Laboratory Investigation (2012) 92, 1342–1357
© 2012 USCAP, Inc All rights reserved 0023-6837/12 $32.00

Evaluating tumor heterogeneity in immunohistochemistry-stained breast cancer tissue

Steven J Potts[1], Joseph S Krueger[1], Nicholas D Landis[1], David A Eberhard[2], G David Young[1], Steven C Schmechel[3] and Holger Lange[1]

Quantitative clinical measurement of heterogeneity in immunohistochemistry staining would be useful in evaluating patient therapeutic response and in identifying underlying issues in histopathology laboratory quality control. A heterogeneity scoring approach (HetMap) was designed to visualize a individual patient's immunohistochemistry heterogeneity in the context of a patient population. HER2 semiquantitative analysis was combined with ecology diversity statistics to evaluate cell-level heterogeneity (consistency of protein expression within neighboring cells in a tumor nest) and tumor-level heterogeneity (differences of protein expression across a tumor as represented by a tissue section). This approach was evaluated on HER2 immunohistochemistry-stained breast cancer samples using 200 specimens across two different laboratories with three pathologists per laboratory, each outlining regions of tumor for scoring by automatic cell-based image analysis. HetMap was evaluated using three different scoring schemes: HER2 scoring according to American Society of Clinical Oncology and College of American Pathologists (ASCO/CAP) guidelines, H-score, and a new continuous HER2 score ($HER2_{cont}$). Two definitions of heterogeneity, cell-level and tumor-level, provided useful independent measures of heterogeneity. Cases where pathologists had disagreement over reads in the area of clinical importance ($+1$ and $+2$) had statistically significantly higher levels of tumor-level heterogeneity. Cell-level heterogeneity, reported either as an average or the maximum area of heterogeneity across a slide, had low levels of dependency on the pathologist choice of region, while tumor-level heterogeneity measurements had more dependence on the pathologist choice of regions. HetMap is a measure of heterogeneity, by which pathologists, oncologists, and drug development organizations can view cell-level and tumor-level heterogeneity for a patient for a given marker in the context of an entire patient cohort. Heterogeneity analysis can be used to identify tumors with differing degrees of heterogeneity, or to highlight slides that should be rechecked for QC issues. Tumor heterogeneity plays a significant role in disconcordant reads between pathologists.

Laboratory Investigation (2012) **92,** 1342–1357; doi:10.1038/labinvest.2012.91; published online 16 July 2012

KEYWORDS: breast cancer; digital pathology; HER2; immunohistochemistry; pathology; tumor heterogeneity

The companion diagnostic approach seeks to dictate therapeutic strategy based on a molecular description of a patient's disease, with drugs targeting the HER2 receptor one of the most widely adopted examples. There are well-established guidelines for selecting patients for anti-HER2 adjuvant therapies in breast cancer treatment. However, even with patient selection, many trastuzumab-treated patients do not benefit from therapy, as their disease progresses or becomes recurrent. For example, about 1/3 of breast cancer patients given Herceptin fail to respond (*de novo* resistance), and about 1/5 of the responsive patients become refractory (acquired resistance).[1] The proportion of patients who are not responsive to therapy, even with the inclusion of a companion diagnostic to predict patient response, indicates that the current approaches to treatment strategy and patient selection may not be as robust as possible.

The current HER2 immunohistochemistry (IHC) score methodology does not account for heterogeneity. Since 2007, the American Society of Clinical Oncology and College of American Pathologists (ASCO/CAP) have recommended

[1]Flagship Biosciences, Westminster, CO, USA; [2]Department of Pathology and Lineberger Comprehensive Cancer Center, University of North Carolina Chapel Hill, NC, USA and [3]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA
Correspondence: Dr SJ Potts, PhD, Flagship Biosciences, 10955 Westmoor Dr., Suite 400, Westminster, CO 80021, USA.
E-mail: steve@flagshipbio.com

Received 30 October 2011; revised 24 March 2012; accepted 2 April 2012

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

specific guidelines for HER2 scoring.[2] These guidelines call for a consistent process of sample preparation and staining (IHC) or hybridization (FISH) approaches, as well as score reporting. The ASCO guidelines also suggest using the terminologies 'positive' (3+), 'equivocal' (2+), or 'negative' (+1 to 0) to define HER2 scoring. According to the immunohistochemical (IHC) scoring methodology, the difference between an 1+ and 2+ score is a description of a 'faint' (0/1+) compared to a 'weak-to-moderate' (2+) membrane staining in more than 10% of the tumor cells. In contrast, a 3+ score is described as a uniform intense membrane staining of >30% of tumor cells. Thus, this widely used scoring approach is semiquantitative, as it relies on a threshold percentage of positive cells to determine the score. Importantly, this overall score does not include any additional information about the percentages of tumor cells that score beyond the threshold levels. The ASCO/CAP guidelines for HER2 FISH scoring also relies on a stratified HER 2/CEP17 ratio for the score.

