

Towards Automatic Protein Co-Expression Quantification in Immunohistochemical TMA Slides

Leslie Solorzano , *Student Member, IEEE*, Carla Pereira , Diana Martins, Raquel Almeida , Fátima Carneiro, Gabriela M. Almeida , Carla Oliveira , and Carolina Wählby 

Abstract—Immunohistochemical (IHC) analysis of tissue biopsies is currently used for clinical screening of solid cancers to assess protein expression. The large amount of image data produced from these tissue samples requires specialized computational pathology methods to perform integrative analysis. Even though proteins are traditionally studied independently, the study of protein co-expression may offer new insights towards patients’ clinical and therapeutic decisions. To explore protein co-expression, we constructed a modular image analysis pipeline to spatially align tissue microarray (TMA) image slides, evaluate alignment quality, define tumor regions, and ultimately quantify protein expression, before and after tumor segmentation. The pipeline was built with open-source tools that can manage gigapixel slides. To evaluate the consensus between pathologist and computer, we characterized a cohort of

142 gastric cancer (GC) cases regarding the extent of E-cadherin and CD44v6 expression. We performed IHC analysis in consecutive TMA slides and compared the automated quantification with the pathologists’ manual assessment. Our results show that automated quantification within tumor regions improves agreement with the pathologists’ classification. A co-expression map was created to identify the cores co-expressing both proteins. The proposed pipeline provides not only computational tools forwarding current pathology practices to explore co-expression, but also a framework for merging data and transferring information in learning-based approaches to pathology.

Index Terms—Co-expression, computational pathology, gastric cancer, image analysis, immunohistochemistry, protein, registration.

Manuscript received November 6, 2019; revised March 13, 2020 and June 10, 2020; accepted July 8, 2020. Date of publication July 13, 2020; date of current version February 4, 2021. The work of Leslie Solorzano and Carolina Wählby was supported by ERC CoG 682810. The work of Carla Pereira was supported by FCT:SRFH/BD/113031/2015. The work of Carla Oliveira was supported in part by FEDER/COMPETE 2020/PORTUGAL2020/ POCI/FCT: Ref. POCI-01-0145-FEDER-007274, 030164, and 016390; and in part by Project NORTE-01-0145-FEDER-000029. (*Corresponding author: Leslie Solorzano.*)

Leslie Solorzano and Carolina Wählby are with the Department of Information Technology and SciLifeLab, Uppsala University, 75236 Uppsala, Sweden (e-mail: leslie.solorzano@it.uu.se; carolina.wahlby@it.uu.se).

Carla Pereira is with i3S - Institute Investigação e Inovação em Saúde and the Institute of Molecular Pathology and Immunology (IPATIMUP), University of Porto, 4099-002 Porto, Portugal (e-mail: carlap@ipatimup.pt).

Diana Martins is with i3S - Institute Investigação e Inovação em Saúde, University of Porto, 4099-002 Porto, Portugal, and also with the Department of Biomedical Laboratory Sciences, Polytechnic Institute of Coimbra, 3045-093 Coimbra, Portugal (e-mail: dianam@ipatimup.pt).

Raquel Almeida is with i3S - Institute Investigação e Inovação em Saúde, the Institute of Molecular Pathology and Immunology (IPATIMUP), the Faculty of Sciences, and the Faculty of Medicine, University of Porto, 4099-002 Porto, Portugal (e-mail: ralmeida@ipatimup.pt).

Fátima Carneiro is with i3S - Institute Investigação e Inovação em Saúde, the Institute of Molecular Pathology and Immunology (IPATIMUP), the Faculty of Sciences, and the Department of Pathology, Faculty of Medicine, University of Porto, 4099-002 Porto, Portugal (e-mail: fcarneiro@ipatimup.pt).

Gabriela M. Almeida and Carla Oliveira are with i3S - Institute Investigação e Inovação em Saúde, the Institute of Molecular Pathology and Immunology (IPATIMUP), and the Faculty of Medicine, University of Porto, 4099-002 Porto, Portugal (e-mail: galmeida@ipatimup.pt; carlaol@ipatimup.pt).

Digital Object Identifier 10.1109/JBHI.2020.3008821

I. INTRODUCTION

GASTRIC cancer (GC) ranks as the fifth most common cancer worldwide and the third most frequent cause of cancer-related mortality [1]. GC patients rarely experience symptoms at early stages of the disease, and more than 80% are diagnosed at an unresectable stage, where the 5-year survival rate is only 20% [2]. A better understanding of tumor heterogeneity and local molecular signatures is crucial to guide better clinical and therapeutic decisions, a fact that is true for many types of cancer [3] where the knowledge of underlying mechanisms is limited.

Immunohistochemical (IHC) analysis of tissue samples is the mainstream approach for diagnosis and therapeutic decision in solid cancers. Pathologists are frequently tasked with the cross-slide analysis of consecutive tissue sections stained to highlight particular features of the tissue or to study a new protein of interest. Despite robust guidelines, IHC is often limited by the subjectivity associated with qualitative visual interpretation of expression levels. In this matter, computational pathology can be an ally, as it objectively assesses, quantifies and relates a large number of features in a systematic and high throughput way.

Advanced commercial and open-source software is starting to gain traction in the world of computational pathology and making its way into the clinical practice. One example is QuPath [4], designed for whole slide image analysis. With the help of Bio-Formats [5], QuPath is capable of opening and reading the most common slide formats and allows for flexible and efficient tissue

analysis, including, e.g., annotation and automated cell detection and counting by pathologists. QuPath can also be programmed via custom macros which can be useful to extract data for further processing. However, QuPath does not include registration tools, making it impossible to quantify co-expression in data coming from different slides and spaces.

Python's scientific computing ecosystem composed of core numeric libraries such as Numpy and Scipy, and domain specific libraries such as scikit-image and scikit-learn provide interfaces to perform image processing and make available several machine learning models that can be used for pixel classification i.e. segmentation.

In order to quantify protein co-expression, we developed and present here an image analysis pipeline to perform image registration on tissue microarray (TMA) slides, define tumor regions and quantify protein co-expression, built on such libraries. As proof of concept, we explored two proteins frequently altered in GC, namely E-cadherin [6] and CD44v6 [7]. Each protein was characterized by IHC, in separate but consecutive TMA sections, distanced a few micrometers apart. Therefore, tissue morphology is expected to be very similar between both tissue sections.

Recent developments in multispectral imaging and multiplex immunohistochemical staining makes it possible to study co-occurrence of several markers in the same tissue core [8]. These methodologies present multiple pre-analytical and analytical challenges such as the lack of standardization of fixation times, dehydration, paraffin impregnation and the cross-reactivity between multiple antibodies. Additionally, the availability of these systems is still very limited in terms of scalability and high throughput. The presented study is focused on well-established protocols commonly used in research and clinical practice.

IHC usually combines two stainings: haematoxylin (H) and diaminobenzidine (DAB). H stains the general scaffold of the tissue in blue/purple, mainly highlighting nuclei morphology. DAB produces a brown precipitate, revealing the expression status of a protein targeted by a specific antibody (i.e. subcellular location, intensity and expression pattern). The morphological features revealed by H can be used as input for automated spatial alignment of multiple parallel tissue sections, including alignment quality assessment, as previously shown in [9]. This motivates the alignment of the cores. To do so, we used a recent registration framework [10] that takes advantage of the spatial information carried by H and also the intensity information of the unmixed stains.

