Investigating the Link between Collagen Deposition Patterns in the TME and Indian Triple-Negative Breast Cancer Patient Outcomes

A Thesis

submitted to

Indian Institute of Science Education and Research, Pune, in partial fulfilment of the requirements for the BS-MS Dual Degree Programme

by

Ingawale Vedant Chandrakant

Indian Institute of Science Education and Research, Pune

Dr. Homi Bhabha Road,

Pashan, Pune 411008, INDIA.

Date: April 2025

Under the supervision of **Dr Madhura Kulkarni,**

at the **Centre for Translational Cancer Research, Pune**

From May 2024 to Mar 2025

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE

Certificate

This is to certify that this dissertation entitled **'Investigating the Link between Collagen Deposition Patterns in the TME and Indian Triple-Negative Breast Cancer Patient Outcomes',** submitted towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune, represents study/work carried out by Ingawale Vedant Chandrakant at the Centre for Translational Cancer Research, Pune under the supervision of Dr Madhura Kulkarni, Head of Translational Research, during the academic year 2024-2025.

Dr Madhura Kulkarni

Committee:

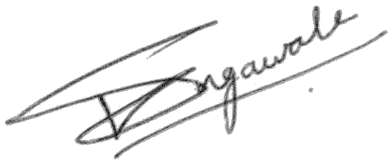
Supervisor: Dr Madhura Kulkarni

Expert: Prof Nagaraj Balasubramanian

This thesis is dedicated to my family and seventeen-year-old Vedant for their love and patience.

Declaration

I hereby declare that the matter embodied in the report entitled **“Investigating the Link between Collagen Deposition Patterns in the TME and Indian Triple-Negative Breast Cancer Patient Outcomes”** are the results of the work carried out by me at the Centre for Translational Cancer Research, Pune, under the supervision of Dr Madhura Kulkarni, and the same has not been submitted elsewhere for any other degree. Wherever others contribute, every effort is made to indicate this clearly, with due reference to the literature and acknowledgement of collaborative research and discussions.



Name of Student - Ingawale Vedant Chandrakant

Reg. no. - 20201201

Table of Contents

List of Tables

List of Figures

Abstract

Triple-Negative Breast Cancers (TNBCs) are characterised by the absence of hormone receptors' expression and disproportionately burden the Indian population. Since TNBCs do not have available targets for treatment, it makes them an aggressive subset, with an unpredictable response to systemic therapy, showing higher rates of recurrence and lower overall survival, posing a significant challenge to clinical management. Therefore, there is a need to understand the factors that govern the response and treatment outcomes.

Recently, the Extracellular Matrix (ECM) and its associated components have emerged as key determinants of breast cancer progression, affecting cellular behaviour, invasion and metastasis. The ECM, a complex dynamic entity, undergoes remodelling through intricate molecular interactions. These alterations are heavily context-dependent and can lead to an increase in the disease's aggressiveness depending on what point in time they occur.

Collagen is a crucial component of the ECM and plays an integral role in determining the mechanical properties of the tissue. With associated enzymes and other proteins, collagen alters matrix stiffness, alignment and density, substantially impacting tumour behaviour. The referred studies suggest that collagen-rich microenvironments correlate with aggressive tumour behaviour.

With a cohort of 139 TNBC patient samples and associated clinical and follow-up data, we analyse collagen deposition in TNBC patient samples by trichrome staining to understand its association with clinical features of the tumour at presentation, treatment response and survival outcomes. By examining patient-derived tissues across pathological stages and correlating collagen deposition patterns with clinical data, our study seeks to unravel the prognostic value of ECM alterations.

Acknowledgements

I express my sincere gratitude to Dr Madhura Kulkarni for being an exceptional mentor. Her guidance and feedback have kept me on track and motivated, which has been crucial throughout this project. Our discussions have immensely aided my academic growth and profoundly influenced how I approach and deal with science.

I thank Prof Nagaraj Balasubramnian for his advice, inputs and suggestions over the year. I thank him for directing and allowing us to try out an analysis pipeline developed in his lab for comparison with ours, as the exercise has been significantly insightful. I thank Dr Varun Ragunathan and Krishnendu Dandapat from IISc, Bengaluru, for acquiring SHG images. I thank Dr Nameeta Shah and Shashank Gupta for working with us to develop a whole slide image analysis pipeline. I would also like to acknowledge Dr Satyajit Rath for his insights on the project.

I express my heartfelt gratitude to my mother, Vaishali, father, Chandrakant, brother, Siddhant and my dearest grandmother, Sushila (Aai), for constantly supporting and loving me throughout my IISER journey. I thank Manav S for helping me with computation and always encouraging and supporting me during my BS-MS. I would like to thank Nikita Gupta, Lokesh Tater, Manasvi Balachandran, Jason Joby, Devapriyan, Bhavesh Vasave, Akash Datta and Krishna Menon for being amazing pillars of support.

I would like to thank Dr Devaki Kelkar and my lab mates at the Tumour Microenvironment Lab, Gomathi V, Khushi S, Sakshi D, Jisha J, Sampada G, Pooja V, S Disha R for constantly helping around in the lab during the period of my project and making science an enjoyable experience.

Lastly, I thank my seventeen-year-old self for dreaming a dream and making it through. You've made it far, and you've made me proud.

Contributions

Contributor name	**Contributor role**
Vedant Ingawale, Madhura Kulkarni	Conceptualisation Ideas
Vedant Ingawale, Krishnendu Dandapat	Methodology
Shashank Gupta, Nameeta Shah	Software
Vedant Inagwale, Madhura Kulkarni	Validation
Vedant Ingawale	Formal analysis
Vedant Inagawale	Investigation
Madhura Kulkarni, Varun Ragunathan, L S Shashidhara, C B Koppiker	Resources
Asha Chote, Vishesha Lalla, Kuhuk Saran	Data Curation
Vedant Ingawale	Writing - original draft preparation
Madhura Kulkarni	Writing - review and editing
Vedant Ingawale	Visualisation
Madhura Kulkarni	Supervision
Madhura Kulkarni	Project administration
Madhura Kulkarni	Funding acquisition

1. Introduction

1.1. The burden of Triple-Negative Breast Cancer

Breast cancer is the most common cancer type in India, with a prevalence of 14.7% of all cancers in 1 year. According to GLOBOCAN 2022, it has the highest incidence rate, accounting for 13.7% of the new cases detected in the year 2022 and a high incidence-to-mortality ratio (51%), making breast cancer the leading cause of cancer-related deaths in India (Bray *et al.*, 2024).

Currently, breast cancer is classified into four molecular phenotypes/subtypes based on the Immunohistochemistry (IHC) expression of Estrogen and Progesterone Hormone Receptors (ER/PR) and Human Epidermal Growth Factor Receptor 2 (HER2). The treatment strategies differ according to the molecular subtype of breast cancer (Howlader *et al.*, 2018).

If >1% of infiltrating tumour cells are positive for the expression of ER/PR in a tumour section by IHC, the specimen is considered ER/PR+ve; else, it is ER/PR-ve, according to the ASCO/CAP guidelines 2010 (Hammond *et al.*, 2010). According to ASCO/CAP guidelines 2018, if >10% of infiltrating tumour cells are stained for HER2 after IHC with a strong intensity, the specimen is considered HER2+ve. If >10% of tumour cells are weakly stained for HER2 after IHC, then the specimen is considered equivocal, and HER2 positivity status needs to be confirmed with a fluorescence in situ hybridisation assay. If the IHC stain for HER2 is incomplete or <10% of cells are stained for HER2, then the specimen is HER2-ve (Ahn *et al.*, 2020).

Breast Cancers that are ER+ve, PR+ve/-ve, and HER2-ve are classified as ER+ve Breast Cancers, which can further be classified as Luminal A and Luminal B; those that are ER-ve, PR-ve, HER2+ve are HER2 Breast Cancers; and those that are ER-ve,

PR-ve, HER2-ve are classified as Triple Negative Breast Cancers (TNBCs; Orrantia-Borunda *et al.*, 2022).

Treating ER+ve and HER2+ve Cancer Subtypes involves targeting the hormone receptors and the HER2 oncogene. However, these therapies do not help treat TNBCs due to the absence of targetable hormone receptors and HER2 protein overexpression or gene amplification. Treatment for TNBCs relies on systemic chemotherapy and surgery (Lehmann *et al.*, 2011). The lack of molecular targets, high recurrence rate, and worse survival of patients with TNBCs compared to patients with ER+ve or HER+ve breast cancers (Lin *et al.*, 2012) make TNBCs a more aggressive pathology, posing a significant challenge to clinical management.

The incidence of TNBCs in Western populations is about 12.2%-13% of all breast cancer incidences (Hammond *et al.*, 2010; Plasilova *et al.*, 2016). However, in Indian populations, the incidence of TNBCs is 24%-31% of all breast cancer incidences (Kulkarni *et al.*, 2020), making the burden of TNBCs on the Indian population alarmingly higher. Therefore, there is an urgent need to understand the factors that govern the response and treatment outcomes of TNBCs.

1.2. The Tumour Microenvironment and Extracellular Matrix

Historically, cancer research has overarchingly focused on identifying and understanding dominant oncogenes' upregulation and tumour suppressor genes' loss of function, which imparts abnormal cell characteristics, making them cancerous. However, over the last couple of decades, the role of the tumour cells' immediate surroundings, as Stephen Paget calls it - the tumour microenvironment (TME), in cancer progression has increasingly come under investigation (Hanahan and Coussens, 2012; Laplane *et al.*, 2018). Studies have suggested that tumour-suppressive signals are provided by this microenvironment surrounding the tumour as long as the conditions for tissue homeostasis are controlled. Once tissue homeostasis is lost, however, the altered TME itself can be a tumour promoter (Bissell and Hines, 2011). Understanding how the TME

plays a role in tumour behaviour could provide avenues to control disease progression and malignancy.

The tumour microenvironment comprises multiple cellular components, including fibroblasts, endothelial cells, adipocytes and immunocytes, and a major acellular component: the extracellular matrix (ECM). The extracellular matrix forms a scaffold of proteins around the cells to support and organise them. However, it is much more than just a scaffold; it plays a role in multiple mechanisms. It provides signals that control cell proliferation and how likely they are to adhere together or migrate. The composition of the matrix is heterogeneous and varies in abundance and composition across patient samples as well as cancer type and can be remodelled depending on the context of the tissue, which induces a vast range of biochemical and biophysical changes affecting cell signalling, ECM stiffness, migration, and tumour progression (Oskarsson, 2013; Yuan *et al.*, 2023).

ECM remodelling is an essential, tightly regulated physiological process in developing and restoring homeostasis in wound repair. The matrix is remodelled through four dynamic processes: ECM deposition, posttranslational modifications of ECM proteins, proteolytic degradation, and force-mediated physical remodelling (Winkler *et al.*, 2020). Dysregulation of these processes is known to happen in pathologies like inflammatory diseases, tissue fibrosis and cancer. Recent studies have highlighted that aberration in the ECM mediated by the tumour has a role in metastasis (Martinez and Smith, 2021).

1.3. Collagen in the TME

The core ‘matrisome,’ defined by the ensemble of ECM and ECM-associated proteins, consists of glycoproteins, collagens and proteoglycans. Of these, collagens are the most abundant protein in the ECM and play an integral role in determining the mechanical properties of the tissue, like mammary density and stiffness. There is a marked increase in collagen deposition in tumour tissue compared to normal tissue (Acerbi *et al.*, 2015). An increase in deposition increases the mammary density of the tissue, and higher mammary density reflects more aggressive tumours (Northey *et al.*,

2020; Bodelon *et al.*, 2021). It is shown that a lot of the collagens of the intracellular matrix (Collagens I, III, V) and proteins involved in crosslinking (LOX family proteins) have more peptides in ECMs of metastatic tumours (Naba *et al.*, 2014). This increase in these ECM proteins is called fibrosis, and in the context of cancers, it is referred to as a *desmoplastic response*. Disruption in signalling pathways in tumour cells leads to cells in the tissue and tumour cells themselves depositing and remodelling collagen, causing fibrosis. Fibroblasts, particularly cancer-associated fibroblasts (CAFs) and sometimes tumour-associated macrophages (TAMs), along with tumour cells, are the main cell types involved in collagen deposition and bring about fibrosis (Xu *et al.*, 2019). Fibrosis is also seen to be a marker of poor prognosis in patients with invasive ductal carcinoma (Hasebe *et al.*, 2002). Thus, a collagen-rich ECM is a major contributor to tumour mechanics.

In addition to *how much* collagen, *how* collagen is deposited can also give us insights into the tumour's behaviour; the structure determines cellular processes and movements. With associated enzymes and other proteins, collagen alters matrix stiffness, alignment and density, substantially impacting tumour behaviour. In the case of breast cancer, it is observed that the collagen surrounding normal breast tissue is typically curly and smooth. However, it progressively thickens, linearises and stiffens parallel to tumour development, promoting metastasis (Fang *et al.*, 2014). These remodelled stiff collagens can be exploited by cancer cells, facilitating invasion. A dense deposition structure also determines how well immune cells infiltrate the microenvironment. In an in vivo study, T cells were seen to migrate preferentially along collagen fibres, suggesting the orientation of the fibres could control migration (Pruitt *et al.*, 2020; Rømer *et al.*, 2021).

Immune cells present in the tumour microenvironment, or *Tumour Infiltrating Lymphocytes* (TILs), as they are called, work as prognostic markers for the disease. TILs include lymphocytes and plasma cells (Salgado *et al.*, 2015; Angelico *et al.*, 2023). TNBCs, specifically, are more immunogenic than other subtypes and have a higher infiltration of lymphocytes than others. TNBCs also have a higher frequency of being positive for PD-L1 expression; given this, higher TIL scores serve as positive prognostic

markers in TNBCs, showing better survival and response to therapy (Kitano *et al.*, 2017). High lymphocyte recruitment is an adaptive mechanism of the host that combats tumour progression and metastasis. Previous work from our lab on studying the spatial distribution of TILs in different molecular subtypes of Indian patients was in line with the literature in this aspect (Vaid *et al.*, 2022).

1.4. Previous Studies on Collagen Deposition

Multiple studies have tried to define and quantify collagen deposition patterns for prognostication. As previously mentioned, the amount of collagen deposition or density gives insights into mammary density, a clinically relevant parameter. The amount of collagen deposition is proportional to mammary density, a negative prognostic marker in breast cancers (Northey *et al.*, 2020). Along with deposition density, *disorder* or *anisotropy* of collagen fibre deposition has also been quantified. Studies have shown that patients with a shorter disease recurrence-free period (or *disease-free survival - DFS* in literature) had more aligned collagen fibres in the tumour tissue (Conklin *et al.*, 2011; Li *et al.*, 2021). Spatially complex structures, like deposited collagen fibre networks, are analysed using fractal dimensions. This metric describes the amount of space and *self-similarity* of the structure and can differentiate between healthy and damaged tissue in the case of wound healing (Frisch *et al.*, 2012). Compared to normal tissue, another study reported that the fractal dimension was significantly lower in scar tissue two weeks after the wound, which increased by week 4, indicating wound remodelling (Tian *et al.*, 2021). Tumour-associated collagen signatures (TACS) have been defined by considering parameters like alignment with respect to the tumour cells and density. Broadly, TACS1 was marked by an increase in collagen deposition density around small tumours, TACS2 had stretched fibres along the tumour boundary as the tumour size increased, and TACS3 was identified in large tumours undergoing invasion where fibres were arranged normally to the tumour boundary (Provenzano *et al.*, 2006).

The “gold standard” for studies on fibrillar collagen has been using Second Harmonic Generation Microscopy (SHG) for imaging and quantifying collagen fibres. This laser scanning technique detects non-centrosymmetric objects and generates high-resolution

fibre-only images (Provenzano *et al.*, 2006; Conklin *et al.*, 2011). Despite the sub-micron resolution imaging, SHG microscopy requires specialised equipment and training along with computational demands, which do not facilitate the easy adoption of the method in clinical settings (Ouellette *et al.*, 2021); access remains a significant hurdle. Acquiring Whole Slide Images (WSIs) is now increasingly standard practice in pathology, and WSI analysis is thought to give more insight into intra-tumoural heterogeneity (Smith *et al.*, 2021). The analysis of SHG microscopy-based studies has also been on *regions of interest* (ROIs) and not WSIs.

1.5. Objectives and Scope

This study uses Masson Goldner's Trichrome Stain to visualise and quantify collagen in tissue sections of formalin-fixed paraffin-embedded patient samples. It is a cost-effective and accessible staining method to differentiate between collagenous and non-collagenous tissue. It stains differentially due to different molecular sizes and is used in diagnostics and accredited laboratories (Gruber, 1992; Ouellette *et al.*, 2021). Whole slide images of these stained tissue sections were acquired and used for analysis. The analysis pipeline will eventually be used to study these WSIs. We look at spatial and orientation metrics of collagen fibres, including parameters like density, fibre length, number, branch points, fractal dimension, and coherency in alignment. WSI files are large, and the pipeline development is done on ROIs. Due to a limitation in computation power, we are developing a pipeline that is Python-based but uses tools and algorithms from established image analysis platforms like FIJI and QuaPath, an open-source digital pathology platform and comparing the outputs from these with each other.

Since our pipeline is for brightfield images, we aim to validate the concordance of the outputs with outputs from an analysis of SHG images. We are also interested in seeing how the metrics differ as we move closer to the tumour from regions not in the immediate vicinity of the tumour. We hypothesise that quantifying the above-mentioned spatial metrics would help differentiate between these regions. We are interested in how quantified patterns in WISs and regions with different proximities to the tumour correlate

with clinical parameters acquired from the biobank. We hypothesise that the listed parameters in the tumour microenvironment would give us insights into tumour size, lymph node metastasis positivity, tumour grade, overall survival and disease-free survival of patients. We suspect differences in deposition patterns of collagen would impact infiltration of immune cells and, therefore, look for correlations between quantified metrics and TIL scores of the patient samples.

In summary, the objectives for this project are:

- To biochemically validate and visualise collagen deposition using Masson Goldner's trichrome stain in formalin-fixed paraffin-embedded samples of Indian patients with TNBC.

- Quantify spatial metrics of deposited collagen to define deposition patterns using brightfield images and check the concordance of outputs with SHG image metrics.

- To check if deposition patterns differ in regions near and away from the tumour.

- To analyse WSIs and see how they differ from an ROI-based analysis.

- To see how patterns in:
    - Regions near and far from the tumour,
    - WSIs

  correlate with clinical parameters like tumour size, lymph node metastasis positivity, grade, survival, recurrence, and TIL scores.

