**AUTHOR:** Your article was submitted with a Data Supplement and/or a Protocol, which is provided with your proof package *for your reference only*. Unlike your article text, tables, figures, acknowledgment, and appendix materials (when applicable), Data Supplement and Protocol files are not copyedited or composed. They will be included with your published article exactly as submitted. *We are unable to make any corrections to these supplemental materials.*

If any changes to a Data Supplement or Protocol are required, you must make the changes and return the modified file as an attachment with your corrected proof.

A Cautionary Note on using Cross-validation for Molecular Classification

Li-Xuan Qi et al

Supplementary Figure 2. Scatterplot comparing the classification error based on cross-validation when the training data were normalized once before-hand versus when the training data were re-normalization within K-fold cross-validation (that is, the K-1 folds data was re-normalized and the Kth fold was frozen normalized to the K-1 fold).
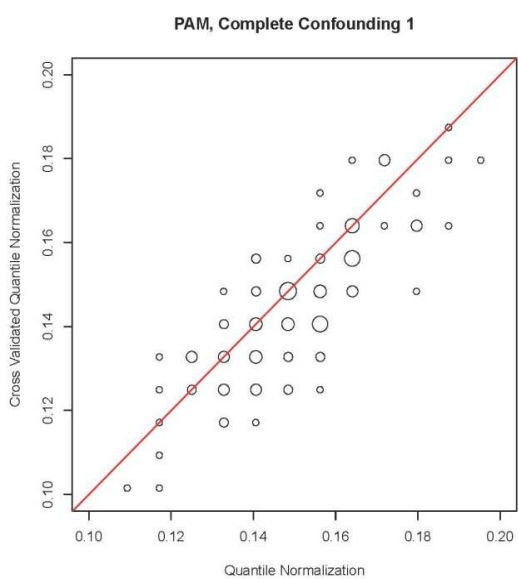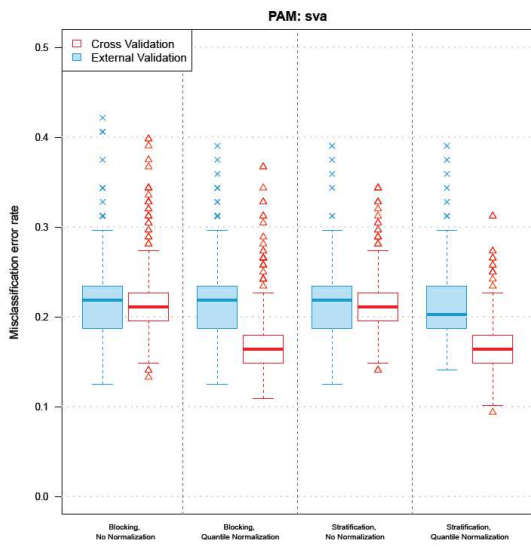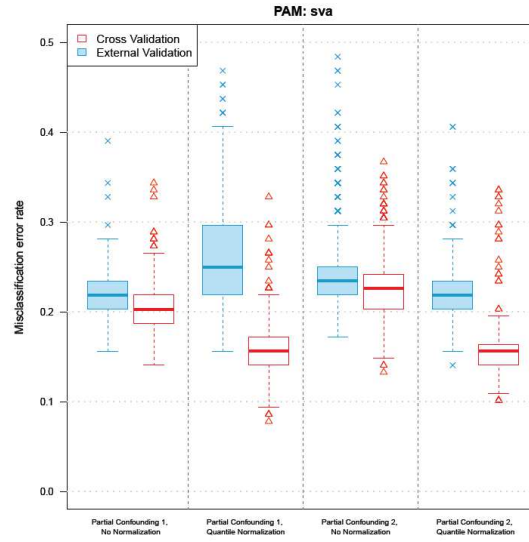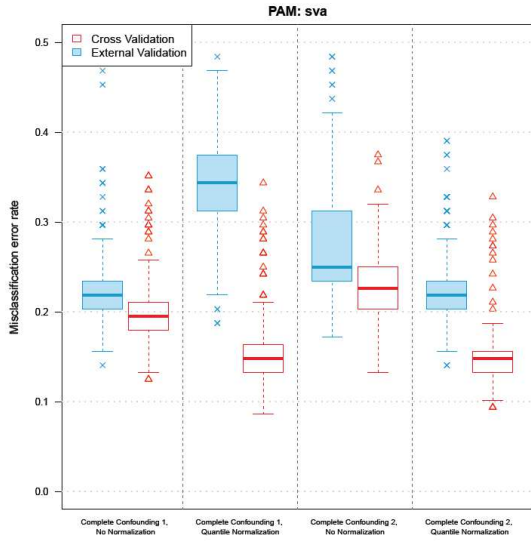
Supplementary Figure 3. Boxplots of classification error rates based on 1,000 simulated datasets under a confounding or balanced design using the PAM method, after applying a batch effect correction method (A. ComBat; B. SVA; C. RUV4) either alone or followed by quantile normalization.
