doi.org/10.36721/PJPS.2026.39.12.331.1

PGR expression as a pharmacogenomic companion biomarker to GENE70-derived genomic risk in ER-positive/HER2-negative breast cancer

Orcun Can
Acibadem Mehmet Ali Aydinlar University, Acibadem Maslak Hospital, Faculty of Medicine, Department of Medical Oncology, Istanbul, Turkey, 34408, Istanbul, Turkey.

Abstract: **Background**: The biology of the estrogen receptor-positive (ER+) and human epidermal growth factor receptor 2-negative (HER2-) breast cancers is heterogeneous even when they are categorized by their risk via genomics. Transcriptomic PGR expression reflects endocrine pathway activity and may provide complementary biological information within established GENE70-derived genomic-risk categories. Whether this molecular marker improves the biological interpretation of genomic-risk stratification beyond conventional clinicopathological assessment remains uncertain. **Objectives**: The aim of this study was to determine whether transcriptomic PGR expression provides complementary biological and prognostic information within reconstructed GENE70-derived genomic-risk categories and refines the characterization of endocrine-related tumour biology in ER-positive/HER2-negative breast cancer. **Methods**: This study analysed publicly available transcriptomic and clinical data from three cohorts: METABRIC (discovery cohort), GSE96058/SCAN-B cohort (validation cohort) and TCGA-BRCA cohort (molecular validation cohort). The GENE70-derived genomic-risk score was reconstructed for each cohort using matched genes. Cox regression, Kaplan-Meier analysis and subgroup comparisons were used to assess relationships between PGR expression, clinicopathologic variables, molecular features and survival outcomes. **Results**: Across the three independent cohorts, low transcriptomic PGR expression was consistently associated with higher GENE70-derived genomic risk, increased MKI67 expression, reduced ESR1 expression and enrichment of the Luminal B subtype. Survival findings differed between cohorts. In the discovery METABRIC cohort, transcriptomic PGR expression showed heterogeneous associations with survival, particularly within GENE70-derived high-risk subgroups, whereas the external GSE96058/SCAN-B validation cohort demonstrated consistent associations between low PGR expression and poorer overall survival in both the overall ER-positive/HER2-negative population and GENE70-derived high-risk subgroups. **Conclusion**: These findings suggest that transcriptomic PGR provides complementary biological and prognostic information within GENE70-derived genomic-risk categories. However, because treatment response was not evaluated in the present study, the findings should not be interpreted as evidence of predictive or pharmacogenomic utility and prospective studies incorporating treatment-response analyses are required before such applications can be established.

Keywords: ER-positive/HER2-negative breast cancer; Endocrine resistance; GENE70-derived risk; Genomic risk; GSE96058; Luminal breast cancer; MammaPrint; METABRIC; Progesterone receptor; PGR expression; SCAN-B; TCGA-BRCA

Submitted on 17-06-2026 – Revised on 20-07-2026 – Accepted on 27-07-2026

INTRODUCTION

With the extensive advances made in screening, surgery, systemic therapy and molecular characterization of breast cancer, breast cancer continues to be the most commonly diagnosed cancer in women globally despite these advances (Bray *et al*., 2024; Giaquinto *et al*., 2024). In addition to providing the most definitive way to determine which subtypes of breast cancer are likely to respond to endocrine therapy, HR+/HER2− breast cancer accounts for the majority of cases within this subtype.

Management of the majority of patients with HR+/HER2− breast cancer involves an endocrine-based therapeutic approach (Tutt *et al*., 2021). For some patients with HR+/HER2− breast cancer, additional therapy may be indicated, depending on the stage of the disease and the potential for recurrence; these additional therapies may include, but are not limited to: chemotherapy, ovarian ablation, CDK4/6 inhibitors, PARP inhibitors and other targeted agents (Loibl *et al*., 2024; Slamon *et al*., 2024). Although nearly all ER+/HER2− breast cancer has been shown to be subject to endocrine therapy as a primary treatment strategy, significant phenotypic heterogeneity has been observed among ER+/HER2- breast cancer, with a number of patients experiencing disease recurrence or relapse despite receiving standard endocrine therapy as part of their management for these disease processes (Loibl *et al*., 2024; Johnston *et al*., 2023; Henry *et al*., 2022). Thus, identifying biomarkers that will assist physicians in characterizing the biology of endocrine-responsive tumours remains a primary goal of clinical research.

**Corresponding author:* E-mail: drorcuncan@gmail.com

The inclusion of multigene assays into the management of patients with early ER+/HER2- breast cancer has facilitated better risk stratification. These assays represent a valuable addition to common clinicopathological variables and help identify patients unlikely to benefit from adjuvant chemotherapy while aiding treatment decisions for selected node-positive and node-negative populations (Kalinsky *et al*., 2021). Among these assays, the 70-gene signature has been extensively validated as a prognostic marker to discriminate patients at risk of recurrence, independent of traditional clinicopathological factors. Notably, however, tumours with comparable genomic-risk classifications can vary significantly in terms of their biological characteristics, suggesting that additional biomarkers could further improve the interpretation of genomic-risk classifications (Sparano *et al*., 2018).

Multigonate markers associated with biological activity encompass much more than simply proliferation because they monitor many elements, including endocrine interaction, cell cycle and common stroma activity in addition to the proliferative activity of the tumour (Lloyd *et al*., 2024; Martin *et al*., 2024). Therefore, when calculating the level of risk to develop a return rate, it is best thought of as a composite of tumour biology rather than a singular measure. Across all ER+/HER2 disease, variability in how much the tumours have differentiated into endometrium and in activity within the endometrial pathway is likely a cause of varying clinical outcomes among tumours displaying similar genomic risk levels (Sparano *et al*., 2019; Henry *et al*., 2022).

With respect to the importance of PR biology as it relates to the ER+ and HER2- subgroup, it is noteworthy that PR expression is consistently assessed in breast cancer pathology and serves as a marker for the effect of ER on the tumour. It has been shown that PR is not merely a passive indicator of ER activity, but also has an active role in controlling the ability of ER to bind to chromatin and regulate transcriptional programs of luminal breast cancer (Allison and Rendi, 2022; Purdie *et al*., 2024) Beyond basic biological observation, clinicopahthologically, it is well established that low levels of PR correlate with poor prognostic indicators such as high tumour grade, increased activity, poor luminal B characteristics and poor overall outcomes among ER+ patients when receiving treatment methods directed against ER (Horwitz and Sartorius, 2020; Purdie *et al*., 2024). This presents strong evidence that the loss of PR signaling defines a biologically unique subset of tumours with reduced ductal differentiation and more aggressive behavior.

Clinical progesterone receptor (PR) status assessed by immunohistochemistry and transcriptomic PGR expression represent related but biologically distinct measurements. Clinical PR is routinely evaluated as a binary or semi-quantitative protein biomarker. In contrast, transcriptomic PGR reflects continuous gene-expression levels that may capture variation in endocrine pathway activity not fully represented by immunohistochemical assessment. Whether transcriptomic PGR provides complementary biological or prognostic information beyond conventional PR assessment remains uncertain and has not been comprehensively evaluated within GENE70-derived genomic-risk categories. The present study was designed to investigate this biological relationship rather than to establish transcriptomic PGR as a replacement for routine clinical PR assessment.

Although both genomic-risk signatures and progesterone receptor status are established prognostic indicators in ER-positive/HER2-negative breast cancer, an important knowledge gap remains. Previous studies have consistently shown that low clinical PR expression is associated with Luminal B disease, increased proliferation, endocrine resistance and poorer clinical outcomes. However, it remains unclear whether continuous transcriptomic PGR expression provides additional biological or prognostic information in tumours already stratified by genomic risk. Specifically, it is unknown whether transcriptomic PGR can identify biologically distinct subgroups among tumours with similar GENE70-derived genomic-risk profiles, thereby improving the interpretation of genomic-risk classifications rather than replacing existing clinicopathological biomarkers.

This is the first study to integrate a reconstructed GENE70-derived genomic risk assessment with continuous transcriptomic PGR expression across three independent public breast cancer cohorts (METABRIC, GSE96058/SCAN-B and TCGA-BRCA) to evaluate whether transcriptomic PGR provides complementary biological and prognostic information within established genomic risk categories. Rather than evaluating PGR as a replacement for conventional progesterone receptor assessment, transcriptomic PGR was investigated as a complementary molecular indicator of endocrine pathway activity and tumour biology within established genomic-risk categories. The relationships between PGR expression, GENE70-derived genomic risk, proliferation, estrogen-signaling activity, luminal subtype distribution and survival outcomes were examined using METABRIC as the discovery cohort, GSE96058/SCAN-B as the external clinical validation cohort and TCGA-BRCA as the molecular validation cohort. It was hypothesized that reduced transcriptomic PGR expression would identify biologically more aggressive tumours within the same genomic-risk category, thereby refining the biological interpretation of GENE70-derived genomic-risk stratification.

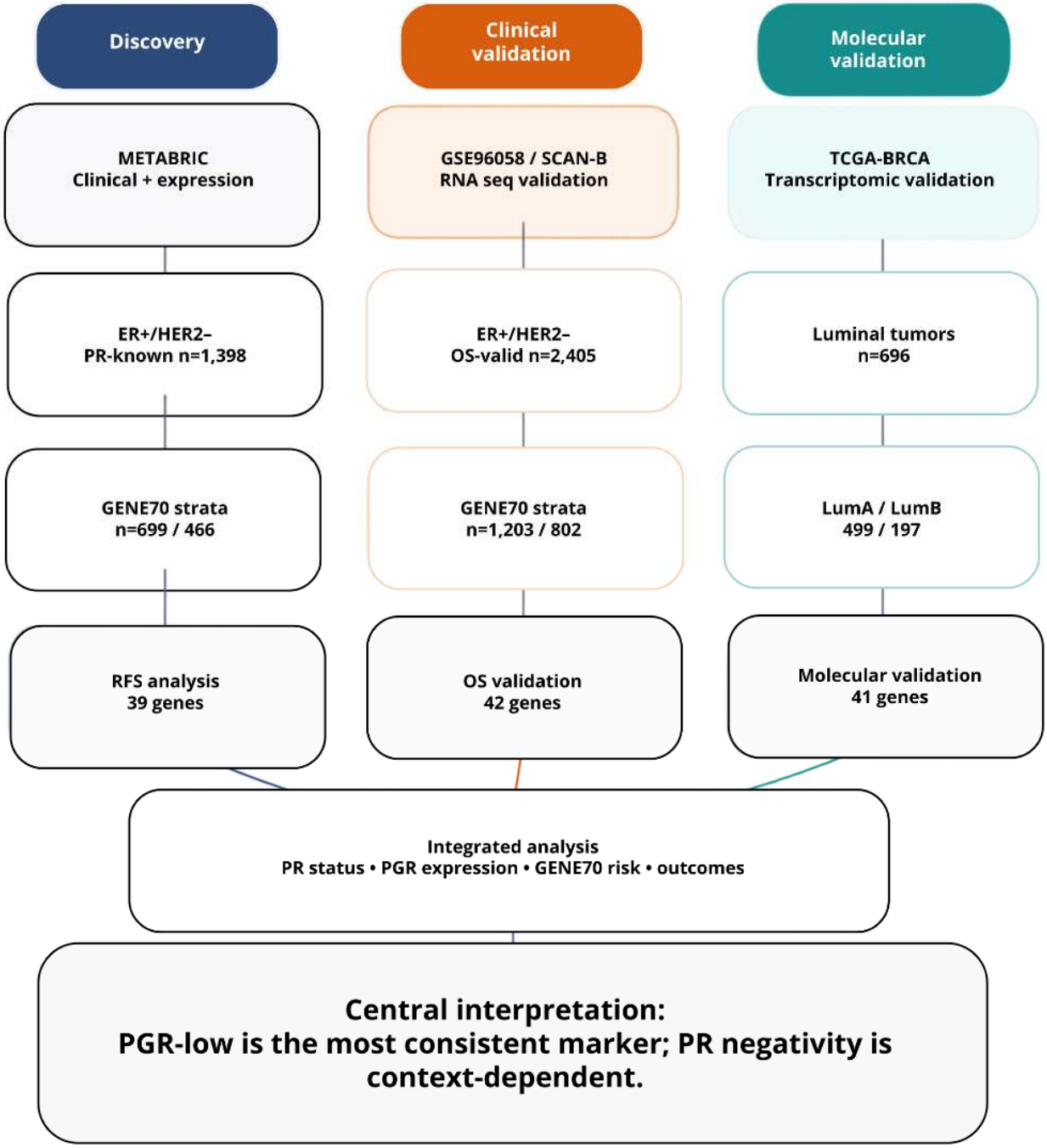

Fig. 1: Study workflow and analytics summary.

