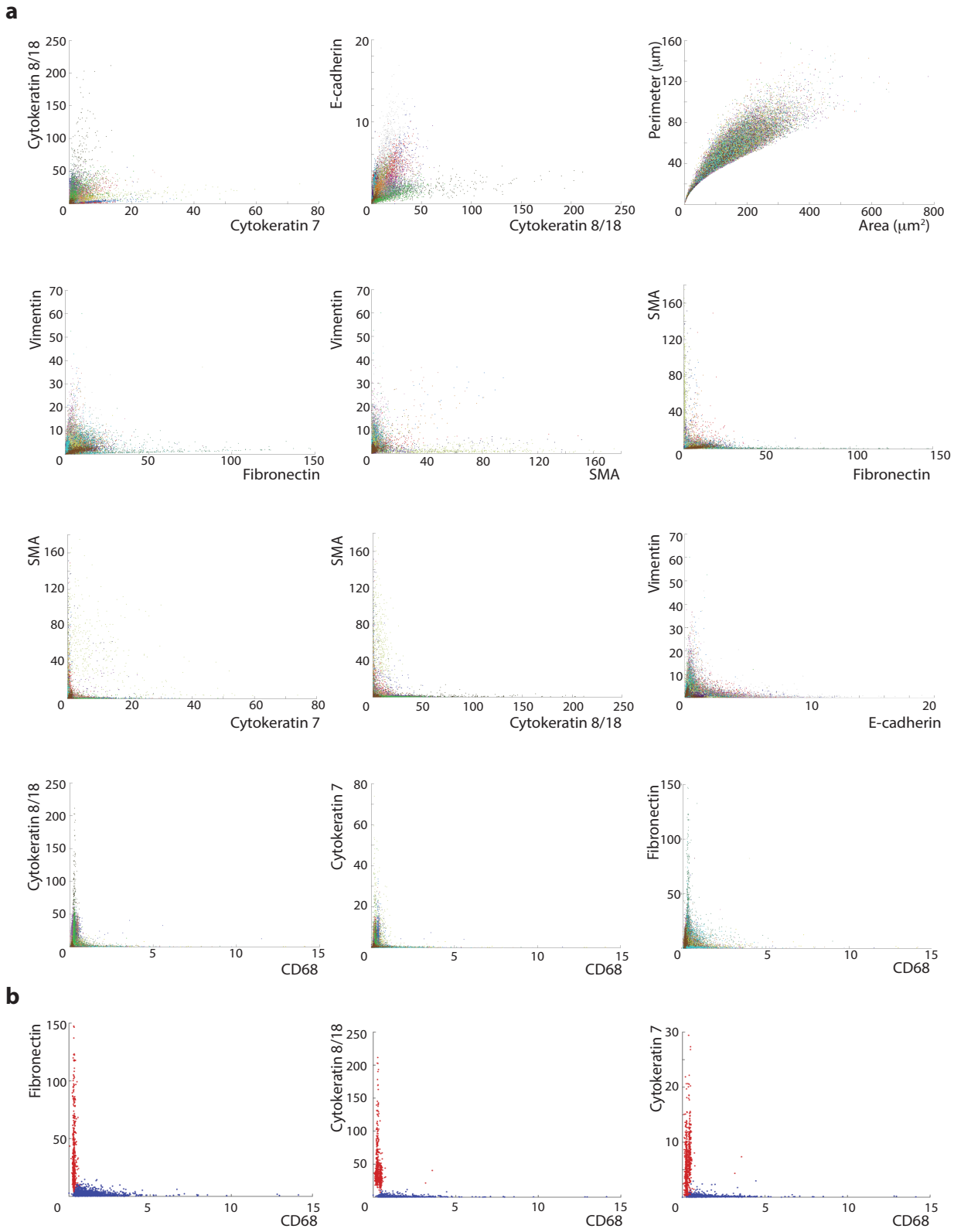
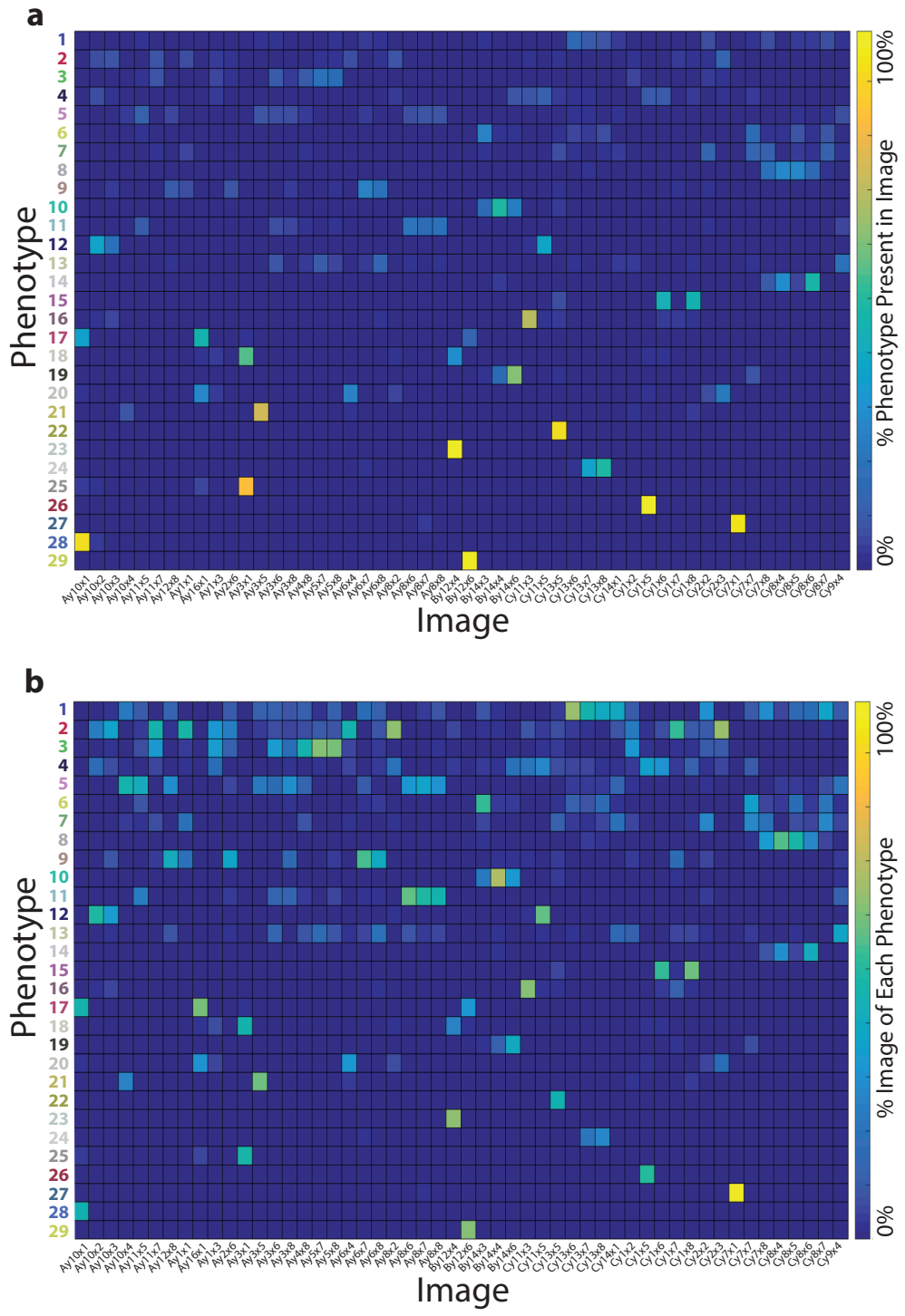
	Cytometry Analysis	Image Analysis	Multiplexed Data Analysis	"Round-trip"	Open Source	Neighborhood Analysis
miCAT	✔	✔	✔	✔	✔	✔
CellProfiler Analyst[16]	✔	✔	✖	✖	✔	✖
CYT[17]	✔	✖	✔	✖	✔	✖
ACCENSE[18]	✔	✖	✔	✖	✔	✖
C-PATH[15]	✖	✔	✖	✖	✖	✖
GRAPHIE[14]	✖	✔	✖	✖	✖	✖

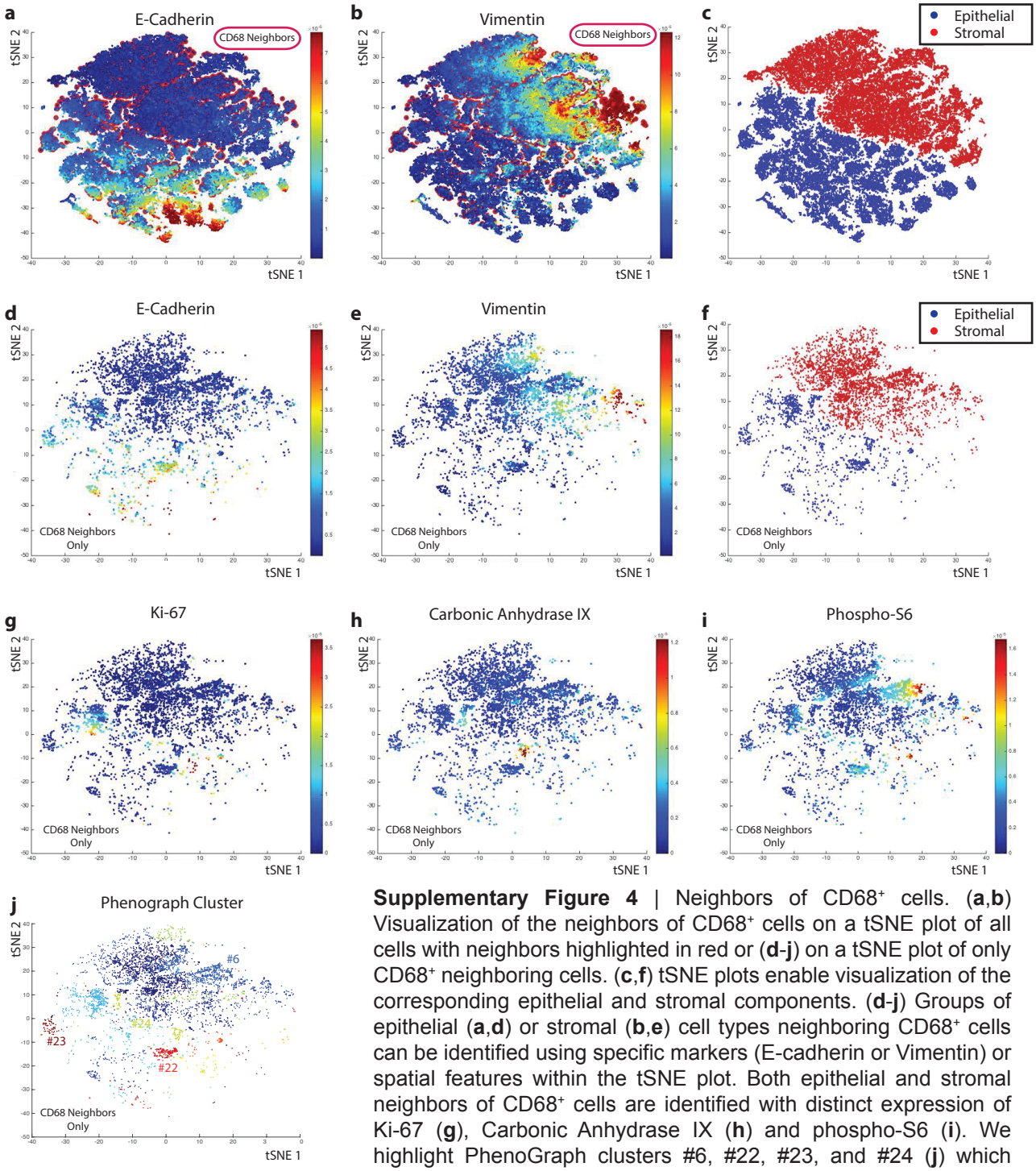
Supplementary Fig 1 | Summary of features available in current analysis software tools. miCAT combines features from multiplexed cytometry and image analysis in one user-friendly open source toolbox. Proprietary commercial software is excluded from this summary.



