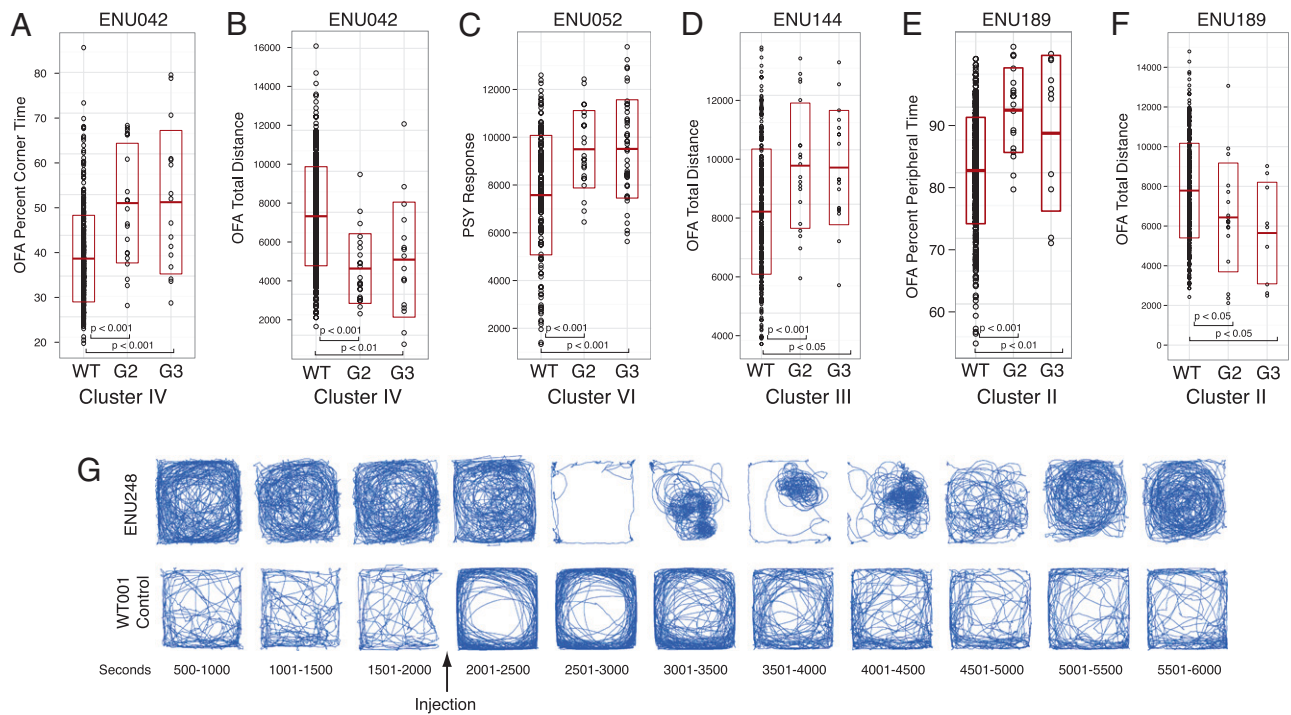
Fig. 5. Heritability testing. Representative lines from several mutant classes identified in Fig. 3 were bred for one more generation to test for heritability. Test data from individual animals are plotted in the scatterplot. Red bar represents mean ± 1 SD of each family. G2 is the screening population and G3 are progeny generated by intercrossing G2 animals. ENU042 belongs to cluster IV and has significant differences in OFA percent time in corners and OFA total distance traveled (*A* and *B*, respectively). ENU052 belongs to cluster VI, which has a phenotype specifically in response to cocaine (*C*). ENU144 (*D*) and ENU189 (*E* and *F*) belong to cluster II and have a phenotype in several OFA measures, two of which are shown here. ENU248 is a line that displays stereotypy followed by high-frequency circling after injection of cocaine. Each box represents 500 s of tracking data for wild-type control line (*G, Lower*) and an ENU248 progeny (*G, Upper*). The animals are injected after 2,000 s, and circling is seen between 3,000 and 4,000 s. Velocity data for this ENU248 animal are provided in Fig. S7, and the videos are available as supplemental data (Movie S1 and Movie S2).

not heritable represent five of the six mutant clusters (Fig. S6*B*). These nonheritable lines were significantly different from control animals in G2 but not significantly different in G3. There does not seem to be a trend as to which lines proved heritable and which lines were not heritable. All heritable (bold, italics text) and nonheritable (plain text) lines have been marked in Fig. 4. Two heritable lines, WT248 and WT054, are not marked in Fig. 4 because WT248 has a circling phenotype and WT054 was not significant at a global level when all 230 lines were used for multiple comparisons.

