OXFORD

JNCI J Natl Cancer Inst (2021) 113(1): djaa051

doi: 10.1093/jnci/djaa051
First published online April 23, 2020
Article

Computational Image Analysis of T-Cell Infiltrates in Resectable Gastric Cancer: Association with Survival and Molecular Subtypes

Benjamin R. Challoner, BMBS [1,†] Katharina von Loga, MD [1,2,†] Andrew Woolston, PhD [1] Beatrice Griffiths, PhD,[1] Nanna Sivamanoharan, BSc [1,2] Maria Semiannikova, BSc [1] Alice Newey, BSc,[1] Louise J. Barber, PhD [1] David Mansfield, BSc,[3] Lindsay C. Hewitt, PhD [4] Yuichi Saito, PhD [5] Naser Davarzani, PhD,[4,6] Naureen Starling, MD [7] Alan Melcher, PhD [8] Heike I. Grabsch, PhD [4,9,‡] Marco Gerlinger, MD [1,7,*,‡]

[1]Translational Oncogenomics Laboratory, Division of Molecular Pathology, The Institute of Cancer Research, London, UK; [2]Translational Immuno-Oncology Team, Centre for Molecular Pathology, The Royal Marsden Hospital NHS Foundation Trust and The Institute of Cancer Research, Sutton, UK; [3]Targeted Therapy Team, Division of Radiotherapy and Imaging, The Institute of Cancer Research, London, UK; [4]Department of Pathology, Maastricht University Medical Center, Limburg, The Netherlands; [5]Department of Surgery, Teikyo University School of Medicine, Tokyo, Japan; [6]Biosystems Data Analysis, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands; [7]Gastrointestinal Cancer Unit, The Royal Marsden Hospital NHS Foundation Trust, London, UK; [8]Translational Immunotherapy Team, Division of Radiotherapy and Imaging, The Institute of Cancer Research, London, UK and [9]Pathology & Data Analytics, Leeds Institute of Medical Research at St James's, University of Leeds, St James's University Hospital, Leeds, UK

[†]These authors contributed equally.

[‡]These authors contributed equally.

[*]**Correspondence to:** Marco Gerlinger, MD, FRCP, Translational Oncogenomics Laboratory, The Institute of Cancer Research, London, 237 Fulham Road, London SW3 6JB, UK (e-mail: marco.gerlinger@icr.ac.uk).

Abstract

Background: Gastric and gastro-esophageal junction cancers (GCs) frequently recur after resection, but markers to predict recurrence risk are missing. T-cell infiltrates have been validated as prognostic markers in other cancer types, but not in GC because of methodological limitations of past studies. We aimed to define and validate the prognostic role of major T-cell subtypes in GC by objective computational quantification. **Methods:** Surgically resected chemotherapy-naïve GCs were split into discovery (n = 327) and validation (n = 147) cohorts. CD8 (cytotoxic), CD45RO (memory), and FOXP3 (regulatory) T-cell densities were measured through multicolor immunofluorescence and computational image analysis. Cancer-specific survival (CSS) was assessed. All statistical tests were two-sided. **Results:** CD45RO-cell and FOXP3-cell densities statistically significantly predicted CSS in both cohorts. Stage, CD45RO-cell, and FOXP3-cell densities were independent predictors of CSS in multivariable analysis; mismatch repair (MMR) and Epstein–Barr virus (EBV) status were not statistically significant. Combining CD45RO-cell and FOXP3-cell densities into the Stomach Cancer Immune Score showed highly statistically significant (all $P \leq .002$) CSS differences (0.9 years median CSS to not reached). T-cell infiltrates were highest in EBV-positive GCs and similar in MMR-deficient and MMR-proficient GCs. **Conclusion:** The validation of CD45RO-cell and FOXP3-cell densities as prognostic markers in GC may guide personalized follow-up or (neo)adjuvant treatment strategies. Only those 20% of GCs with the highest T-cell infiltrates showed particularly good CSS, suggesting that a small subgroup of GCs is highly immunogenic. The potential for T-cell densities to predict immunotherapy responses should be assessed. The association of high FOXP3-cell densities with longer CSS warrants studies into the biology of regulatory T cells in GC.

Gastric and gastro-esophageal junction cancers (GCs) are the third most common cause of cancer-related death worldwide (1). Even localized GCs that are treated aggressively with surgery and perioperative chemotherapy recur in approximately 50% of cases (2). Tumor staging is the only prognostic tool in routine clinical use for resectable GCs (3). These tumors are morphologically heterogenous with diffuse to intestinal types and well to poorly differentiated phenotypes, but these offer limited prognostic information (4,5). Molecular characterization identified 4 distinct GC subtypes (6). The most common chromosomally instable GCs often harbor driver gene amplifications, followed by genomically stable GCs with often diffuse-

Received: December 23, 2019; **Revised:** March 5, 2020; **Accepted:** April 2, 2020

ARTICLE

type growth patterns. Microsatellite instable and/or DNA mismatch repair deficient (MMRd) GCs, harboring high mutation loads, and Epstein–Barr virus positive (EBV+) GCs are less common and had a better prognosis than chromosomally instable and genomically stable GCs in some series (7).

In colorectal cancer (CRC), tumor-infiltrating lymphocytes have been validated as prognostic markers, independent of stage and microsatellite instable status (8–10). The so-called immunoscore systematically grades T-cell infiltrates in CRCs, which can for example be used to personalize adjuvant treatment or follow-up strategies.

The prognostic relevance of immune cell infiltrates is less clear in GC. Meta-analyses found associations of high cytotoxic (CD8), helper (CD4), and memory (CD45RO) T-cell infiltrates with better survival (11,12). However, the survival differences between high vs low infiltrate GCs were generally modest. The role of regulatory (FOXP3) T cells, which are considered immunosuppressive, remains unclear with some studies showing an association with longer and others with shorter survival (13).

Immune infiltrates in GC have not been validated as prognostic biomarkers for clinical use because of small cohort sizes in most studies, the use of poorly reproducible manual and semiquantitative T-cell counting, and the lack of validation cohorts (11–13). Moreover, patients in most studies had been treated with a range of different (neo)adjuvant therapies, and whether T-cell infiltrates are truly prognostic in early stage GCs or predictive of (neo)adjuvant treatment success remains unknown. Furthermore, studies were predominated by Asian patients whose tumors can differ from Western patients in their immunological profile (14). This questions the relevance to Western populations.

Immunotherapy with PD1/PDL1 inhibitors showed responses in approximately 10%–15% of GCs (15–18). Defining biomarkers that predict who will benefit is critical to avoid unnecessary toxicities and costs. PDL1-positive GCs had higher response rates, but PDL1-negative tumors also responded (15), so better predictive biomarkers are a major need. T-cell infiltrates correlated with response to checkpoint inhibitors in other cancer types (19). Developing computational approaches for the objective quantification of T-cell subtypes in GC and defining their relevance as markers of immunogenicity should not only lead to new prognostic tools but also may support the development of predictive immunotherapy biomarkers.

We used multicolor immunofluorescence staining and computational image analysis to objectively quantify T cells in 474 GCs resected from Western patients who did not receive (neo)-adjuvant therapy. Splitting cases into discovery and validation cohorts allowed us to identify and subsequently validate T-cell subtypes that associate with cancer-specific survival (CSS) and finally investigate associations with DNA mismatch repair (MMR) and Epstein–Barr virus (EBV) molecular subtypes.

