



Concordance of immunohistochemistry for predictive and prognostic factors in breast cancer between biopsy and surgical excision: a single-centre experience and review of the literature

Chiara Rossi¹ · Sara Fraticelli¹ · Marianna Fanizza² · Alberta Ferrari³ · Elisa Ferraris⁴ · Alessia Messina¹ · Angelica Della Valle³ · Chiara Annunziata Pasqualina Anghelone³ · Angioletta Lasagna⁴ · Gianpiero Rizzo⁴ · Lorenzo Perrone⁴ · Maria Grazia Sommaruga² · Giulia Meloni² · Silvia Dallavalle² · Elisabetta Bonzano⁵ · Marco Paulli¹ · Giuseppe Di Giulio² · Adele Sgarella³ · Marco Lucioni¹

Received: 10 November 2022 / Accepted: 31 January 2023 / Published online: 21 February 2023
© The Author(s) 2023

Abstract

Purpose Accurate evaluation of breast cancer on bioptic samples is of fundamental importance to guide therapeutic decisions, especially in the neoadjuvant or metastatic setting. We aimed to assess concordance for oestrogen receptor (ER), progesterone receptor (PR), c-erbB2/HER2 and Ki-67. We also reviewed the current literature to evaluate our results in the context of the data available at present.

Methods We included patients who underwent both biopsy and surgical resection for breast cancer at San Matteo Hospital, Pavia, Italy, between January 2014 and December 2020. ER, PR, c-erbB2, and Ki-67 immunohistochemistry concordance between biopsy and surgical specimen was evaluated. ER was further analysed to include the recently defined ER-low-positive in our analysis.

Results We evaluated 923 patients. Concordance between biopsy and surgical specimen for ER, ER-low-positive, PR, c-erbB2 and Ki-67 was, respectively, 97.83, 47.8, 94.26, 68 and 86.13%. Cohen's κ for interobserver agreement was very good for ER and good for PR, c-erbB2 and Ki-67. Concordance was especially low (37%) in the c-erbB2 1+ category.

Conclusion Oestrogen and progesterone receptor status can be safely assessed on preoperative samples. The results of this study advise caution in interpreting biopsy results regarding ER-low-positive, c-erbB2/HER2 and Ki-67 results due to a still suboptimal concordance. The low concordance for c-erbB2 1+ cases underlines the importance of further training in this area, in the light of the future therapeutic perspectives.

Keywords Breast cancer · Preoperative diagnosis · Immunohistochemistry · Biomarkers · Er-low-positive · her2-low

✉ Chiara Rossi
chiara.rossi05@universitadipavia.it

¹ Department of Molecular Medicine, Unit of Anatomic Pathology, University of Pavia, IRCCS San Matteo Hospital Foundation, Pavia, Italy

² Unit of Breast Radiology, IRCCS San Matteo Hospital Foundation, Pavia, Italy

³ Department of Surgical Sciences, General Surgery 3—Breast Surgery, IRCCS San Matteo Hospital Foundation, Pavia, Italy

⁴ Unit of Medical Oncology, IRCCS San Matteo Hospital Foundation, Pavia, Italy

⁵ School in Experimental Medicine, Unit of Radiational Oncology, University of Pavia, IRCCS San Matteo Hospital Foundation, Pavia, Italy

Introduction

Assessment of breast cancer biomarkers has become a staple of routine histopathology for every colleague working in this field. Assessment and quantification of oestrogen receptor (ER), progesterone receptor (PR), and c-erbB2/HER2 are used daily by the clinicians making fundamental therapeutic choices for the patients. On the other hand, Ki-67 has struggled to join this established trifecta in the routine management and risk stratification of breast cancer patients [1–4] and its prognostic and predictive value is restricted to very specific settings in breast cancer; recently, the results from the monarchE study has suggested a prognostic role for Ki-67 $\geq 20\%$ in patients with early breast cancer treated

with cyclin-dependent kinase 4 and 6 (CDK4/6) inhibitors in combination with endocrine therapy [5].

Recently, changes in the stratification of ER positivity have been put forward by ASCO/CAP through an update of the recommendations for ER and PR determination, with the formal introduction of the ER-low-positive (ER-LP, defined as 1–10% of nuclear positivity in the tumour) [6]; this category has been known for some time to share more similarities with basal-type triple-negative breast cancer than with the luminal group [7–10], in morphological aspects, molecular signature and clinical behaviour.

In the same way, the results of the DESTINY-Breast03 trial [11, 12], and the identification of the HER2-low category [13, 14] as a new group of patients who may significantly benefit from anti-HER2 target therapy administration, have underlined once again the importance of the c-erbB2 status and its assessment with both immunohistochemistry and molecular techniques.

Core needle biopsy (CNB) is the most common method for diagnosis of breast cancer, and it has been demonstrated to be a reliable indicator of the surgical specimen results [15, 16]. Assessment of biomarkers on the preoperative specimen is required to administer neoadjuvant therapy to the selected patients that benefit from it and, in case of a complete pathological response, it represents the only sample of that tumour available; also for metastatic patients, who are not eligible for surgical resection, the bioptical sample is the only one that can be used to take life-changing clinical decisions.

It is clear, then, the importance of a reliable assessment of these biomarkers on biopsy. Given the steadily increasing request for precise molecular characterization of breast cancer to ensure correct patient management, we aimed to retrospectively evaluate our patient cohort for reproducibility of the biopsy results for ER, PR, c-erbB2, and Ki-67, focusing also on the recently defined ER-LP and HER2-low groups. We also review the recent literature on the topic to validate our results in the context of the international available results.

