Analysis Techniques for Probe-level Microarray Data

Humberto Ortiz Zuazaga

University of Puerto Rico
High Performance Computing facility

August 20, 2009
Analysis Techniques for Probe-level Microarray Data

Humberto Ortiz Zuazaga

General Introduction

Error Correction and Clustering

Introduction

Methods

Results

Conclusions

Future work

Motivation

Methods

Preliminary results

or, “I’ve got all the genes, now what do I do with them?”
1. General Introduction
2. Error Correction and Clustering
   - Introduction
   - Methods
   - Results
   - Conclusions
3. Probe-level data
   - Introduction
   - Methods
   - Results
   - Conclusions
4. Future work
   - Motivation
   - Methods
   - Preliminary results
   - Summary
Conditioned taste aversion (CTA)

- associative aversive conditioning paradigm
- Animals are exposed to a novel taste, the *conditioned stimulus*
- An *unconditioned stimulus* induces malaise
- The animals develop a long lasting aversion to the conditioned stimulus
CTA Dataset

- two controls, the pre-treatment group and the one hour saline group
- four time points, 1, 3, 6, and 24 hours after conditioning
- 1185 genes on each spotted array
- 5 biological replicates of each array

cAMP Responsive Element Binding protein (CREB)

■ Transcription factor, known to regulate gene expression
■ Plays important role in memory formation
■ Binds to a DNA element called cAMP-response element (CRE)
■ We will use this gene as an example of a canalyzing gene

Outline

1. General Introduction
2. Error Correction and Clustering
   - Introduction
   - Methods
   - Results
   - Conclusions
3. Probe-level data
   - Introduction
   - Methods
   - Results
   - Conclusions
4. Future work
   - Motivation
   - Methods
   - Preliminary results
   - Summary
### Discretization

**A01a glypican 1; HSPG M12; nervous system cell-surface heparan sulfate proteoglycan**

<table>
<thead>
<tr>
<th>Repetition</th>
<th>Pre</th>
<th>Sal</th>
<th>1 h</th>
<th>3 h</th>
<th>6 h</th>
<th>24h</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.172</td>
<td>0.099</td>
<td>0.176</td>
<td>0.142</td>
<td>0.062</td>
<td>0.152</td>
</tr>
<tr>
<td>2</td>
<td>0.274</td>
<td>0.168</td>
<td>0.126</td>
<td>0.114</td>
<td>0.104</td>
<td>0.276</td>
</tr>
<tr>
<td>3</td>
<td>0.003</td>
<td>0.119</td>
<td>0.552</td>
<td>0.178</td>
<td>0.193</td>
<td>0.114</td>
</tr>
<tr>
<td>4</td>
<td>0.114</td>
<td>0.139</td>
<td>0.6</td>
<td>0.311</td>
<td>0.179</td>
<td>0.181</td>
</tr>
<tr>
<td>5</td>
<td>0.04</td>
<td>0.006</td>
<td>0.172</td>
<td>0.103</td>
<td>0.036</td>
<td>-0.047</td>
</tr>
<tr>
<td>average</td>
<td>0.121</td>
<td>0.106</td>
<td>0.325</td>
<td>0.17</td>
<td>0.115</td>
<td>0.135</td>
</tr>
<tr>
<td>control</td>
<td>0.113</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>diffs</td>
<td></td>
<td></td>
<td>0.212</td>
<td>0.057</td>
<td>0.002</td>
<td>0.022</td>
</tr>
<tr>
<td>epsilon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.022</td>
</tr>
<tr>
<td>calls</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

**Future work**

- Motivation
- Methods
- Preliminary results
### Majority logic

<table>
<thead>
<tr>
<th>Repetition</th>
<th>1 h</th>
<th>3 h</th>
<th>6 h</th>
<th>24h</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+</td>
<td>0</td>
<td>−</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>4</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>5</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>−</td>
</tr>
</tbody>
</table>

