Genomic organization and single-nucleotide polymorphism map of desmuslin, a novel intermediate filament protein on chromosome 15q26.3

ABSTRACT

The absence of mutations in the desmuslin gene did not impact its function. Nevertheless, the single-nucleotide polymorphisms mapped in this study are highly disequilibrated and can be used for disambiguation studies of this region of chromosome 15q26.3.

INTRODUCTION

Drosophila melanogaster is a significant model organism. Its anatomy has undergone extensive research, and many brain functions have been linked to specific substructures. The adult brain is composed of around 200,000 neurons that are organized into distinct substructives. Among the other regions, the optic lobe is responsible for processing visual information from the photoreceptors and sending that information to the central brain. The antennal limbs are responsible also for handling smelly information, while the mushroom bodies are involved in learning and memory and other complex behaviors. Around six neurons in the lateral protocerebrum are responsible for driving circadian rhythms in locomoter activity. The central complex is not fully understood, but it appears to be involved in motor coordination as well. While we are becoming more familiar with the anatomy and function of Drosophila brains, we have limited knowledge about the specific brain molecules that regulate function and behavior. Traditional methods for identifying genes involved in neural function include behavioral screening of mutagenized flies, then rescreening candidate lines for potential pleiotropic effects caused by developmental defects, which are time-consuming and laborious. However, sequencing random cDNAs has emerged to identify genes expressed in a specific cell type, providing an alternative method to address this issue. The collection of information for complex mRNA populations has been facilitated by the analysis of expressed sequence tags (ESTs). This approach utilizes sequence information from one or both ends of a cDNA to catalog the complexity of an resulting population. By using modified EST, we can sequence entirely novel cDNAs and use shotgun sequencing to identify concatenated transcripts expressed in the brain. These transcripts are expected to contain more clones that are involved in neuronal function due to the location of expression. Many Drosophila head libraries have been used to isolate cDNAs that correspond to genes identified by genetic screens for their involvement in brain function. A number of the transcripts expressed at a relatively low level in this manner (dunce, CREB, Dco, period, timeless, dissonance) are not present in high quantity; only the Drosophila brain makes up around 14% head tissue (dry weight). By eliminating non-brain tissues, we increase the relative representation of rare neural transcript types in our unique library. By cataloguing the genes expressed in the brain of adult Drosophila, we were able to advance conventional methods of understanding brain function. This has led to an improvement in molecular science by storing sequence information and publishing the data through electronic databases. With this set of tools, it is possible to identify a gene that was sequenced by another laboratory in just ten minutes and potentially deduce the function of the isolated clone in any manner relevant to virtually every branch of biological sciences. The report outlines the creation, quality evaluation, and initial testing of a novel library derived from adult Drosophila brains. It also highlights that 29 clones chosen for analysis in the random selection process do not match any EST sequence generated in support of the Droseophilus genome project (as of

10 October 2000). In addition, the genes that make up 59% of the novel ESTs are not predicted by fly genome annotation algorithms. Based on our analysis of a sample of 13,601 rephrased genes already annotated, we estimate that the number of genes in the Drosophila genome may be underestimated by 10-15% (approximately 1,300 to 2,000 genes).

CONCLUSION

The tree harvest procedure is a useful and efficient method of supervised learning from gene expression data. Its primary objective is to identify the additive and interaction structure of clusters of genes, in relation to an outcome measure. However, this technique requires significant samples for successful analysis and any other procedures with similar objectives. This method had some flaws as there were not enough samples for the actual data, so we decided to try tree harvesting on ever-larger gene expression datasets when they are available; this time, we proceeded with a stepwise approach whereby we used the sum and products of the average gene Expression of chosen clusters—then selecting models that are interpretable and biologically plausible. The gene expression of clusters' average could be used to construct alternative models, such as tree-based models or boosting methods (as demonstrated by Friedman et al.). Additional data: Additional information about clustering from the lymphoma data-driven harvest model and other available data can be found in the online version of this article.