This lack of information about variability within the tumor, or between tumors with the same score, blinds clinicians to a potential readout that could represent a biology responsible for non-effective responses to therapy. It is intuitive that differential cell populations within or between tumors could contribute to clinical refraction to therapy and thereby affect patient outcomes. All potential factors within individual patients that contribute to a lack of response are not known, but cancer biologists have long hypothesized that such disparate populations within the tumor can be selected for outgrowth and emerge as a resistant tumor. This concept of tumor heterogeneity leading to drug resistance was debated as early in the 1950s as the 'Greenstein Hypothesis', and has become part of cancer biology doctrine.[3] In more recent times, as more targeted therapies are being developed, the issue of tumor heterogeneity has re-emerged as a factor significant to clinical strategy. Thus, there is a need for clinical evaluation of tumor heterogeneity that is aligned with the emerging understanding of cancer biology.

Studies of intratumoral heterogeneity from the same site demonstrate that heterogeneity can affect prognosis in 2+ scored tissues.[2,4] Another study found 16% of 3+ score cases exhibiting tumor heterogeneity.[5] A recent case[6] documented the personal significance of tumor heterogeneity, where a patient with invasive breast carcinoma demonstrated HER2 gene amplification on core biopsy, but relapsed while on adjuvant trastuzumab therapy after mastectomy, dying 15 months after diagnosis. Often, metastases harvested at autopsy demonstrated no evidence of HER2 gene amplification, but retrospective examination of the carcinoma in the patient's mastectomy specimen revealed only focal HER2 amplification within the tumor, localized to the region of the prior core biopsy site, highlighting the importance of both adequate sampling and awareness of heterogeneity issues. Another case was noted[7] where a patient with breast cancer had areas of the tumor that were 3+ positive and negative for HER2/neu by IHC, adjacent to each other. These cases represent an underlying biology of tumor heterogeneity, which contributes to the clinical outcome.

The assessment of HER2 protein expression status in breast cancer provides a useful working example of tumor heterogeneity for future biomarker studies. There are substantial biological and clinical implications of intratumor clonal heterogeneity.[8,9] This heterogeneity may reside within a single tumor (intratumoral), or between tumors at different sites (intertumoral). Consequently, researchers have attempted to identify the levels of clinically observed heterogeneity in multiple studies of HER2/neu in breast carcinoma, the results of which are summarized in Table 1. Eight different studies of HER2 heterogeneity between primary breast tumor and metastasis demonstrate the low disconcordance rates between these: 0 and 13%, with the majority of studies under 5% disconcordant. Thus, determining the disparity between primary tumor and metastases may not be of high clinical priority. However, one recent study found disconcordance rates of 14% between core needle and excisional biopsies, suggesting that tumor heterogeneity could contribute to misclassification utilizing needle biopsies.[10] The ASCO/CAP guidelines define HER2 genetic heterogeneity in FISH testing as >5%, and noted that the incidence of intratumor heterogeneity by this definition ranged in the literature from 5 to 30%.[11]

Accordingly, the ability to measure tumor heterogeneity may assist clinicians in verifying the predictive value of the HER2 score. It is critically important that the profession begin to develop improved approaches of reporting heterogeneity in samples. In the discipline of stereology, unbiased sampling is obtained by utilizing an entire tissue block, and randomly sampling both the sections and regions within a section to eliminate bias.[12] However, a heterogeneity measurement seeks to start with the entire population, and then sample in an unbiased manner to then determine a representative variation. In addition, in clinical trials, it is difficult and nearly impossible to obtain the blocks required for stereology sampling, so the industry is left with dealing with one or several tissue slides as the specimen from which to obtain heterogeneity assessments.

