

## ARTICLE OPEN



# Proteogenomic analysis dissects early-onset breast cancer patients with prognostic relevance

Kyong-Ah Yoon<sup>1,12</sup>, Youngwook Kim<sup>2,12</sup>, So-Youn Jung<sup>2,3,12</sup>, Jin-Sun Ryu<sup>4,5</sup>, Kyung-Hee Kim<sup>2,6</sup>, Eun-Gyeong Lee<sup>3</sup>, Heejung Chae<sup>7,8</sup>, Youngmee Kwon<sup>3</sup>, Jaegil Kim<sup>9</sup>, Jong Bae Park<sup>10,13</sup> and Sun-Young Kong<sup>10,11,13</sup>✉

© The Author(s) 2024

Early-onset breast cancer is known for its aggressive clinical characteristics and high prevalence in East Asian countries, but a comprehensive understanding of its molecular features is still lacking. In this study, we conducted a proteogenomic analysis of 126 treatment-naïve primary tumor tissues obtained from Korean patients with young breast cancer (YBC) aged  $\leq 40$  years. By integrating genomic, transcriptomic, and proteomic data, we identified five distinct functional subgroups that accurately represented the clinical characteristics and biological behaviors of patients with YBC. Our integrated approach could be used to determine the proteogenomic status of HER2, enhancing its clinical significance and prognostic value. Furthermore, we present a proteome-based homologous recombination deficiency (HRD) analysis that has the potential to overcome the limitations of conventional genomic HRD tests, facilitating the identification of new patient groups requiring targeted HR deficiency treatments. Additionally, we demonstrated that protein–RNA correlations can be used to predict the late recurrence of hormone receptor-positive breast cancer. Within each molecular subtype of breast cancer, we identified functionally significant protein groups whose differential abundance was closely correlated with the clinical progression of breast cancer. Furthermore, we derived a recurrence predictive index capable of predicting late recurrence, specifically in luminal subtypes, which plays a crucial role in guiding decisions on treatment durations for YBC patients. These findings improve the stratification and clinical implications for patients with YBC by contributing to the optimal adjuvant treatment and duration for favorable clinical outcomes.

*Experimental & Molecular Medicine* (2024) 56:2382–2394; <https://doi.org/10.1038/s12276-024-01332-w>

## INTRODUCTION

Breast cancer is the most common malignancy among women globally<sup>1</sup> and is the leading cancer among women in Korea<sup>2</sup>. The occurrence of young breast cancer (YBC), i.e., breast cancer in patients aged 40 years or younger at diagnosis, is rare; however, its proportional incidence in Korea is approximately 10%<sup>3</sup>, which is twice the incidence rate in the US<sup>4</sup>. Patients with YBC have a genetic susceptibility caused by *BRCA1/2* pathogenic variants, showing aggressive clinical phenotypes with advanced stages, high-grade and triple-negative breast cancer (TNBC), and an unfavorable prognosis, including late recurrence<sup>5–8</sup>. However, the molecular features of YBC underlying these more aggressive tumor characteristics and worse clinical outcomes are unclear. Several studies have revealed the characteristics of YBC. With respect to somatic mutations, the Cancer Genome Atlas (TCGA) and Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) databases revealed that premenopausal (PreM) patients with breast cancer have greater numbers of mutations in 5 genes—*CDH1*, *GATA3*, *MLL3*, *GPS2*, and *PI3KCA*.

Moreover, compared to post-menopausal (PostM) tumors, gene expression in PreM tumors is enriched in the integrin and laminin signaling pathways, EGFR signaling activation, and TGF- $\beta$ , especially in estrogen receptor-positive (ER+) breast cancer<sup>9</sup>. Another study revealed that, compared with Western patients, Asian PreM patients presented a characteristic increase in tumor-infiltrating lymphocytes (TILs) and a decrease in TGF- $\beta$  signaling, which suggested that Asian YBCs harbor a more immune-active microenvironment<sup>10</sup>. An age-specific gene expression study of breast cancer in Middle Eastern women identified 63 genes that were specific to YBC ( $\leq 45$  years) and 2 genes (*TIAM1* and *VANGL2*) whose expression was significantly lower in breast tumors of very young ( $\leq 35$  years) women than in breast tumors from other age groups<sup>11</sup>.

In recent studies, the application of proteogenomic (PG) analysis has improved our understanding of breast cancer by providing comprehensive molecular signatures relevant to clinical features<sup>12–14</sup>. Stromal-enriched clusters, G protein-coupled receptor clusters, and several targetable biological pathways were also

<sup>1</sup>Department of Biochemistry, College of Veterinary Medicine, Konkuk University, Seoul, Korea. <sup>2</sup>Department of Cancer Biomedical Science, Graduate School of Cancer Science and Policy, National Cancer Center, Goyang, Korea. <sup>3</sup>Center for Breast Cancer, National Cancer Center, Goyang, Korea. <sup>4</sup>Division of Translational Science, Research Institute, National Cancer Center, Goyang, Korea. <sup>5</sup>Laboratory Animal Research Facility, Research Institute, National Cancer Center, Goyang, Korea. <sup>6</sup>Proteomics Core Facility, Research Core Center, Research Institute, National Cancer Center, Goyang, Korea. <sup>7</sup>Cancer Data Center, Control Institute, National Cancer Center, Goyang, Korea. <sup>8</sup>Division of Medical Oncology, Hospital, National Cancer Center, Goyang, Korea. <sup>9</sup>GlaxoSmithKline, Waltham, MA, USA. <sup>10</sup>Department of Laboratory Medicine, Research Institute, National Cancer Center Korea, Goyang, Korea. <sup>11</sup>Department of Targeted Therapy Branch, Research Institute, National Cancer Center, Goyang, Korea. <sup>12</sup>These authors contributed equally: Kyong-Ah Yoon, Youngwook Kim, So-Youn Jung. <sup>13</sup>These authors jointly supervised this work: Jong Bae Park, Sun-Young Kong. ✉email: jbp@ncc.re.kr; ksy@ncc.re.kr

Received: 24 November 2023 Revised: 5 July 2024 Accepted: 29 July 2024

Published online: 1 November 2024

identified. Although this approach revealed new clusters and refined subgroups with potential clinical benefits in breast cancer, there is a need to investigate the molecular landscape in young Asian patients.

Therefore, this study aimed to evaluate and analyze the molecular characteristics of YBC in Korean patients ( $\leq 40$  years) via integrated multiomics studies, including genomic, transcriptomic, and proteomic features, and to identify pathways for predicting the clinical outcomes of patients with YBC.

## MATERIALS AND METHODS

### Study objectives and specimens

A total of 178 participants with histologically defined breast cancer, aged 40 years or younger, treated at the National Cancer Center (NCC) in Korea were included in this study. Tumor and adjacent normal tissue samples were obtained from surgically resected specimens, and blood samples were collected from the patients. This study was approved by the Institutional Review Board of the NCC (IRB nos. NCCNCS 13717, NCC2017-0127, and NCC2020-0135). Since some patients had limited samples, only 126 patients for whom all proteogenomic data were fully generated were used for analysis. We retrospectively reviewed the medical and pathological records of the patients, including histological diagnoses of surgical specimens, tumor staging, histological grade, treatment history (type of surgery, use of chemotherapy, hormone therapy, anti-HER2 therapy, and radiotherapy), and follow-up data (recurrence, metastasis, and death).

Tissue specimens and blood samples were provided by the NCC Bio Bank of the National Cancer Center, Korea. The frozen tissue samples were weighed and washed with cold phosphate-buffered saline (PBS) to remove blood contamination, placed in a tube (Covaris, Woburn, MA, USA), snap-frozen in liquid nitrogen, and pulverized using a cryoPREP tissue disruption system (CP02; Covaris). The pulverized tissue powder was divided into 10–20 mg aliquots for DNA, RNA, and protein extraction. Genomic DNA and RNA were extracted from cryopreserved tissue samples and blood samples using the AllPrep DNA/RNA Mini Kit, the QIAamp DNA Blood Mini Kit, the DNeasy Blood & Tissue Kit, and the RNeasy Micro Kit (Qiagen, Valencia, CA, USA) according to the manufacturers' instructions. The concentration and integrity of the extracted RNA were assessed via a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

### Whole exome sequencing (WES) analysis

To generate standard exome capture libraries, the SureSelect Human All Exon V6 + UTR probe set (Agilent Technologies) was used according to the manufacturer's instructions. Briefly, genomic DNA extracted from the samples was fragmented and processed for adaptor ligation and PCR amplification. The exome capture libraries were constructed and amplified. The final purified products were quantified according to the qPCR Quantification Protocol Guide (KAPA Library Quantification kits for Illumina Sequencing platforms) and qualified using a TapeStation DNA Screentape D1000 (Agilent). The products were sequenced using the NovaSeq platform (Illumina Inc., San Diego, CA, USA). Quality-calibrated fastq files were aligned to the reference genome hg19 using BWA-MEM<sup>15</sup>. Each read group was aligned to the reference genome separately, and all read group alignments that belonged to a single aliquot were merged using Picard Tools, SortSam and MergeSamFiles. Duplicate reads were flagged to prevent downstream variant call errors. Local realignment of insertions and deletions was performed using GATK:IndelRealigner<sup>16</sup>. A base quality score recalibration (BQSR) step was then performed using GATK:BaseRecalibrator<sup>17</sup>. Somatic mutation detection was performed via MuTect and GATK:Mutect2<sup>18</sup>, and the mutations were annotated with an oncotator<sup>19</sup>. The GATK Haplotype caller was used to detect germline variants<sup>20,21</sup>.

### Whole-transcriptome sequencing (WTS)

Only high-quality RNA (RNA integrity  $\geq 7.0$ ) generated using the Illumina TruSeq Stranded mRNA Sample Prep Kit (Illumina) was used for RNA library construction. Qualified libraries were indexed, and paired-end 100-bp reads were sequenced using the Illumina NovaSeq platform. The quality of the sequencing reads was evaluated and examined for primer/adaptor sequence contamination via FastQC (v0.11.7)<sup>22</sup>. Trim Galore<sup>23</sup> was employed for trimming the reads, applying a threshold of an average sequence quality  $> 30$ . The trimmed reads were subsequently aligned to

the human reference genome using STAR (STAR-2.7.0). Gene expression profiling was conducted with the HTseq-count algorithm, which uses the raw read counts as input data. Finally, edgeR was used to convert the count data to reads per kilobase million (RPKM), or the transcript per million (TPM) algorithm from the bioinfokit Python package (v1.0.4)<sup>24</sup> was used to adjust different sequencing throughputs between the samples.