2. Materials and Methods

Resources	Source	Identifier
Chemicals and Reagents		
Paraffin	Sigma-Aldrich	
Xylene	Qualigens, Fischer Scientific	Cat#Q32297
Ethanol	From Bio-store, Department of Biology, IISER Pune	
Picric acid	Sigma-Aldrich	Cat#197378
Formaldehyde	Qualigens, Fischer Scientific (From Bio-store, Department of Biology, IISER Pune)	Cat#Q2400
Glacial acetic acid	Qualigens, Fischer Scientific (From Bio-store, Department of Biology, IISER Pune)	Cat#Q2105
Ferric Chloride	From Dr Pramod Pillai's Lab, Department of Chemistry, IISER Pune	
Hydrochloric acid	Qualigens, Fischer Scientific (From Bio-store, Department of Biology, IISER Pune)	
Hematoxylin	Sigma-Aldrich	Cat#H3136
Masson Goldner's Staining Kit	Sigma-Aldrich	Cat#100485
DPX Mountant	Qualigens, Fischer Scientific (From Bio-store, Department of Biology, IISER Pune	Cat#Q18404
Instruments and Consumables		
Microtome	Leica	R2265

Positively Charged Microscope Slides	PathnSitu	Cat#PS011, Cat#PS016
Mantra Quantitative Pathology Workstation	Akoya	
Brightfield Digital Scanner	OptraSCAN Pvt. Ltd	OS-15
Biological Samples		
FFPE Breast Cancer Patient Samples	Biobank, Prashanti Cancer Care Mission, Pune	Busheri *et al.*, 2021
Software and Algorithms		
QuPath	Bankhead *et al.*, 2017	https://github.com/qupath/qupath
FIJI	Schindelin *et al.*, 2012	https://imagej.net/downloads
TWOMBLI	Wershof *et al.*, 2021	https://github.com/wershofe/TWOMBLI
NB Workflow	Kanade *et al.*, 2024	
VSCode	Microsoft	https://code.visualstudio.com/
Amaranth Platform	Amaranth Medical Analytics Pvt. Ltd.	https://www.amaranth-insights.com/platform
GraphPad Prism		v.8.
IBM SPPS Statistics		v.21.0.0.0.

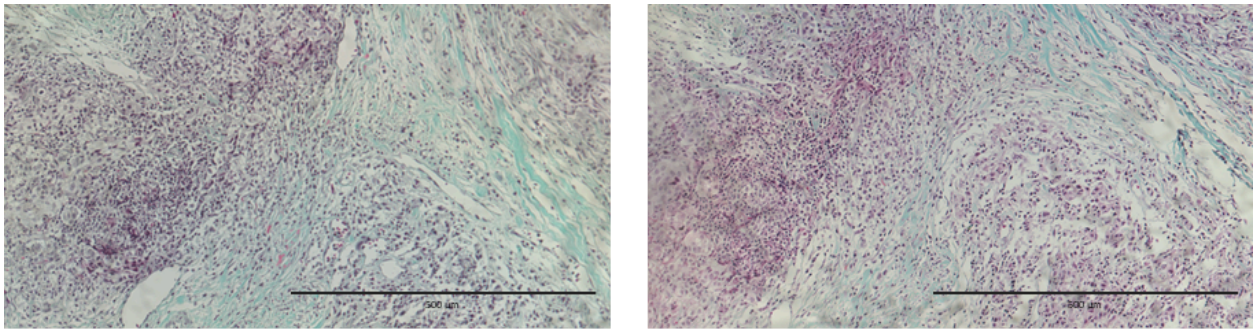
Table 1. Table of Resources and Reagents used.

2.1. Sample Preparation and Masson Goldner’s Trichrome Stain

The Formalin-fixed, paraffin-embedded breast cancer biopsy blocks and surgery blocks with the highest percentage of tumour tissue were selected based on the samples' previous Hematoxylin and Eosin Staining. A pathologist validated these. Selected FFPE tissue blocks were sectioned into three to five-micrometre (3 – 5 µm) thick sections using the Leica Microtome R2265 and adhered to positively charged microscope slides (size 1×3 inch; thickness 1-2 mm; PathnSitu catalogue numbers - PS011 and PS016 for the last 17 stained slides). Sectioning was performed by fellow lab members Pooja V, Gomathi V, Sampada G, Irene B, Sakshi D, Jisha J, Bhavesh O and Malavika S. Patient samples were received in IISER in batches. Samples in earlier batches (1-4) were sectioned from 2021 to 2023, coated in paraffin and stored at 4°C. Samples in batches five and onwards were sectioned from 2023 to 2025 and coated and stored if kept longer than two weeks; otherwise, they were used immediately the next day.

Tissue sections were kept at 62°C in a hybridisation oven overnight. Deparaffinisation was ensured by twice submerging the sections in xylene at room temperature for 15 minutes each, since paraffin is soluble in xylene. The xylene was gradually displaced with ethanol by rehydrating the sections in a descending ethanol dilution series (100%, 90%, 70%, and 50% ethanol in water) at room temperature, 3 minutes per dilution, since xylene is soluble in ethanol and ethanol. They were then transferred to water to replace the ethanol, thoroughly hydrating the tissue sections.

Sections were incubated in Bouin's solution (mordant) at 56°C for one hour. The composition of Bouin's solution is as follows: Saturated Picric Acid Solution (37.5 ml), 40% Formaldehyde (12.5 ml) and Glacial Acetic Acid (2.5 ml). This step makes the tissue sections more permeable to the dyes used in the staining and was included as a modification of the recommended protocol (of the kit).



a) b)

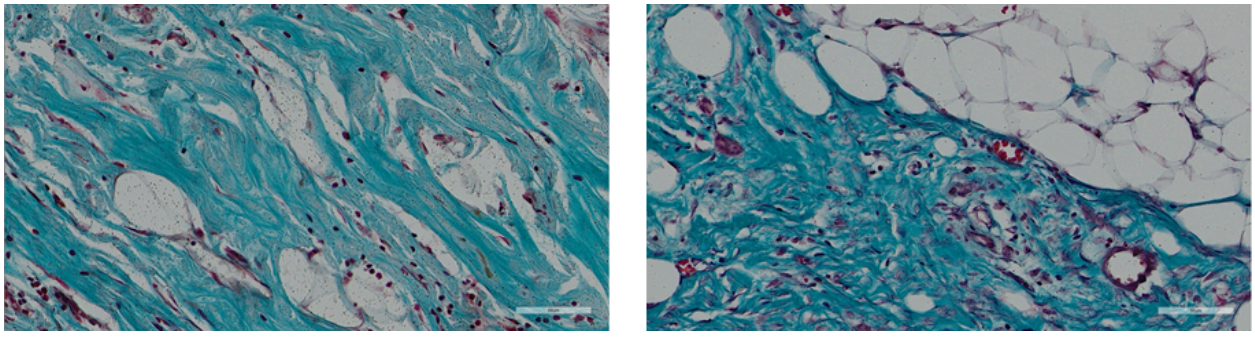
Fig. 1. The figures are roughly the same ROI of two separately stained sections of sample 786, TNBC subtype at 10Xe. a) Section stained without mordant step; b) Sectioned stained with mordant step (1 hour in Bouin's Solution at 56°C).

After a wash, they were incubated in Weigert's iron hematoxylin for ten minutes at room temperature. Weigert's Hematoxylin is a freshly prepared 1:1 solution of working solutions A and B. The composition of working solution A is as follows: Hematoxylin (0.5 gm; Sigma-Aldrich catalogue number: H3136-25G) and 95% Ethyl Alcohol (50 ml); and that of working solution B is 29% Ferric Chloride (4 ml; courtesy of Dr Pramod Pillai's Lab - Department of Chemistry, IISER Pune), Distilled Water (47.5 ml) and Concentrated HCl (0.5 ml). The iron hematoxylin stains the nuclei, giving them a dark brown/blue to black colour.

After washing off the hematoxylin with running tap water for 5 minutes, the sections were washed with 1% acetic acid (diluted from 10% (vol.), Product Identifier: Millipore - 221040) for 30 seconds and then incubated in an Azophloxine solution (Product Identifier: Millipore - 273742) for 7 minutes, a Tungstophosphoric Acid Orange G solution (Product Identifier: Millipore - 273744) for 6 minutes and then a Light Green SF solution (Product Identifier: Millipore - 273745) for 8 minutes. The tissue sections were washed with 1% acetic acid after each incubation period for 30 seconds. Due to its small molecular size, the Azophloxine binds to all acidophilic tissue components, giving it a red colour; the Orange G dye, being larger in size than Azophloxine, displaces the red dye selectively from collagen, as collagen tissue is more porous than non-collagenous tissue. Light Green SF then displaces Orange G and gives collagen a

green colour (Bancroft and Alan Stevens, 1982). Reagents were part of Masson Goldner's Staining Kit, Sigma Aldrich - Merck Millipore (catalogue no. 100485), received in 2021, and all solutions were stored at room temperature, unexposed to light.

Initially, the incubation time for the tissue in the Orange G solution and light green SF was 1 minute and 2 minutes, respectively. However, over time, the quality of the solution degraded (Expiration date - May 2025); hence, the incubation times were increased to the times specified above post-May 2024. In staining rounds from June 2024 to September 2024, some deposition was observed on the slides and the sections. It was later identified that this deposition was due to the precipitation of the Orange G solution (Figure 2). In the subsequent staining rounds, the precipitate was left to settle and undisturbed, and the suspension was used for staining. This fix yielded satisfactory staining without the deposition.



a) b)

Fig. 2. The figures are representative images of a) a section with Orange G solution precipitate on the slide - Sample 305, TNBC subtype at 40X; b) a section without Orange G solution precipitate on the slide, the precipitate was allowed to settle, and supernatant was used for staining - Sample 1022, TNBC subtype at 40X.

The stained sections underwent dehydration in an ascending ethanol dilution series (70%, 96% and 100% ethanol in water), 30 seconds per dilution. The sections were then submerged in 100% ethanol for 2 minutes. Some light green SF is washed off by ethanol; this was particularly enhanced post-May 2024. Therefore, the total time in ethanol was reduced to 1 minute from 3 and a half minutes and only 50% Ethanol was

used. After this, they were air-dried thoroughly for up to 10 minutes to ensure dehydration before incubating in xylene at room temperature for 5 minutes. The sections were then mounted in DPX, a non-aqueous mountant (Thermo Fischer Scientific - Qualigens, catalogue number Q18404). After placing the coverslip, the stained tissue sections were left to air dry overnight.

Vedant I standardised the staining protocol on a set of 27 samples encompassing both TNBC and non-TNBC subtypes from June 2023 to November 2023. Gomathi V and Sharon T helped with the initial trials.

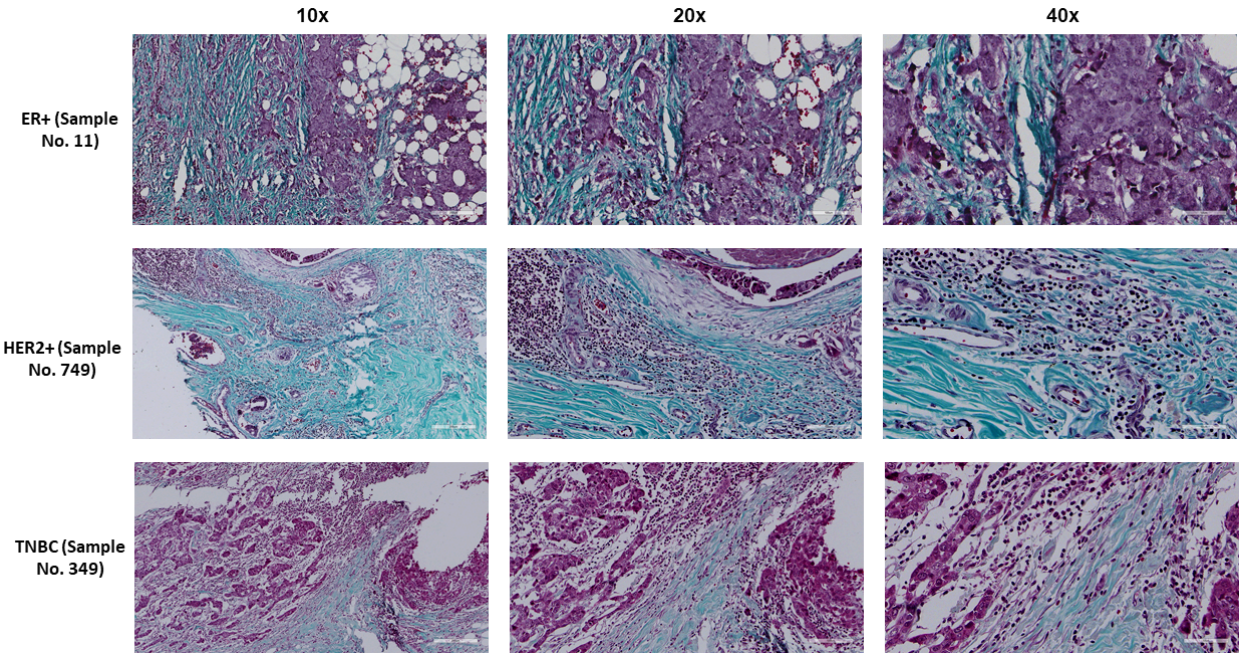


Fig. 3. The figures are representative images of breast cancer tissue of different subtypes from the standardisation set. The green-stained tissue represents deposited collagen, the red represents the non-collagen part (cells, necrotic debris), and the dark stain represents nuclei (of which TILs are most apparent).

2.2. Biobank Details, Clinical Data and Cohort Selection

This section describes the process of collecting patient samples and clinical data according to an audit of the clinic (Baptist *et al.*, 2017; Busheri *et al.*, 2021).

Patients who approach the Orchid Breast Health Clinic, Pune or associated clinics are registered in the patient database. Personal details, including data about medical and family history, are collected. Each patient undergoes a clinical breast examination by an onco-surgeon, who then advises radio-diagnosis if the clinical findings are abnormal. Women above 40 years of age and women aged less than 40 but are symptomatic undergo a Full-Field Digital Mammography (FFDM) with 3D Tomosynthesis. This technique uses X-rays to capture multiple views of the breast from different angles in a few seconds. The findings from the mammogram are reported in the Breast Imaging Reporting and Data System (BIRADS) Lexicon. Since mammographies may not be able to detect lesions in younger women due to predominantly dense breast tissue, women below the age of 40 and those who have inconclusive or abnormal mammogram findings undergo an Automated Breast Volume Scanner (ABVS) Ultrasound, since ultrasonography has been demonstrated to detect lesions missed by mammograms. ABVS Ultrasound uses Elastography, which defines the elasticity of the breast tissue.

A radiologist performs an ultrasound/mammogram-guided breast biopsy on the same day; this ensures that the tumour core is sampled and the tissue is fixed at the same time as the clinical examinations. These representative slivers of tissue are then sent to NABL-accredited pathology laboratories for histological investigation, including a Nottingham Score, a way to rate the aggressiveness of the breast cancer, and IHC to determine the molecular subtypes according to ASCO-CAP guidelines. All confirmed BC cases underwent appropriate surgery at a network hospital. The surgically excised healthy and breast cancer tissue is sent to an accredited pathology lab for further investigations.

Based on the clinical staging of the disease, standardised chemotherapy regimens were applied as per NCCN guidelines, either pre-surgery Neo-adjuvant chemotherapy to

decrease tumour size if more than 5 cm or post-surgery Adjuvant Chemotherapy. Radiotherapy is administered if there are tumour cells in multiple neighbouring lymph nodes.

The clinical data generated in this process includes demography (age, menopausal status), medical data (radiology findings, breast density, biopsy and surgery histopathological exam findings including grade and TNM staging, and chemotherapy and radiotherapy details) and survival status (death, recurrence, follow-up data). All data is stored after de-identifying the patient's name. All patient samples sent to the pathology lab are formalin-fixed and paraffin-embedded, which prevents the decomposition of the tissue and fixes it at the point when it is collected.
The de-identified patient data and FFPE samples were received from the PCCM biobank upon getting patients' consent in their local language and after appropriate ethical approval (details). After this, our lab at IISER carried out all processes.

We defined the cohort for the study on the following inclusion criteria: Patients should be diagnosed with Invasive Ductal Carcinoma of the TNBC subtype since we want to study characteristics of tissues with breast cancer; Patients should be Indian since we would want to study TNBC in Indian patients; Patients whose primary (untreated) breast tissue is available in the biobank, since we do not want to study tissue affected by NACT regimens; Patient samples which had enough tissue to be studied and could be sectioned or patients samples which had sections available. Ductal carcinoma in situ (DCIS) and other non-invasive ductal carcinoma (non-IDC) and recurrent patient samples were excluded, resulting in 139 shortlisted patient samples.

The criteria are tabularised below:

Criteria for inclusion	Criteria for Exclusion
- Patients with confirmed Invasive Ductal Carcinoma - Patients from India - Patients with Triple Negative Breast Cancer - Patients whose primary breast tissue sample is available in the biobank - Patient samples with enough tissue - Patient samples with sections available if tissue is inadequate	- Patients not from India - Patient with non-IDC breast disease – DCIS, benign, likewise - Patients who have undergone NACT whose diagnosis biopsies are unavailable. - Patients with only recurrent samples are available. - Samples without enough tissue - Samples with sections are unavailable in case of inadequate tissue.

Table 2. Table of inclusion and exclusion.

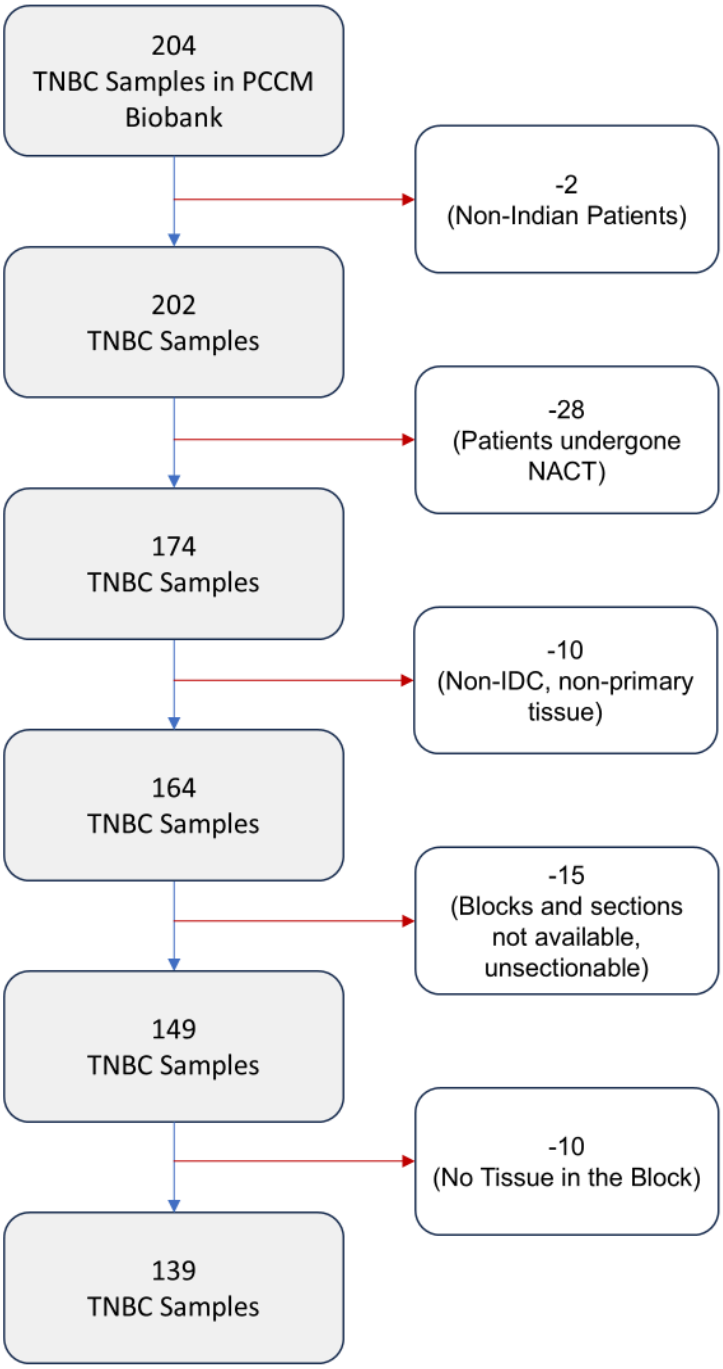


Fig. 4. Flowchart detailing the cohort selection process.