Methods

Patients and clinicopathological characteristics

We reviewed the records of 1654 breast cancer patients who underwent both biopsy and surgical excision at the Department of Breast Surgery, San Matteo Hospital, between January 2014 and December 2020. We excluded patients for which ER, PR, c-erbB2 or Ki-67 values were missing, patients with in situ or microinvasive-only breast cancer, patients with multifocal tumours, metastatic disease and patients who underwent neoadjuvant chemotherapy. Our final cohort comprised 923 patients and their clinical

characteristics are summarized in Table 1. The study was conducted according to the guidelines of the Declaration of Helsinki.

Pathology evaluation

Samples were fixed in 10% neutral buffered formalin and embedded in paraffin before histopathological evaluation. 4-to-5 μm -thick sections were cut and stained with hematoxylin and eosin (HE), and unstained sections were used for immunohistochemistry with antibodies anti-ER (clone EP1, Dako Omnis), anti-PR (clone PgR 1294, Dako Omnis), c-erbB2 (clone A0485, Dako Omnis) and Ki-67 (clone MIB-1, Dako Omnis). All immunoreactions were carried out on a Dako Omnis platform (Dako, Glostrup, Denmark). All cases were seen by at least one pathologist (M.L. and/or C.R.) expert in breast pathology, who also revised all the discrepant cases prior to the final diagnosis.

ER and PR were defined positive when $\geq 1\%$ of the tumour cell nuclei showed immunostaining, according to the 2010 ASCO/CAP guidelines [17]. ER was further stratified into LP and positive using the 10% cut-off, according to the 2020 ASCO/CAP guidelines update [6]. Ki-67 was scored ‘high’ when $\geq 20\%$ of the tumour nuclei were positive, taking into account the cut-off clinically used to define the Luminal B class according to the 2013 St Gallen International Breast Cancer Conference experts Panel opinion [18]. Cells positive for Ki-67 were scored over 100 cells in both ‘cold’ and ‘hot’ tumour areas, and the final value represented an average between those of the different areas. This method is similar to the one recommended by the International Ki-67 in Breast Cancer Working Group (IKWG) in their 2021 updated recommendations [19], with the only difference that the recommended online scoring app was not used. c-erbB2 was scored according to the ASCO/CAP 2013 guidelines and 2018 Focused Update [20, 21] as 0, 1+,

Table 1 Clinicopathological characteristics of the cohort

Characteristic	n (%)
Sex	
Female	918 (99.4)
Male	5 (0.6)
Median age	55
Histology	
No special type	707 (77)
Lobular	126 (14)
Other special histotypes	90 (9)
Grade	
G1	76 (8)
G2	537 (58)
G3	310 (34)

2 + or 3 + depending on intensity and completeness of the membrane staining. FISH was performed in all the equivocal cases, but the results are not reported in this paper since we focus on the immunohistochemical evaluation alone.

Cohen's kappa (κ) was used to measure the interobserver agreement between biopsy and surgical specimen; weighted κ was used when concordance between more than one result was evaluated to account for close matches. κ values < 0.20 were interpreted as poor agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 good agreement, 0.81–0.99 very good agreement and 1 perfect agreement. P-values were calculated with Fisher's exact test and chi-squared test, when appropriated, on GraphPad Prism 5, and *p*-values < 0.05 were deemed significant.

Results

ER

ER was positive in 829/923 (89%) biopsies and in 831/923 (90%) surgical resection specimens, with a concordance of 97.83% ($n = 903/923$), a Cohen's κ of 0.880 (very good agreement) and a *p*-value < 0.0001. The concordant and discrepant results are reported in Table 2, with 9 positive results (9/923, 1%) later reported as negative on the surgical specimen and 11 negative results (11/923, 1.2%) that were upgraded to positive on the final pathology report. These changes would have had a clinical impact with either withholding or administration of endocrine therapy in these patients, if the decision was based only on the biopsy results.

When the positive results were further stratified into ER-LP and ER-positive (Table 3) the overall agreement was still very good (97.18%, weighted $\kappa = 0.924$, *p* value < 0.0001), but the concordance for the ER-LP category itself was low ($n = 11/23$, 47.8%). Of the 12 discordant ER-LP results on biopsy, 8/12 (66.7%) were ER-negative on the surgical specimen and 4/12 (33.3%) were ER-positive. Of these four cases, three cases were only slightly above the

Table 2 Concordance between biopsy and surgical specimen for oestrogen receptor, progesterone receptor and Ki-67. ER, oestrogen receptor, PR, progesterone receptor

Biopsy	Surgical specimen		Total
	+	–	
ER			
+	820	9	829
–	11	83	94
PR			
+	742	17	759
–	36	128	164
Ki-67			
+	235	21	256
–	107	560	667

Table 3 Concordance between biopsy and surgical specimen for oestrogen receptor-negative, oestrogen receptor-LP and oestrogen receptor-positive cases. ER, oestrogen receptor

Biopsy	Surgical specimen			Total
	ER-negative	ER-LP	ER-positive	
ER-negative	83	10	1	94
ER-LP	8	11	4	23
ER-positive	1	2	803	806

cut-off for ER-LP (15%), whilst one showed positivity in 60% of the cells (Fig. 1).

PR

PR was positive in 759/923 (82%) biopsies and in 778/923 (84%) surgical resection specimens. Concordance was 94.26% ($n = 870/923$), Cohen's κ was 0.794 (good agreement) and *p* value was < 0.0001. Most ($n = 36/53$, 67.9%) of the discordant results were biopsies reported as PR negative that were later reported as PR positive on the surgical specimen. The complete results are summarized in Table 2.

c-erbB2/HER2

c-erbB2 was found to be concordant in 68% of cases ($n = 631/923$), and Table 4 details the results for each category. Cohen's weighted κ was 0.675 (good agreement) and *p*-value was < 0.0001. Breaking down the results for each category, 1 + was the least concordant group (37% vs 83%, 79% and 97% for 0, 2 + and 3 +, respectively), with 72% ($n = 136/188$) of the discordant results being diagnosed as 0 on the surgical specimen.