As pathology evolves into a more digital and quantitative discipline, the challenge of quantifying tumor heterogeneity comes more clearly into focus. Whole slide imaging and quantification techniques for the evaluation of IHC biomarkers facilitate an approach for measuring tumor heterogeneity. The ability to distinguish and score individual cells across the whole tissue provides sufficient content to assess reliably diversity of a biomarker within the sample. Combined with a mathematical approach to describe a measure of variation within the sample, a heterogeneity index can be created. In this report, a novel, functional approach is described that assigns a numerical value to HER2 score diversity within a tumor sample, and thus serves to quantify

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

Table 1 Studies of tumor heterogeneity in HER2/neu

Tumor type and stage	Analytical approach and tissue extraction technique	Result
Evaluation of heterogeneity on different samples from the same tumor site taken at the same time		
38 Invasive breast carcinomas grade 2 tumors evenly split on FISH gene amplification (4)	Evaluated IHC and FISH on additional slides of the same tumor	For the 2+ cases with no FISH gene amplification, 72% had a 1+ IHC score on at least one additional slide, 22% remained 2+ and 6% had one slide scored 1+. For the 2+ cases with FISH gene amplification, 55% had a 3+ IHC score on at least one additional slide, 40% remained 2+ in IHC, and 5% had a slide scored 1+. The authors noted that significant intratumoral heterogeneity accounts for many breast cancers with 2+ HER2 protein expression
44 Breast carcinomas (all grades) and 5 normal breast tissues (10)	1-mm TMA cores in triplicate, evaluating seven proteins by IHC: HER2, ER, PR, E-cad, EGFR, p53, and MIB-1	Intratumoral heterogeneity was seen with ER, PR, HER2, p53, and MIB-1, but not with E-cad and EGFR. Results indicate that core needle biopsies are problematic as indicators of status for the entire tumor
21 Breast cancer tumors (11)	HER2, topoisomerase IIα, c-myc, and cyclinD1 were evaluated with q-PCR in macroscopically and microscopically separate areas of individual tumors	HER2 mRNA expression by q-PCR had much lower levels of heterogeneity than the other genes, with heterogeneity occurring in 36% of amplified cases. C-myc and cyclinD1 exhibited heterogeneity in 100% and 83% of cases, respectively
48 Cases of grade 3 breast cancer were analyzed (5)	Multiple sites of morphologically similar tumor were analyzed with HER2 FISH	Intratumoral heterogeneity for HER2/neu gene amplification was demonstrated in only 5 (16%) of 31 cases, where morphologically similar areas of a single tumor were analyzed
Heterogeneity between needle core biopsy and excisional biopsy		
100 Patients with needle core biopsies and subsequent excisional biopsy samples (12)	HER2 FISH and IHC	The concordance rate between FISH results determined on the needle core biopsy and subsequent excisional biopsy of the same tumor was 86%. Of the 15 patients who received neoadjuvant chemotherapy, 93 and 87% had no change in HER2/neu status as determined by IHC or FISH, respectively, in the excisional biopsy specimen when compared with that determined on the prior core biopsy sample
Frequency of heterogeneity in a test population, determined by FISH		
Single institution results of all testing in calendar year 2006, total of 742 consecutive cases of breast carcinoma (13)	Reported on clinical results as run by CAP guidelines	Genotypic heterogeneity, defined as $>5\%$ but $<50\%$ of the tumor cells demonstrating HER2 gene amplification, was observed in 5% (40/7242) of the cases
Heterogeneity comparison between DCIS and cancer		
Multiple foci from 23 breast tumors with DCIS only and 20 cases with synchronous DCIS and infiltrating cancer (14)	Multiple foci extracted and microsatellite markers by PCR for allelic losses in individual foci for loci in chromosomes 6q, 9p, 11q, 13q, 16q, 17q, and 17p	Patterns of allelic losses were generally conserved in the synchronous infiltrating tumors, supporting the paradigm that infiltrating tumors are clonally derived from the *in situ* lesions. However, in 8 (40%) of the 20 DCIS cases with invasive cancer, heterogeneous patterns of allelic loss at one or more loci were observed
Concordance between primary site and metastasis		
44 Breast cancer patients with asynchronous metastasis or recurrence (15)	Samples from primary breast cancers and metastatic lesions were analyzed for p53, ER, PR, and HER2 IHC and HER2 FISH	Discordance rates between primary and secondary tumor were 4.5% for HER2 with IHC, with FISH results consistent with IHC