Materials and Methods

Patients and Samples

The use of archival tissue specimens and of clinicopathological data for research had been approved by the Leeds Research Ethics Committee (CA01/122); the need for patient consent was waived by the ethics committee. At the Leeds Teaching Hospital, 0.6 mm cores from archival formalin-fixed and paraffin-embedded (FFPE) GCs resected between 1985 and 2004 had been embedded into tissue microarrays (TMAs). EBV and MMR status had been assessed by RNA in situ hybridization and immunohistochemistry (20).

Multicolor Immunofluorescence Staining

Multicolor immunofluorescence staining of one slide per TMA block was performed with the Opal tumor-infiltrating lymphocyte kit (PerkinElmer, Waltham, Massachusetts) using CD8, CD4, CD45RO, FOXP3, and pan-cytokeratin antibodies and 4′,6-diamidino-2-phenylindole (DAPI) (Supplementary Table 1, available online).

Computational Image Analysis

Slides were scanned with a PerkinElmer Vectra using a 20x objective to detect emission at 520nm, 570nm, 670nm, 620nm and 690nm wavelengths (CD4, CD8, CD45RO, FOXP3 and pan-cytokeratin, respectively; Supplementary Table 2, available online). Signals were unmixed and images exported with PerkinElmer InForm. A pathologist reviewed all 1903 cores to exclude damaged cores and regions with nonmalignant epithelium. Cell quantification was performed on tagged image file format (TIFF) images with the HALO Highplex 3.0 software (Indica Labs, Albuquerque, New Mexico). Following fluorescence image acquisition, slides were hematoxylin and eosin stained and scanned on a Hamamatsu slide scanner with a 40x objective.

Validation of Computational Cell Quantification

Tissue cores were identified and their surface area quantified with the HALO random forest classifier function. Two pathologists defined thresholds in the HALO software for the computational detection of DAPI stained nuclei, FOXP3 cells based on Opal620 fluorophore nuclear detection area setting, and CD4 cells, CD8 cells, and CD45RO cells based on cytoplasmic and nuclear detection area settings with Opal520, 570, and 650 fluorophores, respectively (Settings: Supplementary Table 3, available online). For validation, a pathologist who was blinded to the computational counts manually annotated all cells stained with a T-cell marker using the HALO annotation function.

The fluorescent FOXP3 staining was also validated against a clinically established chromogenic stain (antigen retrieval in CC1, anti-FOXP3 staining with clone 236 A-E7 [eBioscience, San Diego, California] at 1:50 dilution) and scanned on a Hamamatsu slide scanner. Cells were quantified using QuPath (21) (Settings: Supplementary Table 4, available online).

Statistical Analysis

The Spearman rank test was used to measure correlation. CSS was calculated from surgery to GC-related death and analyzed with the Kaplan–Meier method and the log-rank test. Follow-up was calculated for patients alive at last follow-up. A Cox regression analysis with stepwise selection was used for multivariable analyses. Statistical test details are provided in the figures. P values are 2-tailed, and $P < .05$ was considered statistically significant. Statistical tests were performed with R 3.6.1, SPSS 25, or Graphpad Prism.

Table 1. Clinical and pathological characteristics of the discovery and validation cohorts

Variables	Discovery cohort (n = 341)	Validation cohort (n = 154)	P[a]
Median year of resection	1997	1992	
Median age (range), y	72.0 (29.4–90.0)	70.5 (33.8–90.5)	.30
Sex			
Male	63.3% (216)	66.2% (102)	
Female	36.1% (123)	33.8% (52)	.59
pT (UICC TNM 7th edition)			
pT1	8.8% (30)	5.8% (9)	
pT2	7.9% (27)	10.4% (16)	
pT3	28.4% (97)	36.4% (56)	
pT4	54.8% (187)	47.4% (73)	.30
pN (UICC TNM 7th edition)			
pN0	25.9% (88)	25.3% (39)	
pN1 to pN3b	74.1% (252)	74.7% (115)	.79
pM (UICC TNM 7th edition)			
pM0/Mx	97.4% (332)	95.5% (147)	
pM1	2.6% (9)	4.5% (7)	.40
Stage (UICC TNM 7th edition)			
I	12.6% (43)	9.1% (14)	
II	24.4% (83)	28.6% (44)	
III	60.3% (205)	57.8% (89)	
IV	2.6% (9)	4.5% (7)	.69
Lauren classification			
Intestinal	56.0% (191)	75.2% (115)	
Diffuse	27.3% (93)	13.7% (21)	
Mixed	16.4% (56)	11.1% (17)	<.001
MMR status			
Proficient	91.7% (299)	84.4% (130)	
Deficient	8.3% (27)	15.6% (24)	.02
No MMR data available	4.3% (15)	0.0% (0)	
EBV status			
Negative	97.5% (306)	92.2% (141)	
Positive	2.5% (8)	7.8% (12)	.02
No EBV data available	7.7% (27)	0.6% (1)	

[a]Two-sided, χ^2 tests. EBV = Epstein–Barr virus; MMR = mismatch repair; TNM = Tumour, Regional Lymph Node and Metastasis Classification of Malignant Tumours; UICC = Union for International Cancer Control.

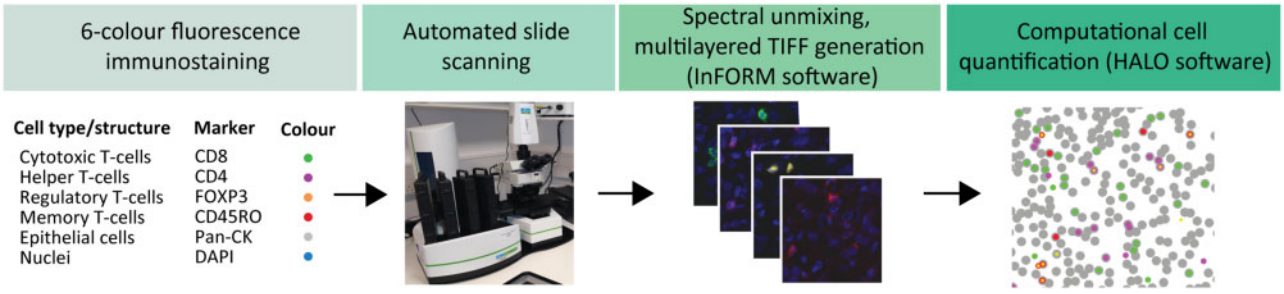

Figure 1. Immune staining and computational analysis workflow.

Results

Clinical Characteristics of the Discovery and Validation Cohorts

FFPE samples from 503 resected GCs from the Leeds Teaching Hospital had been embedded in 14 TMA blocks. Each GC was represented by a minimum of 2 cores from the area of highest tumor cell density.