### Proteomic analysis

Proteomic analysis was performed according to the standard protocol reported by the Clinical Proteomic Tumor Analysis Consortium (CPTAC)<sup>25</sup>. Briefly, cryopreserved tissue powder samples were solubilized in sodium dodecyl sulfate (SDS) solubilization buffer (5% SDS, 50 mM TEAB; pH 8.5) using an S220 focused ultrasonicator (Covaris). Proteins were digested using S-Trap™ spin columns (ProtiFi, Huntington, NY, USA), cleaned with C18 spin columns (Thermo Fisher Scientific, Rockford, IL, USA), and the desalted peptide samples were labeled using TMT11plex reagents (Thermo Fisher Scientific). Pooled 11-plex tandem mass tag (TMT)-labeled samples were processed for peptide fractionation using an Agilent 1260 Infinity HPLC system (Agilent)<sup>26</sup> and then prepared for global proteome and phosphoproteome analyses. Phosphopeptide enrichment was performed using immobilized metal affinity chromatography<sup>27</sup>. The TMT-labeled peptides were loaded onto a trap column (Acclaim PepMap™ 100, 75 mm  $\times$  2 cm), separated on an analytical column (EASY-Spray column, 75 mm  $\times$  50 cm; Thermo Fisher Scientific) via an Ultimate 3000 RSLCnano system (Thermo Fisher Scientific), and analyzed using the top 10 data-dependent methods using a Q Exactive HF-X hybrid quadrupole-orbitrap mass spectrometer (Thermo Fisher Scientific).

### Cell lines and reagents

Human breast cancer cell lines were cultured according to the distributors' instructions. We obtained breast cancer cell lines from the American Type Culture Collection (ATCC) in Manassas, Virginia, USA; the Korean Cell Line Bank in Seoul, Korea; and the DSMZ in Braunschweig, Germany. MDA-MD-453, BT-474, and HCC-1954 cells were maintained in RPMI 1640 medium supplemented with 10% fetal bovine serum (FBS). SKBR3 cells were cultured in McCoy's 5A medium supplemented with 10% FBS. JIMT-1 cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS. Penicillin/streptomycin (1%) was added to all of the culture media, and the cells were cultured at 37 °C with 5% CO<sub>2</sub>.

### Droplet digital PCR (ddPCR)

The copy number of the *HER2* gene was analyzed using droplet digital PCR (ddPCR) on a QX200 Droplet Digital PCR System by Bio-Rad Laboratories. Fluorescent probes (FAM and HEX) were prepared from the PrimePCR™ ddPCR™ Copy Number Assay for ddPCR (dHsaCP1000116 for *HER2*, dHsaCP2500349 for *EIF2C1*, and dHsaCNS516206038 for *POLR2A* as the reference control) (Bio-Rad Laboratories Inc., Hercules, CA, USA). A total of 20  $\mu$ l of PCR mixture was prepared with 10 ng of genomic DNA using 2X ddPCR supermix for the probe (10  $\mu$ l) and 20X *HER2* and *EIF2C1* or *POLR2A* probes (FAM/HEX) (1  $\mu$ l). PCRs were run on a Mastercycler nexus gradient Thermal Cycler (Eppendorf, Hamburg, Germany) at 95 °C for 10 min, followed by 40 cycles of 94 °C for 30 s, 55 °C for 60 s, and a 10 min incubation at 98 °C. Then, the PCR plates were read with a Bio-Rad QX200 droplet reader (Bio-Rad Laboratories, Inc.) and QuantaSoft™ version 1.4.0 software (Bio-Rad Laboratories, Inc.) to assess the number of droplets positive for *HER2* and *EIF2C1* or *POLR2A*.

### Viability assay

Breast cancer cell lines were seeded in triplicate at 2–4  $\times 10^3$  cells per well, depending on the cell line, in 96-well plates. We also treated breast cancer cell lines with neratinib (Selleckchem, Houston, TX, USA) or lapatinib (Selleckchem). After 48 h of treatment with the test compounds, cell viability was examined using a Cell Counting Kit-8 (CCK-8, Dojindo Laboratories, Kumamoto, Japan). Dimethyl sulfoxide (DMSO) (0.1%) was used as the control. The half-maximal inhibitory concentration (IC<sub>50</sub>) was analyzed using GraphPad Prism (version 5.03, GraphPad Software, San Diego, CA, USA).

### Western blot

Total protein was extracted from cell lines or tissues using RIPA lysis buffer (Thermo Fisher Scientific) with protease and phosphatase inhibitor cocktails (GenDEPOT, Baker, TX, USA) following the manufacturer's

instructions. The protein concentrations were determined via a protein assay solution from Bio-Rad Laboratories. Proteins were separated via sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE), and the separated proteins were transferred to polyvinylidene fluoride (PVDF) membranes (Merck Millipore, Billerica, MA, USA). The blots were then blocked for 1 h with 5% skim milk (BD Biosciences, San Diego, CA, USA) in Tris-buffered saline containing 0.1% Tween 20 and incubated overnight with the indicated primary antibodies. Antibodies against the following proteins were used: phospho-HER2-Tyr1248 (Cell Signaling Technology, CST), HER2 (Abcam, Cambridge, UK), phospho-HER2-Ser1054 (Thermo Fisher Scientific), and  $\beta$ -actin (CST, used as the loading control). Proteins were visualized via a horseradish peroxidase-conjugated secondary antibody (CST) and an enhanced chemiluminescence (ECL) reagent (Bio-Rad Laboratories).

### Multiplex immunofluorescence staining

Multiplex immunofluorescence staining for cytokeratin (CK), PD-L1, CD68, CD8, FOXP-3, and PD-1 proteins was performed using a Leica Bond Rx Automated Stainer (Leica Biosystems, Newcastle, UK). Slides with tumor tissue sections were incubated with antibodies against the following proteins: CK (NBP2-29429; Novus Bio; dilution 1:300), PD-L1 (136845; CST; dilution 1:300), CD68 (ab192847; Abcam; dilution 1:300), CD8 (MCA1817; Bio-Rad; dilution 1:300), FOXP3 (ab20034; Abcam; dilution 1:100), and PD-1 (ab137132; Abcam; dilution 1:500). Then, the sections were incubated with the Polymer HRP Ms+Rb (ARH1001EA, AKOYA Biosciences) and subsequently treated with an Opal fluorophore (Opal690, Opal570, Opal520, Opal620, Opal480, or Opal780). Nuclei were subsequently visualized with DAPI, and the sections were coverslipped with ProLong Gold antifade reagent (P36934, Invitrogen). The slides were scanned using the Vectra Polaris Automated Quantitative Pathology Imaging System (Akoya Biosciences, Marlborough, MA, USA), and regions of interest were selected for all available tumor-containing areas. Images were analyzed using Inform 2.5 software (Akoya Biosciences, Marlborough, MA, USA) and TIBCO Spotfire™ (TIBCO, Palo Alto, CA, USA).

### Homologous recombination deficiency (HRD) score

The HRD score was determined as a simple sum of three factors: NtAI (number of telomeric allelic imbalances), LST (large-scale state transitions), and HRD-LOH (HRD loss of heterozygosity). Tumors with HRD scores  $\geq 42$  were defined as HRD-high. For the determination of HRD scores via WES, BAM files of tumor samples were applied to the Sequenza followed by the scarHRD and/or HRDetect package using the default parameters as previously described<sup>28–30</sup>.

### Differentially Expressed Gene (DEG) analyses

DEG analyses were conducted using edgeR<sup>31</sup>, Cuffdiff2<sup>32</sup>, and DESeq2<sup>33</sup> following previously published protocols with default parameters. edgeR employed generalized linear models with tagwise dispersion, and raw counts were used as the primary input. As Cuffdiff2 does not work with count matrices, aligned transcriptomic reads were assembled into transcripts using Cufflinks with quartile normalization, bias correction, multi-read correction, and a reference gene model. Cuffdiff2 identified DEGs via geometric library normalization and per-condition dispersion estimation. Adjusted *p* values were calculated with Benjamini–Hochberg false discovery correction (5%) for all methods.

### Significance of mutation

The MutSigCV algorithm (v.1.3.5)<sup>34</sup> was employed to discern significantly mutated genes in our cohort. The default parameters of three inputs, coverage table files, covariate table files and mutation type dictionaries, were used to run MutSigCV, which uses a statistical framework to evaluate the observed mutation patterns against background mutation rates while considering covariates, including sequence context and replication timing.

### Integrative data analysis

Subtypes were identified on the basis of individual types of data (mRNA sequencing data, whole proteome, phosphoproteome) expressed in at least 70% of patient samples. Median absolute deviations (MADs) were calculated for the individual data, and the top 10% of the most variably expressed genes were subjected to subsequent statistical analyses. On the basis of the MAD values, we then selected the molecules with the

top 10% of MADs: 1,468 genes, 723 proteins, and 128 phosphopeptides. Next, we applied network integration based on consensus clustering to conduct unsupervised classification and to identify proteogenomic features that represent characteristic expression patterns for each cluster. The three data types were aggregated to generate an integrated network and similarity matrix, followed by the designation of consensus clusters. The criteria for determining the number of clusters were as follows: relatively high consistency within clusters, a relatively low variation coefficient, and no appreciable rise in the area under the cumulative distribution function (CDF) curve. The samples most representative of the clusters were identified on the basis of their positive silhouette width.

To determine the representation of intrinsic subtypes of our cohort, the PAM50 model was applied to TPM-normalized RNA-seq data using the “genefu” R package (v3.12)<sup>35</sup>.

Single-sample gene set enrichment analysis (ssGSEA) was performed for each sample using log<sub>2</sub>-transformed and quantile-normalized RNA-seq data and the global proteome for the molecular signatures database (MSigDB) GeneSets<sup>36</sup> via GenePattern 2.0<sup>37</sup> and Gene Ontology (GO) biological process signatures. The gene set scores represent the enrichment scores obtained from ssGSEA. To generate the receiver operating characteristic (ROC) curve and calculate the area under the curve (AUC) from the ssGSEA results and binary phenotype classification, we used ssGSEA\_ROC from GenePattern 2.0. Statistical significance was determined using a false discovery rate (FDR) of less than 0.01. To obtain individual enrichment scores for each pairing of a sample and a gene set, we used the ssGSEA projection on the GenePattern server using *t* test statistics with a *p*-value less than 0.001.

### Immune profiling and downstream analysis

To calculate RNA-based tumor immune scores and estimate immune-cell-specific contributions to each tumor, TPM data were analyzed using ESTIMATE<sup>38</sup>, CIBERSORT in absolute mode<sup>39</sup>, xCell<sup>40</sup>, and the MCP counter<sup>41</sup>. We also inferred the immune cell infiltration by ssGSEA using a published immune gene signature<sup>42</sup>. Protein-based immune scores for stimulatory and inhibitory immune modulators, as well as the set of HLA proteins, were calculated as the mean of the protein log ratios in each set defined in Thorsson et al.<sup>43</sup>.

### Kinase phosphorylation outliers

To identify the kinase activity characteristic of each PAM50 and major integrative clusters (iClusters), we used BlackSheep’s differential extremes value analysis module<sup>44</sup>. For each phosphosite, the median and interquartile range (IQR) were calculated across all tumors. A site was defined as an outlier if it was more than 1.5 times the IQR above the median. The phosphosites were then collapsed into proteins by counting outlier and nonoutlier values per sample. For each group of interest (e.g., PAM50 or iCluster), proteins that were not enriched in outliers in that group and proteins without at least 30% of samples with an outlier were removed. Following filtering, outlier and nonoutlier sites per gene were counted for each group of interest, and Fisher’s exact test was used to calculate a *p*-value with multiple hypothesis correction via the Benjamini–Hochberg (BH) procedure. For additional insight into kinase activity, we visualized the enrichment of kinase activation loop phosphorylation, which was calculated using a rank sum test.

