

RESEARCH ARTICLE

Open Access

# A target based approach identifies genomic predictors of breast cancer patient response to chemotherapy

Robin M Hallett<sup>1</sup>, Gregory Pond<sup>2</sup> and John A Hassell<sup>1,3\*</sup>

## Abstract

**Background:** The efficacy of chemotherapy regimens in breast cancer patients is variable and unpredictable. Whether individual patients either achieve long-term remission or suffer recurrence after therapy may be dictated by intrinsic properties of their breast tumors including genetic lesions and consequent aberrant transcriptional programs. Global gene expression profiling provides a powerful tool to identify such tumor-intrinsic transcriptional programs, whose analyses provide insight into the underlying biology of individual patient tumors. For example, multi-gene expression signatures have been identified that can predict the likelihood of disease recurrence, and thus guide patient prognosis. Whereas such prognostic signatures are being introduced in the clinical setting, similar signatures that predict sensitivity or resistance to chemotherapy are not currently clinically available.

**Methods:** We used gene expression profiling to identify genes that were co-expressed with genes whose transcripts encode the protein targets of commonly used chemotherapeutic agents.

**Results:** Here, we present target based expression indices that predict breast tumor response to anthracycline and taxane based chemotherapy. Indeed, these signatures were independently predictive of chemotherapy response after adjusting for standard clinic-pathological variables such as age, grade, and estrogen receptor status in a cohort of 488 breast cancer patients treated with adriamycin and taxotere/taxol.

**Conclusions:** Importantly, our findings suggest the practicality of developing target based indices that predict response to therapeutics, as well as highlight the possibility of using gene signatures to guide the use of chemotherapy during treatment of breast cancer patients.

**Keywords:** Gene expression profiling, Breast cancer, Chemotherapy response, Gene signatures, Chemotherapy

## Background

Oncologists are faced with the challenging task of selecting the most effective therapies for individual cancer patients to achieve the best possible outcome. Indeed, the latter is the central goal of personalized cancer medicine. Trastuzumab is one of the best examples illustrating the importance of tailoring treatment to the characteristics of an individual's tumor. In the absence of patient selection only 10% of breast cancer patients

derive clinical benefit from trastuzumab treatment [1]. However, when patients are selected for trastuzumab therapy based on ERBB2/HER2 gene amplification, their response rate rises to as high as 50%. The development of global gene expression profiling technologies, such as DNA microarrays, has provided additional avenues to identify the molecular features of tumors that are associated with clinical variables, such as tumor grade or outcome. In fact, sets of genes, commonly called gene signatures, have been identified that predict patient prognosis, and are already used in various clinical settings [2-8].

Many current studies focus on identifying similar gene signatures to guide the selection of appropriate chemotherapy regimens [9-12]. However, gene signatures that

\* Correspondence: hassell@mcmaster.ca

<sup>1</sup>Department of Biochemistry and Biomedical Sciences, Centre for Functional Genomics, McMaster University, 1200 Main Street West, Hamilton, Ontario L8N 3Z5, Canada

<sup>3</sup>Centre for Functional Genomics, McMaster University, 1280 Main St West, Hamilton, Ontario L8S 4K1, Canada

Full list of author information is available at the end of the article

predict tumor sensitivity or resistance to chemotherapy are not currently clinically available. The development and clinical implementation of gene signatures that predict response to commonly used chemotherapeutic agents could facilitate selecting the most efficacious therapeutic regimen given the molecular characteristics of an individual's tumor. Furthermore, therapy-predictive gene signatures could ensure patients do not receive ineffective and potentially deleterious chemotherapeutic regimens.

Measuring the inherent chemosensitivity of a tumor can be accomplished by assessment of pathological response following neoadjuvant treatment with a given treatment regimen. In this fashion, patients in which no invasive or metastatic breast cancer can be detected following treatment are classified as having achieved complete pathological response (pCR), whereas patients that fail to achieve pCR are classified as having residual disease (RD). Importantly, neoadjuvant chemotherapy has been found to be as efficacious as chemotherapy given in the adjuvant setting, and patients who achieve complete pathological response after neoadjuvant intervention generally have an excellent probability of experiencing long-term survival [13-15]. Taken together, these data suggest that response to neoadjuvant chemotherapy (pCR/RD) provides a relevant clinical model to develop and validate gene signature based predictors of breast tumor response to chemotherapy.

We sought to test whether TOP2A and  $\beta$ -tubulin transcript expression indices could predict response to commonly used chemotherapeutic agents, as the protein products of these genes represent the respective molecular targets of commonly used anthracycline- and taxane-related drugs [16-18]. We hypothesized that such target based expression indices would provide a biologically comprehensive measurement of either TOP2A or  $\beta$ -tubulin activity in a patient's tumor, and thus its likely dependence on either of these targets. Importantly, these analyses establish an effective method for identifying predictive drug response signatures, and highlight the use of predictive gene signatures to guide the selection of anthracycline and taxane based chemotherapy regimens for breast cancer patients.

## Results

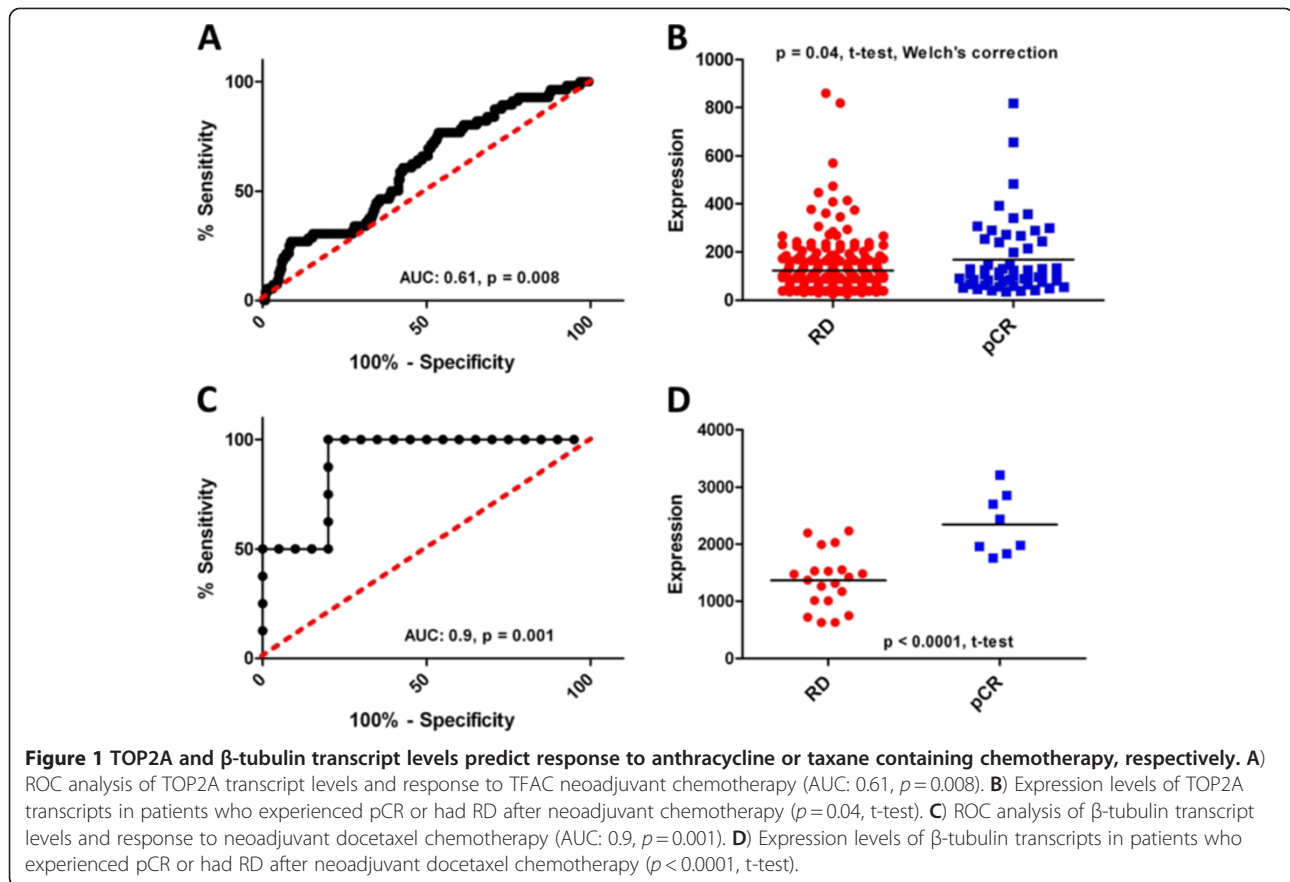
### TOP2A and $\beta$ -tubulin expression are associated with complete pathological response in breast cancer patients treated with chemotherapy

