medRxiv preprint doi: https://doi.org/10.1101/2020.08.14.20174920; this version posted March 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

The clinical and prognostic significance of Protein Arginine Deiminase 2 and 4 (PADI2 & PADI4) in colorectal cancer

Mohamed Gijon[1], Rachael L Metheringham[1], Michael S Toss[3], Samantha J Paston[1], Lindy G Durrant[1, 2*]

[1] Scancell Limited, Biodiscovery Institute, University Park, Science Road, Nottingham NG7 2RD.
[2]University of Nottingham, Biodiscovery Institute, Division of Cancer and Stem Cells, University Park, Science Road, Nottingham NG7 2RD.
[3]Nottingham Breast Cancer Research Centre, Division of Cancer and Stem Cells, School of Medicine, The University of Nottingham, Nottingham City Hospital, Nottingham, UK.

*Corresponding author: Prof. L.G. Durrant, University of Nottingham, Biodiscovery Institute, Division of Cancer and Stem Cells, University Park, Science Road, Nottingham NG7 2RD; Lindy.Durrant@nottingham.ac.uk.

Short title: PADI2/PADI4 are independent predictors of good prognosis in colorectal cancer

Clinical data:

Written, informed consent was obtained for all patients in accordance with the Declaration of Helsinki (1964, amended most recently in 2008), with ethical approval from Nottingham Research Ethics committee (Q1020402).

Conflict of interest

LG Durrant is a director and shareholder in Scancell Ltd. All authors are employees of Scancell Ltd.

The authors have no conflicts of interest to declare.

Word count: 3128 (excluding abstract, acknowledgements, figures, tables and references).

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.14.20174920; this version posted March 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

ABSTRACT

Aims

Protein arginine deiminase (PADs) are a family of enzymes that catalyse the post translational modification (PTM) of proteins. Association between PAD expression with clinicopathology, protein expression and outcome was determined.

Methods

PADI2 and PADI4 expression was assessed immunohistochemically in a cohort of CRC patients.

Results

CRC tissues expressed variable levels of PADI2 which was mainly localised in the cytoplasm and correlated with patient survival (p=0.005); high expression increased survival time from 43.5 to 67.6 months. Expression of cytoplasmic PADI2 correlated with expression of nuclear β catenin, PADI4 and alpha-enolase. In contrast expression of nuclear PADI2 correlated with a decrease in survival (p=0.010), with high expression decreasing survival from 76.4 to 42.9 months. CRC tissues expressed variable levels of PADI4 in both the nucleus and cytoplasm. Expression of cytoplasmic PADI4 correlated with survival (p=0.001) with high expression increasing survival time from 48.1 to 71.8 months. Expression of cytoplasmic PADI4 correlated with expression of, nuclear β catenin, alpha-enolase ($p \leq 0.0001$, p=0.002) and the apoptotic related protein, Bcl-2. Expression of nuclear PADI4 also correlated with survival (p=0.011) with high expression of nuclear PADI4 increasing survival time from 55.4 to 74 months. Expression of nuclear PADI4 correlated with p53, alpha-enolase and Bcl-2. Multivariate analysis showed that TNM stage, cytoplasmic PADI2 and PADI4 remained independent prognostic factors in CRC. Both PADI2 and PADI4 are good prognostic factors in CRC.

Conclusions

High expression of cytoplasmic PADI2, PADI4 and nuclear PADI4 were associated with an increase in overall survival.

Keywords: PADI2, PADI4, citrullination, colorectal cancer, Immunohistochemistry,

medRxiv preprint doi: https://doi.org/10.1101/2020.08.14.20174920; this version posted March 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

INTRODUCTION

Colorectal cancer (CRC) is the third most common diagnosed cancer worldwide Despite substantial improvements in the diagnosis and treatment of CRC it still has a relatively low survival rate which estimated as 50-65% within 5 years depending on the healthcare systems [1]. The early detection of CRC could improve the outcome; however, early stage CRC often do not exhibit symptoms or have non-specific symptoms such which are usually misdiagnosed with other medical conditions [2]. However, most colon cancer patients will be diagnosed with regional or distant metastasis, thus adjuvant chemotherapy after surgery or palliative chemotherapy will often be needed [3].

The majority of CRCs (85%) arise from adenomas [4], the identification of the primary tumour, *BRAF* gene mutation status and tumour microsatellite instability (MSI) need to be determined to help guide clinician on the best treatment [5], which often being determined according to the progression of the disease, timing of the prior therapy, mutation profile and the toxicity profile of the corresponding drugs [6].

Approximately 4-5% CRC patients have tumours that are microsatellite instability-high (MSI-H) or mismatch repair deficient (dMMR), although these patients have a poor response to chemotherapy [7] they do have a clinical benefit following treatment with immune checkpoint inhibitors of PD-1/PD-L1 or in combination ipilimumab [8], [9]. A new approach for categorising CRC is to generate the Immune Score, which is defined by the number of lymphocytes infiltrating or surrounding the tumour. This score appears a strong prognostic factor for disease free and overall survival particularly in early-stage cancers. It also provides a tool or a target for novel therapeutic approaches, including immunotherapy [10]. T cell infiltration was a biomarker of good prognosis in CRC [11].

PAD enzymes catalyse the post translational modification (PTM) of proteins, in a process called citrullination. The exact biological function of citrullination remains obscure. Citrullination can also occur as a result of a degradation and recycling process called autophagy, and is induced in stressed cells [12], including cancer cells. Citrullination is catalysed exclusively by the family of Ca^{2+} dependent PAD enzymes that convert positively charged arginine residues within substrate proteins to neutrally charged citrulline, this modification can impact the protein structure, induce changes that result in protein denaturation and potentially alter the function of the modified proteins [13], [14]. Arginine deimination is recognised to be central for translational regulation of gene expression; to initiate tumour degradation, for formation of neutrophil extracellular traps (NET), and presentation of

medRxiv preprint doi: https://doi.org/10.1101/2020.08.14.20174920; this version posted March 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

antigens within stressed cells to the immune system. Five highly conserved PAD enzymes, (1-4 and 6) have been reported in mammals, each exhibiting tissue specific distribution, subcellular localisation and substrate specificity.

PADI2 and PADI4 have been associated with several type of cancers such as CRC, breast or lung playing a critical role in cancer cells migration [15]–[19]. Both have widespread distribution and are the predominant types in immune cells [20]. The substrate for PADI2 includes the intermediate filament protein vimentin that is expressed by all mesenchymal cells, myelin basic protein (MBP), actin and keratins [21]. It has been shown that PADI2 negatively regulates β-catenin signalling [22]. β-catenin is critical for tumour development and is commonly activated in CRC. PADI4 can be translocated to the nucleus to citrullinate histones and some other transcription factors [23], [24]. PADI4 acts mainly as a transcriptional co-regulator; its substrates include multiple proteins involved in gene regulation of the *p53* pathway such as *ING4* [25], *p300* [26], *p21* [27] and *histones* (*H3, H2A and H4*) [28]–[30]. Other substrates for PADI4 includes nucleophosmin [31] and nuclear lamin C [32], both involved in apoptosis. Increased expression of PADI2 and PADI4 has been detected in malignant tumours, suggesting citrullination has induced tumourigenesis [33]. To date the prognostic value of PADI2 and PADI4 in CRC remains unclear. We have hypothesised that PADI2 and PADI4 are associated with a good outcome in CRC through their role in the citrullination process. Herein, we have investigated the prognostic value of PADI2 and PADI4 expression in a large cohort of CRC.