### Kinase-substrate enrichment analysis (KSEA)

KSEA was performed using the KSEA App web-based tool<sup>45</sup> with phosphosite data with a cutoff of *p* < 0.05 and a substrate count of  $\geq 1$ . Student’s *t*-test was used to calculate the *p* value, and the mean ratio was used to determine the fold change compared with the other four subsets. KSEA computes a normalized score to measure relative kinase activity in cluster 1 versus others using the difference in the mean log fold changes of predefined substrate groups of the kinase, as referenced from the PhosphoSitePlus<sup>46</sup> and NetworkKIN databases<sup>47</sup>. In the latter database, kinase–substrate annotations with a NetworkKIN score of less than five are discarded. The mean log fold changes of all substrates were divided by the standard deviation of log fold changes across all phosphosites, followed by an adjustment to the score for the number of substrates in the data that were annotated to the kinase. The statistical significance of the scores was calculated using a one-tailed *z* test, followed by BH multiple testing correction. Kinases with *z* scores and corresponding *p* values less than 0.05 were considered significant.

## RESULTS

### Patient Cohort Description

The baseline characteristics of the 126 patients with YBC are summarized in Supplementary Table 1. The median patient age was 37 years (range: 25–40). Seventy-six patients (60.3%) had stage II disease, and 64 patients (50.8%) had node metastasis at the time of diagnosis. Sixty-five patients (51.6%) had a histological grade of 3. ER was positive in 84 patients (66.7%), PR was positive in 73 patients (57.9%), and HER2 was positive in 25 patients (19.8%). Based on the clinical classification of breast cancer with ER, PR, HER2, and Ki-67<sup>48</sup> statuses, 33 patients were classified as luminal A type (26.2%), 55 as luminal B type (43.7%), 10 as HER2 type (7.9%), and 25 as TNBC type (19.8%). All patients were diagnosed with invasive ductal carcinoma. Overall, 113 patients (89.7%) underwent systemic chemotherapy, 104 patients (82.5%) underwent adjuvant radiotherapy, and 89 patients (70.6%) underwent adjuvant hormone therapy. The median follow-up period was 94.3 months (range: 4.0–254.3), 37 patients (29.4%) experienced recurrence, and 17 (13.5%) died.

All of the patients in this study ( $n = 126$ , 100%) were initially diagnosed under the age of 40 and premenopausal, and TCGA data largely consists of patients with breast cancer over the age of 40 ( $n = 1,998$ , 91.1%). Compared with the TCGA cohort ( $n = 1116$ ), our patients demonstrated a greater proportion of the basal subtype (28.6% KNCC-YBC vs. 16.5% TCGA) and HER2-enriched subtype (15.9% vs. 6.3%). This difference was maintained when compared to young patients (age  $\leq 40$ ) in the TCGA cohort (basal type 17.7%, HER2 type 4.4% in TCGA\_YBC [ $n = 158$ ]). This finding is comparable to the findings of a previous study in which Asian patients with breast cancer had a greater prevalence of HER2-positive cancers than Caucasian women did<sup>10,49</sup>.

### Genomic landscape of early-onset breast cancer

We performed an integrated analysis of molecular data from patients through WES, WTS, and mass spectrometry-based proteomics, which provided robust quantification of proteins and phosphoproteins (Fig. 1a). Whole-exome sequencing of 126 samples revealed 30,193 somatic mutations in the study cohort. Each patient harbored a median of 119 somatic mutations per sample (range: 2–2858 somatic mutations). Significance analysis of the somatic mutations, employing covariates related to the mutational processes, identified the key mutated genes, including *TP53*, *PIK3CA*, *GATA3*, *MAP2K4*, and *ARID1A* (Fig. 1b). *TP53* (60%), *PIK3CA* (43%), and *GATA3* (36%) were the most frequently mutated genes in this cohort. *GATA3* somatic mutations were mostly mutually exclusive with *TP53* mutations (Fisher's exact test,  $P < 0.01$ ), with *PIK3CA* mutations exhibiting similar mutual exclusivity patterns ( $P < 0.01$ ). Overall, approximately 68% of the patients possessed at least one mutation in one of these three genes. The prevalence of somatic mutations in *MAP2K4* and *ARID1A* in breast cancer has also been documented in other studies<sup>50–53</sup>, indicating the presence of a congruent mutational profile within the early-onset Asian breast cancer demographic.

### Molecular subtype classification on the basis of multiomic profiles

In addition to classical pathological and clinical laboratory-based stratification of the molecular subtypes of the cohorts using hormone receptor and HER2 immunohistochemistry (IHC) analyses, we attempted a robust molecular classification of YBC cohorts based on a multiomics dataset. Unsupervised clustering of the three data platforms (proteome, transcriptome, and phosphoproteome) resulted in discordant stratification unique to each molecular layer. A consensus integrative classification encompassing the three molecular data layers produced five iClusters within the young Korean breast cancer cohort (Fig. 1c).

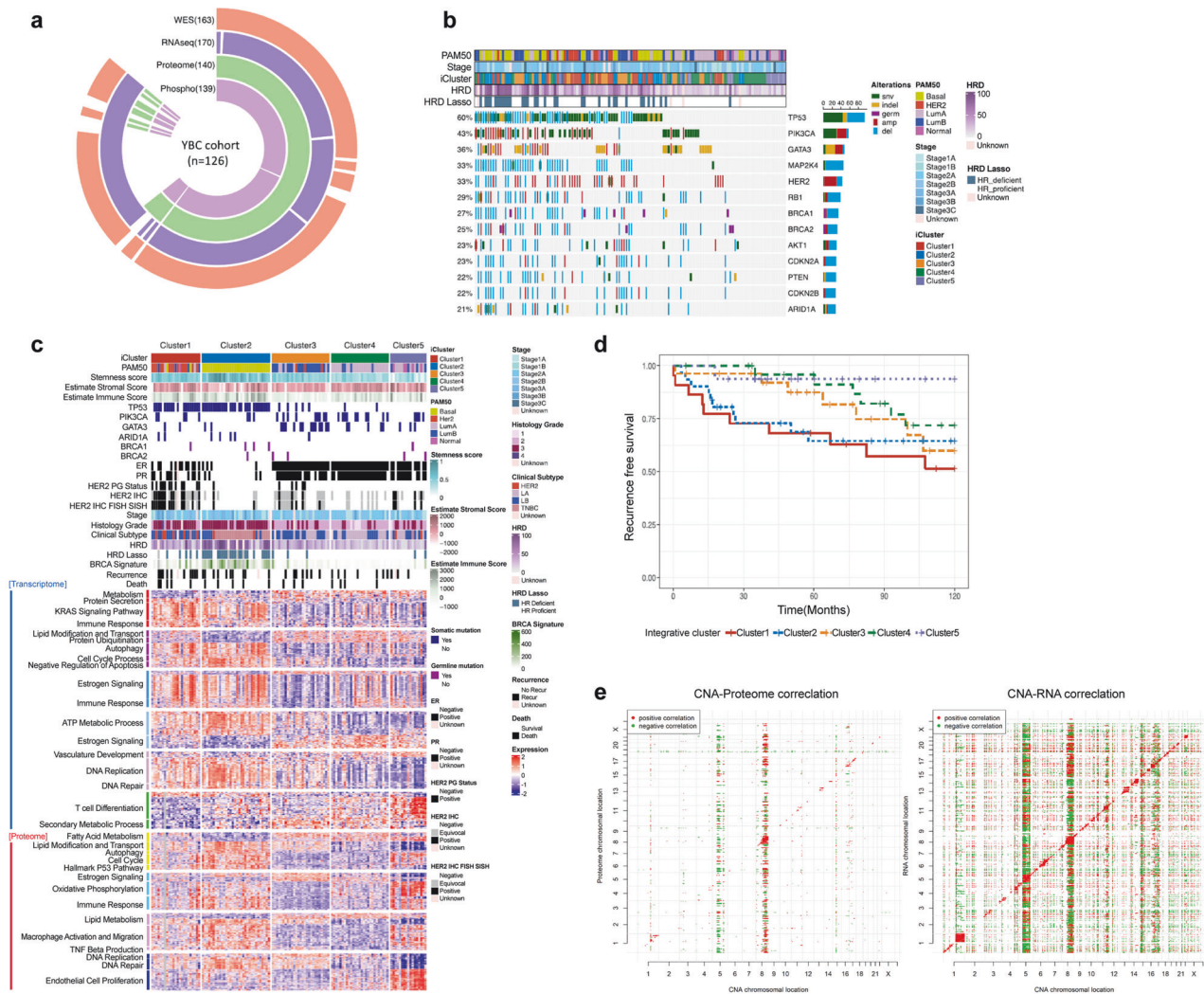
Cluster 1 was significantly concordant with the classical PAM50-based HER2-enriched subtype. This cluster was characterized by a

high prevalence of *HER2* amplification and *TP53* mutations (69.6% and 82.6% in cluster 1, respectively). The key pathways in this cluster included protein secretion, DNA replication, and repair. Cluster 2 was almost invariably consistent with the basal-like cancer type and presented a high frequency of *TP53* mutations and an absence of hormone receptor and *HER2* mutations. Cluster 2 was associated with the activation of pathways such as *KRAS* signaling, autophagy, and cell cycle processes and the negative regulation of apoptosis. Clusters 1 and 2 demonstrated the worst prognosis, with recurrence and metastasis events frequently observed within 5 years after surgical resection (Fig. 1d). Cluster 3 largely overlapped with the classical luminal B subtype, whereas cluster 4 demonstrated closer alignment with the classical luminal A subtype. The important pathways of cluster 3 and cluster 4 included estrogen signaling, the cell cycle, DNA replication and repair, and fatty acid metabolism (Fig. 1c). Those clusters of young Korean patients with breast cancer experienced more frequent recurrence and/or death events 5 or more years post-diagnosis of breast cancer (Fig. 1d). All cases of late relapse in these clusters were ipsilateral (11/11,  $P < 0.00001$ , Fisher's exact test); thus, the late recurrent tumors seemed to have originated from residual and/or resistant tumor cells left from their primary tumors after a sustained period of dormancy rather than from the manifestation of a second primary tumor. In contrast, basal-type and *HER2*-type Korean YBC demonstrated completely opposite biphasic patterns of survival probability; these groups' relapse and consequent death events were more clustered in the first 5 years, followed by a plateau in survival probability. The dichotomy of biphasic survival probability between luminal and nonluminal Korean patients with breast cancer is substantially distinct from the clinical course of Caucasian breast cancer, where luminal A breast cancer shows the most indolent phenotype, followed by luminal B breast cancer and *HER2* breast cancer, with the basal group having the highest risk patterns<sup>54</sup>. The results of the integrated analysis revealed a distinct subtype, cluster 5, which had a very good prognosis, unlike the other four clusters (Fig. 1d). The significant pathways of cluster 5 included vasculature development, endothelial cell proliferation, and oxidative phosphorylation.

