

ARTICLE



Interobserver variability in the assessment of stromal tumor-infiltrating lymphocytes (sTILs) in triple-negative invasive breast carcinoma influences the association with pathological complete response: the IVITA study

Mieke R. Van Bockstal¹ , Aline François¹, Serdar Altınay^{2,41}, Laurent Arnould^{3,41}, Maschenka Balkenhol^{4,41}, Glenn Broeckx^{5,41}, Octavio Burguès^{6,41}, Cecile Colpaert^{7,41}, Francesca Dedeurwaerdere^{8,41}, Benjamin Dessauvage^{9,10,41}, Valérie Duwel^{11,41}, Giuseppe Floris^{12,13,41}, Stephen Fox^{14,41}, Clara Gerosa^{15,41}, Delfyne Hastir^{16,41}, Shabnam Jaffer^{17,41}, Eline Kurpershoek^{18,41}, Magali Lacroix-Triki^{19,41}, Andoni Laka^{20,41}, Kathleen Lambein^{21,41}, Gaëtan Marie MacGrogan^{22,41}, Caterina Marchiò^{23,24,41}, Maria-Dolores Martin Martinez^{25,41}, Sharon Nofech-Mozes^{26,41}, Dieter Peeters^{27,28,41}, Alberto Ravarino^{15,41}, Emily Reisenbichler^{29,41}, Erika Resetkova^{30,41}, Souzan Sanati^{31,41}, Anne-Marie Schelfhout^{32,41}, Vera Schelfhout^{27,41}, Abeer Shaaban^{33,41}, Renata Sinke^{18,41}, Claudia M. Stanciu-Pop^{34,41}, Carolien H. M. van Deurzen^{35,41}, Koen K. Van de Vijver^{36,41}, Anne-Sophie Van Rompuy^{12,41}, Anne Vincent-Salomon^{37,41}, Hannah Y. Wen^{38,41}, Serena Wong^{29,41}, Caroline Bouzin³⁹ and Christine Galant^{1,40}

© The Author(s), under exclusive licence to United States & Canadian Academy of Pathology 2021

High stromal tumor-infiltrating lymphocytes (sTILs) in triple-negative breast cancer (TNBC) are associated with pathological complete response (pCR) after neoadjuvant chemotherapy (NAC). Histopathological assessment of sTILs in TNBC biopsies is characterized by substantial interobserver variability, but it is unknown whether this affects its association with pCR. Here, we aimed to investigate the degree of interobserver variability in an international study, and its impact on the relationship between sTILs and pCR. Forty pathologists assessed sTILs as a percentage in digitalized biopsy slides, originating from 41 TNBC patients who were treated with NAC followed by surgery. Pathological response was quantified by the MD Anderson Residual Cancer Burden (RCB) score. Intraclass correlation coefficients (ICCs) were calculated per pathologist duo and Bland–Altman plots were constructed. The relation between sTILs and pCR or RCB class was investigated. The ICCs ranged from -0.376 to 0.947 (mean: 0.659), indicating substantial interobserver variability. Nevertheless, high sTILs scores were significantly associated with pCR for 36 participants (90%), and with RCB class for eight participants (20%). *Post hoc* sTILs cutoffs at 20% and 40% resulted in variable associations with pCR. The sTILs in TNBC with RCB-II and RCB-III were intermediate to those of RCB-0 and RCB-I, with lowest sTILs observed in RCB-I. However, the limited number of RCB-I cases precludes any definite conclusions due to lack of power, and this observation therefore requires further investigation. In conclusion, sTILs are a robust marker for pCR at the group level. However, if sTILs are to be used to guide the NAC scheme for individual patients, the observed interobserver variability might substantially affect the chance of obtaining a pCR. Future studies should determine the ‘ideal’ sTILs threshold, and attempt to fine-tune the patient selection for sTILs-based de-escalation of NAC regimens. At present, there is insufficient evidence for robust and reproducible sTILs-guided therapeutic decisions.

Modern Pathology (2021) 34:2130–2140; <https://doi.org/10.1038/s41379-021-00865-z>

INTRODUCTION

Triple-negative breast cancers (TNBCs) lack the expression of estrogen receptor (ER), progesterone receptor (PR) and HER2 [1], and are associated with a higher risk of regional recurrence, lower distant recurrence-free survival and lower overall survival in comparison with other molecular subtypes [2, 3]. The majority of TNBCs are invasive carcinomas of no special type (NST), and the most frequent special type TNBC is metaplastic carcinoma [4].

TNBC patients who present with clinically node-positive and/or at least T1c disease are generally treated with anthracycline- and taxane-based neoadjuvant chemotherapy (NAC), with optional addition of carboplatin, according to the ASCO guideline [5]. Pathological complete response (pCR) after NAC guides subsequent clinical decision-making, and is defined as the absence of residual invasive carcinoma in the breast and lymph nodes [5]. Achieving a pCR is an independent predictor of better disease-free

A full list of author affiliations appears at the end of the paper.

Received: 28 March 2021 Revised: 21 June 2021 Accepted: 21 June 2021

Published online: 3 July 2021

survival in TNBC [6, 7]. Many classification systems were developed to objectify the post-NAC therapeutic response. The well validated MD Anderson Residual Cancer Burden (RCB) applies an equation which contains information on both the cellularity and the size of residual carcinoma in the breast and lymph nodes [7]. It is considered the gold standard for assessment of pathological response in NAC clinical trials, shows excellent interobserver agreement, and is characterized by a highly reproducible long-term prognostic significance [8, 9].

Two randomized clinical trials showed that high levels of stromal tumor-infiltrating lymphocytes (sTILs) are predictive for achieving a pCR in TNBC [10, 11]. This was confirmed in retrospective studies beyond trial setting [12–14]. High TILs levels also provide prognostic information, as they are associated with better distant recurrence-free survival in TNBC patients treated with and without NAC [10, 15]. The International Immunology Biomarkers Working Group developed a method to quantify the amount of sTILs in the peri-tumoral stroma of solid tumors such as breast cancer [16, 17]. This method evaluates sTILs for the stromal compartment within the borders of the invasive tumor, and the area of stromal tissue serves as the denominator to determine the percentage of sTILs [17].

Small-scale studies on interobserver variability among two to four pathologists reported variable concordance rates, ranging from substantial agreement to a relatively high level of imprecision [18–20]. Larger studies, wherein nine to thirty-two pathologists evaluated sTILs in a predefined set of breast cancers, consistently reported acceptable and moderate agreement [21–23]. However, none of these studies investigated the impact of interobserver variability on the predictive value of sTILs for achieving a pCR. We, therefore, aimed to investigate the interobserver agreement and association of individual pathologists' sTILs scores with the therapeutic response, defined as either pCR or RCB class. We organized a large-scale international study on 'interobserver variability in TILs assessment' (IVITA), by using a consecutive real-life set of TNBC biopsies outside the randomized clinical trial setting.

MATERIALS AND METHODS

Tissue samples and clinic-pathological data

Archived hematoxylin and eosin (HE) stained slides of the pre-NAC biopsy and post-NAC resection specimen were collected for a consecutive series of TNBC patients at the Cliniques universitaires Saint-Luc (Brussels, Belgium). All patients included in this study were diagnosed with TNBC and underwent surgery between 1 January 2015 and 30 September 2020. Hormone receptor status and HER2 status were defined according to the ASCO/CAP guidelines [24, 25]. The standard NAC scheme included anthracyclines and cyclophosphamide, followed by paclitaxel. Patients with a poor response after anthracyclines and cyclophosphamide also received carboplatin. Information on patient age at diagnosis, type of surgery, time interval between the biopsy and surgery, post-NAC nodal status, macroscopic and microscopic tumor bed size, hormone receptor status and HER2 status was retrieved from the electronic histopathological reports (LIS DaVinci, MIPS, Ghent, Belgium). The institutional ethics committee approved this study (file number: RETRO-TNBC-15-2019/03JUL/297).

Histopathological central review

All biopsies were immediately fixed in 10% neutral-buffered formalin for 6–72 h. Macroscopic examination of post-NAC lumpectomy and mastectomy specimens was performed according to the MD Anderson residual cancer burden (RCB) protocol [7]. All resection specimens were sliced at 5 mm intervals and fixed in 10% neutral-buffered formalin for 6–72 h, in line with the ASCO/CAP guidelines [24]. Histopathological assessment of the biopsies and the resection specimens was performed as previously described [12], and comprised the Nottingham grade, and presence of ductal carcinoma in situ (DCIS) component and unequivocal lymphovascular invasion. The H&E stained slides of all resection specimens were reviewed by two pathologists (AF and MRVB). Archived immunohistochemical stains for p63 and smooth muscle

myosin heavy chain (SMMHC) were available to discern residual DCIS from invasive carcinoma. The therapeutic response after neoadjuvant chemotherapy was objectified by using an online calculator for the RCB score (<http://www3.mdanderson.org/app/medcalc/index.cfm?pagename=jsonconvert3>) [7]. For each patient, the RCB score and corresponding RCB class were noted. An RCB score of zero (RCB-0) was considered as a pCR.

sTILs assessment

The extent of the stromal inflammatory infiltrate in the pre-NAC biopsy was assessed according to the standardized method as described in detail by the International Immunology Biomarkers Working Group [16]. The number of sTILs was noted as the percentage of mononuclear inflammatory cells related to the total peri- and intra-tumor stromal surface area, which served as a denominator [16]. The number of fields was not specified: participants had to evaluate the entire area occupied by invasive carcinoma. No training set was provided, but all participants were provided with the appropriate literature [16, 17, 21], as well as the tutorial of the website www.tilsinbreastcancer.org, which served as a guideline during the sTILs assessment. A similar method has been applied before [21]. All participants evaluated the same set of digitalized pre-NAC core needle biopsy slides. For each patient, one biopsy slide was digitalized by an automated slide scanner with Z-stack feature (NanoZoomer 2.0-RS, Hamamatsu Photonics K.K., Hamamatsu City, Japan). Evaluation of the post-NAC resection specimen was not requested.

