1. Results for smaller sample sizes

We also evaluate the classification performance of the proposed method under smaller sample sizes. Results are shown in Table 2. Sample sizes are $n = m = 20$ for training data and $n = m = 10$ for the testing data. Differences of the means of the two groups are $\delta = 1$, 1.5 and 2 for the differential features. Fractions of differential features, π, are 0.01 and 0.05. Again, we used $\lambda = 1$ and 5-fold cross validation for determining the threshold parameter τ. From Table 2, the number of selected features and misclassification rates decrease with δ for a given fraction of differential features. For example, the number decreases from 4.695 to 2.843 when the group difference increases from 1 to 2 for $\pi = 0.01$. Correspondingly, the misclassification rate decreases from 0.045 to 0.011 for training data and from 0.301 to 0.073 for testing data.

Table 1: Performance of the proposed PTIFS method when sample sizes are small ($n = m = 20$ for training data, $n = m = 10$ for testing data).

π	δ	Number of selected features	Misclassification rate Training	Testing
	1	4.695(1.570)*	0.045(0.038)	0.301(0.123)
0.01	1.5	3.652(1.184)	0.022(0.027)	0.130(0.088)
	2	2.843(0.975)	0.011(0.020)	0.073(0.071)
	1	3.856(1.172)	0.021(0.029)	0.186(0.097)
0.05	1.5	2.856(0.825)	0.011(0.022)	0.125(0.085)
	2	1.869(0.968)	0.022(0.025)	0.118(0.089)

*Standard errors are in parentheses.

2. Results for $\pi = 0.01$ when features are correlated (simulation settings are the same as in Table 2 in the manuscript except that $\pi = 0.01$).

Results for $\pi = 0.01$ for correlated features are presented in Table 2. It can be seen that the TGDR methods with different maximal numbers of iteration performs similar in prediction accuracy. The PTIFS method has slight larger misclassification rate but selects fewer features.

In general, for correlated features, when proportion of differentially expressed features π is larger, PTIFS method is better than TGDR in both the number of selected features and in misclassification rates. But when π is smaller, TGDR performs slightly better in misclassification rates at the cost of selecting more features.

Table 2: Performance of the proposed PTIFS and TGDR methos when features are correlated (maximal correlation coefficient is 0.5).

δ	Method	Number of selected features	Misclassification rate	
			Training	Testing
1.5	PTIFS	5.337 (2.088)	0.138 (0.024)	0.291 (0.077)
	TGDR100	6.629 (4.456)	0.173 (0.050)	0.247 (0.069)
	TGDR200	9.930 (8.824)	0.152 (0.069)	0.248 (0.069)
	TGDR500	19.932 (20.819)	0.121 (0.093)	0.253 (0.071)
2.0	PTIFS	3.004 (1.244)	0.121 (0.023)	0.196 (0.068)
	TGDR100	6.908 (4.126)	0.108 (0.040)	0.169 (0.059)
	TGDR200	12.670(9.202)	0.078 (0.057)	0.168 (0.062)
	TGDR500	32.089(19.265)	0.038 (0.059)	0.170 (0.060)

*Standard errors are in parentheses.