



**QUEEN'S  
UNIVERSITY  
BELFAST**

## **Immune-derived PD-L1 gene expression defines a subgroup of stage II/III colorectal cancer patients with favorable prognosis that may be harmed by adjuvant chemotherapy**

Dunne, P. D., McArt, D. G., O'Reilly, P. G., Coleman, H. G., Allen, W. L., Loughrey, M., ... Johnston, P. G. (2016). Immune-derived PD-L1 gene expression defines a subgroup of stage II/III colorectal cancer patients with favorable prognosis that may be harmed by adjuvant chemotherapy. *Cancer Immunology Research*. DOI: 10.1158/2326-6066.CIR-15-0302

**Published in:**  
Cancer Immunology Research

**Document Version:**  
Peer reviewed version

**Queen's University Belfast - Research Portal:**  
[Link to publication record in Queen's University Belfast Research Portal](#)

**Publisher rights**  
Copyright © 2016 American Association for Cancer Research

**General rights**  
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**  
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact [openaccess@qub.ac.uk](mailto:openaccess@qub.ac.uk).

**Immune-derived PD-L1 gene expression defines a subgroup of stage II/III colorectal cancer patients with favorable prognosis that may be harmed by adjuvant chemotherapy**

Philip D. Dunne<sup>1</sup>, Darragh G. McArt<sup>1</sup>, Paul G. O'Reilly<sup>1</sup>, Helen G. Coleman<sup>2</sup>, Wendy L. Allen<sup>1</sup>, Maurice Loughrey<sup>1, 3</sup>, Sandra Van Schaeybroeck<sup>1</sup>, Simon McDade<sup>1</sup>, Manuel Salto-Tellez<sup>1</sup>, Daniel B. Longley<sup>1</sup>, Mark Lawler<sup>1\*^</sup>, Patrick G. Johnston<sup>1\*</sup>

<sup>1</sup> Centre for Cancer Research and Cell Biology, Queen's University Belfast, UK

<sup>2</sup> Centre for Public Health, Queen's University Belfast, UK

<sup>3</sup> Department of Histopathology, Royal Victoria Hospital, Belfast Health and Social Care Trust, Belfast, UK

\* Joint Senior Author (Contributed equally to this manuscript)

^ Corresponding Author: mark.lawler@qub.ac.uk

This work was supported by a Programme Grant from Cancer Research UK (C212/A13721) and Stratified Medicine Programme grant from MRC/CRUK MRM016587/1

Running Title: Significance of PD-L1 gene expression in colorectal cancer

Key Words: Colorectal cancer; adjuvant disease; PD-L1 gene expression; transcriptomics; PD-1 checkpoint

Conflict of Interest Disclosure: **PDD**: None; **DGMcA**: None; **POR**: None; **HGC**: None; **WLA**: None; **MBL**: None; **SVS**: None; **SMcD**: None; **MST**: None; **DBL**: None; **ML**: None; **PGJ**: Previous Founder and Shareholder of Almac Diagnostics; CV6 Therapeutics: Expert Advisor and Shareholder; Chugai Pharmaceuticals: Consultant.

**Abstract:**

A recent phase 2 study of metastatic colorectal carcinoma (CRC) patients showed that mismatch repair gene status was predictive of clinical response to PD-1-targeting immune checkpoint blockade. Further examination revealed strong correlation between PD-L1 protein expression and microsatellite instability (MSI) in stage IV CRC, suggesting that the amount of PD-L1 protein expression could identify late stage patients who may benefit from immunotherapy. To assess whether the clinical associations between PD-L1 gene expression and MSI identified in metastatic CRC are also present in stage II/III CRC, we used *in silico* analysis to elucidate the cell types expressing the PD-L1 gene. We found a significant association of PD-L1 gene expression with MSI in early stage CRC ( $P < 0.001$ ) and show that unlike in non-CRC tumors, PD-L1 is derived predominantly from the immune infiltrate. We demonstrate that PD-L1 gene expression has positive prognostic value in the adjuvant disease setting (PD-L1<sup>low</sup> v PD-L1<sup>high</sup> HR = 9.09; CI, 2.11–39.10). PD-L1 gene expression had predictive value, as patients with high PD-L1 expression appear to be harmed by standard-of-care treatment (HR = 4.95; CI, 1.10–22.35). Building on the promising results from the metastatic CRC PD-1–targeting trial, we provide compelling evidence that PD-L1<sup>high</sup>/MSI/immune<sup>high</sup> stage II/III CRC patients should not receive standard chemotherapy. This conclusion supports the rationale to clinically evaluate this patient subgroup for PD-1 blockade treatment.

**Introduction:**

Stroma-derived factors have long been known to influence cancer progression (1), and the importance of the microenvironment for molecular classification of colorectal cancer (CRC) tumors has been confirmed (2, 3). These studies highlight the influence of the non-neoplastic component of the tumor on patient prognosis. Expression of PD-L1, the immune checkpoint inhibitor, has been primarily detected on the surface of epithelial neoplastic cells in a number of cancers; however, in CRC, immunohistochemistry (IHC)-based studies of small cohorts have detected high PD-L1 expression in the stromal and immune compartments (4, 5). Although upregulation of PD-L1 in the tumor microenvironment is a recognized tumor immune- defense mechanism, these findings suggest a different origin for PD-L1 protein expression in CRC.

The mismatch repair system (MMR) helps preserve the fidelity of the genome (6, 7). CRCs which harbor defects in MMR demonstrate high microsatellite instability (MSI) and account for 12-15% of CRCs. MSI tumors are generally defined by their large number of somatic mutations, compared to microsatellite stable (MSS) tumors. These tumors also exhibit heavy peritumoral/intratumoral lymphocytic infiltration, most likely due to a large number of mutated antigenic epitopes at the cell surface; this has been previously correlated with good prognosis in early stage disease (8).

A number of adjuvant trials have questioned the value of chemotherapy for defined CRC molecular subtypes in early stage disease, with some studies suggesting potential harm, particularly to the overall good prognosis MSI group (9). Although preclinical data inferred that MSI tumors would not respond to 5FU-based treatment (10), the first large adjuvant study published using MSI status to stratify patients revealed that patients with MSI tumors did benefit from addition of chemotherapy following surgery (11). However, 11 subsequent studies have shown no benefit from 5FU-based treatment for patients with MSI CRC in the adjuvant setting (9).

Recent clinical studies in melanoma, renal cell carcinoma, and non–small cell lung cancer have reported significant positive responses to PD-1 checkpoint targeting (12). In contrast, results in CRC have been disappointing (12). Interrogation of factors associated with response to PD-1 blockade suggested that MSI status was a predictor of response, underpinning a phase 2 clinical trial (13) in patients with metastatic CRC stratified by MSI status. The disease control rate in this study was 90% (CI: 55-100) for patients with MSI tumors and 11% for patients with MSS tumors (CI: 1-35), supporting the hypothesis of strong predictive value of MSI status for positive response to PD-1 blockade in advanced CRC. In addition, many CD8<sup>+</sup> infiltrating T lymphocytes were detected at the invasive front regions of these tumors, corresponding to increased PD-L1 expression levels at the tumor margin.

