Multiplex analysis of intratumoural immune infiltrate and prognosis in patients with stage II–III colorectal cancer from the SCOT and QUASAR 2 trials: a retrospective analysis

Anja L Frei, Anthony McGuigan, Ritik R A K Sinha, Faiz Jabbar, Luciana Gneo, Tijana Tomasevic, Andrea Harkin, Tim Iveson, Mark P Saunders, Karin A Oien, Noori Maka, Francesco Pezzella, Leticia Campo, Molly Browne, Mark Glaire, Wanja Kildal, Havard E Danielsen, Jennifer Hay, Joanne Edwards, Owen Sansom, Caroline Kelly, Ian Tomlinson, Rachel Kerr, David Kerr, Enric Domingo, TransSCOT consortium, David N Church†, Viktor H Koelzer†*

Lancet Oncol 2024; 25: 198–211

See **Comment** page 151

*Members listed in the appendix (p 3)

†Contributed equally

Department of Pathology and Molecular Pathology, University Hospital Zurich, University of Zurich, Zurich, Switzerland (A L Frei MSc, V H Koelzer MD)**; Life Science Zurich Graduate School, PhD Program in Biomedicine, University of Zurich, Zurich, Switzerland** (A L Frei)**; Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK** (A McGuigan MPhil, R R A K Sinha MSc, F Jabbar MBBS, L Gneo PhD, T Tomasevic MBiomedSci, M Glaire DPhil, D N Church DPhil)**; Nuffield Department of Medicine** (V H Koelzer)**, Nuffield Division of Clinical and Laboratory Sciences** (Prof F Pezzella FRCPath, Prof H E Danielsen PhD*, D Kerr FRCP)**, and Department of Oncology** (L Campo DipHE, M Browne MSc, Prof I Tomlinson PhD, R Kerr FRCP, E Domingo PhD, V H Koelzer)**, University of Oxford, Oxford, UK; Cancer Research UK Glasgow Clinical Trials Unit** (A Harkin BA, C Kelly MSc) **and School of Cancer Sciences** (Prof K A Oien FRCPath, Prof J Edwards PhD, Prof O Sansom PhD)**, University of Glasgow, Glasgow, UK; Southampton University, Southampton, UK** (Prof T Iveson FRCP)**; The Christie NHS Foundation Trust, Manchester, UK** (Prof M P Saunders FRCR)**; Glasgow Tissue Research Facility, University of Glasgow, Queen Elizabeth University**

Summary

Background Tumour-infiltrating CD8⁺ cytotoxic T cells confer favourable prognosis in colorectal cancer. The added prognostic value of other infiltrating immune cells is unclear and so we sought to investigate their prognostic value in two large clinical trial cohorts.

Methods We used multiplex immunofluorescent staining of tissue microarrays to assess the densities of CD8⁺, CD20⁺, FoxP3⁺, and CD68⁺ cells in the intraepithelial and intrastromal compartments from tumour samples of patients with stage II–III colorectal cancer from the SCOT trial (ISRCTN59757862), which examined 3 months versus 6 months of adjuvant oxaliplatin-based chemotherapy, and from the QUASAR 2 trial (ISRCTN45133151), which compared adjuvant capecitabine with or without bevacizumab. Both trials included patients aged 18 years or older with an Eastern Cooperative Oncology Group performance status of 0–1. Immune marker predictors were analysed by multiple regression, and the prognostic and predictive values of markers for colorectal cancer recurrence-free interval by Cox regression were assessed using the SCOT cohort for discovery and QUASAR 2 cohort for validation.

Findings After exclusion of cases without tissue microarrays and with technical failures, and following quality control, we included 2340 cases from the SCOT trial and 1069 from the QUASAR 2 trial in our analysis. Univariable analysis of associations with recurrence-free interval in cases from the SCOT trial showed a strong prognostic value of intraepithelial CD8 ($CD8^{IE}$) as a continuous variable (hazard ratio [HR] for 75th *vs* 25th percentile [75*vs*25] 0·73 [95% CI 0·68–0·79], $p=2·5\times10^{-16}$), and of intrastromal FoxP3 ($FoxP3^{IS}$; 0·71 [0·64–0·78], $p=1·5\times10^{-13}$) but not as strongly in the epithelium ($FoxP3^{IE}$; 0·89 [0·84–0·96], $p=1·5\times10^{-4}$). Associations of other markers with recurrence-free interval were moderate. $CD8^{IE}$ and $FoxP3^{IS}$ retained independent prognostic value in bivariable and multivariable analysis, and, compared with either marker alone, a composite marker including both markers ($CD8^{IE}$-$FoxP3^{IS}$) was superior when assessed as a continuous variable (adjusted [a]HR^{75vs25} 0·70 [95% CI 0·63–0·78], $p=5·1\times10^{-11}$) and when categorised into low, intermediate, and high density groups using previously published cutpoints (aHR for intermediate *vs* high 1·68 [95% CI 1·29–2·20], $p=1·3\times10^{-4}$; low *vs* high 2·58 [1·91–3·49], $p=7·9\times10^{-10}$), with performance similar to the gold-standard Immunoscore. The prognostic value of $CD8^{IE}$-$FoxP3^{IS}$ was confirmed in cases from the QUASAR 2 trial, both as a continuous variable (aHR^{75vs25} 0·84 [95% CI 0·73–0·96], p=0·012) and as a categorical variable for low versus high density (aHR 1·80 [95% CI 1·17–2·75], p=0·0071) but not for intermediate versus high (1·30 [0·89–1·88], p=0·17).

Interpretation Combined evaluation of $CD8^{IE}$ and $FoxP3^{IS}$ could help to refine risk stratification in colorectal cancer. Investigation of $FoxP3^{IS}$ cells as an immunotherapy target in colorectal cancer might be merited.

Funding Medical Research Council, National Institute for Health Research, Cancer Research UK, Swedish Cancer Society, Roche, and Promedica Foundation.

Introduction

Colorectal cancer is the third most common malignancy in the world, with nearly 2 million cases diagnosed each year.[1] Approximately 75% of cases are localised (stage I–III) at diagnosis. Such cases are typically managed by surgical resection, and adjuvant cytotoxic chemotherapy (ie, with a fluoropyrimidine with or without oxaliplatin) is recommended in stage III and high-risk stage II cases.[2] Unfortunately, such chemotherapy benefits only a minority of patients, most of whom have tumours that are either

Research in context

Evidence before this study

We searched PubMed on May 19, 2023, without restrictions on language or date, using the terms "XXX AND (colorectal OR colon) AND cancer AND (prognosis OR recurrence OR outcome)", where XXX was each of "CD8", "FoxP3", "CD20" and "CD68", which returned 1149 results (15 meta-analyses), 264 results (eight meta-analyses), 102 results (no meta-analyses), and 164 results (one meta-analysis), respectively. Relevant meta-analyses and primary studies were reviewed. Most were of modest size (<500 patients). None used clinical trial samples. None analysed markers as continuous variables, instead using heterogeneous cutpoints including present versus absent, median, and "optimum" levels. Most did not separate intraepithelial and intrastromal infiltrate. Covariables such as DNA mismatch repair (MMR) status were variably included, and no study included tumour sidedness or stromal proportion. No study examined oxaliplatin-treated patients. CD8 and FoxP3 individually conferred better prognosis; however, a single study that examined both in a single regression model lacked covariables and spatial resolution as detailed above.

Added value of this study

By analysis of two large clinical trials, this study demonstrates that quantification of intrastromal FoxP3 improves upon the known prognostic value of intraepithelial CD8 in stage II–III colorectal cancer, with performance as good as or better than the consensus Immunoscore. Furthermore, it demonstrates that these markers share non-overlapping predictors, with deficient MMR predictive for CD8 infiltrate, and high tumour stroma predictive of FoxP3.

Implications of all the available evidence

Both CD8 and FoxP3 positive immune cells confer favourable prognosis in colorectal cancer. Combined analysis of both markers holds promise to improve risk stratification in early-stage disease. Understanding the mechanisms of tumour suppression by FoxP3-positive cells in colorectal cancer might reveal novel targets for immunotherapy.

Hospital, Glasgow, UK (Prof K A Oien, N Maka FRCPath, J Hay PhD); **Institute for Cancer Genetics and Informatics, Oslo University Hospital, Oslo, Norway** (W Kildal PhD, Prof H E Danielsen); **Cancer Research UK Beatson Institute of Cancer Research, Glasgow, UK** (Prof O Sansom); **Cancer Research UK Scotland Centre, Glasgow and Edinburgh, UK** (Prof O Sansom, E Domingo); **Oxford NIHR Comprehensive Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, Oxford, UK** (D N Church)

*Prof Danielsen died in October, 2023

Correspondence to: Dr David N Church, Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7BN, UK
david.church@well.ox.ac.uk

See **Online** for appendix

cured by surgery alone or have micrometastases resistant to these drugs. Thus, attention has been focused on the identification of biomarkers capable of improving risk stratification and of new therapeutic targets for clinical investigation. Successful examples include genomic alterations, such as deficient DNA mismatch repair (MMR) or microsatellite instability (MSI), which portend good prognosis[3] and sensitivity to immune checkpoint blockade,[4,5] and *KRAS* and *BRAF* mutations, which portend worse outcomes[6] and are the targets of specific inhibitors in the context of metastatic disease.[7,8] Other examples relate to the antitumour immune response, predominantly the density of CD8$^+$ cytotoxic tumour-infiltrating lymphocytes, which correlates with improved outcomes in colorectal cancer.[9–11] Whether quantification of additional immune cell types provides additional prognostic value is less clear. CD68$^+$ macrophage infiltration has been reported to correlate with both better and worse outcomes[12,13] and to depend on functional polarisation,[14] whereas CD20$^+$ B cells have been little studied. Another interesting cell type is FoxP3$^+$ regulatory T cells, which, despite their typically immunosuppressive function, have been associated with improved prognosis in colorectal cancer in several studies.[15,16] However, the extent to which the association of FoxP3$^+$ regulatory T cells with prognosis is independent of CD8$^+$ and other immune cell types is unknown. In this study we sought to investigate the prognostic values of these markers by multiplex immunostaining of colorectal cancers from two large clinical trials.

