The regulatory role of the CD2AP/TKS4 interaction in EMT and its potential as a biomarker for colon cancer

Anita Kurilla [1*], Loretta László [1,2], Tamás Takács [1,2], Álmos Tilajka [1,2], Laura Lukács[1], Julianna Novák[1], Rita Pancsa[1], László Buday [1], Virag Vas [1*]

Abstract

Colon cancer is a leading cause of death worldwide. Although several molecular biomarkers have been developed for colon cancer, more sensitive and specific methods are needed for its early detection. The TKS4 and CD2AP scaffold proteins have been linked to dynamic actin assembly-related processes, such as cell migration and epithelial-mesenchymal transition (EMT), although their co-instructive role during colon cancer development remains unknown. Therefore, this study was designed to investigate the interdependence of TKS4 and CD2AP during colon cancer progression as well as their involvement in EMT. This is the first study demonstrating a TKS4-CD2AP protein-protein interaction in vitro and in vivo. We describe that CD2AP and TKS4 interdependently induce a mesenchymal transition-like process and impair the migration ability of HCT116 colon cancer cells. Here we show that CD2AP is a novel TKS4 interacting partner, and together they form a scaffold complex that potentially regulates EMT-related signalling. Based on our analysis of the relative TKS4 and CD2AP expression levels, we propose that their levels might cooperatively serve as a reliable biomarker for colon cancer detection.

Keywords: scaffold proteins, CD2AP, TKS4, epithelial mesenchymal transition, colon cancer biomarker

* Correspondence: kurilla.anita@ttk.hu; vas.virag@ttk.hu

[1] Institute of Enzymology, Research Centre for Natural Sciences, 1117 Budapest, Hungary;

Full list of author information is available at the end of the article.

1. Introduction

Colon (colorectal) cancer is one of the most common cancer types and is the second leading cause of cancer death worldwide. Mortality and survival rates mostly depend on appropriate screening and treatment techniques. As these methods have improved in the last decade, death rates have dropped significantly (seer.cancer.gov). Currently, several blood-based and molecular biomarkers have been developed and used clinically, including the presence of *KRAS* mutations as well as the presence of circulating tumour DNA (ctDNA) or CTCs circulating tumour cells (CTCs). However, more sensitive and specific methods are needed to detect colon cancer at its early stages [1].

Several molecular events are known to drive colon cancer formation, including the accumulation of oncogenic mutations, development of drug resistance, and initiation of the cellular invasion cascade, leading to the appearance of CTCs and metastasis [2]. However, most of the molecular mechanisms underlying oncogenesis are not yet fully understood; therefore, identifying novel molecular players that regulate colon cancer invasiveness will help us understand its formation and progression, ultimately contributing to the discovery of novel biomarkers and therapeutic targets.

CD2-associated protein (CD2AP) and tyrosine kinase substrate with four SH3 domains (TKS4) have recently emerged as potential cancer formation-regulating factors. These proteins, which lack enzymatic activity, are scaffolds/adaptors that serve as molecular hubs within cells by binding and releasing multiple signalling molecules via phosphorylation-dependent or -independent interactions mainly mediated by SH3 domains. CD2AP was originally implicated in dynamic actin

remodelling via direct binding of filamentous actin. Later, it was found that CD2AP also plays a regulatory role in epithelial cell junction formation [3], and it was also linked to gastric cancer development as it can inhibit metastasis by promoting cellular adhesion and cytoskeleton assembly [4].

TKS4 also serves as a docking site for molecules involved in cell movement, e.g., cortactin or N-WASP, and it regulates podosome formation in normal cells and invadopodia formation in cancer cells [5][6]. The involvement of TKS4 in EGF-EGFR-SRC signalling has been previously described, and regulatory roles in mesenchymal stem cell differentiation, adipose tissue, and bone homeostasis have recently been discovered. Recent progress in TKS4-related research has revealed that TKS4 might also be involved in several cancer types, including breast cancer [7], melanoma [6], acute myeloid leukaemia [8], and prostate cancer [9]. Moreover, our group also demonstrated that the lack of TKS4 in a colon cancer cell line leads to EMT-like processes [10].

Here we show that CD2AP and TKS4 can form a complex in several different cancer cell lines. The fact that both proteins have been implicated in colon cancer cell biology raises the possibility that their interaction is fundamental to colon cancer development; therefore, we analysed both the individual and relative expression levels of CD2AP and TKS4 in colon cancer samples to shed light on the potential effects of their interplay during cancer development. Moreover, we used siRNA-based silencing to downregulate their transcript levels in HCT116 colon cancer cells and measured the resulting changes in EMT process-related protein expression levels and cell migration potential.

We found that CD2AP and TKS4 individually and interdependently induce mesenchymal transition and impair colon cancer cell migration, suggesting that the lack of these proteins provokes partial EMT in cancer cells. We found that the ratio of CD2AP and TKS4 expression levels has an instructive role in tumour phenotypes and that CD2AP/TKS4 ratios are outstanding biomarkers for the identification of tumorous colon tissue.

2. Materials and Methods

2.1. Cell culture

All cell lines were obtained from ATCC and were tested for mycoplasma contamination using MycoAlert Mycoplasma Detection Kit (Lonza, Basel, Switzerland). HCT116 (RRID:CVCL_0291) cells were grown in McCoy's 5A Medium (Thermo Fisher Scientific, 16600082). DLD-1 (RRID:CVCL_0248) cells were grown in RPMI medium. MCF-7 (RRID:CVCL_0031), HEK293 (RRID:CVCL_0045), HPAC (RRID:CVCL_3517), and NIH3T3 (RRID:CVCL_0594) cells were grown in DMEM. All media were supplemented with 10% FBS (Thermo Fisher Scientific, 26400044) and penicillin/streptomycin, and cells were incubated at 37 °C in 5% CO_2.

2.2. siRNA and DNA plasmid transfection

Small interfering RNA (siRNA) transfection was performed with Lipofectamine RNAiMax Transfection Reagent (Thermo Fisher Scientific, 13778030) according to the manufacturer's instructions. SiRNAs targeting CD2AP and TKS4 were purchased from Thermo Fisher Scientific and used at 10 nM. After transfection, cells were incubated for 48 h prior to RNA or protein isolation.

For DNA plasmid transfection, X-tremeGENE Hp DNA Transfection Reagent was used (Roche) following the manufacturer's instructions. Cells were incubated with transfection reagent for 48 h before cell lysis.

2.3. RNA isolation and quantitative real-time polymerase chain reaction (qRT-PCR)

RNA was extracted using Direct-zol RNA Miniprep Kit (Zymo Research). First Strand cDNA Synthesis Kit (Roche) and Fast SYBR Green Master Mix (Thermo Fisher Scientific) were used in qRT-PCR. Data were normalized to the *GAPDH* and *PUM1* transcript levels, and the $2^{-\Delta\Delta Ct}$ method was used to calculate relative gene expression levels. Primers used in this study are listed in Supplementary Table 1.

2.4. Western blotting and immunoprecipitation

For protein extraction, cells were first washed with ice-cold phosphate-buffered saline (PBS). Next, ice-cold harvest buffer (30 mM Tris pH 7.5, with 100 mM NaCl, 1% Triton X-100, 10 mM NaF, 1 mM Na_3VO_4, 1 mM EGTA, 2 mM 4-nitrophenyl phosphate, 10 mM benzamidine, 1 mM phenylmethylsulphonyl fluoride, 25 μg/mL Pepstatin A, 25 μg/mL trypsin inhibitor, and 25 μg/mL aprotinin) was added to the cells. Lysates were centrifuged at 14,000 rpm, 10 min at 4 °C.

Immunoprecipitation was performed using Protein-A Sepharose from *Staphylococcus aureus* (Merck Millipore) and anti-TKS4 antibody [11]. Cell lysates, agarose beads, and anti-TKS4 antibody were incubated for 1 hour at 4 °C and then washed three times with PBS+0.4% Triton-X at 14000 rpm for 1 min at 4 °C.

Sample loading buffer was added to the supernatants, followed by incubation at 95 °C for 5 min. Each protein sample was subjected to 10% SDS-PAGE gel (BioRad).

Nitrocellulose membranes were blocked and incubated for 60 min with anti-TKS4 (1:5000) and anti-CD2AP (1:2000) antibody at room temperature or overnight at 4 °C with anti-E-cadherin, or anti-vimentin antibody. After three washes, the membranes were incubated for 60 min with horseradish peroxidase-conjugated secondary antibody (GE Healthcare) and then washed four times for 15 min each. The proteins of interest were visualized using enhanced chemiluminescence (ECL) detection reagents (Amersham, Little Chalfont, UK). Chemiluminescent imaging was performed with a ChemiDoc MP system (Bio-Rad). The analysed protein levels were normalized to total protein staining.

2.5. Mass spectrometry analyses

After immunoprecipitation, samples were loaded onto 10% SDS-PAGE gels. Next, proteins were visualized via Coomassie Brilliant Blue dye, and gels were fractionized horizontally based on molecular weight (300-100 kDa, 100-50 kDa, 50-25 kDa, 25-10 kDa). Mass spectrometry analyses were performed by UD-GenoMed Medical Genomic Technologies Kft.

2.6. Wound healing assays

HCT116 cells were forward-transfected with siRNAs or plasmid DNA. Next, $5x10^5$ cells were cultured in Ibidi Culture-Insert well. After 24 hours, cells reached confluency and the silicone inserts were removed. The dishes were filled with McCoy's Medium, and images were acquired every 24 hours.

2.7. DuoLink Proximity Ligation Assays

In situ protein-protein interactions between CD2AP and TKS4 were detected via proximity ligation assays (Duolink Mouse/Rabbit, DUO92101−1KT, Sigma-Aldrich).

The DPLA probe anti-rabbit plus binds to the TKS4 primary antibody (HPA036471, Sima Aldrich, 1:100 dilution), whereas the DPLA probe anti-mouse minus binds to the CD2AP antibody (clone 2A2.1, Millipore, 1:250 dilution). The DPLA secondary antibodies generate a signal only when the two DPLA probes interact, indicating a CD2AP-TKS4 interaction. HCT116 cells were plated in 12-well Ibidi removable silicone cultivation chambers and cultured for 2 days. Next, the cells were fixed in 4% paraformaldehyde-PBS for 15 min on ice and further processed according to the manufacturer's instructions: permeabilization with 0.1% Triton X-100 dissolved in PBS for 5 min and blocking with Duolink Blocking Solution for 30 min at 37 °C. The secondary antibodies were conjugated to oligonucleotides that form a closed DNA circle via base pairing when the antibodies are in close proximity. Signal detection was conducted by ligation using Duolink ligase and rolling circle amplification using DNA polymerase and fluorescently labelled nucleotides hybridized into the amplification products. The amplified fluorescent DNA resulted in bright red dots in the samples, with each dot representing an individual CD2AP/TKS4 interaction event. Control samples were processed similarly without addition of the primary antibodies. Staining and image acquisition were performed as previously described [10]. Briefly, nuclei were visualized via DAPI staining (with DuoLink mounting medium), and images were acquired on a Zeiss LSM710 inverted confocal microscope with a 40× objective (Carl Zeiss).

2.8. Peptide array

PepStar Peptide Microarray experiments were performed based on the method of Harnos and colleagues (2018) with some modifications [12]. Microarrays were purchased from JPT (JPT Peptide Technologies GmbH). The TKS4 protein sequence (A1X283) was printed on a glass slide (25x75 mm) in the form of 15-mer peptides

with 11 overlapping residues. The peptide microarrays were printed in three identical subarrays. Several controls were also printed on the glass slides to exclude non-specific binding events or to confirm positive signals (Myc-tag, anti-Myc antibody, human serum albumin, human IgG, rabbit IgG). When screening the binding ability of CD2AP to TKS4 peptides, Myc-tagged CD2AP protein was used. Myc-DDK tagged human CD2-associated protein (RC210191) was purchased from Origene Technologies. The screening experiments were performed using two glass slides. As a control, a slide was incubated with Alexa Fluor® 488-conjugated Myc-Tag (71D10) rabbit monoclonal antibody purchased from CST (Cell Signaling Technology). For the experimental slides, peptides were incubated with Myc-CD2AP recombinant protein. The screening experiments and signal visualization were performed by Diagnosticum Zrt using a sandwich-like format in a 300 ul reaction volume in an incubation chamber. First, slides were blocked with SmartBlock solution (CANDOR Bioscience, 113125) for 1 h at 30 °C. Next, the experimental slides were incubated with recombinant Myc-CD2AP protein for 1 h at 30 °C in SmartBlock solution in a final volume of 300 ul. Slides were washed four times in 300 ul TBS buffer for 10 min at room temperature then incubated with 1 ug/ml Myc-tag Alexa Fluor 488 antibody in SmartBlock solution in a final volume of 300 ul volume for 1 h at 30 °C. To remove unbound fluorescent antibodies, slides were washed four times with TBS buffer for 10 min at room temperature, and the slides were dried via centrifugation in 50 ml falcon tubes at 1200 rcf for 2 min. Recombinant protein binding was detected by measuring the fluorescence intensity of the peptide spots using a Molecular Devices Axon Genepix 4300A laser scanner. The results were analysed with PepSlide Analyzer Software (Sicasys). Peptide sequence of TKS4 and the results of the peptide array are in Supplementary Table 2.

2.9. Tissue cDNA array and database analyses

The Colon Cancer Tissue cDNA Array IV panel (HCRT104) was purchased from Origene Technologies (Origene Technologies Inc., Rockville, MD). Each plate contains 48 samples from patients with cancer whose clinical pathological features are freely available at the following link from Origene: https://www.origene.com/catalog/tissues/tissuescan/hcrt104/tissuescan-colon-cancer-cdna-array-iv. The three cecum samples (one from a normal group and two from a tumorous group) were excluded from the analyses. qRT-PCR was performed with FastStart Green PCR Master Mix, and gene expression levels were normalized to the *GAPDH* and *Pum1* levels [13]based on the method of Hellemans and colleagues [14]. Database analyses were performed using Gent2 database [15] and The Cancer Genome Atlas Colon Adenocarcinoma (TCGA-COAD) using the Xena Platform [16]. *CD2AP* and *TKS4 (SH3PXD2B)* mRNA expression level data were collected from tumour or matched normal tissues. From Gent2 database, datasets including the same nomenclatures (normal, stage I, stage II, stage III and stage IV samples) were exported. Thus, GSE21510 and GSE32323 datasets were omitted from our analyses. Statistically significant differences were calculated using Student's t-test ($p<0.05$). Analysis of variance (ANOVA) followed by the Tukey HSD (Honest Significant Difference) test were used for statistical analysis. For correlation studies, Pearson correlation coefficients were estimated. ROC curve analyses were performed using the Wilson/Brown method. All statistical analyses were performed in GraphPad Prism 8.0.1 (GraphPad Prism, Inc. San Diego, CA). Protein interaction network of TKS4-interacting partners were created using the STRING database [17].

3. Results

3.1. CD2AP is a novel TKS4 interacting partner

To identify specific TKS4-interacting partners, a high-throughput approach was performed using immunoprecipitation (IP) and mass spectrometry (MS) (Figure 1A). We identified various novel interacting partners, including transmembrane receptors, RNA-binding proteins, phosphatases, and cytoskeletal motor proteins. Six proteins consistently appeared as TKS4-interacting partners in all analysed cancer cells, i.e., breast (MCF-7), colon (HCT 116, DLD-1), and pancreas (HPAC) cancer types, namely CD2AP, CAPZA, CAPZB, KIF5B, CCAR1, and WARS (Figure 1B). Furthermore, these proteins were also present in the precipitate in two studied non-cancerous cell lines (NIH 3T3, HEK 293). Among them, CD2AP was previously shown to be a cancer-related scaffold protein [4]; thus, we focused on the CD2AP-TKS4 interaction in this study. The TKS4-CD2AP interaction was also validated by IP-western blot (WB) (Figure 1B).

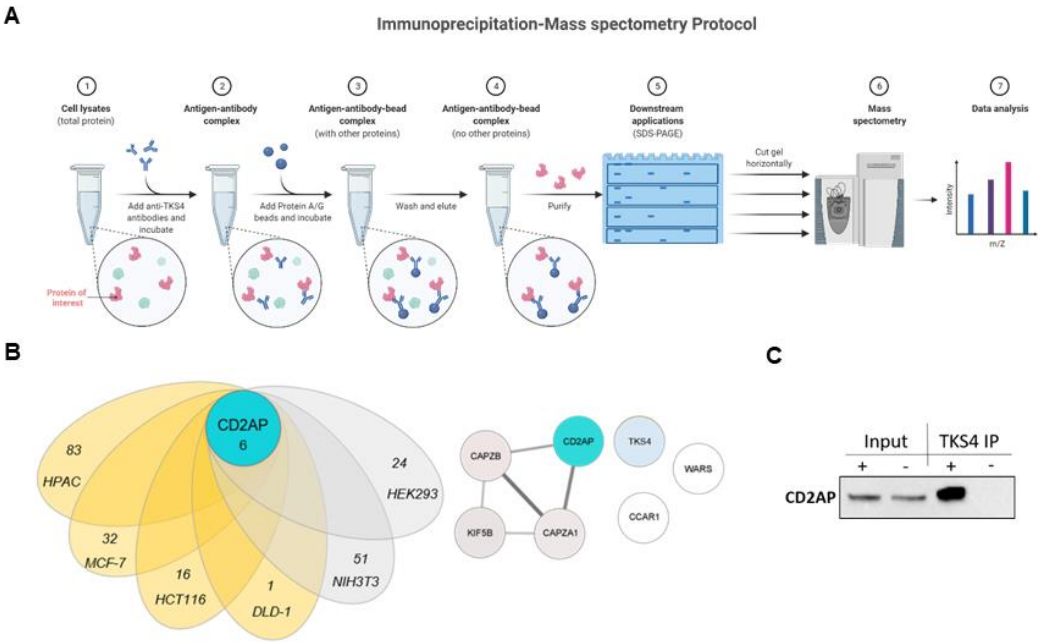


Figure 1. A combination of co-IP and MS was used to identify new TKS4-interacting partners, and IP-WB was used to validate the CD2AP-TKS4 interaction. (A) Step-by-step representation of the IP/MS method (BioRender) **(B)** Left: Numbers above the cell lines indicate the number of identified proteins in the precipitate. Right: Six consistently detected TKS4-interacting partners and their interaction network created via the STRING database. Lines indicate previously identified interactions. **(C)** Validation of CD2AP as a TKS4-interacting partner via IP-WB.