This cohort was split approximately 2 to 1 into a discovery cohort (n = 349, the younger tissue samples) and a validation cohort (n = 154, older tissue samples). Splitting by tissue age allowed to assess and control for a potential decline of antigenicity over time (22). Patient age, sex, and tumor stage were balanced between the cohorts (Table 1). EBV+, MMRd, and intestinal-type tumors were more common in the validation cohort. This may be because of changes in GC biology over the last decades (23) and random variation when analyzing small subgroups. Eight cases that received

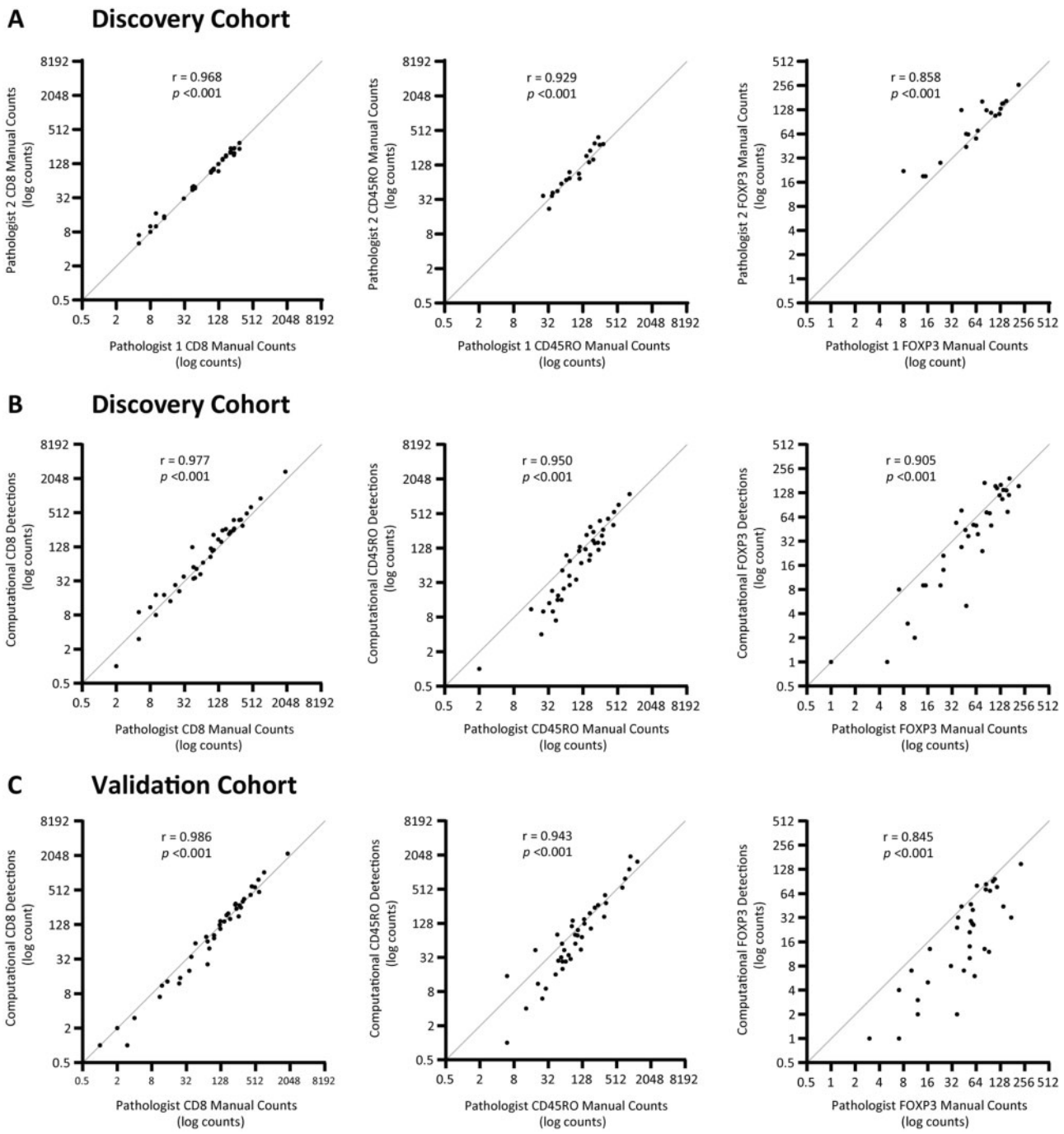

Figure 2. Correlation of pathologist and computational immune cell quantification. **A)** Correlation of CD8 cells, CD45RO cells, and FOXP3 cells counted independently by 2 pathologists (n = 20 cores). **B)** Correlation of computational quantification with counts by a pathologist in the discovery cohort (n = 40). **C)** Correlation of computational quantification with counts by a pathologist in the validation cohort (n = 40). The **grey** 45-degree line indicates where identical counts lie; where computational counts were greater than manual counts, the data points are above the line, and where computational counts were lower than manual counts, the data points are below. The Spearman correlation coefficient and *P* values are shown. All tests were two-sided.

chemotherapy were excluded, leaving only GCs treated with surgery alone. CSS was available for 327 cases in the discovery cohort (median follow-up 6.6 years) and for 147 in the validation cohort (median follow-up 7.3 years). These constituted the final analysis groups. CSS was lower in the discovery cohort than in the validation cohort (Supplementary Figure 1, available online). Consistent with recently published data for Western patients (24), survival of MMRd and MMR proficient (MMRp) GC cases was similar (Supplementary Figure 2, available online), and EBV+ cases had a better survival (7). The higher proportion of EBV+ GCs may therefore contribute to the better survival of the validation cohort.

Multicolor Immunofluorescence Staining

Each cohort was batch stained for CD8 (cytotoxic), CD4 (helper), CD45RO (memory), and FOXP3 (regulatory) T cells and pan-cytokeratin (epithelial cells) (workflow: Figure 1).