Participating pathologists

Participating pathologists with a special interest in breast disease had to actively work as reporting pathologist, either in academic or non-academic laboratories. As an inclusion criterion, all participants had to assess a minimum of 50 primary (oncologic) breast cancer resection specimens per year, in line with the EUSOMA-criteria for dedicated breast pathologists [26]. Most participants previously participated in the digital DCISion study [27]. The following data on the observers were collected via a questionnaire with twenty questions: number of years in practice (including training), the work environment (academic or non-academic laboratory), the daily work method (conventional light microscopy or digital pathology), and the weekly breast pathology workload expressed as a percentage of a full-time week schedule. Information on the habits of evaluating and reporting sTILs was also collected. All participants had digital access to the 41 scanned H&E slides, which were available on the password-protected Cytomine platform [28]. The identity of each participant was anonymized as P1, P2, P3, etc by one pathologist (MRVB), who collected all participants' sTILs scores.

Statistical analysis

The questionnaire results were analyzed, and pie charts and radar diagrams were constructed in Excel (Excel Windows 10, Microsoft Corporation, Redmond, WA, USA). Statistical analyses were performed with IBM SPSS statistics 26.0 (IBM Chicago, IL, USA). Tests for normality were performed with the Shapiro–Wilk test, which showed that the sTILs scores of each participant were not normally distributed ($p < 0.05$; Supplementary Table 1). Therefore, the median (instead of the average) sTILs value was selected for each case to serve as the 'gold standard', based on the assessment of all participants. This 'median' (nonexistent) pathologist was designated 'Px', and a histogram and stem-and-leaf plot were constructed to illustrate the non-normal distribution. Associations between the median Px sTILs scores and different histopathological characteristics were investigated by applying Mann–Whitney *U* and Kruskal–Wallis tests, depending on the number of categories of the characteristic of interest. Mann–Whitney *U* tests and Kruskal–Wallis tests were also performed to investigate associations between the individual sTILs scores (as a continuous variable) and either pCR or RCB class, respectively. Box-and-whisker plots visualized these associations. Next, all sTILs scores were dichotomized *post hoc* according to seven different thresholds (5, 10, 20, 30, 40, 50, and 60%), which included previously reported cutoffs for dichotomization [10, 16]. Low TILs were defined as sTILs lower than or equaling (\leq) each threshold. High TILs were defined as sTILs greater than ($>$) each threshold. Chi-square tests were performed to investigate associations between these sTILs estimates and pCR, and both absolute numbers and column percentages were reported in cross tables. Lastly, the range between the 25th and 75th percentile of the sTILs scores was calculated for each case as a 'surrogate' measure for interobserver variability, and the association of this range with the different histopathological features was investigated, by using Mann–Whitney *U* and Kruskal–Wallis tests. All tests were two-sided and

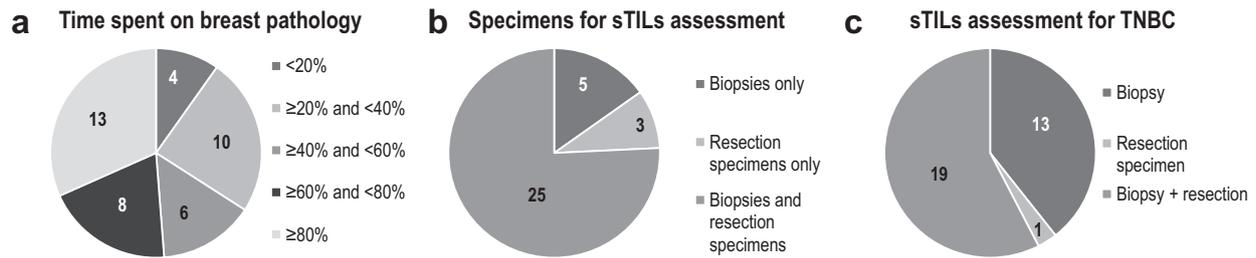


Fig. 1 Pie charts. **a** Distribution of the time spent on breast pathology (a), as reported by each pathologist based on a full-time week schedule. **b** Specimens used for sTILs assessment in general, regardless of the molecular subtype, as reported by 33 participants. **c** Specimens used for sTILs assessment in TNBC, as reported by 33 participants.

the significance level was set at $p < 0.05$, except for Kruskal–Wallis tests, where we applied a *post hoc* Bonferroni correction for multiple testing ($p < 0.0083$).

Interobserver variability was quantified by calculation of the intraclass correlation coefficients (ICC) for sTILs scores, as previously described [27]. The interpretation was performed according to Koo and Li [29]. ICC settings were: two-way random, single measures, absolute agreement. Bland–Altman plots were constructed to visualize the degree of deviation from the median sTILs score P_x , by using both the mean of and the difference between each pathologist's sTILs scores and P_x sTILs scores.

RESULTS

Profile of the participants

Forty-one pathologists were invited to participate. All pathologists completed the questionnaire, and forty pathologists (98%) assessed sTILs in the series of digitalized biopsy slides. The participants represented thirty-four laboratories from eleven countries (Australia, Belgium, Canada, France, Italy, Spain, Switzerland, The Netherlands, Turkey, the United Kingdom, and the United States of America). The participants had been practicing pathology for 18.6 years on average (range 3–35 years). Twenty-eight pathologists (68%) worked in academic laboratories; eleven pathologists (27%) worked in non-academic laboratories and two pathologists (5%) worked in both settings. Conventional light microscopy and digital pathology were used on a daily basis by thirty (73%) and four (10%) pathologists, respectively. Seven pathologists (17%) used both techniques in routine practice. The estimated time spent on breast pathology, based on a full-time working schedule, is shown in Fig. 1a. Thirty-five participants (85%) were aware of the 'International Immuno-Oncology Biomarker Working Group on Breast Cancer' before their participation in the IVITA study, while five (12%) had not yet heard about the Working Group and one (2%) was uncertain. Thirty-one participants (76%) had already visited the website of the Working Group before participating in IVITA, whereas (24%) ten participants did not. One participant (2%) reported to have never assessed the post-NAC therapeutic response in TNBC; four (10%) and two (5%) participants reported using the Pinder regression score or the Miller–Payne system, respectively. Twenty-five participants (61%) applied the MD Anderson RCB score in routine practice. In addition, three participants (7%) combined the RCB score and the Pinder regression score, and two participants (5%) used both the RCB score and the Miller–Payne system. One participant (2%) mentioned the use of the 'Residual Disease in Breast and Nodes' system, whereas two participants (5%) mentioned the EUSOMA recommendations. One participant (2%) indicated 'other classification system', without further specifications. None of the participants used the Chevallier classification, Sataloff's classification, or Nottingham Clinico-Pathological Response Index.

sTILs reporting practice of the participants

Eight pathologists (20%) never mentioned sTILs in the reports of invasive breast cancer patients. Eighteen (44%) and fifteen (37%) pathologists always or sometimes assessed sTILs in invasive breast

cancer, respectively. In this subgroup of 33 pathologists, 25 (76%) reported sTILs for all molecular subtypes. One pathologist (3%) only mentioned sTILs in TNBC, whereas four pathologists (12%) assessed sTILs in both TNBC and HER2-positive breast cancer. Two pathologists (6%) stated that they only mentioned sTILs when the stromal immune infiltrate is marked, regardless of the molecular subtype. The specimen type used for sTILs assessment, in general, is displayed in Fig. 1b. Reporting practices for sTILs in TNBC according to specimen type are shown in Fig. 1c. Nineteen pathologists (46%) did not report sTILs in DCIS, fourteen (34%) pathologists sometimes mentioned sTILs in pure DCIS, whereas six (15%) pathologists always reported TILs in DCIS.

Twenty-one pathologists (64%) assessed sTILs as a percentage of the stromal surface area, as described by the 'International Immuno-Oncology Biomarker Working Group on Breast Cancer' [16]. Ten pathologists (30%) provided a semi-quantitative score based on their own personal interpretation of the degree of stromal inflammation, and two pathologists (6%) only added a comment when the stromal inflammatory infiltrate was marked. When pathologists mentioned sTILs as a percentage, twenty-three participants (82%) did not use a cutoff, whereas five (18%) did use a threshold to indicate whether a particular case has 'low TILs', 'intermediate TILs' or 'high TILs'. Each of these five participants used different thresholds, ranging from 5 to 50%.

Perception of sTILs assessment and its consequences

All participants were asked to estimate the difficulty of sTILs assessment on a scale from 0 to 10, which was most often reported to be moderate (Fig. 2a). The need for standardization of sTILs assessment in daily routine practice was questioned in a similar way and was estimated to be rather high (Fig. 2b).

