



Brunel
University
London

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

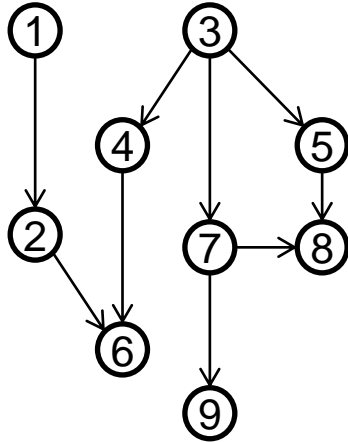
Motivations

- 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*

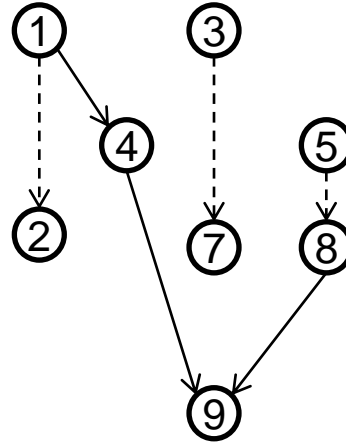
Unique networks & unique genes

Unique-networks

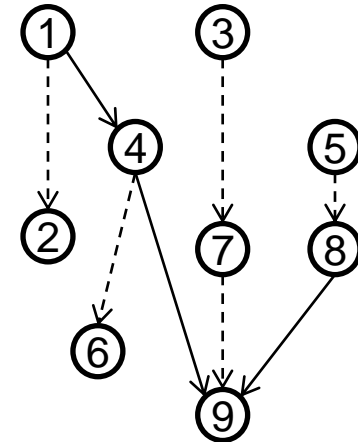
Network 1



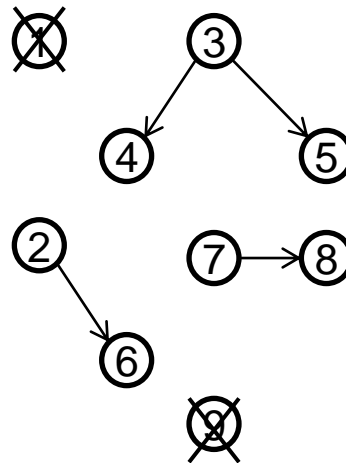
Network 2



Network 3



Unique Network 1

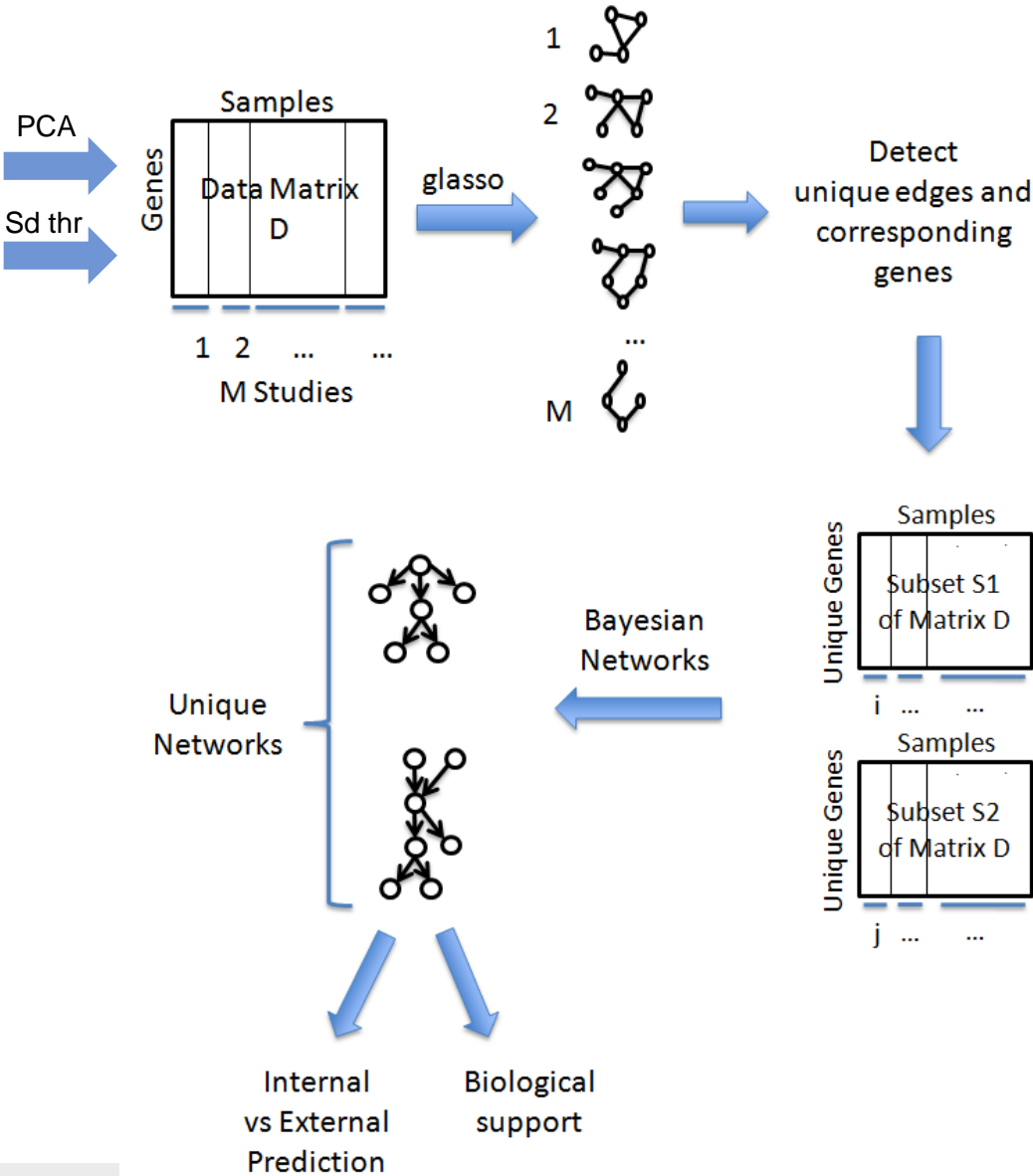


Unique-networks

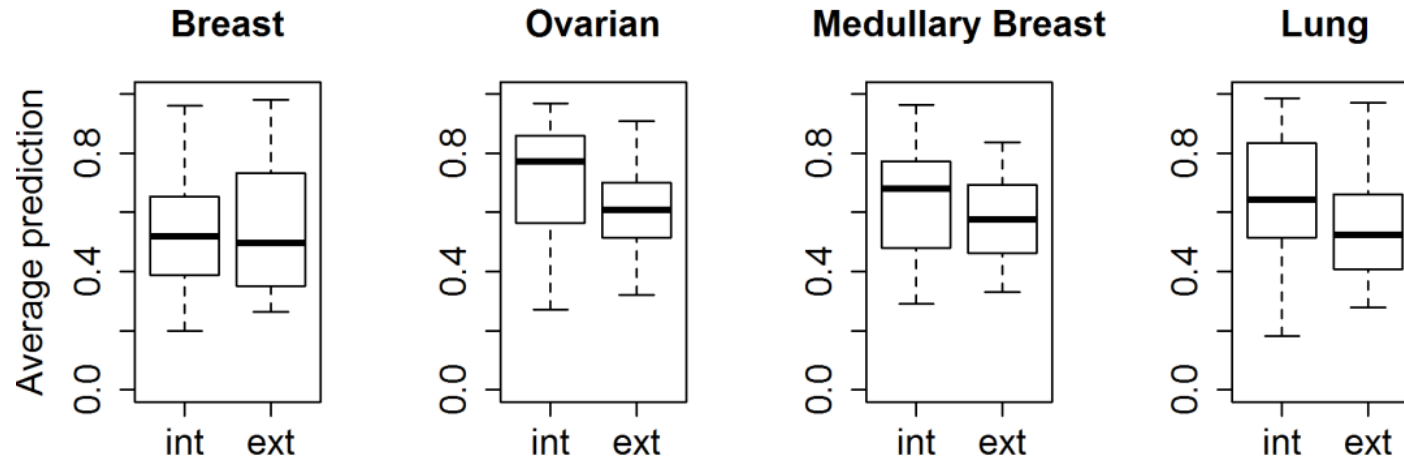


Study ID	Study title	Samples
GSE18864	Triple Negative Breast Cancer	84
GSE9891	Ovarian Tumour	285
GSE21653	Medullary Breast Cancer	266
GSE10445	Adenocarcinoma and large cell Lung Carcinoma	72

