Biology of Human Tumors

Clinical Cancer Research

Intratumor Heterogeneity Affects Gene Expression Profile Test Prognostic Risk Stratification in Early Breast Cancer

Rekha Gyanchandani[1], Yan Lin[2], Hui-Min Lin[2], Kristine Cooper[2], Daniel P. Normolle[2], Adam Brufsky[3], Michael Fastuca[1], Whitney Crosson[1], Steffi Oesterreich[1], Nancy E. Davidson[3], Rohit Bhargava[4], David J. Dabbs[4], and Adrian V. Lee[1]

Abstract

Purpose: To examine the effect of intratumor heterogeneity (ITH) on detection of genes within gene expression panels (GEPs) and the subsequent ability to predict prognostic risk.

Experimental Design: Multiplexed barcoded RNA analysis was used to measure the expression of 141 genes from five GEPs (Oncotype Dx, MammaPrint, PAM50, EndoPredict, and Breast Cancer Index) in breast cancer tissue sections and tumor-rich cores from 71 estrogen receptor (ER)-positive node-negative tumors, on which clinical Oncotype Dx testing was previously performed. If the tumor had foci of high Ki67 ($n = 26$), low/negative progesterone receptor (PR; $n = 13$), or both ($n = 5$), additional cores were obtained. In total, 181 samples were processed. Oncotype Dx recurrence scores were calculated from NanoString nCounter gene expression data.

Results: Hierarchical clustering using all GEP genes showed that majority (61 of 71) of tumor samples clustered by patient, indicating greater interpatient heterogeneity (IPH) than ITH. We found a strikingly high correlation between Oncotype Dx recurrence scores obtained from whole sections versus tumor-rich cores ($r = 0.94$). However, high Ki67 and low PR cores had slightly higher but not statistically significant recurrence scores. For 18 of 71 (25%) patients, scores were divergent between sections and cores and crossed the boundaries for low, intermediate, and high risk.

Conclusions: Our study indicates that in patients with highly heterogeneous tumors, GEP recurrence scores from a single core could under- or overestimate prognostic risk. Hence, it may be a useful strategy to assess multiple samples (both representative and atypical cores) to fully account for the ITH-driven variation in risk prediction. *Clin Cancer Res; 22(21); 5362–9. ©2016 AACR.*

Introduction

Breast cancer, the most common malignancy in women, is a heterogeneous disease characterized by distinct molecular subtypes (1–3). In the past decade, gene expression profiling has enabled development of a wide variety of multigene prognostic signatures such as Oncotype Dx (4), MammaPrint (5), PAM50 (Prosigna) (6), EndoPredict (7), and Breast Cancer Index (BCI; refs. 8, 9). Clinical studies on large patient cohorts have demonstrated that these gene expression panels (GEP) may serve as tools to identify patients who are most likely to benefit from adjuvant systemic therapies while sparing others of the unwanted side effects and treatment-related cytotoxicity. Although there is growing recognition that GEPs are clinically relevant in breast cancer management, they have not been fully embedded into routine clinical practice (10–12).

Among the many commercially available GEPs, Oncotype Dx and EndoPredict are the only tests that are supported by level I evidence on the basis of the marker utility grading system (13). Oncotype Dx (Genomic Health Inc.) measures the expression of 21 genes and calculates a recurrence score that predicts the risk of relapse in patients with estrogen receptor (ER)-positive lymph node–negative early-stage breast cancer (4). Oncotype Dx is so far the most widely used GEP in clinical practice likely based upon its clinical validation (14) and approval by the National Comprehensive Cancer Network (NCCN), American Society of Clinical Oncology (ASCO), and St. Gallen European Society for Medical Oncology (ESMO).

In the past decade, there have been an increasing number of reports of intratumor heterogeneity (ITH) of gene expression and somatic DNA mutations (15–20). Immunohistochemical and FISH studies using multiple areas of a breast tumor have shown significant ITH in ER gene expression levels and *HER2* amplification (15, 16). However, very few studies have examined the effect of ITH on measurement of GEPs and the accuracy to predict

[1]Women's Cancer Research Center, Department of Pharmacology and Chemical Biology, University of Pittsburgh Cancer Institute, Magee Womens Research Institute, Pittsburgh, Pennsylvania. [2]Department of Biostatistics, University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania. [3]Department of Medicine, University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania. [4]Department of Pathology, Magee-Womens Hospital, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

Corresponding Author: Adrian V. Lee, Womens Cancer Research Center and Department of Pharmacology & Chemical Biology, University of Pittsburgh Cancer Institute, Magee Womens Research Institute, 204 Craft Avenue, Room A412, Pittsburgh, PA 15213. Phone: 412-641-8554; Fax: 412-641-2458; E-mail: leeav@upmc.edu

AACR

Translational Relevance

Recent studies show tremendous transcriptomic and genomic heterogeneity not only between breast cancers but also within a single breast cancer. This study examines the clinical importance of this heterogeneity, showing that prognostic risk scores derived from gene transcript levels deviate when taken from different regions (cores) of a breast tumor. Importantly, use of single cores can under- or overestimate prognostic risk and highlight the importance of understanding intratumor heterogeneity for breast cancer prognosis.

patient prognosis (21). In a pilot study of 4 patients, Drury and colleagues reported high concordance in Oncotype Dx recurrence scores between 0.6-mm breast tumor cores and whole sections but showed high variability in recurrence score between individual cores resulting in prognostic misclassification (22). Barry and colleagues examined the influence of ITH on precision of microarray-based assays in multiple core needle biopsies and showed high variance in recurrence risk predictions in 1 of 18 patients with breast cancer due to global variation in gene expression (23). Another study by Gerlinger and colleagues in a small number of renal cell carcinomas (RCC) identified gene expression signatures of both good and poor prognosis in different regions of the same tumor (18). In a recent report, they reanalyzed this gene expression data in 63 regions from 10 RCCs and identified ITH in 8 of 10 tumors (24), thus highlighting the importance of multiregion assessment in prediction of prognosis, response to therapy, and risk of relapse. However, more research is needed to fully understand the degree of ITH-driven variation in risk prediction and design novel tumor sampling strategies that provide more reliable risk estimates.

