

Corresponding author mail id: sigrid.hatse@kuleuven.be

Running title: Scoring immune cell populations in breast tumors

Computerized scoring protocol for identification and quantification of different immune cell populations in breast tumor regions using QuPath software

AUTHORLIST

Lieze Berben¹, Hans Wildiers^{1,2}, Lukas Marcelis³, Asier Antoranz Martinez³, Francesca Bosisio^{3,4}, Sigrid Hatse^{1*}, Giuseppe Floris^{3,4*}

*Both authors contributed equally to this manuscript

¹Laboratory of Experimental Oncology, Department of Oncology, KU Leuven, Leuven, Belgium

²Department of General Medical Oncology, University Hospitals Leuven, Leuven, Belgium

³Department of Pathology, Laboratory of Translational Cell and Tissue Research, KU Leuven, Leuven, Belgium

⁴Department of Pathology, University Hospitals Leuven, KU Leuven, Leuven, Belgium

*These authors contributed equally to this work

The authors declare no potential conflicts of interest.

Word count: 2769

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/his.14108](https://doi.org/10.1111/his.14108)

This article is protected by copyright. All rights reserved

ABSTRACT

Background: Important prognostic and predictive information can be obtained from the composition, functionality and spatial arrangement of different immune cell subtypes in breast tumors. Patients and methods: Tumor infiltrating lymphocytes (TILs) in 62 patients with luminal B-like breast cancer were characterized by immunohistochemical staining with standard markers and were subsequently classified and quantified using the QuPath software. In different delineated tumor regions, proportion and density of CD3⁺, CD4⁺, CD5⁺, CD8⁺, CD20⁺ and FOXP3⁺ cells were assessed. Results of the software analysis were compared to manual counting for CD8 and CD20 stainings. Results: The QuPath scoring protocol slightly overestimated positive, negative, total lymphocyte counts and density while minimally underestimating the proportion of positively stained lymphocytes. However, for density and proportion no real differences were observed compared to manual counting. For all markers density of positively stained immune cells was higher in the invasive front than in the tumor center, pointing to an accumulation of immune cells near the tumor boundaries. When we looked at the proportion of IHC positive immune cells, we observed enrichment of CD5 ($p=0.025$) and CD20 ($p<0.001$) at the periphery and FOXP3 enrichment in the center ($p<0.001$), respectively. Conclusion: The QuPath scoring protocol can adequately identify positively stained immune cells in breast tumors and allows to evaluate differences in immune cell proportion and density within different tumor regions. The entire tumor section can be quantitatively assessed quite rapidly, which is a major advantage over manual counting.

KEYWORDS

Tumor immune infiltration, QuPath, scoring protocol, breast cancer

INTRODUCTION

Breast cancer (BC) has long been considered as non-immunogenic, but accumulating evidence suggests that the tumor immune infiltrate present in the surrounding stroma, is of much higher importance than previously thought [1]. For the evaluation of tumor infiltrating lymphocytes (TILs) in BC, the first recommendations were published in 2014 by the international TILs working group [2]. An update of these recommendations was

published in 2017 [3] and a web-based training tool is now available [4]. According to these guidelines, a high-quality hematoxylin and eosin (H&E) staining of the tumor tissue is sufficient to assess the mononuclear immune infiltrate in the tumor invasive front's stroma. This score provides the percentage of stromal TILs (i.e. area of stroma occupied by TILs) and thus it is a surrogate of the extent of tumor immune infiltration at only one specific tumor region. However, it does not provide any information about immune cell distribution within different tumor regions nor about TILs composition. The prognostic Immunoscore was validated in colorectal cancer but not in BC [5-7]. Although many different methods have been described in literature to evaluate the tumor immune infiltrate composition, no standardized scoring methods are available in BC [8]. Moreover, most of the reported methods are semi-quantitative or depend on time consuming manual counting. Here we have evaluated a semi-automated quantitative scoring protocol using QuPath (further referred to as the QuPath protocol), an open source software for digital pathology and whole slide image analysis [9]. This scoring protocol was applied after immunohistochemical labeling of different immune cell markers on whole breast tumor sections.

METHODS

Patients and tumor tissue specimens

This study was performed on tumor tissue specimens from the 65 patients included in the IMAGE (Immunity and aging) study conducted at our institution (University Hospitals Leuven), ClinicalTrials.gov identifier: NCT02327572, approved by the Ethical committee of the University Hospitals Leuven on the 14 of March 2014: B322201420510/S56278). Eligible patients, were newly diagnosed with early BC: grade II/III invasive carcinoma on core needle biopsy, estrogen receptor (ER)-positive, human epidermal receptor 2 (HER2)-negative and estimated tumor size 1.5 cm or larger. They were all treated at the University Hospitals Leuven and scheduled for primary surgery. A written informed consent was obtained from all patients. The excised tumor tissue of the patients was formalin-fixed and paraffin-embedded (FFPE). **Next, a representative tumor section of the resection specimen was selected to perform immunostainings on whole slide sections.**

Immunohistochemistry

Tumor immune infiltrate characterization was performed by evaluating six immune cell markers (CD3, CD4, CD5, CD8, CD20, and FOXP3) *via* immunohistochemistry on whole tumor sections. The stainings were performed following the manufacturer's instructions, more details can be found in Supplementary 1.

QuPath scoring method

After immunohistochemical staining, the slides were scanned using the Philips Ultra Fast Scanner version 1.6, digitalized slides were converted in BigTIFF format and were then imported into QuPath for further image analysis.