MATERIALS AND METHODS

Study design

This study was retrospective and multi-cohort, based on transcriptomic PGR-low status and clinical PR negativity, to assess whether these two biomarkers provide complementary biological & prognostic information to resolve the GENE70 genomic risk category in ER+/HER2- breast cancer. The study comprised analysis of publicly available datasets in a discovery-validation approach: the METABRIC cohort for recurrence-free survival, GSE96058/SCANB for overall survival validation and TCGA-BRCA for molecular validation to develop an integrated model. Reporting of the results followed REMARK guidelines and the overall study workflow and analytical strategy are summarized in Fig. 1.

Data sources

The data shown here are from the METABRIC study for breast cancer, available from the cBioPortal/DataHub and annotated as “Breast Cancer (METABRIC, Nature 2012

& Nat Commun 2016)" and contains clinical and molecular information on 2,509 primary breast tumours. The METABRIC dataset included variables for recurrence-free survival, clinical receptor status, clinicopathologic features and Illumina microarray mRNA expression data. The integrated METABRIC clinical dataset included 2,509 records and the expression dataset included 20,603 genes for 1,982 expression-profiled samples in the local analysis files. Clinical and expression characteristics were matched and the inclusion criteria (ER-positive, HER2-negative and PR-known) were applied to the primary METABRIC analytic cohort, comprising 1,398 tumours.

The GSE96058/SCAN-B dataset was obtained from the NCBI Gene Expression Omnibus. The GEO record for GSE96058 is a population-based Sweden Cancerome Analysis Network (SCAN) cohort of breast cancer patients with RNA-sequencing data, conventional breast cancer biomarker annotations, PAM50 subtyping and survival analyses. There are 3,409 samples in the GEO series and in the validation cohort, there were 3,273 breast cancers with a median follow-up of 52 months. Tumours included in the analytic cohort for this analysis were those from the GSE96058 cohort that had available PR/PgR status, RNA-seq expression data and were considered valid for overall survival (OS) data. There were 2,405 tumours in the final GSE96058 validation cohort, 2,271 of which were PR-positive and 134 PR-negative.

The cBioPortal/DataHub PanCancer Atlas breast invasive carcinoma study was used as an independent molecular-validation cohort to obtain TCGA-BRCA data. RNA-seq expression data, PAM50-like subtype classification and survival endpoints were obtained from the TCGA-BRCA PanCancer Atlas clinical and expression files available through cBioPortal/DataHub. Within this selected cBioPortal PanCancer Atlas resource, the downloaded clinical tables did not contain complete and usable ER, PR and HER2 receptor annotations required for clinical biomarker analyses. Alternative TCGA clinical resources were considered; however, these resources differed in data structure, harmonization and sample matching with the expression matrices used in the present study. To maintain a single, internally consistent dataset for integrated transcriptomic and survival analyses, only the cBioPortal PanCancer Atlas dataset was used. Consequently, TCGA-BRCA was included exclusively for molecular validation of transcriptomic PGR biology and was not used as a clinical PR-status validation cohort. The molecular-validation cohort consisted of 696 luminal tumours (499 LumA and 197 LumB).

Cohort definitions

The main clinical phenotype of interest was ER+ (hormone-receptive)/HER2- (human epidermal growth factor receptor 2-negative) breast cancer. In METABRIC and GSE96058/SCAN-B, tumours were included in the PR status analyses if they were annotated as ER-positive and HER2-negative, had known PR status and had matching expression data. METABRIC provided the discovery recurrence-free survival cohort and the external clinical overall survival validation cohort was provided by GSE96058/SCAN-B.

METABRIC comprised 1,398 ER-positive/HER2-negative PR-known expression-matched tumours. One thousand three hundred and ninety-seven (973) of these were PR-positive, while 425 were PR-negative. Recurrence-event information was missing for one case in METABRIC; 1397 tumours had 535 recurrence events for which complete RFS information was required for RFS analyses. The primary external validation cohort consisted of 2405 ER-positive/HER2-negative OS-valid tumours (2271 PR-positive and 134 PR-negative, with 213 OS events observed) for GSE96058/SCAN-B.

TCGA-BRCA filtering by clinical ER/PR/HER2 wasn't used since this set of receptor fields wasn't available in the selected clinical tables of the cBioPortal PanCancer Atlas. However, TCGA-BRCA was limited to recognizing only Luminal tumours in the tumour variable present. This resulted in a molecularly validated luminal sample of 696 tumours. TCGA analyses were thus considered transcriptomic and molecular confirmation of the PGR-low phenotype, rather than of clinical PR-negative status.

Accurate gene signature extraction and gene matching

In the present study, the term "GENE70-derived genomic risk" refers to a reconstructed research genomic-risk score generated from publicly available transcriptomic datasets using the published GENE70 signature. Because the score was reconstructed from matched genes available in each dataset rather than obtained using the proprietary commercial assay, it is referred to throughout this manuscript as "GENE70-derived genomic risk" and should not be interpreted as a commercial MammaPrint result.

The GENE70 signature was obtained from the Bioconductor genefu package as the object sig.gene70. Download of the full R package was not necessary for the final Python workflow and the genefu source archive was downloaded and the sig.gene70.rda object was extracted directly. The extracted signature consisted of 70 probe-level entries and comprised the following information: probe identifier, correlation, average good prognosis profile weight, Entrez identifier, NCBI gene symbol, HUGO gene symbol, cytoband, alternative symbol and gene description.

For cross-platform gene matching, probe-level annotations were collapsed to unique HUGO gene symbols, yielding 52 unique GENE70 genes that could be evaluated across the public datasets. Gene recovery was

39 of 52 genes in METABRIC, 42 of 52 genes in GSE96058/SCAN-B and 41 of 52 genes in TCGA-BRCA. The recovery of GENE70-matched genes across the three expression datasets is summarized in the supplementary figures. The incomplete recovery reflects differences in transcriptomic platforms rather than selective exclusion of genes. Reconstruction of published multigene signatures from the subset of genes shared across independent public datasets is a commonly used strategy when the complete proprietary assay cannot be reproduced. Accordingly, the reconstructed score was used only as a research approximation of GENE70-derived genomic-risk biology and was not intended to reproduce or replace the commercial MammaPrint assay. All analyses are therefore described as GENE70-derived genomic-risk analyses throughout the manuscript.

Reconstructing the GENE70-derived risk score

Because the proprietary MammaPrint algorithm and complete assay implementation are not publicly available, the GENE70-derived genomic-risk score was reconstructed using the genes that were consistently identifiable across the analyzed datasets. Harmonized gene symbols were standardized within each cohort and combined using the published GENE70 signature weights available through the genefu reference object. This reconstruction was intended to preserve the biological direction of the published GENE70 signature while acknowledging that analytical equivalence with the commercial assay cannot be assumed. Consequently, all reported analyses represent a research-based GENE70-derived genomic-risk score rather than commercial MammaPrint classifications.

GENE70-derived risk-score calculation

The expression values for each cohort for the matched GENE70 genes were retrieved and mapped to clinical sample identifiers. If more than one expression row mapped to the same gene symbol, the values were averaged at the gene level. Publicly available processed expression matrices were used for all three cohorts. The downloaded expression data were retained in their published normalized formats provided by the original data resources (METABRIC, GSE96058/SCAN-B and TCGA-BRCA) and no additional log transformation of the expression matrices was performed. For cross-cohort comparability during GENE70 reconstruction, matched gene-expression values were converted to numeric format and standardized within each cohort by calculating gene-wise z-scores before weighted score calculation. Standardized GENE70 reference weights were then applied to generate the reconstructed GENE70-derived genomic-risk score.

The score obtained was re-oriented so that a higher score indicated a higher level of genomic risk derived from the GENE70. Recurrence-event coding was corrected in METABRIC to ensure that higher values are associated with higher recurrence risk, not lower risk. GSE96058/SCAN-B and TCGA-BRCA were then oriented in the same direction. GSE96058/SCAN-B and TCGA-BRCA were then aligned in the same direction. High-risk strata were established in each relevant cohort according to the score distribution of each cohort, which were derived from the GENE70. The median high-risk stratum were tumours with scores equal to or greater than the median score of the cohort and the upper-tertile high-risk stratum were tumours with scores equal to or greater than the upper tertile score of the cohort. METABRIC comprised 699 tumours in the median high-risk cohort and 466 tumours in the upper tertile high-risk cohort. Median high-risk cohort (GSE96058/SCAN-B): 1203 tumours; upper tertile high-risk cohort (GSE96058/SCAN-B): 802 tumours.

The primary methods for analyzing survival were using the lower third of PGR expression to achieve maximum biological specificity for the endocrine-related low-expression state and using median-based stratification to obtain balanced molecular comparisons and validation analyses.

Biomarker definitions

The clinical PR status was analyzed as a binary marker in METABRIC and GSE96058/SCAN-B; tumours annotated as PR-positive were taken as the reference group and tumours annotated as PR-negative were coded as PR-negative. Since there were no clinical PR receptor fields in the clinical tables selected for TCGA-BRCA, clinical PR-negative status was not assessed.

The expression of molecular markers of PGR signalling was extracted directly from the expression matrices and analysed both as a continuous variable and as a categorical biomarker. Continuous transcriptomic PGR expression was used for correlation analyses with GENE70-derived genomic risk, MKI67 expression and ESR1 expression to evaluate the biological relationships without applying arbitrary thresholds. For categorical analyses, cohort-specific cut-offs were selected according to the objective of each analysis. The lower tertile of PGR expression was used as the primary definition of PGR-low in survival analyses because it identifies tumours with the greatest reduction in endocrine-related signalling and maximizes biological specificity for evaluating adverse prognostic associations. Median-based cut-offs were used for cross-cohort molecular comparisons and subgroup analyses because they provide balanced group sizes and improve statistical stability across independent datasets with different expression distributions. Additional subgroup-specific median cut-offs were applied only where appropriate to preserve adequate sample size within subgroup analyses. The consistency of the continuous and categorical analyses was evaluated to determine whether the observed biological associations were robust and not dependent on a single threshold

definition. The expression values of additional molecular markers of proliferation (MKI67), endocrine signalling (ESR1) and HER2-related biology (ERBB2) were also extracted. When subgroup analyses required dichotomization of MKI67 expression, cohort-specific median values were used.

