2016 Drosophila White Paper

The first Drosophila White Paper was written in 1999. Revisions to this document were made in 2001, 2003, 2005, 2007, 2009 and 2012; these versions are available at: http://flybase.org/wiki/FlyBase:Fly_Board. Here, the Drosophila Board of Directors presents an updated White Paper identifying and prioritizing current and future needs of the Drosophila research community, based on input from community leaders and comments received from community members. This White Paper was approved by the Drosophila Board in 2016.

Part I *Drosophila* as an experimental system for research: past, present, and future

Drosophila melanogaster is a leading animal model for biomedical research and understanding the basic biology of animal systems. Lessons acquired from studies in *Drosophila* directly impact our understanding of evolutionarily distant metazoans, including humans and other vertebrates, as well as invertebrates such as mosquitoes that are of medical or agricultural importance.

Our understanding of the basic principles of genetics, including the nature of the gene, genetic linkage, meiotic chromosome segregation, and recombination, all arose from studies in *Drosophila*. Pioneering studies that linked molecular lesions in the genome with mutant phenotypes led to the identification of many proteins that play essential, conserved roles in development and physiology. For example, many of the components of systems that cells use to communicate with each other and respond to their environment, including the Notch, Wnt, Hedgehog, Hippo, and Toll signaling pathways, and Trp channels, were first discovered and characterized in *Drosophila*. Components of these pathways are now recognized as central contributing factors to major human diseases including cancer, cardiovascular disease, and neurological disorders, and drugs targeting these pathways are in use or in clinical trials today. Thus *Drosophila* research provides an essential pipeline for discovery of drug targets and, in some cases, direct identification of lead compounds and drugs.

Drosophila research has defined not only molecules and pathways but also fundamental biological processes, including the innate immune response, stem cell determination and maintenance, cell and tissue polarity, growth control, pattern formation, organ morphogenesis and physiology, circadian rhythms, sensory biology and animal behavior, learning and memory, neuronal pathfinding, and synaptic transmission. *Drosophila* thus serves as an outstanding organism for understanding animal biology and modeling human disease, including identifying molecular mechanisms and new therapeutic strategies. The enormous contributions of *Drosophila* research have been acknowledged in part through recognition of many *Drosophila* researchers with major scientific prizes, including several Nobel prizes.

Drosophila will, with adequate funding, continue to play a key role in future research, providing insights into both fundamental biological processes and human disease, as *Drosophila* presents a unique and overwhelming combination of strengths as an experimental model. These include the wealth of information accumulated during a century of research on its genetics, development, physiology, ecology, and evolution, a vigorous and collaborative community of researchers, relatively low maintenance costs, short generation time, simple genome, and an extensive and accessible toolkit that provides diverse strategies for manipulation and visualization of gene function. The unique position of *Drosophila* as a complex, yet easily manipulated and analyzed, animal model makes it well suited for a broad range of studies including investigations of organ development and physiology, neural function across scales from molecules to neural networks to behaviors, transcriptional regulation including *cis*-regulation, nuclear architecture, gene regulatory networks, and epigenetics, and the genetic basis of complex traits. *Drosophila* studies also provide insight into the importance of gene-gene interactions, and powerful tools to identify genes and pathways relevant to orthologous complex traits in humans, gene-environment interactions, including interaction between the microbiome and animal physiology, metabolomics and pharmacogenetics, and identification and characterization of human disease genes. In addition, the genus *Drosophila* has been a key model system for understanding population biology, the molecular basis of speciation, and evolution. *Drosophila* also serves as the closest genetic model for the major insect vectors of disease, including as *Anopheles gambiae* (malaria), *Aedes aegypti* (zika, dengue fever, yellow fever), and *Culex pipiens* (West Nile fever), as well as many agriculturally important insects, including pollinators such as honeybees, and pests that include many species of beetles and aphids.

