Published in final edited form as:

Breast Cancer Res Treat. 2018 August ; 171(1): 199–207. doi:10.1007/s10549-018-4816-9.

Gene expression in triple-negative breast cancer in relation to survival

Shuyang Wang¹, Alicia Beeghly-Fadiel^{1,5}, Qiuyin Cai¹, Hui Cai¹, Xingyi Guo¹, Liang Shi², Jie Wu¹, Fei Ye³, Qingchao Qiu¹, Ying Zheng⁴, Wei Zheng¹, Ping-Ping Bao², Xiao-ou Shu¹

¹Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Vanderbilt-Ingram Cancer Center, Nashville, TN, USA

²Shanghai Municipal Center for Disease Control and Prevention, Shanghai, China

³Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, USA

⁴Shanghai Cancer Hospital, Fudan University, Shanghai, China

⁵Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Institute for Medicine and Public Health, Vanderbilt University Medical Center, Vanderbilt-Ingram Cancer Center, 2525 West End Avenue, 838-A, Nashville, TN 37203, USA

Abstract

Purpose—The identification of biomarkers related to the prognosis of triple-negative breast cancer (TNBC) is critically important for improved understanding of the biology that drives TNBC progression.

Methods—We evaluated gene expression in total RNA isolated from formalin-fixed paraffin-embedded tumor samples using the NanoString nCounter assay for 469 TNBC cases from the Shanghai Breast Cancer Survival Study. We used Cox regression to quantify Hazard Ratios (HR) and corresponding confidence intervals (CI) for overall survival (OS) and disease-free survival (DFS) in models that included adjustment for breast cancer intrinsic subtype. Of 302 genes in our discovery analysis, 22 were further evaluated in relation to OS among 134 TNBC cases from the Nashville Breast Health Study and the Southern Community Cohort Study; 16 genes were further evaluated in relation to DFS in 335 TNBC cases from four gene expression omnibus datasets. Fixed-effect meta-analysis was used to combine results across data sources.

Results—Twofold higher expression of *EOMES* (HR 0.90, 95% CI 0.83–0.97), *RASGRP1* (HR 0.89, 95% CI 0.82–0.97), and *SOD2* (HR 0.80, 95% CI 0.66–0.96) was associated with better OS.

Alicia Beeghly-Fadiel alicia.beeghly@vanderbilt.edu.

Data availability Summary statistics from SBCSS, SCCS, and NBHS data analysis will be provided upon request; GEO data are publically available through the NCBI (<https://www.ncbi.nlm.nih.gov/geo/>).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

Ethical approval All participants of the SBCSS, SCCS, and NBHS provided informed consent; institutional approval was garnered from all appropriate review boards, and all experiments were conducted in compliance with relevant federal and state laws.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10549-018-4816-9>) contains supplementary material, which is available to authorized users.

Twofold higher expression of *EOMES* (HR 0.89, 95% CI 0.81–0.97) and *RASGRP1* (HR 0.87, 95% CI 0.81–0.95) was also associated with better DFS. On the contrary, a doubling of *FA2H* (HR 1.14, 95% CI 1.06–1.22) and *GSPT1* (HR 1.33, 95% CI 1.14–1.55) expression was associated with shorter DFS.

Conclusions—We identified five genes (*EOMES*, *FA2H*, *GSPT1*, *RASGRP1*, and *SOD2*) that may serve as potential prognostic biomarkers and/or therapeutic targets for TNBC.

Keywords

Triple-negative breast cancer; TNBC; Survival; Progression; Gene expression

Introduction

Triple-negative breast cancer (TNBC) is a heterogeneous group of breast malignancies that lack expression of the estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor (HER)2 [1]. Without targets for tailored therapy, cytotoxic chemotherapy is currently the only approved treatment for TNBC, and with high rates of relapse, recurrence, metastasis, and mortality, the prognosis of TNBC is generally poor [2–4]. TNBC accounts for approximately 15% of the 1,676,600 women diagnosed annually with breast cancer worldwide [1, 5], with considerable variation by both geography and race/ethnicity [6, 7]. A better understanding of the factors that contribute to the aggressiveness and poor prognosis is needed to elucidate potential new therapeutic targets and treatment options for TNBC [2, 3].

Gene expression has been used to classify breast cancer into intrinsic subtypes that include luminal A, luminal B, HER2-enriched, normal-like, and basal-like tumors [8–10]. TNBC is frequently thought to be mostly basal-like, but the disease is actually heterogeneous and can include all intrinsic subtypes [1, 4]. In relation to TNBC disease outcomes, expression of both individual and groups of genes has been evaluated. For example, we previously found that low dual specificity phosphatase 4 (*DUSP4*) expression was associated with worse breast cancer specific and overall mortality among 469 TNBC cases [11]. Another study evaluated five human epidermal growth factor receptor genes and found that high *ERBB4* expression was associated with significantly shorter time to distant recurrence-free survival (DRFS) [12]. In addition, examination of 250 cancer-related genes among the same 203 cases indicated that increased expression of five genes (*SMAD2*, *HRAS*, *KRT6A*, *TP63*, and *ETV6*) and decreased expression of two genes (*NFKB1* and *MDM4*) was associated with TNBC prognosis [13]. Another study measured expression of 579 immunology-related genes among 55 Korean TNBC cases and found 92 and 42 genes nominally associated with disease-free survival (DFS) in unadjusted and multivariable adjusted models, respectively; by also comparing gene expression in tumors that did and did not respond to neoadjuvant chemotherapy, three genes (*NFKB1*, *TRAF1*, and *CXCL13*) were identified as better prognostic factors for DFS [14]. However, neither this study, nor other prior studies, included adjustment for intrinsic subtype of breast cancer. Therefore, we undertook this investigation to evaluate gene expression in relation to survival outcomes with multivariable adjustment for intrinsic subtype in order to identify independent prognostic factors and potential new therapeutic targets for TNBC.