PATIENTS AND METHODS

Study cohort

This study was conducted on a cohort of CRC (n=247) from patients undergoing elective surgical resection of histologically proven primary CRC at Nottingham University Hospitals, Nottingham between January 1994 and December 2000. Written informed consent was obtained for all patients in accordance with the Declaration of Helsinki, ethical approval was obtained from the Nottingham Research Ethics committee (Q1020402). The AJCC cancer staging system was used (https://www.cancer.gov/about-cancer/diagnosis-staging/staging), where 3 TNM parameters are described and then TNM combinations are grouped into five less-detailed stages (stage 0-IV). For the analysis of clinicopathological variables for cytoplasmic and nuclear PADI2 and PADI4 expression the patient cohort was grouped into stages I-II and III-IV to allow the comparison of early tumours verses larger and metastatic tumours. Tumours histological subtypes were merged into two groups to obtain more informative statistical details. These groups comprised mucinous adenocarcinoma and

medRxiv preprint doi: https://doi.org/10.1101/2020.08.14.20174920; this version posted March 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

adenocarcinoma which included all other carcinoma subtypes. The clinical details of the study cohort have been previously described [11], [36] and are summarised in Table 1. Clinicopathology information was obtained from pathology reports. Patients with lymph node positive disease routinely received adjuvant chemotherapy with 5- fluorouracil and folinic acid. The mean follow up period was 42 months (range 1-116). Disease specific survival (DSS) was determined from the date of resection of the primary tumour until December 31st 2003 when any survivors were censored.

Immunohistochemistry (IHC)

Formalin fixed paraffin embedded (FFPE) tissue samples of the study cohort were retrieved from tissue archive where 4um thick sections were cut and stained with haematoxylin and eosin (H&E) stains for tumour burden assessment. Tissue microarray (TMA) was constructed from these samples prior to immunohistochemistry (IHC). To validate the TMAs for immuno-phenotyping, full-face sections of 40 randomly selected cases were stained and the protein expression levels were compared. The concordance between TMAs and full-face sections was compared using Cohen's kappa test and reliability was excellent (kappa=0.8). IHC staining was performed using a streptavidin–biotin peroxidase method in conjunction with a Leica autostainer XL for tissue dewaxing and rehydration and the Novolink polymer detection system kit (Leica Biosystems, Newcastle-upon-Tyne, UK). Full face and TMA sections were first deparaffinised with xylene, rehydrated through graded alcohol and immersed in Peroxidase block (Novolink) for 5 minutes. The antigen retrieval was performed by immersing sections in citrate buffer at pH 6.0 and heating in a 800W microwave for 10 minutes at high power followed by 10 minutes at low power. Endogenous avidin/biotin binding was blocked (Protein blocking kit, Novolink), followed by addition of 100 µL Protein Block (Novolink) for 5 minutes. Sections were incubated for 1 hour with polyclonal rabbit anti-PADI2 antibody (Abcam, cat no 16478, 1:150 dilution), monoclonal rabbit anti-PADI4 antibody (Abcam cat no ab128086 1:700 dilution) for a 1 hour incubation at room temperature, monoclonal mouse anti-HLA-Class-II (DP, DQ, DR antigen) (Dako, cat no M0775, 1:1000 dilution) for 45 minutes at room temperature, monoclonal mouse anti-Vimentin antibody (Dako, cat no M7020, 1:150 dilution) for a 1 hour incubation at room temperature or monoclonal rabbit anti-ENO1 (Abcam, cat no ab155955, 1:100 dilution) for a 1 hour incubation at room temperature. The rest of the markers (Table 3) were stained in previous studies (P53 and Bcl2 [37], β-catenin [38], Ki-67, VEGFA and CEA [39]). After washing with tris buffered saline (TBS), sections were incubated with 100 µL of post primary antibody (Novolink kit) for 30 mins. After washing in TBS sections were incubated with 100 µL Novolink polymer for 30 mins. Sections were washed in TBS. Visualisation was achieved using 3, 3′-diaminobenzidine tetra hydrochloride (DAB,

medRxiv preprint doi: https://doi.org/10.1101/2020.08.14.20174920; this version posted March 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Novolink), and then counterstained with haematoxylin (Novolink), dehydrated in alcohol, cleared in xylene (Genta Medica, York, UK), and mounted with distyrene, plasticiser and xylene (DPX–BDH, Poole, UK).

Slide scanning, viewing and scoring.

TMA stained slides were digitalised using NanoZoomer (Hamamatsu, Bridgewater, NJ, USA) high throughput scanner. Images were visualised using Aperio ImageScope v11.2.0.780 software. Cytoplasmic PADI2 and cytoplasmic and nuclear PADI4 proteins expression was assessed using the percentage of positivity (0-100%) and staining intensity (Negative: Score 0, Weak: Score 1, Moderate: Score 2, and Strong: Score 3) following the Histo-score system (H-score 0 to 300. Negative=0; weak=1-100; moderate=101-200 and strong=201-300). Observers were blind to clinical and pathological parameters.

Statistical analysis

Statistical analysis was performed using SPSS21 statistical software (SPSS Inc., Chicago, IL, USA). Pearson's $\chi 2$-tests and Fisher's exact tests were used to determine the significance of associations between categorical variables. The cohort was categorised into two groups by defining the optimal cut-off points for cytoplasmic PADI2 (H-score 150), nuclear PADI2 (H-score 160), cytoplasmic PADI4 (H-score 100) and nuclear PADI4 (H-score 260) expression by X-tile software (X-Tile Bioinformatics Software, Yale University, version 3.6.1). Survival rates were calculated using the Kaplan–Meier method; differences between survival curves were examined using the log-rank test. The Cox regression model was used for multivariate analysis in order to calculate the Hazard ratios and independent significance of individual factors. In all cases two-sided P-values of <0.05 were considered as statistically significant.

RESULTS

Frequency and distribution of PADI2 and PADI4 expression in the study cohort

After exclusion of uninformative cores (i.e., loss or folding of cores during tissue sectioning and processing), there was 247 and 231 cores suitable for scoring of PADI2 and PADI4 expression respectively. Clinicopathological details of each marker are described in Table 1.

In this cohort 99% of the cases expressed PADI2 in the cytoplasm, where 28% showed nuclear and cytoplasmic co-expression. PADI4 showed cytoplasmic and nuclear in all cases. Occasional

medRxiv preprint doi: https://doi.org/10.1101/2020.08.14.20174920; this version posted March 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

surrounding inflammatory cells and fibroblasts showed weak expression to PADI2 and PADI4, however, this was not evaluated further in this study.

According to the H-score assessment. After categorisation of each marker using the optimal cut offs there was 62% of the tumours showed high expression of cytoplasmic PADI2, 28% moderate expression, 9% weak expression and only 1% were negative for PADI2, representative staining is shown in Figure 1A-C. For nuclear PADI2 30% tumours showed strong expression, 32% moderate expression, 34% weak expression and only 4% were negative for nuclear PADI2. The H-score assessment for cytoplasmic PADI4 showed all tumours expressed PADI4, 11% strong expression, 61% moderate expression and 28% weak expression, representative staining is shown in Figure 1D-F. All tumours expressed nuclear PADI4, 77% strong expression, 19% moderate expression and 4% showed weak expression.

Correlation of PADI2 and PADI4 expression with the clinicopathological variables

High nuclear expression of PADI2 was significantly correlated with the TNM stage (p=0.017), while no significant association were observed between nuclear PADI4, cytoplasmic PADI2 or PADI4 expression with any of the other clinicopathological parameters (Table 2).

