

Synplex: In Silico Modeling of the Tumor Microenvironment From Multiplex Images

Daniel Jiménez-Sánchez, *Fellow, IEEE*, Mikel Ariz^{1b}, *Fellow, IEEE*, Carlos E. De Andrea, and Carlos Ortiz-De-Solórzano^{1b}, *Senior Member, IEEE*

Abstract—Multiplex immunofluorescence is a novel, high-content imaging technique that allows simultaneous *in situ* labeling of multiple tissue antigens. This technique is of growing relevance in the study of the tumor microenvironment, and the discovery of biomarkers of disease progression or response to immune-based therapies. Given the number of markers and the potential complexity of the spatial interactions involved, the analysis of these images requires the use of machine learning tools that rely for their training on the availability of large image datasets, extremely laborious to annotate. We present Synplex, a computer simulator of multiplexed immunofluorescence images from user-defined parameters: i. cell phenotypes, defined by the level of expression of markers and morphological parameters; ii. cellular neighborhoods based on the spatial association of cell phenotypes; and iii. interactions between cellular neighborhoods. We validate Synplex by generating synthetic tissues that accurately simulate real cancer cohorts with underlying differences in the composition of their tumor microenvironment and show proof-of-principle examples of how Synplex could be used for data augmentation when training machine learning models, and for the *in silico* selection of clinically relevant biomarkers. Synplex is publicly available at <https://github.com/djimenezsanchez/Synplex>.

Index Terms—Tumor microenvironment, multiplex imaging, simulation, modeling, spatial interactions.

I. INTRODUCTION

SOLID tumors can be understood as abnormal organs with a microenvironment populated by transformed cancer cells and normal immune and stromal cells [1]. Understanding how the composition of this complex microenvironment relates

Manuscript received 4 April 2023; accepted 1 May 2023. Date of publication 8 May 2023; date of current version 2 October 2023. This work was supported by Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación (MCIU/AEI/10.13039/501100011033) and FEDER funds UE under Grant RTI2018-094494-B-C22, Grant RTC-2017-6218-1, Grant PDI2021-122409OB-C22 and Grant TED2021-131300B-I00. (*Corresponding author: Carlos Ortiz-De-Solórzano.*)

Daniel Jiménez-Sánchez and Carlos Ortiz-De-Solórzano are with the Biomedical Engineering Program and Ciberonc, Center for Applied Medical Research, University of Navarra, 31009 Pamplona, Spain (e-mail: danijimnzs@gmail.com; codesolorzano@unav.es).

Mikel Ariz is with the Biomedical Engineering Program and Ciberonc, Center for Applied Medical Research, University of Navarra, 31009 Pamplona, Spain, and also with the Department of Electrical, Electronic and Communications Engineering, Public University of Navarra, 31006 Pamplona, Spain (e-mail: mikelariz@unav.es).

Carlos E. De Andrea is with the Department of Pathology, Clínica Universidad de Navarra, 31008 Pamplona, Spain (e-mail: ceandrea@unav.es).

Digital Object Identifier 10.1109/TMI.2023.3273950

to tumor status or progression is the key to personalized anticancer therapies [2]. This requires answering questions such as: which cell phenotypes populate the tumor or the stromal compartment? How do these phenotypes spatially interact? Is this interaction related to the patient's response to therapy?

Novel multiplexed imaging (MI) technologies allow simultaneous visualization of a high number of protein markers in tissue sections. By using signature panels of tumor and immune biomarkers [3], [4], MI can for the first time reveal *in situ*, the cellular composition of the tumor microenvironment [5]. The quantification of these cellular phenotypes and their interactions, what is commonly known as the *spatial phenotyping* of the tumor, exceeds the capabilities of the human brain. Therefore, automated machine learning (ML)-based image analysis algorithms are being developed with the aim of learning the spatial tumor phenotype and linking it to the patients' diagnosis or prognosis.

Supervised ML methods can be used to study the role of specific tumor microenvironment elements (TMEs) using conventional histopathology (H&E) or immunostained tissue sections. This requires expert manual annotations of those TMEs, which are fed to a model that automatically learns their underlying tissue patterns [6]. Unsupervised methods, in turn, can be used to discover clinically relevant TMEs [7]. To this end, topological networks are created using the information extracted from the tissues (e.g., the intensity of the biomarkers and morphological features) [8], [9]. ML methods must be trained and/or validated using annotated datasets. Indeed, annotated H&E and conventional immunohistochemistry sections have been used for cell segmentation [10], cell proliferation analysis [11], tumor segmentation [12], or tissue classification and grading [13]. In MI, however, due to the extreme complexity of the manual annotation of single cells and cell-to-cell interactions between numerous cell phenotypes, properly curated MI datasets are still lacking.

An alternative to the use of annotated datasets is the use of synthetic datasets generated from patient data [14], [15]. Filogen [16], a simulator of 3D confocal time-lapse sequences of migrating cells, has been used to train cell tracking algorithms [17]. Generative adversarial networks (GANs) have also been trained with annotated real fluorescence microscopy data to create realistic 3D stacks of fluorescent cells [18]. Moreover, *in silico* computational models of the tumor microenvironment [19] have been used for virtual drug screening [20],

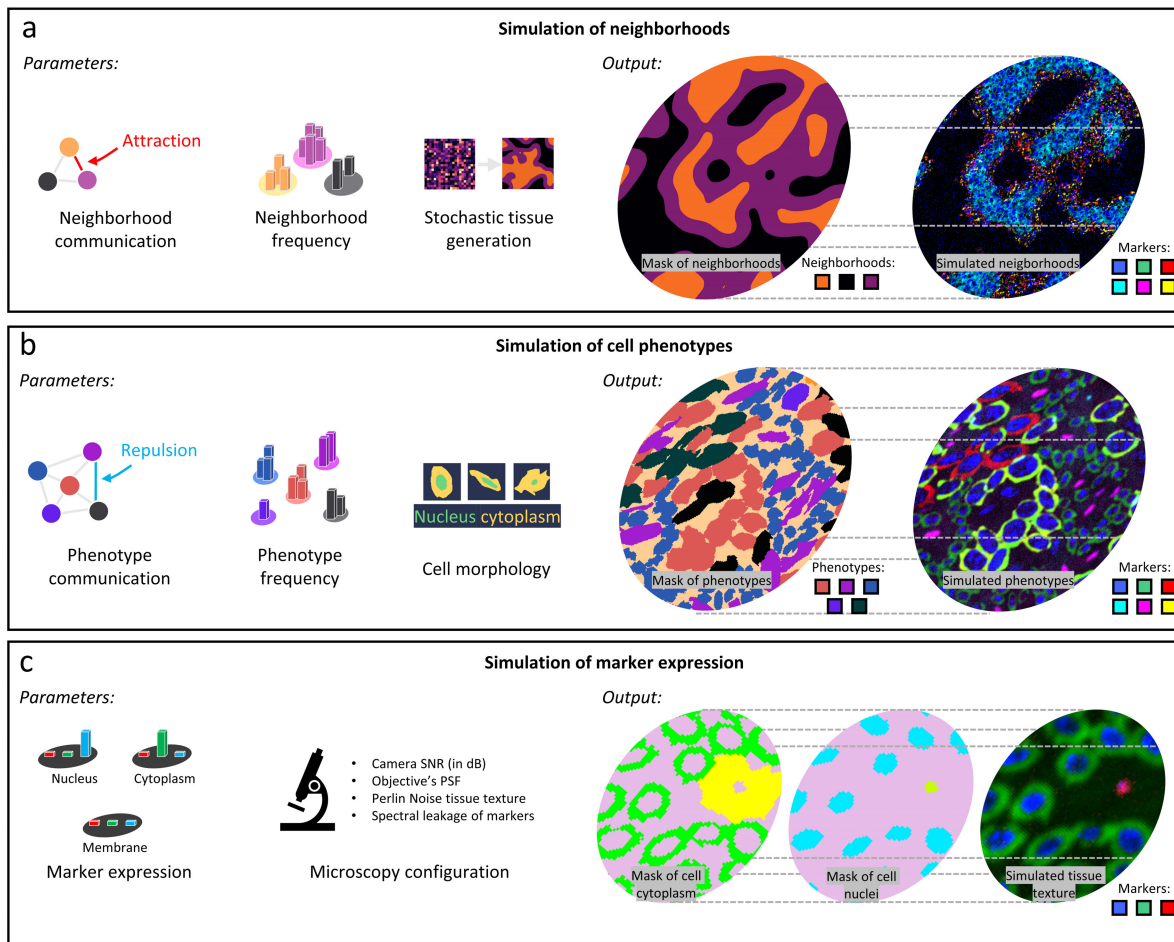


Fig. 1. Overview of Synplex. **a.** The Neighborhood simulation module takes user-defined neighborhood interactions and neighborhood frequencies to generate a mask of stochastically simulated neighborhoods. **b.** The Cell Phenotype simulation module takes user-defined phenotype interactions, phenotype frequencies, and phenotype morphology to generate a simulated mask of phenotypes. **c.** The Marker Expression simulation module takes user-defined marker expressions localized in the nucleus and/or cytoplasm, and the configuration of the microscope used in the simulation, to generate a realistic tissue texture.

drug repurposing [21], or artificial gene synthesis [22]. These models integrate known cellular phenomena [23] to guide biologists in their scientific discovery [24]. PhysiBoSS [25] for instance, simulates the population-wide effect of environmental and genetic alterations of individual cells, helping with the design of *in vitro* and *in vivo* experiments [26].