Unique-networks



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

Gene ID	Gene Name	Gene Type	Gene Class	Gene Function	Gene Description
443 20676_at	7959 tissue factor TFF	GO:007396 TAS	BP	blood coagulation	0.6133333
443 20676_at	7959 tissue factor TFF	GO:007396 TAS	BP	blood coagulation	entrez pub 0.6133333
443 20676_at	7959 tissue factor TFF	GO:002051 TAS	BP	negative regulation of cell cycle	0.5
702 110467_at	4111 melanoma H	L1 MAM6411	GO:0000374 ND	MF	melanocyte_n 0.52
702 110467_at	4111 melanoma H	L1 MAM6411	GO:0000375 ND	MF	cellular_n 0.52
702 110467_at	4111 melanoma H	L1 MAM6411	GO:0000376 ND	MF	hemiplegic_p 0.52

List 2

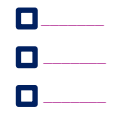
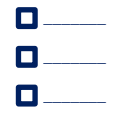
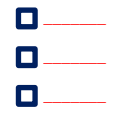
Gene ID	Gene Name	Gene Type	Gene Class	Gene Function	Gene Description
280 20426_at	10619 K4015 avian RAD51L1	GO:0000704 NAS	BP	double-strand break repair	0.8482761
280 20426_at	10619 K4015 avian RAD51L1	GO:0000704 NAS	BP	single-strand break repair	0.8482761
280 20426_at	10619 K4015 avian RAD51L1	GO:0000704 NAS	BP	single-strand break repair	0.8482761
280 20426_at	10619 K4015 avian RAD51L1	GO:0000704 NAS	BP	single-strand break repair	0.8482761
280 20426_at	10619 K4015 avian RAD51L1	GO:0000704 NAS	BP	single-strand break repair	0.8482761

List 3

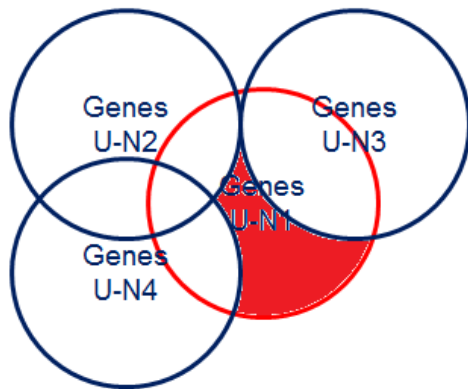
Gene ID	Gene Name	Gene Type	Gene Class	Gene Function	Gene Description
829 213302_at	6205 ribosomal protein RPS11	GO:0009308 TAS	BP	viral infection	0.8866667
829 213302_at	6205 ribosomal protein RPS11	GO:0009308 TAS	BP	viral transcription	0.8866667
829 213302_at	6205 ribosomal protein RPS11	GO:0009308 TAS	BP	viral transcription	0.8866667
829 213302_at	6205 ribosomal protein RPS11	GO:0009308 TAS	BP	viral transcription	0.8866667
829 213302_at	6205 ribosomal protein RPS11	GO:0009308 TAS	BP	viral transcription	0.8866667

List 4

Gene ID	Gene Name	Gene Type	Gene Class	Gene Function	Gene Description
447 20982_at	6440 surfactant protein SP1FC	GO:0005515 PR	MF	protein binding	0.9545454
447 20982_at	6440 surfactant protein SP1FC	GO:0005515 PR	MF	protein binding	0.9545454
447 20982_at	6440 surfactant protein SP1FC	GO:0005515 PR	MF	protein binding	0.9545454
447 20982_at	6440 surfactant protein SP1FC	GO:0005515 PR	MF	protein binding	0.9545454
447 20982_at	6440 surfactant protein SP1FC	GO:0005515 PR	MF	protein binding	0.9545454

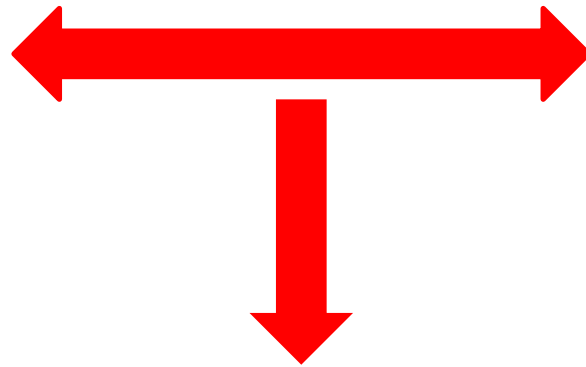
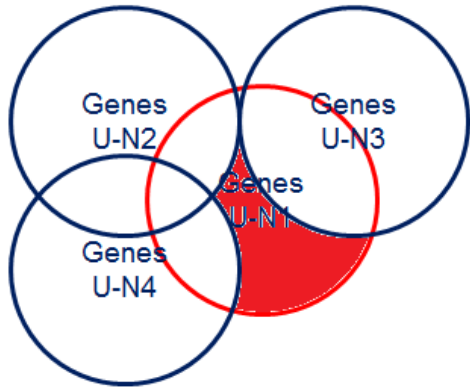


