Analytic Variability in Immunohistochemistry Biomarker Studies

Valsamo K. Anagnostou[1], Allison W. Welsh[1], Jennifer M. Giltnane[1], Summar Siddiqui[1], Camil Liceaga[1], Mark Gustavson[4], Konstantinos N. Syrigos[2], Jill L. Reiter[3], and David L. Rimm[1]

Abstract

Background: Despite the widespread use of immunohistochemistry (IHC), there are no standardization guidelines that control for antibody probe variability. Here we describe the effect of variable antibody reagents in the assessment of cancer-related biomarkers by IHC.

Methods: Estrogen receptor (ER), epidermal growth factor receptor (EGFR) 1, and human epidermal growth factor receptor 3 (HER3) were evaluated by quantitative immunofluorescence. Correlations between ER clones 1D5, SP1, F10, and ER60c, and EGFR monoclonal 31G7, 2-18C9, H11, and 15F8, and polyclonal 2232 antibodies were assessed in 642 breast cancer patients. HER3 was measured by RTJ1, RTJ2, SGP1, M7297, RB-9211, and C-17 antibodies in 42 lung cancer patients. Survival analysis was done with the use of multiple cutoff points to reveal any prognostic classification.

Results: All ER antibodies were tightly correlated (Pearson's $r^2 = 0.94\text{-}0.96$; $P < 0.0001$) and western blotting confirmed their specificity in MCF-7 and BT474 cells. All EGFR antibodies but 2232 yielded specific results in western blotting; however, only 31G7 and 2-18C9 were strongly associated (Pearson's $r^2 = 0.61$; $P < 0.0001$). HER3 staining was nonspecific and nonreproducible. High EGFR–expressing patients had a worse prognosis when EGFR was measured with H11 or 31G7 (log rank $P = 0.015$ and $P = 0.06$). There was no statistically significant correlation between survival and EGFR detected by 2-18C9, 15F8, or polyclonal 2232 antibodies.

Conclusions: Antibody validation is a critical analytic factor that regulates IHC readings in biomarker studies. Evaluation of IHC proficiency and quality control are key components toward IHC standardization.

Impact: This work highlights the importance of IHC standardization and could result in the improvement of clinically relevant IHC protocols. *Cancer Epidemiol Biomarkers Prev; 19(4); 982–91. ©2010 AACR.*

Introduction

Immunohistochemistry (IHC) is frequently used in clinical diagnosis and the classification of neoplasms (1). Despite the widespread use of IHC in routine diagnostic practice, no universally accepted standardization guidelines have been developed. IHC is subject to variable preanalytic, analytic, and postanalytic factors that decrease its reproducibility, including tissue preparation and fixation, diverse reagents, different staining methods, scoring systems, and the definition of a "positive" result (2). Considerable variability has been attributed to the selection and insufficient validation of the primary antibodies, poor quality controls, and lack of rigorous assay optimization (1, 3).

Epidermal growth factor receptors (EGFR) are among the most studied cancer biomarkers because of their oncogenic activity in diverse tumor types (4). EGFR in particular was the first receptor to be proposed as a target for cancer therapy, and several EGFR-targeted therapies are now available in the clinic (5). EGFR IHC results highly depend on the type of antibody, protocols, scoring system, and cutoffs used (6), rendering the prognostic role of EGFR protein expression as assessed by IHC highly controversial (7). This lack of reproducibility may be responsible for its failure as a companion diagnostic test, and studies done to date, largely nonstandardized, have shown that detection of high EGFR expression by IHC does not reliably predict the clinical outcome of EGFR-targeted treatment (8). Therefore, IHC is not routinely used to determine EGFR status.

In contrast, IHC on formalin-fixed paraffin-embedded (FFPE) tissue is the current standard for evaluating estrogen receptor (ER) status in breast cancer, and IHC measurements directly affect the management of breast cancer patients (9). Standardized and validated IHC assays for ER include those that use monoclonal antibody (mAb) clones 1D5 and SP1 (10, 11), and efforts to establish

Authors' Affiliations: [1]Department of Pathology, [2]Section of Medical Oncology, Department of Internal Medicine, Yale Cancer Center, and [3]Department of Obstetrics, Gynecology and Reproductive Sciences, Yale University School of Medicine; and [4]HistoRx Inc., New Haven, Connecticut

Note: Supplementary data for this article are available at Cancer, Epidemiology, Biomarkers and Prevention Online (http://cebp.aacrjournals.org/).

Corresponding Author: Valsamo Anagnostou, Department of Pathology, BML 112A, Yale University School of Medicine, 310 Cedar Street, New Haven, CT 06520. Phone: 203-215-7847; Fax: 203-737-5089. E-mail: valsamo.anagnostou@yale.edu

doi: 10.1158/1055-9965.EPI-10-0097

©2010 American Association for Cancer Research.

Downloaded from http://aacrjournals.org/cebp/article-pdf/19/4/982/2271112/982.pdf by guest on 17 March 2025

best practice guidelines for ER IHC testing are ongoing (12). Nevertheless, high false negative rates of tissue samples evaluated by IHC have been reported when laboratory quality control was inadequate and proficiency testing requirements were not met (3); this highlights the critical need for the development of standardized IHC assays.

Here we analyze, validate, and compare ER, EGFR, and human epidermal growth factor receptor (HER) 3 protein expression detected by commonly used ER, EGFR, and HER3 antibodies in breast and lung cancer. Ultimately, we show the effects of variable EGFR antibody selection and validation on outcome in a breast cancer cohort of 642 patients.

Materials and Methods

Cohorts

FFPE primary breast cancer specimens from 642 patients that underwent surgery at Yale-New Haven Hospital (New Haven, CT) between 1962 and 1983 were obtained from the archives of the Pathology Department of Yale University (New Haven, CT). A smaller cohort was retrospectively collected from 42 non–small cell lung cancer patients from Yale-New Haven Hospital between January 1995 and May 2003. The demographics, tumor characteristics, and routine IHC scoring for ER, progesterone receptor, and HER2 in the breast cancer cohort are shown in Supplementary Table S1. The study was approved by the Institutional Review Board.

Tissue microarrays

Tissue specimens were prepared in a tissue microarray format containing one 0.6-mm representative FFPE tumor core for each primary tumor. FFPE cell line pellets were used as controls: A431, H1666, H1355, MCF-7, MB468, BT474, CHO, BaF3, MB436, and HER3-transfected BaF3 cells were purchased from the American Type Culture Collection (Manassas, VA) or donated by other laboratories. FFPE EGFR–stable transfected CHO cell pellets were generated as described in detail elsewhere (13).

Western blotting

Equivalent amounts of protein (25 μg) were resolved by SDS-PAGE in 4% to 12% bis-TRIS gels (150V for 1h) and transferred at 45V for 2 h to a nitrocellulose membrane. Immunoblots were probed with primary antibodies diluted 1:1,000, except for ER 1D5 and SP1 clones, which were diluted 1:500, and ER60c, which was diluted 1:2,000, followed by anti-rabbit or anti-mouse horseradish peroxidase–conjugated secondary antibodies (Santa Cruz Biotechnology, Santa Cruz, CA) diluted 1:4,000 and detected with the use of enhanced chemiluminescence (SuperSignal West Pico, Thermo Scientific, Rockford, IL). β-Tubulin (rabbit polyclonal, Cell Signaling Technology, Danvers, MA) was used to visualize total protein loading. Band intensity was quantified with the use of the ImageJ software (14). The scores generated were normalized to the maximum β-tubulin signal.

