The Prognostic Value of FoxP3+ Tumor-Infiltrating Lymphocytes in Cancer: A Critical Review of the Literature

Ronald J. deLeeuw1,2, Sara E. Kost1,2, Juzer A. Kakal1, and Brad H. Nelson1,2,3

Abstract

CD8+ tumor-infiltrating lymphocytes (TIL) are associated with survival in a variety of cancers. A second subpopulation of TIL, defined by forkhead box protein 3 (FoxP3) expression, has been reported to inhibit tumor immunity, resulting in decreased patient survival. On the basis of this premise, several groups are attempting to deplete FoxP3+ T cells to enhance tumor immunity. However, recent studies have challenged this paradigm by showing that FoxP3+ T cells exhibit heterogeneous phenotypes and, in some cohorts, are associated with favorable prognosis. These discrepant results could arise from differences in study methodologies or the biologic properties of specific cancer types. Here, we conduct the first systematic review of the prognostic significance of FoxP3+ T cells across nonlymphoid cancers (58 studies from 16 cancers). We assessed antibody specificity, cell-scoring strategy, multivariate modeling, use of single compared with multiple markers, and tumor site. Two factors proved important. First, when FoxP3 was combined with one additional marker, double-positive T cells were generally associated with poor prognosis. Second, tumor site had a major influence. FoxP3+ T cells were associated with poor prognosis in hepatocellular cancer and generally good prognosis in colorectal cancer, whereas other cancer types were inconsistent or understudied. We conclude that FoxP3+ T cells have heterogeneous properties that can be discerned by the use of additional markers. Furthermore, the net biologic effects of FoxP3+ T cells seem to depend on the tumor site, perhaps reflecting microenvironmental differences. Thus, depletion of FoxP3+ T cells might enhance tumor immunity in some patient groups but be detrimental in others. Clin Cancer Res; 18(11); 3022–9. ©2012 AACR.

Introduction

Many studies across a wide variety of human cancers have shown a clear association between the presence of tumor-infiltrating lymphocytes (TIL) and patient survival (1–4). To further understand this phenomenon, additional immune markers have been used to subdivide CD3+ T cells into functional subsets, with special emphasis on cytotoxic (e.g., CD8+, nucelolysin TIA-1 isoform p40+) and regulatory [e.g., CD4+, interleukin 2 receptor subunit alpha (CD25)+, FoxP3+] phenotypes (3, 5). Whereas TIL-expressing cytotoxic markers are generally associated with favorable prognosis, TIL-expressing regulatory markers (referred to as Tregs) were initially reported to correlate with poor prognosis (5). This finding fit with the general notion that Tregs suppress adaptive immune responses and led many groups to pursue strategies to deplete Tregs from patients with cancer as a means to enhance tumor immunity (6–8).

In the past decade, much effort has been devoted to finding molecular markers that uniquely define Tregs. Initially, these cells were characterized as CD4+ and CD25high (9). Further investigation revealed that Tregs express and functionally depend on the transcription factor forkhead box protein P3 (FoxP3; ref. 10). Indeed, humans and mice that lack an intact FOXP3 gene suffer a severe autoimmune syndrome known as immune dysregulation/polyendocrinopathy/enteropathy/X-linked syndrome in humans or the Scarp phenotype in mice (10, 11). Given its essential role in Treg development and function, FoxP3 became a popular single marker for Treg studies in cancer. Intriguingly, studies of the prognostic value of FoxP3+ T cells have lead to highly discrepant findings. In some studies, tumor-infiltrating FoxP3+ T cells have been associated with poor prognosis, consistent with the initial hypothesis that FoxP3+ Tregs inhibit antitumor immunity. In contrast, other studies have found that FoxP3+ T cells are associated with a favorable prognosis.

