ASAH1 rGBM Manuscript ELISA Figure

Justin Sing, University of Toronto

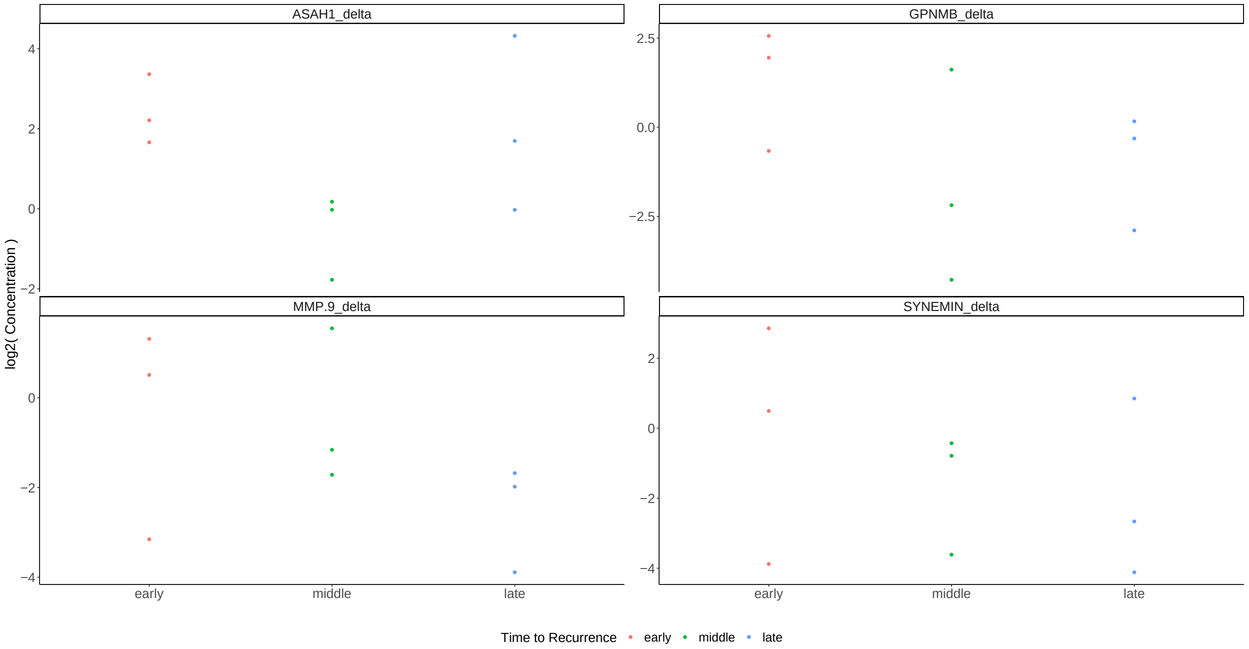
Thu December 15 11:01:11 AM EST 2022

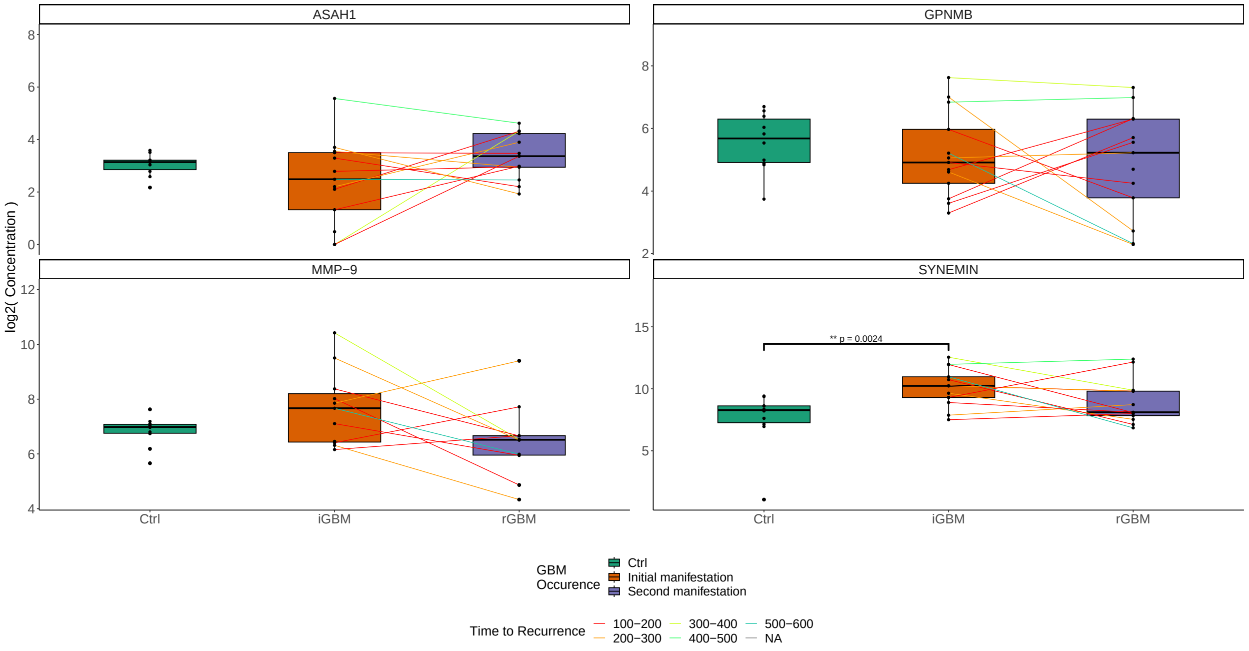
Contents

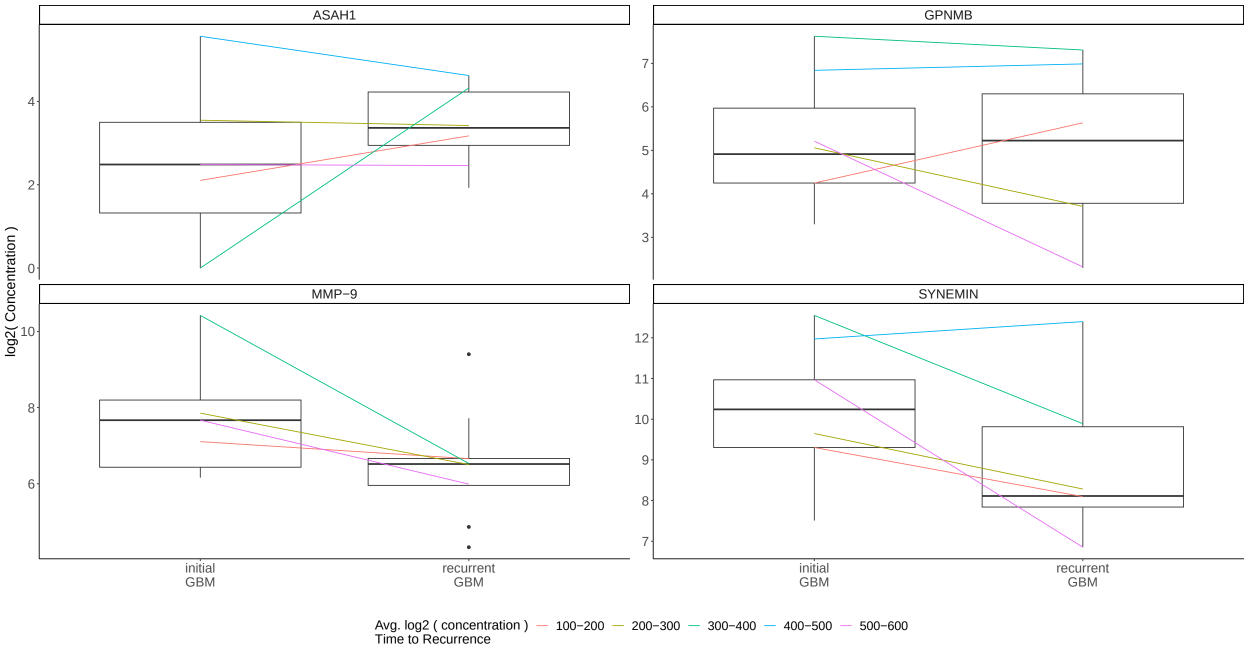
```
##
## Call:
## lm(formula = Time.to.reccurence..days. ~ ASAH1_delta + SYNEMIN_delta +
##     GPNMB_delta + MMP.9_delta, data = ttr_delta)
##
## Residuals:
##        1        2        3        4        5        6        7        8
##   58.150  -60.848  -22.659   -8.495 -104.509  -52.616   55.801  -11.958
##        9
##  147.133
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     170.74      55.27   3.089   0.0366 *
## ASAH1_delta      10.28      29.43   0.349   0.7446
## SYNEMIN_delta   -12.01      26.90  -0.446   0.6785
## GPNMB_delta     -37.57      19.36  -1.941   0.1243
## MMP.9_delta     -26.99      38.94  -0.693   0.5263
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 107.5 on 4 degrees of freedom
## Multiple R-squared:  0.6284, Adjusted R-squared:  0.2569
## F-statistic: 1.691 on 4 and 4 DF,  p-value: 0.3116
```