Optimization of antigen retrieval for EGFR antibodies

Heat-induced antigen retrieval with the use of sodium citrate (pH 6) resulted in reproducible staining when tissue microarray sections were incubated with mAb clones H11 and 15F8, as well as polyclonal 2232; however, heat-induced antigen retrieval did not yield specific EGFR staining with mAb clones 31G7 and 2-18C9 in breast cancer tissue, nor in EGFR–stable transfected CHO cells or A431 cells (data not shown). Digestion with proteinase K was done instead at room temperature and 37°C at four time points (5, 10, 15, and 45 min) to identify the optimal temperature and incubation period. Incubation for >5 min at room temperature or any incubation at 37°C resulted in the destruction of the tissue architecture and appearance of an edge artifact; moreover, cytokeratin multiplexed on the same slides as a positive control was similarly affected (Fig. 1). Based on these results, we selected pretreatment with proteinase K for 5 min at room temperature as the optimal antigen retrieval method for staining tissue microarray sections with mAb clones 31G7 and 2-18C9.

Immunohistochemistry

Sections (5 μm thick) were deparaffinized, rehydrated, and antigen retrieved by pressure cooking in 10 mmol/L citrate (pH 6) or 10 mmol/L Tris/1 mmol/L EDTA buffer (pH 9) for all primary antibodies except anti-EGFR clones 2-18C9 (pharmDx; Dako, Carpinteria, CA) and 31G7 (Zymed/Invitrogen, Carlsbad, CA), for which proteinase K (Dako) digestion was done instead. Slides were incubated with a cocktail of the primary antibody and a mouse monoclonal cytokeratin (clone AE1/AE3, M3515; Dako) or a polyclonal rabbit cytokeratin antibody (Z0622; Dako) overnight at 4°C for all antibodies except ER clone 1D5 and EGFR clone 2-18C9, which were incubated for 1 h at room temperature (Table 1). This was followed by 1-h incubation with Alexa 546–conjugated goat anti-mouse (A11003; Molecular Probes, Eugene, OR) or Alexa 546–conjugated goat anti-rabbit secondary antibody (A11010; Molecular Probes) diluted 1:100 in rabbit (K4003; Dako) or mouse EnVision (K4001; Dako). Cyanine 5 tyramide (FP1117; Perkin-Elmer, Boston, MA) diluted 1:50 was used for target detection. ProLong mounting medium (ProLong Gold, P36931; Molecular Probes) containing 4′,6-diamidino-2-phenylindole was used to identify nuclei. The protocol specifications for all antibodies used are described in detail in Table 1.

Standardization

Serial sections of a smaller specialized tissue microarray (control array) were stained alongside all cohorts to confirm assay reproducibility; this allows for the construction of a normalization standard curve to adjust for run-to-run variability. Tumor heterogeneity for each marker was assessed in duplicate cores of the same tumor. MCF-7 and A431, EGFR-transfected CHO cells, and HER3-transfected BaF3 cells were used as positive controls for ER, EGFR,

Downloaded from http://aacrjournals.org/cebp/article-pdf/19/4/982/2271112/982.pdf by guest on 17 March 2025

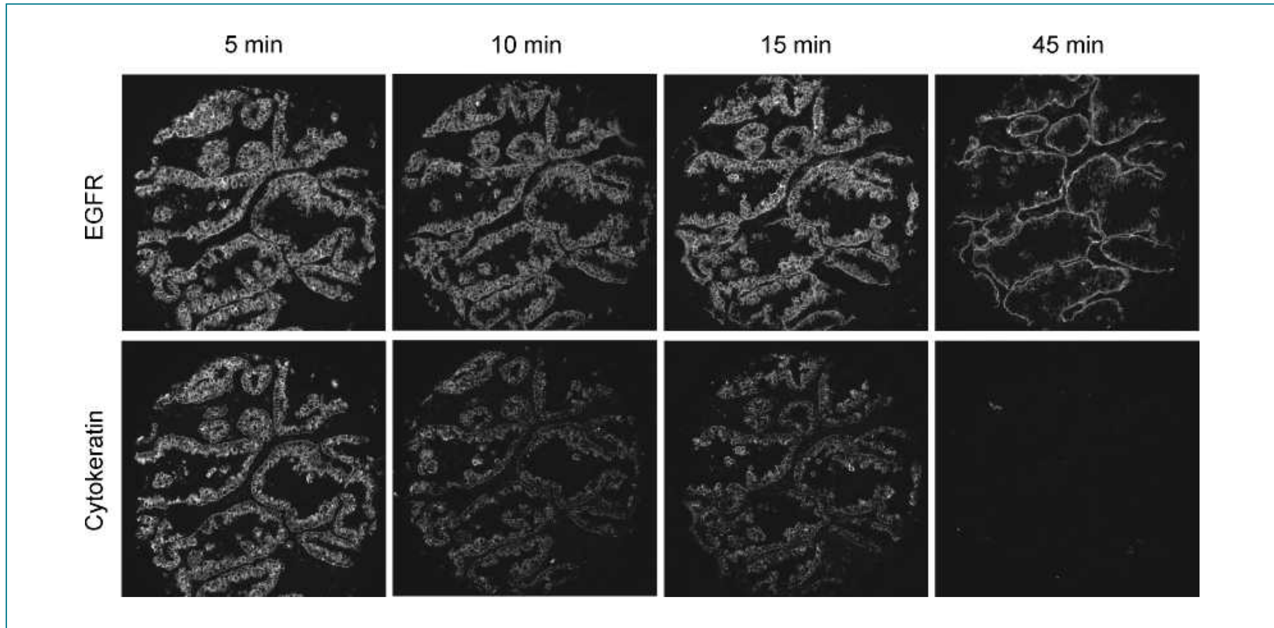

Figure 1. Representative AQUA output for 31G7 after progressive proteinase K digestion for 5, 10, 15, and 45 min at room temperature. Analysis of pan-cytokeratin used as a control for tumor integrity revealed similar results.

and HER3 assays, respectively (Table 1). Negative control sections, in which the primary antibody was omitted, were used for each immunostaining run. In addition MB436, CHO, and BaF3 FFPE cell pellets that are known to not detect ER, human EGFR, and HER3, respectively, were analyzed by quantitative immunofluorescence as negative controls.

Image collection and quantitative analysis

Automated quantitative analysis (AQUA) allows the exact measurement of the protein concentration within subcellular compartments (15). In brief, high resolution 4′,6-diamidino-2-phenylindole, cytokeratin–Alexa 546 and target–cyanine 5 monochromatic images were captured by a PM-2000 microscope (HistoRx, New Haven, CT). Tumor was distinguished by creating an epithelial tumor "mask" from the cytokeratin signal on the basis of an intensity threshold set by visual inspection of histospots. The AQUA scores of target proteins in the tumor mask for EGFR and HER3, and the nuclear compartment for ER were calculated by dividing the sum of the target compartment pixel intensities by the area of the compartment in which they were measured. Specimens with <5% tumor area per histospot were not included in the AQUA for not being representative of the corresponding tumor specimen.

Statistical analysis

AQUA scores were log normalized for run-to-run variability. Pearson's correlation coefficient (r) was used to assess correlations between ER, EGFR, and HER3 measurements with different antibodies, as well as AQUA and western blot readings for the same ER and EGFR antibodies on cell line controls. Assay reproducibility was assessed by comparing AQUA scores on serial sections of the control array; a Pearson's $r > 0.9$ was indicative of minimal assay variability. Survival curves were constructed with the use of the Kaplan-Meier method, and survival differences were analyzed by the log rank test. All P-values were based on two-sided testing, and differences were considered significant at $P < 0.05$. All statistical analyses were done with the use of the SPSS software program (version 13.0 for Windows; SPSS Inc., Chicago, IL).

