

The background of the slide is a complex, abstract network of interconnected sticks or rods. These rods are colored in various shades including yellow, orange, red, blue, and green. They are arranged in a way that creates a sense of depth and a three-dimensional lattice structure, with some rods crossing over others. The overall effect is a dense, interconnected web of lines, symbolizing biological networks or data analysis.

Pathway mechanisms in lung cancer

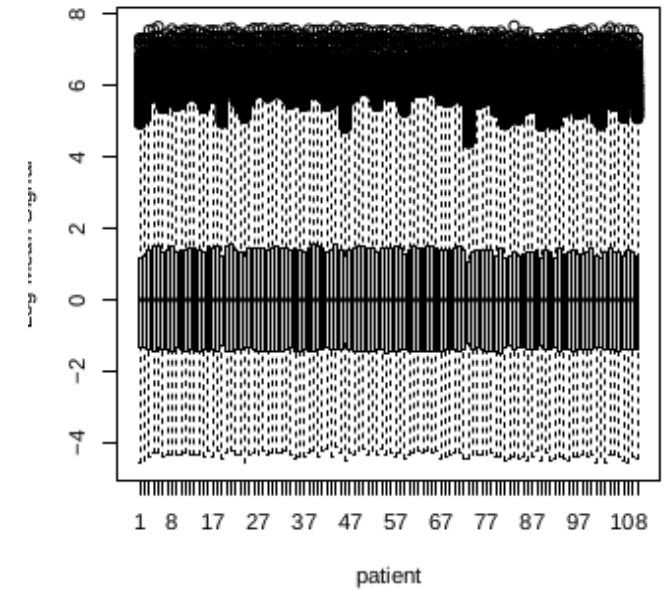
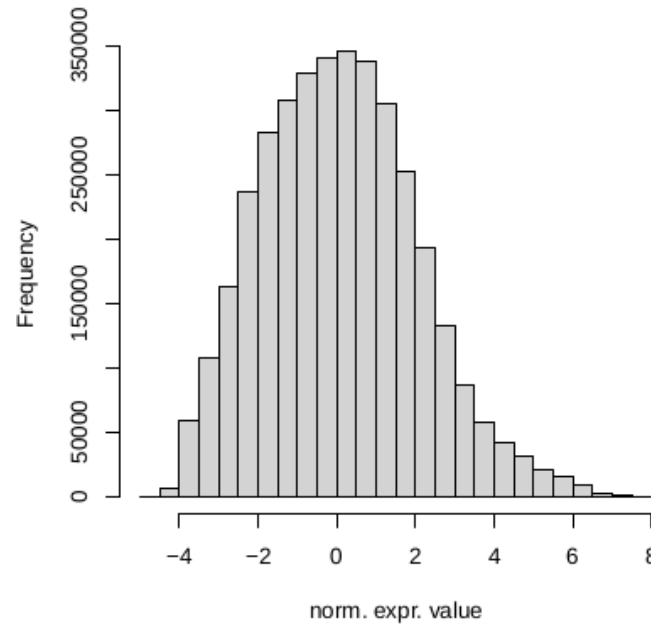
Giuliani Riccardo, Biological networks and data analysis project

Dataset

Data already log2 scale normalized, only median normalization applied

110 samples in total

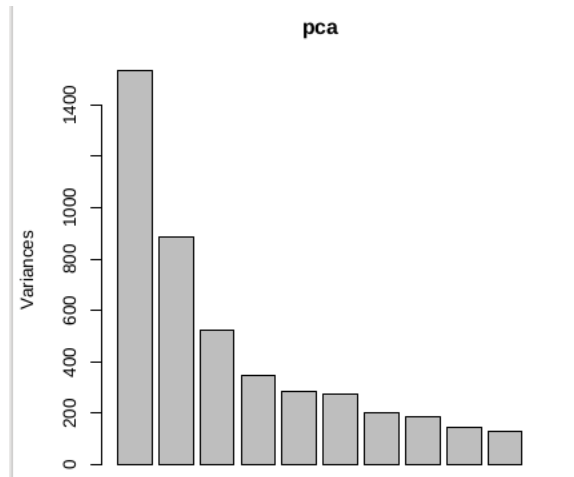
- 40 Tumor and smokers
- 40 Tumor and not-smokers
- 30 Normal controls



