Recommendations for 'Adequate Evaluation of Hormone Receptors': A report of the task force of the Japanese Breast Cancer Society

SHINOBU UMEMURA[1], MASAFUMI KUROSUMI[2], TAKUYA MORIYA[3], TETSUNARI OYAMA[4], KOHJI ARIHIRO[5], HIROKO YAMASHITA[6], YOSHIHISA UMEKITA[7], YOSHIFUMI KOMOIKE[8], CHIKAKO SHIMIZU[9], HISAKI FUKUSHIMA[10], HIROSHI KAJIWARA[1] and FUTOSHI AKIYAMA[11]

[1]Department of Pathology, Tokai University School of Medicine; [2]Department of Pathology, Saitama Cancer Center; [3]Department of Pathology, Kawasaki Medical School; [4]Department of Pathology, Gumma University School of Medicine; [5]Department of Anatomical Pathology, Hiroshima University Hospital; [6]Oncology and Immunology, Nagoya City University Graduate School of Medical Sciences; [7]Department of Pathology, Faculty of Medicine, Kagoshima University; [8]Department of Surgery, Osaka Medical Center for Cancer and Cardiovascular Diseases; [9]Department of Medical Oncology, National Cancer Center Hospital; [10]Department of Surgery, Jikei University School of Medicine; [11]Division of Pathology, The Cancer Institute, Japanese Foundation for Cancer Research, Japan

Abstract. The task force of the Japanese Breast Cancer Society was assembled to examine variable factors related to the immunohistochemical evaluation of hormone receptors in breast cancer, and to provide recommendations for adequate handling of specimens and accurate evaluations for hormone receptors. The various factors examined were: i) the adequate handling of breast cancer tissue for IHC, ii) the concordant rate between EIA and IHC assays using 5 different staining methods, iii) the inter-observer diversity for evaluation, and iv) the threshold for the predictive value of endocrine therapy of primary and recurrent/metastatic breast cancers. The conducted studies found that a 10% threshold was stable and reliable in spite of different validations including the concordance between IHC and EIA assays, inter-observer diversity and disease-free survival rates for patients who received tamoxifen for primary breast cancers. It was also found that 1% of threshold was useful in limited situations including the predictive value of endocrine therapy for recurrent/metastatic breast cancers. Based on these results for technical, pathological and clinical validation studies, a recommendation was proposed. Herein, we summarize the evidence, on which the recommendations were made, and customize the recommendations suitable for the current status.

Contents

1. Background
2. Status of hormone receptor detection in 2003
3. Research plans
4. Summary of results
5. Issues to be considered
6. Changing concepts in hormone receptor evaluation
7. Current status and future directions for 2010

1. Background

It is important that the evaluation of hormone receptors by immunohistochemical analysis be correctly performed so that breast cancer patients and physicians are provided with accurate information. In a research study conducted by the Japanese Breast Cancer Society (JBCS) titled, 'Adequate Evaluation of Hormone Receptors', recommendations for sample handling and a practically useful evaluation system were discussed in 2006 (1). It is now necessary to summarize the evidence, on which the recommendations were made, and to customize them to the current status.

2. Status of hormone receptor detection in 2003

In 2003, a questionnaire was distributed to determine the general status of immunohistochemical examinations performed for the detection of hormone receptors. In 2003, reagent supplies for the enzyme immunoassay (EIA) detection method were stopped, and immunohistochemistry (IHC)
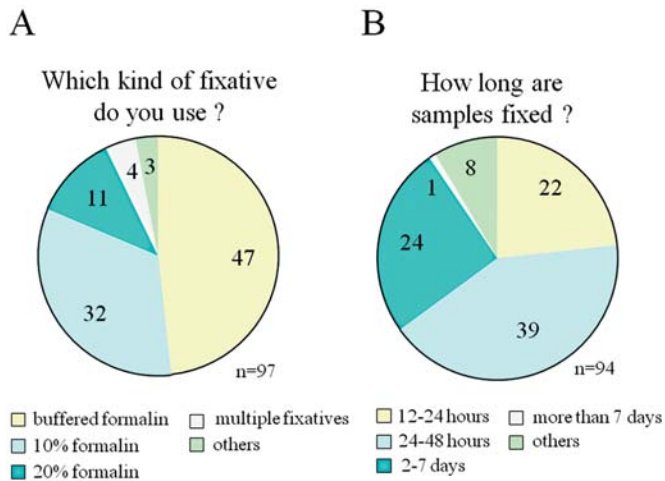
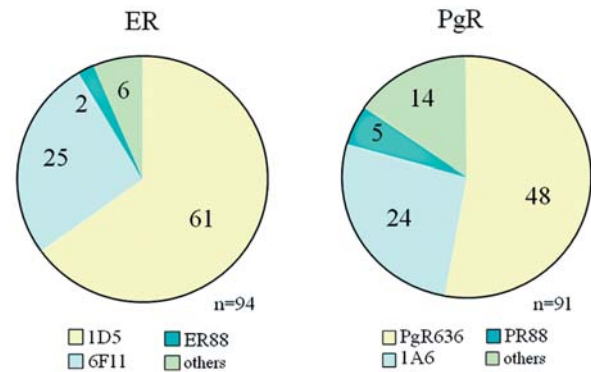Figure 1. Questionnaire results regarding the fixative used for immuno-histochemical examination. The type of fixative (A) and duration of fixation (B) used.