In the present study, we investigated ITH of GEPs using multiplexed barcoded hybridization to measure the expression of 141 genes from 5 GEPs (Oncotype Dx, MammaPrint, PAM50, EndoPredict, and BCI) in whole tumor sections compared with multiple cores from 71 ER-positive node-negative tumors. In addition to selecting a tumor-rich core, areas with high Ki67 and/or low progesterone receptor (PR) were also used to punch cores, potentially representing aggressive areas of a tumor. In addition to examining the effect of ITH on measurement of GEPs, we also examined the effect of ITH on prediction of prognosis in the Oncotype Dx assay.

Materials and Methods

Breast tumor specimens

We previously reported interobserver agreement among pathologists for hormone receptor scoring in 74 cases of ER-positive early breast cancer (25). We used the same cohort to examine GEPs but removed 3 cases because of inadequate tissue, for a total of 71 cases (Supplementary Table S1). All patients had clinical Oncotype Dx recurrence risk scoring performed at Genomic Health. Studies were performed with Institutional Review Board approval PRO09100201. Immunohistochemical expression levels for ER, PR, and Ki67 were scored according to the ASCO/CAP guidelines (Supplementary Table S2).

For each patient, we selected the single formalin-fixed, paraffin-embedded (FFPE) block that was used for the clinical Oncotype Dx test and cut a 5-µm section and a 0.6-mm core from a tumor-rich representative part of the block. Additional cores were cut from tumors that had foci of high Ki67 ($n = 26$), low PR ($n = 13$), or both ($n = 5$) to test the hypothesis that high Ki67 and/or low PR may indicate aggressive areas of a tumor. High Ki67 area in a tumor was defined as 10% or higher labeling index compared with the overall Ki67 labeling index for the whole section (Supplementary Table S2). Low PR area in a tumor was defined as more than 50% cells negative for PR. Five patients had all 4 types of tumor samples (including section, tumor-rich core, high Ki67 core, and low PR core), 39 patients had 3 tumor samples (section, tumor-rich core, and either high Ki67 core or low PR core), and 27 patients had 2 tumor samples (section and tumor-rich core). In total, we processed 181 samples for NanoString nCounter analysis.

RNA isolation

The selected tumor blocks were enriched for invasive tumor and the most predominant noninvasive tissue component in the blocks was adipose tissue. For the tumor section, the whole section was scraped (without macrodissection), and the paraffin shavings were used to isolate RNA. In addition, 1 to 3 cores from each sample were also used for RNA isolation. RNA was isolated using RNeasy FFPE Kit (Qiagen) and quantified using UV spectroscopy (Nanodrop Technologies).

Gene expression

Barcoded probes to measure the expression of genes comprising 5 GEPs and their respective housekeeping genes were manufactured by NanoString Technologies. This included Oncotype Dx (16 genes, 5 housekeeping genes), MammaPrint (66 genes), PAM50 (50 genes, 5 housekeeping genes), EndoPredict (8 genes, 3 housekeeping genes), and BCI (7 genes, 4 housekeeping genes). Since there were some genes that overlapped among the 5 GEPs, gene expression was measured for a total of 141 unique genes (127 endogenous genes and 14 housekeeping genes; Supplementary Table S3). The nCounter assay also included 6 positive controls and 8 negative controls. nCounter analysis was performed according to the manufacturer's instructions using 100 ng of total RNA. Data were collected using the nCounter Digital Analyzer and initially processed using nSolver Analysis Software. QC metrics, including positive control linearity and limit of detection, were assessed using the positive and negative control probes (Supplementary Fig. S1A). Raw intensities were normalized to the geometric mean of the positive controls and housekeeping genes using the R NanoStringNorm package (Supplementary Fig. S1B). Data reproducibility was also assessed using technical replicates (Supplementary Fig. S1C). NanoString data are available in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE79378.

Statistical analysis

All data analyses were performed using R version 3.1.2 (http://www.r-project.org). The NanoString raw intensities were normalized to the geometric mean of positive controls and housekeeping genes using the R NanoStringNorm package (26). Hierarchical clustering (using the Ward method on the Manhattan metric) was performed to visualize the clustering pattern of the NanoString gene expression data. To evaluate the effect of ITH on the GEPs, we clustered all samples by genes belonging to each GEP and

Gyanchandani et al.

calculated the percentage of subjects with all samples clustered together as an indication of the robustness of the GEP to the ITH. Since the clustering results heavily depend upon the number of clusters, we performed the clustering analysis by the number of clusters from 1 to 80.

Clinical Oncotype Dx recurrence scores were available for all patients. The Oncotype Dx score is based on an RT-PCR assay utilizing RNA isolated from macrodissected tissue sections, which is most comparable to the whole sections used in our study. As RT-PCR and nCounter use different technologies for mRNA detection (amplification versus hybridization), the Oncotype Dx recurrence score was calculated from NanoString data (nstringRS) by fitting the following model to data from the "section" samples (1):

$$clinRS = \beta_0 + \sum_{i=1}^{16} \beta_i X_i + \varepsilon \qquad (1)$$

where clinRS denotes the clinical Oncotype Dx recurrence score and X_i represent the expression level of the 16 Oncotype Dx genes in the sections measured by Nanostring. The overall fit of the model was good ($R^2 = 0.87$). The coefficients (β_0 and β_i) estimated from the linear model were then used to calculate the nstringRS for each type of tissue using (2).

$$nstringRS = \beta_0 + \sum_{i=1}^{16} \beta_i X_i \qquad (2)$$

Here, the X_i represent the expression level of the Oncotype Dx genes measured by Nanostring for each sample.

