

# G OPEN ACCESS

**Citation:** Mandal S, Baker A-M, Graham TA, Bräutigam K (2025) The tumour histopathology "glossary" for Al developers. PLoS Comput Biol 21(1): e1012708. <u>https://doi.org/10.1371/journal.</u> pcbi.1012708

**Editor:** B. F. Francis Ouellette, Origin Bioinformatics, CANADA

Published: January 23, 2025

**Copyright:** © 2025 Mandal et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** TAG and AMB are funded by Cancer Research UK (DRCNPG-May21\_100001). KB was funded by the Swiss National Science Foundation (P500PM\_217647 / 1). SM acknowledges funding from the Institute of Cancer Research's Data Science Initiative. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** I have read the journal's policy and the authors of this manuscript have the following competing interests: TAG and AMB are named as coinventors on a patent application that describe a method for TCR sequencing (GB2305655.9), and TAG is named on a method to measure evolutionary dynamics in cancers using

EDUCATION

# The tumour histopathology "glossary" for Al developers

## Soham Mandal<sup>1,2</sup>\*, Ann-Marie Baker<sup>1</sup>, Trevor A. Graham<sup>1</sup>, Konstantin Bräutigam<sup>1</sup>\*

1 Centre for Evolution and Cancer, Institute of Cancer Research, London, United Kingdom, 2 Data Science Team, Institute of Cancer Research, London, United Kingdom

\* soham.mandal@icr.ac.uk (SM); konstantin.braeutigam@icr.ac.uk (KB)

# Abstract

The applications of artificial intelligence (AI) and deep learning (DL) are leading to significant advances in cancer research, particularly in analysing histopathology images for prognostic and treatment-predictive insights. However, effective translation of these computational methods requires computational researchers to have at least a basic understanding of histopathology. In this work, we aim to bridge that gap by introducing essential histopathology concepts to support AI developers in their research. We cover the defining features of key cell types, including epithelial, stromal, and immune cells. The concepts of malignancy, precursor lesions, and the tumour microenvironment (TME) are discussed and illustrated. To enhance understanding, we also introduce foundational histopathology techniques, such as conventional staining with hematoxylin and eosin (HE), antibody staining by immunohistochemistry, and including the new multiplexed antibody staining methods. By providing this essential knowledge to the computational community, we aim to accelerate the development of AI algorithms for cancer research.

# Introduction

Histopathology—the microscopic analysis of tissue samples to diagnose and study diseases—is the mainstay of cancer diagnosis, and most clinical practices revolve around expert human pathologists examining very thin tissue slices ("sections") mounted on glass slides under traditional light microscopes.

The last decades have experienced a huge rise in the application of bioinformatics, artificial intelligence (AI), and machine learning (ML) in cancer research [1–3]. A plethora of AI tools have been developed specifically for histopathology, which aim to improve its diagnostic reliability and accuracy [4–6]. Tissue, i.e., the sum and interplay of cells, and cellular morphologies are the key variables in histopathology [7] and are the foundation for any supervised modelling attempt. Thus, we believe that successful AI development requires an understanding of tumour histology.

Building accurate AI models for medicine is an interdisciplinary task [8] and requires a complement of different expertise, i.e., computational and mathematical skills combined with clinical knowledge. This is often hampered by computer scientists and biologists/expert healthcare professionals not physically working in the same environment, where they cannot

DNA methylation (GB2317139.0). TAG has received honorarium from Genentech and DAiNA therapeutics.

easily access the other's expertise. The lack of interdisciplinary communication results in significant inefficiencies and misunderstandings, leading many AI models to remain prototypes rather than being integrated into clinical practice [9,10]. In light of the rise of AI in the clinic, teaching medics AI basics has become essential [11]; here, we propose that a similarly reciprocal understanding of biology is important for AI developers to build better models.

Recent advancements in AI research have significantly automated a range of diagnostic (defining disease), predictive (predicting the response to a certain treatment [12]), and prognostic (stratifying patients at risk and determining outcome/patient prognosis) tasks in oncology [13–15]. However, these models often operate as "black boxes," providing little transparency regarding their decision-making processes. Physicians, on the other hand, base their diagnoses on well-defined biological features and established criteria [16], especially in the case of histopathologists, whose cornerstone is tissue and cell morphology [17]. Therefore, the interpretability of both the design and the results of AI-based methods are crucial for gaining the trust and acceptance of the medical community and accelerating the clinical implementation of in-silico approaches. A good level of understanding of tumour biology instructs AI developers to incorporate relevant biological features into their computational models [18]. This understanding not only improves the accuracy and relevance of the models but also facilitates the presentation of results in a manner that physicians can validate and understand. For example, a study on early-stage oestrogen receptor-positive (ER+) breast cancer demonstrated how considering relevant biological features is crucial in the development of AI methods for survival prediction [19]. The method described in this work leveraged understanding of nuclear pleomorphism (variance in the appearance of the cell nucleus), which is a crucial factor in breast cancer grading.

This article aims to bridge the gap between the AI development and the translation to routine clinical application by emphasising the importance of relevant biological knowledge, which would be helpful in enhancing model interpretability and the subsequent clinical validation. Only solid biological understanding would enable modellers to define sets of relevant features and implement their morphological properties into algorithms.

While recently published literature aims at bestowing (cancer) healthcare professionals with expertise on AI, e.g., developing image analysis and modelling skills for clinicians [20,21], the opposite—giving AI developers an understanding of cancer histopathology fundamentals —is rare. Thus, in this work we introduce some of the essential concepts of tumour histopathology—the most frequent cell types, the concepts of "neoplasia," "tumours," and the tumour microenvironment (TME). We also illustrate routine histopathology protocols and special stainings that provide the visual representation of the above concepts.

# Concepts of tumour histopathology and representative examples of modelling them

In order to model disease, firstly, a solid understanding of cell types, their physiological function, overall architecture and interplay with other cells is necessary. Parameters for image analysis and neural network training are best derived by applying knowledge of their defining morphology and distinct, if not unique, individual features (the common proverb of "the eyes can't see what the mind doesn't know" applies). In the first section, we introduce different cell types and the concept of "neoplasia."

# Morphological diversities in cell types and tissues

Most human cells consist of nucleus and cytoplasm (Fig 1), both of which are organised into different compartments and organelles and surrounded by the cell membrane. The size of the



**Fig 1. The morphological spectrum of different cell types (representative example using HE staining of a human colon).** Regions of interest with different cell types are highlighted with coloured boxes, i.e., inflammatory cells (lymphocytes, macrophages, eosinophils, neutrophils, and plasma cells), epithelial cells (here, an enterocyte forming the cellular unit of crypts), endothelium, and stromal fibroblasts. Prominent cytoplasm (asterisk) in a representative macrophage. See <u>Table 1</u> and section "The hematoxylin and eosin (HE) stain" for further description. Overview scanned on a *Hamamatsu Nanozoomer* whole slide scanner (Hamamatsu Photonics, Hamamatsu, Japan), magnification: 40×; individual cell types captured on a *Zeiss Axio Imager Z2* microscope platform (Carl Zeiss AG, Oberkochen, Germany), magnification: 63×. *Arrow*: Blood vessel lined with endothelial cells (grey box). *Asterisks*: Cytoplasm. *Crosses*: nucleus. Partly created with *BioRender.com*.

