**Article**

# Robust Method for Unsupervised Scoring of Inmunohistoquemical Staining

Iván Durán-Díaz [*] , Auxiliadora Sarmiento , Irene Fondón , Clément Bodineau , Mercedes Tomé ,
Raúl V. Durán

*Article*

# Robust Method for Unsupervised Scoring of Inmunohistochemical Staining

**Iván Durán Díaz** [1],*[iD]**, Auxiliadora Sarmiento** [1][iD]**, Irene Fondón** [1][iD]**, Clément Bodineau** [2][iD]**, Mercedes Tomé** [3][iD] **and Raúl V. Durán** [3][iD]

[1]   Departamento de Teoría de la Señal y Comunicaciones, Universidad de Sevilla, Avda. Descubrimientos S/N, Seville, Spain; {sarmiento, iduran, irenef}@us.es
[2]   Department of Genetics, Harvard Medical School, Boston, MA, USA; cbodineau@bwh.harvard.edu
[3]   Centro Andaluz de Biología Molecular y Medicina Regenerativa-CABIMER, Consejo Superior de Investigaciones Científicas, Universidad de Sevilla, Universidad Pablo de Olavide, Seville, Spain; mercedes.tome@cabimer.es, raul.duran@cabimer.es
*   Correspondence: iduran@us.es; Tel.: +34-95-448-7334

**Abstract:** Immunohistochemistry is a powerful technique, widely used in biomedical research and clinics, that allows to determine the expression levels of some proteins of interest in tissue samples by means of the intensity of color due to the expression of biomarkers using specific antibodies. As such, immunohistochemical images are complex and problematic to be quantified. Recently we proposed a novel method including a first separation stage based on nonnegative matrix factorization (NMF) that achieved good results. However, that method was highly dependent on sparseness and non-negativity parameter choice and on algorithm initialization. Furthermore, the previously proposed method needed a reference image as starting point for the NMF algorithm. In the present work, we propose a new, simpler and robust method for the automated unsupervised scoring of immunohistochemical images based on bright field. Our work is focused on images from tumor tissues marked with blue (nuclei) and brown (protein of interest). The new proposed method represents a simpler approach that, in one side, avoids the use of NMF for the separation stage, and in the other side, circumvents the need for a control image. This new approach determines the subspace spanned by the two colors of interest by means of principal component analysis (PCA) with dimension reduction. This subspace is a two-dimensional space, allowing the finding of color vectors by considering the peaks of density of points. The method also develops a new scoring stage by avoiding, again, reference images, making the procedure more robust and less dependent on the parameters. Experiments for the semi-quantitative scoring of images in five categories exhibit promising and consistent results when compared to manual scoring carried out by experts.

**Keywords:** histopathological images; principal component analysis; unsupervised stain separation; semi-quantitative scoring

---

## 1. Introduction

Detection of biomarker expression in tissue images is a common technique both in research laboratories and in the clinic. This detection is carried out through the use of chromogens of different colors that bind to the antigens of interest using an antibody-antigen detection system, a technique known as Immunohistochemistry (IHC) [1]. This allows the presence of the chromogen and the protein to which it binds by observing the tissue using bright field microscopy. Capturing images of the tissue allows the subsequent preservation and analysis of the tissue expression pattern of analyzed proteins. Usually, the quantification of the presence, concentration and distribution of the chromogen is carried out manually by viewing and comparing the images by experts.

Two widely used chromogens are 3,3'-Diaminobenzidine (DAB) and Hematoxylin (H) [2–6]. While the presence of DAB is detected as a brown stain, H appears as a blue one. When combined, H stain marks cells nuclei, whereas DAB reveals the greater or lesser presence of the protein of interest.

The IHC image evaluation and quantification procedure lacks objectivity, as it is a subjective method done by experts, exposed to observer variations. Indeed, even the same observer can make different evaluations of the same image at different times [7]. Furthermore, manual scoring is affected by visual distortion due to the perception of colors and their intensity depending on the surrounding colors [8]. This is particularly important in protocols that use two dyes. Finally, IHC visual scoring it is a time-consuming procedure for experts. For this reason, demand for automated scoring procedures is increasing in laboratories and in the clinic [9]. This type of protocols must include a stage for separating both stains to perform the scoring with greater reliability.

Scoring systems can be classified on those based on the perceived intensity of staining, and those based on the percentages of stained cells, while some others mix both criteria. Regarding the most used categories, they range from the simplest systems that divide tissue samples into positive or negative, depending on whether or not the observed staining exceeds certain thresholds; to those that use a set of scoring levels depending on the greater or lesser presence of stain, the so-called semi-quantitative scoring systems [8]. Although some platforms and applications for semiquantitative scoring of IHC images already exist, they require the intervention of researchers as well as reference images. Examples of these systems are QuPath [10], IHC Profiler [11] or DAB-quant [12].

The stage of stain separation is necessary in automated scoring systems, since it allows the obtention of image features based on each stain. The separation stage cannot be performed in the RGB color space, because the concentration of the light-absorbing material follows the Beer-Lambert law (although not exactly, due to some degree of dispersion). For this reason, a transform based on this law is applied to the RGB image, obtaining the coordinates of each pixel in a new space: the optical density (OD). Within this space, the intensity of each coordinate can be considered as a linear combination of the intensities due to each staining [13].

Separation methods in the literature can be classified into two categories: supervised and unsupervised (or blind) methods [14,15]. In the first case, the coordinates of each used stain are known in the OD space; otherwise, they require the intervention of the observer or the use of reference images to complete their calculation. In the second case, unsupervised methods use techniques like independent component analysis (ICA) [16,17], non-negative matrix factorization (NMF) [13,18] or non-negative least squares (NNLS). These techniques suffer from a strong dependence on initialization. Thus, these methods need a previous estimation of the stain vectors coordinates to achieve a correct separation, guiding the search towards the correct solution. Some approaches which are based on deep learning [4,19] need a large number of training images, so they cannot be considered as unsupervised techniques.

Some approaches have exploited singular value decomposition (SVD) in order to calculate the subspace where the OD data lie [20,21] in the context of supervised methods. Principal component analysis (PCA) works in the same direction as SVD, searching for the principal direction of the data: those that preserve the greatest quantity of variance or power (when uncentred PCA is performed) [22].

In this work, we propose an unsupervised method for comparative biomarker quantification that improves the method presented in our previous work [14]. Accordingly, this new methodology represents a simpler and more robust method for color separation, with very low dependence on the selected parameters, and a more accurate quantification without the need for reference or control images.

We circumvent the use of NMF for the stain separation stage, whose results are very dependent on the starting point. Instead, the proposed method exploits the concentration of DAB (brown) and H (blue) to estimate the vectors associated to the color of both stains. This stage has been provided with robustness, in order to achieve consistent results for the estimation of the stain vectors with no prior information, even avoiding the use of reference images that the method in [14] needs.