- **consensus**: + + ? +
Substituting averaged controls

<table>
<thead>
<tr>
<th>Repetition</th>
<th>1 h</th>
<th>3 h</th>
<th>6 h</th>
<th>24h</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>5</td>
<td>+</td>
<td>0</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>cvac</td>
<td>+</td>
<td>+</td>
<td>?</td>
<td>+</td>
</tr>
</tbody>
</table>
## Pruning extreme values

<table>
<thead>
<tr>
<th>Repetition</th>
<th>Pre</th>
<th>Sal</th>
<th>1 h</th>
<th>3 h</th>
<th>6 h</th>
<th>24h</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>—</td>
<td>0.099</td>
<td>0.176</td>
<td>0.142</td>
<td>—</td>
<td>0.152</td>
</tr>
<tr>
<td>2</td>
<td>—</td>
<td></td>
<td>0.126</td>
<td>0.114</td>
<td>0.104</td>
<td>—</td>
</tr>
<tr>
<td>3</td>
<td>0.003</td>
<td>0.119</td>
<td>—</td>
<td>—</td>
<td>0.193</td>
<td>0.114</td>
</tr>
<tr>
<td>4</td>
<td>0.114</td>
<td>0.139</td>
<td>—</td>
<td>—</td>
<td>0.179</td>
<td>0.181</td>
</tr>
<tr>
<td>5</td>
<td>0.04</td>
<td></td>
<td>0.172</td>
<td>0.103</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>new average</td>
<td>0.052</td>
<td>0.119</td>
<td>0.158</td>
<td>0.12</td>
<td>0.159</td>
<td>0.149</td>
</tr>
<tr>
<td>new control</td>
<td></td>
<td>0.086</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>new diffs</td>
<td></td>
<td></td>
<td>0.072</td>
<td>0.034</td>
<td>0.073</td>
<td>0.063</td>
</tr>
<tr>
<td>new epsilon</td>
<td>0.063</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>new calls</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
</tbody>
</table>

Motivation
Methods
Preliminary results
Future work
Consistent calls

1. at least two of the above set of calls agrees in the last 4 columns of data (1 h, 3 h, 6 h, and 24h)

2. either the 1 h or the 24 h columns is a “0”

3. across the last 4 columns of data, the column exhibits the consecutive zeros property (i.e., values do not oscillate between “0” and “+” or “−”)
A01a is not consistent

<table>
<thead>
<tr>
<th></th>
<th>1 h</th>
<th>3 h</th>
<th>6 h</th>
<th>24h</th>
</tr>
</thead>
<tbody>
<tr>
<td>average calls</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>consensus</td>
<td>+</td>
<td>+</td>
<td>?</td>
<td>+</td>
</tr>
<tr>
<td>cvac</td>
<td>+</td>
<td>+</td>
<td>?</td>
<td>+</td>
</tr>
<tr>
<td>new calls</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>0</td>
</tr>
</tbody>
</table>
Clustering

- Categorizing each timepoint for each gene into coarse divisions yields a clustering of genes.
- In our current experiment there are $3^4 = 81$ possible clusters that a gene may fall into.
- Longer time series or larger fields will allow finer grained division of the genes into clusters.
CTA Dataset

- 127 consistent genes in CTA dataset
- Grouping genes with same calls in 1 h – 24 h timepoints yields 23 clusters
Clusters

