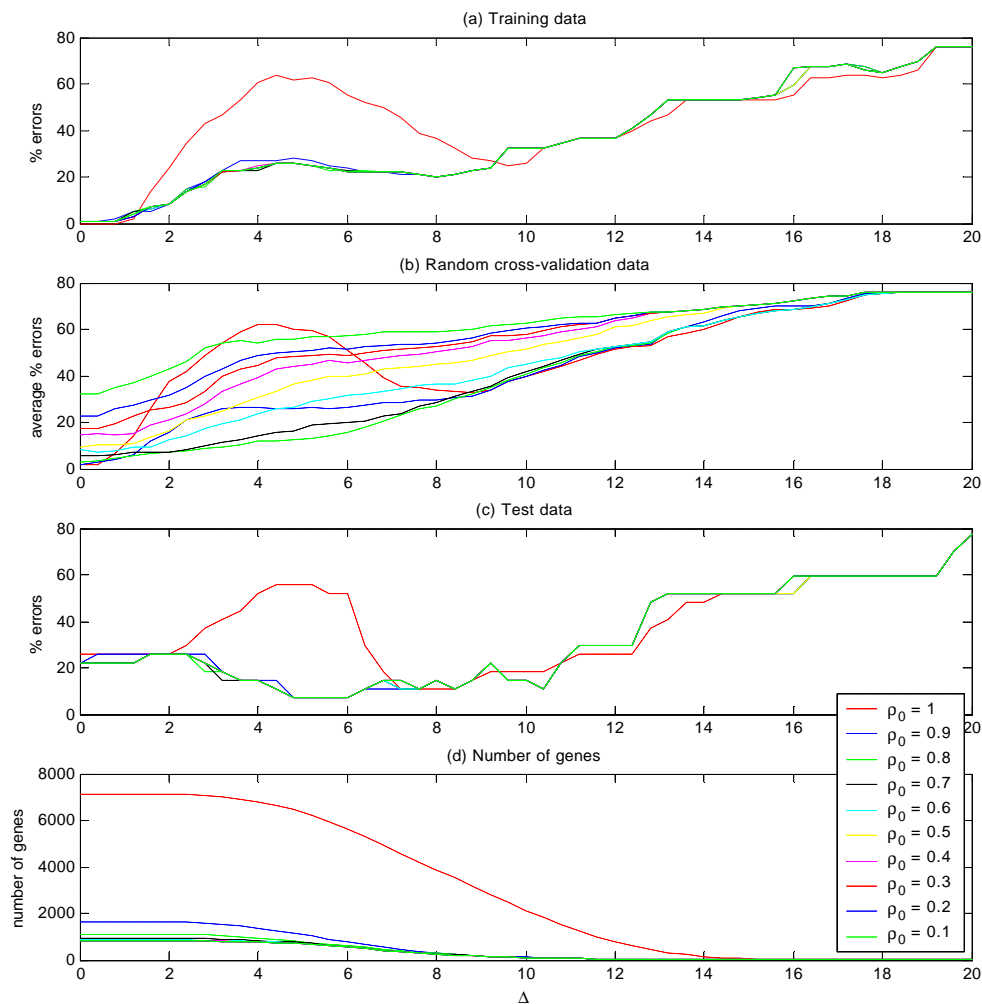


# Supplementary results – multi-class classification of microarray data with repeated measurements: application to cancer

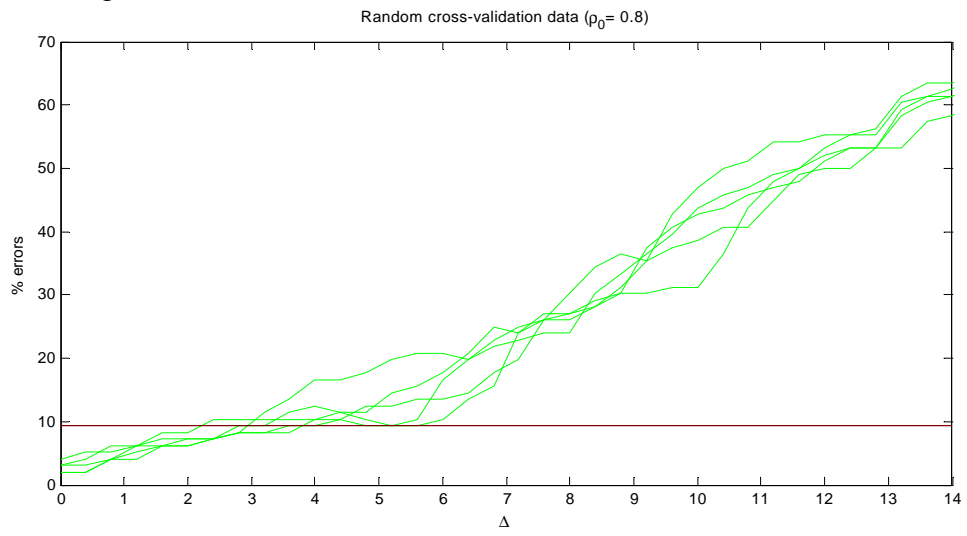
Ka Yee Yeung, Roger E. Bumgarner

**Figure S1 - Prediction accuracy results on the multiple tumor data using the EWUSC algorithm over the range of  $D$  from 0 to 20.**

The percentage of classification errors for  $\rho_0 < 0.6$  are shown in (a) – (d). This is information supplementary to Figure 1 in the main paper. The percentage of errors for each of the 5 random runs of 4-fold cross validation at  $\rho_0 = 0.8$  is plotted against  $\Delta = 0$  to 14 in (e).

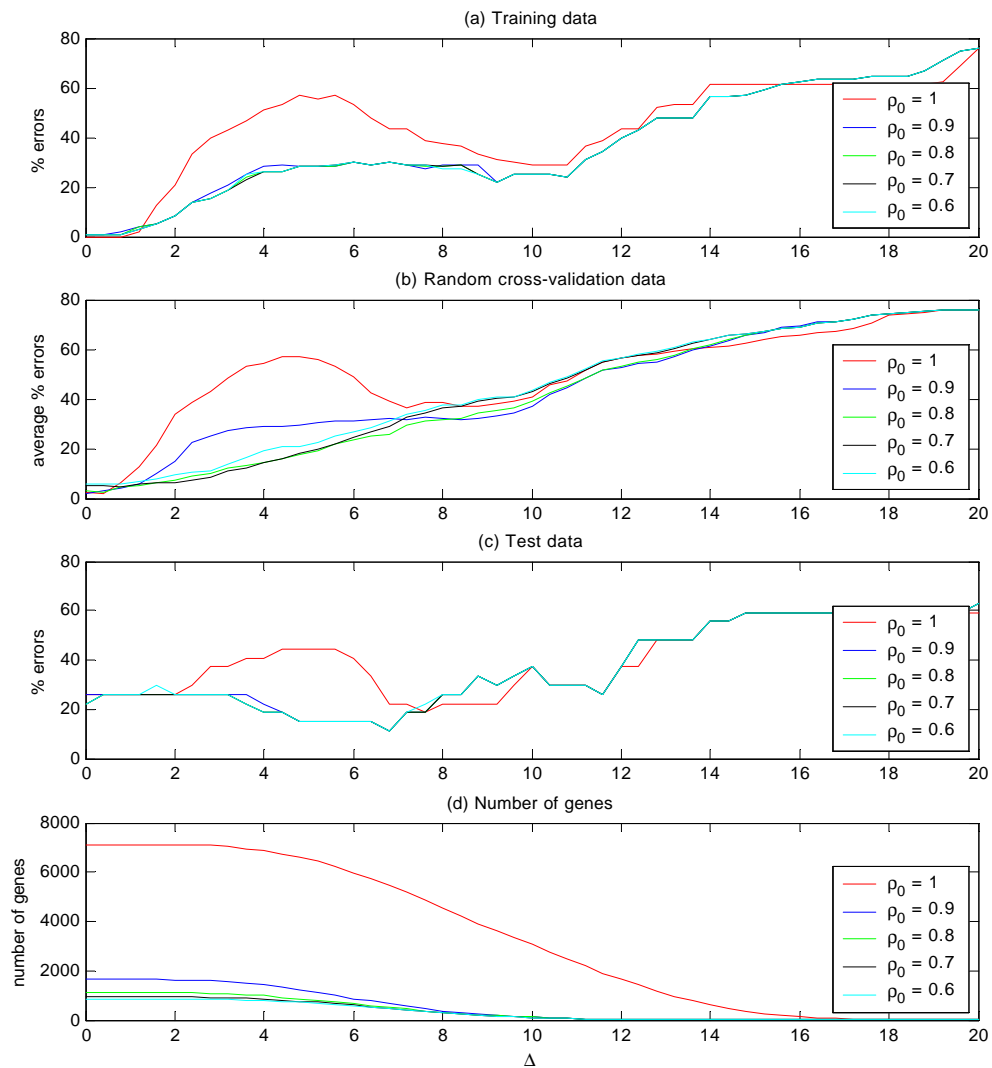


(e) Zoomed-in figure showing each of the 5 random runs of 4-fold cross validation at  $\rho_0 = 0.8$  is plotted against  $\Delta = 0$  to 14.

