

TCGA multi-modal data compilation and pan- cancer data analysis

Presented Ricky Nguyen (z5309282)
by:

Supervisor: Fatemeh Vafaei

Assessor: Sara Ballouz

Content Structure

- **Project Recap**
 - Background and Project Goals
 - Project Outline
 - Overview of entire project pipeline
 - Overview of cancers and data modalities
 - Overview of data preprocessing and ML algorithms
- **Single-omics pipeline**
 - Evaluation of Individual FS Methods
 - Feature Selection - CV
 - Feature Selection - Bootstrapping
 - Performance evaluation + current results
- **Multi-omics pipeline**
 - Early Fusion
 - Late Fusion
- **Pan-cancer Analysis**
 - Overlapping Features + DA + GSEA
 - KM Plots
- **Conclusion**

Background

- Cancer is a complex disease, and analysing a single omics data type is insufficient.
 - Heterogeneity
 - Provide only partial insights.
- Combining multiple omics data types is essential for a holistic understanding of cancer's complex biology and for personalized treatment strategies.
- Survival analysis can help identify factors that influence survival outcomes

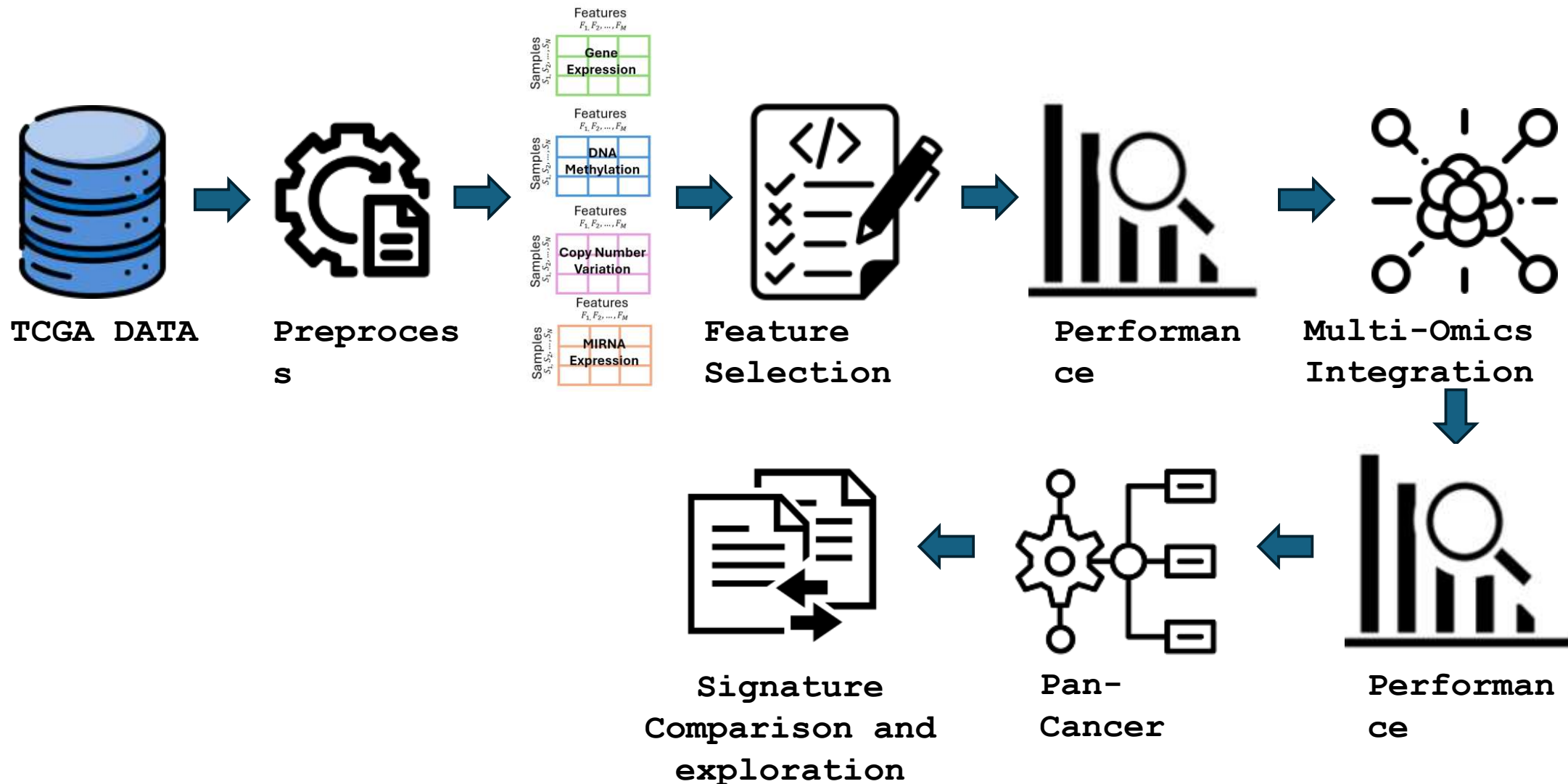
Project Goal

- This project focuses on utilising multi-omics integration methods on The Cancer Genome Atlas (TCGA) data to improve survival analysis and predictions across various women-related cancers.
- By employing machine learning methodologies, the project aims to identify important features and multi-modal signatures that are shared across different cancer types.
- These signatures will be crucial for enhancing our understanding of the factors that contribute to survival outcomes in cancer patients, potentially leading to better diagnostic and therapeutic strategies.

Project Outline

- Preprocess multiple omics data types (Gene expression, DNA methylation, miRNA expression, and Copy Number Variations).
- Utilise machine learning algorithms combined with feature selection methods for survival analysis on single-omics TCGA data.
- From the features obtained each single-omics after feature selection, perform multi-omics integration (early + late fusion). Investigate modality combination with the highest performance.
- Repeat this for all cancers, investigate any commonalities and explore these signatures further.

Entire Project Pipeline



Overview of data across cancers investigated

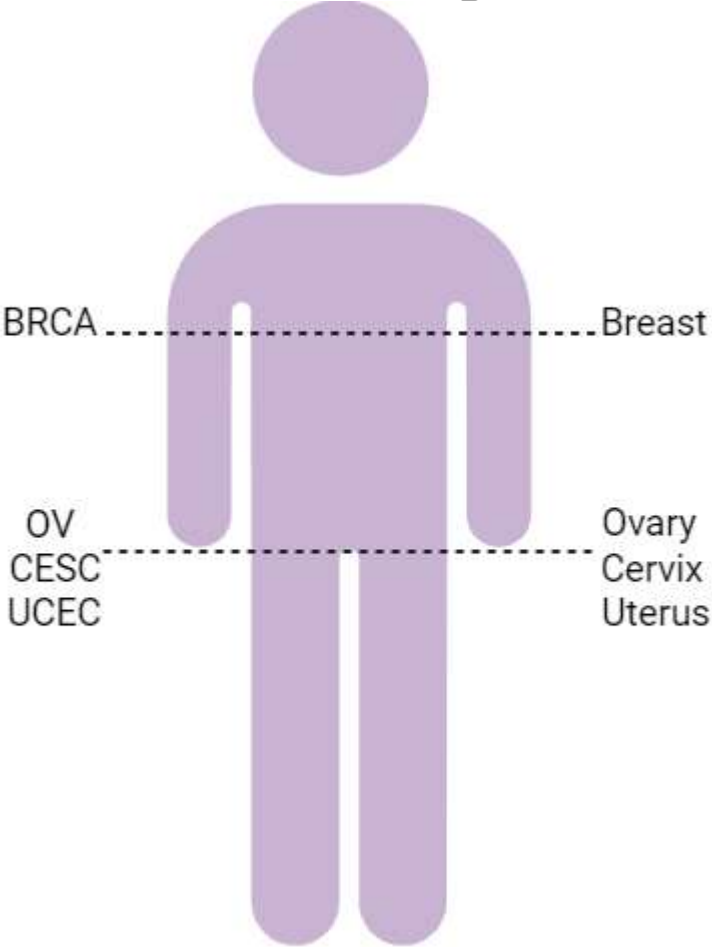
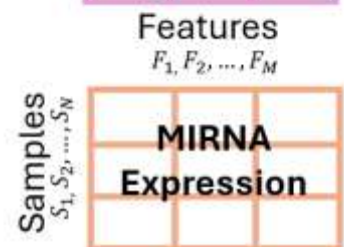
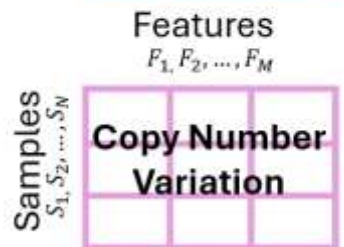
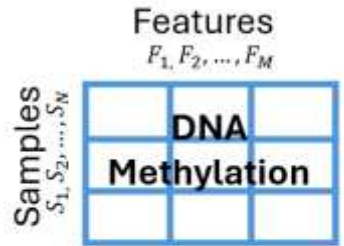
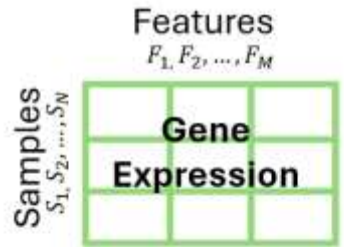


Table 2 Overview of number of samples for
~~all omics data~~

Cancer	GE	CNV	DM	ME	Clinical	Common
BRCA: Breast Invasive Carcinoma	1111	105 0	109 6	1096	1098	739
OV: Ovarian Serous Cystadenocarcinoma	421	557	582	490	608	397
CESC: Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma	304	294	307	307	307	291
UCEC: Uterine Corpus Endometrial Carcinoma	553	536	438	545	560	421

Overview of omics data modalities



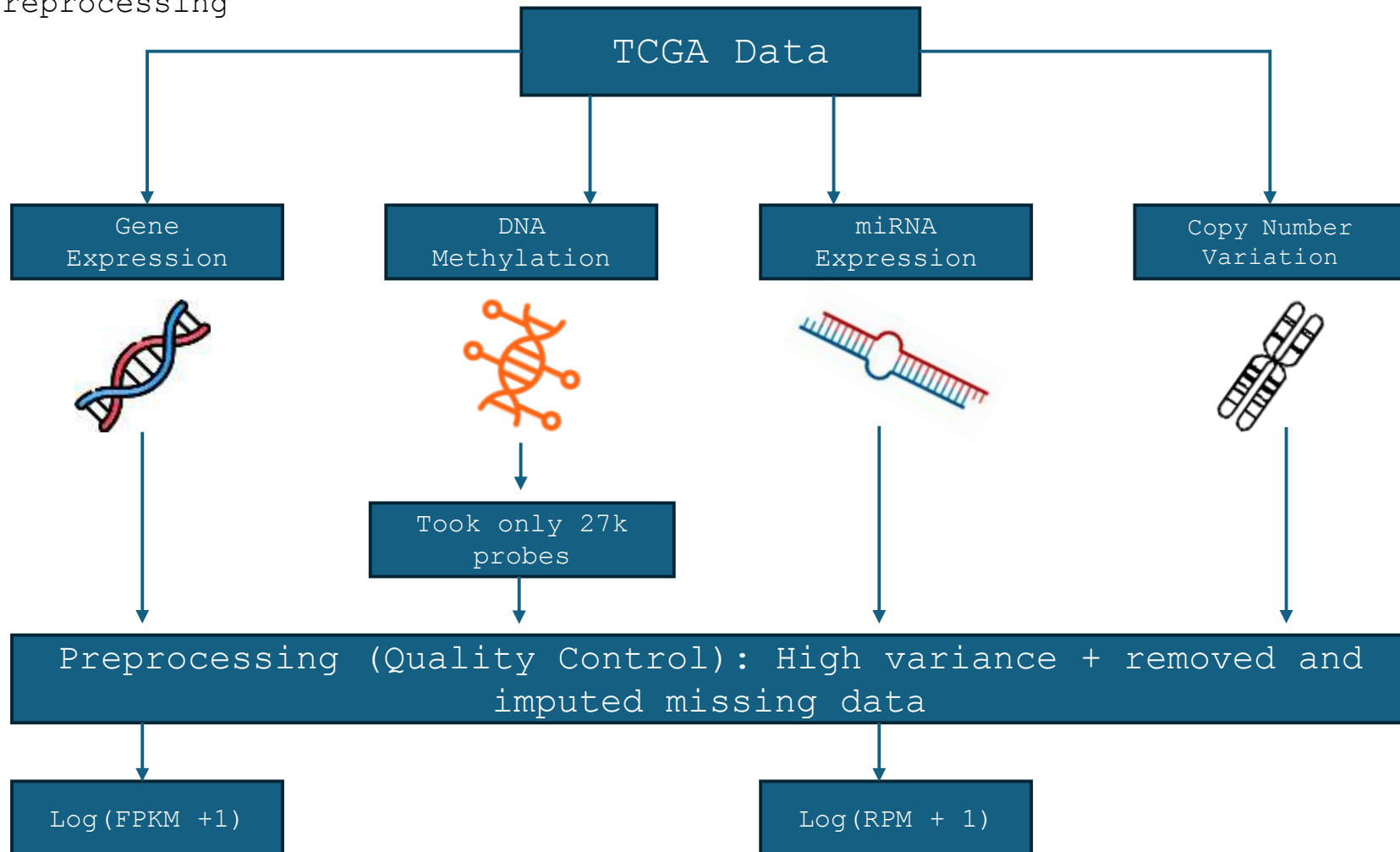
The Multi-omics data in this project were obtained from The Cancer Genome Atlas (TCGA).

Table 1 Overview of four omics data modalities

Data Modality	Gene Expression (GE)	DNA Methylation (DM)	Copy Number Variation (CNV)	MiRNA expression (ME)
Measure	Fragments per kilobase of transcript per million mapped reads (FPKM)	Beta Value	Gain/Loss/Neutral	Reads per million mapped reads (RPM)
Type	Continuous	Continuous	Discrete	Continuous
Range	[0, Billions]	[0,1]	[Loss < 0; Neutral = 0; Gain > 0]	[0, Millions]
Features	Ensembl Gene ID	cg Probe ID	Ensembl Gene ID	MiRNA ID

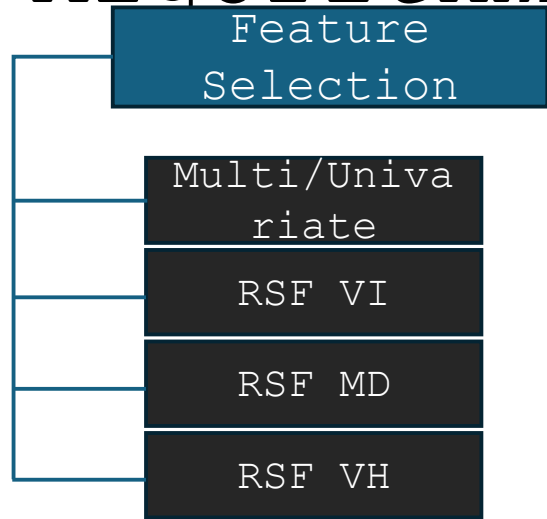
Data Preprocessing

Figure 2 Overview of omics data preprocessing



- Only using samples representing "Primary Solid Tumour" .
- Averaging for multiple signals
- Integrate clinical survival data

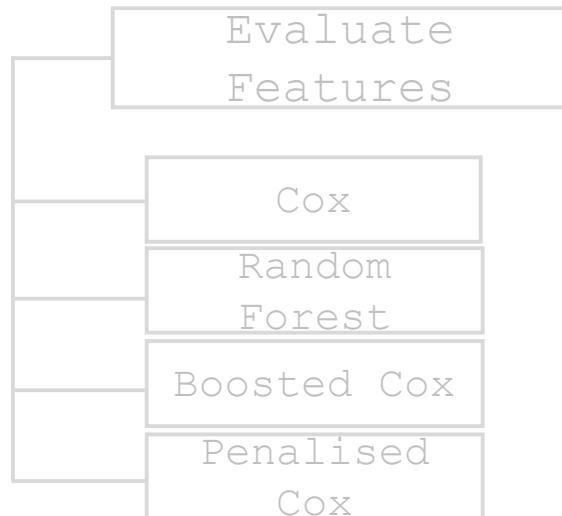
Overview of Feature Selection and ML Algorithms



Feature selection process that selects a subset of relevant features, useful for analysing high-dimensional data

Multi/Univariate:

- Fits a Cox proportional hazards model. Multi considers all features together, whereas uni looks at each feature separately.
- Ranked by C-index



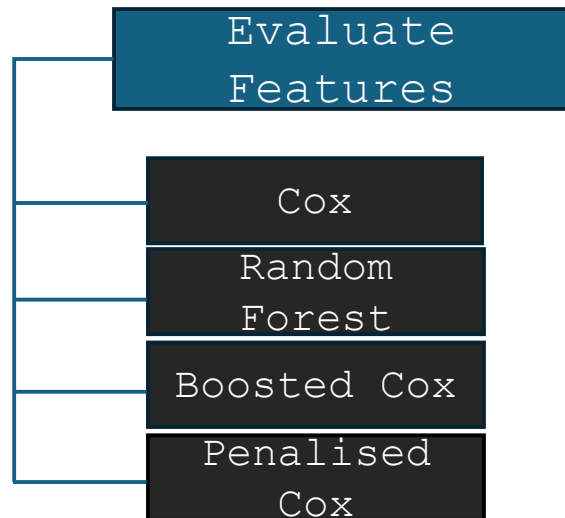
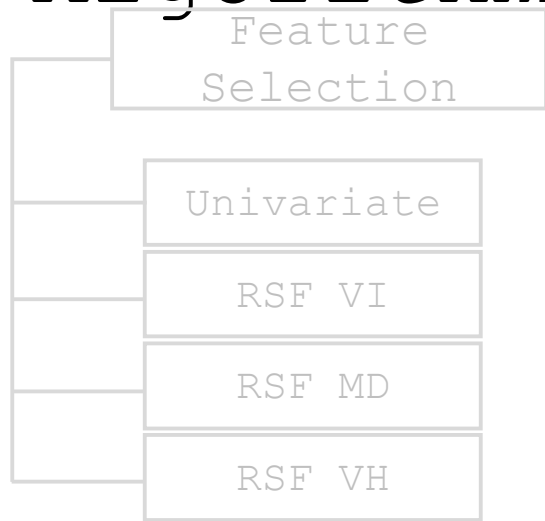
Random Forest (Variable Importance):

- Adding noise to a feature and looking at performance after
- If performance significantly low = highly predictive feature

Random Forest (Minimal depth):

- Looks at which feature's maximal subtree is closest to the root of the tree.
- Closer = more important.

Overview of Feature Selection and ML Algorithms



Cox Proportional Hazards Model:

- Standard tool for analysing survival data
- Evaluates the effect of features on the time to an event of interest

Random Forest:

- Aggregates results from many decision trees
- Feature and split point chosen on the one that maximises survival difference.

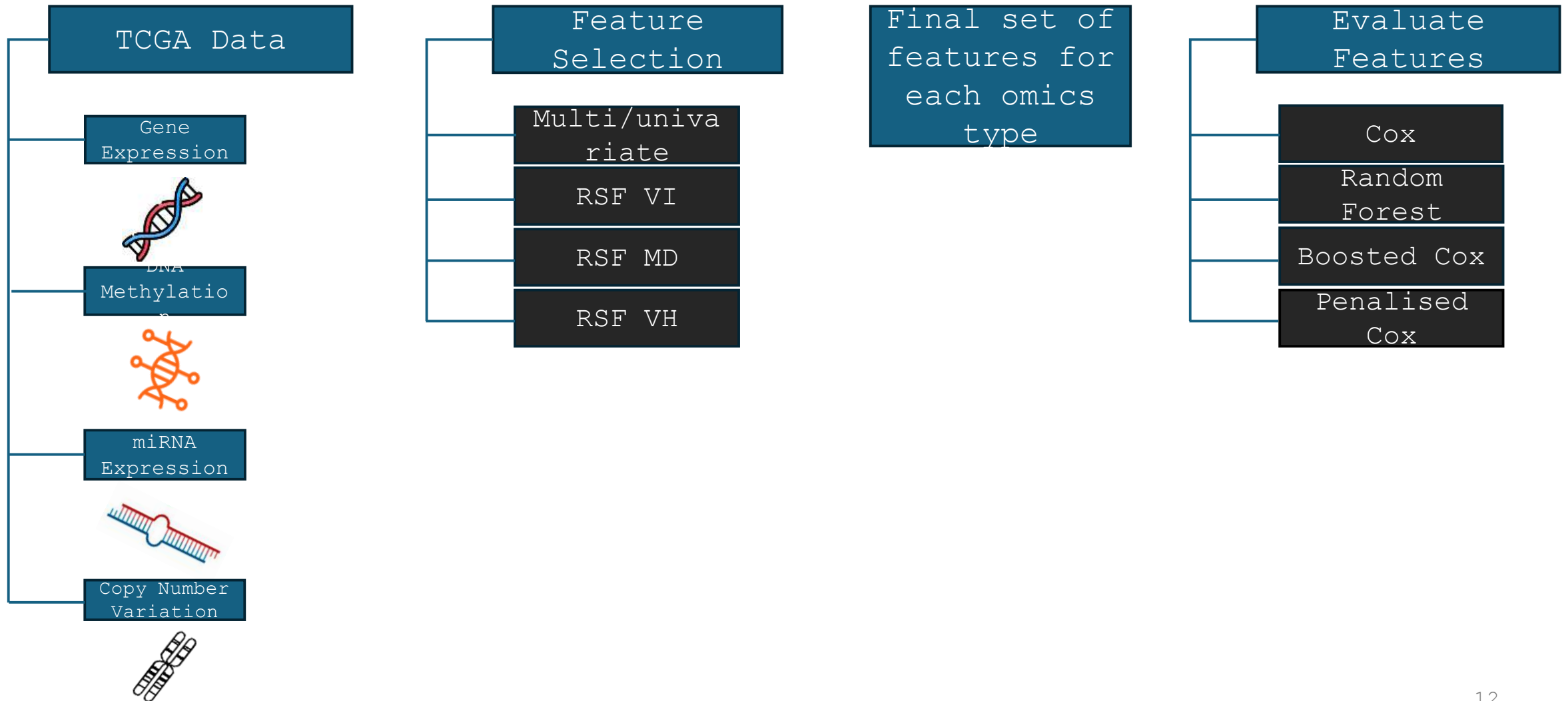
Boosted Cox:

- Ensemble technique that learns from the mistakes of previous models
- GLMBoost - uses a penalised cox as its base learner

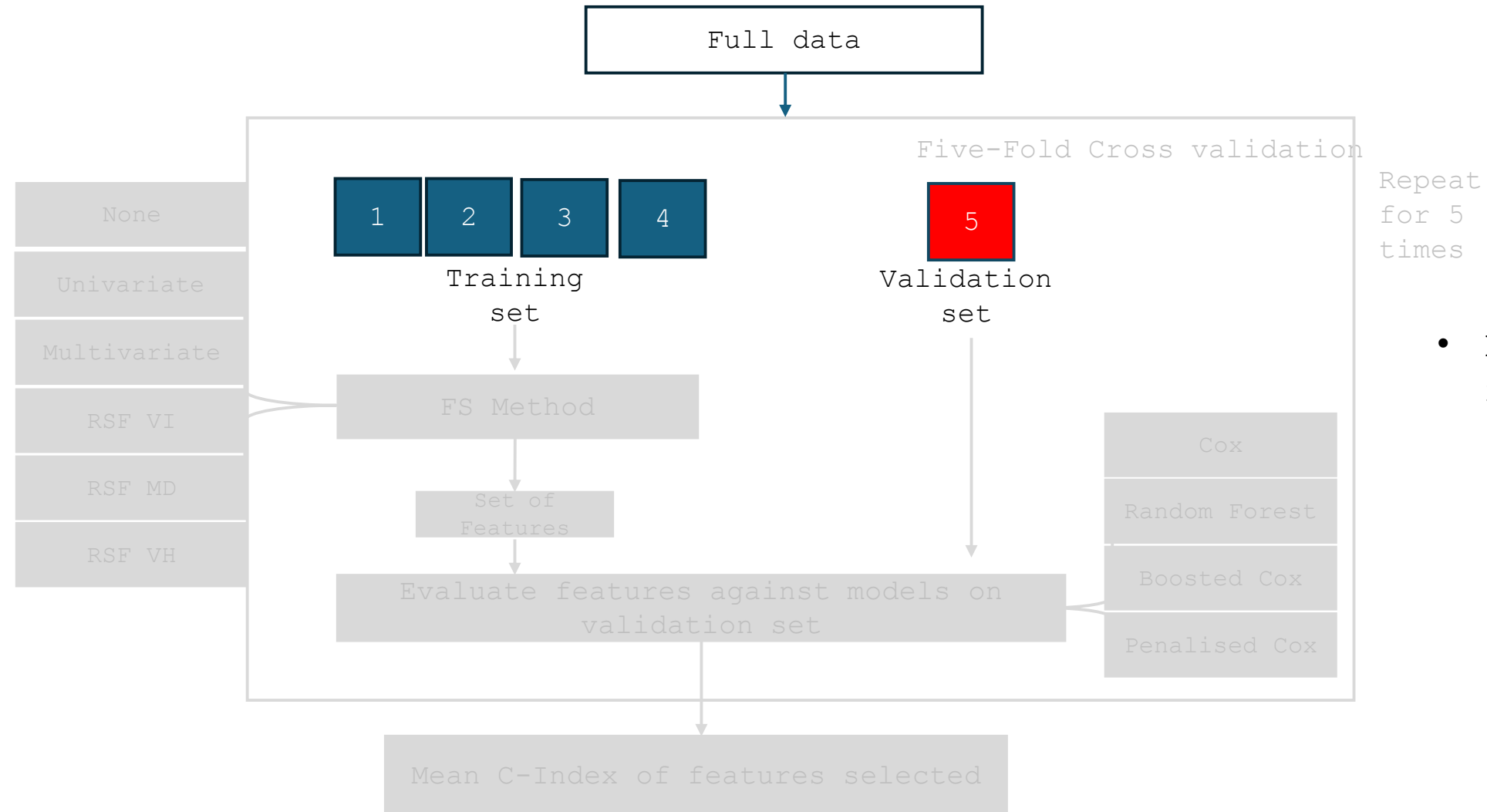
Penalised Cox:

- Regularises the model by reducing the coefficients towards zero
- Less important features have less impact.