The association of gene expression or nstringRS between two different types of tumor samples was described by the Spearman correlation coefficient. To investigate the intratumor variability of each gene and nstringRS, we fit a random-effects model for each gene by subject. Samples from the same subject are considered to be within a cluster. From the linear mixed-effect model, we derived the intratumor variability (1-ICC), where $ICC = \frac{\text{inter-patient variance}}{\text{total variance}}$. (1-ICC) scores ranged between 0 and 1 and were grouped as low (0–0.2), fair (0.2–0.4), moderate (0.4–0.6), and high agreement (0.6–1.0). Kruskal–Wallis tests were used to compare gene expression levels and recurrence scores across all 4 sample types and Wilcoxon rank-sum tests were used for comparisons between 2 groups (Supplementary Table S4).

Results

Hierarchical clustering analysis of genes from five GEPs shows greater inter-patient heterogeneity than ITH

We used nCounter analysis to measure the expression of genes from 5 GEPs in FFPE tumor sections compared to cores taken from tumor blocks of 71 ER-positive tumors (Fig. 1A). Figure 1B shows the distribution of gene expression intensities for 141 measured genes (from 5 GEPs) and their variability across 181 tumor samples (from 71 patients). We investigated the extent of inter-tumor heterogeneity and ITH in gene expression for GEPs in this patient cohort by performing hierarchical clustering using all 141 genes (Fig. 1C and Supplementary Fig. S2). We found that majority of tumor samples clustered by patient (61 of 71), indicating greater interpatient heterogeneity (IPH) than ITH (Fig. 1C). Ten of 71 (14%) patients showed discordant tumor samples, indicating ITH. A heatmap shows 2 distinct patient clusters (rows; Supplementary Fig. S2). On the basis of the relative gene expres-

sion values of proliferation and survival genes, the top and bottom clusters likely represent patients with low and high recurrence risk, respectively. We also performed clustering using the genes from individual GEPs, which generally showed greater ITH compared with the combined analysis using all GEP genes (Fig. 1D, Table 1, and Supplementary Fig. S3A–S3E). Mamma-Print contained the highest number of genes ($n = 66$) among all tests and showed the least number of discordant samples (14 of 71; Table 1 and Supplementary Fig. S3A–S3E). In contrast, BCI, which had the lowest number of genes ($n = 7$) showed the highest number of discordant samples (52 of 71), suggesting that the number of genes in a GEP may influence susceptibility to ITH.

Several genes involved in proliferation and invasion display high ITH

We next calculated the ICC for each gene as a measurement of heterogeneity (Fig. 2). 1-ICC values represented ITH (Fig. 2A). 1-ICC scores ranging between 0 and 1 were grouped as 0–0.2 (low heterogeneity), 0.2–0.4 (fair heterogeneity), 0.4–0.6 (moderate heterogeneity), and 0.6–1.0 (high heterogeneity). Among the 127 genes that were analyzed, 73 genes showed low, 36 genes showed fair, 11 genes showed moderate, and 7 genes showed high ITH (Fig. 2A and Table 2). Hence, a small proportion of genes (18 of 127, 14%) showed elevated variability in gene expression among the tumor samples. Figure 2B shows the distribution of 1-ICC scores for genes in the individual GEPs. Three of 5 GEPs (including Oncotype Dx, MammaPrint, and PAM50) showed genes with moderate-to-high ITH. Supplementary Table S5 lists these heterogeneous genes and the correlation of expression values between different samples. This list includes several proliferation- and invasion-related genes including *MYC*, *FOXC1*, *EGFR*, *FGF18*, *CTSL2*, and *MMP9*. When we clustered patient samples using the genes with low ITH (0–0.2 and 0.2–0.4 intravariance scores), the majority of samples clustered by patient (Fig. 2C and Supplementary Fig. S4A–S4D). Conversely, when we clustered using the genes with 0.4–0.6 and 0.6–1.0 intravariance scores, only a few patients had all samples clustered together, thus confirming their high ITH.

ITH of gene expression affects Oncotype Dx recurrence risk stratification

Oncotype Dx is the most widely used GEP for evaluating ER-positive early breast cancer prognosis. It uses a weighting algorithm to calculate the risk of recurrence score, which is divided into low-risk (<18), intermediate-risk (18–30), and high-risk (≥31) categories. On the basis of the clinical Oncotype Dx recurrence scoring (clinRS) performed at Genomic Health, 28 patients showed low-risk with 1 case of disease recurrence; 30 patients showed intermediate-risk with 5 cases of recurrence; and 13 patients showed high-risk with 1 case of recurrence (Supplementary Table S6). To study the effect of ITH arising from sampling a tumor-enriched area or regions of high Ki67 and low PR, we estimated a predicted RS on the basis of NanoString data (nstringRS), as described in Statistical analysis. Supplementary Figure S5A and S5B shows the expression of Ki67 and PR mRNA in the different types of tumor samples. To compare Ki67 and PR expression levels between tumor-rich cores and high Ki67/low PR cores, we selected only those patients for which both types of cores were available ($n = 26$ and $n = 13$, respectively; Supplementary Fig. S5C and S5D). As expected, the expression of Ki67 mRNA was significantly higher in cores from focal areas of high Ki67, and PR

Figure 1.