As proof-of-principle that target expression could be linked to chemotherapy response, we tested whether transcripts of TOP2A were associated with pCR in a relatively large (GSE21094,  $n = 278$ ,  $n = 56$  pCR,  $n = 222$  RD) cohort of breast cancer patients treated with a neoadjuvant chemotherapy regimen that contained an

anthracycline (TFAC; paclitaxel [T], 5-FU [F], adriamycin [A] and cyclophosphamide[C]). TOP2A transcript levels were associated with increased response to chemotherapy (Figure 1A, AUC: 0.61,  $p = 0.008$ ), and the mean expression of TOP2A transcripts was higher in patients who showed a complete pathological response than those who displayed residual disease after treatment (Figure 1B, RD: 122.5, pCR: 167.7,  $p = 0.04$ , t-test, Welch's correction). Importantly, these data suggested that elevated expression levels of TOP2A transcripts are associated with response to anthracycline based chemotherapy regimens, and are consistent with previous studies that link TOP2A expression with response to anthracycline therapy [19-21]. To extend these findings, we also tested whether  $\beta$ -tubulin expression was associated with pCR in a cohort of breast cancer patients treated with docetaxel.  $\beta$ -tubulin transcript levels were linked with increased response to docetaxel (Figure 1C, AUC: 0.9,  $p = 0.001$ ), and the mean expression of  $\beta$ -tubulin transcripts was higher in complete responders than those whose tumors did not achieve a complete response (Figure 1D, RD: 2344, pCR: 1364,  $p < 0.0001$ , t-test). Indeed, these findings are consistent with previous studies that link  $\beta$ -tubulin expression with sensitivity to taxane chemotherapy [22]. In total, these experiments suggest that target expression is associated with response to their chemotherapeutic agent counterparts.

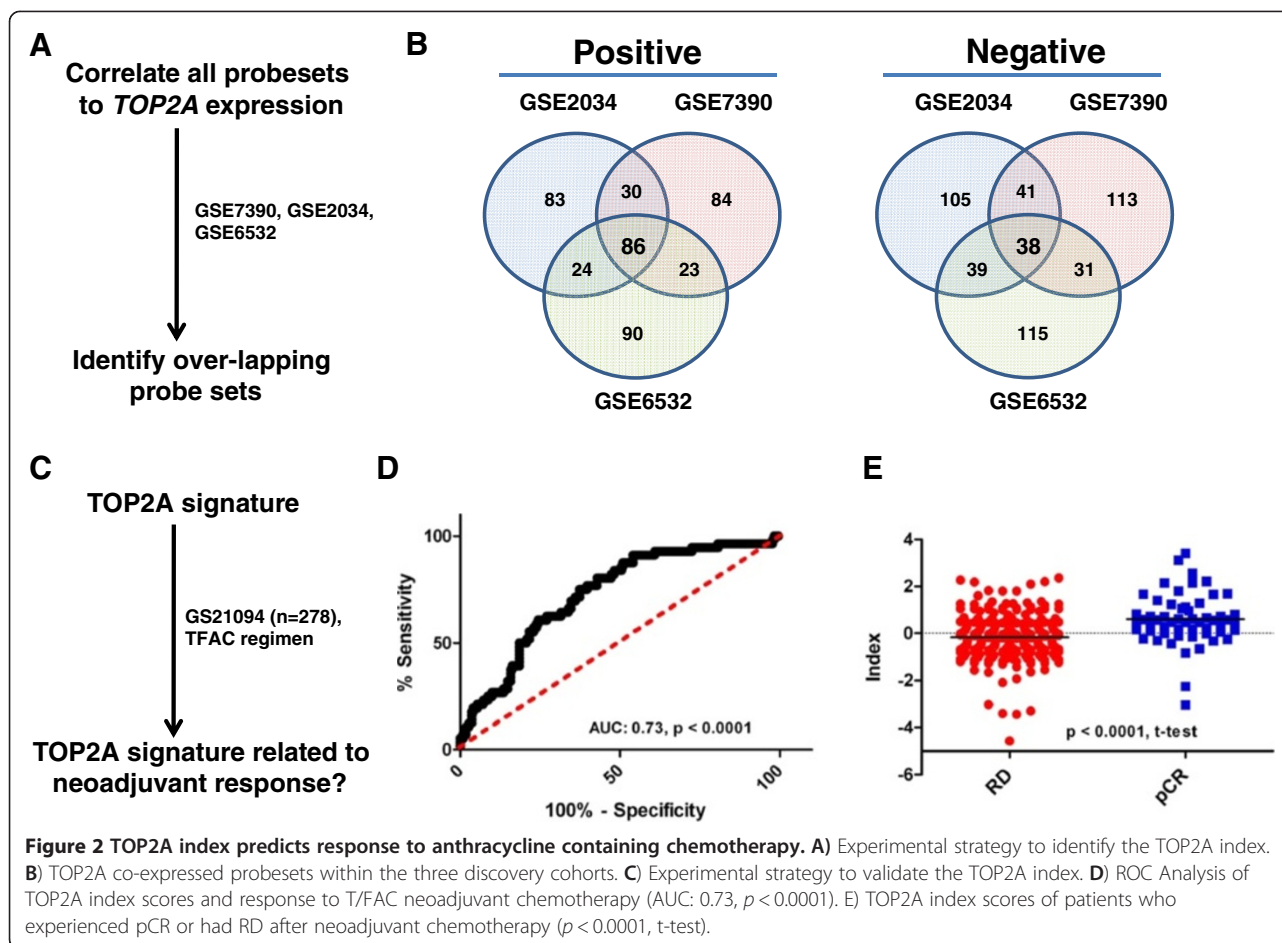
### The TOP2A index is associated with complete pathological response in breast cancer patients treated with anthracyclines

Our previous observations suggested the expression of transcripts encoding the protein targets of therapeutic agents was associated with response to these same therapies. To extend these findings, we calculated a TOP2A transcript expression index by identifying genes whose expression displayed either positive or negative correlation to the expression of TOP2A transcripts in 3 independent publicly available data sets (GSE2034, GSE7390, GSE6532) (Figure 2A). In short, we measured the correlation of all probe sets to TOP2A expression and selected those that ranked among the top and bottom 1% of all probe sets in each data set. The latter approach identified 124 TOP2A associated probe sets, 86 that displayed positive association to TOP2A transcripts and 34 that were negatively associated with TOP2A transcripts (Figure 2B, Additional file 1: Table S1). Using the same 278 patient cohort we employed previously, the TOP2A index was significantly associated with pCR (Figure 2D, AUC: 0.73,  $p < 0.0001$ ) and the TOP2A index scores were higher in patients who demonstrated a pCR than those who retained RD after treatment (Figure 2E, RD: -0.15, pCR: 0.60,  $p < 0.0001$ , t-test). Notably, the TOP2A index included genes linked to DNA repair, including