Another line that we isolated with an interesting phenotype is ENU248. This line originally attracted our attention because animals showed a sudden drop in activity after cocaine injection. However, further examination of the video showed that animals were going into stereotypy followed by high-frequency circling after cocaine injection. When a female displaying this behavior was placed into mating to produce G3, we noticed animals with the same behavior in the successive generation. Fig. 5*G* shows tracking data from a G3 ENU248 animal showing high-frequency circling com-

pared with a control female. Circling can be seen at 3,000–5,000 s of the test (Fig. 5*G* and Movies S1 and S2). This female had extremely high baseline activity and after injection went into rapid circling behavior. The velocity data are provided in Fig. S7. Although this was not a phenotype we were screening for, this type of behavior is seen in 6-hydroxydopamine lesioned rats (38) and has been used as a quantitative measure of sensitization to psychostimulants (39). Thus, this line may display extreme sensitivity to cocaine, which manifests as rotational behavior instead of increase in general locomotor activity.

Discussion

We describe here the results of a forward genetic screen carried out in mice to discover behavioral mutants that deviate by only 1 to 2 SDs from wild type (*z* score 1 to 2). Because forward genetic screens have traditionally searched for qualitative mutants that deviate more than 3 SDs, these mutants would be considered subtle quantitative mutants. However, an examination of the

Table 1. Description of significant clusters

Cluster	Phenotype	Phenotype details
I	Open field	Percent time and distance in corner. This cluster consists of a subset of traits PC1 and is specific to distance traveled and time spent in corners of matrix.
II	Open field	All percent time, percent distance, and total distance traveled. PC1 and PC3 traits. These mutants are significantly different in how much they travel and where they travel.
III	Open field	OFA distance traveled and crossing. PC3 traits, specific for distance traveled in OFA only.
IV	Open field	Subset of distance traveled, percent time, and percent distance. PC1 and PC3 traits.
V	Open field and psychostimulant response	OFA distance traveled and PSY. PC2 and PC3 traits. These mutants are significantly different in OFA activity as well as PSY response.
VI	Psychostimulant response	PSY traits only (PC2). These are normal for open field measures but are significantly affected in PSY response.

literature in which reverse genetic approaches have been used to study the loss of function of genes central in pathways regulating behaviors such as psychostimulant response, fear conditioning, and open field response have shown that 1 to 2 SD mutations are typical (11, 15). Although "emotionality" in rodents has been extensively studied as a model of anxiety disorder in humans, forward genetic screens in mice for behavior have not been as successful as in other phenotypic domains. There have been numerous large-scale screens carried out for neurobehavioral mutants; however, only a handful of mutants have been isolated using this approach (14, 25, 27, 40, 41), probably owing to the imprecise nature of many behavioral assays. We have also found that mutants isolated using a high threshold of detection have a high rate of false positives. For instance, in the National Institutes of Health neurogenomics screen conducted at Northwestern University, only 17% of the putative mutants found in the initial screen passed heritability testing (27). In another G1 dominant screen, Wada et al. (41) found that only 23% of putative mutants passed their criteria for heritability test. In the G2 screen conducted here, we confirmed inheritance in 56% of the lines we placed into breeding to produce G3. Our criterion for inheritance is strict in that the G3 animals must have a mean that is statistically different from control animals. Thus, the G2 screen is much more efficient in finding heritable mutants than G1. The G2-based screening described here is more labor intensive and does not allow the screening of as many mutagenized genomes as a G1 screen; however, the results are much more reliable, with much fewer lines that fail heritability tests.