How can these widely discrepant prognostic claims be explained? On the one hand, they could reflect technical differences among studies, including the specific FoxP3 antibody used, scoring strategy, and statistical methods. Alternatively, the differing claims could reflect biologic differences among studies, including the specific FoxP3 antibody used, scoring strategy, and statistical methods. Alternatively, the differing claims could reflect biologic differences among studies, including the specific FoxP3 antibody used, scoring strategy, and statistical methods.
Translational Relevance

Although forkhead box protein P3+ (FoxP3+) T cells are conventionally thought to suppress tumor immunity, this idea has been challenged by recent studies showing that, in some patient cohorts, tumor-infiltrating FoxP3+ T cells are associated with favorable prognosis. To investigate this apparent discrepancy, we did a comprehensive review of the literature on the prognostic significance of tumor-infiltrating FoxP3+ T cells in human cancer. We conclude that FoxP3 is inadequate as a single functional or prognostic marker. Moreover, the prognostic significance of FoxP3+ T cells can vary according to tumor site. Thus, the original view that FoxP3+ T cells invariably suppress tumor immunity is oversimplified. We require better understanding of the functional subtypes of FoxP3+ T cells and their biologic properties in different tumor microenvironments if we wish to rationally modulate their behavior to enhance tumor immunity.

Antibody Specificity

Different FoxP3 antibodies can yield different staining patterns, indicating that some antibodies may have suboptimal sensitivity or specificity (72, 73). Although 18 of the 58 reviewed studies failed to state which specific FoxP3 antibody was used, within the remaining 40 studies, 11 different FoxP3 antibodies were used (Table 1). The most commonly used antibody was a monoclonal designated 236A/E7. In the 23 studies that used 236A/E7, the prognostic significance of FoxP3+ T cells ranged from poor (n = 23), to neutral (n = 23), to good (n = 12). To better understand why the prognostic value of FoxP3+ T cells varies so widely, we assessed each study for technical factors (including the specific FoxP3 antibody used, scoring strategy, and the use of multivariate modeling) and biologic factors (including the use of additional markers to define Tregs and the tumor site studied).

Cell-Scoring Strategy

We investigated 4 aspects of the scoring strategies used to categorize tumors as positive or negative for FoxP3+ T cells: cutoff points, intratumoral location, use of tissue microarrays compared with whole sections, and computerized compared with manual counting (Table 1). Although there is no standard cutoff point for TIL studies, 32 out of 58 of the reviewed studies used the median number of FoxP3+ T cells as the cutoff point. Within these 32, the distribution among poor, neutral, and good prognostic claims was 16, 11, and 5 studies, respectively. The remaining studies used a variety of scoring strategies, including the presence compared with absence of FoxP3+ T cells, the mean number of FoxP3+ T cells, or other criteria. A fairly even distribution among poor, neutral, and good prognostic claims was observed regardless of the cutoff point used (Table 1). Thus, differing scoring strategies do not account for the variable claims of FoxP3 prognostic significance.

TIL can reside in tumor epithelium, stroma, or both, and this may influence their prognostic significance. Among the 58 reviewed studies, 15 did not discriminate between the...
epithelial and stromal location of FoxP3+ T cells and, instead, provided a general count; 19 counted only FoxP3+ T cells in the epithelium; and 24 counted FoxP3+ T cells from epithelial and stromal compartments independently (Table 1). Regardless of the location of enumerated FoxP3+ T cells, a fairly even distribution was seen among poor, neutral, and good prognostic claims.

We next examined the use of tumor tissue microarrays (TMA) compared with whole sections (Table 1). TMAs were used in 18 of the 58 studies, and prognostic claims ranged from poor (n = 7), to neutral (n = 6), to good (n = 5). A similar range of prognostic claims was seen with studies using whole tissue sections. Regarding cell counting, 38 studies used manual counting by one or more investigators, 6 studies used a computer-based quantification method, and 14 studies did not state the counting method (Table 1). Studies that used manual counting showed an unbiased spread among poor (n = 16), neutral (n = 14), and good (n = 8) prognostic claims, regardless of the number of investigators who carried out cell counting. Definitive conclusions could not be drawn regarding the use of computerized counting, as only 6 studies used such methods, 3 of which involved colorectal cancer (see below).

