

RESEARCH ARTICLE

CD8+ T cell infiltration in breast and colon cancer: A histologic and statistical analysis

James Ziai^{1*}, Houston N. Gilbert², Oded Foreman¹, Jeffrey Eastham-Anderson¹, Felix Chu¹, Mahrukh Huseni³, Jeong M. Kim⁴

1 Department of Pathology, Genentech, Inc., South San Francisco, California, United States of America, **2** Department of Biostatistics, Genentech, Inc., South San Francisco, California, United States of America, **3** Department of Oncology Biomarker Development, Genentech, Inc., South San Francisco, California, United States of America, **4** Department of Cancer Immunology, Genentech, Inc., South San Francisco, California, United States of America

* ziai.james@gene.com



OPEN ACCESS

Citation: Ziai J, Gilbert HN, Foreman O, Eastham-Anderson J, Chu F, Huseni M, et al. (2018) CD8+ T cell infiltration in breast and colon cancer: A histologic and statistical analysis. PLoS ONE 13(1): e0190158. <https://doi.org/10.1371/journal.pone.0190158>

Editor: Joseph Najbauer, University of Pécs Medical School, HUNGARY

Received: June 13, 2017

Accepted: December 8, 2017

Published: January 10, 2018

Copyright: © 2018 Ziai et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: We also confirm here that while all funding sources for this study were internal to Genentech, Inc., the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: While all authors were employees of Genentech at the time of the original manuscript preparation, this does not alter our

Abstract

The prevalence of cytotoxic tumor infiltrating lymphocytes (TILs) has demonstrated prognostic value in multiple tumor types. In particular, CD8 counts (in combination with CD3 and CD45RO) have been shown to be superior to traditional UICC staging in colon cancer patients and higher total CD8 counts have been associated with better survival in breast cancer patients. However, immune infiltrate heterogeneity can lead to potentially significant misrepresentations of marker prevalence in routine histologic sections. We examined step sections of breast and colorectal cancer samples for CD8+ T cell prevalence by standard chromogenic immunohistochemistry to determine marker variability and inform practice of T cell biomarker assessment in formalin-fixed, paraffin-embedded (FFPE) tissue samples. Stained sections were digitally imaged and CD8+ lymphocytes within defined regions of interest (ROI) including the tumor and surrounding stroma were enumerated. Statistical analyses of CD8+ cell count variability using a linear model/ANOVA framework between patients as well as between levels within a patient sample were performed. Our results show that CD8+ T-cell distribution is highly homogeneous within a standard tissue sample in both colorectal and breast carcinomas. As such, cytotoxic T cell prevalence by immunohistochemistry on a single level or even from a subsample of biopsy fragments taken from that level can be considered representative of cytotoxic T cell infiltration for the entire tumor section within the block. These findings support the technical validity of biomarker strategies relying on CD8 immunohistochemistry.

Introduction

Breast and colorectal cancers are the second and third most common malignancies worldwide, respectively. Between 2008 and 2015, breast cancer incidence has increased by more than 20%, while mortality has increased by 14% and it now represents one in four of all cancers in women [1]. Colorectal cancer, at the same time, has decreased 3.2% per year in incidence although mortality has remained stable [2]. Although new chemotherapeutic regimens have

adherence to PLOS ONE policies on sharing data and materials.

Abbreviations: ANOVA, analysis of variance; FFPE, formalin-fixed, paraffin-embedded; FPR, false positive rate; ICC, intra-class correlation coefficient; ROC, receiver operator characteristic; ROI, region of interest; SD, standard deviation; TIL(s), tumor infiltrating lymphocyte(s); TPR, true positive rate; UICC, Union for International Cancer Control.

improved overall and disease-free survival in breast and colorectal cancer patients over the past decade, the benefits in overall survival are heterogeneous and the outlook remains suboptimal, particularly in colorectal cancer.

Disease progression in cancer patients is determined not only by the histologic and molecular features of the tumor but also by the host response, particularly the immune response. Observations that inflammatory cell density is higher in tumor tissue than in non-adjacent normal tissue in breast cancer patients [3] and that histologic distributions of CD4 and CD8 T cell subsets in, for instance liver metastases of colorectal cancer, can be correlated with clinical stage and outcome in colorectal cancer patients [4] emphasize that there is a topography to tumor control by the immune system. As such, much attention has been given to characterizing the immune contexture of tumors and determining the relationship of immune populations within the tumor microenvironment to clinical behavior, prognosis and therapeutic response. While associations between prevalence and outcome for some immune populations, such as B cells, remain unclear, the positive prognostic value of cytotoxic (CD8) tumor infiltrating lymphocytes (TIL) has been demonstrated in multiple solid tumor types including breast [5,6] and colorectal cancer [7,8,9] as well as melanoma [10,11,12,13], bladder [14,15], prostate [16,17], ovarian [18,19], pancreatic [20,21], and, when associated with proliferative index, renal cell carcinoma [22]. Studies in colorectal cancer have shown CD8+ and CD45RO+ cell prevalence and distribution to be superior to traditional UICC clinical staging in predicting colorectal cancer patient outcomes [8].

In addition to the prognostic impact of CD8 levels across multiple cancer indications, densities of intratumoral CD8+ T cells are associated with response to anti-PD-1 (pembrolizumab) treatment in melanoma and mismatch repair deficient cancers, including colon cancer [23,24]. Therefore, enumerating CD8+ T cell densities from archival tumor samples may serve as a predictive biomarker for melanoma and colon cancer patients who benefit from PD-1: PD-L1 blockade. Since CD8+ T cells are hypothesized to be the final effector cells that mediate tumor cell killing, changes in intratumoral CD8+ T cell dynamics serve as a pharmacodynamic marker of clinical activity in certain indications. Together, intratumoral CD8 densities are employed as a predictive and pharmacodynamic biomarker in melanoma and colon cancer, and a prognostic marker in multiple indications.

Although most studies show improved prognosis with increased tumor-associated cytotoxic T cells, some reports have demonstrated conflicting results. For non-small cell lung cancer, several studies associated increased CD8 with improved patient outcomes [25,26,27], but others showed no effect on survival [28], or decreased amounts CD8+ infiltrates have been correlated with improved survival [29]. In addition, CD8+ T cell infiltration and its association with survival in a renal cell carcinoma patients has not demonstrated significance [30,31]. While the heterogeneity of these findings is partly attributable to the variety of methods and endpoints assessed, the heterogeneity of immune infiltration in the tumor microenvironment is also potentially confounding. Examining the heterogeneity of PD-L1 staining and variability in interpretation, Rehman et al. have demonstrated that the variability in PD-L1 signal within a block is greater than between blocks from the same tumor and that one block is sufficient to represent the heterogeneity for a tumor sample [32]. As such, it is possible that significant misrepresentations of lymphocyte populations can occur in single routine histologic sections due to uneven distribution of immune cells throughout the tumor tissue within one block. This issue can be further complicated in evaluation of smaller tumor biopsies compared to full-section slides, where representation of the tumor and its immune contexture is even more limited.

When a biomarker such as CD8 is employed in a clinical setting, the effects of assay variability, whether attributable to analytical characteristics of the assay itself or due to pre-analytic

sample properties such as sample preparation and sample type (e.g., resection, slide, or core), could potentially have broader implications for clinical trial study design, study conduct, and the interpretation of results at the end of trial [33,34,35,36,37]. In practice, clinical trials, particularly those in earlier stages of drug development, are rarely powered to distinguish prognostic from predictive effects of a biomarker [35,37,38], and using an underperforming assay may only exacerbate this endeavor. In the case of an assay that dichotomizes a patient population into distinct subgroups of “diagnostic-negative” and “diagnostic-positive,” an assay with relatively reduced ability to accurately identify patients in a diagnostic-positive subpopulation may result in longer clinical trial recruitment and readout timelines due to a need for enhanced screening to enroll a sufficient number of diagnostic-positive patients. Inclusion of diagnostic-positive patients who benefit from a targeted therapy into the analysis of a diagnostic-negative subgroup may in some cases lead to an inflated estimate of treatment benefit in a patient subgroup where no treatment effect exists, thereby altering the overall perceived benefit-risk ratio of the therapy under evaluation. Finally, and perhaps most important for practitioners sizing clinical trials, differential misclassification of diagnostic-negative patients can dilute the treatment effect of a predictive biomarker in a diagnostic-positive subgroup, resulting in a net loss of power as more patients will be needed to detect the same amount of clinical benefit around which the trial was originally designed [39,40]. Power considerations are made even more difficult in the setting of a continuous biomarker or in any setting where the assay cut-off may be unknown [36,39,40].

In light of these practical considerations: what constitutes an adequate histologic evaluation of immune contexture, particularly in characterizing cytotoxic T cell infiltration of a tumor for prognosis, predicting response to immunotherapy, and for monitoring on-treatment pharmacodynamics? To inform practice of histologic T cell biomarker interpretation, we examined step sections of breast and colorectal cancer samples for CD8 T cell prevalence by immunohistochemistry, and performed *in silico* biopsy sampling of these sections to determine CD8 variability within a tumor block and between simulated biopsy samples. Numbers of CD8+ T cells were obtained by digital image analysis for each level and simulated biopsy. CD8 counts between levels of a tissue sample were fit to a linear variance model to determine the statistical significance of the CD8 count variability. Similarly, in simulated core biopsies, the degree of variability in CD8 counts was assessed and related to the overall ‘true’ block count to determine the number of cores necessary for adequate sampling. Our results show that CD8+ T cell numbers are highly homogenous throughout colon and breast cancer tissue section and that a single virtual core biopsy can potentially adequately represent tumor-associated cytotoxic T cell prevalence, though estimates improve with increasing sample number.

