Expression of *DUSP4* transcript variants as a potential biomarker for colorectal cancer

Tatiana Varela[1,2], Vincent Laizé[1], Natércia Conceição[*,1,3,4], Paulo Caldeira[5], Ana Marreiros[3,4,6], Horácio Guerreiro[5] & M Leonor Cancela[1,3,4,6]

[1]Centre of Marine Sciences, University of Algarve, Faro, Portugal

[2]PhD in Biomedical Sciences, Department of Biomedical Sciences & Medicine, University of Algarve, Faro, Portugal

[3]Department of Biomedical Sciences & Medicine, University of Algarve, Faro, Portugal

[4]Algarve Biomedical Centre, University of Algarve, Faro, Portugal

[5]Department of Gastroenterology, Algarve's University Hospital Centre, Faro, Portugal

[6]Centre for Biomedical Research, University of Algarve, Faro, Portugal

*Corresponding author

Department of Biomedical Sciences and Medicine, University of Algarve, Campus de Gambelas, 8005-139 Faro, Portugal

Email: nconcei@ualg.pt

Tel.: +351 289 800 057

ABSTRACT

Aims

To provide novel data on the expression of *DUSP4* transcripts in colorectal cancer (CRC) tissues and to explore their potential as biomarkers.

Materials & Methods

DUSP4 transcripts expression was determined by qPCR in tissues from 28 CRC patients. Their association with clinicopathological factors and survival analysis was performed. Data from 380 CRC patients available at Cancer Genome Atlas project (TCGA) were also analysed.

Results

All transcripts were overexpressed in CRC tissues. Variant X1 was the most up-regulated and associated with *KRAS* mutations and poorly differentiated tumour. Overexpression of *DUSP4* transcripts could distinguish all tumour stages from normal tissues. Similar results were found in TCGA cohort.

Conclusions

DUSP4 transcripts have the potential to serve as diagnostic biomarkers for CRC, particularly variant X1.

Keywords

Colorectal cancer, DUSP4, Transcript variants, Gene expression, Biomarker.

INTRODUCTION

The prevalence of colorectal cancer (CRC) has increased in the last decades and novel cases of CRC are diagnosed in over 1.3 million people worldwide each year [1]. CRC is currently ranked fourth in terms of the number of deaths caused by cancer [2]. The probability of cure is higher if the cancer is diagnosed and treated in an early stage [3] and it is therefore essential to establish a biological marker that would allow an early detection of the disease, while predicting accurately its prognosis [4]. The pathogenesis of CRC is complex, being influenced by several factors and regulated through different molecular mechanisms [5]. Most of CRCs arise from adenomatous polyps that develop on the inner lining of the colon or rectum as a consequence of the progressive accumulation of genetic (e.g. Kirsten rat sarcoma viral oncogene homolog-(*KRAS*) mutations) and epigenetic alterations leading to the transformation of normal colonic epithelium into an adenoma and ultimately into an adenocarcinoma [6].

Recent data has suggested that Dual-specificity protein phosphatase 4 (DUSP4) plays a central role in cancer development [7] and its expression was abnormally regulated in some tumours such as liver carcinoma, breast cancer, lung cancer, colorectal cancer, pancreatic cancer and glioblastoma [7–9]. DUSP4, also known as mitogen-activated protein kinase phosphatase 2 (MKP2) [10], belongs to a subgroup of the dual specificity phosphatases family [11] responsible for the inactivation of the mitogen-activated protein kinases (MAPKs), through the dephosphorylation of both the phosphoserine/threonine and phosphotyrosine residues of the conserved T-X-Y motif [12]. DUSP4 specifically inactivates the three major MAPKs (ERK, JNK and p38) [13] which are involved in proliferation, differentiation, apoptosis and inflammation processes [14].

DUSP4 expression is up-regulated in colorectal cancer [12], however the contribution of the individual expression of each transcript variant remains unknown. In the present study,

tissues of patients diagnosed with CRC were collected and the relative expression of all *DUSP4* transcript variants was determined. The association of expression levels with clinicopathological factors was also tested to explore the potential of *DUSP4* transcript variants as biomarkers to differentiate tumour from normal tissues and to predict patient prognosis. To further support our findings, data from the Cancer Genome Atlas project (TCGA) dataset were also analysed.

MATERIALS AND METHODS

Patient disposition and clinical data

A total of 28 patients with CRC (10 females and 18 males from 36 to 92 years; see Supplementary Table S1 for the complete data set) and nine healthy patients (2 females and 7 males from 51 to 81 years) from Algarve's University Hospital Centre (CHUA) were enrolled in this study (thereafter referred as CHUA cohort cohort). Their clinical information includes age, gender, tumour location, tumour histology, metastasis, tumour staging, *KRAS* mutations, family history and survival rate. Clinical and pathologic stage of the tumour was determined according to TNM staging criteria and all surgical specimens were reviewed by the Pathology Service of the CHUA. Medical examination revealed malignant neoplasms of the rectum (n = 11), the rectosigmoid junction (n = 3), the sigmoid colon (n = 5), the transverse colon (n = 1), the hepatic flexure (n = 3), the ascending colon (n = 4) and of the cecum (n = 1). The histopathological study showed a majority of well differentiated tumours (n = 13). In most patients no mutations were found in *KRAS* gene (n = 17) and staging revealed cancer at stage I (n = 4), II (n = 9), III (n = 8) and IV (n = 7). All samples were collected with the agreement of the patients from 2014 to 2016, as part of a scientific research project from the Portuguese Society of Gastroenterology and approved by CHUA ethics committee.

Collection of *DUSP4* sequences

Sequences related to human *DUSP4* transcripts and gene were retrieved from GenBank (www.ncbi.nlm.nih.gov) and Ensembl (www.ensembl.org) using on-site Blast facilities. Gene structure was determined using Splign tool available at the National Center for Biotechnology Information (NCBI).