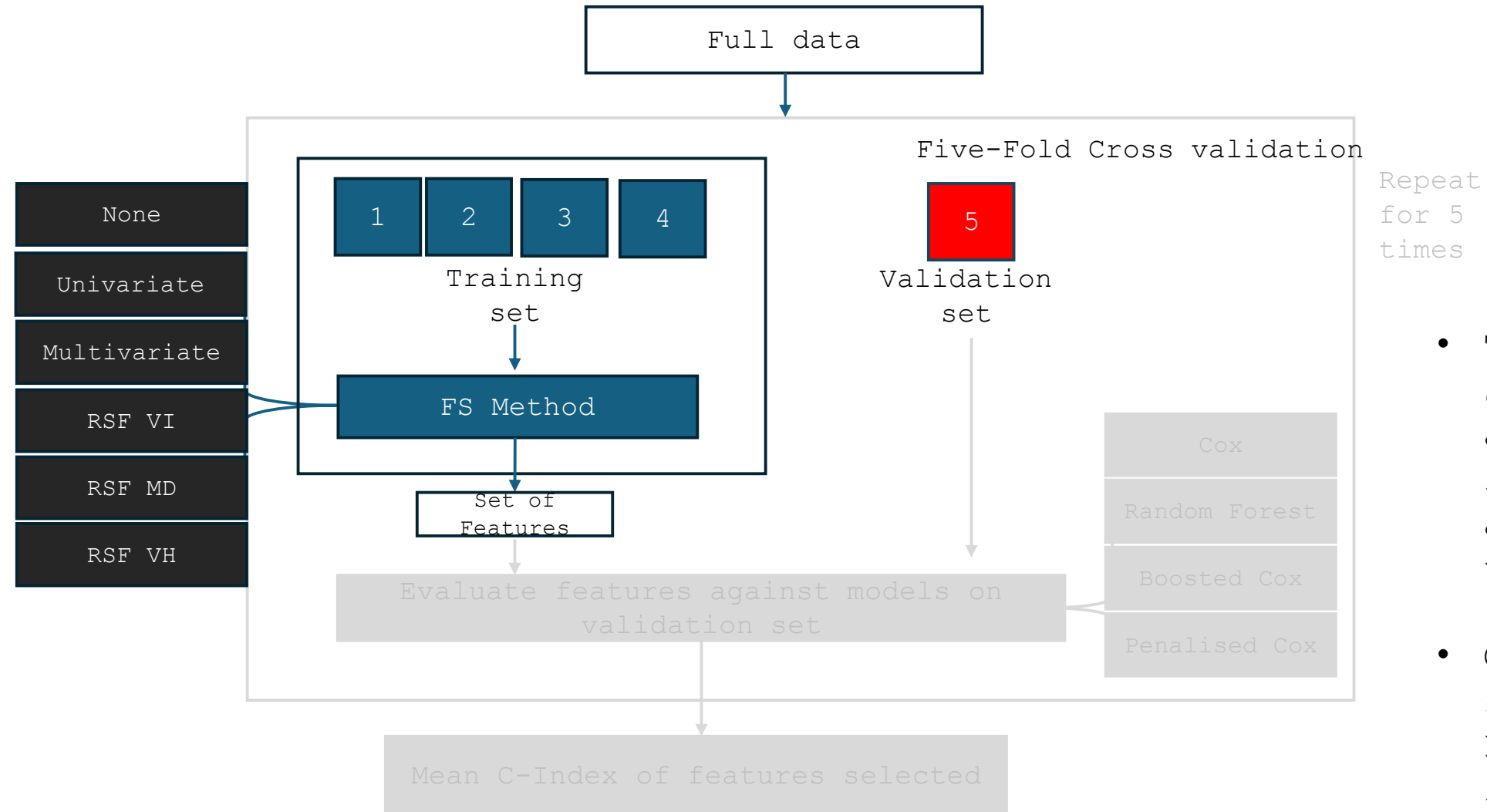
Overview of Single-Omics Pipeline



Evaluation of Individual FS Methods Pipeline

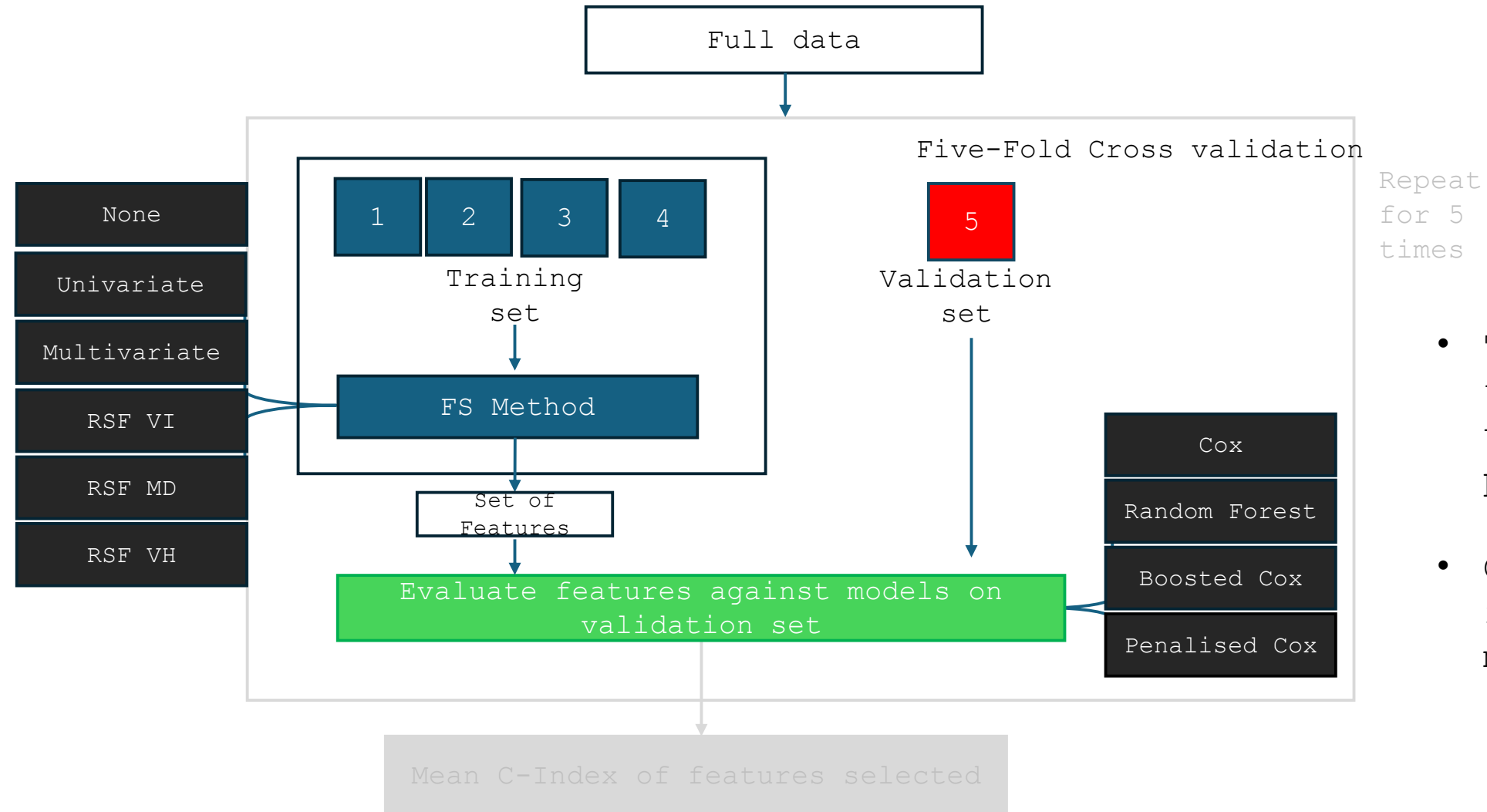


Evaluation of Individual FS Methods Pipeline



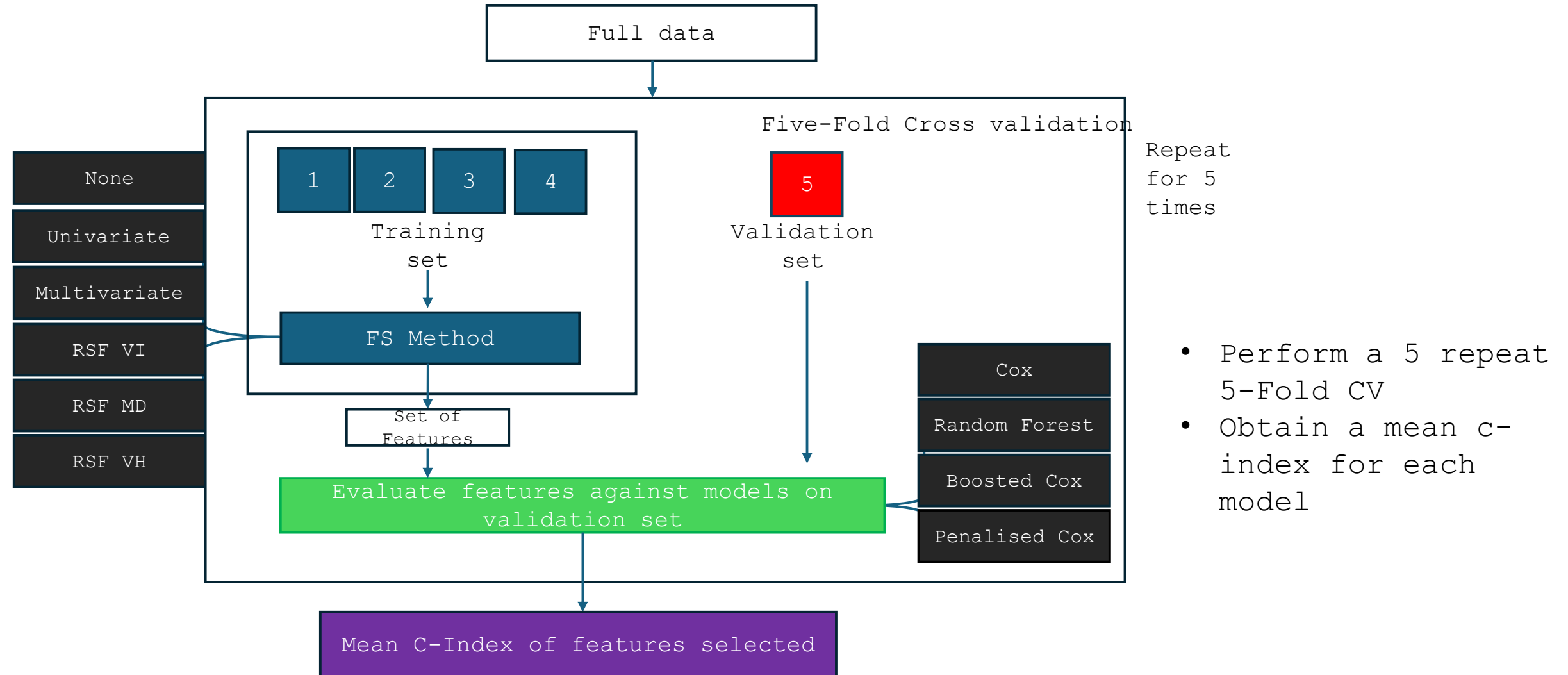
- The feature selection methods are then individually applied on the training set
- Obtain a set of features that the FS method deemed important

Evaluation of Individual FS Methods Pipeline



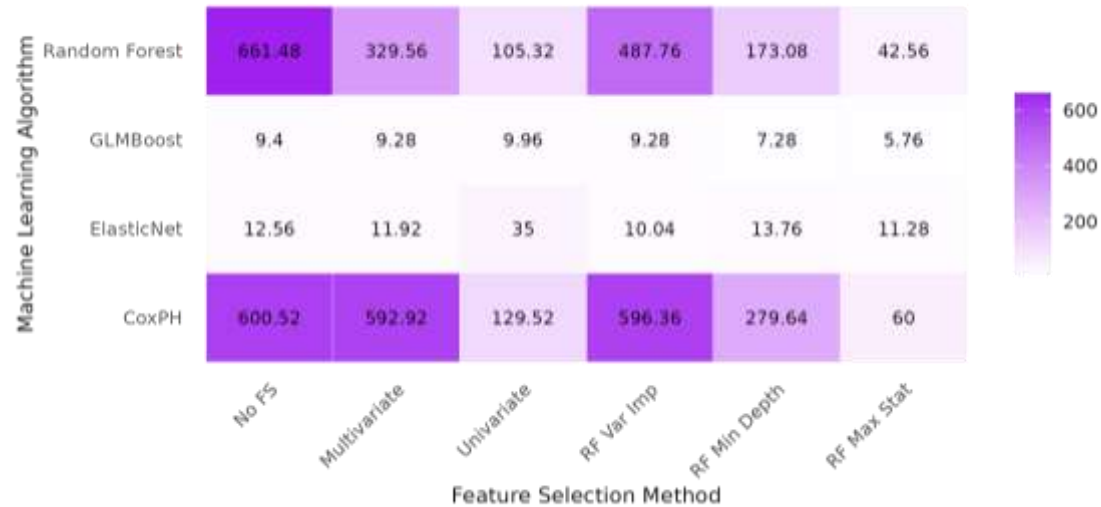
- These features are then evaluated using all four predictive models
- Obtain a c-index from each of the models

Evaluation of Individual FS Methods Pipeline

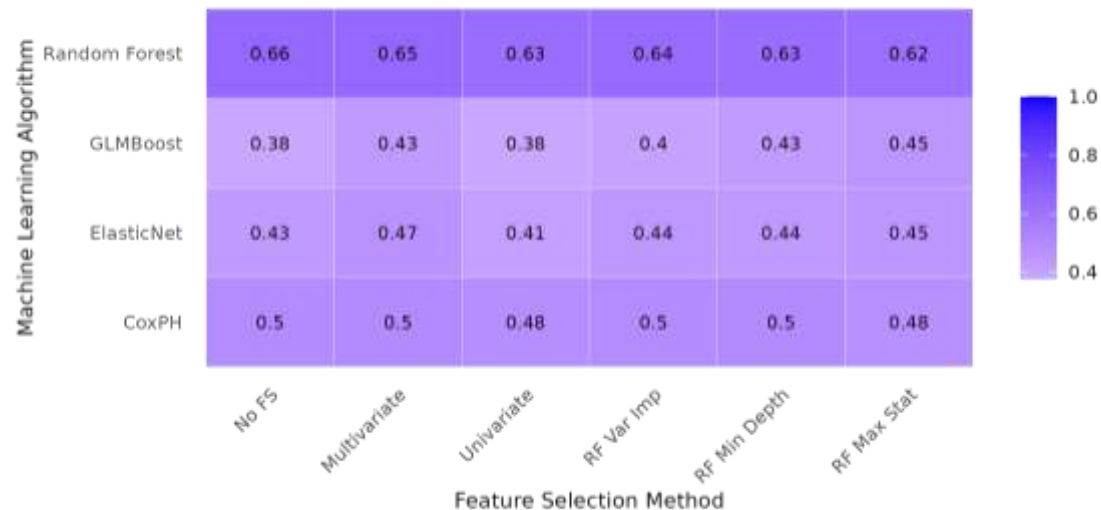


Evaluation of Individual FS Methods (BRCA)

Mean Features Selected by Models + Feature Selection



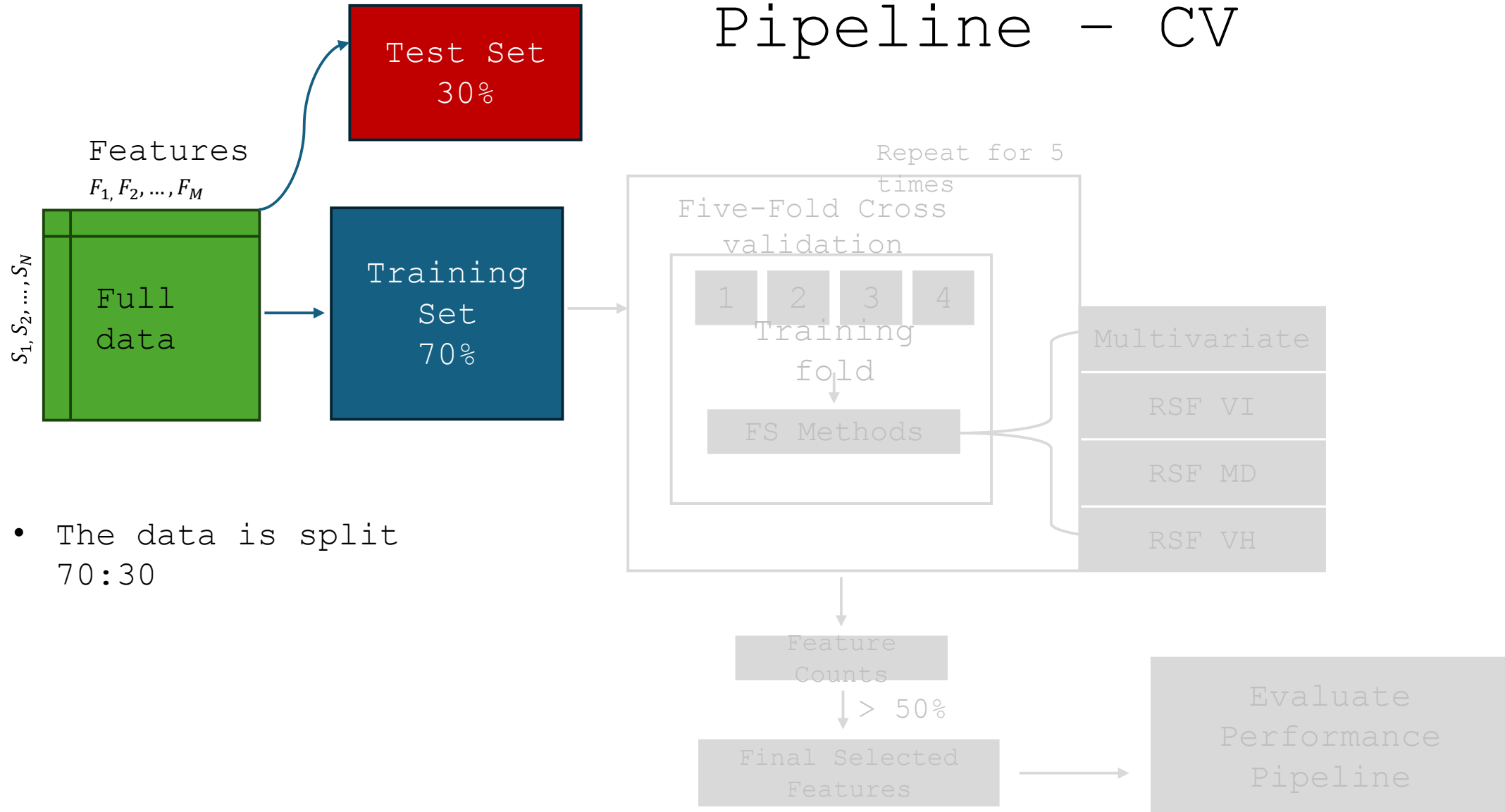
Mean Performance of Models + Feature Selection



