Influences of laboratory environment on behavior

TO THE EDITOR—Advances in genomics have led to much excitement about the potential to identify genes influencing behavioral traits. Given that the proportion of variation due to genotype (heritability) of most behavioral traits is less than 50% (ref. 1), an exclusive focus on genetic determinants will not fully explain individual differences. Mouse genotype interacts with the laboratory environment[2,3], and systematic variation and/or standardization of laboratory conditions has been advocated[4]. Generalization of results requires the assumption that they are not particular to a restricted set of standard laboratory conditions, but poor regulation of important variables can prevent replication. The major challenge therefore lies in determining which factors need to be regulated. Many sources of laboratory-related variability remain unidentified, and the relative impact of known factors is unclear.

We used a computational approach to retrospectively identify and rank sources of variability in pain responses on a common assay of thermal nociception, the 49°C hot water tail-flick/withdrawal test. Data were collected in the normal course of our ongoing study of the genetic mediation of pain and analgesia[5] (**Fig. 1**). The results are consistent with modeling in a subset of the data, and we were able to confirm the results in independent experiments that will be published elsewhere, which also partitioned the variance among genetic, environmental and genetic × environmental interactions.

Mice of varied genotypes were tested using a consistent procedure through natural fluctuations in the laboratory environment. The archival data set analyzed consists of baseline tail-withdrawal latencies for 8,034 naive adult mice, along with the following information (where available) recorded at testing: genotype (strain, substrain and vendor, among 40 inbred, outbred, hybrid and mutant strains), sex, age, weight, testing facility, cage density, season, time of day, experimenter, within-cage order of testing, and animal colony conditions including temperature and humidity. Eight factors were amenable to analysis (**Table 1**).

Despite profound differences in the mean responses of tested strains, the broad-sense heritability estimate (intraclass correlation estimated from variance components from inbred strains only) was 24%, so genotype alone did not account for most of the variance. Hypothesis-driven assessment of effects for individual factors using typical inferential statistics would be biased by the presence of the other factors, which also prohibits simultaneous ranking of their relative impact by common analytic methods. Thus, we used a technique suitable for unbalanced data sets of high dimensionality: classification and regression tree (CART) analysis[6], an automated machine-learning technique. CART is often used in medical applications to develop decision trees for diagnostic classification. Its value as a non-parametric tool for association of a large pool of predictors with a continuous variable has been largely untapped in neuroscience.

In brief, the CART technique develops rules to partition data based on predicting factors, producing a decision tree that can be used to predict the value of tail-withdrawal latency from the modeled factors. CART exhaustively tests all possible splits by each predictor to identify the split that gives the most improvement, defined as the difference between variance in the parent node and mean variance in the resulting two child nodes. The search is performed on each successive node until the data are split completely. The resulting tree is then pruned using a cross-validation technique to select an optimal tree. Split rules are printed in each parent node and indicate which levels of each factor go to the left and right child nodes.

The optimal tree selected by CART (**Fig. 2**) accounted for 42% of overall variance in tail-withdrawal latency (based on cross-validation) and had a resubstitution relative error of 49%, analogous to a multiple r^2 of 51%. These model-fit statistics may represent underestimates, because we took measures to reduce the bias toward selection of high-level categorical predictors and continuous predictors in the generation of tree-growing rules, including the conversion of continuous predictors to categorical ones and the penalization of factors by the number of levels they contain. This was done because we were not interested in the predictive value of the tree *per se*, but in the relative influence of the factors given equalized chances of consideration.

In agreement with previous findings[7], in every split by sex, female mice were more sensitive than males to thermal nociception. This finding confirms that the sex difference, although limited in magnitude, is robust across multiple testing contexts. In virtually every split by order of testing, the first mouse tested had a higher latency than all other mice. In addition, late-day testing, spring testing and higher humidity were usually associated with increased nociceptive sensitivity (lower latencies).

This technique also allows us to rank factors that are most important in reducing variance in many contexts. The

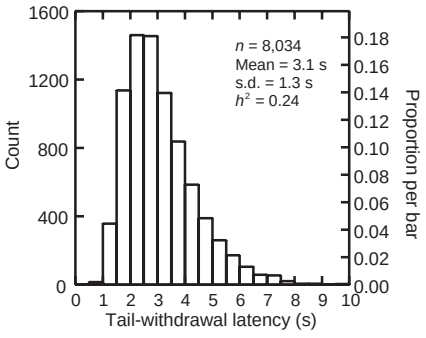

Fig. 1. Frequency histogram of responses on the 49°C tail-withdrawal assay of 8,034 mice tested from 1993 to 2001. Mice were individually removed from their home cage and introduced to a cloth/cardboard 'pocket', which they freely entered. Thus lightly restrained, the distal half of its tail was immersed in water thermostatically controlled at 49 ± 0.2°C. Latency to a vigorous, reflexive tail withdrawal was measured to the nearest 0.1 s with a handheld stopwatch.

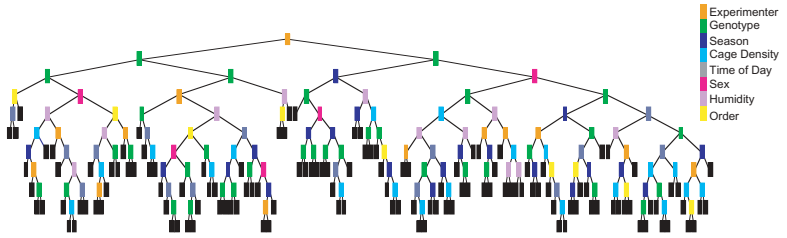

Fig. 2. Regression tree topology for sensitivity on the 49°C tail-withdrawal assay. The optimal tree shown here was selected after 10-fold cross validation. Each parent node is colored to indicate the primary splitter for that node. Variable importance rankings (**Table 1**) are calculated based on the variance reduction attributed to each factor used as a split criterion at each node and its surrogates. Black, terminal nodes. For full-size CART tree, with split rules and variance values, see **Supplementary Figure** online.