Here we present Synplex [27], the first synthetic tissue simulator of MI-stained tumors. Essentially different from the above-mentioned models that simulate isolated cells, cell cultures, or simple inter-related tissue societies made of one or few cell phenotypes, Synplex realistically simulates the *in situ* tumor microenvironment at three increasing levels of spatial complexity: i. cell phenotypes defined by the expression of markers and cell morphology; ii. cellular neighborhoods defined by their cell phenotypes and their interactions; and iii. areas described by the spatial interaction between cellular neighborhoods. Synplex is fed with user-defined parameters that describe the properties of the tumor microenvironment. From these parameters, Synplex produces sets of MI images and their corresponding ground truths, i.e., masks that identify regions with specific cell phenotypes and neighborhoods. We show that this process produces realistic, statistical

variability between images that can simulate a disease, and present two proof of principle experiments of the potential use of Synplex for data augmentation and *in silico* selection of biomarkers.

II. METHODS

Synplex consists of three independent modules (Fig. 1). The first module creates a tissue mask with a predefined presence of cellular neighborhoods and fixed rules of interaction between neighborhoods. The second module populates each neighborhood with different cell phenotypes -predefined by the expression of specific markers and morphological features- and phenotype interactions. The third module simulates the image acquisition process of the tissue through a properly parametrized optical system.

A. Modeling of Cellular Neighborhoods (Fig. 1a)

To create a tissue neighborhood mask, Synplex takes four input parameters: *i*) the number of cellular neighborhoods N ; *ii*) the relative abundance of each neighborhood in the tissue ($N_{ab} \in \mathbb{R}^N$); *iii*) the pairwise interactions between neighborhoods ($N_{int} \in \mathbb{R}^{N \times N}$) with values ranging from 1 (attraction)

to 0 (repulsion); and *iv*) the radius $N_{Ctx} \in \mathbb{Z}$, (pixels) of the neighboring context of a pixel.

Synplex randomly initializes an image neighborhood mask $M_{Nb} \in [0, 1, \dots, N]^{m_x \times m_y}$, being m_x and m_y the image dimensions. It also initializes with 1's (unassigned) an indexing image $R_{Nb} \in [0, 1]^{m_x \times m_y}$ that keeps track of the pixels that have been permanently assigned to a neighborhood. The mask values of M_{Nb} are updated iteratively by the constrained minimization of ℓ_N :

$$\ell_N = \text{sum} \left(\text{abs} \left(N_{int} \odot \left(\frac{N_{ab}}{M_{Nb_{ab}}} \right)^2 \right) \right) \quad (1)$$

where \odot is an element-wise multiplication. In each iteration, Synplex optimizes a $N_{Ctx} \times N_{Ctx}$ region surrounding a randomly selected pixel $s_{i,j}$. To this end, a submatrix $M_c \in \mathbb{R}^{N_{Ctx} \times N_{Ctx}}$ is created with the values of M_{Nb} contained within the target region, whose values are updated in M_{Nb} as follows:

- i. If more than 95% of the pixels of M_c belong to a given neighborhood type, the simulator permanently assigns $s_{i,j}$ to that neighborhood type and sets $R_{Nb}(i, j)$ to zero.
- ii. Otherwise, the simulator updates each value of M_c using the following update rule: $U_{Nb} = N_{int} \odot (N_{ab} / M_{Nb_{ab}})^2$, which minimizes ℓ_N .

This iterative process continues until all pixels of M_{Nb} have been permanently assigned to a neighborhood, i.e. $\sum R_{Nb} = 0$. This process is summarized in Algorithm 1.

Algorithm 1 Pseudocode for Modeling Cellular Neighborhoods. **Inputs:** N_{ab} , Neighborhood Abundance Matrix. N_{int} , Neighborhood to Neighborhood Interaction Matrix, $\{m_x, m_y\}$ Tissue Image Size in x and y-Axis, Respectively. N , Number of Neighborhoods. **Output:** M_{Nb} , Cellular Neighborhood Mask

```

1: Random neighborhood mask:  $M_{Nb} = \text{randomMatrix}(m_x, m_y, N)$ 
2: Initialize vector of set indices:  $R_{Nb} = \text{matrix\_of\_ones}(m_x, m_y)$ 
3: while  $\sum R_{Nb} > 0$ 
4:   Update optimization rule:  $U_{Nb} = N_{int} \odot \left( \frac{N_{ab}}{M_{Nb_{ab}}} \right)^2$ 
5:   Select one random pixel  $s \in R_{Nb} = 1$ 
6:   Get context of  $s$ :  $M_c = \{M_{Nb} | M_{Nb} : \text{dist}(M_{Nb}, s) < N_{Ctx}/2 \text{ px.}\}$ 
7:   if  $\max(\text{histcounts}(M_{cN})) > 95\%$  then
8:     Most abundant neighb:  $M_{Nb} = \text{argmax}(\text{hist}(M_{cN})) : M_N \cap M_{cN}$ 
9:     Reduce pool of pixels to iterate:  $R_{Nb} = 0 : R \cap M_{cN}$ 
10:  else
11:    Update set of neighborhoods:  $M_{Nb} = U_{Nb}(M_{cN}) : M_{Nb} \cap M_{cN}$ 
12:  end while

```

B. Modeling of Cellular Phenotypes (Fig. 1b)

To populate with cells the N neighborhoods created by the previous module Synplex requires eight parameters: *i*) the number of cell phenotypes P ; *ii*) a per-neighborhood cell phenotype abundance matrix, $P_{ab} \in \mathbb{R}^{P \times N}$; *iii*) a matrix $P_{int} \in \mathbb{R}^{P \times P \times N}$ containing the interactions between the cell phenotypes within each neighborhood with values ranging from 1 (attraction) to 0 (repulsion); then for each cell phenotype: *iv*) the average eccentricity of the cells $P_{ecc} \in \mathbb{R}^P$, ranging from 0 (round cells) to 1 (maximally elongated cells); *v*) the average diameter (in pixels) of the cells $P_{siz} \in \mathbb{R}^P$;

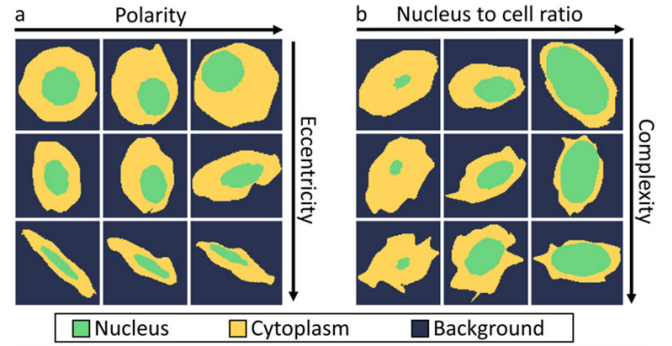


Fig. 2. Modeling of cell morphologies. a, b. Examples of modeled cells with increasing levels of (a) polarity and eccentricity and (b) nucleus to cell ratio and cell complexity.

vi) the average ratio between the size of the cell nucleus and the total cell size $P_{R_NucCell} \in \mathbb{R}^P$, ranging from 0 (absence of nucleus) to 1 (absence of cell cytoplasm); *vii*) the average cell morphological complexity $P_{cmplx} \in \mathbb{R}^P$, ranging from 0 (uniform ellipsoidal shape) to 1 (highly irregular shape); and *viii*) the average cell polarity $P_{polrty} \in \mathbb{R}^P$, ranging from 0 (nucleus at the center of the cell) to 1 (nucleus touching the contour of the cell). Examples of cells synthetically generated with different morphology parameters are shown in Fig. 2.

Synplex creates and initializes five image masks: *i*) a cell phenotype mask, $M_{Pn} \in [0..P]^{m_x \times m_y}$, where each pixel is assigned to a cell phenotype $[0..P]$, randomly initialized from the user-defined values for each neighborhood in P_{ab} ; *ii*) a mask of individual cells $M_{Ph_Cells} \in \mathbb{N}^{m_x \times m_y}$, where each pixel is part of one cell, or part of the background, initialized with 0's; *iii*) a mask of individual nuclei $M_{Ph_Nuc} \in \mathbb{N}^{m_x \times m_y}$, where each pixel is part of one nucleus, cytoplasm, or background, initialized with 0's; *iv*) a matrix of updated indices $R_{Pn} \in [0, 1]^{m_x \times m_y}$, initialized with 1's, that is used to keep track of the pixels that have been permanently assigned to a cell phenotype; and *v*) a mask of cellular orientation $M_{Ornttn} \in [0..1]^{m_x \times m_y}$, where values indicate how cells should be oriented. M_{Ornttn} . In our case, stromal cells form a rim around the tumor, as previously described [28].

