MOLECULAR ONCOLOGY 9 (2015) 115–127

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/molonc

A tumor DNA complex aberration index is an independent predictor of survival in breast and ovarian cancer

Hans Kristian Moen Vollan[a,c,d,e,1], *Oscar M. Rueda*[a,1], *Suet-Feung Chin*[a], *Christina Curtis*[f], *Gulisa Turashvili*[g,h], *Sohrab Shah*[g,h], *Ole Christian Lingjærde*[d,i,j], *Yinyin Yuan*[s], *Charlotte K. Ng*[t], *Mark J. Dunning*[a], *Ed Dicks*[k], *Elena Provenzano*[l,m], *Stephen Sammut*[m], *Steven McKinney*[h], *Ian O. Ellis*[n], *Sarah Pinder*[o,p], *Arnie Purushotham*[o,p], *Leigh C. Murphy*[q], *Vessela N. Kristensen*[c,d,r], *METABRIC Group*[3] *James D. Brenton*[a,b,l,m], *Paul D.P. Pharoah*[b,k,l,m], *Anne-Lise Børresen-Dale*[c,d], *Samuel Aparicio*[g,h,**,2], *Carlos Caldas*[a,b,l,m,*,2]

[a]*Cancer Research UK, Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK*
[b]*Department of Oncology, University of Cambridge, Hills Road, Cambridge CB2 2XZ, UK*
[c]*Department of Genetics, Institute for Cancer Research, Oslo University Hospital, The Norwegian Radium Hospital, Montebello, 0310 Oslo, Norway*
[d]*The K.G. Jebsen Center for Breast Cancer Research, Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, Norway*
[e]*Department of Oncology, Division for Surgery, Cancer and Transplantation, Oslo University Hospital, The Norwegian Radium Hospital, Montebello, 0310 Oslo, Norway*
[f]*Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA*
[g]*Department of Pathology and Laboratory Medicine, University of British Colombia, Vancouver, British Colombia V6T 2B5, Canada*
[h]*Molecular Oncology, British Colombia Cancer Research Center, Vancouver, British Columbia V5Z 1L3, Canada*
[i]*Biomedical Informatics Division, Department of Computer Science, University of Oslo, Oslo, Norway*
[j]*Center for Cancer Biomedicine, University of Oslo, Norway*
[k]*Strangeways Research Laboratories, University of Cambridge, Cambridge CB1 9RN, UK*
[l]*Cambridge Experimental Cancer Medicine Centre, Cambridge CB2 0RE, UK*
[m]*Cambridge Breast Unit, Addenbrooke's Hospital, Cambridge University Hospital NHS Foundation Trust and NIHR Cambridge Biomedical Research Centre, Cambridge CB2 2QQ, UK*
[n]*Department of Histopathology, School of Molecular Medical Sciences, University of Nottingham, Nottingham, NG5 1PB, UK*

Abbreviations: BCSS, Breast cancer specific survival; CAAI, Complex arm-wise aberration index; CNA, Copy number alterations; ER, Estrogen receptor; HR, Hazard ratio; HGSOC, High-grade serous ovarian cancer; MIP, Molecular inversion probe; OS, Overall survival; PFS, Progression free survival.

* *Corresponding author.* Cancer Research UK, Cambridge Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. Tel.: +44 1223 769 650.
** *Corresponding author.* Department of Pathology and Laboratory Medicine, University of British Colombia, Vancouver, British Colombia V6T 2B5, Canada. Tel.: +1 604 675 8201.
E-mail addresses: saparicio@bccrc.ca (S. Aparicio), carlos.caldas@cruk.cam.ac.uk (C. Caldas).
[1] These authors contributed equally to this work.
[2] Senior and co-corresponding authors.
[3] Full list of members at end of text.

http://dx.doi.org/10.1016/j.molonc.2014.07.019

[o]King's College London, Breakthrough Breast Cancer Research Unit, London WC2R 2LS, UK
[p]NIHR Comprehensive Biomedical Research Centre at Guy's and St. Thomas NHS Foundation Trust and King's College London, London WC2R 2LS, UK
[q]Manitoba Institute of Cell Biology, CancerCare Manitoba, University of Manitoba, Manitoba R3E 0V9, Canada
[r]Department of Clinical Molecular Biology (EpiGen), Medical Division, Akershus University Hospital, Lørenskog, Norway
[s]Division of Molecular Pathology, The Institute of Cancer Research, 237 Fulham Road, SW3 6JB, London, UK
[t]Department of Pathology, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA

ARTICLE INFO

Article history:
Received 24 June 2014
Received in revised form
23 July 2014
Accepted 25 July 2014
Available online 8 August 2014

Keywords:
Breast cancer
Ovarian cancer
Prognostic markers
Biomarker
Genomics
Genomic instability
DNA copy number

ABSTRACT

Complex focal chromosomal rearrangements in cancer genomes, also called "firestorms", can be scored from DNA copy number data. The complex arm-wise aberration index (CAAI) is a score that captures DNA copy number alterations that appear as focal complex events in tumors, and has potential prognostic value in breast cancer. This study aimed to validate this DNA-based prognostic index in breast cancer and test for the first time its potential prognostic value in ovarian cancer. Copy number alteration (CNA) data from 1950 breast carcinomas (METABRIC cohort) and 508 high-grade serous ovarian carcinomas (TCGA dataset) were analyzed. Cases were classified as CAAI positive if at least one complex focal event was scored. Complex alterations were frequently localized on chromosome 8p ($n = 159$), 17q ($n = 176$) and 11q ($n = 251$). CAAI events on 11q were most frequent in estrogen receptor positive (ER+) cases and on 17q in estrogen receptor negative (ER−) cases. We found only a modest correlation between CAAI and the overall rate of genomic instability (GII) and number of breakpoints ($r = 0.27$ and $r = 0.42$, $p < 0.001$). Breast cancer specific survival (BCSS), overall survival (OS) and ovarian cancer progression free survival (PFS) were used as clinical end points in Cox proportional hazard model survival analyses. CAAI positive breast cancers (43%) had higher mortality: hazard ratio (HR) of 1.94 (95%CI, 1.62–2.32) for BCSS, and of 1.49 (95%CI, 1.30–1.71) for OS. Representations of the 70-gene and the 21-gene predictors were compared with CAAI in multivariable models and CAAI was independently significant with a Cox adjusted HR of 1.56 (95%CI, 1.23–1.99) for ER+ and 1.55 (95%CI, 1.11–2.18) for ER− disease. None of the expression-based predictors were prognostic in the ER− subset. We found that a model including CAAI and the two expression-based prognostic signatures outperformed a model including the 21-gene and 70-gene signatures but excluding CAAI. Inclusion of CAAI in the clinical prognostication tool PREDICT significantly improved its performance. CAAI positive ovarian cancers (52%) also had worse prognosis: HRs of 1.3 (95%CI, 1.1–1.7) for PFS and 1.3 (95%CI, 1.1–1.6) for OS. This study validates CAAI as an independent predictor of survival in both ER+ and ER− breast cancer and reveals a significant prognostic value for CAAI in high-grade serous ovarian cancer.

1. Introduction

Breast and ovarian cancer are major causes of morbidity and death (Jemal et al., 2010). Current treatment for breast cancer includes a combination of surgery, radiotherapy, chemotherapy, endocrine agents (tamoxifen or aromatase inhibitors) in ER+ cases and trastuzumab in HER2-positive cases (Goldhirsch et al., 2009). Although clinical outcome has improved dramatically with modern therapy, the challenge remains to identify patients that could be spared overtreatment, since at 15 years 58% of patients are alive despite receiving no adjuvant chemotherapy (Early Breast Cancer Trialists' Collaborative Group (EBCTCG), 2005). Markers that could more precisely predict better clinical outcome are therefore of utmost importance.

Expression profiling of breast cancers has provided important insights into tumor biology and clinical behavior, even though clinical implementation is not yet routine (Perou et al., 2000; Sørlie et al., 2001, 2003). Based on the global pattern of gene expression, five subgroups (Luminal A, Luminal B, HER2-enriched, Basal-like and Normal-like subgroups) were identified, with different biology and clinical course. By combining DNA copy number alterations with gene expression using integrative clustering we identified 10 subtypes of breast cancer with clearly distinct genomic drivers (Curtis et al., 2012). IC1 is a moderate-risk group composed mostly of Luminal B tumors. IC2 is a high-risk group defined by an amplicon at 11q13/14, and composed of a mixture of Luminal A and B. IC3 is a low-risk Luminal A group. IC4 is a CNA-devoid group composed by a mixture of ER− tumors with lymphocytic

infiltration and ER+ tumors with abundant stroma. IC5 identifies most of the *HER2* amplified tumors. IC6 is a ZNF-703 driven poor prognosis luminal group. IC7 and IC8 are intermediate-risk Luminal A groups. IC9 is a Luminal B group with 8q gain. Finally IC10 are the core basal-like tumors with high genomic instability. One of our aims was to determine how CAAI affects prognosis in these different groupings.

