OXFORD

Briefings in Bioinformatics, 20(6), 2019, 2130–2140

doi: 10.1093/bib/bby073
Advance Access Publication Date: 3 September 2018
Review article

Breast cancer prognosis signature: linking risk stratification to disease subtypes

Fulong Yu*, Fei Quan*, Jinyuan Xu*, Yan Zhang*, Yi Xie, Jingyu Zhang, Yujia Lan, Huating Yuan, Hongyi Zhang, Shujun Cheng, Yun Xiao and Xia Li

Corresponding authors: Shujun Cheng, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China; State Key Laboratory of Molecular Oncology, Department of Etiology and Carcinogenesis, Cancer Institute and Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100021, China. Tel: 0451-86615922; Fax: 0451-86615922; E-mail: chengshj@263.net.cn; Xia Li, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang 150081, China. Tel: 0451-86615922; Fax: 0451-86615922; E-mail: lixia@hrbmu.edu.cn; Yun Xiao, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang 150081, China. Tel: 0451-86615922; Fax: 0451-86615922; E-mail: xiaoyun@ems.hrbmu.edu.cn
*These authors contributed equally to this work.

Abstract

Breast cancer is a very complex and heterogeneous disease with variable molecular mechanisms of carcinogenesis and clinical behaviors. The identification of prognostic risk factors may enable effective diagnosis and treatment of breast cancer. In particular, numerous gene-expression-based prognostic signatures were developed and some of them have already been applied into clinical trials and practice. In this study, we summarized several representative gene-expression-based signatures with significant prognostic value and separately assessed their ability of prognosis prediction in their originally targeted populations of breast cancer. Notably, many of the collected signatures were originally designed to predict the outcomes of estrogen receptor positive (ER+) patients or the whole breast cancer cohort; there are no typical signatures used for the prognostic prediction in a specific population of patients with the intrinsic subtype. We thus attempted to identify subtype-specific prognostic signatures via a computational framework for analyzing multi-omics profiles and patient survival. For both the discovery and an independent data set, we confirmed that subtype-specific signature is a strong and significant independent prognostic factor in the corresponding cohort. These results indicate that the subtype-specific prognostic signature has a much higher resolution in the risk stratification, which may lead to improved therapies and precision medicine for patients with breast cancer.

Key words: breast cancer; prognosis signature; subtype; integrated analysis

Fulong Yu is a PhD student in College of Bioinformatics Science and Technology at Harbin Medical University.
Quan Fei is an undergraduate student in College of Bioinformatics Science and Technology at Harbin Medical University.
Jinyuan Xu is a PhD student in College of Bioinformatics Science and Technology at Harbin Medical University.
Yan Zhang is a PhD student in College of Bioinformatics Science and Technology at Harbin Medical University.
Yi Xie is an undergraduate student in College of Bioinformatics Science and Technology at Harbin Medical University.
Jingyu Zhang is an undergraduate student in College of Bioinformatics Science and Technology at Harbin Medical University.
Yujia Lan is a PhD student in College of Bioinformatics Science and Technology at Harbin Medical University.
Huating Yuan is a postgraduate student in College of Bioinformatics Science and Technology at Harbin Medical University.
Hongyi Zhang is a PhD student in College of Bioinformatics Science and Technology at Harbin Medical University.
Shujun Cheng is a professor for College of Bioinformatics Science and Technology at Harbin Medical University.
Yun Xiao is a professor for College of Bioinformatics Science and Technology at Harbin Medical University.
Xia Li is a professor and head of the Chair for College of Bioinformatics Science and Technology at Harbin Medical University.
Submitted: 26 November 2017; **Received (in revised form):** 14 July 2018

Introduction

Breast cancer is one of the most frequently diagnosed cancers in women worldwide [1]. The average 5-year survival rate among people with breast cancer is 90% [2]. The significant survival improvement achieved in the past decades greatly benefits from the advances in early diagnosis and appropriate therapy [3]. Although clinicopathological factors, such as tumor size, axillary lymph-node status and histological grade, have been used most frequently and are still important for diagnosis and prognosis evaluation of breast cancer patients, their use alone was insufficient for clinical guidance in the choice of therapeutic strategy [4]. A major step has been made toward the development and testing of relevant molecular prognostic and/or predictive signatures, which could be used to complement clinicopathological factors. In particular, gene-expression-profiling studies of primary breast tumors produce numerous distinct prognostic gene sets [5, 6]. These multigene prognostic signatures usually discriminate patients with good and poor prognosis, which allows more effective decision-making in clinical trials and treatments.

Given that the current expression-based prognosis studies differ considerably with respect to their methodology and analytical and clinical validations, we thus attempt to summarize a number of highly reliable signatures. We selected 24 highly reliable prognostic signatures using a set of rigorous screen filters (Tables 1 and 2 and Supplementary Table S1, see Materials and Methods). Overall, we found that 14 of 24 signatures were subtype-specific signatures, which include Mammaprint (ER+, LN-), Oncotype DX (ER+, LN-), Endopredict (ER+, HER2-), GGI-97 (ER+), Rotterdam signatures (LN-), MS-14 (ER+, LN-), Mammostrat (ER+), 14-gene Yau (ER-), TNBC-related metagenes (ER-, PR-, HER2-), 5-gene Yau (ER-), 95-GC (ER+, LN-), EMT-related gene signature (LN-), RR-related signature (ER+) and 95-gene Naoi (ER+, LN-). Most of these subtype-specific signatures aimed at ER+ breast cancer patients.

1st- and 2nd-generation prognostic signatures

We collected six classic prognostic gene-expression-based signatures from the so-called 1st- and 2nd-generation signatures [7]. All of them were marketed, and two of those were approved by US Food and Drug Administration (FDA). Despite differences in the genes that compose each of the signatures, 1st-generation prognostic signatures are usually more accurate to predict recurrence risk within the 1st 5 years and associated with proliferation-related genes. For comparison, the 2nd-generation signatures focus more on the late recurrences, while some of them were also associated with benefit from adjuvant chemotherapy. We highlighted three most widely adopted platforms (MammaPrint [8], Oncotype DX [9] and Endopredict [10]). For more detailed information, please refer to references [11, 12].

MammaPrint (70-gene assay) is the 1st successful prognostic signature that was marketed by Agendia (the Netherlands). The signature was developed from microarray analysis of 78 young (<55 years) ER+ and LN- breast cancer patients. This microarray test outperformed than the traditional clinicopathological factors in an independent cohort of 295 invasive breast cancers and then received approval as a diagnostic assay by the FDA in February 2007 [8, 13].

Oncotype DX is a 21-gene prognostic predictor developed by using quantitative real-time polymerase chain reaction (qRT-PCR)-based expression profles. Recurrence score was calculated based on this signature's prognosis for the risk of distant relapse at 10 years for patients with ER+ and LN−. A number of retrospective clinical studies have also demonstrated the predictive value of this signature for distant recurrence risk, overall survival (OS) and response to adjuvant chemotherapy in early breast cancer [9, 14, 15].

