scientific reports

OPEN

Loss of SATB2 expression correlates with cytokeratin 7 and PD-L1 tumor cell positivity and aggressiveness in colorectal cancer

Jan Hrudka[1]✉, Radoslav Matěj[1,2,3], Andrej Nikov[4], Igor Tomyak[4], Hana Fišerová[1], Karolína Jelínková[1] & Petr Waldauf[5]

Colorectal carcinoma (CRC) is a disease that causes significant morbidity and mortality worldwide. To improve treatment, new biomarkers are needed to allow better patient risk stratification in terms of prognosis. This study aimed to clarify the prognostic significance of colonic-specific transcription factor special AT-rich sequence-binding protein 2 (SATB2), cytoskeletal protein cytokeratin 7 (CK7), and immune checkpoint molecule programmed death-ligand 1 (PD-L1). We analyzed a cohort of 285 patients with surgically treated CRC for quantitative associations among the three markers and five traditional prognostic indicators (i.e., tumor stage, histological grade, variant morphology, laterality, and mismatch-repair/MMR status). The results showed that loss of SATB2 expression had significant negative prognostic implications relative to overall survival (OS) and cancer-specific survival (CSS), significantly shortened 5 years OS and CSS and 10 years CSS in patients with CRC expressing CK7, and borderline insignificantly shortened OS in patients with PD-L1 + CRC. PD-L1 showed a significant negative impact in cases with strong expression (membranous staining in 50–100% of tumor cells). Loss of SATB2 was associated with CK7 expression, advanced tumor stage, mucinous or signet ring cell morphology, high grade, right-sided localization but was borderline insignificant relative to PD-L1 expression. CK7 expression was associated with high grade and SATB2 loss. Additionally, a separate analysis of 248 neoadjuvant therapy-naïve cases was performed with mostly similar results. The loss of SATB2 and CK7 expression were significant negative predictors in the multivariate analysis adjusted for associated parameters and patient age. In summary, loss of SATB2 expression and gain of CK7 and strong PD-L1 expression characterize an aggressive phenotype of CRC.

In 2020, there were 1.93 million cases of colorectal carcinoma (CRC), making it the third most common human malignancy. The associated 935,000 deaths make CRC the second most significant cancer-related cause of death worldwide[1] and comprise about 10% of human malignant tumors and cancer-related deaths[2]. Localized tumors and the subset of tumors with lymph node metastases can be successfully treated with surgical resection; patients with distant metastases may benefit from (neo)adjuvant treatment. New biomarkers are needed to identify high-risk tumors and stratify patients in terms of prognosis.

Special AT-rich sequence-binding protein 2 (SATB2) is an evolutionarily conserved protein that binds to the matrix attachment regions activating gene transcription in a matrix attachment region-dependent manner, as described in 2003[3,4]. Matrix attachment tumor regions are regulatory DNA sequences important for higher-order chromatin organization and extension of chromatin modifications. Two homologous proteins, SATB1 and SATB2

[1]Department of Pathology, 3rd Faculty of Medicine, Charles University, University Hospital Kralovske Vinohrady, Prague, Czech Republic. [2]Department of Pathology and Molecular Medicine, 3rd Faculty of Medicine, Charles University, Thomayer University Hospital, Prague, Czech Republic. [3]Department of Pathology, 1st Faculty of Medicine, Charles University, General University Hospital, Prague, Czech Republic. [4]Department of General Surgery, 3rd Faculty of Medicine, Charles University, University Hospital Kralovske Vinohrady, Prague, Czech Republic. [5]Department of Anaesthesia and Intensive Care Medicine, 3rd Faculty of Medicine, Charles University, University Hospital Kralovske Vinohrady, Prague, Czech Republic. ✉email: jan.hrudka@lf3.cuni.cz

bind to these DNA sequences with various regulatory functions in gene expression. In mice, SATB2 transcripts have been identified in pre-B cells, the brain, kidneys, thymus, and testis[4]. SATB2 is constitutively expressed in humans and has developmental roles in craniofacial, neural, and osteoblastic differentiation[3]. Haploinsufficiency, caused by the deletion of several genes, has been reported to induce cleft palates and craniofacial malformations[5]. It has been shown that SATB2 is constitutively expressed in colonic mucosa. A recent murine study showed that colonic stem cell identity was lost when SATB2 expression was lost; inversely, the gain of SATB2 expression in small bowel stem cells leads to the conversion into the colonic phenotype[6]. In human pathology, SATB2-immunohistochemistry has gained importance over the last two decades as a marker of osteosarcoma[7,8]. SATB2 is constitutively expressed in the physiological colorectal mucosa and a majority of colorectal adenocarcinomas[9–14] and is widely used as a routine immunohistochemical marker indicating a colorectal origin to differentiate colorectal adenocarcinomas from other adenocarcinoma primaries. Previous studies described an association between diminished SATB2 expression and a poor CRC prognosis[15–18].

Cytokeratins (CK7) are cytoskeletal structural proteins present in epithelia and in epithelial tumors. CRC mostly expresses CK20 like normal colon mucosa. CK7 occurs in various glands (breast, skin adnexa, salivary, pancreatobiliary ducts) and in many types of adenocarcinomas (breast, lung, pancreatobiliary, salivary). It is also in a minority of CRC, although the CK20-CK7 + expression profile is often thought to indicate a non-colorectal origin. The rate of CK7 + CRCs varies between 0 and 22% in published studies[19–35]. According to several reports, including our recent study, CK7-expression in CRC is associated with aggressive tumor properties relative to more advanced stages[23,26,36] and shorter survival times[27,32,34,35].

The programmed death-ligand 1 (PD-L1) is a cell surface molecule expressed in various immune and tumor cell types. PD-L1 expression allows tumor cells to escape anti-tumor cytotoxic immunity and has been described to worsen the prognosis of several gastrointestinal malignancies, including gastric cancer[36–39], esophageal cancer[40,41], and specific subtypes of pancreatic cancer[42,43]. Despite the small subset of CRCs expressing PD-L1, evidence documents its negative impact on the prognosis[44–56].

To identify particularly aggressive CRC phenotypes, we analyzed a cohort similar to our previous study[57] with nearly 10 years of follow-up focusing on the association between SATB2, CK7, and PD-L1 expression and correlations with traditional prognostic parameters such as stage, grade, anatomical site, and mismatch-repair (MMR) status along with prognostic implications. As a secondary aim, we analyzed the prognostic impact of PD-L1 expression according to tumor cell expression rate.

Material and methods

Study cohort.   We selected the medical records from the pathology department, from 2010 to 2013, of 285 patients with surgically resected histopathologically verified adenocarcinoma of the colon and rectum with known follow-ups and formalin-fixed paraffin-embedded (FFPE) resection specimen tumor tissue available. In first step, cases of all stages were included without selection according to neoadjuvant/adjuvant therapy. Additionally, 37 cases which underwent neoadjuvant radio- and/or chemotherapy (prior to surgery) were excluded and a separate analysis of 248 neoadjuvant therapy-naïve cases was performed. No patients received immune-checkpoint inhibitor treatment. Among these cases were conventional adenocarcinoma, mucinous adenocarcinoma, and signet-ring adenocarcinoma. The grade and stage of the tumor were recorded based on medical records. Stages I–IV was assigned using the TNM Classification[58] and the Union for International Cancer Control (UICC).

Tissue microarray and immunohistochemistry.   Tissue microarray (TMA) techniques were used to make paraffin blocks for immunohistochemical slides using a TMA Master 3D Histech manual tissue arrayer. From each paraffin block containing invasive adenocarcinoma tissue, two cylindrical cores measuring 2 mm were taken from random tumor tissue sites. All cores were collected in a recipient TMA paraffin block. Each recipient block contained 20 samples from 10 cases. For immunohistochemistry, four µm-thick tissue sections were stained in a Ventana BenchMark ULTRA auto-stainer (Ventana Medical Systems, Tucson, Arizona). Monoclonal antibodies against SATB2 (CellMarque, EP281, 1:200), CK7 (clone OV-TL, BioSB, 1:500), PD-L1 (clone SP263, Roche, diagnostic kit), MutS homolog 2 (MSH2, clone G219-1129, Roche, ready to use), postmeiotic segregation 2 (PMS2, clone A16-4, Roche, ready to use), MutS homolog 6 (MSH6, clone 44, Roche, ready to use), and MutL homolog 1 (MLH1, clone M1, Roche, ready to use) were used. The positive reactions were visualized using the Ultraview Detection System (Ventana Medical Systems); slides were counterstained with hematoxylin. Stained slides were dehydrated and covered in a xylene-based mounting medium.

Microscopical analysis, cut points.   Using a microscope, two experienced routine pathologists (JH and RM) assessed all immunohistochemical examinations. For SATB2, only nuclear staining was regarded as positive in both weak and strong staining; the percentage of positive tumor cells in both cores was recorded separately. In cases with discordant finding, a consensus value among both cores was stated and recorded using multihead microscope. The cohort was binarized according to an optimal cut-point calculated using optimization of the log-rank test resulting in 40% of SATB2-positive cells being used as the cut-off value. The surv_cut-point function (survminer) determined the optimal cut-point using the maximally selected rank statistics from the `maxstat` R package (see below). According to the cut-point, the cohort was classified into SATB2 high (hi) and SATB2 low (lo). For CK7, the percentage of positive tumor cells in both cores was recorded separately. In case of discordant finding, a consensus value among both cores obtained using multihead microscope was discussed and recorded; staining in ≥ 10% of tumor cells was considered a positive sample. PD-L1 was considered positive only in membranous staining. The percentage of positive neoplastic cells (tumor proportion score—TPS) was recorded following standardized recommendations in both cores separately by both pathologists. In case of a
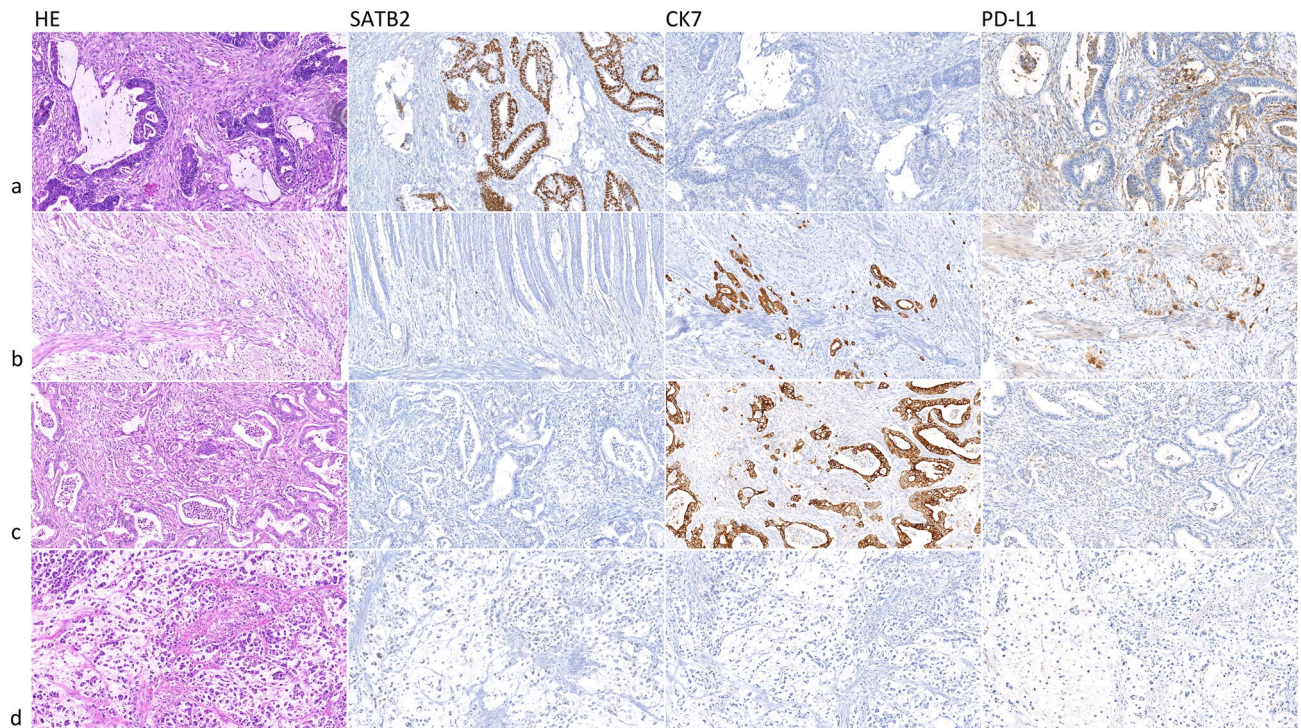
Figure 1. Histological slides showing colorectal carcinoma (CRC), hematoxylin–eosin (HE) stain, and immunohistochemistry, magnification 20×: (**a**) conventional adenocarcinoma with strong nuclear SATB2 positivity, CK7 negativity, and PD-L1 positivity in immune cells and negativity in tumor cells; (**b**) conventional adenocarcinoma with SATB2 negativity (staining in ca 1% tumor cells), strong cytoplasmic CK7 expression in 100% tumor cells, and strong membranous PD-L1 expression in ca 60% tumor cells; (**c**) conventional adenocarcinoma with SATB2 negativity, strong cytoplasmic CK7 positivity in 100% tumor cells, and PD-L1 negativity; (**d**) signet ring cell carcinoma with intracellular and extracellular mucin production showing SATB2 negativity (weak staining in ca 20% tumor cells), CK7 negativity, and PD-L1 negativity.

