A graphical user interface to DNA microarray data analysis using R and Bioconductor

Jarno Tuimala
CSC, the Finnish IT Center for Science
jarno.tuimala@csc.fi
Chipster

- Goal: Easy access to up-to-date analysis methods (R/Bioconductor)

- Features
  - Easy to use graphical interface
  - Comprehensive selection of analysis tools and visualizations
  - Support for different array types (Affymetrix, Agilent, Illumina, cDNA)
  - Possibility to construct (and share) analysis workflows
  - Compatible with Windows, Linux and Mac OS X
  - Easy to install and update
  - Graphical client program on user’s computer sends analysis jobs to central computing servers

- Availability
  - Guest accounts for evaluation
  - Open source (http://chipster.sourceforge.net)

- http://chipster.csc.fi
System architecture

servers

internet

desktop

auth.

brokers

SSL

client

web services

user

local resources

ANALYSIS

VISUALISATION

comp. nodes
Analysis tool content

- In addition to the usual analysis tools, such as, normalization, filtering, statistics, clustering and annotations, Chipster provides:
  - Quality control methods for all supported chiptypes
  - Linear (mixed) models for normalization and analysis
  - Bootstrapping of hierarchical clustering results
  - Alternative mappings of the probes on Affymetrix arrays (altCDF)
  - “Gene set enrichment analysis”
  - Promoter analysis

- Support for chiptypes:
  - Affymetrix (3’ expression, exon and SNP arrays)
  - Agilent (1-color & 2-color systems)
  - Illumina (expression arrays)
  - Generic cDNA
Acknowledgements

- Aleksi Kallio
- Jarno Tuimala
- Taavi Hupponen
- Petri Klemelä, Mikko Koski, Janne Käki, Mika Rissanen

- All the pilot users
- Department of Computer Science, University of Helsinki
- Tekes, the Finnish Funding Agency for Technology and Innovation
Chipster
Software for DNA microarray data analysis

Getting started with Chipster

Chipster allows you to perform DNA microarray data analysis with the R/Bioconductor and other tools through an intuitive graphical user interface. Chipster supports all major DNA microarray platforms (Affymetrix, Illumina, Agilent, and cDNA) and, being a Java program, it is compatible with Windows, Linux and MacOS.

Please note that a user account is required to use the system. Finnish users can use their regular CSC username and password (user accounts can be obtained by filling in this application form and ticking the Chipster box). Short-term evaluation accounts are available for others by contacting us directly.

Launch Chipster (version 1.0.4)

Announcements:

- Registration for Chipster courses is open
- Announcement archive...
### Analysis tools

**Two groups tests**

Tests for comparing the mean gene expression of two groups. Specify the test type, multiple testing correction method and a p-value cut-off for significance. All the test assume that the difference in group means is equal to be zero. LPE only works, if the whole data is used, i.e., the data should not be pre-filtered, if LPE is used.

### Workflow view

- **4 files**
  - **Norm**
  - **Prep**
  - **Stat**
  - **Path**
  - **Phe**
  - **Qual**

