MOUSE PHENOME PROJECT

Observer-rated ataxia: rating scales for assessment of genetic differences in ethanol-induced intoxication in mice

Pamela Metten, Karyn L. Best, Andy J. Cameron, Alisha B. Saultz, Jessica M. Zuraw, Chia-Hua Yu, Douglas Wahlsten, and John C. Crabbe.

Observer-rated ataxia: rating scales for assessment of genetic differences in ethanol-induced intoxication in mice. J Appl Physiol 97: 360–368, 2004. First published March 19, 2004; 10.1152/japplphysiol.00086.2004.—Identification of genetic and physiological mechanisms underlying a drug’s or mutation’s effects on motor performance could be aided by the existence of a simple observation-based rating scale of ataxia for mice. Rating scales were developed to assess ataxia after ethanol (2.75, 3.0, and 3.25 g/kg) in nine inbred mouse strains. Each scale independently rates a single behavior. Raters, blinded to dose, scored four behaviors (splay of hind legs, wobbling, nose down, and belly drag) at each of four time points after injection. The severities of hind leg splaying and wobbling were quantifiable, whereas nose down and belly dragging were expressed in all-or-none fashion. Interrater reliabilities were substantial (0.75 ≤ r ≤ 0.99). Splay scores (rated 0–5) displayed significant effects of strain, dose, and time point. Wobbling (rated 0–4) was dependent on strain and time point. Ethanol affected wobbling (most strains scored >0 at some time), but all doses were equally effective. Incidence of nose down and belly dragging behaviors increased strain dependently after ethanol, but strains did not differentially respond to dose. Ethanol-induced splaying was modestly, and negatively, genetically correlated with wobbling. Nose down and belly dragging tended to be associated with splaying and wobbling at later times. Four distinct ataxia-related behaviors were sensitive to ethanol. Strains differed in ethanol sensitivity for all measures. Modest strain mean correlations among behaviors indicate that these behaviors are probably under control of largely different genes and that ataxia rating scales should rate separate behaviors on discrete scales.

However, it is clear that caution should be employed before electronic data are accepted as meaningful (11). Other tests such as the balance beam (3, 11, 17), hindpaw footprint test (2, 7), and screen test (6, 10, 33) are relatively advantageous because expensive equipment is not required and data recording is done manually. Any measure of “ataxia” is likely to measure only some aspects of ataxia (e.g., loss of balance, gait irregularity, or loss of muscle strength) because of the complex nature of this phenomenon.

Intoxication rating scales have been used to determine doses of sedative hypnotics to administer to maintain intoxication over time and induce dependence in cats and rats (25, 28, 36). These scales rated gait or motor coordination impairment as components of ataxia. They have in common that ataxia was rated as a subjective assessment of “mild” or “slight” to “marked” or “clearly” impaired, allowing differences in interpretation among observers and/or laboratories. The scales also collapse different behaviors into the same scale point. For example, Majchrowicz and Hunt (26) employed a three-point ataxia scale for rats: 1 = slight gait impairment and slight motor incoordination but able to elevate abdomen and pelvis; 2 = clearly impaired staggering gait and impaired motor coordination, with some elevation of abdomen and pelvis; and 3 = slowed righting reflex, heavily impaired motor coordination, and no elevation of abdomen and pelvis. This approach presents two problems: first, the physiological mechanisms underlying a drug’s effects on gait impairment (staggering, for example) may be different from those underlying ability to elevate the abdomen during movement; and second, one may be faced with a subject that could be given two different scores at a single time point. Moreover, different drugs could produce the same score even though they differ dramatically in their capacity to affect each behavior taken separately.

Rating scales have also been employed for chemical-induced ataxia in animal models of diseases such as epilepsy, parkinsonism, and schizophrenia (22, 23, 35). These scales also combine different components of ataxia in the same scale. For example, Sturgeon et al. (35) developed a rating scale for phencyclidine effects on ataxia that encompassed coordination of movement, loss of balance, impairment of antigravity reflex, inability to move, and impairment of weight distribution.

Address for reprint requests and other correspondence: P. Metten, VA Medical Center, Research Service, R&D-D-12, 3710 SW US Veterans Hospital Road, Portland, OR 97239 (E-mail: mettenp@ohsu.edu).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
Furthermore, each individual scale point itself incorporated multiple behaviors.