EXO1[23], ERCC6L[24,25], and RAD51[26], suggesting that the TOP2A index has a functional connection to the mechanisms of action of anthracycline drugs. To look for similarities between the TOP2A index and other predictors of chemotherapy response, we examined whether TOP2A index probes were also present within the DLDA30 predictor. Importantly, the DLDA30 is a validated gene expression based predictor of response to taxanes and anthracyclines, as well as to 5-fluorouracil and cyclophosphamide [9,10,12]. Interestingly, only one probe set from the TOP2A index also comprised an element in the DLDA30 predictor, suggesting minimal overlap between these two gene expression based predictors. Taken together, these data provide proof-of-principle that a target based expression index can predict response to small molecules that inhibit the activity of the given target under investigation. Importantly, we show here that the TOP2A index was associated with complete pathological response in breast cancer patients treated with an anthracycline containing chemotherapy regimen in the neoadjuvant setting. A  $\beta$ -tubulin index is associated with complete pathological response in breast cancer patients treated with taxanes. Given the development of the TOP2A index, we explored the possibility of identifying and developing a similar index to predict

response to taxane drugs (Figure 3A). In a similar fashion as reported above, we identified an expression index of  $\beta$ -tubulin associated transcripts.  $\beta$ -tubulin is the molecular target of taxane drugs, a commonly used class of chemotherapeutic drugs that includes docetaxel and paclitaxel (Taxotere<sup>®</sup> and Taxol<sup>®</sup> respectively)[27-29]. In short, we identified 42  $\beta$ -tubulin transcript-associated probe sets, 28 that displayed positive association with  $\beta$ -tubulin transcripts and 14 that were negatively associated with  $\beta$ -tubulin transcript levels (Figure 3B). We evaluated the association between the  $\beta$ -tubulin index and response to docetaxel using a 14 patient (profiled in replicate,  $n = 28$ , RD: 20, pCR: 8) cohort that was treated with neoadjuvant docetaxel (Figure 3C). The  $\beta$ -tubulin index was associated with complete pathological response (Figure 3D, AUC: 0.89,  $p = 0.002$ ) and  $\beta$ -tubulin index scores were higher in patients who achieved complete pathological response than those who experienced residual disease after docetaxel therapy (Figure 3E, RD: -0.5, pCR: 1.3,  $p < 0.0001$ , t-test). Interestingly, the  $\beta$ -tubulin index included genes linked to cytoskeleton processes, including TPX2[30], and DBN1 [31], suggesting a linkage between the  $\beta$ -tubulin index and the mechanism of action of taxanes. As completed previously, we also looked for similarities between the

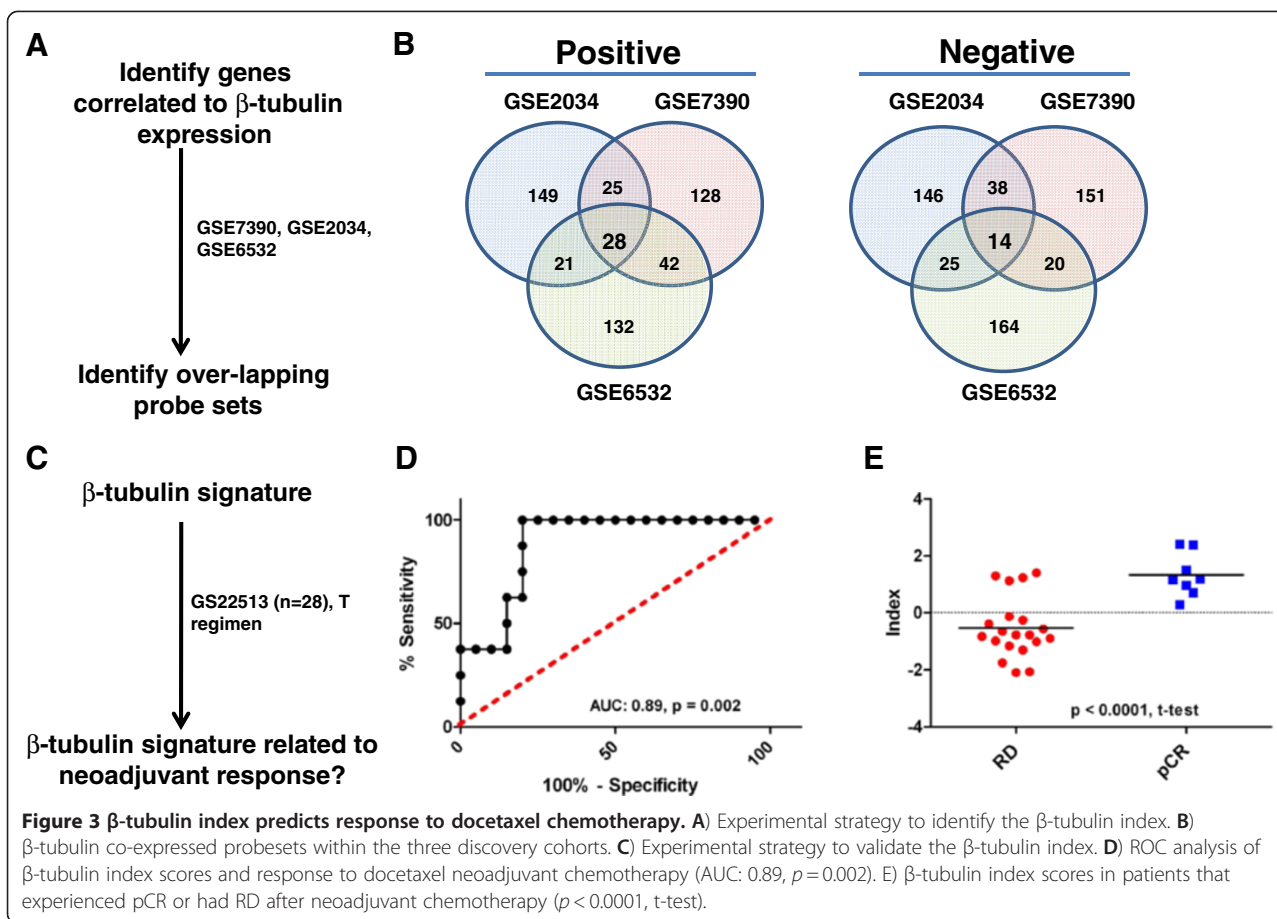


$\beta$ -tubulin index and the DLDA30 predictor. Notably, no probe sets from the  $\beta$ -tubulin index comprised an element within the DLDA30 predictor, suggesting minimal overlap between these two gene expression based predictors. In total, these data provide evidence that target based expression indices can predict response to neoadjuvant chemotherapy. In each case, the area under curve (AUC) was significantly greater than 0.5, confirming the predictive capacity of the indices, as well as the overall validity of the approach.

