

CA23125 - The mETamaterial foRmalism approach to recognize cAncer (TETRA)

TETRA WP1. Optimising methodologies for experimental visualisation of biomedical tissues

Deliverable 1.1 requirements for images of biological tissue to be used for automated cancer detection.

Contributors*

R. Arabi Belaghi¹, A. Aydin², N. Carbó³, M. Durmuş⁴, A. Espona^{5,6}, P. Fernandez^{3,5,6}, G. Fuster^{5,6}, Ö. Gürünlü-Alma⁷, P. Loza-Alvarez⁸⁺, I. Meglinski⁹, I. Miler¹⁰, J. J. Ruiz-Gonzalez⁸.

1. Swedish University

2. Ondokuz Mayıs University.

3. Universitat de Barcelona

4. Samsun University

5. Universitat de Vic – Universitat Central de Catalunya

6. Institut de Recerca i Innovació en Ciències de la Vida i de la Salut de la Catalunya Central

7. Muğla Sıtkı Kocman University

8. ICFO - The Institute of Photonic Sciences

9. Aston University

10. BioSense Institute, University of Novi Sad

*All contributors appear in alphabetic order according to surname.

+Corresponding: pablo.loza@icfo.eu

Abstract

This document consolidates sample-preparation practices across modalities and aligns them with the analytical and AI and modelling workflows used in the project. Sample preparation—fixation, embedding, sectioning, staining, and substrate selection—directly determines the structural, chemical, and optical signals each imaging technique can capture. As a result, rigorous documentation and consistent execution of these steps are fundamental to interpretation, cross-modal correlation, and reproducibility.

Introduction

Understanding how a sample is prepared for each imaging technique is critical as it directly influences the structural, chemical, and optical properties captured during imaging. For example, paraffin embedding introduces residual wax that, if not in the hands of an experienced researcher, can interfere with Raman spectroscopy signals. On the other hand,

while OCT embedding avoids such interference, it requires cryoprotection steps that preserve lipids and antigens. These differences may result in very different features during imaging and interpretation. Furthermore, when new computational models are applied to infer biological states, such as distinguishing healthy tissue from malignant tumors, the accuracy of predictions depends on the consistency and optical properties of the sample and the preparation components used. Models trained on images from one preparation method may fail when applied to data from another because staining, fixation, and substrate choices alter contrast, spectral profiles, and morphology. For instance, Raman-based classifiers rely on biochemical signatures that can be masked by paraffin residues (if this is not properly excluded before hand), while deep learning models for histology depend on color and texture patterns introduced during staining. Therefore, documenting and understanding preparation protocols is essential for reproducibility, cross-modal correlation, and ensuring that AI-driven or model predictions reflect true biological differences rather than artifacts introduced during sample handling.

Below, there is a compilation of the different steps that are normally carried out when preparing a sample (see Table 1 and 2). These procedures, although similar, may differ, depending on the type of imaging techniques to be used. Furthermore, because the differences, the same tissue processed in one place may give different results if processed in a different place with a different protocol. A similar situation is expected to occur if the right number of variables is not taken into account when training the AI or similar approaches (see Table 3). The intention of the tables is to make the different scientists aware of the different methodologies so that these can be taken into account. Ideally, a consensus between biologist, medical doctors, microscopists and modelers should take place so that results are robust in their final interpretation

Sample Preparation and Computational Workflow Tables

Table 1: Sample Preparation

Type of Tissue	Embedding Media	Sample Pre-Processing	Types of Substrates	Sample Preparation Used	Image Modality Used	Image Processing Methods	Correlation with Standard Techniques	Contributors
Breast cancer, Normal breast, Skin tissue	Paraffin	Dewaxing: PFA 4% O/N fixation	Glass	IF, IHC, H&E staining	Contrast-phase microscopy Fluorescence; Confocal	ImageJ and Area and Integrated Density	Correlation with H&E staining to localize tumor ROI	G. Fuster. A. Espona; (U. Vic, U. Central de Catalunya. IRIS-CC); M. N. Carbo, P. Fernandez (U. de Barcelona)
Breast cancer Normal breast	OCT	PFA 4% O/N fixation + 20% solution of sucrose to cryoprotect O/N	Glass	IF, IHC, H&E staining	Contrast-phase microscopy Fluorescence; Confocal	ImageJ and Area and Integrated Density	Correlation with H&E staining to localize tumor ROI	G. Fuster . A. Espona; (U. Vic, U. Central de Catalunya. and IRIS-CC);); M. N. Carbo P. Fernandez (U. de Barcelona)
Biopsy sections (breast, retina)	OCT	silanisation of slides; 20% solution of sucrose to cryoprotect O/N + PFA 4% O/N fixation	Quartz, CaF ₂ , super-mirror stainless steel	H&E staining	Raman	Python, Image J, Matlab	Correlation with H&E staining on same or consecutive section	J. J. Ruiz, P. Loza-Alvarez (ICFO)
Breast cancer and Head and	OCT-embedded	PFA 4% O/N fixation +	Glass	IF, IHC, H&E staining	Contrast-phase	ImageJ and	No correlation or	G. Fuste. A. Espona (U. Vic, U. Central de