Unique genes



- _____
- _____
- _____

Unique genes

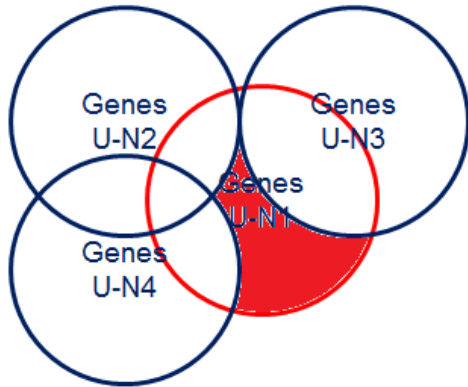


- _____
- _____
- _____

Unique genes for condition 1

- _____
- _____
- _____

Unique genes



- _____
- _____
- _____

Unique genes for condition 1

- _____
- _____
- _____



NBH probability score *

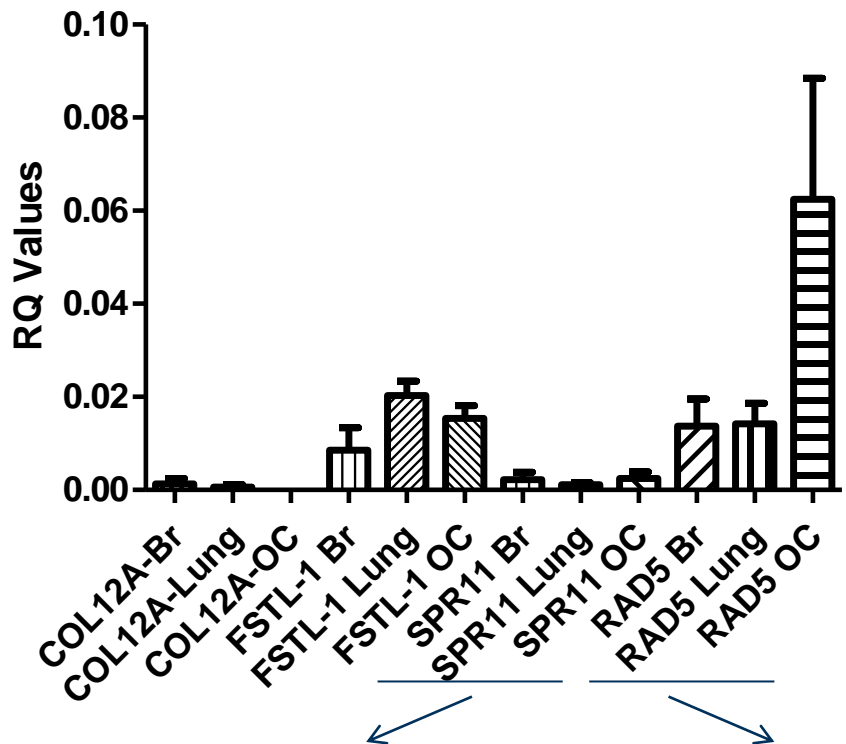
Given cluster, i of size s_i , contains x genes from a defined functional group of size k_j , then the chance of this occurring by chance follows a binomial distribution. But when k_j and x are very large \rightarrow normal approximation

