Systematically defining single gene determinants of response to neoadjuvant chemotherapy reveals specific biomarkers

Agnieszka K. Witkiewicz*, Uthra Balaji, Erik S. Knudsen*

Simmons Cancer Center, UT Southwestern Medical Center, Dallas TX

*For correspondence:
Agnieszka K. Witkiewicz, MD
Department of Pathology
UT Southwestern Medical Center
Dallas TX
Ph: 214-648-5149
E-mail: agnes.witkiewicz@utsouthwestern.edu

Erik Knudsen, PhD
Department of Pathology
UT Southwestern Medical Center
Dallas TX
Ph: 214-648-4115
E-mail: erik.knudsen@utsouthwestern.edu
TRANSLATIONAL RELEVANCE:

Currently, there are no clinically employed markers to define patients that will benefit from neoadjuvant chemotherapy. Here an unbiased systematic approach was used to define pathways and specific markers associated with the response to neoadjuvant chemotherapy in breast cancer. These analyses revealed that genes involved in cell cycle control processes that are regulated by the RB/E2F pathway are significantly associated with response to chemotherapy in both ER-positive and ER-negative breast cancer. However, additional genes were identified that were predictive of response, particularly across different therapeutic regimens. Importantly, identified genes associated with pathological complete response or residual disease were evaluated in independent cohorts by gene expression and immunohistochemistry demonstrated strong predictive power. Together these data suggest that a relatively small number of biomarkers can be identified to predict response to neoadjuvant chemotherapy.
ABSTRACT:

**Purpose:** We sought to systematically define determinants of the response to neoadjuvant chemotherapy to elucidate predictive biomarkers for breast cancer.

**Experimental Design:** An unbiased systematic analysis was performed in multiple independent datasets to define genes predictive of complete pathological response (pCR) following treatment with neoadjuvant chemotherapy. These genes were interrogated across ER-positive and ER-negative breast cancer and those in common across three different treatment regimens were analyzed for optimal predictive power. Subsequent validation was performed on independent cohorts by gene expression and immunohistochemical analyses.

**Results:** Genes that were highly associated with the response to neoadjuvant chemotherapy in breast cancer were readily defined using a computational method ranking individual genes by their respective receiver operating characteristics. Such predictive genes of the response to taxane associated therapies were strongly enriched for cell cycle control processes in both ER-positive and ER-negative breast cancer and correlated with pCR. However, other genes that were specifically associated with residual disease were also identified under other treatment conditions. Using the intersection between treatment groups, nine genes were identified that harbored strong predictive power in multiple contexts and validation cohort. In particular the nuclear oncogene DEK was strongly associated with pCR, whereas the cell surface protein BCAM was strongly associated with residual disease. By immunohistochemical staining these markers exhibited potent predictive power that remained significant in multivariate analysis.

**Conclusion:** Systematic computational approaches can define key genes that will be able to predict the response to chemotherapy across multiple treatment modalities yielding a small collection of biomarkers that can be readily deployed by immunohistochemical analyses.
INTRODUCTION:

While breast cancer is treated with a variety of targeted agents, conventional cytotoxic chemotherapy remains a mainstay of therapy (1-4). At present, complex chemotherapy regimens are applied in multiple distinct clinical scenarios in the treatment of breast cancer. It is well appreciated that triple negative breast cancer is treated largely exclusively with chemotherapy (2, 5, 6); however, other forms of breast cancer are also treated with chemotherapy. For example, luminal B breast cancer is often treated with adjuvant chemotherapy in conjunction with ER targeted therapeutics (7-10). Similarly, Her2 positive cancers are treated with Trastuzumab in conjunction with taxane based chemotherapy (11). In all of these contexts it is critically important to elucidate determinants of the response to chemotherapy.

One means to evaluate the response to chemotherapy in clinical specimens involves the analyses of the response to neoadjuvant chemotherapy (2, 12, 13). While historically surgery has preceded treatment with adjuvant therapy, there has been a significant increase in neoadjuvant therapy (18, 19). Studies have shown that the response to neoadjuvant therapy is effective at predicting the ultimate course of tumor behavior and specific determinants of that response are being sought (2, 12, 14, 15). Importantly, pathological response in neoadjuvant studies reveals tumor response to a given therapy independent of other prognostic features of disease, and therefore markers defined in the analyses of neoadjuvant treatment could be inferred to portend activity in the adjuvant setting as well.