#### Combining indices is predictive of response to multi-agent chemotherapy

Neoadjuvant chemotherapy generally comprises multiple different chemotherapeutic agents. To determine whether we could combine our individual target indices and predict response to multi-agent chemotherapy we tested the TOP2A and  $\beta$ -tubulin indices as a combination index with the 278-patient cohort (TFAC), as well as with two different cohorts of patients treated with neoadjuvant therapy comprising an anthracycline and a taxane (AT) (GSE25055 [n = 310] & GSE25065 [n = 198]). The individual TOP2A and  $\beta$ -tubulin indices were associated with response among

the TFAC-treated patients, and patients who experienced pCR had significantly higher individual index scores than those who retained residual disease (Figure 4A,  $p < 0.0001$  both cases, t-test). After combining the indices for the 278 TFAC-treated patient cohort, we observed a significant association between the combined indices and tumor response (Figure 4B, AUC: 0.76,  $p < 0.0001$ ), and the combined index scores were significantly higher in responders (pCR) than non-responders (RD) (Figure 4C, RD: -0.4, pCR: 1.7,  $p < 0.0001$ , t-test). Interestingly, the AUC value for the combined TOP2A and  $\beta$ -tubulin index was nominally higher (AUC: 0.78) than that produced by the TOP2A index alone (Figure 2D, AUC: 0.73). We also tested the TOP2A and  $\beta$ -tubulin combination index on an additional two independent cohorts of patients treated with AT neoadjuvant chemotherapy (GSE25055 [n = 310] & GSE25065 [n = 198]). In the 310-patient cohort the individual TOP2A and  $\beta$ -tubulin index scores were significantly higher in responders than non-responders (Figure 4D,  $p < 0.0001$  both cases, t-test), the combination of the TOP2A and  $\beta$ -tubulin indices were associated with increased response to treatment (Figure 4E, AUC: 0.76,  $p < 0.0001$ ), and the TOP2A/ $\beta$ -tubulin combination index scores

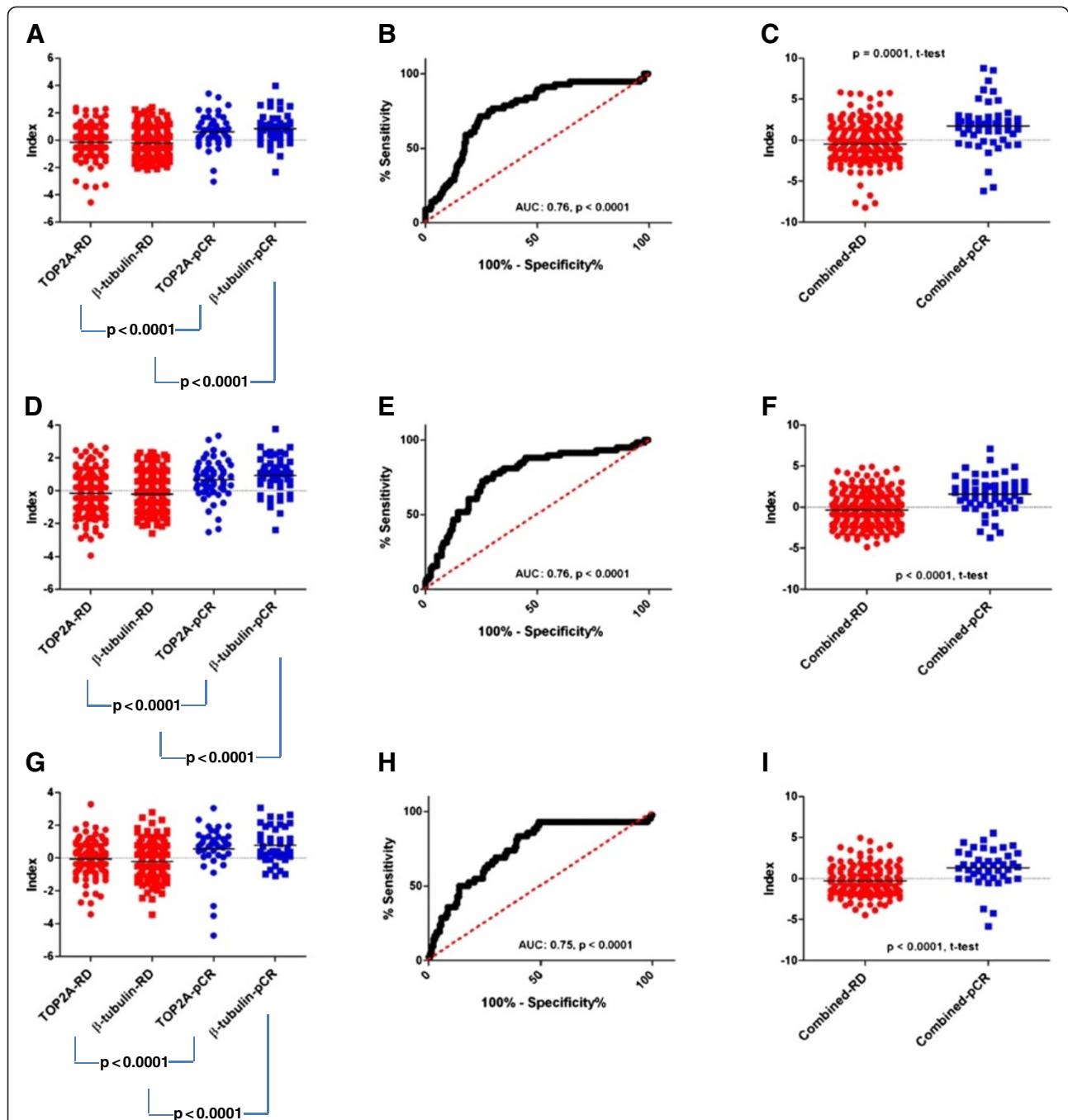


were higher in responders than non-responders (Figure 4F, RD:  $-0.4$ , pCR:  $1.6$ ,  $p < 0.0001$ , t-test). We made similar observations in the 198-patient cohort, where individual TOP2A and  $\beta$ -tubulin index scores were significantly higher in responders than in non-responders (Figure 4G,  $p < 0.0001$  both cases, t-test), and the combination index score was associated with complete pathological response (Figure 4H, AUC:  $0.75$ ,  $p < 0.0001$ ). Again, combined index scores were higher in responders than in non-responders (Figure 4I, RD:  $-0.3$ , pCR:  $1.3$ ,  $p < 0.0001$ , t-test). Taken together these data reveal the feasibility of combining individual target indices to predict response to multi-agent chemotherapy regimens.

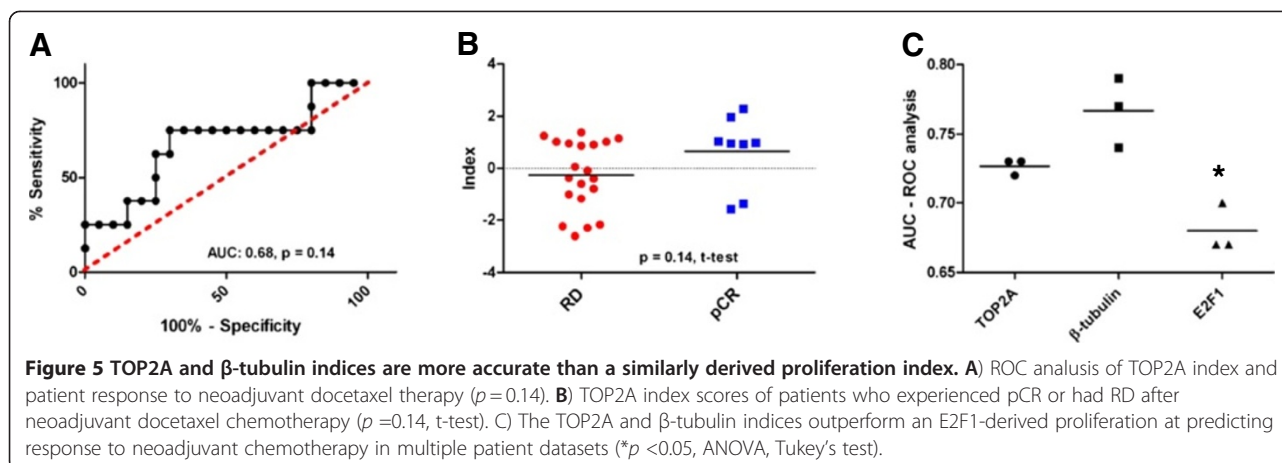
#### TOP2A and $\beta$ -tubulin indices are more accurate than a similarly derived proliferation index

Previous work suggests that the prognostic/predictive power of many breast cancer gene expression signatures is derived from their capacity to measure proliferation [2,4,7,32-36]. Hence, it is possible that the TOP2A and  $\beta$ -tubulin indices described here predict response based on their capacity to measure proliferation rather than providing a biologically relevant measurement of target.