Endopredict (11-gene assay) is another marketed prognostic test, which predicts risk of distant recurrence in patients with ER+ and HER2-. Many subsequent studies showed that this signature could add prognostic value to classic clinicopathological variables and associated with benefit from adjuvant chemotherapy [16, 17].

Biological pathway-based prognostic signatures

We also assigned the other collected prognostic signatures to several relevant biological pathways. These pathways include estrogen receptor, proliferation and metastasis, immune function, cell cycle and metabolism process. The according studies all shown good performance in classifying breast cancer and estimating clinical outcomes. Detailed descriptions can be seen in Table 2 and Supplementary Table S1.

Breast cancer shows a high inter-individual heterogeneity at clinical and molecular levels. This heterogeneity poses an acute challenge for not only the accurate diagnosis and suitable treatment of patients but also in-depth understanding of the underlying tumor biology. Stratification of breast tumors into distinct subtypes has been shown an effective strategy to overcome the heterogeneity within these populations [18]. The currently subtyping of breast cancer is mainly based on the molecular, histopathological and clinical levels with different therapeutic implications. Clinicopathological criteria are a traditional strategy for breast cancer classification and assessment. Progressively, new classifications such as PAM50 method [19]

Table 1. An overview of the 1st- and 2nd-generation prognostic signatures

Signature	Year	Genes	Approval	PMID	HR	P
1st generation						
Mammaprint	2002	70	CE, FDA 2007	11823860	4.6	<.001
Oncotype DX	2004	21	CE	15591335	3.21	<.001
Rotterdam signatures	2005	76	CE	15721472	5.67	<.0001
GGI-97	2006	97	CE	16478745	3.61	<.001
2nd generation						
PAM50	2009	50	CE, FDA 2013	19204204	NA	<.001
Endopredict	2011	11	CE	21807638	1.28	<.001

Abbreviations: FDA, US Food and Drug Administration; CE, European community marking; HR, hazard ratio from original literature; *P*, P-value of Cox analysis from original literature; NA, not available, PMID, Pubmed ID.

Table 2. An overview of biological pathway-based prognostic signatures

Signature	Year	Genes	Pathway	PMID	HR	P
IGS	2007	186	PM	17229949	1.4	<.001
MS-14	2008	14	ER	19025599	4.02	<.05
3D-signature	2008	22	ER	18714348	5.5	<.0001
14-gene Yau.	2010	14	IF	20946665	3.93	<.00001
Mammostrat	2010	5	IF	20615243	1.62	<.00001
TNBC-related metagenes	2011	16	MP	22220191	4.03	<.001
95-gene Naoi.	2011	95	PM	20803240	NA	<.001
5-gene Ascierto ML.	2012	5	IF	21479927	NA	<.001
5-gene Yau.	2013	5	PM	24172169	1.5	<.05
95-GC	2014	72	ER	24461457	1.95	<.001
EMT-related gene signature	2014	51	PM	25060555	2.61	<.0001
CTC	2015	6	PM	25529931	NA	<.001
RR-related signature	2015	18	ER	26527319	3.99	<.001
M-Sig	2015	146	PM	25974184	>2	<.001
IgSF genes	2017	10	IF	27911271	2.78	<.001
GC-18	2017	18	PM	28886126	5.1	<.001
12-gene signature	2017	12	CC	28122328	3.95	<.0001
23-gene Li.	2017	23	PM	28529601	2.1	<.01

Abbreviations: FDA, US Food and Drug Administration; CE, European community marking; HR, hazard ratio; P, P-value from Cox analysis; NA, not available; ER, estrogen receptor; PM, proliferation and metastasis; IF, immune function; CC, cell cycle; MP, metabolism process, PMID, Pubmed ID.

have emerged to define intrinsic breast cancer subtypes, which is informative for prognosis and responsiveness to various therapies. The intrinsic molecular subtypes are gradually becoming part of the lexicon of breast cancer researchers, oncologists and pathologists [20–22]. Identification of molecular subtypes of breast cancer opened new perspectives for personalized diagnosis and therapy. Yet, a number of studies have pointed that even in the patient group with particular subtype, clinic features and patient's outcome could be inherently different [23–25].

A substantial proportion of breast cancer patients received under- or over-treatment because of insufficiently accurate prognosis predictions [26, 27]. Although the current prognosis signatures are usually well tested in the respective studies, their original designs were for the complete population of breast cancer or only a specific patient cohort, such as ER+ patients. Given the complex and heterogeneity within breast cancer, it is not clear how the potential applicability of these representative signatures in the well-classified individual subgroups, e.g. tumor samples stratified based on molecular profiling. Using gene expression profiles of breast cancer patients from The Cancer Genome Atlas (TCGA) [4], we evaluated prognostic ability of six expression-based signatures on individual intrinsic subgroups and the whole cohort. We noted that most of selected prognostic signatures were generally not suitable to evaluate the prognosis of intrinsic subgroups despite that they generally shown significant prognostic power within their originally target patient population. To create an effective predictive model aiming at a particular group with distinct and intrinsic tumor characteristics, we integrated gene expression with DNA methylation, somatic copy number variation and mutation data to develop a computational pipeline for group-specific prognostic signatures discovery.

Materials and Methods

Prognosis signatures collection

A total of 24 highly reliable breast cancer prognostic signatures were obtained. We first performed a systematic review of the literature between 2007 and 2017 by searching for the keywords 'breast cancer', 'gene expression', 'diagnosis', 'signature' and 'prognosis' to clarify the present state of knowledge regarding gene-expression-based prognostic signatures. The choice of the signatures used in this study was based on several criteria. (i) Only gene-expression-based signature associated with patients' outcome were included. (ii) We tried to select signatures with high known prognostic value in terms of OS, disease-free survival, distant metastasis-free survival and disease relapse-free survival. (iii) We tried to select signatures derived from the study of a large sample group (e.g. sample size, >500) with a long-term follow-up time (e.g. median follow-up time is 5 years) ensuring the adequate statistical power. (iv) Signatures with insufficient descriptions of diagnostic and prognostic performance were excluded. In addition, we also collected several signatures that did not satisfy the filtering criteria but have been validated by clinical trials.

