Title: **Constructing Data-Driven Fate-maps for Drosophila Melanogaster Imaginal discs via High-throughput Imaging Methods**

Parvez Ahammad[1], Cyrus L. Harmon, Ann Hammonds[3], Susan E. Celniker[3], S. Shankar Sastry[1], and Gerald M. Rubin[2]

1 Department of Electrical Engineering and Computer Science, University of California, Berkeley, Berkeley, CA.
2 Howard Hughes Medical Institute, Ashburn, VA.
3 Department of Genome Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA.

Many genes are known to have interesting non-trivial spatial expression patterns in the imaginal discs of dipteral organisms. Imaginal discs are the primordial tissues that generate important parts of the adult insect exoskeleton, such as the eyes, wings, and legs. Intricate patterns of regulation of gene expression, in both time and space, are used to pattern the Imaginal disc and to guide cell fate decisions. We present high-throughput methods and tools for data-driven generation and comparative analysis of large numbers of spatial patterns of gene expression in the imaginal discs of Drosophila melanogaster.

Existing methods are limited in either resolution (as in the case of microarrays, which while able to operate on large numbers of genes provide little information about where genes are expressed), or in throughput (as in the case of in situ hybridization, which yields precise spatial information, but only for a single gene). To characterize and explore the spatial patterns of a large number of genes with previously uncharacterized patterns, we have developed methods for the automated identification, representation and characterization of these patterns using a novel semi-supervised joint pattern alignment approach. We have adopted the high-throughput probe generation and staining protocol used in the generation of gene expression patterns in Drosophila embryos to work with mass-isolated third instar larval imaginal discs and have generated patterns for over 130 different genes. We manually segmented a small number of shapes from background for each imaginal disc class and used a joint alignment procedure to automatically learn the canonical shape of distinct classes of imaginal discs from the data. These shapes were then used as the models for a parametric alignment procedure that automatically extracts and aligns an imaginal disc shape from an image. Using these tools, we have developed an integrated, automated pipeline for analyzing gene expression data to select genes for spatial expression analysis, processing image data to learn imaginal disc shapes, to automatically extract instances of these learned shapes in images, to determine the spatial expression pattern of these genes in the registered images and to compare and classify these genes based on their global gene expression profile in Drosophila development and their spatial expression pattern in imaginal discs. Further, we used an unsupervised approach based on hierarchical agglomerative clustering procedure to cluster the tissue classes in the imaginal discs based on their gene expression profile in order to generate consistent data-driven fate-maps of the imaginal discs.