

Figure S6. Heatmaps of MR16 gene expression in the Curtis and Guedj cohorts. Green and red colors indicate low and high gene expression, respectively. The same Group 1 MRs, as defined in the TCGA data (Figure 1A), are over-expressed in the ER-positive tumors (blue bar at the bottom), but lowly expressed in the ER-negative tumors (red bar at the bottom). In contrast, the same Group 2 MRs are over-expressed in the ER-negative tumors, but lowly expressed in the ER-positive tumors. **A:** The MR16 gene expression patterns in the Curtis cohort. Each column is a tumor sample. Due to the too large sample size, we used only the validation data that is approximately half of the 2000 samples. (Another half is the discovery data and results are similar.) Among the MR16 genes, *TP63* gene failed to the annotation as described in the Methods section, thus 15 genes are present in the heatmap. The samples display two large clusters that correspond to the ER-positive and ER-negative tumors. In each of the two clusters, there exist two sub-clusters. In general, the four clusters correspond well to the four major subtypes. **B:** The MR16 gene expression patterns in the Guedj cohort. Each column is a tumor sample. The samples also exhibit two large clusters that correspond to the ER-positive and ER-negative tumors. In each of the two clusters, the sub-clusters are more complicated, compared to the TCGA and the Curtis cohorts. This may suggest more subtypes in this cohort.

