Quantitative In Situ Measurement of Biomolecules for Companion Diagnostics

David L. Rimm M.D., Ph.D
Professor, Director of Translational Pathology
Dept. of Pathology
Yale University School of Medicine
Disclosures

• I am a consultant to Amgen, Applied Cellular Diagnostics, Avida Labs, Biocept, BMS, Cernostics, Genoptix/Novartis, Metamark Genetics, MD Agree, OptraScan, and Perkin Elmer

• Cepheid, Genoptix, Gilead Sciences, Kolltan and OncoplexDx fund research in my lab.
Outline for Presentation

• Using immunohistochemistry (IHC) or quantitative immunofluorescence (QIF) to measure protein on slides

• The challenge of a continuous biomarker – the HER2 example

• The challenge of defining the threshold of detection – The ER example
Immunoperoxidase stain

Uses of IPOX:
• Identification (binary)
• Reading/estimation (ordinal)
• Quantification (continuous)
The human eye is not a great tool for assessment of intensity.
Different Intensities of HER2 IHC Staining Observed Within HER2(+) Patient Population

Intensities of HER2 IHC Staining Observed in HER2+ Patients From the HERA Trial

Although there was a high proportion of positive staining tumor cells in all of the 3+ samples the range of staining intensity varied.

Slide provided by Mitch Dowsett

Zabaglo L, et al. Presented at 33rd SABCS; December 8-12, 2010; San Antonio, TX. Abstract PD10-01.
Inter-Observable Reproducibility
The inter-observer reproducibility was assessed using 40 invasive breast cancer cases (resection specimens) that were sectioned and provided to the 3 sites for staining and interpretation. The sections were blinded and randomized at each site prior to scoring. Interobserver agreement between the 2 independent study sites, was 87.5% (95% CI = 73.3% to 95.8%). The agreement between the independent study sites and LBN was 92.5% (95% CI = 79.6 to 98.4%) and 85% (95% CI = 70.1% to 94.3%), at site 1 and site 2, respectively.

Conclusion: We are willing to accept an error rate of up to 15%!?
“Real” Pathologist Reproducibility: 3 different pathologists read Allred scores on 100 cases of breast cancer

Path 1 v. Path 2: Kappa = 0.482 (p<0.001)
Path 1 v. Path 3: Kappa = 0.444 (p<0.001)
Path 2 v. Path 3: Kappa = 0.400 (p<0.001)

*Positive/Negative concordance: 92-95%

In RED, 9% misclassification rate

Mark Gustavson
Artifactual Grouping by Pathologists

Estrogen Receptor

Pathologist vs AQUA

Pathologist-based Score - Array 1

B

R = 0.884

Pathologist-based Score - Array 1

MET (HGF Receptor)

MET AQUA vs DAB IHC scoring by eye

y = 0.0234x + 32.376

R² = 0.1308

0 1000 2000 3000 4000 5000 6000

membrane cytoplasmic H-scor

Kaplan-Meier Cum. Survival Plot
for total membrane protein - score binarized at 50

Kaplan-Meier for YTMA 250-Z;
NSCLC stratified for cMet expression
5 year follow up

Membrane H-score below 60
Membrane H-score equal or above 50

Disease Specific Survival

Followup Time in Months

negative for cMet SP44
positive for cMet SP44

Time in months

Logrank (Mantel-Cox): p=0.4584
Commercially Available Pathology-Focused Image Analysis Platforms

- Definiens
- Visiopharm
- Perkin Elmer Vectra/INform
- Hamamatsu Nanozoomer
- Leica/Aperio
- Tissuegnostics
- Ventana (formerly Bioimagene)
- Optra SCAN
- Genoptix (formerly HistoRx) AQUA
- Many others sold as research platforms
AQUA®: objective analyte measurement on a tissue slide based on co-localization

Step 1: Mask (define region of interest, exclude stroma, blank space, etc) = colocalization with Cytokeratin for carcinoma

Step 2: Define the numerator (target) and denominator (compartment)

\[
\text{Concentration} = \frac{\text{Numerator}}{\text{Denominator}} = \frac{\sum \text{target intensity in compartment pixels}}{\sum \text{compartment pixel area}} = \text{AQUA score}
\]

Step 3: Calculate the AQUA score

Step 4: Convert to absolute concentration or normalize to set of uniform standards
Generating the AQUA® score

Combine DAPI image and cytokeratin image then cluster to assign each pixel to a subcellular compartment.

\[ \Sigma \text{compartment pixel area} \times \Sigma \text{target intensity in compartment pixels} = \text{AQUA score} \]
ER antibody used is 1D5