Results

Specificity of ER antibodies

ER antibody specificity was confirmed by western blotting on MCF-7 and BT474 cells; each of the primary antibodies tested yielded a single 67-kDa band, and the band intensity was comparable with protein measurements by AQUA in FFPE cell pellets (Supplementary Fig. S1). MB436 cells that do not express any ER were negative by AQUA (data not shown).

Comparison of ER antibodies

We used AQUA to measure ER protein expression detected by clones 1D5, SP1, F10, and ER60c on a tissue microarray of 642 breast cancer specimens. Clones 1D5 and ER60c recognize epitopes in the N terminal domain of ER-α, whereas SP-1 and F10 bind epitopes at the C terminus (Fig. 2). ER AQUA scores ranged from 21.5 to 2,727.3 (n = 261; mean ± SE, 438.98 ± 31.23), 5.7 to

Downloaded from http://aacrjournals.org/cebp/article-pdf/19/4/982/2271112/982.pdf by guest on 17 March 2025

1,865.3 (*n* = 366; mean ± SE, 347.71 ± 24.98), 11 to 3,503 (*n* = 389; mean ± SE, 490.9 ± 35.7), and 2.6 to 1,016.7 (*n* = 396; mean ± SE, 135.74 ± 9.8) for ER 1D5, SP1, F10, and ER60c, respectively. Parametric comparisons between different ER antibodies revealed a tight correlation with a Pearson's r^2 of 0.98 ($P < 0.0001$) between SP1 and F10, F10 and ER60c, and SP1 and ER60c; 0.97 ($P < 0.0001$) between 1D5 and ER60c or SP1; and 0.96 ($P < 0.0001$) between F10 and 1D5 clones (Fig. 2). Quantitative immunofluorescence ER measurements were strongly correlated with semi-quantitative pathologist-based ER scores (Spearman's rho = 0.68, ρ = 0.69, ρ = 0.69, and ρ = 0.68; $P < 0.001$ between pathologist-based ER and AQUA F10, SP1, 1D5, and ER60c, respectively). Evaluation of the assay reproducibility did not reveal significant differences between redundant tumor cores or serial sections of the control array stained in each IHC run (Pearson's $r^2 > 0.9$ for all runs; $P < 0.0001$; Supplementary Fig. S2).

Specificity of EGFR antibodies

Western blotting was done for EGFR antibodies on A431, MB468, H1355, and H1666 cell lines that express different levels of EGFR. MAbs 31G7 and 15F8 detected a single 170-kDa band. MAb H11 also detected a 170-kDa band in A431 and MB-468 cells; however, it detected a 140-kDa band in H1666 (Supplementary Fig. S1). The rabbit polyclonal 2232 revealed multiple bands, in addition to the expected 170-kDa band, at various molecular weights (Supplementary Fig. S1). The mAb 2-18C9 recognizes a conformation-specific epitope dependent on

Table 1. Primary antibody characteristics

	Host/Isotype	Antigen retrieval	Concentration	Incubation	Positive control	Source
ER-α						
1D5	Mouse/IgG1, κ	HIAR-pH = 6, 15 min	7.9 μg/mL	1 h, RT	MCF-7 cells	Dako, Carpinteria, CA
F10	Mouse/IgG2a	HIAR-pH = 6, 15 min	0.04 μg/mL	ON, 4°C	MCF-7 cells	Santa Cruz, Santa Cruz, CA
SP1	Rabbit/IgG	HIAR-pH = 6, 15 min	1:1,000*	ON, 4°C	MCF-7 cells	Thermo Fisher, Fremont, CA
60c	Rabbit/IgG	HIAR-pH = 6, 15 min	1:5,000*	ON, 4°C	MCF-7 cells	Millipore, Billerica, MA
EGFR						
31G7	Mouse/IgG1	Proteinase K, 5 min RT	3 μg/mL	ON, 4°C	EGFR-transfected CHO and A431 cells	Zymed/Invitrogen, Carlsbad, CA
2-18C9	Mouse	Proteinase K, 5 min RT	Prediluted	1 h, RT	EGFR-transfected CHO and A431 cells	Dako, Carpinteria, CA
H11	Mouse/IgG1, κ	HIAR-pH = 6, 15 min	6.14 μg/mL	ON, 4°C	EGFR-transfected CHO and A431 cells	Dako, Carpinteria, CA
15F8	Rabbit/IgG	HIAR-pH = 6, 15 min	1:50*	ON, 4°C	EGFR-transfected CHO and A431 cells	Cell Signaling, Danvers, MA
2232	Rabbit polyclonal	HIAR-pH = 6, 15 min	1:200*	ON, 4°C	EGFR-transfected CHO and A431 cells	Cell Signaling, Danvers, MA
HER3						
RTJ2	Mouse/IgG1	HIAR-pH = 6, 15 min	2 μg/mL	ON, 4°C	HER3-transfected BaF3 cells	Santa Cruz, Santa Cruz, CA
RB_9211	Rabbit polyclonal	HIAR-pH = 6, 15 min	0.2 μg/mL	ON, 4°C	HER3-transfected BaF3 cells	Thermo Scientific, Fremont, CA
C_17	Rabbit polyclonal	HIAR-pH = 9, 15 min	0.2 μg/mL	ON, 4°C	HER3-transfected BaF3 cells	Santa Cruz, Santa Cruz, CA
M7297	Mouse/IgG2a, κ	HIAR-pH = 9, 15 min	13 μg/mL	ON, 4°C	HER3-transfected BaF3 cells	Dako, Carpinteria, CA
SGP1	Mouse/IgG1, κ	HIAR-pH = 6, 15 min	4 μg/mL	ON, 4°C	HER3-transfected BaF3 cells	Thermo Scientific, Fremont, CA
RTJ1	Mouse/IgM	HIAR-pH = 6, 15 min	0.5 μg/mL	ON, 4°C	HER3-transfected BaF3 cells	Novocastra, Newcastle upon Tyne, UK

Abbreviations: HIAR, heat-induced antigen retrieval; RT, room temperature; ON, overnight.
*The antibody concentration was not provided by the manufacturer.

Downloaded from http://aacrjournals.org/cebp/article-pdf/19/4/982/2271112/982.pdf by guest on 17 March 2025