However, quantifying co-expression patterns from TMA slides presents additional challenges in the management of gigapixel images. TMA images are stored in resolution pyramidal formats that require specialized programming libraries. We therefore developed custom tools for slide visualization, registration, and analysis within the scope of the open source TisSUUmaps project [11]. TisSUUmaps is built on top of OpenSeadragon [12], meaning that an arbitrary number of slides can be visualized and annotated in parallel, on a browser, which enables easy collaboration and sharing of information. The pipeline is modular and built with open-source tools, and modules can be changed to suit the tools used in any already existing pipelines.

Overall, using our image analysis pipeline, we were able to quantify protein co-expression in TMAs that can hopefully unveil novel insights in the clinical and molecular impact of multiple protein markers.

II. PREVIOUS WORK

The workflow presented in this paper involves several image analysis methods developed over the years. These are introduced in the following section while the methods section provides the details of the final workflow.

A. Color Unmixing

Bright-field tissue slide scanning outputs an RGB image, where each channel represents the absorption of each of the three colors (red, green and blue). In consequence, the resulting TMA slide has to be processed to find, separate and quantify the protein of interest, i.e., provide a way to find how much contribution in an RGB channel is given to the colors of the desired stains. The relationship between the concentration of a stain and its absorbance or optical density depends on the chemical and physical properties of the stains. There is an ongoing debate on the interpretation of a pixel intensity and saturation with respect to the amount of stain [13]. This means that while intensities provide information about the presence of stain, they might not be linearly related with the amount of stain it represents. This is important to remember when considering a pixel as positive or negative for a certain stain. There are specially hard cases when the DAB stain is very faint and there can be a high discrepancy between decisions.

Color unmixing (also referred to as color deconvolution) [14], is the process of finding a matrix of optical densities or absorbances from samples of the image. Several ways exist to do color unmixing [15], [16]. The goal is always to find a way to transform from the RGB color representation to a space spanned by the stains in the tissue.

B. Image Registration

Image registration is the process of spatially aligning two different images. We need to align the core images so that their corresponding features are aligned in space, enabling analysis of protein co-expression. After alignment, structures composed by cells in the tissue will overlap and the images will be considered spatially aligned. The first image is referred to as the reference, it is in a reference space. The second image is referred to as the floating image. Once registered, all the main features, the tissue scaffold, will be aligned.

There are different types of transformations, features, measures and evaluation methods for different registration methods and they are always application dependent; there is no one-size-fits-all. There are linear and non-linear transformations. Linear transformations have a smaller number of parameters and a smaller search space with fewer local optima and are usually easier to interpret physically [17]. Non-linear transformations, also called elastic or deformable transformations, have a very

high number of parameters that need regularization terms to achieve any kind of physical or anatomical interpretability.

Different types of registration have been used in high resolution tissue images, and registration examples with feature extraction include [18]–[20], while [21]–[23] exemplify non-linear approaches. A detailed study of registration of consecutive tissue sections can be found in [24] where linear transformations are performed on images of the same slide (washed and re-stained). The same study shows that much of the alignment in the pathology field is done manually which is arduous and inaccurate.

There are several automated registration methods based on Mutual Information, both as a measure to guide the registration and as an evaluation metric [25], [26]. However, these methods do not make full use of the spatial information. A recent registration framework uses a measure that combines intensity and spatial information in the same regularization term [10], typically exhibiting a larger region of convergence, increasing the chance of finding the correct transformation parameters under larger rotations, translations and scaling.

More formally, the spatial alignment of two images is modelled as a minimization process where the distance between the overlapped images is the main criterion. This distance d is measured iteratively as several transformations from a valid set Ω are applied to an image A to match its counterpart B . It is formulated as:

$$\hat{T} = \arg \min_{T \in \Omega} d(T(A), B) \quad (1)$$

For affine transformations, Ω includes rotations, translation, scaling, shearing, and reflections. The distance d combines both intensity and spatial information, with the benefit of few local optima as compared to intensity based approaches, and no need for feature matching as compared to feature based methods.

C. Tumor Segmentation

TMA slides represent pieces of a transversal section in a tumor sample. Tumor samples are three dimensional structures composed of tumor cells but also stromal non-tumor components. Therefore, regions highly enriched with tumor cells within TMA cores must be found. As manual labeling is labor intensive, automatic segmentation with the aid of fewer to no labels is preferred.

Different tumors exhibit different morphological features depending on the histological type. With daily practice, pathologists learn to identify tumor-specific features and interpret them under a specific clinical context. In image analysis, tumor segmentation becomes a problem of pixel classification where each pixel is assigned to a class e.g. tumor, non-tumor, and background.

To distinguish tumor *vs.* non-tumor cells, the pathologist often evaluates the H staining that highlights cell nuclei features. More often H&E staining is used to locate tumor. Visually speaking, cancer cell nuclei are recognized by a decreased staining intensity, while normal cells display an homogeneous purple color, which translates to textures in image analysis. A pathologist can then manually label a few pixels representing different cells types and this information can be used in any of a great number of pixel classifiers. Not all the DAB staining

is biologically meaningful, even when disregarding technical artifacts. In specific, normal epithelial cells of the stomach lack CD44v6 expression, but it is expressed in pre-malignant and malignant lesions of gastric mucosa. On the other hand, E-cadherin is expressed at the membrane of normal epithelial cells. The loss of its function, either complete or aberrant protein expression, reflects a pathological condition in the gastric mucosa. Therefore it is crucial to perform tumor segmentation to ensure a reliable automated estimation of protein expression in tumor situ.

There are both classical and modern learning-based methods that allow for pixel classification, from graphical models, thresholding, K-means, naive Bayes, support vector machines, decision trees, random forest, to deep learning, as reviewed in [27]. One example of pixel classification for TMAs can be found in [28], where tumors in lung tissue are detected using Markov Random fields.

Random forests (RF) are a very powerful ensemble method that creates several decision trees (hence the forests) and these trees make a decision together. All trees try to find the best features that represent the data, which makes them an ideal choice for pixel classification. RF are the default choice in the interactive Ilastik software [29].

D. Protein Co-Expression

The spatial overlap of two pixels (protein 1 and protein 2) allows to infer the co-expression of both proteins. Consecutive TMA slides in close proximity are expected to represent the same structures. This proximity allows the quantification of colocalization to be interpreted as co-expression within a level of certainty. Co-expression can then be assessed by finding the distances between the nearest pixels that best represent the protein. But even if proteins are spatially distributed over a similar region, some co-expression could be the result of random overlap, specially from pixels with lower intensities.

The work in [30] has been crucial for establishing the mathematical framework for colocalization quantification borrowing many elements from statistics. Originally used to quantify colocalization in fluorescence microscopy [31], colocalization is quantified using the Pearson Correlation Coefficient.