Standardized Index Array
Lowest positive vs. highest negative
Expanded “levels” to visualize threshold
contracted dynamic range of grayscale (max RGB input level 255→16)
Discordant classification of ER status in YTMA 130 cohort

19.7% AQUA positive/IHC negative (n=46)
0.4% AQUA negative/IHC positive (n=1)
Double negative (n=40)
Double positive (n=147)

Welsh et al, JCO 2011
Precision Results (ER-alpha)

<table>
<thead>
<tr>
<th></th>
<th>Pearson R</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 1 v. Day 2</td>
<td>.97</td>
<td>.97</td>
</tr>
<tr>
<td>Day 1 v. Day 3</td>
<td>.97</td>
<td>1.01</td>
</tr>
<tr>
<td>Day 2 v. Day 3</td>
<td>.98</td>
<td>1.04</td>
</tr>
</tbody>
</table>

%CV = 4.2

Mark Gustavson and Jason Christiansen
Comparison Between Methods (reproducibility)

Pathologist vs Pathologist

DAB semi vs DAB semi

QIF (AQUA) vs QIF (AQUA)

Elizabeth Zarrella and Veronique Neumeister
Comparison Between Methods

DAB semi vs QIF(AQUA)

Pathologist vs DAB semi

Pathologist vs QIF(AQUA)

Elizabeth Zarrella and Veronique Neumeister
Regression of IHC vs QIF scores for CB11 and SP3 in YTMA 263

**CB11**

Regression of HER2 CB11 chromogenic IHC and QIF scores in YTMA 263

\[ y = 31.383\ln(x) - 127.35 \]

Spearman's rho: 0.781

**SP3**

Regression of HER2 SP3 IHC vs QIF scores in YTMA 263

\[ y = 31.691\ln(x) - 103.84 \]

Spearman's rho: 0.708
Outline for Presentation

• Using immunohistochemistry (IHC) or quantitative immunofluorescence (QIF) to measure protein on slides
• The challenge of a continuous biomarker – the HER2 example
• The challenge of defining the threshold of detection – The ER example
FDA Cleared Companion Dx antibodies:

<table>
<thead>
<tr>
<th>Drug</th>
<th>Antigen</th>
<th>Company</th>
<th>Antibody</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trastuzumab</td>
<td>HER2</td>
<td>Ventana</td>
<td>4B5</td>
</tr>
<tr>
<td>Trastuzumab</td>
<td>HER2</td>
<td>Leica Biogenex</td>
<td>CB11</td>
</tr>
<tr>
<td>Trastuzumab</td>
<td>HER2</td>
<td>Dako</td>
<td>A0485</td>
</tr>
<tr>
<td>Endocrine Rx</td>
<td>Estrogen Receptor</td>
<td>Ventana</td>
<td>SP1</td>
</tr>
<tr>
<td>Endocrine Rx</td>
<td>Estrogen Receptor</td>
<td>Dako</td>
<td>1D5</td>
</tr>
</tbody>
</table>
What are we using for IHC?

<table>
<thead>
<tr>
<th>Antibody</th>
<th>N</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>4B5</td>
<td>568</td>
<td>55%</td>
</tr>
<tr>
<td>A0485</td>
<td>54</td>
<td>5%</td>
</tr>
<tr>
<td>CB11</td>
<td>24</td>
<td>2%</td>
</tr>
<tr>
<td>Herceptest</td>
<td>296</td>
<td>29%</td>
</tr>
<tr>
<td>SP3</td>
<td>52</td>
<td>5%</td>
</tr>
<tr>
<td>other</td>
<td>37</td>
<td>4%</td>
</tr>
<tr>
<td>Total</td>
<td>1031</td>
<td></td>
</tr>
</tbody>
</table>
Studies done “by eye” cannot see discordance
Data from NCCTG9831 from ongoing collaboration with Edith Perez and Karla Ballman
DAB shows same effect, but more subtle

CB11 (ICD)  SP3 (ECD)
# HER2 Standardization Array (YTMA 263)

<table>
<thead>
<tr>
<th></th>
<th>1+ AMP</th>
<th>1+, NOT AMP</th>
<th>2+, AMP</th>
<th>2+, NOT AMP</th>
<th>3+, AMP</th>
<th>3+, NOT AMP</th>
<th>0, NOT AMP</th>
<th>2+, NOT AMP</th>
<th>0, NOT AMP</th>
<th>1+, NOT AMP</th>
<th>NORMAL BREAST TISSUE (NBT)</th>
</tr>
</thead>
</table>
SP3 vs CB11 in YTMA 263 (Average AQUA Scores from 4 builds)

\[ y = 0.2584x + 73.894 \]

\[ R^2 = 0.6856 \]
Distribution of average HER2 CB11 (1:10000) scores in YTMA 263