### Data visualisation

<table>
<thead>
<tr>
<th>Method</th>
<th>Spreadsheet</th>
<th>Redraw</th>
</tr>
</thead>
<tbody>
<tr>
<td>Showing 67 rows of 67</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>chip.microarray1.cel</th>
<th>chip.microarray2.cel</th>
<th>chip.microarray3.cel</th>
<th>chip.microarray4.cel</th>
<th>flag.microarray1.cel</th>
<th>flag.microarray2.cel</th>
</tr>
</thead>
<tbody>
<tr>
<td>201650_at</td>
<td>8.06</td>
<td>10.63</td>
<td>5.12</td>
<td>3.83</td>
<td>P</td>
</tr>
<tr>
<td>201744_s_at</td>
<td>9.74</td>
<td>8.73</td>
<td>7.25</td>
<td>4.0</td>
<td>P</td>
</tr>
<tr>
<td>201842_s_at</td>
<td>10.18</td>
<td>8.71</td>
<td>6.66</td>
<td>5.11</td>
<td>P</td>
</tr>
<tr>
<td>201909_at</td>
<td>10.38</td>
<td>10.75</td>
<td>5.39</td>
<td>5.58</td>
<td>A</td>
</tr>
<tr>
<td>202018_s_at</td>
<td>10.44</td>
<td>7.57</td>
<td>5.88</td>
<td>4.95</td>
<td>A</td>
</tr>
<tr>
<td>202036_s_at</td>
<td>9.97</td>
<td>7.97</td>
<td>9.11</td>
<td>4.77</td>
<td>A</td>
</tr>
<tr>
<td>202112_at</td>
<td>8.96</td>
<td>8.68</td>
<td>5.57</td>
<td>10.89</td>
<td>A</td>
</tr>
<tr>
<td>202237_at</td>
<td>10.54</td>
<td>11.14</td>
<td>6.4</td>
<td>10.33</td>
<td>A</td>
</tr>
<tr>
<td>202266_s_at</td>
<td>10.79</td>
<td>8.6</td>
<td>9.15</td>
<td>4.24</td>
<td>A</td>
</tr>
<tr>
<td>202310_s_at</td>
<td>10.59</td>
<td>11.18</td>
<td>5.23</td>
<td>7.49</td>
<td>P</td>
</tr>
<tr>
<td>202404_s_at</td>
<td>10.13</td>
<td>9.84</td>
<td>4.54</td>
<td>6.14</td>
<td>P</td>
</tr>
<tr>
<td>202440_s_at</td>
<td>10.8</td>
<td>9.65</td>
<td>6.47</td>
<td>5.71</td>
<td>P</td>
</tr>
<tr>
<td>203021_at</td>
<td>10.39</td>
<td>8.52</td>
<td>9.83</td>
<td>5.28</td>
<td>A</td>
</tr>
<tr>
<td>203453_at</td>
<td>9.99</td>
<td>8.52</td>
<td>8.53</td>
<td>4.87</td>
<td>A</td>
</tr>
<tr>
<td>204259_at</td>
<td>11.14</td>
<td>9.93</td>
<td>8.53</td>
<td>4.87</td>
<td>A</td>
</tr>
<tr>
<td>204269_at</td>
<td>11.01</td>
<td>7.67</td>
<td>9.59</td>
<td>5.49</td>
<td>P</td>
</tr>
<tr>
<td>204304_s_at</td>
<td>9.49</td>
<td>8.48</td>
<td>5.82</td>
<td>4.25</td>
<td>A</td>
</tr>
<tr>
<td>204607_at</td>
<td>5.59</td>
<td>5.52</td>
<td>9.01</td>
<td>9.49</td>
<td>A</td>
</tr>
<tr>
<td>204704_s_at</td>
<td>6.72</td>
<td>3.56</td>
<td>9.17</td>
<td>4.11</td>
<td>P</td>
</tr>
</tbody>
</table>

Connected to broker at nani.csc.fi
<table>
<thead>
<tr>
<th>Probeset ID</th>
<th>UniProt Gene Name</th>
<th>UniProt</th>
<th>RefSeq</th>
<th>Entrez Gene</th>
<th>Entrez Gene Symbol</th>
<th>Entrez Gene Description</th>
<th>UniProt Description</th>
<th>GO Term</th>
<th>GO ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>201744_s_at</td>
<td>TACSTD2 (synonym: GA733-1, M1S1, TROP2)</td>
<td>P09758</td>
<td>NM_002353</td>
<td>4070</td>
<td>TACSTD2</td>
<td>tumor-associated calcium signal transducer 2 precursor (Pancreatic carcinoma marker protein GA733-1) (Cell surface glycoprotein Trop-2)</td>
<td></td>
<td>cell proliferation, cell surface receptor linked signal transduction, cytosol, integral to plasma membrane, receptor activity, visual perception</td>
<td>0008283, 0007166, 0008529, 0008887, 0004872, 0007601</td>
</tr>
<tr>
<td>201842_s_at</td>
<td>SFRP1 (synonym: FRP, FRP1, SARP2)</td>
<td>Q8N474</td>
<td>NM_003012</td>
<td>6422</td>
<td>SFRP1</td>
<td>secreted frizzled-related protein 1</td>
<td>Secreted frizzled-related protein 1 precursor (SFRP-1) (Frizzled-related protein 1) (SFRP-1) (Secreted apoptosis-related protein 2) (SARP-2)</td>
<td>Wnt receptor signaling pathway, anti-apoptosis, extracellular space, membrane, morphogenesis, signal transduction, transmembrane receptor activity</td>
<td>0016055, 0006916, 0005615, 0016020, 0009653, 0007165, 0004888</td>
</tr>
</tbody>
</table>
Workflow – reusing your analysis pipeline

- Creates a "macro" that can be applied to another normalized dataset and phenodata
- Choose a dataset, and workflow records the analysis steps starting from it
- Workflows can be shared with other users
Wizard for Affymetrix data

- Ready-made workflow to find differentially expressed genes
  - Normalization
  - Phenodata creation
  - Statistical test
  - Hierarchical clustering