Quantification of Prognostic Immune Cell Markers in Colorectal Cancer Using Whole Slide Imaging Tumor Maps

Niels Halama, M.D., Inka Zoernig, Ph.D., Anna Spille, B.T.A., Sara Michel, M.Sc., Matthias Kloor, M.D., Silke Grauling-Halama, M.D., Kathi Westphal, Ph.D., Peter Schirmacher, M.D., Dirk Jäger, M.D., and Niels Grabe, Ph.D.

OBJECTIVE: To analyze intratumoral heterogeneity of immune cells and the resulting impact of heterogeneity on the level of individual patient prediction.

STUDY DESIGN: Using whole slide imaging by virtual microscopy, we present the first spatial quantitative study of immune cells in a set of colorectal cancer primary tumors. We generated “tumor maps” based on cell densities in fields of 1 mm², visualizing intratumoral heterogeneity. In this example, cutoffs of marker-based cell stains identified by tissue microarray (TMA) led to ambiguous decisions in 11 of the 20 patients studied. Classic TMA analysis can be used in large patient cohorts to generate clinically significant predictors. The transfer of these predictors from large-scale TMA to individualized predictions thus far has not been investigated. In colorectal cancer, TMA-based quantitative immune cell counts using immune cell surface molecules (CD3, CD8, Granzyme B, and CD45RO) have been shown to be potentially better predictors for patient survival than the classical TNM system.

RESULTS: Our results make clear that for individualized prognostic evaluations, whole slide imaging by virtual microscopy is irreplaceable during identification of prognostic markers as well as in their subsequent application.

CONCLUSION: In the future, spatial marker signatures could contribute to individual patient classifiers.

Keywords: colorectal cancer, immune cells, prognosis, virtual microscopy.

Histologic sampling allows the efficient study of target molecules in large numbers of patients, correlating molecular in situ findings with clinical pathologic information. Sampling histologic probes using tissue microarrays (TMAs) is especially problematic in the case of spatial heterogeneity of the target molecule as it is frequently observed in cancer tissue. Current TMAs typically use two randomly selected cores for analysis and analyze large cohorts of patients. Both measures reduce statistical variance and increase the reliability of the derived classification results under the assumption...
of large patient cohorts. However, taking multiple tissue cylinders from the same paraffin block is laborious and increases the risk for destroying the block. Also, although the TMA approach efficiently leads to statistically valid results, the main clinical intention behind histologic sampling is to determine cutoffs (thresholds) that enable a decision concerning the individual patient. Unfortunately, it is not at all clear in how far—or whether at all—the statistical aggregation of mass data from TMAs will lead to reliable cutoffs for diagnostic or prognostic decisions on the level of the individual patient.

Recently, a technologic breakthrough in histology occurred by the introduction of virtual microscopy (VM). It allows the automated high throughput microscopy of microscopic slides in the form of full tissue sections. Thus massive histologic data delivering morphologic and molecular tissue data with high spatial resolution is generated, especially with fluorescence. The ability to scan full tissue slides makes VM a crucial tool for determining the spatial heterogeneity of histologic parameters, and even more important, it may produce clinically relevant classifiers. Using VM we set out to determine spatial heterogeneity of immunohistologic parameters can influence diagnostic and prognostic decisions.

We performed an analysis of heterogeneity of different lymphocyte subsets in colorectal cancer (CRC) primary tumors by immunohistologic staining. CRC causes 655,000 deaths worldwide per year, including about 55,000 in the United States alone, where it is the second most common tumor entity to cause cancer death. Immune cells infiltrating the tumor were shown to have an impact on the prognosis of patients with CRC. Not only can the primary tumor exhibit structural heterogeneity but also the densities of infiltrating immune cells. Clinical outcome of CRC is influenced by immune responses at the primary tumor site. Using a recently published set of immune cell surface stainings, we studied the spatial heterogeneity of these parameters in CRC primary tumors. We quantitatively analyzed the surface molecules CD3, CD8, Granzyme B, and CD45RO, which can be found on specific subsets of tumor-infiltrating lymphocytes. This provides insight into the extent of primary tumor heterogeneity in CRC in general and finally shows the potential relevance of VM for cancer diagnosis and prognosis.