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

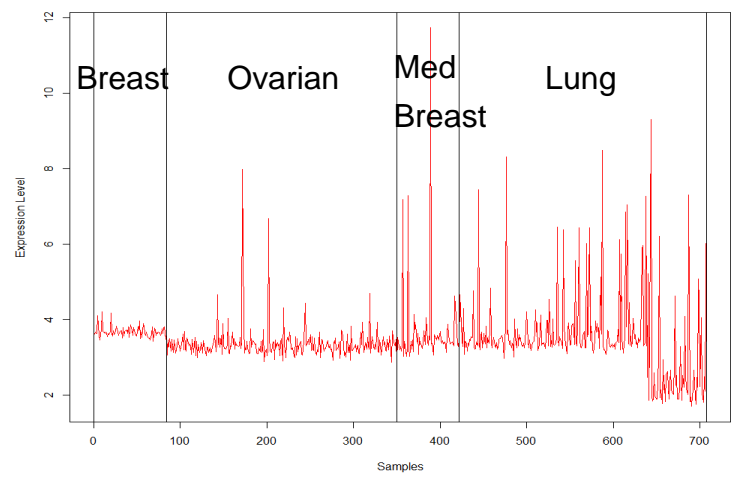
Study ID	s_i	k_j	x	n	z -score	p-value
GSE18864	117	2982	11	54675	1.83	$\leq 3.4\%$
GSE9891	61	692	4	54675	3.68	$\leq 1\%$
GSE21653	89	0	0	54675	N/A	$\leq 1\%$
GSE10445	80	240	3	54675	4.47	$\leq 1\%$