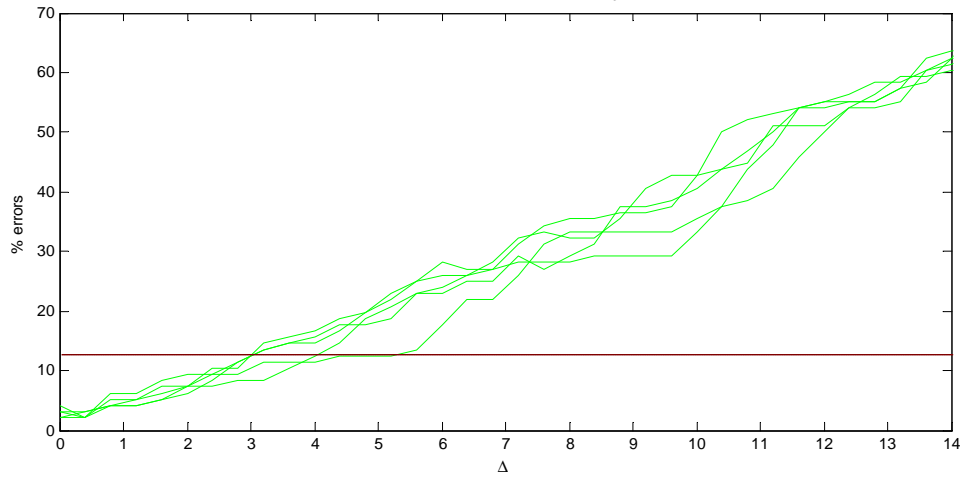


**Figure S2 - Prediction accuracy results on the multiple tumor data using the USC algorithm over the range of  $D$  from 0 to 20.**

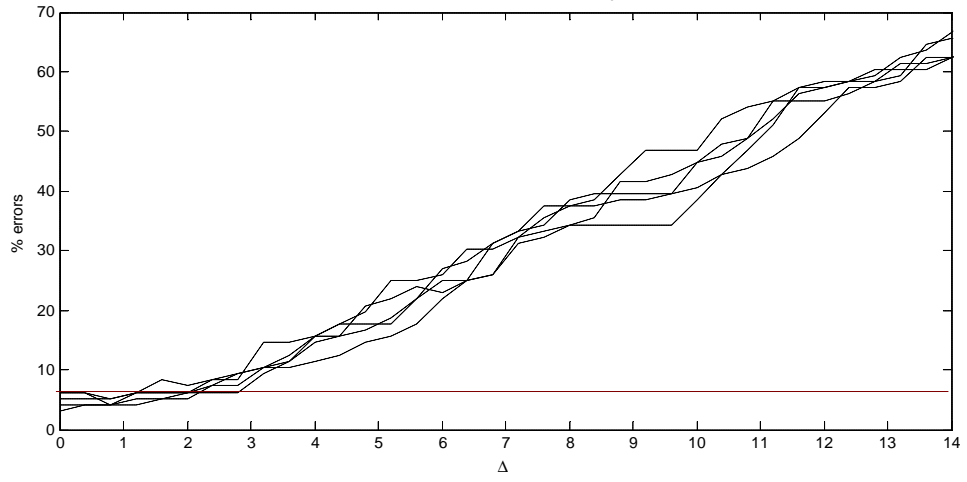
The percentage of classification errors is plotted against  $\Delta$  on (a) the full training set (96 samples) and (c) the test set (27 samples). In (b), the average percentage of errors is plotted against  $\Delta$  on the 4-fold cross validation data over five different random 4-fold splits of the training set. In (d), the number of relevant genes is plotted against  $\Delta$ . Different colors are used to specify different correlation thresholds ( $\rho_0 = 0.6, 0.7, 0.8, 0.9$  or 1). The number of errors for each of the five random 4-fold cross validation data sets is plotted over  $\Delta = 0$  to 14 at  $\rho_0 = 0.8$  in (e) and at  $\rho_0 = 0.7$  in (f).



(e) Random cross-validation data ( $\rho_0 = 0.8$ )

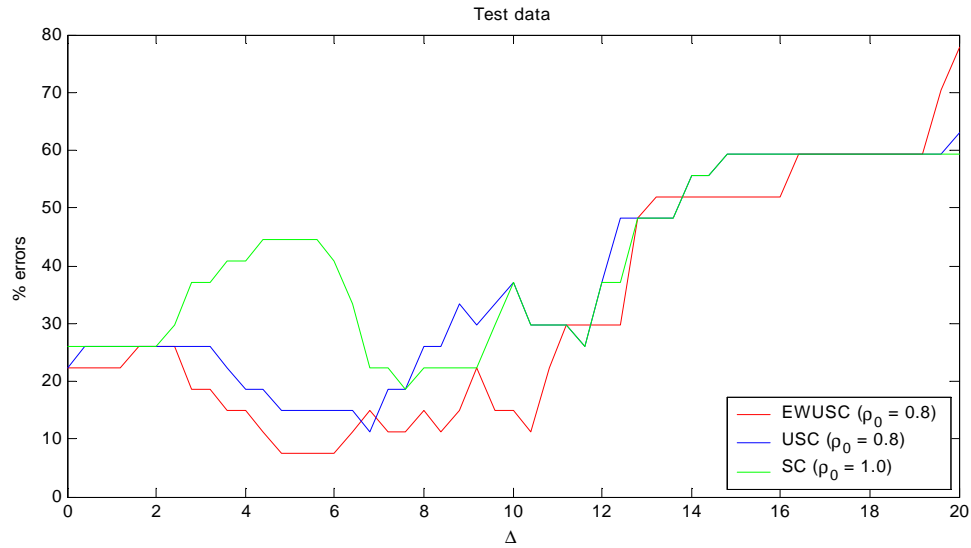


(f) Random cross-validation data ( $\rho_0 = 0.7$ )

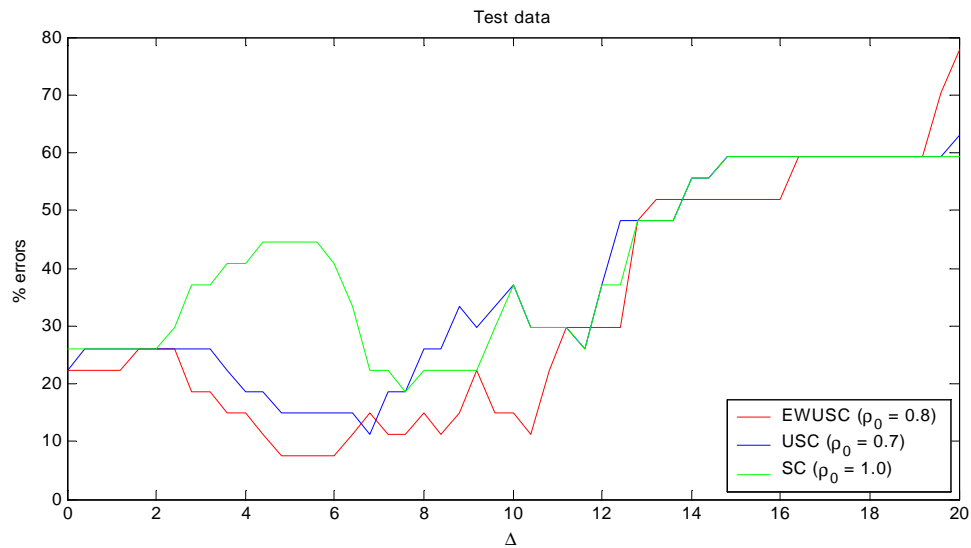


**Figure S3 – Comparing the prediction accuracy results of EWUSC and USC on the multiple tumor data.**

(a) The percentage of classification errors is plotted against  $\Delta$  on the test set using the EWUSC algorithm at  $\rho_0 = 0.8$  and the USC algorithm at  $\rho_0 = 0.8$  and  $\rho_0 = 1.0$  (which is equivalent to the original shrunken centroid algorithm).

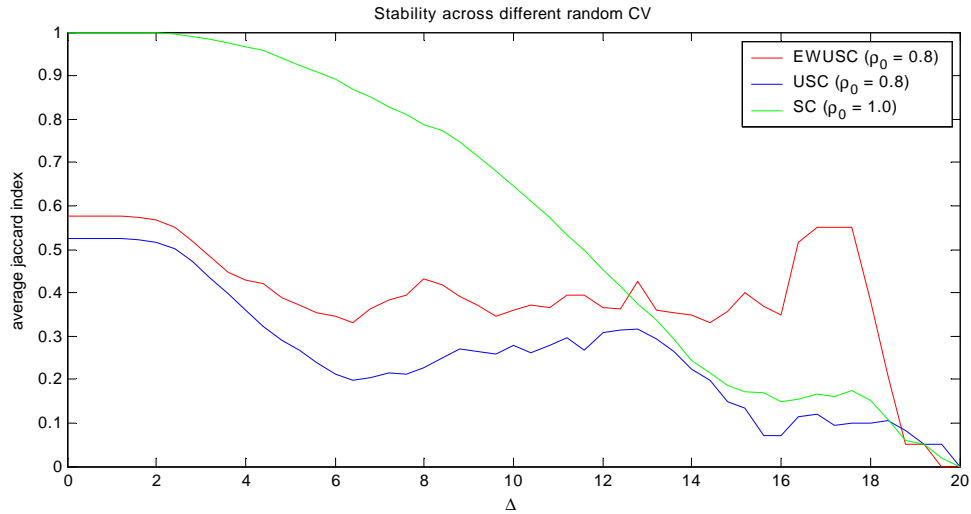


(b) The percentage of classification errors is plotted against  $\Delta$  on the test set using the EWUSC algorithm at  $\rho_0 = 0.8$  and the USC algorithm at  $\rho_0 = 0.7$  and  $\rho_0 = 1.0$ .

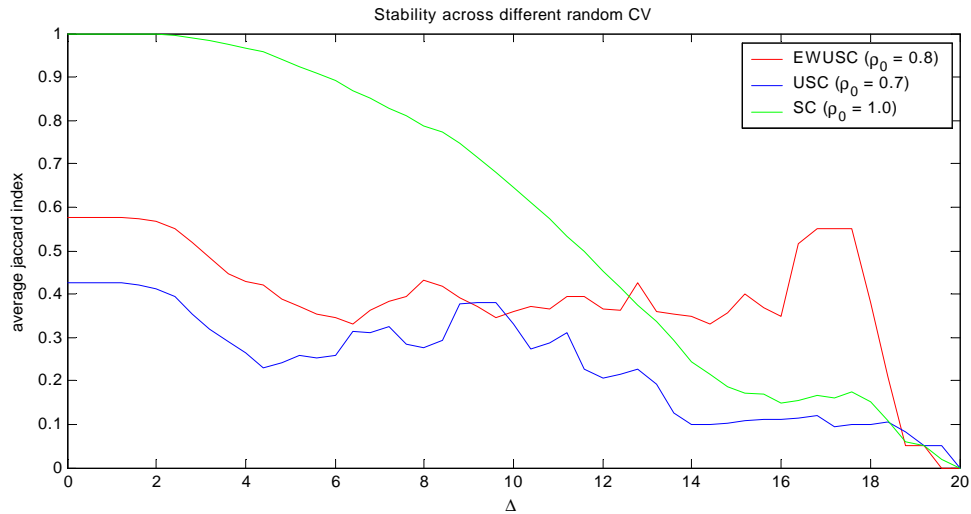


**Figure S4 – Comparing the feature stability of EWUSC vs. USC on the multiple tumor data.**

(a) The average Jaccard index is plotted against  $\Delta$  over the five different random runs of 4-fold cross validation using the EWUSC algorithm at  $\rho_0 = 0.8$  and the USC algorithm  $\rho_0 = 0.8$ . The EWUSC algorithm produces more stable features over most values of  $\Delta$ .