Several studies have analyzed the gene expression programs associated with response to neoadjuvant chemotherapy (14-18). Our group and others have analyzed specific gene expression programs associated with response to chemotherapy. These studies have indicated that gene expression programs involved in RB/E2F biology or proliferation associated properties are associated with pathological complete response (17-20). In contrast, others have used data...
sets to infer predictive markers using supervised computational approaches (14, 15, 21, 22). Here we sought to use a simple method to identify individually predictive genes that can be used singly or in combination across chemotherapy regimens and disease subtypes that could be used to direct therapy. These small number of genes returned by such a method can be individually analyzed by immunohistochemistry or other methods that are readily amenable to clinical utilization.
MATERIALS AND METHODS

Datasets: Raw .CEL files and platform annotations for gene expression datasets GSE20194, GSE20271, GSE22093, GSE23988, and GSE25066, GSE41998, GSE2226 were downloaded from the Gene Expression Omnibus (GEO). A comprehensive summary of the cohorts and related citations are provided in the supplemental data (Supplemental Table 4). Datasets were normalized by the robust multiarray average algorithm (RMA) from limma Bioconductor packages in R. We assumed that the probes annotation supplied by the array manufacturer were accurate. For genes with multiple probe sets, we averaged the gene expression levels. Patients without response information will be excluded from analysis. We pooled the neoadjuvant patients from the GEO datasets GSE20194, GSE20271, GSE22093, GSE23988, and GSE25066 to develop the “NEO” dataset. The datasets GSE41998 and GSE2226 were employed for independent validation (Supplemental Table 4).

ROC screening: The area under receiver operator characteristic curve (AUC) was used to screen the genes according to their ability to distinguish between two phenotypes, and the AUC was calculated by R-package ROCR. We ran ROCR for all the genes in the microarray from different treatment types (TA, TFAC and FAC), different subtypes (ER-/+) and pooled NEO dataset. We ranked genes from different categories based on the patients who either had a complete response (pCR) or retained residual disease (RD) after chemotherapy. The ranking of the genes was performed for genes predictive of prognosis.

Gene module summarization: To study associations between gene signature modules and clinical responses, we summarized the signatures to a single feature for classification. This approach was applied to the RB-signature, Genome Grade Index, Mammaprint, OncotypeDx, and CIN70 gene signatures. We averaged the group of genes expression values, which were
then used as a feature for ROC screening. For the analysis of the Theraprint genes, single genes were analyzed for their association with clinical outcome by evaluating their ROC characteristics individually.

**UTSW Cohort and Immunohistochemical Staining:** A cohort of 74 consecutive cases for which tissue was available, and who were treated with neoadjuvant chemotherapy between 2010-2012 were employed for immunohistochemical validation analysis. Clinical-pathological features of the cohort are summarized in Supplementary Table 3. Immunohistochemical stains for DEK (Santa Cruz, cat #: sc-30213, dil 1:200) and BCAM (Santa Cruz, cat #: sc-46795, dil 1:100) were performed on pretreatment biopsy. The stains were performed using a BenchMark XT stainer (Ventana Medical Systems).
RESULTS

Cell cycle regulatory genes are potent predictors of response to neoadjuvant chemotherapy: To define the key determinants of chemotherapy response data from several treatment groups were utilized. Because most patients with Her2-positive disease are also treated with trastuzumab (23), such patients were removed from the analysis; thus, these cohorts represent ER-positive disease Her2-negative breast cancer and triple negative breast cancer. Gene expression data of pre-treatment biopsies from patients treated with taxane and anthracycline (TA) and taxane, 5-fluorouracil, anthracycline, and cytoxan (TFAC) were evaluated (Cohort information, Supplemental Table 1). Recognizing that each individual gene could have some intrinsic predictive value, a simple method of ranking each gene based on its receiver operating curve characteristic was employed in individual datasets (Figure 1). This analysis provided a rank ordering of individual genes associated with response (Figure 1A). The top 20 predictive AUC values were comparable between the two data sets (0.73-0.79 vs. 0.75-0.81). Interestingly, although the TFAC and TA datasets are independent, the same gene, the nuclear oncogene DEK, was found to have the top predictive power (Figure 1A). Gene ontology analysis of the top 150 genes by AUC ranking demonstrated that in both TA and TFAC there was a significant over-representation of genes associated with cell cycle (Figure 1B). Notably, many of these genes are regulated by the RB/E2F pathway that has been independently associated with response to neoadjuvant chemotherapy (24, 25). Consistent with these findings, analysis of multiple signatures associated with cell cycle genes including the GGI signature (26), the CIN70 signature (27), the OncotypeDx proliferation genes (28), Mammaprint (70 gene signature) (29), and the RB-signature (30) all exhibited potent predictive values in the TA data set and TFAC data set (Figure 1C and Supplemental Figures S1 and S2). Unsupervised analyses demonstrated a clear association of a high-signature value with pCR
(Figure 1C). Additionally, each signature had potent ROC characteristics in the datasets (Figure 1D). Thus, these findings reinforce the concept that cell cycle regulated genes can have profound influence related to the response to neoadjuvant chemotherapy in breast cancer.