In a previous study, Bazalii et al. [18] has already demonstrated that CIN85 (also called SH3KBP1), a close paralog of CD2AP, interacts with TKS4 through its three SH3 domains; however, they did not identify the exact interacting residues within TKS4. Since SH3 domains typically bind proline-rich short linear motifs (SLiMs) within disordered regions of their partner proteins, we decided to map the interaction sites and determine the location of the CD2AP-binding sites in TKS4 by performing a Pepstar Peptide Microarray analysis. Binding of full-length CD2AP protein was screened against 15mer overlapping TKS4 peptides, therefore, we did not expect to find interactions potentially mediated by the folded domains of TKS4, rather, our analysis was suited to identify potential SLiMs (shorter than 15 residues) within the disordered regions of TKS4. CD2AP and CIN85 have 3 SH3 domains each, which are special in the sense that they recognize non-canonical PxP/A/V/IxPR motifs instead of the canonical PxxP SH3 binding sites [19] [20]. Since 15mer peptides cannot preserve conformational features, we calculated the relative accessible surface area (rASA) of each amino acid of the TKS4 peptides that showed binding to see if those are accessible in the folded protein. The TKS4 peptides showing the strongest interactions included peptides 188 and 189, which both contain a suitable CD2AP/CIN85 binding site ^{755}PVVPPR^{760}; in addition, these peptides had the highest surface accessibility values (10 of 15 peptides with >50% rASA) (Figure 2A). Thus,

we propose that binding of the CD2AP SH3 domains to this motif represent the primary interaction mechanism between the two proteins. The other TKS4 peptides that bound CD2AP mainly reside within the four SH3 domains of TKS4 with low or medium accessibility values (Figure 2A). As the binding cleft of SH3 domains comprises approximately 60 residues and most of them are less accessible in the context of the whole protein (Figure 2A), these binding events are probably artefacts stemming from domain fragmentation and exposure of their hydrophobic cores.

Based on our results, we conclude that CD2AP is a novel direct TKS4-interacting partner and that it binds to the ^{755}PVVPPR^{760} motif in TKS4 through its SH3 domains.

To confirm the presence of the CD2AP-TKS4 interaction in colon cancer cells and to visualize the resulting complexes, proximity ligation assays (PLA) were used (DuoLink). Micrographs of the assays revealed that CD2AP and TKS4 colocalized and are in close proximity (at a distance of less than 40 nm) as indicated by the appearance of red fluorescent dots in HCT116 cells (Figure 2B).

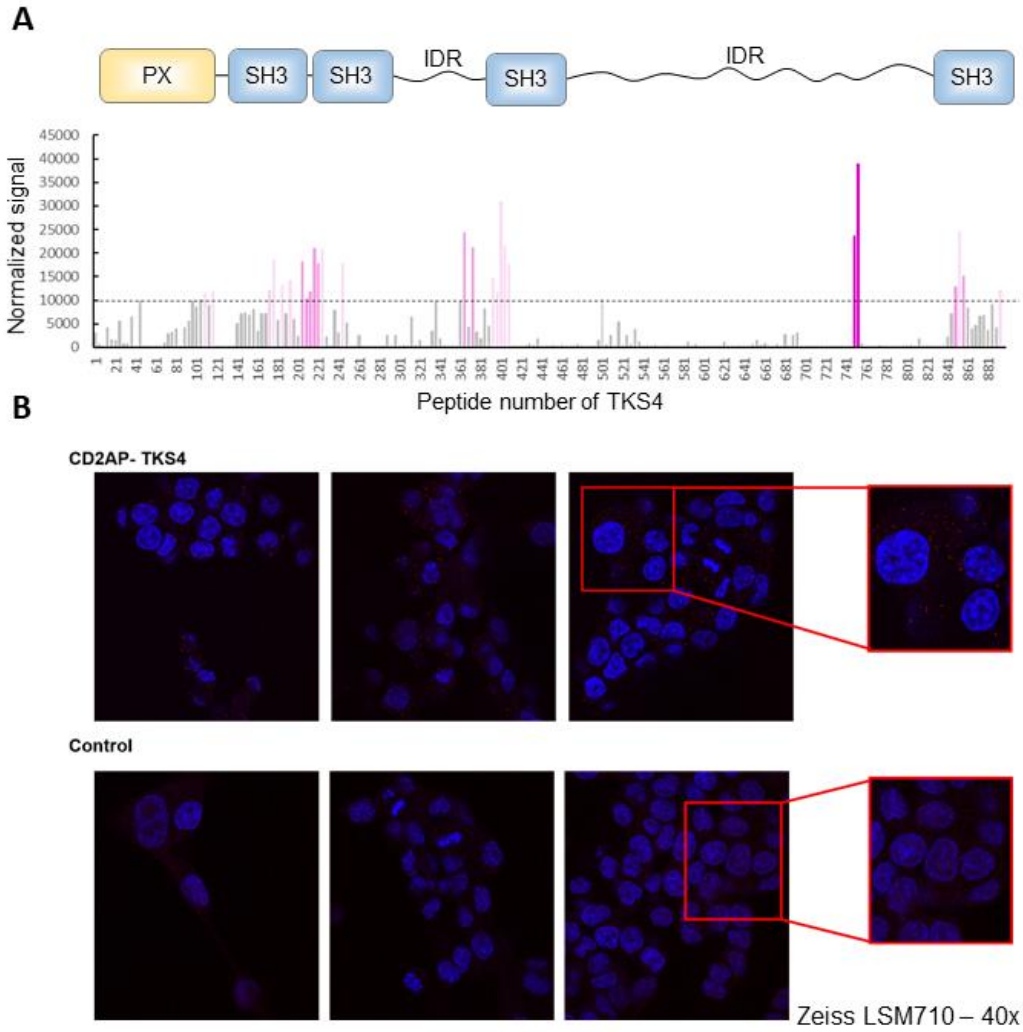


Figure 2. (A) Identification of potential CD2AP binding sites within TKS4. Columns indicate the detected normalized fluorescent signal from the peptide microarray, and the dashed line represents the threshold signal intensity (10000 U). The numbers on the X axis show the analyzed peptides represented by the position of their first residue in the TKS4 sequence. Pink columns represent normalized signal above threshold (dashed line) and the shading indicates the relative surface accessibility value (# of residues with >50% rASA) of each 15mer peptide spot (light pink: 0-<5 residues, medium pink: 5-<10 residues, dark pink: ≥10 residues) **(B)** Detecting the CD2AP-TKS4 interaction in HCT116 cells: fluorescence signal is generated only when the two DPLA probes are in proximity. Colocalization appears as red dots in the cytoplasm and close to the cell membrane. No colocalization was seen in control staining that did not include the primary antibodies in the DPLA reaction. Colours: red

dots = CD2AP-TKS4 colocalization, green = actin (phalloidin), blue = DAPI-stained nuclei;

3.2. Absence of CD2AP or TKS4 promotes colon cancer cell migration separately through an intermediate epithelial-mesenchymal transition (EMT) process

The CD2AP/TKS4 molecular interaction has not been previously described, and the consequences of this adaptor complex formation have not been previously studied. It was already reported that the lack of TKS4 promotes cell migration and induces an EMT-like process in colon cancer cells [10], and CD2AP has also been connected to cancer development, the inhibition of metastasis in gastric cancer via interactions with actin, and the regulation of epithelial cell junction formation. Therefore, we hypothesized that the TKS4/CD2AP complex cooperatively regulates cytoskeletal rearrangements during cancer progression and participates in EMT process initiation. To investigate the cooperative role of the CD2AP/TKS4 adhesion-related adapter proteins in colon cancer progression, *CD2AP* and *TKS4* were silenced via siRNA treatment either separately or together in HCT116 colon cancer cells, and cell motility assays were performed in the temporarily modified cells. We confirmed that CD2AP and TKS4 protein levels were efficiently knocked down after the siRNA treatments (Figure 4B). To explore whether CD2AP levels modulate cell migration, as was previously reported in the absence of TKS4 protein, wound healing assays were performed. *CD2AP* silencing alone or in combination with *TKS4* silencing enhanced HCT116 cell migration compared with that of the control cells (Figure 3).

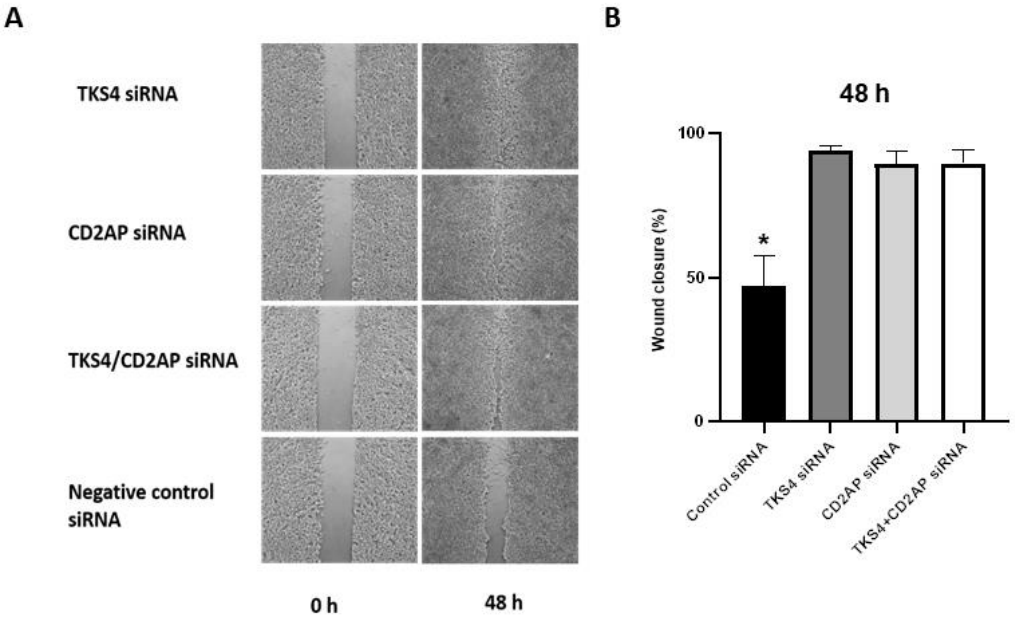


Figure 3. Wound healing assay in *CD2AP*-, *TKS4*-, and *CD2AP/TKS4*-silenced HCT116 cells. Between-group comparisons were performed using ANOVA and Tukey's test, and statistical significance was set at *P<0.05.

To further examine whether increased cell migration stems from an EMT process, we measured the expression levels of EMT-related genes. Interestingly, the RNA and protein levels of the epithelial marker E-cadherin increased in *CD2AP*-silenced cells (Figure 4A,B). In parallel, *Zeb1* RNA was downregulated, suggesting that the repression of *E-cadherin* transcription by Zeb1 is weakened, thereby resulting in an elevated E-cadherin level and an epithelial cell phenotype. Similarly, the E-cadherin level also increased in *TKS4*-silenced cells, suggesting a similar effect on EMT (Figure 4A,B). By contrast, stable *TKS4*-/- knock out cells [10] have a decreased E-cadherin level compared with that of wild type cells. These data suggest that modulation of the TKS4 level (differing between reduced TKS4 levels and

complete absence of TKS4) might drive molecular changes or different stages of the EMT process while exerting the same effect on migratory phenotypes.

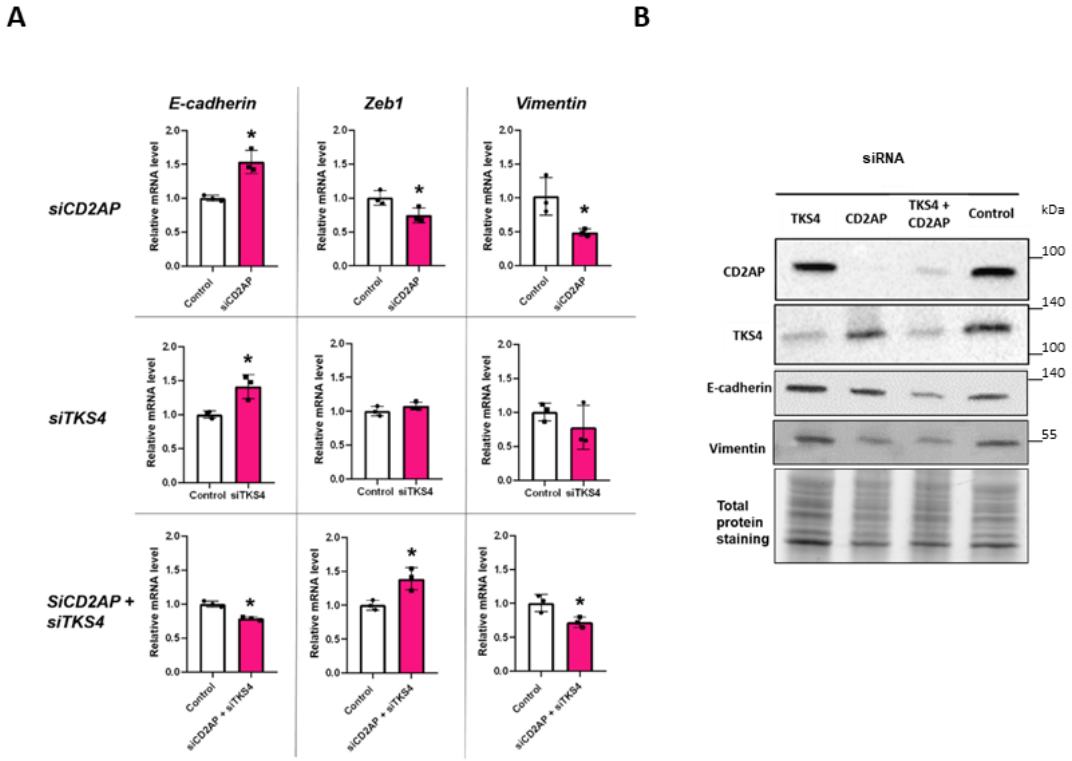


Figure 4. qPCR and western blot analyses of EMT markers in *CD2AP*-, *TKS4*-, and *CD2AP/TKS4*-silenced HCT116 cells. (A) qPCR data showing EMT marker gene expression levels. **(B)** EMT marker protein levels as measured via western blotting. Protein levels of EMT markers (E-cadherin and vimentin) normalized to total protein staining. P values were calculated with Student's t-test, and statistical significance was set at *P<0.05.

To further examine the roles of the CD2AP/TKS4 adaptor complex in EMT processes, the mRNA and protein levels of the mesenchymal marker vimentin were also examined. Both the qPCR and WB results showed decreased vimentin levels in *CD2AP*-silenced cells, further confirming that cells lacking CD2AP retain their

epithelial phenotype (Figure 4A,B). By contrast, vimentin levels were unchanged in *TKS4*- silenced cells. It may be assumed that both TKS4 and CD2AP induces mesenchymal transition and impair the migratory ability of tumour cells in parallel, suggesting that cells undergo a partial EMT program, namely a transition to an intermediate cell state (ICS) or hybrid E/M phenotype [21][22].

To test the hypothesis that CD2AP and TKS4 levels modify cancer cell motility, HCT116 cells were transduced with CD2AP- and/or TKS4-overexpressing plasmids. We confirmed that CD2AP and TKS4 expressions were efficiently enhanced after transfection (Figure 5C). The results showed that CD2AP and TKS4 overexpression resulted in slower cell migration and that CD2AP-, TKS4-, and CD2AP/TKS4-overexpressing HCT116 cells both have normal E-cadherin and vimentin protein levels (Figure 5AB). Based on these results, it appears that either TKS4 or CD2AP inhibits cell migration and induces partial EMT in colon cancer cells.

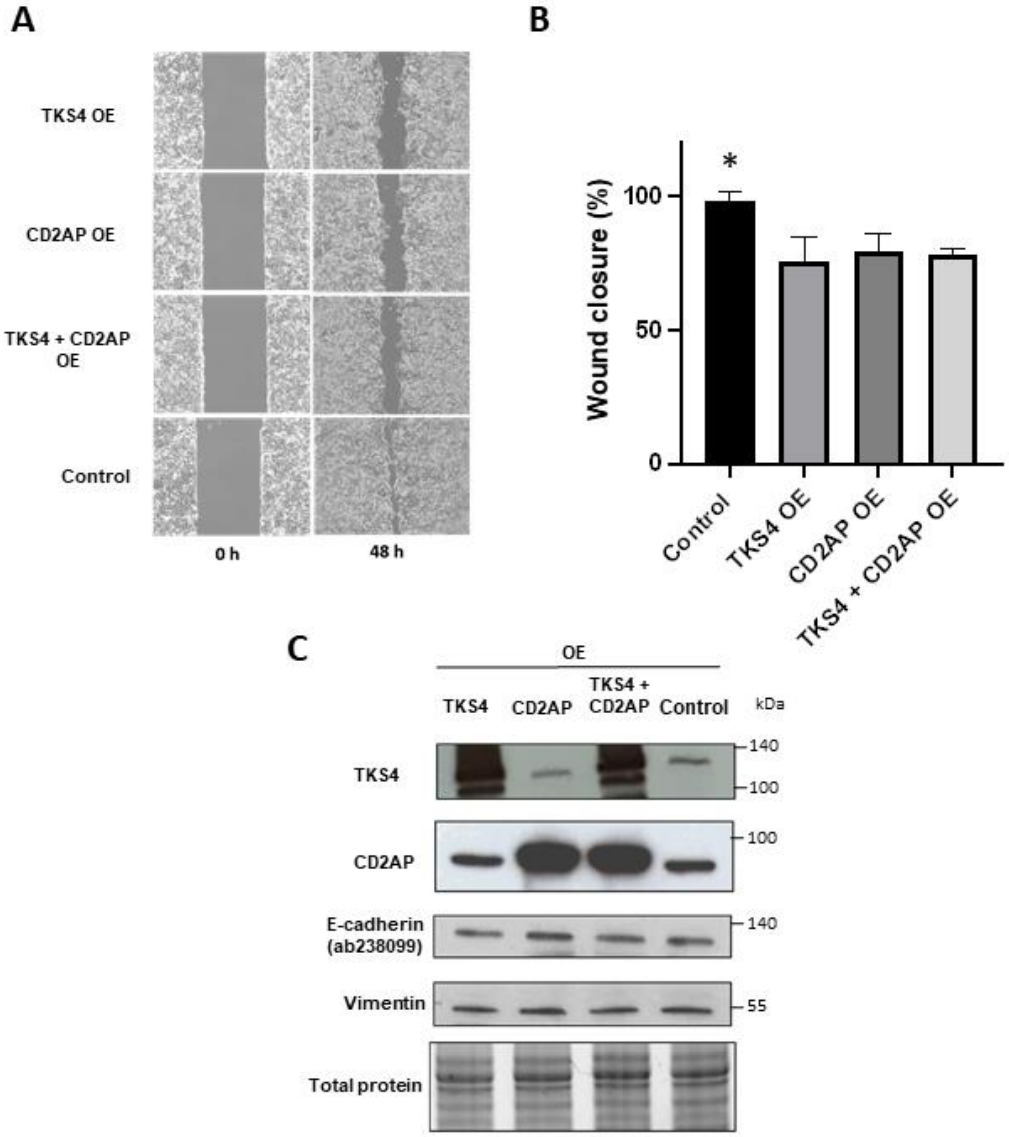


Figure 5. Wound healing assays with CD2AP-, TKS4-, and CD2AP/TKS4-overexpressing HCT116 cells. (A) Wound healing assays. **(B)** Results of wound closure after 48 hours (%). Between-group comparisons were performed using ANOVA and Tukey's test, and statistical significance was set at *P<0.05. **(C)** EMT marker protein levels as measured via western blotting. Protein levels of EMT markers (E-cadherin and vimentin) normalized to total protein staining.