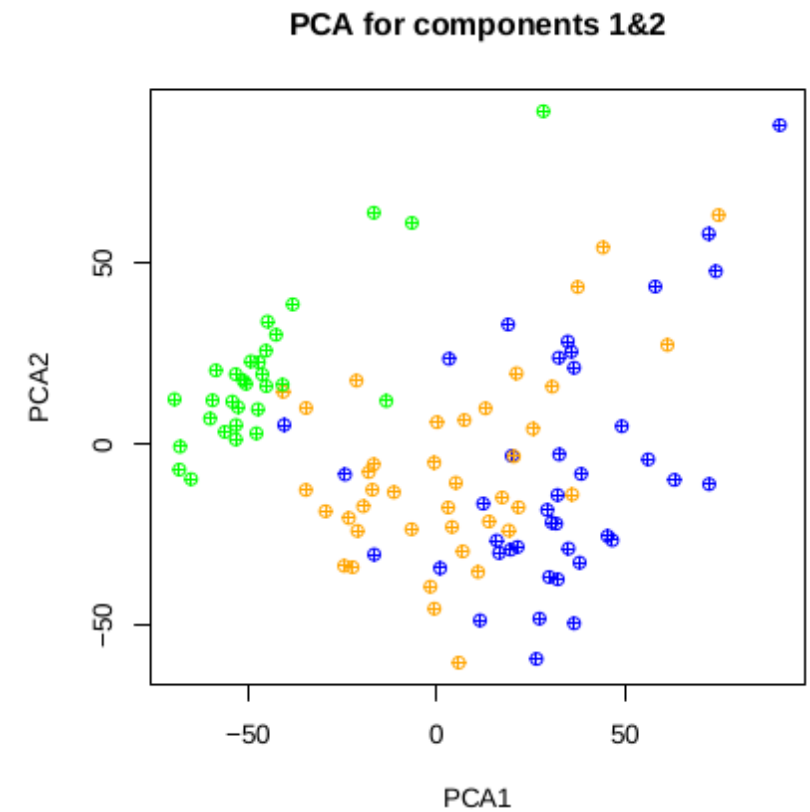
median normalized data

PCA

Unsupervised analysis to have a quick look at the dataset



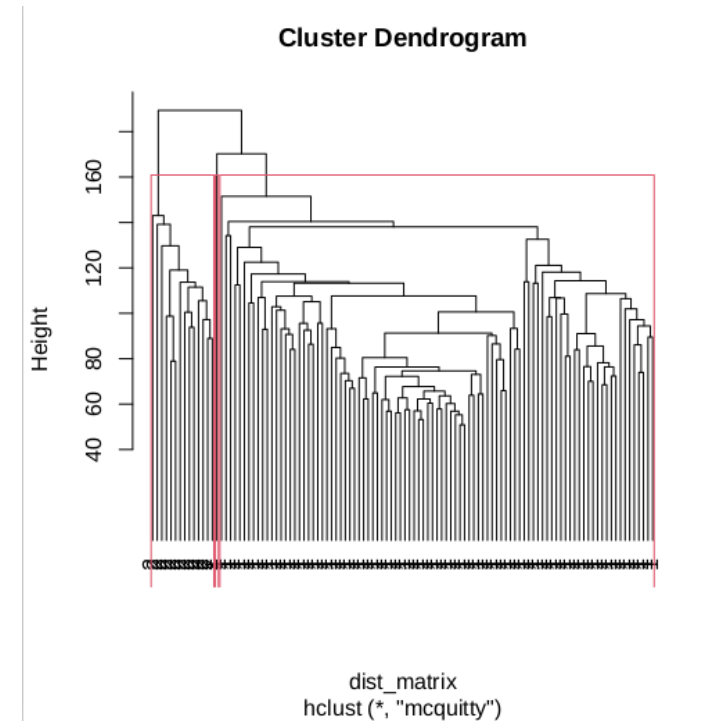
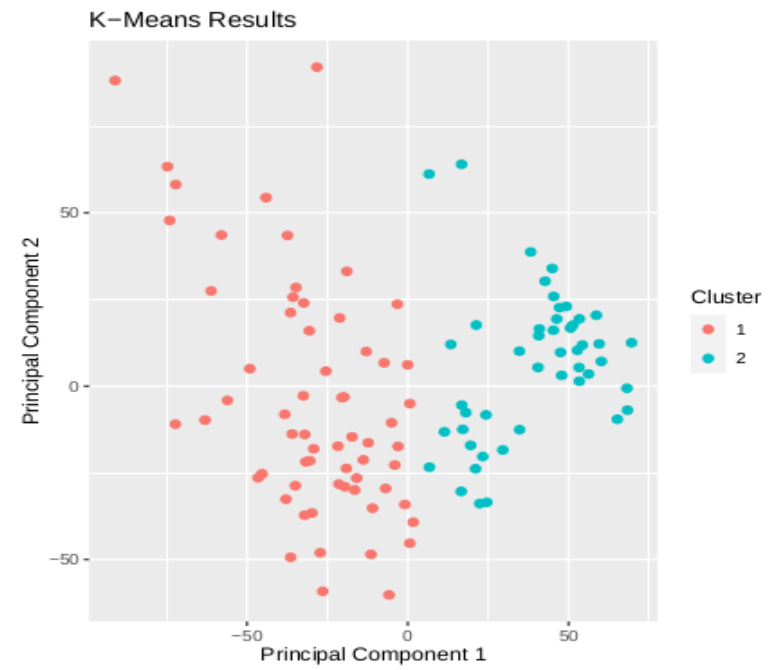
First two components only were chosen