Materials and methods

Case selection and sectioning

Formalin-fixed paraffin-embedded (FFPE) breast cancer and colon cancer samples were procured. Breast cancer samples were procured from Avaden Biosciences and included both ductal (8) and medullary (4) carcinoma cases. Cases contained between 20% and 80% tumor in the regions analyzed by visual estimate of area. Patient ages for breast samples ranged from 40–82 years and AJCC stages I-IV were represented. Colorectal cancer samples were procured from Conversant Bio and included moderately differentiated adenocarcinoma (9), mucinous carcinoma (1) and rectal (2) carcinoma cases. Samples contained between 20% and 60% tumor in the regions analyzed by visual estimate of area. Patient ages ranged from 43–89 years, and stages I and II disease were represented.

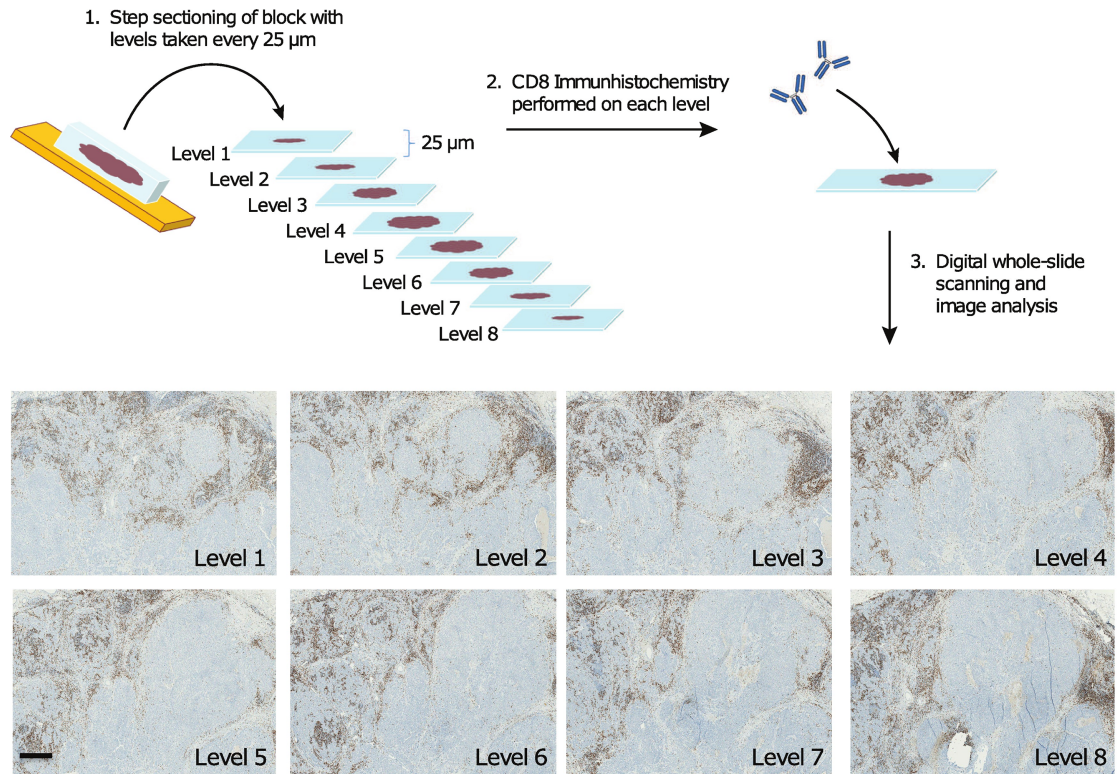


Fig 1. Sample workflow for immunohistochemistry and image analysis. Tumor blocks were sectioned and levels for CD8 immunohistochemistry taken at 25 µm intervals. Stained slides were scanned and the tumor area and immediately adjacent stroma was manually designated by a pathologist on all slides. All nucleated cells as well as CD8+ cells within the defined area were identified and counted by image analysis. Simulated core biopsies were identified by creating a grid of rectangular regions over the entire image, each approx. 2mm² in size. Rectangular regions that overlapped with at least 0.7 mm² of manually identified region were analyzed. Scale bar illustrated in “Level 1” panel equals 500 µm.

<https://doi.org/10.1371/journal.pone.0190158.g001>

Tissue blocks were serially sectioned and four serial levels were taken approximately every 10 µm until the block was exhausted or tissue significantly reduced yielding eight sets of 4 serial sections per case available for staining. For any given block, the second level of each set of serial sections was stained for CD8, leading to an assessment of CD8 prevalence every 25 µm (Fig 1).

Immunohistochemical staining

Immunohistochemistry (IHC) was performed on 4 µm thick formalin-fixed, paraffin-embedded tissue sections mounted on glass slides. All IHC steps were carried out on the Ventana Discovery XT automated platform (Ventana Medical Systems; Tucson, AZ). Sections were treated with Cell Conditioner 1, standard time, and then incubated in primary antibody, CD8 (C8-144B, Dako, cat. # M7103) at a working concentration of 0.157 mg/ml for 60 minutes at room temperature. Specifically bound primary antibody was detected by the OmniMap anti-rabbit HRP detection kit, followed by ChromoMap DAB (Ventana Medical Systems; Tucson, AZ). The sections were counterstained with Hematoxylin II (Ventana Medical Systems; Tucson, AZ), dehydrated, and cover-slipped. Positive staining controls were performed in parallel using human tonsil and negative controls were performed using IgG1 isotype antibody.

Digital imaging and analysis

Whole slide images were acquired with a Nanozoomer 2.0-HT automated slide scanning platform (Hamamatsu, Hamamatsu City, Shizuoka Pref., Japan) at 200x final magnification. Scanned slides were analyzed in the Matlab software package (version R2012b by Mathworks, Natick, MA) as 24-bit RGB images. Tumor regions were manually identified at a macroscopic level. Simulated core biopsies were identified by creating a grid of rectangular regions over the entire image, each approx. 2 mm² in size. Rectangular regions that overlapped with at least 0.7 mm² of manually identified region were analyzed. Individual cells were identified using an algorithm based on radial symmetry [41]. Each cell was then scored as positive or negative for DAB staining using a blue-normalized algorithm to identify brown pixels [42].

Statistical analyses

All statistical analyses were conducted in the R environment for statistical computing [43]. In addition to computing summary statistics (means, medians, ranges), CD8 staining data were also analyzed using a linear model framework both pooled across tissue types and within each tissue type individually. The data were first analyzed using a fixed effects linear model (ANOVA) with patient and section both treated as factors. In this manner, the contribution of handling section as a factor in the model could be assessed relative to the more parsimonious submodel in which section was not considered. The percent of variability in the model explained by the patient and section effects was estimated. As a second pass, variance component analysis was continued using a linear mixed effects model where a patient-level random effect shared by measurements (sections) taken on the same subject (tumor sample or block) was incorporated into a model for CD8+ percent staining (random intercept model). The intraclass correlation coefficient (ICC), bounded by 0 and 1, and estimated as the ratio of the variance of the patient random effect to the total variability in the model, was calculated as a summary measure for the similarity of sections taken from a given tumor sample.

Resampling methods were employed to answer questions about the utility of using a smaller number of sections for determining CD8+ status. The mean CD8-positive percent staining value across sections within a tumor sample was treated as the true expression value for that specimen, and the median value across all tumor samples was selected a candidate cut-off for CD8+ staining. Resampling was performed 1000 times, (i) by resampling smaller numbers of sections—e.g., one or two from the original 8—with replacement within the original, fixed set of 25 tumor samples to assess the effects of within-tumor heterogeneity, and (ii) by resampling with replacement first tumors and then sections within tumors to better assess the effects of biological and technical variability in determining CD8+ status in future samples of patient tumor blocks.

A total of 198 slides sectioned from the 25 blocks were further subdivided into simulated smaller biopsy fragments ("cores") of maximum size ~2 mm², each with a minimum tumor area of at least 33% (median number of simulated biopsy fragments per slide: 57, range: 2–161). A key question was how many simulated smaller biopsies would be needed to adequately approximate the CD8 IHC staining and assay performance of (i) the entire slide and (ii) the entire tumor block (pooled across slides). Sampling 1000 times without replacement, the mean CD8 percent positive staining was estimated using increasing numbers of cores (1–5 cores) and by taking either the mean or maximum of the subsampled biopsy fragments as an estimate summary measure. To allow for less-biased resampling without replacement of up to 5 cores, analysis (i) was restricted to 183 slides with at least 15 identified fragments (analysis-restricted median number of simulated biopsy fragments per slide = 59, range: 15–161, S1 and S2 Figs). The difference between the mean (of maximum) of the core biopsies selected at each

round of resampling and the mean (or maximum) of the cores *not* selected (the “out-of-bag” sample) was calculated as a summary measure. Local regression (loess) fits were also obtained in order to characterize the variability associated with sample fragments as a function of mean CD8 percent staining [44]. Assay performance was also evaluated in terms of accuracy, sensitivity and specificity of the summary measure estimates relative to the slide or block mean intensity (i.e., “truth”) across a variety of potential candidate cut-offs for CD8-positive percent staining, e.g., 1%, 2%, 5% and 10% positivity.