The results from the univariate analysis for cytoplasmic PADI2 with protein expression showed that there is a significant correlation with nuclear β-catenin (p=0.020), vascular endothelial growth factor A (VEGFA) (p=0.001) and carcinoembryonic antigen (CEA) (p=0.042), cytoplasmic and nuclear PADI4 ($p \leq 0.0001$, p=0.048) and cytoplasmic alpha-enolase ($p \leq 0.001$). The results from the analysis for nuclear PADI2 showed only a correlation with the tumour suppressor protein, p53 (p=0.024). The results from the univariate analysis for cytoplasmic PADI4 showed there is a significant correlation with nuclear β-catenin (p=0.004), nuclear PADI4 ($p \leq 0.0001$), cytoplasmic and nuclear alpha-enolase ($p \leq 0.0001$, $p \leq 0.002$) and the apoptotic related protein, Bcl-2 (p=0.032). The results from the analysis for nuclear PADI4 showed there is a significant correlation with nuclear alpha-enolase (p=0.011), p53 (p=0.016) and Bcl-2 (p=0.022) (Table 3).

PADI2 and PADI4 expression and outcome

High cytoplasmic expression of PADI2 showed a significant correlation with better survival (p=0.005) (Figure 2A). In contrast, high nuclear expression of PADI2, (Figure 2B) was associated with a decrease in survival (p=0.010). Higher cytoplasmic and nuclear PADI4 (Figures 2C and 2D) showed there was

medRxiv preprint doi: https://doi.org/10.1101/2020.08.14.20174920; this version posted March 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

a significant correlation with survival with tumours expressing cytoplasmic PADI4 and nuclear PADI4 (p=0.001 and p=0.011, respectively).

Cytoplasmic PADI2 and PADI4 expression by colorectal tumours are independent prognostic markers

Multivariate Cox regression model including other factors known to be associated with poor outcome in CRC (TNM stage and vascular invasion), showed high cytoplasmic expression of PADI2 (Table 4) was an independent prognostic protective marker (p<0.001). The expression of cytoplasmic PADI4 (Table 4) also conferred independent tumour protection.

DISCUSSION

To improve the survival in patients with CRC its early diagnosis and treatment is critical. A number of prognostic makers have been investigated to identify patients at high risk of relapse. The identification of a reliable prognostic marker has been challenging. Post translational modifications of the human proteome can have a major impact on health and disease, increasing the functional diversity of proteins by adding new functional groups. There is accumulating evidence that citrullination is implicated in cancer development and can be an excellent target to alert the immune response to recognise and remove stressed cells including cancer cells. This study aimed to investigate the prognostic value of PADI2 and PADI4 in CRC.

PAD enzymes are activated by a high concentration of calcium, typically in the millimolar range [12]. In healthy viable cells the PAD enzymes are maintained in an inactive state due to the low concentrations of Ca^{2+} [13], in double membrane vesicles within viable cells the calcium concentration can be high leading to the activation of the PAD enzymes. The PAD enzymes play a key role in the regulation of gene expression through the citrullination of histones and transcription factors and have been shown to have a role in tumour progression [21]. Citrullination can occur within autophagosomes as a result of autophagy, here the activated PAD enzymes citrullinate engulfed proteins from the cytoplasm [40], [41], this process is induced in stressed cells [42] including cancer cells. During inflammation, epitopes are citrullinated and can be subsequently presented by MHC class II (MHC-II) molecules for recognition by CD4 T cells [42]. A number of studies in autoimmune patients have demonstrated that CD4 T cell responses can be detected to citrullinated proteins such as vimentin and enolase [43]–[46]. We have already shown that citrullinated α-enolase and vimentin are excellent

medRxiv preprint doi: https://doi.org/10.1101/2020.08.14.20174920; this version posted March 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

targets for anti-tumour immunity inducing potent immune responses [43], [47], [48] and *in vivo* tumour rejection [49].

Our data shows that 99% of CRC tumours express PADI2 and all tumours in our cohort expressed PADI4. Increasing evidence suggests that the PAD enzymes play a role normal cell transformation and tumour progression. We also show that in the majority of cases PADI4 was expressed in both the cytoplasm and nucleus, whereas PADI2 was mainly located in the cytoplasm with only 28% of tumours showing nuclear expression. Analysis of the clinicopathological parameters for the cohort stained with PADI2 and PADI4 revealed that only nuclear PADI2 showed a correlation with a clinical parameter, notably TNM stage. Both PADI2 and PADI4 target different cellular substrates [50], [51], defined by the cell type and the subcellular localisation of the substrate, this indicates that PADI2 and PADI4 acts as an independent prognosis markers.

We investigated if the expression of PADI2 or PADI4 correlated with other markers involved in cancer, cell survival and progression. Nuclear PADI2 correlated with p53 expression and both conferred a worse prognosis [37]. In contrast the expression of nuclear PADI4 conferred a good prognosis and inversely correlated with p53 expression. Anwar *et al.* analysed the results from 35 studies where p53 was found to predict of a worse outcome [52]. Multiple interactions between PADI4 and p53 have been reported, suggesting the importance of PAD-induced citrullination during apoptosis. The expression of the major p53 target gene OKL38 was suppressed by the p53-mediated recruitment of PADI4 to the promoter of OKL38 and the subsequent removal of histone arginine methylation mark, thus directly modulating apoptosis [53]. PADI4 was found to bind and then citrullinate the inhibitor of growth 4 (ING4), another tumour suppressor protein that is known to bind to p53. The PADI4 mediated citrullination of ING4 at the nuclear localisation sequence region prevents p53-to-ING4 binding, suppresses p53 acetylation, and subsequently inhibited downstream p21 expression [25]. Nuclear and cytoplasmic PADI4 conferred a good prognosis and correlated with expression of Bcl-2 which is also a predictor of good prognosis [37]. Despite the defined role of Bcl-2 protein in suppression of apoptosis, in CRC the evidence suggests that Bcl-2 expression is correlated with favourable parameters [54] and a better prognosis [55]. Although Bcl-2 inhibits apoptosis it has also been shown to suppress tumour cell growth [56] and better differentiation of the tumours, possibly resulting in a more favourable outcome. It has been reported that Bcl-2 contains an anti-proliferative domain, distinct from the domains required for its anti-apoptotic activity [57], and also Bcl-2 can be converted to a BAX-like death effector by cleavage of a regulatory loop domain by caspases [58]. Cytoplasmic PADI2 and PADI4 expression correlated with nuclear β-catenin expression and both

medRxiv preprint doi: https://doi.org/10.1101/2020.08.14.20174920; this version posted March 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

correlated with good prognosis in this cohort. Wnt/β-catenin signalling is important for carcinogenesis of certain tumours, and PADI2-mediated citrullination was recently linked to this pathway [22]. The anti-parasitic drug temozolomide (NTZ) inhibited Wnt/β-catenin signalling through directly targeting PADI2 as demonstrated by co-immunoprecipitation studies. The addition of NTZ resulted in a significant increase of β-catenin citrullination by PADI2 and also substantially extended the half-life of PADI2 protein, suggesting that NTZ enhances the stability of PADI2. Furthermore, the results from colony growth assays have suggested that PADI2- mediated citrullination of β-catenin could limit the proliferation of CRC cells by inhibiting the Wnt pathway. As a result, this elegant study uncovered a hitherto unrecognised mechanism of Wnt signalling inhibition via PADI2.