**Distinct predictive genes emerge from different treatment cohorts:** Interestingly, in the analysis of another cohort of patients, treated with 5-fluorouracil, anthracycline, and Cytoxan (FAC) a completely distinct cadre of genes with top AUC was observed (Figure 2A). In this cohort, while the range of top AUC values were similar to those in TA and TFAC, there were surprisingly very few genes in common. Additionally, in the analysis of gene ontology there was no enrichment for cell cycle associated processes as was observed in the TA and TFAC data (Figure 2B). In part this is because the majority of the high-ranked AUC genes are associated with RD as opposed to pCR as is observed with cell cycle genes in the TA and TFAC cohorts (Figure 2C). These data were recapitulated at the single gene level. For example, the top performing gene in the FAC dataset (IFI16) had little predictive value in TA/TFAC (Figure 2C). In contrast, TTK was highly predictive in TA and TFAC but not FAC (Figure 2D). These data suggest that it may be particularly challenging to define biomarkers of response that would be useful across multiple manifestations of disease subtypes and treatment approaches.

**Identification of genes that are predictive of response to chemotherapy in both ER-positive and ER-negative breast cancer:** While both ER-positive and ER-negative breast cancers are treated with similar chemotherapy regimens in the neoadjuvant setting, there is clearly a distinction in the relative response between these two forms of disease (6, 20, 31). Therefore, we initially evaluated determinants that may be specific for ER-positive vs. ER-negative breast cancer. To this end, we combined all treatment groups to build a large single data set (Supplemental Table 1). In this cohort of 994 cases, cell cycle associated genes such
as DEK, ANP32E, and MCM3 were particularly potent markers of response (Figure 3A). Consistent with this analysis, gene ontology demonstrated the corresponding enrichment for terms associated with cell cycle control (Figure 3A). To interrogate genes selective to ER-positive, ER-negative, and either subtype of breast cancer simple AUC ranking was applied to each subtype independently (Figure 3B). Interestingly, in both cases cell cycle associated genes were still highly represented, although in ER-negative DNA replication and DNA repair terms illustrated the highest representation, while in ER-positive disease mitosis related processes were highly over-represented (Figure 3B). Interestingly, when established cell cycle signatures were evaluated in these cohorts based on ER-status they exhibited differential predictive power (Supplement Figure S3). For example, the GGI and OncotypeDx proliferation signatures performed best in ER-positive cases, while the RB-signature performed best in ER-negative cases. The differential behaviors of specific genes was clearly apparent in the analysis of single genes (Figure 3C and 3D), ILF2 is particularly relevant for ER-negative cancer, while ASPM is relevant largely for ER-positive breast cancer. To approach the question of overlap between these groups we simply evaluated the union of the top performing 150 genes in each subtype. Surprisingly only 19 genes were in common between ER-positive and ER-negative breast cancer and of these genes 18 were associated with pCR (Figure 3E and 3F). As shown by GMMN, such genes were relatively effective predictors in both ER-positive and ER-negative breast cancer (Figure 3G).

Identification of genes with potent predictive power across treatment groups: To define the presence of predictive genes independent of treatment, a similar interaction analysis was performed between the TA, TFAC, and FAC datasets. Here 9 genes emerged that exhibited strong AUC values irrespective of treatment (Figure 4A). These genes were evenly distributed for their association with pCR (DEK, DONSON, LBR, YEATS) and RD (BCAM, MTRN, FOXA1, SLC22A5, ANXA9) (Figure 4B). In tertile and ROC analysis using this 9 gene signature there
was clearly predictive value of these genes (Figure 4B and 4C). The two genes with the strongest AUC associated with pCR (DEK) and RD (BCAM) were evaluated and shown to be differentially expressed in pCR vs. RD cases, and associated with outcome irrespective of treatment or ER-status (Figure 4C, D and E and Supplemental figure S4).

