Relationship Between Quantitative GRB7 RNA Expression and Recurrence after Adjuvant Anthracycline Chemotherapy in Triple Negative Breast Cancer

Joseph A. Sparano¹, Lori J. Goldstein², Barrett H. Childs³, Steven Shak⁴, Diana Brassard², Sunil Badve⁵, Frederick L. Baehner⁴, Roberto Bugarini⁴, Steve Rowley³, Edith Perez⁶, Lawrence N. Shulman⁷, Silvana Martino⁸, Nancy E. Davidson⁹, Paraic A. Kenny¹, George W. Sledge, Jr.⁵, and Robert Gray¹⁰

From the Albert Einstein College of Medicine, Bronx, NY ¹, Fox Chase Cancer Center, Philadelphia, PA ², sanofi-aventis, Cambridge, MA ³, Genomic Health, Inc, Redwood City, CA ⁴, Indiana University Simon Cancer Center, Indianapolis, IN ⁵, Mayo Clinic, Jacksonville, FL ⁶, Dana Farber Cancer Institute, Boston, MA ⁷, The Angeles Clinic and Research Institute, Santa Monica, CA ⁸, University of Pittsburgh Cancer Institute, Pittsburgh, PA ⁹, and the Eastern Cooperative Oncology Group, Brookline, MA ¹⁰.

Acknowledgement: The authors thank Adekunle Raji and other staff members at the Eastern Cooperative Oncology Group Pathology Coordinating Office at the Robert H. Lurie Comprehensive Cancer Center, Chicago, IL.

Supported in part by grants from the Department of Health and Human Services and the National Institutes of Health (CA23318 to the Eastern Cooperative Oncology Group [ECOG]statistical center, CA66636 to the ECOG data management center, CA21115 to the ECOG coordinating center, CA25224 to North Central Cancer Treatment Group, CA32291 to Cancer and Leukemia Group B, CA32012 to Southwest Oncology Group), and by a grant from sanofi-aventis.

Presented in part on as oral presentations on December 11, 2008 at the CTRC-AACR San Antonio Breast Cancer Symposium, San Antonio, TX, and on June 7, 2009 at the 45th Annual Meeting of the American Society of Clinical Oncology Orlando, FL.

Running Title: Quantitative GRB7 expression and recurrence in triple negative breast cancer

Key words: Breast cancer, GRB7 RNA expression, recurrence

Word count: Abstract 241, Translational Relevance – 150, Manuscript 4535

Tables – 3, Figures -3

Supplemental tables – 3, Supplemental Figures -2;

References - 46

Corresponding author: Joseph A. Sparano, MD, Albert Einstein College of Medicine, Montefiore Medical Center-Weiler Division, 1825 Eastchester Rd. Bronx, NY 10461; Phone 718-904-2555; Fax 718-904-2892; Email: jsparano@montefiore.org
ABSTRACT

Purpose: To perform an exploratory analysis of the relationship between gene expression and recurrence in patients with operable triple negative breast cancer (TNBC) treated with adjuvant doxorubicin-containing chemotherapy.

Experimental design: RNA was extracted from archived tumor samples derived from 246 patients with stage I-III TNBC treated with adjuvant doxorubicin-containing chemotherapy, and was analyzed by quantitative RT-PCR for a panel of 374 genes. The relationship between gene expression and recurrence was evaluated using weighted Cox proportional hazards model score tests.

Results: GRB7 was the only gene for which higher expression was significantly associated with increased recurrence in TNBC (Korn’s adjusted p value=0.04). In a Cox proportional hazards model adjusted for clinicopathologic features, higher GRB7 expression was associated with an increased recurrence risk (HR 2.31, p=0.04 using the median as the split). The 5-year recurrence rates were 10.5% (95% confidence intervals [CI] 7.8%, 14.1%) in the low and 20.4% (95% CI 16.5%, 25.0%) in the high GRB7 groups. External validation in other datasets indicated that GRB7 expression was not prognostic in two adjuvant trials including variable systemic therapy, but in two other trials showed that high GRB7 expression was associated with resistance to neoadjuvant doxorubicin and taxane therapy.

Conclusions: GRB7 was associated with an increased risk of recurrence in TNBC, suggesting that GRB7 or GRB7-dependent pathways may serve as potential biomarkers for therapeutic targets. Therapeutic targeting of one or more factors identified which function as interaction nodes or effectors should also be considered.
Translational Relevance

GRB7 RNA expression was associated with a significantly increased risk of recurrence in operable triple negative breast cancer (TNBC) patients treated with adjuvant anthracycline chemotherapy in our analysis, and in external validation studies was associated with resistance to neoadjuvant anthracycline therapy but was not prognostic in operable TNBC patients who received no adjuvant therapy. GRB7 belongs to a small family of mammalian SH2 domain adapter proteins that are known to interact with a number of receptor tyrosine kinases and signaling molecules (including HER1, HER2, and ephrin receptors), with the integrin signaling pathway, and focal adhesion kinase (FAK). GRB7 shares sequence homology with the mig-10 gene of C. elegans, which is required for migration of embryonic neurons, suggesting an important role in cell motility. These findings suggest that high GRB7 expression has a role as a potential biomarker for resistance to anthracycline therapy and serve as a therapeutic target in TNBC.
INTRODUCTION