A flowchart of the QuPath protocol can be found in Figure 1. First, the *Simple tissue detection* tool was used to create annotation of the tissue region to be analyzed and to subordinate and link subsequent passage for further analysis of different regions of interest. In a second step, the tumor border of each sample was manually outlined using the *Polygon tool*. The tumor border was determined as the boundary between tumor cells and normal tissue. Thirdly, the tumor border was selected and a software script (developed by the QuPath developer [10]) was run to create additional boundaries, i.e. 500 μm outwards and 500 μm inwards of the tumor border (Fig. 1 in green and blue). Afterwards, these were used to divide the tumor into three different regions: outer margin, inner margin and tumor center (Fig. 1 in black). The combination of outer and inner margin is further referred to as the invasive front (Fig. 1 in yellow). The tumor center starts at the internal border created by the software script and comprises all the tissue enclosed inside the invasive front. The whole tumor region was defined as the combination of the tumor center with invasive front. When all tumor regions were defined, the annotation created by *Simple tissue detection* was removed. The *Cell detection* tool was used in the fourth step. During this process, QuPath detects every cell in the tumor *via* a built-in cell segmentation algorithm. Depending on tissue type and specific staining applied, other settings are needed for proper cell detection. Table 1 shows an overview of the applied cell detection settings. In the fifth step, the *Add smoothed features (25 μm)* tool calculates a new measurement by taking a weighted average of cell measurements within the 25 μm range whereby the image is segmented homogeneously. Hereby the classification of groups of similar cells facilitated. Once all cells in the tumor regions were

detected and smoothed features were added, several groups of cells of the same cell type were manually outlined with the *Polygon tool* thus generating different cell classes that were annotated: “tumor cell” (Fig. 1 in light green), “immune cell” (Fig.1 in purple), “necrosis” and “other” (step 6, not shown in Fig 1). Thereafter, in the seventh and eighth step, a *Detection classifier* was created via the *Create detection classifier* function. To make this *detection classifier* function operational, the user needs to train the software by annotating a sufficiently high number of cells based on previously assigned parameters. Subsequently, after a phase of trial and error, the user must check the software’s ability to correctly recognize and assign diverse cells. At the end of the training and trial and error phase, the *Detection classifier* is ready to be used. Therefore, although the *Detection classifier* can be saved and potentially used for all samples, we preferred to create a new *Detection classifier* for each single sample and staining. This was necessary to compensate small technical artefacts coming from the manual staining procedure. To be able to distinguish between positively and negatively stained cells, the *Intensity feature* in the *Detection classifier* was used. A representative example of a non-processed and processed CD3-stained tumor section can be found in Figure 1A/B/C.

Lastly, each tumor region’s area was computed and numbers of tumor cells, immune cells (both positively and negatively stained), necrotic and other cells were automatically counted within each region. All data were extracted from QuPath and further calculations were performed in MS Excel. The *proportion* of positive lymphocytes was defined as the ratio of positively stained lymphocytes *versus* the total number of infiltrating lymphocytes counted in that region after digital segmentation. The *density* of positive lymphocytes was defined as the number of positively stained lymphocytes per mm².

Statistics

All statistical analyses were performed using GraphPad Prism 8 software. The non-parametric Mann-Whitney U test was used to evaluate differences in proportion and density of positively stained cells between the tumor center and invasive front. All reported p-values are two-sided, significance threshold was set at 5%. To compare the QuPath protocol and manual counting, the relative error (RE) of the QuPath protocol was calculated: (QuPath-Manual)/Manual. The threshold for a ‘real difference’ between the two methods was set above 0.200 (20,0%). The threshold of 20% was chosen to spot

outliers, and was based on functional sensitivity [11] described by Armbruster and Pry. A threshold of 20% is commonly used to evaluate interobserver variability. Additionally, a Pearson correlation between the QuPath protocol and manual counting was performed as well. All reported p-values are two-sided, significance threshold was set below 5% for all tests.

RESULTS

Comparison between automated QuPath scoring and manual counting

To confirm QuPath's accuracy, a comparison with manual counting was performed. The manual counting is considered as the gold standard for the evaluation of lymphocyte infiltration. In total, 5 randomly selected CD8 and CD20 stained tumor sections were evaluated. For each sample, 5 regions of interest (ROI) of 0.25 mm² were scored using both the QuPath protocol and manual counting. Positively stained immune cells as well as negatively stained immune cells were quantified from which the proportion and density of positively stained cells were calculated. Results of both methods are summarized in Table 2; some real differences were found between both methods, however the QuPath protocol proved to be quite accurate. For both CD8 and CD20 the same trend was observed: positive, negative, total lymphocyte counts and density were slightly overestimated while the proportion of positively stained cells was marginally underestimated. When analyzing the data in more detail, it was noticed that especially the negative lymphocyte counts were overestimated and this was a real difference (RE>20%, as defined in the methods section) for CD8 but not for CD20. Consequently, the total lymphocyte counts were overestimated as well, again more pronounced for CD8. Nevertheless, when looking at the most important output parameters: density and proportion, no real differences were seen which is reflected by the small RE's. Hence, the impact of the minor overestimation of the positive and negative lymphocytes counts have a relatively low impact on the density and proportion measured by QuPath. Next, the correlation analysis between the QuPath Protocol and manual counting showed very strong and highly significant associations between both methods for all evaluated parameters as shown in Table 3. Additionally, we compared the manual proportions within the selected regions of 0.25 mm² against the QuPath proportions in the same ROI and the QuPath proportions in the whole tumor section. Here, we also noticed no real

difference between the different methods and regions (Table 4), hence the proportion does not change between the methods and not in the total tumor area.

Using the QuPath protocol for immune infiltrate characterization in breast tumors

Tumor tissue was available for 62 patients of the 65 patients included in the IMAGE study. The full IHC panel was performed in all tumor samples but one, for which CD3, CD4 and FOXP3 lacked because of insufficient tumor tissue on the paraffin block. With the QuPath protocol, proportion and density of positively stained immune cells could be evaluated for each marker in different tumor regions. An overview of the obtained results can be found in Figure 2. Regardless of the region considered, about 50% of lymphocytes in the tumor were CD3⁺ T-cells. Approximately 30% and 25% of TILs stained positive for CD4 and CD8, respectively. CD20⁺ cells represented a smaller fraction of TILs (13-20%) and the regulatory T-cell marker FOXP3 was expressed by ≤10% of TILs. Comparison of the infiltration in tumor center and invasive front revealed some interesting spatial differences. All evaluated immune cell markers showed a significantly higher density in the invasive front than in the tumor center (Table 5, Figure 2). Moreover, compared to TILs in the tumor center, higher proportions of TILs in the invasive front stained positive for the T-cell marker/B-cell marker CD5 ($p=0.025$) and B-cell marker CD20 ($p<0.001$), whereas the tumor center contained a higher proportion of FOXP3⁺ lymphocytes ($p<0.001$). Proportions of CD3⁺, CD4⁺ and CD8⁺ TILs did not significantly differ between both regions.

