SUPPLEMENTARY MATERIAL

A. TIS example

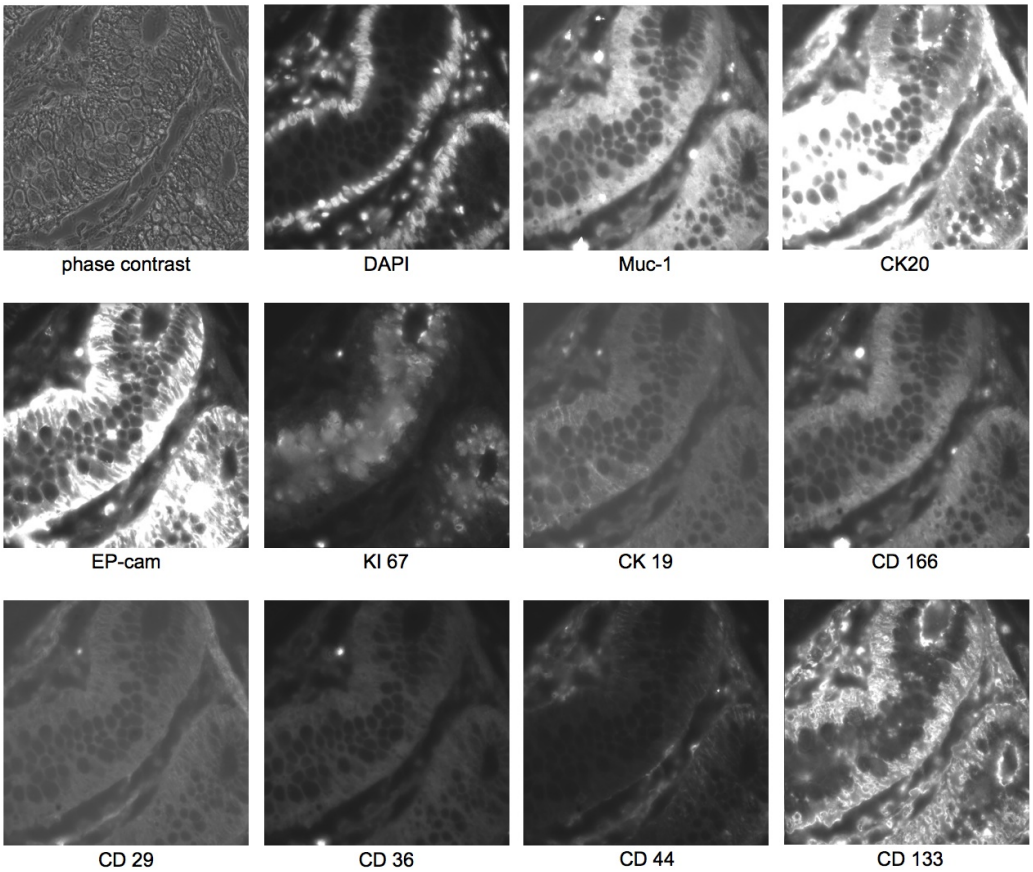

FIGURE 1. Example TIS image obtained from a field of view in a colon tissue section classified as normal, i. e. according to pathological criteria the tissue showed no signs of a tumor development. The figure shows the $N = 11$ fluorescence micrographs plus one phase contrast image. The other TIS images can be inspected in the BioIMAX system using the provided access information.

B. The Hierarchical Hyperbolic Self-Organizing Map (H^2SOM)

In this section we give a brief motivation for implementing SOMs in the hyperbolic space, give a brief introduction into the foundations of hyperbolic space and describe some of its features.

Foundations. In the beginning, the hyperbolic space was just an idea which was born by the question for non-Euclidean geometry, which was basically driven by questioning the 5th axiom of Euclid. In the beginning of the 19th century, Lobachevsky, Gauss and Bolyai were the first to deny this axiom and created the theoretical notion of what is called today *hyperbolic geometry*. Later on, based on Riemannian geometry, different models have been proposed to represent the hyperbolic space in euclidean 3D with euclidean coordinates (x, y, z) computed from its original coordinates by a mapping function $x = f_x(u, v), y = f_y(u, v), z = f_z(u, v)$ from the hyperbolic space coordimates (u, v).