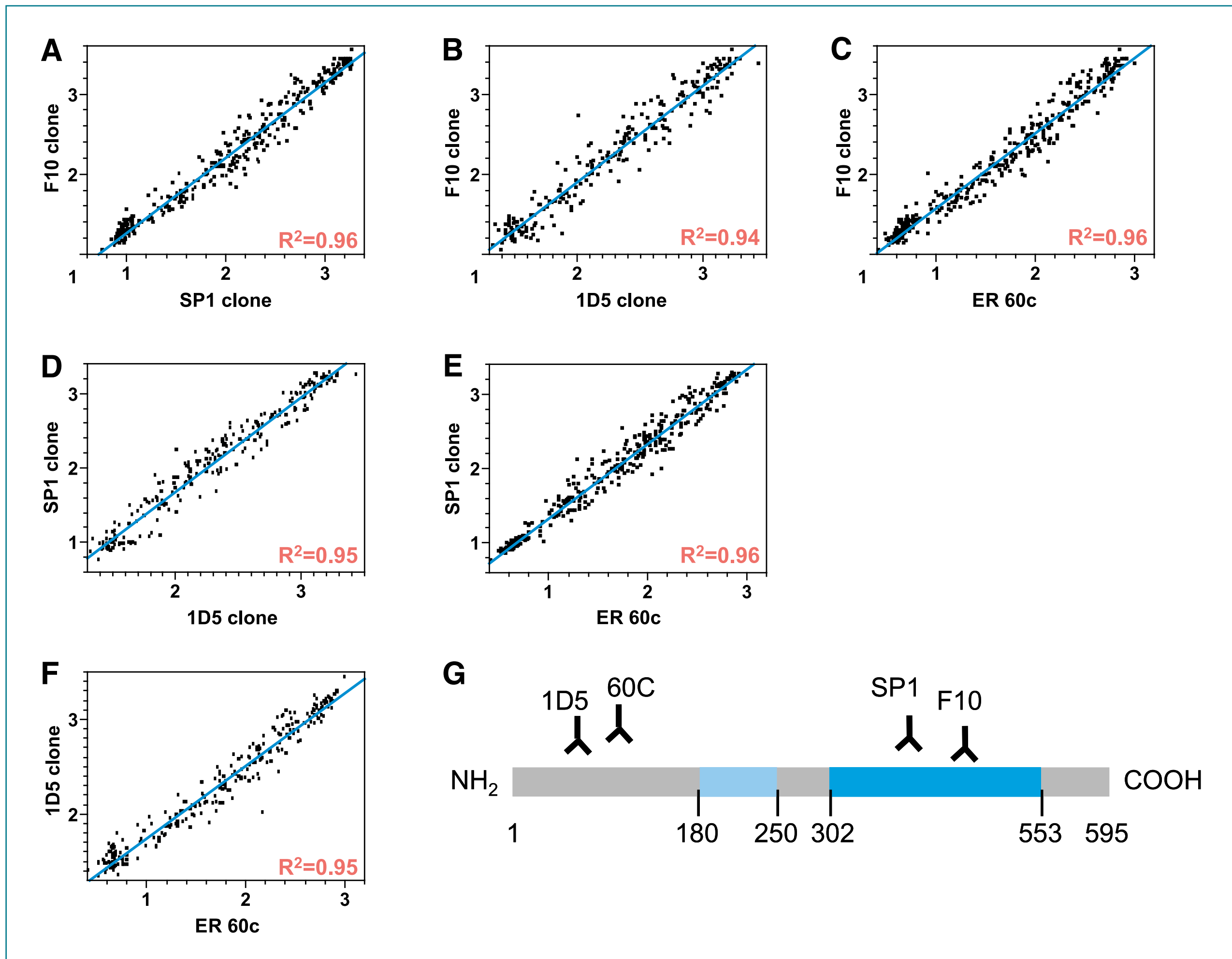

Figure 2. Parametric correlations between ER 1D5, SP1, F10, and ER60c clones (A-F) mapping epitopes in the N terminus (1D5 and ER60c) and C terminus (SP1 and F10) of the 67kD ER protein (G).

disulfide bonds and therefore was not useful for western blotting analysis. We also did AQUA on matched FFPE cell pellets as well as FFPE EGFR-transfected CHO cell pellets that revealed the same range of expression shown in western blot analysis. Comparison of quantified EGFR bands and AQUA scores of the same cell lines revealed an excellent correlation for 31G7 (Pearson's r^2 = 0.9), a good correlation for the 15F8 clone (r^2 = 0.62), and a moderate correlation for clone H11 (r^2 = 0.34). Polyclonal 2232 bands were not quantified because the blotting was not apparently specific for 170-kDa EGFR. CHO cells that do not express any human EGFR were negative by AQUA (data not shown).

Comparison of EGFR antibodies

EGFR protein levels were measured by AQUA, and ranged from 1.44 to 146 (n = 554; mean ± SE, 5.77 ± 0.41), 1.41 to 146 (n = 608; mean ± SE, 13.31 ± 0.61), 7.25 to 121.74 (n = 478; mean ± SE, 32.28 ± 0.71), 4.43 to 165.66 (n = 549; mean ± SE, 9.64 ± 0.41), and 0.01 to 167.5 (n = 395; mean ± SE, 39.22 ± 1.16) for clones 31G7, 2-18C9, H11, 15F8, and polyclonal 2232, respectively. Clones 31G7, 2-18C9, and H11 recognize epitopes on the extracellular domain of EGFR protein as well as the 145-kDa variant III form of EGFR. The 15F8 clone binds to the C terminus, whereas polyclonal 2232 recognizes residues around tyrosine 1068 of the EGFR protein (Fig. 3). In contrast to the strong correlation observed among the ER antibodies, we found that the EGFR antibodies resulted in higher analytic variability. The highest correlation was identified between EGFR measurements by the 2-18C9 and 31G7 clones (r^2 = 0.61; P < 0.00001; Fig. 4), whereas a weak correlation was observed between 15F8 and 2-18C9 (r^2 = 0.21; P = 0.01) or 31G7 (r^2 = 0.17; P = 0.01), and no association was found between 15F8 measurements and H11 (Pearson's r^2 = 0.006; P = 0.5) or polyclonal 2232 (r^2 = 0.05; P = 0.1). H11 was not associated with any other EGFR antibody (r^2 = 0.03 and P = 0.1; r^2 = 0.0007

Downloaded from http://aacrjournals.org/cebp/article-pdf/19/4/982/2271112/982.pdf by guest on 17 March 2025

and $P = 0.59$; and $r^2 = 0.003$; $P = 0.25$ for comparisons with polyclonal 2232, 2-18C9, and 31G7 clones, respectively). There was also no association between EGFR protein levels assessed by the polyclonal 2232 antibody and mAb 2-18C9 ($r^2 = 0.0005$; $P = 0.71$) or 31G7 ($r^2 = 0.008$; $P = 0.2$). A significant correlation was found between EGFR 31G7 and 2-18C6 scores in redundant tumor cores (Pearson's $r = 0.64$ and $r = 0.77$; $P < 0.0001$; Supplementary Fig. S2). No significant differences were observed between serial sections of the control array stained in each run (Pearson's $r^2 > 0.9$ for all runs; $P < 0.0001$; Supplementary Fig. S2).

Comparison of HER3 antibodies

HER3 expression was measured by six antibodies in a smaller lung cancer tissue microarray, and AQUA scores ranged from 12 to 52.9 ($n = 42$; mean ± SE, 26.9 ± 1.8), 111

