

NETWORK-BASED DATA ANALYSIS

ELEONORA GIULIANI

2024/2025

DATASET GSE43837

38 samples

Genes: 61359



19 human breast
cancer brain
metastases

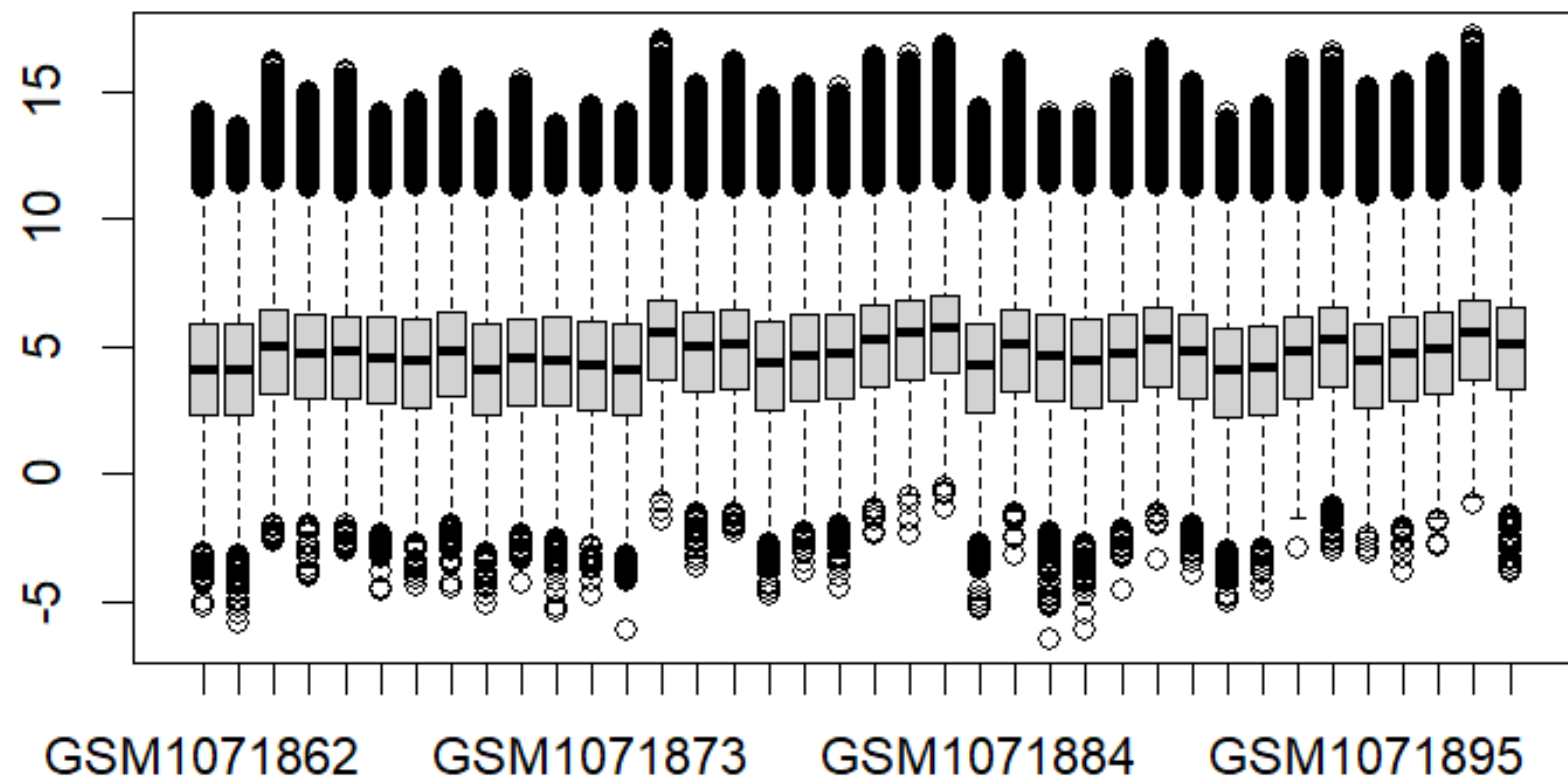
GSM1071862 brain_met_J3
GSM1071863 brain_met_BT25
GSM1071864 brain_met_J12
GSM1071865 brain_met_BT16
GSM1071866 brain_met_J8
GSM1071867 brain_met_BT34
GSM1071868 brain_met_J1
GSM1071869 brain_met_BT11
GSM1071870 brain_met_J2
GSM1071871 brain_met_BT30
GSM1071872 brain_met_J4
GSM1071873 brain_met_BT57
GSM1071874 brain_met_J5
GSM1071875 brain_met_BT51
GSM1071876 brain_met_BT31
GSM1071877 brain_met_J7
GSM1071878 brain_met_J10
GSM1071879 brain_met_J11
GSM1071880 brain met BT40