For the scoring stage, features whose values mostly increase with the concentration of the stains helped us to develop a procedure for the initialization of the clustering algorithm (k-means). While

in our previous work a set of three parameters, together with the use of two reference images were needed, in the present work the initialization procedure provides excellent results for common values of the two parameters that we use and, again, avoiding the use of reference images. This makes totally unnecessary any intervention of the observer and allows a totally automated procedure.

## 2. Methods

We developed an automated and unsupervised procedure for the scoring of IHC images with two stains: Hematoxylin and DAB. The quantification of the set of images is done in three steps: stain separation, feature extraction and, finally, clustering and scoring. A semi-quantitative scoring system clusters the images in five groups and assigns scores from "1+" to "5+" for each image, being "1+" the lowest level of DAB expression.

### 2.1. Stain separation

The stain separation step is carried out within the color space given by the transform provided by the Beer-Lambert Law. Let $\mathbf{I}_m \in \mathbb{R}^{3 \times N}$ be the matrix of RGB intensities of the $m$th analyzed image, where $N$ is the total number of pixels of each color plane. Then the relative optical density matrix for this image, $\mathbf{X}_m$, is obtained as

$$\mathbf{X}_m = -\log\left(\frac{\mathbf{I}_m}{I_o}\right), \tag{1}$$

where $I_o$ is the maximum level of intensity at each pixel (255 for 8 bit images) and the function $\log(\mathbf{I}_m)$ is applied to each entry of the matrix $\mathbf{I}_m$. This provides a representation of each pixel in a $\mathbb{R}^3$ space where each color corresponds to a direction, being a higher darkness of the color represented by a higher norm of the vector. So, each color (including its darkness) can be represented by the coordinates of its corresponding vector within this space. Since the image was obtained by means of an IHC procedure, the color vector of each pixel lies, mainly, within the subspace generated by the color vectors associated to both stains: that due to H (blue) and that due to DAB (brown). This allows to estimate the matrix $\mathbf{X}_m$ as [13]

$$\hat{\mathbf{X}}_m = \mathbf{W}_m \mathbf{H}_m, \tag{2}$$

where $\hat{\mathbf{X}}_m$ is the matrix that estimates $\mathbf{X}_m$, $\mathbf{W}_m$ is a $3 \times 2$ matrix whose columns are the color vectors associated to both stains for $m$th image and $\mathbf{H}_m$ is a $2 \times N$ matrix whose rows are the intensities or activations of each column of $\mathbf{W}_m$ at each pixel of the $m$th image.

We exploit this fact in order to simplify the problem by means of dimension reduction. The subspace generated by the columns of $\mathbf{W}_m$ is found by means of the application of uncentred Principal Component Analysis (PCA) to the matrix $\mathbf{X}_m$. This provides the principal directions of the data, that lie within the above mentioned subspace (although the principal directions do not have to coincide with the columns of $\mathbf{W}_m$) as can be seen in Figure 1.
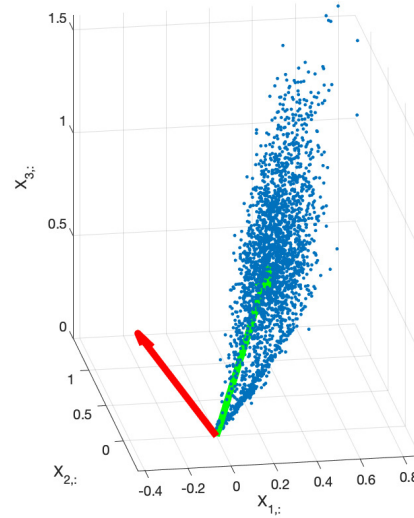
**Figure 1.** Scatter plot (blue) of the columns of matrix $\mathbf{X}_m$ for one example image with DAB and H stains. For a better visualization a sampling of the entries was made before to depict the scatter plot. As can be seen, the points lies mainly at the plain defined by the stain vectors. The principal directions (green and red), along the two first principal components are depicted. These orthogonal directions lie at the same plain defined by the color stain vectors, although they do not coincide. The principal components allow us to simplify the problem of searching the stain vectors to a 2D problem.

Let

$$\mathbf{Q}_m = \left[ \begin{array}{ccc} \mathbf{q}_{m1} & \mathbf{q}_{m2} & \mathbf{q}_{m3} \end{array} \right] \tag{3}$$

be the unitary matrix whose columns are the eigenvectors of $\mathbf{R}_m = \mathbf{X}_m \mathbf{X}_m^T / N$ and

$$\mathbf{L}_m = \mathrm{diag}\{L_{m1}, L_{m2}, L_{m3}\} \tag{4}$$

be the diagonal matrix whose nonzero entries are the eigenvalues of $\mathbf{R}_m$. The superscript $T$ stands for transpose. Both the eigenvectors and the eigenvalues are in descending order. Since $\mathbf{X_m} \approx \hat{\mathbf{X}}_m$, the principal subspace is given by the two first columns of $\mathbf{Q}_m$, as the third eigenvalue, $L_{m3}$, is very small compared to $L_{m1}$ and $L_{m2}$. The columns of $\mathbf{X}_m$ are projected onto the principal subspace (the plain defined by the first two principal direction), thus obtaining the $2 \times N$ matrix of principal components

$$\mathbf{Z}_m = \left[ \begin{array}{c} \sqrt{L_{m1}}^{-1} \cdot \mathbf{q}_{m1}^T \\ \sqrt{L_{m2}}^{-1} \cdot \mathbf{q}_{m2}^T \end{array} \right] \cdot \mathbf{X}_m. \tag{5}$$

Since all columns of $\mathbf{X}_m$ lie at the first octant, the color vectors are relatively close at $\mathbb{R}^3$ Euclidean space. A simple projection onto the $\mathbb{R}^2$ subspace defined by $\mathbf{q}_{m1}$ and $\mathbf{q}_{m2}$, keeping the scales of the principal components would lead to a 2D scattering where it would be difficult to precisely distinguish the color vectors associated to each stain. This is the reason why the columns of $\mathbf{X}_m$ are projected onto the subspace defined by $\mathbf{q}_{m1}\sqrt{L_{m1}}$ and $\mathbf{q}_{m2}\sqrt{L_{m2}}$, i.e., PCA with unit-power principal components is performed [22]. This expands the scattering on the second principal component direction, as we show in Figure 2, and allows to recover the stain color vectors more accurately.
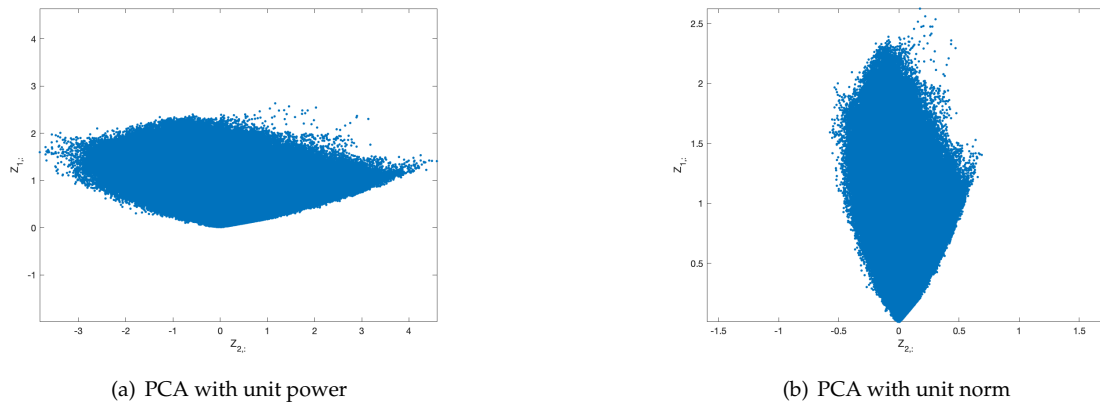
(a) PCA with unit power

