

Here is a sample of an ExpressionView XML file (ALL test data).

To improve readability, we only show the first entry of the lists.

Comments are written in red.

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<ged>
```

```
  # summary of data
```

```
    <summary>
```

```
      <description>ExpressionView data  
file</description>
```

```
      <version>1.0</version>
```

```
      <nmodules>8</nmodules>
```

```
      <ngenes>3522</ngenes>
```

```
      <nsamples>128</nsamples>
```

```
    </summary>
```

```
  # description of experiment
```

```
    <experimentdata>
```

```
      <title>Gene expression profile of adult T-  
cell acute lymphocytic leukemia identifies distinct  
subsets of patients with different response to  
therapy and survival.</title>
```

```
      <name>Chiaretti et al.</name>
```

```
      <lab>Department of Medical Oncology,  
Dana-Farber Cancer Institute, Department of  
Medicine, Brigham and Women's Hospital,
```

Harvard Medical School, Boston, MA 02115,
USA.</lab>

<abstract>Gene expression profiles were examined in 33 adult patients with T-cell acute lymphocytic leukemia (T-ALL). Nonspecific filtering criteria identified 313 genes differentially expressed in the leukemic cells. Hierarchical clustering of samples identified 2 groups that reflected the degree of T-cell differentiation but was not associated with clinical outcome. Comparison between refractory patients and those who responded to induction chemotherapy identified a single gene, interleukin 8 (IL-8), that was highly expressed in refractory T-ALL cells and a set of 30 genes that was highly expressed in leukemic cells from patients who achieved complete remission. We next identified 19 genes that were differentially expressed in T-ALL cells from patients who either had a relapse or remained in continuous complete remission. A model based on the expression of 3 of these genes was predictive of duration of remission. The 3-gene model was validated on a further set of T-ALL samples from 18 additional patients treated on the same clinical protocol. This study demonstrates that gene expression profiling can identify a limited number of genes that are predictive of response to induction therapy and

remission duration in adult patients with T-ALL.</abstract>

<url></url>

<annotation>hgu95av2</annotation>

<organism>Homo sapiens</organism>

</experimentdata>

list of genes

<genes>

column names and titles for genes
(shown in the datagrid)

<genetags>

<id>#</id>

<score>Score</score>

<name>Name</name>

<symbol>Symbol</symbol>

<entrezid>EntrezID</entrezid>

</genetags>

individual gene

<gene>

<id>1</id>

<score/>

<name>33500_i_at</name>

<symbol>LOC100126583</symbol>

<entrezid>100126583</entrezid>

</gene>

...

</genes>

list of samples

<samples>

column names and titles for samples
(shown in the datagrid)

<sampletags>

<id>#</id>

<score>Score</score>

<name>Name</name>

<diagnosis> Date of
diagnosis</diagnosis>
<sex> Gender of the patient</sex>
<age> Age of the patient at
entry</age>

</sampletags>

individual sample

<sample>

<id>128</id>

<score/>

<name>LAL4</name>

<diagnosis>NA</diagnosis>

<sex>NA</sex>

<age>NA</age>

</sample>

...

</samples>

list of modules

<modules>

column names and titles for modules
(shown in the datagrid)

<moduletags>

<id>#</id>

<name>Name</name>

<iterations>iterations</iterations>

<oscillation>oscillation</oscillation>

<thr_row>thr.row</thr_row>

<thr_col>thr.col</thr_col>

<freq>freq</freq>

<rob>rob</rob>

<rob_limit>rob.limit</rob_limit>

</moduletags>

column names and titles for go
categories (shown in the datagrid)

<gotags>

<id>#</id>

<go>GO</go>

<term>Term</term>

```
<ontology>Ontology</ontology>
<pvalue>PValue</pvalue>
<oddsratio>OddsRatio</oddsratio>
<expcount>ExpCount</expcount>
<count>Count</count>
<size>Size</size>
</gotags>
```

column names and titles for kegg pathways (shown in the datagrid)

```
<keggtags>
  <id>#</id>
  <kegg>KEGG</kegg>
  <pathname>Path Name</pathname>
  <pvalue>PValue</pvalue>
  <oddsratio>OddsRatio</oddsratio>
  <expcount>ExpCount</expcount>
  <count>Count</count>
  <size>Size</size>
</keggtags>
```

individual module

```
<module>
  <id>1</id>
  <name>module 1</name>
  <iterations>22</iterations>
  <oscillation>0</oscillation>
  <thr_row>2.7</thr_row>
```

<thr_col>1.4</thr_col>

<freq>1</freq>

<rob>21.98</rob>

<rob_limit>21.98</rob_limit>

list of genes contained in module

<containedgenes>214, 215, 216, 217,
218, 219, 220, 221, 222, 223, 224, 225, 226, 227,
228, 229, 230, 231, 232, 233, 234, 235, 236, 237,
238, 239, 240, 241, 242, 243, 244, 245, 246, 247,
248, 249, 250, 251</containedgenes>

gene socres

<genescores>-0.94, -0.88, 0.74, -0.76,
-1.00, -0.84, -0.74, -0.76, -0.85, -0.88, 0.78, -0.80,
-0.95, 0.78, 0.79, 0.79, 0.83, 0.92, 0.83, -1.00,
0.79, 0.89, -0.84, -0.91, -0.82, -0.83, -0.80, 0.73,
-0.86, -0.78, -0.74, -0.80, -0.73, 0.87, -0.82, -0.92,
0.77, 0.95</genescores>

list of samples contained in module

<containedsamples>63, 64, 65, 54,
66, 67, 55, 56, 53, 57, 29, 58, 62, 30, 59, 68, 69,
60, 70, 61, 71</containedsamples>

sample scores

<samplescores>-0.62, -1.00, -0.77,
-0.40, -0.28, -0.70, -0.48, -0.14, -0.36, -0.63, -0.26,
-0.48, -0.22, -0.27, -0.46, -0.34, -0.07, -0.76, -0.25,
-0.08, -0.36</samplescores>

list of intersecting modules

<intersectingmodules>7</intersectingmodules>

list of go categories

<gos>

individual go category

<go>

<id>1</id>

<go>GO:0006955</go>

<term>immune

response</term>

<ontology>BP</ontology>

<pvalue>6.78e-07</pvalue>

<oddsratio>10.13</oddsratio>

<expcount>2.72</expcount>

<count>16</count>

<size>213</size>

</go>

...

</gos>

list of kegg pathways

<keggs>

individual kegg pathway

<kegg>

<id>1</id>

<kegg>05310</kegg>

<pathname>Asthma</pathname>

<pvalue>9.67e-05</pvalue>

<oddsratio>36.43</oddsratio>

<expcount>0.26</expcount>

<count>5</count>

<size>14</size>

</kegg>

...

</keggs>

</module>

...

</modules>

gene expression data

rounded to two digits, encoded in Base64

loop over genes, loop over samples

<data>

9/oH7P8U9hPu/xj47e7p9iHp8PL+

+/ft6v8YN+3p8AEMBvb07fX2C/gG4wgg7/f9AxbJB
Qfv7fHm

6ej94AX48Pbu6+AA7OH29fPu7enu5uv89R3+8Qg
m8vcH6vn89/H16u7zAgAEA+716fjk5+zq6AXu

...

</data>

</ged>