To examine the impact of somatic copy number alterations (SCNAs) on the levels of mRNAs and proteins, we conducted univariate correlation analysis to identify features that exhibited statistically significant correlations (Fig. 1e). This analysis considered changes in 'cis'—affecting genes within the altered loci—and 'trans'—impacting genes located elsewhere. We evaluated the genes in this cohort, each with corresponding SCNA, mRNA, and protein data, by calculating the correlation and its statistical significance for all CNA–mRNA and CNA–protein pairings. Notably, 61% of the CNA–mRNA pairs and 30% of the CNA–protein pairs demonstrated significant positive 'cis' correlations. We further scrutinized the subset of these pairs that included well-characterized oncogenes and tumor suppressor genes. 'Trans' effects appeared as prominent vertical bands in the analysis, with frequency histograms highlighting significant 'hot spots' of 'trans' correlation. Interestingly, correlations between CNA and proteins were generally weaker than those observed for CNA and mRNAs. The most pronounced 'trans' relationships at the protein level were identified in specific CNA regions, including regions of chromosome 5 and chromosome 8, in early-onset Korean breast cancer.

### Targetable elements associated with clusters and genomic alterations

Next, we examined the functional impact of genetic alterations on the abundance of both *cis*-acting (cognate gene products) and *trans*-acting (other gene products) proteins and their phosphorylation sites (Fig. 2a, b). Several *cis* and *trans* associations were discernible. These include enhanced expression of

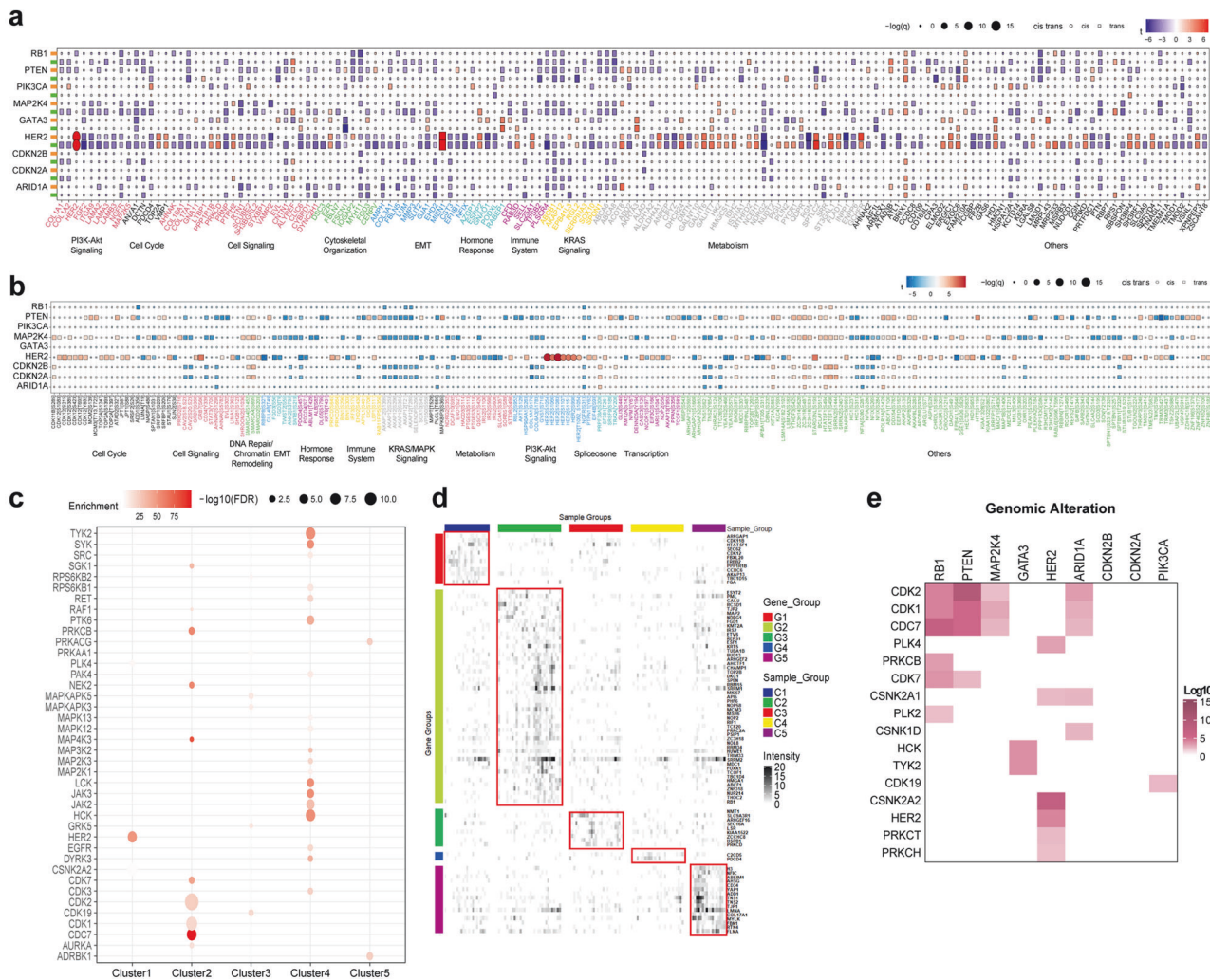


**Fig. 1 Molecular portraits of early-onset breast cancer.** **a** Data structure outlining the molecular profiles of 126 patients with early-onset breast cancer, categorized according to each type of molecular data generated in the current study. WES ( $n = 163$ ), RNA-seq ( $n = 170$ ), proteome ( $n = 140$ ), and phosphoproteome ( $n = 139$ ) data were generated for 178 young- breast cancer patients. Among these patients, multiomic data were generated for 126 individuals. **b** Genomic landscape of major breast cancer driver genes on the basis of mutational and somatic copy number status. Genomic alterations are color-coded in accordance with the type of mutation. The bar graph on the right denotes the total count of each specific genomic alteration observed. PAM50: PAM50 breast cancer subtypes; Stage: Pathological stage of breast cancer; iCluster: 5 integrative clusters based on proteogenomic analysis of the current study cohort; HRD: the level of homologous recombination deficiency on the basis of next-generation sequencing data; HRD LASSO: classification of HRD status from next-generation sequencing data; **c** Heatmap illustrating supervised clustering of differentially expressed genes (upper panel) and proteins (lower panel) across integrative molecular clusters (Kruskal–Wallis test,  $FDR p < 5 \times 10^{-5}$ ). The pathways enriched by the differentially expressed genes and proteins are annotated on the left. **d** Kaplan–Meier curves showing the progression-free survival outcomes of patients in the integrative clusters in the dataset. Stratification of patients on the basis of comprehensive multiomics data yielded five distinct molecular clusters, each associated with a different clinical outcome. Clusters 1 (red), 2 (blue), 3 (yellow), 4 (green), and 5 (violet) are shown. **e** Correlations of somatic copy number alterations (SCNAs, x-axis) with mRNA (left) and protein (right) abundances (y-axis).

HER2 proteins (Fig. 2a) and strong *cis*-phosphorylation (Fig. 2b) in HER2-positive breast cancers, cementing their role as potent drivers of alterations in a subset of breast cancers. Migration and invasion enhancer 1 (*MIEN1*) and growth factor receptor-bound protein 7 (*GRB7*) are genes located proximal to the *ERBB2* gene and are concomitantly overexpressed in HER2-positive breast cancer subgroups (Fig. 2a, b). *MIEN1* has been shown to functionally increase the invasive and migratory phenotypes of various types of cells in breast cancer<sup>55</sup>. Additionally, *GRB7* overexpression facilitates the phosphorylation of both AKT and HER2/neu in HER2/neu-overexpressing cells, promoting tumor growth *in vivo*<sup>56</sup>. These findings suggest additional roles for these proximal genes in breast cancer. Several phosphorylation sites associated with common breast cancer oncogenes were

identified and clustered in the cell cycle and DNA repair activities, which is consistent with their established roles in cancer.

We investigated which kinase activities were enriched in the molecular subgroups defined in this study (Fig. 2c). As expected, HER2 kinase activity was upregulated in cluster 1, which was highly correlated with the classical PAM50 Her2 subgroup. Cluster 2 (basal subgroup) was enriched with cell cycle kinases, including *CDC7*, *CDK1*, and *CDK2*. Hormone receptor-positive cancer groups (clusters 3–4) presented increased activity of kinases involved in immune regulation. To identify potential therapeutic targets specific to each iCluster subtype, we used phosphoproteomic data to determine kinase activation (Fig. 2d). We identified phosphorylated kinases that were enriched in each iCluster



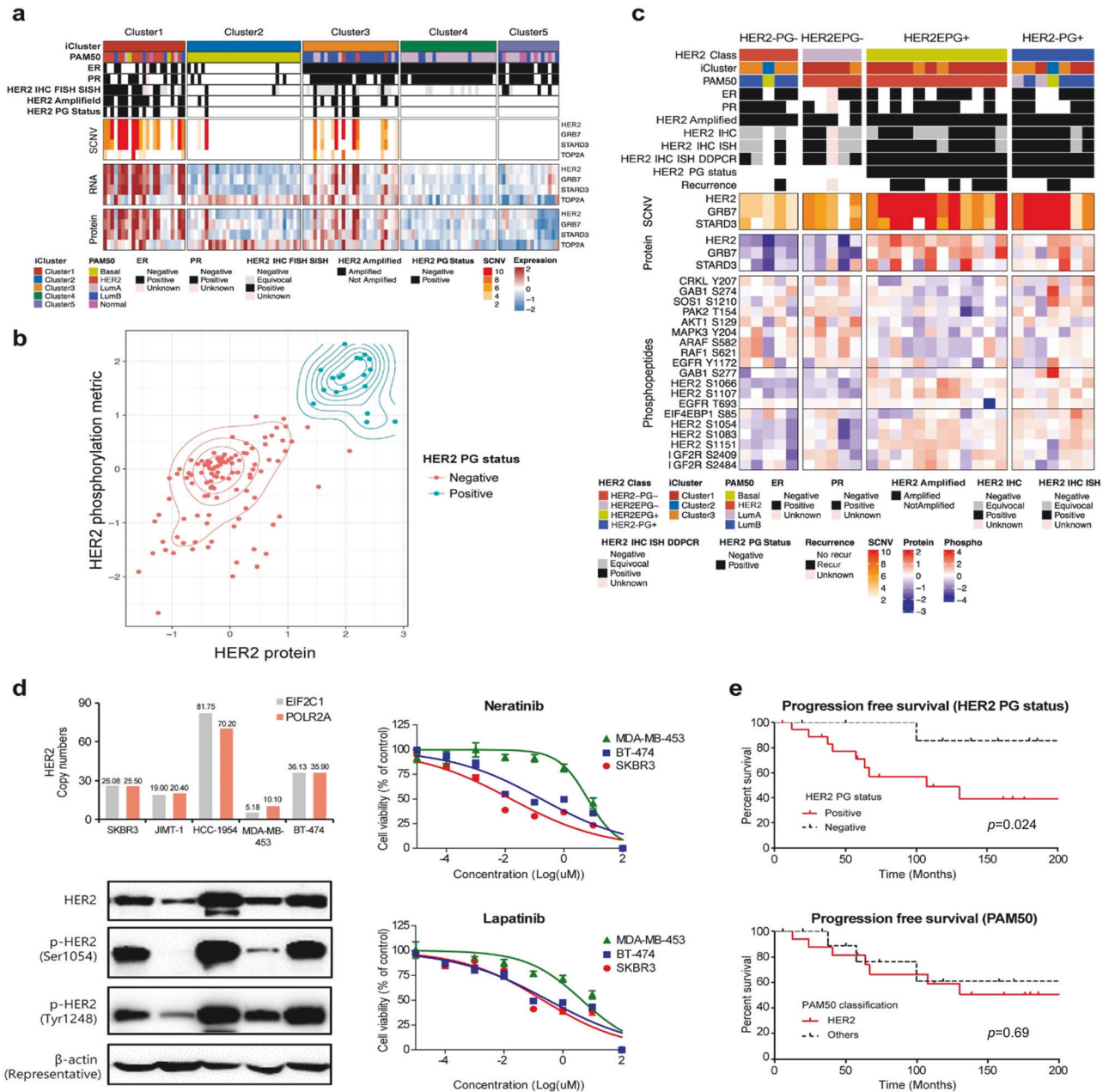
**Fig. 2 Targetable elements associated with clusters and genomic alterations from proteogenomic analysis of early-onset East Asian breast cancer.** *Cis*- and *trans*-effects of major genomic alterations, including mutations and copy number variations, on protein (a) and phosphoprotein (b) levels. The *cis* and *trans* effects of each genomic alteration can be categorized into major cancer hallmark pathways. c Plot depicting kinases that exhibited preferential substrate phosphorylation within each integrated molecular cluster. d Heatmap showing the fraction of outlier values in each sample per protein. The proteins shown are kinases highly phosphorylated in each cluster, with an FDR of less than 0.01 according to BlackSheep. The top panel shows the classifier based on the five major integrative clusters (iClusters). The right panels depict the abundance of the kinase activation loop and kinase substrate enrichment. e Heatmap showing q values from kinase substrate enrichment analysis for enrichment of phosphorylation outliers (y-axis) in samples with the indicated mutated gene (x-axis). Kinases with an FDR of less than 0.01 are shown.