neck cancer (animal or in-ovo models inoculated with human cell lines)		O/Nin 20% sucrose for cryoprotection			microscopy Fluorescence; Confocal	Area and Integrated Density	correlation with H&E staining to localize tumor ROI	Catalunya. IRIS-CC); M. N. Carbo P. Fernandez (U. de Barcelona)
Tumor biopsy, animal tissue, normal control tissue	Paraffin-embedded	Washed in cold PBS and Kept on ice. Fixation in 10% neutral buffered formalin (12–24 h), PBS wash, deparaffinization (xylene, ethanol gradients),	Silanized glass, Quartz, CaF ₂ , (Raman) ITO-coated, quartz, (Maldi) gold/platinum-coated (SEM)	H&E, IHC, IF, Raman dye-conjugated antibodies	Brightfield, Fluorescence, Confocal, Raman, Multiphoton (SHG, CARS, SRS), SEM/TEM	FIJI/R ImageJ, Python (scikit-image), MATLAB,	Validated against H&E, IHC, molecular assays	Prof. Dr. Özlem Gürnlü Alma (Muğla / Türkiye) R. Arabi Belaghi (Swedish University/Sweden)
Skin	Paraffin and Araldite embeded	See Table 2 for extended details *	Glass slides, copper grids	H&E staining, BF&MB staining, unstained and fixed for TPEF/SHG, contrasted for TEM	Light microscopy, TPEF, SHG, pSHG, TEM, Raman spectroscopy	iTEM software for collagen fiber measurements; R (hyperSpec) for Raman spectra processing; PCA analysis ADD: Image J	Correlation between TPEF/SHG and H&E morphology; Raman spectra compared across groups	I. Miler BioSense Institute, University of Novi Sad
Rat brain (motor cortex, piriform)	Paraffin	Perfusion fixation with 10% neutral formalin, silver nitrate pretreatment (3–	Glass slides	Silver impregnation, cresyl violet Nissl staining, GFAP	Brightfield microscopy (Axio Imager 2), immunohis	ZEN image analytic system; MATLAB for 3D	Comparison of silver impregnation vs Nissl vs GFAP IHC	I. Meglinski (Aston U.)

cortex, striatum)		4 days), dehydration (ethanol gradients), Deparaffinized in Xylene (5min), ethyl (5min) and distilled water (2min)		immunohistoc hemistry	tochemistr y	reconstructi on; Origin for statistical analysis		
----------------------	--	--	--	--------------------------	-----------------	---	--	--

Table 2. SKIN Tissue preparation:

	LM (Light Microscopy)	TEM (Transmission Electron Microscopy)
Fixative	10% Neutral Buffered Formalin ($\approx 4\%$ formaldehyde)	2.5% Glutaraldehyde + 1% Osmium tetroxide (OsO_4)
Sample size	3 – 5 mm	$\sim 1\text{mm}^3$
Fixation times	12–24 h	2–4 h + 1–2 h (OsO_4)
Temperature	Room temperature or 4 °C	4 °C
Subsequent processing	Paraffin embedding	Epoxy resin embedding
Purpose	Preservation of histological morphology	Preservation of cellular ultrastructure

O/N: Over night

IF: Immunofluorescence

IHC: Immunohistochemistry

H&E: Hematoxinilin and Eosin staining

ROI: Region of Interest

PFA: Paraformaldehyde

OCT: Optimal Cutting Temperature compound

PBS: Phosphate-Buffered Saline

BF: basic fuchsine

MB: methylene blue

FFPE: Formalin Fixed paraffin embedded

Table 3. Computational and AI-based workflows:

Type of Tissue	Sample Pre-Processing	Types of Substrates	Sample Preparation Used	Image Modality Used	Image Processing Methods	Correlation with Standard Techniques	Contributors
<i>Digital datasets (H&E, Raman, SHG, OCT)</i>	• Metadata parsing and harmonization (JSON, XML) • DICOM → OME-TIFF standardization with checksum validation • Illumination and color normalization (Macenko & Reinhard methods) • Artifact removal via morphological filters & CNN-based denoising (DnCNN, Noise2Void) • Patch extraction with adaptive tiling and context padding • Automated tissue segmentation (U-Net, SAM, or Mask R-CNN) • Feature extraction (GLCM, SIFT, ORB, and deep embeddings) • Quality control: blur detection, stain inconsistency scoring, exposure histogram analysis • Data balancing and	<i>OME-TIFF digital slides; HDF5 spectral cubes</i>	<i>AI-ready preprocessing pipeline automated in Python; batch logging with MLflow; dataset versioning using DVC</i>	<i>Digital histology; hyperspectral; multiphoton</i>	• Classical: ImageJ/FIJI macros, OpenCV filters (CLAHE, morphological ops) • ML/AI-based: PyTorch, TensorFlow, scikit-image, Detectron2, MONAI, albumentations • Segmentation: U-Net++, DeepLabV3+, Vision Transformers (Swin-UNet) • Feature extraction: self-supervised encoders (SimCLR, BYOL), embeddings via CLIP/ResNet-50 • Analysis: PCA, t-SNE, UMAP for cluster visualization • Explainability: Grad-CAM, SHAP, LIME for pathology region attribution	<i>Cross-validation with histopathological expert annotations; model performance metrics (F1, IoU, ROC-AUC); statistical correlation between morphological and spectral features; explainability maps verified by pathologists</i>	<i>M. Durmuş (Samsun University)</i>