Results

Prevalence and distribution of CD8+ T cell infiltrates in breast and colon cancer cases

In order to determine whether sampling error contributes to variation in intratumoral T cell counts, we sectioned through and enumerated CD8+ T cells in 13 primary colorectal adenocarcinoma and 12 breast carcinoma (6 ductal adenocarcinoma, 6 medullary carcinoma) cases (1 FFPE block per case). Step sections of breast and colorectal cancer samples were taken at approximately 25 μm intervals from FFPE samples and stained for CD8 using DAB-based chromogenic immunohistochemistry (Fig 1). Image analysis was performed on manually identified tumor regions, which included tumor margins. Samples were manually evaluated for adequacy and no significant background staining or labeling of non-lymphocyte cell populations were noted on any CD8 immunohistochemistry slide. CD8+ cells were identified and calculated as a percentage of total cells identified by hematoxylin counterstain.

From the 12 breast cancer cases, 3 were noted to have significant tissue loss on deeper levels. Specifically, deepest sections (level 8) from two cases showed less than 40% of tissue remaining compared to the initial level. A separate case showed approximately 30% of tissue compared to the initial level on levels 7 and 8. Only one colon cancer sample showed significant tissue loss on one of eight levels. Since these changes could potentially introduce artifactual variability, statistical analyses that both included and excluded these levels were run.

Breast cancer samples showed a wide range of cytotoxic T cell infiltration varying from less than 1% (0.41%, ductal adenocarcinoma) to greater than 50% (55.9%, medullary carcinoma) marker-positive lymphocytes. Histology showed diffuse T-cell infiltration with most breast cancer cases showing CD8+ cells in comparable amounts at the tumor edge and center. Medullary carcinoma cases showed characteristically high lymphocyte infiltration and high CD8+ area (mean 24.0%, median 19.7%, pooled inpatient SD: 1.7%), compared to ductal adenocarcinoma (mean 8.8%, median 5.9%, pooled inpatient SD: 1.2%).

Colorectal cases showed, in general, somewhat less cytotoxic T cell infiltration than breast cancer cases (mean 3.8%, median 2.9%, inpatient SD: 0.4%). The majority of CRC cases (12/13) showed less than 10% CD8+ T cells on any level in contrast to breast cancer samples where 50% (6/12) of cases showed less than 10% CD8+ prevalence. The maximum amount of CD8+ T cells was 13.6%. In contrast to breast cancer tissues, CD8+ T cells in colon cancer cases were confined largely the stroma with few infiltrating lymphocytes among epithelial cells.

CD8 IHC assay performance characteristics

Assessment of full-face sections showed greater variability in CD8+ T cell prevalence between cases than between levels from an individual case. This observation also held when cases were divided by histologic type, such that cytotoxic T cell prevalence on serial sections varied more greatly between cases of ductal adenocarcinoma than within a case of ductal adenocarcinoma

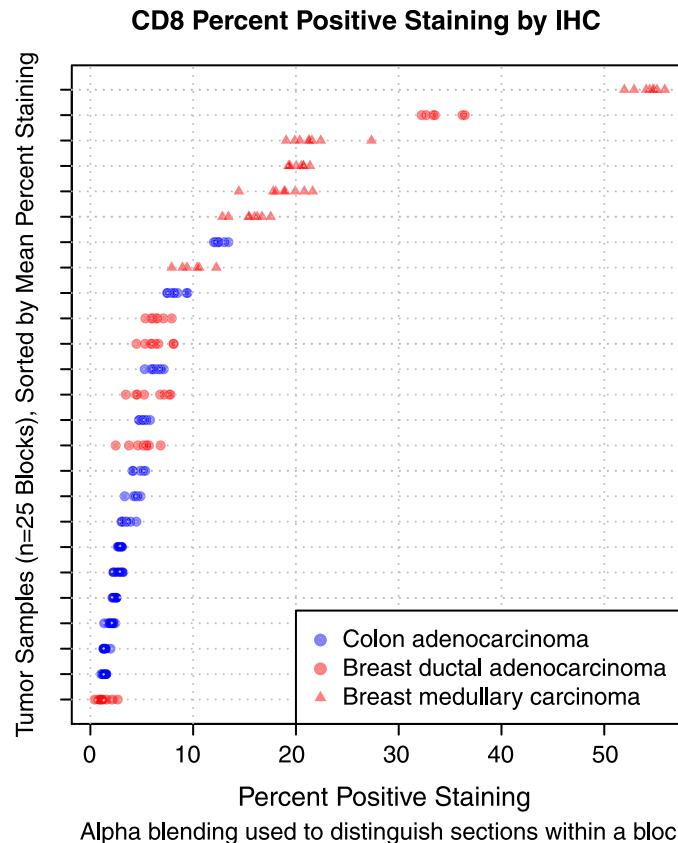


Fig 2. CD8 IHC repeated measures staining results for 25 breast (red) and colorectal (blue) carcinoma samples. Each tumor block was sectioned 8 times or until the sample was exhausted.

<https://doi.org/10.1371/journal.pone.0190158.g002>

(Fig 2). However, minimal variation in cytotoxic T cell counts between levels within any given tumor block was observed over the entire observed dynamic range of CD8 staining (Fig 2). In a pooled analysis of all colon and breast samples, a linear model was fit using tumor block and slide section as explanatory variables for percent staining. Patient variability at the tumor block level accounted for 99.2% of the (biological) variability in the model. Even when the two most extreme tumor samples containing CD8+ expression > 30% were removed from the analysis (Fig 2), the patient-level effect still explained 95.8% of the total variability in the model. The effect of section variability failed to reach statistical significance in any analyses subgroup explored. The ICC calculated for percent CD8+ cells between sections within a tumor sample in the linear mixed effects model variance component analysis was 0.99. In order to characterize the variability within tumor type, separate models were also fit stratified by tumor type, and for the breast samples, within histologic subtype. From this analysis, percent variability (%) in CD8 counts between step sections from colonic adenocarcinoma (0.2%), medullary breast carcinoma (0.1%) and ductal adenocarcinoma (0.1%) blocks were minor. Together, these results suggest that variation in CD8 counts between cases for breast or colon samples is driven largely by interpatient biological effects and not due to inadequate sampling.

To determine whether a single slide is sufficient to assess the status of CD8+ T cell abundance, we next assessed the performance of this assay benchmarked against a median cut-off. In the resampling-based analysis, pooling breast and colon samples, a median (50th percentile)

cutoff of 5.9% percent CD8+ staining was calculated for determining a patient's CD8+ status. With a patient's average CD8 expression computed across all sections being considered "truth," patients were classified as CD8-high or CD8-low depending on whether or not their mean expression was above or below the median cut-off determined by the larger sample. Relative to the CD8 status determined by the mean CD8 expression, classifying patient samples by taking even just one slide at random resulted in 94.8% (95% CI: 88.0% -100%) positive agreement (concordance), with 94.5% simulated sensitivity in detecting "CD8-high" patients and 95.1% simulated specificity in detecting "CD8-low" patients (S3 Fig). Similar results were obtained when resampling both patients and sections within patients [95.2% positive agreement (95%CI: 84.0%-100%); 94.9% sensitivity and 95.5% specificity] (S3 Fig). No appreciable differences in assay performance were observed when significant tissue loss was excluded versus when they were included (data not shown). Therefore, for classifying tumor blocks using CD8 counts, the usage of an individual slide is not predicted to lead to misclassification due to insufficient tumor block sampling.

Compared to archival tumor blocks utilized for prognosis and predictive diagnostic studies, needle biopsies acquired for post-treatment pharmacodynamics analysis represent a smaller sample of the entire tumor, and may be prone to technical variation (S1 and S2 Figs). In practice, most investigators and patients are willing to provide a limited number of core biopsies, and not all of those core biopsies may be designated to measure the same biomarker. In order to assess the cost, as measured by variability, associated with using a core biopsy sample instead of a tumor cross-section, we performed virtual biopsy sampling from within slides and compared biopsy-based CD8+ cell prevalence to slide- and block-based measurements of staining. Rectangular fields of view of between 0.5 and 2.0 mm² were digitally generated within manually selected tumor regions and the percentage of CD8+ cells were calculated for each field. Surprisingly, sampling just 1–5 biopsy fragments from an individual slide, and across slides, recapitulated CD8 staining observed both at the slide and tumor block level, respectively (Figs 3 and 4). Using the set of cores for a slide as a proxy for parent data with known mean and variance, sampling just one biopsy led to an estimate of CD8 staining that was within 1 standard deviation (SD) of the total slide staining in greater than 70% of simulations for 141 of the 183 slides analyzed (77.0%) and in over 80% of simulations for 100 slides (Fig 3A). Taking the mean of two biopsies produced estimates of CD8 staining that were within 1SD of the slide staining intensity in 182 slides (99.4%) in over 80% of simulations, with 42 (23.0%) of slides landing within the 1 SD range in over 90% of simulations (Fig 3A). Incremental performance gains were observed with more intensive sampling, with gains after sampling three cores becoming less pronounced (Fig 3A and 3C).