(b) PCA with unit norm

**Figure 2.** Scatter plot of the principal components for an example image when a) PCA with unit power principal components or b) PCA with unit norm loading vectors are performed. Normalization of the principal components power allows to spread the angle of the scattering, so the angle between the stain vectors is greater, improving the accuracy of the estimate. In the other hand, normalization of the norm of loading vectors keep the distance between the stain vectors, which reduces the accuracy. The second principal component was depicted along horizontal axis for visual convenience.

### 2.1.1. Stain color basis estimation

The next step of the procedure consists in the estimation of the two basis vectors for the stain colors, i.e., the columns of the matrix $\mathbf{W}_m$. Due to the projection in (5) the columns of this matrix are projected into the same subspace defined by the principal directions

$$\mathbf{B}_m = \left[ \begin{array}{c} \sqrt{L_{m1}}^{-1} \cdot \mathbf{q}_{m1}^T \\ \sqrt{L_{m2}}^{-1} \cdot \mathbf{q}_{m2}^T \end{array} \right] \cdot \mathbf{W}_m. \tag{6}$$

Let $\theta_m(n)$ and $r_m(n)$ be the angle and norm of the $n$th column of $\mathbf{Z}_m$. Since, in many pixels, only one stain is dominant, the values of $r_m(n)$ are higher around the basis vectors. So the columns of $\mathbf{B}_m$ are estimated by means of the angles where large concentrations of high levels of $r_m(n)$ are found. The range of values of $\theta_m(n)$ is divided into 1000 bins for a sufficient angle resolution. For the $k$th bin, whose center is given by $\phi_m(k)$, the mean of the values of $r_m(n)$ such that $\theta_m(n)$ lies within the bin is computed. Then, this function is smoothed by a lowpass eighth-order Butterworth filter with digital cutoff frequency equal to 0.035 to get the function $v_m(k)$, as illustrated in Figure 3.
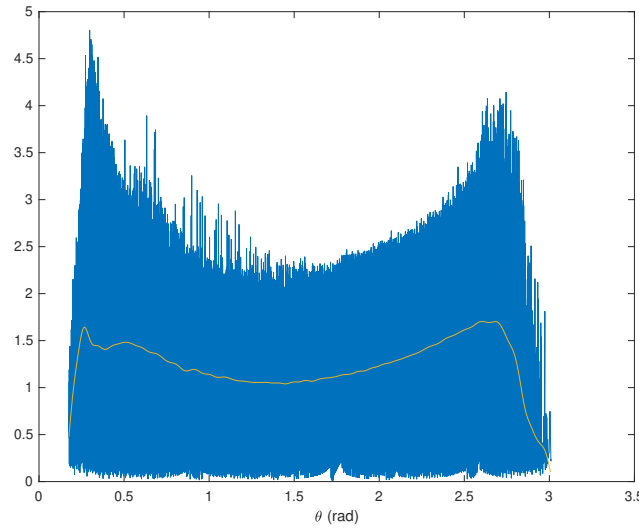
**Figure 3.** Norm $r_m(n)$ (blue) of the columns of $\mathbf{Z}_m$ for an example image versus the angle of the columns together with the smoothed mean of concentration, $v_m(k)$ (yellow). Searching for the peaks of this last function leads to the angles corresponding to the stain vectors within the subspace of principal components.

Therefore, $\mathbf{B}_m$ is estimated by means of $\widehat{\mathbf{B}}_m$, defined as

$$\widehat{\mathbf{B}}_m = \begin{bmatrix} \cos(\phi_{m_1}) & \cos(\phi_{m_2}) \\ \sin(\phi_{m_1}) & \sin(\phi_{m_2}) \end{bmatrix} \tag{7}$$

where $\phi_{m_1}$ and $\phi_{m_2}$ are the main peaks in the filtered version of $r_m(n)$. Since it is possible the existence of more than two peaks, in order to avoid a bad selection of the true angles of both basis vectors, a procedure was implemented. Due to the nature of the data, the basis vectors are always near the extrema of $\phi_m(n)$ (in an ideal case, the points of the scattering are a linear combination of the basis vectors with nonnegative coefficients). So, the maximum of $v_m(k)$ is selected as the position for one the basis vectors. The other peak is selected from the angle bins that lies within the opposite half of the total range of $\phi_m(k)$. This ensures that the selected angles correspond to a good estimate of the basis vectors.

It is easy to see that the estimate of $\mathbf{W}_m$ corresponding to $\widehat{\mathbf{B}}_m$ is given by the representation of the columns of this matrix in the whole 3D space as

$$\widehat{\mathbf{W}}_m = \begin{bmatrix} \mathbf{q}_{m_1}\sqrt{L_{m1}} & \mathbf{q}_{m_2}\sqrt{L_{m2}} \end{bmatrix} \cdot \widehat{\mathbf{B}}_m. \tag{8}$$

2.1.2. Color deconvolution

From (2) we can write the decomposition of $\mathbf{X}_m$ in terms of $\widehat{\mathbf{W}}_m$ and $\mathbf{H}_m$ as

$$\mathbf{X}_m = \widehat{\mathbf{W}}_m \cdot \mathbf{H}_m. \tag{9}$$

This describes, for each column of $\mathbf{H}_m$, an overdetermined system of equations as $\mathbf{W}_m$ and $\mathbf{X}_m$ are known. Therefore, the stain concentration matrix for the $m$th is estimated by means of a linear least squares problem as

$$\widehat{\mathbf{H}}_m = \widehat{\mathbf{W}}_m^+ \mathbf{X}_m \tag{10}$$

where $\widehat{\mathbf{W}}_m^+ = (\widehat{\mathbf{W}}_m^T \widehat{\mathbf{W}}_m)^{-1} \cdot \widehat{\mathbf{W}}_m$ is the Moore-Penrose pseudoinverse of the full-rank matrix $\widehat{\mathbf{W}}_m$ [23]. This procedure is known as color deconvolution [24]. Since the concentration of each stain at every pixel must be nonnegative, the values $[\widehat{\mathbf{H}}_m]_{ij} < 0$ are set to 0.