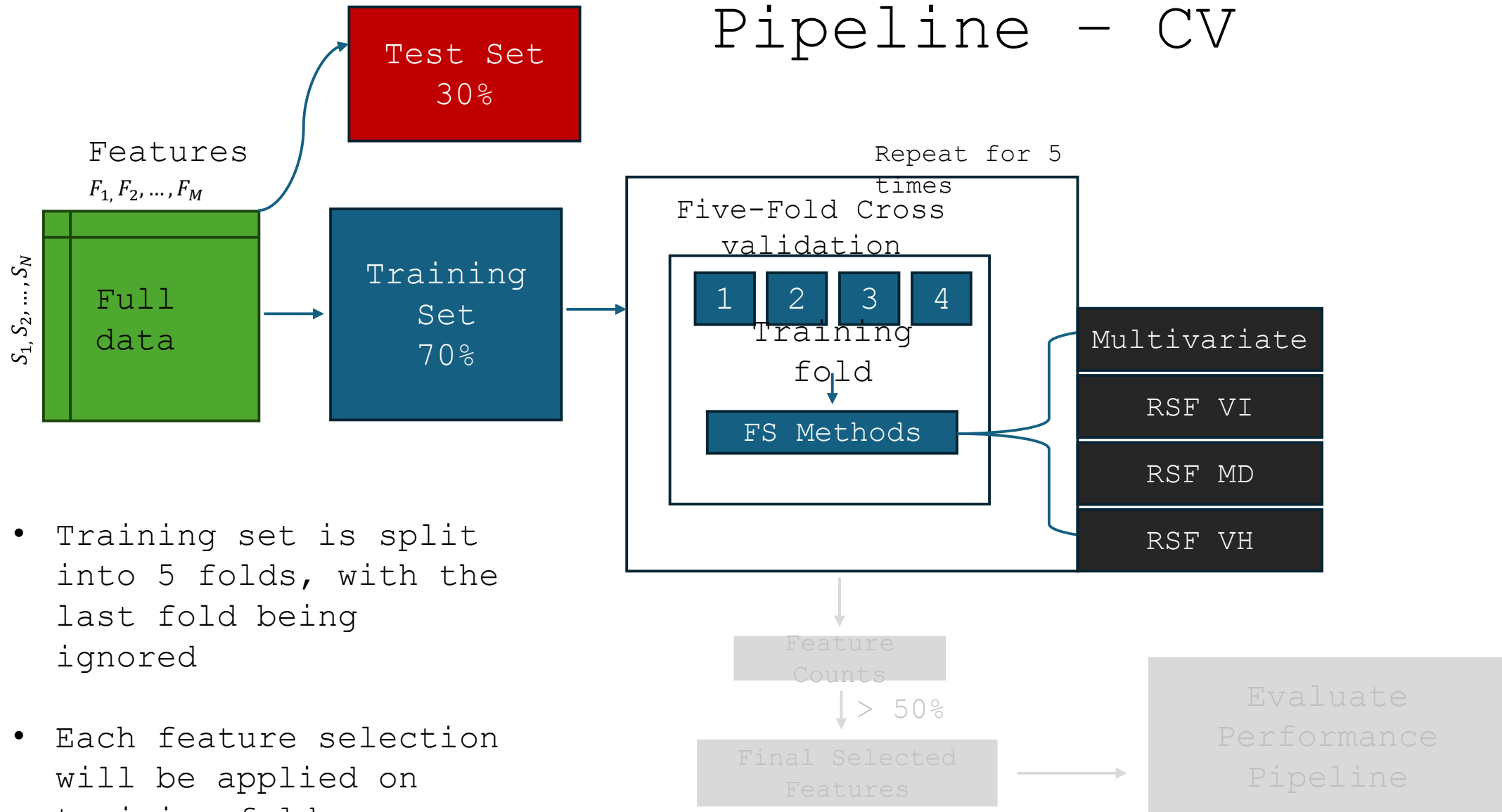
Example of BRCA DM performance and features selected

- Expect performance of each feature selection to be similar
- Great reduction of redundant features
- Transition from univariate to multivariate

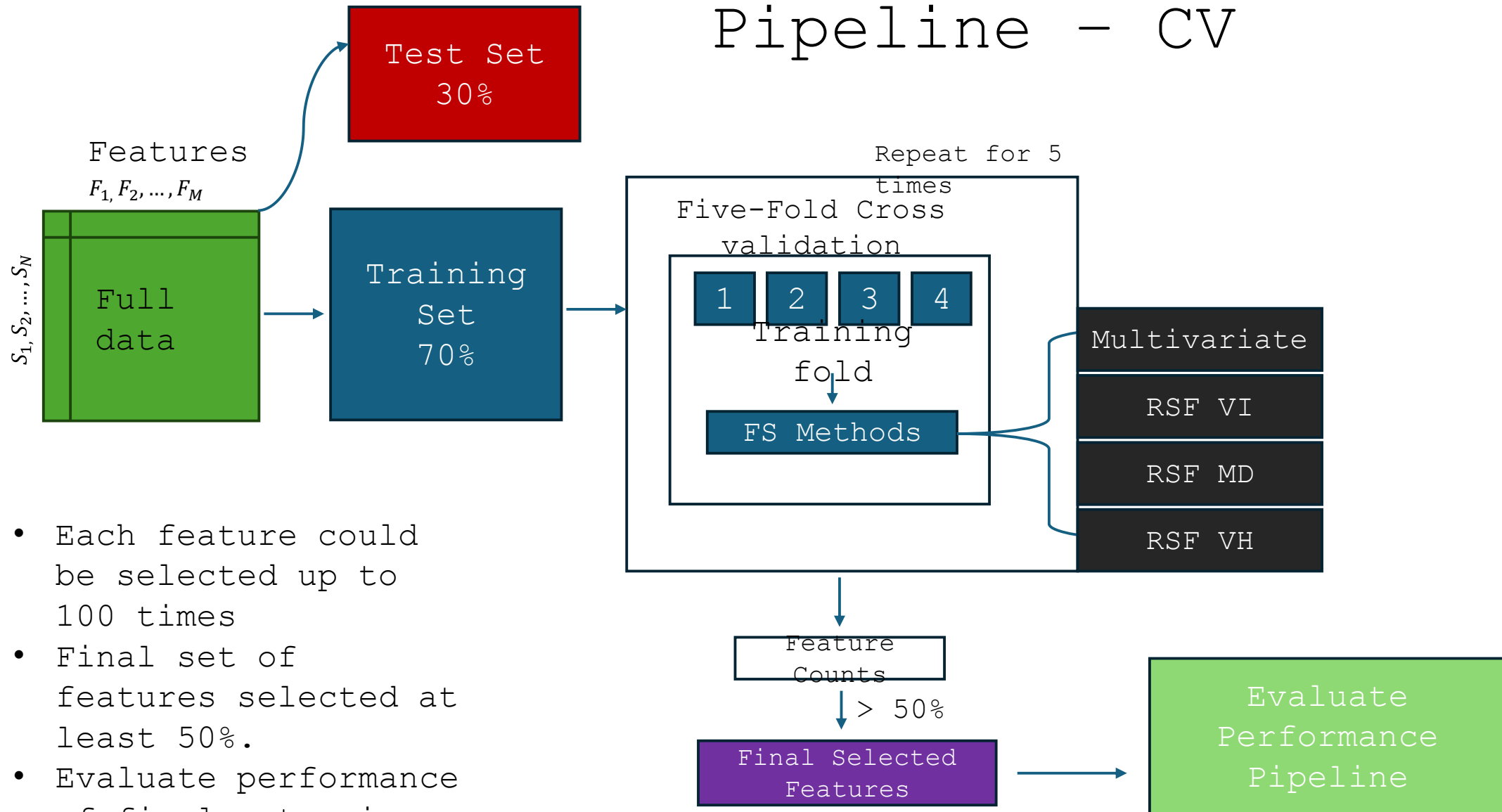
Feature Selection Pipeline - CV



Feature Selection Pipeline - CV

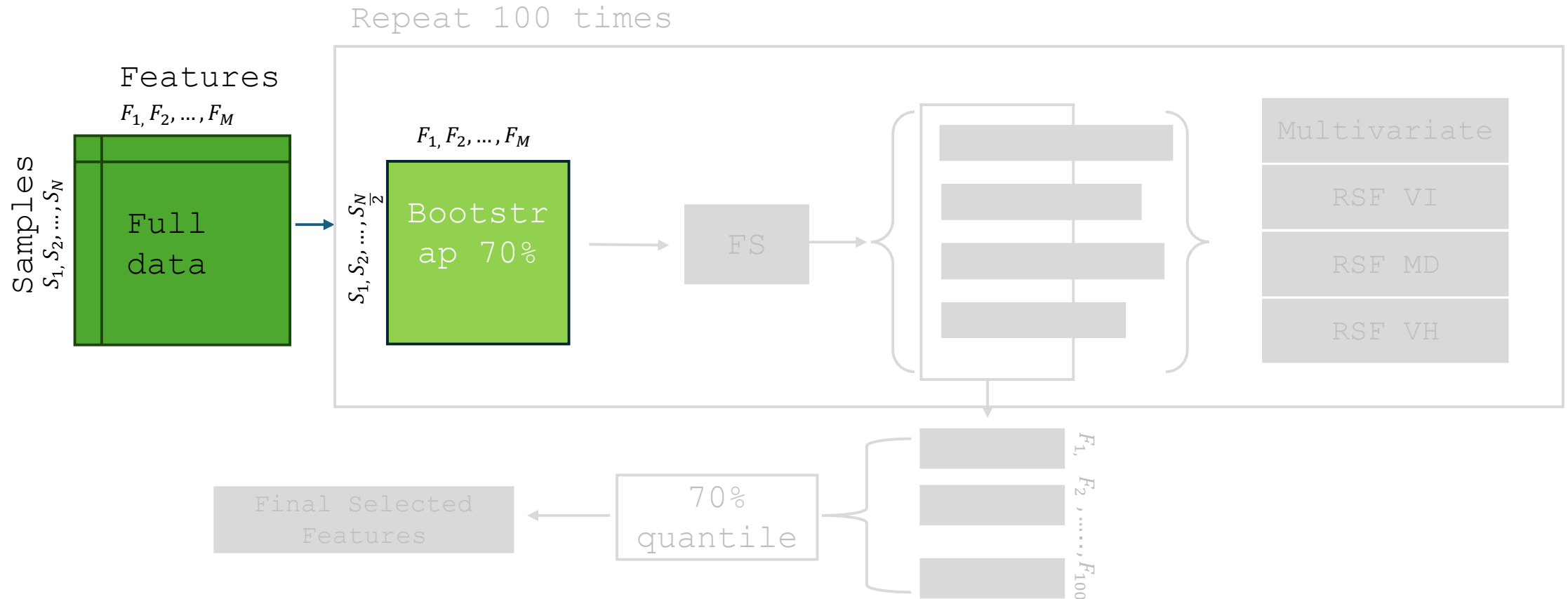


Feature Selection Pipeline - CV



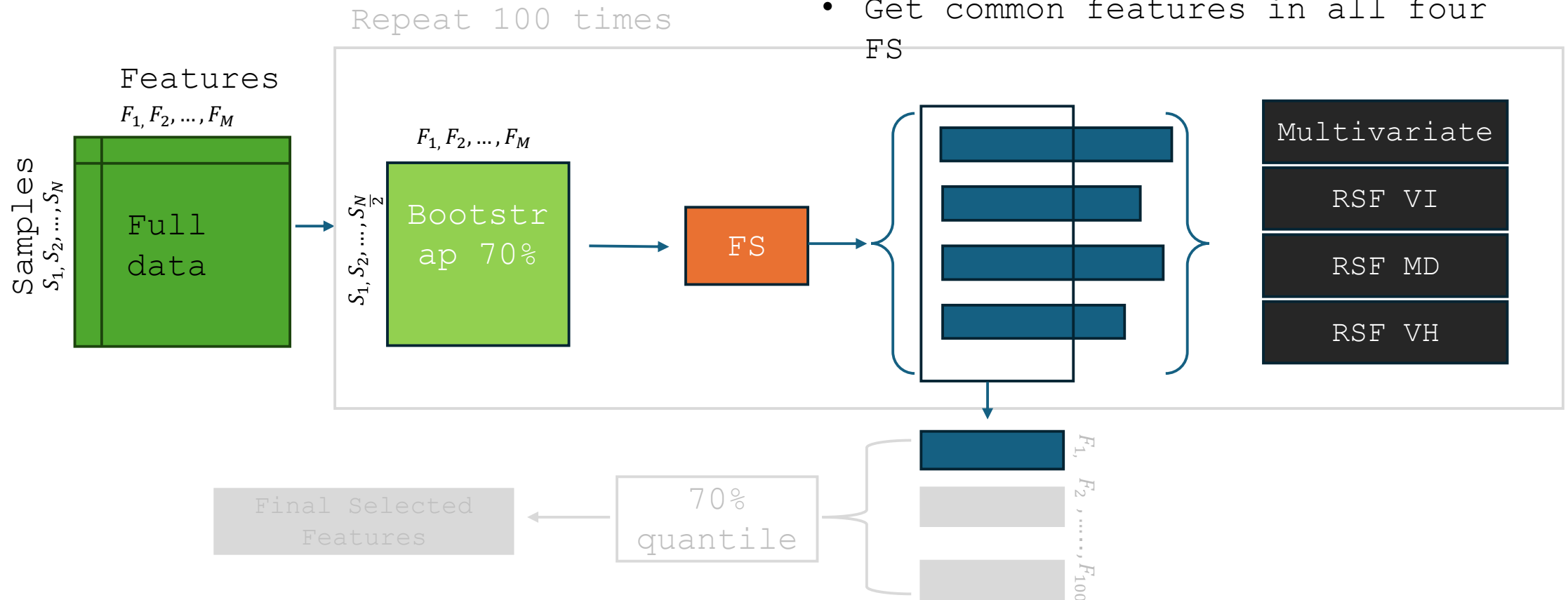
Feature Selection Pipeline - Bootstrapping

- Bootstrap 70% of the data



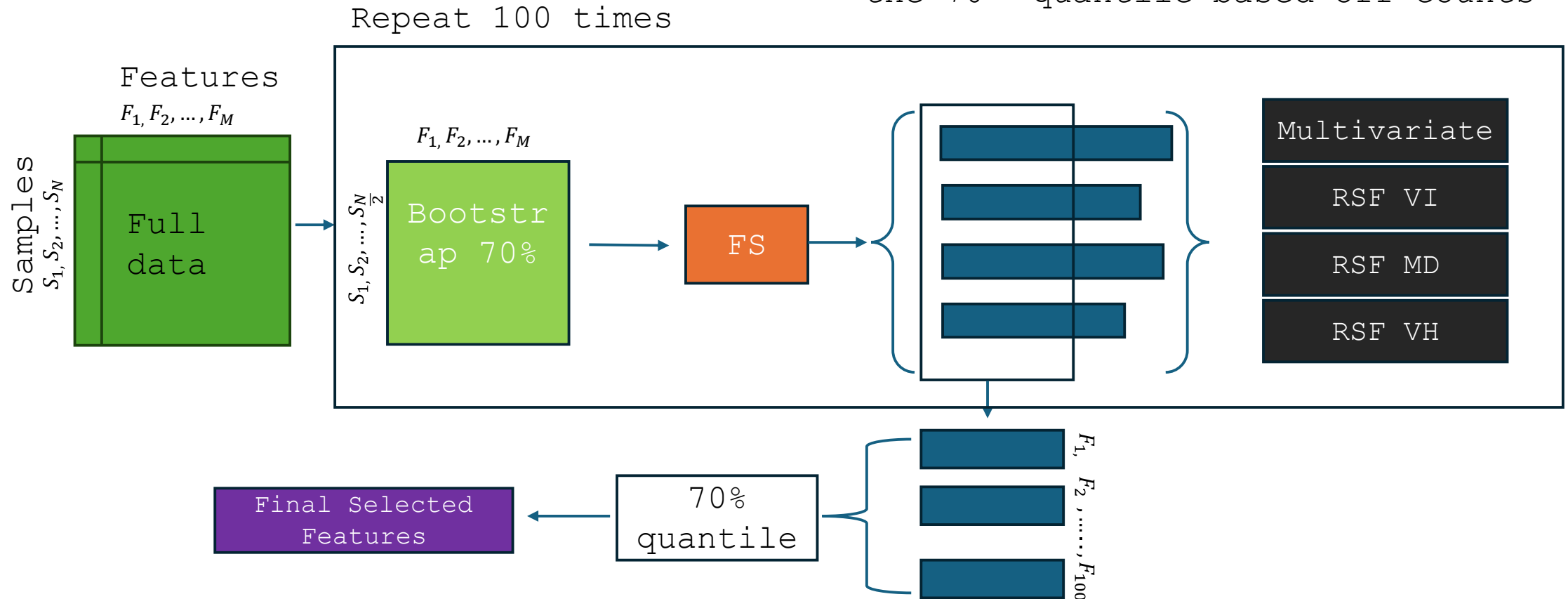
Feature Selection Pipeline - Bootstrapping

- Run each of the feature selection methods on the bootstrapped data
- Get common features in all four FS

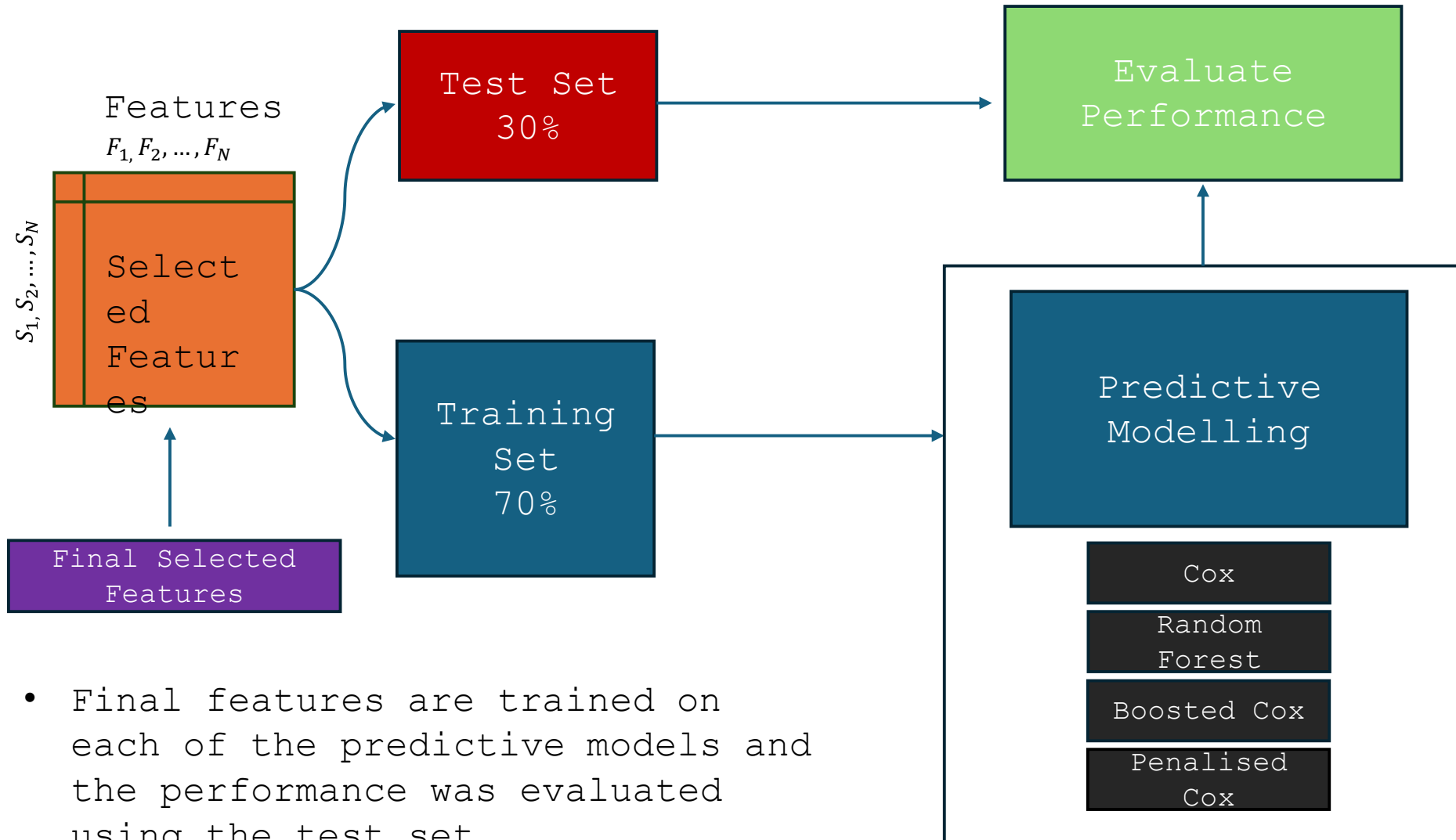


Feature Selection Pipeline - Bootstrapping

- Bootstrap 100 times
- Get the features that appear in the 70th quantile based off counts



Performance Pipeline



Current Results - Feature Selection (BRCA)

Table 3 Overview of BRCA performance for each omics using FS pipeline

	Omics	None	CV	BS
CoxPH (1)	GE	0.463	0.394	0.363
	ME	0.509	0.581	0.392
	DM	0.496	0.461	0.490
	CNV	0.473	0.497	0.421
Ranger (2)	GE	0.725	0.655	0.635
	ME	0.619	0.654	0.587
	DM	0.658	0.640	0.634
	CNV	0.611	0.629	0.579
GLMBoost (3)	GE	0.323	0.406	0.423
	ME	0.405	0.406	0.523
	DM	0.376	0.507	0.509
	CNV	0.422	0.420	0.431
ElasticNet (4)	GE	0.325	0.407	0.340
	ME	0.422	0.4	0.493
	DM	0.430	0.5	0.517
	CNV	0.425	0.426	0.435

Feature Selection

- Overall, decent results for CV as feature selection
- Ranger has highest overall performance
- Bootstrapping method not as good as CV.