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Gene coordinate</th>
</tr>
</thead>
<tbody>
<tr>
<td>- 0 0 0</td>
<td>A05f, A12k, A14g, B07m, B07n</td>
</tr>
<tr>
<td>0 + + +</td>
<td>B01n, B04i, B05l, B06j, B06l, C10l, D01j, D10l, E01k, E03j, E09c, E10e, E13i, E13l, F01e</td>
</tr>
<tr>
<td>+ 0 0 0</td>
<td>B07e, D04f, D09l</td>
</tr>
<tr>
<td>- - 0 0</td>
<td>A02m, A14e, B03i, C08i, F08c, F13e</td>
</tr>
<tr>
<td>- - + 0</td>
<td>C14k, E03l</td>
</tr>
<tr>
<td>0 0 + +</td>
<td>B13n, D14g, E06i, E10m, E11l, E13e, E14d</td>
</tr>
<tr>
<td>- + + 0</td>
<td>B06m, E02i</td>
</tr>
<tr>
<td>0 0 - -</td>
<td>A05n, D02a, F12e</td>
</tr>
<tr>
<td>0 - + +</td>
<td>B04j, E10l, F11j</td>
</tr>
<tr>
<td>+ - - 0</td>
<td>C01f, F01a</td>
</tr>
<tr>
<td>0 - + -</td>
<td>A11d, C09m</td>
</tr>
<tr>
<td>- + 0</td>
<td>D12a</td>
</tr>
<tr>
<td>0 + - +</td>
<td>D09i</td>
</tr>
<tr>
<td>- - 0</td>
<td>A02l, A03h, A09c, B10l, C02m, C04d, C04f, C06e, D02b, F02b, F03c, F09n, F11e</td>
</tr>
<tr>
<td>+ - + 0</td>
<td>B13a, F02a</td>
</tr>
<tr>
<td>0 0 0 -</td>
<td>A10l, C08a, C14g, C14l, D13e</td>
</tr>
<tr>
<td>+ + 0 0</td>
<td>A08l, B08b, C08j, F11k</td>
</tr>
<tr>
<td>+ + - 0</td>
<td>D04l</td>
</tr>
<tr>
<td>0 0 0 +</td>
<td>A07i, B09h, C10c, D08n, E03m, E04i, E13h, E14f</td>
</tr>
<tr>
<td>0 - - +</td>
<td>C01e, F12j</td>
</tr>
<tr>
<td>0 - -</td>
<td>B05b, C13a, C13i, C14n, E02e, F04l, F06a, F11l</td>
</tr>
<tr>
<td>- + 0 0</td>
<td>A12m</td>
</tr>
<tr>
<td>+ + 0 0</td>
<td>B01i, B07f, B10c, B10d, B14h, D08f, D09e, E03i, E14k, F02n, F05k, G15</td>
</tr>
</tbody>
</table>
Genes equivalent to canalyzing genes

- A canalyzing gene is one that is required for proper function of the organism
- We hypothesize that the canalyzing gene does not directly affect the organism, but acts through canalyzed genes
- Our clustering algorithm groups canalyzed and canalyzing genes together
Genes equivalent to CREB

- We search for genes with similar patterns of expression to CREB
- Obtained upstream sequences for “000+” cluster (1020 bp, 800 bp before start of transcription) expression most similar to CREB

<table>
<thead>
<tr>
<th>Coord.</th>
<th>Accession</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A07i</td>
<td>L24388</td>
<td>galactosyltransferase-associated protein kinase (GTA); CDC2-related protein kinase (CDC2L1)</td>
</tr>
<tr>
<td>B09h</td>
<td>L10362</td>
<td>synaptic vesicle protein 2B</td>
</tr>
<tr>
<td>C10c</td>
<td>L33869</td>
<td>ceruloplasmin (CERP; CP); ferroxidase</td>
</tr>
<tr>
<td>D08n</td>
<td>X63255</td>
<td>N-methyl-D-aspartate receptor subtype 1 (NMDAR1; NR1); glutamate receptor subunit zeta 1 (GRIN1)</td>
</tr>
<tr>
<td>E03m</td>
<td>M29712</td>
<td>melanin-concentrating hormone (PMCH; MCH)</td>
</tr>
<tr>
<td>E04i</td>
<td>V01228</td>
<td>calcitonin</td>
</tr>
<tr>
<td>E13h</td>
<td>M20713</td>
<td>guanine nucleotide-binding protein G(K) alpha 3 subunit (G(I) alpha 3 (GNAI3))</td>
</tr>
<tr>
<td>E14f</td>
<td>X06890</td>
<td>ras-related protein RAB4A</td>
</tr>
</tbody>
</table>
Transcription factor sites

- Searched sequences for transcription factor binding sites with TESS [1]
- Found two very interesting genes: Pmch and Calca, both have CRE sites

Cyclic neuropeptide
Affects appetite or metabolism
Induces hippocampal synaptic transmission

Calca

- Vasodilator
- May be involved in axonal regeneration
- May be involved in synaptogenesis

We have developed an error correction procedure for microarray data.