Multivariate Correction for Stage or Grade of Disease

In principle, the density of FoxP3+ T cells could reflect the stage and/or grade of disease, which could influence prognosis. Of the 45 studies that correlated FoxP3+ T cells from both epithelial and stromal compartments independently (Table 1). Regardless of the location of enumerated FoxP3+ T cells, a fairly even distribution was seen among poor, neutral, and good prognostic claims.

Table 1. Characteristics of 58 studies

<table>
<thead>
<tr>
<th>FoxP3 prognosis claim</th>
<th>Poor</th>
<th>Neutral</th>
<th>Good</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>23</td>
<td>23</td>
<td>12</td>
</tr>
</tbody>
</table>

Specific antibody

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clone 42</td>
<td>Custom</td>
<td></td>
</tr>
<tr>
<td>BioLegend</td>
<td>Abcam</td>
<td></td>
</tr>
<tr>
<td>mAbcam2509</td>
<td>Novus Biologicals</td>
<td>FJK-16s</td>
</tr>
<tr>
<td>206D</td>
<td>eBioscience</td>
<td></td>
</tr>
<tr>
<td>236A/E7</td>
<td>mAbcam2510</td>
<td></td>
</tr>
<tr>
<td>259D</td>
<td>eBio7979</td>
<td></td>
</tr>
<tr>
<td>221D/D3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Scoring strategy

<table>
<thead>
<tr>
<th>Cutoff point</th>
<th>Absence and/or presence</th>
<th>Mean cutoff</th>
<th>Median cutoff</th>
<th>Other cutoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Counting location

<table>
<thead>
<tr>
<th>General count</th>
<th>Tumor only</th>
<th>Tumor and stroma</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Tissue used

<table>
<thead>
<tr>
<th>Whole sections</th>
<th>Tissue microarray</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>17</td>
</tr>
</tbody>
</table>

Counting strategy

<table>
<thead>
<tr>
<th>Investigator(s)</th>
<th>Computer program</th>
<th>Not reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>14</td>
<td>8</td>
</tr>
</tbody>
</table>

Multivariate correction for stage or grade

| Yes: 42 | 20 | 11 | 11 |
| No: 16  | 3  | 12 | 1  |

Multivariate correction for other TIL subsets

| Yes: 8 | 2  | 3  | 3  |
| No: 50 | 21 | 20 | 9  |

Use of multiple markers

| Yes: 8 | 4  | 4  |
| No: 50 | 19 | 19 | 12 |

Tumor site

<table>
<thead>
<tr>
<th>Hepatocellular</th>
<th>Cervical</th>
<th>Head and neck</th>
<th>Pancreatic</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

(Continued on the following page)
This issue was addressed in 42 studies by use of multivariate models that included stage, grade, and other clinicopathologic features (Table 1). Among these studies, the prognostic significance of FoxP3+ T cells ranged from poor (n = 20), to neutral (n = 11), to good (n = 11). Notably, in 4 studies, FoxP3+ T cells were a significant univariate prognostic indicator, only to be removed during multivariate analysis. Of the 16 studies that did not use multivariate analysis, the potentially confounding effects of stage and grade were mitigated in most by the fact that (i) FoxP3+ T cells showed no prognostic significance even in univariate analysis or (ii) only specific stages or grades of disease were included in the study. In summary, even though FoxP3+ T cells are frequently associated with the stage and/or grade of disease, we found that this factor was well controlled in most studies and does not account for the variability of FoxP3 prognostication.

Multivariate Correction for Other Tumor-Infiltrating Lymphocyte Subsets

FoxP3+ T cells are usually found together with other TIL subsets, which can make it difficult to discern their independent prognostic effect. Although multivariate analysis can solve this problem, it requires that all TIL subtypes significant in univariate analysis be included in the multivariate model. In the 8 studies that included all TIL subsets in multivariate analysis, the prognostic value of FoxP3+ T cells ranged from poor (n = 2), to neutral (n = 3), to good (n = 3; Table 1). Thus, although the number of studies is low, it seems that the prognostic significance of FoxP3+ T cells is not solely attributable to the presence of other TIL subpopulations.