However, it is natural that there exists no representation of $\mathbb{H}^2$ in $\mathbb{R}^3$ which preserves all distances and angles. One of the most prominent representation models is the Poincaré Disk Model [10], which is quite similar to the Klein-Beltrami Model from 1871. The Poincaré Disk maps the a point on a hyperbolic plane, given in polar coordinates (r, θ) to the open unit disk $\mathcal{D}$ [11] using the following mapping functions:

$$f_x(r, \theta) = \tanh(\frac{r}{2}) \cos(\theta)$$
$$f_y(r, \theta) = \tanh(\frac{r}{2}) \sin(\theta)$$

The distance (also called the line segment) of two points (r, θ) and $(r + \triangle r, \theta + \triangle\theta)$ is computed by

$$ds^2 = 4\frac{dr^2 + r^2 d\theta^2}{(1 - r^2)^2}. \tag{1}$$

We chose the Poincaré model because of the following properties:

i. It maps the infinite large area of the hyperbolic plane $\mathbb{H}^2$ entirely onto the unit euclidean disc.
ii. The mapping is conformal, i. e. angles are preserved.
iii. While the angles (i. e. shape features) are preserved, the projection non-isometric and exhibits a strong fish-eye effect. The origin of the $\mathbb{H}^2$ is represented almost faithfully but with growing distance from the center, the data display gets more and more squeezed due to the tanh function.

To be able to access all details at the squeezed borders we need to implement a procedure to manipulate the Poincaré projection to allow focussing on selected areas of the $\mathbb{H}^2$.

Möbius transform. The Möbius transform is bijective and homomorphic and it describes the group of isometries of the Poincaré disk $\mathcal{D}$ [1]. The transform maps a point z in the disc $\mathcal{D}$ to new cordinates $M(z)$ using the following formula:

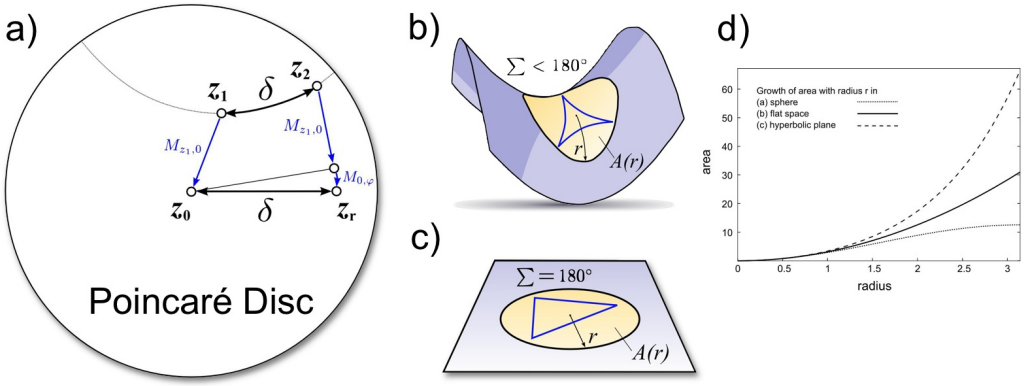

FIGURE 2. Geometric properties of the hyperbolic space. Please see the text for details.

$$M_{c,\phi}(z) = e^{i\varphi}\frac{z-c}{1-\bar{c}z},\ c \in \mathbb{C},\ |c| < 1. \tag{2}$$

For $c = 0$ the transform $M_{0,\phi}(z)$ just describes a rotation of $\mathbb{D}$ with angle ϕ and for $\phi = 0$ a translation is achieved, mapping c to $-c$.

Properties of the hyperbolic space. Since we choose the Poincaré disc for representing structures from the $\mathbb{H}^2$ in a euclidean plane, we need methods to compute distances of points in $\mathcal{D}$ in $\mathbb{H}^2$ as well as areas in relation to this projection as displayed in figure 2. The hyperbolic distance $\delta(z_1, z_2)$ of two points z_1 and z_2 is given by integrating the line segment (eq. (1)) along the circular path displayed in Figure 2 a). In practice, two Möbius transforms ($M_{z_1,0}$ (translation) and $M_{0,\varphi}$ (rotation)) are applied to move z_1 to the origin z_0 and z_2 to z_r:

$$M_{0,\varphi}(M_{z_1,0(z_2)}) = e^{i\varphi}\frac{z_2 - z_1}{1-\bar{z}_1 z_2} = z_r$$