The key to conducting a screen for subtle mutants is the use of multiple individuals arising from a single G1. Testing a family of G2 animals from each G1 allows multiple sampling of each mutant genome. In this screen we tested more than 2,300 G2 animals from 229 mutant G1s and extracted 16 different behavioral scores from the OFA and PSY of these animals. PCA places these behaviors into three groups: psychostimulant response, distance traveled during OFA, and time spent and distance traveled in various regions during OFA. We compared the behavioral scores from the measured and calculated phenotypes for significant differences from control animals. Using Dunnett's test to control for multiple testing, families that were significantly different were identified. Several families were significantly different in multiple phenotypes, and upon hierarchical clustering based on P values we discovered six classes of mutants. These classes are combinations of the three independent factors that were discovered in PCA. We bred several of these mutant lines for another generation and confirmed that the phenotypes were indeed heritable. PCA and factor analysis have been extensively used to study rodent behavior and have also been applied to QTL analysis for behavioral measures (30, 42–44). Use of multivariate analysis has led to the discovery of unique QTLs regulating latent behaviors that cannot be directly measured, such as fearfulness and anxiety (45, 46). Hierarchical clustering has been applied to chemical genetics screen for sleep/wake behavior in larval zebrafish, yielding information about drugs affecting shared behaviors (47). The combination of these statistical techniques allowed us to screen for subtle mutants.

In this G2 screen of 229 lines 10% of the lines had a significant difference at α of 0.01 and 18% at α of 0.05. It is interesting to note that in our screens we obtained many more mutants in OFA than in PSY, perhaps reflecting the genetic architecture of these phenotypes. Our heritability tests suggests a false-positive rate of 44%, which is much better than the false-positive rates of 77% and 83% reported in G1- and G3-based screens (27, 41). If half these lines are heritable, then at an α of 0.05, 9% of the initial putative mutants are true hits. A 10% mutant discovery rate is similar to what has been observed for reverse genetic screens in mice for behavioral phenotypes and for forward genetic screens in Drosophila for behavior. Silva and colleagues (15) screened 54 random knockout, transgenic, and point mutant mouse lines for learning and memory tasks. They discovered that 15 of 54 lines (28%) had significant differences in some aspect of learning task, such as remote memory, freezing, or short-term memory. In another screen of 206 knockout mice from Lexicon and Deltagen (48), 19% of the lines had abnormal pheno-

type in OFA. Similarly, in Drosophila screens for learning and memory mutants by Boyton and Tully (49), using P-element insertions yielded 170 positives from 1,016 lines screened (17%). In this screen only 23% of initial hits held up in heritability testing—a rate much lower than ours. Thus, in three screens that are similar to ours the mutant discovery rate is between 17% and 28%, within range of what we are observing.

Another consideration given the 10% mutant discovery rate is the probability of having multiple mutations causing the abnormal behavior in the same line. The initial estimate of mutation rates determined by Russell and colleagues (50) using specific locus mouse tester strain was 1/700. Using traditional sequencing as well as next-generation sequencing efforts, the mutation rate has been estimated at ≈1.4 per million base pair for C57BL/6J at the dose of ENU used in this project (51–53). Because 1.5% of the 2.5-Gb mouse genome is coding and ≈70% of coding mutations would be nonsynonymous (51), we expect 37 nonsynonymous mutations per haploid genome [$(2.5 \times 10^{9}) \times (1.4 \times 10^{-6}) \times 0.015 \times 0.7$]. Experimentally derived mutation rates have estimated 45 nonsynonymous coding mutations per G1 genome (52), close to our estimates. If a G1s carries 37 mutations, then every G2 would carry ≈19 nonsynonymous mutations on average. If we assume 22,000 genes in the mouse genome, then 19 of 22,000 or 1 of every 1,158 genes should contain a mutation. If 10% of our screened progeny are true mutants, then we can calculate the number of genes that cause the phenotype of interest as 116 (10% of 1,158). Given our rough calculations, the probability of having 2 of these 116 genes mutated simultaneously is $1/116^{2}$, extremely unlikely. It is possible that subtle phenotypes identified in this screen are due to noncoding mutations that are in promoters, enhancers, or noncoding RNA, in which case our assumption that only mutations in 1.5% of the genome is meaningful would have to be revised. Because we propose to map these mutants using QTL approaches, two interacting loci that contribute to the phenotype can be resolved and would not pose a technical challenge to genetic mapping.