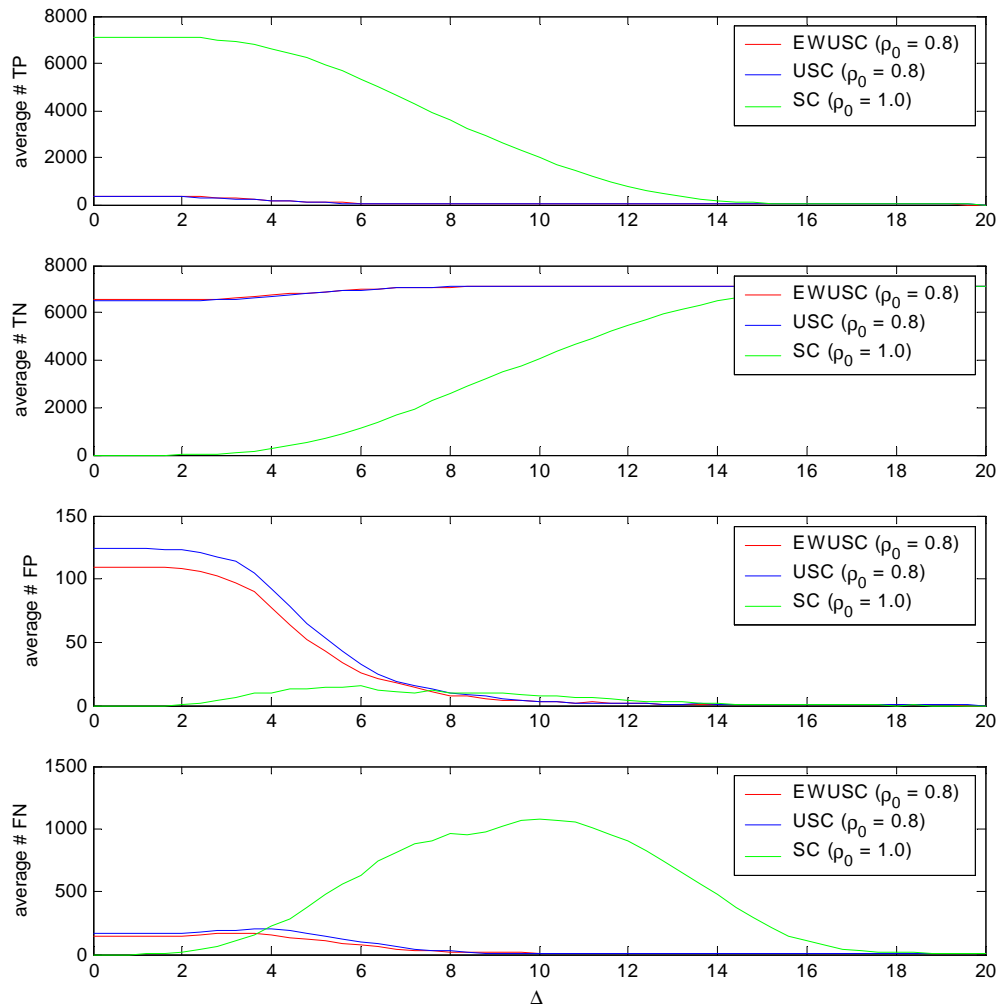


(b) The average Jaccard index is plotted against  $\Delta$  over the five different random runs of 4-fold cross validation using the EWUSC algorithm at  $\rho_0 = 0.8$  and the USC algorithm at  $\rho_0 = 0.7$ . The EWUSC algorithm produces more stable features over most values of  $\Delta$ .



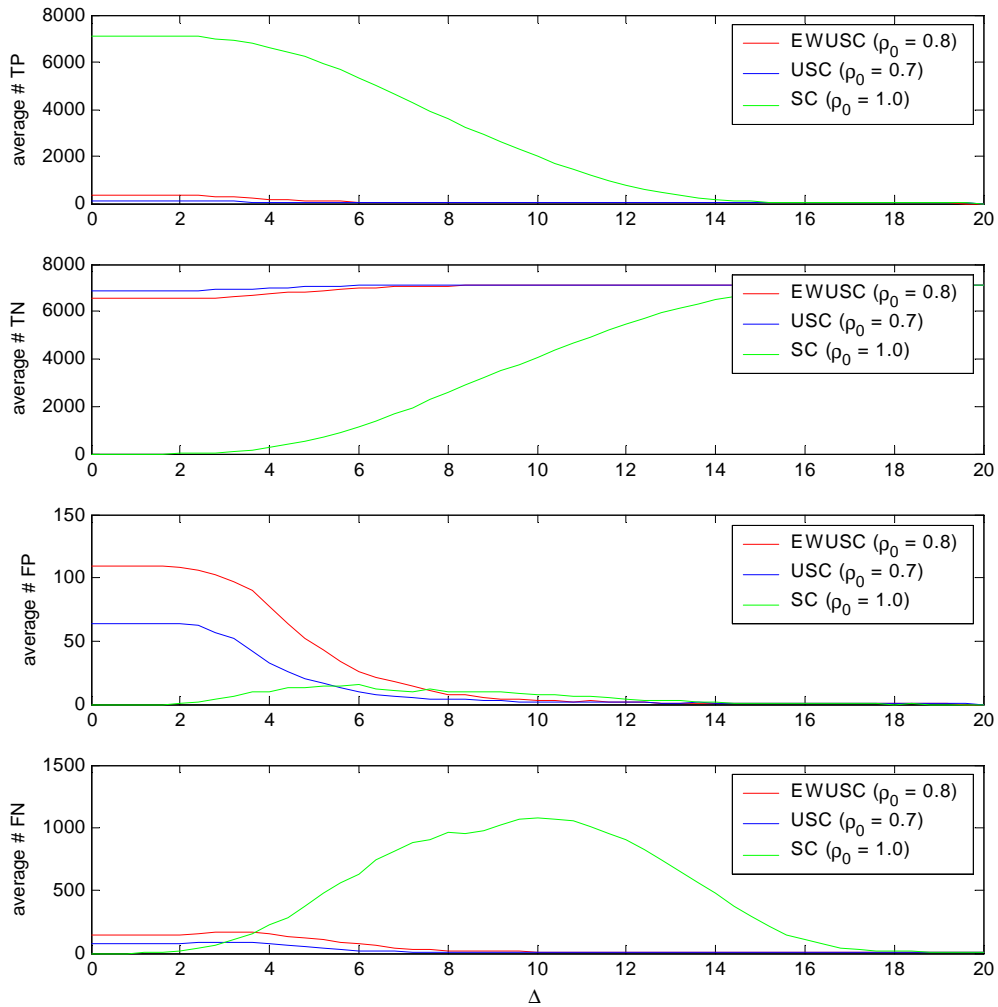
**Figure S5 - Comparison of feature stability results on the multiple tumor data using the EWUSC and USC algorithms.**

The average numbers of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are plotted against  $\Delta$  for both EWUSC and USC at  $\rho_0 = 0.8$ .



**Figure S6 - Comparison of feature stability results on the multiple tumor data using the EWUSC and USC algorithms.**

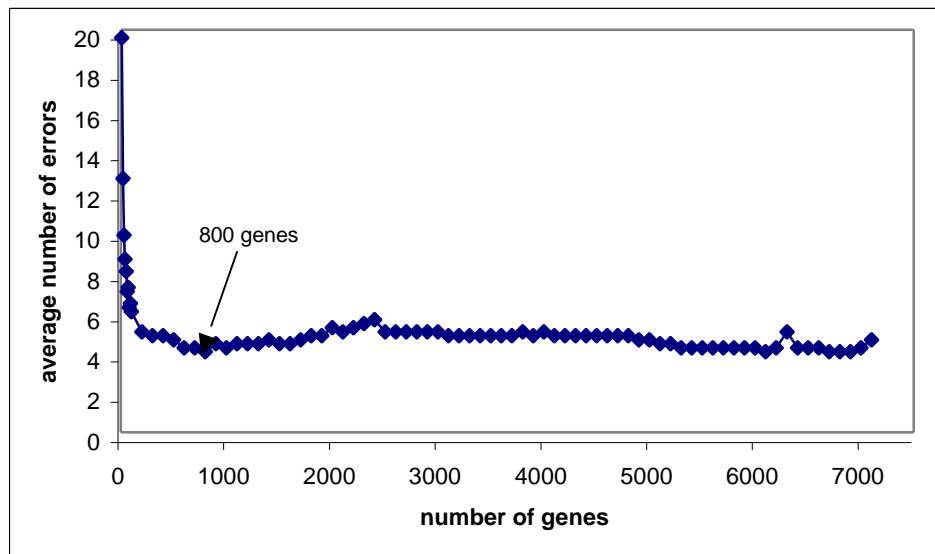
The average numbers of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are plotted against  $\Delta$  for both EWUSC at  $\rho_0 = 0.8$  and USC at  $\rho_0 = 0.7$ .



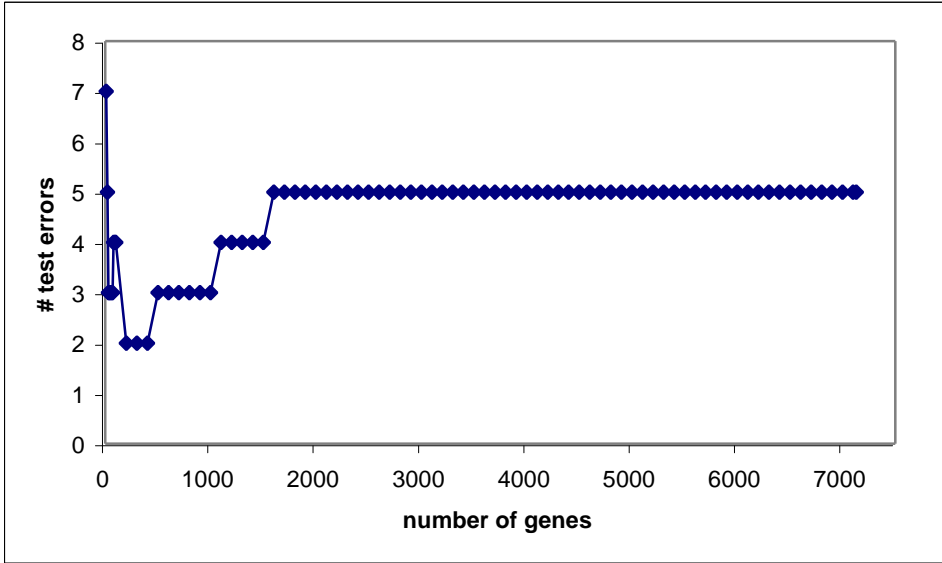
**Figure S7 – Classification accuracy results on the multiple tumor data using SVM.**

Figure S7(a) shows the average number of classification errors on the cross validation data plotted against the number of relevant genes chosen in each of the 11 binary SVM classifiers. The minimum average number of errors is achieved at 800 genes for each SVM, which are equivalent to a total of 5050 genes over all 11 SVM classifiers because most genes are chosen in only one classifier. Figure S7(b) shows the number of classification errors on the test set plotted against the number of relevant genes chosen in each of the 11 binary SVM classifiers. With 800 genes per classifier, SVM produces 3 errors on the test set. The minimum number of errors on the test set attained by SVM is 2, when 200 relevant genes are chosen for each binary classifier. However, these 200 genes per classifier are equivalent to a total of 1699 genes over all 11 SVM classifiers.

