Transcriptomics and proteomics reveal distinct biology for lymph node metastases and tumour deposits in colorectal cancer

Brouwer, Nelleke Pm ; Webbink, Loth ; Haddad, Tariq S ; Rutgers, Natasja ; van Vliet, Shannon ; Wood, Colin S ; Jansen, Pascal Wtc ; Lafarge, Maxime W ; imCMS Consortium ; de Wilt, Johannes Hw ; Hugen, Niek ; Simmer, Femke ; Jamieson, Nigel B ; Tauriello, Daniele Vf ; Kölzer, Viktor H ; Vermeulen, Michiel ; Nagtegaal, Iris D

Abstract: Both lymph node metastases (LNMs) and tumour deposits (TDs) are included in colorectal cancer (CRC) staging, although knowledge regarding their biological background is lacking. This study aimed to compare the biology of these prognostic features, which is essential for a better understanding of their role in CRC spread. Spatially resolved transcriptomic analysis using digital spatial profiling was performed on TDs and LNMs from 10 CRC patients using 1,388 RNA targets, for the tumour cells and tumour microenvironment. Shotgun proteomics identified 5,578 proteins in 12 different patients. Differences in RNA and protein expression were analysed, and spatial deconvolution was performed. Image-based consensus molecular subtype (imCMS) analysis was performed on all TDs and LNMs included in the study. Transcriptome and proteome profiles identified distinct clusters for TDs and LNMs in both the tumour and tumour microenvironment segment, with upregulation of matrix remodelling, cell adhesion/motility, and epithelial-mesenchymal transition (EMT) in TDs (all $p < 0.05$). Spatial deconvolution showed a significantly increased abundance of fibroblasts, macrophages, and regulatory T-cells ($p < 0.05$) in TDs. Consistent with a higher fibroblast and EMT component, imCMS classified 62% of TDs as poor prognosis subtype CMS4 compared to 36% of LNMs ($p < 0.05$). Compared to LNMs, TDs have a more invasive state involving a distinct tumour microenvironment and upregulation of EMT, which are reflected in a more frequent histological classification of TDs as CMS4. These results emphasise the heterogeneity of locoregional spread and the fact that TDs should merit more attention both in future research and during staging.

DOI: https://doi.org/10.1002/path.6196

Femke; Jamieson, Nigel B; Tauriello, Daniele Vf; Kölzer, Viktor H; Vermeulen, Michiel; Nagtegaal, Iris D (2023). Transcriptomics and proteomics reveal distinct biology for lymph node metastases and tumour deposits in colorectal cancer. Journal of Pathology, 261(4):401-412.
DOI: https://doi.org/10.1002/path.6196

Journal of Pathology
J Pathol December 2023; **261:** 401–412
Published online 4 October 2023 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/path.6196

ORIGINAL ARTICLE

Transcriptomics and proteomics reveal distinct biology for lymph node metastases and tumour deposits in colorectal cancer

Nelleke PM Brouwer[1]*, Loth Webbink[1], Tariq S Haddad[1], Natasja Rutgers[1], Shannon van Vliet[1], Colin S Wood[2,3], Pascal WTC Jansen[4], Maxime W Lafarge[5], imCMS Consortium†, Johannes HW de Wilt[6], Niek Hugen[7], Femke Simmer[1], Nigel B Jamieson[2,3], Daniele VF Tauriello[8], Viktor H Kölzer[5], Michiel Vermeulen[4,9] and Iris D Nagtegaal[1]

[1] *Department of Pathology, Radboud University Medical Centre, Nijmegen, The Netherlands*
[2] *Wolfson Wohl Cancer Research Centre, School of Cancer Sciences, University of Glasgow, UK*
[3] *Academic Unit of Surgery, Glasgow Royal Infirmary, University of Glasgow, UK*
[4] *Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Oncode Institute, Radboud University Nijmegen, The Netherlands*
[5] *Department of Pathology and Molecular Pathology, University and University Hospital of Zürich, Zürich, Switzerland*
[6] *Department of Surgery, Radboud University Medical Centre, Nijmegen, The Netherlands*
[7] *Department of Surgery, Netherlands Cancer Institute, Amsterdam, The Netherlands*
[8] *Department of Medical Biosciences, Research Institute for Medical Innovation, Radboud University Medical Centre, Nijmegen, The Netherlands*
[9] *The Netherlands Cancer Institute, Amsterdam, The Netherlands*

*Correspondence to: NPM Brouwer, Department of Pathology, Radboud University Medical Centre, Nijmegen, The Netherlands.
E-mail: nelleke.brouwer@radboudumc.nl

†The imCMS Consortium group members' information is included in the supplementary material, File S1.

Abstract

Both lymph node metastases (LNMs) and tumour deposits (TDs) are included in colorectal cancer (CRC) staging, although knowledge regarding their biological background is lacking. This study aimed to compare the biology of these prognostic features, which is essential for a better understanding of their role in CRC spread. Spatially resolved transcriptomic analysis using digital spatial profiling was performed on TDs and LNMs from 10 CRC patients using 1,388 RNA targets, for the tumour cells and tumour microenvironment. Shotgun proteomics identified 5,578 proteins in 12 different patients. Differences in RNA and protein expression were analysed, and spatial deconvolution was performed. Image-based consensus molecular subtype (imCMS) analysis was performed on all TDs and LNMs included in the study. Transcriptome and proteome profiles identified distinct clusters for TDs and LNMs in both the tumour and tumour microenvironment segment, with upregulation of matrix remodelling, cell adhesion/motility, and epithelial–mesenchymal transition (EMT) in TDs (all $p < 0.05$). Spatial deconvolution showed a significantly increased abundance of fibroblasts, macrophages, and regulatory T-cells ($p < 0.05$) in TDs. Consistent with a higher fibroblast and EMT component, imCMS classified 62% of TDs as poor prognosis subtype CMS4 compared to 36% of LNMs ($p < 0.05$). Compared to LNMs, TDs have a more invasive state involving a distinct tumour microenvironment and upregulation of EMT, which are reflected in a more frequent histological classification of TDs as CMS4. These results emphasise the heterogeneity of locoregional spread and the fact that TDs should merit more attention both in future research and during staging.

Keywords: colorectal cancer; lymph node metastases; tumour deposits; spatial transcriptomics; proteomics; algorithm; imCMS; tumour microenvironment; cell biology; locoregional spread

Received 8 June 2023; Revised 12 July 2023; Accepted 2 August 2023

No conflicts of interest were declared.

Introduction

Colorectal cancer (CRC) patients are staged according to the tumor node metastasis (TNM) system, which is based on the sequential progression hypothesis that metastases are the direct result of lymphatic spread. Therefore, there is a major focus on lymph node metastases (LNMs) [1]. However, recent studies have shown that other forms of locoregional spread also provide valuable prognostic information, with tumour deposits (TDs) having the greatest impact [2,3]. TDs

are aggregates of tumour cells in the fat surrounding the bowel, found in 20–25% of CRC patients [3]. In the current staging system, TDs are considered a less important form of locoregional spread, only deemed clinically relevant in the absence of LNMs [4]. However, recent studies have shown that TDs add relevant prognostic information, which might be a reflection of distinct biology [2,3].

In recent years, research has mainly focused on elucidating the biology of different aspects of primary CRC tumours, identifying extensive inter- and intra-tumour heterogeneity with prognostic implications [5]. Some studies have identified biological processes and molecular markers in the primary tumour bulk that have prognostic impact or are associated with different kinds of locoregional spread. Efforts to classify tumours according to molecular markers have yielded consensus molecular subtypes (CMSs), which have shown significant prognostic impact with poor survival rates for patients with the more mesenchymal CMS4 subtype [6]. In addition, genetic, transcriptomic, and proteomic profiles have also been established for LNMs and distant metastases [7–9]. Comparing the biological background of different types of locoregional spread, such as TDs and LNMs, would improve our understanding of cancer spread and might explain differences in prognostic impact.