Effective use of PD-1–targeting checkpoint inhibitors requires reliable biomarkers/companion diagnostics. Immunohistochemical detection of PD-L1 is currently confounded by technical variation, fluctuations in detection levels, and reproducible cut-off thresholds. Most importantly, marked intratumoral staining heterogeneity greatly hinders reproducibility of immunohistochemical scoring systems. Thus, alternative approaches for assessing PD-L1 are required.

The extremely promising results in stage IV disease prompted us to evaluate the potential for immune checkpoint–targeting in the adjuvant setting. We performed extensive bioinformatics analyses employing well-characterized independent transcriptional profiling datasets to determine (i) whether PD-L1 gene expression is associated with specific cell lineage compartments within the CRC tumor microenvironment; (ii) ability to stratify patients in early stage CRC using PD-L1 gene expression and determine its association with MSI status/immune infiltration; and (iii) clinical relevance of PD-L1 gene expression to both prognosis and potential for benefit from adjuvant chemotherapy.

## **Materials and Methods**

### ***Independent datasets***

Gene expression profiles from independent CRC datasets were downloaded from NCBI Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>) under accession numbers GSE14333, GSE35602, GSE13294, GSE39396 and GSE39582. GSE14333 contains microarray profiles of surgically resected specimens in 290 CRC patients; 185 have additional treatment and survival data and are employed in this study. GSE35602 contains microarray profiles separately profiled from micro-dissected stroma or epithelium regions from thirteen CRC tissues. GSE13294 contains 155 CRC microarray profiles, with MSI status, from surgically resected specimens. GSE39396 contains microarray profiles from fresh colorectal specimens where FACS has been used to divide cells into specific endothelial (CD45<sup>+</sup>EPCAM<sup>-</sup>CD31<sup>-</sup>FAP<sup>-</sup>), epithelial (CD45<sup>-</sup>EPCAM<sup>+</sup>CD31<sup>-</sup>FAP<sup>-</sup>), leukocyte (CD45<sup>-</sup>EPCAM<sup>-</sup>CD31<sup>+</sup>FAP<sup>-</sup>) and fibroblast (CD45<sup>-</sup>EPCAM<sup>-</sup>CD31<sup>-</sup>FAP<sup>+</sup>) populations prior to microarray profiling. GSE39582 contains 566 stage I-IV profiles from a large CRC series, of which the stage II/III profiles were selected for analysis.

### ***Transcriptional analysis***

Partek Genomics Suite was used for independent dataset analysis. For the purpose of clustering, data matrices were standardized to the median value of probe sets expression. Standardization of the data allows for comparison of expression levels for different probe sets. Following standardization, 2-dimensional hierarchical clustering was performed (samples x probe sets/genes). Euclidean distance was used to calculate the distance matrix, a multidimensional matrix representing the distance from each data point (probe set-sample pair) to all the other data points. Ward's linkage method was subsequently applied to join samples and genes together, with the minimum variance, to find compact clusters based on the calculated distance matrix.

**Statistics**

Median and tertile stratification was performed on GSE13294 by calculation of mean expression values from both CD274 probe-sets. These values were then classified as high and low based on 77:78 sample distributions or as high, medium, and low based on 52:51:52 sample distributions. Student *t* tests and Fishers exact tests were carried out using GraphPad Prism version 5 for Windows, GraphPad Software.

**Survival analysis**

Survival curves, comparing expression and treatment subgroups were estimated with the Kaplan–Meier method and compared by the log-rank test, using GraphPad Prism version 5 for Windows, GraphPad Software. Cox Proportional Hazards analysis, using Stata version 11.2, was applied to evaluate recurrence-free survival according to PD-L1 gene expression levels within the indicated subgroups, prior to and after adjustment for age, sex, tumor stage and location, and receipt of adjuvant treatment. Categorical and continuous variables were compared between individuals within the overall cohort and also in PD-L1<sup>high</sup>/MSI versus PD-L1<sup>low</sup>/MSS tumors using chi-squared tests and t-tests, respectively.

**Results:*****PD-L1 gene expression associated with immune component of TME***

Unlike other cancers, PD-L1 expression in the colorectal tumor microenvironment (TME) may not be exclusive to epithelial tumor cells (4). To further investigate this preliminary finding, we utilized 2 CRC gene expression transcriptomic datasets, derived from samples either laser-microdissected to purify stromal and epithelial regions (GSE35602), or separated into epithelial, leukocyte, endothelial or fibroblast components using FACS (GSE39396) prior to microarray profiling of the separated cells. Scatterplot and boxplot assessment of gene expression levels of PD-L1 (CD274), according to region of origin, revealed significantly increased gene expression in the stroma compared to neoplastic epithelial cells in the laser-microdissected dataset (two-tailed Student *t*-test  $P < 0.0001$ , Fig. 1A and C). In the FACS-derived microarray dataset, significantly higher levels of PD-L1 gene expression were observed in tumor-associated leukocytes compared to epithelial cells (two-tailed Student *t*-test  $P = 0.0071$ , Fig. 1B and D).

In order to assess the purity of the microdissection of the samples from the GSE35602 dataset, we utilized a previously published gene expression list derived exclusively from non-epithelial cells. These 213 genes were defined from 3 signatures specifically expressed by cancer-associated fibroblasts (CAFs) ( $n = 131$ ), leukocytes ( $n = 47$ ) and endothelial cells ( $n = 35$ ) (3). Using hierarchical clustering and principal component analyses, we robustly identified 2 distinct groups in the GSE35602 dataset that corresponded to each “region of origin” with 100% accuracy (Supplementary Fig. S1A and C). When the same approach was applied to GSE39396 (FACS-sorted dataset), region of origin was again identified with 100% accuracy (Supplementary Fig. S1B and D).

In addition to PD-L1, expression of other immune therapy targets (CTLA4, LAG3, and IDO1) has also been reported to be upregulated in immune infiltrating cells of MSI tumors compared to MSS tumors(4), although their gene expression levels in individual cell



compartments have not been assessed. Using region-specific and cell-specific gene expression profiles, we found CTLA4 (6.3 fold,  $P < 0.001$ ), LAG3 (4.5 fold,  $P < 0.001$ ), PD-L1 (4.2 fold,  $P < 0.001$ ), and IDO1 (9.2 fold,  $P < 0.001$ ) are all elevated in stroma compared to epithelium in CRC tumor samples (Fig. 1D). Whereas IDO1 expression was elevated in all stromal compartments compared to epithelial cells, the elevated expression of CTLA4, LAG3, and PDL1 was confined to the leukocyte-specific compartment (Fig. 1E and F). We also confirm that expression of interferon- $\gamma$  (IFNG) is specific to the immune-derived compartment (Supplementary Fig. S1E), consistent with previous findings (14).

Collectively, these analyses provide compelling evidence that PD-L1 in colorectal tumors is predominantly derived from infiltrating immune cells rather than neoplastic epithelial cells.