Since $z_r = r + 0i$ is of real value it follows

$$r = \left|\frac{z_2 - z_1}{1-\bar{z}_1 z_2}\right|$$

and integration of the line segment gives the following distance function:

$$\delta(z_1, z_2) = \int_0^r \frac{2}{1-x^2}dx = 2\arctan(r) = 2\arctan\left(\left|\frac{z_2 - z_1}{1-\bar{z}_1 z_2}\right|\right) \tag{3}$$

The area $A(\rho)$ of a hyperbolic circle with radius ρ (see yellow patch in Figure 2 c) is computed by

$$A(\rho) = \int_0^r \int_0^{2\pi} \frac{4r}{(1-r^2)^2} dr d\theta = 4\pi \frac{r^2}{1-r^2} = 4\pi \sinh^2\left(\frac{\rho}{2}\right). \tag{4}$$

From eq. 4 one can see that the circular area in hyperbolic space shows to have a feature which is of particular relevance to the self-organizing map (SOM) algorithm. For a small radius ($r < 1$) the space is almost flat and $A(r) \approx \pi r^2$. But for larger r the area grows asymptotically exponential in contrast to the quadratic growth in the flat euclidean plane (see Figure 2 c). For comparison, we display the growth rates for increasing r for a sphere, a flat space and a hyperbolic plane in Figure 2 d. So the hyperbolic plane offers more space for a SOM in a low dimensional embedding.

The Self-Organizing Map in hyperbolic space. The classic SOM consists of a a set of formal neurons $(\mathbf{u}^{(k)}, z_k)_{k=1\ldots K}$ which are locally arranged on a regular lattice (like a kartesian grind or a hexagonal grid) with z_k representing the lattice coordinates of the kth neuron and $\mathbf{u}^{(k)}$ as the corresponding prototype vector. In the hyperbolic plane the lattice to lay out the neurons $(\mathbf{u}^{(k)}, z_k)_{k=1\ldots K}$ is constructed in the following way. First, the center node is placed (see blue node in Figure 3) in the center of $\mathcal{D}$ and a set of b nodes is placed around it. These nodes are placed as corners of $b-1$ triangles with angle $\alpha = 360/(b-1)$ and side length

$$l = \tanh\left(\frac{1}{2}\text{arccosh}\left(\frac{\cos(\alpha)}{1-\cos(\alpha)}\right)\right)$$

which is practically achieved by setting the first node (green node in Figure 3a) and applying the Möbius transform $M_{0,\varphi}$ (with $\varphi = \cos(\alpha) + i\sin(\alpha)$) $(b-1)$ times so a ring of b nodes is created. In the next step, another ring of nodes is created by expanding each perimeter node with $b-3$ nodes, again using the Möbius transform (see Figure 3 b). This step can be repeated for each new perimeter ring of nodes until the desired number of nodes is initiated. The final number of nodes depends on the number of neighbors b and the number of rings R and grows exponentially with R. In Figure 3c) we show a grid for $b = 7$ and $R = 2$. In 3d) we show a large HSOM with $b = 7$ and $R = 6$ to display the strong degree of squeezing the nodes in the outer perimeter at the border of the disc. Using the Möbius transform, one node can be selected as focus (red node in Figure 3d)-f) and moved to the center so the area around this node can now be inspected with a high level of detail (see Figure 3 f) while the rest of the HSOM grid is squeezed on the left side of the disc.

The HSOM is trained basically in the same way as a "classic" SOM, but with a new neighborhood function $h(k, k^\star)$ (for two neurons k and $k^\star$) which results from exchanging the Euclidean distance in the standard Gaussian neighborhood function with the hyperbolic distance shown above (see eq. 3):

$$h(k, k^\star) = \exp\left(-\frac{\arctan\left(\left|\frac{z_k - z_{k^*}}{1-\bar{z}_k z_{k^*}}\right|\right)}{\sigma^2(t)}\right). \tag{5}$$