Our laboratories have been interested in genetic and environmental influences on many mouse behavioral traits affected by drugs of abuse, including ethanol-induced ataxia (41). To assess behavior in different genotypes across laboratories reliably, optimized test conditions are necessary (12, 31, 41). Our strategy has been to assess different measures of ataxia systematically in genetically outbred mice to determine a range of optimal and reliable test conditions (32) and then to follow up these tests in panels of standard inbred mouse strains; we have found strains to differ in response to ethanol on several different ataxia-related behaviors (10, 11, 31). As part of the Mouse Phenome Project (29), a large project designed to collect data on up to 40 inbred strains of mice, we have also extended some of this work to examine interlaboratory reliability of differences among 20 inbred strains (31, 40, 42).

Crabbe et al. (8, 9) showed differences among 20 inbred mouse strains in ethanol-induced ataxia (30 and 60 min after 3 g/kg), observed while the animal was in its home cage, using a three-point scale from slight (rating of 1), to marked (rating of 2) locomotor incoordination, to loss of righting reflex (rating of 3). These studies found that strains differed in severity of ataxia (not surprisingly) and that 129P3/J (then designated 129/J) and C57BL/10N mice were rated most ataxic, whereas SWR/J and SJL/J mice were rated least ataxic. This rating scale also collapsed multiple behaviors into single scale points. Thus our present goal was to establish a new set of rating scales that also collapsed multiple behaviors into single scale points. Moreover, the 129P3/J and C57BL/10N mice were the two most ataxic, whereas SWR/J and SJL/J mice were the two least ataxic. This rating scale also collapsed multiple behaviors into single scale points. Thus our present goal was to establish a new set of rating scales that would separately describe easily observed abnormalities in body positions and movement. While designing these scales, we sought indexes sensitive to ethanol-induced impairment that would detect strain and dose differences reliably. We were successful for two signs, splaying of hind legs and wobbling, and were able to score presence or absence of two additional signs, nose down and belly dragging (Table 1).

MATERIALS AND METHODS

Animals. Male and female mice (approximately equal numbers of each sex across strain and dose groups, except where noted), 62–77 days of age at the time of testing were used. Animals from nine inbred strains [129P3/J (129P3; n = 12), 129S1/SvImJ (129S1; n = 24), AJ (n = 24), BALB/cByJ (BALB; n = 36), BTBR T1 cytokine congenic (B6; n = 24), C57BL/6J (B6; n = 36), C3H/HeJ (C3H; n = 24), DBA/2J (D2; n = 32), and FVB/NJ (FVB; n = 24); total N = 236] were purchased from The Jackson Laboratory (Bar Harbor, ME). Strains were chosen from the “A” list of the Mouse Phenome Project (http://aretha.jax.org/pub-cgi/phenome/mpdcgi?rtn=docs/prstrains) because they are widely used, fertile, relatively inexpensive, and offer diverse pedigrees (29). The exceptions were the BTBR strain, which once was on the A list, and the 129P3 strain; both were studied here because we have data on them for several other relevant tasks. Some data were also collected from the F1 hybrids of B6 and D2, but we did not receive the full complement for testing (data available on request). Three shipments of male and female animals were tested in this experiment. The first shipment comprised four strains (only female D2 mice were available for this shipment). All 129P3 mice came in this shipment. Two shipments each comprised eight strains. On arrival, mice were acclimated to our facility for ~3 wk before testing. All animals were housed in standard caging with Bed-o-Cob corn cob bedding in isosexual groups, three per cage with food (Purina 5001) and water provided ad libitum. Mice were maintained on a 12:12-h light-dark cycle with lights on at 6:00 AM. Both colony and procedure rooms were maintained at 22 ± 1°C. Principles of laboratory animal care (http://www.nap.edu/readingroom/books/labrats) were followed in conducting these experiments, and all procedures were approved by the Department of Veterans Affairs Medical Center Institutional Animal Care and Use Committee in accordance with the National Institutes of Health’s guidelines.