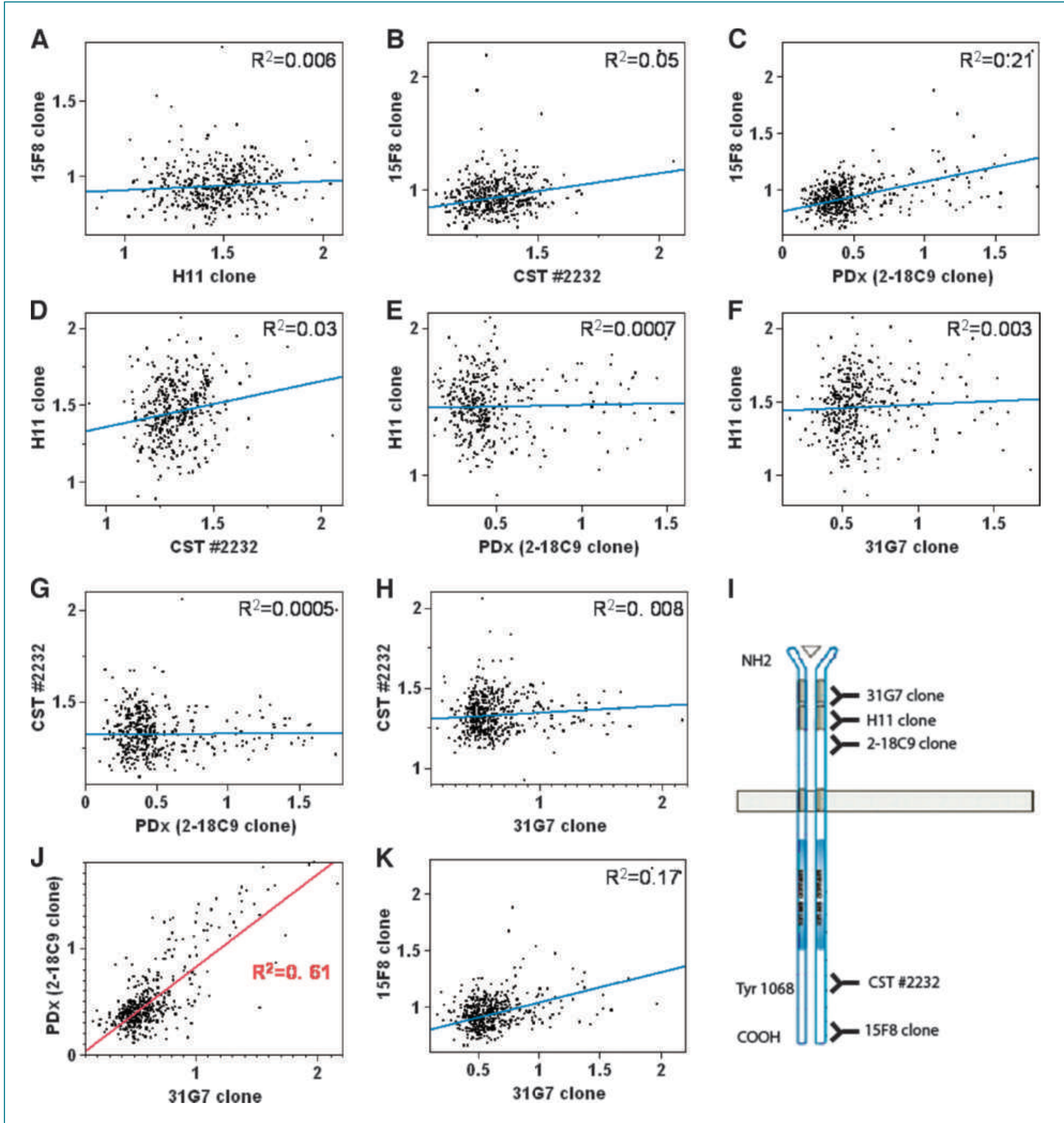

Figure 3. Correlations between EGFR primary antibodies (A-K). EGFR 34G7, 2-18C9, and H11 clones target the N terminal of the wild-type receptor, polyclonal 2232 binds an epitope around tyrosine 1068, and 15F8 clone binds to the C terminus (I).

Downloaded from http://aacrjournals.org/cebp/article-pdf/19/4/982/2271112/982.pdf by guest on 17 March 2025

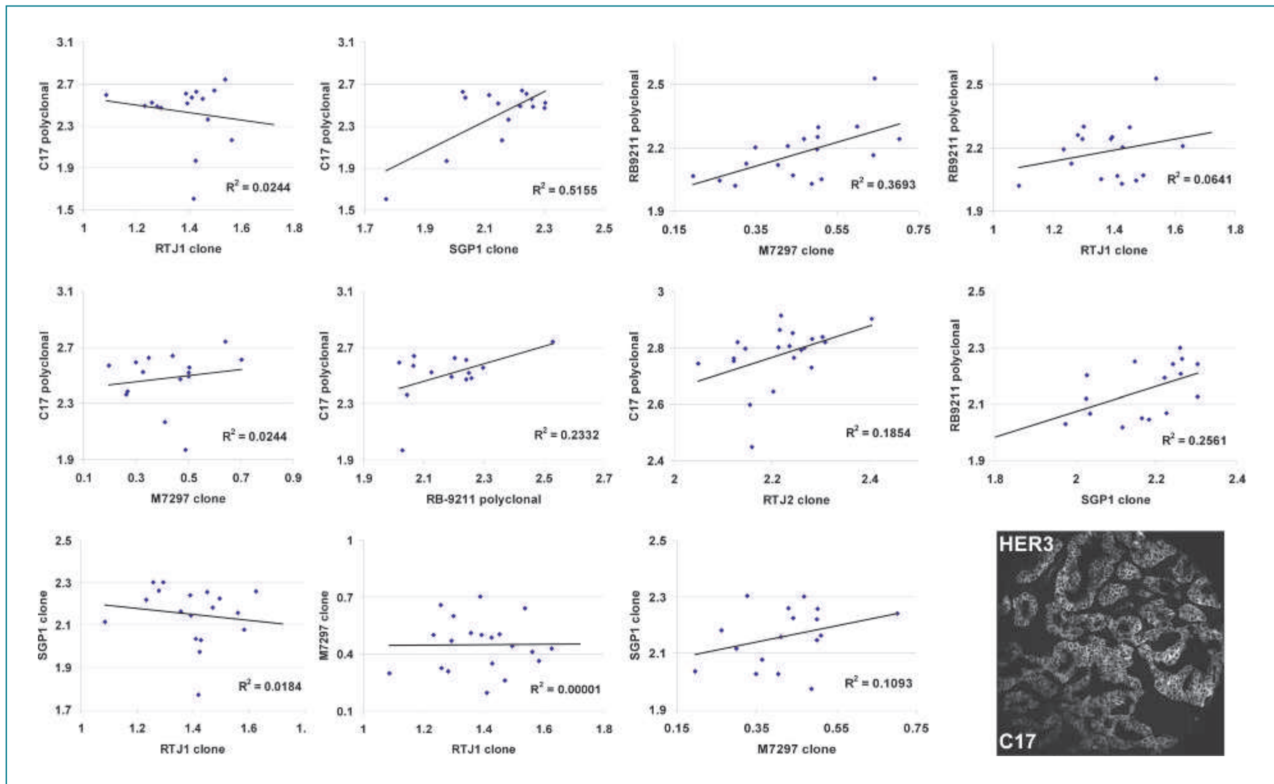

Figure 4. Correlations between different HER3 antibodies measured by AQUA in lung cancer specimens. No association was identified for HER3 measurements by clones RTJ1, RTJ2, SGP1, and M7297, and polyclonal C17 and RB_9211, except for C17 and SGP1 (Pearson's r^2 = 0.5155). Specific cytoplasmic and membranous staining was observed only for C17. Inset, representative expression pattern.

to 244 (*n* = 42; mean ± SE, 168.3 ± 6.21), 58.9 to 200 (*n* = 42; mean ± SE, 145.4 ± 7.8), and 15 to 50 (*n* = 42; mean ± SE, 29.3 ± 0.8), to 40 to 552 (*n* = 42; mean ± SE, 309 ± 25.8) and 104 to 338 (*n* = 42; mean ± SE, 157.8 ± 20) for clones RTJ1, RTJ2, SGP1, and M7297, and polyclonals C17 and RB-9211, respectively. MAbs RTJ1, RTJ2, and M7297, and polyclonal C17 recognize epitopes on the intracellular domain of HER3, whereas the SGP1 clone and RB-9211 bind to the extracellular domain. All antibodies except for C17 failed to detect the expected membranous and/or cytoplasmic localization of HER3, and comparisons between HER3 measurements did not reveal strong correlations between C17 and RTJ1 (r^2 = 0.02), C17 and M7297 (r^2 = 0.02), C17 and RB-9211 (r^2 = 0.28), C17 and RTJ2 (r^2 = 0.18), RB-9211 and RTJ1 (r^2 = 0.08), RB-9211 and SGP1 (r^2 = 0.26), RB-9211 and M7297 (r^2 = 0.38), SGP1 and RTJ1 (r^2 = 0.01), SGP1 and M7297 (r^2 = 0.1), or RTJ1 and M7297 (r^2 = 0.00001; Fig. 4). A tight correlation was observed between C17 and SGP1 (r^2 = 0.61). However, C17 yielded membranous and cytoplasmic staining (Fig. 4), whereas SGP1 yielded nuclear staining without any cytoplasmic localization (data not shown).