According to current clinical practice, only four (4/923, 0.4%) patients had changes in the diagnosis that could significantly impact their treatment choice (two 1 + biopsies later upgraded to 3 and two 3 + biopsies downgraded to 1 +). They were all, except for one locally advanced cancer, early breast cancers that would have not been candidates for neoadjuvant treatment. However, on the account of the results of the biopsy alone, they would have been denied a potentially life-saving treatment or subjected to the toxicities of an ineffective one.

ki-67

Ki-67 was high in 256/923 (28%) biopsies and in 357/923 (38%) surgical resection specimens. Concordance was 86.13% ($n = 759/923$), Cohen's κ was 0.686 (good agreement), and *p* value was < 0.0001. Concordance and discordance are reported in Table 2; most of the discordant results ($n = 107/128$, 84%) were biopsies in which Ki-67 was

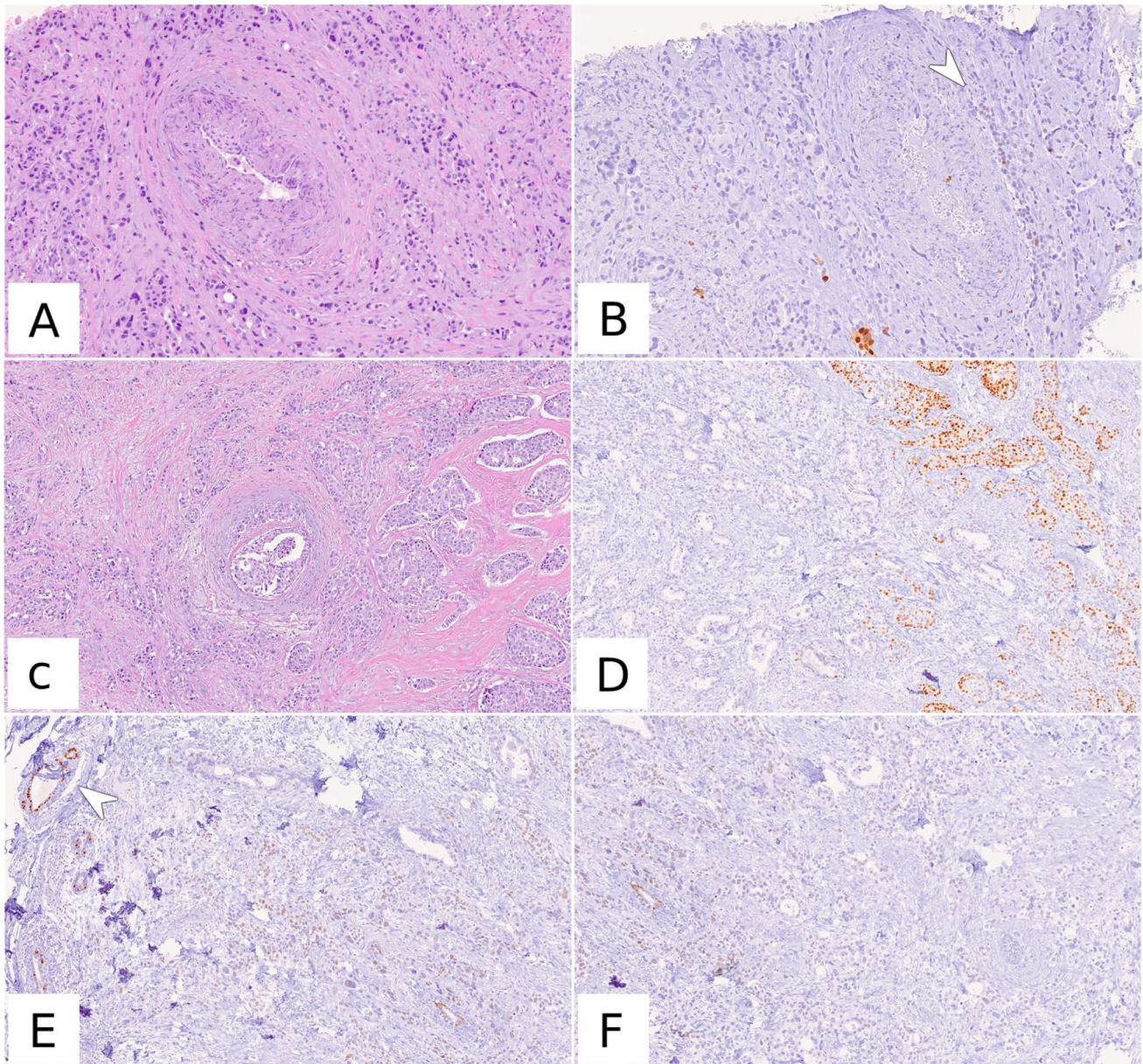


Fig. 1 Example of ER-LP discordance between biopsy and surgical specimen. On biopsy, (A, HE, 10x, B, ER immunohistochemistry, 10x) the tumour showed only faint, very focal (arrowhead) positivity for ER, that was quantified at 1%. The surgical specimen (C, HE, 5x,

D–F, ER immunohistochemistry, 5x) revealed a dishomogeneous and faint ER positivity. Note the positive internal control (arrowhead) in the top-left corner of panel (E)

Table 4 Concordance between biopsy and surgical specimen for c-erbB2 status

Biopsy	Surgical Specimen				Total
	0	1+	2+	3+	
0	321	54	14	0	389
1+	136	112	50	2	300
2+	2	29	130	3	164
3+	0	2	0	68	70

reported as low that were later upgraded on the excision specimen.