III. METHODS AND DATA

This section presents the proposed image analysis pipeline and an overview on the data analyzed. The presented pipeline is based on open source-source code, it is modular, and each module can be implemented according to the tools available to any laboratory searching to incorporate computational pathology pipelines. Detailed video tutorials and an implementation of the pipeline are available at github.com/wahlby-lab/TMA-studies.

A. Protein IHC Image Data From a GC Patient Cohort

Tumor samples from a cohort of gastric adenocarcinoma patients surgically treated between January 2008 and December 2014 at Centro Hospitalar Universitário de São João (CHUSJ, Porto, Portugal). A TMA was assembled from paraffin-embedded tumor material of these patients. Representative areas

TABLE I
PATHOLOGIST CLASSIFICATION SCHEME FOR CD44v6

Score	Estimated tumor area coverage (%)	Expression pattern
0	0%	Negative or cytoplasmic
1+	< 10%	Membrane
2+	10 – 50%	Membrane
3+	> 50%	Membrane

TABLE II
PATHOLOGIST CLASSIFICATION SCHEME FOR E-CADHERIN

Score	Estimated tumor area coverage (%)	Expression pattern
0	0%	Negative
1+	< 25%	Cytoplasmic or aberrant membrane
2+	25 – 50%	Cytoplasmic or aberrant membrane
3+	> 50%	Cytoplasmic or aberrant membrane

of the tumors were selected based on H&E stained tissue specimens and one tissue core with 2 mm diameter was obtained from each selected specimen.

Nine TMA blocks were constructed and in each block a normal gastric mucosa core and a GC cell line core were included as controls, as well as a core of a non-gastric tissue sample for orientation purposes. IHC staining of E-cadherin and CD44v6 was performed in 3-micrometer TMA sections, in close proximity. The assay was carried out on the automated Ventana Benchmark ULTRASTaining System, using the OptiView DAB IHC Detection Kit (Roche/Ventana Medical Systems, Tucson, AZ, USA). H staining is used to reveal all tissue. Positive and negative staining controls were performed in parallel with paraffin sections. To monitor E-cadherin normal expression, either normal gastric mucosa, intestinal metaplasia or cell lines included at the TMA, while normal skin was used to control CD44v6 normal expression. Each TMA slide was scanned using NanoZoomer 2.0HT (Hamamatsu) whole slide imaging scanner. TMA image sizes range between 150,000 pixels in width and 100,000 pixels in height, resulting in sizes of 5000 pixels² for each core with a average resolution of 0.226 micrometers/pixel. Overall, our image data set sums up cores from 18 TMA slides (9 for each protein) and the respective pathologist grade per TMA core, per protein. After excluding patients with no clinical-pathological or treatment data available, patients lost to follow-up and without available paraffin-embedded material for analysis, in total, 261 cases were eligible for analysis.

One expert pathologist graded all cores for one of the proteins, while another expert pathologist independently graded all cores stained for the other protein. In both cases one of four classes (from 0 to 3+) were assigned based on the estimated tumor area and the expression pattern. Due to the distinct biological context of each protein, each class represents different criteria. Specifically, CD44v6 is not expressed in normal tissue but becomes expressed in pre-malignant and malignant cells, and exerts its function at the membrane (Table 1). On the other hand, E-cadherin is expressed at the membrane of normal cells but, in a tumor context, is often abnormally expressed (cytoplasmic or aberrant membrane, i.e. dotted or incomplete membrane) or absent (Table 2). In this dataset, a fraction of cases (118/261

(45%)) preserved membranous E-cadherin expression. As we are interested in exploring E-cadherin and CD44v6 dysfunction, cases preserving complete membranous E-cadherin expression were excluded and the remaining counterpart (142/261 (55%)) proceeded for analysis.

B. Methods in Image Analysis

The pipeline modules of our work are shown in Fig. 1, serving as a visual guideline to the outcomes of every module. Our methods begin by separating the core images into its main components and all the necessary information will then be extracted from these separated stains.

The documentation for the API of the scientific python libraries used is available in their respective pages online. We present our methodology and the parameters used.

1) *Color Unmixing*: Fig. 1 shows a TMA slide containing multiple IHC-stained cores where the blue/violet color represents the H staining and the brown color represents the DAB staining of E-cadherin and CD44v6. Color unmixing was done separately for E-cadherin and CD44v6.

First, representative patches were manually selected and RGB values were grouped into 255 clusters by k-means clustering. These 255 colors (representative of all RGB values expected from a given IHC staining) were thereafter sorted by their value (in HSV space), and colors of very high value were removed due to their closeness to the white and very light background colors. Starting from the color samples found in this reduced RGB space, an absorbance model was extracted as explained in [14]. Solving a least squares problem between the strongest absorbance and the remaining ones, creates the orthogonal basis M whose pseudo inverse is used to create grayscale images representing H and DAB. The result is two grayscale images, H_1 and DAB_1 for protein 1 (E-cadherin) and correspondingly two images H_2 and DAB_2 for protein 2 (CD44v6), as shown in Fig. 1, where intensity corresponds to presence of the respective stains. M was created once for each of the two stains, and thereafter used in the full experiment.

2) *Image Registration*: The H images from the cores carry information about the common structure of the core; a common scaffold, and are therefore used for image registration (while the DAB images may differ as they represent different proteins that may have different spatial distribution within the tissue). Using the recent registration framework proposed by [10], the main transformation between cores and its parameters is obtained by using H_1 and H_2 in equation 1:

$$\hat{T} = \arg \min_{T \in \Omega} d(T(H_1), H_2) \quad (2)$$

Where the distance d is Alpha-AMD, a distance that incorporates both intensities and spatial information to guide the registration.

The transformation T is chosen to be affine, meaning that Ω only includes scaling, rotation and translation. Since the aligned scaffolds represent very similar structures but not exactly the same, any elastic deformation could potentially create undesirable false structures. The transformation T is stored and applied