19 nonmetastatic
primary human
breast tumors

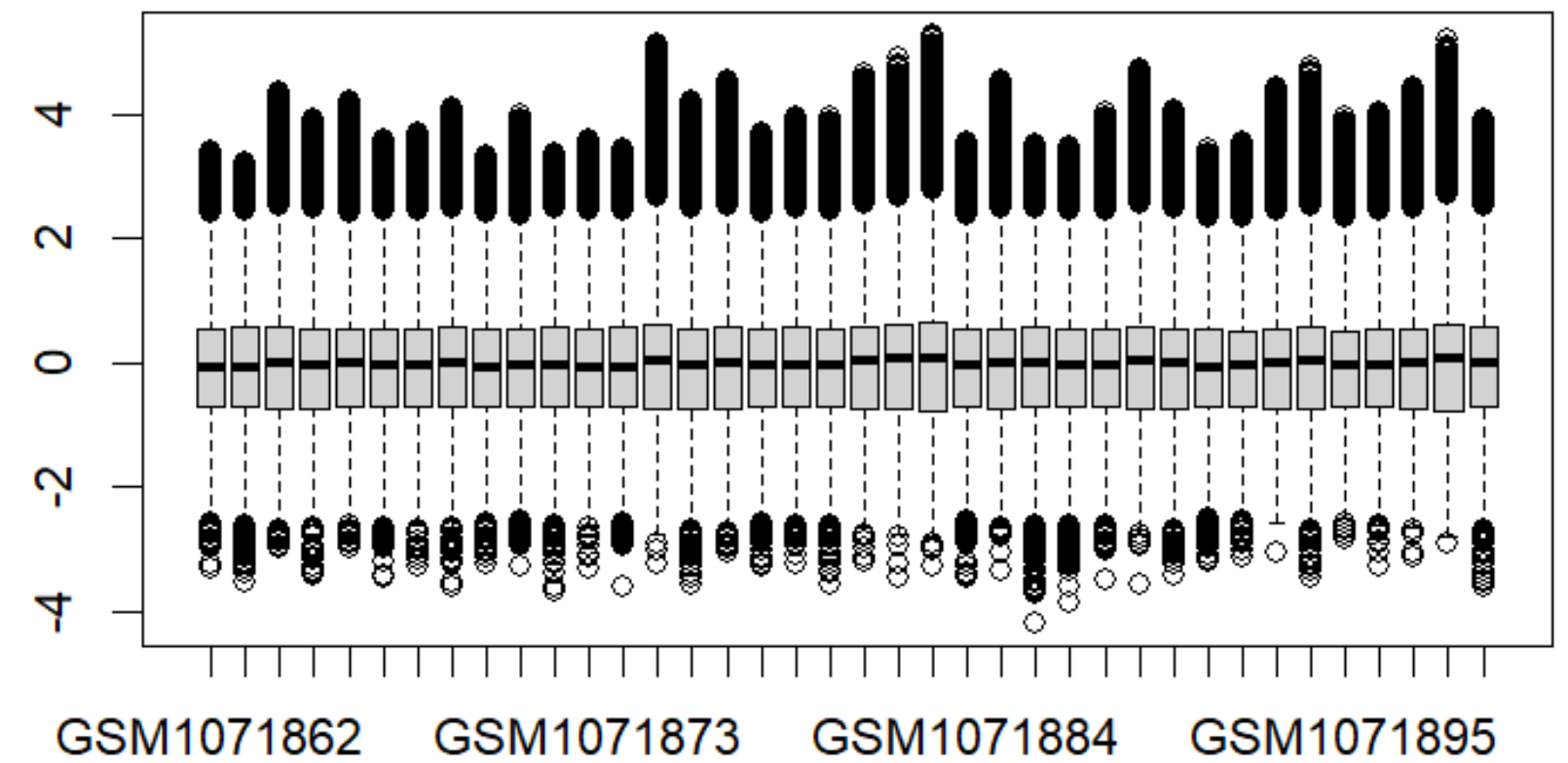
GSM1071881 breast_primary_MGH16
GSM1071882 breast_primary_MGH42
GSM1071883 breast_primary_MGH46
GSM1071884 breast_primary_MGH133
GSM1071885 breast_primary_MGH153
GSM1071886 breast_primary_MGH351
GSM1071887 breast_primary_MGH1104
GSM1071888 breast_primary_MGH574
GSM1071889 breast_primary_MGH434
GSM1071890 breast_primary_MGH450
GSM1071891 breast_primary_MGH421
GSM1071892 breast_primary_MGH482
GSM1071893 breast_primary_MGH963
GSM1071894 breast_primary_MGH455
GSM1071895 breast_primary_MGH1084
GSM1071896 breast_primary_MGH1038
GSM1071897 breast_primary_MGH1057
GSM1071898 breast_primary_MGH674
GSM1071899 breast_primary_MGH1102