Materials and methods

Study populations

TNBC cases included in the current study were participants of the Shanghai Breast Cancer Survival Study (SBCSS), a population-based cohort study that has been previously described [15]. Briefly, 5042 incident cases, aged 20–75 years in 2002–2006 were recruited approximately 6 months after a diagnosis of breast cancer in Shanghai, China. In-person interviews took place at 18-month, 36-month, 5-year and 10 years post-diagnosis and periodic linkages with the Shanghai Vital Statistics Registry (VSR) and Shanghai Cancer Registry (SCR) were conducted to collect information on demographics, lifestyle factors, clinical characteristics, and recurrence and mortality outcomes. Tumor sections were collected from diagnostic hospitals, where medical charts were also reviewed to gather information on clinical characteristics (including stage and grade of disease), first-line treatments received, and tumor ER/PR status. HER2 status was assessed in the Vanderbilt Molecular Epidemiology Laboratory following the DAKO Envision™ kit protocol (DAKO, Cat# K4011) for the HER2 cytoplasmic domain (DAKO, Cat# A0485, 1:100). Detailed descriptions of demographic and clinical predictors for breast cancer survival among SBCSS TNBC cases have been previously reported [11, 16].

Additional TNBC cases were drawn from the participants of the Nashville Breast Health Study (NBHS) and the Southern Community Cohort Study (SCCS). Briefly, the NBHS included 2,694 primary incident breast cancer cases age 25–75 who resided in the Nashville metropolitan area of Tennessee between 2001 and 2011; demographic information was ascertained through telephone interviews and clinical information including histopathology, tumor stage and grade, and hormone receptor status was obtained by reviewed medical and pathology records [17]. The SCCS is a prospective cohort of 85,806 participants, aged 40–79 years, recruited from 12 southeastern states in the US between 2002 and 2009 [18]. Ascertainment of incident breast cancer cases among SCCS participants was obtained through annual linkage with state cancer registries that span the SCCS catchment area; tumor registry information included first-line treatment and tumor ER, PR, and HER2 status was obtained. Information on vital status for NBHS and SCCS TNBC cases was obtained by linkage with the National Death Index. In these three cohorts, the proportion of TNBC cases was approximately 11, 16, and 13%, respectively.

Targeted gene expression

Gene expression levels were measured in total RNA isolated from formalin-fixed paraffin-embedded (FFPE) breast cancer tissues using Nanostring technology. Briefly, FFPE tissue sections and Haematoxylin and Eosin (H&E) stained slides were reviewed by a study pathologist to select areas that contained at least 80% tumor cells for RNA extraction. Total RNA was isolated and purified using Qiagen's miRNeasy FFPE Kit (Qiagen, Valencia, CA), following manufacturer's instructions. RNA quantity and quality were checked by Nanodrop using a wavelength spectrum of 220–320 nm, evaluating the 260/280 ratio, and by separation on an Agilent BioAnalyzer. Of 525 SBCSS tumor samples, those with insufficient tumor tissue for RNA extraction ($N=19$) or a low concentration of resulting RNA ($N=10$) were excluded, leaving 496 SBCSS samples from which to assay gene expression (94.5%).

A custom-designed nCounter Gene Expression Code-Set of 311 selected target mRNA was conducted using NanoString nCounter technology according to standard protocol. Target genes included those from the PAM50 prognostic assay [10], drug metabolism genes, genes selected from the literature, as well as 10 housekeeping genes from which 5 were selected for reference (*ACTB*, *RPLP0*, *MRPL19*, *SF3A1*, and *PSMC4*). Gene expression normalization and quality control were conducted using the R package NanoStringNorm (version 1.1.16). We performed the following data normalization steps: (1) reducing technical variation by adjusting each sample's counts based on its relative value to the geometric mean of all samples; (2) correcting background count levels by subtracting from each sample's count the mean plus two standard deviations of counts from negatives controls; and (3) normalizing for sample RNA content using the geometric mean of expression levels from the five reference genes. Quality control resulted in excluding samples with (1) a positive normalization factor outside of the acceptable range (0.3–3.0) ($N = 6$); (2) a background level > 3 standard deviations from the mean ($N = 4$); (3) 90% gene expression missing after background level correction ($N = 11$); or (4) RNA levels > 3 standard deviations from the mean ($N = 6$). This left 469 (94.6%) SBCSS samples with expression data that passed quality control for inclusion in the current analysis. The same assay and data normalization protocol were implemented to measure gene expression in 86 NBHS and 48 SCCS TNBC samples. To classify intrinsic molecular subtype, we applied a centroid-based clustering algorithm based on the classification developed by Parker et al. known as PAM50 [10].

Publically available clinical and gene expression data for TNBC cases were also obtained from the Gene Expression Omnibus (GEO) of the National Center for Biotechnology Information (NCBI). To be included in our analysis, GEO datasets were required to have 10 events among 50 TNBC cases, as well as outcomes data available for 60% of all subjects. Of 10 NCBI breast cancer datasets with molecular subtype and survival information identified, three independent datasets (GSE21653 [19], GSE25055 [20], and GSE25065 [20]) and 1 combined dataset (GSE10886 [10], GSE6128 [21], GSE3165 [22], and GSE3521 [23]) met our inclusion criteria and were included in our analysis.