The ability of *Drosophila* research to continue to pioneer our understanding of general principles underlying the biology of animals including humans depends both on the availability of funding, and on continual reassessment of the resources necessary to support *Drosophila* research. We prioritize continued funding of investigator-initiated research into both basic and applied problems in biological sciences. We also encourage better integration of *Drosophila* researchers during the planning stages of larger projects, much like

our community's participation in the Genome and ENCODE projects. We encourage support for community identified shared resources, as outlined in this document.

Part II Maximizing contributions of *Drosophila* research

Here, we outline current resource priorities of the *Drosophila* research community, in order of importance.

1) Informatics Resources for *Drosophila* Research

To ensure that *Drosophila* continues to play its essential role in both basic and translational biomedical research, it is crucial that there be a central bioinformatics resource that captures, organizes and presents core information on *Drosophila* genomics and genetics, both from the primary literature and from large-scale data- and resource-generation projects. The primary resource for this is currently Flybase, and there is universal agreement that continued support for the curated resources exemplified in Flybase is essential to all *Drosophila* research. Key informatics resources include genome and transcriptome sequence information, up-to-date gene annotations, the characterization of mutant phenotypes, RNA and protein expression profiles, and interacting gene, protein, RNA and small molecule networks, and catalogs of *Drosophila* stocks and molecular reagents, as well as databases for new classes of information such as gene expression atlases, neural connectivity, and metabolomics. Whereas capture of some classes of information from the literature may be automated, organizing and presenting most classes of information requires manual curation. All these data classes require community input, direction and oversight. Generic genetics and genomics databases are not a viable substitute.

To enhance the accessibility and utility of *Drosophila* bioinformatic resources, both for *Drosophila* researchers and for those working with other systems, it is essential to link resources dedicated to *Drosophila* with those dedicated to other organisms. Evolution is a powerful genomics tool that informs research on organisms throughout the tree of life. Nascent interactions among databases supporting the well-established model systems and human genomic and genetic disease information must be strengthened and made more accessible. Not only will this promote more rapid progress in *Drosophila* research, it will significantly enhance progress in functional genomics overall by promoting cross-talk among scientists working in different fields. Up-to-date and well-organized electronic databases are essential conduits to translate information from *Drosophila* research to other areas of study, including the study of human biology, genetic disease and biomedicine, cellular responses to infectious pathogens, and dipteran disease vectors. Maintaining a current and organized database requires not only an investment in effectively linking databases, while preserving their essential and diverse contents, but also creating interfaces that make them accessible to varied user groups. At the same time, it is essential that the unique classes of information fundamental to *Drosophila* research be preserved and enhanced so that these databases continue to benefit future research. We are concerned about fragmented NIH policy on database support and the lack of international efforts to support this infrastructure. Our community would like to play a more active role in establishing these programs, rather than having decisions imposed on us.

2) Resources for analysis of genes and phenotypes

Resources that facilitate functional analysis of genes and phenotypes are a high priority for *Drosophila* researchers. A powerful advantage of *Drosophila* as a model system lies in the wide repertoire of genetic manipulations that are possible; continued enhancement of this genetic toolkit should include expanding the set of genes with loss-of-function mutations, including null alleles created by gene deletion or disruption for genes not already represented in existing mutant collections, and resources that facilitate replacement of genomic loci with allelic variants. CRISPR/Cas9 technology makes it possible to target any gene, and an expanded collection of mutations that covers most or all genes, including genes without large ORFs (encoding peptides or small RNAs), and hence underrepresented in gene disruption collections, will be a valuable resource for a wide range of studies. Development of genetic resources should advance strategies and genome-wide resources for manipulating the activity and expression of genes with tight spatial and temporal control, including expression of wild-type or variant alleles, and fly lines that enable targeted knock-out or knock-down of gene expression. This can be done through RNAi, strategies based on CRISPR/Cas9 and its derivatives, or protein degradation strategies, in combination with independent systems for spatial and temporal manipulation of expression (*e.g.* GAL4, LexA, QF) to allow conditional and reversible removal of genes, mRNA or proteins in any tissue at any time. Insertional mutations created by targeting GAL4 or LexA to

knock down gene function, combined with expression of cDNAs under GAL4 or LexA control, will enable proper spatial and temporal expression for rescue experiments, including expressing altered genes for structure-function studies, expressing tagged proteins for analysis of protein localization, and expressing homologous genes from humans or other species.