Goal: Comparison of the gene expression between the two groups

PRE-PROCESSING



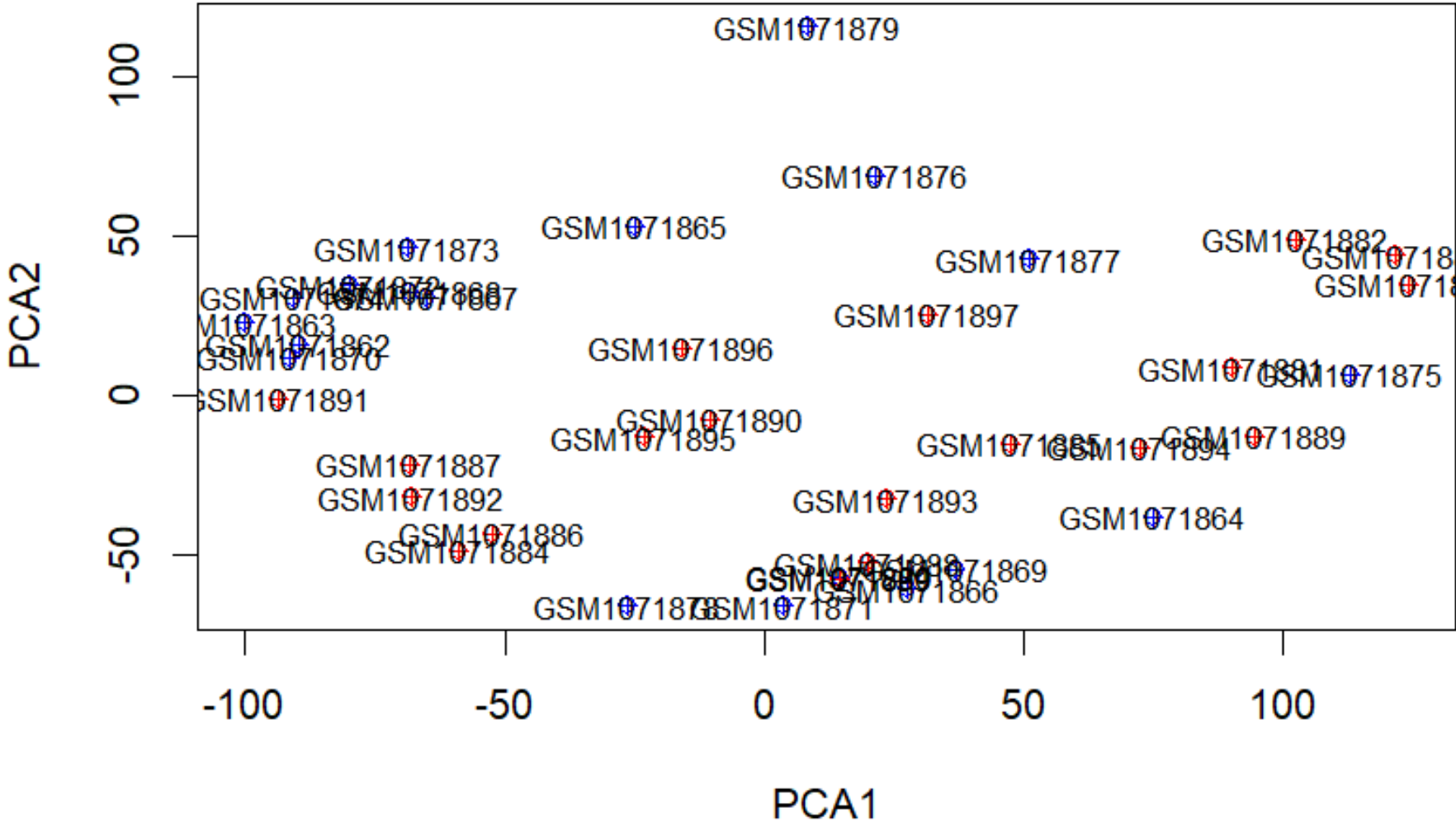
non-normalized dataset



normalized dataset

UNSUPERVISED LEARNING METHODS

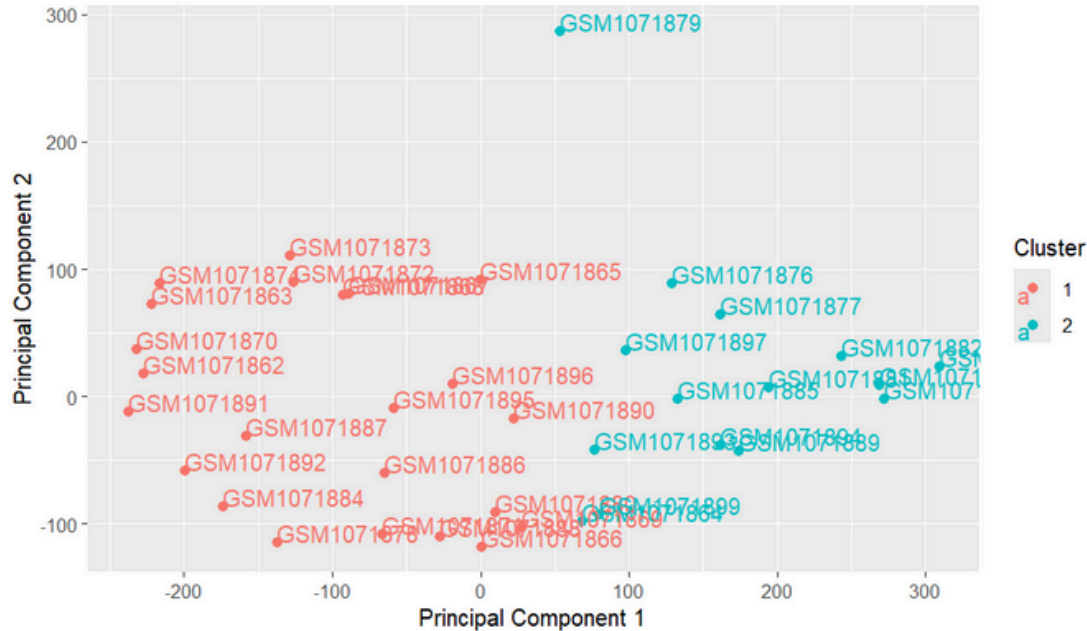
PCA for components 1&2



Red : nonmetastatic primary breast cancer (GSM1071881-GSM1071899)

Blue: breast cancer brain metastases (GSM1071862-GSM1071880)