2.3. Whole Slide Brightfield Imaging

The stained tissue sections were sent to OptroScan, India, Pvt. Ltd. Brightfield Whole Slide Images were acquired at 40X magnification using the OS-15 brightfield digital scanner. The images were acquired in a JP2 format and later converted to TIFF for analysis.

Standardisation slides were sent for imaging in January 2024 as part of the OptraSCAN round 17. Images were received in February 2024. Forty-five slides from the cohort were stained from December 2023 to March 2024 and sent for imaging in March 2024 as part of round 18; they were received in October 2024. There were delays in this round due to logistical and colour balance issues. To rule out the suspicion that the staining caused the issue, we sent two sets of slides to OptraScan's once-pathologist, Dr Kolte; one set of slides was from the standardisation set (whose stain was satisfactory), and the other was that of 3 slides from the 45 slides sent. The oncopathologist rated the stain quality for both sets to be satisfactory and comparable. To show that the issue was that of the scanner's colour balance, we compared the mean RGB values of the backgrounds and saw that the later round had blue values relatively higher than red and green, giving the images a blue tint (Figure 5). The slides were then rescanned in August-September 2024.

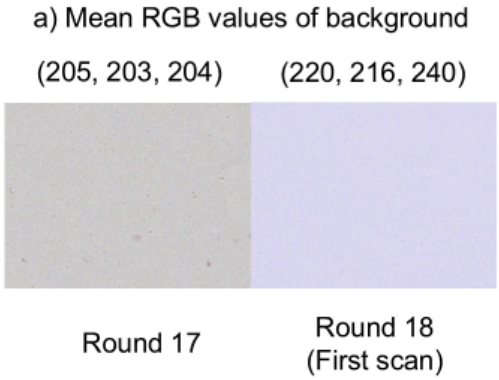


Fig 5. a) Representative backgrounds of tissue samples imaged in Round 17 (correct colour balance) and the first scan of Round 18 (offset colour balance);

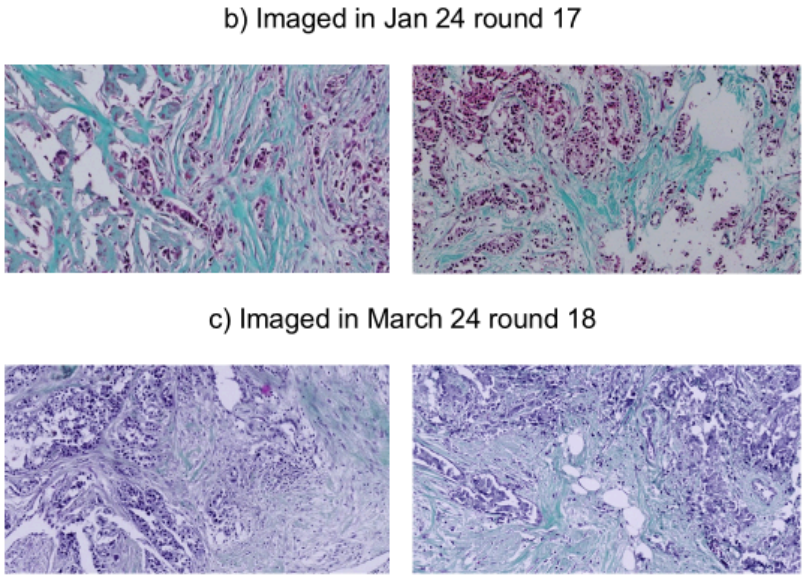


Fig. 5. b) Representative ROIs from WSIs of samples imaged in Round 17; c) Representative ROIs from WSIs of samples imaged in Round 18

Seventy-four slides from the cohort were stained from May 2024 to September 2024 as part of round 22, and 72 were sent for imaging in November 2024. Out of these, 12 slides had thick sections, possibly due to miscalibration of the microtome, and 18 had a considerable amount of orange G solution deposition, as described in the previous section. These factors caused the images to be out of focus. The faulty samples were resectioned, restrained and sent to be reimaged. The images were received in March 2025. A last set of 16 slides was sent for imaging in March 2025.

2.4. SHG microscopy as proof-of-principle

SHG imaging of three samples (benign sample, TNBC sample with low collagen deposition, and TNBC sample with high collagen deposition) was done on a non-optical microscopy (NLO) setup by Krishnendu D, from Dr Varun Raghunath's Lab, Electrical Communication Engineering (ECE), Indian Institute of Science (IISc), Bengaluru.

An optical parametric oscillator (APE Levante-IR) is pumped by a fixed wavelength femtosecond pulsed laser source (Fidelity-HP) centred at 1040 nm to generate a signal beam in the wavelength at 1600 nm with a 200-fs pulse width and 80 MHz repetition rate. The signal beam at 1600 nm is subsequently directed to a harmonic generator (APE HarmoniXX SHG) to generate a harmonic wavelength at 800 nm, which is used as the source of excitation for the measurements. The sample is mounted in an inverted position on an x-y stage of the inverted optical microscope (Olympus IX73). A water immersion objective (60x/1.2 NA) is used to focus the input laser onto the sample, and the same objective is used to collect the backwards-emitted SHG signal at 400 nm. The SHG signal is detected by a photomultiplier tube (Hamamatsu R3896) with a set of appropriate short-pass (550 nm cut-off) and band-pass (400 ± 40 nm) filters to reject any residual input source and minimise background signal. The sample images are acquired by scanning the laser beam with a pair of galvanometric mirrors (Thorlabs GVS002). From each sample, we acquire 25 images of size 110 × 110 μm2 with an overlap region of 10 μm and stitch to generate a larger FOV of size 500 × 500 μm2.

Three sections of each sample were sent for imaging, one HnE stained, one Trichrome stained and one unstained section to compare the quality of SHG output across stains. 500×500 μm^2 patches were acquired from the same regions of interest annotated by Dr Madhura Kulkarni for the HnE and Trichrome sections. This exercise is not possible for the unstained section since the section is not visible.

2.5. Preprocessing, Segmentation and Post-Processing

We are building an analysis pipeline based on tile analysis and plan to upgrade this to WSI analysis. Currently, processing power is a limitation for whole slide analysis; the image ranges from 700 MB to 4 GB in some cases. Hence, we analyse 500×500 μm^2 tiles instead of the whole slide image to create and test algorithms. The WSIs were tiled using the *openslide* and *tiffile* libraries on Python. Tiles were then arranged and visualised on a wide-screen desktop of the Mantra Pathology Workstation to match the WSI, and then the tumour core was categorised near and far from it. Tiles with regions within 500μm from the lesion boundary were classified as near the tumour core, and regions outside 500 μm from the lesion boundary were classified as far. If the tile has an immediate neighbour with a lesion boundary and a considerable amount of tissue falls within 500 μm, it is classified as within 500 μm or near the lesion. A tile with no intermediate neighbour with a lesion boundary can be classified as far from the tumour core (Figure 6).

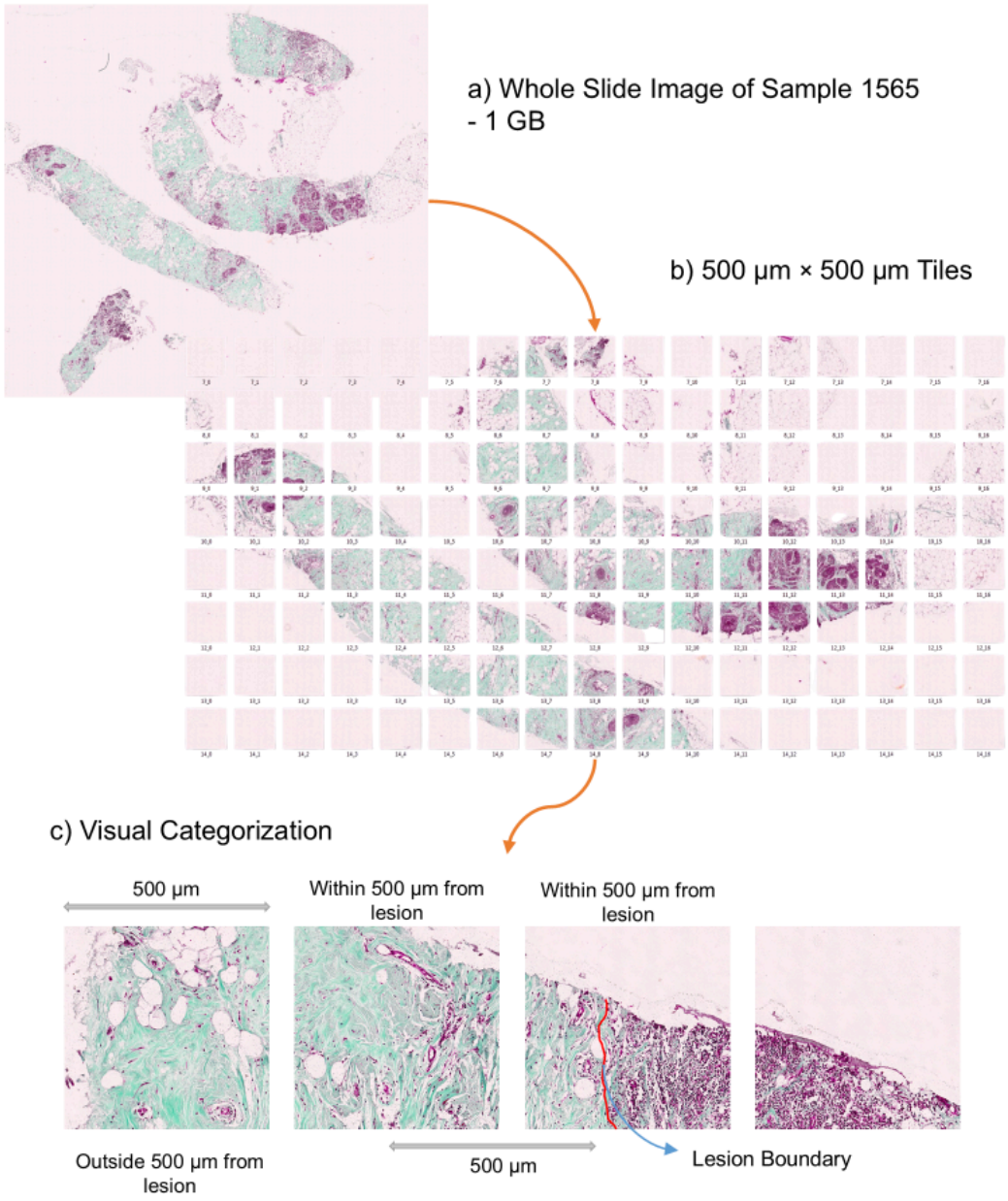


Fig. 6. a) The thumbnail of the WSI of sample 1565, which is 1GB, is b) WSI tiled into 500 µm × 500 µm ROIs; c) Depiction of visual classification method.

The analysis pipeline has five broad steps: Pre-processing, Segmentation, post-processing, quantification and eventual scoring. For segmentation, the selected regions undergo deconvolution based on an algorithm described by Ruifrok and

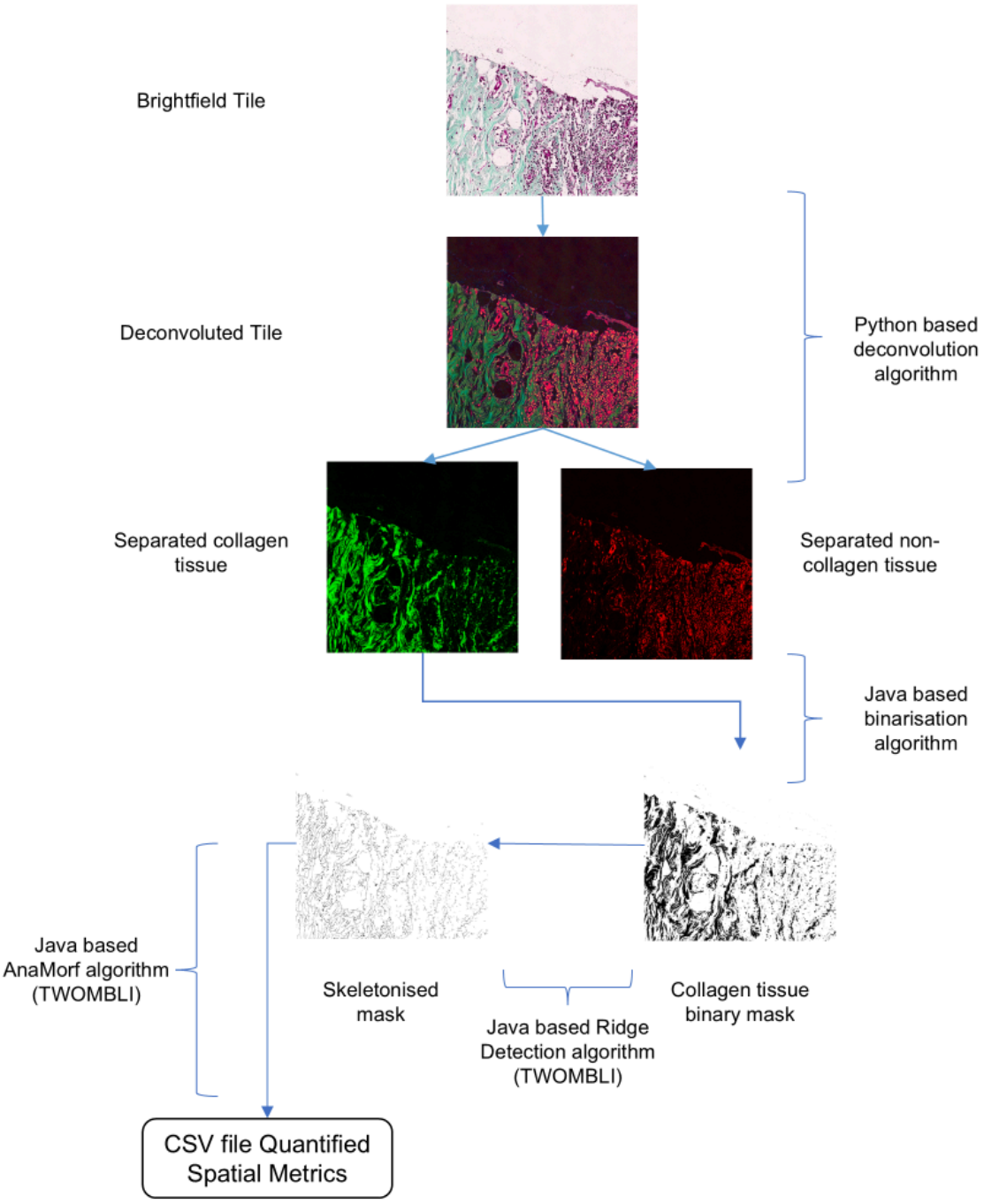
Johnston (Ruifrok and Johnston, 2001), which is detailed in the Appendix. The deconvolution algorithm was coded into Python with help from Manav S.

Once deconvolved, stains are separated, and a fibres-only image is obtained, which is further used for analysis. Noise due to the Orange G precipitate is removed using the outlier removal and then converted to 8-bit TIFF files on ImageJ. Outlier removal is used again to remove possible noise from the dark nuclei, which were not appropriately deconvolved.

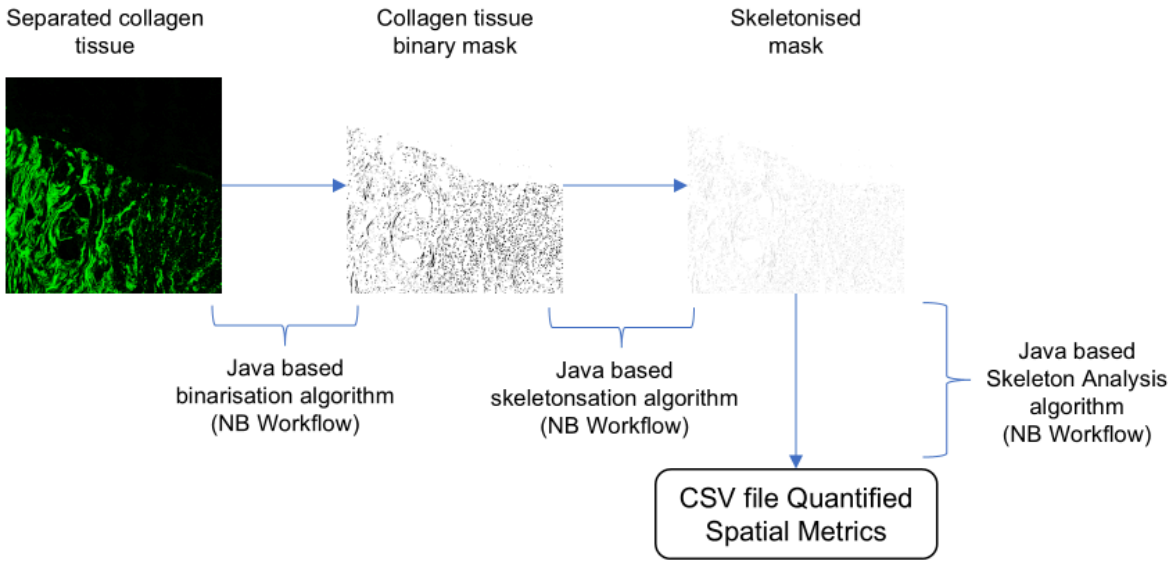
We were interested in comparing outputs from three workflows to see which gives satisfactory outputs. First, we use TWOMBLI, an ImageJ plugin written by Wershof et al., 2021 for tissue microarray analysis. This plugin relies on other established plugins to carry out its processes. It uses Ridge Detection to create a skeleton mask, Anamorf to quantify spatial metrics of the skeleton (except orientation), and OrientationJ to quantify alignment metrics (Wershof *et al.*, 2021). A directory is created with the post-processed tiles as inputs, and the plugin runs based on pre-existing parameters, which were arrived at after trial and error. Skeleton masks are saved as outputs along with a CSV file that quantifies spatial parameters (Figure 7a). The second is the ‘NB workflow’, a set of ImageJ macros written by Kanade S et al., 2024 from Dr Nagaraj Balasubramanian’s Adhesion Lab, Department of Biology, IISER Pune. This workflow was made to quantify changes in 3D collagen deposition at the cellular level and uses built-in processes like Skeletonise 2D/3D to create a skeleton mask and Analyse Skeleton 2D to quantify spatial metrics of the skeleton (Figure 7.b). Details about these plugins are in the appendix.

2.6. Disease-Free Survival Analysis

Disease-free survival (DFS) was calculated as time in months from the date of surgery till the date of recurrence or the last follow-up date in the case of no event. Kaplan-Meier survival plots for DFS for up to 5 years of follow-up were plotted using IBM SPSS Statistics v 21.0.0.0.


Brightfield Tile
Deconvoluted Tile
Python based deconvolution algorithm
Separated collagen tissue
Separated non-collagen tissue
Java based binarisation algorithm
Java based AnaMorf algorithm (TWOMBLI)
Skeletonised mask
Java based Ridge Detection algorithm (TWOMBLI)
Collagen tissue binary mask
CSV file Quantified Spatial Metrics

a)



b)

.