Sampling of CRC tissues

Tumour tissue and non-tumour tissue (sampled 5 cm away of the tumour tissue; referred to as TAN - tumour associated normal tissue) were collected in 28 patients with colorectal cancer. Sampling was performed during colonoscopic examinations at the CHUA. Tissues from colon/rectum of nine healthy individuals were also collected. All tissues were stored in 1 ml of RNAlater (Sigma-Aldrich, St. Louis, MO, USA) and maintained at 4ºC up to 24 h.

RNA extraction and cDNA synthesis

Total RNA was extracted from tissue samples following the protocol described by Chomczynski and Sacchi [15] and further purified using the GeneJET RNA Purification kit (Thermo Scientific, Waltham, MA, USA), according to manufacturer's instructions. RNA quality and integrity were assessed using an Experion Bioanalyzer (Bio-Rad, Hercules, CA, USA). Total RNA (1 µg) was subjected for 30 min at 37ºC to RQ1 RNase-free DNase (Promega, Madison, WI, USA) and cDNA was synthetized using the Moloney murine leukemia virus reverse transcriptase (Invitrogen, Carlsbad, CA, USA) using an oligo-dT primer.

Quantitative real-time polymerase chain reaction (qPCR)

Amplifications by qPCR were performed in a CFX Connect Real-Time PCR thermocycler (Bio-Rad) as follows: an initial denaturation step at 95ºC for 2 min, then 40 cycles of amplification (each cycle is 30 s at 95ºC, 20 s at 63ºC). qPCR reactions (20 µl final volume) were composed of 1X SensiFAST SYBR No-ROX mix (Bioline, London, UK), 0.4 µM of each forward and reverse gene-specific primers and 2 µl of cDNA prepared from patient RNA. Gene-specific primers (Table 1) were designed using PerlPrimer program [16] and only those exhibiting an efficiency of amplification greater than 91% were used.

The relative expression (RE) was calculated using the ΔΔCt method [17] and normalized using expression levels of peptidylprolyl isomerase A (*PPIA*) and ribosomal protein lateral stalk subunit P0 (*RPLP0*) reference genes, previously validated for CRC samples by Krzystek-Korpacka et al. [18], Sørby et al. [19] and Kheirelseid et al. [20]. Because of technical limitations in primer design, expression levels for variant 2 were determined by subtracting the RE value for variant X1 to the combined RE value for variants 2 and X1 (calculated using primers that amplify both variants). Expression levels for calpain-10 (*CAPN10*) – a gene not differentially expressed in tumour versus TAN samples [21] – were also determined and used as control.

TCGA dataset

A total of 380 CRC patients (169 females and 207 males) from the TCGA (cancergenome.nih.gov) database (thereafter referred as TCGA cohort-unpaired samples) were also included in this study. Clinicopathological information (age, gender, staging, *KRAS* mutations and survival rate) and *DUSP4* expression data (level 3 data, RNA-seq Version 2 Illumina; gene-level transcription estimates, as in log2 [x + 1] transformed RNASeq by Expectation Maximization normalized count) were retrieved. The median age at initial

pathologic diagnosis was 65 years (range 40–90 years) and a mutation in *KRAS* gene was found in 50% of patients. The vast majority of patients (n = 249) was classified stage II and III on the TNM classification system. Among these patients, a total of 32 patients have expression data in both tumour and TAN tissues (thereafter referred as TCGA cohort-paired samples). Data available from the TCGA database does not individualize the expression of the three *DUSP4* transcript variants.

Statistical analysis

Expression data was analysed using SPSS Statistics V22.0 (IBM, Armonk, NY, USA) and the non-parametric test Kruskal-Wallis. The associations between *DUSP4* expression and clinicopathological variables were analysed using the Chi-square test (χ^2). For this analysis, the cohort was divided according to *DUSP4* expression (in fold change of tumour over TAN for paired samples) into a high-level subgroup (patients whose expression is above the median) and a low-level subgroup (patient whose expression is below the median). The Kaplan-Meier algorithm was used to build survival curves for these two subgroups and the significance of the differences was evaluated using a log-rank test. Mann-Whitney U and Kruskal-Wallis tests were used to assess the associations between continuous and categorical variables for two or more than two groups, respectively. Some of the outliers are not shown in the graphs to allow a better view of the differences between the groups by maximizing the graphs scale. Differences were considered statistically significant for $p < 0.05$.

RESULTS

Diversity of *DUSP4* transcripts in sequence databases

The analysis of *DUSP4* expression in CRC samples is only useful and informative if all the transcripts that contribute to this expression are identified. GenBank and Ensembl databases

were thoroughly searched for transcript sequences (e.g. Expressed Sequence Tags – ESTs – and annotated mRNAs) related to *DUSP4* and consensus sequences were reconstructed (data not shown). Three alternative spliced transcripts identical to annotated variant 1 (accession no. NM_001394) and variant 2 (accession no. NM_057158), and to predicted variant X1 (accession no. XM_011544428) were identified (Figure 1A). The translation initiation site is in exon 1 for variant 1 and in exon 3 for variant 2. Variant X1 only differs from variant 2 in the 5'-untranslated region (UTR) with the retention of the intron 2. Predicted prior this work, the existence of variant X1 was confirmed by PCR amplification followed by sequencing. Transcripts with alternative 3'-UTRs resulting from the alternative usage of two polyadenylation signals (see Figure 1A) were also identified for variants 1 and 2 (data not shown). At a protein level, the C-terminus, which contains the catalytic domain, is identical in both isoforms but the N-terminus is drastically different; it contains a nuclear localization signal (NLS) and a kinase-interaction motif (KIM) region in variant 1, while both domains are absent in variant 2 (Figure 1B).

