

Supplementary Material

Nomogram for Early Prediction of Parkinson's Disease Based on microRNA Profiles and Clinical Variables

Supplementary Table 1. Baseline characteristics comparison among Healthy control, Parkinson's disease cases and Prodromal cases

Variables	Healthy controls (n=573)	Cases (n=711)		<i>p</i>
		Early PD (n=656)	Prodromal PD (n=55)	
Age, y	59.4 (48.8,66.4)	63.8 (55.6,70.2)	67.4 (64.7,73.0)	<0.001
MDS-UPDRS total score	5.0 (2.0,9.0)	32.0 (23.0,43.0)	10.0 (6.0,17.0)	<0.001
Gender				<0.001
Man	272 (47.5)	390 (59.5)	44 (80.0)	
Woman	301 (52.5)	266 (40.5)	11 (20.0)	
Education level				<0.001
Less than 12, y	35 (6.1)	55 (8.4)	10 (18.2)	
12-16, y	235 (41.0)	339 (51.7)	20 (36.4)	
Greater than 16, y	283 (49.4)	231 (35.2)	20 (36.4)	
Missing	20 (3.5)	31 (4.7)	5 (9.1)	
Hoehn and Yahr staging				<0.001 [#]
0	561 (97.9)	4 (0.6)	52 (94.5)	
1	10 (1.7)	232 (35.4)	2 (3.6)	
2	2 (0.3)	420 (64.0)	1 (1.8)	

[#]The comparison was completed by Fisher's exact test.

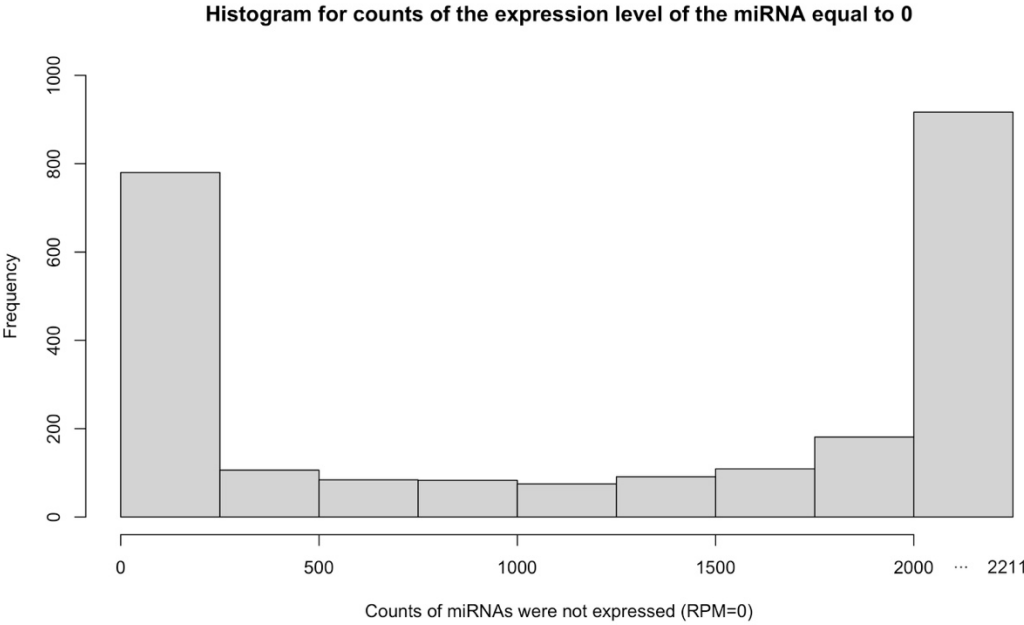
The HY Stage 0 indicates no signs of PD symptoms, whereas the HY Stage 1 and 2 mean that tremor, rigidity, reduced arm swing, and slowness are present on one or both sides of the body.

Supplementary Table 2. The AUC of ROC curves for comparisons of the nomogram, clinical, and miRNA panel

ROC models	AUC (95% CI)	<i>p</i>
Training		
Nomogram	0.70(0.65, 0.74)	Reference
miRNA panel	0.62(0.57, 0.67)	<0.001
Clinical	0.64(0.59, 0.69)	0.002
Validation		
Nomogram	0.74(0.68, 0.80)	Reference
miRNA panel	0.64(0.57, 0.71)	0.002
Clinical	0.68(0.61, 0.75)	0.004
Testing		
Nomogram	0.72(0.68, 0.77)	Reference
miRNA panel	0.64(0.59, 0.69)	<0.001
Clinical	0.69(0.64, 0.73)	0.023