discordant finding, a consensus value was discussed using multihead microscope. In all three examined markers, the consensus value was used in statistical analysis. Regarding MMR status, tumors with any apparent nuclear staining with MSH2, MSH6, PMS2, and MLH1 were considered MMR-proficient. Tumors with an obvious loss of nuclear staining of anti-MMR antibodies with control positivity in stroma and lymphocytes were considered MMR-deficient. All microscopic analyses (Fig. 1) were performed without knowledge of clinical data and patient follow-up.

Statistics. Overall survival (OS) and cancer-specific survival (CSS) were calculated from the date of surgery to the date of recorded death or the last known follow-up date (censoring). Concerning CSS, patients with non-CRC-related causes of death were censored on the date of death. Separate analyses for 5-year and 10-year follow-ups were performed. For survival analysis, we performed a univariate Kaplan–Meier analysis with the log-rank test and confidence intervals calculated using the log–log method. For estimation of the restricted mean (rmean) survival in each subgroup, a restricted mean survival time (RMST) analysis with a 95% confidence interval (CI) was used. To calculate the hazard ratio (HR) for each parameter, univariate Cox regressions with 95% CIs were performed. Logistic regressions (Pearson's chi-squared test) were used to assess associations between measured variables.

Furthermore, we performed separate 5-year and 10-year OS and CSS analyses on three subgroups, i.e., PD-L1 negative, weak and moderate expressors (CRCs expressing PD-L1 in 1–49% of tumor cells), and strong expressors (50–100%). Survival analysis, RMSE, Cox regression, and logistic regression were performed separately for the binarized cohort as follows: SATB2hi (> 40%) vs. SATB2lo (≤ 40%), CK7+ (≥ 10%) vs. CK7− (< 10%), PD-L1 + (≥ 1%) vs. PD-L1− (< 1%), low grade (grade 1 + 2) vs. high grade (grade 3), conventional adenocarcinoma vs. adenocarcinoma with variant morphology (mucinous + signet ring carcinoma), MMR-proficient vs. MMR-deficient, right sided tumors (cecum, ascendens, hepatic flexure, transversum) versus left sided tumors (lienal flexure, descendens, sigmoid, rectum), and localized tumors (UICC stages 1 + 2) vs. advanced tumors (UICC stages 3 + 4).

In the next step, multivariate Cox regressions adjusting associated parameters and adjusting analyzed variables on the patient's age were performed. To evaluate reliability rates among two tumor tissue cores and among both pathologists, unweighted Cohen`s Kappa value was calculated. The Cohen´s kappa test was used to analyze binarized variables using cut-off values as described above (SATB2hi vs. SATB2lo, CK7+ vs. CK7−, PD-L1+ (≥ 1%) vs. PD-L1− (< 1%)). P values < 0.05 were considered statistically significant. All analyses were performed in R version 4.0.3 (2020-10-10)[59]; survival analyses were done using package survival version 3.2-7[60].

	CK7− SATB2lo PD-L1negative	CK7+ SATB2lo PD-L1negative	CK7− SATB2hi PD-L1 negative	CK7+ SATB2hi PD-L1negative	CK7− SATB2lo PD-L1 ≥ 1%	CK7+ SATB2lo PD-L1 ≥ 1%	CK7− SATB2hi PD-L1 ≥ 1%	CK7+ SATB2hi PD-L1 ≥ 1%	Total
Total	34	11	206	6	7	2	19	0	285 (100%)
Age group									
< 50	1	1	10	0	0	0	1	0	13 (4.6%)
50–59	4	3	29	2	0	0	2	0	40 (14.0%)
60–69	10	4	74	1	0	0	5	0	94 (33.0%)
70–79	12	1	56	2	5	2	7	0	85 (29.8%)
80+	7	2	37	1	2	0	4	0	53 (18.6%)
Sex									
F	11	8	88	1	4	1	11	0	124 (43.5%)
M	23	3	118	5	3	1	8	0	161 (56.5%)
UICC stage									
I	3	1	22	2	0	0	5	0	33 (11.6%)
II	10	4	82	3	1	0	10	0	110 (38.6%)
III	14	5	77	1	5	0	4	0	106 (37.2%)
IV	7	1	25	0	1	2	0	0	36 (12.6%)
Situs									
Cecum	6	4	23	0	1	1	3	0	38 (13.3%)
Ascendens	4	3	23	1	1	0	2	0	34 (11.9%)
Hepatic flexure	3	0	9	0	1	0	1	0	14 (4.9%)
Transversum	2	0	10	0	0	0	2	0	14 (4.9%)
Lienal flexure	1	0	8	1	0	0	0	0	10 (3.5%)
Descendens	0	1	9	1	0	0	1	0	12 (4.2%)
Sigmoideum	8	1	49	1	0	0	7	0	66 (23.2%)
Rectosigmoideum	2	0	18	0	0	0	3	0	23 (8.1%)
Rectum	8	2	55	2	3	1	0	0	71 (24.9%)
Multiple	0	0	20	0	1	0	0	0	3 (1.1%)
MMR									
Deficient	4	1	13	1	2	0	4	0	25 (8.8%)
Proficient	30	10	193	5	5	2	15	0	260 (91.2%)
Grade									
Low grade (1 + 2)	23	5	158	3	3	1	12	0	205 (73.0%)
High grade (3)	11	6	46	3	2	1	7	0	76 (27.0%)
Unknown	0	0	2	0	2	0	0	0	4
Morphology									
Adenocarcinoma NOS	30	9	198	6	6	2	18	0	269 (94.4%)
Mucinous + signet ring	4	2	8	0	1	0	1	0	16 (5.6%)

Table 1. Entire cohort. Summary of all cases and all examined variables, and the number of subjects according to their CK7/SATB2/PD-L1 profile.

Results

Cohort description. The entire cohort consisted of 161 male and 124 female patients, with a mean age of 68.55 years, a median age of 69 years, a mode of 70 years, and a range of 30–94 years. The neoadjuvant therapy-naïve cohort consisted of 136 male and 112 female patients, with a mean age of 69.2 years, and a range of 30–94 years. 37 cases which underwent neoadjuvant radio- and/or chemotherapy had a tumor situated in rectum or rectosigmoid. All variables analyzed in the study are summarized in Table 1 (entire cohort) and Table 2 (neoadjuvant therapy-naïve cohort) and Supplementary Table 1. Results of the 5-year OS and CSS analyses of the entire cohort are summarized in Supplementary Table 2, and the 10-year OS and CSS analyses of the entire in Supplementary Table 3. The results of 5- and 10-year survival analysis of the neoadjuvant therapy-naïve cohort are summarized in Supplementary Tables 4 and 5, respectively. Kaplan–Meier curves showing the 5-year OS and

CSS in the entire cohort according to immunohistochemical profile and traditional variables are shown in Figs. 2 and 3, respectively. Kaplan–Meier curves showing the 10-year OS and CSS in the entire cohort are shown in Supplementary Figs. 1 and 2, respectively. Kaplan–Meier curves showing survival in the neoadjuvant therapy-naïve cohort are shown in Supplementary Figs. 3–6.

The unweighted Cohen's kappa test showed almost perfect intercore reliability in SATB2 (K = 0.85) and CK7 (K = 0.913); in PD-L1, there was a substantial agreement among both pathologists in core 1 (K = 0.795), and nearly perfect agreement among both pathologists in core 2 (K = 0.884). All findings in individual cores are listed in Supplementary Table 1, the kappa values are listed in Supplementary Table 6.

Entire cohort. *SATB2 survival analysis.* The patients with SATB2lo tumors (n = 54) had significantly shorter 5-year OS (rmean = 3.157 vs. 4.008 years, HR = 2.17, p = 0.00019, Fig. 2a) and 5-year CSS (rmean = 3.494 versus 4.171 years, HR = 2.16, p = 0.0012, Fig. 3a) compared to those with SATB2hi tumors (n = 231). Analyzing the 10-year follow-up, there was significantly shorter OS (restricted mean/rmean = 4.921 years vs. 6.943 years, hazard ratio = 1.96, p = 0.00042, Sup. Fig. 1a) and CSS (rmean = 5.633 years versus 7.648 years, HR = 2.21, p = 0.00027, Sup. Fig. 2a) in the patients with SATB2lo tumors.