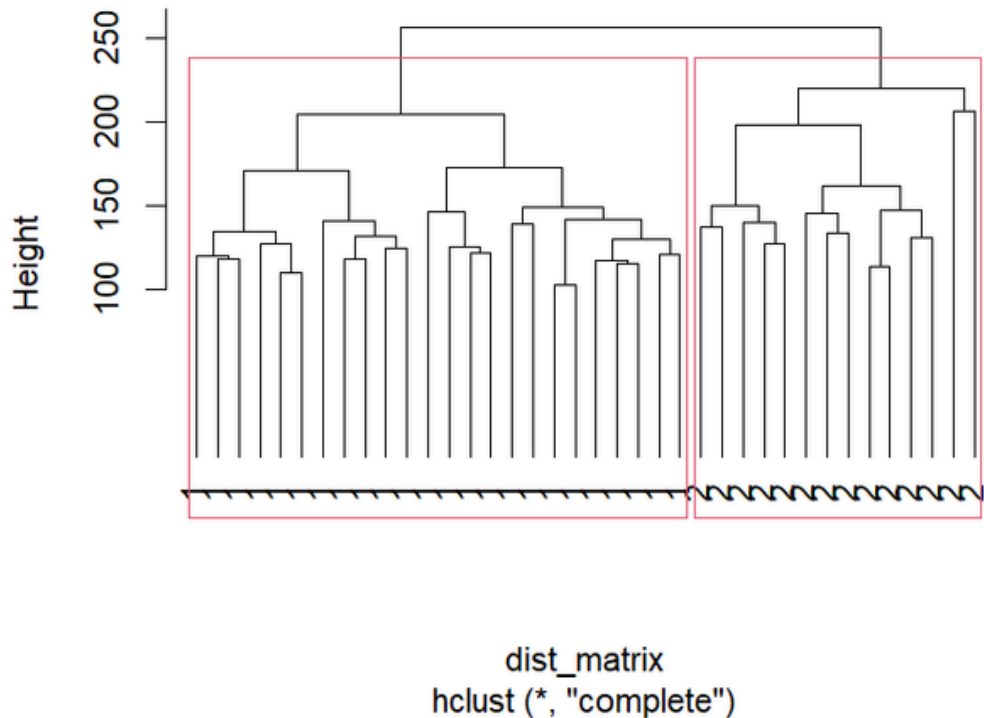
K-Means Results



g1: 23

g2: 15

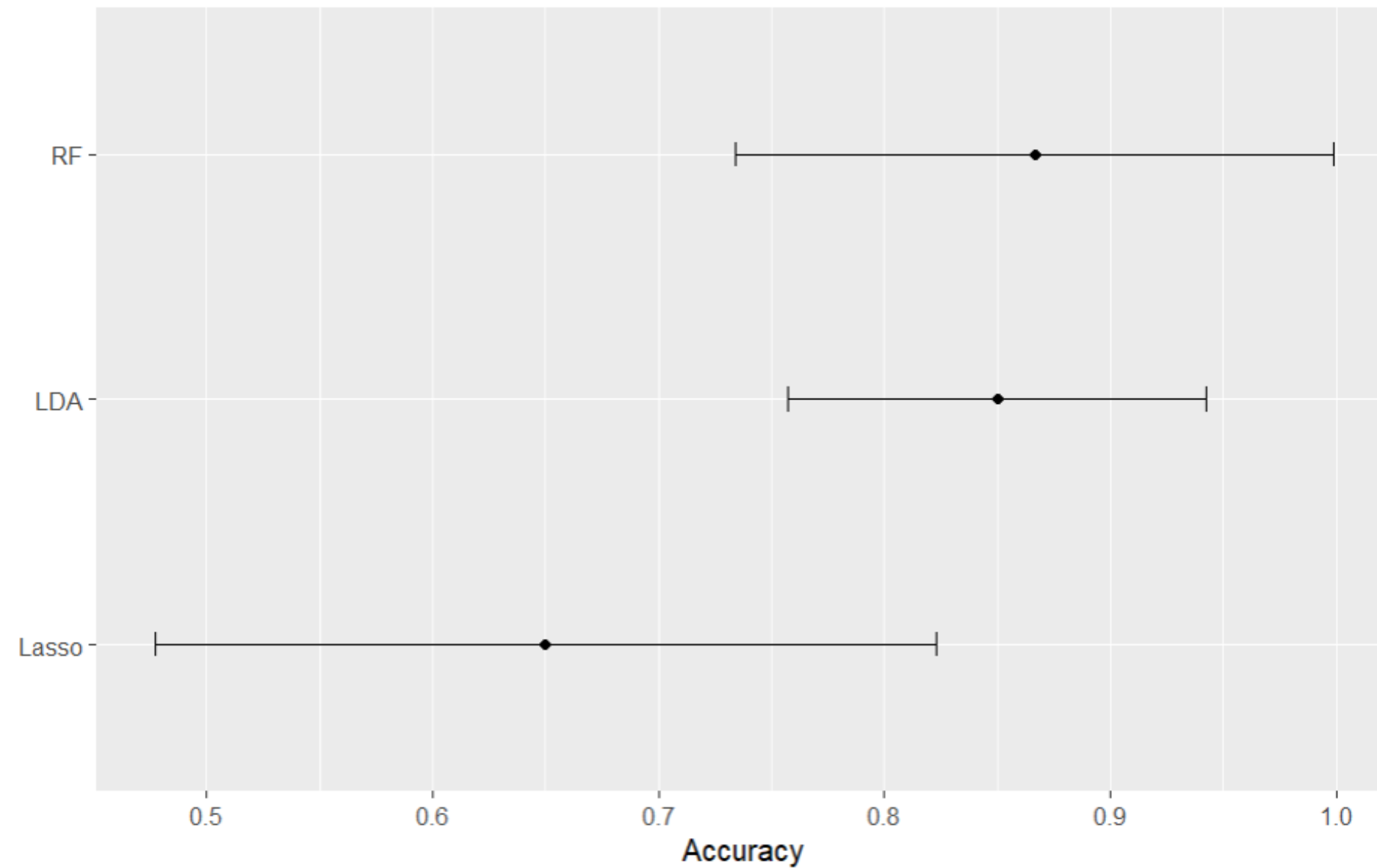
Cluster Dendrogram



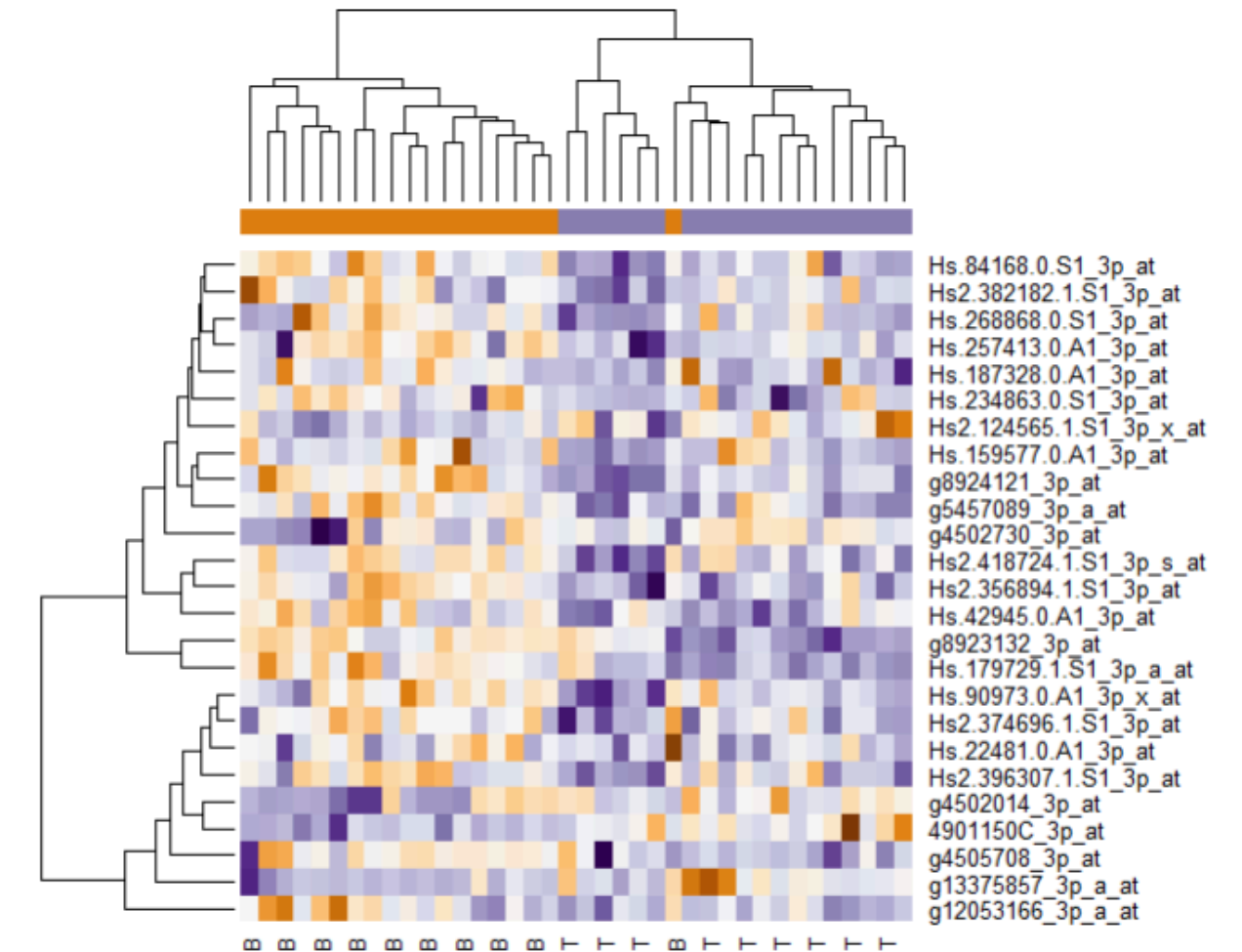
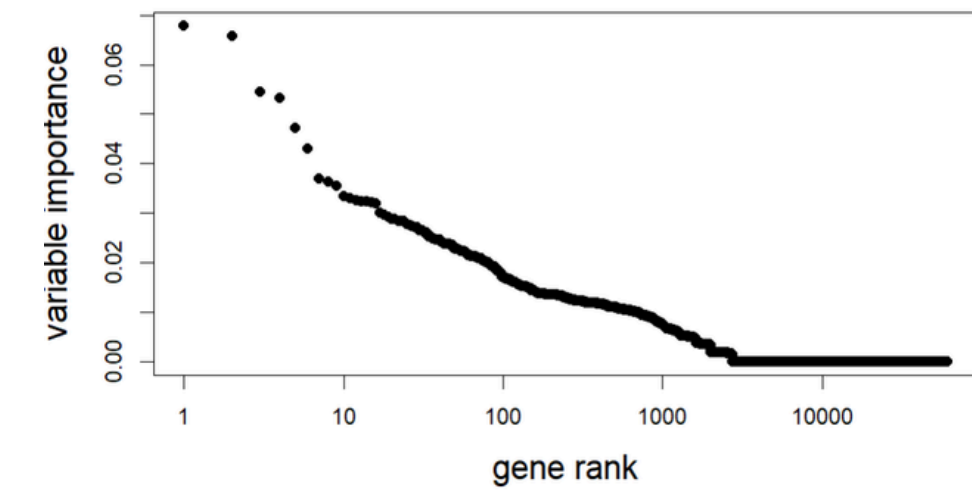
g1:24

g2:14

SUPERVISED LEARNING METHODS

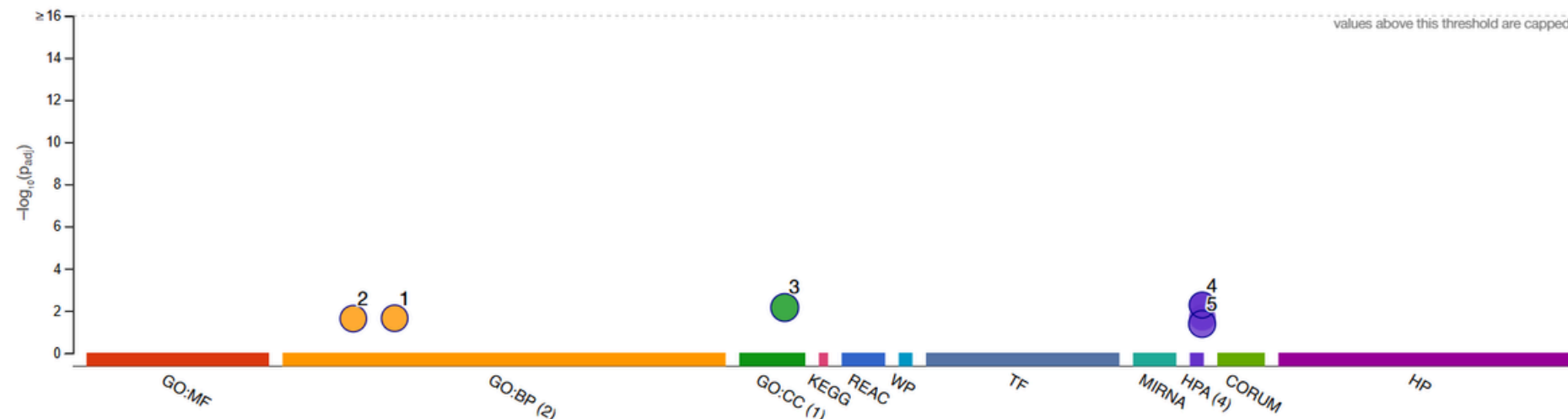


ALL subset results



FUNCTIONAL ENRICHMENT ANALYSIS

gProfiler



ID	Source	Term ID	Term Name	p _{adj} (query_1)
1	GO:BP	GO:0023051	regulation of signaling	2.308×10 ⁻²
2	GO:BP	GO:0010646	regulation of cell communication	2.407×10 ⁻²
3	GO:CC	GO:0071944	cell periphery	7.146×10 ⁻³
4	HPA	HPA:0500652	soft tissue 1; fibroblasts[≥Medium]	5.436×10 ⁻³
5	HPA	HPA:0500651	soft tissue 1; fibroblasts[≥Low]	4.298×10 ⁻²