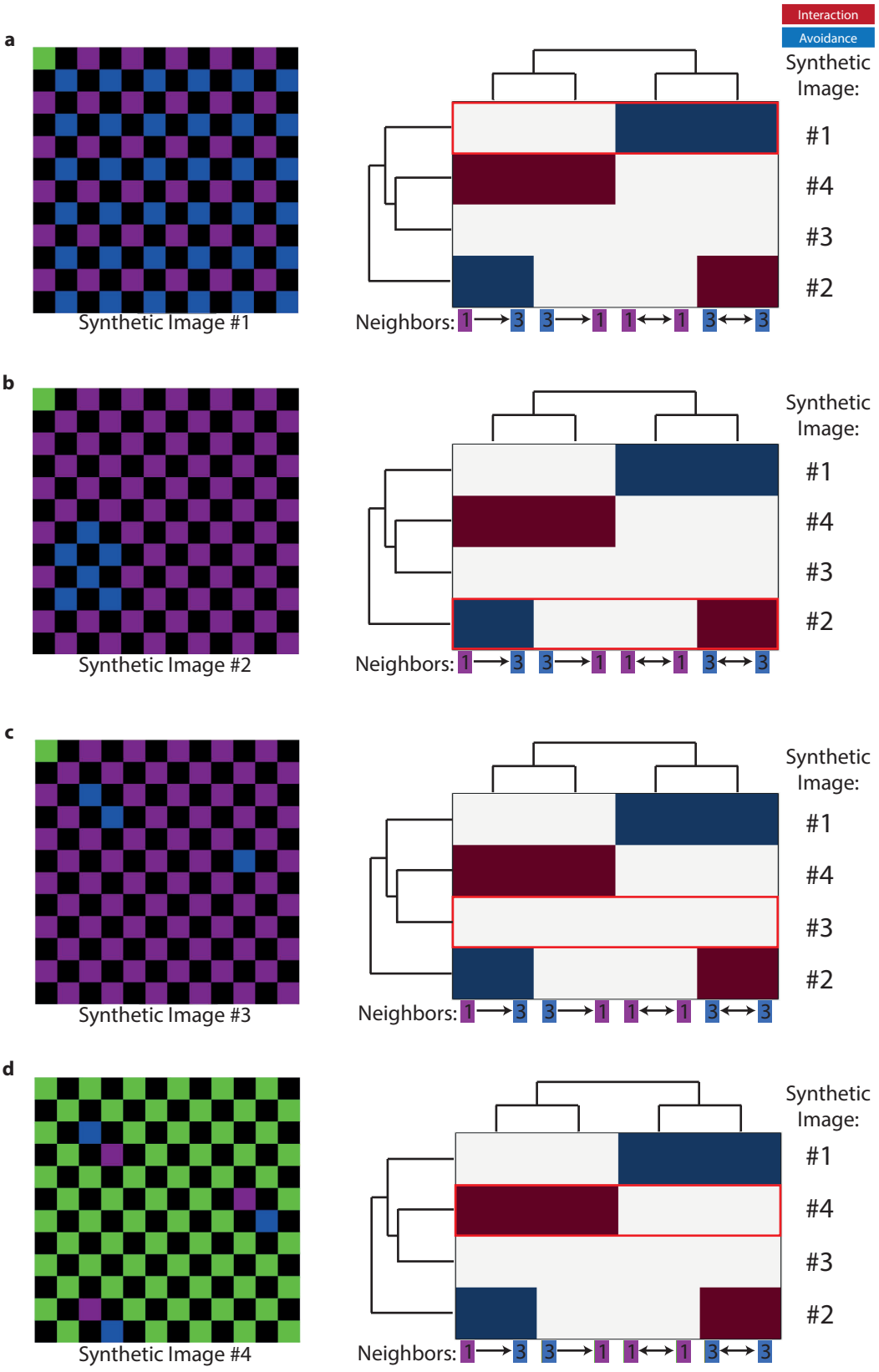
Supplementary Figure 2 | Representative biaxial plots of single cell data from Figure 2. (**a**) All cells colored by source image, and (**b**) PhenoGraph cell phenotype #7 $CD68^{+}$ (blue) and Fibronectin+ phenotype #26 or Cytokeratin 8/18+ and Cytokeratin 7+ phenotype #20 (red).

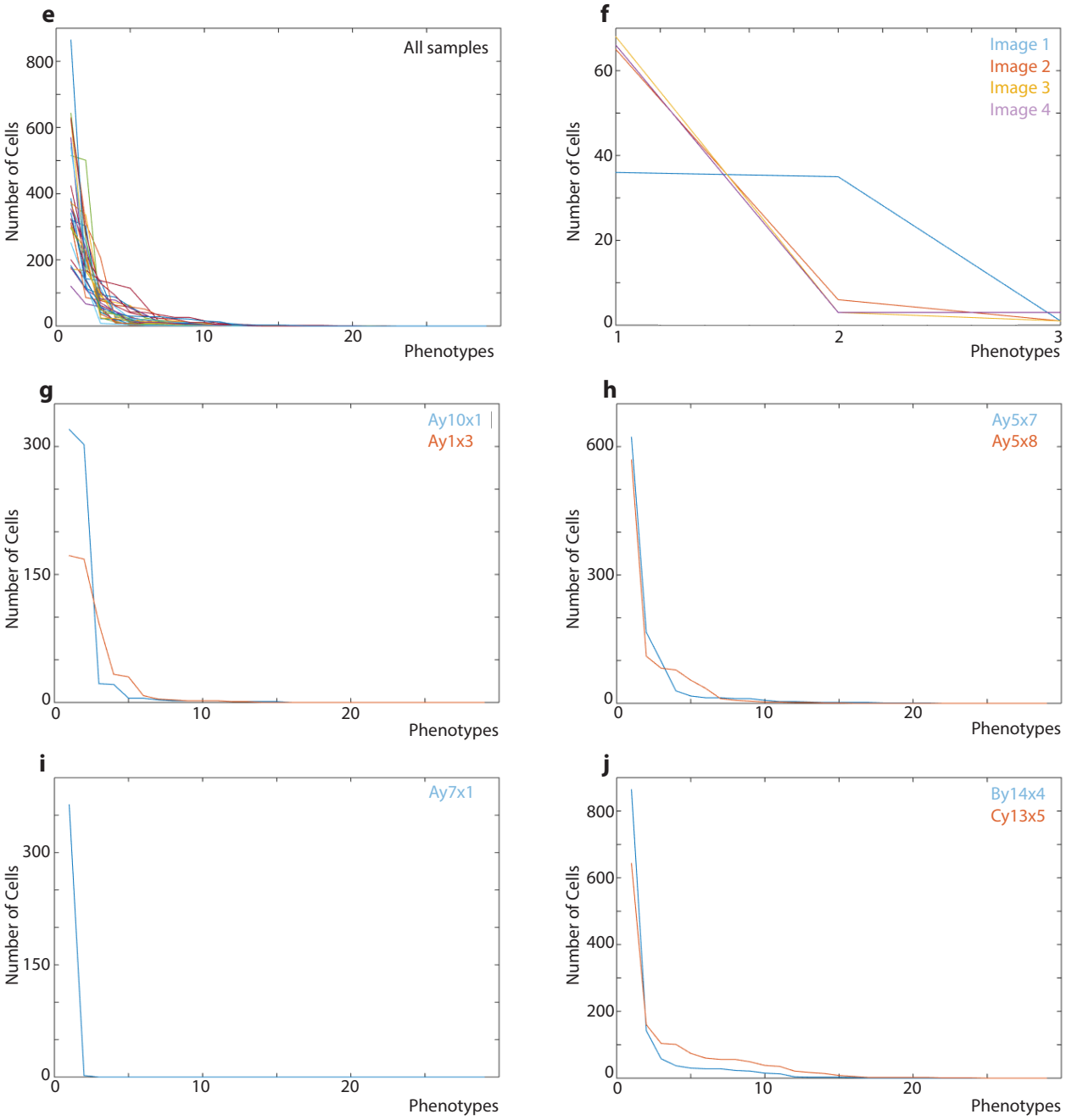


Supplementary Figure 3 | Distribution of cell phenotypes. (**a**) Heatmap depicting the source images of each phenotype where color intensity depicts the percentage of a cell phenotype present in each image. (**b**) Heatmap depicting the variety of cell phenotypes in each image where color intensity depicts the percentage of cells in the image from each cell phenotype.



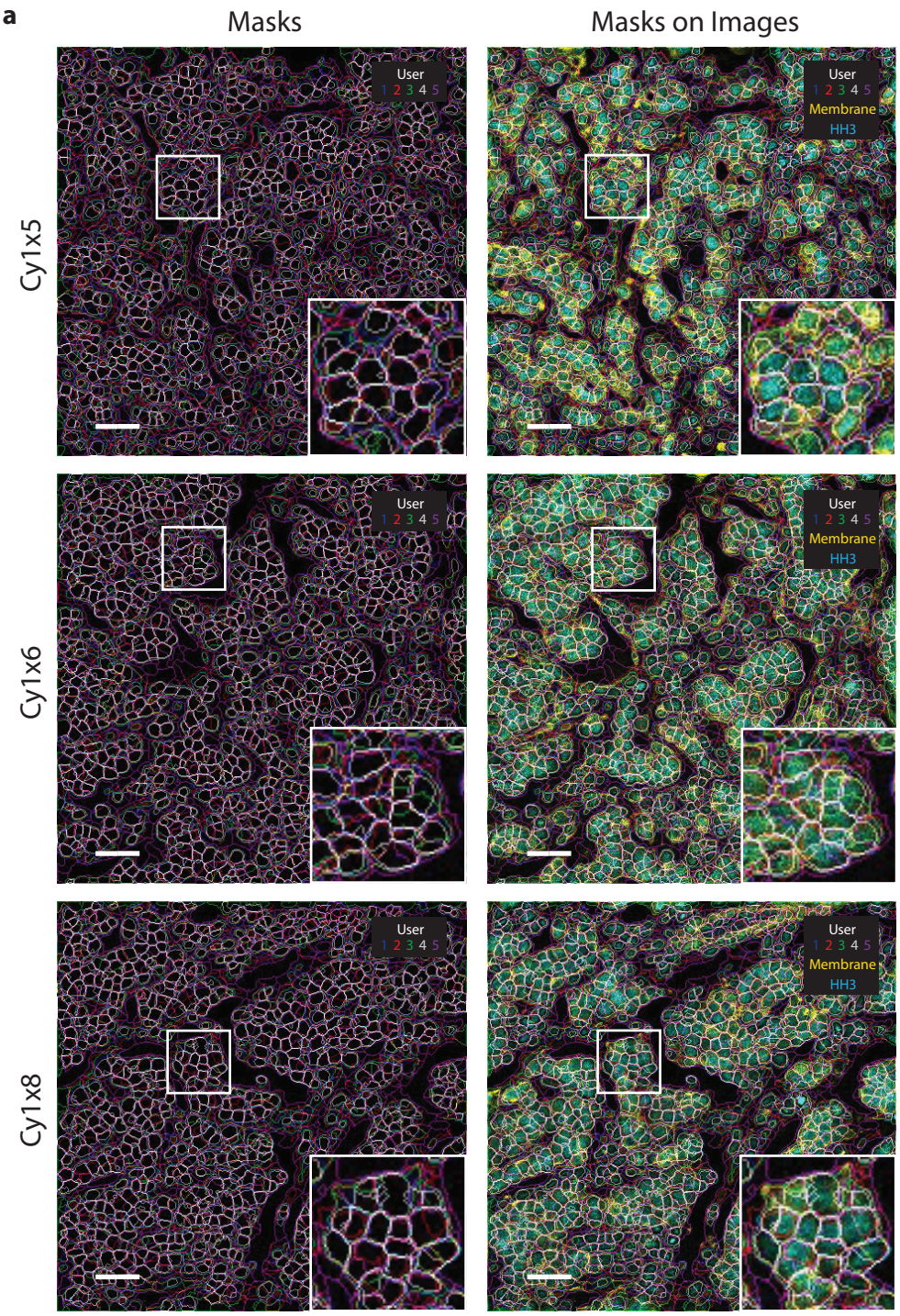
Supplementary Figure 4 | Neighbors of CD68⁺ cells. (**a**,**b**) Visualization of the neighbors of CD68⁺ cells on a tSNE plot of all cells with neighbors highlighted in red or (**d**-**j**) on a tSNE plot of only CD68⁺ neighboring cells. (**c**,**f**) tSNE plots enable visualization of the corresponding epithelial and stromal components. (**d**-**j**) Groups of epithelial (**a**,**d**) or stromal (**b**,**e**) cell types neighboring CD68⁺ cells can be identified using specific markers (E-cadherin or Vimentin) or spatial features within the tSNE plot. Both epithelial and stromal neighbors of CD68⁺ cells are identified with distinct expression of Ki-67 (**g**), Carbonic Anhydrase IX (**h**) and phospho-S6 (**i**). We highlight PhenoGraph clusters #6, #22, #23, and #24 (**j**) which suggests a relationship and cellular crosstalk between CD68⁺ cells and these specific cell types.