Discussion

ER

In the published literature ER concordance between biopsy and surgical excision averages at 93.3% (range 78.7–99.1%,

Table 5). Our concordance was 97.83%, slightly higher than the average value but in keeping with the known results. This variance can also be attributed to the different cut-offs and scoring methods used in the single laboratories, reported in the table that impair their universal reproducibility.

Several factors are recognized to influence concordance between biopsy and surgical specimen, with the pre-analytical phase being the most important to standardize intra-laboratory results, trying to avoid under-fixation [17, 22]. The choice of the antibody clone and the immunostaining platform has also been demonstrated to be relevant in the over- or underestimation of the ER percentage [23], especially when comparing the widely used Dako and Ventana clones.

The high percentage (66.7% of ER-LP cases) of ER-LP that tested ER-negative on the surgical specimen suggests that ER-LP results more frequently represent an “overcalling” of a non-luminal carcinoma, in line with the current understanding of ER-LP tumour biology, that relates them more closely to this group. Recent published data suggest also that artefactual reduced intensity of staining of ER in normal breast tissue adjacent to the neoplasia may concur

to report these cases as ER-LP [24], therefore underlying the importance of the presence of internal and/or on-slide controls in the assessment of these borderline cases.

In summary, our results suggest that caution should be taken when calling ER-LP on biopsy and for subsequent decision-making, and further studies are needed to define if this category can be safely defined on a biopsy specimen.

PR

Our final concordance for PR was 94.26%. From a review of the literature, the average reported concordance is 87.3% (range 73.5–95%, Table 6), and our results fall on the upper side of this range.

The lower concordance of PR assessment on biopsy and surgical specimen reflects its naturally occurring dishomogeneity in breast normal tissue and tumours, owing to its nature as a down-stream ER effector and therefore requiring an intact ER pathway to be strongly expressed. Our results with a higher proportion of upgrades rather than downgrades on the surgical specimen (67.9% vs 32.1%) suggest that

Table 5 Review of the literature regarding oestrogen receptor concordance

	No of samples	Concordance (%)	IHC clone
Badoual, 2005 [25]	103	90.3	1D5, Immunotech
Burge, 2006 [26]	87	95	SF11, Biocare Medical
Hodi, 2007* [27]	338	98.8	1D5, Dako
Wood, 2007 [28]	100	95.8	ID11, Dako
Arnedos, 2009 [29]	336	98.2	SP1, Ventana
Tamaki, 2010 [30]	353	94.9	SF11, Ventana
Uy, 2010§ [31]	160	81.9	1D5, Dako
Lorgis, 2011° [32]	175	84	6F11, Novocastra
Ough, 2011 [33]	209	88	Not reported
Li, 2012# [34]	2450	92.8	
Seferina, 2012° [35]	526	89.5	1D5, Dako
Dekker, 2013 [36]	122	99.1	1D5, Dako
Greer, 2013 [37]	205	89	6F11, Leica 1D5, Dako
Munch-Petersen, 2013 [38]	89	98	SP1, Dako
Motamedolshariati, 2014 [39]	30	96.7	Not reported
Chen, 2017 [40]	996	78.8	SP1, Ventana
Ensani, 2017 [41]	100	90	1D5, Biogenex
Kombak, 2017 [42]	284	93.3	6F11, Leica
Meattini, 2017° [43]	101	94.1	SF11, Ventana
You, 2017§ [44]	1219	97.1	SF11, Novocastra
Berghuis, 2019 [45]	684	95.5	Multiple
Jeong, 2020§ [46]	623	96.5	SP1, Ventana
Shanmugalingam, 2022 [47]	484	96.7	SP1, Ventana
Sløstad, 2022 [48]	961	90.8	LDT

*McCarty's h scoring system, #meta-analysis, ° positivity threshold: 10%, § Allred score, IHC, immunohistochemistry, LDT, laboratory-developed test.

Table 6 Review of the literature regarding PR concordance

	No of samples	Concordance (%)	IHC clone
Badoual, 2005 [25]	103	89.3	PR10A9, Immunotech
Burge, 2006 [26]	87	89	PgR 636, Dako
Wood, 2007 [28]	100	90.3	PgR 636, Dako
Arnedos, 2009 [29]	336	85	1E2, Ventana
Tamaki, 2010 [30]	353	89.5	6, Ventana
Uy, 2010§ [31]	160	85.6	PgR 636, Dako
Lorgis, 2011° [32]	175	78.3	PgR 636, Dako
Ough, 2011 [33]	209	78	Not reported
Li, 2012# [34]	2448	84.8	
Seferina, 2012° [35]	526	83.6	PgR 636, Dako
Greer, 2013 [37]	205	89	PgR 1294, Dako PgR 636, Dako
Motamedolshariati, 2014 [39]	30	90	Not reported
Chen, 2017 [40]	985	73.5	1E2, Ventana
Ensani, 2017 [41]	100	81	PR88, Biogenex
Kombak, 2017 [42]	284	89.4	16, Leica
Meattini, 2017° [43]	101	88.1	16, Ventana
You, 2017§ [44]	1219	95	16, Novocastra
Berghuis, 2019 [45]	890	84.8	Multiple
Jeong, 2020§ [46]	623	93	1E2, Ventana
Shanmugalingam, 2022 [47]	484	93.2	1E2, Ventana
Slostad, 2022 [48]	961	87.2	LDT

meta-analysis, ° positivity threshold: 10%, § Allred score, *IHC* immunohistochemistry, *LDT* laboratory-developed test

indeed tumour heterogeneity, with a negative spot sampled on biopsy, may be a significant issue in PR assessment.