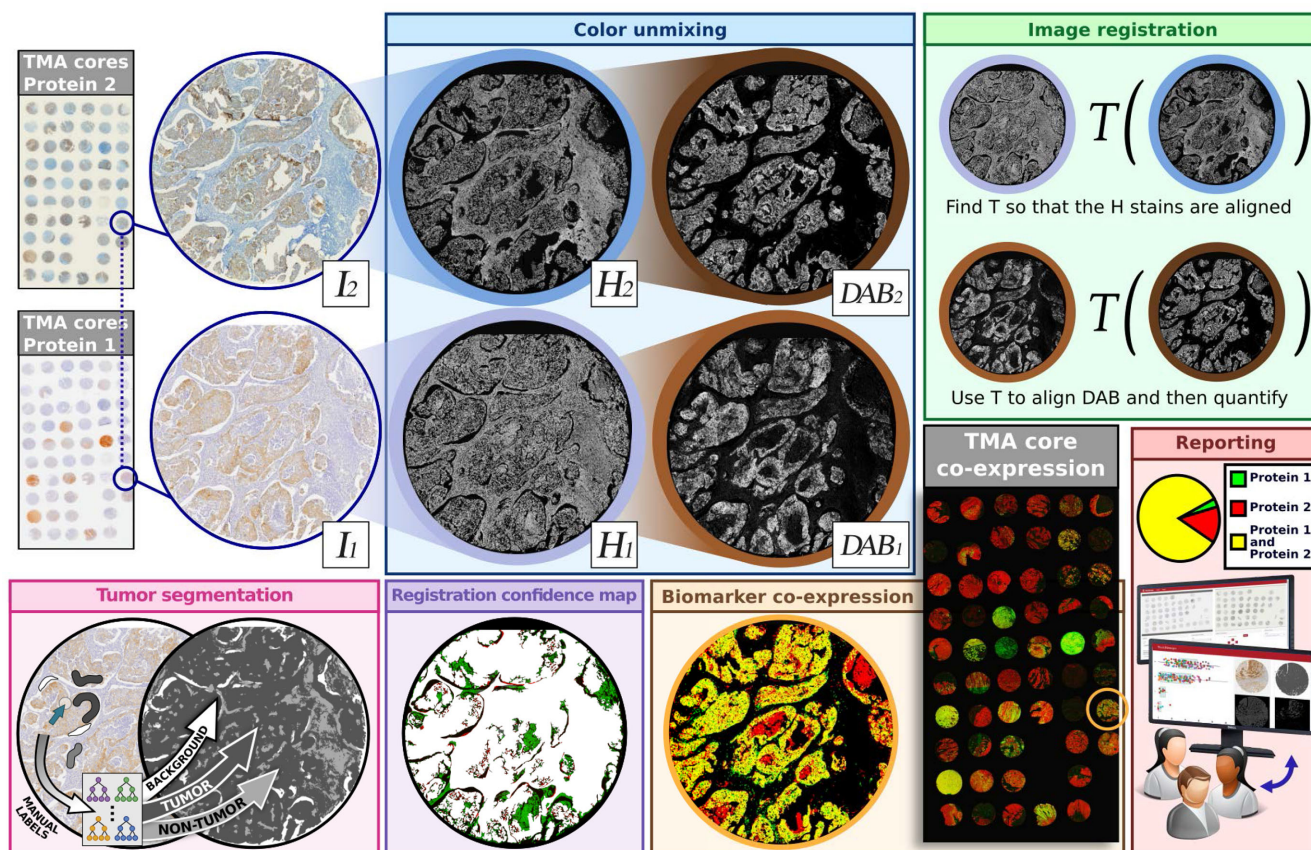


Fig. 1. Overview of the stages to quantify co-expression in TMA cores. Two individual proteins stainings of the core images (I_1, I_2) are unmixed to reveal tissue scaffolding (H_1, H_2) and protein expression (DAB_1, DAB_2). Each pair of TMA cores expected to come from the a very near spatial location are registered using the H image to find the transformation T , which is applied to the DAB image. Using image I and manual labels, tumor segmentation is found in each core. A registration confidence map (RCM) ensures H stained tissue is present and correlates across the two cores (white = high correlation, green/red = tissue missing in H_1 or H_2). Only white regions are kept for analysis of co-expression. DAB_1 and transformed DAB_2 stains are superimposed and co-expression quantified using the tumor segmentation and registration confidence maps as masks.

to DAB_2 in order to bring it to the same space as DAB_1 where they can be effectively compared.

The Alpha-AMD method requires a few parameters to work. For the level of pyramids we use (32, 16, 8, 4) which means that the first computations for the registration will be done on an image 32 times smaller all the way until 4. We used gaussian blur with sigmas of (15.0, 8.0, 4.0, 2.0) which means that each of the previously mention levels will be blurred applying a gaussian filter with the selected sigmas. The method allows for the input of masks from where to sample so we input binary masks that cover the core. We use 300 iterations for the registration.

We used CD44v6 as the reference image meaning that all the cores from E-cadherin were transformed to fit in the same space of the CD44v6 cores.

In general, we observed that the H images contain enough information to guide the transformation, particularly because the registration is robust enough to use the information coming from nucleus and cytoplasm in H despite them not being exactly equal in both images.

The transformation T is stored and kept to be applied to any information that has to be sent to the CD44v6 space.

3) Registration Confidence Map: In order to trust quantification of co-expression from registered images, we must first be sure that similar tissue actually exists in all parts of the sample. Images I_1 and I_2 come from different slides which can present artifacts in different locations such as rips, folds, dust, noise etc. When bringing I_2 to the space of I_1 all the artifacts have to be masked out. This means that information present in I_1 might be missing from I_2 and vice versa. If information is missing in one of the images, no attempt of quantifying co-expression should be done in the area.

For this purpose, a Registration Confidence Map (RCM) is created by finding the colocalization between $\max(H_1, DAB_1)$ and $\max(H_2, DAB_2)$ with artifacts masked out [9]. This maximum intensity projection (MIP) of both H and DAB from the same core indicates the presence of tissue. The RCM will then indicate the pixels that are trusted to provide information on expression and co-expression of proteins. In Fig. 1 H and DAB are shown and the colocalization of the MIPs becomes the RCM, shown in the box with the same name. White regions represent regions with high confidence.

4) Tumor Segmentation: An RF classifier was used to segment each core into tumor, non-tumor and background. Due to

the lack of H&E staining we performed tumor segmentation on the CD44v6 image, and as the E-cadherin images are registered with CD44v6, we later applied the same tumor mask when quantifying both proteins. Sparse labeling was performed per category (background, tumor, non-tumor) manually drawn by the pathologist in 25 cores to train the RF classifier. This results in hundreds of thousands of features per class. Fig. 1 shows an illustration of this process.

Using classical image processing, features are extracted from the pixel information under the labels. The features extracted are intensities, edges and texture information. In our case, the most meaningful features were Gaussian smoothing GS(10.0), Laplace of Gaussian LoG(3.5), Gaussian gradient magnitude GGM(10.0), Hessian Matrix Eigenvalues HME(1.6, 3.5, 10.0). GS features remove high frequencies and blur the image. GS is the input for LoG and GGM which find edges. HME serves as corner detector. The features are selected by their gini importance, or mean decrease in impurity (MDI) meaning how much they contribute to the purity of the tree. This is computed automatically along with the RF and the chosen features are those that have a MDI over the mean.

To train the RF we used the Scikit-Learn implementation version 0.22.1 from the “ensemble” module. From the RandomForestClassifier class we set the `n_estimators` to 100 which is the number of trees in the forest. We also set parameter `warm_start` which allows for a continued training if needed, even after the script has finished.

Using the module “io” the images I and their corresponding labels are loaded. For each channel and each pixel from I that lies under a label, the modules “filters” and “features” compute GS, LoG, GGM and HME which are then organized in an array of size $(n_samples, n_features)$ along with an array of size $(n_samples)$, which is used as input for the fit method from the classifier.

To perform tumor segmentation, features are extracted from unlabeled pixels and classes (tumor, non-tumor, background) are predicted.

5) Protein Expression Quantification: The developed image analysis methods to quantify protein expression are inspired by the pathologist’s method for visual assessment.

From a computational perspective, once stains are separated and tumor segmented, the images DAB_1 and DAB_2 contain the information of the protein expression. In a first approach, the expression of H and DAB images was quantified by direct thresholding of the images. However, in [13] it is suggested that caution should be taken when quantifying protein expression directly from the unmixed images; they must undergo additional processing before quantification correlating with visual assessment can be done.