a
Interaction
Avoidance
Synthetic Image:
#1
#4
#3
#2
Synthetic Image #1
Neighbors: 1 → 3 3 → 1 1 ↔ 1 3 ↔ 3
b
Synthetic Image:
#1
#4
#3
#2
Synthetic Image #2
Neighbors: 1 → 3 3 → 1 1 ↔ 1 3 ↔ 3
c
Synthetic Image:
#1
#4
#3
#2
Synthetic Image #3
Neighbors: 1 → 3 3 → 1 1 ↔ 1 3 ↔ 3
d
Synthetic Image:
#1
#4
#3
#2
Synthetic Image #4
Neighbors: 1 → 3 3 → 1 1 ↔ 1 3 ↔ 3



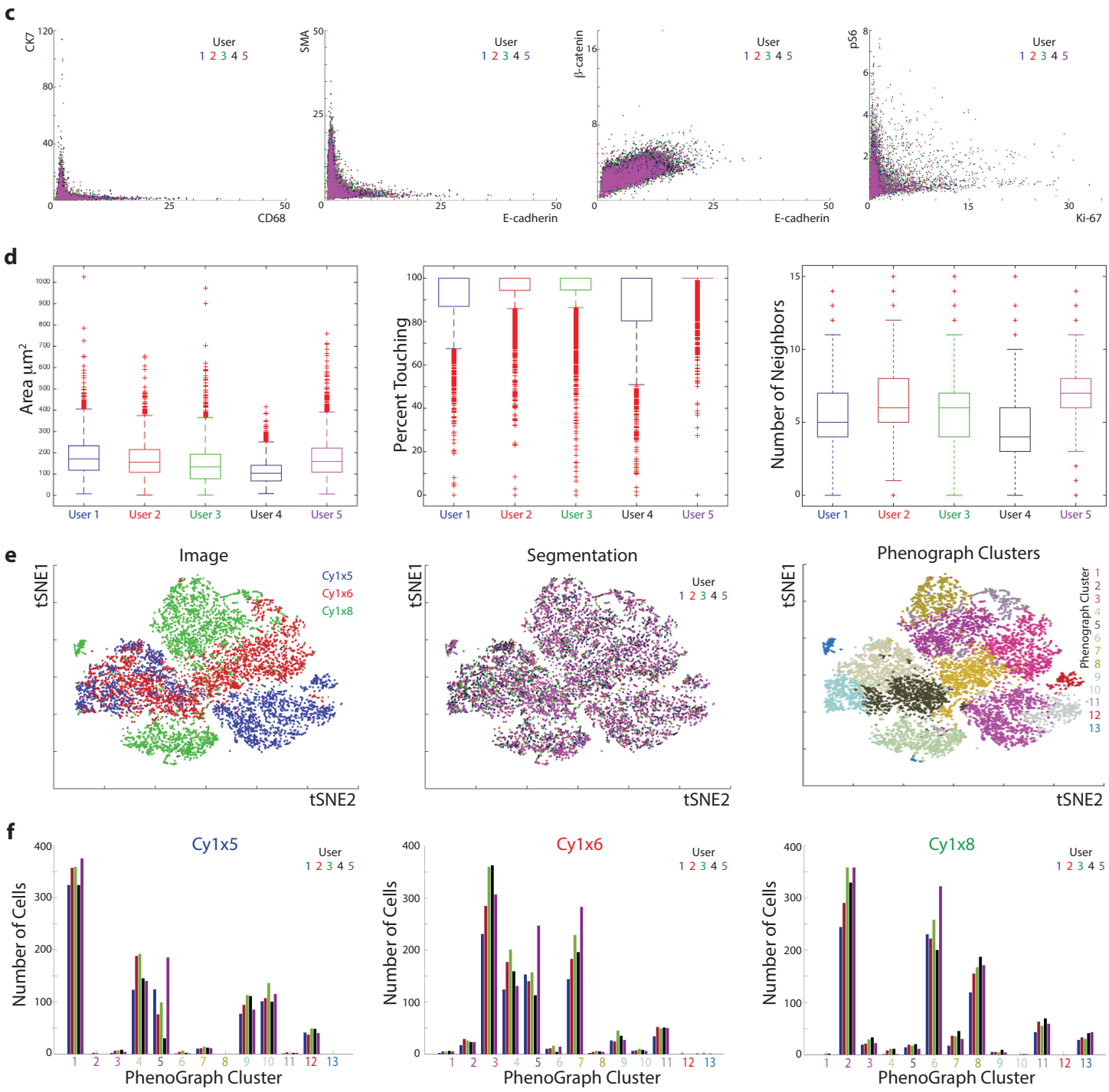
Supplementary Figure 5 | Neighborhood analysis validation using synthetic data. Four synthetic images were created containing three synthetic, square cell types (1 = Purple, 2 = Green, 3 = Blue) that only touch at their corners with black depicting the absence of a cell. The interactions in these four images are represented in a single hierarchically clustered heatmap. Cell type 2 is not in any significant interactions and therefore does not appear in the hierarchically clustered heatmap. (**a**) In synthetic image #1 cells do not neighbor other cells of the same phenotype. The hierarchically clustered heatmap indicates that interactions between cells of the same type (1←→1, 3←→3) are less likely than a random distribution of cells. (**b**) Synthetic image #2 has a distinct, large cluster of cell type 3 . The hierarchically clustered heatmap indicates significant interactions between cells of type 3 and a decreased number of interactions between cells of type 1 and 3 compared to a random distribution throughout. (**c**) In Synthetic image #3, rare, small clusters of cells do not result in statistically significant interactions ($p<0.05$) and are therefore not displayed in the hierarchically clustered heatmap. (**d**) Distinct pairs of cells in synthetic image #4 result in significant interactions in both directions between cell types 1 and 3. The distribution of cell frequencies is displayed as the quantity of each cell phenotype ordered from the most to the least abundant for (**e**) all tissue samples and (**f**) synthetic images. Specific breast cancer image examples that have similar relative cell frequencies as presented in the synthetic data are shown in (**g**)-(**j**). (**g**) Breast cancer images with two abundant cell types akin to synthetic data image #1. (**h**) Breast cancer images with abundant and rare cell types akin to synthetic data image #2. (**i**) Breast cancer images with extremely rare cell types akin to synthetic data image #3. (**j**) Breast cancer images with small clusters of cells akin to synthetic data image #4.