**Top performing predictive makers retain predictive value in independent validation cohorts with distinct therapeutic regimens:** The analysis performed supported the potential utilization of a small set of 9 markers for the prediction of therapeutic response. To validate the performance two additional cohorts were utilized, wherein patients were treated with AC and taxane or the microtubule poison ixabepilone or AC with or without taxane (Summarized in Supplemental Table 2). In these cohorts, the 9 genes identified were appropriately associated with pCR and RD as determined in the discovery cohorts (Figure 5A). Importantly the genes in these validation cohorts exhibited potent predictive value in tertile analysis and ROC analysis comparable to that observed in the discovery cohorts (Figure 5B and 5C). Thus, the small panel of genes defined in this fashion could have utility in generally predicting therapeutic response.

**Immunohistochemical validation of the robust predictive value of DEK and BCAM in an additional neoadjuvant cohort:** While molecular approaches are being progressively employed in clinical care, immunohistochemistry remains the mainstay of breast cancer biomarker analysis to guide treatment (e.g. ER-staining). Therefore, we optimized immunohistochemical staining of the two top performing markers, DEK and BCAM. DEK stains nuclei and exhibited a range of staining from negative to high staining in 100% of cells (Figure 6A). To quantify the staining a modified histo-score was utilized where the intensity (0-3) was multiplied against the percent of cells staining positive divided into quartile groups. This approach yielded a range of products from 0 to 12 that were employed in tertiles and as a
“semi-continuous” variable. BCAM exhibits membrane staining and the percent of cells exhibiting robust membrane staining was utilized both with tertile cutpoints, and as a continuous variable. These marker levels were evaluated by ROC analysis (Supplemental Figure S5). These data demonstrated potent predictive value of both markers. Using defined tertile cutpoints, there was a clear association of DEK and BCAM with pCR and RD respectively (Figure 6B). Importantly, in addition to pathologically defined pCR and RD, there is the quantitative association of response to chemotherapy as measured by residual cancer burden (RCB). The RCB is a measure of the cellularity and tumor bed of the tumor post-treatment resection. As shown, high DEK levels were strongly associated with a low RCB indicative of preferred response to chemotherapy (Figure 6C). In contrast high BCAM was associated with a high RCB indicative of a poor response to chemotherapy (Figure 6C). To evaluate whether the predictive value of DEK and BCAM would add value beyond standard pathological features of response, univariate and multi-variate statistical analyses were performed (Figure 6D). The data indicate that both BCAM and DEK provided added statistical value beyond grade and nodal status. Together these data indicate that DEK and BCAM are associated with the response to chemotherapy and could serve as important independent markers of response to neoadjuvant chemotherapy.
DISCUSSION:

The improved treatment of cancer is tied to a more targeted approach to therapy. Typically, this is viewed in the context of drugs that have a specific molecular target; however, one of the most important areas in clinical care is to improve the delivery of chemotherapy. Across breast cancer subtypes chemotherapy is routinely utilized to reduce the burden of disease in the neoadjuvant setting, or prevent disease recurrence in the adjuvant setting (2, 12). Chemotherapy can provide long-term clinical benefit in breast cancer, as patients with tumors that experience a complete pathological response to neoadjuvant chemotherapy have a particularly good long-term prognosis (7). This finding is highly relevant in the area of triple negative breast cancer, wherein a pathological complete response denotes the same prognosis as patients with ER-positive breast cancer (7). In contrast, tumors that progress while undergoing neoadjuvant chemotherapy have a poor prognosis, and are generally resistant to chemotherapy. For these reasons, it would be ideal to have a means to predict response to chemotherapy. This would allow chemotherapy to be targeted to patients whose tumors are most likely to benefit from such treatment, and consider surgery in combination with other modalities for patients that would be predicted to have an unfavorable response to treatment.