Fig. 7. a) Workflow showing segmentation, post-processing and quantification (using TWOMBLI). Parts of the pipeline are across languages, which will be brought into Python; b) The Workflow continues from segmentation, with post-processing and quantification done using the NB workflow.

2.6. Comparison of Methods of Analysis

Three samples were selected for comparison between the TWOMBLI workflow and NB workflow methods of analysis. One was benign, one TNBC sample had low collagen, and one had high collagen. On comparing the binarisation outputs of the two, the binarisation of the pipeline, TWOMBLI could capture more deposition areas compared to the NB pipeline, which had data loss in terms of area. However, the NB binarisation picked up on finer details of the deposition, which was reflected in the pipeline's skeleton output. The NB skeleton's details aligned with how collagen was seen in brightfield images, whereas the TWOMBLI skeleton lost these finer details and picked up trends in deposition patterns. The finer details would prove crucial while quantifying spatial parameters like fibre counts and junction points; on the other hand, overall trends give better quantification of parameters like the orientation of fibres.

We chose to move forward with the TWOMBLI tool since it gave the most kinds of outputs and, most importantly, orientation quantifications. The skeleton output of the NB output is currently not compatible with the TWOMBLI workflow quantification algorithm. However, we are still exploring the possibility of bettering the NB workflow output to

resemble the collagen deposition in Brightfield by minimising data loss of area and fine-tuning skeletonisation. Rajalaxmi B, from the Adhesion Lab, is working on this.

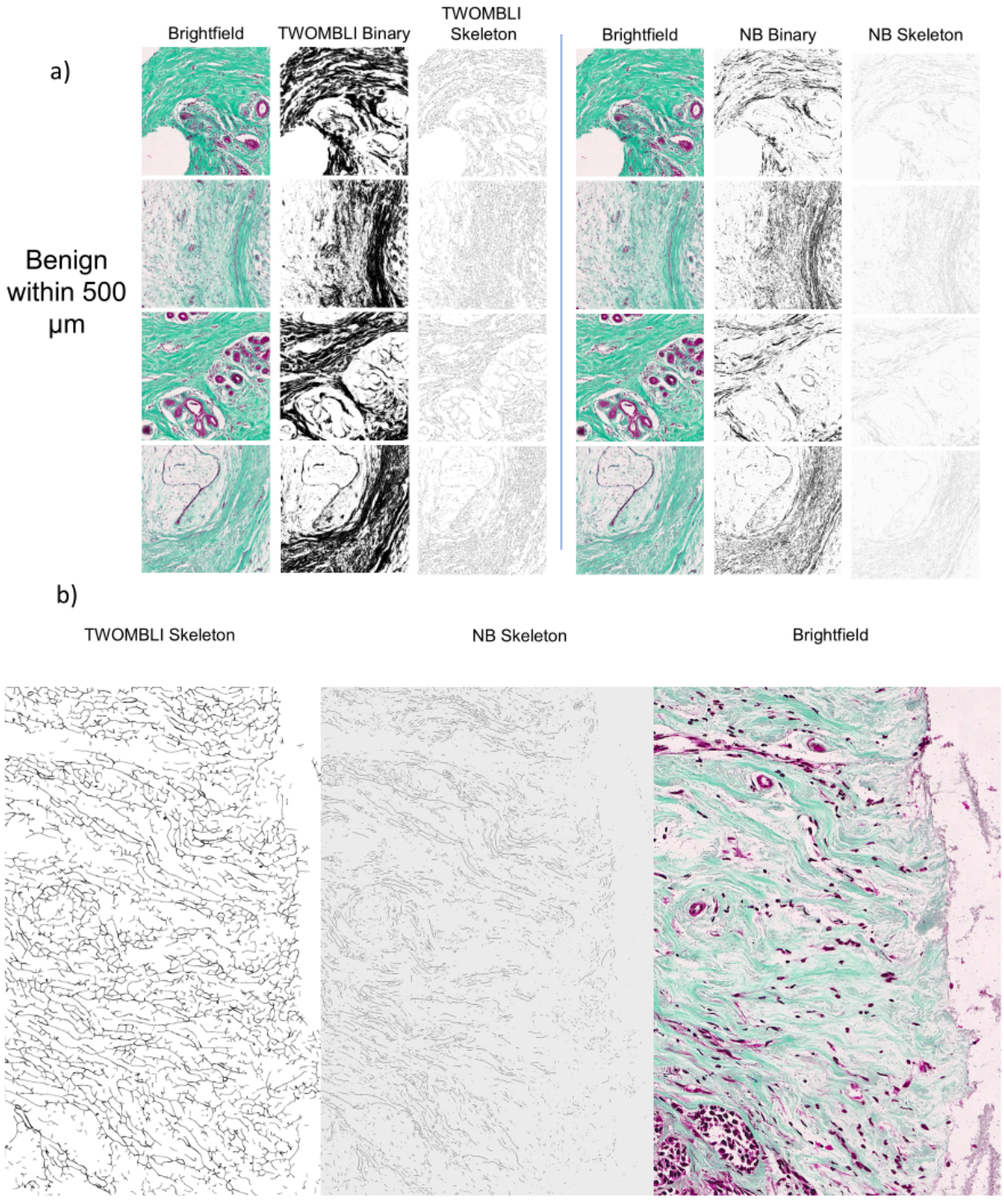


Fig. 8. a) Comparison between brightfield, binary and skeletons from both workflows. b) Close-up of skeleton outputs of a tile from a TNBC sample with high collagen from both workflows

2.7. Statistics

All statistics were done on GraphPad Prism v.8. The difference in spatial parameters between regions was compared using a Mann-Whitney test. IBM SPPS Statistics v.21.0.0.0. was used for the ROC and Kaplan-Meier Survival Analysis. The distribution of clinical characteristics and test variables of spatial parameters was analysed using a 2 × 2 Chi-square contingency test.

3. Results

3.1. SHG Signal from Trichrome section comparable to unstained section standard

We sent the physical stained sections of the same samples on which the comparison between TWOMBLI and NB workflows was performed to our collaborator, Dr Varun R, IISc, Bengaluru. A total of nice slides were sent, one Trichrome-stained section, one HnE-stained section and one unstained slide of each sample. Krishnendu D from his lab tried out exploratory imaging to see if trichrome-stained sections could be satisfactorily imaged, like the imaging of unstained sections, which is the standard. The SHG signal was strong from sections stained for collagen and was very similar to the signal from the unstained tissue section; thus, we concluded that the same trichrome section we have stained as part of my cohort could be used for a validation analysis of spatial parameter outputs from both SHG and Brightfield Images. We are awaiting SHG images of the three samples (benign, TNBC with High Collagen and TNBC with Low Collagen) to compare results across the two imaging techniques to see if they are concordant.

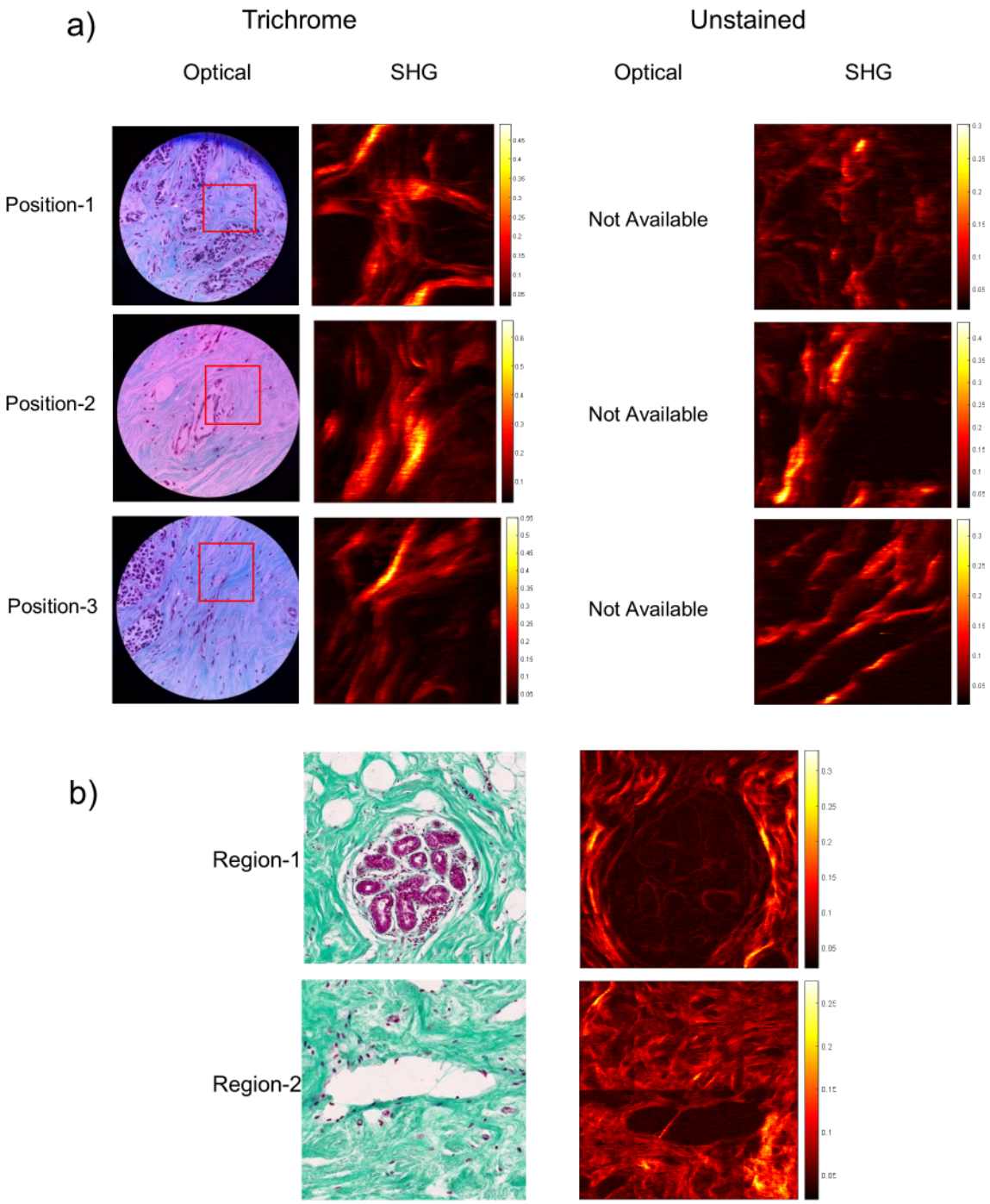


Fig. 11. a) Comparison of signal from SHG microscopy between Trichrome and Unstained sections of the TNBC sample with High Collagen; b) Representative SHG image field of view of the benign sample

3.2. Demographic and Clinical Characteristics of the Cohort

From the cohort of 139 Indian TNBC Patients, 46 primary tumour tissue samples were randomly selected and analysed. The demographic and clinical characteristics, as of March 2025, are summarised in Table 3. The mean age of the patients at the time of diagnosis is 50.5 years, ranging from 27 to 80 years, and was almost equally distributed in the below 50 (early onset of disease) and over 50 (late onset of disease). Patients diagnosed post-menopause (56.1%) were slightly more represented than pre-menopausal patients (43.9%). The cohort consists of a higher proportion of patients presenting a high-grade tumour (59.26%) than lower (40.9%) and has a slightly higher proportion of patients with a higher breast density (Type C-D, 53.7%) than lower (Type A-B, 46.3%). The proportion of bigger tumours (T2-T3, 79.55%) was higher than that of those with a lower tumour size (T1, 18.42%). 26.67% of the patients received NACT, and the proportion of responders and non-responders was equal. Parameters like node positivity and lymph vascular invasion are yet to be received from the biobank. Follow-up information of 45 patients was available, and the median follow-up period was 30.3 months. Seven patients in this randomly selected batch of patients showed recurrence of the disease, while two did not survive.

		All cohort
No. of patients		46
Age (n=43)	**(Mean ± S.D)**	50.512 ± 13.71
	Early (<50)	51.16% (22)
	Late (>= 50)	48.84% (21)
Menopausal status (n=41)	**Pre**	43.90% (18)
	Post	56.10% (23)
Grade (n=44)	**Low (I/II)**	40.90% (18)
	High (III)	59.09% (26)
Tumor size (cT) (n=38)	**T1**	18.42% (7)
	T2	76.92% (30)
	T3	2.63% (1)
Breast Density (n=41)	**Type A and B**	46.34% (19)
	Type C and D	53.66% (22)
pT (primary tissue, no NACT) (n=28)	**T0**	3.57% (1)
	T1	14.28% (4)
	T2	78.57% (22)
	T3	0.00% (0)
	T4	3.57% (1)
pN (primary tissue, no NACT) (n=28)	**Negative**	67.85% (19)
	Positive	2.14% (9)
NACT (n=45)	**No**	73.33% (33)
	Yes	26.67% (12)
PCR status after NACT (n= 12)	**pCR**	41.67% (5)
	RD	41.67% (5)
Survival outcomes	**No. followed-up**	45
	Median months	30.27
	Follow-up in Months (Range)	0.17-197.53
	# Recurred (local, distant)	7
	# Death	2

Table 3. Demographic Table, with available clinical characteristics, of the 46 patients whose tissues were analysed.

3.3. Spatial parameters differ for regions with different proximities to the tumour

For each sample processed and analysed, the image of the tissue was tiled into 500 μm × 500 μm-sized patches and tiles were categorised into within 500 μm of the tumour

core boundary and outside 500 μm from the tumour core boundary. Up to 16 tiles were selected from each category, provided those regions were available, as some samples with lower deposition of collagen regions far away from the tumour were absent.

These tiles were analysed using the TWOMBLI pipeline, and outputs were compared between regions. Along with the outputs of the tiles, those averaged over each sample were compared between regions. The estimate of the number of fibres normalised over the total tissue area was significantly less in regions closer to the tumour, in both the tile-wise and cross-sample comparisons. The Estimated Average length was slightly but significantly closer to the tumour, but this difference was lost in the cross-sample comparison. The number of branch points normalised to the total tissue area was also close to the tumour.

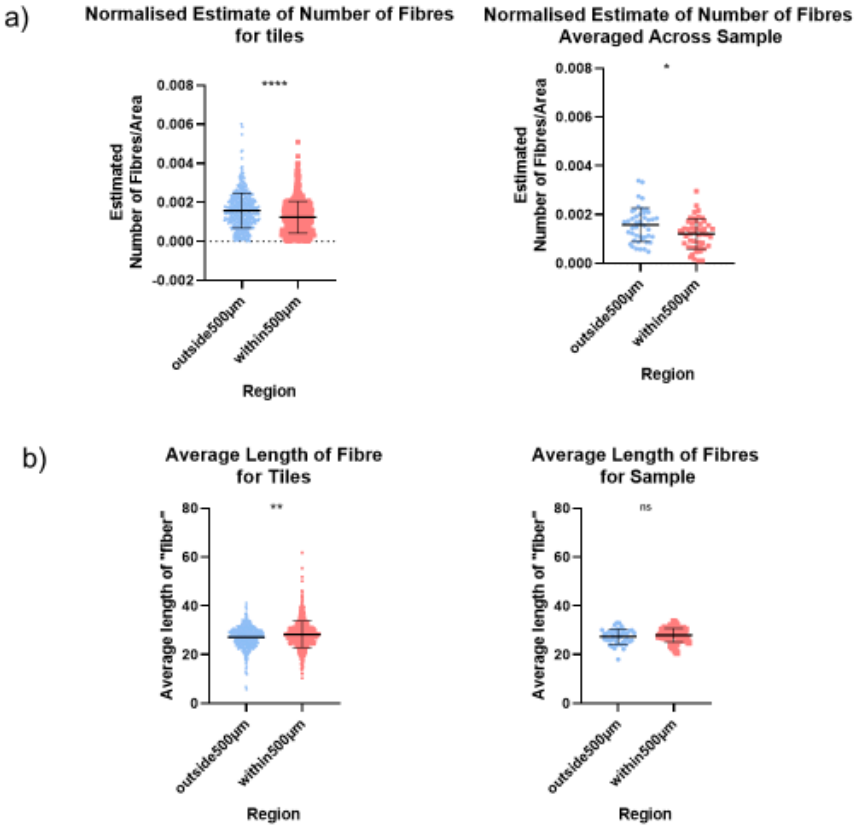


Fig. 9. a) Normalised Estimate of the Number of fibres tile-wise (p<0.0001) and cross-sample = 0.0146.0146); b) Average Length of Fibres tile-wise (p = 0.0038) and cross-sample (p = 0.3559)

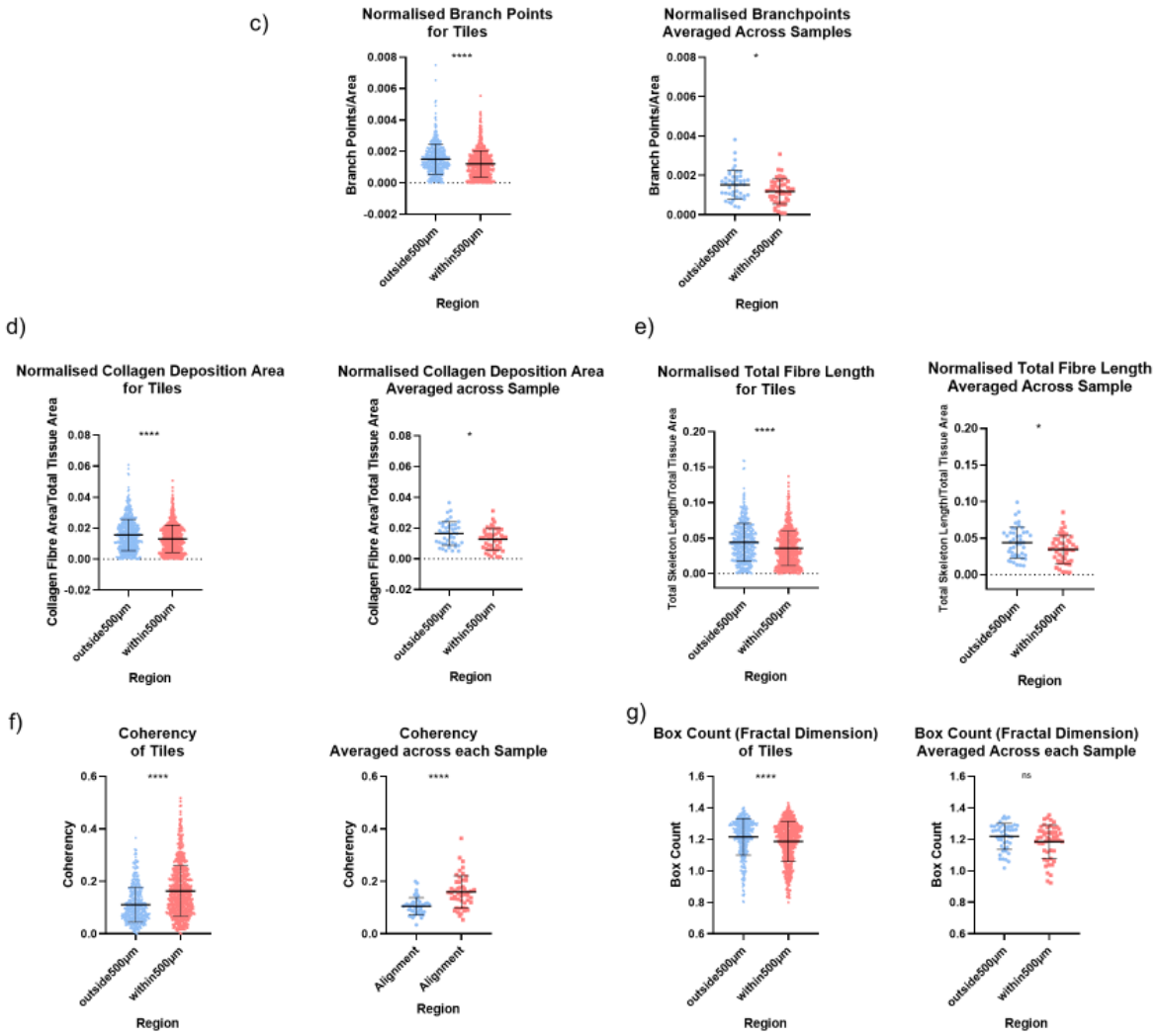


Fig.9. c) Normalised Number of Branch Points tile-wise ($p < 0.0001$) and crass-sample ($p = 0.0239$); d) Normalised Collagen Deposition Fibre Area tiles-wise ($p < 0.0001$) and cross-sample ($p = 0.0250$); e) Normalised Total Length of Fibre Skeleton tile-wise ($p < 0.0001$) and cross-sample ($p = 0.0379$); f) Coherency of Alignment tile-wise ($p < 0.0001$) and cross-sample ($p < 0.0001$); Box Count (Fractal Dimension) tile-wise ($p < 0.0001$) and cross-sample ($p = 0.0988$). The Mann-Whitney test was performed to check for significance.