#### ***Association between MSI, immune infiltration, and PD-L1 gene expression***

To assess whether the clinical associations between PD-L1 gene expression and MSI identified in metastatic CRC are also present in stage II/III CRC, we evaluated transcriptomic data derived from a cohort of 155 patients (GSE13294), enriched to contain approximately equal numbers of MSI ( $n = 78$ ) and MSS ( $n = 77$ ) tumors. Using a median- and tertile-stratification approach based on mean PD-L1 gene expression, we differentiated samples based on high or low PD-L1 gene expression (Fig. 2A) and high, medium, or low gene expression (supplementary Fig. S2A). Using a median-stratified approach, PD-L1 gene expression is significantly associated with MSI status (Fishers exact two-tailed test  $P < 0.0001$ , Fig. 2B). In addition, this PD-L1<sup>high</sup>/MSI-rich subgroup is significantly associated with the 47 leukocyte-specific non-epithelial gene signature (Fishers exact two-tailed test  $P < 0.0001$ , Fig. 2C). When the intermediate group ( $n = 51$ ) is removed from the tertile analysis, we again find a significant correlation between PD-L1 gene expression and both MSI and the leukocyte-specific signature (Supplementary Fig. 2B, C, and D) These results reveal the

relationship between high PD-L1 gene expression, MSI, and immune infiltration in stage II/III disease.

### ***Identification of a subgroup of patients with high PD-L1 gene expression***

Utilizing hierarchical clustering of microarray gene expression profiles from a large stage II/III CRC dataset (GSE39582), and employing Euclidean and Ward metrics, we identified a distinct subgroup of patients with high PD-L1 gene expression relative to the remaining population (Fig. 3A). This PD-L1<sup>high</sup> subgroup accounted for 20% of the overall cohort, which we use as our threshold for all subsequent analyses; further investigation highlighted a strong correlation between elevated PD-L1 gene expression and MSI genotype (Fishers exact two-tailed test  $P < 0.0001$ ; Fig. 3A and B, Supplementary Fig. S3A), further validating our earlier findings (Fig. 2).

Stratification of the data was performed to facilitate an evaluation of the available clinicopathologic factors using two different comparisons; PD-L1<sup>high</sup> gene expression subgroup (PD-L1<sup>high</sup>) versus PD-L1<sup>low</sup> gene expression subgroup (PD-L1<sup>low</sup>) in the entire cohort and MSI versus MSS in the PD-L1<sup>high</sup> subgroup only. Within the entire cohort, individuals with PD-L1<sup>high</sup> tumors did not differ from those with PD-L1<sup>low</sup> tumors in terms of age, sex or stage distribution. PD-L1<sup>high</sup> tumors were less likely to be treated with adjuvant chemotherapy, although this did not reach statistical significant ( $P = 0.07$ ). PD-L1<sup>high</sup> tumors were more likely to be right-sided, MSI, CpG island methylator phenotype (CIMP) positive, chromosome instability (CIN) negative, and protein kinase BRAF mutant than PD-L1<sup>low</sup> tumors, but not p53 or KRAS mutant (Table 1).

Findings from the PD-L1<sup>high</sup> subgroup stratified by MSI or MSS, confirm that MSS/PD-L1<sup>high</sup> tumors were less likely to be right-sided, CIMP<sup>+</sup>, CIN<sup>-</sup> and BRAF mutant than MSI/PD-L1<sup>high</sup> tumors. This analysis again highlighted that whereas MSI status is significantly associated with high PD-L1 gene expression, a subgroup of 13% of MSS tumors were also classified as

PD-L1<sup>high</sup>. Utilizing the previously described leukocyte-specific signature (3) that is solely attributed to the leukocyte compartment of the TME, we found strong overlap between those patients in the PD-L1<sup>high</sup> subgroup and those with a gene expression profile indicative of an increased immune infiltrate (Fishers exact two-tailed test  $P < 0.001$ , Fig. 3C). This finding further confirmed that the PD-L1<sup>high</sup> subgroup, which is significantly enriched for MSI ( $P < 0.001$ ), was also associated with more tumor-infiltrating immune cells, and highlights PD-L1 expression as a robust transcriptional marker for this subgroup. Although we do find significant clinicopathological differences between MSI and MSS in the PD-L1<sup>high</sup> subgroup, in agreement with our data presented in Figs. 1 and 2, it is the biological signature indicative of a large immune infiltration that appears to dictate the level of PD-L1 gene expression.

Further analysis confirmed co-expression and elevated expression of CTLA4, LAG3, and IDO1 in stage II/III tumors samples that have high expression of PD-L1 (Supplementary Fig. S3B), in addition to significant upregulation of IFN $\gamma$  (two-tailed Student  $t$ -test  $P < 0.001$ , Supplementary Fig. S3C).

### ***PD-L1 is a significant positive prognostic marker in early stage disease***

To investigate the clinical relevance of PD-L1 gene expression, we used relapse follow-up data associated with the well-characterized GSE39582 dataset. Patients ( $n = 201$ ) were stratified based on PD-L1 subgroup, stage and treatment. In the untreated stage III population, we found a clear difference in relapse-free survival (RFS) between low and high PD-L1 subgroups, with the PD-L1<sup>low</sup> subgroup having a significantly worse outcome ( $P = 0.0003$ ; HR = 9.09; 95% CI, 2.11–39.10; Fig. 4A and B; Table 2). However, in the treated cohort, the correlation between survival and high PD-L1 expression was lost ( $P = 0.6514$ ; HR = 0.86; 95% CI, 0.45–1.66), suggesting that PD-L1 gene expression also has value for predicting benefit from standard adjuvant chemotherapy (Fig. 4B, Table 2).

To address this question, we performed treatment interaction analyses and found that, whereas patients with low expression of the PD-L1 gene significantly benefit from adjuvant chemotherapy ( $P = 0.0062$ ; HR = 0.49; 95%CI, 0.29–0.83), patients in the PD-L1<sup>high</sup> subgroup have poorer overall survival following treatment ( $P = 0.0208$ ; HR = 4.95; 95%CI, 1.10–22.35; Fig. 4A and B; Table 2). An adjusted analysis for the known confounders and covariates of PD-L1 gene expression (Table 1) again confirmed that PD-L1 gene-expression could be considered as an independent biomarker for patient stratification, as although the prognostic and predictive trend remained the same, the adjusted multivariate significance was lost (Table 2).

In order to confirm these findings in an independent patient cohort, we interrogated a further early stage CRC dataset (GSE14333). Patients ( $n = 185$ ) were again stratified into high and low PD-L1 subgroups in similar proportions as identified using our initial dataset. In Dukes' B patients within this cohort, high PD-L1 gene expression was significantly associated with better disease-free survival (DFS) in the untreated population compared to those who received adjuvant treatment ( $P = 0.0371$ ; HR = 10.18; CI, 1.15–90.14). This trend was also observed in the combined Dukes B/C cohort, but failed to reach significance, most likely due to the small number of patients in this combined cohort compared to the original dataset (Fig. 4C).

These data indicate that TME-derived PD-L1 transcription levels are both a positive prognostic marker for improved relapse/disease-free survival in early stage disease, but importantly also are a negative predictive marker for chemotherapy in the adjuvant setting.