Experimental procedure. In pilot studies with genetically heterogeneous mice and early versions of the rating scales, alcohol doses <2.5 g/kg were ineffective in producing reliable ataxia, whereas doses >3.5 g/kg produced loss of righting reflex in many mice (data not shown). We next tested BALB, B6, D2, and 129P3 mice (4–6 per dose) after 2.75, 3.0, or 3.5 g/kg to determine the final test conditions. Because ataxia after 3.5 g/kg appeared to represent a ceiling effect for most of these strains, the final conditions employed used 8–13 mice per strain per dose (total of 224 mice) of eight strains × three doses of ethanol (2.75, 3.0, or 3.25 g/kg). Where possible, we also analyzed nine strains × two doses (2.75 or 3.0 g/kg). The two exceptions to cell sample size were at the 30-min time point after 3 g/kg, where trial length was inadvertently shortened for three FVB mice (leaving five mice in this cell), and at the 5-min time point after 3.25 g/kg, where shortened trials occurred for three B6 mice (leaving seven mice in this cell).
Products, 20% vol/vol, in 0.9% saline, at doses of 2.75, 3.0, 3.25, or 3.5 g/kg, and then placed into an individual holding cage. Ethanol solution was made fresh on the day of testing and administered at a volume of 17.5–22.25 ml/kg body wt. Mice within a cage were assigned pseudorandomly to different dose groups. Mice were injected by one technician at intervals of 75 s and rated by others, who were blinded to dose at the time of scoring. Because wobbling and splaying severity ratings were made after the fact from black and white videotape, raters were also largely blinded to strain, as there were four albino strains, two black, one dilute brown, and two agouti. Each mouse was placed in the enclosure 5, 10, 15, and 30 min after injection and allowed to move unhindered and rated as before (Table 1). After the experiment, the videotapes were evaluated and scored for greatest degree of ataxia observed for each mouse during each 45-s test. To assess inter-rater reliability of the splay and wobble rating scales, mice of each shipment were rated by one person (i.e., K. L. Best rated all mice) and also “guest rated” by a separate observer. Raters recorded their scores independently. Line crossings were also scored by dividing the video monitor view of the platform into nine equal square segments.

Data analyses. SYSTAT (Chicago, IL) version 10 was used for all statistical analyses. All reported analyses for wobbling and splaying are based on severity ratings by K. L. Best, whereas analyses for belly dragging and nose down and belly dragging behaviors are based on presence/absence evaluations at the time of testing by A. B. Saultz. The binary presence/absence data for nose down and belly dragging behaviors were analyzed by logistic regression, with strain, sex, and dose entered as categorical variables. Data were tested for significant differences using the $\chi^2$ likelihood ratio test. Splay and wobble data were analyzed by separate ANOVAs on dose, strain, and time, with differences interpreted as significant at $P < 0.05$. Because sex generally did not interact with other factors, the data reported are collapsed on sex. Significant differences among strains or doses or significant interactions were analyzed with one-way ANOVAs and post hoc tests with Tukey’s honestly significant difference (HSD) statistic. Splay and wobble rating data were also analyzed to determine whether a given strain, dose, or time point had a rating significantly greater than zero (unidirectional $t$-test vs. null hypothesis of mean = 0). Interrater reliability of rating scale data was assessed by correlation of the mean strain ratings by K. L. Best with a composite of three guest raters’ scores (each rated at least 4 strains and about 1/3 of the mice). Inbred strain data from one-way ANOVAs on strain were used to estimate strain effect sizes as $\omega^2$ (the proportion of individual differences attributable to genetic sources) for specific conditions (34). Correlations of strain means were used to estimate lower bounds of the contribution of common genetic determinants between pairs of variables (21). One of our goals for the analyses was to identify specific dose \times time conditions that provided a good summary measure of strain sensitivity.

RESULTS

No effects of sex were found, so data were collapsed on this factor for further analysis. Thus we next analyzed strain and dose at each time point. Significant main effects of strain and dose were identified at each time point (both main effects $P < 0.001$ except at 5 min when both showed $P < 0.05$), but the interaction terms were never significant (all $P > 0.14$). Dose effects were as expected (increasing splaying with increasing dose; Fig. 1). The lack of a significant interaction between strain and dose indicated that strains responded similarly to doses, although it is possible that there was insufficient power to detect an interaction with our sample sizes (37, 38). Very few cases were identified where mean splay scores were not significantly greater than zero. All such cases occurred at the 15- or 30-min time points and, except for the FVB strain, were always after 2.75 g/kg.