Methods

Study design and cohorts

In this retrospective analysis, we used data and tissue samples from the SCOT (ISRCTN59757862) and QUASAR 2 (ISRCTN45133151) trials, details of which have been reported previously[17,18] and are provided in the appendix (p 4). To assess the prognostic value of immune markers, we used a discovery-validation approach, with the SCOT trial serving as the discovery cohort, and the QUASAR 2 trial the validation cohort.

In the SCOT trial, patients aged 18 years or older with an Eastern Cooperative Oncology Group (ECOG) performance status of 0–1 were recruited between March 27, 2008, and Nov 29, 2013, and randomly assigned to receive either 3 months or 6 months of adjuvant oxaliplatin-based chemotherapy following resection of stage III or high-risk stage II colorectal cancer (colon or rectum). Analysis of the primary endpoint of disease-free survival confirmed that the shortened duration (ie, 3 months rather than 6 months of adjuvant chemotherapy) was adequate for most patients.[17]

In the QUASAR 2 trial, patients aged 18 years and older with an ECOG performance status of 0–1 were recruited between April 25, 2005, and Oct 12, 2010, and randomly assigned to receive capecitabine or capecitabine plus bevacizumab after resection of stage III or high-risk stage II colorectal cancer. Analysis of the primary endpoint of disease-free survival showed no benefit of bevacizumab.[18]

3076 of 6088 patients in the SCOT trial and 1195 of 1952 patients in the QUASAR 2 trial donated samples for research; characteristics of these subsets of patients were similar to those of the parent trials.[17,18] For the present analysis, donated samples were included provided that they had adequate tumour content. Tissue microarrays were constructed from 0·6 mm punched cores from formalin-fixed paraffin-embedded blocks from tumour samples from 2350 participants from SCOT and 1·0 mm

cores from tumour samples from 1195 participants from QUASAR 2, under guidance of the study pathologists (KAO, NM, and FP), and were stored at room temperature protected from air. 1786 patients from SCOT had two cores taken from the centre of the tumour and two from the invasive margin, and a further 564 had four cores from each region. For QUASAR 2 samples, all had three cores taken, without selection for location. Additional details of patients and tumour samples are provided in the appendix (p 4).

Ethical approval and patient consent for recruitment and sample collection in the SCOT and QUASAR 2 trial were obtained centrally and at all recruiting centres and for all participants (reference numbers 07/S0703/136 and 04/MRE/11/18, respectively). Ethical approval for anonymised tumour molecular analysis was granted by Oxfordshire Research Ethics Committee B (05\Q1605\66).

Procedures

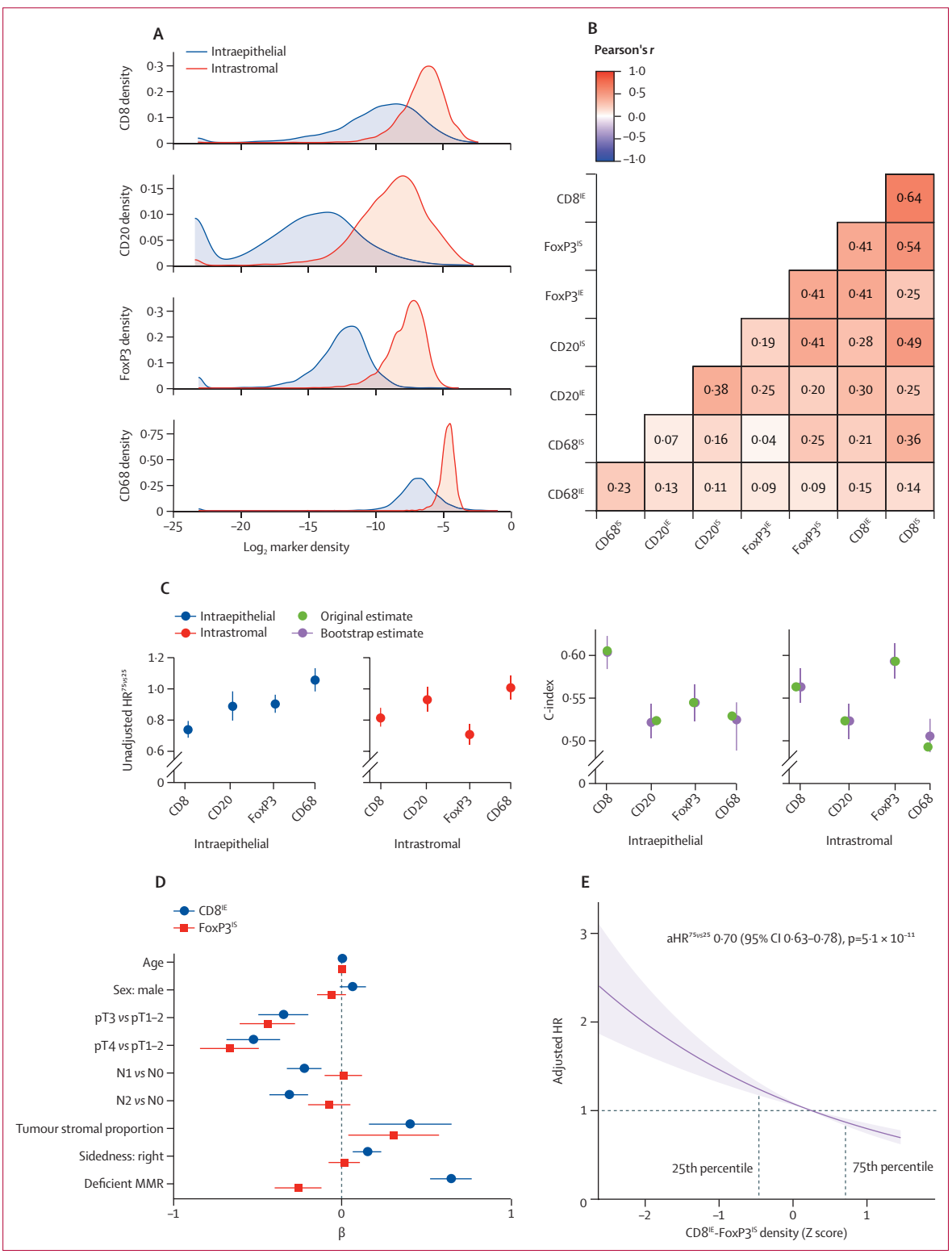
Multiplex immunofluorescence staining for CD8, CD4, CD20, FoxP3, CD68, pan-cytokeratin, and DAPI (appendix pp 4–6, 26) was done on tissue microarray sections using the OPAL protocol (AKOYA Biosciences, Marlborough, MA, USA) on the BOND RXm autostainer (Leica Microsystems, Wetzlar, Germany) in the Oxford Cancer Centre Good Clinical Practice-approved laboratory (LC, MB), in accordance with the Society for Immunotherapy of Cancer statement on best practices for immunofluorescence staining and validation.[19] Full details of antibodies, reagents, and methods have been previously published[20] and are provided in the appendix (p 5) and a related publication.[21] Multispectral images from stained slides were obtained on the AKOYA Biosciences Vectra Polaris, with batch analysis done with the inForm 2.4.8 software provided. Batched, analysed multispectral images were fused for analysis. All stains were visually reviewed and verified by a board-certified gastrointestinal pathologist (VHK). Autofluorescence in the 520 nm channel precluded analysis of CD4, and this marker, as well as a subset of slides with increased artefactual bleed-through between pan-cytokeratin (Opal 650) and CD68 (Opal 690), were excluded from further study. Details of CD8 and CD3 immunostaining of the QUASAR 2 samples have been published previously.[10] Following detailed quality control, immune markers were quantified and localised to the epithelial and stromal compartments by a machine-learning-based tissue classifier (HALO AI DenseNet v2) trained for this purpose, which also quantified tumour stromal proportion (see appendix pp 6–7 for further details). Marker density was quantified as the marker-positive area as a proportion of the total area analysed, expressed in μm² across informative tissue microarray cores. This area-based metric was highly concordant with an alternative metric based on individual cell positivity for CD8 and FoxP3 in intraepithelial and intrastromal compartments (r_s 0·55–0·92), and with CD8 measured by immunohistochemistry in a previous study.[10] Further details are available in a related technical report.[21] Methods for automated detection of CD8 and CD3 cells in samples from QUASAR 2 have been previously reported.[10] The present analysis was done with blinding to clinical information for all cases. Demographic and clinicopathological factors were taken from trial databases. Methods for molecular analysis in QUASAR 2 were reported previously.[6,10] Identification of deficient MMR in cases from the SCOT trial was done with artificial intelligence (AI)-based image analysis and expert pathologist review.[22]

Outcomes

The study primary endpoint was the association between immune marker density and spatial distribution and colorectal cancer recurrence-free interval, defined as the time from randomisation to relapse, with censoring at last contact or death in case of no recurrence. The secondary endpoint was association with colorectal cancer-specific survival, defined as time from randomisation to death from colorectal cancer, with censoring at last contact or death from cause other than colorectal cancer. Both endpoints were to be assessed in the discovery and validation cohorts. Exploratory analyses included the association of the composite immune marker across colorectal cancer subgroups in the pooled study population, comparison with Immunoscore[9] using endpoints of recurrence-free interval and disease-free survival in individual cohorts and the pooled study population, and post-hoc analyses of individual markers in each cohort (appendix p 10). Biomarker analyses done in this study are detailed in the appendix (p 10) along with their objectives, endpoints, population, and methods, and details of where the results of each analysis are reported.

