Introduction to genomic microarray (aCGH) technology

Jonathan Pevsner, Ph.D.
SNP group at the Genome Cafe
August 3, 2005
Genomic microarrays are synonymous with array comparative genome hybridization (aCGH).

The standard protocol for aCGH is to isolate genomic DNA from two sources (e.g. a patient sample and a control pool), and label the DNA (Cy3 or Cy5; “green” or “red”).

A microarray is manufactured, consisting of bacterial artificial chromosomes (BACs) immobilized to a glass microscope slide or filter. Other versions of aCGH use yeast artificial chromosomes (YACs), cosmids, complementary DNA (cDNA), or oligonucleotides.
Introduction

The two samples are competitively hybridized to the array. Regions of chromosomal duplication and/or deletion are observed when the signal is relatively more green or red; most regions having no chromosomal copy number change appear an intermediate color (e.g. orange).
Scales of DNA

Chromosome 10: 135 Mb (983 genes)

Chromosome 1: 245 Mb (2580 genes)

Human genome: 3 billion base pairs

Chromosome 21: 46 Mb (337 genes)
Obtain BAC clones (200,000 base pairs) with chromosomal assignments that span the genome...

3,000 BAC clones x 1 Mb intervals = 3 billion base pairs
15,000 BACs x 0.2 Mb intervals = 3 billion base pairs
Genomic microarrays are available from commercial sources (e.g. Spectral Genomics and Vysis) and from academic and NIH-sponsored facilities (e.g. Roswell Park).

Some commercial arrays have about 100 to 400 BAC clones, spanning regions known to harbor deletions or duplications in clinical syndromes.
Genomic microarray methodology