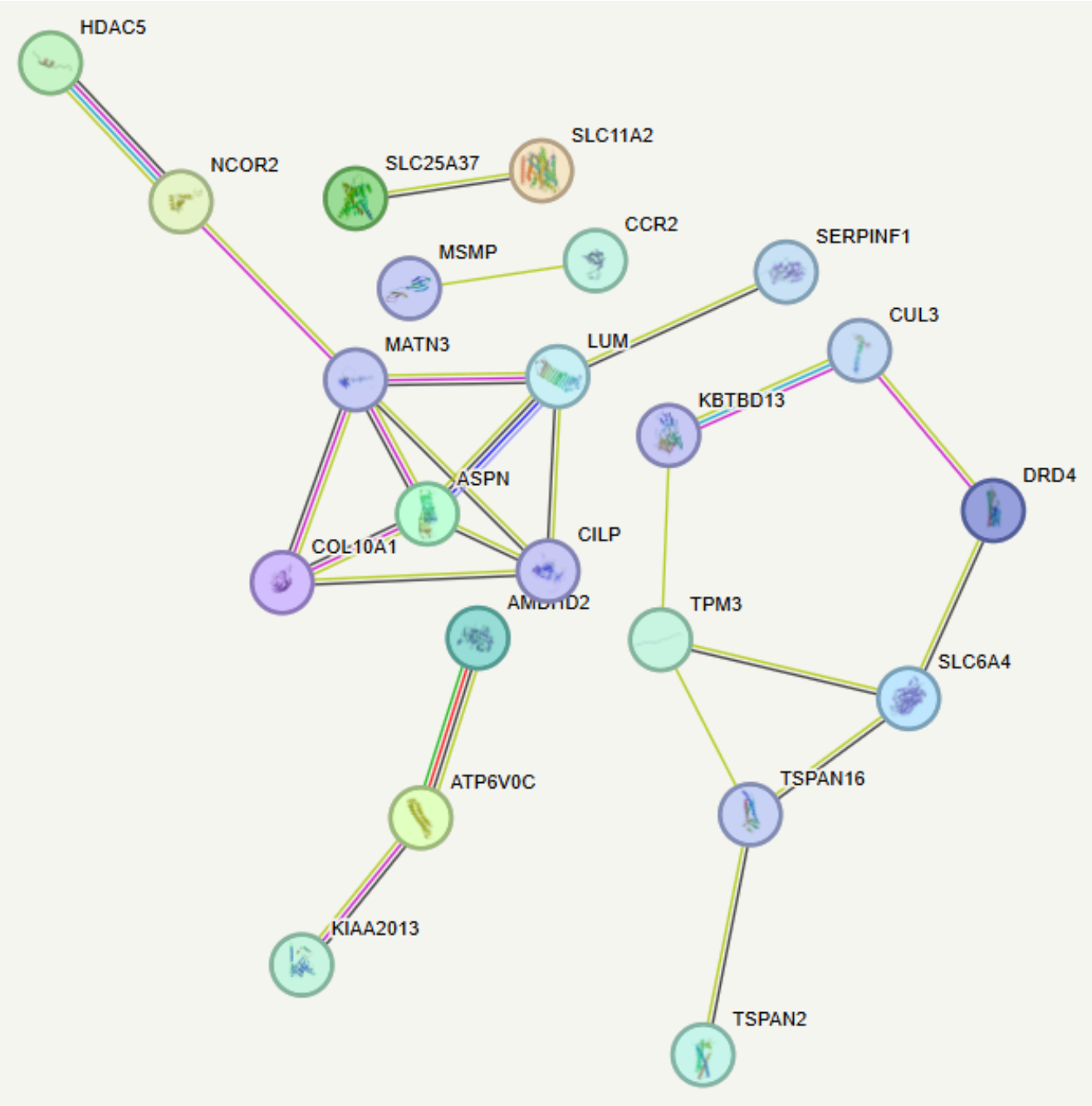
DAVID

44 chart records [Download File](#)