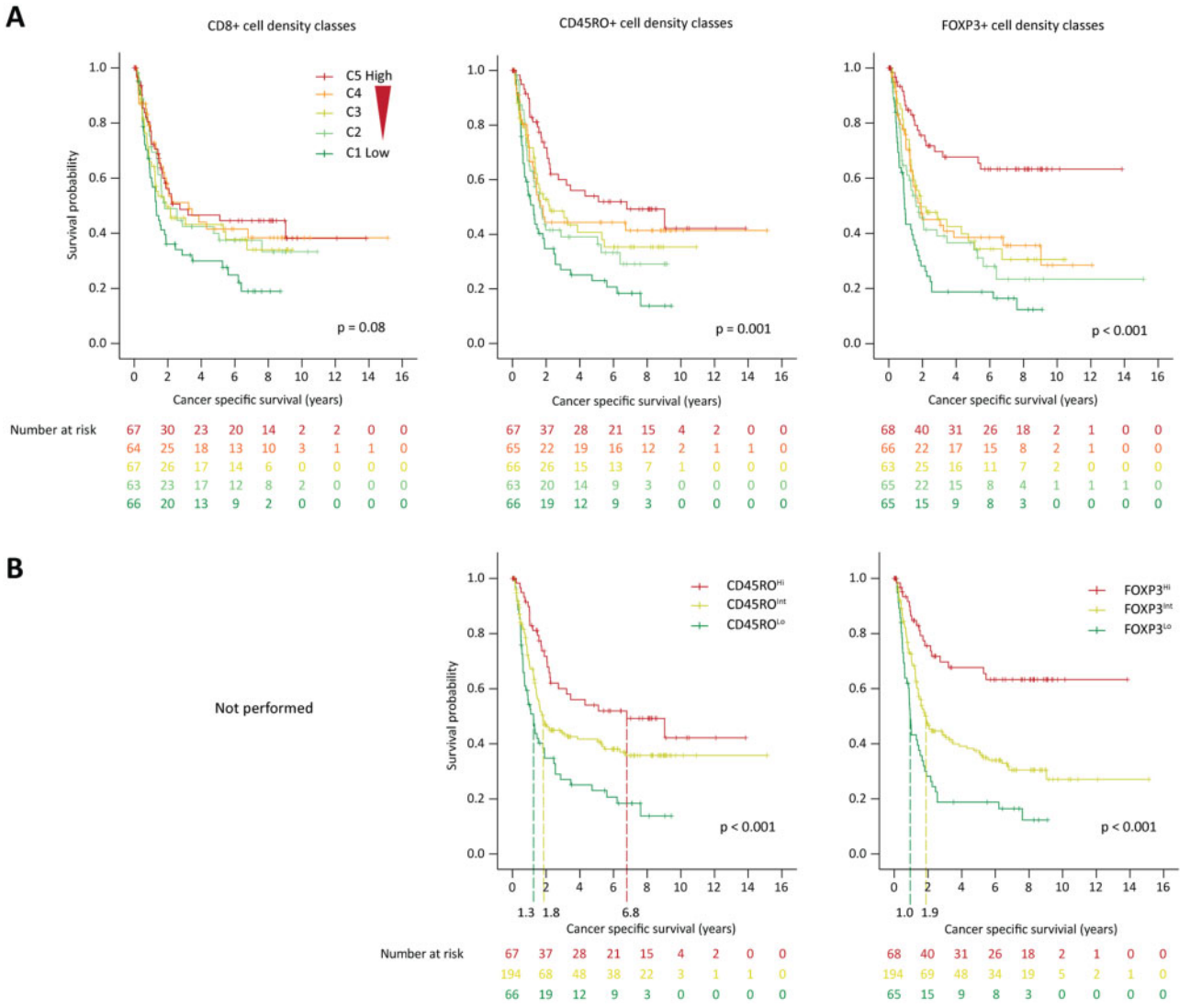

Figure 3. Kaplan–Meier analysis of cancer-specific survival by CD8-cell, CD45RO-cell, and FOXP3-cell density in the discovery cohort. **A)** Cancer-specific survival for each of the 5 equal-sized groups. **B)** Cancer-specific survival for the 3 density groups. **Dashed lines** indicate the median survival time for individual groups. *P* values were calculated with a log-rank test. All tests were two-sided.

Training and Validation of the Computational Image Analysis

Following scanning with an automated microscope, 2 pathologists defined the threshold settings for computational cell detection. Autofluorescence of elastin fibers and nonspecific staining led to high false-positive numbers in the CD4 channel. A threshold for reliable CD4-cell detection could therefore not be defined, and these were not included in the analysis. CD8 and CD45RO stains showed specific membranous staining of cells identifiable as lymphocytes on subsequent hematoxylin and eosin staining of the same slides. FOXP3 showed dim to intense levels of nuclear staining. All threshold settings were optimized to avoid false-positive detection.

For each marker, stained cells were also counted in 20 randomly chosen cores independently by 2 pathologists. Manual counts from both pathologists showed a high correlation (Spearman r = 0.858–0.968; all $P < .001$), demonstrating that the densities of these cells can be reproducibly determined (Figure 2A). A pathologist who was blinded to the computational analysis results subsequently counted cells in 40 cores of each cohort. Comparison with the cell counts from the optimized computational quantification showed a high correlation (Spearman r = 0.845–0.986; all *P* values < .001; Figure 2, B and C). Computational quantification had a tendency to underestimate cell numbers (apparent in Figure 2, B and C, where data points deviate below the 45-degree line, which indicates perfect agreement), particularly when there were few immune cells per core and more pronounced for FOXP3 cells and CD45RO cells than for CD8 cells. However, the high Spearman correlation coefficient shows that this does not substantially impair the ranking of samples relative to each other. This validated the computational cell quantification.

CD8-cell, CD45RO-cell, and FOXP3-cell densities per square millimeter were calculated for each core, and the average density across all cores per GC case was used for analysis. All immune cell types showed higher densities in the discovery cohort compared with the validation cohort (Supplementary Figure 3, available online). This could be a consequence of the higher tissue age in the validation cohort (Table 1), which can

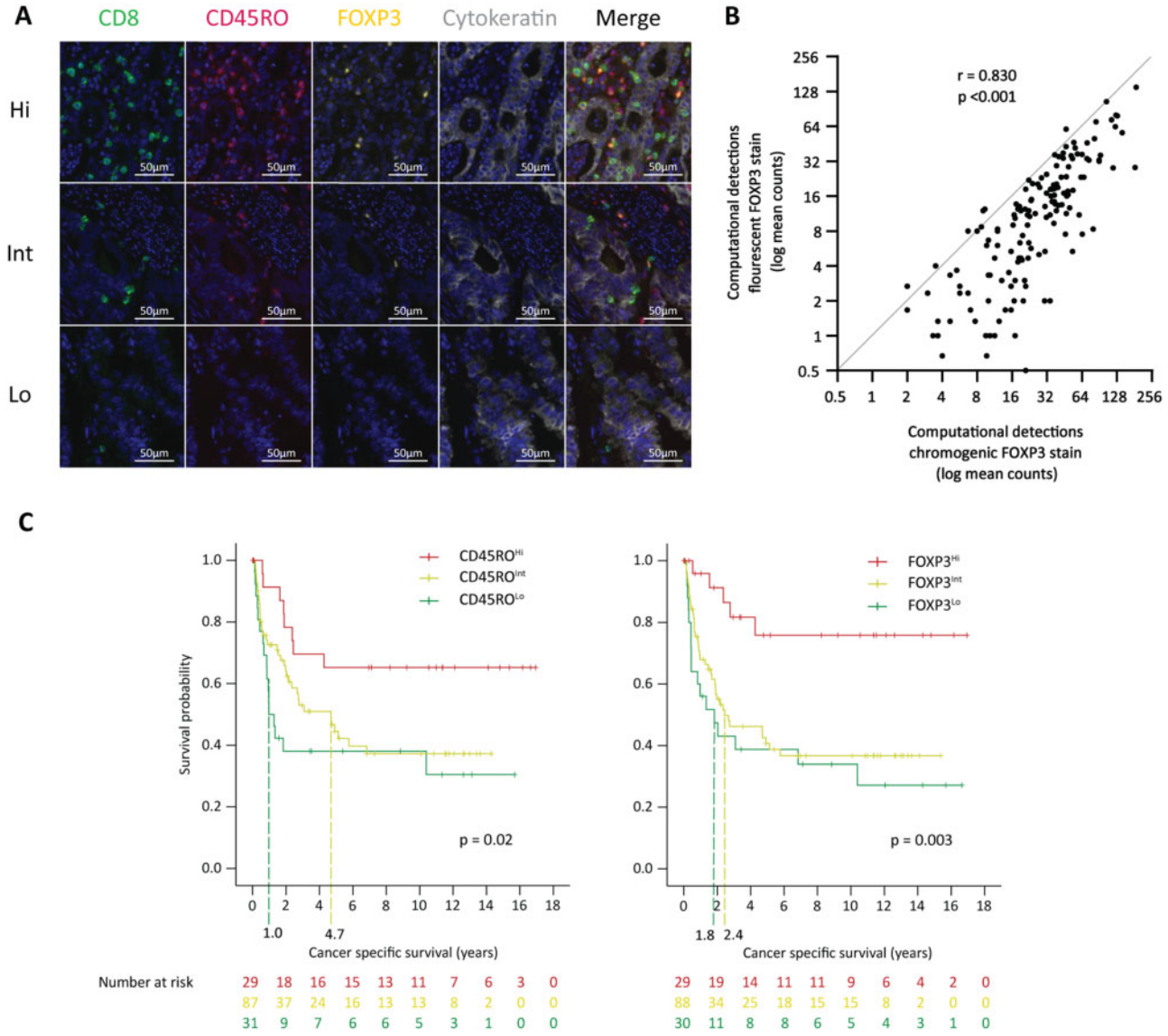

Figure 4. Multimodal data validation. **A)** Representative multicolor fluorescence images of TMA cores with high (Hi), intermediate (Int), and low (Lo) density infiltrates of CD8 cells (**green**), CD45RO cells (**red**), FOXP3 cells (**yellow**). All scale bars = 50 μm. **B)** Correlation of computationally counted FOXP3 cells stained with chromogenic vs fluorescent immunohistochemistry from 167 patients. The **grey** 45-degree line indicates where identical counts lie. The Spearman correlation coefficient and *P* values are shown. **C)** Kaplan–Meier analysis of cancer-specific survival by CD45RO-cell and FOXP3-cell density in the validation cohort. **Dashed lines** indicate the median survival time for individual groups. *P* values were calculated with a log-rank test. All tests were two-sided.

impair antigen stability or of batch effects of multicolor immunofluorescence staining.