“Triple negative breast cancer” (TNBC) is defined as breast cancer which lacks expression of the estrogen receptor (ER), progesterone receptor (PR), and HER2/neu protein. Population-based studies indicate that TNBC accounts for about 15% of all breast cancers in the United States, and occurs more commonly in younger women, and women of black race or Hispanic ethnicity.\(^1\) TNBC is associated with a higher risk of distant recurrence, earlier time to recurrence, and worse prognosis after recurrence.\(^2\) \(^3\) About 80% of TNBC are characterized as being of a basal-like breast cancer genotype identified by gene expression profiling.\(^4\) \(^5\) \(^6\) A panel of antibodies which includes cytokeratin markers may more accurately classify basal subtypes than relying on ER, PR, and HER2/neu expression alone.\(^7\) \(^8\) Although inhibitors of poly (ADP-ribose) polymerase may hold promise\(^9\), therapeutic approaches are currently limited to cytotoxic chemotherapy

In the current study that is the subject of this report, we evaluated gene expression patterns from tumors derived from a cohort of patients with stage I-III breast cancer treated with adjuvant doxorubicin-containing chemotherapy. In addition to performing gene expression profiling, we defined breast cancer subsets by standard IHC for ER, PR, and HER2/neu protein expression in a central laboratory.\(^10\) We evaluated differences in gene expression patterns between triple negative disease and HR-positive, HER2/neu negative disease. We also performed an exploratory analysis evaluating the relationship between gene expression and recurrence within the triple negative group. Our objectives were to identify RNA expression biomarkers associated with recurrence within the
TNBC group, and potential therapeutic targets for the TNBC group which may not have been previously recognized.

**METHODS**

**Study Population and Treatment**

The study utilized tumor specimens and clinical information from patients enrolled on trial E2197 (ClinicalTrials.gov identifier NCT00003519), coordinated by the Eastern Cooperative Oncology Group (ECOG), details of which have been reported elsewhere. Briefly, patients were randomly assigned to receive four 3-week cycles of doxorubicin 60 mg/m2 and cyclophosphamide 600 mg/m2 (AC) or docetaxel 60 mg/m2 (AT). Methods for selection of cases included in the genomic analysis have been previously described, and summarized in the CONSORT diagram shown in **supplemental Figure 1**. The characteristics of the sample cohort were comparable to the excluded cohort, also as previously described. The clinical protocol was approved by the institutional review boards of all participating institutions and was carried out in accordance with the Declaration of Helsinki, Food and Drug Administration Good Clinical Practices, and local ethical and legal requirements. The use of specimens for this project was approved the North American Intergroup Correlative Science Committee and by the Northwestern University Institutional Review Board (which oversees the ECOG Pathology Coordinating Office, where the specimens were banked and evaluated).

**Specimen Selection, Processing, and Gene Expression Analysis**

All specimens underwent analysis for tumor grade, and for ER, PR, and HER2/neu protein expression in a central lab as previously described. Quantitative
RNA expression levels were measured by real-time reverse transcriptase polymerase chain reaction (RT-PCR) using gene-specific primers. The panel of 374 genes was assembled by searching the published literature, genomic databases, pathway analysis, and microarray-based gene expression profiling experiments performed in fresh frozen tissue to identify genes likely to be associated with prognosis or response to chemotherapy, as previously reported.

A total of 246 cases were defined as having TNBC by IHC using the methods described above, of whom 15% had a recurrence. This cohort includes only one patient (who did not recur) whose tumor was HER2/neu positive by the Genomic Health RT-PCR assay (≥ 11.5 units), which has been found to exhibit 97% concordance with HER2/neu gene amplification by fluorescent in situ hybridization. If hormone receptor expression were defined by RT-PCR using the Genomic Health cutoff of for ER (≥ 6.5 units) and PR (≥ 5.5 units), then there were 258 triple negative cases (instead of 246; 227 of these are the same) and the estimated slope of continuous GRB7 (linear effect in log expression level) is 0.692 (instead of 0.637 as in the analysis presented in Figure 1), indicating that the results would be similar irrespective of which definition was used. The results presented here are based upon cases selection using the central IHC definition for ER, PR, and HER2/neu, since this most accurately reflects actual clinical practice, and optimizes the methodology for ER/PR testing by IHC (reflected by somewhat higher concordance with central RT-PCR).
Statistical analyses

The relationship between gene expression and recurrence was evaluated using weighted Cox proportional hazards model score tests used to rank genes by their individual significance for predicting recurrence risk as previously described.\textsuperscript{14} Recurrence was defined as distant and/or local-regional recurrence of disease. Among the 59 recurrences, 39 were distant and 20 were local-regional. Adjusted p-values controlling the false discovery proportion (FDP) at $\leq 10\%$ were computed using algorithm B* in Korn et al.\textsuperscript{16} (using 500 permutations) and were applied in a ‘step-down’ fashion.\textsuperscript{17} A weighted algorithms were used to correct for differential sampling of relapse and non-relapse cases.\textsuperscript{18} All p values are two-sided. Differences in gene expression were evaluated between the HR-negative (ie, triple negative; N=246) and HR-positive, HER2-negative groups (N=383) by a weighted t test. Pathway analysis was performed using Ingenuity Pathway Analysis Version 7.6 on the top 40 genes overexpressed in TNBC versus HR-positive, HER2-negative tumors.