subtype via outlier enrichment analysis. The dataset revealed several druggable targets and kinases, including ERBB2 (trastuzumab), in the cluster 1 subtype. The cluster 2 subtype includes isolated PML (tretinoin) and RCSD1 (imatinib). PDCC4 (paclitaxel) was found in cluster 4, whereas CD34 (prednisolone) and TJP1 (dexamethasone) were present in cluster 5.

These proteins and kinases are potential targets for subtype-specific treatment. We also investigated the associations between specific genomic alterations and kinase activity (Fig. 2e). We identified 16 kinases significantly associated with genomic alterations. *RB1*, *PTEN*, and *MAP2K4* mutations formed a strong subcluster with increased activity of several canonical cell cycle kinases, including the CDK1, CDK2, CDK7, and CDC7 proteins. These associations were also detected in *ARID1A*-mutant breast cancers. The *PLK4*, *CSNK2A2*, *HER2*, *PRKCT*, and *PRKCH* proteins are associated with *HER2* genomic alterations. *GATA3* genomic alteration was associated with increased activity of the HCK and *TYK2* kinases, whereas *PIK3CA* genomic alteration was linked with the *CDK19* kinase.

### Proteogenomic analysis of HER2-type YBC

We assessed the *HER2* driver status by integrating genomic and proteomic data to refine its classification and clinical significance. *HER2* status was also grouped by classical IHC and in situ hybridization testing following ASCO-CAP guidelines (*HER2*+: IHC score of 3+ or an IHC score of 2+ fluorescence in situ hybridization (FISH); silver-enhanced in situ hybridization (SISH); *HER2* equivocal status: IHC score of 2+ without FISH results or amplification by FISH without IHC results)<sup>57</sup>. Proteogenomic analysis revealed that 19 patients who presented high levels of both *HER2* protein and phosphorylated *HER2* were designated as *HER2* proteogenomic status positive (*HER2* PG+) (Fig. 3a). *HER2* PG+ cases were enriched mostly in cluster 1, followed by cluster 3. The majority of *HER2* PG+ samples (16 out of 19) demonstrated largely consistent pathological designations with confirmed *HER2*-positive expression according to IHC, FISH, and SISH tests. The remaining three patients who were grouped according to *HER2* equivocal status demonstrated a high *HER2* copy number as a result of droplet digital PCR analysis. Although we identified 28



**Fig. 3 Proteogenomic classification of HER2 status in YBC. a** Proteogenomic analysis of the HER2 locus in the current cohort. The heatmap displays the clinical and molecular data (top panel), SCNA data (center upper panel), RNA expression data (center lower panel), and protein expression data (bottom panel) of genes located near HER2 on chromosome 17q in the corresponding samples. HER2 IHC FISH SISH: pathological index of HER2 status from immunohistochemistry, fluorescence in situ hybridization or silver-enhanced in situ hybridization, HER2 PG Status: Her2 proteogenomic status. **b** Proteogenomic classification of the HER2 breast cancer subtype. Proteogenomic status of HER2 combined with HER2 protein levels (x-axis) and HER2 phosphorylation status (y-axis). HER2 proteogenomic status was stratified and depicted as either blue (positive) or red (negative). **c** Phosphopeptide levels of components of the HER2 signaling pathway according to the refined classification of HER2. The top panel of the heatmap outlines the subtype classifications and clinical marker status for each sample, whereas the center panels denote SCNAs and protein levels for genes in the amplicon closely associated with HER2, followed by the corresponding protein levels. The bottom panel illustrates the abundances of phosphopeptides, such as serine residues 1066, 1107, 1054, 1083, and 1151 from the HER2 pathway. **d** HER2 protein levels and drug response to HER2 inhibitors in breast cancer cell lines. (Left upper) Distribution of HER2 copy number in HER2-positive cell lines. The HER2 copy number was measured in HER2-positive breast cancer cell lines via droplet-digital PCR (ddPCR). EIF2C1 and POLR2A were used as reference genes to determine the ratios of HER2 to EIF2C1 or POLR2A. (Left lower) Western blot analysis was conducted to assess the expression and phosphorylation status of HER2 in HER2-positive breast cancer cell lines. The HER2-positive breast cancer cell lines SKBR3, JIMT-1, HCC-1954, MDA-MB-453, and BT-474 were separated via SDS-PAGE and immunoblotted. Western blotting was performed for total HER2, p-HER2 (Ser1054, Tyr1248), and  $\beta$ -actin. (Right) Drug sensitivity test of HER2 inhibitors (neratinib, lapatinib) in breast cancer cells phosphorylated at serine 1054. MDA-MB-453, BT-474, and SKBR3 cells were treated with neratinib or lapatinib for 48 h. Data are presented as the mean  $\pm$  SEM. Drug sensitivity assays were performed independently in triplicate. **e** Kaplan–Meier curves showing progression-free survival outcomes according to HER2 PG status or PAM50 class.

patients with *HER2* gene amplification on the basis of copy number alterations, only 19 of these patients were positive for *HER2* by PGstatus. Consistent with the proximal location of the *HER2* gene on chromosome 17q12, the copy number, RNA, and protein expression of *GRB7* and *STAR3* were concomitantly increased in patients with *HER2* amplification (Fig. 3a). The protein levels and phosphorylation metrics of *HER2* more robustly and clearly segregated the designation between patients with positive (*HER2*-PG+) and negative (*HER2*-PG-) *HER2* PG status (Fig. 3b).

In the stratification of patients with breast cancer with *HER2* driver dependency, there was a notable incompatibility between conventional PAM50-based classification, classical pathological *HER2* type designation, and proteogenomic-based integrative clustering. To gain deeper insights into the biological characteristics that led to sample clustering within the *HER2*-enriched subtype (*HER2E*) group despite inconsistent *ERBB2* status, we conducted an analysis of phosphosites that are known to canonically relay kinase signaling from the human *ERBB2* protein. Patients assigned to the *HER2E* group were divided into *HER2EPG*- and *HER2EPG*+ groups on the basis of their PG status (Fig. 3c). As anticipated, all PAM50 *HER2EPG*+ samples presented elevated levels of *ERBB2* phosphopeptides. Five additional cases of the PAM50 *HER2*-enriched subtype were classified into the *HER2EPG*- subgroup. A close examination of the five cases revealed that these samples shared common activated receptor tyrosine kinase (RTK) downstream pathway signatures, including *ARAF*, *MAPK3*, and *AKT1* (Fig. 3c). This finding indicates that the PAM50 simplification scheme mistakenly designates some samples with activated RTKs as belonging to the *HER2* subtype, regardless of their subtype-specific nature. Indeed, some samples demonstrated an activated phosphorylation profile of *IGF2R*, further suggesting that some breast cancer samples might be targetable with precise segregation of RTKs with a suitable driver dependency nature<sup>14</sup>.

Quantitative phosphoproteomic analysis revealed that *HER2* PG+ cases exhibited elevated levels of *HER2* serine phosphopeptides, including serine residues 1054 and 1066, in comparison to PG- cases (Fig. 3c). Given the close association between the phosphorylation status at serine 1054 (S-1054) and *HER2* PG status, we investigated its potential as a predictive biomarker for the sensitivity of breast cancer cell lines to *HER2* inhibitors. Despite the limited number of cell lines analyzed, the phosphorylation of S-1054 was found to be correlated with the response to lapatinib and neratinib treatments (Fig. 3d). We then compared the effects of a *HER2* inhibitor on the phosphorylation status of *HER2*-positive breast cancer cell lines. The phosphorylation level of *HER2* at tyrosine 1248 (Y-1248) decreased following treatment with neratinib. The phosphorylation of S-1054 decreased in *HER2*-positive breast cancer cell lines following neratinib treatment (Supplementary Fig. 1). These results suggest that decreased phosphorylation of S-1054 on *HER2* could serve as a predictive marker for sensitivity to *HER2* inhibitors.

Among patients with amplified *HER2*, the proteogenomic status of *HER2* was potentially prognostic, whereas PAM50 *HER2* enrichment was not (Fig. 3e). The patients with *HER2* PG+ status tended to have poorer clinical outcomes among the patients with amplified *HER2* ( $p = 0.024$ ).

