Effect of single nucleotide polymorphisms on Affymetrix® match-mismatch probe pairs

Eric C. Rouchka1,*, Abhijit W. Phatak1, Amar V. Singh2,†

1Department of Computer Engineering and Computer Science,
University of Louisville, Louisville, KY, USA
2Department of Molecular, Cellular, and Craniofacial Biology,
University of Louisville, Louisville, KY, USA
Microarray Technology

• Microarray:
    • Allows study of thousands of genes at same time
  – Solid Support (Glass or nylon)
  – Slide of DNA molecules
    • Molecule: string of bases (25 bp – 500 bp)
    • uniquely identifies gene or unit to be studied
Fabrications of Microarrays

- Size of a microscope slide

Images: http://www.affymetrix.com
Differing Conditions

• **Ultimate Goal:**
  – Understand expression level of genes under different conditions

• **Helps to:**
  – Determine genes involved in a disease
  – Pathways to a disease
  – Used as a screening tool
Gene Conditions

- Cell types (brain vs. liver)
- Developmental (fetal vs. adult)
- Response to stimulus
- Gene activity (wild vs. mutant)
- Disease states (healthy vs. diseased)
Affymetrix Technology

- Biotin (one dye) instead of 2 colors
- One treatment per chip
- 11, 16, or 20 gene markers pairs per gene
Affymetrix Design of Probes

- Each marker has two corresponding probes: one perfect match (PM) and one mismatch (MM).

- MM has a single base difference (complementary base in the middle).

  - PM to maximize hybridization
  - MM to ascertain the degree of cross-hybridization
Example probe pair

- PM GCACAGCTTGCA**A**AGGATATTGCCA
- MM GCACAGCTTGCA**T**AGGATATTGCCA
Single Nucleotide Polymorphisms

• Single base differences between individuals in a population

• 11,751,216 catalogued in dbSNP for Humans (one every 300 bases)

• Causes for differences and disease
  – Sickle cell anemia; cystic fibrosis; muscular dystrophy; type II diabetes
SNPs and Affy Probes

• Question to be studied:
  – Does presence of SNP in 13th Base pose problems?
Datasets

• Human HU-133A platform from Affymetrix®
  – 22,000 probe sets
  – 247,965 probe pairs (PM/MM)
• dbSNP
  – build 124 (now up to 128)
• Human genome sequences
  – goldenpath build hg17 (genome.ucsc.edu)
• GEO dataset GDS1758 (12 experiments)
Methods

• Step 1: Map Affy Probes to genomic locations
• Step 2: Filter out only unique probes
• Step 3: Map SNPs to genomic locations
• Step 4: Find overlaps between affy probes and SNPs
• Step 5: Filter out overlaps where SNP occurs in 13th base (MM position)
Results

- 540 probe sequences with SNP in 13\textsuperscript{th} base (536 probe sets represented)
MM > PM?

• Looked at 12 experiments
• Took 536 probe sets
  – Two groups:
    • SNP containing probe pairs
    • Non-SNP containing probe pairs
  – How often was MM > PM in each of these?
<table>
<thead>
<tr>
<th># observations of Mismatch probe expression &gt; Match</th>
<th>Probes without SNP</th>
<th>Probes with SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percent Occurrence</td>
<td>Percent occurrence</td>
</tr>
<tr>
<td>0</td>
<td>47.0%</td>
<td>58.4%</td>
</tr>
<tr>
<td>1</td>
<td>8.9%</td>
<td>6.4%</td>
</tr>
<tr>
<td>2</td>
<td>5.4%</td>
<td>4.3%</td>
</tr>
<tr>
<td>3</td>
<td>4.3%</td>
<td>3.4%</td>
</tr>
<tr>
<td>4</td>
<td>3.8%</td>
<td>2.3%</td>
</tr>
<tr>
<td>5</td>
<td>3.5%</td>
<td>2.4%</td>
</tr>
<tr>
<td>6</td>
<td>3.1%</td>
<td>1.8%</td>
</tr>
<tr>
<td>7</td>
<td>2.8%</td>
<td>1.8%</td>
</tr>
<tr>
<td>8</td>
<td>2.6%</td>
<td>2.4%</td>
</tr>
<tr>
<td>9</td>
<td>2.5%</td>
<td>1.9%</td>
</tr>
<tr>
<td>10</td>
<td>2.5%</td>
<td>1.9%</td>
</tr>
<tr>
<td>11</td>
<td>3.7%</td>
<td>3.0%</td>
</tr>
<tr>
<td>12</td>
<td>10.0%</td>
<td>9.9%</td>
</tr>
</tbody>
</table>