3.3. CD2AP/TKS4 double silencing increases the migratory ability of colon cancer cells accompanied by an intermediate EMT phenotype

Co-silencing of CD2AP and TKS4 in HCT116 cells also induced faster wound closure; therefore, CD2AP and TKS4 might both inhibit cell migration (Figure 4). As E-cadherin RNA and protein levels were also lower in these double silenced cells, it can be assumed that the decreased levels of these proteins provoke EMT-related cell migration. However, the level of vimentin was also lower in these cells (like in CD2AP-silenced cells) (Figure 4). Therefore, our results suggest that CD2AP and TKS4 have coordinated roles in inducing a unique ICS in colon cancer cells.

3.4. CD2AP and TKS4 expression levels are a useful colon cancer biomarker

To investigate the expression levels of CD2AP and TKS4 in colon cancer patients, we collected and analysed mRNA data from TCGA and the Gent2 database. We found lower CD2AP levels and higher TKS4 levels in tumour tissue compared to normal tissue from cancer patients (Figure 6A). To test if further expressional changes present during progression, GENT2 datasets were grouped to stage I, stage II, stage III and stage IV. While CD2AP levels were equal in all stages, TKS4 expressions were significantly increased in Stage IV compared to Stage II cancer (Figure 6A). We also wanted to validate the results of the database analysis and to further explore the effects of the expression levels on disease progression in human colon cancer patients. Therefore, mRNA expression levels were measured using an qRT-PCR method with a Human Colon Cancer Tissue cDNA Array. Similar to the database results, the tissue cDNA array analysis revealed that the *CD2AP* mRNA level was decreased in cancerous colon tissue compared to normal tissue (Figure 6B). Furthermore, significantly increased *TKS4* expression levels were detected in Stage II and Stage IV cancer, and the *TKS4* level increased during disease progression, similar to findings gleaned from the database analyses (Figure

6B). In summary, the *CD2AP* level decreases while the *TKS4* level increases in colon cancer tissue.

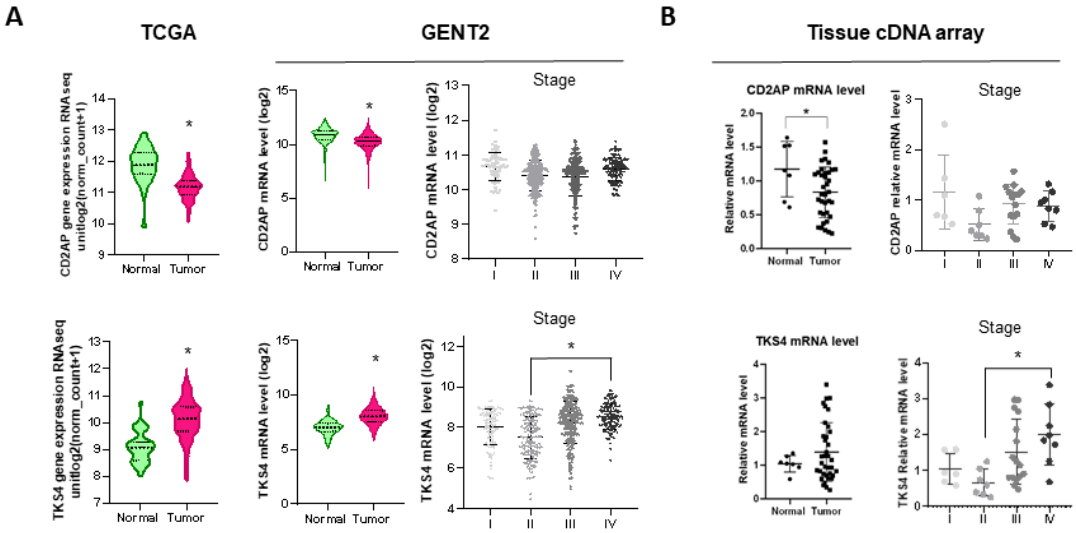


Figure 6. (A) *CD2AP* and *TKS4* expression levels in colon cancer patients taken from TCGA and the Gent2 database (left panel). *CD2AP* and *TKS4* expression levels in Stage I-IV from GENT2 (right panel) **(B)** Colon Cancer Tissue cDNA array results of *CD2AP* and *TKS4* mRNA expression levels in normal and tumour samples (left panel). *CD2AP* and *TKS4* expression levels in tumour tissue from stage I, II, III and IV cancers (right panel). P values were calculated with Student's t-test or Tukey's test (with an ANOVA post-hoc test in the comparison of stages). Statistical significance was set at *P<0.05.

3.5. CD2AP and *TKS4* mRNA expression levels are negatively correlated in normal but not in tumour tissue from colon cancer patients

It was previously reported that the gene expression levels of interacting partners might be correlated [23] [24], implying that members of the TKS4 interactome might be co-regulated, particularly CD2AP. To investigate whether the opposite expression level changes observed in cancer tissue (see above) affect the

relationship between *CD2AP* and *TKS4* expression levels, we performed correlation analyses of the *CD2AP* and *TKS4* levels in cancer and normal samples using TCGA-COAD (Figure 7A,B). In normal tissue, we found a significant negative correlation between the *CD2AP* and *TKS4* levels (R = -0.5801; p = 0.0001) (Figure 7A, green dots), suggesting that high *TKS4* expression was accompanied by low *CD2AP* expression in non-cancerous tissues. By contrast, no such a correlation was found in tumour tissue (R = -0.078; p =0.188) (Figure 7B, pink dots). Thus, this altered relationship between the CD2AP and TKS4 levels might be related to the formation of tumour cells in the colon and could be a hallmark of cancer progression. Moreover, merged expression data from TCGA-COAD were plotted to explore whether there is any clustering based on the *CD2AP* and *TKS4* levels. This analysis revealed three clusters defined by different *CD2AP* and *TKS4* levels, namely CD2AP/TKS4 ratios (Figure 7C). Normal samples formed a single “low*TKS4* - high*CD2AP*” cluster (high ratio), while tumour samples formed two distinct clusters: a “low*TKS4* - low*CD2AP*” (mean ratio) and a “high*TKS4* - low*CD2AP*” cluster (low ratio) (Figure 7C,E). The *CD2AP* and *TKS4* expression data from the GENT2 database and the human tissue cDNA array were also analysed on a dot-plot (Figure 7D) to determine whether similar clustering patterns could be observed. Indeed, normal and cancerous tissues clustered similarly based on their relative TKS4 and CD2AP expression levels (Figure 7D). Thus, the discrimination of tumorous and normal tissue based on analysis of *CD2AP* or *TKS4* expression alone might result in a false negative detection, while co-analysis of *CD2AP* and *TKS4* expression might represent a useful diagnostic marker.

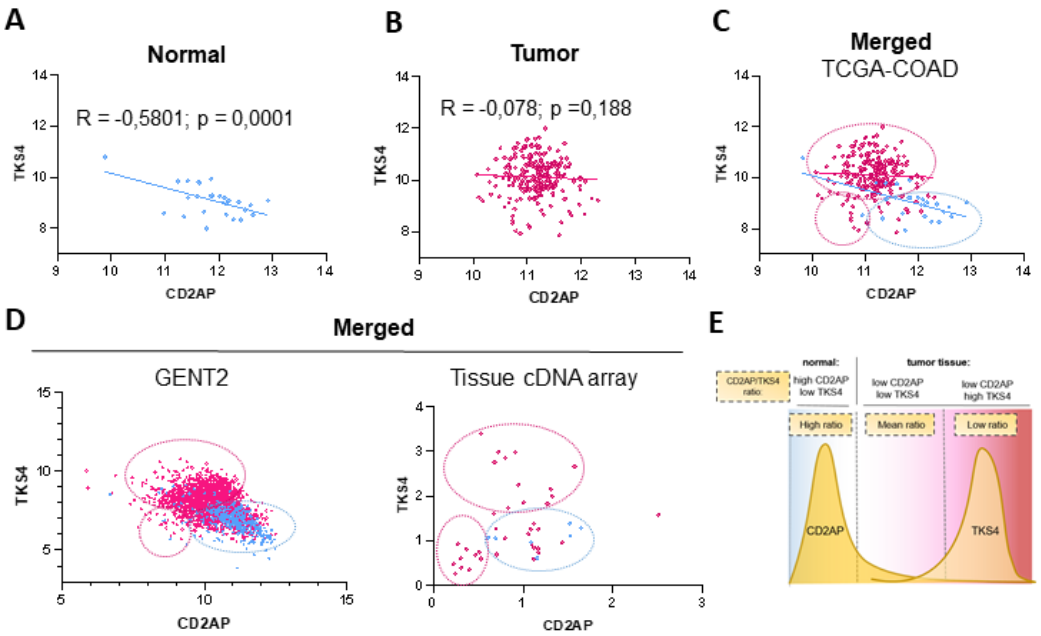
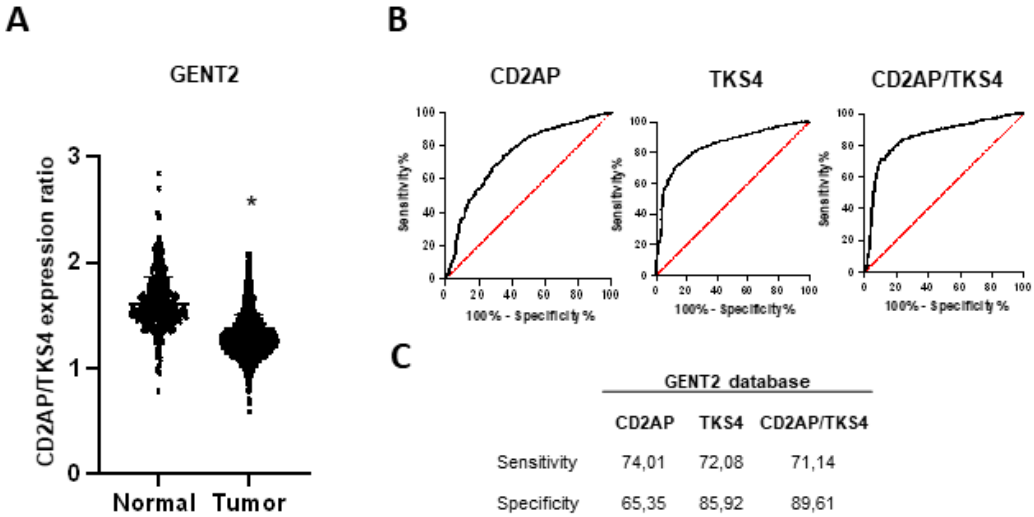


Figure 7. *CD2AP* and *TKS4* expression level correlations in colon cancer patient samples. (A) Correlation analysis of *CD2AP* and *TKS4* expression levels in non-tumorous (blue dots) **(B)** and tumorous colon (pink dots) (COAD, colon adenocarcinoma) tissue from the TCGA-COAD database. Transcript levels are shown as gene expression RNAseq unitlog2(norm_count+1) values. R = Pearson correlation coefficient. Statistical significance was set at *P<0.05. **(C)** Merged dot-plot of the *CD2AP* and *TKS4* expression levels from the TCGA-COAD database (left panel), **(D)** GENT2 database (middle panel) and samples from tissue cDNA array (right panel), **(E)** CD2AP/TKS4 ratio clusters in normal and tumorous colon tissues.

3.6. The CD2AP/TKS4 expression ratio is an outstanding paired biomarker for colon cancer diagnosis

Our results suggest that CD2AP and TKS4 expression levels are associated with tumour formation. As loss of CD2AP or loss of TKS4 alone resulted in different expression patterns of EMT markers *in vitro* (Figure 4), we hypothesized that the *CD2AP/TKS4* expression ratio might have an instructive role in tumour phenotypes.

To test this possibility, *CD2AP/TKS4* gene expression ratios were calculated using GENT2 data (Figure 6A) and our tissue cDNA array data (Figure 6B). The results showed that the CD2AP/TKS4 gene expression ratio is consistently decreased in tumour tissue (Figure 8A; Suppl Fig. 1.).



C

GENT2 database

	CD2AP	TKS4	CD2AP/TKS4
Sensitivity	74,01	72,08	71,14
Specificity	65,35	85,92	89,61
Optimal cutoff value	≤10.67	>7.618	≤1.36
AUC	0,7461	0,8470	0,8529
p-value	**<0,0001**	**<0,0001**	**<0,0001**

Figure 8. (A) The CD2AP/TKS4 gene expression ratio in colon cancer patients from GENT2 data. P values were calculated with Student's t-test, and statistical significance was set at *P<0.05. **(B)** ROC curve analyses of TKS4, CD2AP and CD2AP/TKS4 ratios from GENT2 data. **(C)** Results of ROC curve analyses from GENT2 datasets

To further test the power of the *CD2AP/TKS4* ratio as a biomarker, ROC curve analyses were performed based on our database analyses and our colon cancer tissue cDNA array data to compare the clinical sensitivity and specificity of *CD2AP* or

TKS4 expression alone as well as the *CD2AP/TKS4* expression ratio in distinguishing normal and tumorous tissue (Figure 8B). The ROC curve analysis of TCGA-COAD data showed that the *CD2AP* and *TKS4* levels as well as the *CD2AP/TKS4* ratio are excellent markers for distinguishing tumour and normal phenotypes. For the *CD2AP* and *TKS4* expression levels and the *CD2AP/TKS4* ratio, the area under curve (AUC) values were 0.7461, 0.8470, and 0.8529, respectively (Figure 8C). Of these three features, the *CD2AP/TKS4* ratio was the most sensitive and specific, with an AUC of nearly 0.9 (Figure 8C). For the tissue cDNA array data, neither *CD2AP* expression nor *TKS4* expression alone could distinguish between tumour and normal tissue (Suppl. Fig. 1.); however, the ROC curve based on the *CD2AP/TKS4* ratio had an AUC of 0.7529. Similar results have been observed using the TCGA datasets (Suppl. Fig. 1.). Thus, analysis of paired expression levels had an acceptable significance level to support the use of the *CD2AP/TKS4* ratio as a colon cancer biomarker (Figure 8B,C).

In conclusion, the *CD2AP/TKS4* expression ratio is an outstanding biomarker for the identification of tumorous colon tissue (AUC=0.8981).

4. Discussion

Several biomarkers have been developed and used for colon cancer detection; however, due to the high prevalence of this cancer and its poor prognosis, more sensitive and specific biomarkers are needed. Here we show that the previously identified gastric cancer-related CD2AP adaptor protein also has a role in colon cancer development. In addition, we found that CD2AP and the EMT regulator TKS4 form a complex in several different cancer cell types. Via peptide array analysis, we

confirmed that CD2AP binds to region between the last two SH3 domains in TKS4. In TKS4, the ^{755}PVVPPR^{760} sequence represents an optimal recognition motif for the first two SH3 domains of CD2AP. Our results regarding this interaction are reinforced by the results of a previous study that confirmed an interaction between TKS4 and CIN85 [18]. CIN85 is a paralog of CD2AP and both have three SH3 domains [25], which are special in the sense that they recognize atypical polyproline/arginine motifs (PxP/A/V/IxPR) instead of the canonical PxxP SH3 binding sites [19][20]. Based on these similarities, we propose that CD2AP and CIN85 likely interact with TKS4 via the same mechanism. Furthermore, via Duolink Proximity Ligation Assays we confirmed that CD2AP and TKS4 are in close proximity.

We also showed that TKS4 and CD2AP have roles in cell motility and can regulate EMT-related processes in colon cancer. These results raise the possibility that the interaction between these two proteins is fundamental to signalling during colon cancer development. Therefore, we investigated the cooperative roles of CD2AP and TKS4 in tumour progression. We found that these proteins parallelly induce mesenchymal traits and modify colon cancer cell migration; thus, it appears that reduced levels of these proteins provoke a partial EMT process in colon cancer.

Our results suggest that as adaptor proteins, TKS4 and CD2AP form a complex that organises spatially distinct signal transduction pathways by bringing signalling molecules into close proximity. This type of signal control is not unique, as previous studies have also demonstrated that adaptor proteins can functionally interact with each other to co-organize the members of different signalling pathways [26][27]. For instance, multiple interactions between the Grb2, SOS, LAT, SLP-76, ADAP, and Gads adaptor proteins create a web of interactions, creating a “bridge” between various signalling pathways and might fine tune various signalling events

[26]. Furthermore, Park and co-workers showed that the Crk and CrkL adaptor proteins have overlapping roles in the regulation of cell motility and EMT in breast and colorectal cancer [28]. Thus, our results further reinforce the paradigm that adaptor proteins can functionally interact and that they might have cooperative roles in tumour development. As both TKS4 and CD2AP bind to the actin cytoskeleton rearranging proteins, we established a model in which TKS4 and CD2AP cooperatively regulate actin cytoskeleton and actin-based cell motility processes (Figure 9). It was previously confirmed that TKS4 could be membrane-bound (via the PX-domain) and that it co-localizes with cortactin, an actin-binding scaffold protein [5][11][29]. In addition, upon Src activation, TKS4 binds to the C-terminal region of cortactin [11]. In EGF-treated cells, CD2AP can also interact with cortactin through its SH3 domains [30][31]. In this study, we showed that TKS4 and CD2AP interact under non-phosphorylated conditions, as we did not treat cells with EGF or transfect cells with Src. Thus, we hypothesize that CD2AP and TKS4 form a stable complex at the cell membrane and that they both conditionally interact with cortactin dependent on EGF phosphorylation. This web of interactions brings actin filaments, actin-regulating proteins and the membrane together to create membrane protrusions, which are necessary for the formation of membrane ruffles, lamellipodia, or invadopodia to support efficient metastasis. Our results also show that TKS4 and CD2AP independently decrease but cooperatively increase the E-cadherin level. E-cadherin is a cell adhesion molecule that binds to the Arp2/3 complex, and it has a role in metastasis [32][33]. As TKS4 and CD2AP function synergistically to promote a unique ICS, we conclude that the TKS4/CD2AP complex might regulate E-cadherin-mediated actin rearrangement and actin-based cell motility processes, including EMT.