The ultimate challenge in discovery of subtle mutants is the ability to clone these mutations. Certainly classic linkage approaches that are used for Mendelian phenotypes will not be adequate for such quantitative mutants. QTL approaches have been successful in mapping and cloning ENU-induced mutations (12–14, 54, 55). In several cases cloning has been facilitated by lethality exhibited by the homozygous mutant, turning a quantitative behavioral trait into a qualitative viability trait that can be used for genetic mapping (12, 13). We believe a combination of low-resolution mapping using QTL analysis with the recent advances in next-generation sequencing technology will provide a feasible route to cloning of these mutants (11). Although 1 to 2 SDs might be subtle for ENU screens, a single point mutation affecting a trait by 1 SD is considered quite strong for a QTL. Use of new sequencing technologies has already been applied to the cloning of mutants in Drosophila, Caenorhabditis elegans, Arabidopsis, and yeast (56–60). Recently Bruce Beutler and colleagues (61) used bulk segregation mapping along with next-generation sequencing to identify ENU mutation with coat color and circling phenotypes in mice. In addition, this sequencing technology has made feasible the use of closely related mouse substrains such as C57BL/10 as well as C57BL/6 substrains as a mapping partner. Use of closely related strains allows mapping at low resolution but limits the number of genetic modifiers that might exist for subtle behavioral phenotypes such as those seen in this screen. We predict that recent technological progress will greatly facilitate the cloning of the mutants identified in this screen.

Materials and Methods

Animal Husbandry and Testing. Mouse strain C57BL/6J (stock no. 000664) were obtained from Jackson Laboratory. All mice were housed in barrier Specific Pathogen Free (SPF) conditions with Light:Dark (LD 12:12) according to University of Texas Southwestern Medical Center Institutional Animal Care and Use Committee guidelines. Animals (10–14 wk of age) were allowed to acclimate to the testing room for half an hour before behavioral testing. Mice were weighed and visually inspected for low body weight or any other developmental problems, such as malocclusion. Any mouse deemed to be ab-

normal was removed from the screen. Mice were placed in a 55 cm × 55 cm × 36 cm (width × length × height) matrix (Phenome Technologies) for OFA for 30 min, followed by i.p. injection with 20 mg/kg cocaine (Sigma C5776) solution in 0.9% saline. Fluorescent light levels were measured at 700 lx ± 56 lx (mean ± SD of all 32 chambers in our testing facility). Ultrafine insulin syringes were used for injection of cocaine (3/10 mL BD Lo-Dose, catalog no. 328438). After each test the test chamber was cleaned with dilute Quatricide solution (Pharmacal Research Labs). Testing was carried out between Zeitgeber Time (ZT) ZT4 and ZT10. For behavioral recording and analysis we used Actimetrics LimeLight software.

ENU Mutagenesis. C57BL/6J mice (6–8 wk old) were injected with ENU (Sigma, catalog no. N3385) as previously described (62). After recovery from

sterility, these G0 mice were mated according to the breeding scheme described in Fig. 1A.

Statistical Analysis. All analyses were carried out using R (33). For multiple comparison the multcomp package was used (31). For PCA the SciViews package was used, and for plots the ggplot2 package was used (63). Clustering was carried out using the MeV software package (64).

