**Supplementary Appendix**

***Selection of genes included in the multigene assay***

Genes were selected using gene expression data from microarray data of 1,754 estrogen receptor (ER)-positive and lymph node (LN)-negative breast cancer samples in seven Gene Expression Omnibus (GEO) datasets including GSE2034, GSE2990, GSE3494, GSE4922, GSE6532, GSE7390, and GSE12093 (1), and RNA-seq data of 1,025 ER-positive breast cancer samples in The Cancer Genome Atlas (TCGA) (2) and a previous study conducted at Seoul National University Hospital (SNUH) (3). Affymetrix GPL 96 was the platform used for the microarray data.

Microarray data were first used to derive the 21-gene recurrence score (RS), as previously described (4). In total, 135 genes with a mean Pearson’s correlation coefficient of > 0.5 with the calculated RS among individual samples among seven GEO datasets were selected. For the RNA-seq data, log2 transformation was performed and values of 0 to 15 were rescaled. Similarly, 97 genes with a mean Pearson’s correlation coefficient of > 0.5 with the calculated RS in TCGA and SNUH datasets were selected, all of which were included in the 135 genes from the microarray datasets. Next, fourteen of the 16 non-reference genes constituting the 21-gene assay that are not included in the set of 135 genes were added (5).

Reference genes were selected from the RNA-seq data in TCGA. The distant recurrence rate of the samples was considered for deriving the coefficient of variation for three groups of high-, intermediate-, and low-expression genes among 19,317 genes in RNA-seq data of 785 samples. Top 10 genes with the lowest variations in each group were selected to comprise 30 reference genes in total. Finally, a panel of 179 genes was designed for analysis.

***Sample preparation and targeted RNA-seq***

Formalin‑fixed paraffin‑embedded (FFPE) specimens from patients with early‑stage estrogen receptor (ER)‑positive, human epidermal growth factor receptor 2 (HER2)-negative, lymph node‑negative breast cancer were used for RNA extraction. The FFPE blocks and paired hematoxylin and eosin (H&E)-stained slides were obtained from the Department of Pathology of each hospital. The pathologists observed the H&E-stained slides and marked the invasive tumor portions on it. In total, 10 (10‑μm-thick) sections with at least 30% tumor fraction were deparaffinized with xylene with a heating block, followed by ethanol washing, and total RNA was subsequently isolated using the RNeasy® FFPE kit (QIAGEN GmbH, Hilden, Germany) in accordance with the manu­facturer's protocol, with DNase I treatment. The quantity, purity, and integrity of RNA were determined with the Qubit RNA HS Assay kit on QubitTM 4 Fluorometer, NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and RNA ScreenTape/Sample Buffer on 4200 TapeStation (Agilent Technologies, Santa Clara, CA, USA), respectively. To improve the sequencing performance, ≥ 90% ribosomal RNAs (rRNA) were depleted along with ≥ 1 μg of total RNA prior to library construction, and cDNA libraries were generated using the KAPA Stranded RNA-Seq Kit with RiboErase (HMR) (Kapa Biosystems, Wilmington, MA, USA) via the following steps: fragmentation for 1 min at 65˚C, 1st and 2nd strand cDNA synthesis, A-tailing, Adapter ligation, and dual-indexed library amplification. The eluted libraries were quantified with the Qubit dsDNA HS Assay kit (Thermo Fisher Scientific, Waltham, MA, USA) and the library size was evaluated using D1000 ScreenTape/Reagents (Agilent Technologies, Santa Clara, CA, USA) for further processing. Successful cDNA synthesis and amplification displayed a distinct peak ranging 200–700 bp, with an average size of 300–400 bp.