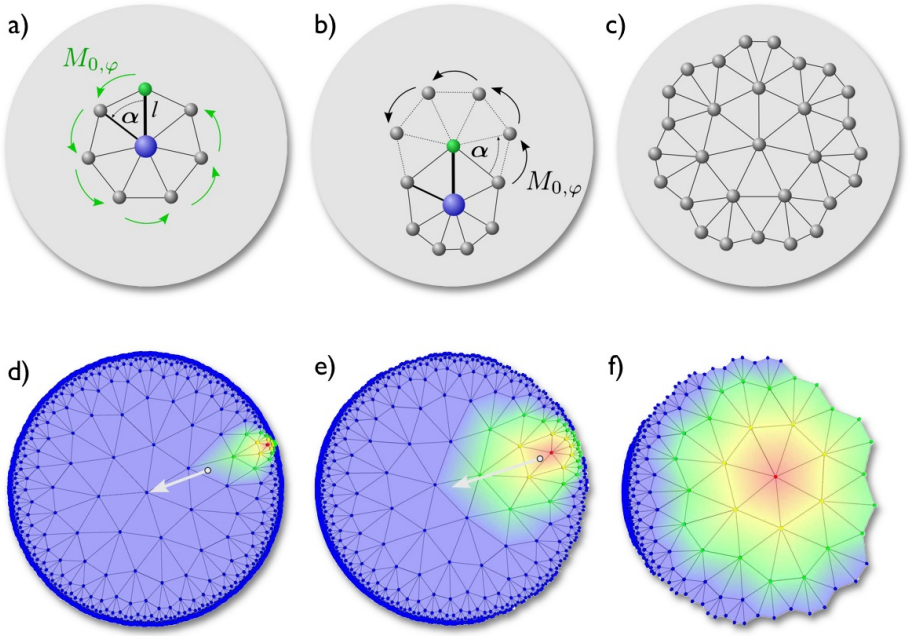

FIGURE 3. The upper row shows the construction of a HSOM grid with $K' = 7$ and $R = 2$. The lower row visualizes the application of the Möbius transform to move one node in the outer periphery (red dot in d) to the center of $\mathcal{D}$ so the nodes in its vicinity can be inspected in high detail as shown in f).

The hierarchical HSOM: H^{2}SOM. The grid construction concept of the HSOM results in a huge number nodes due to the exponential growth rate. While on the one hand, a large number of nodes has the benefit of a more trustworthy embedding due to a low reconstruction error, on the other hand the training and mapping time increases painfully because of a time-consuming search for the best matching unit (BMU) in each learning step. Thus, we apply a special version of the HSOM, the hierarchically learning HSOM (H^{2}SOM) which uses a beam search strategy when searching for the best match unit. This means, that for each training step the BMU is searched iteratively starting with the inner 1st ring. When the BMU in the 1st ring is found, only the successors (i. e. child) in the 2nd ring are analyzed regarding the BMU criterion, and so forth. In [8] it was shown, that these H^{2}SOM have the same potential to perform complex unsupervised learning and projection tasks compared to the original SOM but have log scale training time.

Empirical evaluation: continuity and trustworthiness. To empirically evaluate the quality of the H^{2}SOM learning result we apply the methods proposed in ([12]). A good H^2SOM quality is necessary if the result, i. e. the assignment of MCEPs to clusters and the projection to the low dimensional grid, is to be used for a pseudocoloring. If the learning

result is of low quality, different MCEPs are mapped to similar or even identical colors or vice versa. To assess the quality of a H^{2}SOM learning after a number of n training steps, Venna and Kaski proposed to compute to values which give an empirical estimation for the absences of two kinds of errors, the trustworthiness

$$T(k) = 1 - \frac{2}{Nk(2N - 3k - 1)} \sum_{i=1}^{N} \sum_{x_j \in U_k(x_i)} (r^{(k)}(x_i, x_j) - k)$$

and the continuity $C(k)$

$$C(k) = 1 - \frac{2}{Nk(2N - 3k - 1)} \sum_{i=1}^{N} \sum_{x_j \in V_k(x_i)} (\hat{r}^{(k)}(x_i, x_j) - k), \text{with}$$

- $x_i \in \mathbb{R}^N, i = 1, N$
- $C_k(x_i)$ the set of those k data vectors that are closest to x_i in the original data space
- $\hat{C}_k(x_i)$ the set of those k data vectors that are closest to x_i after projection
- $U_k := x_j : x_j \in \hat{C}_k(x_i) \wedge x_j \notin C_k(x_i)$
- $V_k := x_j : x_j \notin \hat{C}_k(x_i) \wedge x_j \in C_k(x_i)$
- $r^{(k)}(x_i, x_j), i \neq j$ the rank of x_j when the data vectors are ordered based on their distance from the data vector x_i in original data space
- $\hat{r}^{(k)}(x_i, x_j), i \neq j$ the rank of x_j when the data vectors are ordered based on their distance from the data vector x_i after projection