CK7 survival analysis. In the 5-year follow-up, there was a negative prognostic impact associated with CK7 expression with significantly shorter OS (rmean = 3.124 vs. 3.898, HR = 2.13, p = 0.012, Fig. 2b) and CSS (rmean = 3.378 vs. 4.096 years, HR = 2.28, p = 0.012, Fig. 3a) comparing CK7 + (n = 19) with CK7 − (n = 266) cases, respectively. In the 10-year follow-up, CK7 expression had a detrimental prognostic impact with borderline insignificantly shorter OS (rmean 4.9 years vs. 6.677 years, HR = 1.70, p = 0.077, Sup. Fig. 1b) and significantly shorter CSS (rmean = 5.468 years vs. 7.412, HR = 2.0, p = 0.035, Sup. Fig. 2b).

PD-L1 survival analysis. In the 5-year follow up, PD-L1 + tumors (n = 28) displayed insignificantly shorter OS (rmean = 3.261 vs. 3.91, HR = 1.52, p = 0.156, Fig. 2c) and no differences in CSS (rmean = 3.948 vs. 4.063, HR-1.02, p = 0.97, Fig. 3c) compared to PD-L1- (n = 257) cases. In the 10-year follow up, PD-L1 expression showed insignificantly shorter OS (rmean = 5.527 years vs. 6.670 years, HR = 1.46, p = 0.15, Sup. Fig. 1c) and no significant difference in CSS (rmean = 7.29 years vs. 7.329 years, HR = 0.98, p = 0.97, Sup. Fig. 2c) between PD-L1+ and PD-L1− CRCs.

PD-L1 was further analyzed by categorizing the patients based on the percentage of tumor cells expressing PD-L1 with the following results (Supplementary Table 7): strong PD-L1 expressors (expression in 50–100% of tumor cells, n = 4) compared to weak and moderate (1–49%, n = 24) and negative cases (0% n = 257) displayed borderline insignificantly shorter 5-year OS (rmean = 3.910 vs. 3.446 vs. 2.148 years, respectively, p = 0.062, Fig. 4a) and significantly shorter 10-year OS (rmean = 2.641 vs. 6.033 vs. 6.670 years, respectively, p = 0.01). There were no significant differences in both 5-year CSS (rmean = 3.838 vs. 3.976 vs. 4.063 years, p = 0.94, Fig. 4b) and in 10-year CSS (rmean = 7.171 vs. 7.392 vs. 7.29 years, p = 0.96).

Traditional prognosticators. Among traditional prognostic variables studied in the 5-year survival analysis, OS was significantly shorter in case of advanced UICC stage (p < 0.0001, Fig. 2e), grade 3 (p = 0.0031, Fig. 2g), right-sided tumors (p = 0.012, Fig. 2h), and carcinomas with mucinous or signet ring cell morphology (p = 0.046, Fig. 2f), whereas CSS was worse in tumors with advanced UICC stages (p < 0.0001, Fig. 3e), grade 3 (p = 0.012, Fig. 3g), right-sided tumors (p = 0.012, Fig. 3h), and MMR-proficient CRCs (p = 0.028, Fig. 3d)—see Supplementary Table 2. In the 10-year OS, there was a significant negative prognostic impact of advanced UICC stages (p < 0.0001, Sup. Fig. 1e) and grade 3 (p = 0.013, Sup. Fig. 1g). Concerning the 10-year follow-up for CSS, there was a significant worsening effect of the MMR-proficient status (p = 0.0091, Sup. Fig. 2d), advanced UICC stage (p < 0.0001, Sup. Fig. 2e), grade 3 (p = 0.021, Sup. Fig. 2g), and borderline insignificant effect of right-sided tumor sites (p = 0.064, Sup. Fig. 2h)—see Supplementary Table 3.

Logistic regression. The results of the logistic regression analyses (Pearson's chi-squared test) are summarized in Table 3. The analyses revealed significant associations: SATB2lo tumors were prone to be in a more advanced stage (Odds ratio = 0.468, p = 0.016), high-grade (OR = 0.518, p = 0.042), with variant mucinous or signet-ring cell morphology (OR = 3.674, p = 0.014), right-sided (OR = 2.068, p = 0.017), and cytokeratin 7 positive (OR = 0.084, p < 0.001). Furthermore, there was a borderline insignificant association of SATB2 negativity with PD-L1 expression in CRC (OR = 0.448, p = 0.066, Fig. 5). CK7 positive tumors were more frequently high-grade (OR = 3.3, p = 0.013) and borderline insignificantly more right-sided (OR = 0.445, p = 0.093). No association of CK7 expression with UICC stage, morphology, PD-L1 status, or MMR status was found. PD-L1+ CRCs were more frequently MMR-deficient compared to PD-L1− tumors (OR = 0.293, p = 0.018). No association of PD-L1 positivity with UICC stage, grade, morphology, laterality, and CK7 status was identified. A significant association between CK7 expression and SATB2 low status and the borderline insignificant association of PD-L1 expression with SATB2 low status were described above.

Neoadjuvant therapy-naïve cohort. *SATB2 survival analysis.* After exclusion of patients after neoadjuvant therapy, SATB2lo status (n = 49) remained a significant predictor of shorter 5-year OS (rmean = 3.022 vs. 3.961 years, HR = 2.28, p = 0.00011, Sup. Fig. 3a), 5-year CSS (rmean = 3.383 versus 4.133 years, HR = 2.26, p < 0.001, Sup. Fig. 4a), 10-year OS (rmean = 4.621 years vs. 6.811 years, hazard ratio = 2.04, p = 0.00029, Sup. Fig. 5a), and 10-year CSS (rmean = 5.376 years versus 7.574 years, HR = 2.32, p = 0.00022, Sup. Fig. 6a) compared to those with SATB2hi tumors (n = 199).
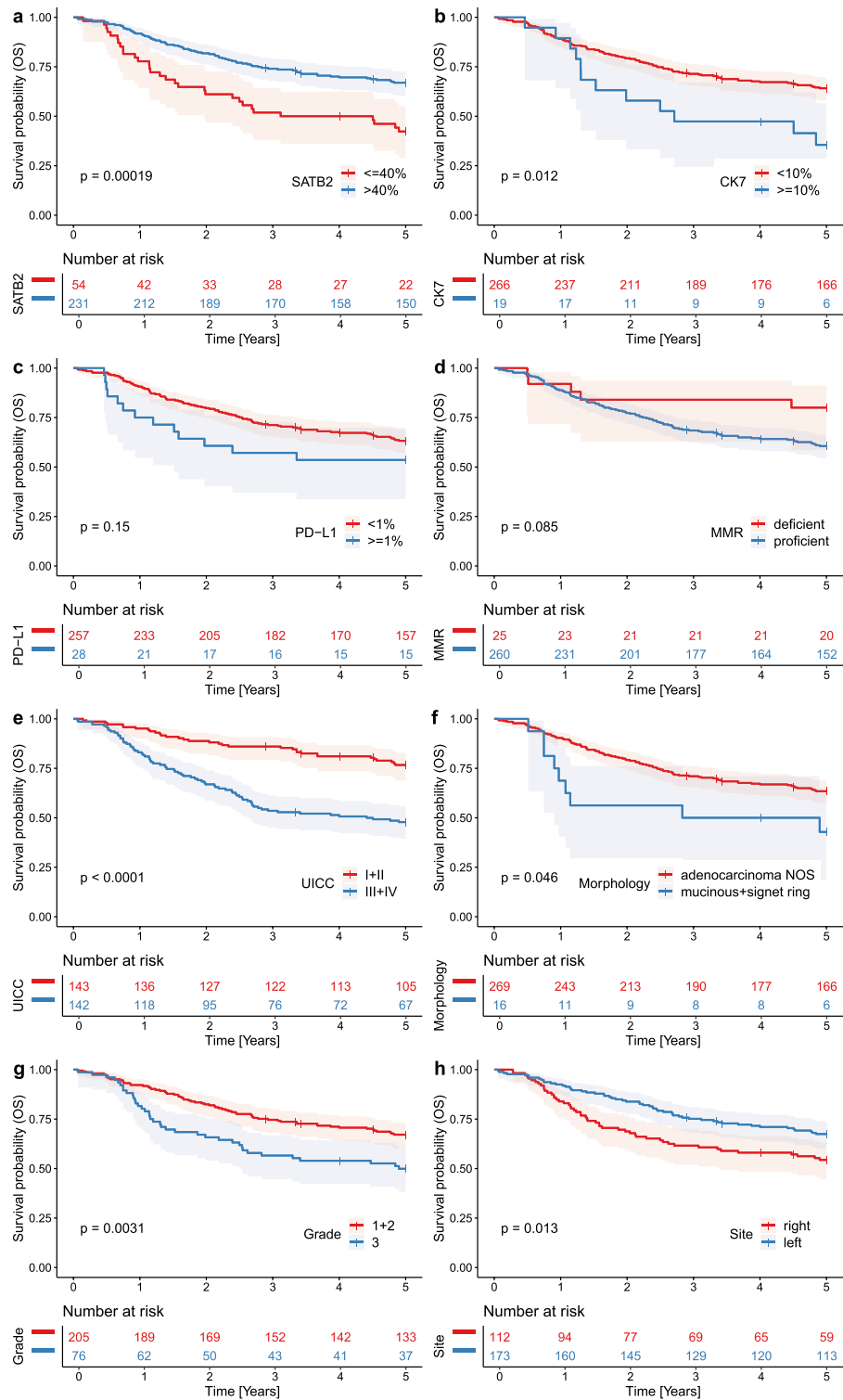
Figure 2. Entire cohort, Kaplan Meier curves documenting 5-year overall survival (OS) in patients with CRC. Note decreased survival in patients with CRC with SATB2 low expression (**a**), CK7 high expression (**b**), PD-L1 expression (**c**), mismatch-repair proficient status (**d**), advanced stage (**e**), mucinous or signet ring cell morphology (**f**), high-grade tumors (**g**), and right-sided tumor localization (**h**). P values from the log-rank test < 0.05 indicate a significant difference in survival.
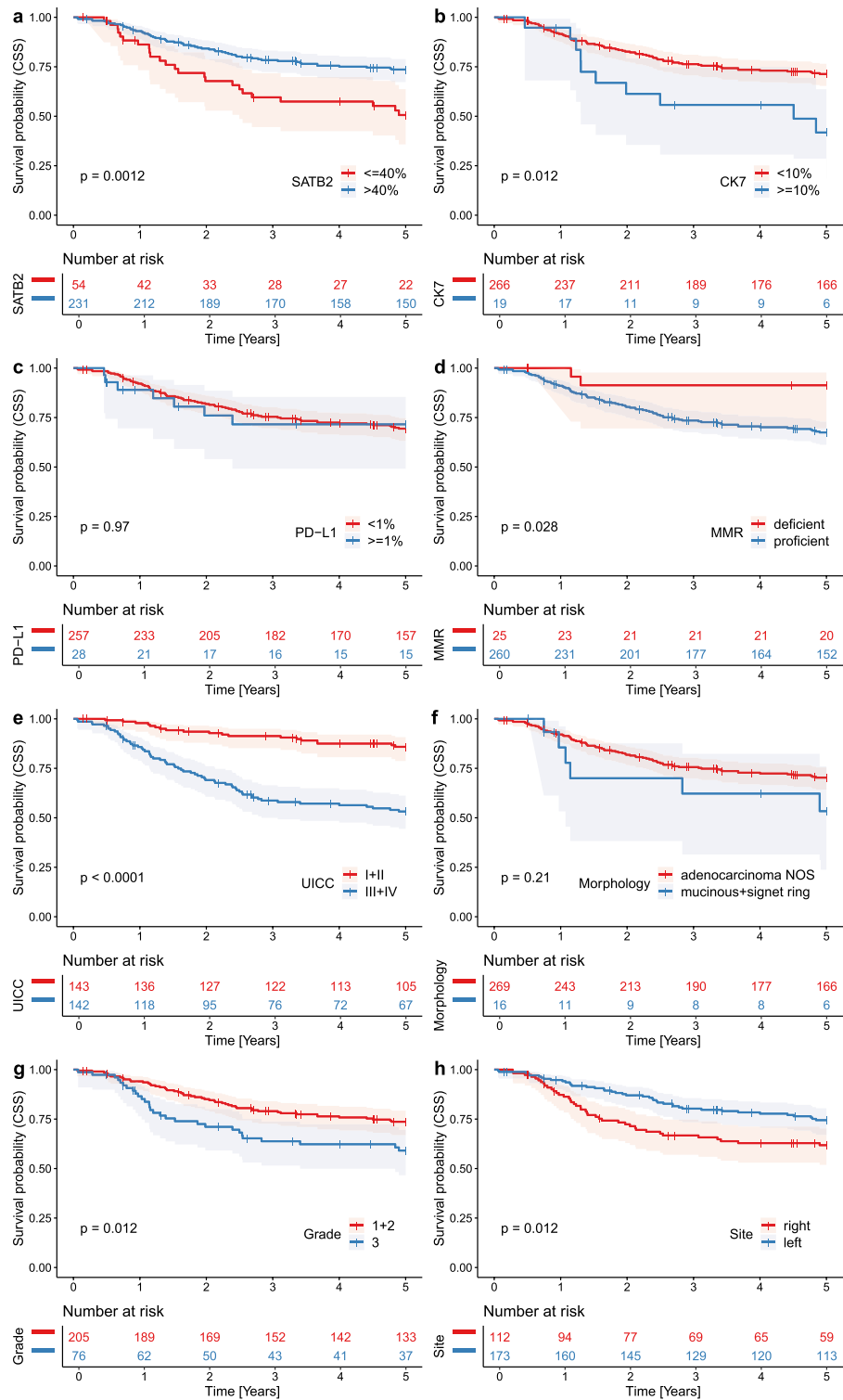
Figure 3. Entire cohort, Kaplan Meier curves documenting worse 5-year cancer-specific survival (CSS) in patients with CRC with SATB2 low expression (**a**), CK7 high expression (**b**), mismatch-repair proficient status (**d**), advanced stage (**e**), high-grade tumors (**g**), and right-sided tumor localization (**h**). There is little and no impact of a mucinous or signet ring morphology (**f**), and PD-L1 expression (**c**), respectively. P values from the log-rank test < 0.05 indicate a significant difference in survival.

