

MISS HELEN JONES (Orcid ID : 0000-0001-8238-7779)

DR VIKTOR HENDRIK KOELZER (Orcid ID : 0000-0001-9206-4885)

Article type : Original Article

# Stromal composition predicts recurrence of early rectal cancer after local excision

Short running title: Stromal composition predicts cancer recurrence

Helen JS Jones<sup>1</sup>, Chris Cunningham<sup>1</sup>, Hanne A Askautrud<sup>2</sup>, Håvard E Danielsen<sup>2,3,4</sup>, David J Kerr<sup>4</sup>, Enric Domingo<sup>5</sup>, Tim Maughan<sup>5</sup>, Simon J Leedham<sup>6#</sup>, Viktor H Koelzer<sup>7,8#</sup>

# co-senior

<sup>1</sup> Department of Colorectal Surgery, Oxford University Hospitals NHS Trust, Oxford

<sup>2</sup> Institute for Cancer Genetics and Informatics, Oslo University Hospital, Oslo

<sup>3</sup> Department of Informatics, University of Oslo, Oslo

<sup>4</sup> Nuffield Division of Clinical Laboratory Sciences, University of Oxford, Oxford, UK

<sup>5</sup> MRC Oxford Institute for Radiation Oncology, Department of Oncology, University of Oxford, Oxford, UK

<sup>6</sup> Intestinal Stem Cell Biology Laboratory, Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK

<sup>7</sup> Department of Pathology and Molecular Pathology, University and University Hospital Zürich, Zürich, Switzerland

<sup>8</sup> Department of Oncology and Nuffield Department of Medicine, University of Oxford, Oxford, UK

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/HIS.14438](https://doi.org/10.1111/HIS.14438)

This article is protected by copyright. All rights reserved

Accepted Article

Corresponding author: Helen Jones, Department of Colorectal Surgery, Oxford University Hospitals NHS Foundation Trust, Churchill Hospital, Oxford, tel: 01865235657, helen.jones3@ouh.nhs.uk , ORCID ID 0000-0001-8238-7779

There are no conflicts of interest.

Word count: 2,500

# Abstract

## Aims

After local excision of early rectal cancer, definitive lymph node status is not available. An alternative means for accurate assessment of recurrence risk is required to determine the most appropriate subsequent management. Currently used measures are suboptimal. We assess three measures of tumour stromal content to determine their predictive value after local excision in a well-characterized cohort of rectal cancer patients without prior radiotherapy.

## Methods and results

143 patients were included. H&E sections were scanned for 1) deep neural network (DNN, a machine learning algorithm) tumour segmentation into compartments including desmoplastic stroma and inflamed stroma; and 2) digital assessment of tumour stromal fraction (TSR) and optical DNA ploidy analysis. 3' mRNA sequencing was performed to obtain gene expression data, from which stromal and immune scores were calculated using the ESTIMATE method.

Full results were available for 139 samples and compared with disease-free survival. All three methods were prognostic. Most strongly predictive was a DNN-determined ratio of desmoplastic to inflamed stroma  $>5.41$  ( $p < 0.0001$ ). A ratio of ESTIMATE stromal to immune score  $<1.19$  was also predictive of disease-free survival ( $p = 0.00051$ ), as was stromal fraction  $>36.5\%$  ( $p = 0.037$ ).

## Conclusions

The DNN-determined ratio of desmoplastic to inflamed ratio is a novel and powerful predictor of disease recurrence in locally excised early rectal cancer. It can be assessed on a single H&E section so could be applied in routine clinical practice to improve the prognostic information available to patients and clinicians to inform the decision about further management.

Key words: early rectal cancer, local excision, prognostic factor, stroma, desmoplastic stroma, inflamed stroma

## Introduction

Local excision (LE) for early rectal cancer (ERC) is gaining in popularity as patients and clinicians seek to avoid traditional radical surgery with its morbidity. However, by not removing the mesorectum, the definitive nodal status remains unknown, and there is a higher risk of local recurrence due to occult metastases<sup>1</sup>. This may be mitigated by close surveillance to detect any recurrence early when amenable to salvage surgery, or more aggressively, by adjuvant chemoradiotherapy (CRT) or completion radical surgery. These latter options carry some morbidity so prognostic features more accurate than currently used<sup>2</sup> are sought to better guide their use.

In many types of solid cancer, a higher proportion of stroma correlates with poorer prognosis<sup>3</sup>. The usual measure of stromal contribution is tumour stromal ratio (TSR); this can be obtained from H&E slides. It has been proposed that TSR should be included in the TNM (tumour, node, metastasis) staging algorithm to improve prognostic information<sup>4</sup>. One drawback is low interobserver agreement between pathologists<sup>5</sup>. Promising digital methods have been developed; in a study of over 1,800 colorectal cancers, tumours with over 65% stroma had almost double the risk of recurrence (42% vs 22%) in 10 years compared with those with less than 50% stroma<sup>6</sup>. In early colorectal cancer a combined measure of TSR and epithelial ploidy may be better than TSR alone<sup>7</sup>.

An artificial intelligence-based method of tissue segmentation within the tumour has been developed and shown to accurately measure tissue compartments in colorectal cancer<sup>8,9</sup>. This segmentation includes two types of stroma: desmoplastic and inflamed. Desmoplastic stroma is characterised by disorganised production of connective tissue, comprising mainly collagen fibres and has been assessed in colorectal cancer<sup>10</sup>. The main cells are cancer-associated fibroblasts (CAF)<sup>11</sup>. Inflamed stroma is characterised by a lymphocyte-rich infiltrate.

Gene expression has also been used to assess the stroma and immune components of the tumour. The ESTIMATE stromal and immune scores use expression data for 141 genes each to infer the fraction of stromal and immune cells in a tumour<sup>12</sup> and have shown prognostic utility in colorectal cancer<sup>13</sup>.

Previous studies have assessed colorectal cancer in general and it is not clear whether these results are applicable to patients with ERC suitable for organ preserving surgery. After LE, currently used prognostic features stratify patients' recurrence risk and are used to advise on further management, but these are less than adequate: some patients undergo radical completion surgery to find no residual tumour<sup>14</sup>, while others undergo surveillance for low-risk disease yet develop recurrence<sup>15</sup>. The aim of this study is to

assess the value of stromal content in locally excised ERC as a biomarker for recurrence risk. An accurate and clinically applicable biomarker would be a valuable adjunct to inform the decision-making process about subsequent management after LE.

## Materials and methods

### *Ethics*

This study was approved by West Midlands – South Birmingham Research Ethics Committee as ‘An observational study to correlate the results of ploidy and stroma analysis with prognosis in early rectal cancer’ (16/WM/0443, 28/10/2016) and as ‘Pre-treatment molecular stratification and the histogenic origins of rectal cancer’ as an umbrella project approved by Oxford University Research tissue bank ethics reference 11/YH/0020 and IBD Cohort 09/H1204/30.

### *Patient cohort*

The Oxford TEM database prospectively collects data on all patients undergoing LE for rectal cancer. All patients were considered to have early rectal cancer suitable for LE based on pre-operative imaging. Data includes demographics, operative details, histopathological data and follow up. All those who had surgery between 2007 and 2017 and consented to tissue use for ethically approved research were eligible; any patient who had prior radiotherapy was excluded. Formalin-fixed paraffin-embedded tissue blocks containing tumour were retrieved. Sequential sections were cut; two 5µm sections were stained with haematoxylin and eosin (H&E) and further sections used for RNA extraction.