This is particularly relevant to the TOP2A index, as the protein product of the TOP2A gene is directly involved in DNA synthesis. To address this issue we first tested the capacity of the TOP2A index to predict response in the docetaxel-only treated cohort. In this fashion, the capacity of the TOP2A index to predict response to docetaxel would suggest that the predictive capacity of the TOP2A index is not target specific and may measure a more general phenomenon related to chemotherapy sensitivity, such as proliferation. Within this patient cohort, the TOP2A index was not significantly associated with patient response to docetaxel therapy (Figure 5A&B, AUC:  $0.68$ ,  $p = 0.14$ , RD:  $-0.26$ , pCR:  $0.65$ ,  $p = 0.12$ , t-test). Indeed, these results suggest that the predictive capacity of the TOP2A index is based on measurement of target rather than proliferation. We also generated a 'proliferation index' built using the same methodology as the TOP2A and  $\beta$ -tubulin indices, around the well characterized proliferation gene E2F1 [8,37-41]. As expected, the E2F1 index was related to chemotherapy response in the 3 large cohorts of breast cancer patients 3 tested previously (GSE21094, GSE25055, GSE25065) (Additional file 2: Figure S1). Importantly, these results indicate the gene expression signatures that measure



**Figure 4 TOP2A and  $\beta$ -tubulin combination index predicts response to multi-agent anthracycline/taxane chemotherapy.** **A)** Individual target index scores (A: TOP2A index, T:  $\beta$ -tubulin index) in patients who experienced complete pathological response or had residual disease after TFAC neoadjuvant chemotherapy (GSE21094,  $p < 0.0001$  both cases, t-test). **B)** ROC analysis of combination index scores and response to TFAC neoadjuvant chemotherapy (AUC: 0.76,  $p < 0.0001$ ). **C)** Combination index scores in patients who experienced pCR or had RD after neoadjuvant chemotherapy ( $p < 0.0001$ , t-test). **D)** Individual target index scores of patients who experienced pCR or had RD after AT neoadjuvant chemotherapy (GSE25055,  $p < 0.0001$  both cases, t-test). **E)** ROC Analysis of combination index scores and response to AT neoadjuvant chemotherapy for GSE25055 (AUC: 0.78,  $p < 0.0001$ ). **F)** Combination index scores of patients who experienced pCR or had RD after AT neoadjuvant chemotherapy ( $p < 0.0001$ , t-test). **G)** Individual target index scores of patients who experienced pCR or had RD after AT neoadjuvant chemotherapy (GSE25065,  $p < 0.0001$  both cases, t-test). **H)** ROC analysis of combination index scores and response to AT neoadjuvant chemotherapy for GSE25065 (AUC: 0.75,  $p < 0.0001$ ). **I)** Combination index scores of patients who experienced complete pathological response or had residual disease after AT neoadjuvant chemotherapy ( $p < 0.0001$ , t-test).



proliferation are associated with patient response to chemotherapy. However, the performance of the E2F1 predictor was inferior to either the TOP2A and  $\beta$ -tubulin predictors (Figure 5C,  $*p < 0.05$ , ANOVA, Tukeys). Taken together, these data suggest that the predictive capacity of the TOP2A index is specific to anthracycline drugs, and moreover, that the TOP2A and  $\beta$ -tubulin indices are superior to gene signatures that measure proliferation.

#### Comparison of the TOP2A and $\beta$ -tubulin indices with clinico-pathologic parameters

Reported genomic tests for response to chemotherapy have generally failed to outperform the predictive capacity of standard clinic-pathological measurements [10]. In this regard we tested whether the TOP2A/ $\beta$ -tubulin combination index was related to tumor response after adjusting for standard clinical pathological variables. Combination of the two previously used data sets (GSE25055 & GSE25065, 20 patients discarded as pCR/RD data was unavailable) yielded a data set comprising 488 patients for the analysis. In a univariate analysis that included estrogen receptor status, tumor grade, nodal status, patient age, as well as the TOP2A/ $\beta$ -tubulin combination index score, only estrogen receptor status (AUC: 0.68,  $p < 0.001$ ), tumor grade (AUC: 0.69,  $p < 0.001$ ) and the combination index score (Range [-5.829, 7.120], AUC: 0.76,  $p < 0.001$ ) were found to be statistically significantly related to tumor response, whereas age and nodal status were not (Table 1). In a multivariate model (AUC: 0.78), the TOP2A/ $\beta$ -tubulin combination index score remained statistically significantly related to tumor response ( $p < 0.001$ ) as well as estrogen receptor status ( $p = 0.014$ ) and grade ( $p = 0.016$ ) (Table 1). In the multivariate model the odds ratio for the TOP2A/ $\beta$ -tubulin index was 1.33, which indicated that for each unit increase (i.e. from -3 to -2, from 0 to 1, or from 4 to 5, etc) in the AT index, there was a 1.33

times increase in the odds of a patient experiencing complete pathological response. Importantly, these data suggest that the TOP2A/ $\beta$ -tubulin combination index score was related to patient response to chemotherapy even after adjusting for standard clinical-pathological variables.

#### Discussion

Here we describe the identification of TOP2A and  $\beta$ -tubulin transcript expression indices that predict complete pathological response to neoadjuvant chemotherapy regimens containing anthracycline and taxane drugs. Complete pathological response represents an appropriate clinical endpoint for these studies as patients who experience pCR also experience improved survival compared to those patients who retain RD [13-15]. Notably, TOP2A or  $\beta$ -tubulin, the respective targets of anthracycline and taxane drugs, have been linked to anthracycline and taxane response in previous studies, respectively [17,18,21,42-44]. However, the expression of

**Table 1** Logistic regression analysis of the GSE25055 & GSE25065 validation set

Characteristic	Odds ratio (95% CI)	AUC	p-value
Univariate Analyses			
AT Index	1.57 (1.39, 1.77)	0.76	<0.001
Age (/10 years)	0.84 (0.68, 1.04)	0.55	0.10
ER Status Positive	0.23 (0.14, 0.37)	0.68	<0.001
Grade	5.28 (2.95, 9.46)	0.69	<0.001
Node Positive	1.16 (0.72, 1.88)	0.52	0.54
Multivariate Analyses			
Age (/10 years)	0.89 (0.70, 1.13)	0.78	0.33
ER Status Positive	0.48 (0.27, 0.86)		0.014
Grade	2.24 (1.16, 4.33)		0.016
Node	0.95 (0.54, 1.67)		0.86
AT Index	1.33 (1.14, 1.55)		<0.001

either of these genes has failed to become a useful clinical predictor of anthracycline or taxane response. We hypothesized that measurement of target-associated transcripts in a tumor sample might provide a more comprehensive measure of molecular target activity, and thus the tumor's likelihood of response to therapy. Indeed, based on the datasets explored for the studies presented here, this appears to be the case. Moreover, a combination index derived from the TOP2A and  $\beta$ -tubulin expression indices was statistically significantly related to pathological response in a multivariate model that also included age, nodal status, tumor grade and estrogen receptor status in a group of 488 patients treated with anthracycline and taxane based chemotherapy.