PGR expression was determined to be a transcriptomic indicator of activity in the endocrine pathway and to convey information regarding pharmacogenomically relevant data on patients diagnosed with ER-positive/HER2-negative malignancies.

Outcomes

Recurrence-free survival (RFS) was the main survival endpoint in the METABRIC study. The available RFS month variable was used to determine the RFS time and the RFS status field was used to code for recurrence events. The "1" was coded as an event and the "0 Recurred" as censored in the RFS event coding. Prior to all final survival analyses this correction was applied.

Overall survival was the major validation end point for the primary GSE96058/SCAN-B. OS time and event status were available from the survival variables supplied with GSE96058 and only tumours with valid OS time or event status were included in the OS analyses. METABRIC RFS and GSE96058 OS were reported separately as they measure different clinical outcomes and had different event burdens.

Exploratory validation of the endpoints of OS, DSS, DFS and PFS (where available) was performed for TCGA-BRCA. Exploratory validation of the endpoints of OS, DSS, DFS and PFS (where available) was performed for TCGA-BRCA. TCGA-BRCA was therefore not used as a clinical survival-validation cohort, as clinical receptor fields were not available and there were limited numbers of events in several luminal and high-risk subgroups. The primary endpoint of TCGA-BRCA was then the molecular phenotyping validation of low PGR expression. Exploratory validation of the endpoints of OS, DSS, DFS and PFS (where available) was performed for TCGA-BRCA. TCGA-BRCA was therefore not used as a clinical survival-validation cohort, as clinical receptor fields were not available and there were limited numbers of events in several luminal and high-risk subgroups. The primary endpoint of TCGA-BRCA was then the molecular phenotyping validation of low PGR expression.

Statistical analysis

Statistical analyses were performed using Python statistical and survival-analysis libraries. Continuous variables were compared using the Mann–Whitney U test, categorical variables using Fisher's exact test and correlations were assessed using Spearman's rank correlation. Survival analyses were performed using Kaplan–Meier curves with log-rank tests and Cox proportional hazards regression models. Hazard ratios (HRs) with 95% confidence intervals (CIs) were calculated and statistical significance was defined as a two-sided p-*value* <0.05.

Availability of data and data reproducibility

All datasets analyzed in this study were publicly available through cBioPortal/DataHub, GEO and the Bioconductor Gene Set Resource. Processed datasets and analytical workflows were used to ensure reproducibility. The reconstructed GENE70-derived genomic-risk score represents a research approximation rather than the commercial MammaPrint assay.

RESULTS

Cohort feasibility

Downstream scoring and validation of the GENE70-derived scores were possible in all three cohorts, as there was enough gene recovery. 1398 METABRIC ER-positive/HER2-negative PR-known tumours, 2,405 GSE96058/SCAN-B ER-positive/HER2-negative OS-valid tumours and 696 TCGA-BRCA luminal tumours were included in the analytic cohorts. All datasets contained matched GENE70 genes with 39/52 matched genes from METABRIC, 42/52 matched genes from GSE96058/SCAN-B and 41/52 matched genes from TCGA-BRCA. The event burden observed in the different cohorts was also found to be quite different for the two endpoints, with a higher burden of recurrence events in METABRIC compared to the observed burden of OS events in GSE96058/SCAN-B, therefore, RFS and OS endpoints were reported separately from one another, rather than assuming they are directly interchangeable. The corresponding PR-status group sizes and crude event rates across the two clinical cohorts are summarized in the cohort characteristics. The analytic sample sizes available, endpoint availability and feasibility of gene-matching are summarized in Fig. 2 and Table 1. Details of cohort selection, sequential exclusion criteria, final analytical sample sizes and missing data across the three cohorts are provided in table S1. A cross-cohort summary of the associations between PR/PGR status, GENE70-derived genomic risk, proliferation, luminal subtype and survival outcomes is provided in Table S2.

The primary analyses focused on evaluating the relationship between clinical PR status, transcriptomic PGR expression, GENE70-derived genomic risk score and survival outcomes in the overall ER-positive/HER2-negative cohorts. All subgroup analyses, alternative biomarker cut-offs, attenuation analyses, molecular phenotype evaluations and additional multivariable models were performed to determine the consistency of the primary findings across different biological and clinical settings. These secondary analyses should therefore be considered exploratory and interpreted as supportive rather than confirmatory evidence.

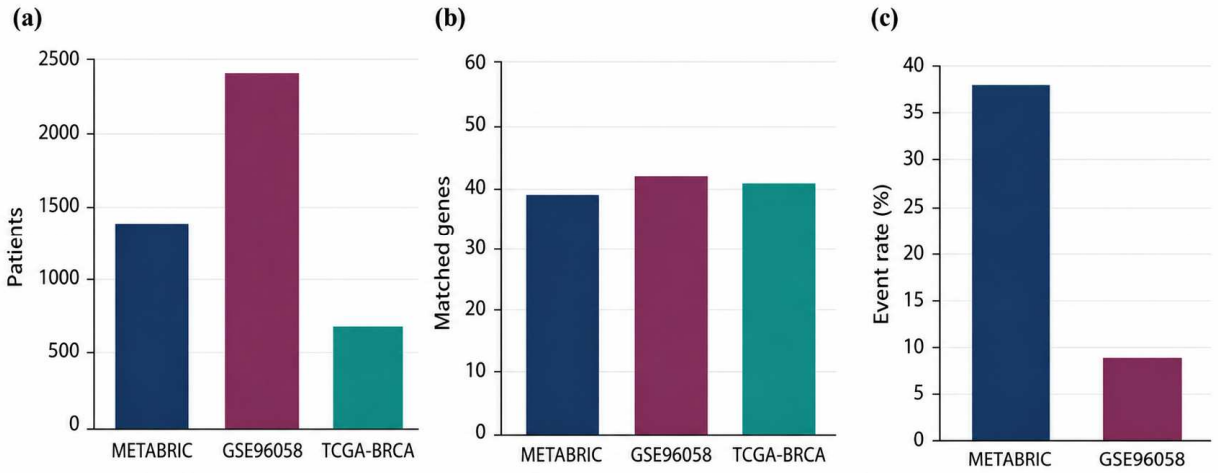

Fig. 2: Cohort characteristics and feasibility of GENE70-derived genomic-risk scoring across the three study cohorts. (a) Analytic cohort sizes for METABRIC, GSE96058/SCAN-B and TCGA-BRCA; (b) Number of GENE70 genes recovered in each cohort from the 52-gene signature; (c) Observed event burden for the primary survival endpoints in METABRIC (recurrence-free survival) and GSE96058/SCAN-B (overall survival). METABRIC contributed the primary recurrence-free survival analysis, GSE96058/SCAN-B provided external overall survival validation, and TCGA-BRCA was used for molecular validation.

Table 1: Analytic cohort characteristics and GENE70-derived scoring feasibility.

Dataset	Role	Analytic cohort	N	Endpoint/use	Events, n (%)	PR-negative, n	GENE70 matched	Strata used
METABRIC	Discovery / RFS analysis	ER+/HER2−, PR-known, expression-matched	1,398	RFS	536 (38.3)	425	39/52	Median high-risk: 699; upper-tertile: 466
GSE96058	External clinical validation	ER+/HER2−, PR/PgR-known, OS-valid	2,405	OS	213 (8.9)	134	42/52	Median high-risk: 1,203; upper-tertile: 802
TCGA-BRCA	Molecular validation	Luminal surrogate transcriptomic cohort	696	Molecular phenotype	—	—	41/52	LumA: 499; LumB: 197

Note. Events are reported for the primary endpoint used in each survival cohort. TCGA-BRCA was used for molecular validation; therefore, event and PR-negative counts are not reported in this main cohort table. GENE70 matched genes are shown out of 52 valid genes used for scoring.

METABRIC phenotype

The molecular phenotype associated with clinical PR negativity in METABRIC was characterized by significantly higher GENE70-derived genomic-risk scores, significantly lower PGR expression, significantly higher MKI67 expression and significantly lower ESR1 expression in PR-negative tumours compared with PR-positive tumours. The analysis included 425 PR-negative tumours and 973 PR-positive tumours (Fig. 3; Table 1).

METABRIC recurrence

METABRIC recurrence-free survival (RFS) analyses showed a heterogeneous association between clinical PR status and recurrence across the evaluated GENE70-derived genomic-risk strata. Kaplan–Meier curves demonstrated a non-significant trend toward poorer recurrence-free survival for PR-negative tumours in the overall ER-positive/HER2-negative cohort and no statistically significant differences were observed in the median or upper-tertile GENE70-derived high-risk subgroups (Fig. 4; Table 2). The corresponding forest plot of METABRIC PR-status survival sensitivity models is shown in Fig. S1. The corresponding GSE96058/SCAN-B validation models are shown in figure. S2. Exploratory TCGA-BRCA survival analyses across the available OS, DSS, DFS and PFS endpoints are shown in Fig. S3.