Survival analysis for EGFR in breast cancer

To investigate whether the prognostic value of EGFR expression in breast tumors was dependent on the antibody used for IHC detection, we applied different cut points, including tertiles, quartiles, and the top deciles of the continuous EGFR AQUA scores to reproduce the cutoff points most commonly used in the EGFR literature. The *P*-values shown were not normalized for multiple cut-point testing because this was an exploratory analysis to illustrate the effect of cut-point selection. The top 66.66% EGFR-expressing tumors had a worse prognosis compared with the bottom 33.33% expressing group when EGFR protein levels were assessed by mAb H11 (*n* = 290 versus *n* = 135, respectively; log rank *P* = 0.015; Fig. 5). The same trend was observed when EGFR was detected by mAb 31G7 (top 66.66% *n* = 333 versus bottom 33.33% *n* = 167; log rank *P* = 0.126; Fig. 5), and the prognostic stratification was more significant when the median AQUA score was used as a cutoff point (top 50% *n* = 250 versus bottom 50% *n* = 250; log rank *P* = 0.059; Fig. 5C). Interestingly, survival analysis with the use of the median AQUA score for H11 failed to reach significance (top 50% *n* = 212 versus bottom 50% *n* = 213; log rank *P* = 0.63; Fig. 5). Survival analysis with the use of all cutoff points described above for EGFR protein levels assessed by the 15F8 clone, the 2-18C9 clone, and the 2232 polyclonal did not reveal any statistically significant association between EGFR and disease-specific survival (Fig. 5 and data not shown). Thus, the statistical significance,

Downloaded from http://aacrjournals.org/cebp/article-pdf/19/4/982/2271112/982.pdf by guest on 17 March 2025

direction of trend, and optimal cut point seem to be highly dependent on antibody selection.

Discussion

IHC is among the most commonly used research tools in biomedical research. A 3-fold increase in immunohistochemical studies over the last 20 years is indicative of the increased popularity of IHC tests (16). Efforts toward standardization and quality controls have been made (1, 17-20); however, there are no universally accepted guidelines that primarily address IHC laboratory practice (21). Antibody specificity and validation, use of appropriate controls, type of antigen retrieval, and automation are critical analytic factors that affect IHC results (22, 23). Here we measured three commonly studied cancer-related biomarkers, ER, EGFR, and HER3, with the use of different primary antibodies to assess the extent of analytic variability in IHC biomarker studies.

Specificity has been considered the criterion standard for antibody selection, and antibodies are often validated by western blotting when the antibody recognizes the denatured antigen. We and others have noted that most antibodies produced by immunization with synthetic peptides fall into this category (24). Antibodies recognizing ≥2 bands might still identify the same target protein, which might be present as a splice variant, deletion mutant, degradation product, or one that contains different post-translational modifications in the cells of interest. However, additional cross-validation studies with antibodies recognizing different epitopes would then be required. We identified a single band at the expected molecular weight (67 kDa for ER and 170 kDa for EGFR) with each of the antibodies tested for ER and EGFR, except for the polyclonal 2232 and mAb clone H11 anti-EGFR antibodies. Nevertheless, comparisons between all EGFR antibodies recognizing different epitopes on the full-length 170-kDa protein did not reveal strong associations between any of the EGFR antibodies, with the exception of clones 31G7 and 2-18C9. These results are consistent with previous studies that reported strong correlations between mAb clones 2-18C9 and 31G7 with the use of conventional chromagenic IHC (7), as well as between AQUA and IHC-based EGFR detection (25).

Validation of the antibodies by a second independent method in a panel of cell lines with known target expression has been used to confirm specificity (23, 26-29). We used quantitative immunofluorescence to confirm western blotting results and found an excellent correlation between the two methods. MCF-7 and A431, which overexpress endogenous ER and EGFR, respectively, and EGFR-transfected CHO cells were used as positive controls. MB436 breast cancer cells and nontransfected CHO cells were used as negative controls for the ER and EGFR IHC runs. IHC negative controls including

Downloaded from http://aacrjournals.org/cebp/article-pdf/19/4/982/2271112/982.pdf by guest on 17 March 2025

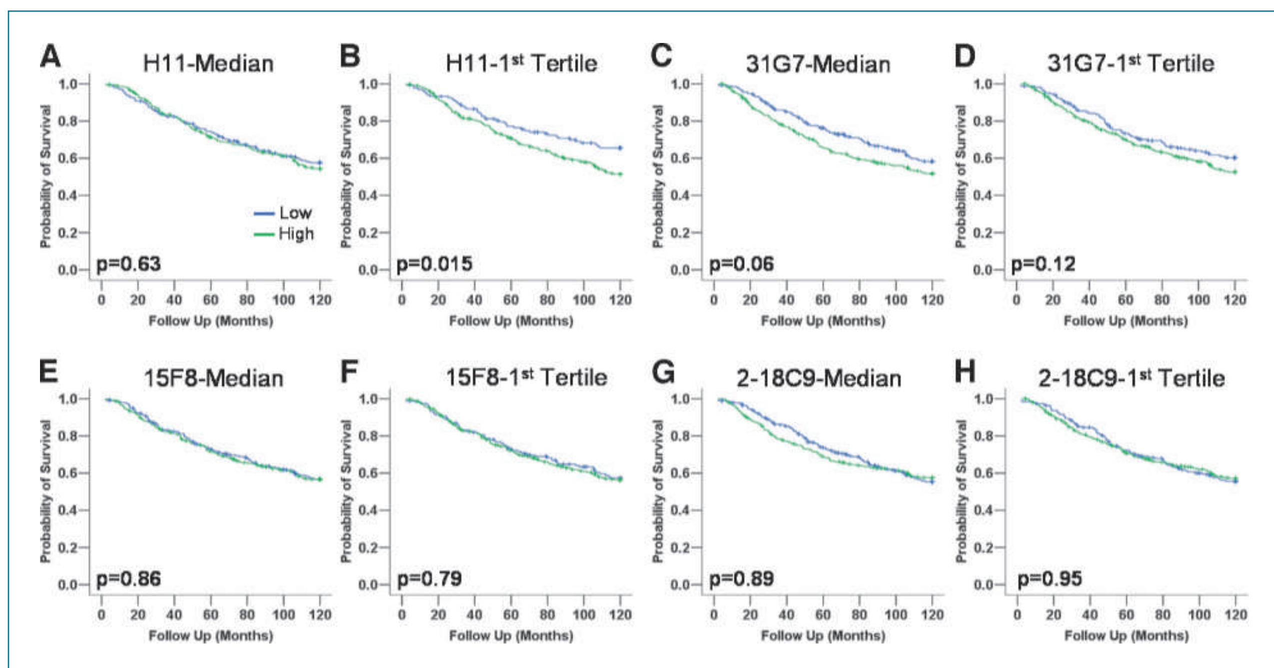

Figure 5. Survival curves based on cohort division by the 1st tertile, and median EGFR AQUA score detected by H11 (A and B), 31G7 (C and D), 15F8 (E and F), and 2-18C9 (G and H). High-EGFR expressers have a worse prognosis compared with the low-expressing group when EGFR is detected by H11, and the 1st tertile is used as a cutoff point (log rank $P = 0.015$). The same trend is observed for 31G7 when the median (log rank $P = 0.06$) and 1st tertile ($P = 0.12$) are used for stratification of patients into high/low expressers.

no primary antibody or use of an unimmunized antibody from the same species are valuable but insufficient. Here we used in addition to conventional negative controls cell lines that are known to not express the protein of interest.