(a) On the cross validation data: the *average* number of classification errors (over five 4-fold random splits of the training set) on the cross validation data is plotted against the number of genes chosen in *each* binary classifier.

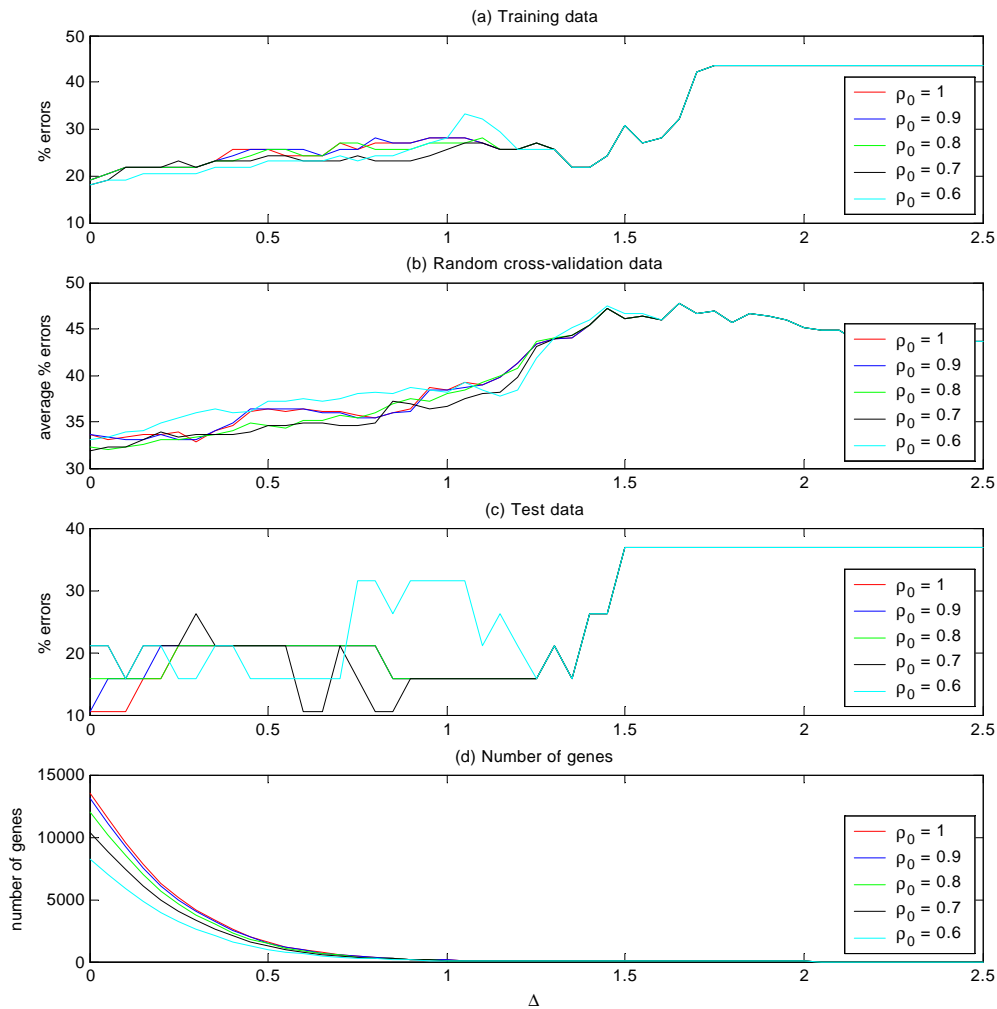


(b) On the test set: the number of classification errors on the test set is plotted against the number of genes chosen in *each* binary classifier.

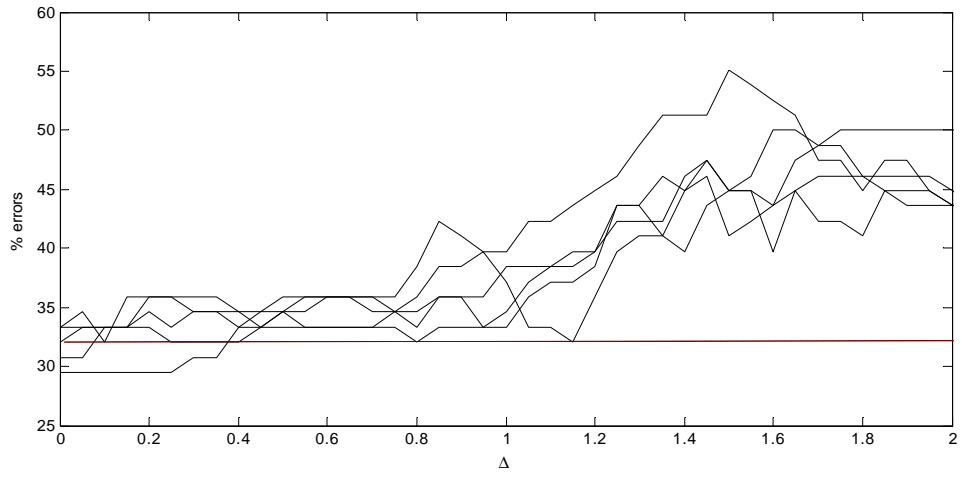


**Figure S8 - Prediction accuracy results on the breast cancer data using the EWUSC algorithm over the range of  $\Delta$  from 0 to 2.5.**

The percentage of classification errors is plotted against  $\Delta$  on (a) the full training set (78 samples) and (c) the test set (19 samples). In (b), the average percentage of errors is plotted against  $\Delta$  on the 10-fold cross validation data over five different random 10-fold splits of the training set. In (d), the number of relevant genes is plotted against  $\Delta$ . Different colors are used to specify different correlation thresholds ( $\rho_0 = 0.6, 0.7, 0.8, 0.9$  or 1). The number of errors for each of the 5 random 10-fold cross validation data sets at  $\rho_0 = 0.7$  is plotted against  $\Delta = 0$  to 2 in (e).

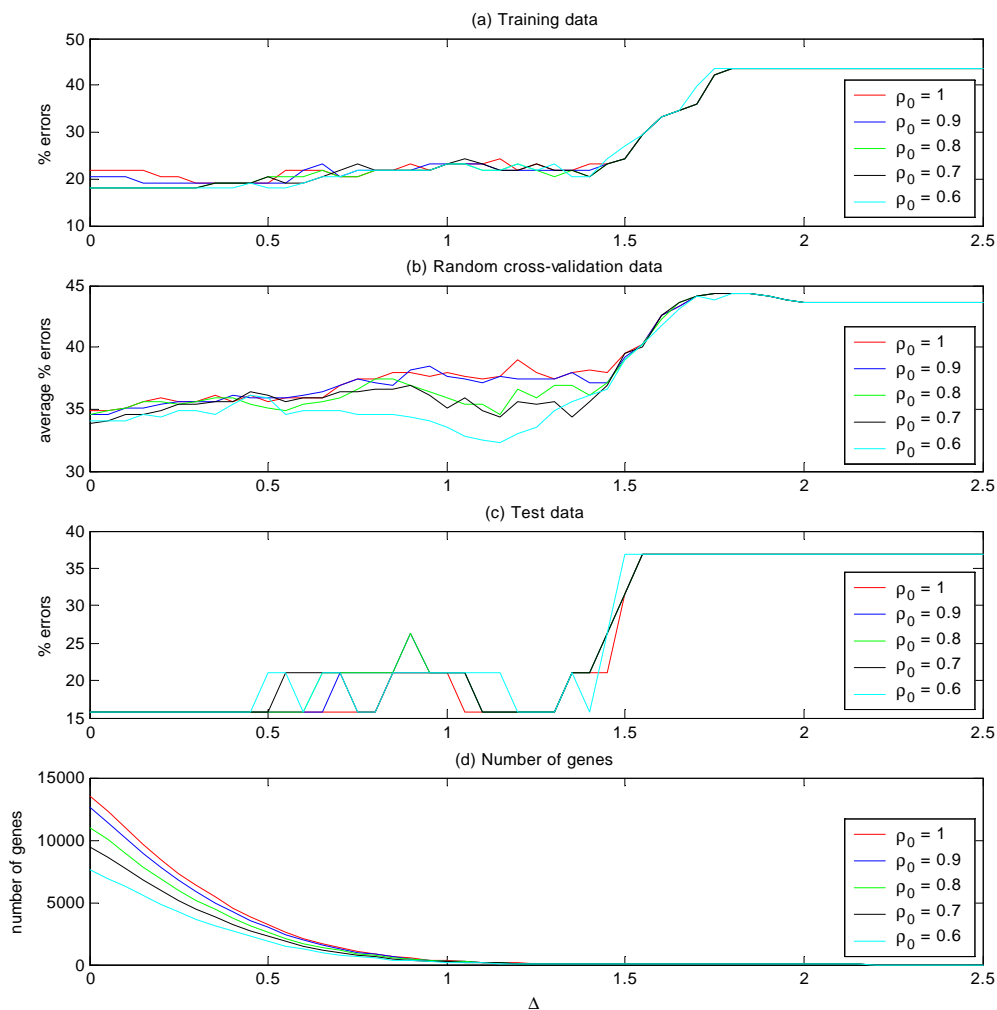


(e) Random cross-validation data ( $\rho_0 = 0.7$ )