External Validation

We evaluated the relationship between \textit{GRB7} expression and outcomes in several publicly available datasets, which used various methods to examine gene expression, including populations with operable disease $^5, ^{19, 20}$ in which we evaluated the relationship between \textit{GRB7} expression and prognosis, and in two neoadjuvant datasets in which we evaluated the relationship between \textit{GRB7} expression and response to anthracycline and taxane-containing neoadjuvant chemotherapy $^{21, 22}$. From studies in which the depositing investigators had annotated their samples as “Basal-like” we used these
designations as an approximation for TNBC. The studies used included the following: (1) Bonnefoi et al data (GSE6861): All patients received neoadjuvant chemotherapy. TNBC status was assigned to samples negative for ER and PR and non-amplified for HER2 as follows: Analysis of the probe for ESR1 (g4503602_3p_at) and ERBB2 (Hs.323910.2.A1_3p_a_at) revealed a clear bimodal distribution. The higher expressing samples for each probe were considered to be ER-positive and HER2-amplified, respectively. The ER-negative, HER2 non-amplified sample set was then further curated to remove samples expressing high levels of PR (g4505766_3p_at). This yielded 76 cases which we designated TNBC, of whom 32 had a pathologic complete response. (2) Tabchy et al data (GSE20271): The data provided by these investigators were annotated for ER, PR and HER2 status. Using their annotation, we selected the ER-negative, PR-negative and HER2-negative samples (58 samples) as our TNBC set. Analysis of the expression levels of ER, PR and HER2 revealed one sample (GSM508138) with very high HER2 levels (within the range of the tumors annotated as HER2-amplified, and the fifth highest level of HER2 in the dataset overall using the probe 210930_s_at for ERBB2). This sample was excluded from our analysis, leaving 57 we considered TNBC. Of these 57, there were 12 pCRs and 45 patients with residual disease (RD). Data from these 57 samples were presented in the manuscript, however the data were also analyzed without excluding the suspicious HER2-high sample, and the difference in GRB7 levels between the pCRs and the RDs remained significant (P<0.05). (3) NKI295/Van de Vijver data: Data provided by these investigators were already annotated by breast tumor molecular subtype. 46 cases annotated as “basal” by Van
de Vijver et al. were considered TNBC for the purpose of our analysis. There was no difference in recurrence rate between $GRB7$-high and $GRB7$-low tumors ($N = 46$, 24 of whom experienced recurrence). We also examined outcomes for the subset of these patients who received systemic chemotherapy ($N = 18$) and there was no difference in outcomes by $GRB7$ levels. (4) Wang et al data (GSE2034)$^{19}$: The data provided by these investigators included ER status. We used the expression values of the probe for ERBB2 (216836_s_at) to exclude tumors likely to be ERBB2-amplified. This left 51 tumors we considered TNBC. There were 17 relapses. (5) Parker et al$^5$: This study included several independent datasets. The specific data we analyzed were obtained from Supplemental Tables 1 and 4 in the online supplementary material for this paper. These data consisted of RT-PCR gene expression data on 279 tumors, designated “UBC-PCR” in Supplementary Table 4. The investigators annotated 61 of the samples as basal-like. We examined the relationship between GRB7 and recurrence free survival in 55 of these samples (3 samples lacked follow-up data for recurrence-free survival and 3 samples lacked expression data for GRB7). There were 26 relapses.” To summarize, in the Parker et al. study$^5$, we analyzed the Basal-like tumors from the 289 patient qRT-PCR dataset presented in the supplementary material. In the remaining studies, we used the gene expression levels reported by the probes for ER, PR and ERBB2 to define subgroups that were negative for ER and PR and not ERBB2-amplified. Samples without follow-up and one sample annotated as Basal-like but with very high levels of ERBB2 mRNA were excluded. This approach provided 285 cases for further
analysis, broken down as follows: 76 cases (32 pCRs), 57 cases (12 pCRs), 46 cases (24 recurrences), 51 cases (17 recurrences) and 55 cases (26 recurrences).