Cell phenotype mask values in M_{Pn}, M_{Ph_Cells} , and M_{Ph_Nuc} are updated iteratively by a constrained minimization of ℓ_P :

$$\ell_P = \text{sum} \left(\text{abs} \left(P_{int} \odot \left(\frac{P_{ab}}{M_{Pn_{ab}}} \right)^2 \right) \right) \quad (2)$$

where $M_{Pn_{ab}} \in \mathbb{R}^{P \times N}$ is the current phenotype abundance in M_{Pn} , and P_{ab} and P_{int} constrain the function loss to user-defined cell phenotype abundances and interactions. Note that \odot is an element-wise multiplier. To minimize ℓ_P , the tissue simulator optimizes a subset of pixel values in each iteration and subsequently updates $M_{Pn_{ab}}$ using $U_P = P_{int} \odot (P_{ab} / M_{Pn_{ab}})^2 \in \mathbb{R}^{P \times P \times N}$ as the update rule.

In each iteration, the optimization process randomly selects an unassigned pixel $s_{i,j}$, i.e., $R_{Pn}(i, j) = 1$, selects a region of $P_{Ctx} \times P_{Ctx}$ pixels surrounding $s_{i,j}$ in M_{Pn} and M_{Nb} and stores it in $M_{CP} \in \mathbb{R}^{P_{Ctx} \times P_{Ctx}}$ and $M_{cN} \in \mathbb{R}^{P_{Ctx} \times P_{Ctx}}$, respectively. To calculate the phenotype of $s(i, j)$ we use the

rule matrix $U_P \in \mathbb{R}^{P \times P \times N}$. First, the simulator applies a majority voting strategy on M_{CN} to determine the most frequent neighborhood (Ng) contained in the region surrounding $s(i, j)$, which is used to extract the Ng^{th} plane from U_P creating U'_P . Then, the simulator inputs the existing M_{CP} into U'_P to obtain the updated phenotype.

Once the new cell phenotype is defined, a cell M'_{CP} is created: to this end, first a cell shape mask is created consisting of an ellipse with $P_{ecc}(M_P)$, $P_{siz}(M_P)$, and orientation $M_{Orntn}(i, j)$; then the complexity term is applied by adding morphological sparse variations expanding and/or constraining the cell structure by a factor $P_{cmplx} \in \mathbb{R}^P$; finally, a nucleus M''_{CP} is generated with its size calculated using the ratio of the nucleus to the cell size $P_{R_NucCell} \in \mathbb{R}^P$, and it is placed inside the cell according to its polarity $P_{polrty} \in \mathbb{R}^P$. Each parameter is multiplied by a random variable that takes value from 0.9 to 1, to ensure the required cellular variability. The decision on whether to permanently assign or not the generated cell to the tissue is based on the distribution of values of R_{Pn} in the $P_{Cix} \times P_{Cix}$ region:

- i. If more than 95% of M'_{CP} cell pixel values are yet to be set, the simulator automatically adds the new cell to the tissue. This way, the morphology of the cell (M'_{CP}) is assigned to the mask of cells (M_{Ph_Cells}), the phenotype of the cell (M'_{CP}) to the phenotype mask (M_{Pn}), and the nucleus of the cell (M''_{CP}) to the nucleus mask (M_{Ph_Nuc}). Additionally, the matrix that stores which pixel values have already been permanently assigned is updated setting to 0's all pixels of M'_{CP} .
- ii. If less than 95% of M'_{CP} values remain to be set, i.e., there is no reasonable space to insert a new cell in the tissue, the simulator has two options: to expand the cytoplasm of existing neighboring cells, filling the empty intercellular space, or to fill this space with stromal fibers. This decision is based on the user-defined phenotype abundance of stromal fibers, i.e., $P_{ab}(stromalfibers)$, and the actual abundance of stromal fibers in the tissue, i.e., $M_{Pn_{ab}}(stromalfibers)$. If $P_{ab}(stromalfibers)$ is higher than $M_{Pn_{ab}}(stromalfibers)$, the pixels of M'_{CP} are set to the stromal phenotype. Otherwise, when $M_{Pn_{ab}}(stromalfibers) > P_{ab}(stromalfibers)$, the pixels of M'_{CP} are used to expand the cytoplasm of neighboring cells. Therefore, the predefined abundance of stromal fibers controls the level of compactness of tissues.

This iterative process, illustrated in Algorithm 2, continues until all pixel values are permanently assigned to a phenotype, i.e. $\sum R_{pn} = 0$.

C. Tissue Texture and Virtual Microscopy (Fig. 1c)

This module has three objectives: i) assign the predefined marker expression intensity to each cell compartment: nucleus, cytoplasm, and membrane; ii) create the tissue and cell texture; and iii) simulate the effect of the acquisition by a fluorescence microscope, accounting for both signal noise and spectral leakage. To this end, Synplex needs seven user-defined parameters: i) the average marker expression, $M_{ph} \in \mathbb{R}^{P \times Mk}$, where each row specifies the intensity of

Algorithm 2 Pseudocode for Modeling Cell Phenotypes.

Inputs: P_{ab} , Phenotype Abundance Matrix; P_{int} , Cell Phenotype to Cell Phenotype Interaction Matrix; P_{ecc} , Eccentricity of Cell Phenotypes; P_{siz} , Size of Cell Phenotypes; $P_{R_NucCell}$, Ratio of Nucleus Size to Cell Size of Phenotypes; P_{cmplx} , Complexity of Cell Phenotypes; P_{polrty} , Polarity of Cell Phenotypes; $\{m_x, m_y\}$, Tissue Image Size in x and y-Axis, Respectively; M_{Nb} , Cellular Neighborhood Mask; and P , Number of Phenotypes. **Output:** M_{Pn} , Cell Phenotype Mask; M_{Ph_Cells} , Individual Cell Mask; M_{Ph_Nuc} , Individual Cell Nuclei Mask

```

1: Initialize cell phenotype mask:  $M_{Pn} = \text{WeightedrandomMatrix}(m_x, m_y, P_{ab})$ 
2: Initialize individual cell mask:  $M_{Ph\_Cells} = \text{matrixOfZeros}(m_x, m_y)$ 
3: Initialize individual nucleus mask:  $M_{Ph\_Nuc} = \text{matrixOfZeros}(m_x, m_y)$ 
4: Initialize cell orientation mask:  $M_{Orntn} = \text{CellOrientation\_Tum}(m_x, m_y, M_{Nb})$ 
5: Initialize matrix of permanently assigned pixels:  $R = \text{matrixOfOnes}(m_x, m_y)$ 
6: Initialize index of cells:  $Cell\_Index = 0$ 
7: while  $\sum R > 0$ 
8:   Update optimization rule:  $U_P = P_{int} \odot \left( \frac{P_{ab}}{M_{Pn_{ab}}} \right)^2$ 
9:   Select random pixel  $s_{i,j} | R(i, j) = 1$ 
10:  Get pheno. of  $s$ :  $M_{CP} = M_{Pn}(i - P_{Cix}/2, i + P_{Cix}/2, [j - P_{Cix}/2, j + P_{Cix}/2])$ 
11:  Get neighb:  $Ng = \text{majorityVoting}(M_{Nb}[i - P_{Cix}/2, i + P_{Cix}/2], [j - P_{Cix}/2, j + P_{Cix}/2])$ 
12:  Get cell phenotype:  $Ph = \text{majorityVoting}(U'_P(M_C)) | U'_P = U_P(:, :, Ng)$ 
13:  Generate cell:  $M'_{CP} = \text{CellGenerator}(P_{ecc}, P_{siz}, P_{Orntn}, P_{cmplx}, Ph)$ 
14:  Generate nucleus:  $M''_{CP} = \text{NucleusGenerator}(M'_{CP}, P_{R\_NucCell}, P_{polrty}, Ph)$ 
15:  if  $\text{mean}(R([i - P_{Cix}/2, i + P_{Cix}/2], [j - P_{Cix}/2, j + P_{Cix}/2]) \cap M'_{CP}) > 95\%$  then
16:    Set new phenotype:  $M_{Pn} = Ph : M_{Pn} \cap M_{CP}$ 
17:    Set new cell:  $M_{Ph\_Cells} = Cell\_Index : M_{Pn} \cap M'_{CP}$ 
18:    Set new nucleus:  $M_{Ph\_Nuc} = Cell\_Index : M_{Pn} \cap M''_{CP}$ 
19:    Reduce pool of pixels to iterate:  $R = 0 : R \cap M'_{CP}$ 
20:    Increment cell index:  $Cell\_Index = Cell\_Index + 1$ 
21:  elseif  $\text{mean}(R \cap M'_{CP}) < 50\%$  then
22:    if  $P_{ab}(stromalfibers) > M_{Pn_{ab}}(stromalfibers)$  then
23:      Set stromal fibers:  $M_{Pn} = stromalfibers : M_{Pn} \cap M'_{CP}$ 
24:    else
25:      Increase cell cytoplasm:  $M_{Pn} = \text{majorityVoting}(M_{Pn}) : M_{Pn} \cap M'_{CP}$ 
26:  end while

```

each of the Mk expected markers for one cell phenotype; ii) the cell phenotype marker localization, $M_{loc} \in [0, 1]^{Mk \times 3}$, where each marker can be expressed in the nucleus, cytoplasm, and/or the membrane; iii) the microscope's signal to noise ratio, $M_{SNR} \in \mathbb{R}^{Mk}$, measured in decibels (dB); iv) spectral leakage, $M_{leak} \in \mathbb{R}^{Mk \times 2}$, measured as the percentage of spectral bleed-through between two contiguous markers (i.e., one with longer fluorescent emission and the other with a shorter emission); v) Perlin noise parameters to generate the tissue texture for each marker, $M_{Perlin} \in \mathbb{R}^{Mk \times 3}$, where the rows refer to markers, and columns contain the Perlin noise initial and final frequencies and persistence values [29]; vi) the Perlin noise parameters for the background, $M_{PerlinBckg} \in \mathbb{R}^3$; vii) the microscope's point spread function width (in μm), $M_{PSF} \in \mathbb{R}$.

