RESEARCH

Accuracy and Reproducibility of HER2 Status in Breast Cancer Using Immunohistochemistry: A Quality Control Study in Tuscany Evaluating the Impact of Updated 2013 ASCO/CAP Recommendations

**S. Bianchi · S. Caini · M. Paglierani · C. Saieva ·
V. Vezzosi · G. Baroni · A. Simoni · D. Palli ·
On behalf of the Tuscany Breast Cancer Study Group**

Abstract The correct identification of HER2-positive cases is a key point to provide the most appropriate therapy to breast cancer (BC) patients. We aimed at investigating the reproducibility and accuracy of HER2 expression by immunohistochemistry (IHC) in a selected series of 35 invasive BC cases across the pathological anatomy laboratories in Tuscany, Italy. Unstained sections of each BC case were sent to 12 participating laboratories. Pathologists were required to score according to the Food and Drug Administration (FDA) four-tier scoring system (0, 1+, 2+, 3+). Sixteen and nineteen cases were HER2 non-amplified and amplified respectively on fluorescence in situ hybridization. Among 192 readings of the 16 HER2 non-amplified samples, 153 (79.7 %) were coded as 0 or 1+, 39 (20.3 %) were 2+, and none was 3+ (false positive rate 0 %). Among 228 readings of the 19 HER2 amplified samples, 56 (24.6 %) were scored 0 or 1+, 79 (34.6 %) were 2+, and 93 (40.8 %) were 3+. The average sensitivity was 75.4 %, ranging between 47 % and 100 %, and the overall false negative rate was 24.6 %. Participation of pathological anatomy laboratories performing HER2 testing by IHC in external quality assurance programs should be made mandatory, as the system is able to identify laboratories with suboptimal performance that may need technical advice. Updated 2013 ASCO/CAP recommendations should be adopted as the widening of IHC 2+ "equivocal" category would improve overall accuracy of HER2 testing, as more cases would be classified in this category and, consequently, tested with an in situ hybridisation method.

Introduction

The introduction of trastuzumab had a remarkable impact on the treatment of epidermal growth factor receptor 2 (HER2)-positive BC patients. Trastuzumab targets the extracellular domain of HER2, which is a transmembrane receptor tyrosine kinase encoded by a gene located on the long arm of the chromosome 17 [1] and amplified and/or over-expressed in 15–20 % of primary BC [2]. The correct identification of HER2-positive BC cases is therefore a key point to provide the most appropriate therapy to these patients [3]. Recommendations on HER2 testing were developed in 2007 by the American Society of Clinical Oncology (ASCO) and the College of American Pathologists (CAP) [4].

Two diagnostic techniques are currently available to assign HER2 status in clinical practice: immunohistochemistry (IHC), which detects HER2 protein expression, and in-situ hybridization (ISH), which quantifies gene amplification. Current testing algorithm recommend IHC upfront and scoring HER2 expression according to the Food and Drug Administration (FDA) four-tier system: scores 0 and 1+ are considered as HER2- negative (i.e., not eligible to

S. Bianchi (✉) · M. Paglierani · V. Vezzosi · G. Baroni · A. Simoni
Pathological Anatomy Unit, Department of Surgery and
Translational Medicine, University of Florence, Florence, Italy
e-mail: simonetta.bianchi@unifi.it

S. Caini · C. Saieva · D. Palli
Molecular and Nutritional Epidemiology Unit, Cancer Research and
Prevention Institute (ISPO), Florence, Italy

anti-HER2 treatment), score 3+ is considered as HER2-positive (i.e., eligible to anti-HER2 treatment), whereas score 2+ constitutes a gray zone in which a reflex ISH testing is needed [5].

Three ISH techniques are currently available to assess HER2 gene amplification: fluorescence in situ hybridization (FISH), chromogenic in situ hybridization (CISH), and silver in situ hybridization (SISH). Of these, FISH represents the most validated ISH technique. According to current guidelines [6, 7], HER2 gene amplification is defined as HER/chromosome 17 centromere (CEP17) ratio ≥2, or HER2 copy number ≥6.

The need for accurate and reproducible determination of HER2 status using standardized assays, with strict adherence to quality control and quality assurance programs, are widely addressed in recent recommendations [4, 6–8], and key points in pre-analytical, analytical and post-analytical phases of HER2 testing have been identified [4]. In particular, the performance of the distinct available antibodies can affect accuracy and reproducibility of results during the analytical phase. Anti-HER2 antibodies for IHC are the A0485 rabbit polyclonal, the 4B5 rabbit monoclonal and the CB11 mouse monoclonal antibodies. These antibodies can be used with in-house protocols or as part of commercial kit preparations, such as: HercepTest™ (clone A0485; Dako, Glostrup, Denmark), Bond Oracle® HER2 (clone CB11; Leica Microsystems GmbH, Wetzlar, Germany), and Pathway® HER2 (clone 4B5; Ventana Medical Systems Inc., Tucson, AZ, USA).