|	CK7− SATB2lo PD-L1negative	CK7+ SATB2lo PD-L1negative	CK7− SATB2hi PD-L1 negative	CK7+ SATB2hi PD-L1negative	CK7− SATB2lo PD-L1≥1%	CK7+ SATB2lo PD-L1≥1%	CK7− SATB2hi PD-L1≥1%	CK7+ SATB2hi PD-L1≥1%	Total
Total	31	11	175	5	5	2	19	0	248 (100%)
Age group									
<50	1	1	10	0	0	0	1	0	13 (5.2%)
50–59	4	3	25	1	0	0	2	0	35 (14.1%)
60–69	7	4	57	1	0	0	5	0	74 (29.8%)
70–79	12	1	46	2	3	2	7	0	73 (29.4%)
80+	7	2	37	1	2	0	4	0	53 (21.4%)
Sex									
F	11	8	77	1	3	1	11	0	112 (45.2%)
M	20	3	98	4	2	1	8	0	136 (54.8%)
UICC stage									
I	1	1	18	2	0	0	5	0	27 (10.9%)
II	10	4	69	2	1	0	10	0	96 (38.7%)
III	13	5	63	1	4	0	4	0	90 (36.3%)
IV	7	1	25	0	0	2	0	0	35 (14.1%)
Situs									
Cecum	6	4	23	0	1	1	3	0	38 (15.3%)
Ascendens	4	3	23	1	1	0	2	0	34 (13.7%)
Hepatic flexure	3	0	9	0	1	0	1	0	14 (5.6%)
Transversum	2	0	10	0	0	0	2	0	14 (5.6%)
Lienal flexure	1	0	8	1	0	0	0	0	10 (4.0%)
Descendens	0	1	9	1	0	0	1	0	12 (4.8%)
Sigmoideum	8	1	48	1	0	0	7	0	65 (26.2%)
Rectosigmoideum	2	0	16	0	0	0	3	0	21 (8.5%)
Rectum	5	2	27	1	1	1	0	0	37 (14.9%)
Multiple	0	0	2	0	1	0	0	0	3 (1.2%)
MMR									
Deficient	4	1	13	1	2	0	4	0	25 (10.1%)
Proficient	27	10	162	4	3	2	15	0	223 (89.9%)
Grade									
Low grade (1+2)	20	5	130	3	3	1	12	0	174 (74.7%)
High grade (3)	9	6	36	2	1	1	7	0	59 (25.3%)
Unknown	0	0	2	0	1	0	0	0	4
Morphology									
Adenocarcinoma NOS	27	9	168	5	4	2	18	0	233 (94.0%)
Mucinous + signet ring	4	2	7	0	1	0	1	0	15 (6.0%)

Table 2. Neoadjuvant therapy-naïve cases and all examined variables, and the number of subjects according to their CK7/SATB2/PD-L1 profile.

CK7 survival analysis. After elimination of pretreated patients, CK7 expression retained it's negative prognostic impact in 5-year OS (rmean = 3.019 vs. 3.835, HR = 2.16, p = 0.01, Sup. Fig. 3b), 5-year CSS (rmean = 3.281 vs. 4.048 years, HR = 2.35, p = 0.0098, Sup. Fig. 4b), and 10-year CSS (rmean = 5.178 years vs. 7.317, HR = 2.12, p = 0.023, Sup. Fig. 5b) in CK7+ (n = 18) versus CK7− (n = 230) tumors, respectively. In 10-year OS, the effect was borderline insignificant (rmean 4.606 years vs. 6.515 years, HR = 1.75, p = 0.062, Sup. Fig. 6b), similarly to the entire cohort.

PD-L1 survival analysis. In the neoadjuvant therapy-naïve patients, PD-L1 expression showed insignificantly shorter 5-year OS (rmean = 3.227 vs. 3.84 years, HR = 1.42, p = 0.26, Sup. Fig. 3c) and 10-year OS (rmean = 5.616 years vs. 6.468 years, HR = 1.29, p = 0.37, Sup. Fig. 5c) and no significant difference in both 5-year CSS (rmean = 3.982 vs. 4.0, HR = 0.88, p = 0.77, Sup. Fig. 4c) and 10-year CSS (rmean = 7.678 years vs. 7.133 years, HR = 0.76, p = 0.51, Sup. Fig. 6c) comparing PD-L1+ (n = 26) and PD-L1− (n = 222) tumors, respectively.

PD-L1 was analyzed according to the percentage of tumor cells expressing PD-L1 (Supplementary Table 8). Similarly, to the entire cohort, the patients with strong PD-L1 expression (50–100%, n = 4) compared to both weak and moderate (1–49%, n = 22) and negative tumors (0%, n = 222) displayed significantly shorter 10-year
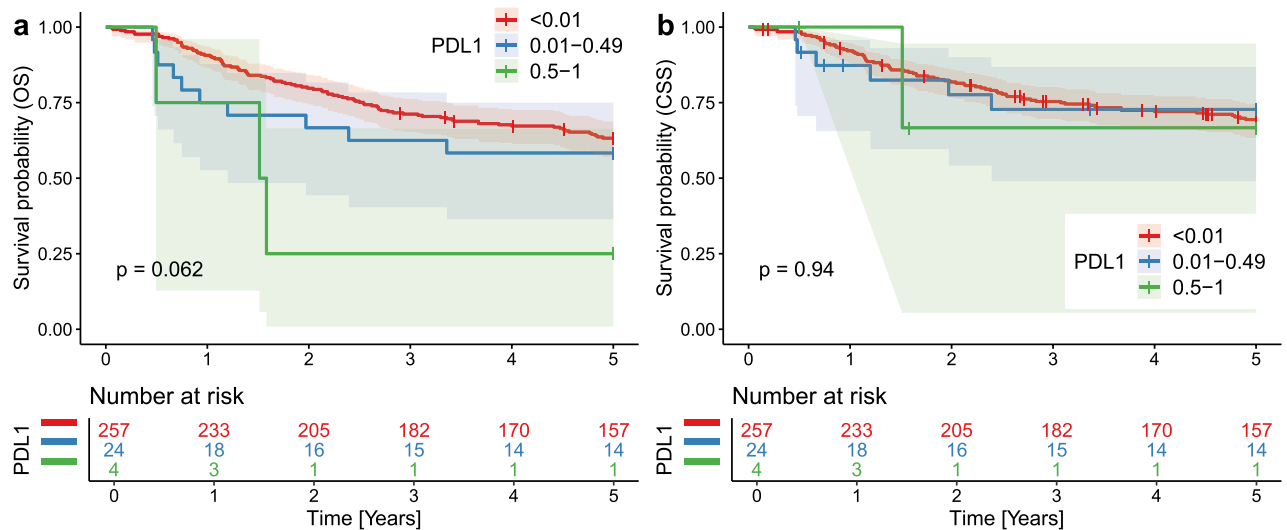
Figure 4. Entire cohort, Kaplan Meier curves documenting (**a**) worse 5-year overall survival (OS) in patients with CRC strongly (50–100% tumor cells) expressing PD-L1 compared to weak and moderate expressors and negative cases; (**b**) no significant differences in cancer-specific survival (CSS). P values from the log-rank test < 0.05 indicate a significant difference in survival.

OS (rmean = 6.468 vs. 6.19 vs. 2.641 years, respectively, p = 0.019) and borderline insignificantly shorter 5-year OS (rmean = 6.873 vs. 6.378 vs. 3.398, respectively, p = 0.099). There were almost no differences in both 10-year CSS (rmean = 7.133 vs. 7.781 vs. 7.171, p = 0.74) and 5-year CSS (rmean = 7.367 vs. 7.781 vs. 7.171, p = 0.89).