A total of 24 signatures passed our criteria. Because of inadequate information, eight representative signatures were selected, including five subtype-specific signatures and three non-subtype-specific signatures (Supplementary Table S2). These five subtype-specific signatures contain the 1st- (Rotterdam signatures, Oncotype DX) and 2nd- (Endopredict) generation prognostic signatures, as well as pathway-based signature (MS-14, Mammostrat). The three non-subtype-specific signatures contain IgSF genes, 12-gene signature and 18-GC. Among these eight signatures, 'Oncotype DX' signature was generated from a 250-candidate genes panel through computing integrative risk scores of patients (e.g. ER group score and proliferation group score) and the 18-GC signature that was developed by generating classification trees and optimized by using Bayesian statistical method. The remaining six signatures (including Endopredict, Rotterdam signatures, Mammostrat, 12-gene signature, MS-14 and IgSF signature) were developed generally through building a Cox proportional hazard regression model with slight differences during the gene selection process. 'Endopredict' selected genes by the rank order of Cox P-values. 'Rotterdam signatures' kept the robustness of gene selection by a bootstrapping of patients in the training set. 'Mammostrat' iteratively pruned candidate genes obtained from a univariate Cox model, until the removal caused a significant reduction in the fit of the model. Following a meta-analysis and genes filtration, '12-gene signature' was developed by Cox regression analysis. 'MS-14' employed a semi-supervised principal component method to optimize the Cox scores and Least Absolute Shrinkage and Selection Operator (LASSO) regression to select 14 genes. After the Cox regression analysis, prognostic genes of IgSF signature were identified by integrating protein–protein interaction network and immunoglobulin superfamily genes.

Data sets

The mRNA gene expression, DNA methylation, segmented copy number profiles and clinical data of primary breast cancer patients were downloaded from the TCGA data portal. The TCGA gene expression profiles were treated as the discovery data set including 813 tumors, which was composed of 136 (16.7%) basal-like, 90 (11.1%) HER2-enriched, 411 (50.6%) Luminal A and 176 (21.6%) Luminal B samples. In order to ensure no systematic biases in the multi-omics analysis, the DNA methylation and copy number profiles with matched gene expression profiles were retained.

Gene expression data generated with Affymetrix U133A arrays for 454 breast cancer patients (GSE25066) was downloaded from GEO database as the independent validation data set. This data set includes 163 (35.9%) basal-like, 76 (16.7%) HER2-enriched, 149 (32.8%) Luminal A and 66 (14.6%) Luminal B samples. As previously described [28], gene expressions were called from the raw Affymetrix.cel files with RMA and ComBat normlization methods [29].

All the intrinsic subtype classifiers of patients were based on PAM50 annotation. We focused on four major breast cancer subtypes, namely, basal-like, HER2-enriched, Luminal A and Luminal B, where the Normal-like samples were assigned into HER2-enriched group.

Identification of group-specific prognostic signatures

Candidate prognostic genes

The normalized Reads Per Kilobase of transcript, per Million mapped reads (RPKM) values of TCGA profiles were used as gene expression levels. For patients from the whole cohort or a particular subtype, gene expression levels more than 1 in at least 50% samples were kept for further analysis. A candidate prognostic gene was determined if its expression was associated with OS. The candidate prognostic genes should satisfy the following two criteria. First, the gene significantly affected survival in the univariate analysis of Cox proportional hazard model ($P < .05$). Second, there was a significant difference between the survival times of two groups divided according to the median of this gene expression ($P < .05$).

Filtering for the multi-omics data analysis

Aberrant methylation of gene promoter is an alternative way to inactivate tumor oncogenes or suppressor genes in cancer [30]. DNA methylation level of gene promoter may be a good indication for understanding molecular mechanism of prognosis gene. CpG methylation was measured by the Illumina Infinium 450K platform. For simplicity, the CpG sites occurred in the promoter region of candidate prognosis gene were retained. In order to search for gene mostly influenced by DNA methylation, Pearson correlation coefficients (PCC) between mRNA expression of candidate prognostic gene and methylation levels of its promoter CpG sites were calculated. We defined 'Epigenetic regulation' gene whose expression was significantly associated with at least one CpG methylation level ($PCC < 0$, $P < .05$).

Somatic copy number alterations (SCNAs) commonly occurred in breast cancer. Subtype-specific frequent alteration region may represent specific driver event explaining the intrinsic fundamental pathological mechanisms. We performed group-specific 'driver' events discovery using Genomic Identification of Significant Targets in Cancer 2.0 [31] for the segmented copy number data of the particular cohort. The significantly altered regions were identified at q value < 0.25. The 'SCNA' gene set was defined as the candidate prognosis genes falling within the group-specific significantly altered regions.

Somatic mutation files were downloaded from The Catalogue Of Somatic Mutations In Cancer database [32]. 'Point mutation' gene set was defined as candidate prognosis genes whose coding sequence contain any validated somatic mutation.

The risk gene set was identified by the candidate prognostic genes satisfied any filtering of the omics data analysis, that is, a union gene set composed of 'Epigenetic regulation', 'SCNA' and 'Point mutation' gene sets.

Variable selection based on random survival forest

Random survival forest is an important extension of random forests and a commonly used method for variables selection in high-dimensional survival settings [33]. For each subtype, we obtained a risk gene set and then the corresponding gene expression profiles and clinical data (including survival time and events) were used for training the random survival forest model. The response variables are mainly based on the survival time and event (alive or death) of patients and the predictor variables are these expression levels of the risk genes. During the construction of random survival forest model, tree node splits according to maximizing survival differences between child nodes. In each tree, survival time and status of the patients were considered as response variables. The random survival forest analyses provide measures including variable importance and minimal depth for each variable. A larger variable importance and a smaller minimal depth mean that the variable is more predictive in the survival model. To refine the risk gene set, we fit a random survival forest model using 1000 trees for the according sample group. Random survival forest analyses were carried out using R package 'randomForestSRC'. We built the model using the function *rfsrc()* and selected genes using function *var.select()* with the most conservative threshold.

The group-specific risk score was calculated for each patient as a linear combination of the according signature genes weighted by the Cox regression coefficients, which were obtained from univariate analysis of candidate prognostic genes. We used the median of the risk scores as the threshold to discretize scores into high- and low-risk groups, as above- and below-median scores in the cohort were associated with poor and favorable survival.

Statistical analysis

All statistical analysis were carried out using R 3.3.1 with the package 'survival' [34] and package 'randomForestSRC' [33]. Univariate Cox proportional hazard models were fit to identify factors significantly related to OS. Survival curves were constructed using the Kaplan–Meier method [35] and the log-rank test [36] was used for comparison between groups. In multivariate

Table 3. Evaluation of non-subtype-specific signatures within the whole cohort

Signature	Low-/high-risk group (n)	HR (95% CI)	Significance in the corresponding cohort
IgSF genes	407/406	1.81 (1.16, 2.81)	**0.0079**
18-GC	407/406	1.29 (0.84, 1.99)	0.23
12-gene	407/406	1.53 (0.99, 2.36)	**0.049**