Hierarchical clustering analysis of genes from 5 GEPs shows greater IPH than ITH. **A,** nCounter analysis was used to measure the expression of genes from 5 GEPs in FFPE tumor sections compared with cores taken from tumor blocks for 71 ER-positive node-negative tumors. Cores were also obtained from foci of high Ki67 ($n = 26$), low PR ($n = 13$), or both ($n = 5$). **B,** mean versus SD plot of gene expression intensities for all measured genes (including 127 endogenous genes, 14 housekeeping genes, 6 positive controls, and 8 negative controls). Housekeeping genes show modest to high levels of gene expression with very low variation. **C,** hierarchical clustering by the Ward method using the Manhattan metric was performed on all GEP genes. The heatmap represents gene expression from 71 tumors ($n = 181$ samples) profiled for 5 GEPs (127 endogenous genes). Red indicates high, and green indicates low relative gene expression. Genes (columns) are clustered, and tumors (rows) are clustered. **D,** clustering analysis for individual GEPs indicating the proportion of patients with all samples within the same cluster for a range of clusters (1–80).
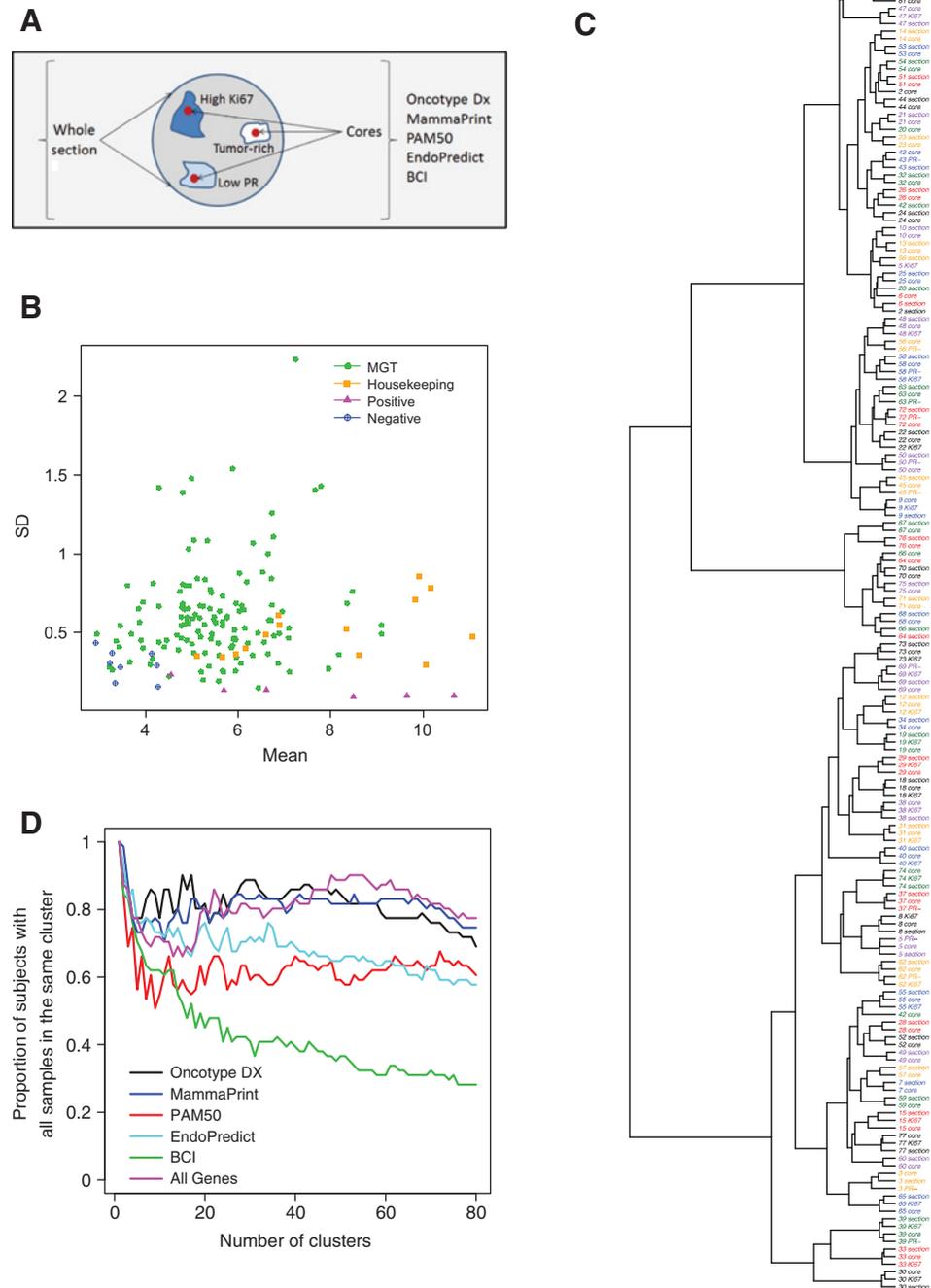
mRNA was significantly lower in cores from areas of low PR (Supplementary Fig. S5C and S5D). Figure 3A and Supplementary Table S6 shows nstringRS scores for the different types of samples along with the clinical Oncotype DX recurrence scores (clinRS). Overall, the nstringRS scores derived from sections and tumor-rich cores correlated well with the clinRS (Spearman's $\rho = 0.92$ and $\rho = 0.90$, respectively; Supplementary Fig. S6A and S6B). However, the Oncotype Dx risk categories differed in 14 of 71 (19.7%) sections and 16 of 71 (22.5%) tumor-rich cores. We next looked at the agreement of nstringRS between the different types of samples (Fig. 3B and Supplementary Fig. S6C–S6F). nstringRS was strikingly similar between a whole tumor section and a tumor-rich core (Spearman's $\rho = 0.94$; Fig. 3B). The risk categories based on the tumor-rich cores and high Ki67 cores differed in 7 of 26 (27%) samples, whereas those derived from low PR cores differed in 3 of 13 (23%) samples. Although, the high Ki67 and low PR cores showed a trend toward higher median nstringRS scores compared with tumor-rich cores (Supplementary Fig. S6E

Gyanchandani et al.