DAB stains appears in membranes as well as some other tissue structures, and while the strength of the staining in the membranes is variable, yet lower intensities count as much as strong ones at visual scoring, if the stain is located in a membrane. Inspired by the visual assessment method, we apply the Robert’s cross edge filters on the DAB images and combine this information with the original DAB image. We let DAB be the input image and E be the Robert’s cross edge information

and construct the membrane image B :

$$B = \begin{cases} 2 \times DAB \times E & \text{if } E \leq 0.5 \\ 1 - 2 \times (1 - DAB) \times (1 - E) & \text{if } E > 0.5 \end{cases}$$

This means that weak edges, i.e. $E \leq 0.5$, take the information from reduced DAB intensities so that the B image will be darker (as $2 \times E$ is always less than 1.0). For pixels with strong edges, i.e. $E > 0.5$, the intensity is enhanced so that the B image will be close to 1 if E or DAB have strong values.

Next, we removed background noise by thresholding B at a fixed value (0.03) applied across both stains in all images. A fixed value, rather than automated thresholding, was applied to avoid enhancement of background in absence of signal. The resulting binary image represents the final presence/absence of the protein. The final percentages of tumor area coverage were calculated by finding the fraction of positive pixels in the part of each core fulfilling the RCM criteria and either including or excluding the tissue regions outside the tumor segmentation mask.

The protein co-expression was quantified for each pair of aligned cores fulfilling the RCM criteria and excluding the tissue regions outside the tumor segmentation mask. All overlapping pixels in the aligned DAB_1 and $T(DAB_2)$ images were aggregated and their total number over the number of pixels in the tumor returns the percentage of co-expression. An example a co-expression map can be observed in Fig. 3.

IV. RESULTS

In this study, we developed an image analysis pipeline to align and quantify co-expression of two proteins in GC TMA cores. Prior to this work, experienced pathologists visually inspected the expression profile (both extension and pattern) of E-cadherin and CD44v6 and grouped cores in four categories (from 0 to 3+), called scores.

To evaluate the reliability of our co-expression quantification, we first evaluated CD44v6 and E-cadherin expression individually. Our image analysis pipeline is inspired by the pathologist’s classification methodology, where focus is on expression within tumor areas.

The box plots in Fig. 2 compare the manual and automated classification of two proteins in 142 cores, prior to and after tumor segmentation. Within each class, we can observe that the automated measurements are widely distributed for both proteins. To relate manual and automated classifications, we determine the quartiles from the data and calculated thresholds based on the weighted medians between the groups. These thresholds allow the categorization of the predicted percentages and compare classes with the pathologists’ classification as shown in the confusion matrices. To evaluate the quantification of CD44v6 and E-cadherin expression, we present confusion matrices to compare the number of cases assigned to each class according to manual *versus* automated classification. Classes are imbalanced so we colored each cell to reflect the percentage of cores within the predicted classes with respect to the pathologist class.

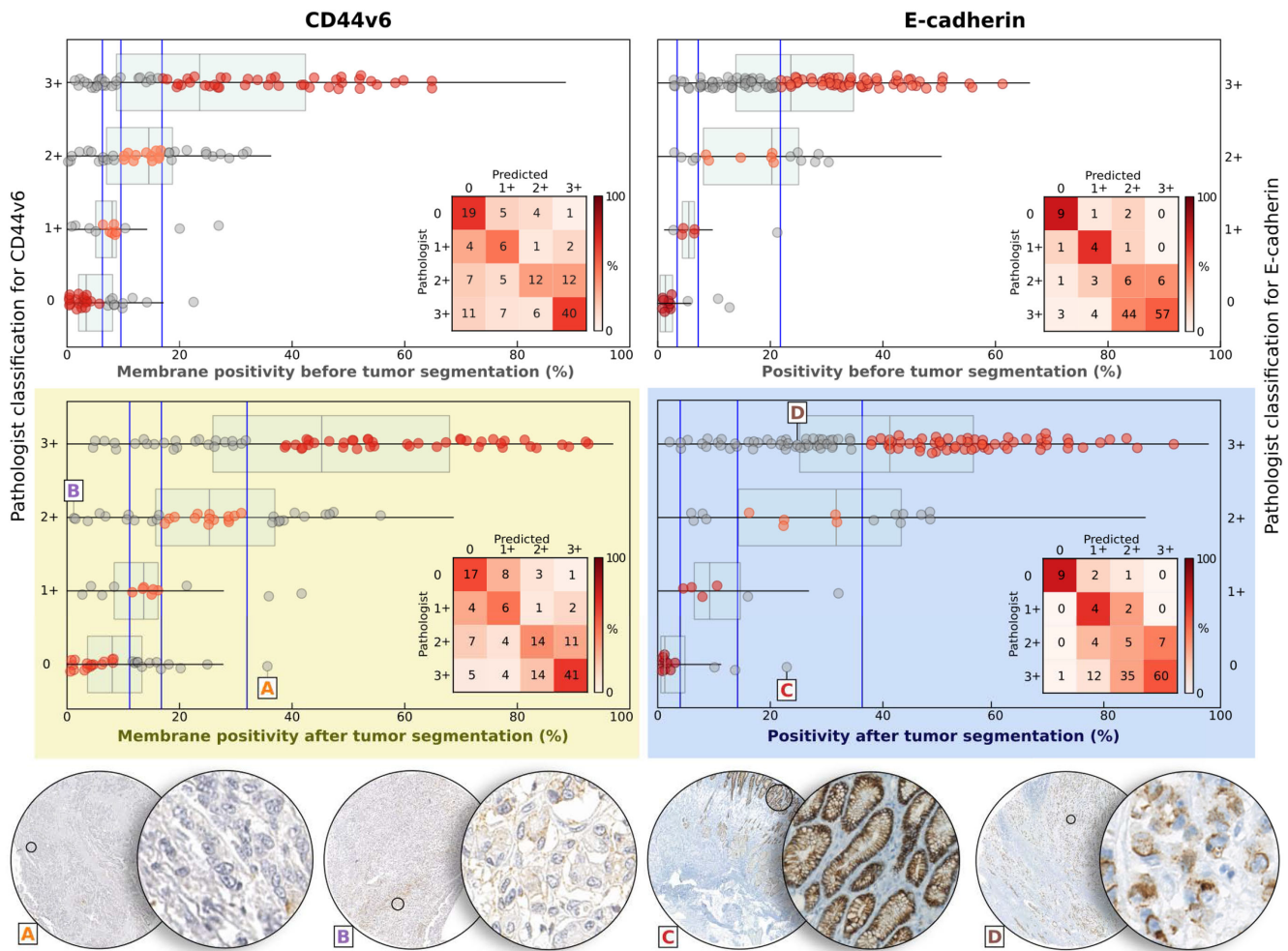


Fig. 2. Comparison of manual and automated classification of E-cadherin and CD44v6 per core. Confusion matrices show manual classification score vs. automated classification score based on the weighted median of the continuous automated classification results. Colors in the matrices reflect the percentage of cases in the predicted classes with respect to the pathologist class. In the bottom area, the yellow background highlights the classification of CD44v6 within tumor and the blue background highlights E-cadherin classification within tumor. Four cases (A, B, C and D) illustrate some examples of cores that have been misclassified due to complex expression patterns, less accurate tumor segmentation, or poor color unmixing.