```
## Analysis of Variance Table
##
## Response: Time.to.reccurence..days.
##               Df Sum Sq Mean Sq F value Pr(>F)
## ASAH1_delta    1    172     172  0.0149 0.9088
## SYNEMIN_delta  1  33047   33047  2.8572 0.1662
## GPNMB_delta    1  39473   39473  3.4128 0.1384
## MMP.9_delta    1   5557    5557  0.4805 0.5263
## Residuals      4  46265   11566
```

```
## [1] "The estimated sample size needed for 80% power is: 29"
```


ASAH1_delta
GPNMB_delta
MMP.9_delta
SYNEMIN_delta
log2(Concentration)
4
2
0
−2
2.5
0.0
−2.5
−4
early
middle
late
Time to Recurrence
early
middle
late


ASAH1
GPNMB
MMP−9
SYNEMIN
log2(Concentration)
** p = 0.0024
8
6
4
2
0
12
10
15
5
Ctrl
iGBM
rGBM
GBM Occurence
Ctrl
Initial manifestation
Second manifestation
Time to Recurrence
100−200
200−300
300−400
400−500
500−600
NA


ASAH1
GPNMB
MMP−9
SYNEMIN
log2(Concentration)
0
2
4
3
4
5
6
7
6
8
10
7
8
9
10
11
12
initial GBM
recurrent GBM
initial GBM
recurrent GBM
Avg. log2 (concentration)
Time to Recurrence
100−200
200−300
300−400
400−500
500−600