Hybridization of each library and biotynilated-179 RNA oligonucleotide probes was carried out using Target Capture Solution Box 1 and 3 (Celemics Inc. Seoul, South Korea) in accordance with the manufacturer's protocols. Target capture probes encompassing all coding exon regions were also generated by Celemics Inc. (Seoul, South Korea). Target captured libraries were selected using streptavidin-coated magnetic beads (Celemics Inc. Seoul, South Korea) using Target Capture Solution Box 2 (Celemics Inc. Seoul, South Korea). On-bead captured libraries were amplified via PCR using the KAPA HiFi HotStart ReadyMix (2X) and primer mix to the linked Illumina adapters (KAPABIOSYSTEMS) and subsequently, the amplified captured DNA libraries were finally purified using clean-up beads (Celemics Inc. Seoul, South Korea). For library validation for the sequencing run, each captured library was carefully quantified, and the average library size was determined to be 300–400 bp via the 4200 TapeStation instrument before all target captured samples were pooled. It is important to maintain a unique index for each sample during library preparation. Individual sample indices should not overlap. Upon accurate determination of the target size and the amount of data needed and the concentration of each sample, libraries were pooled to fit the amount of data output. Pooled libraries were diluted to a final concentration of 4 nM and denatured in accordance with the Illumina NextSeq System user guide. Paired-end targeted RNA-sequencing was performed using Mid Output Reagent Cartridge v2 and Mid Output Flow Cell Cartridge v2 on Illumina NextSeqTM 500 platform (Illumina, San Diego, CA, USA). Upon completion of the sequencing run, raw FASTQ files were analyzed for further study and the run performance was assessed. The raw reads were trimmed using Trimmomatic (version 0.33) (6) to eliminate Illumina adapter sequences. Thereafter, the trimmed reads were aligned to the human genome reference (hg19) with the STAR (version 2.5.2a, default parameters) (7), and read count quantifications were estimated using HTSeq-Count (version 0.11.0) (8). After read trimming, an average number of 2,445,983 reads per sample was obtained, and 93.91% were uniquely mapped to the reference genome. The mean sequencing depth and on-target ratio over the target regions were 354.65× and 86.95%, respectively.

***The conserved-exon panel (CEP) framework:***

1. Each RNA-seq data were processed to enumerate reads mapped into the CEP regions, where the CEP refers to the exonic regions included in more than a majority of the splicing isoforms expressed at the corresponding genetic loci (Figure 1). This approach aimed to exclude the minor exons potentially causing biases in samples with different isoform usages.
2. Considering the number of reads in CEP regions as the expression levels for each gene, scaling factors for each sample were determined using the trimmed mean of M-values (TMM) method. Thereafter, normalized expression levels for each gene were determined by multiplying the corresponding scaling factors. Consequently, a sample was generated through a gene matrix of normalized gene expression levels.



Figure 1. Conserved Exon Panel (CEP) approach

***RS prediction model construction using cross-validation***

 Cross-validation (CV) is a method frequently used to select genes and to discover hyper-parameters for optimization of models. We used the same 250 samples in the training set to perform a five-fold CV to develop a CV model. We first confirmed that there was no over-fitting in the original model by performing the five-fold CV. Among 21 genes with coefficients in the original model, 19 genes were included in the CV model. The two genes not included in the CV model were KIF14 and RRM2 with very small coefficient values of 0.0430 and 0.0391, respectively. The NGS-PS derived from the original model and the CV model showed a high correlation of r = 0.96 (Figure 2). In the verification set of 93 independent samples, Pearson’s correlation coefficient between RS and the original and CV models were 0.84 and 0.88, respectively, with no statistical difference (Figure 3). On setting the dichotomization cut-off as > 25 (high) vs ≤ 25 (low) for both NGS-PS and RS, the original model and the CV model showed a risk category concurrence to RS in 91.4% (85/93) and 90.3% (84/93) of samples, respectively (Figure 3). The predictive ability of the new CV model was not superior to the original model developed with Lasso. When a model is developed using CV, some portion of the data is contaminated when used for training, and thus we tried to guarantee the independent training and verification sets. Accordingly, we adopted the original model for the NGS-based multigene assay in this study.



Figure 2. Correlation between NGS-PS derived from original and five-fold CV models.



Figure 3. Pearson’s correlation coefficient between RS and the original and CV models in the verification set.

***Validity of the cut-off value assessed by two-fold cross validation.***

 The validity of the estimated cut-off value is assessed by first randomly segregating the dataset into two subsets. An optimal cut-off value is then estimated in each group. The cut-off value determined in subset 1 is used to categorize samples in subset 2, and the cut-off value determined in subset 2 is used to categorize samples in subset 1. Finally, samples in subset 1 and subset 2 are pooled for Cox regression analysis to determine whether groups categorized as low- and high-risk show different distant metastasis-free survival (DMFS) distribution according to the cut-off value. Two-fold cross validation showed statistical difference in DMFS distribution between low- and high-risk groups categorized in each subset, confirming the validity of the cut-off value in this cohort (Table 1).

|  |
| --- |
| **Table 1. Two-fold cross validation of cut-off value** |
| NGS-PS | Subset 1(N=207) | Subset 2(N=206) |
| Mean ± SD | 18.20 ± 7.79 | 19.54 ± 7.21 |
| Median (Min,Max) | 15.90 (0, 55.44) | 18.65 (6.93, 41.73) |
| Optimal cut-off value | 20.5 | 20 |
| Number of distant recurrence | 37 | 45 |
| Log-rank (p-value) | <0.0001 |
| NGS-PS: next-generation sequencing-Prognosis Score; SD: standard deviation. |

