linear and logistic regression models, were used to account for possible sources of experimental variation in phenotyping, including sex, strain, litter, day of analysis, equipment used and blood sample handling. The model indicated that much of the unwanted experimental variation observed across studies can be explained by day of analysis, suggesting a clear way forward for improving future study designs.

Results from Hrabě de Angelis *et al.*[1] provide valuable insights into the roles of genes in normal physiology and disease. Although homozygous mutants were the primary subjects for evaluation, in almost 35% of tested lines, mouse embryos with two copies of the mutant allele did not survive. An additional ~12% were subviable, producing fewer homozygotes than expected (≥13% of 28 progeny). In these cases, heterozygotes were tested. In a small set of lines for which both homozygotes and heterozygotes were tested, homozygotes had about four times more annotations and larger effect sizes for annotations common to both zygosities. This analysis provides insight into gene function with respect to gene dosage, further informing comparison of mouse models to human disease. These results compare well to those produced by a similar large-scale effort reported by the Sanger Institute Mouse Genetics Project, which surveyed 489 targeted alleles[4].

Hrabě de Angelis *et al.*[1] further examine genes involved in multidomain disease phenotypes, including metabolism, bone and skeleton, and neurological and behavioral phenotypic domains. Previously uncharacterized genes comprise about half the genes analyzed, and of these, 88% presented divergent phenotypes. The authors also found extensive pleiotropic effects, with 65% of lines affecting more than one phenotype.

By systematically examining and adjusting for the potential confounding factors that occur in multicenter analyses, this study paves the way for future coordinated phenotyping efforts[5], not just in mice, but also in humans. As the human genome sequence continues to be interrogated with elegant new informatics approaches, and new information about deleterious and silent human 'knockouts' is revealed[6,7], mouse models will align with such analyses and will continue to contribute valuable insight into gene function.

COMPETING FINANCIAL INTERESTS
The author declares no competing financial interests.

1. Hrabě de Angelis, M.H. *et al. Nat. Genet.* **47**, 969–978 (2015).
2. Bradley, A. *et al. Mamm. Genome* **23**, 580–586 (2012).
3. Mallon, A.M., Blake, A. & Hancock, J.M. *Nucleic Acids Res.* **36**, D715–D718 (2008).
4. White, J.K. *et al. Cell* **154**, 452–464 (2013).
5. Brown, S.D. & Moore, M.W. *Mamm. Genome* **23**, 632–640 (2012).
6. Callaway, E. *Nature* **514**, 548 (2014).
7. Sulem, P. *et al. Nat. Genet.* **47**, 448–452 (2015).

Research Highlights

Papers from the literature selected by the Nature Biotechnology editors. (Follow us on Twitter, @NatureBiotech #nbtHighlight)

Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system
Zetsche, B. *et al. Cell* doi:10.1016/j.cell.2015.09.038 (25 September 2015)

An integrated map of structural variation in 2,504 human genomes
Sudmant, P.H. *et al. Nature* **526**, 75–81 (2015)

Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments
Sun, K. *et al. PNAS* **112**, E5503–E5512 (21 September 2015)

Sca/eS: an optical clearing palette for biological imaging
Hama, H. *et al. Nat. Neurosci.* **18**, 1518–1529 (2015)

Genetic determinants of *in vivo* fitness and diet responsiveness in multiple human gut *Bacteroides*
Wu, M. *et al. Science* **350**, 6256 (2015)