

### Figure S4. Classification accuracy as a function of number of genes selected.

We used an algorithm named “random-forest coupled with recursive gene elimination” to select a small number of genes that can best discriminate the training samples into the four subtypes. In order to avoid sample selection bias, bootstrap resampling within the training data was implemented. The x-axis is the number of genes to be selected. The y-axis is the averaged gene classification accuracy across the bootstrap samples.