1. Morgan TH (1911) The origin of five mutations in eye color in *Drosophila* and their modes of inheritance. *Science* 33:534–537.
2. Morgan TH (1911) The origin of nine wing mutations in *Drosophila*. *Science* 33:496–499.
3. Nüsslein-Volhard C, Wieschaus E (1980) Mutations affecting segment number and polarity in *Drosophila*. *Nature* 287:795–801.
4. Brenner S (1974) The genetics of *Caenorhabditis elegans*. *Genetics* 77:71–94.
5. Vitaterna MH, et al. (1994) Mutagenesis and mapping of a mouse gene, Clock, essential for circadian behavior. *Science* 264:719–725.
6. Takahashi JS, Pinto LH, Vitaterna MH (1994) Forward and reverse genetic approaches to behavior in the mouse. *Science* 264:1724–1733.
7. Nurse P, Thuriaux P, Nasmyth K (1976) Genetic control of the cell division cycle in the fission yeast *Schizosaccharomyces pombe*. *Mol Gen Genet* 146:167–178.
8. Benzer S (1967) Behavioral mutants of *Drosophila* isolated by countercurrent distribution. *Proc Natl Acad Sci USA* 58:1112–1119.
9. Hirsch J (1962) Individual differences in behavior and their genetic basis. *Roots of Behavior: Genetics, Instinct, and Socialization in Animal Behavior*, ed Bliss EL (Hafner, New York).
10. Benzer S (1971) From the gene to behavior. *JAMA* 218:1015–1022.
11. Takahashi JS, Shimomura K, Kumar V (2008) Searching for genes underlying behavior: Lessons from circadian rhythms. *Science* 322:909–912.
12. Keays DA, et al. (2007) Mutations in alpha-tubulin cause abnormal neuronal migration in mice and lissencephaly in humans. *Cell* 128:45–57.
13. Speca DJ, et al. (2010) Conserved role of unc-79 in ethanol responses in lightweight mutant mice. *PLoS Genet* 6:e1001057.
14. Reijmers LG, et al. (2006) A mutant mouse with a highly specific contextual fear-conditioning deficit found in an N-ethyl-N-nitrosourea (ENU) mutagenesis screen. *Learn Mem* 13:143–149.
15. Matynia A, et al. (2008) A high through-put reverse genetic screen identifies two genes involved in remote memory in mice. *PLoS ONE* 3:e2121.
16. Hall CS (1934) Emotional behavior in the rat. I. Defecation and urination as measures of individual differences in emotionality. *J Comp Psychol* 18:385–403.
17. Hall CS (1938) The inheritance of emotionality. *Sigma Xi Q* 26:17–27.
18. Hall CS (1941) Temperament: A survey of animal studies. *Psychol Bull* 38:909–943.
19. Laakso A, Mohn AR, Gainetdinov RR, Caron MG (2002) Experimental genetic approaches to addiction. *Neuron* 36:213–228.
20. Robinson TE, Berridge KC (2003) Addiction. *Annu Rev Psychol* 54:25–53.
21. Goldman D, Oroszi G, Ducci F (2005) The genetics of addictions: Uncovering the genes. *Nat Rev Genet* 6:521–532.
22. Plomin R, McClearn GE, Gora-Maslak G, Neiderhiser JM (1991) Use of recombinant inbred strains to detect quantitative trait loci associated with behavior. *Behav Genet* 21:99–116.
23. Crabbe JC, Phillips TJ, Buck KJ, Cunningham CL, Belknap JK (1999) Identifying genes for alcohol and drug sensitivity: Recent progress and future directions. *Trends Neurosci* 22:173–179.
24. Phillips TJ (1997) Behavior genetics of drug sensitization. *Crit Rev Neurobiol* 11:21–33.
25. Gondo Y (2008) Trends in large-scale mouse mutagenesis: From genetics to functional genomics. *Nat Rev Genet* 9:803–810.
26. Acevedo-Arozena A, et al. (2008) ENU mutagenesis, a way forward to understand gene function. *Annu Rev Genomics Hum Genet* 9:49–69.
27. Vitaterna MH, Pinto LH, Takahashi JS (2006) Large-scale mutagenesis and phenotypic screens for the nervous system and behavior in mice. *Trends Neurosci* 29:233–240.
28. Friendly M (2002) Corrgrams: Exploratory displays for correlation matrices. *Am Stat* 56:316–324.
29. Gabriel KR (1971) Biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58:453.
30. Ramos A, Mormède P (1998) Stress and emotionality: A multidimensional and genetic approach. *Neurosci Biobehav Rev* 22:33–57.
31. Bretz F, Hothorn T, Westfall P (2011) *Multiple Comparisons Using R* (Chapman and Hall/CRC, Boca Raton).
32. Dunnett CW (1955) A multiple comparison procedure for comparing several treatments with a control. *J Am Stat Assoc* 50:1096–1121.
33. R Development Core Team (2011) R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna).
34. Naik UD (1975) Some selection-rules for comparing P-processes with a standard. *Commun Stat* 4:519–535.
35. Dunnett CW, Tamhane AC (1991) Step-down multiple tests for comparing treatments with a control in unbalanced one-way layouts. *Stat Med* 10:939–947.
