Human Pathology (2019) **85**, 27–32

ELSEVIER

Human PATHOLOGY

www.elsevier.com/locate/humpath

Original contribution

Hotspot enumeration of CD8+ tumor-infiltrating lymphocytes using digital image analysis in triple-negative breast cancer yields consistent results☆,☆☆

Patrick J. McIntire MD[a,*], Elaine Zhong MD[a], Ami Patel MD[a], Francesca Khani MD[a], Timothy M. D'Alfonso MD[c], Zhengming Chen PhD, MPHMS[b], Sandra J. Shin MD[d], Paula S. Ginter MD[a]

[a]*Weill Cornell Medicine, Department of Pathology and Laboratory Medicine, New York, NY 10065, USA*
[b]*Weill Cornell Medicine, Department of Public Healthcare Policy and Research, Division of Biostatistics and Epidemiology, New York, NY 10065, USA*
[c]*Memorial Sloan Kettering Cancer Center, Department of Pathology, New York, NY 10065, USA*
[d]*Albany Medical College, Department of Pathology and Laboratory Medicine, Albany, NY 12208, USA*

Received 25 May 2018; revised 16 October 2018; accepted 18 October 2018

Keywords:
TILs;
TNBC;
CD8;
DIA;
Hotspots

Summary Tumor-infiltrating lymphocytes (TILs) have emerged as prognostic in triple-negative breast cancer (TNBC). We aimed to assess the consistency of hotspot placement and TIL enumeration among multiple pathologists. Additionally, we assessed hotspot TIL count consistency by comparing hotspot counts in 3 separate locations within a single whole-tissue section. Anti-CD8 immunohistochemistry was performed on a representative section from 66 cases of primary TNBC, which were then scanned as whole-slide images. Quantification of the tissue area and combined stromal and intratumoral CD8+ TILs was performed using digital image analysis (DIA) within 2.2 mm–diameter circle hotspots. TIL counts were quantified as absolute counts and densities (absolute count/tissue area in micrometers^2). For each case, 6 pathologists placed a single *hotspot*, defined as an area with the subjectively highest CD8+ immunoreactivity, within the tumor bed. Separately for each case, a single pathologist placed hotspots in 3 different locations within a single tumor section. Intraclass correlation coefficients (ICCs) were generated following TIL enumeration via DIA. ICCs for single hotspot placement by 6 pathologists were 0.96 for density and 0.97 for absolute counts, respectively. In 32% of cases (21/66), all the hotspots placed by the 6 pathologists were in the same location. When evaluating hotspots in 3 different locations within a tumor, the ICC was 0.95 for both density and absolute counts. Hotspot evaluation by DIA is a reproducible method for CD8+ TIL quantification, and the use of hotspots may reduce TIL count variation caused by intratumoral TIL heterogeneity.

☆ Disclosures: The authors have no financial disclosures or conflicts of interest.
☆☆ The findings were presented (in abstract form) at the 2018 USCAP Annual Meeting in Vancouver, British Columbia.
* Corresponding author at: Department of Pathology and Laboratory Medicine, New York-Presbyterian Hospital, Weill Cornell Medicine, 525 E 68th St, Starr 1031E, New York, NY 10065, USA.
E-mail address: pmcinti2@gmail.com (P. J. McIntire).

https://doi.org/10.1016/j.humpath.2018.10.014
0046-8177/

1. Introduction

Advancements in immunotherapies for cancer have renewed interest in the role of inflammatory infiltrates in breast cancer and other tumor types. Studies have shown that tumor-infiltrating lymphocytes (TILs) are prognostic, predictive, and prevalent in triple-negative breast cancer (TNBC), a breast cancer subtype associated with poor clinical outcomes [1-8].

An International TIL Working Group (ITWG) has established guidelines for enumerating TILs on hematoxylin and eosin (H&E) sections [9]. Subsequent studies have shown the guidelines are adequately prognostic [10] and that interobserver concordance is moderate [11,12]. According to the guidelines, TIL quantification entails the evaluation of the whole tumor area on a single section. However, TILs are heterogeneously distributed throughout the tumor and are variably concentrated at the periphery [13,14]. This intratumoral heterogeneity has been reported to be responsible for outliers in studies evaluating interobserver concordance using the ITWG guidelines [11]. To improve interobserver concordance of TIL quantification, Denkert et al [11] asked pathologists to evaluate at least 3 areas of a predefined size (1mm^2) per tumor, and the mean of the recorded values for each region was used for statistical evaluation. Denkert et al demonstrated that evaluating smaller tissue areas can effectively reduce intratumoral variance as a source of outliers and increase concordance.

