A Novel Characterization of Canalizing Genes
H. Ortiz-Zuazaga1, R. Chiesa2, M. Guerrero2, A. Vazquez2, I. Chévere2, S. Peña de Ortiz2, O. Moreno3
1HighPerformance Computing Facility
2Department of Biology, Rio Piedras Campus
3Department of Computer Science, Rio Piedras Campus
University of Puerto Rico

Abstract

lDNA microarrays are a technique for measuring the abundance of RNA from many thousands of genes simultaneously in an inexpensive experiment. The analysis of microarrays for clustering and error correction of microarray expression measurements. The techniques developed in close collaboration between computer scientists and molecular biologists, allowing a natural description of genes, and permitted the identification of candidate genes that had been unnoticed in earlier analyses. The techniques use block-bounding and majority logic device to discretize continuum expression measurements into a set of integer values or with a finite field. This concept of analyzing expression data, such as a network reverse engineering. We develop a novel statistical characterization of canalizing functions and the concept of analyzing Boolean functions. We have applied the block-bounding method to a data set from a behavioral training experiment on rats, conditioned taste aversion. An error correction procedure produces 127 consistent genes divided into 25 clusters. We focused on a group of genes expressing in the CTA group, which is known to be required for expression throughout the experiment. Within this group, we look for genes known to have a canalizing function in CTA and found two such genes: Peh1 and CalcB. A cycle nonneuronal that induces hippocampal sprouting. CalcB is principally a vandinostat, but seems to have a role in basal regulation or synaptic pruning. The analysis resulted in the discovery of two genes believed to be involved in learning and memory. The significance supports the idea that the hippocampus plays a major role in these processes. These genes have been identified in a preclinical animal that had conditional essential tests for differential expression and clustering. This close collaboration between our research groups leads to opportunities and insights that would go unnoticed by either group working alone.

Acknowledgments

This work supported in part by grants [CORE1-F31HD015853, RO1-NS48010] and [20-1044].

Methods

The gene profiling experiment was replicated five times. Each animal was used per condition for each replicate. Thus, a total of twenty rats were used per condition. Animals were sacrificed by decapitation at 1, 3, 6, and 24 hours after conditioning. The CTA data set has two controls, the pre-treatment group and the one hour value group, and four time points, 1, 3, 6, and 24 hours after conditioning. Both groups had 1000 genes, and we focused on the first two points.

We have developed an error correction and clustering procedure, which we illustrate with a gene from the CTA data set:

Alta glioma 1.3; HD2-MU: nervous system cell-surface heparin sulfate proteoglycan (1.3).

The data set consists of 127 consistent genes divided into 25 clusters. We focused on a group of genes expressing in the CTA group, which is known to be required for expression throughout the experiment. Within this group, we look for genes known to have a canalizing function in CTA and found two such genes: Peh1 and CalcB. A cycle nonneuronal that induces hippocampal sprouting. CalcB is principally a vandinostat, but seems to have a role in basal regulation or synaptic pruning. The analysis resulted in the discovery of two genes believed to be involved in learning and memory. The significance supports the idea that the hippocampus plays a major role in these processes. These genes have been identified in a preclinical animal that had conditional essential tests for differential expression and clustering. This close collaboration between our research groups leads to opportunities and insights that would go unnoticed by either group working alone.

Results

We have performed the analysis described above on the CTA data set. In this data set, there are 127 consistent genes that we divide into clusters by grouping together the genes that have the same calls of 1 in the 3- to 6-hour timepoints. Thus, there are 25 clusters.

In Dr. Sandra Palón’s lab they study the role of CRB in CTA gene known to be required for long-term memory acquisition. We have focused on the expression of this gene, and other genes with the same pattern of expression. CRB is a DNA element known to bind to a DNA element called CRB-responsive element (CREB) in the target genes, and in conjunction with a co-activator initiates transcription of the target genes.

We focused on the cluster labeled as “CTA.” The confusion of the calls for these genes represents an average over the 3, 1, and 6 hour timepoints, followed by correction at the 24 hour timepoint. This cluster consists of genes whose expression changes most clearly reflect the expression profile of CRB. We investigated the genes in this cluster in depth, removing the false positives that arise due to the baseline Genome Version 32.

From Knebel we obtained genomic sequence for each gene. We sequence the last 1000 bases of the transcription factor binding sites. We look for the CRB element, a DNA sequence that is the target site for CREB in the upstream region of interest to CRB.

Two genes in particular caught our interest: Peh1 and CalcB. Both genes have CREB elements in their upstream regions. According to the Rat Genome Database, Peh1 is a cycle nonneuronal that induces hippocampal sprouting. CalcB is principally a vandinostat, but seems to have a role in basal regulation or synaptic pruning. Thus these genes exhibit a pattern of expression consistent with the expression of CTA. Both have CREB elements upstream of their transcription start site, and seem to have a role in strengthening or maintaining new synapses. Thus they are strongly implicated as important genes for the formation of memories. Our collaborator, Dr. Sandra Palón’s lab, has previously shown that CRB activity is linked to the expression of CREB in the hippocampus. We will confirm the changes in expression of those genes and investigate their role in memory.

Conclusions

- Interaction between biologists and computer scientists allows us to analyze data that incorporate biological knowledge to be developed.
- Our contribution is to discretize clustering and error correction in a natural description of genes, and permit the identification of candidate genes that were unnoticed in earlier analyses.
- Canalizing Boolean Functions:

- Canalizing Boolean Functions are the ones that are important for regulatory networks.
- A Boolean function is called canalizing if it stabilizes in its inputs if there exist - and such that for all .

A Boolean function is called canalizing if it stabilizes in its inputs if there exist - and such that for all .

If a Boolean function is called canalizing then it is called canalizing if it stabilizes in its inputs if there exist - and such that for all .

If a Boolean function is called canalizing then it is called canalizing if it stabilizes in its inputs if there exist - and such that for all .

A Boolean function is called canalizing if it stabilizes in its inputs if there exist - and such that for all .

A Boolean function is called canalizing if it stabilizes in its inputs if there exist - and such that for all .