Statistical analysis

Full details of statistical methods are provided in the appendix (pp 7–8) together with the TRIPOD and REMARK checklists for prognostic studies (appendix pp 24–25). Analyses were done and reported in line with REMARK guidelines.[23] Analyses used all available samples from participants identified retrospectively on the basis of available tissue and data, without selection for time of study recruitment or length of follow-up. Comparisons of baseline characteristics of participants in each study between excluded and included cases were done using the Wilcoxon rank sum test or Pearson's χ^2 test. Following confirmation of strong positive skewness, immune marker densities were $\log_2$ transformed and analysed as continuous or categorical variables, with the continuous variables scaled to permit between-marker comparisons. Differences in immune marker densities between tumour regions was assessed by the Kolmogorov-Smirnov (KS) test. Correlations were analysed with use of Pearson's (r) or Spearman's (r_s)



(Figure 1 continues on next page)

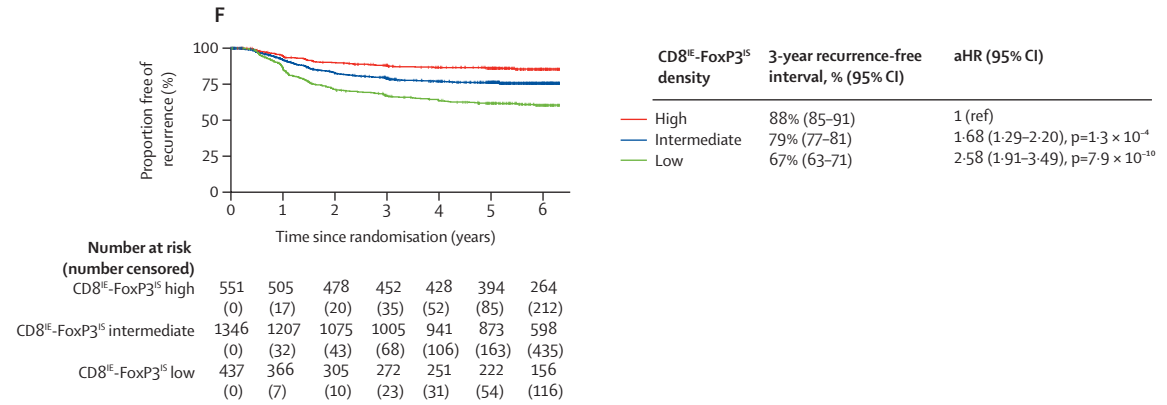


Figure 1: **Multiplex immunoprofiling of 2340 cases from the SCOT trial**
(A) Kernel density plots showing intratumoural density of CD8+ (cytotoxic T cells), CD20+ (B cells), FoxP3+ (regulatory T cells), and CD68+ (macrophages) in the intraepithelial and intrastromal compartments in SCOT trial cases. (B) Correlations (Pearson's *r*) between immune markers according to intraepithelial (IE) or intrastromal (IS) localisation. (C) Univariable prognostic value and model discrimination for immune marker densities according to intraepithelial or intrastromal localisation. Left panel shows unadjusted HRs for recurrence-free interval between the 75th and 25th percentiles ($HR^{75 vs 25}$). Right panel shows corresponding model discrimination (C index) on the original cohort of 2340 cases and on an internal validation cohort obtained by 1000 bootstrap resamples. Error bars are 95% CIs. Model estimates and metrics are shown in the appendix (p 17). (D) Independent predictors or correlates of $CD8^{IE}$ and $FoxP3^{IS}$ densities as analysed by multiple linear regression. Model estimates are shown in the appendix (p 17). β is the change in marker density (z score). Error bars are 95% CIs. (E) Risk of colorectal cancer recurrence according to composite $CD8^{IE}$-$FoxP3^{IS}$ marker density, assessed by multivariable analysis (including age, sex, primary T stage, N stage, tumour sidedness, deficient MMR, tumour stromal proportion, chemotherapy regimen, and chemotherapy duration). The shaded area represents 95% CI. (F) Kaplan-Meier curves showing recurrence-free interval of colorectal cancer by $CD8^{IE}$-$FoxP3^{IS}$ marker density as categorical variables (high, intermediate, or low), with multivariable HR adjusted for the same variables as in panel E. aHR=adjusted hazard ratio. $CD8^{IE}$=intraepithelial CD8. $FoxP3^{IS}$=intrastromal FoxP3. HR=hazard ratio. MMR=DNA mismatch repair.

coefficients. Predictors of immune infiltrate were analysed with multiple linear regression.

Time-to-event analyses were done using Kaplan-Meier survival curves, compared by the log-rank test, adjusted for multiple testing in case of multiple comparisons, and univariable and multivariable Cox models, with multivariable models adjusted for prespecified baseline and prognostic factors in each trial cohort (ie age, sex, primary tumour stage, nodal stage, tumour stromal proportion, sidedness, MMR status, and chemotherapy regimen and chemotherapy duration in SCOT trial; and age, sex, primary tumour stage, nodal stage, lymphovascular space invasion, tumour stromal proportion, MSI status, *KRAS* mutation, *BRAF* mutation, and chemotherapy regimen in the QUASAR 2 trial). Multivariable models incorporated baseline and known prognostic factors as forced-entry variables and used data from participants with complete data with no imputation done; cases with missing data were excluded. Censoring in time-to-event analyses was non-informative. Proportionality of hazards was confirmed by analysis of scaled Schoenfeld residuals. For continuous variables, hazard ratios (HRs) were obtained for comparison of cases at the 75th percentile to those at the 25th percentile (exploratory analyses of Cox models using restricted cubic splines demonstrated no significant non-linearity). The prognostic value of the composite marker was evaluated as a continuous variable, with HRs calculated for cases of the 75th versus 25th percentile of marker distribution, and as a three-group categorical variable using a method similar to that employed by the consensus Immunoscore (ie, split at the 25th and 70th percentiles into low, intermediate, and high density groups;[9,11] appendix p 27).

The sample size for this study was not predetermined. All statistical tests were two-sided, and the threshold for statistical significance was set at α=0·05, with the exception of interaction testing, for which α=0·1 was used for consistency with relevant recent literature.[11] Analyses were done in R (version 4.2.2) using R Studio (version 2022.07.1, build 554). The R packages used are listed in the appendix (pp 7–8).

Role of the funding source

The funders of this study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

After exclusion of cases without tissue microarrays and with technical failures, multiplex immunofluorescence staining was done successfully for the immune cell markers CD8 (cytotoxic T cells), FoxP3 (regulatory T cells), CD68 (macrophages), CD20 (B cells), and pan-cytokeratin (cancer cells) on stage II–III colon and rectal cancers from the SCOT[17] and QUASAR 2[18] trials (trial profile on appendix p 28). CD4 immunofluorescent staining was unsuccessful because of tissue autofluorescence. Following quality control,[25] 2340 cases from the SCOT trial and 1069 from the QUASAR 2 trial were informative for and thus included in immune marker analysis; characteristics of these informative

cases were similar to those not informative for biomarker analysis, with the exception of modest but significant differences in disease stage for both studies and treatment regimen for cases from the SCOT trial (appendix pp 11–14). Cases in the final SCOT and QUASAR 2 cohorts were similar in age (median 65 years [IQR 20–84] *vs* 65 years [21–85]) and sex (1412 [60%] of 2340 *vs* 616 [58%] of 1069 male), although the SCOT trial had a greater proportion of patients with stage III disease (1868 [80%] *vs* 697 [65%]; appendix pp 11–12). We developed a machine-learning method to define intraepithelial and intrastromal marker densities,[17] and used this to define infiltrate in the tumour centre and invasive margin in SCOT cases, and in the tumour centre in QUASAR 2 cases. Individual marker densities were similar in the tumour centre and invasive margin in SCOT cases (appendix p 29), with the strongest correlation between regions observed for intraepithelial CD8 ($CD8^{IE}$; r_s=0·50), intrastromal CD8 ($CD8^{IS}$; r_s=0·44), and intrastromal FoxP3 ($FoxP3^{IS}$; r_s=0·41; appendix p 30). Intrastromal immune cell infiltrate was greater than intraepithelial infiltrate in all cases in the tumour centre and invasive margin, when considered separately (appendix p 31) and when combined ($p<2·2\times10^{-16}$, KS test; figure 1A; appendix p 32). For the SCOT trial, the correlations between the densities of various markers in intraepithelial and intrastromal compartments across both the tumour centre and invasive margin varied substantially, from weak ($FoxP3^{IE}$ *vs* $CD68^{IS}$, r=0·04) to strongly positive ($CD8^{IE}$ *vs* $CD8^{IS}$, r=0·64; figure 1B), with similar results when the tumour centre and invasive margin were analysed separately (appendix p 34). Similar

