Supplementary legends

Supplement 1: List of eligible data sets. A Untreated cohort (11 data sets, 1188 patients); B Treated cohort (16 data sets; 2334 patients)

Supplement 2: List of the 9 prognostic gene signatures evaluated in this study. A Proliferation-related gene signatures (GGI, GENE70, GENE76); B Stroma-related gene signatures (DCN, SDDP, PLAU); C Immune-related gene signatures (STAT1, IRM, Immunomodulatory)

Supplement 3: Distribution of breast cancer subtypes according to age. Subtypes are defined according to gene expression (see methods).

Supplement 4: Differences in relapse-free survival (RFS) according to age in the untreated cohort. A all available patients with relapse data; B ER+/HER2- subtype; C HER2-overexpressing subtype; D ER-/HER2- subtype

Supplement 5: Prognostic evaluation of gene signatures according to age groups in ER+/HER2- tumors. Dotted line represents a hazard ratio (HR) of 1.0, and error bars represent 95% confidence intervals (CI). All HR shown have been adjusted for AOL. A. ER+/HER2- SDPP; B ER+/HER2- DCN; C ER+/HER2- PLAU; D ER+/HER2-IRM; E ER+/HER2- STAT1; F ER+/HER2- Immunomodulatory. The p-value of an interaction between age as a continuous variable and the gene signature in a Cox-model and corresponding false discover rate (FDR) value is shown for each gene signature.
**Supplement 6:** Prognostic evaluation of stroma-related gene signatures according to age groups in HER2+ tumors. Dotted line represents a hazard ratio (HR) of 1.0; and error bars represent 95% confidence intervals (CI). All HR shown have been adjusted for AOL. A. PLAU; B DCN; C SDPP.

**Supplement 7:** Prognostic evaluation of immune-related gene signatures according to age groups in HER2+ and ER-/HER2- tumors. Dotted line represents a hazard ratio (HR) of 1.0; and error bars represent 95% confidence intervals (CI). All HR shown have been adjusted for AOL. A. HER2+ IRM; B HER2+ STAT1; C HER2+ Immunomodulatory; D ER-/HER2- IRM; E ER-/HER2- STAT1; F ER-/HER2- Immunomodulatory. The p-value of an interaction between age as a continuous variable and the gene signature in a Cox-model and corresponding false discovery rate (FDR) value is shown for each gene signature.

**Supplement 8:** List of genes and gene sets that were investigated in a linear regression model to determine their association with age (41 genes, 13 gene sets). For each gene or gene set, adjustments according to data set and covariates are provided along with false discovery rate (FDR).