**Discussion:**

Since the FDA's approval of the first immune checkpoint therapy (the CTLA4-specific antibody ipilimumab), a number of clinical trials have demonstrated the potential for targeting this pathway in a variety of cancers. However, immune therapy has had surprisingly little impact in CRC. A recent phase 2 study gave the first indication that PD-1 targeting of CRC in metastatic disease significantly benefited patients with MSI tumors when compared to those with MSS disease (13). This study also indicated that PD-L1 expression (assessed by IHC) was strongly associated with MSI, suggesting that expression of PD-L1 may be a useful predictive biomarker of response to PD-1 immune checkpoint-targeting in this setting.

Using an *in silico* approach, we assessed whether PD-L1 gene expression was associated with MSI in early stage CRC, which would provide a rationale for pursuing PD-1 checkpoint-targeting in the adjuvant disease setting. We found that PD-L1 transcription levels are significantly elevated in the immune cells present in the TME, in agreement with an earlier study in a small patient cohort (4). Given that high immune infiltration can occur in CRC, these findings may explain the difference in response rates to immune-checkpoint targeting in CRC compared to other tumors, e.g. lung, melanoma, where PD-L1 expression has been detected in the membrane of epithelial neoplastic cells. Using a transcriptomic dataset from a 155 CRC patient cohort, enriched to include ~50% MSI/MSS, we found that PD-L1<sup>high</sup> tumors are significantly associated with the MSI genotype. Additionally, a significant correlation between high PD-L1 gene expression and substantial immune cell infiltration was found, further supporting the hypothesis that these patients would benefit from PD-1-targeting agents. This significant association with MSI was also evident in a large well characterized stage II/III clinical cohort, which also confirms that PD-L1 gene expression is significantly associated with right-sided, CIMP and BRAF mutant tumors.

A recent metastatic CRC clinical trial uncovered a subgroup of patients with MSI genotype and high PD-L1 levels using IHC; we identified a distinct subgroup of stage II/III CRCs, this

time identifiable by high PD-L1 transcription levels. While this subgroup was significantly associated with MSI, it was not exclusive to this genotype, with a small number of MSS tumors also being PD-L1<sup>high</sup>. Conversely, a small proportion of MSI patients were classified as PD-L1<sup>low</sup>. These results suggest that while the MSI genotype results in high mutation rates which promote high levels of immune infiltration, the expression of PD-L1, and indeed immune infiltration levels, can also be upregulated by MSI independent mechanisms. Thus, PD-L1 gene expression rather than MSI status may be a more powerful predictive biomarker for response to PD-1/PD-L1 checkpoint inhibition in CRC. Previous studies have shown that patients with MSI CRC generally have a good overall prognosis in early stage disease, however there is still debate as to the benefit of adjuvant chemotherapy in this group (9). Recently, a large meta-analysis concluded that there was no effect of adjuvant treatment for MSI patients, whereas there was a significant benefit in MSS patients (15). Data presented here show for the first time that high PD-L1 transcription levels, which is significantly associated with the MSI genotype, identify a subgroup of patients with a significantly better prognosis in early stage disease. In addition, we also show that this PD-L1<sup>high</sup> subgroup derives no clinical benefit and indeed may be harmed by adjuvant 5-FU-based chemotherapy using an unadjusted analysis.

Although this dataset, and the independent validation set, were not generated from material collected for randomized control trials, nonetheless, our findings on high PD-L1 transcriptional levels have clinical implications above and beyond MSI status alone, for further stratifying patients into those likely to benefit from standard adjuvant chemotherapy and those who may potentially be harmed. Our analyses suggest that the PD-L1<sup>high</sup> subgroup should not be given adjuvant 5-FU based chemotherapy following surgery, whereas patients with low PD-L1 gene expression significantly benefit from adjuvant treatment, in both unadjusted and adjusted models for survival analysis. Moreover, although data presented here strongly indicate that PD-L1<sup>high</sup> patients may not need any systemic therapy, their PD-L1 levels, MSI status, and immune infiltrate levels confirm that it is this

clinical subgroup, (based on the recent clinical trial in the metastatic setting), which may instead benefit from PD-1/PD-L1 immune checkpoint inhibitors, notably in stage III disease. This PD-L1<sup>high</sup> subgroup also displayed increased expression of a number of other immunotherapy targets, namely CTLA4, LAG3, and IDO1. Similar to the findings presented here for PD-L1, expression of each of these targets is confined to the stroma, in particular to the immune compartment, with the exception of the metabolic regulator IDO1, which is found in all stromal compartments. These findings highlight the potential for combination immunotherapies in this immune checkpoint overexpressing subgroup. Although PD-L1 gene expression had significant prognostic and predictive value in 2 independent cohorts, final validation requires transcriptional data, detailed treatment information, and clinical follow up from an independent, well-balanced cohort, enriched for MSI stage II/III CRC patients enrolled in a prospective clinical trial. This type of patient stratification approach is ongoing in current clinical trials (16, 17) that will enable further biomarker-based validation in this setting. Ongoing debate surrounds the definition of a clinically relevant companion diagnostic threshold for assessing PD-L1 protein levels by IHC, to predict benefit from PD-1 blockade. This will only be possible by retrospective outcome-supervised analysis of the tumor tissue from ongoing and completed PD-1/PD-L1 checkpoint inhibition trials in CRC. Matched IHC and microarray profiling from the same tissue would allow the generation of these urgently required thresholds for prospective use.

The transcriptional profiles we have analyzed are representative of primary CRC tumors prior to therapy, and as such provide insight into the cell populations present and their signaling activities during development of the primary tumor. It is conceivable that the majority of these tumors, which we now show have a paucity of immune cells, initiate and develop by circumventing a widespread immune response (18). It is only tumors that have escaped immunosurveillance, allowing development of invasive malignancy and subsequent metastatic spread, which have high relapse rates and poor survival. The small proportion of

tumors that we have identified with an inherently large immune infiltrate [resulting from high numbers of mutated antigenic epitopes due to their MSI status,(8)] can be held in a state of equilibrium by this response (19), despite PD-L1–mediated immune checkpoint activation. This may explain why these tumors have a relatively good prognosis if left untreated (9). In the post-treatment setting, we know that addition of 5-FU based adjuvant therapy following surgery results in loss of tumor infiltrating immune cells (20). Thus, patients with immune-rich tumors would be harmed by exposure to 5-FU, therefore explaining the negative predictive value between high PD-L1 transcription and response to standard-of-care chemotherapy.

In conclusion, data presented here, alongside data from the metastatic trial (13) identify a subgroup that is defined by an underlying biology consisting of increased PD-L1 transcription, MSI genotype, and large immune infiltrates in stage II/III disease. We now demonstrate the prognostic value of PD-L1 gene expression in early stage CRC and highlight the potentially harmful effects of standard-of-care chemotherapy in this clinically relevant and PD-L1–definable subgroup.