Table 1. Patients with discordant tumor samples identified by hierarchical clustering of GEP genes

	GEP	Number of endogenous genes	Patients with discordant tumor samples
1	BCI	7	52/71
2	Endo	8	30/71
3	Oncotype Dx	16	26/71
4	PAM50	50	25/71
5	MammaPrint	66	14/71
6	All genes	127	10/71

and S6F), when we calculated for each patient, the difference in nstringRS scores between tumor-rich cores and high Ki67 or low PR cores, we observed no significant change in the median of this quantity ($P = 0.095$ and $P = 0.675$, respectively). Finally, when we compared clinRS with the nstringRS for all types of samples, we found that for majority of tumors (53 of 71, 75%) the different samples showed similar nstringRS scores and risk stratification. However, in 18 of 71 (25%) tumors, the recurrence scores diverged enough to cause differential classification, as the scores crossed the boundaries for low, intermediate, and high risk (Fig. 3C, Table 3, Supplementary Tables S6 and S7). One of 18 tumors with discordant scores (associated with a decrease in risk due to a section) was from a patient with recurrent disease (Supplementary Table S6). In addition, in these tumors with discordant scores, cases where at least 3 types of samples (section, tumor-rich core, high Ki67 core, and/or low PR core) were available, 6 of 10 tumors

showed an increase in risk due to a high Ki67 or low PR core, 3 of 10 tumors showed an increase in risk due to a tumor-rich core, and 1 of 10 tumors showed an increase in risk due to both low PR core and a tumor-rich core (Supplementary Table S7).

Discussion

This is the first study to comprehensively examine the effect of ITH on measurement of clinically used GEPs and their ability to predict prognostic risk in early breast cancer. The study utilized the Nanostring nCounter platform, which is ideally suited for gene expression detection in FFPE tissue. We described the ITH for each gene ($n = 127$) in the 5 most commonly used GEPs. Hierachical clustering of tumors using all of the genes in the 5 GEPs showed relatively low ITH as individual samples from each patient clustered together for the majority of patients. However, when clustering tumors using genes in the individual GEPs, higher rates of ITH were found. An in-depth analysis of Oncotype Dx showed a strikingly high correlation between the Oncotype Dx recurrence score from a whole section (without macrodissection) and a representative tumor-rich core, suggesting little influence of the tumor microenvironment on the genes in this test. However, when measuring multiple cores within a tumor, ITH resulted in prognostic misclassification in 25% of patients.

Recent genome-wide genomic and transcriptomic studies have indicated high ITH in cancer, with some tumors having regions of both indolent and aggressive diseases (15–20). However, these studies are generally designed to identify the greatest level of ITH,
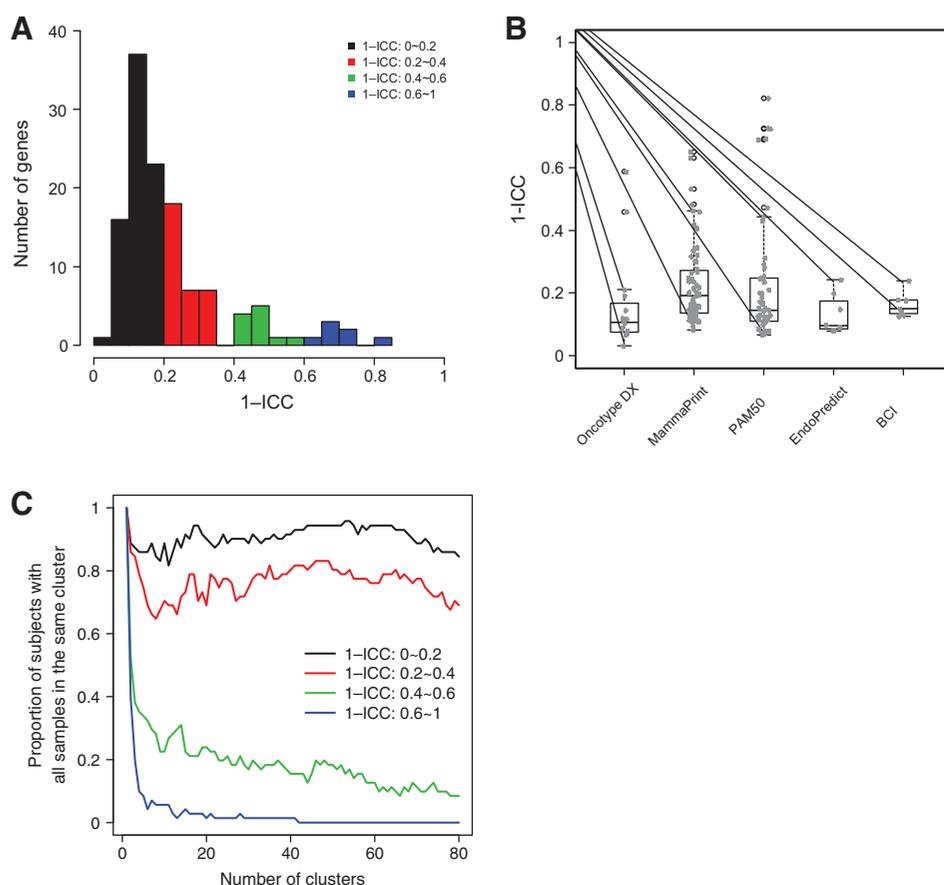
Figure 2.
ICC for GEP genes as a measurement of heterogeneity. **A,** frequency distribution of 1-ICC scores representing ITH is shown for all GEP genes. 1-ICC scores range between 0 and 1 [0–0.2 (low), 0.2–0.4 (fair), 0.4–0.6 (moderate), and 0.6–1.0 (high)]. A tail to the right indicates a subgroup of the genes that are more heterogeneous among different types of tumor samples. **B,** 1-ICC distribution by gene signatures. Oncotype Dx, MammaPrint, and PAM50 tests show genes with higher heterogeneity compared with EndoPredict and BCI. **C,** clustering analysis was done using genes with varying ITH. Number of patients with all samples in the same cluster is plotted against the number of clusters. Lower number of patients with all samples clustered together indicates higher ITH.