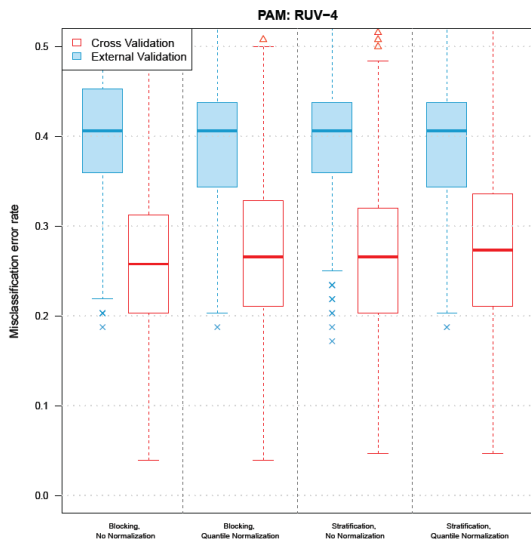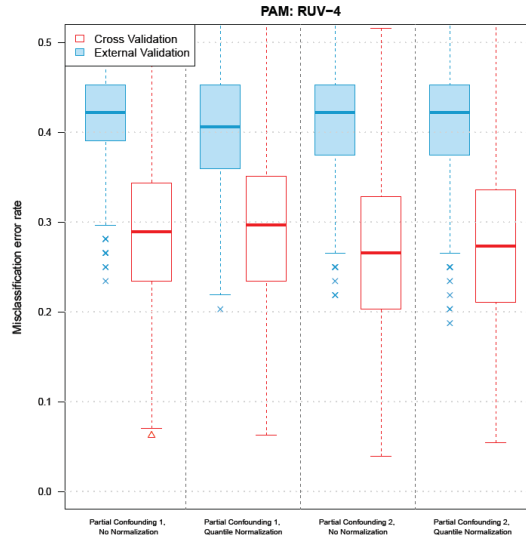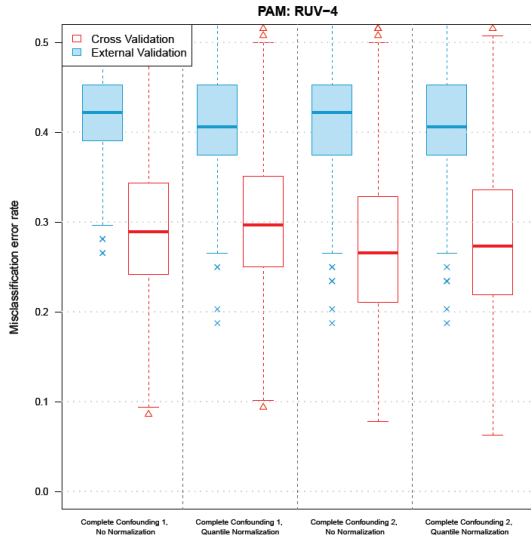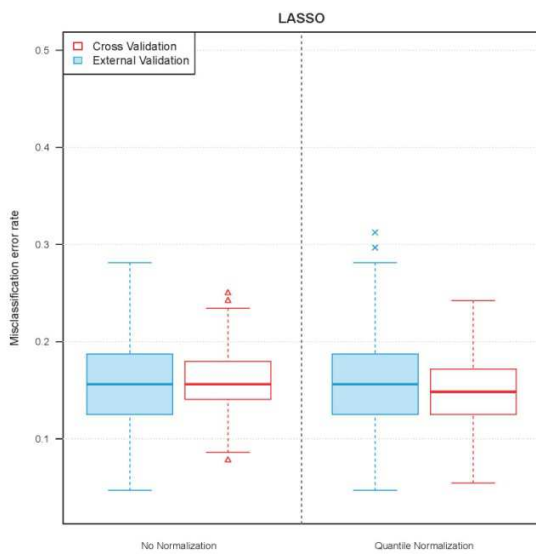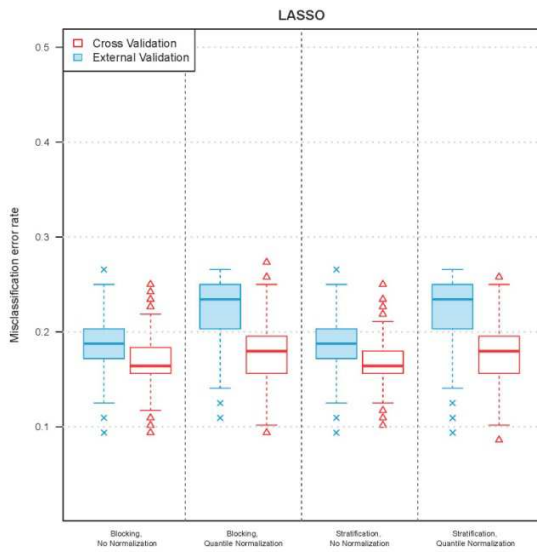
A. ComBat
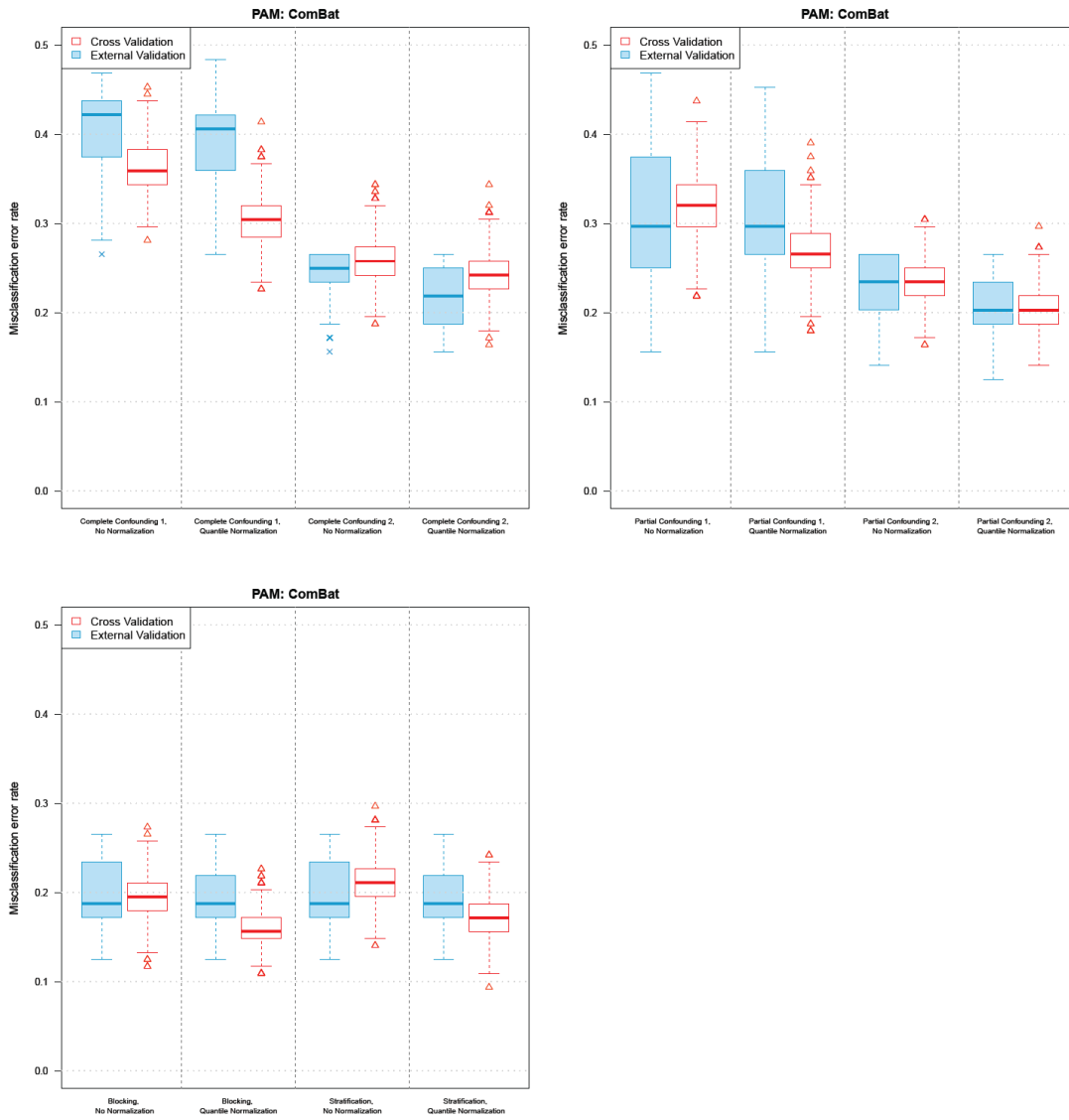
B. SVA

C. RUV-4

Supplementary Figure 4. Boxplots of classification error rates based on 1,000 simulated datasets under a confounding or balanced design using the LASSO method. The error rate was estimated with (1) cross-validation and (2) external validation using the corresponding test set (with the same training-versus-test sample-split) from the uniformly-handled data. The x-axis indicates the array-assignment design and the normalization status of the training data.

   A.   **Analysis of the uniformly-handled data**
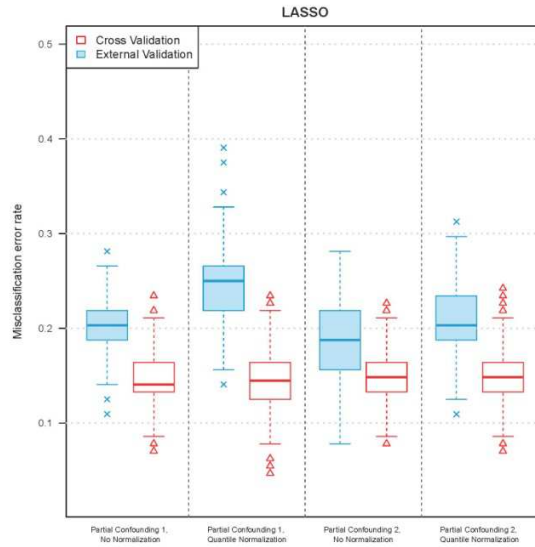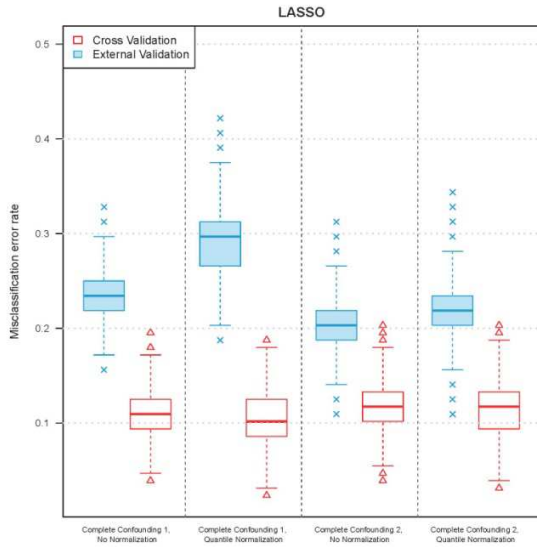
B. Analysis of the simulated data

Supplementary Figure 5. Boxplots of classification error rates based on 1,000 simulated datasets under a confounding or balanced design using the PAM method, when the biological signal was strong. The error rate was estimated with (1) cross-validation and (2) external validation using the corresponding test set (with the same training-versus-test sample-split) from the uniformly-handled data. The x-axis indicates the array-assignment design and the normalization status of the training data.