	augmentation (rotation, elastic deformation, stain transfer, CutMix, RandAugment)				• Anomaly detection: Autoencoders and One-Class SVM for outlier patches • Pipeline orchestration: Jupyter + Airflow integration for reproducibility		
		<i>DCM (DICOM)</i>	anonymization and transformation. Minimum 2048x2048 pixels) or resolution / density (600 PPI)				A. Aydın ² Ondokuz Mayıs University.

Harmonized Methods Derived from the Table

Across tissues and contributors, the table shows a common backbone: (i) pre-processing (e.g., PBS handling, fixation in 10% neutral buffered formalin or cryoprotection with sucrose for OCT), (ii) embedding (paraffin for routine histology, OCT for Raman-compatible frozen sections, Araldite for EM), (iii) sectioning matched to modality (3–5 μm for LM/IHC/IF; ~ 80 nm for TEM; 1 μm for resin LM), and (iv) substrate selection that fits the physics of each technique (silanized glass for histology/IF; quartz/ CaF_2 or stainless-steel mirrors for Raman; conductive grids for TEM). Staining and labeling (H&E, IHC/IF, silver–Nissl) serve as morphological and molecular references, while label-free SHG/TPEF complements them for collagen/elastin structure.

Application by Modality and Use Case

- FFPE Histology, IHC/IF (Breast, Head & Neck): Dewaxing and antigen retrieval enable robust morphology (H&E) and protein localization. ImageJ quantifies area and integrated density, and H&E provides tumor ROI guidance.
- Raman Spectroscopy (OCT, Unstained): Unstained frozen sections on low-background substrates minimize parasitic signals; spectra are processed with baseline correction, normalization, and PCA. Correlation to H&E is achieved on the same or consecutive section.
- Label-Free Nonlinear Optics (SHG/TPEF/pSHG): Paraffin sections are imaged prior to staining to preserve label-free contrast; serial sections and pSHG metrics support comparison of collagen organization with H&E/TEM.
- Electron Microscopy (Araldite/TEM): Resin embedding with heavy-metal contrasting preserves ultrastructure; ultrathin sections provide nanoscale context that explains optical observations.
- Neuro Architectonics (Silver + Nissl + GFAP): Silver pretreatment and impregnation, followed by Nissl and optional GFAP IHC, delineate cortical/striatal layers; ZEN, MATLAB, and statistics (e.g., Origin) support quantitative comparisons.

Correlative Strategy and Validation

The workflow consistently reserves serial sections to enable direct correlation: (i) acquire label-free or spectroscopic data first to avoid dye interference, (ii) stain the same or adjacent section with H&E and/or IHC, and (iii) register regions of interest across modalities. This ensures that biochemical (Raman) and structural (SHG/TPEF, TEM) readouts are anchored to accepted histopathological references.

Image Analysis and AI Readiness

Image processing stacks are matched to the modality: ImageJ for immunostaining metrics; ZEN/iTEM/MATLAB for microscopy quantification and 3D reconstruction; R (hyperSpec) with PCA for spectra. For AI, radiology and computational imaging pipelines employ ethical approvals and anonymization, DICOM→PNG conversion, expert-validated annotations (Labelme/COCO), K-fold training on architectures such as Detectron, and comprehensive performance reporting. These practices create datasets suitable for robust, transferable models.

Common Issues and Mitigations

Key pitfalls include paraffin residues that bias Raman signals, autofluorescence from substrates or tissue, photobleaching in IF, mechanical artifacts (shrinkage or cracking), and degradation during storage. Mitigations include thorough dewaxing, background correction/spectral unmixing, minimizing light exposure, gentle dehydration/staining steps, and strict SOP-driven storage and QA/QC.

Towards a consensus for AI-Ready Preparation

Selecting the optimal preparation is not one-size-fits-all. A consensus among biologists, imaging specialists, and AI modelers is required to balance molecular fidelity, optical contrast, and data uniformity. Results should be benchmarked against H&E, IHC, and validated spectroscopic/optical protocols and be consistent with modern imaging practices to ensure that computational inferences reflect genuine tissue biology rather than preparation artifacts.