We then looked for strain differences at each dose and time point, allowing inclusion of the 129P3 strain in the analyses at the two lowest doses. Significant strain differences were identified at all time points after injection of 2.75 g/kg and at 10 min or later after injection of 3.0 g/kg (all $P < 0.05$). The only significant strain difference after 3.25 g/kg occurred at 30 min ($P = 0.001$). Although most strains showed severe splaying at 5 min that decreased over time, some strains (e.g., 129S1) showed marked effects of ethanol but little change over time. The BTBR strain showed little dose dependence of splaying at the earlier times but showed dose-dependent splaying at 30 min. In contrast, FVB mice differed dramatically by dose at the earlier times but not at 30 min. Tukey’s HSD post hoc test revealed patterns of strain differences. Generally, the results show that the 129 strains were significantly more sensitive than FVB and BALB, whereas the other strains were intermediate.

Correlation of mean strain splaying scores revealed that within-dose correlations were generally significant across adjacent time points (e.g., $0.67 \leq r \leq 0.96$, $P < 0.05$ for the 2.75 g/kg dose). Significant strain correlations across adjacent doses also were identified at single time points. At 5 min, assessment after 2.75 or 3.0 g/kg produced nearly the same strain sensitivity ($r = 0.85$, $P < 0.01$). At both 10 and 15 min after injection, 2.75 and 3.0 g/kg produced nearly identical strain sensitivity ($r \geq 0.96$, $P < 0.001$), but strain rank orders differed from the 5-min time point. At both 15 and 30 min after injection, 3.0 and 3.25 g/kg produced similar strain sensitivity ($r = 0.78$, $P < 0.05$), but strain rank orders differed between these two time points. Together, these interdose and intertime point correlations argue that strain rank splaying scores at 10 or 15 min after 2.75 or 3.0 g/kg differed from those at 15 or 30 min after 3.25 g/kg. Although strain differences were seen at 5 min after the lowest dose, higher doses appeared to produce a ceiling effect at 5 min. Interrater reliabilities for splay scores (Pearson’s $r$) were $0.88 \leq r \leq 0.996$ ($P < 0.01$). Genetic effect sizes, calculated as the proportion of variance attributable to strain (34), were strongest after 10 min following injection of 2.75 and 3.0 g/kg (0.13 $\leq \omega^2 \leq 0.27$). With the 3.25 g/kg dose, the genetic effect was marginal until 30 min, when $\omega^2 = 0.28$. For further analyses, we therefore selected the later time points to index strain sensitivity.

Wobbling scale. Preliminary analysis of wobbling showed significant effects of strain, time, and their interaction ($P < 0.001$, 0.001, and 0.05, respectively; see Fig. 2). A trend toward a significant main effect of sex was observed with females more impaired ($P = 0.06$), but no interactions of sex...
Fig. 1. Splay scores for all strains after ethanol injection. Strains are divided into 3 groups for clarity. 

A: 129P3, 129S1, and A/J mice.
B: BALB, BTBR, and C3H mice.
C: B6, D2, and FVB mice.

Means ± SE are shown. See text for mouse strain definitions.
with other factors were present. Interestingly, no effects of dose were significant ($P > 0.28$) probably because of the narrow dose range suited to strain difference analysis. Thus data were collapsed on ethanol dose and sex for further analysis (Fig. 2). Data from 129P3 mice were included. Because our intention was to determine which time points were most suited to identification of strain differences, we next looked for strain differences at each time point. Significant effects of strain were found at every time point ($P < 0.0001$). Tukey’s HSD post hoc test revealed that the C3H strain displayed greater wobble scores than other strains at most time points (Fig. 2). The FVB strain also was very sensitive, but the onset for FVB was delayed until the 10-min time point.

Ethanol induced significant wobbling in all strains at some time points, with each animal’s maximum wobbling score occurring most often at 5 min (data not shown). Scores ranged from 0 to 4 at every time point. Correlation of strain means between adjacent time points revealed that the 10- and 15-min time points were virtually identical ($r = 0.94$, $P < 0.001$), indicating that either of these times would give the same strain rank order of severity of wobbling scores. A significant correlation between 5 and 15 min was also found ($r = 0.67$, $P < 0.05$), and this correlation improved when the FVB strain was removed ($r = 0.87$, $P < 0.01$). Interrater reliabilities for wobbling scores were $0.75 \leq r \leq 0.98$ ($P < 0.05$). Genetic effect sizes ranged from $0.11 \leq \omega^2 \leq 0.21$ and were strongest at 10 and 30 min; therefore, we selected these time points for further analyses.