Figure 2. Questionnaire results regarding primary antibodies used for detection of estrogen and progesterone receptors.

assays substituted EIA for the detection of hormone receptors. The questionnaire was sent to 335 counselors of JBCS, and while not all of the counselors replied, 162 responses were analyzed.

Before July 2003, 50/158 (31.6%) institutions were using IHC as the sole method for detecting hormone receptor expression (HR-IHC), vs. 32/158 (20.3%) institutions that were using a combination of IHC and EIA. In contrast, after July 2003, almost all of the institutions had converted to only using IHC. Although primary tumors were the majority of samples examined, physicians in 78.6% of institutions (125/159) considered recurrent or metastatic tumor may be examined, if these were available. Prior to July 2003, 102/155 (65.8%) institutions performed HR-IHC using in-house laboratories. These laboratories used either a manual method (42/102, 41.2%), an automated machine (46/102, 45.1%) or both methods (13/102, 12.7%). In contrast, 25/108 (23.1%) institutions commissioned examinations to be performed by contracted laboratories. After July 2003, the use of contracted laboratories by 25 institutions increased to 53 institutions (34.2%).

For samples that were examined by in-house laboratories, the conditions for fixation were found to vary between institutions. For example, fixatives included buffered formalin (47/97, 48.5%), 10% formalin (32/97, 33.0%) and 20% formalin (11/97, 11.3%) (Fig. 1A). The time from tumor resection to fixation was also found to vary, with fixation times being within 1 h at 43.3% (43/90) of institutions, 2 h at 28.9% (26/90) of institutions and >2 h at 27.8% (25/90) of institutions. Once the samples were placed in fixative, the duration of fixation varied from within 24 h at 23.4% (22/94) of institutions, 24-48 h at 41.5% (39/94) of institution or >48 h at 35.1% (33/94) of institutions (Fig. 1B).

The primary antibodies used for detection of hormone receptors included antibodies that were approved as diagnostic reagents, except for a few institutions which used alternative antibodies. For example, 6/94 (6.4%) institutions used alternative antibodies for detection of ER, while 14/91 (15.4%) institutions used alternative antibodies for detection of PgR

(Fig. 2). In addition, 26/88 (29.5%) institutions performed modified protocols. For example, in protocols using clone 1D5 for detection, the antigen retrieval methods listed in the automated vs. manual methods included variations in the retrieval solution used (76.9 and 40.1%), in the heating method used (69.2 and 29.4%) and in the duration of heating applied (76.9 and 52.9%), respectively.

For the evaluation of samples analyzed, differences in the thresholds for positive vs. negative (25/89, 28.1%), the population of positive cells detected (29/89, 32.6%) and the scoring system based on population and staining intensity (24/89, 27.0%) were found to vary among individual institutions. For example, thresholds for the positive cell population were found to be set at 10% for 47/79 institutions (59.5%), at 5% for 17/79 institutions (21.5%) and at 1% for 8/79 institutions (10.1%).

For controls, a majority of the institutions surveyed (79/89, 88.8%) examined external or internal controls simultaneously. Similarly, two-thirds of the institutions certified their immunohistochemical techniques by a comparison with EIA methods or external assessments. Unfortunately, more than half of the institutions observed discordant results between the IHC and EIA assays performed. Causes for this discrepancy were speculated to include inadequate sampling (32/48, 66.7%) and technical problems (2/48, 2.1%) for the EIA assays. However, fixation (6/48, 12.5%), threshold (6/48, 12.5%) and technical problems (1/48, 2.1%) for IHC assays were also speculated to be sources of discrepancy.