	Univariable analysis			Multivariable analysis		
	Cases (events)	Hazard ratio (95% CI)	p value	Cases (events)	Adjusted hazard ratio (95% CI)	p value
Age, years (continuous, per year)	2340 (559)	1·00 (0·99–1·01)	0·54	1951 (466)	0·99 (0·98–1·00)	0·31
Sex						
Female	928 (220)	1 (ref)	..	790 (190)	1 (ref)	..
Male	1412(339)	1·00 (0·85–1·19)	0·99	1161 (276)	1·0 (0·83–1·20)	0·98
Primary tumour stage						
pT1–2	175 (15)	1 (ref)	..	145(11)	1 (ref)	..
pT3	1383(274)	2·34 (1·39–3·9)	$1·4\times10^{-4}$	1173 (240)	2·58 (1·40–4·74)	0·0022
pT4	782 (270)	4·57 (2·71–7·68)	$1·1\times10^{-8}$	633 (215)	4·44 (2·39–8·24)	$2·3\times10^{-6}$
Nodal stage						
N0	472 (70)	1 (ref)	..	389 (61)	1 (ref)	..
N1	1282 (263)	1·40 (1·08–1·83)	0·012	1064 (221)	1·60 (1·19–2·13)	0·0017
N2	586 (226)	2·99 (2·29–3·91)	$1·4\times10^{-15}$	498 (184)	2·70 (2·01–3·63)	$4·0\times10^{-11}$
Tumour stroma proportion, 75th *vs* 25th percentile	2340 (559)	1·18 (1·06–1·33)	0·0027	1951 (466)	1·20 (1·05–1·37)	0·0059
Sidedness						
Left	1369 (291)	1 (ref)	..	1173 (249)	1 (ref)	..
Right	938 (259)	1·38 (1·17–1·63)	$1·7\times10^{-4}$	778 (217)	1·34 (1·10–1·62)	0·0036
DNA mismatch repair status						
Proficient	1751 (434)	1 (ref)	..	1722 (426)	1 (ref)	..
Deficient	229 (40)	0·70 (0·51–0·97)	0·034	229 (40)	0·67 (0·47–0·94)	0·020
Chemotherapy regimen						
CAPOX	1652 (389)	1 (ref)	..	1369 (317)	1 (ref)	..
FOLFOX	688 (170)	1·04 (0·87–1·24)	0·70	582 (149)	1·04 (0·86–1·27)	0·69
Chemotherapy duration						
24 weeks	1160 (268)	1 (ref)	..	969 (222)	1 (ref)	..
12 weeks	1180 (291)	1·09 (0·92–1·29)	0·32	982 (244)	1·11 (0·92–1·33)	0·27
$CD8^{IE}$-$FoxP3^{IS}$ density (continuous), 75th *vs* 25th percentile (HR^{75vs25})	2334 (558)	0·68 (0·63–0·74)	$<2·2\times10^{-16}$	1951 (466)	0·70 (0·63–0·78)	$5·1\times10^{-11}$
$CD8^{IE}$-$FoxP3^{IS}$ density (categorical)						
High	551 (76)	1 (ref)	..-	480 (68)	1 (ref)	..
Intermediate	1346 (317)	1·79 (1·39–2·30)	$5·3\times10^{-6}$	1152 (279)	1·68 (1·29–2·20)	$1·3\times10^{-4}$
Low	437 (165)	3·14 (2·39–4·12)	$<2\times10^{-16}$	319 (119)	2·58 (1·91–3·49)	$7·9\times10^{-10}$

Multivariable results for covariables are from a model including $CD8^{IE}$-$FoxP3^{IS}$ as a continuous variable. Performance and comparison of models is provided in the appendix (p 17). $CD8^{IE}$=intraepithelial CD8. $FoxP3^{IS}$=intrastromal FoxP3. CAPOX=capecitabine and oxaliplatin. FOLFOX=leucovorin, fluorouracil, and oxaliplatin. HR=hazard ratio.

Table 1: **$CD8^{IE}$-$FoxP3^{IS}$ density and colorectal cancer recurrence-free interval in cases from the SCOT trial (n=2340)**

findings were seen for cases in the QUASAR 2 study (appendix p 33).

We examined the association between intratumoural immune cell infiltrates (as continuous variables) and colorectal cancer recurrence-free interval in the 2340 SCOT cases after scaling (Z score transformation) to enable between-marker comparisons. Preliminary univariable analyses of marker densities in the tumour centre alone and the invasive margin alone revealed similar or lower prognostic value and model discrimination than when marker densities were analysed across both regions (using the mean density; appendix p 15). We therefore used the combination of both regions for all subsequent analyses (except where stated otherwise). As anticipated, CD8 was highly prognostic, with the strongest association observed for intraepithelial ($CD8^{IE}$) localisation (HR for cases at 75th *vs* 25th percentile [HR^{75vs25}] 0·73 [95% CI 0·68–0·79], $p=2·5\times10^{-16}$; figure 1C; appendix p 15). Notably, tumour-infiltrating $FoxP3^{+}$ cells held similar prognostic value to $CD8^{IE}$ when localised intrastromally ($FoxP3^{IS}$; HR^{75vs25} 0·71 [0·64–0·78], $p=1·5\times10^{-13}$), but not when localised within the epithelium (0·89 [0·84–0·96], $p=1·5\times10^{-4}$). Other markers showed weaker or no association with recurrence, which was reflected in significantly lower model discrimination (C index; figure 1C; appendix p 15). $CD8^{IE}$ and $FoxP3^{IS}$ densities had some discordant independent associations with other tumour characteristics (figure 1D; appendix p 16). $CD8^{IE}$ density was lower with increasing primary T stage and N stage, and higher with increasing tumour stromal proportion, deficient MMR, in right-sided tumours, and independent of MMR status. $FoxP3^{IS}$ density was also lower with increasing T stage and higher in stroma-rich tumours; however, it was not associated with N stage or sidedness, and, in contrast to $CD8^{IE}$, was significantly lower in MMR-deficient tumours.

After demonstrating a lack of problematic collinearity by calculation of variance inflation factor (1·65), we confirmed that $CD8^{IE}$ (HR^{75vs25} 0·79 [95% CI 0·73–0·86], $p=8·7\times10^{-8}$) and $FoxP3^{IS}$ (0·80 [0·72–0·88], $p=1·5\times10^{-5}$) held independent prognostic value for recurrence-free interval in a bivariable model and following adjustment for confounders in multivariable analysis (appendix p 17). A composite variable incorporating both markers, $CD8^{IE}$-$FoxP3^{IS}$ (calculated as the mean of marker Z scores) had greater discrimination than either marker alone in univariable and multivariable models as a continuous variable (adjusted [a]HR^{75vs25} 0·70 [95% CI 0·63–0·78], $p=5·1\times10^{-11}$; table 1; figure 1E; appendix p 17). Similarly, a three-group $CD8^{IE}$-$FoxP3^{IS}$ categorical variable based on marker densities in the combined tumour centre and invasive margin using cutpoints similar to the consensus Immunoscore[9] had greater discrimination than one based on densities in the tumour centre and invasive margin separately (appendix p 35), and was strongly prognostic (aHR for intermediate *vs* high 1·68 [95% CI 1·29–2·20], $p=1·3\times10^{-4}$; low *vs* high 2·58 [1·91–3·49], $p=7·9\times10^{-10}$) independently of T and N stage, tumour sidedness, deficient MMR, and tumour stromal proportion (table 1; figure 1F; appendix p 17). Lower $CD8^{IE}$-$FoxP3^{IS}$ density also correlated with reduced cancer-specific survival as both a continuous variable (aHR^{75vs25} 0·70 [95% CI 0·61–0·80], $p=3·7\times10^{-7}$) and as a categorical variable (intermediate *vs* high aHR 1·64 [1·14–2·37], p=0·0077; low *vs* high 2·90 [1·94–4·33], $p=1·8\times10^{-7}$; appendix pp 18–19, 36).

We sought to validate our results in the 1069 eligible cases in the QUASAR 2 trial. Multiple regression confirmed the associations between pT4 stage tumours and reduced $CD8^{IE}$ and $FoxP3^{IS}$ density; between MSI and increased $CD8^{IE}$ (but not $FoxP3^{IS}$ density); and tumour stromal proportion and increased $FoxP3^{IS}$ density, although it did not confirm the association of tumour stromal proportion with $CD8^{IE}$ density as observed in SCOT cases; figure 2A; appendix p 20). Notably, in a model that excluded *KRAS* and *BRAF* mutation (to allow comparison with SCOT, for which these variables were not available), $CD8^{IE}$ was significantly associated with right-sided tumour location (appendix p 37), although not when these factors were included (figure 2A). Neither $CD8^{IE}$ density nor $FoxP3^{is}$ density were associated with *KRAS* or *BRAF* mutation (figure 2A; appendix p 37).

As in SCOT, $CD8^{IE}$ and $FoxP3^{IS}$ were each strongly associated with recurrence-free interval on univariable analysis (appendix p 21). Similarly, the composite $CD8^{IE}$-$FoxP3^{IS}$ variable showed greater discrimination than either marker alone, although neither marker was prognostic in a model including both as independent variables, possibly owing to collinearity or fewer events in this cohort (appendix p 21). $CD8^{IE}$-$FoxP3^{IS}$ was also independently associated with recurrence-free interval in multivariable analyses both as a continuous variable (aHR^{75vs25} 0·84 [0·73–0·96], p=0·012) and as a categorical variable for low versus high density (aHR 1·80 [95% CI 1·17–2·75], p=0·0071), but not for intermediate versus high (1·30 [0·89–1·88], p=0·17; table 2; figure 2B, C; appendix p 21). As in SCOT, $CD8^{IE}$-$FoxP3^{IS}$ was strongly associated with cancer-specific survival in the QUASAR 2 cohort, both as a continuous and a categorical variable for low versus high density, but not for intermediate versus high density tumours (appendix pp 22–23, 38).

We compared the prognostic value of $CD8^{IE}$-$FoxP3^{IS}$ with that of alternative immunoprofiling methods in similar cohorts (appendix p 8). Differences in recurrence-free interval between low and high groups in 1864 stage III cases from the SCOT trial (high *vs* low HR 0·32 [95% CI 0·24–0·42]) were larger than those reported for three-group Immunoscore (IS3) in a study of 763 stage III non-trial tumours[24] (high *vs* low 0·48 [95% CI 0·32–0·71]; appendix p 39), while differences in disease-free survival between these groups in the current study (low *vs* high aHR 2·15 [1·62–2·85]) were very