## References

1. Jass JR, Morson BC. Reporting colorectal cancer. *J Clin Pathol*. 1987 Sep;40(9):1016-23.
2. Calon A, Lonardo E, Berenguer-Llergo A, Espinet E, Hernando-Momblona X, Iglesias M, et al. Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat Genet*. 2015 Feb 23.
3. Isella C, Terrasi A, Bellomo SE, Petti C, Galatola G, Muratore A, et al. Stromal contribution to the colorectal cancer transcriptome. *Nat Genet*. 2015 Feb 23.
4. Llosa NJ, Cruise M, Tam A, Wicks EC, Hechenbleikner EM, Taube JM, et al. The vigorous immune microenvironment of microsatellite instable colon cancer is balanced by multiple counter-inhibitory checkpoints. *Cancer Discov*. 2015 Jan;5(1):43-51.
5. Taube JM, Klein A, Brahmer JR, Xu H, Pan X, Kim JH, et al. Association of PD-1, PD-1 ligands, and other features of the tumor immune microenvironment with response to anti-PD-1 therapy. *Clin Cancer Res*. 2014 Oct 1;20(19):5064-74.
6. Aaltonen LA, Peltomaki P, Leach FS, Sistonen P, Pylkkanen L, Mecklin JP, et al. Clues to the pathogenesis of familial colorectal cancer. *Science*. 1993 May 7;260(5109):812-6.
7. Thibodeau SN, Bren G, Schaid D. Microsatellite instability in cancer of the proximal colon. *Science*. 1993 May 7;260(5109):816-9.
8. Saeterdal I, Bjorheim J, Lislerud K, Gjertsen MK, Bukholm IK, Olsen OC, et al. Frameshift-mutation-derived peptides as tumor-specific antigens in inherited and spontaneous colorectal cancer. *Proc Natl Acad Sci U S A*. 2001 Nov 6;98(23):13255-60.
9. Boland CR, Goel A. Microsatellite instability in colorectal cancer. *Gastroenterology*. 2010 Jun;138(6):2073,2087.e3.
10. Carethers JM, Chauhan DP, Fink D, Nebel S, Bresalier RS, Howell SB, et al. Mismatch repair proficiency and in vitro response to 5-fluorouracil. *Gastroenterology*. 1999 Jul;117(1):123-31.
11. Elsaleh H, Joseph D, Grieu F, Zeps N, Spry N, Iacopetta B. Association of tumour site and sex with survival benefit from adjuvant chemotherapy in colorectal cancer. *Lancet*. 2000 May 20;355(9217):1745-50.
12. Postow MA, Callahan MK, Wolchok JD. Immune Checkpoint Blockade in Cancer Therapy. *J Clin Oncol*. 2015 Jun 10;33(17):1974-82.
13. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med*. 2015 Jun 25;372(26):2509-20.
14. Bach EA, Aguet M, Schreiber RD. The IFN gamma receptor: a paradigm for cytokine receptor signaling. *Annu Rev Immunol*. 1997;15:563-91.
15. Des Guetz G, Schischmanoff O, Nicolas P, Perret GY, Morere JF, Uzzan B. Does microsatellite instability predict the efficacy of adjuvant chemotherapy in colorectal cancer? A systematic review with meta-analysis. *Eur J Cancer*. 2009 Jul;45(10):1890-6.

16. Lawler M, Kaplan R, Wilson RH, Maughan T, S-CORT Consortium. Changing the Paradigm-Multistage Multiarm Randomized Trials and Stratified Cancer Medicine. *Oncologist*. 2015 Aug;20(8):849-51.
17. Schneider D, Bianchini G, Horgan D, Michiels S, Witjes W, Hills R, et al. Establishing the Evidence Bar for Molecular Diagnostics in Personalised Cancer Care. *Public Health Genomics*. 2015;18(6):349-58.
18. Dunn GP, Bruce AT, Ikeda H, Old LJ, Schreiber RD. Cancer immunoediting: from immunosurveillance to tumor escape. *Nat Immunol*. 2002 Nov;3(11):991-8.
19. Koebel CM, Vermi W, Swann JB, Zerafa N, Rodig SJ, Old LJ, et al. Adaptive immunity maintains occult cancer in an equilibrium state. *Nature*. 2007 Dec 6;450(7171):903-7.
20. Mitchell MS, DeConti RC. Immunosuppression by 5-fluorouracil. *Cancer*. 1970 Oct;26(4):884-9.

## Figure Legends

**Figure 1: PD-L1 gene expression is higher in the stromal immune compartment of the tumor microenvironment.** **A.** Scatterplot indicating higher PD-L1 gene expression in stromal cells compared to the epithelium of 13 microdissected primary tumor samples (GSE35602). **B.** PD-L1 gene expression was higher in the leukocyte population compared to fibroblast, endothelial, and epithelial populations isolated by FACS (GSE39396). **C + D.** Dot plot with associated box and whisker plots representing the mean expression values for PD-L1 in each sample set. 10% lower and 90% upper values are indicated. **E + F.** Euclidean clustering for gene expression profiles of CTLA4, LAG3, PD-L1, and IDO1 in each sample set. Overlay bar indicates the region- or cell-of-origin as used in A–D.

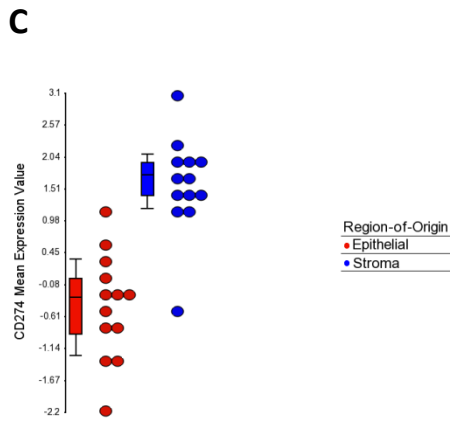
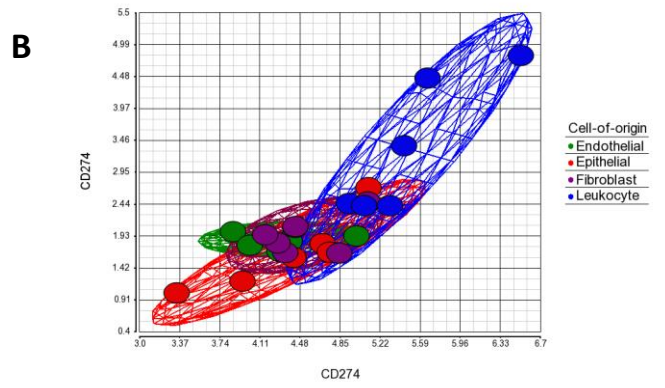
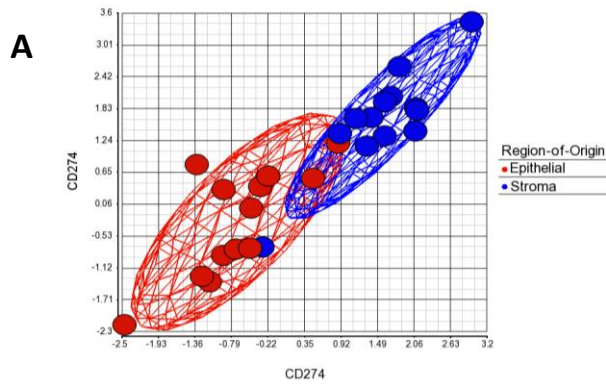
**Figure 2: Association of high PD-L1 gene expression with MSI subtype and substantial immune infiltrate.** **A.** (Left) Dot plot with associated box plots representing the mean gene expression values for PD-L1 in each subgroup. (Right) uniform probability plot of PD-L1 expression values highlight the cutoff between PD-L1<sup>high</sup> and PD-L1<sup>low</sup> subgroups (GSE13294) **B + C.** (Left) Dot plot with associated box plots in PD-L1<sup>high</sup> and PD-L1<sup>low</sup> subgroups further stratified by MSI status (B) and immune infiltrate signature (C). (Right) Fishers exact two-tailed test confirms a statistically significant correlation between high PD-L1 levels and either MSI status (B) or immune infiltrate signature (C). Whiskers on boxplots represent 10% lower and 90% upper values.