3. Research plans

The aim of the present study was to examine how to minimize the diversity among laboratories in the detection of hormone receptor status, and how to provide consistent, reproducible, and reliable results in routine practice. For this purpose, various factors were examined: i) the adequate handling of breast cancer tissue for IHC, ii) the concordant rate between EIA and IHC assays using 5 different staining methods, iii) the inter-observer diversity for evaluation and iv) the threshold

Table I. Summary of best thresholds for estrogen receptor immunohistochemistry in studies examined.

Scoring system[a]	A (Allred score)[a]		B (J-score) and C[b]	
Studies for staining procedures[a,b]				
Best thresholds and concordance between EIA and IHC (%)				
Procedure (I)	TS3	(88.6)	any positive cell or 1% (87.5)	
Procedure (II)	TS4	(92.0)	10%	(93.2)
Procedure (III)	TS5	(90.9)	10%	(94.3)
Procedure (IV)	TS2 or 3	(89.8)	1%	(89.8)
Procedure (V)	TS5	(94.3)	10%	(94.3)
Studies for evaluation system				
Pathological validation[b]				
Best thresholds and concordance between EIA and IHC (%)				
Scoring system A (coefficient r=0.524)	TS3	(89.9)		
Scoring system B (coefficient r=0.383)			10%	(92.1)
Scoring system C (coefficient r=0456)			10%	(92.1)
Best thresholds and inter-observer concordance (%)				
Scoring system A	TS3	(93.3)		
Scoring system B			1%	(96.6)
Scoring system C			10%	(89.9)
Clinical validation				
Best thresholds and p-value for endocrine therapy for primary breast cancer				
Scoring system A	TS5	(p=0.00049)		
Scoring system B			10%	(p=0.04)
Best thresholds and p-value for endocrine therapy for recurrent/metastatic breast cancer				
Endocrine therapy	TS4	(p=0.020)	10%	(p=0.011)
Tamoxifen	TS4	(p=0.047)	10%	(p=0.03)
Post-relapse survival	TS3	(p=0.0005)	1%	(p=0.0005)

[a]See research plan; [b]highest concordant rates between EIA and IHC and inter-observer concordance are shown in this table.

for the predictive value of endocrine therapy of primary and recurrent/metastatic breast cancers. Details of each study were individually published as original articles (2-6).

Adequate tissue handling was examined with a focus on the fixation (2). Surgical specimens were cut to a size of 5x5x3 mm and fixed in 10% formalin for various lengths of time including 30 min, 3, 6, 24 and 48 h, 4, 7 days and 3 weeks. The condition before fixation was also examined with samples left in distilled water (DW) for 24 h.

Staining methods and evaluation systems were examined using the same series of 89 cases obtained from the Saitama Cancer Center (3). Antibodies used in these studies included clone 1D5 (DakoCytomation), clone 6F11 (Ventana Japan), clone ER88 (Kyowa Medix) for ER and PgR636 (DakoCytomation), 16 (Ventana Japan) (which is the next generation of clone 1A6) and PR88 (Kyowa medix) for PgR. All of these antibodies, except clone 16, are approved in Japan. For clones 1D5, ER88, PgR636 and PR88, staining methods are available for both automated and manual protocols. However, for clones 6F11, 1A6 and 16, only staining

methods compatible with an automated machine are approved. Therefore, the 5 staining procedures that were evaluated for the detection of ER were: clone ER88 (I) and 1D5 (II) using manual protocols, and 6F11 (III), PR88 (IV), and 1D5 (V) using automated protocols. The five staining procedures were also evaluated for their detection of PgR as follows: clone PR88 (I) and PgR636 (II) using manual protocols, and 16 (III), PR88 (IV), and PgR636 (V) using automated protocols. For each of these protocols, formalin-fixed, paraffin-embedded sections were submitted along with enzyme immunoassay results and IHC was performed strictly according to the manu-facturer's protocols for each antibody. The scoring systems examined included the Allred score (A) (8), the (B) system which includes scores of 0, 1 (<1%), 2 (1% ≤10%) and 3 (≥10%) or the (C) system which includes scores of 0, 1 (<10%), 2 (10% ≤66%) and 3 (≥66%).

To examine the predictive value and threshold for endocrine therapy, 486 primary breast cancer patients who received tamoxifen as an adjuvant hormone therapy at the Cancer Institute Ariake Hospital were analyzed (4,5). Seventy-

five patients with recurrent/metastatic breast cancer who received endocrine therapy and were evaluated at the Nagoya City University Hospital were also analyzed (6).