Given this heterogeneity, PR concordance is especially sensitive to sampling artefacts, especially undersampling of the target lesion. From the current literature, four represents the minimum number of biopsy cores that should be retrieved to ensure a correct preoperative evaluation [49].

c-erbB2/HER2

In the published literature, concordance for c-erbB2 when evaluated with immunohistochemistry alone averages at 85.4% (range 56–98.8%, Table 7); our series demonstrates a concordance in 68% of cases, lower than the reported values, but still in the published range.

Breaking down the results, the 1+ category was the one with the lowest concordance, with only 37% of results getting confirmation on the surgical specimens. Whilst we had previously reported that this discordance for the 1+ category was not likely to have a significant therapeutic impact for the patients [50], the recent introduction of the HER2-low category as a subset of patients that could benefit from the administration of targeted anti-HER2 therapy will radically change that in the upcoming years.

The 2013 ASCO/CAP guidelines define 1+ as incomplete membrane staining that is faint or barely perceptible and within > 10% of tumour cells and readily appreciated in a contiguous population using a low-power objective, whereas 0 is defined as absence of staining or incomplete membrane staining that is faint/barely perceptible and within ≤ 10% of tumour cells [20]. This diagnosis must be made on immunohistochemistry alone, with no help coming from ancillary molecular studies.

However, in everyday practice, this distinction is not an easy one to make especially on biopsy, where crushing or technical artefacts and tumour heterogeneity make it a particularly challenging task. In the light of the growing importance of distinguishing between 0 and 1+, we stress the importance of a dedicated, up-to-date breast pathologist examining these specimens, especially in those difficult cases in which the diagnosis is not immediately clear.

Ki-67

Ki-67 represents an unfulfilled promise in the field of breast cancer; despite being widely used as a marker of proliferation, it has failed time and time again to reach prognostic and predictive significance.

Table 7 Review of the literature regarding c-erbB2/HER2 concordance

	No of samples	Concordance (%)	IHC
Burge, 2006 [26]	81	96	Dako
Wood, 2007 [28]	100	86.6	Dako
Arnedos, 2009 [29]	331	98.8	Ventana
Lebeau, 2010 [51]	500	90.4	Dako
Tamaki, 2010 [30]	353	89.3	Ventana
Lorgis, 2011* [32]	175	98.3	Ventana
Ough, 2011 [33]	209	56	Not reported
Lee, 2012* [52]	300	98	Dako
Seferina, 2012° [35]	526	80.6	Dako
Dekker, 2013 [36]	122	96.4	Dako
Munch-Petersen, 2013 [38]	89	84.0	Ventana
Motamedolshariati, 2014 [39]	30	93.3	Dako
Chen, 2017 [40]	941	62.6	Ventana
Ensani, 2017 [41]	100	97.3	Not reported
Kombak, 2017 [42]	243	90.1	Leica
You, 2017 [44]	1219	84.6	Ventana
Slostad, 2022 [48]	961	73.4%	LDT

*IHC + FISH, °IHC + SISH (Ventana), IHC immunohistochemistry, LDT laboratory-developed test

This is at least in part ascribable to the still unclear nature of this biomarker that, despite the best efforts that has been put into it, still does not have a fixed biological cut-off; guidelines for what qualifies as ‘high’ and ‘low’ Ki-67, and whether these definitions truly represent different biological entities, are still unclear.

Recent recommendations from the IKWG [19, 53] report that sufficient levels of evidence of the prognostic value of Ki-67 exists only in the setting of ER-positive early-stage breast cancer, where levels $\geq 5\%$ and $\geq 30\%$, respectively, may favour withholding or administration of chemotherapy. Historically, the 2013 St Gallen International Breast Cancer Conference suggested a 20% cut-off for the definition of ‘high’ Ki-67 in the definition of the surrogate intrinsic subtypes of breast cancer [18], but still advised that different

cut-offs could be adopted by single laboratories. As reported in literature [40, 43, 44, 46, 54, 55], most laboratories use either a 14% or a 20% cut-off.

Our concordance was in line with the values reported in literature (Table 8), especially with those reported by You et al. [44], whose cut-off and population are similar to those of this work. Similar values are reported also with a 14% cut-off by Kim et al., [55], Ahn et al., [54] and Meattini et al. [43], whilst lower concordance levels are reported by studies in which a Ventana antibody is used for immunohistochemistry [40, 46] irrespectively of the cut-off used.

To our knowledge, this work represents one of the largest single-centre series present in literature and the first one where issues within the ER-LP and HER2-low categories were specifically addressed. Moreover, all the

Table 8 Review of the literature regarding Ki-67 concordance

	No of samples	Concordance (%)	IHC clone
Ough, 2011 [33]	209	59	Not reported
Greer, 2013 [37]	205	73	MIB-1, Dako
Kim, 2016* [55]	310	85.8	MIB-1, Dako
Ahn, 2017* [54]	89	82	MIB-1, Dako
Chen, 2017* [40]	696	70.3	30–9, Ventana
Kombak, 2017* [42]	236	80.9	K2, Leica
Meattini, 2017* [43]	101	88.1	Mib-1, Immunotech
You, 2017§ [44]	1219	87	MIB-1, Dako
Jeong, 2020§ [46]	623	78.7	MIB-1, Ventana
Shanmugalingam, 2022# [47]	484	70.5	Not reported

*Cut-off high expression: 14%, § cut-off high expression: 20%, #, stratified into three categories (low, intermediate, high), IHC, immunohistochemistry

immunohistochemistry reactions were carried out with the same antibodies on the same platform, and a breast pathologist (M.L. and/or C.R.) was involved in the diagnosis of all cases, ensuring a high degree of homogeneity in the test results. Possible limitations include the retrospective nature of the study and the lack of data regarding the HER2 2+ amplification status; however, this study aimed to evaluate only the immunohistochemical concordance for HER2, with particular attention to the HER2-low categories.

