

Figure S1. Unsupervised classification of all patients with BC. 23 variables, including sex, age and tumor size, were inputted into the model and the model attempted to find the eigenvectors and eigenvalues in the 23-dimension space. These eigenvectors are a linear combination of the original 23 variables and are also referred to as transformed features. The corresponding eigenvalues stand for their contribution to the data variance. In general, features with high eigenvalues may be used to distinguish data sub-classes. (A) The top 5 transformed features and the corresponding variance they indicate. (B) Weights of the clinical characteristics that were the top two transformed features; each datapoint stands for an original feature, with tumor stage at the top right. (C) Unsupervised classification of overall BC using the top two transformed features. Each datapoint stands for one patient; the red triangles indicate patients with SPM and the blue circles patients with OOPM. Patients of the SPM and OOPM cohorts were not able to be distinguished from the figure. SPM, second primary malignancy; OOPM, only one primary malignancy; BC, breast cancer.