Unsupervised algorithms

Work without any user supervision to clusterize data

K-means and **Hierarchical clustering**: they base on distance between data and clusters to perform data partitioning



None of them able to infer the presence of 3 clusters

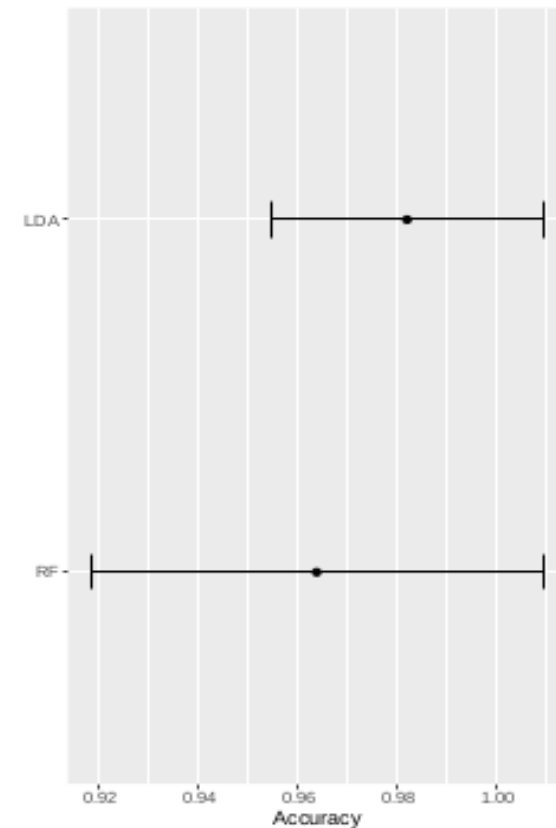
Supervised methods

Require a set of labelled data provided by the user to perform the analysis

Random Forest, LDA, LASSO and SCUDO: methods based on different mathematical concepts to build a predictive model





	RF	LDA	LASSO	SCUDO
sensitivity	0.98	0.99	1	0.95
specificity	0.97	0.97	1	0.90









Based on ability to divide data from tumoral to normal



Functional enrichment analysis

DAVID and **g:Profiler**: similar methods that given in input a sequence of genes query different databases to find possible relations

Enrichment Score: 5.01	G		Count	P_Value	Benjamini
Fatty acid degradation	RT		16	1.4E-17	2.1E-15
alcohol dehydrogenase activity, zinc-dependent	RT		6	1.6E-10	2.7E-8
ethanol oxidation	RT		7	2.0E-10	3.6E-8
metal ion-binding site:Zinc 1; catalytic	RT		7	2.3E-10	3.5E-8
alcohol dehydrogenase (NAD) activity	RT		6	5.4E-10	4.7E-8







<input type="checkbox"/>	alcohol dehydrogenase activity, zinc-dependent	GO:0004024	6.009×10 ⁻¹³	
<input type="checkbox"/>	adenylate cyclase activity	GO:0004016	6.765×10 ⁻¹³	
<input type="checkbox"/>	adrenergic receptor activity	GO:0004935	4.520×10 ⁻¹²	
<input type="checkbox"/>	small molecule binding	GO:0036094	9.665×10 ⁻¹²	
<input type="checkbox"/>	alcohol dehydrogenase (NAD+) activity	GO:0004022	2.137×10 ⁻¹¹	
<input type="checkbox"/>	alcohol dehydrogenase [NAD(P)+] activity	GO:0018455	7.080×10 ⁻¹¹	
<input type="checkbox"/>	alpha-adrenergic receptor activity	GO:0004936	8.577×10 ⁻¹¹	
<input type="checkbox"/>	cyclase activity	GO:0009975	1.225×10 ⁻⁹	

Only pathway databases selected for the analysis

200 most significant genes from random forest used

Network analysis

Cytoscape and **enrichnet** employ network analysis to explore the most promising pathways involving our genes

