Incorporation of TILs in Daily Breast Cancer Care: How Much Evidence can we Bear?

INVITED REVIEW

Incorporation of TILs in Daily Breast Cancer Care: How Much Evidence can we Bear?

Anne-Vibeke Laenkholm[1], Grace Callagy[2], Marcelo Balancin[3], John M. S. Bartlett[4], Christos Sotiriou[5], Caterina Marchio[6] Marleen Kok[7], Carlos Henrique Dos Anjos[8], Roberto Salgado[9]

[1]Department of Surgical Pathology, Zealand University Hospital, Roskilde, Denmark.

[2]Discipline of Pathology, NUI Galway, Lambe Institute for Translational Research, Galway, Ireland.

[3]Department of Pathology, University of São Paulo Medical School (USP), São Paulo, Brazil.

[4] Cancer Research UK Edinburgh Centre, Institute of Genetics and Cancer, The University of Edinburgh, Edinburgh, UK.

[5] Medical Oncology Department, Institut Jules Bordet, Université Libre de Bruxelles, Brussels, Belgium.

[6] Division of Pathology, Candiolo Cancer Institute FPO-IRCCS, Candiolo, Italy; Department of Medical Sciences, University of Turin, Turin, Italy.[7] Departments of Medical Oncology and Tumor Biology and Immunology, The Netherlands Cancer Institute, Amsterdam, The Netherlands.

[8] Oncology Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, US.

[9]Department of Pathology, GZA-ZNA Hospitals, Antwerp, Belgium; Division of Research, Peter Mac Callum Centre, Melbourne, Australia.

Corresponding author

Roberto Salgado

Department of Pathology, GZA-ZNA Hospitals, Antwerp, Belgium; Division of Research, Peter Mac Callum Centre, Melbourne, Australia.

roberto@salgado.be

Word-count: 5418

Key words: TILs, Breast Cancer, Prognosis, Immunotherapy.

Abstract

One of the most important developments in the breast cancer field has been an improved understanding of prognostic and predictive biomarkers, of which TILs are increasingly gaining importance. The evaluation of TILs by light microscopy on a H&E-stained section is workable in a daily practice setting. Reproducibility of reporting TILs is good, but heterogeneity is a cause of variation. TILs provide clinicians with important prognostic information for patients with TNBC, as early stage TNBC with high TILs have >98% 5-year survival and TILs predict benefit to immunotherapy. Importantly, while TILs do not have level of evidence IA, TILs should be used as a prognostic factor with caution and with other accepted prognostic variables, such as tumor size and lymph node status, to inform clinicians and patients on their treatment options. A framework on how to use the TILs in daily practice is proposed, including a co-assessment with PD-L1 for its predictive role to immunotherapy.

Introduction

One of the most important developments in the breast cancer field has been a better comprehension of prognostic and predictive biomarkers. By definition, a prognostic factor can provide information on clinical outcome at the time of diagnosis, independent of therapy, thus reflecting the natural history of the tumour [1]. By contrast, a predictive factor provides information on response to a given treatment. Several biomarkers in breast cancer, such as ER and HER2, are both prognostic and predictive.

Triple negative breast cancer (TNBC) is an aggressive subtype of breast cancer that represents approximately 15% to 20% of all cases. It is associated with a poor prognosis in both the early and advanced settings, in part because of fewer systemic treatment options as compared with estrogen receptor–positive or HER2–enriched breast cancer. There are no validated prognostic biomarkers available for TNBC other than tumour size, lymph node involvement and pathological complete response (pCR) to neoadjuvant chemotherapy and, as we will argue, the strength of the immune system. TNBC is considered the most immunogenic breast cancer subtype, with a higher median number of tumor-infiltrating lymphocytes (TILs) and PD-L1 expression - both markers associated with tumor microenvironment (TME) immune activity. In the last decade, much evidence has accumulated on the importance of the immune system as a determinant of outcome and response to therapy in TNBC. This review provides an overview of the immune response in TNBC, focusing on TILs and detailing an important role for the pathologist. The assessment of TILs in the studies mentioned in this review was performed using the Guidelines from the International Immuno-Oncology Biomarker Working Group (WG), also called the TIL WG (www.tilsinbreastcancer.org) [2].

The immune Microenvironment and Breast Cancer

Tumor infiltrating lymphocytes.

TILs are mononuclear immune cells that infiltrate tumor tissue and have been described in almost all solid tumours, including breast cancer (Figure 1 till 3). The relationship between the mere quantity of TILs and outcome in breast cancer was first reported in 1922 [3]. The relationship between higher levels of TILs and improved prognosis in patients with early stage TNBC has now been confirmed in >25000 patients that includes many prospective-retrospective phase 3-, phase 2- and phase 1-studies [4]. Histologic evaluation of TILs has now reached level IB-evidence as a prognostic marker in TNBC [5].

Furthermore, the current predictive evidence of TILs in the neoadjuvant setting is Level of Evidence 2A, GR level B [6].

Expert committees such as St Gallen Breast Cancer Expert Committee and European Society for Medical Oncology (ESMO) recognize the prognostic importance of TILs with the caveat that TILs should not direct treatment as an independent variable [7, 8]. These suggest, as will be explained later, that TILs should be used together with other prognostic variables like tumor size and lymph node to give clinicians all the prognostic information required to discuss treatment options with their patients. Nevertheless, a careful approach is recommended until biomarker level IA evidence is reached for TILs.

In this context, a careful appreciation of the definitions of levels of evidence by clinicians and pathologists is needed for the correct use of biomarkers in daily practice. The levels of evidence are defined as follows: level IA "Prospective randomized controlled trials designed to address the tumor marker utility", and level IB as "Prospective trials not designed to address the tumor marker, but the design accommodates tumor marker utility". For a predictive marker, the trial must be a randomized controlled trial, with ≥ 1 validation study in order to obtain level of evidence IB [9, 10].

Clinical significance of sTILs

TILs as a prognostic biomarker in triple negative breast cancer

An interesting way to define the potential of a prognostic biomarker in the natural history of the tumour is a systematically untreated patient population only managed with local treatment (surgery +/- adjuvant radiotherapy). De Jong and colleagues correlated TILs with patient outcome in early-stage, node-negative TNBC in patients younger than 40-years who did not get any (neo)-adjuvant systemic treatment [11]. The study included 481 patients of whom 90% had pT1c or pT2 tumours. TILs < 30% were found in 51% of cases and TILs of 30-75% and ≥ 75% were present in 26% and 22% of cases, respectively [11]. TILs were associated with overall survival (OS) as well as distant metastasis free survival at 15-years follow-up. Patients with sTILs ≥ 75% (n = 107 patients) had a distant recurrence rate of only 1.9% at 15-years follow-up, while patients with sTILs 30-75% (n = 127 patients) and < 30% (n = 247 patients) had a distant recurrence rate of 16% and 39%, respectively [11]. In another systemically untreated patient population study, Park and colleagues evaluated the correlation between TILs and patient outcome in a pooled analysis of nearly 500 early-stage TNBC patients [12]. TILs were significantly associated with outcome, and in the pathologic stage I subpopulation, according to the AJCC 8th edition [13], an excellent survival was

observed in patients with TILs > 30%, with an estimated 5-year invasive disease-free survival (iDFS) of 91%, distant disease-free survival (DDFS) of 97%, and OS of 98% [12]. Both studies suggest that in early-stage TNBC, there is a subset of patients for whom, based on the strength of the immune response as exemplified by TILs, the added benefit of chemotherapy is very limited [11, 12]. This information may be important for clinicians as well as for patients.

In the adjuvant setting, a pooled analysis of several randomized adjuvant clinical trials of early-stage TNBC with more than 2000 patients also showed the strong prognostic importance of TILs [5]. By multivariable analysis, TILs added independent prognostic information for all endpoints, including iDFS, DDFS and OS. Each 10% increment in TILs corresponded to an iDFS hazard ratio (HR) of 0.87 (95% CI, 0.83 to 0.91) for iDFS, 0.83 (95% CI, 0.79 to 0.88) for DDFS, and 0.84 (95% CI, 0.79 to 0.89) for OS [5]. As for the untreated patient population, patients with node-negative disease, T1/T2 tumors and TILs ≥ 30% had excellent outcomes with a 3-year iDFS of 92%, DDFS 97% and OS 99% [5]. A combined analysis of TILs and prognostic stage in this same patient population confirm that TILs can up- or downgrade the prognostic stage of patients with TNBC (Figure 5) [14].

Considering the level IB evidence for the prognostic value of TILs, the expert panels at St Gallen 2019, ESMO and the 2019 edition of the World Health Organization (WHO) Classification of Tumors of the Breast endorsed the prognostic information of TILs in TNBC, with a caution that it should not be used as an independent variable in daily practice [7, 8, 15, 16]. However, at St. Gallen 2021 the data on TILs were still considered inadequate for the purposes of choosing specific chemotherapy regimens and deciding whether to withhold chemotherapy treatment or not. Therefore, it is currently not recommended that clinicians change their treatment decision based on the TILs ***only***, as level IA evidence would be needed for this. So, why then should pathologists score TILs in their daily practice?

TILs as a predictive biomarker for chemotherapy benefit in triple negative breast cancer

A pCR to neoadjuvant chemotherapy, defined as absence of invasive disease in breast and lymph nodes, has been proposed as a surrogate endpoint for long-term clinical benefit, such as DFS and OS in early stage TNBC [17]. Several meta-analyses have shown that patients with a pCR have lower disease recurrence and lower breast cancer specific mortality [18-20].

Several studies have documented the association between TILs and the likelihood of achieving pCR in early stage TNBC treated with neoadjuvant chemotherapy. For example, Denkert and colleagues evaluated 906 patients with early stage TNBC treated with neoadjuvant chemotherapy in 6 randomized

trials performed by the German Breast Cancer (GBG) Group [21]. Increased TILs on the core-biopsies were linked to increased pCR. A pCR was achieved in 80 (31%) of 260 patients with low TILs (0-10%), 117 (31%) of 373 patients with intermediate TILs (11-59%) and in 136 (50%) of 273 patients with high TILs (≥60%). The correlation between TILs and pCR was independent of the chemotherapy regimen. Neoadjuvant studies in early-stage TNBC that have used platin agents, anthracyclines, taxanes or 5-FU, have all shown similar associations between TILs and pCR [22, 23].

Conceptually, pCR is an important endpoint for the high-risk TNBC population for whom attaining a pCR is important for patient outcome. In those patients who have tumours with high TILs, less systemic chemotherapy may be needed to achieve a pCR [24].

TILs as a predictor biomarker for immune therapy or combined Immune therapy/chemotherapy benefit in triple negative breast cancer

Previous phase 1b/2 studies suggest that TILs predict pCR to neoadjuvant immune therapy combined with chemotherapy in early stage TNBC [25, 26]. Baseline TILs were predictive of pCR in both the durvalumab (anti-PD-L1) plus chemotherapy and chemotherapy plus placebo arms of GeparNuevo [25]. In addition, overall T-cell density was associated with pCR in response to pembrolizumab in the randomized phase II I-SPY 2 trial that added pembrolizumab to chemotherapy [26]. In addition, KEYNOTE-173, a multicohort phase 1b study evaluating the safety and preliminary antitumor activity of neoadjuvant chemotherapy plus pembrolizumab in high-risk, early-stage TNBC, showed that the pCR rate had a significant correlation with TILs measured at baseline and on-treatment [27]. Hence stratifying patients into different treatments arms based on TILs may become important, as conceptually less intense chemotherapy may be an option for those patients with high TILs and the converse would be the case for those with low TILs.

In the metastatic setting, atezolizumab has received conditional FDA-approval for patients with PD-L1 positive advanced/metastatic TNBC, based on results of the Impassion130-trial [28]. However, the Impassion131-trial did not confirm these findings [29], which has led Genentech to voluntary withdraw its application for accelerated approval of the drug for use in metastatic TNBC in the US [30]. An exploratory biomarker analysis of the Impassion130 trial showed that the HR for benefit of atezolizumab for the PD-L1 immune positivity/any TILs patient population (Progression Free Survival (PFS)), 0.65 and OS, 0.71) was very similar to the HR for any PD-L1 immune-positivity and TILs >10% patient population (PFS, 0.64 and OS, 0.75) [31] . This suggests that the PD-L1-inhibition driven benefit is mainly determined

by the number of immune cells. This is not surprising, as all PD-L1 assays in breast cancer rely on the number of immune cells within the tumour, which are mainly the TILs [28, 31].