Correlation of T-Cell Densities With Cancer-Specific Survival in the Discovery Cohort

The discovery cohort was split into 5 equal-sized groups based on the density of each of the immune cell subtypes (C1-lowest to C5-highest densities). CSS did not statistically significantly differ for CD8 cells ($P = .08$; Figure 3A), although tumors with the lowest densities (C1) showed a trend toward inferior survival. CD45RO-cell ($P = .001$) and FOXP3-cell ($P < .001$) densities were both statistically significantly associated with CSS and showed similar patterns; tumors with the highest densities (C5) had the best survival, those with the lowest densities (C1) had the poorest CSS with a rapid decline over the first 2 years, and groups C2–C4 showed intermediate survival.

We therefore reclassified groups C2–C4 into 60% of cases with intermediate ($CD45RO^{Int}$, $FOXP3^{Int}$), 20% of cases with low ($CD45RO^{Lo}$, $FOXP3^{Lo}$), and 20% of cases with high densities ($CD45RO^{Hi}$, $FOXP3^{Hi}$). These consolidated groups showed highly statistically significant CSS differences ($P < .001$; Figure 3B) with clinically meaningful differences in median CSS. Examples of immune infiltrates in GCs with low, intermediate, and high immune cell densities are shown (Figure 4A).

Validation of the FOXP3 Staining

The strong association of higher FOXP3-cell infiltrates with better CSS was surprising as regulatory T cells are immunosuppressive and predict for a poor prognosis in some cancer types

Table 2. Final statistically significant variables of the multivariable Cox regression analysis of the discovery cohort and assessment of these in the validation cohort

	Discovery cohort		Validation cohort	
Variable	HR (95% CI)	P[a]	HR (95% CI)	P[a]
CD45RO-cell density				
Hi	1.00 (Reference)	—	1.00 (Reference)	—
Int	1.71 (1.11 to 2.65)	.02	1.59 (0.74 to 3.44)	.24
Lo	2.09 (1.25 to 3.48)	.005	2.59 (1.09 to 6.18)	.03
FOXP3-cell density				
Hi	1.00 (Reference)	—	1.00 (Reference)	—
Int	2.00 (1.21 to 3.29)	.007	3.19 (1.25 to 8.15)	.02
Lo	2.79 (1.54 to 5.08)	.001	3.28 (1.19 to 9.06)	.02
Stage (UICC TNM 7th Edition)				
pT3/4	1.00 (Reference)	—	1.00 (Reference)	—
pT1/2	0.22 (0.10 to 0.45)	<.001	0.29 (0.10 to 0.80)	.02
pN1–3	1.00 (Reference)	–	1.00 (Reference)	—
pN0	0.45 (0.29 to 0.70)	<.001	0.37 (0.18 to 0.74)	.005

[a]Two-sided, Cox regression analysis. CI = confidence interval; HR = hazard ratio; TNM = Tumour, regional lymph node and metastasis classification of malignant tumours; UICC = UICC = Union for International Cancer Control.

(25). We therefore validated the fluorescent FOXP3 staining against a clinically established chromogenic FOXP3 stain. Staining of 472 cores from 167 GCs with this assay followed by computational quantification showed a high correlation (Spearman $r = 0.830$; $P < .001$) with fluorescent staining (Figure 4B). Computational cell counts of fluorescent FOXP3 staining were systematically lower than computational counts of the chromogenic stain, revealing systematic biases between the methods. Yet, the high correlation coefficient shows that the ranking of samples remains consistent and validates fluorescent FOXP3-cell quantification.

Validation of the Prognostic Role of CD45RO- and FOXP3-Cell Infiltrates

We next assessed whether CD45RO- and FOXP3-cell densities were also prognostic in the validation cohort. Identically to the approach used in the discovery cohort, cases were split into tumors with the highest 20%, intermediate 60%, and lowest 20% of immune cell infiltrates. CD45RO-cell and FOXP3-cell densities were also statistically significantly associated with CSS in this cohort ($P = .02$ and $P = .003$, respectively; Figure 4C), validating them as prognostic markers in Western GC patients. CSS of the $CD45RO^{Lo}$ and $FOXP3^{Lo}$ groups was less distinct from that of the $CD45RO^{Int}$ and $FOXP3^{Int}$ groups in the validation cohort compared with the discovery cohort. Yet, the rapid decline over the first 2 years remained apparent for $CD45RO^{Lo}$ and $FOXP3^{Lo}$ cases, suggesting that these are important subgroups.

Multivariable Analysis

We next investigated whether CD45RO cells and FOXP3 cells were independent predictors of CSS by analyzing them with tumor stage, Lauren classification, EBV, and MMR subtypes in a multivariable analysis. CD45RO-cell and FOXP3-cell densities, as well as pathological tumour (pT) and pathological regional lymph node (pN) stage, were the only statistically significant predictors of CSS in the discovery and validation cohorts (Table 2; Supplementary Table 5, available online).

Combining CD45RO- and FOXP3-Cell Densities Into the Stomach Cancer Immune Score

Because CD45RO-cell and FOXP3-cell densities were independent prognostic factors, we investigated whether they could be combined to further refine CSS prediction. A 3x3 contingency table defined 9 possible combinations of CD45RO-cell and FOXP3-cell density groups (Figure 5A). The most divergent combinations of $CD45RO^{Hi}FOXP3^{Lo}$ and $CD45RO^{Lo}FOXP3^{Hi}$ comprised very few cases ($n = 11$ and $n = 2$, respectively), which precluded meaningful analysis, and were excluded. CSS was similar for some groups, allowing consolidation into 4 categories, termed the Stomach Cancer Immune-Score (STIM-score): $CD45RO^{Int}FOXP3^{Lo}$, $CD45RO^{Lo}FOXP3^{Int}$, and $CD45RO^{Lo}FOXP3^{Lo}$ were combined into STIM1; $CD45RO^{Hi}FOXP3^{Int}$ and $CD45RO^{Int}FOXP3^{Hi}$ into STIM3. $CD45RO^{Hi}FOXP3^{Hi}$ showed the best CSS and were defined as STIM4, and the largest group of $CD45RO^{Int}FOXP3^{Int}$ tumors was defined as STIM2. Re-analysis by STIM-score was highly statistically significant in the discovery ($P < .001$; Figure 5B) and validation cohorts ($P = .002$; Figure 5C).