Statistical analysis

We evaluated overall survival (OS) and disease-free survival (DFS), defined as the intervals between the date of diagnosis and death or disease recurrence, respectively. Intervals without outcome events were censored at either the date of last linkage (12/21/15) or last follow-up (09/20/16), as appropriate. Associations between gene expression and DFS and OS were quantified by Hazard Ratios (HRs) and corresponding 95% confidence intervals (CIs) derived from Cox proportional hazards regression. Expression data were $\log_2$ transformed for normality, and then analyzed as continuous variables. Regression models included adjustment for age at diagnosis (continuous), stage of disease (0/I, II, III/IV, and unknown), and intrinsic subtype. In addition, we also adjusted for *ALDH1A1* and *DUSP4* expression, since we previously reported that they were significantly related to TNBC outcomes [11, 16]. For genes that had significant association in the SBCSS, we further evaluated associations with OS among NBHS and SCCS cases and associations with DFS using GEO data. Results across datasets were combined using fixed-effects meta-analysis regression

methods. Unless otherwise stated, analyses were conducted using SAS (version 9.4; SAS Institute, Inc., Cary, NC).

Results

A total of 938 TNBC cases were evaluated in the current study, including 469 from the SBCSS, 134 from the SCCS and NBHS, and 335 cases from GEO datasets (Table 1). Most TNBC cases were high grade (55.4% in the SBCSS to 71.6% in the SCCS and NBHS) and had basal-like intrinsic subtypes (41.8% in the SBCSS to 83.6% in the SCCS and NBHS). Distributions of characteristics and associations with TNBC outcomes were generally as expected, with shorter OS among African Americans and shorter OS and DFS among those with late stage disease (III/IV).

Among 469 TNBC cases from the SBCSS, nominally significant associations with either DFS or OS were found for 55 genes (Supplemental Table 1). After including adjustment for age, stage of disease, PAM50 intrinsic subtype, and *ALDH1A1* and *DUSP4* expression, nominally significant associations with OS were found for 22 and 16 genes, respectively.

We further evaluated OS among 134 TNBC cases from the NBHS and SCCS; of 22 genes tested, two had expression levels that were nominally significant after fixed-effect meta-analysis across the two study populations (Table 2). In these replication analyses, twofold higher expression of *EOMES* (HR 0.76, 95% CI 0.60–0.97) and *SOD2* (HR 0.41, 95% CI 0.18–0.94) was associated with worse OS. Because information on time to disease progression was not available for NBHS and SCCS participants, associations with DFS could not be further evaluated in these study populations.

We further evaluated DFS among 335 TNBC cases from four publically available GEO datasets; of 16 genes tested, six had expression levels that were nominally associated with DFS by fixed-effect meta-analysis across GEO data- sets (Table 3). In these replication analyses, twofold higher expression of *EOMES* (HR 0.46, 95% CI 0.29–0.74), *RASGRP1* (HR 0.71, 95% CI 0.58–0.88), *SLC9A2* (HR 0.62, 95% CI 0.50–0.76), and *TNSI* (HR 0.74, 0.60–0.92) was associated with worse DFS, while twofold higher expression of *FA2H* (HR 1.45, 95% CI 1.19–1.75) and *GSPT1* (HR 2.07, 95% CI 1.44–2.99) was associated with better DFS.

For the seven genes with nominally significant replication associations with either OS or DFS, findings from all analyses were combined by fixed-effect meta-analysis (Table 4). Three genes had significant associations with OS and consistent associations across study populations (*EOMES*, *RASGRP1*, and *SOD2*), while four genes had significant associations with DFS, but somewhat heterogeneous associations across datasets (*EOMES*, *FA2H*, *GSPT1*, and *RASGRP1*). Each doubling of *EOMES* expression was associated with better OS (HR 0.90, 95% CI 0.83–0.97) and DFS (HR 0.89, 95% CI 0.81–0.97). Similarly, each doubling of *RASGRP1* expression was associated with better OS (HR 0.89, 95% CI 0.82–0.97) and DFS (HR 0.87, 95% CI 0.81–0.95). Twofold higher *SOD2* expression was associated with better OS (HR 0.80, 95% CI 0.66–0.96), but the association with DFS (HR 0.86, 95% CI 0.73–1.01) did not reach statistical significance ($p = 0.074$). Similarly, twofold

higher *GSPT1* expression was associated with shorter DFS (HR 1.33, 95% CI 1.14–1.55), but the association with OS (HR 1.16, 95% CI 1.00–1.36) did not reach statistical significance (p-value = 0.054). Each doubling of *FA2H* expression was associated with shorter DFS (HR 1.14, 95% CI 1.06–1.22), but not with OS. *SLC9A2* and *TNSI* expression was not significantly associated with TNBC survival outcomes in analyses that included all available data. Additional adjustment for disease grade did not materially influence any of these results.

To further understand these associations, we first examined correlations between expression in TNBC tumor samples from the SBCSS for all five genes with significant OS or DFS associations (*EOMES*, *FA2H*, *GSPT1*, *RASGRP1*, and *SOD2*). While *FA2H*, *RASGRP1* and *SOD2* were significantly associated with all four other genes, *EOMES* and *GSPT1* were each significantly correlated with three other genes tested (Supplemental Table 2). Significant positive correlations ranged from 0.12 (*RASGRP1* and *GSPT1*) to 0.63 (*EOMES* and *RASGRP1*), while *SOD2* expression was negatively correlated with *FA2H* (– 0.15) and *GSPT1* (– 0.17). Next, mutually adjusted regression models were employed to evaluate independent influences on TNBC outcomes (Supplemental Table 3). When all five genes were included in single regression models, only *RASGRP1* retained significance for OS (HR 0.89, 95% CI 0.81–0.98), while both *FA2H* (HR 1.10, 95% CI 1.02–1.19) and *RASGRP1* (HR 0.90, 95% CI 0.81–0.99) retained significance for DFS. Multiple additional gene combinations were modeled, with many that included significant results. The only combination of three genes where all three were independently associated with DFS after mutual adjustment were *FA2H* (HR 1.10, 95% CI 1.01–1.19), *GSPT1* (HR 1.23, 95% CI 1.03–1.46), and *RASGRP1* (HR 0.87, 95% CI 0.79–0.95).

