Supplementary Figure Legends

Figure S1. Comparison of subtype calls between “Intrinsic subtype” and “PAM50”. (A) The distributions of the PAM50 calls within individual Intrinsic subtypes. The classification agreements for individual subtypes (in percentage) between Intrinsic and PAM50 were: 86.5% for Basal-like, 74.4% for HER2-enriched, 52.4% for LumA, 58.2% for LumB and 60.1% for Normal-like. Cohen's kappa coefficient $\kappa$ on discrete subtype calls was used to measure the overall agreement on the subtype assignments. (B) Correlation for Intrinsic signature and PAM50. Each dot represents the centroid correlation score for one sample from Intrinsic (X-axis) and PAM50 (Y-axis). Pearson correlation coefficient on subtype-centroid correlations was used to measure the stability of a specific subtype. (C) A distance measurement $\Delta$ per sample was defined to gauge the discriminant ability for individual subtypes in a signature and the degree of inconsistency for the same subtype in another signature. The D score was essentially the distance between the two subtype assignments on a sample, and it was computed as the difference of the corresponding centroid correlations. For a tumor per signature, the distance between the top two potential subtype assignments was measured: the larger the D, the more distinguishable the assigned subtype on the sample by the signature. In the case of inconsistent assignments for the same sample across the signatures, the distance between the previous assignment and the new assignment was measured to quantify the degree of the inconsistency. Upper panel: Samples were ranked by their distance measurements from intrinsic signature. Lower panel: X-axis indicates ranks from the intrinsic signature for individual samples. Y-axis indicates the D scores from PAM50 for the corresponding samples. Basal subtype was fairly robust in the two signatures,
as only the samples with borderline calls in the intrinsic signature had inconsistent calls in PAM50.

**Figure S2. Evaluation of time- & ER-dependency in predicting Distant Metastasis Free Survival by gene signatures on the complete dataset (n=912).** (A-B) Test for non-proportional hazard assumption for DMFS prediction by gene signatures in ER-positive group (A; n=692) and ER-negative group (B; n=220) separately. A univariate Cox model was fitted for each signature, and the scaled Schoenfeld residuals were tested against transformed time (Kaplan-Meier estimates). Residuals distributed randomly around 0 would satisfy the PH assumption. (C-D) Cumulative regression plots of the estimate along with 95 percent pointwise confidence interval from a univariate additive regression model for individual gene signatures. The curve in each plot reflects the cumulative effect of a signature covariate on survival over time. The slope of the curve reflects the effects of the signature on survival at any given time. Hence, a time-independent effect should result in a curve with a constant slope. (C) In ER-positive group (n=692). (D) In ER-negative group (n=220).

**Figure S3. Evaluation of time- & ER-dependency in predicting Distant Metastasis Free Survival by gene signatures on systemically untreated patients (A-C) and systemically treated patients (D-F), separately.** (A-B; D-E) Cumulative regression plots of the estimate along with 95 percent pointwise confidence interval from a univariate additive regression model for each individual gene signature. (A) In ER-positive & systemically untreated group (n=274). (B) In ER-negative & systemically untreated group (n=121). (D) In ER-positive & systemically treated
Estimated effect (standardized hazard ratios, $e^\beta$, with 95% confidence intervals) of gene signatures for survival prediction within different time intervals stratified by ER status. The X-axis indicates the follow-up time intervals: up to 5-year, 5-10 year, and beyond 10 year. Within each subinterval, a univariate Cox model per signature was fitted. The Y-axis indicates the estimated hazard ratios (HR) on a logarithmic scale corresponding to a 1 standard deviation increase in the signature. The null, HR=1, is indicated by the blue line. Solid dots indicate HRs significantly different from 1 ($P<0.05$). ER+ is denoted as red and ER− is denoted as blue.

**Figure S4. Analysis on additional parameters for breast cancer prognostication in the studied cohort.** (A-B) Known prognostic parameters predict Distant Metastasis Free Survival. (A) Kaplan Meier plot up to 15 years for: Node, Tumor size and Histological grade, respectively. (B) Kaplan Meier plot up to 15 years for Histological grade in ER+ subgroup and ER- subgroup, respectively. (C) TP53 mutated tumors associated with elevated hypoxic score in ER-negative group. Analysis is based on the 335 breast cancers that had been characterized for the presence or absence of TP53 mutations. One-tailed t-test was performed to assess the significance of increases in Hypoxia score for samples with TP53 mutated over the wild-type ones. (D) Univariate analysis on gene signatures with G1, G2, G3 separately in ER+ group. Risk scores for individual gene signatures were standardized, and prognostic effects of signatures were compared using Hazard Ratio from univariate Cox model: Risk = Signature in ER+ group (n= 715; panel ER+), G1 samples of the ER+ group, (n= 153; panel ER+/G1), G2 samples of the ER+ group, (n= 311; panel ER+/G2) and G3 samples of the ER+ group, (n=149; panel ER+/G3).
X-axis indicates the Standardized Hazard Ratio. (E) Univariate analysis on gene signatures with T1, T2, T3 separately in ER+ group. Risk scores for individual gene signatures were standardized, and prognostic effects of signatures were compared using Hazard Ratio from univariate Cox model: Risk = Signature in ER+ group (n=715; panel ER+), T1 samples of the ER+ group, (n= 349; panel ER+/T1), T2 samples of the ER+ group, (n= 335; panel ER+/T2) and T3 samples of the ER+ group, (n=26; column ER+/T3). X-axis indicates the Standardized Hazard Ratio.