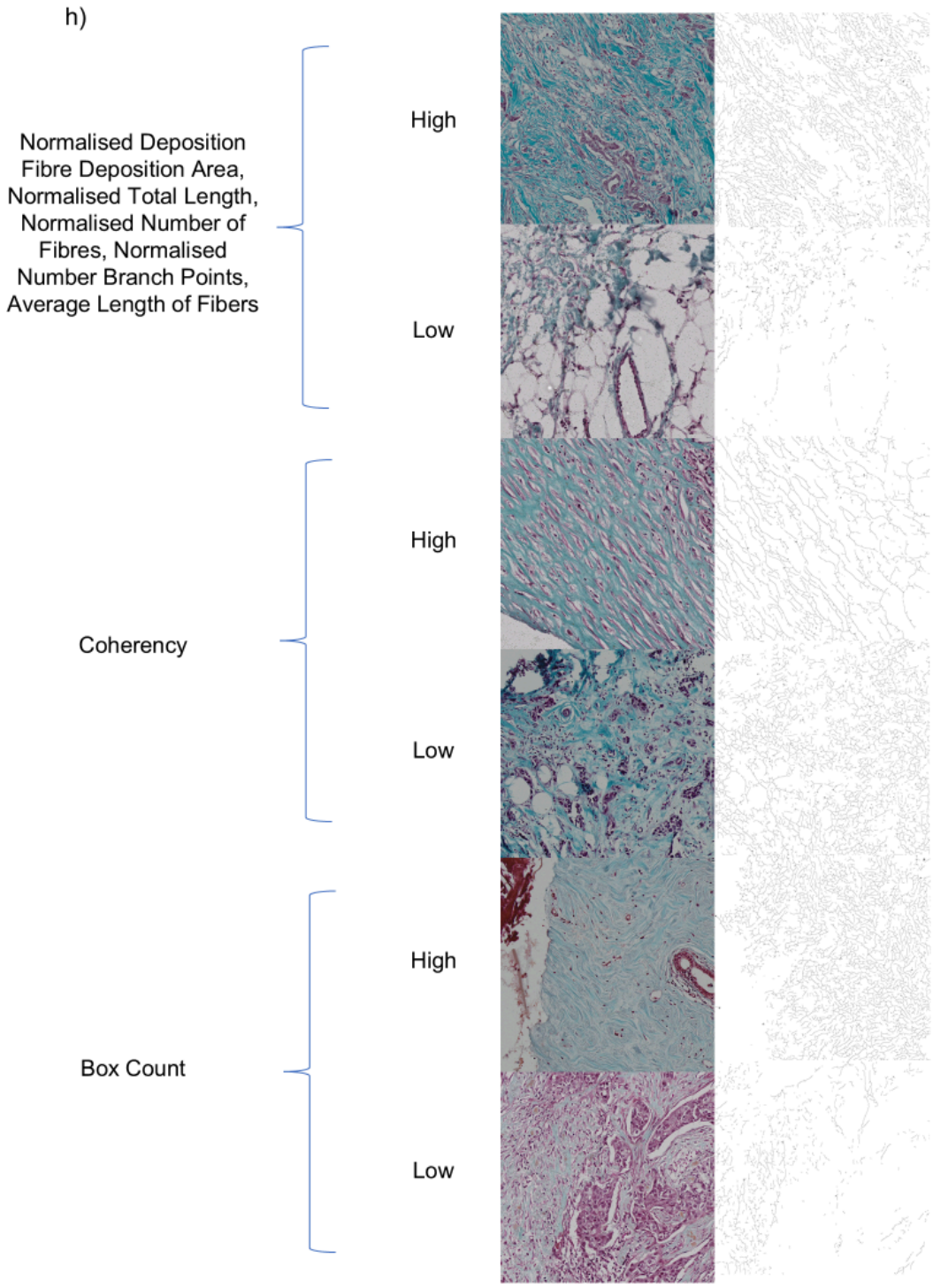


Fig.9. h) Representative tiles for different levels of each quantified parameter.

We also looked at the area and total length of the fibre network normalised to the total tissue area; both seemed to decrease in regions within 500 μm from the tumour

boundary in the tile-wise and cross-sample comparisons (Figure 10.d and e). Coherency and box count (Fractal Dimension) reflect the *nature* of the collagen deposition. Coherency (of alignment) measures how aligned the fibres are to each other. Fibres seem to be significantly more aligned with each other in regions near the tumour than in regions far away. The fractal dimension, a complexity measure, is lower in regions near the tumour; however, the cross-sample comparison loses its significance. These results confirm that spatial metrics of collagen deposition differ in regions with different proximity to the tumour.

3.4. Parameter reflecting the nature of deposition in a specific region better discriminator than just the amount of deposition

For now, we have looked at how two spatial parameters discriminate in disease-free survivorship. The variables we were interested in were: collagen fibre deposition area, based on the skeleton map of the deconvoluted image, and coherency outputs of tiles averaged over the sample; then separated outputs regions near and far from the tumour core boundary, to see if measuring these regions gives us better separation; and finally, the ratio of the outputs for regions far from the tumour to near, to see if the difference in the spatial parameter separates the outcomes better.

Receiver Operating Characteristic Analysis was performed on these variables, with the state variable being the status of the recurrence of the disease. The areas under the curve have been summarised in Table 4. The coherency of collagen fibres in regions away from the tumour was surprisingly the best discriminator in disease-free survivorship compared to other variables, better than any deposition area variable representing the deposition amount. The second is the ratio of the coherencies of regions far from the tumour to regions near, followed by the area of the collagen fibre deposition across the region. Thus, separating the regions of the sample based on the proximity to the tumour while studying the characteristics of collagen deposition can provide more insightful information.

	Collagen Fibre Deposition Area	Coherency
Test Variable	Area Under Curve	
Across Regions	0.560 ± 0.11	0.324 ± 0.12
Within 500 um	0.4886 ± 0.14	0.320 ±0.12
Outside 500 um	0.537 ± 11	0.621 ± 0.12
Outside-to-Within Ratio	0.444 ± 0.14	0.587 ± 0.14

Table 4. Summary of the ROC Analysis

3.5. Disease-Free Survival Analysis and Comparison between Spatial Parameters and Clinical Characteristics

Cutoffs for high variable values were set using these ROC curves. Each variable gave two plausible cutoffs, one cutoff value giving at least 0.5 Sensitivity, and the other where the 1-Specificity was less than 0.5. Preference was given to the cutoff value where the Sensitivity was at least 0.5 as long as the 1 - Specificity was below 0.5. However, the final selection was made based on how good the separation is in the disease-free survival curves (Figure 10). Bases of cutoffs are summarised as,

	Collagen Fibre Deposition Area	Coherency
Test Variable	Cutoff Basis	
Across Regions	Sensitivity	Specificity
Within 500 um	Sensitivity	Sensitivity
Outside 500 um	Sensitivity	Sensitivity
Outside-to-Within Ratio	Specificity	Sensitivity

Table 5. Summary of the Bases of Cutoffs set

ROC for Collagen Deposition Area Across Regions

a)

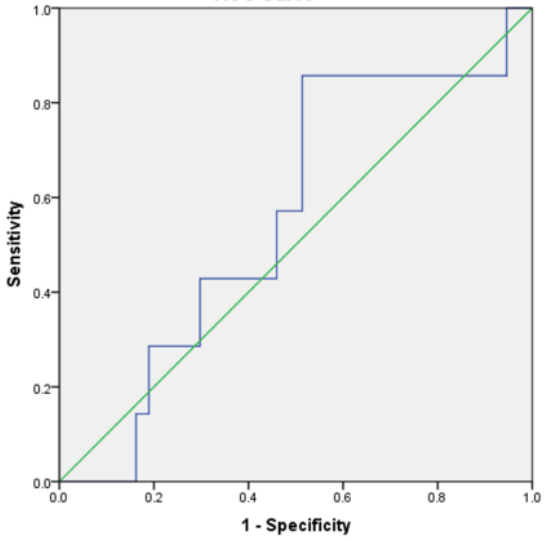
Case Processing Summary

RecurrenceStatusCoded	Valid N (listwise)
Positive[a]	7
Negative	37
Missing	2

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is 1.00.

ROC Curve



Area Under the Curve

Test Result Variable(s): CollagenAreaTotalTissueArea

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.560	.110	.619	.344	.776

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Specificity Cutoff – 0.138

Sensitivity Cutoff – 0.1445

Coordinates of the Curve

Test Result Variable(s): CollagenAreaTotalTis

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
.0000	1.000	1.000
.0041	1.000	.973
.0048	1.000	.946
.0054	.857	.946
.0060	.857	.919
.0061	.857	.892
.0063	.857	.865
.0065	.857	.838
.0071	.857	.811
.0080	.857	.784
.0087	.857	.757
.0091	.857	.730
.0096	.857	.703
.0101	.857	.676
.0106	.857	.649
.0116	.857	.622
.0123	.857	.595
.0126	.857	.568
.0128	.857	.541
.0132	.857	.514
.0135	.714	.514
.0137	.571	.514
.0139	.571	.486
.0143	.571	.459
.0146	.429	.459
.0150	.429	.432
.0153	.429	.405
.0159	.429	.378
.0169	.429	.351
.0174	.429	.324
.0181	.429	.297
.0189	.286	.297
.0195	.286	.270
.0201	.286	.243
.0204	.286	.216
.0206	.286	.189
.0209	.143	.189
.0211	.143	.162
.0219	.000	.162
.0229	.000	.135
.0233	.000	.108
.0242	.000	.081
.0255	.000	.054
.0265	.000	.027
1.0000	.000	.000

a. The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

Fig. 10. a) ROC analysis output for test variable - Collagen Fibre Deposition Area

ROC for Coherency Outside 500 µm

b)

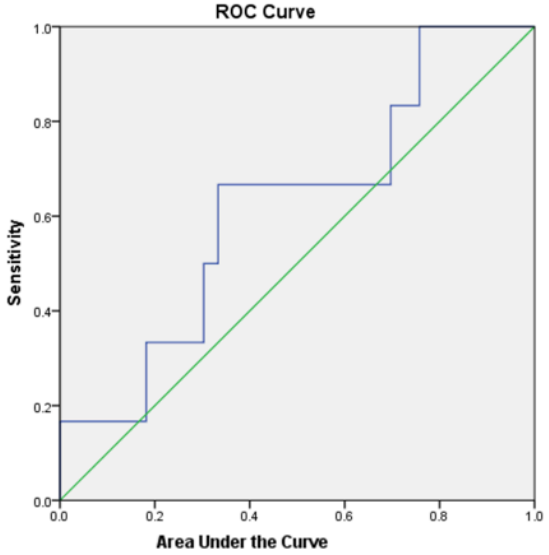
Case Processing Summary

RecurrenceStatusCoded	Valid N (listwise)
Positive[a]	6
Negative	33
Missing	7

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is 1.00.

ROC Curve



Test Result Variable(s): CoherencyOutside500um

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.621	.123	.350	.381	.862

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Coordinates of the Curve

Test Result Variable(s): CoherencyOutside50

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
.0000	1.000	1.000
.0614	1.000	.970
.0640	1.000	.939
.0669	1.000	.909
.0688	1.000	.879
.0709	1.000	.848
.0729	1.000	.818
.0754	1.000	.788
.0777	1.000	.758
.0808	.833	.758
.0836	.833	.727
.0851	.833	.697
.0870	.667	.697
.0876	.667	.667
.0899	.667	.636
.0921	.667	.606
.0932	.667	.576
.0973	.667	.545
.1004	.667	.515
.1009	.667	.485
.1037	.667	.455
.1063	.667	.424
.1081	.667	.394
.1104	.667	.364
.1115	.667	.333
.1125	.500	.333
.1141	.500	.303
.1155	.333	.303
.1164	.333	.273
.1187	.333	.242
.1212	.333	.212
.1244	.333	.182
.1314	.167	.182
.1407	.167	.152
.1472	.167	.121
.1528	.167	.091
.1619	.167	.061
.1800	.167	.030
.1966	.167	.000
1.0000	.000	.000

a. The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

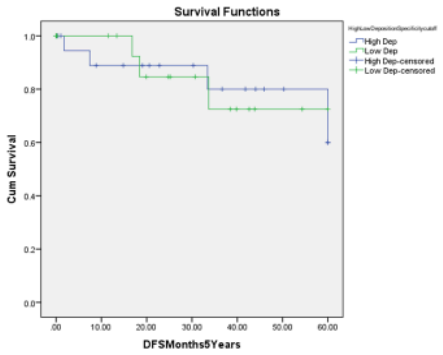
Specificity Cutoff – 0.10075

Sensitivity Cutoff – 0.1141

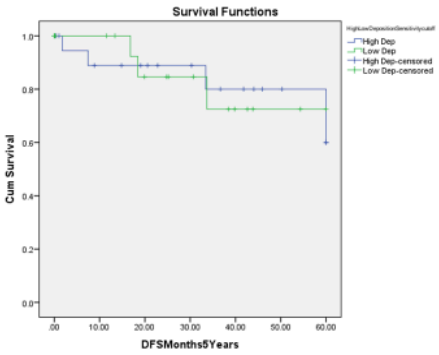
Fig. 10. b) ROC analysis output for test variable - Coherency of Fibre Alignment far from the tumour.

c) DFS plot of Collagen Fibres Deposition Area Across Regions

Specificity Cut-off Curve



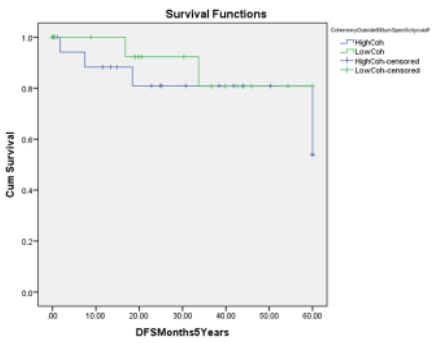
Sensitivity Cut-off Curve



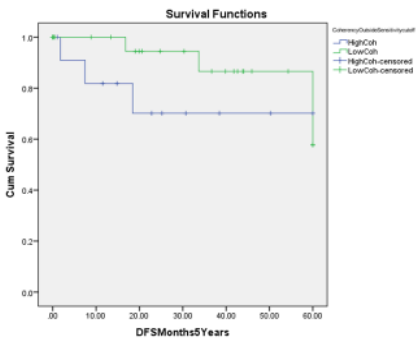
Specificity Cutoff – 0.138
Sensitivity Cutoff – 0.1445

d) DFS of Coherency outside 500µm

Specificity Cut-off Curve



Sensitivity Cut-off Curve



Specificity Cutoff – 0.10075
Sensitivity Cutoff – 0.1141

Fig. 10. c) Disease-free Survival plot based on Collagen Fibre Deposition Area cutoff, blue lines represent high deposition and green low (Chi square = 0.004, p = 0.947 for specificity cutoff; Chi square = 0.004, p = 0.947 for sensitivity cutoff); d) Disease-free Survival plot based on Coherency far from the tumour, blue lines represent high coherency and green low (Chi square = 0.650, p = 0.723 for specificity cutoff; Chi square = 1.679, p = 0.432 for sensitivity cutoff)

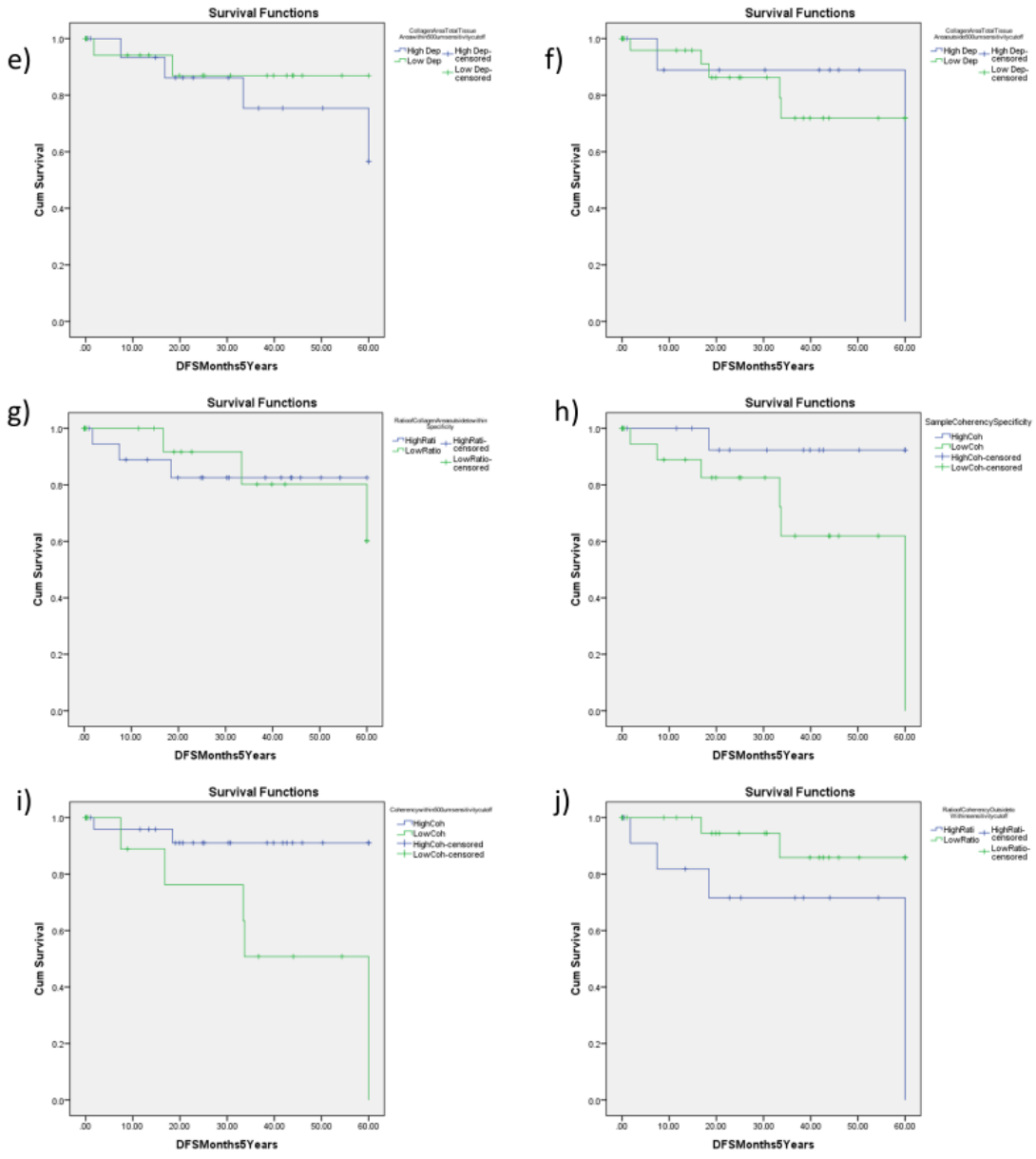


Fig. 10. Disease-free Survival plots based on e) Collagen Fibre Deposition Area - within 500 um blue representing high deposition area, Chi square = 2.619, p = 0.270), f) Collagen Fibre Deposition Area - outside 500 um (blue line representing high deposition area, Chi square = 0.009, p = 0.925), g) Ratio of Collagen Fibre Deposition Area - outside to inside (blue line representing high ratio, Chi square = 2.234, p = 0.327), h) Coherency of Fibres across regions (blue line representing high coherency, Chi square = 4.735, p = 0.030), i) Coherency of Fibres - within 500 um (blue line representing high coherency, Chi square = 6.995, p = 0.008), j) Ratio of Coherency of Fibres - outside to within (blue line representing high ratio, Chi Square = 5.600, p = 0.061).