For the neoadjuvant datasets, a major issue was the relative insensitivity of the Affymetrix probesets compared to the sensitivity of the RT-PCR assay used in our analysis. The RT-PCR assay for GRB7 detected differences across a 190-fold range among TNBC cases in our study, while the Affymetrix X3P probe used in the study by Bonnefoi et al had a dynamic range of 3-fold in the TNBC samples, and the Affymetrix U133A probe used in the study by Tabchy et al had a dynamic range of only 5-fold. To try to understand the inefficiency of these probesets we used the Affy package of Bioconductor to extract the raw probe-level data from each microarray used in the Bonnefoi study. This permitted us to individually analyze the signal values for each of the 11 probes that make up the GRB7 probeset across 160 microarrays. In general, the correlation coefficient between the individual probes and the aggregate value reported for GRB7 in each array was relatively good, with Pearson correlation coefficients ranging from 0.79 to 0.93. However, when we stratified the data by into ERBB2/GRB7 amplified (i.e. high GRB7 mRNA) and non-amplified (i.e. low GRB7 mRNA), an interesting difference emerged. For the high GRB7 expressors, the performance of the individual probes remained good, with Pearson Correlation coefficients ranging from 0.80 to 0.91. However, in the low GRB7 expressors (of which the TNBC samples in this study are a substantial subgroup) the correlation between the individual GRB7 probes the overall value reported for the GRB7 probeset was significantly reduced, with Pearson’s correlation coefficients ranging from -0.38 to 0.29). From this analysis, we concluded that although the GRB7 probeset on these arrays works very well when GRB7 mRNA is
abundant, at lower GRB7 levels it suffers from a substantial amount of noise, resulting in a substantial loss of linearity. For this reason, our validation studies have not assumed that the signals in this range are drawn from a Gaussian distribution, and we have used the rank-based Mann-Whitney U-test to attempt to discern whether there are differences in outcome depending on GRB7 levels.

RESULTS

Characteristics of Triple Negative Population

The characteristics of patients with TNBC are shown in Table 1. Most patients were 65 or younger (93%), had tumors with poor histologic grade (90%) that were associated with negative axillary lymph nodes (81%), and occurred in white subjects (87%). When compared with 60 patients who had HR-negative, HER2-positive disease, patients with TNBC were younger (41% vs. 27% less than 46, p=0.04) and more likely to have negative axillary nodes (81% vs. 66%, p=0.01), but were otherwise similar with regard to tumor size, tumor grade, and race.

Genes Associated with Increased Recurrence in Triple Negative Disease

We evaluated the relationship between gene expression and recurrence in the 246 patients with TNBC. There were 6 genes significantly associated with recurrence (adjusted p-value < 0.05), including 1 gene for which increased expression was associated with increased recurrence, and 5 with decreased recurrence. In order to show the relationship between gene expression and recurrence, we show the hazard ratio per
standard deviation (HR/SD), which is the hazard ratio for a difference of one standard deviation of the log expression level of the gene, where the standard deviation refers to the distribution of log expression levels of the gene in the sample. This is a measure of effect that is invariant to rescaling (or shifting) the log expression levels (and invariant to the base used in the logs). The only gene associated with increased recurrence was GRB7 (growth factor receptor bound protein 7) (p=0.04, estimated hazard ratio per standard deviation of gene expression [HR/SD] = 1.74). Genes associated with decreased recurrence included APOC1 (apolipoprotein C1) (p=0.03, HR/SD = 0.59), ESR2 (estrogen receptor beta) (p=0.03, HR/SD = 0.53), PIM2 (PIM2 oncogene, HR/SD = 0.59) (p=0.04), CD68 (macrophage antigen, HR/SD = 0.67) (p=0.05), and BIRC3 (baculoviral IAP repeat 3) (p=0.05, HR/SD = 0.60). There were only six genes whose expression correlated (r>0.4) with GRB7 (found on chromosome 17q12), including ERBB2 (r=0.70, HR/SD = 1.30, chromosome 17q21.1), DDR1 (discoidin domain receptor tyrosine kinase 1; r=0.53, HR/SD = 1.32, chromosome 6p21.3), KRT19 (keratin 19; r=0.49, HR/SD = 1.56, chromosome 17q21.2), ERBB3 (r=0.48, HR/SD = 1.18, chromosome 12q13), GPR56 (G protein-coupled receptor 56; r=0.48, HR/SD = 1.29, chromosome 16q13) and PHB (prohibitin; r=0.42, HR/SD = 1.01, chromosome 17q21). Expression levels were evaluated for genes located on chromosome 17q, including ERBB2, GRB7, PHB, and KRT19, and were significantly lower for ERBB2 and GRB7 in TNBC compared with HER2/neu overexpressing breast cancer (Supplemental Figure 2).

**Relationship Between GRB7 Expression and Recurrence as a Continuous or Categorical Variable**
In order to further characterize the relationship between GRB7 expression and recurrence in TNBC, we evaluated this relationship as a continuous variable using a spline model for the log hazard ratio as shown in Figure 1. GRB7 RNA expression ranged from a low of 2.4 to as high as 10 units, which corresponds to about a 190-fold difference in RNA expression between the highest and lowest values. There was a highly significant relationship between the risk of recurrence and GRB7 expression (p=0.001). The median value of 6.5 was chosen for the additional categorical analyses, which falls on the maximal slope of the curve. The estimated hazard ratio for high versus low expression based on the median split was 2.24 (p=0.006). If tertile splits were used, the difference between the low and intermediate groups was not significant (HR 0.85, 95% CI 0.39,1.88), but was significant for high vs. intermediate is (HR 2.41, 95% CI 1.27, 4.59, p=0.007). Therefore, the median split was used to define GRB7 as a categorical variable in subsequent analyses. Higher GRB7 expression was also associated with significantly higher risk of recurrence in the subset of 60 patients with ER/PR-negative, HER2/neu positive disease (hazard ratio for high vs. low expression 1.75, 95% C.I. 1.02, 3.00, p=0.04).

