CLINICAL TRIAL

Association between *FGFR1* copy numbers, *MAP3K1* mutations, and survival in axillary node-positive, hormone receptor-positive, and HER2-negative early breast cancer in the PACS04 and METABRIC studies

Dimitri Carene[1,2] · Alicia Tran-Dien[2] · Jérôme Lemonnier[3] · Florence Dalenc[4] · Christelle Levy[5] · Jean-Yves Pierga[6] · William Jacot[7] · Jean-Luc Canon[8] · Catherine Richon[9] · Ludovic Lacroix[2,9] · Christophe Caux[10] · Fabrice André[1,2,11] · Stefan Michiels[12,13,14]

Abstract

Purpose Hormone receptor-positive (HR+) and human epidermal growth factor receptor 2 negative (HER2−) early breast cancer (BC) is the most prevalent BC subtype with substantial biological heterogeneity. Although clinicopathological (CP) characteristics have a clear prognostic value, additional biomarkers could refine survival prediction and guide treatment decision.

Methods Copy number aberrations and somatic driver mutations were obtained with OncoScan CGH array and sequencing of 36 genes on HR+/HER2− node-positive early BC patients treated with chemotherapy from the PACS04 trial. We built a two-gene genomic score (GS) associated with distant disease-free survival (DDFS), whose prognostic value was assessed on the external METABRIC data ($n = 1413$) using overall survival (OS) and breast cancer-specific survival (BCSS).

Results In the PACS04 trial ($n = 327$), the median follow-up for DDFS (65 events) was 9.6 years. *FGFR1* amplifications ($HR_{Amplification} = 2.44$, 95% CI [1.25; 4.76], $p = 0.009$) and *MAP3K1* mutations ($HR_{Mutation} = 0.10$, [0.01; 0.78], $p = 0.03$) were associated with DDFS beyond CP characteristics. A prognostic GS combining *FGFR1* amplifications and *MAP3K1* mutations added more information to CP model ($\chi^2_{DDFS} = 12.97$, $p_{DDFS} < 0.001$ and $\chi^2_{OS} = 11.52$, $p_{OS} < 0.001$). In the METABRIC study ($n = 1413$), *FGFR1* amplifications ($HR_{Amplification} = 2.00$ [1.40; 2.87], $p < 0.001$) and *MAP3K1* mutations ($HR_{Mutation} = 0.58$ [0.41; 0.83], $p = 0.003$) were significantly associated with BCSS beyond CP characteristics. The prognostic GS added significant prognostic information to CP model ($\chi^2_{BCSSS} = 15.39$, $p_{BCSS} < 0.001$ and $\chi^2_{OS} = 5.62$, $p_{OS} = 0.02$).

Conclusion In axillary node-positive, HR+, and HER2− early BC, amplifications of *FGFR1* gene were strongly associated with increased risk for distant disease, while mutations of MAP3K1 gene were significantly associated with decreased risk.

Keywords Copy number aberrations (CNA) · Mutations · Breast cancer (BC) · Biomarkers · Cox regression

Fabrice André and Stefan Michiels contributed equally to this work.

✉ Stefan Michiels
Stefan.michiels@gustaveroussy.fr

Extended author information available on the last page of the article

Introduction

Hormone receptor-positive (HR+) and human epidermal growth factor receptor 2 negative (HER2−) early breast cancer (BC) is the most common BC subtype, accounting for 73% of all BC in the Unites States in 2010 [1]. In this BC subtype, patients with node-positive disease have a worse prognosis than node-negative patients. Patients with early HR+/HER2− node-positive BC are usually treated with adjuvant therapy to reduce the risk of distance disease and recurrence: chemotherapy to treat micro-metastatic disease, radiotherapy to destroy any cancer cells that remain in the
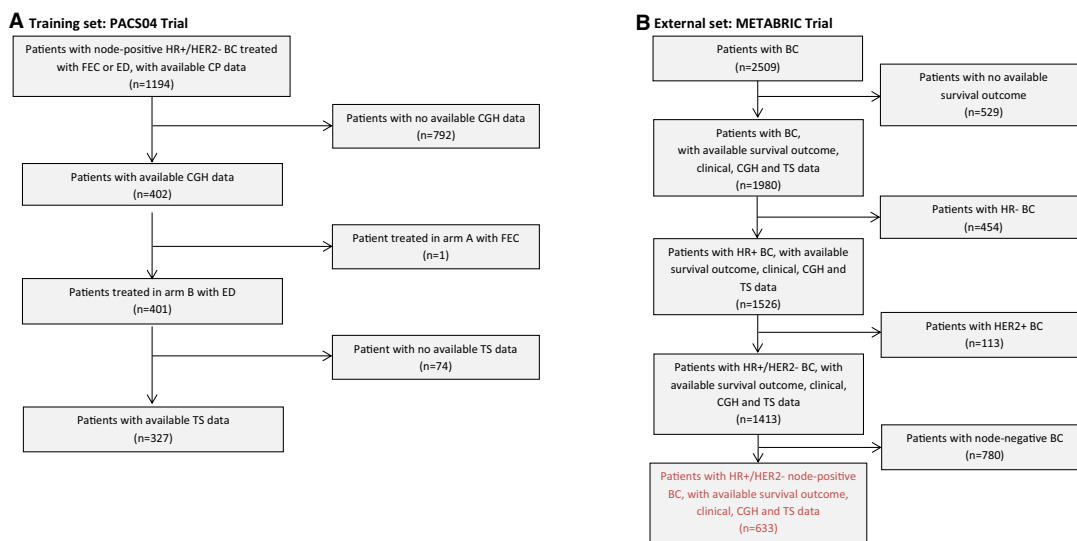
Fig. 1 Flow chart for the training and the external data. *PACS04* Programme Adjuvant Cancer du Sein, *METABRIC* Molecular Taxonomy of Breast Cancer International Consortium, *HR+* hormone receptor positive, *HER2−* human epidermal growth factor receptor 2 negative,

BC breast cancer, *TS* targeted sequencing, *CGH* comparative genomic hybridization, *FEC* 5-fluorouracil (5-FU) + epirubicin + cyclophosphamide, *ED* epirubicin + docetaxel

breast area after surgery, and hormone therapy that is given during at least 5 years to stop estrogen receptor signaling.

Currently, to estimate an individual prognosis for clinical decision-making, clinicopathological (CP) characteristics such as age, tumor size, lymph node status, tumor grade, *HER2* status, proliferation, and percentage of cells expressing estrogen and progesterone receptors are usually considered [2]. Although CP characteristics are prognostic, they do not explain all the variations in survival, as early BC patients may develop distant metastasis [3, 4] during or after chemotherapy or hormone therapy. Hence, in addition to relevant classical CP characteristics, it is of interest to study the prognostic value of additional biomarkers.

Previous studies have shown that high amplifications of *CCND1* (11q13) [5, 6], *FGFR1* (8p12) [7, 8], *ZNF2017* (20q13) [9–11], and *ERBB2* (17q12) [12–14] are associated with a poor outcome and may contribute to hormone therapy resistance. Amplifications of these genes are observed, respectively, in approximately 19.6%, 22.8%, 8.5%, and 9.9% [15] of cases. Other studies have also shown that *PIK3CA* [16], *TP53* [17], and *MAP3K1* [18] mutations are significantly associated with survival and are the most frequently mutated genes in BC with a prevalence of 36%, 37%, and 8%, respectively [19]. The aim of this study was to evaluate which genomic alterations could refine survival prediction. Using data from the PACS04 trial, we evaluated the prognostic value of the copy number of predefined well-known genes (*CCND1*, *FGFR1*, *ZNF217*, and *ERRB2*) and mutations of a panel of driver genes in addition to conventional CP characteristics. We built a genomic score to discriminate

patients with different levels of risk and investigated the prognostic value in the external METABRIC cohort.

Materials and methods

Patients

Training data: PACS04 [20] was a phase III trial that included 1194 patients with node-positive, HR+, and HER2− early BC. After conservative surgery, patients were randomized to receive either six cycles of 5-fluorouracil, cyclophosphamide, and epirubicin (FEC) or six cycles of epirubicin and docetaxel 75 mg/m^2 (ED). Radiotherapy was given after adjuvant chemotherapy followed by hormone therapy. Three hundred and twenty-seven patients treated with ED had clinical, DNA copy number, and somatic mutation information available. Copy numbers were obtained by the OncoScan Comparative Genomic Hybridization (CGH) array and somatic mutations by targeted sequencing (TS).

Validation data: METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) [21] was a Canada–UK project that included 2509 patients with early BC. One thousand four hundred and thirteen patients had HR+ and HER2− primary tumors and had survival outcome, clinical, DNA copy number, and somatic mutation information available. Copy number and somatic driver mutations were generated using Affymetrix SNP 6.0 CGH array and TS of 173 genes. Patient selection for PACS04 and METABRIC data is summarized in Fig. 1.