Recently, the phase III KN355 trial of metastatic TNBC showed an improved PFS and OS for patients treated with pembrolizumab in combination with chemotherapy with a CPS>10 compared to those treated with chemotherapy with placebo. The randomized phase III KEYNOTE-119 trial, a study that randomized metastatic TNBC patients, in the second- or third-line therapy to chemotherapy or single drug pembrolizumab, did not show an overall survival benefit of pembrolizumab over chemotherapy in the pre-specified population of patients with PDL1 positivity with combined positive score (CPS) > 1 or CPS > 10, the primary endpoints of this study. An exploratory analysis of this same study showed that patients with TILs >5% had better OS with pembrolizumab monotherapy arm over chemotherapy [32]. The results of this trial suggest that TILs predict survival benefit for pembrolizumab, over chemotherapy in this group of patients - a benefit that was not predicted by PD-L1 CPS with cutoffs of > 1 or > 10.

In the phase III Keynore-522 trial, in the neo-adjuvant setting, the frequency of pathological complete response was significantly higher (64.8%) among those who received pembrolizumab and chemotherapy compared to those who received placebo and chemotherapy (51.2%). In addition, a recent analysis showed that the immunotherapy-arm had a 37% reduction in EFS-events: the 3-year EFS-rate was 84.5% compared with 76.8% with chemotherapy alone [33]. A TIL-analysis in both KN522 and KN355 is ongoing and will be very informative for the scientific community as well as for future trial designs [33, 34].

Taken together, these data suggest that the level of TILs gives predictive information of benefit from immune check point inhibitors (ICI) and ICI combined with chemotherapy for patients with advanced/metastatic TNBC. Consequently, clinical trials evaluating the benefit of ICI in TNBC have started to include TILs levels as a stratification factor [35] Considering the predictive importance of TILs in KEYNOTE-119 as well as Impassion130, may TILs be considered to have level of evidence IB as a predictive marker for immunotherapy in the metastatic TNBC-setting? If yes, should TILs be used to select patients for immunotherapy?

Can TILs help in the real-world evaluation of PD-L1 assays in triple negative breast cancer?

The two randomized phase 3 trials, the IMpassion130 and Keynote-355 have demonstrated treatment benefit with ICI in TNBC [28, 34]. Both trials included PD-L1 as an inclusive biomarker, and both

used different assays. In IMpassion130, the Ventana PD-L1 (SP142)-assay was used and KEYNOTE-355 used the PD-L1 IHC 22C3 pharmDx-assay [28, 34].

Currently, PD-L1 assays are used with different cut-points and quantification algorithms depending on the type of cancer and choice of ICI [36]. PD-L1 expression varies according to the immunohistochemical protocol, including the visualization system and platform used, with the SP142-assay having a documented lower sensitivity compared to other PD-L1 assays [37, 38]. There are also concerns about the reproducibility of PD-L1 reporting by pathologists [39, 40]. For PD-L1 assays to be "interchangeable", it would be necessary for the specific assay to produce the same clinical outcomes reported in the trial [41]. Unfortunately, none of the assay comparisons have been performed in the setting of a prospective clinical trial, only as exploratory post-hoc analyses [42, 43]. Performing these types of comparative analysis ***before*** an assay is co-approved with a drug by FDA is considered crucial for the correct interpretation of trial results in which an assay is used (see later).

In the IMpassion130, a post-hoc biomarker analysis [42] showed a significantly improved PFS for all investigated assays: with HRs for SP142 of 0.60 (95% CI, 0.47-0.78); for 22C3 of 0.68 (95% CI, 0.56-0.82); and for SP263 of 0.64 (95% CI, 0.53 - 0.79). The results for OS were also nearly similar for the three assays SP142 (HR 0.74, 95% CI, 0.54 - 1.01); 22C3 (HR 0.78, 95% CI, 0.62 - 0.99); and SP263 (HR, 0.75 (95% CI, 0.59- 0.96). Thus, a treatment benefit with ICI was observed with all three assays, as the HRs were comparable. Additional subgroup analyses combining several assays together, indicated that the treatment effect was most significant in the group of SP142 positive patients. However, this interpretation needs to be taken with much caution, as conceptually combining different assays with different sensitivities in the context of the small size of the subgroups may lead to underpowered interpretations.

The HR associated with benefit to a therapy and the median OS are not the most appropriate methods to make inferences on the performance of an assay. The predictive value of an assay is more appropriately examined by performing a treatment-biomarker interaction analysis between the biomarker-positive and biomarker-negative population.

Thus, a reasonable conclusion is that all PD-L1-assays predict benefit to ICI in the advanced/metastatic TNBC-setting to some extent. In general, it is unknown what the impact on the drug-label would be if a drug-company would formally state that other biomarkers may perform comparably to the FDA-approved Companion Diagnostic of the approved drug. It is important that biomarker-analysis of trials that were used to develop any Companion Diagnostic Assay inform in an unbiased manner on the performance of different assays used to detect the biomarker. [44].

Currently, FDA-approved assays are only meticulously analyzed by the pathology community in terms of their analytical performance ***after*** the formal FDA-approval of the companion diagnostic used in the trial in which the drug was approved. When a trial is considered "positive", it is often assumed that the assay used in that trial should be the assay of choice for worldwide use. This may not necessarily be true or workable. Recently, an expert panel of representatives of international pathology organizations have described why this current assay-approval narrative is not helping our patients [45]. An urgent revision of the current narrative of assay- and combined drug-approval is needed and should be based on the concept that a clinical trial serves to validate a biomarker, and not an assay of a particular vendor. PD-L1 is a reliable predictive biomarker but there is confusion surrounding these assays probably being one of the reasons why confirmatory trials such as Impassion131 did not confirm the findings of the Impassion130 trial that was used for the accelerated approval of the assay and the drug. This limits the potential use of ICI in TNBC [46, 47]. This situation is avoidable. Solutions to solve the current confusion on assays are proposed in Table 1.

In the real-world setting not all laboratories will use the SP142-assay. Indeed, when the KEYNOTE 355-regimen gets approved by the regulatory authorities, both the SP142-assay and the 22C3-assay will potentially be in use with different platforms and scoring systems required, which is a non-workable situation for most pathology laboratories. Acknowledging that all PD-L1 assays in breast cancer rely mostly on TILs, the TILs-WG have proposed that TILs and PD-L1 should be evaluated together, and this would streamline assessment [36]. In this context, if there are no immune cells, as in Figure 2, any PD-L1 assay will be negative; if there are many TILs, all PD-L1 assays will probably be positive. As an illustration of this concept, Figure 3 is a case from daily practice that shows a high level of TILs but was negative for SP142 (score 0). This patient, despite having a negative PD-L1 assay, could probably respond to immune therapy if the disease were advanced/metastatic. In addition, PD-L1 may be positive if another assay is used. What would you or your oncologist do? 1. Nothing?; 2. Reconsider the staining?; 3. Consider another assay?

Taking everything together, the question of PDL1 assay interchangeability unavoidably arises, given the difference in scoring methods, combined with the analytical differences between assays. Unfortunately, each randomized controlled trial for one of the checkpoint inhibitors is prospectively linked with only one assay rather than the biomarker. Retrospective analyses linking outcome data with other assays are seen as susceptible to bias. So, from a strict evidenced-based approach, the assays are not interchangeable, as confirmed by several studies [41, 48, 49]. However, it must be acknowledged that practical considerations may preclude maintenance of drug-diagnostic pairings in the real world. Although

far from perfect, there is a high degree of analytical correlation between most of the various assays. Although this imperfect correlation theoretically might compromise the predictive value of an assay that has been specifically validated for a particular drug, an evidence bases to assess this linkage is largely lacking. The practice of medicine often requires the use of judgment and experience in the absence of an evidence base, and this is no exception. Critically all these trials validate the use of a biomarker, PDL1, to target a drug in the class of checkpoint inhibitors. The principle of biomarker targeted therapy is therefore confirmed in multiple trials. Ultimately, the responsibility for selecting the most appropriate PDL1 assay should rest with the pathologist who weighs the available evidence with local practical considerations in making that decision.

Methodology for Evaluating TILs and Reproducibility of Reporting

The guidance developed by the TIL-WG is the recommended methodology for quantifying TILs in breast cancer [2, 16]. It recommends that TILs are assessed on routine H&E slides by light microscopy, mainly in treatment naive BC. Only TILs within the tumour stroma (stromal TILs) and within the tumour boundary are quantified as a percentage of the intra-tumoural stromal area and is reported as a continuous measurement. Quantification of sub-populations of TILs by immunohistochemistry in daily practice is not recommended at present, but immunohistochemistry can of course be used if the TILs are not easily recognized, for example in small and partly crushed biopsies. The TIL-WG has led the efforts to standardize and validate reporting TILs. The Group has developed a comprehensive training and educational resource around TILs in cancer that is accessible on its website (www.tilsinbreastcancer.org). Their site includes videos (TILs Education: What They Are and What They Do - YouTube) and demonstrations on how to score TILs in different settings, an interactive TIL scoring tool (Login - TILs Training Tool - International TILS Working Group (virtuelle-mikroskopie.de)) and a repository of images of the range of pitfalls and static calibrated reference images (Home - International TILS Working Group (tilsinbreastcancer.org)) that can aid the pathologist scoring TILs in daily practice. A tutorial on how to score TILs can also be accessed HERE (https://www.dropbox.com/s/dqule4ru6b122uv/Tutorial_Virchows%20Archive_2021_Rebuttal.pptx?dl=0).

TILs assessment by light microscopy is straightforward in that it requires no additional testing other than a routine H&E but as for all our morphological biomarkers it is subject to variability, especially in the selection of tumour areas to score by the pathologist. Several studies have examined the

reproducibility of TILs assessment (Table 3). These have in general reported good inter-observer agreement for TILs as a continuous variable with intraclass coefficient (ICC) values ranging from 0.66 to 0.89 [50-53]. The largest reproducibility studies reported to date were conducted by the TIL-WG in which good inter-observer agreement (ICC 0.70, 95% CI 0.62-0.78) was achieved between 34 pathologists that impressively improved to excellent (ICC 0.89, 95% CI 0.85-0.92) when the interactive scoring aid developed by the TIL-WG was used for assessment [51]. Most of these studies evaluated TILs in core biopsies of TNBCs but a similar level of agreement is reported in resection specimens [52] and across all breast cancer subtypes [53] . Agreement for TILs reported in categories defined by cutpoints is generally less than for continuous measurements with moderate agreement across a range of values reported in several studies [50, 51, 53, 54]. However, substantial concordance was achieved in three TIL WG RING studies at five cutpoint values from 1% to 75% [51, 55] which is encouraging in the context of data in TNBC cohorts demonstrating strong prognostic significance for TILs above a threshold of 30% [5] .

Notwithstanding good overall agreement reported in the reproducibility studies, the range of scores for many cases was wide with frequent outlier scores observed. The clinical impact of this discordance is unclear and, in practice, would depend on the endpoint being examined. An indication of the potential impact of discordance on prognostication can be gleaned from a prognostic tool developed by the TIL WG for early TNBC (Online TIL and Prognosis Tool - International TILS Working Group (tilsinbreastcancer.org)). The tool integrates the TILs score as a continuous variable, without consideration of a cutpoint, with clinicopathological parameters and suggests discordance may have only a modest effect on prognostication. For example, a 30% difference in TILs would result in less than 10% difference in predicted invasive disease-free survival in a 45-year-old patient with node-negative, grade 3 TNBC changing from 74% (95% CI 0.74-0.78) at 20% TILs to 82% (95% CI 0.79-0.86) at 50% TILs. However, it is likely that the impact of discordance may be more significant for predicting response to therapy, and this will depend on the cutpoint that is being used.