*Swift S. et al. Genome biology 2004

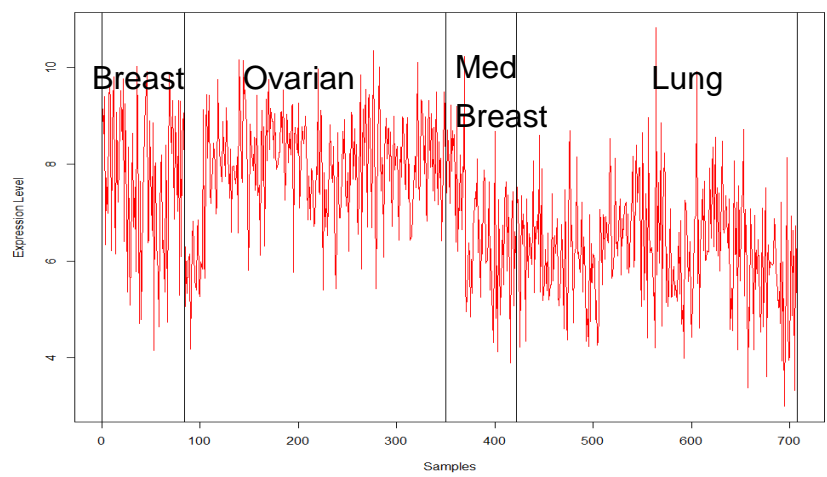
Tissue Microarray



SPRR1A



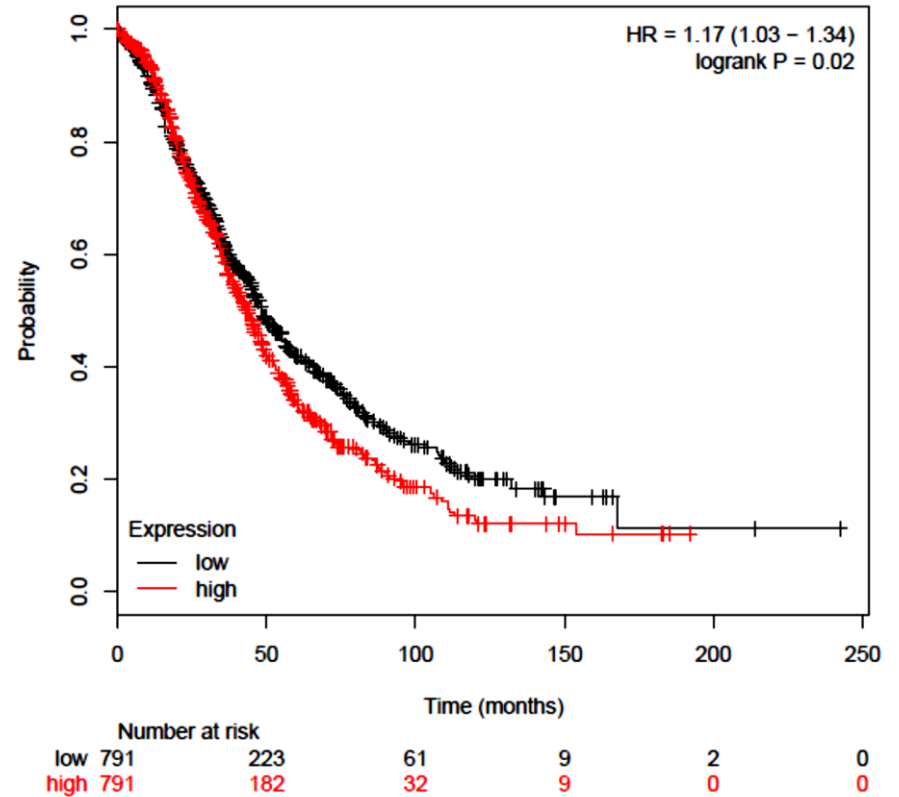
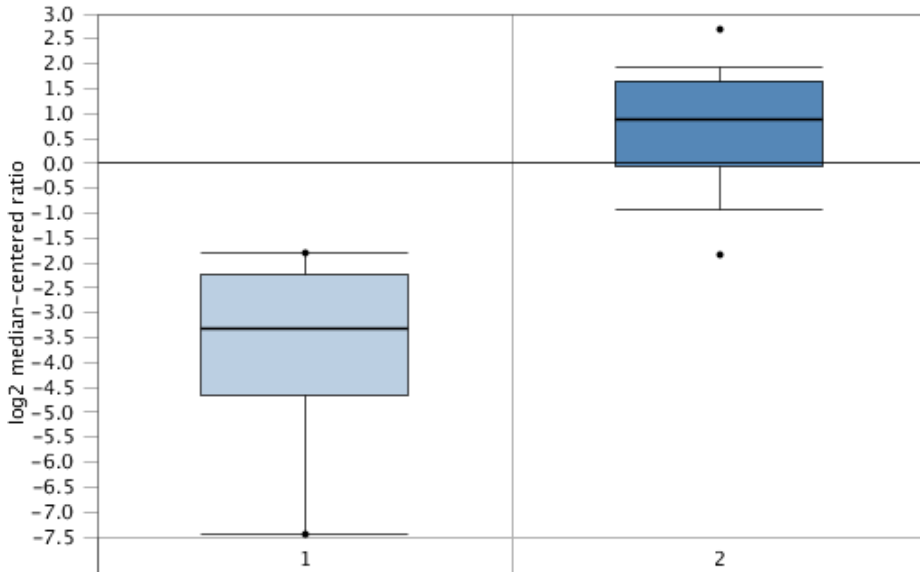