Thirty-five participants (85%) reported to regularly attend multidisciplinary meetings to discuss the clinical management of breast cancer patients. Twenty-four participants (59%) indicated that clinicians actively ask for sTILs assessment during these meetings, either on a regular basis or occasionally. Fifteen pathologists (37%) reported that clinicians never ask for sTILs during these multidisciplinary meetings, and three participants had no opinion (7%). According to fourteen participants (34%), sTILs scores never influenced the NAC treatment scheme for TNBC patients, whereas two additional participants (5%) indicated that this was not yet the case, but very likely to happen in the near future. Seven (17%) and fourteen (34%) participants responded that sTILs influenced the NAC treatment scheme in TNBC on a regular basis, or occasionally, respectively.

Histopathological characteristics

The TNBC dataset contained two biopsies (5%) of pleomorphic invasive lobular carcinoma and 39 cases (95%) of invasive ductal carcinoma of no special type (NST). The mean age at diagnosis was 55 years (range 31–83). The mean interval between the biopsy and the surgical resection was 5.8 months (range 2.5–10.3 months).

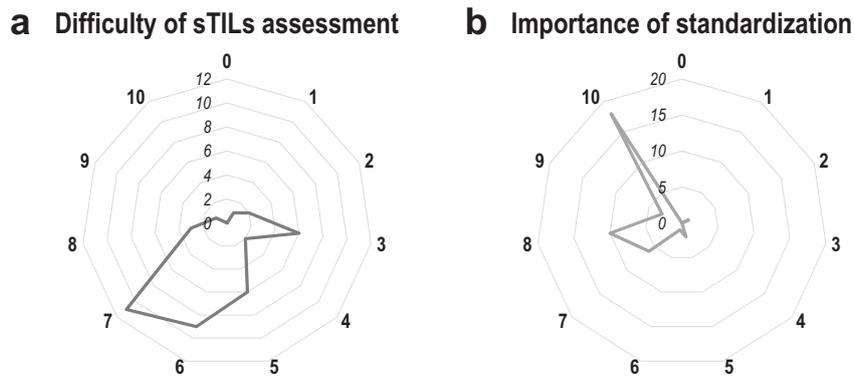


Fig. 2 Radar diagrams. Radar diagrams illustrating the perceived difficulty of sTILs assessment (a) and the perceived importance of standardization of sTILs assessment in daily routine practice (b), as reported by 41 pathologists. The scale ranged from 0 (very low) to 10 (very high).

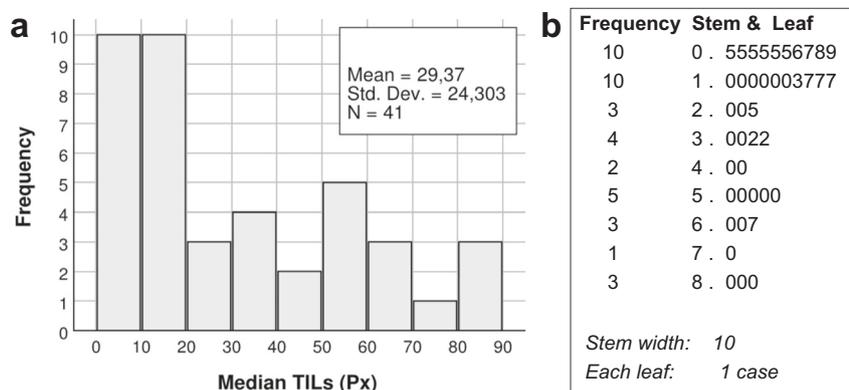


Fig. 3 Non-normal distribution of sTILs in TNBC biopsies. Histogram (a) and stem-and-leaf plot (b) illustrating the non-normal distribution of the median sTILs scores (Px) in this series of 41 TNBC biopsies.

This interval did not significantly correlate with pCR ($p = 0.262$). Ten TNBC (24%) were of grade 2, and thirty-one (76%) were grade 3. Three TNBC (7%) presented with lymphovascular invasion in the biopsy, and seven TNBC (17%) contained DCIS. The RCB classes in this dataset were as follows: sixteen cases of RCB-0 (39%), five RCB-I (12%), thirteen RCB-II (32%), and seven RCB-III (17%). The sTILs dataset contained three missing values, represented by two cases that were not assessed by two pathologists because they were considered as extensive DCIS without clear invasion. These cases were not excluded from the analysis.

Figure 3 contains a histogram and corresponding stem-and-leaf plot that illustrate the non-normal distribution of the median sTILs score (Px) for each biopsy included in this study (Shapiro–Wilk test: $p < 0.001$). Median Px sTILs were not associated with grade ($p = 0.346$), the presence of lymphovascular invasion ($p = 0.629$), the presence of an in situ component in the biopsy ($p = 0.176$), or age at diagnosis ($p = 0.775$).

Quantification of interobserver variability

Supplementary Table 2 contains the ICC values for each pathologist duo. The ICCs range from -0.376 to 0.947 , with a mean value of 0.659 , indicating an overall substantial interobserver variability [29]. Based on the mean of each pathologist's sTILs scores and Px, as well as the difference between each pathologist's sTILs scores with Px, Bland–Altman plots were constructed to visualize the degree of discordance (Supplementary Fig. 1; Fig. 4). Overall, 'low' sTILs cases show less variability than cases with 'intermediate' or 'high' sTILs. TNBC with higher

sTILs levels is generally characterized by a wider range among the different sTILs ratings by the participants. However, the observed interobserver variability was not related to any of the histopathological characteristics. For instance, the range between the 25th and 75th percentile of Px was not associated with the presence of a DCIS component ($p = 0.543$) or tumor grade ($p = 0.394$). The interobserver variability was not associated with any of the laboratory settings or sTILs reporting habits ($p > 0.05$).

Associations between sTILs and therapeutic response

Table 1 contains the descriptive values for the sTILs scores for each individual pathologist and the median Px. We observed a statistically significant association between high sTILs scores and the presence of a pCR for 36 out of forty pathologists (90%). The sTILs scores of one pathologist (2%) were inversely associated with pCR, i.e. high sTILs scores were associated with lack of a pCR. Similar analyses were performed for associations with the RCB class, wherein 'absent pCR' was represented by RCB-I, -RCB-II and RCB-III. Here, a *post hoc* Bonferroni correction for multiple testing was applied, i.e. the level of significance was set at 0.0083 . sTILs were associated with RCB class in only eight out of forty (20%) pathologists. Box-and-whisker plots (Supplementary Fig. 2) show that TNBC with RCB-II and RCB-III usually have sTILs levels that are intermediate to those of RCB-0 and RCB-I, with the highest sTILs levels observed in RCB-0 and the lowest observed in RCB-I. This was also observed for the median Px sTILs (Fig. 5).

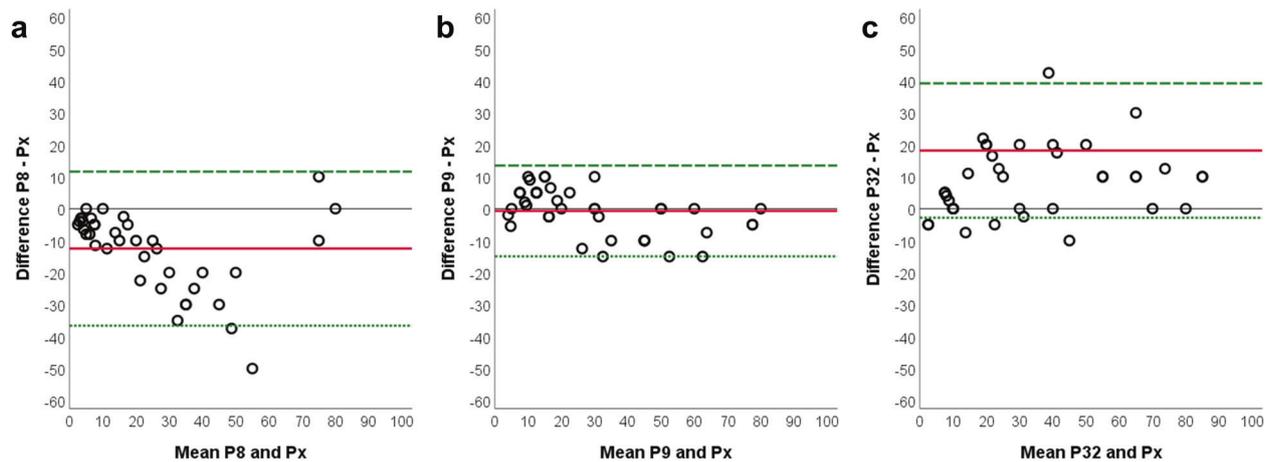


Fig. 4 Bland-Altman plots based on the ratings of three participants of the IVITA study. Example of three Bland-Altman plots, showing a substantial lower rating of P8 when compared with Px (a), near-perfect agreement between P9 and Px (b), and a substantial higher rating of P32 when compared with Px (c). Other Bland-Altman plots are shown in Supplementary Fig. 1. The full red line is the mean difference, and the dashed and dotted green lines represent the upper and lower limits of the 95% confidence interval of the mean.

Post hoc dichotomization using different sTILs thresholds

To identify a cutoff that could be used to select patients who are more likely to achieve a pCR in routine clinical practice, seven thresholds were explored. All sTILs scores of each pathologist were dichotomized as low sTILs versus high sTILs. The 5% cutoff resulted in a significant association between sTILs classification and pCR for only 9 pathologists (23%), whereas the 10% cutoff resulted in a similar association for 19 pathologists (48%; Table 2 and Supplementary Table 3). The 20%, 30, and 40% thresholds resulted in a significant association between sTILs and pCR for 30, 31, and 28 out of 40 pathologists, respectively (75%, 78, and 70%). The 50 and 60% cutoff resulted in a similar association for 25 and 22 out of 40 pathologists, respectively (63 and 55%). Overall, pathologists who generally limit their sTILs score in a narrow range in the lower half of the spectrum do not benefit from a high threshold such as the 40% or 50% cutoff, as too many pCR cases are considered to have low TILs. This was the case for pathologists P1, P8, P21, P26, P30, P31, and P33. On the other hand, pathologists who tend to give high sTILs estimates show a correlation with pCR at a higher sTILs threshold, such as pathologists P13, P15, P17, P32, and P36 (Supplementary Table 3), because low threshold results in few TNBC being designated as having low TILs.