### *DNN digital pathology*

A stained section was annotated to indicate the cancer then scanned. Artificial intelligence (AI)-based histomorphologic tissue classification was undertaken using a deep neural net algorithm (DNN)<sup>8</sup> to quantify tissue composition across the whole lesion. DNN automatically segments the tumour into the following compartments (Figure 1) and quantifies the area (mm<sup>2</sup>) with a maximum resolution of 50µm<sup>2</sup> for a single area:

- Background (whitespace, excluded from subsequent analysis)
- Necrosis
- Epithelium (tumour area)
- Desmoplastic stroma
- Inflamed stroma

- Mucin
- Non-neoplastic mesenchymal components of bowel wall.

### *Stromal Fraction (TSR)*

A stained section was scanned and digitally annotated. Two masks were created, for background and connective tissue. After further image processing the masks were combined and TSR calculated according to previously published methodology<sup>7</sup> (Figure 2).

### *Ploidy*

A section was used for optical density ploidy analysis according to previously published methodology<sup>16</sup> and classification criteria<sup>17</sup>. The DNA ploidy histogram for each sample was classified as diploid or non-diploid. The ploidy status was combined with the TSR, using a 50% stromal fraction cut-off, into four groups with the intermediate two combined: diploid low stroma, diploid high stroma + non-diploid low stroma, non-diploid high stroma<sup>7</sup>.

### *Gene expression*

Five slides were deparaffinised then dissected using a 21-gauge needle and the marked H&E slide as a guide, to extract tumour tissue. RNA was extracted using Roche High Pure FFPE RNA isolation kit (v3, October 2012, modified protocol). The extract was treated with Invitrogen Amplification Grade DNase treatment (ThermoFisher, catalogue number: 18068015) to remove DNA. The extracted RNA was submitted for 3' mRNA sequencing, using reference genome GRCh37.EBVB95-8wt.ERCC.

### *Signature scores*

ESTIMATE scores<sup>12</sup> for stromal and immune signature were calculated using the Estimate R package<sup>18</sup>. Each signature is based on 141 genes. The sum of the stromal and immune scores is used to infer tumour purity.

### *Statistical analysis*

Most results were available for most patients, but for a few one or more of the results was unavailable; data are given for all available patients for each technique. Data were collated and descriptive statistics obtained in Excel (Microsoft). Summary data are reported as median and interquartile range. Analysis was undertaken using R statistical software ([www.r-project.org](http://www.r-project.org)). Mann-Whitney was used to test for group differences. Receiving Operating Characteristic (ROC) curves were used to assess the area under the curve (AUC) and optimal cut-points, using the Youden index. Cox proportional-hazards regression was used to assess outcome and Kaplan-Meier survival estimates obtained. Akaike's information criterion (AIC) was

used to compare fit of survival models. Univariable analysis assessed association with disease recurrence. Disease-free survival (DFS) was calculated from date of LE to date of detection of local or distant recurrence, or censoring. Patients without recurrence were censored at date of last follow-up or death.

## Results

Suitable TEM tissue samples were available from 150 patients. TSR results were available for 143 and ploidy for 140 of these. DNN analysis results were available for 140; there was an overlap of 139 between these groups. ESTIMATE stroma and immune scores were available for all patients. Table 1 shows demographic, tumour, and outcome data.

### *DNN analysis*

The proportional classified areas are shown in Figure 3. The median proportion of epithelium was 0.52 (IQR 0.42-0.59). Considering only the cellular areas, the median epithelial tumour content was 56% (IQR 49-64%). For desmoplastic stroma, the median proportional area was 0.30 (IQR: 0.22-0.40) and for inflammatory stroma 0.05 (IQR: 0.02-0.09).

The ratio of desmoplastic to inflamed (D:I) stromal area was calculated and compared with the occurrence of disease recurrence. Figure 4A shows the log of the ratio for the two groups; there was significantly more desmoplastic stroma relative to inflamed in those with recurrence ( $p=0.00067$ ).

Assessing the value of this ratio as a predictor of recurrence produced the ROC curve in Figure 4B. This determines a D:I cut-off at 5.41 (0.73 for log ratio). AUC is 0.71 (95% CI: 0.63-0.80). At this cut-off, the sensitivity is 0.92, specificity 0.50, positive predictive value (PPV) 0.30 and negative predictive value (NPV) 0.97. This model defined a poor prognosis group with high D:I comprising 58% (81/140) of the total. The DFS curve shows significant survival advantage in those with low ratio (Figure 4C). The 5-year DFS for the high D:I group is 65% (95% CI: 54-78) compared with 96% (95% CI: 90-100) for the low D:I group. AIC for this model was 221. The analysis was repeated using cell count rather than area and similar results obtained.

### *Stroma Fraction (TSR)*

TSR for this group varied from 19% to 62% stroma, with a median of 35%. Using the conventional 50% cut-off, 11 (7.7%) tumours were classified as high stroma and had worse DFS (Figure 5A). These comprised three T1 tumours, five T2 and three T3 tumours. ROC suggested an alternate cut-off at 36.5%; this selected 56 (39%) into the high-risk group (Figure 5B). The estimated 5-year DFS in the 50% cut-off high-

risk group was 51% (95% CI: 25-100) compared with 81% (95% CI: 73-89) for low stromal tumours, model AIC=239. Corresponding figures using 36.5% cut-off were 70% (95% CI: 58-85) 5-year DFS for the high-risk group compared with 84% (95% CI: 75-93) for low stromal tumours; AIC was 238.

As previous work<sup>7</sup> suggested that in early-stage cancer a combination of ploidy and TSF may provide better prognostic value than TSR alone this was also assessed, categorising samples as high and low stroma (50% cut-off), and diploid or non-diploid. Figure 5C shows a significant difference in DFS but again with a small high-risk group (10/140, 7%), AIC=229. Using the 36.5% cut-off yielded a higher proportion in the non-diploid, high stroma group (42/140, 30%) but did not give a significant difference in DFS ( $p=0.16$ ).

The proportional tumour area of stroma (desmoplastic and inflamed) obtained via DNN methodology was compared with the stromal fraction. Figure 5D shows a weak positive correlation,  $R^2 = 0.23$ .

### *ESTIMATE scoring for tumour compartments*

The median stromal score was -465 (IQR -994 – 539) and immune score -432 (IQR -817 – 251). These scores were compared with the occurrence of disease recurrence; no statistically significant association for either scores was observed. However, the S:I (stroma: immune) ratio showed a significant association with recurrence ( $p=0.03$ ) with median S:I 1.15 in the disease-free group and 0.96 in the recurrence group. ROC curve (Figure 6A) showed AUC 0.64 and indicated an optimal cut-off of 1.24. This classified 85 (59%) tumours as low S:I (high risk); this group had an estimated 5-year DFS of 69% (95% CI: 58-81) compared with 92% (95% CI: 85-100) for the high S:I group. The sensitivity was 0.92, specificity 0.45, PPV 0.27 and NPV 0.96. DFS curves are shown in Figure 6B, AIC=228.

The ESTIMATE stromal and immune scores were compared with DNN cell count proportion in desmoplastic and inflamed stroma (Supplementary figure S2); these showed a modest positive correlation ( $R^2=0.37$  and  $0.21$  respectively). A comparison with proportional area of the stromal compartments showed weaker correlations than the cell count proportion. The median tumour purity, calculated from the ESTIMATE scores, was 0.89 (IQR 0.87-0.91). This is higher than the DNN-measured epithelial area and the two measures had a moderate correlation ( $R^2=0.38$ ).