Discordance in HER2 IHC results between local and central laboratories participating in clinical trials has been well documented [2]. This is especially relevant for low volume laboratories, where up to≈20 % of HER2 IHC assays may prove to be false positive findings after retesting in a central high volume laboratory [9]. The literature has mainly focused so far on false positivity of HER2 IHC assays; on the contrary, only a few data are reported concerning false negative HER2 IHC results. In the VIRGO observational cohort study of patients with primarily HER2 negative breast cancer, 4 % of specimens rated as HER2 negative by local laboratories were found to be HER2 positive centrally [10]. In most clinical trials published so far, a re-evaluation by expert pathologists at central level was performed that focused mainly on laboratory performance rather than on microscopic evaluation of results [11].

We aimed at investigating the accuracy and reproducibility of HER2 expression by IHC in a selected series of invasive BC cases with a standardized pre-analytical phase across the pathological anatomy laboratories in Tuscany, Italy. In particular, we assessed both the performance of pathological anatomy laboratories and the microscopic interpretation of BC cases by comparing the self-reported HER2 IHC results with the HER2 status on FISH.

Materials and Methods

Thirty-five surgical specimen cases of invasive BC were selected from the archives of the Pathological Anatomy Unit of the Careggi Hospital, Florence, Italy (referred to as "reference laboratory" thereinafter). A disproportionate number of IHC 2+ cases were included to enhance the number of discrepancies that usually occur in borderline cases.

The pre-analytical phase was standardized for all BC cases, with a cold ischemia time less than 1 h and a formalin fixation time, in 10 % neutral buffered formalin, within a time frame comprised between 24 and 48 h. Representative formalin-fixed paraffin-embedded blocks were cut into 4-µm-thick sections, mounted on coated glass slides, and baked overnight at 56 °C.

Pathologists from the 14 public pathological anatomy laboratories of Tuscany were invited to join the study. Two laboratories were excluded due to their routine use of fixatives other than buffered formalin in their daily routine. This left with 12 participating laboratories: of these, 2 used A0485 rabbit polyclonal antibody with in-house protocols, 2 used HercepTest™, 2 used Bond Oracle® HER2, and 6 used Pathway® HER2. All participating laboratories were using internal quality assurance procedures.

Thirty five sets (1 for each BC case) of unstained sections were sent to each participating laboratory in 4 separate shipments during May-August 2013, with the aim of obtaining freshly cut sections and preventing the potential reduction of intensity of IHC staining due to slide storage [12]. Three unstained sections were made available to each participating laboratory for each selected BC case: 1 for H&E, 1 for IHC and 1 as reserve. The first and the last section from each paraffin block were re-tested on IHC by reference laboratory to exclude variances of HER2 expression.

The slides of each case were anonymous and identified with randomly assigned numerical labels.

Participating pathologists were required to report for each case: the percentage of stained cells; the intensity (weak, moderate, strong) and pattern (complete or incomplete) of membrane staining; and the score according to the FDA four-tier scoring system (0, 1+, 2+, 3+). Scores 0 or 1+ were considered negative for HER2 expression; score 2+ was considered as equivocal; and score 3+ was considered positive.

The HER2 status of each specimen was confirmed by dual-color FISH in the reference laboratory, using the PathVysion HER-2 DNA Probe Kit (Abbott Laboratories, Abbott Park, IL, USA). FISH-stained sections were scanned and three separate invasive carcinoma areas identified. The number of CEP17 and HER2 signals was counted in 60 non-overlapping nuclei using at least three distinct tumor fields. In cases with an average of CEP17 signals less than three, the HER2/CEP17 ratio was calculated and those BC cases that had a ratio ≥2 were considered amplified. In consideration that the true

polisomy (average of CEP17 signals ≥3) is a rare event [13] the average of HER2 copy number was used in cases with an average of CEP17 signals ≥3, and cases were interpreted as amplified when the average of HER2 copy number was ≥6 [3, 14]. Cases with HER2 genetic heterogeneity [15] were excluded from the study as their concordance when tested on IHC in different laboratories and on IHC vs. FISH is known to be low.

HER2 status on FISH was used as "gold standard": 16 and 19 cases resulted non-amplified and amplified respectively.

