Strengths and weaknesses in behavioral phenotyping of genetically modified mice

Wolfer DP (1,2,3), Lipp HP (2)

(1) Zurich Center for Integrative Human Physiology (ZIHP), University of Zurich, Switzerland
(2) Institute of Anatomy, University of Zurich, Switzerland
(3) Institute for Human Movement Sciences, ETH Zurich, Switzerland

Transgenic mouse models of Alzheimer's disease have become very popular in the last decade. However, the interpretation of behavioral phenotyping studies has produced controversial data. Perhaps most importantly, there appears to be no consistent mouse AD phenotype. Consequently, the evaluation of treatments and rescue experiments faces considerable obstacles. Yet, these problems are relatively common for many mutant mice, but remain more discrete because typically behavioral analysis of mouse mutants is rarely replicated in genetic models with less widespread interest to the biomedical community. The main conceptual problem is that many researchers expect a mechanistic phenotype reflecting predictable consequences of induced transgenes or gene deletions. This is probably an illusion because what is measured behaviorally reflects the compensatory processes of the brain. This is particularly evident in constitutive mutants in which there is both, developmental compensation and system homeostasis. In fact, careful analysis of the behavioral variation in genetic mouse models often reveals that a significant phenotype is caused by "outliers", while the majority shows comparable scores to the wildtype controls. More disturbing is the fact that multiple testing in a battery often shows significant group differences in different tests, but the scores of the mutants in one test do often not predict the scores in other tests, while inter-test predictability is usually better in wildtype mice. This implies that the genetic manipulation has pleiotropic effects, and that the way mice compensate a genetic bias has a considerable stochastic component. The result is then a syndrome rather than a predictable deficit.
From a practical point of view, this situation implies that behavioral phenotyping of clinically critical mouse models needs larger samples subjected to discriminant analysis rather than standard ANOVA, and that we must search for ways of how to pool data for a meta-analysis. If this is not possible for reasons of economy, the minimal approach is to subject a sample of mice to different behavioral tests, and look for between-test correlations separately in mutants and wildtypes in order to obtain a rough estimate whether one is dealing with a common underlying factor (indicated by decent between-test correlations), or whether the phenotype reflects stochastic developmental compensation.
Supp. Swiss National Science Foundation NCCR "Neural Plasticity and Repair", EUROSPIN