

Evaluating Tumor Heterogeneity in Immunohistochemistry Stained Breast Cancer Tissue

Steven J. Potts,¹ Joseph S. Krueger², Nicholas Landis³, David A. Eberhard², Stephen C. Schmeche³, G. David Young¹, and Holger Lange¹

¹Flagship Biosciences, Flagstaff, AZ, United States: steve@flagshipbio.com, holger@flagshipbio.com, nlandis@flagshipbio.com, jrueger@flagshipbio.com, dave@flagshipbio.com

²Dept. of Pathology and the Lineberger Cancer Center, University of North Carolina, Raleigh, North Carolina, United States: daveberhard@gmail.com

³Lab Medicine and Pathology, University of Minnesota, Minneapolis, Minnesota, United States: schme004@umn.edu

San Antonio Breast Cancer Symposium— December 6–10, 2011

Introduction

- 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, or their disease progresses or becomes recurrent. 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 are insufficient.
- One primary reason may be that the current HER2 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 specific guidelines for HER2 scoring on a scale of 0-3+, depending on the overall percent of cells which are positive for HER2. This score does not include any additional information about the percentages of tumor cells that score beyond the threshold levels.
- This lack of information about variability within the tumor, or between tumors who have the same score, blinds clinicians to a potential readout which could represent a biology responsible for effective responses to therapy. It is intuitive that differential cell populations within or between tumors could contribute to clinical refractoriness to therapy and thereby affect patient outcomes. This idea is supported by several published studies of the effect of intratumoral heterogeneity on prognosis, the results of which are summarized below.
- Accordingly, the ability to clinically 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.
- 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 image quantification techniques for the evaluation of IHC biomarkers facilitates an approach for measuring tumor heterogeneity.

Study Context: Quantitative clinical measurement of heterogeneity in immunohistochemistry staining would be useful in both evaluating patient therapeutic response and identifying underlying issues in histopathology laboratory quality control.

Study Objective: To create a heterogeneity scoring approach (HetMap) that allows visualization of a patient's tumor heterogeneity in the context of an IHC score.

Study Design: We combined the use of ecology diversity statistics with HER2 scoring 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). We evaluated the approach using 200 specimens from clinical breast cancer cases (even distribution across slides), across two different CLIA labs. Three board-certified MD pathologists each blindly drew 9-30 regions per slide of "representative" tumor to run an Aperio cleared algorithm of HER2 on each region, with CLIA controls and validation. The numerical HER2 scores were determined according to ASCO/CAP guidelines, standard H-score, as well as a new continuous HER2 score (HER2_{cont}). Heterogeneity was evaluated using different scoring schemes as described, and the influence of heterogeneity on pathologists score as well as the influence of pathologist scoring region on the various scoring approaches determined.

Defining Heterogeneity

Variability

Tissue	Count	Number of sections sampled in three	Number of regions	Biologic variability present in three additional samples of control
Variable (different controlling regions averaged across a single tissue section)				
Human breast	HER2 IHC staining	150 (50 regions x 3)	180	17%
Human breast	ER2 percent positive cells	150 (50 regions x 3)	180	11%
Human breast	PR percent positive cells	150 (50 regions x 3)	180	33%

Table 1. Tissue cross-sections exhibit high levels of biological variability. The table shows typical coefficients of variation by sampling multiple small related sections across a large tissue section using a pathologist. CV values depend on the marker, and are high even when the comparison diagnostic scoring rules are applied.

Heterogeneity

Tumor Level Heterogeneity: Het_{tumor}

Cell Level Heterogeneity: Het_{cell}

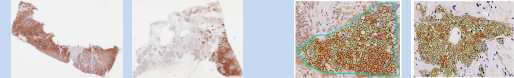


Figure 1. Tumor heterogeneity. Tumor level heterogeneity is the differential expression of a marker within the entire tumor section. The tumor on the left shows very strong HER2 staining, with low heterogeneity across the entire section. The tumor on the right has high HER2 staining in the lower right region, with variation in staining across the remainder of the tumor. Both regions received a summary +3 HER2 score according to ASCO/CAP scoring guidelines, despite clearly having differences in tumor profiles.

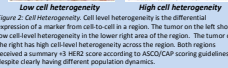


Figure 2. Cell heterogeneity. Cell level heterogeneity is the differential expression of a marker from cell-to-cell in a region. The tumor on the left shows low cell-level heterogeneity in the lower right area of the region. The tumor on the right has high cell-level heterogeneity across the region. Both regions received a summary +3 HER2 score according to ASCO/CAP scoring guidelines, despite clearly having different population dynamics.