From a clinical standpoint, predicting response to anthracycline and taxane based chemotherapy may be useful to identify breast cancer patients who have a high likelihood of benefiting from such regimens. Conversely, patients predicted to be resistant to anthracycline- and taxane-based chemotherapy may benefit from enrollment in clinical trials investigating the efficacy of novel treatments [45]. Many issues remain to be addressed to confirm the clinical utility of the TOP2A and  $\beta$ -tubulin indices. In this study our conclusions are based on the analysis of retrospective data, which limits its clinical value. Moreover, we did not establish or optimize a threshold that would serve to separate patients predicted likely to respond to therapy from those likely to be resistant. Additionally, we did not test the capacity of the TOP2A index to predict response to neoadjuvant chemotherapy that consisted of only of an anthracycline, suggesting that the TOP2A index may be predictive of general chemotherapy response. Although we did observe that the TOP2A index was not predictive of patient response to docetaxel. Based on our multivariate analysis, our data suggests the TOP2A and  $\beta$ -tubulin indices remain predictive even after adjusting for clinical parameters such as tumor grade and estrogen receptor status, indicating that these indices likely have clinical value. Strictly speaking however, a true estimate of the usefulness of these indices would require a prospective clinical trial comparing randomly selected with index selected chemotherapy regimens.

An advantage of the approach taken here is our use of publicly available data, as well as the efficient use of patient samples for validation purposes. For example, the traditional approach for gene signature identification [2,6,7,9,46], commonly called the top-down approach, multiple datasets are required that comprise both tumor gene expression profiles as well as knowledge of the clinical variables under investigation, for the purposes of signature identification and subsequent independent validation. Other approaches, such as large-scale functional based RNA interference screens, have also yielded predictive

signatures, although these experiments are relatively labour intensive and expensive [47]. Here, we calculated target indices using datasets for which response to chemotherapy is not known. In this fashion, we maintained the independence of datasets for which response to neoadjuvant chemotherapy was measured as a clinical variable, thus maintaining the availability of multiple independent datasets for validation.

The identification of gene signatures that predict response to chemotherapy also have potential to offer new insights into the biology of breast tumors, particularly the transcriptional programs that govern therapy response. In this regard, it may be possible to identify molecular signaling pathways that either augment chemotherapy resistance or enhance sensitivity. Indeed, the latter strategy provides a rational approach to identifying new drug regimens, where a signaling pathway inhibitor/activator is included with the original chemotherapy regimen. In this fashion, tumors predicted to be therapy resistant might be rendered sensitive to the original therapy and treatment efficacy could be increased.

Another important implication of this study is that it highlights the identification of target based expression indices as a means to predict response to therapeutics. For example, it might be possible to generate a target based expression index for additional molecular targets, such as the HER2/Neu receptor tyrosine kinase, which is the molecular target of the humanized monoclonal antibody trastuzumab [1] as well as the small molecule Her2/Neu kinase inhibitor, lapatinib [48,49]. Using such an approach, therapeutic response to the latter agents might then be predicted using transcriptional target based signatures. Indeed, this approach could be tested for multiple new experimental molecularly targeted therapies.

## Conclusion

Importantly, these findings suggest the practicality of developing and testing target based indices that predict response to therapeutics. Moreover, our data highlights the possibility of using gene signatures to guide the use of chemotherapy during treatment of breast cancer patients.

## Methods

### Patients and Samples

All data was publicly available and downloaded from the gene expression omnibus (<http://www.ncbi.nlm.nih.gov/geo/>). Multiple discovery cohorts (GSE2034, GSE7390, GSE6532) were independently evaluated to determine target indices for TOP2A,  $\beta$ -tubulin, and E2F1. Together these cohorts comprised 811 patient tumor gene expression profiles derived from the Affymetrix U133A microarray platform (Table 2). Multiple validation cohorts



were independently studied to test whether target indices were related to pathological response to neoadjuvant chemotherapy (GSE21094 (T/FAC), GSE22513 (T), GSE25055 & GSE25065 (AT)). These cohorts comprised 800 patient tumor gene expression profiles from the Affymetrix U133A and Affymetrix U133 Plus 2.0 microarray platforms (Table 2). The clinical characteristics of the validation cohorts are summarized in Table 3.

The raw intensity files (.CEL) comprising each dataset were download and normalised using the Robust Multi-chip Algorithm (RMA) to generate probe set intensities [50].

### Identification of Target Related Genes

Target index genes were identified by their co-expression with either TOP2A (TOP2A, 201292\_at),  $\beta$ -tubulin (TUBB, 212320\_at), or E2F1 (204947\_at) based on a Pearson distance function [51]. We filtered these results such that only probe sets appearing in the most and least 1% of co-expressed probe sets within each identification cohort were included in the target index. The final TOP2A index comprised 86 probe sets with positive and 38 probe sets with negative correlation to TOP2A transcript levels (Additional file 1: Table S1). The  $\beta$ -tubulin index comprised 28 probe sets with positive and 14 probe sets with negative correlation to  $\beta$ -tubulin transcript levels (Additional file 1: Table S2).

To evaluate the target index, the expression values for each probe set were transformed such that the mean and standard deviation were set to 0 and 1 in each dataset, respectively. A target index was calculated for each patient as follows:

$$\frac{\sum_{i \in P} x_i}{n_P} - \frac{\sum_{i \in N} x_j}{n_N}$$

Where  $x$  is the transformed expression,  $n$  is the number of probe sets,  $P$  is the set of probes with reported positive correlation to the target probe set, and  $N$  is the set of probes with reported negative correlation to the target probe set [46,52].

**Table 3 Characteristics of the validation cohorts**

Characteristic	GSE21094		GSE25055		GSE25065		GSE22513	
	#	%	#	%	#	%	#	%
<b># patients</b>	<b>278</b>		<b>310</b>		<b>198</b>		<b>14</b>	
<b>Age, years</b>								NA
<=50	133	48	168	54	109	55		
>50	144	52	142	46	89	45		
NA	1	0						
Mean	52		50		49			
STD	11		10		11			
<b>Nodal status</b>	NA							NA
Positive			223	72	128	65		
Negative			87	28	71	35		
NA								
<b>Grade</b>	NA							NA
1			19	6	13	7		
2			117	38	63	32		
3			151	49	108	55		
NA			23	7	14	7		
<b>Response</b>								
pCR	56	20	58	19	42	21	4	29
RD	222	80	248	80	140	71	10	71
NA			4	1	16	8		
<b>ER status</b>								NA
Positive	164	49	131	42	123	62		
Negative	114	41	174	56	74	37		
NA			5	2	1	1		

Abbreviations: STD: Standard deviation; pCR: complete pathological response; RD: Residual disease. ER: Estrogen receptor; NA: Not available.