Table 2 summarises the predictive performance for each of the three stromal measures considered.

Supplementary Table S1 shows the results of univariable analysis for association with disease recurrence and finds only pT stage, positive resection margin and D:I to be significant.

## Discussion

This study has addressed a particular patient group – those undergoing LE for ERC. While all tumours were assessed as suitable for LE on pre-operative staging, with patients fully informed of the implications, histopathology sometimes shows the tumour to be more advanced or aggressive than expected<sup>2</sup>, or the resection margin to be less than ideal, especially if the tumour location is low on the sphincter or high where there is risk of peritoneal breach. Following surgery, patient and clinician must decide how to proceed; whether surveillance alone will suffice, or adjuvant treatment be employed. The present results offer a refinement to the current assessment of recurrence risk based on tumour stage, size, and lymphovascular invasion<sup>19</sup>. The additional consideration of stromal composition may give a more individualised risk assessment.

In this series of 143 patients, the D:I ratio provided good discrimination between good and poor prognosis tumours with 5-year estimated DFS of 96% and 65% respectively. High desmoplastic stromal content is often associated with activated fibroblasts, TGF- $\beta$  signalling activation and poor prognosis<sup>20</sup>. In contrast, inflamed stroma is rich in lymphocytes; immune activation tends to be associated with a favourable outcome<sup>21</sup>.

TSR provided discrimination between high and low risk groups, with 5-year DFS estimates of 70% and 84%. Tumours with over 36.5% stroma had poorer prognosis, in keeping with Scheer's<sup>5</sup> finding of significantly worse disease-free and overall survival with TSR over 30% among 154 patients with rectal cancer. It is worth noting that only 13% of tumours in our series had stromal fraction over 50%, lower than the 34% of 377 rectal cancers in the QUASAR trial<sup>6</sup> probably reflecting the earlier cancers in the current series, which may not yet have produced a great volume of stroma. Furthermore, in early cancers more subtle features of the stroma such as the presence of inflammatory infiltrates rather than simply volume may have a greater role. The addition of ploidy to TSR identified a small group (7%) of non-diploid, high stroma tumours which had poorer outcomes, but this stratification was weaker than for more advanced colorectal cancers<sup>7</sup>.

Although the individual stromal and immune ESTIMATE scores did not show a significant association with survival, a ratio of the two discriminated high and low risk groups. The disadvantage of this technique is the requirement for gene expression data which involves time and cost, making it less useful for current clinical practice. The ESTIMATE scores can be used to infer tumour purity, or epithelial content. However, this ignores other normal tumour content such as the vascular cells and muscle, which are recognised

separately in the DNN mesenchyme component; this may result in over-estimation and explain the higher values for tumour purity compared to DNN epithelial area in this series.

The DNN AI-based segmentation of tumour compartments is a novel technique with the practical advantage that it requires only an H&E section so can generate data cost-effectively with minimal delay following surgery; this is its first application to a series of locally excised rectal cancers. Once the algorithm has been trained, the technique is fully reproducible, however when used on standard H&E sections as in this clinical setting the sections may be thicker with a less standardised staining protocol than in the research setting where the algorithm was developed, which could contribute some variability to results. D:I could be considered along with the standard histopathological features in deciding on management following surgery – with high D:I additional treatment, either completion surgery or adjuvant CRT, should be considered or if neither of these is favoured, meticulous surveillance can be instituted. Alternatively, with low D:I, patients can be reassured and follow surveillance with greater confidence. DNN can also be used on biopsy specimens<sup>8</sup> so could contribute to the initial decision for LE. Further work with larger patient numbers is required to determine whether D:I can also indicate radiosensitivity when considering adjuvant CRT.

The limitations of this study are the relatively small number of patients. We therefore recommend validation of the DNN-based assessment method in independent cohorts and the prospective clinical trial setting. This is facilitated by rapid adoption of digital pathology for the assessment of primary resection specimens. The algorithm generated here will be made publicly available to facilitate translation. A few of the tumours, considered 'early' prior to surgery, were T3. The techniques discussed here require further study in more advanced cancers.

This paper has assessed three measures of stromal content and found some promising prospects for better stratification of locally excised rectal cancers. All techniques were prognostic, emphasising the importance of complex interactions between epithelium, stroma, and immune components in determining tumour behaviour. In contrast to colorectal cancer in general<sup>22</sup>, TSR alone is not so useful in this group. Adding ploidy increases its value, but less than in more advanced cancers. A ratio of ESTIMATE scores is promising but not immediately useful. AI-based histomorphology offers a quick and simple addition to standard histopathological assessment with good prognostic value in this patient group. This could provide a valuable extra tool to inform the discussion about subsequent management with patients following local excision.

## Acknowledgements and funding

This research was funded by grants from the Bowel Disease Research Foundation (BDRF) and Oxford Colon Cancer Trust (Occtopus) and was supported by the NIHR Biomedical Research Centre, Oxford.

VHK gratefully acknowledges funding by the Swiss National Science Foundation (P2SKP3\_168322/1 and P2SKP3\_168322/2), the Werner-Hedy Berger Janser Foundation and the Promedica Foundation (F-87701-41-01).

This work was supported by Wellcome Trust Senior Clinical Research Fellowship (206314/Z/17/Z) to SJL.

Core funding to the Wellcome Centre for Human Genetics was provided by the Wellcome Trust (090532/Z/09/Z and 203141/Z/16/Z). We thank the Oxford Genomics Centre and John Broxholme and Ben Wright of the Bioinformatics & Statistical Genetics Core at the Wellcome Centre for Human Genetics for the generation and initial processing of the sequencing data.

## Author contributions

Helen Jones: conception, design, data acquisition, analysis and interpretation, drafting, revising, final approval; Chris Cunningham: conception, design, data interpretation, revising, final approval; Hanne Askautrud: data acquisition, analysis and interpretation, revising, final approval; Håvard Danielsen: data interpretation, revising, final approval; David Kerr: data interpretation, revising, final approval; Enric Domingo: data interpretation, revising, final approval; Tim Maughan: data interpretation, revising, final approval; Simon Leedham: conception, design, data interpretation, revising, final approval; Viktor Koelzer: conception, design, data interpretation, revising, final approval

## List of abbreviations

D:I	ratio of desmoplastic to inflamed stroma determined by DNN
DFS	disease-free survival
DNN	deep neural network
ERC	early rectal cancer
LE	local excision

S:I ratio of ESTIMATE stromal to immune scores

TEM transanal endoscopic microsurgery

TSR tumour stromal ratio

## List of online Supporting Information

### TABLE S1: UNIVARIABLE ANALYSIS FOR DISEASE RECURRENCE

**FIGURE S1: SCATTER PLOTS SHOWING CORRELATION BETWEEN DNN MEASURE AND GENE EXPRESSION SCORES IN RECTAL CANCERS. A.** COMPARISON OF ESTIMATE STROMAL SCORE WITH THE CELL COUNT PROPORTION OF DESMOPLASTIC STROMA CELLS DETERMINED BY DNN IN 139 RECTAL CANCERS,  $R^2 = 0.37$ . **B.** COMPARISON OF ESTIMATE IMMUNE SCORE WITH THE CELL COUNT PROPORTION OF INFLAMED STROMA CELLS DETERMINED BY DNN IN 139 CANCERS,  $R^2 = 0.21$ .