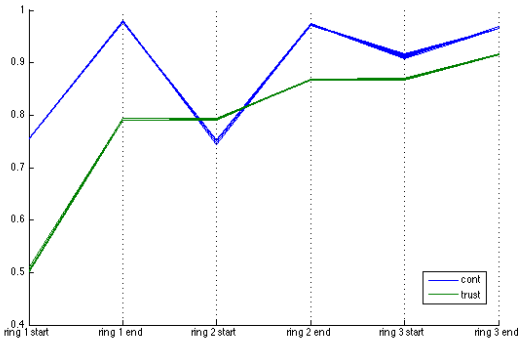

FIGURE 4. To estimate the quality of the H^{2}SOM learning, the trustworthiness $T(k)$ and the continuity $C(k)$ were computed during the training. At each time point, 10 supsamplings of the training data were used to compute $T(k)$ and $C(k)$ and the reulst were plotted. One can see, that for different samplings the values show a very low variance and the values of $T(k)$ and $C(k)$ increase for growing number of training steps.

Frankly speaking, the trustworthiness gives an estimate for how many of the neighbors in the projection are also neighbors in the original space. The continuity gives an estimate about how many of the neighbors in the original space are kept after the projection.

Because of computational complexity of these measures we subsampled the data set $1:400$ (training was still done with all samples) for computing $T(k)$ and $C(k)$. In each ring training period, $T(k)$ and $C(k)$ were computed two times, at the beginning and one at the end of each training period. This procedure was redone ten times. The entire procedure was repeated ten times for different saplings and the results are plotted shown in Fig. 4.

C. TIS image analysis with PCA and k-means

One way to visualize a TIS image could be of course to compute a principal component analysis (PCA) [9, 4, 5] and select the eigenvectors of the largest eigenvalues as a new low-dimensional coordinate system. Each pixel is mapped to this new low-dimensional space by projecting its co-location vector $\mathbf{g}_{x,y} = (g_1, g_2, ..., g_N)_{x,y}$ on the eigenvectors. This approach has two drawbacks. First, the PCA is limited to resolve linear features of the N-dimensional point cloud of the image. Second, if the spectrum of eigenvalues shows no strong decrease for the highest eigenvalues, i. .e. an elbow shape, a large number of eigenvectors is needed to account for a considerable percentage of the data variance. Third, subsets of variables (proteins) may be discarded by the PCA so that not the whole combinatorial power can be analyzed.

Another way to analyze TIS images is to first identify the MCEPs and display them with random colors like in the CMP approach. To compare the WHIDE tool with this established approach we applied a k-means algorithm to the same data set as the H^{2}SOM. The Lloyd k-means [6] was applied and the number of clusters was set to $k = 161$, so the same number of prototypes was used as in the H^{2}SOM approach. After training each cluster was mapped to a random color and each pixel belonging to this cluster was displayed

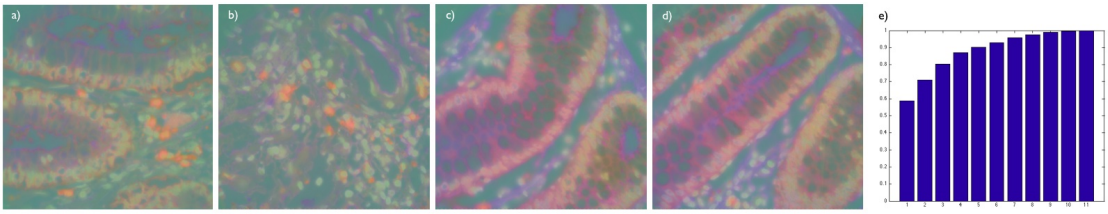

FIGURE 5. A PCA was performed for the entire set of fluorescence co-location feature vectors $\Gamma_\cup = \{\mathbf{x}^{(\xi)}\}$. Afterwards, each feature vector was projected with three eigenvectors corresponding to the three largest eigenvalues to new coordinates $\mathbf{v}^{(\xi)} = (v_1, v_2, v_3)^{(\xi)}$ and these were scaled to $[0;1]$ and used as color coordinates in a RGB space to render pseudocolor maps for each TIS image (see a)-d)). In e) we show the accumulation plot for the sum of eigenvalues $f(i) = \sum_{j=0}^{i} \lambda_j / \sum_{j=0}^{N} \lambda_j$, with $\lambda_i < \lambda_j, \forall i, j$ with $j < i$