**Figure 3: PDL1 gene expression profile in stage II/III CRC.** **A.** (Left) Hierarchical clustering of stage II/III CRC patient cohort (GSE39582) based on expression profiles of PDL1 (CD274) identifies a strong positive subgroup accounting for 20% of the overall population. Overlay bar indicates assigned PD-L1 subgroup (PD-L1<sup>high</sup>, red; PD-L1<sup>low</sup>, blue) and MSI/MSS status (MSI, black; MSS, gray; NA, white). (Right) Scatterplot of PD-L1 gene expression values highlight the positive and negative groups identified in the heatmap. **B.** (Left) Scatterplot of PD-L1 gene expression values according to MSI/MSS status. (Right)

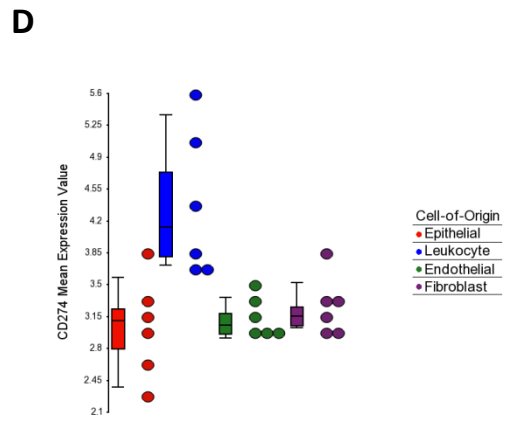
Fishers exact test confirms a statistically significant correlation between high PDL1 transcript amounts and MSI status. **C.** Hierarchical clustering of patient cohort based on leukocyte-specific gene signature stratifies population into 3 groups based on their immune infiltrate content. Overlay bar indicates that the high immune infiltrate subgroup overlaps strongly with the high PDL1 gene expression subgroup.

**Figure 4: PD-L1 expression is a strong positive prognostic and negative predictive marker to chemotherapy in CRC.** **A.** Survival curve using Kaplan-Meier estimation comparing PD-L1 transcript expression in untreated tumors (left) and comparing treated and untreated for PD-L1<sup>high</sup> (right) stage III CRC patients (GSE39582). **B.** Unadjusted hazard ratio analysis of the cohort based on PD-L1 expression and/or receipt of adjuvant treatment. Hazard ratio is plotted on a logarithmic base 10 scale. **C.** Further confirmation of prognostic and predictive value for PD-L1 transcription in Dukes B (left) and combined Dukes B/C (right) in GSE41333.

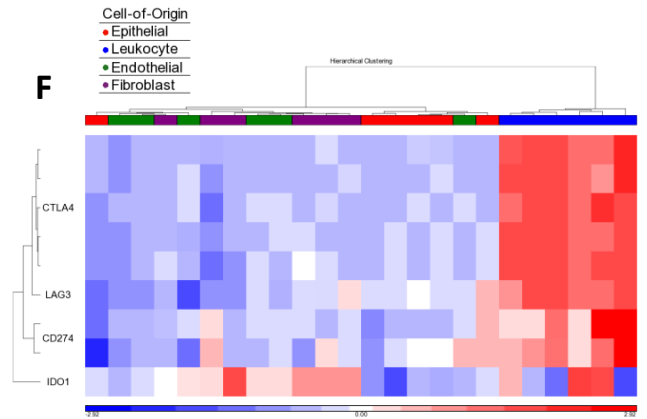
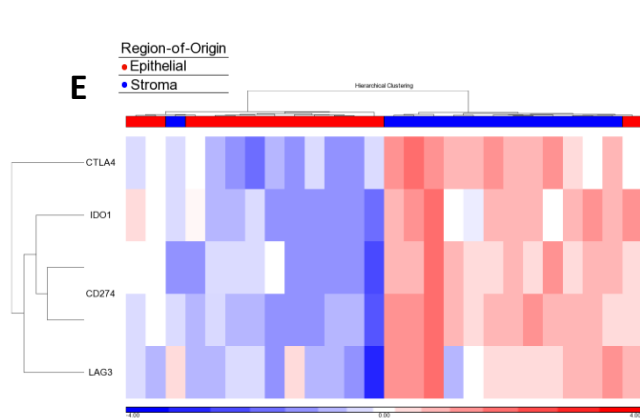
**Figure 1**



Epithelial v Stroma  $P < 0.0001$

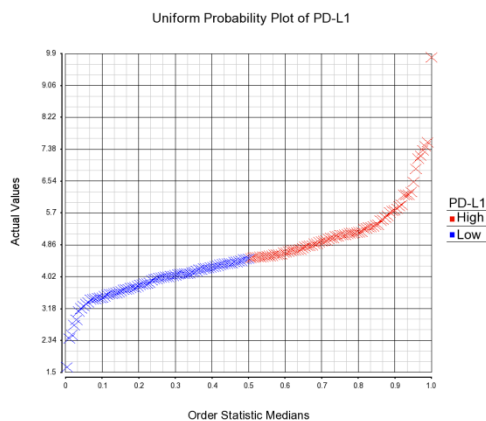
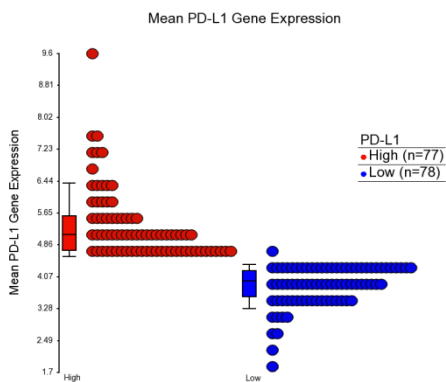