Statistical Analysis

We used a weighted kappa statistic [16] to quantify the between-reader agreement for all possible pairs of readers. We used the following weights: 1 for perfect agreement, 2/3 for a disagreement by 1 score value (0 vs. 1+, 1+ vs. 2+, 2+ vs. 3+), 1/3 for a disagreement by 1 score values (0 vs. 2+, 1+ vs. 3+), and 0 for the disagreement by 3 score values (0 vs. 3+). Values for kappa statistics range between 0 and 1, and the magnitude of agreement by the kappa is as follows: 0–0.20 very low, 0.20–0.40 low, 0.40–0.60 moderate, 0.60–0.80 good, 0.80–1.00 excellent [17].

We calculated the sensitivity (for each reading pathologists and overall) as the proportion of actual (i.e., on FISH) HER2-amplified BC cases that were assigned a 2+ or 3+ score by IHC. We used the chi-square test to assess whether the average sensitivity differed across used antibodies and annual caseload (number of BC samples per year: ≤100, 101–200, 201–300, >300).

All analyses were performed using STATA version 11 (STATA Corp., TX, USA).

Results

Out of the 420 readings (35 breast cancer cases for each of 12 participating pathologists), the distribution of scores was as follows: 107 (25.5 %) were 0; 102 (24.3 %) were 1+; 118 (28.1 %) were 2+; and 93 (22.1 %) were 3+.

In Table 1, we report the values of the weighted kappa statistics for the 66 possible pairs of pathologists. The agreement was excellent for 3 pairwise comparisons (5 %); good for 31 pairs (47 %); moderate for 22 pairs (33 %); and low for 10 pairs (15 %). The three kappa values higher than 0.80 were observed for the between-reader comparisons of pathologists E, I and J with each other.

In Table 2, we report the number of samples that were read by each pathologist as 0, 1+, 2+ or 3+, according to the status of HER2 on FISH, along with values of sensitivity for each individual pathologist and overall. Among 192 readings of the 16 HER2 non-amplified samples, 153 (79.7 %) were coded as

0 or 1+, 39 (20.3 %) were 2+, and none was 3+ (false positive rate 0 %). The only pathologist (B) who classified all 16 non-HER2-amplified BC samples as either 0 or 1+ also scored 10 of the 19 HER2-amplified breast cancer cases as either 0 or 1+, showing the lowest sensitivity (47 %).

On the other hand, among 228 readings of the 19 HER2-amplified samples, 56 (24.6 %) were scored 0 or 1+, 79 (34.6 %) were 2+, and 93 (40.8 %) were 3+. The average sensitivity was 75.4 %, ranging between 47 % and 100 %, and the overall false negative rate was 24.6 %. The three pathologists (F, H and L) who achieved a 100 % sensitivity also scored, respectively, 4, 6 and 8 HER2 non-amplified cases as 2+.

The average sensitivity was not affected by the antibody used (p=0.35) or the annual caseload of the laboratory (p=0.97). In particular, the three 100 %-sensitivity pathologists used 3 different antibodies and reported different annual caseloads (300+ and101–200 cases/year for 1 and 2 centers, respectively).

We report in Table 3 the distribution of scores given by the 12 pathologists to each BC sample according to the HER2 status on FISH. Of the 16 HER2 non-amplified BC samples, 6 were read as either 0 or 1+ by all pathologists, and none was read as 3+. Among the 19 HER2-amplified BC samples, 8 were read as either 2+ or 3+ by all pathologists, while the remaining 11 were read as 0 or 1+ by at least 1 (up to 9) pathologist; in this latter group 7 BC cases were never scored 3+.

In Table 4 we report the distribution of false negative readings along with antibody used,

IHC score, % of stained tumor cells and intensity of incomplete membrane staining. FISH-amplified samples read as 1+ considerably differed in terms of percentage of stained cells and intensity of incomplete membrane staining, which points towards analytical phases and poor overall performance being the main determinants of inconsistencies across laboratories (Fig. 1). In BC cases 14, 16, 21 and 23, the false-negative readings were made by most pathologists, and their frequency did not depend on the antibody used. The HER2/CEP17 ratio on FISH was above 2 in cases 14 and 21, and the average of CEP17 signals was ≥3 in cases 16 (Fig. 2) and 23; all these cases resulted amplified with the average of HER2 copy number ≥6.

All false negative readings were reviewed centrally. The 13 readings scored as 0 were confirmed, demonstrating that false negative results were due to analytical variability, particularly to antigen retrieval. Among the 43 false negative readings scored as 1+, 28 were attributable to problems during the microscopic evaluation, mainly due to the stringency of applying the scoring criteria by the pathologists (all these cases were scored as 2+ centrally). The remaining 15 cases read as 1+ were confirmed at central revision, thus showing that false negative results were due to laboratory performance concerning analytical phase.