We confirmed a very good agreement for ER assessment on biopsy, and a still satisfactory, albeit lower, concordance for PR that falls just short of the very good agreement cut-off. Ki-67 evaluation instead was confirmed to be slightly less reliable than ER and PR, with a significantly lower concordance, and could warrant evaluation also on the surgical specimen, if available, in the light of potential treatment options [5].

ER-LP analysis revealed that, even if the global ER concordance is still satisfactory, the concordance for the specific category is still low and must be further investigated to define the boundaries within which it can be safely assessed on a biopsy sample.

The low concordance for the 1+c-erbB2 immunohistochemistry is particularly relevant in the context of the new therapeutic advances involving it and highlights the technical difficulties in consistently implementing the current diagnostic criteria available for the diagnosis.

In the light of the quick and exciting evolution of the therapeutic landscape of breast cancer, it is important that caution is taken in evaluating biopsy samples, especially for those predictive factors that may significantly impact the therapeutic options of the patient. Although only a very small number of patients in our series would have received an inappropriate treatment based on biopsy results alone, our data underline the relevance of surgical sample retesting, at least in selected cases, including large tumours and cases with discrepant histological characteristics. Specific training needs should be addressed by the national and international pathology societies, especially in those still grey areas of ER-LP and HER2-low categories, where the difference between a negative and a positive result may withhold a target therapy, or cause patients to undergo unnecessary, and often quite burdensome, aggressive therapy.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Chiara Rossi and Marco Lucioni. The first draft of the manuscript was written by Chiara Rossi and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open access funding provided by Università degli Studi di Pavia within the CRUI-CARE Agreement. The authors declare that no

funds, grants, or other support were received during the preparation of this manuscript.

Data availability The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Amin MB, American Joint Committee on Cancer, American Cancer Society (2017) AJCC cancer staging manual, Eight edition/ editor-in-chief, Mahul B. Amin, MD, FCAP ; editors, Stephen B. Edge, MD, FACS [and 16 others] ; Donna M. Gress, RHIT, CTR-Technical editor ; Laura R. Meyer, CAPM-Managing editor. American Joint Committee on Cancer, Springer, Chicago
- Harris LN, Ismaila N, McShane LM et al (2016) Use of biomarkers to guide decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer: American society of clinical oncology clinical practice guideline. *JCO* 34:1134–1150. <https://doi.org/10.1200/JCO.2015.65.2289>
- Gennari A, André F, Barrios CH et al (2021) ESMO clinical practice guideline for the diagnosis, staging and treatment of patients with metastatic breast cancer. *Ann Oncol* 32:1475–1495. <https://doi.org/10.1016/j.annonc.2021.09.019>
- Cardoso F, Kyriakides S, Ohno S et al (2019) Early breast cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 30:1194–1220. <https://doi.org/10.1093/annonc/mdz173>
- Harbeck N, Rastogi P, Martin M et al (2021) Adjuvant abemaciclib combined with endocrine therapy for high-risk early breast cancer: updated efficacy and Ki-67 analysis from the monarchE study. *Ann Oncol* 32:1571–1581. <https://doi.org/10.1016/j.annonc.2021.09.015>
- Allison KH, Hammond MEH, Dowsett M et al (2020) Estrogen and progesterone receptor testing in breast cancer: ASCO/CAP guideline update. *JCO* 38:1346–1366. <https://doi.org/10.1200/JCO.19.02309>
- Gloyeske NC, Dabbs DJ, Bhargava R (2014) Low ER+ breast cancer. *Am J Clin Pathol* 141:697–701. <https://doi.org/10.1309/AJCP34CYSATWFDPO>
- Prabhu JS, Korlimarla A, Desai K et al (2014) A majority of low (1–10%) ER positive breast cancers behave like hormone receptor negative tumors. *J Cancer* 5:156–165. <https://doi.org/10.7150/jca.7668>