4. Summary of results

Tissue handling for IHC. Immunohistochemical signals decreased when samples were left in DW for 24 h. An extended delay between specimen collection and fixation should be avoided. Immunohistochemical signals were markedly decreased in specimens under the unsuitable conditions, over-fixation for 7 days and under-fixation for 30 min (2).

Concordant rate between EIA and IHC assays using five different staining methods. Five different IHC staining procedures (I-V), were compared with EIA assays using 3 different evaluation systems (A-C). The best concordant rates are 93.2% (II, clone 1D5), 94.3% (III, clone 6F11) and 94.3% (V, clone 1D5) at a 10% threshold (3). The sensitivities of procedures (I) and (IV) were found to be lower than for the other procedures, and the best concordance rate was set at a threshold of 1%. When evaluation system (A) was used, thresholds for the best concordance rates for each staining procedure varied widely from TS2 or 3 for procedure (IV), to TS5 for procedures (III) and (V).

Inter-observer diversity. Pathological studies for each evaluation system were performed using sections stained by procedure (II). The best correlation coefficient between IHC and EIA assays was associated with evaluation system (A) (r=0.524), followed by (C) (r=0.456) and (B) (r=0.383). The highest concordance rates between EIA and IHC assays were achieved when the threshold was set at 10% (92.1%) and TS3 (89.9%) (Table I). Inter-observer diversity was found to be the greatest when scoring system (A) was used, and was the smallest when scoring system (B) was used. Specifically, concordance rates among observers was 96.6% at a threshold of 1%, 89.9% at a threshold of 10% and 93.3% with a threshold of TS3.

Threshold for the predictive value of endocrine therapy of primary breast cancers. For 486 primary breast cancer specimens analyzed, IHC was found to more accurately predict the therapeutic effects than EIA (p=0.04 vs. 0.62, respectively) (4). There is no significant difference when evaluation was limited for invasive area (p=0.004) or including intraductal component (p=0.004). Disease-free survival (DFS) of patients with breast cancers expressing ER ≥10% (0.04) or ≥TS5 (p=0.00049) was significantly better than those expressing ER <10% or TS5 (4). When the threshold was set at 1% for evaluation systems (B) and (C), and at TS4 for (A), the difference between DFS rates was not significant.

Threshold for the predictive value of endocrine therapy of recurrent/metastatic breast cancers. Responsiveness for the patient with recurrent or metastatic breast cancers was most significantly correlated with ER expression with a threshold of 10% (p=0.011 and 0.03) and TS4 (p=0.02 and 0.047) for any endocrine therapy or tamoxifen only, respectively (6).

Post-relapse survival was most significantly correlated with ER expression with a threshold of 1% (p=0.0005) and TS3 (p=0.0005).

All of these results are summarized in Table I and demonstrate that a threshold of 10% provided consistent pathological and clinical validations. Considering the therapeutic effect of hormone therapy for patients with metastatic or recurrent cancers, a threshold of 1% should not be neglected. Based on these results, a recommended working protocol for IHC evaluation of hormone receptor expression is shown in Table II (1,7).

5. Issues to be considered

The conducted studies found that a 10% threshold was stable and reliable in spite of different validations including the concordance between IHC and EIA assays, inter-observer diversity, and disease-free survival rates for patients who received tamoxifen for primary breast cancers. It was also found that 1% of threshold was useful in limited situations including for staining procedures using ER88, for inter-observer diversity and for the predictive value of endocrine therapy for recurrent/metastatic breast cancers.

The Allred scoring system is a well-known and clinically validated scoring system (8,9). The present study showed that there are several issues to be considered to recommend the Allred scoring system utilized in global multi-institutions. First issue is the differences of primary antibody and staining protocols. The original staining procedure published by Allred *et al* was clone 6F11 for detection of ER, and clone 1294 for detection of PgR. The pretreatment and staining protocol for clone 6F11 recommended in Japan differed from the Allred protocol and the 5 staining procedures currently approved in Japan showed different degrees of sensitivity that affected the distribution of the total score. It is necessary that the staining procedures as sensitive as the original procedure and globally standardized.