Table 2. Distribution of genes with low, fair, moderate, and high ITH

		1-ICC				
		0~0.2	0.2~0.4	0.4~0.6	0.6~1	Total
1	Oncotype	13	1	2	0	16
2	MammaPrint	35	21	6	2	64
3	PAM50	32	9	3	5	49
4	EndoPredict	7	1	0	0	8
5	BCI	6	1	0	0	7
6	All genes	73	36	11	7	127

as they include all transcribed genes. This is in contrast to GEPs which all use a small number of selected genes. Indeed, the effect of ITH on GEP tests seems to be a balance between the number of genes in the test, and the ITH of each of these genes. For example, when we clustered tumors on the basis of all of the genes that were measured, we found that most tumor samples clustered by patient, indicating greater IPH than ITH. However, selecting genes from each individual GEP resulted in lower numbers of tumors clustering per patients and higher apparent ITH. An additional level of consideration is the actual genes themselves. We show the level of ITH of each gene in each test and find that choosing genes with low ITH results in apparent low IPH. It should be noted that none of the current GEPs used ITH as a determinant in inclusion/exclusion of genes in the test.

Our data are consistent with two previous smaller reports on the level of ITH and its effect upon risk prediction in early breast cancer. Drury and colleagues. reported high concordance in recurrence scores between 0.6-mm cores and whole tumor sections, similar to our data, but showed high variability in recurrence scores between the individual cores, resulting in prognostic

misclassification in 1 of 4 patients (25%; ref. 22). Barry and colleagues. examined the influence of ITH on the precision of microarray-based assays in multiple core needle biopsies and showed high variance in recurrence risk predictions in 1 of 18 patients due to global variation in gene expression (23). In the current study, the clinical Oncotype Dx recurrence scores correlated well with the nstringRS scores derived from sections and tumor-rich cores, but the Oncotype Dx risk categories differed in 20% of sections and 23% of representative cores. Similarly, the risk categories based on the high Ki67 and low PR expression cores differed in 27% and 23% of patients, respectively. Overall, comparing the nstringRS from sections and all 3 cores, we found a 25% (18 of 71) divergence in risk categories. In our analysis, we also found that an increase in risk was more commonly observed with a high Ki67 or low PR core, whereas a decrease in risk occurred more frequently due to a tumor-rich core. However, given the similarities in risk scores and differences in risk categories, the high variation observed in risk categories might be due to a limitation of the risk category cutpoints that lie in high-density areas of patient scores, even if those cutpoints have been precisely and objectively determined. Importantly, all differences in classifications were between adjacent risk groups (e.g., low to intermediate and intermediate to high), and no tumors showed divergence from low to high. Another limitation is that the Oncotype Dx and NanoString nCounter use different technologies for mRNA detection (amplification vs. hybridization). Although they show a good concordance using our linear model ($R^2 = 0.87$), differences in the technologies might contribute to some of the variations in the risk categories. Hence, on the basis of the inclusion of atypical cores potentially representing aggressive