TILs heterogeneity has been identified as the main cause of variation in reporting (Figure 4A-C). Heterogeneity can be observed between the leading edge and the center of the tumour; within the tumour; and between densely infiltrated spaced-apart tumour clusters separated by sparsely infiltrated stroma [55]. It complicates assessment in both core biopsies and excision specimens because the multiple tumour cores and fragments are included in the former. The interpretation of TILs hot-spots also contributes to variation because, while current guidance recommends excluding hot-spots from evaluation [2], this is difficult to do for core biopsies. The excellent agreement achieved by the TIL WG

using their web-based interactive scoring tool suggests that the elements of the scoring aid, namely the requirement to evaluate multiple separate tumour areas and match them against calibrated sTILs reference images, mitigate the effect of heterogeneity on variability and these two elements are emphasized for TILs assessment [55]. However, the variability associated with selecting tumour areas to score is difficult to overcome [56]. Other factors that impact on consistency to a lesser extent include technical issues e.g. crush artefact and pre-analytic factors that may be most relevant in the evaluation of resection specimens; difficulties in selecting the cells or area to score e.g. scoring apoptotic clusters, histiocytic or neutrophils; difficulty delineating the tumour boundary; and the presence of limited tumour stroma for evaluation [50, 52, 55]

The extent to which the consistency of visual TILs assessment can be improved is uncertain and consequently computational machine learning methods are being explored as a more efficient and reliable approach. This has been facilitated by advances in machine learning algorithms and hardware and by the availability of large publicly available datasets for training with considerable success achieved in areas of pathology to date [57]. Computational methods can provide more exact TILs measurements than visual assessment and have the added potential to evaluate the spatial distribution of TILs and TILs-tumour interactions in BC [58-60] with preliminary data suggesting that TILs spatial distribution may provide added prognostic information [58]. Computational methods are still experimental and until these are optimized and validated and their potential becomes clear, the responsibility of scoring TILs lies with the pathologist. Indeed, comparing different machine learning tools with each other will probably reveal the same level of variability as visual assessment, as each assay has been developed in its own setting, with its own slides, and inherent biases. The TILs-WG is partnering with other groups, under the leadership of the FDA to develop reference materials for validation of machine learning tools [61]. In the meantime, rigorous training can improve the reproducibility of visual assessment by pathologists [62, 63]. A focus on scoring challenging cases, recognizing pitfalls and on approaches to mitigate heterogeneity are key to improving consistency of TILs reporting and the TIL WG educational and training resources are invaluable in this regard.

Why should Pathologist's score TILs in their daily practice?

Clinicians use tumour stage to determine treatment options for their patients. Considering the prognostic importance of TILs in early TNBC, what would be the impact, if any, of TILs on stage? Recently,

it was shown in the pooled dataset of >2000 TNBC-patients and using a TILs cutpoint of 30%, that TILs significantly up- or downgraded tumour stage, as defined by the AJCC 8th edition staging system [5, 13]. Stage IB patients with >30% TILs have >95% 5-year survival, so it may be considered that adjuvant chemotherapy may have limited benefit for these patients, which was confirmed in early-stage TNBC-patients who were not given systemic chemotherapy. Furthermore, patients with stage IIA-tumours and >30% TILs have a better 5-year survival than stage IB tumours with <30% TILs (Figure 6). In addition, histological grade was not prognostic in this pooled analysis. Considering that most TNBCs are high grade, it may be that high grade is intrinsic to the TNBC subtype and that the immune response is a more important determinant of outcome in these tumours; whereas for luminal breast cancers the tumour cell characteristics are more important predictors of outcome and histological grade is prognostic in luminal disease. This may also partly explain why patients with TNBC benefit from immunotherapy while its benefit for patients with luminal disease is still unclear. As the systemic immune response drives outcome in patients with TNBC, so strengthening the endogenous immune system with immunotherapy explains the relative success of immunotherapy in TNBC to date. Nevertheless, histological grade should be reported in TNBC but merely to identify patients with the rarer so-called low-grade TNBCs, for example adenoid cystic carcinoma, that have low TILs but an excellent outcome [16]. In Table 2 a framework is proposed for how TILs can be integrated in daily practice and combined with PD-L1. The starting point for the pathologist is the H&E-stained section, and this should also be the starting for any immune biomarker being analyzed, including when other technologies, such as multiplex technologies and Artificial Intelligence-tools are employed.

It needs to be emphasized that TILs, in accordance with recent recommendations including the St. Gallen 2021 statement, should not be used as a standalone biomarker to choose specific chemotherapy regimens or to decide whether to withhold chemotherapy treatment. However, decisions of international committees such as St. Gallen depend on how the clinical question is worded. So far, the questions were framed as for example “Should TILs be used to define treatments in TNBC?”, and the answer is a definitive “No”. If the questions are worded as for example “Do TILs inform clinicians on the risk-profile of TNBC-patients, in combination with other prognostic variables?”, the answer should be “Yes”. Nevertheless, neither is it recommended to use TILs as a standalone predictive variable for immunotherapy, as only PDL1 has been shown to be a predictive biomarker in prospective trials. However, in current daily practice most prognostic variables are used in combination to inform clinical decisions and are not used as stand-alone variables. For example, in TNBC, current evidence indicates that tumor size combined with lymph node status and TILs accurately inform the clinician on the risk-profile of their patients. It is only for gene-

expression profiles like OncotypeDx and Mammaprint, used for luminal disease that we have level of evidence 1A. Interestingly, gene-expression assays were accepted in clinical practice guidelines before the level 1A-evidence generating trials such as Mindact and Tailor-X were published. Furthermore, many of the prognostic variables used routinely by pathologists over many decades, such as lymphovascular invasion, do not have level of evidence 1A, and these are used in combination to determine the risk-profile of patients. Thus, TILs should also be used in this way, not in isolation, but combined with all other relevant prognostic variables.

Conclusion

The evaluation of TILs using a microscope and a H&E-stained section is workable in a daily practice setting but is not perfect; this is the case for all morphological biomarkers - and modesty is always a good value to abide for. Nonetheless, the assessment is implementable in daily practice of anatomic pathologists worldwide considering the criteria commonly used by pathology laboratories (Figure 6) and it provides clinicians with important prognostic information for their patients with TNBC. TILs can also help pathologists in the assessment of PD-L1, if deemed necessary by the pathologist. The incorporation of TILs in daily practice will also facilitate the inclusion of this variable in national cancer registries, as exemplified by the Danish Breast Cancer Group (Figure 7A-B). Routine reporting of TILs will provide data on how best to use this variable, not as a stand-alone prognostic factor, but together with other accepted prognostic variables in current use, such as tumor size and nodal status and pCR after neoadjuvant treatment, to inform clinicians and patients in their discussions on treatment-options. Currently, some clinicians use TILs to provide information on the likelihood of achieving a pCR. It is important to emphasize that the evidence for TILs as a prognostic factor is much more substantial than for most other prognostic factors, such as lymphovascular tumor invasion, that we have used in our daily practice for decades. Furthermore, the excellent outcome observed at 15-years follow-up in young TNBC-patients with >75% TILs in the tumour validates the management of patients who, in the past, would have had a diagnosis of medullary breast cancer; these patients did not always receive chemotherapy because of the excellent outcome associated with this tumour type. In the recent WHO-breast tumor classification, the term medullary breast cancer has been replaced with TNBC-NST with medullary patterns, prompting many clinicians to give chemotherapy to these patients. This highlights how the naming of tumour types by the pathology community has important consequences, in this case illustrated by TILs in breast cancer.

Finally, prospective trials are needed to obtain Level IA-evidence for TILs to be used as a stand-alone variable for de-escalation of chemotherapy or as a predictive factor for chemotherapy or

immunotherapy. Figure 8 details some scenarios for consideration by the academic-, the patients-, the trial- and the industry-community. Figure 9 provides a proposed framework, using a Web-based central repository, that includes local pathology laboratories in clinical trial practices. This is in contrast to how it is done today, where the central laboratory performs an assay during a trial, assuming subsequently that this assay can then be implemented in daily practice worldwide. The example of PDL1 has proven that this is not the case. For TILs, and for many other biomarkers, we can and should be able to do better.

Legends of Figures (Figure 1, 2 and 3) are kindly provided by the International Immuno-Oncology Biomarker Working Group.

Figure 1A: A breast cancer case containing a high number of TILs. If this case was a TNBC, it may be assumed that this patient has a high likelihood of response to immunotherapy, any PD-L1 assay will be positive, and that this patient will probably have an excellent outcome.

Figure 1B: The same image as Figure 1 but illustrating that the TILs are not only in the stroma, but also in between the cancer cells. These are the so-called intra-epithelial TILs.

Figure 2: A breast cancer with no TILs. If this case was a TNBC, it may be assumed that this patient has a very low likelihood of a response to immunotherapy, any PD-L1 assay will be negative, and that this patient will probably have an adverse outcome.

Figure 3: A breast cancer from a patient with TNBC in which TILs are high but with no expression of PD-L1 using the SP142-assay. What would you or your oncologist do? 1. Nothing?; 2. Reconsider the staining? 3. Consider another assay?

Figure 4: TILs in core biopsies of triple negative breast cancers. The distribution of TILs can be relatively homogeneous, as seen in (A). Heterogeneity is common and can be observed as gradient from low to high TILs within the tumour (B) and as dense TILs aggregates surrounding tumour nests intermingled with less densely infiltrated stromal areas (C).

Figure 5: Kaplan-Meier curves of overall survival of triple-negative breast cancer patients treated with anthracycline-based chemotherapy with or without taxanes, according to clinical prognostic stage and TILs. Clinical Prognostic Stage assigned stage for all patients based on history, physical examination, imaging studies performed (but not required) and relevant biopsies. Clinical Prognostic Stage is determined by T, N, M, tumor grade, as well as subtype information using human epidermal growth factor receptor (HER2), estrogen receptor (ER), and progesterone receptor (PR)- status. Figure 5 is used with permission from Loi et al., NPJ Breast Cancer « Tumor Infiltrating Lymphocyte Stratification of Prognostic Staging of Early Stage Triple Negative Breast Cancer », in press [14].

Figure 6: Comparison of TILs to other pathological variables used in daily practice with respect to its usability in the daily practice of pathologists

Figure 7: Illustration of the Danish Guidelines for Breast Cancer from the Danish Breast Cancer Group (A), including the reporting in the Danish Breast Cancer Group database (B) [65]

Figure 8: Potential role for the use of TILs in de-escalating adjuvant and neoadjuvant chemotherapy in early-stage TNBC. This figure is used by permission from Brown et al., Cancer J. 2021 Jan-Feb 01;27(1):25-31 [4].

Figure 9: Organization of a workflow for reliable and timely biomarker scoring in a general single-center or multi-center trial. Personnel at individual centers scan the slides after processing by the local pathology department. Digital slides are uploaded to a central web-based repository. A study-specific identifier is assigned to each sam²ple. The central manager is notified by the system when new slides are available and requests pathologists to review it. When a consensus score is obtained, the trial office is notified for randomization of the patient. Figure 9 is used with permission from Hudeček et al., NPJ Breast Cancer. 2020 May 12; 6:15 [66].

Funding: no funding was received for the development of this manuscript.

Conflicts of interest/Competing interests:

MK reports being an advisory board member for BMS, MSD, Roche, Daiichi. The institute of MK receives research funding from BMS, Roche, AZ.

CS reports Advisory Board (Receipt of honoraria or consultations fees): ASTELLAS, CEPHEID, VERTEX, SEATTLE GENETICS, PUMA, AMGEN, Merck & Co., Inc. Participation in a company sponsored speaker's bureau: EISAI, PRIME ONCOLOGY, TEVA, , EXACT SCIENCES. Other support (travel, accommodation expenses): ROCHE, GENENTECH, PFIZER.

CM reports receiving personal consultancy fees by Bayer, Roche, Daiichi Sankyo, Astra Zeneca.

RS reports non-financial support from Merck and Bristol Myers Squibb; research support from Merck, Puma Biotechnology, and Roche; and personal fees from Roche and Exact Sciences for advisory boards.

All other authors report no conflicts of interest/declarations.

Availability of data and material (data transparency)
Not applicable

Code availability (software application or custom code)
Not applicable

Ethics approval: not

applicable.

Author contributions/ consent to participate/consent for publication: All authors made a substantial contribution to the conception or design of the work. All authors participated in either drafting the work or revising it critically for important intellectual content. All authors have approved the final completed version of this paper and assume accountabilities for all aspects of the work.

Acknowledgements: RS is supported by a grant from the Breast Cancer Research Foundation (BCRF, grant No. 17-194). Gert van den Eynden is thanked for willing to share figure 7.