Before tumor segmentation, the predicted amount of CD44v6 and E-cadherin expression was underestimated for all the scores which makes the discrimination of classes difficult.

After tumor segmentation, the discrimination between classes improved for both proteins. We observed that our method performed better quantifying *de novo* CD44v6 expression, particularly when CD44v6 membranous expression was present in the majority (>50%) of tumor cells (class 3+, 41 out of 64 ~ 64%). In contrast, predicting E-cadherin aberrant expression was not a trivial task for the computer, nor was it for the pathologists, due to the difficulty of estimating multiple of patterns. Nevertheless, our method performed well in classifying E-cadherin complete loss of expression (class 0, 9 out of 12 ~ 75%), which are often regarded the most aggressive GC tumors.

The comparison between the pathologists' classification and the computer prediction gives a Cohen's Kappa of 0.362 for CD44v6 and 0.241 for E-cadherin within tumor areas. The

values are reduced when tumor segmentation is not used, resulting in 0.355 for CD44v6 and 0.236 for E-cadherin. While the interpretation of Cohen's Kappa is somewhat ambiguous, numbers higher than 0.2 indicate some measure of agreement. We can argue that unbalanced distribution across the classes usually results in lower Kappa statistics.

Considering that ambiguous cases can be challenging for either pathologist or computer, we developed a core explorer visualization tool for IHC data. We performed a depth analysis of four cases (A and B for CD44v6 and C and D for E-cadherin) and identified a few sources of discrepancies between pathologists and computers' assessment.

For case A, given the absence of CD44v6 staining, the pathologist scored 0. However, the computer scored it as a 3+ because, during color unmixing step, very light browns had to be included (which can be confused with light purples/blues) resulting in the overestimation of the amount of DAB staining. In case B, our method scored a lower percentage of CD44v6 compared to the

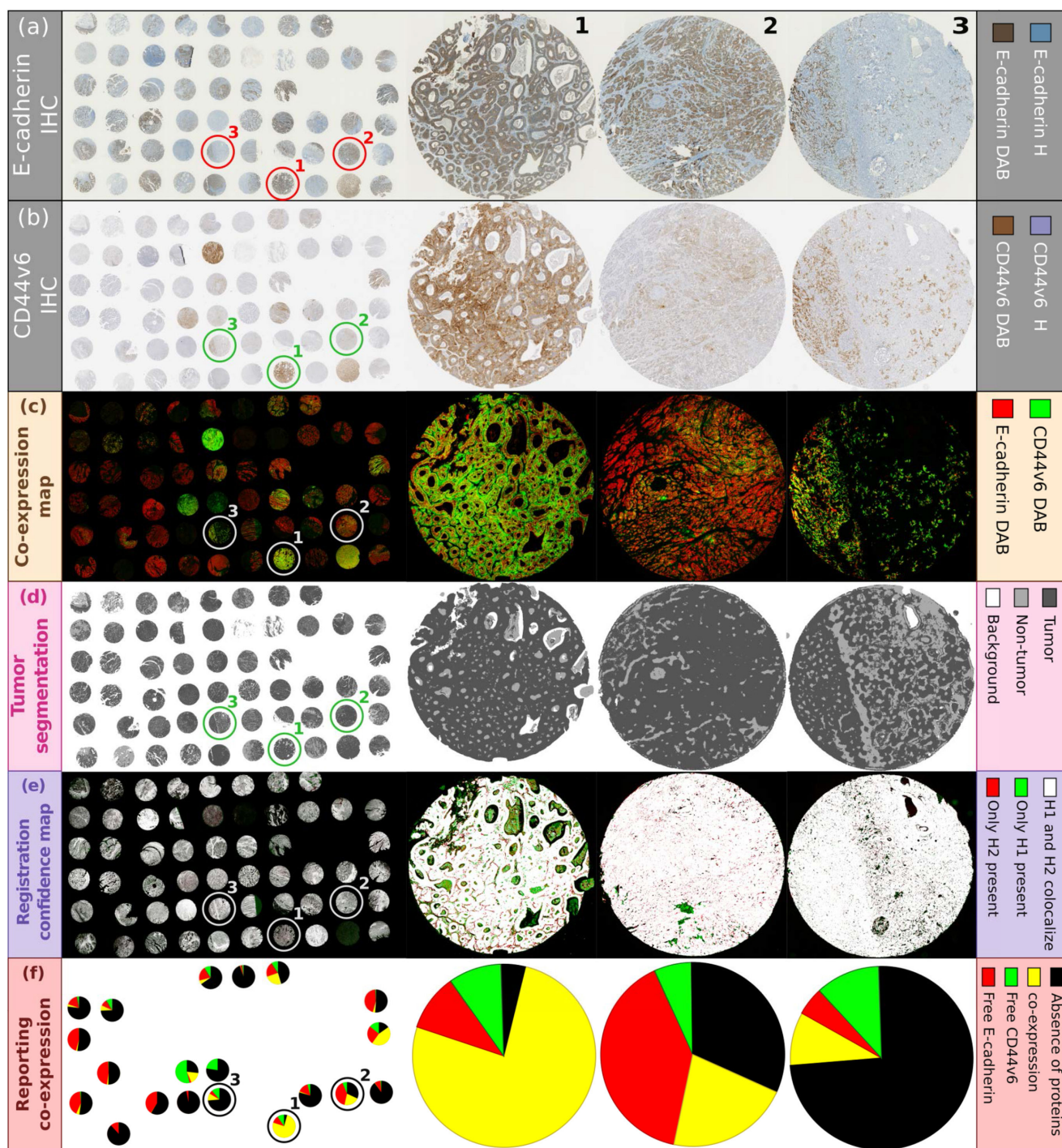


Fig. 3. Overview of the output from each of the steps to quantify and visualize E-cadherin, CD44v6, and their co-expression. (a, b) Original core array for E-cadherin and CD44v6. (c) Co-expression map showing E-cadherin in red and CD44v6 in green. (d) Tumor segmentation, darkest gray shows tumor. (e) The registration confidence map (RCM) highlights tissue colocalization zones in white. (f) Overview of co-expression (yellow), free proteins (free E-cadherin in red, free CD44v6 in green), and absence of protein (black) within tumor and RCM. The cores preserving E-cadherin membranous expression or without available paraffin-embedded material were excluded from the co-expression quantification analysis.

pathologist classification of 2+. Even though CD44v6 is present at the cell membrane, its expression is focal and faint which results in an underestimation of DAB staining.

Case C was classified by the pathologist as 0, but our method reports a significant amount of E-cadherin. Even though DAB staining is visibly present at the membrane, it belongs to a

non-tumor region which was incorrectly considered as tumor. Improving tumor segmentation would clear this problem. In case D, E-cadherin protein expression was aberrant in more than 50% of the cells and thus the pathologist classified it as 3+. Despite its vast extension, the pattern of E-cadherin aberrant expression is dotted and incomplete membranous, which resulted in the

underestimation of DAB staining predicted by the computer, scored 2+.