Figure 3.
ITH in gene expression affects Oncotype Dx recurrence risk stratification. **A,** NanoString-derived Oncotype Dx recurrence scores (nstringRS) are indicated for patients with different sample types; section, tumor-rich core, high Ki67 core, and low PR core, along with clinical Oncotype Dx recurrence scores (clinRS). **B,** correlation of nstringRS between whole sections and representative cores (Spearman's $\rho = 0.94$). **C,** clinRS was compared with the nstringRS for all types of samples for changes in risk stratification. For 18 of 71 patients, recurrence scores crossed the boundaries for low, intermediate, and high risk.
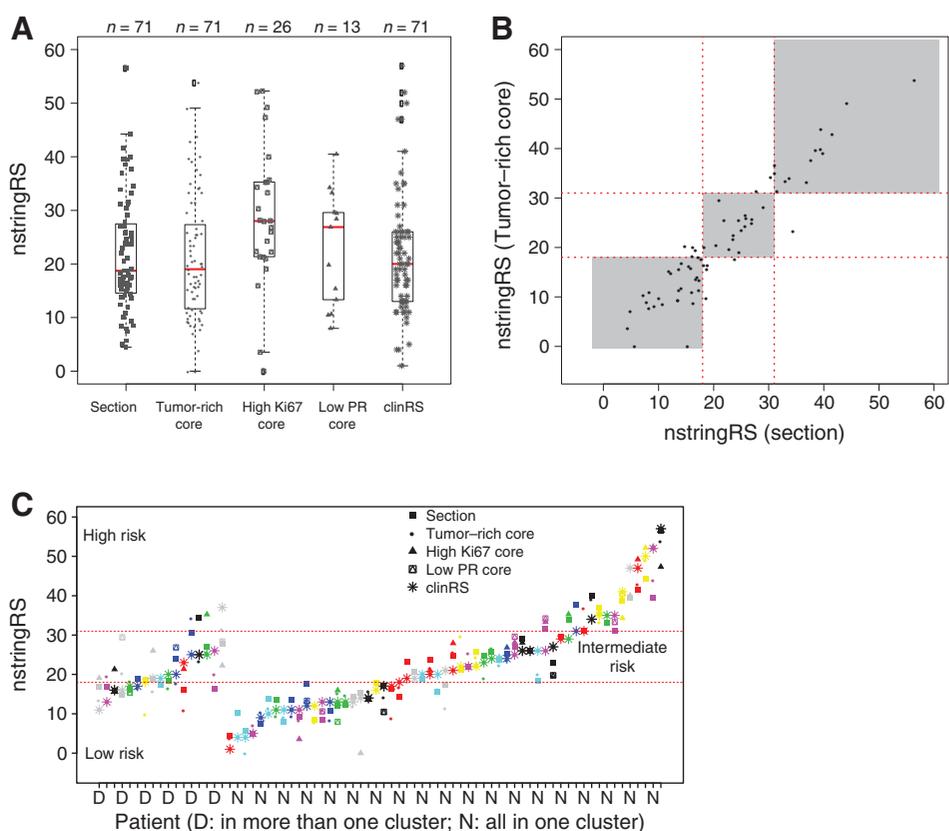
Table 3. Differential risk stratification of patients with discordant scores based on Clinical Oncotype Dx and NanoString Oncotype Dx recurrence scores

Number of patients	Risk from clinical oncotype Dx recurrence scores (clinRS)	Risk from nanoString oncotype Dx recurrence scores (nstringRS)
7/71 (9.9%)	Low	Intermediate
4/71 (5.6%)	Intermediate	High
5/71 (7.0%)	Intermediate	Low
1/71 (1.4%)	High	Intermediate

areas of the tumor, the 25% variation in prognostic risk stratification provides an overestimation of the magnitude of existing ITH. Even so, this is a clinically significant finding and underscores the need to better understand the analytic variables affecting performance of currently used GEPs. This suggests that in patients with highly heterogeneous tumors, multiple cores might be required to estimate risk prediction and that it might be a useful strategy to include both representative and atypical cores while selecting multiple samples to fully account for the ITH-driven variation in risk prediction. However, given the fact that in our study, only 7 of 71 patients showed recurrent disease, the clinical meaning of this ITH-driven variation in prognostic risk stratification for the individual patients could not be fully addressed. Hence, future studies utilizing even larger cohorts of patients with higher number of prognostic and nonprognostic cores will be required to further confirm these findings.

The striking similarity between Oncotype Dx nstringRS scores obtained from a whole section and a tumor-rich core suggests that the tumor microenvironment may have little effect on this GEP and likely represents the fact that the genes in this GEP are quite tumor-specific (e.g., *ER*, *PR*, *ERBB2*, etc.). Furthermore, the tumor blocks included in this study were invasive tumor–enriched and the most predominant noninvasive tissue was adipose tissue. It has been shown that a mitotically active tumor stroma (27) and a biopsy cavity with activated lymphocytes (28) can spuriously elevate the risk scores. However, the concordance between Oncotype Dx scores in a section and representative core was not due to the fact that these early breast cancers were very homogenous, as histopathology showed a dynamic range of stromal and immune involvement (data not shown). Other GEPs, such as MammaPrint, contain many genes that are involved in the tumor microenvironment (e.g., *VEGF*) and thus may be more affected by the amount of stroma in a tumor.

In summary, a direct analysis of ITH for the 5 main GEPs in early breast cancer reveal intratumor variation, with ITH being inversely proportional to the number of genes in the test. For Oncotype Dx, the nstringRS scores strongly correlated between a tumor-rich core and whole section and suggest that a representative core may be used to determine recurrence score. However, in instances where tumors do have ITH, the differences in gene expression can affect risk prediction. This was apparent in 25% of patients and suggests that GEPs might best be measured using knowledge and estimates of ITH from prior histopathology.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: A.M. Brufsky, S. Oesterreich, N.E. Davidson, R. Bhargava, D.J. Dabbs, A.V. Lee
Development of methodology: R. Gyanchandani, K.L. Cooper, S. Oesterreich, R. Bhargava, A.V. Lee
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): R. Gyanchandani, A.M. Brufsky, W. Crosson, D.J. Dabbs, A.V. Lee
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): R. Gyanchandani, Y. Lin, H.-M. Lin, K.L. Cooper, D.P. Normolle, A.M. Brufsky, A.V. Lee
Writing, review, and/or revision of the manuscript: R. Gyanchandani, Y. Lin, K.L. Cooper, D.P. Normolle, A.M. Brufsky, S. Oesterreich, N.E. Davidson, R. Bhargava, D.J. Dabbs, A.V. Lee
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): R. Gyanchandani, A.M. Brufsky, M. Fastuca, A.V. Lee
Study supervision: R. Bhargava, A.V. Lee

References

1. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature 2000;406:747–52.
2. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869–74.
3. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 2003;100:8418–23.
4. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004;351:2817–26.
5. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530–6.
6. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 2009;27:1160–7.
7. Filipits M, Rudas M, Jakesz R, Dubsky P, Fitzal F, Singer CF, et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. Clin Cancer Res 2011;17:6012–20.
8. Ma XJ, Salunga R, Dahiya S, Wang W, Carney E, Durbecq V, et al. A five-gene molecular grade index and HOXB13:IL17BR are complementary prognostic factors in early stage breast cancer. Clin Cancer Res 2008;14:2601–8.
9. Goetz MP, Suman VJ, Ingle JN, Nibbe AM, Visscher DW, Reynolds CA, et al. A two-gene expression ratio of homeobox 13 and interleukin-17B receptor for prediction of recurrence and survival in women receiving adjuvant tamoxifen. Clin Cancer Res 2006;12:2080–7.