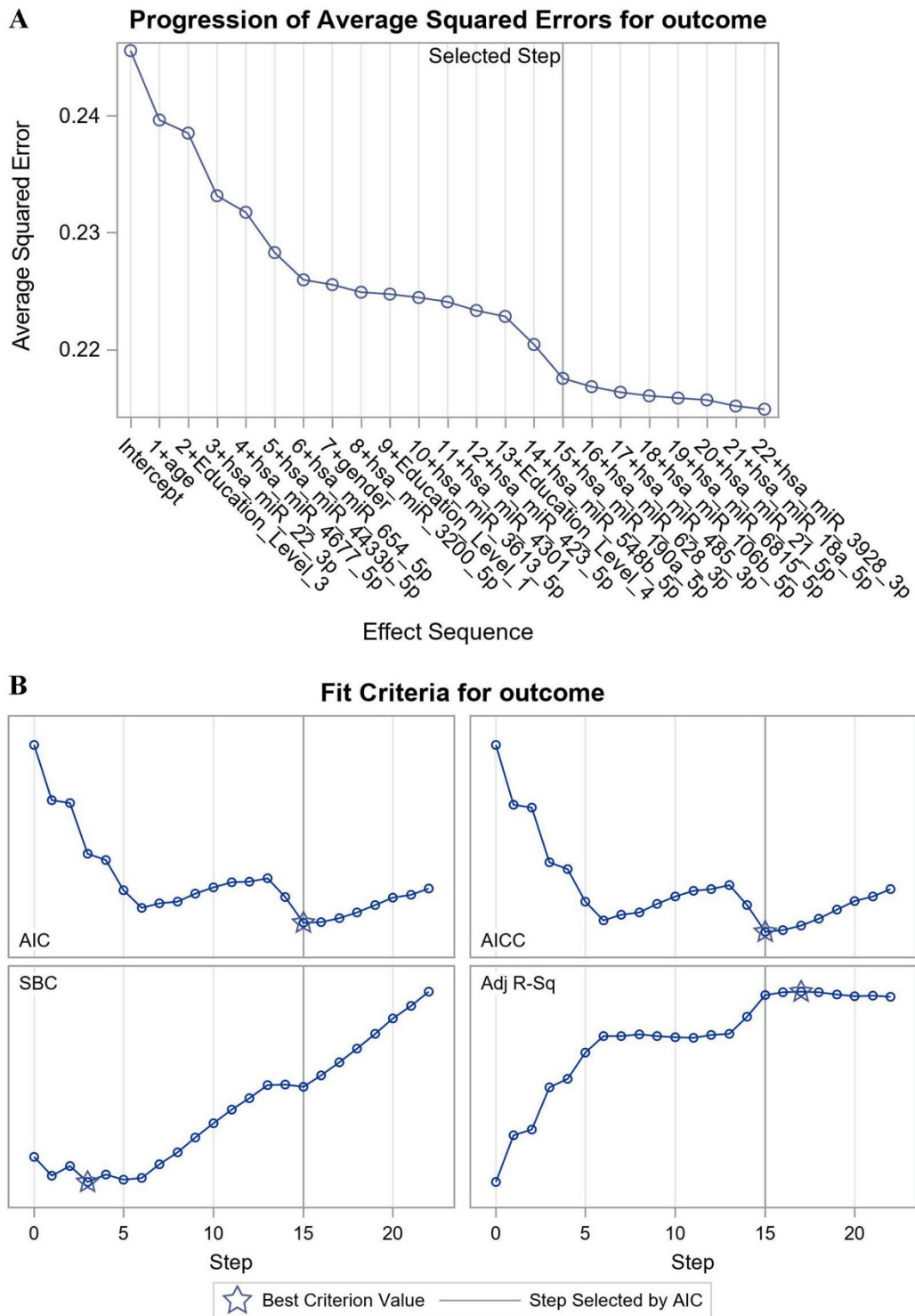
The comparison of AUC between the nomogram and clinical or miRNA panel was conducted by bootstrap methods test. The clinical model consists of age, gender, and education level, while the miRNA panel includes 10 miRNAs.

Supplementary Figure 1. Histogram for counts of the microRNA not expressed.



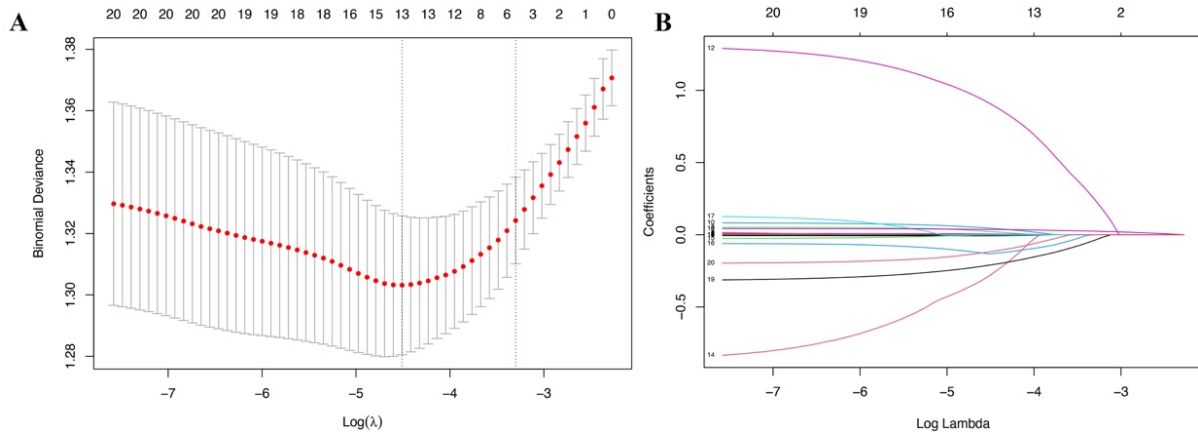
RPM, Reads per million mapped to miRNA normalized read counts

Supplementary Figure 2. Elastic net model for feature selection by 8-fold cross-validation (A) and model fit statistics (B).