DISCUSSION

Composition, location and functionality of different immune cell subtypes within the tumor microenvironment may have a large impact on prognosis as well as on response to cancer therapy. Therefore, immune infiltrate characterization in breast tumors has gained enormous interest over the past years [3, 12]. This may be of particular importance for immunotherapy, as this treatment approach often relies on immune cells already present in the tumor [13]. However, to date no validated standardized scoring methods are available to characterize and quantify different immune cell populations in the breast tumor microenvironment.

Here, we have evaluated a computerized scoring QuPath protocol, for its ability to identify and enumerate positively stained immune cells in different regions of breast tumors. Because no standardized scoring method is available, manual cell counting was performed to validate the QuPath protocol's utility and accuracy. The QuPath protocol was able to correctly identify positively stained immune cells for the assessed markers (i.e. CD8 and CD20), while some overestimation of negatively stained immune cells was observed (more pronounced for CD8, RE>20%). No real difference between the manual counting and QuPath protocol were noticed for the most important output parameters: density and proportion. Moreover, very strong correlations were found between the QuPath protocol and manual counting for all parameters assessed, confirming the utility and accuracy of the QuPath protocol. Additionally, the proportion did not change between the methods and not in the total tumor area. Time necessary to evaluate a tumor section with the QuPath protocol depends on sample size and on the manual annotation session (step 6 and 7 of Fig. 1). With this protocol, even the largest whole tumor section (390 mm²) in our cohort could be analyzed within 2 hours. In contrast, absolute quantification of total TILs within a whole tumor section of only 100 mm² by manual counting, depending on the amount of infiltration, could require up to 150-fold more time (i.e. about 300 hours per staining). Thus, by using the QuPath protocol the time to evaluate the immune infiltrate in the whole tumor section can be dramatically reduced, rendering whole slide quantifications much more feasible. Hence, potential selection bias, which results from restricted counting of preselected tumor areas, which might influence results interpretation, could be reduced. Furthermore, after training of the classifier all measurements are done automatically, making this method less labor intensive than manual counting, even though training had to be repeated for each sample and staining. Notably, using tissue micro array slides could reduce the analysis time even more. As they are stained simultaneously (same staining intensity in all samples,...), the training step is potentially required only once per staining. However, like manual counting, performance of the QuPath protocol is also user dependent [14]. The amount of manual annotations, made by the user in order to teach the software how to identify different cell types (tumor cells, immune cells, other), as well as the accuracy by which this is done, has a large impact on cell classification by QuPath.

Recently, the immune landscape has gained interest in BC, however, mainly in the triple negative and HER2 positive breast cancers. **Nevertheless, the luminal B-like BC subtype is a frequent BC subtype and has a clearly inferior prognosis compared to luminal A-like BC [15].** Thus, there is a clinical need for improved understanding of this highly under explored BC subtype. For this reason, the computerized QuPath protocol was used to characterize the immune infiltrate in patients carrying tumors of the luminal B-like BC. For several classical immune subset markers (i.e. CD3, CD4, CD5, CD8, CD20 and FOXP3), proportion as well as density of positively stained cells were assessed in different tumor regions (i.e. tumor center, invasive front and whole tumor region) of a whole tumor section. Hereby insights were gained not only in abundance and composition of the immune infiltrate but also in spatial distribution of different lymphocyte subtypes. However, in this study, we only performed single stainings of each marker separately and on different tissue sections, which could explain discrepancies between the percentages. On average, approximately 40-50% of TILs in the whole tumor region consisted of T-cells, as confirmed by CD3 and CD5 staining. Nevertheless, CD5 is also expressed by a subset of B-cells. Furthermore, CD4 and CD8 stainings showed that 32% and 26% of total TILs could be classified as CD4⁺ helper T-cells and CD8⁺ cytotoxic T-cells, respectively. However, it should be noted that macrophages also express CD4 and that co-expression of CD4 and CD8 can occur. Only a small fraction (7.5%) of TILs could be identified as regulatory T-cells (FOXP3⁺). Compared to the predominant T-cell compartment, B-cells (CD20⁺) were less abundant in the immune infiltrate, with on average only 18% of TILs in the whole tumor region. These findings are in agreement with earlier publications reporting on the composition of the immune infiltrate in BC [16-18]. When comparing different tumor regions, we observed some striking spatial differences. For all markers assessed, the density was higher in the invasive front than in tumor center, indicating that TILs are accumulating near the tumor edge and that the local immune response mainly takes place at the boundary, rather than in the tumor core. This spatial difference was most pronounced for CD20⁺ B-cells, which not only showed a higher density in the invasive front, but also constituted a higher percentage of TILs in this region (21% in the invasive front *versus* 13% in tumor center). Conversely, the infiltrate in the tumor center contained proportionally more FOXP3⁺ cells. Apparently, either regulatory T-cells tend to penetrate deeper into the tumor than other lymphocyte

subtypes or lymphocytes are converted into regulatory T-cells *in situ* by the tumor itself. This may be of importance with regard to immunotherapy, as these cells are supposed to have an immunosuppressive function.

CONCLUSION

The semi-automatic QuPath protocol is a much faster and equally reliable method compared to manual counting. It allows a more extensive and more detailed local tumor immune response analysis in BC, using whole tissue sections stained with diverse markers. Implementation and validation of this QuPath protocol in other tumor types warrants further investigation.

ACKNOWLEDGEMENTS

We would like to show our gratitude to the developer of the QuPath software, Peter Bankhead, PhD, University of Edinburgh for developing the script that enabled us to create the different tumor regions.