## Reference list

1. Stornes T, Wibe A, Nesbakken A, Myklebust TÅ, Endreseth BH. National Early Rectal Cancer Treatment Revisited. *Dis Colon Rectum*. 2016;59(7):623-629. doi:10.1097/DCR.0000000000000591
2. Patel SA, Chen YH, Hornick JL, et al. Early stage rectal cancer: clinical and pathologic prognostic markers of time to local recurrence and overall survival after resection. *Dis Colon Rectum*. 2014;57(4):449-459. doi:0.1097/DCR.0b013e3182a70709
3. Wu J, Liang C, Chen M, Su W. Association between tumor-stroma ratio and prognosis in solid tumor patients: a systematic review and meta-analysis. *Oncotarget*. 2016;7(42):68954–68965. doi:10.18632/oncotarget.12135
4. van Pelt GW, Sandberg TP, Morreau H, et al. The tumour-stroma ratio in colon cancer: the biological role and its prognostic impact. *Histopathology*. Aug 2018;73(2):197-206. doi:10.1111/his.13489
5. Scheer R, Baidoshvili A, Zoidze S, et al. Tumor-stroma ratio as prognostic factor for survival in rectal adenocarcinoma: A retrospective cohort study. *World J Gastrointest Oncol*. 2017;9(12):466-474. doi:10.4251/wjgo.v9.i12.466

6. Hutchins G, Treanor D, Wright A, et al. Intratumoral stromal morphometry predicts disease recurrence but not response to 5-fluorouracil—results from the QUASAR trial of colorectal cancer. *Histopathology*. 2017;72(3):391-404. doi:10.1111/his.13326
7. Danielsen HE, Hveem TS, Domingo E, et al. Prognostic markers for colorectal cancer: estimating ploidy and stroma. *Ann Oncol*. 2018;29(3):616-623. doi:10.1093/annonc/mdx794
8. Domingo E, Chatzipli A, Richman S, et al. Assessment of tissue composition with digital pathology in colorectal cancer. *Cancer Reserch*. 2019;79(13 Suppl):A4446. doi:10.1158/1538-7445.AM2019-4446
9. Kleeman S, Koelzer V, Jones H, et al. Exploiting Differential Wnt Target Gene Expression to Generate a Molecular Biomarker for Colorectal Cancer Stratification. *Gut*. 2020;69(6):1092-1103. doi:10.1136/gutjnl-2019-319126
10. Nearchou I, Ueno H, Kajiwarra Y, et al. Automated Detection and Classification of Desmoplastic Reaction at the Colorectal Tumour Front Using Deep Learning. Article. *Cancers*. 2021;13(7):1615. doi:10.3390/cancers13071615
11. Coulson-Thomas V, Coulson-Thomas Y, Gesteira T, et al. Colorectal cancer desmoplastic reaction up-regulates collagen synthesis and restricts cancer cell invasion. *Cell and Tissue Research*. 2011;346(2):223-236. doi:10.1007/s00441-011-1254-y
12. Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013;4:2612. doi:10.1038/ncomms3612
13. Liu W, Ye H, Liu YF, et al. Transcriptome-derived stromal and immune scores infer clinical outcomes of patients with cancer. *Oncol Lett*. 2018;15(4):4351-4357. doi:10.3892/ol.2018.7855
14. Jones H, Cunningham C, Nicholson G, Hompes R. Outcomes Following Completion and Salvage Surgery for Early Rectal Cancer: A Systematic Review. *European J Surg Oncol*. 2018;44(1):15-23. doi:10.1016/j.ejso.2017.10.212
15. Jones H, Hompes R, Mortensen N, Cunningham C. Modern management of T1 rectal cancer by transanal endoscopic microsurgery: a 10-year single-centre experience. *Colorectal Disease*. 2018;20(7):586-592. doi:10.1111/codi.14029
16. Pradhan M, Abeler V, Danielsen H, Sandstad B, Tropé CK, GB, Risberg A. Prognostic importance of DNA ploidy and DNA index in stage I and II endometrioid adenocarcinoma of the endometrium. *Ann Oncology*. 2012;23(5):1178-1184. doi:10.1093/annonc/mdr368
17. Pradhan M, Abeler V, Danielsen H, Tropé C, Risberg B. Image cytometry DNA ploidy correlates with histological subtypes in endometrial carcinomas. OriginalPaper. *Modern Pathology*. 2006;19(9):1227-1235. doi:doi:10.1038/modpathol.3800641

18. *estimate: Estimate of Stromal and Immune Cells in Malignant Tumor Tissues from Expression Data v 1.0.13/r21*. 2016. <https://rdr.io/rforge/estimate/>
19. Bach SP, Hill J, Monson JR, et al. A predictive model for local recurrence after transanal endoscopic microsurgery for rectal cancer. *Br J Surg*. 2009;96(3):280-90. doi:10.1002/bjs.6456
20. Ueno H, Shinto E, Shimazaki H, et al. Histologic categorization of desmoplastic reaction: its relevance to the colorectal cancer microenvironment and prognosis. *Ann Surg Oncol*. 2015;22(5):1504-1512. doi:10.1245/s10434-014-4149-9
21. Idos G, Kwok J, Bonthala N, Kysh L, Gruber S, Qu C. The Prognostic Implications of Tumor Infiltrating Lymphocytes in Colorectal Cancer: A Systematic Review and Meta-Analysis. Original Paper. *Scientific Reports*. 2020;10(1):1-14. doi:doi:10.1038/s41598-020-60255-4
22. West N, Dattani M, McShane P, et al. The proportion of tumour cells is an independent predictor for survival in colorectal cancer patients. *Br J Cancer*. 2010;102(10):1519-1523. doi:10.1038/sj.bjc.6605674
23. WHO Classification of Tumours Editorial Board. *Digestive System Tumours*. 5 ed. WHO Classification of Tumours. WHO International Agency for Research on Cancer; 2019.

# Tables

**TABLE 1: DEMOGRAPHIC, TUMOUR AND OUTCOME DATA FOR THE PATIENT GROUP**

	<b>Number (%)</b>
<b>Number of patients</b>	143
<b>Age, median (IQR)</b>	69 (64-78)
<b>Sex Male (%): Female</b>	87 (61%): 56
<b>Tumour height above anal verge</b>	
<6.0 cm	68 (48%)
>6 cm	75 (52%)
<b>Tumour stage</b>	
<b>pT1</b>	76 (53%)
pT1 sm1	11 (8%)
pT1 sm2	21 (15%)
pT1 sm3	44 (31%)
<b>pT2</b>	58 (41%)
<b>pT3</b>	9 (6%)
<b>Tumour diameter (cm)</b>	
<3.0	113 (79%)
3-5	26 (18%)
>5	4 (3%)
<b>Histological subtype<sup>23</sup></b>	
<b>Adenocarcinoma NOS</b>	113 (79%)
Mucinous adenocarcinoma	1 (1%)
Adenocarcinoma with mucinous component	13 (9%)
Serrated adenocarcinoma	12 (8%)
Adenoma-like adenocarcinoma	4 (3%)
<b>Differentiation</b>	
Well differentiated	12 (9%)
Moderately differentiated	123 (86%)
Poorly differentiated	8 (8%)

<b>Lympho-vascular invasion</b>	44 (31%)
<b>Resection margin &lt;1mm</b>	41 (29%)
<b>Post-TEM management</b>	
<b>Completion surgery</b>	13 (9%)
<b>Adjuvant radiotherapy</b>	42 (29%)
<b>Surveillance alone</b>	88 (62%)
<b>Local recurrence</b>	18 (13%)
<b>Distant metastases</b>	14 (10%)
<b>Length of follow-up, years, median (IQR)</b>	4.0 (2.2 – 5.3)