Current Results – Feature Selection (OV)

Table 3 Overview of OV performance for each omics using FS pipeline

	Omics	None	CV
CoxPH (1)	GE	0.515	0.496
	ME	0.495	0.434
	DM	0.483	0.449
	CNV	0.471	0.524
Ranger (2)	GE	0.530	0.530
	ME	0.557	0.603
	DM	0.534	0.543
	CNV	0.536	0.532
ElasticNet (3)	GE	0.480	0.491
	ME	0.448	0.384
	DM	0.422	0.467
	CNV	0.482	0.479
GLMBoost (4)	GE	0.500	0.485
	ME	0.445	0.395
	DM	0.5	0.460
	CNV	0.484	0.489

Feature Selection

- Similarly, decent results for CV as feature selection
- However, for OV overall lower C-index in general < 0.60

Current Results - Feature Selection (CESC)

Table 3 Overview of CESC performance for each omics using FS pipeline

	Omics	None	CV
CoxPH (1)	GE	0.550	0.400
	ME	0.677	0.428
	DM	0.540	0.520
	CNV	0.441	0.560
Ranger (2)	GE	0.626	0.525
	ME	0.651	0.688
	DM	0.553	0.496
	CNV	0.471	0.471
ElasticNet (3)	GE	0.386	0.355
	ME	0.441	0.412
	DM	0.576	0.595
	CNV	0.520	0.573
GLMBoost (4)	GE	0.360	0.372
	ME	0.473	0.427
	DM	0.5	0.583
	CNV	0.497	0.570

Feature Selection

- Same trend with CESC

Current Results – Feature Selection (UCEC)

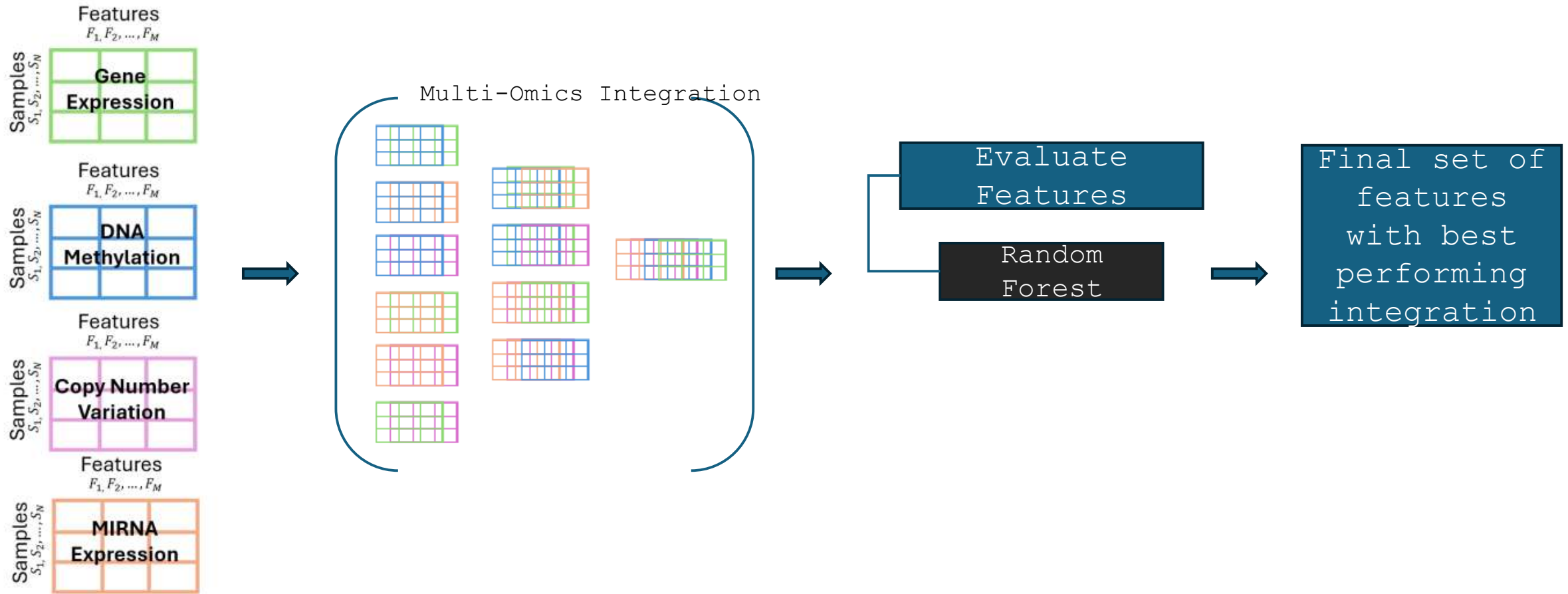
Table 3 Overview of UCEC performance for each omics using FS pipeline

	Omics	None	CV
CoxPH (1)	GE	0.447	0.465
	ME	0.411	0.371
	DM	0.470	0.491
	CNV	0.387	0.382
Ranger (2)	GE	0.666	0.598
	ME	0.686	0.676
	DM	0.636	0.632
	CNV	0.696	0.685
ElasticNet (3)	GE	0.369	0.373
	ME	0.315	0.341
	DM	0.360	0.361
	CNV	0.383	0.357
GLMBoost (4)	GE	0.359	0.370
	ME	0.317	0.354
	DM	0.359	0.342
	CNV	0.392	0.5

Feature Selection

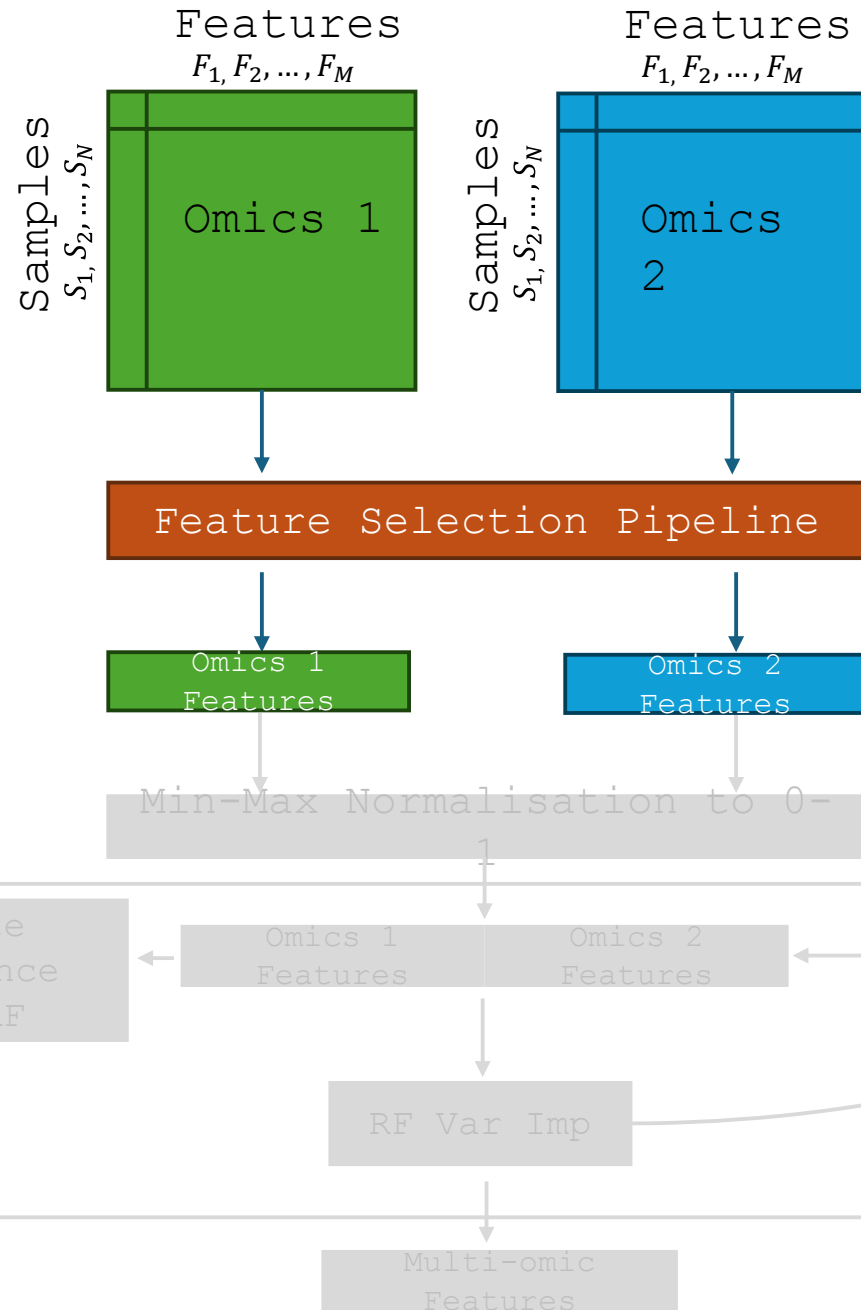
- Overall, Ranger appears to be the highest performing of our predictive models for all four cancers.

Overview of Multi-Omics Pipeline



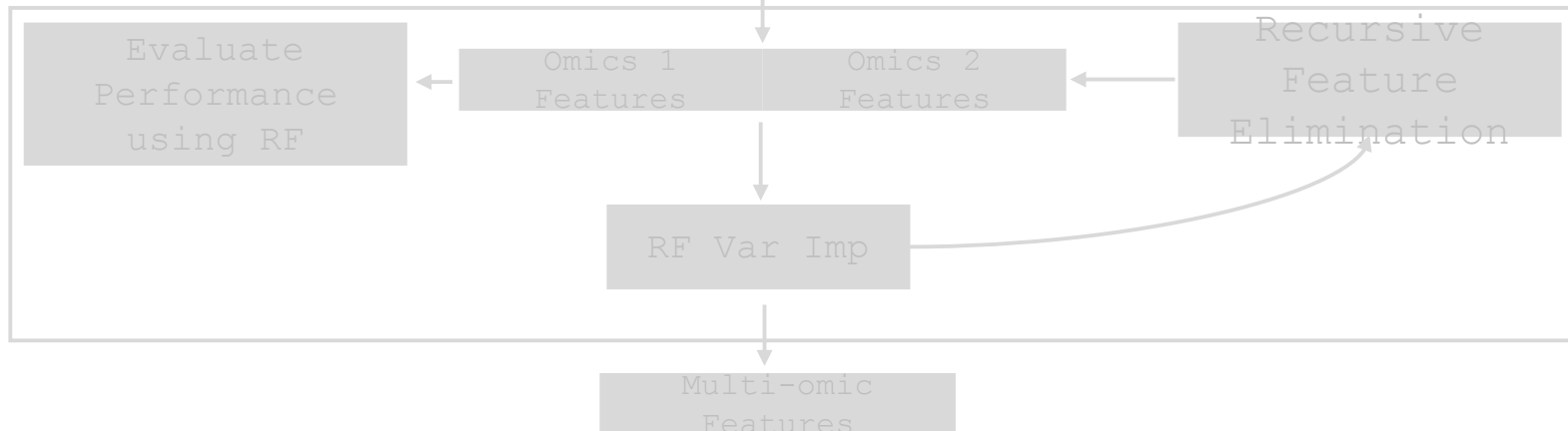
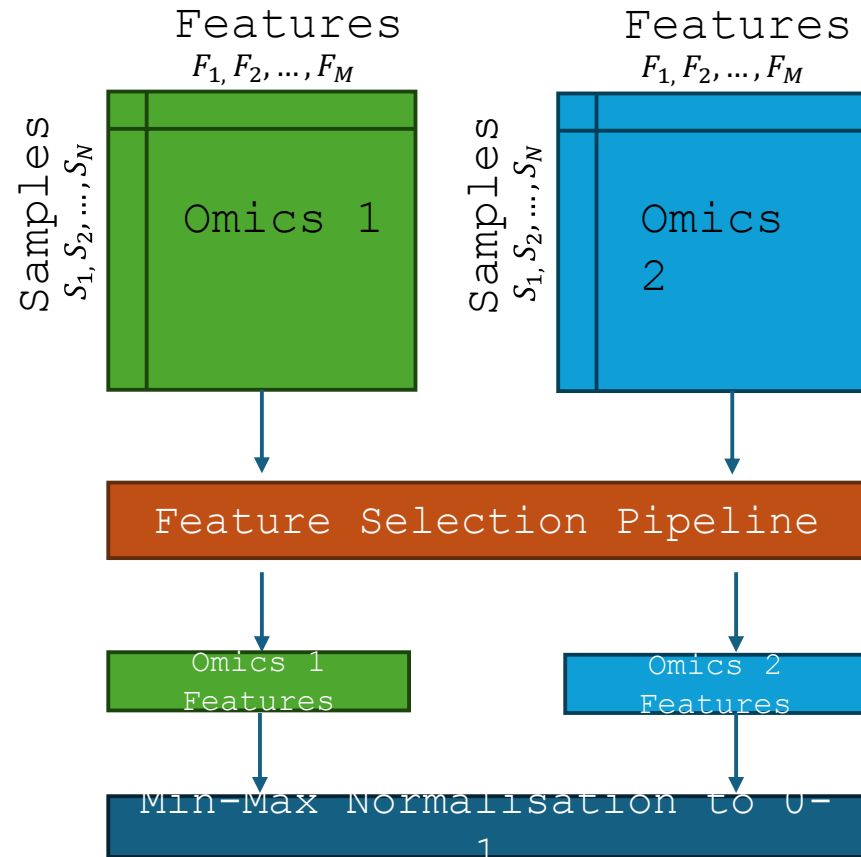
Multi-Omics Integration Pipeline - Early fusion integration

- Run each omics data through feature selection
- Obtain list of features for each omics



Multi-Omics Integration Pipeline - Early fusion integration

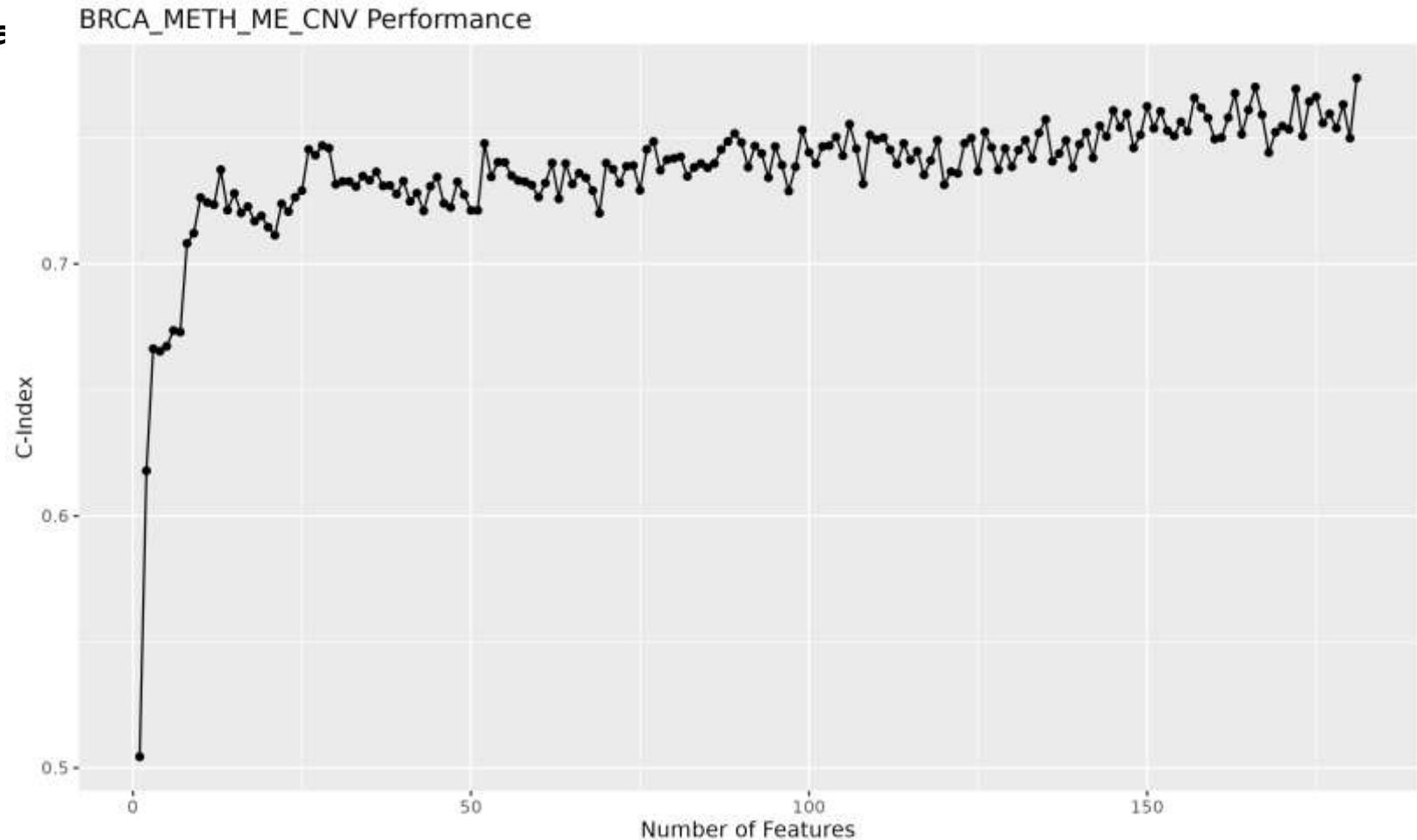
- Each omics data will have a different scale
- Normalise each of the data using min-max

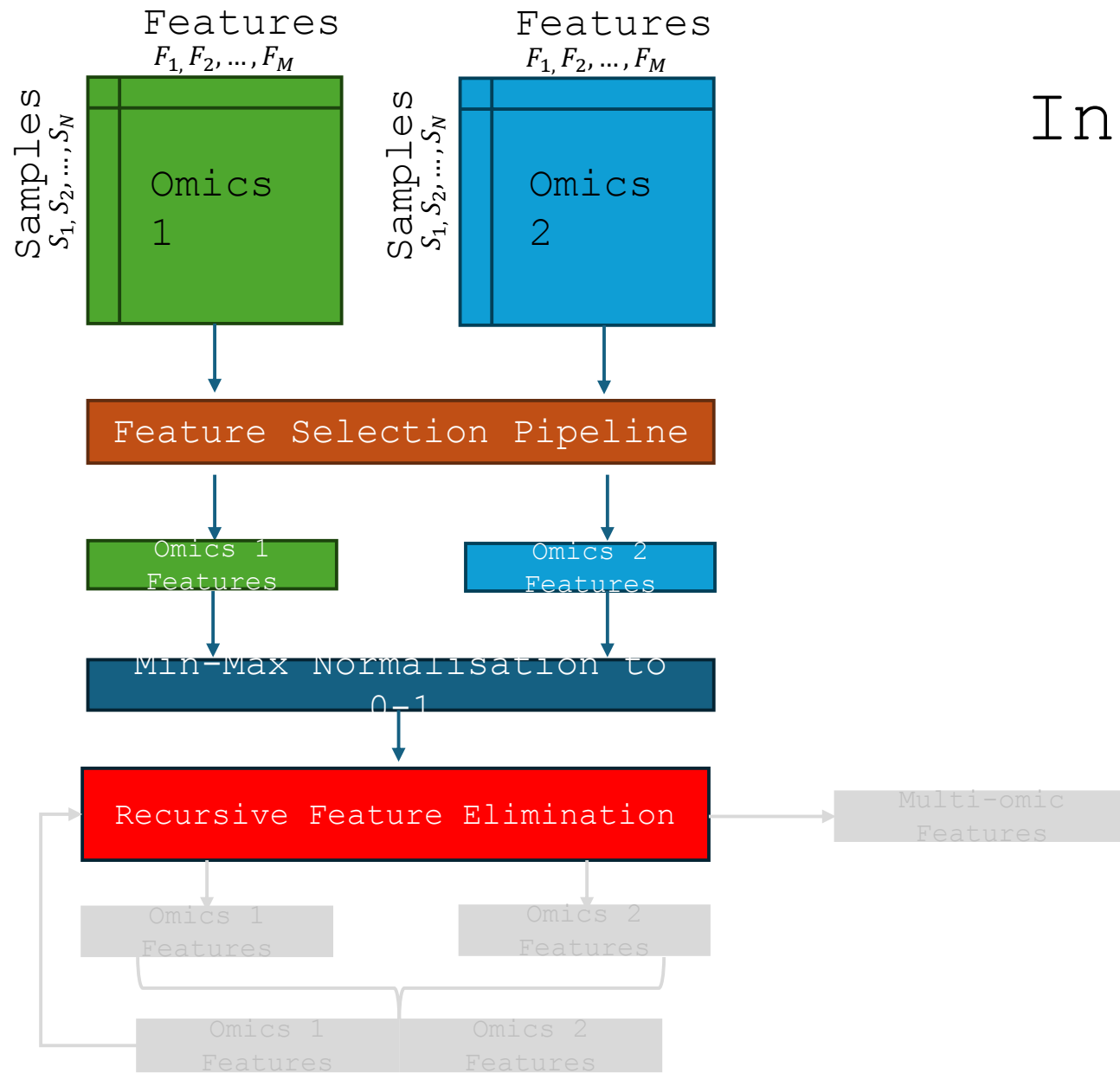


Current Results – Multi Omics (BRCA)

Recursive Feature Elimination

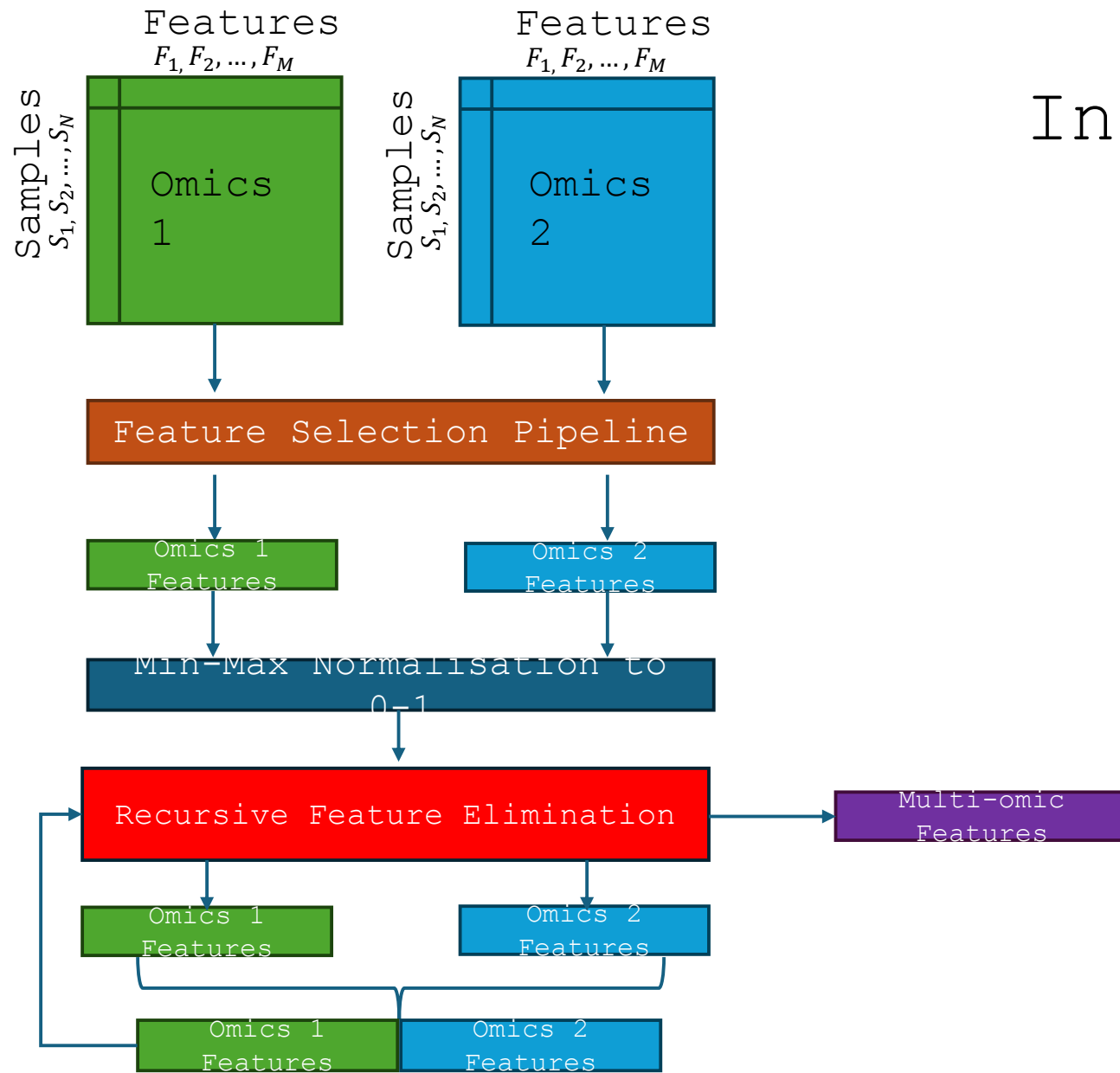
- Graphs showing performance of each iteration of RFE
- Example for DnaMeth + miRNA + CNV





Multi-Omics Integration Pipeline - Late fusion integration

- Initial process is the same, then each of the different omic features are run through RFE
- This will provide the highest performing set of features from each omics source



Multi-Omics Integration Pipeline - Late fusion integration

- The new set of features will then be concatenated, and the combined set will undergo RFE again

Multi-Omics Integration- Late vs Early (CESC)

Using Early Fusion:

Performance	ME + GE	ME + DM	GE + DM	GE + CNV	ME + CNV	DM + CNV	DM + GE + ME	DM + GE + CNV	DM + ME + CNV	ME + GE + CNV	ME + GE + CNV + DM
C-index	0.728	0.740	0.653	0.608	0.672	0.523	0.730	0.713	0.718	0.673	0.653

Common trend where Late Fusion is better across all cancers

Using Late Fusion:

Performance	ME + GE	ME + DM	GE + DM	GE + CNV	ME + CNV	DM + CNV	DM + GE + ME	DM + GE + CNV	DM + ME + CNV	ME + GE + CNV	ME + GE + CNV + DM
C-index	0.654	0.640	0.699	0.738	0.632	0.718	0.822	0.737	0.835	0.693	0.720

Current Results – Multi Omics

Performance	(BRCA)	ME + GE	ME + DM	GE + DM	GE + CNV	ME + CNV	DM + CNV	DM + GE + ME	DM + GE + CNV	DM + ME + CNV	ME + GE + CNV	ME + GE + CNV + DM
C-index		0.768	0.702	0.694	0.714	0.718	0.654	0.692	0.707	0.774	0.752	0.655

Omics	Features
DM	cg15520279, cg18087514, cg02776251, cg24262376, cg01671575, cg14717170, cg20786074, cg09656934, cg20308679, cg10634551, cg12626411, cg01664666, cg07837085, cg20261167, cg15147516, cg08205865, cg25583174, cg12958813, cg15761405, cg25167447, cg04574507, cg20676475, cg11902458, cg00807586, cg01637734, cg02836529, cg26583078, cg16232126, cg17214107, cg08422599, cg09952204, cg10978355, cg05619712, cg21137417, cg09563216, cg02983451, cg15077070, cg11657808, cg20311501, cg05275752, cg02431964, cg17453778, cg18335243, cg18182399, cg19616230, cg10777851, cg03679305, cg23300372, cg01612158, cg12815916, cg03614513, cg21591742, cg18006568, cg27403635, cg23282674, cg23557926, cg22176895, cg13320626, cg06933072, cg18110483, cg22359606, cg09220361, cg13847070, cg22855405, cg07038400, cg13986130, cg07354440, cg01114088, cg18794577, cg16557944, cg18482268, cg11428724, cg20613889, cg00995327, cg02774439, cg01216369, cg24512973, cg17701886, cg16516400, cg11819637, cg09478478, cg05200628, cg12105450, cg17525406, cg04098585, cg08474603, cg00884221, cg25462303, cg17105014, cg17542385, cg17020834, cg07548313, cg01868128, cg14925024, cg05768141, cg17108819, cg08826839, cg05909475, cg25057743, cg25631352, cg08047907, cg10861599, cg08658594, cg10691387, cg18277754, cg08578023, cg12880658, cg17281600, cg08690031, cg19228118, cg10031651, cg03702236, cg08057475, cg22605643, cg25437385, cg03852144, cg21948655, cg03329572, cg06168050, cg06274159, cg23363832, cg10370591, cg11724759, cg05248781, cg10774440, cg07664856, cg21972382, cg20535085, cg07694025, cg24928687, cg12864235, cg20579480, cg14419187, cg06043114, cg19884658, cg17430393, cg19526600, cg00498604, cg03199651, cg11976166, cg16139316, cg06940792, cg00567749, cg10946435, cg21991396, cg18908499, cg01367992, cg26117023
ME	hsa.mir.31, hsa.mir.30a, hsa.mir.22, hsa.mir.150, hsa.mir.4742, hsa.mir.221
CNV	ADAM2, AC139365.1, RAB11FIP1, CYP4F44P, AC239800.2, U3, RNU1.124P, HSPA8P13, POMK, AC110275.1, AC139365.2, AC123767.1, RNU6.356P, AC103726.1, AC067817.2, ZNF703, SLC20A2, RPS20P22, AC048387.1, MIR1204, ADAM3A, RN7SL709P, AC118650.1, SNORD65B, SFRP1, MIR548AO, AC091182.1

Overall Performance of each integrated omics

- Integration of MiRNA, DNA Methylation and Copy Number variation data achieves the best overall performance
- Outperforms baseline single-modality Ranger model using only GE (C-Index of 0.725)
- These are the features from each omics set that contributed to the highest performance after integration

Current Results – Multi Omics (OV)

Performance	ME + GE	ME + DM	GE + DM	GE + CNV	ME + CNV	DM + CNV	DM + GE + ME	DM + GE + CNV	DM + ME + CNV	ME + GE + CNV	ME + GE + CNV + DM
C-index	0.673	0.683	0.646	0.593	0.559	0.616	0.666	0.585	0.630	0.633	0.637

Omics	Features	Overall Performance of each integrated omics									
ME	hsa.mir.1301, hsa.mir.3200, hsa.let.7a.1, hsa.let.7a.3, hsa.mir.135b, hsa.mir.139	<ul style="list-style-type: none">Integration of MiRNA and DNA Methylation achieves the best overall performanceOutperforms baseline single-modality Ranger model									
DM	cg04533291, cg03874199, cg27342801, cg20676475, cg06101324										

Current Results – Multi Omics (CESC)

Performance	ME + GE	ME + DM	GE + DM	GE + CNV	ME + CNV	DM + CNV	DM + GE + ME	DM + GE + CNV	DM + ME + CNV	ME + GE + CNV	ME + GE + CNV + DM
C-index	0.654	0.640	0.699	0.738	0.632	0.718	0.822	0.737	0.835	0.693	0.720

Omics	Features	Overall Performance of each integrated omics
DM	cg25514503, cg16607065, cg01612158, cg12958813	<ul style="list-style-type: none">Similarly to BRCA, Integration of MiRNA, DNA Methylation and Copy Number variation data achieves the best overall performanceSignificantly outperforms baseline single-modality Ranger model
ME	hsa.mir.502, hsa.mir.101.2, hsa.mir.140, hsa.mir.150, hsa.mir.500b, hsa.mir.1306, hsa.mir.142, hsa.mir.204, hsa.mir.144, hsa.mir.188, hsa.mir.155, hsa.mir.205, hsa.mir.335, hsa.mir.148a, hsa.mir.196a.1, hsa.mir.151a, hsa.mir.193a, hsa.let.7e	
CNV	FAM91A2P, RNU6.488P	

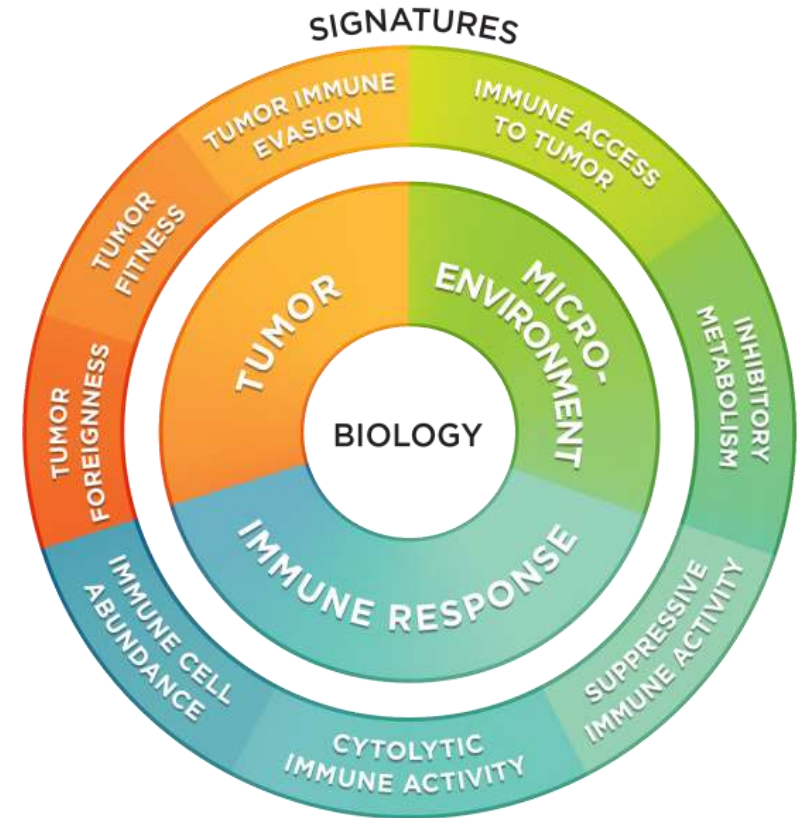
Current Results – Multi Omics (UCEC)

Performance	ME + GE	ME + DM	GE + DM	GE + CNV	ME + CNV	DM + CNV	DM + GE + ME	DM + GE + CNV	DM + ME + CNV	ME + GE + CNV	ME + GE + CNV + DM
C-index	0.654	0.723	0.684	0.650	0.650	0.609	0.678	0.606	0.766	0.667	0.687

Omics	Features	Overall Performance of each integrated omics
DM	cg25514503, cg16607065, cg01612158, cg12958813	<ul style="list-style-type: none">Interestingly, 3 out of 4 cancer types have the combination of Methylation, miRNA and copy number as the highest performing.However, all four cancers did have Methylation and miRNA as a part of their final integration
ME	hsa.mir.502, hsa.mir.101.2, hsa.mir.140, hsa.mir.150, hsa.mir.500b, hsa.mir.1306, hsa.mir.142, hsa.mir.204, hsa.mir.144, hsa.mir.188, hsa.mir.155, hsa.mir.205, hsa.mir.335, hsa.mir.148a, hsa.mir.196a.1, hsa.mir.151a, hsa.mir.193a, hsa.let.7e	
CNV	FAM91A2P, RNU6.488P	

Overview of Pan-cancer Analysis

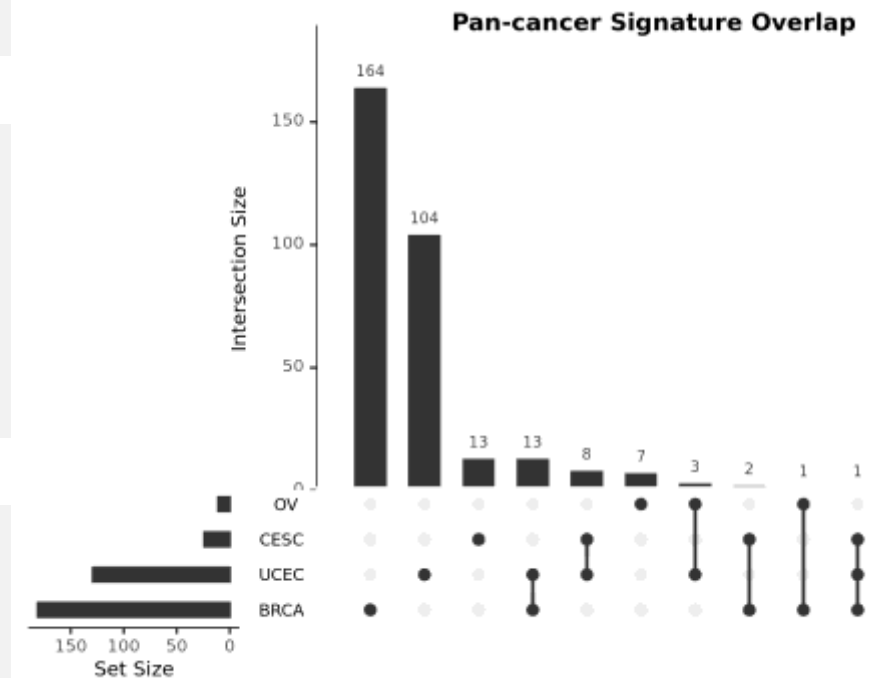
- Looked at any overlapping features across cancers
- Further investigated other potential overlaps by exploring miRNA and Methylation Targets
- Exploration included disease association networks for both miRNA gene Targets and the overlapping miRNA
- GSEA on the Targets from each cancers, to later identify any overlapping GO Terms or KEGG Pathways
- Deeper exploration of interesting features using KM-plots



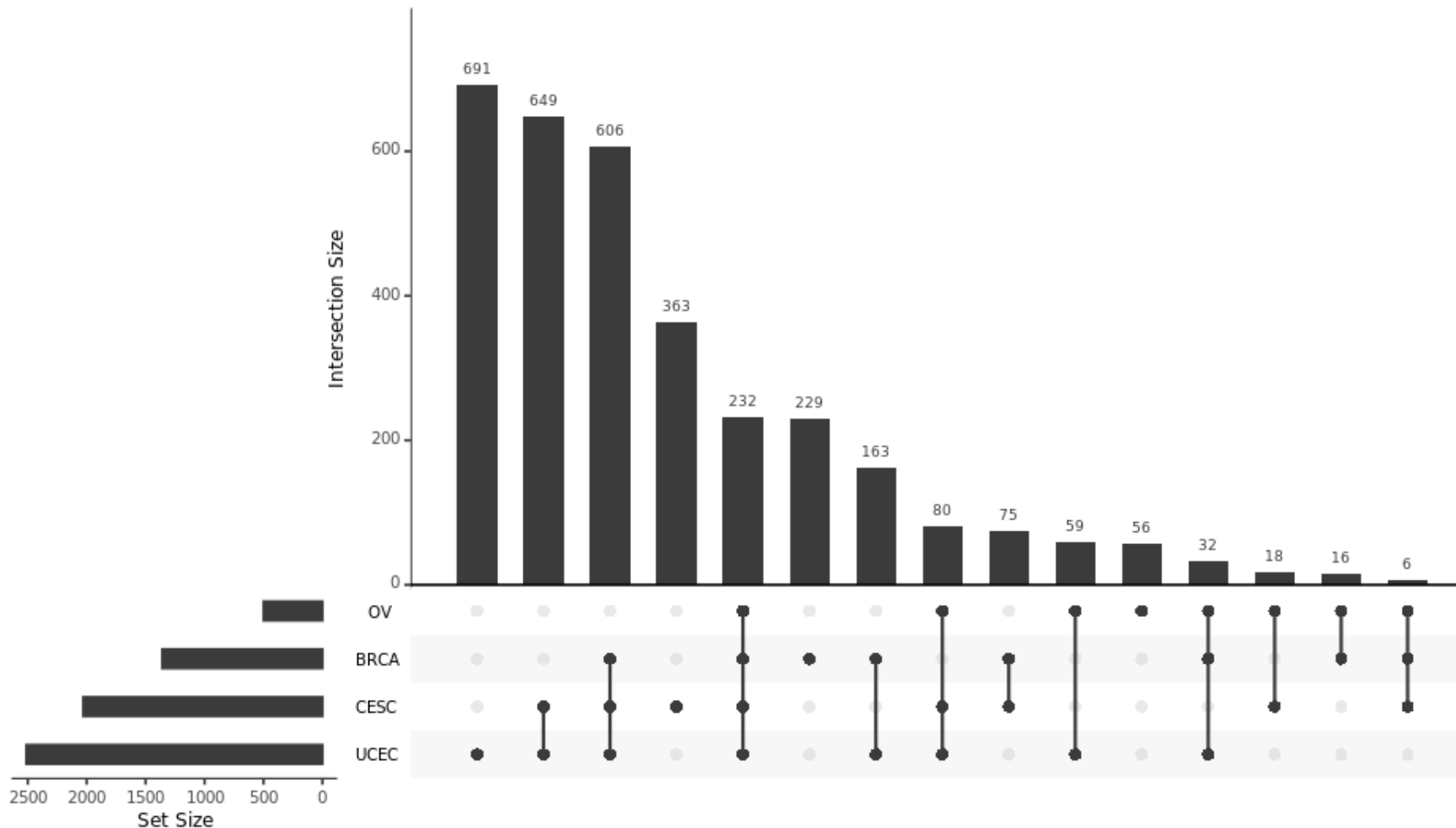
Pan-cancer Analysis - Overlapping Features

Cancer Types	Overlapping Features
BRCA, OV	METH_cg20676475
BRCA, CESC	METH_cg01612158, METH_cg12958813, ME_hsa.mir.150
BRCA, UCEC	CNV_MIR1204, CNV_U3, METH_cg07548313, METH_cg09478478, METH_cg11657808, METH_cg12105450, METH_cg14419187, METH_cg17525406, METH_cg18277754, METH_cg22605643, ME_hsa.mir.150, ME_hsa.mir.22, ME_hsa.mir.30a ME_hsa.mir.31
OV, UCEC	ME_hsa.let.7a.1, ME_hsa.let.7a.3, ME_hsa.mir.135b
CECSC, UCEC	METH_cg16607065, ME_hsa.mir.140, ME_hsa.mir.142, ME_hsa.mir.144, ME_hsa.mir.148a, ME_hsa.mir.150, ME_hsa.mir.155, ME_hsa.mir.196a.1, ME_hsa.mir.335
BRCA, CESC, UCEC	ME_hsa.mir.150

METH = Methylation, CNV = Copy Number Variation,
ME = miRNA



Pan-cancer Analysis – MiRNA Targets



MiRNA Gene Target

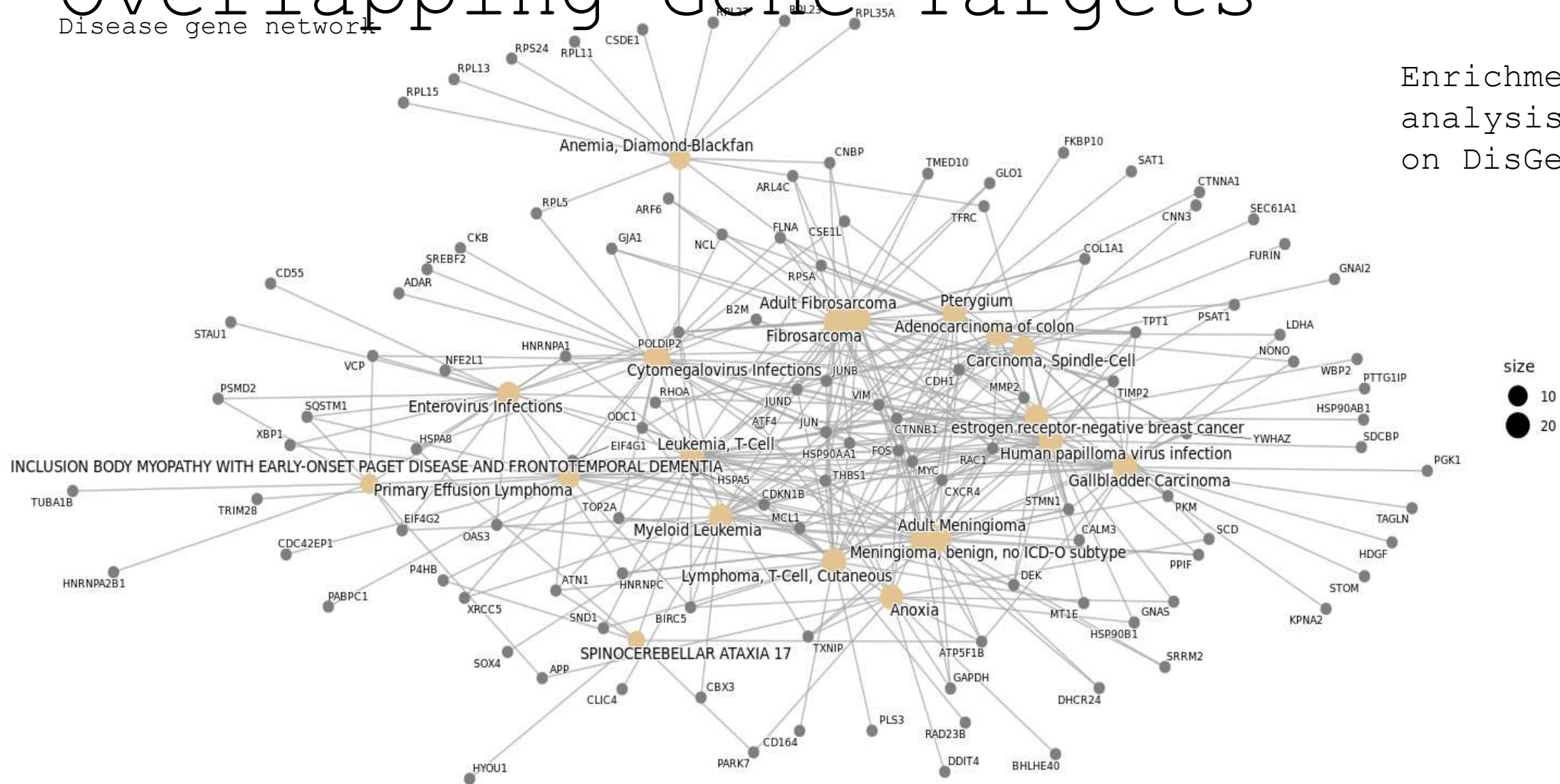
Overlap: Using the multiMiR Package to extract validated miRNA-target interactions

- Filtered the gene list to have only expressed genes within each cancer

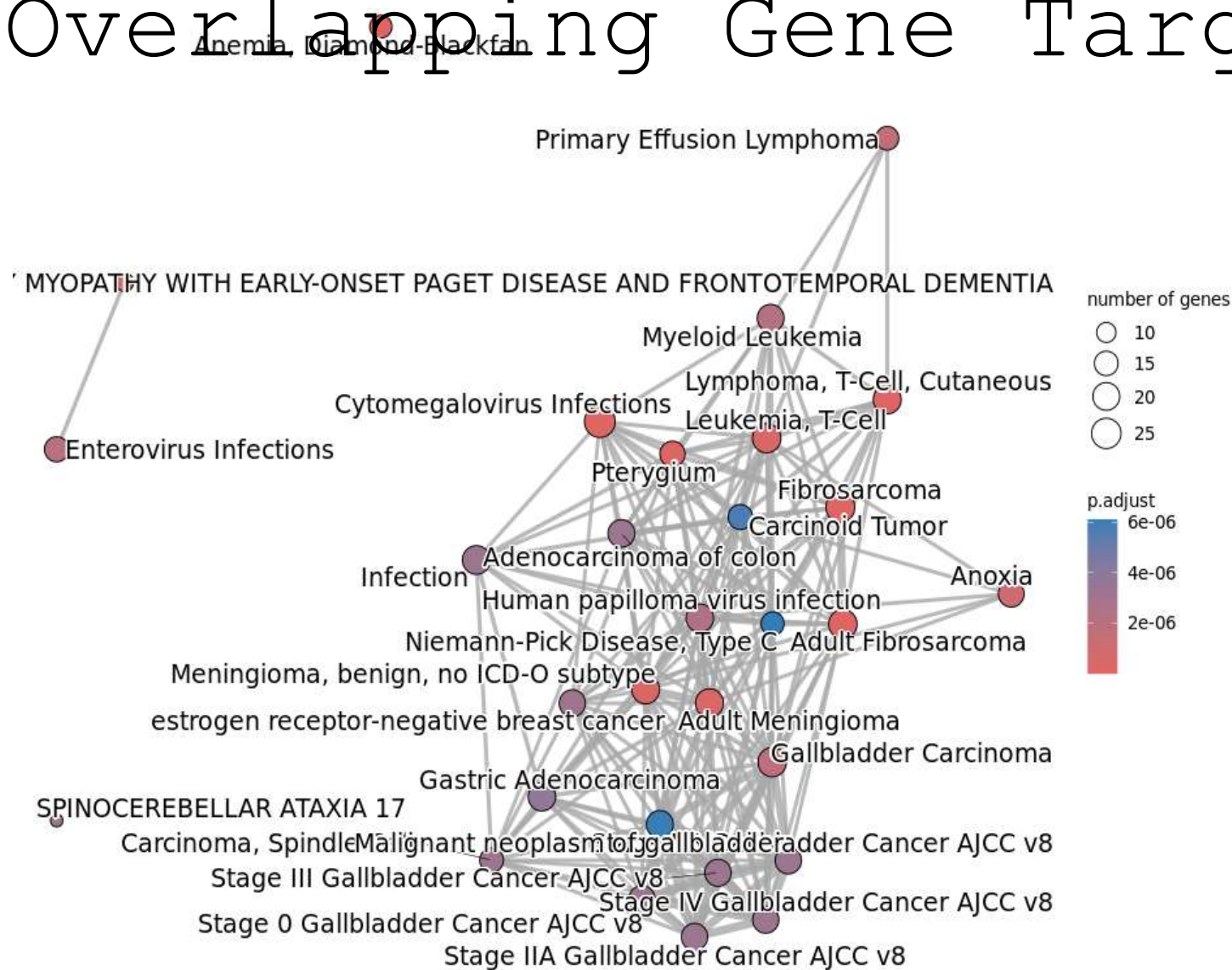
Pan-cancer Analysis – Overlapping Gene Targets