Antibody specificity has also been evaluated in the context of biologically relevant subcellular localization. We have assessed the importance of subcellular localization and its correlation with analytic variability in a series of experiments with the use of different HER3 antibodies in a small lung cancer tissue microarray. We found that 50% of the antibodies used yielded a nuclear pattern for HER3, and no association in terms of specific staining was found among the clones tested. In contrast, all EGFR and ER antibodies yielded membranous and nuclear localization, respectively, in both cell-line positive controls, and EGFR-positive and ER-positive clinical cases, consistent with the biological relevance of these proteins.

One of the critical analytic factors is antigen retrieval, which is used to unmask antigens in FFPE tissue sections. Most antibodies show increased staining when heat-induced antigen retrieval is done (30, 31); however, heat-induced antigen retrieval is not effective for all antibodies, and loss of specific immunostaining (1) as well as "retrieval" of misleading cross-reacting sites (24) have been reported. Enzymatic digestion yields superior results for specific proteins (32-34); however, it has been shown to destroy the targeted epitopes (35, 36), and the lack of a standardized protocol in terms of incubation time and temperature may result in tissue breakdown, loss of morphology, and high levels of background (37). We used heat-induced antigen retrieval for all ER, EGFR H11, 15F8, and 2232 antibody detection, and an optimized protocol for proteinase K digestion for EGFR 31G7 and 2-18C9 clones. Not surprisingly, we observed the destruction of the tissue morphology and loss of EGFR-specific staining with prolonged proteinase K incubation, consistent with previous findings. Heat-induced antigen retrieval for the same clones did not yield any specific EGFR immunoreactivity; this is not consistent with the study of Derecskei et al. showing that microwave antigen retrieval in citrate buffer (pH 6) converted EGFR-negative into EGFR-positive tumors (38).

EGFR-related analytic variability precludes firm conclusions on the prognostic and predictive potential of EGFR. Lung cancer studies failed to consistently show a prognostic (39-42) or predictive (43, 44) role for EGFR protein levels, and the same heterogeneity in study results was also observed in breast cancer studies in which the clinical application of an EGFR IHC test is highly controversial (25, 45-47). Interestingly, only H11 IHC analysis yielded a statistically significant prognostic classification, and a trend toward prolonged survival was observed for the low EGFR–expressing group when 31G7 was used to detect EGFR protein expression. We believe that our results reflect the heterogeneity seen in the EGFR literature attributed to the variation in the study methodology and interpretation of the results. Quantitative immunofluorescence has been shown to have accuracy comparable with an ELISA (48), and has been broadly used for tissue microarray and whole tissue section studies (49). The reproducibility and comparison of AQUA-derived clinical cutoff points with conventional IHC for ER and EGFR have been described in two previous studies (25, 50). We assessed outcome as a function of tumor heterogeneity on different constructs of the same cohort and observed similar prognostic classifications for both ER and EGFR (data not shown).

In conclusion, our work emphasizes on the impact of analytic processing on the validity of biomarker measurements by IHC. Evaluation of IHC proficiency by external agencies is a key first step toward IHC standardization. In the future, we believe rigorous standardization methods will be adopted in the leading laboratories that will include a series of index cases or some other method to show robust reproducibility in the assessment of protein expression on FFPE tissue.

Disclosure of Potential Conflicts of Interest

M. Gustavson: Employment, HistoRx, Inc. D. Rimm: Consultant/Advisory Board, HistoRx, Inc.

Grant Support

Department of Public Health, State of Connecticut #2009-0097 and NIH R33 CA 110511 (D.L. Rimm); and the Susan G. Komen Breast Cancer Foundation, the Breast Cancer Alliance, and the Milheim Foundation for Cancer Research (J.L. Reiter).

Received 01/26/2010; revised 02/17/2010; accepted 02/19/2010; published OnlineFirst 03/23/2010.

Downloaded from http://aacrjournals.org/cebp/article-pdf/19/4/982/2271112/982.pdf by guest on 17 March 2025