ROC for Collagen Deposition Area within 500µm

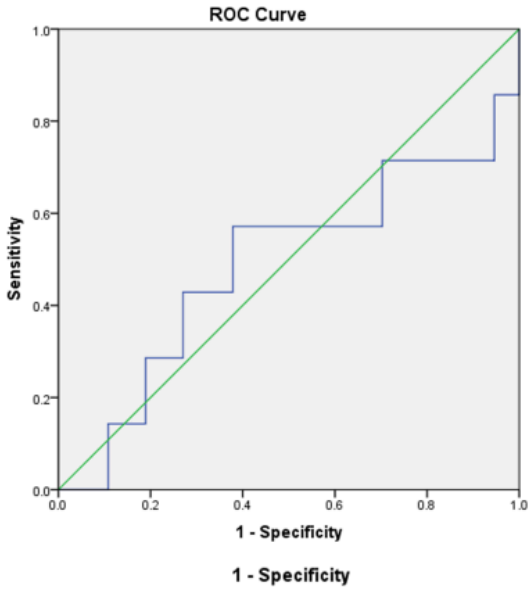
k)

Case Processing Summary

RecurrenceStatusCoded	Valid N (listwise)
Positive[a]	7
Negative	37
Missing	2

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is 1.00.



1 - Specificity

Area Under the Curve

Test Result Variable(s): CollagenAreaTotalTissueAreawithin500um

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.486	.136	.911	.221	.752

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Coordinates of the Curve

: CollagenAreaTotalTissueAreawithin500um

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.0006	.857	1.000
.0013	.857	.973
.0020	.857	.946
.0031	.714	.946
.0037	.714	.919
.0049	.714	.892
.0060	.714	.865
.0062	.714	.838
.0062	.714	.811
.0066	.714	.784
.0070	.714	.757
.0074	.714	.730
.0079	.714	.703
.0083	.571	.703
.0087	.571	.676
.0090	.571	.649
.0096	.571	.622
.0104	.571	.595
.0114	.571	.568
.0122	.571	.541
.0124	.571	.514
.0129	.571	.486
.0135	.571	.459
.0139	.571	.432
.0140	.571	.405
.0145	.571	.378
.0151	.429	.378
.0153	.429	.351
.0156	.429	.324
.0159	.429	.297
.0161	.429	.270
.0167	.286	.270
.0174	.286	.243
.0181	.286	.216
.0190	.286	.189
.0194	.143	.189
.0197	.143	.162
.0202	.143	.135
.0209	.143	.108
.0217	.000	.108
.0231	.000	.081
.0249	.000	.054
.0285	.000	.027
1.0312	.000	.000

a. The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

Specificity Cutoff – 0.0126

Sensitivity Cutoff – 0.0148

Fig. 10. k) ROC analysis output for test variable - Collagen Deposition Area near the tumour.

ROC for Collagen Deposition Area outiside500µm

I)

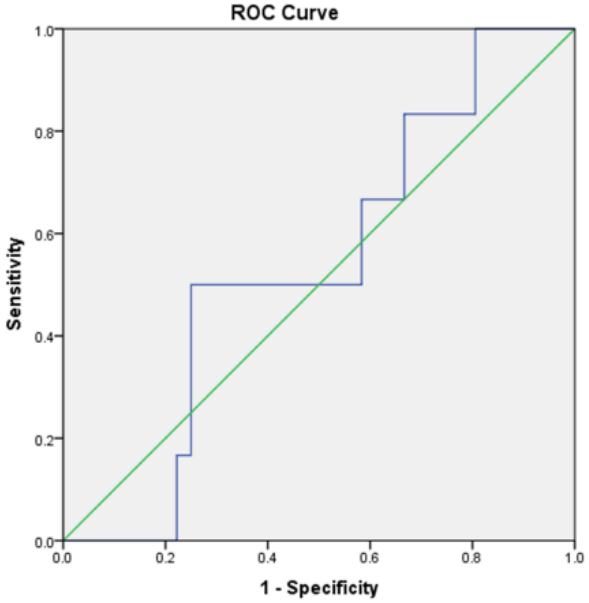
Case Processing Summary

RecurrenceStatusCoded	Valid N (listwise)
Positive[a]	6
Negative	36
Missing	4

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is 1.00.

ROC Curve



Area Under the Curve

Test Result Variable(s): CollagenAreaTotalTissueAreaoutside500um

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.537	.113	.774	.316	.758

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Specificity Cutoff – 0.0163

Sensitivity Cutoff – 0.0205

Coordinates of the Curve

Test Result Variable(s): CollagenAreaTotalTis

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.0025	1.000	.917
.0051	1.000	.889
.0053	1.000	.861
.0059	1.000	.833
.0067	1.000	.806
.0070	.833	.806
.0075	.833	.778
.0083	.833	.750
.0088	.833	.722
.0094	.833	.694
.0103	.833	.667
.0107	.667	.667
.0108	.667	.639
.0111	.667	.611
.0124	.667	.583
.0137	.500	.583
.0139	.500	.556
.0147	.500	.528
.0159	.500	.500
.0167	.500	.472
.0170	.500	.444
.0170	.500	.417
.0173	.500	.389
.0180	.500	.361
.0188	.500	.333
.0195	.500	.306
.0200	.500	.278
.0203	.500	.250
.0208	.333	.250
.0215	.167	.250
.0221	.167	.222
.0228	.000	.222
.0237	.000	.194
.0244	.000	.167
.0250	.000	.139
.0264	.000	.111
.0287	.000	.083
.0308	.000	.056
.0340	.000	.028
1.0366	.000	.000

a. The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

Fig. 10. I) ROC analysis output for test variable - Collagen Deposition Area far from the tumour

ROC for Ratio of Collagen Deposition Area Outside to Inside

m)

Case Processing Summary

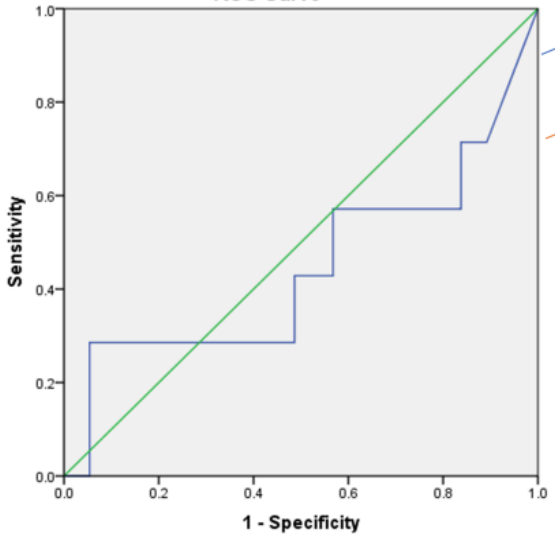
RecurrenceStatusCoded[a]	Valid N (listwise)
Positive[b]	7
Negative	37
Missing	2

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The test result variable(s): RatioofCollagenAreaoutsidetowithin has at least one tie between the positive actual state group and the negative actual state group.

b. The positive actual state is 1.00.

ROC Curve



Diagonal segments are produced by ties.

Area Under the Curve

Test Result Variable(s): RatioofCollagenAreaoutsidetowithin

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.444	.141	.642	.168	.720

The test result variable(s): RatioofCollagenAreaoutsidetowithin has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Specificity Cutoff – 1.2659

Sensitivity Cutoff – 1.1699

Coordinates of the Curve

Test Result Variable(s): RatioofCollagenArea

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.2944	.714	.892
.6437	.714	.865
.7056	.714	.838
.7487	.571	.838
.7926	.571	.811
.8153	.571	.784
.8924	.571	.757
.9646	.571	.730
.9810	.571	.703
.9972	.571	.676
1.0320	.571	.649
1.0648	.571	.622
1.0930	.571	.595
1.1403	.571	.568
1.1995	.429	.568
1.2344	.429	.541
1.2465	.429	.514
1.2853	.429	.486
1.3354	.286	.486
1.3584	.286	.459
1.3721	.286	.432
1.4490	.286	.405
1.5167	.286	.378
1.5744	.286	.351
1.6483	.286	.324
1.6748	.286	.297
1.6961	.286	.270
1.7662	.286	.243
1.8552	.286	.216
1.9069	.286	.189
1.9328	.286	.162
2.0285	.286	.135
2.1939	.286	.108
2.3595	.286	.081
2.4825	.286	.054
2.5663	.143	.054
3.0748	.000	.054
6.1532	.000	.027
9.7732	.000	.000

The test result variable(s): RatioofCollagenAreaoutsidetowithin has at least one tie between the positive actual state group and the negative actual state group.

a. The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

Fig. 10. m) ROC analysis output for test variable - Ratio of Deposition Area - outside to within.

ROC for Coherency Across Sample

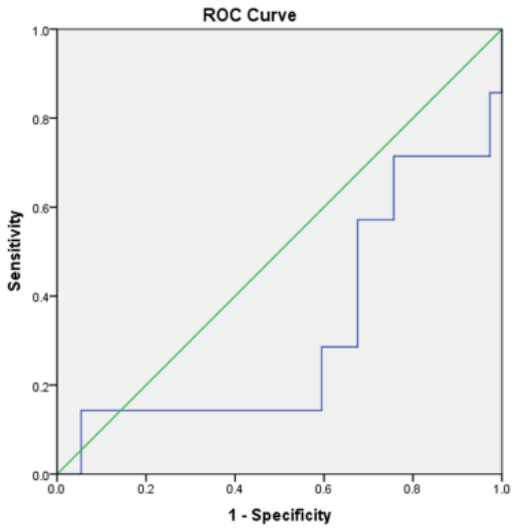
n)

Case Processing Summary

RecurrenceStatusCoded	Valid N (listwise)
Positive[a]	7
Negative	37
Missing	2

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is 1.00.



Area Under the Curve

Test Result Variable(s): SampleCoherency

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.324	.119	.144	.091	.558

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Coordinates of the Curve

Test Result Variable(s): SampleCoherency

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
.0000	1.000	1.000
.0655	.857	1.000
.0781	.857	.973
.0786	.714	.973
.0816	.714	.946
.0847	.714	.919
.0892	.714	.892
.0934	.714	.865
.0943	.714	.838
.0966	.714	.811
.0992	.714	.784
.1005	.714	.757
.1016	.571	.757
.1103	.571	.730
.1185	.571	.703
.1196	.571	.676
.1210	.429	.676
.1224	.286	.676
.1243	.286	.649
.1262	.286	.622
.1272	.286	.595
.1273	.143	.595
.1301	.143	.568
.1354	.143	.541
.1388	.143	.514
.1404	.143	.486
.1421	.143	.459
.1448	.143	.432
.1487	.143	.405
.1514	.143	.378
.1535	.143	.351
.1551	.143	.324
.1554	.143	.297
.1566	.143	.270
.1619	.143	.243
.1660	.143	.216
.1692	.143	.189
.1785	.143	.162
.1851	.143	.135
.1920	.143	.108
.1986	.143	.081
.2002	.143	.054
.2301	.000	.054
.2718	.000	.027
1.0000	.000	.000

a. The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

Specificity Cutoff – 0.1396

Sensitivity Cutoff – 0.1210

Fig. 10. n) ROC analysis output for test variable - Coherency of Fibre Alignment near the tumour.

ROC for Coherency within500µm

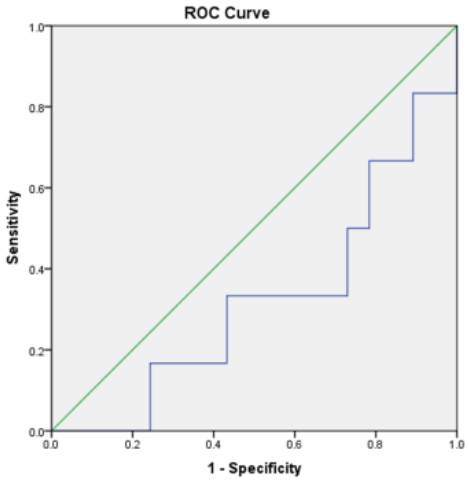
o)

Case Processing Summary

RecurrenceStatusCoded	Valid N (listwise)
Positive[a]	6
Negative	37
Missing	3

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is 1.00.



Area Under the Curve

Test Result Variable(s): Coherencywithin500um

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.320	.117	.161	.090	.550

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Coordinates of the Curve

Test Result Variable(s): Coherencywithin500

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
.0000	1.000	1.000
.0608	.833	1.000
.0723	.833	.973
.0789	.833	.946
.0850	.833	.919
.0892	.833	.892
.0933	.667	.892
.0968	.667	.865
.0976	.667	.838
.0988	.667	.811
.1087	.667	.784
.1189	.500	.784
.1236	.500	.757
.1279	.500	.730
.1304	.333	.730
.1334	.333	.703
.1358	.333	.676
.1389	.333	.649
.1429	.333	.622
.1459	.333	.595
.1498	.333	.568
.1539	.333	.541
.1564	.333	.514
.1579	.333	.486
.1582	.333	.459
.1591	.333	.432
.1626	.167	.432
.1665	.167	.405
.1696	.167	.378
.1719	.167	.351
.1756	.167	.324
.1881	.167	.297
.1975	.167	.270
.2007	.167	.243
.2043	.000	.243
.2102	.000	.216
.2171	.000	.189
.2218	.000	.162
.2331	.000	.135
.2470	.000	.108
.2646	.000	.081
.2832	.000	.054
.3269	.000	.027
1.0000	.000	.000

a. The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

Specificity Cutoff – 0.1571

Sensitivity Cutoff – 0.1291

Fig. 10. o) ROC analysis output for test variable - Coherency of Fibre Alignment near the tumour.

ROC for Coherency Ratio Outside to Within

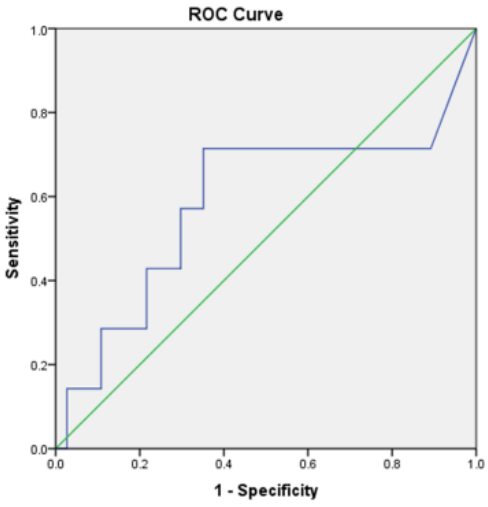
p)

Case Processing Summary

RecurrenceStatusCoded	Valid N (listwise)
Positive[a]	7
Negative	37
Missing	2

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is 1.00.



Diagonal segments are produced by ties.

Area Under the Curve

Test Result Variable(s): RatioofCoherencyOutsidetoWithin

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.587	.139	.470	.314	.860

The test result variable(s): RatioofCoherencyOutsidetoWithin has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Specificity Cutoff – 0.6

Sensitivity Cutoff – 0.7905

Coordinates of the Curve

Test Result Variable(s): RatioofCoherencyOu

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.1368	.714	.892
.3071	.714	.865
.3435	.714	.838
.3687	.714	.811
.4093	.714	.784
.4323	.714	.757
.4492	.714	.730
.4694	.714	.703
.4816	.714	.676
.4911	.714	.649
.5016	.714	.622
.5161	.714	.595
.5283	.714	.568
.5361	.714	.541
.5731	.714	.514
.6270	.714	.486
.6560	.714	.459
.6728	.714	.432
.6865	.714	.405
.6911	.714	.378
.7118	.714	.351
.7476	.571	.351
.7664	.571	.324
.7832	.571	.297
.7978	.429	.297
.8068	.429	.270
.8386	.429	.243
.8791	.429	.216
.8999	.286	.216
.9057	.286	.189
.9283	.286	.162
.9604	.286	.135
.9765	.286	.108
1.0558	.143	.108
1.1586	.143	.081
1.1987	.143	.054
1.2261	.143	.027
1.3949	.000	.027
2.5480	.000	.000

The test result variable(s): RatioofCoherencyOutsidetoWithin has at least one tie between the positive actual state group and the negative actual state group.

a. The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

Fig. 10. p) ROC analysis output for test variable - Ratio of Coherency - outside to within.

The survival curves show that patients with high coherency in alignment far from the tumour showed more recurrence (Figure 10.d). Similar trends were seen for patients with a high coherence ratio, outside to within. (Figure 10.j) No specific trend was seen for the survival plot for the Deposition Area across regions (Figure 10.d), or any other deposition area-related curve. Thus, the test variables indicating the nature of deposition showed a better separation in DFS and the parameter quantifying the amount, despite having comparable AUC in the ROC analysis. Low coherency across regions and near the tumour, performing worse is a peculiar trend (Figure 10.h and i).