Disease gene network

Enrichment
analysis based
on DisGeNET



Pan-cancer Analysis – Overlapping Gene Targets

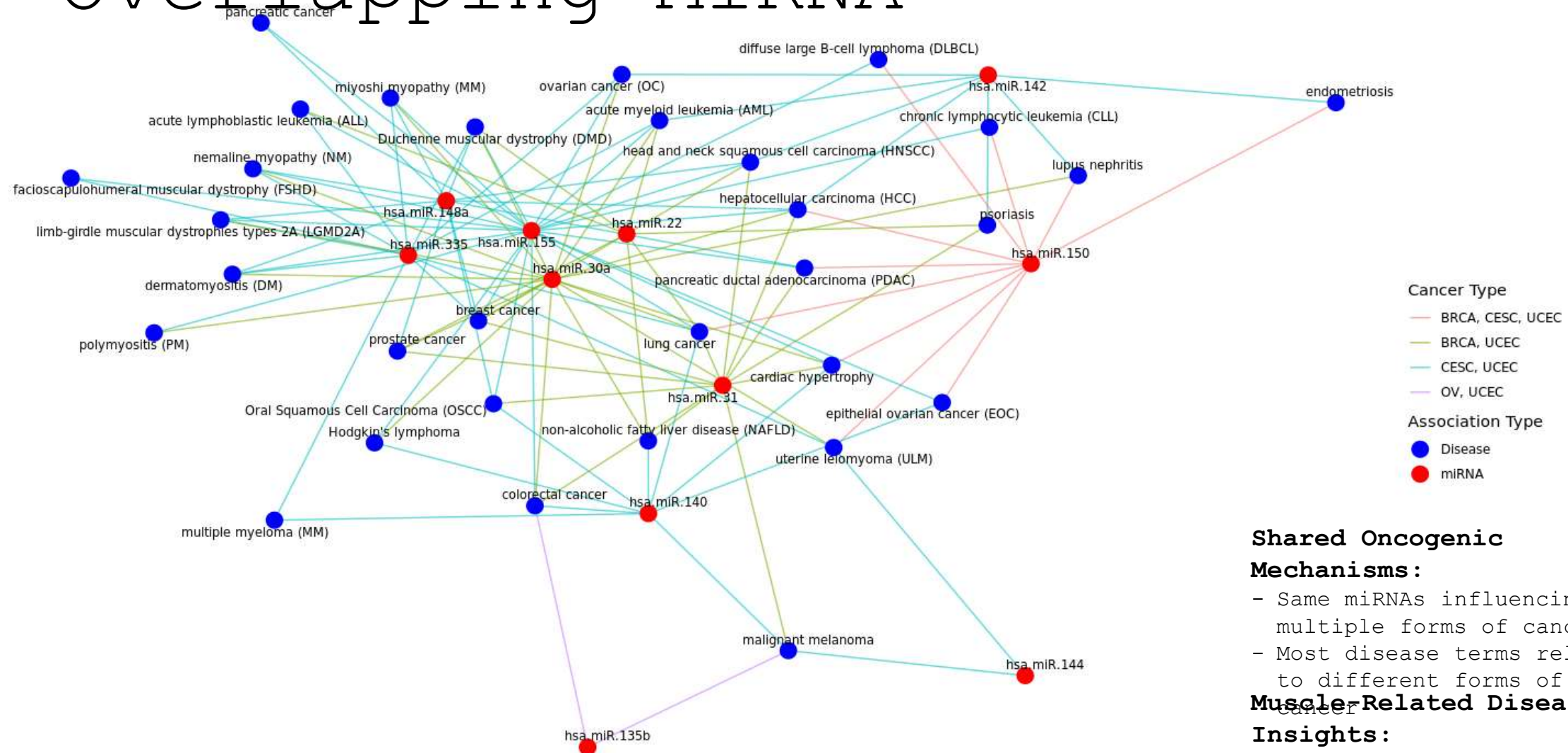


Enrichment Map:

- Mutually overlapping gene sets tend to cluster together, making it easy to identify functional module

Pan-cancer Analysis - Overlapping MiRNA

miRNA-Disease Associations Across Cancers



Gene Set Enrichment Analysis (GSEA)

Gene Ontology (GO)

Used to describe gene function, and relationships along these three aspects:

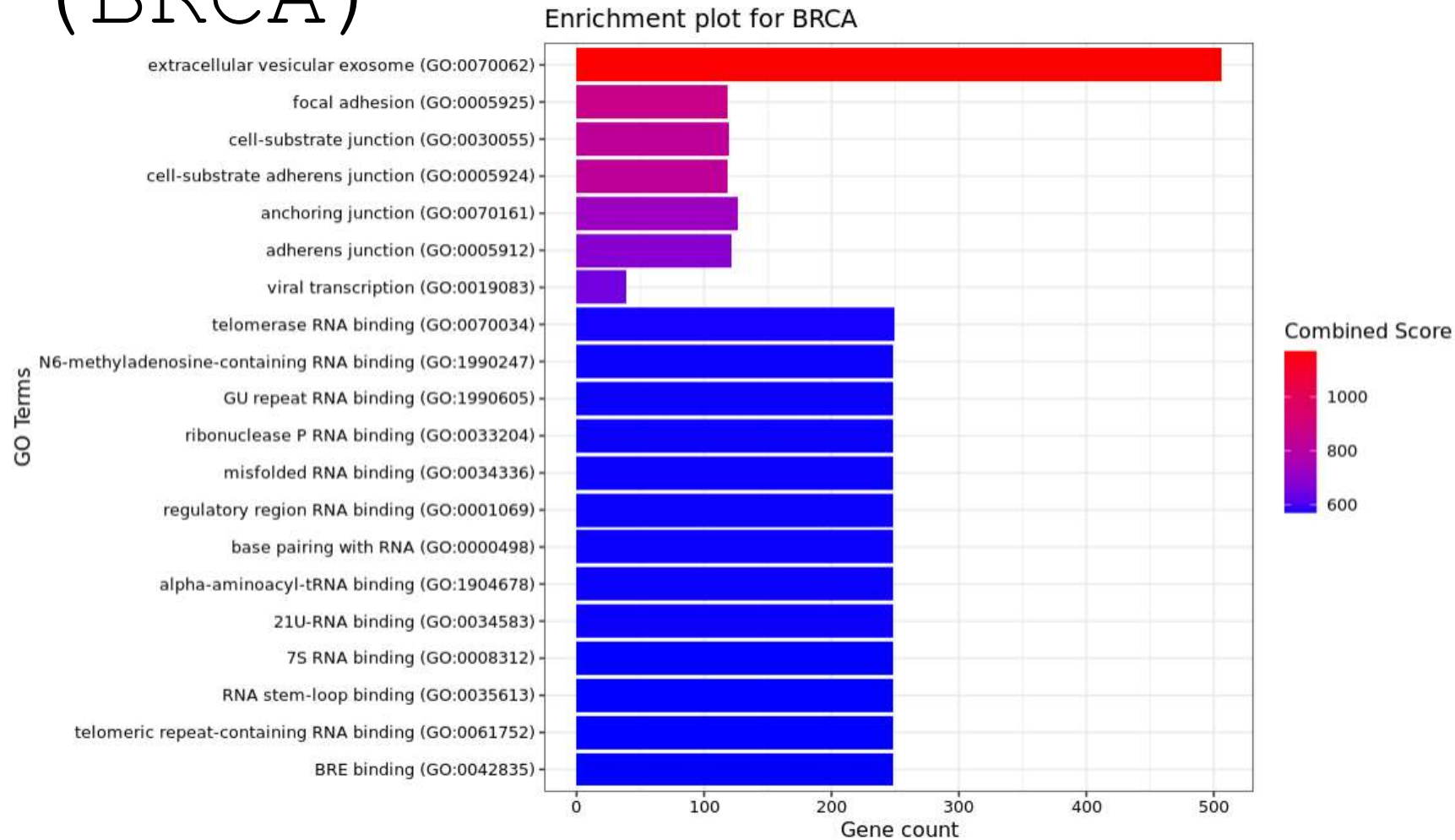
- Molecular Function:
molecular activities of gene products
- Cellular Component:
where gene products are active
- Biological Process:
pathways and larger processes made up of the activities of multiple gene products

Kyoto Encyclopedia of Genes and Genomes (KEGG)

Represents molecular interaction and reaction networks. These pathways cover a wide range of biochemical processes:

1. Metabolism
2. Genetic information processing
3. Environmental information processing
4. Cellular processes
5. Organismal systems
6. Human diseases
7. Drug development.

MiRNA Targets – GO Terms (BRCA)

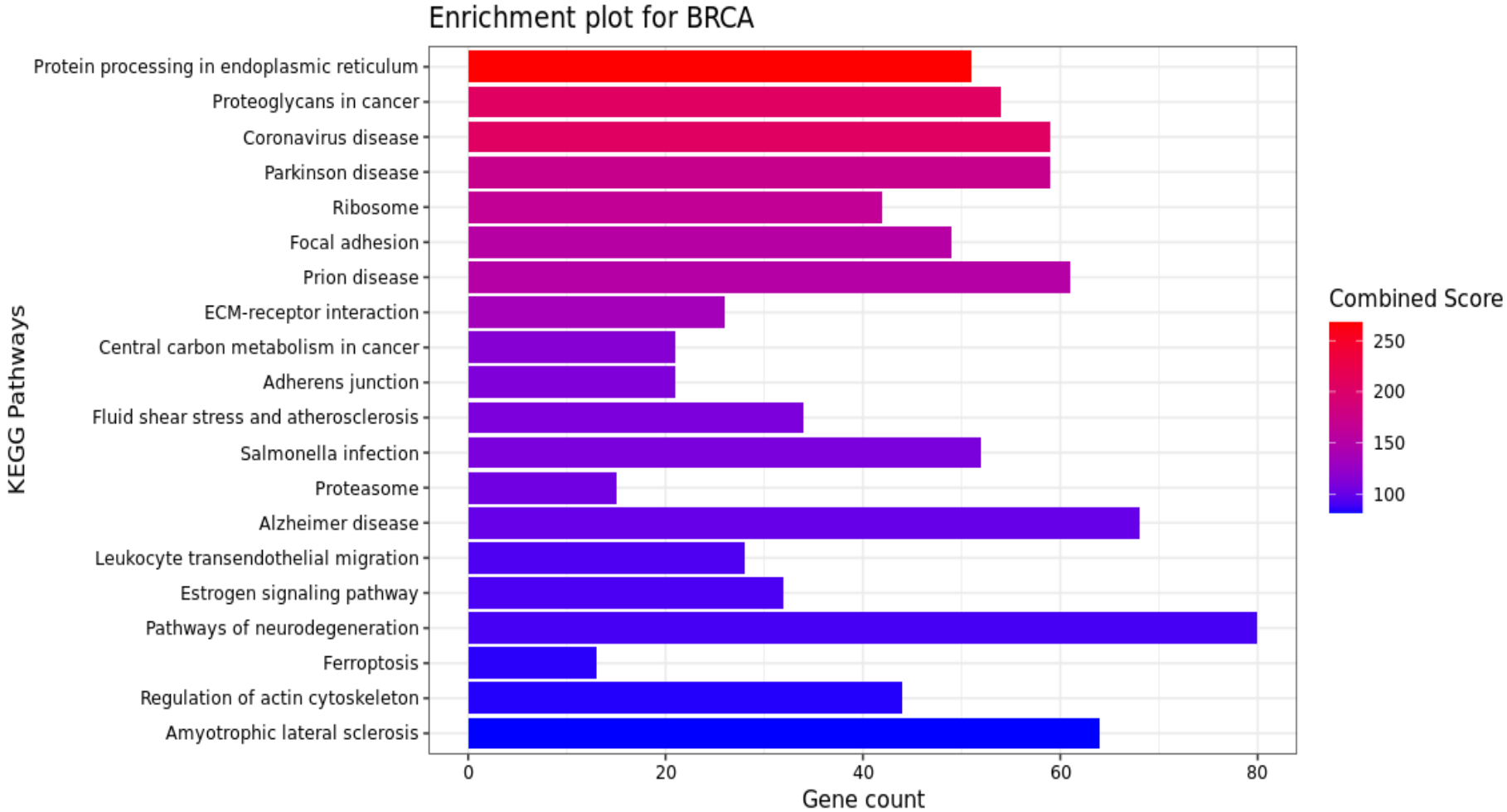


Significant GO Terms BRCA:

- Reveals significant involvement in extracellular vesicular transport, cell adhesion, and RNA binding/processing
- Suggest that miRNAs may regulate key processes related to tumour progression, metastasis, and RNA dynamics in breast cancer

MiRNA Targets – KEGG Pathway

(BRCA)

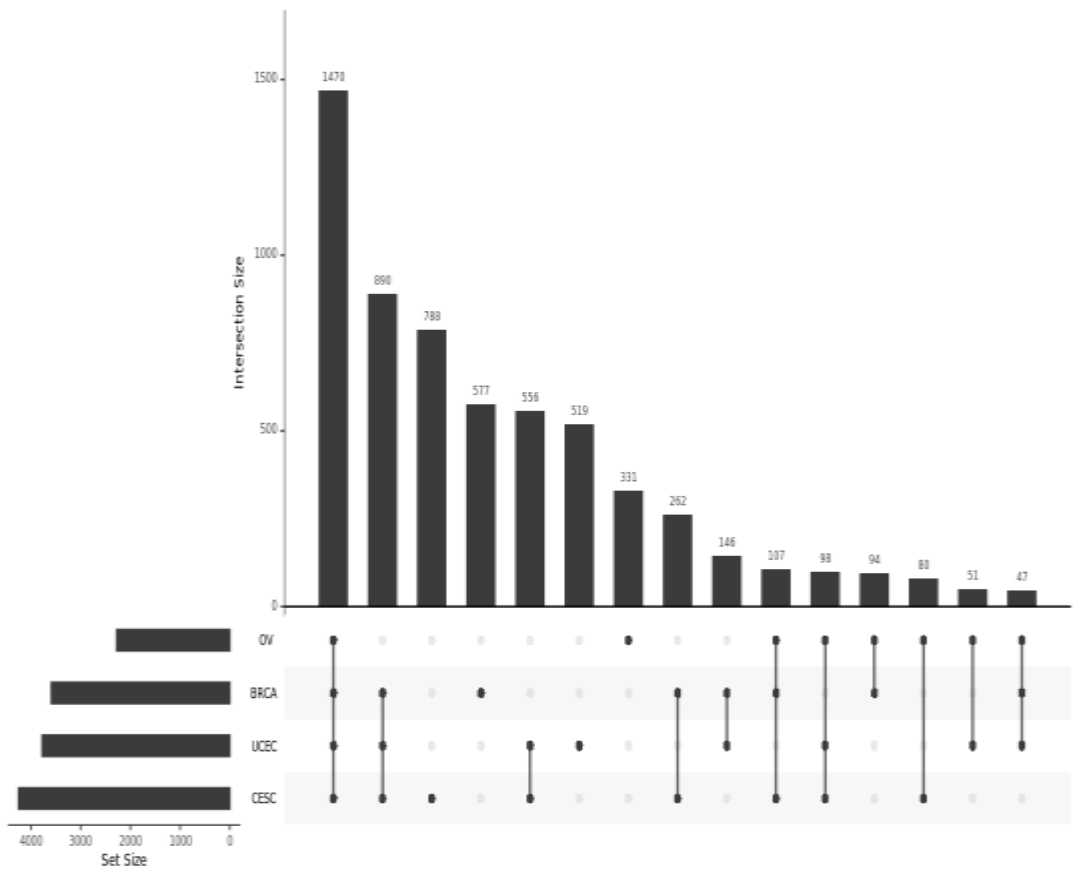


Significant KEGG Pathway

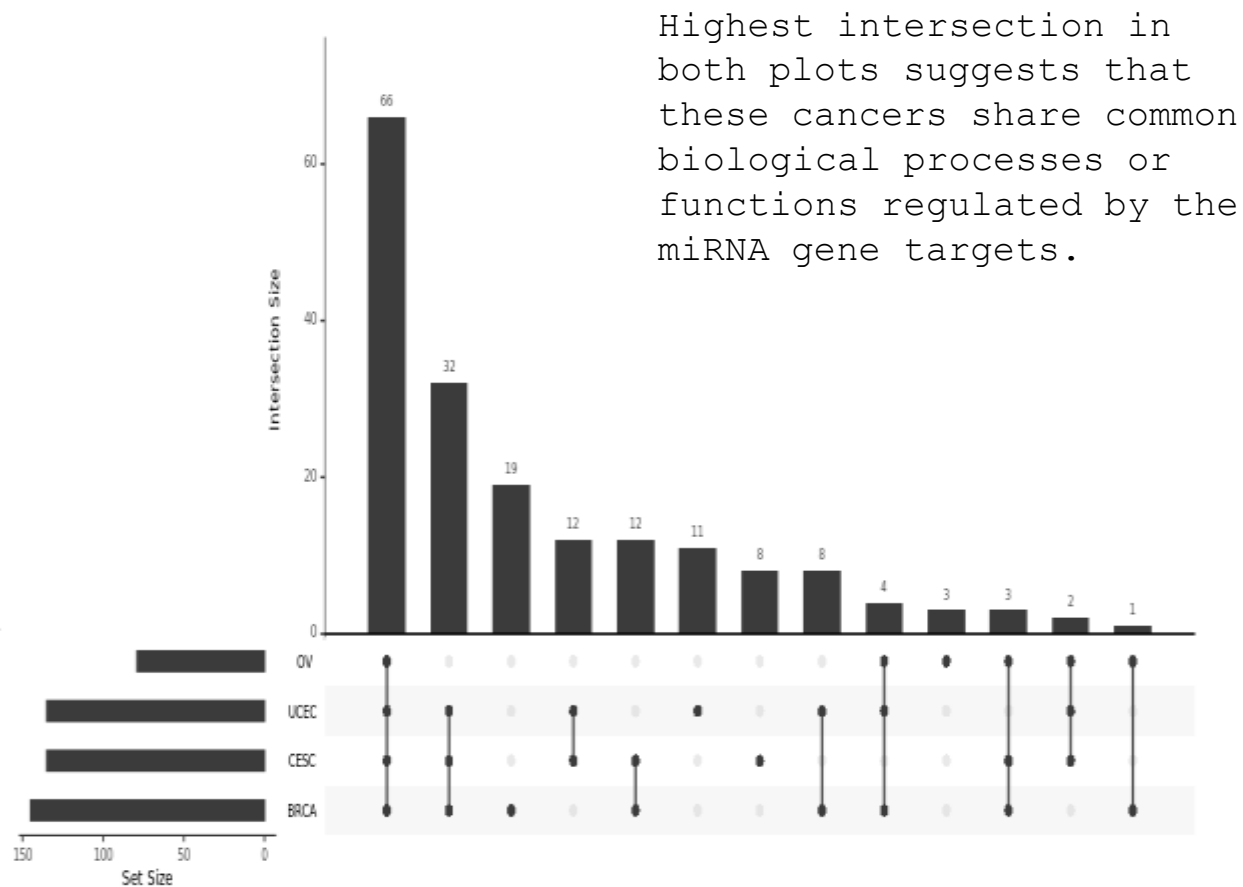
BRCA:

- Reveals significant involvement in protein processing, pathways significant to cancer progression and metastasis, infectious and neurodegenerative disease pathways.