DISCUSSION

In the present study, we demonstrate substantial interobserver variability in sTILs assessment, although the ICC values strongly vary among the different participants. As the participating pathologists work in different countries, employ different laboratory settings (academic versus non-academic, digital versus conventional microscopy, etc) and differ in their reporting habits (quantifying therapeutic response, routine sTILs reporting or not, etc), several factors might have influenced the observed degree of discordance. The variation in practice of TILs reporting from the survey is an interesting finding and calls for more standardization, as was acknowledged by the participants. Unfortunately, the heterogeneous characteristics of the participants do not allow extensive statistical analysis due to lack of power. Similarly, it was impossible to investigate a potential 'training center effect'. In addition, various pitfalls in the sTILs assessment may also have contributed to increased discordance, including crush artifacts, section artifacts due to blunt microtome knives, overstained specimens, extensive tumor necrosis, solid TNBC architecture mimicking pure DCIS, limited intra- and peri-tumoral stroma, and

extensive neutrophilic infiltration (Fig. 6), as previously described [17]. Although we aimed to obtain a 'real-life' biopsy dataset, the evaluation of a single digitalized archived H&E slide does not correspond to the 'real-life' setting. In routine practice, deeper levels are available to cope with technical artifacts, and immunohistochemical stains for myoepithelial markers are available to distinguish in situ from invasive components. Most participants did not use digital pathology on a daily basis, which might also have influenced the sTILs scores.

Interestingly, the individual sTILs scores were statistically significantly associated with the therapeutic response for 90% of all participants, despite the presence of substantial interobserver variability and despite the limited size of the evaluated TNBC cohort. This observation indicates that high sTILs are a robust predictive marker for achieving a pCR after NAC in TNBC, at least at the population level. The 2019 Saint Gallen International Consensus Panel recommended that sTILs be routinely assessed in TNBC because of their prognostic value [30], although this has not been widely adopted in international guidelines. Nevertheless, the 2021 Saint Gallen International Consensus Panel voted against the routine use of sTILs in early TNBC, as evidence on sTILs for the guidance of NAC regimens in TNBC patients is lacking [31, 32]. This contrasts with the perception of twenty-one participants in the present study, who inadvertently assumed that sTILs in the pre-NAC biopsy influenced the NAC treatment at least occasionally.

The above variation in sTILs assessment to identify patients likely to achieve a pCR might impact the clinical decision-making if sTILs would be used one day to guide the NAC regimen for individual patients. At present, sTILs are reported as a continuous variable, but any future clinical decision-making will require a particular threshold. Although there is insufficient evidence to de-escalate NAC at present [31, 32], future studies should determine this 'ideal' sTILs threshold, i.e. how much sTILs in the pre-NAC biopsy are sufficient to de-escalate the NAC regimen, without compromising the chance of achieving a pCR for a significant number of patients?

The introduction of a particular threshold to guide clinical decision-making will have to be accompanied by education of pathologists to render sTILs assessment more uniform. Computational assessment by the use of machine learning models might aid to objectify sTILs levels in TNBC in the future [33]. In the present study, we explored seven different *post hoc* thresholds for sTILs assessment, which affect the number of TNBC that are designated as 'high sTILs' and 'low sTILs', as well as the association with pCR. The total number of statistically significant associations between

Table 1. Descriptive statistics and associations between TILs and either pCR or RCB class per pathologist.

Rater	TILs versus pCR <i>p</i> value	TILs versus RCB class <i>p</i> value	Mean TILs	Min TILs	Max TILs	Range TILs	Pe25 TILs	Pe50 TILs	Pe75 TILs
P01	0.020 ^a	0.073	23	0	80	80	5	20	35
P02	0.003 ^a	0.031	22	0	77	77	5	10	40
P03	0.001 ^a	0.004 ^a	24	2	90	88	5.5	10	42.5
P04	0.001 ^a	0.004 ^a	42	5	100	95	20	40	70
P05	0.101	0.394	36	2	90	88	7.5	30	60
P06	0.002 ^a	0.011	35	5	80	75	10	30	55
P07	0.003 ^a	0.019	43	5	90	85	17.5	40	70
P08	0.003 ^a	0.006 ^a	17	0	80	80	2	10	20
P09	0.001 ^a	0.007 ^a	29	2	80	78	15	20	40
P10	0.004 ^a	0.021	20	1	80	79	5	10	30
P11	0.014 ^a	0.083	31	1	80	79	5	20	30
P12	0.004 ^a	0.021	29	5	90	85	8	15	50
P13	0.007 ^a	0.020	31	1	90	89	10	30	40
P14	0.002 ^a	0.011	17	0	90	90	0	5	40
P15	0.011 ^a	0.045	30	3	80	77	8	18	55
P16	0.003 ^a	0.021	44	5	100	95	20	40	75
P17	0.005 ^a	0.034	38	1	90	89	10	30	65
P18	0.120	0.333	25	4	80	76	15	25	62.5
P19	0.019 ^a	0.081	27	0	90	90	50	10	45
P20	0.009 ^a	0.041	25	1	80	79	5	20	45
P21	0.024 ^a	0.054	26	2	90	88	5	10	30
P22	0.002 ^a	0.018	26	2	90	88	12.5	25	55
P23	0.013 ^a	0.016	20	1	70	69	5	17.5	40
P24	0.001 ^a	0.009	31	5	90	85	10	20	70
P25	0.006 ^a	0.030	31	2	90	88	10	40	70
P26	0.016 ^a	0.036	25	0	80	80	5	10	35
P27	0.006 ^a	0.044	29	5	80	75	5	30	70
P28	0.014 ^a	0.092	38	1	95	94	1	10	80
P29	0.004 ^a	0.018	29	2	95	93	10	30	55
P30	0.001 ^a	0.008 ^a	23	2	90	88	7.5	15	30
P31	0.013 ^b	0.055	30	1	90	89	20	30	75
P32	0.002 ^a	0.006 ^a	27	0	90	90	10	30	60
P33	0.024 ^a	0.049	25	0	80	80	5	20	40
P34	0.002 ^a	0.007 ^a	30	1	90	89	7.5	20	60
P35	0.159	0.561	28	1	90	89	5	10	35
P36	0.005 ^a	0.009	20	5	75	70	10	30	50
P37	0.003 ^a	0.009	32	3	100	97	10	40	75
P38	0.014 ^a	0.075	35	10	80	70	20	30	50
P39	0.002 ^a	0.007 ^a	38	0	100	100	5	10	35
P40	0.001 ^a	0.009	34	1	90	89	7.5	20	70
Px	0.004 ^a	0.020	29	5	80	75	9.5	20	50

^aStatistically significant association, with the significance level set at 0.05 (for TILs versus pCR, Mann–Whitney U test) or 0.0083 (for TILs versus RCB class, Kruskal–Wallis test with *post hoc* Bonferroni correction).

^bInverse statistically significant association, i.e., pCR was associated with lower TILs levels.

Max maximum TILs value, *Min* minimum TILs value, *Pe25* 25th percentile, *Pe50* 50th percentile or median, *Pe75* 75th percentile, *pCR* pathological complete response, *Range* difference between maximum and minimum TILs, *TILs* tumor-infiltrating lymphocytes.

pCR and individual sTILs assessments did not substantially differ between the 20%, 30 and 40% thresholds: 30, 31 and 28 out of 40 pathologists, respectively. However, the association depended on the ‘stringency’ of the sTILs assessment. For instance, pathologists who gave low sTILs estimates did not benefit from the thresholds above 40%, which assigned too many TNBC cases to the ‘low sTILs’

category. Pathologists who gave high sTILs estimates benefited from the higher sTILs thresholds, as the thresholds below 30% assigned too many non-pCR TNBC to the ‘high sTILs’ category (Table 2; Supplementary Table 3). Of note, the participants were not aware of these thresholds at the time of the assessment, and therefore, the use of ad hoc thresholds would likely provide