We support continued development of tools to study human genes and their disease variants in *Drosophila*, facilitating emerging strategies in precision medicine, and accelerating characterization of undiagnosed diseases. Creation of a library of human cDNAs in fly-ready vectors allows all researchers to quickly obtain, modify and study human genes, and we advocate creation of a collection of transgenic fly stocks that carry tagged UAS-human cDNAs. This will permit testing of function of human genes in *Drosophila*, and provide a basis for the functional testing of human disease variants, an increasingly common need in medical genomics.

We advocate support of community facilities and resources for high-throughput screening, including RNAi or CRISPR/Cas9-based screening, and pharmacological screening, both in cell lines and in whole animals. While the ability to analyze genes and phenotypes *in vivo*, in an intact animal, is a particular strength of *Drosophila*, some classes of experiments can be more easily performed on cultured cells, and expanding the collection of available *Drosophila* cell lines to include more diverse cell and tissue types, and improving on methods to culture cells and tissues *in vitro*, will facilitate live imaging studies, and biochemical and pharmacological characterization and screening of cells and tissues.

We advocate for resources that enable, enhance and expand physiological and phenotypic characterization of *Drosophila*. These will provide a deeper understanding of responses to environmental perturbations, gene-environment interactions, and polygenic traits. This should include annotation of the *Drosophila* metabolome, and the establishment of standardized protocols and resources to permit comparisons of the metabolome across tissues, genotypes, and species. It should also include analysis of the *Drosophila* microbiome and its contribution to physiology, including resources to characterize microbiomes from diverse genetic backgrounds and environments.

Tools and resources to determine expression patterns of *Drosophila* RNAs and proteins at high temporal and spatial resolution, together with sub-cellular localization profiles, provide essential insights into function and valuable markers for phenotypic characterization. To extend the expression analysis tool-kit, we advocate two complementary approaches: the creation of collections of tagged genes and the production of antibodies against *Drosophila* proteins. Antibodies are a foundational resource in molecular biology, as they enable the study of protein localization, modifications, and interactions, *in situ,* with genes under endogenous regulatory controls, without any potential for impairment of gene function by tags. A repository of highly specific, high affinity, and sustainable antibodies will be a valuable resource, and in addition to immunization, synthetic techniques, including recombinant antibodies, nucleic acid aptamers and non-immunoglobulin protein scaffolds should be expanded. Tags are needed as an efficient, reliable, and inexpensive way to study protein localization and characterize protein function, given current limitations of antibody resources. Limited sets of tagged genes are currently available, but broader gene sets need to be generated, along with stable fly lines, and the activity of tagged proteins needs to be confirmed by genetic rescue experiments. These collections should include tagging endogenous genes with markers (*e.g.* GFP) at their genomic loci, without disrupting gene function, to assess expression patterns of genes and subcellular localization of proteins in wild-type and mutant backgrounds, and provide reagents for GFP-based knock-down or immunoprecipitation experiments. Collections of tagged transgenes carrying tagged cDNAs (*e.g.* UAS-cDNA-tag) can also be used for localization and interaction studies, and are valuable for structure-function studies and comparisons to human UAS-cDNA collections. Many genes produce multiple transcript isoforms via RNA processing mechanisms, including regulated alternative splicing, and future analysis of expression patterns should include the spatial and temporal distribution of alternative transcripts and protein isoforms.

Support for functional analysis of the *Drosophila* genes and phenotypes must be coupled to bioinformatic efforts that will establish atlases and databases of the resulting data sets, and make them accessible to all researchers, as described above in part 1. It must also be coupled to mechanisms for making tools and resources widely available, as described below in parts 3 and 4.