The PAD enzymes have been described as playing a role in early tumour cell transformation and tumour progression. In addition, the PAD enzymes also catalyse the PTM of proteins under conditions of cellular stress, in the process of citrullination. Citrullination occurs as a result of the degradation and recycling process of autophagy. In the presence of inflammation, epitopes are citrullinated and can be subsequently presented by MHC-II molecules for recognition by CD4 T cells [42]. We have recently shown that citrullinated α-enolase and vimentin are excellent targets for anti-tumour immunity inducing potent immune responses [43], [47], [48] and *in vivo* tumour rejection [49]. It was therefore of interest that cytoplasmic PADI2 correlated with cytoplasmic and nuclear alpha-enolase, and cytoplasmic PADI4 correlated with vimentin and cytoplasmic and nuclear alpha-enolase suggesting that these may be good targets for citrullination in colorectal cancer, these tumour cells can then be recognised and removed by the immune system.

In conclusion, our results show that high expression of cytoplasmic PADI2, PADI4 and nuclear PADI4 were associated with an increase in overall survival. Nuclear expression of PADI4 may confer a good prognosis by regulating expression of genes such as the p53/Bcl-2. Cytoplasmic PADI2 and or PADI4 may confer a good prognosis by altering cell signalling pathways such as via β-catenin or may be involved in protein degradation which can lead to presentation of citrullinated epitopes for detection by cytotoxic CD4 T cells. Cancer vaccines targeting citrullinated vimentin and/or enolase maybe a good therapy for CRC patients expressing high levels of PAD enzymes [12].

Acknowledgements

medRxiv preprint doi: https://doi.org/10.1101/2020.08.14.20174920; this version posted March 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

The authors would also like to thank Dr Tina Parsons for critical reading of the manuscript, Maria Diez and Chris Nolan for their technical support to this research.

Funding: This work was funded by Scancell Ltd.

Statement of authors contributions

MG developed the methodology, performed experiments, interpreted IHC staining data and collected clinical data. LD, MG, RM performed statistical analysis. LD contributed to the study design. MG wrote the draft manuscript. MT, SP, LD provided critical review of the data and contributed to the final manuscript. LD supervised the study. All the authors reviewed and approved the final manuscript.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.14.20174920; this version posted March 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

References

[1] H. Brenner, M. Kloor, and C. P. Pox, "Colorectal cancer," *Lancet*, vol. 383, no. 9927, pp. 1490–1502, 2014, doi: 10.1016/S0140-6736(13)61649-9.

[2] P. Vega, F. Valentín, and J. Cubiella, "Colorectal cancer diagnosis: Pitfalls and opportunities," *World J. Gastrointest. Oncol.*, vol. 7, no. 12, p. 422, 2015, doi: 10.4251/wjgo.v7.i12.422.

[3] C. Wu, "Systemic Therapy for Colon Cancer," *Surg. Oncol. Clin. N. Am.*, vol. 27, no. 2, pp. 235–242, 2018, doi: 10.1016/j.soc.2017.11.001.

[4] M. L. Dwyer, A. Colombo, and G. Bozzi, "A Genetic Model for Colorectal Tumorigenesis," *G. Ital. Cardiol.*, vol. 19, no. 2, pp. 170–172, 1989.

[5] A. B. Benson *et al.*, "NCCN Guidelines ® Insights Colon Cancer, Version 2.2018 Featured Updates to the NCCN Guidelines," *JNCCN J. Natl. Compr. Cancer Netw.*, vol. 16, no. 4, pp. 359–369, 2018, doi: 10.6004/jnccn.2018.0021.

[6] A. B. Benson *et al.*, "Colon cancer, version 1.2017: Clinical practice guidelines in oncology," *JNCCN J. Natl. Compr. Cancer Netw.*, vol. 15, no. 3, pp. 370–398, 2017, doi: 10.6004/jnccn.2017.0036.

[7] J. Taieb *et al.*, "473OThree versus six months' adjuvant oxaliplatin-based chemotherapy for patients with stage III colon cancer: Per-protocol, subgroups and long-lasting neuropathy results," *Ann. Oncol.*, vol. 28, no. suppl_5, p. 2017, 2017, doi: 10.1093/annonc/mdx393.

[8] L. Marcus, S. J. Lemery, P. Keegan, and R. Pazdur, "FDA approval summary: Pembrolizumab for the treatment of microsatellite instability-high solid tumors," *Clin. Cancer Res.*, vol. 25, no. 13, pp. 3753–3758, 2019, doi: 10.1158/1078-0432.CCR-18-4070.

[9] S. Kamatham, F. Shahjehan, and P. M. Kasi, "Immune Checkpoint Inhibitors in Metastatic Colorectal Cancer: Current Status, Recent Advances, and Future Directions," *Curr. Colorectal Cancer Rep.*, vol. 15, no. 4, pp. 112–121, 2019, doi: 10.1007/s11888-019-00437-z.

[10] H. K. Angell, D. Bruni, J. Carl Barrett, R. Herbst, and J. Galon, "The immunoscore: Colon cancer and beyond a C," *Clin. Cancer Res.*, vol. 26, no. 2, pp. 332–339, 2020, doi: 10.1158/1078-0432.CCR-18-1851.

[11] J. A. D. Simpson, A. Al-Attar, N. F. S. Watson, J. H. Scholefield, M. Ilyas, and L. G. Durrant, "Intratumoral T cell infiltration, MHC class I and STAT1 as biomarkers of good prognosis in colorectal cancer," *Gut*, vol. 59, no. 7, pp. 926–933, 2010, doi: 10.1136/gut.2009.194472.

[12] V. A. Brentville *et al.*, "Combination vaccine based on citrullinated vimentin and enolase mediated peptides induces potent CD4-mediated- anti--tumor responses," pp. 6–9, 2020, doi: 10.1136/jitc-2020-000560.

[13] M. Heilbron Ferrer, "Protein Arginine Deiminases and Associated Citrullination: Physiological Functions and Diseases Associated with Dysregulation," *Curr Drug Targets. 2015 ;*, vol. 16, no. 57, pp. 199–227, 2002, doi: 10.1016/j.bbamem.2015.02.010.Cationic.

[14] M. Alghamdi *et al.*, "An overview of the intrinsic role of citrullination in autoimmune disorders," *J. Immunol. Res.*, vol. 2019, 2019, doi: 10.1155/2019/7592851.

[15] L. Shi, H. Yao, Z. Liu, M. Xu, A. Tsung, and Y. Wang, "Endogenous PAD4 in Breast Cancer Cells Mediates Cancer Extracellular Chromatin Network Formation and Promotes Lung Metastasis," *Mol. Cancer Res.*, vol. 18, no. 5, pp. 735–747, 2020, doi: 10.1158/1541-7786.MCR-19-0018.

[16] Y. Wang, R. Chen, Y. Gan, and S. Ying, "The roles of PAD2- and PAD4-mediated protein citrullination catalysis in cancers," *Int. J. Cancer*, vol. 148, no. 2, pp. 267–276, 2021, doi:

medRxiv preprint doi: https://doi.org/10.1101/2020.08.14.20174920; this version posted March 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

10.1002/ijc.33205.

[17] A. E. Yuzhalin *et al.*, "Colorectal cancer liver metastatic growth depends on PAD4-driven citrullination of the extracellular matrix," *Nat. Commun.*, vol. 9, no. 1, 2018, doi: 10.1038/s41467-018-07306-7.