Once the cell phenotype masks (M_{Ph_Cells} and M_{Ph_Nuc}) are generated, we extract two masks: a cell membrane mask, M_{Ph_Mem} , created by applying a Sobel edge detection operator on M_{Ph_Cells} ; and a mask of the cytoplasm, M_{Ph_Cyt} , created by applying an XOR filter between M_{Ph_Cells} and $(M_{Ph_Nuc} \cup M_{Ph_Mem})$. Then, the cell phenotype marker localization M_{loc} , and marker expression M_{ph} , are used to determine if one pixel should be stained by one or several

markers and to apply the appropriate level of marker expression. This process generates a multispectral $m_x \times m_y \times Mk$ image

To create final realistic images, we add tissue texture using Perlin noise. To this end a background/autofluorescence mask is created using the parameters predefined in $M_{PerlinBckg}$, that simulates imaging artifacts, such as uneven illumination. Then, for each marker, texture is added using the parameters defined in M_{Perlin} . Finally, to simulate the effect of a real optical system, the image is convolved with the microscope's point spread function (PSF), approximated by a Gaussian function, M_{PSF} . Finally, the camera photodetection uncertainty is applied as photon shot noise to achieve a specific SNR per marker M_{SNR} , $mein$ dB.

III. EXPERIMENTS AND RESULTS

To validate the potential of Synplex for generating artificial images containing realistic tumor microenvironments we carried out two experiments. First, we generated a synthetic dataset of 45 multiplex immunostained images from a set of specific user-defined parameters and compared the synthetic microenvironment features measured in the generated images with the predefined values. Second, we used a set of experimental parameters measured from an endometrial cancer cohort, stained with a 6-marker panel [30] to generate 300 synthetic tissue images, and compared the synthetically generated cellular populations with those quantified in the real tissues. Finally, we present two proof-of-principles examples: the use of artificial tumor simulations for ML data augmentation and for *in silico* selection of clinically relevant markers.

A. First Validation Experiment: Simulation of Tissue Samples With User-Defined Parameters

Forty-five $1500 \times 1500 \times 6$ (pixels \times pixels \times markers) MI tissue images were simulated that contained 6 cellular neighborhoods (Nb1-Nb6) with predefined abundance and neighborhood-to-neighborhood interactions. Each neighborhood contained a specific combination of 7 cell phenotypes (Ph1-Ph7) and local phenotype interactions. Each phenotype had a user-defined size, eccentricity, polarity, nucleus-to-cell ratio, and complexity, as well as predefined intensities of 6 markers (Mk1-Mk6), expressed in the nucleus, cytoplasm and/or membrane.

Finally, each marker had predefined Perlin noise, signal-to-noise ratio (SNR) and spectral leakage to simulate realistic images acquired through an optical system represented by a point spread function.

1) *Validation of Neighborhood Abundance and Interactions:* We first measured the relative abundance of each neighborhood in the synthetic images, and compared those values with the predefined abundances, obtaining a low Mean Squared Error (MSE) of 0.98%. Then, we compared the neighborhood interactions in the synthetic images with the user-defined values (Fig. 3a). To this end we measured the level of attraction between each pair of synthetic neighborhoods as the percentage of contour/boundary pixels shared by both neighborhoods. Fig. 3b shows the interactions measured between Nb2

and other neighborhoods, confirming the predefined attraction between Nb2 and Nb3. Fig. 3c likewise shows the interactions measured for Nb5, showing the expected predefined repulsion between Nb5 and Nb6. Finally, Fig. 3d shows, as a representative example, the neighborhood mask of one of the artificial images, showing the expected interactions. In summary, Synplex accurately modeled the predefined neighborhood interactions.

2) *Validation of Phenotype Abundance and Interactions:* Then we compared the relative abundance of each cell phenotype in each simulated neighborhood, with the user-defined abundances. The MSE, averaged for all phenotypes and neighborhoods was low (-1.46%). We illustrate this with the predefined phenotype abundances of two representative neighborhoods, Nb4 (Fig. 3e) and Nb6 (Fig. 3j) along with the abundances measured in the synthetic images (Fig. 3f, k). As shown, the phenotype abundances are comparable.

Next, we compared the simulated spatial interactions between phenotypes with those defined by the user. Thus, each simulated cell was considered 'connected' to another cell if the latter lay within a 30-pixel radius. Then, the interaction between each pair of phenotypes was measured as the percentage of 1-hop connected cells of both phenotypes relative to the total number of neighbors of the cell. Fig. 3g and l show the phenotype interactions predefined for Nb4 and Nb6, respectively. Fig. 3h shows the quantified phenotype interactions between Ph2 and the other phenotypes in Nb4 displaying, as expected, high attraction between Ph2 and Ph7. Similarly, Fig. 3m shows the quantified interactions between Ph5 and the other phenotypes within Nb6 where, as predefined, Ph5 displays strong repulsion to Ph6. Finally, we present two qualitative examples of Nb4 and Nb6: Fig. 3i shows evidence of attraction between Ph2 (green) and Ph7 (orange); Fig. 3n shows evidence of repulsion between Ph5 (magenta) and Ph6 (cyan). In summary, we show that Synplex can create synthetic tissues with predefined interacting phenotypes.

3) *Validation of Phenotype Marker Expression:* To validate the accuracy of the simulations in regard to the texture of the images, we measured the marker expression intensity in all the cells of all synthetic images, finding high agreement between user-defined and synthetic expression levels. We illustrate this with two representative examples (Ph1 and Ph7). For Ph1 (predefined in Fig. 3o) we obtained an average Mk1 intensity of $72.7 \pm 12.1\%$ in the simulated images (Fig. 3p). This number agrees with the predefined value, and its variability is consistent with the SNR of this marker, preset to 45dB. Furthermore, we obtained an intensity of $31.2 \pm 7.6\%$ for Mk2, which is consistent with the predefined bleed through (20%) between Mk1 and Mk2 (Fig. 3o). Examples of all relevant Ph1 marker expression effects can be visualized in Fig. 3q. Indeed, Ph1 cells express Mk1 in the nucleus, and spectral leakage from Mk1 can be seen in marker Mk2. Cell phenotype Ph7 consists of cells with nuclear expression of Mk1 and cytoplasmic expression of Mk3 and Mk4. As expected for this phenotype (predefined in Fig. 3r), the expression levels measured in the simulated images showed a low standard deviation for Mk3 ($71.1 \pm 1.5\%$) and Mk4 ($39.1 \pm 3.2\%$), consistent with a