Each gene is described as upregulated, unchanged or downregulated, corresponding to a natural division of expression values.

Additional heuristics reduce the set of genes being examined to a small group of candidate genes.

The procedure results in genes being grouped into discrete clusters.

We found two candidate genes, Pmch and Calca.

Multiple sources of evidence support the theory that these genes are involved in biological processes related to memory.
Acknowledgments

Portions of this work were supported by a SCORE grant (S06GM08102), an INBRE grant (P20RR016470) and an IDeA Program grant (P20RR15565), from the National Institutes of Health.

Data provided by Dr. Sandra Peña de Ortiz.
Microarray feature density continues to increase
- Oligo arrays have shorter probe sequences (cost)
- Many arrays utilize multiple probes per gene or feature
- probe sets are summarized to obtain a single value
Affymetrix array image
Experiment designed to test microarray analysis methods

Two separate microarray technologies (Affymetrix, NimbleGen)

Time course data, 0, 6 hours and 24 hours after training

Massed and Spaced training

Affymetrix: 14 probes per probeset, 14,000 probesets

NimbleGen: about 10 probes per probeset, 12,000 probesets
Outline

1. General Introduction
2. Error Correction and Clustering
   - Introduction
   - Methods
   - Results
   - Conclusions
3. Probe-level data
   - Introduction
   - Methods
   - Results
   - Conclusions
4. Future work
   - Motivation
   - Methods
   - Preliminary results
   - Summary
Discretization

- Very similar to methods in prior chapter
- Take the mean over all repetitions
- Take average of controls
- Select discretization threshold $\epsilon$ so that one of the remaining points is within $\epsilon$ of the control
- perform consensus, consensus vs mean, mean and trimmed mean analyses
- will define a new measure of “consistent” genes
Majority Logic Decoding Probes

```
[['0', '-', '0', '+'],
 ['0', '-', '+', '+'],
 ['0', '-', '0', '+'],
 ['0', '-', '+', '+'],
 ['0', '-', '+', '+'],
 ['-', '-', '0', '0'],
 ['0', '-', '0', '+'],
 ['0', '-', '0', '+'],
 ['-', '-', '-', '0'],
 ['0', '-', '0', '+'],
 ['0', '-', '0', '+'],
 ['0', '-', '+', '+']]

'mld': ['0', '-', '0', '+']
```
Weighted Mutual Information

We want to measure concordance and informativeness of probes.

\[ I(X, Y) = \sum_{y \in Y} \sum_{x \in X} w(x, y)p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right) \]
Weights

We set the weights such that similar patterns of expression are given higher weights, and opposite expression is given lower weight:

\[
    w(x, y) = \begin{cases} 
    1.0 & \text{if } x = y, \ x, y \neq ? \\
    0.5 & \text{if } x = ? \text{ or } y = ? \\
    0.1 & \text{otherwise}
    \end{cases}
\]
Example WMI

\[ ac = ['0', '-', '0', '+'] \]
\[ nc = ['0', '-', '-', '-'] \]

\[
\begin{align*}
I(ac, nc) &= w(-, -)p(-, -) \log \left( \frac{p(-, -)}{p(-)p(-)} \right) + \\
&\quad w(0, -)p(0, -) \log \left( \frac{p(0, -)}{p(0)p(-)} \right) + \\
&\quad w(+, -)p(+, -) \log \left( \frac{p(+, -)}{p(+)p(-)} \right) + \\
&\quad w(0, 0)p(0, 0) \log \left( \frac{p(0, 0)}{p(0)p(0)} \right) \\
&= 1 \cdot 1/4 \cdot \log \left( \frac{1/4}{1/4 \cdot 3/4} \right) + 0.1 \cdot 1/4 \cdot \log \left( \frac{1/4}{2/4 \cdot 3/4} \right) + \\
&\quad 0.1 \cdot 1/4 \cdot \log \left( \frac{1/4}{1/4 \cdot 3/4} \right) + 1 \cdot 1/4 \cdot \log \left( \frac{1/4}{2/4 \cdot 1/4} \right) = 0.35
\end{align*}
\]
SWAMI Score