[18] F. Li *et al.*, "Inhibiting PAD2 enhances the anti-tumor effect of docetaxel in tamoxifen-resistant breast cancer cells," *J. Exp. Clin. Cancer Res.*, vol. 38, no. 1, pp. 1–15, 2019, doi: 10.1186/s13046-019-1404-8.

[19] S. Horibata *et al.*, "Role of peptidylarginine deiminase 2 (PAD2) in mammary carcinoma cell migration," *BMC Cancer*, vol. 17, no. 1, pp. 1–14, 2017, doi: 10.1186/s12885-017-3354-x.

[20] Y. Zhou *et al.*, "Spontaneous secretion of the citrullination enzyme PAD2 and cell surface exposure of PAD4 by neutrophils," *Front. Immunol.*, vol. 8, no. SEP, pp. 1–16, 2017, doi: 10.3389/fimmu.2017.01200.

[21] A. E. Yuzhalin, "Citrullination in cancer," *Cancer Res.*, vol. 79, no. 7, pp. 1274–1284, 2019, doi: 10.1158/0008-5472.CAN-18-2797.

[22] Y. Qu *et al.*, "Small molecule promotes b-catenin citrullination and inhibits Wnt signaling in cancer," *Nat. Chem. Biol.*, vol. 14, no. 94, 2018, doi: 10.1038/nchembio.2510.

[23] H. Asaga, K. Nakashima, T. Senshu, and A. Ishigami, "Immunocytochemical localization of peptidylarginine deiminase in human eosinophils and neutrophils Abstract : Peptidylarginine deiminase , registered banks , is expressed in HL-60 cells differentiated blood cell fractions . PAD activity with similar sub- ," *J. Leukoc. Biol.*, 2001.

[24] K. Nakashima, T. Hagiwara, and M. Yamada, "Nuclear localization of peptidylarginine deiminase V and histone deimination in granulocytes," *J. Biol. Chem.*, vol. 277, no. 51, pp. 49562–49568, 2002, doi: 10.1074/jbc.M208795200.

[25] Q. Guo and W. Fast, "Citrullination of Inhibitor of Growth 4 (ING4) by Peptidylarginine Deminase 4 (PAD4) disrupts the interaction between ING4 and p53," *J. Biol. Chem.*, vol. 286, no. 19, pp. 17069–17078, 2011, doi: 10.1074/jbc.M111.230961.

[26] Y. H. Lee, S. A. Coonrod, W. L. Kraus, M. A. Jelinek, and M. R. Stallcup, "Regulation of coactivator complex assembly and function by protein arginine methylation and demethylimination," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 10, pp. 3611–3616, 2005, doi: 10.1073/pnas.0407159102.

[27] P. Li *et al.*, "Regulation of p53 Target Gene Expression by Peptidylarginine Deiminase 4," *Mol. Cell. Biol.*, vol. 28, no. 15, pp. 4745–4758, 2008, doi: 10.1128/mcb.01747-07.

[28] T. Hagiwara, Y. Hidaka, and M. Yamada, "Deimination of histone H2A and H4 at arginine 3 in HL-60 granulocytes," *Biochemistry*, vol. 44, no. 15, pp. 5827–5834, 2005, doi: 10.1021/bi047505c.

[29] G. L. Cuthbert *et al.*, "Histone deimination antagonizes arginine methylation," *Cell*, vol. 118, no. 5, pp. 545–553, 2004, doi: 10.1016/j.cell.2004.08.020.

[30] Y. Wang *et al.*, "Human PAD4 regulates histone arginine methylation levels via demethylimination," *Science (80-.).*, vol. 306, no. 5694, pp. 279–283, 2004, doi: 10.1126/science.1101400.

[31] T. Hagiwara, K. Nakashima, H. Hirano, T. Senshu, and M. Yamada, "Deimination of arginine residues in nucleophosmin/B23 and histones in HL-60 granulocytes," *Biochem. Biophys. Res. Commun.*, vol. 290, no. 3, pp. 979–983, 2002, doi: 10.1006/bbrc.2001.6303.

[32] C. Tanikawa *et al.*, "Regulation of histone modification and chromatin structure by the p53-

medRxiv preprint doi: https://doi.org/10.1101/2020.08.14.20174920; this version posted March 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

PADI4 pathway," *Nat. Commun.*, vol. 3, 2012, doi: 10.1038/ncomms1676.

[33] X. Chang, J. Han, L. Pang, Y. Zhao, Y. Yang, and Z. Shen, "Increased PADI4 expression in blood and tissues of patients with malignant tumors," *BMC Cancer*, vol. 9, pp. 1–11, 2009, doi: 10.1186/1471-2407-9-40.

[34] R. Funayama *et al.*, "Protein-arginine deiminase 2 suppresses proliferation of colon cancer cells through protein citrullination," *Cancer Sci.*, vol. 108, no. 4, pp. 713–718, 2017, doi: 10.1111/cas.13179.

[35] N. Cantarino *et al.*, "Downregulation of the deiminase PADI2 is an early event in colorectal carcinogenesis and indicates poor prognosis," *Mol. Cancer Res.*, vol. 14, no. 9, pp. 841–848, 2016, doi: 10.1158/1541-7786.MCR-16-0034.

[36] N. F. S. Watson *et al.*, "Immunosurveillance is active in colorectal cancer as downregulation but not complete loss of MHC class I expression correlates with a poor prognosis," *Int. J. Cancer*, vol. 118, no. 1, pp. 6–10, 2006, doi: 10.1002/ijc.21303.

[37] N.F.S. Watson et al., "Evidence that the p53 negative / Bcl-2 positive phenotype is an independent indicator of good prognosis in colorectal cancer: A tissue microarray study of 460 patients," *Tandlægebladet*, vol. 9, no. 10, pp. 790–796, 2005, doi: 10.1186/1477-7819-3-Received.

[38] G. G. Chung, E. Provost, E. P. Kielhorn, L. A. Charette, D. L. Rimm, and B. L. Smith, "Tissue microarray analysis of β-catenin in colorectal cancer shows nuclear phospho-β-catenin is associated with a better prognosis," *Clin. Cancer Res.*, vol. 7, no. 12, pp. 4013–4020, 2001.

[39] L. G. Durrant, R. A. Robins, and R. W. Baldwin, "Flow cytometric screening of monoclonal antibodies for drug or toxin targeting to human cancer," *J. Natl. Cancer Inst.*, vol. 81, no. 9, pp. 688–696, 1989, doi: 10.1093/jnci/81.9.688.

[40] J. M. Ireland and E. R. Unanue, "Autophagy in antigen-presenting cells results in presentation of citrullinated peptides to CD4 T cells," *J. Exp. Med.*, vol. 208, no. 13, pp. 2625–2632, 2011, doi: 10.1084/jem.20110640.

[41] M. Sorice *et al.*, "Autophagy generates citrullinated peptides in human synoviocytes: A possible trigger for anti-citrullinated peptide antibodies," *Rheumatol. (United Kingdom)*, vol. 55, no. 8, pp. 1374–1385, 2016, doi: 10.1093/rheumatology/kew178.

[42] V. A. Brentville, M. Vankemmelbeke, R. L. Metheringham, and L. G. Durrant, "Post-translational modifications such as citrullination are excellent targets for cancer therapy," *Semin. Immunol.*, vol. 47, no. October 2019, p. 101393, 2020, doi: 10.1016/j.smim.2020.101393.