TABLES

Table 1: Overview of the applied cell detection settings

	Manual staining	Automated staining
Set up parameters		
Detection image	Hematoxylin OD	Hematoxylin OD
Requested pixel size	0.5 μm	0.5 μm
Nucleus parameters		
Background radius	8 μm	8 μm
Median filter radius	0 μm	0 μm
Sigma	1.2	1.2
Minimum area	3 μm^2	3 μm^2
Maximum area	300 μm^2	300 μm^2
Intensity parameters		
Threshold	0.1	0.05
Max background intensity	2	2
Split by shape	V	V
Exclude DAB (membrane staining)	/	/
Cell parameters		
Cell expansion	1	1

Include cell nucleus	V	V
General parameters		
Smooth boundaries	V	V
Make measurements	V	V

Table 2: QuPath and manual counting of the CD8 and CD20 staining with positively stained (+), negatively stained (-) cells, total number of lymphocytes (Lymphocytes), the proportion of the positively stained cells (%) and the density of the positively stained cells (positively stained cells/mm²) in 5 different tumor samples and in 5 different tumor regions of 0.25 mm². The mean, standard deviation (SD) and relative error of the difference between the methods are reported. The relative error was calculated using the following formula: (QuPath-Manual)/Manual, the threshold of a real difference between the two methods was set above 0.200 (20%) (marked in dark grey).

Sample	+			-			Lymphocytes			Density			Proportion		
	Manual	QuPath	RE	Manual	QuPath	RE	Manual	QuPath	RE	Manual	QuPath	RE	Manual	QuPath	RE
CD8 1.1	347	381	0.098	439	532	0.212	786	913	0.162	694	762	0.098	44.1	41.7	-0.055
CD8 1.2	297	325	0.094	233	332	0.425	530	657	0.240	594	650	0.094	56.0	49.5	-0.117
CD8 1.3	209	231	0.105	243	315	0.296	452	546	0.208	418	462	0.105	46.2	42.3	-0.085
CD8 1.4	138	163	0.181	247	331	0.340	385	494	0.283	276	326	0.181	35.8	33.0	-0.079
CD8 1.5	35	36	0.029	47	43	-0.085	82	79	-0.037	70	72	0.029	42.7	45.6	0.068
CD8 2.1	405	409	0.010	777	991	0.275	1182	1400	0.184	810	818	0.010	34.3	29.2	-0.147
CD8 2.2	260	273	0.050	688	899	0.307	948	1172	0.236	520	546	0.050	27.4	23.3	-0.151
CD8 2.3	337	341	0.012	1238	1500	0.212	1575	1841	0.169	674	682	0.012	21.4	18.5	-0.134
CD8 2.4	207	233	0.126	269	380	0.413	476	613	0.288	414	466	0.126	43.5	38.0	-0.126
CD8 2.5	319	336	0.053	510	610	0.196	829	946	0.141	638	672	0.053	38.5	35.5	-0.077
CD8 3.1	132	128	-0.030	374	468	0.251	506	596	0.178	264	256	-0.030	26.1	21.5	-0.177
CD8 3.2	195	218	0.118	442	627	0.419	637	845	0.327	390	436	0.118	30.6	25.8	-0.157
CD8 3.3	97	94	-0.031	357	449	0.258	454	543	0.196	194	188	-0.031	21.4	17.3	-0.190
CD8 3.4	105	103	-0.019	451	575	0.275	556	678	0.219	210	206	-0.019	18.9	15.2	-0.196
CD8 3.5	81	69	-0.148	383	471	0.230	464	540	0.164	162	138	-0.148	17.5	12.8	-0.268
CD8 4.1	294	301	0.024	396	563	0.422	690	864	0.252	588	602	0.024	42.6	34.8	-0.182
CD8 4.2	218	235	0.078	362	551	0.522	580	786	0.355	436	470	0.078	37.6	29.9	-0.205
CD8 4.3	122	146	0.197	342	407	0.190	464	553	0.192	244	292	0.197	26.3	26.4	0.004
CD8 4.4	65	73	0.123	142	179	0.261	207	252	0.217	130	146	0.123	31.4	29.0	-0.077

CD8 4.5	447	574	0.284	674	908	0.347	1121	1482	0.322	894	1148	0.284	39.9	38.7	-0.029
CD8 5.1	82	88	0.073	227	301	0.326	309	389	0.259	164	176	0.073	26.5	22.6	-0.148
CD8 5.2	155	114	-0.265	558	755	0.353	713	869	0.219	310	228	-0.265	21.7	13.1	-0.397
CD8 5.3	237	284	0.198	968	1135	0.173	1205	1419	0.178	474	568	0.198	19.7	20.0	0.018
CD8 5.4	222	264	0.189	563	635	0.128	785	899	0.145	444	528	0.189	28.3	29.4	0.038
CD8 5.5	234	257	0.098	855	959	0.122	1089	1216	0.117	468	514	0.098	21.5	21.1	-0.016
CD8 total mean	209.6	227.0	0.083	471.4	596.6	0.266	681.0	823.7	0.210	419.2	454.1	0.083	32.0	28.6	-0.107
CD8 total SD	110.0	128.4		273.6	325.4		349.5	413.2		220.0	256.8		10.3	10.3	
CD20 1.1	36	23	-0.361	383	398	0.039	419	421	0.005	72	46	-0.361	8.6	5.5	-0.364
CD20 1.2	261	278	0.065	467	632	0.353	728	910	0.250	522	556	0.065	35.9	30.5	-0.148
CD20 1.3	11	9	-0.182	263	259	-0.015	274	268	-0.022	22	18	-0.182	4.0	3.4	-0.164
CD20 1.4	47	45	-0.043	362	416	0.149	409	461	0.127	94	90	-0.043	11.5	9.8	-0.151
CD20 1.5	21	16	-0.238	260	294	0.131	281	310	0.103	42	32	-0.238	7.5	5.2	-0.309
CD20 2.1	123	144	0.171	252	341	0.353	375	485	0.293	246	288	0.171	32.8	29.7	-0.095
CD20 2.2	410	415	0.012	495	658	0.329	905	1073	0.186	820	830	0.012	45.3	38.7	-0.146
CD20 2.3	1108	1255	0.133	588	834	0.418	1696	2089	0.232	2216	2510	0.133	65.3	60.1	-0.080
CD20 2.4	335	379	0.131	619	781	0.262	954	1160	0.216	670	758	0.131	35.1	32.7	-0.070
CD20 2.5	237	280	0.181	439	538	0.226	676	818	0.210	474	560	0.181	35.1	34.2	-0.024
CD20 3.1	32	32	0.000	341	434	0.273	373	466	0.249	64	64	0.000	8.6	6.9	-0.200
CD20 3.2	123	125	0.016	370	494	0.335	493	619	0.256	246	250	0.016	24.9	20.2	-0.191
CD20 3.3	28	29	0.036	332	414	0.247	360	443	0.231	56	58	0.036	7.8	6.5	-0.158
CD20 3.4	14	10	-0.286	140	113	-0.193	154	123	-0.201	28	20	-0.286	9.1	8.1	-0.106
CD20 3.5	123	127	0.033	450	565	0.256	573	692	0.208	246	254	0.033	21.5	18.4	-0.145
CD20 4.1	4	5	0.250	225	226	0.004	229	231	0.009	8	10	0.250	1.7	2.2	0.239
CD20 4.2	6	3	-0.500	223	213	-0.045	229	216	-0.057	12	6	-0.500	2.6	1.4	-0.470
CD20 4.3	181	186	0.028	641	682	0.064	822	868	0.056	362	372	0.028	22.0	22.0	0.000