Traditional prognosticators. After exclusion of pretreated cases, 5-year OS was significantly shorter along with advanced UICC stage (p < 0.0001, Sup. Fig. 3e), and grade 3 (p = 0.0033, Sup. Fig. 3g), and borderline insignificantly shorter in right-sided tumors (p = 0.057, Sup. Fig. 3h), in tumors with mucinous and signet ring cell morphology (p = 0.17, Sup. Fig. 3f), and in MMR-proficient CRCs (p = 0.053, Sup. Fig. 3d). 5-year CSS was shorter in tumors with advanced UICC stages (p < 0.0001, Sup. Fig. 4e), grade 3 (p = 0.038, Sup. Fig. 4g), right-sided tumors (p = 0.036, Sup. Fig. 4h), and MMR-proficient CRCs (p = 0.019, Sup. Fig. 4d)—see Supplementary Table 4. In the 10-year OS, there was a significant negative prognostic effect of advanced UICC stages (p < 0.0001, Sup. Fig. 5e) and borderline insignificant impact of MMR-proficient status (p = 0.094, Sup. Fig. 5d), and grade 3 (p = 0.12, Sup. Fig. 5g). In the 10-year CSS, there was significantly worse survival in the MMR-proficient tumors (p = 0.0072, Sup. Fig. 6d), CRCs in advanced UICC stage (p < 0.0001, Sup. Fig. 6e), and borderline insignificantly worse survival in grade 3 tumors (p = 0.087, Sup. Fig. 6g) and right-sided tumors (p = 0.1, Sup. Fig. 6h)—see Supplementary Table 5.

Logistic regression. The results of the logistic regression analyses (Pearson's chi-squared test) after exclusion of cases after neoadjuvant therapy are summarized in Supplementary Table 9. The analyses showed following significant associations: SATB2lo tumors were more frequently in advanced stage (OR = 0.466, p = 0.022), with variant mucinous or signet-ring cell morphology (OR = 5.979, p = 0.011), right-sided (OR = 2.027, p = 0.029), and cytokeratin 7 positive (OR = 0.071, p < 0.001, Fig. 6). CK7 positive tumors were more frequently high-grade, the association was borderline insignificant (OR = 2.486, p = 0.087). PD-L1+ CRCs were again more frequently MMR-deficient compared to PD-L1− tumors (OR = 0.312, p = 0.026).

Multivariate Cox regression. In the next step, the prognostic impact of SATB2 and CK7 on the 5-year CSS was analyzed using a multivariate Cox regression with 95% CIs adjusting for associated variables in the neoadjuvant therapy-naïve cohort. The associations were known from previous logistic regression. The 5-year survival was analyzed using a multivariate analysis relative to the higher level of statistical significance compared to the 10-year survival. In this multivariate analysis, SATB2 low status remained a significant predictor (HR = 1.92, p = 0.011) of poor 5-year CSS if adjusted on patient´s age (HR = 1.023, p = 0.032), UICC stage (HR = 5.49, p < 0.001), mucinous or signet ring cell morphology (HR = 1.156, p = 0.759) but SATB2 was insignificant (HR = 1.56, p = 0.131) if adjusted on CK7 positivity (HR = 2.08, p = 0.061). CK7 remained a significant (HR = 2.12, p = 0.038) negative prognostic factor if adjusted on patient´s age (HR = 1.02, p = 0.111) and histopathological grade (HR = 1.6, p = 0.07) but CK7 was insignificant (HR = 1.56, p = 0.262) if adjusted on SATB2 (HR = 1.77, p = 0.05).

Summary of results. In summary, our study revealed significantly shorter OS and CSS in patients with SATB2lo and CK7+ CRC and borderline insignificantly shorter OS in patients with PD-L1+ CRC in the neoadjuvant therapy-naïve and in the entire cohort. SATB2 was significant stage- and age-independent negative predictor in the multivariate Cox regression. SATB2 low status was associated with CK7 expression, right-sided site, mucinous

	SATB2<=40% (n=54)	SATB2 >40% (n=231)	CK7 >=10% (n=19)	CK7 negative (n=266)	PD-L1>= 1% (n=19)	PD-L1 negative (n=266)
SATB2<=40% (n=54)			13 (24.0%)	41 (76.0%)	9 (16.7%)	45 (83.3%)
SATB2 >40% (n=231)			6 (2.6%) OR=0.084	225 (97.4%) **p<0.001**	19 (8.2%) OR=0.448	212 (91.8%) **p=0.066**
CK7 >=10% (n=19)	13 (68.4%)	6 (31.6%)			2 (10.5%)	17 (89.5%)
CK7 negative (n=266)	41 (15.4%) OR=0.084	225 (84.6%) **p<0.001**			26 (9.8%) OR=1.086	240 (90.2%) p=0.915
PD-L1>= 1%(n=19)	9 (16.7%)	19 (8.2%)	2 (7.1%)	26 (92.9%)		
PD-L1 negative (n=266)	45 (83.3%) OR=0.448	212 (91.8%) **p=0.066**	17 (6.6%) OR=1.086	240 (93.4%) p=0.915		
MMR-proficient (n=260)	47 (18.1%)	213 (81.9%)	17 (6.6%)	240 (93.4%)	22 (8.5%)	238 (91.5%)
MMR-deficient (n=25)	7 (28.0%) OR=1.762	18 (72.0%) p=0.232	2 (7.1%) OR=0.805	26 (92.9%) p=0.78	6 (24.0%) OR=0.293	19 (76.0%) **p=0.018**
UICC I+II (n=143)	19 (13.3%)	124 (86.7%)	10 (7.0%)	133 (93.0%)	16 (11.2%)	127 (88.8%)
UICC III+IV (n=142)	35 (24.6%) OR=0.468	107 (75.4%) **p=0.016**	9 (6.3%) OR=0.9	133 (93.7%) p=0.825	12 (8.5%) OR=0.733	130 (91.5%) p=0.439
Adenocarcinoma NOS (269)	47 (17.5%)	222 (82.5%)	17 (6.3%)	252 (93.7%)	26 (9.7%)	243 (90.3%)
Mucinous+signet ring carcinoma (n=16)	7 (43.8%) OR=3.674	9 (56.2%) **p=0.014**	2 (12.5%) OR=0.472	14 (87.5%) p=0.346	2 (12.5%) OR=0.749	14 (87.5%) p=0.712
Grade 1+2 (n=205)	32 (15.6%)	173 (84.4%)	9 (4.4%)	196 (95.6%)	16 (7.8%)	189 (92.2%)
Grade 3 (n=76)	20 (26.3%) OR=0.518	56 (73.7%) **p=0.042**	10 (13.2%) OR=3.3	66 (86.8%) **p=0.013**	10 (13.2%) OR=7.79	66 (86.8%) p=0.174
Right sided CRC (n=112)	29 (25.9%)	83 (74.1%)	11 (9.8%)	101 (90.2%)	13 (11.6%)	99 (88.4%)
Left sided CRC (n=173)	25 (14.5%) OR=2.068	148 (85.5%) **p=0.017**	8 (4.6%) OR=0.445	165 (90.2%) **p=0.093**	15 (8.7%) OR=0.723	158 (91.3%) p=0.417

Table 3. Results of logistic regression, significant and borderline insignificant associations are in bold, the main findings of the study are in bold and underlined—association of CK7 expression and SATB2 loss.

or signet ring cell morphology, and advanced UICC stages. CK7 expression was significant predictor of dismal prognosis independently from patient´s age and tumor grade in the multivariate analysis. Both strongly associated SATB2 low status and CK7 expression were insignificant predictors in the multivariate analysis if adjusted on CK7 and SATB2, respectively. The association of SATB2 loss with PD-L1 expression was borderline insignificant (p = 0.066) in the entire cohort and insignificant after exclusion of pretreated patients. PD-L1 expression was a borderline insignificant negative OS predictor without predictive value for CSS; PD-L1 expression was associated with MMR-deficient status; strong PD-L1 expression (> 49% of tumor cells) was a significant poor OS predictor but with no impact on CSS in both analyses.
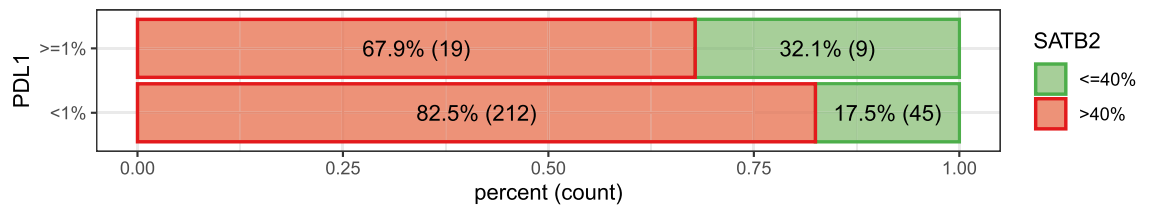
Figure 5. Bar chart documenting borderline insignificant enrichment of SATB2 negative CRCs with PD-L1 expression in ≥ 1% tumor cells (logistic regression, p = 0.066).
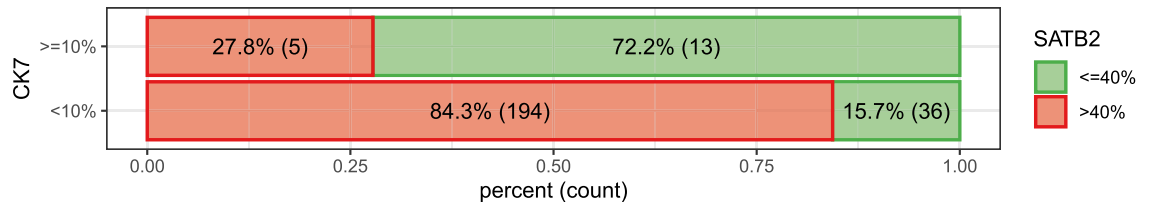
Figure 6. Bar chart documenting significant enrichment of SATB2 negative CRCs with CK7 expression in the neoadjuvant therapy-naïve cohort (logistic regression, p < 0.001).

Discussion

SATB2 is widely used by pathologists as a sensitive and specific marker of a colorectal origin in adenocarcinomas. As expressed in the majority of CRCs, loss of SATB2 has been identified as a negative prognostic marker in several studies. Wang et al. described an association between SATB2 low expression and the presence of lymph node and distant metastases, advanced Dukes' stage, shorter overall, and disease-free survival in CRC[15]. Mezheyeuski et al. linked strong SATB2 expression to left-sided tumor localization, low-grade, non-mutated BRAF status, longer overall survival, and better responsiveness to chemotherapy[17]. Schmitt et al. recently compared the negative prognostic roles of SATB2 loss and caudal type homeobox transcription factor 2 (CDX2) loss; the authors found a negative prognostic relevance for SATB2 loss in both univariate and multivariate analyses and confirmed its prognostic superiority compared to the loss of CDX2[18]. Eberhart et al. described SATB2 expression as an independent favorable prognostic marker in CRC and as a predictor of response to adjuvant and neoadjuvant chemotherapy[16]. Relative to SATB2 being a response to chemotherapy predictor, the favorable value of its expression in CRC[16] sharply contrasts with its negative therapy-predictive value in different cancers, e.g., head and neck squamous carcinoma[61]. This contradiction illustrates the complex role of SATB2 in cancer biology, where it has been primarily described as a transcription factor in craniofacial embryogenesis[62]. In CRC, SATB2 should be regarded as a differentiation marker and, thus, as a beneficial feature of a low-risk phenotype since it may act as an oncogene in non-colonic cell populations. SATB1 is a member of the same protein family that has been described as a cancer promoter in breast tumors and has been linked to poor prognosis[63]. Opposite of SATB2, SATB1 has been linked to rectal cancer progression[64] but without survival analysis.