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

Table 1 Continued

Tumor type and stage	Analytical approach and tissue extraction technique	Result
Concordance between primary site and metastasis		
21 Breast cancer patients with metastasis (16)	Samples from primary breast cancers and metastatic lesions were analyzed with IHC for HER2, p53, ER, and PR	Discordance rate between primary and secondary tumors was 0% for HER2 and p53. Expression levels in breast cancer cells were almost unchanged as the disease progressed, regardless of hormone receptor status
47 Breast cancer patients with metastasis along with literature review of other similar studies (17)	Samples from primary breast cancers and lymph node metastatic lesions were analyzed with IHC and CISH/FISH for HER2	No cases of drastic changes in HER2 expression between primary and lymph node metastasis were reported. Authors conclude that breast cancer lymph node metastases generally overexpress HER2 to the same extent as the corresponding primary, including distant metastases
58 Breast cancer patients with metastasis (18)	Samples from primary breast cancers and metastatic lesions were analyzed by IHC and FISH for HER2	Discordance rate between primary and metastatic tumors was 14% (8 of 58 patients), with the majority (7) positive in metastasis and negative in primary. FISH results were concordant with IHC for the data set
789 Breast cancer patients with metastasis (19)	Samples from primary breast cancers and recurrent tumors were analyzed by IHC for ER, PR, and HER2, and by FISH for HER2	Discordance rate for ER, PR, and HER2 was 18.4%, 40.3%, and 13.6%, respectively. Patients with concordance have significantly better post-recurrence survival than discordant cases
205 Breast cancer cases from 20 institutions with matching primary and recurrent tumor (20)	Samples from primary and recurrent were analyzed for ER, PR, and HER2 by IHC	Discordance rates for ER, PR, and HER2 were 10.2%, 24.8%, and 2.9%, respectively, with no significant difference in locoregional or distant recurrence. The switch in receptor status led to a change in the subsequent treatment plan for 17.5% of the patients
107 Patients with primary breast cancer and at least one distant metastatic lesion (21)	HER2 levels were analyzed by IHC and FISH	Discordance rate of 6% with IHC, with all six discordant cases showing greater HER2 overexpression in the metastatic tumor. By FISH the discordant rate was 5%, and the discordant cases were split between under- and overexpression in the metastatic tumor
Breast cancer primary sites and matched metastatic lymph nodes (22)	HER2, p53, bcl-2, topoisomerase IIα, HSP27, and HSP70 were evaluated by IHC	Disconcordance rates were 2% for HER2, 6% for p53, 15% for bcl-2, 19% for topoisomerase IIα, 24% for HSP27, and 30% for HSP70
Heterogeneity before and after treatment		
39 Patients with locally advanced breast cancers who received neoadjuvant chemotherapy and 60 breast cancer patients who did not receive neoadjuvant chemotherapy (23)	IHC for HER2 was performed on paraffin sections of the core biopsy before treatment and the excised specimen following chemotherapy	HER2 IHC scores decreased in 28.5% (15/39) of patients receiving neoadjuvant chemotherapy compared to 11.7% (7/60) of patients in the control ($P<0.013$). HER2/neu IHC status changed from strongly positive to negative (3+ to 0) in 5 of 39 (12.5%) in the study group and in 2 of 60 (3.3%) in the control group ($P=0.104$)

heterogeneity. This output can be included with other digital pathology-based measurements of IHC biomarkers to provide a more contextual value to the numerical score. Two definitions are introduced to further assist with describing heterogeneity cell-level and tumor-level heterogeneity (Figure 1). Cell-level heterogeneity (Het_{cell}) is the variability of cells within a nest of tumor, and tumor-level heterogeneity (Het_{tumor}) is the variability of nests of cells across an entire tumor. There is only one score per slide for Het_{tumor}, but as each nest or sampled region in a tumor has its own Het_{cell} score, it is challenging to combine these into a single measure for a given slide. Thus, several approaches are examined to aggregate measures of cell-level heterogeneity across a slide.