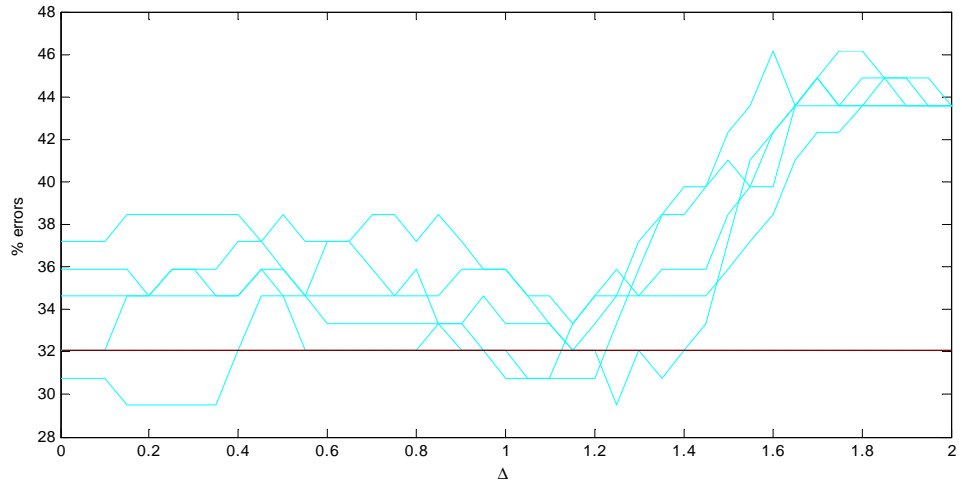


**Figure S9 - Prediction accuracy results on the breast cancer data using the USC algorithm over the range of  $D$  from 0 to 2.5.**

The percentage of classification errors is plotted against  $\Delta$  on (a) the full training set (78 samples) and (c) the test set (19 samples). In (b), the average percentage of errors is plotted against  $\Delta$  on the 10-fold cross validation data over five different random 10-fold splits of the training set. In (d), the number of relevant genes is plotted against  $\Delta$ . Different colors are used to specify different correlation thresholds ( $\rho_0 = 0.6, 0.7, 0.8, 0.9$  or 1). The number of errors for each of the 5 random 10-fold cross validation data sets at  $\rho_0 = 0.7$  is plotted over  $\Delta = 0$  to 2 in (e).

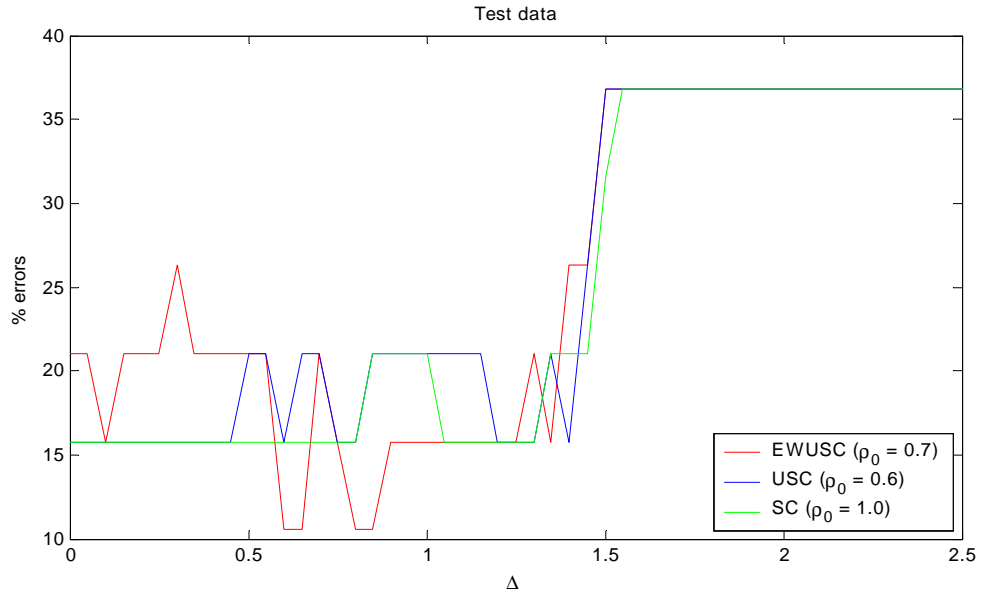


(e) Random cross-validation data ( $\rho_0=0.6$ )



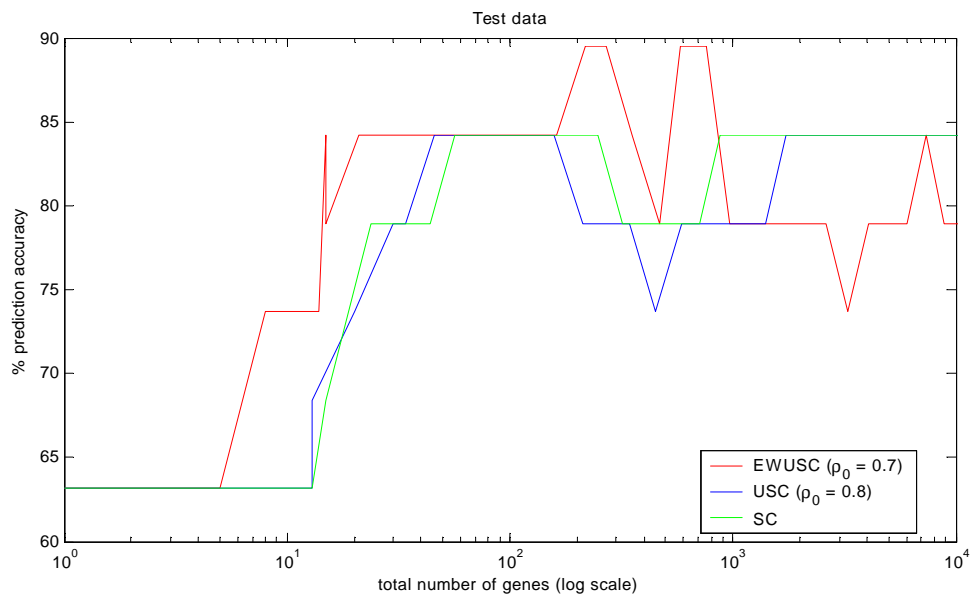
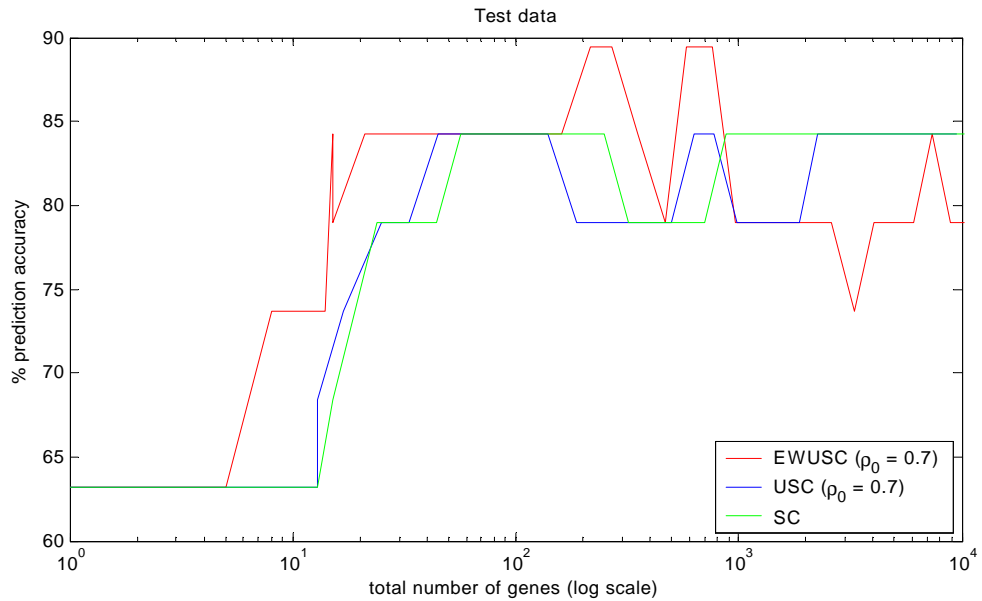
**Figure S10 – Comparing prediction accuracy of EWUSC and USC on the breast cancer data.**

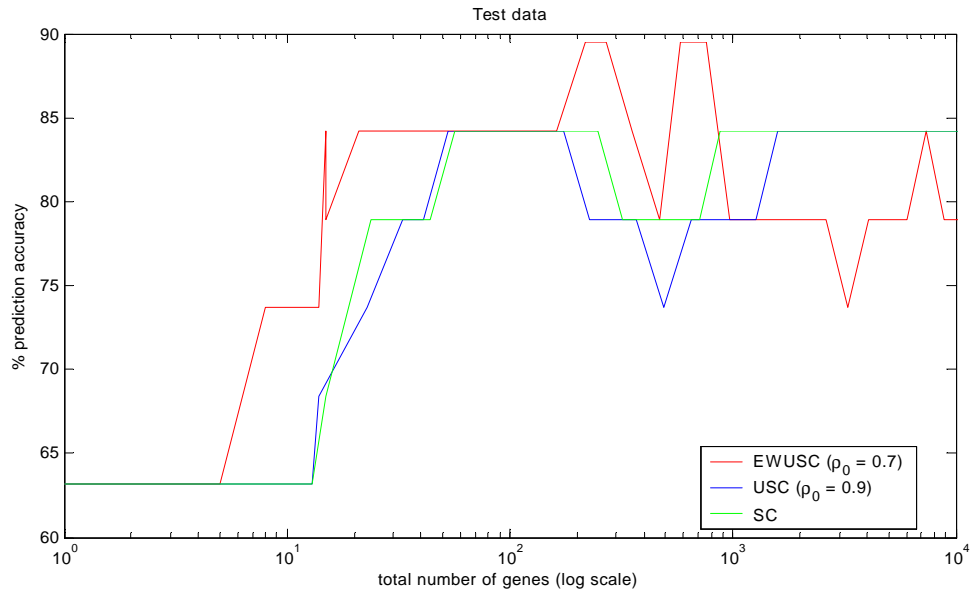
The percentage of classification errors is plotted against  $\Delta$  on the test set using the EWUSC algorithm at  $\rho_0 = 0.7$ , the USC algorithm at  $\rho_0 = 0.6$  and the original shrunken centroid algorithm (USC at  $\rho_0 = 1.0$ ).