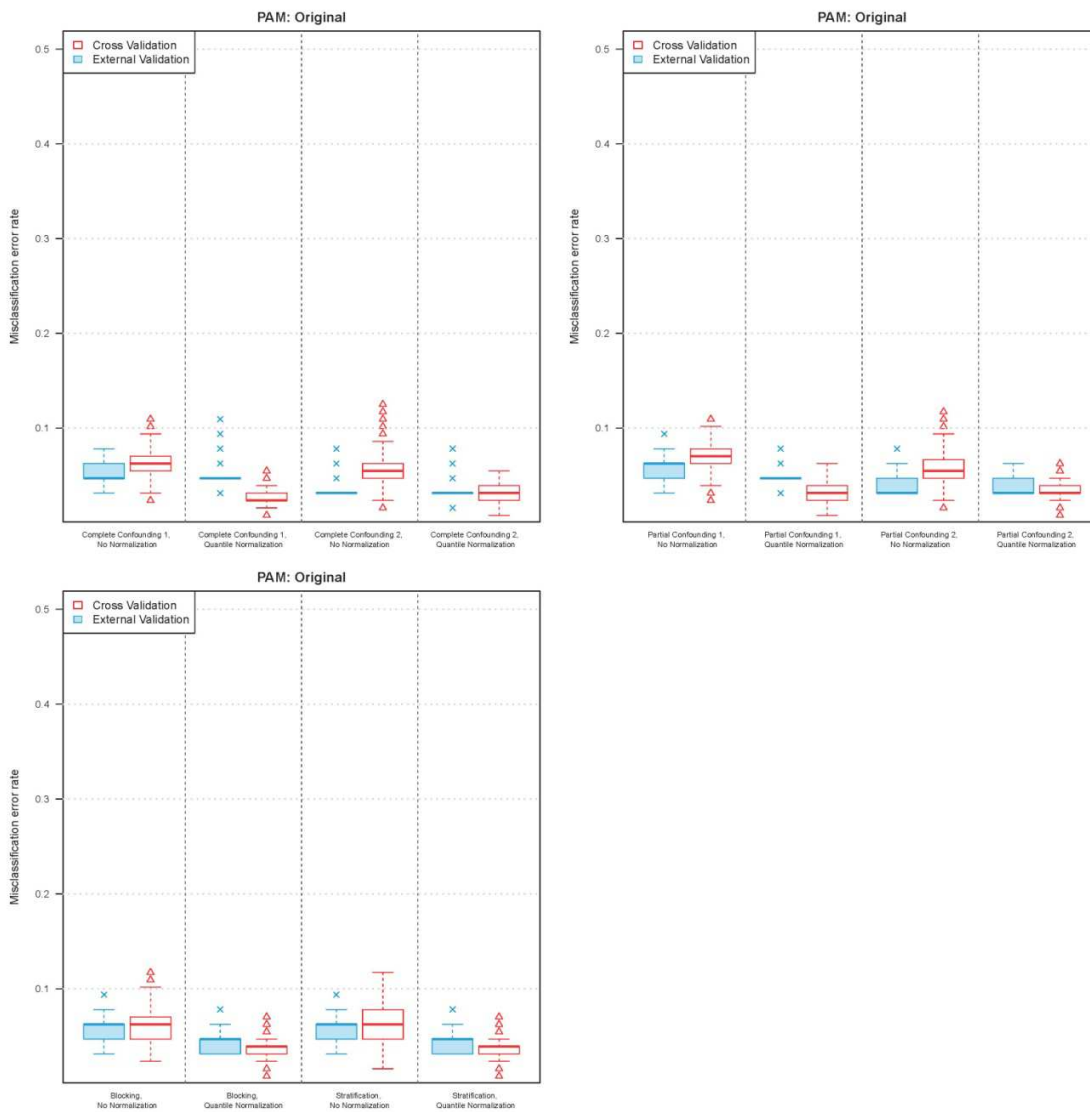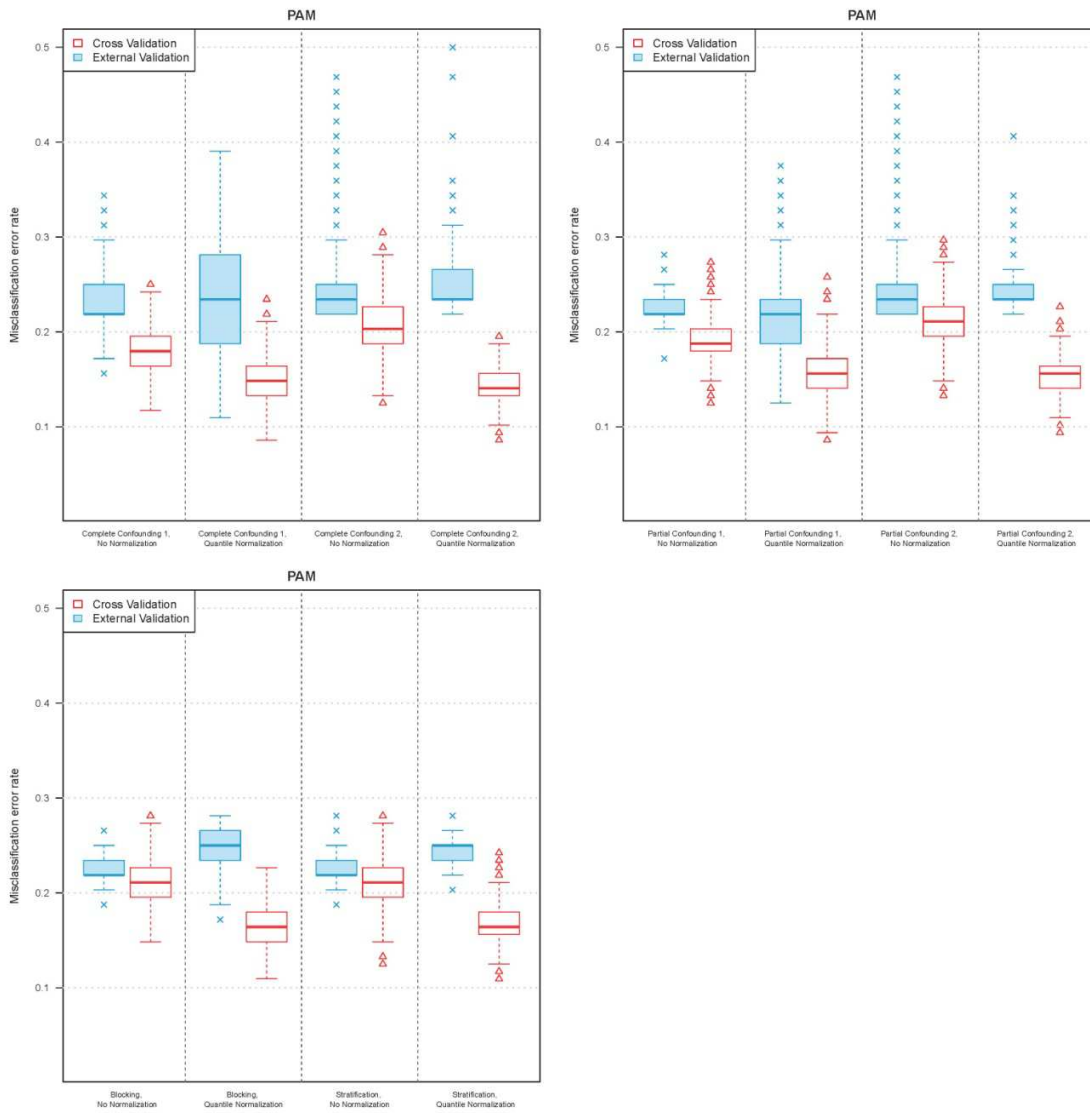
Supplementary Figure 6. Boxplots of classification error rates based on 1,000 simulated datasets under a confounding or balanced design using the PAM method, when the amount of handling effects was amplified. The error rate was estimated with (1) cross-validation and (2) external validation using the corresponding test set (with the same training-versus-test sample-split) from the uniformly-handled data. The x-axis indicates the array-assignment design and the normalization status of the training data.