The estimate is improved by means of the following procedure. The values $[\widehat{\mathbf{H}}_m]_{ij}$ for columns of $\widehat{\mathbf{H}}_m$ that have only one nonzero entry are recalculated by using the corresponding column of $\widehat{\mathbf{W}}_m$. After this, a new estimate of the basis vectors is computed as the mean of the columns of $\mathbf{Z}_m$ for which one stain is dominant, i.e., when the value for a concentration is, at least, 10 times that for the other. A new estimate of $\widehat{\mathbf{H}}_m$ is computed following the procedure described from equation (10).

### 2.1.3. Average basis vectors

Once the previous procedure is carried out, the average of the matrices $\widehat{\mathbf{W}}_m$ is computed

$$\overline{\mathbf{W}} = \left\langle \widehat{\mathbf{W}}_m \right\rangle_m. \tag{11}$$

Then the concentration matrix is estimated for each image as $\overline{\mathbf{H}}_m = \overline{\mathbf{W}}^+ \mathbf{X}_m$. Again, the values $[\overline{\mathbf{H}}_m]_{ij} < 0$ are set to 0 and then the values $[\overline{\mathbf{H}}_m]_{ij}$ for columns of $\overline{\mathbf{H}}_m$ that have only one nonzero entry are recalculated.

### 2.2. Feature extraction

A set of four features is computed for each image. Since the columns of $\widehat{\mathbf{W}}_m$ and $\overline{\mathbf{W}}$ were not normalized, we take into account them together with $\widehat{\mathbf{H}}_m$ and $\overline{\mathbf{H}}_m$. The $j$th column of $\widehat{\mathbf{W}}_m$ (resp. $\overline{\mathbf{W}}$) is denoted as $\widehat{\mathbf{W}}_{m:,j}$ (resp. $\overline{\mathbf{W}}_{:,j}$) and the $i$th row of $\widehat{\mathbf{H}}_m$ (resp. $\overline{\mathbf{H}}_m$) is denoted as $\widehat{\mathbf{H}}_{m_{i,:}}$ (resp. $\overline{\mathbf{H}}_{m_{i,:}}$). Let $M_1(m)$ be the mean of the the intensity average of all the pixels of DAB stain for the $m$th image according to the results for the average basis vectors

$$M_1(m) = \frac{1}{3N} \left\| \overline{\mathbf{W}}_{:,2} \cdot \overline{\mathbf{H}}_{m_{2,:}} \right\|_1 = \frac{1}{3N} \left\| \overline{\mathbf{W}}_{:,2} \right\|_1 \cdot \left\| \overline{\mathbf{H}}_{m_{2,:}} \right\|_1 = \frac{1}{3N} \sum_{i=1}^{3} [\overline{\mathbf{W}}]_{i2} \sum_{j=1}^{N} [\overline{\mathbf{H}}_m]_{2j} \tag{12}$$

where the 1-norm of the matrix is computed entry-wise, i.e., it is the sum of the entries of the matrix (they are all nonnegative). For row vectors we computed the norm as for column vectors. Then the first feature for the $m$th image is defined as the average of the DAB intensity in this image, normalized by the maximum value that this quantity reaches among the entire set of images

$$f_1(m) = \frac{M_1(m)}{\max_m \{M_1(m)\}}. \tag{13}$$

The second feature is based on the relation between the robust maxima (99th percentile) of both stains: H and DAB. Let $R_1(m)$ (resp. $R_2(m)$) be the robust maximum of the intensities matrix for H stain (resp. DAB stain), according to the results for the average basis vectors, i.e., the robust maximum of the matrix $\overline{\mathbf{W}}_{:,1} \cdot \overline{\mathbf{H}}_{m_{1,:}}$ (resp. $\overline{\mathbf{W}}_{:,2} \cdot \overline{\mathbf{H}}_{m_{2,:}}$). Then, we define

$$\bar{R}_2(m) = \frac{R_2(m)}{\max_m R_2(m)} \tag{14}$$

and

$$\bar{R}_1(m) = \frac{R_1(m)}{R_1(m = \arg\max_m R_2(m))}. \tag{15}$$

The second feature is then defined as

$$f_2(m) = \frac{\bar{R}_2(m)}{\bar{R}_1(m)}. \tag{16}$$

The third feature is analogue to the first one, but using 2-norm, i.e., based on the average power of the intensity. Let $M_2(m)$ be

$$M_3(m) = \frac{1}{\sqrt{3N}} \left\| \overline{\mathbf{W}}_{:,2} \cdot \overline{\mathbf{H}}_{m_2,:} \right\|_F = \frac{1}{\sqrt{3N}} \left\| \overline{\mathbf{W}}_{:,2} \right\|_2 \cdot \left\| \overline{\mathbf{H}}_{m_2,:} \right\|_2$$

$$= \frac{1}{\sqrt{3N}} \sqrt{\sum_{i=1}^{3} [\overline{\mathbf{W}}]_{i2}^2} \sqrt{\sum_{j=1}^{N} [\overline{\mathbf{H}}_m]_{2j}^2} \quad (17)$$

where $\|\cdot\|_F$ is Frobenius norm of the matrix. Then we define the third feature, $f_3(m)$ as

$$f_3(m) = \frac{M_3(m)}{\max\limits_{m}\{M_3(m)\}}. \quad (18)$$

Although very goods results are achieved when using the above defined features, the inclusion of a fourth one improves the behavior of the clustering procedure. The feature $f_4(m)$ is defined as the square root of the average power of the intensity of the DAB stain at every pixel according to the results for the basis vectors estimated for each image,

$$f_4(m) = \frac{1}{\sqrt{3N}} \left\| \widehat{\mathbf{W}}_{m:,2} \cdot \widehat{\mathbf{H}}_{m_2,:} \right\|_F = \frac{1}{\sqrt{3N}} \left\| \widehat{\mathbf{W}}_{m:,2} \right\|_2 \cdot \left\| \widehat{\mathbf{H}}_{m_2,:} \right\|_2 \quad (19)$$

*2.3. Clustering and scoring*

As a general behavior the values of the four defined features increase when the concentration of DAB stain increases. Consequently, a simple, automated an efficient procedure was used to initialize the algorithm k-means for clustering. Also, this property allows us to assign the scores in a increasing order of the centroids norms. Let $\mathbf{f}(m)$ be the vector of features whose $i$th entry is $f_i(m)$. In order to initialize the clustering algorithm we computed the $p$-norm of the vector

$$\|\mathbf{f}(m)\|_p = \left( \sum_{i=1}^{T} |f_i(m)|^p \right)^{1/p} \quad (20)$$

where $T$ is the length of the vector of features $\mathbf{f}(m)$. A certain range of values can lead to good initialization and results for the clustering algorithm. In section 3 we discuss about these values.