Recently, hotspot CD8+ TIL counts have been shown to be associated with a survival benefit in TNBC [15]. CD8+ cytotoxic T cells are consistently the most prognostic and predictive of the TIL subtypes. Other TIL subtypes, including CD4+ regulatory T cells, CD20+ B cells, and plasma cells, show more conflicting data [16-22]. Furthermore, advances in image recognition software have made digital image analysis (DIA) an instrument to reduce interobserver variance [23-25].

The aim of this study was to assess the consistency of hotspot placement and TIL enumeration among multiple pathologists. Additionally, we assessed the consistency of hotspot TIL enumeration within a tumor by comparing counts from 3 hotspots in 3 different locations within a single whole-tissue section of TNBC.

2. Materials and methods

2.1. Tumor samples

Patients with primary TNBCs were identified at New York-Presbyterian Hospital/Weill Cornell Medicine from the years 2000-2014. Tumors measuring <0.5 cm in greatest dimension were excluded from this study. Institutional review board approval was obtained for all parts of this study.

2.2. Immunohistochemistry methods and interpretation

Immunohistochemistry was performed on 4 μm–thick sections prepared from representative tumor blocks. Slides were deparaffinized in sequential baths of xylene, transferred to sequential baths of 100% ethanol followed by sequential baths of 95% ethanol, and then rinsed in deionized water prior to staining. Immunohistochemistry was performed using mouse anti-human monoclonal anti-CD8 antibody (4B11, Dilution: RTU, Leica Biosystems, cat. # PA0183; Buffalo Grove, IL). Following antigen retrieval, the slides were stained using the Bond Polymer Refine Detection system (Leica Biosystems, Buffalo Grove, IL), dehydrated in 100% ethanol, and mounted in Cytoseal XYL (Richard-Allan Scientific, Kalamazoo, MI) by a Leica CV5030 cover slipper (Leica Biosystems, Buffalo Grove, IL).

2.3. Whole slide scanning

Slides were scanned at a 20× magnification using a single *z*-plane via an Aperio AT2 whole slide scanner (Leica Biosystems, San Diego, CA). Scanned images were evaluated for quality and to ensure that they were in focus. Digital files in ".svs" format were stored on an image server for remote evaluation using eSlide Manager Healthcare Network application (Leica Biosystems, Buffalo Grove, IL). The scanned images were loaded onto the HALO DIA platform (Indica Labs, Corrales, NM).

2.4. Hotspot placement training

Six pathologists, ranging in clinical experience from 1 to 10 years (P. J. M., A. P. P., E. W. Z., P. G., F. K., T. D.), participated in an educational session detailing the mechanics of the HALO DIA software. Inclusion and exclusion criteria were reviewed at this time. Subsequently, each pathologist reviewed 20 practice cases, which were not included in the study, with a knowledgeable DIA user (P. J. M.).

2.5. Hotspot placement and TIL and tissue area enumeration via DIA

The whole slide images were evaluated at scanning magnification for each case. A *hotspot* was defined as an area within the tumor with the highest subjective number of immunoreactive CD8+ TILs. The hotspot size chosen for evaluation was a 2.2 mm–diameter circle (equivalent to a 10× objective field of view), as previously described [15]. The pathologists manually excluded areas of in situ carcinoma, prior biopsy site, necrosis, and regressed hyalinized stroma from DIA. Any areas devoid of tissue (ie, prior Tissue Microarray site or beyond the tissue edge) were automatically excluded by the DIA software. The DIA settings for TIL enumeration, including size and staining thresholds, were created in