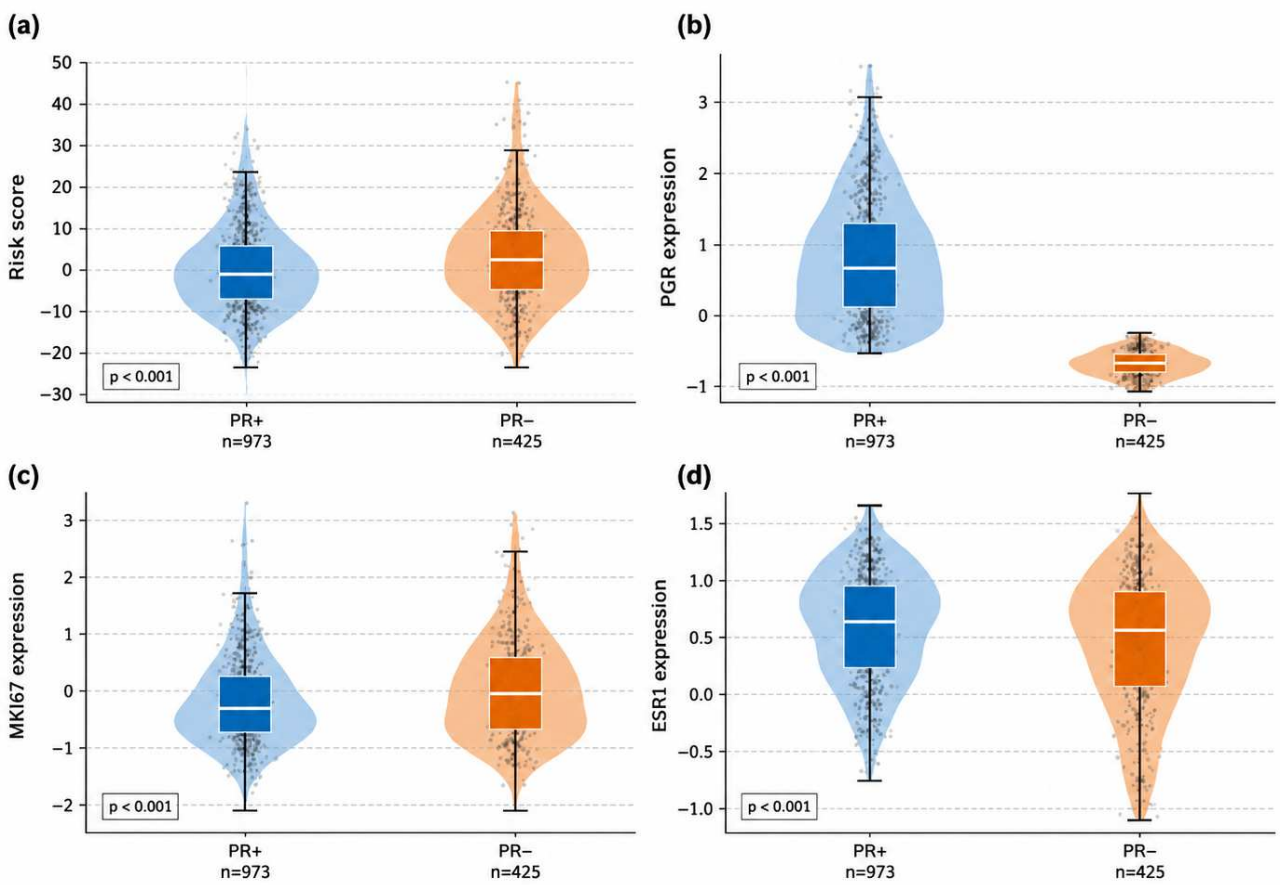

Fig. 3: Molecular phenotype associated with PR loss in METABRIC.
(a) GENE70-derived risk score; (b) PGR expression; (c) MKI67 expression; and (d) ESR1 expression according to PR status. PR-negative tumours showed higher GENE70-derived risk scores and MKI67 expression and lower PGR and ESR1 expression than PR-positive tumours.

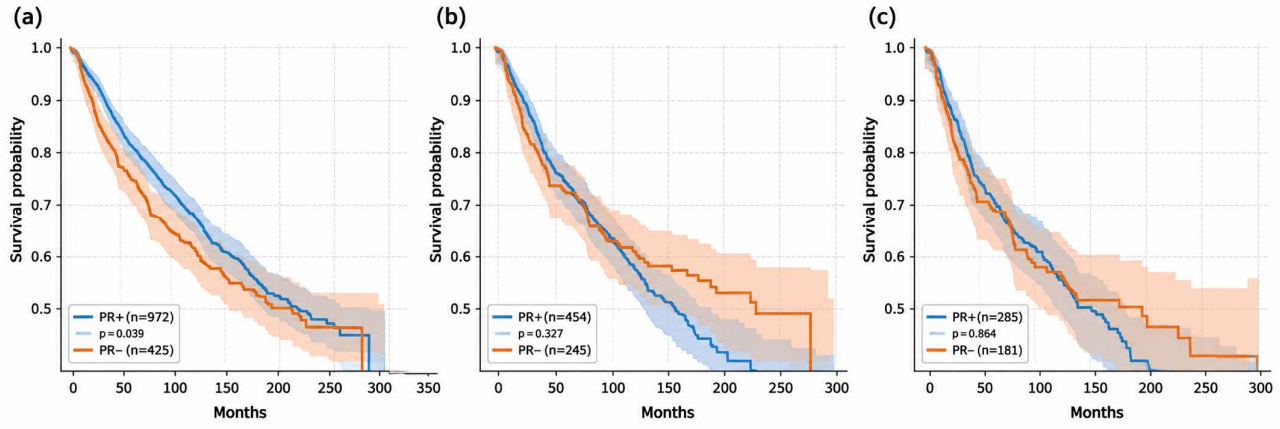

Fig. 4: Context-dependent recurrence-free survival according to PR status in METABRIC.
(a) Full ER-positive/HER2-negative cohort; (b) GENE70-derived median high-risk subgroup; and (c) GENE70-derived upper-tertile high-risk subgroup. Kaplan–Meier curves show recurrence-free survival according to PR status in each subgroup.

Table 2: Key survival analyses across METABRIC and GSE96058/SCAN-B.

Dataset	Endpoint	Analysis stratum / model	Marker comparison	N / events	HR (95% CI)	p-value
METABRIC	RFS	Full ER+/HER2−	PR-negative vs PR-positive	1397/535	1.19 (0.99–1.42)	0.063
METABRIC	RFS	Full ER+/HER2−; adjusted	PR-negative vs PR-positive	964/372	0.92 (0.74–1.15)	0.474
METABRIC	RFS	GENE70 median high-risk	PR-negative vs PR-positive	699/312	0.89 (0.70–1.13)	0.331
METABRIC	RFS	GENE70 upper-tertile high-risk	PR-negative vs PR-positive	466/226	0.94 (0.72–1.23)	0.667
GSE96058	OS	Full ER+/HER2−	PR-negative vs PR-positive	2405/213	1.97 (1.26–3.07)	0.003
GSE96058	OS	Full ER+/HER2−; GENE70-adjusted	PR-negative vs PR-positive	2405/213	1.62 (1.03–2.56)	0.038
GSE96058	OS	GENE70 median high-risk	PR-negative vs PR-positive	1203/114	1.99 (1.18–3.36)	0.009
GSE96058	OS	GENE70 upper-tertile high-risk	PR-negative vs PR-positive	802/90	1.77 (1.01–3.11)	0.045
GSE96058	OS	Full ER+/HER2−	PGR-low tertile vs higher	2405/213	1.62 (1.25–2.09)	<0.001
GSE96058	OS	GENE70 median high-risk	PGR-low tertile vs higher	1203/114	1.98 (1.39–2.83)	<0.001
GSE96058	OS	GENE70 upper-tertile high-risk	PGR-low tertile vs higher	802/90	1.83 (1.22–2.75)	0.004

Note: Hazard ratios greater than 1 indicate higher event risk in the marker-positive comparison group. METABRIC analyses use recurrence-free survival, whereas GSE96058/SCAN-B analyses use overall survival. The adjusted METABRIC row includes clinicopathologic variables plus continuous GENE70-derived risk.

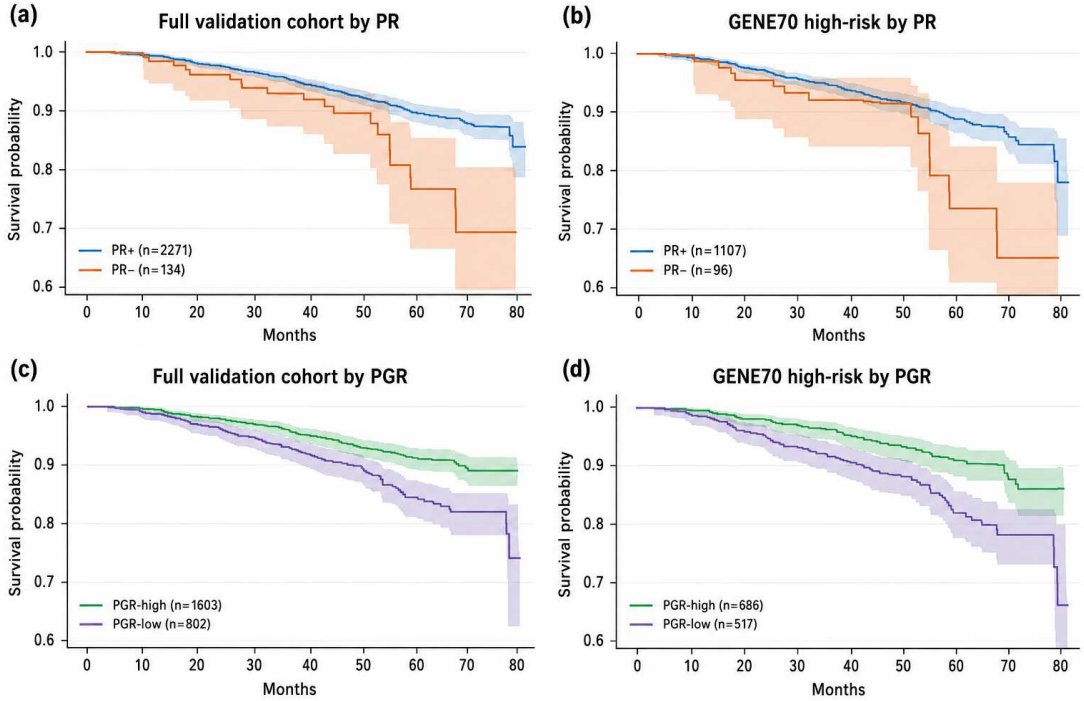

Fig. 5: External clinical validation of PR status and transcriptomic PGR expression in GSE96058/SCAN-B. (a) Overall survival in the full validation cohort according to PR status; (b) overall survival in the GENE70-derived high-risk subgroup according to PR status; (c) overall survival in the full validation cohort according to PGR expression; and (d) overall survival in the GENE70-derived high-risk subgroup according to PGR expression.