Sample size and availability have limited phenotypic analysis of small locoregional nodules. Techniques such as single-cell transcriptomics require fresh tissue and lack information on spatial context. The distinction between TDs and LNMs is dependent on their spatial context, which is only assessable in processed tissue, which is now possible with the recent development of spatial gene expression [10,11]. Similarly, it has become possible to perform proteomics on formalin-fixed, paraffin-embedded (FFPE) tissue, providing an opportunity to validate transcriptome profiles and to determine the unique definite phenotype of TDs and LNMs in an untargeted approach [12].

Because it is essential to understand the biological heterogeneity of locoregional spread to further our understanding of the earliest steps towards distant metastases in CRC, this study aimed to conduct a comparative analysis of the biology of LNMs and TDs. By applying digital spatial profiling (DSP) combined with shotgun proteomics, we show that TDs display a more invasive phenotype, based on differences in gene expression and pathway activation. Furthermore, we validated these transcriptomic differences in protein profiles and correlated these molecular profiles with morphological features automatically detected in histology images by machine learning models. These results provide a biological and scientific basis to further investigate the role of locoregional spread in cancer progression as well as to optimise staging of CRC patients.

Materials and methods

Human samples

This study included resection material from 22 patients with colorectal adenocarcinoma with TDs and LNMs, obtained from the pathology archives of Radboud University Medical Centre. Clinicopathological details of the study cohort can be found in supplementary material, Table S1. The study was performed according to the Dutch 'Federa, Human Tissue and Medical Research: Code of Conduct for Responsible Use (2011)' regulations, not requiring patient informed consent. This study was approved by the Ethics Committee (CMO) of Radboud University Medical Centre (approval no. 2017-3603).

In situ hybridisation and DSP

Ten patients were included for DSP, and at least one TD and one LNM were included from every patient in the tissue microarrays used for analysis. Details on *in situ* hybridisation and DSP are available in Supplementary materials and methods.

DSP data processing and normalisation

Raw data were available in the GeoMx DSP Analysis Suite. The segmental biological probe quality control (QC) and area of illumination (AOI) QC was performed using the GeomxTools 3.1.1 package developed by NanoString in R version 4.2.0 [13]. The following settings were used: raw read threshold <1,000 reads, aligned reads <80%, sequencing saturation <50%, negative probe count geomean < three, minimum nuclei <100, and minimum surface area 1,000 μm^2. A total of two AOIs were flagged with low reads and five with low saturated reads. Probe QC was performed with default settings, excluding probes from target count calculation in all segments if (geomean probe in all segments)/(geomean probes within target) ≤ 0.1 and if Grubbs' outlier test failed in ≥20% of the segments. If a probe failed Grubbs' outlier test in a specific segment, the probe was excluded from target count calculation in that segment. The limit of quantitation was calculated using two SDs of the geomean of the negative probes.

To determine whether flagged AOIs needed to be removed from the study, background to on-target signal strength was compared. Quartile 3 count (Q3) normalisation was performed to account for technical effects between AOIs. Following QC, the resulting gene expression matrix consisted of 246 AOIs that passed QC and a total of 1,377 genes. All downstream analyses were performed in R version 4.2.0 or later.

Cell deconvolution analysis was performed using the SpatialDecon version 1.3 package (https://github.com/Nanostring-Biostats/SpatialDecon/). One-way analysis of variance (ANOVA) was applied to determine statistical significance of difference in cell populations between TDs and LNMs.

Proteomics sample preparation and shotgun mass spectrometry

Twelve patients were selected for proteomics, and one LNM and one TD were included from each patient. Details on proteomics sample preparation and shotgun

mass spectrometry (MS) are available in Supplementary materials and methods.

Proteomics data analysis

Protein identification and quantification were done in MaxQuant version 1.6.0.1 [14] with default settings, with match-between-runs, iBAQ, and label-free quantification enabled. The tandem MS spectra were searched against the human Uniprot database including reverse peptide sequences for false discovery rate (FDR) estimation downloaded in June 2017. Mass tolerance was set at 4.5 and 20 ppm for precursor ion and fragment ions, respectively. FDR was set at 0.01 for both the peptide and protein levels. A minimum of two ratio counts were required for protein quantification.

Common contaminants and decoy database hits were removed from the resulting MaxQuant protein groups file, and alias gene names were replaced with official gene symbols using the Limma package. If this resulted in duplicate entries, the entry with the highest number of the razor + unique peptides was retained. Protein groups were required to have a least two assigned peptides, of which at least one was a unique peptide. Differentially enriched protein analysis was performed using the DEP package. All protein groups that were detected in all replicates of at least one condition were considered for downstream analysis. Imputation of missing values was performed using the MinProb method with the default settings. This yielded a total of 5,578 unique proteins for downstream analysis in R version 4.2.0 or later.

Immunohistochemistry and semi-quantification

Details on immunohistochemical detection of selected proteins of interest are available in Supplementary materials and methods.

Image-based CMS analyses

Morphological features associated with the CMSs were automatically quantified by an imCMS deep learning ensemble model applied to whole slide images (WSIs) of H&E-stained tissue specimens of all the nodules included in the DSP and proteomics experiments. This ensemble model was trained to predict CMSs from input H&E-stained WSIs using a multicohort dataset of WSIs paired with transcriptional CMS calls, covering all cancer stages among resection and biopsy specimens [15]. Validation performance in holdout cohorts were in line with the performances previously reported [15], supporting the application of the trained model to this dataset.

Manually annotated tumour and tumour microenvironment (TME) regions were divided into overlapping tiles of size 318×318 px (at magnification $\times 5$) and processed to produce tile-level probability scores for each CMS class. In every annotated region, the scores of all contained tiles were averaged to produce slide-level imCMS scores, and imCMS calls were made by selecting the class with the highest slide-level score. To produce robust imCMS calls for all nodules, the imCMS call made by the majority of the five individual models that constituted the ensemble model was reported. Predictions were treated as mixed for cases with an absence of majority. Classes predicted at the tile level weighted by ensemble agreement were concatenated to produce visualisation maps and used to estimate the spatial heterogeneity of CMS-associated morphology.

Statistical analyses

Differential gene expression across groups was analysed using linear mixed-effects models, and the corresponding FDR was calculated by the Benjamini-Hochberg method. Differentially expressed genes (DEGs) were defined as $\log_2$ fold-change >0.5 and FDR ≤0.05. For the tumour and TME segment of the DSP experiment as well as the proteomics experiment, heatmaps with unsupervised hierarchal clustering were generated for the top 100 DEGs. The top 100 gene list was curated by selecting the 100 most DEGs from a list of only statistically significant DEGs (FDR < 0.05). For the proteomics cohort, all proteins that overlapped between the genes in the epithelial–mesenchymal transition (EMT) hallmark gene set from Gene Set Enrichment Analysis (GSEA MsigDB M5930) and the proteomics data were included in an additional heatmap, yielding a total of 105 genes. The statistical significance of the difference in clusters yielded by unsupervised hierarchal clustering was analysed using the chi-squared (χ^2) test.

GSEA was performed on log2-transformed, normalised expression matrices of both the DSP and proteomics cohort. Pathways were curated by Nanostring and tailored to the targets in the GeoMx CTA. The Wilcoxon rank-sum test was used to analyse the statistical significance of the difference in expression of specific genes between TDs and LNMs over all patients in the DSP and proteomics experiment, as well as to analyse differences in immunohistochemistry (IHC) scores.

Results

In total, 22 CRC cases with at least one TD and one LNM were used for this study. For DSP, 10 cases were used, and from every case at least one TD and one LNM were included. For the proteomic analysis, one TD and one LNM per case were included for 12 cases. No significant differences were found between the cohorts (supplementary material, Table S1). Figure 1 shows the distinct appearance of TDs and LNMs on histopathology. TDs (Figure 1A,B) are clusters of tumour cells located in the fat, in the vicinity of vessels and nerves, without clear remnants of a lymph node, vascular invasion, or perineural invasion, whereas LNMs (Figure 1C,D) show a clear lymph node structure with tumour cells.