CD20 4.4	13	14	0.077	176	170	-0.034	189	184	-0.026	26	28	0.077	6.9	6.9	0.000
CD20 4.5	89	72	-0.191	1038	1085	0.045	1127	1157	0.027	178	144	-0.191	7.9	7.9	0.000
CD20 5.1	134	147	0.097	217	263	0.212	351	410	0.168	268	294	0.097	38.2	35.9	-0.061
CD20 5.2	70	73	0.043	223	236	0.058	293	309	0.055	140	146	0.043	23.9	23.6	-0.011
CD20 5.3	91	93	0.022	207	245	0.184	298	338	0.134	182	186	0.022	30.5	27.5	-0.099
CD20 5.4	225	249	0.107	371	455	0.226	596	704	0.181	450	498	0.107	37.8	35.4	-0.063
CD20 5.5	306	324	0.059	505	622	0.232	811	946	0.166	612	648	0.059	37.7	34.2	-0.092
CD20 total mean	161.1	173.3	0.076	383.5	454.7	0.186	544.6	628.0	0.153	322.2	346.6	0.076	22.5	20.3	-0.098
CD20 total SD	228.3	257.8		196.2	236.4		358.0	436.6		456.6	515.7		16.5	15.2	

Table 3: Pearson correlation analysis between the QuPath protocol and manual counting. Results are shown for the CD8 and CD20 staining with positively stained (+), negatively stained (-) cells, total number of lymphocytes (Lymphocytes), the proportion of the positively stained cells (%) and the density of the positively stained cells (positively stained cells/mm²). The Pearson correlation coefficient (ρ) and the p-value are shown and significant results are marked in dark grey.

	Rho	p-value
CD8		
+	0.982	<0.0001
-	0.990	<0.0001
Lymphocytes	0.993	<0.0001
Density	0.982	<0.0001
Proportion	0.963	<0.0001
CD20		
+	0.999	<0.0001
-	0.969	<0.0001
Lymphocytes	0.992	<0.0001
Density	0.999	<0.0001
Proportion	0.996	<0.0001

Table 4: QuPath and manual proportions of the CD8 and CD20 positively stained cells were compared in different regions i.e. the selected regions of interest (ROI) of 0.25 mm² and the whole tumor section (WS). The mean, standard deviation (SD) and relative error of the difference between the methods and regions are reported. The relative error was calculated using the following formula: (QuPath-Manual)/Manual, the threshold of a real difference between the two methods was set above 0.200 (20%) (marked in dark grey).

Sample	Proportion (Manual_ROI vs QuPath_ROI)			Proportion (Manual ROI vs QuPath_WS)			Proportion (Manual ROI vs QuPath_WS)		
	Manual_ROI	QuPath_ROI	Relative error	Manual_ROI	QuPath_WS	Relative error	QuPath_ROI	QuPath_WS	Relative error
CD8 1.1	44.1	41.7	-0.055	44.1			41.7		
CD8 1.2	56.0	49.5	-0.117	56.0			49.5		
CD8 1.3	46.2	42.3	-0.085	46.2			42.3		
CD8 1.4	35.8	33.0	-0.079	35.8			33.0		
CD8 1.5	42.7	45.6	0.068	42.7			45.6		
Mean	45.0	42.2	-0.061	45.0	44.6	-0.009	42.2	44.6	0.056
CD8 2.1	34.3	29.2	-0.147	34.3			29.2		
CD8 2.2	27.4	23.3	-0.151	27.4			23.3		
CD8 2.3	21.4	18.5	-0.134	21.4			18.5		
CD8 2.4	43.5	38.0	-0.126	43.5			38.0		
CD8 2.5	38.5	35.5	-0.077	38.5			35.5		
Mean	33.0	28.9	-0.124	33.0	28.7	-0.130	28.9	28.7	-0.007
CD8 3.1	26.1	21.5	-0.177	26.1			21.5		
CD8 3.2	30.6	25.8	-0.157	30.6			25.8		
CD8 3.3	21.4	17.3	-0.190	21.4			17.3		
CD8 3.4	18.9	15.2	-0.196	18.9			15.2		
CD8 3.5	17.5	12.8	-0.268	17.5			12.8		
Mean	22.9	18.5	-0.192	22.9	17.5	-0.236	18.5	17.5	-0.054
CD8 4.1	42.6	34.8	-0.182	42.6			34.8		
CD8 4.2	37.6	29.9	-0.205	37.6			29.9		

