

Figure 1

Hu et al.

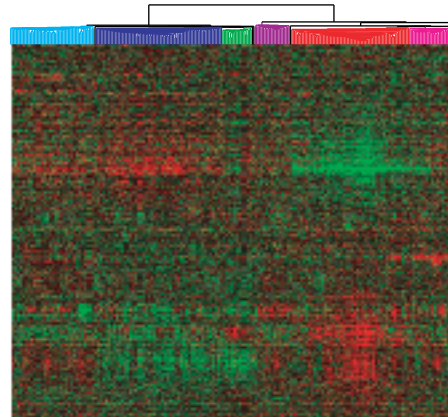
Training Set

A dataset of 105 breast tumor samples, 9 normal breast samples, and 26 sample pairs (each pair of samples is taken from the same patient), represented by 146 arrays, is used to derive the 1300-gene "Intrinsic/UNC" gene set.



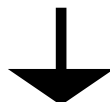
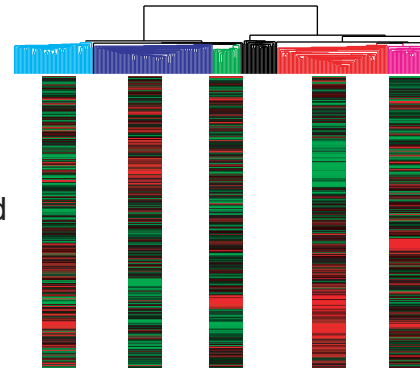
Combined Test Set

A test set of 311 tumors and 4 normal breast samples represented by 315 arrays and 2800 genes in common, was created by combining the datasets of Sorlie et al. (2001; 2003), van't Veer et al. (2002) and Sotiriou et al. (2003). This "combined test set" was analyzed by hierarchical clustering using the subset of "Intrinsic/UNC" genes that were present within the combined test set (306 genes).



Single Sample Predictor (SSP)

The hierarchical clustering of the "combined test set" is used to create 5 Subtype Mean expression profiles (i.e. Centroids) based upon the expression of the 306 Intrinsic/UNC genes. New samples are then assigned to the nearest subtype/centroid as determined by Spearman correlation.



Validation of the SSP using 2 test datasets

The SSP is used to make subtype predictions on 2 test sets of homogeneously treated patients. The resulting classifications were then analyzed using Kaplan-Meier Survival plots.