DNA isolation, hybridization, and sequencing

Tumor DNA was successfully extracted from formalin-fixed and paraffin-embedded (FFPE) samples from PACS04 trial patients. Five hundred samples were collected and hybridization was performed on 433. DNA copy numbers were assessed using Oncoscan FFPE assay kit. The data were normalized and segmented via rCGH [22] algorithm, and the copy number threshold to determine amplification and deletion was, respectively, greater than 4 copies (log-2-ratio > 1) [23] and less than 1 copy (log-2-ratio < − 1) [23]. Targeted next-generation sequencing of a panel of 36 genes was performed on 394 primary breast tumor tissues from PACS04 trial. Only mutations with variant allele frequency (VAF) greater than 10% were considered. A filter method was applied to selected genes with a prevalence of mutation greater than 3% to make sure that the statistical analyses converged.

Statistical analysis

In the training data (PACS04 trial), to compare the quantity of lost information, the differences in patient characteristics between the patients included and excluded from the analysis were tested using, respectively, Kruskal–Wallis and Fisher's exact test for continuous and categorical variables. The primary endpoint for the statistical analysis was distant disease-free survival (DDFS), defined as the time from randomization to the first recurrence (distant site) or death from any cause. The secondary endpoint was overall survival (OS), defined as the time from randomization to death from any cause. Patients who did not experience events were censored at the date of last follow-up. The correlation between CP and genomic features was analyzed using Spearman correlation and correlation coefficients value and 95% confidence intervals (CI) were obtained using 5000 bootstrap repetitions. We checked the hypothesis of log-linearity of continuous covariates in the Cox model using splines with 2 degrees of freedom. A power calculation of this study was performed. Assuming a genomic biomarker with a prevalence of 25% of alteration in the analysis population at a two-sided significance level of 0.05, if the hazard ratio (HR) for DDFS is equal to 2, then the power to reject the null hypothesis that the HR is equal to 1 with a 0.05 significance level will be 77%. In univariable analysis, the prognostic value of each genomic characteristic on DDFS was performed using Cox regression and Kaplan–Meier curves. The threshold of two-sided significance was fixed at 5%. In multivariable analysis, we estimated the CP Cox model containing age at randomization (years), tumor size (mm), tumor grade (SBR), lymph node status, estrogen receptor (ER, %), and progesterone receptor (PR, %). The CP model was compared with the model containing CP and relevant genomic (RG) characteristics using the likelihood ratio test to make sure that RG added complementary information to the traditional CP model. We also calculated the relative contribution of each variable in the Cox regression model. For internal validation, the area under the receiver operating characteristic (ROC) curve (AUC) [24] value at 5 and 10 years was used for evaluating the discrimination of prognostic CP and CP + RG models.

The genomic score (GS) was defined as the linear combination of the copy number and mutation genes in the CP + RG model. We also used the likelihood ratio test to evaluate the ability of the GS to predict OS by comparing CP model with CP + GS model.

In the validation data (METABRIC study), the DDFS endpoint was not available. Only the OS endpoint—defined as the time from diagnosis to death from any cause—and breast cancer-specific survival (BCSS) endpoint—defined as the time from diagnosis to death from BC—were available. Patients who did not experience the event were censored at the date of last follow-up. First, we built the CP + RG model found in the training data (PACS04) to ensure that the results were consistent in both data sets. Second, the GS built in the training data (PACS04) was validated in the validation data (METABRIC) to evaluate its ability to predict BCSS and OS. The CP model was compared to the CP + GS model, using the likelihood ratio test and AUC values at 5 and 10 years. We used R software version 3.5.1 for the statistical analysis and threshold of two-sided significance was fixed at 5%.

Data availability

The data sets generated during and/or analyzed during the current study are not publicly available in respect with the patient consent forms. Data access can be requested from the corresponding author.

Results

Description of PACS04 study population (training population)

Three hundred and twenty-seven patients (27%) with node-positive, HR+/HER2− BC treated with ED were selected. The other 867 patients (73%) did not have CGH data (66%) or TS data (6%) available or were not treated with ED (~ 1%). No significant differences were noted in the patient characteristics between selected ($n = 327$) and unselected ($n = 867$) patients (Table 1). The proliferation marker Ki-67 was excluded from the rest of the analyses because of too many missing values.

Table 1 Study population (embedded in the PACS04 trial)

Characteristics	Patients with node-positive HR+/HER2− BC, with clinical, CGH, and targeted sequencing data, treated with ED		Overall (n = 1194)	p-value
	Yes (n = 327)	No (n = 867)		
Demographic and tumor information				
Age at randomization (years)				0.4$^{\varphi}$
Mean (SD)	51.2 (7.9)	50.9 (8)	51 (8)	
Median (Q1–Q3)	52 (46–57)	51 (45–57)	52 (45–57)	
Min–max	27–65	27–66	27–66	
N (missing)	327 (0)	867 (0)	1194 (0)	
Tumor size (mm)				0.3$^{\varphi}$
Mean (SD)	25.8 (16.9)	25.1 (16.9)	25.3 (16.9)	
Median (Q1–Q3)	20 (15–30)	20 (15–30)	20 (15–30)	
Min–max	7–150	4–180	4–180	
N (missing)	325 (2)	858 (9)	1183 (11)	
Tumor SBR grade				0.7$^{\chi}$
Well differentiated (I)	59 (18.4%)	142 (16.9%)	201 (17.3%)	
Moderately differentiated (II)	183 (57%)	474 (56.4%)	657 (56.5%)	
Poorly differentiated (III)	79 (24.6%)	225 (26.8%)	304 (26.2%)	
N (Missing)	321(6)	841 (26)	1162 (32)	
Lymph node status				0.3$^{\chi}$
1–3	224 (68.5%)	627 (72.3%)	851 (71.3%)	
4–9	78 (23.9%)	188 (21.7%)	266 (22.3%)	
≥ 10	25 (7.6%)	52 (6%)	77 (6.4%)	
N (missing)	327 (0)	867 (0)	1194 (0)	
Biomarkers				
ER: estrogen receptor (%)				0.1$^{\varphi}$
Mean (SD)	87.4 (21.2)	87.5 (24.3)	87.5 (23.5)	
Median (Q1–Q3)	100 (80–100)	100 (87.8–100)	100 (87–100)	
Min–Max	0–100	0–100	0–100	
N (missing)	325 (2)	860 (7)	1185 (9)	
PR: Progesterone Receptor (%)				0.3$^{\varphi}$
Mean (SD)	56.3 (40.7)	58.5 (41)	57.9 (40.9)	
Median (Q1–Q3)	70 (10–97)	73 (10–100)	70 (10–100)	
Min–max	0–100	0–100	0–100	
N (missing)	316(11)	841 (26)	1157 (37)	
Ki-67: proliferation marker (%)				0.1$^{\varphi}$
Mean (SD)	15.9 (16.2)	14.2 (15.4)	14.7 (15.6)	
Median (Q1–Q3)	10 (5–20)	10 (2–20)	10 (2–20)	
Min–max	0–70	0–100	0–100	
N (missing)	169 (158)	410 (457)	579 (615)	

Data are mean (SD), median (Q1–Q3), min–max, *N* (missing), or *N* (%), *SD* standard deviation, *Q1* 25th percentile, *Q3* 75th percentile, *Min* minimum, *Max* maximum, *N* number, *Missing* number of missing values, % percentage, *HR+* hormone receptor positive, *HER2−* human epidermal growth factor receptor negative, *BC* breast cancer, *CGH* comparative genomic hybridization, *SBR* Scarff Bloom Richardson grading, *ED* Epirubicin + Docetaxel

$^{\varphi, \chi}$, correspond, respectively, to *p*-value associated with Kruskal–Wallis test, Fisher's exact test