with this color as shown in Figure 6 in this supplementary. Examining the k-means results for the four TIS images one can see, that with this coloring approach the tissue morphology is much more difficult to resolve. Second, it is very difficult (or impossible) to say, if two colors of two distant regions are just similar or identical. In the WHIDE approach, this not a serious problem since even if the colors are not identical, their MCEPs will be very similar due to the neighborhood preserving projection of the H^2SOM. Third, the nova-like artifacts have a much bigger influence on the clustering and on the visualization. Fourth, the PCA and the k-means approach do not allow a dynamic manipulation of the color mapping, which is demonstrated for WHIDE in the example shown in Figure 3 in the main manuscript. Here, a few interesting cells have been identified in the initial projection. In a next step, the color contrast has been enhanced for these cells to improve the color contrast of these cells' MCEPs. While these cells can be detected in the PCA projection, their contrast cannot be enhanced in this visualization. In the k-means visualizations, it is hard (or maybe impossible) to identify these cells in the projection shown. In summary, the WHIDE approach enables a more comprehensive and more flexible analysis of the TIS data, compared to the two other approaches PCA and k-means in combination with random colors.

D. Additional application examples

To illustrate the advantages of WHIDE and BioIMAX regarding flexibility, we applied WHIDE to two additional data sets and the results can be selected and explored in the "WHIDE Demo" project. In the following we give a short description of the data sets and background.

- MaldiImage_Barley:
  The data set was recored in a study focused on the development of MALDI-Imaging (MI) techniques [3] for plant tissues under stress. Target compounds in the centre of our research will be small molecules (metabolites), especially compounds of the phenylpropanoid pathway.To study morphological features of barley together

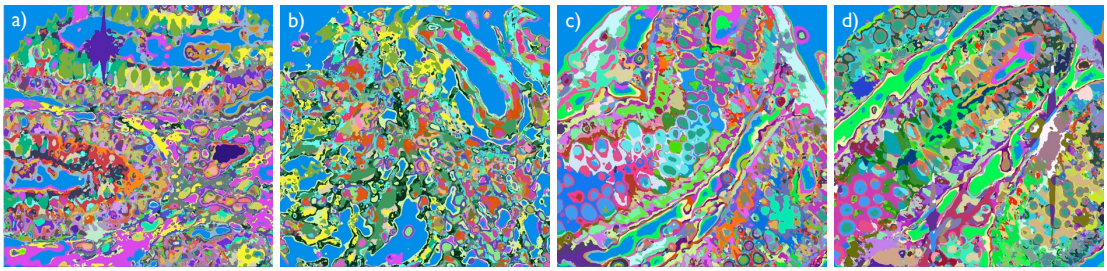

Figure 6. A k-means was applied to the feature vectors of the four TIS images from our study. To demonstrate the benefits of WHIDE's non-random, topology preserving color mapping, the $k = 161$ cluster prototypes were mapped to random colors for the k-means result.

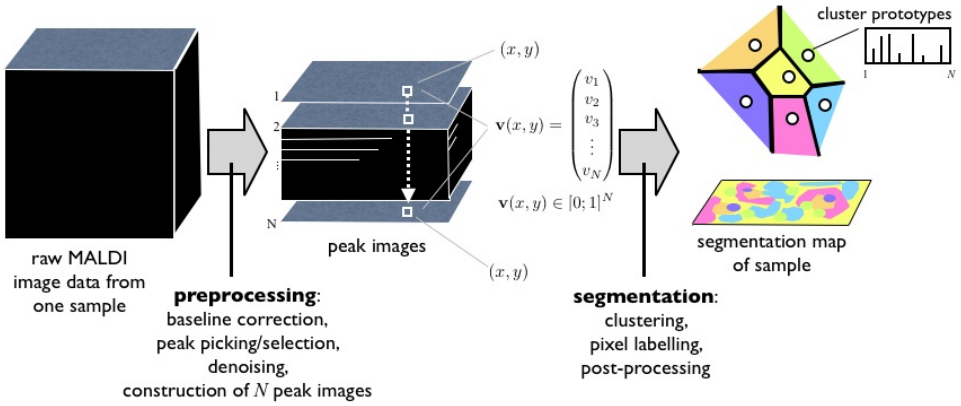

FIGURE 7. The MALDI image analysis pipeline: the original data set represents an MBI of high dimension ($N = 26$)

with abundances of peptide residues we applied MALDI imaging and extracted a $N = 26$-dimensional MBI from the original data set following a common processing pipeline shown in Figure 7.