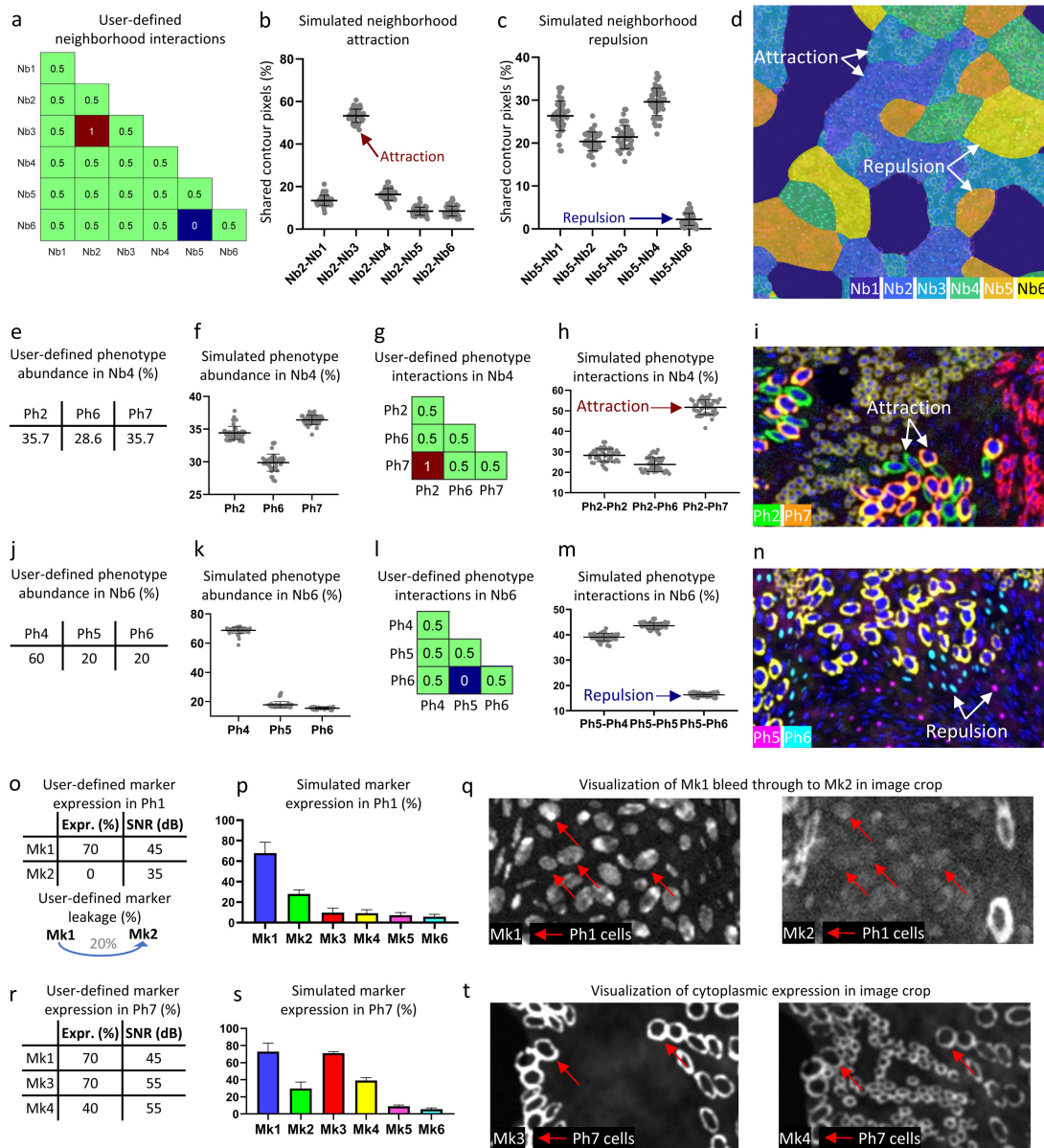


Fig. 3. Quantitative and qualitative assessment of the first validation experiment: Simulation of tissue samples from user-defined parameters. **a.** User-defined neighborhood interactions (0 represents repulsion, 0.5 represents random interaction, 1 represents attraction) between pairs of neighborhoods. **b, c.** Boxplots of the simulated neighborhood attraction and repulsion between pairs of neighborhoods. **d.** Example tissue image with an overlaid neighborhood mask visualizing neighborhood attraction and repulsion. **e, j.** User-defined phenotype abundances in neighborhoods Nb4 and Nb6. **f, k.** Simulated phenotype abundances in Nb4 and Nb6. **g, l.** User-defined phenotype interactions in Nb4 and Nb6, respectively, (0 represents repulsion, 0.5 represents random interaction, 1 represents attraction) between pairs of phenotypes. **h, m.** Boxplots of the simulated phenotype attraction and repulsion between pairs of phenotypes. **i, n.** Visual assessment of attraction and repulsion between phenotypes. **o-q.** Quantitative and qualitative assessment of marker bleed through between Mk1 and Mk2 in Ph1 cells. **r-t.** Quantitative and qualitative assessment of marker colocalization in the cytoplasm of Ph7 cells.

higher SNR of 55dB for both markers, whereas the standard deviation was higher for Mk1 ($72.9 \pm 9.8\%$), consistent with a lower predefined SNR of 45dB. As expected, Mk1 bleed through to Mk2 is clearly observed (Fig. 3s). Finally, Fig. 3t show examples of cytoplasmic expression of Mk3 and Mk4. In summary, Synplex accurately simulates cell phenotypes with customizable single-cell features.

B. Second Validation Experiment: Simulation of Tissue Samples From a Real Multiplex Image Dataset

Twelve high-grade endometrial carcinomas were stained using a six-color multiplex panel targeting key elements of

the tumor microenvironment: CD8+ T cells, the transcription factor FoxP3, the bona fide T cell activation marker CD137, the programmed cell death-1 (PD-1), cytokeratin (CK), and the nucleus (DAPI). Tissue sections were scanned using a PhenoImager HT, capturing 391 1468 × 1876 × 6 images. Clinical information of these tumors, was available [31], most relevant the classification of the tumors into two groups: ‘POLE-mutated’ with high levels of immunological activity and good response to treatment [32] (77 images from 3 patients), and ‘POLE-WT’ (wild-type) with low levels of immunological activity and poor response to treatment (314 images from 9 patients) (Fig. 4a, b).

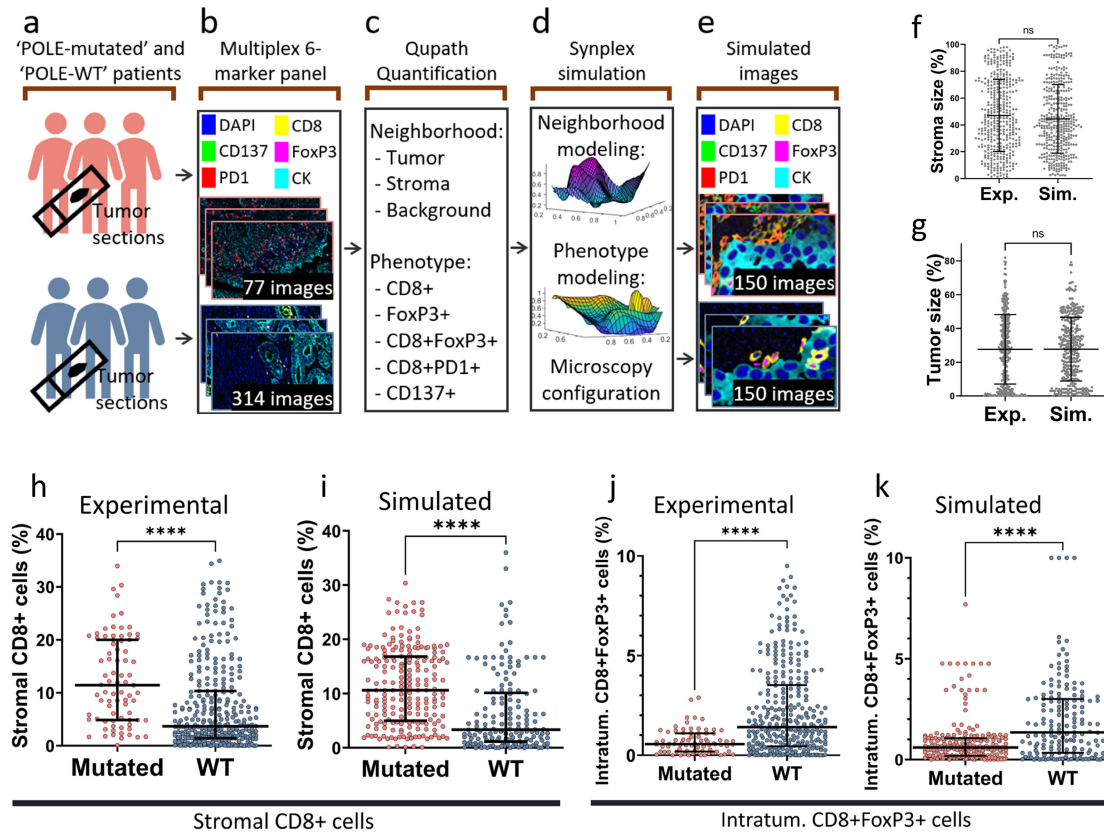


Fig. 4. Second validation experiment: Simulation of tissue samples from a real Multiplex image dataset. a-e. Simulation process from a 6-marker real adenocarcinoma cancer image dataset from two patient types (a) from which a total of 391 image fields were acquired (b). Neighborhoods and phenotypes were quantified using QuPath (c) and were used by Synplex (d) to simulate 300 synthetic images (e). f, g. Scatterplots comparing experimental and simulated neighborhood abundances (one point per image). h-k. Scatterplots comparing data type (experimental vs. simulated) and patient type (Mutated vs. WT) for stromal CD8+ (h, i) and intratumoral CD8+/FoxP3+ (j, k) populations. Statistical differences were assessed by means of a Mann-Whitney's U test with a significance value of 0.05.

To characterize the tissue microenvironment of these endometrial carcinomas, we quantified the cellular neighborhoods and cell phenotypes of the tissues using QuPath [6], [33] (Fig. 4c). Three types of neighborhoods were defined based on intensity thresholds applied to the CK and DAPI markers to distinguish *i.* a tumor neighborhood consisting of nucleated CK+ cells, *ii.* a stromal neighborhood consisting of nucleated CK- cells, and *iii.* a background neighborhood devoid of cells. Then, six relevant cell phenotypes were defined using a random tree classifier implemented in QuPath, to distinguish between CD8+, FoxP3+, CD8+FoxP3+, CD8+PD1+, CD137+ populations, as well as cells with no marker expression.