All the TMA slides were processed to create co-expression maps for visual assessment of the relationship between CD44v6 and E-cadherin expression. In the majority of the cases, core alignment was successfully achieved and the morphological structures matched. Despite thin TMA slices, protein co-localization with a single cell resolution is not achievable since subcellular details are not shared between slides. Nonetheless, the overlap of both proteins brings important insights regarding the heterogeneity degree of a tumor core. In Fig. 3 we highlight three cores which illustrate the level of heterogeneity that a single core may harbor. For instance, the first core shows a unique and homogeneous population that co-expresses both proteins, while the other cores harbor multiple populations, the majority expressing one of both proteins and others that co-express both.

At a glance, this type of map helps the quick identification of cores co-expressing both proteins and the assessment of the level of tumor heterogeneity that was previously challenging to do by visual assessment. For example, to evaluate protein co-expression, the pathologist starts by visually identifying the tumor areas in the core and then evaluates the pattern for each individual protein in the tumor area. Only after that, the pathologist displays two separate images side by side, mentally divides the image in quarters and starts his/her estimation of expression level of both proteins simultaneously. On average, the manual assessment of concomitant protein expression takes between 5-8 minutes per step. Even with the support of another operator/researcher, it is unlikely that all the information mentioned by the pathologist during his/her assessment is properly documented. In contrast, our pipeline processes one slide containing 60 cores in 30 minutes (30 seconds per core), saving all the information processed in each step of the analysis. Therefore, full images, tumor segmentation, protein intensity and patterns are documented and can be revisited and re-analyzed at any time. Note that our pipeline is not an end-to-end software product; each module of the pipeline should be executed independently which gives the opportunity to verify the result of each step. The clinical value of the information has to be evaluated and verified by an expert, and further studies have to be carried out to relate the measurements to relevant biological and clinical information.

Research is moving away from categorical scores to continuous scores such as the results of our pipeline. Either way, we believe that the contribution of pathologist and computer should not be independent of each other but complementary.

V. DISCUSSION AND FUTURE WORK

Classifying tissue automatically is never a trivial task. We have presented here an initial approach with a methodology that may be improved by using more advanced components and a continuous feedback from pathologist to researcher. Also, an evaluation that includes intra and inter pathologist variation would give a better recognition of difficult-to-grade cases and more fairly define faults of the computational approach.

Faint and misty protein stainings are one of the most frequent sources of equivocal information that lead to mistakes in the subsequent steps of the automated method. In the future, there is room to improve the performance of our method by refining the distinction of multiple aberrant patterns that E-cadherin can assume in the tumor nest and explore the intensity patterns of a protein.

The computational quantification may be improved by better tumor segmentation, and taking into account the expression patterns of both proteins. Our method reports better results for CD44v6 compared to E-cadherin, because its patterns are less complex and simpler to quantify.

Data analysis is tightly coupled with good data visualization where efficient visualization methods are able to summarize multiple facets of data and can bring new insights. Our process and interface with the pathologists was improved by developing a *core explorer* visualization tool for IHC data to explore cores according to their pathologist given grade and the percentage calculated with our method. The information generated throughout the computer quantification process (as color unmixing and tumor segmentation) is displayed in the interface. At a glance, researchers and pathologists can quickly observe the distribution of the cores, inspecting the outliers or any technical mistake or bias in a specific slide. Inspecting the cores was crucial to understand the difficulties of the pipeline and even help with the correction in a few scores.

Beyond the scope of this paper, we can ask new questions such as: can the co-expression maps be developed further for the study of multiple markers? Can the present method be used to reveal co-localization at the subcellular level? Can any particular co-expression patterns within the tumor be used as reference for analysis of other markers? If so, they can become labels for other pieces of tissue aligned to the same context.

We believe that the proposed pipeline for registration and quality control for TMA data could be highly valuable in learning-based approaches to explore the clinical and molecular impact of multiple protein markers and improve our knowledge in intra- and inter-tumor/patient heterogeneity. Given the variability associated with manual assessment and reporting of IHC results, the key strength of our method is to extract, register and easily document the analysis of multiple protein markers. Using our pipeline, researchers are able to run preliminary analysis in a high throughput manner, on the direction of their biological or clinical question. This strategy is expected to save pathologists' time, while serving as the basis for faster verification, validation, adjustment and documentation of data. In future work, combining rare markers with common markers can provide invaluable and unbiased information for network training.

ACKNOWLEDGMENT

The authors thank Dina Leitão and Gilza Gonçalves for the technical support with TMA slides processing, Mário Seixas for technical support with TMA slides digitalization and Johan Öfverstedt for his help with the registration framework.