Materials and Methods

Patients

Archival specimens were obtained from 20 patients at the University Clinic Heidelberg who signed informed consent documents. Patients were diagnosed with stage UICC III or IV CRC without microsatellite instability and treated at our facilities. Mean age was 62 years, with 38% of the patients being women. The study was approved by the Medical Ethical Committee of the University of Heidelberg.

Immunohistochemical Staining

Sections were cut from paraffin-embedded blocks. Mouse monoclonal antibodies used were antihuman CD3 (DAKO, Carpinteria, California, U.S.A.), CD8, granzyme B (Novocastra, Newcastle, U.K.) and CD45RO (clone UCHL-1, Biologo, Kronshagen, Germany). Immunohistochemistry of paraffin sections was carried out using a two-step protocol (Novolink Polymer Detection System, Novocastra) according to the manufacturer's instructions and as described previously. Sections were developed in diaminobenzidine solution and counterstained with hematoxylin. Negative control slides with omitted primary antibodies were included in all assays.

Evaluation of Immunohistochemical Variables

Slide scanning was performed using a NDP Nanozoomer from Hamamatsu Photonics. In contrast to...
conventional digital microscopy systems, the NDP Nanozoomer generates digital images of full tissue sections, allowing large-scale histologic evaluations with high precision across the complete section. Manual cell counting was performed by independent observers on virtual slides. In the center (without invasive margins) of the tumor a grid of fields (each being 1 mm²) was generated, and each field was numbered sequentially. Then (dependent on the spatial extent of the tumor tissue) four to eight fields were selected randomly and evaluated by the observers. Variations in the enumeration within a range of 5% were re-evaluated and a consensus decision made. For the CD3 tumor maps we performed automated counting using the VisioMorph software system from Visiopharm (Hoersholm, Denmark). Briefly, images of 1 mm² regions were produced from complete tissue sections. Tumor-containing images were processed with segmentation and target cell identification (the algorithm was based on spectral and morphologic aspects). Summarized over all patients, a total of 867 mm² of tumor tissue was analyzed for immune cell density.

Results

Using VM we produced a total of 80 virtual slides of primary CRC samples from 20 patients, showing the spatial distribution of CD3, CD8, Granzyme B, and CD45RO. CD3 (Figures 1 and 2) is used as a marker for mature lymphocytes. CD8 acts as a coreceptor for the T cell receptor and is expressed on the surface of cytotoxic T cells, but can also be found on natural killer cells. CD45RO is expressed on most thymocytes and on ~45% of peripheral blood T cells and identifies memory T cells. Granzyme B is crucial for the rapid induction of target cell apoptosis by cytotoxic T lymphocytes and natural killer cells. In the first step, on each virtual slide we randomly selected up to eight windows of 1 mm² of the primary tumor. Positively stained cells were manually counted on virtual slides. For each tissue slide we determined the window with the minimum and maximum number of stained cells, as well as the average over all windows of each section (Figure 3).

For all markers we tested whether a single field evaluation of a patient exceeds the given cutoff, as published by Galon et al.¹³ The analysis of the absolute values of the cell counts obtained for each field for each patient for each biomarker shows, unfortunately, that single field counts can yield clear diagnostic decisions for only a fraction of the analyzed patients (Table I). Analyzing the cell counts for CD3 (suggested cutoff is 370 cells/mm²) resulted in 9 patients (45%) who could not be unambiguously classified. For these patients it was at random (depending on the selected evaluation field) whether they were classified as “good prognosis” or “poor prognosis” patients. For CD8 we identified 12 (60%) patients, for CD45RO 3 (15%) and for Granzyme B 8 (40%) patients who could not be unambiguously classified according to the suggested cutoffs.