To mitigate against bias when sampling a subset of cores from a finite set of fragments, we estimated the differences in means of the fragments selected at each round of sampling with the means of the fragments not selected (the "out-of-bag" sample), and were thus able to approximate the standard deviations associated with sampling the subsets of cores relative to the rest of the tumor sample. Differences in means were centered around zero (unbiased), with decreases in variability again observed with increased sampling (Fig 3B) and incremental gains in performance observed after sampling three cores. The standard deviations were also plotted as a function of mean CD8 staining for the full set of cores and fitted with loess smoothers to highlight that the improvements in performance were observed over the full dynamic range of CD staining (Fig 3C). Similar trends were detected when using the same numbers of fragments to estimate CD8 staining pooled over all slides for each of the 25 blocks (Fig 4A and 4C).

While it may not be surprising that with increased sampling, the mean of an increasing number of core biopsies will approach the value which was measured on its parent slide or in an entire block with "square root n "-like convergence, some researchers may alternatively use

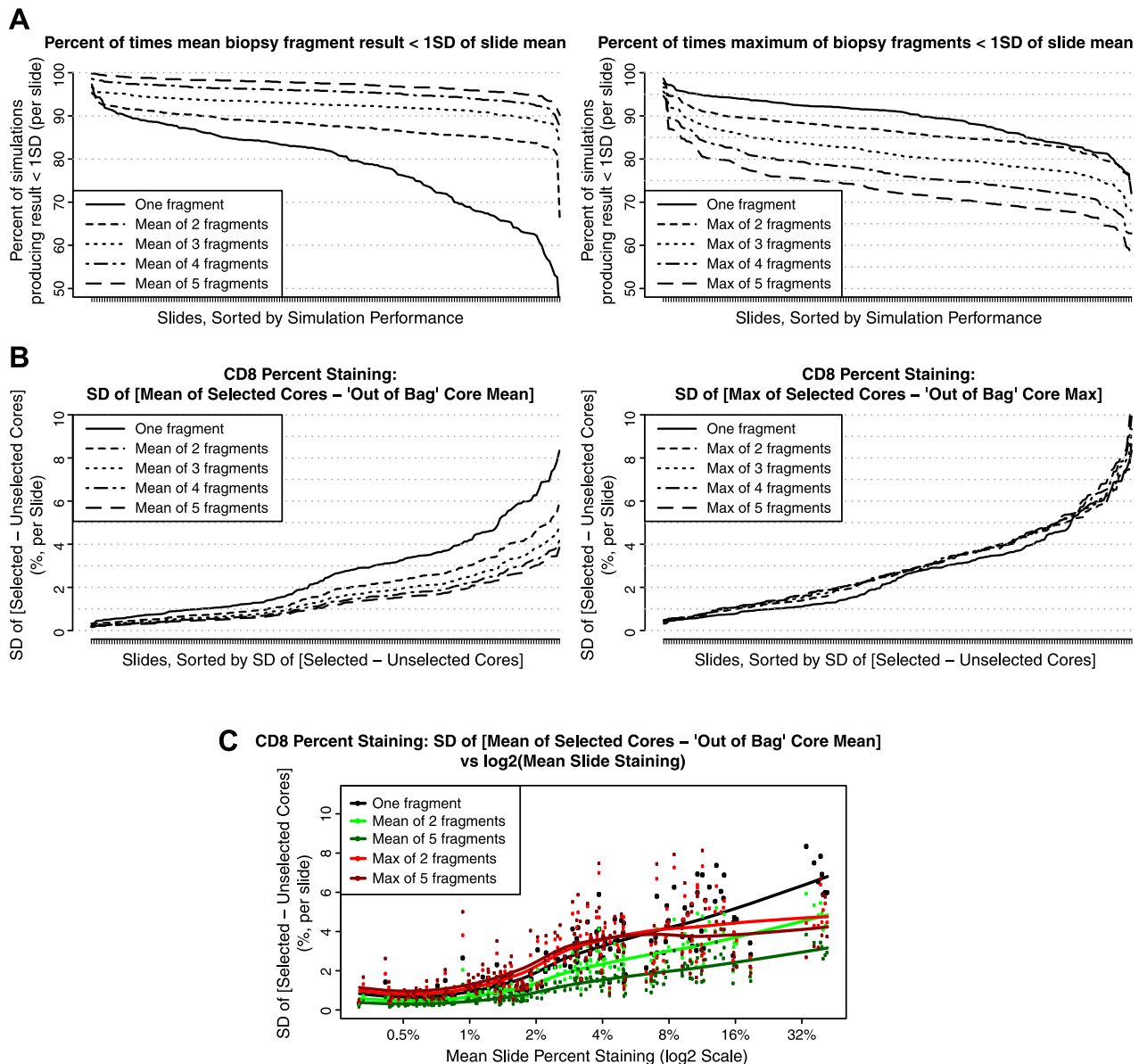


Fig 3. Slide-level biopsy simulation results. (A) Percent of times over 1000 rounds of simulations that the values obtained from sampling increased numbers of biopsy fragments produced a result within 1SD of the mean CD8 staining for a given slide. Calculating the mean over increased numbers of biopsies led to better estimates of the mean, while calculating maxima over the sample biopsies led to overestimates of a slide's CD8 levels. Samples are sorted by increasing performance in terms of being able to produce an estimate within 1SD of total CD8 staining for that slide. (B) Estimates of the standard deviation of the difference between the mean or maximum of selected core biopsies and the mean or maximum of the out-of-bag or unselected cores on a given slide. Increased sampling leads to improvements in variability for when using means but not when using order statistics. (C) Estimates of the standard deviation of the difference between the mean or maximum of selected core biopsies and the mean or maximum of the out-of-bag or unselected cores on a given slide as a function of the mean CD8 percent positive staining for a slide. Loess fits are used to highlight mean performance over the observed dynamic range on both the log₂ and observed percent staining scales. Increased sampling leads to improvements in variability when using means but not when using order statistics.

<https://doi.org/10.1371/journal.pone.0190158.g003>

the maximum percent staining value obtained on from a small number of biopsies as a summary measure. Our results demonstrate that the use of order statistics such as maxima should be carefully considered or avoided altogether if one is trying to get a representative assessment of overall staining within a tumor or individual. Using the maximum over increasing numbers of biopsies resulted in biased overestimates of CD8 staining relative to the mean CD8 positivity