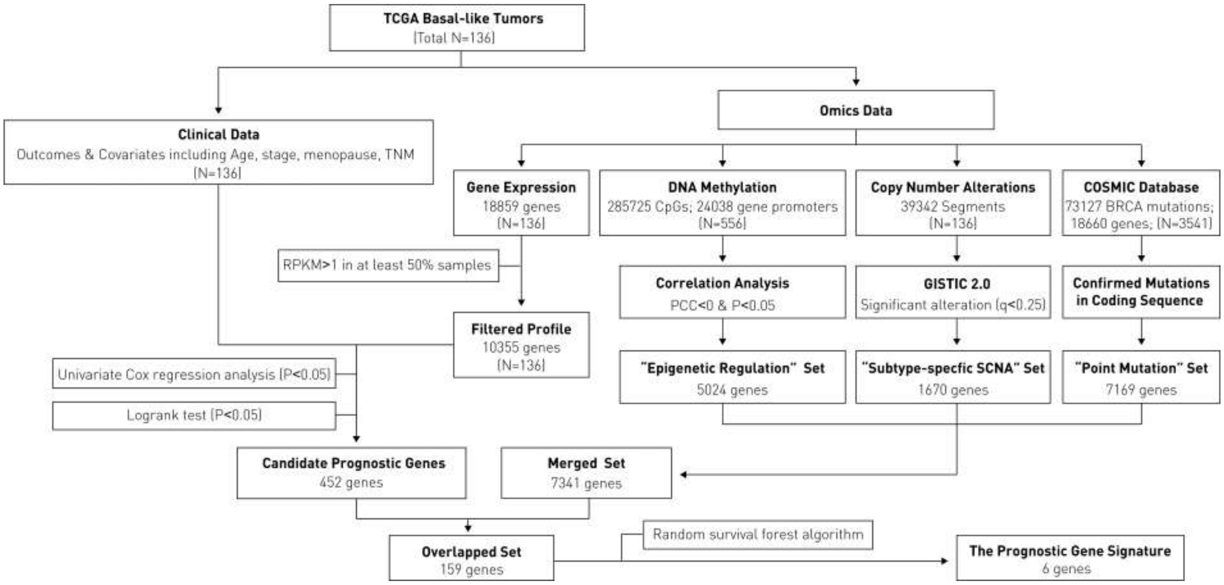

Figure 1. Work chart of identification of subtype-specific prognostic signature.

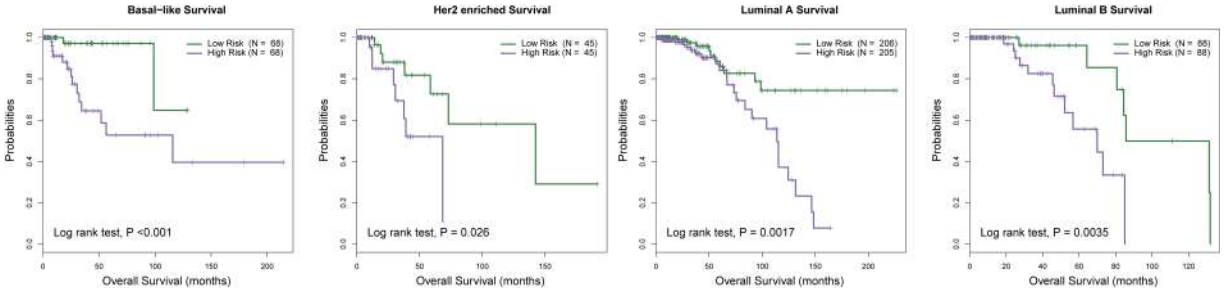

Figure 2. KM plots for subtype-specific signatures in the according subgroup from the training set.

models, we estimated the adjusted hazard ratios (HRs) of breast cancer subgroups with standard clinicopathologic variables: age at diagnosis, histologic grade, menopausal status at diagnosis, tumor size (T stage), lymph node invasion (N stage) and metastatic spread (M stage). Samples with missing values were excluded from the analysis. In order to assess the prediction error rate, we calculated C-index, a commonly used measure to quantify the discriminatory power of a predictive model, for prognosis models using the R package 'survcomp' [37]. Wald's test was used to evaluate the significance of HRs. All tests were two-sided and $P < .05$ was considered statistically significant.

Results

Assessment of the prognostic power of expression-based signatures in breast cancer subtypes

The tests for signature development are generally different in the patient cohorts and experimental and computational methods of analysis. A previous study has pointed that the significant agreement shown the outcome predictions using the current gene-expression-based prognostic signature for individual patients [38]. Given that the significant complexity and heterogeneity within the tumor, however, an important and unanswered question is what is the performance of these predictors across distinct subtypes of breast tumors. We thus attempted to evaluate the prognostic value for the selected, highly reliable signatures within individual breast cancer subtypes.

We first selected three representative non-subtype-specific signatures, which were designed for the whole breast cancer cohort, to access their prognostic value within individual breast cancer intrinsic subtypes (see materials and methods). The TCGA breast cancer patients were used as the evaluation cohort and the according risk scores were calculated (Supplementary Table S2). In concordance with the previous studies, we found that these signatures generally remained to be good predictors of risk stratification in the whole breast cancer patient cohort (Table 3).

Table 4. Univariable and multivariable Cox regression analysis of the subtype-specific prognosis signatures in the training set

		Univariable analysis		**Multivariable analysis**	
		HR (95% CI)	***P***	**HR (95% CI)**	***P***
Basal-like (n = 136)					
Age	<=50/>50	1.46 (0.53, 4.04)	.4705		
Disease stage	<=II/>II	3.96 (1.46, 10.8)	**.007**		
Premenopausal	Yes/no	0.96 (0.25, 3.61)	.94		
T stage	<=T2/>T2	1.30 (0.40, 4.26)	.66		
Lymph node involvement	<=N1/>N1	8.43 (3.00,23.6)	<**.001**	26.0 (3.27,206)	**.0021**
M stage	M0/M1	1.74 (0.39, 7.79)	.47		
Signature	Low risk/high risk	8.28 (1.87,36.6)	**.0053**	17.5 (2.19,139)	**.0070**
HER2 enriched (n = 90)					
Age	<=50/>50	2.06 (0.64, 6.66)	.23		
Disease stage	<=II/>II	7.45 (2.19, 25.4)	**.0013**	35.3 (3.58, 348)	**.0022**
Premenopausal	Yes/no	NA	NA		
T stage	<=T2/>T2	3.74 (1.31, 10.6)	**.013**		
Lymph node involvement	<=N1/>N1	3.76 (1.25, 11.3)	**.018**		
M stage	M0/M1	1.33 (0.29, 6.00)	.71		
Signature	Low risk/high risk	3.45 (1.08, 11.0)	**.036**	6.29 (1.48, 26.7)	**.012**
Luminal A (n = 411)					
Age	<=50/>50	1.32 (0.61, 2.83)	.48		
Disease stage	<=II/>II	1.40 (0.69, 2.83)	.35		
Premenopausal	Yes/no	0.97 (0.35, 2.68)	.96		
T stage	<=T2/>T2	0.60 (0.25, 1.45)	.25	0.18 (0.04, 0.78)	**.022**
Lymph node involvement	<=N1/>N1	1.35 (0.56, 3.28)	.50		
M stage	M0/M1	2.70 (1.08, 6.74)	**.033**		
Signature	Low risk/high risk	3.13 (1.48, 6.60)	**.0028**	4.13 (1.50, 11.4)	**.0061**
Luminal B (n = 176)					
Age	<=50/>50	1.06 (0.37, 3.03)	.91		
Disease stage	<=II/>II	3.80 (1.34,10.8)	**.012**	11.5 (1.28, 103)	**.029**
Premenopausal	Yes/no	3.13 (0.39, 24.8)	.28		
T stage	<=T2/>T2	1.87 (0.68, 5.14)	.22		
Lymph node involvement	<=N1/>N1	2.32 (0.90,5.97)	.080		
M stage	M0/M1	1.76 (0.60,5.22)	.31		
Signature	Low risk/high risk	5.04 (1.55, 16.4)	**.0073**	8.25 (1.97, 34.4)	**.0038**

Abbreviations: HR, hazard ratio; CI, confidence interval; significant *P* values (<0.05) are shown in bold.