**Nose down, belly drag, and line crossings.** Data for these three behaviors are presented at selected times for the 3.0 g/kg dose only (Table 2). Logistic regression analyses of the nose down behavior showed significant effects of strain at every time point ($\chi^2 \geq 23.97$, degrees of freedom (df) = 8, $P < 0.01$), as well as a significant main effect of dose at 5 and 30 min ($\chi^2 \geq 14.54$, df = 2, $P < 0.01$). The proportions of animals displaying nose down behavior decreased with increasing dose at 5 min but increased with dose at 30 min (data not shown). Adding sex as a factor in the regression did not change the pattern of results; no effects of sex were significant.

Belly dragging was analyzed similarly. Significant main effects of strain were detected at every time point (all $\chi^2 \geq 27.52$, df = 8, $P \leq 0.001$), but only at 5 min was there a significant effect of dose ($\chi^2 = 13.51$, df = 2, $P < 0.01$). No interaction terms were significant. At 5 min, strains generally showed more belly dragging at low doses than at higher (A/J, BALB, D2, and FVB strains; data not shown).

Not surprisingly, significant strain differences were detected in line crossings before drug injection ($P < 0.001$; Table 2). After ethanol injection, repeated-measures ANOVA on line crossings revealed main effects of strain, dose, and time and a significant interaction of time and strain ($P < 0.01$). There was also a main effect of sex (females > males, $P < 0.05$), but no interactions were detected; therefore, data are presented collapsed on sex. Significant main effects of strain were detected at every time point (all $P < 0.05$), but the main effect of dose was detected only at 10, 15, and 30 min (all $P < 0.05$). No other effects were significant. 129 substrains differed in baseline activity but not after injection. Mice of the A/J strain were one of the most active strains by 30 min after injection, whereas BTBR mice became essentially inactive after the 5-min time point.

**Relationship of rating scale data with pharmacokinetics.** We correlated the strain mean rating scale data with strain mean blood ethanol concentrations (BECs) collected from eight of these strains (excluding 129P3) in other studies in our laboratory (10, 30). BECs collected from the lateral tail vein at 35 min after 2.5 g/kg ethanol were not significantly correlated.
with either splaying or wobbling scores ($-0.16 \leq r \leq +0.57$, $P \geq 0.14$). However, splaying 5 min after 2.75 and 3.0 g/kg and 30 min after 2.75 g/kg were significantly correlated with BEC at 5 min after the 3 g/kg injection ($r = 0.72, P < 0.05$; Ref. 30). Inspection of the scatterplots of these correlations suggested that the 129S1 strain was an extreme scorer on both the splay and BEC variables. Removal of this strain eliminated all significant correlations ($r \leq 0.39, P > 0.39$) except at the 5-min time point, where splay scores after 2.75 remained significantly correlated ($r = 0.82, P < 0.05$) and a trend remained after 3.0 g/kg ($r = 0.67, P = 0.10$). These results suggest that, soon after injection, BEC may play a role in the severity of splaying, but not wobbling, and that genetic differences in splaying and wobbling at later times are not related to pharmacokinetic differences. No significant correlations were found between BEC and nose down, belly dragging, or line crossings.

Therefore, when looking for relationships between behaviors, we examined correlations among splaying and wobbling at the 10- and 30-min time points after injection of 2.75, 3.0, and 3.25 g/kg and nose down and belly dragging at the 5- and 30-min time points after injection of 3.0 g/kg.