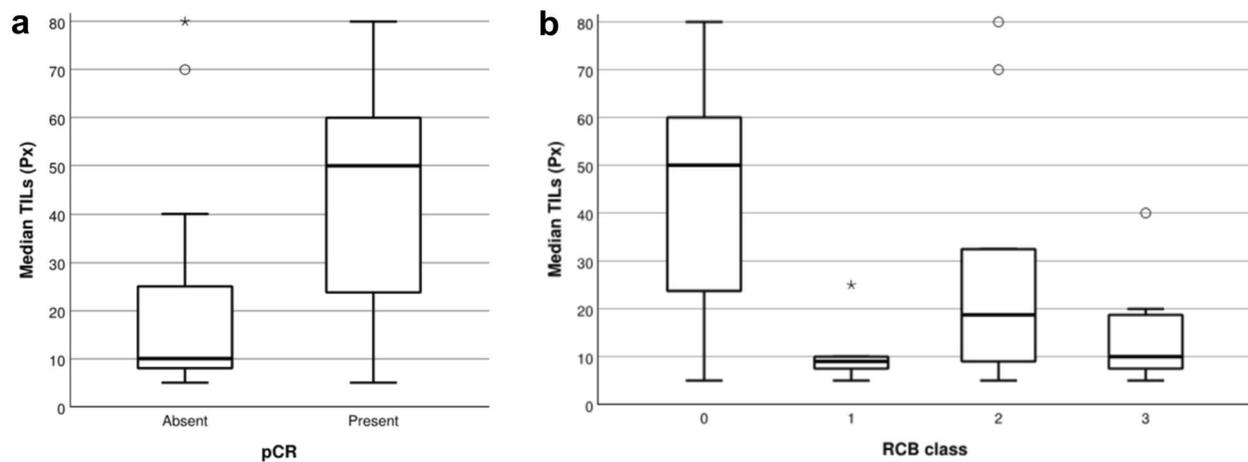


Fig. 5 Box-and-whisker plots. These plots illustrate the association between median sTILs (Px) scores and the absence or presence of pCR (**a**), and the association between median sTILs (Px) scores and the RCB class (**b**). Circles represent outliers; asterisks represent extremes. The bold line within each box represents the median value (50th percentile), the upper and lower limits of the boxes represent the 75th and 25th percentiles, respectively.

different results. Future studies should investigate ad hoc which sTILs threshold is characterized by acceptable interobserver variability among a large community of pathologists. Simultaneously, the selected threshold should have an acceptable 'degree of error', i.e. how many 'false-negative' high sTILs TNBC and 'false-positive' low sTILs TNBC patients are tolerated? The former will not be treated with a de-escalated NAC regimen and are exposed to potential side effects, whereas the latter are inadvertently undertreated by a de-escalated NAC regimen and have smaller chances of achieving a (near) pCR. Additional research is required to explore this difficult equilibrium.

The interobserver variability observed in sTILs assessment in TNBC shows striking similarities with Ki-67 assessment in early hormone receptor-positive, HER2-negative breast cancer, which shows substantial inter-laboratory and interobserver variability as well [34, 35]. Similar to sTILs, Ki-67 was associated with pCR both as a continuous variable and as a dichotomized variable at several thresholds, in the neoadjuvant GeparTrio trial [36]. Pathologists and oncologists will have to face similar challenges in sTILs assessment, but the experience with the issues in Ki-67 assessment might provide useful information for the implementation of sTILs as a quantitative biomarker in TNBC.

Although we observed a strong association between high sTILs and high pCR rates in TNBC for most participants, this was not the case when the individual sTILs scores were correlated with the RCB class: a statistically significant association was observed for only 20% of the participants. Heterogeneously distributed sTILs are unlikely to be responsible for this phenomenon, as Cha et al. have shown that sTILs in core needle biopsies strongly correlated with sTILs in subsequent resections [37]. In addition, Althobiti et al. reported no significant difference between sTILs across different tumor blocks of the same case [38]. In the present cohort, the reduced association with RCB class was mainly due to the RCB-II and RCB-III cases, which showed sTILs levels intermediate to those observed in RCB-0 and RCB-I. This peculiar observation may suggest that pCR is multifactorial. There might be a role for failing immune responses, as several of these RCB-II/III cases contained an almost similar number of sTILs than some TNBC with post-NAC pCR. However, the limited size of the present TNBC cohort precludes any strong conclusion regarding sTILs levels in RCB-I cases, due to a lack of power. Our observation requires validation in larger, independent patient cohorts to exclude findings merely due to chance.

Although assessment of sTILs in residual disease was beyond the scope of the present study, sTILs in residual post-NAC TNBC could add further prognostic information to RCB class, as high residual sTILs levels are associated with improved recurrence-free and overall survival [39].

Future studies should explore whether additional analyses can fine-tune the prognostic and predictive value of sTILs. Immunohistochemical subtyping of sTILs may elucidate which immune cell subtypes stimulate an anti-tumor response during NAC. For instance, high post-NAC levels of CD4-positive lymphocytes in RCB-II and RCB-III TNBC seem to be associated with longer distant recurrence-free survival, and their prognostic value is independent of the RCB class [40]. High pre-NAC levels of CD4-positive lymphocytes are also associated with higher rates of pCR in a breast cancer cohort containing various molecular subtypes [41]. Inflammatory breast cancer patients with high numbers of intra-tumor CD20-positive and CD8-positive lymphocytes respond better to treatment (Badr et al.—submitted manuscript). New technologies such as multiplex immunofluorescent profiling of the immune microenvironment and whole transcriptome RNA sequencing may also aid the future fine-tuning of sTILs as a predictive marker for pCR. Immunomodulatory mRNA signatures and the PAM50 basal-like profile are associated with significantly higher pCR rates in TNBC [42]. Immune-associated mRNA signatures were associated with pCR after NAC in the GeparNuevo trial, although they were of limited use to predict the response to additional immune checkpoint blockade by durvalumab [43].

Patients with metastatic or locally advanced TNBC are eligible for treatment with immune checkpoint inhibitors such as atezolizumab, on the condition that the PD-L1 expression on immune cells occupies $\geq 1\%$ of the tumor area [44]. Atezolizumab represents the first targeted therapy for TNBC patients [45]. The addition of neoadjuvant pembrolizumab to the NAC regimen for stage II/III TNBC patients significantly increased the chance of obtaining a pCR in the phase 3 KEYNOTE-522 trial, regardless the PD-L1 status [46]. Other immune checkpoint inhibitors such as durvalumab are currently being evaluated in a clinical trial setting. Despite the poor reproducibility of PD-L1 assessment in a prospective multi-institutional assessment [47], the interobserver variation seems more limited within a single institution [48]. PD-L1 expression in sTILs might be useful to identify patients at high risk for poor therapeutic response. Consequently, these patients may

Table 2. *p* values illustrating the association between sTILs and pCR per pathologist by applying seven different cutoffs to discern low sTILs from high sTILs.

	Applied threshold for low sTILs versus high sTILs						
	5%	10%	20%	30%	40%	50%	60%
TILs (P1)	0.010 ^a	0.028 ^a	0.087	0.118	0.305	0.636	0.834
TILs (P2)	0.059	0.007 ^a	0.001 ^a	0.007 ^a	0.002 ^a	0.305	0.305
TILs (P3)	0.030 ^a	0.007 ^a	<0.001 ^a	<0.001 ^a	0.002 ^a	0.305	0.744
TILs (P4)	0.418	0.141	0.033 ^a	0.002 ^a	<0.001 ^a	<0.001 ^a	0.001 ^a
TILs (P5)	0.054	0.124	0.202	0.072	0.323	0.323	0.706
TILs (P6)	0.150	0.154	0.018 ^a	0.001 ^a	<0.001 ^a	0.002 ^a	0.020 ^a
TILs (P7)	0.246	0.086	0.005 ^a	0.015 ^a	0.014 ^a	0.017 ^a	0.020 ^a
TILs (P8)	0.018 ^a	0.001 ^a	0.007 ^a	0.636	0.834	0.834	0.834
TILs (P9)	0.150	0.086	0.003 ^a	0.001 ^a	0.007 ^a	0.133	0.308
TILs (P10)	0.054	0.007 ^a	0.006 ^a	0.017 ^a	0.123	0.008 ^a	0.023 ^a
TILs (P11)	0.096	0.015 ^a	0.003 ^a	0.010 ^a	0.002 ^a	0.020 ^a	0.020 ^a
TILs (P12)	0.922	0.033 ^a	0.001 ^a	0.001 ^a	<0.001 ^a	0.002 ^a	0.025 ^a
TILs (P13)	0.150	0.050	0.051	0.006 ^a	0.007 ^a	0.054	0.120
TILs (P14)	0.007 ^a	0.002 ^a	<0.001 ^a	0.001 ^a	0.025 ^a	0.025 ^a	0.070
TILs (P15)	0.242	0.323	0.055	0.006 ^a	<0.001 ^a	<0.001 ^a	0.005 ^a
TILs (P16)	0.246	0.052	0.018 ^a	0.005 ^a	0.010 ^a	0.029 ^a	0.044 ^a
TILs (P17)	0.224	0.010 ^a	0.121	0.021 ^a	0.005 ^a	0.036 ^a	0.002 ^a
TILs (P18)	0.305	0.819	0.154	0.192	0.058	0.096	0.156
TILs (P19)	0.096	0.021 ^a	0.029 ^a	0.006 ^a	0.021 ^a	0.133	0.305
TILs (P20)	0.096	0.028 ^a	0.002 ^a	0.001 ^a	0.002 ^a	0.007 ^a	0.636
TILs (P21)	0.121	0.029 ^a	0.020 ^a	0.054	0.281	0.133	0.025 ^a
TILs (P22)	0.242	0.156	0.005 ^a	0.010 ^a	<0.001 ^a	0.002 ^a	0.020 ^a
TILs (P23)	0.013 ^a	0.041 ^a	0.140	0.060	0.102	0.278	0.191
TILs (P24)	0.242	0.003 ^a	0.001 ^a	0.001 ^a	0.002 ^a	<0.001 ^a	<0.001 ^a
TILs (P25)	0.545	0.098	0.028 ^a	0.015 ^a	0.003 ^a	<0.001 ^a	0.001 ^a
TILs (P26)	0.192	0.055	0.006 ^a	0.021 ^a	0.129	0.305	0.636
TILs (P27)	0.098	0.058	0.028 ^a	0.003	0.001 ^a	0.002 ^a	0.007 ^a
TILs (P28)	0.028 ^a	0.003 ^a	0.005 ^a	0.005 ^a	0.006 ^a	0.006 ^a	0.007 ^a
TILs (P29)	0.692	0.236	0.015 ^a	0.001 ^a	<0.001 ^a	0.002 ^a	0.054
TILs (P30)	0.030 ^a	0.033 ^a	0.001 ^a	0.129	0.054	0.003 ^a	0.008 ^a
TILs (P31)	0.045 ^a	0.250	0.002 ^a	0.015 ^a	0.051	0.033 ^a	0.154
TILs (P32)	0.246	0.154	0.058	<0.001 ^a	0.001 ^a	0.001 ^a	0.020 ^a
TILs (P33)	0.098	0.028 ^a	0.010 ^a	0.007 ^a	0.250	0.478	0.819
TILs (P34)	0.030 ^a	0.058	0.007 ^a	0.007 ^a	0.002 ^a	0.002 ^a	0.002 ^a
TILs (P35)	0.051	0.055	0.154	0.942	0.922	0.922	0.757
TILs (P36)	0.056	0.350	0.218	0.001 ^a	0.002 ^a	0.025 ^a	0.025 ^a
TILs (P37)	0.224	0.059	0.033 ^a	0.072	0.021 ^a	0.001 ^a	<0.001 ^a
TILs (P38)	§	0.141	0.028 ^a	0.021 ^a	0.036 ^a	0.054	0.133
TILs (P39)	0.154	0.003 ^a	<0.001 ^a	0.002 ^a	0.007 ^a	0.007 ^a	0.007 ^a
TILs (P40)	0.156	0.009 ^a	0.007 ^a	0.002 ^a	<0.001 ^a	<0.001 ^a	0.001 ^a