Another consideration is inter-observer diversity. A subjective bias is typically introduced into the estimation process, especially when we estimate the staining intensity. Heterogeneity in the distribution of positive cells and their staining intensity can be significant sources of observation diversity. Adequate training for estimation process is a possible resolution. Observation diversity was also detected in histological grading systems. The pathological board of Japan National Surgical Adjuvant Study of Breast Cancer (NSAS-BC) provides opportunities for individuals to practice grading nuclear atypia at their meetings. It was found that inter-observer diversity was successfully minimized when repeated trainings were completed (10,11). Similar training programs for evaluation of histology staining would also be expected to minimize inter-observer diversities.

The third issue was that threshold settings was the inconsistency between the studies conducted. In this study, the threshold for primary breast cancer patients who received tamoxifen was TS5, although Harvey *et al* reported a threshold of TS3 (>TS2) in a study of 777 patients who received endocrine therapy (which mostly included tamoxifen) (9). There is no definitive reason for this difference, although differences of prognosis of the primary breast cancers

Table II. Recommendation for working protocol.

___

Handling of sample
    Slice the samples in adequate size for cassette container
    Immerse in an enough volume of fixative (10% formalin, 10% buffered formalin)
    Fix the samples within 48 h

Staining procedure
    Keep the staining protocols provided from antibody distributing companies

Evaluation process
    Observe HE sections to distinguish carcinoma and the non-neoplastic mammary gland
    Check internal, positive and negative controls to estimate the staining conditions
    Observe immunostained sections throughout at lower magnification to check technical problems
    Evaluate whole of the lesion. If there is a remarkable discrepant staining finding between invasive and ductal component, it should be described in the report

Evaluation system (J-Score)
    Score

Score 0	No positive cells
Score 1	Population of positive cells, <1%
Score 2	Population of positive cells, ≥1% <10%
Score 3	Population of positive cells, ≥10%

    Division

Negative	Score 0
Borderline	Score 1, 2
Positive	Score 3

___

Table III. Modified J-Score proposed.

___

Score	
Score 0	No positive cells
Score 1	Population of positive cells, <1%
Score 2	Population of positive cells, ≥1% <10%
Score 3	Population of positive cells, ≥10%
3a	≥10% <50%
3b	≥50%
Division	
Negative	Score 0
Borderline	Score 1, 2
Positive	Score 3

___

between USA and Japan, differences of staining procedures are considerable.

6. Changing concepts in hormone receptor evaluation

The circumferential situation for evaluation system has developed domestically and internationally since 2005. At the 9th consensus meeting at St. Gallen for early breast cancers in 2005, 3 categories for endocrine responsiveness defined were: 'endocrine responsive', 'endocrine non-responsive' and 'endocrine response uncertain' (12). Respon-

siveness to endocrine therapy can be influenced by factors other than expression of hormone receptors, however, the expression level of hormone receptors is currently considered the most predictive factor for endocrine therapy responsiveness. The 'endocrine response uncertain' category can refer to patients with breast cancers that exhibit low levels (usually <10% of positive cells) of steroid hormone receptor immunoreactivity and an absence of PgR expression.

The three categories of endocrine responsiveness have remained fundamentally unchanged through the 10th annual St. Gallen meeting in 2007 (13), although the characterization of these categories was improved. For example, 'highly endocrine responsive' tumors were shown to express high levels of both ER and PgR in a majority of cells, while 'incompletely endocrine responsive' tumors were found to express lower levels or an absence, of ER or PgR. In general, breast cancers with higher expression levels of hormone receptors are predicted to be more responsive to endocrine therapy.

At the 11th annual St. Gallen consensus meeting (14), a major alteration was made in that the concept or 'category' was changed to 'threshold'. The panel recommended that adjuvant endocrine therapy was recommended for all patients with tumors exhibiting any degree of positive ER expression, and a majority of the panelists supported the use of percentages rather than categories in pathological reports. Additionally, a consideration for whether hormone therapy alone, or in combination with chemotherapy, was based on clinicopathological features and/or multigene assays for patients with ER

positive/HER-2 negative cancers. High level expression of ER and PgR was described as >50%.

7. Current status and future directions for 2010

Expression of hormone receptors is a consecutive variable, therefore, the reporting system that describes the percentage of positive cells for hormone receptors is reasonable. Yaziji *et al* (15) recommended a reporting system for hormone receptor describing percentage of positive cells of each category of staining intensity, e.g., no, weak, moderate and strong signal. Automated staining and counting systems are one of the ways by which clinical requirements can be fulfilled routinely. Currently, however, not all institutions are able to perform fully automated IHC assays and evaluations. Therefore, from the practical aspects, the proposed scoring system is useful with minimal diversities and no major conflict with the 9th, 10th and 11th editions of the St. Gallen recommendations. Small modifications regarding the concept of 'highly endocrine-responsive' samples would improve the usefulness of the current routine practice (Table III).