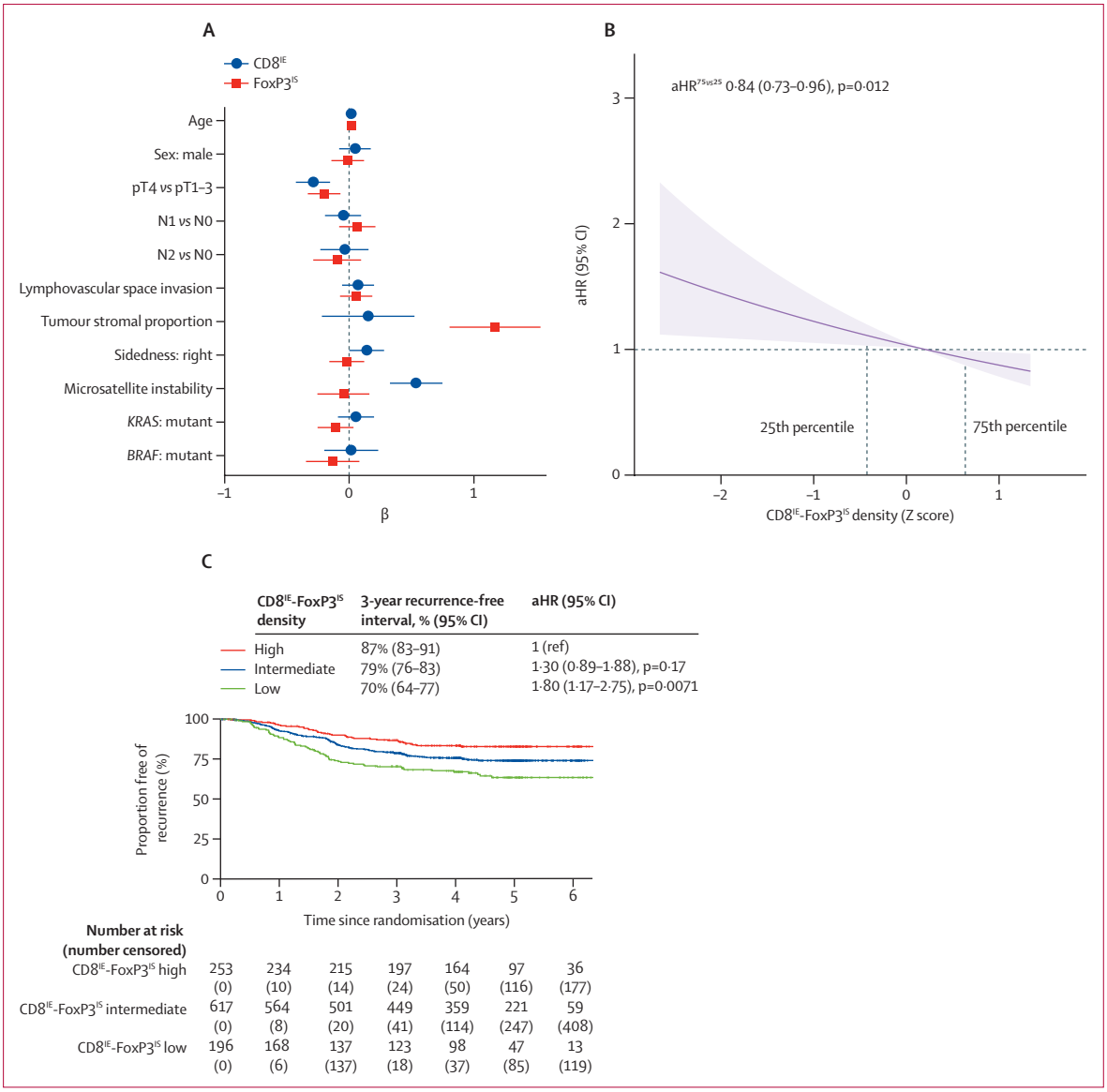


Figure 2: **Validation of predictors and prognostic value of CD8^{IE} and FoxP3^{IS} in cases from the QUASAR 2 trial**
(A) Independent predictors or correlates of CD8^{IE} and FoxP3^{IS} densities as analysed by multiple linear regression. Model estimates are shown in the appendix (p 23). β is the change in marker density (z score). Error bars are 95% CIs. (B) Risk of colorectal cancer recurrence by composite CD8^{IE}-FoxP3^{IS} marker density, assessed by multivariable analysis (including age, sex, pT stage, N stage, sidedness, lymphovascular invasion, tumour stromal proportion, microsatellite instability, *KRAS* mutation, *BRAF* mutation, and treatment regimen). Adjusted HR is shown for recurrence-free interval between the 75th and 25th percentiles (HR^{75vs25}). The shaded areas represent 95% CIs. (C) Kaplan-Meier curves showing recurrence-free interval for colorectal cancer by CD8^{IE}-FoxP3^{IS} marker density as a categorical variable (high, intermediate, or low), with multivariable HRs adjusted for the same variables as in panel B. aHR=adjusted hazard ratio. CD8^{IE}=intraepithelial CD8. FoxP3^{IS}=intrastromal FoxP3. HR=hazard ratio.

similar to those between IS3 groups in 1062 cases from the IDEA-France study[11] (2·22 [1·31–3·77]; appendix p 39). Similarly, comparison of recurrence-free interval between CD8^{IE}-FoxP3^{IS} groups in 3400 mainly stage III (75·3%) tumours from the SCOT and QUASAR 2 trials (high *vs* low aHR 0·41 [0·33–0·51]) revealed very similar HRs to those for IS3 in 2648 mainly stage I–II (71·2%) tumours in the landmark Immunoscore publication[9] (0·40 [0·30–0·54]; appendix p 39). We used data from a previous study[10] to directly compare CD8^{IE} and FoxP3^{IS} to the Immunoscore markers CD8 and CD3 in 868 QUASAR 2 cases, first confirming the expected strong concordance of CD8 (by immunohistochemistry) and CD8^{IE} densities (r_s=0·58, p<2×10^{-16}; appendix p 40). Categorisation of cases by CD8 and CD3 density using IS3 cutpoints confirmed worse recurrence-free interval (HR 1·65 [95% CI 1·12–2·44], p=0·011) and cancer-specific survival (2·09 [1·23–3·57], p=0·0066) for CD8-CD3 low versus high tumours (appendix p 41). However, the categorical CD8^{IE}-FoxP3^{IS} variable was more strongly

	Univariable analysis			Multivariable analysis		
	Cases (events)	Hazard ratio (95% CI)	p value	Cases (events)	Adjusted hazard ratio (95% CI)	p value
Age, years (continuous, per year)	1069 (256)	1·01 (1·00–1·02)	0·22	874 (209)	1·00 (0·98–1·01)	0·71
Sex						
Female	453 (102)	1 (ref)	..	367 (84)	1 (ref)	..
Male	616 (154)	1·08 (0·84–1·39)	0·53	507 (125)	1·06 (0·80–1·41)	0·68
Primary tumour stage						
pT1–3	677 (126)	1 (ref)	..	546 (98)	1 (ref)	..
pT4	392 (130)	2·06 (1·61–2·64)	$1·1\times10^{-8}$	328 (111)	2·25 (1·70–2·98)	$1·6\times10^{-8}$
Nodal stage						
N0	372 (55)	1 (ref)	..	315 (47)	1 (ref)	..
N1	531 (115)	1·52 (1·10–2·10)	0·012	421 (89)	1·84 (1·28–2·66)	0·0010
N2	166 (86)	4·84 (3·45–6·80)	$<2·0\times10^{-16}$	138 (73)	5·05 (3·46–7·36)	$<2·0\times10^{-16}$
Lymphovascular space invasion						
Absent	551 (114)	1 (ref)	..	471 (97)	1 (ref)	..
Present	465 (132)	1·44 (1·12–1·85)	0·0043	403 (112)	1·32 (1·00–1·74)	0·054
Tumour stroma proportion (continuous), 75th *vs* 25th percentile	1069 (256)	1·05 (0·90–1·24)	0·52	874 (209)	1·10 (0·91–1·33)	0·30
Sidedness						
Left	584 (139)	1 (ref)	..	503 (121)	1 (ref)	..
Right	444 (106)	1·04 (0·80–1·34)	0·76	371 (88)	0·95 (0·70–1·28)	0·74
Microsatellite instability status						
Microsatellite stable	886 (222)	1 (ref)	..	758 (190)	1 (ref)	..
Microsatellite unstable	136 (25)	0·74 (0·49–1·12)	0·15	116 (19)	0·60 (0·35–1·01)	0·055
KRAS status*						
Wild type	664 (153)	1 (ref)	..	595 (137)	1 (ref)	..
Mutant	313 (80)	1·12 (0·86–1·47)	0·41	279 (72)	1·27 (0·93–1·74)	0·13
BRAF status†						
Wild type	856 (195)	1 (ref)	..	755 (173)	1 (ref)	..
Mutant	130 (38)	1·40 (0·99–1·99)	0·056	119 (36)	1·92 (1·25–2·96)	0·0031
Chemotherapy regimen						
Capecitabine	526 (112)	1 (ref)	..	420 (85)	1 (ref)	..
Capecitabine and bevacizumab	543 (144)	1·28 (1·00–1·64)	0·052	454 (124)	1·29 (0·98–1·71)	0·070
$CD8^{IE}$-$FoxP3^{IS}$ (continuous), 75th *vs* 25th percentile (HR^{75vs25})	1066 (255)	0·79 (0·71–0·89)	$1·5\times10^{-4}$	874 (209)	0·84 (0·73–0·96)	0·012
$CD8^{IE}$-$FoxP3^{IS}$ (categorical)						
High	253 (40)	1 (ref)	..	213 (37)	1 (ref)	..
Intermediate	617 (150)	1·58 (1·18–2·25)	0·0097	501 (118)	1·30 (0·89–1·88)	0·17
Low	196 (65)	2·41 (1·63–3·58)	$1·2\times10^{-5}$	160 (54)	1·80 (1·17–2·75)	0·0071

Multivariable results for covariables are from a model including $CD8^{IE}$-$FoxP3^{IS}$ as continuous variable. Performance and comparison of models is provided in the appendix (p 21). $CD8^{IE}$=intraepithelial CD8. $FoxP3^{IS}$=intrastromal FoxP3. *Mutations at codons 12, 13, 61, and 146. †*BRAF* V600E mutation.

Table 2: **$CD8^{IE}$-$FoxP3^{IS}$ density and colorectal cancer recurrence-free interval in QUASAR 2 trial cases**

prognostic than the CD8-CD3 variable for both endpoints, evidenced by larger differences between groups (HR 2·00 [95% CI 1·31–3·06], p=0·0014 for recurrence-free interval; 3·05 [1·77–5·27], $p=6·3\times10^{-5}$ for cancer-specific survival) and greater model discrimination (appendix p 41). The variation in the prognostic value of marker combinations was mirrored by differences in their correlation, which was substantially greater between CD8 and CD3 (r_s=0·67) than between $CD8^{IE}$ and $FoxP3^{IS}$ (r_s=0·41; appendix p 40).