- RED: 3+, AMPLIFIED
- BLUE: 2+, AMPLIFIED
- GREEN: HER2 amplified cell line

Cut-point is around 500

Daniel Carvajal
HER2 CB11 analysis

All : 1 Joinpoint

- Observed
- 1-54 Slope = 11.01
- 54-92 Slope = 350.84

Test For Number of Joinpoints

<table>
<thead>
<tr>
<th>Test Number</th>
<th>Null Hypothesis</th>
<th>Alternate Hypothesis</th>
<th>Numerator Degrees of Freedom</th>
<th>Denominator Degrees of Freedom</th>
<th>Number of Permutations</th>
<th>P-Value</th>
<th>Significance Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>0 Joinpoint(s)</td>
<td>1 Joinpoint(s) *</td>
<td>2</td>
<td>88</td>
<td>4500</td>
<td>0.0002222</td>
<td>0.0500000</td>
</tr>
</tbody>
</table>

Final Selected Model 1 Joinpoint(s)

Daniel Carvajal
Using Joinpoint to Assess Ab Sensitivity and Specificity to Predict FISH amplification in YTMA 263

<table>
<thead>
<tr>
<th>HER2 antibody</th>
<th>Joinpoint cut-point (AU)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Positive predictive value (%)</th>
<th>Negative predictive value (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB11</td>
<td>1299.3</td>
<td>94.12</td>
<td>85.48</td>
<td>78.05</td>
<td>96.36</td>
</tr>
<tr>
<td>A0485</td>
<td>1495.9</td>
<td>90.62</td>
<td>85.71</td>
<td>78.38</td>
<td>94.12</td>
</tr>
<tr>
<td>SP3</td>
<td>525.8</td>
<td>88.24</td>
<td><strong>98.41</strong></td>
<td>96.77</td>
<td>93.94</td>
</tr>
<tr>
<td>D8F12</td>
<td>767.6</td>
<td>90.91</td>
<td><strong>93.88</strong></td>
<td>90.91</td>
<td>93.88</td>
</tr>
</tbody>
</table>
# HeCOG 10/05 patient subgroups

## Table 2. Biological subgroups

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER(+) / PgR(+) / HER2(-)</td>
<td>366 (51.9)</td>
</tr>
<tr>
<td>ER(+) / PgR(+) / HER2(+)</td>
<td>90 (12.8)</td>
</tr>
<tr>
<td>ER(+) / PgR(-) / HER2(-)</td>
<td>66 (9.4)</td>
</tr>
<tr>
<td>ER(+) / PgR(-) / HER2(+)</td>
<td>23 (3.3)</td>
</tr>
<tr>
<td>ER(-) / PgR(-) / HER2(-) (TNBC)</td>
<td>68 (9.6)</td>
</tr>
<tr>
<td>ER(-) / PgR(-) / HER2(+)</td>
<td>69 (9.8)</td>
</tr>
<tr>
<td>ER(-) / PgR(+) / HER2(-)</td>
<td>15 (2.1)</td>
</tr>
<tr>
<td>ER(-) / PgR(+) / HER2(+)</td>
<td>7 (0.9)</td>
</tr>
<tr>
<td>Not determined</td>
<td>1 (0.1)</td>
</tr>
</tbody>
</table>

## Table 3. Treatment subgroups

<table>
<thead>
<tr>
<th>Treatment group</th>
<th>N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER(+) with hormonotherapy</td>
<td>527 (96.7 of ER[+])</td>
</tr>
<tr>
<td>ER(+) without hormonotherapy</td>
<td>18 (3.3 of ER[+])</td>
</tr>
<tr>
<td>HER2(+) with trastuzumab</td>
<td>186 (98.4 of HER2[+])</td>
</tr>
<tr>
<td>HER2(+) without trastuzumab</td>
<td>3 (1.6 of HER2[+])</td>
</tr>
<tr>
<td>Hormonotherapy and trastuzumab</td>
<td>111 (15.7 of total)</td>
</tr>
<tr>
<td>ER(-) with hormonotherapy</td>
<td>0 (0)</td>
</tr>
<tr>
<td>HER2 (-) with trastuzumab</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>
Relationship between CB11 and SP3 on HeCOG 10/05

$y = 0.2436x + 64.946$

$R^2 = 0.4912$
HER2 ICD/ECD protein expression by quadrants

- Cut-points obtained using YTMA 263 slides stained alongside HeCOG and analyzed using Joinpoint software.