**Relationships among variables.** Results of correlational analyses for the 3.0 g/kg dose at selected times are presented in Table 3. Lower wobbling ratings were significantly correlated with increased splaying after 3.0 g/kg at the 10-min time point (Table 3, Fig. 3) and at 30 min after 3.25 g/kg ($r = -0.78, P < 0.05$). Trends were observed in the same direction at some other doses and times (not shown). As expected, increased splaying was generally significantly correlated with decreases in line crossings at most time points and doses ($-0.39 \leq r \leq -0.93$). Splaying at 30 min after 3 g/kg was significantly correlated with nose down behavior at 15 and 30 min after 2.75 g/kg and at 10 min after 3.25 g/kg ($r \geq 0.71, P < 0.05$) but not at other times or doses. At early times after 2.75 g/kg, increased splaying scores were generally correlated with less belly dragging (some $r = -0.69, P < 0.05$). In contrast, increased splaying at several doses and times was significantly correlated with more belly dragging at 30 min after 2.75 g/kg and 10 min after 3.25 g/kg ($0.69 \geq r \geq 0.89$). Wobbling was significantly correlated with nose down behavior after injection of 3 g/kg at 5 and 10 min ($r \geq 0.69, P < 0.05$) and belly dragging behavior and also with line crossings at 5 min ($r \geq 0.67, P < 0.05$) but not at other times. Finally, nose down behavior was significantly correlated with belly dragging at 5 min after all three doses ($r \geq 0.84, P < 0.01$), at 10 min after 3.0 and 3.25 g/kg ($r \geq 0.79, P \leq 0.01$), and at 15 min after 3.25 g/kg ($r = 0.77, P < 0.05$).

**DISCUSSION**

The splaying scale clearly showed strain and dose sensitivity, with the effective ethanol dose range being quite narrow (Fig. 1). Strains also differed significantly in wobbling after ethanol injection, but all doses were equally effective. The rather modest negative genetic correlation of these two traits indicates that they are likely mediated in part by only some common genes (Fig. 3). However, the wobbling ratings require that the mouse is walking, whereas scores of 3–5 on the splay scale are given to mice that cannot get up on their hind feet (and thus usually receive a wobble score of 0, unless they wobble before severe splaying). We interpret these data to indicate that wobbling is a more mild sign of intoxication than splaying at the doses employed. For example, the C3H strain had the most severe wobbling scores (Fig. 2) but only mild

---

**Table 3. Strain mean correlations among ataxia behaviors after 3.0 g/kg dose**

<table>
<thead>
<tr>
<th>Behavior</th>
<th>Time, min</th>
<th>Splay 10 min</th>
<th>Splay 30 min</th>
<th>Wobble 10 min</th>
<th>Wobble 30 min</th>
<th>Nose 5 min</th>
<th>Nose 30 min</th>
<th>Belly 5 min</th>
<th>Belly 30 min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Splay</td>
<td>30</td>
<td>0.85†</td>
<td>-0.47</td>
<td>0.43</td>
<td>-0.33</td>
<td>0.28</td>
<td>0.01</td>
<td>0.07</td>
<td>-0.24</td>
</tr>
<tr>
<td>Wobble</td>
<td>30</td>
<td>-0.69*</td>
<td>-0.13</td>
<td>0.36</td>
<td>0.27</td>
<td>0.07</td>
<td>0.04</td>
<td>0.11</td>
<td>0.04</td>
</tr>
<tr>
<td>Nose</td>
<td>5</td>
<td>-0.33</td>
<td>-0.18</td>
<td>0.51</td>
<td>0.27</td>
<td>0.12</td>
<td>0.39</td>
<td>0.31</td>
<td>0.35</td>
</tr>
<tr>
<td>Nose</td>
<td>30</td>
<td>0.42</td>
<td>0.60</td>
<td>0.28</td>
<td>0.01</td>
<td>0.07</td>
<td>0.04</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>Belly</td>
<td>5</td>
<td>-0.47</td>
<td>-0.31</td>
<td>0.36</td>
<td>0.12</td>
<td>0.07†</td>
<td>-0.24</td>
<td>0.11</td>
<td>-0.24</td>
</tr>
<tr>
<td>Belly</td>
<td>30</td>
<td>0.24</td>
<td>0.25</td>
<td>-0.06</td>
<td>0.15</td>
<td>0.08</td>
<td>0.63</td>
<td>0.25</td>
<td>-0.20</td>
</tr>
<tr>
<td>Line crossing</td>
<td>5</td>
<td>-0.35</td>
<td>-0.06</td>
<td>0.19</td>
<td>0.07</td>
<td>0.04</td>
<td>0.25</td>
<td>0.13</td>
<td>0.09</td>
</tr>
<tr>
<td>Line crossing</td>
<td>10</td>
<td>-0.93†</td>
<td>-0.84</td>
<td>0.56</td>
<td>0.29</td>
<td>0.27</td>
<td>-0.60</td>
<td>0.44</td>
<td>-0.28</td>
</tr>
<tr>
<td>Line crossing</td>
<td>30</td>
<td>-0.68*</td>
<td>-0.84</td>
<td>0.31</td>
<td>-0.11</td>
<td>0.17</td>
<td>-0.57</td>
<td>0.26</td>
<td>-0.11</td>
</tr>
</tbody>
</table>

Values for wobble only were collapsed on doses 2.75–3.25 g/kg (see RESULTS). *Significant at $P < 0.05$; †significant at $P < 0.01$. 