Discussion

In this two-stage study of gene expression in TNBC in relation to survival outcomes, we found that higher *EOMES*, *RASGRP1*, or *SOD2* expression in tumor tissue was associated with reduced total mortality after adjustment for known prognostic factors, including intrinsic subtype and gene expression of *DUSP4* and *ALDH1*. Similarly, higher *EOMES* or *RASGRP1* and lower *FA2H* or *GSPT1* expression in tumor tissue was associated with reduced TNBC disease-free survival. Mutually adjusted regression analyses suggest that *RASGRP1*, *FA2H*, and *GSPT1* may drive these associations.

Existing literature provides biologic plausibility to support our findings for *RASGRP1*, *FA2H*, and *GSPT1*. RAS guanyl releasing protein 1 (*RASGRP1*) is a guanine nucleotide exchange factor (GEF) that can activate Ras and the Erk/ MAP kinase cascade; it is important for T-cell and B-cell development, differentiation, and homeostasis. Altered *RASGRP1* expression has been implicated in systemic lupus erythematosus [24], and *RASGRP1* variants have been associated with diabetes [25, 26]. With regard to cancer, *RASGRP1* expression was higher in normal colon than adenocarcinoma samples, and higher expression was reported to be correlated with better clinical outcomes for colorectal cancer [27]. In the current study, we found that high *RASGRP1* expression was associated with favorable outcomes for TNBC. Fatty acid 1-hydroxylase (*FA2H*) is the enzyme responsible for free fatty acid hydroxylation prior to incorporation into sphingolipids or glycosylceramides.

FA2H was shown to influence cell cycle and cellular migration in laboratory experiments [28]. Cancer cells with high *FA2H* expression were found to be more sensitive to the plasma membrane-disrupting effects of elisidepsin (Irvalec®, PM02734) [29], a therapeutic agent that has shown promising antineoplastic effects in phase I trials for advanced stage solid tumors [30, 31]. Our finding that high FA2H expression was associated with shorter time to disease progression among TNBC cases suggests that these patients may benefit from elisidepsin; this hypothesis should be further examined starting with both in vitro and in vivo studies. The *G1 to S phase transition 1 (GSPT1)* gene encodes the essential eukaryotic peptide chain release factor GTP-binding subunit ERF3A. With a role in terminating protein translation, this multifunctional GTPase is important for mitosis and apoptosis. In colorectal cancer cells with high *GSPT1* expression, knockdown resulted in reduced proliferation and cell migration [32]. Variation in the amino terminal of the protein has been linked to breast cancer susceptibility [33, 34] and gene expression [35], but ours is the first study to report that *GSPT1* expression is associated with worse TNBC outcomes.

EOMES, or eomesodermin, is a member of the T-box family that encodes a transcription factor [36]. *EOMES* is a key regulatory gene for cell-mediated immunity, and has been shown to be critical for T-cell activation and antitumor response [37, 38]. Breast cancer cell lines that were resistant to taxanes and other treatments had lower *EOMES* expression than sensitive cells [39]. Among 41 patients with metastatic renal cell cancer, higher *EOMES* expression was associated with more favorable performance status and survival outcomes [40], but among 175 colorectal cancer cases, higher *EOMES* was associated with worse overall survival [41]. In the current analysis, we found that higher *EOMES* expression was associated with better TNBC outcomes, although this association was frequently attenuated when regression models included adjustment for genes with correlated expression.

Superoxide dismutase 2 (SOD2), or *manganese superoxide dismutase (MnSOD)*, encodes an essential free radical scavenging enzyme that governs the types of intracellular reactive oxygen species (ROS) produced by mitochondria. Migrating cancer cells were shown to have higher *SOD2* expression than primary tumors, and higher *SOD2* expression was correlated with worse breast cancer survival in unadjusted analyses [42]. Other studies on *SOD2* have focused on a functional variant (rs4880, previously known as rs1799725) where a substitution of valine by alanine (Val16Ala) influences protein expression and function, possibly due to decreased mRNA stability [43]. Breast cancer survival studies have been inconsistent, but a recent meta-analysis combined results across eight studies and found no relationship between this functional *SOD2* variant and breast cancer survival [44]. In the current analysis, we found that higher *SOD2* expression was associated with better OS, and no other studies have specifically evaluated TNBC outcomes.

Strengths of the current study include a relatively large sample size with a total of 938 TNBC cases across discovery and replication stages. We also employed a systematic approach and followed REMARK criteria for reporting tumor biomarker study findings [45]. The largest strength of the current study is that we classified tumors based on PAM50 intrinsic subtype and included adjustment for this using multivariable regression models. We also adjusted for two previously identified prognostic biomarkers, *DUSP4* and *ALDH1* [11, 16]. Major limitations of our study are that not all populations had sufficient information to

evaluate both DFS and OS or to adjust for type of treatments received. Some differences may therefore be present across our study populations, but the majority of tumor samples that we evaluated were collected prior to receiving treatment (89.1% in the SBCSS). Another caveat is that our study populations were ethnically diverse, and this may influence both breast cancer survival and gene expression. But because of our discovery and replication study design, and inclusion of GEO data, we were unable to adjust for race/ethnicity in our analyses. However, when expression of 15,929 genes was compared in tumors from 1047 Caucasian and 1057 Chinese breast cancer cases, there were far fewer significant differences by ethnicity ($N=11$) than by breast cancer subtype ($N=1619$) [46]. Multiple comparisons are another possible concern, but Bonferroni correction would be too conservative as many of the genes included in our study were found to be significantly correlated. To reduce the rate of false positive findings, we applied a two-stage study design. Meta-analysis yielded three consistent significant associations for OS, but the four significant DFS associations were found to have some heterogeneity across the SBCSS discovery and four GEO datasets used for validation. Nevertheless, given that our final analyses included 22 genes for OS and 16 genes for DFS, three associations (*FA2H*, *GSPT1*, and *RASGRP1*) surpassed a Bonferroni corrected significance threshold for associations with DFS ($p=0.003$).