We further examined the prognostic and predictive utility of $CD8^{IE}$-$FoxP3^{IS}$ in the individual trials and within tumour subgroups in pooled analysis of all 3400 patients with informative data. The prognostic value of $CD8^{IE}$-$FoxP3^{IS}$ was similar across patient groups by age, sex, and T and N stage, but stronger in left-sided than right-sided tumours (aHR^{75vs25} 0·66 [95% CI 0·60–0·73] *vs* 0·84 [0·75–0·96], adjusted $p_{interaction}$=0·010), stroma-high *vs* stroma-low tumours (0·68 [0·61–0·77] *vs* 0·82 [0·74–0·92]; adjusted $p_{interaction}$=0·018), and in

MMR-proficient versus MMR-deficient tumours (0·72 [0·66–0·78] *vs* 0·97 [0·73–1·29], adjusted $p_{interaction}$=0·079; appendix p 42; see figure 3 for unadjusted estimates), with similar associations for $CD8^{IE}$-$FoxP3^{IS}$ as a categorical variable (appendix p 43). Notably, MMR-proficient $CD8^{IE}$-$FoxP3^{IS}$-high tumours had a similar recurrence rate to MMR-deficient $CD8^{IE}$-$FoxP3^{IS}$-high and $CD8^{IE}$-$FoxP3^{IS}$-intermediate tumours (appendix p 42). $CD8^{IE}$-$FoxP3^{IS}$ was similarly prognostic in unadjusted (figure 3) and multivariable adjusted analysis of stage III tumours of clinical low risk (stage T1–3, N1; aHR^{75vs25} 0·74 [95% CI 0·64–0·86]) and high risk (stage T4 or N2 [or both]; 0·73 [0·66–0·82]; appendix p 44), although absolute differences between high, intermediate, and low $CD8^{IE}$-$FoxP3^{IS}$ density were larger in high-risk cases (appendix p 44). Clinically low-risk $CD8^{IE}$-$FoxP3^{IS}$-low cases had a similar probability of recurrence at 3 years to high-risk $CD8^{IE}$-$FoxP3^{IS}$-high cases (79% [95% CI 74–85] *vs* 78% [73–84]; Benjamini-Hochberg-adjusted log-rank test p=0·53; appendix p 44). $CD8^{IE}$-$FoxP3^{IS}$ did not predict a differential benefit of chemotherapy duration in the SCOT trial, or of the addition of bevacizumab to adjuvant capecitabine in the QUASAR 2 trial (figure 3).

We leveraged our large sample size to examine the prognostic value of $CD8^{IE}$-$FoxP3^{IS}$ by T stage and N stage. $CD8^{IE}$-$FoxP3^{IS}$ was prognostic in all N stages in univariable analysis (figure 3), identifying subgroups with differing prognosis within identical T and N categories (figure 4), and was the most important predictor of recurrence after T stage in N0 and N1 tumours (figure 4A, B). In N2 disease, $CD8^{IE}$-$FoxP3^{IS}$-high was the most important prognostic factor, and identified subgroups of T4 tumours with 3-year recurrence-free probability ranging from 78% (95% CI 67–91) to 37% (28–49) with differences greater still in left-sided tumours (figure 4; appendix p 45). These results contrasted with those from previous, smaller studies of the consensus Immunoscore,[9] in which prognostic value appeared to be attenuated with increasing N stage ($p_{interaction}$=0·080; appendix p 39).

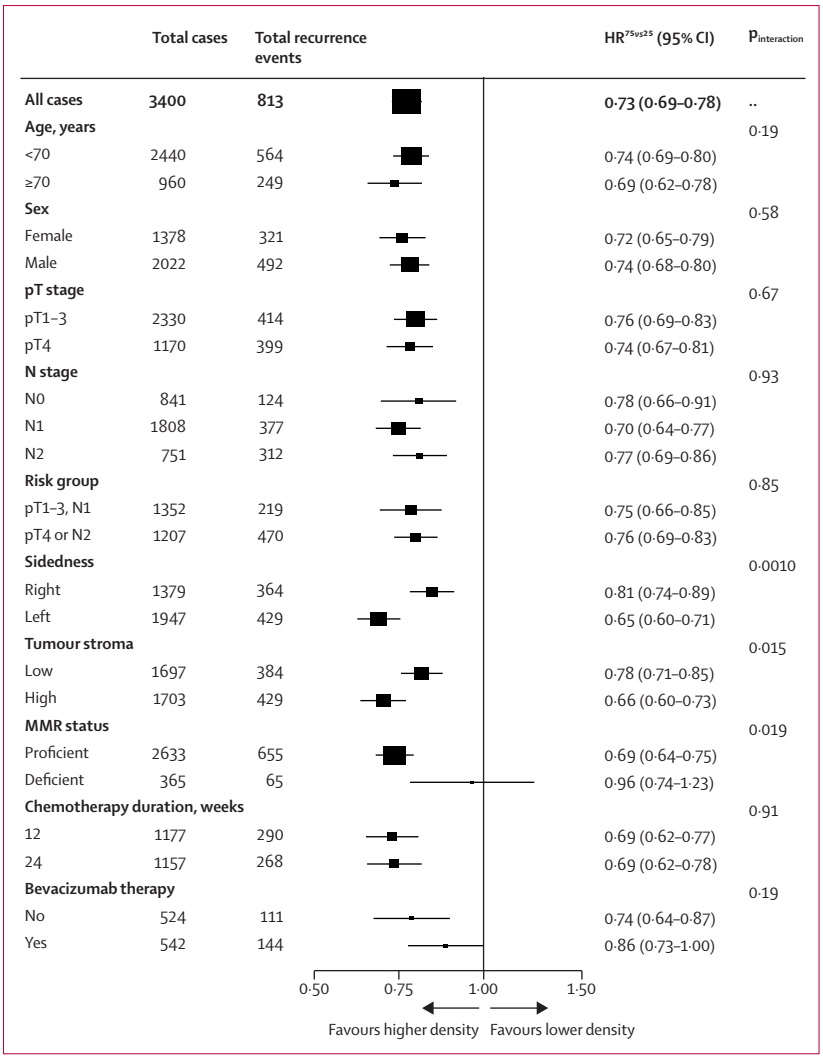


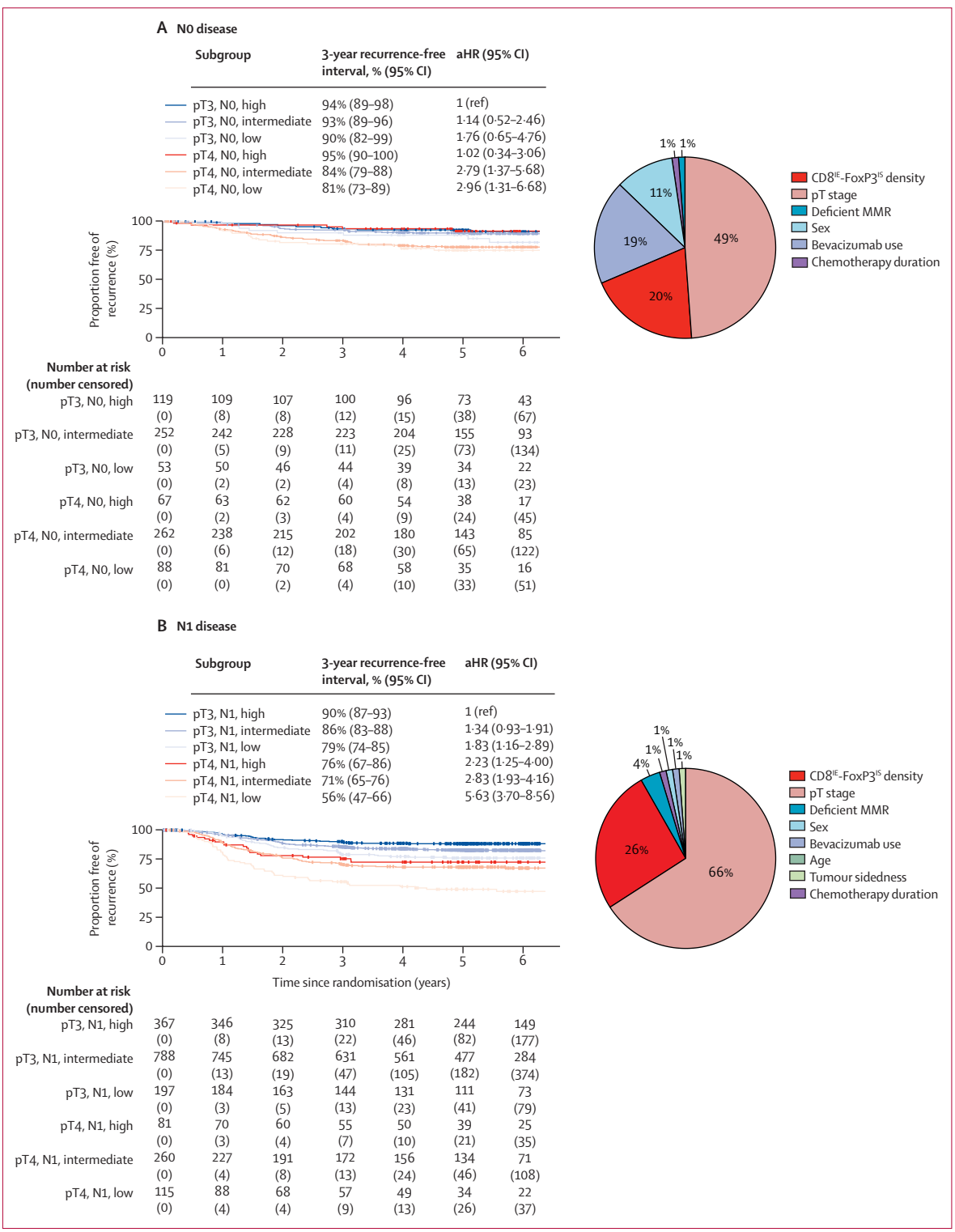
Figure 3: **Variation in prognostic value of $CD8^{IE}$-$FoxP3^{IS}$ for recurrence by subgroups (pooled analysis)**
Forest plot showing colorectal cancer recurrence risk by $CD8^{IE}$-$FoxP3^{IS}$ density according to patient and tumour factors. Estimates are from an unadjusted pooled analysis of 3400 cases, stratified by trial, with the exception of treatment duration (from analysis of 2334 SCOT cases) and bevacizumab use (from analysis of 1066 QUASAR 2 cases). Hazard ratios indicate the risk of colorectal cancer recurrence for cases at the 75th percentile versus those at the 25th percentile of $CD8^{IE}$-$FoxP3^{IS}$ density (HR^{75vs25}). Results from a corresponding subgroup analysis of $CD8^{IE}$-$FoxP3^{IS}$ density as a categorical variable are shown in the appendix (p 43). $CD8^{IE}$=intraepithelial CD8. $FoxP3^{IS}$=intrastromal FoxP3. MMR=DNA mismatch repair.