36. Marcus R, Peritz E, Gabriel KR (1976) Closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 63:655–660.
37. Marcus R (1976) Powers of some tests of equality of normal means against an ordered alternative. *Biometrika* 63:177–183.
38. Ungerstedt U, Arbuthnott GW (1970) Quantitative recording of rotational behavior in rats after 6-hydroxy-dopamine lesions of the nigrostriatal dopamine system. *Brain Res* 24:485–493.
39. Badiani A, Morano MI, Akil H, Robinson TE (1995) Circulating adrenal hormones are not necessary for the development of sensitization to the psychomotor activating effects of amphetamine. *Brain Res* 673:13–24.
40. Bućan M, Abel T (2002) The mouse: Genetics meets behaviour. *Nat Rev Genet* 3:114–123.
41. Wada Y, et al. (2010) ENU mutagenesis screening for dominant behavioral mutations based on normal control data obtained in home-cage activity, open-field, and passive avoidance tests. *Exp Anim* 59:495–510.
42. Flint J (2003) Analysis of quantitative trait loci that influence animal behavior. *J Neurobiol* 54:46–77.
43. Korol AB, Ronin YI, Itskovich AM, Peng J, Nevo E (2001) Enhanced efficiency of quantitative trait loci mapping analysis based on multivariate complexes of quantitative traits. *Genetics* 157:1789–1803.
44. Takahashi A, Kato K, Makino J, Shiroishi T, Koide T (2006) Multivariate analysis of temporal descriptions of open-field behavior in wild-derived mouse strains. *Behav Genet* 36:763–774.
45. Turri MG, DeFries JC, Henderson ND, Flint J (2004) Multivariate analysis of quantitative trait loci influencing variation in anxiety-related behavior in laboratory mice. *Mamm Genome* 15:69–76.
46. Henderson ND, Turri MG, DeFries JC, Flint J (2004) QTL analysis of multiple behavioral measures of anxiety in mice. *Behav Genet* 34:267–293.
47. Rihel J, et al. (2010) Zebrafish behavioral profiling links drugs to biological targets and rest/wake regulation. *Science* 327:348–351.
48. Flint J, Mott R (2008) Applying mouse complex-trait resources to behavioural genetics. *Nature* 456:724–727.
49. Boynton S, Tully T (1992) latheo, a new gene involved in associative learning and memory in *Drosophila melanogaster*, identified from P element mutagenesis. *Genetics* 131:655–672.
50. Hitotsumachi S, Carpenter DA, Russell WL (1985) Dose-repetition increases the mutagenic effectiveness of *N*-ethyl-*N*-nitrosourea in mouse spermatogonia. *Proc Natl Acad Sci USA* 82:6619–6621.
51. Takahasi KR, Sakuraba Y, Gondo Y (2007) Mutational pattern and frequency of induced nucleotide changes in mouse ENU mutagenesis. *BMC Mol Biol* 8:52.
52. Arnold CN, et al. (2011) Rapid identification of a disease allele in mouse through whole genome sequencing and bulk segregation analysis. *Genetics* 187:633–641.
53. Concepcion D, Seburn KL, Wen G, Frankel WN, Hamilton BA (2004) Mutation rate and predicted phenotypic target sizes in ethylnitrosourea-treated mice. *Genetics* 168:953–959.
54. Speca DJ, Rabbee N, Chihara D, Speed TP, Peterson AS (2006) A genetic screen for behavioral mutations that perturb dopaminergic homeostasis in mice. *Genes Brain Behav* 5:19–28.
55. Furuse T, et al. (2010) Phenotypic characterization of a new Grin1 mutant mouse generated by ENU mutagenesis. *Eur J Neurosci* 31:1281–1291.
56. Blumenstiel JP, et al. (2009) Identification of EMS-induced mutations in *Drosophila melanogaster* by whole-genome sequencing. *Genetics* 182:25–32.
57. Schneeberger K, Weigel D (2011) Fast-forward genetics enabled by new sequencing technologies. *Trends Plant Sci* 16:282–288.
58. Doitsidou M, Poole RJ, Sarin S, Bigelow H, Hobert O (2010) *C. elegans* mutant identification with a one-step whole-genome-sequencing and SNP mapping strategy. *PLoS ONE* 5:e15435.
59. Birkeland SR, et al. (2010) Discovery of mutations in *Saccharomyces cerevisiae* by pooled linkage analysis and whole-genome sequencing. *Genetics* 186:1127–1137.
60. Laitinen RA, Schneeberger K, Jelly NS, Ossowski S, Weigel D (2010) Identification of a spontaneous frame shift mutation in a nonreference *Arabidopsis* accession using whole genome sequencing. *Plant Physiol* 153:652–654.
61. Xia Y, et al. (2010) Bulk segregation mapping of mutations in closely related strains of mice. *Genetics* 186:1139–1146.
62. Siepka SM, Takahashi JS (2005) Forward genetic screens to identify circadian rhythm mutants in mice. *Methods Enzymol* 393:219–229.
63. Wickham H (2009) *ggplot2: Elegant Graphics for Data Analysis* (Springer, London).
64. Saeed AI, et al. (2006) TM4 microarray software suite. *Methods Enzymol* 411:134–193.