In conclusion, this study found that *EOMES*, *FA2H*, *GSPT1*, *RASGRP1*, and *SOD2* were associated with total or disease-free survival among TNBC patients. In addition to potential prognostic biomarkers, our finding also suggests potential new therapeutic targets for TNBC.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This study was supported by grants from the Department of Defense Breast Cancer Research Program (DAMD 17-02-1-0607) and the National Institutes of Health (R01CA118229; P50CA098131). RNA sample preparation was conducted at the Survey and Biospecimen Shared Resources facility that is supported in part by the Vanderbilt-Ingram Cancer Center (P30 CA068485). The authors thank participants and research team members of the Shanghai Breast Cancer Survival Study for their dedication to the study; Ms. Regina Courtney and Dr. Bo Huang for their help with RNA sample preparation.

References

1. Sharma P (2016) Biology and Management of Patients With Triple-Negative Breast Cancer. *Oncologist* 21:1050–1062 [PubMed: 27401886]
2. Palma G, Frasci G, Chirico A, Esposito E, Siani C, Saturnino C et al. (2015) Triple negative breast cancer: looking for the missing link between biology and treatments. *Oncotarget* 6:26560–26574
3. Hurvitz S, Mead M (2016) Triple-negative breast cancer: advancements in characterization and treatment approach. *Curr Opin Obstet Gynecol* 28:59–69 [PubMed: 26694831]
4. Kumar P, Aggarwal R (2016) An overview of triple-negative breast cancer. *Arch Gynecol Obstet* 293:247–269 [PubMed: 26341644]
5. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A (2015) Global cancer statistics, 2012. *CA Cancer J Clin* 65:87–108 [PubMed: 25651787]

6. DeSantis CE, Fedewa SA, Goding Sauer A, Kramer JL, Smith RA, Jemal A (2016) Breast cancer statistics, 2015: convergence of incidence rates between black and white women. *CA Cancer J Clin* 66:31–42 [PubMed: 26513636]
7. Wang C, Kar S, Lai X, Cai W, Arfuso F, Sethi G et al. (2018) Triple negative breast cancer in Asia: an insider's view. *Cancer Treat Rev* 62:29–38 [PubMed: 29154023]
8. Perou CM, Sørbye T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA et al. (2000) Molecular portraits of human breast tumours. *Nature* 406:747–752 [PubMed: 10963602]
9. Sørbye T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H et al. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 98:10869–10874
10. Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T et al. (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 27:1160–1167 [PubMed: 19204204]
11. Baglia ML, Cai Q, Zheng Y, Wu J, Su Y, Ye F et al. (2014) Dual specificity phosphatase 4 gene expression in association with triple-negative breast cancer outcome. *Breast Cancer Res Treat* 148:211–220 [PubMed: 25281216]
12. Kim JY, Jung HH, Do I-G, Bae S, Lee SK, Kim SW et al. (2016) Prognostic value of ERBB4 expression in patients with triple negative breast cancer. *BMC Cancer* 16:138 [PubMed: 26907936]
13. Park YH, Jung HH, Do I-G, Cho EY, Sohn I, Jung SH et al. (2015) A seven-gene signature can predict distant recurrence in patients with triple-negative breast cancers who receive adjuvant chemotherapy following surgery. *Int J Cancer* 136:1976–1984 [PubMed: 25537444]
14. Lee HJ, Lee JJ, Song IH, Park IA, Kang J, Yu JH et al. (2015) Prognostic and predictive value of NanoString-based immune-related gene signatures in a neoadjuvant setting of triple-negative breast cancer: relationship to tumor-infiltrating lymphocytes. *Breast Cancer Res Treat* 151:619–627 [PubMed: 26006068]
15. Shu XO, Zheng Y, Cai H, Gu K, Chen Z, Zheng W et al. (2009) Soy food intake and breast cancer survival. *JAMA* 302:2437–2443 [PubMed: 19996398]
16. Liu Y, Baglia M, Zheng Y, Blot W, Bao P-P, Cai H et al. (2015) ALDH1A1 mRNA expression in association with prognosis of triple-negative breast cancer. *Oncotarget* 6:41360–41369
17. Cui Y, Deming-Halverson SL, Beeghly-Fadiel A, Lipworth L, Shrubsole MJ, Fair AM et al. (2014) Interactions of hormone replacement therapy, body weight, and bilateral oophorectomy in breast cancer risk. *Clin Cancer Res* 20:1169–1178 [PubMed: 24423614]
18. Signorello LB, Hargreaves MK, Blot WJ (2010) The Southern Community Cohort Study: investigating health disparities. *J Health Care Poor Underserved* 21:26–37
19. Sabatier R, Finetti P, Cervera N, Lambaudie E, Esterni B, Mamessier E et al. (2011) A gene expression signature identifies two prognostic subgroups of basal breast cancer. *Breast Cancer Res Treat* 126:407–420 [PubMed: 20490655]
20. Hatzis C, Pusztai L, Valero V, Booser DJ, Esserman L, Lluch A et al. (2011) A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer. *JAMA* 305:1873–1881 [PubMed: 21558518]
21. Hoadley KA, Weigman VJ, Fan C, Sawyer LR, He X, Troester MA et al. (2007) EGFR associated expression profiles vary with breast tumor subtype. *BMC Genom* 8:258
22. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z et al. (2007) Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol* 8:R76 [PubMed: 17493263]
23. Hu Z, Fan C, Livasy C, He X, Oh DS, Ewend MG et al. (2009) A compact VEGF signature associated with distant metastases and poor outcomes. *BMC Med* 7:9 [PubMed: 19291283]
24. Yasuda S, Stevens RL, Terada T, Takeda M, Hashimoto T, Fukae J et al. (2007) Defective expression of Ras guanyl nucleotide-releasing protein 1 in a subset of patients with systemic lupus erythematosus. *J Immunol* 179:4890–4900 [PubMed: 17878389]
25. Qu HQ, Grant SFA, Bradfield JP, Kim C, Frackelton E, Hakonarson H et al. (2009) Association of RASGRP1 with type 1 diabetes is revealed by combined follow-up of two genome-wide studies. *J Med Genet* 46:553–554 [PubMed: 19465406]