Test Sample

Test

Reference

Cy3

Cy5

Exp. 1

Test Sample

Exp. 1

Test Sample
Genomic microarray methodology

Test Sample

Test Sample

Cy3

Cy5

Exp. 1

Cy5

Cy3

Exp. 2

pter

qter

Cy3/Cy5 Ratios

Ratios
<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>RP11-430E19</td>
<td>p36.33</td>
<td>3</td>
<td>chr1</td>
<td>92526.5</td>
<td>92526.5</td>
<td>0.217606567</td>
<td>1.162802694</td>
<td>11.18438</td>
</tr>
<tr>
<td>RP11-421C4</td>
<td>p36.33</td>
<td>3</td>
<td>chr1</td>
<td>1279023</td>
<td>1279023</td>
<td>0.041905374</td>
<td>1.02950467</td>
<td>10.82757</td>
</tr>
<tr>
<td>RP11-201E15</td>
<td>p36.33</td>
<td>2</td>
<td>chr1</td>
<td>1742026</td>
<td>1742026</td>
<td>0.175608425</td>
<td>1.129362337</td>
<td>7.59006</td>
</tr>
<tr>
<td>RP11-41H8</td>
<td>p36.33</td>
<td>3</td>
<td>chr1</td>
<td>1742026</td>
<td>1742026</td>
<td>0.002465973</td>
<td>1.001710744</td>
<td>10.81911</td>
</tr>
<tr>
<td>RP11-82D16</td>
<td>p36.33</td>
<td>3</td>
<td>chr1</td>
<td>2008810</td>
<td>2008810</td>
<td>-0.007704713</td>
<td>0.954673735</td>
<td>9.539708</td>
</tr>
<tr>
<td>RP11-737N8</td>
<td>p36.32</td>
<td>3</td>
<td>chr1</td>
<td>2711369</td>
<td>2711369</td>
<td>-0.016655833</td>
<td>0.986319597</td>
<td>8.286518</td>
</tr>
<tr>
<td>RP11-62M23</td>
<td>p36.32</td>
<td>3</td>
<td>chr1</td>
<td>3433793</td>
<td>3433793</td>
<td>-0.169398667</td>
<td>0.889213054</td>
<td>10.42137</td>
</tr>
<tr>
<td>RP11-51D17</td>
<td>p36.32</td>
<td>3</td>
<td>chr1</td>
<td>3960058</td>
<td>3960058</td>
<td>0.010970802</td>
<td>1.007633367</td>
<td>10.41297</td>
</tr>
<tr>
<td>RP11-237N15</td>
<td>p36.32</td>
<td>3</td>
<td>chr1</td>
<td>3994321</td>
<td>3994321</td>
<td>0.059699427</td>
<td>1.042248695</td>
<td>10.3252</td>
</tr>
<tr>
<td>RP11-110O5</td>
<td>p36.32</td>
<td>3</td>
<td>chr1</td>
<td>4265132</td>
<td>4265132</td>
<td>-0.011780806</td>
<td>0.991867416</td>
<td>11.61258</td>
</tr>
<tr>
<td>RP11-230F23</td>
<td>p36.31</td>
<td>0</td>
<td>chr1</td>
<td>5262547</td>
<td>5262547</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RP11-51B4</td>
<td>p36.31</td>
<td>6</td>
<td>chr1</td>
<td>6069162</td>
<td>6069162</td>
<td>-0.039100635</td>
<td>0.97326155</td>
<td>10.90924</td>
</tr>
<tr>
<td>RP11-58A11</td>
<td>p36.31</td>
<td>3</td>
<td>chr1</td>
<td>6323686</td>
<td>6323686</td>
<td>-0.058079947</td>
<td>0.960541634</td>
<td>10.34621</td>
</tr>
<tr>
<td>RP11-60J11</td>
<td>p36.32</td>
<td>3</td>
<td>chr1</td>
<td>6912363</td>
<td>6912363</td>
<td>0.019498133</td>
<td>1.013606818</td>
<td>10.61542</td>
</tr>
<tr>
<td>RP11-748N23</td>
<td>p36.23</td>
<td>3</td>
<td>chr1</td>
<td>7877043</td>
<td>7877043</td>
<td>-0.02676486</td>
<td>0.961619028</td>
<td>11.05003</td>
</tr>
<tr>
<td>RP11-81J7</td>
<td>p36.23</td>
<td>3</td>
<td>chr1</td>
<td>8678095</td>
<td>8678095</td>
<td>0.118617777</td>
<td>1.065694178</td>
<td>10.97545</td>
</tr>
<tr>
<td>RP11-90B14</td>
<td>p36.22</td>
<td>3</td>
<td>chr1</td>
<td>9290576</td>
<td>9290576</td>
<td>0.01685125</td>
<td>1.011748678</td>
<td>9.351194</td>
</tr>
<tr>
<td>RP11-874A11</td>
<td>p36.22</td>
<td>3</td>
<td>chr1</td>
<td>10166618</td>
<td>10166618</td>
<td>0.098549043</td>
<td>1.070690082</td>
<td>9.469107</td>
</tr>
<tr>
<td>RP11-199O1</td>
<td>p36.22</td>
<td>3</td>
<td>chr1</td>
<td>10263767</td>
<td>10263767</td>
<td>0.127539737</td>
<td>1.052429166</td>
<td>11.17051</td>
</tr>
<tr>
<td>RP11-28P4</td>
<td>p36.22</td>
<td>0</td>
<td>chr1</td>
<td>11013076</td>
<td>11013076</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RP11-426M1</td>
<td>p36.22</td>
<td>3</td>
<td>chr1</td>
<td>11891866</td>
<td>11891866</td>
<td>0.091609113</td>
<td>1.06570572</td>
<td>9.842095</td>
</tr>
<tr>
<td>RP11-106F7</td>
<td>p36.22</td>
<td>3</td>
<td>chr1</td>
<td>12041922</td>
<td>12041922</td>
<td>0.033407643</td>
<td>1.023426605</td>
<td>10.97711</td>
</tr>
<tr>
<td>RP11-90D15</td>
<td>p36.21</td>
<td>3</td>
<td>chr1</td>
<td>12668345</td>
<td>12668345</td>
<td>0.11014089</td>
<td>1.079333636</td>
<td>9.451463</td>
</tr>
<tr>
<td>RP11-90N5</td>
<td>p36.21</td>
<td>3</td>
<td>chr1</td>
<td>13200402</td>
<td>13200402</td>
<td>-0.014417443</td>
<td>0.990056358</td>
<td>9.49052</td>
</tr>
<tr>
<td>RP11-178M15</td>
<td>p36.21</td>
<td>3</td>
<td>chr1</td>
<td>13349191</td>
<td>13349191</td>
<td>0.03171834</td>
<td>1.02222894</td>
<td>11.16751</td>
</tr>
<tr>
<td>RP11-251N10</td>
<td>p36.21</td>
<td>3</td>
<td>chr1</td>
<td>13521555</td>
<td>13521555</td>
<td>0.112248397</td>
<td>1.080911422</td>
<td>10.98198</td>
</tr>
<tr>
<td>RP11-219F4</td>
<td>p36.21</td>
<td>3</td>
<td>chr1</td>
<td>14351549</td>
<td>14351549</td>
<td>0.114643707</td>
<td>1.062707623</td>
<td>10.41272</td>
</tr>
<tr>
<td>RP11-265F14</td>
<td>p36.21</td>
<td>5</td>
<td>chr1</td>
<td>15172959</td>
<td>15172959</td>
<td>0.090762372</td>
<td>1.064932782</td>
<td>10.10825</td>
</tr>
<tr>
<td>RP11-276H7</td>
<td>p36.13</td>
<td>3</td>
<td>chr1</td>
<td>15861063</td>
<td>15861063</td>
<td>0.06983472</td>
<td>1.049596431</td>
<td>10.95166</td>
</tr>
<tr>
<td>RP11-54F24</td>
<td>p36.13</td>
<td>1</td>
<td>chr1</td>
<td>16642970</td>
<td>16642970</td>
<td>-0.3492005</td>
<td>0.765019012</td>
<td>7.216903</td>
</tr>
<tr>
<td>RP11-79D15</td>
<td>p36.13</td>
<td>3</td>
<td>chr1</td>
<td>17443992</td>
<td>17443992</td>
<td>0.0654096</td>
<td>1.046382124</td>
<td>10.18863</td>
</tr>
<tr>
<td>RP11-91O22</td>
<td>p36.13</td>
<td>3</td>
<td>chr1</td>
<td>18162764</td>
<td>18162764</td>
<td>-0.020692573</td>
<td>0.966622726</td>
<td>11.39608</td>
</tr>
</tbody>
</table>
Potential of Genomic Microarrays