References

1. Hayes, D.F., B. Trock, and A.L. Harris, *Assessing the clinical impact of prognostic factors: when is "statistically significant" clinically useful?* Breast Cancer Res Treat, 1998. **52**(1-3): p. 305-19.
2. Salgado, R., et al., *The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014.* Ann Oncol, 2015. **26**(2): p. 259-71.
3. Sistrunk, W.E. and W.C. Maccarty, *Life Expectancy Following Radical Amputation for Carcinoma of the Breast: A Clinical and Pathologic Study of 218 Cases.* Ann Surg, 1922. **75**(1): p. 61-9.
4. Brown, L.C., et al., *Tumor-Infiltrating Lymphocyctes in Triple-Negative Breast Cancer: Update for 2020.* Cancer J, 2021. **27**(1): p. 25-31.
5. Loi, S., et al., *Tumor-Infiltrating Lymphocytes and Prognosis: A Pooled Individual Patient Analysis of Early-Stage Triple-Negative Breast Cancers.* J Clin Oncol, 2019. **37**(7): p. 559-569.
6. *AGO 2021E_05_Prognostic_and_Predictive_Factors_MASTER_final_20210302.pdf (ago-online.de).*
7. Burstein, H.J., 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**(10): p. 1541-1557.
8. Cardoso, F., et al., *Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-updagger.* Ann Oncol, 2019. **30**(8): p. 1194-1220.
9. Harris, L.N., et al., *Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline.* J Clin Oncol, 2016. **34**(10): p. 1134-50.
10. Simon, R.M., S. Paik, and D.F. Hayes, *Use of archived specimens in evaluation of prognostic and predictive biomarkers.* J Natl Cancer Inst, 2009. **101**(21): p. 1446-52.
11. De Jong, V.M.T.W.Y.O.M.e.a., *Prognostic value of tumour infiltrating lymphocytes in young triple negative breast cancer patients who did not receive adjuvant systemic treatment; by the PARADIGM study group.* Annals of Oncology, 2020. **31**(supplement 4): p. S303.
12. Park, J.H., et al., *Prognostic value of tumor-infiltrating lymphocytes in patients with early-stage triple-negative breast cancers (TNBC) who did not receive adjuvant chemotherapy.* Ann Oncol, 2019. **30**(12): p. 1941-1949.
13. Cancer, A.J.C.O., *AJCC Cancer Staging Manual, 8th Edition.* 2017.
14. Sherene Loi1, R.S., Sylvia Adams, Giancarlo Pruneri, Prudence A Francis , Magali Lacroix-Triki, Heikki Joensuu, Maria Vittoria Dieci, Sunil Badve, Sandra Demaria, Robert Gray, Elisabetta Munzone, Damian Drubay, Jerome Lemonnier, Christos Sotiriou, Pirkko Liisa Kellokumpu-Lehtinen, Andrea Vingiani, Kathryn Gray, Fabrice André, Carsten Denkert, Martine Piccart, Elvire Roblin, Stefan Michiels, *Tumor Infiltrating Lymphocyte Stratification of Prognostic Staging of Early Stage Triple Negative Breast Cancer.* NPJ Breast Cancer, 2021. **In Press**.
15. Tan, P.H., et al., *The 2019 World Health Organization classification of tumours of the breast.* Histopathology, 2020. **77**(2): p. 181-185.
16. Board, I.A.f.R.o.C.L., *Breast Tumours, WHO Classification of Tumours, 5th Edition.* 2019. **2**.
17. Provenzano, E., et al., *Standardization of pathologic evaluation and reporting of postneoadjuvant specimens in clinical trials of breast cancer: recommendations from an international working group.* Mod Pathol, 2015. **28**(9): p. 1185-201.
18. Huang, M., et al., *Association of Pathologic Complete Response with Long-Term Survival Outcomes in Triple-Negative Breast Cancer: A Meta-Analysis.* Cancer Res, 2020. **80**(24): p. 5427-5434.
19. Cortazar, P., et al., *Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis.* Lancet, 2014. **384**(9938): p. 164-72.

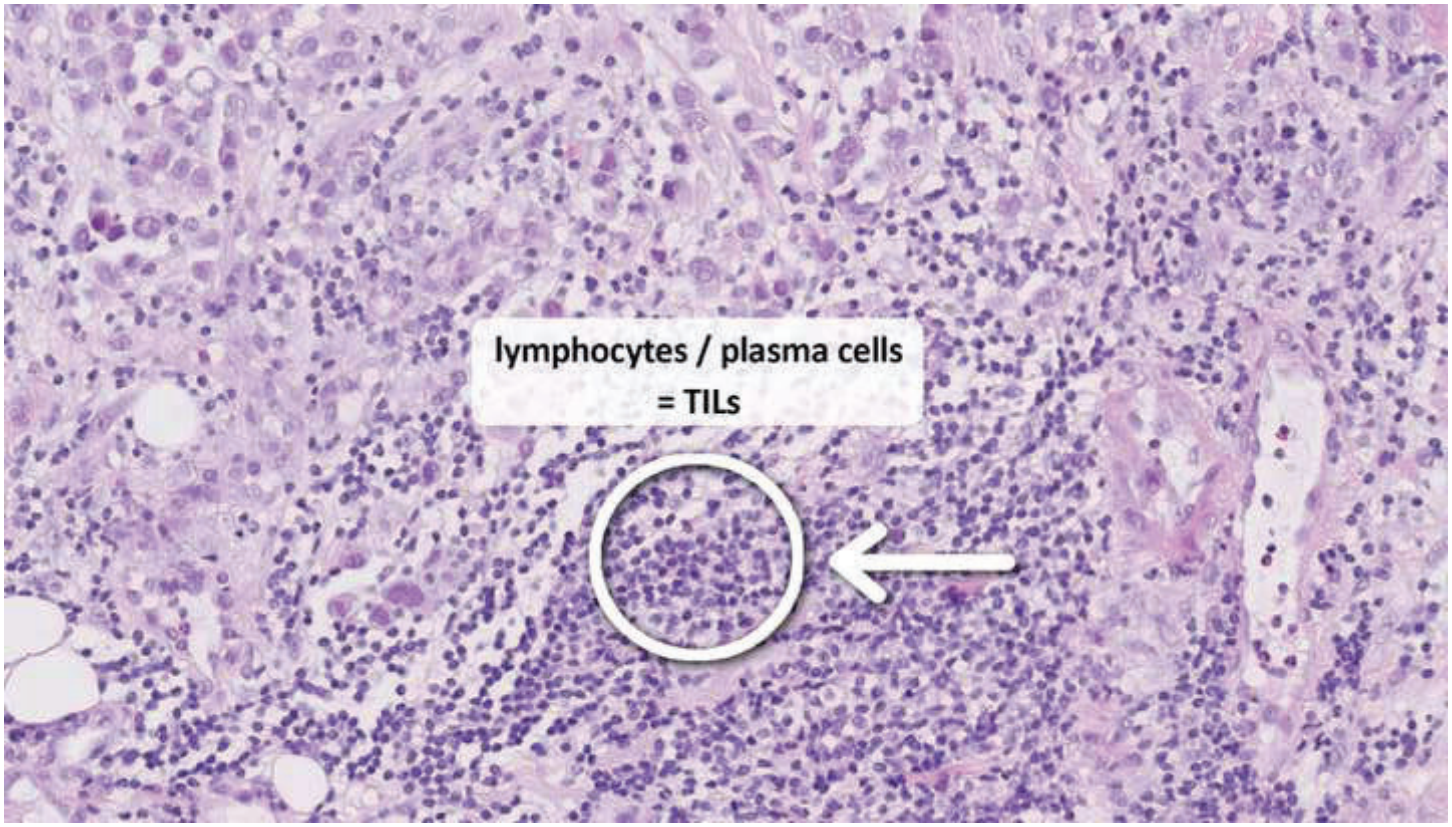
20. Samiei, S., et al., *Axillary Pathologic Complete Response After Neoadjuvant Systemic Therapy by Breast Cancer Subtype in Patients With Initially Clinically Node-Positive Disease: A Systematic Review and Meta-analysis.* JAMA Surg, 2021: p. e210891.
21. Denkert, C., et al., *Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy.* Lancet Oncol, 2018. **19**(1): p. 40-50.
22. Lee, J.S., S.E. Yost, and Y. Yuan, *Neoadjuvant Treatment for Triple Negative Breast Cancer: Recent Progresses and Challenges.* Cancers (Basel), 2020. **12**(6).
23. Miyashita, H., et al., *Neo-adjuvant therapy for triple-negative breast cancer: Insights from a network meta-analysis.* Breast J, 2020. **26**(9): p. 1717-1728.
24. Savas, P., R. Salgado, and S. Loi, *Seeing the forest and the tree: TILs and PD-L1 as immune biomarkers.* Breast Cancer Res Treat, 2021. **189**(3): p. 599-606.
25. Loibl, S., 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**(8): p. 1279-1288.
26. Chien, A.J., et al., *MK-2206 and Standard Neoadjuvant Chemotherapy Improves Response in Patients With Human Epidermal Growth Factor Receptor 2-Positive and/or Hormone Receptor-Negative Breast Cancers in the I-SPY 2 Trial.* J Clin Oncol, 2020. **38**(10): p. 1059-1069.
27. Schmid, P., et al., *Pembrolizumab plus chemotherapy as neoadjuvant treatment of high-risk, early-stage triple-negative breast cancer: results from the phase 1b open-label, multicohort KEYNOTE-173 study.* Ann Oncol, 2020. **31**(5): p. 569-581.
28. Schmid, P., et al., *Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer.* N Engl J Med, 2018. **379**(22): p. 2108-2121.
29. Miles, D., et al., *Primary results from IMpassion131, a double-blind, placebo-controlled, randomised phase III trial of first-line paclitaxel with or without atezolizumab for unresectable locally advanced/metastatic triple-negative breast cancer.* Ann Oncol, 2021. **32**(8): p. 994-1004.
30. Elias, D., et al., *Correction: Gene expression profiling identifies FYN as an important molecule in tamoxifen resistance and a predictor of early recurrence in patients treated with endocrine therapy.* Oncogene, 2020. **39**(10): p. 2229-2230.
31. Emens, L.A., et al., *Atezolizumab and nab-Paclitaxel in Advanced Triple-Negative Breast Cancer: Biomarker Evaluation of the IMpassion130 Study.* J Natl Cancer Inst, 2021. **113**(8): p. 1005-1016.
32. Winer, E.P., et al., *Pembrolizumab versus investigator-choice chemotherapy for metastatic triple-negative breast cancer (KEYNOTE-119): a randomised, open-label, phase 3 trial.* Lancet Oncol, 2021. **22**(4): p. 499-511.
33. Schmid P, C.J., Dent R, et al., *KEYNOTE-522: Phase 3 study of neoadjuvant pembrolizumab plus chemotherapy versus placebo plus chemotherapy, followed by adjuvant pembrolizumab versus placebo for early-stage triple-negative breast cancer.* ESMO Virtual Plenary. Abstract VP7-2021., 2021.
34. Cortes, J., et al., *Pembrolizumab plus chemotherapy versus placebo plus chemotherapy for previously untreated locally recurrent inoperable or metastatic triple-negative breast cancer (KEYNOTE-355): a randomised, placebo-controlled, double-blind, phase 3 clinical trial.* Lancet, 2020. **396**(10265): p. 1817-1828.
35. El Bairi, K., et al., *The tale of TILs in breast cancer: A report from The International Immuno-Oncology Biomarker Working Group.* NPJ Breast Cancer, 2021. **7**(1): p. 150.
36. Gonzalez-Ericsson, P.I., 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**(5): p. 667-684.