- Quadrants:
  - Q1 (ICD$_{low}$/ECD$_{high}$): 5/159 patients (3.1%).
  - Q2 (ICD$_{high}$/ECD$_{high}$): 59/159 patients (37.1%).
  - Q3 (ICD$_{high}$/ECD$_{low}$): 24/159 patients (15.1%).
  - Q4 (ICD$_{low}$/ECD$_{low}$): 71/159 patients (44.7%).
Disease-free survival in HER2-positive, Trastuzumab-treated patients from HeCOG 10/05: Q2 vs Q3

\[ P=0.027 \]

Q2 (ICD\textsubscript{high}/ECD\textsubscript{high}) n=59

Q3 (ICD\textsubscript{high}/ECD\textsubscript{low}) n=24

Log rank \( P=0.027 \); HR=0.23 (95% C.I.: 0.037 to 0.815)
Outline for Presentation

• Using immunohistochemistry (IHC) or quantitative immunofluorescence (QIF) to measure protein on slides
• The challenge of a continuous biomarker – the HER2 example
• The challenge of defining the threshold of detection – The ER example
FDA Cleared Companion Dx antibodies:

<table>
<thead>
<tr>
<th>Drug</th>
<th>Antigen</th>
<th>Company</th>
<th>Antibody</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trastuzumab</td>
<td>HER2</td>
<td>Ventana</td>
<td>4B5</td>
</tr>
<tr>
<td>Trastuzumab</td>
<td>HER2</td>
<td>Leica Biogenex</td>
<td>CB11</td>
</tr>
<tr>
<td>Trastuzumab</td>
<td>HER2</td>
<td>Dako</td>
<td>A0485</td>
</tr>
<tr>
<td>Endocrine Rx</td>
<td>Estrogen Receptor</td>
<td>Ventana</td>
<td>SP1</td>
</tr>
<tr>
<td>Endocrine Rx</td>
<td>Estrogen Receptor</td>
<td>Dako</td>
<td>1D5</td>
</tr>
</tbody>
</table>
Immunohistochemical Detection Using the New Rabbit Monoclonal Antibody SP1 of Estrogen Receptor in Breast Cancer Is Superior to Mouse Monoclonal Antibody 1D5 in Predicting Survival


A Comparison of Estrogen Receptor SP1 and 1D5 Monoclonal Antibodies in Routine Clinical Use Reveals Similar Staining Results

Jane E. Brock, MBBS, PhD, Jason L. Hornick, MD, PhD, Andrea L. Richardson, MD, PhD, Deborah A. Dillon, MD, and Susan C. Lester, MD, PhD

Key Words: Estrogen receptor; SP1; 1D5; Monoclonal antibodies; Breast carcinoma

DOI: 10.1309/AJCPSDFWOLPMEU9
Using an index TMA to define the Threshold of Detection

Should we be following the rules for limits of detection and quantification of analytical chemistry?
Index array 184

ER Nuclear AQUA

ER Nuclear = 112.65
ER Nuclear = 88.03
ER Nuclear = 65.96

ER Nuclear = 70.96
Comparison of SP1 to 1D5 using QIF (AQUA) on a panel of cell line and patient controls – Higher Affinity leads to better Signal to Noise.
Comparison of antibodies by QIF assay suggests SP1 is more sensitive than 1D5


Quantitative Analysis of Estrogen Receptor Expression Shows SP1 Antibody Is More Sensitive Than 1D5

Allison W. Welsh, PhD, Malini Harigopal, MD, Hallie Wimberly, BS, Manju Prasad, MD, and David L. Rimm, MD, PhD
The Goal: To Measure Protein on a slide with the accuracy of Analytical Chemistry

LOD = limit of detection
LOQ = limit of quantification
LOL = limit of linearity
“Lessons Learned”

- Standardization is critical for reproducibility
- Cell lines can help establish cut-points or thresholds
- The human eye is great for many things, but not assessing subtle differences in intensity
- The binding domain and the affinity of the antibody are critical variables
Thanks to:

Rimm Group:
Kurt Schalper
Daniel Carvajal
Jason Brown
Susan Combs
James Smithy
Lauren Moore
Joe McLaughlin
Mehmet Altan
Nemanja Rodic
John McGuire
Joanna Hu
Vasiliki Pelakanou
Mengyao Feng
Nikita Mani
Yan Song
Maria Toki