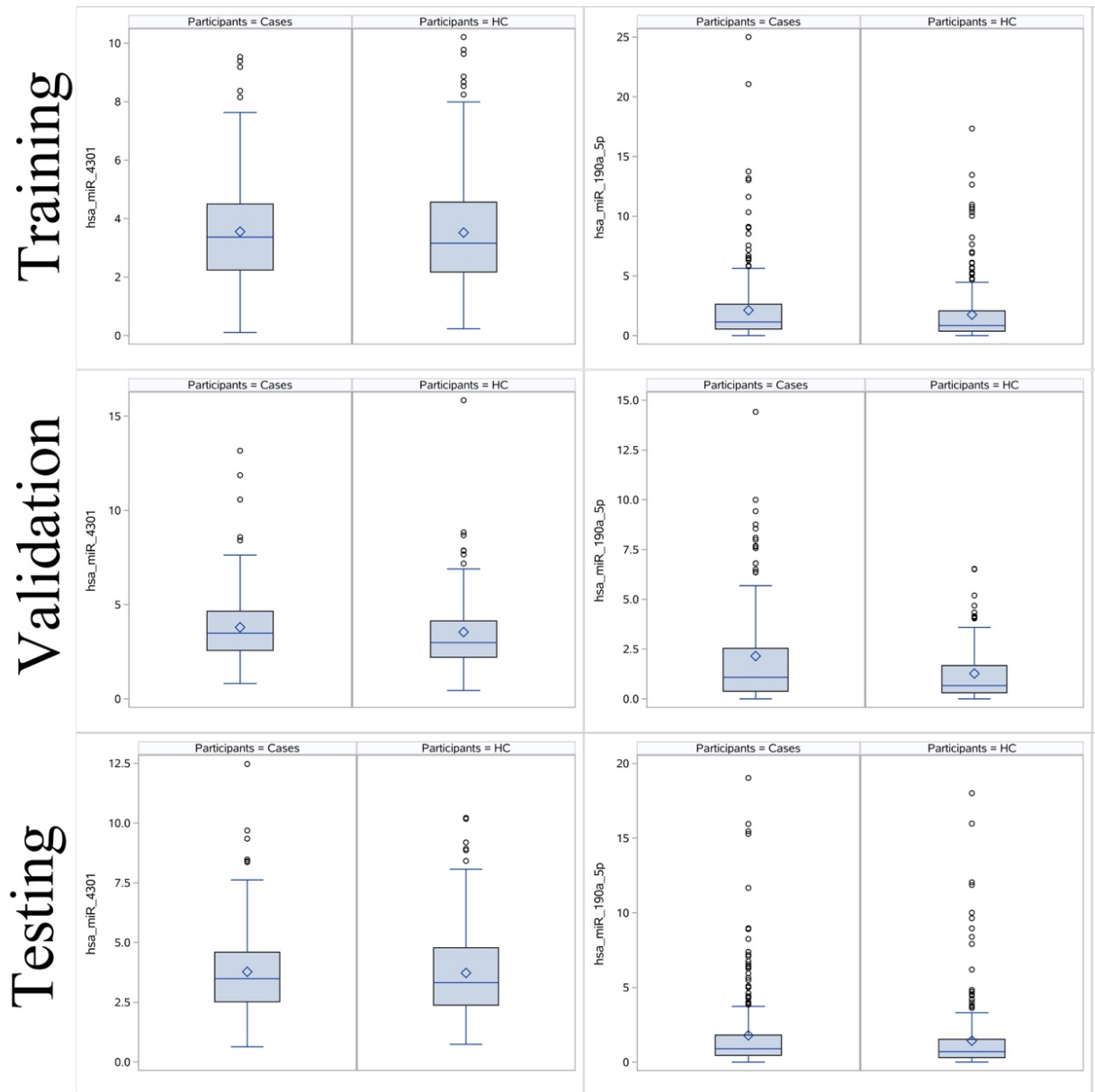


AIC, Akaike information criterion; AICC, the sample-size adjusted AIC; SBC, Schwarz information criterion; adj R-Sq, the adjusted R-squared (R^2)

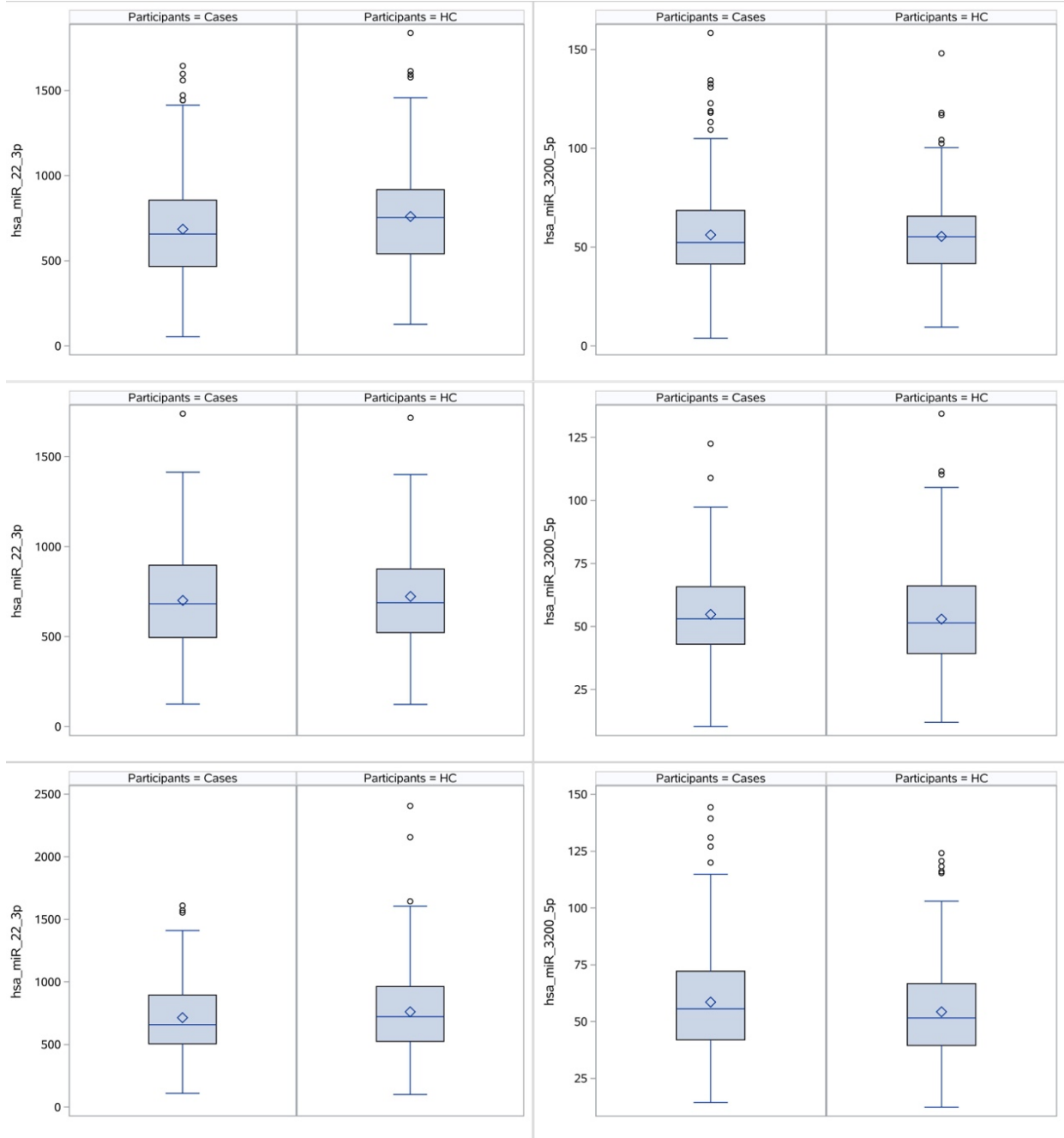
Supplementary Figure 3. LASSO regression model for feature selection by 8-fold cross-validation. Panel A shows the binomial deviance of the LASSO model with different log-lambda ($\log-\lambda$), the left dashed line is the $\log-\lambda$ value of the minimum mean value of binomial deviance in the model, and the right dashed line is the $\log-\lambda$ value of the simplest model obtained within the variance range of the minimum mean value of binomial deviance; Panel B represents the screening path of the LASSO regression model. The best LASSO model with 13 risk factors was selected by the $\log-\lambda$ at which the minimal binomial deviance was achieved. LASSO, least absolute shrinkage and selection operator.