In addition to non-subtype-specific signatures, 14 of our collected signatures were subtype-specific. Notably, we found that the subtype-specific signatures were almost all developed using technologies such as RT-PCR and/or microarray (Supplementary Table S1) and most of the subtype-specific signatures using a fixed cut-off rather than the median of risk scores to define risk groups. Thus, it is very difficult to apply these subtype-specific signatures to the data from RNA-seq that is a widely used technology, because the original thresholds of signatures could not be directly used for the risk stratification of patients detected by different technologies. A comprehensive method correcting the difference between RT-PCR, microarray and RNA-seq is needed for effective evaluation of subtype-specific signatures and application to sequencing-based transcriptome data.

Identification of subtype-specific prognostic signatures using diverse molecular data

To investigate molecular prognostic signature within individual intrinsic subgroups, we developed an approach that integrated clinical and multiple-omics data (including transcriptome, epigenome and genome data) of TCGA breast cancer tumors to identify group-specific expression-based signatures correlated with survival of patient belonging to individual intrinsic subtype. Taking the basal-like subtype as an example (Figure 1), we selected the group-specific signature containing genes that not only were significantly associated with patients' survival but also could lead us to better understand the complex biology of breast cancer. A detailed description of our computational workflow can be seen in materials and methods.

For each intrinsic subgroup, we used the according subtype-specific signature to divide patients of the training set into a high-risk group or a low-risk group (see materials and methods, Supplementary Table S3–S8). Overall, the patients within the low-risk group showed significantly longer OS than those within the high-risk group in all of the four subtypes (Figure 2): basal-like group (median survival not reached versus 116 months; P = .0016), HER2-enriched group (median survival, 142.4 months versus 68.4 months; *P* = .027), Luminal A group (median survival, 147 months versus 115 months; P = .0017) and Luminal B group (median survival, 85.8 months versus 69.9 months, P = .0035). We found the HRs for the four subgroups were all greater than 3, with the range from 3.13 to 8.28 (Table 4).

To determine whether the survival prediction performance of the subtype-specific signatures is independent of clinical and pathological factors of patients within each individual subgroup, we performed multivariable Cox regression analysis. We observed that all of the signatures retained strong and significant independent prognostic factors in the four cohorts (Table 4). We observed that the subtype-specific signature was the only significant risk factor in the multivariate Cox model of the Basal-like cohort. For the other three cohorts, except

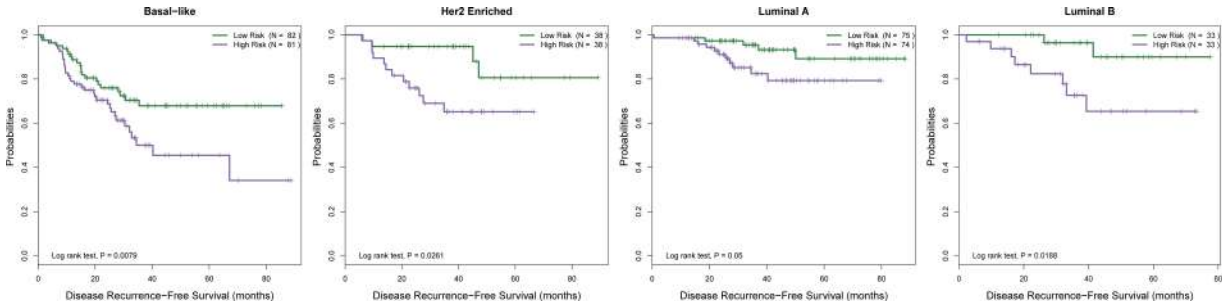

Figure 3. KM plots for subtype-specific signatures in the according subgroup from the validation set.

Table 5. Univariable and multivariable Cox regression analysis of the subtype-specific prognosis signatures in the validation set

		Univariable analysis		Multivariable analysis	
		HR (95% CI)	*P*	HR (95% CI)	*P*
Basal-like (n = 163)					
Age	<=50/>50	1.54 (0.91, 2.60)	.11		
Grade	1/>1	NA	NA		
T stage	<=T2/>T2	2.32 (1.35,3.97)	**.0022**		
Lymph nodal status	<=N1/>N1	1.70 (0.97,2.97)	.062		
Signature	Low risk/high risk	2.07 (1.20,3.57)	**.0093**	2.24 (1.24,4.05)	**.0076**
HER2 enriched (n = 76)					
Age	<=50/>50	2.79 (0.97, 8.05)	.058	4.31 (1.14, 16.3)	**.031**
Grade	1/>1	2.08 (0.27, 15.9)	.48		
T stage	<=T2/>T2	1.02 (0.38, 2.75)	.97		
Lymph nodal status	<=N1/>N1	4.23 (1.57, 11.4)	**.0044**	5.58 (1.72, 18.0)	**.0041**
Signature	Low risk/high risk	3.36 (1.08, 10.4)	**.036**	4.81 (1.27, 18.2)	**.021**
Luminal A (n = 149)					
Age	<=50/>50	0.36 (0.11, 1.13)	.079		
Grade	1/>1	NA	NA		
T stage	<=T2/>T2	2.78 (1.03, 7.47)	**.043**	2.82 (1.01, 7.85)	**.047**
Lymph nodal status	<=N1/>N1	3.61 (1.31, 9.97)	**.013**	4.79 (1.62, 14.1)	**.0046**
Signature	Low risk/high risk	2.77 (0.96, 7.99)	.060	3.47 (1.15, 10.5)	**.027**
Luminal B (n = 66)					
Age	<=50/>50	0.58 (0.16, 2.04)	.39		
Grade	1/>1	NA	NA		
T stage	<=T2/>T2	0.97 (0.27, 3.44)	.96		
Lymph nodal status	<=N1/>N1	0.60 (0.08, 4.78)	.63		
Signature	Low risk/high risk	5.29 (1.12, 25.0)	**.036**	5.89 (1.19, 29.1)	**.029**

Abbreviations: HR, hazard ratio; CI, confidence interval; significant *P* values (<0.05) are shown in bold; NA, not available because of too much missing data.

for subtype-specific signatures, only one clinical variable was the significant risk factor in the multivariate Cox models, respectively. In addition, we found that our predictive models showed substantial predictive power, with C-indexes ranged from 0.76 to 0.88 (Supplementary Table S9).