9. Dieci MV, Griguolo G, Bottosso M et al (2021) Impact of estrogen receptor levels on outcome in non-metastatic triple negative breast cancer patients treated with neoadjuvant/adjuvant chemotherapy. *NPJ Breast Cancer* 7:101. <https://doi.org/10.1038/s41523-021-00308-7>
10. Sheffield BS, Kos Z, Wang X et al (2015) Molecular profiling of ER weakly-positive breast cancer. *JCO* 33:525–525. https://doi.org/10.1200/jco.2015.33.15_suppl.525
11. Cortés J, Kim S-B, Chung W-P et al (2022) Trastuzumab deruxtecan versus trastuzumab emtansine for breast cancer. *N Engl J Med* 386:1143–1154. <https://doi.org/10.1056/NEJMoa2115022>
12. Modi S, Jacot W, Yamashita T et al (2022) Trastuzumab Deruxtecan in previously treated HER2-low advanced breast cancer. *N Engl J Med NEJ*. <https://doi.org/10.1056/NEJMoa2203690>
13. Marchiò C, Annaratone L, Marques A et al (2021) Evolving concepts in HER2 evaluation in breast cancer: heterogeneity, HER2-low carcinomas and beyond. *Semin Cancer Biol* 72:123–135. <https://doi.org/10.1016/j.semcancer.2020.02.016>
14. Tarantino P, Hamilton E, Tolaney SM et al (2020) HER2-low breast cancer: pathological and clinical landscape. *JCO* 38:1951–1962. <https://doi.org/10.1200/JCO.19.02488>
15. Focke CM, Decker T, van Diest PJ (2017) Reliability of the Ki67-labelling index in core needle biopsies of luminal breast cancers is unaffected by biopsy volume. *Ann Surg Oncol* 24:1251–1257. <https://doi.org/10.1245/s10434-016-5730-1>
16. Clark BZ, Onisko A, Assylbekova B et al (2019) Breast cancer global tumor biomarkers: a quality assurance study of intratumoral heterogeneity. *Mod Pathol* 32:354–366. <https://doi.org/10.1038/s41379-018-0153-0>
17. Hammond MEH, Hayes DF, Dowsett M et al (2010) American society of clinical oncology/college of American pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *JCO* 28:2784–2795. <https://doi.org/10.1200/JCO.2009.25.6529>
18. Goldhirsch A, Winer EP, Coates AS et al (2013) Personalizing the treatment of women with early breast cancer: highlights of the St Gallen international expert consensus on the primary therapy of early breast cancer 2013. *Ann Oncol* 24:2206–2223. <https://doi.org/10.1093/annonc/mdt303>
19. Nielsen TO, Leung SCY, Rimm DL et al (2021) Assessment of Ki67 in breast cancer: updated recommendations from the international Ki67 in breast cancer working group. *JNCI* 113:808–819. <https://doi.org/10.1093/jnci/djaa201>
20. Wolff AC, Hammond MEH, 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. *JCO* 31:3997–4013. <https://doi.org/10.1200/JCO.2013.50.9984>
21. Wolff AC, Hammond MEH, Allison KH et al (2018) Human epidermal growth factor receptor 2 testing in breast cancer: American society of clinical oncology/college of American pathologists clinical practice guideline focused update. *Arch Pathol Lab Med* 142:1364–1382. <https://doi.org/10.5858/arpa.2018-0902-SA>
22. Allred DC (2008) Problems and solutions in the evaluation of hormone receptors in breast cancer. *JCO* 26:2433–2435. <https://doi.org/10.1200/JCO.2007.15.7800>
23. Kornaga EN, Klimowicz AC, Guggisberg N et al (2016) A systematic comparison of three commercial estrogen receptor assays in a single clinical outcome breast cancer cohort. *Mod Pathol* 29:799–809. <https://doi.org/10.1038/modpathol.2016.74>
24. Caruana D, Wei W, Martinez-Morilla S et al (2020) Association between low estrogen receptor positive breast cancer and staining performance. *NPJ Breast Cancer* 6:5. <https://doi.org/10.1038/s41523-020-0146-2>
25. Badoual C, Maruani A, Ghorra C et al (2005) Pathological prognostic factors of invasive breast carcinoma in ultrasound-guided large core biopsies—correlation with subsequent surgical excisions. *The Breast* 14:22–27. <https://doi.org/10.1016/j.breast.2004.07.005>
26. Burge CN, Chang HR, Apple SK (2006) Do the histologic features and results of breast cancer biomarker studies differ between core biopsy and surgical excision specimens? *The Breast* 15:167–172. <https://doi.org/10.1016/j.breast.2005.06.004>
27. Hodi Z, Chakrabarti J, Lee AHS et al (2006) The reliability of assessment of oestrogen receptor expression on needle core biopsy specimens of invasive carcinomas of the breast. *J Clin Pathol* 60:299–302. <https://doi.org/10.1136/jcp.2006.036665>
28. Wood B, Junckerstorff R, Sterrett G et al (2007) A comparison of immunohistochemical staining for oestrogen receptor, progesterone receptor and HER-2 in breast core biopsies and subsequent excisions. *Pathology* 39:391–395. <https://doi.org/10.1080/00313020701444465>
29. Arnedos M, Nerurkar A, Osin P et al (2009) Discordance between core needle biopsy (CNB) and excisional biopsy (EB) for estrogen receptor (ER), progesterone receptor (PgR) and HER2 status in early breast cancer (EBC). *Ann Oncol* 20:1948–1952. <https://doi.org/10.1093/annonc/mdp234>
30. Tamaki K, Sasano H, Ishida T et al (2010) Comparison of core needle biopsy (CNB) and surgical specimens for accurate preoperative evaluation of ER, PgR and HER2 status of breast cancer patients. *Cancer Sci* 101:2074–2079. <https://doi.org/10.1111/j.1349-7006.2010.01630.x>
31. Uy GB, Laudico AV, Carnate JM et al (2010) Breast cancer hormone receptor assay results of core needle biopsy and modified radical mastectomy specimens from the same patients. *Clin Breast Cancer* 10:154–159. <https://doi.org/10.3816/CBC.2010.n.021>
32. Lorgis V, Algros MP, Villanueva C et al (2011) Discordance in early breast cancer for tumour grade, estrogen receptor, progesterone receptors and human epidermal receptor-2 status between core needle biopsy and surgical excisional primary tumour. *Breast* 20:284–287. <https://doi.org/10.1016/j.breast.2010.12.007>
33. Ough M, Velasco J, Hieken TJ (2011) A comparative analysis of core needle biopsy and final excision for breast cancer: histology and marker expression. *Am J Surg* 201:692–694. <https://doi.org/10.1016/j.amjsurg.2010.02.015>
34. Li S, Yang X, Zhang Y et al (2012) Assessment accuracy of core needle biopsy for hormone receptors in breast cancer: a meta-analysis. *Breast Cancer Res Treat* 135:325–334. <https://doi.org/10.1007/s10549-012-2063-z>
35. Seferina SC, Nap M, van den Berkmortel F et al (2013) Reliability of receptor assessment on core needle biopsy in breast cancer patients. *Tumor Biol* 34:987–994. <https://doi.org/10.1007/s13277-012-0635-5>
36. Dekker TJA, Smit VTHBM, Hooijer GJK et al (2013) Reliability of core needle biopsy for determining ER and HER2 status in breast cancer. *Ann Oncol* 24:931–937. <https://doi.org/10.1093/annonc/mds599>
37. Greer LT, Rosman M, Mylander WC et al (2013) Does breast tumor heterogeneity necessitate further immunohistochemical staining on surgical specimens? *J Am Coll Surg* 216:239–251. <https://doi.org/10.1016/j.jamcollsurg.2012.09.007>
38. Munch-Petersen HD, Rasmussen BB, Balslev E (2014) Reliability of histological malignancy grade, ER and HER2 status on core needle biopsy vs surgical specimen in breast cancer. *APMIS* 122:750–754. <https://doi.org/10.1111/apm.12213>
39. Motamedolshariati M, Memar B, Aliakbaian M et al (2014) Accuracy of prognostic and predictive markers in core needle breast biopsies compared with excisional specimens. *Breast Care (Basel)* 9:107–110. <https://doi.org/10.1159/000360787>
40. Chen J, Wang Z, Lv Q et al (2017) Comparison of core needle biopsy and excision specimens for the accurate evaluation of breast