Approach to Scoring Heterogeneity in Tissue Sections

Ecological Diversity Applied to Tumor Heterogeneity

Consider a forest:

Each species of tree has a proportional abundance in a sample of that forest

A forest may be high in diversity (deciduous trees) or low in diversity (tropical trees).

However, there is an important difference between the types of trees within a forest

Consider a breast biopsy:

Each type of cancer cell has a proportional abundance in a sample of that tissue section

A region may be high in diversity (HER2⁺ cells) or low in diversity (HER2⁻ cells).

However, there is an important difference between the types of cells within a tissue

Use of a Continuous HER2 Score (HER2_{cont})

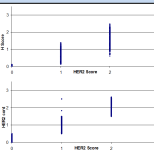


Figure 3. Use of a HER2 continuous (HER2_{cont}) score. Since the Aperio HER2 scoring analysis provides a score for each individual cell, a score which conveys cell-by-cell information was sought. The H-scores of the regions were non-linear with the HER2 score, so we developed the HER2_{cont} score to retain linearity. The non-linearity of the H-score is due to the heavier weighting of higher intensity staining over lower intensity staining to calculate the score.

Measuring Heterogeneity of a Tissue Section

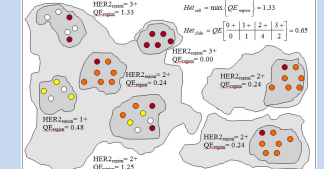


Figure 4. Visual representation of patient heterogeneity (Het_{tumor}) versus the slide level heterogeneity (Het_{cell}) scoring approach. Het_{tumor} is the QE value of all these regions across the whole tissue section. The variability in cell types as well as the dynamic range of scores for each cell determine the QE score for each region. Het_{cell} is the QE value of all these regions across the whole tissue section. The variability and dynamic range of Het_{cell} values determines the QE score for the entire tumor section, or Het_{tumor}.

Patient Molecular Profiling with HetMap™

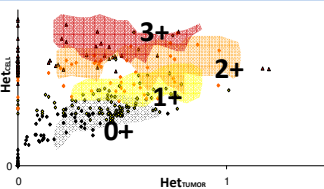


Figure 5. A representation of the data using HetMap. HetMap is a graphical representation of the type of heterogeneity observed in each patient as it relates to a HER2 score. In general, the Het_{tumor} increases with the HER2 score, due to increased complexity of the sample (more 3+ cells). In contrast, the Het_{cell} score is independent of the HER2 and Het_{tumor} scores. However, this graph indicates that there are specific subsets of patients within each score who either have an abnormally high degree of heterogeneity across the whole tumor. The current understanding of tumor heterogeneity on clinical response suggests that this measure has potential clinical value, as discussed in figure 7.

What is a Good Way to Measure Heterogeneity?

- Heterogeneity may occur between neighboring cells, or in different parts of the tumor.
- Normally, pathologists will take an average expression across a slide by selecting several random regions to score and combined the scores from each region for a final tumor score.
- Important information is lost about the richness of diversity of individual cells or regions of the tumor. Richness of diversity also considers not only the number of species, but how different the species are from each other (taxonomic distance).
- Ecological measurements of diversity consider functional differences in species. As a palm tree and oak tree are vastly different, so are HER2- and HER2+ cells. Considerations of these differences are significant to tumor biology and clinical outcome.

Rao's Quadratic Entropy (QE) Calculation of Rao's Quadratic Entropy compares the proportional abundance of a species and factors in the difference between species as it pertains to an attribute. This approach is well suited for evaluating differences in cells (species) characterized by a molecular diagnostic marker (IHC stain) in an ecosystem (tumor).

$$QE = \sum_{i,j} p_i p_j$$

Combining Heterogeneity and HER2 Scoring Methods

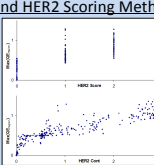


Figure 6. Use of Mean (QE_{cont}) with the HER2_{cont} score. Utilizing the maximum QE value for each HER2_{cont} score showed the best dynamic range of values. Utilizing the maximum score, by definition, captures the maximum entropy observed, enabling clearer visualization of the relationship between HER2 score and heterogeneity. The dataset consisted of 8549 tumor regions.