### Proteogenomic analysis of the tumor immune microenvironment

Using transcriptome, proteome, and phosphoproteome datasets, we attempted to classify patients according to their immune signals to understand their tumor immune microenvironments at the molecular and cellular levels (Fig. 4a). Clusters 1 and 2 were enriched in immune signatures, and the mRNA and protein expression levels of immune checkpoint targets, such as *CD276*, *TIGIT*, *LAG3*, and *CTLA4*, were also elevated relative to those in the other clusters. Cluster 3 included an increased proportion of

patients with fewer immunogenic features, with decreased numbers of T, B, NK, and monocytic lineages, as well as immune checkpoint targets. Multiplex IHC staining of five different immune cell markers (*CD8*, *CD68*, *FOXP3*, *PD1*, and *PD-L1*) confirmed the validity of inferring tumor immune microenvironment features from a multiomics dataset (Fig. 4b). Correlation analyses of multiplexed IHC staining data revealed that YBC immune cells have highly intercorrelated molecular features, confirming the established correlative nature of immune checkpoint molecules, such as augmented *PD-L1* staining, which is highly correlated with an increased number of exhausted *CD8*+ Tregs and immunosuppressive macrophages in the tumor parenchyma (Supplementary Fig. 2). *CD8*-positive T-cell IHC further validated the inferred active immunogenic tumor microenvironment (Fig. 4c). We also identified three immune-based subtypes with distinct immune characteristics (Fig. 4d). Among the three clusters, immune-cluster 2 was assigned an immune hot phenotype, immune-cluster 3 an immune cold phenotype, and immune-cluster 1 an immune-intermediate phenotype. There was a positive correlation between the stromal and immune scores according to the three immune-based subtypes ( $R^2 = 0.4246$ ,  $p < 0.0001$ ; Fig. 4e).

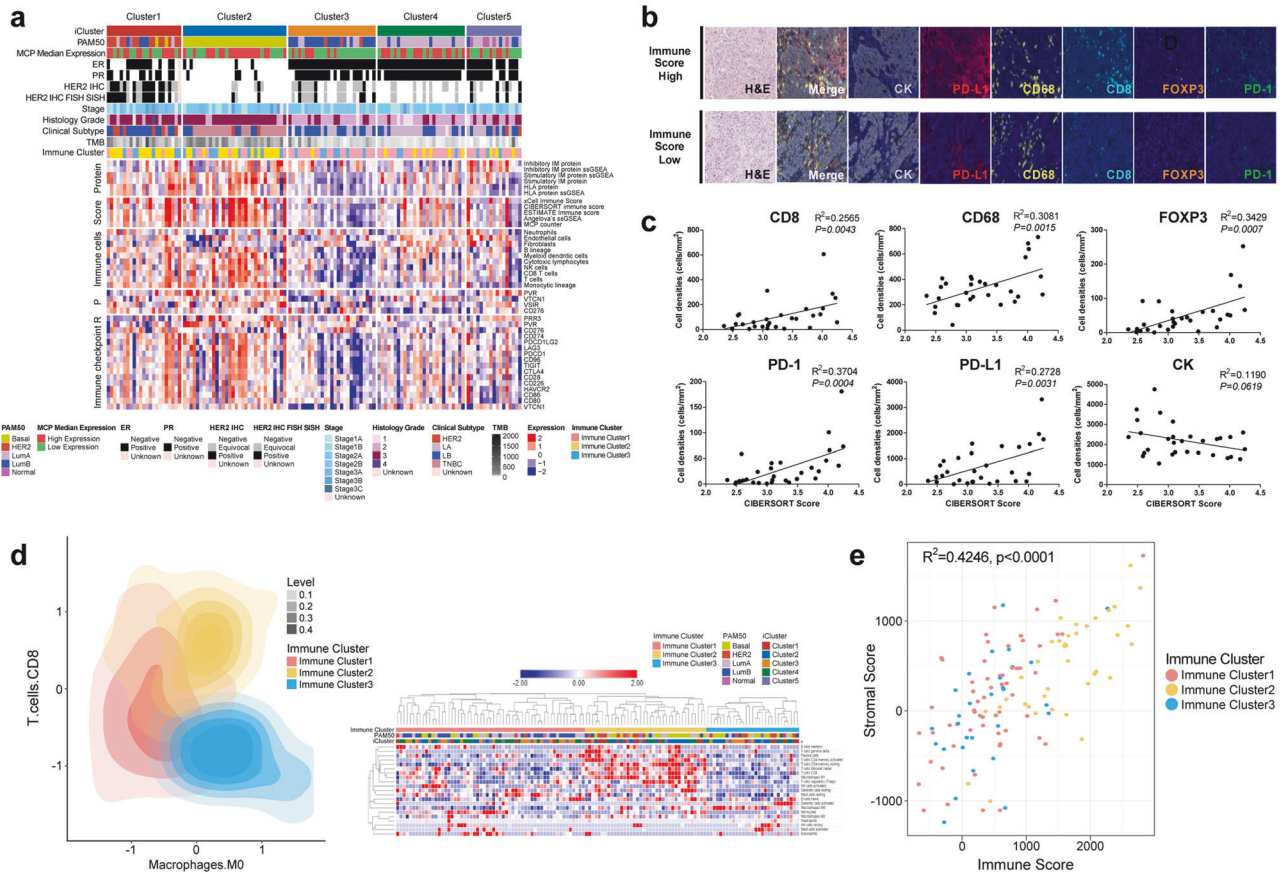
Classical RNA-seq-based PAM50 classification demonstrated that each subtype of YBC comprises a subset of patients with an active immune signature (Supplementary Fig. 3), albeit with higher immune scores in the basal PAM50 subtype (Supplementary Fig. 4) (ANOVA,  $P = 0.007$ ). The active immune tumor microenvironment consisted of elevated levels of stimulatory immune proteins and HLA proteins and increased proportions of tumor-infiltrating lymphocytes (TILs) and myeloid dendritic cells (Supplementary Fig. 3).

In contrast to the PAM50 stratification, integrative multiomics-based clustering generated immune microenvironment features that segregated cluster 3 from the other clusters (ANOVA,  $P = 0.001$ ) (Supplementary Fig. 5). The evaluation of the clinical characteristics of the poorly immunogenic subgroup tended to be associated with a poor clinical prognosis (Supplementary Fig. 6). The clinical association of poor immunogenic features with a poor prognosis was observed in the luminal B PAM50 subtypes in this study cohort (Fig. 4a), suggesting that immune microenvironment features in specific subgroups of early-onset breast cancer identified through multiomics integration may serve as prognostic molecular markers for stratifying high-risk late-onset luminal B subtype patients.

### Proteogenomic analysis of homologous recombination-deficient (HRD) breast cancer

To investigate the mutational processes involved in early-onset East Asian breast cancer, we deconvoluted the mutational signatures in the cohort. Four prevalent mutational signatures—*BRCAness*, *APOBEC*, mismatch repair defects, and aging signatures—were observed (Fig. 5a). We also examined germline pathogenic variations, defined as mutations that were reported as the most likely disease-causing variants in ClinVar or mutations that are truncating proteins in 10 genes reported to increase breast cancer susceptibility with a relatively strong penetrance<sup>10,58,59</sup>. Among the 10 breast cancer susceptibility genes, 10.3% (13/126) of patients harbored pathogenic germline variations, whereas one patient was observed to possess pathogenic variations in both *RAD50* and *BRCA1*. Pathogenic *BRCA1/2* germline mutations were more prevalent in Korean patients with YBC, affecting 8.7% of patients in the study cohort compared with 3.8% of patients in the TCGA database (Fisher's test,  $P < 0.02$ ). *BRCA* germline mutations, which are classical biomarkers of *BRCAness* and *PARP* inhibitors, only partially explained the proportion of patients with elevated genomic HRD levels (Fig. 5b). We interrogated the proteome data to search for proteins closely associated with the features of HRD (Fig. 5c). Among these, 20 highly ranked proteins were chosen whose





**Fig. 4 Immunological landscape of early-onset Korean breast cancer.** **a** Heatmap showing the wide range of expression levels for immune-related features in each integrative molecular cluster. Protein-derived signatures for immune modulator gene sets are depicted in the top panel. Z scores of RNA-based immune signatures from xCell, CIBERSORT, ESTIMATE, gene sets from Angelova, and the MCP counter are shown in the second data panel. The third to fifth data panels show log<sub>2</sub> ratios for normalized RNA-seq and proteomics data (the phosphoprotein is the median for all sites on a given protein) for immune cells and immune checkpoint targets, such as CD276, TIGIT, LAG3, and CTLA4. MCP median expression: ER: pathological staining of estrogen receptor, PR: pathological staining of progesterone receptor. **b** Multiplex imaging portrays the immune microenvironment of the tumor. The upper and lower panels represent samples that correspond to high and low computational immune scores, respectively. The samples were probed for CK (cytokeratin), PD-L1, CD68, CD8, FOXP-3, and PD-1. The morphological features are indicated by H&E staining. **c** Graphs depict the linear regression for correlations between cell counts per total area (cells/mm<sup>2</sup>) of each patient from all available ROIs for multiplex imaging probes and the CIBERSORT absolute immune score.  $*P < 0.05$  **d** YBC samples are classified into immune-hot (yellow), immune-intermediate (pink), and immune-cold (blue) groups. The heatmap on the right illustrates various immune components (y-axis) as per the immune clusters. **e** Collinearity between immune and stromal scores in early-onset breast cancer. The scatter plot shows the relationship between the computational immune score (x-axis) and the stromal score (y-axis), with colors corresponding to the three different immune clusters. Spearman rank correlation coefficient analysis of the immune score (x-axis) and stromal score (y-axis) for immune clusters 1 ~ 3 ( $R^2 = 0.4246$ ,  $p < 0.0001$ ).

combination could segregate genomic HRD-positive and HRD-negative cancers (Fig. 5d).