[43] A. L. Feitsma *et al.*, "Identification of citrullinated vimentin peptides as T cell epitopes in HLA-DR4-positive patients with rheumatoid arthritis," *Arthritis Rheum.*, vol. 62, no. 1, pp. 117–125, 2010, doi: 10.1002/art.25059.

[44] C. Gerstner *et al.*, "Functional and structural characterization of a novel HLA-DRB1*04: 01-restricted α-enolase T cell epitope in rheumatoid arthritis," *Front. Immunol.*, vol. 7, no. NOV, pp. 1–12, 2016, doi: 10.3389/fimmu.2016.00494.

[45] E. A. James *et al.*, "Citrulline-specific Th1 cells are increased in rheumatoid arthritis and their frequency is influenced by disease duration and therapy," *Arthritis Rheumatol.*, vol. 66, no. 7, pp. 1712–1722, 2014, doi: 10.1002/art.38637.

[46] O. Snir *et al.*, "Identification and functional characterization of T cells reactive to citrullinated vimentin in HLA-DRB1*0401-positive humanized mice and rheumatoid arthritis patients,"

medRxiv preprint doi: https://doi.org/10.1101/2020.08.14.20174920; this version posted March 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Arthritis Rheum., vol. 63, no. 10, pp. 2873–2883, 2011, doi: 10.1002/art.30445.

[47] V. A. Brentville *et al.*, "Citrullinated vimentin presented on MHC-II in tumor cells is a target for CD4+ T-Cell-mediated antitumor immunity," *Cancer Res.*, vol. 76, no. 3, pp. 548–560, 2016, doi: 10.1158/0008-5472.CAN-15-1085.

[48] L. G. Durrant, R. L. Metheringham, and V. A. Brentville, "Autophagy, citrullination and cancer," *Autophagy*, vol. 12, no. 6, pp. 1055–1056, 2016, doi: 10.1080/15548627.2016.1166326.

[49] V. A. Brentville *et al.*, "Citrullinated vimentin presented on MHC-II in tumor cells is a target for CD4+ T-Cell-mediated antitumor immunity," *Cancer Res.*, vol. 76, no. 3, pp. 548–560, 2016, doi: 10.1158/0008-5472.CAN-15-1085.

[50] E. Darrah, A. Rosen, J. T. Giles, and F. Andrade, "Peptidylarginine deiminase 2, 3 and 4 have distinct specificities against cellular substrates: Novel insights into autoantigen selection in rheumatoid arthritis," *Ann. Rheum. Dis.*, vol. 71, no. 1, pp. 92–98, 2012, doi: 10.1136/ard.2011.151712.

[51] B. Amin and W. Voelter, *Human Deiminases: Isoforms, Substrate Specificities, Kinetics, and Detection*, vol. 106. 2017.

[52] S. Anwar, I. M. Frayling, N. A. Scott, and G. L. Carlson, "Systematic review of genetic influences on the prognosis of colorectal cancer," *Br. J. Surg.*, vol. 91, no. 10, pp. 1275–1291, 2004, doi: 10.1002/bjs.4737.

[53] H. Yao *et al.*, "Histone Arg modifications and p53 regulate the expression of OKL38, a mediator of apoptosis," *J. Biol. Chem.*, vol. 283, no. 29, pp. 20060–20068, 2008, doi: 10.1074/jbc.M802940200.

[54] F. A. Sinicrope, J. Hart, F. Michelassi, and J. Jack Lee, "Prognostic Value of bcl-2 Oncoprotein Expression in Stage II Colon Carcinoma," *Clin. Cancer Res.*, vol. 1, no. 10, pp. 1103–1110, 1995.

[55] D. Öfner *et al.*, "Immunohistochemically detectable bcl-2 expression in colorectal carcinoma: Correlation with tumour stage and patient survival," *Br. J. Cancer*, vol. 72, no. 4, pp. 981–985, 1995, doi: 10.1038/bjc.1995.446.

[56] J. A. Pietenpol, N. Papadopoulos, S. Markowitz, J. K. V Willson, K. W. Kinzler, and B. Vogelstein, "Paradoxical Inhibition of Solid Tumor Cell Growth by bcl2 ′," *Cancer Res.*, vol. 54, no. 14, pp. 3714–3717, 1994.

[57] P. Theodorakis, C. D'Sa-Eipper, T. Subramanian, and G. Chinnadurai, "Unmasking of a proliferation-restraining activity of the anti-apoptosis protein EBV BHRF1," *Oncogene*, vol. 12, no. 8, pp. 1707–1713, 1996.

[58] E. H. Cheng *et al.*, "Conversion of Bcl-2 to a Bax-like Death Effector by Caspases," *Science (80-.).*, vol. 278, no. December, pp. 2–5, 1997.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.14.20174920; this version posted March 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure legends

Figure 1. Representative immunohistochemistry staining of PADI2 (A-C) and PADI4 (D-F) in colorectal tissue microarray images. The level of expression ranged from weak (A, D), moderate (B, E) and strong expression (C, F) according to the H-score.

Figure 2. Kaplan-Meier curves for disease survival for CRC patients from the time of tumour resection showing high (green line) or low expression (blue line) of PADI2 or PADI4, (A) cytoplasmic PADI2, low = 42 cases, high = 205 cases (B) nuclear PADI2, low = 33 cases, high = 40 cases (C) cytoplasmic PADI4, low = 64 cases, high = 167 cases (D) nuclear PADI4, low = 111 cases, high = 120. Significance determined using log-rank test.

Table 1: Clinicopathological details of the study cohort

	***PADI2* (n)**	***(%)***	***PADI4* (n)**	***(%)***
Total cases:	247		231	
Age at operation (years):				
Median	72		71	
Range	57-88		57-87	
Gender:				
Male	149	60	136	59
Female	98	40	95	41
Histologic tumour type				
Adenocarcinoma	213	86	198	86
Mucinous	31	13	29	13
Unknown	3	1	4	1
Tumour differentiation				
Well	19	8	15	6
Moderate	175	71	169	73
Poor	49	20	43	19
Unknown	4	1	4	2
Tumour site				
Right colon	76	31	75	32
Left colon	55	22	44	19
Rectal	94	38	87	38
Unknown	22	9	25	11
Stage grouping				
I	35	14	30	13
II	95	39	85	37
III	85	34	84	36
IV	28	11	27	12
*Unknown**	4	2	5	2

*Unknown cases were excluded from analysis

medRxiv preprint doi: https://doi.org/10.1101/2020.08.14.20174920; this version posted March 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

1 **Table 2.** Clinicopathological variables for the patient cohort stained for cytoplasmic and nuclear
2 PADI2 and PADI4