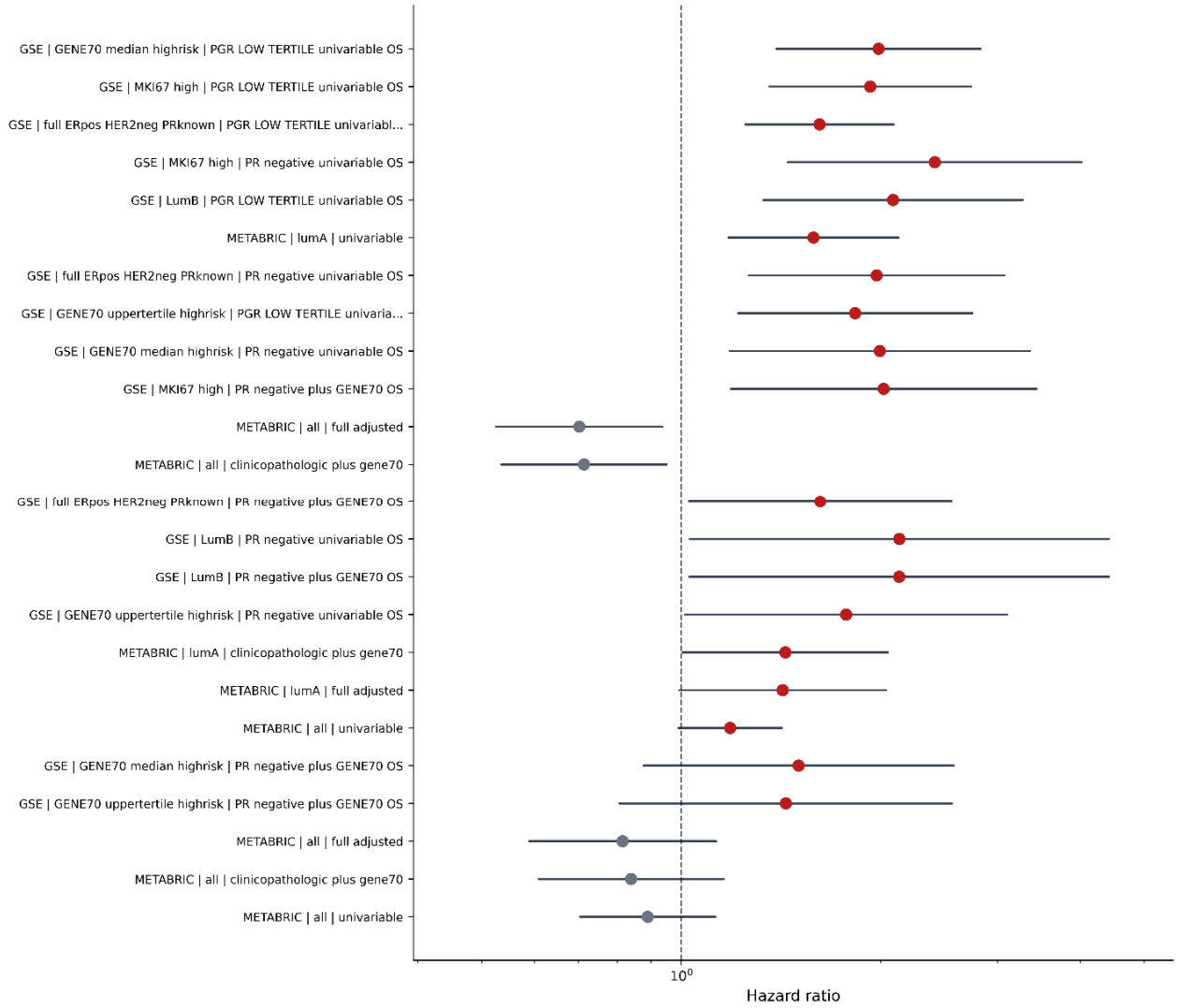

Fig. 6: Integrated survival evidence across cohorts. Forest plot of selected hazard ratios for PR-negative and PGR-low effects across METABRIC recurrence-free survival analyses and GSE96058/SCAN-B overall survival validation models. GSE96058/SCAN-B shows consistent adverse survival effects, especially for PGR-low status, while METABRIC recurrence findings are more heterogeneous.

The corresponding PR-positive and PR-negative group sizes and crude event rates across the METABRIC and GSE96058/SCAN-B cohorts are shown in Fig. S4. The complete METABRIC PR-status survival sensitivity analyses are provided in Table S3. Additional METABRIC recurrence-free survival analyses using alternative PGR expression cut-offs are provided in table S4. Detailed clinicopathologic and molecular comparisons between PR-positive and PR-negative tumours are provided in Table S5. Interaction analyses evaluating whether the PR-negative association differed according to continuous GENE70-derived risk, node-positive status, or MKI67-high status are provided in table S6. The complete external validation results across the evaluated clinical and molecular subgroups are provided in Table S7.

GSE96058 validation

Both clinical PR negativity and low PGR expression were associated with an adverse survival signal with external clinical validation in GSE96058/SCAN-B (Fig. 5). For the OS validation set (full cohort, PR-positive/HER2-negative), PR-negative tumours had inferior OS compared to PR-positive tumours (134 PR-negative cases, 2271 PR-positive cases). This same trend towards worse survival was observed in the GENE70 median high-risk subgroup, in which survival again was worse for PR-negative tumours. Survival was even worse in PGR-low tumours, with significantly worse overall survival in the PGR-low tumours in both the full validation cohort and the PGR median-high risk cohort (GENE70). The visual separation was greater and more uniform in PGR-low status as compared to binary clinical PR negativity, especially in the high-risk validation stratum.

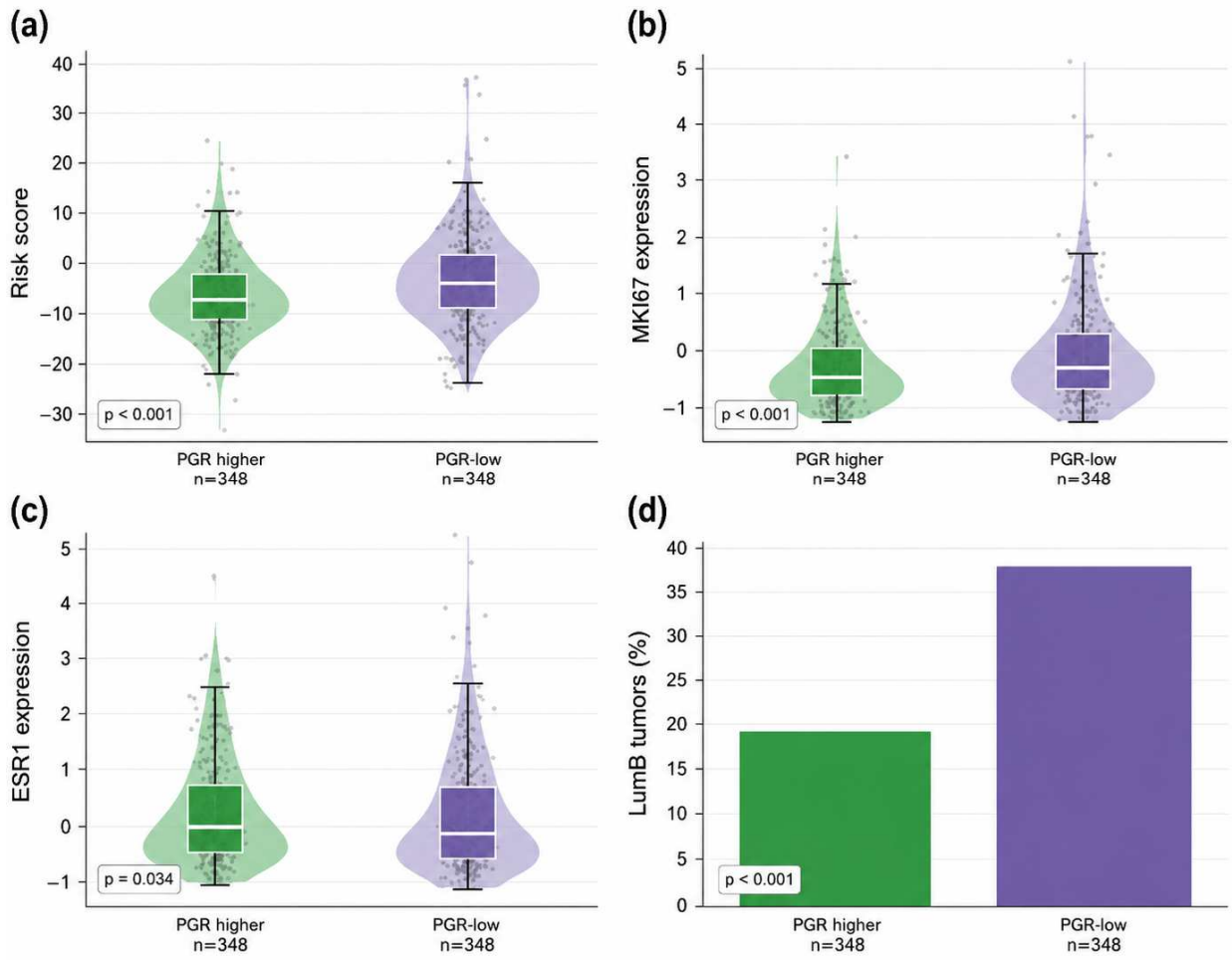

Fig. 7: Independent molecular validation of the PGR-low phenotype in TCGA-BRCA. (a) GENE70-derived genomic-risk score according to PGR expression category; (b) MKI67 expression according to PGR expression category; (c) ESR1 expression according to PGR expression category; and (d) Luminal B tumour enrichment according to PGR expression category.

The distribution of GENE70-derived genomic-risk scores according to PGR-low status across the three cohorts is shown in Fig. S5. The corresponding distributions of MKI67 expression between PGR-low and PGR-higher tumours across the three cohorts are shown in Fig. S6. The corresponding ESR1 expression distributions between PGR-low and PGR-higher tumours across the three cohorts are shown in Fig. S7. Additional Kaplan–Meier analyses of overall survival according to clinical PR status across selected GSE96058/SCAN-B subgroups are shown in Fig. S8. Additional Kaplan–Meier analyses of overall survival according to PGR-low status across selected GSE96058/SCAN-B subgroups are shown in Fig. S9. The complete external validation results across the evaluated clinical and molecular subgroups are provided in Table S7. The primary external clinical evidence supporting the prognostic relevance of PR/PGR loss is shown in figure 5 and Table 2: PGR-low is the more common expression variant.Integrated survival evidence The integrated cross-cohort survival forest plot revealed that GSE96058/SCAN-B had the most robust and significant negative evidence for survival, particularly for the PGR-low status in the whole validation cohort, the median high-risk subgroup, the LumB subgroup and the MKI67-high subgroup. The PR-negative effects in GSE96058/SCAN-B were also mostly negative, with no difference in precision between the full and the high-risk validation strata. METABRIC revealed a more complex pattern: the LumA group and some full-cohort comparisons were directionally contrary, while several models that were adjusted (full cohort or high-risk) models did not provide uniform support for a PR-negative adverse recurrence effect. Transcriptomic PGR-low expression demonstrated consistent prognostic associations across the analyzed cohorts, whereas the prognostic associations of clinical PR negativity were more heterogeneous across datasets and clinical endpoints. Because formal predictive model comparisons were not performed, these findings should not be interpreted as evidence of superior prognostic performance of transcriptomic PGR. These are summarized in Fig. 6 and Table 2.

Table 3: TCGA-BRCA molecular validation of PGR-low luminal tumours.

Molecular feature	Comparison / test	PGR-low median	PGR-higher median	Effect estimate	Odds ratio	p-value
GENE70-derived risk score	PGR-low vs PGR-higher	-3.68	-7.17	3.48	—	<0.001
MKI67 expression	PGR-low vs PGR-higher	-0.25	-0.42	0.17	—	<0.001
ESR1 expression	PGR-low vs PGR-higher	-0.11	0.02	-0.13	—	0.034
ERBB2 expression	PGR-low vs PGR-higher	0.11	-0.05	0.15	—	0.051
LumB subtype enrichment	PGR-low vs PGR-higher	—	—	—	2.66	<0.001
PGR correlation with GENE70-derived risk score	Continuous PGR	—	—	ρ -0.19	—	<0.001
PGR correlation with MKI67 expression	Continuous PGR	—	—	ρ -0.12	—	0.001
PGR correlation with ESR1 expression	Continuous PGR	—	—	ρ 0.09	—	0.014

Note. Median differences are calculated as PGR-low minus PGR-higher. Spearman rho values are shown for continuous PGR expression correlations. TCGA-BRCA is reported as molecular validation because clinical ER/PR/HER2 receptor fields were not available in the selected clinical tables.

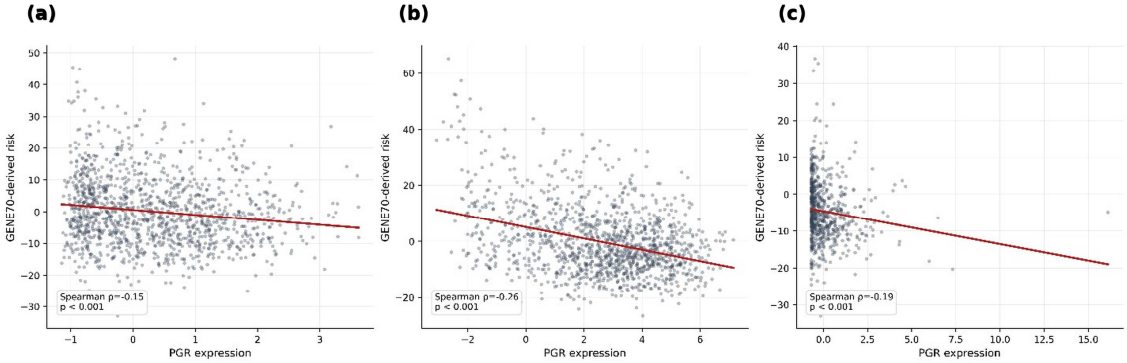

Fig. 8: Inverse association between PGR expression and GENE70-derived genomic risk across three breast cancer cohorts. (a) METABRIC; (b) GSE96058/SCAN-B; (c) TCGA-BRCA. Scatter plots with fitted trend lines show an inverse association between continuous PGR expression and GENE70-derived genomic-risk scores across all three cohorts.