RAD51AP1



ONCOMINE data for RAD51AP1

KM Plots for RAD51AP1

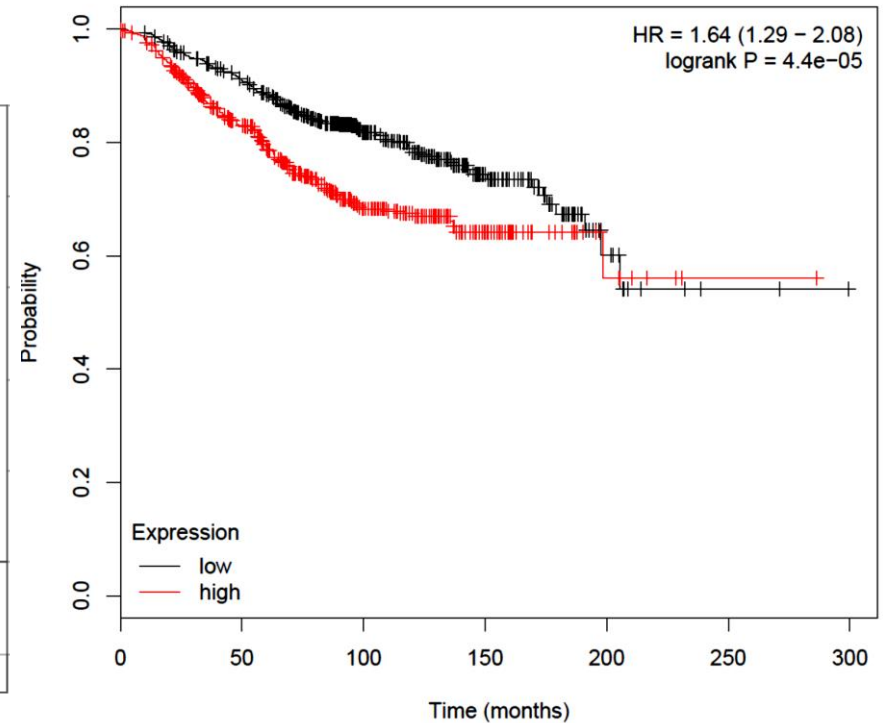
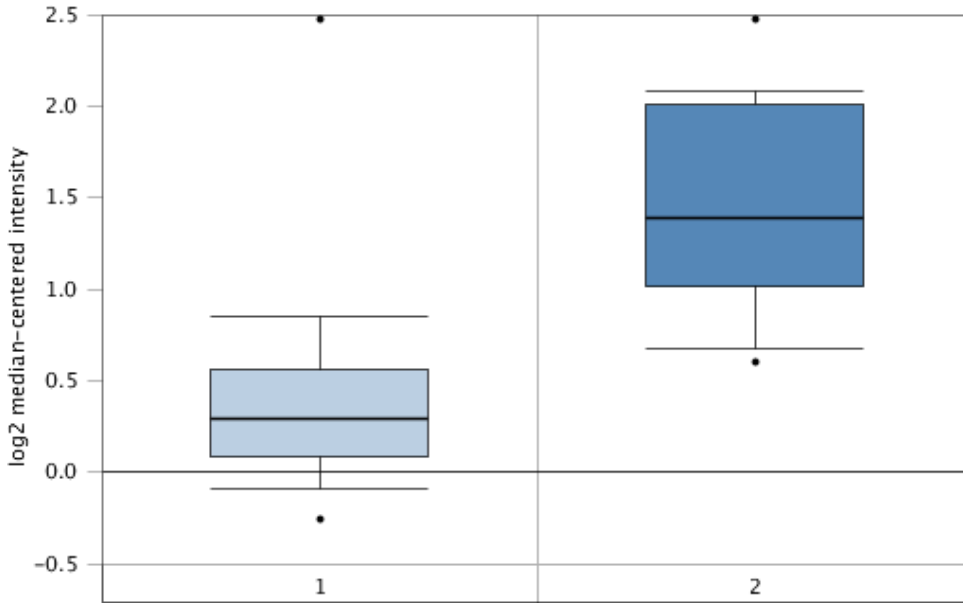
Ovarian Cancer