We measured the number of cells with those six phenotypes in each of the three tissue neighborhoods. These quantifications had been previously used to associate cell populations with POLE mutation status [30]. Relevant interactions between these phenotypes had been previously discovered [34] namely, attraction between CD8+ and FoxP3+ cells, CD8+FoxP3+ and FoxP3+ cells, CD8+ and tumor cells and, finally, CD137+ and tumor cells.

Using the abundance and interactions between the quantified phenotypes, the information about the optical setup, and some image quality parameters (marker expression, marker SNR, marker leakage, PSF width, Perlin Noise persistence and

frequency), Synplex created 300 ($2000 \times 2000 \times 6$) synthetic images (Fig. 5d, e), 150 simulating the microenvironment of POLE-Mutated patients, and 150 simulating the microenvironment of POLE-WT patients.

Validation of the Simulation of the Patient Cohort: We first compared the neighborhood abundances (tumor and stromal areas) between the simulated and real images (Fig. 4f, g). No statistically significant differences were found. We then compared the abundance of the simulated cell phenotypes located in the tumor and stromal compartments with the experimental values quantified with QuPath (Table I). The values are comparable, and more importantly, the differences in the abundance of the cells with the immunological phenotypes of interest follow the same trend. We illustrate this with two phenotypes that successfully discriminate between patient types: Fig. 4h, i show that CD8+ cells are statistically more abundant in POLE-mutated patients ($p < 0.001$), both in the simulated and experimental cohorts. Fig. 4j, k show that Synplex accurately simulated the significantly higher abundance of CD8+/FoxP3+ cells in WT vs. mutated patients.

Fig. 6 shows both artificial and real images from POLE-mutated patients (Fig. 6a, b), displaying a tumor microenvironment populated by a high number of PD1+ and CD8+ cells located in peritumoral areas, implying high levels of inflammation, whereas artificial and real images of

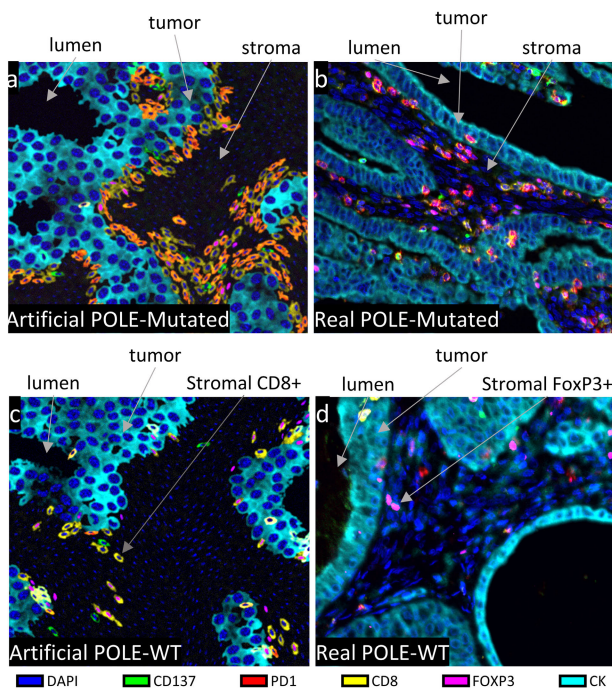


Fig. 5. Representative examples from the real and simulated endometrial carcinoma MI tissue samples. Please notice that the orange color means co-localization of PD1 (red) and CD8 (yellow).

TABLE I

COMPARISON BETWEEN REAL AND SIMULATED CELL PHENOTYPES

| Region | Phenotype | Experimental (%) | | | Simulated (%) | | |
|--------|-------------|------------------|------|------------|---------------|------|------------|
| | | Mut. | WT | p-val. | Mut. | WT | p-val. |
| Tumor | CD8 | 6.22 | 8.19 | 0.03 | 5.92 | 7.28 | 0.04 |
| | FoxP3 | 0.09 | 0.24 | 10^{-5} | 0.13 | 0.28 | 10^{-5} |
| | CD137 | 0.19 | 0.29 | 0.23 | 0.31 | 0.32 | 0.58 |
| | CD8 PD1 | 2.19 | 0.95 | 10^{-6} | 2.04 | 0.78 | 10^{-11} |
| | CD8 & FoxP3 | 0.56 | 1.41 | 10^{-8} | 0.59 | 1.34 | 10^{-7} |
| Stroma | CD8 | 11.4 | 3.65 | 10^{-9} | 10.6 | 3.33 | 10^{-13} |
| | FoxP3 | 0.18 | 0.70 | 10^{-5} | 0.29 | 0.74 | 10^{-5} |
| | CD137 | 0.26 | 0.13 | 0.04 | 0.38 | 0.23 | 0.01 |
| | CD8 PD1 | 6.48 | 1.04 | 10^{-15} | 7.70 | 1.14 | 10^{-19} |
| | CD8 & FoxP3 | 1.77 | 0.91 | 10^{-4} | 2.28 | 0.98 | 10^{-6} |

POLE-WT patients (Fig. 5c, d) show lower level of immune active cells in the stroma. In summary, Synplex was able to generate sets of images from experimental data, realistically modeling the tumor microenvironment of the real tissue images.

C. First Proof-of-Principle Experiment: Use of Tumor Simulations for Augmenting Training Datasets in AI-Based Patient Predictions Systems

ML pipelines must be trained with large patient cohorts. In this section, we study if tumor simulations generated by Synplex could increase the predictive performance of such pipelines.

For this purpose, we trained NaroNet [30], a state-of-the-art weakly-supervised interpretable DL pipeline, using real tissues and the synthetically generated endometrial carcinoma dataset described in the previous section, to predict patient

mutation status, i.e., POLE-mutated vs. POLE-WT. The real dataset contained 12 patients, each composed, on average, of $5\,1876 \times 1404$ pixel images. Synplex simulated 50 patients, each represented by an artificial 2000×2000 pixel tissue image. To evaluate the patient prediction accuracy, we used a leave-one-out cross-validation strategy where, iteratively, images from all but one of the patients were used for training and the remaining patient was used for testing. As shown (Fig. 6a), the image-wise prediction accuracy increased from an AUC of 0.89 when trained only with real data, to 0.92 when trained using the combination of real and simulated data.

Next, given the importance of the image texture in the learning process, we tested if the performance of our model could be improved by using generative adversarial networks (GANs) [35], [36] to learn the texture parameters from real datasets instead of entering a parametrized set of values. To this end, we used the pix2pix framework [35], which translates images from one style to another, and has been used in H&E-stained histopathology images before, translating segmentation mask images to realistic tissue textures [36]. When this strategy is applied to our environment, we trained a GAN using 50 real images, where the input was a 24-dimensional segmentation mask each channel representing the nucleus or the whole cell of 12 simulated cell phenotypes. The output of the GAN consisted of the 6-marker immunofluorescence image. The model was trained with the hyperparameters proposed in ref. [35] during 600 epochs (~ 12 hours). Using the trained model, we obtained the realistic tissue texture for the 50 synthetic images used in the augmentation experiment. As shown, when NaroNet was trained with both real and synthetic tissue images with GAN-generated tissue texture, the predictive performance increased to an AUC of 0.95. This increase is especially relevant given the small, imbalanced sample size and the complexity of the classification.

We show a real POLE-mutated image (Fig. 6b) where patches with the highest relevance to predict the POLE mutation correspond to a specific phenotype that consists of stromal CD8 cells. These cells, as shown in Fig. 6c, are significantly more abundant in patients with the POLE mutation in the real images. Consistent with this finding, Fig. 6d shows the same behavior in our simulated dataset. This confirms the similarity between the real and simulated datasets and is consistent with our QuPath quantifications (see Table I), where the stromal CD8 population was associated with the POLE mutation.