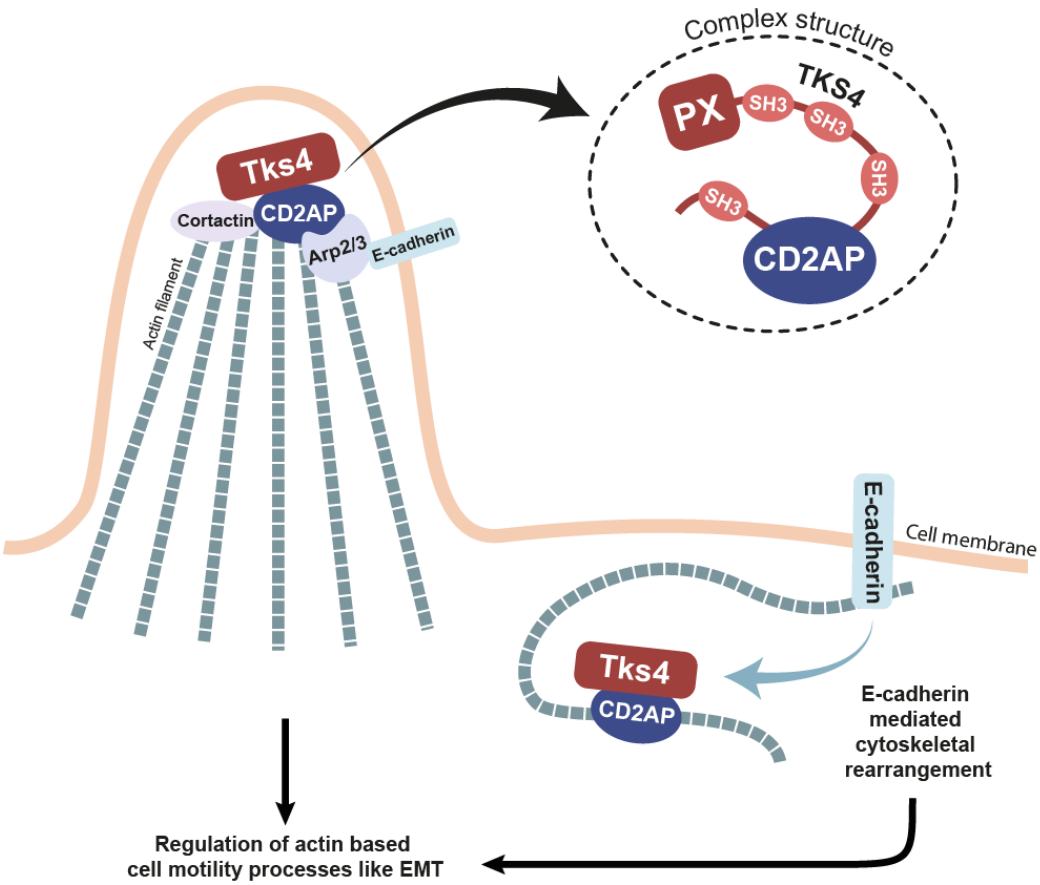


Figure 9. **Model of the role of the TKS4/CD2AP complex in cytoskeleton rearrangement and actin-based cell motility process**.

Previous studies showed that molecular interactions between EMT-related proteins can induce metastasis. For instance, Ghahhari et al. reported that an interaction between the ZEB1 transcription factor and ERα (oestrogen receptor α) induces breast cancer cell metastasis to bone tissue [34]. Furthermore, it has been shown that ZEB1 can create a transactivation complex with participants of the AP-1 (adaptor protein complex-1) and the Hippo pathways, leading to the formation of more invasive and aggressive breast cancer cells. It was proposed that impairing the molecular interactions between adaptor proteins involved in metastasis might be a novel approach for treating aggressive cancer [35]. Similarly, our results revealed that the interaction between CD2AP and TKS4 might have a cooperative role in tumour development, particularly in promoting a unique partial EMT state in colon

cancer cells (Figure 10). Thus, targeting this protein-protein interaction might be a novel approach for inhibiting colon cancer metastasis.

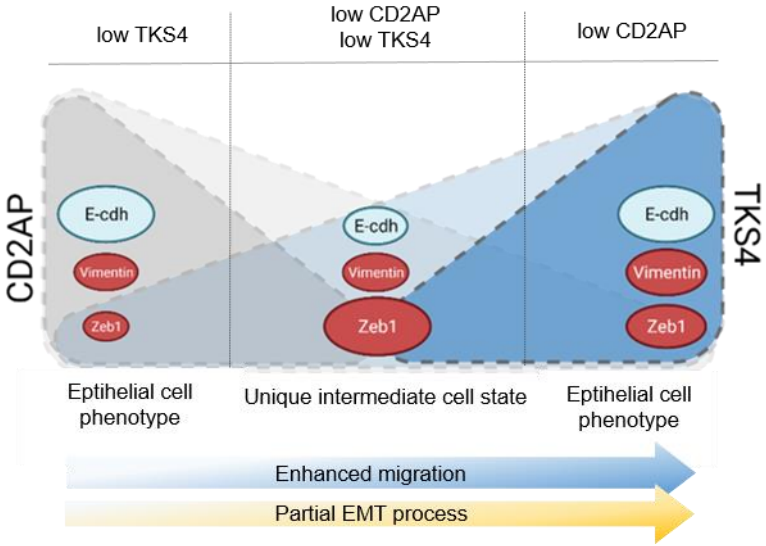


Figure 10. Epithelial and mesenchymal cell phenotypes, cell migrations and expression levels of EMT-related genes grouped according to the levels of the CD2AP and TKS4 proteins based on gene silencing experiments. E-cdh: E-cadherin, Vimentin, Zeb1.

As we detected the CD2AP/TKS4 interaction in various cancer types, we hypothesize that the cooperative function of CD2AP and TKS4 in metastasis might be conserved in other cancerous and normal cells. Using cDNA samples from colon tumour samples and RNAseq data from databases, we found that the *CD2AP* level decreases as the *TKS4* level increases in colon cancer and that further expression levels changes can be detected during disease progression. Correlation analysis of *CD2AP* and *TKS4* levels revealed a significant negative correlation presents in normal tissues but it is absent in tumour samples. Thus, this dysregulation might be linked to colon cancer cell formation and could be a hallmark of cancer progression. We further found that the CD2AP/TKS4 expression level ratio has an instructive role in tumour phenotypes. We hypothesize that disturbed EMT signalling regulation induced by the decreased CD2AP/TKS4 expression ratio in colon cancer cells might

enhance their migratory capacity. Wound healing assays support this hypothesis, as a low CD2AP/TKS4 ratio indeed resulted in faster colon cancer cell migration.

Our ROC curve analyses suggest that the CD2AP/TKS4 expression ratio is an outstanding biomarker for the identification of tumorous colon tissue. Further exploration of the molecular interaction between CD2AP and TKS4, their cooperative roles in EMT, as well as the implications of the decreased CD2AP/TKS4 expression ratio in tumours might reveal additional details that could be used to develop novel approaches for detecting colon cancer. From a broader perspective, our study highlights the possibility that gene expression ratios might be more reliable biomarkers than the expression levels of single genes for the diagnosis of early-stage colon cancer and perhaps additional cancer types, such as breast or pancreatic adenocarcinoma.

Conclusions

Our study revealed that CD2AP is a novel TKS4-interacting protein and together they form a scaffold complex. They function synergistically to promote a unique ICS as they potentially regulate EMT-related signalling in HCT116 cells. Furthermore, our work provided a new insight into that gene expression ratio –as CD2AP/TKS4 ratio - might serve as a more reliable biomarker than single gene expression for the diagnosis of early-stage colon cancer. At last, as the interaction between CD2AP and TKS4 might have a cooperative role in tumour development, targeting this protein-protein interaction might be a novel approach for inhibiting colon cancer metastasis.

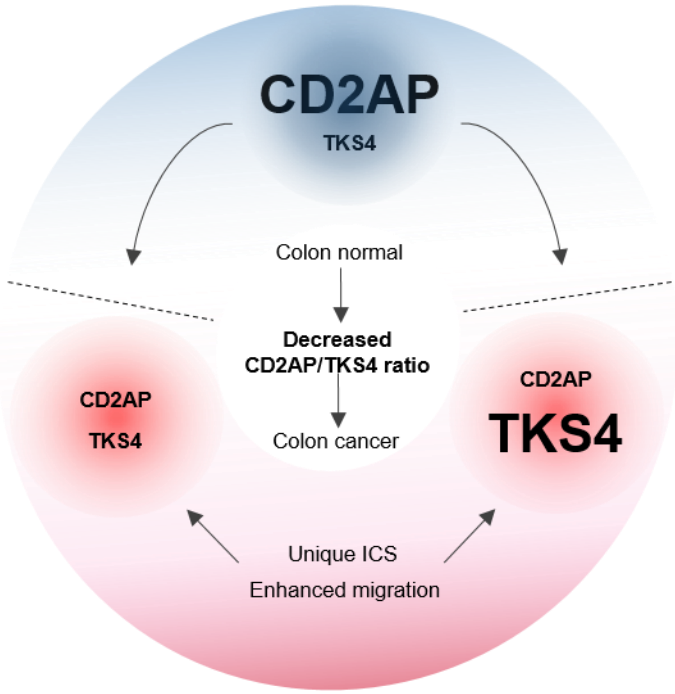


Figure 11. Integrated model depicting the tumour-promoting effect of CD2AP/TKS4 gene expression ratio during colon cancer formation.

Additional files

Additional file 1: Additional Tables 1-2.

Additional file 2: Figure S1.

Ethical approval

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author contributions

AK initiated the project, designed the study, carried out the experiments, analysed the data, supervised students, and wrote the manuscript. VV participated in the coordination, conceptualization, and implementation of the study, helped in data analysis as well as writing and editing of the manuscript. LL contributed to the experimental work, especially the DuoLink assay, data interpretation, and manuscript preparation. TT and JN helped in cell-based experiments and in data visualization. ÁT contributed to the analysis steps and supported manuscript preparation. LL contributed to the western blot experiments and data interpretation. RP contributed to data interpretation and visualization, as well as editing of the manuscript. LB supervised the project and obtained the funding. All authors read and approved the final manuscript.

Acknowledgments

We thank Zita Haidar for technical assistance and Éva Csősz for valuable help with the mass spectrometry experiments. The authors thank Rita Lózsa, Dávid Szüts, Gyöngyi Bézsenyi and László Homolya for the technical help and the tasks related to cell lines. We also thank János Fekete for his help with bioinformatic analyses. We are especially grateful to Balázs Győrffy for his useful comments about the manuscript.

Funding

This research was funded by grants from the National Research, Development, and Innovation Fund of Hungary K124045 (L.B.), FIEK_16-1-2016–0005, HunProtExc 2018-1.2.1-NKP-2018–00005 (L.B.) and. The authors are thankful for the financial support received as being a Centre of Excellence of the Hungarian Academy of Sciences. The work of TT was supported by the KDP-2021 program of the Ministry of Innovation and Technology from the source of the National Research, Development and Innovation Fund. LL is especially thankful for the support from an award from the Dr. Bodzsár Éva Foundation.

Availability of data and materials

The data and materials supporting the conclusions of this study are included within the article and its additional files.

Author details

[1] Institute of Enzymology, Research Centre for Natural Sciences, 1117 Budapest, Hungary; [2] Doctoral School of Biology, Institute of Biology, ELTE Eötvös Loránd University,1117 Budapest, Hungary

References

1 Jelski W & Mroczko B (2020) Biochemical Markers of Colorectal Cancer – Present and Future. *Cancer Manag Res* **Volume 12**, 4789–4797.

2 László L, Kurilla A, Takács T, Kudlik G, Koprivanacz K, Buday L & Vas V (2021) Recent Updates on the Significance of KRAS Mutations in Colorectal Cancer Biology. *Cells* **10**, 667.

3 Tang VW & Brieher WM (2013) FSGS3/CD2AP is a barbed-end capping protein that stabilizes actin and strengthens adherens junctions. *J Cell Biol* **203**, 815–833.

4 Xie W, Chen C, Han Z, Huang J, Liu X, Chen H, Zhang T, Chen S, Chen C, Lu M, Shen X & Xue X (2020) CD2AP inhibits metastasis in gastric cancer by promoting cellular adhesion and cytoskeleton assembly. *Mol Carcinog* **59**, 339–352.

5 Buschman MD, Bromann PA, Cejudo-Martin P, Wen F, Pass I & Courtneidge SA (2009) The novel adaptor protein Tks4 (SH3PXD2B) is required for functional podosome formation. *Mol Biol Cell* **20**, 1302–11.

6 Iizuka S, Abdullah C, Buschman MD, Diaz B & Courtneidge SA (2016) The role of Tks adaptor proteins in invadopodia formation, growth and metastasis of melanoma. *Oncotarget* **7**, 78473–78486.

7 Maziveyi M, Dong S, Baranwal S & Alahari SK (2018) Nischarin regulates focal adhesion and Invadopodia formation in breast cancer cells. *Mol Cancer* **17**, 21.

8 Li Z-H, Liu Y & Gao S-Y (2019) Correlation between IL-7 genomic protein methylation level and acute myeloid leukemia. *Eur Rev Med Pharmacol Sci* **23**,

1196–1202.

9 Caires-dos-Santos L, da Silva S V., Smuczek B, de Siqueira AS, Cruz KSP, Barbuto JAM, Augusto TM, Freitas VM, Carvalho HF & Jaeger RG (2020) Laminin-derived peptide C16 regulates Tks expression and reactive oxygen species generation in human prostate cancer cells. *J Cell Physiol* **235**.

10 Szeder B, Tárnoki-Zách J, Lakatos D, Vas V, Kudlik G, Merő B, Koprivanacz K, Bányai L, Hámori L, Róna G, Czirók A, Füredi A & Buday AL (2019) Absence of the Tks4 Scaffold Protein Induces Epithelial-Mesenchymal Transition-Like Changes in Human Colon Cancer Cells. *Cells* **8**.

11 Lányi Á, Baráth M, Péterfi Z, Bőgel G, Orient A, Simon T, Petrovszki E, Kis-Tóth K, Sirokmány G, Rajnavölgyi É, Terhorst C, Buday L & Geiszt M (2011) The Homolog of the Five SH3-Domain Protein (HOFI/SH3PXD2B) Regulates Lamellipodia Formation and Cell Spreading. *PLoS One* **6**, e23653.

12 Harnoš J, Ryneš J, Víšková P, Foldynová-Trantírková S, Bajard-Ešner L, Trantírek L & Bryja V (2018) Analysis of binding interfaces of the human scaffold protein AXIN1 by peptide microarrays. *J Biol Chem* **293**, 16337–16347.

13 Vermani L, Kumar R & Senthil Kumar N (2020) GAPDH and PUM1: Optimal Housekeeping Genes for Quantitative Polymerase Chain Reaction-Based Analysis of Cancer Stem Cells and Epithelial-Mesenchymal Transition Gene Expression in Rectal Tumors. *Cureus*.

14 Hellemans J, Mortier G, De Paepe A, Speleman F & Vandesompele J (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* **8**, R19.

15 Park S-J, Yoon B-H, Kim S-K & Kim S-Y (2019) GENT2: an updated gene expression database for normal and tumor tissues. *BMC Med Genomics* **12**, 101.

16 Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, Banerjee A, Luo Y, Rogers D, Brooks AN, Zhu J & Haussler D (2020) Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol* **38**, 675–678.

17 Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, Jensen LJ & von Mering C (2021) The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* **49**, D605–D612.

18 Bazalii A V., Samoylenko AA, Petukhov DM, Rynditch A V., Redowicz MJ & Drobot LB (2015) Interaction between adaptor proteins Ruk/CIN85 and Tks4 in normal and tumor cells of different tissue origins. *Biopolym Cell* **30**, 37–41.

19 Kowanetz K, Szymkiewicz I, Haglund K, Kowanetz M, Husnjak K, Taylor JD, Soubeyran P, Engstrom U, Ladbury JE & Dikic I (2003) Identification of a Novel Proline-Arginine Motif Involved in CIN85-dependent Clustering of Cbl and Down-regulation of Epidermal Growth Factor Receptors. *J Biol Chem* **278**, 39735–39746.

20 Rouka E, Simister PC, Janning M, Kumbrink J, Konstantinou T, Muniz JRC, Joshi D, O'Reilly N, Volkmer R, Ritter B, Knapp S, von Delft F, Kirsch KH & Feller SM (2015) Differential Recognition Preferences of the Three Src Homology 3 (SH3) Domains from the Adaptor CD2-associated Protein (CD2AP) and Direct

Association with Ras and Rab Interactor 3 (RIN3). *J Biol Chem* **290**, 25275–25292.

21 Schaeffer D, Somarelli JA, Hanna G, Palmer GM & Garcia-Blanco MA (2014) Cellular Migration and Invasion Uncoupled: Increased Migration Is Not an Inexorable Consequence of Epithelial-to-Mesenchymal Transition. *Mol Cell Biol* **34**, 3486–3499.

22 Pasani S, Sahoo S & Jolly MK (2020) Hybrid E/M Phenotype(s) and Stemness: A Mechanistic Connection Embedded in Network Topology. *J Clin Med* **10**.

23 Grigoriev A (2001) A relationship between gene expression and protein interactions on the proteome scale: analysis of the bacteriophage T7 and the yeast Saccharomyces cerevisiae. *Nucleic Acids Res* **29**, 3513–9.

24 Ge H, Liu Z, Church GM & Vidal M (2001) Correlation between transcriptome and interactome mapping data from Saccharomyces cerevisiae. *Nat Genet* **29**, 482–6.

25 Gaidos G, Soni S, Oswald DJ, Toselli PA & Kirsch KH (2007) Structure and function analysis of the CMS/CIN85 protein family identifies actin-bundling properties and heterotypic-complex formation. *J Cell Sci* **120**.

26 Yablonski D (2019) Bridging the Gap: Modulatory Roles of the Grb2-Family Adaptor, Gads, in Cellular and Allergic Immune Responses. *Front Immunol* **10**.