37. Buttner, R., et al., *Programmed Death-Ligand 1 Immunohistochemistry Testing: A Review of Analytical Assays and Clinical Implementation in Non-Small-Cell Lung Cancer.* J Clin Oncol, 2017. **35**(34): p. 3867-3876.
38. Martinez-Morilla, S., et al., *Quantitative assessment of PD-L1 as an analyte in immunohistochemistry diagnostic assays using a standardized cell line tissue microarray.* Lab Invest, 2019.
39. Rimm, D.L., et al., *Reanalysis of the NCCN PD-L1 companion diagnostic assay study for lung cancer in the context of PD-L1 expression findings in triple-negative breast cancer.* Breast Cancer Res, 2019. **21**(1): p. 72.
40. Reisenbichler, E.S., 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**(9): p. 1746-1752.
41. Torlakovic, E., et al., *"Interchangeability" of PD-L1 immunohistochemistry assays: a meta-analysis of diagnostic accuracy.* Mod Pathol, 2020. **33**(1): p. 4-17.
42. Rugo, H.S., *Performance of PD-L1 immunohistochemistry (IHC) assays in unresectable locally advanced or metastatic triple-negative breast cancer (mTNBC).* Annals of Oncology, 2019. **30 (suppl_5)**: p. v851-v934.
43. Emens, L.A., et al., *Atezolizumab and nab-Paclitaxel in Advanced Triple-Negative Breast Cancer: Biomarker Evaluation of the IMpassion130 Study.* J Natl Cancer Inst, 2021.
44. Lundh, A., et al., *Industry sponsorship and research outcome.* Cochrane Database Syst Rev, 2017. **2**: p. MR000033.
45. Salgado, R., et al., *How current assay approval policies are leading to unintended imprecision medicine.* Lancet Oncol, 2020. **21**(11): p. 1399-1401.
46. Franzoi, M.A. and E. de Azambuja, *Atezolizumab in metastatic triple-negative breast cancer: IMpassion130 and 131 trials - how to explain different results?* ESMO Open, 2020. **5**(6): p. e001112.
47. Schmid, P., et al., *Atezolizumab plus nab-paclitaxel as first-line treatment for unresectable, locally advanced or metastatic triple-negative breast cancer (IMpassion130): updated efficacy results from a randomised, double-blind, placebo-controlled, phase 3 trial.* Lancet Oncol, 2020. **21**(1): p. 44-59.
48. Grabenstetter, A., et al., *PD-L1 Expression in Metaplastic Breast Carcinoma Using the PD-L1 SP142 Assay and Concordance Among PD-L1 Immunohistochemical Assays.* Am J Surg Pathol, 2021. **45**(9): p. 1274-1281.
49. Noske, A., et al., *A multicentre analytical comparison study of inter-reader and inter-assay agreement of four programmed death-ligand 1 immunohistochemistry assays for scoring in triple-negative breast cancer.* Histopathology, 2021. **78**(4): p. 567-577.
50. O'Loughlin, M., 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): p. 1-9.
51. Denkert, C., 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**(10): p. 1155-64.
52. Kim, R.S., et al., *Stromal Tumor-infiltrating Lymphocytes in NRG Oncology/NSABP B-31 Adjuvant Trial for Early-Stage HER2-Positive Breast Cancer.* J Natl Cancer Inst, 2019. **111**(8): p. 867-871.
53. Tramm, T., et al., *Standardized assessment of tumor-infiltrating lymphocytes in breast cancer: an evaluation of inter-observer agreement between pathologists.* Acta Oncol, 2018. **57**(1): p. 90-94.

54. Swisher, S.K., 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**(7): p. 2242-8.
55. Kos, Z., et al., *Pitfalls in assessing stromal tumor infiltrating lymphocytes (sTILs) in breast cancer.* NPJ Breast Cancer, 2020. **6**: p. 17.
56. Kilmartin, D., et al., *Intra-Tumour Heterogeneity Is One of the Main Sources of Inter-Observer Variation in Scoring Stromal Tumour Infiltrating Lymphocytes in Triple Negative Breast Cancer.* Cancers (Basel), 2021. **13**(17).
57. Klauschen, F., et al., *Scoring of tumor-infiltrating lymphocytes: From visual estimation to machine learning.* Semin Cancer Biol, 2018. **52**(Pt 2): p. 151-157.
58. Nawaz, S., et al., *Beyond immune density: critical role of spatial heterogeneity in estrogen receptor-negative breast cancer.* Mod Pathol, 2015. **28**(6): p. 766-77.
59. Heindl, A., et al., *Relevance of Spatial Heterogeneity of Immune Infiltration for Predicting Risk of Recurrence After Endocrine Therapy of ER+ Breast Cancer.* J Natl Cancer Inst, 2018. **110**(2).
60. Skriver, S.K., et al., *Tumour-infiltrating lymphocytes and response to neoadjuvant letrozole in patients with early oestrogen receptor-positive breast cancer: analysis from a nationwide phase II DBCG trial.* Breast Cancer Res, 2020. **22**(1): p. 46.
61. Dudgeon SN, W.S., Hanna MG, Gupta R, Amgad M, Sheth M, Marble H, Huang R, Herrmann MD, Szu CH, Tong D, Werness B, Szu E, Larsimont D, Madabhushi A, Hytopoulos E, Chen W, Singh R, Hart SN, Sharma A, Saltz J, Salgado R, Gallas BD., *A pathologist-annotated dataset for validating artificial intelligence: A project description and pilot study.* J Pathol Inform 2021;12:45 2021.
62. Polley, M.Y., et al., *An international study to increase concordance in Ki67 scoring.* Mod Pathol, 2015. **28**(6): p. 778-86.
63. Nielsen, T.O., et al., *Assessment of Ki67 in Breast Cancer: Updated Recommendations from the International Ki67 in Breast Cancer Working Group.* J Natl Cancer Inst, 2020.
64. Dieci, M.V., et al., *Update on tumor-infiltrating lymphocytes (TILs) in breast cancer, including recommendations to assess TILs in residual disease after neoadjuvant therapy and in carcinoma in situ: A report of the International Immuno-Oncology Biomarker Working Group on Breast Cancer.* Semin Cancer Biol, 2018. **52**(Pt 2): p. 16-25.
65. www.DBCG.dk.
66. Hudecek, J., et al., *Application of a risk-management framework for integration of stromal tumor-infiltrating lymphocytes in clinical trials.* NPJ Breast Cancer, 2020. **6**: p. 15.