Gene expression-based prognostic classifiers have been proposed to risk stratify patients (Paik et al., 2004; van de Vijver et al., 2002; van't Veer et al., 2002), and are already used in some clinical environments, but their implementation in practice awaits results of ongoing trials and will always be problematic given the intrinsic instability of RNA (Borgan et al., 2011). Furthermore it remains questionable if these genomic tests are cost effective (Hall et al., 2012), or if they improve the performance of current prognostic tools that integrate clinical and pathological parameters routinely used in the clinic, such as Adjuvant! (Ravdin et al., 2001) or PREDICT (Wishart et al., 2010, 2011, 2012). PREDICT is an online tool based on population-based cancer registry data, similar to Adjuvant!, but that in addition incorporates mode of detection, HER2-status and trastuzumab benefit (Wishart et al., 2012, 2011, 2010).

Ovarian cancer is the fifth leading cause of cancer related death in US women (Jemal et al., 2010). The majority of these deaths occur in patients with advanced stage, high-grade serous ovarian carcinomas (HGSOC) (Koonings et al., 1989; Seidman et al., 2004), despite optimal cytoreduction by debulking surgery and adjuvant chemotherapy (Ozols, 1997). Major prognostic factors are age at diagnosis, performance status, histology and residual tumor size (Winter et al., 2007, 2008). Recent studies have sought to develop molecular classifications, but these do not significantly improve prognostic performance (Cancer Genome Atlas Research Network, 2011; Etemadmoghadam et al., 2009).

Genomic rearrangements in breast cancer have distinct patterns, possibly reflecting different mechanisms of genomic instability (Hicks et al., 2006). Analysis of somatically acquired copy number alterations (CNA) has identified distinct types of structural changes (e.g. whole-arm alterations and firestorms) (Hicks et al., 2006; Russnes et al., 2010). Clustered narrow peaks of high copy number gains characterize 'firestorms'. We developed a score, the Complex Arm-wise Aberration Index (CAAI), to quantify such complex events (Russnes et al., 2010). In our original report CAAI was shown to have independent prognostic power (Russnes et al., 2010), but this finding needs to be independently validated. Here we report such independent validation of the prognostic value of CAAI in 1950 breast carcinomas (Curtis et al., 2012), and show that it adds to the gene expression-based prognostic classifiers (the 70-gene and the 21-gene signatures) (Paik et al., 2004; van de Vijver et al., 2002). We also evaluate CAAI as a prognostic marker in 508 advanced stage HGSOC (Cancer Genome Atlas Research Network, 2011).

2. Materials and methods

2.1. Breast cancer cases

A total of 1950 breast cancer cases from the METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) cohort (Curtis et al., 2012) were included in this study. Female breast cancer cases were selected on the basis of invasive histology (in-situ and benign cases were excluded, as well as cases with unknown histology). Using an eQTL-based approach (Lynch et al., 2012) a few cases with mismatched DNA/RNA were identified, and excluded. This resulted in 1950 cases with gene expression, SNP-array and clinical data available for analysis (flow chart of included samples in Appendix A). Clinical and pathological variables were collected from hospital reports. Estrogen receptor status by IHC (immunohistochemistry) was available for 1921 samples and ER status for the remaining 29 samples was scored by the expression value of *ESR1* (Lehmann et al., 2011). IHC progesterone receptor status was not available, hence expression values for *PGR* were used to score PR status (Lehmann et al., 2011). HER2 status was obtained from segmented copy number data as described in the original METABRIC report (Curtis et al., 2012). Clinical variables are presented in Supplementary Table 1.

2.2. Serous ovarian carcinomas

A total of 508 high-grade serous ovarian adenocarcinomas, recruited at the time of primary surgery, were obtained from TCGA (the Cancer Genome Atlas) (NCBI/TCGA project number 2459) (Cancer Genome Atlas Research Network, 2011). A description of the clinical variables is presented in Supplementary Table 5.

2.3. Bioinformatic and statistical analyses

CNA profiles were obtained from Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara, CA, USA). Estimation of raw copy number from arrays was performed using the CRMAv2 method implemented in the "aroma.affymetrix" R-package (Bengtsson et al., 2009). For the SNP-probes the two alleles were summarized to obtain an estimate of the total copy number. Normalized probe intensity ratios were obtained with matched normal DNA for cases where such data was available, otherwise an average of a pool of 473 adjacent normal tissue samples and 270 HapMap samples were used. Raw normalized copy number estimates were segmented into regions of constant copy number with PCF (Piecewise Constant Fit) with the smoothing parameters kmin = 5 and gamma = 200 (Nilsen et al., 2012). CAAI was then scored as previously described (Russnes et al., 2010). The method computes scores that capture complex rearrangements for each chromosomal arm separately, based on the segmented copy number data (Nilsen et al., 2012), and calls the tumor CAAI positive if the score exceeds the set threshold in at least one chromosomal arm. Gene expression data was available for the METABRIC cases and used to derive the PAM50 subtype classifications (Curtis et al., 2012), as well as representations of the 21-gene (Paik et al., 2004) and 70-gene prognostic classifiers (van't Veer et al., 2002). All statistical analyses were performed using R version 2.15.0 with the packages 'copynumber' and 'rms' (Gentleman et al., 2004; Harrell, 2014; Nilsen et al., 2012). A subset of 32 samples from METABRIC was also profiled with Molecular Inversion Probe (MIP) arrays, a method for copy number profiling optimized for use with paraffin-

embedded material. CAAI was also tested as a prognostic marker using the breast cancer prognostic tool PREDICT (www.predict.nhs.uk) (Wishart et al., 2012, 2011, 2010). The predictive performance of CAAI in the ovarian cohort was assessed by internal validation to correct for potential overfitting, given no adequate independent dataset for validation was available, using the bootstrap resampling technique (Halabi et al., 2003; Harrell, 2001). The study complies with the Reporting Recommendations for Tumor Markers (REMARK) (McShane et al., 2005). A detailed description of materials and methods can be found in Appendix A (Supplementary material and methods).

3. Results

3.1. Breast cancer

3.1.1. CAAI distribution and correlation with overall genomic instability

A total of 835 (43%) of the 1950 breast cancer cases were CAAI positive (Supplementary Figure 1). CAAI positivity was significantly associated with larger tumors, higher grade, negative ER and PR (progesterone receptor), and amplification of the *HER2* gene (Supplementary Table 1). CAAI events were most frequent on 11q in ER+ cases and 17q in ER− cases. Hierarchical clustering of the CAAI positive cases based on binary dissimilarity and Ward's method, revealed groups dominated by 1q, 8p, 11q or 17q complex alterations, and with a tendency for mutually exclusivity (Figure 1A). The proportion of CAAI positive cases in the 1950 breast cancer cases was variable across both the ten integrative subtypes recently described (Curtis et al., 2012) and across the five expression-based intrinsic subtypes (Perou et al., 2000) (Figure 1B and C).

Cancer genomes show a wide range of copy number changes and patterns, both within and between tumor types (Beroukhim et al., 2010). To explore how CAAI was related to the overall instability of the genome, we calculated the fraction of the genome affected by copy number change (Genomic instability index, GII) and counted the number of DNA breakpoints. In Figure 2 the results are presented and show only modest correlation with $r = 0.28$ and $r = 0.42$ for GII and breakpoints respectively ($p < 0.001$). CAAI positive cases can be found in tumors with a low number of breakpoints and tumors with low degree of genomic instability. This suggests CAAI and overall genomic instability capture distinct processes and are the result of different underlying biological mechanisms.

3.1.2. CAAI is validated as a prognostic index

CAAI status was significantly associated with outcome, both BCSS and OS (Figure 3). This association persisted when cases were stratified for lymph node and ER status. Importantly, the separation between good and poor prognosis was independent of adjuvant systemic therapy (Supplementary Figure 2). Results from univariable Cox regression analyses are summarized in Supplementary Table 2. Positive CAAI status was prognostic for both BCSS (HR 1.94 (95%CI, 1.62–2.32; $p < 0.001$)) and OS (HR 1.49 (95%CI, 1.30–1.71; $p < 0.001$)). CAAI status was significantly different distributed in the intrinsic subtypes, with CAAI positive tumors most frequently found in HER2-enriched (61.8%) and Luminal B (59.9%) subgroups. The frequency of CAAI positive tumors in Basal-like, Luminal A and Normal-like subgroups were 40.8%, 30.4% and 26.3% respectively. CAAI status was associated with higher mortality within each of the intrinsic subtypes, except for Luminal B tumors (Supplementary Figure 3). CAAI was also prognostic in IC4, IC7 and borderline significant in IC10 (Supplementary Figure 4). The HER2-enriched Pam50 subgroup overlap with, but is not equal to the clinical *HER2* amplified group. In IC5 all tumors have *HER2* gene amplification, but not all HER2+ tumors belong to IC5. CAAI was prognostic in *HER2* amplified tumors (by SNP-array) and in the Pam50 HER2-enriched group, but not in IC5 subgroup (Supplementary Figure 5). A total of 64.4% of the *HER2* amplified tumors by SNP-array were CAAI positive. Interestingly only 139 of the 279 CAAI positive *HER2* amplified tumors (49.8%) had a complex alteration affecting chromosome 17q, the remainder had complex events elsewhere in the genome, most frequently affecting 11q and 8p. This shows that not all *HER2* amplified cases have a complex structural alteration affecting the *HER2* amplicon, but rather a more simple amplification.