https://doi.org/10.1371/journal.pcbi.1012708.g001

Cell type	Distinguishing morphological, cytological, and spatial characteristics	Immunohistochemistry	Disease setting	
Epithelial cell	Distinct shape (polyhedral geometry), pink cytoplasm, ovoid shape	Cytokeratin/s	e.g., colorectal adenoma, carcinoma (Fig 2)	
Lymphocyte	Small, round, very intense stained large nuclei (condensed chromatin), scant amount of cytoplasm. Usually scattered but can aggregate (e.g., in "follicles"). Can cluster at inflammation sites	e.g., CD3, CD4, CD8, CD20, CD45, FOXP3 ( <b>Table 3</b> )	e.g., lymphoma, (chronic) inflammation, tumour-infiltrating lymphocytes (TILs)	
Neutrophil	Multi-lobed nucleus and (pale looking) granular cytoplasm. Scattered, can cluster at inflammation sites	e.g., Myeloperoxidase (MPO)	e.g., acute inflammation, tumour- associated neutrophils (TANs)	
Eosinophil	Bi-lobed nucleus, large and bright red-stained granules. Scattered, can cluster at inflammation sites	e.g., Major basic protein	e.g., acute and chronic inflammation, Eosinophilic Pneumonia, TME component [78]	
Fibroblast	Spindle-shaped, elongated nucleus. Scattered, part of connective tissue	e.g., Vimentin	e.g., fibrosis, cancer-associated fibroblasts (CAFs)	
Endothelial cell	Elongated, and flattened, clear cytoplasm, lining the lumen of (blood and lymphatic) vessels	e.g., CD31, CD34, D2-40, ERG	e.g., part of the TME (microvessels), neoplastic vasculature	
Macrophage	Abundant cytoplasm, large size, phagocytosed particles	e.g., CD11b, CD68, CD163	Tumour-associated macrophages (TAMs), Whipple disease, Rosai– Dorfman disease, etc.	
Plasma cell	Eccentric nucleus, "clock-face" chromatin, abundant cytoplasm	e.g., CD38, CD138, MUM1	e.g., multiple myeloma, chronic inflammation	
Mast cell	Relatively large ovoid cells, cytoplasmic granules, central nucleus	e.g., cKIT (CD117), Tryptase	e.g., mastocytosis	

Table 1. Cell types, their morphology (Fig 1), detection, and disease association. *CD*, Cluster of Differentiation; *HE*, hematoxylin and eosin; *MUM1*, Multiple Myeloma Oncogene 1; *SMA*, smooth muscle antigen; *TME*, tumour microenvironment.

https://doi.org/10.1371/journal.pcbi.1012708.t001

nucleus exhibits a considerate amount of variability. Normal cells mostly display smooth nuclear contours and smaller size. Cancerous cells, on the other hand, tend to exhibit larger and pleomorphic (i.e., bizarre looking) nuclei with a prominent nucleolus (the spherical site for ribosome biogenesis). Further, there is variation in nuclear shape among different cell types. Fibroblasts, which are key components of connective tissue, have a spindle shape, whereas epithelial cells tend to be more round or oval (Fig 1). Second, the cytoplasm varies significantly in size and composition. Eosinophilic granulocytes typically feature a bilobed (two-lobed, spectacle-shaped) nucleus, while macrophages can be recognised by their large cytoplasm (Table 1). Overall, standard cellular morphology reveals distinct sets of features to build AI models upon for cell phenotyping. Usually, these AI models [22,23] consist of 2 sub-models, one for cell segmentation and the other for cell classification using the segmentation results. Recently, substantial progress has been made to develop unified models [24,25] for cell segmentation and classification simultaneously. However, cell phenotyping still faces significant challenges, which include but are not limited to the scarcity of annotated large scale data sets, the significant morphological heterogeneity within cell types, and the complex spatial relationships between cell types and their microenvironment [26].

Human tissues, i.e., functional units of synergistically working cells, are composed of collections of cells that in a non-diseased state have an ordered arrangement in space. Roughly, tissue can be subsumed into 2 major compartments; parenchyma, i.e., the functional part composed of specialised cells, and stroma, i.e., the supporting part, mainly connective tissue, extracellular matrix and (micro)vessels. While stroma is morphologically similar across tissue types, the architecture of the parenchyma can have drastic differences. As an example, breast parenchyma consists of lobules and ducts for lactation, whereas parenchyma in the heart is mainly cardiac muscle. In short, tissue function defines the composition of the parenchyma and vice versa.

### What are "tumours"?

"Tumour" (Latin for "swelling") is an ill-defined term, in principle designating an increase in tissue volume. It refers to a neoplastic process ("neoplasia" being the abnormal and excessive growth of cells and tissue), whose biological "potential" is in most cases dichotomously classified as either benign (localised without metastatic potential, e.g., a minute hyperplastic polyp in the colon) or malignant (invading neighbouring tissue and/or moving to distant organs, e.g., colorectal cancer). The cells of origin for a neoplasia can be classed as epithelial, lymphoid (blood cells), or mesenchymal (connective tissue). Understandably, a large number of AI models focus on the most frequent cancer types [27], which are epithelial in origin and solid, i.e., mass-forming. Non-solid neoplasia are, for instance, cancers of the blood system, which are not localised and not confined to a single organ, e.g., leukemia where neoplastic cells are in circulation in the blood. In the following, we concentrate on solid neoplasia, due to its epidemiologic relevance and localised anatomy.

Histologically, solid neoplasia is composed of the tumour parenchyma, i.e., the neoplastic cells themselves, and the tumour stroma (Fig 2A and 2B). The tumour stroma has gained more and more attention in cancer research [28], and is mainly composed of cells of the tumour microenvironment (TME, see below), extracellular matrix (structural and specialised proteins surrounding units of cancer cells), and connective tissue (mainly collagen fibres, fibroblasts, and microvessels).

# What is "tumour invasion" and its predecessors?

Tumour invasion is the critical step to a malignant phenotype in epithelial neoplasia (crudely referred to as "cancer"). Illustrative examples are the malignant transformation of (colorectal) adenomas to carcinomas [29], loss of basal cells in prostate cancer [30], loss of myoepithelial cells in breast cancer [31], or crossing anatomical barriers, e.g., the basal layer in skin squamous carcinoma (Fig 2C). For those interested in modelling diagnostic AI support, knowledge of these anatomical structures is critical. The features that are truly unique to invasive cancers should be exploited—for example, the presence of features such as necrosis (dead cells) or an abundance of mitoses (dividing cells) do not imply malignancy, as they can be frequent in benign neoplasms.

Invasion is usually the final step in a sequence of malignant transformation. In epithelial tumours (e.g., gastric adenocarcinoma [32]), it is frequently preceded by, first metaplasia, and second dysplasia (definitions in S1 Table). Epithelial dysplasia, i.e., simply put the presence of abnormal but yet not cancerous cells [33], is a frequent precursor lesion, e.g., in upper and lower gastrointestinal tract, genital tract, skin, and the head and neck region. However, the defining morphological criteria of dysplasia differ by tissue type, and modelling approaches need to take this into account (Table 2 and Fig 3). For instance, while hyperchromasia (darker staining) is a feature of dysplastic nuclei in the colon, this does not hold true for squamous dysplasia, where architectural disorders and mitoses are more diagnostic. The modeller needs a solid knowledge about which morphological criteria are defining in a respective organ, as concepts are not necessarily transferable. We provide a few examples of such features in Table 2.