Annotation (pathway/process) ^	Significance of network distance distribution (XD-Score) ^	Significance of overlap (Fisher-test, q-value) ^	Dataset size (uploaded gene set) ^	Dataset size (pathway gene set) ^	Dataset size (overlap) ^
Fatty acid metabolism					
 compute graph visualization	3.0e+00	6.0e-14	129	41	15 (show)
 see mapped genes					
Riboflavin metabolism					
 compute graph visualization	2.0e+00	2.7e-03	129	16	4 (show)
 see mapped genes					
Propanoate metabolism					
 compute graph visualization	1.4e+00	6.1e-04	129	32	6 (show)
 see mapped genes					

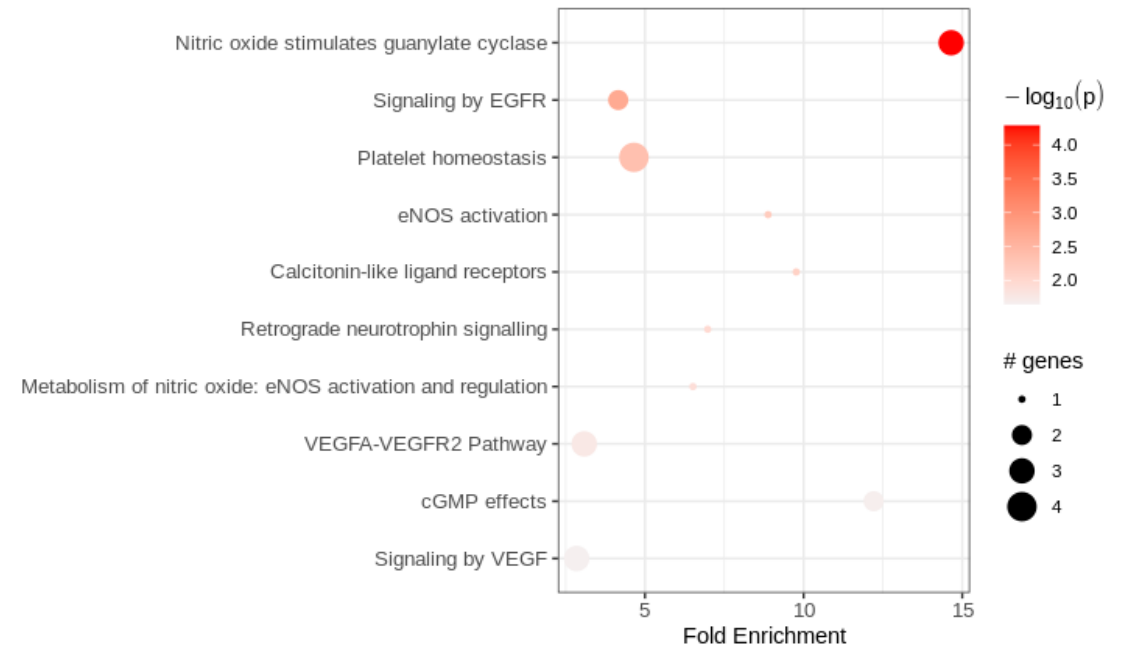
Pathway or Process	XD-score
Fatty acid metabolism	1.74552
Retinol metabolism	1.61090
Drug metabolism - cytochrome P450	1.09620
Tyrosine metabolism	1.09620
Riboflavin metabolism	0.95181
Propanoate metabolism	0.91354

Fatty acid, propanoate, riboflavin, tyrosine metabolisms

PathfindR

PathfindR: another way to perform network analysis, has provided different results!

ID	Term_Description	Fold_Enrichment	occurrence	lowest_p	highest_p
R-HSA-392154	Nitric oxide stimulates guanylate cyclase	14.650633	10	5.2e-05	5.2e-05
R-HSA-177929	Signaling by EGFR	4.156208	10	2.1e-03	2.1e-03
R-HSA-418346	Platelet homeostasis	4.650995	10	4.3e-03	4.3e-03
R-HSA-203615	eNOS activation	8.879171	10	6.5e-03	6.5e-03
R-HSA-419812	Calcitonin-like ligand receptors	9.767089	10	8.0e-03	8.0e-03



Only 113 over 196 genes correctly converted

Conclusions

- Unsupervised learning didn't seem to perform very well in that case
- Supervised algorithms provided high level accuracy in the analysis
- Discordant results in gene enrichment and network analysis

--> Further analysis required to validate the results

Thank you
for
the attention

ADDITIONAL FIGURES