The CAAI score is a continuous numeric value (range 0.0–22.4 in the 1950 breast cancer cases). To classify samples as CAAI positive and negative (dichotomous variable) we used the same threshold value of 0.5 as in our original report (Russnes et al., 2010). Using a log2 transformation of the continuous CAAI as a variable in a Cox regression model showed an HR of 1.46 (95%CI, 1.33–1.61; $p < 0.001$). This significant association between the continuous numeric CAAI score and outcome illustrates that results are not dependent on the chosen threshold, but since the present study is an independent validation of the prognostic value of CAAI, we used the dichotomized variable for all the multivariable models (see below).

3.1.3. Multivariable survival analysis

Results from multivariable Cox regression models for BCSS are presented in Table 1 (and for OS are shown in Supplementary Table 3). When comparing ER+ and ER− tumors the proportional hazard varies over time (Blows et al., 2010). We therefore fitted separate Cox models for ER+ and ER− cases. In the 1412 ER+ cases CAAI status, tumor size, lymph node status, age, histologic type (ILC vs. IDC) and the surrogate 21-gene predictor were significant after adjustment for covariates. Positive CAAI had an HR of 1.56 (95%CI, 1.23–1.99; $p < 0.001$). A total of 421 ER− cases were available for multivariable analysis, and only axillary lymph node status, histological type and CAAI status remained significant (HR 1.55 (95%CI, 1.11–2.18; $p = 0.011$)). Neither of the two expression-based signatures (that is surrogates to resemble Mamma Print and OncoTypeDX, see below), was found to be significant in ER− disease.

3.1.4. CAAI provides additional stratification to gene expression classifiers

The prognostic value of the two gene expression prognostic classifiers (Paik et al., 2004; van de Vijver et al., 2002) was

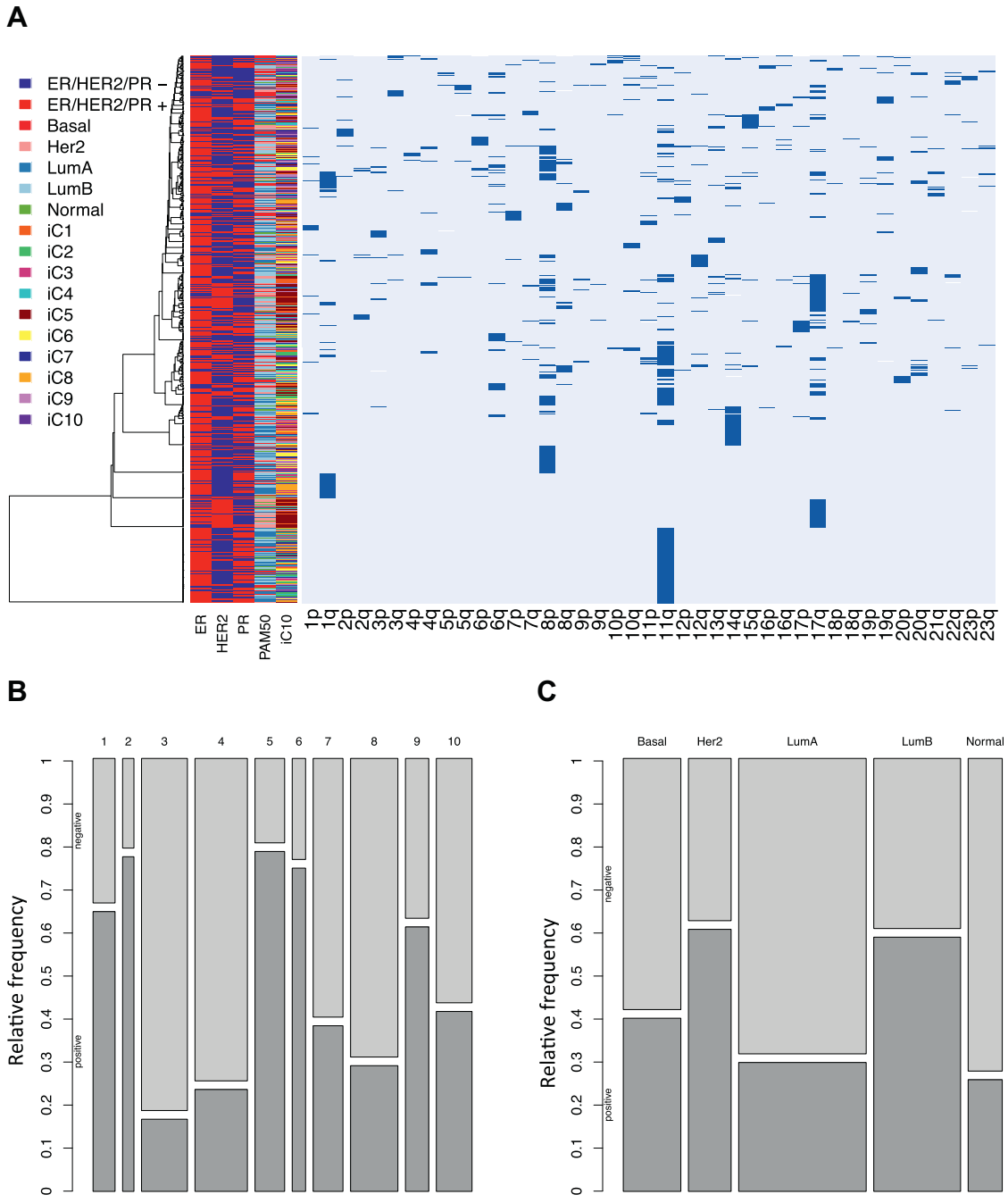

Figure 1 – *Distribution of complex events in breast cancer* **Hierarchical clustering of the CAAI positive cases is shown in panel A, with CAAI positive events in blue and negative in gray. The distribution of CAAI positive cases in molecular subtypes of breast cancer is shown in B (integrative cluster subtypes) and C (PAM50 subtypes).**

tested in the 1950 breast cancer cases using surrogate 21-gene and 70-gene signatures as described in supplementary information and presented in Supplementary Figure 6. CAAI status was able to further stratify each of the gene expression risk groups (Supplementary Figure 7). CAAI provides added prognostic information in all subsets, with the strongest effect in the high-risk groups. To evaluate the added predictive ability of a new biomarker, such as CAAI, is a complex task, as statistical significance is not always the same as clinical significance (Pencina et al., 2008). For this reason, we compared the performance of prognostic models in the complete dataset, and in ER+ and ER– disease separately, using three different statistical measures proposed in the literature: the likelihood ratio test, the C-index (Harrell et al., 1996), a measure of the concordance between predicted and observed survival times, and finally U-statistics based on net classification improvements (Pencina et al., 2011), to formally test the significance of the increase in the C-index. The results of these analyses, presented in Supplementary Table 4, confirm that a model including CAAI and the two expression-based