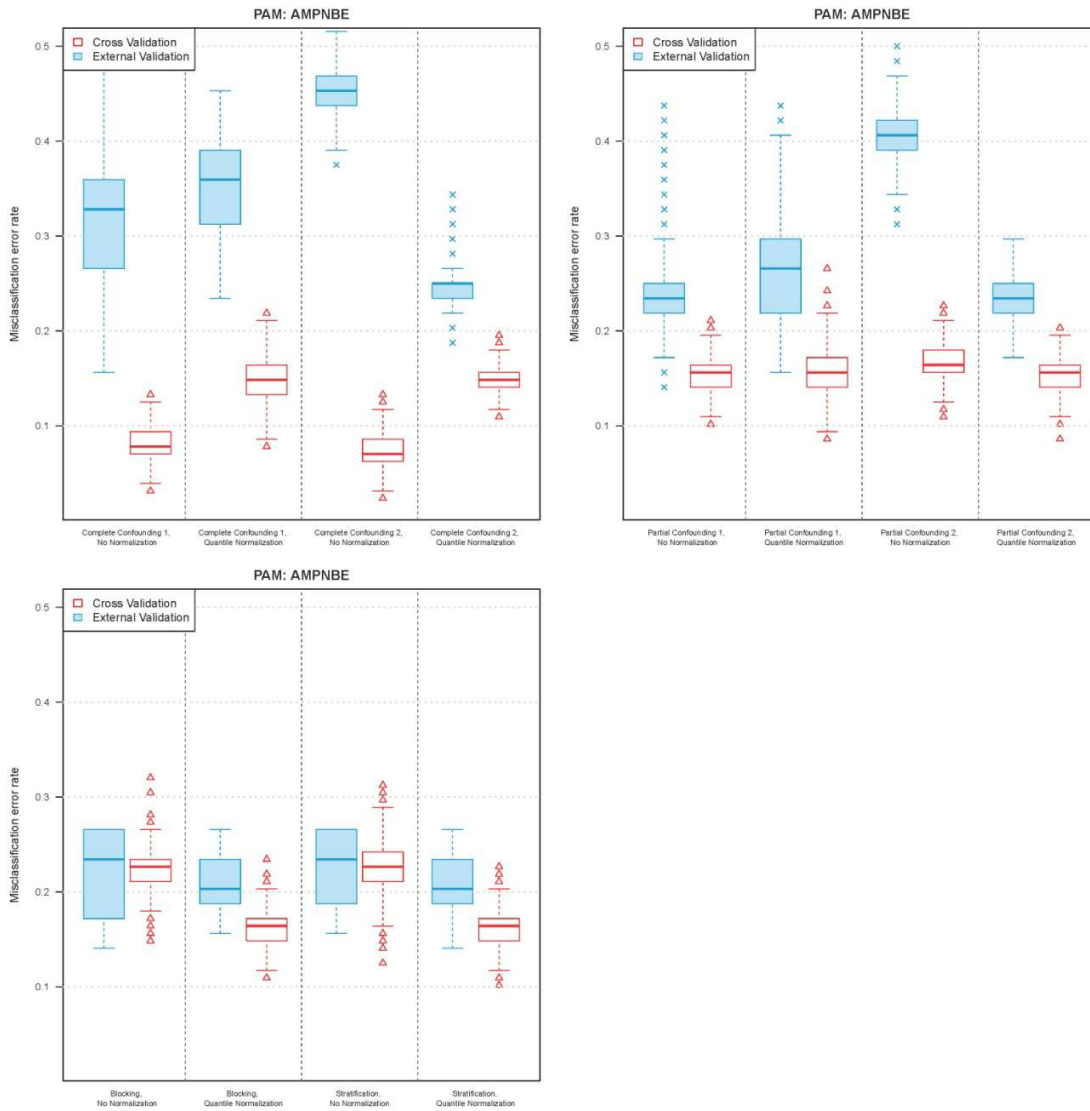
Supplementary Figure 7. Simulation results for a second split of samples to training set and test set.

A. PAM

B. LASSO