Table 1 Values of the weighted kappa statistics for the between-reader agreement. Values below 0.4 and above 0.8 are reported in bold and italics, respectively

	B	C	D	E	F	G	H	I	J	K	L
A	**0.353**	**0.285**	0.660	0.608	0.484	**0.397**	0.617	0.532	0.610	0.538	0.576
B		0.624	0.457	0.531	**0.379**	0.603	**0.272**	0.579	0.570	0.567	**0.326**
C			**0.397**	0.431	0.445	0.602	**0.281**	0.523	0.469	0.509	**0.381**
D				0.748	0.649	0.599	0.595	0.720	0.795	0.676	0.611
E					0.730	0.627	0.643	*0.842*	*0.823*	0.752	0.652
F						0.511	0.764	0.657	0.640	0.612	0.715
G							**0.397**	0.682	0.713	0.631	0.493
H								0.566	0.599	0.526	0.708
I									*0.842*	0.721	0.624
J										0.796	0.653
K											0.579

Discussion

Because HER2-positive status correlates with clinical efficacy of trastuzumab, false-negative results of HER2 status testing may lead to under-treatment and deny eligible BC patients a potentially life-extending targeted therapy. False-positive HER2 results are an issue as well, as over-treatment of HER2-negative patients may have considerable side effects and represents a waste of resources. Accurate and reproducible diagnostic testing of HER2 status is therefore a key aspect for the appropriate use of trastuzumab in clinical practice.

Although IHC and ISH can be used interchangeably, IHC is widely available in most pathological anatomy laboratories, while ISH is generally centralized at experienced, well-equipped reference laboratories. Therefore, most laboratories rely on IHC to determine the HER2 status of BC cases, followed by ISH for equivocal cases (i.e., those scored as 2+ on IHC).

The importance of accuracy and reproducibility in the determination of HER2 IHC testing in BC has been frequently acknowledged [4, 6, 8, 18], and participation in an EQA program is, in some countries such as UK, Canada and US, mandatory for all laboratories performing HER2 testing by IHC [4, 6, 8, 18]. In Germany, where participation in histopathology EQA program is currently not mandatory, annual or bi-annual nationwide trials (QualitatInitiative Pathologie, QuIP) for tissue-based markers in breast cancer (i.e. ER, PgR and HER2) have been set up on tissue microarrays; the participation of laboratories in these trials was on a voluntary basis [19].

In our study we aimed to eliminate assay variation due to pre-analytical factors using BC cases from a single reference laboratory with a standardized pre-analytical phase and

Table 2 Number and score (0, 1+, 2+ and 3+) of readings that each pathologist ($n=12$) gave to 35 BC samples that tested non-amplified or amplified on the FISH test for the HER2 status. Individual and mean sensitivity estimates (combining 2+ and 3+ scores)

Reader	HER2 non-amplified BCs ($n=16$)				HER2 amplified BCs ($n=19$)				Sensitivity
	0	1+	2+	3+	0	1+	2+	3+	
A	4	5	7	0	1	3	5	10	0.79
B	10	6	0	0	4	6	5	4	0.47
C	10	5	1	0	3	6	7	3	0.53
D	5	9	2	0	1	3	5	10	0.79
E	9	5	2	0	0	4	5	10	0.79
F	8	4	4	0	0	0	10	9	1.00
G	10	5	1	0	3	6	3	7	0.53
H	3	7	6	0	0	0	10	9	1.00
I	9	4	3	0	0	6	5	8	0.68
J	9	4	3	0	1	4	5	9	0.74
K	11	3	2	0	0	5	8	6	0.74
L	6	2	8	0	0	0	11	8	1.00
Total	94	59	39	0	13	43	79	93	0.75
	(49 %)	(31 %)	(20 %)	(0 %)	(6 %)	(19 %)	(34 %)	(41 %)	

Table 3 Distribution of scores (0, 1+, 2+ and 3+) of all readings given by the twelve pathologists to each individual BC sample, according to HER2 status on FISH