- cancer molecular markers: a report of 1003 cases. *Pathol Oncol Res* 23:769–775. <https://doi.org/10.1007/s12253-017-0187-5>
41. Ensani F, Omranipour R, Jahanzad I et al (2017) The core needle and surgical biopsy concordance to detect estrogen, progesterone, and Her-2 receptors in breast cancer: a comparative study. *Iran J Pathol* 12:202–208
 42. Kombak FE, Şahin H, Mollamemişoğlu H et al (2017) Concordance of immunohistochemistry between core needle biopsy and surgical resection of breast cancer. *Turk J Med Sci* 47:1791–1796. <https://doi.org/10.3906/sag-1702-152>
 43. Meattini I, Bicchierai G, Saieva C et al (2017) Impact of molecular subtypes classification concordance between preoperative core needle biopsy and surgical specimen on early breast cancer management: single-institution experience and review of published literature. *Eur J Surg Oncol (EJSO)* 43:642–648. <https://doi.org/10.1016/j.ejso.2016.10.025>
 44. You K, Park S, Ryu JM et al (2017) Comparison of core needle biopsy and surgical specimens in determining intrinsic biological subtypes of breast cancer with immunohistochemistry. *J Breast Cancer* 20:297. <https://doi.org/10.4048/jbc.2017.20.3.297>
 45. Berghuis AMS, van Deurzen CHM, Koffijberg H et al (2019) Real-world data on discordance between estrogen, progesterone, and HER2 receptor expression on diagnostic tumor biopsy versus tumor resection material. *Breast Cancer Res Treat* 175:451–458. <https://doi.org/10.1007/s10549-019-05141-y>
 46. Jeong YS, Kang J, Lee J et al (2020) Analysis of the molecular subtypes of preoperative core needle biopsy and surgical specimens in invasive breast cancer. *J Pathol Transl Med* 54:87–94. <https://doi.org/10.4132/jptm.2019.10.14>
 47. Shanmugalingam A, Hitos K, Hegde S et al (2022) Concordance between core needle biopsy and surgical excision for breast cancer tumor grade and biomarkers. *Breast Cancer Res Treat* 193:151–159. <https://doi.org/10.1007/s10549-022-06548-w>
 48. Slostad JA, Yun NK, Schad AE et al (2022) Concordance of breast cancer biomarker testing in core needle biopsy and surgical specimens: a single institution experience. *Cancer Med*. <https://doi.org/10.1002/cam4.4843>
 49. Sun T, Zhang H, Gao W, Yang Q (2021) The appropriate number of preoperative core needle biopsy specimens for analysis in breast cancer. *Medicine* 100:e25400. <https://doi.org/10.1097/MD.00000000000025400>
 50. Rossi C, Fraticelli S, Boveri E et al (2021) PS-02-026 immunohistochemistry alone may represent a cost-effective alternative for HER2 status assessment on biopsy. *Virchows Arch* 479:S72. <https://doi.org/10.1007/s00428-021-03157-8>
 51. Lebeau A, Turzynski A, Braun S et al (2010) Reliability of human epidermal growth factor receptor 2 immunohistochemistry in breast core needle biopsies. *J Clin Oncol* 28:3264–3270. <https://doi.org/10.1200/JCO.2009.25.9366>
 52. Lee AHS, Key HP, Bell JA et al (2012) Concordance of HER2 status assessed on needle core biopsy and surgical specimens of invasive carcinoma of the breast: HER2 in breast cancer core biopsies. *Histopathology* 60:880–884. <https://doi.org/10.1111/j.1365-2559.2011.04144.x>
 53. Reis-Filho JS, Davidson NE (2021) Ki67 assessment in breast cancer: are we there yet? *JNCI* 113:797–798. <https://doi.org/10.1093/jnci/djaa202>
 54. Ahn S, Lee J, Cho M-S et al (2018) Evaluation of Ki-67 index in core needle biopsies and matched breast cancer surgical specimens. *Arch Pathol Lab Med* 142:364–368. <https://doi.org/10.5858/arpa.2017-0014-OA>
 55. Kim HS, Park S, Koo JS et al (2016) Risk factors associated with discordant Ki-67 levels between preoperative biopsy and postoperative surgical specimens in breast cancers. *PLoS One* 11:e0151054. <https://doi.org/10.1371/journal.pone.0151054>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.