**TABLE 2: SUMMARY OF PROGNOSTIC VALUE OF THREE STROMAL MEASURES IN EARLY RECTAL CANCER**

Measure	AUC	Proportion at high risk	Estimated 5-year DFS (95% CI)	Akaike IC	p value
Stromal fraction >36.5%	0.58	56%	70% (58-85) vs 84% (75-93)	238	0.037
Stroma fraction >50% + ploidy		7%	65% (39-100) vs 86% (79-94) vs 86% (75-100)*	229	0.032
Desmoplastic: inflamed stroma area ratio >5.41	0.71	56%	65% (54-78) vs 96% (90-100)	221	0.00067
ESTIMATE stromal: immune score <1.24	0.64	59%	69% (58-81) vs 92% (85-100)	228	0.00051

AUC=AREA UNDER CURVE, DFS=DISEASE-FREE SURVIVAL, CI=CONFIDENCE INTERVAL, AKAIKE IC= AKAIKE INFORMATION CRITERION, \*3-YEAR SURVIVAL RESULTS SHOWN

## Figure legends

**FIGURE 1: EXAMPLE OF A TUMOUR ANNOTATED USING A DEEP NEURAL NET ALGORITHM (DNN), SHOWING OVERVIEW AND ZOOMED VIEW. A. UNANNOTATED ORIGINAL, B. DNN ANNOTATION, C. CELL SEGMENTATION.** TISSUE COMPARTMENTS: RED = EPITHELIUM; GREEN = DESMOPLASTIC STROMA; PURPLE = INFLAMED STROMA; BLUE = MUCIN; YELLOW = WHITESPACE; PINK = NECROSIS.

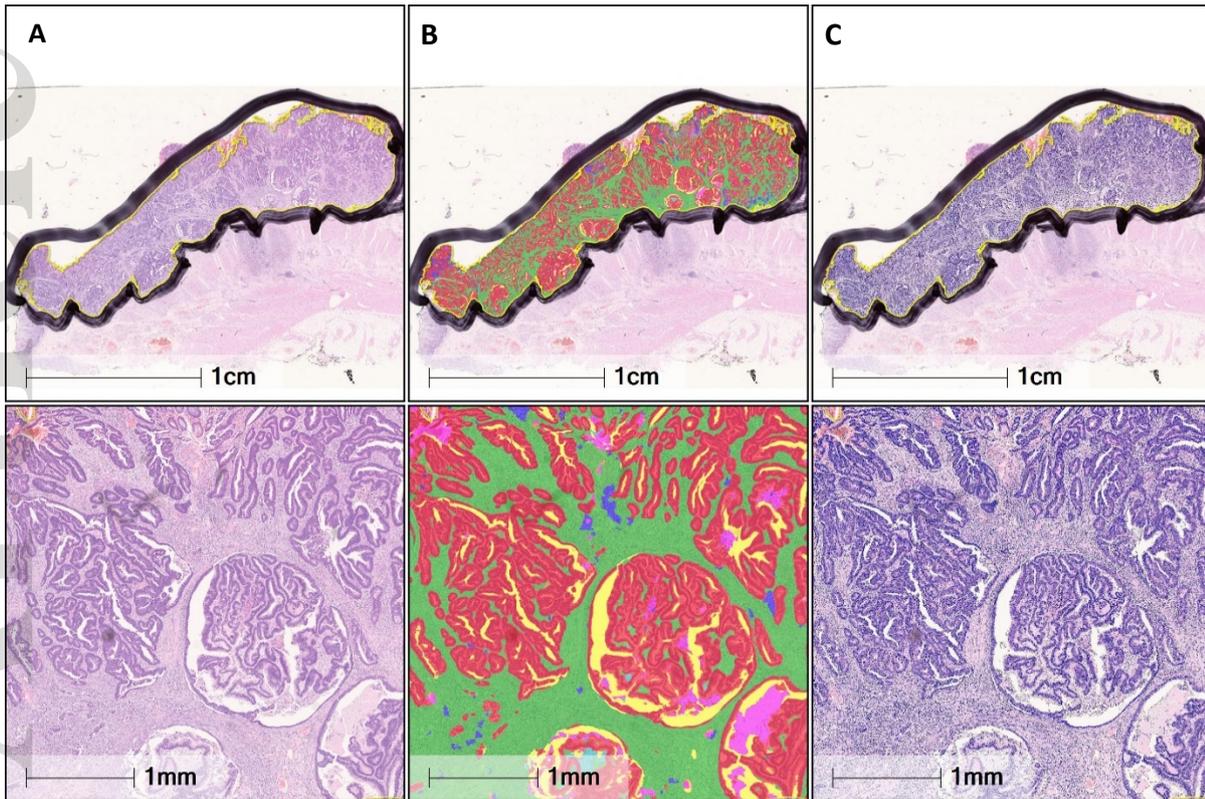
**FIGURE 2: EXAMPLE OF A TUMOUR WITH A HIGH STROMAL FRACTION (60%) BY TSR. THE TUMOUR WAS PT3. THE PATIENT UNDERWENT ADJUVANT CHEMORADIOTHERAPY BUT WAS DIAGNOSED WITH MULTIPLE LIVER METASTASES 9 MONTHS AFTER LOCAL EXCISION. A. SCAN OF H&E SLIDE FROM THE TUMOUR. B. DIGITALLY PROCESSED SCAN SHOWING SEGMENTED STROMA IN LIGHT GREEN COLOUR. C. MASK OF TUMOUR PRODUCED BY DIGITAL PROCESSING, WITH STROMA IN GREEN. RED INDICATES TUMOUR CELLS AND BLUE IS ARTEFACT. THIS MASK WAS USED TO CALCULATE THE TUMOUR STROMAL FRACTION.**

**FIGURE 3: PROPORTIONAL AREA OF THE TUMOUR COMPONENTS DETERMINED BY DNN DIGITAL PATHOLOGY IN 140 RECTAL CANCERS.** IN MOST TUMOURS, EPITHELIUM FORMS THE MAJOR PART AND EPITHELIAL CELLS PREDOMINATE. DESMOPLASTIC STROMA FORMS A SIGNIFICANT PART OF MOST TUMOURS WHILE INFLAMED STROMA MAKES UP A SMALLER, VARIABLE PROPORTION. FEW TUMOURS ARE MUCINOUS. NON-NEOPLASTIC TISSUE AND NECROSIS FORM ONLY A MINOR PART IN MOST ANNOTATED TUMOUR REGIONS.