REFERENCES

- [1] F. Bray *et al.*, “Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” *CA Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, Nov. 2018, doi: [10.3322/caac.21492](https://doi.org/10.3322/caac.21492).
- [2] E. C. Smyth, M. Verheij, W. Allum, D. Cunningham, A. Cervantes, and D. Arnold, “Gastric cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up,” *Ann. Oncol.*, vol. 27, no. 5, pp. v38–v49, Sep. 2016, doi: [10.1093/annonc/mdw350](https://doi.org/10.1093/annonc/mdw350).
- [3] C. R. Taylor, “Introduction to predictive biomarkers: Definitions and characteristics,” in *Predictive Biomarkers in Oncol.: Applications in Precision Med.*, 1st ed. Berlin, Germany: Springer, 2019, ch.1 pp. 3–18.
- [4] P. Bankhead *et al.*, “QuPath: Open source software for digital pathology image analysis,” *Sci. Rep.*, vol. 7, 2017, Art. no. 16878, doi: [10.1038/s41598-017-17204-5](https://doi.org/10.1038/s41598-017-17204-5).
- [5] M. Linkert *et al.*, “Metadata matters: Access to image data in the real world,” *J. Cell Biol.*, vol. 189, no. 5, pp. 777–782, doi: [10.1083/jcb.201004104prognostic](https://doi.org/10.1083/jcb.201004104prognostic).
- [6] P. Carneiro *et al.*, “E-cadherin dysfunction in gastric cancer-Cellular consequences, clinical applications and open questions,” *FEBS Lett.*, vol. 586, no. 18, pp. 2981–2989, Aug. 2012, doi: [10.1016/j.febslet.2012.07.045](https://doi.org/10.1016/j.febslet.2012.07.045).
- [7] C. B. da Cunha *et al.*, “De novo expression of CD44 variants in sporadic and hereditary gastric cancer,” *Lab. Investigation*, vol. 90, no. 11, pp. 1604–1614, Nov. 2010, doi: [10.1038/labinvest.2010.155](https://doi.org/10.1038/labinvest.2010.155).
- [8] A. Mezheyski *et al.*, “Multispectral imaging for quantitative and compartment-specific immune infiltrates reveals distinct immune profiles that classify lung cancer patients,” *J. Pathol.*, vol. 244, no. 4, pp. 421–431, 2018, doi: [10.1002/path.5026](https://doi.org/10.1002/path.5026).
- [9] L. Solorzano *et al.*, “Whole slide image registration for the study of tumor heterogeneity,” in *Proc. OMIA, COMPAY*, 2018, vol. 11039, pp. 95–102, doi: [10.1007/978-3-030-00949-6_12](https://doi.org/10.1007/978-3-030-00949-6_12).
- [10] J. Öfverstedt, J. Lindblad, and N. Sladoje, “Fast and robust symmetric image registration based on distances combining intensity and spatial information,” *IEEE Trans. Image Process.*, vol. 28, no. 7, pp. 3584–3597, Jul. 2019, doi: [10.1109/TIP.2019.2899947](https://doi.org/10.1109/TIP.2019.2899947).
- [11] L. Solorzano *et al.*, “TissUMaps: Interactive visualization of large-scale spatial gene expression and tissue morphology data,” *Bioinformatics*, May 2020, doi: [10.1093/bioinformatics/btaa541](https://doi.org/10.1093/bioinformatics/btaa541).
- [12] OpenSeadragon An open-source, web-based viewer for high-resolution zoomable images, implemented in pure JavaScript, Accessed on: Oct. 2019 [Online]. Available: <https://openseadragon.github.io>
- [13] C. M. van der Loos, “Multiple immunoenzyme staining: Methods and visualizations for the observation with spectral imaging,” *J. Histochem. Cytochem.*, vol. 56, no. 4, pp. 313–328, 2008, doi: [10.1369/jhc.2007.950170](https://doi.org/10.1369/jhc.2007.950170).
- [14] A. Ruifrok and D. Johnston, “Quantification of histochemical staining by color deconvolution,” *Analytical Quantitative Cytol. Histol.*, vol. 4, no. 4, pp. 291–299, Aug. 2001.
- [15] C. Wemmert, J. M. Krüger, G. Forestier, L. Sternberger, F. Feuerhake, and P. Gañçarski, “Stain unmixing in brightfield multiplexed immunohistochemistry,” in *Proc. IEEE Int. Conf. Image Process.*, 2013, pp. 1125–1129, doi: [10.1109/ICIP.2013.6738232](https://doi.org/10.1109/ICIP.2013.6738232).
- [16] J. A. van Der Laak *et al.*, “Hue-saturation-density (HSD) model for stain recognition in digital images from transmitted light microscopy,” *Cytometry*, vol. 39, no. 4, pp. 275–284, 2000.
- [17] B. Zitová *et al.*, “Image registration methods: A survey,” *Image Vision Comput.*, vol. 21, no. 11, pp. 977–1000, Oct. 2003, doi: [10.1016/S0262-8856\(03\)00137-9](https://doi.org/10.1016/S0262-8856(03)00137-9).
- [18] N. Trahearn *et al.*, “A fast method for approximate registration of whole-slide images of serial sections using local curvature,” in *Proc. SPIE 9041, Med. Imag. Digit. Pathol.*, Mar. 2014, Art. no. 90410E, doi: [10.1117/12.2043308](https://doi.org/10.1117/12.2043308).
- [19] A. Sarkar, Q. Yuan, and C. Srinivas, “A robust method for inter-marker whole slide registration of digital pathology images using lines based features,” in *Proc. IEEE 11th Int. Symp. Biomed. Imag.*, Jul. 2014, pp. 762–765, doi: [10.1109/isbi.2014.6867982](https://doi.org/10.1109/isbi.2014.6867982).
- [20] L. Cooper *et al.*, “Feature-based registration of histopathology images with different stains: An application for computerized follicular lymphoma prognosis,” *Comput. Methods Programs Biomed.*, vol. 96, no. 3, pp. 182–192, Dec. 2009, doi: [10.1016/j.cmpb.2009.04.012](https://doi.org/10.1016/j.cmpb.2009.04.012).
- [21] D. Mueller *et al.*, “Real-time deformable registration of multi-modal whole slides for digital pathology,” *Computerized Med. Imag. Graph.*, vol. 35, no. 7–8, pp. 542–556, Dec. 2011, doi: [10.1016/j.compmedimag.2011.06.006](https://doi.org/10.1016/j.compmedimag.2011.06.006).
- [22] J. Pinchat, “A survey of methods for 3d histology reconstruction,” *Med. Image Anal.*, vol. 46, pp. 73–105, May 2018, doi: [10.1016/j.media.2018.02.004](https://doi.org/10.1016/j.media.2018.02.004).
- [23] G. Mandalain *et al.*, “Fusion of autoradiographs with an MR volume using 2-D and 3-D linear transformations,” *Neuroimage*, vol. 23, no. 1, pp. 111–127, Sep. 2004, doi: [10.1016/j.neuroimage.2004.04.038](https://doi.org/10.1016/j.neuroimage.2004.04.038).
- [24] X. Lopez *et al.*, “Registration of whole immunohistochemical slide images: An efficient way to characterize biomarker colocalization,” *J. Amer. Med. Informat. Assoc.*, vol. 22, no. 1, pp. 86–99, Jan. 2015, doi: [10.1136/amiajnl-2014-002710](https://doi.org/10.1136/amiajnl-2014-002710).
- [25] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, “Mutual-information-based registration of medical images: A survey,” *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 986–1004, Aug. 2003, doi: [10.1109/TMI.2003.815867](https://doi.org/10.1109/TMI.2003.815867).
- [26] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, “Image registration by maximization of combined mutual information and gradient information,” *IEEE Trans. Med. Imag.*, vol. 19, no. 8, pp. 809–814, Aug. 2000, doi: [10.1109/42.876307](https://doi.org/10.1109/42.876307).
- [27] M. Fernández-Delgado *et al.*, “Do we need hundreds of classifiers to solve real world classification problems?” *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3133–3181, 2014.
- [28] C. W. Wang, “Robust automated tumor segmentation on histological and immunohistochemical tissue images,” *PLOS One*, vol. 6, no. 2, Art. no. e15818, doi: [10.1371/journal.pone.0015818](https://doi.org/10.1371/journal.pone.0015818).
- [29] S. Berg *et al.*, “ilastik: interactive machine learning for (bio)image analysis,” *Nat. Methods*, vol. 16, pp. 1226–1232, 2019, doi: [10.1038/s41592-019-0582-9](https://doi.org/10.1038/s41592-019-0582-9).
- [30] S. V. Costes *et al.*, “Automatic and quantitative measurement of protein-protein colocalization in live cells,” *Biophys. J.*, vol. 86, no. 6, pp. 3993–4003, Jun. 2004, doi: [10.1529/biophysj.103.038422](https://doi.org/10.1529/biophysj.103.038422).
- [31] J. W. D. Comeau *et al.*, “A guide to accurate fluorescence microscopy colocalization measurements,” *Biophys. J.*, vol. 91, no. 12, pp. 4611–4622, Dec. 2006, doi: [10.1529/biophysj.106.089441](https://doi.org/10.1529/biophysj.106.089441).
- [32] S. Nembrini *et al.*, “The revival of the Gini importance?” *Bioinformatics*, vol. 34, no. 21, pp. 3711–3718, May 2018.