Discussion

By spatially resolving immune markers into intraepithelial and intrastromal compartments in more than 3000 colorectal cancers, we showed that the combined quantification of $CD8^{IE}$ and $FoxP3^{IS}$ provides superior prognostication to either marker alone, with performance similar to the gold-standard Immunoscore,[9,11,24] validating and extending previous smaller studies.[15,25] To the best of our knowledge, this study is the largest multiparameter immunoprofiling study in cancer to date. We found that this prognostic value varies by tumour sidedness, tumour stromal proportion, and MMR status. Additionally, we found that $CD8^{IE}$ and $FoxP3^{IS}$ have non-overlapping predictors: deficient MMR or MSI predicted increased $CD8^{IE}$ but not $FoxP3^{IS}$ density, whereas tumour stromal proportion was more predictive of $FoxP3^{IS}$ than $CD8^{IE}$ density. Finally, we showed that, compared with $CD8^{IE}$ and $FoxP3^{IS}$, the prognostic values of CD20 and CD68 densities are limited, at least according to our methods.

Our study raises important preclinical questions. While $CD8^{+}$ cytotoxic T cells kill malignant cells via release of cytolytic factors such as perforin and granzyme B, the mechanism by which $FoxP3^{+}$ cells suppress colorectal cancer is unknown. FoxP3 is a marker for regulatory T cells, which are immunosuppressive and pro-tumorigenic in most cancer types.[26] However, although the data are somewhat inconsistent, previous smaller

studies have also shown better outcomes with increasing FoxP3⁺ cell density in colorectal cancer,[15,25] albeit mostly without defining its localisation or confirming its independence from CD8 infiltrate. The apparent contradiction between these studies and ours with an immunosuppressive role of FoxP3⁺ cells in colorectal



(Figure 4 continues on next page)

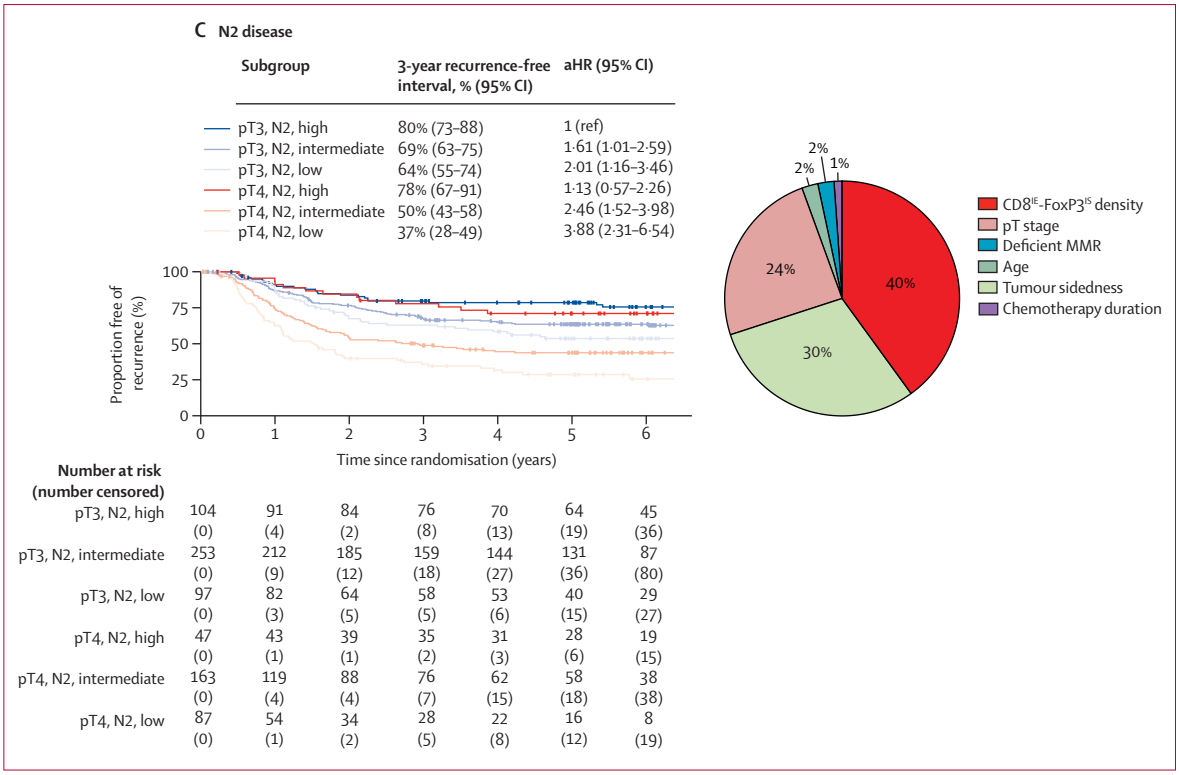


Figure 4: **Stratification of risk by CD8^{IE}-FoxP3^{IS} density and N stage**
Kaplan-Meier plots show the proportion of patients free from recurrence by CD8^{IE}-FoxP3^{IS} density as a three-group variable (high *vs* intermediate *vs* low) in N0 (A), N1 (B), and N2 (C) disease. aHRs were obtained from multivariable pooled analysis of cases with available data and are adjusted for age, sex, tumour sidedness, tumour stromal proportion, MMR and microsatellite instability status, treatment duration, and use of adjuvant bevacizumab. Pie charts show the relative importance of individual variables on multivariable regression of cases with N0 (A), N1 (B), and N2 (C) disease, according to the proportion of the χ^2 statistic. Corresponding plots stratified by tumour sidedness are shown in the appendix (p 45). CD8^{IE}=intraepithelial CD8. FoxP3^{IS}=intrastromal FoxP3. aHR=adjusted hazard ratio. MMR=DNA mismatch repair.

cancer might be due to a subpopulation of non-suppressive, IFN-δ-secreting FoxP3$^+$ cells,[27] or to transient FoxP3 expression in activated T cells.[28,29] Notably, the prognostic value of FoxP3 density was greater in the tumour stroma than in the epithelium, suggesting that the antitumour action of FoxP3$^+$ cells might be indirect. Determining the mechanism of antitumour action and the potential of these cells as immunotherapy targets—particularly in MMR-proficient tumours—is an important question for future studies. Although co-immunofluorescence is unsuited to clinical application, CD8 and FoxP3 are readily detectable by immunohistochemistry, and advances in digital pathology and AI-based image analysis mean that their quantification is likely to be feasible in the clinic within the next decade.[30,31] Development of methods by which to quantify these markers (and do so in the clinic) and to validate our results in additional cohorts appears to be worthwhile, ideally in combination with other promising methods for risk stratification, such as measurement of ctDNA.

Our study has several strengths. The SCOT and QUASAR 2 trial cohorts had carefully curated clinical and pathological data and comprehensive follow-up. Our AI-based method quantified multiple markers in both intraepithelial and intrastromal compartments, providing unprecedented resolution of colorectal cancer immune infiltrate at scale.[17] In turn, this enabled us to identify independent predictors of CD8^{IE} and FoxP3^{IS} densities, which has not been possible in previous studies using manual scoring. Our study design and size enabled discovery and validation in independent trial cohorts and previously unfeasible subgroup analyses.

Our study also has limitations. Tumour samples were not available for all patients in either trial. The clinical, formalin-fixed paraffin-embedded material precluded more granular immunophenotyping to define FoxP3$^+$ cell subsets in our samples. Absence of *KRAS* and *BRAF* mutation status in cases from the SCOT trial precluded definitive analysis of their relationships with immune infiltrate. The expense of multiplex co-immunofluorescence on more than 3000 whole slides led us to use tissue microarray cores as a cost-effective alternative. Although concordance between CD8 cell density quantified by immunofluorescence with that by immunohistochemistry was excellent,[25] testing of the prognostic value of CD8^{IE}-FoxP3^{IS} in whole tissue sections by immunohistochemistry appears to be worthwhile. Similar considerations of cost meant that we were also

unable to analyse other markers of interest, such as PD-L1 and CTLA-4. Finally, the inclusion of cases from the SCOT trial used for variable selection in the pooled analysis might have led to some overfitting, highlighting the need for validation in further trials and real-world cohorts.

In conclusion, we have shown that analysis of FoxP3^{IS} improves upon the well established prognostic value of CD8^{IE} infiltrate in colorectal cancer. Validation of this result could help to improve colorectal cancer risk stratification, and investigation of the underlying mechanisms might reveal novel targets for therapeutic modulation.