The association of these spatial parameters was checked with the clinical characteristics of the cohort. So far, none of the clinical characteristics showed any significant differences between high and low-deposition areas and coherency across regions, regions far and near the tumour, or outside to within the ratio of the parameter. (Table 6a to h). However, it is worth noting that the association of pathological node positivity status (determined post-surgery) for patients with high low fibre deposition area is close to significance.

a)

		All cohort	High Collagen Deposition Area	Low Collagen Deposition Area	p-value
No. of patients		46	23	23	
Age (n=43)	Early (<50)	51.16% (22)	43.48% (10	52.17% (12)	0.5467
	Late (>= 50)	48.84% (21)	52.17% (12)	39.13% (9)	
Menopausal status (n=41)	Pre	43.90% (18)	34.78% (8)	43.48% (10)	0.355
	Post	56.10% (23)	60.87% (14)	39.13% (9)	
Grade (n=44)	Low (I/II)	40.90% (18)	39.13% (9)	39.13% (9)	>0.9999
	High (III)	59.09% (26)	56.52% (13)	56.52% (13)	
Tumor size (cT) (n=38)	T1	18.42% (7)	21.74% (5)	8.70% (2)	0.2983
	T2	76.92% (30)	60.87% (14)	69.57% (16)	
	T3	2.63% (1)	0.00% (0)	4.35% (1)	
Breast Density (n=41)	Type A and B	46.34% (19)	47.83% (11)	34.78% (8)	0.3543
	Type C and D	53.66% (22)	39.13% (9)	56.52% (13)	
pT (primary tissue, no NACT) (n=28)	T0	3.57% (1)	4.35% (1)	4.35% (1)	0.8174
	T1	14.28% (4)	13.04% (3)	4.35% (1)	
	T2	78.57% (22)	60.87% (14)	34.78% (8)	
	T3	0.00% (0)	0.00% (0)	0.00% (0)	
	T4	3.57% (1)	4.35% (1)	0.00% (0)	
pN (primary tissue, no NACT) (n=28)	Negative	67.85% (19)	43.48% (10)	39.13% (9)	0.098
	Positive	2.14% (9)	34.78% (8)	4.35% (1)	
NACT (n=45)	No	73.33% (33)	86.96% (20	56.52% (13)	
	Yes	26.67% (12)	13.04% (3)	39.13% (9)	
PCR status after NACT (n= 12)	pCR	41.67% (5)	4.35% (1)	17.39% (4)	>0.9999
	RD	41.67% (5)	4.35% (1)	17.39% (4)	
Survival outcomes	No. followed-up	45	22	23	
	Median months	30.27	25.26	31.01	
	Follow-up in Months (Range)	0.17-197.53	0.17-197.53	0.46-67.93	
	# Recurred (local, distant)	7	4	3	
	# Death	2	2	0	

Table 6.a) Table of clinical characteristics against high and low collagen deposition area across regions. Significance is tested by the Chi-square test.

b)

		All cohort	High Collagen Deposition Area Within 500um	Low Collagen Deposition Area Within 500um	p-value
No. of patients		46	18	27	
Age (n=43)	**Early (<50)**	51.16% (22)	38.89% (7)	51.85% (14)	0.5303
	Late (>= 50)	48.84% (21)	55.56% (10)	40.74% (11)	
Menopausal status (n=41)	**Pre**	43.90% (18)	27.78% (5)	44.44% (12)	0.2024
	Post	56.10% (23)	66.67% (12)	40.74% (11)	
Grade (n=44)	**Low (I/II)**	40.90% (18)	33.33% (6)	44.44% (12)	0.5408
	High (III)	59.09% (26)	61.11% (11)	51.85% (14)	
Tumor size (cT) (n=38)	**T1**	18.42% (7)	27.78% (5)	7.41% (2)	0.1863
	T2	76.92% (30)	61.11% (11)	66.67% (18)	
	T3	2.63% (1)	0.00% (0)	3.70% (1)	
Breast Density (n=41)	**Type A and B**	46.34% (19)	50.00% (9)	37.04% (10)	0.5197
	Type C and D	53.66% (22)	38.89% (7)	51.85% (14)	
pT (primary tissue, no NACT) (n=28)	**T0**	3.57% (1)	0.00% (0)	3.70% (1)	0.5476
	T1	14.28% (4)	11.11% (2)	7.41% (2)	
	T2	78.57% (22)	66.67% (12)	33.33% (9)	
	T3	0.00% (0)	0.00% (0)	0.00% (0)	
	T4	3.57% (1)	5.56% (1)	0.00% (0)	
pN (primary tissue, no NACT) (n=28)	**Negative**	67.85% (19)	38.89% (7)	40.74% (11)	0.0192
	Positive	2.14% (9)	44.44% (8)	3.70% (1)	
NACT (n=45)	**No**	73.33% (33)	88.89% (16)	59.26% (16)	
	Yes	26.67% (12)	11.11% (2)	37.04% (10)	
PCR status after NACT (n= 12)	**pCR**	41.67% (5)	5.56% (1)	14.81% (4)	>0.9999
	RD	41.67% (5)	5.56% (1)	14.81% (4)	
Survival outcomes	**No. followed-up**	45	17	27	
	Median months	30.27	34.03	30.27	
	Follow-up in Months (Range)	0.17-197.53	0.3-197.53	0.17-67-93	
	# Recurred (local, distant)	7	4	2	
	# Death	2	1	1	

Table 6. b) Table of clinical characteristics against high and low collagen deposition area near the tumour. Significance is tested by the Chi-square test.

c)

		All cohort	High Collagen Deposition Area Outside 500um	Low Collagen Deposition Area Outside 500um	p-value
No. of patients		46	23	23	
Age (n=43)	Early (<50)	51.16% (22)	43.48% (10)	52.17% (12)	0.5467
	Late (>= 50)	48.84% (21)	52.17% (12)	39.13% (9)	
Menopausal status (n=41)	Pre	43.90% (18)	34.78% (8)	43.48% (10)	0.355
	Post	56.10% (23)	60.87% (14)	39.13% (9)	
Grade (n=44)	Low (I/II)	40.90% (18)	39.13% (9)	39.13% (9)	>0.9999
	High (III)	59.09% (26)	56.52% (13)	56.53% (13)	
Tumor size (cT) (n=38)	T1	18.42% (7)	21.74% (5)	8.70% (2)	0.4103
	T2	76.92% (30)	65.22% (15)	65.22% (15)	
	T3	2.63% (1)	0.00% (0)	4.35% (1)	
Breast Density (n=41)	Type A and B	46.34% (19)	43.48% (10	34.78% (8)	0.7557
	Type C and D	53.66% (22)	47.83% (11)	52.17% (12)	
pT (primary tissue, no NACT) (n=28)	T0	3.57% (1)	4.35% (1)	0.00% (0)	0.3596
	T1	14.28% (4)	13.04% (3)	4.35% (1)	
	T2	78.57% (22)	47.83% (11)	43.48% (10	
	T3	0.00% (0)	0.00% (0)	0.00% (0)	
	T4	3.57% (1)	4.35% (1)	0.00% (0)	
pN (primary tissue, no NACT) (n=28)	Negative	67.85% (19)	47.83% (11)	34.78% (8)	
	Positive	2.14% (9)	26.09% (6)	13.04% (3)	>0.9999
NACT (n=45)	No	73.33% (33)	82.61% (19)	60.87% (14)	
	Yes	26.67% (12)	13.04% (3)	39.13% (9)	
PCR status after NACT (n= 12)	pCR	41.67% (5)	4.35% (1)	17.39% (4)	>0.9999
	RD	41.67% (5)	4.35% (1)	17.39% (4)	
Survival outcomes	No. followed-up	45	23	22	
	Median months	30.27	20.41	32.55	
	Follow-up in Months (Range)	0.17-197.53	0.17-103.67	0.76-197.53	
	# Recurred (local, distant)	7	3	4	
	# Death	2	2	0	

Table 6. c) Table of clinical characteristics against high and low collagen deposition area far from the tumour. Significance is tested by the Chi-square test.

d)

		All cohort	High Collagen Deposition Ratio Outside to Within	Low Collagen Deposition Ratio Outside to Within	p-value
No. of patients		46	23	22	
Age (n=43)	**Early (<50)**	51.16% (22)	52.17% (12)	40.91% (9)	0.7579
	Late (>= 50)	48.84% (21)	43.48% (10)	50.00% (11)	
Menopausal status (n=41)	**Pre**	43.90% (18)	43.48% (10	31.82% (7)	0.5378
	Post	56.10% (23)	47.83% (11)	54.55% (12)	
Grade (n=44)	**Low (I/II)**	40.90% (18)	47.38% (11)	31.82% (7)	0.358
	High (III)	59.09% (26)	47.38% (11)	63.64% (14)	
Tumor size (cT) (n=38)	**T1**	18.42% (7)	13.04% (3)	18.18% (4)	0.6928
	T2	76.92% (30)	69.57% (16)	59.09% (13)	
	T3	2.63% (1)	0.00% (0)	4.55% (1)	
Breast Density (n=41)	**Type A and B**	46.34% (19)	34.78% (8)	50.00% (11)	0.3419
	Type C and D	53.66% (22)	56.52% (13)	36.36% (8)	
pT (primary tissue, no NACT) (n=28)	**T0**	3.57% (1)	4.35% (1)	0.00% (0)	0.3259
	T1	14.28% (4)	13.04% (3)	4.55% (1)	
	T2	78.57% (22)	39.13% (9)	54.55% (12)	
	T3	0.00% (0)	0.00% (0)	0.00% (0)	
	T4	3.57% (1)	4.35% (1)	0.00% (0)	
pN (primary tissue, no NACT) (n=28)	**Negative**	67.85% (19)	65.33% (15)	27.27% (6)	0.0721
	Positive	2.14% (9)	17.39% (4)	31.82% (7)	
NACT (n=45)	**No**	73.33% (33)	69.57% (16)	72.73% (16)	
	Yes	26.67% (12)	26.09% (6)	27.27% (6)	
PCR status after NACT (n= 12)	**pCR**	41.67% (5)	8.70% (2)	13.64% (3)	>0.9999
	RD	41.67% (5)	13.04% (3)	9.09% (2)	
Survival outcomes	**No. followed-up**	45	23	21	
	Median months	30.27	28.65	35.1	
	Follow-up in Months (Range)	0.17-197.53	0.17-67.93	0.8-103.67	
	# Recurred (local, distant)	7	3	3	
	# Death	2	2	0	

Table 6. d) Table of clinical characteristics against high and low ratio of collagen deposition - outside to within. Significance is tested by the Chi-square test.

e)

		All cohort	High Sample Coherency	Low Sample Coherency	p-value
No. of patients		46	20	26	
Age (n=43)	**Early (<50)**	51.16% (22)	60.00% (12)	38.46% (10)	0.364
	Late (>= 50)	48.84% (21)	40.00% (8)	50.00% (13)	
Menopausal status (n=41)	**Pre**	43.90% (18)	40.00% (8)	38.46% (10)	>0.9999
	Post	56.10% (23)	55.00% (11)	46.15% (12)	
Grade (n=44)	**Low (I/II)**	40.90% (18)	50.00% (10)	30.77% (8)	0.3588
	High (III)	59.09% (26)	50.00% (10)	61.54% (16)	
Tumor size (cT) (n=38)	**T1**	18.42% (7)	25.00% (5)	7.69% (2)	0.7558
	T2	76.92% (30)	65.00% (13)	65.38% (17)	
	T3	2.63% (1)	5.00% (1)	0.00% (0)	
Breast Density (n=41)	**Type A and B**	46.34% (19)	40.00% (8)	42.31% (11)	0.7558
	Type C and D	53.66% (22)	55.00% (11)	42.31% (11)	
pT (primary tissue, no NACT) (n=28)	**T0**	3.57% (1)	5.00% (1)	0.00% (0)	>0.9999
	T1	14.28% (4)	5.00% (1)	11.54% (3)	
	T2	78.57% (22)	60.00% (12)	38.46% (10)	
	T3	0.00% (0)	0.00% (0)	0.00% (0)	
	T4	3.57% (1)	0.00% (0)	3.85% (1)	
pN (primary tissue, no NACT) (n=28)	**Negative**	67.85% (19)	55.00% (11)	30.77% (8)	0.4197
	Positive	2.14% (9)	15.00% (3)	23.08% (6)	
NACT (n=45)	**No**	73.33% (33)	75.00% (15)	69.23% (18)	
	Yes	26.67% (12)	20.00% (4)	30.77% (8)	
PCR status after NACT (n= 12)	**pCR**	41.67% (5)	10.00% (2)	11.54% (3)	0.755
	RD	41.67% (5)	5.00% (1)	15.38% (4)	
Survival outcomes	**No. followed-up**	45	20	24	
	Median months	30.27	34.78	27.73	
	Follow-up in Months (Range)	0.17-197.53	0.47-103.67	0.17-197.53	
	# Recurred (local, distant)	7	1	6	
	# Death	2	0	2	

Table 6. e) Table of clinical characteristics against high and low coherence across regions. Significance is tested by the Chi-square test.

f)

		All cohort	High Coherency Within 500um	LowCoherency Within 500um	p-value
No. of patients		46	31	15	
Age (n=43)	**Early (<50)**	51.16% (22)	54.84% (17)	33.33% (5)	0.3319
	Late (>= 50)	48.84% (21)	41.94% (13)	53.33% (8)	
Menopausal status (n=41)	**Pre**	43.90% (18)	41.94% (13)	33.33% (5)	0.7417
	Post	56.10% (23)	48.39% (15)	53.33% (8)	
Grade (n=44)	**Low (I/II)**	40.90% (18)	45.16% (14)	26.67% (4)	0.3327
	High (III)	59.09% (26)	51.61% (16)	66.67% (10)	
Tumor size (cT) (n=38)	**T1**	18.42% (7)	16.13% (5)	13.33% (2)	>0.9999
	T2	76.92% (30)	64.52% (20)	66.67% (10)	
	T3	2.63% (1)	3.23% (1)	0.00% (0)	
Breast Density (n=41)	**Type A and B**	46.34% (19)	45.16% (14)	33.33% (5)	0.7437
	Type C and D	53.66% (22)	48.39% (15)	46.67% (7)	
pT (primary tissue, no NACT) (n=28)	**T0**	3.57% (1)	3.23% (1)	0.00% (0)	>0.9999
	T1	14.28% (4)	9.68% (3)	6.67% (1)	
	T2	78.57% (22)	48.39% (15)	46.67% (7)	
	T3	0.00% (0)	0.00% (0)	0.00% (0)	
	T4	3.57% (1)	0.00% (0)	6.67% (1)	
pN (primary tissue, no NACT) (n=28)	**Negative**	67.85% (19)	45.16% (14)	33.33% (5)	0.4074
	Positive	2.14% (9)	16.13% (5)	26.67% (4)	
NACT (n=45)	**No**	73.33% (33)	67.74% (21)	80.00% (12)	
	Yes	26.67% (12)	29.03% (9)	20.00% (3)	
PCR status after NACT (n= 12)	**pCR**	41.67% (5)	9.68% (3)	13.33% (2)	0.4444
	RD	41.67% (5)	12.90% (4)	0.00% (0)	
Survival outcomes	**No. followed-up**	45	30	14	
	Median months	30.27	26.82	35.1	
	Follow-up in Months (Range)	0.17-197.53	0.17-103.67	0.3-197.53	
	# Recurred (local, distant)	7	2	5	
	# Death	2	1	1	

Table 6. f) Table of clinical characteristics against high and low coherence near the tumour. Significance is tested by the Chi-square test.

g)

		All cohort	High Coherency Outside 500um	Low Coherency Outside 500um	p-value
No. of patients		46	13	30	
Age (n=43)	**Early (<50)**	51.16% (22)	46.15% (6)	53.33% (16)	0.747
	Late (>= 50)	48.84% (21)	53.85% (7)	46.67% (14)	
Menopausal status (n=41)	**Pre**	43.90% (18)	30.77% (4)	40.00% (12)	0.4946
	Post	56.10% (23)	69.23% (9)	46.67% (14)	
Grade (n=44)	**Low (I/II)**	40.90% (18)	46.15% (6)	36.67% (11)	>0.9999
	High (III)	59.09% (26)	53.85% (7)	56.67% (17)	
Tumor size (cT) (n=38)	**T1**	18.42% (7)	23.08% (3)	10.00% (3)	0.3781
	T2	76.92% (30)	69.23% (9)	66.67% (20)	
	T3	2.63% (1)	0.00% (0)	3.33% (1)	
Breast Density (n=41)	**Type A and B**	46.34% (19)	46.15% (6)	43.33% (13)	>0.9999
	Type C and D	53.66% (22)	46.15% (6)	50.00% (15)	
pT (primary tissue, no NACT) (n=28)	**T0**	3.57% (1)	0.00% (0)	3.33% (1)	>0.9999
	T1	14.28% (4)	7.69% (1)	10.00% (3)	
	T2	78.57% (22)	53.85% (7)	43.33% (13)	
	T3	0.00% (0)	0.00% (0)	0.00% (0)	
	T4	3.57% (1)	7.69% (1)	0.00% (0)	
pN (primary tissue, no NACT) (n=28)	**Negative**	67.85% (19)	46.15% (6)	40.00% (12)	>0.9999
	Positive	2.14% (9)	23.08% (3)	16.67% (5)	
NACT (n=45)	**No**	73.33% (33)	76.92% (10)	66.67% (20	
	Yes	26.67% (12)	23.08% (3)	30.00% (9)	
PCR status after NACT (n= 12)	**pCR**	41.67% (5)	15.38% (2)	10.00% (3)	>0.9999
	RD	41.67% (5)	7.69% (1)	13.33% (4)	
Survival outcomes	**No. followed-up**	45	13	29	
	Median months	30.27	27.73	31.07	
	Follow-up in Months (Range)	0.17-197.53	1.1-103.67	0.17-101.23	
	# Recurred (local, distant)	7	4	3	
	# Death	2	2	0	

Table 6. g) Table of clinical characteristics against high and low coherence far from the tumour. Significance is tested by the Chi-square test.

h)

		All cohort	High Coherency Ratio	Low Coherency Ratio	p-value
No. of patients		46	15	30	
Age (n=43)	**Early (<50)**	51.16% (22)	40.00% (6)	50.00% (15)	0.7442
	Late (>= 50)	48.84% (21)	53.33% (8)	43.33% (13)	
Menopausal status (n=41)	**Pre**	43.90% (18)	33.33% (5)	40.00% (12)	>0.9999
	Post	56.10% (23)	53.33% (8)	50.00% (15)	
Grade (n=44)	**Low (I/II)**	40.90% (18)	33.33% (5)	43.33% (13)	0.7438
	High (III)	59.09% (26)	60.00% (9)	53.33% (16)	
Tumor size (cT) (n=38)	**T1**	18.42% (7)	13.33% (2)	16.67% (5)	0.6869
	T2	76.92% (30)	80.00% (12)	56.67% (17)	
	T3	2.63% (1)	0.00% (0)	3.33% (1)	
Breast Density (n=41)	**Type A and B**	46.34% (19)	40.00% (6)	43.33% (13)	0.7475
	Type C and D	53.66% (22)	53.33% (8)	43.33% (13)	
pT (primary tissue, no NACT) (n=28)	**T0**	3.57% (1)	0.00% (0)	3.33% (1)	0.6361
	T1	14.28% (4)	6.67% (1)	10.00% (3)	
	T2	78.57% (22)	46.67% (7)	46.67% (14)	
	T3	0.00% (0)	0.00% (0)	0.00% (0)	
	T4	3.57% (1)	6.67% (1)	0.00% (0)	
pN (primary tissue, no NACT) (n=28)	**Negative**	67.85% (19)	40.00% (6)	40.00% (12)	>0.9999
	Positive	2.14% (9)	20.00% (3)	20.00% (6)	
NACT (n=45)	**No**	73.33% (33)	73.33% (11)	70.00% (21)	
	Yes	26.67% (12)	26.67% (4)	26.67% (8)	
PCR status after NACT (n= 12)	**pCR**	41.67% (5)	20.00% (3)	6.67% (2)	0.1667
	RD	41.67% (5)	0.00% (0)	13.33% (4)	
Survival outcomes	**No. followed-up**	45	15	28	
	Median months	30.27	27.73	31.02	
	Follow-up in Months (Range)	0.17-197.53	0.3-79.4	0.17-103.67	
	# Recurred (local, distant)	7	4	2	
	# Death	2	11	0	

Table 6. g) Table of clinical characteristics against high and low ratio of Coherency - outside to within. Significance is tested by the Chi-square test.