27 Buday L, Khwaja A, Sipeki S, Faragó A & Downward J (1996) Interactions of Cbl with two adaptor proteins, Grb2 and Crk, upon T cell activation. *J Biol Chem* **271**.

28 Park T (2021) Crk and CrkL as Therapeutic Targets for Cancer Treatment. *Cells* **10**.

29 Bögel G, Gujdár A, Geiszt M, Lányi Á, Fekete A, Sipeki S, Downward J & Buday L (2012) Frank-ter Haar syndrome protein Tks4 regulates epidermal growth factor-dependent cell migration. *J Biol Chem* **287**, 31321–31329.

30 Lynch DK, Winata SC, Lyons RJ, Hughes WE, Lehrbach GM, Wasinger V, Corthals G, Cordwell S & Daly RJ (2003) A Cortactin-CD2-associated Protein (CD2AP) Complex Provides a Novel Link between Epidermal Growth Factor Receptor Endocytosis and the Actin Cytoskeleton. *J Biol Chem* **278**, 21805–21813.

31 Zhao J, Bruck S, Cemerski S, Zhang L, Butler B, Dani A, Cooper JA & Shaw AS (2013) CD2AP Links Cortactin and Capping Protein at the Cell Periphery To Facilitate Formation of Lamellipodia. *Mol Cell Biol* **33**, 38–47.

32 Bracke ME, Van Roy FM & Mareel MM (1996) The E-cadherin/catenin complex in invasion and metastasis. *Curr Top Microbiol Immunol* **213 I**, 123–161.

33 Kovacs EM, Goodwin M, Ali RG, Paterson AD & Yap AS (2002) Cadherin-directed actin assembly: E-cadherin physically associates with the Arp2/3 complex to direct actin assembly in nascent adhesive contacts. . *Curr Biol* **12**, 379–382.

34 Mohammadi Ghahhari N, Sznurkowska MK, Hulo N, Bernasconi L, Aceto N & Picard D (2022) Cooperative interaction between ERα and the EMT-inducer ZEB1 reprograms breast cancer cells for bone metastasis. *Nat Commun* **13**, 2104.

35 Feldker N, Ferrazzi F, Schuhwerk H, Widholz SA, Guenther K, Frisch I, Jakob K, Kleemann J, Riegel D, Bönisch U, Lukassen S, Eccles RL, Schmidl C, Stemmler MP, Brabletz T & Brabletz S (2020) Genome-wide cooperation of EMT

transcription factor ZEB 1 with YAP and AP -1 in breast cancer . *EMBO J* **39**.

Figure titles and legends

Figure 1. A combination of co-IP and MS was used to identify new TKS4-interacting partners, and IP-WB was used to validate the CD2AP-TKS4 interaction. (A) Step-by-step representation of the IP/MS method (BioRender) **(B)** Left: Numbers above the cell lines indicate the number of identified proteins in the precipitate. Right: Six consistently detected TKS4-interacting partners and their interaction network created via the STRING database. Lines indicate previously identified interactions. **(c)** Validation of CD2AP as a TKS4-interacting partner via IP-WB.

Figure 2. (A) Identification of potential CD2AP binding sites within TKS4. Columns indicate the detected normalized fluorescent signal from the peptide microarray, and the dashed line represents the threshold signal intensity (10000 U). The numbers on the X axis show the analyzed peptides represented by the position of their first residue in the TKS4 sequence. Pink columns represent normalized signal above threshold (dashed line) and the shading indicates the relative surface accessibility value (# of residues with >50% rASA) of each 15mer peptide spot (light pink: 0-<5 residues, medium pink: 5-<10 residues, dark pink: ≥10 residues) **(B)** Detecting the CD2AP-TKS4 interaction in HCT116 cells: fluorescence signal is generated only when the two DPLA probes are in proximity. Colocalization appears as red dots in the cytoplasm and close to the cell membrane. No colocalization was seen in control staining that did not include the primary antibodies in the DPLA reaction. Colours: red dots = CD2AP-TKS4 colocalization, green = actin (phalloidin), blue = DAPI-stained nuclei;

Figure 3. Wound healing assay in *CD2AP*-, *TKS4*-, and *CD2AP/TKS4*-silenced HCT116 cells. Between-group comparisons were performed using ANOVA and Tukey's test, and statistical significance was set at *P<0.05.

Figure 4. qPCR and western blot analyses of EMT markers in *CD2AP*-, *TKS4*-, and *CD2AP/TKS4*-silenced HCT116 cells. (A) qPCR data showing EMT marker gene expression levels. **(B)** EMT marker protein levels as measured via western blotting. Protein levels of EMT markers (E-cadherin and vimentin) normalized to total protein staining. P values were calculated with Student's t-test, and statistical significance was set at *P<0.05.

Figure 5. Wound healing assays with CD2AP-, TKS4-, and CD2AP/TKS4-overexpressing HCT116 cells. (A) Wound healing assays. **(B)** Results of wound closure after 48 hours (%). Between-group comparisons were performed using ANOVA and Tukey's test, and statistical significance was set at *P<0.05. **(C)** EMT marker protein levels as measured via western blotting. Protein levels of EMT markers (E-cadherin and vimentin) normalized to total protein staining.

Figure 6. (A) *CD2AP* and *TKS4* expression levels in colon cancer patients taken from TCGA and the Gent2 database (left panel). *CD2AP* and *TKS4* expression levels in Stage I-IV from GENT2 (right panel) **(B)** Colon Cancer Tissue cDNA array results of *CD2AP* and *TKS4* mRNA expression levels in normal and tumour samples (left panel). *CD2AP* and *TKS4* expression levels in tumour tissue from stage I, II, III and IV cancers (right panel). P values were calculated with Student's t-test or Tukey's test (with an ANOVA post-hoc test in the comparison of stages). Statistical significance was set at *P<0.05.

Figure 7. *CD2AP* and *TKS4* expression level correlations in colon cancer patient samples. (A) Correlation analysis of *CD2AP* and *TKS4* expression levels in non-tumorous (blue dots) **(B)** and

tumorous colon (pink dots) (COAD, colon adenocarcinoma) tissue from the TCGA-COAD database. Transcript levels are shown as gene expression RNAseq unitlog2(norm_count+1) values. R = Pearson correlation coefficient. Statistical significance was set at *P<0.05. **(C)** Merged dot-plot of the *CD2AP* and *TKS4* expression levels from the TCGA-COAD database (left panel), **(D)** GENT2 database (middle panel) and samples from tissue cDNA array (right panel), **(E)** CD2AP/TKS4 ratio clusters in normal and tumorous colon tissues.

Figure 8. (A) The CD2AP/TKS4 gene expression ratio in colon cancer patients from GENT2 data. P values were calculated with Student's t-test, and statistical significance was set at *P<0.05. **(B)** ROC curve analyses of TKS4, CD2AP and CD2AP/TKS4 ratios from GENT2 data. **(C)** Results of ROC curve analyses from GENT2 datasets

Figure 9. **Model of the role of the TKS4/CD2AP complex in cytoskeleton rearrangement and actin-based cell motility process**.

Figure 10. Epithelial and mesenchymal cell phenotypes, cell migrations and expression levels of EMT-related genes grouped according to the levels of the CD2AP and TKS4 proteins based on gene silencing experiments. E-cdh: E-cadherin, Vimentin, Zeb1.

Additional files

Supplementary Tables 1-2

Supplementary Table S1. Oligos used in the study

ID	Oligo name	Oligo sequence 5'-3'
1	GAPDH Forward	TGCACCACCAACTGCTTAGC
2	GAPDH Reverse	GGCATGGACTGTGGTCATGAG
3	PUM1 Forward	TGAGTTTATTCCTTCAGACCAGCAG
4	PUM1 Reverse	GCAAATACCTGTCCCTTAAACGCA
5	Zeb1 Forward	GGGAGGAGCAGTGAAAGAGA
6	Zeb1 Reverse	TTTCTTGCCCTTCCTTTCTG
7	Vimentin Forward	TACAGGAAGCTGCTGGAAGG
8	Vimentin Reverse	ACCAGAGGGAGTGAATCCAG
9	E-cadherin Forward	TGCCCAGAAAATGAAAAAGG
10	E-cadherin Reverse	GTGTATGTGGCAATGCGTTC
11	TKS4 Forward	GAGGTGAAGGTGCTAGACGT
12	TKS4 Reverse	GGTCCTTCTGTCCTCCTTCC
13	CD2AP Forward	AGGAATTCAGCCACATCCAC
14	CD2AP Reverse	TTGAGGGAAACAGTCCCAAC

Supplementary Table S2. TKS4 and CD2AP peptide array. PX and SH3 domaines are shown with bold letters, signals up to 10000 UA are grey.

Nr	Normal Scan	Start	End	Overlap	Normalized signal		# of residues with >30% RSA	# of residues with >50% RSA
1	MPPRRSIVEVKVLDV	1	15	0	2987,667		-	-
2	**RSIVEVKVLDVQKRR**	5	19	11	525	PX domain	-	-
3	**EVKVLDVQKRRVPNK**	9	23	11	173	PX domain	-	-
4	**LDVQKRRVPNKHYVY**	13	27	11	4051,333	PX domain	-	-
5	**KRRVPNKHYVYIIRV**	17	31	11	1561,667	PX domain	-	-
6	**PNKHYVYIIRVTWSS**	21	35	11	1361,333	PX domain	-	-
7	**YVYIIRVTWSSGSTE**	25	39	11	5448	PX domain	-	-
8	**IRVTWSSGSTEAIYR**	29	43	11	723,6667	PX domain	-	-
9	**WSSGSTEAIYRRYSK**	33	47	11	668,6667	PX domain	-	-
10	**STEAIYRRYSKFFDL**	37	51	11	6294,333	PX domain	-	-
11	**IYRRYSKFFDLQMQM**	41	55	11	8	PX domain	-	-
12	**YSKFFDLQMQMLDKF**	45	59	11	9811,667	PX domain	-	-
13	**FDLQMQMLDKFPMEG**	49	63	11	56,33333	PX domain	-	-
14	**MQMLDKFPMEGGQKD**	53	67	11	2	PX domain	-	-
15	**DKFPMEGGQKDPKQR**	57	71	11	2	PX domain	-	-
16	**MEGGQKDPKQRIIPF**	61	75	11	2	PX domain	-	-
17	**QKDPKQRIIPFLPGK**	65	79	11	2	PX domain	-	-
18	**KQRIIPFLPGKILFR**	69	83	11	852	PX domain	-	-
19	**IPFLPGKILFRRSHI**	73	87	11	2900	PX domain	-	-
20	**PGKILFRRSHIRDVA**	77	91	11	3125,333	PX domain	-	-
21	**LFRRSHIRDVAVKRL**	81	95	11	3930,333	PX domain	-	-
22	**SHIRDVAVKRLIPID**	85	99	11	97,33333	PX domain	-	-
23	**DVAVKRLIPIDEYCK**	89	103	11	4131	PX domain	-	-
24	**KRLIPIDEYCKALIQ**	93	107	11	5449,667	PX domain	-	-
25	**PIDEYCKALIQLPPY**	97	111	11	9696,333	PX domain	-	-
26	**YCKALIQLPPYISQC**	101	115	11	8525,667	PX domain	-	-
27	**LIQLPPYISQCDEVL**	105	119	11	10131,33	PX domain	-	-
28	**PPYISQCDEVLQFFE**	109	123	11	11305	PX domain	6	3
29	**SQCDEVLQFFETRPE**	113	127	11	8778,667	PX domain	-	-
30	**EVLQFFETRPEDLNP**	117	131	11	11681,67	PX domain	7	5
31	**FFETRPEDLNPPKEE**	121	135	11	203,3333	PX domain	-	-
32	**RPEDLNPPKEEHIGK**	125	139	11	70,66667	PX domain	-	-
33	LNPPKEEHIGKKKSG	129	143	11	2		-	-
34	KEEHIGKKKSGGDQT	133	147	11	2		-	-
35	IGKKKSGGDQTSVDP	137	151	11	2		-	-
36	**KSGGDQTSVDPMVLE**	141	155	11	4966,333	SH3 1	-	-
37	**DQTSVDPMVLEQYVV**	145	159	11	7169,667	SH3 1	-	-
38	**VDPMVLEQYVVVANY**	149	163	11	7313	SH3 1	-	-
39	**VLEQYVVVANYQKQE**	153	167	11	6774,667	SH3 1	-	-
40	**YVVVANYQKQESSEI**	157	171	11	8059,333	SH3 1	-	-
41	**ANYQKQESSEISLSV**	161	175	11	3370	SH3 1	-	-
42	**KQESSEISLSVGQVV**	165	179	11	7068,333	SH3 1	-	-
43	**SEISLSVGQVVDIIE**	169	183	11	7144	SH3 1	-	-
44	**LSVGQVVDIIEKNES**	173	187	11	12069,67	SH3 1	7	0
45	**QVVDIIEKNESGWWF**	177	191	11	18487,67	SH3 1	4	0

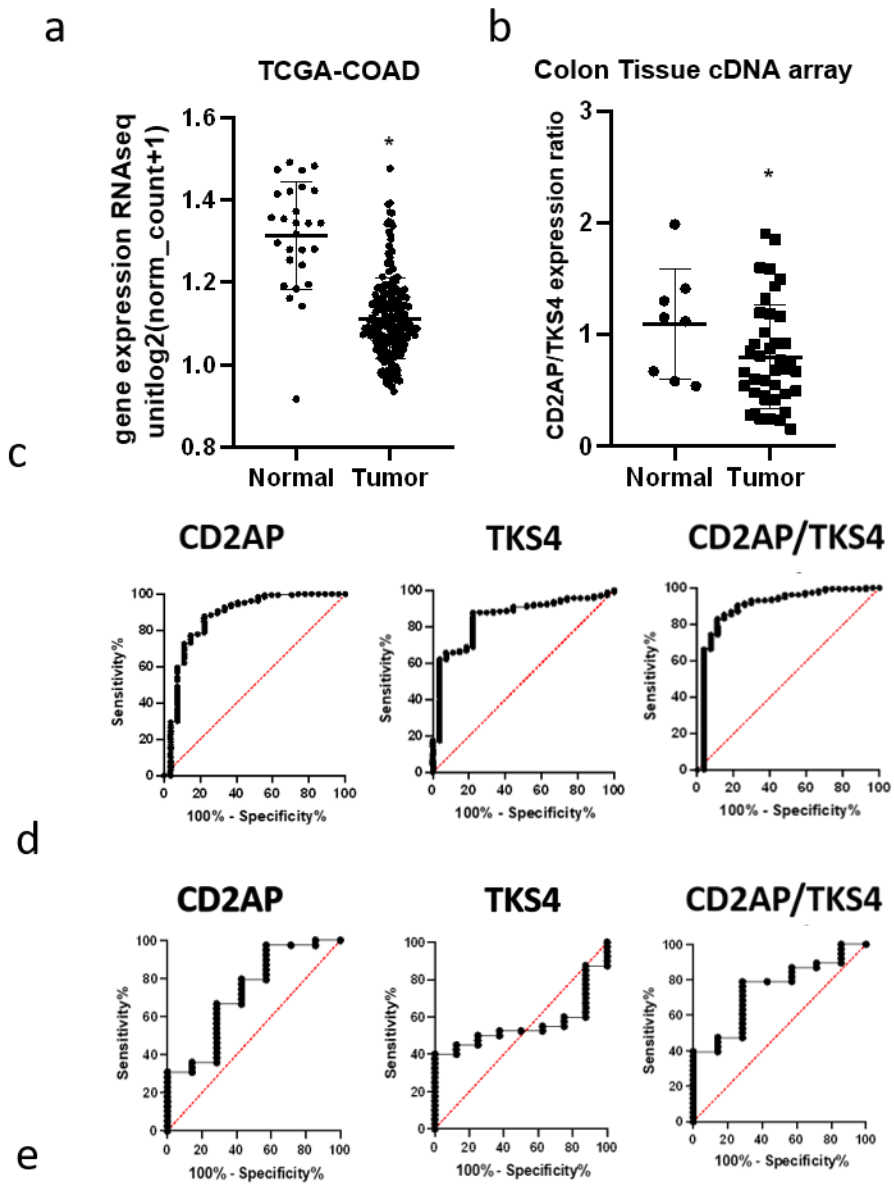
46	**IIEKNESGWWFVSTA**	181	195	11	5744	SH3 1	-	-
47	**NESGWWFVSTAEEQG**	185	199	11	13133,67	SH3 1	4	2
48	**WWFVSTAEEQGWVPA**	189	203	11	7080,333	SH3 1	-	-
49	**STAEEQGWVPATCLE**	193	207	11	14226,33	SH3 1	3	3
50	**EQGWVPATCLEGQDG**	197	211	11	5918	SH3 1	-	-
51	VPATCLEGQDGVQDE	201	215	11	2327,667		-	-
52	CLEGQDGVQDEFSLQ	205	219	11	18002,67		11	8
53	QDGVQDEFSLQPEEE	209	223	11	10212,67		15	9
54	**QDEFSLQPEEEEKYT**	213	227	11	11724,33	SH3 2	12	6
55	**SLQPEEEEKYTVIYP**	217	231	11	20852,67	SH3 2	10	5
56	**EEEEKYTVIYPYTAR**	221	235	11	17800,33	SH3 2	8	5
57	**KYTVIYPYTARDQDE**	225	239	11	20815	SH3 2	8	4
58	**IYPYTARDQDEMNLE**	229	243	11	2028,667	SH3 2	-	-
59	**TARDQDEMNLERGAV**	233	247	11	311,3333	SH3 2	-	-
60	**QDEMNLERGAVVEVI**	237	251	11	7725	SH3 2	-	-
61	**NLERGAVVEVIQKNL**	241	255	11	2916	SH3 2	-	-
62	**GAVVEVIQKNLEGWW**	245	259	11	17648	SH3 2	3	0
63	**EVIQKNLEGWWKIRY**	249	263	11	5197,667	SH3 2	-	-
64	**KNLEGWWKIRYQGKE**	253	267	11	2	SH3 2	-	-
65	**GWWKIRYQGKEGWAP**	257	271	11	246	SH3 2	-	-
66	**IRYQGKEGWAPASYL**	261	275	11	2493,667	SH3 2	-	-
67	**GKEGWAPASYLKKNS**	265	279	11	2	SH3 2	-	-
68	**WAPASYLKKNSGEPL**	269	283	11	2	SH3 2	-	-
69	**SYLKKNSGEPLPPKP**	273	287	11	2	SH3 2	-	-
70	**KNSGEPLPPKPGPGS**	277	291	11	2	SH3 2	-	-
71	EPLPPKPGPGSPSHP	281	295	11	2		-	-
72	PKPGPGSPSHPGALD	285	299	11	2		-	-
73	PGSPSHPGALDLDGV	289	303	11	2474		-	-
74	SHPGALDLDGVSRQQ	293	307	11	2		-	-
75	ALDLDGVSRQQNAVG	297	311	11	2397		-	-
76	DGVSRQQNAVGREKE	301	315	11	46,33333		-	-
77	RQQNAVGREKELLSS	305	319	11	368,6667		-	-
78	AVGREKELLSSQRDG	309	323	11	2		-	-
79	EKELLSSQRDGRFEG	313	327	11	6436		-	-
80	LSSQRDGRFEGRPVP	317	331	11	324,3333		-	-
81	RDGRFEGRPVPDGDA	321	335	11	1313,667		-	-
82	FEGRPVPDGDAKQRS	325	339	11	2		-	-
83	PVPDGDAKQRSPKMR	329	343	11	2		-	-
84	GDAKQRSPKMRQRPP	333	347	11	3294		-	-
85	QRSPKMRQRPPPRRD	337	351	11	9482,667		-	-
86	KMRQRPPPRRDMTIP	341	355	11	1697,333		-	-
87	RPPPRRDMTIPRGLN	345	359	11	322,3333		-	-
88	RRDMTIPRGLNLPKP	349	363	11	162		-	-
89	TIPRGLNLPKPPIPP	353	367	11	2		-	-
90	**GLNLPKPPIPPQVEE**	357	371	11	247,6667	SH3 3	-	-
91	**PKPPIPPQVEEEYYT**	361	375	11	9667,333	SH3 3	-	-