### Statistical Analysis

pCR or RD following treatment with neoadjuvant chemotherapy was used as the clinical endpoints for this study. The predictive capacities of the target indices were evaluated using receiver-operator characteristic curve (ROC) analysis and both univariate and multivariate logistic regression. T-tests were used to compare indices between responders and non-responders. Welch's correction was used when the variance of the index was

**Table 2 Summary of samples used to identify and validate target indices**

Characteristic	Discovery cohorts			Validation cohorts			
	GSE2034	GSE7390	GSE6532	GSE21094	GSE25055	GSE25065	GSE22513
Samples	286	198	327	278	310	198	14
Regimen	N/A	N/A	N/A	TFAC	AT	AT	T
							U133 Plus
Array type	U133A	U133A	U133A	U133A	U133A	U133A	2.0
	<b>Total arrays: 811</b>			<b>Total arrays: 800</b>			

Abbreviations: TFAC: paclitaxel/docetaxel, 5-fluorouracil, doxorubicin, cyclophosphamide; AT: doxorubicin, paclitaxel/docetaxel; T: docetaxel.

unequal in these two patient groups. All tests were two-sided and a p-value of 0.05 or less was considered statistically significant. ANOVA and Tukeys multiple comparison test was used to test for differences between multiple groups ( $n > 2$ ), and p-values of 0.05 or less were considered significant.

## Additional files

**Additional file 1: Table S1 and Table S2:** Elements comprising the *TOP2A* and  $\beta$ -tubulin indices. Probe sets comprising the *TOP2A* and  $\beta$ -tubulin indices.

**Additional file 2: Figure S1:** The E2F1 index is predictive of chemotherapy response in multiple datasets **A)** GSE21094 ( $n = 278$ , TFAC). **B)** GSE25055 ( $n = 310$ , AT). **C)** GSE25065 ( $n = 198$ , AT).

## Competing interests

The author's declare no competing interests.

## Author's contributions

Study design and conception of project, RMH and JAH. Completion of research, RMH. Statistical analysis, RMH and GP. Writing of manuscript RMH, GP and JAH. All authors read and approved the final manuscript.

## Acknowledgements

This work was generously supported by grants from the Canadian Stem Cell Network, the Ontario Institute for Cancer Research and the Canadian Breast Cancer Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Author details

<sup>1</sup>Department of Biochemistry and Biomedical Sciences, Centre for Functional Genomics, McMaster University, 1200 Main Street West, Hamilton, Ontario L8N 3Z5, Canada. <sup>2</sup>Department of Oncology, Juravinski Hospital and Cancer Centre, 711 Concession Street, Hamilton, Ontario L8V 1C3, Canada. <sup>3</sup>Centre for Functional Genomics, McMaster University, 1280 Main St West, Hamilton, Ontario L8S 4K1, Canada.

Received: 25 January 2012 Accepted: 20 April 2012

Published: 11 May 2012

## References

- Vogel CL, Cobleigh MA, Tripathy D, Gutheil JC, Harris LN, Fehrenbacher L, Slamon DJ, Murphy M, Novotny WF, Burchmore M, et al: **Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer.** *J Clin Oncol* 2002, **20**(3):719-726.
- van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, et al: **A gene-expression signature as a predictor of survival in breast cancer.** *N Engl J Med* 2002, **347**(25):1999-2009.
- van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, et al: **Gene expression profiling predicts clinical outcome of breast cancer.** *Nature* 2002, **415**(6871):530-536.
- Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, et al: **A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer.** *N Engl J Med* 2004, **351**(27):2817-2826.
- Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, Cronin M, Baehner FL, Watson D, Bryant J, et al: **Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer.** *J Clin Oncol* 2006, **24**(23):3726-3734.
- Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, Nordgren H, Farmer P, Praz V, Haibe-Kains B, et al: **Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis.** *J Natl Cancer Inst* 2006, **98**(4):262-272.
- Hallett RM, Dvorkin A, Gabardo CM, Hassell JA: **An algorithm to discover gene signatures with predictive potential.** *J Exp Clin Cancer Res* 2010, **29**(1):120.
- Hallett RM, Hassell JA: **E2F1 and KIAA0191 expression predicts breast cancer patient survival.** *BMC Res Notes* 2011, **4**:95.
- Hess KR, Anderson K, Symmans WF, Valero V, Ibrahim N, Mejia JA, Booser D, Theriault RL, Buzdar AU, Dempsey PJ, et al: **Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer.** *J Clin Oncol* 2006, **24**(26):4236-4244.
- Lee JK, Coutant C, Kim YC, Qi Y, Theodorescu D, Symmans WF, Baggerly K, Rouzier R, Pusztai L: **Prospective comparison of clinical and genomic multivariate predictors of response to neoadjuvant chemotherapy in breast cancer.** *Clin Cancer Res*, **16**(2):711-718.
- Rouzier R, Perou CM, Symmans WF, Ibrahim N, Cristofanilli M, Anderson K, Hess KR, Stec J, Ayers M, Wagner P, et al: **Breast cancer molecular subtypes respond differently to preoperative chemotherapy.** *Clin Cancer Res* 2005, **11**(16):5678-5685.
- Hatzis C, Pusztai L, Valero V, Booser DJ, Esserman L, Lluch A, Vidaurre T, Holmes F, Souchon E, Wang H, et al: **A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer.** *JAMA* 2011, **305**(18):1873-1881.
- Bear HD, Anderson S, Smith RE, Geyer CE Jr, Mamounas EP, Fisher B, Brown AM, Robidoux A, Margolese R, Kahlenberg MS, et al: **Sequential preoperative or postoperative docetaxel added to preoperative doxorubicin plus cyclophosphamide for operable breast cancer: National Surgical Adjuvant Breast and Bowel Project Protocol B-27.** *J Clin Oncol* 2006, **24**(13):2019-2027.
- Wolmark N, Wang J, Mamounas E, Bryant J, Fisher B: **Preoperative chemotherapy in patients with operable breast cancer: nine-year results from National Surgical Adjuvant Breast and Bowel Project B-18.** *J Natl Cancer Inst Monogr* 2001, **30**:96-102.
- Liedtke C, Mazouni C, Hess KR, Andre F, Tordai A, Mejia JA, Symmans WF, Gonzalez-Angulo AM, Hennessy B, Green M, et al: **Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer.** *J Clin Oncol* 2008, **26**(8):1275-1281.
- Abal M, Andreu JM, Barasoain I: **Taxanes: microtubule and centrosome targets, and cell cycle dependent mechanisms of action.** *Curr Cancer Drug Targets* 2003, **3**(3):193-203.
- Mozzetti S, Ferlini C, Concolino P, Filippetti F, Raspaglio G, Prislei S, Gallo D, Martinelli E, Ranelli FO, Ferrandina G, et al: **Class III beta-tubulin overexpression is a prominent mechanism of paclitaxel resistance in ovarian cancer patients.** *Clin Cancer Res* 2005, **11**(1):298-305.
- Desmedt C, Di Leo A, de Azambuja E, Larsimont D, Haibe-Kains B, Selleslags J, Delaloge S, Duhem C, Kains JP, Carly B, et al: **Multifactorial approach to predicting resistance to anthracyclines.** *J Clin Oncol* 2011, **29**(12):1578-1586.
- Nikolenyi A, Sukosd F, Kaizer L, Csorgo E, Voros A, Uhercsak G, Ormandi K, Lazar G, Thurzo L, Brodowicz T, et al: **Tumor topoisomerase II alpha status and response to anthracycline-based neoadjuvant chemotherapy in breast cancer.** *Oncology* 2011, **80**(3-4):269-277.
- Burgess DJ, Doles J, Zender L, Xue W, Ma B, McCombie WR, Hannon GJ, Lowe SW, Hemann MT: **Topoisomerase levels determine chemotherapy response in vitro and in vivo.** *Proc Natl Acad Sci U S A* 2008, **105**(26):9053-9058.
- Tanner M, Isola J, Wiklund T, Erikstein B, Kellokumpu-Lehtinen P, Malmstrom P, Wilking N, Nilsson J, Bergh J: **Topoisomerase IIalpha gene amplification predicts favorable treatment response to tailored and dose-escalated anthracycline-based adjuvant chemotherapy in HER2/neu-amplified breast cancer: Scandinavian Breast Group Trial 9401.** *J Clin Oncol* 2006, **24**(16):2428-2436.
- Aoki D, Oda Y, Hattori S, Taguchi K, Ohishi Y, Basaki Y, Oie S, Suzuki N, Kono S, Tsuneyoshi M, et al: **Overexpression of class III beta-tubulin predicts good response to taxane-based chemotherapy in ovarian clear cell adenocarcinoma.** *Clin Cancer Res* 2009, **15**(4):1473-1480.
- Tsubouchi H, Ogawa H: **Exo1 roles for repair of DNA double-strand breaks and meiotic crossing over in *Saccharomyces cerevisiae*.** *Mol Biol Cell* 2000, **11**(7):2221-2233.
- Kokavec J, Podskocova J, Zavadi J, Stopka T: **Chromatin remodeling and SWI/SNF2 factors in human disease.** *Front Biosci* 2008, **13**:6126-6134.
- Huang C, Sloan EA, Boerkoel CF: **Chromatin remodeling and human disease.** *Curr Opin Genet Dev* 2003, **13**(3):246-252.
- Baumann P, West SC: **Role of the human RAD51 protein in homologous recombination and double-stranded-break repair.** *Trends Biochem Sci* 1998, **23**(7):247-251.
- Manfredi JJ, Horwitz SB: **Taxol: an antimetabolic agent with a new mechanism of action.** *Pharmacol Ther* 1984, **25**(1):83-125.