# The "face" of malignancy—Morphology and pitfalls

Not every malignancy follows the conventional benign-intermediate-malignant trajectory described above, examples being de novo (i.e., not arising from a precursor tumour, such as a skin mole ("naevus")) malignant melanoma [34] or sarcoma (relatively rare malignant tumours of the soft tissue or bone). Briefly, malignant tumours can be classified according to their cell(s) of origin, e.g., epithelial, mesenchymal (simply put, connective tissue), or



**Fig 2. The morphology of "cancer," HE.** (A) Nonneoplastic epithelium (left) versus adenocarcinoma glands/cells (right, separated by dotted line) in a *The Cancer Genome Atlas* (TCGA) sample of colorectal adenocarcinoma (COAD, *TCGA-A6-2678*). Dashed *inset* (corresponding to the dashed rectangle): Higher magnification showing malignant COAD glands with malignant nuclear features and a focus of "dirty" necrosis within the cancer gland (asterisk). *Insets*: On the one hand, enterocytes (e, colonic epithelium) with regular contours, polarised towards the lumen with a goblet (mucus-containing) cell (G) on top. On the other hand, cancer cell (c) with a prominent nucleolus (cross), large size and irregular contours. *Arrow*: Regular colonic crypts, rich in goblet cells (arrow on goblet cell). *Arrowhead*: High-grade dysplasia at the surface

mucosa. (**B**) Tumour parenchyma (red) and tumour stroma (blue, green) in pancreatic adenocarcinoma (PAAD, *TCGA-2J-AAB6*). PAAD glands are surrounded by dense "desmoplasia" (blue, inset) with pinkish collagen deposition (black arrowhead) and immune cell infiltrate (green, inset), mainly lymphocytes (white arrowheads) among microvessels (arrows). (**C**) Phenotype of invasion. From left to right: Invasive breast cancer (BRCA) clusters (BRCA, *TCGA-3C-AALJ*), prostatic adenocarcinoma (PRAD, *TCGA-EJ-7792*) glands with loss of basal cell layer (*arrows*: benign prostatic parenchyma), invasive squamous cell carcinoma of the cervix (CESC, *TCGA-C5-A7XC*) with adjacent high-grade squamous intraepithelial lesion (HSL, white asterisk) and of the head and neck region (HNSC, *TCGA-BA-4076*). See <u>Table 2</u> and section "The hematoxylin and eosin (HE) stain" for further description. *Asterisks*: Surface squamous epithelium, *arrowheads*: invasive SC. *Crosses*: Extracellular matrix/stroma between BRCA clusters.

https://doi.org/10.1371/journal.pcbi.1012708.g002

lymphoid (see above). It's reasonable to assume that for developers and the healthcare system, epidemiologically frequent malignant tumours are most relevant, these being malignant epithelial tumours, namely (adeno)carcinomas, e.g., of the prostate, breast, colon (rectum), and lung [21]. Adenocarcinomas are a morphological (and also molecularly distinct) subset of carcinomas (e.g., in contrast to squamous cell carcinomas) and have a typical morphology with gland-like (i.e., circularly arranged with a central lumen) growth with malignant nuclear features such as bizarre cell forms, mitotic figures, and a prominent nucleolus (Fig 2A); however, there are always exceptions and this can make modelling tricky and algorithms incompatible with routine diagnostic practice. Illustrative exceptions are deviant subtypes or atypical cancer growth patterns, such as high-grade prostatic adenocarcinoma with diffuse growth [35], or invasive-lobular breast cancer with discohesive tumour cells [36]. Not incorporating these features into algorithms is severely limiting and might jeopardise clinical conclusion from a computer-assisted diagnostic setup. Non-small cell lung cancer (NSCLC), second most frequent cancer in both sexes in the United States, has a known multitude of growth patterns [37] and also colorectal cancer can exhibit a very deviant morphology, even if rare [38]. In addition,

Biological object	Cellular detail	Architectural detail	Stromal detail	Additional identifiers	Relevant modelling considerations
Dysplasia in the gastrointestinal tract	Nuclear hyperchromasia [79], cell form (e.g., "pencillate" in intestinal adenomas [80])	Can be polypoid or flat	Evtl. associated inflammation		Staining intensity of epithelial cells to quantify hyperchromasia. Eccentricity, circularity, and elongation to detect pencil-like nuclei.
Dysplasia in the genital and head and neck region	Mitoses, irregular cell form, loss of cell polarity [81]	Loss of stratification of squamous epithelium	Evtl. associated inflammation	Atypical keratinisation	Detection of mitotic cells [82,83]. Cell- based morphological features (such as eccentricity, convex area, contour area, extent, perimeter, solidity, and orientation) [84]. Quantification of cell polarity [85].
Tumour invasion	Invasive "phenotype" (atypical nucleus, nucleolus), mitoses	(Partial) loss of anatomic border, e.g., a basal membrane [86]	"Desmoplasia," potential immune response		Proportion of continuous versus disrupted basal membrane/layer, "desmoplastic" stroma [87].
Adenocarcinoma	Gland-forming, mucus	Complex clustering of carcinoma glands	"Desmoplasia," potential immune response	"Dirty necrosis" (not specific for a cancer type) [88]	Glandular shape descriptors [89]. Nuclear shape descriptors (bizarre form, nucleolus). Mucus can be used as a proxy.
Special subtypes and growth patterns of adenocarcinoma	Signet ring cells (intracellular mucus, compressed nucleus), giant cells	e.g., (micro)papillary, tubular, acinar, lepidic	"Desmoplasia," potential immune response		Detection of signet ring cells [90]. Detection of specific and unique growth patterns (such as lepidic, papillary, acinar, cribriform, micropapillary and solid [91]).
Squamous cell carcinoma	Intercellular bridges	Keratin "pearls," specific growth pattern (e.g., basaloid)	"Desmoplasia," potential immune response		Detection of keratin pearls [92]. Detection of intercellular bridges [93].

https://doi.org/10.1371/journal.pcbi.1012708.t002



Fig 3. Example of computational modelling of an oncologic precursor lesion (low-grade dysplasia (LGD), human colon) with feature extraction. (A) Conventional features of (sporadic) colonic LGD, i.e., hyperchromasia, next to normal colonic mucosa (separated by black line) with inconspicuous enterocytes and crypts (representative example of an inflammatory bowel disease cohort, HE stain). *Arrows*: Hyperchromatic, pseudostratified (pencillate) nuclei in LGD.
(B) Nuclei detection, produced using a deep neural network [77]. *Inset*: High magnification. (C) Nuclei intensity mapping. Computed from the segmentation mask in B by normalising the nuclei intensities across the whole slide image. LGD shows higher intensities, histologically corresponding to hyperchromasia (darker staining in HE).

https://doi.org/10.1371/journal.pcbi.1012708.g003

adenocarcinoma is frequently accompanied by a strong stromal response ("desmoplasia"), which can be seen by collagen deposition and extracellular matrix (ECM) recomposition [39] (Fig 2B and Table 2). Squamous cell carcinoma, most prominently in the head and neck region (Fig 2C), genital tract (cervical cancer, anal cancer) and lung is characterised by keratin "pearls" (whorl-shaped accumulations of keratin, a structural protein), and intercellular bridges (specialised connections between adjacent cells). It has to be kept in mind that even very typical diagnostic features might not be apparent in poorly differentiated tumours which have lost much of their resemblance to the tissue of origin ("dedifferentiation" is the process of losing tissue specialisation, returning to a less specialised state). Mixed carcinomas, e.g., adenos-quamous, adeno-neuroendocrine (e.g., mixed neuroendocrine-nonneuroendocrine neoplasms [40]), add further to the complexity. Lastly, malignancy of other lineages, e.g., non-epithelial shows a considerable amount of variation, too. In particular, malignant melanoma is known for its plethora of "morphological faces" [41].

# Composition of, and modelling the tumour microenvironment (TME)

The TME is the complex biological ecosystem surrounding a tumour. It is composed of tumour-infiltrating lymphocytes ("TILs") [42], (cancer-associated) fibroblasts and (tumour-associated) neutrophils (CAFs, TANs) [43,44], macrophages, extracellular matrix, and supportive elements such as microvessels. All of those have gained significant attention due to their prognostic, tumour-promoting or -suppressive impact. While macrophages have traditionally been dichotomously subclassified (i.e., M1- and M2-polarised [45]), TILs can be more deeply sub-stratified.

Modelling the TME needs a comprehensive strategy due to its inherent level of complexity and set of "players." Nevertheless, modelling using morphology is possible to a certain degree as the "players" generally have distinct cellular and architectural features (<u>Table 1</u>). In the following, we introduce routine stainings and immunohistochemistry that can facilitate the morphological characterisation of tumours and their TME.