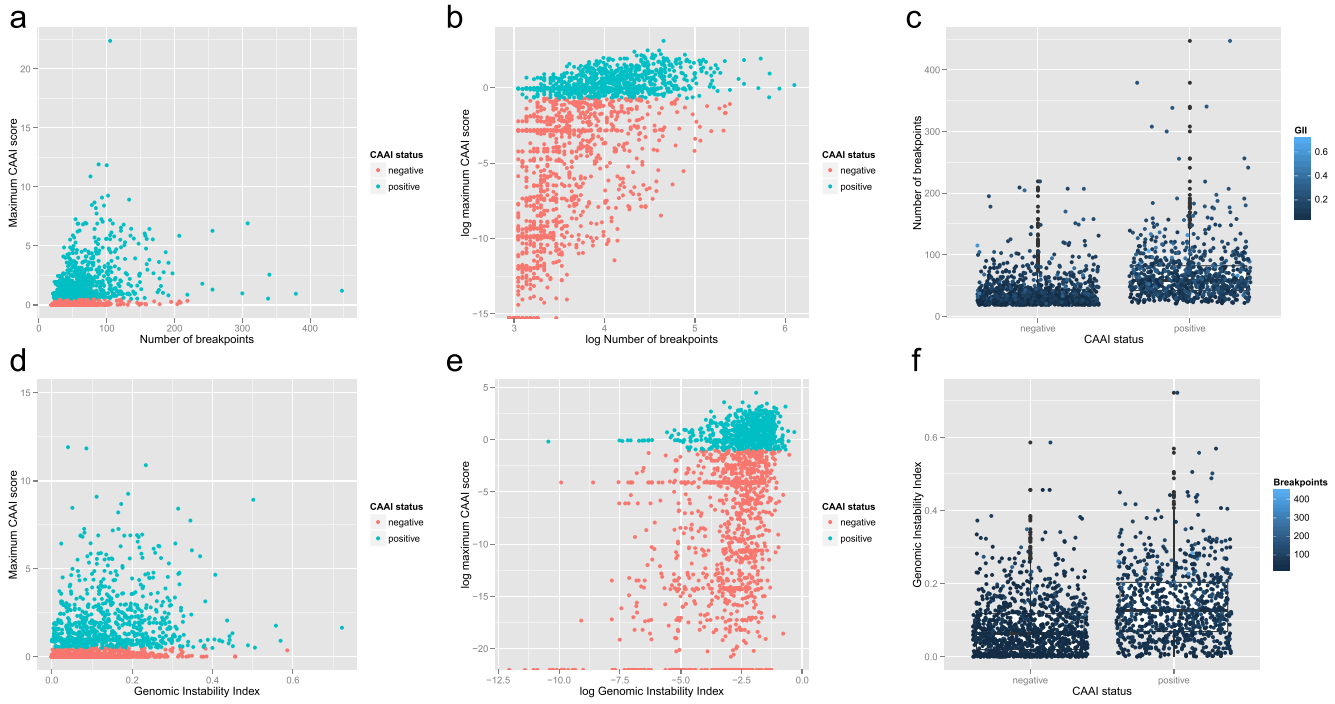

Figure 2 – *CAAI and overall Genomic Instability* **Scatter plot of maximum CAAI score and Genomic Instability Index on the actual (A) and a logarithmic (B) scale, colored by CAAI status. Boxplot of GII in CAAI negative and positive cases, colored by number of breakpoints (C). Scatter plot of maximum CAAI score and the number of DNA breakpoints (D), and on a logarithmic (E) scale, colored by CAAI status. Boxplot of number of DNA breakpoints in CAAI negative and positive cases, colored by GII (F).**

prognostic signatures outperforms a model including the 21-gene and 70-gene signatures but excluding CAAI.

3.2. *Ovarian cancer*

In HGSOC CAAI positivity was detected in 262 of the 508 (52%) available cases. The distribution of clinical variables in the data set is shown in Supplementary Table 5. The CAAI scores were generally higher in the ovarian cohort than in breast cancer. Univariable Cox regression of the continuous CAAI-score (log2 transformed) showed a significant HR of 1.3 (95%CI, 1.2–1.5; $p < 0.01$). We chose a threshold of 1.0 for the subsequent analyses to reflect the more rearranged genome of ovarian cancer. Kaplan–Meier plots illustrate the prognostic impact of CAAI status on progression free survival (PFS) ($p = 0.013$) and OS ($p < 0.001$) (Figure 4). Univariable Cox regression models for all clinical variables are presented in Supplementary Table 6. CAAI positive cases had an increased risk of progression with HR 1.3 (95%CI, 1.1–1.6; $p = 0.01$) and for death of any cause with HR 1.4 (95%CI, 1.2–1.7; $p \leq 0.01$).

The results from multivariable Cox regression analyses are presented in Table 2. Age at diagnosis did not meet the proportional hazards (PH) assumption and the model was stratified by a categorical representation of age (≤60 vs. >60). Positive CAAI increased risk of relapse, independent of other covariates with a HR of 1.3 (95%CI, 1.1–1.7; $p = 0.01$). CAAI was together with histological grade a predictor of OS, with HR of 1.3 (95%CI, 1.1–1.6; $p < 0.01$). Since no adequate external/independent dataset was available, an internal validation of the multivariable models was performed using bootstrapping to correct for potential over-fitting (Halabi et al., 2003; Harrell, 2001). Supplementary table 7 shows the result of the validation using several indexes (details in supplementary material). The amount of over-fitting in the models was modest, as reflected by the estimated optimism. Supplementary Figure 8 shows the predictions from the models at 1 and 2 years of follow up compared with the actual survival probability (calibration accuracy). The agreement between the observed and predicted survival was highly concordant; mean optimism 0.002 for PFS at both time points and 0.003 and −0.002 for OS at 1 and 2 years of observation respectively.

3.3. *Possible clinical utility of CAAI in breast cancer*

3.3.1. *CAAI in FFPE tumor material*

Most aCGH-based analyses rely on availability of fresh frozen tumor material. Molecular inversion probe (MIP) arrays have been developed for analysis of FFPE material and this could facilitate clinical implementation of CAAI analysis. We therefore compared 32 samples hybridized to both Affymetrix SNP 6.0 and MIP platforms. Supplementary Figure 9 shows the good correlation between CAAI scores obtained with the two platforms in matched fresh-frozen and paraffin-embedded samples. Using an optimized threshold for the MIP data the agreement between methods was substantial: Cohen's Kappa value of 0.75. These preliminary results suggest that CAAI scoring could be implemented using formalin fixed paraffin-embedded tumor material.

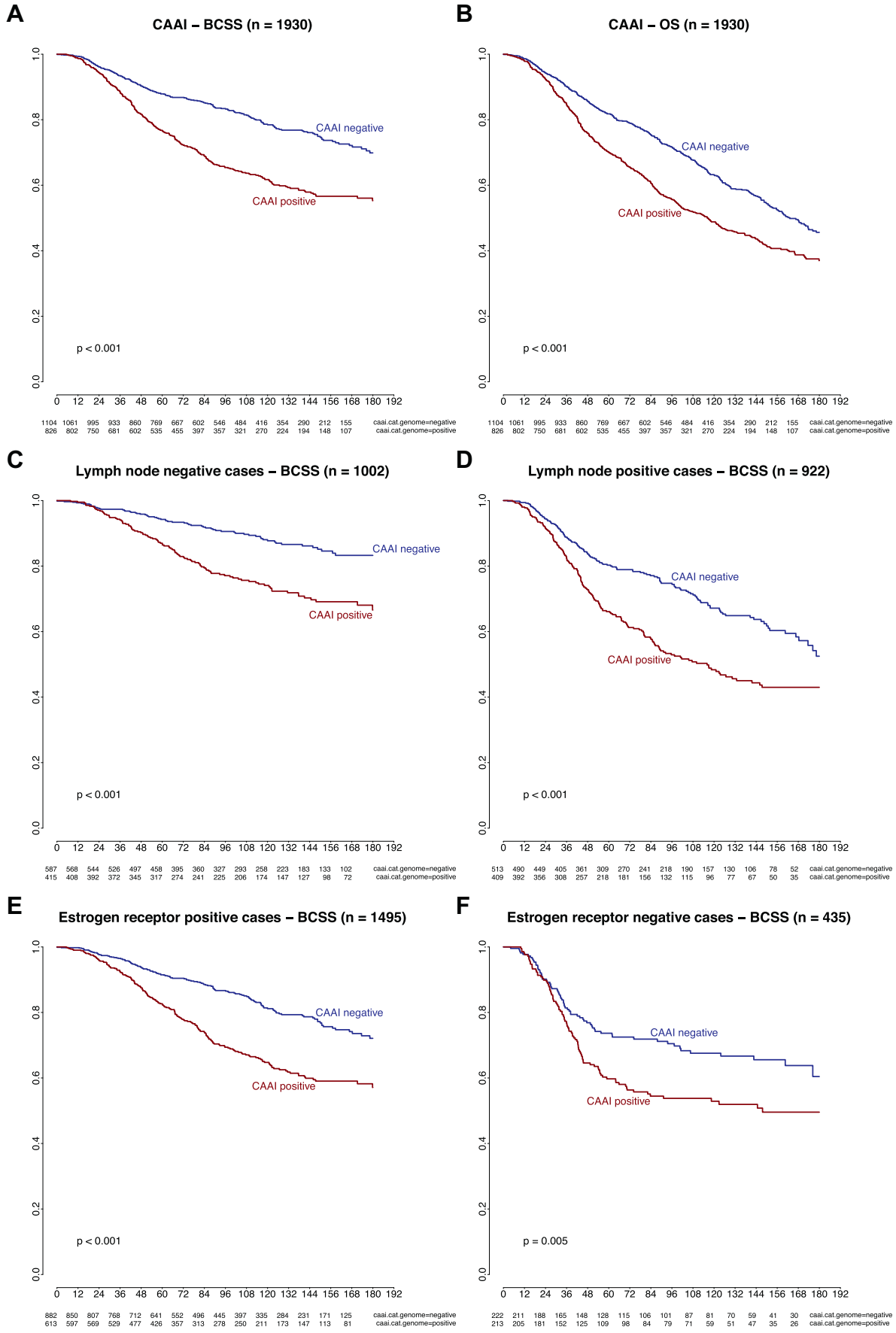

Figure 3 — ***Kaplan–Meier estimates of outcome in breast cancer*** **In A and B survival estimates for CAAI positive and negative cases are shown for breast cancer specific survival (BCSS) and overall survival (OS) respectively. Outcome for lymph node negative and lymph node positive cases are shown in C and D. E and F shows survival estimates for ER+ and ER− cases.**

Table 1 – Multivariable Cox regression model with breast cancer specific survival as outcome variable.