We further asked how the results of the individual markers indicated a tendency to lie above or below the cutoffs from Galon et al.¹³ Each patient and marker were counted if the value was above (1) or below (0) the respective threshold. We then tested whether the unambiguous decisions for the individual markers would lead to a consistent decision for the patient (U). We counted 9 of 20 patients (45%) who could be classified unambiguously over all markers when just including unambiguous marker decisions in the concluding decision for each patient.

This led us to the question of whether there may be particular spatial patterns of immune cells of the individual tumors that may explain the diverse immunologic reactions against the tumor in the individual patient. We analyzed the cell count variation across the tumor tissue on complete slides in form of immunologic tumor maps. Figure 4 shows the density distributions within the primary tumors. Here, CD3 is shown as a representative staining with broad heterogeneity over all samples. For all
20 primary CRC samples we evaluated the complete intratumoral cell density of the CD3 staining automatically. Each 1 mm² field contained tumor tissue. In total, an area of 867 mm² was automatically evaluated, with an average of 48 mm² per patient slide. The chosen color scale emphasizes the regional differences across the complete range of immune cell numbers. Figure 4 shows various degrees of heterogeneity across all patient samples, with lower heterogeneity found only in samples with lower cell counts. No samples with a homogeneous high cell density distribution were observed.

**Discussion**

The work of Galon et al.\(^{13}\) sent ripples through the field of oncology in the year 2006. The immune cell densities at the tumor site were shown to have a better predictive power than the classical TNM system. These findings, however, have not been introduced in clinical day-to-day routine, despite clear cutoff values given by the authors. On averaged large patient cohorts, these cutoff values allow prediction of more favorable clinical courses in patients with high immune cell densities at the tumor site. We therefore set out to determine the spatial het-
heterogeneity of these prognostic immune cell surface molecules.\textsuperscript{13,14,16,19-21}

We observed marked differences among individual windows of evaluation in a single patient. In one patient, depending on the selected region, CD3 cell counts varied from 214 cells/mm\textsuperscript{2} to 789 cells/mm\textsuperscript{2}, for example. CD45RO showed in terms of statistical deviation to be the most homogeneous marker in this evaluation, with a median lower deviation (times the median) of 0.42 and an upper deviation of only 0.61 (times the median). Independent of the degree of heterogeneity we observed strongly varying spatial distributions in all patient samples in all four immune cell markers.

Single field evaluations with the given cutoff revealed unambiguous decisions for 8–17 patients depending on the marker. When combining the decisions taken based on the individual markers, only 9 of 20 (45\%) patients could be unambiguously assigned to a group. The results of our study therefore strongly motivate two questions concerning (1) marker identification (the value of the quantitative results from TMAs) and (2) marker application (how heterogeneous markers can be deployed for patient individual classifiers and clinical decisions).

**Marker Identification.** Sampling of multiple cores can allow deducing cutoff values for large cohorts,\textsuperscript{8,22} but further investigation showed that even four or more core biopsies could be required for “compensation” of tumor heterogeneity.\textsuperscript{7} Clearly, using multiple cores may yield better cutoffs when referring to the whole patient cohort. But regardless of how many cores are sampled, such cutoffs may be of little practical value for a diagnostic decision in an individual patient. Moreover, sam-

<table>
<thead>
<tr>
<th>Patients</th>
<th>CD3 (%)</th>
<th>CD8 (%)</th>
<th>CD45RO (%)</th>
<th>Granzyme B (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All fields below or above cutoff</td>
<td>11 (55)</td>
<td>8 (40)</td>
<td>17 (85)</td>
<td>12 (60)</td>
</tr>
</tbody>
</table>

\textbf{Table I} \textit{Unambiguous Identification of Individual Patients Based on TMA Analysis Defined Cutoffs}
pling of multiple cores destroys information that may be hidden in the spatial organization of the tumors, thus making interpretation of results from 1-mm² fields even more erratic. An alternative to TMA is full-tissue slide microscopy for data acquisition. This procedure has three advantages over TMA: (1) the laborious work of repeated TMA sampling is avoided; (2) the statistical uncertainty of number of cores needed is avoided because the whole tissue is scanned; and (3) the spatial context of each area imaged is preserved.