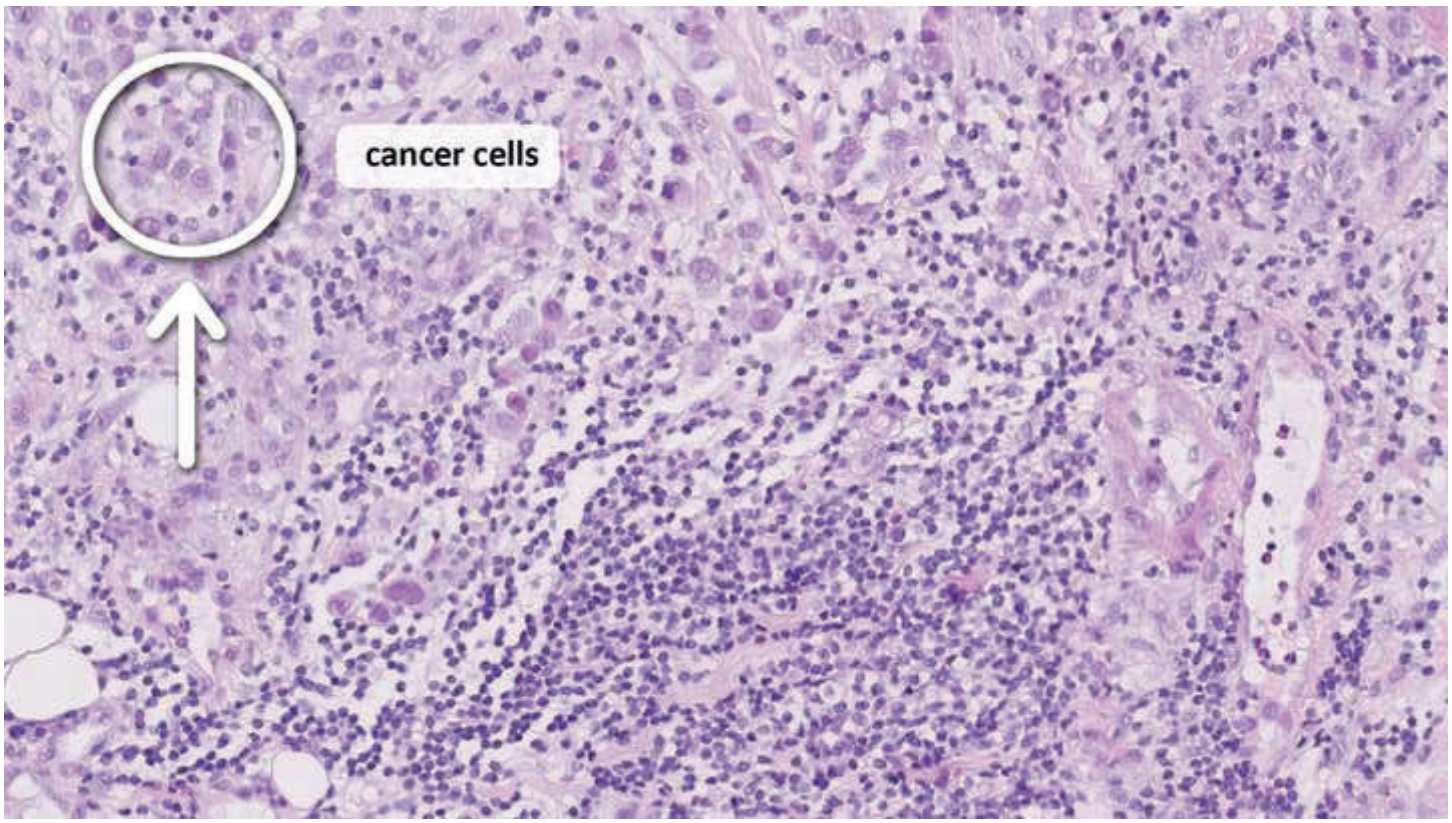
Figure

Click here to access/download;Figure;Figure 1A.jpg



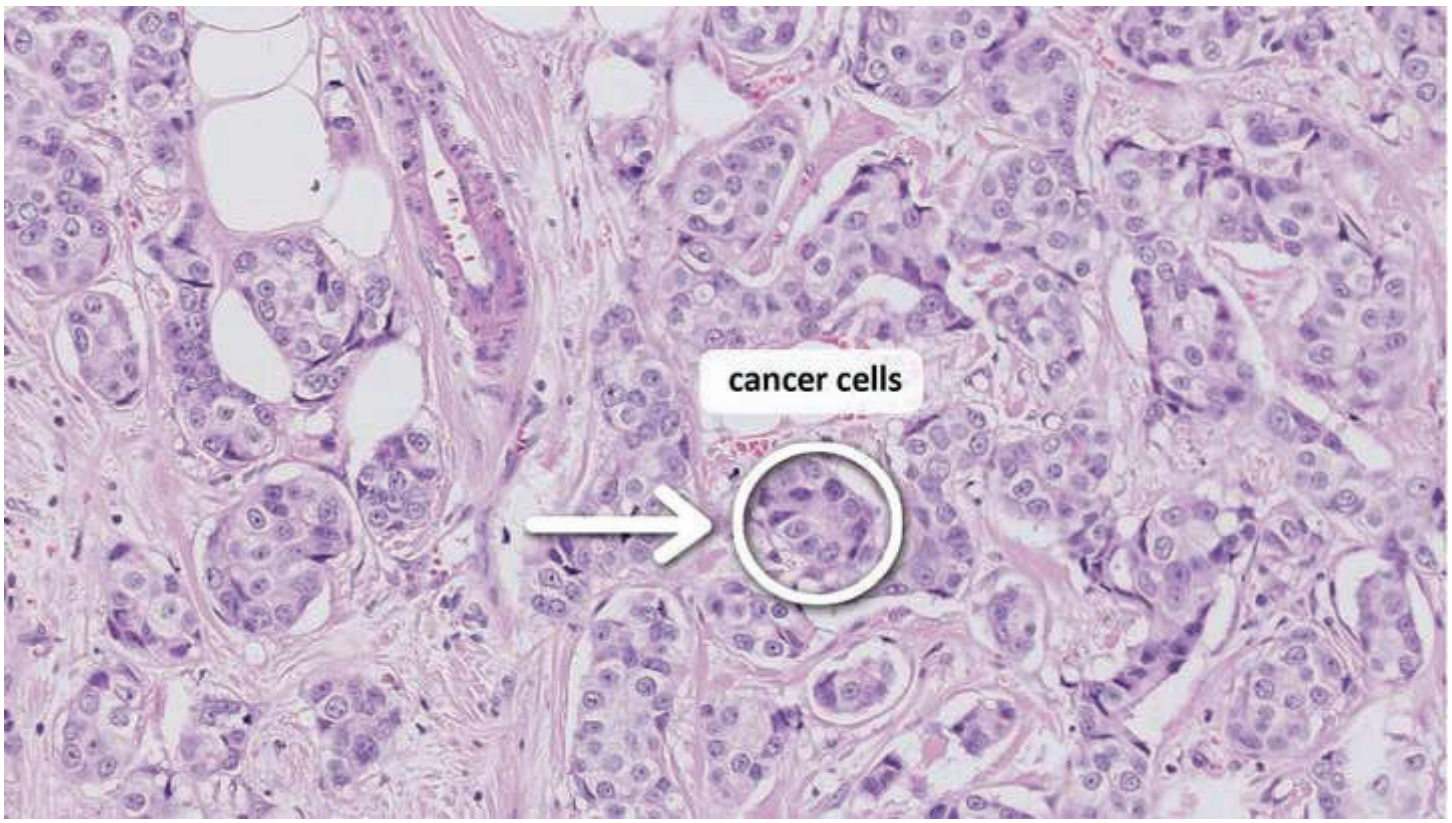
Figure

Click here to access/download;Figure;Figure 1B.jpg



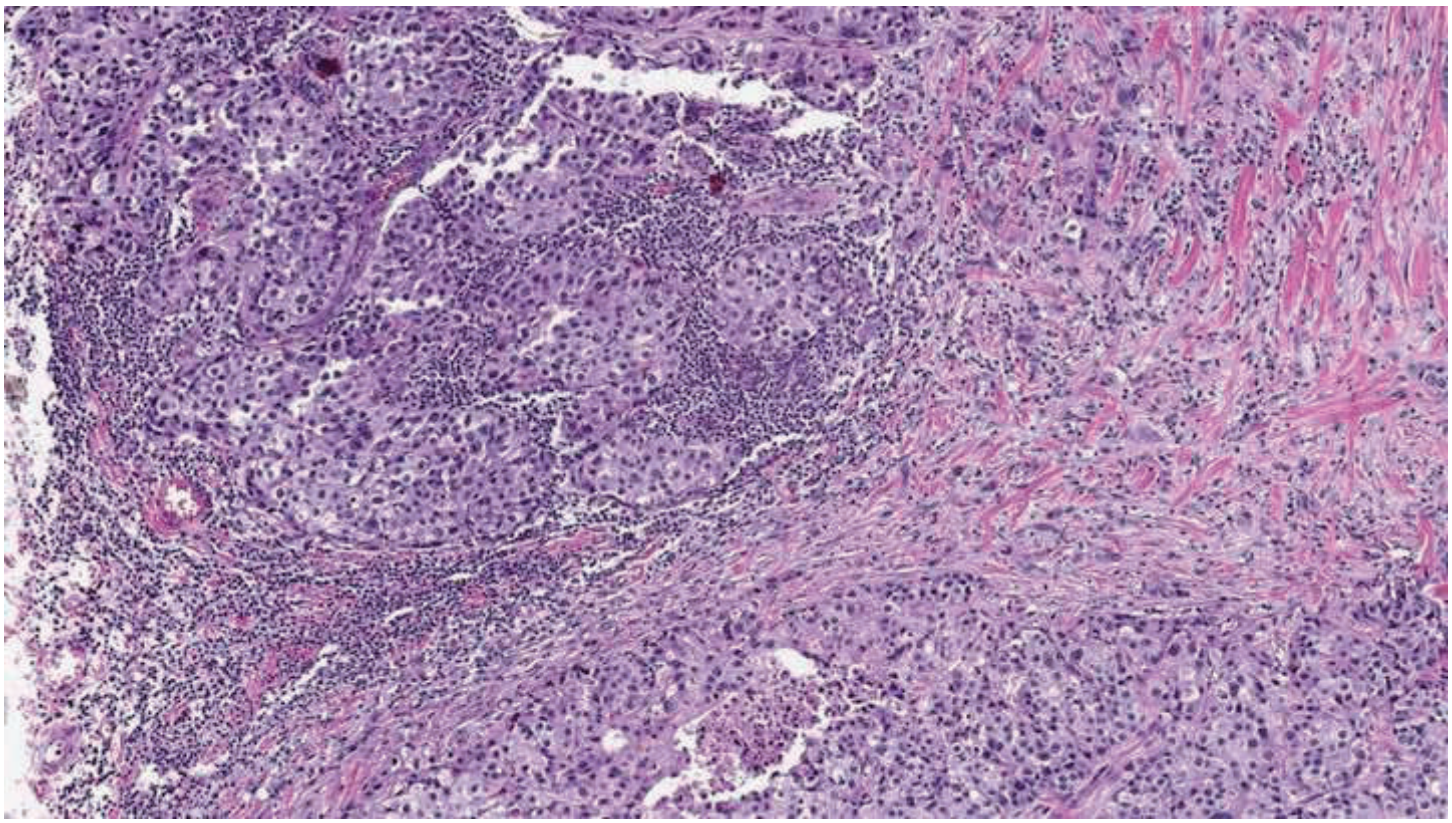
Figure

Click here to access/download;Figure;Figure 2.jpg



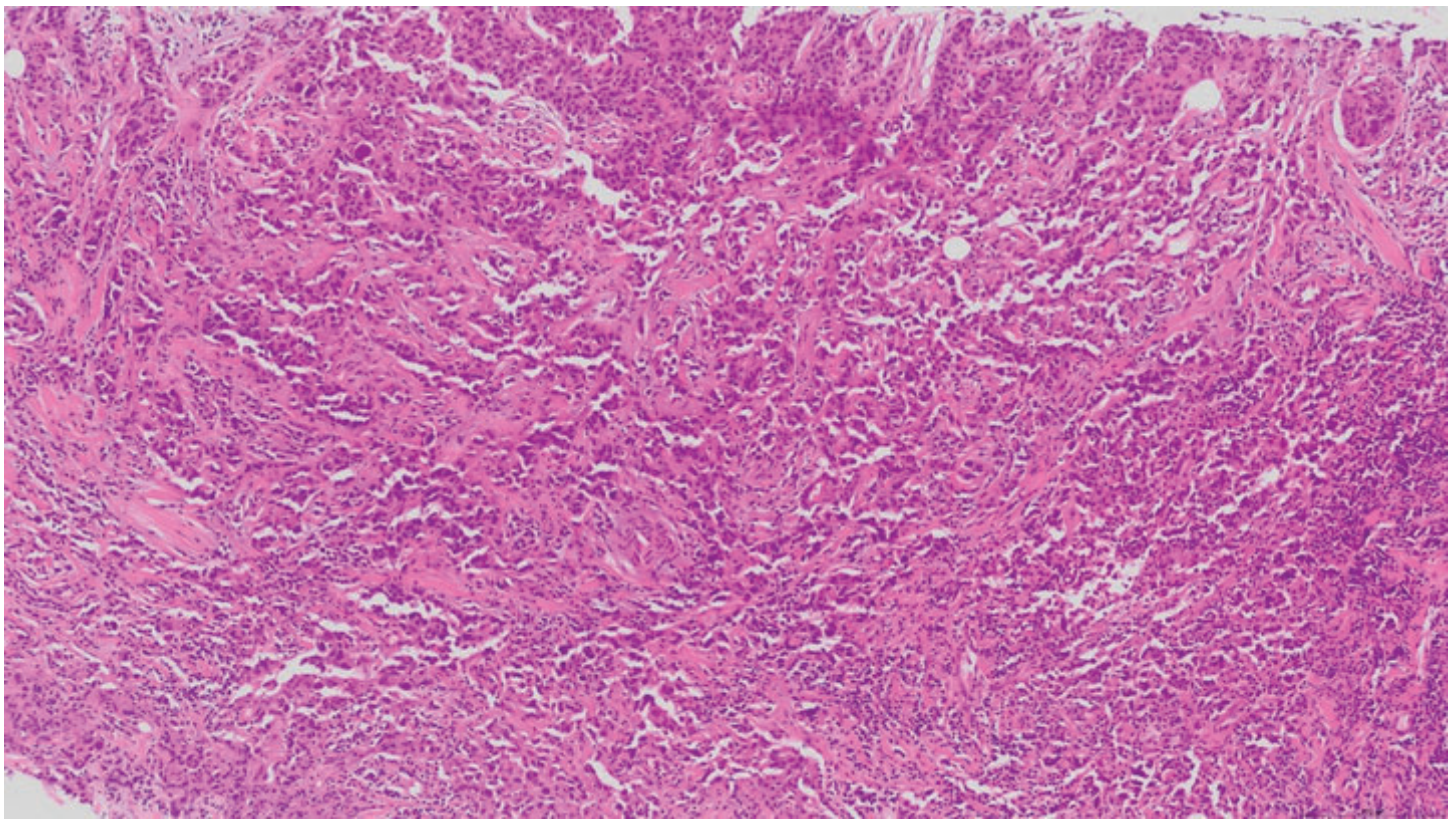
Figure

Click here to access/download;Figure;Figure 3.jpg



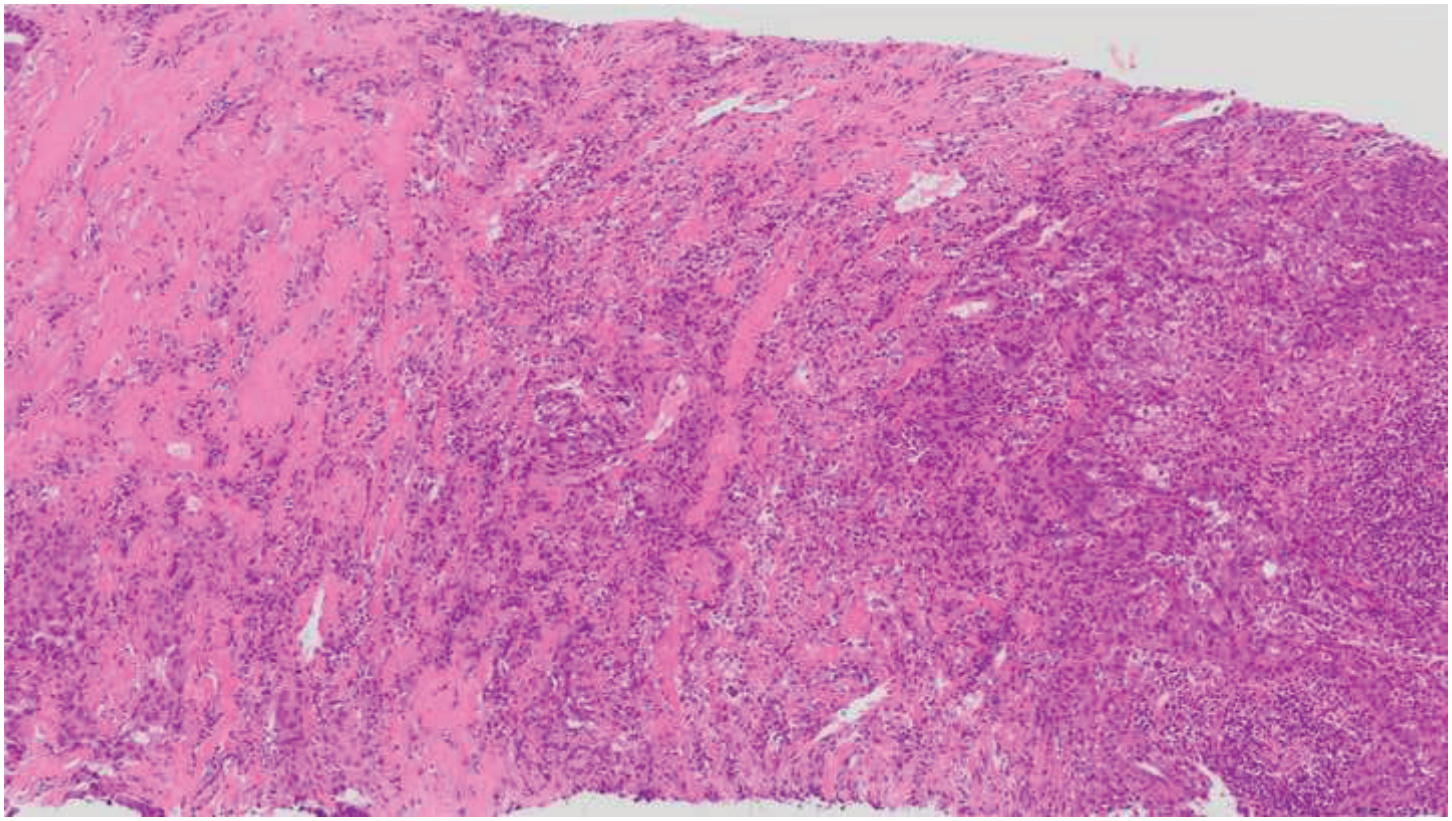
Figure

Click here to access/download;Figure;Figure 4A.jpg



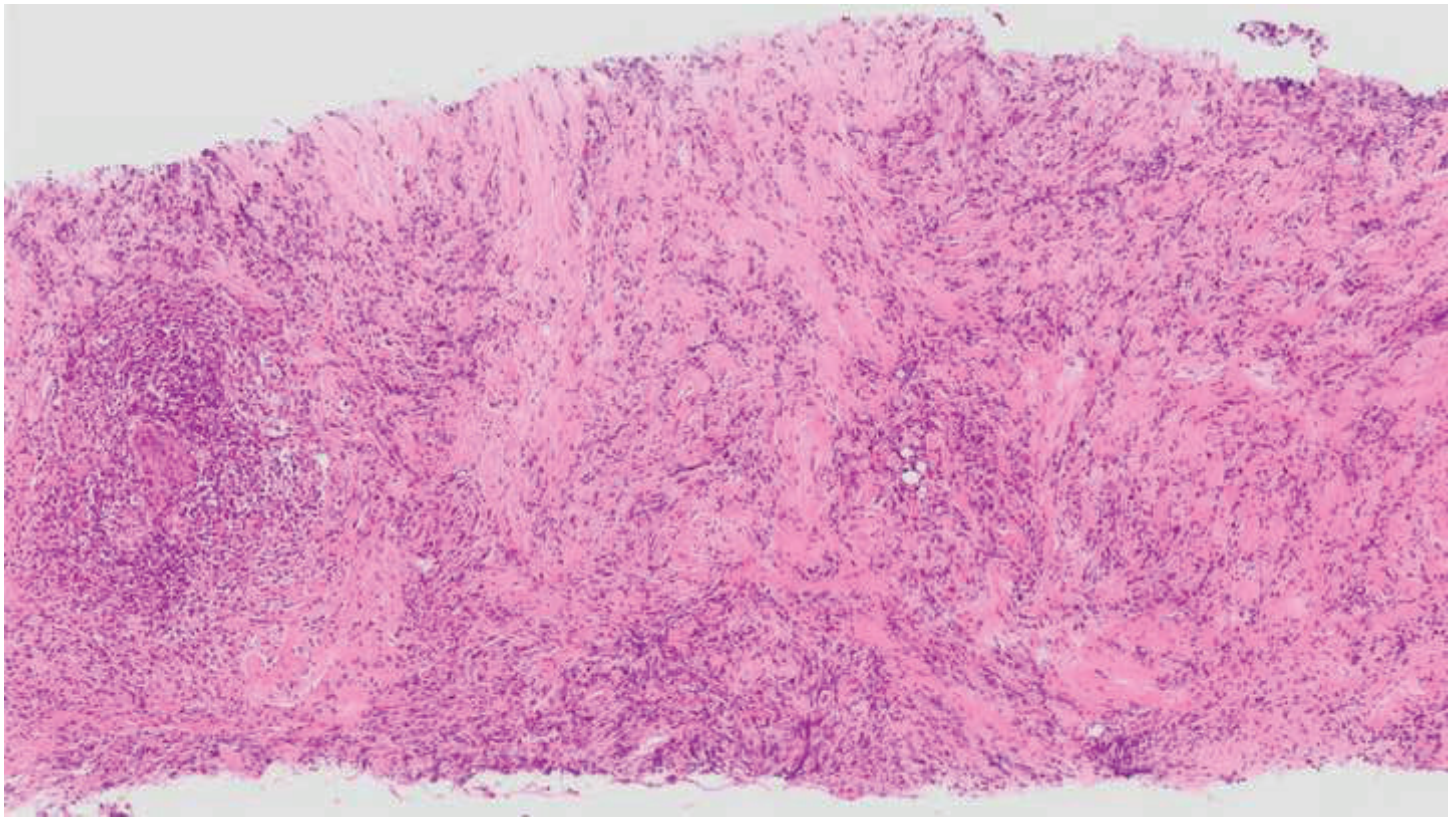
Figure

Click here to access/download;Figure;Figure 4B.jpg



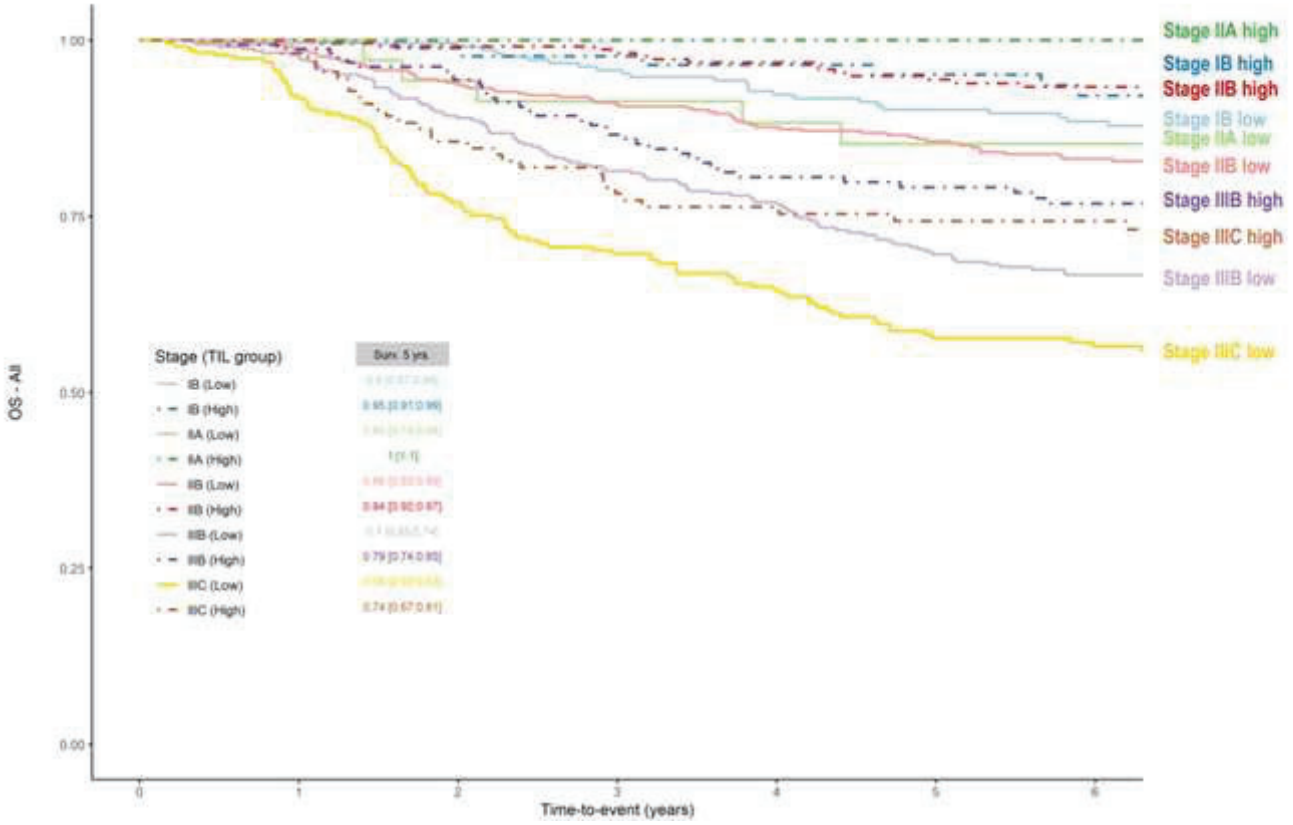
Figure