DUSP4 **expression in colorectal cancer tissues**

The relative expression of the three *DUSP4* transcripts was determined by qPCR in normal (TAN) and tumour tissue samples of CRC patients and in colorectal tissue samples of healthy individuals (CHUA cohort; Figure 2A). The median value of *DUSP4* expression was always significantly higher ($p<0.001$) in tumour tissues versus TAN tissues, i.e. 1.39 vs. 0.08 for variant X1 (>17 times), 23.26 versus 2.31 for variant 1 (approximately 10 times) and 4.42 vs. 0.58 for variant 2 (>7 times). *DUSP4* expression was also significantly higher for all transcript variants in tumour tissues versus colorectal tissues of healthy individuals ($p=0.021$ for variant 2 and $p=0.001$ for variants 1 and X1). As expected from previous report [21], the expression of *CAPN10* was unaltered in tumour tissues versus TAN and healthy tissues

(Figure 2B), further strengthening the overall analysis of *DUSP4* expression. The analysis of the TCGA cohort also showed a significantly higher expression of *DUSP4* (total) in tumour tissues versus TAN tissues ($p<0.001$) in both paired and unpaired samples (Figure 2C), further strengthening our qPCR data on individualized expression levels.

Expression data also indicated that all *DUSP4* transcripts are expressed in colorectal tissues, although apparently to different extents (variant 1 > variant 2 > variant X1), and are overexpressed in tumour tissues, suggesting that all of them may have a role in the development of CRC.

Correlation between *DUSP4* expression and clinicopathological parameters

To assess the pathological significance of the higher expression of *DUSP4* transcripts in CRC tissues from CHUA cohort, the correlation between overexpression and several clinicopathological parameters was evaluated (Table 2). The overexpression of variant 1 was significantly associated with the family history of cancer ($p=0.014$), while the overexpression of variant X1 was significantly associated with *KRAS* mutations ($p=0.025$), but not with the type of mutation ($p=0.266$). On the contrary, the overexpression of variant 2 was not significantly associated to any of the clinicopathological parameters. The correlation between *DUSP4* expression (total) and clinicopathological parameters was also assessed in the CRC cohort of the TCGA (using unpaired samples since only two patients in the paired samples had information about the existence of *KRAS* mutations) and an increase in *DUSP4* expression was significantly associated with *KRAS* mutations (Chi-square test; $p=0.005$; data not shown).

These associations were further evaluated using other statistic tests. Patients harbouring *KRAS* mutations had significantly higher expression levels of variant X1 when compared to the *KRAS* wild-type patients (Mann-Whitney U test; $p=0.011$; Figure 3A). In TGCA cohort-

unpaired samples, a significantly higher expression of *DUSP4* was also observed in patients with *KRAS* mutations (Mann-Whitney U test; p=0.001; Figure 3A). Furthermore, an overexpression of variant X1 was also significantly associated with a poorly differentiated tumour (Kruskal-Wallis test; p=0.013; Figure 3B). This association could not be evaluated using TCGA data since no information on tumour histology is available in the database.

DUSP4 expression throughout disease progression

The individual expression of *DUSP4* transcripts was found to be statistically higher in tumour vs. TAN tissues independently of the disease stages (I-IV; Figure 4A) and thus can distinguish all tumour stages, including early stage I disease, from TAN tissues. A similar correlation was observed in the TCGA cohort using both paired and unpaired samples (Figure 4B), indicating that *DUSP4* expression – total or individualized by transcripts – may be used as an early diagnostic biomarker for CRC. Although not statistically different from levels in stages I, II and III tumours, levels of *DUSP4* expression determined by qPCR were much higher in stage IV tumours for all transcript variants, suggesting that *DUSP4* expression levels could also be indicative of CRC severity and used to monitor disease progression. This trend was however not observed in the TCGA cohort and this particular correlation remains to be further investigated.

DUSP4 expression and survival

To determine whether *DUSP4* expression could be used to predict the likely outcome of CRC, survival curves were created for each transcript variant following the Kaplan-Meier method (Figure 5). Despite the lack of significant correlation between *DUSP4* expression and overall survival (OS) (p>0.05, log-rank test), the mean survival time was always higher for the group of patients with low expression than for the group of patients with high expression,

and this for all the transcript variants (i.e. 54.7 versus 47.2 months for variant 1, 53.3 versus 48.8 months for variant 2 and 58.1 versus 45.1 months for variant X1, for patients with low and high *DUSP4* expression, respectively; results not shown). This trend was further evidenced when the expression of *DUSP4* transcript variants was considered as a whole (p=0.089; data not shown) and when the expression for both variant 1 and variant X1 was analysed together (p=0.053; data not shown). Survival curves for *DUSP4* expression in TCGA cohorts showed no significant correlation between *DUSP4* expression and overall survival (p>0.05, log-rank test: results not shown) indicating that the correlation between a higher expression of *DUSP4* – total or individualized by transcripts – and a shorter overall survival should be further investigated.

DISCUSSION

There are increasing evidences that DUSP4 may be involved in the pathophysiology of several types of cancer, in particular colorectal cancer. However, current knowledge on DUSP4 cancer-related mechanisms of action is scarce and based mainly on expression data [7,22]. In this regard, deciphering the role of the alternative splicing variants that have been recently identified and still not characterized appears to be essential toward a better understanding of DUSP4 contribution to tumour progression [23]. Novel data on the expression of three *DUSP4* transcript variants in CRC tissues from 28 patients is provided here. This is, as far as we know, the first set of data that assesses the expression of *DUSP4* transcript variants 1, 2 and X1 individually and not globally, in an attempt to elucidate their specific role in CRC development and their suitability as CRC biomarkers. While previous studies have reported an aggregate overexpression of *DUSP4* in CRC tissues/cell lines compared with normal tissues/cell lines [12,24,25], this new data revealed that all the three transcript variants are individually overexpressed in tumour tissues, not only confirming

previous data but also suggesting that, although different in their N-terminal part, the peptides encoded by these variants may all have a role in colorectal cancer. This hypothesis should be further investigated using isoform-specific antibodies (e.g. antibodies developed against the N-terminal region distinct in DUSP4 isoforms) that still remain to be produced.

A correlation between the overexpression of transcript variant X1 and mutations in the proto-oncogene *KRAS* has been evidenced, confirming and amending previous association data [24,26]. This is of particular importance since *KRAS* mutations are found in almost half of the CRC cases, playing a central role in cancer development [26]. The overexpression of transcript variant X1 was also associated with poorly differentiated tumours, which is indicative of a poorer prognosis [27]. While it may seem in contradiction with the data reported by Sim et al. and Saigusa et al. showing that tumour histology was not associated with the aggregate *DUSP4* expression [12,28], our data provides interesting insights on the importance of variant X1 and on the necessity to individually assess the role of the different transcript variants. Variants 1 and 2 account for most of *DUSP4* expression and may have masked the association of variant X1 with tumour histology in these two studies. Altogether, our results suggest that higher expression of transcript X1 may be related with a poorer prognosis, supporting its usefulness as a biomarker for CRC prognosis.