When the values of $\|\mathbf{f}(m)\|_p$ are sorted in ascending order the result is not a linear function of $m$, but rather a power function. For an automated initialization of the clustering algorithm, the range of values of $\|\mathbf{f}(m)\|_p^q$ are divided into 5 intervals with the same length, which provides a good starting point. Therefore, the $k$th interval is $I_k = \left[ E_k^{(i)}, E_k^{(s)} \right]$ where

$$E_k^{(i)} = \frac{k-1}{5} \left( \max \|\mathbf{f}(m)\|_p^q - \min \|\mathbf{f}(m)\|_p^q \right) + \min \|\mathbf{f}(m)\|_p^q \quad (21)$$

$$E_k^{(s)} = \frac{k}{5} \left( \max \|\mathbf{f}(m)\|_p^q - \min \|\mathbf{f}(m)\|_p^q \right) + \min \|\mathbf{f}(m)\|_p^q. \quad (22)$$

Although the value of $q$ could change for different sets of features, square root is a good selection for the four proposed. We followed the criterion of fitting the first and second half of the sorted values of $\|\mathbf{f}(m)\|_p$ to a power function and set $q$ as the average of both resulting powers. For the selected 4 features, the value is close to 2. In section 3 we discuss about the possible values of this parameter. The

initial clusters are assigned based on the norm of the features vector; if this norm lies in the $k$th interval, then the corresponding vector belongs to the $k$th cluster. So the initial $k$th cluster, $\Omega_k^{(0)}$, is set as

$$\Omega_k^{(0)} = \left\{ m : \ \|\mathbf{f}(m)\|_p^q \in I_k \right\} \quad \forall k = 1, \ldots, 5. \tag{23}$$

The initial $k$th centroid is set as the mean of the features vectors that belong to the $k$th cluster

$$\mathbf{c}_k = \sum_{m \in \Omega_k} \mathbf{f}(m) \quad \forall k = 1, \ldots, 5. \tag{24}$$

These initial centroids were used for k-means algorithm equiped with a measure of distance based on $p$-norm. Only 3 iterations of the algorithm are needed to achieve the scores of the images.

## 3. Results and discussion

Our dataset was the same as in [14] and consisted of 94 images taken from stained xenograph tumors, using HCT116 cells implanted in mice treated with either vehicle or with a cell-permeable $\alpha$-ketoglutarate derivative, dimethyl$\alpha$-KG (DMKG), followed by treatment with the mTORC1 inhibitor temsirolimus (TEM) or with metformin (MET). Samples were processed as described in [25]. All procedures were approved by the corresponding institutional organizations (APAFIS# 10090 2017052409402562 v2). Omission of the primary antibody in the immunostaining procedure was used as a negative control. Images were acquired in TIFF format with a Leica DM6000B microscope using $\times 20$ or $\times 40$ objective lenses and a Leica DFC500 digital camera. All images were independently scored from "1+" to "5+" by four expert observers. The agreement among observers measured as the percentage of images that have been annotated with the same score by all experts is 70.21%. This makes that the maximum mean agreement of a scoring with all observers be 90.96%.

### 3.1. Results for stain separation step

The stain separation procedure achieves satisfactory results, providing a good decomposition of DAB and H planes, showing consistent solutions along the whole set of images. In Figure 4 the estimated color stain vectors for each image, together with the mean of these color vectors, are depicted. Note that the direction of the vectors is very similar for all the images. In Figures 5 and 6 the results for the separation procedure are show for two example images from the set. In the first case the processed image has a high concentration of both (DAB and H) stains, whereas the second image has a very low concentration of DAB satin. In both figures two results are shown: that for individual analysis without considering the rest of images, based on the estimate from (8); and the estimation of stain concentrations from the average stain vectors estimated in (11). In all cases the color of H stain is correctly estimated. Furthermore, even if the color of DAB stain for the second image is not very accurate at the first analysis, the calculated concentrations of this stain are similar to that obtained from the average stain vectors.
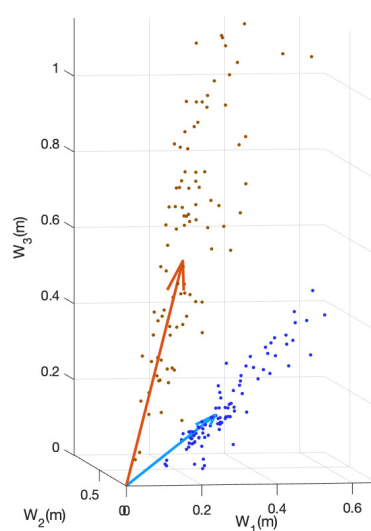
**Figure 4.** Scatter plot of the estimates $\widehat{\mathbf{W}}_m$ for the whole set of analyzed images and for both H (blue) and DAB (brown) stain colors, together with the columns of the mean $\overline{\mathbf{W}}$ also for both colors. The consistency of the results for nearly all images can be noted.
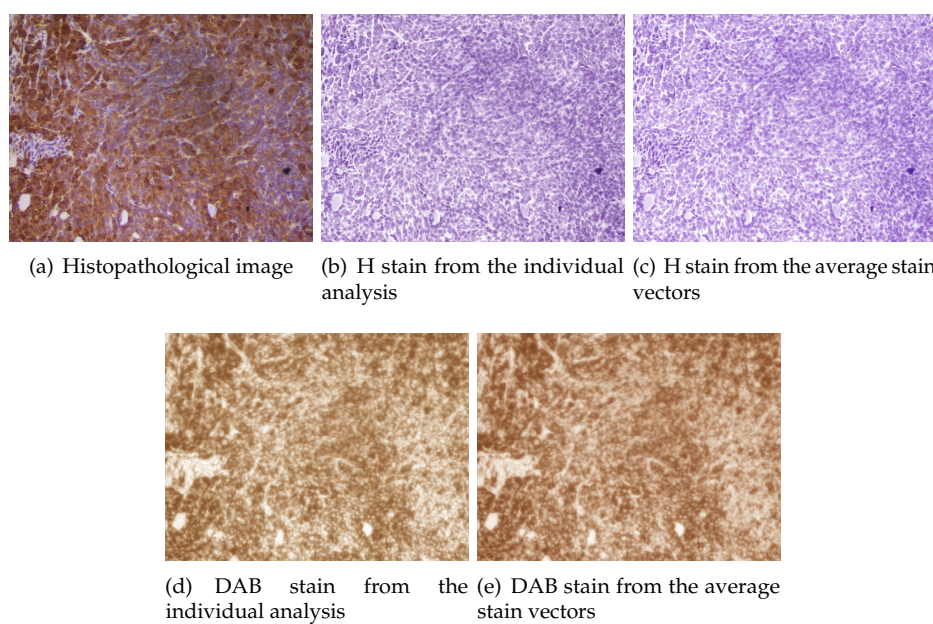


