Drosophila Board White Paper 2005
December 2005

Explanatory Note: The first Drosophila White Paper was written in 1999. Revisions to this document were made in 2001 and 2003.
http://flybase.net/.data/docs/CommunityWhitePapers/DrosBoardWP2001.html
http://flybase.net/.data/docs/CommunityWhitePapers/DrosBoardWP2003.html
At our 2004 meeting, the Drosophila Board of Directors decided to write a new White Paper to take stock of the progress made in the preceding two years and to assess current and future needs of the Drosophila research community. This draft was prepared by the Board, and modified according to feedback received from the Drosophila research community.

The importance of the Drosophila model for understanding basic biological mechanisms is ever more evident. The union of the powerful genetic manipulations built on a century of genetic research on *D. melanogaster* with modern genomic technologies (many made possible through the foresight and support of NIH) makes *D. melanogaster* an unparalleled model for understanding animal biology. Further, the genus Drosophila has been an important model for understanding animal populations and evolution. The striking level of conservation of many genes, proteins and pathways between fly and human have ensured that Drosophila research is directly relevant to human disease, and indeed, the vertebrate biological community often looks to Drosophila research to identify candidate genes for pathways or diseases of interest, and to provide insights into the mechanisms underlying these vertebrate processes. Key insights have been gained in recent years into the genetic and cellular mechanisms of processes such as neurodegeneration, vasculogenesis, the innate immune response, stem cell determination and maintenance, cell and tissue polarity, signal transduction, growth control, neural control of behavior and organogenesis.

In addition to insights gained for basic biology and human disease, Drosophila research also impacts on human health by serving as the closest genetic model for the insect vectors of disease, such as *Anopheles gambiae* (malaria), *Aedes aegypti* (dengue fever, yellow fever), *Culex pipiens* (West Nile fever), *Rhodnius proxilus* (Chagas disease) and *Glossina morsitans* (African sleeping sickness). It is now becoming clear that Drosophila research is producing the technologies and information sets that will inform genomic/genetic research in these medically important species, and that the Drosophila community is also likely to provide the training for future researchers in these species.

Studies of Drosophila have provided fertile testing ground for new approaches in genomic research. Continued and even greater success relies on the maintenance and expansion of key projects and facilities and on the development of new technologies. To this end, the Drosophila research community has identified current bottlenecks to rapid progress and defined its most critical priorities for the next two years. We begin by first noting recent achievements that have been most important for the community-at-large:

- Progress toward completion of the *Drosophila melanogaster* genome through refinement in the sequencing of some of areas of the euchromatic arms (Release_4) and progress toward high-quality sequence of heterochromatin.
- Updates to the *Drosophila melanogaster* gene annotation set (Release_4.2).
- Insights gained into gene and genome organization and evolution through the sequencing of the euchromatin of eleven additional *Drosophila* species: *ananassae*, *erecta*, *grimshawi, mojavensis*, *persimilis*, *pseudoobscura, sechelia, simulans, virilis, willistoni, yakuba.*

- An expanding library of complete cDNAs.
- An expanding collection of mutant strains with transposable element insertions or point mutations disrupting over 50% of the nearly 14,000 annotated genes.
- Ten-fold expansion of the number of Drosophila cell lines available for study.
- Progress toward the goal of complete coverage of the genome with chromosomal deficiencies mapped to the sequence.
- Development of RNA-interference (RNAi) technologies for cultured cells and whole animals.
- Continued improvement of genetic techniques such as targeted gene disruption.
- Transcriptional profiling of the complete life cycle and many tissue types.
- Progress toward genome-wide tiling arrays for complete transcriptional profiling and genome-wide protein binding site mapping by ChIP-array.
- Database development to integrate genome and genetic resources for Drosophila.