The copy number status of some well-known genes (*CCND1*, *FGFR1*, *ZNF217*, and *ERBB2*) was obtained from CGH data. 21.4%, 15%, 12.5%, and 5.2% of patients, respectively, had amplifications for *CCND1*, *FGFR1*, *ZNF217*, and *ERBB2* (Fig. 2). In the TS data based on a panel of 36 sequenced genes (*PIK3CA*, *TP53*, *MAP3K1*,
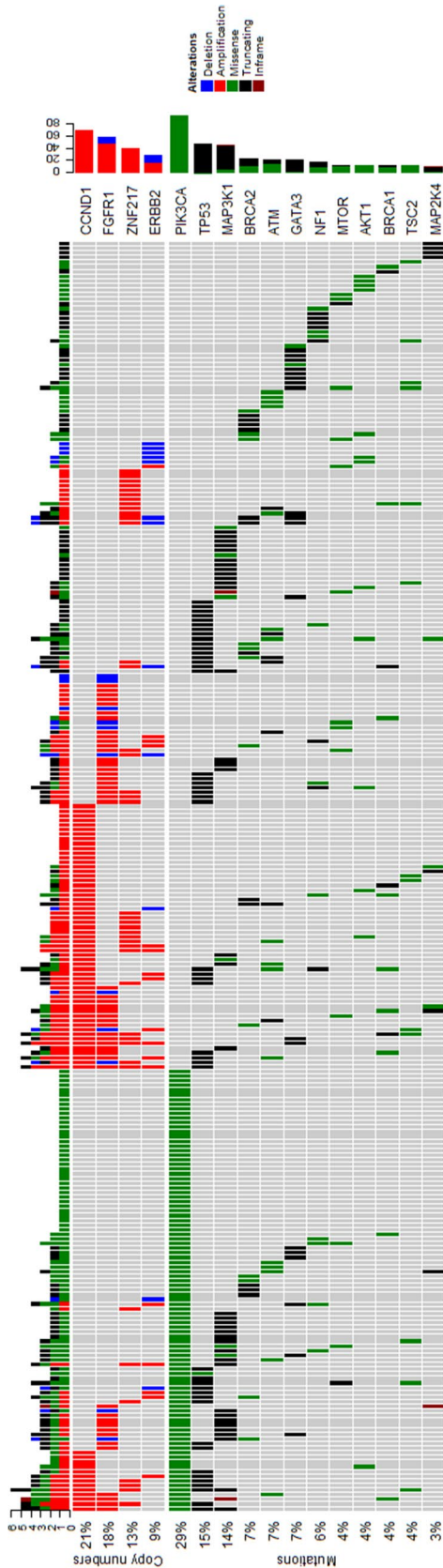
Fig. 2 Landscape of somatic driver alterations in node-positive HR+/HER2− breast cancer (PACS04 trial). *CCND1* Cyclin D1, *FGFR1* fibroblast growth factor receptor 1, *ZNF217* zinc finger protein 217, *ERBB2* receptor tyrosine kinase 2. Normal, copy number equal to 2. Amplification, copy number greater than 4. Deletion, copy number less than 1. Only mutations with a prevalence of 3% or greater are displayed. *PIK3CA*, Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic. *TP53*, Tumor Protein p53. *MAP3K1*, Mitogen-Activated Protein Kinase Kinase Kinase 1. *BRCA2*, Breast Cancer 2. *ATM*, Ataxia Telangiectasia Mutated. *GATA3*, GATA binding protein 3. *NF1*, Neurofibromin 1. *MTOR*, Mechanistic Target of Rapamycin Kinase. *AKT1*, Serine-threonine protein Kinase 1. *BRCA1*, Breast cancer 1. *TSC2*, Tuberous Sclerosis 2 protein. *MAP2K4*, Mitogen-Activated Protein Kinase Kinase 4

BRCA2, *ATM*, *GATA3*, *NF1*, *MTOR*, *AKT1*, *BRCA1*, *TSC2*, *MAP2K4*, *PTEN*, *TSC1*, *ESR1*, *PALB2*, *EGFR*, *ERBB2*, *FBXW7*, *KEAP1*, *RAD51B*, *RAD51C*, *STK11*, *FGFR4*, *PIK3R1*, *SF3B1*, *ALK*, *CTNNB1*, *DDR2*, *ERBB3*, *ERBB4*, *FGFR2*, *KDR*, *MET*, *RET*, and *RHOA*), 12 genes (33%) were selected and 24 genes (67%) were excluded because the prevalence of mutation was less than 3% (Supplementary Table S1). *PIK3CA*, *TP53*, and *MAP3K1* were the most mutated genes and the prevalence was, respectively, 29.1%, 14.7%, and 14% (Fig. 2). All genomic characteristics are described in supplementary data (Supplementary Tables S1 and S2).

Correlation between CP and genomic features in the training data (PACS04 trial)

The correlation between CP and genomic features is given in Table 2.

Copy numbers and mutations associated with DDFS in the training data (PACS04 trial)

After a median follow-up of 9.6 years, we observed 65 events for 327 patients in the analysis population. In the univariable analysis, copy number alterations of *FGFR1* ($HR_{Amplification} = 2.18$, 95% CI [1.21; 3.91], $p < 0.01$ and $HR_{Deletion} = 3.22$, 95% CI [1.28; 8.12], $p = 0.01$) and *ZNF217* ($HR_{Amplification} = 1.99$, 95% CI [1.08; 3.65], $p = 0.03$) were associated with an increased risk of distant disease, whereas *MAP3K1* mutations ($HR_{Mutation} = 0.09$, 95% CI [0.01; 0.63], $p = 0.02$) were associated with a decreased risk of distant disease (Table 3). The Kaplan–Meier analysis showed the similar results for FGFR1 ($p_{log-rank} = 0.0024$), ZNF217 ($p_{log-rank} = 0.024$) and MAP3K1 ($p_{log-rank} = 0.0023$) (Figs. 3 and 4). The verification of the log-linearity hypothesis is provided in Supplementary Figs. S1, S2, and S3.

In multivariable analysis, the model containing CP and relevant genomic characteristics selected during the univariable analysis step (CP + *FGFR1* + *ZNF217* + *MAP3K1*) showed that *ZNF217* copy numbers ($HR_{Amplification} = 1.67$, 95% CI [0.79; 3.53], $p = 0.2$) were not associated with DDFS (Table 4) and did not add further prognostic information ($\chi^2 = 1.62$, $p = 0.2$) compared to the CP + *FGFR1* + *MAP3K1* model (Table 5). The final Cox model selected was CP + *FGFR1* + *MAP3K1* and showed that *FGFR1* amplifications ($HR_{Amplification} = 2.44$, 95% CI [1.25; 4.76], $p = 0.009$) were associated with a poor prognosis and *MAP3K1* mutations ($HR_{Mutation} = 0.10$ [0.01; 0.78], $p = 0.03$) were associated with a good prognosis (Table 4). *MAP3K1* mutations alone added information to the CP model ($\chi^2 = 7.49$, $p = 0.006$) whereas *FGFR1* copy numbers alone did not ($\chi^2 = 4.96$, $p = 0.081$). It was the combination of *FGFR1* copy numbers and *MAP3K1* mutations that added relevant

Table 2 Spearman Copy numbers, mutations and CP characteristics correlation coefficients and 95% CI (PACS04 trial)

	Age at randomization (years)	Tumor size (mm)	Tumor grade (SBR)	Lymph node status	ER (%)	PR(%)
Copy numbers						
CCND1	− 0.002 [− 0.111; 0.114]	0.118 [0.01; 0.225]	0.216 [0.114; 0.318]	− 0.11 [− 0.208; − 0.009]	0.017 [− 0.095; 0.122]	− 0.006 [− 0.112; 0.102]
FGFR1	− 0.018 [− 0.132; 0.097]	0.059 [− 0.049; 0.162]	− 0.01 [− 0.125; 0.108]	− 0.011 [− 0.117; 0.094]	− 0.017 [− 0.124; 0.092]	− 0.1 [− 0.206; 0.003]
ZNF217	− 0.035 [− 0.146; 0.078]	0.047 [− 0.06; 0.153]	0.212 [0.104; 0.316]	0.031 [− 0.076; 0.142]	− 0.047 [− 0.16; 0.064]	− 0.051 [− 0.154; 0.054]
ERBB2	0.035 [− 0.09; 0.168]	0.058 [− 0.046; 0.157]	0.092 [− 0.019; 0.202]	0.033 [− 0.087; 0.147]	− 0.005 [− 0.114; 0.1]	0.003 [− 0.105; 0.105]
Somatic mutations						
PIK3CA	0.037 [− 0.069; 0.142]	− 0.021 [− 0.127; 0.088]	− 0.104 [− 0.206; − 0.002]	− 0.102 [− 0.202; 0]	− 0.068 [− 0.178; 0.042]	0.089 [− 0.021; 0.2]
TP53	0.005 [− 0.107; 0.116]	0.089 [− 0.025; 0.194]	0.137 [0.027; 0.244]	− 0.096 [− 0.188; 0.001]	− 0.11 [− 0.22; 0]	− 0.067 [− 0.178; 0.044]
MAP3K1	0.055 [− 0.05; 0.159]	− 0.03 [− 0.134; 0.078]	− 0.121 [− 0.224; − 0.014]	− 0.063 [− 0.159; 0.046]	0.009 [− 0.104; 0.116]	0.131 [0.021; 0.24]
BRCA2	− 0.018 [− 0.118; 0.086]	0.034 [− 0.067; 0.131]	0.01 [− 0.108; 0.126]	− 0.004 [− 0.108; 0.112]	0.071 [− 0.031; 0.164]	− 0.035 [− 0.153; 0.082]
ATM	0.081 [− 0.02; 0.174]	0.089 [− 0.002; 0.18]	0.047 [− 0.063; 0.154]	− 0.059 [− 0.147; 0.043]	− 0.045 [− 0.156; 0.069]	− 0.104 [− 0.206; 0.003]
GATA3	0.002 [− 0.103; 0.106]	0.052 [− 0.057; 0.16]	− 0.08 [− 0.207; 0.054]	− 0.02 [− 0.119; 0.092]	0.034 [− 0.078; 0.134]	− 0.071 [− 0.19; 0.048]
NF1	− 0.019 [− 0.119; 0.09]	0.026 [− 0.099; 0.149]	0.043 [− 0.079; 0.162]	0.041 [− 0.075; 0.152]	− 0.012 [− 0.126; 0.098]	0.016 [− 0.099; 0.126]
MTOR	− 0.084 [− 0.174; 0.018]	− 0.082 [− 0.185; 0.027]	0.004 [− 0.099; 0.111]	− 0.053 [− 0.132; 0.045]	− 0.028 [− 0.139; 0.079]	0.112 [0.001; 0.214]
AKT1	− 0.115 [− 0.205; − 0.016]	0.008 [− 0.085; 0.107]	0.027 [− 0.064; 0.12]	0.028 [− 0.081; 0.144]	− 0.038 [− 0.152; 0.071]	0.106 [0.012; 0.189]
BRC A1	− 0.028 [− 0.146; 0.097]	0.002 [− 0.108; 0.111]	0.126 [0.034; 0.218]	0.11 [− 0.018; 0.232]	− 0.016 [− 0.132; 0.094]	− 0.081 [− 0.155; − 0.006]
TSC2	− 0.067 [− 0.168; 0.04]	− 0.03 [− 0.15; 0.095]	0.031 [− 0.098; 0.158]	0.008 [− 0.102; 0.133]	0.035 [− 0.076; 0.131]	− 0.009 [− 0.123; 0.114]
MAP2K4	− 0.02 [− 0.128; 0.085]	− 0.005 [− 0.126; 0.118]	− 0.074 [− 0.151; − 0.011]	− 0.092 [− 0.147; − 0.017]	− 0.042 [− 0.169; 0.083]	− 0.07 [− 0.172; 0.042]