Several studies made prognostic claims on the basis of the ratio of FoxP3+ T cells to other lymphocyte subsets, including CD3+/FoxP3+ (n = 2), CD4+/CD25+FoxP3+ (n = 1), CD68+/FoxP3+ (n = 2), CD8+/FoxP3+ or FoxP3+/CD8+ (n = 18), CD8+/CCR4+FoxP3+ (n = 1), FoxP3+/CD4+ (n = 2), FoxP3+/CD3+/CD45RO+ (n = 1), and Granzyme-B+/FoxP3+ (n = 1; Supplementary Table S1). Among these 28 studies, prognostic claims for FoxP3+ TIL ranged from poor (n = 12), to neutral (n = 11), to good (n = 5). Thus, the use of lymphocyte ratios has been inconsistently applied and yielded inconsistent prognostic claims.

Clinical Significance and Publication Bias

We next evaluated whether the magnitude of the prognostic effect was similar for studies claiming good compared with poor prognosis. Of the 58 studies, 32 reported multivariate hazard ratios for overall survival. A funnel plot revealed no significant difference between the magnitude of hazard ratios for studies claiming poor compared with good prognosis (Fig. 1). Furthermore, there was no evidence of publication bias, as the studies were evenly distributed throughout the plot.

Use of Multiple Markers to Define FoxP3+ T Cells

Although FoxP3 was originally thought to uniquely define conventional CD4+ Tregs (75), more recent studies indicate that, in some circumstances, FoxP3 can also be expressed by effector T cells (16, 18). We assessed whether studies that subdivided FoxP3+ T cells using a second marker yield more consistent prognostic results. Of the 58 reviewed studies, 50 used FoxP3 as a sole marker, which resulted in variable prognostic claims ranging from poor (n = 19), to neutral (n = 19), to good (n = 12; Table 1). The remaining 8 studies measured at least one marker in addition to FoxP3, including CD4, CD8, CD25, and C-C chemokine receptor 4 (CCR4). Four of these 8 studies showed that FoxP3+ T cells that coexpressed a second marker were associated with poor prognosis. The remaining 4 claimed that the identified subset did not have any prognostic significance. Of note, none of the 8 studies claimed an association with good prognosis. On the basis of the above findings, we investigated more closely which markers were used in addition to FoxP3. Shah and colleagues used 2-color IHC to identify both CD4+FoxP3+ and CD8+FoxP3+ T cells in cervical cancer. Intriguingly, they found CD8+FoxP3+ T cells at a mean number of 3.32 per high-power field and CD4+FoxP3+ T cells at a mean number of 11.45 per high-power field (30). Thus, had they used FoxP3 as a single marker, only ~75% of the cells they measured would have been CD4+ T cells, which underscores the fact that not all FoxP3+ T cells are conventional Tregs. In another study, Watanabe and colleagues used coexpression of CCR4 to delineate a subset of FoxP3-expressing T cells in oral cancer (62). An average of 58% of FoxP3+ cells were found to coexpress CCR4. Whereas total FoxP3+ T cells had no prognostic value (similar to 3 other studies of oral cancer; refs. 60, 61, 63), CCR4+FoxP3+ T cells showed a highly significant association with survival. These studies highlight the importance of using additional markers to account for the heterogeneity of FoxP3+ T cells.

