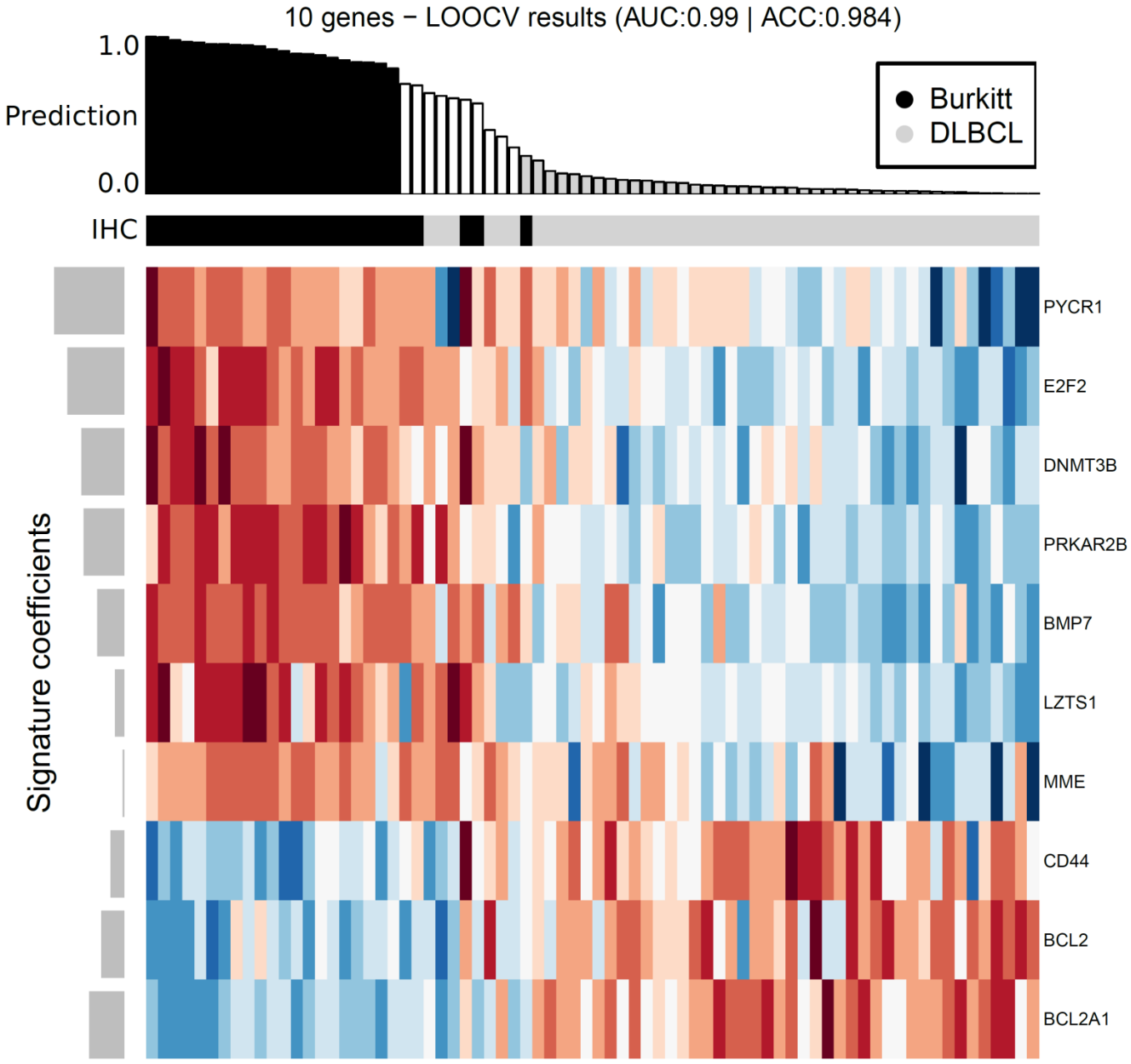

Figure S1. Parsimonious 10-gene classifier for DLBCL versus BL. Red indicates up-regulation and blue down-regulation of gene expression. Each row represents a gene and each column represents a sample. Given the robust results of the 21-gene assay using FFPE, we evaluated whether or not an even more parsimonious assessment of the transcriptome could feasibly replace IHC to refine diagnosis in samples from SSA. We applied a variance filter and used the top 10 genes to develop the classification model. The 26 BL and 48 DLBCL samples, defined by morphology and IHC, were used, leaving out the 6 BCL-U cases. An elastic net model (Hui Zou, T. H. (2005). Regularization and variable selection via the Elastic Net. Journal of the Royal Statistical Society, (Series B), 301–320. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.124.4696) with alpha parameter $0{\cdot}1^{15}$ in combination with leave-one-out cross-validation (LOOCV) was used to predict the lymphoma types. The output of our elastic net model is a probability value between 0·0 (DLBCL) and 1·0 (BL), which reflects the certainty of the model. Since we are only interested in predictions with high confidence we only called samples with a probability higher than 0·75 or lower than 0·25. This resulted in a relatively small intermediate class of uncertain samples (n=10), whereas the remaining 64 cases achieved certainty. The prediction performance was quantified by both accuracy and area under the ROC curve, for which we used only samples that were called by the classifier. Thus, the 10-gene classifier distinguished approximately 85% of pathology-defined BL from DLBCL.