4. Discussion

Given the aggressive nature and lack of therapeutic targets, along with the disproportionate burden on Indian women, there is a pressing need to find more prognostic markers for the disease. In recent years, cancer research has shifted towards studying the role of the tumour microenvironment and how it interacts with the tumour to understand the disease better. To this end, we attempted to explore the role of the TME, particularly collagen deposition in the TME, in determining TNBC prognosis in Indian patients. Several studies have pointed towards the role of collagen deposition in the behaviour of the tumour. In this study, we aimed to build an analysis pipeline, based on standard image analysis tools like FIJI and QuPath, to quantify deposition patterns of WSI images of FFPE tissue sections of Indian TNBC patients stained with Masson-Goldner's trichrome stain.

We have stained all 139 from our defined cohort and acquired whole slide images of 126 patient samples. Keeping brightfield WSIs at the centre of our analysis is crucial since these give a well-rounded picture of the entire tissue architecture. Brightfield WSIs are also an emerging standard in diagnostics and studies on tissues. However, while working with WSIs, processing power has proved to be a limitation. Therefore, in this thesis, we have relied on tiles of different tissue regions to validate the pipeline we are constructing. We have made multiple comparisons between the methods of analysis of different pipelines. We chose to move forward with the TWOMBLI workflow since it gave the most number of outputs, including crucial measurements of alignment of the fibres. We are awaiting SHG microscopy images to see how concordant the use of the pipeline is with the gold standard for collagen fibre studies.

Of the 126 processed and imaged samples, we have performed an analysis on a set of 46 randomly selected samples. The demographic and clinical characteristics distribution indicates that this set is evenly distributed across ages, and has more representation of post-menopausal status, high-grade, and high tumour size. We have yet to receive curated data about the node-positivity and LVI status from the biobank; hence, these parameters have been excluded from the study.

We saw that the spatial parameters we have quantified can differentiate between regions near the tumour and far from it. It is difficult to comment on the biological reasons for the trends in fibre skeleton quantification since, currently, the resolution of the skeleton is quite low, and the significance seems to be reduced in all parameters, barring the coherence of alignment of fibres. This tells us that the pipeline, in its current state, quantifies changes in the deposition's nature.

The ROC analysis of our chosen test variable confirmed that it is indeed beneficial to separate regions based on the proximity to the tumour. Interestingly, the coherency of fibres in regions far from the tumour was the best separator for disease-free survivorship. The ratio between the outputs of the parameter of two different regions is also a good separator, indicating that exploring intra-tissue heterogeneity in deposition would be insightful. In the disease-free survivor analysis, the high coherency far from the tumour seemed to have more instances of recurrence. This could be due to more aggressive tumours influencing broader areas. In the case of coherency near the tumour, low coherency patients performed worse, and the same was true for coherency across samples. Placing these results in the context of the literature, it is seen that higher alignment of fibres around the tumour boundary has worse survivorship (Conklin *et al.*, 2011; Ouellette *et al.*, 2021); however, these studies have looked at collagen fibre alignment concerning the tumour boundary. Tumours with fibres aligned but perpendicular to the tumour boundary had more aggressive disease than tissues with less aligned fibres and more aligned fibres but parallel to the tumour boundary. Thus, considering alignment coherency and their orientation with respect to the tumour boundary could resolve the data. We have not done the same, which is reflected in the low AUC of coherency near the tumour and across regions in the ROC. Similarly, concerning the collagen deposition area in the tissue, there is considerable data loss in quantifying the deposition area based on the area covered by the fibre skeleton. The separation by these parameters should be revisited after re-quantifying the area based on the binary images of the deconvoluted images.

The clinical parameters were not significantly different when samples above a cutoff of test variables were compared with those below for this set of analysed samples. The

number of samples in categories of clinical parameters was also low (e.g., tumours with low tumour sizes, and NACT, yes, in the case of response status). Further comments on the association with clinical parameters should be made after the data curation of all samples and the completion of analysis for all available samples in the cohort.

Along with completing the analysis of all samples in the cohort, fine-tuning steps in the pipeline, which may cause data loss, would provide a much better picture. Now that the pipeline outline is complete and working, the spatial parameters with TIL scores could be compared to see the deposition pattern's relationship with immune cells in the TME. We plan to validate our results with an ML model for analysis (currently being developed in collaboration with Dr Namita S and Shashank G, Amaranth Medical Analytics Pvt. Ltd, Bengaluru) and then upscale the same pipeline for WSIs. Exploring how concordant the pipeline is with HnE images should also be explored. Comparing outputs from analysis done on trichrome-stained tissues and hematoxylin and eosin-stained (HnE) tissue would explore if the pipeline can be extrapolated to the HnE stain, which, although not specific to collagen, is used in routine morphometric histopathological exams. This provides scope for expanding the compatibility with the pipeline for future analyses.

5. Conclusion

In this thesis, we managed to stain the FPPE tissue of Indian TNBC patients with Masson Goldner's Trichrome Stain to validate and visualise collagen deposition. We acquired brightfield WSIs, and quantified deposition patterns in regions near the tumour and far from it in samples from a set of 46 randomly selected samples from our defined cohort of 139 patients.. We saw that most spatial parameters quantified differed in regions near and far from the tumour and that separating between regions based on proximity to the tumour gives us insights about patient outcomes better than looking at the whole tissue’s collagen deposition spatial parameters and has consequences toward the prognosis of patients.

6. References

Acerbi, I., Cassereau, L., Dean, I., Shi, Q., Au, A., Park, C., Chen, Y.Y., Liphardt, J., Hwang, E.S., and Weaver, V.M. (2015). Human breast cancer invasion and aggression correlates with ECM stiffening and immune cell infiltration. Integr. Biol. 7, 1120–1134. https://doi.org/10.1039/c5ib00040h.

Ahn, S., Woo, J.W., Lee, K., and Park, S.Y. (2020). HER2 status in breast cancer: changes in guidelines and complicating factors for interpretation. J. Pathol. Transl. Med. 54, 34–44. https://doi.org/10.4132/jptm.2019.11.03.

Angelico, G., Broggi, G., Tinnirello, G., Puzzo, L., Vecchio, G.M., Salvatorelli, L., Memeo, L., Santoro, A., Farina, J., Mulé, A., et al. (2023). Tumor Infiltrating Lymphocytes (TILS) and PD-L1 Expression in Breast Cancer: A Review of Current Evidence and Prognostic Implications from Pathologist's Perspective. Cancers 15, 4479. https://doi.org/10.3390/cancers15184479.

Bancroft, J.D. and Alan Stevens (1982). Theory and practice of histological techniques / editors, John D. Bancroft, Alan Stevens ; foreword by I.M.P. Dawson. Seventh edition. (Churchill Livingstone).

Bankhead, P., Loughrey, M.B., Fernández, J.A., Dombrowski, Y., McArt, D.G., Dunne, P.D., McQuaid, S., Gray, R.T., Murray, L.J., Coleman, H.G., et al. (2017). QuPath: Open source software for digital pathology image analysis. Sci. Rep. 7, 16878. https://doi.org/10.1038/s41598-017-17204-5.

Baptist, J., Dixit, S., Busheri, L., Krishnan, L., Alhat, R., Verghese, B., and Koppiker, C. (2017). Evaluating the Performance of an Advanced Breast Cancer Diagnosis Unit in India. Indian J. Public Health Res. Dev. 8, 598. https://doi.org/10.5958/0976-5506.2017.00403.X.

Bissell, M.J., and Hines, W.C. (2011). Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. Nat. Med. 17, 320–329. https://doi.org/10.1038/nm.2328.

Bodelon, C., Mullooly, M., Pfeiffer, R.M., Fan, S., Abubakar, M., Lenz, P., Vacek, P.M., Weaver, D.L., Herschorn, S.D., Johnson, J.M., et al. (2021). Mammary collagen architecture and its association with mammographic density and lesion severity among women undergoing image-guided breast biopsy. Breast Cancer Res. 23, 105. https://doi.org/10.1186/s13058-021-01482-z.

Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I., and Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA. Cancer J. Clin. 74, 229–263. https://doi.org/10.3322/caac.21834.

Busheri, L., Dixit, S., Nare, S., Alhat, R., Thomas, G., Jagtap, M., Navgire, R., Shinde, P., Banale, R., Unde, R., et al. (2021). Breast cancer biobank from a single institutional cohort in an urban setting in India: Tumor characteristics and survival outcomes. Cancer Treat. Res. Commun. 28, 100409. https://doi.org/10.1016/j.ctarc.2021.100409.

Conklin, M.W., Eickhoff, J.C., Riching, K.M., Pehlke, C.A., Eliceiri, K.W., Provenzano, P.P., Friedl, A., and Keely, P.J. (2011). Aligned Collagen Is a Prognostic Signature for Survival in Human Breast Carcinoma. Am. J. Pathol. 178, 1221–1232. https://doi.org/10.1016/j.ajpath.2010.11.076.

Fang, M., Yuan, J., Peng, C., and Li, Y. (2014). Collagen as a double-edged sword in tumor progression. Tumor Biol. 35, 2871–2882. https://doi.org/10.1007/s13277-013-1511-7.

Frisch, K.E., Duenwald-Kuehl, S.E., Kobayashi, H., Chamberlain, C.S., Lakes, R.S., and Vanderby, R. (2012). Quantification of collagen organization using fractal dimensions and Fourier transforms. Acta Histochem. 114, 140–144. https://doi.org/10.1016/j.acthis.2011.03.010.

Gael Latour et al., "Correlative nonlinear optical microscopy and infrared nanoscopy reveals collagen degradation in altered parchments," Scientific Reports 6, 26344 (2016)

Gruber, H. (1992). Adaptations of Goldner's Masson Trichrome Stain for the Study of Undecalcified Plastic Embedded Bone: Biotechnic & Histochemistry: Vol 67, No 1 - Get Access. https://www.tandfonline.com/doi/epdf/10.3109/10520299209110002?needAccess=true.

Hammond, M.E.H., Hayes, D.F., Wolff, A.C., Mangu, P.B., and Temin, S. (2010). American Society of Clinical Oncology/College of American Pathologists Guideline Recommendations for Immunohistochemical Testing of Estrogen and Progesterone Receptors in Breast Cancer. J. Oncol. Pract. 6, 195–197. https://doi.org/10.1200/JOP.777003.

Hanahan, D., and Coussens, L.M. (2012). Accessories to the Crime: Functions of Cells Recruited to the Tumor Microenvironment. Cancer Cell 21, 309–322. https://doi.org/10.1016/j.ccr.2012.02.022.

Hasebe, T., Sasaki, S., Imoto, S., Mukai, K., Yokose, T., and Ochiai, A. (2002). Prognostic Significance of Fibrotic Focus in Invasive Ductal Carcinoma of the Breast: A Prospective Observational Study. Mod. Pathol. 15, 502–516. https://doi.org/10.1038/modpathol.3880555.

Howlader, N., Cronin, K.A., Kurian, A.W., and Andridge, R. (2018). Differences in Breast Cancer Survival by Molecular Subtypes in the United States. Cancer Epidemiol. Biomarkers Prev. 27, 619–626. https://doi.org/10.1158/1055-9965.EPI-17-0627.

Kanade, S., Desai, M., Bhatavadekar, N., and Balasubramanian, N. (2024). Quantitative analysis of fibrillar collagen organization in the immediate proximity of embedded fibroblasts in 3D collagen hydrogels. J. Biosci. 49, 65. https://doi.org/10.1007/s12038-024-00449-4.

Kitano, A., Ono, M., Yoshida, M., Noguchi, E., Shimomura, A., Shimoi, T., Kodaira, M., Yunokawa, M., Yonemori, K., Shimizu, C., et al. (2017). Tumour-infiltrating lymphocytes

are correlated with higher expression levels of PD-1 and PD-L1 in early breast cancer. ESMO Open 2, e000150. https://doi.org/10.1136/esmoopen-2016-000150.

Kulkarni, A., Kelkar, D.A., Parikh, N., Shashidhara, L.S., Koppiker, C.B., and Kulkarni, M. (2020). Meta-Analysis of Prevalence of Triple-Negative Breast Cancer and Its Clinical Features at Incidence in Indian Patients With Breast Cancer. JCO Glob. Oncol., 1052–1062. https://doi.org/10.1200/GO.20.00054.

Laplane, L., Duluc, D., Larmonier, N., Pradeu, T., and Bikfalvi, A. (2018). The Multiple Layers of the Tumor Environment. Trends Cancer 4, 802–809. https://doi.org/10.1016/j.trecan.2018.10.002.

Lehmann, B.D., Bauer, J.A., Chen, X., Sanders, M.E., Chakravarthy, A.B., Shyr, Y., and Pietenpol, J.A. (2011). Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J. Clin. Invest. 121, 2750–2767. https://doi.org/10.1172/JCI45014.

Li, H., Bera, K., Toro, P., Fu, P., Zhang, Z., Lu, C., Feldman, M., Ganesan, S., Goldstein, L.J., Davidson, N.E., et al. (2021). Collagen fiber orientation disorder from H&E images is prognostic for early stage breast cancer: clinical trial validation. Npj Breast Cancer 7, 1–10. https://doi.org/10.1038/s41523-021-00310-z.

Lin, N.U., Vanderplas, A., Hughes, M.E., Theriault, R.L., Edge, S.B., Wong, Y.-N., Blayney, D.W., Niland, J.C., Winer, E.P., and Weeks, J.C. (2012). Clinicopathologic features, patterns of recurrence, and survival among women with triple-negative breast cancer in the National Comprehensive Cancer Network. Cancer 118, 5463–5472. https://doi.org/10.1002/cncr.27581.

Naba, A., Clauser, K.R., Lamar, J.M., Carr, S.A., and Hynes, R.O. (2014). Extracellular matrix signatures of human mammary carcinoma identify novel metastasis promoters. eLife 3, e01308. https://doi.org/10.7554/eLife.01308.

Northey, J.J., Barrett, A.S., Acerbi, I., Hayward, M.-K., Talamantes, S., Dean, I.S., Mouw, J.K., Ponik, S.M., Lakins, J.N., Huang, P.-J., et al. (2020). Stiff stroma increases

breast cancer risk by inducing the oncogene ZNF217. https://doi.org/10.1172/JCI129249.

Orrantia-Borunda, E., Anchondo-Nuñez, P., Acuña-Aguilar, L.E., Gómez-Valles, F.O., and Ramírez-Valdespino, C.A. (2022). Subtypes of Breast Cancer. In Breast Cancer, H. N. Mayrovitz, ed. (Exon Publications).

Oskarsson, T. (2013). Extracellular matrix components in breast cancer progression and metastasis. The Breast 22, S66–S72. https://doi.org/10.1016/j.breast.2013.07.012.

Ouellette, J.N., Drifka, C.R., Pointer, K.B., Liu, Y., Lieberthal, T.J., Kao, W.J., Kuo, J.S., Loeffler, A.G., and Eliceiri, K.W. (2021). Navigating the Collagen Jungle: The Biomedical Potential of Fiber Organization in Cancer. Bioengineering 8, 17. https://doi.org/10.3390/bioengineering8020017.

Plasilova, M.L., Hayse, B., Killelea, B.K., Horowitz, N.R., Chagpar, A.B., and Lannin, D.R. (2016). Features of triple-negative breast cancer: Analysis of 38,813 cases from the national cancer database. Medicine (Baltimore) 95, e4614. https://doi.org/10.1097/MD.0000000000004614.

Provenzano, P.P., Eliceiri, K.W., Campbell, J.M., Inman, D.R., White, J.G., and Keely, P.J. (2006). Collagen reorganization at the tumor-stromal interface facilitates local invasion. BMC Med. 4, 38. https://doi.org/10.1186/1741-7015-4-38.

Pruitt, H.C., Lewis, D., Ciccaglione, M., Connor, S., Smith, Q., Hickey, J.W., Schneck, J.P., and Gerecht, S. (2020). Collagen fiber structure guides 3D motility of cytotoxic T lymphocytes. Matrix Biol. 85–86, 147–159. https://doi.org/10.1016/j.matbio.2019.02.003.

Robert W. Boyd, “Nonlinear Optics,” J. Chem. Inf. Model. 53, 1689–1699, (2013)

Rømer, A.M.A., Thorseth, M.-L., and Madsen, D.H. (2021). Immune Modulatory Properties of Collagen in Cancer. Front. Immunol. 12. https://doi.org/10.3389/fimmu.2021.791453.

Ruifrok, A.C., and Johnston, D.A. (2001). Quantification of histochemical staining by color deconvolution. Anal. Quant. Cytol. Histol., 291–299.

Salgado, R., Denkert, C., Demaria, S., Sirtaine, N., Klauschen, F., Pruneri, G., Wienert, S., Van den Eynden, G., Baehner, F.L., Penault-Llorca, F., et al. (2015). The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. Ann. Oncol. 26, 259–271. https://doi.org/10.1093/annonc/mdu450.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682. https://doi.org/10.1038/nmeth.2019.

Smith, B., Hermsen, M., Lesser, E., Ravichandar, D., and Kremers, W. (2021). Developing image analysis pipelines of whole-slide images: Pre- and post-processing. J. Clin. Transl. Sci. 5, e38. https://doi.org/10.1017/cts.2020.531.

Tian, F., Jiang, Y., Liu, Y., Lu, S., Yang, J., and Cao, Y. (2021). Can fractal dimension analysis be used in quantitating collagen structure? Exp. Dermatol. 30, 1825–1828. https://doi.org/10.1111/exd.14417.

Vaid, P.M., Puntambekar, A.K., Jumle, N.S., Banale, R.A., Ansari, D., Reddy, R.R., Unde, R.R., Namewar, N.P., Kelkar, D.A., Shashidhara, L.S., et al. (2022). Evaluation of tumor-infiltrating lymphocytes (TILs) in molecular subtypes of an Indian cohort of breast cancer patients. Diagn. Pathol. 17, 91. https://doi.org/10.1186/s13000-022-01271-y.

Wershof, E., Park, D., Barry, D.J., Jenkins, R.P., Rullan, A., Wilkins, A., Schlegelmilch, K., Roxanis, I., Anderson, K.I., Bates, P.A., et al. (2021). A FIJI macro for quantifying pattern in extracellular matrix. Life Sci. Alliance 4. https://doi.org/10.26508/lsa.202000880.

Winkler, J., Abisoye-Ogunniyan, A., Metcalf, K.J., and Werb, Z. (2020). Concepts of extracellular matrix remodelling in tumour progression and metastasis. Nat. Commun. 11, 5120. https://doi.org/10.1038/s41467-020-18794-x.

Xu, S., Xu, H., Wang, W., Li, S., Li, H., Li, T., Zhang, W., Yu, X., and Liu, L. (2019). The role of collagen in cancer: from bench to bedside. J. Transl. Med. 17, 309. https://doi.org/10.1186/s12967-019-2058-1.

Yuan, Z., Li, Y., Zhang, S., Wang, X., Dou, H., Yu, X., Zhang, Z., Yang, S., and Xiao, M. (2023). Extracellular matrix remodeling in tumor progression and immune escape: from mechanisms to treatments. Mol. Cancer 22, 48. https://doi.org/10.1186/s12943-023-01744-8.