Master Figure

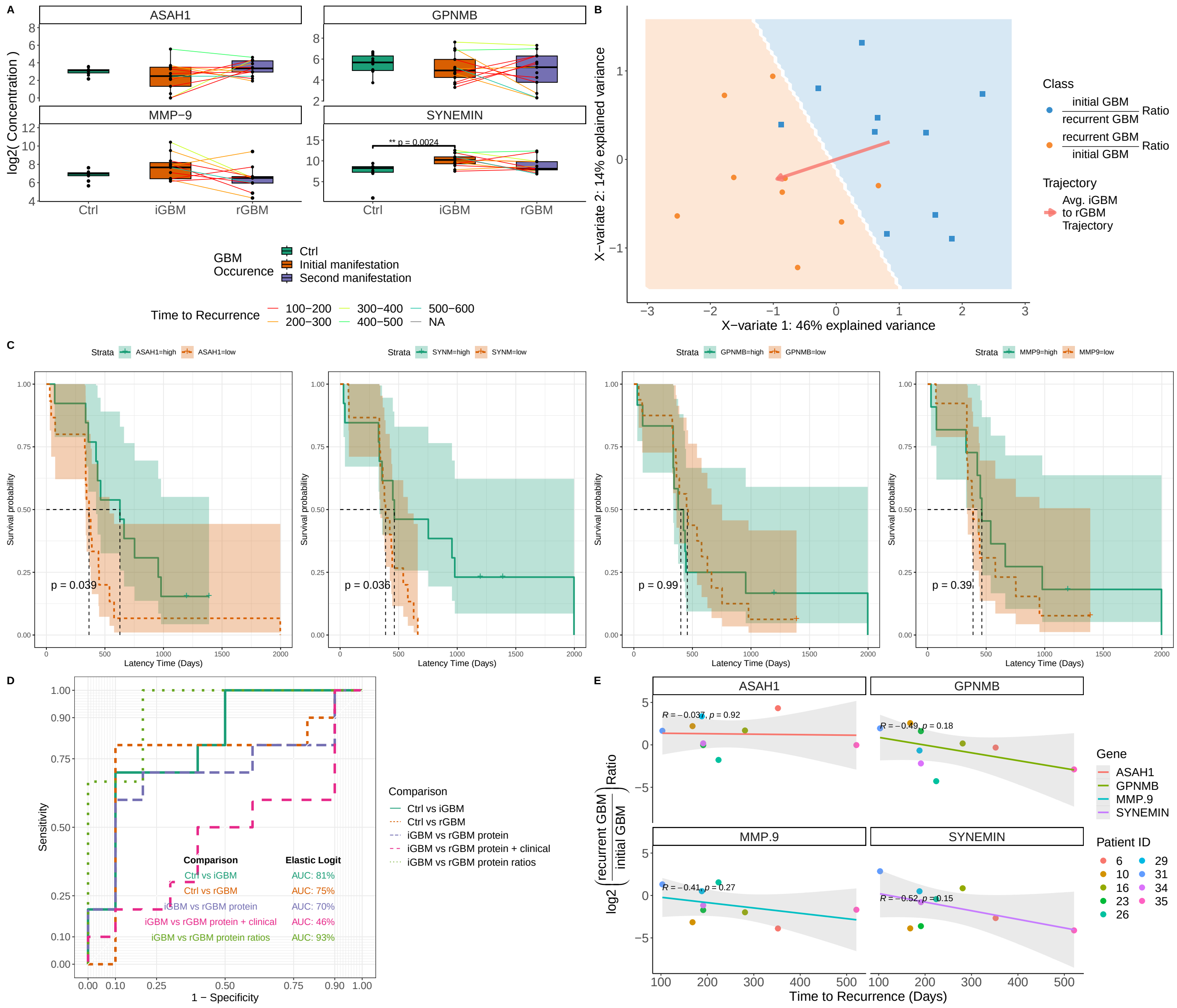
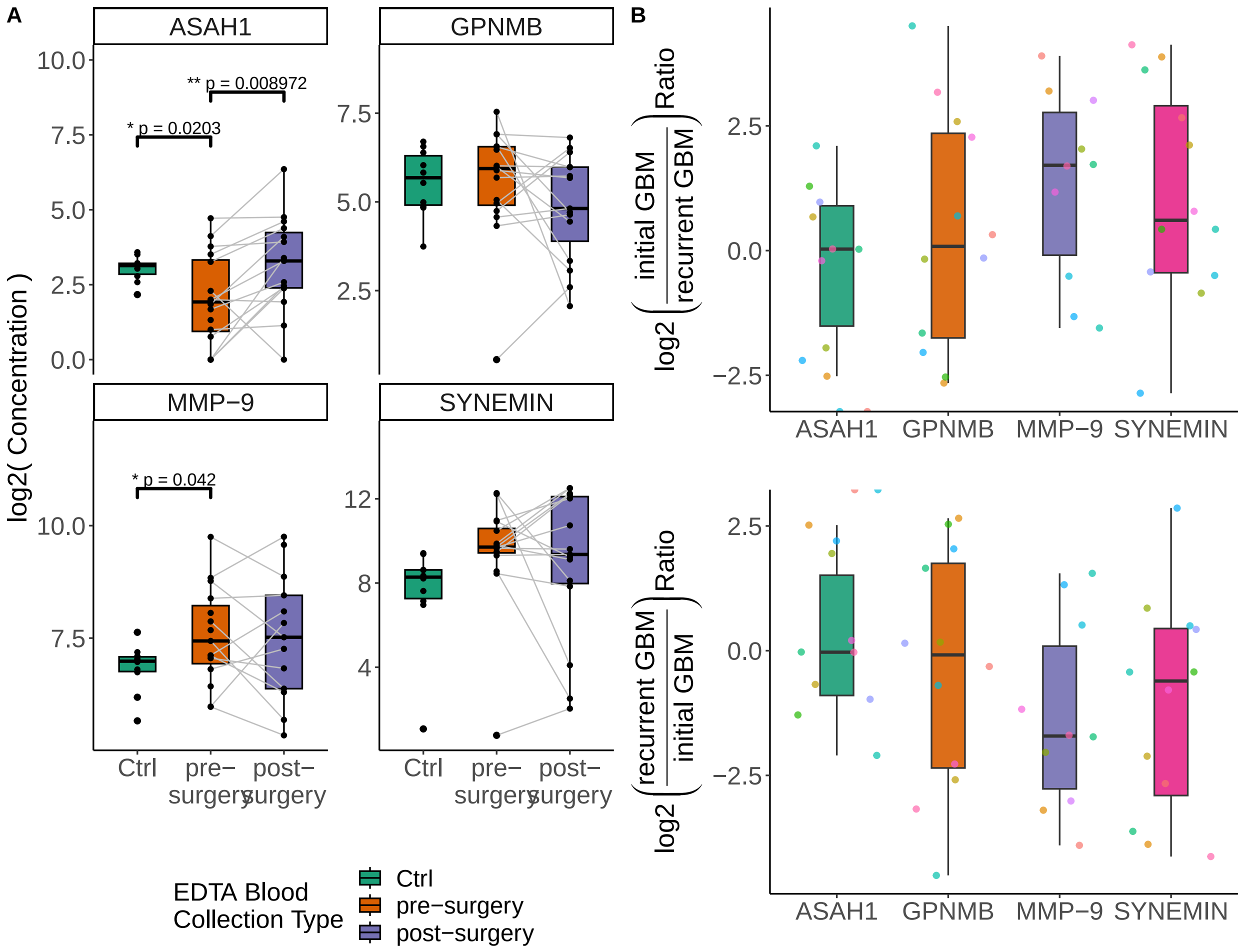