The primary objective of neoadjuvant chemotherapy is to improve surgical outcomes, and as with any systemic chemotherapy, it is used to reduce risk of distant recurrence. In principle, any tumor that is a candidate for adjuvant systemic chemotherapy could be treated with neoadjuvant chemotherapy. In breast cancer, the use of systemic adjuvant chemotherapy is largely evaluated based on disease subtype. For example, triple negative breast cancers will almost universally receive chemotherapy, while for ER-positive breast cancer the use of molecular signatures (e.g. PAM50, Mammaprint or OncotypeDx) are used to evaluate the benefit from adjuvant chemotherapy beyond endocrine therapy alone (28, 32-35). Here we identified genes that had high-predictive value in ER-positive and ER-negative cohorts. Due to the frequent use of Herceptin in Her2-positive breast cancer, those cases were excluded from
the analysis. In the consideration of ER-positive breast cancer, as may be expected proliferation associated genes were associated with therapeutic response since these genes differentiate luminal A and luminal B subtypes. In general, all proliferation signatures tested (ie. GGI, Oncotype, CIN, RB-signature) had similar activity in predicting response, although the RB-signature had marginally better performance characteristics. Interestingly, such proliferation gene signatures were also effective within ER-negative breast cancer and individual proliferation genes harbored the top AUC. Unexpectedly, the individual genes associated with response were largely distinct from those in ER-positive disease. Genes involved in mitosis (e.g. Cyclin B2, MAD2L1, UBE2C) represented top ER-positive genes, while genes involved in DNA replication and DNA repair (MCM3, MSH2, FANCL) were most predictive in ER-negative breast cancer. Interspersed within the dominant cell cycle/proliferation associated gene programs were genes that have been implicated in the response to chemotherapy. For example, LDHB has been recently reported as a determinant of chemotherapy response and was identified herein (36). In spite of the overall similarity of gene function, investigating the intersection between ER-positive and ER-negative revealed a small number of genes which maintained robust predictive power and suggested that a “general” set of markers could be identified that would be actionable within either the ER-positive or ER-negative subtypes. Interestingly, this cadre of genes was considerably over-represented for association with pCR that was reflective of the fact that the vast majority of the top predictive genes in ER-negative breast cancer are associated with pCR not RD.

Most chemotherapy used in the neoadjuvant setting represents an anthracycline in combination with a taxane (TA), cytoxan (AC), or both (ACT) (2, 12, 37). In the cohorts employed herein, there was a significant difference in the top predictive markers between the TA and TFAC cohorts vs. the FAC cohort. At present it is impossible to determine if this is a specific feature of the therapy utilized or represents some form of technical bias within the independent cohorts. However, the data from these analyses reinforces the concept that
investigating multiple cohorts and treatments is important, and defining markers that are predictive under multiple independent contexts are presumptively critical for delineating utility under clinical conditions that may be “less than ideal”. For example, the interrogation of genes specific to the FAC cohort would yield many genes that have limited predictive power in other contexts (e.g. IFI16). In the analysis of the intersection between treatment groups we defined only 9 genes that were effective in all treatment settings. Importantly we defined genes that were associated with cell cycle and pCR (e.g. DEK and DONSON) as well as genes associated with RD (e.g. BCAM and METRN). DEK is a nuclear oncogene that is implicated in DNA damage repair and apoptosis (38). LBR is the lamin B receptor, while YEATS2 and DONSON have largely unknown functions in mammalian cells. Of these genes only a subset are cell cycle regulated (LBR, DONSON, and DEK); therefore, the inclusion of the other genes ostensibly provides complementary biological information that would be expected to improve performance. Interestingly, none of these genes have been identified as being particularly relevant markers for the response to neoadjuvant chemotherapy. However, in our additional validation cohorts these markers continued to be predictive of chemotherapy response. These findings contrasted with the analyses of Theraprint (Agendia Inc.) that has been presented as a means to judge potential response to chemotherapy. The Theraprint genes, and similar targeted panels, are based on individual studies of single genes and functional associations between proteins and drugs. An example, is that the levels of ribonucleotide reductase will infer response to anti-metabolites. We evaluated all genes within Theraprint individually and found that very few of these genes harbored significant predictive power (Supplemental Figure 6). In this setting ESR1 (estrogen receptor) was the most potent predictive marker. These findings could represent the fact that neoadjuvant chemotherapy is not delivered as a single agent; or more likely, that due to the complexity of cancer simple gene inferences based on functional data do not uniformly hold true.
Over the last several years new molecular tests have emerged to guide breast cancer treatment. Most notably PAM50 has recently received approval, thereby joining OncotypeDx and MammaPrint in providing guidance to the treatment of ER-positive disease. While there is growing acceptance for the use of RNA based predictive tools, immunohistochemistry remains the key tool in the context of evaluating the treatment of breast cancer. This is because the standard of care markers evaluated on the diagnostic core biopsy are performed by IHC. In recognition of this issue, we interrogated the two strongest performing markers across treatment groups, DEK and BCAM, in pre-treatment diagnostic biopsies. The antibodies were optimized on clinical strainers and employed on a cohort of consecutive cases at UTSW. The data demonstrated that both markers harbor strong predictive value and the simple combination of the two markers was particularly effective at predicting response to therapy. In addition to pCR/RD, tumors with low DEK and high BCAM had a particularly poor RCB score. RCB is associated with long-term prognosis and is a quantitative measure of residual disease. The finding that these markers were effective in terms of the fraction of residual disease is particularly important in defining patients for whom an alternative treatment is preferable. An RCB score of 3 indicates that the tumor was largely refractory to treatment, and perhaps alternative approaches to neoadjuvant chemotherapy should be effectively considered. In spite of the robust testing here in multiple independent retrospective cohorts with multiple independent approaches, it is important to interrogate predictive power prospectively. To this end an observational study of biomarkers in the response to neoajduvant chemotherapy is ongoing.
REFERENCES:


FIGURE LEGENDS

Figure 1.  **Top predictive markers of pathological response in patients treated with taxane/anthracycline based therapies:**  A. Gene expression data was mined to define genes with the top-ranked receiver operating curve (ROC) characteristics in cases treated with TA and TFAC. The top 20 genes and their associated AUC values are shown for two cohorts. B. Gene ontology analysis was performed on the top 150 predictive genes identified in each cohort. C. The indicated gene expression signatures were employed to cluster cases based on signature expression value. The Kolmogorov-Smirnov (KS) statistic was used to determine the association of the signature value with clinical response. D. ROC curves of the indicated gene signatures that are enriched for cell cycle/proliferation associated genes.

Figure 2.  **Distinct top predictive markers in cohorts treated with anthracycline/cytoxan:**  A. Gene expression data was mined to define genes with the top-ranked receiver operating curve (ROC) characteristics in cases treated with FAC. The top 20 genes and their associated AUC values are shown for the cohort. B. Gene ontology analysis was performed on the top 150 predictive genes identified in the cohort. C. Genes predictive in the FAC cohort were employed to cluster cases based on signature expression values. The Kolmogorov-Smirnov (KS) statistic was used to determine the association of the signature value with clinical response. D. Example of a gene that is predictive in TA/TFAC cohorts, but not FAC cohort. E. Example of a gene that is predictive in FAC cohort, but not the TA/TFAC cohorts.

Figure 3.  **Definition of genes that harbor strong predictive value in discrete breast cancer subtypes:**  Gene expression data was mined to define genes with the top-ranked receiver operating curve (ROC) characteristics across all treatment groups. The top 20 genes
and their associated AUC values are shown. Gene ontology analysis was performed on the top 150 predictive genes identified in the cohort. B. The integrated cohort was sub-divided and top predictive genes unique to ER-positive and ER-negative breast cancer was determined by ROC analysis. Top 20 genes are shown for each subtype. Gene ontology analysis was performed on the top 150 predictive genes identified in each disease sub-type. C. Example of a gene that is predictive in ER-positive but not ER-negative breast cancer. D. Example of a gene is predictive in ER-negative but not ER-positive breast cancer. E. Intersection of the predictive genes in ER-positive and ER-negative sub-types and genes with the top AUC values are shown. F. The genes predictive in both ER-positive and ER-negative disease were employed to cluster cases based on signature expression value (all cases, ER-positive, and ER-negative). The Kolmogorov-Smirnov (KS) statistic was used to determine the association of the signature value with clinical response. G. Example of a gene predictive in both ER-negative and ER-positive breast cancer.

**Figure 4. Definition of genes that harbor strong predictive value across treatment groups:** A. Intersection analysis of predictive genes across 3 treatment groups reveals only 9 genes that are in common between the cohorts. Genes and their representative AUC values are shown for each cohort. B. The genes predictive in all treatment groups were employed to cluster cases based on signature expression value. The Kolmogorov-Smirnov (KS) statistic was used to determine the association of the signature value with clinical response. The association of defined tertiles with pathological response is shown. C. The ROC predictive behavior is shown based on therapeutic intervention and ER-status. D. The top performing gene associated with pCR, DEK, is shown as a single determinant of pathological response. E. The top performing gene associated with RD, BCAM, is shown as a single determinant of pathological response.
Figure 5. Validation of markers of chemotherapy response in independent cohorts. A. The 9 genes predictive of response across treatments were employed to cluster cases based on signature expression value in the two independent cohorts indicated. The Kolmogorov-Smirnov (KS) statistic was used to determine the association of the signature value with clinical response. B. The association of defined tertiles with pathological response are shown. C. ROC curves of the association between gene expression and clinical outcomes are shown.