complementary information to CP model ($\chi^2 = 15.08$, $p < 0.01$) in determining DDFS (Table 5). These results were also confirmed by the AUC criteria. The AUC at 5 and 10 years were higher for CP + FGFR1 + MAP3K1 model in comparison with the CP model. The AUC values were, respectively, 81.39 and 67.47 at 5 and 10 years for CP model and were, respectively, 84.99 and 72.86 at 5 and 10 years for CP + FGFR1 + MAP3K1 model. The ROC curves are given in Supplementary Fig. S4 and Table S3. The relative contribution of FGFR1 (rank = 3, $\chi^2 = 8.50$, $p = 0.014$) and MAP3K1 (rank = 4, $\chi^2 = 4.71$, $p = 0.03$) was significant in determining DDFS, after the usual important prognostic variables tumor size and lymph node status ranked 1st and 2nd (Table S4).

Development of a genomic score based on the training data (PACS04 trial)

From the CP + FGFR1 + MAP3K1 model, we established a Genomic Score (GS) using only the coefficients associated with FGFR1 copy numbers and MAP3K1 mutations. The GS was split into three groups for biological relevance. The GS distribution is given in Fig. 5. Low risk corresponds to MAP3K1 mutated (GS < 0); moderate risk corresponds to MAP3K1 not mutated and FGFR1 not amplified (GS = 0); high risk corresponds to MAP3K1 not mutated and FGFR1 amplified (GS > 0).

The Kaplan–Meier figure of DDFS according to this grouping shows a clear separation (Fig. 6). As per

Table 3 Univariable Cox models on DDFS (PACS04 trial)

Characteristics	No. patients (No. events)	HR (95% CI) p-value
Copy number		
CCND1 (ref: normal)	257 (47)	
Amplification	70 (18)	1.44 (0.84; 2.48) p = 0.2
FGFR1 (ref: normal)	267 (45)	
Amplification	49 (15)	2.18 (1.21; 3.91) p < 0.01
Deletion	11 (5)	3.22 (1.28; 8.12) p = 0.01
ZNF217 (ref: normal)	286 (52)	
Amplification	41 (13)	1.99 (1.08; 3.65) p = 0.03
ERBB2 (ref: normal)	298 (57)	
Amplification	17 (4)	1.24 (0.45; 3.44) p = 0.7
Deletion	12 (4)	1.85 (0.67; 5.09) p = 0.2
Mutation		
PIK3CA (ref: normal)	232 (43)	
Mutation	95 (22)	1.17 (0.7; 1.96) p = 0.5
TP53 (ref: normal)	279 (52)	
Mutation	48 (13)	1.55 (0.84; 2.86) p = 0.2
MAP3K1 (ref: normal)	281 (64)	
Mutation	46 (1)	0.09 (0.01; 0.63) p = 0.02
BRCA2 (ref: normal)	303 (60)	
Mutation	24 (5)	1.03 (0.41; 2.57) p = 0.9
ATM (ref: normal)	304 (59)	
Mutation	23 (6)	1.46 (0.63; 3.39) p = 0.4
GATA3 (ref: normal)	305 (60)	
Mutation	22 (5)	1.15 (0.46; 2.86) p = 0.8
NF1 (ref: normal)	309 (62)	
Mutation	18 (3)	0.9 (0.28; 2.85) p = 0.8
MTOR (ref: normal)	313 (62)	
Mutation	14 (3)	1 (0.31; 3.18) p = 1
AKT1 (ref: normal)	314 (61)	
Mutation	13 (4)	1.67 (0.60; 4.59) p = 0.3
BRCA1 (ref: normal)	314 (62)	
Mutation	13 (3)	1.33 (0.42; 4.24) p = 0.6
TSC2 (ref: normal)	314 (62)	
Mutation	13 (3)	1.21 (0.38; 3.86) p = 0.7
MAP2K4 (ref: normal)	316 (64)	

Table 3 (continued)

Characteristics	No. patients (No. events)	HR (95% CI) p-value
Mutation	11 (1)	0.49 (0.07; 3.52) p = 0.5

the definition of the score, it added additional information ($n_{DDFS} = 307$, n - $event_{DDFS} = 58$, $\chi^2_{DDFS} = 12.97$, $p_{DDFS} < 0.001$ and $n_{OS} = 307$, n - $event_{OS} = 45$, $\chi^2_{OS} = 11.52$, $p_{OS} < 0.001$) to the CP model for DDFS and OS. The model description is given in Supplementary Table S5. The comparison between a CP + KI67 + GS model versus a CP + KI67 model on the small subset of 158 patients with KI67 data showed that the GS added additional information to this model especially for OS ($n_{DDFS} = 158$, n - $event_{DDFS} = 29$, $\chi^2_{DDFS} = 2.92$, $p_{DDFS} = 0.083$ and $n_{OS} = 158$, n - $event_{OS} = 24$, $\chi^2_{OS} = 4.06$, $p_{OS} = 0.041$).

Validation of the genomic score in the external data (METABRIC trial)

In the external data set (METABRIC), we first checked the log-linearity hypothesis of continuous covariates (Supplementary Figure S5) and evaluated the prognostic value of *FGFR1* copy numbers and *MAP3K1* mutations on BCSS and OS. In univariable analysis (Fig. 7), *FGFR1* copy numbers were significantly associated with the BCSS ($p_{BCSS} < 0.001$) and OS ($p_{OS} < 0.001$) whereas *MAP3K1* mutations were only significantly associated with BCSS ($p_{BCSS} < 0.01$) but not with OS ($p_{OS} = 0.28$). In addition to CP parameters, *FGFR1* copy numbers were significantly associated with BCSS ($HR_{Amplification} = 2.00$, 95% CI [1.40; 2.87], $p < 0.001$) and OS ($HR_{Amplification} = 1.54$, 95% CI [1.15; 2.07], $p = 0.004$); *MAP3K1* mutations were strongly associated with BCSS ($HR_{Mutation} = 0.58$, 95% CI [0.41; 0.83], $p = 0.003$) but were not significantly associated with OS ($HR_{Mutation} = 0.82$, 95% CI [0.65; 1.02], $p = 0.08$) (Supplementary Table S7). *FGFR1* and *MAP3K1* added information ($\chi^2_{BCSS} = 21.03$, $p_{BCSS} < 0.001$ and $\chi^2_{OS} = 10.36$, $p_{OS} = 0.005$) to the CP model for BCSS and OS *(Supplementary Table S8)*. The detail of the relative contribution of *FGFR1* and *MAP3K1* is given in *Supplementary* Table S9. The AUC values at 5 and 10 years for BCSS and OS were higher for CP + *FGFR1* + *MAP3K1* model ($AUC(5 years)_{BCSS} = 74.47$, $AUC(10 years)_{BCSS} = 73.35$, $AUC(5 years)_{OS} = 71.77$ and $AUC(10 years)_{OS} = 73.34$) than CP model ($AUC(5 years)_{BCSS} = 74.12$, $AUC(10 years)_{BCSS} = 72.24$, $AUC(5 years)_{OS} = 71.29$ and $AUC(10 years)_{OS} = 72.66$). The ROC curves are given in Supplementary Figures S6 and S7, and Table S10. Copy number alterations of *FGFR1* gene increased the risk of death from BC whereas mutations of *MAP3K1* decreased
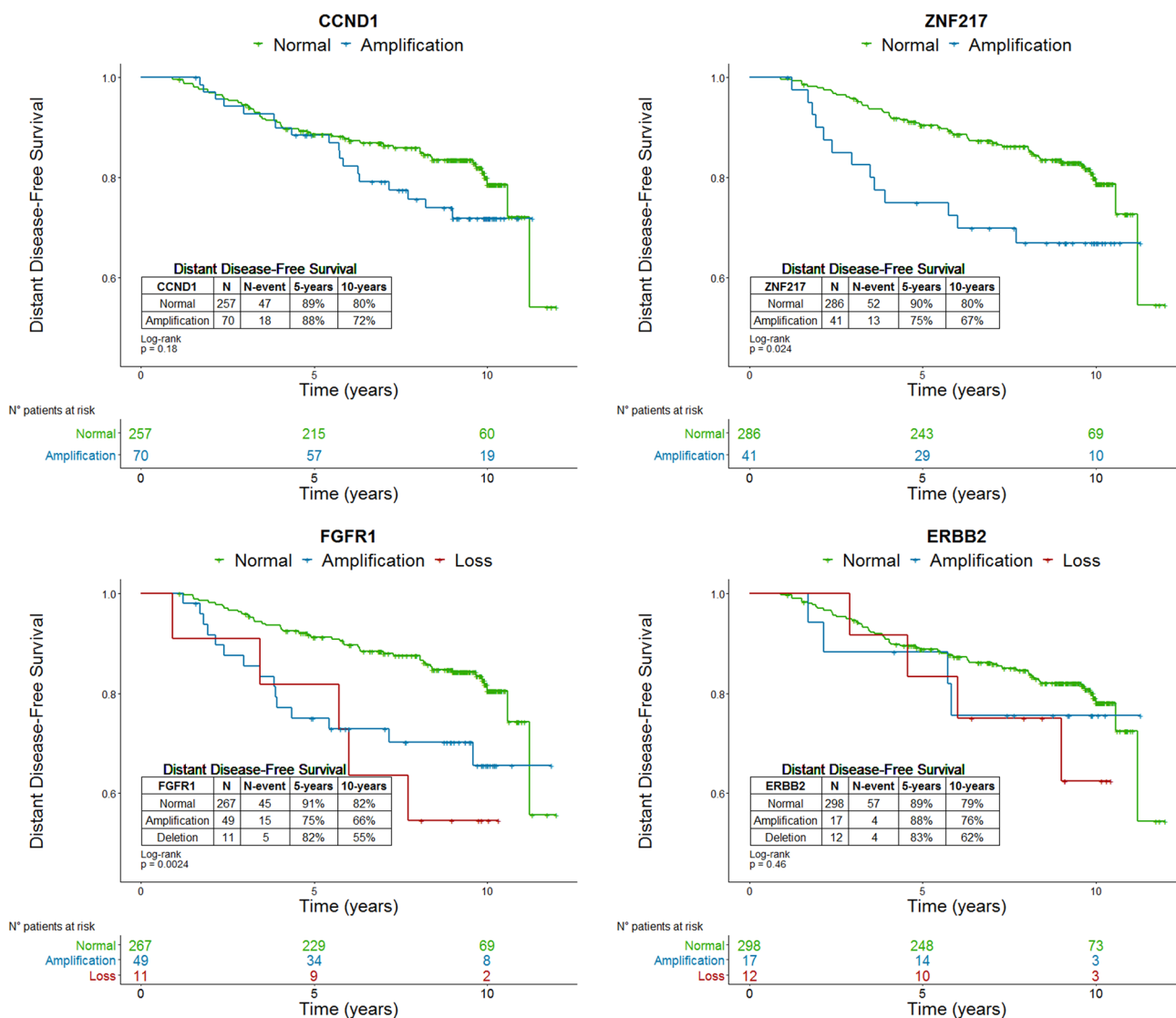
Fig. 3 Survival curves according to copy number status of well-known genes in the training data (PACS04 trial). *N* number of patients, *N-event* number of events. Normal: copy number equal to 2;