Figure 3: A Promising Plasma Proteomics Signature for Stratification and Prediction of Recurrent Glioblastoma Multiforme. Concentration of four proteins identified from TMT-MS tissue proteomics, were measured in patients' plasma samples using Enzyme-Linked Immunosorbent Assay (ELISA). (**a**) The log2 transformed protein concentration measurements demonstrates the difference in log2 concentration distributions between control, initial GBM and paired recurrent GBM populations. (**b**) Reducing the data of the four proteins ratios demonstrates an average trajectory and separation of $\frac{initial \quad GBM}{recurrent \quad GBM}$ vs $\frac{recurrent \quad GBM}{initial \quad GBM}$. (**c**) Kaplan-Meyer plots showing demonstrating high concentration of ASAH1 and high concentration of SYNEMIN contributing to a higher probability of longer survival. (**d**) The receiver operating characteristic (ROC) shows the potential of using plasma proteomics for separating each population, driven mostly from information contained in the proteins. (**e**) The potential of recurrent GBM / initial GBM protein ratios for predicting time to recurrence. Note, each point represents a single patient, so we may not be able to directly state a specific ratio can be used to infer to time to recurrence. However, we can still state given a specific patients ratio, we can predict that patients time to recurrence.

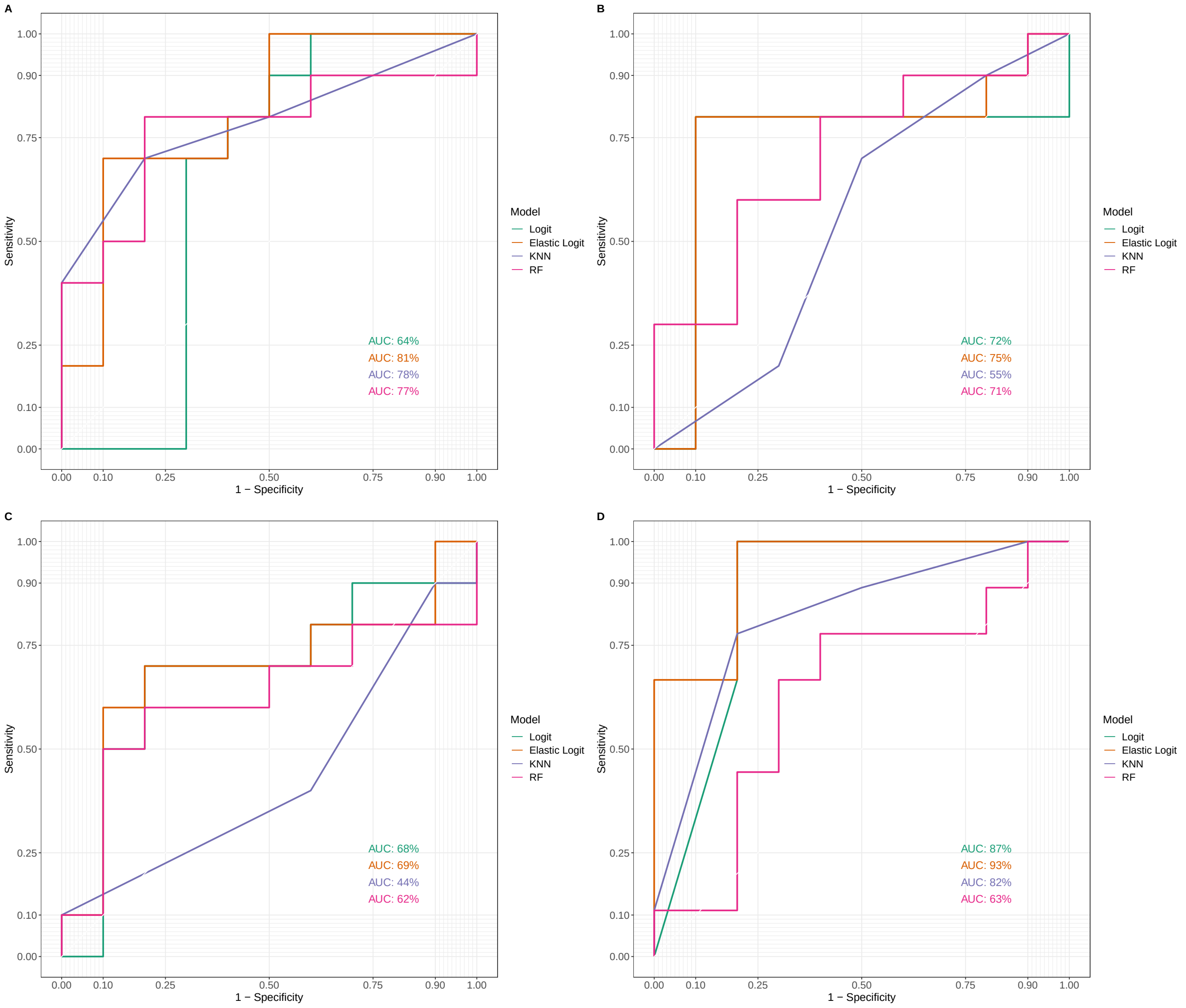
Note: I'm not entirely sure about plot e, and if my thinking/interpretation is right, but let me know what you think.