Relationship Between GRB7 Expression and Clinical Variables

In order to evaluate the relationship between GRB7 expression and clinical variables, we compared the clinical characteristics of patients with high versus low expression levels (supplemental Table 1). There were no significant differences in any clinical characteristic examined between patients who exhibited high versus low GRB7
expression levels, except for fewer patients over 65 years of age with high GRB7 expression (3% vs. 10%, p=0.03).

**Multivariate Analysis: Relationship Between GRB7 Expression and Recurrence**

In order to evaluate the relationship between *GRB7* expression and recurrence adjusted for clinicopathologic variables, Cox proportional hazards models were fit to examine the joint effects of factors on recurrence rates, as shown in Table 2. The models included age, nodal status, centrally determined tumor grade, and tumor size. In Model I, which did not include *GRB7* expression, features associated with an increased risk of recurrence included 1 positive axillary lymph node (versus none) and large tumor size (more than 2 cm). Model II added *GRB7* as a continuous linear variable to Model I; *GRB7* x +2 vs. x was used for the hazard ratio (corresponding to an approximately 4-fold increase in gene expression), where x is an arbitrary value of *GRB7* (comparable to the analysis of Recurrence Score as continuous variable in the report by Paik et al,13). In this model, *GRB7* expression was a highly significant predictor for recurrence (hazard ratio 3.41, 95% CI 1.78, 6.53; p=0.002). Model III added *GRB7* as a dichotomous variable (high vs. low, using the median split) to Model I. In this model, there was also a significant relationship between *GRB7* expression and recurrence (hazard ratio 2.31, 95% CI 1.30, 4.11; p=0.004).

**Pathway Analysis of Differentially Expressed Genes**
Comparing gene expression in TNBC with HR-positive, HER2-negative disease revealed 269 genes (73%) with significantly different expression (p<0.0001). The top 40 genes showing significantly higher expression and lower expression in the TNBC group are shown in supplementary Table 2 and supplementary Table 3, respectively. The top 40 genes showing higher expression in the TNBC group included genes associated with nucleosome assembly (CENPA), kinase activity (TTK), invasion (CTSL2), DNA damage response (CHEK1), transcriptional regulation (MYBL2), transmembrane amino acid transport (SLC7A5), transcription (FOXM1), cell division (CDC20, KIFC2, AURKB, PLK1) and the cell cycle (KIFC1, DEPDC1, CDCA8). Pathway analysis was performed including the 40 top genes showing higher expression in TNBC and showed substantial interaction between the proteins that they encode (Figure 2). Some of the encoded proteins appeared to serve as nodes for interaction, including cytoplasmic proteins such as the anti-apoptotic protein survivin (BIRC5), kinases involved in mitosis such as Aurora Kinase B (AURKB) and in the cell cycle such as cyclin dependent kinase 1 (CDC2), and nuclear transcription factors such as forkhead box M1 protein (FOXM1) and Myb-related protein B (MYBL2).

Validation in Other Datasets

We evaluated the relationship between GRB7 expression and outcomes in several publicly available datasets, which used various methods to examine gene expression. With regard to prognosis, we found no relationship between GRB7 expression and recurrence in patients with node-negative breast cancer who received no adjuvant chemotherapy (excluding patients with high ERBB2 and ER expression)\textsuperscript{19}, and in patients
with basal breast cancer subtype with node-negative or positive disease (some of whom received adjuvant chemotherapy) (data not shown).\textsuperscript{5,20} With regard to prediction, \textit{GRB7} expression level was evaluated in 76 patients with TNBC treated with neoadjuvant sequential epirubicin and docetaxel-containing chemotherapy.\textsuperscript{21} There was no significant difference in the proportion with pathologic complete response (pCR) who had an elevated \textit{GRB7} expression at or above the median compared with below the median (13 of 38 [34\%] vs. 19 of 38 [50\%], Fisher’s Exact Test: \(p = 0.17\)). On the other hand, patients who did not have a pCR had significantly higher median \textit{GRB7} expression (Mann-Whitney test for difference between medians: \(p = 0.0397\), \textbf{Figure 3a}), which was consistent with our findings indicating an association between higher \textit{GRB7} expression and resistance to adjuvant anthracycline therapy (given with or without concurrent docetaxel). Likewise, in a second neoadjuvant dataset in which 12 of 57 patients (21\%) with TNBC had a pCR after treatment with neoadjuvant 5-flourouracil, doxorubicin, and cyclophosphamide (FAC) alone or preceded by paclitaxel (T/FAC)\textsuperscript{22}, median \textit{GRB7} expression levels were significantly higher in the non-responders (\(p = 0.0044\), Mann-Whitney test, \textbf{Figure 3b}).