# **Routine tissue protocols**

# Formalin-fixed paraffin-embedded (FFPE) and fresh frozen (FF) tissue

Formalin-fixed, paraffin-embedded (FFPE) tissue preservation is the gold standard in histopathology for maintaining tissue integrity. This technique, first introduced by German pathologist Friedrich Blum in 1896 [46], involves fixing tissue samples in formalin, which preserves their cellular structure by cross-linking proteins. The fixed tissues are then dehydrated, embedded in paraffin wax, and formed into a solid, archivable tissue block. These blocks can be sectioned into thin slices (usually between 2 and 10  $\mu$ m [47]), mounted onto glass slides, and stained allowing for microscopic examination. Fresh frozen (FF) tissue on the contrary is immediately preserved by snap-freezing at -196°C in liquid nitrogen (e.g., cancer tissue within 1 h from surgery), without (formalin) fixation. As FF is tissue in its purest form, it is more accurate for genomic analysis than FFPE. However, FFPE preserves better structural integrity and is much more standardised (and affordable) for conventional staining and immunohistochemistry (see below). Despite these advantages, it is crucial to consider that the number of (viable) cells of tissue samples is highly heterogeneous and depends not only on the tissue (and biopsy site) itself but also on how it is retrieved [48]. In general, surgical resection specimens collected in conventional FFPE tissue cassettes (approximately  $1 \times 1 \times 0.5$  cm) tend to contain the most cells, while the amount of cells in tissue biopsies is much reduced and limited by several factors such as, for instance, the gauge of the biopsy needle [49], the anatomical site (e.g., soft tissue is less cellular than bone marrow), and the expertise of the operator [50,51].

# The hematoxylin and eosin (HE) stain

The hematoxylin and eosin (HE) stain is the standard staining that has been used in (diagnostic) histopathology for many years [52]. While hematoxylin stains acidic structures, e.g., the nucleus, in different degrees of blue-purple, eosin stains basophilic structures in red-pink, such as the cytoplasm and ECM (**Fig 1**). This allows for the identification of common cell types and their arrangement in space. The HE stain is cheap, widely used and well accepted in the diagnostic community [53]. A low amount of staining variability is critical for both diagnostics and AI algorithms [54].

## Immunohistochemistry (IHC) and immunofluorescence (IF)

HE stains allow for a vast amount of tissue interpretability, but in order to address cells and their interplay more granularly, auxiliary information can be obtained from immunohistochemistry (IHC). (Single-plex) IHC has revolutionised histopathology in the 20th century [55] and continues to be an indispensable tool. In principle, IHC detects a target antigen of interest (e.g., membrane transporters, enzymes) by using a chromogen-linked commercial antibody that binds to the antigen of interest and "staining" it a particular colour—usually brown (Fig 4A and Table 3) [56]. The target of interest could be in tumour cells or in cells of the TME.

While IHC uses enzymes as chromogens, immunofluorescence (IF) uses fluorescent dyes (fluorophores) conjugated to antibodies. Advantages of IF are higher resolution and an improved visualisation of co-localised antigens. On the contrary, IHC stainings are long-lasting, cheaper and can be viewed by light microscopy. Some antigens are of particular relevance for diagnostic and (consequently also) AI-developing purposes, namely the proliferation marker *Ki-67* [57,58] or lineage markers such as cytokeratins [59] (Table 3). "Clusters of differentiation" (CDs) are surface proteins that can help with subtyping cells, particularly immune cells (refer to https://ftp.uniprot.org/pub/databases/uniprot/knowledgebase/ complete/docs/cdlist.txt). This is useful as there is little potential to identify an immune cell subpopulation from an HE alone. IHC staining should be validated extensively, as some antibodies tend to cross-react among different targets leading to lack of specificity and misleading results [60]. Internal on-slide controls can be helpful as a quality control, such as cross-reacting (stromal) cells.

Tissue microarrays (TMAs) (Fig 4A), i.e., assembling a multitude of usually 0.6 to 1 mm sized tissue cores into a single slide, have allowed for a high-throughput setup [61]. Depending on the tissue type and anatomic site, a TMA core usually captures between a few hundred and a few thousand cells per tissue core [62]. TMAs allow multiple stainings and tissues (from different patients) can be analysed under standardised conditions [63,64].

# The advent of multiplexing

Recently, antigen visualisation reached a new era in which we can detect up to hundreds of markers of interest in one section of tissue (recent comprehensive review in [65]) (Table 3). Basic panels visualise cells of interest and key anatomic structures [66], for instance, an epithelial, a pan-leukocyte marker and vessels (e.g., a cytokeratin, CD45, CD31). Multiplexing allows for multiple panels which represent different compartments (Fig 4B and 4C and Table 3) characterising the TME [67] and its neighbourhoods [68]. This increasing plexity allows for the interrogation of biology in more detail, but results in more complex data sets than HE. This could present significant challenges for AI models, which include the demand for larger training data sets (and thus higher computational power) to reach a similar level of performance seen for HE, higher likelihood of technical artefacts due to more complex wet lab



**Fig 4. Immunohistochemistry (IHC) and multiplexing.** (A) Singleplex IHC and tissue microarrays (TMAs). Left: Preserved nuclear expression of the mismatch repair protein MSH6 in a colorectal adenocarcinoma sample (asterisks). *Arrow*: Preserved expression in stromal fibroblasts. On the right, the principle of TMA construction: Tissue cores, usually 6 to 10 mm in diameter, are punched and transferred to a new FFPE block, which is composed of multiple patients and tissues. *Inset*: Overview of colorectal cancer tissue and illustrative punches. (B) Basic multiplex immunofluorescence (mIF) panel on colorectal mucosa consisting of 2 markers and DAPI (blue) as a standard nuclear stain. CD20 (light blue) labels B-cells and a lymphoid follicle (arrowheads). Beta-catenin stains colonic crypts (red,

arrows). (C) Extended mIF panel with 6 markers and DAPI. Beta-catenin (red) highlights colorectal adenocarcinoma glands (asterisks), while the space between malignant glands is composed of immune cells, such as B-cells (CD20, light blue), and macrophages (CD68, green, black arrowheads). CD31 (orange) highlights blood vessels (grey arrowheads, intermixed with SMA, yellow) and might suggest vascular invasion. The stroma is represented by Vimentin (purple, grey arrowheads)-positive fibroblasts and smooth muscle (SMA, yellow. arrow). *CD*, Cluster of Differentiation; *SMA*, smooth muscle actin.

https://doi.org/10.1371/journal.pcbi.1012708.g004

protocols, and the difficulty of jointly modelling multiple cellular markers and their spatial relationships. Further, different markers often co-localise, which introduces additional difficulties to the modelling process [69]. Thus, compared to HE, analysing both IHC and IF data is usually harder and more expensive. A challenge further exacerbated by increasing plexity.

# **Challenges and outlook**

To be implemented into routine practice, an AI algorithm needs several indispensable properties, i.e., clinical relevance, high accuracy, rapid implementation, fast computation, and last but not least, user-friendliness. Perfect accuracy is desired for pathological diagnostics, such as differentiating between tumour invasion and benign disease; anything less could put patients' lives at risk. False negatives that lack a prognostic biomarker may lead to reduced therapeutic

**Table 3. Representative markers for multiplexing panels (Fig 4).** *ALK*, anaplastic lymphoma kinase; *AR*, androgen receptor; *ARID1A*, AT-rich interactive domain-containing protein 1A; *Bax*, Bcl-2-associated X protein; *CA*, carboan-hydrase; *CD*, Cluster of Differentiation; *CK*, cytokeratin; *EGFR*: Epidermal Growth Factor Receptor; *ER*, *oestrogen receptor*; ERG, ETS-Related Gene; *FOXP3*, forkhead box P3; *GLUT1*, glucose transporter member 1; *INSM1*, insulinoma-associated protein 1; *LCA*, leukocyte common antigen; *LDH*, lactate dehydrogenase; *MCT*, moncarboxylate transporter; *mdm2*, mouse double minute 2 homolog; *MMR*, mismatch repair; *MEK*, mitogen-activated protein kinase kinase enzyme; *MPO*, myeloperoxidase; *MUM1 (IRF4)*, multiple myeloma 1 (interferon regulatory factor 4); *PD-1/PD-L1*, programmed death/ligand-1; *PHH3*, phospho-histone H3; *PgR*, progesterone receptor; *PTEN*, phosphatase and tensin homolog; *Rb1*, retinoblastoma protein; *SMA*, smooth muscle antigen; *SOX10*, SRY related HMG box 10.