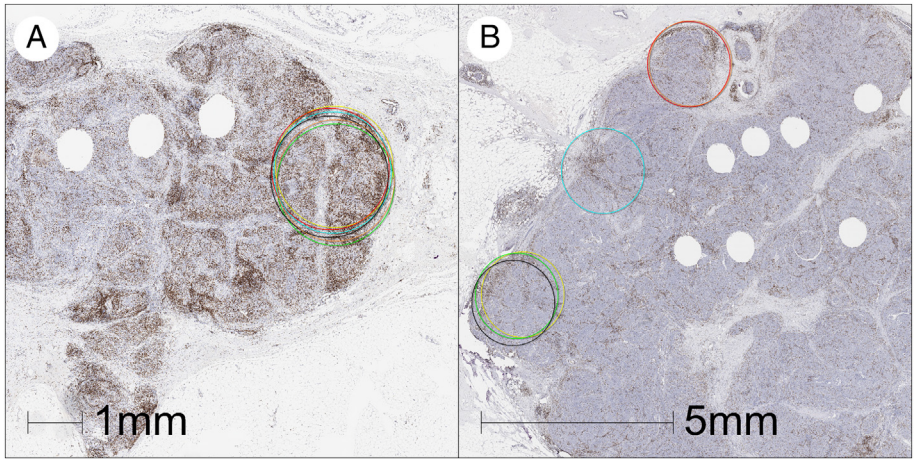

Fig. 1 Examples of tumor annotation by 6 pathologists in HALO on anti-CD8–stained tissue sections. A, Low-power view of whole-tissue section with 2.2 mm–diameter circle (ie, 10× objective field of view) hotspot annotations from 6 pathologists in the same location (overlapping hotspots). B, An example with diverse hotspot selection between the 6 pathologists in which concordance was maintained.

conjunction with software engineers from HALO, as previously described [15]. Cells staining with an intensity exceeding the threshold was deemed as positive and counted. The settings were set to include the full range of staining intensities (weak to strong). DIA enumerated the CD8+ cells and the area of the tissue analyzed.

2.6. Selection of 1 hotspot by 6 different pathologists within a whole-tissue section of TNBC

For the interobserver consistency portion of the study, each of the 6 pathologists independently reviewed each case in a blinded manner. For each case, the pathologists placed a single hotspot within the invasive tumor borders on a whole-tissue section (Fig. 1). After all the pathologists had placed their hotspots, placement of the 6 hotspot annotations was evaluated in each case to determine the proximity of placed hotspots among the 6 pathologists. When all 6 hotspots overlapped, the hotspots were deemed to be in the same location.

2.7. Selection of 3 different hotspots within a whole-tissue section of TNBC

For the intratumoral consistency portion of the study, 1 pathologist (P. J. M.) placed 3 separate hotspots in 3 different locations within the invasive tumor borders on a single whole-tissue section for each case (Fig. 2). The pathologist could place the hotspot anywhere within the tumor provided none of the 3 hotspots overlapped.

2.8. Statistical analysis

After TIL and tissue area enumeration by DIA, statistical analyses were performed using absolute CD8+ TIL counts

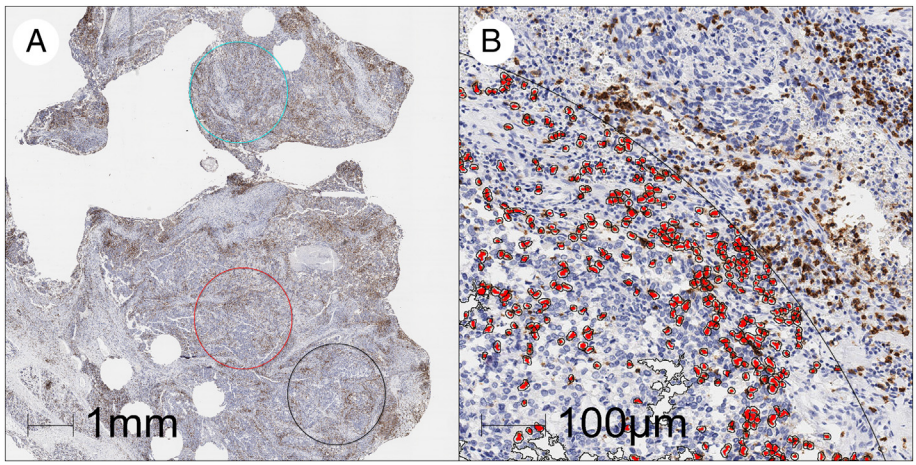

Fig. 2 Multiple equivalent hotspots within a single tumor section. A, Low-power view of an anti-CD8–stained whole-tissue section with 3 separate hotspot annotations. B, High-power view of HALO TIL enumeration within an annotated hotspot. CD8+ cells enumerated by HALO are red within the area of annotation, and CD8+ cells outside the annotation border are brown.

and densities (absolute count/tissue area in micrometers^2). Using 1-way analysis of variance, intraclass correlation coefficient (ICC) was calculated from the chromogenic immunohistochemistry data to evaluate the consistency between pathologists [26]. ICCs were also separately calculated to evaluate the consistency of different hotspot (3 per case) TIL enumeration within a whole-tissue section. All analyses were performed using statistical software SAS Version 9.4 (SAS Institute, Cary, NC).