Figure S5. (A) Evaluation of time- & HER2-dependency in predicting Disease-specific Survival by gene signatures on the complete dataset (n=912). Estimated effect (standardized hazard ratios, $e^\beta$, with 95% confidence intervals) of gene signatures for survival prediction within different time intervals stratified by HER2 status. The X-axis indicates the follow-up time intervals: up to 5 years, 5-10 years, and beyond 10 years. Within each subinterval, a univariate Cox model per signature was fitted. The Y-axis indicates the estimated hazard ratios (HR) on a logarithmic scale corresponding to a 1 standard deviation increase in the signature. The null, HR=1, is indicated by the blue line. Solid dots indicate HRs significantly different from 1 (P<0·05). HER2+ (n=114) is denoted as red and HER2− (n= 798) is denoted as blue. (B) Evaluation of time- & HER2 / ER dependency in predicting Disease-specific Survival by gene signatures on the complete dataset (n=912).

Figure S6. Evaluation of time- & ER-dependency in predicting Disease-specific Survival by gene signatures on METABRIC discovery set (n=996). (A-B) Cumulative regression plots of the estimate along with 95 percent pointwise confidence interval from a univariate additive regression model for each individual
gene signature. (A) In ER-positive group of the complete METABRIC discovery set (n=801). (B) In ER-negative group of the complete METABRIC discovery set (n=195). (C-E) Estimated effect (standardized hazard ratios, $e^\beta$, with 95% confidence intervals) of gene signatures for survival prediction within different time intervals stratified by ER status. The X-axis indicates the follow-up time intervals: up to 5-year, 5-10 year, and beyond 10 year. Within each subinterval, a univariate Cox model per signature was fitted. The Y-axis indicates the estimated hazard ratios (HR) on a logarithmic scale corresponding to a 1 standard deviation increase in the signature. The null, HR=1, is indicated by the blue line. Solid dots indicate HRs significantly different from 1 (P<0.05). ER+ is denoted as red and ER− is denoted as blue. (C) On the complete METABRIC discovery set. (D) On the systemically untreated patients and (E) On the systemically treated patients.
Figure S2

A

Intrinsic_RORs: p=0.00317

PAM50_RORs: p=0.000498

gene70: p=0.0232

gene76: p=0.00916

GGI: p=3.88e-05

WR: p=7.65e-05

Hypoxia: p=0.772

RS: p=0.00189

EP: p=0.000372

Intrinsic_RORs: p=0.00692

PAM50_RORs: p=0.368

gene70: p=0.0565

gene76: p=0.0702

GGI: p=0.0296

WR: p=0.000246

Hypoxia: p=0.713

RS: p=0.0628

EP: p=0.246

Time to distant metastasis (year)

Cumulative regression function

B

C

D

Figure S2
Figure S4

(A) Survival probability over time to distant metastasis (year) for different nodes:
- Node+: Blue solid line, Node−: Red dashed line.
  - Node+: p = 9.3e−07
  - Node−: p = 0.00033

(B) Survival probability over time to distant metastasis (year) for different histological grades:
- ER+: Green solid line, ER−: Red dashed line.
  - ER+: p = 0.00019
  - ER−: p = 0.57

(C) Hypoxia score distribution for TP53:
- TP53 wt: Blue dots, TP53 mut: Red circles.
  - TP53 wt: p = 0.29
  - TP53 mut: p = 0.029

(D) Gene signature comparison for different hormone receptor statuses:
- ER+/G1
- ER+/G2
- ER+/G3
- Intrinsic
- PAM50
- G1
- G2
- G3
- G4
- 70-gene
- 76-gene

(E) Univariate Cox model: Standardized Hazard Ratio for different hormone receptor statuses:
- ER+/T1
- ER+/T2
- ER+/T3
- TP53 wt
- TP53 mut
- pvalue
  - p ≤ 0.05
  - p > 0.05
Figure S5

A

Intrinsic  PAM50  70−gene  76−gene  GGI  WR  Hypoxia  RS  EP

Follow−up time interval (year)

Standardized hazard ratio

• $p \leq 0.05$  ○ $p > 0.05$

HER2+  HER2−

B

HER2−ER−  HER2−ER+  HER2+ER−  HER2+ER+

Standardized hazard ratio

Follow−up time interval (year)