Integrating Gene Regulatory Networks to identify cancer-specific genes

Valeria Bo¹, Dimple Chudasama², Emmanouil Karteris² and Allan Tucker¹

¹Department of Computer Science, Brunel University, London, UK
²Department of Life Sciences, Brunel University, London, UK
Motivations

- Microarrays: simultaneous expression of thousands of genes under different conditions
- Improving techniques and data storage → large increase of data
- Several computational tools to decrypt information
- Analysis of multiple studies → consensus network

Unique networks & unique genes
**Unique-networks**

Network 1

Network 2

Network 3

Unique Network 1

# Unique-networks

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Study title</th>
<th>Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE18864</td>
<td>Triple Negative Breast Cancer</td>
<td>84</td>
</tr>
<tr>
<td>GSE9891</td>
<td>Ovarian Tumour</td>
<td>285</td>
</tr>
<tr>
<td>GSE21653</td>
<td>Medullary Breast Cancer</td>
<td>266</td>
</tr>
<tr>
<td>GSE10445</td>
<td>Adenocarcinoma and large cell Lung Carcinoma</td>
<td>72</td>
</tr>
</tbody>
</table>
Unique-networks

Gene Expression Omnibus (GEO)

PCA
Sd thr

Data Matrix D

Samples

1 2 ... ...
M Studies

glasso

Detect unique edges and corresponding genes

Internal vs External Prediction

Biological support

Subset S1 of Matrix D

Subset S2 of Matrix D

Unique Genes

Unique Networks

Brunel University London
Unique-networks

Unique-network for Medullary-breast cancer

Unique-network for Lung cancer
Unique-networks

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Study title</th>
<th>Samples</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE18864</td>
<td>Triple Negative Breast Cancer</td>
<td>84</td>
<td>0.55</td>
</tr>
<tr>
<td>GSE9891</td>
<td>Ovarian Tumour</td>
<td>285</td>
<td>0.00</td>
</tr>
<tr>
<td>GSE21653</td>
<td>Medullary Breast Cancer</td>
<td>266</td>
<td>0.02</td>
</tr>
<tr>
<td>GSE10445</td>
<td>Adenocarcinoma and large cell Lung Carcinoma</td>
<td>72</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Unique genes

Same logic behind unique-networks
Unique genes

GeneCards®
The Human Gene Compendium
Unique genes

Unique genes for condition 1

- U-N1
- U-N2
- U-N3
- U-N4
Unique genes

- Genes U-N2
- Genes U-N3
- Genes U-N4

Unique genes for condition 1

- Gene U-N1
NBH probability score *

Given cluster, $i$ of size $s_i$, contains $x$ genes from a defined functional group of size $k_j$, then the chance of this occurring by chance follows a binomial distribution. But when $k_j$ and $x$ are very large $\rightarrow$ normal approximation

$$z = \frac{x - \mu}{\sigma}, \mu = k_j p, \sigma = \sqrt{k_j pq}$$

<table>
<thead>
<tr>
<th>Study ID</th>
<th>$s_i$</th>
<th>$k_j$</th>
<th>$x$</th>
<th>$n$</th>
<th>$z$-score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE18864</td>
<td>117</td>
<td>2982</td>
<td>11</td>
<td>54675</td>
<td>1.83</td>
<td>$\leq 3.4%$</td>
</tr>
<tr>
<td>GSE9891</td>
<td>61</td>
<td>692</td>
<td>4</td>
<td>54675</td>
<td>3.68</td>
<td>$\leq 1%$</td>
</tr>
<tr>
<td>GSE21653</td>
<td>89</td>
<td>0</td>
<td>0</td>
<td>54675</td>
<td>N/A</td>
<td>$\leq 1%$</td>
</tr>
<tr>
<td>GSE10445</td>
<td>80</td>
<td>240</td>
<td>3</td>
<td>54675</td>
<td>4.47</td>
<td>$\leq 1%$</td>
</tr>
</tbody>
</table>

*Swift S. et al. Genome biology 2004
Tissue Microarray

RQ Values

- COL12A-Br
- COL12A-Lung
- COL12A-OC
- FSTL-1 Br
- FSTL-1 Lung
- FSTL-1 OC
- SPR11 Br
- SPR11 Lung
- SPR11 OC
- RAD5 Br
- RAD5 Lung
- RAD5 OC

Comparisons:
- Breast
- Ovarian
- Med
- Lung
ONCOMINE data for RAD51AP1

KM Plots for RAD51AP1
ONCOMINE data for RAD51AP1

Breast Cancer

KM Plots for RAD51AP1

HR = 1.64 (1.29 - 2.08)
logrank P = 4.4e-05

Expression
- low
- high

Number at risk
- low: 561, 486, 282, 84, 14, 2, 0
- high: 556, 397, 192, 45, 7, 1, 0

Time (months)
ONCOMINE data for RAD51AP1

Lung Cancer

KM Plots for RAD51AP1
GUI
Conclusions

• Identify unique sub-networks and genes based upon a number of microarray studies

• Support results using prediction accuracy and NBH probability score

• Application interface to combine different studies through AND and NOT logic operators

• Highlights structures and nodes that could be potential targets for further research

• Pipeline code available soon!

• valeria.bo@cruk.cam.ac.uk