These achievements have been accomplished through a collaboration of the research community to recognize and prioritize its most pressing needs, and the funding agencies to provide the resources and coordination necessary to meet these needs. Further progress in Drosophila research depends upon a continuation of this most important collaboration. This White Paper represents an updated view of the most important priorities for near term future community resources. It is written with knowledge that a separate white paper on a Drosophila ENCODE project to define the DNA elements in the entire *Drosophila melanogaster* genome has been submitted to NHGRI. The community enthusiastically endorses the concept of the *D. melanogaster* ENCODE project.

There is overwhelming agreement that the following three resources must continue to be supported to serve the entire research community.

1) Stock centers that provide a comprehensive range of genetically defined and wild-type stocks at affordable costs. Existing capacity of 25,000 *D. melanogaster* strains at the Bloomington Stock Center is expected to meet community needs for the next 2 to 3 years only. This number takes into account current efforts to accumulate functionally defined mutant alleles for every gene, deficiencies that provide extensive coverage of the genome, transposable element insertion alleles being generated by the on-going gene disruption projects and lines with fluorescent protein fusions to endogenous proteins. Plans are underway for the generation of 10,000 lines expressing RNAi constructs, which will bring the needed capacity for *D. melanogaster* strains to 35,000 over the next five years. Therefore, additional capacity must be developed in new or existing stock centers.

The NHGRI species sequencing and BAC projects have driven increasing demands for stocks of the twelve sequenced species and their relatives from the Tucson Stock Center. Tucson currently maintains approximately 1500 different stocks of about 250 species. Given the envisioned acquisition of fresh wild type and newly created genetically marked stocks of these other species, this number will double in the next two to three years. While Tucson's space and infrastructure are adequate to accommodate the increase, the Center is already understaffed.

2) Expanded and improved electronic databases to capture and organize Drosophila data, and integrate the information with other databases used by the research community. It is essential to support efforts that can keep pace with the enormous acquisition rate and increasing complexity of data being generated by Drosophila

researchers, including the sequence of eleven new Drosophila species, up-to-date gene annotations and the characterization of mutant phenotypes, RNA and protein expression profiles, and interacting gene, protein, RNA and small molecule networks. These efforts must also include effectively linking Drosophila databases with those of other organisms, including other well-established model systems and emerging systems for genome research. Not only will this development promote more rapid progress in Drosophila research, it should significantly enhance progress in functional genomics overall by promoting crosstalk among scientists working in different fields. Up-to-date and well-organized electronic databases are essential conduits to translate information from fly research to human research.

3) Continued support for a molecular stock center that provides the community with fair and equal access to an expanding set of key molecular resources at affordable costs. These resources include commonly used vectors, full-length cDNA clones, EST clones, cell lines, genomic libraries and microarrays. A well-run molecular stock center is cost effective for grant dollars, serves multiple research communities and plays a catalytic role by making available resources that might otherwise remain closely held. Moreover, the importance of a molecular stock center is magnified by new NIH guidelines requiring investigators to make materials available through such centers.

In addition to the resources described above, certain research projects that require large infrastructures and investments over several years must be in place to realize the full potential of Drosophila as a model system for functional and comparative genomics. Several of these projects are ongoing, use existing technologies, and require adequate funding for their successful completion. Others are projects that require the development of new technologies. The research community considers the following high priority projects.

4) Functional analysis of the Drosophila genome. The most powerful advantage of *Drosophila* as a model system lies in the wide repertoire of genetic manipulations possible. Key to all genetic approaches is the ready availability of loss of function mutations in all genes. An ongoing NIH-funded project will provide for the generation and sequencing of nearly 10,000 unique P-element insertions for an anticipated 75% coverage of the annotated genes. Because many genes will be refractory to mutagenesis by transposable elements, alternatives to P element gene disruption techniques should also be considered a high priority. Developing technologies such as TILLING, PCR-based deletion screening, and SNP mapping of point mutants are important to accomplish the functional analysis of the entire genome by mutations. RNAi screening is another powerful approach for functional analysis of the Drosophila genome. The value of a centralized facility has already become clear from the experience of the NIGMS-supported Drosophila RNAi Screening Center (DRSC). Over 7,000 genes have been linked to a phenotype in one or more assays developed in Drosophila cells. Important improvements include: replacing ~5% of the dsRNAs in the existing library that have off-target effects, generating dsRNA libraries targeting specific classes of genes, improving screen automation, data acquisition and data normalization, and integration of the DRSC database with existing ones to enable cross validation and high confidence references for data mining purposes. We encourage continued support for centralized RNAi screening, as well as distribution of validated RNAi resources to the community.