Target of interest	Antigen/Marker				
Basic anatomical structures	Blood/Lymphatic vessels: CD31, CD34, ERG / D2-40 (Podoplanin) Connective tissue: e.g., collagens (Smooth) muscle: Desmin, SMA Nerves: S-100, SOX10				
Immune infiltrate (see <u>Table 1</u> )	B-cells: e.g., CD19, CD20 Cytokines: e.g., Interleukin/s (II) such as II-17 for Th17 cells Macrophages: e.g., CD11b, CD68, CD163 Neutrophils: e.g., CD15, MPO Natural Killer (NK)-cells: e.g., CD16, CD56, CD57 Plasma cells: e.g., MUM1 (IRF4) T-cells: e.g., CD3, CD4 (T-helper), CD8 (cytotoxic), CD45 (LCA), Granzyme B (activated T- and NK-cells) T-regulatory cells: e.g., FOXP3				
Lineage	Epithelial: e.g., Cytokeratin/s Melanocytic: e.g., Melan-A, S-100, SOX10 Mesenchymal: e.g., Desmin, SMA Neuroendocrine: e.g., Chromogranin A, INSM1, Synaptophysin				
Tumour invasion	Calponin, CK5/6, p63 (basal cell layer loss in prostate cancer, myoepithelial cell loss in breast cancer), collagens (basal membrane)				
Proliferation/Apoptosis	e.g., proliferation index Ki67, PHH3 (mitoses)/Caspases (such as Caspase-3) or Bax, Fas (CD95)				
Drug response and prediction	e.g., Her-2, hormone receptors (e.g., AR, ER, PgR), MMR proteins, PD1/PD-L1				
Surrogates for molecular alterations	e.g., ALK, ARID1A, BRAF (V600E), EGFR, mdm2, MEK, p53, PTEN, Rb1				
Hypoxia and metabolism	e.g., CA9, GLUT1, LDH, MCT1, MCT4				

https://doi.org/10.1371/journal.pcbi.1012708.t003

options. A variety of technical difficulties, such as staining differences, scanner variability, image modalities, and image size [70,71] hinder the performance of AI models, including their generalisation to different datasets. For these models to achieve robust generalisation across different datasets, several key standardisation approaches are required throughout the imaging and analysis pipeline, such as introducing standardisation of tissue processing, sectioning thickness, reagents, fixation protocols, scanner calibration, and performing stain normalisation [72-74]. Further, multi-centre validation data sets that are able to represent real-world technical variations could also help in developing and validating more generalised AI models. Aside from the need of standardisation, the AI developer is frequently confronted with the profound problem of lacking in-depth biological, morphological, and structural knowledge. It is our hope that this work enables the developer to leverage biologically relevant features into designing computational models. It is our hope that this work enables the developer to leverage biologically relevant features into designing computational models. Further, this becomes helpful in explaining the output from "black-box" AI models, by correlating the results with known biological features. With a common knowledge level, the design of "pathologist in-theloop" approaches [75,76] in training AI models are facilitated.

# **Supporting information**

**S1 Table.** A glossary of different "-plasias" with examples. (DOCX)

# Acknowledgments

The figures shown here are in whole or part based upon data generated by the TCGA Research Network: <u>https://www.cancer.gov/tcga</u>. The authors thank Chirine Sakr for providing access to multiplex immunofluorescence images, Louise Howell and Emma Westlund (ICR Light Microscopy Facility) for assistance with high-resolution light microscopy, and Ashley Wong for proofreading and feedback.

# **Author Contributions**

**Conceptualization:** Soham Mandal, Ann-Marie Baker, Trevor A. Graham, Konstantin Bräutigam.

Funding acquisition: Trevor A. Graham, Konstantin Bräutigam.

Investigation: Soham Mandal, Ann-Marie Baker, Trevor A. Graham, Konstantin Bräutigam.

Methodology: Soham Mandal, Ann-Marie Baker, Trevor A. Graham, Konstantin Bräutigam.

Supervision: Trevor A. Graham, Konstantin Bräutigam.

Visualization: Soham Mandal, Ann-Marie Baker, Konstantin Bräutigam.

Writing – original draft: Soham Mandal, Ann-Marie Baker, Trevor A. Graham, Konstantin Bräutigam.

Writing – review & editing: Soham Mandal, Ann-Marie Baker, Trevor A. Graham, Konstantin Bräutigam.

#### References

1. Foronda M. The AI revolution in cancer. Nat Res [Internet]. 2020 Dec 10 [cited 2024 Jul 11]; Available from: https://www.nature.com/articles/d42859-020-00082-9.

- Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012 Jul; 9(7):671–675. https://doi.org/10.1038/nmeth.2089 PMID: 22930834
- Landhuis E. Deep learning takes on tumours. Nature. 2020 Apr 21; 580(7804):551–553. <u>https://doi.org/10.1038/d41586-020-01128-8</u> PMID: 32317799
- Shmatko A, Ghaffari Laleh N, Gerstung M, Kather JN. Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. Nat Cancer. 2022 Sep; 3(9):1026–1038. https://doi.org/10.1038/ s43018-022-00436-4 PMID: 36138135
- Echle A, Rindtorff NT, Brinker TJ, Luedde T, Pearson AT, Kather JN. Deep learning in cancer pathology: a new generation of clinical biomarkers. Br J Cancer. 2021 Feb; 124(4):686–696. https://doi.org/ 10.1038/s41416-020-01122-x PMID: 33204028
- Unger M, Kather JN. A systematic analysis of deep learning in genomics and histopathology for precision oncology. BMC Med Genomics. 2024 Feb 5; 17(1):48. https://doi.org/10.1186/s12920-024-01796-9 PMID: 38317154
- Rosai J. Why microscopy will remain a cornerstone of surgical pathology. Lab Invest. 2007 May; 87 (5):403–408. https://doi.org/10.1038/labinvest.3700551 PMID: 17401434
- Martín-Noguerol T, Paulano-Godino F, López-Ortega R, Górriz JM, Riascos RF, Luna A. Artificial intelligence in radiology: relevance of collaborative work between radiologists and engineers for building a multidisciplinary team. Clin Radiol. 2021 May 1; 76(5):317–324. <u>https://doi.org/10.1016/j.crad.2020.11</u>. 113 PMID: 33358195
- van de Sande D, Genderen MEV, Smit JM, Huiskens J, Visser JJ, Veen RER, et al. Developing, implementing and governing artificial intelligence in medicine: a step-by-step approach to prevent an artificial intelligence winter. BMJ Health Care Inform. 2022 Feb 1; 29(1):e100495. <u>https://doi.org/10.1136/bmjhci-2021-100495 PMID: 35185012</u>
- Reis-Filho JS, Kather JN. Overcoming the challenges to implementation of artificial intelligence in pathology. JNCI J Natl Cancer Inst. 2023 Jun 1; 115(6):608–612. https://doi.org/10.1093/jnci/djad048 PMID: 36929936
- Ötleş E, James CA, Lomis KD, Woolliscroft JO. Teaching artificial intelligence as a fundamental toolset of medicine. Cell Rep Med. 2022 Dec 20; 3(12):100824. <u>https://doi.org/10.1016/j.xcrm.2022.100824</u> PMID: 36543111
- 12. Oldenhuis CNAM Oosting SF, Gietema JA, de Vries EGE. Prognostic versus predictive value of biomarkers in oncology. Eur J Cancer. 2008 May 1; 44(7):946–953.
- Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ, et al. Foundation models for generalist medical artificial intelligence. Nature. 2023 Apr; 616(7956):259–265. <u>https://doi.org/10.1038/s41586-023-05881-4</u> PMID: 37045921
- Chen RJ, Ding T, Lu MY, Williamson DFK, Jaume G, Song AH, et al. Towards a general-purpose foundation model for computational pathology. Nat Med. 2024 Mar; 30(3):850–862. <u>https://doi.org/10.1038/</u> s41591-024-02857-3 PMID: 38504018
- Vorontsov E, Bozkurt A, Casson A, Shaikovski G, Zelechowski M, Severson K, et al. A foundation model for clinical-grade computational pathology and rare cancers detection. Nat Med. 2024 Jul 22:1– 12. https://doi.org/10.1038/s41591-024-03141-0 PMID: 39039250
- Glatzer M, Panje CM, Sirén C, Cihoric N, Putora PM. Decision Making Criteria in Oncology. Oncology. 2018 Sep 18; 98(6):370–378.
- 17. The Pathologist [Internet]. 2021 [cited 2024 Jul 14]. On the Value of Morphology in Pathology. Available from: https://thepathologist.com/diagnostics/on-the-value-of-morphology-in-pathology.
- Yanai I, Chmielnicki E. Computational biologists: moving to the driver's seat. Genome Biol. 2017 Nov 23; 18(1):223. https://doi.org/10.1186/s13059-017-1357-1 PMID: 29169371
- Lu C, Romo-Bucheli D, Wang X, Janowczyk A, Ganesan S, Gilmore H, et al. Nuclear shape and orientation features from H&E images predict survival in early-stage estrogen receptor-positive breast cancers. Lab Invest. 2018 Nov 1; 98(11):1438–1448.
- Perez-Lopez R, Ghaffari Laleh N, Mahmood F, Kather JN. A guide to artificial intelligence for cancer researchers. Nat Rev Cancer. 2024 Jun; 24(6):427–441. https://doi.org/10.1038/s41568-024-00694-7 PMID: 38755439
- 21. Gu Q, Patel A, Hanna MG, Lennerz JK, Garcia C, Zarella M, et al. Bridging the Clinical-Computational Transparency Gap in Digital Pathology. Arch Pathol Lab Med 2024 Jun 14;(Epub ahead of print.). https://doi.org/10.5858/arpa.2023-0250-RA PMID: 38871349
- 22. Sharma H, Zerbe N, Heim D, Wienert S, Behrens HM, Hellwich O, et al. A Multi-resolution Approach for Combining Visual Information using Nuclei Segmentation and Classification in Histopathological Images: In: Proceedings of the 10th International Conference on Computer Vision Theory and Applications [Internet]. Berlin, Germany: SCITEPRESS—Science and and Technology Publications; 2015.

p. 37–46. Available from: http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/ 0005247900370046

- Wang P, Hu X, Li Y, Liu Q, Zhu X. Automatic cell nuclei segmentation and classification of breast cancer histopathology images. Signal Process. 2016 May 1; 122:1–13.
- Graham S, Vu QD, Raza SEA, Azam A, Tsang YW, Kwak JT, et al. Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med Image Anal. 2019 Dec 1; 58:101563. https://doi.org/10.1016/j.media.2019.101563 PMID: 31561183
- 25. Graham S, Vu QD, Jahanifar M, Raza SEA, Minhas F, Snead D, et al. One model is all you need: Multitask learning enables simultaneous histology image segmentation and classification. Med Image Anal. 2023 Jan 1; 83:102685. https://doi.org/10.1016/j.media.2022.102685 PMID: 36410209
- Bareham B, Dibble M, Parsons M. Defining and modeling dynamic spatial heterogeneity within tumor microenvironments. Curr Opin Cell Biol. 2024 Oct 1; 90:102422. <u>https://doi.org/10.1016/j.ceb.2024</u>. 102422 PMID: 39216233
- 27. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin. 2024; 74(1):12–49. https://doi.org/10.3322/caac.21820 PMID: 38230766
- Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov. 2022 Jan 12; 12(1):31–46. <a href="https://doi.org/10.1158/2159-8290.CD-21-1059">https://doi.org/10.1158/2159-8290.CD-21-1059</a> PMID: 35022204
- Nguyen LH, Goel A, Chung DC. Pathways of Colorectal Carcinogenesis. Gastroenterology. 2020 Jan 1; 158(2):291–302. https://doi.org/10.1053/j.gastro.2019.08.059 PMID: 31622622
- Goldstein AS, Huang J, Guo C, Garraway IP, Witte ON. Identification of a Cell of Origin for Human Prostate Cancer. Science. 2010 Jul 30; 329(5991):568–571. <u>https://doi.org/10.1126/science.1189992</u> PMID: 20671189
- Adriance MC, Inman JL, Petersen OW, Bissell MJ. Myoepithelial cells: good fences make good neighbors. Breast Cancer Res. 2005 Jul 12; 7(5):190. https://doi.org/10.1186/bcr1286 PMID: 16168137
- Correa P, Piazuelo MB. The gastric precancerous cascade. J Dig Dis. 2012; 13(1):2–9. https://doi.org/ 10.1111/j.1751-2980.2011.00550.x PMID: 22188910
- Schwab M, editor. Dysplasia. In: Encyclopedia of Cancer [Internet]. Berlin, Heidelberg: Springer; 2011. p. 1178–1178. Available from: https://doi.org/10.1007/978-3-642-16483-5\_1765
- Cymerman RM, Shao Y, Wang K, Zhang Y, Murzaku EC, Penn LA, et al. De Novo vs Nevus-Associated Melanomas: Differences in Associations With Prognostic Indicators and Survival. JNCI. J Natl Cancer Inst. 2016 Oct 1; 108(10):djw121.
- 35. Hansum T, Hollemans E, Verhoef EI, Bangma CH, Rietbergen J, Osanto S, et al. Comedonecrosis Gleason pattern 5 is associated with worse clinical outcome in operated prostate cancer patients. Mod Pathol. 2021 Nov; 34(11):2064–2070. https://doi.org/10.1038/s41379-021-00860-4 PMID: 34175896
- Reed AEM, Kutasovic JR, Lakhani SR, Simpson PT. Invasive lobular carcinoma of the breast: morphology, biomarkers and 'omics. Breast Cancer Res. 2015 Jan 30; 17(1):12. <u>https://doi.org/10.1186/</u> s13058-015-0519-x PMID: 25849106
- Solis LM, Behrens C, Raso MG, Lin HY, Kadara H, Yuan P, et al. Histologic patterns and molecular characteristics of lung adenocarcinoma associated with clinical outcome. Cancer. 2012; 118(11):2889– 2899. https://doi.org/10.1002/cncr.26584 PMID: 22020674
- Ono Y, Yilmaz O. Emerging and under-recognised patterns of colorectal carcinoma morphologies: a comprehensive review. J Clin Pathol. 2024 Jul 1; 77(7):439–451. <u>https://doi.org/10.1136/jcp-2023-208816 PMID: 38448211</u>
- DeClerck YA. Desmoplasia: A Response or a Niche? Cancer Discov. 2012 Sep 10; 2(9):772–774. https://doi.org/10.1158/2159-8290.CD-12-0348 PMID: 22969116
- La Rosa S, Sessa F, Uccella S. Mixed Neuroendocrine-Nonneuroendocrine Neoplasms (MiNENs): Unifying the Concept of a Heterogeneous Group of Neoplasms. Endocr Pathol. 2016 Dec 1; 27(4):284– 311. https://doi.org/10.1007/s12022-016-9432-9 PMID: 27169712
- Magro CM, Crowson AN, Mihm MC. Unusual variants of malignant melanoma. Mod Pathol. 2006 Feb; 19(2):S41–S70. https://doi.org/10.1038/modpathol.3800516 PMID: 16446716
- Brummel K, Eerkens AL, de Bruyn M, Nijman HW. Tumour-infiltrating lymphocytes: from prognosis to treatment selection. Br J Cancer. 2023 Feb; 128(3):451–458. <u>https://doi.org/10.1038/s41416-022-02119-4</u> PMID: 36564565
- **43.** Shaul ME, Fridlender ZG. Tumour-associated neutrophils in patients with cancer. Nat Rev Clin Oncol. 2019 Oct; 16(10):601–620. https://doi.org/10.1038/s41571-019-0222-4 PMID: 31160735
- Yang D, Liu J, Qian H, Zhuang Q. Cancer-associated fibroblasts: from basic science to anticancer therapy. Exp Mol Med. 2023 Jul; 55(7):1322–1332. https://doi.org/10.1038/s12276-023-01013-0 PMID: 37394578