10. Zelnak AB, O'Regan RM. Genomic subtypes in choosing adjuvant therapy for breast cancer. Oncology (Williston Park) 2013;27:204–10.

11. Weigelt B, Reis-Filho JS, Swanton C. Genomic analyses to select patients for adjuvant chemotherapy: trials and tribulations. Ann Oncol 2012;23Suppl10:x211–8.

12. Gyorffy B, Hatzis C, Sanft T, Hofstatter E, Aktas B, Pusztai L. Multigene prognostic tests in breast cancer: past, present, future Breast Cancer Res 2015;17:11.

13. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. J Natl Cancer Inst 2009;101: 1446–52.

14. Mamounas EP, Tang G, Fisher B, Paik S, Shak S, Costantino JP, et al. Association between the 21-gene recurrence score assay and risk of locoregional recurrence in node-negative, estrogen receptor-positive breast cancer: results from NSABP B-14 and NSABP B-20. J Clin Oncol 2010; 28:1677–83.

15. Pertschuk LP, Axiotis CA, Feldman JG, Kim YD, Karavattayhayyil SJ, Braithwaite L. Marked intratumoral heterogeneity of the proto-oncogene her-2/neu determined by three different detection systems. Breast J 1999;5:369–74.

16. Chung GG, Zerkowski MP, Ghosh S, Camp RL, Rimm DL. Quantitative analysis of estrogen receptor heterogeneity in breast cancer. Lab Invest 2007;87:662–9.

17. Navin N, Krasnitz A, Rodgers L, Cook K, Meth J, Kendall J, et al. Inferring tumor progression from genomic heterogeneity. Genome Res 2010;20: 68–80.

18. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 2012;366:883–92.

19. Szerlip NJ, Pedraza A, Chakravarty D, Azim M, McGuire J, Fang Y, et al. Intratumoral heterogeneity of receptor tyrosine kinases EGFR and PDGFRA amplification in glioblastoma defines subpopulations with distinct growth factor response. Proc Natl Acad Sci U S A 2012;109:3041–6.

20. Marusyk A, Tabassum DP, Altrock PM, Almendro V, Michor F, Polyak K. Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. Nature 2014;514:54–8.

21. Jankowitz RC, Lee AV. The evolving role of multi-gene tests in breast cancer management. Oncology (Williston Park) 2013;27:210, 212, 214.

22. Drury S, Salter J, Baehner FL, Shak S, Dowsett M. Feasibility of using tissue microarray cores of paraffin-embedded breast cancer tissue for measurement of gene expression: a proof-of-concept study. J Clin Pathol 2010;63:513–7.

23. Barry WT, Kernagis DN, Dressman HK, Griffis RJ, Hunter JVD, Olson JA, et al. Intratumor heterogeneity and precision of microarray-based predictors of breast cancer biology and clinical outcome. J Clin Oncol 2010; 28:2198–206.

24. Gulati S, Martinez P, Joshi T, Birkbak NJ, Santos CR, Rowan AJ, et al. Systematic evaluation of the prognostic impact and intratumour heterogeneity of clear cell renal cell carcinoma biomarkers. Eur Urol 2014;66: 936–48.

25. Cohen DA, Dabbs DJ, Cooper KL, Amin M, Jones TE, Jones MW, et al. Interobserver agreement among pathologists for semiquantitative hormone receptor scoring in breast carcinoma. Am J Clin Pathol 2012;138: 796–802.

26. Waggott D, Chu K, Yin S, Wouters BG, Liu FF, Boutros PC. NanoString-Norm: an extensible R package for the pre-processing of NanoString mRNA and miRNA data. Bioinformatics 2012;28:1546–8.

27. Acs G, Esposito NN, Kiluk J, Loftus L, Laronga C. A mitotically active, cellular tumor stroma and/or inflammatory cells associated with tumor cells may contribute to intermediate or high Oncotype DX Recurrence Scores in low-grade invasive breast carcinomas. Mod Pathol 2012;25: 556–66.

28. Baehner F, Quale C, Pomeroy C, Cherbavaz C, Shak S. Biopsy cavities in breast cancer specimens: their impact on quantitative RT-PCR gene expression profiles and recurrence risk assessment. Mod Pathol 2009: 28A–9A.

Clinical Cancer Research

AACR American Association for Cancer Research

Intratumor Heterogeneity Affects Gene Expression Profile Test Prognostic Risk Stratification in Early Breast Cancer

Rekha Gyanchandani, Yan Lin, Hui-Min Lin, et al.

Updated version	Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-15-2889
Supplementary Material	Access the most recent supplemental material at: http://clincancerres.aacrjournals.org/content/suppl/2016/05/14/1078-0432.CCR-15-2889.DC1

Cited articles	This article cites 27 articles, 11 of which you can access for free at: http://clincancerres.aacrjournals.org/content/22/21/5362.full#ref-list-1
Citing articles	This article has been cited by 5 HighWire-hosted articles. Access the articles at: http://clincancerres.aacrjournals.org/content/22/21/5362.full#related-urls

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, use this link http://clincancerres.aacrjournals.org/content/22/21/5362. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.