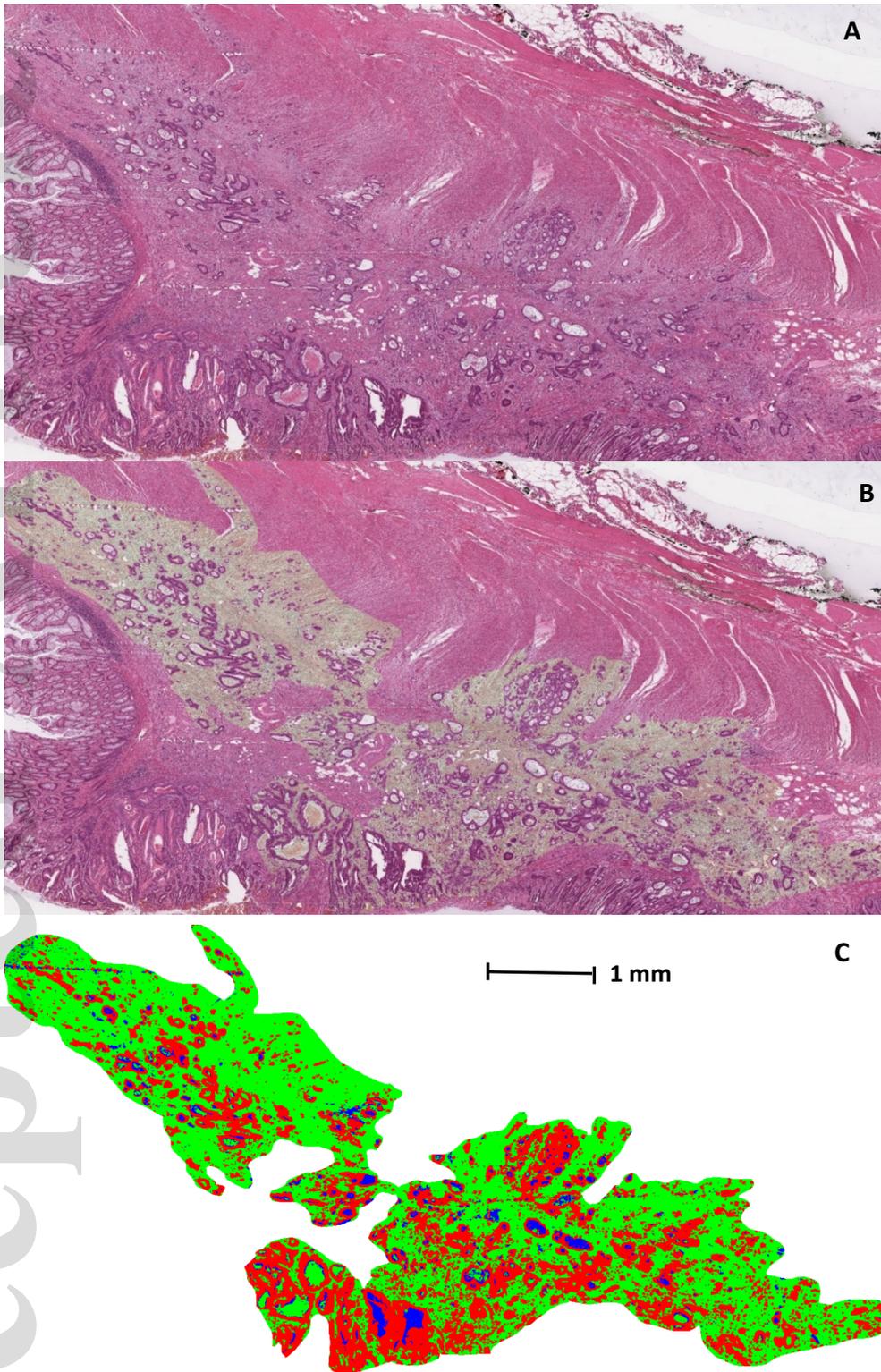
**FIGURE 4: LOG RATIO OF DESMOPLASTIC TO INFLAMED STROMA AREA IN RECTAL CANCER. A: BOX AND WHISKER PLOT SHOWING LOG D:I FOR STROMAL AREA IN 140 RECTAL CANCERS. THE MEDIAN LOG D:I WAS 1.01 IN THE 26 WHO DEVELOPED RECURRENCE AND 0.74 IN THE 114 WHO REMAINED DISEASE FREE. THE LOG D:I WAS SIGNIFICANTLY HIGHER IN THE RECURRENCE GROUP,  $P=0.00067$ . B: ROC CURVE ASSESSING LOG D:I AS PREDICTOR OF RECURRENCE. USING THE YODEN CRITERION, THE OPTIMAL CUT-OFF IS LOG D:I=0.73, WITH AUC 0.71. C. DISEASE-FREE SURVIVAL BY D:I, WITH HIGH D:I DEFINED AS  $\geq 5.41$ , THE CUT-OFF VALUE ON ROC CURVE.**

**FIGURE 5: TUMOUR STROMAL FRACTION IN RECTAL CANCER. A. DISEASE-FREE SURVIVAL BY STROMAL FRACTION USING A CUT-OFF OF 50% STROMAL FRACTION. B. DISEASE-FREE SURVIVAL BY STROMAL FRACTION USING A CUT-OFF OF 36.5% STROMAL FRACTION. C. DISEASE-FREE SURVIVAL BY STROMAL FRACTION AND PLOIDY, D=DIPLOID, ND=NON-DIPLOID, LS=LOW STROMA, HS=HIGH STROMA. D. SCATTER PLOT SHOWING CORRELATION OF DNN DESMOPLASTIC STROMA PROPORTION WITH TUMOUR STROMAL FRACTION.**

**FIGURE 6: RATIO OF ESTIMATE STROMA AND IMMUNE SCORES IN RECTAL CANCERS. A: ROC CURVE ASSESSING S:I AS PREDICTOR OF RECURRENCE. USING THE YODEN CRITERION, THE OPTIMAL CUT-OFF IS S:I=1.24, WITH AUC 0.64. B: DISEASE-FREE SURVIVAL BY S:I, WITH LOW S:I DEFINED AS  $\geq 1.24$ , THE CUT-OFF VALUE ON ROC CURVE.**



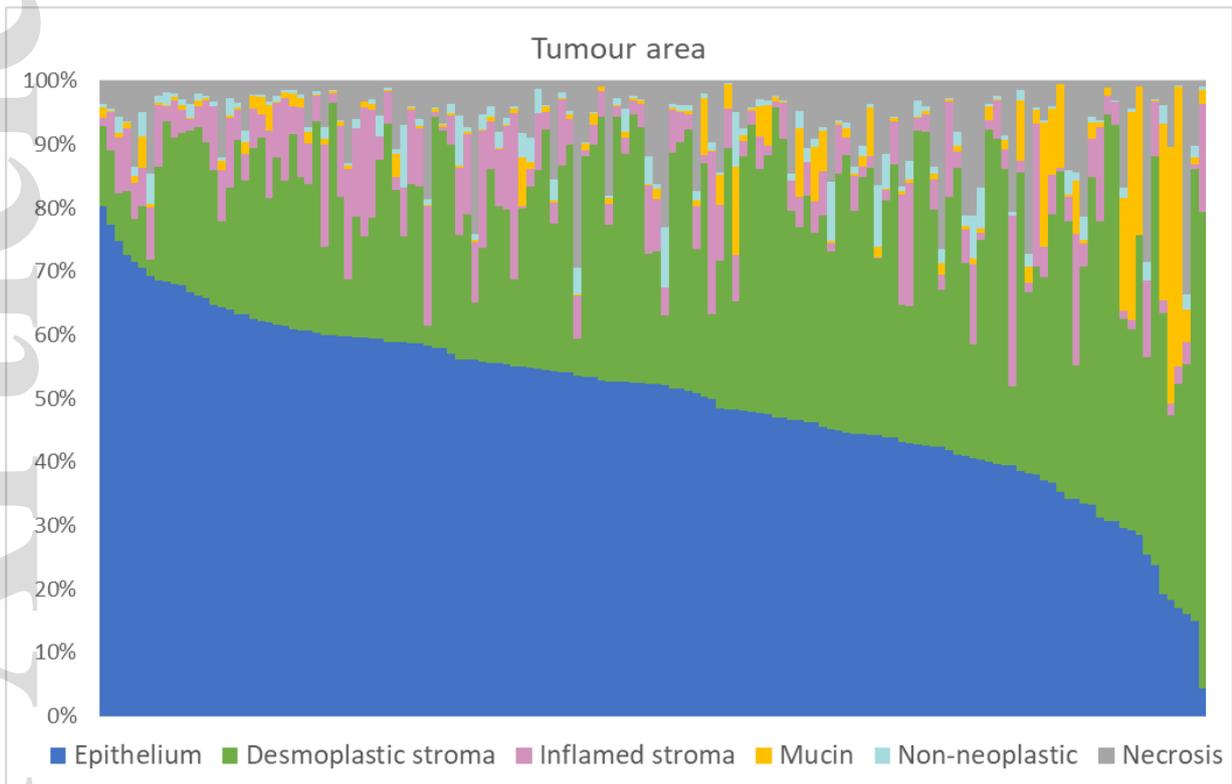
**FIGURE 1: EXAMPLE OF A TUMOUR ANNOTATED USING A DEEP NEURAL NET ALGORITHM, SHOWING OVERVIEW AND ZOOMED VIEW. A. UNANNOTATED ORIGINAL, B. DNN ANNOTATION, C. CELL SEGMENTATION. TISSUE COMPARTMENTS: RED = EPITHELIUM; GREEN = DESMOPLASTIC STROMA; PURPLE = INFLAMED STROMA; BLUE = MUCIN; YELLOW = WHITESPACE; PINK = NECROSIS.**