- Chen S, Saeed AFUH, Liu Q, Jiang Q, Xu H, Xiao GG, et al. Macrophages in immunoregulation and therapeutics. Signal Transduct Target Ther. 2023 May 22; 8(1):1–35.
- 46. Blum. Der Formaldehyd als Härtungsmittel. Zschr f Mikro und mikrosk Tech. 1896; 10:314–315.
- Kim SW, Roh J, Park CS. Immunohistochemistry for Pathologists: Protocols, Pitfalls, and Tips. J Pathol Transl Med. 2016 Oct 13; 50(6):411–418. https://doi.org/10.4132/jptm.2016.08.08 PMID: 27809448
- **48.** Fishman JE, Milikowski C, Ramsinghani R, Velasquez MV, Aviram G. US-guided Core-Needle Biopsy of the Breast: How Many Specimens Are Necessary? Radiology. 2003 Mar; 226(3):779–782. <u>https://doi.org/10.1148/radiol.2263011622</u> PMID: 12601206
- 49. Moore W, Sawas A, Lee C, Ferretti J. Needle Gauge and Cytological Yield in CT-Guided Lung Biopsy. Int Sch Res Not. 2011; 2011(1):970813.
- Schneider F, Smith MA, Lane MC, Pantanowitz L, Dacic S, Ohori NP. Adequacy of Core Needle Biopsy Specimens and Fine-Needle Aspirates for Molecular Testing of Lung Adenocarcinomas. Am J Clin Pathol. 2015 Feb 1; 143(2):193–200. https://doi.org/10.1309/AJCPMY8UI7WSFSYY PMID: 25596245
- Ljung BM, Drejet A, Chiampi N, Jeffrey J, Goodson WH III, Chew K, et al. Diagnostic accuracy of fineneedle aspiration biopsy is determined by physician training in sampling technique. Cancer Cytopathol. 2001; 93(4):263–268. https://doi.org/10.1002/cncr.9040 PMID: 11507700
- Chan JKC. The Wonderful Colors of the Hematoxylin–Eosin Stain in Diagnostic Surgical Pathology. Int J Surg Pathol. 2014 Feb 1; 22(1):12–32. https://doi.org/10.1177/1066896913517939 PMID: 24406626
- de Haan K, Zhang Y, Zuckerman JE, Liu T, Sisk AE, Diaz MFP, et al. Deep learning-based transformation of H&E stained tissues into special stains. Nat Commun. 2021 Aug 12; 12(1):4884.
- Azevedo Tosta TA, de Faria PR, Neves LA, do Nascimento MZ. Computational normalization of H&Estained histological images: Progress, challenges and future potential. Artif Intell Med. 2019 Apr 1; 95:118–132.
- Salto-Tellez M, Maxwell P, Hamilton P. Artificial intelligence—the third revolution in pathology. Histopathology. 2019; 74(3):372–376. https://doi.org/10.1111/his.13760 PMID: 30270453
- Magaki S, Hojat SA, Wei B, So A, Yong WH. An Introduction to the Performance of Immunohistochemistry. In: Yong WH, editor. Biobanking: Methods and Protocols [Internet]. New York, NY: Springer; 2019. p. 289–298. Available from: https://doi.org/10.1007/978-1-4939-8935-5 25
- 57. Inwald EC, Klinkhammer-Schalke M, Hofstädter F, Zeman F, Koller M, Gerstenhauer M, et al. Ki-67 is a prognostic parameter in breast cancer patients: results of a large population-based cohort of a cancer registry. Breast Cancer Res Treat. 2013 Jun 1; 139(2):539–552. <u>https://doi.org/10.1007/s10549-013-2560-8 PMID: 23674192</u>
- McCall CM, Shi C, Cornish TC, Klimstra DS, Tang LH, Basturk O, et al. Grading of Well-differentiated Pancreatic Neuroendocrine Tumors Is Improved by the Inclusion of Both Ki67 Proliferative Index and Mitotic Rate. Am J Surg Pathol. 2013 Nov; 37(11):1671. <u>https://doi.org/10.1097/PAS</u>. 00000000000089 PMID: 24121170
- Kandukuri SR, Lin F, Gui L, Gong Y, Fan F, Chen L, et al. Application of Immunohistochemistry in Undifferentiated Neoplasms: A Practical Approach. Arch Pathol Lab Med. 2017 Aug 1; 141(8):1014–1032. https://doi.org/10.5858/arpa.2016-0518-RA PMID: 28745568
- Buchwalow I, Samoilova V, Boecker W, Tiemann M. Non-specific binding of antibodies in immunohistochemistry: fallacies and facts. Sci Rep. 2011 Jul 1; 1(1):28. https://doi.org/10.1038/srep00028 PMID: 22355547
- Bubendorf L, Nocito A, Moch H, Sauter G. Tissue microarray (TMA) technology: miniaturized pathology archives for high-throughput in situ studies. J Pathol. 2001; 195(1):72–79. https://doi.org/10.1002/path. 893 PMID: 11568893
- Milanes-Yearsley M, Hammond MEH, Pajak TF, Cooper JS, Chang C, Griffin T, et al. Tissue Micro-Array: A Cost and Time-Effective Method for Correlative Studies by Regional and National Cancer Study Groups. Mod Pathol. 2002 Dec; 15(12):1366–1373. https://doi.org/10.1097/01.MP.0000036345. 18944.22 PMID: 12481019
- Camp RL, Neumeister V, Rimm DL. A Decade of Tissue Microarrays: Progress in the Discovery and Validation of Cancer Biomarkers. J Clin Oncol. 2008 Dec; 26(34):5630–5637. https://doi.org/10.1200/ JCO.2008.17.3567 PMID: 18936473
- Kallioniemi OP, Wagner U, Kononen J, Sauter G. Tissue microarray technology for high-throughput molecular profiling of cancer. Hum Mol Genet. 2001 Apr 1; 10(7):657–662. <u>https://doi.org/10.1093/hmg/</u> 10.7.657 PMID: 11257096
- Bollhagen A, Bodenmiller B. Highly Multiplexed Tissue Imaging in Precision Oncology and Translational Cancer Research. Cancer Discov. 2024 Nov 1; 14(11):2071–2088. <u>https://doi.org/10.1158/2159-8290</u>. CD-23-1165 PMID: 39485249