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

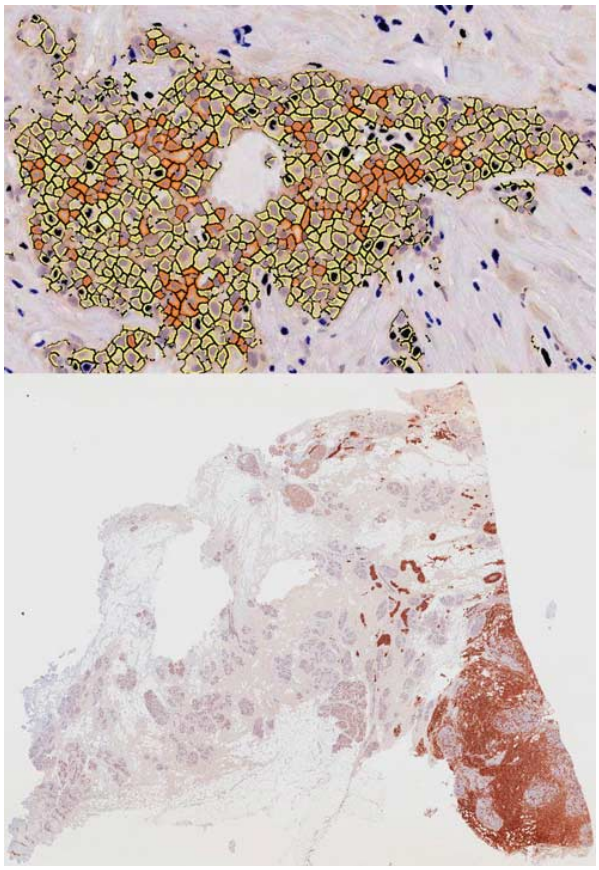

Figure 1 Definitions of cell-level (above) and tumor-level heterogeneity (below). Slide-level heterogeneity is a sampling substitute for tumor-level heterogeneity. The below figure also illustrates some contributions of anatomic heterogeneity, as parts of the lesser stained areas are ductal carcinoma *in situ* (DCIS).

Numerical Indices of Tumor Cell Diversity

Diversity measurement is a well-established field in the ecological sciences, and numerous approaches to quantifying the variability of species have been utilized in this discipline. Ecologists will describe diversity in terms of richness and evenness, and each can be ranked differently depending on the weighting of these concepts. For example, one area might have only two species, each covering half the area. The second area might have six different species, with one dominant species covering 95% of the area, and the other five each only covering 1%. Defined in terms of richness, the second area with eight different species would be considered more diverse. Defined in terms of evenness of distribution, the first area would be more diverse as it avoids having one type dominating over all others. Two commonly used diversity indices are the Shannon index[13] and the Simpson index,[14] for measuring plant and animal species diversity. The Shannon index of diversity is defined as:

$$\text{Shannon} = \sum_{i=1}^{N} p_i \times \ln p_i$$

where N is the number of biological types and p_i the proportional abundance of the ith type. This index, ranging in theory from 0 to infinity, estimates the average uncertainty in predicting to which species type a randomly selected subunit of area belongs. The Simpson index is defined as:

$$\text{Simpson} = \sum_{i=1}^{N} p_i \times p_i$$

Producing values from 0 to 1, Simpson's index defines the probability that two randomly selected equal-sized subunits of terrain belong to different species. A recent evaluation of tumor heterogeneity pioneered the use of both Shannon and Simpson indices in evaluating 8q24 copy number gain in both CD24+ and CD44+ cell populations in ductal carcinoma *in situ* and invasive regions of tumors.[15] Copy numbers at each of three levels were considered as separate 'species' and the indices applied to deliver a measure of heterogeneity within each sample. Two distinct tumor subtypes of high and low diversity of 8q24 copy number, as measured by the Shannon index, and the group with lower diversity contained fewer samples of HER2+ tumors. There was no difference between diversity of the luminal A tumors and the normal cells, although basal-like tumors tended to have higher diversity scores. In this study, few qualitative differences were seen between Shannon and Simpson indices, although the data set were small. The Shannon index tends to blur distinctions of species richness and evenness, while the Simpson index can be dominated by the most abundant species in the population.

The disadvantage of both Shannon and Simpson indices is that they do not account for taxonomic distance between species. In the world of clinical anatomic pathology, most cells are binned and scored as one of three or four classes. In HER2 scoring methodology, pathologists (or pathologist-trained computer programs) score cells as populations of either 0+, 1+, 2+, or 3+ intensity. Consider two regions: Region A with ten 0+ cells and ten 3+ cells, and region B with ten 1+ cells and ten 2+ cells. Clearly, Region A has a higher level of heterogeneity than Region B, but Shannon and Simpson indices would score these as equal heterogeneity. To overcome this problem, an ecological diversity approach known as Rao's quadratic entropy (QE)[16] was used. A distance matrix is incorporated in the diversity index, where, for example, a difference between a 0+ and 3+ cell would be weighted a '3', and a 1+ to 2+ would be weighted a '1'. When all weights are the same, the scoring schemes tend to be equivalent to those mentioned previously.

The equation is as follows:

$$\text{QE} = \sum_{i>j=1}^{N} d_{ij} p_i p_j$$

$\bar{D}$ =

	3+	2+	1+	0+
3+	0	1	2	3
2+	1	0	1	2
1+	2	1	0	1
0+	3	2	1	0

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.