Amplification: copy number greater than 4; Deletion: copy number less than 1; p: *p* value associated with the log-rank test

this risk. Applying our genomic score in the external data set, the Kaplan–Meier analysis indicated that it was significantly associated ($p_{BCSS} < 0.001$ and $p_{OS} = 0.0015$) with survival (Fig. 8). It provides additional information ($n_{BCSS} = 1333$, n - event$_{BCSS} = 392$, $\chi^2_{BCSS} = 15.39$, $p_{BCSS} < 0.001$ and $n_{OS} = 1333$, n - event$_{OS} = 763$, $\chi^2_{OS} = 5.62$, $p_{OS} = 0.02$) to CP model. The detail of CP + GS model is given in *Supplementary* Table S11. The results were consistent with the results found in the training data. The patients' characteristics from the METABRIC study are

described in *supplementary* Table S6. In the validation data set, the genomic score was validated on all patients with HR+ and HER2- BC without distinguishing patients with lymph node invasion to maximize statistical power. A sensitivity analysis was conducted looking only at the population of patients with lymph node-positive disease. In univariable analysis, the genomic score was significantly associated ($p_{BCSS} < 0.01$) with BCSS but less strongly with OS ($p_{OS} = 0.07$) (Supplementary Figure S8). In the multivariable analysis, in addition to CP parameters, it
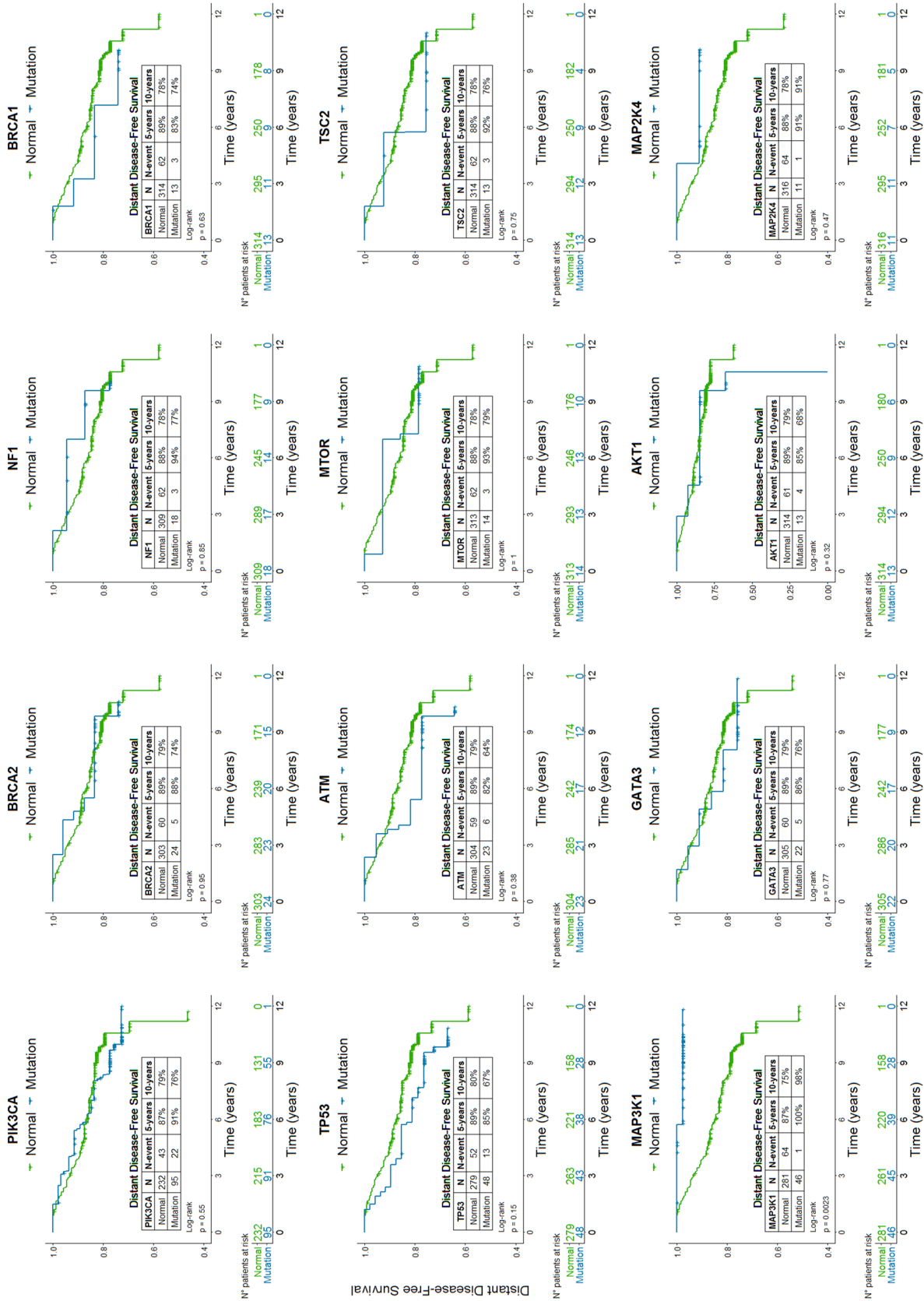
Fig. 4 Survival curves according to driver genes' mutation status in the training data (PACS04 trial). *Mutation* sum of all types of mutations, *N* number of patients, *N-event* number of events, *p* log-rank *p*-value

Table 4 Multivariable Cox models on DDFS (PACS04 trial)