TCGA molecular validation

TCGA-BRCA molecularly validated the PGR-low phenotype of luminal tumours. There were 696 tumours in the TCGA luminal validation cohort, of which 499 were LumA and 197 were LumB. The corresponding correlations between continuous PGR expression and selected molecular markers in the TCGA-BRCA luminal cohort are shown in Fig. S10. The direction and magnitude of the molecular differences between PGR-low and PGR-higher luminal tumours are additionally illustrated in Fig. S11. The recovery of GENE70-matched genes across the three expression datasets is summarized in Fig. S12. The strongest model signals across the survival and molecular analyses are summarized in Fig. S13. Sensitivity analyses comparing median-based and tertile-based PGR-low definitions in the GSE96058/SCAN-B cohort are shown in Fig. S14. There was a significant difference between the scores derived from the GENE70 and the expression of the tumour markers MKI67 and ESR1, with PGR-low tumours having higher values than PGR-higher tumours and LumB tumours being more abundant. The proportion of LumB tumours was greater in the PGR-low group than in the PGR-higher group and the LumB enrichment was visually significant. These results demonstrate the molecular reproducibility of the PGR-low phenotype across multiple datasets and further reveal that PGR-low status predicts increased genomic risk and proliferative luminal biology, even when TCGA is not used as a clinical PR-status survival validation cohort. Results from the molecular validation analyses done by the TCGA are shown in Fig. 7 and Table 3.

PGR and genomic risk

In all 3 cohorts (METABRIC, GSE96058/SCAN-B and TCGA-BRCA), continuous PGR expression showed an inverse correlation with genomic risk as determined by GENE70. All three cohorts showed an inverse relationship, with a Spearman correlation of METABRIC: ρ = -0.15, GSE96058: ρ = -0.26 and TCGA-BRCA: ρ = −0.19. This cross-cohort pattern was seen when PGR expression was lower and GENE70-derived risk was higher; regardless of the dataset platform or the cohort role.

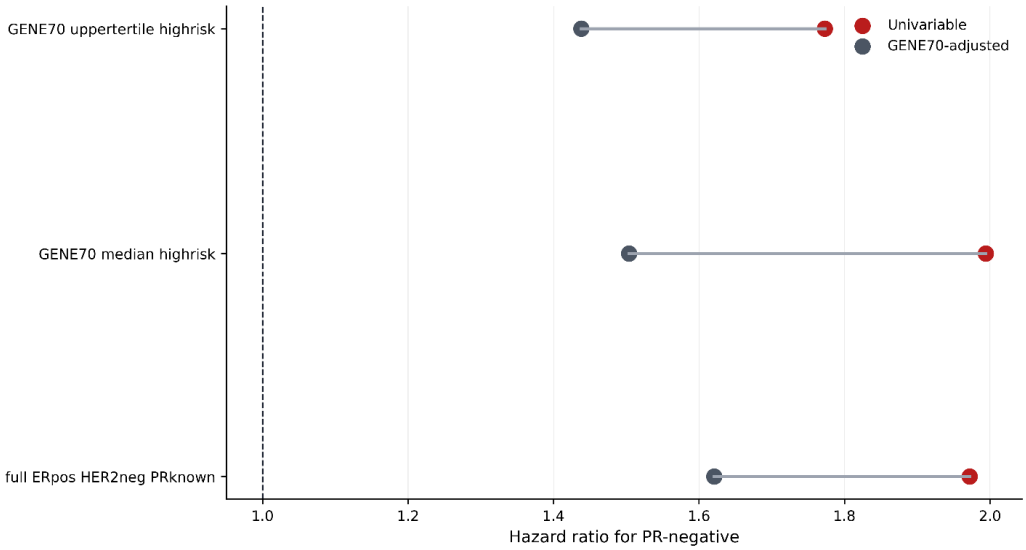

Fig. 9: PR-negative effect attenuates after GENE70 adjustment. Dumbbell plot comparing univariable and GENE70-adjusted hazard ratios for PR-negative status in GSE96058/SCAN-B full and high-risk validation strata. PR-negative hazard ratios move toward the null after continuous GENE70-derived risk adjustment, indicating partial overlap between PR loss and genomic-risk burden.

Table 4: Effect attenuation after continuous GENE70-derived risk adjustment in GSE96058/SCAN-B.

GSE96058 validation stratum	N / events	Univariable HR (95% CI)	p-value	GENE70-adjusted HR (95% CI)	p-value	Log-HR attenuation (%)
Full ER+/HER2−	2405/213	1.97 (1.26–3.07)	0.003	1.62 (1.03–2.56)	0.038	28.9
GENE70 median high-risk	1203/114	1.99 (1.18–3.36)	0.009	1.50 (0.88–2.58)	0.138	40.9
GENE70 upper-tertile high-risk	802/90	1.77 (1.01–3.11)	0.045	1.44 (0.81–2.56)	0.218	36.6

Note. Log-HR attenuation is calculated as the percentage reduction in log(HR) after adding continuous GENE70-derived risk to the Cox model. HRs refer to PR-negative versus PR-positive status.

This relationship is consistent, further supporting the thought that PGR-low status is not just a categorical clinical diagnosis, but a transcriptomic marker associated with the genomic-risk phenotype. Figure 8 and Table 3 support a cross-cohort molecular association between PGR expression and GENE70-derived risk.

GENE70 adjustment

The hazard ratio for clinical PR negativity was reduced after adjustment for the continuous GENE70-derived risk in GSE96058/SCAN-B in the entire ER-positive/HER2-negative validation cohort and in both GENE70-derived high-risk strata. The univariable PR-negative HRs were uniformly >1; however, after adjusting for the continuous GENE70-derived risk, the HRs approached the null. This attenuation pattern suggests that there is overlap between clinical PR negativity and the continuous genomic-risk signal measured by the GENE70-based score (Fig. 9; Table 4). The result, however, does not eliminate the PR-negative signal, it only shows that there is some commonality between the PR-negative signal and the higher genomic risk, as measured by GENE70. The attenuation analysis is summarized in figure 9 and Table 4, with the complete univariable and GENE70-adjusted estimates provided in Table S8. The exploratory TCGA-BRCA survival analyses across the available OS, DSS, DFS and PFS endpoints are provided in Table S9.

DISCUSSION

Principal findings

The three-cohort study demonstrates that the loss of progesterone receptors in ER-positive/HER2-negative breast cancer has biological significance, but the meaning of progesterone receptor loss varies according to the method of measurement and the clinical context in which the evaluation is performed. The biological associations of reduced transcriptomic PGR expression with higher GENE70-derived genomic risk, increased proliferation, reduced endocrine signaling and Luminal B enrichment were highly reproducible across the three cohorts. In contrast, the survival findings were cohort-dependent. METABRIC demonstrated only a non-significant trend toward poorer recurrence-free survival for PR-negative tumours and this association was not maintained after multivariable adjustment or within GENE70-derived high-risk subgroups, whereas the independent GSE96058/SCAN-B cohort demonstrated consistent associations between low transcriptomic PGR expression

and poorer overall survival. Accordingly, the findings support the reproducibility of the biological phenotype associated with reduced PGR expression, while indicating that its prognostic associations may vary according to cohort characteristics, clinical endpoints and analytical context. Transcriptomic PGR expression demonstrated consistent biological associations across the analyzed cohorts, whereas the prognostic associations of clinical PR status were more heterogeneous across the available clinical datasets. Because TCGA-BRCA was used exclusively for molecular validation and did not contain complete clinical ER, PR and HER2 receptor annotations in the selected dataset, it was not used for clinical PR-status validation or comparison. Furthermore, the present study was not designed to formally compare the predictive or prognostic performance of transcriptomic PGR and clinical PR status. Therefore, the findings should be interpreted as demonstrating that transcriptomic PGR provides complementary biological information within GENE70-derived genomic-risk categories rather than evidence that it performs better than conventional clinical PR assessment (Bidard *et al*., 2022; Turner *et al*., 2023).

That the loss of PGR correlates with the more aggressive luminal biology is in line with the accepted breast cancer molecular taxonomy. The breast cancer classifications based on the expression of these markers showed that ER-positive disease does not form a biologically homogeneous group and that different subtypes of luminal breast cancer vary in their proliferation, endocrine signaling and outcome (Parker *et al*., 2009; Perou *et al*., 2000). This was followed by the clinically important intrinsic subtype classification of PAM50 work, which further emphasizes the LumA and LumB lower/higher risk distinction (Piccart *et al*., 2021). These findings support this classification, with PGR-low tumours consistently showing greater enrichment in the LumB subtype, higher MKI67 expression and higher GENE70-derived genomic-risk scores. These findings suggest that reduced transcriptomic PGR expression reflects a less endocrine-differentiated and more proliferative tumour phenotype. However, the present analyses cannot determine whether reduced PGR expression acts as an independent biological driver of tumour progression or primarily represents a downstream marker of diminished endocrine pathway activity. Because PGR is a well-established estrogen receptor–regulated gene, decreased transcriptomic PGR expression may reflect impaired endocrine signaling and loss of luminal differentiation rather than exerting a direct causal effect on tumour aggressiveness. Accordingly, the observed associations should be interpreted as evidence that transcriptomic PGR functions as a biological indicator of endocrine pathway activity, while mechanistic studies are required to determine whether PGR itself contributes independently to disease progression.

Relationship to GENE70-derived genomic risk

The present study deliberately uses a GENE70-derived genomic risk score rather than the commercial MammaPrint result. Public transcriptomic datasets do not contain the complete proprietary assay or its implementation algorithm, making direct reproduction of the commercial test impossible. Therefore, a reconstructed score based on the published GENE70 genes available within each dataset was generated to investigate biological associations rather than to replicate the clinical assay. Similar reconstruction strategies have been widely applied in secondary analyses of public transcriptomic cohorts when complete commercial genomic signatures are unavailable. Consequently, the findings should be interpreted as reflecting the biological behavior of a reconstructed GENE70-derived genomic risk model rather than as analytical validation of the commercial MammaPrint assay. Continuous PGR expression showed a consistent inverse correlation with GENE70-derived genomic risk across all three cohorts despite differences in transcriptomic platforms. These findings support the reproducibility of the biological association between reduced PGR expression and a higher genomic risk burden.

Clinical PR status versus transcriptomic PGR expression

Clinical PR status determined by immunohistochemistry and transcriptomic PGR expression represent related but biologically distinct biomarkers. Clinical PR remains an established component of routine pathological evaluation, whereas transcriptomic PGR reflects continuous variation in endocrine pathway activity at the gene-expression level. In the present study, transcriptomic PGR demonstrated consistent associations with genomic risk burden, proliferative activity and survival across multiple cohorts. However, because formal comparative analyses such as likelihood-ratio testing, concordance-index evaluation, or multivariable model comparison including both biomarkers were not performed, the findings should not be interpreted as evidence that transcriptomic PGR provides superior prognostic performance compared with conventional clinical PR assessment (Sorlie *et al*., 2001; Sparano *et al*., 2018).