**Prognostic significance of protein–RNA correlations**

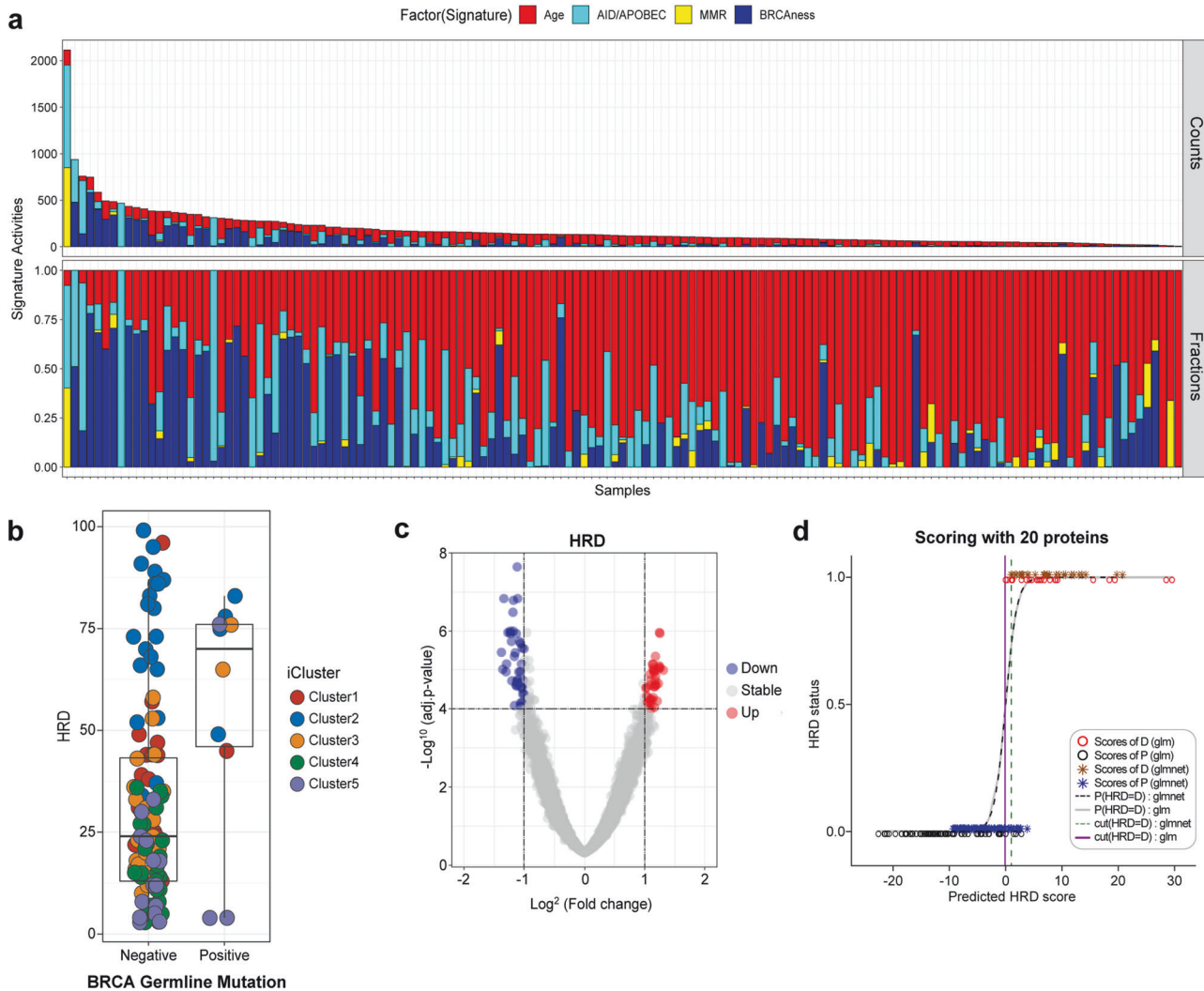
Positive correlations existed between the majority of the quantified proteins (>90%) and their corresponding mRNA transcripts within the present study cohort, with a median of 0.34 (Supplementary Fig. 7a). These correlations were not contingent upon technical factors such as the average protein precursor area or the number of peptide spectral matches. Structural ribosomal proteins and those involved in translational processes presented the lowest correlations between protein and transcript abundance (Supplementary Fig. 7a), highlighting their function and regulatory control, primarily at the protein level.

Notably, proteins derived from transcripts identified in breast cancer drivers, including estrogen receptor responses, exhibited the strongest mRNA–protein correlations (Supplementary Fig. 7a), underscoring a direct connection to the protein phenotype

in the major canonical breast cancer driver signaling pathway. In hormone receptor-positive breast cancer patients, the subgroup with a high RNA–protein correlation was associated with a poorer prognosis than the subgroup with a low RNA–protein correlation (Supplementary Fig. 7b,  $p = 0.0049$ ). This is reinforced by the greater correlation of genes that respond to estrogen receptor signaling blockade (Supplementary Fig. 7a). These findings suggest that protein–RNA correlations could serve as promising prognostic markers for predicting late recurrence in the luminal subtype, necessitating comprehensive external validation accounting for potential confounders in future investigations.

**DISCUSSION**

Patients with YBC have a poor prognosis not only in Western countries<sup>7,60</sup> but also in Korea<sup>5</sup>, particularly those with luminal A and luminal B/HER2-type breast cancer<sup>61</sup>. This phenomenon has been attributed to tamoxifen resistance<sup>5,61</sup>. Our proteogenomic



**Fig. 5 Proteogenomic analysis of homologous recombination-deficient YBC. a** Mutational signature in early-onset breast cancer displaying the quantity (upper) and proportion (lower) of somatic mutations per sample belonging to each mutational signature. These include aging (red), APOBEC (cyan blue), mismatch repair defects (MMR, yellow), and BRCAness (blue) signatures. **b** Relationship between the homologous recombination defect (HRD) index and BRCA germline mutations. Germline BRCA mutations have a limited correlation with the degree of HRD in breast cancer samples. **c** Identification of protein elements that are highly correlated with the degree of HRD in breast cancer samples. **d** Predictive accuracy of different scoring methods in forecasting HRD status on the basis of scores derived from 20 proteins associated with HRDness. The data are presented using two different modeling techniques: a generalized linear model and elastic net regularization. The x-axis represents the predicted HRD score values for each sample. The y-axis indicates HRD status, with 0 representing non-HRD (HRP) and 1 representing HRD. Scores of D, P (glm): scores derived from the generalized linear model for Dataset D or P, represented by red and black circles. Scores of D, P (glmnet): scores derived from the elastic net model for Dataset D or P, represented by brown and blue asterisks. P (HRD = D): glm/glmnet: predicted logistic probability of HRD using the glm or glmnet model, depicted by a gray solid line or black dashed line. Threshold (HRD = D): glm/glmnet: The threshold value for HRD is D using the glm or glmnet model, shown as a solid or dotted line estimated from the maximum ROC. The thresholds help distinguish between HRD and non-HRD samples, providing a visual comparison of the predictive power and accuracy of the glm and glmnet models.

analysis stratified luminal breast cancer patients into a YBC group with late recurrence (clusters 3 and 4) and very good survival (cluster 5). This classification system can be used to identify patients requiring long-term endocrine therapy.

According to the genomic expression analysis, YBC had significantly lower ER $\alpha$ , ER $\beta$ , and PR mRNA expression but higher HER-2 and EGFR expression. In addition, lower ER $\alpha$  and higher EGFR mRNA expression were significant predictors of inferior disease-free survival<sup>52</sup>.

In our Korean YBC cohort, *TP53*, *PIK3CA* and *GATA3* were the most frequently mutated genes, which is consistent with the results of the TCGA cohort. The TCGA cohort provides three age groups ( $\leq 45$ , 46–69, and  $\geq 70$ ), and *PIK3CA* and *TP53* were the most common somatic mutations in all three age groups while *GATA3*

was the third most common and independent somatic mutation in young patients, affecting 15.2%<sup>63</sup>.

Proteogenomic analysis integrating the phosphorylation status of HER2 can be used to identify true HER2+ patients. Traditionally, HER2 has been tested for gene amplification or protein over-expression by IHC and ISH. Through proteomic profiling, not only quantitative expression but also posttranslational modification of HER2 could be used for a more precise classification. We identified the proteogenomic status of HER2 as a potential prognostic marker among patients with amplified HER2, as patients with HER2 PG+ status tended to have poorer clinical outcomes than PG- patients. Especially for Asian patients with breast cancer who show a high proportion of HER2 enrichment, HER2 PG status may help predict the clinical outcomes of HER2-targeted therapies.

Furthermore, compared with PG- patients, HER2 PG+ patients presented increased phosphorylation of the HER2 protein. Phosphorylation of the HER2 protein has been extensively studied as an essential regulatory mechanism for protein function, but relatively little is known about the effects of serine residues compared with those of tyrosine residues. Our results suggest that the phosphorylation status of the serine 1054 residue is a good marker for estimating the PG status of HER2.

PAM50 is a classification tool used to characterize and predict the outcomes of patients with breast cancer<sup>64–66</sup>. However, our results demonstrated that the proteogenomic status of HER2 was a potential prognostic factor, whereas PAM50 HER2 enrichment was not. Patients with a HER2 PG+ status tended to have poorer clinical outcomes. The HER2-positive subtypes show ethnic disparities, and Korean patients with breast cancer have a high degree of HER2 enrichment<sup>10,49,67</sup>. Our results suggest that HER2 PG status is required to select true-HER2+ patients who respond well to anti-HER2 treatment. As expected, all of the PAM50 HER2EPG+ samples had high levels of HER2 phosphopeptides, whereas the PAM50 HER2EPG- samples had markedly lower levels but elevated levels of phosphorylation of other ERBB family members and of the mitogen-activated protein kinase (MAPK) signaling pathway compared with PG+ samples. These findings suggest that complementary pathways that can replace HER2 signaling may be targetable in PAM50 HER2E tumors without HER2 amplification.

The tumor microenvironment (TME) is a critical factor that influences the response to immunotherapy. The cytotoxic subset of CD8+ T cells can eliminate cancer cells and is associated with improved patient survival. Conversely, immunosuppressive regulatory CD4+ T cells (Tregs) or macrophages are associated with a poorer prognosis<sup>68</sup>. “Hot” tumors characterized by high levels of T-cell infiltration generally exhibit more favorable responses to immunotherapies than “cold” tumors, which have limited T-cell infiltration<sup>69</sup>. The existence of TILs within the TME indicates the presence of an inherent immune response against the tumor, which is linked to a more favorable prognosis and increased responsiveness to chemotherapy<sup>70</sup>.

Compared with more immunogenic cancers, breast cancer, which is traditionally considered to be poorly immunogenic, or “cold” with a modest tumor mutation burden (TMB) and a low TIL count, shows a notable prevalence of APOBEC mutational signatures<sup>71–73</sup>. The mutational load in BC exhibits significant variability, and tumors with higher TMB may exhibit a more favorable response to immune checkpoint inhibitors (ICIs)<sup>74</sup>. Although PD-L1 expression is commonly used as a biomarker, it alone is not sufficient as a predictor, demonstrating its predictive value mainly in metastatic TNBC. TNBC and HER2-positive BC are considered more immunogenic, whereas hormone receptor-positive BC is considered less immunogenic. However, certain patients with HER2-positive or hormone receptor-positive BC may also show promise for immunotherapy in addition to TNBC<sup>75</sup>.

The study cohort revealed a clinical association between an unfavorable prognosis and poor immunogenic characteristics, specifically within the luminal B PAM50 subtypes. This implies that, in cases where hot immune groups are identified within specific subtypes, it is essential to consider immunotherapeutic interventions.

Our proteome-based analysis identified HRD-associated proteins and an HRDness that could facilitate the adaptive and versatile identification of patients with YBC who may benefit from treatment with PARP inhibitors. Additionally, we demonstrated that protein–RNA correlations can be used to predict the late recurrence of hormone receptor-positive breast cancer. These findings have valuable clinical implications for improving survival outcomes and reducing treatment-related toxicity by guiding the optimal duration and selection of adjuvant treatment for patients with breast cancer. These

findings can be clinically applied to guide the optimal duration and selection of adjuvant treatment for patients with breast cancer, which is particularly crucial for patients diagnosed with breast cancer at a young age. Within each standard molecular subtype of breast cancer, we identified functionally significant protein groups whose differential abundance closely correlated with the clinical progression of breast cancer. Moreover, we derived a recurrence-predictive cluster capable of predicting late recurrence, specifically for the luminal subtype, and validated its efficacy in various patient cohorts. Prognostication currently used for breast cancer depends largely on clinical subtypes and mRNA-based information. This proteome-integrated prognostic index plays a crucial role in guiding treatment duration decisions for early-onset breast cancer, thereby contributing to improved patient stratification and personalized treatment approaches.

## DATA AVAILABILITY

All data that support the findings of this study are available from the corresponding authors upon IRB approval.