Sample ID	HER2 status (on FISH)	0	1+	2+	3+	Sample ID	HER2 status (on FISH)	0	1+	2+	3+
3	non-amplified	10	2	0	0	1	amplified	0	0	2	10
4	non-amplified	0	4	8	0	2	amplified	0	0	0	12
9	non-amplified	8	2	2	0	5	amplified	0	3	8	1
10	non-amplified	10	2	0	0	6	amplified	2	2	8	0
11	non-amplified	7	3	2	0	7	amplified	0	0	2	10
12	non-amplified	6	5	1	0	8	amplified	1	5	6	0
17	non-amplified	10	2	0	0	13	amplified	0	4	7	1
18	non-amplified	12	0	0	0	14	amplified	0	7	5	0
19	non-amplified	0	7	5	0	15	amplified	0	0	4	8
20	non-amplified	1	5	6	0	16	amplified	2	6	4	0
25	non-amplified	0	5	7	0	21	amplified	2	7	3	0
26	non-amplified	3	6	3	0	22	amplified	0	2	7	3
29	non-amplified	1	8	3	0	23	amplified	4	3	5	0
30	non-amplified	11	1	0	0	24	amplified	0	0	1	11
32	non-amplified	4	6	2	0	27	amplified	2	0	3	7
34	non-amplified	11	1	0	0	28	amplified	0	4	8	0
						31	amplified	0	0	1	11
						33	amplified	0	0	5	7
						35	amplified	0	0	0	12

concentrated our attention on analytical phase, i.e. distinct antibodies used, and on post-analytical variables, i.e. interpretation of results. We found that, with a standardized pre-analytical phase, no false positive HER2 IHC testing occurred, as all HER2 IHC 3+ cases showed HER2 amplification on FISH.

On the other end of the spectrum, clinical experience and recent literature [7] indicated that false-negative HER2 test results must be carefully considered. The revised criteria of ASCO-CAP recommendations published in 2007 [4] raised the HER2-positive threshold by FISH (HER2/CEP17 ratio from 2.0 to 2.2 or HER2 copy number from 4 to 6 copies/cell) and by IHC (strong circumferential staining from >10 % to >30 % of cells), leading to a potential under-estimation of false negative. To avoid this potential source of bias, we opted to use the US FDA threshold values for both IHC and FISH. Nevertheless, our study reports 56 readings that were scored as negative (13 and 43 readings scored as 0 and 1+ respectively) on IHC among those amplified by FISH. These false-negative results are attributable to analytical variability (13 of 13 scored as 0, 15 of 43 scored as 1+) and to interpretation criteria (28 of 43 scored as 1+).

In analytical phase, the availability of distinct antibodies and their specificity can take part in affecting reproducibility of results [20]. In our study, however, the average sensitivity was not affected by the antibody used, and the three laboratories that achieved a sensitivity of 100 % used each a different antibody.

In terms of analytical variability, it is well known as accuracy and reproducibility of results is highly dependent upon staining methodology, particular antigen retrieval [2, 5, 21]. In our study, 10 out 12 laboratories used commercial kit preparations (2 HercepTest™; 2 Bond Oracle® HER2 and 6 Pathway® HER2) according to the manufacturer's instructions. Commercial kits are used as semi-automated (HercepTest™) automated (Pathway® HER2) or fully-automated (Bond Oracle® HER2) system with open protocol with FDA certified antibody (Pathway® HER2) or FDA certified closed protocol (HercepTest™, Bond Oracle® HER2). In case of open protocol, optimization of antigen retrieval and incubation times is still required in order to adequately detect epitopes in paraffin-embedded tissue [2, 5, 21]. Moreover antigen retrieval is done in a variety of ways in different commercial kits and this could produce marked differences in staining reactions, determining significant differences in IHC staining patterns that could partly explain the false negative results we reported. Laboratories are usually unaware of these variations if they do not participate in EQA programs. The different chemical composition of the retrieval solutions may affect the efficacy of the antigen retrieval process. In addition, duration and temperature are other two variables that are critical to the process of heat-induced antigen retrieval and can have an impact on tissue staining patterns. Although antigen-retrieval protocols have the potential to be standardized, they continue to vary not only between laboratories [21] but also between distinct commercial kit preparations. An analysis specifically aimed at identifying the technical aspects responsible of false negative

Table 4 Details on false negative readings (0 and 1+ with amplification on FISH): participating center, antibody used, score at IHC, % of stained tumor cells and intensity of incomplete membrane staining