3. Results

3.1. Interobserver consistency in TIL enumeration

Six pathologists each placed a single hotspot within a whole-tissue section of TNBC in a blinded manner. CD8+ TILs were enumerated from each hotspot. The consistency of CD8+ TIL counts by hotspot placement among 6 pathologists was determined. The ICCs were 0.97 (95% confidence interval [CI]: 0.96-0.98) for absolute CD8+ TIL counts and 0.96 (95% CI: 0.95-0.97) for CD8+ TIL density (Fig. 3). Retrospectively, the hotspots placed by the 6 pathologists were evaluated in each case to assess the number of cases in which all 6 pathologists placed the hotspot in the same location. For 32% of cases (21/66), all 6 pathologists selected hotspots in the same location (overlapping hotspots).

3.2. Intratumoral consistency in TIL enumeration

Three hotspots per case were enumerated and subsequently evaluated for hotspot TIL count consistency within a single whole tumor section. The ICCs generated from TIL enumeration of hotspots placed in 3 different locations within a single whole tumor section by a single pathologist were 0.95 (95% CI: 0.93-0.96) for both CD8+ absolute and density counts.

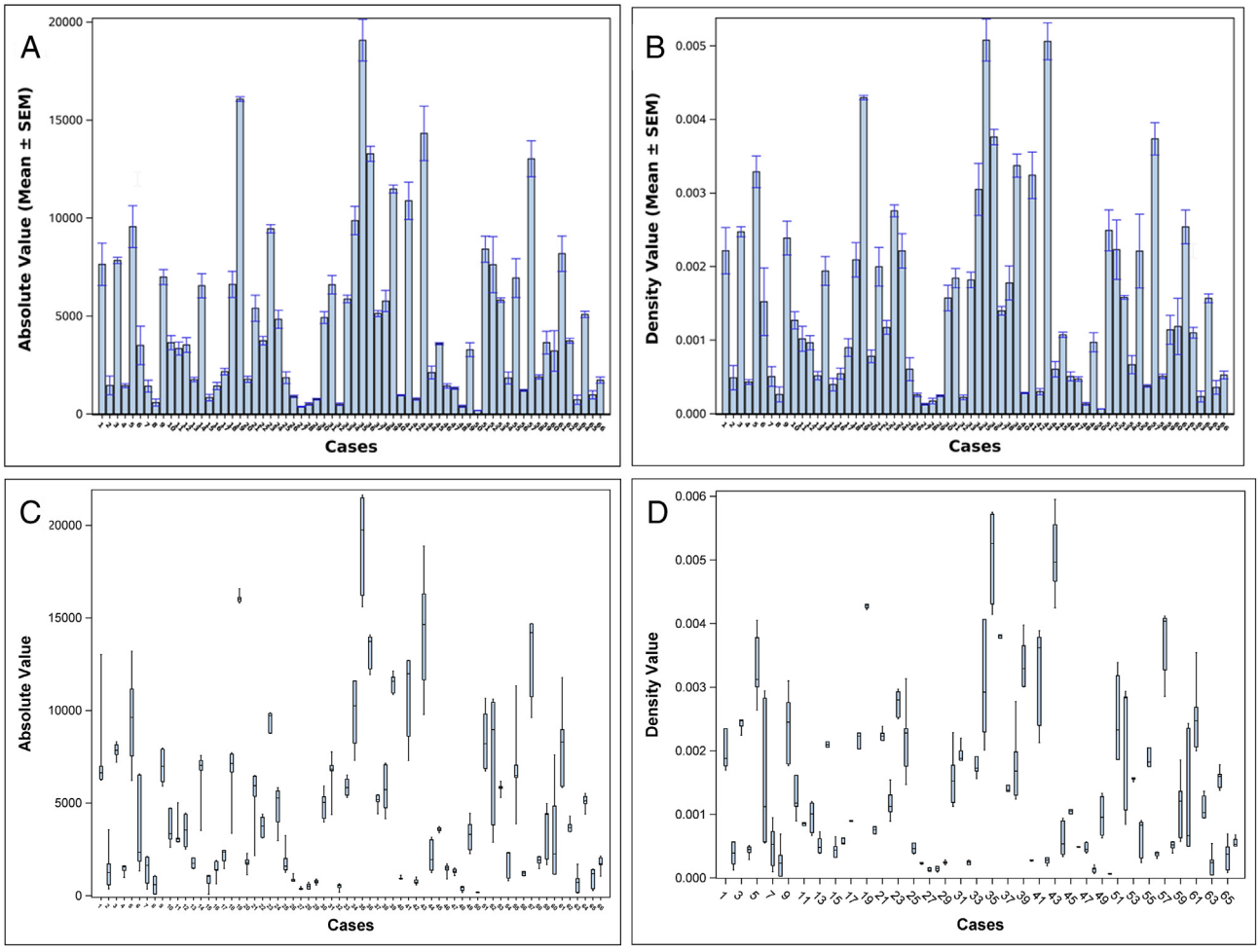

Fig. 3 Distribution of average absolute CD8+ TIL count among 6 pathologists. A, Bar graph of the average absolute CD8+ TIL count. B, Bar graph of the CD8+ TIL density count. C, The distribution of absolute CD8+ TIL counts provided by each of the 6 pathologists' hotspot placement (6 data points per case). Each boxplot represents 25th%, median, and 75th% scores with the whiskers denoting the minimum and maximum scores of the 6 data points. D, The distribution of CD8+ TIL density.

4. Discussion

The immune response to cancer is a new and evolving field of translational research that has gained momentum in the era of effective immunotherapies. Immune infiltrates have been shown to influence response to therapy and prognosis, particularly in triple-negative and HER2-positive breast cancers [1,5-7]. The aim of this study was to assess the consistency of hotspot placement and TIL enumeration among multiple pathologists. Additionally, we assessed the consistency of hotspot TIL counts within a tumor by comparing counts in 3 hotspots in 3 different locations within a whole-tissue section of TNBC.