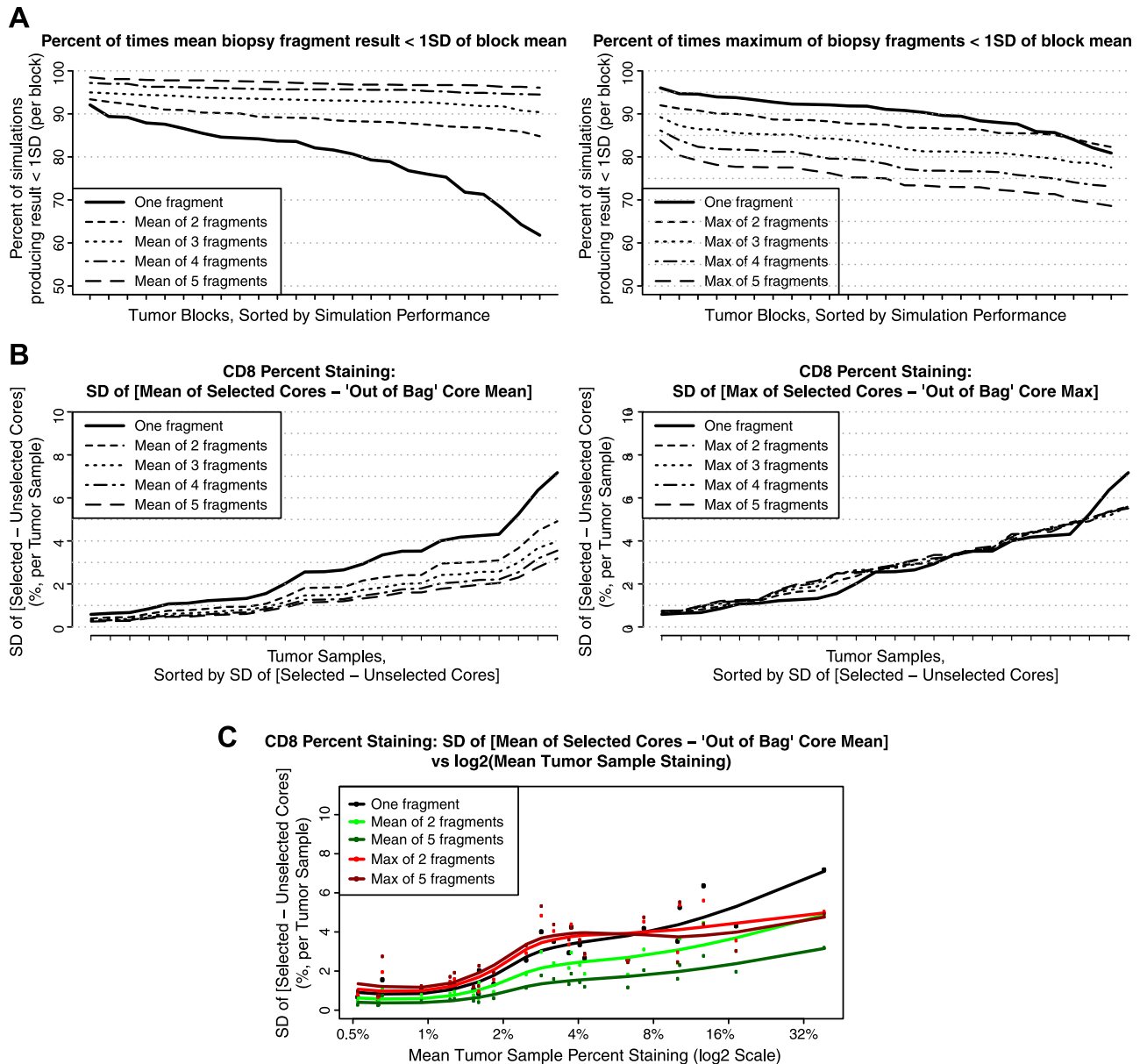


Fig 4. Block-level biopsy simulation results. (A) Percent of times over 1000 rounds of simulations that the values obtained from sampling increased numbers of biopsy fragments produced a result within 1SD of the mean CD8 staining for an entire tumor block. Calculating the mean over increased numbers of fragments led to better estimates of the mean, while calculating maxima over the sample biopsies led to overestimates of a sample's CD8 levels. Samples are sorted by increasing performance in terms of being able to produce an estimate within 1SD of total CD8 staining for that tumor block. (B) Estimates of the standard deviation of the difference between the mean or maximum of selected core biopsies and the mean or maximum of the out-of-bag or unselected cores for a given tumor sample. Increased sampling leads to improvements in variability for when using means but not when using order statistics. (C) Estimates of the standard deviation of the difference between the mean or maximum of selected core biopsies and the mean or maximum of the out-of-bag or unselected cores for a given tumor sample as a function of the mean CD8 percent positive staining for that block. Loess fits are used to highlight mean performance over the observed dynamic range on both the \log_2 and observed percent staining scales. Increased sampling leads to improvements in variability when using means but not when using order statistics.

<https://doi.org/10.1371/journal.pone.0190158.g004>

and in estimates which rapidly diverged from the mean +/- 1SD range of CD8 staining of the slide or block (Figs 3A and 4A). In addition to producing biased estimates, the use of the maximum did not result in improvements in variability relative to the out-of-bag mean or maximum of the unselected cores (Figs 3B, 3C, 4B and 4C)

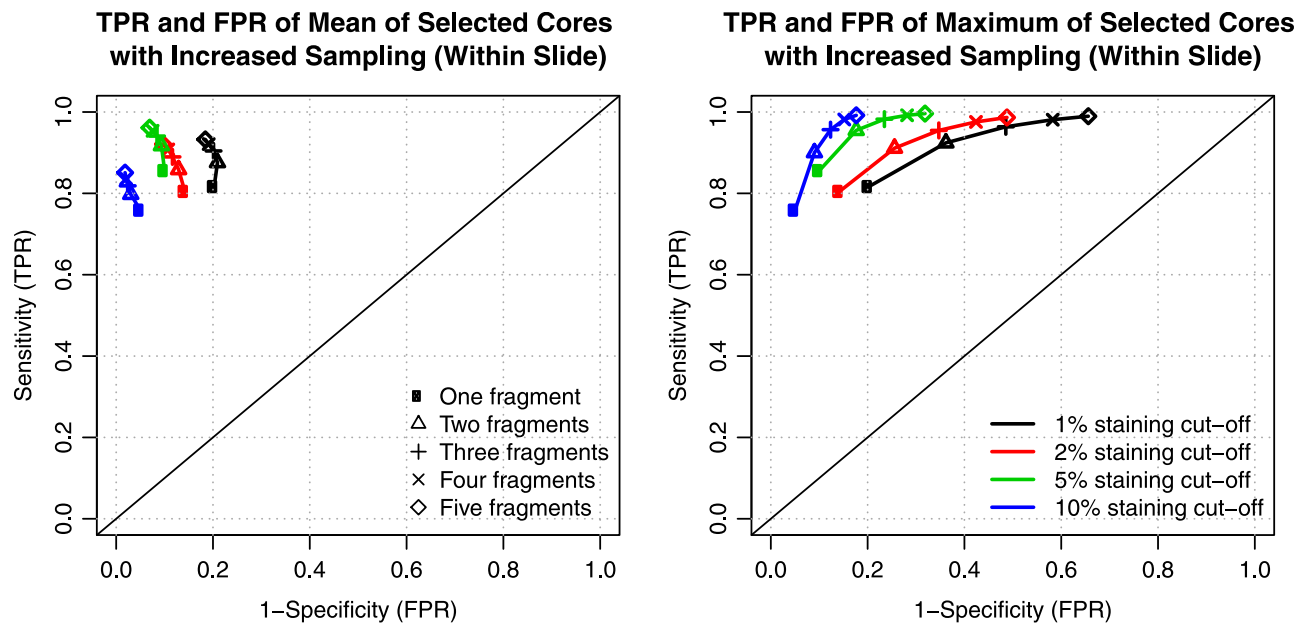


Fig 5. Simulation receiver-operator characteristic (ROC) curves. With increased sampling, there is increased performance of the CD8 IHC assay to classify as positive or negative for a given cut-off (different colors) when using the means of different numbers of sampled biopsies (left) and decreased performance when using maxima (right). Over the set of cut-offs (1%, 2%, 5%, 10%), the analysis treated the mean CD8 staining as the true intensity against which the results of the subsamples of biopsy fragments were evaluated. Similar results were obtained when benchmarking the staining result from core biopsies against the mean staining of the entire block (data not shown).

<https://doi.org/10.1371/journal.pone.0190158.g005>

In addition to characterizing the bias and variability over the range of CD8 staining in our sample, receiver-operator characteristic (ROC) curves were used to display assay performance of the CD8 staining estimates obtained from the biopsies over a range of potential candidate staining cut-offs that could be considered in a clinical trial setting (1%, 2%, 5% and 10%). Here, the mean CD8 staining of a slide or block was used as the gold standard or "truth" against which CD8 positivity or negativity as determined by a set of simulated core biopsies was evaluated for a given cut-off. In terms of being able to correctly classify patients as being CD8-positive or CD8-negative, a high level of sensitivity and specificity was observed when comparing results from a single biopsy fragment to a sample's status based on evaluation of the whole slide or block. For example, using a 5% cut-off, relative to the result obtained from a slide, the value for the biopsy fragment displayed on average 85.6% sensitivity and 89.9% specificity over 1000 rounds of simulation (Fig 5). Taking the mean of two biopsy fragments increased the sensitivity to 91.7% while maintaining specificity at 90.7% (Fig 5). Over the set of candidate cut-offs examined, as evidenced by movement towards the upper left corners of the ROC plots, increased sampling of cores led to increases in sensitivity and specificity of the biopsies' ability to classify slides as CD8-positive or CD8-negative when using the mean of the biopsies as a summary measure (Fig 5). As expected, taking the maximum of the biopsies resulted in an overestimation of the sample's mean CD8 staining, leading to higher occurrences of false positives (Fig 5). Together, our data suggest that although increased number of core biopsies resulted in enhanced CD8 assay sensitivity a single core biopsy is sufficient to represent CD8 levels, regardless of whether median or candidate cutoff is employed.

Discussion

Predictive diagnostics for cancer immunotherapies have focused on the tumor-immune microenvironment and its composition. As such, histologic assessment of tumors and their

immune contexture is becoming central to prognosticating clinical behavior in patients with solid tumors both by determining the amount of immune infiltrate and its composition, as well as prevalence of actionable immunotherapy targets, e.g. PD-L1, and prognostically relevant immune cell subtypes, including CD8+ T cells. However, variable sampling and interpretation methods including scoring algorithms and inter-pathologist variability can potentially confound histology as a prognostic method. As well, interpretation of more limited samples, such as core biopsies, raises questions about sampling adequacy and error. However, the extent and potential effect(s) of sample type and spatial heterogeneity of tumor immune infiltrates has not been detailed. The clinical and scientific importance of TIL assessment in breast cancer has been underscored by recent efforts to standardize histologic interpretation of TILs in patient samples, but it has been noted that TIL heterogeneity and adequacy of core biopsy samples versus tissue sections has not been formally characterized [45].

CD8+ cytotoxic T cells in particular have received attention not only because of their known role as cytolytic agents and demonstrated reactivity to tumor-derived self-epitopes [46] but also their prevalence and apparent positive prognostic effect in breast cancer as well as other tumor types. In particular, the presence of CD8+ cells in the tumor prior to the onset of neoadjuvant chemotherapy has been shown to predict response to therapy including pathological complete response particularly in triple-negative and HER2-expressing breast cancers [47,48,49,50]. In colon cancer, immunohistochemical assessment of CD3+ T cells and CD8+ T cells in the tumor microenvironment has been shown to more accurately predict patient outcome than traditional UICC staging and its utility as an adjunct to traditional UICC staging has been proposed [51]. Additionally, in colon cancer patients with microsatellite instability, CD8+ T cell density predicts response to anti-PD-1 treatment [24]. Given these findings and the lack of published evidence documenting the extent of TIL heterogeneity in whole tumor sections or biopsies, we evaluated CD8+ T cell prevalence in step sections from breast and colon tumor blocks and simulated core biopsy samples to determine the variability of tumor-associated CD8+ T cell prevalence between levels within a block and between simulated core biopsies and whole tissue sections as well as block means.