Sample ID	Reader	Antibody used for IHC	Score at IHC	% of stained tumour cells	intensity of incomplete membrane staining
6	B	A0485 rabbit polyclonal	0	0	
6	C	CB11	0	0	
8	C	CB11	0	0	
16	A	A0485	0	< 10	weak
16	B	A0485 rabbit polyclonal	0	0	
21	B	A0485 rabbit polyclonal	0	0	
21	G	A0485	0	0	
23	B	A0485 rabbit polyclonal	0	0	
23	D	4B5	0	0	
23	G	A0485	0	0	
23	L	4B5	0	0	
27	C	CB11	0	0	
27	G	A0485	0	0	
5	A	A0485	1+	70	weak
5	B	A0485 rabbit polyclonal	1+	>10	weak
5	C	CB11	1+	30	weak
6	G	A0485	1+	10	weak
6	I	4B5	1+	20	weak
8	B	A0485 rabbit polyclonal	1+	>10	weak
8	E	4B5	1+	45	weak
8	G	A0485	1+	30	weak
8	I	4B5	1+	50	moderate
8	L	4B5	1+	80	weak
13	B	A0485 rabbit polyclonal	1+	>10	weak
13	E	4B5	1+	15	weak
13	G	A0485	1+	40	weak
13	M	4B5	1+	>10	weak
14	B	A0485	1+	20	weak
14	C	CB11	1+	30	weak
14	D	4B5	1+	>90	weak
14	G	A0485	1+	30	weak
14	I	4B5	1+	30	moderate
14	L	4B5	1+	>10	moderate
14	M	4B5	1+	>10	weak
16	D	4B5	1+	40	weak
16	E	4B5	1+	15	weak
16	G	A0485	1+	40	weak
16	I	4B5	1+	>10	moderate
16	L	4B5	1+	20	weak
16	M	4B5	1+	>10	weak
21	A	A0485	1+	60	weak
21	C	CB11	1+	20	weak
21	D	4B5	1+	30	weak
21	E	4B5	1+	80	weak
21	I	4B5	1+	15	weak
21	L	4B5	1+	>10	weak
21	M	4B5	1+	>10	weak
22	B	A0485	1+	>10	weak
22	C	CB11	1+	50	weak

Table 4 (continued)

Sample ID	Reader	Antibody used for IHC	Score at IHC	% of stained tumour cells	intensity of incomplete membrane staining
23	A	A0485	1+	30	weak
23	C	CB11	1+	15	weak
23	M	4B5	1+	>10	weak
28	B	A0485	1+	30	moderate
28	C	CB11	1+	50	weak
28	G	A0485	1+	40	moderate
28	I	4B5	1+	>10	moderate

A0485, CB11 and 4B5: antibodies used as part of commercial kits
A0485 rabbit polyclonal: antibody used in house protocol
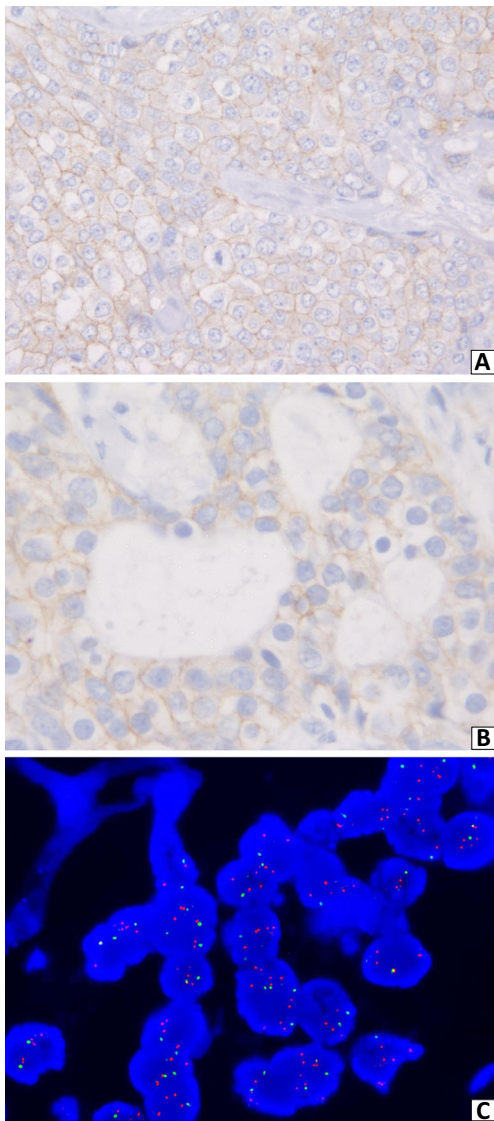
Fig. 1 Case 28. **a**: in eight laboratories this case was scored as moderate, complete, membrane staining in more than 10 % of tumor cells, equivocal, 2+. Case 28. **b**: in four laboratories this case was scored as moderate/weak, incomplete, membrane staining in more than 10 % of tumor cells, negative, 1+. Case 28. **c**: Presence of amplification, the average of CEP17 signals is less than three and the HER2/CEP17 ratio is ≥2
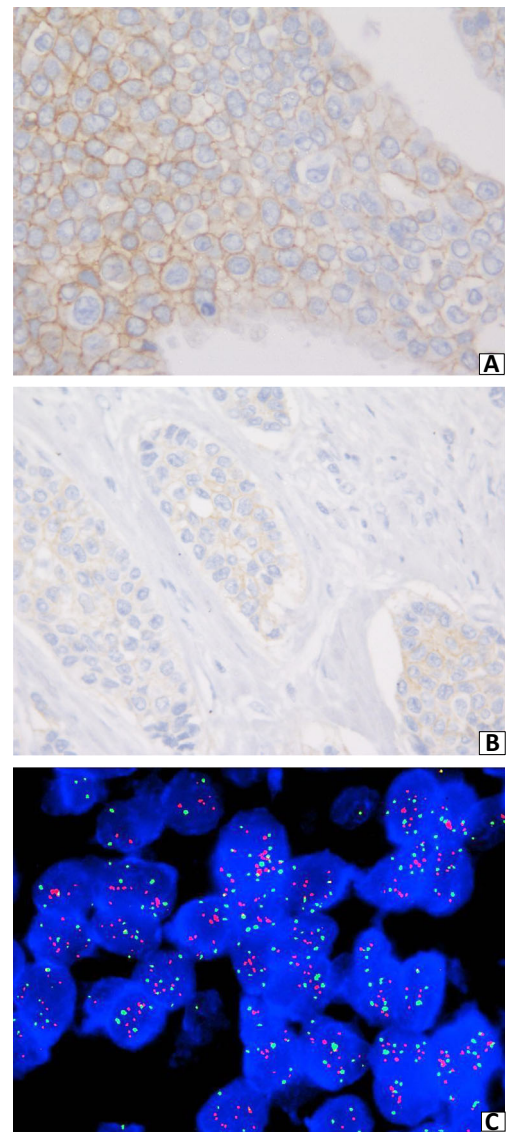
Fig. 2 Case 16. **a**: in four laboratories this case was scored as moderate, complete, membrane staining in more than 10 % of tumor cells, equivocal, 2+. Case 16. **b**: in six laboratories this case was scored as weak/moderate, incomplete, membrane staining in more than 10 % of tumor cells, negative, 1+. Case 16. **c**: Presence of amplification, the average of CEP17 signals is ≥3 and the average of HER2 copy number as ≥6