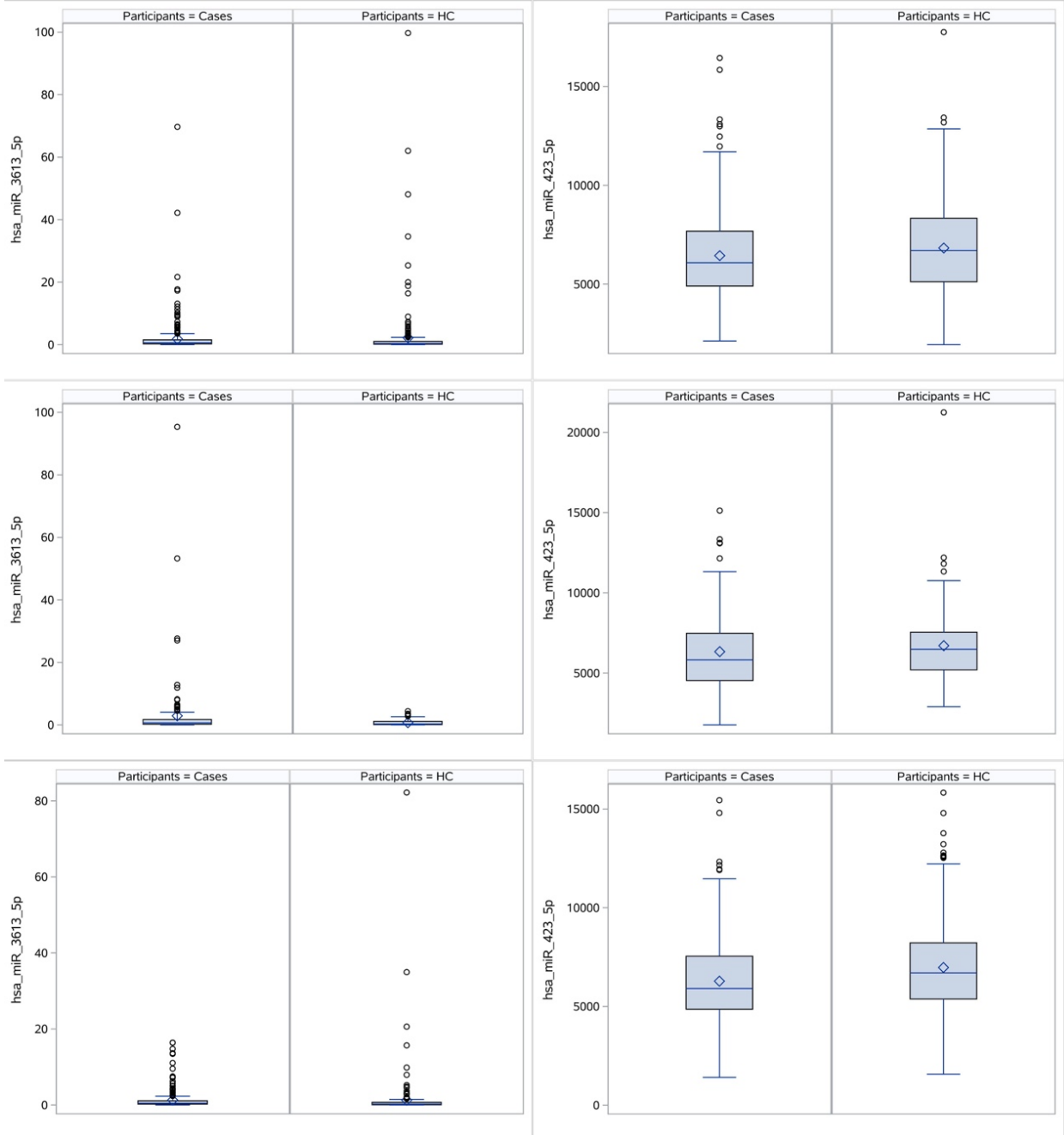
Supplementary Figure 4. Box plots for the expression of the 10 identified miRNA based on control and cases groups in training, validation, and testing datasets separately.