Epithelial v Leukocyte  $P = 0.0071$

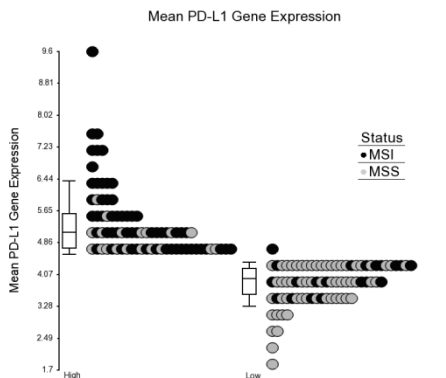


# Figure 2

**A**

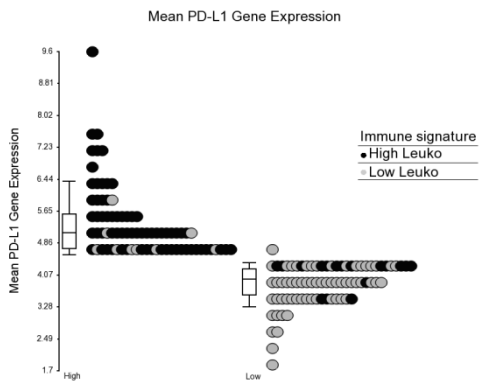


**B**



	MSI	MSS	Total
PD-L1 <sup>high</sup>	58	19	77
PD-L1 <sup>low</sup>	20	58	78
<b>Total</b>	<b>78</b>	<b>77</b>	<b>155</b>
<b>Fishers Exact two-tailed test <math>P &lt; 0.0001</math></b>			

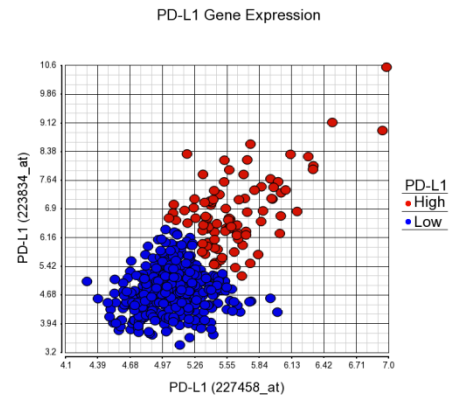
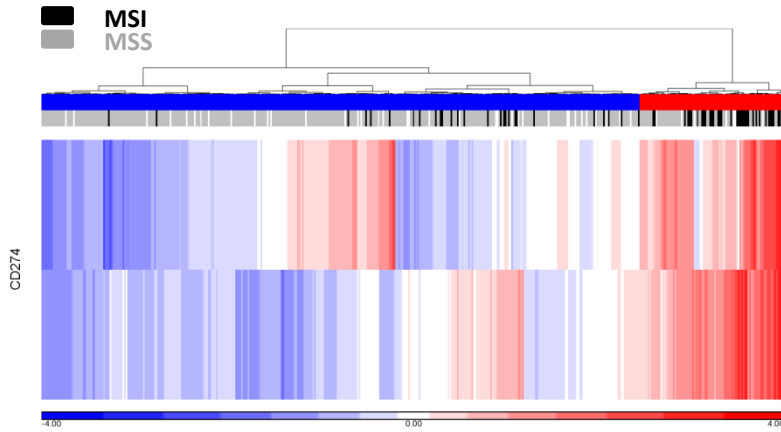
**C**



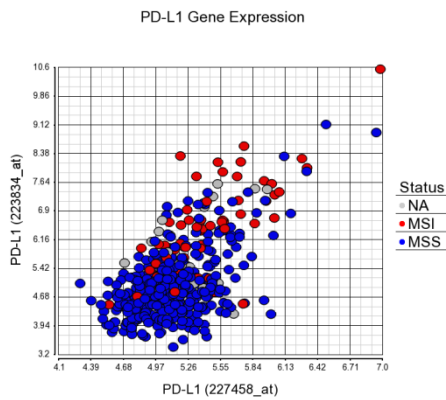
	High-Leuko	Low-Leuko	Total
PD-L1 <sup>high</sup>	68	9	77
PD-L1 <sup>low</sup>	17	61	78
<b>Total</b>	<b>85</b>	<b>70</b>	<b>104</b>
<b>Fishers Exact two-tailed test <math>P &lt; 0.0001</math></b>			

**Figure 3**

**A**

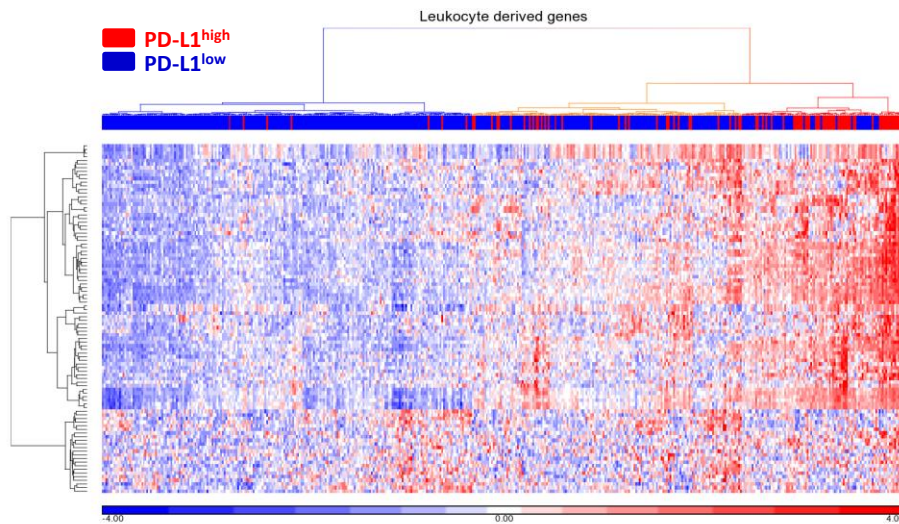


**B**



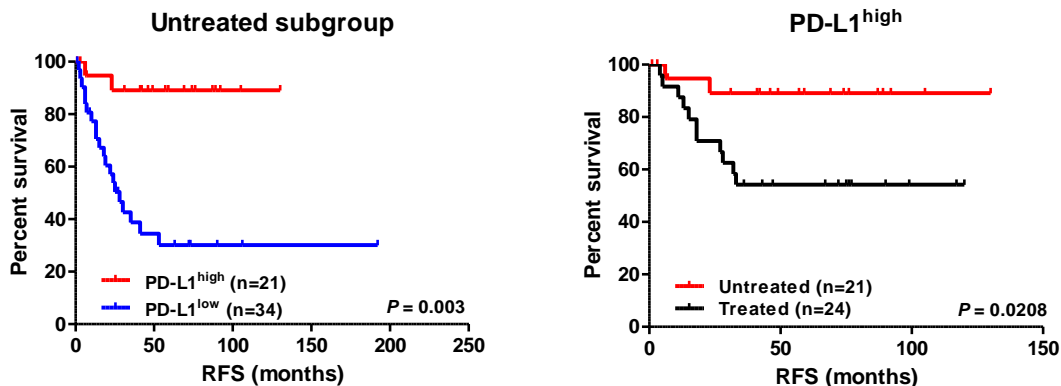
	MSI	MSS	Total
PD-L1 <sup>high</sup>		36	48
PD-L1 <sup>low</sup>	27	313	340
<b>Total</b>	<b>63</b>	<b>361</b>	<b>424</b>
<b>Fisher Exact two-tailed <math>P &lt; 0.0001</math></b>			

