6120 Executive Boulevard, Suite 550
Rockville, MD 20852
society@genetics-gsa.org

6120 Executive Boulevard, Suite 750
Rockville, MD 20852
ascbinfo@ascb.org

Francis S. Collins, MD, PhD
Director of the National Institutes of Health

Diana W. Bianchi, MD
Director of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development

Eric D. Green, MD, PhD
Director of the National Human Genome Research Institute

Walter J. Koroshetz, MD
Director of the National Institute of Neurological Disorders and Stroke

Jon R. Lorsch, PhD
Director of the National Institute of General Medical Sciences

Griffin P. Rodgers, MD, MACP
Director of the National Institute of Diabetes and Digestive and Kidney Diseases

Dear Dr. Collins, Dr. Bianchi, Dr. Green, Dr. Koroshetz, Dr. Lorsch, and Dr. Rodgers:

We are contacting you about NIH funding that supports model organism research. We first want to thank you for your past support, including your positive response to the community statement of support presented to you at the Genetics Society of America's TAGC meeting in 2016. We are now contacting you because we are very concerned about the ongoing and severe cuts to funding of Model Organism Databases (MODs), and the large, negative impact these cuts will have on biomedical discovery and human health.

Model organism research is a crucial aspect of the NIH's mission to improve human health and is currently funded by the NIH at over $15 billion dollars per year. A key aspect of the power of model organisms is the collective expertise, reagents, and knowledge that have accumulated across decades of research using these unique genetic resources. Because of the evolutionary conservation of basic biological processes, model organism data are especially important to researchers trying to gain insight into the genetic basis and mechanisms of human disease.

The Model Organism Databases (MODs) have been the primary means of organizing these data and are a core foundation of the collective resources across communities. The MODs have exponential effects on research because they facilitate integrative and synthetic understanding of gene function. The primary MODs for biomedical research include (but are not limited to): the Saccharomyces Genome Database (SGD), WormBase, FlyBase, the Zebrafish Information Network (ZFIN), Mouse Genome Informatics (MGI), and the Rat Genome Database (RGD). Each of the MODs collaborates closely with many other important bioinformatics resources, particularly the Gene Ontology (GO), and also integrates a suite of genomic data repositories and computational biology analysis platforms.

While the mission of these MODs was focused initially on representing genetic and genomic information, their scope has expanded to include cataloging key reagents (*e.g.* mutant lines, transgenic lines, DNA constructs, antibodies). This task is crucial for the NIH's mission to validate reagents and increase reproducibility. MODs also provide searchable curated information about diverse biological properties of organisms and genes, including disease associations, phenotypes, interaction pathways, and gene and protein expression. The key aspect of this work for the scientific community is standardizing and integrating data from multiple resources, including sources of high-throughput experimental data and hundreds of biomedical journals. The expert curation supported by the MODs is essential to make these data and scientific publications conform to FAIR (findable, accessible, interoperable, reuseable) standards. Investment in the MODs pays off many-fold in terms of efficiency, reproducibility, and rigor of research.

The impact of MODs has been transformational for biomedical research. The ontologies and annotations generated by the MODs are integrated into hundreds of bioinformatics resources and have been foundational to data science innovations that rely on semantic reasoning to support predictive biology. It was collaboration among the MODs that led to the formation of the Gene Ontology, a resource that forever changed how genome-scale data are analyzed and interpreted. The MODs and GO are constantly evolving and adapting in response to data generated by the research community and to new techniques in data science, including development and use of natural language processing programs for curation. **These aspects of their mission rely on software development and expert human curators who comb newly published literature, properly tagging and integrating it, and converting it to a form amenable for computation.** Given the emerging impact of data science on biomedicine, expertly-curated knowledgebases are needed now more than ever for future development of human-AI interfaces.

Despite the fact that these community resources support research funded by all of the NIH institutes, most of the funding for MODs and GO has been provided by only a few institutes that have been paying an outsized portion of the costs. **Model organism**

database funding is now undergoing a severe decrease: five of the MODs listed above, as well as GO, have been cut by 25% and will sustain another 25% cut in the near future for a total decrease of approximately $8M in annual funding (~50% of 2016 funding, not including inflation). These cuts threaten the sustainability of the MODs and GO. As a result of the sustained reductions in funding, many staff members have already been laid off and more departures are anticipated. We are facing significant reductions or cessation of the curation of new information, resources, and literature THIS YEAR.

Some new funding from NHGRI and NHLBI was awarded to develop centralized infrastructure for the MODs and GO. The new knowledge commons platform—the Alliance of Genome Resources—aims to reduce redundancy in software development across these databases and to make them more interoperable. The Alliance is making good progress and could be a model for efforts to modernize the data ecosystem that supports biomedical research.

But it is crucial to point out that funding for the Alliance of Genome Resources does **not** support curation activities. The Alliance draws information from individual model organism databases. If the individual MODs and GO are not sustained (because of cuts in curation activities), the Alliance of Genome Resources will ALSO fail, as it will be starved of the very information it is designed to serve to the scientific community. Moreover, the Alliance does not provide all the organism-specific data crucial for researchers to design experiments and exploit the power of each distinct model. Thus, it is essential that the NIH continue to fund the model organism databases and their curators as they continue to catalog new information and support organism-specific resources.

It would be difficult to overstate how devastating the loss of these resources would be for basic research. The MODs and GO pay for themselves in terms of (a) time and money saved by centrally collating information, avoiding duplication of effort by numerous scientists, (b) time and money saved repeating experiments that have already been performed, and (c) time and money saved creating reagents that existed elsewhere in the literature (e.g. transgenic or mutant lines). Furthermore, in addition to maximizing the potential output of dollars committed to research, the curation of information by the MODs also serves a second NIH mission: making information paid for by taxpayers publicly accessible. Importantly, a lack of adequate support for the MODs will have a large, negative impact on model organism research and biomedical discovery.

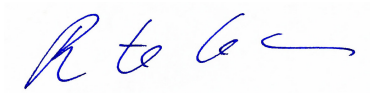
The cost of continuing full support of the MODs and GO (restoration of ~$8M / year), constitutes only ~0.05% of the NIH's yearly expenditures on model organism research. Given the low cost of the model organism databases relative to their important support of multiple crucial NIH missions, we urge you to help restore the funding of these critical resources. Because the MODs and GO support researchers in all fields of biomedical research, the ideal approach would be a pan-NIH institute funding

mechanism that results in stable and sustainable support for these core community resources. Quick and decisive action is necessary, given that these organizations are faced with soon having to terminate MOD employees that have carried out vital work to support the scientific community for many years, a crucial expertise that will be irrevocably lost from this research infrastructure.

Thank you for your prompt attention to this matter that affects a very large proportion of the NIH-funded scientific research community.

Hugo Bellen, *President, GSA*

On behalf of GSA's Executive Committee:
E. Jane Hubbard, *Vice President, GSA*
Denise Montell, *Past President, GSA*
Erika Matunis, *Secretary, GSA*
Michael Buszczak, *Treasurer, GSA*
Steven Munger, *Executive Committee Member at Large, GSA*



Ruth Lehmann, *President, ASCB*

On behalf of ASCB's Executive Committee:
Eva Nogales, *Past President, ASCB*
Marty Chalfie, *President-Elect, ASCB*
Kerry Bloom, *Secretary, ASCB*
Malcom Campbell, *Treasurer, ASCB*