Whole genome perspective of karyotyping

Increased resolution of FISH

~3-10Mb

~40-100kb
## Four case studies

Characterization of four cases seen at KKI.

<table>
<thead>
<tr>
<th>Case #1</th>
<th>Identifier</th>
<th>Cytogenetic Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case #2</td>
<td>TC02-3387</td>
<td>2q deletion</td>
</tr>
<tr>
<td>Case #3</td>
<td>L96-1928</td>
<td>22q deletion, 6p duplication</td>
</tr>
<tr>
<td>Case #4</td>
<td>L99-2297</td>
<td>3p deletion</td>
</tr>
</tbody>
</table>
Case #1: genomic arrays suggest deletion on chromosome (2)(q32.2-q34)

190.5 to 213.8 Mb
Case #4: genomic arrays confirm deletion of distal chromosome 22q
Case #4: genomic arrays suggest possible duplication of chromosome 3q and duplication of 6p
X and Y for SNP chips

For normal males, there should be three possible SNP calls for the X chromosome:
[1] homozygous (A is interpreted by the algorithm as AA)
[2] homozygous (B is interpreted by the algorithm as BB)
[3] no call
Heterozygous (AB) calls should only happen as an error

For normal females, there are four possible SNP calls for the X chromosome:
[1] homozygous (AA)
[2] homozygous (BB)
[3] heterozygous (AB)
[4] no call
For normal males, there should be three possible SNP calls for the X chromosome:
[1] homozygous (A is interpreted by the algorithm as AA)
[2] homozygous (B is interpreted by the algorithm as BB)
[3] no call
Heterozygous (AB) calls should only happen as an error

For normal females, there are four possible SNP calls for the X chromosome:
[1] homozygous (AA)
[2] homozygous (BB)
[3] heterozygous (AB)
[4] no call
Chromosome Y probes are “lost” in a female patient
Case #1 – A female patient with a chromosome 7 deletion
X chromosome probes are “gained” in a female patient
Case #1 - Chromosome 7 probes are down-regulated
Genomic microarray analysis of case 1928 (22del)

- X chromosome probes (blue circles)
- 6p duplicated region (red circles)
- Chromosome 22 deleted probes
- Y chromosome probes
Integration of SNP chip and genomic microarray data

A

B

Normalized dye ratios

Consecutive homozygotes

Physical location, chromosome 22 (megabases)
Principal components analysis (PCA): visualizing highly dimensional data

Columns: 25 samples
Rows: 22,000 genes

25 x 25 Correlation matrix

Two points that are close to each other in PCA space have similar overall patterns of gene expression.