pCR Pathological complete response, *sTILs* Stromal tumor-infiltrating lymphocytes.

§ No *p* value was calculated as none of the sTILs scores was <5%.

^aStatistically significant result as determined by Chi-Square test.

be eligible for additional immune checkpoint blockade in the neoadjuvant setting. Foldi et al. recently reported promising results in a phase I/II trial, wherein PD-L1-positive TNBC were associated with higher pCR rates than PD-L1-negative TNBC, independent of the pre-NAC sTILs levels [49]. The GeparNuevo trial suggested similar results, as the addition of durvalumab before the start of anthracycline/taxane-based NAC seemed to

increase pCR rates in TNBC patients [50]. The International Immuno-Oncology Biomarker Working Group developed a risk management framework for the implementation of combined PD-L1 and TILs assessment in breast cancer [44], as several studies reported a strong correlation between PD-L1 positive immune cells and high sTILs levels [49, 51–54]. Biologically, TNBCs require infiltration by sTILs to be designated as PD-L1 positive.

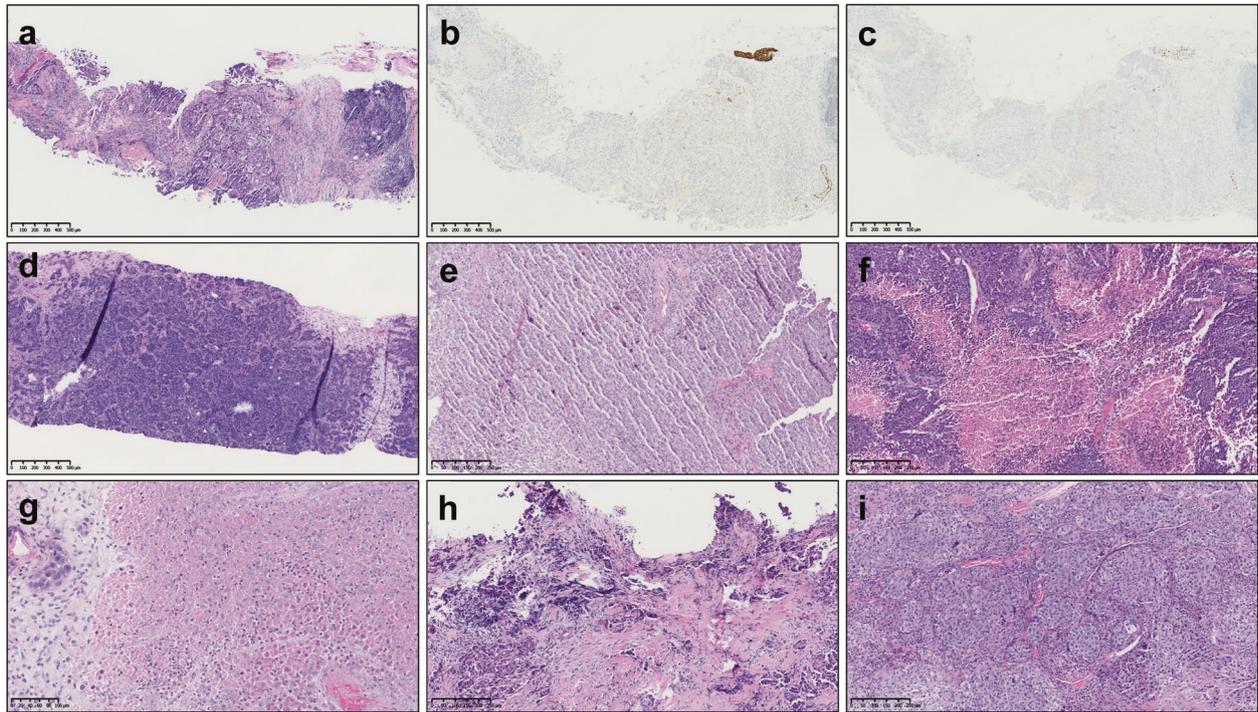


Fig. 6 Photomicrographs of TNBC biopsies. These images illustrate several potential pitfalls which can hamper sTILs assessment, such as DCIS-like TNBC with solid architecture (a–c), an overstained biopsy specimen with folds (d), section artifacts caused by a blunt microtome knife (e), extensive necrosis (f), extensive neutrophilic infiltration in necrotic areas (g), ample crush artifacts (h) and limited amounts of peri- and intra-tumoral stroma (i). Hematoxylin and eosin stainings (a, d–i); SMM-HC immunohistochemistry (b); p63 immunohistochemistry (c).

In conclusion, sTILs are a robust marker for pCR at the group level, despite substantial interobserver variability among pathologists. However, if sTILs are to be used to guide de-escalation of the NAC regimen in individual patients, interobserver discordance might significantly impact the chance of obtaining a pCR. Future studies should therefore explore the impact of training, as well as the 'ideal' sTILs threshold for dichotomization, as clinical decision-making will demand a particular cutoff. Although sTILs can be considered as a prognostic marker, there is currently insufficient evidence to modify NAC regimens based on pre-NAC sTILs levels. Intriguingly, patients with RCB-II and RCB-III in this cohort often had intermediate sTILs, which may suggest failing immune responses. Hence, future research should focus on fine-tuning patient selection for sTILs-based de-escalation of NAC regimens.

DATA AVAILABILITY

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

REFERENCES

- Nofech-Mozes S, Trudeau M, Kahn HK, Dent R, Rawlinson E, Sun P, et al. Patterns of recurrence in the basal and non-basal subtypes of triple-negative breast cancers. *Breast Cancer Res Treat.* 2009;118:131–7.
- van Maaren MC, de Munck L, Strobbe LJA, Sonke GS, Westenend PJ, Smidt ML, et al. Ten-year recurrence rates for breast cancer subtypes in the Netherlands: a large population-based study. *Int J Cancer.* 2019;144:263–72.
- Wang Y, Yin Q, Yu Q, Zhang J, Liu Z, Wang S, et al. A retrospective study of breast cancer subtypes: the risk of relapse and the relations with treatments. *Breast Cancer Res Treat.* 2011;130:489–98.
- Balkenhol MCA, Vreuls W, Wauters CAP, Mol SJJ, van der Laak JAWM, Bult P. Histological subtypes in triple negative breast cancer are associated with specific information on survival. *Ann Diagn Pathol.* 2020;46:151490.
- Korde LA, Somerfield MR, Carey LA, Crews JR, Denduluri N, Hwang ES, et al. Neoadjuvant chemotherapy, endocrine therapy, and targeted therapy for breast cancer: ASCO guideline. *J Clin Oncol.* 2021;39:1485–505.
- Bonnefoi H, Litière S, Piccart M, MacGrogan G, Fumoleau P, Brain E, et al. Pathological complete response after neoadjuvant chemotherapy is an independent predictive factor irrespective of simplified breast cancer intrinsic subtypes: a landmark and two-step approach analyses from the EORTC 10994/BIG 1-00 phase III trial. *Ann Oncol.* 2014;25:1128–36.
- Symmans WF, Peintinger F, Hatzis C, Rajan R, Kuerer H, Valero V, et al. Measurement of residual breast cancer burden to predict survival after neoadjuvant chemotherapy. *J Clin Oncol.* 2007;25:4414–22.
- Peintinger F, Sinn B, Hatzis C, Albarracín C, Downs-Kelly E, Morkowski J, et al. Reproducibility of residual cancer burden for prognostic assessment of breast cancer after neoadjuvant chemotherapy. *Mod Pathol.* 2015;28:913–20.
- Bossuyt V, Provenzano E, Symmans WF, Boughey JC, Coles C, Curigliano G, et al. Recommendations for standardized pathological characterization of residual disease for neoadjuvant clinical trials of breast cancer by the BIG-NABCG collaboration. *Ann Oncol.* 2015;26:1280–91.
- Denkert C, von Minckwitz G, Darb-Esfahani S, Lederer B, Heppner BI, Weber KE, et al. Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. *Lancet Oncol.* 2018;19:40–50.
- Denkert C, Von Minckwitz G, Brase JC, Sinn BV, Gade S, Kronenwett R, et al. Tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy with or without carboplatin in human epidermal growth factor receptor 2-positive and triple-negative primary breast cancers. *J Clin Oncol.* 2015;33:983–91.
- Van Bockstal MR, Noel F, Guiot Y, Duhoux FP, Mazzeo F, Van Marcke C, et al. Predictive markers for pathological complete response after neo-adjuvant chemotherapy in triple-negative breast cancer. *Ann Diagn Pathol.* 2020;49:151634.
- Ruan M, Tian T, Rao J, Xu X, Yu B, Yang W, et al. Predictive value of tumor-infiltrating lymphocytes to pathological complete response in neoadjuvant treated triple-negative breast cancers. *Diagn Pathol.* 2018;13:66.
- Hwang HW, Jung H, Hyeon J, Park YH, Ahn JS, Im YH, et al. A nomogram to predict pathologic complete response (pCR) and the value of tumor-infiltrating lymphocytes (TILs) for prediction of response to neoadjuvant chemotherapy (NAC) in breast cancer patients. *Breast Cancer Res Treat.* 2019;173:255–66.