5) Capturing temporal and spatial expression patterns for all Drosophila genes and proteins. With the addition of microarray and RNAi screens to genetic screens, the need

for information on gene expression at the cellular and subcellular level is increasingly acute. Ongoing efforts have demonstrated the utility of genome-wide analysis of RNA expression patterns using RNA *in situ* hybridization to embryos. Thus far, over 5,000 genes have been analyzed and these efforts have demonstrated an economy of scale. This analysis should be completed for all genes and extended to other tissues at different stages of the life cycle. Particularly powerful is the protein-trap technology using a transposable element with a GFP-containing exon to mark proteins and analyze tissue and sub-cellular distribution of proteins *in vivo*. Support to generate, maintain and provide these lines to the community is considered a high priority since *in vivo* applications are broad and powerful. The development of sophisticated imaging methods that could capture dynamic expression patterns in multi-dimensions and with sub-cellular resolution will add substantially to the utility of this information. Antibodies are invaluable tools for expression profiling, as well as biochemical analyses; however antibody production is inefficient for individual labs. One of the most important goals for in the next few years should be the production of a large collection of antibodies against Drosophila proteins. New approaches to generating panels of polyclonal and monoclonal antibodies are needed to accelerate availability of these powerful reagents. Large scale peptide generation, recombinant protein production and purification, immunization and bleeding schedules, screening for titers and applicability for blotting, IP and histochemical studies are all best organized and executed in a streamlined manner in a dedicated facility. Systematic approaches to making antibodies to classes of proteins such as membrane proteins, secreted proteins and transcription factors would be particularly useful.

6) Production of comprehensive cDNA resources. cDNA sequencs for the majority, if not all, of the genes of *Drosophila melanogaster* will be of enormous use for gene annotations and expression studies at the level of individual genes or on global scales using microarrays. Ongoing efforts to obtain and sequence full-length cDNAs should be supported. In addition, the insertion of the complete cDNA set into appropriate vectors for proteome and ribonome studies is a high priority. Such studies may include analysis of protein-protein, DNA-protein and RNA-protein interactions. In addition to these studies, the complete cDNA set could be used as a tool for the production of antibodies against Drosophila proteins. Well-characterized cDNAs, which have been corrected for amplification-mediated mutations, need to be placed in vectors that can be manipulated for various proteomics applications. This would allow these tools to be efficiently produced and made available to the community at reasonable costs.

7) Annotation of genome sequence from additional Drosophila species. Thanks to four separate National Human Genome Research Institute (NHGRI) funded initiatives, the sequence of 11 additional species of Drosophila is well underway and assemblies should be available soon. These new data present an unparalleled opportunity for rapid progress in a range of areas including (1) using comparative sequence analysis to improve the annotations of *D. melanogaster*, (2) understanding genome evolution including the functional evolution of genetic pathways, (3) describing variation at a genome scale, (4) identifying non-coding genes and regulatory elements, and (5) investigating differences between recently diverged species that produce interfertile hybrids. To fully realize the potential of this unique resource, continuing support is needed for assembling, aligning and annotating these genomes. In addition, projects aimed at sequencing EST's and cDNA clones for selected species will be invaluable for refining annotations.