Marker Application. The motivation behind identifying clinical markers by TMA or other techniques is that they are to be applied for the individual patient. Unfortunately, in the case of heterogeneous marker distributions, using TMA during marker acquisition may hinder subsequent clinical application of these markers because the information link between the spatial context of marker patterns and the respective clinical patient outcome is not preserved. This is emphasized by our results, shown graphically in Figure 4, which demonstrates the spatial immune cell heterogeneity in CRC primary tumors. These individual tumor maps show that clinical decisions based on a single 1-mm² field are not feasible, given the observed heterogeneity. If a larger area is evaluated (equivalent here to several 1-mm² fields) the question arises as to how the overall spatial cell density distribution is to be interpreted. Because of the complex patterns observed, it does not appear promising to simply average all field densities into a single cell density value for an individual patient. In contrast, it appears that tumors may be—similar to genetic signatures—characterized by spatial signatures. Whether this is the case may be a matter of debate. If the spatial context of the expression patterns to be

<table>
<thead>
<tr>
<th>Table II</th>
<th>Combination of Cutoffs for Each Patient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stain/Field</td>
<td>1</td>
</tr>
<tr>
<td>CD3</td>
<td>0</td>
</tr>
<tr>
<td>CD8</td>
<td>–</td>
</tr>
<tr>
<td>CD45</td>
<td>–</td>
</tr>
<tr>
<td>GR.B</td>
<td>0</td>
</tr>
<tr>
<td>U</td>
<td>0</td>
</tr>
</tbody>
</table>

0 = all cell counts below cutoff, 1 = all cell counts above cutoff, – = ambiguous distribution, U = unambiguous decisions for this patient.

Figure 4 Density distributions across colorectal primary tumors on complete tissue slides. Pictograms 1–20 show primary CRC tumor tissue differentiated into 1-mm² fields for each of the patients. Each field was analyzed for CD3 positively stained cells and colored respectively.
sampled is not recorded during marker identification, it is irrevocably lost to the task of marker application. Instead, using VM to image full tissue sections preserves this spatial context in digital image form and thus may later allow development of individual patient classifiers incorporating spatial information in a quantitatively based diagnostic decision. Although previous studies did undoubtedly identify significant cutoffs referring to whole patient cohorts, they did not deliver dedicated algorithms to process the complex spatial patterns from full slide imaging into patient individual prognostic or therapeutic decisions.

We suggest tumor maps similar to the ones described here as a tool for incorporating the complex spatial information hidden in the spatial context of the individual tumor into diagnostic algorithm development. In our view, only VM can reliably generate the necessary data basis for developing such spatially based statistical decision tools because it allows the scanning of full tissue slides. Current VM scanners provide performance sufficient to process even large numbers of patient samples. Also, because VM implicitly generates data regarding marker heterogeneity, potential cohorts do not necessarily need to include many hundreds of patients. In this study we did not demonstrate a specific diagnostic algorithm producing a parameter able to interpret the spatial heterogeneity of the tumor. Until this diagnostic algorithm is created and a follow-up study is performed, we cannot know if the tumor map will have prognostic value in the single patient.

In summary, we strongly suggest the integration of VM of full tissue sections in the identification and routine analysis of predictive markers. The fast and reliable evaluation of large tissue sections allows their comprehensive analysis on the level of the individual patient, elegantly circumventing the hurdles of spatial heterogeneity in tissues.

References