CD8 4.3	26.3	26.4	0.004	26.3			26.4		
CD8 4.4	31.4	29.0	-0.077	31.4			29.0		
CD8 4.5	39.9	38.7	-0.029	39.9			38.7		
Mean	35.6	31.8	-0.107	35.6	29.4	-0.174	31.8	29.4	-0.075
CD8 5.1	26.5	22.6	-0.148	26.5			22.6		
CD8 5.2	21.7	13.1	-0.397	21.7			13.1		
CD8 5.3	19.7	20.0	0.018	19.7			20.0		
CD8 5.4	28.3	29.4	0.038	28.3			29.4		
CD8 5.5	21.5	21.1	-0.016	21.5			21.1		
Mean	23.5	21.2	-0.098	23.5	23.1	-0.017	21.2	23.1	0.090
CD8 total mean	32.0	28.6	-0.107	32.0	28.7	-0.104	28.6	28.7	0.003
CD20 1.1	8.6	5.5	-0.364	8.6			5.5		
CD20 1.2	35.9	30.5	-0.148	35.9			30.5		
CD20 1.3	4.0	3.4	-0.164	4.0			3.4		
CD20 1.4	11.5	9.8	-0.151	11.5			9.8		
CD20 1.5	7.5	5.2	-0.309	7.5			5.2		
Mean	13.5	10.9	-0.193	13.5	12.0	-0.111	10.9	12.0	0.101
CD20 2.1	32.8	29.7	-0.095	32.8			29.7		
CD20 2.2	45.3	38.7	-0.146	45.3			38.7		
CD20 2.3	65.3	60.1	-0.080	65.3			60.1		
CD20 2.4	35.1	32.7	-0.070	35.1			32.7		
CD20 2.5	35.1	34.2	-0.024	35.1			34.2		
Mean	42.7	39.1	-0.084	42.7	34.4	-0.194	39.1	34.4	-0.120
CD20 3.1	8.6	6.9	-0.200	8.6			6.9		
CD20 3.2	24.9	20.2	-0.191	24.9			20.2		
CD20 3.3	7.8	6.5	-0.158	7.8			6.5		

CD20 3.4	9.1	8.1	-0.106	9.1			8.1		
CD20 3.5	21.5	18.4	-0.145	21.5			18.4		
Mean	14.4	12.0	-0.167	14.4	9.7	-0.326	12.0	9.7	-0.192
CD20 4.1	1.7	2.2	0.239	1.7			2.2		
CD20 4.2	2.6	1.4	-0.470	2.6			1.4		
CD20 4.3	22.0	22.0	0.000	22.0			22.0		
CD20 4.4	6.9	6.9	0.000	6.9			6.9		
CD20 4.5	7.9	7.9	0.000	7.9			7.9		
Mean	8.2	8.1	-0.012	8.2	11.3	0.378	8.1	11.3	0.395
CD20 5.1	38.2	35.9	-0.061	38.2			35.9		
CD20 5.2	23.9	23.6	-0.011	23.9			23.6		
CD20 5.3	30.5	27.5	-0.099	30.5			27.5		
CD20 5.4	37.8	35.4	-0.063	37.8			35.4		
CD20 5.5	37.7	34.2	-0.092	37.7			34.2		
Mean	33.6	31.3	-0.068	33.6	34.2	0.018	31.3	34.2	0.093
CD20 total mean	22.5	20.3	-0.098	22.5	20.3	-0.097	20.3	20.3	0.001

Table 5: Results of the QuPath analysis with the mean proportion (%) (positivity among total lymphocytes) and mean density (number of positively stained cells/mm²) for each marker in the different regions. P-value of the difference between tumor center and invasive front are reported for proportion and density for each marker.

	Tumor center	Invasive front	P-value
CD3 mean proportion	46.7	48.4	0.595
CD4 mean proportion	29.5	32.9	0.071
CD5 mean proportion	36.8	41.2	0.025
CD8 mean proportion	26.7	25.4	0.483
CD20 mean proportion	12.7	20.9	<0.001
FOXP3 mean proportion	10.2	6.2	<0.001
CD3 mean density	287.7	445.8	<0.001
CD4 mean density	210.5	401.0	<0.001
CD5 mean density	222.7	391.2	<0.001
CD8 mean density	194.1	274.6	0.004
CD20 mean density	91.3	192.5	<0.001
FOXP3 mean density	38.5	54.7	0.046

FIGURE LEGENDS

Figure 1: Flowchart of the analysis pipeline in the QuPath software. 1. Simple tissue detection was performed to detect all tissue on the scan. 2. The tumor border (red) was manually outlined. 3. A script divided the tumor into different regions: outer margin 500 μm in width (green), inner margin 500 μm in width (blue) and the tumor center (black). C. The combination of the outer and inner margin was defined as the invasive front (yellow). 4. Segmentation analysis of QuPath software allows recognition of diverse types of cells. 5. Smoothed features were added to get a more homogenous segmentation. 6. Annotations of the different cell types were created by assignment of different color codes for each cell type. 7. A detection classifier was created. 8. Intensity features were chosen. Bottom images show an example of an unanalyzed (A) and QuPath analyzed (B+C) CD3 stained section of the tumor, the latter with coloring of the tumor cells (purple), positively stained immune cells (red) and negatively stained immune cells (blue).

Figure 2: Bar plots: tumor immune infiltration in the different tumor regions as analyzed by Qu-path. The different immune cell markers (CD3, CD4, CD5, CD8, CD20 and FOXP3) are represented on the x-axis. The mean density (number of positively stained lymphocytes/ mm^2) of the positively stained immune cells can be seen on barplot A. The mean proportion (%) can be seen on barplot B. The tumor center is represented in grey, invasive front in yellow and the whole tumor region in pink. Level of significance: *: $p < 0.05$; **: $p < 0.01$ and ***: $p < 0.001$.

REFERENCE LIST

1. Dushyanthen, S., et al., *Relevance of tumor-infiltrating lymphocytes in breast cancer*. BMC Med, 2015. **13**: p. 202.
2. Salgado, R., et al., *The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014*. Annals of Oncology, 2015. **26**(2): p. 259-271.
3. Hendry, S., et al., *Assessing Tumor-Infiltrating Lymphocytes in Solid Tumors: A Practical Review for Pathologists and Proposal for a Standardized Method from the International Immuno-Oncology Biomarkers Working Group: Part 2: TILs in Melanoma, Gastrointestinal Tract Carcinomas, Non-Small Cell Lung Carcinoma and Mesothelioma, Endometrial and Ovarian Carcinomas, Squamous Cell Carcinoma of the Head and Neck, Genitourinary Carcinomas, and Primary Brain Tumors*. Adv Anat Pathol, 2017. **24**(6): p. 311-335.
4. Cancer, I.I.-O.B.W.G.o.B. *Everything you need to know about TILs in cancer*. 2019; Available from: <https://www.tilsinbreastcancer.org/>.
5. Galon, J., et al., *Type, density, and location of immune cells within human colorectal tumors predict clinical outcome*. Science, 2006. **313**(5795): p. 1960-4.