***Multivariate Cox proportional hazards model***

A multivariate Cox proportional hazards model including the NGS-PS, age, tumor size, histologic grade, and PR status was developed. Multivariate analysis revealed that NGS-PS was independently associated with distant recurrence in the model on adjusting for other variables with an HR of 5.15 (95% CI 3.01–8.80) and 4.99 (95% CI 2.60–9.58) in overall patients and in those without chemotherapy, respectively (both p < 0.001) (Table 2).

|  |
| --- |
| **Table 2. Multivariate Cox proportional analysis of age, tumor size, histologic grade, and PR status in predicting distant recurrence in validation cohort** |
| All patients (N=413) | 　 | 　 | 　 |
| Analysis without NGS-PS | 　 | Hazard ratio (95% CI) | P-value |
| 　 | Age | 1.00 (0.98-1.03) | 0.925 |
| 　 | Tumor size | 1.50 (1.31-1.73) | < 0.001 |
| 　 | Histologic grade | 2.64 (1.59-4.39) | < 0.001 |
| 　 | PR status | 0.63 (0.34-1.18) | 0.147 |
| Analysis with NGS-PS | 　 | 　 | 　 |
| 　 | Age | 1.00 (0.98-1.03) | 0.940 |
| 　 | Tumor size | 1.44 (1.25-1.66) | < 0.001 |
| 　 | Histologic grade | 1.72 (1.02-2.88) | 0.040 |
| 　 | PR status | 1.09 (0.58-2.05) | 0.792 |
| 　 | NGS-PS | 5.15 (3.01-8.80) | < 0.001 |
| No Chemotherapy (N=377) | 　 | 　 | 　 |
| Analysis without NGS-PS | 　 | Hazard ratio (95% CI) | P-value |
| 　 | Age | 1.04 (1.01-1.07) | 0.012 |
| 　 | Tumor size | 1.37 (1.01-1.86) | 0.042 |
| 　 | Histologic grade | 2.23 (1.06-4.69) | 0.035 |
| 　 | PR status | 0.78 (0.34-1.78) | 0.549 |
| Analysis with NGS-PS | 　 | 　 | 　 |
| 　 | Age | 1.03 (1.00-1.06) | 0.059 |
| 　 | Tumor size | 1.38 (0.98-1.95) | 0.068 |
| 　 | Histologic grade | 1.44 (0.68-3.07) | 0.339 |
| 　 | PR status | 1.35 (0.58-3.15) | 0.489 |
| 　 | NGS-PS | 4.99 (2.60-9.58) | < 0.001 |
| NGS-PS: Next-generation sequencing-Prognosis Score; PR: progesterone receptor |

***Comparison of integrated RS and clinical risk with NGS-PS in the training and verification sets.***

An exploratory analysis was performed using samples no metastasis in the training and verification sets. Among 280 patients with an integrated RS and clinical risk (CR), 33.6% (94/280) and 76.4% (186/280)) were categorized into integrated RS & CR high and low risk, respectively. Among the same 280 patients, 32.1% (90/280) and 77.9% (190/280) were categorized into high and low risk NGS-PS groups, respectively. Comparable proportion of patients in the risk categories determined by the integrated RS & PS and NGS-PS suggests that the NGS-PS would not result in a higher proportion of high risk patients compared to the integrated RS & CR categorization applied today for patients who is tested with Oncotype DX.

A Cohen’s kappa analysis was comparing the integrated RS & CR category to NGS-PS category based on cut-off value 20 showed a kappa of 0.497, indicating a moderate strength of agreement between the two risk stratification models. In addition, an ROC curve plotted using NGS-PS values classified as high vs low integrated RS & CR showed an AUC of 0.857, suggesting good accuracy of NGS-PS given that the Oncotype DX and integrated RS and clinical risk is being used widely for routine practice.

***References***

1. Györffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q*, et al.* An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat* **2010**;123:725–31.

2. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. *Nature* **2012**;490:61-70.

3. Lee J-H, Zhao X-M, Yoon I, Lee JY, Kwon NH, Wang Y-Y*, et al.* Integrative analysis of mutational and transcriptional profiles reveals driver mutations of metastatic breast cancers. *Cell Discov* **2016**;2:1-14.

4. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M*, et al.* A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* **2004**;351:2817–26.

5. Kim J, Kim A, Kim C. Examination of the Biomark assay as an alternative to Oncotype DX for defining chemotherapy benefit. *Oncol Lett* **2019**;17:1812-8.

6. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**;30:2114-20.

7. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S*, et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**;29:15-21.

8. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **2015**;31:166-9.