Digital image analysis has been used as a tool to assess biomarkers in breast cancer and has been shown to outperform manual biomarker assessment [23-25,27]. In a recent study evaluating manual H&E-stained sections for stromal TIL density, the authors observed a systematic shift of mean TIL values between the pathologists, suggesting that pathologists have different individual cut points for a given TIL percentage [11]. Similar results have been described when evaluating Ki67 in breast cancer [28,29]. Digital image analysis evades such sources of error because it has the capability of setting a single quantification threshold for immunoreactive cells, resulting in consistent counting and variance reduction.

Although the ITWG guidelines suggest evaluation of stromal TIL density within the entire tumor bed on H&E-stained sections, in practice, it may prove difficult to perform. In fact, the studies that have evaluated interobserver variability using the proposed ITWG recommendations, which have shown moderate to strong concordance, did not fully adopt the guidelines [11,12]. To reach a predefined acceptable concordance value, pathologists evaluated at least 3 areas of a predefined size (1 mm^2) per tumor for each case [11]. The mean values recorded for each region were used for statistical analysis. This addressed variation due to tumor heterogeneity and random errors [11]. In a separate study which evaluated concordance among pathologists in manual H&E stromal TIL estimation, TILs were categorized into 3 categories (<10%, 10%-50%, and >50%) rather than counted as a continuous variable because "in practice, most pathologists will not report specific values" [12]. This study also identified pitfalls in H&E TIL identification including individual cell necrosis, marked individual tumor cell apoptosis, and presence of reactive plasma cells mimicking tumor cells and plasmacytoid-appearing tumor cells mimicking plasma cells [12]. In response, the authors stated that IHC staining with anti-CD45 or anticytokeratin can be used to differentiate TILs from carcinoma cells, respectively, which is not in keeping with the ITWG guidelines [9,12]. Finally, they explained that TIL quantification in specimens with focal areas of dense inflammatory infiltrate was problematic, a finding previously reported to account for individual outliers [11,12]. With the use of immunohistochemistry, quantification of TILs by DIA is not subject to these interpretation pitfalls, as only immunoreactive cells are enumerated.