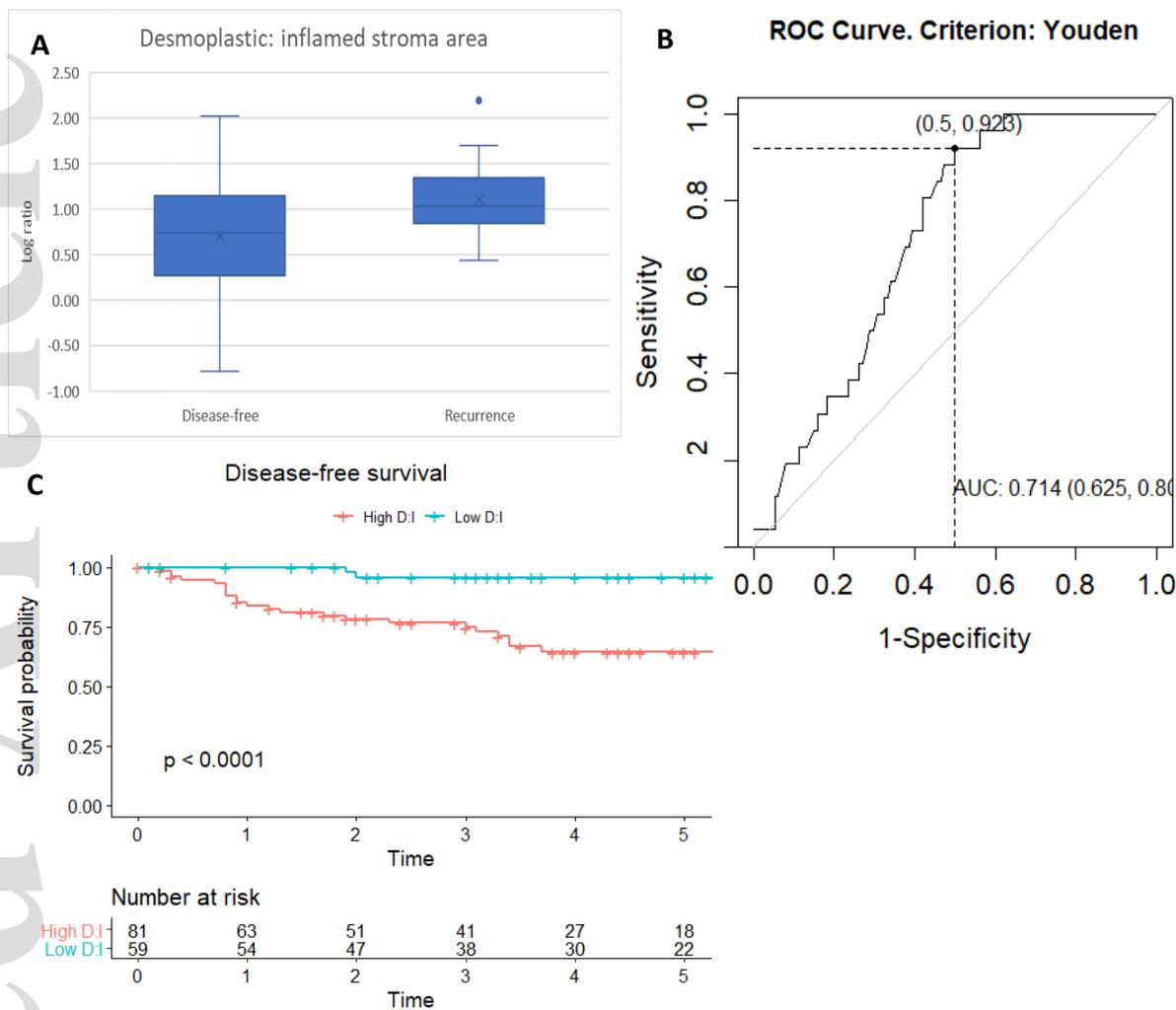
**FIGURE 2: EXAMPLE OF A TUMOUR WITH A HIGH STROMAL FRACTION (60%) BY TSR. A. SCAN OF H&E SLIDE FROM THE TUMOUR. B. DIGITALLY PROCESSED SCAN SHOWING SEGMENTED STROMA IN LIGHT GREEN COLOUR. C. MASK OF TUMOUR**

PRODUCED BY DIGITAL PROCESSING, WITH STROMA IN GREEN. RED INDICATES TUMOUR CELLS AND BLUE IS ARTEFACT. THE MASK WAS USED TO CALCULATE THE TUMOUR STROMAL FRACTION.