92	**IPPQVEEEYYTIAEF**	365	379	11	24300,67	SH3 3	10	6
93	**VEEEYYTIAEFQTTI**	369	383	11	4183	SH3 3	-	-
94	**YYTIAEFQTTIPDGI**	373	387	11	21021	SH3 3	8	5
95	**AEFQTTIPDGISFQA**	377	391	11	3174,333	SH3 3	-	-
96	**TTIPDGISFQAGLKV**	381	395	11	1804,333	SH3 3	-	-
97	**DGISFQAGLKVEVIE**	385	399	11	8070,333	SH3 3	-	-
98	**FQAGLKVEVIEKNLS**	389	403	11	4385	SH3 3	-	-
99	**LKVEVIEKNLSGWWY**	393	407	11	14586,67	SH3 3	9	2
100	**VIEKNLSGWWYIQIE**	397	411	11	11530	SH3 3	8	2
101	**NLSGWWYIQIEDKEG**	401	415	11	30918,33	SH3 3	8	4
102	**WWYIQIEDKEGWAPA**	405	419	11	21478,33	SH3 3	4	3
103	**QIEDKEGWAPATFID**	409	423	11	17619	SH3 3	6	4
104	**KEGWAPATFIDKYKK**	413	427	11	2	SH3 3	-	-
105	APATFIDKYKKTSNA	417	431	11	2		-	-
106	FIDKYKKTSNASRPN	421	435	11	2		-	-
107	YKKTSNASRPNFLAP	425	439	11	2		-	-
108	SNASRPNFLAPLPHE	429	443	11	777		-	-
109	RPNFLAPLPHEVTQL	433	447	11	421		-	-
110	LAPLPHEVTQLRLGE	437	451	11	1786,333		-	-
111	PHEVTQLRLGEAAAL	441	455	11	160,6667		-	-
112	TQLRLGEAAALENNT	445	459	11	48,66667		-	-
113	LGEAAALENNTGSEA	449	463	11	2		-	-
114	AALENNTGSEATGPS	453	467	11	257,3333		-	-
115	NNTGSEATGPSRPLP	457	471	11	14,33333		-	-
116	SEATGPSRPLPDAPH	461	475	11	313		-	-
117	GPSRPLPDAPHGVMD	465	479	11	27,33333		-	-
118	PLPDAPHGVMDSGLP	469	483	11	2		-	-
119	APHGVMDSGLPWSKD	473	487	11	2		-	-
120	VMDSGLPWSKDWKGS	477	491	11	453,3333		-	-
121	GLPWSKDWKGSKDVL	481	495	11	55,66667		-	-
122	SKDWKGSKDVLRKAS	485	499	11	2		-	-
123	KGSKDVLRKASSDMS	489	503	11	2		-	-
124	DVLRKASSDMSASAG	493	507	11	2		-	-
125	KASSDMSASAGYEEI	497	511	11	1444,333		-	-
126	DMSASAGYEEISDPD	501	515	11	10176		-	-
127	SAGYEEISDPDMEEK	505	519	11	589		-	-
128	EEISDPDMEEKPSLP	509	523	11	2483,667		-	-
129	DPDMEEKPSLPPRKE	513	527	11	2		-	-
130	EEKPSLPPRKESIIK	517	531	11	5320		-	-
131	SLPPRKESIIKSEGE	521	535	11	2		-	-
132	RKESIIKSEGELLER	525	539	11	2509,333		-	-
133	IIKSEGELLERERER	529	543	11	559		-	-
134	EGELLERERERQRTE	533	547	11	3798,667		-	-
135	LERERERQRTEQLRG	537	551	11	993		-	-
136	RERQRTEQLRGPTPK	541	555	11	2		-	-
137	RTEQLRGPTPKPPGV	545	559	11	2		-	-

138	LRGPTPKPPGVILPM	549	563	11	2		-	-
139	TPKPPGVILPMMPAK	553	567	11	260,3333		-	-
140	PGVILPMMPAKHIPP	557	571	11	2		-	-
141	LPMMPAKHIPPARDS	561	575	11	2		-	-
142	PAKHIPPARDSRRPE	565	579	11	2		-	-
143	IPPARDSRRPEPKPD	569	583	11	2		-	-
144	RDSRRPEPKPDKSRL	573	587	11	2		-	-
145	RPEPKPDKSRLFQLK	577	591	11	2		-	-
146	KPDKSRLFQLKNDMG	581	595	11	2		-	-
147	SRLFQLKNDMGLECG	585	599	11	1003		-	-
148	QLKNDMGLECGHKVL	589	603	11	2		-	-
149	DMGLECGHKVLAKEV	593	607	11	602,6667		-	-
150	ECGHKVLAKEVKKPN	597	611	11	2		-	-
151	KVLAKEVKKPNLRPI	601	615	11	2		-	-
152	KEVKKPNLRPISKSK	605	619	11	2		-	-
153	KPNLRPISKSKTDLP	609	623	11	2		-	-
154	RPISKSKTDLPEEKP	613	627	11	2		-	-
155	KSKTDLPEEKPDATP	617	631	11	2		-	-
156	DLPEEKPDATPQNPF	621	635	11	1075		-	-
157	EKPDATPQNPFLKSR	625	639	11	2		-	-
158	ATPQNPFLKSRPQVR	629	643	11	2		-	-
159	NPFLKSRPQVRPKPA	633	647	11	2		-	-
160	KSRPQVRPKPAPSPK	637	651	11	2		-	-
161	QVRPKPAPSPKTEPP	641	655	11	2		-	-
162	KPAPSPKTEPPQGED	645	659	11	2		-	-
163	SPKTEPPQGEDQVDI	649	663	11	579,6667		-	-
164	EPPQGEDQVDICNLR	653	667	11	1484		-	-
165	GEDQVDICNLRSKLR	657	671	11	397,6667		-	-
166	VDICNLRSKLRPAKS	661	675	11	617		-	-
167	NLRSKLRPAKSQDKS	665	679	11	2		-	-
168	KLRPAKSQDKSLLDG	669	683	11	15,66667		-	-
169	AKSQDKSLLDGEGPQ	673	687	11	597,3333		-	-
170	DKSLLDGEGPQAVGG	677	691	11	2		-	-
171	LDGEGPQAVGGQDVA	681	695	11	2703		-	-
172	GPQAVGGQDVAFSRS	685	699	11	26,66667		-	-
173	VGGQDVAFSRSFLPG	689	703	11	2419		-	-
174	DVAFSRSFLPGEGPG	693	707	11	3056,333		-	-
175	SRSFLPGEGPGRAQD	697	711	11	2		-	-
176	LPGEGPGRAQDRTGK	701	715	11	2		-	-
177	GPGRAQDRTGKQDGL	705	719	11	2		-	-
178	AQDRTGKQDGLSPKE	709	723	11	2		-	-
179	TGKQDGLSPKEISCR	713	727	11	2		-	-
180	DGLSPKEISCRAPPR	717	731	11	2		-	-
181	PKEISCRAPPRPAKT	721	735	11	57,33333		-	-
182	SCRAPPRPAKTTDPV	725	739	11	2		-	-
183	PPRPAKTTDPVSKSV	729	743	11	2		-	-

184	AKTTDPVSKSVPVPL	733	747	11	2		-	-
185	DPVSKSVPVPLQEAP	737	751	11	195,3333		-	-
186	KSVPVPLQEAPQQRP	741	755	11	2		-	-
187	VPLQEAPQQRPVVPP	745	759	11	10,66667		-	-
188	EAPQQRPVVPPRRPP	749	763	11	23601		12	10
189	QRPVVPPRRPPPPKK	753	767	11	38921,67		12	10
190	VPPRRPPPPKKTSSS	757	771	11	639		-	-
191	RPPPPKKTSSSSRPL	761	775	11	53		-	-
192	PKKTSSSSRPLPEVR	765	779	11	2		-	-
193	SSSSRPLPEVRGPQC	769	783	11	2		-	-
194	RPLPEVRGPQCEGHE	773	787	11	272,6667		-	-
195	EVRGPQCEGHESRAA	777	791	11	2		-	-
196	PQCEGHESRAAPTPG	781	795	11	2		-	-
197	GHESRAAPTPGRALL	785	799	11	2		-	-
198	RAAPTPGRALLVPPK	789	803	11	2		-	-
199	TPGRALLVPPKAKPF	793	807	11	2		-	-
200	ALLVPPKAKPFLSNS	797	811	11	2		-	-
201	PPKAKPFLSNSLGGQ	801	815	11	2		-	-
202	KPFLSNSLGGQDDTR	805	819	11	386,3333		-	-
203	SNSLGGQDDTRGKGS	809	823	11	247		-	-
204	GGQDDTRGKGSLGPW	813	827	11	1718,667		-	-
205	DTRGKGSLGPWGTGK	817	831	11	42,66667		-	-
206	KGSLGPWGTGKIGEN	821	835	11	26		-	-
207	GPWGTGKIGENREKA	825	839	11	2		-	-
208	TGKIGENREKAAAAS	829	843	11	2		-	-
209	GENREKAAAASVPNA	833	847	11	2		-	-
210	EKAAAASVPNADGLK	837	851	11	2		-	-
211	**AASVPNADGLKDSLY**	841	855	11	2151,667	SH3 4	-	-
212	**PNADGLKDSLYVAVA**	845	859	11	7170,333	SH3 4	-	-
213	**GLKDSLYVAVADFEG**	849	863	11	12682,67	SH3 4	6	5
214	**SLYVAVADFEGDKDT**	853	867	11	24445,67	SH3 4	5	4
215	**AVADFEGDKDTSSFQ**	857	871	11	15045,67	SH3 4	8	5
216	**FEGDKDTSSFQEGTV**	861	875	11	8304,333	SH3 4	-	-
217	**KDTSSFQEGTVFEVR**	865	879	11	3849	SH3 4	-	-
218	**SFQEGTVFEVREKNS**	869	883	11	4679,333	SH3 4	-	-
219	**GTVFEVREKNSSGWW**	873	887	11	6588	SH3 4	-	-
220	**EVREKNSSGWWFCQV**	877	891	11	6668,667	SH3 4	-	-
221	**KNSSGWWFCQVLSGA**	881	895	11	3578,333	SH3 4	-	-
222	**GWWFCQVLSGAPSWE**	885	899	11	9014,667	SH3 4	-	-
223	**CQVLSGAPSWEGWIP**	889	903	11	4079,333	SH3 4	-	-
224	**SGAPSWEGWIPSNYL**	893	907	11	11970,67	SH3 4	6	4
225	**SWEGWIPSNYLRKKP**	897	911	11	61,66667	SH3 4	-	-

Supplementary Figure S1.



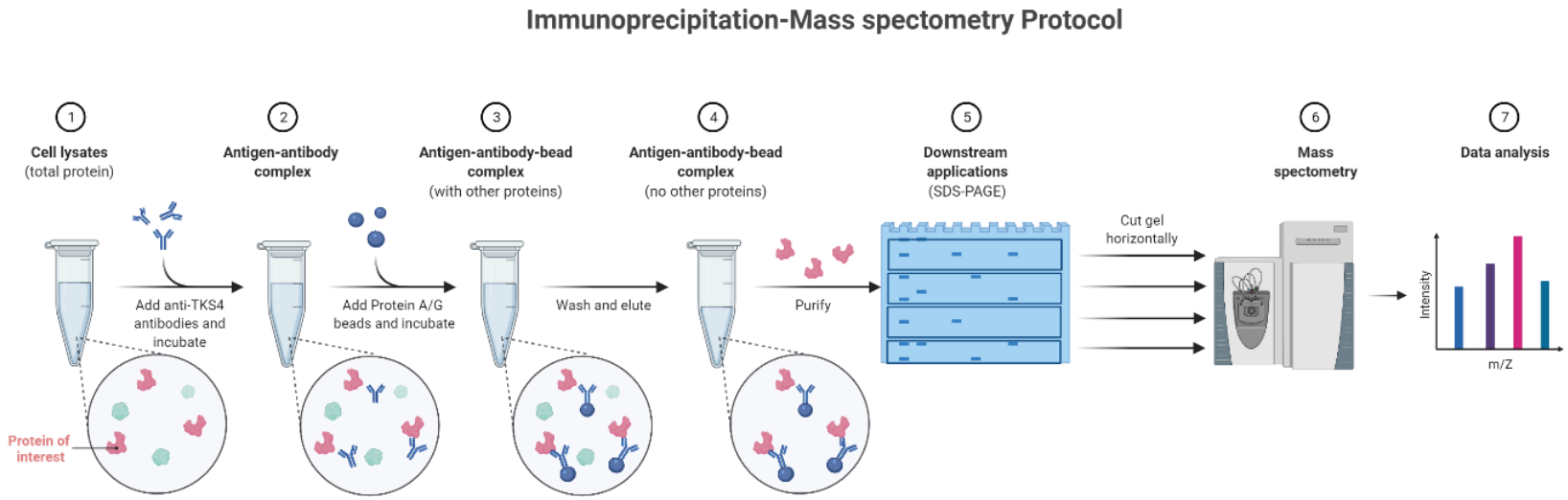
	TCGA-COAD database			Colon Tissue cDNA array		
	CD2AP	**TKS4**	**CD2AP/TKS4**	**CD2AP**	**TKS4**	**CD2AP/TKS4**
Sensitivity	87,89	87,89	83,16	97,3	40,54	81,08
Specificity	77,78	77,78	88,89	42,86	100	71,43
Optimal cutoff value	≤11.59	>9.274	≤1.1836	≤1.4316	>1.3519	>0.8914
AUC	0,8733	0,8487	0,8981	0,7181	0,5598	0,7529
p-value	**<0,0001**	**<0,0001**	**<0,0001**	0,0698	0,6189	**0,0356**

Supplementary Figure S1. ***a)*** *The CD2AP/*TKS4 *gene expression ratio in colon cancer patients from TCGA-COAD database and* ***(b)*** *based on the Colon Cancer Tissue cDNA array results.* P values were calculated with Student's t-test, and statistical significance was set at *P<0.05. ***(c)*** *ROC curve analyses of TKS4, CD2AP and CD2AP/TKS4* ratios from TCGA-COAD database. ***(d)*** ROC curve analyses of

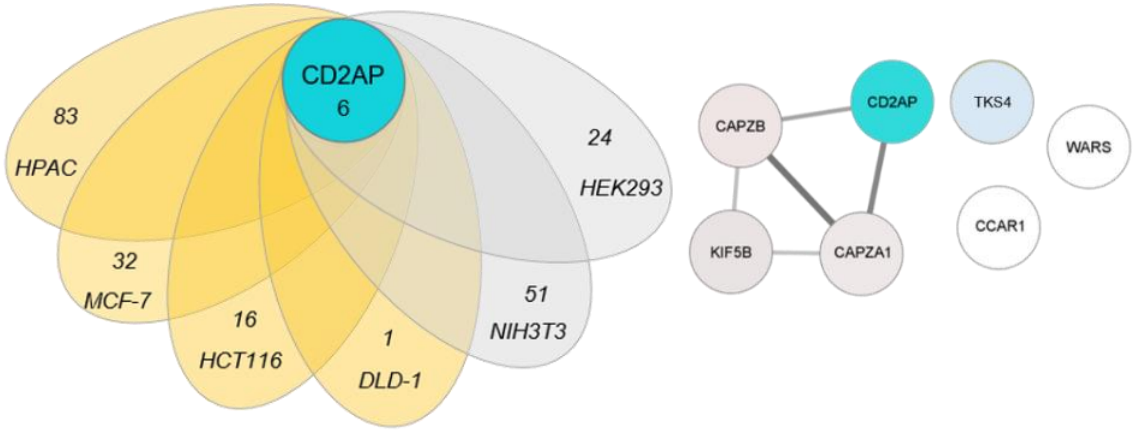
TKS4, CD2AP, and TKS4/CD2AP ratios to discriminate normal and tumour tissue from the Colon Cancer Tissue cDNA array. **(e)** Results of ROC curve analyses from TCGA or Tissue cDNA array datasets.

Fig. 1.

A


Immunoprecipitation-Mass spectometry Protocol
1
Cell lysates
(total protein)
2
Antigen-antibody complex
3
Antigen-antibody-bead complex
(with other proteins)
4
Antigen-antibody-bead complex
(no other proteins)
5
Downstream applications
(SDS-PAGE)
6
Mass spectometry
7
Data analysis
Add anti-TKS4 antibodies and incubate
Add Protein A/G beads and incubate
Wash and elute
Purify
Cut gel horizontally
Protein of interest
Intensity
m/Z

B


CD2AP
6
83
HPAC
32
MCF-7
16
HCT116
1
DLD-1
51
NIH3T3
24
HEK293
CAPZB
CD2AP
TKS4
WARS
KIF5B
CAPZA1
CCAR1

C

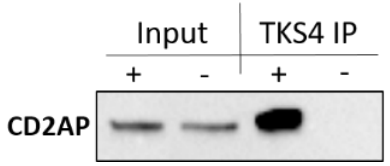

Input
TKS4 IP
+
-
+
-
CD2AP

Fig. 2.