(a) Histopathological image

(b) H stain from the individual analysis

(c) H stain from the average stain vectors



(d) DAB stain from the individual analysis

(e) DAB stain from the average stain vectors

**Figure 5.** Example of stain separation of an IHC image form the set: a) Original image, b) H stain from the exclusive analysis of this image, c) H stain from the average stain vectors computed in (11), d) DAB stain from the analysis of this image and e) DAB stain from the average stain vectors. This is an example where the analysis of the image basically coincides with the average, since the intensities of both stains are high.
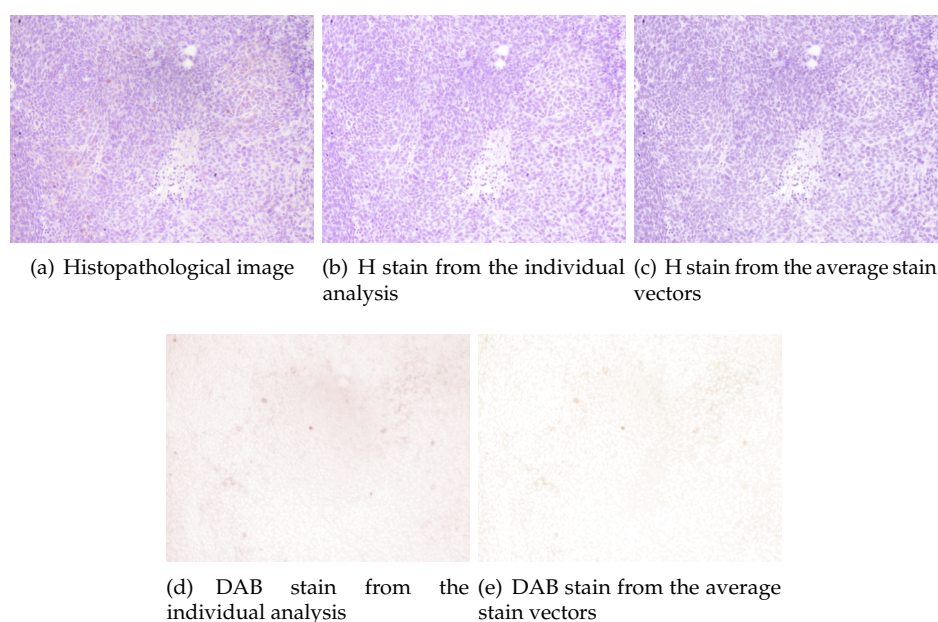
(a) Histopathological image

(b) H stain from the individual analysis

(c) H stain from the average stain vectors

(d) DAB stain from the individual analysis

(e) DAB stain from the average stain vectors

**Figure 6.** Example of stain separation of another IHC image form the set: a) Original image, b) H stain from the exclusive analysis of this image, c) H stain from the average stain vectors computed in (11), d) DAB stain from the analysis of this image and e) DAB stain from the average stain vectors. This is an example where the exclusive analysis of the image does not provides a good estimate of the DAB stain vector, due to the very low concentration of this stain, although the intensity of the separated DAB stain is very similar to that obtained for the average DAB vector.

### 3.2. Performance of the scores prediction

In [14] we showed that 1-norm of the DAB concentration stain correlates better than average threshold method (ATM) score and pixel-wise H-score to the scorings of the observers. There we used two other features (related to that denoted as $f_3(m)$ in the present work, but not the same) that improved the correlation. Here we show how the correlation for 1-norm increases when we add the features $f_2(m)$, $f_3(m)$ and $f_4(m)$. All the features that we propose increase (in general) when the score increases, which allows us to use the norm of the feature vector for an automated initialization of the clustering algorithm, due to the high correlation between this norm and the score of each image, as it can be seen in Figure 7.
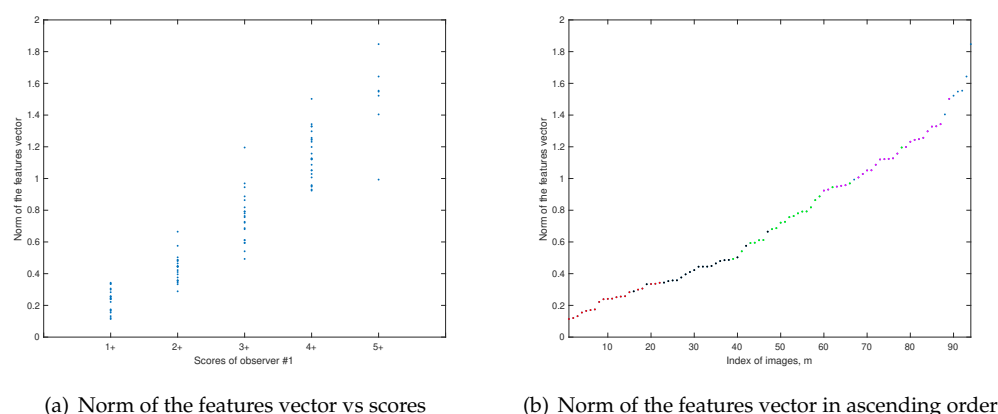
(a) Norm of the features vector vs scores

(b) Norm of the features vector in ascending order

**Figure 7.** Correlation between the norm of the features vector and the scores of the observer #1 shown as a plot of a) the norm versus the scores and b) the sorted norm versus the indices of the images. In this last figure the scores are shown in different colors in ascending order. In both cases the high correlation between the norm and the scores can be observed. This is used for the automated initialization of the clustering algorithm.

Furthermore, the use of only the first three features allows to achieve excellent results by using the above mentioned automated procedure for the initialization. Figure 8 shows a comparison of the prediction scores and the median of the observers' scores for each triple of feature values. Despite small differences the automated scoring achieves a high level of similarity with the reference truth.

Although the procedure was conceived to use four features, we also tested its behavior when only two or three features are exploited. Table 1 shows the main results for 2, 3 and 4 features in terms of percentage of coincidences with each observer and with the median of the observers' scores. Features $f_1$ and $f_2$ were the couple used for testing our method with 2 features. After, we added $f_3$ and, finally, $f_4$. A high level of coincidence is achieved for 3 and 4 features, whereas the level of coincidences for two features was acceptable.
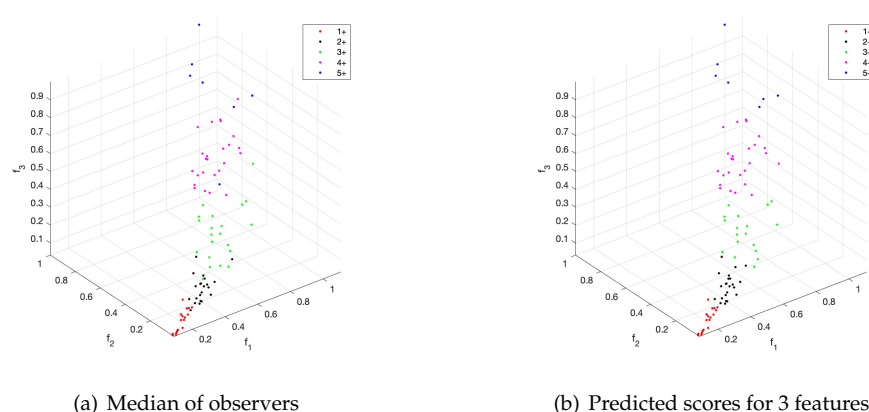


(a) Median of observers

(b) Predicted scores for 3 features

**Figure 8.** Scatter plot of the fisrt 3 features together with: a) median of the scores assigned by observers and b) predicted scores for these features after clustering algorithm.