26. Li H, Gan W, Lu L, Dong X, Han X, Hu C et al. (2013) A genome- wide association study identifies GRK5 and RASGRP1 as type 2 diabetes loci in Chinese Hans. *Diabetes* 62:291–298 [PubMed: 22961080]
27. Depeille P, Henricks LM, van de Ven RAH, Lemmens E, Wang CY, Matli M et al. (2015) RasGRP1 opposes proliferative EGFR- SOS1-Ras signals and restricts intestinal epithelial cell growth. *Nat Cell Biol* 17:804–815 [PubMed: 26005835]
28. Alderson NL, Hama H (2009) Fatty acid 2-hydroxylase regulates cAMP-induced cell cycle exit in D6P2T schwannoma cells. *J Lipid Res* 50:1203–1208 [PubMed: 19171550]
29. Herrero AB, Astudillo AM, Balboa MA, Cuevas C, Balsinde J, Moreno S (2008) Levels of SCS7/FA2H-mediated fatty acid 2-hydroxylation determine the sensitivity of cells to antitumor PM02734. *Cancer Res* 68:9779–9787 [PubMed: 19047157]
30. Salazar R, Jones RJ, Oaknin A, Crawford D, Cuadra C, Hopkins C et al. (2012) A phase I and pharmacokinetic study of elisidepsin (PM02734) in patients with advanced solid tumors. *Cancer Chem-other Pharmacol* 70:673–681
31. Goel S, Viteri S, Morán T, Coronado C, Dios JLI, Miguel-Lillo B et al. (2016) Phase I, dose-escalating study of elisidepsin (Irva-lec®), a plasma membrane-disrupting marine antitumor agent, in combination with erlotinib in patients with advanced malignant solid tumors. *Invest New Drugs* 34:75–83 [PubMed: 26627080]
32. Xiao R, Li C, Chai B (2015) miRNA-144 suppresses proliferation and migration of colorectal cancer cells through GSPT1. *Biomed Pharmacother* 74:138–144 [PubMed: 26349975]
33. Malta-Vacas J, Chauvin C, Gonçalves L, Nazaré A, Carvalho C, Monteiro C et al. (2009) eRF3a/GSPT1 12-GGC allele increases the susceptibility for breast cancer development. *Oncol Rep* 21:1551–1558 [PubMed: 19424636]
34. Miri M, Hemati S, Safari F, Tavassoli M (2012) GGCn poly- morphism of eRF3a/GSPT1 gene and breast cancer susceptibility. *Med Oncol* 29:1581–1585 [PubMed: 22101789]
35. Malta-Vacas J, Ferreira P, Monteiro C, Brito M (2009) Differen- tial expression of GSPT1 GGCn alleles in cancer. *Cancer Genet Cytogenet* 195:132–142 [PubMed: 19963113]
36. Probst S, Arnold SJ (2017) Eomesodermin-at dawn of cell fate decisions during early embryogenesis. *Curr Top Dev Biol* 122:93–115 [PubMed: 28057273]
37. Pearce EL, Mullen AC, Martins GA, Krawczyk CM, Hutchins AS, Zediak VP et al. (2003) Control of effector CD8 + T cell function by the transcription factor eomesodermin. *Science* 302:1041–1043 [PubMed: 14605368]
38. Zhu Y, Ju S, Chen E, Dai S, Li C, Morel P et al. (2010) T-bet and eomesodermin are required for T cell-mediated antitumor immune responses. *J Immunol* 185:3174–3183 [PubMed: 20713880]
39. Chang H, Jeung H-C, Jung JJ, Kim TS, Rha SY, Chung HC (2011) Identification of genes associated with chemosensitivity to SAHA/taxane combination treatment in taxane-resistant breast cancer cells. *Breast Cancer Res Treat* 125:55–63 [PubMed: 20224928]
40. Dielmann A, Letsch A, Nonnenmacher A, Miller K, Keilholz U, Busse A (2016) Favorable prognostic influence of T-box transcription factor eomesodermin in metastatic renal cell cancer patients. *Cancer Immunol Immunother* 65:181–192 [PubMed: 26753694]
41. Wang R, Kang Y, Löhr CV, Fischer KA, Bradford CS, Johnson G et al. (2016) Reciprocal regulation of BMF and BIRC5 (Survivin) linked to Eomes overexpression in colorectal cancer. *Cancer Lett* 381:341–348 [PubMed: 27539959]
42. Kamarajugadda S, Cai Q, Chen H, Nayak S, Zhu J, He M et al. (2013) Manganese superoxide dismutase promotes anoikis resistance and tumor metastasis. *Cell Death Dis* 4:e504 [PubMed: 23429290]
43. Sutton A, Imbert A, Igoudjil A, Descatoire V, Cazanave S, Pessayre D et al. (2005) The manganese superoxide dismutase Ala- 16Val dimorphism modulates both mitochondrial import and mRNA stability. *Pharmacogenet Genom* 15:311–319
44. Wang C, Liu Y, Zhou J, Ye L, Chen N, Zhu M et al. (2017) There is no relationship between SOD2 Val-16Ala polymorphism and breast cancer risk or survival. *Mol Clin Oncol* 7:579–590 [PubMed: 29046792]

45. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM (2005) REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer* 93:387–391 [PubMed: 16106245]
46. Huang X, Dugo M, Callari M, Sandri M, De Cecco L, Valeri B et al. (2015) Molecular portrait of breast cancer in China reveals comprehensive transcriptomic likeness to Caucasian breast cancer and low prevalence of luminal A subtype. *Cancer Med* 4:1016–1030 [PubMed: 25787708]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1
Five year survival rates by patient and tumor characteristics among triple-negative breast cancer cases

Characteristic	N (%) [*]	Discovery in SBCSS			Validation in SCCS and NBHS			Validation with GEO ^{**}			
		5-year OS (%)	p-value	5-year DFS (%)	p-value	N (%) [*]	5-year OS (%)	p-value	N (%) [*]	5-year DFS (%)	p-value
Number	469 (100.0)	82.1		80.8		134 (100.0)	85.3		335 (100.0)	59.1	
Age at Diagnosis			< 0.01		0.13			0.27			0.59
<50	202 (43.1)	84.2		83.0		44 (32.8)	81.0		157 (46.9)	60.2	
50–60	126 (43.8)	83.3		80.9		59 (44.0)	92.8		85 (25.4)	60.0	
60	141 (30.1)	78.0		77.6		31 (23.1)	76.8		93 (27.8)	56.1	
Stage of disease			< 0.01		< 0.01			< 0.01			< 0.01
0–I	148 (31.6)	89.9		89.8		59 (44.0)	96.6		32 (9.6)	80.0	
II	258 (55.0)	81.4		80.8		33 (24.64)	84.7		185 (55.2)	65.8	
III–IV	47 (10.0)	59.6		52.1		28 (20.9)	74.7		118 (25.2)	42.1	
Unknown	16 (3.4)	87.5		81.3		14 (10.5)	59.4		0 (0.0)	NA	
Grade			0.06		0.14			0.76			NA
1 low	56 (11.9)	94.6		89.1		2 (1.5)	100.0		0 (0.0)	NA	
2 moderate	153 (32.6)	81.7		80.3		18 (13.4)	88.5		0 (0.0)	NA	
3 high	260 (55.4)	79.6		79.4		96 (71.6)	85.7		0 (0.0)	NA	
Unknown	0 (0.0)					18 (13.5)	77.8		335 (100.0)	100.0	
PAM50 subtype			0.30		0.08			0.60			0.16
Luminal A	128 (27.3)	83.6		83.4		5 (3.7)	100.0		6 (1.8)	100.0	
Luminal B	42 (9.0)	83.3		78.2		2 (1.5)	100.0		3 (0.9)	100.0	
Basal-like	196 (41.8)	77.6		76.4		112 (83.6)	83.3		230 (68.7)	56.2	
Her2-rich	63 (13.4)	90.4		90.2		6 (4.5)	100.0		21 (6.3)	47.5	
Normal-like	40 (8.5)	85.0		82.3		9 (6.7)	88.9		18 (5.4)	67.1	
Unknown	0 (0.0)					0 (0.0)			57 (17.0)	65.9	
Race			NA		NA			0.03			NA
Asian	469 (100.0)	82.1		80.8		0 (0.0)	NA		0 (0.0)	NA	
African American	0 (0.0)	NA		NA		58 (43.6)	78.2		0 (0.0)	NA	
Caucasian	0 (0.0)	NA		NA		75 (56.4)	91.7		0 (0.0)	NA	
Unknown	0 (0.0)	NA		NA		0 (0.0)	NA		335 (100.0)	59.1	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Bold values denote statistical significance at P -value 0.05
* Column percents may not sum to 100 due to rounding error
** Data from four combined GEO datasets, as further described in “Materials and methods”

Table 2

Associations between gene expression and overall survival among TNBC cases from the SBCSS, NBHS, and SCCS

Gene	Discovery: SBCSS (<i>n</i> =469)		Replication: NBHS and SCCS(<i>n</i> =134)		
	HR (95% CI)*	<i>p</i> -value*	HR (95% CI)**	<i>p</i> -value**	<i>p</i> -het**
<i>BAG1</i>	1.12 (1.01–1.24)	0.030	1.21 (0.76–1.92)	0.430	0.746
<i>CCNB1</i>	1.14 (1.01–1.28)	0.034	1.15 (0.46–2.84)	0.766	0.001
<i>CCNE1</i>	1.13 (1.03–1.24)	0.011	1.24 (0.85–1.82)	0.257	0.284
<i>CD44</i>	0.78 (0.64–0.95)	0.013	0.67 (0.34–1.31)	0.242	0.651
<i>CXXC5</i>	1.17 (1.03–1.32)	0.014	0.89 (0.55–1.44)	0.624	0.113
<i>EOMES</i>	0.91 (0.84–0.99)	0.029	0.76 (0.60–0.97)	0.030	0.131
<i>ESR1</i>	1.08 (1.00–1.15)	0.042	1.11 (0.85–1.44)	0.436	0.718
<i>FKBP1A</i>	1.16 (1.00–1.35)	0.045	1.68 (0.55–5.12)	0.358	0.688
<i>FOXA1</i>	1.08 (1.00–1.16)	0.042	1.10 (0.89–1.38)	0.380	0.739
<i>GADD45B</i>	1.27 (1.05–1.53)	0.012	0.54 (0.26–1.10)	0.089	0.464
<i>GPR160</i>	1.11 (1.01–1.22)	0.028	1.02 (0.71–1.48)	0.900	0.733
<i>HSPD1</i>	1.34 (1.01–1.77)	0.043	1.44 (0.61–3.40)	0.408	0.641
<i>MIA</i>	0.94 (0.88–1.00)	0.042	0.94 (0.75–1.17)	0.561	0.167
<i>MRPL19</i>	1.22 (1.00–1.48)	0.048	0.91 (0.50–1.68)	0.770	0.342
<i>PCGF2</i>	1.12 (1.00–1.24)	0.047	1.22 (0.79–1.90)	0.369	0.925
<i>RASGRP1</i>	0.90 (0.83–0.98)	0.012	0.75 (0.55–1.03)	0.073	0.423
<i>SERPINE1</i>	1.13 (1.01–1.26)	0.034	0.79 (0.56–1.11)	0.170	0.623
<i>SIT1</i>	0.92 (0.85–1.00)	0.045	0.86 (0.66–1.12)	0.256	0.301
<i>SMARCE1</i>	1.38 (1.03–1.86)	0.031	0.78 (0.23–2.62)	0.689	0.180
<i>SOD2</i>	0.82 (0.68–1.00)	0.048	0.41 (0.18–0.94)	0.034	0.250
<i>TNS1</i>	1.18 (1.01–1.39)	0.039	0.88 (0.45–1.74)	0.714	0.467
<i>TUBB3</i>	1.10 (1.02–1.19)	0.009	1.18 (0.91–1.53)	0.205	0.968