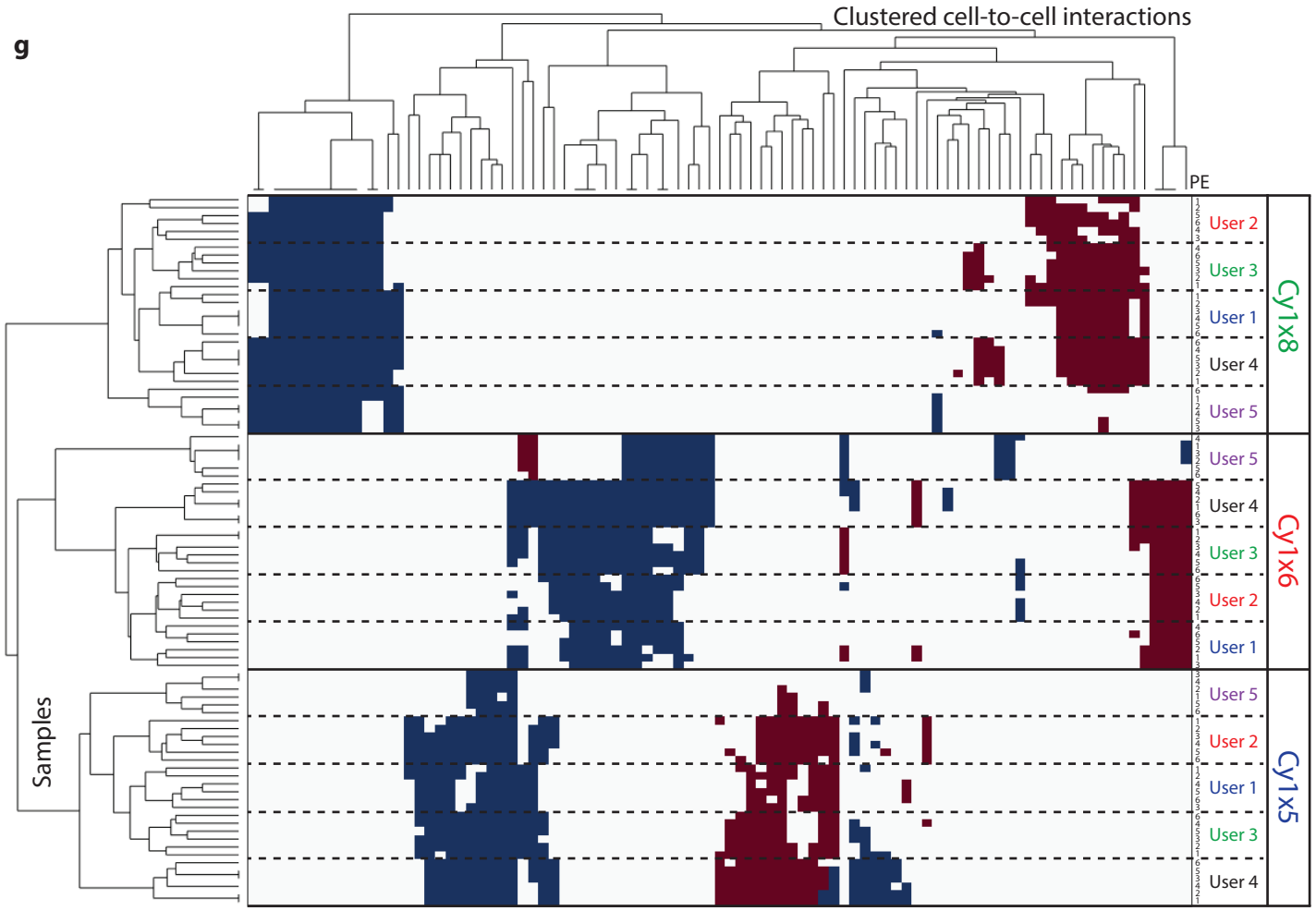
a



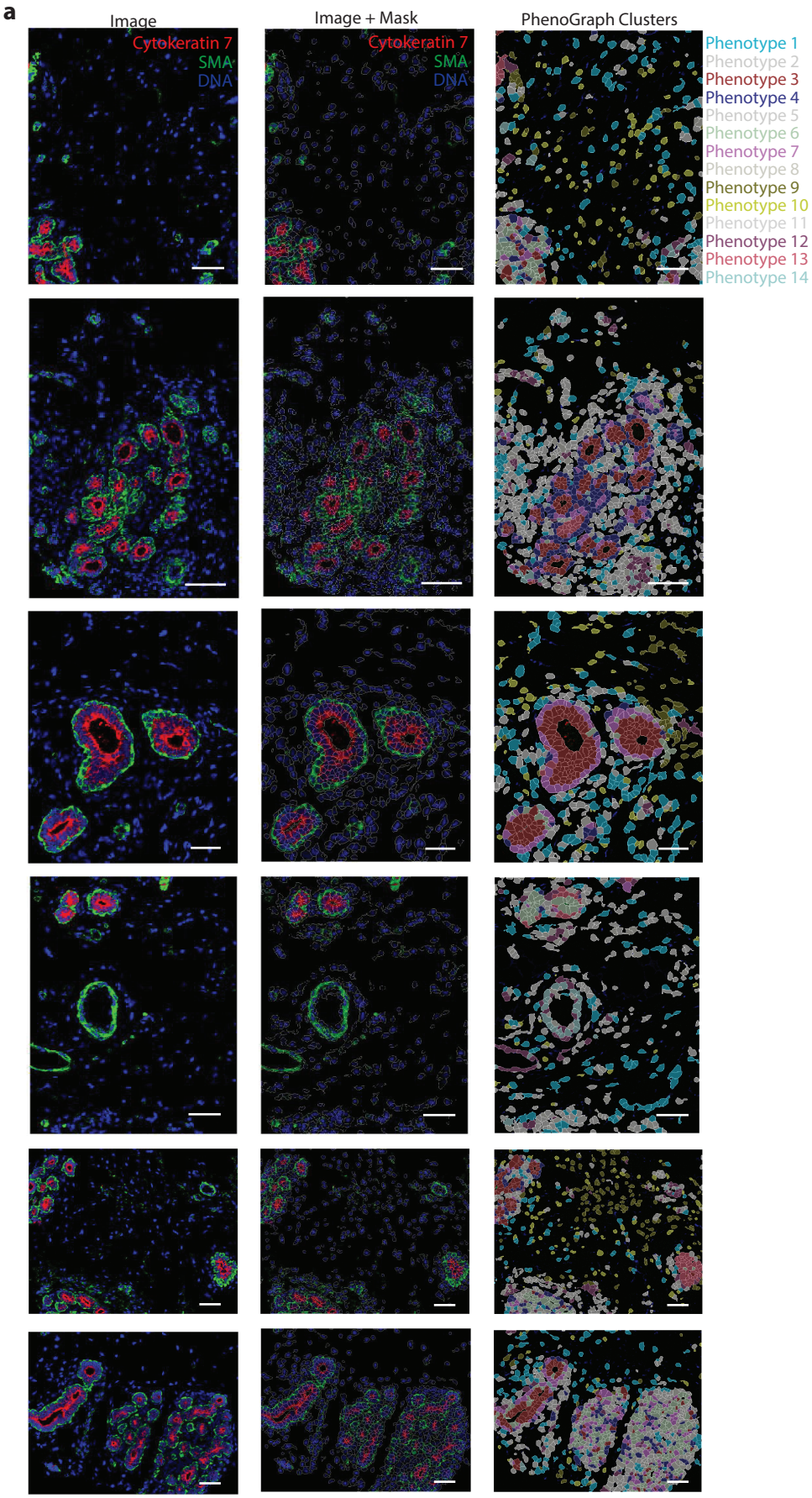
b

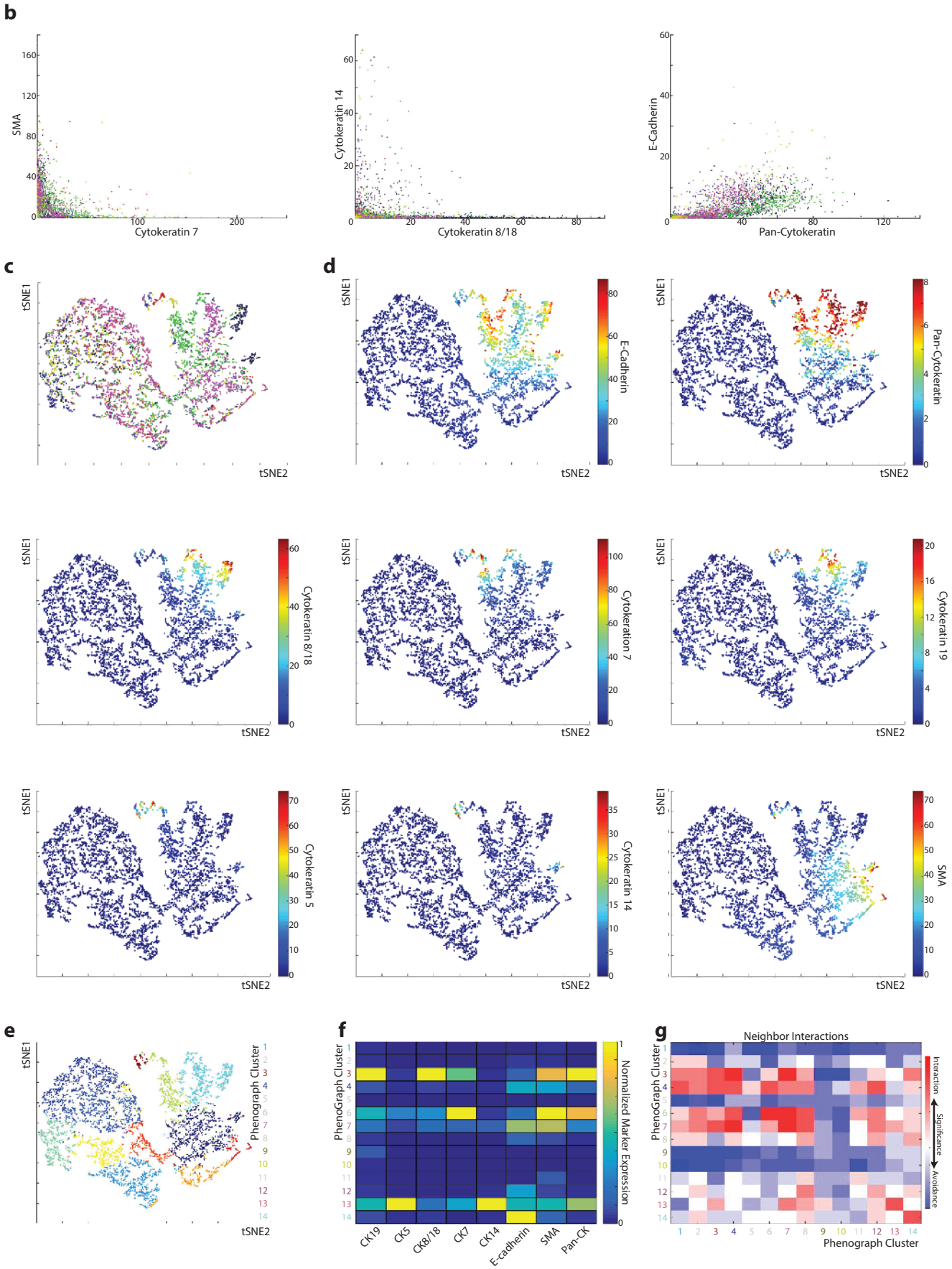
	Cy1x5	Cy1x6	Cy1x8
User 1	0.45	**0.46**	0.50
User 2	0.42	0.44	0.49
User 3	0.43	0.42	0.50
User 4	**0.47**	0.45	**0.51**
User 5	0.41	0.40	0.48


c
CK7
CD68
SMA
E-cadherin
β-catenin
E-cadherin
pS6
Ki-67
User
1 2 3 4 5
d
Area µm²
Percent Touching
Number of Neighbors
User 1
User 2
User 3
User 4
User 5
e
Image
Segmentation
Phenograph Clusters
tSNE1
tSNE2
Cy1x5
Cy1x6
Cy1x8
Phenograph Cluster
f
Cy1x5
Cy1x6
Cy1x8
Number of Cells
PhenoGraph Cluster

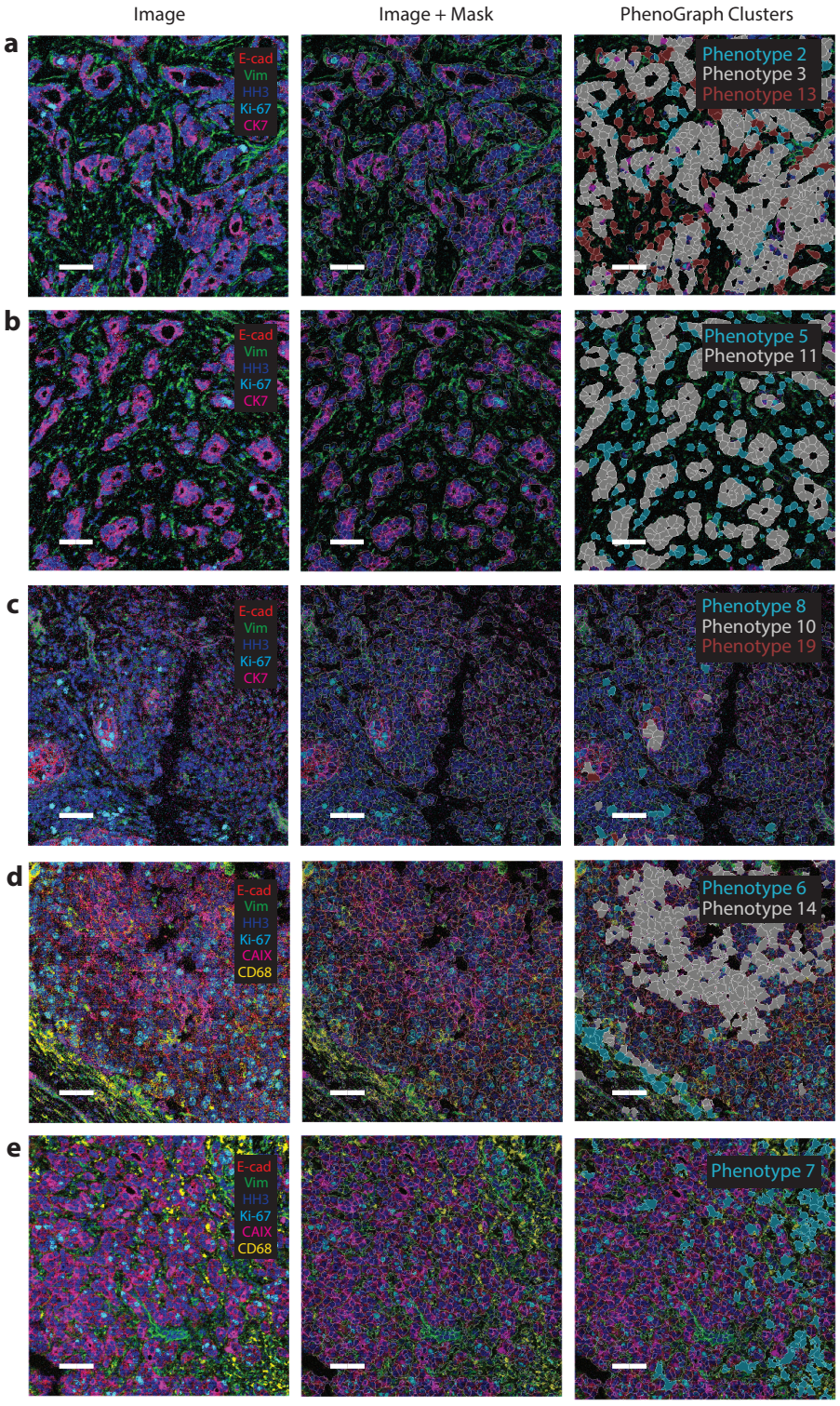


Supplementary Figure 6 | Segmentation Comparison. **(a)** Visualization of single cell segmentation masks created by five independent users alone (Left) and overlaid on top of IMC tissue image (Right, Membrane = E-cadherin + CK8/18 + CK7 + Her2 + SMA (yellow), DNA Intercalator (blue)). Scale bar 50µm. **(b)** Table of mask segmentation scores with the best score highlighted in red. No single user provides the best score for all images and User 5 consistently performs worse than the other users. **(c)** Biaxial plots of single cell data from five independent segmentations of three images colored by user. **(d)** Boxplots of single cell spatial features from five independent segmentations of three images. **(e)** tSNE maps of single cell data from five independent segmentations of three images colored by source image, segmentation user, and PhenoGraph cluster. **(f)** Histograms of the number of cells from each PhenoGraph cluster in each image that were identified by each user. **(g)** Agglomerative clustering of all cell-to-cell interactions identified by pixel expansion (PE) of 1- 6 pixels in the three images by each user according to the presence of significant ($p < 0.01$) phenotype interaction (red) or avoidance (blue). White represents interactions that are not present or not significant. Extended, less specific single cell mask by inexperienced User 5 results in outlier neighbor analysis.