Characteristics	No. patients (No. events) or Median [Q1-Q3]	CP HR (95% CI) p-value	CP+MAP3K1+FGFR1+ZNF217 HR (95% CI) p-value	CP+MAP3K1+FGFR1 HR (95% CI) p-value
Age at randomization (years), linear	52 [45–57]	0.99 (0.96; 1.02) p=0.5	1.003 (0.97; 1.03) p=0.9	1.00 (0.97; 1.03) p=1
Age at randomization (years), nonlinear		– p=0.3	– p=0.3	– p=0.2
Tumor size (mm)	20 [15–30]	1.02 (1.01; 1.03) $p<0.001$	1.02 (1.01; 1.03) $p<0.001$	1.02 (1.01; 1.03) $p<0.001$
Tumor grade (ref: Well differentiated)	58 (5)			
Moderately differentiated	175 (32)	1.69 (0.64; 4.45) p=0.3	1.57 (0.59; 4.18) p=0.4	1.58 (0.59; 4.20) p=0.4
Poorly differentiated	74 (21)	2.5 (0.9; 6.97) p=0.08	2.08 (0.72; 5.94) p=0.2	2.26 (0.80; 6.40) p=0.1
Lymph node status (ref: 1–3)	211 (28)			
4–9	72 (21)	2.19 (1.23; 3.9) p=0.007	2.59 (1.42; 4.73) p=0.002	2.56 (1.40; 4.68) p=0.002
≥	24 (9)	2.73 (1.24; 5.98) p=0.01	3.47 (1.55; 7.79) p=0.003	3.24 (1.46; 7.21) p=0.004
log-transformed ER	4.6 [4.4–4.6]	1.15 (0.64; 2.08) p=0.6	0.97 (0.49; 1.90) p=0.9	0.95 (0.48; 1.88) p=0.9
log-transformed PR, linear	4.26 [2.4–4.6]	0.98 (0.82; 1.17) p=0.8	1.04 (0.86; 1.24) p=0.7	1.04 (0.87; 1.25) p=0.6
log-transformed PR, nonlinear		– p=0.2	– p=0.5	– p=0.5
MAP3K1 (ref: normal)	265 (57)			
Mutation	42 (1)		0.10 (0.01; 0.82) p=0.03	0.10 (0.01; 0.78) p=0.03
FGFR1 (ref: normal)	251 (41)			
Amplification	47 (14)		2.37 (1.21; 4.64) p=0.01	2.44 (1.25; 4.76) p=0.009
Deletion	9 (3)		3.08 (0.81; 11.80) p=0.1	3.23 (0.88; 11.88) p=0.08
ZNF217 (ref: normal)	271 (48)			
Amplification	36 (10)		1.67 (0.79; 3.53) p=0.2	
N-patients		307	307	307
N-events		58	58	58
Concordance		0.75	0.77	0.78
Likelihood ratio test		$p<0.001$	$p<0.001$	$p<0.001$

No. patients number of patients, *No. event* number of events, *Q1* first quartile, *Q3* third quartile, *CP* clinicopathological Cox model including age, tumor size, tumor grade, lymph node status, ER, and PR. HR (95% CI), hazard ratio (95% confidence interval)

Table 5 Likelihood ratio test for multivariable Cox models on DDFS (PACS04 trial)

Model	No. patients	No. events	Likelihood Ratio (χ^2)	p-value
CP *VS* CP+FGFR1+ZNF217+MAP3K1	307	58	16.69	0.002
CP *VS* CP+FGFR1+MAP3K1	307	58	15.08	0.001
CP+FGFR1+MAP3K1 *VS* CP+FGFR1+MAP3K1+ZNF217	307	58	1.62	0.2
CP *VS* CP+MAP3K1	307	58	7.49	0.006
CP *VS* CP+FGFR1	307	58	4.96	0.081
CP+MAP3K1 *VS* CP+MAP3K1+FGFR1	307	58	7.58	0.02

No. patients number of patients, *No. event* number of events, *CP* clinicopathological Cox model, including age, tumor size, tumor grade, lymph node status, ER, and PR. *NULL* null model, *VS* versus

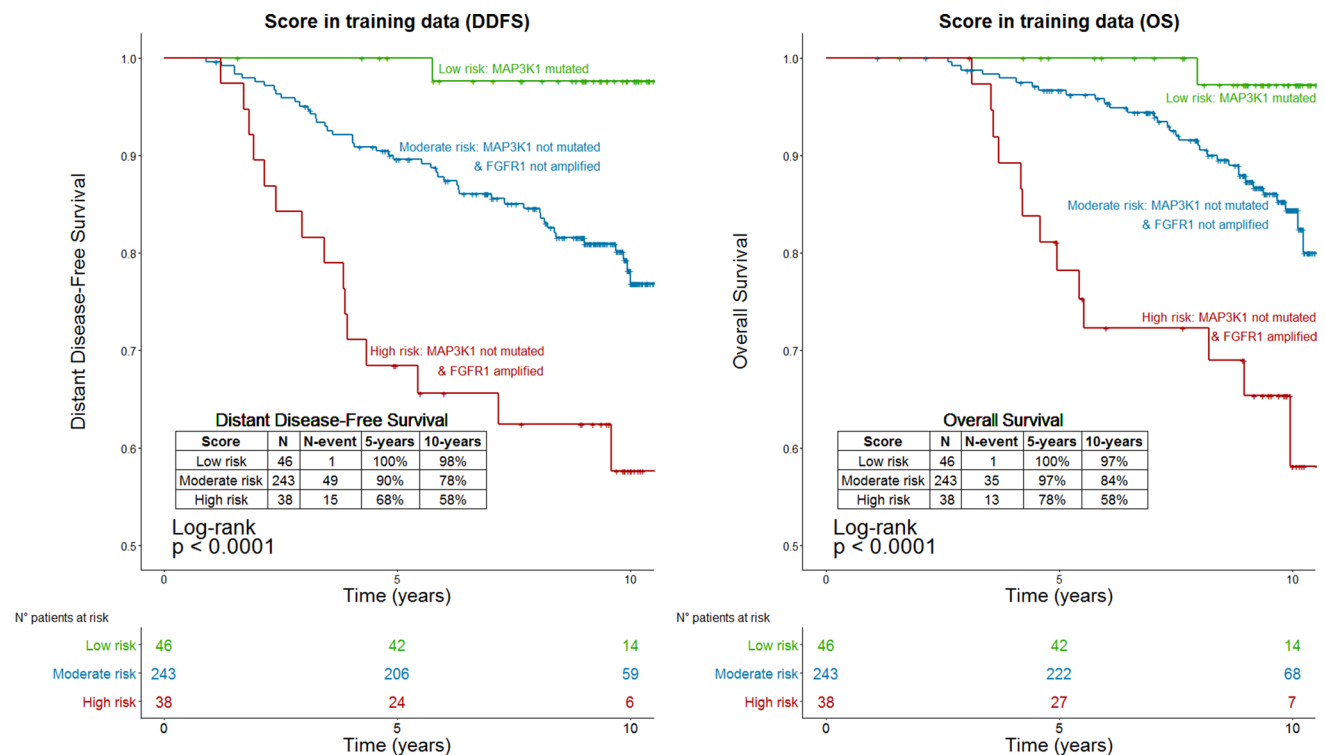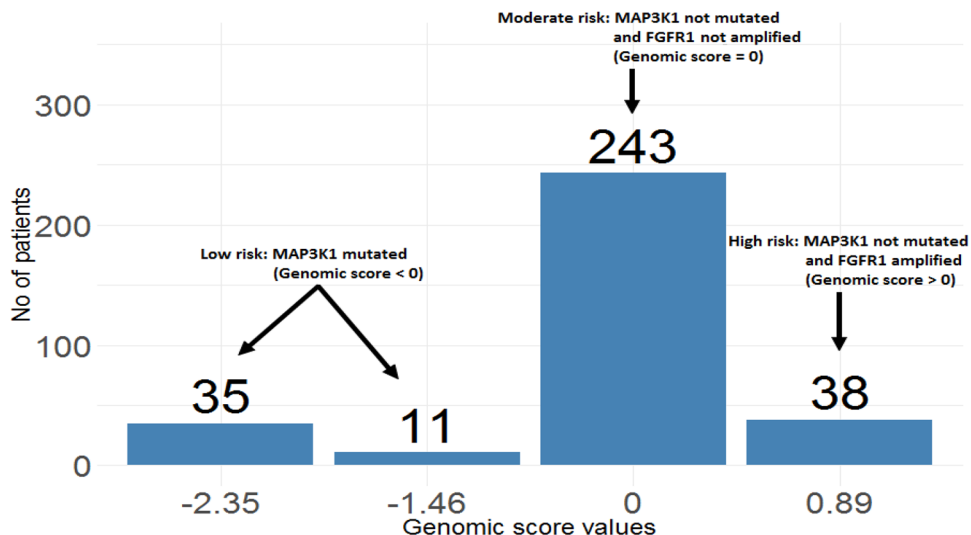**Fig. 5** Genomic score distribution (PACS04 trial)

Fig. 6 Survival curves according to the Genomic Score in the training data (PACS04 trial). *DDFS* distant disease-free survival, *OS* overall survival *N* number of patients, *N-event* number of events, *p* log-rank *p* value. Low risk, *MAP3K1* mutated. Moderate, *MAP3K1* not mutated and *FGFR1* not amplified. High risk, *MAP3K1* not mutated and *FGFR1* amplified

was strongly associated (HR = 1.25, 95% CI [1.04; 1.49], $\chi^2_{BCSS} = 5.88$, $p_{BCSS} = 0.01$) with the BCSS but to a lesser extent with OS (HR = 1.09, 95% CI [0.97; 1.2], $\chi^2_{OS} = 1.91$, $p_{OS} = 0.2$) (Supplementary Table S14 and Table S15). The details of *CP + FGFR1 + MAP3K1* model are given in supplementary Table S12 and Table S13.