**Figure S11 - Comparing prediction accuracy of EWUSC, USC and SC on the breast cancer data at different correlation thresholds.**

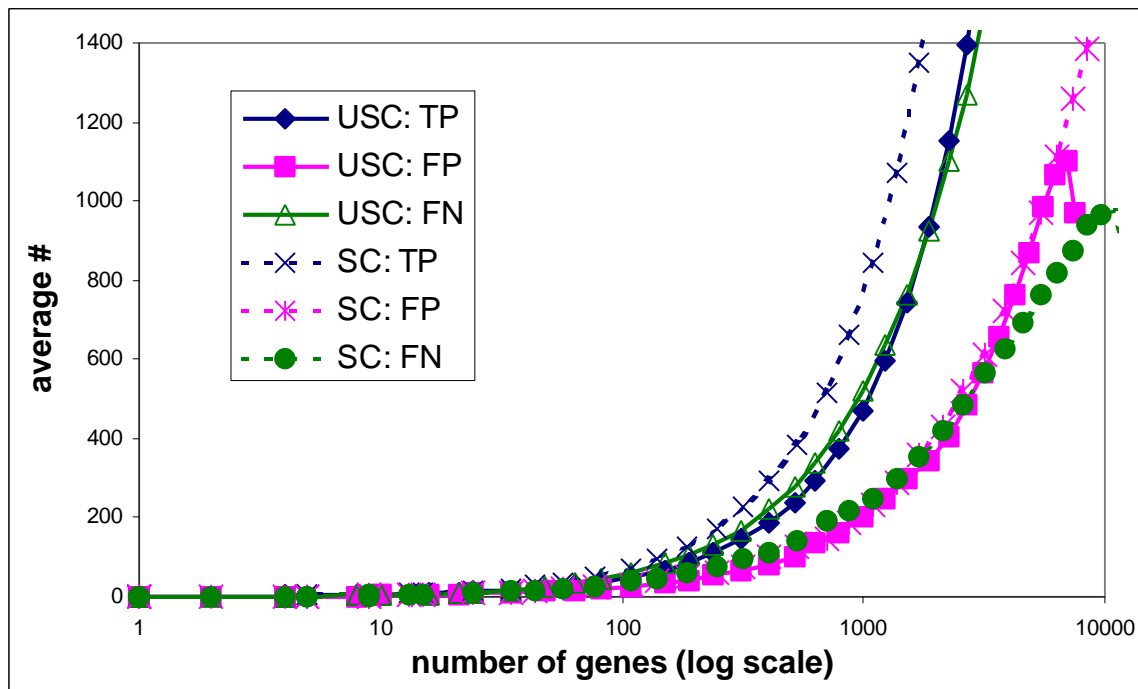
The following figures complement Figure 5 in the manuscript. The percentage of prediction accuracy is plotted against the number of relevant genes using the EWUSC algorithm at  $\rho_0 = 0.7$ , the USC algorithm at  $\rho_0 = 0.7, 0.8, 0.9$  and the SC algorithm (USC at  $\rho_0 = 1.0$ ). Note that the horizontal axis is shown on a log scale.





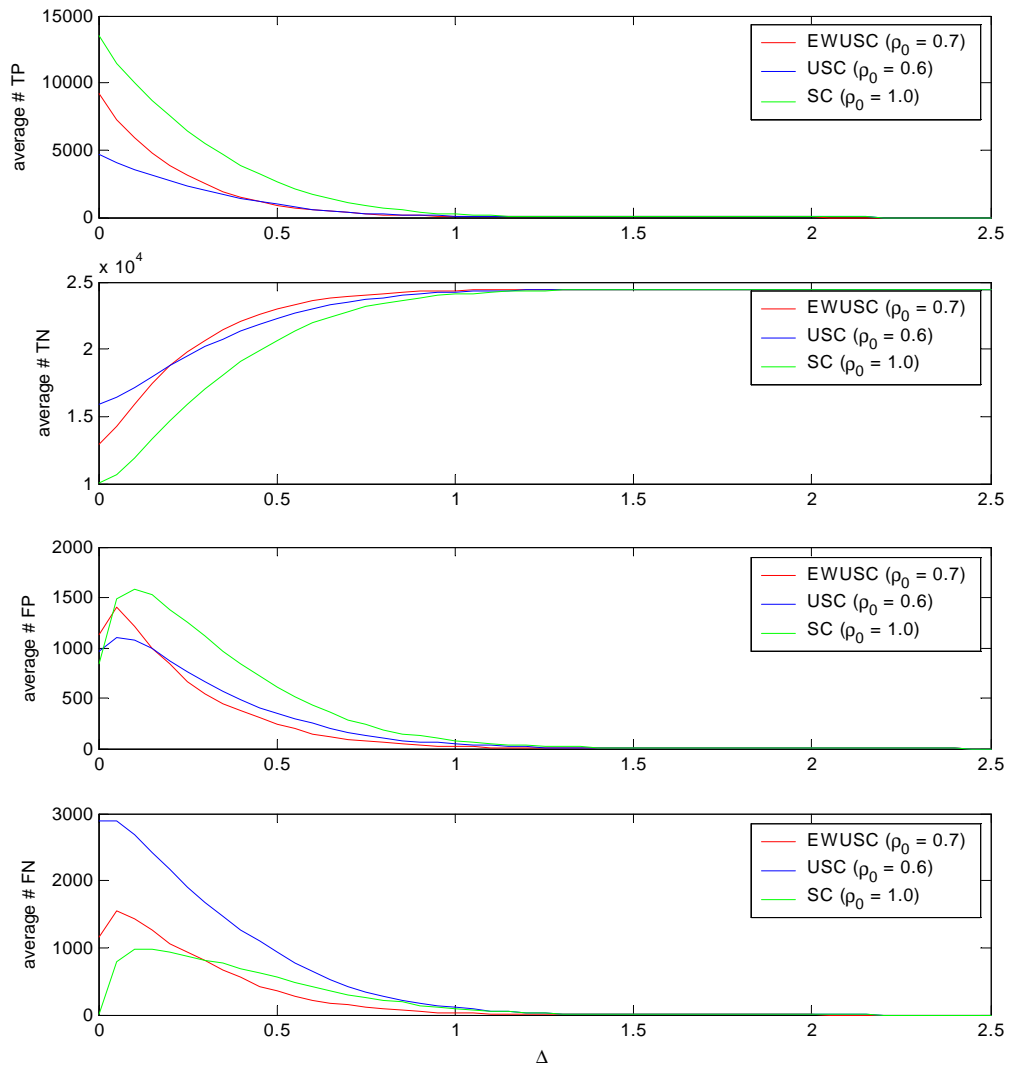
**Figure S12 – Comparing feature stability of USC vs. SC on the breast cancer data in terms of TP, FP and FN.**

The average numbers of true positives (TP), false positives (FP) and false negatives (FN) are plotted against the number of selected genes (shown in log scale) over five random runs of 10-fold cross validation using the EWUSC algorithm at  $\rho_0 = 0.7$  and the USC algorithm at  $\rho_0 = 0.6$ .



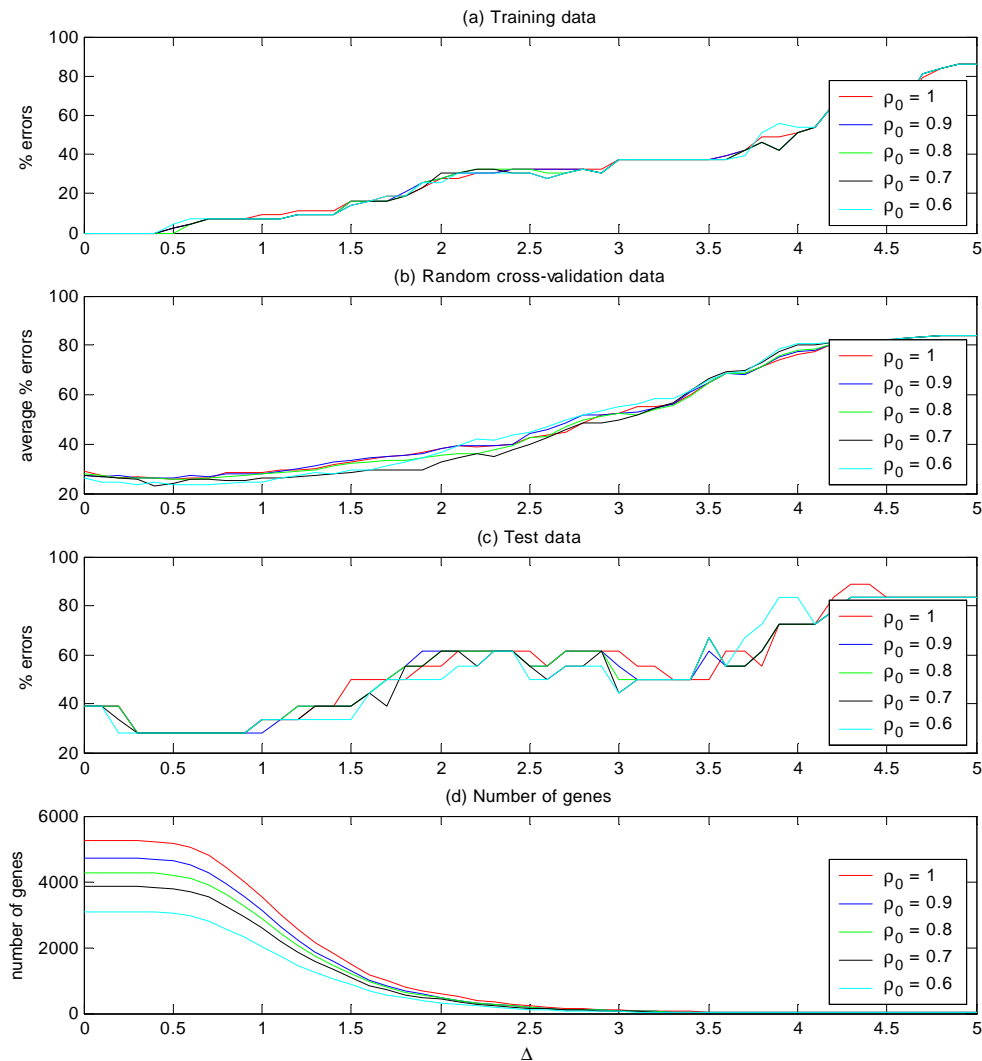
**Figure S13 - Comparison of feature stability results on the breast cancer data using the EWUSC and USC algorithms.**

The average numbers of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are plotted against  $\Delta$  for EWUSC at  $\rho_0 = 0.7$ , USC at  $\rho_0 = 0.6$  and SC.