Note: I excluded the pre-surgery vs post-surgery distribution boxplots, because I didn't do any prediction analyses on that set. We can still include it in the supplement if we thing there is anything interesting there to show?

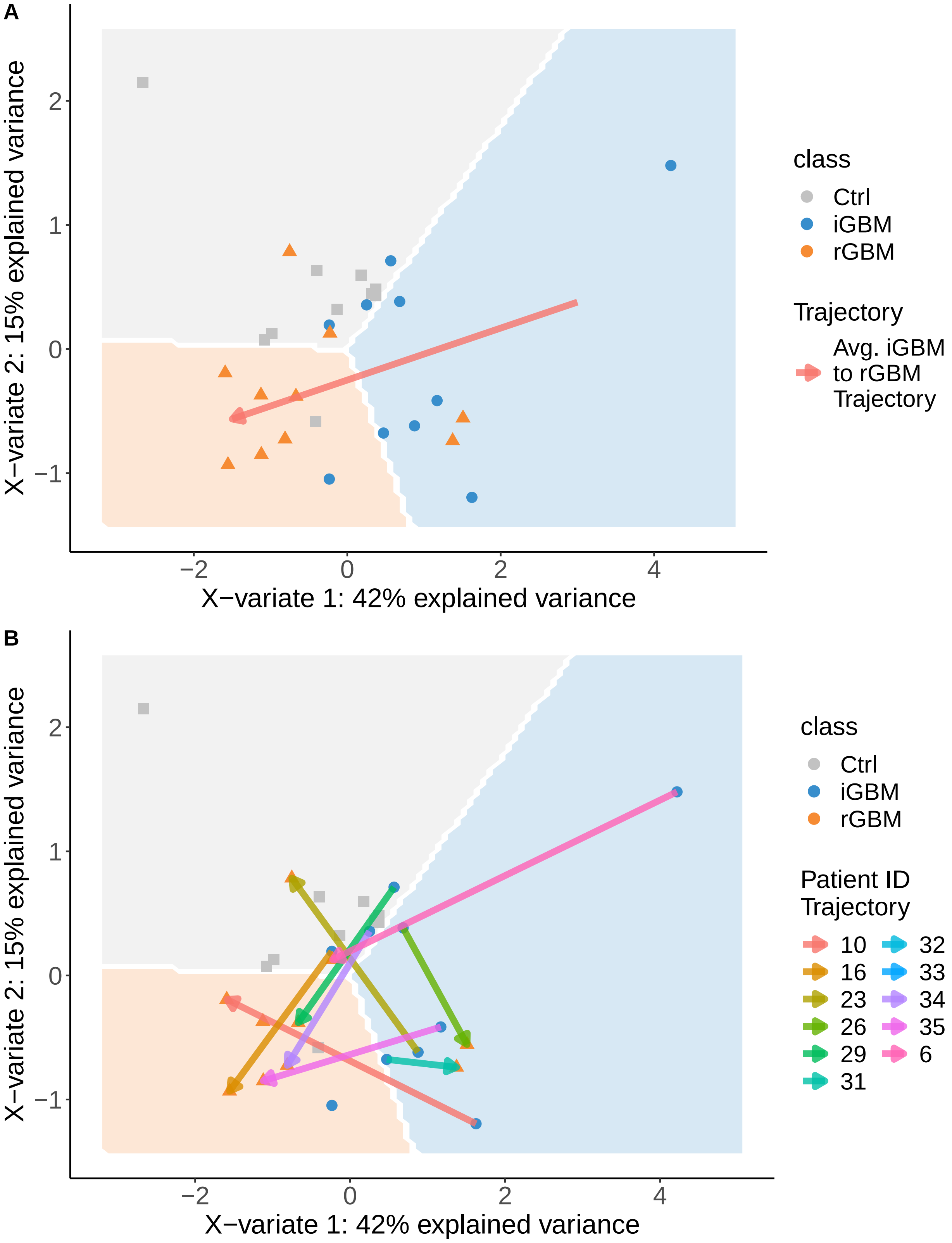
Supplemental Figures and Tables



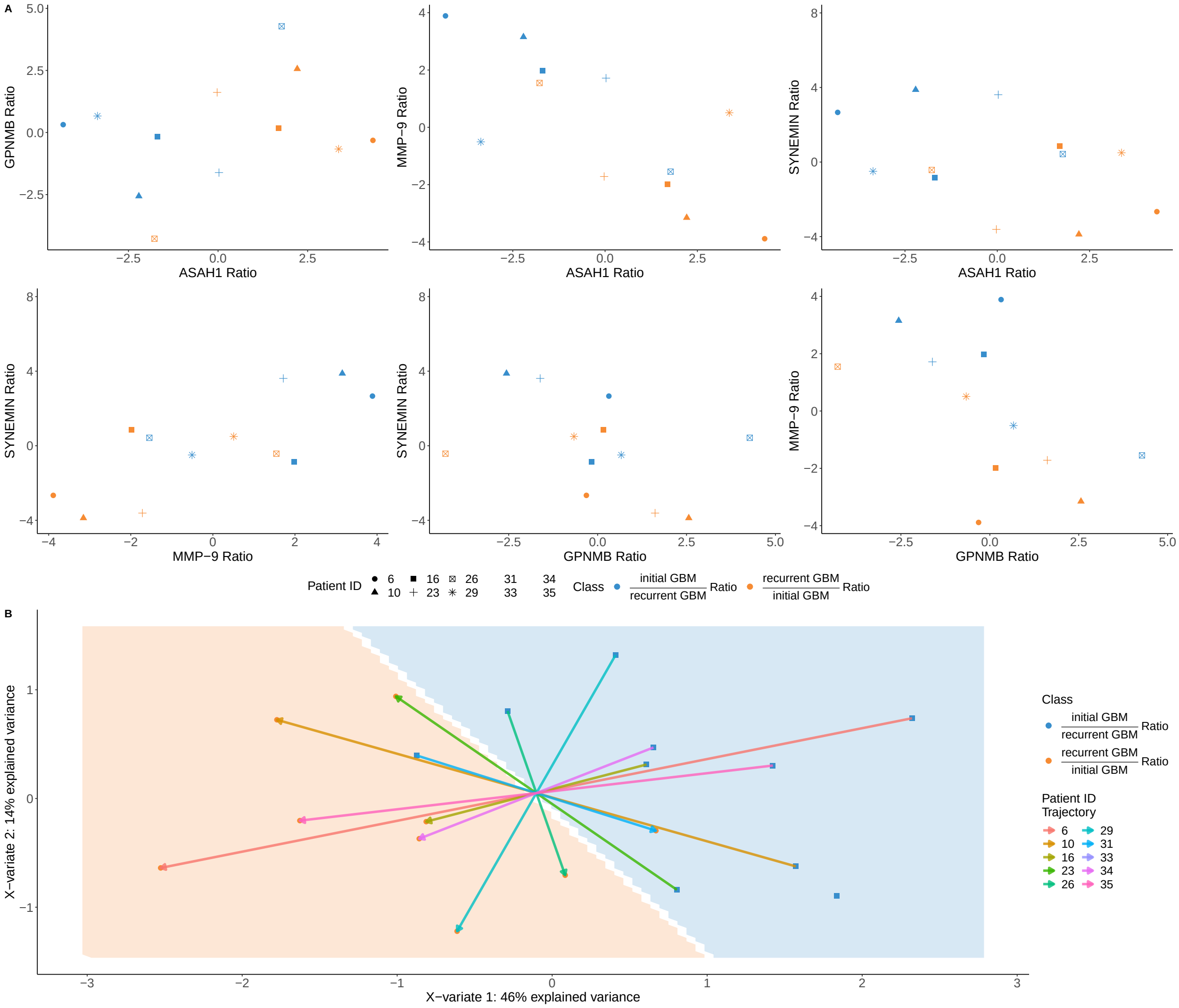
Supplemental Figure 1 : Distribution of Pre- vs Post-Surgery Plasma Proteomics and Ratios of Initial and Recurrent GBM Plasma Proteomics. (**a**) The log2 transformed protein concentration measurements demonstrates the difference in log2 concentration distributions between control, pre-surgery and post-surgery populations. (**b**) Demonstrates the log transformed $\frac{initial \quad GBM}{recurrent \quad GBM}$ ratio (top) as well as the inverse ratio (bottom) for the potential use of identifying a patients state to recurrence using these ratios.