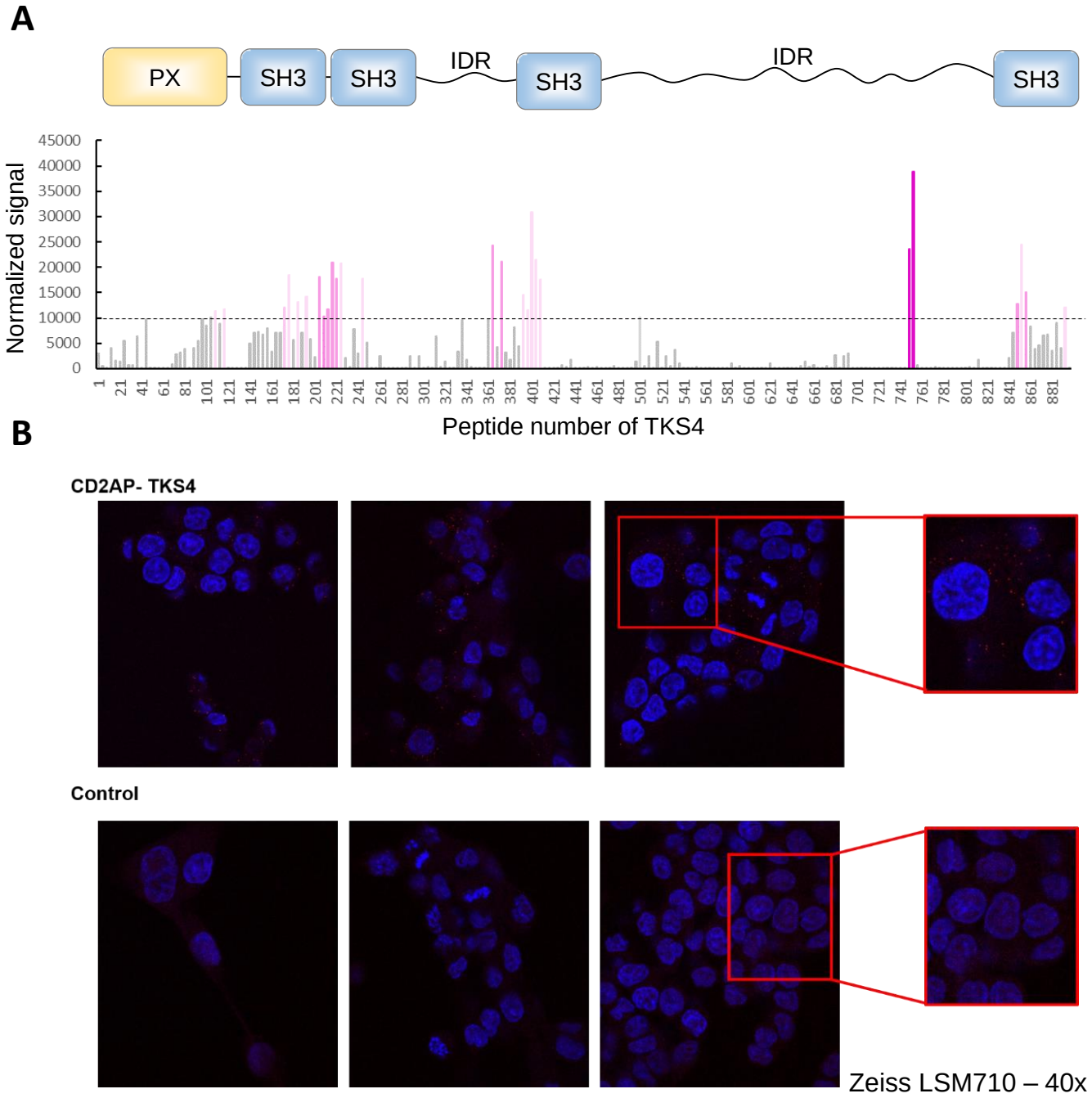

A
PX
SH3
SH3
IDR
SH3
IDR
SH3
Normalized signal
45000
40000
35000
30000
25000
20000
15000
10000
5000
0
Peptide number of TKS4
B
CD2AP- TKS4
Control
Zeiss LSM710 – 40x

Fig. 3.

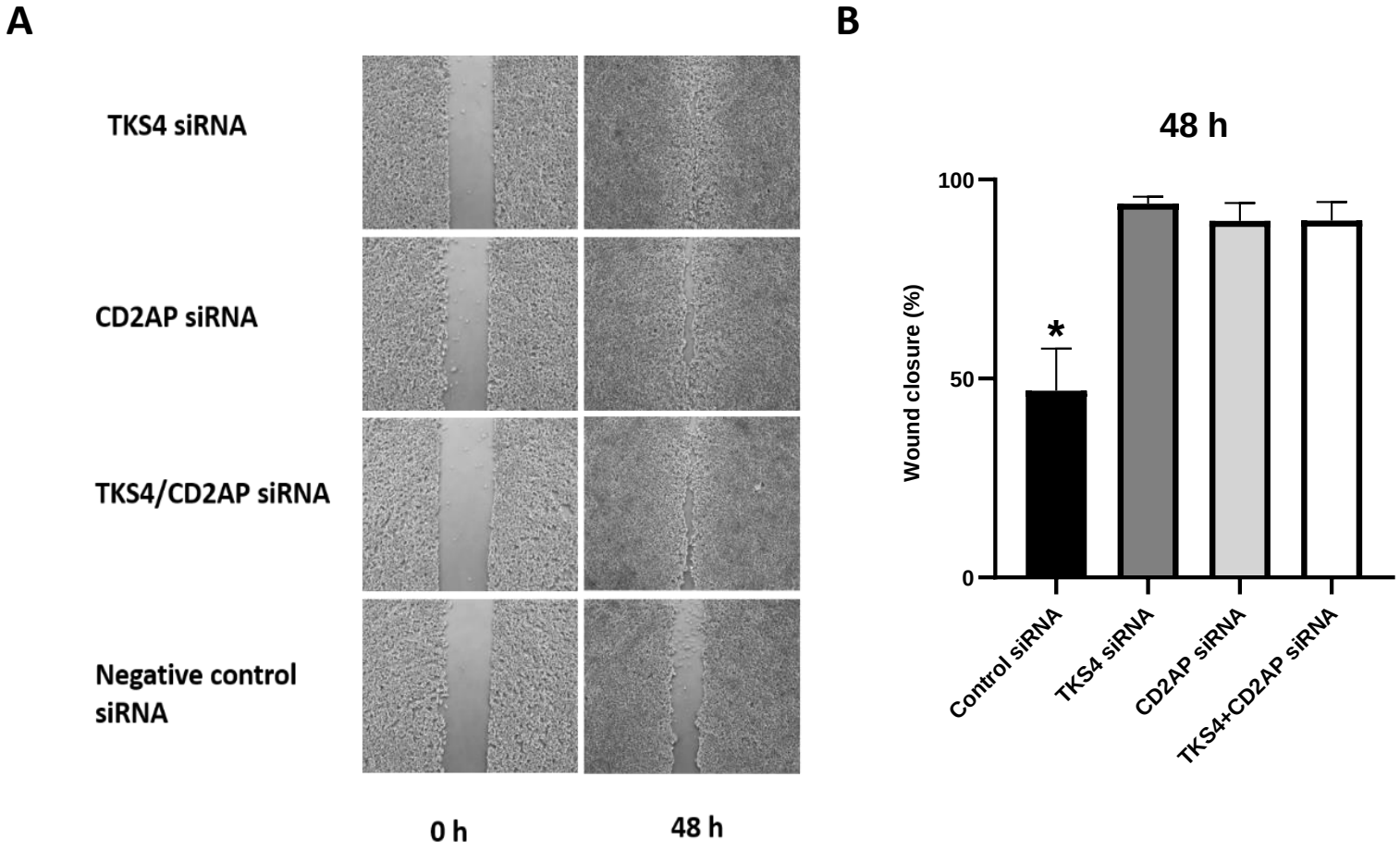
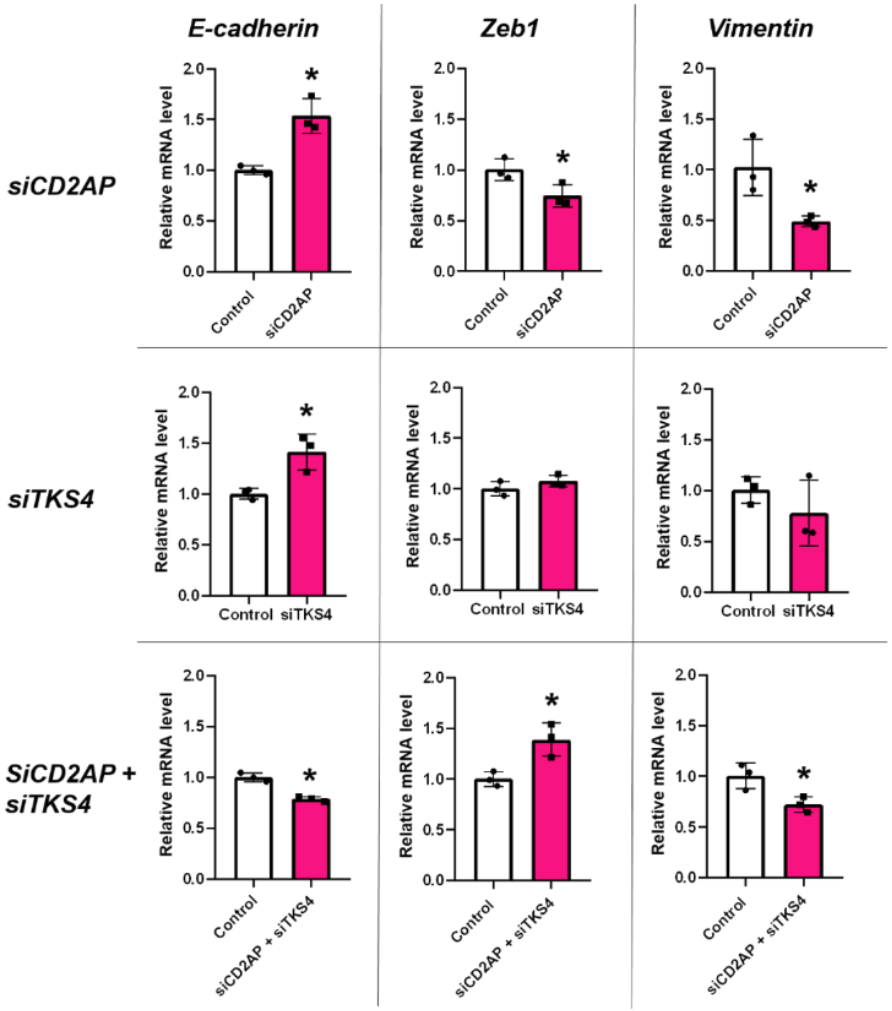

A
TKS4 siRNA
CD2AP siRNA
TKS4/CD2AP siRNA
Negative control siRNA
0 h
48 h
B
48 h
Wound closure (%)
100
50
0
*
Control siRNA
TKS4 siRNA
CD2AP siRNA
TKS4+CD2AP siRNA

Fig. 4.

A


E-cadherin
Zeb1
Vimentin
siCD2AP
siTKS4
SiCD2AP + siTKS4
Relative mRNA level
2.0
1.5
1.0
0.5
0.0
*
Control
siCD2AP
siTKS4
siCD2AP + siTKS4

B

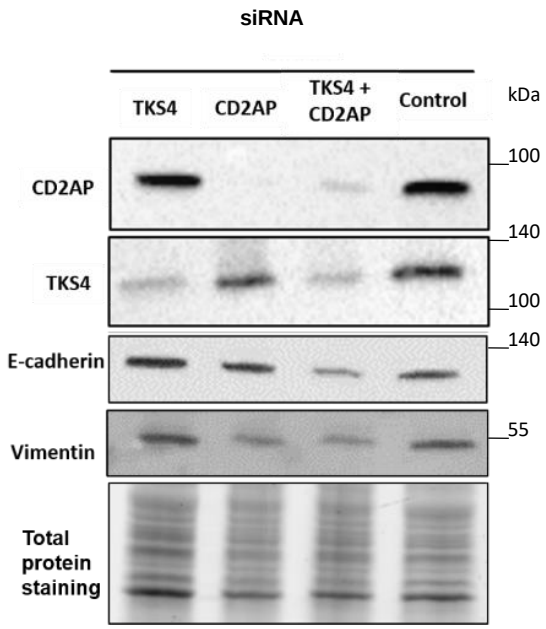

siRNA
TKS4
CD2AP
TKS4 + CD2AP
Control
kDa
CD2AP
TKS4
E-cadherin
Vimentin
Total protein staining
100
140
100
140
55

Fig. 5.

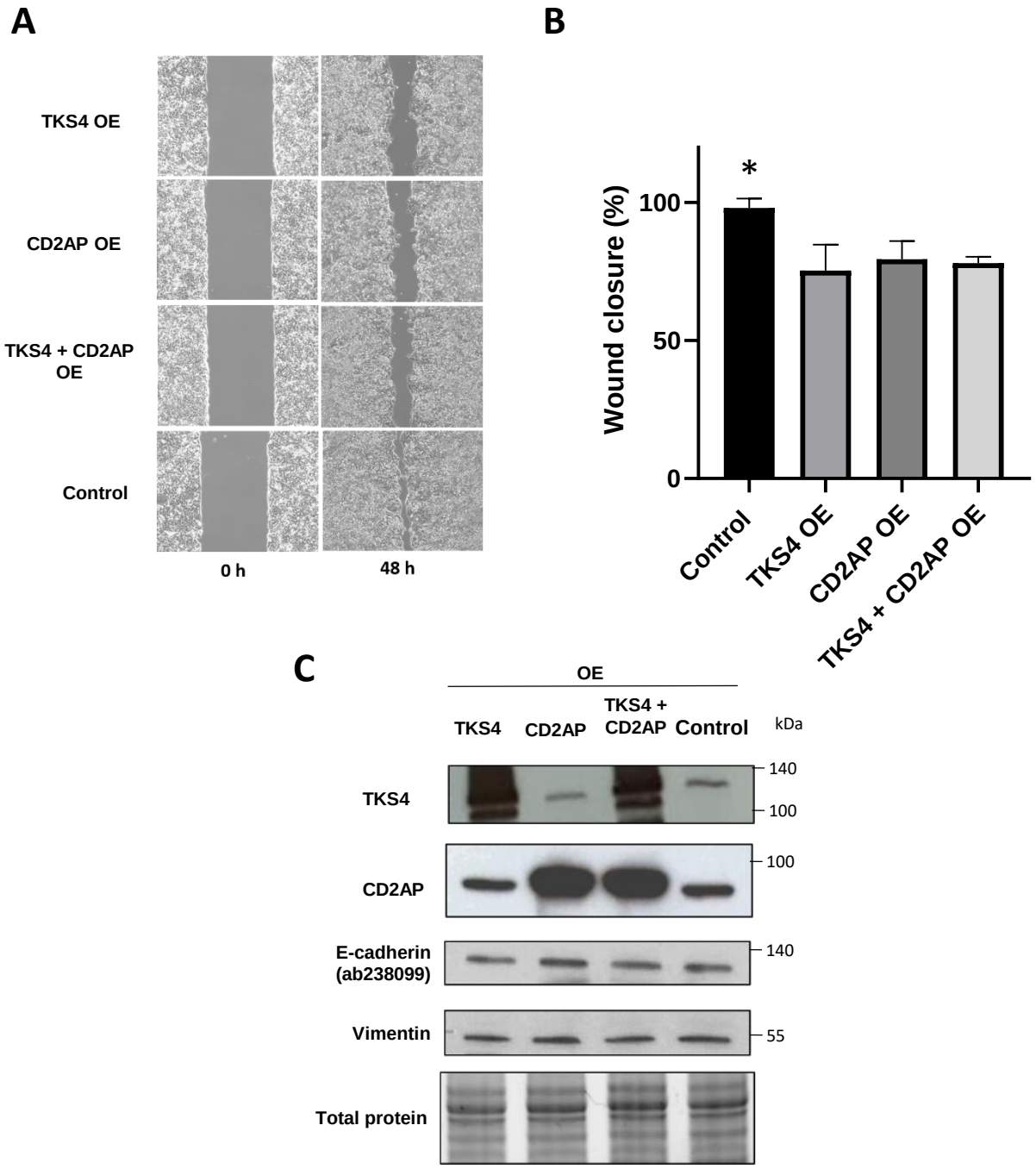

A
TKS4 OE
CD2AP OE
TKS4 + CD2AP OE
Control
0 h
48 h
B
Wound closure (%)
100
50
0
*
Control
TKS4 OE
CD2AP OE
TKS4 + CD2AP OE
C
OE
TKS4
CD2AP
TKS4 + CD2AP
Control
kDa
TKS4
140
100
CD2AP
100
E-cadherin (ab238099)
140
Vimentin
55
Total protein

Fig. 6.