The possible association of *DUSP4* transcript overexpression with a reduced overall survival is supported by the data recently published by Sim et al. and Vriendt et al., [12,25] but is contradicted by those reported by Saigusa et al. suggesting that patients with higher *DUSP4* expression have a better overall survival [28]. Reasons for these discrepancies are currently unknown and must be further investigated if *DUSP4* overexpression is to be used as a prognosis marker for CRC.

Unfortunately, no information on the status of microsatellite instability (MSI) was available for CHUA cohort. This is a parameter that will need to be determined in the future

as the association between MSI status and the overexpression of *DUSP4* transcript variants should be tested. Previous studies showed that overall *DUSP4* overexpression was associated with high MSI in colorectal cancer and may be responsible for the increased cell proliferation [24].

To further explore *DUSP4* as possible biomarker for CRC we analysed and compared its expression in each disease stage with the TAN tissues. *DUSP4* transcript variants overexpression in stage-related groups of patients revealed its potential as CRC diagnostic biomarker. This study showed that *DUSP4* overexpression is an early event in CRC and therefore, can be used as a diagnostic biomarker, even for early stage of the disease since it is capable of distinguish all tumour stages, including early stage I disease, from TAN tissues. In this case, the aggregated expression of *DUSP4* transcripts may be used since all variants showed the same profile of overexpression across tumour stages.

Because the CHUA cohort used in this study has a limited sample size and because all samples were collected from the same hospital, the TCGA database was used to complement and support some of our findings with data from a larger and diversified cohort. Although TCGA data does not individualize the expression of *DUSP4* transcripts, a higher expression of *DUSP4* in CRC tissues was significantly associated with all tumour stages and with the presence of *KRAS* mutations in CRC patients.

A larger cohort of CRC patients and probably more quantitative methods should also be used to further confirm the specific potential of variant X1 overexpression as a biomarker. While it appeared to be the least expressed (i.e. lower levels) of the three variants, it was the most upregulated transcript but also the variant exhibiting more associations with CRC clinicopathological parameters (e.g. *KRAS* mutations and tumour histology). Because variant X1 encodes for a N-terminus truncated protein (it misses a large part of the MAPK binding domain, the kinase interaction motif and one of the two nuclear localization signals present in

the DUSP4 archetype protein coded by variant 1), future studies should aim at the better characterization of this variant. In this regard, variant X1 expression should be determined in a variety of adult tissues (sampled from healthy and diseased patients) and throughout development, factors involved in transcriptional and post-transcriptional regulation of variant X1 should be identified, and substrate specificity should also be studied. At a cellular level, it would be interesting to gain insights on the role of *DUSP4* transcript variants in the proliferation, invasion and migration of cancer cells. In this regard, gain- and loss-of-function studies in CRC-related cell lines such as HT-29 should be performed.

It is still controversial whether DUSP4 is a tumour suppressor or a tumour promoter in CRC. In favour of a tumour promoter activity, Gröschl et al. reported that *DUSP4* overexpression increased the proliferation and survival of colorectal cancer cell lines, and altered the expression of mismatch repair genes (down-regulation) and growth promoting genes (up-regulation) [24]. The overall survival of patients with high *DUSP4* expression was also significantly reduced compared with OS of patients with low *DUSP4* expression [25]. In favour of a tumour suppressor activity, Saigusa et al. reported that *DUSP4* expression was negatively associated with characteristic factors of tumour progression [28]. Because *DUSP4* overexpression was found to be possibly associated with a shorter OS but also to be a marker of adverse prognosis, our data supports the hypothesis that DUSP4 may function as a tumour promoter in colorectal cancer.

CONCLUSION

In conclusion, our study suggests that expression of *DUSP4* transcripts, especially variant X1, could be a potential biomarker for CRC, facilitating its earlier diagnosis and future studies should aim to confirm it. A more detailed investigation with larger sample size will be necessary to better understand the role of each *DUSP4* transcript variant in CRC.

SUMMARY POINTS

- *DUSP4* transcript variants are individually overexpressed in colorectal cancer tissues
- *DUSP4* overexpression is an early event in CRC
- Higher expression of *DUSP4* variant X1 was associated with *KRAS* mutations and poorly differentiated tumours
- Expression of *DUSP4* transcripts, especially variant X1, could be a potential biomarker for CRC

Compliance with Ethical Standards

Conflict of Interest: The authors declare that they have no conflict of interest.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent: Informed consent was obtained from all individual participants included in the study.

REFERENCES

1. Haggar FA, Boushey RP. Colorectal cancer epidemiology: Incidence, mortality, survival, and risk factors. *Clin. Colon Rectal Surg.* 6(212), 191–197 (2009).
2. Aran V, Victorino AP, Thuler LC, Ferreira CG. Colorectal cancer: Epidemiology, disease mechanisms and interventions to reduce onset and mortality. *Clin. Colorectal Cancer*. 15(3), 195–203 (2016).
3. Chandran SP, Natarajan SB, Chandraseharan S, Mohd Shahimi MSB. Nano drug delivery strategy of 5-fluorouracil for the treatment of colorectal cancer. *J. Cancer Res. Pract.* 4(2), 45–48 (2017).
4. Tanaka T, Tanaka M, Tanaka T, Ishigamori R. Biomarkers for colorectal cancer. *Int. J. Mol. Sci.* 11(9), 3209–3225 (2010).

5. Colussi D, Brandi G, Bazzoli F, Ricciardiello L. Molecular pathways involved in colorectal cancer: Implications for disease behavior and prevention. *Int. J. Mol. Sci.* 14(8), 16365–16385 (2013).