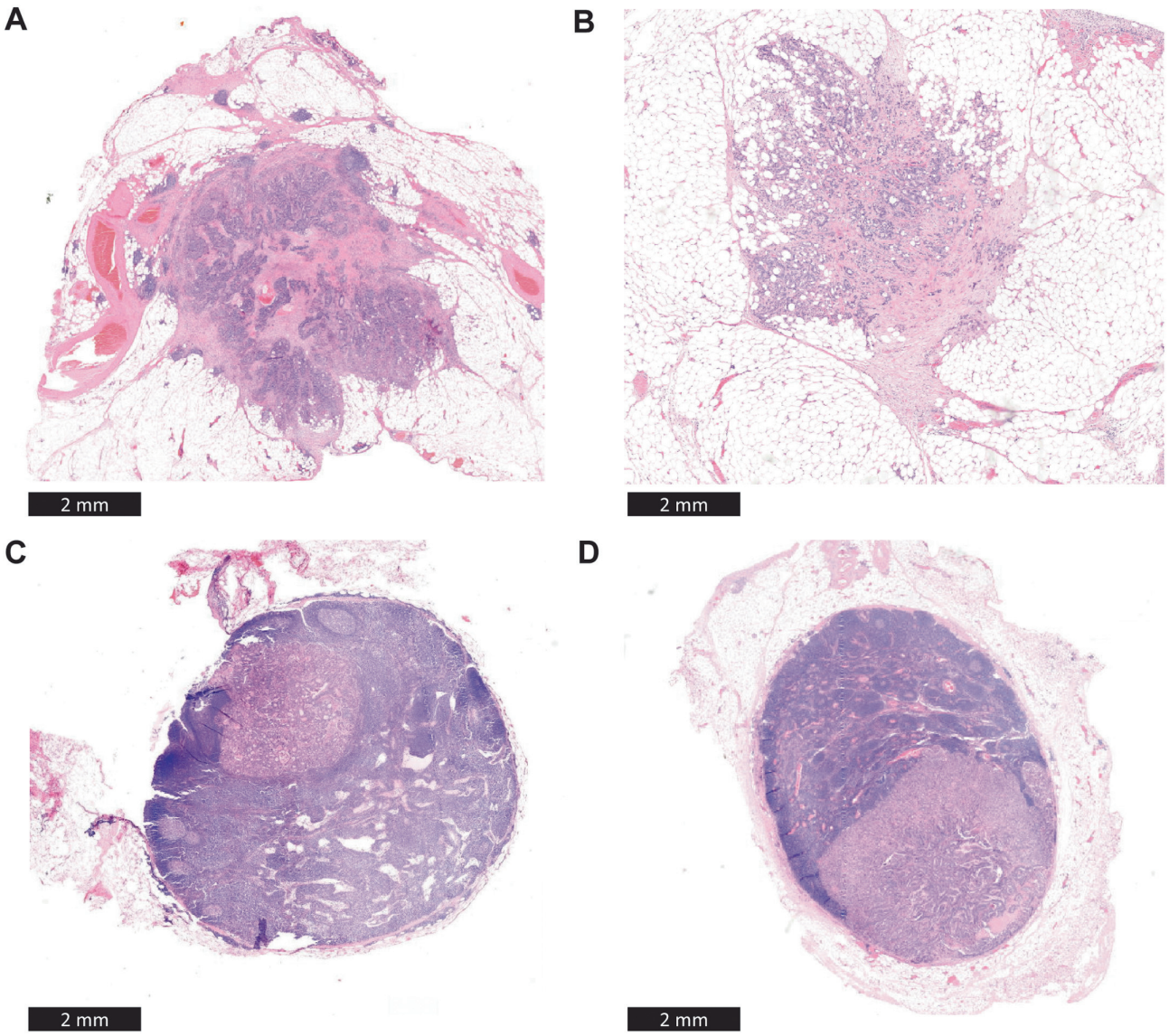

Figure 1. Examples of TD and LNM as seen on histopathology. (A and B) Example of tumour deposit in H&E. (C and D) Example of lymph node metastasis in H&E.

TD and LNM display a distinct transcriptomic signature

To investigate whether the histological differences seen between TDs and LNMs represent a distinct biology, we compared the transcriptomic profiles of TDs and LNMs from 10 CRC cases using DSP. To specifically examine the transcriptional alterations within the tumour cells and TME, we analysed cytokeratin+ (tumour) and cytokeratin− (TME) segments (i.e. the AOIs) separately within the regions of interest (ROIs) (supplementary material, Figure S1A). The data showed an expected differential expression of genes between the tumour and TME segments (supplementary material, Figure S1B,C), confirming the quality of the methodology.

Unsupervised hierarchal clustering of the top 100 DEGs in the tumour segments of TD and LNM identified three distinct clusters. Rather than clustering by patient origin, samples largely segregated by nodule type (supplementary material, Figure S3). Clusters 1 and 2 represent mostly LNMs, whereas cluster 3 is strongly enriched for TDs (χ^2 test: $p < 0.001$; Figure 2A). In the TME, the distinction between TDs and LNMs was even more pronounced, with two clusters each dominating with either structure ($p < 0.001$) (Figure 2A).

The overexpressed genes in LNMs are associated mostly with immune cells and inflammatory signalling, illustrated by *CD22* and *CCL21* (tumour segment) and *TRAF1* and *MS4A1* and (TME segment) (Figure 2B). In TDs, overexpressed genes are mostly involved in migration and invasion as well as matrix remodelling (e.g. *FOS* and *CEACAM6* in the tumour segment, and *SFRP2*, *COL11A1*, *FN1*, and *COMP* in the TME segment). GSEA confirmed these differences in biological processes between TDs and LNMs, with more matrix remodelling, cell motility, and *MET* and *PDGF* signalling in TDs, whereas LNMs showed upregulated B- and T-cell pathways (FDR < 0.001) (Figure 2C). The full list of GSEA on DSP data is provided in supplementary material, Table S3A,B.