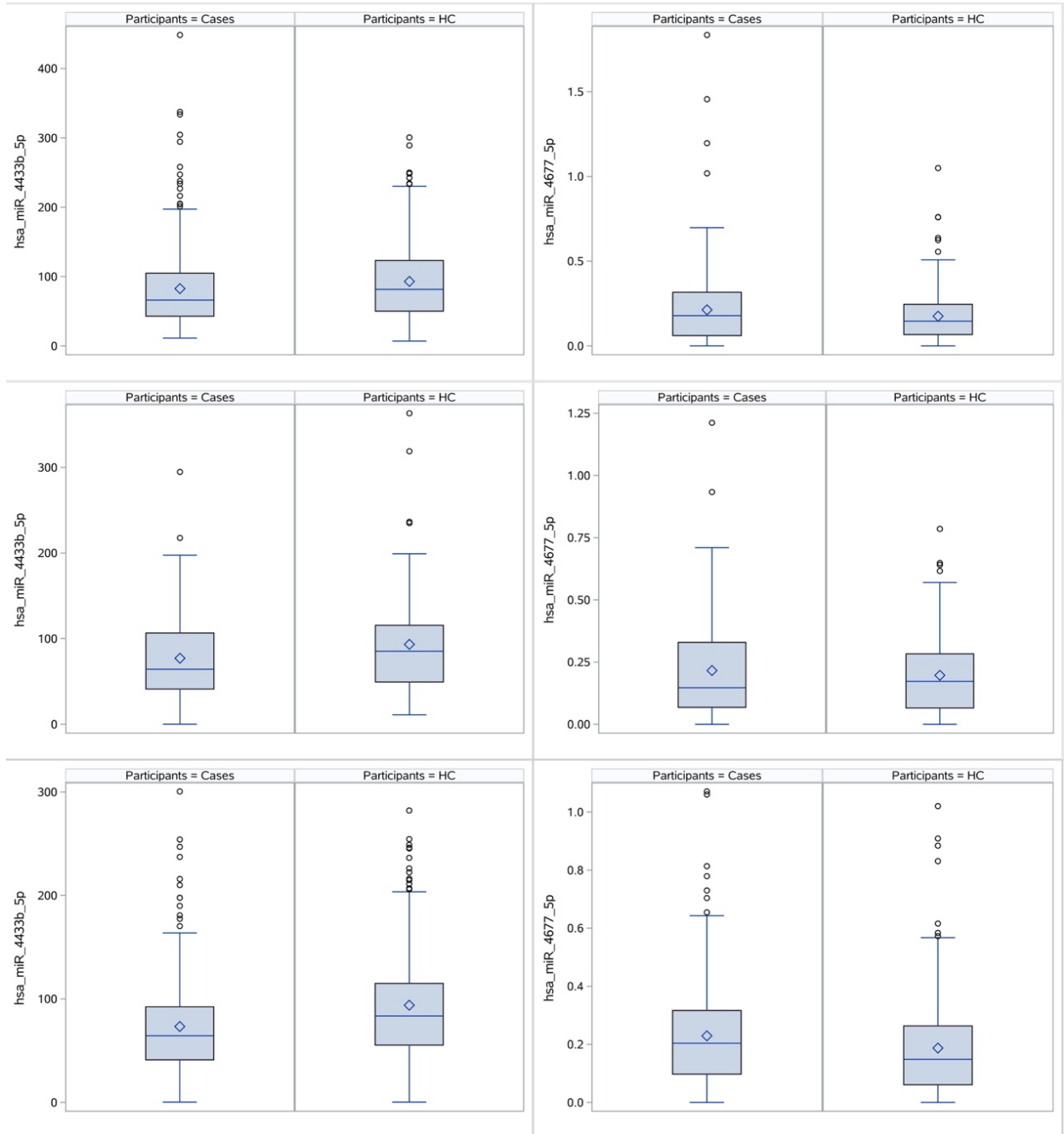
Continued



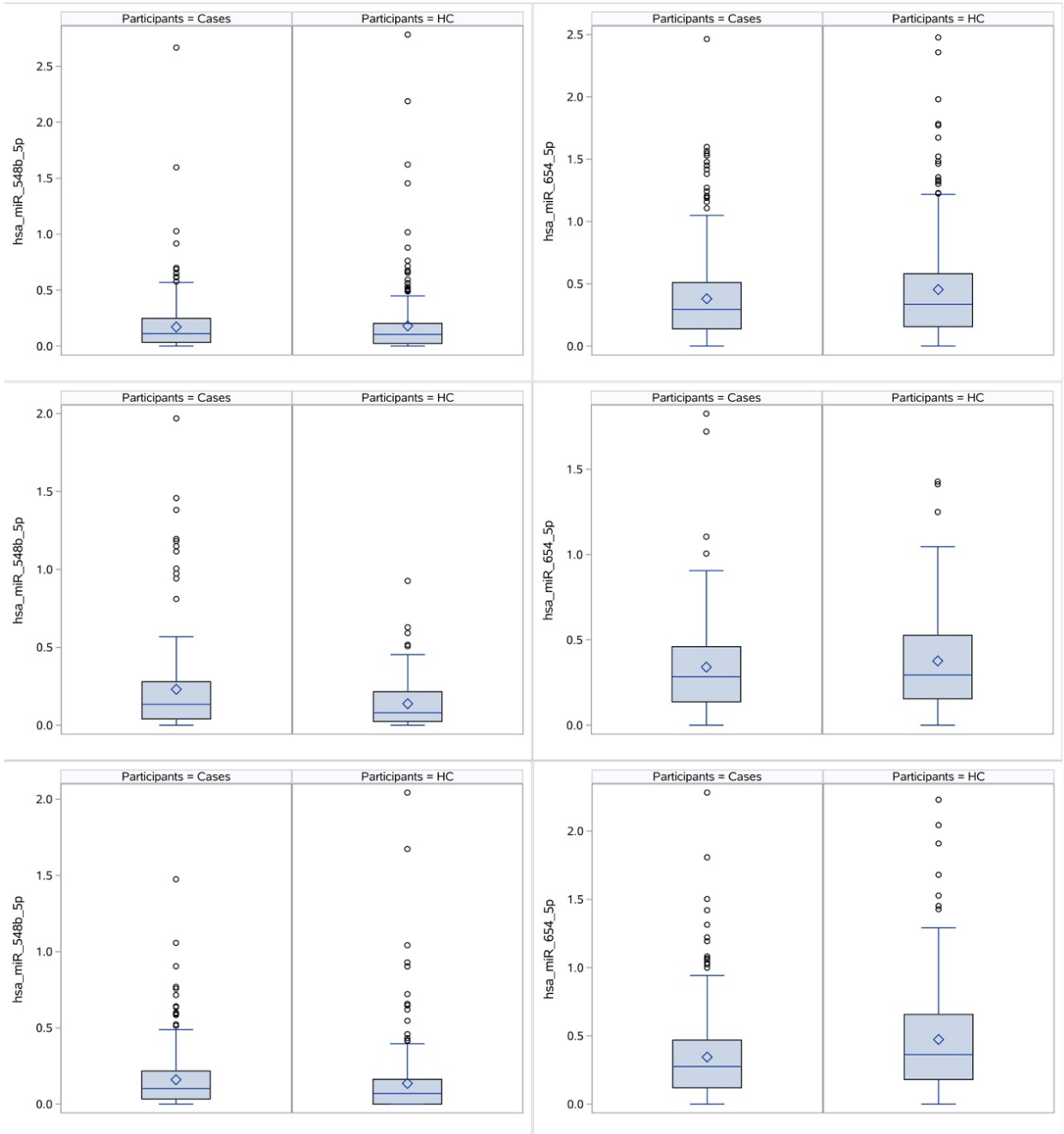
Continued



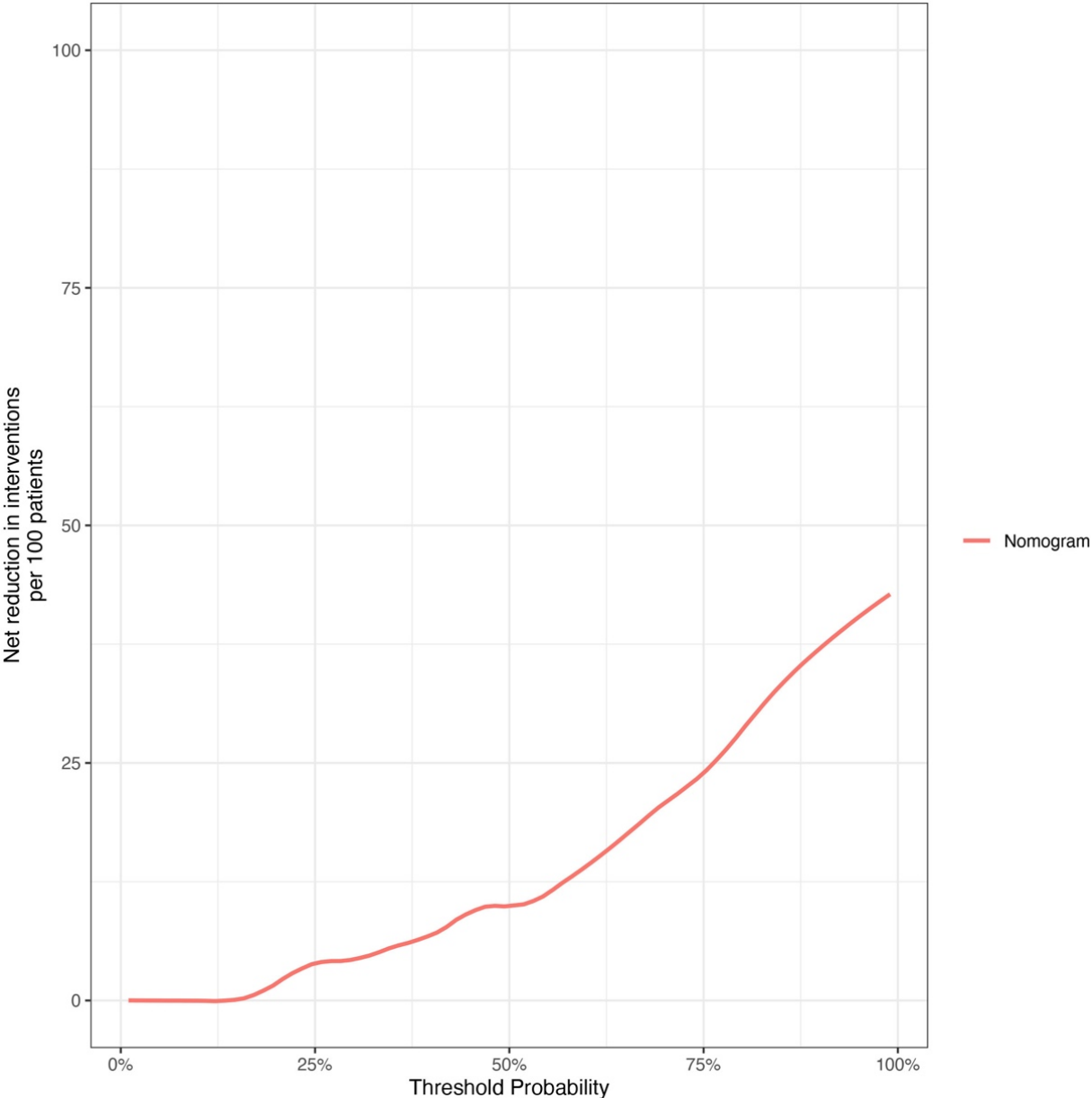
Continued



Continued

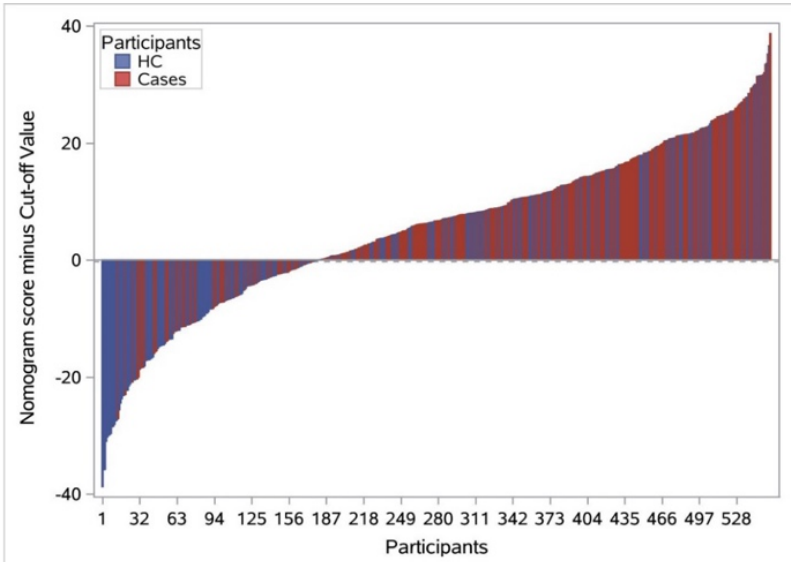


Supplementary Figure 5. DCA curves for a net reduction in nomogram based on training datasets. Shown are the threshold probability on the x-axis and the net reduction on the y-axis.

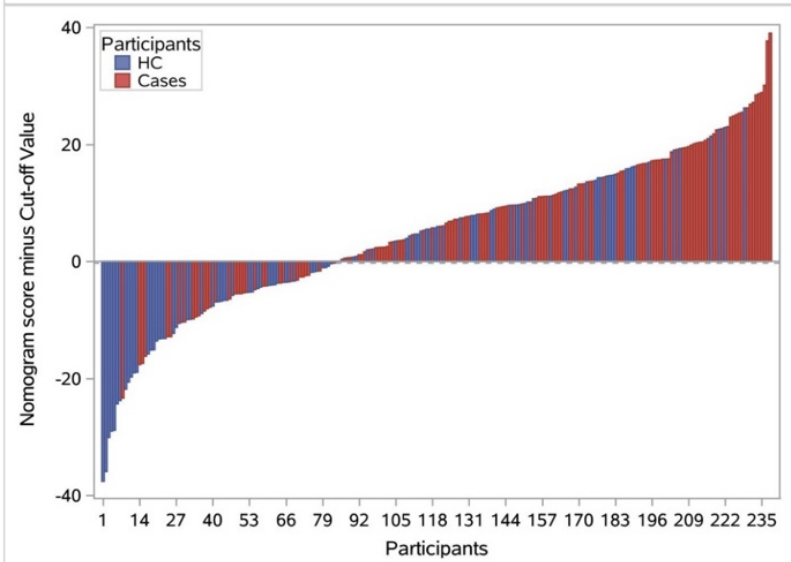


Supplementary Figure 6. Waterfall plots for comparison risk scores between healthy control and cases in the training, validation, and testing sets. Shown are the risk score (nomogram scores minus cut-off value) on the y-axis and individual distribution on the x-axis. The cut-off point (42.1) was detected by the ROC analysis. Waterfall plots with a score above 0 represent the high-risk group, while waterfall plots with a score below 0 represent the low-risk group.