---

**Fig. 3.** Genetic correlation scatterplot of ataxia rating scale responses at 10 min. x-Axis: strain mean splay score after 3 g/kg ethanol. y-Axis: strain mean wobble score [collapsed on dose (2.75–3.25 g/kg), see RESULTS]. Strains showing more severe splay scores show milder wobble scores ($P < 0.05$).
splaying (Fig. 1B). Conversely, the 129 substrains had profound splaying scores (Fig. 1A) but had mostly scores of 0 for wobbling because they could not walk. The 129S1 strain’s splaying persisted into the 30-min time point at all doses tested, and this strain’s score after 2.75 g/kg exceeded scores of all other strains (all \( P < 0.05 \)), indicating that this strain is remarkably sensitive to this measure. The 2.75 g/kg dose differentiated the two 129 substrains, in that 129P3 had completely recovered by 30 min after this dose (Fig. 1A).

Although the nose down and belly dragging behaviors were both significantly correlated with splaying and wobbling at 5 min and after some doses, these correlations were opposite in sign (wobbling was associated with more nose down and belly dragging, whereas splaying tended to be associated with less). However, this finding is also likely because many strains were immobile at the 5-min time point and thus were not able to show belly dragging while walking. It would be interesting to quantify belly dragging and nose down behaviors. For example, it should be possible to quantify the amount of the belly being dragged by measuring the size of the spot on a fluorescent pathway such as that employed in gait analysis (19, 20).

Comparison of strains with such a measure would be useful and should be sensitive to lower doses than those employed to detect splaying of the hindlimbs.

We have found the C3H strain to be relatively sensitive to ethanol-induced ataxia at a wide range of doses for several other behaviors. For example, C3H mice were so sensitive to low doses of ethanol (between 1 and 1.4 g/kg) on the balance beam that they straddled the narrowest beams and simply dragged one or more hindlimbs rather than walk along the beam and make missteps (11). At 2.5 g/kg, a dose more similar to that employed here, the C3H strain was also among the most sensitive of those tested on the grid test of ataxia, which measures foot slips through a grid floor relative to distance traveled (11). It is interesting that this sensitivity is similar to that seen here for wobbling rather than splaying of hind legs, which might be perceived from the experimenter’s point of view as the more similar behavior. We tested another set of mice after 2 g/kg. Although wobbling and splaying scores decreased overall, C3H mice still scored the highest wobble ratings of all the strains (data not shown). These results support the conclusion that wobbling and splaying are different behaviors and under the control of largely different mechanisms. On the low ends of both scales, mice sometimes splayed and wobbled during the same moment. C3H mice were also among the most sensitive to nose down and belly dragging behaviors after ethanol (Table 2).

The FVB strain was interesting for several reasons. At the 3.25 g/kg dose, this strain was maximally sensitive to ethanol-induced splaying but, by 30 min, had almost completely recovered (Fig. 1C, third panel), suggesting that substantial acute functional tolerance (AFT) had developed (14). Moreover, by 10 min, the FVB strain had recovered sufficiently to begin to show severe wobbling scores (Fig. 2). AFT to loss of righting reflex after 3.0 g/kg has been assessed in eight of these inbred strains, including FVB. In that study, which compared BEC at recovery with that at loss of righting reflex, FVB ranked only fourth in AFT: the A/J, C3H, and D2 strains had greater changes in BEC (30). Mean strain correlations between AFT to the loss of righting reflex and splay scores at any time point were not significant. Thus, if splay scores at 30 min generally reflect AFT to this effect of ethanol, different mechanisms must underlie AFT to splaying and to loss of righting reflex.