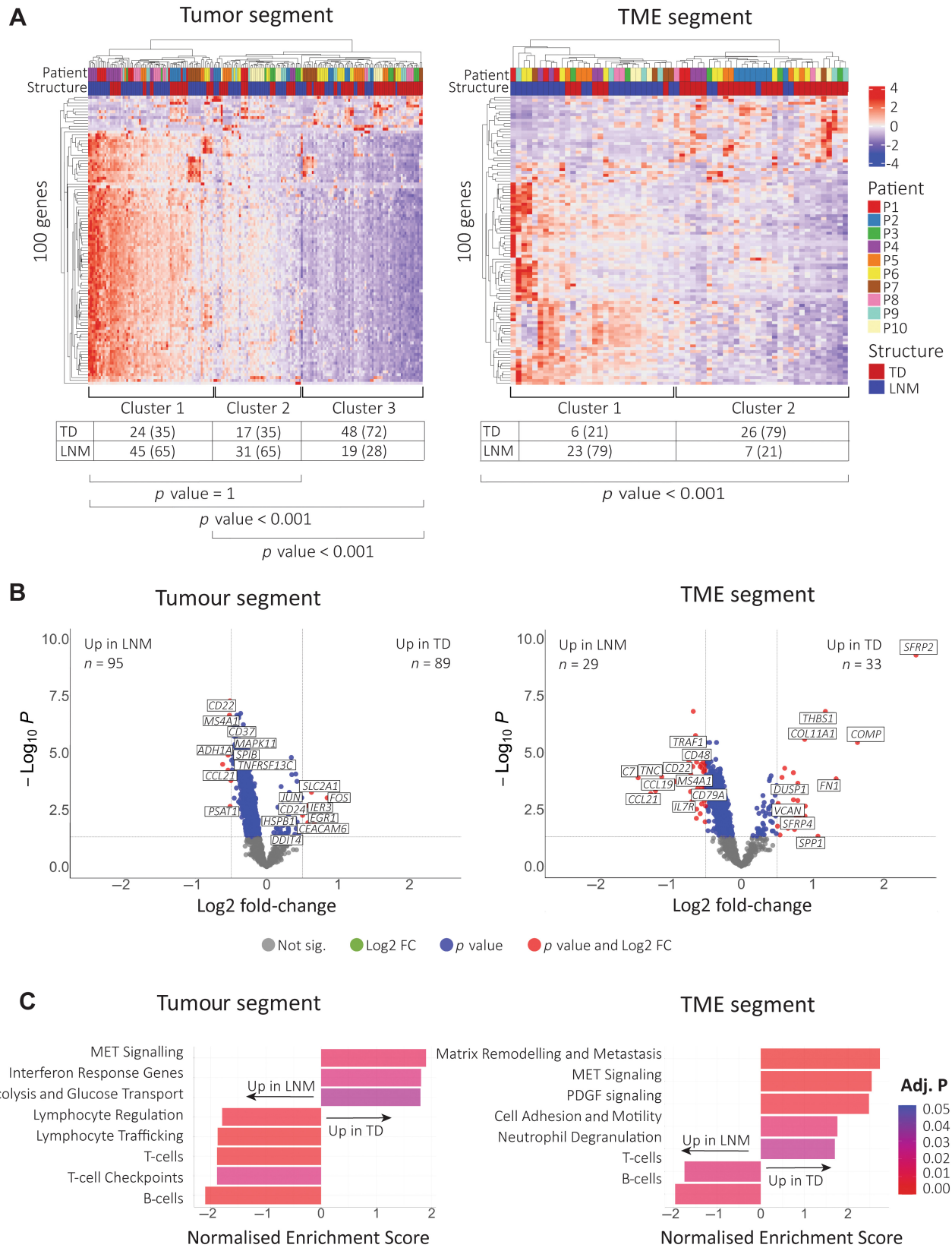

Figure 2. Differences in transcriptome between TD and LNM. (A) Heatmap of top 100 DEGs between TD and LNM with unsupervised hierarchal clustering, for tumour segment and TME segment separately. (B) Volcano plots showing DEGs between TD and LNM (unpaired *t*-test) for tumour segment and TME segment separately. FDR is calculated using Benjamini–Hochberg method. Complete gene expression data can be found in supplementary material, Table S2A,B. (C) Bar graphs showing selected significant pathways calculated using gene sets curated by Nanostring ($p < 0.001$) with genes upregulated in either TD or LNM, for the tumour and stromal segment separately. Complete pathway analysis can be found in supplementary material, Table S3A,B.

Proteomic analysis confirms a more invasive phenotype of TDs

To investigate whether differences in gene expression translated into differences in protein abundance, we performed an untargeted proteomic analysis using shotgun MS on tissue from 12 patients. The protein profiles, consisting of 5,578 unique proteins, represented the combination of tumour cells and TME and were in line with the results from DSP. The top 100 heatmap identified three clusters, with two predominantly containing either TDs or LNMs and the third a mixture (Figure 3A). The biological processes in the proteomics data are also similar to those identified in the DSP analysis (Figure 3B,C), with these biological processes of matrix remodelling and invasion pointing towards the process of EMT. Some proteins in the pathways upregulated in TDs (collagens, MMPs, *FAP*, *LAMA3*, *LAMC2*, *SERPINH1* and *SPP1*) are also part of the curated GSEA-MSigDB EMT hallmark gene set. Analysing the EMT hallmark gene set in the proteomics data (Figure 3D) showed upregulation of EMT-related proteins in the TD-rich cluster compared to the cluster composed mainly of LNMs ($p < 0.001$), highlighting the association of TDs with this process.

As a final validation of the DSP and proteomics data, we used IHC for two biologically interesting gene products. *SFRP2* is one of the most highly overexpressed genes in both the DSP and proteomic analyses, and MXRA5 was found in our proteomics dataset (Figure 4A). In agreement with reports that *SFRP2* is expressed in fibroblasts, we observed stromal staining in patient samples, and MXRA5 – its gene included in the EMT geneset – was present in tumour cells (Figure 4B). IHC quantification confirmed upregulated protein abundance for both targets in TDs compared to LNMs ($p < 0.001$, Figure 4C).

Distinct composition of TME in TDs and LNMs

To estimate the relative abundance of cell types in the TME, we next performed spatial deconvolution of the DSP transcriptome data. Multiple cell types showed a significantly different abundance between TDs and LNMs. The former consisted of more fibroblasts, macrophages, and regulatory T-cells compared to the latter (Figure 5A). As expected, several immune cell types, including CD4+ T cells, are more abundant in LMNs, even when taking into account the selection of ROIs that preferred tumour reactive stroma, rather than predominantly lymphoid areas. These findings regarding differences in cell composition, in combination with the DSP and proteomics data, suggest a phenotype in TDs that promotes CRC progression and immune evasion [16].

When results from the DSP and proteomics are combined, a model can be constructed for the cellular and functional differences between TDs and LNMs (Figure 5B). Compared to LNMs, TDs generally show more upregulation of EMT, migration and invasion of tumour cells, and a potentially more immunosuppressive TME with a higher abundance of fibroblasts, macrophages, and regulatory T-cells.

TDs represent mainly the CMS4 phenotype

Some of the markers upregulated in the TDs and LNMs, such as *SFRP2*, are part of the CMS classification of CRC [6]. To translate the molecular findings from DSP and proteomics to the histological level, H&E WSIs of all the nodules included in the study were classified by imCMS, and the spatial heterogeneity of tile-level CMS-associated visual features was estimated for each nodule (Figure 6A). Both LNMs and TDs showed heterogeneous phenotypes, with TDs showing a higher proportion of CMS4-associated morphology in the nodules (Figure 6B). When comparing the nodule-level imCMS classifications, TDs showed a higher proportion of imCMS4 nodules (36.4% of LNM versus 61.8% of TD, $p < 0.05$), and LNMs showed a higher proportion of imCMS2 nodules (30.3% of LNM versus 8.8% of TD, $p < 0.05$).

Discussion

In this study, we combined transcriptomic profiling and shotgun proteomics to describe the molecular biology of TDs and LNMs in CRC. Furthermore, our molecular findings were translated to the morphological level using an image analysis algorithm in the form of the imCMS classifier. This multi-omic combination of techniques provides unique insights into the biology of locoregional spread in CRC, which more traditional strategies could not have provided.

DSP transcriptome analyses showed clear differences in gene expression between TDs and LNMs, with TDs exhibiting signs of increased cell motility, matrix remodelling, and EMT, as well as a distinct composition of the TME. In contrast, LNMs showed a more immune-cell-enriched TME with more immunogenic signalling and interaction between immune cells and tumour cells. These results were confirmed by untargeted proteomics in an independent patient cohort.