D. Second Proof-of-Principle Experiment: In Silico Selection of Biomarkers

Biomarker discovery requires preselecting markers optimally related to a specific tumor characteristic. Although a large amount of knowledge about protein expression and cell-cell interactions exists in publicly available databases [37], [38], experimentally selecting a predictive marker subset is high time and resource consuming. We propose to use Synplex to test *in silico* the best marker combination for specific predictive tasks. To demonstrate this, we performed a simple proof-of-principle experiment, to evaluate *in silico*

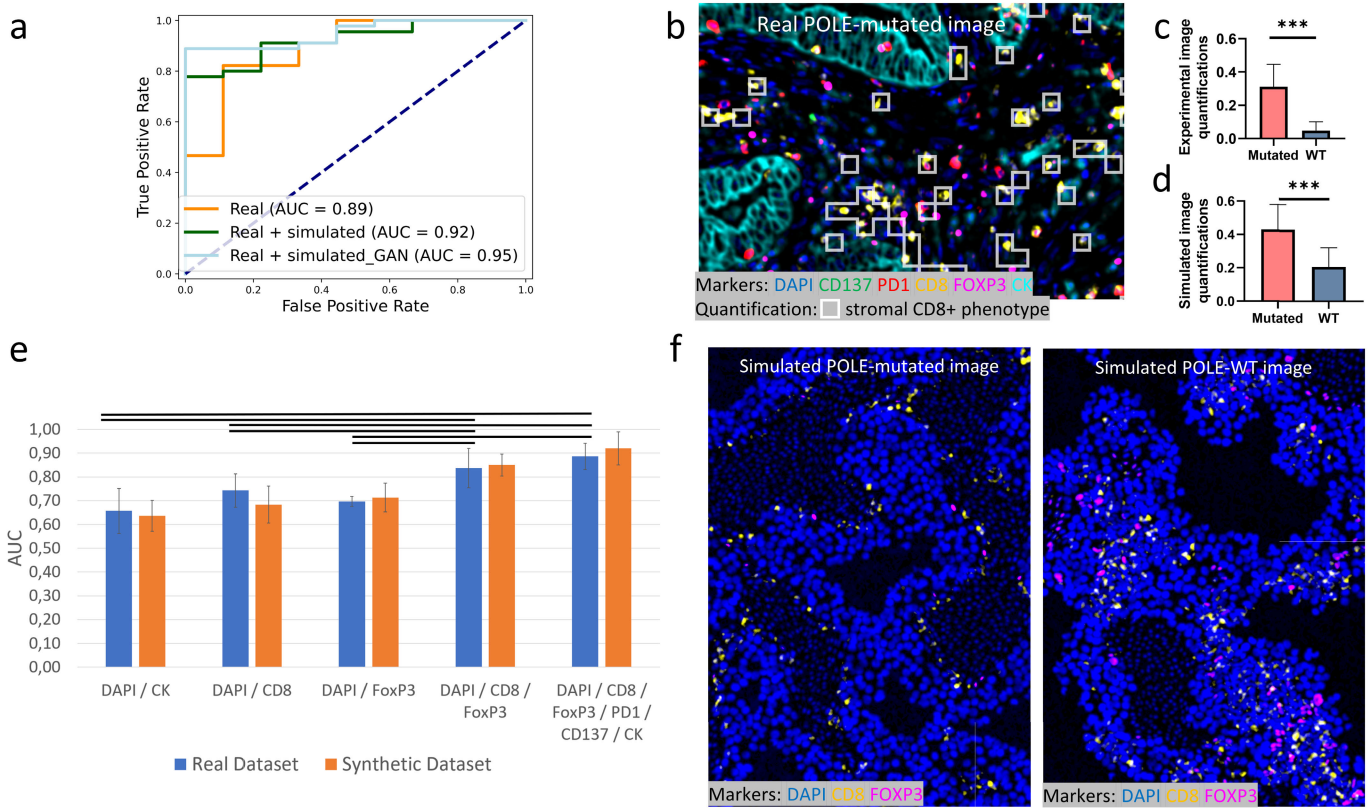


Fig. 6. Proof of principle for the possible applications of Synplex using a real endometrial carcinoma dataset and its simulated counterpart. **a** ROC curves showing prediction performance of NaroNet when combining real and simulated data for image-level prediction. **b-d** NaroNet's interpretability on **(b)** a real image showing the quantification of the stromal CD8+ phenotype that differentiates patients in both **(c)** real and **(d)** simulated images. **e** Prediction performance of NaroNet classifying simulated (POLE-Mutated and POLE-WT) images when using different combinations of markers (statistically significant differences are shown with black horizontal bars). **f** Examples of POLE-mutated and POLE-WT simulated images displaying CD8 and FOXP3 markers, which were tested *in silico* and showed a high predictive power.

the ability of ML to predict patient types using different combinations of markers. To this end, we trained NaroNet using the 12 synthetically generated endometrial carcinoma tumors, using different combinations of markers. *In silico* prediction performance was then tested in the 12 real patients with the expectation of seeing similar performances, with the same set of NaroNet parameters. To ensure a fair comparison, we used the same number of training epochs both in real and synthetic experiments. Each leave-one-out experiment was repeated three times and the resulting AUCs were averaged, to ensure statistical robustness. We used groups of marker candidates expected to have predictive value based on our previous QuPath quantifications (Fig. 4h, j). The results are shown in Fig. 6e. We performed a two-way ANOVA, using the marker combination and the nature of the dataset (synthetic or real) as factors in the analysis. The statistical results showed no difference between using the real or the synthetic database in the experiments, confirming the ability of Synplex to accurately simulate a real disease paradigm. Furthermore, statistical differences were found between some marker combinations: using DAPI+CK, DAPI+CD8 or DAPI+FOXP3 alone resulted in a significantly poorer prediction performance than using the combination of FOXP3 and CD8 or using all 5 markers. Interestingly, there was no significant prediction performance difference when expanding the number

of markers from 2 (CD8 and FOXP3) to 5 (CD137, CD8, PD1, FOXP3 and CK), meaning that our ML framework can predict endometrial carcinoma POLE-mutated patients using a 2-plex staining without requiring the a priori available 5-plex staining. As an example of this, we show (Fig. 6f) how CD8 and FOXP3 are differently distributed in POLE-mutated and POLE-WT images. As a conclusion of this simple proof-of-principle experiment, we envision Synplex as a discovery tool to help predicting *in silico* which combination(s) of available markers could provide *a priori* better accuracy for a particular predictive task.

IV. DISCUSSION

We have presented Synplex, a novel simulator of synthetic MI images that recreates tumor microenvironments containing predefined cell phenotypes and cellular neighborhoods. Synplex consists of three sequential independent modules. The first creates a tissue mask with a predefined presence of cellular neighborhoods and fixed rules of interaction between these neighborhoods. Each neighborhood is in turn defined by a specific composition of cell phenotypes and interactions between them. The second module populates each neighborhood with the predefined cell phenotypes, defined by the expression of specific markers and morphological features. The third module creates the final images, giving them a

realistic tissue texture by simulating the acquisition process of the tissue by an optical microscope. As Synplex' modules are independent pieces of code they could be easily extended and tuned for simulating other organs' histology from acquired MI images, such as whole-brain tissue mapping [39].

Synplex has been validated with two experiments. In the first, synthetic tumor microenvironments were created following user-defined parameters. Then the simulated cell populations and interactions were quantified to confirm the accuracy of the simulation, by comparing those values with the predefined values. This experiment confirmed that indeed, Synplex successfully produces images containing the expected tumor microenvironment. In the second experiment, we quantified phenotypes and neighborhoods from images acquired using a 6-marker panel on an endometrial carcinoma dataset and fed the quantified values to Synplex to produce a synthetic patient cohort that successfully replicated the disease paradigm observed in the real image dataset. We showed that Synplex simulates the behavior of entire tissue cohorts as well as the tissue microenvironment elements that have higher predictive value, showing the same statistical behavior both in the real and simulated cohorts.

These results show that Synplex may help with the development of new image analysis pipelines designed to analyze MI tissue images. Indeed, artificial datasets generated using Synplex could be used for the augmentation of annotated image datasets destined to train ML pipelines. To support this idea we have presented, as proof-of-principle, an example of how the combination of synthetic and real images in the training phase of a weakly supervised DL model boosts the accuracy of the prediction of POLE mutation status in a very small, unbalanced cohort of endometrial carcinomas. Therefore, the use of synthetic samples, for a modality such as MI in which the availability of image datasets is limited and annotation is often impracticable, can indeed help training machine learning models to perform cell segmentation, tissue segmentation, or patient classification. In contrast to GAN-based augmentation methods that generate data with little control of the simulated images, Synplex generates images with quantifiable parameters accompanied by masks at the cell and neighborhood level. Indeed, we used Synplex phenotype and neighborhood output to generate GAN-based tissue texture, which is possible when real images are available. Furthermore, Synplex can be used for benchmarking of new bioimage analysis tools, a task which is often complicated due to the scarcity of annotated data. This is especially true in the case of MI imaging data. Here, the interactivity between cell phenotypes or neighborhoods is often weak, and thus, tools that aim to make relevant biomedical discoveries by quantifying those relationships must be systematically validated. In this context, Synplex allows researchers to create images with adjustable parameters, generating artificial patient cohorts, where patient types can be defined by specific variations in elements from the tumor microenvironment. Therefore, it is possible to fine-tune the abundance of a phenotype in the tissue or the strength of the interaction between two cell populations, thus creating different levels of difficulty to validate algorithms.

Finally, the fact that Synplex simulates complex tumor microenvironments fed with experimental data obtained from quantitative studies that quantified tumor features individually (e.g., fluorescent expression of each marker, the abundance of each cell phenotype, etc.) could allow researchers to test *in silico* which potential targets are more relevant for ML methods designed to differentiate between tumors. We have shown a proof-of-principle example of this, in which we have determined *in silico*, using synthetic images, which combination of markers from a 6-plex immune panel are the most decisive for distinguishing between POLE mutated vs POLE WT endometrial tumors, and we have confirmed this finding using the corresponding real image dataset. This could be straightforwardly extended to other experimental design parameters (e.g., the number of tissue samples used per patient, the number of images per tissue, the location of images – peripheral or central tumor areas-, etc.). Therefore, artificial patient cohorts could be created following different tissue selection strategies to study *in silico* which design decisions are more likely to produce better data for a given ML-based predictive task. These *in silico* strategies would result in significant time and monetary savings. This is especially true in a MI setup, given the high number of possible combinations that exist.