Our results demonstrate that the prevalence of CD8+ T cells is highly consistent within a standard sample block. Comparing whole slide counts of step sections from a given tumor block, CD8+ T cell percentage varied within 8% between all levels of a given breast cancer sample and within 2% between all levels of a given colorectal cancer sample. Also, we find, based on a linear-effects model and variance component analysis, that of any observed variability across all samples or within a tumor type, less than 1% can be attributed to section-to-section variability. We conclude that, for this particular IHC assay, CD8+ cell prevalence on a single slide can be representative of the CD8+ status of the entire block.

While evaluation of CD8 prevalence on a standard slide may generate consistent results, the adequacy of biopsy samples for evaluation of immune cell prevalence has not been extensively characterized. Biopsy sampling, particularly core biopsy sampling, has traditionally been examined in the context of solid tumor diagnosis and phenotyping. Additionally, core biopsies are routinely employed for assessing changes in immune cell numbers and activity in immunotherapy studies. In breast cancer, core biopsies have been shown to correlate poorly with whole slide sections for assessment of some histologic markers such as Ki-67 but well with others such as ER and HER2 [52]. As well, standard colon biopsies have been shown to harbor neoplastic lesions in deeper block levels in a minority of cases [53] and core biopsy samples of metastatic colorectal cancer lesions have shown significant molecular heterogeneity from the parent lesion, consistent either with tumor evolution or sampling bias [54]. By random *in silico* designation and analysis of 2 mm² sections of tumor area to simulate biopsy sampling, we found that (1) one biopsy sample can approximate CD8 prevalence to within one standard

deviation of the block mean in greater than 75% of all analyzed cases, (2) using multiple cores improves performance up to three cores (best), and (3) the mean CD8+ cell percentage for multiple biopsies was a less variable and less biased predictor of CD8 status for a given tumor sample than the maximum value. For three of four cases in which a single simulated core biopsy failed to estimate the CD8 prevalence to within 1 standard deviation of the corresponding slide or block, a heterogeneous distribution of CD8+ cells, concentrated primarily at the tumor margins, was noted in each slide of each case. Significant tissue loss was noted in one remaining case. This suggests that spatial heterogeneity can potentially confound interpretation of biopsies for immune cell prevalence but only in a minority of cases. However, we find that additional biopsy samples can significantly improve accuracy. For instance, where single core simulations in one case generated estimates to within 1 SD in only approximately 65% of instances, addition of a second core and calculation of mean CD8+ cell percentage generated accuracy to within 1 SD of the slide mean in approximately 85% of instances. Addition of a third core further improved accuracy to over 90%. Such improvements were observed in all cases where single biopsy simulations were accurate to within 1 SD less than 80% of the time.

However, if multiple biopsies are available for interpretation, is calculation of a mean CD8+ cell percentage the best for interpretation? We explored the utility of taking the maximum CD8+ cell percentage value obtained from multiple cores but found calculation of mean CD8+ cell percentage to be more accurate. With increasing numbers of samples, the mean CD8+ cell percentage reduces variability over the dynamic range of the assay and improves sensitivity with minimal change in a low diagnostic false positive rate across multiple diagnostic cut-offs (1%, 2%, 5%, 10%). As such, while mean CD8+ cell percentage is a highly specific method, sensitivity can be reduced at higher diagnostic cutoffs. Conversely, when using maximum CD8+ cell percentage, increasing the number of samples analyzed improves sensitivity but dramatically reduces specificity across all cut-offs. These relationships hold true for biopsy sample estimates relative to both corresponding slides as well as blocks.

Since samples for the current study were procured, the availability of tumor sections of adequate amount and integrity was limited and only less than 15 cases for each tumor type could be evaluated. Expanding the number of samples could better characterize any potential variability in step sections or biopsy samples. However, given the high degree of consistency in the existing samples, we speculate that any further differences would be minor and that the limited number of samples in the current study does not significantly detract from the results. Also, in the current study we could only examine one block from a given tumor. While the degree of cytotoxic T cell heterogeneity was not significant within a given block, we cannot make conclusions regarding the homogeneity of CD8 infiltration within an entire tumor. At the time of this writing, the degree of histologic CD8 heterogeneity from standard pathology sampling of an entire tumor specimen has not been documented. Performing the current analysis across multiple blocks/sections from a tumor would add further insight into the prognostic value of CD8 immunohistochemistry in tumors as well as potentially inform sampling methods by pathology services for routine histologic analysis. In addition, our methods may be more widely applicable and inform development of other tissue-based biomarkers. For instance, Foxp3+ cell prevalence in tumor samples either alone or in ratio to CD8+ cell counts has gained attention as a potential tissue-based prognostic marker in multiple tumor types. Studies in breast cancer have documented improved pathologic complete remission (pCR) and disease-free survival in HER2+ or triple negative breast cancer (TNBC) patients with CD8+ to Foxp3+ cell ratio (CFR) greater than 1 when assessed by immunohistochemistry [55]. As well, improved recurrence-free survival and breast cancer specific survival in TNBC patients without pCR has been documented [56]. For colon cancer, prognostic value of the CFR has been variable with some studies citing improved overall (OS) (Suzuki) and/or disease-free (DFS)

survivals [57, 58] but others finding no association between the CFR and outcome [59]. In addition, immunohistochemical studies examining the prognostic significance of Foxp3+ cell [prevalence alone have been similarly variable finding either no association with outcome [58] or improved outcome for increased numbers of Tregs in tumor tissue and impaired outcome for increased numbers of Tregs in normal mucosa [60]. The variability of these results is attributable to multiple factors including different patient populations, biopsy type, disease stages, treatments, and metastatic status. The utility of Foxp3 as a Treg-specific marker has also been questioned since it has been shown to be expressed in non-suppressive activated T cells [61]. However, studies have not currently ruled out the possibility that spatial heterogeneity of FoxP3+ cells in tumor samples may also be a contributing factor. Histologic and digital image analyses similar to those we describe could clarify these findings and inform use of the CFR as a prognostic marker.

Despite the limitations associated with the number of tumor samples included in this analysis, our results highlight the practical importance of understanding *a priori* the analytical performance of any biomarker assay that may be deployed in clinical practice. An assessment of assay performance can also include the effect of biopsy type (resections, slides or core biopsies) on the results of the assay. Additionally, information about assay performance can also be included or accounted for when designing a clinical trial. While variability over the dynamic range of the assay in a patient population may be informative, it is of particular interest to understand assay performance around a given cut-off. Simulations can be run in order to assess if the assumed biomarker-treatment effect relationship of a therapy being interrogated is robust to assay performance issues given the cut-off and the range of expression in the patient population of interest. Once the clinical trial is ongoing, (blinded) assay data may also be evaluated in order to ensure that biomarker prevalence estimates and assay performance is as expected as during the design stages of the clinical trial and to guide potential risk mitigation strategies during enrollment.

Conclusions

Taken together, our findings show that CD8+ cell prevalence is highly homogeneous within a standard tissue sample and that evaluation of any slide from a given block can be considered representative of the CD8+ cell prevalence. Also, through *in silico* biopsy modeling, we find that immune cell prevalence estimates can be accurately represented with just one biopsy sample and that accuracy improves with increasing sample number when the mean CD8+ cell percentage among the biopsy samples is assessed. These findings not only support the technical validity of biomarker strategies relying on CD8 immunohistochemistry but also suggest standards for sample number adequacy and interpretation.

Supporting information

S1 Fig. CD8 cores by slide. CD8 IHC repeated measures staining results for simulated core needles biopsies on 183 slides sectioned from 25 breast (red, left) and colorectal (blue, right) carcinoma samples.

(EPS)

S2 Fig. CD8 cores by patient. CD8 IHC repeated measures staining results for simulated core needles biopsies for 25 breast (red) and colorectal (blue) carcinoma samples.

(EPS)

S3 Fig. CD8 sampling and classification. Violin plots of overall percent agreement between CD8-positive and CD8-negative calls at the tumor (block) level when (i) using the original set

of samples and resampling sections only (top), and (ii) when resampling both patients and then sections within patient (bottom). Scenario (i) captures technical reproducibility in this data set, while scenario (ii) attempts to capture variability in agreement calls in future samples of the same size. Treating the block-level mean as the gold standard "true" value of CD8 expression for that sample, using even just one slide (as opposed to averaging or taking the max of two slides) resulted in high agreement when applying a median (50th percentile) cut-off. Red points indicate a subsample of results over rounds of simulation. (EPS)

Acknowledgments

The authors wish to thank Lisa Damico-Beyer, PhD and Ina Rhee, MD, PhD for their thoughtful review of this manuscript and suggestions.

Author Contributions

Conceptualization: Oded Foreman, Mahrukh Huseni, Jeong M. Kim.

Data curation: James Ziai, Houston N. Gilbert.

Formal analysis: James Ziai, Houston N. Gilbert.

Methodology: Houston N. Gilbert, Jeffrey Eastham-Anderson, Felix Chu.

Software: Jeffrey Eastham-Anderson.

Supervision: Jeong M. Kim.

Validation: Jeffrey Eastham-Anderson, Felix Chu.

Writing – original draft: James Ziai, Jeong M. Kim.

Writing – review & editing: James Ziai, Houston N. Gilbert, Oded Foreman, Jeffrey Eastham-Anderson, Mahrukh Huseni, Jeong M. Kim.