- One Affymetrix probeset may map to several NimbleGen probesets
- Average WMI over all equivalent NimbleGen probesets
- Sum all average WMI scores over all Affymetrix probesets
- Obtain Summed Weighted Averaged Mutual Information or SWAMI score
- A single numerical value characterizing an analysis method
- Replaces consistency test from previous chapter
### SWAMI Scores for several transformation and regression methods

<table>
<thead>
<tr>
<th>Transformation</th>
<th>Regression</th>
<th>SWAMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>log2</td>
<td>Huber</td>
<td>182</td>
</tr>
<tr>
<td>log2</td>
<td>fair</td>
<td>186</td>
</tr>
<tr>
<td>log2</td>
<td>Cauchy</td>
<td>169</td>
</tr>
<tr>
<td>sqrt</td>
<td>Huber</td>
<td>212</td>
</tr>
<tr>
<td>sqrt</td>
<td>fair</td>
<td>230</td>
</tr>
<tr>
<td>sqrt</td>
<td>Cauchy</td>
<td>200</td>
</tr>
<tr>
<td>cuberoot</td>
<td>Huber</td>
<td>207</td>
</tr>
<tr>
<td>cuberoot</td>
<td>fair</td>
<td>216</td>
</tr>
<tr>
<td>cuberoot</td>
<td>Cauchy</td>
<td>202</td>
</tr>
</tbody>
</table>
SWAMI scores for error correction methods

<table>
<thead>
<tr>
<th>Method</th>
<th>summarized</th>
<th>probes</th>
</tr>
</thead>
<tbody>
<tr>
<td>trimmed mean</td>
<td>2753</td>
<td>3657</td>
</tr>
<tr>
<td>mean</td>
<td>2610</td>
<td>2535</td>
</tr>
<tr>
<td>consensus</td>
<td>1920</td>
<td>3058</td>
</tr>
<tr>
<td>consensus vs mean control</td>
<td>1888</td>
<td>1525</td>
</tr>
</tbody>
</table>
We have designed a technique to validate microarray analysis methods.

The technique measures concordance between two data sets derived from different microarray technologies on the same experimental conditions.

We have performed the experiment using data from an odor avoidance training time course on fruit flies.

The results demonstrate our prior error correction techniques produce the most concordance between the data sets in this experiment.
Acknowledgments

This work was supported in part by a SCORE grant (S06GM08102) and an INBRE grant (P20RR016470) from the National Institutes of Health. The data was kindly provided by Dart Neuroscience LLC.
Outline

1. General Introduction
2. Error Correction and Clustering
   - Introduction
   - Methods
   - Results
   - Conclusions
3. Probe-level data
   - Introduction
   - Methods
   - Results
   - Conclusions
4. Future work
   - Motivation
   - Methods
   - Preliminary results
   - Summary
“Moore’s law” for microarray features

Feature count on commercially available arrays

Feature density increases exponentially.
“Moore’s law” for microarray data
Extrachromosomal DNA

Figure from *Immunobiology*. Janeway, Charles A.; Travers, Paul; Walport, Mark; Shlomchik, Mark. New York and London: Garland Science; 2001.
Probes on the genome

Analysis
Techniques for Probe-level Microarray Data

Humberto Ortiz Zuazaga

General Introduction

Error Correction and Clustering

Introduction
Methods
Results
Conclusions

Probe-level data

Introduction
Methods
Results
Conclusions

Future work

Motivation
Methods
Preliminary results
Probes on the sequence
1. General Introduction
2. Error Correction and Clustering
   - Introduction
   - Methods
   - Results
   - Conclusions
3. Probe-level data
   - Introduction
   - Methods
   - Results
   - Conclusions
4. Future work
   - Motivation
   - Methods
   - Preliminary results
   - Summary
Differentially expressed genes
Genome wide probe abundance

Analysis
Techniques for
Probe-level
Microarray
Data
Humberto
Ortiz Zuazaga

General
Introduction
Error
Correction and
Clustering
Introduction
Methods
Results
Conclusions