Survival findings and cohort heterogeneity

Survival findings differed between cohorts. METABRIC demonstrated heterogeneous recurrence-free survival associations, whereas GSE96058/SCAN-B consistently showed poorer overall survival among tumours with low transcriptomic PGR expression. These differences are likely attributable to variations in study design, transcriptomic platforms, clinical endpoints, follow-up duration and patient characteristics, rather than to contradictory biological findings. GSE96058/SCAN-B provided external clinical validation of the adverse prognostic association with low transcriptomic PGR expression, whereas METABRIC and TCGA-BRCA

served complementary roles in discovery and molecular validation. Despite methodological differences among the cohorts, the biological associations between reduced PGR expression, higher GENE70-derived genomic risk, increased proliferation, reduced estrogen-signaling activity and Luminal B enrichment were consistently observed.

Molecular validation in TCGA-BRCA

TCGA-BRCA provided independent molecular validation of the transcriptomic PGR phenotype. PGR-low luminal tumours consistently demonstrated higher GENE70-derived genomic risk, increased MKI67 expression, reduced ESR1 expression and enrichment of the Luminal B subtype (Fig. 7; Table 3). These findings are supported by the supplementary molecular analyses, including the distribution of GENE70-derived genomic-risk scores, MKI67 expression and ESR1 expression according to PGR expression category. Additional TCGA-BRCA molecular comparisons and correlation analyses are provided in Table S10. Although TCGA-BRCA could not be used to validate clinical PR status because complete receptor annotations were unavailable, the molecular findings independently supported the biological reproducibility of the PGR-low phenotype across datasets (Sparano *et al*., 2019).

Biological interpretation

Biologically, what this means is that PGR-low tumours could be a more proliferative and less endocrine differentiated group of ER-positive/HER2-negative breast cancer. Progesterone receptor expression is an estrogen-regulated marker and absence of PR in ER-positive disease is often considered a sign of down-regulation of the ER pathway, activation of growth factor pathways, or endocrine resistance biology. Endocrine resistance has been reviewed and highlights that there are several ways breast cancer can become resistant to endocrine therapy, such as alteration of ER signaling, activation of growth factor pathways, activation of PI3K/AKT/mTOR pathways and activation of cell cycle pathways (Tutt *et al*., 2021). This observation is congruent with this model, as PGR-low tumours showed greater risk (GENE70) and greater expression of MKI67 in the present study, though these were not directly tested as causal mechanisms for endocrine resistance.

The results also correlate with the clinical landscape of ER+ breast cancer being heterogeneous and the increasing use of clinicopathologic features, receptor status, proliferation and multigene assays in the treatment decision. Endocrine therapy, chemotherapy, targeted therapies, menopausal status, nodal status, tumour burden and the risk of recurrence are all complex factors that need to be taken into consideration in hormone receptor-positive disease (Turner *et al*., 2023; Loibl *et al*., 2024; Turner *et al*., 2023). No new clinical decision rule is proposed; rather, the findings provide evidence that PGR-low status may contribute to the more aggressive biological behavior observed in a subset of ER-positive/HER2-negative tumours within this clinically similar population (Johnston *et al*., 2023; Loibl *et al*., 2024).

Clinical and research implications

The first thing that arises in the mind of the clinical researcher is that PGR-low status should not be used for eliminating known genomic assays or receptor testing. Instead, the data support the hypothesis that PGR-low expression may be used to better interpret risk, particularly in ER-positive/HER2-negative disease, where a tumour may already have a higher GENE70-derived genomic risk. External validation showed that PGR-low was a negative predictor of overall survival in the GSE96058/SCAN-B cohort and in high-risk subsets identified by GENE70. This implies that PGR expression may be included in future models that integrate the expression of multiple genes with endocrine receptor biology. Nevertheless, before clinical translation, any model with PGRs would need to be validated in a prospective cohort, have established thresholds for each assay, be evaluated for each treatment and determine if such a model would be more useful in making a decision than the currently used multigene model.

The study offers a helpful research-based framework for distinguishing between clinical PR status and the transcriptomic expression of PGRs. Future research is needed to assess whether PGR-low status enhances prediction of response to endocrine therapy, benefit from chemotherapy and patterns of late and/or metastatic recurrence. The correlation of PGR-low status with cell-cycle activity also suggests that experiments combining PGR expression with proliferation signatures, endocrine therapy exposure, ESR1 alteration, PI3K pathway alteration and PGR inhibitor response may be beneficial. This type of work could help determine whether PGR-low is simply associated with aggressive luminal biology or can also be used to inform treatment choices.

Pharmacological implications

Low progesterone receptor (PGR)-expressing tumours tend to have characteristics of a greater proliferation rate, decreased levels of the estrogen receptor (ESR1) and a predominance of the more aggressive Luminal B subtype, as well as an increased risk rating according to the GENE70 molecular signature, all of which would indicate a biological state with diminished endocrine differentiation. Although treatment-response data were not available for direct evaluation, the observed associations between transcriptomic PGR expression, genomic-risk burden, proliferation and endocrine-signaling activity are biologically consistent with mechanisms that have previously been linked to endocrine resistance. Nevertheless, the present study evaluated biological and prognostic associations only and was not designed to determine whether transcriptomic PGR predicts response to endocrine therapy, CDK4/6

inhibitors, chemotherapy, or other targeted treatments. Therefore, transcriptomic PGR should currently be regarded as a complementary biological and prognostic biomarker, while its potential role as a pharmacogenomic or predictive companion biomarker requires confirmation in prospective studies that incorporate standardized treatment-response data.

Limitations

This study has several limitations. First, it was a retrospective analysis based on publicly available datasets. Second, the GENE70-derived genomic risk score was reconstructed from publicly available transcriptomic data and should not be interpreted as equivalent to the commercial MammaPrint assay. Third, treatment-response data were unavailable, preventing evaluation of predictive or pharmacogenomic utility. Finally, prospective studies using standardized clinical cohorts are required to validate these findings before clinical implementation.

CONCLUSION

Reduced transcriptomic PGR expression was consistently associated with higher GENE70-derived genomic risk, increased proliferative activity, reduced endocrine signaling and enrichment of the Luminal B subtype across three independent cohorts of ER-positive/HER2-negative breast cancer. Although survival associations varied across datasets, the external validation cohort confirmed that low PGR expression was associated with poorer overall survival, supporting its biological and prognostic relevance within established genomic risk categories. These findings indicate that transcriptomic PGR provides complementary biological information that may improve the interpretation of GENE70-derived genomic risk stratification rather than replace conventional clinical progesterone receptor assessment. Because treatment-response outcomes were not evaluated and the genomic-risk score was reconstructed from publicly available datasets, the present findings should not be interpreted as evidence of predictive or pharmacogenomic utility. Prospective studies incorporating standardized treatment-response data and validated genomic assays are required before clinical application.

Acknowledgements

The authors acknowledge the investigators and participants of the METABRIC, GSE96058/SCAN-B and TCGA-BRCA studies for making clinical and molecular breast cancer data publicly available. The authors also acknowledge cBioPortal/DataHub, the Gene Expression Omnibus, The Cancer Genome Atlas and the Bioconductor genefu resource for enabling reproducible secondary analysis of publicly available datasets. No individually identifiable patient-level data were accessed during this study.

Author's contributions

Orcun Can: Conceptualized the study, designed the methodology, performed data collection, statistical analyses, interpretation of the results, visualization and drafted the original manuscript. Reviewed, revised and approved the final version of the manuscript and agrees to be accountable for all aspects of the work.

Funding

There was no funding.

Data availability statement

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethical approval

This study used publicly available, de-identified secondary datasets and did not involve new patient recruitment, intervention, biospecimen collection, or access to identifiable patient information. Therefore, additional institutional ethics approval and informed consent were not required for the present secondary analysis. This study was performed in adherence with the REMARK and TRIPOD guidelines. See supplementary file for the REMARK and TRIPOD checklist.

Conflict of interest

The authors declare no conflict of interest.

Availability of data and materials

All data analysed in this study are publicly available. METABRIC data were obtained from cBioPortal at https://www.cbioportal.org/study/summary?id=brca_metabric
and are also available through the cBioPortal DataHub at https://github.com/cBioPortal/datahub/tree/master/public/brca_metabric
TCGA-BRCA data were obtained from cBioPortal at https://www.cbioportal.org/study/summary?id=brca_tcga_pan_can_atlas_2018 and are available through the cBioPortal DataHub at https://github.com/cBioPortal/datahub/tree/master/public/brca_tcga_pan_can_atlas_2018 GSE96058/SCAN-B data were obtained from the Gene Expression Omnibus under accession GSE96058: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE96058 .
The GENE70 signature was extracted from the public Bioconductor genefu package source available at https://bioconductor.org/packages/genefu/ and https://github.com/bhklab/genefu. Processed analysis tables, figure source data and generated outputs are available from the corresponding author upon reasonable request.

Code availability

The analysis was performed using Python-based scripts in Google Colab. The original datasets analysed in this study

are publicly available through cBioPortal and the Gene Expression Omnibus (GSE96058), subject to their respective terms of use. The Python scripts used for data processing, GENE70-derived genomic-risk reconstruction, statistical analyses, survival analyses, molecular validation, figure generation and processed analysis outputs are available from the corresponding author upon reasonable request. The authors acknowledge that deposition in a public repository would further enhance reproducibility and intend to make the analysis workflow publicly available in a future repository update. Formal assessment of the proportional hazards assumption using Schoenfeld residuals showed no evidence of a clinically meaningful violation in the Cox regression models used in this study.

Supplementary data

https://pjps.pk/uploads/2026/08/SUP1787479351.pdf

REFERENCES

Allison KH and Rendi MH (2022). Updates on breast biomarkers. *Virchows Arch.*, **480**(3): 453-468.

Allison KH, Hammond MEH, Dowsett M, McKernin SE, Carey LA, Fitzgibbons PL, Hayes DF, Lakhani SR, Chavez-MacGregor M, Perlmutter J, Perou CM, Regan MM, Rimm DL, Symmans WF, Torlakovic EE, Varella L, Viale G, Weisberg TF, McShane LM and Wolff AC (2020). Estrogen and progesterone receptor testing in breast cancer: ASCO/CAP guideline update. *J. Clin. Oncol.*, **38**(12): 1346-1366.

Andre F, Ismaila N, Allison KH, Barlow WE, Collyar DE, Damodaran S, Henry NL, Jhaveri K, Kalinsky K, Kuderer NM, Litvak A, Mayer EL, Pusztai L, Raab R, Wolff AC and Stearns V (2022). Biomarkers for adjuvant endocrine and chemotherapy in early-stage breast cancer: ASCO guideline update. *J. Clin. Oncol.*, **40**(16): 1816-1837.