Bold values denote statistical significance at *P*-value 0.05

* Proportional hazards regression; adjusted for age at diagnosis, stage of disease, intrinsic subtype, and ALDH1A1 and DUSP4 expression

** HR and *p*-value derived from fixed-effect meta-analysis regression; *p*-heterogeneity

Table 3

Association between gene expression and disease-free survival among TNBC cases from the SBCSS and GEO datasets

Gene	Discovery: SBCSS (<i>n</i> =469)		Replication: NBHS and SCCS(<i>n</i> =134)		
	HR (95% CI) *	<i>p</i> -value *	HR (95% CI) **	<i>p</i> -value **	<i>p</i> -het **
<i>BAG1</i>	1.19 (1.06–1.34)	0.003	1.19 (0.91–1.56)	0.195	0.054
<i>CCNE1</i>	1.17 (1.06–1.30)	0.002	0.89 (0.69–1.14)	0.343	0.614
<i>CD44</i>	0.67 (0.53–0.84)	< 0.001	1.16 (0.88–1.53)	0.293	0.003
<i>CYLD</i>	0.83 (0.71–0.97)	0.023	0.81 (0.57–1.16)	0.259	0.170
<i>EOMES</i>	0.91 (0.83–0.99)	0.032	0.46 (0.29–0.74)	0.001	0.938
<i>ESR1</i>	1.09 (1.01–1.18)	0.031	0.95 (0.59–1.51)	0.817	0.584
<i>FA2H</i>	1.09 (1.01–1.18)	0.029	1.45 (1.19–1.75)	< 0.001	0.105
<i>GADD45B</i>	1.27 (1.04–1.55)	0.020	1.18 (0.87–1.61)	0.285	0.726
<i>GPR160</i>	1.13 (1.02–1.26)	0.015	0.90 (0.67–1.22)	0.502	0.976
<i>GSPT1</i>	1.21 (1.02–1.43)	0.032	2.07 (1.44–2.99)	< 0.001	0.400
<i>IL2RB</i>	0.90 (0.81–0.99)	0.033	0.77 (0.57–1.05)	0.101	0.005
<i>RASGRP1</i>	0.91 (0.83–0.99)	0.026	0.71 (0.58–0.88)	0.002	0.020
<i>SLC9A2</i>	1.10 (1.01–1.21)	0.034	0.62 (0.50–0.76)	< 0.001	0.110
<i>SMARCE1</i>	1.41 (1.01–1.95)	0.041	1.04 (0.68–1.58)	0.862	0.362
<i>TNSI</i>	1.20 (1.00–1.43)	0.048	0.74 (0.60–0.92)	0.007	0.536
<i>TUBB3</i>	1.14 (1.05–1.24)	0.001	1.03 (0.76–1.38)	0.857	0.008

Bold values denote statistical significance at *P*-value 0.05

* Proportional hazards regression; adjusted for age at diagnosis, stage of disease, intrinsic subtype, and ALDH1A1 and DUSP4 expression

** HR and *p*-value derived from fixed-effect meta-analysis regression; *p*-heterogeneity

Table 4

Meta-analysis of gene expression and TNBC survival associations

Gene	OS meta-analysis (<i>n</i> = 603)			DFS meta-analysis (<i>n</i> = 804)		
	HR (95% CI) [*]	<i>p</i> -value [*]	<i>p</i> -het [*]	HR (95% CI) [*]	<i>p</i> -value [*]	<i>p</i> -het [*]
<i>EOMES</i>	0.90 (0.83–0.97)	0.006	0.127	0.89 (0.81–0.97)	0.006	0.020
<i>FA2H</i>	1.05 (0.98–1.12)	0.153	0.397	1.14 (1.06–1.22)	0.001	0.010
<i>GSPT1</i>	1.16 (1.00–1.36)	0.054	0.787	1.33 (1.14–1.55)	< 0.001	0.033
<i>RASGRP1</i>	0.89 (0.82–0.97)	0.005	0.394	0.87 (0.81–0.95)	0.001	0.007
<i>SLC9A2</i>	1.07 (0.99–1.16)	0.083	0.171	1.01 (0.93–1.10)	0.775	0.000
<i>SOD2</i>	0.80 (0.66–0.96)	0.017	0.141	0.86 (0.73–1.01)	0.074	0.816
<i>TNSI</i>	1.17 (1.00–1.36)	0.052	0.542	0.98 (0.86–1.13)	0.824	0.006

Bold values denote statistical significance at *P*-value = 0.05^{*} HR and *p*-value derived from fixed-effect meta-analysis regression; *p*-heterogeneity