Variable	ER positive cases ($n = 1412$)				ER negative cases ($n = 421$)			
	HR	95% CI		p-value	HR	95% CI		p-value
		Lower	Upper			Lower	Upper	
Any genomic complex events								
CAAI pos vs. neg	1.56	1.23	1.99	**< 0.001**	1.55	1.11	2.18	**0.011**
Histological grade								
Categories as ordinal	1.13	0.92	1.38	0.247	0.77	0.49	1.23	0.281
Tumor size								
pT2 vs. pT1	1.50	1.17	1.93	**0.001**	1.36	0.94	1.96	0.105
pT3 vs. pT1	3.62	2.30	5.70	**< 0.001**	1.08	0.55	2.11	0.819
Axillary lymph node status								
Positive vs. negative	2.27	1.78	2.88	**< 0.001**	2.07	1.37	3.14	**0.001**
HER2 status (from arrays)								
Positive vs. negative	0.99	0.75	1.31	0.941	0.99	0.70	1.40	0.967
Progesterone receptor status								
Positive vs. negative	1.05	0.81	1.35	0.731	0.86	0.44	1.66	0.646
Age at diagnosis								
Continous variable	1.01	1.00	1.02	**0.024**	0.99	0.97	1.00	0.167
Histological type								
ILC vs. IDC	1.73	1.15	2.60	**0.009**	1.25	0.49	3.14	0.641
Other invasive vs. IDC	1.05	0.68	1.60	0.838	0.36	0.15	0.86	**0.021**
70-gene classifier								
Poor vs. good prognosis	1.25	0.95	1.64	0.110	1.00	0.54	1.86	0.999
21-gene classifier								
Moderate vs. good prognosis	1.23	0.91	1.68	0.179	1.21	0.35	4.23	0.763
Poor vs. good prognosis	2.35	1.64	3.36	**< 0.001**	1.14	0.35	3.69	0.824

Separate models for ER positive and negative cases. Significant p-values (<0.05) in bold. Both models were stratified for site of inclusion.

3.3.2. *Implementation of CAAI in the prognostic tool PREDICT (Wishart et al., 2012, 2011, 2010)*

CAAI as a prognostic marker was added to the prognostic model PREDICT (www.predict.nhs.uk) (Wishart et al., 2012, 2011, 2010), which is based on a Cox proportional hazard model. PREDICT was developed using a case-cohort of breast cancer cases of unknown CAAI status and so the underlying, baseline hazard is representative of cases of average CAAI status. The CAAI hazard ratio estimate is for CAAI positive compared to CAAI negative cases, and so these were rescaled to give an average hazard ratio of unity using an estimated prevalence of CAAI positivity of 40 percent. We then compared the performance of PREDICT with and without the addition of CAAI in predicting breast cancer specific mortality at five years after diagnosis using calibration, discrimination and reclassification as measures of model performance. Model calibration is a comparison of the predicted mortality estimates from each model with the observed mortality. Model discrimination was evaluated by calculating the area under the receiver-operator-characteristic curve (AUC) calculated for 5-years breast cancer specific mortality. This is a measure of how well the models identify those patients with poorer survival. The AUC is the probability that the predicted mortality from a randomly selected patient who died will be

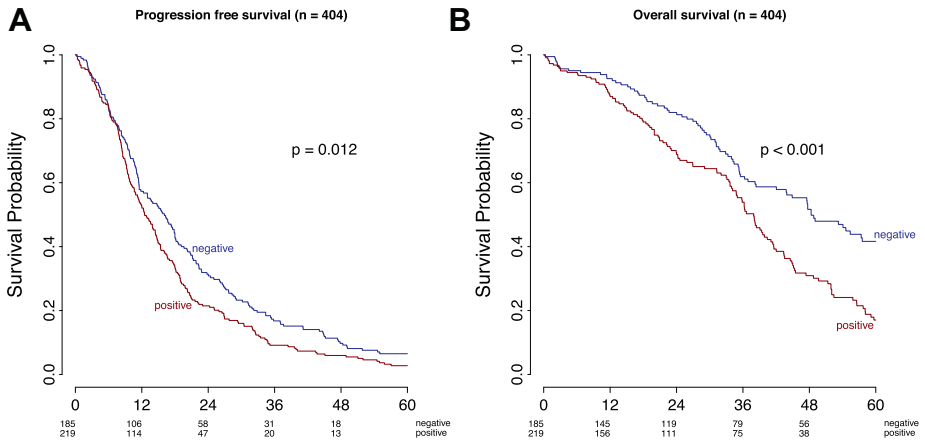

Figure 4 – *Kaplan–Meier estimates of outcome in ovarian cancer* Survival estimates for CAAI positive vs. CAAI negative cases with progression free survival (A) and overall survival (B).

Table 2 – Multivariable Cox regression models in ovarian cancer.

Variable	Progression free survival				Overall survival			
	HR	95% Confidence interval		p-value	HR	95% Confidence interval		p-value
		Lower	Upper			Lower	Upper	
Any complex genomic events (CAAI)								
Positive vs. negative	1.3	1.1	1.6	**0.01**	1.8	1.3	2.4	**<0.01**
FIGO stage								
Stage IV vs. stage III	1.2	0.9	1.6	0.13	1.2	0.9	1.8	0.26
Histological grade								
Grade 3 vs. Grade 2	1.4	1.0	2.0	**0.05**	1.2	0.8	1.9	0.38
Residual disease								
Suboptimal (>10 mm residual tumor) vs. optimal (<10 mm)	0.9	0.7	1.2	0.49	1.1	0.8	1.5	0.69

The model was stratified for age. Significant *p*-values (<0.05) in bold.
$n = 368$ (38 excluded due to missing data).

higher than the predicted mortality from a randomly selected survivor. Reclassification is the extent to which a new model classified individuals into specified risk groups. PREDICT is used in a clinical setting to identify patients most likely to benefit from adjuvant chemotherapy. The Cambridge Breast Unit multi-disciplinary team uses the estimated absolute benefit of chemotherapy at ten years to define three patient groups: those with a predicted absolute benefit of less than 3 percent in whom chemotherapy is not recommended, those with a predicted absolute benefit of 3–5 percent in whom the balance of risks and benefits are discussed with the patients, and those with an absolute benefit of greater than 5 percent for whom chemotherapy is generally recommended. The absolute benefit is approximately proportional to the absolute risk. We therefore classified the 1950 cases from the breast cancer cohort into three groups based on thresholds for estimated absolute benefit at five years of 1.5 percent and 2.5 percent.

We tested whether the addition of CAAI to the PREDICT prognostic model (modified PREDICT) could improve both model calibration and model discrimination in the breast cancer cohort of 1950 cases. In the first five years following diagnosis there were 299 breast cancer deaths, this compared with 276 predicted using PREDICT and 283 using the modified PREDICT. The improved calibration was seen for both ER+ and ER− disease. The discrimination for the modified PREDICT (AUC = 0.775) was significantly better ($p = 0.005$) than for PREDICT (AUC = 0.764). PREDICT is typically used in the clinical setting to estimate the absolute benefit of adjuvant chemotherapy. Supplementary Table 8 shows the stratification of the breast cancer cases into the three absolute benefit groups: less than 1.5% (when chemotherapy is not recommended), 1.5–2.5% (when chemotherapy is discussed), and greater than 2.5% (when chemotherapy is recommended). 1706 patients remained in the same group under both models. Of the 244 cases that changed risk groups using modified PREDICT, 134 moved into a group with a smaller predicted benefit, potentially sparing them chemotherapy, and 110 moved to a higher risk group, potentially resulting in them receiving the added benefit of chemotherapy.

4. Discussion

In this study we validate CAAI, a measure of complex structural alterations in cancer genomes, as an independent prognostic biomarker in breast cancer, and reveal its significant prognostic value also in ovarian cancer. The size of the METABRIC breast cancer cohort (1950 cases) and the comprehensive clinical data available enabled stratification by relevant subgroups, and showed that CAAI is an independent predictor of outcome in both ER− and ER+ disease. The prognostic effect was found in both chemotherapy naïve and chemotherapy treated patients. The novel finding of the prognostic effect of complex focal DNA copy number change in ovarian cancer, captured by CAAI, is interesting as the method was developed for breast cancer. Since it has proven challenging to identify prognostic markers in ovarian cancer, the modest effect of CAAI needs independent validation.