Bidard FC, Kaklamani VG, Neven P, Streich G, Montero AJ, Forget F, Mouret-Reynier MA, Sohn JH, Taylor D, Harnden KK, Khong H, Kocsis J, Dalenc F, Dillon PM, Babu S, Waters S, Deleu I, Garcia Saenz JA, Bria E, Cazzaniga M, Lu J, Aftimos P, Cortes J, Liu S, Tonini G, Laurent D, Habboubi N, Conlan MG and Bardia A (2022). Elacestrant (oral selective estrogen receptor degrader) versus standard endocrine therapy for estrogen receptor-positive, human epidermal growth factor receptor 2-negative advanced breast cancer: Results from the randomized phase III EMERALD trial. *J. Clin. Oncol.*, **40**(28): 3246-3256.

Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I and Jemal A (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.*, **74**(3): 229-263.

Burstein HJ, Curigliano G, Thurlimann B, Weber WP, Poortmans P, Regan MM, Senn HJ, Winer EP and Gnant M (2021). Customizing local and systemic therapies for women with early breast cancer: The St. Gallen International Consensus Guidelines for treatment of early breast cancer 2021. *Ann. Oncol.*, **32**(10): 1216-1235.

Cardoso F, van't Veer LJ, Bogaerts J, Slaets L, Viale G, Delaloge S, Pierga JY, Brain E, Causeret S, DeLorenzi M, Glas AM, Golfinopoulos V, Goulioti T, Knox S, Matos E, Meulemans B, Neijenhuis PA, Nitz U, Passalacqua R, Ravdin P, Rubio IT, Saghatchian M, Smilde TJ, Sotiriou C, Stork L, Straehle C, Thomas G, Thompson AM, van der Hoeven JM, Vuylsteke P, Bernards R, Tryfonidis K, Rutgers E and Piccart M (2016). 70-gene signature as an aid to treatment decisions in early-stage breast cancer. *N. Engl. J. Med.*, **375**(8): 717-729.

Drukker CA, Bueno-de-Mesquita JM, Retel VP, van Harten WH, Tinteren HV, Wesseling J, Rutgers EJT, van de Vijver MJ and van't Veer LJ (2013). A prospective evaluation of a breast cancer prognosis signature in the observational RASTER study. *Int. J. Cancer*, **133**(4): 929-936.

Giaquinto AN, Sung H, Newman LA, Freedman RA, Smith RA, Star J, Jemal A and Siegel RL (2024). Breast cancer statistics, 2024. *CA Cancer J. Clin.*, **74**(6): 477-495.

Hanker AB, Sudhan DR and Arteaga CL (2020). Overcoming endocrine resistance in breast cancer. *Cancer Cell*, **37**(4): 496-513.

Harbeck N, Penault-Llorca F, Cortes J, Gnant M, Houssami N, Poortmans P, Ruddy K, Tsang J and Cardoso F (2019). Breast cancer. *Nat. Rev. Dis. Primers*, **5**(1): 66.

Henry NL, Somerfield MR, Dayao Z, Elias A, Kalinsky K, McShane LM, Moy B, Park BH, Shanahan KM, Sharma P, Shatsky R, Stringer-Reasor E, Telli M, Turner NC and DeMichele A (2022). Biomarkers for systemic therapy in metastatic breast cancer: ASCO guideline update. *J. Clin. Oncol.*, **40**(27): 3205-3221.

Horwitz KB and Sartorius CA (2008). Progestins in hormone replacement therapies reactivate cancer stem cells in women with preexisting breast cancers: A hypothesis. *J. Clin. Endocrinol. Metab.*, **93**(9): 3295-3298.

Johnston SRD, Toi M, O'Shaughnessy J, Rastogi P, Campone M, Neven P, Huang CS, Huober J, Garnica Jaliffe G, Cicin I, Tolaney SM, Goetz MP, Rugo HS, Senkus E, Testa L, Del Mastro L, Shimizu C, Wei R, Shahir A, Munoz M, San Antonio B andré V, Harbeck N and Martin M (2023). Abemaciclib plus endocrine therapy for hormone receptor-positive, HER2-negative, node-positive, high-risk early breast cancer (monarchE): Results from a preplanned interim analysis of a randomised, open-label, phase 3 trial. *Lancet Oncol.*, **24**(1): 77-90.

Kalinsky K, Barlow WE, Gralow JR, Meric-Bernstam F, Albain KS, Hayes DF, Lin NU, Perez EA, Goldstein

LJ, Chia SKL, Dhesy-Thind S, Rastogi P, Alba E, Delaloge S, Martin M, Kelly CM, Ruiz-Borrego M, Gil-Gil M, Arce-Salinas CH, Brain EGC, Lee ES, Pierga JY, Bermejo B, Ramos-Vazquez M, Jung KH, Ferrero JM, Schott AF, Shak S, Sharma P, Lew DL, Miao J, Tripathy D, Pusztai L and Hortobagyi GN (2021). 21-gene assay to inform chemotherapy benefit in node-positive breast cancer. *N. Engl. J. Med.*, **385**(25): 2336-2347.

Lloyd MR, Jhaveri K, Kalinsky K, Bardia A and Wander SA (2024). Precision therapeutics and emerging strategies for hormone receptor-positive metastatic breast cancer. *Nat. Rev. Clin. Oncol.*, **21**: 743-761.

Loibl S andre F, Bachelot T, Barrios CH, Bergh J, Burstein HJ, Cardoso MJ, Carey LA, Dawood S, Del Mastro L, Denkert C, Fallenberg EM, Francis PA, Gamal-Eldin H, Gelmon K, Geyer CE, Gnant M, Guarneri V, Gupta S, Kim SB, Krug D, Martin M, Meattini I, Morrow M, Janni W, Paluch-Shimon S, Partridge A, Poortmans P, Pusztai L, Regan MM, Sparano J, Spanic T, Swain SM, Tjulandin S, Toi M, Trapani D, Tutt A, Xu B and Curigliano G (2024). Early breast cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Ann. Oncol.*, **35**(2): 159-182.

Nielsen TO, Leung SCY, Rimm DL, Dodson A, Acs B, Badve S, Denkert C, Ellis MJ, Fineberg S, Flowers M, Kreipe HH, Laenkholm AV, Pan H, Penault-Llorca F, Polley MY, Salgado R, Smith IE, Sugie T, Bartlett JMS, McShane LM, Dowsett M and Hayes DF (2021). Assessment of Ki67 in breast cancer: Updated recommendations from the International Ki67 in Breast Cancer Working Group. *J. Natl. Cancer Inst.*, **113**(7): 808-819.

Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, Quackenbush JF, Stijleman IJ, Palazzo J, Marron JS, Nobel AB, Mardis E, Nielsen TO, Ellis MJ, Perou CM and Bernard PS (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.*, **27**(8): 1160-1167.

Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale AL, Brown PO and Botstein D (2000). Molecular portraits of human breast tumours. *Nature*, **406**(6797): 747-752.

Piccart M, van 't Veer LJ, Poncet C, Cardoso F, Delaloge S, Pierga JY, Vuylsteke P, Nitz U, Pusztai L, Saghatchian M, Forbes J, Tryfonidis K, Litière S, Aftimos P, Bolla M, Bogaerts J and Rutgers E (2021). 70-gene signature as an aid for treatment decisions in early breast cancer: Updated results of the phase 3 randomised MINDACT trial with an exploratory analysis by age. *Lancet Oncol.*, **22**(4): 476-488.

Prat A, Cheang MCU, Martín M, Parker JS, Carrasco E, Caballero R, Tyldesley S, Gelmon K, Bernard PS, Nielsen TO and Perou CM (2013). Prognostic significance of progesterone receptor-positive tumour cells within immunohistochemically defined luminal A breast cancer. *J. Clin. Oncol.*, **31**(2): 203-209.

Purdie CA, Quinlan P, Jordan LB, Ashfield A, Ogston S, Dewar JA and Thompson AM (2014). Progesterone receptor expression is an independent prognostic variable in early breast cancer: A population-based study. *Br. J. Cancer*, **110**(3): 565-572.

Slamon DJ, Lipatov O, Nowecki Z, McAndrew N, Kukielka-Budny B, Stroyakovskiy D, Yardley DA, Huang CS, Fasching PA, Crown J, Bardia A, Chia S, Im SA, Ruiz-Borrego M, Loi S, Xu B, Hurvitz S, Barrios C, Untch M, Moroose R, Visco F, Afenjar K, Fresco R, Severin I, Ji Y, Ghaznawi F, Li Z, Zarate JP, Chakravartty A, Taran T and Hortobagyi GN (2024). Ribociclib plus endocrine therapy in early breast cancer. *N. Engl. J. Med.*, **390**(12): 1080-1091.

Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Lønning PE and Børresen-Dale AL (2001). Gene expression patterns of breast carcinomas distinguish tumour subclasses with clinical implications. *Proc. Natl. Acad. Sci. U.S.A.*, **98**(19): 10869-10874.

Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, Geyer CE Jr, Dees EC, Goetz MP, Olson JA Jr, Lively T, Badve SS, Saphner TJ, Wagner LI, Whelan TJ, Ellis MJ, Paik S, Wood WC, Ravdin PM, Keane MM, Gomez Moreno HL, Reddy PS, Goggins TF, Mayer IA, Brufsky AM, Toppmeyer DL, Kaklamani VG, Berenberg JL, Abrams J and Sledge GW Jr (2018). Adjuvant chemotherapy guided by a 21-gene expression assay in breast cancer. *N. Engl. J. Med.*, **379**(2): 111-121.

Sparano JA, Gray RJ, Ravdin PM, Makower DF, Pritchard KI, Albain KS, Hayes DF, Geyer CE Jr, Dees EC, Goetz MP, Olson JA Jr, Lively T, Badve SS, Saphner TJ, Wagner LI, Whelan TJ, Ellis MJ, Paik S, Wood WC, Keane MM, Gomez Moreno HL, Reddy PS, Goggins TF, Mayer IA, Brufsky AM, Toppmeyer DL, Kaklamani VG, Berenberg JL, Abrams J and Sledge GW Jr (2019). Clinical and genomic risk to guide the use of adjuvant therapy for breast cancer. *N. Engl. J. Med.*, **380**(25): 2395-2405.

The Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, **490**(7418): 61-70.

Tutt ANJ, Garber JE, Kaufman B, Viale G, Fumagalli D, Rastogi P, Gelber RD, de Azambuja E, Fielding A, Balmaña J, Domchek SM, Gelmon KA, Hollingsworth SJ, Korde LA, Linderholm B, Bandos H, Senkus E, Suga JM, Shao Z, Pippas AW, Nowecki Z, Huzarski T, Ganz PA, Lucas PC, Baker N, Loibl S, McConnell R, Piccart M, Schmutzler R, Steger GG, Costantino JP, Arahmani A, Wolmark N, McFadden E, Karantza V,

Lakhani SR, Yothers G, Campbell C and Geyer CE Jr (2021). Adjuvant olaparib for patients with BRCA1- or BRCA2-mutated breast cancer. *N. Engl. J. Med.*, **384**(25): 2394-2405.

Turner NC, Oliveira M, Howell SJ, Dalenc F, Cortes J, Gomez Moreno HL, Hu X, Jhaveri K, Krivorotko P, Loibl S, Morales Murillo S, Okera M, Park YH, Sohn J, Toi M, Tokunaga E, Yousef S, Zhukova L, de Bruin EC, Grinsted L, Schiavon G, Foxley A and Rugo HS (2023). Capivasertib in hormone receptor-positive advanced breast cancer. *N. Engl. J. Med.*, **388**(22): 2058-2070.