We sought to evaluate the consistency of TIL counts among multiple pathologists using hotspots and DIA. For a single hotspot placement by 6 pathologists, the ICCs were 0.96 (density) and 0.97 (absolute) TIL counts, respectively. In fact, we found that, in 32% of cases (21/66), all 6 pathologists chose a hotspot in the same location (overlapping hotspots). Studies evaluating the reproducibility of manual stromal TIL density and Ki67 indices have yielded varied results with ICCs ranging from 0.70 to 0.89 and 0.71 to 0.94, respectively. In these studies, to reach a clinically acceptable ICC, each required a second trial with additional techniques to improve both stromal TIL density and Ki67 index estimation [11,27,28]. With increased availability and utilization of digital pathology, one can foresee a future in which DIA is accessible for clinical use. In fact, billing codes differ based on whether morphometric analysis is performed manually or using computer-assisted technology [30].

Mani et al [13] quantitatively assessed the spatial heterogeneity of TIL subtypes across biopsies, within a single section, and among multiple sections. They found that CD8+ and CD3+ TIL heterogeneity was broad within a single section. In our study, we were able to reduce the effect of CD8+ TIL heterogeneity on variance by selectively focusing on areas with the highest density of immunoreactive CD8+ cells (hotspots). When evaluating 3 hotspots in 3 different locations of a single whole-tissue section, the ICC yielded 0.95 for both density and absolute TIL counts. Thus, within a single section, there are multiple hotspots with similar numbers of immunoreactive cells per unit area from which a pathologist may choose.

We acknowledge that DIA is not clinically available for the majority of pathologists today. In addition, our cohort is composed of only TNBCs, with a significant portion of cases being lymphocytic predominant breast cancer; however, Fig. 3 demonstrates that our cohort is comprised of tumors with a variety of absolute CD8+ TIL counts (range: 176-19 069 cells). We also only assessed a single section per case, but this is less concerning given that Mani et al [13] found that there was more CD8+ TIL heterogeneity within a single block of tumor than between multiple blocks.

5. Conclusion

Hotspot TIL enumeration by DIA is a highly reproducible method for quantification of CD8+ TILs. Given that 3 hotspots in 3 different locations of a single whole-tissue section yielded excellent ICCs, the use of hotspots may reduce TIL count variation because of intratumoral TIL heterogeneity.

Acknowledgment

This work was supported by the Translational Research Program at Weill-Cornell Medicine Pathology and Laboratory Medicine. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] Adams S, Gray RJ, Demaria S, et al. Prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancers from two phase III randomized adjuvant breast cancer trials: ECOG 2197 and ECOG 1199. J Clin Oncol 2014;32:2959-66.

[2] Denkert C, Loibl S, Noske A, et al. Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer. J Clin Oncol 2010;28:105-13.

[3] Ibrahim EM, Al-Foheidi ME, Al-Mansour MM, Kazkaz GA. The prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancer: a meta-analysis. Breast Cancer Res Treat 2014;148:467-76.

[4] Iglesia MD, Vincent BG, Parker JS, et al. Prognostic B-cell signatures using mRNA-seq in patients with subtype-specific breast and ovarian cancer. Clin Cancer Res 2014;20:3818-29.

[5] Loi S, Michiels S, Salgado R, et al. Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: results from the FinHER trial. Ann Oncol 2014; 25:1544-50.

[6] Loi S, Sirtaine N, Piette F, et al. Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02-98. J Clin Oncol 2013;31:860-7.

[7] Matsumoto H, Thike AA, Li H, et al. Increased CD4 and CD8-positive T cell infiltrate signifies good prognosis in a subset of triple-negative breast cancer. Breast Cancer Res Treat 2016;156:237-47.

[8] Pruneri G, Vingiani A, Bagnardi V, et al. Clinical validity of tumor-infiltrating lymphocytes analysis in patients with triple-negative breast cancer. Ann Oncol 2016;27:249-56.

[9] Salgado R, Denkert C, Demaria S, et al. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. Ann Oncol 2015;26: 259-71.

[10] Park HS, Heo I, Kim JY, et al. No effect of tumor-infiltrating lymphocytes (TILs) on prognosis in patients with early triple-negative breast cancer: validation of recommendations by the International TILs Working Group 2014. J Surg Oncol 2016;114:17-21.

[11] Denkert C, Wienert S, Poterie A, et al. Standardized evaluation of tumor-infiltrating lymphocytes in breast cancer: results of the ring studies of the international immuno-oncology biomarker working group. Mod Pathol 2016;29:1155-64.

