Integrating Gene Regulatory Networks to identify cancer-specific genes

Valeria Bo\textsuperscript{1}, Dimple Chudasama\textsuperscript{2}, Emmanouil Karteris\textsuperscript{2} and Allan Tucker\textsuperscript{1}

\textsuperscript{1}Department of Computer Science, Brunel University, London, UK
\textsuperscript{2}Department 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

Bo V. et al. Discovering Study-Specific Gene Regulatory Networks. PLoS ONE. 2014.
| Study ID   | Study title                                      | Samples |
|-----------|--------------------------------------------------|---------|
| GSE18864  | Triple Negative Breast Cancer                    | 84      |
| GSE9891   | Ovarian Tumour                                   | 285     |
| GSE21653  | Medullary Breast Cancer                          | 266     |
| GSE10445  | Adenocarcinoma and large cell Lung Carcinoma     | 72      |
Unique-networks

Gene Expression Omnibus (GEO)

PCA
Sd thr

Genes

Data Matrix D

Samples

1 2 ... ...
M Studies

glasso

Detect unique edges and corresponding genes

Unique Networks

Internal vs External Prediction

Biological support

Subset S1 of Matrix D

Subset S2 of Matrix D

Unique Genes

Samples

i ...

j ...

M

...
Unique-networks

Unique-network for Medullary-breast cancer

Unique-network for Lung cancer
# Unique-networks

| Study ID   | Study title                                           | Samples | P-value |
|------------|-------------------------------------------------------|---------|---------|
| GSE18864   | Triple Negative Breast Cancer                        | 84      | 0.55    |
| GSE9891    | Ovarian Tumour                                       | 285     | 0.00    |
| GSE21653   | Medullary Breast Cancer                              | 266     | 0.02    |
| GSE10445   | Adenocarcinoma and large cell Lung Carcinoma         | 72      | 0.00    |
Unique genes

Same logic behind unique-networks

List 1

List 2

List 3

List 4
Unique genes
Unique genes

Unique genes for condition 1

GeneCards
The Human Gene Compendium

GeneALaCart
GeneCards Batch Queries
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 → normal approximation

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

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

*Swift S. et al. Genome biology 2004
ONCOMINE data for RAD51AP1

Ovarian Cancer

KM Plots for RAD51AP1

|   | Number at risk |
|---|----------------|
| low | 791 223       |
| high| 791 182       |

Expression
- low
- high

HR = 1.17 (1.03 - 1.34)
logrank P = 0.02
ONCOMINE data for RAD51AP1

Breast Cancer

KM Plots for RAD51AP1

HR = 1.64 (1.29 – 2.08)
log rank P = 4.4e-05
ONCOMINE data for RAD51AP1

KM Plots for RAD51AP1

Lung Cancer

Box plots and Kaplan-Meier (KM) plots illustrating the expression levels of RAD51AP1 in lung cancer. The box plots show the distribution of log2 median-centered intensity across different expression levels, while the KM plots highlight the survival analysis with hazard ratios and log-rank statistics.
Logic Application

Choose the original data file .RData File
Choose File: display passed data.RData
Upload complete

Choose the adjacency matrix .RData File
Choose File: cency studies thr.RData
Upload complete

Choose the studies description .csv File
Choose File: shiny_display/studies.csv
Upload complete

AND studies
1 4

NOT studies
2

Study Description
1 1 Breast Cancer
2 2 Ovarian Cancer
3 3 Medullary Breast Cancer
4 4 non small cell Lung Cancer

Save sub-networks

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