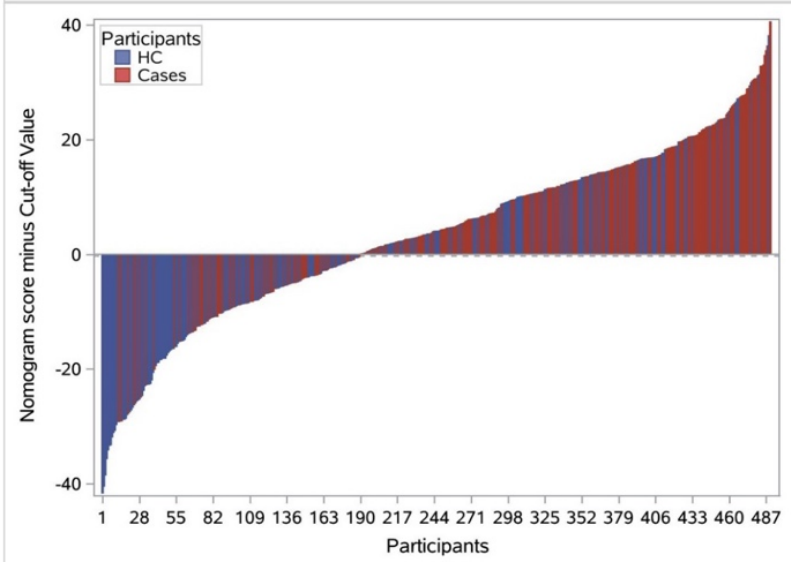
Training



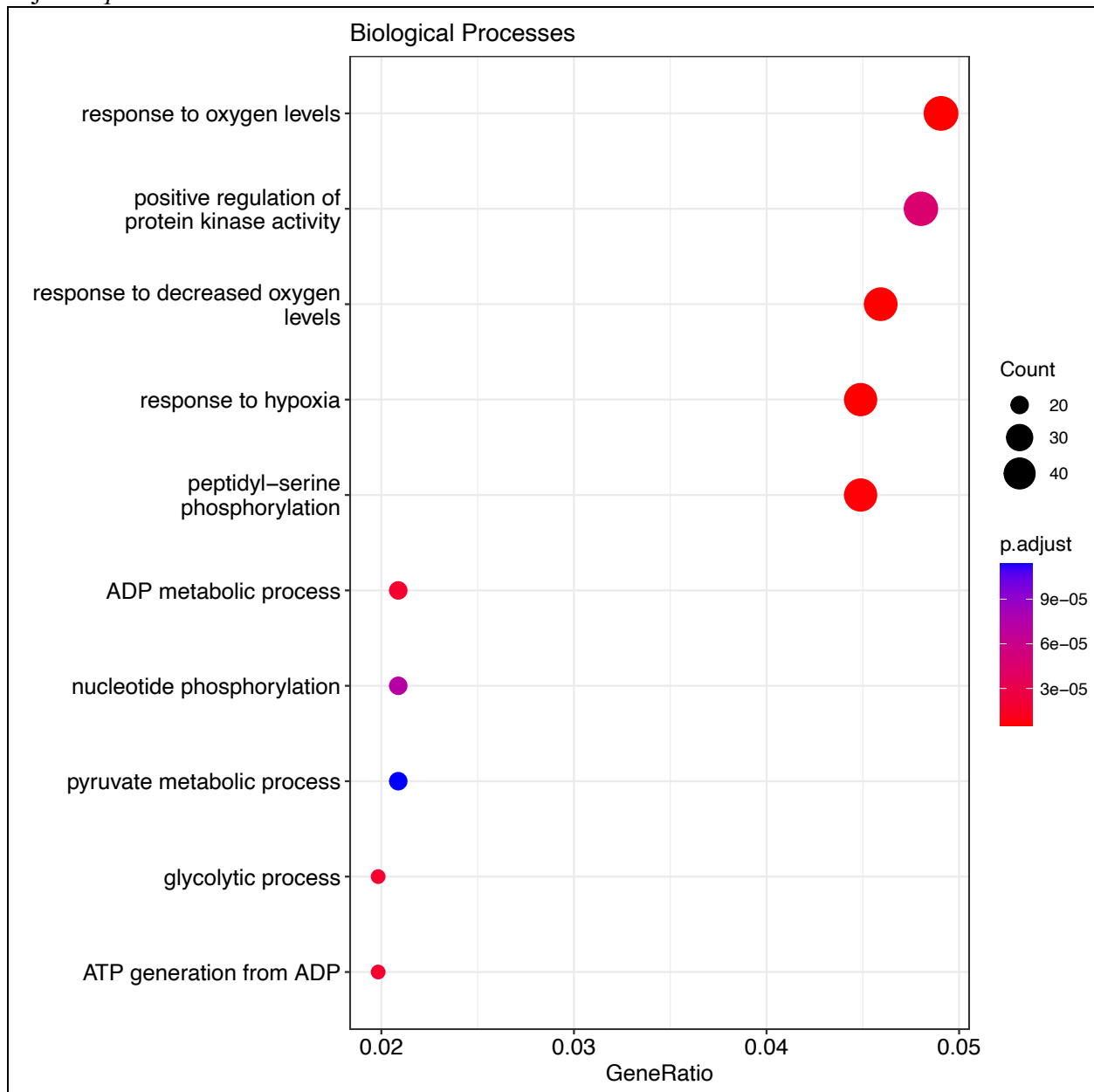
Validation

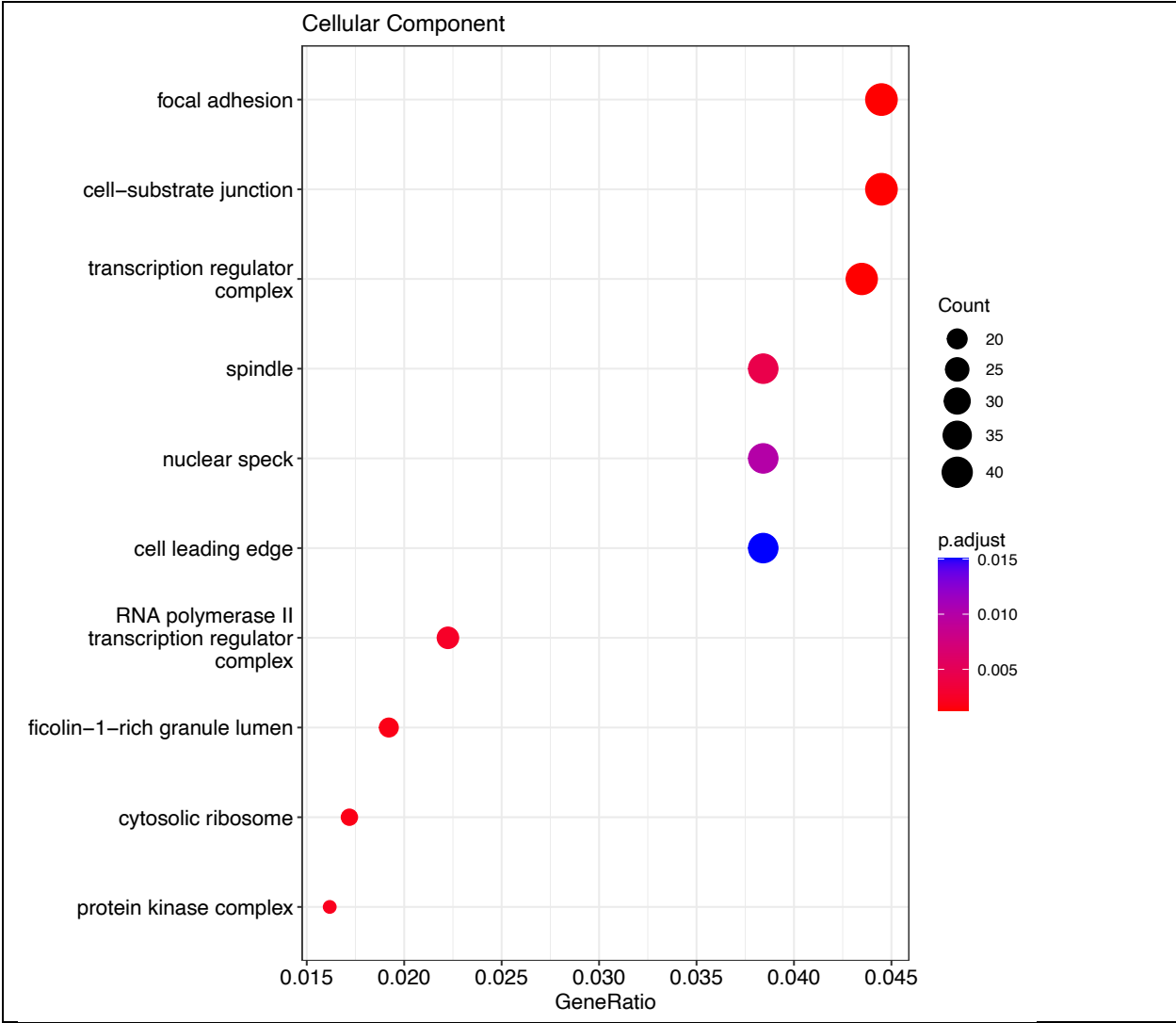


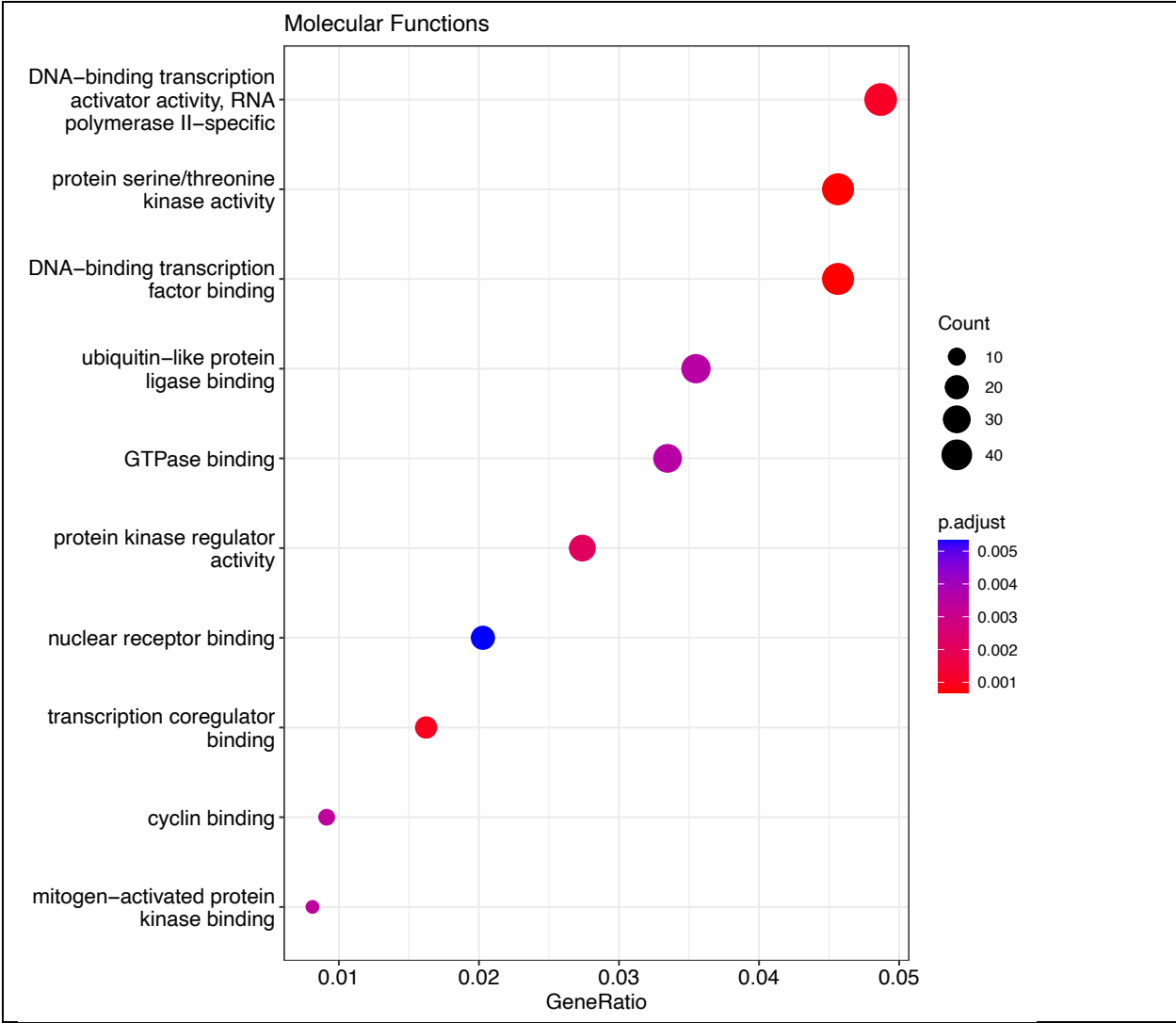
Testing



Supplementary Figure 7. GO analysis results of the predicted 1,084 target genes enrichment dot plot from the 10-miRNA signature. Shown are the gene ratio on the x-axis and the simplified enriched term list on the y-axis. Each dot represents a gene set of specific GO terms. The dot size indicates the counts of genes implicated with specific GO terms. The dot color represents the adjusted *p* values.







Supplementary Figure 8. KEGG analysis results of the predicted 1,084 target genes enrichment dot plot from the 10-miRNA signature. Shown are the gene ratio on the x-axis and the enriched term list on the y-axis. Each dot represents a gene set of specific KEGG terms. The dot size indicates the counts of genes implicated with specific KEGG terms. The dot color represents the adjusted *p* values.