Figure 6. Immunohistochemical analysis validate DEK and BCAM as predictive markers of response to neoadjuvant chemotherapy. A. The staining of DEK and BCAM was optimized in a clinical laboratory. Representative images of low and high staining of the markers is shown. B. Association of high, medium, and low DEK and BCAM levels were evaluated as a function of pCR/RD. C. Association of DEK and BCAM levels with the residual cancer burden (RCB). D. Univariate and multi-variate analysis of DEK and BCAM in the neoadjuvant cohort.
Figure 1

A

<table>
<thead>
<tr>
<th>Gene</th>
<th>AUC</th>
<th>TFAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEK</td>
<td>0.7856</td>
<td>0.8130</td>
</tr>
<tr>
<td>MCM5</td>
<td>0.7862</td>
<td>0.7988</td>
</tr>
<tr>
<td>MCM2</td>
<td>0.7619</td>
<td>0.7877</td>
</tr>
<tr>
<td>MCMB</td>
<td>0.7651</td>
<td>0.7792</td>
</tr>
<tr>
<td>NELK1</td>
<td>0.7214</td>
<td>0.7762</td>
</tr>
<tr>
<td>ANP32E</td>
<td>0.7509</td>
<td>0.7725</td>
</tr>
<tr>
<td>TTK</td>
<td>0.7668</td>
<td>0.7683</td>
</tr>
<tr>
<td>CENPA</td>
<td>0.7144</td>
<td>0.7625</td>
</tr>
<tr>
<td>FEN1</td>
<td>0.7444</td>
<td>0.7594</td>
</tr>
<tr>
<td>CHEK1</td>
<td>0.7441</td>
<td>0.7574</td>
</tr>
<tr>
<td>CCNA2</td>
<td>0.7432</td>
<td>0.7571</td>
</tr>
<tr>
<td>MCM3</td>
<td>0.7432</td>
<td>0.7568</td>
</tr>
<tr>
<td>SKA1</td>
<td>0.7431</td>
<td>0.7567</td>
</tr>
<tr>
<td>NFIB</td>
<td>0.7431</td>
<td>0.7542</td>
</tr>
<tr>
<td>TPX2</td>
<td>0.7427</td>
<td>0.7556</td>
</tr>
<tr>
<td>EZF5</td>
<td>0.7416</td>
<td>0.7538</td>
</tr>
<tr>
<td>KLF19B</td>
<td>0.7395</td>
<td>0.7520</td>
</tr>
<tr>
<td>AURKB</td>
<td>0.7376</td>
<td>0.7516</td>
</tr>
<tr>
<td>NASP</td>
<td>0.7348</td>
<td>0.7516</td>
</tr>
<tr>
<td>SLC7A8</td>
<td>0.7343</td>
<td>0.7513</td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th>Gene Ontology Term</th>
<th>P-value</th>
<th>Benjamini</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0007149-cell cycle</td>
<td>1.36E-09</td>
<td>4.0E-03</td>
</tr>
<tr>
<td>GO:0006952-cell cycle phase</td>
<td>7.9E-09</td>
<td>3.0E-04</td>
</tr>
<tr>
<td>GO:0006725-cell cycle process</td>
<td>7.9E-09</td>
<td>3.0E-04</td>
</tr>
<tr>
<td>GO:0006725-cell cycle process</td>
<td>7.9E-09</td>
<td>3.0E-04</td>
</tr>
<tr>
<td>GO:0015605-cell cycle</td>
<td>7.9E-09</td>
<td>3.0E-04</td>
</tr>
<tr>
<td>GO:0006725-cell cycle process</td>
<td>7.9E-09</td>
<td>3.0E-04</td>
</tr>
</tbody>
</table>

C

Oncoptype DX Proliferation Genes

RB Signature

Mammaprint Signature

D

ROC curves for different gene sets.
Figure 2

A

Gene | AUC
--- | ---
IFI16 | 0.8021
PLSCR1 | 0.7998
OSBPRL1 | 0.7896
TP53TG1 | 0.7896
RRAS | 0.7798
UBA1 | 0.7794
DECR2 | 0.7776
C6orf92 | 0.7772
PPP1R13B | 0.7753
LBR | 0.7733
CTSC | 0.7704
NEIL1 | 0.7704
SREBF1 | 0.7700
TMEM168 | 0.7688
NECAB1 | 0.7608
NF2 | 0.7637
IFTR8 | 0.7604
DOK7 | 0.7602
TRIM45 | 0.7600
CELSR1 | 0.7609