References

1. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11
2. <http://seer.cancer.gov/statfacts/html/colorect.html>
3. Ruffel B, Au A, Rugo HS, Esserman LJ, Hwang ES, Coussens LM. Leukocyte composition of human breast cancer. *Proc. Natl Acad. Sci. USA* 2012 Feb 12; 109(8):2796–2801 <https://doi.org/10.1073/pnas.1104303108> PMID: 21825174
4. Halama N, Michel S, Kloor M, Zoernig I, Benner A, Spille A, et al. Localization and density of immune cells in the invasive margin of human colorectal cancer liver metastases are prognostic for response to chemotherapy. *Cancer Res.* 2011; 71:5670–5677 <https://doi.org/10.1158/0008-5472.CAN-11-0268> PMID: 21846824
5. Mahmoud S, Lee A, Ellis I, Green A. CD8+ T lymphocytes infiltrating breast cancer: A promising new prognostic marker? *Oncoimmunology.* 2012 May 1; 1(3):364–365 <https://doi.org/10.4161/onci.18614> PMID: 22737616
6. Liu S, Lachapelle J, Leung S, Gao D, Foulkes WD, Nielsen TO. CD8+ lymphocyte infiltration is an independent favorable prognostic indicator in basal-like breast cancer. *Breast Cancer Res.* 2012 Mar 15; 14(2):R48 <https://doi.org/10.1186/bcr3148> PMID: 22420471
7. Mlecnik B, Tosolini M, Kirilovsky A, Berger A, Bindea G, Meatchi T, et al. Histopathologic-based prognostic factors of colorectal cancers are associated with the state of the local immune reaction. *J Clin Oncol.* 2011 Feb 20; 29(6): 610–8 <https://doi.org/10.1200/JCO.2010.30.5425> PMID: 21245428
8. Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pagès C, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* 2006; 313:1960–1964 <https://doi.org/10.1126/science.1129139> PMID: 17008531

9. Pagès F, Kirilovsky A, Mlecnik B, Asslaber M, Tosolini M, Bindea G, et al. In situ cytotoxic and memory T cells predict outcome in patients with early-stage colorectal cancer. *J Clin Oncol* 2009; 27:5944–5951 <https://doi.org/10.1200/JCO.2008.19.6147> PMID: 19858404
10. Clemente CG, Mihm MC Jr, Bufalino R, Zurrida S, Collini P, Cascinelli N. Prognostic value of tumor infiltrating lymphocytes in the vertical growth phase of primary cutaneous melanoma. *Cancer* 1996; 77:1303–1310 PMID: 8608507
11. Mackensen A, Ferradini L, Carcelain G, Triebel F, Faure F. Evidence for in situ amplification of cytotoxic T-lymphocytes with antitumor activity in a human regressive melanoma. *Cancer Res.* 1993; 53:3569–3573 PMID: 8339262
12. Tefany FJ, Barnetson RS, Halliday G. M., McCarthy SW, McCarthy WH. Immunocytochemical analysis of the cellular infiltrate in primary regressing and non-regressing malignant melanoma. *J. Invest. Dermatol.* 1991; 97:197–202 PMID: 1712819
13. van Oijen M, Bins A, Elias S, Sein J, Weder P, de Gast G, et al. On the role of melanoma-specific CD8+ T-cell immunity in disease progression of advanced-stage melanoma patients. *Clin Cancer Res* 2004; 10:4754–4760 <https://doi.org/10.1158/1078-0432.CCR-04-0260> PMID: 15269149
14. Nakakubo Y, Miyamoto M, Cho Y, Hida Y, Oshikiri T, Suzuoki M, et al. Clinical significance of immune cell infiltration within gallbladder cancer. *Br. J. Cancer* 2003; 89:1736–1742 <https://doi.org/10.1038/sj.bjc.6601331> PMID: 14583778
15. Sharma P, Shen Y, Wen S, Yamada S, Jungbluth AA, Gnjatic S, et al. CD8 tumor-infiltrating lymphocytes are predictive of survival in muscle-invasive urothelial carcinoma. *Proc. Natl Acad. Sci. USA* 2007; 104:3967–3972 <https://doi.org/10.1073/pnas.0611618104> PMID: 17360461
16. Karja V, Aaltomaa S, Lipponen P, Isotalo P, Talja M, Mokka R, et al. Tumour-infiltrating lymphocytes: a prognostic factor of PSA-free survival in patients with local prostate carcinoma treated by radical prostatectomy. *Anticancer Res.* 2005; 25:4435–4438 PMID: 16334122
17. Richardsen E, Uglehus R.D, Due J, Busch C, Busund L T. The prognostic impact of M-CSF, CSF-1 receptor, CD68 and CD3 in prostatic carcinoma. *Histopathology* 2008; 53:30–38 <https://doi.org/10.1111/j.1365-2559.2008.03058.x> PMID: 18510570
18. Sato E, Olson SH, Ahn J, Bundy B, Nishikawa H, Qian F, et al. (2005) Intraepithelial CD8+ tumor-infiltrating lymphocytes and a high CD8+/regulatory T cell ratio are associated with favorable prognosis in ovarian cancer. *Proc Natl Acad Sci U S A* 102:18538–18543 <https://doi.org/10.1073/pnas.0509182102> PMID: 16344461
19. Hamanishi J, Mandai M, Iwasaki M, Okazaki T, Tanaka Y, Yamaguchi K, et al. Programmed cell death 1 ligand 1 and tumor-infiltrating CD8+ T lymphocytes are prognostic factors of human ovarian cancer. *Proc. Natl Acad. Sci. USA* 2007; 104:3360–3365 <https://doi.org/10.1073/pnas.0611533104> PMID: 17360651
20. Hiraoka N, Onozato K, Kosuge T, Hirohashi S. Prevalence of FOXP3+ regulatory T cells increases during the progression of pancreatic ductal adenocarcinoma and its premalignant lesions. *Clin. Cancer Res.* 2006; 12:5423–5434 <https://doi.org/10.1158/1078-0432.CCR-06-0369> PMID: 17000676
21. Fukunaga A, Miyamoto M, Cho Y, Murakami S, Kawarada Y, Oshikiri T, et al. (2004) CD8+ tumor-infiltrating lymphocytes together with CD4+ tumor-infiltrating lymphocytes and dendritic cells improve the prognosis of patients with pancreatic adenocarcinoma. *Pancreas* 28:e26–e31 PMID: 14707745
22. Nakano O, Sato M, Naito Y, Suzuki K, Orikasa S, Aizawa M, et al. Proliferative activity of intratumoral CD8(+) T-lymphocytes as a prognostic factor in human renal cell carcinoma: Clinicopathologic demonstration of antitumor immunity. *Cancer Res.* 2001; 61:5132–5136 PMID: 11431351
23. Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJ, Robert L, et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature.* 2014 Nov 27 2014; 515(7528):568–71 <https://doi.org/10.1038/nature13954> PMID: 25428505
24. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med.* 2015 Jun 25; 372(26):2509–20 <https://doi.org/10.1056/NEJMoa1500596> PMID: 26028255
25. Al-Shibli K I, Donnem T, Al-Saad S, Persson M, Bremnes RM, Busund LT. Prognostic effect of epithelial and stromal lymphocyte infiltration in non-small cell lung cancer. *Clin. Cancer Res.* 2008; 14:5220–5227 <https://doi.org/10.1158/1078-0432.CCR-08-0133> PMID: 18698040
26. Ito N, Suzuki Y, Taniguchi Y, Ishiguro K, Nakamura H, Ohgi S. Prognostic significance of T helper 1 and 2 and T cytotoxic 1 and 2 cells in patients with non-small cell lung cancer. *Anticancer Res.* 2005; 25:2027–2031 PMID: 16158940
27. Kawai O, Ishii G, Kubota K, Murata Y, Naito Y, Mizuno T, et al. Predominant infiltration of macrophages and CD8(+) T cells in cancer nests is a significant predictor of survival in stage IV non-small cell lung cancer. *Cancer.* 2008; 113:1387–1395 <https://doi.org/10.1002/cncr.23712> PMID: 18671239