MiRNA Targets - GSEA Overlaps



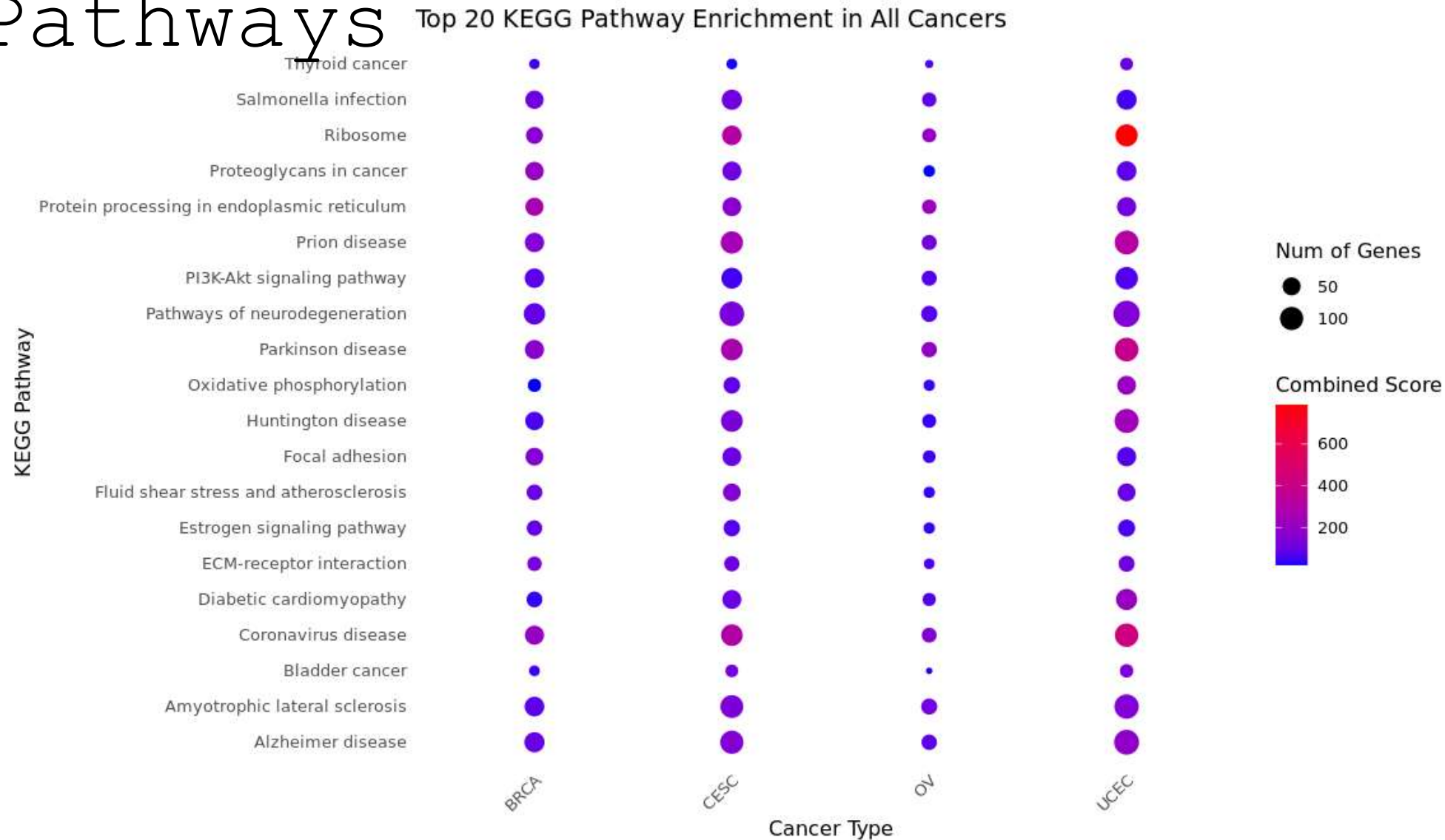
GO Term Overlaps



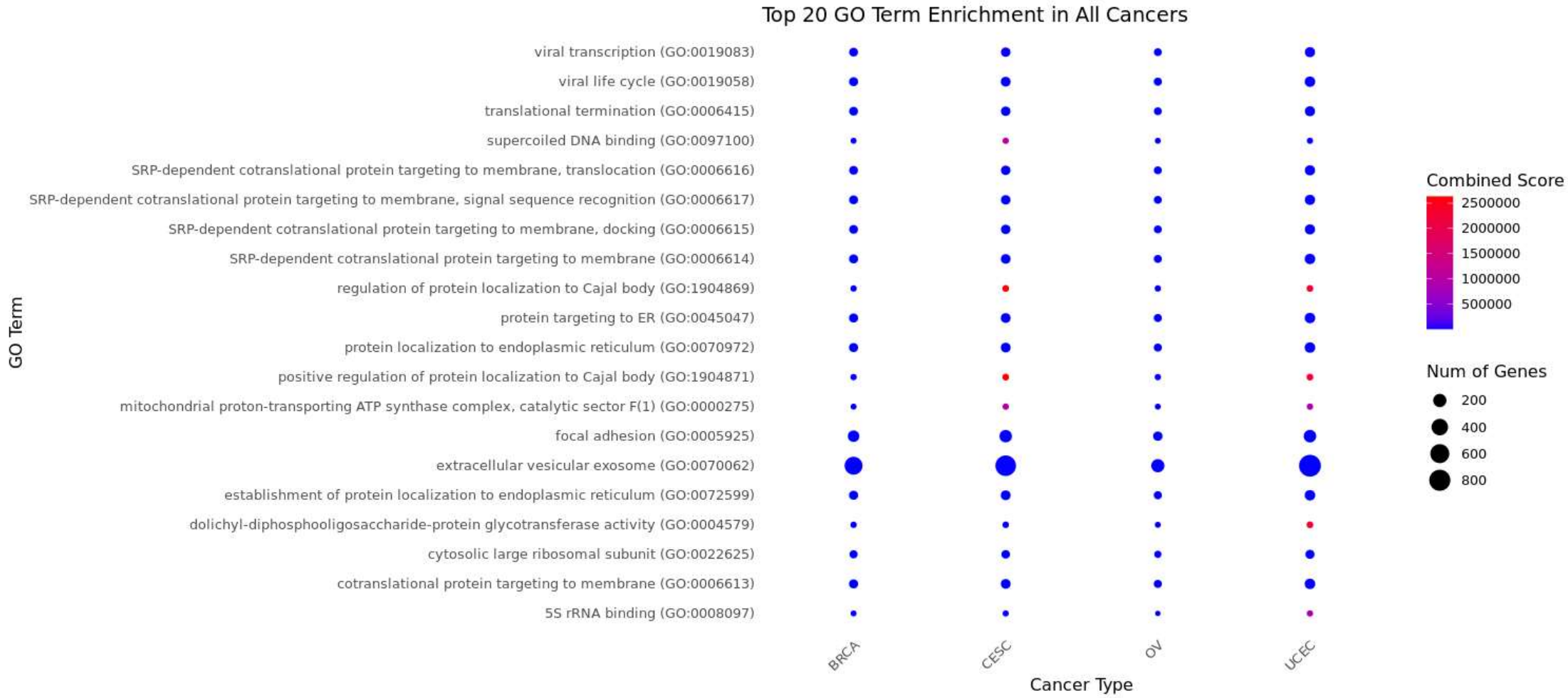
KEGG Pathway Overlaps

Highest intersection in both plots suggests that these cancers share common biological processes or functions regulated by the miRNA gene targets.

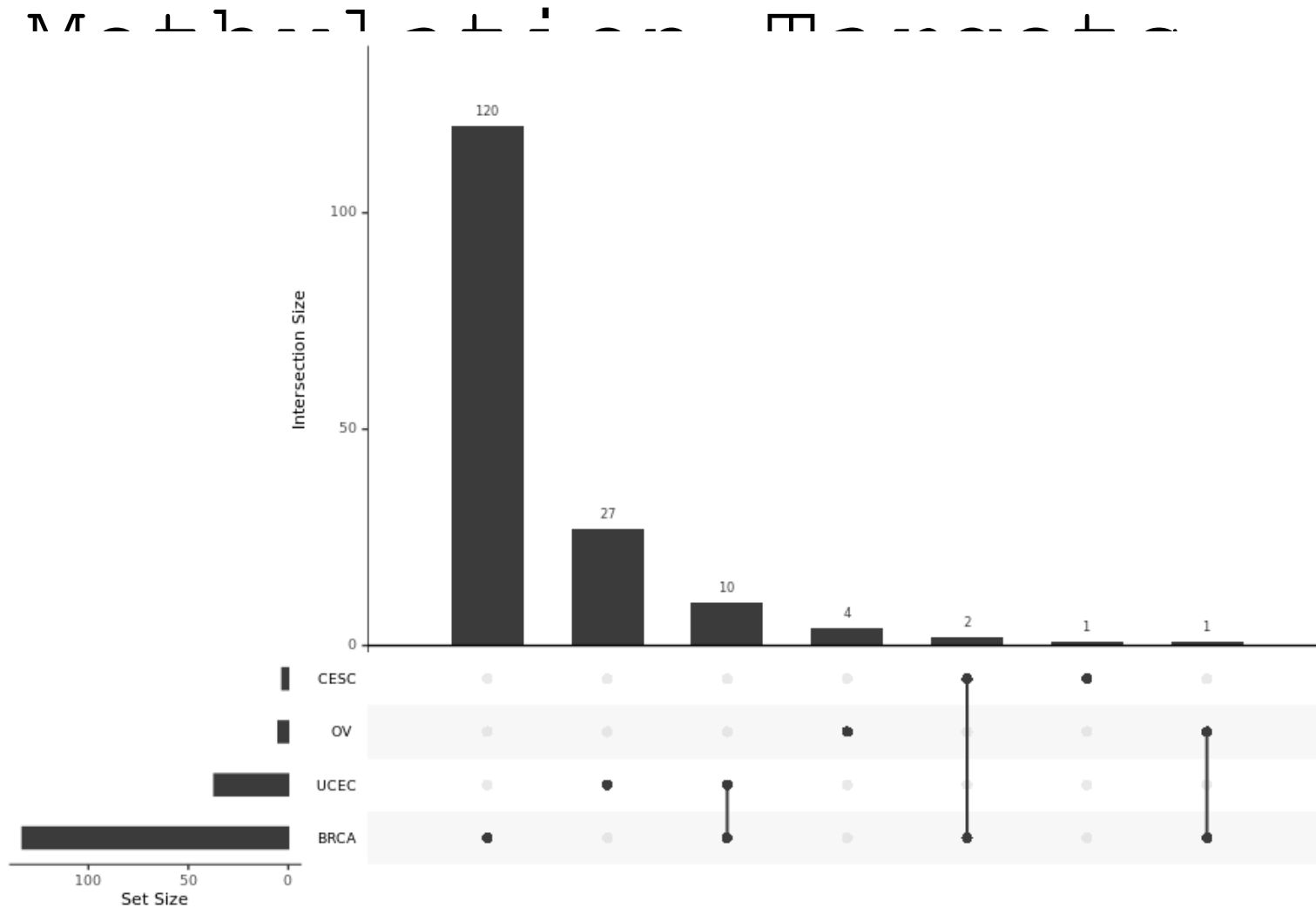
MiRNA Targets – Common Pathways



MiRNA Targets – Common GO Terms



Pan-cancer Analysis -



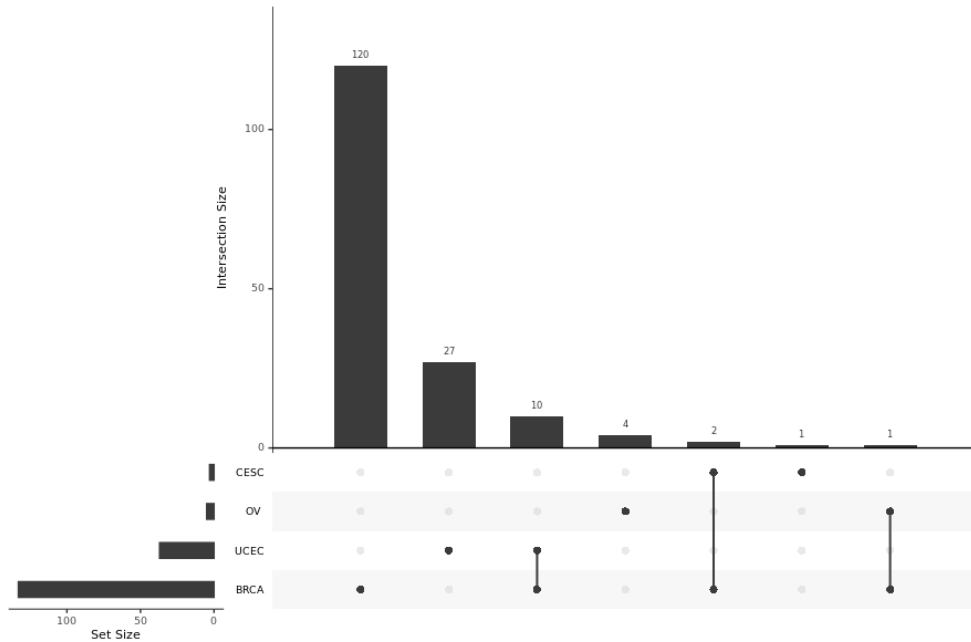
Probe Gene Target

Overlap:

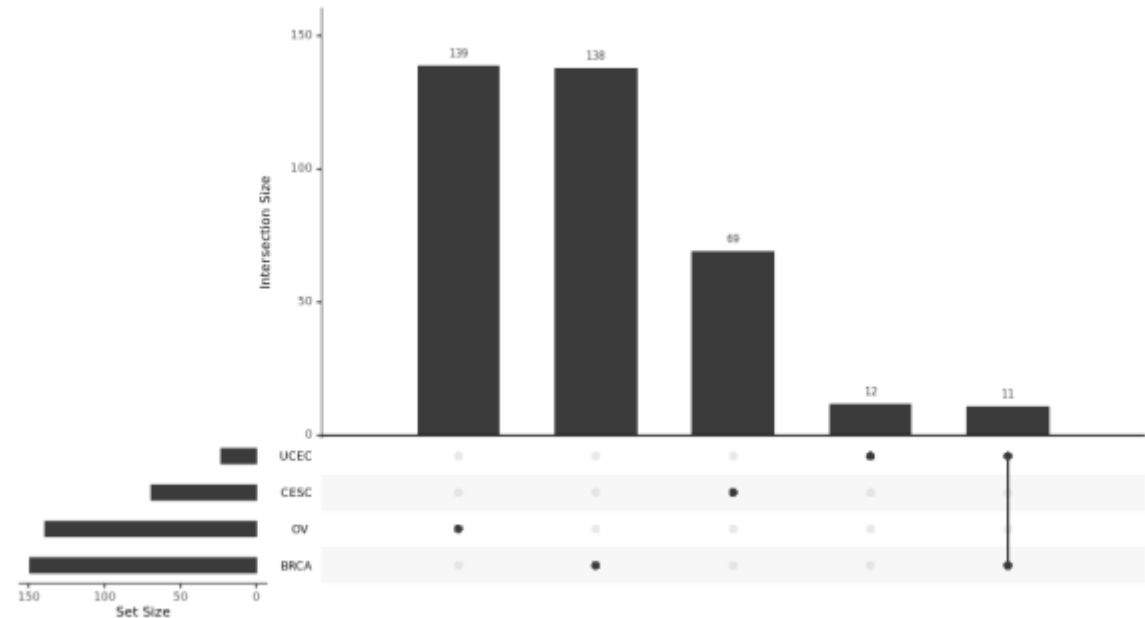
Using the Illumina Human Methylation 27k database to get annotated cg probes sites

- 1 to 1 gene association oppose to 1 to many with miRNA
- Very few genes compared to miRNA, and none common for all cancers

Methylation Targets – GSEA Overlaps



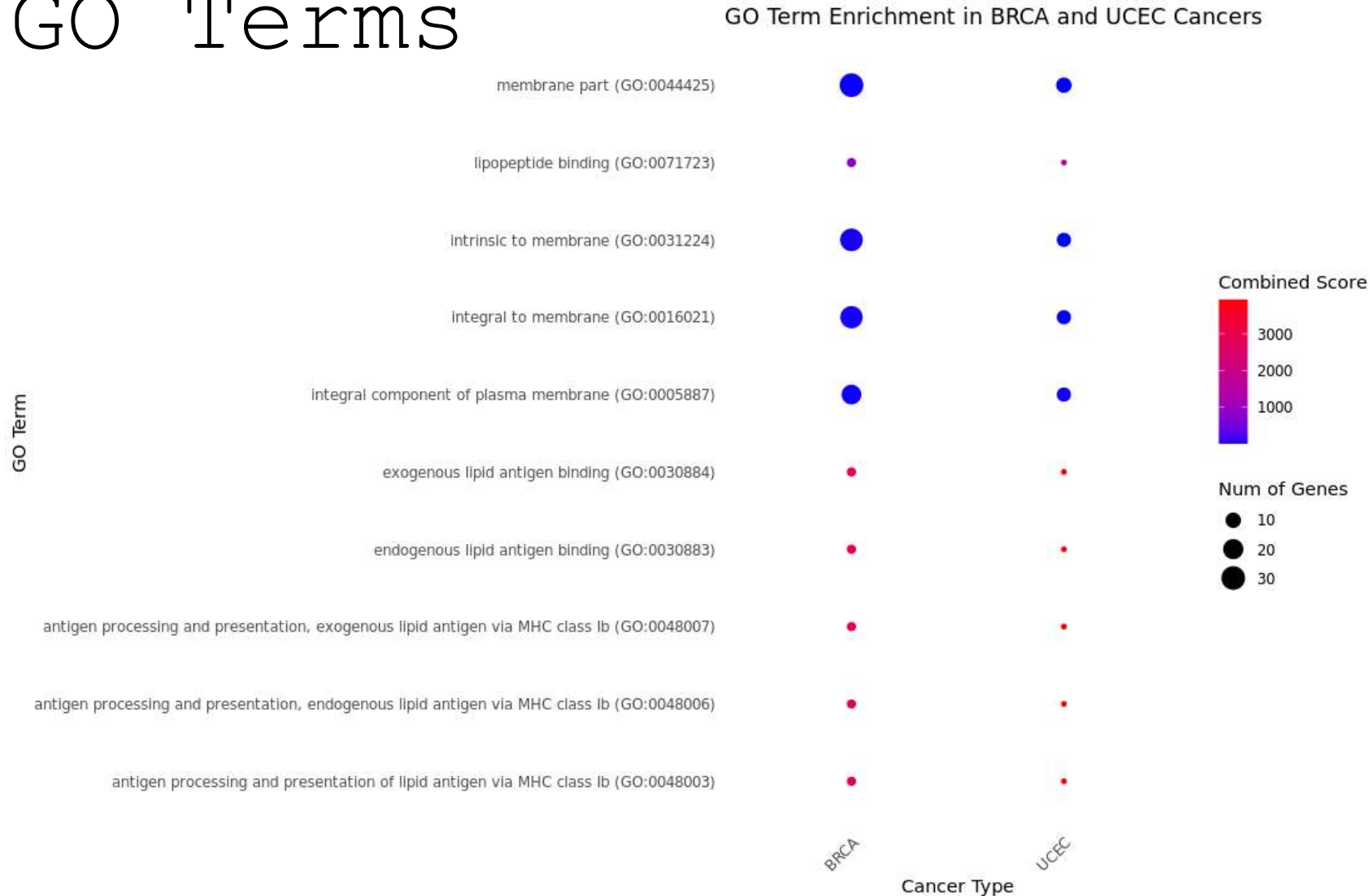
KEGG Pathways



GO Terms

Not much overlaps between Pathways and GO Terms across different cancers for each methylation target. Only UCEC and BRCA have shared GO Terms.

Methylation Targets – Common GO Terms



Significant KEGG Pathway UCEC:

Suggest these targets are involved in processes related to membrane structure and function and involvement in specific immune-related responses, related to lipid antigens.

- Might indicate immune evasion in the context of cancer, and dysregulation of these process could contribute to cancer.
- Few genes = lack of diversity

Survival Analysis – KM Plots

Kaplan-Meier Curve UCEC for miRNA: hsa.mir.22

Strata + High + Low

100%

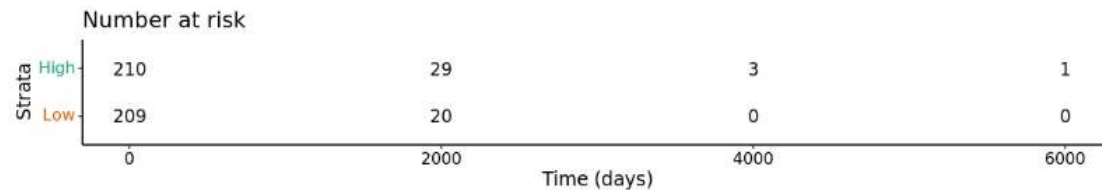
> [Mol Biol Rep.](#) 2023 Jun;50(6):5185-5193. doi: 10.1007/s11033-023-08458-6. Epub 2023 Apr 29.

miR-22-enriched breast cancer cells display repressed glycolytic metabolism, increased glycogen synthesis, and reduced survival in low glucose conditions

[Costas Koufaris](#)¹, [Margarita E Papandreou](#)², [James K Ellis](#)², [Vicky Nicolaidou](#)^{3, 4}, [Hector C Keun](#)²

Affiliations + expand

PMID: 37119413 DOI: [10.1007/s11033-023-08458-6](#)



Kaplan-Meier Curve BRCA for miRNA: hsa.mir.22

Strata + High + Low

100%

> [PLoS One.](#) 2023 Feb 7;18(2):e0281536. doi: 10.1371/journal.pone.0281536. eCollection 2023.

Hsa-miR-22-3p inhibits liver cancer cell EMT and cell migration/ invasion by indirectly regulating SPRY2

[Shuaishuai Cui](#)¹, [Yuanyuan Chen](#)¹, [Yunfei Guo](#)¹, [Xing Wang](#)², [Dahu Chen](#)¹

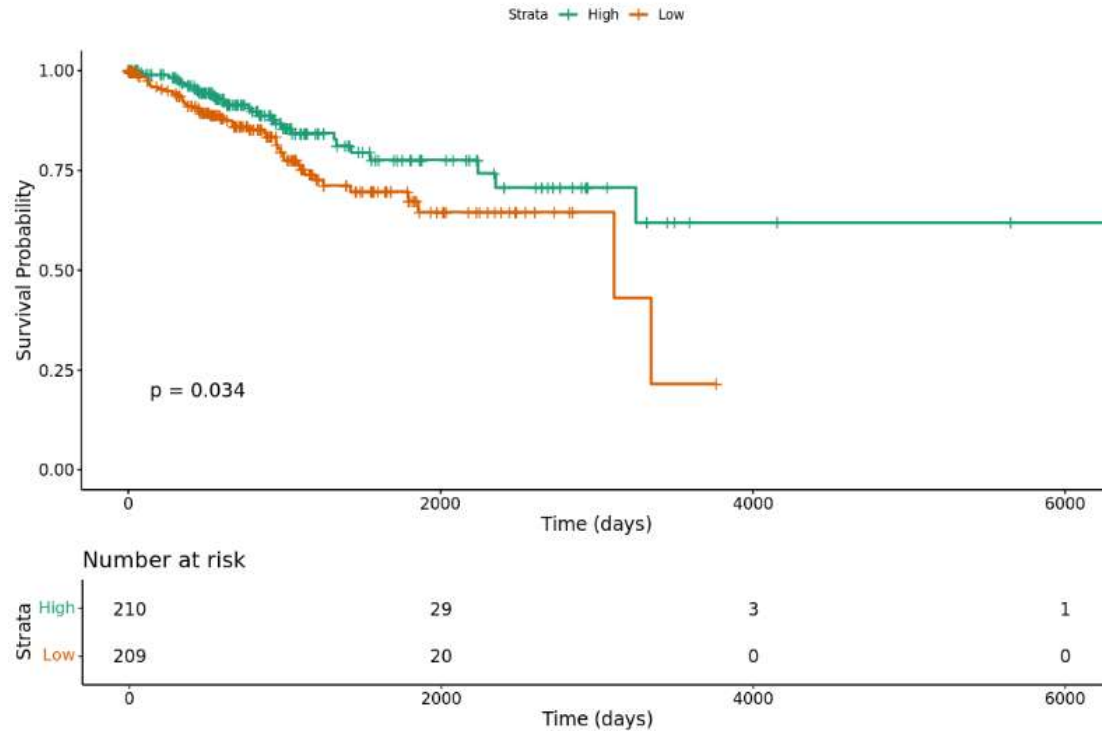
Affiliations + expand

PMID: 36749775 PMCID: [PMC9904474](#) DOI: [10.1371/journal.pone.0281536](#)

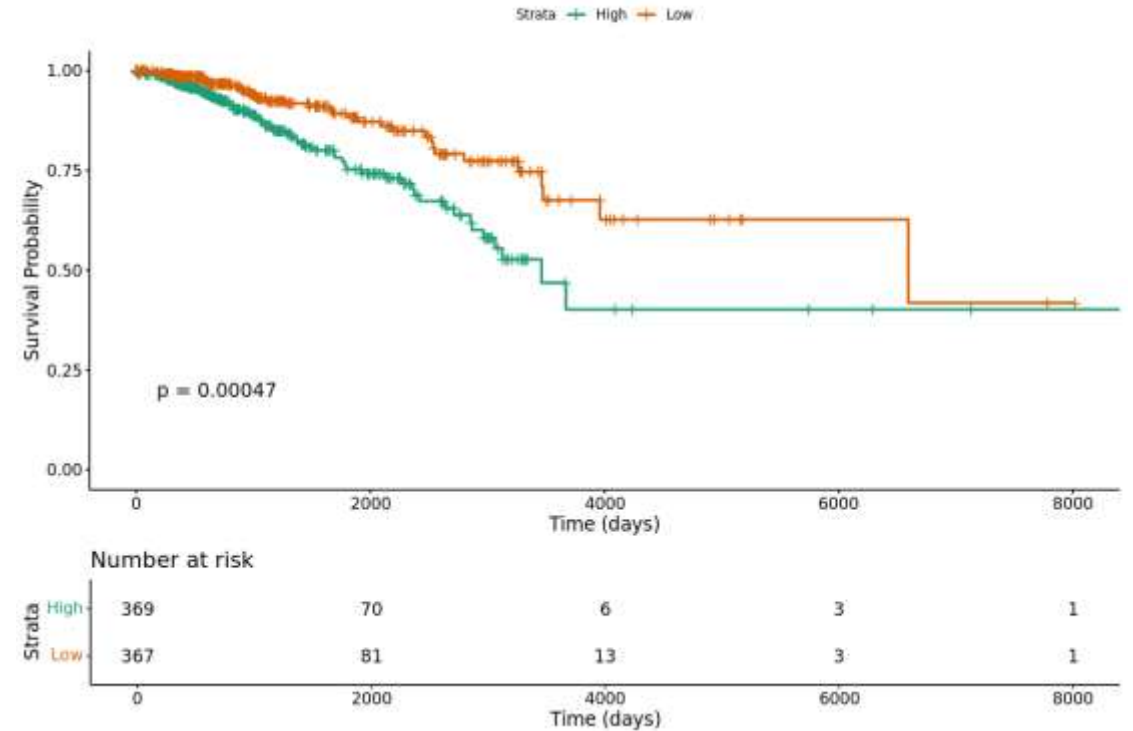


Survival Analysis – KM Plots

Kaplan-Meier Curve UCEC for miRNA: hsa.mir.22



Kaplan-Meier Curve BRCA for miRNA: hsa.mir.22



Survival Analysis – KM Plots

Oncol Lett. 2021 Sep; 22(3): 685.