ONCOMINE data for RAD51AP1

KM Plots for RAD51AP1

Breast Cancer

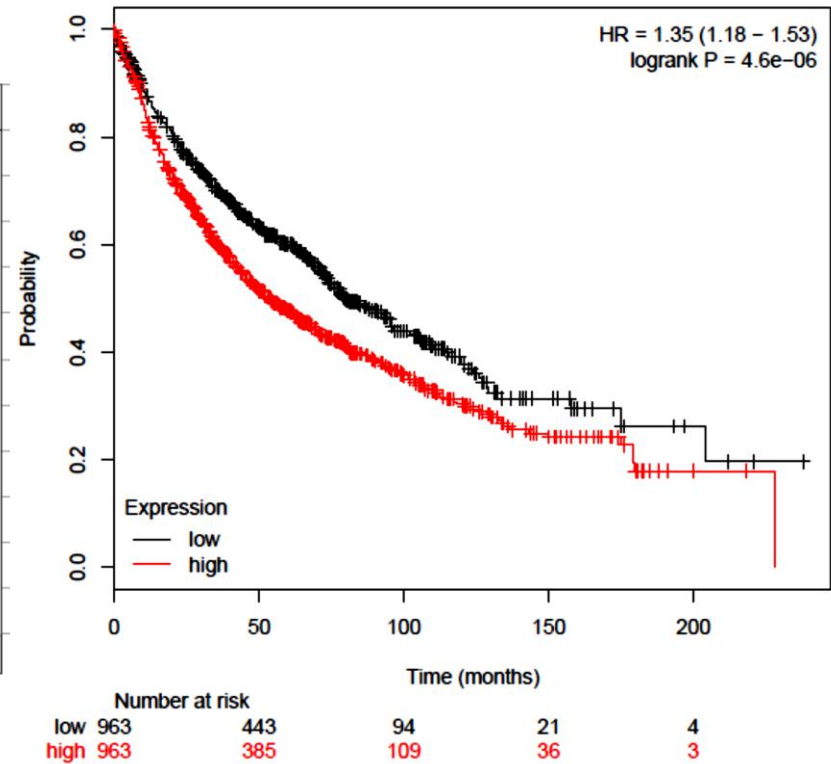
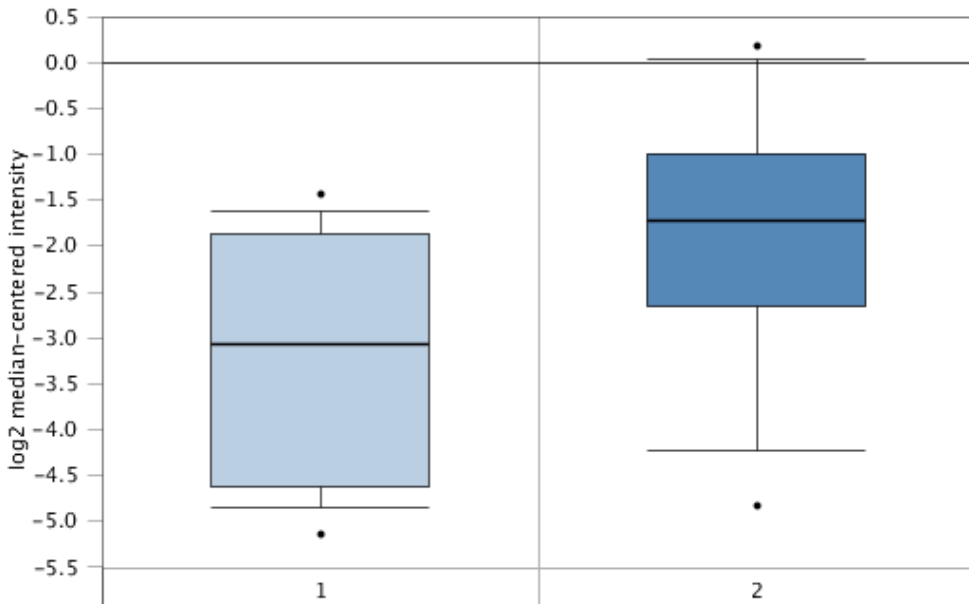