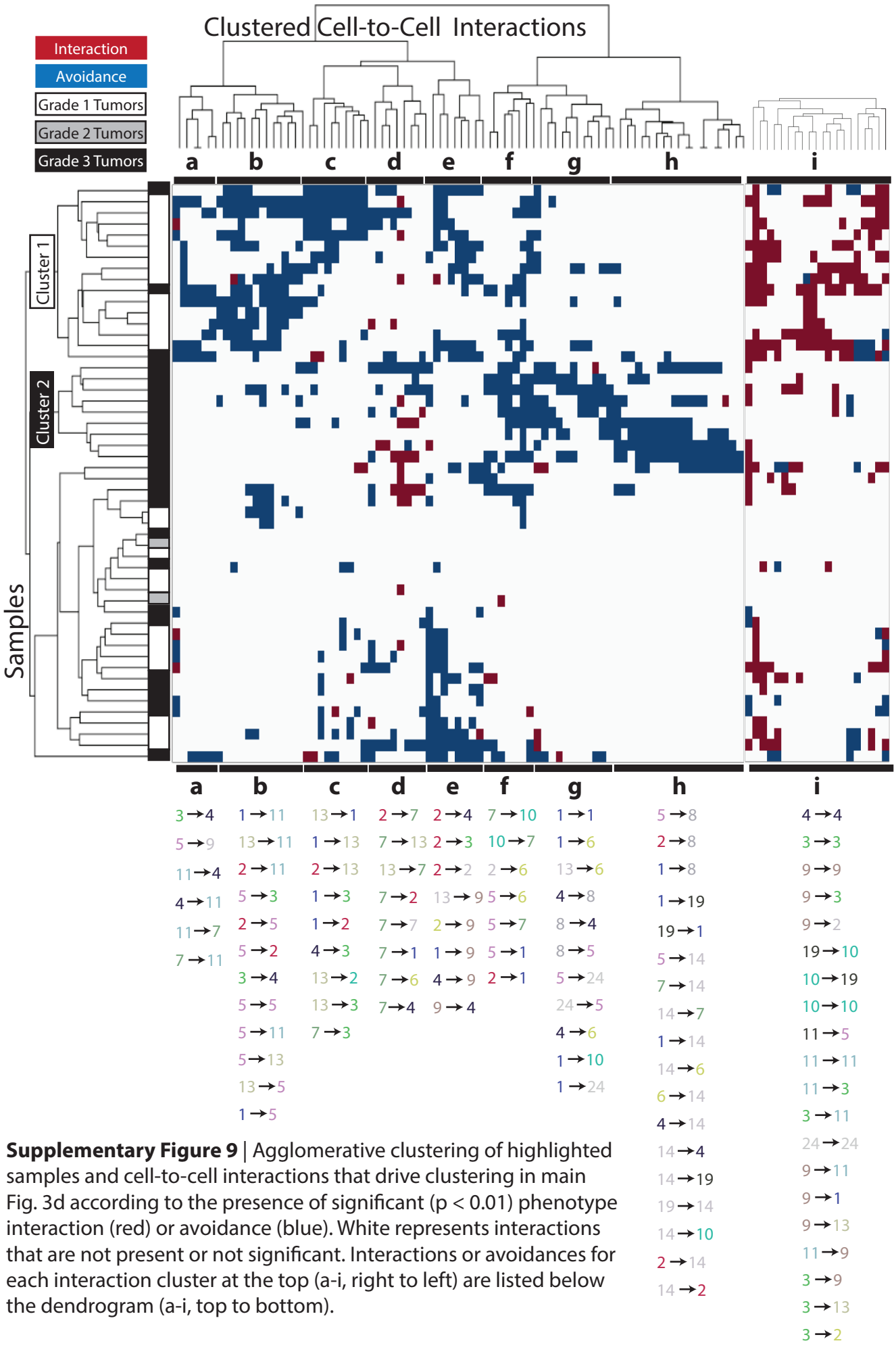

a
Image
Image + Mask
PhenoGraph Clusters
Cytokeratin 7
SMA
DNA
Phenotype 1
Phenotype 2
Phenotype 3
Phenotype 4
Phenotype 5
Phenotype 6
Phenotype 7
Phenotype 8
Phenotype 9
Phenotype 10
Phenotype 11
Phenotype 12
Phenotype 13
Phenotype 14



Supplementary Figure 7 | Neighbor Analysis of Healthy Mammary Tissue. (**a**) Raw Image, Image with single cell mask, and Phenograph Phenotype - labelled high-dimension IMC images of normal mammary tissue. Cytokeratin 7 (red), SMA (Smooth Muscle Actin, green), DNA Intercelator (blue). Scale bar 50μm. (**b**) Biaxial plots of single cell data from segmented health mammary tissue colored by source image. (**c**) tSNE plot of high-dimension single cell data colored by source image. (**d**) tSNE plot of high-dimension single cell data with the quantification of the marker of interest heatmapped for each cell. (**e**) Phenograph clustering of all cell phenotypes visualized as a distinct color on the tSNE plot. (**f**) Heatmap visualizing the median marker intensity for each Phenograph defined cell phenotype. (**g**) Heatmap of all neighboring cell interactions between all cell phenotypes in which square color indicates the prevalence of samples with significant interactions (permutation test: $p < 0.01$) in which the cell type in the row is significantly neighbored (red) or avoided (blue) by the cell type in the column. White represents a prevalence less than 10%.



Supplementary Figure 8 | Raw Image, Image with single cell mask, and phenoGraph phenotype for visualization of observed interactions. Scale bar 50µm. (**a**) Highlighted directional interaction from main Fig. 3c: highly interactive tumor cell phenotype #3 is surrounded by stromal cell phenotype #13 (main Fig. 3c row 3, column 13, red square), but phenotype #3 is not significantly enriched in the surroundings of phenotype #13 (main Fig. 3c row 13, column 3, blue square). Also, cluster 1 interactions depicted in Fig. 3e of self interacting phenotype #3 cells with phenotype #2 and #13 interactions. (**b**) Example from Neighbor Heatmap main Fig. 3c and Network main Fig. 3e: Directional interactions between phenotype #11 and #5. Phenotype #11 is surrounded by #5, while #11 is avoiding phenotype #5. (**c**) Avoidant proliferative cell phenotype #8 and interactive proliferative cell phenotypes #10 and #19. (**d**) Avoidance of active stroma phenotype #6 and hypoxic cell phenotype #14 identified in cluster #2 (main Fig. 3d,e; Supplementary Fig. 8) (**e**) $CD68^{+}$ cell phenotype #7 which has interactions identified in the Grade 3 enriched cluster 2 (main Fig. 3d,e).



Supplementary Figure 9 | Agglomerative clustering of highlighted samples and cell-to-cell interactions that drive clustering in main Fig. 3d according to the presence of significant ($p < 0.01$) phenotype interaction (red) or avoidance (blue). White represents interactions that are not present or not significant. Interactions or avoidances for each interaction cluster at the top (a-i, right to left) are listed below the dendrogram (a-i, top to bottom).

Metal Tag	Target	Clone	Marker	Vendor	Lot
La139	phosphoCreb	J151-21	Signaling	Becton Dickinson	558359
Pr141	GATA3	L50-823	Transcription Factor	Becton Dickinson	558686
Nd145	Twist	ABD29	Transcription Factor	Millipore	2039819
Nd146	CD68	KP1	Immune	E Biosciences	14-0688-82
Nd148	SMA	1A4	Stromal	Abcam	ab76549
Eu151	c-erbB-2, HER2	3B5	Receptor Tyrosine Kinase	Becton Dickinson	2223593
Gd158	Progesterone Receptor A/B	EP2	Hormone Receptor	Epitomics	N/A
Gd160	CD44	IM7	Adhesion	Becton Dickinson	550538
Dy161	EpCAM	9C4	Epithelial	Biolegend	B166689
Dy162	Vimentin	D21H3	Stromal	Cell Signaling	5741BF
Dy163	Fibronectin	10/Fibronectin	Stromal	Becton Dickinson	N/A
Dy164	Cytokeratin 7	RCK105	Epithelial	Becton Dickinson	550507
Ho165	β-Catenin	D13A1	Epithelial	Cell Signaling	8814BF
Er166	Carbonic Anhydrase IX	AF2188	Hypoxia	R&D Systems	AF2188
Er167	E-Cadherin	36/E-Cadherin	Epithelial	Becton Dickinson	610182
Er168	Ki-67	8D5	Proliferation	Cell Signaling	2
Er170	phosphoS6	D57.2.2E	Growth, Signaling	Cell Signaling	4858BF
Yb174	Cytokeratin 8/18	C51	Epithelial	Cell Signaling	2
Yb176	Histone H3	D1H2	Nuclei	Cell Signaling	6

Supplementary Table 1a | Antibodies and stains used to measure breast cancer samples.

Metal Tag	Target	Clone	Marker	Vendor	Lot
Pr141	Cytokeratin 5	EP1601Y	Basal	Abcam	GR233399-1
Nd143	Cytokeratin 19	Troma-III	Luminal	Dev Studies Hybridoma Bank	9/28/15
Nd144	Cytokeratin 8/18	C51	Luminal	Cell Signaling	2
Sm147	Cytokeratin 14	Polyclonal	Basal	Thermo Fischer	QI2091522
Nd148	SMA	1A4	Basal / Endothelial	Abcam	033M4768
Er167	E-Cadherin	36/E-Cadherin	Epithelial	Becton Dickinson	610182
Yb174	Cytokeratin 7	RCK105	Luminal	Becton Dickinson	5259670
Lu175	Pan-Cytokeratin	AE3	Epithelial	Millipore	2607604
Ir191/Ir193	DNA Intercelator		DNA	Fluidigm	

Supplementary Table 1b | Antibodies used to stain healthy breast tissue samples.

Supplementary Note 1 – miCAT Manual

miCAT – Getting started

1. Install miCAT

miCAT is automatically installed from the web when running the miCATAppInstaller.exe file. Windows users must have Visual Studio installed for features like PhenoGraph to function. If Visual Studio is not already installed on your computer download it from https://www.visualstudio.com/downloads/.

2. Open miCAT

Double click the icon generated during installation to open miCAT.

3. Data requirements

In order for raw image data to be successfully loaded into miCAT, all files associated with a specific image need to be stored in a separate, uniquely named folder (see example data set). Multiple folders, each containing the data corresponding to one image, can be loaded simultaneously. The different markers measured must be stored as individual (unstacked) tiff files in the image folder. miCAT requires that the tiff files be saved as unsigned 16-bit integers (uint16). If there is single-cell information, the image folder can also contain a mask, such as that generated after segmentation in CellProfiler, that identifies individual cells. The mask can be saved as a mat or tiff file in uint16 or int32 format.

4. Load samples into miCAT

After opening the miCAT GUI, the first step is to load the image data for subsequent analysis. Click on the “Load” button on the upper left of the interface. A drop-down menu with two options will appear. Choose “Load Samples” in order to start a new miCAT session (Figure 1).

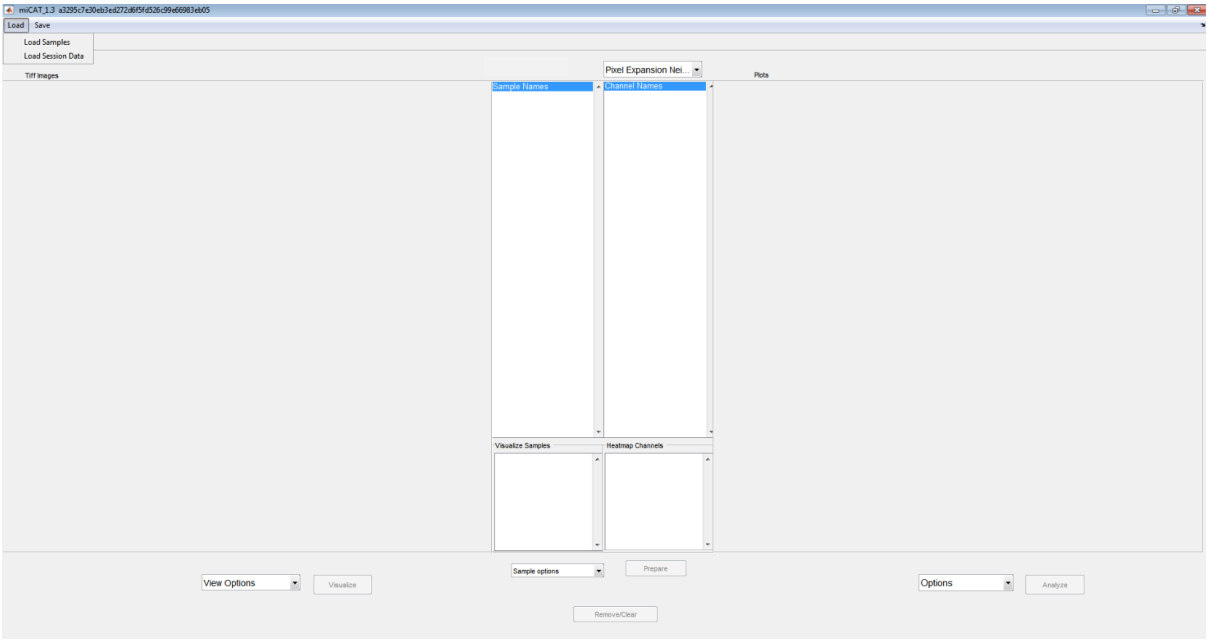


Figure 1

You will be prompted with a folder selection box (Figure 2). Navigate to the folders containing the images to be loaded. Click “Done” once you have added all the desired samples to the list on the right of the window (Figure 2).

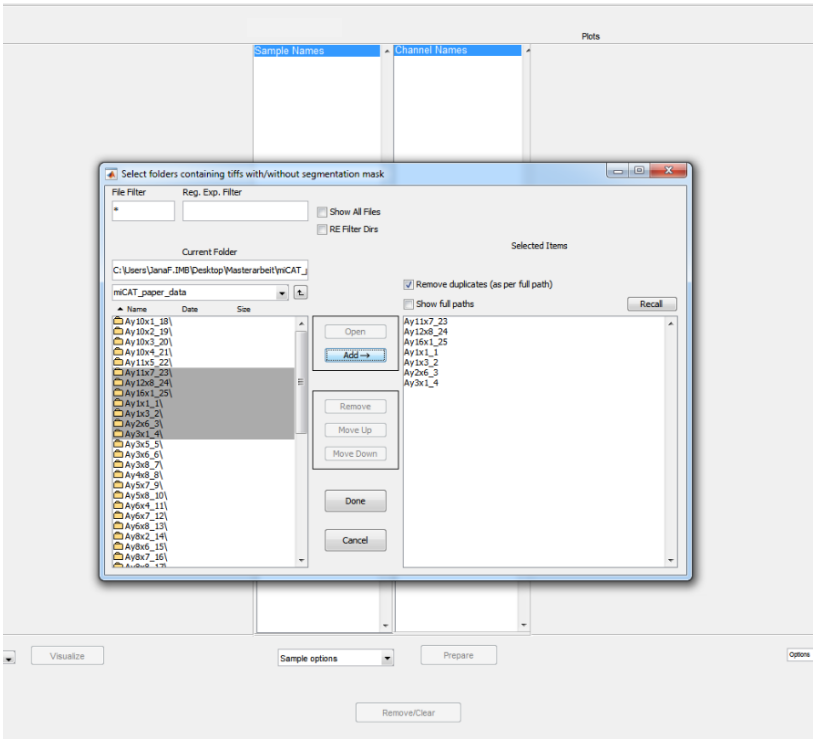


Figure 2

Once miCAT has finished loading the image data, you will be asked to specify the number of pixels to expand from each cell in order to look for neighboring cells (Figure 3). Choose an integer between two and ten. This pixel expansion can later be set to a different value. If you would like to arcsinh

transform your data, enter a suitable cofactor in the corresponding question box, otherwise just press cancel. It may take from a couple of seconds up to one minute per image (500x500 pixels) for the single-cell information to be updated. Once the folder selection prompt appears, browse to where you wish to store the “Custom Gates Folder”. This is where any files that are automatically generated during the miCAT session will be saved. Folder names should not include any spaces or special characters.