6. Pritchard CC, Grady WM. Colorectal cancer molecular biology moves into clinical practice. *Gut*. 60(1), 116–29 (2011).

7. Low HB, Zhang Y. Regulatory roles of MAPK phosphatases in cancer. *Immune Netw.* 16(2), 85–98 (2016).

8. Bermudez O, Pagès G, Gimond C, *et al.* The dual-specificity MAP kinase phosphatases: Critical roles in development and cancer. *Am. J. Physiol. cell Physiol.* 299, 189–202 (2013).

9. Nunes-Xavier C, Romá-Mateo C, Ríos P, *et al.* Dual-specificity MAP kinase phosphatases as targets of cancer treatment. *Anticancer Agents Med. Chem.* 11(1), 109–132 (2011).

10. Waha A, Felsberg J, Hartmann W, *et al.* Epigenetic downregulation of mitogen-activated protein kinase phosphatase MKP-2 relieves its growth suppressive activity in glioma cells. *Cancer Res.* 70(4), 1689–1699 (2010).

11. Sloss CM, Cadalbert L, Finn SG, Fuller SJ, Plevin R. Disruption of two putative nuclear localization sequences is required for cytosolic localization of mitogen-activated protein kinase phosphatase-2. *Cell. Signal.* 17(6), 709–716 (2005).

12. Sim J, Yi K, Kim H, *et al.* Immunohistochemical expression of dual-specificity protein phosphatase 4 in patients with colorectal adenocarcinoma. *Gastroenterol. Res. Pract.* 2015, 283764 (2015).

13. Wang Y, Ji Q. DUSP4/MKP2 overexpression is associated with BRAF V600E mutation and aggressive behavior of papillary thyroid cancer. *Onco Targets Ther.* 9, 2255–2263 (2016).

14. Lawan A, Al-Harthi S, Cadalbert L, *et al.* Deletion of the dual specific phosphatase-4 (DUSP-4) gene reveals an essential non-redundant role for MAP kinase phosphatase-2 (MKP-2) in proliferation and cell survival. *J. Biol. Chem.* 286(15), 12933–12943 (2011).

15. Chomczynski P, Sacchi N. The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162(1), 156–159 (1987).

16. Marshall OJ. PerlPrimer: Cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics*. 20, 2471–2472 (2004).

17. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* 3(6), 1101–1108 (2008).

18. Krzystek-korpacka M, Diakowska D, Bania J, Gamian A. Expression stability of common housekeeping genes is differently affected by bowel inflammation and cancer: Implications for finding suitable normalizers for inflammatory bowel disease studies. *Inflamm. Bowel Dis.* 20(7), 1147–1156 (2014).

19. Sørby LA, Andersen SN, Bukholm IRK, Jacobsen MB. Evaluation of suitable reference genes for normalization of real-time reverse transcription PCR analysis in

colon cancer. *J. Exp. Clin. Cancer Res.* 29(1), 144 (2010).

20. Kheirelseid EAH, Chang KH, Newell J, Kerin MJ, Miller N. Identification of endogenous control genes for normalisation of real-time quantitative PCR data in colorectal cancer. *BMC Mol. Biol.* 11, 12 (2010).

21. Pitule P, Vycital O, Bruha JAN, *et al.* Differential expression and prognostic role of selected genes in colorectal cancer patients. *Anticancer Res.* 33, 4855–4865 (2013).

22. Owens DM, Keyse SM. Differential regulation of MAP kinase signalling by dual-specificity protein phosphatases. *Oncogene*. 26(22), 3203–3213 (2007).

23. Cadalbert LC, Sloss CM, Cunningham MR, *et al.* Differential regulation of MAP kinase activation by a novel splice variant of human MAP kinase phosphatase-2. *Cell. Signal.* 22(3), 357–365 (2010).

24. Gröschl B, Bettstetter M, Giedl C, *et al.* Expression of the MAP kinase phosphatase DUSP4 is associated with microsatellite instability in colorectal cancer (CRC) and causes increased cell proliferation. *Int. J. Cancer Expr.* 132(7), 1537–1546 (2013).

25. De Vriendt V, De Roock W, Fabio A, *et al.* DUSP4 expression identifies a subset of colorectal cancer tumors that differ in MAPK activation, regardless of the genotype. *Biomarkers*. 18(6), 516–524 (2013).

26. Cagnol S, Rivard N. Oncogenic KRAS and BRAF activation of the MEK/ERK signaling pathway promotes expression of dual-specificity phosphatase 4 (DUSP4/MKP2) resulting in nuclear ERK1/2 inhibition. *Oncogene*. 32(5), 564–576 (2013).

27. Ishihara S, Watanabe T, Akahane T, *et al.* Tumor location is a prognostic factor in poorly differentiated adenocarcinoma, mucinous adenocarcinoma, and signet-ring cell carcinoma of the colon. *Int. J. Colorectal Dis.* 27(3), 371–379 (2012).

28. Saigusa S, Inoue Y, Tanaka K, *et al.* Decreased expression of DUSP4 is associated with liver and lung metastases in colorectal cancer. *Med. Oncol.* 30(3), 620–627 (2013).

Table 1. List of the qPCR primers used in the present study.

Gene name	GenBank Accession number	Sequence (5' to 3')
DUSP4 variant 1	NM_001394	F: CACCGACATCTGCCTGCTCAAAGG R: CACTGCCGAGGTAGAGGAAGGGAAGG
DUSP4 variant 2 and X1	NM_057158 and XM_011544428	F: AACGAAGAGGAAGCACGATGGGAAGAA R: CATAGCCGCCTTGAGACTCTGAACACA
DUSP4 variant X1	XM_011544428	F: TTTCCCTGTCCCTCGTGAATCCCGTTG R: GCACCTCGCCACCCACCAGAG
PPIA	NM_021130	F: GCAAATGCTGGACCCAACACAAAT R: AATGGTGATCTTCTTGCTGGTCTTG
RPLP0	NM_001002	F: GCTGCTGAACATGCTCAACATCTCC R: CAACATTGCGGACACCCTCCA
CAPN10	NM_023083	F: CTCCTTCACCTGTCGCATTTGGC R: TGGACCTTGGCGTAGACCTTTTCC

Table 2. Correlation (Chi-square test) between the relative expression of *DUSP4* transcript variants and clinicopathological factors in colorectal cancer (CHUA cohort). n=14 for low level; n=14 for high level.