Association of CD45RO- and FOXP3-Cell Densities With Molecular Characteristics

MMRd and EBV+ GCs are considered particularly immunogenic because of high mutation burdens and virus presence, respectively. We therefore assessed how immune cell densities differed between these subtypes. Of EBV+ tumors from the discovery and the validation cohort, 87.5% and 50.0%, respectively, were STIM3 or STIM4 (Figure 5, B and C), supporting higher immune recognition of EBV+ compared with MMRp&EBV- GCs of which only 30.9% in the discovery cohort and 26.2% in the validation cohort were STIM3 or STIM4. The percentage of GCs that were classed as STIM3 or STIM4 was similar for MMRd GCs and MMRp&EBV- GCs (Figure 5, B and C).

Comparing immune cell densities directly between molecular subgroups showed that CD45RO-cell densities were statistically significantly higher in EBV+ than in MMRp&EBV- GCs in the discovery and the validation cohorts (Figure 6, A and B). FOXP3-cell densities were statistically significantly higher in

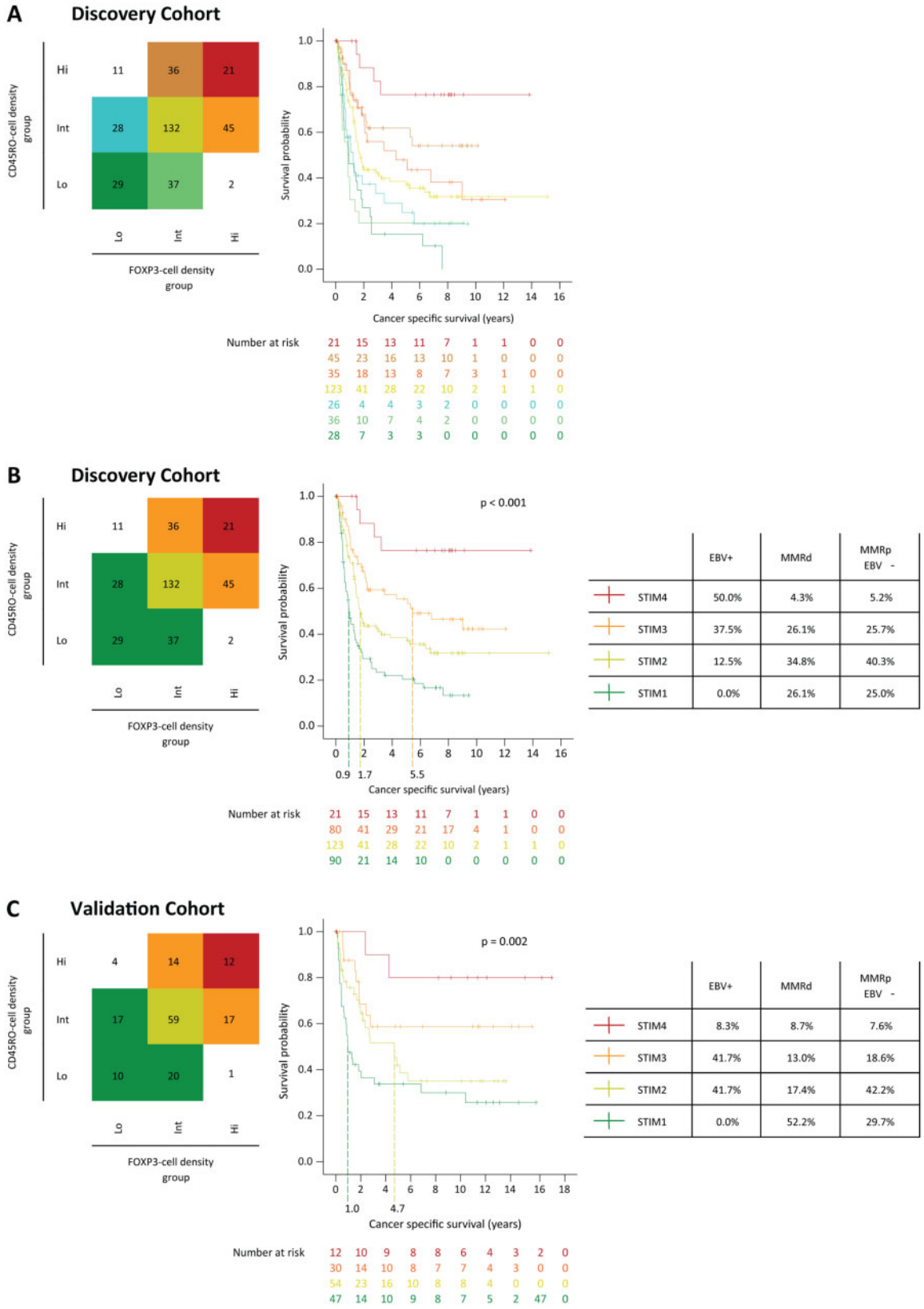

Figure 5. Combination of CD45RO-cell and FOXP3-cell densities into the Stomach Cancer Immune-Score (STIM-score). **A)** A 3x3 contingency table of CD45RO-cell and FOXP3-cell density classes. A Kaplan–Meier analysis of the 7 color-coded groups is shown on the right. **B)** Kaplan-Meier analysis of the consolidated STIM-score in the discovery cohort and **(C)** the validation cohort. **Dashed lines** indicate the median survival time for individual groups. The data table shows the distribution of EBV+, MMRd, and MMRp/EBV- cases according to STIM-score. *P* values were calculated with a log-rank test. All tests were two-sided. EBV = Epstein–Barr virus; MMRd = mismatch repair deficient; MMRp = mismatch repair proficient.