Discussion

In HR+/HER2− BC, a key aim of precision medicine is to tailor adjuvant clinical management based on individual risk of relapse. This BC subtype is a heterogeneous disease based on genomic characteristics. We therefore performed
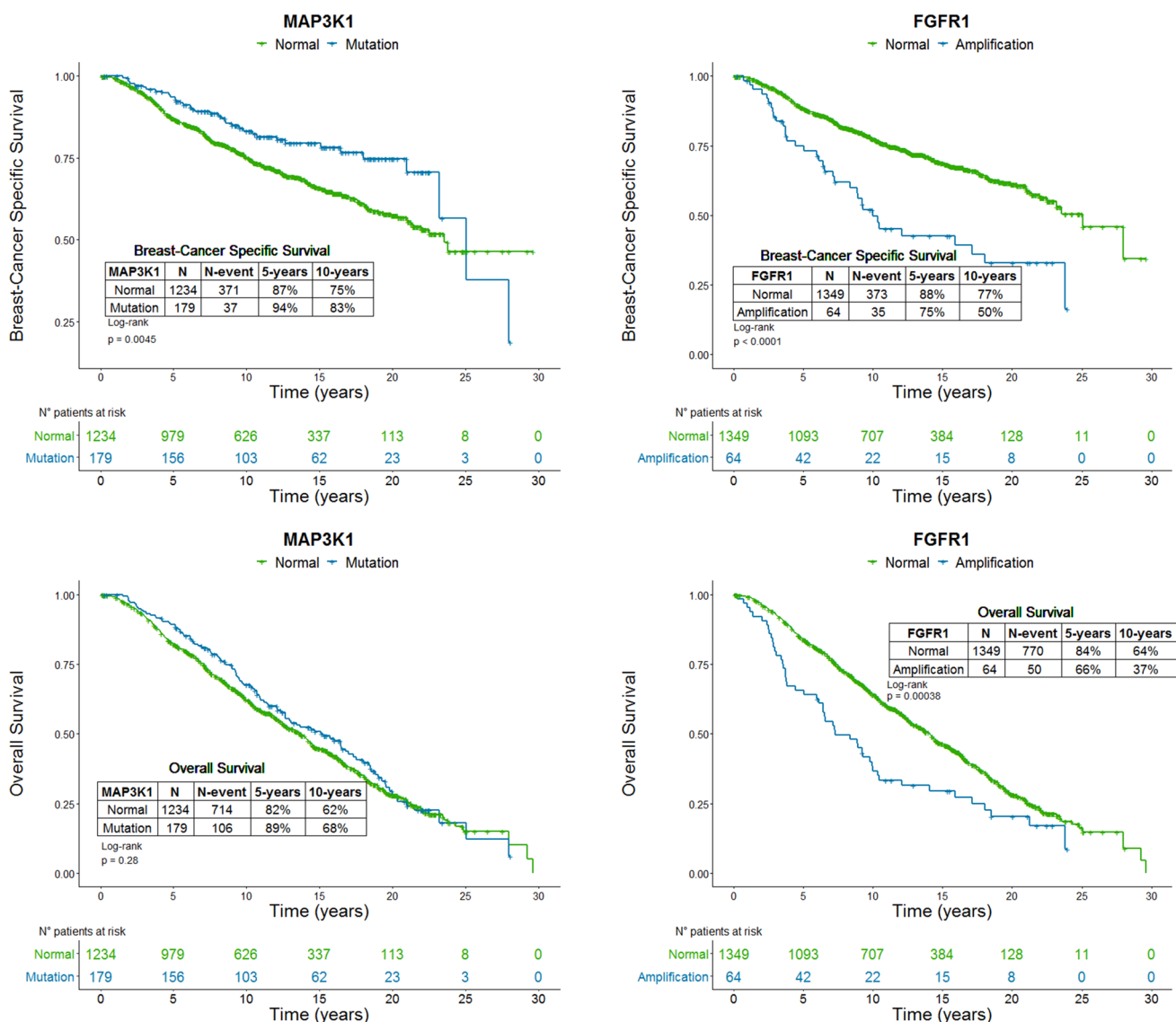
Fig. 7 Survival curves according to genomic characteristics (METABRIC trial). *N* number of patients, *N-event* number of events, *p* log-rank *p*-value

a comprehensive analysis of a large population of patients with HR+/HER2− early BC, describing relevant genomic characteristics of these tumors and their associations with usual CP characteristics as well as their prognosis. The analysis indicated that *FGFR1* copy number and *MAP3K1* mutation added strong complementary information to CP features in determining the risk of distant relapse. FGFR1 amplification was identified as an independent risk factor while *MAP3K1* mutation was associated with a protective effect. These results are consistent with previous studies [25, 26]. *FGFR1* gene encodes a tyrosine kinase receptor that plays an important role in the development

of BC and amplification of *FGFR1* has oncogenic properties. Recently, it has been shown that activation of *FGFR1* drives the invasive behavior of BC cells [25]. *MAP3K1*, a serine-threonine kinase, mediates apoptosis through activation of JNK pro-apoptosis protein; however, *MAP3K1* mutations are associated with a low proliferation index and a low tumor grade [18].

In our analysis, *ERBB2* amplification was not associated with DDFS in univariable analysis. This result was expected because our BC subgroup is HR-positive and HER2-negative. Although *ZNF217* amplification was strongly associated with an increased risk of distant disease in the
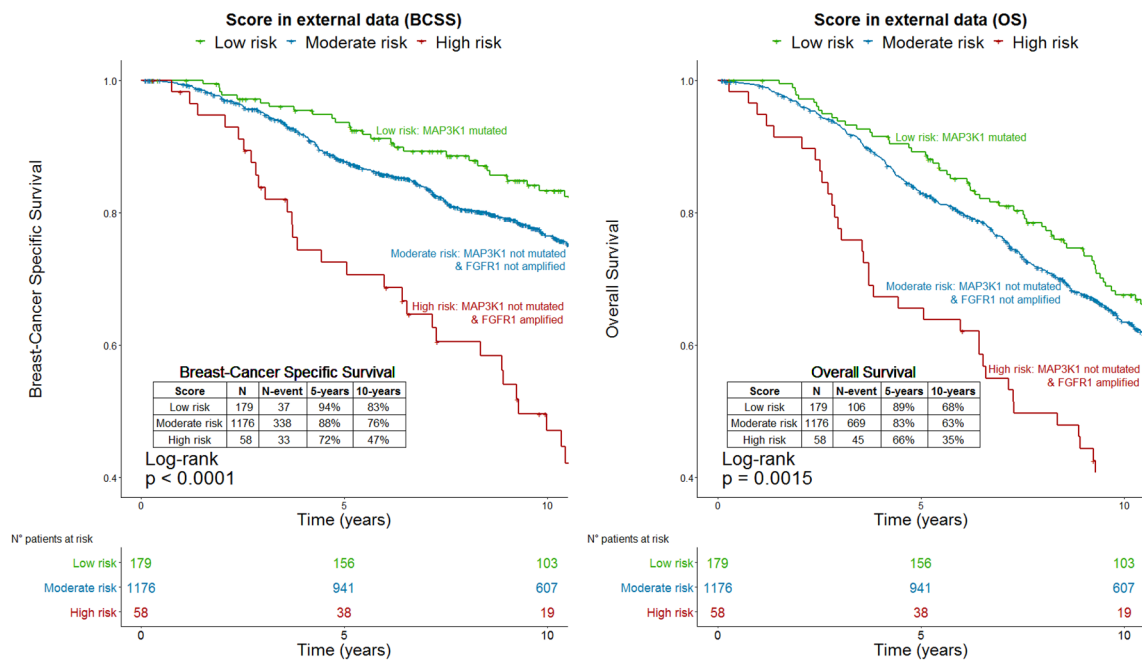
Fig. 8 Survival curves according to the Genomic Score in the validation data (METABRIC trial). *BCSS* breast cancer-specific survival, *OS* overall survival, *N* number of patients. *N-event* number of events. *p* p-value associated with the log-rank test. Low risk, *MAP3K1* mutated. Moderate, *MAP3K1* not mutated and *FGFR1* not amplified. High risk, *MAP3K1* not mutated and *FGFR1* amplified

univariable analysis, it did not remain significant in the multivariable analysis. *ZNF217* did not provide any information to CP + *FGFR1* + *MAP3K1* model, probably because it was slightly correlated with the tumor grade and mutational status of the *MAP3K1* gene.

We retrieved the genomic part of the CP + *FGFR1* + *MAP3K1* model to create a genomic score. This score adjusted with CP features provided additional information to the initial CP model in determining the risk of death. The division of the score into three relevant categories made it possible to distinguish different levels of risk. Patients with a mutation of the *MAP3K1* gene have a very good prognosis. Those with no alteration of *FGFR1* and *MAP3K1* genes have a moderate risk. And finally, those with an amplification of *FGFR1* gene and no mutation of *MAP3K1* gene have a higher risk. In the validation sample (METABRIC), the results were in line with the results found in the training sample (PACS04). *FGFR1* amplification and *MAP3K1* mutation were both associated with BCSS. Using the score that was established in the training data, the results showed that it added relevant information to the CP model in determining the risk of relapse (local and distant) and death. In the subset of patients with lymph node-positive disease in the METABRIC study, the genomic score continued to add additional prognostic information for BCSS.