Clinicopathological factors	Relative expression								
	Variant 1			Variant 2			Variant X1		
	Low level	High level	*p*	Low level	High Level	*p*	Low level	High level	*p*
Age			0.316			0.135			0.316
≤50	1	0		0	1		0	1	
51-60	1	2		1	2		2	1	
61-70	5	2		6	1		5	2	
71-80	2	6		2	6		2	6	
≥81	5	4		5	4		5	4	
Gender			0.430			0.115			0.430
M	10	8		11	7		10	8	
F	4	6		3	7		4	6	
Tumour location			0.585			0.719			0.467
Cecum	1	0		1	0		1	0	
Ascending colon	1	3		2	2		2	2	
Hepatic flexure	2	1		1	2		0	3	
Transverse colon	0	1		0	1		0	1	
Sigmoid colon	2	3		2	3		3	2	
Rectosigmoid junction	1	2		1	2		2	1	
Rectum	7	4		7	4		6	5	
Tumour histology			0.711			0.823			0.065
Well differentiated	8	5		7	6		7	6	
Moderately differentiated	4	5		5	4		7	2	
Poorly differentiated	1	2		1	2		0	3	
Mucinous	0	1		0	1		0	1	
Mucinous well differentiated	1	1		1	1		0	2	

Metastasis			1			1			0.403
Yes	4	4		4	4		3	5	
No	10	10		10	10		11	9	
Metastasis localization			0.736			0.736			0.380
Hepatic	2	2		2	2		2	2	
Pulmonar	1	0		1	0		1	0	
Hepatic/ Pulmonar	1	1		1	1		0	2	
Peritoneal	0	1		0	1		0	1	
Without metastasis	10	10		10	10		11	9	
Tumour staging			0.650			0.968			0.222
Stage I	2	2		2	2		1	3	
Stage II	6	3		5	4		5	4	
Stage III	3	5		4	4		6	2	
Stage IV	3	4		3	4		2	5	
***KRAS* mutations**			0.081			0.345			0.025*
Yes	3	7		4	10		2	8	
No	11	6		6	7		11	6	
Type of *KRAS* mutations			0.488			0.777			0.266
Exon2 Gly12Val	1	1		1	1		1	1	
Exon 2 Gly12Asp	1	3		2	2		1	3	
Exon 2 Gly12Cys	0	1		0	1		0	1	
Exon 4 Ala46Thr	1	1		1	1		0	2	
Exon 2 Gly13Asp	0	1		0	1		0	1	
None	11	6		10	7		11	6	
Survival rate (month)			0.478			0.881			0.421
≤24	3	6		4	5		4	5	
24-36	7	5		6	6		5	7	
>36	4	3		4	3		5	2	
Deceased			0.663			0.663			0.190
Dead	3	4		3	4		2	5	
Alive	11	10		11	10		12	9	
Previous pathologies			0.622			0.622			0.139
Yes	11	12		12	11		13	10	
No	3	2		2	3		1	4	
Familiar history			0.014*			0.622			0.622
Yes	5	0		3	2		2	3	
No	9	14		11	12		12	11	

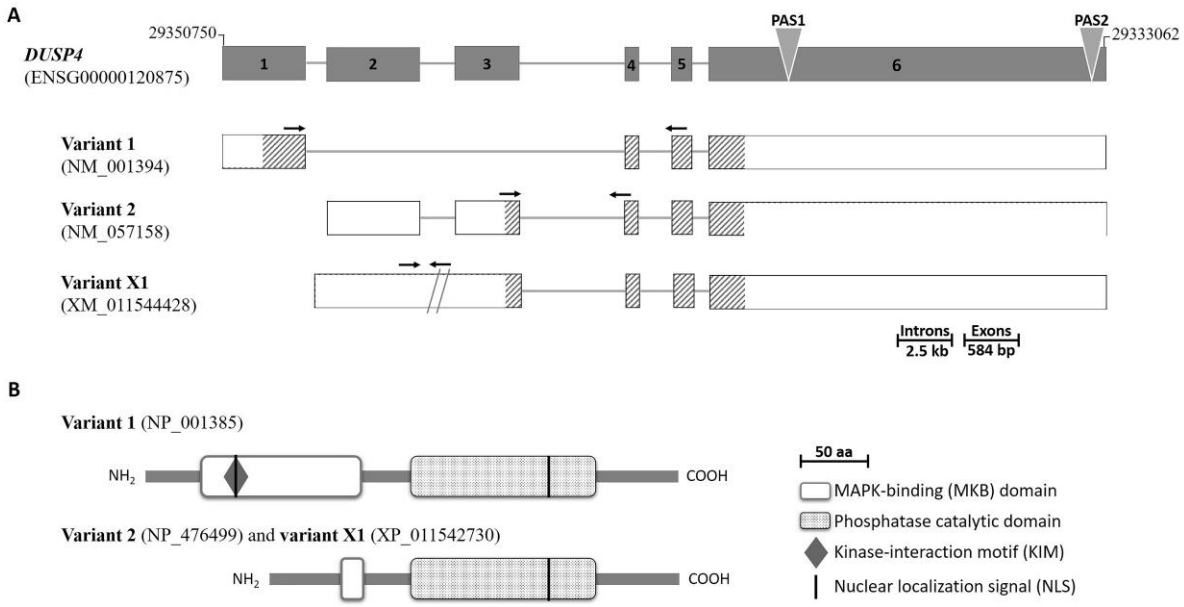

Figure 1. *DUSP4* transcript diversity (A) and deduced protein (B). *DUSP4* gene has six exons and five introns. The three alternatively spliced transcripts previously reported (annotated variants 1 and 2 and predicted variant X1) are indicated below the gene structure. In gene structure, exons are represented by solid dark boxes and introns by solid lines. The numbers delimiting the first and last exons indicate their position within chromosome 8 (Ensembl GRCh38 coordinates). Arrows indicate position of the designed primers. PAS, polyadenylation signal. In transcript structure, pattern-filled boxes indicate the coding region and white boxes the untranslated regions (5' and 3' UTRs).