Analyses of copy number data rely on a robust smoothing of the raw data into estimates of the underlying copy number state. Different algorithms for such analysis exist and the individual methods have different settings to control the sensitivity and the specificity of calls. In this analysis PCF was used after normalization of CEL files, to make it as similar to our previous results as possible. This was essential since our aim was to validate the prognostic value of CAAI in breast cancer. Threshold parameters were kept constant for the validation in breast cancer, but were adjusted in ovarian cancer due to the more aneuploid background of these genomes. As for most novel tests, thresholds for calling of positives over negatives could be optimized and improved. The prognostic effect of CAAI used as a continuous score, both in breast and ovarian cancer, was clear, indicating that results are not dependent on the thresholds used.

Because we had matched gene expression profiling data for all the 1950 breast cancers cases we could use this data to implement surrogate indices for the 21-gene and 70-gene prognostic gene expression classifiers that are currently being validated in two large clinical trials; MINDACT (Cardoso et al., 2007, 2008; Rutgers et al., 2011) (ClinicalTrials.gov nr.

NCT00433589) and TAILORx (Sparano, 2006; Sparano and Paik, 2008) (ClinicalTrials.gov nr. NCT00310180). We could therefore show, using a multivariable Cox regression model, that the prognostic value of CAAI was independent of these gene expression classifiers. Moreover we could also robustly evaluate, using three statistical measures, the added value of CAAI in a model that also included the 21 and 70 gene expression signatures. These results indicate that even if MINDACT or TAILORx demonstrate the clinical utility of gene expression classifiers, CAAI would add further prognostic information and therefore have clinical impact.

CAAI was designed to capture focal complex copy number aberrations. Our comparison with overall genome instability, measured by GII and number of breakpoints, shows that CAAI is only modestly correlated with these. This finding suggests that there are distinct biological processes underlying these distinct patterns. The CAAI index captures localized regions of copy number change, but in a genomic profile that is characterized by extensive structural variation, the index may call other patterns as well. On the other hand the method is sensitive to score genomes that overall have few changes, but with distinct localized clustered events. A total of 35.6% (154/433) of the *HER2* amplified tumors (by SNP-array) are CAAI negative, showing that a single amplicon containing a known driver gene does not necessarily result in a sample being called CAAI positive. Luminal B and the HER2-enriched tumors are more often CAAI positive than others. Intriguingly CAAI was not significantly associated with survival in Luminal B tumors. We noted that curves in the KM-plot joined towards the end of the 15 years observation time, probably reflecting non-breast cancer mortality with very long follow up time.

The exact mechanism by which firestorms affect clinical outcome is not known. Firestorms are thought to arise from breakage-fusion-bridge cycles, either associated with fragile sites, or as result of recombination at palindromic sites at shortened telomeres (Hicks et al., 2006), resulting in focal amplifications, frequently at sites of known oncogenes such as *ERBB2* or *CCND1*. These genomic events could lead not only to increased tumor aggressiveness (proliferation, invasiveness and metastasis) but also underpin clonal evolution as a driver of drug resistance (Aparicio and Caldas, 2013). All of these features would combine to portend worse outcome for CAAI positive breast and ovarian cancers.

The potential clinical utility of CAAI was demonstrated in breast cancer, showing it could be implemented without requiring fresh frozen tumor material, and improving calibration and discrimination of the clinically used PREDICT prognostication tool by the addition of CAAI. A test based on DNA extracted from formalin fixed paraffin-embedded tissue blocks, rather than RNA from fresh frozen tumor material, would significantly facilitate implementation in routine diagnostic labs.

Author contributions

HKMV, OMR, PDPP and CC lead the analysis. HKMV and CC wrote the manuscript with contributions from all authors. S-FC, ChC, GT, SS, OCL, YY, CKN, MD, ED, VNK, SM contributed with data collection and data analysis. GT, EP, StS, IOL, SP, AP and LCM provided clinical and pathological expertise. JDB, PDPP, ALBD, SA and CC supervised data collection and analysis. SA and CC co-conceived and oversaw the METABRIC study and are joint senior authors and project co-leaders.

Acknowledgments

We thank all the patients that contributed to the study. The authors also thank Peter Van Loo, Hege G Russnes and Gro Nilsen for support, help with analyses and fruitful discussions. The results published here are in part based upon data generated by The Cancer Genome Atlas pilot project established by the NCI and NHGRI (project 2459). Information about TCGA and the investigators and institutions who constitute the TCGA research network can be found at http://cancergenome.nih.gov/. Funding: METABRIC was funded by Cancer Research UK and the British Columbia Cancer Foundation. HKMV received funding from European Association for Cancer Research (EACR) travel fellowship award 2010, Kjøpmann Einar Unsgaard og hustru Kitty Unsgaards legat 2010 and The Norwegian Radium Hospital Foundation. This study is a part of the EUROCAN platform (FP7, Grant number 260791). This work was partly supported by the Research Council of Norway through its Centres of Excellence funding scheme, project number 179571. AP has received financial support from the Department of Health via the NIHR Comprehensive Biomedical Research Centre award to Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London and King's College Hospital NHS Foundation Trust. SA has received funding from BC Cancer Foundation, Canadian Breast Cancer Foundation BC/Yukon, and Canada Research Chairs Program.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.molonc.2014.07.019.

The METABRIC Group

Co-chairs Carlos Caldas[1,2], Samuel Aparicio[3,4]

Steering committee James D. Brenton[1,2], Ian Ellis[5], David Huntsman[3,4], Sarah Pinder[6], Arnie Purushotham[6], Leigh Murphy[7], Carlos Caldas[1,2], Samuel Aparicio[3,4]

Tissue and clinical data source sites: University of Cambridge/Cancer Research UK Cambridge Research Institute Carlos Caldas (Principal Investigator)[1,2]; Helen Bardwell[2], Suet-Feung Chin[1,2], Christina Curtis[1,2,*], Zhihao Ding[2], Stefan Gräf[1,2], Linda Jones[8], Bin Liu[1,2], Andy G. Lynch[1,2], Irene Papatheodorou[1,2], Stephen J. Sammut[9], Gordon Wishart[9]; **British Columbia Cancer Agency** Samuel Aparicio (Principal Investigator)[3,4], Steven Chia[4], Karen Gelmon[4], David Huntsman[3,4], Steven McKinney[3,4], Caroline Speers[4], Gulisa Turashvili[3,4], Peter Watson[3,4,7]; **University of Nottingham** Ian Ellis (Principal

Investigator)[5], Roger Blamey[5], Andrew Green[5], Douglas Macmillan[5], Emad Rakha[5]; **King's College London** Arnie Purushotham (Principal Investigator)[6], Cheryl Gillett[6], Anita Grigoriadis[6], Sarah Pinder[6], Emanuele di Rinaldis[6], Andy Tutt[6]; **Manitoba Institute of Cell Biology** Leigh Murphy (Principal Investigator)[7], Michelle Parisien[7], Sandra Troup[7]

Cancer genome/transcriptome characterization centers: University of Cambridge/Cancer Research UK Cambridge Research Institute Carlos Caldas (Principal Investigator)[1,2], Suet-Feung Chin (Team Leader)[1,2], Derek Chan[1], Claire Fielding[2], Ana-Teresa Maia[1,2], Sarah McGuire[2], Michelle Osborne[2], Sara M. Sayalero[2], Inmaculada Spiteri[2], James Hadfield[2]; **British Columbia Cancer Agency** Samuel Aparicio (Principal Investigator)[3,4], Gulisa Turashvili (Team Leader)[3,4], Lynda Bell[4], Katie Chow[4], Nadia Gale[4], David Huntsman[3,4], Maria Kovalik[4], Ying Ng[4], Leah Prentice[4]

Data analysis subgroup: University of Cambridge/Cancer Research UK Cambridge Research Institute Carlos Caldas (Principal Investigator)[1,2], Simon Tavaré (Principal Investigator)[1,2,10,11], Christina Curtis (Team Leader)[1,2,*], Mark J. Dunning[2], Stefan Gräf[1,2], Andy G. Lynch[1,2], Oscar M. Rueda[1,2], Roslin Russell[2], Shamith Samarajiwa[1,2], Doug Speed[2,10]; Florian Markowetz (Principal Investigator)[1,2], Yinyin Yuan[1,2]; James D. Brenton (Principal Investigator)[1,2]; **British Columbia Cancer Agency** Samuel Aparicio (Principal Investigator)[3,4], Sohrab P. Shah (Team Leader)[3,4], Ali Bashashati[3], Gavin Ha[3], Gholamreza Haffari[3], Steven McKinney[3,4]

[1]Department of Oncology, University of Cambridge, Hills Road, Cambridge CB2 2XZ, UK. [2]Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. [3]Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia V6T 2B5, Canada. [4]Molecular Oncology, British Columbia Cancer Research Centre, Vancouver, British Columbia V5Z 1L3, Canada. [5]Department of Histopathology, School of Molecular Medical Sciences, University of Nottingham, Nottingham NG5 1PB, UK. [6]King's College London, Breakthrough Breast Cancer Research Unit, London, WC2R 2LS, UK. [7]Manitoba Institute of Cell Biology, University of Manitoba, Manitoba R3E 0V9, Canada. [8]Cambridge Experimental Cancer Medicine Centre, Cambridge CB2 0RE, UK. [9]Cambridge Breast Unit, Addenbrooke's Hospital, Cambridge University Hospital NHS Foundation Trust and NIHR Cambridge Biomedical Research Centre, Cambridge CB2 2QQ, UK. [10]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Centre for Mathematical Sciences, Cambridge CB3 0WA, UK. [11]Molecular and Computational Biology Program, University of Southern California, Los Angeles, California 90089, USA. *Present address: Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA.