REFERENCES

- [1] R. Baghban et al., "Tumor microenvironment complexity and therapeutic implications at a glance," *Cell Commun. Signaling*, vol. 18, no. 1, pp. 1–19, Dec. 2020.
- [2] I. Melero, A. Teijeira, F. Aranda, and P. Berraondo, "Firefighters for the wrong type of inflammation in tumors," *Cancer Discovery*, vol. 11, no. 10, pp. 2372–2374, Oct. 2021.
- [3] L. Keren and M. Angelo, "Mapping cell phenotypes in breast cancer," *Nature Cancer*, vol. 1, no. 2, pp. 156–157, Feb. 2020.
- [4] C. Farrera et al., "T-cell phenotyping uncovers systemic features of atopic dermatitis and psoriasis," *J. Allergy Clin. Immunol.*, vol. 145, no. 3, pp. 1021–1025, Mar. 2020.
- [5] C. M. Schürch et al., "Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front," *Cell*, vol. 182, no. 5, pp. 1341–1359, Sep. 2020.
- [6] P. Bankhead et al., "QuPath: Open source software for digital pathology image analysis," *Sci. Rep.*, vol. 7, no. 1, pp. 1–7, Dec. 2017.
- [7] C. R. Stoltzfus et al., "CytoMAP: A spatial analysis toolbox reveals features of myeloid cell organization in lymphoid tissues," *Cell Rep.*, vol. 31, no. 3, Apr. 2020, Art. no. 107523.
- [8] K. E. Blise, S. Sivagnanam, G. L. Banik, L. M. Coussens, and J. Goecks, "Single-cell spatial architectures associated with clinical outcome in head and neck squamous cell carcinoma," *NPJ Precis. Oncol.*, vol. 6, no. 1, pp. 1–14, Feb. 2022.
- [9] D. Jimenez-Sanchez, M. Ariz, and C. Ortiz-de-Solorzano, "Unsupervised learning of contextual information in multiplex immunofluorescence tissue cytometry," in *Proc. IEEE 17th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2020, pp. 1275–1279.
- [10] C. Stringer, T. Wang, M. Michaelos, and M. Pachitariu, "Cellpose: A generalist algorithm for cellular segmentation," *Nature Methods*, vol. 18, no. 1, pp. 100–106, 2021.
- [11] L. von Chamier et al., "Democratising deep learning for microscopy with ZeroCostDL4Mic," *Nature Commun.*, vol. 12, no. 1, p. 2276, Apr. 2021.
- [12] K. Lee et al., "Deep learning of histopathology images at the single cell level," *Frontiers Artif. Intell.*, vol. 4, p. 137, Sep. 2021.
- [13] C. Yapp et al., "UnMICST: Deep learning with real augmentation for robust segmentation of highly multiplexed images of human tissues human tumor atlas network," *Commun. Biol.*, vol. 5, 2022, Art. no. 1263.

- [14] V. Ulman, D. Svoboda, M. Nykter, M. Kozubek, and P. Ruusuvoori, "Virtual cell imaging: A review on simulation methods employed in image cytometry," *Cytometry A*, vol. 89, no. 12, pp. 1057–1072, 2016.
- [15] D. Wiesner, D. Svoboda, M. Maška, and M. Kozubek, "CytoPacq: A web-interface for simulating multi-dimensional cell imaging," *Bioinformatics*, vol. 35, no. 21, pp. 4531–4533, Nov. 2019.
- [16] D. V. Sorokin et al., "FiloGen: A model-based generator of synthetic 3-D time-lapse sequences of single motile cells with growing and branching filopodia," *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2630–2641, Dec. 2018.
- [17] C. Castilla, M. Maska, D. V. Sorokin, E. Meijering, and C. Ortiz-de-Solorzano, "3-D quantification of filopodia in motile cancer cells," *IEEE Trans. Med. Imag.*, vol. 38, no. 3, pp. 862–872, Mar. 2019.
- [18] P. Baniukiewicz, E. J. Lutton, S. Collier, and T. Bretschneider, "Generative adversarial networks for augmenting training data of microscopic cell images," *Frontiers Comput. Sci.*, vol. 1, p. 10, Nov. 2019.
- [19] M. Berghoff, J. Rosenbauer, F. Hoffmann, and A. Schug, "Cells in silico—Introducing a high-performance framework for large-scale tissue modeling," *BMC Bioinf.*, vol. 21, no. 1, pp. 1–21, Dec. 2020.
- [20] E. H. B. Maia, L. C. Assis, T. A. de Oliveira, A. M. da Silva, and A. G. Taranto, "Structure-based virtual screening: From classical to artificial intelligence," *Frontiers Chem.*, vol. 8, p. 343, Apr. 2020.
- [21] S. Pushpakom et al., "Drug repurposing: Progress, challenges and recommendations," *Nature Rev. Drug Discovery*, vol. 18, no. 1, pp. 41–58, Jan. 2019.
- [22] L. Wang et al., "Synthetic genomics: From DNA synthesis to genome design," *Angew. Chem. Int. Ed.*, vol. 57, no. 7, pp. 1748–1756, Feb. 2018.
- [23] K. Grzesik, K. Eng, F. Crauste, and S. Battaglia, "In silico methods for studying T cell biology," in *International Review of Cell and Molecular Biology*, vol. 342. New York, NY, USA: Elsevier, 2019, pp. 265–304.
- [24] C. Gong et al., "A computational multiscale agent-based model for simulating spatio-temporal tumour immune response to PD1 and PDL1 inhibition," *J. Roy. Soc. Interface*, vol. 14, no. 134, Sep. 2017, Art. no. 20170320.
- [25] G. Letort et al., "PhysiBoSS: A multi-scale agent-based modelling framework integrating physical dimension and cell signalling," *Bioinformatics*, vol. 35, no. 7, pp. 1188–1196, Apr. 2019.
- [26] B. Aguilar et al., "A generalizable data-driven multicellular model of pancreatic ductal adenocarcinoma," *GigaScience*, vol. 9, no. 7, Jul. 2020, Art. no. g1aa075.
- [27] D. Jimenéz-Sánchez, M. Ariz, and C. Ortiz-de-Solórzano, "Synplex: A synthetic simulator of highly multiplexed histological images," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Jul. 2021, pp. 1–4.
- [28] E.-M. Kamionka et al., "Collagen organization does not influence T-cell distribution in stroma of human pancreatic cancer," *Cancers*, vol. 13, no. 15, p. 3648, Jul. 2021.
- [29] L. Xuemei, L. Huan, and Y. Hui, "Generation of organ texture with Perlin noise," in *Proc. Int. Conf. E-Health Netw. Digit. Ecosyst. Technol. (EDT)*, vol. 1, Apr. 2010, pp. 150–152.
- [30] D. Jiménez-Sánchez, M. Ariz, H. Chang, X. Matias-Guiu, C. E. de Andrea, and C. Ortiz-de-Solórzano, "NaroNet: Discovery of tumor microenvironment elements from highly multiplexed images," *Med. Image Anal.*, vol. 78, May 2022, Art. no. 102384.
- [31] A. León-Castillo et al., "Clinicopathological and molecular characterisation of 'multiple-classifier' endometrial carcinomas," *J. Pathol.*, vol. 250, no. 3, pp. 312–322, Mar. 2020.
- [32] Q. Duan, H. Zhang, J. Zheng, and L. Zhang, "Turning cold into hot: Firing up the tumor microenvironment," *Trends Cancer*, vol. 6, no. 7, pp. 605–618, Jul. 2020.
- [33] M. Abengoza-Muela et al., "Diverse immune environments in human lung tuberculosis granulomas assessed by quantitative multiplexed immunofluorescence," *Mod. Pathol.*, vol. 33, no. 12, pp. 2507–2519, Dec. 2020.
- [34] A. Feichtenbeiner, M. Haas, M. Büttner, G. G. Grabenbauer, R. Fietkau, and L. V. Distel, "Critical role of spatial interaction between CD8⁺ and Foxp3⁺ cells in human gastric cancer: The distance matters," *Cancer Immunol., Immunotherapy*, vol. 63, no. 2, pp. 111–119, Feb. 2014.
- [35] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. 30th IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5967–5976.
- [36] F. Mahmood et al., "Deep adversarial training for multi-organ nuclei segmentation in histopathology images," *IEEE Trans. Med. Imag.*, vol. 39, no. 11, pp. 3257–3267, Nov. 2020.
- [37] HuBMAP Consortium, "The human body at cellular resolution: The NIH human biomolecular atlas program," *Nature*, vol. 574, no. 7777, pp. 187–192, 2019.
- [38] P. J. Thul and C. Lindskog, "The human protein atlas: A spatial map of the human proteome," *Protein Sci.*, vol. 27, no. 1, pp. 233–244, Jan. 2018.
- [39] D. Maric et al., "Whole-brain tissue mapping toolkit using large-scale highly multiplexed immunofluorescence imaging and deep neural networks," *Nature Commun.*, vol. 12, no. 1, pp. 1–12, Mar. 2021.