Click here to access/download;Figure;Figure 4C.jpg



Figure

Click here to access/download;Figure;Figure 5.jpg



Figure

Click here to access/download;Figure;Figure 6.jpg

	FFPE	Easy	Standardized	Reproducible	Well-documented
Histological (sub)type	V	V	V	+/-	V
Histological grade	V	V	V	+/-	V
Tumour burden (size + LN-status)	V	V	V	V	V
Mitotic activity	V	V	V	+/-	V
Perineural Invasion	V	V	+/-	+/-	V
Lymphovascular Invasion	V	V	+/-	+/-	V
Pharmacodiagnostic biomarkers like PD-L1	V	V	V	+/-	V
TILs	V	V	V	V	V

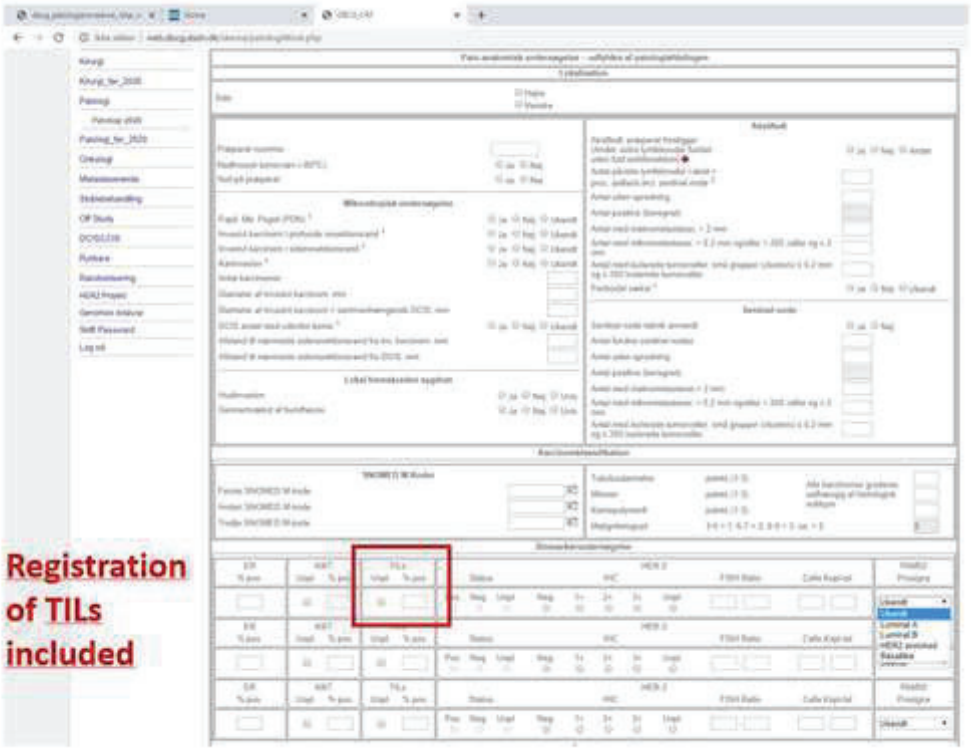
Figure

Click here to access/download;Figure;Figure 7A.jpg

Figure

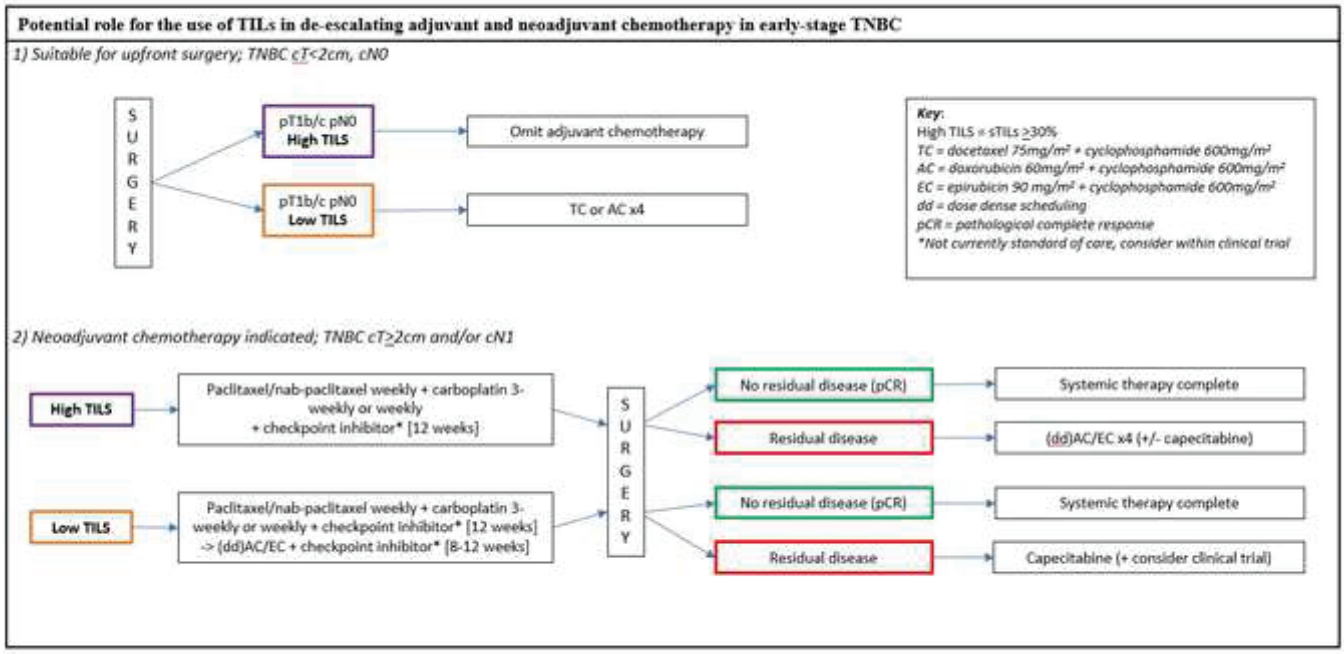
Click here to access/download;Figure;Figure 7B.jpg