Tumor Site and Subtype

It is conceivable that the biologic and prognostic effect of FoxP3+ T cells depends on microenvironmental context, in which case, tumor site and histologic and/or molecular subtype may be important factors. Indeed, when tumor site was taken into consideration, we found clear prognostic associations in some cases. For example, the 5 studies of hepatocellular cancer unanimously concluded that FoxP3+ T cells are associated with a good prognosis (Table 1). Conversely, 4 out of 10 studies investigating colorectal cancer concluded that FoxP3+ T cells correlated with a good prognosis, whereas the remaining 6 studies found no prognostic association. In considering colorectal cancer, Ladoire and colleagues recently hypothesized that the favorable prognostic effect of FoxP3+ T cells may reflect their ability to suppress tumor-promoting inflammatory responses to gut microbes (76).
In contrast to the above examples, the prognostic significance of FoxP3+ T cells remains controversial in several other cancers. In breast cancer, the reported prognostic effect of FoxP3+ T cells ranges from poor (n = 5), to neutral (n = 1), to good (n = 2). Although ovarian cancer was one of the first tumor sites in which CD4+ Tregs were associated with poor prognosis (5), subsequent studies using FoxP3 as a marker are split among poor (n = 1), neutral (n = 4), and good (n = 2) prognostic claims. Similarly, studies looking at gastric cancers show a split among poor (n = 2), neutral (n = 1), and good (n = 2) prognostic claims. For the remaining 10 tumor sites, the number of published studies is insufficient to make definitive conclusions about the prognostic significance of FoxP3+ T cells.

In addition to tissue of origin, tumors can be classified based on their molecular features, as discussed recently by Ogino and colleagues (77). Hence, it is conceivable that the variability of FoxP3+ T cell prognostication could be attributable to the inherent molecular heterogeneity within tumor types. In support of this idea, the prognostic value of FoxP3+ T cells is stronger in mismatch repair-proficient colorectal cancer compared with mismatch repair-deficient colorectal cancer (31). Similarly, FoxP3+ T cells are prognostically significant in estrogen receptor (ER)+ but not ER− breast cancer (22, 27). In uveal melanoma, FoxP3+ T cells provide prognostic significance in cyclooxygenase-2+ cases (58). Although few in number, these studies suggest that the molecular subtype of tumors may influence the prognostic value of FoxP3 T cells.

Conclusions

Having critically reviewed the literature on the prognostic value of FoxP3+ T cells, we can make several recommendations for future studies. (i) We recommend that prognostic marker studies follow a standard reporting structure, such as the REMARK criteria (78). (ii) In many cancers, FoxP3+ T cells are highly correlated with the stage and grade of disease; therefore, it is important to correct for these and other appropriate clinicopathologic factors. (iii) FoxP3+ T cells are invariably found with other lymphocytes; therefore, all TIL subsets with prognostic value should be included in multivariate models. (iv) The use of multiple markers to identify functional subsets of FoxP3+ T cells can lead to greater clarity about their prognostic value. (v) The prognostic value of FoxP3+ T cells seems to depend significantly on tumor site and possibly molecular subtype, suggesting
that the biologic properties of FoxP3+ T cells are influenced by the tumor microenvironment in which they reside. Overall, this study provides a cautionary note for the concept of depleting FoxP3+ cells from patients with cancer as a means to enhance tumor immunity. Our findings suggest that this strategy may be beneficial for some tumor sites (e.g., liver) but detrimental to others (e.g., colorectal). Improved understanding of the different FoxP3+ T cell subsets in human cancer will likely enable the development of more precise and effective immunotherapies.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

References

13. van Nies JG, Sajet A, Tummers QR, Putter H, Osanto S, et al. Recruitment of regulatory T cells is correlated with hypoxia-induced...


Material

<table>
<thead>
<tr>
<th>Updated version</th>
<th>Access the most recent version of this article at:</th>
<th>doi:10.1158/1078-0432.CCR-11-3216</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplementary Material</td>
<td>Access the most recent supplemental material at:</td>
<td><a href="http://clinncancerres.aacrjournals.org/content/suppl/2012/04/17/1078-0432.CCR-11-3216.DC1">http://clinncancerres.aacrjournals.org/content/suppl/2012/04/17/1078-0432.CCR-11-3216.DC1</a></td>
</tr>
</tbody>
</table>

Cited articles

This article cites 75 articles, 19 of which you can access for free at: http://clincancerres.aacrjournals.org/content/18/11/3022.full.html#ref-list-1

Citing articles

This article has been cited by 12 HighWire-hosted articles. Access the articles at: /content/18/11/3022.full.html#related-urls

E-mail alerts

Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions

To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions

To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.