15. Loi S, Michiels S, Salgado R, Sirtaine N, Jose V, Fumagalli D, 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.
16. Salgado R, Denkert C, Demaria S, Sirtaine N, Klauschen F, Pruneri G, 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.
17. Hendry S, Salgado R, Gevaert T, Russell PA, John T, Thapa B, et al. Assessing tumor-infiltrating lymphocytes in solid tumors: a practical review for pathologists and proposal for a standardized method from the international immunooncology biomarkers working group: part 1: assessing the host immune response, TILs in invasive breast carcinoma and ductal carcinoma in situ, metastatic tumor deposits and areas for further research. *Adv Anat Pathol.* 2017;24:235–51.
18. Buisseret L, Desmedt C, Garaud S, Fornili M, Wang X, Van Den Eyden G, et al. Reliability of tumor-infiltrating lymphocyte and tertiary lymphoid structure assessment in human breast cancer. *Mod Pathol.* 2017;30:1204–12.
19. Khoury T, Peng X, Yan L, Wang D, Nagrale V. Tumor-infiltrating lymphocytes in breast cancer: evaluating interobserver variability, heterogeneity, and fidelity of scoring core biopsies. *Am J Clin Pathol.* 2018;150:441–50.
20. Swisher SK, Wu Y, Castaneda CA, Lyons GR, Yang F, Tapia C, 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.
21. Kos Z, Roblin E, Kim RS, Michiels S, Gallas BD, Chen W, et al. Pitfalls in assessing stromal tumor infiltrating lymphocytes (sTILs) in breast cancer. *NPJ Breast Cancer.* 2020;6:17.
22. Tramm T, Di Caterino T, Jylling AMB, Lelkaitis G, Lænkholm AV, Ragó P, et al. Standardized assessment of tumor-infiltrating lymphocytes in breast cancer: an evaluation of inter-observer agreement between pathologists. *Acta Oncol.* 2018;57:90–94.
23. O’Loughlin M, Andreu X, Bianchi S, Chemielik E, Cordoba A, Cserni G, et al. Reproducibility and predictive value of scoring stromal tumour infiltrating lymphocytes in triple-negative breast cancer: a multi-institutional study. *Breast Cancer Res Treat.* 2018;171:1–9.
24. Allison KH, Hammond MEH, Dowsett M, McKernin SE, Carey LA, Fitzgibbons PL, et al. Estrogen and progesterone receptor testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists Guideline Update. *Arch Pathol Lab Med.* 2020;38:1346–66.
25. Wolff AC, Hammond MEH, Allison KH, Harvey BE, Mangu PB, Bartlett JMS, et al. Human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Focused Update. *J Clin Oncol.* 2018;36:2105–22.
26. Wilson AR, Marotti L, Bianchi S, Biganzoli L, Claassen S, Decker T, et al. The requirements of a specialist Breast Centre. *Eur J Cancer.* 2013;49:3579–87.
27. Dano H, Altinay S, Arnould L, Bletard N, Colpaert C, Dedeurwaerdere F, et al. Interobserver variability in upfront dichotomous histopathological assessment of ductal carcinoma in situ of the breast: the DCISion study. *Mod Pathol.* 2020;33:354–66.
28. Marée R, Rollus L, Stévens B, Hoyoux R, Louppe G, Vandaele R, et al. Collaborative analysis of multi-gigapixel imaging data using Cytomine. *Bioinformatics.* 2016;32:1395–401.
29. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;15:155–63.
30. Burstein HJ, Curigliano G, Loibl S, Dubsy P, Gnant M, Poortmans P, et al. Estimating the benefits of therapy for early-stage breast cancer: The St. Gallen International Consensus Guidelines for the primary therapy of early breast cancer 2019. *Ann Oncol.* 2019;30:1541–57.
31. Thomssen C, Balic M, Harbeck N, St Gnant M. Gallen/Vienna 2021: a brief summary of the consensus discussion on customizing therapies for women with early breast cancer. *Breast Care.* 2021;16:135–43.
32. Denkert C. Tumor infiltrating lymphocytes (TILs) as prognostic biomarker in patients with breast cancer. *Breast.* 2021;56:55.
33. Amgad M, Stovgaard ES, Balslev E, Thagaard J, Chen W, Dudgeon S, et al. Report on computational assessment of Tumor Infiltrating Lymphocytes from the International Immuno-Oncology Biomarker Working Group. *Npj Breast Cancer.* 2020;6:16.
34. Leung SCY, Nielsen TO, Zabaglo LA, Arun I, Badve SS, Bane AL, et al. Analytical validation of a standardised scoring protocol for Ki67 immunohistochemistry on breast cancer excision whole sections: an international multicentre collaboration. *Histopathology.* 2019;75:225–35.
35. Polley MY, Leung SC, McShane LM, Gao D, Hugh JC, Mastropasqua MG, et al. An international Ki67 reproducibility study. *J Natl Cancer Inst.* 2013;105:1897–906.
36. Denkert C, Loibl S, Müller BM, Eidtmann H, Schmitt WD, Eiermann W, et al. Ki67 levels as predictive and prognostic parameters in pretherapeutic breast cancer core biopsies: A translational investigation in the neoadjuvant gepartrio trial. *Ann Oncol.* 2013;24:2786–93.
37. Cha YJ, Ahn SG, Bae SJ, Yoon CI, Seo J, Jung WH, et al. Comparison of tumor-infiltrating lymphocytes of breast cancer in core needle biopsies and resected specimens: a retrospective analysis. *Breast Cancer Res Treat.* 2018;171:295–302.
38. Althobiti M, Aleskandarany MA, Joseph C, Toss M, Mongan N, Diez-Rodriguez M, et al. Heterogeneity of tumour-infiltrating lymphocytes in breast cancer and its prognostic significance. *Histopathology.* 2018;73:887–96.
39. Luen SJ, Salgado R, Dieci MV, Vingiani A, Curigliano G, Gould RE, et al. Prognostic implications of residual disease tumor-infiltrating lymphocytes and residual cancer burden in triple-negative breast cancer patients after neoadjuvant chemotherapy. *Ann Oncol.* 2019;30:236–42.
40. Pinard C, Debléd M, Ben Rejeb H, Velasco V, Tunon de Lara C, Hoppe S, et al. Residual cancer burden index and tumor-infiltrating lymphocyte subtypes in triple-negative breast cancer after neoadjuvant chemotherapy. *Breast Cancer Res Treat.* 2020;179:11–23.
41. García-Martínez E, Gil GL, Benito AC, González-Billalabeitia E, Conesa MAV, García TG, et al. Tumor-infiltrating immune cell profiles and their change after neoadjuvant chemotherapy predict response and prognosis of breast cancer. *Breast Cancer Res.* 2014;16:488.
42. Filho OM, Stover DG, Asad S, Ansell PJ, Watson M, Loibl S, et al. Association of immunophenotype with pathologic complete response to neoadjuvant chemotherapy for triple-negative breast cancer. *JAMA Oncol.* 2021;7:603–8.
43. Sinn BV, Loibl S, Hanusch CA, Zahm DM, Sinn H-P, Untch M, et al. Immune-related gene expression predicts response to neoadjuvant chemotherapy but not additional benefit from PD-L1 inhibition in women with early triple-negative breast cancer. *Clin Cancer Res.* 2021;27:2584–91.
44. Gonzalez-Ericsson PI, Stovgaard ES, Sua LF, Reisenbichler E, Kos Z, Carter JM, et al. The path to a better biomarker: application of a risk management framework for the implementation of PD-L1 and TILs as immuno-oncology biomarkers in breast cancer clinical trials and daily practice. *J Pathol.* 2020;250:667–84.
45. Cimino-Mathews A. Novel uses of immunohistochemistry in breast pathology: interpretation and pitfalls. *Mod Pathol.* 2021;34:62–77.
46. Schmid P, Cortes J, Pusztai L, McArthur H, Kümmel S, Bergh J, et al. Pembrolizumab for early triple-negative breast cancer. *N. Engl J Med.* 2020;382:810–21.
47. Reisenbichler ES, Han G, Bellizzi A, Bossuyt V, Brock J, Cole K, et al. Prospective multi-institutional evaluation of pathologist assessment of PD-L1 assays for patient selection in triple negative breast cancer. *Mod Pathol.* 2020;33:1746–52.
48. Hoda RS, Brogi E, D’Alfonso TM, Grabenstetter A, Giri D, Hanna MG, et al. Inter-observer Variation of PD-L1 SP142 Immunohistochemistry Interpretation in Breast Carcinoma: A Study of 79 Cases Using Whole Slide Imaging. *Arch Pathol Lab Med.* 2021. <https://doi.org/10.5858/arpa.2020-0451-0a>.
49. Foldi J, Silber A, Reisenbichler E, Singh K, Fischbach N, Persico J, et al. Neoadjuvant durvalumab plus weekly nab-paclitaxel and dose-dense doxorubicin/cyclophosphamide in triple-negative breast cancer. *NPJ Breast Cancer.* 2021;7:9.
50. Loibl S, Untch M, Burchardi N, Huober J, Sinn BV, Blohmer JU, et al. A randomised phase II study investigating durvalumab in addition to an anthracycline taxane-based neoadjuvant therapy in early triple-negative breast cancer: Clinical results and biomarker analysis of GeparNuevo study. *Ann Oncol.* 2019;30:1279–88.
51. Emens LA, Molinero L, Loi S, Rugo HS, Schneeweiss A, Diéras V, et al. Atezolizumab and nab-Paclitaxel in Advanced Triple-Negative Breast Cancer: Biomarker Evaluation of the IMpassion130 Study. *J Natl Cancer Inst.* 2021. <https://doi.org/10.1093/jnci/djab004>.
52. Noske A, Möbus V, Weber K, Schmatloch S, Weichert W, Köhne CH, et al. Relevance of tumour-infiltrating lymphocytes, PD-1 and PD-L1 in patients with high-risk, nodal-metastasised breast cancer of the German Adjuvant Intergroup Node-positive study. *Eur J Cancer.* 2019;114:76–88.
53. Dieci MV, Tsvetkova V, Griguolo G, Miglietta F, Tascia G, Giorgi CA, et al. Integration of tumour infiltrating lymphocytes, programmed cell-death ligand-1, CD8 and FOXP3 in prognostic models for triple-negative breast cancer: Analysis of 244 stage I–III patients treated with standard therapy. *Eur J Cancer.* 2020;136:7–15.
54. Wimberly H, Brown JR, Schalper K, Haack H, Silver MR, Nixon C, et al. PD-L1 expression correlates with tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy in breast cancer. *Cancer Immunol Res.* 2015;3:326–32.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the help of Mr. Sébastien Godecharles with digitalizing the HE slides used in this study. MRVB received a postdoctoral mandate (grant number 2019–089) from the not-for-profit organization Fondation against Cancer (Brussels, Belgium), and is supported by the “Fonds dr. Gaëtan Lagneaux” of the Fondation Saint-Luc (Brussels, Belgium). OB has received personal consultancy fees from Roche, outside the scope of the present work. GF received a postdoctoral