References

1. Umemura S, Kurosumi M, Moriya T, Oyama T, Arihiro K, Yamashita H, Umekira Y, Komoike Y, Shimizu C, Fukushima H, Kajiwara H and Akiyama F: Immunohistochemical evaluation for hormone receptors in breast cancer: a practically useful evaluation system and handling protocol. Breast Cancer 13: 232-235, 2006.
2. Oyama T, Ishikawa Y, Hayashi M, Arihiro K and Horiguchi J: The effects of fixation, processing and evaluation criteria on immunohistochemical detection of hormone receptors in breast cancer. Breast Cancer 14: 182-188, 2007.
3. Arihiro K, Umemura S, Kurosumi M, Moriya T, Oyama T, Yamashita H, Umekita Y, Komoike Y, Shimizu C, Fukushima H, Kajiwara H and Akiyama F: Comparison of evaluations for hormone receptors in breast carcinoma using two manual and three automated immunohistochemical stainings. Am J Clin Pathol 127: 356-365, 2007.
4. Horii R, Akiyama F, Ito Y and Iwase T: Assessment of hormone receptor status in breast cancer. Pathol Int 57: 784-790, 2007.
5. Horii R, Akiyama F, Itoh Y and Iwase T: Optimal indications of endocrine therapy alone as adjuvant systemic treatment of breast cancer. Br J Cancer 97: 654-658, 2007.
6. Yamashita H, Ando Y, Nishio M, Zhang Z, Hamaguchi M, Mita K, Kobayashi S, Fujii Y and Iwase H: Immunohistochemical evalutation of hormone receptor status for predicting response to endocrine therapy in metastatic breast cancer. Breast Cancer 13: 74-83, 2006.
7. Kurosumi M: Immunohistochemical assessment of hormone receptor status using a new scoring system (J-Score) in breast cancer. Breast Cancer 24: 189-193, 2007.
8. Allred DC, Harvey JM, Berardo M and Clark GM: Prognostic and predictive factors in breast cancer by immunohistochemical analysis. Mod Pathol 11: 155-168, 1998.
9. Harvey JM, Clark GM, Osborne CK and Allred DC: Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. J Clin Oncol 17: 1474-1481, 1999.
10. Tsuda H, Akiyama F, Kurosumi M, Sakamoto G and Watanabe T: Monitoring of interobserver agreement in nuclear atypia scoring of node-negative breast carcinomas judged at individual collaborating hospitals in the National Surgical Adjuvant Study of Breast Cancer (NSAS-BC) protocol. Jpn J Clin Oncol 29: 413-420, 1999.
11. Tsuda H, Akiyama F, Kurosumi M, Sakamoto G and Watanabe T: The efficacy and limitations of repeated slide conferences for improving interobserver agreement when judging nuclear atypia of breast cancer. The Japan National Surgical Adjuvant Study of Breast Cancer (NSAS-BC) Pathology Section. Jpn J Clin Oncol 29: 68-73, 1999.
12. Goldhirsch A, Glick JH, Gelber RD, Coates AS, Thurlimann B and Senn HJ: Meeting highlights: international expert consensus on the primary therapy of early breast cancer. Ann Oncol 16: 1569-1583, 2005.
13. Goldhirsch A, Wood WC, Gelber RD, Coates AS, Thürlimann B and Senn HJ: 10th St. Gallen conference: progress and promise: highlights of the international expert consensus on the primary therapy of early breast cancer 2007. Ann Oncol 18: 1133-1144, 2007.
14. Goldhirsch A, Ingle JN, Gelber RD, Coates AS, Thurlimann B, Senn HJ and Panel Members: Thresholds for therapies: highlights of the St Gallen international expert consensus on the primary therapy of early breast cancer 2009. Ann Oncol 20: 1319-1329, 2009.
15. Yaziji H, Taylor CR, Phil D, Goldstein NS, Dabbs DJ, Hammond EH, Helett B, Floyd AD, Barry TS, Martin AW, Badve S, Baehner F, Cartun RW, Eisen RN, Swanson PE, Hewitt SM, Vyberg M, Hicks DG and Members of the Standardization Ad-Hoc Consensus Committee: Consensus recommendation on estrogen receptor testing in breast cancer by immunohistochemistry. Appl Immunohistochem Mol Morphol 16: 513-520, 2008.