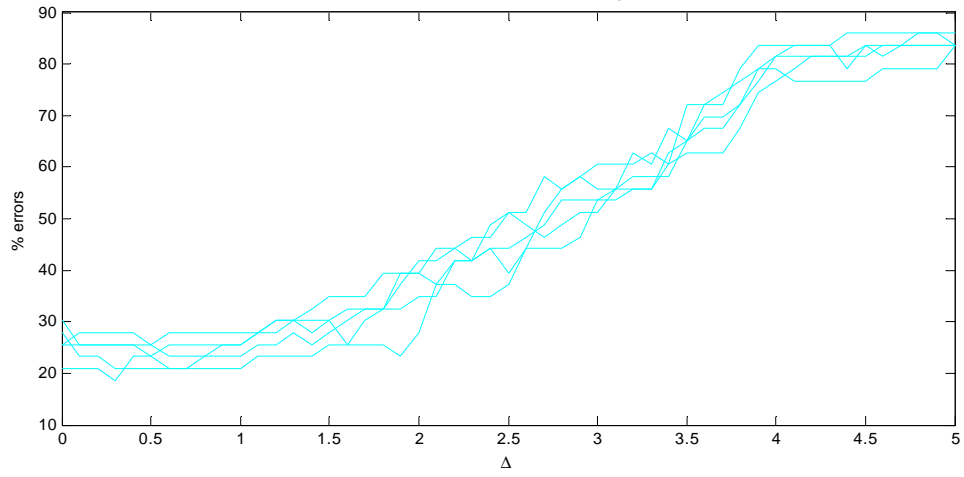


### Figure S14 - Prediction accuracy results on the NCI 60 data using the USC algorithm over the range of $D$ from 0 to 5

The percentage of classification errors is plotted against  $\Delta$  on (a) the training set (43 samples) and (c) the test set (18 samples). In (b), the average percentage of errors is plotted against  $\Delta$  on the 3-fold cross validation data over different random 3-fold splits of the training set. In (d), the number of relevant genes is plotted against  $\Delta$ . Different colors are used to specify different correlation thresholds ( $\rho_0 = 0.6, 0.7, 0.8, 0.9$  or  $1$ ). The percentage of errors for each of the 5 random runs of 3-fold cross validation at  $\rho_0 = 0.6$  is plotted against  $\Delta = 0$  to 5 in (e).

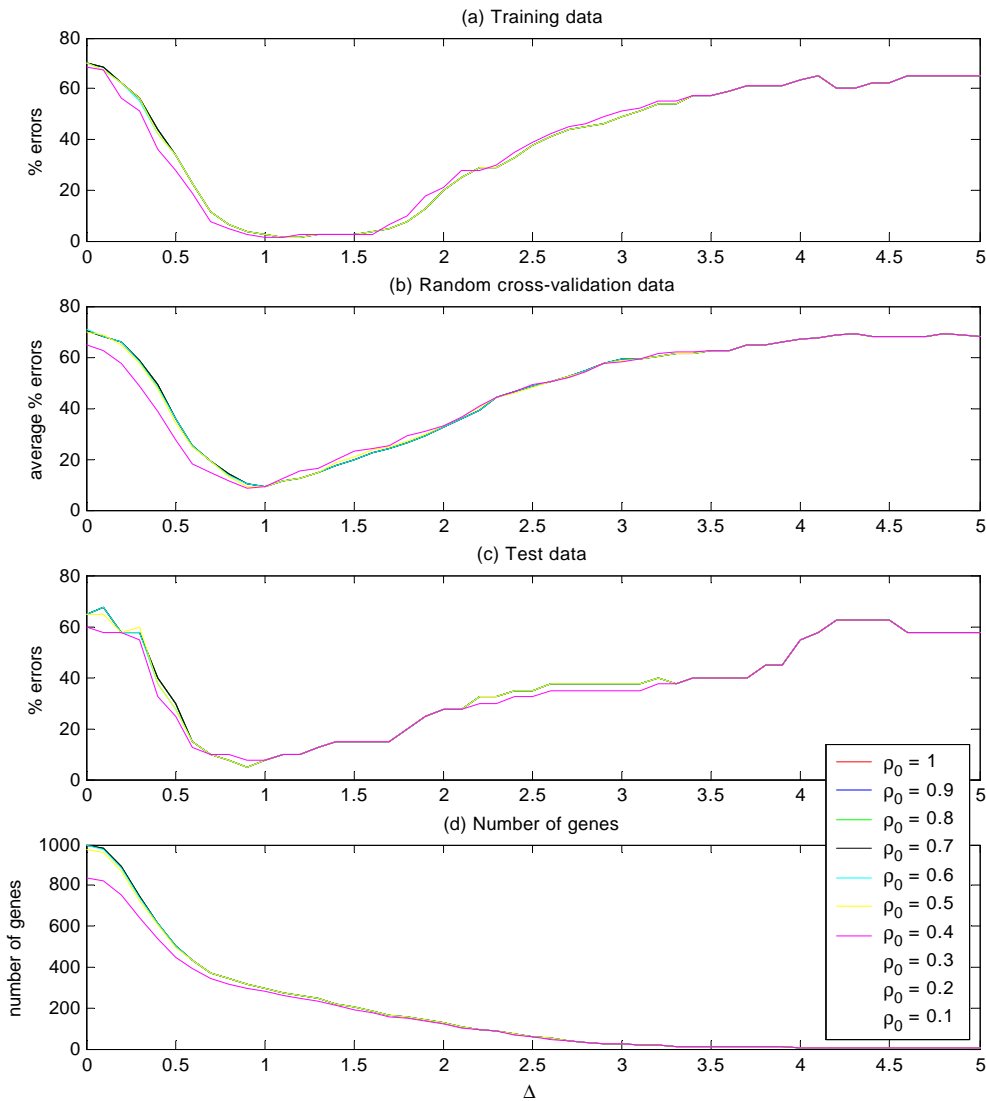


(e) Random cross-validation data ( $\rho_0 = 0.6$ )



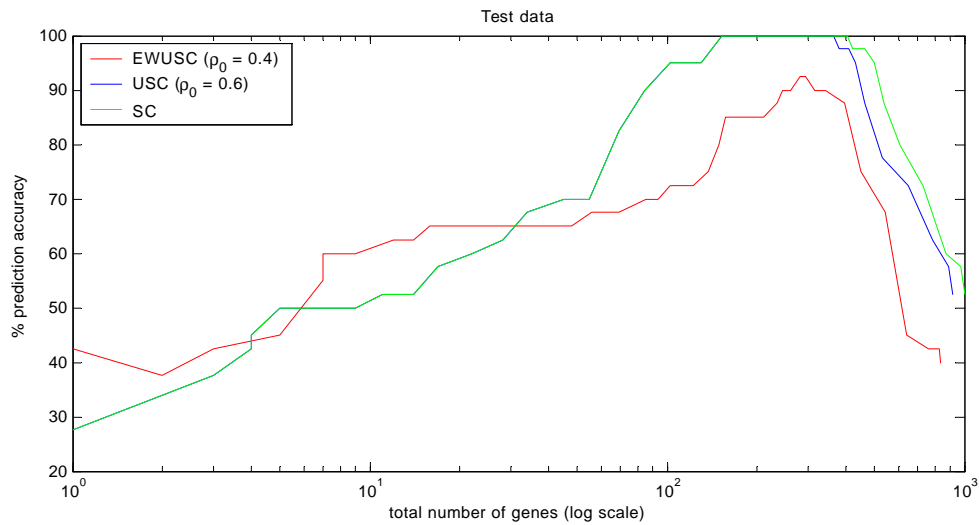
**Figure S15 - Prediction accuracy results on a synthetic dataset with 4 repeated measurements at low technical noise ( $a = 2$ ,  $l = 1$ ) using the EWUSC algorithm over the range of  $D$  from 0 to 5.**

The percentage of classification errors is plotted against  $\Delta$  on (a) the training set (80 samples) and (c) the test set (40 samples). In (b), the average percentage of errors is plotted against  $\Delta$  on the 5-fold cross validation data over different random 5-fold splits of the training set. In (d), the number of relevant genes is plotted against  $\Delta$ . Different colors are used to specify different correlation thresholds ( $\rho_0 = 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$  or  $1$ ).



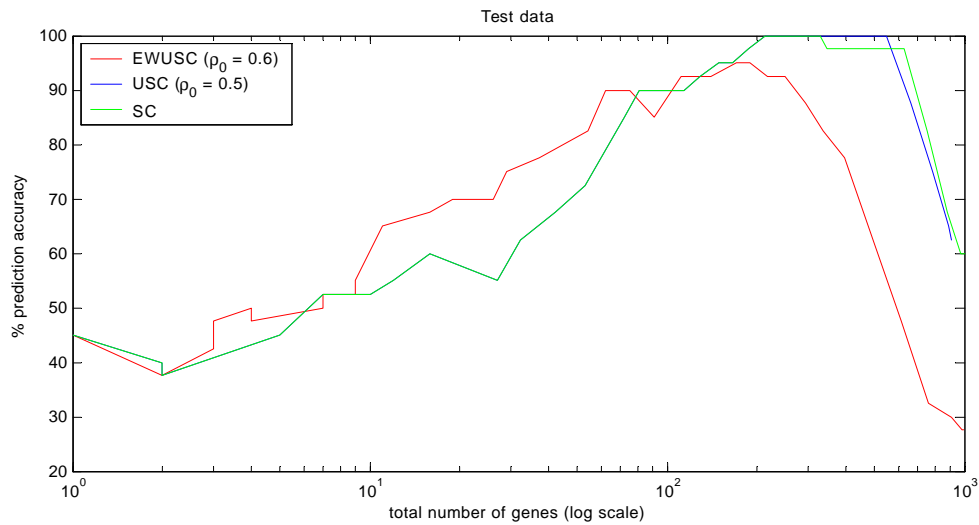
**Figure S16 - Comparing prediction accuracy of EWUSC, USC and SC on a synthetic dataset with 4 repeated measurements and low technical noise ( $\alpha = 2, \lambda = 1$ ).**

The percentage of prediction accuracy is plotted against the number of relevant genes using the EWUSC algorithm at  $\rho_0 = 0.4$ , the USC algorithm at  $\rho_0 = 0.6$  and the SC algorithm (USC at  $\rho_0 = 1.0$ ). Note that the horizontal axis is shown on a log scale. USC and SC tend to produce more accurate predictions.