- Accepted Article
6. Galon, J., et al., *Towards the introduction of the 'Immunoscore' in the classification of malignant tumours*. J Pathol, 2014. **232**(2): p. 199-209.
 7. Pagès, F., et al., *International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study*. The Lancet, 2018. **391**(10135): p. 2128-2139.
 8. Hendry, S., et al., *Assessing Tumor-infiltrating Lymphocytes in Solid Tumors: A Practical Review for Pathologists and Proposal for a Standardized Method From the International Immunooncology Biomarkers Working Group: Part 1: Assessing the Host Immune Response, TILs in Invasive Breast Carcinoma and Ductal Carcinoma In Situ, Metastatic Tumor Deposits and Areas for Further Research*. Adv Anat Pathol, 2017. **24**(5): p. 235-251.
 9. Bankhead, P., et al., *QuPath: Open source software for digital pathology image analysis*. Scientific Reports, 2017. **7**(1): p. 16878.
 10. Bankhead, P. *Creating annotations around the tumor*. 2018; Available from: <https://petebankhead.github.io/qupath/scripts/2018/08/08/three-regions.html>.
 11. Armbruster, D.A. and T. Pry, *Limit of blank, limit of detection and limit of quantitation*. Clin Biochem Rev, 2008. **29 Suppl 1**(Suppl 1): p. S49-52.

- Accepted Article
12. Galon, J., et al., *The immune score as a new possible approach for the classification of cancer*. Journal of translational medicine, 2012. **10**: p. 1-1.
 13. Bonaventura, P., et al., *Cold Tumors: A Therapeutic Challenge for Immunotherapy*. Frontiers in immunology, 2019. **10**: p. 168-168.
 14. Loughrey, M.B., et al., *Validation of the systematic scoring of immunohistochemically stained tumour tissue microarrays using QuPath digital image analysis*. Histopathology, 2018. **73**(2): p. 327-338.
 15. Voduc, K.D., et al., *Breast Cancer Subtypes and the Risk of Local and Regional Relapse*. Journal of Clinical Oncology, 2010. **28**(10): p. 1684-1691.
 16. Meng, S., et al., *Distribution and prognostic value of tumor-infiltrating T cells in breast cancer*. Mol Med Rep, 2018. **18**(5): p. 4247-4258.
 17. Wagner, J., et al., *A Single-Cell Atlas of the Tumor and Immune Ecosystem of Human Breast Cancer*. Cell, 2019. **177**(5): p. 1330-1345.e18.
 18. Konig, L., et al., *Dissimilar patterns of tumor-infiltrating immune cells at the invasive tumor front and tumor center are associated with response to neoadjuvant chemotherapy in primary breast cancer*. BMC Cancer, 2019. **19**(1): p. 120.