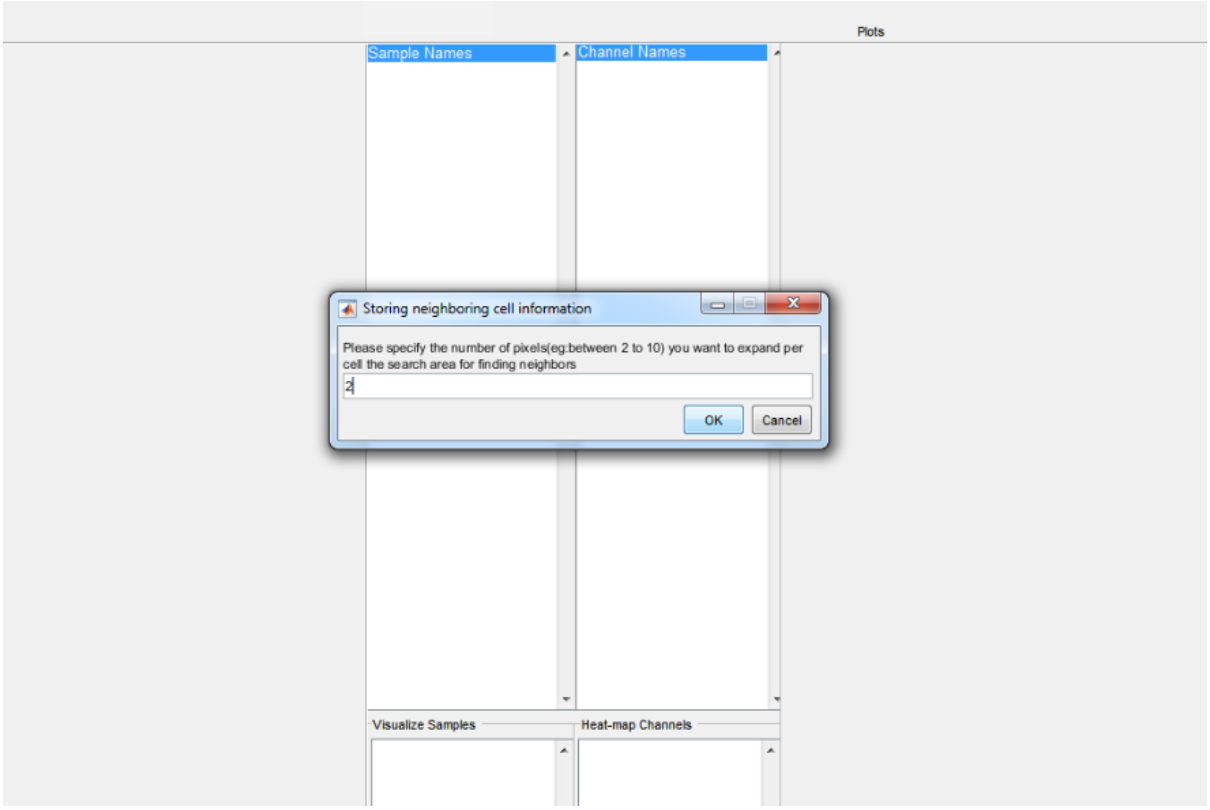


Figure 3

The loaded image data are displayed in two columns in the center of the interface (Figure 4). The box on the left hand side contains the “Samples”, which at this stage in the analysis process are just the individual images. The measured markers for the currently selected image are shown in the “Channels” box on the right. Note that miCAT also calculates image-specific cell identifiers, cell size and shape measurements, the percentage of a cell’s surface touching other cells, and the number of neighboring cells. Figure 4 displays the overall arrangement of the miCAT interface.

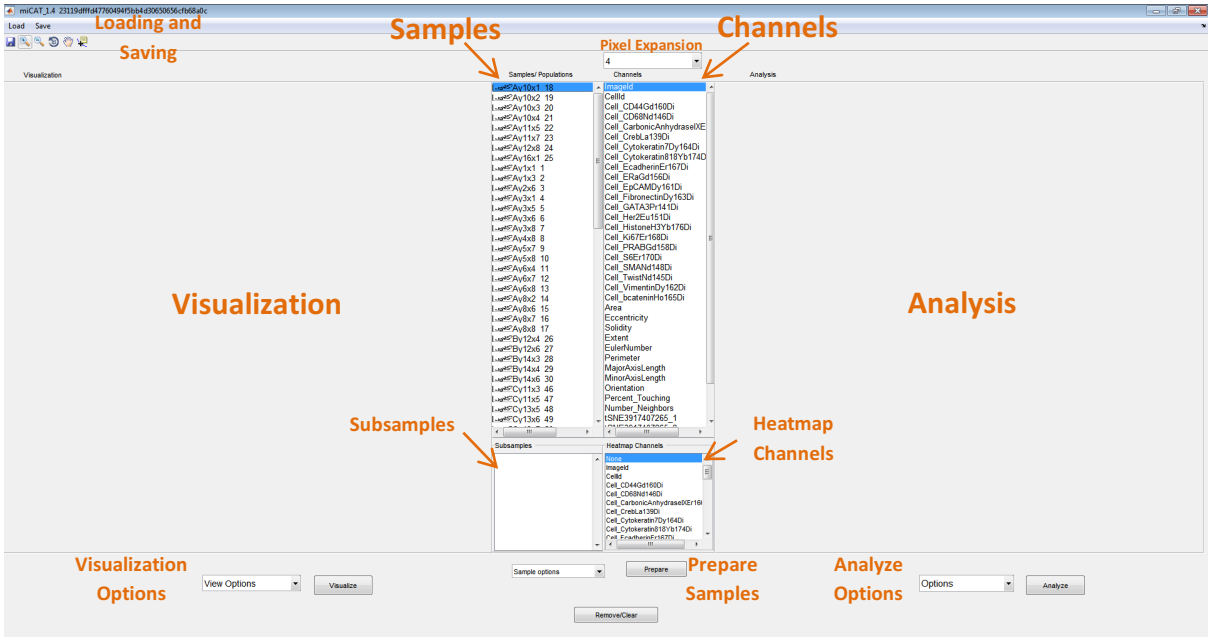


Figure 4

5. Save and reload session data

Once a miCAT analysis has been started, the full session can be saved by clicking the save icon (do not attempt to use the save button in the drop-down menu) on the upper left of the GUI (Figure 5). The session must be saved as a mat file in order to continue the analysis at a later time.

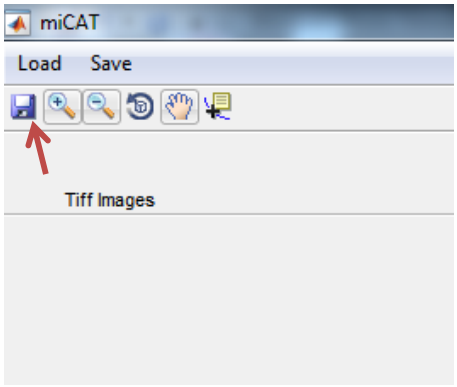


Figure 5

To reopen a previously saved session, choose the “Load Session Data” option in the drop-down menu of the “Load” button and navigate to the mat file containing the session of interest.

6. Visualize images

The left hand side of the GUI is dedicated to different kinds of visualizations of the selected images. The form of visualization can be determined and changed using the “View Options” drop-down

menu on the lower left (Figure 6). Having selected one of the options, the “Visualize” button must be pressed for changes to be applied. The following visualization options are supported:

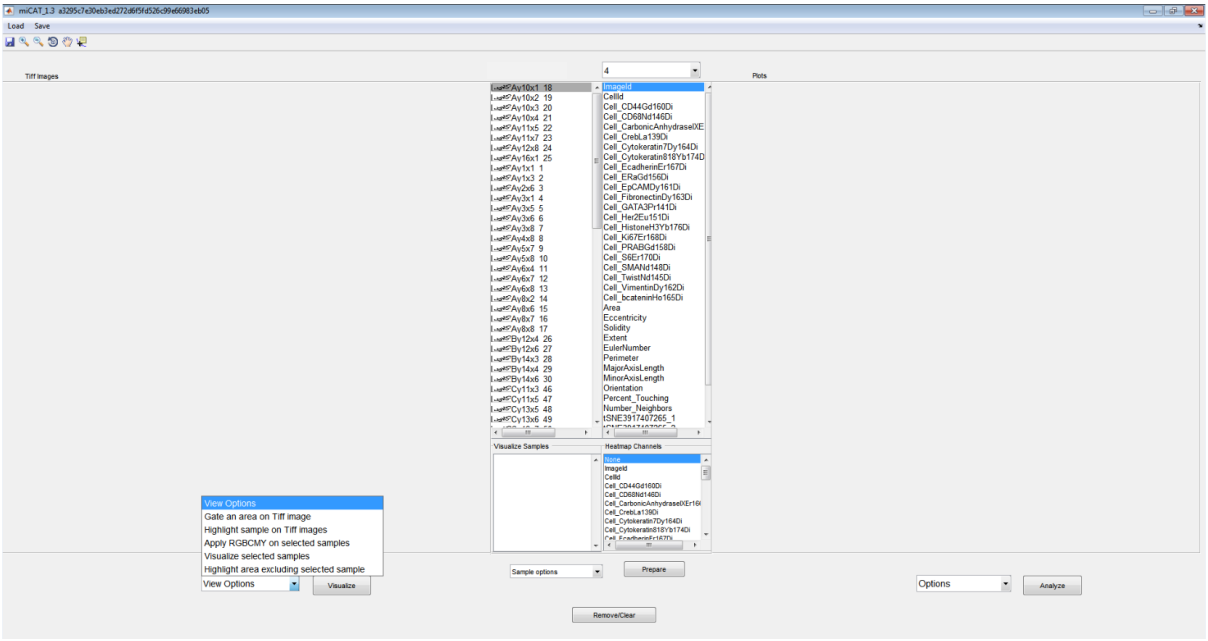


Figure 6

1. Choose “Apply RGBCMY on selected samples” in order to display the signal of the selected markers each in a different color and overlaid into one image (Figure 7a). To select multiple markers or images, please hold down the shift/CMD key. The tabs that appear above the visualization represent each of the selected images. Below the image a slider along with a checkbox for each color appears. Check a color and move the slider in order to increase the intensity of a specific signal. When the mask checkbox on the lower left is selected, the tiff image is overlaid with the segmentation mask, highlighting the individual cells (Figure 7b). The “Plot sample area XY” checkbox marks the centroid of each cell on the image.