Yale Pathology Tissue Services
Lori Charette
Sudha Kumar
Veronique Neumeister
Yalai Bai
(Google YPTS)

Rimm Lab Alums:
Vamsi Velchetti
Elsa Anagnostou
Anastasios Dimou
Alley Welsh
Robert Camp
Maria Vassilikapoulou
Huan Cheng
Jennifer Bordeaux
Vamsi Velchetti
Elizabeth Zarrella
Hallie Wimberly

Yale Collaborators
Lajos Pusztai
Yuval Kluger
Leiping Chen

Outside Yale Collaborators
David Hicks (U Rochester)
Edith Perez and NCCTG
George Fountzilas – HeCOG
Amanda Psyrri – HeCOG
Konstantine Kalogeras -HeCOG

Work supported by grants from the NCI, DOD, BCRF, and the Susan G Komen Foundation for the Cure
Methods

Tissue Microarray slide submitted for staining in 4 different labs
Lab1 = Research Histology
Labs 2-4 = Anonymous CLIA labs

4 cuts from Yale TMA 49-9: A cohort of nearly 600 cases from 1962-1982, less than 2% treated with endocrine therapy, node negative treated with surgery alone

1) Lab-to-lab discordance
2) 1%-to-10% discordance

Each slide scanned with Bioimagene scanner
read by 2 pathologists (MH and DLR) and 1 grad student (AW)
Lab-to-Lab Discordance

Lab3

<table>
<thead>
<tr>
<th></th>
<th>ER pos 1%</th>
<th>ER neg 1%</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lab3</td>
<td>301</td>
<td>83</td>
<td>384</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>111</td>
<td>118</td>
</tr>
<tr>
<td></td>
<td><strong>18.3%</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>308</strong></td>
<td><strong>194</strong></td>
<td><strong>502</strong></td>
</tr>
</tbody>
</table>

Lab4

<table>
<thead>
<tr>
<th></th>
<th>ER pos 1%</th>
<th>ER neg 1%</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lab4</td>
<td>234</td>
<td>140</td>
<td>374</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>101</td>
<td>106</td>
</tr>
<tr>
<td><strong>32.2%</strong></td>
<td><strong>241</strong></td>
<td><strong>241</strong></td>
<td><strong>482</strong></td>
</tr>
</tbody>
</table>

Lab2

<table>
<thead>
<tr>
<th></th>
<th>ER pos 1%</th>
<th>ER neg 1%</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lab2</td>
<td>231</td>
<td>72</td>
<td>303</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>178</td>
<td>186</td>
</tr>
<tr>
<td><strong>18.7%</strong></td>
<td><strong>241</strong></td>
<td><strong>250</strong></td>
<td><strong>491</strong></td>
</tr>
</tbody>
</table>

distribution of % positive in misclassified cases
Some labs show “false negative” results compared to other labs.

30% of cases were called positive from lab 2 but negative from lab 4 show outcome similar to double positives.
Serial section Reproducibility for Domain Specific antibodies

Serial section regression for SP3 1:500 in YTMA 263-2

\[ y = 1.3313x + 3.2272 \]
\[ R^2 = 0.9424 \]

Extra-Cellular Domain (SP3)

Serial sections regression for HER2 CB11 (10.4 ng/ml)

\[ y = 0.9284x - 118.39 \]
\[ R^2 = 0.9821 \]

Intra-Cellular Domain (CB11)
Using Spectrum/Aperio scores from Membrane/Total pixel count for HER2 CB11

**Regression of inverted chromogenic membrane intensity scores for HER2 CB11 in YTMA 263**

\[ y = 1.1031x - 1.0545 \]

\[ R^2 = 0.9335 \]

**HER2 CB11 inverted chromogenic membrane intensity scores in YTMA 263**

- **RED:** 3+, AMPLIFIED
- **BLUE:** 2+, AMPLIFIED
- **GREEN:** HER2 amplified cell line

**Joinpoint:** 71.65 (AU)
Using Joinpoint to Assess Ab Sensitivity and Specificity to Predict FISH amplification in YTMA 263 (Chromogenic IHC)

<table>
<thead>
<tr>
<th>HER2 antibody</th>
<th>Joinpoint cut-point (AU)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Positive predictive value (%)</th>
<th>Negative predictive value (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB11</td>
<td>71.65</td>
<td>96.3</td>
<td>64.29</td>
<td>63.41</td>
<td>96.43</td>
</tr>
<tr>
<td>SP3</td>
<td>75.75</td>
<td>89.66</td>
<td>82.61</td>
<td>76.47</td>
<td>92.68</td>
</tr>
</tbody>
</table>