**DISCUSSION**

We performed an exploratory analysis evaluating the relationship between recurrence and a panel of genes in 246 patients with stage I-III TNBC who received standard doxorubicin-containing chemotherapy and were followed for at least five years. Quantitative RT-PCR was used to measure RNA extracted from paraffin-embedded tumor specimens for a panel of 374 rationally selected genes. All samples were centrally
evaluated for ER, PR, and HER2-expression by immunohistochemistry in a standardized and rigorous manner and confirmed to be triple negative.\textsuperscript{24} GRB7 was the only gene for which higher expression was found to be associated with a significantly elevated risk of recurrence, suggesting that GRB7 may serve as an important biomarker in TNBC, and that perhaps GRB7 or GRB7-dependent pathways may serve as therapeutic targets. Similar to Oncotype DX Recurrence Score in ER-positive breast cancer, the relationship between GRB7 expression and recurrence was evident when evaluated as a continuous variable or dichotomous variable adjusted for other covariates, and the relative risk elevation was comparable. For example, a 50 unit increase in RS (which has a range of 0-100) was associated with a 2-8 fold increase (p<0.001) in risk of distant recurrence in ER-positive disease treated with tamoxifen in the B14 trial\textsuperscript{25}, and 2.1-fold increase (p=0.06) in ER-positive disease treated with adjuvant chemotherapy plus tamoxifen in the E2197 trial\textsuperscript{12}, whereas a 2 unit increase in GRB7 expression (range 2.4-8) was associated with a 3.4-fold increase (p=0.002) in the risk of recurrence in TNBC evaluated in this dataset. When evaluated in four other publicly available datasets, although GRB7 expression was not associated with recurrence in patients who received no adjuvant chemotherapy, median GRB7 expression levels were significantly higher in patients with TNBC who failed to achieve a pCR after preoperative anthracycline and taxane therapy. Although there was no specific GRB7 expression threshold predictive of response in these relatively small neoadjuvant trials, the significantly higher median expression levels in non-responders are nevertheless consistent with a relationship between GRB7 expression and sensitivity to anthracycline and/or taxane therapy, In addition, Ramsey et al have recently reported an association between high GRB7 protein expression and
recurrence in the presence or absence of adjuvant chemotherapy\textsuperscript{26}, providing additional independent evidence supporting our findings.

\textit{GRB7} belongs to a small family of mammalian SH2 domain adapter proteins that are known to interact with a number of receptor tyrosine kinases and signaling molecules (including HER1, HER2, and ephrin receptors), with the integrin signaling pathway, and with focal adhesion kinase (FAK).\textsuperscript{27, 28} \textit{GRB7} shares sequence homology with the \textit{mig-10} gene of \textit{C. elegans}, which is required for migration of embryonic neurons, suggesting an important role in cell motility.\textsuperscript{29} \textit{GRB7} is also included in the 21 gene signature in estrogen-receptor-positive disease\textsuperscript{25}, and in the 512 intrinsic gene set\textsuperscript{4} and PAM50 gene set\textsuperscript{5}, indicating other evidence that is may be an important biomarker. \textit{GRB7} is located on the same amplicon as the \textit{erbB2} gene and thus usually coamplified in HER2/neu overexpressing breast cancers, and we confirmed that \textit{GRB7} expression levels were significantly lower in HER2/neu non-overexpressing tumors. Although \textit{GRB7} expression was correlated with \textit{ERBB2} expression within the TNBC group (r=0.70), \textit{GRB7} but not \textit{ERBB2} expression was significantly associated with recurrence in this population, supporting its role as a prognostic marker in this setting. Supporting its potential as a therapeutic target, several inhibitors of \textit{GRB7} have been developed, some of which have been shown to potentate the effects of cytotoxic therapy and trastuzumab.\textsuperscript{30-33} In addition, inhibitors of \textit{GRB7}-dependent pathways such as FAK, ephrins\textsuperscript{34}, and integrins\textsuperscript{35} offers additional therapeutic potential. Evaluation of a highly specific \textit{GRB7} peptide inhibitor (G7-18NATE) in a panel of 4 TNBC cell lines revealed that \textit{GRB7} inhibition significantly impaired migration and invasion, reduced colony formation in 3D culture by promoting apoptosis, and synergistically sensitized TNBC cell
lines to doxorubicin and docetaxel.\textsuperscript{36} Taken together, these findings \textit{GRB7} as a key mediator of migration, invasion, colony growth and chemoresistance of TNBC, and suggest that in addition to serving as a prognostic or predictive biomarker, \textit{GRB7}-dependent pathways may prove to be important therapeutic target in TNBC.