We looked to see whether the present rating scales correlated genetically with the older data set (8, 9). We used the present data for the 30-min time point after 3.0 g/kg, since these experimental conditions were the same as one time point reported for the older study. Although only one strain (129P3/J) was identical and five strains (A/HeN, BALB/cAnN, C3H/HeN, C57BL/6N, DBA/2N in the older study) were from substrains closely related to the Jackson Laboratory substrains used in the present study, splaying of the hind legs tended to correlate with severity of ataxia on the old scale (\( r = 0.68, df = 4, P < 0.14 \)). This correlation approached significance even though the strains presently employed were distinct substrains from those used earlier (e.g., C57BL/6J vs. C57BL/6N). No other behaviors from the present study approached significant correlation with the older data (all \( r \leq 0.21 \)).

The data presented here for alcohol suggest a dose range and time course for other studies looking for genetic differences in ataxia-related behaviors. It is likely that ataxic effects of other drugs may also be rated with these scales. Many of the behaviors in the present study have been included in descriptions of ataxia rating scales developed or used for other drugs, such as phencyclidine (35), phenytoin and carbamazepine (1), benzodiazepines and β-carbolines (23), and barbiturates (1, 5).

With the use of B6 mice, it has been shown that 3-nitropropionic acid intoxication (15) produces motor symptoms attributed to striatal system damage. Symptoms rated included hindlimb dystonia (space between hindlimbs, hindlimb coordination, crouching posture, gait impairment, all included in a 2-point scale) and truncal dystonia (hunching, rated on a 2-point scale). From the picture provided (Figure 4A in Ref. 15) the hindlimb dystonia appears to be very similar to our splayed hind leg behavior.

These rating scales might also be useful for rating mutants found to display cerebellar ataxias (see Ref. 18 for review). For example, splaying was observed in shaker short-tail mutants (dr+/−), a defect affecting cerebellar development (24, 39). The cerebellum and Purkinje cells are known to be vulnerable to ethanol abuse, and the motor effects of ethanol are mediated at least in part by the cerebellum (see Refs. 18 and 27 for reviews). Some of the same apparatuses have been employed to detect both ataxia mutants and ethanol-induced ataxia [e.g., balance beam (Refs. 3, 17)].

As we developed the scales for splayed hind legs and wobble, we encountered three primary cases of discrepancy between raters that resulted in scale revisions. The first two led to revisions to the splaying scale. One involved clarifying the descriptions between what are now the second and third points of the scale. A second revision was the addition of the fifth scalar point, to clarify whether a mouse was not moving because it could not (inability) or because it simply was not (apparent unwillingness). This issue was resolved by watching the mouse for muscle tone as it was being set down. If the mouse hung limply and subsequently did not move, the scale point assigned was 5. Otherwise, it was scored as 0. The third revision was made to the wobbling rating scale and also involved addition of a point to the scale. In this case, a 4 was
assigned to an animal that fell to its side and did not right itself. After the scale revisions were made, all relevant animals were rescored.

Interrater reliabilities of the wobbling and splaying rating scale data were significant, and the availability on CD of video clips of the different behaviors should assist training of other raters. These scales offer simple ratings that could be performed in addition to other behavioral tasks administered after the same dose of a given drug. The lack of strong genetic relationships among ataxia measures across tasks in this and other studies argues that researchers will likely need to assess multiple tests and conditions to establish a thorough interpretation of genetic differences (10, 11).

ACKNOWLEDGMENTS

The authors thank Jason P. Schlumbohm, Christina J. Cotnam, and Andrew Forster for assistance with these experiments. Dr. John K. Belknap kindly provided thoughtful and insightful comments on an earlier draft of the manuscript. We thank Dr. Nathan R. Rustay for helpful discussions on dosing and Drs. Justin S. Rhodes and Kim Cronise for statistical advice. We are grateful to Hazen Jette, Nathan Weiss, and Steve Gossen of the Department of Veterans Affairs, Dr. Mark Rutledge-Gorman, and Michelle Bobo for assistance with video editing, digital imaging, and formatting of the CDs of the behaviors on video.

The experiments comply with the current laws of the United States of America.

GRANTS

This study was supported by National Institute on Alcohol Abuse and Alcoholism Grants AA-10760, AA-12714, and AA-13519 and the Department of Veterans Affairs.

REFERENCES