mandate from the Klinisch Onderzoek en Opleidingsraad (KOOR) of the University Hospitals Leuven. CM has received personal consultancy fees from Roche, Bayer, Astrazeneca and Daiichi Sankyo, outside the scope of the present work. HW is supported by the Memorial Sloan Kettering Cancer Center Support Grant/Core Grant (P30 CA008748), awarded by the National Cancer Institute. CG is supported by the “Fonds dr. Gaëtan Lagneaux” of the Fondation Saint-Luc (Brussels, Belgium).

AUTHOR CONTRIBUTIONS

MRVB and CG performed study concept and design. With the exception of CB, all authors provided acquisition of the data, including histopathological evaluation and completion of the questionnaire. MRVB and AF developed the methodology, performed the statistical analysis and interpreted the data. CB provided technical and material support. MRVB wrote the first draft of the manuscript. AF, CB, and CG revised the paper. All authors reviewed, read and approved the final paper.

COMPETING INTERESTS

The authors declare no competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study was conducted in accordance with the Declaration of Helsinki and was approved by the ethics committee of the Cliniques universitaires Saint-Luc (Brussels, Belgium). The need for informed consent was waived.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41379-021-00865-z>.

Correspondence and requests for materials should be addressed to M.R.V.B.

Reprints and permission information is available at <http://www.nature.com/reprints>

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

¹Department of Pathology, Cliniques universitaires Saint-Luc Bruxelles, Woluwé-Saint-Lambert, Belgium. ²Department of Pathology, University of Health Sciences, Bakirköy Dr. Sadi Konuk Health Application and Research Center, Istanbul, Turkey. ³Département de Biologie et de Pathologie des Tumeurs, Centre George-François Leclerc, Dijon, France. ⁴Department of Pathology, Radboud University Medical Center, HB, Nijmegen, The Netherlands. ⁵Department of Pathology, University Hospital Antwerp, Edegem, Belgium. ⁶Department of Pathology, Hospital Clínico Universitario de Valencia, València, Spain. ⁷Department of Pathology, AZ Turnhout Campus Sint-Jozef, Turnhout, Belgium. ⁸Department of Pathology, AZ Delta, Roeselare, Belgium. ⁹Division of Pathology and Laboratory Medicine, Medical School, The University of Western Australia, Crawley, WA, Australia. ¹⁰Anatomical Pathology, PathWest Laboratory Medicine WA, Perth, Australia. ¹¹Department of pathology, AZ Kiina Brasschaat, Brasschaat, Belgium. ¹²Department of Pathology, University Hospitals Leuven, KU Leuven – University of Leuven, Leuven, Belgium. ¹³Department of Imaging and Pathology, Laboratory of Translational Cell & Tissue Research, KU Leuven – University of Leuven, Leuven, Belgium. ¹⁴Department of Pathology, Peter MacCallum Cancer Center and the University of Melbourne, Melbourne, VIC, Australia. ¹⁵Department of Pathology, University of Cagliari, AOU San Giovanni di Dio, Cagliari, Italy. ¹⁶Institute of Pathology, Lausanne University Hospital, Lausanne, Switzerland. ¹⁷Department of Pathology, Mount Sinai Hospital and Icahn School of Medicine, New York, NY, USA. ¹⁸Pathan BV, PM, Rotterdam, The Netherlands. ¹⁹Department of Pathology, Gustave-Roussy Cancer Campus, Villejuif, France. ²⁰Department of Pathology, Clinique Notre-Dame de Grâce (CNDG), Gosselies, Belgium. ²¹Department of Pathology, AZ St Lucas Hospital, Ghent, Belgium. ²²Surgical Pathology Unit, Department of Pathobiology, Institut Bergonié, Bordeaux, France. ²³Department of Medical Sciences, University of Turin, Torino, Italy. ²⁴Pathology Unit, FPO-IRCCS, Candiolo Cancer Institute, Candiolo, Italy. ²⁵Institut de Pathologie et de Génétique, Charleroi, Belgium. ²⁶Department of Laboratory Medicine and Molecular Diagnostics, Sunnybrook Health Sciences Center, University of Toronto, Toronto, ON, Canada. ²⁷Department of Pathology, AZ St Maarten, Mechelen, Belgium. ²⁸Histopathology, Imaging and Quantification Unit, HistoGeneX, Antwerp, Belgium. ²⁹Department of Pathology, Yale School of Medicine, Yale New Haven Hospital, New Haven, USA. ³⁰The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ³¹Department of Pathology and Lab Medicine, Cedars-Sinai Medical Center, Los Angeles, USA. ³²Department of Pathology, Onze-Lieve-Vrouwziekenhuis Aalst, Aalst, Belgium. ³³Department of Cellular Pathology, Queen Elizabeth Hospital Birmingham, University of Birmingham, Birmingham, UK. ³⁴Department of Pathology, CHU UCL Namur, Site Godinne, Yvoir, Belgium. ³⁵Department of Pathology, Erasmus Medical Center Rotterdam, Rotterdam, The Netherlands. ³⁶Department of Pathology, Ghent University Hospital, Ghent, Belgium. ³⁷Pôle de Médecine Diagnostique & Théranostique, INSERM U934, Institut Curie, Paris Cedex 05, France. ³⁸Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA. ³⁹ZIP IREC Imaging Platform, Institute of Clinical and Experimental Research (IREC), Université catholique de Louvain, Brussels, Belgium. ⁴⁰Institute of Clinical and Experimental Research (IREC), Université catholique de Louvain, Brussels, Belgium. ⁴¹These authors contributed equally: Serdar Altınay, Laurent Arnould, Maschenka Balkenhol, Glenn Broeckx, Octavio Burguès, Cecile Colpaert, Franceska Dedeurwaerdere, Benjamin Dessauvage, Valérie Duwel, Giuseppe Floris, Stephen Fox, Clara Gerosa, Delyne Hastir, Shabnam Jaffer, Eline Kurpershoek, Magali Lacroix-Triki, Andoni Laka, Kathleen Lambein, Gaëtan Marie MacGrogan, Caterina Marchiò, Maria-Dolores Martin Martinez, Sharon Nofech-Mozes, Dieter Peeters, Alberto Ravarino, Emily Reisenbichler, Erika Resetkova, Souzan Sanati, Anne-Marie Schelfhout, Vera Schelfhout, Abeer Shaaban, Renata Sinke, Claudia M. Stanciu-Pop, Carolien H. M. van Deurzen, Koen K. Van de Vijver, Anne-Sophie Van Rompuy, Anne Vincent-Salomon, Hannah Y. Wen, Serena Wong. [✉]email: mieke.vanbockstal@saintluc.uclouvain.be