**Supplementary Figure 1: Project outline.**

Patient samples which fulfilled QC criteria were selected and annotated. Representative H&E images of central tumor, invasive front and lymph node regions used for nucleic acid extraction are shown. Following microarray profiling, differential gene expression analysis was performed alongside supervised machine learning to generate a classifier for identifying samples from central or invasive regions.

**Supplementary Table 1: Clinico-pathological information for the 25 patients who were the source of the samples employed in this study.** M = Male; F = Female; LVI = Lymphovascular Invasion

**Supplementary Table 2: Differentially expressed gene list contrasting tumor region of origin between CT-IF.**

**Supplementary Table 3: Differentially expressed gene list contrasting tumor region of origin between CT-LN.**

**Supplementary Table 4: Differentially expressed gene list contrasting tumor region of origin between LN-IF.**