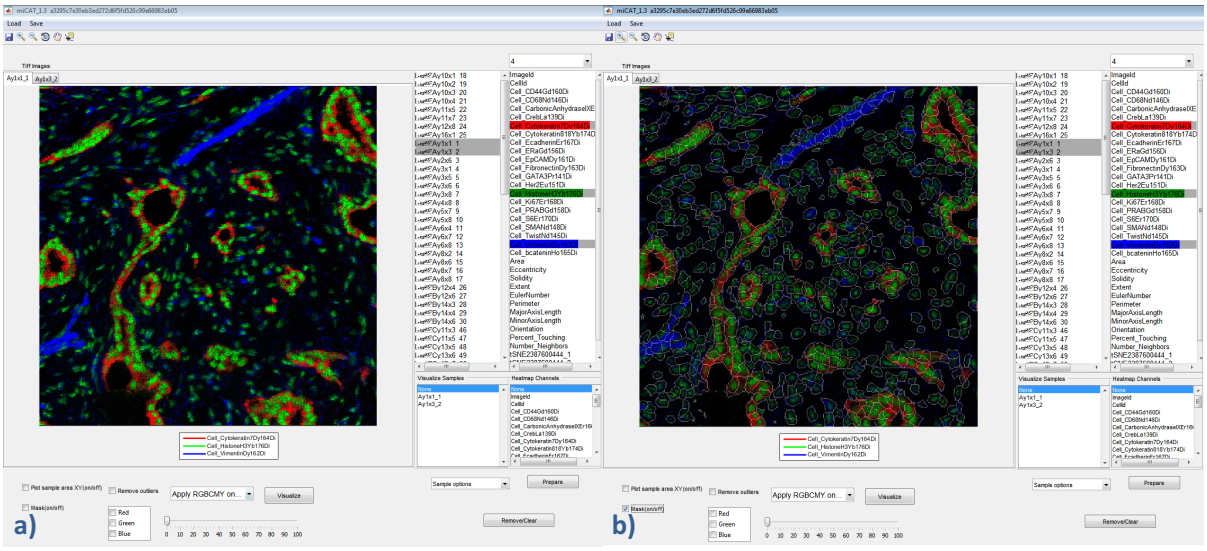


Figure 7

2. The “Heatmap channels on selected samples” option displays a heatmap of the selected channel (Figure 8a) overlaid on the single-cell mask (Figure 8b). In this case, the slider allows you to specify a “percentile cut-off” for very high values. This can be used to prevent the heatmap from being displayed as mostly blue if there are some extremely high values in very few cells. If the slider is set to one, the cut-off will be at the 99th percentile, meaning that the intensities above this value will be set to the value of the cut-off.

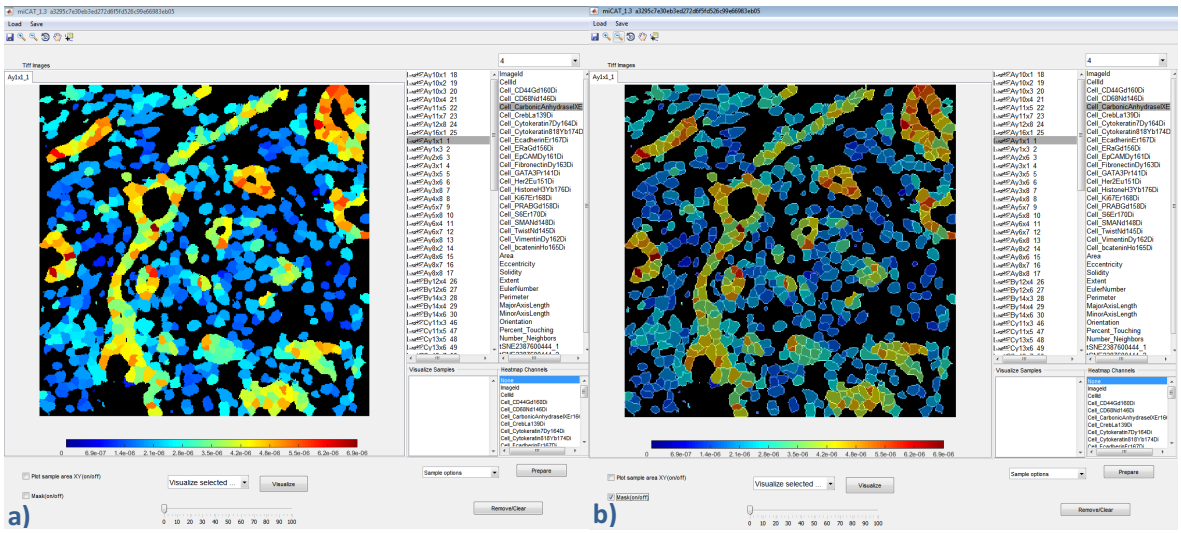


Figure 8

3. Cells of interest can be manually gated with the “Gate an area on tiff” option. Confirm the selection by pressing the “Visualization” button before gating. Afterwards, encircle a certain area with the cursor directly in the tiff image (Figure 9). Hold down the mouse key while selecting and release it when finished. These cells are then saved as a gate, which appears in the samples column bellow the loaded images.

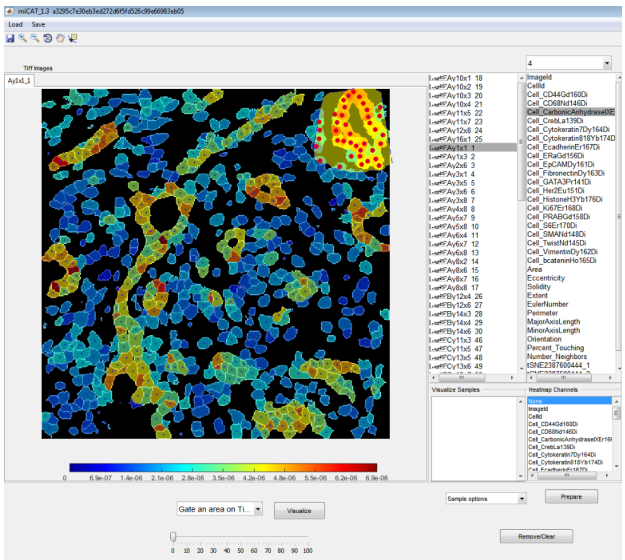


Figure 9

4. A subset of cells (such as a Phenograph cluster or manually gated cells) can also be highlighted on the tiff image using the “Highlight samples on tiff” option. If multiple subsets

are visualized simultaneously, they appear in different colors on the image. These colors are automatically selected such that they are most distinguishable from the colors in the background image (Figure 10).

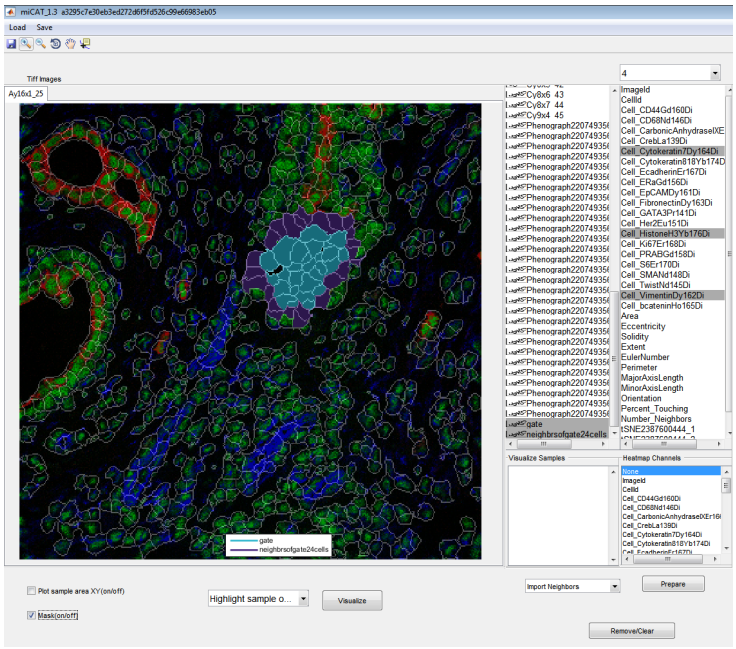


Figure 10

5. “Highlight excluding selected sample” leads to the exact opposite of “Highlight samples on tiff”. All cells except for the ones in the selected gates will be marked (Figure 11).

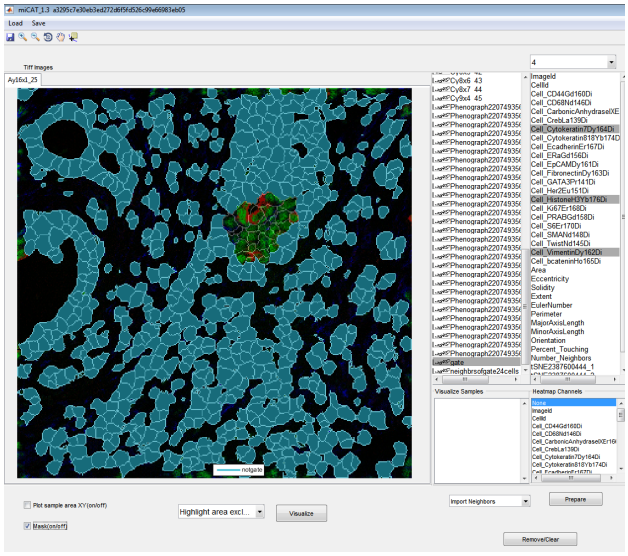


Figure 11

7. Analyze data

The right hand side of the interface is dedicated to data analysis and representation. Next to the “Analyze” button there are several options (Figure 12).

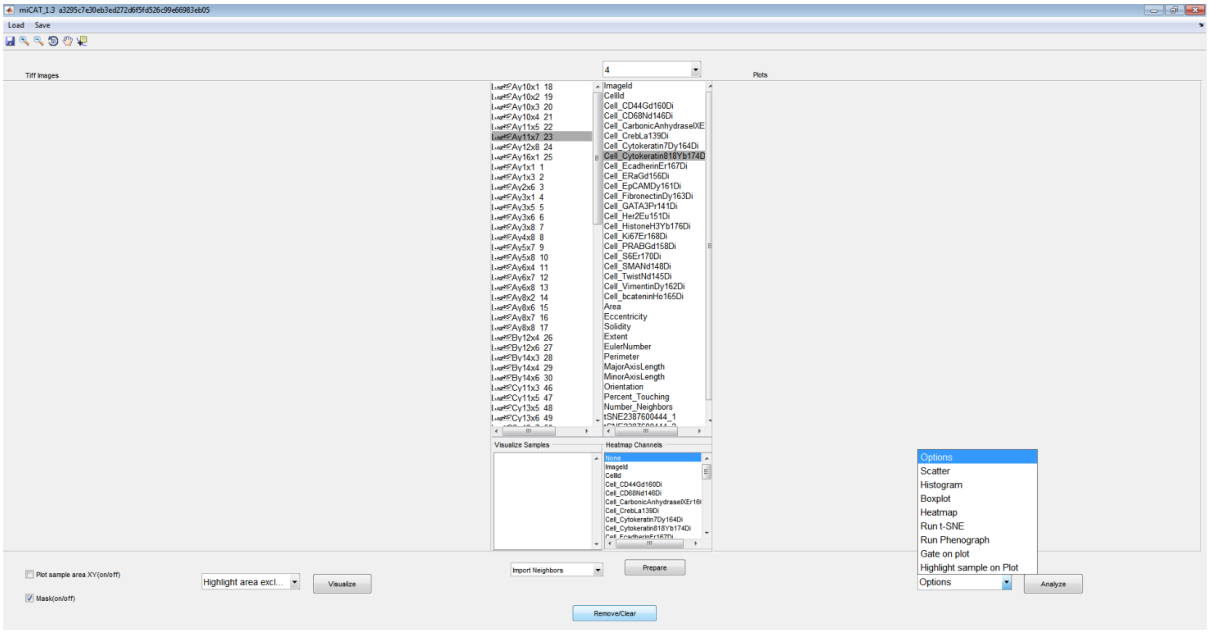


Figure 12

1. The “Scatter” option produces a scatterplot of up to three markers. If multiple images are selected, they are displayed in different colors (Figure 13a). From the channel list either two or three markers can be selected simultaneously to be displayed in a two- or three-dimensional plot. A fourth channel can be visualized in terms of color. Choose this additional channel from the “Heatmap Channels” box below the regular channels. This will overlay the dots in the scatterplot with a heatmap of the selected marker’s intensities (Figure 13b). In two-dimensional plots, a regression line can be added to the scatterplot by checking the “Regression line” box. Additionally, the R-value of Pearson’s correlation will be displayed.