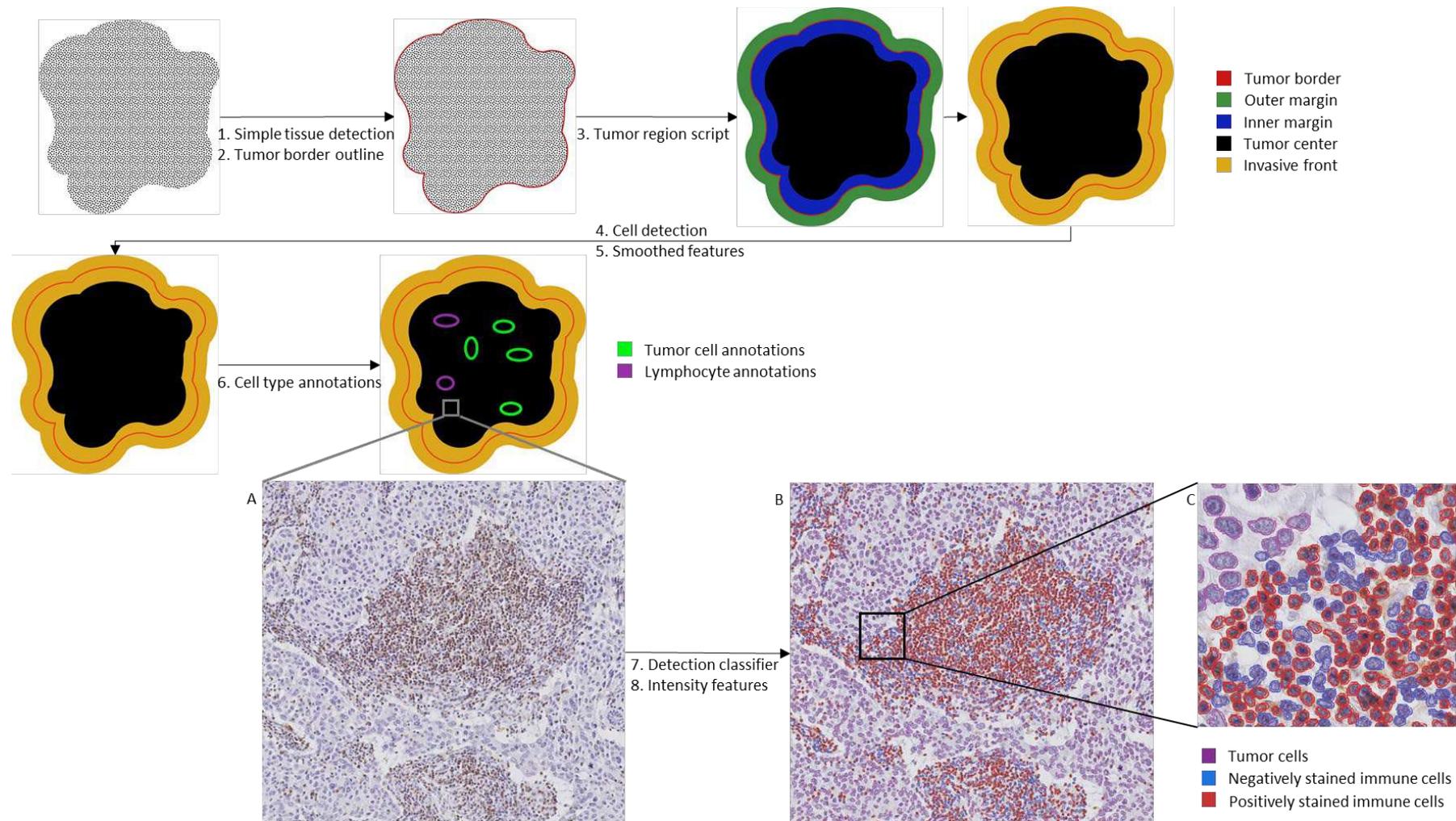


Figure 1: Flowchart of the analysis pipeline in the QuPath software. 1. Simple tissue detection was performed to detect all tissue on the scan. 2. The tumor border (red) was manually outlined. 3. A script divided the tumor into different regions: outer margin 500 μm in width (green), inner margin 500 μm in width (blue) and the tumor center (black). C. The combination of the outer and inner margin was defined as the invasive front (yellow). 4. Segmentation analysis of QuPath software allows recognition of diverse types of cells. 5. Smoothed features were added to get a more homogenous segmentation. 6. Annotations of the different cell types were created by assignment of different color codes for each cell type. 7. A detection classifier was created. 8. Intensity features were chosen. Bottom images show an example of an unanalyzed (A) and QuPath analyzed (B+C) CD3 stained section of the tumor, the latter with coloring of the tumor cells (purple), positively stained immune cells (red) and negatively stained immune cells (blue).

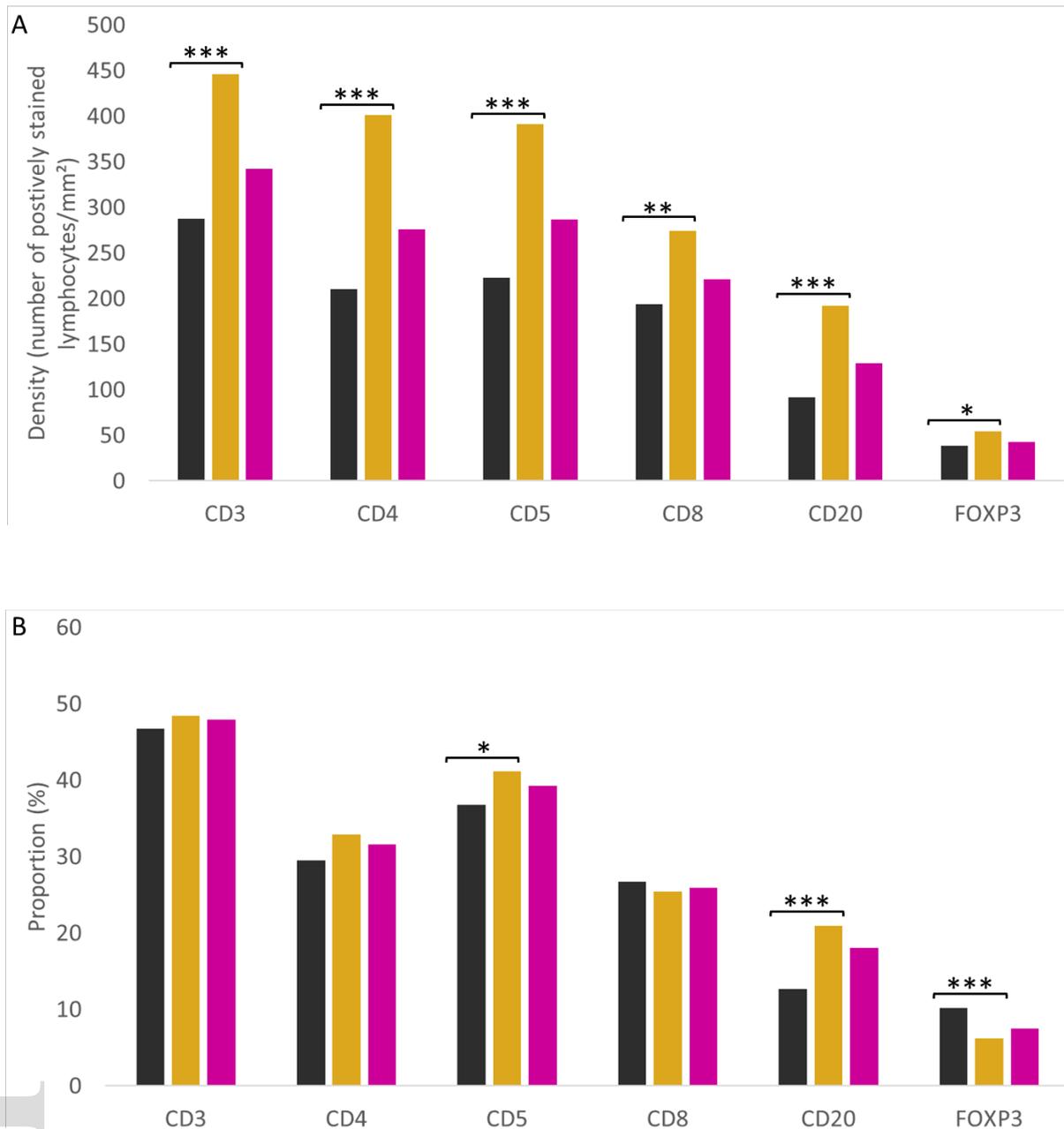


Figure 2: Bar plots: tumor immune infiltration in the different tumor regions as analyzed by Qu-path. The different immune cell markers (CD3, CD4, CD5, CD8, CD20 and FOXP3) are represented on the x-axis. The mean density (number of positively stained lymphocytes/mm²) of the positively stained immune cells can be seen on barplot A. The mean proportion (%) can be seen on barplot B. The tumor center is represented in grey, invasive front in yellow and the whole tumor region in pink. Level of significance: *: $p < 0.05$; **: $p < 0.01$ and ***: $p < 0.001$.