- Tan WCC, Nerurkar SN, Cai HY, Ng HHM, Wu D, Wee YTF, et al. Overview of multiplex immunohistochemistry/immunofluorescence techniques in the era of cancer immunotherapy. Cancer Commun. 2020; 40(4):135–153. https://doi.org/10.1002/cac2.12023 PMID: 32301585
- Baxi V, Edwards R, Montalto M, Saha S. Digital pathology and artificial intelligence in translational medicine and clinical practice. Mod Pathol. 2022 Jan; 35(1):23–32. <u>https://doi.org/10.1038/s41379-021-</u>00919-2 PMID: 34611303
- Schürch CM, Bhate SS, Barlow GL, Phillips DJ, Noti L, Zlobec I, et al. Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. Cell. 2020 Sep 3; 182 (5):1341–1359.e19. https://doi.org/10.1016/j.cell.2020.07.005 PMID: 32763154
- Stack EC, Wang C, Roman KA, Hoyt CC. Multiplexed immunohistochemistry, imaging, and quantitation: A review, with an assessment of Tyramide signal amplification, multispectral imaging and multiplex analysis. Methods. 2014 Nov 1; 70(1):46–58. https://doi.org/10.1016/j.ymeth.2014.08.016 PMID: 25242720
- Duenweg SR, Bobholz SA, Lowman AK, Stebbins MA, Winiarz A, Nath B, et al. Whole slide imaging (WSI) scanner differences influence optical and computed properties of digitized prostate cancer histology. J Pathol Inform. 2023 Jan 1; 14:100321. https://doi.org/10.1016/j.jpi.2023.100321 PMID: 37496560
- Humphries MP, Kaye D, Stankeviciute G, Halliwell J, Wright AI, Bansal D, et al. Development of a multiscanner facility for data acquisition for digital pathology artificial intelligence. J Pathol. 2024; 264(1):80– 89. https://doi.org/10.1002/path.6326 PMID: 38984400
- 72. Reinhard E, Adhikhmin M, Gooch B, Shirley P. Color transfer between images. IEEE Comput Graph Appl. 2001 Jul; 21(5):34–41.
- Macenko M, Niethammer M, Marron JS, Borland D, Woosley JT, Guan X, et al. A method for normalizing histology slides for quantitative analysis. In: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro [Internet]. 2009. p. 1107–1110. Available from: https://ieeexplore.ieee. org/abstract/document/5193250
- Vahadane A, Peng T, Sethi A, Albarqouni S, Wang L, Baust M, et al. Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images. IEEE Trans Med Imaging. 2016 Aug; 35(8):1962–1971. https://doi.org/10.1109/TMI.2016.2529665 PMID: 27164577
- 75. Rau TT, Cross W, Lastra RR, Lo RC, Matoso A, Herrington CS. Closing the loop–the role of pathologists in digital and computational pathology research. J Pathol Clin Res. 2024 Mar 10; 10(2):e12366. https://doi.org/10.1002/2056-4538.12366 PMID: 38462794
- Bodén ACS, Molin J, Garvin S, West RA, Lundström C, Treanor D. The human-in-the-loop: an evaluation of pathologists' interaction with artificial intelligence in clinical practice. Histopathology. 2021; 79 (2):210–218. https://doi.org/10.1111/his.14356 PMID: 33590577
- Schmidt U, Weigert M, Broaddus C, Myers G. Cell Detection with Star-Convex Polygons. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. Medical Image Computing and Computer Assisted Intervention–MICCAI 2018. Cham: Springer International Publishing; 2018. p. 265–273.
- 78. Grisaru-Tal S, Itan M, Klion AD, Munitz A. A new dawn for eosinophils in the tumour microenvironment. Nat Rev Cancer. 2020 Oct; 20(10):594–607. <u>https://doi.org/10.1038/s41568-020-0283-9</u> PMID: 32678342
- 79. Itzkowitz SH, Harpaz N. Diagnosis and management of dysplasia in patients with inflammatory bowel diseases. Gastroenterology. 2004 May 1; 126(6):1634–1648. https://doi.org/10.1053/j.gastro.2004.03. 025 PMID: 15168373
- Liu C, Walker NI, Leggett BA, Whitehall VL, Bettington ML, Rosty C. Sessile serrated adenomas with dysplasia: morphological patterns and correlations with MLH1 immunohistochemistry. Mod Pathol. 2017 Dec; 30(12):1728–1738. https://doi.org/10.1038/modpathol.2017.92 PMID: 28752838
- Ranganathan K, Kavitha L. Oral epithelial dysplasia: Classifications and clinical relevance in risk assessment of oral potentially malignant disorders. J Oral Maxillofac Pathol. 2019 Apr; 23(1):19. <a href="https://doi.org/10.4103/jomfp.JOMFP\_13\_19">https://doi.org/10.4103/jomfp.JOMFP\_13\_19</a> PMID: 31110412
- Paul A, Dey A, Mukherjee DP, Sivaswamy J, Tourani V. Regenerative Random Forest with Automatic Feature Selection to Detect Mitosis in Histopathological Breast Cancer Images. In: Navab N, Hornegger J, Wells WM, Frangi A, editors. Medical Image Computing and Computer-Assisted Intervention—MIC-CAI 2015. Cham: Springer International Publishing; 2015. p. 94–102.
- Irshad H. Automated mitosis detection in histopathology using morphological and multi-channel statistics features. J Pathol Inform. 2013 Jan 1; 4(1):10. <u>https://doi.org/10.4103/2153-3539.112695</u> PMID: 23858385

- **84.** Shephard AJ, Bashir RMS, Mahmood H, Jahanifar M, Minhas F, Raza SEA, et al. A fully automated and explainable algorithm for predicting malignant transformation in oral epithelial dysplasia. Npj Precis Oncol. 2024 Jun 28; 8(1):1–12.
- Tan SE, Tan W, Fisher KH, Strutt D. QuantifyPolarity, a new tool-kit for measuring planar polarized protein distributions and cell properties in developing tissues. Development. 2021 Sep 7; 148(18): dev198952. https://doi.org/10.1242/dev.198952 PMID: 34351416
- Chang J, Chaudhuri O. Beyond proteases: Basement membrane mechanics and cancer invasion. J Cell Biol. 2019 Jul 17; 218(8):2456–2469. https://doi.org/10.1083/jcb.201903066 PMID: 31315943
- Ueno H, Kajiwara Y, Ajioka Y, Sugai T, Sekine S, Ishiguro M, et al. Histopathological atlas of desmoplastic reaction characterization in colorectal cancer. Jpn J Clin Oncol. 2021 Jun 1; 51(6):1004–1012. https://doi.org/10.1093/jjco/hyab040 PMID: 33855369
- Konishi Y, Taki T, Nakai T, Kuroe T, Morisue R, Miyoshi T, et al. Clinicopathological features and prognostic impact of dirty necrosis in metastatic lung cancers from the colon and rectum. Cancer Sci. 2023; 114(5):2169–2177. https://doi.org/10.1111/cas.15647 PMID: 36369892
- Awan R, Sirinukunwattana K, Epstein D, Jefferyes S, Qidwai U, Aftab Z, et al. Glandular Morphometrics for Objective Grading of Colorectal Adenocarcinoma Histology Images. Sci Rep. 2017 Dec 4; 7 (1):16852. https://doi.org/10.1038/s41598-017-16516-w PMID: 29203775
- 90. Li J, Yang S, Huang X, Da Q, Yang X, Hu Z, et al. Signet Ring Cell Detection with a Semi-supervised Learning Framework. In: Chung ACS, Gee JC, Yushkevich PA, Bao S, editors. Information Processing in Medical Imaging. Cham: Springer International Publishing; 2019. p. 842–854.
- Pan X, AbdulJabbar K, Coelho-Lima J, Grapa AI, Zhang H, Cheung AHK, et al. The artificial intelligence-based model ANORAK improves histopathological grading of lung adenocarcinoma. Nat Cancer. 2024 Feb; 5(2):347–363. https://doi.org/10.1038/s43018-023-00694-w PMID: 38200244
- 92. Das DK, Chakraborty C, Sawaimoon S, Maiti AK, Chatterjee S. Automated identification of keratinization and keratin pearl area from in situ oral histological images. Tissue Cell. 2015 Aug 1; 47(4):349–358. https://doi.org/10.1016/j.tice.2015.04.009 PMID: 26150310
- Mathialagan P, Chidambaranathan M. Computer vision techniques for Upper Aero-Digestive Tract tumor grading classification–Addressing pathological challenges. Pattern Recogn Lett. 2021 Apr 1; 144:42–53.