Although \textit{GRB7} was the only gene for which increased expression was associated with an increased risk of recurrence, increased expression of several other genes was found to be associated with a significantly lower risk of recurrence, including CD68, a macrophage antigen likely reflecting host cells infiltrating the tumor. CD68 is also one of the 16 tumor associated genes in the Oncotype DX\textsuperscript{®} Recurrence Score\textsuperscript{®} for HR-positive disease, in which higher expression is likewise associated with a more favorable outcome. This suggests the importance of host response, or specific antigens eliciting a host response, that may play an important role in determining prognosis in TNBC.

The analysis of differential gene expression comparing TNBC with HR-positive, HER2-negative disease provided some additional information suggesting potential therapeutic targets which are summarized in Table 3. These include inhibitors of aurora kinase B\textsuperscript{37-39}, polo-like kinase 1\textsuperscript{40,41}, kinesin family member C1, checkpoint kinase 1\textsuperscript{42}, and forkhead box M1\textsuperscript{43}. The pathway analysis indicates that some of these proteins (encoded by their genes) may be particularly vulnerable targets because they serve as nodes for interaction or transcriptional effectors, including aurora kinase B (\textit{AURKB}), cyclin dependent kinase 1 (\textit{CHEK1}), forkhead box M1 (\textit{FOXM1}), and myb related protein B (\textit{MYBL2}). Other reports have likewise confirmed the important role of some of the genes. For example, Thorner et al reported a significant association between the G/G genotype of a non-synonymous \textit{MYBL2} germline variant (rs2070235, S427G) and an
increased risk of basal-like breast cancer [OR 2.0, 95% CI (1.1-3.8)]; they also showed that MYBL2 is involved in cell cycle control, and that its dysregulation contributes to increased sensitivity specifically to DNA topoisomerase II inhibitors.44

There are several notable strengths of this analysis. First, this is one of the largest training sets specifically evaluating gene expression in a uniformly treated cohort of patients with TNBC; inadequate sample size is recognized as a major limitation of previous studies.45 Second, the TNBC group was defined by standard IHC methods commonly used in clinical practice, but performed in a standardized and rigorous manner in a central lab. Third, we evaluated a limited panel of candidate genes that were rationally selected because of their known or postulated association with prognosis or response to chemotherapy rather than a genome wide approach, and used a standardized RT-PCR method that brings precision and large dynamic range. This offers the potential to reduce the likelihood of identifying falsely positive associations by enriching for candidate genes likely to be associated with recurrence, and provides confidence in assuring reproducibility of the identified genes and the method of measuring their expression. In addition, stringent statistical methods were used to control false discovery.17 Some of the analyses were also adjusted for clinicopathologic variables in order to explore whether specific genes, such as GRB7, provided information beyond standard clinicopathologic measures. Lastly, we performed external validation in four independent data sets and confirmed that high GRB7 expression was associated with resistance to doxorubicin and taxane therapy, but did not provide prognostic information in the absence of systemic chemotherapy.
There were also several limitations of this analysis. A fundamental premise of our study is that increased gene transcription, as reflected by RNA expression levels, may identify potential therapeutic targets, biomarkers predictive of clinical behavior or response to therapy, or both. However, altered transcription may reflect an effect rather than a cause of the malignant phenotype. In addition, searching for activating gene mutations, oncogenes, or inactivated tumor suppressor genes may be a more fruitful strategy for therapeutic targeting.\textsuperscript{46} On the other hand, there is a clear precedent for effectively targeting pathways in breast cancer that are not associated with discernable activating mutations, as exemplified by antiestrogen therapy.

In conclusion, we identified several genes that are novel therapeutic targets in TNBC, and which may have potential clinical utility. Validation in preclinical systems will be required for drug development, and additional validation in other clinical datasets will be required for clinical application.
Table Legends:

Table 1. Characteristics of triple negative patient population included in this analysis.

Table 2. Multivariate model evaluating relationship between GRB7 expression and recurrence (hazard ratios and 95% confidence intervals)

Table 3: Potential therapeutic targets in TNBC identified in this analysis.
**Figure Legends**

**Figure 1.** Natural log hazard ratio for recurrence risk as a function of quantitative *GRB7* expression without adjustment for other factors. Points in red correspond to recurrences, whereas the curved solid lines are at ±2 standard errors. There was a highly statistically significant association between quantitative expression and recurrence (P = 0.001 overall, P = 0.53 for nonlinearity.) Straight line is the estimate from the linear model. Each integer on the y-axis scale indicates a 2.72-fold increase in risk of recurrence.

**Figure 2.** Pathway analysis of genes highly expressed in TNBC compared to HR-positive, HER2-negative breast cancer. The nodes of the network are the proteins encoded by the genes overexpressed in TNBC and their shapes represent their biological activities as shown in the inset. The nodes are connected by edges which indicate direct and indirect biological relationships between these proteins. Highly connected nodes may represent therapeutic targets, inhibition of which could attenuate signaling throughout the network.