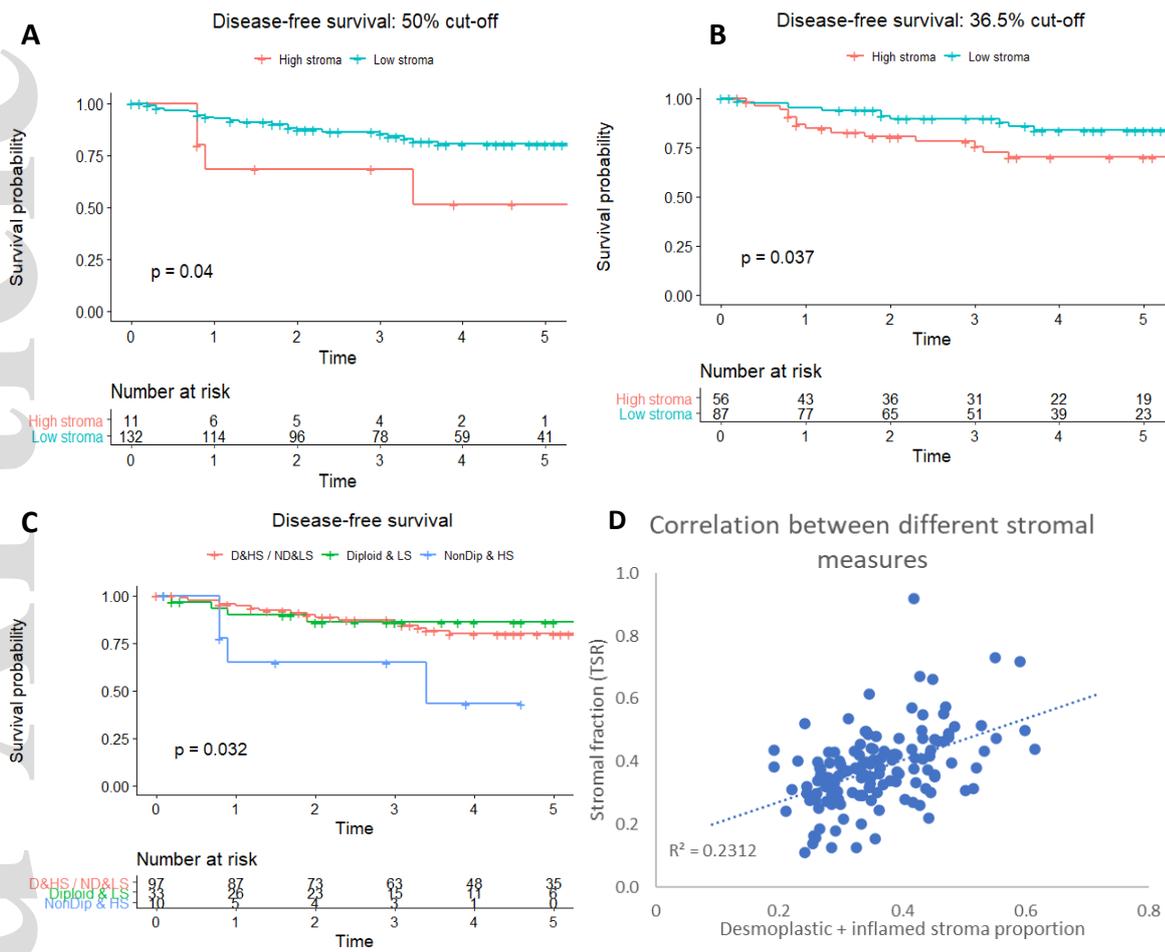
Accepted Article



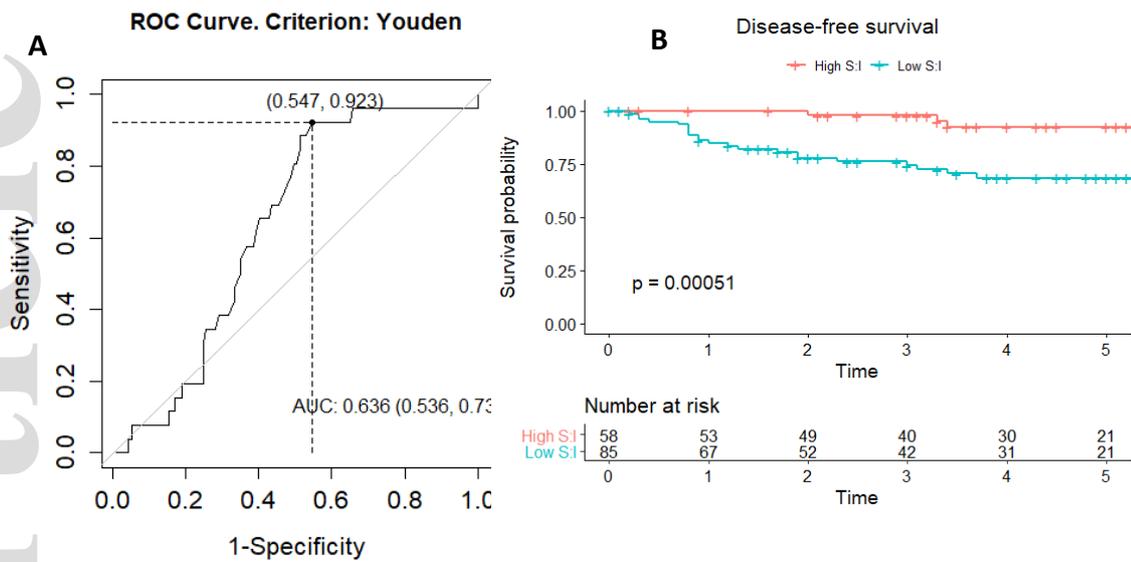
**FIGURE 3: PROPORTIONAL AREA OF THE TUMOUR COMPONENTS DETERMINED BY DNN DIGITAL PATHOLOGY IN 140 RECTAL CANCERS.** IN MOST TUMOURS, EPITHELIUM FORMS THE MAJOR PART AND EPITHELIAL CELLS PREDOMINATE. DESMOPLASTIC STROMA FORMS A SIGNIFICANT PART OF MOST TUMOURS WHILE INFLAMED STROMA MAKES UP A SMALLER, VARIABLE PROPORTION. FEW TUMOURS ARE MUCINOUS. NON-NEOPLASTIC TISSUE AND NECROSIS FORM ONLY A MINOR PART IN MOST ANNOTATED TUMOUR REGIONS.



**FIGURE 4: LOG RATIO OF DESMOPLASTIC TO INFLAMED STROMA AREA IN RECTAL CANCER. A:** BOX AND WHISKER PLOT SHOWING LOG D:I FOR STROMAL AREA IN 140 RECTAL CANCERS. THE MEDIAN LOG D:I WAS 1.01 IN 26 WHO DEVELOPED RECURRENCE AND 0.74 IN 114 WHO REMAINED DISEASE FREE. THE LOG D:I WAS SIGNIFICANTLY HIGHER IN THE RECURRENCE GROUP,  $p=0.00067$ . **B:** ROC CURVE ASSESSING LOG D:I AS PREDICTOR OF RECURRENCE. USING THE YODEN CRITERION, THE OPTIMAL CUT-OFF IS LOG D:I=0.73, WITH AUC 0.71. **C:** DISEASE-FREE SURVIVAL BY D:I, WITH HIGH D:I DEFINED AS  $\geq 5.41$ , THE CUT-OFF VALUE ON ROC CURVE.



**FIGURE 5: TUMOUR STROMAL FRACTION IN RECTAL CANCER. A.** DISEASE-FREE SURVIVAL BY STROMAL FRACTION USING A CUT-OFF OF 50% STROMAL FRACTION. **B.** DISEASE-FREE SURVIVAL BY STROMAL FRACTION USING A CUT-OFF OF 36.5% STROMAL FRACTION. **C.** DISEASE-FREE SURVIVAL BY STROMAL FRACTION AND PLOIDY, D=DIPLOID, ND=NON-DIPLOID, LS=LOW STROMA, HS=HIGH STROMA. **D.** SCATTER PLOT SHOWING CORRELATION OF DNN STROMA (DESMOPLASTIC + INFLAMED) PROPORTION WITH TUMOUR STROMAL FRACTION.



**FIGURE 6: RATIO OF ESTIMATE STROMA AND IMMUNE SCORES IN RECTAL CANCERS. A: ROC CURVE ASSESSING S:I AS PREDICTOR OF RECURRENCE. USING THE YOUDEN CRITERION, THE OPTIMAL CUT-OFF IS S:I=1.24, WITH AUC 0.64. B: DISEASE-FREE SURVIVAL BY S:I, WITH LOW S:I DEFINED AS  $\geq 1.24$ , THE CUT-OFF VALUE ON ROC CURVE.**