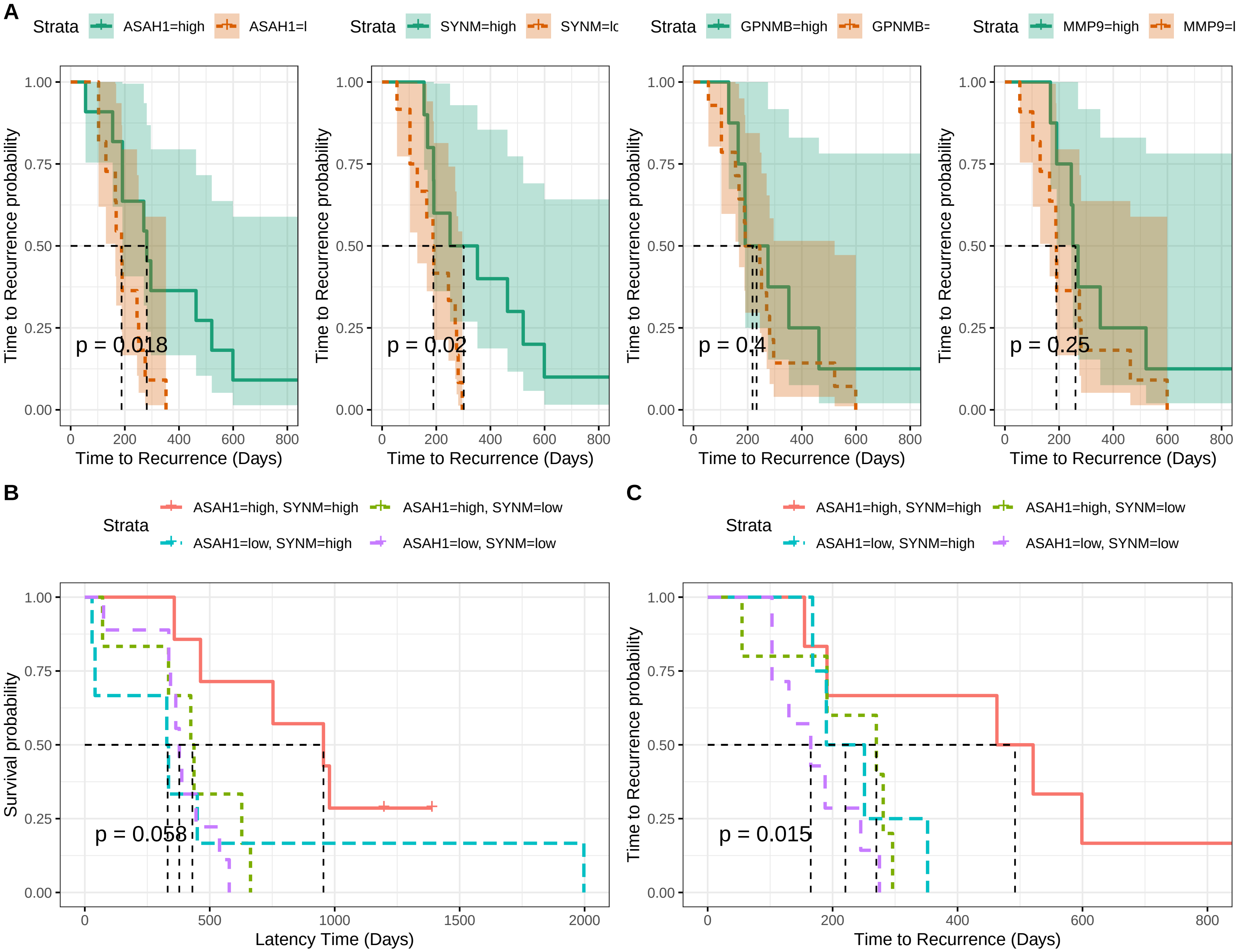
Supplemental Figure 2 : Variation in Predictive Performance using Different Models. Testing parametric (logistic regression, logistic regression with elastic net regularization), non-parametric (k-nearest neighbours) and ensemble (random forest) supervised learning algorithms on different prediction tasks. (**a**) Most models have AUCs greater than 60% when predicting control vs initial GBM. (**b**) Most models have comparable AUCs, except for k-nearest neighbours when predicting for initial GBM vs recurrent GBM. (**c**) Including clinical data for *Tumor localization*, *Sex*, *Age at surgery (years)*, *Type of resection* reduces predictive performance for all models. (**d**) All models, except random forest, have AUCs above 80% when predicting initial GBM vs recurrent GBM using protein ratios.



Supplemental Figure 3 : Individual Patient Trajectory in Low Dimensional Space. (**a**) A sparse partial least squares discriminant analysis shows separation between the different population groups, as well as an overall trajectory of a initial GBM patient going towards a recurrent GBM patient in a low dimensional space. (**b**) Individual trajectories demonstrate the individual variance of patients path to recurrence in a low-dimensional space.



Supplemental Figure 4 : Information Contained in Plasma Proteomic Ratios for Initial GBM vs Recurrent GBM Stratification. (**a**) Pairwise protein ratios shows promise of individual protein-pair ratios being able to stratify initial vs recurrent plasma GBM samples. (**b**) Individual trajectories demonstrate the individual variance of patients path to recurrence in a low-dimensional space using ratios.


A
Strata
ASAH1=high
ASAH1=l
SYNM=high
SYNM=lc
GPNMB=high
GPNMB=
MMP9=high
MMP9=lc
Time to Recurrence probability
1.00
0.75
0.50
0.25
0.00
p = 0.018
p = 0.02
p = 0.4
p = 0.25
0
200
400
600
800
Time to Recurrence (Days)
B
Strata
ASAH1=high, SYNM=high
ASAH1=high, SYNM=low
ASAH1=low, SYNM=high
ASAH1=low, SYNM=low
Survival probability
p = 0.058
0
500
1000
1500
2000
Latency Time (Days)
C
Strata
ASAH1=high, SYNM=high
ASAH1=high, SYNM=low
ASAH1=low, SYNM=high
ASAH1=low, SYNM=low
Time to Recurrence probability
p = 0.015
0
200
400
600
800
Time to Recurrence (Days)

Supple-

mental Figure 5 : Survival and Time to Recurrence Analysis. (**a**) Kaplan-Meyer plots showing demonstrating high concentration of ASAH1 and high concentration of SYNEMIN contributing to a higher probability of tumor recurrence occuring at a later date. (**b**) Kaplan-Meyer plots demonstrating the combination of both ASAH1 and SYNEMIN and their cotnribution of survival probability and time to recurrence.