References

1. Hsi ED. A practical approach for evaluating new antibodies in the clinical immunohistochemistry laboratory. Archiv Pathol Lab Med 2001;125:289–94.
2. Giltnane JM, Rimm DL. Technology insight: identification of biomarkers with tissue microarray technology. Nat Clin Pract 2004;1:104–11.
3. Allred DC, Harvey JM, Berardo M, Clark GM. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. Mod Pathol 1998;11:155–68.
4. Yarden Y, Sliwkowski MX. Untangling the ErbB signalling network. Nat Rev Mol Cell Biol 2001;2:127–37.
5. Scaltriti M, Baselga J. The epidermal growth factor receptor pathway: a model for targeted therapy. Clin Cancer Res 2006;12:5268–72.
6. Dacic S. EGFR assays in lung cancer. Adv Anat Pathol 2008;15:241–7.
7. Mathieu A, Weynand B, Verbeken E, et al. Comparison of four antibodies for immunohistochemical evaluation of epidermal growth factor receptor expression in non-small cell lung cancer. Lung Cancer (Amsterdam, Netherlands) 2009. Epub ahead of print.
8. Siena S, Sartore-Bianchi A, Di Nicolantonio F, Balfour J, Bardelli A. Biomarkers predicting clinical outcome of epidermal growth factor receptor-targeted therapy in metastatic colorectal cancer. J Natl Cancer Inst 2009;101:1308–24.
9. Allred DC, Carlson RW, Berry DA, et al. NCCN Task Force report: estrogen receptor and progesterone receptor testing in breast cancer by immunohistochemistry. J Natl Compr Canc Netw 2009;7 Suppl 6:S1–S21, quiz S2-3.
10. Cheang MC, Treaba DO, Speers CH, et al. 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. J Clin Oncol 2006;24:5637–44.
11. Dako Corporation C, CA. ER/PR pharm DX kit [Package Insert]. 2008.
12. Yaziji H, Taylor CR, Goldstein NS, et al. Consensus recommendations on estrogen receptor testing in breast cancer by immunohistochemistry. Appl Immunohistochem Mol Morphol 2008;16:513–20.
13. Christensen TA, Reiter JL, Baron AT, Maihle NJ. Generation and characterization of polyclonal antibodies specific for human p110 sEGFR. Hybridoma Hybridomics 2002;21:183–9.
14. Available from: http://rsbweb.nih.gov/ij/.
15. Camp RL, Chung GG, Rimm DL. Automated subcellular localization and quantification of protein expression in tissue microarrays. Nat Med 2002;8:1323–7.
16. Kalyuzhny AE. The dark side of the immunohistochemical moon: industry. J Histochem Cytochem 2009;57(12):333–40.
17. Deutsch EW, Ball CA, Berman JJ, et al. Minimum information specification for *in situ* hybridization and immunohistochemistry experiments (MISFISHIE). Nat Biotechnol 2008;26:305–12.
18. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. REporting recommendations for tumor MARKer prognostic studies (REMARK). Breast Cancer Res Treat 2006;100:229–35.
19. Herman GE, Elfont EA. The taming of immunohistochemistry: the new era of quality control. Biotech Histochem 1991;66:194–9.
20. Taylor CR. Quality assurance and standardization in immunohistochemistry. A proposal for the annual meeting of the Biological Stain Commission, June 1991. Biotech Histochem 1992;67:110–7.
21. Fetsch PA, Abati A. Overview of the clinical immunohistochemistry laboratory: regulations and troubleshooting guidelines. Methods Mol Biol 1999;115:405–14.
22. Rhodes A, Jasani B, Balaton AJ, et al. Study of interlaboratory reliability and reproducibility of estrogen and progesterone receptor assays in Europe. Documentation of poor reliability and identification of insufficient microwave antigen retrieval time as a major contributory element of unreliable assays. Am J Clin Pathol 2001;115:44–58.
23. Rhodes A, Jasani B, Balaton AJ, Miller KD. Immunohistochemical demonstration of oestrogen and progesterone receptors: correlation of standards achieved on in house tumours with that achieved on external quality assessment material in over 150 laboratories from 26 countries. J Clin Pathol 2000;53:292–301.
24. Couchman JR. Commercial antibodies: the good, bad, and really ugly. J Histochem Cytochem 2009;57:7–8.
25. Giltnane JM, Ryden L, Cregger M, Bendahl PO, Jirstrom K, Rimm DL. Quantitative measurement of epidermal growth factor receptor is a negative predictive factor for tamoxifen response in hormone receptor positive premenopausal breast cancer. J Clin Oncol 2007;25:3007–14.
26. Anagnostou VK, Bepler G, Syrigos KN, et al. High expression of mammalian target of rapamycin is associated with better outcome for patients with early stage lung adenocarcinoma. Clin Cancer Res 2009;15:4157–64.
27. Wick MR, Swanson PE. Targeted controls in clinical immunohistochemistry: a useful approach to quality assurance. Am J Clin Pathol 2002;117:7–8.
28. Badve SS, Baehner FL, Gray RP, et al. Estrogen- and progesterone-receptor status in ECOG 2197: comparison of immunohistochemistry by local and central laboratories and quantitative reverse transcription polymerase chain reaction by central laboratory. J Clin Oncol 2008;26:2473–81.
29. Rhodes A, Jasani B, Barnes DM, Bobrow LG, Miller KD. Reliability of immunohistochemical demonstration of oestrogen receptors in routine practice: interlaboratory variance in the sensitivity of detection and evaluation of scoring systems. J Clin Pathol 2000;53:125–30.
30. Shi SR, Imam SA, Young L, Cote RJ, Taylor CR. Antigen retrieval immunohistochemistry under the influence of pH using monoclonal antibodies. J Histochem Cytochem 1995;43:193–201.
31. Pileri SA, Roncador G, Ceccarelli C, et al. Antigen retrieval techniques in immunohistochemistry: comparison of different methods. J Pathol 1997;183:116–23.
32. Beach TG, White CL, Hamilton RL, et al. Evaluation of α-synuclein immunohistochemical methods used by invited experts. Acta Neuropathol 2008;116:277–88.
33. Vilches-Moure JG, Ramos-Vara JA. Comparison of rabbit monoclonal and mouse monoclonal antibodies in immunohistochemistry in canine tissues. J Vet Diagn Invest 2005;17:346–50.
34. Ramos-Vara JA, Beissenherz ME. Optimization of immunohistochemical methods using two different antigen retrieval methods on formalin-fixed paraffin-embedded tissues: experience with 63 markers. J Vet Diagn Invest 2000;12:307–11.
35. Liao J, Mortensen V, Norgaard-Pedersen B, Koch C, Brodbeck U. Monoclonal antibodies against brain acetylcholinesterases which recognize the subunits bearing the hydrophobic anchor. Eur J Biochem/FEBS 1993;215:333–40.
36. Leong AS, Gilham PN. The effects of progressive formaldehyde fixation on the preservation of tissue antigens. Pathology 1989;21:266–8.
37. Leong TY, Leong AS. How does antigen retrieval work? Adv Anat Pathol 2007;14:129–31.
38. Derecskei K, Moldvay J, Bogos K, Timar J. Protocol modifications influence the result of EGF receptor immunodetection by EGFR pharmDx in paraffin-embedded cancer tissues. Pathol Oncol Res 2006;12:243–6.
39. Gupta R, Dastane AM, McKenna R, Jr., Marchevsky AM. The predictive value of epidermal growth factor receptor tests in patients with pulmonary adenocarcinoma: review of current "best evidence" with meta-analysis. Hum Pathol 2009;40:356–65.
40. Meert AP, Martin B, Delmotte P, et al. The role of EGF-R expression on patient survival in lung cancer: a systematic review with meta-analysis. Eur Respir J 2002;20:975–81.
41. Clark GM, Zborowski DM, Culbertson JL, et al. Clinical utility of epidermal growth factor receptor expression for selecting patients with advanced non-small cell lung cancer for treatment with erlotinib. J Thorac Oncol 2006;1:837–46.
42. Perez-Soler R, Chachoua A, Hammond LA, et al. Determinants of tumor response and survival with erlotinib in patients with non-small-cell lung cancer. J Clin Oncol 2004;22:3238–47.
43. Tsao MS, Sakurada A, Cutz JC, et al. Erlotinib in lung cancer-molecular and clinical predictors of outcome. N Engl J Med 2005;353:133–44.
44. Eberhard DA, Giaccone G, Johnson BE. Biomarkers of response to epidermal growth factor receptor inhibitors in Non-Small-Cell Lung Cancer Working Group: standardization for use in the clinical trial setting. J Clin Oncol 2008;26:983–94.
45. Gori S, Sidoni A, Colozza M, et al. EGFR, pMAPK, pAkt and PTEN status by immunohistochemistry: correlation with clinical outcome in HER2-positive metastatic breast cancer patients treated with trastuzumab. Ann Oncol 2009;20:648–54.
46. Somlo G, Chu P, Frankel P, et al. Molecular profiling including epidermal growth factor receptor and p21 expression in high-risk breast cancer patients as indicators of outcome. Ann Oncol 2008;19:1853–9.
47. Okunade G, Green AR, Ying M, et al. Biological profile of oestrogen receptor positive primary breast cancers in the elderly and response to primary endocrine therapy. Crit Rev Oncol Hematol 2009;72:76–82.
48. McCabe A, Dolled-Filhart M, Camp RL, Rimm DL. Automated quantitative analysis (AQUA) of *in situ* protein expression, antibody concentration, and prognosis. J Natl Cancer Inst 2005;97:1808–15.
49. 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.
50. Moeder CB, Giltnane JM, Harigopal M, et al. Quantitative justification of the change from 10% to 30% for human epidermal growth factor receptor 2 scoring in the American Society of Clinical Oncology/College of American Pathologists guidelines: tumor heterogeneity in breast cancer and its implications for tissue microarray based assessment of outcome. J Clin Oncol 2007;25:5418–25.

Downloaded from http://aacrjournals.org/cebp/article-pdf/19/4/982/2271112/982.pdf by guest on 17 March 2025