Subtype-specific prognostic signature presents prognosis power within an independent breast cancer cohort

To determine whether our subtype-specific molecular signatures derived from the TCGA data are applicable to a completely independent set of samples, the expression profiles of a cohort of 454 patients were collected. For each subtype, the patients of the independent cohort were classified as high-risk or low-risk groups according to the risk scores calculated using subtype-specific prognostic signatures. Notably, we found each of the four group-specific signature shows a good prognosis power (Figure 3, Table 5). The medians of HR for the four subgroups were all greater than 2, with the range from 2.07 to 5.29. Moreover, we found these models shown good prediction accuracy with C-indexes ranged from 0.66 to 0.79 (Supplementary Table S3). These data demonstrated that each of the prognostic signatures was significantly predictive for outcome in the corresponding breast cancer subtype.

The prognostic power of the 'whole cohort signature' is limited within subgroup

We also identified the 'whole cohort signature' employing our group-specific prognostic approach on the entire training data sets (Supplementary Table S7). Noticeably, the 'whole cohort signature' also shown a good risk stratification in both the training and validation data (Figure 4, Table 6). The association of the 'whole cohort signature' with survival was further investigated in a multivariable Cox analysis including the same covariates considered before. The 'whole cohort signature' retained its significant and favorable prognostic role on the patients' survival. Interestingly, although the 'whole cohort signature' could be a strong and independent risk factor for the data set which is not a distinguished subtype, we found its classification efficiency was substantially decreased when it is applied on the individual

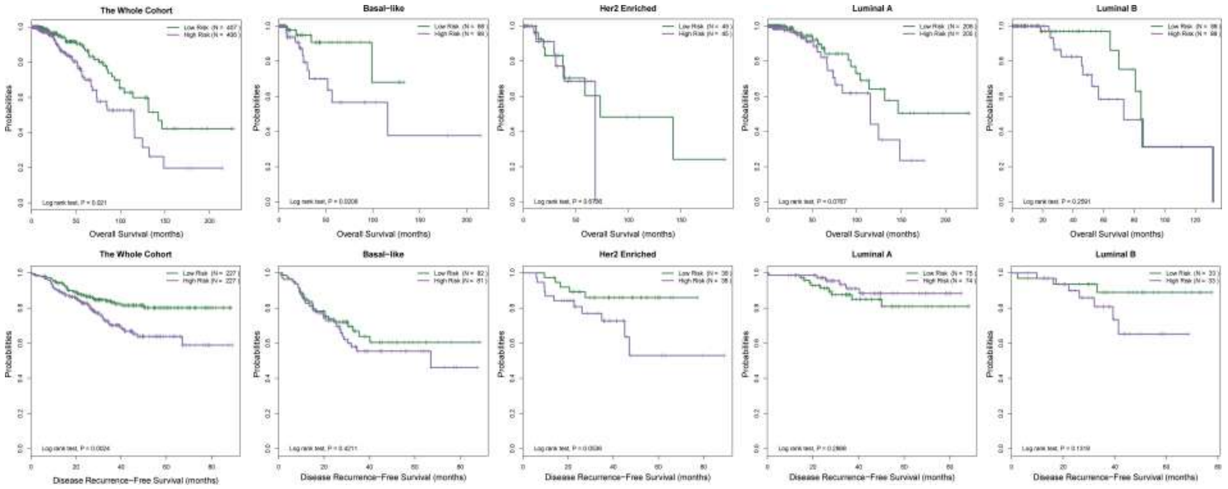

Figure 4. Evaluation of the whole cohort signature.

Table 6. Univariable and multivariable Cox regression analysis of the 'whole cohort signature' in the training and validation set

		Univariable analysis		Multivariable analysis	
		HR (95% CI)	*P*	HR (95% CI)	*P*
Training set (n = 813)					
Age	<=50/>50	1.31 (0.81, 2.10)	.27		
Disease stage	<=II/>II	2.56 (1.67,3.93)	<**.001**	2.69 (1.32, 5.48)	**.0061**
Premenopausal	Yes/No	1.52 (0.74, 3.15)	.26		
T stage	<=T2/>T2	1.17 (0.72, 1.90)	.54		
Lymph node involvement	<=N1/>N1	3.06 (1.95,4.82)	<.001		
M stage	M0/M1	2.15 (1.22,3.79)	**.0079**		
Signature	Low risk/high risk	2.04 (1.31, 3.17)	**.0015**	2.30 (1.46, 3.63)	<.001
Validation set (n = 454)					
Age	<=50/>50	1.14 (0.77, 1.70)	**.50**		
Grade	1/>1	8.26 (1.15, 59.3)	**.036**		
T stage	<=T2/>T2	2.02 (1.36, 3.01)	<**.001**	1.81 (1.21, 2.71)	**.0041**
Lymph nodal status	<=N1/>N1	2.42 (1.60, 3.67)	<**.001**	2.12 (1.39, 3.23)	<**.001**
Signature	Low risk/high risk	1.87 (1.24, 2.81)	**.0028**	1.74 (1.15, 2.63)	**.0087**

Abbreviations: HR, hazard ratio; CI, confidence interval; significant *P* values (<0.05) are shown in bold.

intrinsic subgroups of either the training or validation data. Specifically, the 'whole cohort signature' is significant only in the basal-like cohort from the training set and did not reach the significant threshold in all the four subgroups from the validation set. These results further demonstrated the validity of our signature discovery method.

In addition, we also performed patient survival prediction using cross-validated models, e.g. assessing Basal-like-specific signature separately in the other three individual intrinsic subgroups. We repeated the evaluation separately for the four subtype-dependent signatures. Unfortunately, all of the signatures showed little predictive power across the inter-subtype groups in the both data sets (Supplementary Table S10). Combined with the previous results for evaluation of subtype-specific signatures, we should not only consider the underlying biological variability between subclasses of breast cancer but also better understand and predict the prognosis within the subtype-specific context.

Discussion

One of the most attractive features of molecular biomarker for clinical applications is accurate prognosis for patients with malignant disease, which helps stratify patients into different risk groups and choose the most effective treatment. In this study, we present a brief overview of representative, gene-expression-based prognostic signatures of breast cancer according to various rigorous filtering criteria. We found several well-established prognostic signatures, carrying similar information of prognostication, could barely predict the survival in the intrinsic subgroup of patients with breast cancer. Although several recent studies highlighted the potential limitations of prognostic prediction using whole patient data sets without considering the effect of tumor heterogeneity, the development of subtype-specific prognostic signature need to be paid more attention [39–41]. This study introduces a method of prognostic biomarker discovery with a subtype-specific manner, which simultaneously identified prognostic signatures for each of the four intrinsic subtypes in the TCGA data set. Through the Cox proportional hazards analyses, each of the signatures was confirmed the validity in the corresponding subgroups from a single independent data set.