Based on both the transcriptome and proteome data, TDs and LNMs clustered differently, showing that their distinct biology was stronger than interpatient variation (supplementary material, Figure S3). Interestingly, both the tumour transcriptome data and the proteome data (Figures 2A and 3A) showed three clusters based on expression profiles, suggesting two clusters with opposing phenotypes for either LMNs or TDs, and an intermediate cluster. This intermediate cluster represents cases with a distinct histology: they show either lymphovascular invasion or signs of extranodal extension (supplementary material, Figure S2). This histology suggests a more invasive phenotype of cluster 2 compared to the LNM cluster, as well as a different phenotype from the TD cluster. Further research is needed to determine whether this intermediate cluster

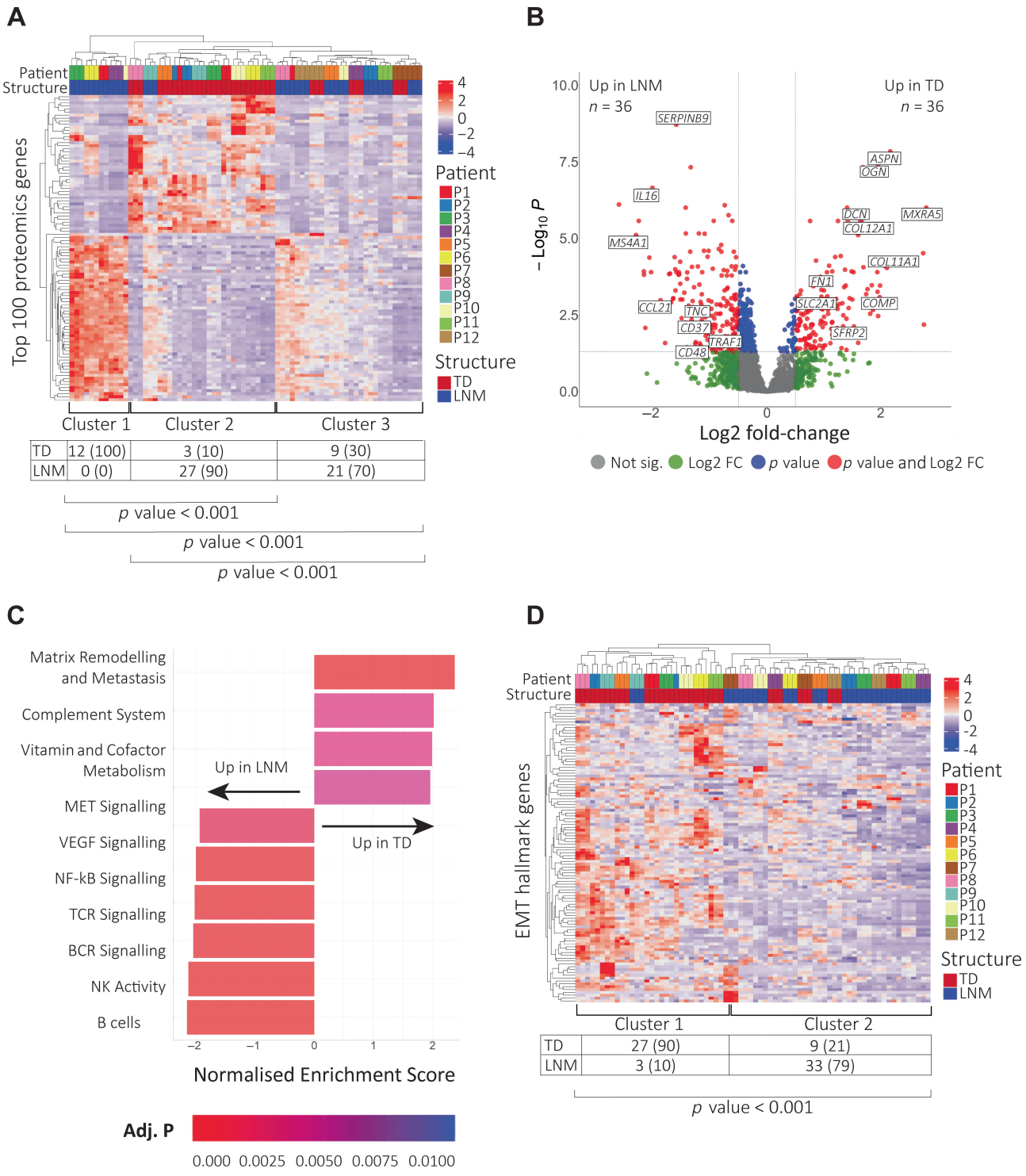

Figure 3. Differences in proteome between TD and LNM, which includes both tumour and TME segment. (A) Heatmap of top 100 differentially expressed proteins between TD and LNM, with unsupervised hierarchal clustering resulting in three clusters. (B) Volcano plot showing differentially expressed proteins between TD and LNM (unpaired *t*-test). FDR is calculated using Benjamini–Hochberg method. Complete protein expression data can be found in supplementary material, Table S2C. (C) Bar graphs showing enrichment for selected significant pathways curated by Nanostring ($p < 0.001$) with proteins upregulated in either TD or LNM. Complete pathway analysis can be found in supplementary material, Table S3C. (D) Heatmap for all EMT hallmark proteins found in proteomics dataset (GSEA M5930), with unsupervised hierarchal clustering resulting in two distinct clusters.

could represent a transition phase from LNMs to TDs, which may help us understand the origin of TDs.

In line with the more invasive phenotype of TDs, several of the DEGs overexpressed in TDs were found to be pro-tumourigenic in previous studies. TDs showed an upregulation of both *FOS* and *JUN*, a transcriptional complex associated with cancer progression and EMT in CRC [17,18]. MXRA5 and CEACAM6 function as a matrix remodeller and a cell adhesion molecule, respectively, and are associated with poor prognosis [19,20].