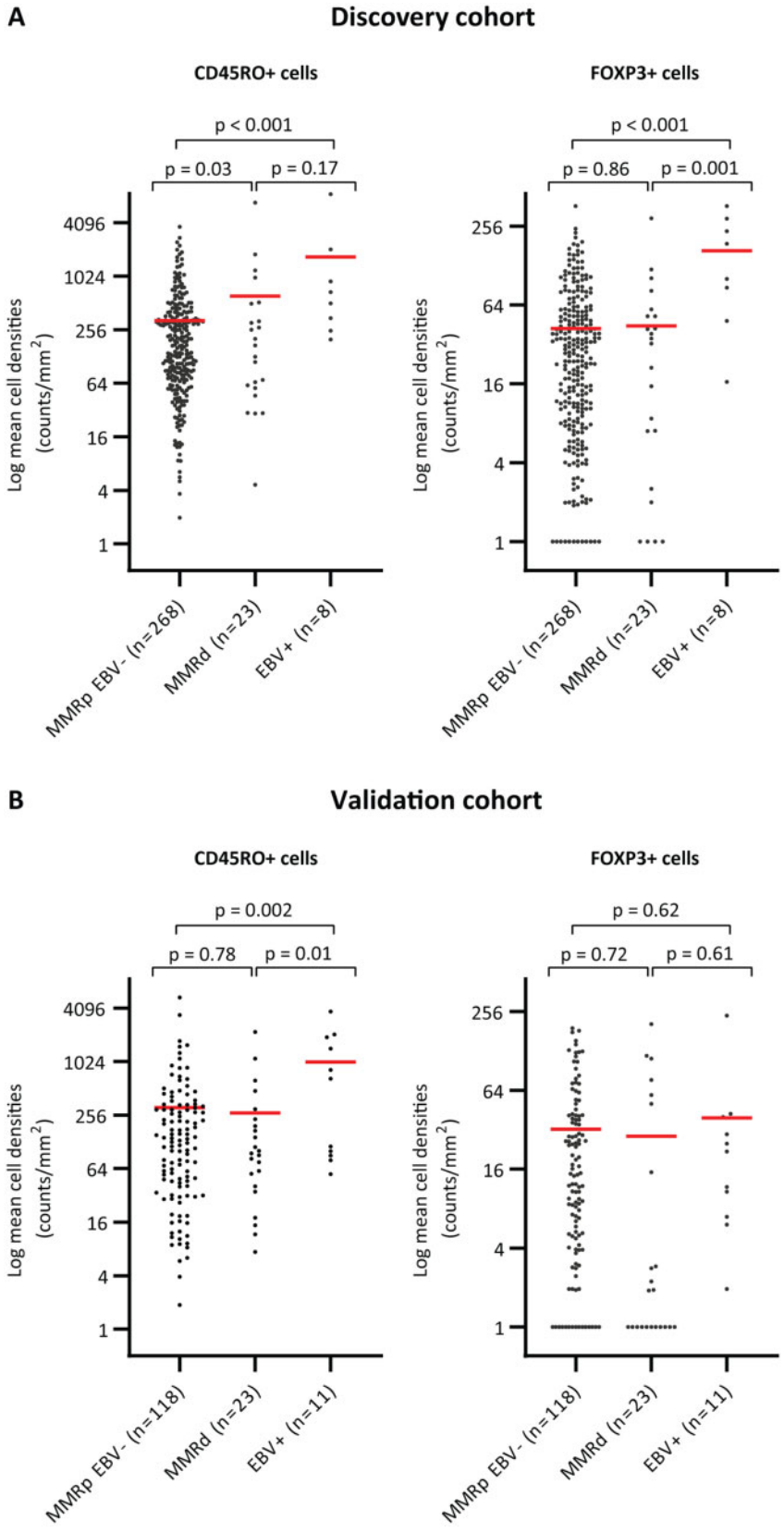

Figure 6. Association of CD45RO-cell and FOXP3-cell densities with MMR and EBV status in **(A)** the discovery cohort and **(B)** the validation cohort. Densities were offset by 1 before log transformation. **Horizonal bars** indicate the mean, and *P* values were calculated with unpaired *t* tests on nontransformed data. All tests were two-sided. EBV = Epstein–Barr virus; MMRd = mismatch repair deficient; MMRp = mismatch repair proficient.

EBV+ GCs compared with MMRp&EBV- tumors in the discovery cohort (Figure 6A) but not in the validation cohort (Figure 6B). MMRd GCs only showed a higher density of CD45RO cells compared with MMRp&EBV- GCs in the discovery cohort (Figure 6A). The small number of EBV+ and MMRd cases may have contributed to these differences between discovery and validation cohorts.

Discussion

Multicolor immunofluorescence coupled with computational image analysis identified and validated CD45RO-cell and FOXP3-cell densities as prognostic markers in Western GC patients, independent of stage, MMR and EBV status, and other pathological features. Those 20% of patients with the highest densities of CD45RO cells and of FOXP3 cells had a particularly good CSS, whereas the 3 groups with intermediate densities showed similar CSS. GCs with the lowest densities showed a rapid early decline in CSS. Most previous GC studies defined high and low T-cell infiltrate groups based on a median cutoff value (11–13), which may explain the weak prognostic effect they found for immune infiltrates compared with the large differences shown by our analysis.

Our results provide important insights into GC immunobiology, suggesting that immunogenicity is low or that immunosuppressive factors constrain immune recognition in the majority of tumors, so that only 20% of cases achieve major survival benefits from T-cell infiltration. CD45RO cells are considered long-lived memory T cells, which are generated in response to cognate antigen recognition (10). Abundant CD45RO cells may, hence, identify GCs that have been actively detected by T cells. In contrast, CD8 cells were not statistically significantly associated with CSS, perhaps indicating that a large proportion of these are passive bystanders that do not recognize cancer cells (26). FOXP3 is a marker for regulatory T cells described as immunosuppressive. The paradoxical association of a cell type that is thought to inhibit antitumor immunity with a good prognosis may point to the existence of distinct subtypes of suppressive and nonsuppressive regulatory T cells, as recently described in CRCs (27). Furthermore, in vitro experiments reported FOXP3 expression as an early activation marker in T cells without suppressive function (28). Our results warrant further investigation of the biology of FOXP3 cells in GC and suggest caution when applying immunotherapies that inhibit or deplete regulatory T cells in GCs.

With response rates for PD1/PDL1 inhibitors of 10%–15% (16–18), immunotherapy sensitivity is also confined to a small subgroup of GCs. In addition, PD1/PDL1 inhibitors are predominantly effective against tumors that are spontaneously recognized by the immune system, which manifests in higher T-cell infiltrates and interferon-gamma signatures, among others (29). Investigating whether high CD45RO-cell or FOXP3-cell infiltrates can identify GCs with high spontaneous immunogenicity that will also respond to immunotherapy will be an important next step, particularly as PDL1 expression is a poor predictive biomarker for checkpoint inhibitors in GC (18,30). Understanding the molecular basis that results in low immune cell infiltrates in 20% of cases with the associated rapid survival decline may lead to novel therapeutic opportunities for this group of patients.

Our data furthermore revealed high CD45RO-cell and FOXP3-cell infiltrates in EBV+ GCs. Surprisingly, T-cell densities in MMRd GCs were similar to MMRp/EBV- GCs. We recently showed that the hypermutator phenotype in MMRd GCs confers extreme intratumor heterogeneity that enables the evolution of multiple genetic immune-evasion events within individual tumors (31). This ability to readily acquire mutations in immune-evasion regulators, and potentially the activity of non-genetic immune escape mechanisms such as high beta-catenin activity (32) or mesenchymal features (33), may perhaps explain low immune infiltrates in many MMRd GCs.

This large study in GC patients who had been treated with surgery alone defines T-cell subtypes that influence the natural history of GCs. This unique cohort can be used as a comparator when assessing the predictive role of immune cells in GC immunotherapy trials that do not include an untreated control group. Comparison to GCs treated with perioperative or adjuvant chemotherapy should be undertaken to further define the predictive role of T-cell infiltrates for chemotherapy outcomes.

We finally devised a strategy combining both cell types into the STIM-score, which is more straightforward to apply in the clinic than 2 separate markers and identifies patient groups with clinically meaningful differences in CSS. This could be useful to identify patients with low recurrence risks who may not require intensive adjuvant or perioperative chemothrapy or chemoradiotherapy or to prioritize those patients with very poor survival outcomes for treatment intensification trials.

Before the STIM-score can be clinically applied, optimal T-cell density cutoffs should be defined in diagnostic GC biopsies and in resected GCs whole slides because these may differ from cutoffs determined in TMAs. Limitations of this study include a lack of CD4-cell analysis, the specimen age of up to 33 years, and the absence of samples from tumor margins in the TMA, which precluded investigating tumor margins and centers similar to the immunoscore approach (8–10). Finally, independent validation in GC cohorts from additional centers should be undertaken.

Funding

MG receives funding from Cancer Research UK, the Schottlander Foundation, and the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement No. 820137). MG, BC, NSt, and KvL receive funding from the Royal Marsden Hospital/ICR NIHR Biomedical Research Centre for Cancer.

Notes

Role of the funder: The funders had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; and the decision to submit the manuscript for publication.