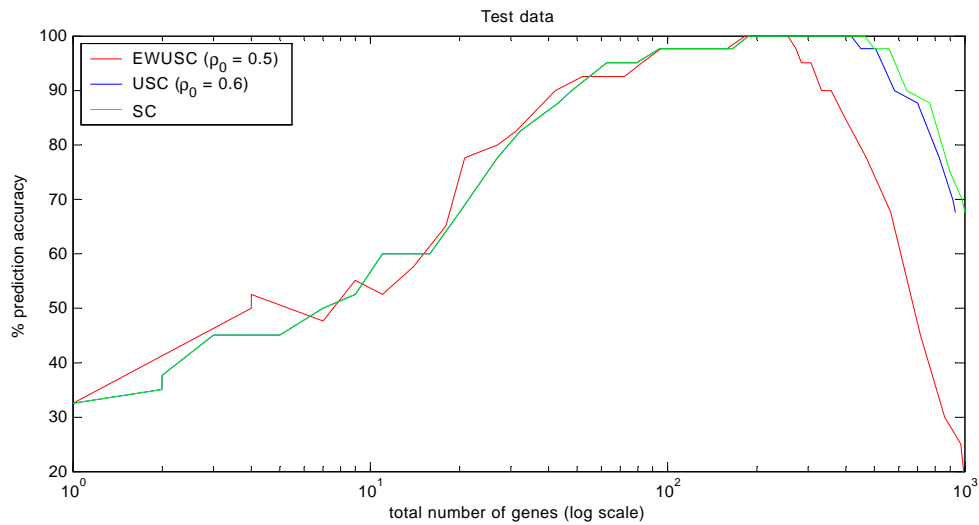
**Figure S17 - Comparing prediction accuracy of EWUSC, USC and SC on a synthetic dataset with 4 repeated measurements at high technical noise ( $\alpha = 2, \lambda = 10$ ).**

The percentage of prediction accuracy is plotted against the number of relevant genes using the EWUSC algorithm at  $\rho_0 = 0.6$ , the USC algorithm at  $\rho_0 = 0.5$  and the SC algorithm (USC at  $\rho_0 = 1.0$ ). Note that the horizontal axis is shown on a log scale.



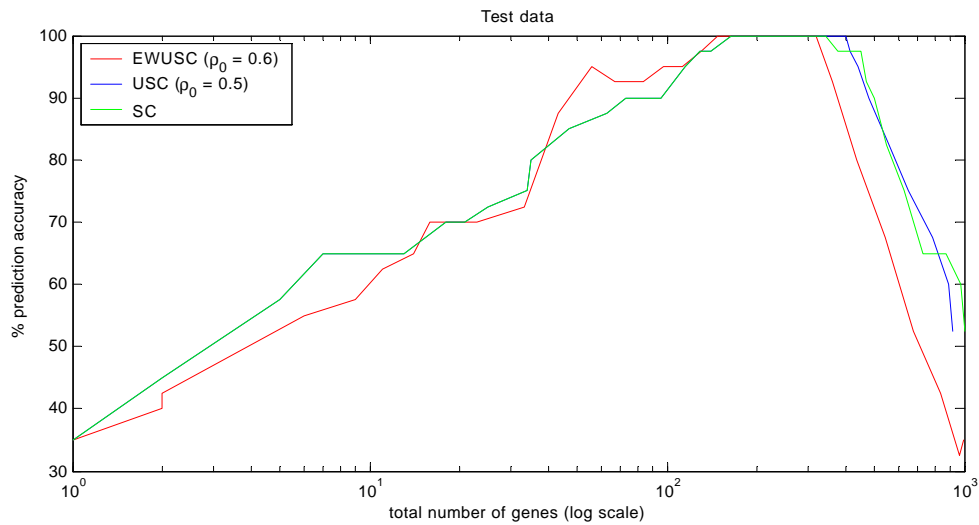
**Figure S18 - Comparing prediction accuracy of EWUSC, USC and SC on a synthetic dataset with 20 repeated measurements at low technical noise ( $\alpha = 2, \lambda = 1$ ).**

The percentage of prediction accuracy is plotted against the number of relevant genes using the EWUSC algorithm at  $\rho_0 = 0.5$ , the USC algorithm at  $\rho_0 = 0.6$  and the SC algorithm (USC at  $\rho_0 = 1.0$ ). Note that the horizontal axis is shown on a log scale.



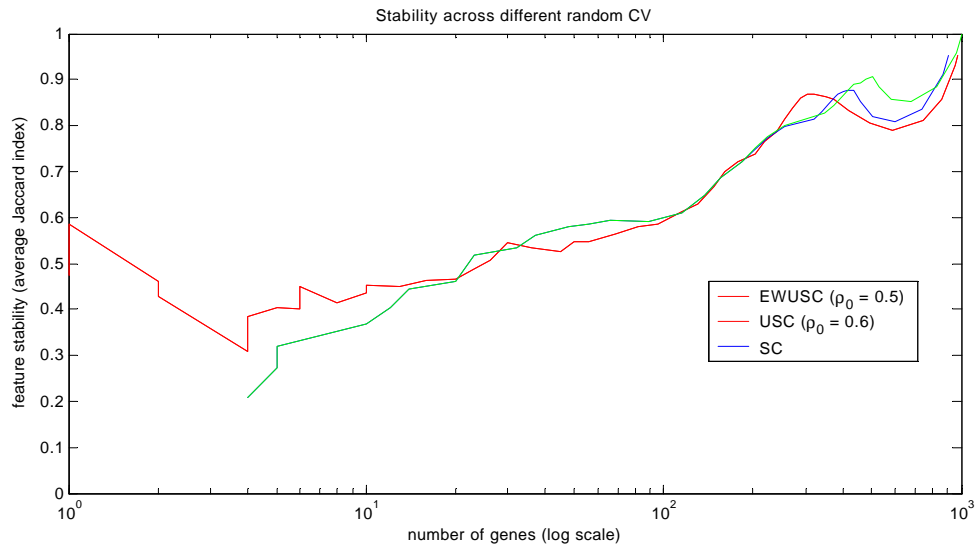
**Figure S19 - Comparing prediction accuracy of EWUSC, USC and SC on a synthetic dataset with 20 repeated measurements at high technical noise ( $\alpha = 2, \lambda = 10$ ).**

The percentage of prediction accuracy is plotted against the number of relevant genes using the EWUSC algorithm at  $\rho_0 = 0.6$ , the USC algorithm at  $\rho_0 = 0.5$  and the SC algorithm (USC at  $\rho_0 = 1.0$ ). Note that the horizontal axis is shown on a log scale.

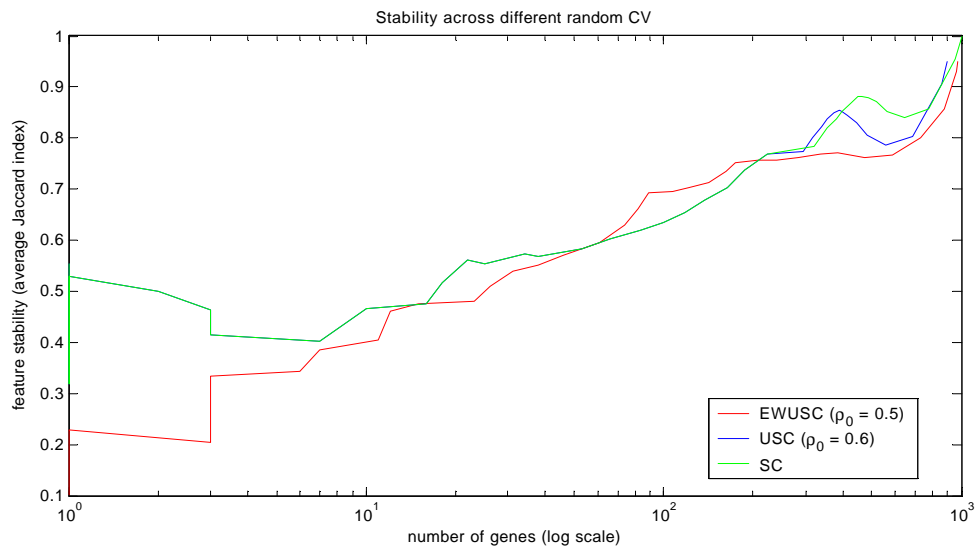


**Figure S20 - Comparing feature stability of EWUSC, USC and SC on a synthetic dataset with 4 repeated measurements at  $\alpha = 2$ .**

(a)  $\lambda = 1$

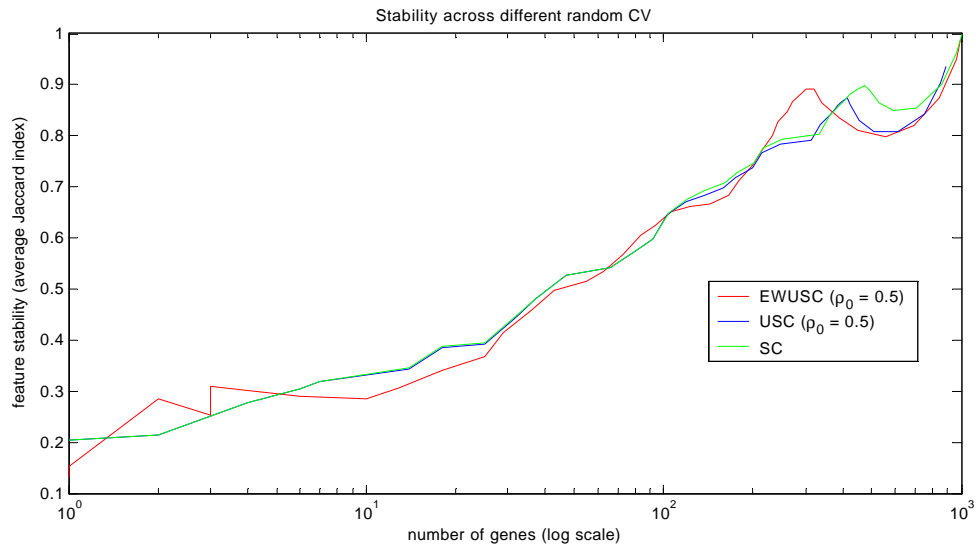


(b)  $\lambda = 10$



**Figure S21 - Comparing feature stability of EWUSC, USC and SC on a synthetic dataset with 20 repeated measurements at  $\alpha = 2$ .**

(a)  $\lambda = 1$



(b)  $\lambda = 10$