28. Wakabayashi O, Yamazaki K, Oizumi S, Hommura F, Kinoshita I, Ogura S, et al. CD4+ T cells in cancer stroma, not CD8+ T cells in cancer cell nests, are associated with favorable prognosis in human non-small cell lung cancers. *Cancer Sci*. 2003 Nov; 94(11):1003–9 PMID: [14611679](#)
29. Hiraoka K, Miyamoto M, Cho Y, Suzuoki M, Oshikiri T, Nakakubo Y, et al. Concurrent infiltration by CD8+ T cells and CD4+ T cells is a favourable prognostic factor in non-small-cell lung carcinoma. *Br J Cancer*. 2006 Jan 30; 94(2):275–80 <https://doi.org/10.1038/sj.bjc.6602934> PMID: [16421594](#)
30. Bromwich EJ, McArdle PA, Canna K, McMillan DC, McNichol AM, Brown M, et al. The relationship between T-lymphocyte infiltration, stage, tumour grade and survival in patients undergoing curative surgery for renal cell cancer. *Br J Cancer*. 2003; 89:1906–1908 <https://doi.org/10.1038/sj.bjc.6601400> PMID: [14612901](#)
31. Lamb GW, McArdle PA, Ramsey S, McNichol AM, Edwards J, Aitchison M, et al. The relationship between the local and systemic inflammatory responses and survival in patients undergoing resection for localized renal cancer. *BJU Int*. 2008; 102:756–761 <https://doi.org/10.1111/j.1464-410X.2008.07666.x> PMID: [18384626](#)
32. Rehman JA, Han G, Carvajal-Haudorf DE, Wasserman BE, Pelekanou V, Mani NL, et al. Quantitative and pathologist-read comparison of the heterogeneity of programmed death-ligand 1 (PD-L1) expression in non-small cell lung cancer. *Mod Pathol*. 2017; 30:340–349 <https://doi.org/10.1038/modpathol.2016.186> PMID: [27834350](#)
33. Carroll KJ. Biomarkers in drug development: friend or foe? A personal reflection gained working within oncology. *Pharmaceutical Statistics* 2007; 6(4):253–260 <https://doi.org/10.1002/pst.269> PMID: [17351904](#)
34. Pennello GA. Analytical and clinical evaluation of biomarkers assays: When are biomarkers ready for prime time? *Clinical Trials* 2013; 10:666–676 <https://doi.org/10.1177/1740774513497541> PMID: [23983159](#)
35. Jenkins M, Flynn A, Smart T, Harbron C, Sabin T, Ratnayake J, et al. on behalf of the PSI Biomarker Special Interest Group. *Pharmaceutical Statistics* 2011; 10:494–507 <https://doi.org/10.1002/pst.532> PMID: [22162336](#)
36. Fridlyand J, Simon R, Walrath JC, Roach N, Buller R, Schenkein DP, et al. Considerations for the successful co-development of targeted cancer therapies and companion diagnostics. *Nat Rev Drug Discov* 2013; 12:743–755 <https://doi.org/10.1038/nrd4101> PMID: [24008432](#)
37. Fridlyand J, Yeh RF, Mackey H, Bengtsson T, Delmar P, Spaniolo G, et al. An industry statistician's perspective on PHC drug development. *Contemporary Clinical Trials* 2013; 36:624–635 <https://doi.org/10.1016/j.cct.2013.04.006> PMID: [23648396](#)
38. McShane LM, Polley MC. Development of omics-based clinical tests for prognosis and therapy selection: The challenge of achieving statistical robustness and clinical utility. *Clinical Trials* 2013; 10:653–665 <https://doi.org/10.1177/1740774513499458> PMID: [24000377](#)
39. Dibben SM, Holt RJ, Davison TS, Wilson CL, Taylor J, Paul I, et al. Implications for powering biomarker discovery studies. *J Mol Diagn* 2012; 14(2):130–139 <https://doi.org/10.1016/j.jmoldx.2011.10.002> PMID: [22251612](#)
40. Mackey HM, Bengtsson T. Sample size and threshold estimation for clinical trials with predictive biomarkers. *Contemporary Clinical Trials* 2013; 36:664–672 <https://doi.org/10.1016/j.cct.2013.09.005> PMID: [24064355](#)
41. Veta M, van Diest PJ, Kornegoor R, Huisman A, Viergever MA, Pluim JPW. Automatic Nuclei Segmentation in H&E Stained Breast Cancer Histopathology Images. *PLoS ONE* 2013; 8(7): e70221 <https://doi.org/10.1371/journal.pone.0070221> PMID: [23922958](#)
42. Brey EM, Lalani Z, Johnston C, Wong M, McIntire LV, Duke PJ, P CW Jr. Automated Selection of DAB-labeled Tissue for Immunohistochemical Quantification. *J Histochem & Cytochem*. 2003; 51(5): 575–584
43. R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>
44. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc* 1979; 74:829–836
45. 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 Feb; 26(2):259–71 <https://doi.org/10.1093/annonc/mdl450> PMID: [25214542](#)
46. Olurinde MO, Shen CH, Drake A, Bai A, Chen J. Persistence of tumor-infiltrating CD8 cells is tumor-dependent but antigen-independent. *Cell Mol Immunol*. Sept 2011; 8(5): 415–423 <https://doi.org/10.1038/cmi.2011.18> PMID: [21666707](#)

47. Seo AN, Lee HJ, Kim EJ, Kim HJ, Jang MH, Lee HE et al. Tumour-infiltrating CD8+ lymphocytes as an independent predictive factor for pathological complete response to primary systemic therapy in breast cancer. *Br J Cancer*. 2013; 109:2705–13 <https://doi.org/10.1038/bjc.2013.634> PMID: 24129232
48. Lee HJ, Seo JY, Ahn JH, Ahn SH, Gong G. Tumor-associated lymphocytes predict response to neoadjuvant chemotherapy in breast cancer patients. *J Breast Cancer*. 2013; 16:32–9 <https://doi.org/10.4048/jbc.2013.16.1.32> PMID: 23593079
49. Mahmoud S, Paish E, Powe D, Macmillan R, Grainge M, Lee A et al. Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer. *J Clin Oncol*. 2011; 29:1949–55 <https://doi.org/10.1200/JCO.2010.30.5037> PMID: 21483002
50. Liu F, Li Y, Ren M, Zhang X, Guo X, Lang R et al. Peritumoral FOXP3(+) regulatory T cell is sensitive to chemotherapy while intratumoral FOXP3(+) regulatory T cell is prognostic predictor of breast cancer patients. *Breast Cancer Res Treat*. 2012; 135:459–67 <https://doi.org/10.1007/s10549-012-2132-3> PMID: 22842982
51. Galon J, Mlecnik B, Bindea G, Angell HK, Berger A, Lagorce C, et al. Towards the introduction of the 'Immunoscore' in the classification of malignant tumours. *J Pathol*. Jan 2014; 232(2): pp 199–209 <https://doi.org/10.1002/path.4287> PMID: 24122236
52. Petrau C, Clatot F, Cornic M, Berghian A, Veresezan L, Callonnet F, et al. Reliability of Prognostic and Predictive Factors Evaluated by Needle Core Biopsies of Large Breast Invasive Tumors. *Am J Clin Pathol*. 2015 Oct; 144(4):555–62 <https://doi.org/10.1309/AJCP9KFVM2GZMNDV> PMID: 26386076
53. Wu ML, Dry SM, Lassman CR. Deeper Examination of Negative Colorectal Biopsies. *Am J Clin Pathol*. 2002; 117:424–428 <https://doi.org/10.1309/VBYP-UT2E-V198-M1TQ> PMID: 11888081
54. Baldus SE, Schaefer KL, Engers R, Hartleb D, Stoecklein NH, Gabbert HE. Prevalence and Heterogeneity of KRAS, BRAF, and PIK3CA Mutations in Primary Colorectal Adenocarcinomas and Their Corresponding Metastases. *Clin Cancer Res*. February 1, 2010 16:790–799 <https://doi.org/10.1158/1078-0432.CCR-09-2446> PMID: 20103678
55. Asano Y, Kashiwagi S, Goto W, Kurata K, Noda S, Takashima T, et al. Tumour-infiltrating CD8 to FOXP3 lymphocyte ratio in predicting treatment responses to neoadjuvant chemotherapy of aggressive breast cancer. *Br J Surg*. 2016 Jun; 103(7):845–54 <https://doi.org/10.1002/bjs.10127> PMID: 26953091
56. Miyashita M, Sasano H, Tamaki K, Hirakawa H, Takahashi Y, Nakagawa S, et al. Prognostic significance of tumor-infiltrating CD8+ and FOXP3+ lymphocytes in residual tumors and alterations in these parameters after neoadjuvant chemotherapy in triple-negative breast cancer: a retrospective multicenter study. *Breast Cancer Res*. 2015 Sep 4; 17:124. <https://doi.org/10.1186/s13058-015-0632-x> PMID: 26341640
57. Suzuki H, Chikazawa N, Tasaka T, Wada J, Yamasaki A, Kitaura Y, et al. Intratumoral CD8(+) T/FOXP3 (+) cell ratio is a predictive marker for survival in patients with colorectal cancer. *Cancer Immunol Immunother*. 2010; 59(5):653–661 <https://doi.org/10.1007/s00262-009-0781-9> PMID: 19908042
58. Zeestraten EC, Van Hoesel AQ, Speetjens FM, Menon AG, Putter H, van de Velde CJ, et al. FoxP3- and CD8-positive Infiltrating Immune Cells Together Determine Clinical Outcome in Colorectal Cancer. *Cancer Microenviron*. 2013 Apr; 6(1):31–9 <https://doi.org/10.1007/s12307-011-0071-x> PMID: 21732187
59. Loddenkemper C, Schernus M, Noutsias M, Stein H, Thiel E, Nagorsen D. In situ analysis of FOXP3+ regulatory T cells in human colorectal cancer. *J Transl Med*. 2006; 4:52 <https://doi.org/10.1186/1479-5876-4-52> PMID: 17166272
60. Salama P, Phillips M, Grieu F, Morris M, Zeps N, Joseph D, et al. Tumor-Infiltrating FOXP3+ T regulatory cells show strong prognostic significance in colorectal cancer. *J Clin Oncol*. 2009 Jan 10; 27(2):186–92 <https://doi.org/10.1200/JCO.2008.18.7229> PMID: 19064967
61. Wang J, Ioan-Facsinay A, van der Voort EI, Huizinga TW, Toes RE. Transient expression of FOXP3 in human activated nonregulatory CD4+ T cells. *Eur J Immunol*. 2007 Jan; 37(1):129–38. <https://doi.org/10.1002/eji.200636435> PMID: 17154262