- All clinico-pathological data are registrered in the DBCG database (web-registration).
- Pathology web-schemes for each procedure, here exemplified by registration of pathology data after lumpectomy.
- Separate web-schemes in the neo-adjuvant setting including reporting of RCB class.
- The biomarker section is identical for all schemes (for up to three tumours in multifocal disease).
- The pathology report and web-scheme also include information of prognostic gene signature.
- **Annual publication regarding quality af the reported data – as part of the The Danish Clinical Quality Program – National Clinical Registries (RKKP) which constitutes the infrastructure of the Danish clinical quality registries and the Danish Multidisciplinary Cancer Groups (DMCG).**



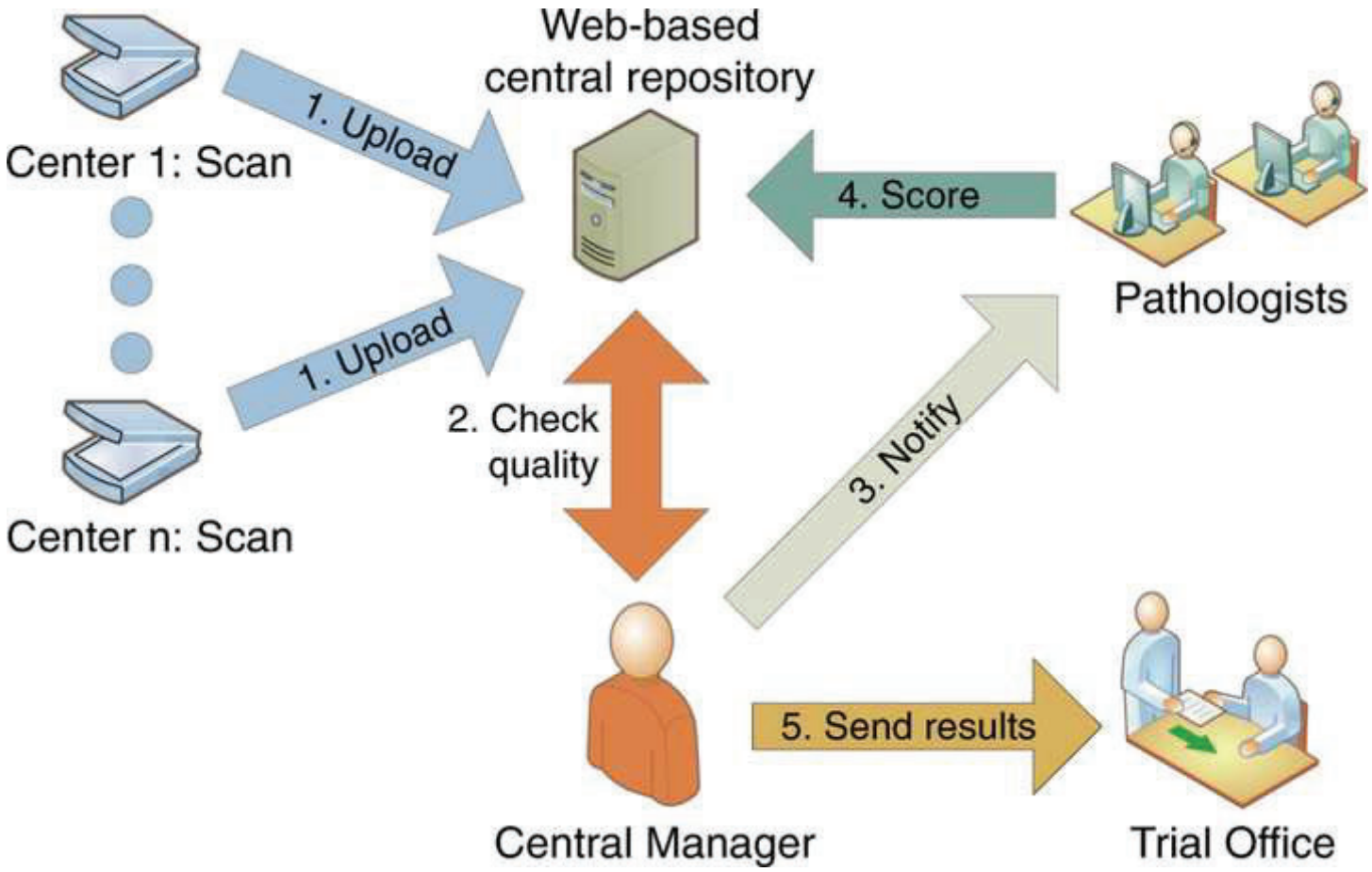
Figure

Click here to access/download;Figure;Figure 8.jpg



Figure

Click here to access/download;Figure;Figure 9.jpg



Table

Table 1 Solutions to improve the current-assay approval narrative. Reprinted from "How current assay approval policies are leading to unintended imprecision medicine. From The Lancet Oncology 2020; 21 (11): 1399-1401, Copyright (2020), with permission from Elsevier [45]"

- ❖ Industry should be mandated to perform concordance studies with other similar assays or standardized controls before a drug is approved.
- ❖ Industry should support, in concert with all stakeholders, relabeling or revising approved companion diagnostics if there is evidence that the labeling may lead to uncertainty in the identification of patients for treatments.
- ❖ Industry should support, in concert with all stakeholders, relabeling or revising of the companion diagnostics if equivalent clinical validity has been demonstrated with other biomarkers or standards, providing access to clinical trial tissues to validate other assays.
- ❖ Industry, when considering the incorporation of assays in their trials, should communicate and share assay-information when using an assay that identifies the same molecule (epitope, antigen, DNA, RNA, etc…) as in other competitive trials. For example, methodological information related to the binding sites of the antibodies used in the companion diagnostic assay, should be made public, even if this information is commercially sensitive.
- ❖ Pathways for regulatory acceptance of other assays that are equivalent, but less expensive and easier to implement in daily practice, should be developed by governments and regulatory agencies ideally before a drug is labeled together with a companion diagnostic.
- ❖ Early engagement by all stakeholders in External Quality Control Schemes to allow rapid development of guidelines and quality standards is essential, preferably before an assay is approved by the regulatory agencies.
- ❖ Clinical practice guidelines developed by professional organizations like ASCO, ESMO, etc... should endorse not just a companion diagnostic assay used in the trial, but any rigorously technically validated equivalent laboratory assays that can define substantially the same population as the companion diagnostic.
- ❖ Regulators should require data confirming the analytical validity of the companion diagnostic in the distributed setting in which it would be applied, at a level of rigor comparable to that required to show efficacy of the drug in question.

Table

Table 2 Framework on how to include TILs in TNBC in the daily practice of Pathologists.

- TILs can be assessed with good reproducibility by Pathologists.
- Pathologists only need a microscope and a H&E-stained section and can be trained using a freely available training-tool (www.tilsinbreastcancer.org).
- TILs scoring can be done at the time of diagnosis.
- Stage I TNBC with high TILs have excellent 5-year survival, irrespective of treatment. Stage II patients with high TILs have a better outcome than stage I patients with low TILs. This is why clinicians and patients need to know on the TILs.
- TILs and PD-L1 are associated with prediction of response to immunotherapy.
- If pathologists score TILs for prognostic purposes in their daily practice, this information is already available in the report if needed for selection for ICI, in a combination with PD-L1, at a later date.
- If the patient develops metastasis, the pathologist may use any PD-L1-antibody, if it is well validated, and used in conjunction with TILs. If there are no TILs, PD-L1 IC will be negative, and if there are many TILs, it may not matter too much which assay is used, as long as the assay is validated.

Table

Table 3 Overview of reproducibility-studies on TILs assessment

	Participants n	Cases n	Cohort	BC Subtype	Inter-Observer Agreement for TILs			
					Continuous	Categorical		
					ICC (95% CI)	Cutpoint (%)	Kappa[c]	Cutpoint (%)	Concordance rate[d]
Denkert, 2016 RS 1[a] [51]	34	60	Digitized slides of NCBs from GeparSixto; One third low, one third intermediate and one third with high TILs	TNBC	0.7(0.62-0.78)	60 50 0-20, 21-49, >50	0.45 0.51 0.46	1 5 10 30 75	0.94 (± 0.08) 0.83 (± 0.09) 0.77 (± 0.08) 0.81 (± 0.08) 0.90 (± 0.06)
Denkert , 2016 RS 2[a] [51]	28	60	Digitized slides of NCBs from GeparSixto; One third low, one third intermediate and one third with high TILs.	TNBC	0.89 (0.85-0.92)	60 50 0-20, 21-49, >50	0.63 0.72 0.65	1 5 10 30 75	0.94 (± 0.04) 0.89 (± 0.05) 0.86 (± 0.05) 0.93 (± 0.03) 0.92 (± 0.03)
Swisher, 2016 [54]	4	75	Glass slides of NCBs from routine practice	TNBC		< 10, 10-50, >50	0.57 (0.04)[e]		
O'Loughlin, 2018[b] [50]	19	84	Digitized slides of NCBs from routine practice	TNBC	0.660 (0.58-0.75)	25 50	0.50 (0.41-0.61)[f] 0.48 (0.39-0.59)		
Tramm, 2018 [53]	9	124	Digitized slides of NCBs from routine practice	All	0.71 (0.65-0.77)	0-10, 11-39, > 40 0-20, 21-49, >50 >50 60	0.41 0.36 0.48 0.44	<10, 11-39,>40 <20, 21-49,>50 50 60	0.79 (0.60-0.09)[g] 0.82 (0.54-0.92) 0.93 (0.81-0.99) 0.95 (0.77-0.99)
Dieci, 2018 [64]	6	50	Digitized slides of whole section/resection cases post NACT	TNBC	0.59 (0.45-0.70)	11 categories	0.58 (0.47-0.70)[h]		
Kim, 2019 [52]	7	100	Digitized slides from whole section/resection cases from NSABP-B31 trial	HER2 Positive	0.76 (0.69-0.83)	60 50 0-20, 21-49, >50	0.63 0.72 0.65	1 5 10 30 75	0.91 (± 0.06) 0.84 (± 0.1) 0.79 (± 0.06) 0.87 (± 0.04) 0.94 (± 0.03)

Kilmartin,[i] [56]	23	49	Digitized slides of NCBs from routine practice	TNBC	0.63 (0.54-0.74)	20	0.48 (0.39-0.60)[f]
						25	0.57 (0.48-0.68)
						30	0.54 (0.48-0.68)
						40	0.49 (0.39-0.61)
						50	0.43 (0.33-0.55)
						60	0.35 (0.26-0.47)

BC, breast cancer; CI, Confidence Interval; ICC, Intraclass coefficient; n, number; NCB, needle core biopsy; NACT, neoadjuvant chemotherapy; RS, Ring Study; TNBC, Triple Negative Breast Cancer.

a Denkert: two ring studies with an independent set of 60 cases used for each. A web-based interactive scoring aid was used to score cases in ring study 2.

b. TILs were scores in 2 rounds with an interval of 4 months: ICCs from second circulation.

c Fleiss kappa except for Tramm et al, Light's kappa; Intraclass coefficient for O'Loughlin et al.

d The concordance of all pairs of pathologists was calculated for five different TIL groups. the values in the table are the sample mean and the sample standard deviation of these concordance rates for all pairs of pathologists in each study.

e Standard error in parenthesis

f ICC, 95% CI in parenthesis

g Range in parenthesis

h Light's kappa 95% confidence interval in parenthesis

i A web-based interactive scoring aid develop by the TIL WG was used to score cases.