Logistic Regression with Elastic Net Regularization Performance Metrics

Comparison	Sensitivity	Specificity	Pos Pred Value	Neg Pred Value	Balanced Accuracy
Ctrl vs iGBM	70%	70%	70%	70%	70%
Ctrl vs rGBM	80%	80%	80%	80%	80%
iGBM vs rGBM protein	70%	70%	70%	70%	70%
iGBM vs rGBM protein + clinical	50%	50%	20%	80%	50%
iGBM vs rGBM protein ratios	80%	88.89%	88.89%	80%	84.44%

Supplemental Table I : Logistic Regression with Elastic Net Regularization Performance Metrics

Statistical Comparisons Between Population Groups per Genes

Comparison	Statistical Test[1]	ASAH1[2]	GPNMB[2]	MMP-9[2]	SYNEMIN[2]
Control vs Pre-Surgery vs Post-Surgery Plasma Samples					
Ctrl vs post- surgery	Permutation Test	0.7487500	0.1669500	0.1272000	0.2579500
Ctrl vs pre- surgery	Permutation Test	0.0203000	0.9942000	0.0420000	0.0594500
pre- surgery vs post- surgery	Paired Permutation Test	0.0089720	0.2421880	0.4755860	0.5427860
Control vs Initial GBM vs Recurrent GBM Plasma Samples					
Ctrl vs iGBM	Permutation Test	0.1836000	0.3916500	0.0839000	0.0024000
Ctrl vs rGBM	Permutation Test	0.2547000	0.2447000	0.4589000	0.1386000
iGBM vs rGBM	Paired Permutation Test	0.1992188	0.6123047	0.1171875	0.1054688

[1] Permutation test using t-test statisics for non-paired samples, or paired permutation test using pair t-test statistics for paired samples.

[2] Values in cells represent the non-adjusted p-value from the corresponding statistical test.

Supplemental Table II : Statistical Comparisons Between Population Groups per Genes

Methods

Statistical Test for Comparison of Log2 Distributions

Log2 transformed distributions of protein concentrations were compared pairwise for control, initial GBM, and recurrent GBM samples for each protein, ASAH1, GNMPB, MMP-9 and SYNEMIN. Statistical p-values were computed using a permutation test, that utilizes t-test statistics but performs random permutations of the data for computing a p-value. Note, for paired sample data, initial GBM vs recurrent GBM, a paired permutation test was performed instead. Visualization and statistical analyses were performed in R (v4.1.2) and RStudio (v2021.09.2+382), permutation tests were performed using broman (v0.80).

Dimensionaly Reduction

To visualize the samples plasma proteomics on a different dimension, we applied a sparse Partial Least Squares Discriminant Analysis (sPLS-DA), which maximizes the covariance between the explanatory and response variable(s). This was performed on the log2 transformed protein concentrations, as well as the log2 transformed ratios ($\frac{initial \quad GBM}{recurrent \quad GBM}$ vs $\frac{recurrent \quad GBM}{initial \quad GBM}$) of each protein. Individual trajectories are displayed as vector arrows connecting initial GBM patient samples to their corresponding recurrent GBM patient samples, in the sPLS-DA space using the first two components that contain the most amount of explained variance. An average trajectory direction can be computed by taking the average of the individual trajectory vectors, to visualize an overall trajectory of initial GBM state to recurrent GBM state. To visualize decision boundaries for class separation (control, initial GBM and recurrent GBM samples), we applied a maximum distance approach. [1] Visualization and statistical analyses were performed in R (v4.1.2) and RStudio (v2021.09.2+382), sPLS-DA analysis was performed using mixOmics (v6.18.1).

Prediction Analysis

Prediction tasks were performed using parametric (logistic regression, logistic regression with elastic net regularization), non-parametric (k-nearest neighbours) and ensemble (random forest) supervised learning algorithms. General predictive tasks were performed on: controls vs initial GBM samples, controls vs recurrent GBM samples, initial GBM samples vs recurrent GBM samples. Data used in prediction tasks are log2 transformed protein concentrations, log2 transformed protein concentrations supplemented with clinical data (*Tumor localization*, *Sex*, *Age at surgery (years)*, *Type of resection*), and log2 transformed ratios of $\frac{initial \quad GBM}{recurrent \quad GBM}$ vs $\frac{recurrent \quad GBM}{initial \quad GBM}$ protein measurements. Model training and prediction performances are performed on the entire data set using leave-one-out cross-validation due to the small sample size. Visualization and statistical analyses were performed in R (v4.1.2) and RStudio (v2021.09.2+382), prediction analyses were performed using caret (v6.0-93).

References

1. Rohart, F, B Gautier, A Singh, and K-A Lê Cao. 2017. “MixOmics: An R Package for ’Omics Feature Selection and Multiple Data Integration.” PLoS Computational Biology 13 (11). Cold Spring Harbor Labs Journals.