Table 1. Factor importance rankings computed by CART.

Factor[a]	Number of factor levels	Score[b]
Experimenter	11	100.0
Genotype	40	78.0
Season	4	35.8
Cage density	7	20.4
Time of day	3[c]	17.4
Sex	2	14.6
Humidity	4[d]	12.0
Order of testing	7	8.7

[a]Some factors (subject age, weight and ambient temperature) were not considered because insufficient biologically relevant within-factor variability existed in the data set. Preliminary models indicated that testing facility may influence the trait as well, but it was excluded from the final model because only one experimenter collected data in multiple facilities. [b]Scores are relative to the highest-ranked factor. [c]Time of day levels: early (09:30–10:55 h), mid-day (11:00–13:55 h), late (14:00–17:00 h). [d]Humidity levels: high (≥60%), medium-high (40-59%), medium-low (20-39%), low (<20%).

rankings are based on the relative variance reduction attributed to each factor when used as a primary splitter or as one of the top five surrogates (factors highly correlated with the splitter, whose importance it may mask) at each node, relative to an arbitrary score of 100 for the highest-ranked factor (**Table 1**). Experimenter identity had the greatest association with tail-withdrawal latency, outweighing genotype, the second-ranked factor. Also varying with behavior were environmental factors not commonly appreciated as affecting pain sensitivity, including season, cage density, time of day (within a 12 h diurnal period), humidity and testing order.

The importance of laboratory environment factors was demonstrated in a report of site-specific responses in behavioral testing[2]. Precisely what differentiated experimenters in the present study remains uncertain. The size and non-additive nature of the effect (by parametric analyses, data not shown) on baseline latencies eliminates the simple explanation of reaction-time differences. Experimenter age, sex and experience level did not correlate with the observed differences. All experimenters were trained by the principal investigator (J.S.M.) or by a graduate student (S.G.W.) trained directly by him. Differential animal handling, perhaps inducing stress differences, is likely to be responsible. Indeed, different types of restraint greatly affect sensitivity on this assay[8].

The large effect of experimenter does not seem to be an artifact of CART analysis, or specific to this data archive. We confirmed that experimenter effects account for more trait variability than genotype in a prospective experiment designed to test this and other hypotheses generated by the CART analysis. (A detailed account of this analysis will be published elsewhere.) Furthermore, this experiment allowed us to partition 87% of the variability in this trait into genotype (27%), environmental (45%) and genotype × environment interaction (15%) sources.

Large projects are often carried out by multiple undergraduate and graduate students, postdocs, technicians and other transient personnel. Furthermore, research, particularly on mutant mice, increasingly involves collaborations. The impact of these 'nuisance' factors becomes greater as data are shared in the growing body of online resources, such as those generated by large-scale phenotyping efforts, including mutagenesis screens and the Mouse Phenome Project[9]. Cautious interpretation of such data is warranted in light of varying within- and between-lab environments, to avoid phenocopies (environmental effects misattributed as genetic effects) or spurious genetic correlations.

By analyzing our data archives, we identified the most salient laboratory environmental factors associated with trait variance, some of which had been noted previously. These effects need to be further explored with mechanistic studies in mice and humans. We expect that stress level may be a common mediator, as environmental stressors can modulate pain sensitivity in either direction[10]. Systematic investigation of gene × environment interactions may yield clinically important information leading to the individualization of pharmacological and behavioral treatment strategies for pain. More generally, we believe that data-mining techniques can be applied to many existing data sets to identify consequential laboratory environmental influences that are robust to experimental validation.

Note: Supplementary information is available on the Nature Neuroscience website.

Elissa J. Chesler[1], Sonya G. Wilson[1], William R. Lariviere[1], Sandra L. Rodriguez-Zas[2] and Jeffrey S. Mogil[1,3]

Depts. of Psychology[1] and Animal Science[2], University of Illinois at Urbana-Champaign, Champaign, Illinois 61820, USA

[3]Dept. of Psychology, McGill University, Montreal, Quebec H3A 1B1, Canada

1. Plomin, R. *Science* **248**, 183–188 (1990).
2. Crabbe, J. C., Wahlsten, D. & Dudek, B. C. *Science* **284**, 1670–1672 (1999).
3. Cabib, S., Orsini, C., Le Moal, M. & Piazza, P. V. *Science* **289**, 463–465 (2000).
4. Wahlsten, D. *Physiol. Behav.* **73**, 695–704 (2001).
5. Mogil, J. S. *Proc. Natl. Acad. Sci. USA* **96**, 7744–7751 (1999).
6. Breiman, L. *Classification and Regression Trees* (Wadsworth, Pacific Grove, CA, 1984).
7. Berkley, K. J. *Behav. Brain Sci.* **20**, 371–380 (1997).
8. Mogil, J. S., Wilson, S. G. & Wan, Y. in *Methods in Pain Research* (ed. Kruger, L.) 11–39 (CRC Press, Boca Raton, Florida, 2001).
9. Paigen, K. & Eppig, J. T. *Mamm. Genome* **11**, 715–717 (2000).
10. Jorum, E. *Pain* **32**, 341–348 (1988).