	Number at risk						
	0	50	100	150	200	250	300
low	561	486	282	84	14	2	0
high	556	397	192	45	7	1	0

ONCOMINE data for RAD51AP1

KM Plots for RAD51AP1

Lung Cancer



GUI

Logic Application

Choose the original data file .RData File

Choose File Upload complete

Choose the adjacency matrix .RData File

Choose File Upload complete

Choose the studies description .csv File

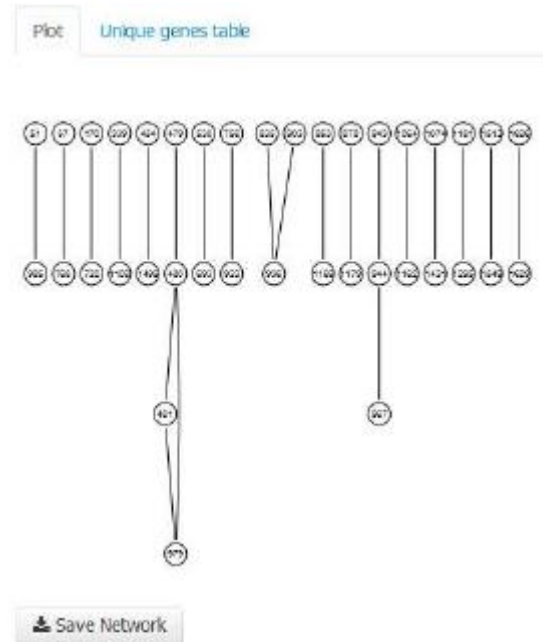
Choose File Upload complete

AND studies

NOT studies

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

Show sub-networks



Plot Unique genes table

id	rd
1	10008_04_at
2	15597_73_at
3	20040_4_at
4	20473_3_at
5	20580_0_at
6	20610_4_at
7	20616_0_at
8	20618_0_at
9	20743_0_at
10	21029_7_at
11	21110_1_at
12	21197_4_at
13	21175_6_at
14	21350_2_at
15	21381_3_at
16	21421_0_at
17	21497_7_at
18	21673_3_at
19	21694_0_at
20	21851_0_at
21	21855_7_at
22	21728_1_at
23	21752_0_at
24	21782_0_at
25	22017_2_at
26	22042_0_at
27	22224_2_at
28	22394_0_at
29	22432_1_at
30	22458_0_at
31	22459_0_at
32	22787_1_at
33	23153_5_at
34	23822_0_at
35	23820_3_at
36	23890_0_at
37	AFFX-HUMRGE/M10098_3_at
38	AFFX-M27830_0_at

Save gene list

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