28. Manfredi JJ, Parness J, Horwitz SB: Taxol binds to cellular microtubules. *J Cell Biol* 1982, **94**(3):688–696.
29. Parness J, Horwitz SB: Taxol binds to polymerized tubulin in vitro. *J Cell Biol* 1981, **91**(2 Pt 1):479–487.
30. Wittmann T, Wilm M, Karsenti E, Vernos I: TPX2, A novel xenopus MAP involved in spindle pole organization. *J Cell Biol* 2000, **149**(7):1405–1418.
31. Ishikawa R, Hayashi K, Shirao T, Xue Y, Takagi T, Sasaki Y, Kohama K: Drebrin, a development-associated brain protein from rat embryo, causes the dissociation of tropomyosin from actin filaments. *J Biol Chem* 1994, **269**(47):29928–29933.
32. Desmedt C, Haibe-Kains B, Wirapati P, Buyse M, Larsimont D, Bontempi G, Delorenzi M, Piccart M, Sotiriou C: Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin Cancer Res* 2008, **14**(16):5158–5165.
33. Wirapati P, Sotiriou C, Kunkel S, Farmer P, Pradervand S, Haibe-Kains B, Desmedt C, Ignatiadis M, Sengstag T, Schutz F, *et al*: Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res* 2008, **10**(4):R65.
34. Mosley JD, Keri RA: Cell cycle correlated genes dictate the prognostic power of breast cancer gene lists. *BMC Med Genomics* 2008, **1**:11.
35. Venet D, Dumont JE, Detours V: Most random gene expression signatures are significantly associated with breast cancer outcome. *PLoS Comput Biol* 2011, **7**(10):e1002240.
36. Dai H, van't Veer L, Lamb J, He YD, Mao M, Fine BM, Bernards R, van de Vijver M, Deutsch P, Sachs A, *et al*: A cell proliferation signature is a marker of extremely poor outcome in a subpopulation of breast cancer patients. *Cancer Res* 2005, **65**(10):4059–4066.
37. DeGregori J, Johnson DG: Distinct and Overlapping Roles for E2F Family Members in Transcription, Proliferation and Apoptosis. *Curr Mol Med* 2006, **6**(7):739–748.
38. DeGregori J, Kowalik T, Nevins JR: Cellular targets for activation by the E2F1 transcription factor include DNA synthesis- and G1/S-regulatory genes. *Mol Cell Biol* 1995, **15**(8):4215–4224.
39. DeGregori J, Leone G, Miron A, Jakoi L, Nevins JR: Distinct roles for E2F proteins in cell growth control and apoptosis. *Proc Natl Acad Sci U S A* 1997, **94**(14):7245–7250.
40. Slansky JE, Farnham PJ: Introduction to the E2F family: protein structure and gene regulation. *Curr Top Microbiol Immunol* 1996, **208**:1–30.
41. Vuaroqueaux V, Urban P, Labuhn M, Delorenzi M, Wirapati P, Benz CC, Flury R, Dieterich H, Spyrtos F, Eppenberger U, *et al*: Low E2F1 transcript levels are a strong determinant of favorable breast cancer outcome. *Breast Cancer Res* 2007, **9**(3):R33.
42. Kavallaris M, Kuo DY, Burkhart CA, Regl DL, Norris MD, Haber M, Horwitz SB: Taxol-resistant epithelial ovarian tumors are associated with altered expression of specific beta-tubulin isoforms. *J Clin Invest* 1997, **100**(5):1282–1293.
43. Slamon DJ, Press MF: Alterations in the TOP2A and HER2 genes: association with adjuvant anthracycline sensitivity in human breast cancers. *J Natl Cancer Inst* 2009, **101**(9):615–618.
44. Tommasi S, Mangia A, Lacalamita R, Bellizzi A, Fedele V, Chiriatti A, Thomssen C, Kendzierski N, Latorre A, Lorusso V, *et al*: Cytoskeleton and paclitaxel sensitivity in breast cancer: the role of beta-tubulins. *Int J Cancer* 2007, **120**(10):2078–2085.
45. Scoggins JF, Ramsey SD: A national cancer clinical trials system for the 21st century: reinvigorating the NCI Cooperative Group Program. *J Natl Cancer Inst* 2010, **102**(17):1371.
46. Yau C, Esserman L, Moore DH, Waldman F, Sninsky J, Benz CC: A multigene predictor of metastatic outcome in early stage hormone receptor-negative and triple-negative breast cancer. *Breast Cancer Res* 2010, **12**(5):R85.
47. Juul N, Szallasi Z, Eklund AC, Li Q, Burrell RA, Gerlinger M, Valero V, Andreopoulou E, Esteva FJ, Symmans WF, *et al*: Assessment of an RNA interference screen-derived mitotic and ceramide pathway metagene as a predictor of response to neoadjuvant paclitaxel for primary triple-negative breast cancer: a retrospective analysis of five clinical trials. *Lancet Oncol*, **11**:358–365.
48. Dang C, Lin N, Moy B, Come S, Sugarman S, Morris P, Abbruzzi A, Chen C, Steingart R, Patil S, *et al*: Dose-dense doxorubicin and cyclophosphamide followed by weekly paclitaxel with trastuzumab and lapatinib in HER2/neu-overexpressed/amplified breast cancer is not feasible because of excessive diarrhea. *J Clin Oncol* 2010, **28**(18):2982–2988.
49. Moy B, Kirkpatrick P, Kar S, Goss P: Lapatinib. *Nat Rev Drug Discov* 2007, **6**(6):431–432.
50. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP: Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003, **4**(2):249–264.
51. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, *et al*: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999, **286**(5439):531–537.
52. Hallett RM, Dvorkin-Gheva A, Anita B, Hassell JA: A Gene Signature for Predicting Outcome in Patients with Basal-like Breast Cancer. *Sci Rep* 2012, **2**:227.

doi:10.1186/1755-8794-5-16

Cite this article as: Hallett *et al*: A target based approach identifies genomic predictors of breast cancer patient response to chemotherapy. *BMC Medical Genomics* 2012 **5**:16.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

