Final Report

DE-FG03-99ER62739: Resources for Biological Annotation of the Drosophila Genome

This project supported seed money for the development of cDNA and genetic resources to support studies of the Drosophila melanogaster genome. The project was co-funded by the NIH NGHRI through grant P50 HG0750 (Gerald M. Rubin, PI). The project was highly successful in not only in contributing significantly to the development of these resources during the course of the grant, but also in providing the foundation for the expansion of these resources which is still ongoing, supported by grants from the NIH's NIGMS (for the genetic resources) and NHGRI (for the cDNA resources).

Key publications supported by this work that provide additional detail:

1, The Drosophila gene collection: identification of putative full-length cDNAs for 70% of D. melanogaster genes. Stapleton M, Liao G, Brokstein P, Hong L, Carninci P, Shiraki T, Hayashizaki Y, Champe M, Pacleb J, Wan K, Yu C, Carlson J, George R, Celniker S, Rubin GM. Genome Res. 2002 Aug; 12(8):1294-300.

Collections of full-length nonredundant cDNA clones are critical reagents for functional genomics. The first step toward these resources is the generation and single-pass sequencing of cDNA libraries that contain a high proportion of full-length clones. The first release of the Drosophila Gene Collection Release 1 (DGCr1) was produced from six libraries representing various tissues, developmental stages, and the cultured S2 cell line. Nearly 80,000 random 5' expressed sequence tags (5' expressed sequence tags [ESTs]from these libraries were collapsed into a nonredundant set of 5849 cDNAs, corresponding to ~40% of the 13,474 predicted genes in Drosophila. To obtain cDNA clones representing the remaining genes, we have generated an additional 157,835 5' ESTs from two previously existing and three new libraries. One new library is derived from adult testis, a tissue we previously did not exploit for gene discovery; two new cap-trapped normalized libraries are derived from 0-22-h embryos and adult heads. Taking advantage of the annotated D. melanogaster genome sequence, we clustered the ESTs by aligning them to the genome. Clusters that overlap genes not already represented by cDNA clones in the DGCr1 were analyzed further, and putative full-length clones were selected for inclusion in the new DGC. This second release of the DGC (DGCr2) contains 5061 additional clones, extending the collection to 10,910 cDNAs representing >70% of the predicted genes in Drosophila.

2. The Berkeley Drosophila Genome Project gene disruption project: Single P-element insertions mutating 25% of vital Drosophila genes. Spradling AC, Stern D, Beaton A, Rhem EJ, Laverty T, Mozden N, Misra S, Rubin GM. Genetics. 1999 Sep; 153(1):135-77.

A fundamental goal of genetics and functional genomics is to identify and mutate every gene in model organisms such as Drosophila melanogaster. The Berkeley Drosophila Genome Project (BDGP) gene disruption project generates single P-element insertion strains that each mutate unique genomic open reading frames. Such strains strongly facilitate further genetic and molecular studies of the disrupted loci, but it has remained unclear if P elements can be used to mutate all Drosophila genes. We now report that the primary collection has grown to contain 1045 strains that disrupt more than 25% of the estimated 3600 Drosophila genes that are

essential for adult viability. Of these P insertions, 67% have been verified by genetic tests to cause the associated recessive mutant phenotypes, and the validity of most of the remaining lines is predicted on statistical grounds. Sequences flanking >920 insertions have been determined to exactly position them in the genome and to identify 376 potentially affected transcripts from collections of EST sequences. Strains in the BDGP collection are available from the Bloomington Stock Center and have already assisted the research community in characterizing >250 Drosophila genes. The likely identity of 131 additional genes in the collection is reported here. Our results show that Drosophila genes have a wide range of sensitivity to inactivation by P elements, and provide a rationale for greatly expanding the BDGP primary collection based entirely on insertion site sequencing. We predict that this approach can bring >85% of all Drosophila open reading frames under experimental control.