**Table 1.** Accuracy achieved by the proposed algorithm (% of coincidences with observers). The use of two, three or four features was compared.

| Reference | | $f_1, f_2$[1] | $f_1, f_2, f_3$[2] | $f_1, f_2, f_3, f_4$[3] |
|---|---|---|---|---|
| Observer 1 | | 86.17 | 89.36 | 90.43 |
| Observer 2 | | 78.72 | 79.79 | 78.72 |
| Observer 3 | | 85.11 | 88.30 | 89.36 |
| Observer 4 | | 86.17 | 89.36 | 90.43 |
| | Mean of results[4] | 84.04 | 86.70 | 87.23 |
| Observer's median[5] | | 88.30 | 91.49 | 92.55 |

[1] $p = 1.5$, $q = 1/1.8$, 1 iteration. [2] $p = 1.9$, $q = 1/1.75$ 2 iterations. [3] $p = 2.5$, $q = 2$ 3 iterations.
[4] Mean of above results
[5] Comparison to the median of the observers' scores.

A wide range of values are suitable for the parameters $p$ ($p$-norm for initialization and distance in clustering algorithm) and $q$ (the power for reshaping the curve of values of the norm of the features vector). Figure 9 shows the percentage of coincidences of the scoring assigned by the proposed algorithm with the median of the scores assigned by the observers. For values of more than 90% low correlation between $p$ and $q$ can be observed. This leads to consider the proposed method quite robust, in terms of the selection of the parameters, which enables an automated procedure for the scoring of the IHC images.
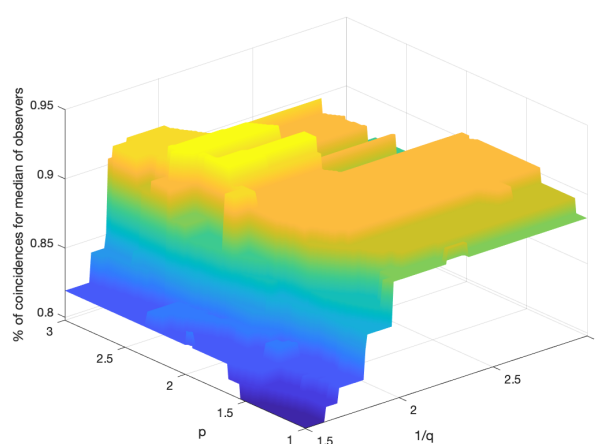


**Figure 9.** Percentage of coincidences of the scoring procedure with the median of the observers for the use of four features versus parameters $p$ (norm) and $1/q$ (root applied to the norm of the features vector for the initialization).

## 4. Conclusions

In recent years, automated scoring methods have been developed for the evaluation of IHC images. However, these methods suffer from problems related to the need for researchers to intervene at a certain point in the process, the use of control or training images or the dependence of the results on the choice of certain parameters.

In the present work we have proposed a completely unsupervised method that allows scoring IHC images without the need for reference images and without dependence on the set up of several parameters. The procedure consists of a first stage of separation of the DAB and H stains. In this

stage, PCA allows working in a lower dimensional space (a plane), which allows finding the vectors associated with both stains in a simple and efficient way. Once these vectors and the matrix of intensities of both stains are obtained, four characteristics are used to cluster the images, although it has been proven that the use of three of them provides very good results and, even, the use of only two achieves acceptable results. The clustering method used (k-means) is based on automated initialization, without the need for reference images or parameter setup, making it a robust unsupervised method.

The proposed method has been tested with a database of 94 images and the reference truth was taken from four expert observers. The scoring was made up of five categories and the algorithm reached a mean percentage of correct predictions of 87.23% for a maximum of 90.96% (due to the experts' opinion percentage of coincidences, which limited the average maximum percentage of correct predictions). Adapting the method to a different number of categories would be immediate, as in our previous work.

As a future line of work, we consider it interesting to find simple, fast and effective methods to locate nuclei and to develop new features related to the DAB stain of nuclei, since this could help to improve the results.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data that support this study are available from the author R.V.D., upon reasonable request.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| IHC | Immunohistochemistry |
| RGB | Red-Green-Blue |
| DAB | 3,3'-Diaminobenzidine |
| H | Hematoxylin |
| CD | Color Deconvolution |
| NMF | Non-Negative Matrix Factorization |
| OD | Optical Density |
| PCA | Principal Component Analysis |

## References

1. Miettinen, M. Immunohistochemistry of soft tissue tumours - review with emphasis on 10 markers. *Histopathology* **2014**, *64*, 101–118. doi:10.1111/his.12298.
2. Filliol, A.; Saito, Y.; Nair, A.; Dapito, D.H.; Yu, L.X.; Ravichandra, A.; Bhattacharjee, S.; Affo, S.; Fujiwara, N.; Su, H.; Sun, Q.; Savage, T.M.; Wilson-Kanamori, J.R.; Caviglia, J.M.; Chin, L.; Chen, D.; Wang, X.; Caruso, S.; Kang, J.K.; Amin, A.D.; Wallace, S.; Dobie, R.; Yin, D.; Rodriguez-Fiallos, O.M.; Yin, C.; Mehal, A.; Izar, B.; Friedman, R.A.; Wells, R.G.; Pajvani, U.B.; Hoshida, Y.; Remotti, H.E.; Arpaia, N.; Zucman-Rossi, J.; Karin, M.; Henderson, N.C.; Tabas, I.; Schwabe, R.F. Opposing roles of hepatic stellate cell subpopulations in hepatocarcinogenesis. *Nature* **2022**, *610*, 356–365. doi:10.1038/s41586-022-05289-6.

3. Grosset, A.A.; Loayza-Vega, K.; Adam-Granger, É.; Birlea, M.; Gilks, B.; Nguyen, B.; Soucy, G.; Tran-Thanh, D.; Albadine, R.; Trudel, D. Hematoxylin and Eosin Counterstaining Protocol for Immunohistochemistry Interpretation and Diagnosis. *Applied Immunohistochemistry & Molecular Morphology* **2019**, *27*.

4. Nielsen, P.S.; Georgsen, J.B.; Vinding, M.S.; Østergaard, L.R.; Steiniche, T. Computer-Assisted Annotation of Digital H&E/SOX10 Dual Stains Generates High-Performing Convolutional Neural Network for Calculating Tumor Burden in H&E-Stained Cutaneous Melanoma. *International Journal of Environmental Research and Public Health* **2022**, *19*. doi:10.3390/ijerph192114327.