- HCS_Listeria_Infection:
  To study bacterial invasion processes in cell cultures, we proposed a semi-automized high content screening platform in [2]. We analysed the process of L. monocytogenes infection in Mouse macrophages RAW 264.7 cell cultures stained with WCSR (Whole Cell Stain Red, Thermo Scientific) to visualize cytoplasm, and Hoechst 33342 to visualize cell nuclei. Bacteria were visualized with GFP as described in [2]. The image acquisition was done with the Scan^R screening station (Olympus). Images were taken with a 40×objective (N.A. 0.9) and standard filter sets for bisbenzimide Hoechst 33342 (ex. 360370 nm, em. 420460 nm), GFP (ex. 451490 nm, em. 500530 nm) and WCS Red (ex. 590650 nm, em. $>$ 650 nm). The software autofocus was used in the bisbenzimide channel. From the three stains, an RGB image was composed ($N = 3$) and analyzed using BioIMAX tools [7].

Of course, especially the MALDI image application can be discussed regarding preprocessing or improved regarding the CIPRA visualization. However, using WHIDE it is instantaneously possible to perceive a morphological structure in the detected clusters of peak patterns.

E. How to log in to BioIMAX and start WHIDE

WHIDE start up steps as shown in Figure 8:

(1) Start up an internet browser and direct it to `http://ani.cebitec.uni-bielefeld.de/BioIMAX` enter Login: *whidetestuser*, Password: *whidetest* and confirm.
(2) You are presented with the welcome screen of BioIMAX. Select the WHIDE Demo project in the *Projects* list on the left (a) and start up the *Project Browser* (b).

FIGURE 8. A series of BioIMAX and WHIDE screenshots explaining how to start WHIDE.

(3) Check *Filter data by: Results* (a), select any result (b). Notice how WHIDE appears in the *Applications* list after you selected a result. Start WHIDE for your selection by clicking on (c).
(4) WHIDE starts in a new window. In case your result consists of mappings for more than one TIS image, a selection menu (a) appears to choose any subset of the available TIS images. After confirmation (b) WHIDE initalizes for a few seconds before you can start exploring.

REFERENCES

[1] J. Anderson. *Hyperbolic Geometry*. Springer Verlag, New York, 2001.

[2] M. Arif, N. Rajpoot, T. Nattkemper, U. Technow, T. Chakraborty, N. Fisch, N. Jensen, and K. Niehaus. Quantification of cell infection caused by listeria monocytogenes invasion. *J Biotechnol.*, 2011. doi:10.1016/j.jbiotec.2011.03.008.

[3] K. Chunghtai and R. Heeren. Mass spectrometric imaging for biomedical tissue analysis. *Chem. Rev.*, 110(5):3237–77, 2010.

[4] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441, 1933.

[5] K. Karhunen. Über Lineare Methoden in der Wahrscheinlichkeitsrechnungna. *Annales Academiae Scientiarum Fennicae*, 1:102–344, 1946.

[6] S. Lloyd. Least squares quantization in pcm. *Information Theory, IEEE Transactions on*, 28(2):129 – 137, 1982.

[7] C. Loyek, D. Langenkmper, J. Kölling, K. Niehaus, and N. T.W. A web 2.0 strategy for the collaborative analysis of complex bioimages. In *IDA (Biennial Symposium on Intelligent data Analysis)*, Porto, Portugal, 2011.

[8] J. Ontrup and H. Ritter. Large-scale data exploration with the hierarchically growing hyperbolic SOM. *Neural Networks*, 19:751–761, 2006.

[9] K. Pearson. On lines and planes of closest fit to a system of point in space. *Philosophical Magazin*, 2:559–572, 1901.

[10] H. Poincaré. Sur les applications de la gométrie non euclidienne la thorie des formes quadratiques. *Compte Rendu de lassociation Francaise pour lAvancement des Sciences*, 10:132138, 1881.

[11] A. Ramsay and R. Richtmyer. *Introduction to Hyperbolic Geometry*. Springer Verlag, New York, 1995.

[12] J. Venna and S. Kaski. Neighborhood preservation in nonlinear projection methods: An experimental study. In *Artificial Neural Networks-ICANN*, pages 485–491, 2001.