REFERENCES

Aparicio, S., Caldas, C., 2013. The implications of clonal genome evolution for cancer medicine. N. Engl. J. Med. 368, 842–851. http://dx.doi.org/10.1056/NEJMra1204892.

Bengtsson, H., Wirapati, P., Speed, T.P., 2009. A single-array preprocessing method for estimating full-resolution raw copy numbers from all affymetrix genotyping arrays including GenomeWideSNP 5 & 6. Bioinformatics 25, 2149–2156. http://dx.doi.org/10.1093/bioinformatics/btp371.

Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., Mc Henry, K.T., Pinchback, R.M., Ligon, A.H., Cho, Y.-J., Haery, L., Greulich, H., Reich, M., Winckler, W., Lawrence, M.S., Weir, B.A., Tanaka, K.E., Chiang, D.Y., Bass, A.J., Loo, A., Hoffman, C., Prensner, J.R., Liefeld, T., Gao, Q., Yecies, D., Signoretti, S., Maher, E., Kaye, F.J., Sasaki, H., Tepper, J.E., Fletcher, J.A., Tabernero, J., Baselga, J., Tsao, M.-S., Demichelis, F., Rubin, M.A., Janne, P.A., Daly, M.J., Nucera, C., Levine, R.L., Ebert, B.L., Gabriel, S., Rustgi, A.K., Antonescu, C.R., Ladanyi, M., Letai, A., Garraway, L.A., Loda, M., Beer, D.G., True, L.D., Okamoto, A., Pomeroy, S.L., Singer, S., Golub, T.R., Lander, E.S., Getz, G., Sellers, W.R., Meyerson, M., 2010. The landscape of somatic copy-number alteration across human cancers. Nature 463, 899–905. http://dx.doi.org/10.1038/nature08822.

Blows, F.M., Driver, K.E., Schmidt, M.K., Broeks, A., van Leeuwen, F.E., Wesseling, J., Cheang, M.C.U., Gelmon, K., Nielsen, T.O., Blomqvist, C., Heikkilä, P., Heikkinen, T., Nevanlinna, H.A., Akslen, L.A., Bégin, L.R., Foulkes, W.D., Couch, F.J., Wang, X., Cafourek, V., Olson, J.E., Baglietto, L., Giles, G.G., Severi, G., McLean, C.A., Southey, M.C., Rakha, E., Green, A.R., Ellis, I.O., Sherman, M.E., Lissowska, J., Anderson, W.F., Cox, A., Cross, S.S., Reed, M.W.R., Provenzano, E., Dawson, S.-J., Dunning, A.M., Humphreys, M.K., Easton, D.F., Garcia-Closas, M., Caldas, C., Pharoah, P.D., Huntsman, D.G., 2010. Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. PLoS Med. 7, e1000279. http://dx.doi.org/10.1371/journal.pmed.1000279.

Borgan, E., Navon, R., Vollan, H.K.M., Schlichting, E., Sauer, T., Yakhini, Z., Lingjaerde, O.C., Sørlie, T., Børresen-Dale, A.-L., 2011. Ischemia caused by time to freezing induces systematic microRNA and mRNA responses in cancer tissue. Mol. Oncol.. http://dx.doi.org/10.1016/j.molonc.2011.08.004.

Cancer Genome Atlas Research Network, 2011. Integrated genomic analyses of ovarian carcinoma. Nature 474, 609–615. http://dx.doi.org/10.1038/nature10166.

Cardoso, F., Piccart-Gebhart, M., van't Veer, L.J., Rutgers, E., TRANSBIG Consortium, 2007. The MINDACT trial: the first prospective clinical validation of a genomic tool. Mol. Oncol. 1, 246–251. http://dx.doi.org/10.1016/j.molonc.2007.10.004.

Cardoso, F., van't Veer, L.J., Rutgers, E., Loi, S., Mook, S., Piccart-Gebhart, M.J., 2008. Clinical application of the 70-gene profile: the MINDACT trial. J. Clin. Oncol. 26, 729–735. http://dx.doi.org/10.1200/JCO.2007.14.3222.

Curtis, C., Shah, S.P., Chin, S.-F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S.A., Yuan, Y., Gräf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S., METABRIC Group, Langerød, A., Green, A., Provenzano, E., Wishart, G.C., Pinder, S.E., Watson, P., Markowetz, F., Murphy, L., Ellis, I., Purushotham, A., Børresen-Dale, A.-L., Brenton, J.D., Tavaré, S., Caldas, C., Aparicio, S., 2012. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486, 346–352. http://dx.doi.org/10.1038/nature10983.

Early Breast Cancer Trialists' Collaborative Group (EBCTCG), 2005. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 365, 1687–1717. http://dx.doi.org/10.1016/S0140-6736(05)66544-0.

Etemadmoghadam, D., DeFazio, A., Beroukhim, R., Mermel, C., George, J., Getz, G., Tothill, R., Okamoto, A., Raeder, M.B.,

Harnett, P., Lade, S., Akslen, L.A., Tinker, A.V., Locandro, B., Alsop, K., Chiew, Y.-E., Traficante, N., Fereday, S., Johnson, D., Fox, S., Sellers, W.R., Urashima, M., Salvesen, H.B., Meyerson, M., Bowtell, D.D.L.AOCS Study Group, 2009. Integrated genome-wide DNA copy number and expression analysis identifies distinct mechanisms of primary chemoresistance in ovarian carcinomas. Clin. Cancer Res. 15, 1417–1427. http://dx.doi.org/10.1158/1078-0432.CCR-08-1564.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A.J., Sawitzki, G., Smith, C., Smyth, G.K., Tierney, L., Yang, J.Y.H., Zhang, J., 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5, R80. http://dx.doi.org/10.1186/gb-2004-5-10-r80.

Goldhirsch, A., Ingle, J.N., Gelber, R.D., Coates, A.S., Thurlimann, B., Senn, H.J., Panel members, 2009. Thresholds for therapies: highlights of the St Gallen International Expert Consensus on the primary therapy of early breast cancer 2009. Presented at the Ann. Oncol.: Official Journal of the European Society for Medical Oncology/ESMO, 1319–1329. http://dx.doi.org/10.1093/annonc/mdp322.

Halabi, S., Small, E.J., Kantoff, P.W., Kattan, M.W., Kaplan, E.B., Dawson, N.A., Levine, E.G., Blumenstein, B.A., Vogelzang, N.J., 2003. Prognostic model for predicting survival in men with hormone-refractory metastatic prostate cancer. J. Clin. Oncol. 21, 1232–1237.

Hall, P.S., McCabe, C., Stein, R.C., Cameron, D., 2012. Economic evaluation of genomic test-directed chemotherapy for early-stage lymph node-positive breast cancer. J. Natl. Cancer Inst. 104, 56–66. http://dx.doi.org/10.1093/jnci/djr484.

Harrell, F.E., 2001. Regression Modeling Strategies. Springer Verlag.

Harrell, F.E., 2014. rms: Regression Modeling Strategies. R package version 4.1-3. http://CRAN.R-project.org/package=rms.

Harrell, F.E., Lee, K.L., Mark, D.B., 1996. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat. Med. 15, 361–387. http://dx.doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2–4.

Hicks, J.B., Krasnitz, A., Lakshmi, B., Navin, N.E., Riggs, M., Leibu, E., Esposito, D., Alexander, J., Troge, J., Grubor, V., Yoon, S., Wigler, M., Ye, K., Børresen-Dale, A.-L., Naume, B., Schlichting, E., Norton, L., Hägerström, T., Skoog, L., Auer, G., Månér, S., Lundin, P., Zetterberg, A., 2006. Novel patterns of genome rearrangement and their association with survival in breast cancer. Genome Res. 16, 1465–1479. http://dx.doi.org/10.1101/gr.5460106.

Jemal, A., Siegel, R., Xu, J., Ward, E., 2010. Cancer statistics, 2010. CA Cancer J. Clin. 60, 277–300. http://dx.doi.org/10.3322/caac.20073.