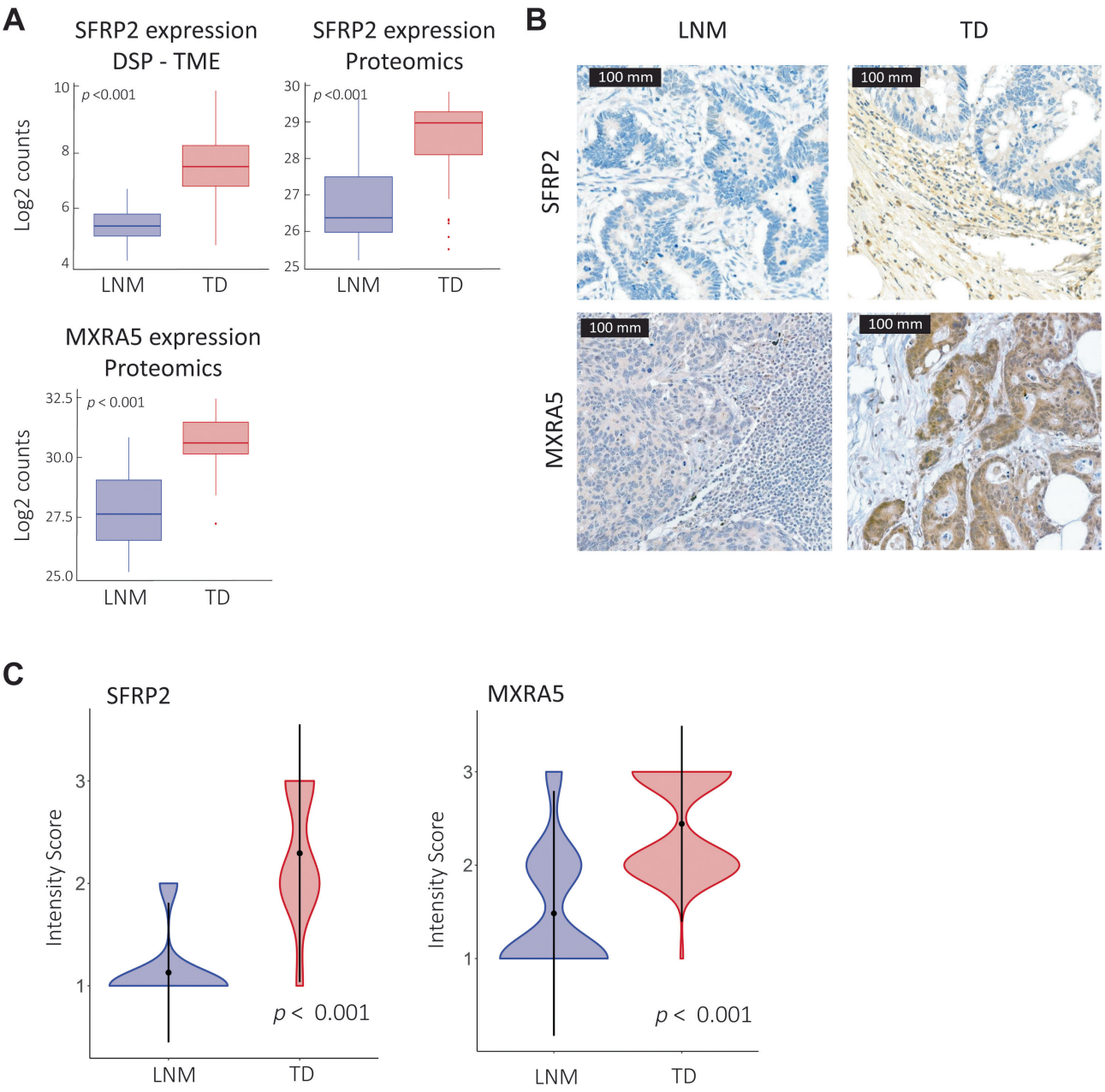

Figure 4. Immunohistochemical validation of differentially expressed genes/proteins identified by DSP and proteomics. (A) Boxplots of log2-transformed, normalised levels of *SFRP2* and MXRA5 detected by DSP and proteomics. Statistical significance was tested by Wilcoxon rank-sum test. (B) Protein abundance of SFRP2 (TME segment) and MXRA5 (tumour segment) in TD and LNM samples that were profiled using DSP. (C) Violin plot showing proportion of samples for each scored IHC intensity for SFRP2 and MXRA5. Statistical significance was tested by Wilcoxon rank-sum test.

TDs also show activation of the c-MET/HGF and PDGF pathways, which promote tumour invasion, metastasis, and angiogenesis [21,22]. In contrast, the tumour cells in LNMs mostly expressed genes involved in interaction with the immune environment, instead of genes indicating signs of invasion and migration. For example, LNMs showed higher levels of *CCR7*, the receptor for *CCL21*. Normally, the CCL21-CCR7 axis is important in lymphocyte homing to lymph nodes but can be hijacked by cancer cells to facilitate their migration to lymph nodes, just like the CXCL21-CXCR4 axis [23]. *PSAT1*, which plays a role in the recruitment and modulation of tumour immune infiltrating cells [24], was also upregulated in LNMs. Additionally, stress effects and influences from the immunogenic environment were found, such as overexpression of *MAPK11*, which is mostly activated in tumour cells as a response to inflammatory cytokines [25]. Activation of the NF-κB pathway points to influences of the immunogenic environment on tumour cells as it is primarily initiated by pro-inflammatory signals and induces proliferation and anti-apoptosis in CRC [26].

In addition to the distinct tumour cell characteristics in TDs and LNMs, we observed substantial differences in DEGs and the cellular composition of the TME. Spatial deconvolution showed a higher abundance of

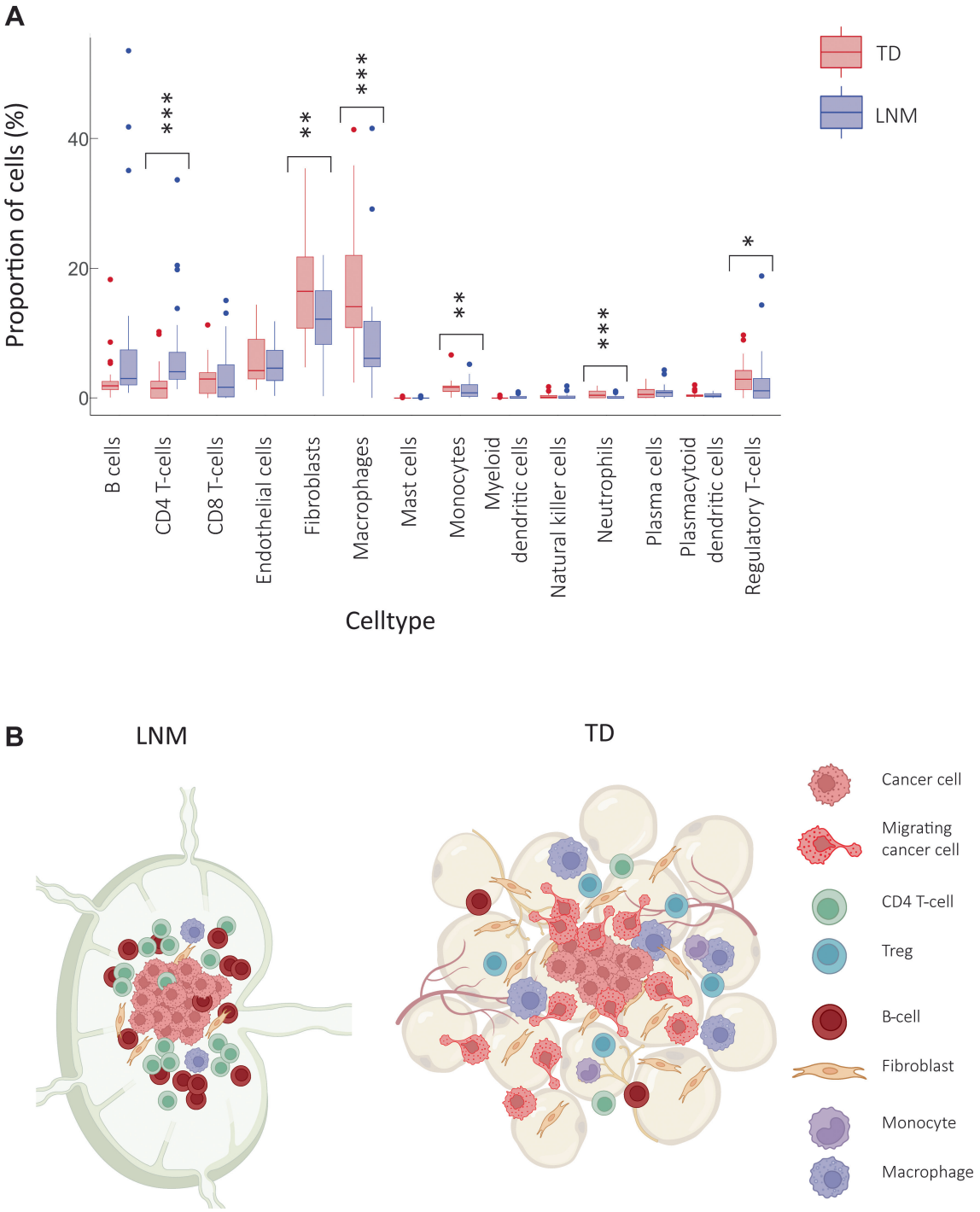

Figure 5. Translation of DSP to differences in cell populations. (A) Spatial deconvolution of cell populations within tumour microenvironment of TD and LNM. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. (Wilcoxon test). (B) Schematic of biological differences between TD and LNM identified using DSP and proteomics.

regulatory T-cells, fibroblasts, and macrophages in the TME of TDs. The results of this method of spatial deconvolution were validated previously by our group using multiplex fluorescent imaging [27]. A cancer-associated fibroblast population with high *SFRP2*, as observed in the TDs in our study, was previously described as being associated with the migration and invasion of colon cancer cells [28]. As an explanation, SFRP2 is thought to cooperate with WNT16B in preventing cell death and increasing proliferation, migration, and invasion [29]. TDs also appear to have a denser extracellular matrix as a result of the increased deposition of a variety of collagens (e.g. COL11A1, COL1A1/2, COL3A1, COL5A1) and fibronectin, which help tumour cells migrate and show poor prognostic impact in CRC [30–33]. The combination of this denser and, thus, stiffer matrix and proteins, such as THBS1, an adhesive glycoprotein that guides cell–cell and cell–matrix interactions, may facilitate more tumour cell migration inside TDs than LNMs [34,35]. In addition, the TME in TDs also shows signs of immunosuppression, with more regulatory