[12] Swisher SK, Wu Y, Castaneda CA, et al. Interobserver agreement between pathologists assessing tumor-infiltrating lymphocytes (TILs) in breast cancer using methodology proposed by the International TILs Working Group. Ann Surg Oncol 2016;23:2242-8.

[13] Mani NL, Schalper KA, Hatzis C, et al. Quantitative assessment of the spatial heterogeneity of tumor-infiltrating lymphocytes in breast cancer. Breast Cancer Res 2016;18:78.

[14] Kotoula V, Chatzopoulos K, Lakis S, et al. Tumors with high-density tumor infiltrating lymphocytes constitute a favorable entity in breast cancer: a pooled analysis of four prospective adjuvant trials. Oncotarget 2016;7:5074-87.

[15] McIntire PJ, Irshaid L, Liu Y, et al. Hotspot and whole tumor enumeration of CD8+ tumor infiltrating lymphocytes utilizing digital image analysis is prognostic in triple negative breast cancer. Clin Breast Cancer 2018;18: 451-458.e1.

[16] An T, Sood U, Pietruk T, Cummings G, Hashimoto K, Crissman JD. In situ quantitation of inflammatory mononuclear cells in ductal infiltrating breast carcinoma. Relation to prognostic parameters. Am J Pathol 1987; 128:52-60.

[17] Bates GJ, Fox SB, Han C, et al. Quantification of regulatory T cells enables the identification of high-risk breast cancer patients and those at risk of late relapse. J Clin Oncol 2006;24:5373-80.

[18] Gobert M, Treilleux I, Bendriss-Vermare N, et al. Regulatory T cells recruited through CCL22/CCR4 are selectively activated in lymphoid infiltrates surrounding primary breast tumors and lead to an adverse clinical outcome. Cancer Res 2009;69:2000-9.

[19] Mahmoud SM, Lee AH, Paish EC, Macmillan RD, Ellis IO, Green AR. The prognostic significance of B lymphocytes in invasive carcinoma of the breast. Breast Cancer Res Treat 2012;132:545-53.

[20] Mohammed ZM, Going JJ, Edwards J, Elsberger B, McMillan DC. The relationship between lymphocyte subsets and clinico-pathological determinants of survival in patients with primary operable invasive ductal breast cancer. Br J Cancer 2013;109:1676-84.

[21] Wei H, Fu P, Yao M, Chen Y, Du L. Breast cancer stem cells phenotype and plasma cell-predominant breast cancer independently indicate poor survival. Pathol Res Pract 2016;212:294-301.

[22] West NR, Kost SE, Martin SD, et al. Tumour-infiltrating FOXP3(+) lymphocytes are associated with cytotoxic immune responses and good clinical outcome in oestrogen receptor-negative breast cancer. Br J Cancer 2013; 108:155-62.

[23] Lloyd MC, Allam-Nandyala P, Purohit CN, Burke N, Coppola D, Bui MM. Using image analysis as a tool for assessment of prognostic and predictive biomarkers for breast cancer: how reliable is it? J Pathol Inform 2010;1: 29-36.

[24] Mofidi R, Walsh R, Ridgway PF, et al. Objective measurement of breast cancer oestrogen receptor status through digital image analysis. Eur J Surg Oncol 2003;29:20-4.

[25] Stalhammar G, Fuentes Martinez N, Lippert M, et al. Digital image analysis outperforms manual biomarker assessment in breast cancer. Mod Pathol 2016;29:318-29.

[26] Müller R, Büttner P. A critical discussion of intraclass correlation coefficients. Stat Med 1994;13(23–24):2465-76.

[27] Diaz LK, Sahin A, Sneige N. Interobserver agreement for estrogen receptor immunohistochemical analysis in breast cancer: a comparison of manual and computer-assisted scoring methods. Ann Diagn Pathol 2004;8:23-7.

[28] Polley MY, Leung SC, Gao D, et al. An international study to increase concordance in Ki67 scoring. Mod Pathol 2015;28:778-86.

[29] Polley MY, Leung SC, McShane LM, et al. An international Ki67 reproducibility study. J Natl Cancer Inst 2013;105:1897-906.

[30] Digital Pathology Association. FAQs. DPA digitalpathologyassociation.org/healthcare-faqs; 2018.