5. Konukiewitz, B.; Schmitt, M.; Silva, M.; Pohl, J.; Lang, C.; Steiger, K.; Halfter, K.; Engel, J.; Schlitter, A.M.; Boxberg, M.; Pfarr, N.; Wilhelm, D.; Foersch, S.; Tschurtschenthaler, M.; Weichert, W.; Jesinghaus, M. Loss of CDX2 in colorectal cancer is associated with histopathologic subtypes and microsatellite instability but is prognostically inferior to hematoxylin-eosin-based morphologic parameters from the WHO classification. *Br J Cancer* **2021**, *125*, 1632–1646. doi:10.1038/s41416-021-01553-0.

6. van der Loos, C.M. Multiple immunoenzyme staining: methods and visualizations for the observation with spectral imaging. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* **2008**, *56*, 313–328. doi:10.1369/jhc.2007.950170.

7. Kitaya, K.; Yasuo, T. Inter-observer and intra-observer variability in immunohistochemical detection of endometrial stromal plasmacytes in chronic endometritis. *Exp Ther Med* **2013**, *5*, 485–488. doi:10.3892/etm.2012.824.

8. Aeffner, F.; Wilson, K.; Martin, N.T.; Black, J.C.; Hendriks, C.L.L.; Bolon, B.; Rudmann, D.G.; Gianani, R.; Koegler, S.R.; Krueger, J.; Young, G.D. The Gold Standard Paradox in Digital Image Analysis: Manual Versus Automated Scoring as Ground Truth. *Archives of Pathology and Laboratory Medicine* **2017**, *141*, 1267–1275, [https://meridian.allenpress.com/aplm/article-pdf/141/9/1267/1449767/arpa_2016-0386-ra.pdf]. doi:10.5858/arpa.2016-0386-RA.

9. David K Meyerholz, A.P.B. Fundamental Concepts for Semiquantitative Tissue Scoring in Translational Research. *ILAR Journal* **2018**, *59*, 13–17. doi:https://doi.org/10.1093/ilar/ily025.

10. Bankhead, P.; Fernández, J.; McArt, D.G.; Boyle, D.P.; Li, G.; Loughrey, M.B.; Irwin, G.W.; Harkin, D.P.; James, J.A.; McQuaid, S.; Salto-Tellez, M.; Hamilton, P.W. Integrated tumor identification and automated scoring minimizes pathologist involvement and provides new insights to key biomarkers in breast cancer. *Laboratory Investigation* **2018**, *98*, 15–26. doi:10.1038/labinvest.2017.131.

11. Varghese, F.; Bukhari, A.B.; Malhotra, R.; De, A. IHC Profiler: an open source plugin for the quantitative evaluation and automated scoring of immunohistochemistry images of human tissue samples. *PLoS One* **2014**, *9*, e96801. doi:10.1371/journal.pone.0096801.

12. Patel S, Fridovich-Keil S, R.S.F.K.J. DAB-quant: An open-source digital system for quantifying immunohistochemical staining with 3,3?-diaminobenzidine (DAB). *PLoS ONE* **2022**, *17*. doi:https://doi.org/10.1371/journal.pone.0271593.

13. Vahadane, A.; Peng, T.; Sethi, A.; Albarqouni, S.; Wang, L.; Baust, M.; Steiger, K.; Schlitter, A.M.; Esposito, I.; Navab, N. Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images. *IEEE Transactions on Medical Imaging* **2016**, *35*, 1962–1971.

14. Sarmiento, A.; Durán-Díaz, I.; Fondón, I.; Tomé, M.; Bodineau, C.; Durán, R.V. A Method for Unsupervised Semi-Quantification of Inmunohistochemical Staining with Beta Divergences. *Entropy* **2022**, *24*. doi:10.3390/e24040546.

15. Roy, S.; kumar Jain, A.; Lal, S.; Kini, J. A study about color normalization methods for histopathology images. *Micron* **2018**, *114*, 42–61. doi:https://doi.org/10.1016/j.micron.2018.07.005.

16. Trahearn, N.; Snead, D.; Cree, I.; Rajpoot, N. Multi-class stain separation using independent component analysis. Medical Imaging 2015: Digital Pathology; Gurcan, M.N.; Madabhushi, A., Eds. International Society for Optics and Photonics, SPIE, 2015, Vol. 9420, pp. 113 – 123. doi:10.1117/12.2081933.

17. Alsubaie, N.; Trahearn, N.; Raza, S.; Snead, D.; Rajpoot, N. Stain Deconvolution Using Statistical Analysis of Multi-Resolution Stain Colour Representation. *PLoS ONE* **2017**, *12(1)*. doi:https://doi.org/10.1371/journal.pone.0169875.

18. Li, X.; Plataniotis, K.N. A Complete Color Normalization Approach to Histopathology Images Using Color Cues Computed From Saturation-Weighted Statistics. *IEEE Transactions on Biomedical Engineering* **2015**, *62*, 1862–1873. doi:10.1109/TBME.2015.2405791.

19.     Bencze, J.; Szarka, M.; Kóti, B.; Seo, W.; Hortobágyi, T.G.; Bencs, V.; Módis, L.V.; Hortobágyi, T. Comparison of Semi-Quantitative Scoring and Artificial Intelligence Aided Digital Image Analysis of Chromogenic Immunohistochemistry. *Biomolecules* **2022**, *12*. doi:10.3390/biom12010019.

20.     Macenko, M.; Niethammer, M.; Marron, J.S.; Borland, D.; Woosley, J.T.; Guan, X.; Schmitt, C.; Thomas, N.E. A method for normalizing histology slides for quantitative analysis. 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2009, pp. 1107–1110. doi:10.1109/ISBI.2009.5193250.

21.     Salvi, M.; Michielli, N.; Molinari, F. Stain Color Adaptive Normalization (SCAN) algorithm: Separation and standardization of histological stains in digital pathology. *Computer Methods and Programs in Biomedicine* **2020**, *193*, 105506. doi:https://doi.org/10.1016/j.cmpb.2020.105506.

22.     Jolliffe, I.; Springer-Verlag. *Principal Component Analysis*; Springer Series in Statistics, Springer, 2002.

23.     Golub, G.H.; van Loan, C.F. *Matrix Computations*, fourth ed.; JHU Press, 2013.

24.     Ruifrok, A.C.; Johnston, D.A. Quantification of histochemical staining by color deconvolution. *Anal Quant Cytol Histol* **2001**, *23*, 291–299.

25.     Bodineau, C.; Tomé, M.; Courtois, S.; Costa, A.S.H.; Sciacovelli, M.; Rousseau, B.; Richard, E.; Vacher, P.; Parejo-Pérez, C.; Bessede, E.; Varon, C.; Soubeyran, P.; Frezza, C.; Murdoch, P.d.S.; Villar, V.H.; Durán, R.V. Two parallel pathways connect glutamine metabolism and mTORC1 activity to regulate glutamoptosis. *Nature Communications* **2021**, *12*, 4814. doi:10.1038/s41467-021-25079-4.