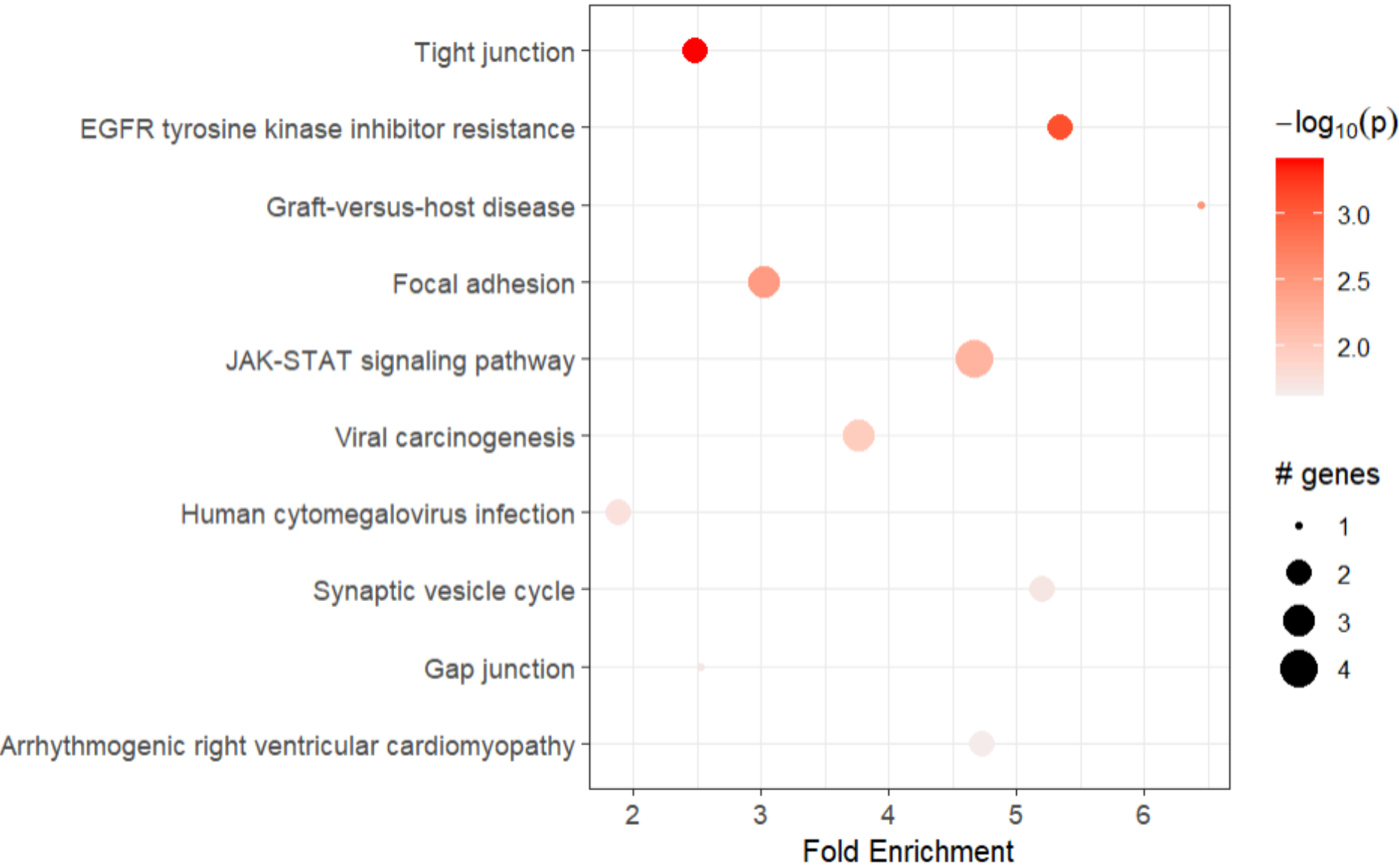
Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
<input type="checkbox"/>	GOTERM_CC_DIRECT	membrane	RT		54	33,5	1,3E-4	3,5E-2
<input type="checkbox"/>	GOTERM_CC_DIRECT	receptor complex	RT		7	4,3	3,9E-3	3,9E-1
<input type="checkbox"/>	GOTERM_CC_DIRECT	endomembrane system	RT		6	3,7	4,5E-3	3,9E-1
<input type="checkbox"/>	GOTERM_CC_DIRECT	nucleoplasm	RT		39	24,2	1,2E-2	7,5E-1
<input type="checkbox"/>	GOTERM_BP_DIRECT	positive regulation of kinase activity	RT		4	2,5	5,8E-3	1,0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	extracellular matrix organization	RT		6	3,7	6,0E-3	1,0E0
<input type="checkbox"/>	WIKIPATHWAYS	Ferroptosis	RT		4	2,5	1,6E-2	1,0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	microtubule cytoskeleton organization	RT		5	3,1	1,8E-2	1,0E0
<input type="checkbox"/>	GOTERM_MF_DIRECT	protein binding	RT		98	60,9	2,2E-2	1,0E0
<input type="checkbox"/>	WIKIPATHWAYS	Regulatory circuits of STAT3 signaling	RT		4	2,5	2,5E-2	1,0E0
<input type="checkbox"/>	GOTERM_CC_DIRECT	synaptic membrane	RT		3	1,9	3,1E-2	1,0E0
<input type="checkbox"/>	GOTERM_MF_DIRECT	protein tyrosine kinase activity	RT		4	2,5	3,3E-2	1,0E0
<input type="checkbox"/>	GOTERM_CC_DIRECT	cell junction	RT		5	3,1	3,6E-2	1,0E0
<input type="checkbox"/>	REACTOME_PATHWAY	MAPK family signaling cascades	RT		7	4,3	3,6E-2	1,0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	cell surface receptor protein tyrosine kinase signaling pathway	RT		4	2,5	3,8E-2	1,0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	cadmium ion transmembrane transport	RT		2	1,2	3,9E-2	1,0E0
<input type="checkbox"/>	GOTERM_MF_DIRECT	ferrous iron transmembrane transporter activity	RT		2	1,2	4,0E-2	1,0E0
<input type="checkbox"/>	GOTERM_CC_DIRECT	endosome membrane	RT		6	3,7	4,2E-2	1,0E0
<input type="checkbox"/>	GOTERM_MF_DIRECT	extracellular matrix structural constituent	RT		4	2,5	4,5E-2	1,0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	cellular defense response	RT		3	1,9	5,0E-2	1,0E0
<input type="checkbox"/>	WIKIPATHWAYS	Microtubule cytoskeleton regulation	RT		3	1,9	5,2E-2	1,0E0
<input type="checkbox"/>	GOTERM_CC_DIRECT	plasma membrane	RT		47	29,2	5,2E-2	1,0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	cobalt ion transport	RT		2	1,2	5,2E-2	1,0E0
<input type="checkbox"/>	GOTERM_MF_DIRECT	iron ion transmembrane transporter activity	RT		2	1,2	5,3E-2	1,0E0
<input type="checkbox"/>	GOTERM_CC_DIRECT	collagen-containing extracellular matrix	RT		7	4,3	5,4E-2	1,0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	positive regulation of protein localization to plasma membrane	RT		3	1,9	5,5E-2	1,0E0
<input type="checkbox"/>	REACTOME_PATHWAY	Transcriptional Regulation by TP53	RT		7	4,3	5,6E-2	1,0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	lung-associated mesenchyme development	RT		2	1,2	5,8E-2	1,0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	manganese ion transmembrane transport	RT		2	1,2	5,8E-2	1,0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	phosphorylation	RT		9	5,6	6,0E-2	1,0E0
<input type="checkbox"/>	EC_NUMBER	2.7.10.1	RT		3	1,9	6,4E-2	1,0E0
<input type="checkbox"/>	GOTERM_MF_DIRECT	unmethylated CpG binding	RT		2	1,2	6,5E-2	1,0E0
<input type="checkbox"/>	WIKIPATHWAYS	Wnt signaling	RT		4	2,5	6,5E-2	1,0E0
<input type="checkbox"/>	GOTERM_MF_DIRECT	antiporter activity	RT		3	1,9	6,9E-2	1,0E0
<input type="checkbox"/>	REACTOME_PATHWAY	Signal Transduction	RT		27	16,8	6,9E-2	1,0E0
<input type="checkbox"/>	GOTERM_CC_DIRECT	endosome	RT		6	3,7	7,0E-2	1,0E0
<input type="checkbox"/>	GOTERM_MF_DIRECT	serotonin binding	RT		2	1,2	7,2E-2	1,0E0
<input type="checkbox"/>	REACTOME_PATHWAY	Interleukin-21 signaling	RT		2	1,2	7,3E-2	1,0E0
<input type="checkbox"/>	GOTERM_CC_DIRECT	peroxisomal membrane	RT		3	1,9	7,4E-2	1,0E0
<input type="checkbox"/>	KEGG_PATHWAY	Regulation of actin cytoskeleton	RT		5	3,1	7,7E-2	1,0E0
<input type="checkbox"/>	KEGG_PATHWAY	Cytoskeleton in muscle cells	RT		5	3,1	7,9E-2	1,0E0
<input type="checkbox"/>	REACTOME_PATHWAY	Disease	RT		20	12,4	8,5E-2	1,0E0
<input type="checkbox"/>	GOTERM_MF_DIRECT	double-stranded RNA binding	RT		3	1,9	8,9E-2	1,0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	positive regulation of calcium ion transport into cytosol	RT		2	1,2	9,5E-2	1,0E0