Conflicts of interest: MG receives research funding from Merck KG and BMS; NSt from BMS, Astra Zeneca, Pfizer, Merck KG and honoraria from Eli-Lilly, Astra Zeneca, MSD, Merck KG and Pierre Fabre. The other authors have no conflicts of interest to declare.

Author contributions: MG, HG, and KvL planned the study. KvL and MG obtained funding. MG led the study. HG provided clinical samples, data, and statistical expertise. LH and YS were involved in data collection. BC, KvL, NSi, and DM optimized and performed staining, scanning, and computational image analysis, and BG provided technical support. BC analyzed all data with help from HG, KvL, LB, MS, AN, and AW. BC, KvL, ND, NSt, AM, HG, and MG interpreted the data. All authors were involved

in manuscript writing and gave final approval of the manuscript.

References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394–424.
2. Al-Batran SE, Homann N, Pauligk C, et al. Perioperative chemotherapy with fluorouracil plus leucovorin, oxaliplatin, and docetaxel versus fluorouracil or capecitabine plus cisplatin and epirubicin for locally advanced, resectable gastric or gastro-oesophageal junction adenocarcinoma (FLOT4): a randomised, phase 2/3 trial. *Lancet*. 2019;393(10184):1948–1957.
3. Brierley J, Gospodarowicz MKe, Wittekind C. *TNM classification of malignant tumours*. 8th ed. New Jersey, NJ: Wiley-Blackwell: 2017.
4. Gullo I, Oliveira P, Athelogou M, et al. New insights into the inflamed tumor immune microenvironment of gastric cancer with lymphoid stroma: from morphology and digital analysis to gene expression. *Gastric Cancer*. 2019; 22(1):77–90.
5. Lee JH, Chang KK, Yoon C, et al. Lauren histologic type is the most important factor associated with pattern of recurrence following resection of gastric adenocarcinoma. *Ann Surg*. 2018;267(1):105–113.
6. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. 2014;513(7517):202–209.
7. Cristescu R, Lee J, Nebozhyn M, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med*. 2015; 21(5):449–456.
8. Anitei MG, Zeitoun G, Mlecnik B, et al. Prognostic and predictive values of the immunoscore in patients with rectal cancer. *Clin Cancer Res*. 2014;20(7):1891–1899.
9. Galon J, Costes A, Sanchez-Cabo F, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science*. 2006;313(5795):1960–1964.
10. Galon J, Pages F, Marincola FM, et al. The immune score as a new possible approach for the classification of cancer. *J Transl Med*. 2012;10:1. doi: 10.1186/1479-5876-10-1.
11. Yu PC, Long D, Liao CC, et al. Association between density of tumor-infiltrating lymphocytes and prognoses of patients with gastric cancer. *Medicine (Baltimore)*. 2018;97(27):e11387.
12. Zheng X, Song X, Shao Y, et al. Prognostic role of tumor-infiltrating lymphocytes in gastric cancer: a meta-analysis. *Oncotarget*. 2017;8(34):57386–57398.
13. Huang Y, Liao H, Zhang Y, et al. Prognostic value of tumor-infiltrating FoxP3+ T cells in gastrointestinal cancers: a meta analysis. *PLoS One*. 2014;9(5): e94376.
14. Lin SJ, Gagnon-Bartsch JA, Tan IB, et al. Signatures of tumour immunity distinguish Asian and non-Asian gastric adenocarcinomas. *Gut*. 2015;64(11): 1721–1731.
15. Fuchs CS, Doi T, Jang RW, et al. Safety and efficacy of pembrolizumab monotherapy in patients with previously treated advanced gastric and gastroesophageal junction cancer: phase 2 clinical KEYNOTE-059 trial. *JAMA Oncol*. 2018;4(5):e180013.
16. Kang YK, Boku N, Satoh T, et al. Nivolumab in patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant of, at least two previous chemotherapy regimens (ONO-4538-12, ATTRACTION-2): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet*. 2017; 390(10111):2461–2471.
17. Muro K, Chung HC, Shankaran V, et al. Pembrolizumab for patients with PD-L1-positive advanced gastric cancer (KEYNOTE-012): a multicentre, open-label, phase 1B trial. *Lancet Oncol*. 2016;17(6):717–726.
18. Wang F, Wei XL, Wang FH, et al. Safety, efficacy and tumor mutational burden as a biomarker of overall survival benefit in chemo-refractory gastric cancer treated with toripalimab, a PD-1 antibody in phase IB/II clinical trial NCT02915432. *Ann Oncol*. 2019;30(9):1479–1486.
19. Tumeh PC, Harview CL, Yearley JH, et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature*. 2014;515(7528):568–571.
20. Hewitt LC, Inam IZ, Saito Y, et al. Epstein-Barr virus and mismatch repair deficiency status differ between oesophageal and gastric cancer: a large multicentre study. *Eur J Cancer*. 2018;94:104–114.
21. Bankhead P, Loughrey MB, Fernandez JA, et al. QuPath: open source software for digital pathology image analysis. *Sci Rep*. 2017;7(1):16878.
22. Combs SE, Han G, Mani N, et al. Loss of antigenicity with tissue age in breast cancer. *Lab Invest*. 2016;96(3):264–269.
23. Arnold M, Park JY, Camargo MC, et al. Is gastric cancer becoming a rare disease? A global assessment of predicted incidence trends to 2035. *Gut*. 2020; 69(5):823–829.
24. Pietrantonio F, Miceli R, Raimondi A, et al. Individual patient data meta-analysis of the value of microsatellite instability as a biomarker in gastric cancer. *J Clin Oncol*. 2019;37(35):3392–3400.
25. deLeeuw RJ, Kost SE, Kakal JA, et al. The prognostic value of FoxP3+ tumor-infiltrating lymphocytes in cancer: a critical review of the literature. *Clin Cancer Res*. 2012;18(11):3022–3029.
26. Simoni Y, Becht E, Fehlings M, et al. Bystander CD8(+) T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature*. 2018; 557(7706):575–579.
27. Saito T, Nishikawa H, Wada H, et al. Two FOXP3(+)CD4(+) T cell subpopulations distinctly control the prognosis of colorectal cancers. *Nat Med*. 2016; 22(6):679–684.
28. Kmieciak M, Gowda M, Graham L, et al. Human T cells express CD25 and Foxp3 upon activation and exhibit effector/memory phenotypes without any regulatory/suppressor function. *J Transl Med*. 2009;7(1):89.
29. Trujillo JA, Sweis RF, Bao R, et al. T cell-inflamed versus non-T cell-inflamed tumors: a conceptual framework for cancer immunotherapy drug development and combination therapy selection. *Cancer Immunol Res*. 2018;6(9): 990–1000.
30. Kelly RJ. Immunotherapy for esophageal and gastric cancer. *Am Soc Clin Oncol Educ Book*. 2017;37:292–300.
31. von Loga K, Woolston A, Punta M, et al. Extreme intratumour heterogeneity and driver evolution in mismatch repair deficient gastro-oesophageal cancer. *Nat Commun*. 2020;11(1):139.
32. Luke JJ, Bao R, Sweis RF, et al. WNT/beta-catenin pathway activation correlates with immune exclusion across human cancers. *Clin Cancer Res*. 2019; 25(10):3074–3083.
33. Kim ST, Cristescu R, Bass AJ, et al. Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer. *Nat Med*. 2018;24(9):1449–1458.