**C**

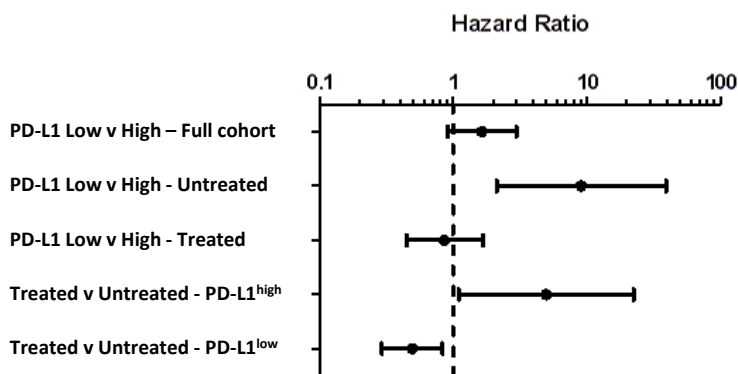


**Figure 4**

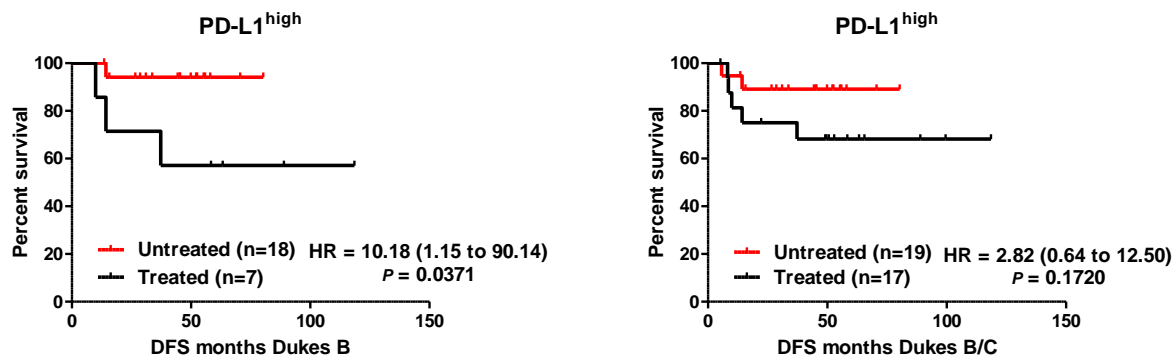
**A**



**B**



**C**





**Table 1**

	All colon cancers			PD-L1 <sup>high</sup> only		
Characteristic	PD-L1 <sup>low</sup> n = 340	PD-L1 <sup>high</sup> n = 84	P value	MSI n = 36	MSS n = 48	P value
Age, years, mean (SD)	67.6 (12.7)	69.5 (13.9)	0.23	70.8(17.0)	68.6 (11.2)	0.47
Sex, n (%)						
Male	193 (56.8)	49 (58.3)		19 (52.8)	30 (62.5)	
Female	147 (43.2)	35 (41.7)	0.80	17 (47.2)	18 (37.5)	0.37
Tumour stage, n (%)						
II	181 (53.2)	41 (48.8)		19 (52.8)	22 (45.8)	
III	159 (46.8)	43 (51.2)	0.47	17 (47.3)	26 (54.2)	0.53
Tumour location, n (%)						
Proximal	125 (36.8)	54 (64.3)		32 (88.9)	22 (45.8)	
Distal	215 (63.2)	30 (35.7)	<0.001	4 (11.1)	26 (54.2)	<0.001
Adjuvant treatment* receipt, n (%)						
No	178 (52.4)	54 (64.3)		27 (75.0)	27 (56.3)	
Yes	161 (47.4)	29 (34.5)	0.07	8 (22.2)	21 (43.8)	0.07
MSI status, n (%)						
MSI	27 (7.9)	36 (42.9)		/	/	/
MSS	313 (92.1)	48 (57.1)	<0.001	/	/	/
CIMP, n (%)						
Negative	246 (72.4)	41 (48.8)		11 (30.6)	30 (62.5)	
Positive	40 (11.8)	30 (35.7)		22 (61.1)	8 (16.7)	
Unknown	54 (15.9)	13 (15.5)	<0.001	3 (8.3)	10 (20.8)	<0.001
CIN, n (%)						
Negative	50 (14.7)	27 (32.1)		22 (61.1)	5 (10.4)	
Positive	218 (64.1)	41 (48.8)		11 (30.6)	30 (62.5)	
Unknown	72 (21.2)	16 (19.1)	0.001	3 (8.3)	13 (27.1)	<0.001
p53, n (%)						
WT	100 (29.4)	33 (39.3)		15 (41.7)	18 (37.5)	
MT	131 (38.5)	26 (31.0)		3 (8.3)	23 (47.9)	
Unknown	109 (32.1)	25 (29.8)	0.20	18 (50.0)**	7 (14.6)**	<0.001**
KRAS, n (%)						
WT	194 (57.1)	53 (63.1)		25 (69.4)	28 (58.3)	
MT	128 (37.7)	28 (33.3)		10 (27.8)	18 (37.5)	
Unknown	18 (5.3)	3 (3.6)	0.56	1 (2.8)	2 (4.2)	0.58
BRAF, n (%)						
WT	278 (81.8)	55 (65.5)		18 (50.0)	37 (77.1)	
MT	18 (5.3)	19 (22.6)		17 (47.2)	2 (4.2)	
Unknown	44 (12.9)	10 (11.9)	<0.001	1 (2.8)	9 (18.8)	<0.001

MSI: Microsatellite instability; MSS: Microsatellite stable; MT: Mutant; WT: Wild-type.  
 \*Adjuvant chemotherapy treatment receipt, unknown for 2 individuals (one PD-L1<sup>high</sup>, one PD-L1<sup>low</sup>).  
 \*\* p53 results confounded by lack of information for mutational status in 50% of PD-L1<sup>high</sup> MSI cases

**Table 1: Characteristics of colon cancer patients and tumours according to PD-L1 gene expression status.**

**Table 2**

		Unadjusted	Adjusted*
<b>Entire cohort</b>	PD-L1 Low v High	HR 1.64 (95% CI 0.91-2.97)	HR 1.32 (95% CI 0.69-2.54)
<b>Untreated Only</b>	PD-L1 Low v High	HR 9.09 (95% CI 2.11-39.10)	HR 4.81 (95% CI 0.79-29.19)
<b>Treated Only</b>	PD-L1 Low v High	HR 0.86 (95% CI 0.45-1.66)	HR 0.80 (95% CI 0.40-1.60)
<b>PD-L1<sup>High</sup> Only</b>	Treated v Untreated	HR 4.95 (95%CI 1.10-22.35)	HR 1.62 (95%CI 0.25-10.36)
<b>PD-L1<sup>Low</sup> Only</b>	Treated v Untreated	HR 0.49 (95%CI 0.29-0.83)	HR 0.44 (95%CI 0.25-0.80)
* Adjusted for all confounders in Table 1, with those that were significant at $P < 0.25$ level kept in the model as covariates. PD-L1 results adjusted for age, sex, CIMP, CIN, and KRAS. Tested for tumour location, BRAF, MSI, and p53.			

**Table 2: Unadjusted and adjusted analyses of relapse-free survival.** RFS analysis was performed using Cox proportional hazards method stratified by PD-L1 levels or treatment expression levels. Analysis was performed both before and following adjustment.