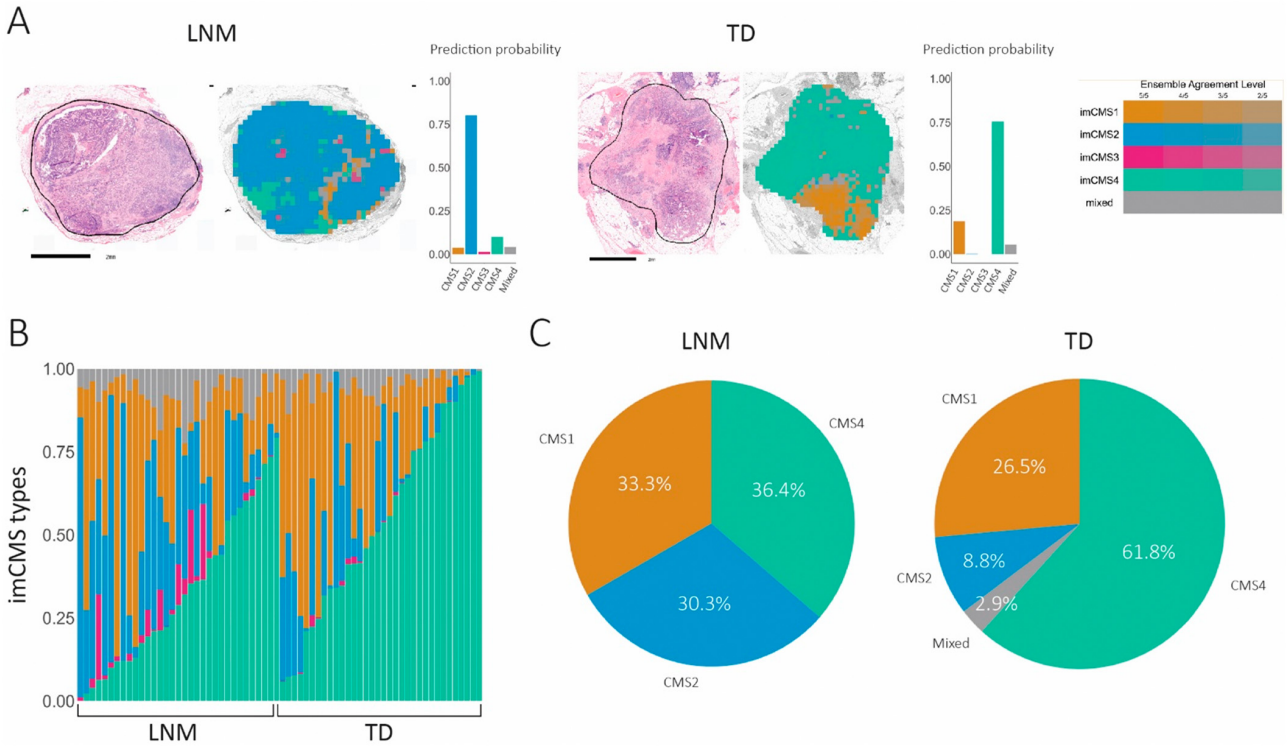

Figure 6. imCMS classification of all TD and LNM included in proteomics and DSP analyses. (A) Example image of overlay showing imCMS class predicted by model for each image tile extracted from annotated tumour regions in either TD or LNM. The model enables estimation of the heterogeneity of CMS-associated morphology in a nodule as the proportion of tiles it contains that are associated to each imCMS class. (B) Overview of heterogeneity of imCMS classes in TD and LNM from proteomics and DSP. (C) Pie chart showing proportions of different imCMS classes in TD and LNM from proteomics and DSP cohort ($p < 0.05$ Fisher's exact test).

T-cells and higher *SPP1*, which polarises cancer-associated macrophages and promotes immune escape and host tumour immune tolerance by suppressing T-cell activation [36–38].

The transcriptomic and proteomic results in this study suggest a more invasive, mesenchymal, and fibrotic phenotype for TDs. An important example where molecular profiles have been matched with histological appearance are the CMS subtypes in CRC, which have a large clinical impact as molecular subgroups with distinct growth patterns and prognostic impact [6]. CMS4 is generally viewed as the mesenchymal subtype with the most aggressive biology and the worst prognosis [6]. The imCMS classifier, which was previously developed and validated on large independent cohorts, uses histology images to identify CMS-associated morphological visual features without the need for molecular data, which enabled us to translate our multi-omic findings to the morphological level in a blinded manner [15]. Indeed, TDs were significantly more often classified as imCMS4, reflecting their more aggressive biology on a morphological level. Our findings on the more invasive biology of TDs contradict the current use of TD status in the TNM classification system for CRC [4]. The status of TDs has been widely debated on the basis of its stronger prognostic impact [39–41]. The invasive phenotype found in our study provides a biological explanation for the poor prognostic impact and complements the evidence that is needed to give TD status a more prominent role in TNM staging.

This study provides a framework for further research on different types of locoregional spread in CRC. Despite the distinctive biological profiles of TDs and LNMs, we did not identify unique markers for TDs or LNMs, and there is overlap in DEGs between tumour segments and TME. This may suggest that the segmentation and lasering of the different segments were not perfect. However, as the different segments consisted of several hundred cells, and the imperfections in segmentation will only be present at segment borders, the impact of this uncertainty will be minimal. Finally, our biological results are in line with the prognostic impact of TDs but do not provide the biological evidence that either TDs or LNMs are essential for distant metastatic spread.

To conclude, TDs are an aggressive form of locoregional spread and are currently underestimated in cancer staging. TDs are characterised by more matrix remodelling, cell motility, and EMT, which explains the poor prognosis of patients with TDs. Appreciation of the heterogeneity and biology of locoregional spread will optimise risk stratification for CRC patients and ultimately improve clinical management and prognosis.

Acknowledgements

The authors would like to acknowledge our colleagues from the Institute of Tissue Medicine and Pathology at

the University of Bern for their assistance with the SFRP2 IHC stain. The Vermeulen Lab is part of the Oncode Institute, which is partly funded by the Dutch Cancer Society (KWF).

Author contributions statement

NPMB, FS, IDN, MWL, VHK and MV conceived the experiments. NPMB, LW, TSH, NR, SV, PWTCJ and MWL carried out the experiments. NPMB, LW, CSW, PWTCJ and NBJ analysed the data. All authors were involved in writing the paper and had final approval of the submitted and published versions.

Data availability statement

The GeoMx data are available via Zenodo at https://doi.org/10.5281/zenodo.7835226. The mass spectrometry proteomics data were deposited at the ProteomeXchange Consortium via the PRIDE partner repository under the dataset identifier PXD041625.