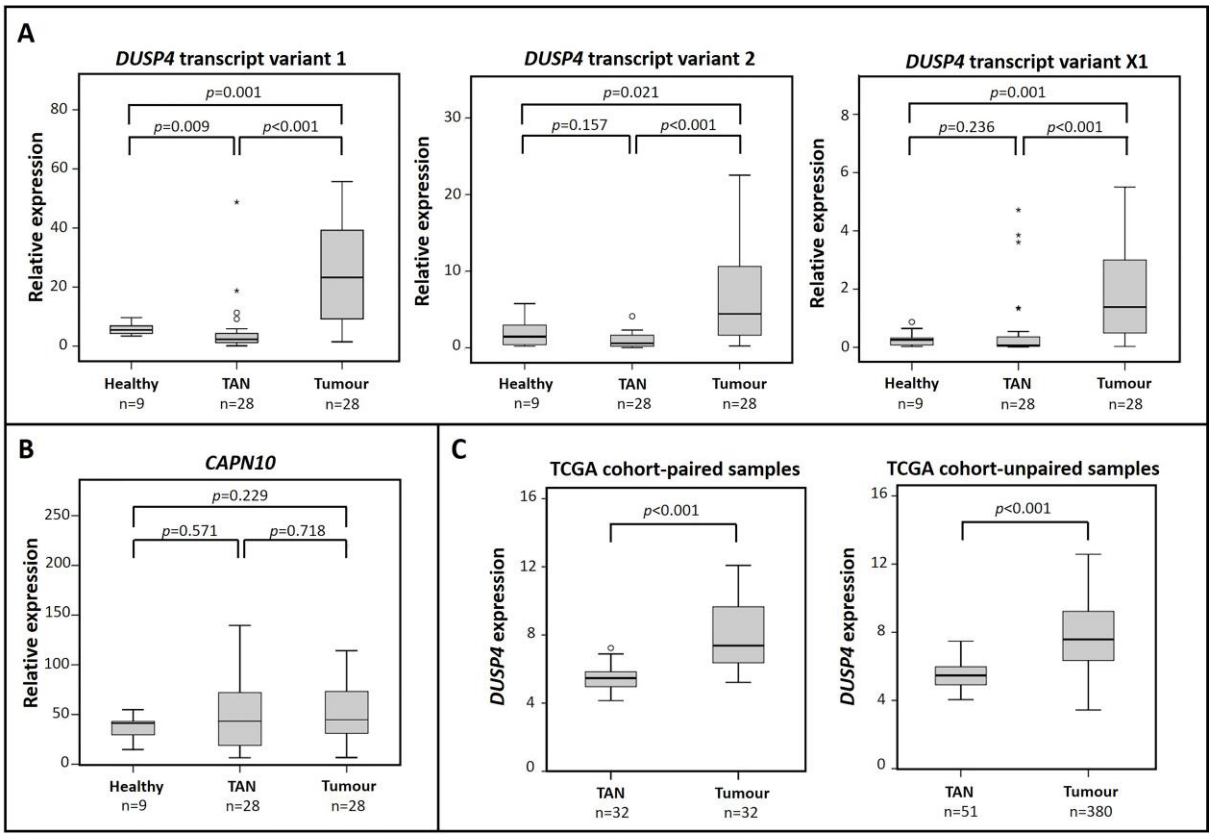

Figure 2. *DUSP4* transcript variants are over-expressed in human CRC tissues. Relative expression of (A) *DUSP4* transcript variant 1, variant 2 and variant X1 and (B) *CAPN10* in healthy (n=9), TAN and tumour (n=28) tissues. (C) *DUSP4* expression in TCGA cohort-paired samples (n=32) and TCGA cohort-unpaired samples (n=380). The black line in boxplots indicates the median. Differences with $p<0.05$ were considered statistically significant. *Stars* and *circles* represent severe and moderate outliers, respectively.

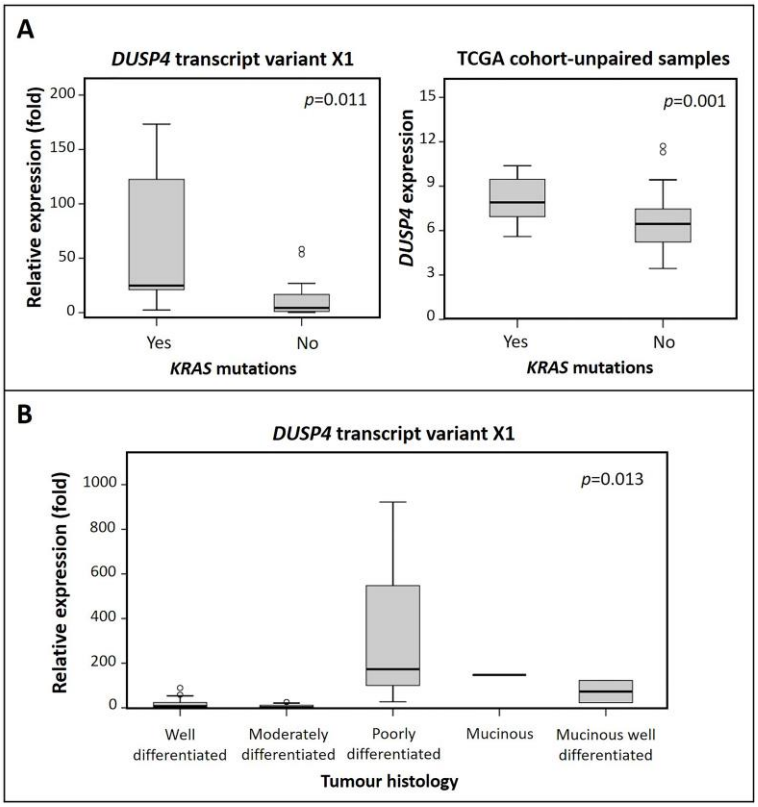

Figure 3. *DUSP4* overexpression was associated with *KRAS* mutations and poorly differentiated tumours. (A) Association between the expression of *DUSP4* and *KRAS* mutations in CHUA (*left*) and TCGA (*right*) cohorts. (B) Association between *DUSP4* transcript variant 1 and tumour histology in CHUA cohort. CRC associations were analysed using the Mann-Whitney U test (A) and the Kruskal-Wallis test (B). Differences with $p<0.05$ were considered statistically significant. *Circles* represent moderate outliers.