A large number of prognostic and predictive signatures have been proposed for breast cancer. Among them, the 1st- and 2nd-generation prognostic signatures have been marketed, which were generally developed by evaluating the difference in disease recurrence for ER+ patients. Notably, some signatures were even trained and tested using the complete cohort without considering the internal heterogeneity of breast cancers. These signatures inevitably classify the high grade and ER-negative patients to the high-risk group, regardless of a large majority of these patients have good prognosis [42, 43]. Therefore, it is not a surprise that the limited ability of prognosis prediction on all of each subgroup was observed. Despite there are no clinically useful prognostic signatures for ER-negative cancers, the basal-like-specific signatures may provide a potential solution because of nearly 70% of the basal-like subtype were ER-negative. To the best of our knowledge, our study is a 1st attempt to investigate subtype-specific signatures simultaneously in four major subgroups determined by the PAM50 subtyping. It may be useful for the prognostic study in the future to recognize significant prognostic biomarkers though analyzing the individual breast cancer subgroups defined by other relevant strategies. In addition, we also found the subtypes defined by the union of subtype-specific signature genes associated with the intrinsic subtypes (Supplementary material).

With the advantage of multidimensional genomic studies, we could know the genomic and epigenetic abnormalities that occurred in cancer cells and their downstream effects on gene expression. Nowadays, it's not uncommon to identify prognosis risk factors via an integrated genomic approach. For example, Auwera *et al.* [44] present an integrated transcriptome analysis of breast cancer to identify a prognostic signature composed of several mRNAs and microRNAs. By integrating mRNA, microRNA and DNA methylation next-generation sequencing data, Volinia and his colleagues [45] performed survival analysis on patients with breast cancer to identify an integrated prognostic signature. In this study, we also proposed an integrating multi-omics method to investigate subtype-specific signatures. In comparison, our prognostic signatures used only mRNA gene expression information for prognosis evaluation, which could be applied more broadly because of the plethora of transcriptome profiling of cancer patients. Furthermore, the signatures identified using this approach on the patients from the subgroup or the entire cohort both have significant prognostic value in the corresponding data sets. We anticipate our computational framework would have more widespread application value in the other cancer prognosis studies.

Using subtype-specific signatures, we observed a difference in prognosis for each of four intrinsic subtypes on both the TCGA and independent validation data sets. Each of the signatures encompasses known cancer genes. Interestingly, we found there were no overlap genes between these signatures, suggesting that the underlying mechanisms related to patients' outcome are not common for the different breast cancer subclasses. Specifically, the basal-like-specific prognostic signature contains six genes. Among them, *LMO2* (a gene encodes a cysteine-rich, two LIM-domain protein) has a most risk weight for basal-like patient prognostic evaluation, which has been reported to be an important transcription factor to promote tumor cell invasion and metastasis in basal-type breast cancer [46]. There are only a few targeted-treatment options for basal-like breast cancers, *LMO2* plays vital roles in breast cancer development, and its inhibitors may be useful targets for the therapy with basal-like patients with a poorer outcome [47]. The HER2-enriched-specific signature is composed of five genes. The gene *ABCC5* could promote breast cancer metastasis to bone [48] and induce chemoresistance in breast tumor initiating cells [49], which may have implications for the better treatment of the HER2-enriched populations with poor prognosis. The Luminal A-specific prognostic signature contains eight genes. We found three genes (*PTGIS*, *ST3GAL1* and *PGAP1*) were involved in metabolic pathway and two genes (*IL1R1* and *LONRF1*) were part of cytokine signaling in immune system pathway. The Luminal B-specific prognostic signature contains 10 genes, which is the largest gene set among the four signatures. A gene named *FBXO4* whose transcriptional level was significantly higher in the luminal-subtype breast cancer cell than in the basal subtype [50]. A recent study also point out that *FBXO4* has the prognostic power in the luminal subtype breast cancer [51].

Rather than adopting the 'one size fits all' approach, we anticipated that the development of subtype-specific prognostic signatures will enable a more effective prediction of patients' survival and may serve as a good starting point for personalized decision-making.

Key points

- Breast cancers are not a single disease but a heterogeneous entity having distinct subtypes with diverse disease biology, clinical behavior and therapy sensitivity. The identification of subtype-specific prognostic biomarkers is important for breast cancer translational research.
- The representative gene-expression-based signatures with significant prognostic value were generally not suitable to evaluate the prognosis of the intrinsic breast cancer subtype; there is a great need for a more subtle risk hierarchy with the outcome of breast cancer.
- We describe a subtype-specific prognostic signature discovery approach and the resulting signature is confirmed as a promising and potential independent prognostic indicator for the corresponding cohort.
- Our study demonstrated that 'one-size-fits-all' solution may not be appropriate for the prognostic prediction of breast cancer individuals belonging to different biological subgroups; it is a necessary step to make a higher resolution for the identification of prognostic signatures in the future.

Supplementary Data

Supplementary data are available online at https://academic.oup.com/bib.

Funding

This work was supported in part by the National Program on Key Basic Research Project (973 Program, 2014CB910504), the National High Technology Research and Development Program of China (863 Program, 2014AA021102), the National Natural Science Foundation of China (61473106 and 61573122), the China Postdoctoral Science Foundation (2016M600260), Wu Lien-Teh Youth Science Fund Project of Harbin Medical University (WLD-QN1407), special funds for the construction of higher education in Heilongjiang Province (UNPYSCT-2016049) and the Heilongjiang Postdoctoral Foundation (LBH-Z16098), the Funds for the Graduate Innovation Fund of Heilongjiang Province (YJSCX2015-5HYD) and Harbin Special Funds for Innovative Talents of Science and Technology Research Project (2017RAGXJ004).