B

FAC Gene Ontology

Term | P-value | Benjamini
--- | --- | ---
GO:0009607 - organ morphogenesis | 4.00E-04 | 0.4433
GO:0046512 - organ development | 6.07E-04 | 0.3588
GO:0006010 - response to virus | 0.0010 | 0.5391
GO:0001665 - urogenital system development | 0.0016 | 0.4455
GO:0042221 - response to chemical stimulus | 0.0018 | 0.4091
GO:0048812 - reproductive structure development | 0.0023 | 0.4570
GO:0018093 - response to organic substance | 0.0027 | 0.4490
GO:0048518 - positive regulation of biological process | 0.0045 | 0.6048
GO:0007099 - plasma membrane organization | 0.0066 | 0.6208

C

Top Genes - FAC cohort

KS P = 8.83E-08

D

IFI16

E

TTK
Figure 4

AUC for Common genes

<table>
<thead>
<tr>
<th>Genes</th>
<th>TA</th>
<th>TFAC</th>
<th>FAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEK</td>
<td>0.7656</td>
<td>0.8130</td>
<td>0.7544</td>
</tr>
<tr>
<td>BCAM</td>
<td>0.7280</td>
<td>0.7313</td>
<td>0.7569</td>
</tr>
<tr>
<td>METRN</td>
<td>0.7191</td>
<td>0.7241</td>
<td>0.7647</td>
</tr>
<tr>
<td>ANXA9</td>
<td>0.7166</td>
<td>0.7569</td>
<td>0.7529</td>
</tr>
<tr>
<td>DONSON</td>
<td>0.7163</td>
<td>0.7340</td>
<td>0.7662</td>
</tr>
<tr>
<td>SLC22A5</td>
<td>0.7159</td>
<td>0.7594</td>
<td>0.7521</td>
</tr>
<tr>
<td>LBR</td>
<td>0.7017</td>
<td>0.7538</td>
<td>0.7733</td>
</tr>
<tr>
<td>FOXA1</td>
<td>0.6999</td>
<td>0.7369</td>
<td>0.7625</td>
</tr>
<tr>
<td>YEATS2</td>
<td>0.6985</td>
<td>0.7431</td>
<td>0.7412</td>
</tr>
</tbody>
</table>

Response KS $P = 9.166e-05$

B

C

D

E

Downloaded from clincancerres.aacrjournals.org on April 13, 2017. © 2014 American Association for Cancer Research.
Figure 6

Univariate Analysis

<table>
<thead>
<tr>
<th></th>
<th>OR</th>
<th>95%CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEK</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>high</td>
<td>0.033</td>
<td>0.006-0.175</td>
<td>5.87x10^{-3}</td>
</tr>
<tr>
<td>BCAM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>high</td>
<td>14.25</td>
<td>2.86-71.06</td>
<td>0.0011</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>old</td>
<td>0.961</td>
<td>0.098-9.35</td>
<td>0.973</td>
</tr>
<tr>
<td>young</td>
<td>1</td>
<td>0.344-4.34</td>
<td>0.7346</td>
</tr>
<tr>
<td>ER Status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>0.255</td>
<td>0.07-0.877</td>
<td>0.0302</td>
</tr>
<tr>
<td>Positive</td>
<td>1.25</td>
<td>0.344-4.34</td>
<td>0.7346</td>
</tr>
<tr>
<td>LN Status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2.9</td>
<td>0.801-10.398</td>
<td>0.102</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>OR</th>
<th>95%CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEK</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>high</td>
<td>0.029</td>
<td>0.004-0.192</td>
<td>0.00022</td>
</tr>
<tr>
<td>BCAM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>high</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>old</td>
<td>11.574</td>
<td>2.083-64.312</td>
<td>0.0051</td>
</tr>
<tr>
<td>young</td>
<td>0.399</td>
<td>0.024-6.516</td>
<td>0.519</td>
</tr>
<tr>
<td>ER Status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>0.398</td>
<td>0.06-2.557</td>
<td>0.33</td>
</tr>
<tr>
<td>Positive</td>
<td>6.54</td>
<td>0.045-9.428</td>
<td>0.755</td>
</tr>
<tr>
<td>LN Status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2.751</td>
<td>0.303-16.737</td>
<td>0.427</td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 &amp; 2</td>
<td>1.808</td>
<td>0.348-9.383</td>
<td>0.48</td>
</tr>
<tr>
<td>3</td>
<td>1.112</td>
<td>0.235-5.253</td>
<td>0.893</td>
</tr>
</tbody>
</table>
Clinical Cancer Research

Systematically defining single gene determinants of response to neoadjuvant chemotherapy reveals specific biomarkers

Agnieszka K Witkiewicz, Uthra Balaji and Erik Knudsen


Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-14-0885

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2014/07/22/1078-0432.CCR-14-0885.DC1

Author Manuscript
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.