results could not be made due to differences in antibodies and procedures across participating laboratories.

Beginning with the 2007 ASCO/CAP guideline recommendations, participation in EQA program has become a mandatory requirement for all US laboratories performing HER2 testing. In Italy, however, participation in EQA program is still left to individual initiative though highly recommended. In our study, only the reference laboratory is routinely participating in an EQA program [NordiQC http://www.nordiqc.org/] for both IHC and FISH for HER2.

Reiner-Concin et al. [22] investigated the accuracy of HER2 IHC testing across 32 laboratories in Austria by using 10 BC cases, 2 of which were amplified on FISH. The proportion of 3+ scores among non-amplified BC cases was 1 % (2 out of 248), while the amplified BC cases that were classified as 0 or 1+ were 6 out of 63 (9.5 %), much lower than in our study. Their overall better performance compared to our study may be due to the disproportionate number of IHC 2+ equivocal cases included in our study that might have enhanced the discrepancies.

In terms of post-analytical phase, it is well known that the HER2 IHC scoring is subjective and inter-observer reproducibility can be problematic, especially for 2+ cases [23, 24].

Even when experienced pathologists are involved, K statistics of only 0.67 and 0.74 were achieved for two HER2 IHC tests [23]. A study involving 94 laboratories from 21 countries found 73 % agreement on staining of tumor samples, and the lack of reproducibility was mainly due to the stringency of applying the scoring criteria by pathologists [24]. Subjectivity in HER2 IHC interpretation represents a major problem as there is no consistent epithelial internal positive control for HER2 within non neoplastic breast tissue [25].

In the recent ASCO/CAP HER2 testing recommendation update [7], published immediately after our study was implemented, the IHC 2+ "equivocal" category has been expanded to include cases that would have previously been classified as 1+ negative (circumferential membrane staining that is incomplete, weak/moderate and within >10 % of tumor cells) or 0 negative (strong complete membrane staining within < or equal 10 % of tumor cells). Applying these revised guidelines to our data, the false negative results would drop to 13 (3 %), as the 43 1+ cases (Table 4) would now be classified as 2+. The adoption of these updated ASCO/CAP recommendations on HER2 testing [7] could therefore reduce false-negative HER2 results and improve consistency of interpretation criteria among pathologists.

In summary, on the basis of our results, a standardization of preanalytical phase could reduce false positive rates in HER2 determination by IHC.