Published online 2021 Jul 28. doi: [10.3892/ol.2021.12946](https://doi.org/10.3892/ol.2021.12946)

PMCID: PMC8335731 or Curve UCEC for CpG Probe: cg17525406

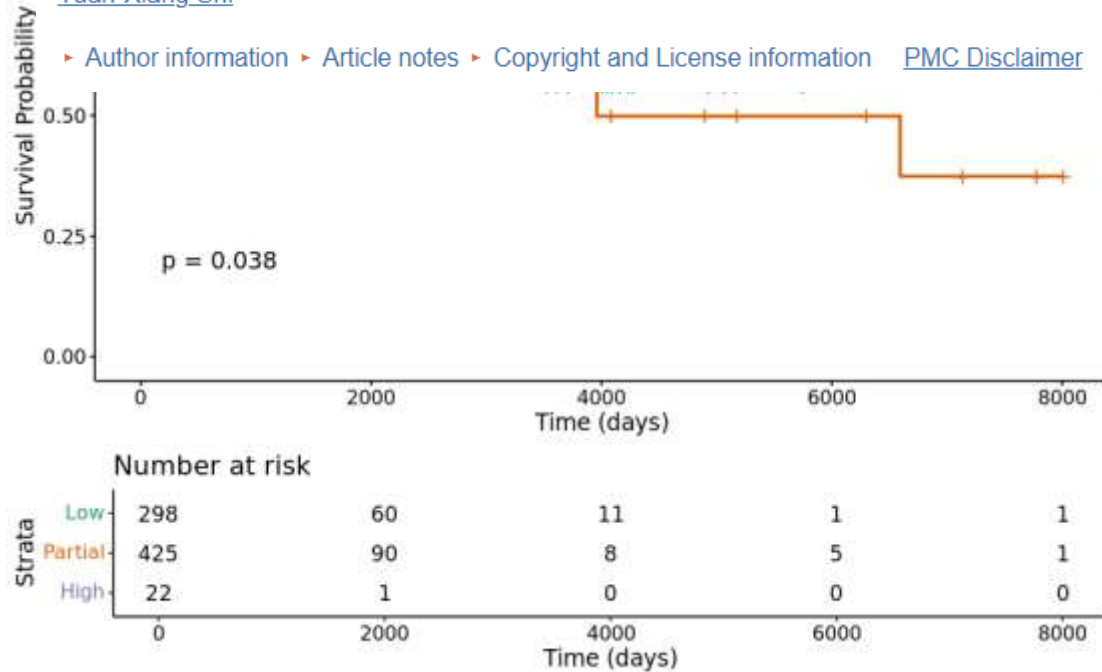
PMID: [34434284](#)

Strata Low Partial High

Identification of the molecular function of tripartite motif containing 58 in human lung cancer

Yuan-Xiang Shi

[▶ Author information](#) ▶ [Article notes](#) ▶ [Copyright and License information](#) [PMC Disclaimer](#)













Front. Oncol., 03 December 2019

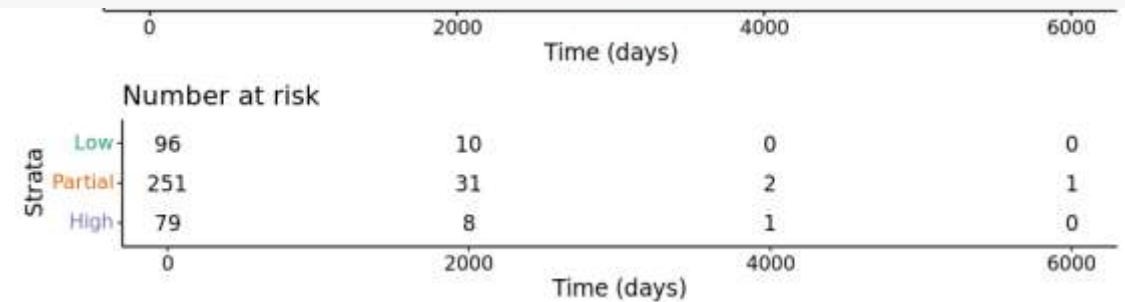
Sec. Thoracic Oncology

Volume 9 - 2019 | <https://doi.org/10.3389/fonc.2019.01281>

A Diagnostic Panel of DNA Methylation Biomarkers for Lung Adenocarcinoma

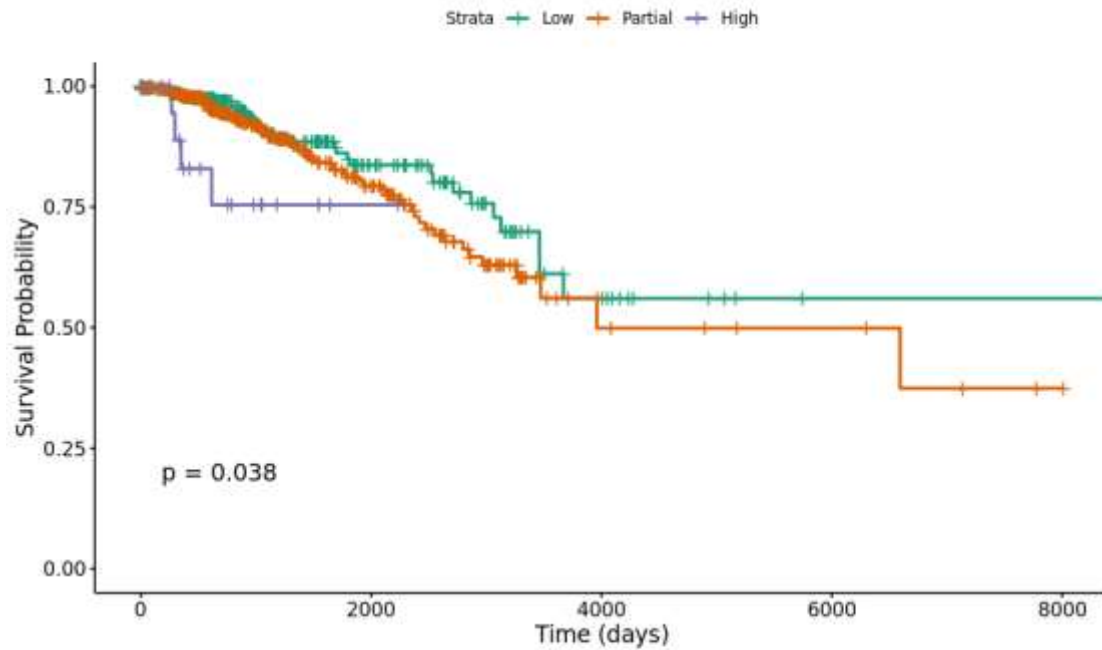
 Nan Shen^{1,2†}
 Jun Du^{3†}
 Hui Zhou^{4†}
 Nan Chen^{5†}
 Yi Pan^{6,7}
 Jörg D. Hoheisel⁶

 Zonghui Jiang⁸
 Ling Xiao^{9*}
 Yue Tao^{2*}
 Xi Mo^{2*}



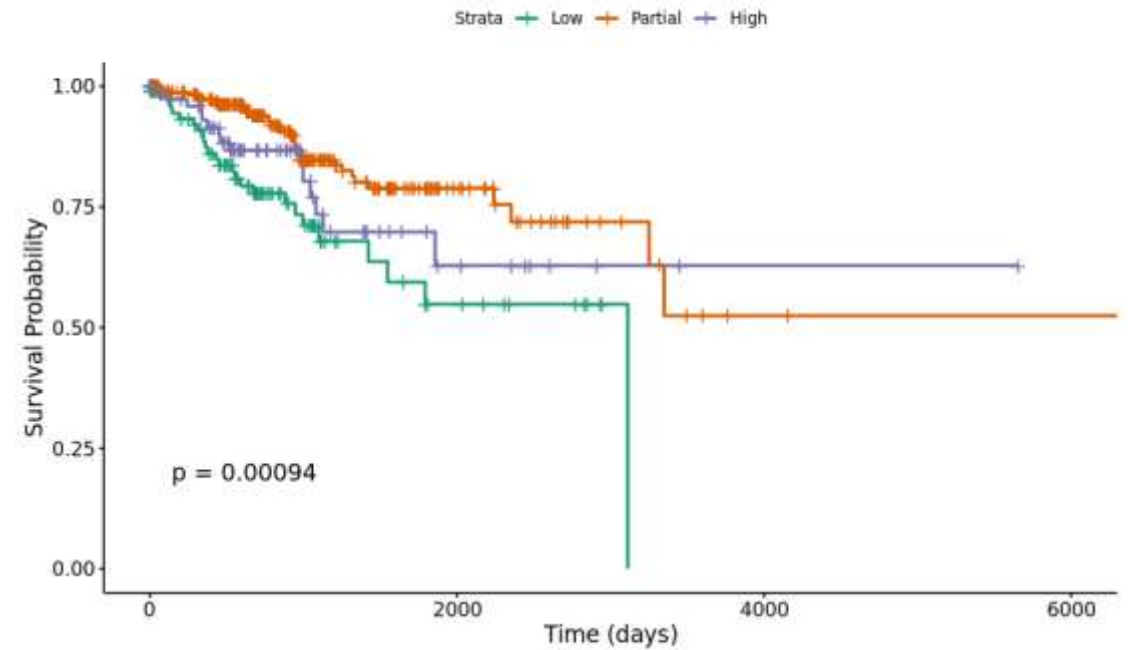
Survival Analysis – KM Plots

Kaplan-Meier Curve BRCA for CpG Probe: cg17525406



Number at risk					
Strata	Low	Partial	High		
Low	298	60	11	1	1
Partial	425	90	8	5	1
High	22	1	0	0	0
	0	2000	4000	6000	8000

Kaplan-Meier Curve UCEC for CpG Probe: cg17525406



Number at risk				
Strata	Low	Partial	High	
Low	96	10	0	0
Partial	251	31	2	1
High	79	8	1	0
	0	2000	4000	6000

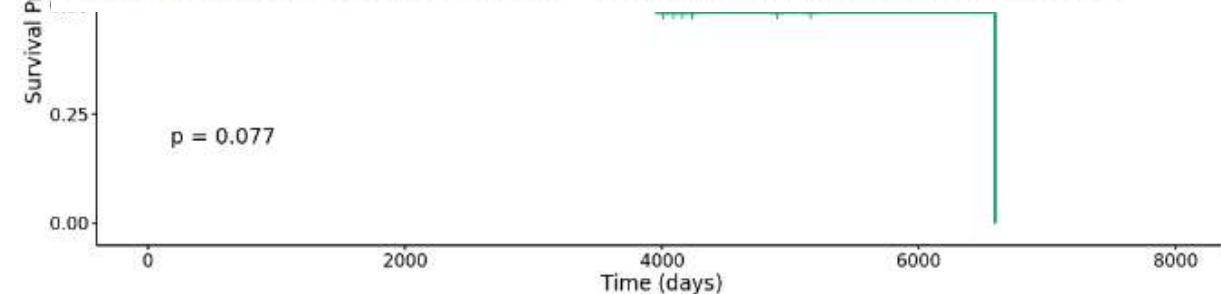
Survival Analysis – KM Plots

Role of microRNA-150 in solid tumors (Review)

Authors: Fang Wang, ✉ Xiubao Ren, ✉ Xinwei Zhang

[View Affiliations](#)

Published online on: April 30, 2015 <https://doi.org/10.3892/ol.2015.3170>



Kaplan-Meier Curve CESC for miRNA: hsa.mir.150

Strata: High Low

[Sci Rep](#) 2017; 7: 12012.

Published online 2017 Sep 20. doi: [10.1038/s41598-017-10695-2](https://doi.org/10.1038/s41598-017-10695-2)

PMCID: PMC5607379

PMID: [28931826](https://pubmed.ncbi.nlm.nih.gov/28931826/)

MicroRNA profiling study reveals miR-150 in association with metastasis in nasopharyngeal carcinoma

[Patrick Ying-Kit Yue](#),^{#1} [Wai-Yan Ha](#),^{#1} [Chi-Chiu Lau](#),¹ [Florence Man-Fung Cheung](#),^{2,3} [Anne Wing-Mui Lee](#),^{2,4} [Wai-Tong Ng](#),^{2,4} [Roger Kai-Cheong Ngan](#),^{2,5} [Chun-Chung Yau](#),^{2,6} [Dora Lai-Wan Kwong](#),^{2,7} [Hong-Lok Lung](#),^{1,2} [Nai-Ki Mak](#),^{1,2} [Maria Li Lung](#),^{2,8} and [Ricky Ngok-Shun Wong](#)^{✉1}

[▶ Author information](#) [▶ Article notes](#) [▶ Copyright and License information](#) [PMC Disclaimer](#)



[Oncol Res](#) 2019; 27(3): 317–323.

Published online 2019 Feb 21. doi: [10.3727/096504018X15228863026239](https://doi.org/10.3727/096504018X15228863026239)

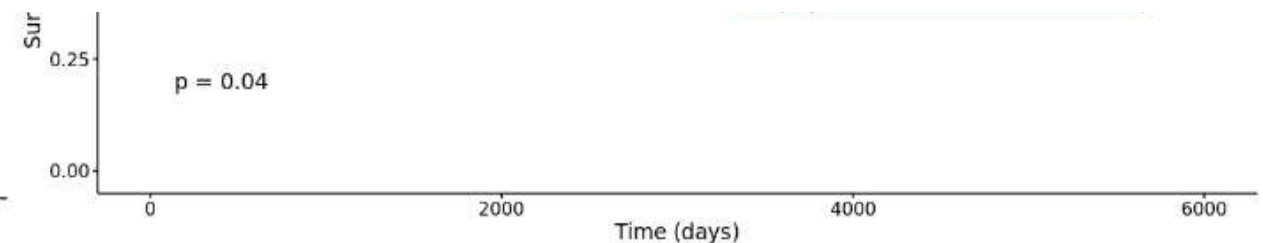
PMCID: PMC7848275

PMID: [29690954](https://pubmed.ncbi.nlm.nih.gov/29690954/)

miR-150 Suppresses Tumor Growth in Melanoma Through Downregulation of MYB

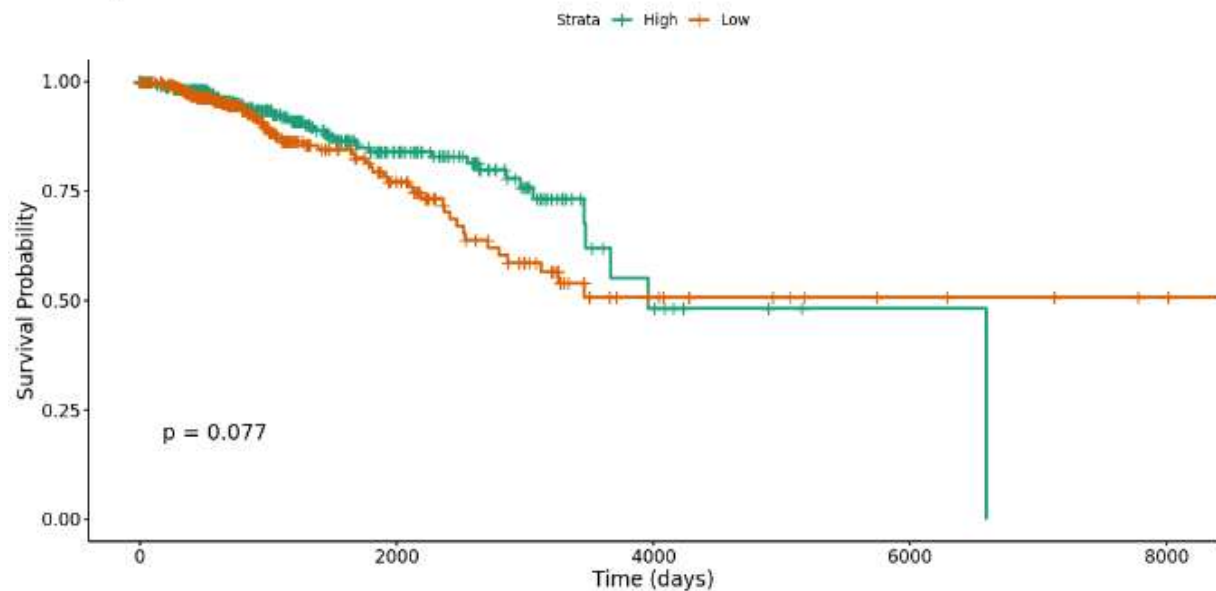
[Xiyan Sun](#),^{*1} [Chao Zhang](#),^{†1} [Yang Cao](#),[†] and [Erbiao Liu](#)^{*}

[▶ Author information](#) [▶ Copyright and License information](#) [PMC Disclaimer](#)

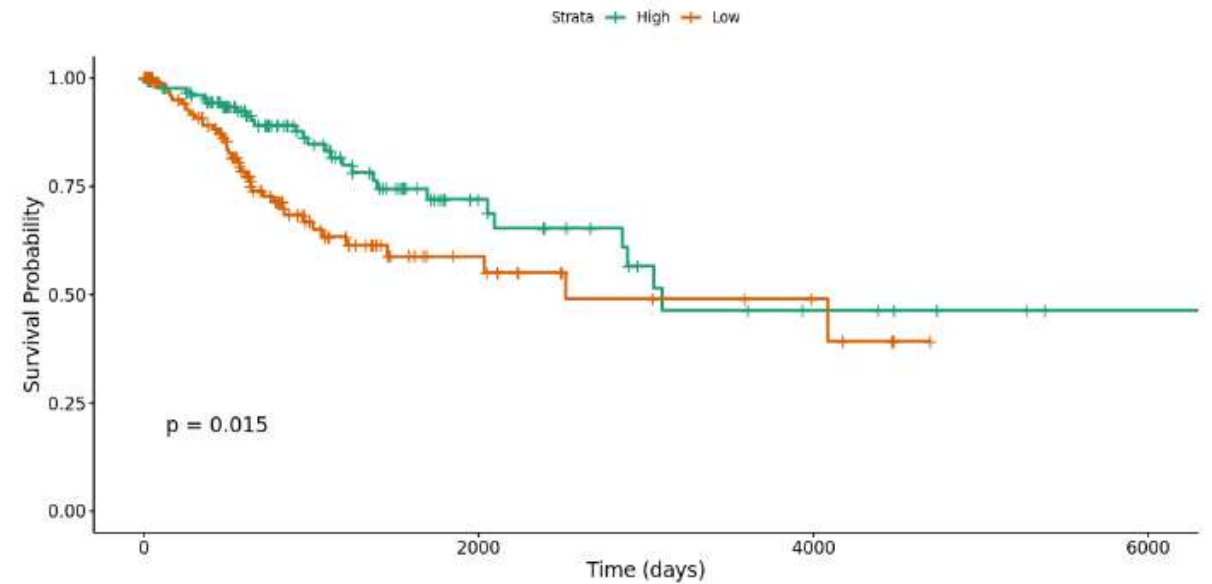


Survival Analysis – KM Plots

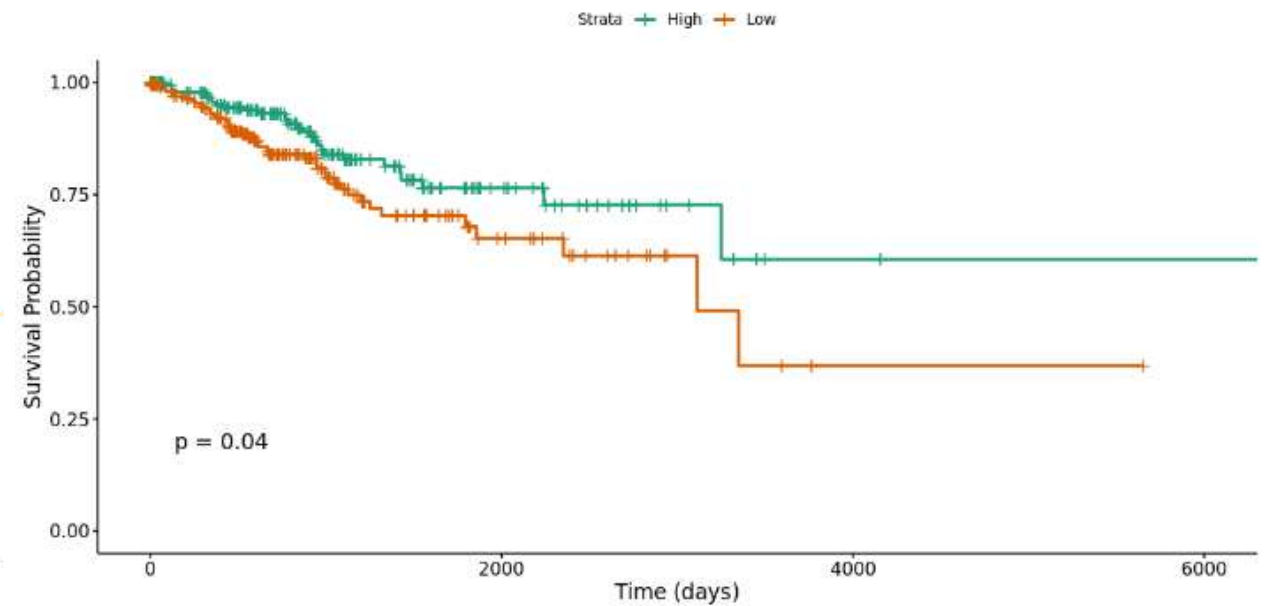
Kaplan-Meier Curve BRCA for miRNA: hsa.mir.150



Kaplan-Meier Curve CESC for miRNA: hsa.mir.150



Kaplan-Meier Curve UCEC for miRNA: hsa.mir.150



Conclusion and summary of findings

- We have shown that multi-modal integration does improve survival predictions, by beating the baseline single-omics survival predictions.
- Random Forest seems to be one of the better survival ML algorithms that can handle High dimensional data.
- The best multi-modal combination was the integration of Methylation, miRNA and copy number variation data across 3 out of 4 cancers. However, all of them did have methylation and miRNA data
- There was not any feature that was in all four cancers, with only mir150 being in 3. However, upon further investigation of gene targets there were significant overlaps in GO Terms and KEGG Pathways.
- Features shared common oncogenic pathways as well as infectious and neurodegenerative disorders, shown in both GSEA and disease-association network analysis.
- The multi-modal signatures explored in the KM analysis did reveal not only survival significance but biological relevance with literature.