Eberhart et al. documented a significant association between SATB2 expression and microsatellite stability[16]. Our data suggest a slight enrichment of MMR-deficient status in SATB2 low tumors, although the difference did not reach statistical significance. Similarly, the loss of widely known colorectal differentiation markers CK20 and CDX2 has been described to be more frequent in MMR-deficient tumors[24]. This may be explained by the genomic instability and high mutation loads in MSI CRC, which lead to the loss of intestinal markers.

Our data are in line with well-established research that led to the identification of SATB2 loss as a robust marker of a poor prognosis in CRC. We found significantly shorter OS and CSS, an inclination toward advanced UICC stage, right-sided localization, and variant (mucinous or signet ring cell) morphology. The prognostic value of SATB2 loss was an age- and stage-independent negative predictor in the multivariate analysis. As a novel finding, we identified a strong association between SATB2 loss with CK7 and a borderline insignificant (p = 0.066) association with PD-L1 expression. Both CK7 and PD-L1 expression were negative prognostic indicators as well. We believe that SATB2 and CK7 immunohistochemistry will allow for the identification of particularly aggressive CRC phenotypes.

CK7, its negative prognostic role, and association with traditional factors indicating a poor prognosis (particularly stage) have been documented in several studies, including ours[26,27,32,34,35,57]. Some recent studies shared the methodology in terms of 10% positivity cut-off value to regard the tumor CK7 positive[34,35,57]. According to Fei et al.[35] and Kirchner et al.[65], CK7 expression may be regarded as a hallmark of retro-differentiation or dedifferentiation related to the re-acquiring of a fetal phenotype, which is linked to epithelial-mesenchymal transition that gives the tumor a capacity to metastasize. Kirchner et al. described a lack of glandular differentiation and a primitive duct-like morphology in gastric mucosa expressing CK7[65]. All these facts characterize CK7 expression in CRC as a dedifferentiation marker linked to more aggressive behavior.

PD-L1 has been widely studied in various types of cancer, and multiple immunohistochemistry assays have been approved as complementary diagnostics for patients with non-small cell lung cancer (NSCLC), melanoma, urothelial bladder cancer, esophageal, and gastric cancer[66,67]. In these tumors, PD-L1 is regarded as the target

molecule of immune-checkpoint therapy (i.e., nivolumab, pembrolizumab), which significantly improves patient survival. Nevertheless, despite the wide use of PD-L1 expression, which has been shown by immunohistochemistry as a biomarker for PD-1/PD-L1 blockade in many types of cancer, there are growing concerns regarding its true predictability[68]. In CRC, the US Food and Drug Administration approved nivolumab as a treatment modality in MMR-deficient/MSI-H metastatic CRC refractory to fluoropyrimidine, oxaliplatin, and irinotecan[69], although PD-L1 has not yet been approved as an immunotherapy predictive marker. Concerning the predictive value of tumor cell PD-L1 expression in CRC, results have been inconsistent in various studies, including meta-analyses. There are at least thirteen studies describing its ability to indicate a poor prognosis[44–56], three studies indicating positive prognostic potential[70–72], and five studies indicating no prognostic potential on CRC survival[73–77]. Among recent meta-analyses, we found two that described a negative impact[56,78] and one describing almost no reliable prognostic value[79].

Our PD-L1 expression analysis in CRC is in line with the ambiguous results of the cited studies: we found a slight (borderline insignificant with p-value = 0.15) negative impact on OS and no impact on CSS, but there was significantly shorter OS in cases that strongly (> 49%) expressed PD-L1 on tumor cell membranes. In terms of CSS, no significant results were obtained. Our CSS analysis failed most likely due to the low number of cases with strong PD-L1 expression.

The contradictory results of various studies can be explained by (1) overall paucity of CRCs expressing PD-L1 even in larger datasets, i.e., usually occurring in < 10% of probands, making it hard to reach statistical significance; (2) different methods of PD-L1 expression assessment either in terms of methodology (tumor proportion score—TPS, combined positive score—CPS, immune cell score—IC), or in terms of different antibodies (e.g., clones 22C3, and SP263), or in terms of different percentage threshold value to regard a tumor as positive. In our study, we used a standardized diagnostic kit (SP263), and TPS was assessed in line with the manufacturer's instructions. The results of our study suggest a minimal prognostic impact for weak PD-L1 expression. Our data clearly document that only membranous PD-L1 expression, in a substantial subset (≥ 50%) of malignant cells, is linked to poor survival. From the biological point of view, this may be explained by PD-L1 on tumor cell surfaces switching off cytotoxic T-lymphocytes, thus enabling the tumor cell to escape anti-tumor immunity. In NSCLC, several studies documented variable effects of therapeutic immune-checkpoint inhibitors (pembrolizumab) depending on the percentage of tumor cells expressing PD-L1 (as assessed through immunohistochemistry and using positivity cut-off values of ≥ 1%[80,81] or ≥ 50%[82]. The results of our research indicate using either higher cut-off value than 1% for indicating a CRC as PD-L1+ for patient prognostic stratification, despite being limited by a small number of patients with CRC expressing PD-L1.

CRC may be categorized into two distinct groups according to molecular genetic pathways of carcinogenesis. The first involves chromosomal instability (CIN) and the second is characterized by microsatellite instability (MSI)[83]. The CIN pathway concerns about 80% of sporadic CRCs that are characterized by frequent allelic imbalance, chromosomal amplifications, deletions and translocations. The minority of CRCs arise due to MSI caused by hereditary or sporadic mutational or epigenetic deactivation of one from the four most common genes involved in the process of DNA reparation coding DNA mismatch repair (MMR) proteins (MLH1, MSH2, MSH6, PMS2). MSI CRCs are characterized by failed DNA repair resulting in high load of tumor mutation burden. As a result of the high mutation burden of MSI CRC, the tumor cells present multiple neoantigens (i.e. PD-L1)—in line with this theory and several previous studies[46,76,84], our data document a significant PD-L1 overexpression in MMR-deficient CRC tumor cells.

Although bringing a novel finding (association of two prognostic markers SATB2 and CK7), there are several limitations in our study. Like in all studies using time and cost sparing TMA technique, the question of tumor heterogeneity cannot be ignored. However, microarray technique showed good correlation with whole section immunohistochemistry in a comparative study, and it was previously validated in CRC for several markers[85]. As described in results section, we obtained very good intercore reliability rate. Interobserver variability in immunohistochemical markers with frequent low expression (particularly PD-L1) may represent a potential pitfall in our study. Nevertheless, PD-L1 showed good interobserver agreement in gastric cancer[86,87], in NSCLC[88], in breast cancer[89], and very recently in multiple tumors[90] in various antibody clones in both TPS and CPS. Although no validation study has been published on CRC, probably due to low number of CRC expressing PD-L1, we can consider PD-L1 clone SP263 a reliable method with good interobserver agreement, as we obtained kappa values around 0.8 documenting substantial interobserver agreement.

As already mentioned, all the three markers examined in our study (SATB2 low status, CK7 expression, and strong PD-L1 positivity) can serve for prognostic stratification of patients suffering from CRCs in all stages. However, the "aggressive CRC phenotype" may be characterized by dedifferentiation, i.e., the loss of the differentiation marker, production of aberrant cytokeratin 7, and eventual high production of a neo-antigen, i.e., PD-L1. The exact link between the three markers is hard to explain and needs further investigation at the molecular level. We suggest that the loss of constitutive colon-specific protein (SATB2) and the gain of aberrant non-colon molecules CK7 and PD-L1, i.e., possibly via genomic instability or a high mutation load, may be involved and needs to be further explored. Our study was limited by the overall paucity of CK7+ and PD-L1+ CRCs; eventual corroboration of these associations in larger patient cohorts will be of great interest.

Data availability

All research data are available in the supplementary (electronic only). If needed, please contact the corresponding author (jan.hrudka@lf3.cuni.cz) to obtain the data from this study.