FUNCTIONAL ENRICHMENT ANALYSIS

STRING



pathfindR



CONCLUSION AND FUTURE PERSPECTIVE

PCA is able to divide the two groups

RF has the best performance

In the enrichment analysis, not many enriched terms were **specifically connected** to breast cancer or brain metastasis.

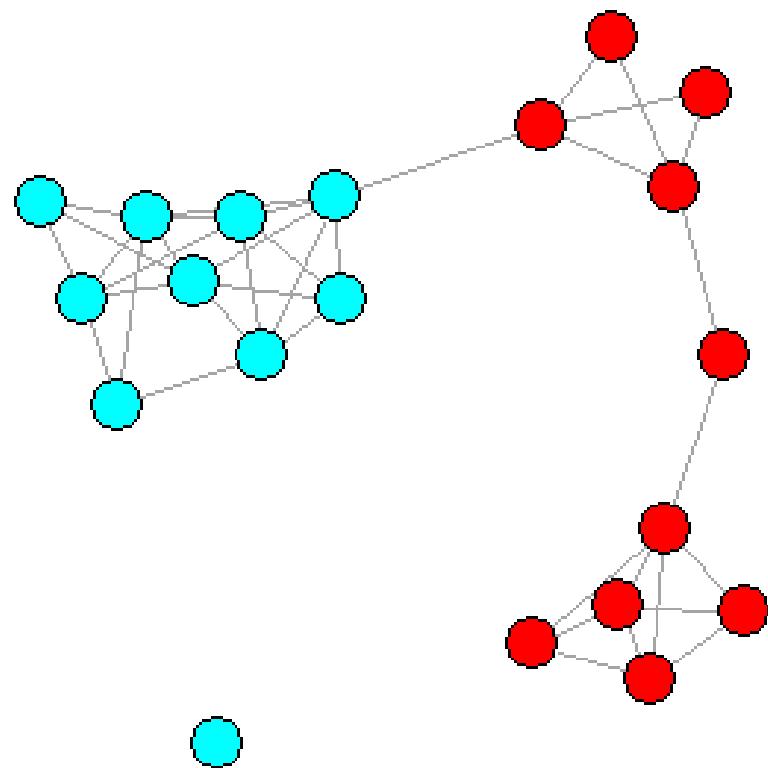
Repeat the analysis considering a **larger number** of samples

Repeat the analysis considering **BRCA1** gene

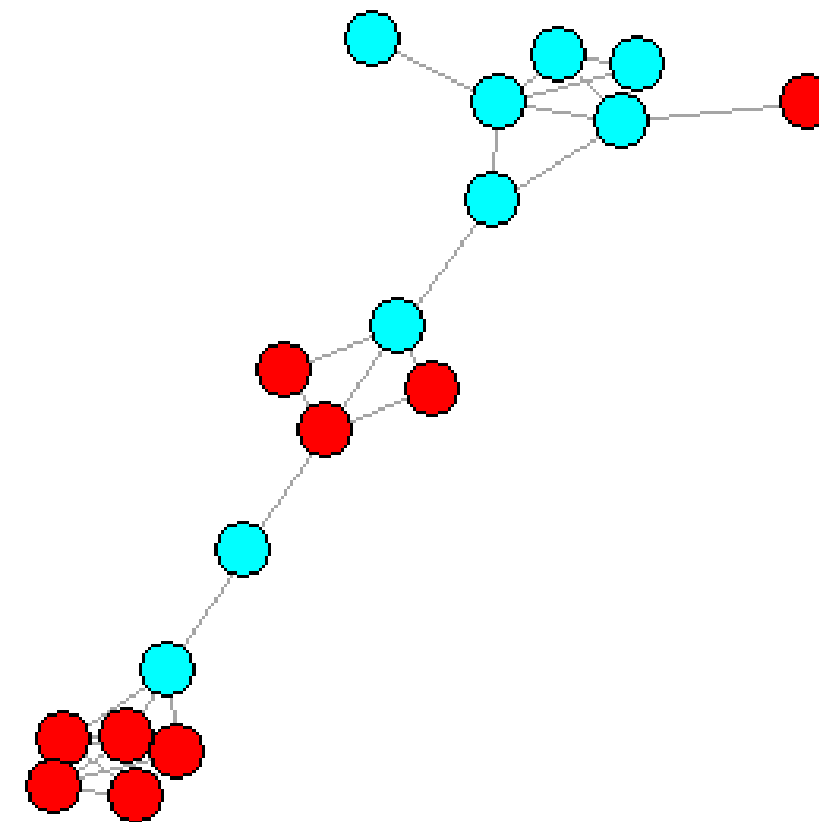
Repeat the analysis focusing on the genes highlighted by **rScudo**

THANK YOU

RSCUDO



■ Affected
■ Control



■ Affected
■ Control