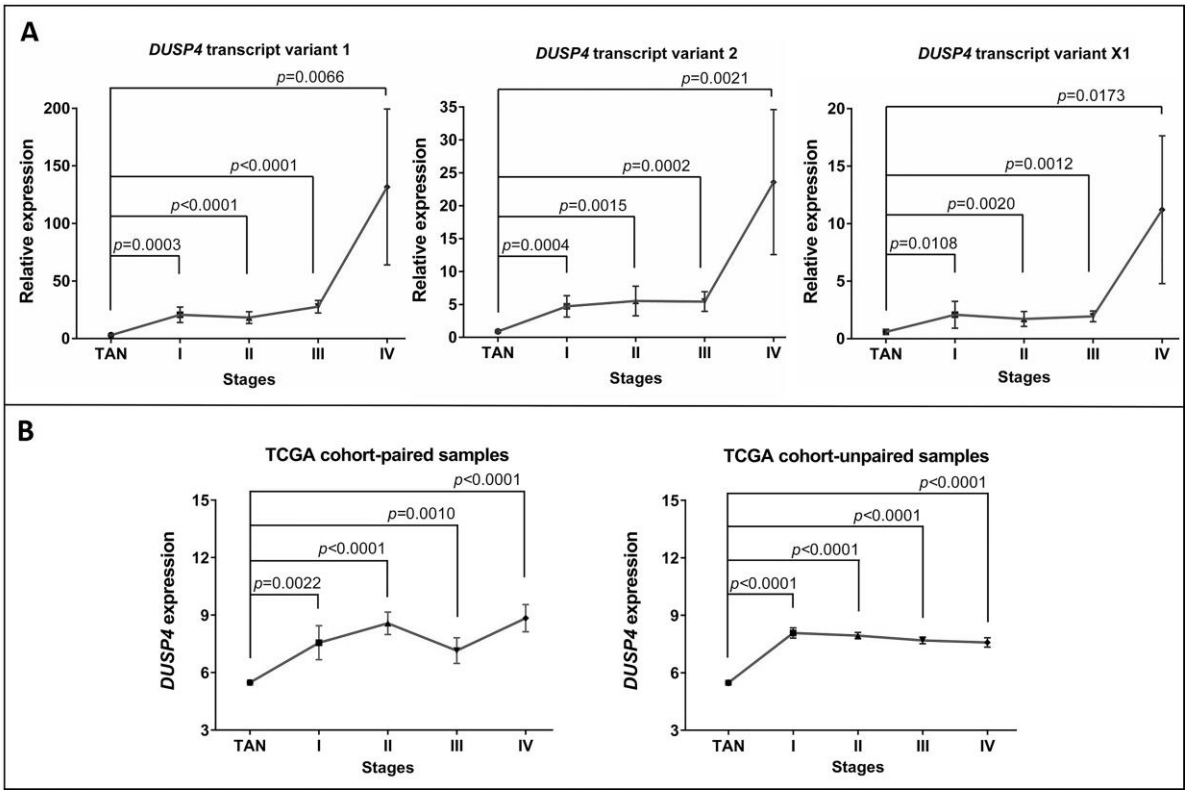

Figure 4. *DUSP4* overexpression is an early event in CRC. (A) Expression of *DUSP4* transcript variant 1, variant 2 and variant X1 in CHUA cohort and (B) expression of *DUSP4* (total) in paired and unpaired samples of the TCGA cohort. Data are presented as mean ± standard error of the mean (SEM). Statistical analysis was performed using the Mann–Whitney U test. Differences with $p<0.05$ were considered statistically significant.

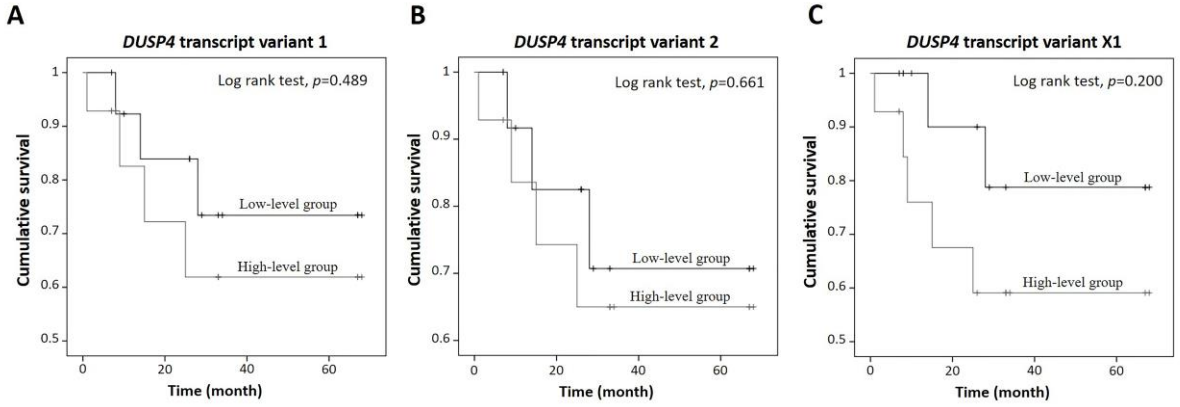

Figure 5. Cumulative overall survival (OS) curves according to the relative expression of *DUSP4* transcript variants 1 (A), variant 2 (B) and variant X1 (C). Kaplan-Meier method combined to a log-rank test was used to create and analyse OS curves. Values of relative expression above the median constituted the high-level group, whereas values below the median constituted the low-level group.