However, our study has several limitations. In PACS04 trial, the sample size was relatively small with few events and the proliferation biomarker (Ki-67) was only available

on a small subset of patients. In the METABRIC study, the DDFS endpoint was not available and the type of chemotherapy was not specified. We can also note that in the three prognostic groups generated, frequencies are relatively high in the training set (PACS04) and lower in the external validation data (METABRIC).

From a clinical standpoint, our study proposes a new way of classifying patients with HR+ and HER2- lymph node-positive BC based on known genes. Taking into account the copy number of *FGFR1* and the mutational status of *MAP3K1* improves the understanding of the prognosis and can help decision-making for the adjuvant treatment.

Conclusion

Currently, only CP parameters are taken into account to predict distant disease and guide HR+ and HER2− lymph node-positive BC adjuvant treatment. Our study showed that in addition to CP parameters, copy number alterations of *FGFR1* gene and mutational status of *MAP3K1* gene could improve distant disease prediction. *FGFR1* amplification was associated with a high risk of distant disease whereas *MAP3K1* mutation was associated with a low risk of distant disease.

The genomic score combining both genomic features classified patients into three subgroups with different prognosis: low risk (*MAP3K1* mutated), moderate risk

(*MAP3K1* not mutated and *FGFR1* not amplified), and high risk (*MAP3K1* not mutated and *FGFR1* amplified). Consideration of *FGFR1* and *MAP3K1* could be a new way to refine the treatment decision-making.

Compliance with ethical standards

Conflict of interest The authors declare no relevant conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the PACS04 study.

References

1. Howlader N, Altekruse SF, Li CI, Chen VW, Clarke CA, Ries LAG et al (2014) US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status. J Natl Cancer Inst. https://doi.org/10.1093/jnci/dju055
2. Cianfrocca M, Goldstein LJ (2004) Prognostic and predictive factors in early-stage breast cancer. Oncologist 9:606–616
3. Rugo HS (2008) The importance of distant metastases in hormone-sensitive breast cancer. Breast 17(Suppl 1):S3–S8
4. Rosano GMC (2007) The importance of distant metastasis and its impact on survival rates on early-stage hormone receptor-positive breast cancer. US Oncol Dis 1:22–25
5. Mohammadizadeh F, Hani M, Ranaee M, Bagheri M (2013) Role of cyclin D1 in breast carcinoma. J Res Med Sci 18:1021–1025
6. Roy PG, Pratt N, Purdie CA, Baker L, Ashfield A, Quinlan P et al (2010) High CCND1 amplification identifies a group of poor prognosis women with estrogen receptor positive breast cancer. Int J Cancer 127:355–360
7. Turner N, Pearson A, Sharpe R, Lambros M, Geyer F, Lopez-Garcia MA et al (2010) FGFR1 amplification drives endocrine therapy resistance and is a therapeutic target in breast cancer. Cancer Res 70:2085–2094
8. Drago JZ, Niemierko A, Spring L, Moy B, Juric D, Isakoff SJ et al (2017) FGFR gene amplification and response to endocrine therapy in metastatic hormone receptor positive (HR+) breast cancer. JCO 35:1013
9. Vendrell JA, Thollet A, Nguyen NT, Ghayad SE, Vinot S, Bièche I et al (2012) ZNF217 is a marker of poor prognosis in breast cancer that drives epithelial-mesenchymal transition and invasion. Cancer Res 72:3593–3606
10. Littlepage LE, Adler AS, Kouros-Mehr H, Huang G, Chou J, Krig SR et al (2012) The transcription factor ZNF217 is a prognostic biomarker and therapeutic target during breast cancer progression. Cancer Discov 2:638–651
11. Krig SR, Miller JK, Frietze S, Beckett LA, Neve RM, Farnham PJ et al (2010) ZNF217, a candidate breast cancer oncogene amplified at 20q13, regulates expression of the ErbB3 receptor tyrosine kinase in breast cancer cells. Oncogene 29:5500–5510
12. Kim Y-S, Won YS, Park KS, Song BJ, Kim JS, Oh SJ et al (2008) Prognostic significance of HER2 gene amplification according to stage of breast cancer. J Korean Med Sci 23:414–420
13. Borg A, Baldetorp B, Fernö M, Killander D, Olsson H, Sigurdsson H (1991) ERBB2 amplification in breast cancer with a high rate of proliferation. Oncogene 6:137–143
14. Carr JA, Havstad S, Zarbo RJ, Divine G, Mackowiak P, Velanovich V (2000) The association of HER-2/neu amplification with breast cancer recurrence. Arch Surg 135:1469–1474
15. Letessier A, Sircoulomb F, Ginestier C, Cervera N, Monville F, Gelsi-Boyer V et al (2006) Frequency, prognostic impact, and subtype association of 8p12, 8q24, 11q13, 12p13, 17q12, and 20q13 amplifications in breast cancers. BMC Cancer 6:245
16. Cizkova M, Susini A, Vacher S, Cizeron-Clairac G, Andrieu C, Driouch K et al (2012) PIK3CA mutation impact on survival in breast cancer patients and in ERα, PR and ERBB2-based subgroups. Breast Cancer Res 14:R28
17. Fernández-Cuesta L, Oakman C, Falagan-Lotsch P, Smoth K-S, Quinaux E, Buyse M et al (2012) Prognostic and predictive value of TP53 mutations in node-positive breast cancer patients treated with anthracycline- or anthracycline/taxane-based adjuvant therapy: results from the BIG 02-98 phase III trial. Breast Cancer Res 14:R70
18. Pham TT, Angus SP, Johnson GL (2013) MAP3K1: genomic alterations in cancer and function in promoting cell survival or apoptosis. Genes Cancer 4:419–426
19. Network Cancer Genome Atlas (2012) Comprehensive molecular portraits of human breast tumours. Nature 490:61–70
20. Roché H, Allouache D, Romieu G, Bourgeois H, Canon J, Serin D et al (2009) Five-year analysis of the FNCLCC-PACS04 trial: FEC100 vs ED75 for the adjuvant treatment of node positive breast cancer. Cancer Res 69:602
21. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ et al (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486:346–352
22. Commo F, Guinney J, Ferté C, Bot B, Lefebvre C, Soria J-C et al (2016) rCGH: a comprehensive array-based genomic profile platform for precision medicine. Bioinformatics 32:1402–1404
23. Adélaïde J, Finetti P, Bekhouche I, Repellini L, Geneix J, Sircoulomb F et al (2007) Integrated profiling of basal and luminal breast cancers. Cancer Res 67:11565–11575
24. Blanche P. TimeROC: time-dependent ROC curve and AUC for censored survival data. https://cran.r-project.org/web/packages/timeROC/timeROC.pdf
25. Gelsi-Boyer V, Orsetti B, Cervera N, Finetti P, Sircoulomb F, Rougé C et al (2005) Comprehensive profiling of 8p11-12 amplification in breast cancer. Mol Cancer Res 3:655–667
26. Griffith OL, Spies NC, Anurag M, Griffith M, Luo J, Tu D et al (2018) The prognostic effects of somatic mutations in ER-positive breast cancer. Nat Commun 9:3476

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

Affiliations

Dimitri Carene[1,2] · Alicia Tran-Dien[2] · Jérôme Lemonnier[3] · Florence Dalenc[4] · Christelle Levy[5] · Jean-Yves Pierga[6] · William Jacot[7] · Jean-Luc Canon[8] · Catherine Richon[9] · Ludovic Lacroix[2,9] · Christophe Caux[10] · Fabrice André[1,2,11] · Stefan Michiels[12,13,14]

[1] Faculté de Médecine, Kremlin-Bicêtre, Université Paris Sud, Villejuif, France

[2] INSERM U981, Université Paris Sud, Villejuif, France

[3] R&D UNICANCER, Paris, France

[4] Institut Claudius Ragaud, Institut Universitaire du Cancer de Toulouse-Oncopole, Toulouse, France

[5] Centre François Baclesse, Caen, France

[6] Institut Curie, Paris, France

[7] Institut régional du Cancer de Montpellier, Montpellier, France

[8] Grand Hôpital de Charleroi, Charleroi, Belgium

[9] Department of Medical Biology and Pathology, Translational Research Laboratory and BioBank, Gustave Roussy, Villejuif, France

[10] Univ Lyon, Université Claude Bernard Lyon 1, INSERM 1052, CNRS 5286, Cancer Research Center of Lyon (CRCL), Centre Léon Bérard, 69008 Lyon, France

[11] Department of Medical Oncology, Gustave Roussy, Villejuif, France

[12] Service de Biostatistique et d'Epidémiologie, Gustave Roussy, Villejuif, France

[13] CESP, Faculté de médecine, Université Paris Sud, Faculté de médecine UVSQ, INSERM, Université Paris Saclay, Villejuif, France

[14] Service de Biostatistique et d'Epidémiologie, Gustave Roussy and INSERM U1018, Université Paris Sud, 114 Rue Edouard Vaillant, 94800 Villejuif, France