Probe-level
data
Introduction
Methods
Results
Conclusions

Future work
Motivation
Methods
Preliminary
results
## Candidate genes

<table>
<thead>
<tr>
<th>Description</th>
<th>Gene Expression Array Combined</th>
<th>logFC</th>
<th>logRatio</th>
<th>Amplification</th>
</tr>
</thead>
<tbody>
<tr>
<td>aCGH 244K</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aCGH 44K</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aCGH 44K</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aCGH COMBINED</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mus musculus a disintegrin and metalloproteinase domain 7 (Adam7), mRNA.</td>
<td>-0.315534</td>
<td>0.183978</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mus musculus a disintegrin and metalloproteinase domain 7 (Adam7), mRNA.</td>
<td>-1.19907</td>
<td>0.183978</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mus musculus a disintegrin and metalloproteinase domain 7 (Adam7), mRNA.</td>
<td>-0.315534</td>
<td>0.183978</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mus musculus a disintegrin and metalloproteinase domain 8 (Adam8), mRNA.</td>
<td>0.401733</td>
<td>0.297703</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mus musculus a disintegrin and metalloproteinase domain 8 (Adam8), mRNA.</td>
<td>-0.091417</td>
<td>0.297703</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mus musculus a disintegrin and metalloproteinase domain 8 (Adam8), mRNA.</td>
<td>0.401733</td>
<td>0.297703</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mus musculus a disintegrin and metalloproteinase domain 8 (Adam8), mRNA.</td>
<td>-0.091417</td>
<td>0.297703</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mus musculus a disintegrin and metalloproteinase domain 10 (Adam10), mRNA [NM_007399]</td>
<td>3.17E-01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mus musculus a disintegrin and metalloproteinase domain 33 (Adam33), mRNA [NM_033615]</td>
<td>2.48E-01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mus musculus a disintegrin and metalloproteinase domain 5 (Adam5), mRNA [NM_007401]</td>
<td>7.90E-01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mus musculus a disintegrin and metalloproteinase domain 5 (Adam5), mRNA [NM_007401]</td>
<td>2.86E-01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mus musculus a disintegrin and metalloproteinase domain 7 (Adam7), mRNA [NM_007402]</td>
<td>0.477659</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mus musculus a disintegrin and metalloproteinase domain 7 (Adam7), mRNA [NM_007402]</td>
<td>4.94E-01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mus musculus a disintegrin and metalloproteinase domain 7 (Adam7), mRNA [NM_007402]</td>
<td>3.61E-01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mus musculus ataxin 7-like 4 (Atxn7l4), mRNA [NM_028139]</td>
<td>2.41E-01</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
- logFC: Log Fold Change
- logRatio: Log Ratio
- Amplification: Amplification factor

**Legend:**
- aCGH: Array Comparative Genomic Hybridization
- Adam: Adam protein
- mRNA: Messenger RNA

**Additional information:**
- The table shows the different genes and their corresponding expression levels in the context of gene expression arrays.
1 General Introduction
2 Error Correction and Clustering
   - Introduction
   - Methods
   - Results
   - Conclusions
3 Probe-level data
   - Introduction
   - Methods
   - Results
   - Conclusions
4 Future work
   - Motivation
   - Methods
   - Preliminary results
   - Summary
Summary

- Microarray probe counts, data, and experiments continue to grow at exponential rates.
- Sequence databases also continue to grow at exponential rates.
- This growth presents a challenge, and opportunity to researchers.
- Combining multiple sources of information is possible, but requires new analysis and visualization techniques to be effective.
This work was supported in part by a SCORE grant (SC1MH086072) and an INBRE grant (P20RR016470) from the National Institutes of Health. The data was kindly provided by Dr. Sandra Peña de Ortiz and Edgardo Castro.
Analysis Techniques for Probe-level Microarray Data

Humberto Ortiz Zuazaga

General Introduction

Error Correction and Clustering

Introduction Methods Results Conclusions

Probe-level data

Introduction Methods Results Conclusions

Future work

Motivation Methods Preliminary results

Fin