Participation of pathological anatomy laboratories performing HER2 testing by IHC in EQA programs should be highly recommended or even made compulsory, as the system is able to identify laboratories with suboptimal performance that may need technical advice. Finally, updated ASCO/CAP recommendations [7] should be adopted as the widening of IHC 2+ "equivocal" category would improve overall accuracy of HER2 testing, as more cases would be classified in the this category and, consequently, also tested with a ISH method.

References

1. Popescu NC, King CR, Kraus MH (1989) Localization of the human erbB-2 gene on normal and rearranged chromosomes 17 to bands q12-21.32. Genomics 4:362–366
2. Perez EA, Cortes J, Gonzalez-Angulo AM, JMS et al (2014) HER2 testing: current status and future directions. Cancer Treat Rev 40: 276–284
3. Hanna WM, Ruschoff J, Bilous M et al (2014) HER2 in situ hybridization in breast cancer: clinical implications of polysomy 17 and genetic heterogeneity. Mod Pathol 27:4–18
4. Wolff AC, Hammond EH, Schwartz JN et al (2007) American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. J Clin Oncol 25:118–145
5. Provenzano E, Johnson N (2009) Overview of recommendations of HER2 testing in breast cancer. Diagn Histopathol 15:478–484
6. Walker RA, Bartlett JM, Dowsett M et al (2008) HER2 testing in UK: further update to recommendations. J Clin Pathol 61:818–824
7. Wolff AC, Hammond EH, Hicks DG et al (2013) Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. J Clin Oncol 31: 3997–4013
8. Hanna W, O'Malley FP, Barnes P et al (2007) Updated recommendations from the Canadian National Consensus Meeting on HER2/neu testing in breast cancer. Curr Oncol 4:149–153
9. Paik S, Bryant J, Tan-Chiu E et al (2002) Real-world performance of HER2 testing-National Surgical adjuvant Breast and Bowel Project Experience. J Natl Cancer Inst 94:852–854
10. Vogel CL, Bloom K, Burris H et al (2011) Discordance between central and local laboratory Her2 testing from a large HER2-negative population in VIRGO, a metastatic breast cancer registry. Cancer Res 71(24):188s
11. Reiner-Concin A (2008) External quality assurance in immunohistochemistry-is it the solution to a complex problem? Breast Care 3:78–79

12. Mirlacher M, Kasper M, Storz M et al (2004) Influence of slide aging on results of translational research studies using immunohistochemistry. Mod Pathol 17:1414–1420

13. Marchiò C, Lambros MB, Gugliotta P et al (2008) Does chromosome 17 centromere copy number predict polysomy in breast cancer? A fluorescence in situ hybridization and microarray-based CGH analysis. J Pathol 219:16–24

14. Vance GH, Barry TS, Bloom KJ et al (2009) Genetic heterogeneity in HER2 testing in breast cancer: panel summary and guidelines. Arch Pathol Lab Med 133:611–612

15. Vandem Bempt I, Van Loo P, Drijkoningen M et al (2008) Polysomy 17 in breast cancer: clinicopathologic significance and impact on HER-2 testing. J Clin Oncol 26:4869–4874

16. Cohen J (1968) Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 70:213–220

17. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174

18. Ellis IO, Dowsett M, Bartlett J et al (2000) Recommendations for HER2 testing in the UK. J Clin Pathol 53:890–892

19. von Wasielewski R, Krusche CA, Ruschoff J et al (2008) Implementation of external quality assurance trials for immunohistochemically determined breast cancer biomarkers in Germany. Breast Care 3:128–133

20. Schrohl AS, Pedersen HC, Jonsen SS et al (2011) Human epidermal growth factors receptors 2 (HER2) immunoreactivity: specificity of three pharmacodiagnostic antibodies. Histopathology 59:975–983

21. Leong TY-M, Cooper K, Leong AS-Y (2010) Immnunohistology - past, present and future. Adv Anat Pathol 17:404–418

22. Reiner-Concin A, Regitnig P, Dinges HP et al (2008) Practice of HER2 immunohistochemistry in breast carcinoma in Austria. Pathol Oncol Res 14:253–259

23. Bartlett JM, Going JJ, Mallon EA et al (2001) Evaluating HER2 amplification and overexpression in breast cancer. J Pathol 195:422–428

24. Rhodes A, Jasani B, Anderson E et al (2002) Evaluation of HER-2/neu immunohistochemical assay sensitivity and scoring on formalin-fixed and paraffin-processed cell lines and breast tumors: a comparative study involving results from laboratories in 21 countries. Am J Clin Pathol 118:408–417

25. Sauter G, Lee J, Bartlett JMS et al (2009) Guidelines for human epidermal growth factor receptor 2 testing: biologic and methodologic considerations. J Clin Oncol 27:1323–1333