References

1. World Health Organization. Cancer. https://www.who.int/news-room/fact-sheets/detail/cancer. Accessed Dec 2021.
2. Sung, H. *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249. https://doi.org/10.3322/caac.21660 (2021).
3. FitzPatrick, D. R. *et al.* Identification of SATB2 as the cleft palate gene on 2q32-q33. *Hum. Mol. Genet.* **12**, 2491–2501. https://doi.org/10.1093/hmg/ddg248 (2003).
4. Dobreva, G., Dambacher, J. & Grosschedl, R. SUMO modification of a novel MAR-binding protein, SATB2, modulates immunoglobulin mu gene expression. *Genes. Dev.* **17**, 3048–3061. https://doi.org/10.1101/gad.1153003 (2003).
5. Britanova, O. *et al.* Satb2 haploinsufficiency phenocopies 2q32-q33 deletions, whereas loss suggests a fundamental role in the coordination of jaw development. *Am. J. Hum. Genet.* **79**, 668–678. https://doi.org/10.1086/508214 (2006).
6. Gu, W. *et al.* SATB2 preserves colon stem cell identity and mediates ileum-colon conversion via enhancer remodeling. *Cell Stem Cell* **29**, 101–115. https://doi.org/10.1016/j.stem.2021.09.004 (2022).
7. Conner, J. R. & Hornick, J. L. SATB2 is a novel marker of osteoblastic differentiation in bone and soft tissue tumours. *Histopathology* **63**, 36–49. https://doi.org/10.1111/his.12138 (2013).
8. Owosho, A. A. *et al.* SATB2 and MDM2 immunoexpression and diagnostic role in primary osteosarcomas of the jaw. *Dent. J. (Basel).* https://doi.org/10.3390/dj10010004 (2021).
9. Magnusson, K. *et al.* SATB2 in combination with cytokeratin 20 identifies over 95% of all colorectal carcinomas. *Am. J. Surg. Pathol.* **35**, 937–948. https://doi.org/10.1097/PAS.0b013e31821c3dae (2011).
10. Dragomir, A., de Wit, M., Johansson, C., Uhlen, M. & Pontén, F. The role of SATB2 as a diagnostic marker for tumors of colorectal origin: Results of a pathology-based clinical prospective study. *Am. J. Clin. Pathol.* **141**, 630–638. https://doi.org/10.1309/AJCPWW2URZ9JKQJU (2014).
11. Lin, F. *et al.* Cadherin-17 and SATB2 are sensitive and specific immunomarkers for medullary carcinoma of the large intestine. *Arch. Pathol. Lab. Med.* **138**, 1015–1026. https://doi.org/10.5858/arpa.2013-0452-OA (2014).
12. Berg, K. B. & Schaeffer, D. F. SATB2 as an immunohistochemical marker for colorectal adenocarcinoma: A concise review of benefits and pitfalls. *Arch. Pathol. Lab. Med.* **141**, 1428–1433. https://doi.org/10.5858/arpa.2016-0243-RS (2017).
13. Ma, C., Lowenthal, B. M. & Pai, R. K. SATB2 is superior to CDX2 in distinguishing signet ring cell carcinoma of the upper gastrointestinal tract and lower gastrointestinal tract. *Am. J. Surg. Pathol.* **42**, 1715–1722. https://doi.org/10.1097/PAS.0000000000001159 (2018).
14. Yang, C. *et al.* Diagnostic utility of SATB2 in metastatic krukenberg tumors of the ovary: An immunohistochemical study of 70 cases with comparison to CDX2, CK7, CK20, chromogranin, and synaptophysin. *Am. J. Surg. Pathol.* **42**, 160–171. https://doi.org/10.1097/PAS.0000000000000951 (2018).
15. Wang, S. *et al.* Down-regulated expression of SATB2 is associated with metastasis and poor prognosis in colorectal cancer. *J. Pathol.* **219**, 114–122. https://doi.org/10.1002/path.2575 (2009).
16. Eberhard, J. *et al.* A cohort study of the prognostic and treatment predictive value of SATB2 expression in colorectal cancer. *Br. J. Cancer.* **106**, 931–938. https://doi.org/10.1038/bjc.2012.34 (2012).
17. Mezheyeuski, A. *et al.* Metastatic colorectal carcinomas with high SATB2 expression are associated with better prognosis and response to chemotherapy: A population-based Scandinavian study. *Acta. Oncol.* **59**, 284–290. https://doi.org/10.1080/0284186X.2019.1691258 (2020).
18. Schmitt, M. *et al.* Loss of SATB2 occurs more frequently than CDX2 loss in colorectal carcinoma and identifies particularly aggressive cancers in high-risk subgroups. *Cancers (Basel).* **13**, 6177. https://doi.org/10.3390/cancers13246177 (2021).
19. Kummar, S., Fogarasi, M., Canova, A., Mota, A. & Ciesielski, T. Cytokeratin 7 and 20 staining for the diagnosis of lung and colorectal adenocarcinoma. *Br. J. Cancer.* **86**, 1884–1887. https://doi.org/10.1038/sj.bjc.6600326 (2002).
20. Park, S. Y., Kim, H. S., Hong, E. K. & Kim, W. H. Expression of cytokeratins 7 and 20 in primary carcinomas of the stomach and colorectum and their value in the differential diagnosis of metastatic carcinomas to the ovary. *Hum. Pathol.* **33**, 1078–1085. https://doi.org/10.1053/hupa.2002.129422 (2002).
21. Zhang, P. J., Shah, M., Spiegel, G. W. & Brooks, J. J. Cytokeratin 7 immunoreactivity in rectal adenocarcinomas. *Appl. Immunohistochem. Mol. Morphol.* **11**, 306–310. https://doi.org/10.1097/00129039-200312000-00005 (2003).
22. McGregor, D. K., Wu, T. T., Rashid, A., Luthra, R. & Hamilton, S. R. Reduced expression of cytokeratin 20 in colorectal carcinomas with high levels of microsatellite instability. *Am. J. Surg. Pathol.* **28**, 712–718. https://doi.org/10.1097/01.pas.0000126757.58474.12 (2004).
23. Hernandez, B. Y. *et al.* CK20 and CK7 protein expression in colorectal cancer: demonstration of the utility of a population-based tissue microarray. *Hum. Pathol.* **36**, 275–281. https://doi.org/10.1016/j.humpath.2005.01.013 (2005) (**PMID: 15791572**).
24. Lugli, A., Tzankov, A., Zlobec, I. & Terracciano, L. M. Differential diagnostic and functional role of the multi-marker phenotype CDX2/CK20/CK7 in colorectal cancer stratified by mismatch repair status. *Mod. Pathol.* **21**, 1403–1412. https://doi.org/10.1038/modpathol.2008.117 (2008).
25. Saad, R. S., Silverman, J. F., Khalifa, M. A. & Rowsell, C. CDX2, cytokeratins 7 and 20 immunoreactivity in rectal adenocarcinoma. *Appl. Immunohistochem. Mol. Morphol.* **17**, 196–201. https://doi.org/10.1097/PAI.0b013e31819268f2 (2009).
26. Bayrak, R., Yenidünya, S. & Haltas, H. Cytokeratin 7 and cytokeratin 20 expression in colorectal adenocarcinomas. *Pathol. Res. Pract.* **207**, 156–160. https://doi.org/10.1186/1746-1596-7-9 (2011).
27. Harbaum, L. *et al.* Keratin 7 expression in colorectal cancer—Freak of nature or significant finding?. *Histopathology* **59**, 225–234. https://doi.org/10.1111/j.1365-2559.2011.03694.x (2011).
28. Bayrak, R., Haltas, H. & Yenidunya, S. The value of CDX2 and cytokeratins 7 and 20 expression in differentiating colorectal adenocarcinomas from extraintestinal gastrointestinal adenocarcinomas: Cytokeratin 7−/20+ phenotype is more specific than CDX2 antibody. *Diagn. Pathol.* **7**, 9. https://doi.org/10.1186/1746-1596-7-9 (2012).
29. Gurzu, S. & Jung, I. Aberrant pattern of the cytokeratin 7/cytokeratin 20 immunophenotype in colorectal adenocarcinomas with BRAF mutations. *Pathol. Res. Pract.* **208**, 163–166. https://doi.org/10.1016/j.prp.2012.01.003 (2012).
30. Yamagishi, H. *et al.* Aberrant cytokeratin expression as a possible prognostic predictor in poorly differentiated colorectal carcinoma. *J. Gastroenterol. Hepatol.* **28**, 1815–1822. https://doi.org/10.1111/jgh.12319 (2013).
31. Landau, M. S., Kuan, S. F., Chiosea, S. & Pai, R. K. BRAF-mutated microsatellite stable colorectal carcinoma: An aggressive adenocarcinoma with reduced CDX2 and increased cytokeratin 7 immunohistochemical expression. *Hum. Pathol.* **45**, 1704–1712. https://doi.org/10.1016/j.humpath.2014.04.008 (2014).
32. Droy-Dupré, L. *et al.* Hierarchical clustering identifies a subgroup of colonic adenocarcinomas expressing crypt-like differentiation markers, associated with MSS status and better prognosis. *Virchows Arch.* **466**, 383–391. https://doi.org/10.1007/s00428-015-1724-9 (2015).
33. Al-Maghrabi, J., Emam, E. & Gomaa, W. Immunohistochemical staining of cytokeratin 20 and cytokeratin 7 in colorectal carcinomas: Four different immunostaining profiles. *Saudi J. Gastroenterol.* **24**, 129–134. https://doi.org/10.4103/sjg.SJG_465_17 (2018).
34. Loupakis, F. *et al.* CK7 and consensus molecular subtypes as major prognosticators in V600EBRAF mutated metastatic colorectal cancer. *Br. J. Cancer.* **121**, 593–599. https://doi.org/10.1038/s41416-019-0560-0 (2019).
35. Fei, F. *et al.* CK7 expression associates with the location, differentiation, lymph node metastasis, and the Dukes' stage of primary colorectal cancers. *J. Cancer.* **10**, 2510–2519. https://doi.org/10.7150/jca.29397 (2019).