8) Completion of the mapping, sequencing, and annotation of *Drosophila melanogaster* heterochromatin. The difficulty of analyzing heterochromatin remains the major roadblock toward the completion of genome projects in most multicellular organisms. Mapping, sequencing, and annotation of heterochromatin is essential for genome-wide analyses, such as mapping the distributions of transcription factors and chromatin components, non-protein coding RNAs, and RNAi-mediated gene disruption screens. In addition, elucidating heterochromatin organization is key to understanding the epigenetic regulation gene expression, with immediate implications in developmental biology and medicine. Important information about the composition and organization of Drosophila heterochromatin has been generated through detailed annotation of existing sequences, including the demonstration that ~3% of all Drosophila protein-coding genes reside in heterochromatin. However, much of the existing sequence is unmapped and unfinished, and reliable annotations require more complete information. The NHGRI has generously funded the Drosophila Heterochromatin Genome Project (DHGP), and we encourage continued support for this project as well as other investigations of heterochromatin sequence and function.

Below we categorize additional high-priority needs of the community that may be best met by R01, investigator-initiated efforts or pilot grants, rather than by large project grants.

- Development of new methodologies that broaden the scope of the use of RNAi in Drosophila cells and whole animals. In particular, the application of RNAi to primary tissue culture cells will facilitate the design of novel cell-based assays that reflect complex in vivo biological processes (i.e., axonal outgrowth, muscle differentiation). In addition, any technological advances that aim to improve the efficiency and miniaturization of the high-throughput RNAi approach (i.e., dsRNA chips) are clearly needed in order to improve the reliability and affordability of the current technology. Improvement of methods to deliver RNAi to whole animals, especially embryos, is also needed. Finally the distribution of validated resources for RNAi screening will greatly expand access to this technology.

- Development of new cell lines and molecular characterization of existing cell lines. Cell lines have found increasing use in Drosophila research, but only a limited number of Drosophila cell lines are available. In particular, there is a need for tissue-specific cell lines that could be used in RNAi screens (for example epithelial cells to screen for genes involved in epithelial cell polarity), and for cell-cell interaction studies (i.e. cell lines that fail to express a certain signaling pathway). Having access to a diverse set of cell lines should facilitate the biochemical purification and analysis of molecular complexes and would complement whole organism approaches.

- Establishment of molecularly defined genomic duplications. The recent availability of molecularly defined deletions has been a major leap forward. In addition, it would be extremely useful to have a set of molecularly defined duplications for the entire genome. Duplications of defined chromosomal intervals are important in mapping genes, identifying molecular lesions, and assaying gene dosage effects. The X chromosome is of the highest priority since duplications will make it possible to carry out complementation analysis of X-linked essential genes. Large segments of genomic DNA flanked by FRT sites can be integrated into the genome. This

approach has the added advantage that genomic DNA inserts can also be excised by Flipase in specific cells, thus permitting the study of gene loss in either pre- or post-mitotic cells (including in adults) in an otherwise wild-type animal.

- Development of methods to understand the evolution of gene function. It is important to understand the functional evolution of genetic pathways, not just sequence evolution. This requires support to develop tools for the sequenced non-*melanogaster* species such as gene replacement and transformation that have been successfully used in *D. melanogaster*.

- Generation of a well-characterized collection of conditional (ts lethal) mutants. Such a collection would be of real value to the Drosophila community for several reasons. First, the majority of available lethal mutations are embryonic lethal; and thus, studying post-embryonic development using these mutants is extremely difficult. Second, even for those lethals that do die later in development, potential embryonic defects can be masked by stores of protein or RNA deposited into the oocyte by heterozygous mothers. In such cases, it is necessary to make germline clones, however, if the protein is also required for germline development, such eggs may stop developing or they may be disorganized and therefore difficult to analyze. Third, even in the best cases, conditional mutants are required to determine the precise temporal requirement of a gene product. One of the best ways to address all of the above limitations and move the field forward is to isolate conditional mutants in as many genes as possible.

- An efficient means of cryopreservation of Drosophila at any stage of development. This has long been a high priority for Drosophila researchers. A successful cryopreservation procedure would reduce the stress on the stock centers, ensure that valuable genetic resources are not lost and could curtail costs involved in running fly kitchens, and constantly maintaining laboratory stocks in all Drosophila labs.