PADI2	**Cytoplasmic PADI2 (%)**			**Nuclear PADI2 (%)**		
Variable	**Low PADI2**	**High PADI2**	**χ2 test (*p* Value)**	**Low PADI4**	**High PADI4**	**χ2 test (*p* Value)**
Age						
≤ 72	22 (16.2)	114 (83.8)	0.702	18 (43.9)	23 (56.1)	0.800
> 72	20 (18.0)	91 (82.0)		15 (46.9)	17 (53.1)	
Gender						
Male	24 (16.1)	125 (83.9)	0.644	20 (51.3)	19 (48.7)	0.264
Female	18 (18.4)	80 (81.6)		13 (38.2)	21 (61.8)	
Histologic tumour type						
Adenocarcinoma	34 (16.0)	179 (84.0)	0.634	20 (50.0)	19 (50.0)	0.264
Mucinous	6 (19.4)	25 (80.6)		13 (20.0)	21 (80.0)	
Tumour differentiation						
Well	2 (10.5)	17 (89.5)		3 (42.9)	4 (57.1)	
Moderate	28 (16.0)	147 (84.0)	0.586	24 (50.0)	24 (50.0)	0.571
Poor	10 (20.4)	39 (79.6)		6 (35.3)	11 (64.7)	
Tumour site						
Right colon	14 (18.4)	62 (81.6)		10 (45.5)	12 (54.5)	
Left colon	10 (18.2)	45 (81.8)	0.898	9 (60.0)	10 (40.0)	0.544
Rectal	15 (16.0)	79 (84.0)		12 (42.9)	16 (57.1)	
Stage grouping						
I - II	20 (15.4)	110 (84.6)	0.475	24 (57.1)	18 (42.9)	**0.017***
III - IV	22 (18.8)	95 (81.2)		9 (29.0)	22 (71.0)	
PADI4	**Cytoplasmic PADI4 (%)**			**Nuclear PADI4 (%)**		
Variable	**Low PADI2**	**High PADI2**	**χ2 test (*p* Value)**	**Low PADI4**	**High PADI4**	**χ2 test (*p* Value)**
Age						
≤ 72	35 (27.1)	94 (72.9)	0.827	60 (46.5)	69 (53.5)	0.598
> 72	29 (28.4)	73 (71.6)		51 (50.0)	51 (50.0)	
Gender						
Male	34 (25.0)	102 (75.0)	0.272	62 (25.0)	74 (54.4)	0.370
Female	30 (31.6)	65 (68.4)		49 (31.6)	46 (48.4)	
Histologic tumour type						
Adenocarcinoma	56 (28.3)	142 (71.7)	0.391	93 (47.0)	105 (53.0)	0.370
Mucinous	6 (20.7)	23 (79.3)		16 (55.2)	13 (44.8)	
Tumour differentiation						
Well	4 (26.7)	11 (73.3)	0.691	7 (46.7)	8 (53.3)	0.336
Moderate	44 (26.0)	125 (74.0)		77 (45.6)	92 (54.4)	
Poor	14 (32.6)	29 (67.4)		25 (58.1)	18 (41.9)	
Tumour site						
Right colon	24 (32.0)	51 (68.0)	0.498	41 (54.7)	34 (45.3)	0.240
Left colon	11 (25.0)	33 (75.0)		21 (47.7)	23 (52.3)	
Rectal	21 (24.1)	66 (75.9)		36 (41.4)	51 (58.6)	
Stage grouping						
I - II	29 (25.2)	86 (74.8)	0.400	50 (43.5)	65 (56.5)	0.166
III - IV	35 (30.2)	81 (69.8)		61 (52.6)	55 (47.4)	

3

medRxiv preprint doi: https://doi.org/10.1101/2020.08.14.20174920; this version posted March 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 3. Univariate analysis of cytoplasmic and nuclear PADI2 and PADI4 expression in correlation with tumour markers or treatment using the χ2 or Fisher's exact test.

Variable	**PADI2 cytoplasmic**	**PADI2 nuclear**	**PADI4 cytoplasmic**	**PADI4 nuclear**
	χ2 test (*p* Value)	**χ2 test (*p* Value)**	**χ2 test (*p* Value)**	**χ2 test (*p* Value)**
PADI2 cytoplasmic		0.394	**<0.0001****	**0.048***
PADI2 nuclear	0.394		0.145	0.336
PADI4 cytoplasmic	**<0.0001****	0.145		**<0.0001****
PADI4 Nuclear	**0.048***	0.336	**<0.0001****	
MHC class II in tumour	0.407	0.476	0.398	0.370
Vimentin	1.000	0.094	0.529	0.554
α-Enolase cytoplasmic	**0.001****	0.729	**<0.0001****	0.469
α-Enolase nuclear	0.088	0.670	**0.002****	**0.011***
Distant metastasis	0.687	0.658	0.059	0.143
Nuclear β-catenin	**0.020***	0.988	**0.004****	0.549
P53	0.645	**0.024***	0.922	**0.016***
Ki-67	0.368	0.474	0.224	0.765
Bcl2	0.975	0.764	**0.032***	**0.022***
VEGFA	**0.001***	0.933	0.363	0.267
carcinoembryonic antigen (CEA)	**0.042**	0.243	0.522	0.091
Chemotherapy	0.840	0.500	0.846	0.599
Radiotherapy	0.938	0.351	0.424	0.123

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

medRxiv preprint doi: https://doi.org/10.1101/2020.08.14.20174920; this version posted March 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

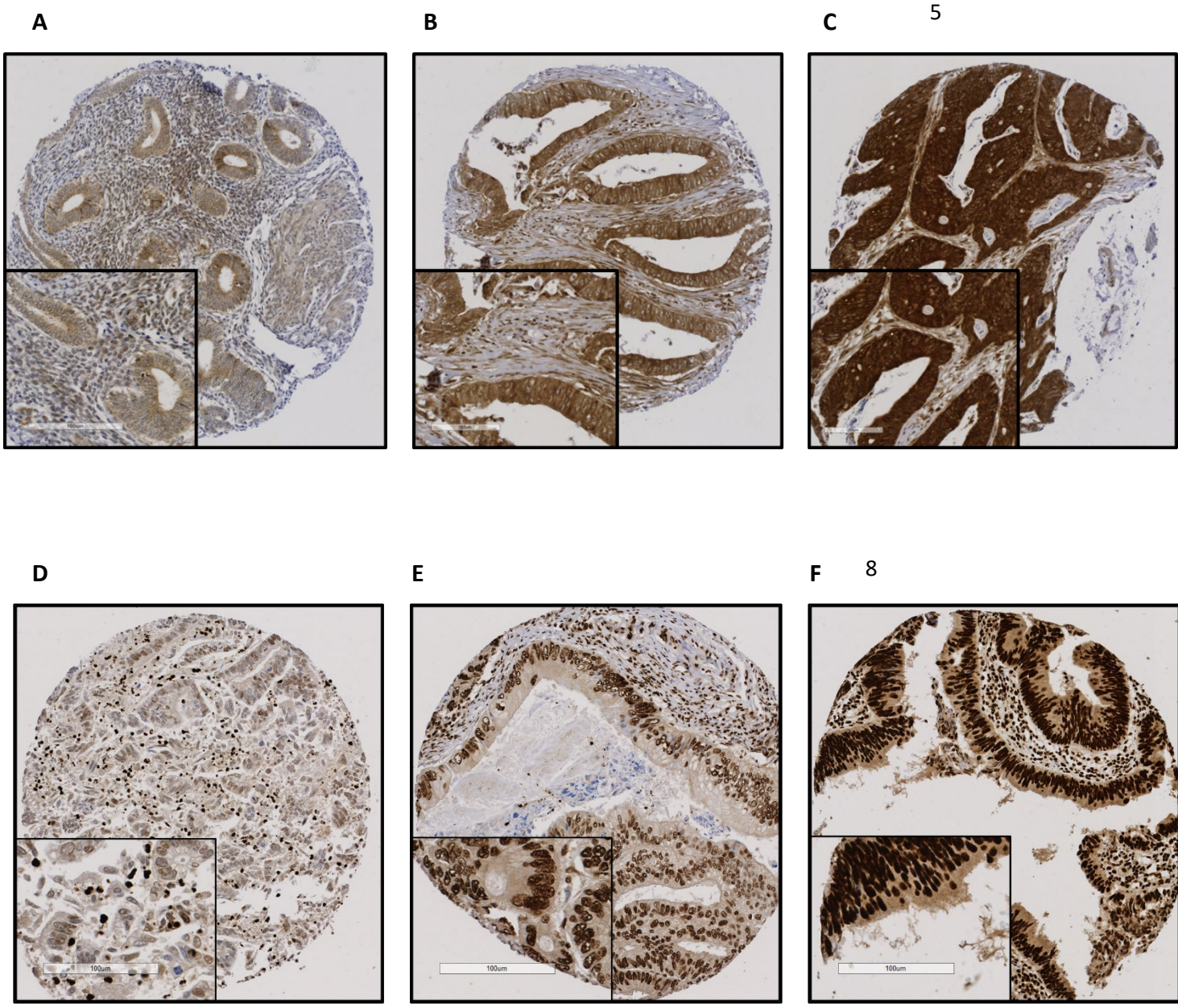
Table 4. Multivariate analysis results (Cox regression model) for TNM stage, Vascular Invasion, Nuclear β-catenin and cytoplasmic and nuclear PADI2 and PADI4 expression.

