Mouse Phenome Database: A resource to address phenotypic heterogeneity, research reproducibility and replicability

M.A. Bogue, S. C. Grubb, V. Philip and E.J. Chesler

The Jackson Laboratory, Bar Harbor, Maine, USA
molly.bogue@jax.org

Mouse Phenome Database (MPD; phenome.jax.org) allows data from complex trait studies to be found, reproduced, reused and placed into the context of other studies

Access to primary data is crucial for three reasons: 1) reproducibility, 2) reanalysis in light of new developments and 3) integrative analysis to find consensus among diverse studies. Unfortunately, data often exist in diverse and sometimes non-computable stores with insufficient documentation and restricted access. MPD is a widely used online resource providing access to primary experimental data and protocols in the predominant genetic model organism, the laboratory mouse [1]. The MPD, in existence for the past 12 years, amasses, annotates, integrates and maintains primary quantitative phenotype data and protocols in a centralized public database. Since the inception of MPD, a wealth of phenotype technologies and mouse resources have led to an expanded scope and refocus of the system – from inbred strain characteristics to a rigorously curated data resource for complex trait and integrative genetic analysis. This resource houses phenotypic, gene expression or genotype data for >1300 strains. MPD provides a catalog of phenotypic assays, analysis tools to explore genetic variation, and a common framework for data access and data dissemination. Data come from investigators around the world (supported by ~130 governmental funding agencies and research foundations) and represent a broad scope of behavioral endpoints and disease-related characteristics in naïve mice and those exposed to drugs, environmental agents or other treatments. MPD provides an important venue for compliance with data sharing policies and facilitates data reuse, saving time and resources while reducing animal use.

Integrating various sources of phenotype data in MPD provides researchers with the resources they need to reproduce experiments, reanalyze genetic studies with new algorithms and genetics maps, understand relationships among traits and elucidate the shared genetics for a multitude of traits. The high level of documentation and curation standards and stability of the program at The Jackson Laboratory have made MPD a primary resource for investigators to archive and retrieve quantitative mouse phenotypic data. This high-quality, standardized data resource enables investigators to select mouse strains for modeling disease, compare results of diverse phenotypic assays and benchmark experimental data using detailed protocols.

The utility of model organism research is dependent on excellent research reproducibility

Inadequate reporting and encumbered access to primary data reduce the impact of studies and act as a barrier to effective translation of scientific discoveries from basic/preclinical findings to human applications including development of new diagnostics and therapies. Substantial recent attention has been given to poor cross-species relevance, replicability and reproducibility of research findings in the laboratory mouse. The reasons for poor reproducibility are many-fold and include inadequate reporting of mouse strains and resources, insufficient statistical power, diversity among experimental protocols and lack of documentation of research resources. It is imperative that studies are reported with detailed experimental information to allow researchers to evaluate reported results, repeat experiments and extend findings.

MPD has consistently provided rigorous curation of mouse experimental data to help alleviate issues associated with reproducibility. By structuring mouse phenotyping studies, annotating them to controlled vocabularies and developing integrative tools that rely on the unique value of this data, MPD facilitates access and reuse of primary phenotype data, enabling cross-species comparisons and ultimately assuring relevance to human studies.

This will help achieve a more holistic understanding of disease processes and mechanisms and maximize the value of the data while leveraging the investments made in basic research.

Interpreting the heterogeneity of behavior

Understanding the genetics of complex disease, and in particular behavioral disorders, requires sensitivity to the challenge of modeling disease in mice in a manner that captures the heterogeneity, complex interrelations and co-occurrences among disorders. Behavioral experiments are quite sensitive to environmental conditions and experimenter effects. Furthermore, researchers often choose a single experimental assay as an endpoint among many possible tests of a behavioral characteristic, despite the different biases or varied aspects of behavior modeled by each. Inconsistencies among these assays may confound the already deep complexity and heterogeneity of behavior. Although mice cannot recapitulate the full complexity of certain psychiatric disorders and diseases, many component features can be independently observed and genetically characterized.

MPD tools help investigators to identify a consensus among diverse assays and varied environments through genetic correlations among traits and underlying molecular mechanisms. This strategy is widely used in behavioral genetics and has made behavior one of the most highly represented fields of study in MPD. We are in the process of enhancing the MPD system by expanding the scope of data and developing new analysis tools to further address phenotypic heterogeneity, research reproducibility and replicability. We have plans to incorporate a more sophisticated suite of tools for integrated multi-dimensional analysis that will better enable correlation analyses and discovery of biological mechanisms. Genetic parameter estimations and population distributions will help researchers to identify those assays and mouse resources that will provide the best experimental characteristics for the assessment of specific behavioral and biological concepts. Integrative analysis methods will enable discovery of coherent signal amidst gene-environment interactions, laboratory environmental diversity and assay diversity.

Reference

[1] Grubb, S.C., Bult, C.J., Bogue, M.A. (2014). Mouse Phenome Database. *Nucleic Acids Research, **42**(Database issue)*, D825-834.