References

1. Dukes CE. The classification of cancer of the rectum. *J Pathol Bacteriol* 1932; **35:** 323–332.
2. Lord AC, D'Souza N, Pucher PH, *et al*. Significance of extranodal tumour deposits in colorectal cancer: a systematic review and meta-analysis. *Eur J Cancer* 2017; **82:** 92–102.
3. Nagtegaal ID, Knijn N, Hugen N, *et al*. Tumor deposits in colorectal cancer: improving the value of modern staging-a systematic review and meta-analysis. *J Clin Oncol* 2017; **35:** 1119–1127.
4. International Union Against Cancer. In *TNM Classification of Malignant Tumours* (8th edn), Brierley JD, Gospodarowicz MK, Wittekind CH (eds). Wiley-Blackwell: Hoboken, 2017.
5. Koncina E, Haan S, Rauh S, *et al*. Prognostic and predictive molecular biomarkers for colorectal cancer: updates and challenges. *Cancers (Basel)* 2020; **12:** 319.
6. Guinney J, Dienstmann R, Wang X, *et al*. The consensus molecular subtypes of colorectal cancer. *Nat Med* 2015; **21:** 1350–1356.
7. Liu J, Cho YB, Hong HK, *et al*. Molecular dissection of CRC primary tumors and their matched liver metastases reveals critical role of immune microenvironment, EMT and angiogenesis in cancer metastasis. *Sci Rep* 2020; **10:** 10725.
8. Milioli HH, Santos Sousa K, Kaviski R, *et al*. Comparative proteomics of primary breast carcinomas and lymph node metastases outlining markers of tumor invasion. *Cancer Genomics Proteomics* 2015; **12:** 89–101.
9. Wong GYM, Diakos C, Hugh TJ, *et al*. Proteomic profiling and biomarker discovery in colorectal liver metastases. *Int J Mol Sci* 2022; **23:** 6091.
10. Merritt CR, Ong GT, Church SE, *et al*. Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nat Biotechnol* 2020; **38:** 586–599.
11. Roelands J, van der Ploeg M, Ijsselsteijn ME, *et al*. Transcriptomic and immunophenotypic profiling reveals molecular and immunological hallmarks of colorectal cancer tumourigenesis. *Gut* 2022; **72:** 1326–1339.
12. Coscia F, Doll S, Bech JM, *et al*. A streamlined mass spectrometry-based proteomics workflow for large-scale FFPE tissue analysis. *J Pathol* 2020; **251:** 100–112.
13. Reeves JDP, Ortogero N, Griswold M, *et al*. GeoMxWorkflows: GeoMx digital spatial profiler (DSP) data analysis workflows, R package version 1.4.0. 2023. [Accessed 28 July 2023]. Available from: https://github.com/Nanostring-Biostats/GeoMxWorkflows.
14. Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* 2016; **11:** 2301–2319.
15. Sirinukunwattana K, Domingo E, Richman SD, *et al*. Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning. *Gut* 2021; **70:** 544–554.
16. Tauriello DVF, Batlle E. Targeting the microenvironment in advanced colorectal cancer. *Trends Cancer* 2016; **2:** 495–504.
17. Huang X, Han C, Zhong J, *et al*. Low expression of the dynamic network markers FOS/JUN in pre-deteriorated epithelial cells is associated with the progression of colorectal adenoma to carcinoma. *J Transl Med* 2023; **21:** 45.
18. Qu X, Yan X, Kong C, *et al*. c-Myb promotes growth and metastasis of colorectal cancer through c-fos-induced epithelial-mesenchymal transition. *Cancer Sci* 2019; **110:** 3183–3196.
19. Wang GH, Yao L, Xu HW, *et al*. Identification of MXRA5 as a novel biomarker in colorectal cancer. *Oncol Lett* 2013; **5:** 544–548.
20. Yang T, Wang H, Li M, *et al*. CD151 promotes colorectal cancer progression by a crosstalk involving CEACAM6, LGR5 and Wnt signaling via TGFβ1. *Int J Biol Sci* 2021; **17:** 848–860.
21. Chen J, Yuan W, Wu L, *et al*. PDGF-D promotes cell growth, aggressiveness, angiogenesis and EMT transformation of colorectal cancer by activation of Notch1/Twist1 pathway. *Oncotarget* 2017; **8:** 9961–9973.
22. Fu J, Su X, Li Z, *et al*. HGF/c-MET pathway in cancer: from molecular characterization to clinical evidence. *Oncogene* 2021; **40:** 4625–4651.
23. Müller A, Homey B, Soto H, *et al*. Involvement of chemokine receptors in breast cancer metastasis. *Nature* 2001; **410:** 50–56.
24. Feng M, Cui H, Tu W, *et al*. An integrated pan-cancer analysis of PSAT1: a potential biomarker for survival and immunotherapy. *Front Genet* 2022; **13:** 975381.
25. Johnson GL, Lapadat R. Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases. *Science* 2002; **298:** 1911–1912.
26. Taniguchi K, Karin M. NF-κB, inflammation, immunity and cancer: coming of age. *Nat Rev Immunol* 2018; **18:** 309–324.
27. Haddad TS, van den Dobbelsteen L, Öztürk SK, *et al*. Pseudobudding: ruptured glands do not represent true tumor buds. *J Pathol* 2023; **261:** 19–27.
28. Kasashima H, Duran A, Martinez-Ordoñez A, *et al*. Stromal SOX2 upregulation promotes tumorigenesis through the generation of a SFRP1/2-expressing cancer-associated fibroblast population. *Dev Cell* 2021; **56:** 95–110.e10.
29. Sun Y, Zhu D, Chen F, *et al*. SFRP2 augments WNT16B signaling to promote therapeutic resistance in the damaged tumor microenvironment. *Oncogene* 2016; **35:** 4321–4334.
30. Galván JA, García-Martínez J, Vázquez-Villa F, *et al*. Validation of COL11A1/procollagen 11A1 expression in TGF-β1-activated immortalised human mesenchymal cells and in stromal cells of human colon adenocarcinoma. *BMC Cancer* 2014; **14:** 867.
31. Tabuso M, Adya R, Stark R, *et al*. Fibrotic phenotype of peritumour mesenteric adipose tissue in human colon cancer: a potential Hallmark of metastatic properties. *Int J Mol Sci* 2021; **22:** 2430.
32. Wu Y, Xu Y. Integrated bioinformatics analysis of expression and gene regulation network of COL12A1 in colorectal cancer. *Cancer Med* 2020; **9:** 4743–4755.
33. Cai X, Liu C, Zhang TN, *et al*. Down-regulation of FN1 inhibits colorectal carcinogenesis by suppressing proliferation, migration, and invasion. *J Cell Biochem* 2018; **119:** 4717–4728.
34. Liu C, Pei H, Tan F. Matrix stiffness and colorectal cancer. *Onco Targets Ther* 2020; **13:** 2747–2755.

35. Liu X, Xu D, Liu Z, *et al*. THBS1 facilitates colorectal liver metastasis through enhancing epithelial-mesenchymal transition. *Clin Transl Oncol* 2020; **22:** 1730–1740.
36. Liu Y, Ye G, Dong B, *et al*. A pan-cancer analysis of the oncogenic role of secreted phosphoprotein 1 (SPP1) in human cancers. *Ann Transl Med* 2022; **10:** 279.
37. Olguín JE, Medina-Andrade I, Rodríguez T, *et al*. Relevance of regulatory T cells during colorectal cancer development. *Cancers (Basel)* 2020; **12:** 1888.
38. Qi J, Sun H, Zhang Y, *et al*. Single-cell and spatial analysis reveal interaction of FAP^+ fibroblasts and $SPP1^+$ macrophages in colorectal cancer. *Nat Commun* 2022; **13:** 1742.
39. Cohen R, Shi Q, Meyers J, *et al*. Combining tumor deposits with the number of lymph node metastases to improve the prognostic accuracy in stage III colon cancer: a post hoc analysis of the CALGB/SWOG 80702 phase III study (Alliance). *Ann Oncol* 2021; **32:** 1267–1275.
40. Lino-Silva LS, Anchondo-Núñez P, Chit-Huerta A, *et al*. Stage I-III colon cancer patients with tumor deposits behave similarly to stage IV patients. Cross-section analysis of 392 patients. *J Surg Oncol* 2019; **120:** 300–307.
41. Wong-Chong N, Motl J, Hwang G, *et al*. Impact of tumor deposits on oncologic outcomes in stage III colon cancer. *Dis Colon Rectum* 2018; **61:** 1043–1052.

SUPPLEMENTARY MATERIAL ONLINE

Supplementary materials and methods

Figure S1. Illustration of workflow for DSP of TD and LNM

Figure S2. Examples of lymphovascular invasion and extranodal extension as seen on histopathology of LNM in phenotypically mixed group

Figure S3. Alluvial plot showing how tumour segment samples from TDs and LNMs included in DSP heatmap analysis are distributed over the three distinct clusters in Figure 2A.

Table S1. Clinicopathological characteristics of patients in proteomics cohort and DSP cohort

Table S2. Differentially expressed genes and between tumour segment of TD and LNM, between stromal TME of TD and LNM, and proteins between TD and LNM

Table S3. Pathway analysis for tumour segment and TME segment from DSP and proteomics

File S1. imCMS Consortium group members