## REFERENCES

- Hong, S. et al. Cancer Statistics in Korea: Incidence, mortality, survival, and prevalence in 2017. *Cancer Res Treat.* **52**, 335–350 (2020).
- Sung, H. et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
- Kang, S. Y. et al. Breast Cancer Statistics in Korea, 2018. *J. Breast Cancer* **24**, 123–137 (2021).
- DeSantis, C. E. et al. Breast cancer statistics, 2019. *CA Cancer J. Clin.* **69**, 438–451 (2019).
- Ahn, S. H. et al. Poor outcome of hormone receptor-positive breast cancer at very young age is due to tamoxifen resistance: nationwide survival data in Korea—a report from the Korean Breast Cancer Society. *J. Clin. Oncol.* **25**, 2360–2368 (2007).
- Azim, H. A. Jr & Partridge, A. H. Biology of breast cancer in young women. *Breast Cancer Res.* **16**, 427 (2014).
- Partridge, A. H. et al. Subtype-dependent relationship between young age at diagnosis and breast cancer survival. *J. Clin. Oncol.* **34**, 3308–3314 (2016).
- Kim, J. et al. Analysis of the tumor characteristics in young age breast cancer patients using collaborative stage data of the Korea Central Cancer Registry. *Breast Cancer Res. Treat.* **187**, 785–792 (2021).
- Liao, S. et al. The molecular landscape of premenopausal breast cancer. *Breast Cancer Res.* **17**, 104 (2015).
- Kan, Z. et al. Multi-omics profiling of younger Asian breast cancers reveals distinctive molecular signatures. *Nat. Commun.* **9**, 1725 (2018).
- Colak, D. et al. Age-specific gene expression signatures for breast tumors and cross-species conserved potential cancer progression markers in young women. *PLoS One* **8**, e63204 (2013).
- Mertins, P. et al. Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534**, 55–62 (2016).
- Johansson, H. J. et al. Breast cancer quantitative proteome and proteogenomic landscape. *Nat. Commun.* **10**, 1600 (2019).
- Krug, K. et al. Proteogenomic landscape of breast cancer tumorigenesis and targeted therapy. *Cell* **183**, 1436–1456.e1431 (2020).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
- McKenna, A. et al. The genome analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
- Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
- Ramos, A. H. et al. Oncotator: cancer variant annotation tool. *Hum. Mutat.* **36**, E2423–E2429 (2015).
- Kanchi, K. L. et al. Integrated analysis of germline and somatic variants in ovarian cancer. *Nat. Commun.* **5**, 3156 (2014).
- Zanti, M. et al. Performance evaluation of pipelines for mapping, variant calling and interval padding, for the analysis of NGS germline panels. *BMC Bioinform.* **22**, 218 (2021).

22. Andrews S. FastQC (v0.11.7), <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. (2010).
23. Krueger F. Trim Galore, [https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/).
24. Bioinfokit Python package (v1.0.4), [pypi.org/project/bioinfokit/](https://pypi.org/project/bioinfokit/). (2020).
25. Mertins, P. et al. Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography-mass spectrometry. *Nat. Protoc.* **13**, 1632–1661 (2018).
26. Wang, Y. et al. Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics* **11**, 2019–2026 (2011).
27. Mun, D. G. et al. Proteogenomic characterization of human early-onset gastric cancer. *Cancer Cell* **35**, 111–124.e110 (2019).
28. Favero, F. et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann. Oncol.* **26**, 64–70 (2015).
29. Sztupinski, Z. et al. Migrating the SNP array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer. *NPJ Breast Cancer* **4**, 16 (2018).
30. Zhao, E. Y. et al. Homologous recombination deficiency and platinum-based therapy outcomes in advanced breast cancer. *Clin. Cancer Res.* **23**, 7521–7530 (2017).
31. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
32. Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L. & Pachter, L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**, 46–53 (2013).
33. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
34. Lawrence, M. S. et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
35. Gendoo, D. M. et al. Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. *Bioinformatics* **32**, 1097–1099 (2016).
36. Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J. P. & Tamayo, P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
37. Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P. & Mesirov, J. P. GenePattern 2.0. *Nat. Genet.* **38**, 500–501 (2006).
38. Yoshihara, K. et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
39. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
40. Aran, D., Hu, Z. & Butte, A. J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220 (2017).
41. Becht, E. et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* **17**, 218 (2016).
42. Angelova, M. et al. Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy. *Genome Biol.* **16**, 64 (2015).
43. Thorsson, V. et al. The immune landscape of cancer. *Immunity* **51**, 411–412 (2019).
44. Blumenberg, L., Kawaler, E. A., Cornwell, M., Smith, S., Ruggles, K. V. & Fenyo, D. BlackSheep: A bioconductor and bioconda package for differential extreme value analysis. *J. Proteome Res.* **20**, 3767–3773 (2021).
45. Wiredja, D. D., Koyuturk, M. & Chance, M. R. The KSEA App: a web-based tool for kinase activity inference from quantitative phosphoproteomics. *Bioinformatics* **33**, 3489–3491 (2017).
46. Hornbeck, P. V. et al. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res.* **40**, D261–D270 (2012).
47. Linding, R. et al. NetworKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res.* **36**, D695–D699 (2008).
48. Guiu, S. et al. Molecular subclasses of breast cancer: how do we define them? The IMPAKT 2012 Working Group Statement. *Ann. Oncol.* **23**, 2997–3006 (2012).
49. Pan, J. W. et al. The molecular landscape of Asian breast cancers reveals clinically relevant population-specific differences. *Nat. Commun.* **11**, 6433 (2020).
50. Encinas, G. et al. Somatic mutations in early onset luminal breast cancer. *Oncotarget* **9**, 22460–22479 (2018).
51. Stephens, P. J. et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486**, 400–404 (2012).
52. Pereira, B. et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat. Commun.* **7**, 11479 (2016).
53. Ellis, M. J. et al. Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* **486**, 353–360 (2012).
54. Kaufmann, M. & Rody, A. Long-term risk of breast cancer recurrence: the need for extended adjuvant therapy. *J. Cancer Res. Clin. Oncol.* **131**, 487–494 (2005).
55. Katz, E. et al. A gene on the HER2 amplicon, C35, is an oncogene in breast cancer whose actions are prevented by inhibition of Syk. *Br. J. Cancer* **103**, 401–410 (2010).
56. Bai, T. & Luoh, S. W. GRB-7 facilitates HER-2/Neu-mediated signal transduction and tumor formation. *Carcinogenesis* **29**, 473–479 (2008).
57. Wolff, A. C. et al. Human epidermal growth factor Receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Focused Update. *J. Clin. Oncol.* **36**, 2105–2122 (2018).
58. Landrum, M. J. et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862–D868 (2016).
59. Shiohvit, S. & Korde, L. A. Genetics of breast cancer: a topic in evolution. *Ann. Oncol.* **26**, 1291–1299 (2015).
60. Lin, C. H. et al. Contrasting epidemiology and clinicopathology of female breast cancer in Asians vs the US population. *J. Natl Cancer Inst.* **111**, 1298–1306 (2019).
61. Johansson, A. L. V., Trewin, C. B., Hjerkind, K. V., Ellingjord-Dale, M., Johannesen, T. B. & Ursin, G. Breast cancer-specific survival by clinical subtype after 7 years follow-up of young and elderly women in a nationwide cohort. *Int J. Cancer* **144**, 1251–1261 (2019).
62. Anders, C. K. et al. Young age at diagnosis correlates with worse prognosis and defines a subset of breast cancers with shared patterns of gene expression. *J. Clin. Oncol.* **26**, 3324–3330 (2008).
63. Azim, H. A. Jr, Nguyen, B., Brohee, S., Zoppoli, G. & Sotiriou, C. Genomic aberrations in young and elderly breast cancer patients. *BMC Med.* **13**, 266 (2015).
64. Perou, C. M. et al. Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
65. Nielsen, T. O. et al. A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. *Clin. Cancer Res.* **16**, 5222–5232 (2010).
66. Liu, M. C. et al. PAM50 gene signatures and breast cancer prognosis with adjuvant anthracycline- and taxane-based chemotherapy: correlative analysis of C9741 (Alliance). *NPJ Breast Cancer* **2**, 15023 (2016).
67. Telli, M. L. et al. Asian ethnicity and breast cancer subtypes: a study from the California Cancer Registry. *Breast Cancer Res Treat.* **127**, 471–478 (2011).
68. Allard, B. et al. Immuno-oncology-101: overview of major concepts and translational perspectives. *Semin Cancer Biol.* **52**, 1–11 (2018).
69. Gajewski, T. F. The next hurdle in cancer immunotherapy: overcoming the non-T-cell-inflamed tumor microenvironment. *Semin Oncol.* **42**, 663–671 (2015).
70. Savas, P. et al. Clinical relevance of host immunity in breast cancer: from TILs to the clinic. *Nat. Rev. Clin. Oncol.* **13**, 228–241 (2016).
71. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
72. Emens, L. A. Breast cancer immunotherapy: facts and hopes. *Clin. Cancer Res.* **24**, 511–520 (2018).
73. Adams, S. et al. Current landscape of immunotherapy in breast cancer: a review. *JAMA Oncol.* **5**, 1205–1214 (2019).
74. Karn, T. et al. Tumor mutational burden and immune infiltration as independent predictors of response to neoadjuvant immune checkpoint inhibition in early TNBC in GeparNuevo. *Ann. Oncol.* **31**, 1216–1222 (2020).
75. Yao, J., Li, S. & Wang, X. Identification of breast cancer immune subtypes by analyzing bulk tumor and single cell transcriptomes. *Front Cell Dev. Biol.* **9**, 781848 (2021).

## ACKNOWLEDGEMENTS

We thank Dr. Henry Rodriguez and Dr. Ana I. Robles from the U.S. National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium (CPTAC) for helpful discussions.

## AUTHOR CONTRIBUTIONS

Conception and/or design: K.-A.Y., S.-Y.J., J.B.P., and S.-Y.K. Acquisition, analysis, and/or interpretation of data: K.-A.Y., Y.K., S.-Y.J., J.-S.R., K.-H.K., E.-G.L., H.C., Y.K., J.K., J.B.P., and S.-Y.K. Review of the manuscript: K.-A.Y., Y.K., S.-Y.J., J.-S.R., K.-H.K., E.-G.L., H.C., Y.K., J.K., J.B.P., and S.-Y.K.

## FUNDING

This work was supported by grant NCC-1810861 from the National Cancer Center of Korea to J.B.P. and 202000580001 from GSK of the USA to S.-Y. K and grant NCC-2210543 from the National Cancer Center of Korea to S.-Y.K. This work was also

supported by the auspices of a Memorandum of Understanding between the proteogenomic Research Project of NCC, Korea, and the U.S. National Cancer Institute's International Cancer Proteogenome Consortium (ICPC). The ICPC encourages international cooperation among institutions and nations in proteogenomic cancer research, in which proteogenomic datasets are made available to the public.

### COMPETING INTERESTS

J.K. is an employee of GSK, USA, and the other authors declare no competing interests.

### ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s12276-024-01332-w>.

**Correspondence** and requests for materials should be addressed to Jong Bae Park or Sun-Young Kong.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024