References

1. Ferlay J, Shin HR, Bray F, *et al*. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010;**127**:2893–917.
2. Miller KD, Siegel RL, Lin CC, *et al*. Cancer treatment and survivorship statistics, 2016. *CA Cancer J Clin* 2016;**66**:271–89.
3. Liu R, Wang X, Chen GY, *et al*. The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N Engl J Med* 2007;**356**:217–26.
4. Ciriello G, Gatza ML, Beck AH, *et al*. Comprehensive molecular portraits of invasive lobular breast cancer. *Cell* 2015;**163**:506–19.
5. Harbeck N, Sotlar K, Wuerstlein R, *et al*. Molecular and protein markers for clinical decision making in breast cancer: today and tomorrow. *Cancer Treat Rev* 2014;**40**:434–44.
6. Symmans WF, Hatzis C, Sotiriou C, *et al*. Genomic index of sensitivity to endocrine therapy for breast cancer. *J Clin Oncol* 2010;**28**:4111–9.
7. Oakman C, Santarpia L, Di Leo A. Breast cancer assessment tools and optimizing adjuvant therapy. *Nat Rev Clin Oncol* 2010;**7**:725–32.
8. van't Veer LJ, Dai H, van de Vijver MJ, *et al*. Gene expression profiling predicts clinical outcome of breast cancer, *Nature* 2002;**415**:530-536.
9. Paik S, Shak S, Tang G, *et al*. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004;**351**:2817–26.
10. Filipits M, Rudas M, Jakesz R, *et al*. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clin Cancer Res* 2011;**17**:6012–20.
11. Reis-Filho JS, Pusztai L. Gene expression profiling in breast cancer: classification, prognostication, and prediction. *Lancet* 2011;**378**:1812–23.
12. Sotiriou C, Pusztai L. Gene-expression signatures in breast cancer. *N Engl J Med* 2009;**360**:790–800.
13. Glas AM, Floore A, Delahaye LJ, *et al*. Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics* 2006;**7**:278.
14. Albain KS, Barlow WE, Shak S, *et al*. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. *Lancet Oncol* 2010;**11**:55–65.
15. Toi M, Iwata H, Yamanaka T, *et al*. Clinical significance of the 21-gene signature (Oncotype DX) in hormone receptor-positive early stage primary breast cancer in the Japanese population. *Cancer* 2010;**116**:3112–8.
16. Dubsky P, Filipits M, Jakesz R, *et al*. EndoPredict improves the prognostic classification derived from common clinical guidelines in ER-positive, HER2-negative early breast cancer. *Ann Oncol* 2013;**24**:640–7.
17. Denkert C, Kronenwett R, Schlake W, *et al*. Decentral gene expression analysis for ER+/Her2- breast cancer: results of ssa proficiency testing program for the EndoPredict assay. *Virchows Arch* 2012;**460**:251–9.
18. Taherian-Fard A, Srihari S, Ragan MA. Breast cancer classification: linking molecular mechanisms to disease prognosis. *Brief Bioinform* 2015;**16**:461–74.
19. Sorlie T, Perou CM, Tibshirani R, *et al*. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001;**98**:10869–74.
20. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. *Nature* 2012;**490**:61–70.
21. Prat A, Pineda E, Adamo B, *et al*. Clinical implications of the intrinsic molecular subtypes of breast cancer. *Breast* 2015;**24**(suppl 2):S26–35.
22. Llombart-Cussac A, Cortes J, Pare L, *et al*. HER2-enriched subtype as a predictor of pathological complete response following trastuzumab and lapatinib without chemotherapy in early-stage HER2-positive breast cancer (PAMELA): an open-label, single-group, multicentre, phase 2 trial. *Lancet Oncol* 2017;**18**:545–54.
23. Sims AH, Howell A, Howell SJ, *et al*. Origins of breast cancer subtypes and therapeutic implications. *Nat Clin Pract Oncol* 2007;**4**:516–25.
24. Parker JS, Mullins M, Cheang MC, *et al*. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 2009;**27**:1160–7.
25. Cheang MC, Voduc D, Bajdik C, *et al*. Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. *Clin Cancer Res* 2008;**14**:1368–76.
26. Bleyer A, Welch HG. Effect of three decades of screening mammography on breast-cancer incidence. *N Engl J Med* 2012;**367**:1998–2005.
27. Zhao Q, Shi X, Xie Y, *et al*. Combining multidimensional genomic measurements for predicting cancer prognosis: observations from TCGA. *Brief Bioinform* 2015;**16**:291–303.
28. Hatzis C, Pusztai L, Valero V, *et al*. A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer. *JAMA* 2011;**305**:1873–81.
29. Yasrebi H. Comparative study of joint analysis of microarray gene expression data in survival prediction and risk assessment of breast cancer patients. *Brief Bioinform* 2016;**17**:771–85.
30. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* 2012;**13**:484–92.
31. Mermel CH, Schumacher SE, Hill B, *et al*. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* 2011;**12**:R41.
32. Bamford S, Dawson E, Forbes S, *et al*. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer* 2004;**91**:355–8.
33. Ishwaran H, Gerds TA, Kogalur UB, *et al*. Random survival forests for competing risks. *Biostatistics* 2014;**15**:757–73.
34. Moreno-Betancur M, Sadaoui H, Piffaretti C, *et al*. Survival analysis with multiple causes of death: extending the competing risks model. *Epidemiology* 2017;**28**:12–9.
35. Bland JM, Altman DG. Survival probabilities (the Kaplan-Meier method). *BMJ* 1998;**317**:1572.
36. Bland JM, Altman DG. The logrank test. *BMJ* 2004;**328**:1073.
37. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med* 2004;**23**:2109–23.
38. Fan C, Oh DS, Wessels L, *et al*. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 2006;**355**:560–9.
39. D'Aiuto F, Callari M, Dugo M, *et al*. miR-30e* is an independent subtype-specific prognostic marker in breast cancer. *Br J Cancer* 2015;**113**:290–8.

40. Liu Z, Zhang XS, Zhang S. Breast tumor subgroups reveal diverse clinical prognostic power. *Sci Rep* 2014;**4**:4002.
41. Iglesia MD, Vincent BG, Parker JS, *et al*. Prognostic B-cell signatures using mRNA-seq in patients with subtype-specific breast and ovarian cancer. *Clin Cancer Res* 2014;**20**: 3818–29.
42. Bueno-de-Mesquita JM, van Harten WH, Retel VP, *et al*. Use of 70-gene signature to predict prognosis of patients with node-negative breast cancer: a prospective community-based feasibility study (RASTER). *Lancet Oncol* 2007;**8**:1079–87.
43. Goldstein LJ, Gray R, Badve S, *et al*. Prognostic utility of the 21-gene assay in hormone receptor-positive operable breast cancer compared with classical clinicopathologic features. *J Clin Oncol* 2008;**26**:4063–71.
44. Van der Auwera I, Limame R, van Dam P, *et al*. Integrated miRNA and mRNA expression profiling of the inflammatory breast cancer subtype. *Br J Cancer* 2010;**103**:532–41.
45. Volinia S, Croce CM. Prognostic microRNA/mRNA signature from the integrated analysis of patients with invasive breast cancer. *Proc Natl Acad Sci U S A* 2013;**110**:7413–7.
46. Liu Y, Wang Z, Huang D, *et al*. LMO2 promotes tumor cell invasion and metastasis in basal-type breast cancer by altering actin cytoskeleton remodeling. *Oncotarget* 2017;**8**: 9513–24.
47. Sang M, Ma L, Sang M, *et al*. LIM-domain-only proteins: multifunctional nuclear transcription coregulators that interacts with diverse proteins. *Mol Biol Rep* 2014;**41**: 1067–73.
48. Mourskaia AA, Amir E, Dong Z, *et al*. ABCC5 supports osteoclast formation and promotes breast cancer metastasis to bone. *Breast Cancer Res* 2012;**14**:R149.
49. Shah NR, Chen H. MicroRNAs in pathogenesis of breast cancer: Implications in diagnosis and treatment. *World J Clin Oncol* 2014;**5**:48–60.
50. Kang JH, Choi MY, Cui YH, *et al*. Regulation of FBXO4-mediated ICAM-1 protein stability in metastatic breast cancer. *Oncotarget* 2017;**8**:83100–13.
51. Akcakanat A, Zhang L, Tsavachidis S, *et al*. The rapamycin-regulated gene expression signature determines prognosis for breast cancer. *Mol Cancer* 2009;**8**:75.