36. Tamura, T. *et al.* Programmed death-1 ligand-1 (PDL1) expression is associated with the prognosis of patients with stage II/III gastric cancer. *Anticancer Res.* **35**, 5369–5376 (2015).
37. Zhang, L. *et al.* Programmed cell death ligand 1 (PD-L1) expression on gastric cancer and its relationship with clinicopathologic factors. *Int. J. Clin. Exp. Pathol.* **8**, 11084–11091 (2015).
38. Eto, S. *et al.* Programmed cell death protein 1 expression is an independent prognostic factor in gastric cancer after curative resection. *Gastric Cancer* **19**, 466–471. https://doi.org/10.1007/s10120-015-0519-7 (2016).
39. Gu, L. *et al.* PD-L1 and gastric cancer prognosis: A systematic review and meta-analysis. *PLoS ONE* **12**, e0182692. https://doi.org/10.1371/journal.pone.0182692 (2017).
40. Tanaka, K. *et al.* Negative influence of programmed death-1-ligands on the survival of esophageal cancer patients treated with chemotherapy. *Cancer Sci.* **107**, 726–733. https://doi.org/10.1111/cas.12938 (2016).
41. Yagi, T. *et al.* PD-L1 expression, tumor-infiltrating lymphocytes, and clinical outcome in patients with surgically resected esophageal cancer. *Ann. Surg.* **269**, 471–478. https://doi.org/10.1097/SLA.0000000000002616 (2019).
42. Luchini, C. *et al.* PD-1, PD-L1, and CD163 in pancreatic undifferentiated carcinoma with osteoclast-like giant cells: expression patterns and clinical implications. *Hum. Pathol.* **81**, 157–165. https://doi.org/10.1016/j.humpath.2018.07.006 (2018).
43. Hrudka, J. *et al.* Negative prognostic impact of PD-L1 expression in tumor cells of undifferentiated (anaplastic) carcinoma with osteoclast-like giant cells of the pancreas: Study of 13 cases comparing ductal pancreatic carcinoma and review of the literature. *Virchows Arch.* **477**, 687–696. https://doi.org/10.1007/s00428-020-02830-8 (2020).
44. Shi, S. J. *et al.* B7–H1 expression is associated with poor prognosis in colorectal carcinoma and regulates the proliferation and invasion of HCT116 colorectal cancer cells. *PLoS ONE* **8**, e76012. https://doi.org/10.1371/journal.pone.0076012 (2013).
45. Zhu, H. *et al.* Clinical significance of programmed death ligand-1 (PD-L1) in colorectal serrated adenocarcinoma. *Int. J. Clin. Exp. Pathol.* **8**, 9351–9359 (2015).
46. Lee, L. H. *et al.* Patterns and prognostic relevance of PD-1 and PD-L1 expression in colorectal carcinoma. *Mod. Pathol.* **29**, 1433–1442. https://doi.org/10.1038/modpathol.2016.139 (2016).
47. Rosenbaum, M. W., Bledsoe, J. R., Morales-Oyarvide, V., Huynh, T. G. & Mino-Kenudson, M. PD-L1 expression in colorectal cancer is associated with microsatellite instability, BRAF mutation, medullary morphology and cytotoxic tumor-infiltrating lymphocytes. *Mod. Pathol.* **29**, 1104–1112. https://doi.org/10.1038/modpathol.2016.95 (2016).
48. Saigusa, S. *et al.* Implication of programmed cell death ligand 1 expression in tumor recurrence and prognosis in rectal cancer with neoadjuvant chemoradiotherapy. *Int. J. Clin. Oncol.* **21**, 946–952. https://doi.org/10.1007/s10147-016-0962-4 (2016).
49. Enkhbat, T. *et al.* Programmed cell death ligand 1 expression is an independent prognostic factor in colorectal cancer. *Anticancer Res.* **38**, 3367–3373. https://doi.org/10.21873/anticanres.12603 (2018).
50. Koganemaru, S. *et al.* Prognostic value of programmed death-ligand 1 expression in patients with stage III colorectal cancer. *Cancer Sci.* **108**, 853–858. https://doi.org/10.1111/cas.13229 (2017).
51. Shao, L. *et al.* Tumor cell PD-L1 predicts poor local control for rectal cancer patients following neoadjuvant radiotherapy. *Cancer Manag. Res.* **9**, 249–258. https://doi.org/10.2147/CMAR.S139889 (2017).
52. Lee, K. S. *et al.* Programmed cell death ligand-1 protein expression and CD274/PD-L1 gene amplification in colorectal cancer: Implications for prognosis. *Cancer Sci.* **109**, 2957–2969. https://doi.org/10.1111/cas.13716 (2018).
53. Ho, H. L. *et al.* PD-L1 is a double-edged sword in colorectal cancer: the prognostic value of PD-L1 depends on the cell type expressing PD-L1. *J. Cancer Res. Clin. Oncol.* **145**, 1785–1794. https://doi.org/10.1007/s00432-019-02942-y (2019).
54. Wu, Z. *et al.* Prognostic impact of adenosine receptor 2 (A2aR) and programmed cell death ligand 1 (PD-L1) expression in colorectal cancer. *Biomed. Res. Int.* **2019**, 8014627. https://doi.org/10.1155/2019/8014627 (2019).
55. Chen, Y. L., Wang, G. X., Lin, B. A. & Huang, J. S. MicroRNA-93-5p expression in tumor tissue and its tumor suppressor function via targeting programmed death ligand-1 in colorectal cancer. *Cell Biol. Int.* **44**, 1224–1236. https://doi.org/10.1002/cbin.11323 (2020).
56. Wang, S. *et al.* Clinicopathological and prognostic significance of PD-L1 expression in colorectal cancer: A meta-analysis. *Int. J. Colorectal. Dis.* **36**, 117–130. https://doi.org/10.1007/s00384-020-03734-4 (2021).
57. Hrudka, J., Fišerová, H., Jelínková, K., Matěj, R. & Waldauf, P. Cytokeratin 7 expression as a predictor of an unfavorable prognosis in colorectal carcinoma. *Sci. Rep.* **11**, 17863. https://doi.org/10.1038/s41598-021-97480-4 (2021).
58. Brierley, J. D., Gospodarowicz, M. K. & Wittekind, C. *TNM Classification of Malignant Tumors* 8th edn, 73–76 (Wiley Blackwell, 2017).
59. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, (Vienna, Austria 2019). https://www.R-project.org/. Accessed Dec 2021.
60. Therneau, T. A Package for Survival Analysis in R. version 3.1–7 (2015). https://CRAN.R-project.org/package=survival. Accessed Dec 2021.
61. Chung, J. *et al.* SATB2 augments ΔNp63α in head and neck squamous cell carcinoma. *EMBO Rep.* **11**, 777–783. https://doi.org/10.1038/embor.2010.125 (2010).
62. Dobreva, G. *et al.* SATB2 is a multifunctional determinant of craniofacial patterning and osteoblast differentiation. *Cell* **125**, 971–986. https://doi.org/10.1016/j.cell.2006.05.012 (2006).
63. Han, H. J., Russo, J., Kohwi, Y. & Kohwi-Shigematsu, T. SATB1 reprogrammes gene expression to promote breast tumour growth and metastasis. *Nature* **452**, 187–193. https://doi.org/10.1038/nature06781 (2008).
64. Meng, W. J. *et al.* Correlation of SATB1 overexpression with the progression of human rectal cancer. *Int. J. Colorectal. Dis.* **27**, 143–150. https://doi.org/10.1007/s00384-011-1302-9 (2012).
65. Kirchner, T. *et al.* Metaplasia, intraepithelial neoplasia and early cancer of the stomach are related to dedifferentiated epithelial cells defined by cytokeratin-7 expression in gastritis. *Virchows Arch.* **439**, 512–522. https://doi.org/10.1007/s004280100477 (2001).
66. Gong, J., Chehrazi-Raffle, A., Reddi, S. & Salgia, R. Development of PD-1 and PD-L1 inhibitors as a form of cancer immunotherapy: A comprehensive review of registration trials and future considerations. *J. Immunother. Cancer.* **6**, 8. https://doi.org/10.1186/s40425-018-0316-z (2018).
67. Lu, S. *et al.* Comparison of biomarker modalities for predicting response to PD-1/PD-L1 checkpoint blockade: A systematic review and meta-analysis. *JAMA Oncol.* **5**, 1195–1204. https://doi.org/10.1001/jamaoncol.2019.1549 (2019).
68. Balar, A. V. & Weber, J. S. PD-1 and PD-L1 antibodies in cancer: Current status and future directions. *Cancer Immunol. Immunother.* **66**, 551–564. https://doi.org/10.1007/s00262-017-1954-6 (2017).
69. Overman, M. J. *et al.* Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): An open-label, multicentre, phase 2 study. *Lancet Oncol.* **18**, 1182–1191. https://doi.org/10.1016/S1470-2045(17)30422-9 (2017).
70. Li, Y. *et al.* Prognostic impact of programed cell death-1 (PD-1) and PD-ligand 1 (PD-L1) expression in cancer cells and tumor infiltrating lymphocytes in colorectal cancer. *Mol. Cancer.* **15**, 55. https://doi.org/10.1186/s12943-016-0539-x (2016).
71. Calik, I. *et al.* Intratumoral cytotoxic T-lymphocyte density and PD-L1 expression are prognostic biomarkers for patients with colorectal cancer. *Medicina (Kaunas).* **55**, 723. https://doi.org/10.3390/medicina55110723 (2019).
72. Chen, T. W. *et al.* Prognostic relevance of programmed cell death-ligand 1 expression and CD8+ TILs in rectal cancer patients before and after neoadjuvant chemoradiotherapy. *J. Cancer Res. Clin. Oncol.* **145**, 1043–1053. https://doi.org/10.1007/s00432-019-02874-7 (2019).

73. Lee, K. S. *et al.* Prognostic implication of CD274 (PD-L1) protein expression in tumor-infiltrating immune cells for microsatellite unstable and stable colorectal cancer. *Cancer Immunol. Immunother.* **66**, 927–939. https://doi.org/10.1007/s00262-017-1999-6 (2017).

74. Berntsson, J. *et al.* Expression of programmed cell death protein 1 (PD-1) and its ligand PD-L1 in colorectal cancer: Relationship with sidedness and prognosis. *Oncoimmunology.* **7**, e1465165. https://doi.org/10.1080/2162402X.2018.1465165 (2018).

75. Ogura, A. *et al.* Pattern of programmed cell death-ligand 1 expression and CD8-positive T-cell infiltration before and after chemo-radiotherapy in rectal cancer. *Eur. J. Cancer.* **91**, 11–20. https://doi.org/10.1016/j.ejca.2017.12.005 (2018).

76. Eriksen, A. C. *et al.* Programmed Death Ligand-1 expression in stage II colon cancer—Experiences from a nationwide popula-tionbased cohort. *BMC Cancer* **19**, 142. https://doi.org/10.1186/s12885-019-5345-6 (2019).

77. Wyss, J. *et al.* Stromal PD-1/PD-L1 expression predicts outcome in colon cancer patients. *Clin. Colorectal Cancer.* **18**, e20–e38. https://doi.org/10.1016/j.clcc.2018.09.007 (2019).

78. Shen, Z., Gu, L., Mao, D., Chen, M. & Jin, R. Clinicopathological and prognostic significance of PD-L1 expression in colorectal cancer: A systematic review and meta-analysis. *World J. Surg. Oncol.* **17**, 4. https://doi.org/10.1186/s12957-018-1544-x (2019).

79. Alexander, P. G., McMillan, D. C. & Park, J. H. A meta-analysis of CD274 (PD-L1) assessment and prognosis in colorectal cancer and its role in predicting response to anti-PD-1 therapy. *Crit. Rev. Oncol. Hematol.* **157**, 103147. https://doi.org/10.1016/j.critrevonc.2020.103147 (2021).

80. Sul, J. *et al.* FDA approval summary: Pembrolizumab for the treatment of patients with metastatic non-small cell lung cancer whose tumors express programmed death-ligand 1. *Oncologist.* **21**, 643–650. https://doi.org/10.1634/theoncologist.2015-0498 (2016).

81. Hui, R. *et al.* Long-term OS for patients with advanced NSCLC enrolled in the KEYNOTE-001 study of pembrolizumab (pembro). *J. Clin. Oncol.* **34**, 15 (2016) (**suppl. 9026**).

82. Reck, M. *et al.* Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N. Engl. J. Med.* **375**, 1823–1833. https://doi.org/10.1056/NEJMoa1606774 (2016).

83. Kreidieh, M., Mukherji, D., Temraz, S. & Shamseddine, A. Expanding the scope of immunotherapy in colorectal cancer: Current clinical approaches and future directions. *Biomed. Res. Int.* **2020**, 9037217. https://doi.org/10.1155/2020/9037217 (2020).

84. Wei, X. L. *et al.* The Clinical and biomarker association of programmed death ligand 1 and its spatial heterogeneous expression in colorectal cancer. *J. Cancer.* **9**, 4325–4333. https://doi.org/10.7150/jca.27735 (2018).

85. Jourdan, F. *et al.* Tissue microarray technology: Validation in colorectal carcinoma and analysis of p53, hMLH1, and hMSH2 immunohistochemical expression. *Virchows Arch.* **443**, 115–121. https://doi.org/10.1007/s00428-003-0833-z (2003).

86. Park, Y. *et al.* PD-L1 testing in gastric cancer by the combined positive score of the 22C3 PharmDx and SP263 assay with clinically relevant cut-offs. *Cancer. Res. Treat.* **52**, 661–670. https://doi.org/10.4143/crt.2019.718 (2020).

87. Ahn, S. & Kim, K. M. PD-L1 expression in gastric cancer: interchangeability of 22C3 and 28–8 pharmDx assays for responses to immunotherapy. *Mod. Pathol.* **34**, 1719–1727. https://doi.org/10.1038/s41379-021-00823-9 (2021).

88. De Marchi, P. *et al.* PD-L1 expression by Tumor Proportion Score (TPS) and Combined Positive Score (CPS) are similar in non-small cell lung cancer (NSCLC). *J. Clin. Pathol.* **74**, 735–740. https://doi.org/10.1136/jclinpath-2020-206832 (2021).

89. Noske, A. *et al.* Interassay and interobserver comparability study of four programmed death-ligand 1 (PD-L1) immunohistochemistry assays in triple-negative breast cancer. *Breast* **60**, 238–244. https://doi.org/10.1016/j.breast.2021.11.003 (2021).

90. Nuti, S. *et al.* High interobserver and intraobserver reproducibility among pathologists assessing PD-L1 CPS across multiple indications. *Histopathology* https://doi.org/10.1111/his.14775 (2022).

Acknowledgements

Author contributions

J.H., A.N., I.T., H.F., and K.J. collected clinical and histopathological data. J.H. and R.M. performed the microscopic evaluation for immunohistochemistry. J.H. and R.M. designed the experiment and wrote the paper. P.W. performed the statistical analysis.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-22685-0.

Correspondence and requests for materials should be addressed to J.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.