Contributors

ALF, DNC, and VHK contributed to conceptualisation of the study. ALF, AM, RRAKS, FJ, LG, TT, AH, TI, MPS, KAO, NM, FP, JH, RK, DK, ED, and DNC contributed to data curation. ALF, AM, RRAKS, FJ, LG, TT, LC, ED, DNC, and VHK contributed to the formal analysis. IT, OS, VHK, and DNC contributed to funding acquisition. ALF, FJ, DNC, and VHK contributed to the investigation. ALF, DNC, and VHK contributed to the methodology. JH, DNC, and VHK contributed to project administration. ALF, AH, TI, MPS, KAO, NM, FP, MG, WK, HED, JH, JE, IT, OS, CK, RK, DK, ED, DNC, and VHK contributed to acquisition of resources. ALF and VHK developed software. VHK and DNC contributed to supervision of the study. ALF, VHK, and DNC contributed to validation. ALF, AM, FJ, ED, VHK, and DNC contributed to data visualisation. DNC wrote the original draft of the manuscript, and all authors contributed to reviewing and editing the manuscript. ALF, AM, DNC and VHK accessed and verified the raw study data. All authors had access to the study data.

Declaration of interests

DK has patents related to use of digital pathology algorithms, and serves as an advisor and has stock in Oxford Cancer Biomarkers. MPS has received honoraria from Merck, Server, Bayer, Takeda, and Amgen for meetings or lectures. WK has received royalties from Room4. DNC has participated in advisory boards for MSD and has received research funding on behalf of the TransSCOT consortium from HalioDx for analyses independent of this study. VHK has served as an invited speaker on behalf of Sharing Progress in Cancer Care and Indica Labs, and has received project-based research funding from The Image Analysis Group and Roche outside of the submitted work. All other authors declare no competing interests.

Data sharing

The datasets pertaining to the SCOT trial used in the current study are available from the TransSCOT collaboration on reasonable request in the absence of scientific conflict of interest. Applications for analysis of TransSCOT samples are welcome and should be addressed to JH (jennifer.hay@glasgow.ac.uk). Datasets and samples from the QUASAR 2 trial are available upon reasonable request and should be addressed to DNC (david.church@well.ox.ac.uk).

Acknowledgments

The SCOT trial was funded by the Medical Research Council (transferred to NETSCC–Efficacy and Mechanism Evaluation; grant number G0601705), the National Institute for Health and Care Research (NIHR) Health Technology Assessment (14/140/84), Cancer Research UK Core Clinical Trials Unit Glasgow funding (C6716/A9894), and the Swedish Cancer Society. The TransSCOT sample collection was funded by a Cancer Research UK Clinical Trials Awards and Advisory Committee—Sample Collection (C6716/A13941). The QUASAR 2 trial was funded by an unrestricted educational grant to DJK from Roche. This biomarker study was funded by the Oxford NIHR Comprehensive Biomedical Research Centre, a Cancer Research UK Advanced Clinician Scientist Fellowship (C26642/A27963) to DC, Cancer Research UK (award A25142) to the Cancer Research UK Glasgow Centre. VHK acknowledges funding by the Promedica Foundation (F-87701-41-01). The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR, or the Department of Health. We thank the patients who participated in the SCOT and QUASAR 2 trials and consented for their samples to be used for correlative research, as well as the recruiting clinicians and study team. We are also grateful to Haitao Wang (Department of Oncology, University of Oxford) for curation of the QUASAR 2 samples, the Translational Histopathology Laboratory (Department of Oncology, University of Oxford) for performing immunostaining and the Glasgow Tissue Research Facility (Glasgow University) for tissue microarray construction and scanning. This Article is dedicated to the memory of Havard Danielsen, who died during review of this study.

References

1 Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021; **71:** 209–49.
2 Argilés G, Tabernero J, Labianca R, et al. Localised colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2020; **31:** 1291–305.
3 Hutchins G, Southward K, Handley K, et al. Value of mismatch repair, *KRAS*, and *BRAF* mutations in predicting recurrence and benefits from chemotherapy in colorectal cancer. *J Clin Oncol* 2011; **29:** 1261–70.
4 Le DT, Uram JN, Wang H, et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl J Med* 2015; **372:** 2509–20.
5 Chalabi M, Fanchi LF, Dijkstra KK, et al. Neoadjuvant immunotherapy leads to pathological responses in MMR-proficient and MMR-deficient early-stage colon cancers. *Nat Med* 2020; **26:** 566–76.
6 Domingo E, Camps C, Kaisaki PJ, et al. Mutation burden and other molecular markers of prognosis in colorectal cancer treated with curative intent: results from the QUASAR 2 clinical trial and an Australian community-based series. *Lancet Gastroenterol Hepatol* 2018; **3:** 635–43.
7 Fakih MG, Kopetz S, Kuboki Y, et al. Sotorasib for previously treated colorectal cancers with *KRAS*^{G12C} mutation (CodeBreaK100): a prespecified analysis of a single-arm, phase 2 trial. *Lancet Oncol* 2022; **23:** 115–24.
8 Kopetz S, Grothey A, Yaeger R, et al. Encorafenib, binimetinib, and cetuximab in *BRAF* V600E-mutated colorectal cancer. *N Engl J Med* 2019; **381:** 1632–43.
9 Pagès F, Mlecnik B, Marliot F, et al. International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. *Lancet* 2018; **391:** 2128–39.
10 Glaire MA, Domingo E, Sveen A, et al. Tumour-infiltrating CD8^{+} lymphocytes and colorectal cancer recurrence by tumour and nodal stage. *Br J Cancer* 2019; **121:** 474–82.
11 Pagès F, André T, Taieb J, et al. Prognostic and predictive value of the Immunoscore in stage III colon cancer patients treated with oxaliplatin in the prospective IDEA France PRODIGE-GERCOR cohort study. *Ann Oncol* 2020; **31:** 921–29.
12 Väyrynen JP, Haruki K, Lau MC, et al. The prognostic role of macrophage polarization in the colorectal cancer microenvironment. *Cancer Immunol Res* 2021; **9:** 8–19.
13 Li J, Li L, Li Y, et al. Tumor-associated macrophage infiltration and prognosis in colorectal cancer: systematic review and meta-analysis. *Int J Colorectal Dis* 2020; **35:** 1203–10.
14 Koelzer VH, Canonica K, Dawson H, et al. Phenotyping of tumor-associated macrophages in colorectal cancer: impact on single cell invasion (tumor budding) and clinicopathological outcome. *OncoImmunology* 2015; **5:** e1106677.
15 Salama P, Phillips M, Grieu F, et al. Tumor-infiltrating FOXP3^{+} T regulatory cells show strong prognostic significance in colorectal cancer. *J Clin Oncol* 2009; **27:** 186–92.
16 Idos GE, Kwok J, Bonthala N, Kysh L, Gruber SB, Qu C. The prognostic implications of tumor infiltrating lymphocytes in colorectal cancer: a systematic review and meta-analysis. *Sci Rep* 2020; **10:** 3360.
17 Iveson TJ, Kerr RS, Saunders MP, et al. 3 versus 6 months of adjuvant oxaliplatin-fluoropyrimidine combination therapy for colorectal cancer (SCOT): an international, randomised, phase 3, non-inferiority trial. *Lancet Oncol* 2018; **19:** 562–78.
18 Kerr RS, Love S, Segelov E, et al. Adjuvant capecitabine plus bevacizumab versus capecitabine alone in patients with colorectal cancer (QUASAR 2): an open-label, randomised phase 3 trial. *Lancet Oncol* 2016; **17:** 1543–57.

19 Taube JM, Akturk G, Angelo M, et al. The Society for Immunotherapy of Cancer statement on best practices for multiplex immunohistochemistry (IHC) and immunofluorescence (IF) staining and validation. *J Immunother Cancer* 2020; **8:** e000155.

20 Viratham Pulsawatdi A, Craig SG, Bingham V, et al. A robust multiplex immunofluorescence and digital pathology workflow for the characterisation of the tumour immune microenvironment. *Mol Oncol* 2020; **14:** 2384–402.

21 Frei AL, McGuigan A, Sinha RR, et al. Accounting for intensity variation in image analysis of large-scale multiplexed clinical trial datasets. *J Pathol Clin Res* 2023; **9:** 449–63.

22 Nowak M, Jabbar F, Rodewald AK, et al. Single cell AI-based detection of DNA mismatch repair deficiency in 1,988 colorectal cancers reveals prognostic and predictive value in the SCOT trial. *medRxiv* 2023; published online Dec 18. https://doi.org/10.1101/2023.12.18.23300137 (preprint).

23 McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer* 2005; **93:** 387–91.

24 Mlecnik B, Bifulco C, Bindea G, et al. Multicenter international Society for Immunotherapy of Cancer study of the consensus Immunoscore for the prediction of survival and response to chemotherapy in stage III colon cancer. *J Clin Oncol* 2020; **38:** 3638–51.

25 Cavalleri T, Bianchi P, Basso G, et al. Combined low densities of FoxP3$^+$ and CD3$^+$ tumor-infiltrating lymphocytes identify stage II colorectal cancer at high risk of progression. *Cancer Immunol Res* 2019; **7:** 751–58.

26 Sakaguchi S, Mikami N, Wing JB, Tanaka A, Ichiyama K, Ohkura N. Regulatory T cells and human disease. *Annu Rev Immunol* 2020; **38:** 541–66.

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:** 679–84.

28 Allan SE, Crome SQ, Crellin NK, et al. Activation-induced FOXP3 in human T effector cells does not suppress proliferation or cytokine production. *Int Immunol* 2007; **19:** 345–54.

29 Tran DQ, Ramsey H, Shevach EM. Induction of FOXP3 expression in naive human CD4$^+$FOXP3 T cells by T-cell receptor stimulation is transforming growth factor-beta dependent but does not confer a regulatory phenotype. *Blood* 2007; **110:** 2983–90.

30 Colling R, Pitman H, Oien K, et al. Artificial intelligence in digital pathology: a roadmap to routine use in clinical practice. *J Pathol* 2019; **249:** 143–50.

31 Koelzer VH, Sirinukunwattana K, Rittscher J, Mertz KD. Precision immunoprofiling by image analysis and artificial intelligence. *Virchows Arch* 2019; **474:** 511–22.