**Figure 3 a-b.** External validation showing significantly higher median *GRB7* expression levels in patients who did not have a pathologic complete response to therapy compared with those who did for the study reported by Bonnefoi et al (Figure 3a)\textsuperscript{21} and Tabchy et al (Figure 3b)\textsuperscript{22}.
References:


Table 1. Characteristics of triple negative patient population included in this analysis

<table>
<thead>
<tr>
<th></th>
<th>No.</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td>246</td>
<td></td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 45 years</td>
<td>101</td>
<td>41%</td>
</tr>
<tr>
<td>45-65 years</td>
<td>128</td>
<td>52%</td>
</tr>
<tr>
<td>&gt; 65 years</td>
<td>17</td>
<td>7%</td>
</tr>
<tr>
<td><strong>Central grade</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Well / Mod</td>
<td>25</td>
<td>10%</td>
</tr>
<tr>
<td>Poor</td>
<td>221</td>
<td>90%</td>
</tr>
<tr>
<td><strong>Pos Axillary Nodes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 positive</td>
<td>200</td>
<td>81%</td>
</tr>
<tr>
<td>1 positive</td>
<td>29</td>
<td>12%</td>
</tr>
<tr>
<td>2-3 positive</td>
<td>17</td>
<td>7%</td>
</tr>
<tr>
<td><strong>Tumor size</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤=2cm</td>
<td>116</td>
<td>47%</td>
</tr>
<tr>
<td>&gt;2, ≤=5 cm</td>
<td>120</td>
<td>49%</td>
</tr>
<tr>
<td>&gt;5cm</td>
<td>10</td>
<td>4%</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>214</td>
<td>87%</td>
</tr>
<tr>
<td>Black</td>
<td>27</td>
<td>11%</td>
</tr>
<tr>
<td>Other</td>
<td>5</td>
<td>2%</td>
</tr>
</tbody>
</table>
Table 2. Multivariate model evaluating relationship between GRB7 expression and recurrence (hazard ratios and 95% confidence intervals)

<table>
<thead>
<tr>
<th></th>
<th>Model I</th>
<th>Model II</th>
<th>Model III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age ≤ 45 vs. &gt;65</td>
<td>0.61 (0.21,1.79)</td>
<td>0.49 (0.17,1.42)</td>
<td>0.49 (0.17,1.46)</td>
</tr>
<tr>
<td>Age 45-65 vs. &gt;65</td>
<td>0.81 (0.29,2.26)</td>
<td>0.67 (0.25,1.84)</td>
<td>0.63 (0.22,1.78)</td>
</tr>
<tr>
<td>Nodes 1 vs. 0</td>
<td>2.37 (1.39,4.03)</td>
<td>2.04 (1.17,3.57)</td>
<td>2.27 (1.32,3.92)</td>
</tr>
<tr>
<td>Nodes 2-3 vs. 0</td>
<td>1.83 (0.83,4.05)</td>
<td>1.57 (0.65,3.83)</td>
<td>1.96 (0.86,4.47)</td>
</tr>
<tr>
<td>Grade Poor vs. Mod/Well</td>
<td>1.53 (0.56,4.19)</td>
<td>1.62 (0.53,4.95)</td>
<td>1.43 (0.51,3.98)</td>
</tr>
<tr>
<td>Tumor size &gt;2 vs. ≤ 2 cm</td>
<td>1.93 (1.09,3.41)</td>
<td>1.97 (1.10,3.55)</td>
<td>1.95 (1.09,3.47)</td>
</tr>
<tr>
<td>GRB7 x+2 vs. x</td>
<td>3.41 (1.78, 6.53)</td>
<td>p=0.002</td>
<td></td>
</tr>
<tr>
<td>GRB7 High vs. Low</td>
<td></td>
<td>2.31 (1.30,4.11)</td>
<td>p=0.004</td>
</tr>
</tbody>
</table>
Table 3. Potential targets in triple negative breast cancer identified in this analysis.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein Name</th>
<th>Protein Function</th>
<th>Drugs</th>
</tr>
</thead>
<tbody>
<tr>
<td>AURKB</td>
<td>Aurora Kinase B</td>
<td>Binds microtubule K fibers near kinetochore</td>
<td>AZD1152, VX-680, AT9283</td>
</tr>
<tr>
<td>PLK1</td>
<td>Polo-like Kinase 1</td>
<td>Regulates G2/M transition</td>
<td>BI6727, ON01910</td>
</tr>
<tr>
<td>KIFC1</td>
<td>Kinesin Family Member C1</td>
<td>Microtubule motor activity</td>
<td>ARRY-250, ispinesib, SB743921</td>
</tr>
<tr>
<td>CHEK1</td>
<td>CHK1 Checkpoint Homologue</td>
<td>Regulates G2/M checkpoint</td>
<td>AZD7762, PF0477736</td>
</tr>
<tr>
<td>FOXM1</td>
<td>Forkhead Box M1</td>
<td>G1–S and G2–M cell cycle phase progression; mitotic spindle</td>
<td>Siomycin A</td>
</tr>
<tr>
<td>CDC2</td>
<td>CDK1</td>
<td>G1-S and G2M cell cycle phase progression</td>
<td>Flavopiridol</td>
</tr>
</tbody>
</table>
Figure 3a.

Figure 3b.

P = 0.0397

P = 0.0044