Influence of Heterogeneity on Scoring

Effect on Pathologist Concordance

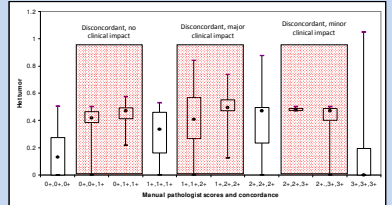


Figure 7. Effect of heterogeneity on pathologist concordance. The box plot demonstrates the effect of heterogeneity on the ability of three different pathologists to agree on HER2 scores across 200 samples. Each box represents the subset of samples given a specific score by the three pathologists. Concordant scores (white areas) occur when all three pathologists were in agreement (0+0+0, 1+1+1, 2+2+2, 3+3+3), and discordant scores (pink boxes) occur when all three pathologists are not in agreement (0+0+1, 0+1+1, 1+1+1, 1+1+2, 1+2+2, 2+2+3, 2+3+3, 3+3+3). As expected, none of the approaches to measure cell-level heterogeneity (Het_{cell}) significant by correlated with discordant scoring of samples (not shown). Tumor level heterogeneity (Het_{tumor}) was a statistically significant (p<0.01) predictor of discordance between pathologists in all score bins. Importantly, discordant scores in the "optimal HER2" range (1+1+2, 2+1+2, 2+2+3) have potentially major clinical impact as this cutoff determines whether or not patients will receive Trastuzumab. Although Het_{cell} was also predictive of discordant scores in the "High HER2" range (2+2+3 or 2+3+3), this has a more modest potential clinical impact as 2+ patients receive HER2 FISH reflex testing prior to trastuzumab therapy. As such, although Het_{tumor} was also predictive of discordant scores in the "low HER2" range (0+0+1, 0+1+1), this does not have clinical impact as both 0+ and 1+ patients do not receive trastuzumab therapy.

Impact of Pathologist Choice of Regions

	Het _{tumor}	HER2 _{cont}	Het _{cell}
Lab A. Sidev	.13	.10	.10
Lab B. Sidev	.09	.11	.09
Lab A. Average	.41	1.2	.64
Lab B. Average	.31	1.1	.57
Lab A. CV	30%	8%	16%
Lab B. CV	25%	9%	15%

Figure 8. Effect of pathologist choice of scoring regions on the various scores. The standard deviation, average, and coefficient of variation values were determined for the values of tumor level heterogeneity (Het_{tumor}), the HER2 continuous score (HER2_{cont}), and cell level heterogeneity (Het_{cell}) in the three sets of regions drawn by the three pathologists for each slide. The variation in HER2_{cont} between pathologists drawing their own regions was low in both labs, validating this scoring methodology. This was lower than the ASCO/CAP recommended concordance rate of 95%, but far better than published concordance rates for HER2 IHC tests around 20%. The CV values for Het_{tumor} ranged much higher, indicating the idea that this particular assessment of heterogeneity is inadequate due to the local nature of its determination, and should not be used. Het_{tumor} was far more impacted by pathologist choice of regions than the other measurements. It's likely that the number of regions sampled (10-15) may not be sufficient to make adequate determinations of tumor level heterogeneity. Relying on a methodology which samples all the tumor on a slide, such as digital quantification, is becoming possible and practical, and may be required for this type of analysis.

Conclusions

The assessment of HER2 status in breast cancer provides a useful working example of the importance for incorporating tumor heterogeneity measurements in biomarker studies.

- In this study we introduced two definitions to appropriately define heterogeneity – cell level (Het_{cell}) and tumor level heterogeneity (Het_{tumor}). Cell level heterogeneity is the variability of cells within a nest of tumor cells, and tumor level heterogeneity (Het_{tumor}) is the variability between the nests of tumor cells or regions across an entire tumor. We applied principles of diversity in ecology to mathematically score tumor heterogeneity using Rao's Quadratic Entropy (QE), which accounts for the difference (distance) between taxonomic species (0+ cell vs a 3+ cell).
- We examined several approaches to aggregating measures of cell level heterogeneity across a slide. We combined whole slide image analysis techniques to score HER2 staining in a tumor. We utilized a novel scoring paradigm, the HER2 continuous score (HER2_{cont}), along with a mathematical approach to describe a measure of variation within a sample to create a heterogeneity index for HER2 scores. The numerical value of this score quantifies the diversity within a tumor sample. This output can be included with other digital pathology-based measurements of IHC biomarkers to provide a more contextual value to the numerical score of a biomarker.
- Cell-level heterogeneity, reported as the maximum area of heterogeneity within a region, had little correlation with discordance in pathologist scoring. In contrast, tumor level heterogeneity, reported as the maximum area of heterogeneity across an entire tissue section, correlated with increased discordance in pathologist scoring. This potentially resulted in significant clinical impact in the cases of discordance of HER2 1+/2+ scores. Tumor-level heterogeneity measurements had dependence on the pathologist choice of regions, indicating that a whole-slide scoring approach, rather than representative regions, is required.
- 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 a useful means to identify tumors with higher degrees of heterogeneity, or to highlight slides that should be rechecked for QC issues. Incorporating a heterogeneity assessment will be critical in determining cutpoints for patient stratification in companion diagnostic development.

This presentation is the intellectual property of the authors. Contact steve@flagshipbio.com for permission to reprint and/or distribute.