PADI2		Cytoplasmic PADI2			Nuclear PADI2		
VARIABLE	**CATEGORY**	**HAZARD RATIO (HR)**	**95% CI**	**P VALUE**	**HAZARD RATIO (HR)**	**95% CI**	**P VALUE**
Stage	I and II III and IV	2.290	1.689 to 3.105	**<0.001***	2.651	1.590 to 4.421	**<0.001***
Vascular Invasion	Negative Positive	1.437	0.937 to 2.204	0.097	2.152	0.934 to 4.957	0.072
PADI2 (Cytoplasmic)	Low High	0.458	0.288 to 0.728	**0.001***	N/A		
PADI2 (nuclear)	Low High	N/A			1.677	0.755 to 3.722	0.204
PADI4		**Cytoplasmic PADI4**			**Nuclear PADI4**		
VARIABLE	**CATEGORY**	**HAZARD RATIO (HR)**	**95% CI**	**P VALUE**	**HAZARD RATIO (HR)**	**95% CI**	**P VALUE**
Stage	I and II III and IV	2.238	1.613 to 3.104	**<0.001***	2.303	1.658 to 3.200	**<0.001***
Vascular Invasion	Negative Positive	1.493	0.953 to 2.339	0.080	1.435	0.917 to 2.247	0.114
PADI4 (Cytoplasmic)	Low High	0.629	0.398 to 0.994	**0.047***	N/A		
PADI4 (nuclear)	Low High	N/A			0.727	0.469 to 1.126	0.153

Significant p values are in bold

medRxiv preprint doi: https://doi.org/10.1101/2020.08.14.20174920; this version posted March 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure 1. Representative immunohistochemistry staining of PADI2 (A-C) and PADI4 (D-F) in colorectal tissue microarray images. The level of expression ranged from weak (A, D), moderate (B, E) and strong expression (C, F) according to the H-score.



medRxiv preprint doi: https://doi.org/10.1101/2020.08.14.20174920; this version posted March 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

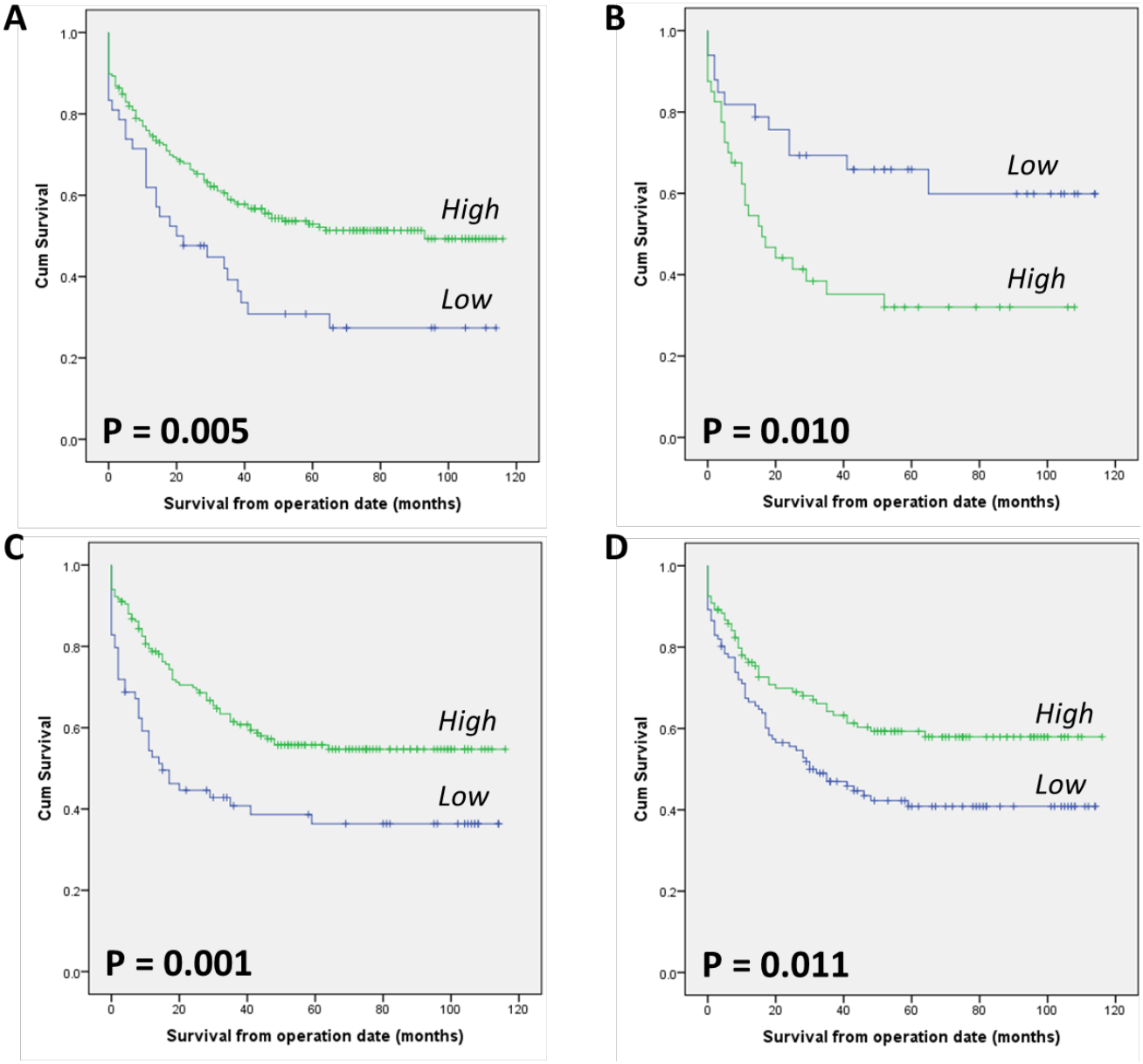



Figure 2. Kaplan-Meier curves for disease survival for CRC patients from the time of tumour resection showing high (green line) or low expression (blue line) of PADI2 or PADI4, (A) cytoplasmic PADI2, low = 42 cases, high = 205 cases (B) nuclear PADI2, low = 33 cases, high = 40 cases (C) cytoplasmic PADI4, low = 64 cases, high = 167 cases (D) nuclear PADI4, low = 111 cases, high = 120. Significance determined using log-rank test.