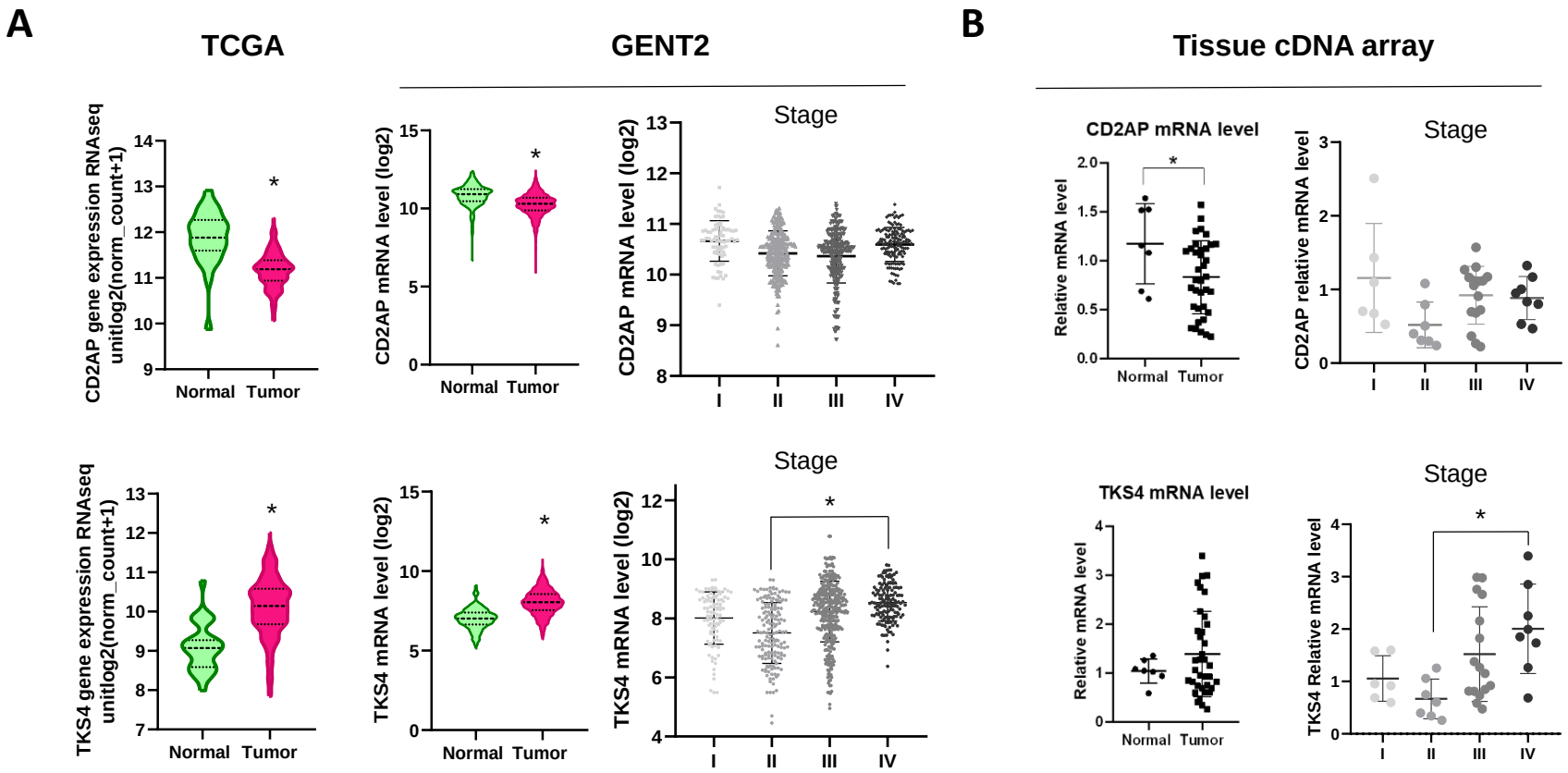

A
TCGA
GENT2
B
Tissue cDNA array
CD2AP gene expression RNAseq unitlog2(norm_count+1)
CD2AP mRNA level (log2)
Stage
CD2AP mRNA level
Relative mRNA level
CD2AP relative mRNA level
TKS4 gene expression RNAseq unitlog2(norm_count+1)
TKS4 mRNA level (log2)
TKS4 mRNA level
TKS4 Relative mRNA level
Normal
Tumor
I
II
III
IV
*

Fig. 7.

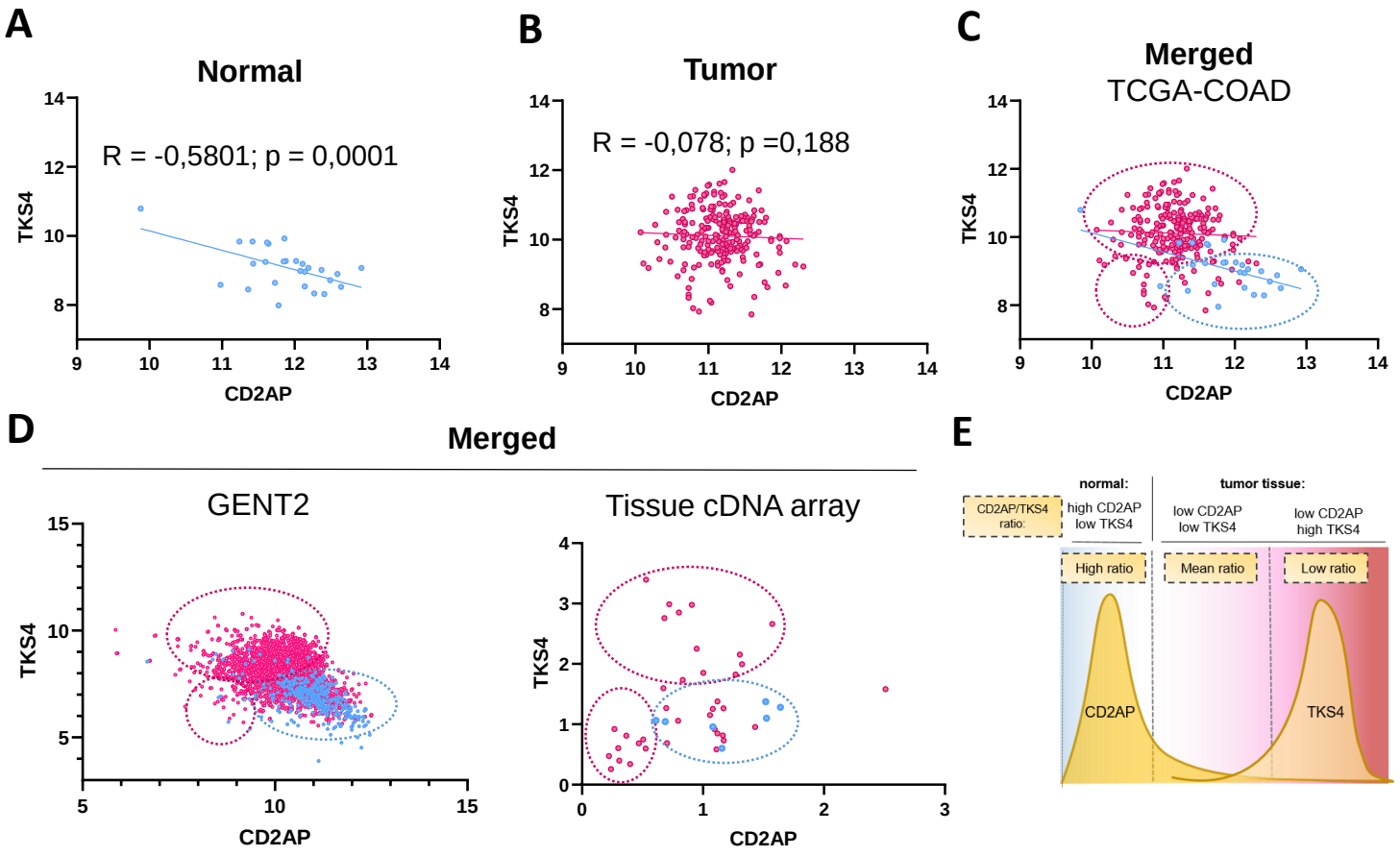
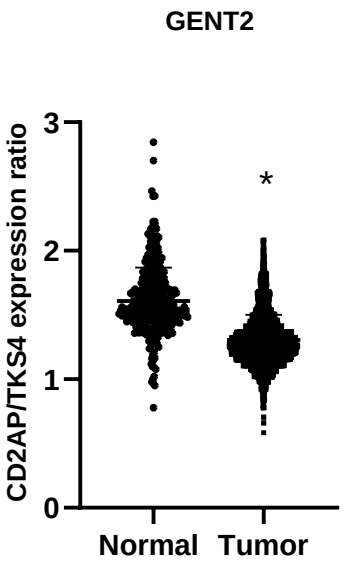
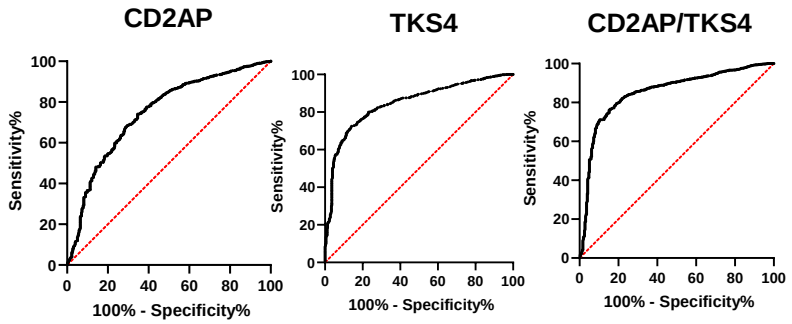

A
Normal
R = -0,5801; p = 0,0001
TKS4
CD2AP
B
Tumor
R = -0,078; p =0,188
TKS4
CD2AP
C
Merged
TCGA-COAD
TKS4
CD2AP
D
Merged
GENT2
TKS4
CD2AP
Tissue cDNA array
TKS4
CD2AP
E
normal:
tumor tissue:
CD2AP/TKS4 ratio:
high CD2AP low TKS4
low CD2AP low TKS4
low CD2AP high TKS4
High ratio
Mean ratio
Low ratio
CD2AP
TKS4

Fig. 8.

A



B



C

	GENT2 database		
	CD2AP	**TKS4**	**CD2AP/TKS4**
Sensitivity	74,01	72,08	71,14
Specificity	65,35	85,92	89,61
Optimal cutoff value	≤10.67	>7.618	≤1.36
AUC	0,7461	0,8470	0,8529
p-value	**<0,0001**	**<0,0001**	**<0,0001**

Fig. 9.

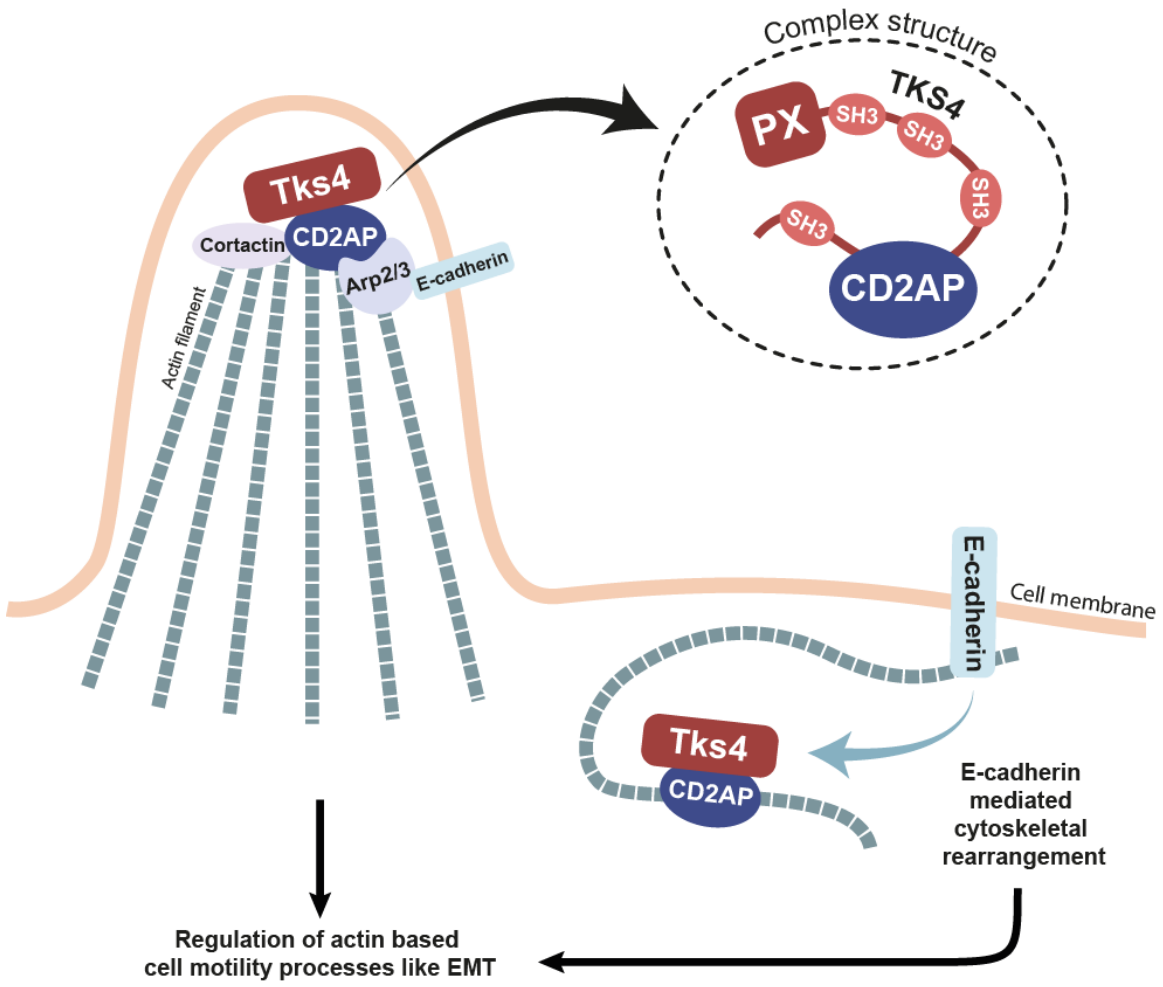

Complex structure
TKS4
PX
SH3
SH3
SH3
SH3
CD2AP
Tks4
Cortactin
CD2AP
Arp2/3
E-cadherin
Actin filament
E-cadherin
Cell membrane
Tks4
CD2AP
E-cadherin mediated cytoskeletal rearrangement
Regulation of actin based cell motility processes like EMT

Fig. 10.

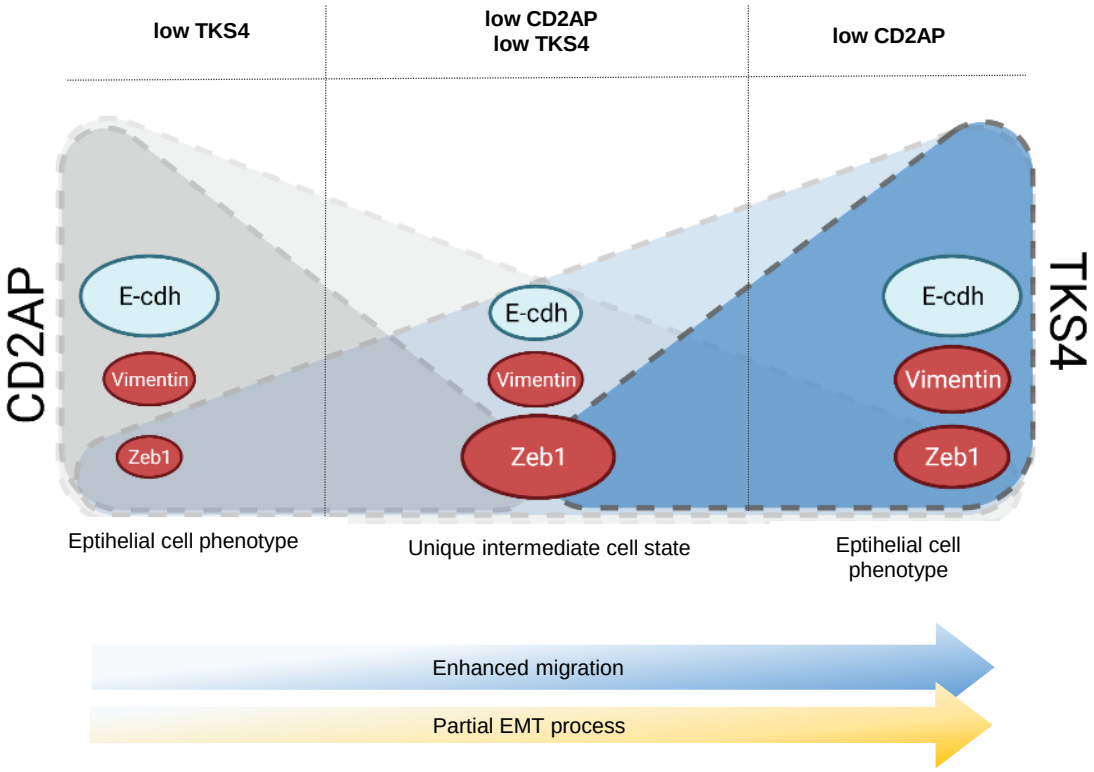
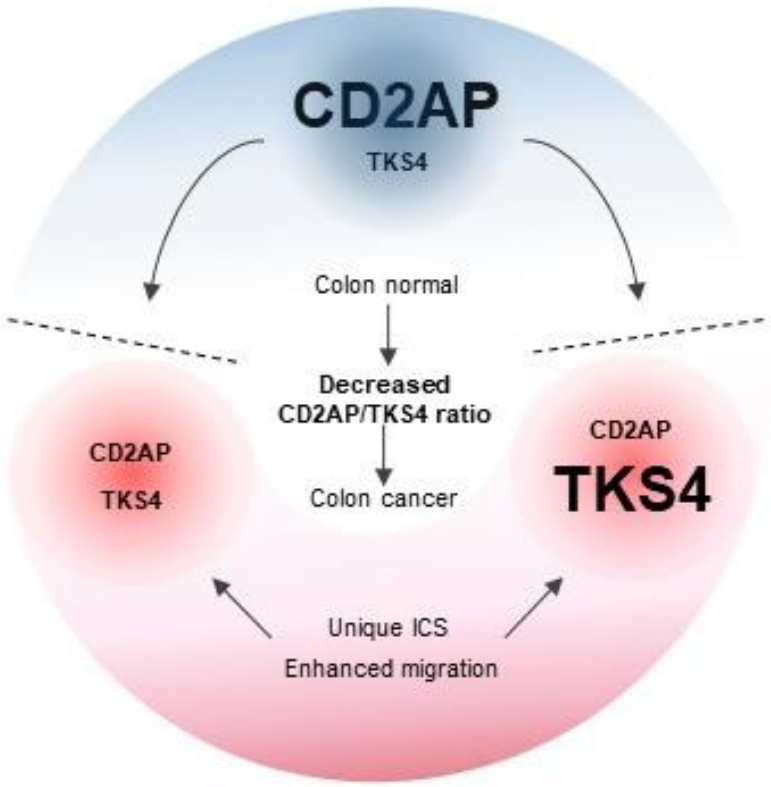

low TKS4
low CD2AP
low TKS4
low CD2AP
CD2AP
TKS4
E-cdh
Vimentin
Zeb1
E-cdh
Vimentin
Zeb1
E-cdh
Vimentin
Zeb1
Eptihelial cell phenotype
Unique intermediate cell state
Eptihelial cell phenotype
Enhanced migration
Partial EMT process

Fig. 11.


CD2AP
TKS4
Colon normal
Decreased
CD2AP/TKS4 ratio
Colon cancer
CD2AP
TKS4
CD2AP
TKS4
Unique ICS
Enhanced migration