The units of the PC axes represent percent of variance captured along that axis.
Three major groups in an aCGH data set (each dot is a BAC ratio value)

PCA performed with Partek software (www.partek.com)
Explanation of three major groups in the data

[1] = a subset of the Y chromosome probes

[2] (the “disc”) = autosomal probes

[3] = most of the X chromosome probes
PCA after filtering out X and Y chromosome probes

Conclusions:

[1] the “disc” appears to show few outliers
[2] the data, viewed as BAC clones, do not have an unusual structure
Principal components analysis (PCA) with samples color-coded:
D1 (red) = date1 = type 1 samples
D2 (blue) = date2 = type 2 samples, one hyb
D3 (green) = date3 = type 2 samples, another hyb
D4 (purple) = date4 = type 1 samples, another set
D5 (orange) = date5 = type 3 samples

SUMMARY: We describe a tool, called aCGH-Smooth, for the automated identification of breakpoints and smoothing of microarray comparative genomic hybridization (array CGH) data. aCGH-Smooth is written in visual C++, has a user-friendly interface including a visualization of the results and user-defined parameters adapting the performance of data smoothing and breakpoint recognition. aCGH-Smooth can handle array-CGH data generated by all array-CGH platforms: BAC, PAC, cosmid, cDNA and oligo CGH arrays. The tool has been successfully applied to real-life data. AVAILABILITY: aCGH-Smooth is free for researchers at academic and non-profit institutions at http://www.few.vu.nl/~vumarray/.

SUMMARY: CGH-Explorer is a program for visualization and statistical analysis of microarray-based comparative genomic hybridization (array-CGH) data. The program has preprocessing facilities, tools for graphical exploration of individual arrays or groups of arrays, and tools for statistical identification of regions of amplification and deletion.

MOTIVATION: Chromosomal copy number changes (aneuploidies) are common in cell populations that undergo multiple cell divisions including yeast strains, cell lines and tumor cells. Identification of aneuploidies is critical in evolutionary studies, where changes in copy number serve an adaptive purpose, as well as in cancer studies, where amplifications and deletions of chromosomal regions have been identified as a major pathogenetic mechanism. Aneuploidies can be studied on whole-genome level using array CGH (a microarray-based method that measures the DNA content), but their presence also affects gene expression. In gene expression microarray analysis, identification of copy number changes is especially important in preventing aberrant biological conclusions based on spurious gene expression correlation or masked phenotypes that arise due to aneuploidies. Previously suggested approaches for aneuploidy detection from microarray data mostly focus on array CGH, address only whole-chromosome or whole-arm copy number changes, and rely on thresholds or other heuristics, making them unsuitable for fully automated general application to gene expression datasets. There is a need for a general and robust method for identification of aneuploidies of any size from both array CGH and gene expression microarray data.
RESULTS: We present ChARM (Chromosomal Aberration Region Miner), a robust and accurate expectation-maximization based method for identification of segmental aneuploidies (partial chromosome changes) from gene expression and array CGH microarray data. Systematic evaluation of the algorithm on synthetic and biological data shows that the method is robust to noise, aneuploidal segment size and P-value cutoff. Using our approach, we identify known chromosomal changes and predict novel potential segmental aneuploidies in commonly used yeast deletion strains and in breast cancer. ChARM can be routinely used to identify aneuploidies in array CGH datasets and to screen gene expression data for aneuploidies or array biases. Our methodology is sensitive enough to detect statistically significant and biologically relevant aneuploidies even when expression or DNA content changes are subtle as in mixed populations of cells. AVAILABILITY: Code available by request from the authors and on Web supplement at http://function.cs.princeton.edu/ChARM/

MOTIVATION: Plots of array Comparative Genomic Hybridization (CGH) data often show special patterns: stretches of constant level (copy number) with sharp jumps between them. There can also be much noise. Classic smoothing algorithms do not work well, because they introduce too much rounding. To remedy this, we introduce a fast and effective smoothing algorithm based on penalized quantile regression. It can compute arbitrary quantile curves, but we concentrate on the median to show the trend and the lower and upper quartile curves showing the spread of the data. Two-fold cross-validation is used for optimizing the weight of the penalties.

RESULTS: Simulated data and a published dataset are used to show the capabilities of the method to detect the segments of changed copy numbers in array CGH data.

SUMMARY: ArrayNorm is a user-friendly, versatile and platform-independent Java application for the visualization, normalization and analysis of two-color microarray data. A variety of normalization options were implemented to remove the systematic and random errors in the data, taking into account the experimental design and the particularities of every slide. In addition, ArrayNorm provides a module for statistical identification of genes with significant changes in expression. AVAILABILITY: The package is freely available for academic and non-profit institutions from http://genome.tugraz.at