3) *Drosophila Stock Centers*

Stock centers that provide universal access to genetically defined stocks are essential for all *Drosophila* research and they remain a high priority for infrastructure funding. They are complex operations that are heavily used by the national and international fly communities. For example, the Bloomington Drosophila Stock

Center, the repository for *Drosophila melanogaster* strains funded by NIH, maintains more than 59,000 genetically distinct stocks and distributed 243,148 samples to approximately 2,000 laboratories during 2015. These centers, whether general or specialized in scope, distribute the "core" stocks necessary for genetic experimentation in *Drosophila*.

Stock centers must have the physical ability to maintain the large number and variety of stocks needed for contemporary genetics research in a safe and reliable manner, and, to retain relevance and impact, they also need the management capacity to assure that collection contents adjust to changing research needs. Stock centers must keep valuable existing stocks while acquiring new stocks from researchers and integrating with or leading large-scale resource development projects. To maximize the benefit of maintaining the strains, stock centers must provide information that will promote their experimental use by integrating stock information into online model organism databases such as FlyBase, emphasizing website development and maintenance, and having staff available for consultation. These efforts to provide information on stock applications are particularly important to investigators new to *Drosophila* research, such as those wishing to pursue discoveries made in vertebrates using the sophisticated genetic approaches available in flies. Stock centers must also have the capacity to deal with the regulatory challenges associated with the distribution of live animals and the administrative challenges of acquiring large proportions of operating budgets from user fees.

We urge funding agencies to recognize that the viability and vitality of stock centers depends on the appropriate balance between grant support and user-generated income. Cost-recovery programs have enabled stock centers to expand beyond the limits of grant funding, but, as public resources important to scientific progress, stock centers need the security and stability provided by continued public investment and oversight. The continued success of stock centers will depend on agencies giving them flexibility in determining staffing, the structures of cost-recovery programs and the uses of fee income. We strongly believe that healthy partnerships between stocks centers and funding agencies will continue to be a key factor in the success of *Drosophila* as a research organism.

4) Molecular and Cell line Stock Centers

Molecular and cell-line stock centers provide the community with access to an expanding set of key resources at affordable costs, enhance research capabilities, enable efficient use of resources, and facilitate exchange of materials. It is important to maintain reliable, central repositories that are able to distribute key reagents to the scientific community expeditiously as it can relieve individual labs of this responsibility and afford the end user with a dependable timeline for receiving materials. A central repository also ensures that these valuable resources are not degraded or lost, and provides technical guidance and ready access to reliable, relevant protocols. In addition, the importance of a molecular stock center is magnified by NIH guidelines that require investigators to make materials widely available and that emphasize reproducibility.

Key resources to be maintained and distributed include cDNA clones and transformation vectors, as well as collections of full-length cDNA and genomic clones for expression in flies, in cell lines, and in yeast or bacteria. Molecular reagents for manipulation of gene expression (*e.g.* by RNAi or CRISPR/Cas9) also need to be maintained and distributed. A molecular stock center needs to be able to accept both resources generated by large-scale projects, as well as donations from individual labs. A reliable, centralized repository of *Drosophila* cell lines also needs to be maintained. Support for antibody repositories is also invaluable. Some *Drosophila* monoclonal antibodies are available from the NIH-supported Developmental Studies Hybridoma Bank, but support for storage and distribution of polyclonal antisera, and antibody reagents created by other techniques such as phage display, would also be valuable.

5) Long-term preservation of *Drosophila* strains

Unlike the strains of most other genetic model organisms, *Drosophila* strains cannot be maintained practically in any form other than living cultures. The development of robust methods for the long-term preservation of *Drosophila* strains would benefit biomedical research by providing more options for maintaining and distributing strains, allowing the preservation of important, but rarely used strains, preventing the accumulation of mutations associated with long-term culture, and helping to secure genetic resources from disaster. Recent advances in cryogenics, dehydration technologies and nutritional and environmental manipulations suggest that new methods for long-term preservation could be developed for *Drosophila* embryos, larvae or sperm. We strongly suggest exploring these possibilities and investing in the development of methods that hold promise.