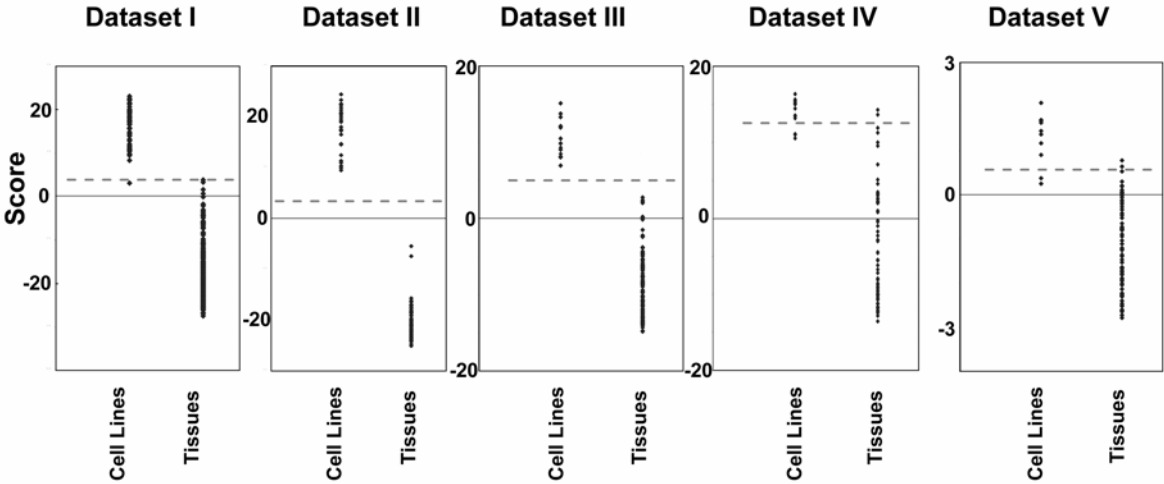


## Additional data file 2



Classification of cell lines and tissue samples across 5 datasets based upon their expression of the differentially expressed genes identified in dataset I and II. The difference in Euclidean distance (score) for each sample dataset respectively is presented and the samples within each dataset was stratified for the true cell lines and tissue samples (x-axis categories). The dashed line represent the objective classification thresholds used that gives equal number of false positives and false negatives. For details upon classification see materials and methods.