Koonings, P.P., Campbell, K., Mishell, D.R., Grimes, D.A., 1989. Relative frequency of primary ovarian neoplasms: a 10-year review. Obstet. Gynecol. 74, 921–926.

Lehmann, B.D., Bauer, J.A., Chen, X., Sanders, M.E., Chakravarthy, A.B., Shyr, Y., Pietenpol, J.A., 2011. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J. Clin. Invest. 121. http://dx.doi.org/10.1172/JCI45014.

Lynch, A.G., Chin, S.-F., Dunning, M.J., Caldas, C., Tavaré, S., Curtis, C., 2012. Calling sample mix-ups in cancer population studies. PLoS ONE 7, e41815. http://dx.doi.org/10.1371/journal.pone.0041815.

McShane, L.M., Altman, D.G., Sauerbrei, W., Taube, S.E., Gion, M., Clark, G.M.Statistics Subcommittee of the NCI-EORTC Working Group on Cancer Diagnostics, 2005. REporting recommendations for tumour MARKer prognostic studies (REMARK). Br. J. Cancer. http://dx.doi.org/10.1038/sj.bjc.6602678.

Nilsen, G., Liestøl, K., Van Loo, P., Vollan, H.K.M., Eide, M.B., Rueda, O.M., Chin, S.-F., Russell, R., Baumbusch, L.O., Caldas, C., Børresen-Dale, A.-L., Lingjærde, O.C., 2012. Copynumber: efficient algorithms for single- and multi-track copy number segmentation. BMC Genomics. http://dx.doi.org/10.1186/1471-2164-13-591.

Ozols, R.F., 1997. Update of the NCCN ovarian cancer practice guidelines. Oncology (Williston Park, N.Y.) 11, 95–105.

Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F.L., Walker, M.G., Watson, D., Park, T., Hiller, W., Fisher, E.R., Wickerham, D.L., Bryant, J., Wolmark, N., 2004. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N. Engl. J. Med. 351, 2817–2826. http://dx.doi.org/10.1056/NEJMoa041588.

Pencina, M.J., D'Agostino, R.B., Steyerberg, E.W., 2011. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat. Med. 30, 11–21. http://dx.doi.org/10.1002/sim.4085.

Pencina, M.J., D'Agostino, R.B., Vasan, R.S., 2008. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat. Med. 27, 157–172. http://dx.doi.org/10.1002/sim.2929. Discussion 207–12.

Perou, C.M., Sørlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S.X., Lønning, P.E., Børresen-Dale, A.-L., Brown, P.O., Botstein, D., 2000. Molecular portraits of human breast tumours. Nature 406, 747–752. http://dx.doi.org/10.1038/35021093.

Ravdin, P.M., Siminoff, L.A., Davis, G.J., Mercer, M.B., Hewlett, J., Gerson, N., Parker, H.L., 2001. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. J. Clin. Oncol. 19, 980–991.

Russnes, H.G., Vollan, H.K.M., Lingjaerde, O.C., Krasnitz, A., Lundin, P., Naume, B., Sørlie, T., Borgen, E., Rye, I.H., Langerød, A., Chin, S.-F., Teschendorff, A.E., Stephens, P.J., Månér, S., Schlichting, E., Baumbusch, L.O., Kåresen, R., Stratton, M.R., Wigler, M., Caldas, C., Zetterberg, A., Hicks, J.B., Børresen-Dale, A.-L., 2010. Genomic architecture characterizes tumor progression paths and fate in breast cancer patients. Sci. Transl. Med. 2, 38ra47. http://dx.doi.org/10.1126/scitranslmed.3000611.

Rutgers, E., Piccart-Gebhart, M.J., Bogaerts, J., Delaloge, S., Veer, L.V., Rubio, I.T., Viale, G., Thompson, A.M., Passalacqua, R., Nitz, U., Vindevoghel, A., Pierga, J.-Y., Ravdin, P.M., Werutsky, G., Cardoso, F., 2011. The EORTC 10041/BIG 03-04 MINDACT trial is feasible: results of the pilot phase. Eur. J. Cancer 47, 2742–2749. http://dx.doi.org/10.1016/j.ejca.2011.09.016.

Seidman, J.D., Horkayne-Szakaly, I., Haiba, M., Boice, C.R., Kurman, R.J., Ronnett, B.M., 2004. The histologic type and stage distribution of ovarian carcinomas of surface epithelial origin. Int. J. Gynecol. Pathol. 23, 41–44. http://dx.doi.org/10.1097/01.pgp.0000101080.35393.16.

Sparano, J.A., 2006. TAILORx: trial assigning individualized options for treatment (Rx). Clin. Breast Cancer 7, 347–350. http://dx.doi.org/10.3816/CBC.2006.n.051.

Sparano, J.A., Paik, S., 2008. Development of the 21-gene assay and its application in clinical practice and clinical trials. J. Clin. Oncol. 26, 721–728. http://dx.doi.org/10.1200/JCO.2007.15.1068.

Sørlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Thorsen, T., Quist, H., Matese, J.C., Brown, P.O., Botstein, D., Eystein Lønning, P., Børresen-Dale, A.-L., 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci.

U.S.A. 98, 10869–10874. http://dx.doi.org/10.1073/pnas.191367098.

Sørlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J.S., Nobel, A.B., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C.M., Lønning, P.E., Brown, P.O., Børresen-Dale, A.-L., Botstein, D., 2003. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc. Natl. Acad. Sci. U.S.A. 100, 8418–8423. http://dx.doi.org/10.1073/pnas.0932692100.

van de Vijver, M.J., He, Y.D., van't Veer, L.J., Dai, H., Hart, A.A.M., Voskuil, D.W., Schreiber, G.J., Peterse, J.L., Roberts, C., Marton, M.J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., van der Velde, T., Bartelink, H., Rodenhuis, S., Rutgers, E.T., Friend, S.H., Bernards, R., 2002. A gene-expression signature as a predictor of survival in breast cancer. N. Engl. J. Med. 347, 1999–2009. http://dx.doi.org/10.1056/NEJMoa021967.

van't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A.M., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., Schreiber, G.J., Kerkhoven, R.M., Roberts, C., Linsley, P.S., Bernards, R., Friend, S.H., 2002. Gene expression profiling predicts clinical outcome of breast cancer. Nature 415, 530–536. http://dx.doi.org/10.1038/415530a.

Winter, W.E., Maxwell, G.L., Tian, C., Carlson, J.W., Ozols, R.F., Rose, P.G., Markman, M., Armstrong, D.K., Muggia, F., McGuire, W.P.Gynecologic Oncology Group Study, 2007. Prognostic factors for stage III epithelial ovarian cancer: a Gynecologic Oncology Group Study. J. Clin. Oncol. 25, 3621–3627. http://dx.doi.org/10.1200/JCO.2006.10.2517.

Winter, W.E., Maxwell, G.L., Tian, C., Sundborg, M.J., Rose, G.S., Rose, P.G., Rubin, S.C., Muggia, F., McGuire, W.P.Gynecologic Oncology Group, 2008. Tumor residual after surgical cytoreduction in prediction of clinical outcome in stage IV epithelial ovarian cancer: a Gynecologic Oncology Group Study. J. Clin. Oncol. 26, 83–89. http://dx.doi.org/10.1200/JCO.2007.13.1953.

Wishart, G.C., Azzato, E.M., Greenberg, D.C., Rashbass, J., Kearins, O., Lawrence, G., Caldas, C., Pharoah, P.D., 2010. PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. Breast Cancer Res. 12, R1. http://dx.doi.org/10.1186/bcr2464.

Wishart, G.C., Bajdik, C.D., Azzato, E.M., Dicks, E., Greenberg, D.C., Rashbass, J., Caldas, C., Pharoah, P.D., 2011. A population-based validation of the prognostic model PREDICT for early breast cancer. Eur. J. Surg. Oncol. 37, 411–417. http://dx.doi.org/10.1016/j.ejso.2011.02.001.

Wishart, G.C., Bajdik, C.D., Dicks, E., Provenzano, E., Schmidt, M.K., Sherman, M., Greenberg, D.C., Green, A.R., Gelmon, K.A., Kosma, V.-M., Olson, J.E., Beckmann, M.W., Winqvist, R., Cross, S.S., Severi, G., Huntsman, D.G., Pylkäs, K., Ellis, I., Nielsen, T.O., Giles, G., Blomqvist, C., Fasching, P.A., Couch, F.J., Rakha, E., Foulkes, W.D., Blows, F.M., Bégin, L.R., van't Veer, L.J., Southey, M., Nevanlinna, H., Mannermaa, A., Cox, A., Cheang, M., Baglietto, L., Caldas, C., Garcia-Closas, M., Pharoah, P.D., 2012. PREDICT Plus: development and validation of a prognostic model for early breast cancer that includes HER2. Br. J. Cancer 107, 800–807. http://dx.doi.org/10.1038/bjc.2012.338.