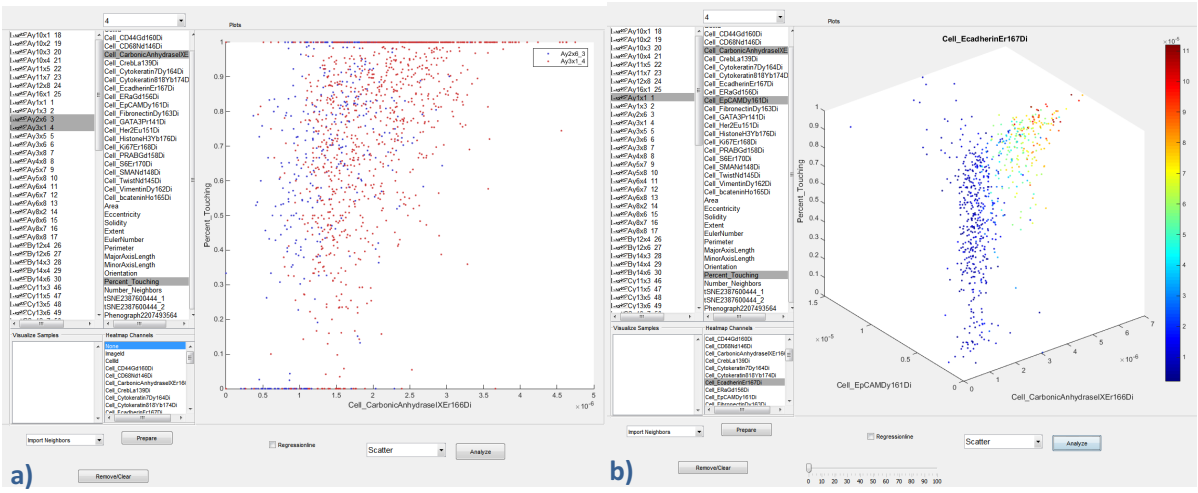


Figure 13

2. The “Histogram” option creates one histogram for each selected channel, displaying the lines for multiple selected images in different colors (Figure 14a). Alternatively, choose “Boxplot” if this representation is better suited for your purpose (Figure 14b). For optimal visualization, this option is best used with multiple images but not too many channels at the time.

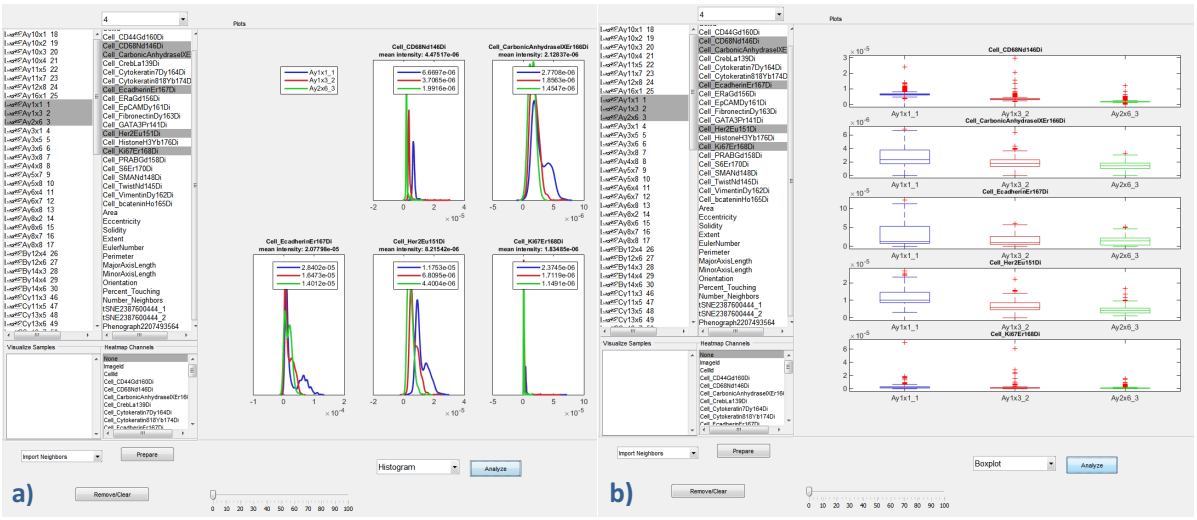


Figure 14

3. Generate a heatmap of the selected gates (y-axis) and channels (x-axis) by choosing the "Heatmap" option (Figure 15a). Checking the "b2r" checkbox will convert the regular heatmap to a more meaningful "anomaly heatmap", where white represents values close to zero and blue or red represent values below or above zero, respectively (Figure 15b). By default the heatmap displays mean values. These can be changed to medians by checking the "median instead of mean" checkbox.

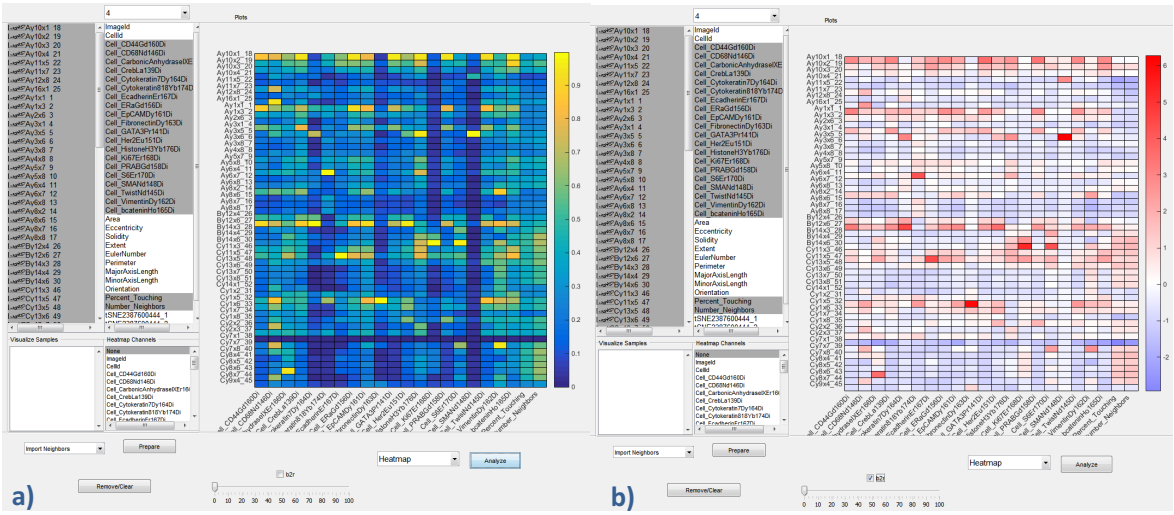


Figure 15

4. A t-SNE dimensionality reduction can be run on cells in the selected gates with the selected channels by choosing the "Run t-SNE" option. A seed is set internally to ensure reproducibility. The resulting cell coordinates necessary to display the t-SNE map are saved as two additional channels at the bottom of the list. Generate a scatterplot of these two channels with the "Scatter" option of the "Analyze Options" drop-down menu in order to visualize the t-SNE map (Figure 16a). If no channel from the "Heatmap Channels" box is selected, the colors on the plot simply represent the different selected images. If a channel is selected, the t-SNE map will be overlaid with the heatmap of the marker intensities from the selected channel. When overlaying cells with a heatmap, the "Percentile cut-off" slider

allows the intensities of the highest outliers to be set to the intensity value of a given percentile (Figure 16b).

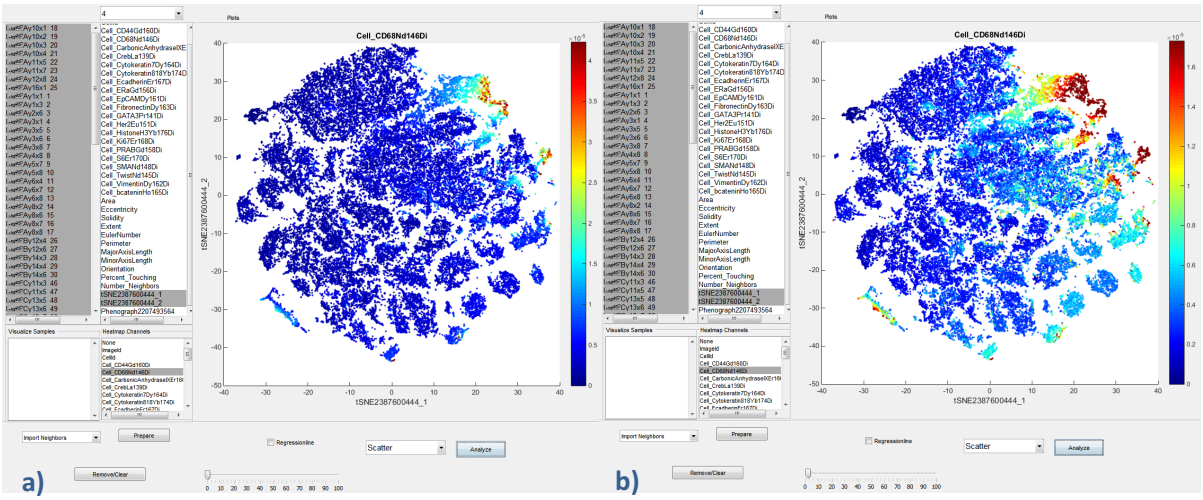


Figure 16

5. The “Phenograph” option will cluster the cells of the selected gates into Phenograph clusters based on the selected channels. Each resulting cluster will be saved as an individual gate and will appear below the rest of the samples. One possibility for visualization of the Phenograph clusters is to overlay the t-SNE map with the differently colored clusters by choosing the Phenograph result from the “Heatmap Channels” box (Figure 17).

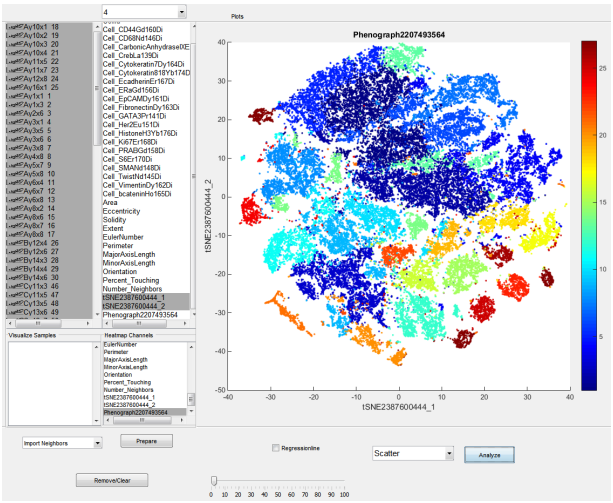


Figure 17

6. Similar to the manual gating on the tiff image, “Gate on plot” allows you to gate on certain cells of interest directly in the scatterplot (Figure 18). The cells in the gated area will be saved as a new gate with a user-specified name.

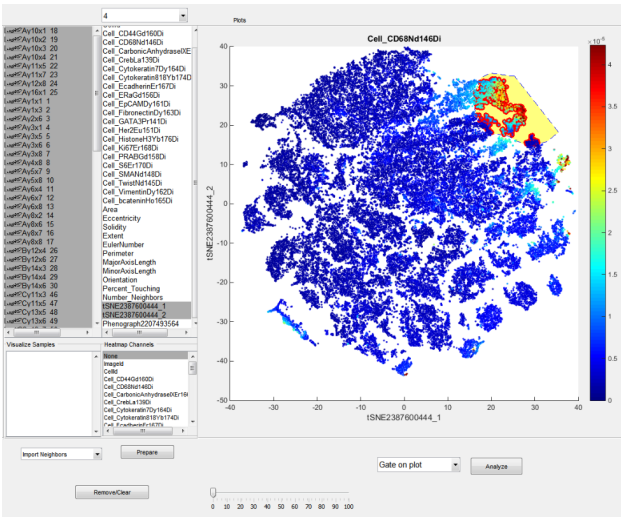


Figure 18

7. The “Highlight samples on plot” option outlines cells of the selected gates on the scatterplot or t-SNE map (Figure 19).

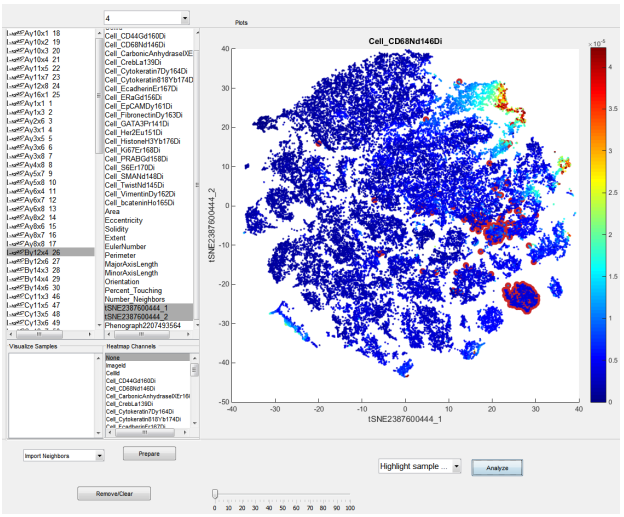


Figure 19

8. A principle component analysis can be run by selecting the “Run PCA” option. As for the t-SNE dimensionality reduction, the resulting coordinates necessary to display the first two components are saved as two new channels at the bottom of the list. Generate a scatterplot of these two channels with the “Scatter” option of the “Analyze Options” drop-down menu in order to visualize the first two principle components. If no channel from the “Heatmap Channels” box is selected, the colors on the plot simply represent the different selected images (Figure 20a). If a channel is selected, the scatter plot will be overlaid with the heatmap of the marker intensities from the selected channel (Figure 20b). When overlaying cells with a heatmap, the “Percentile cut-off” slider allows the intensities of the highest outliers to be set to the intensity value of a given percentile.

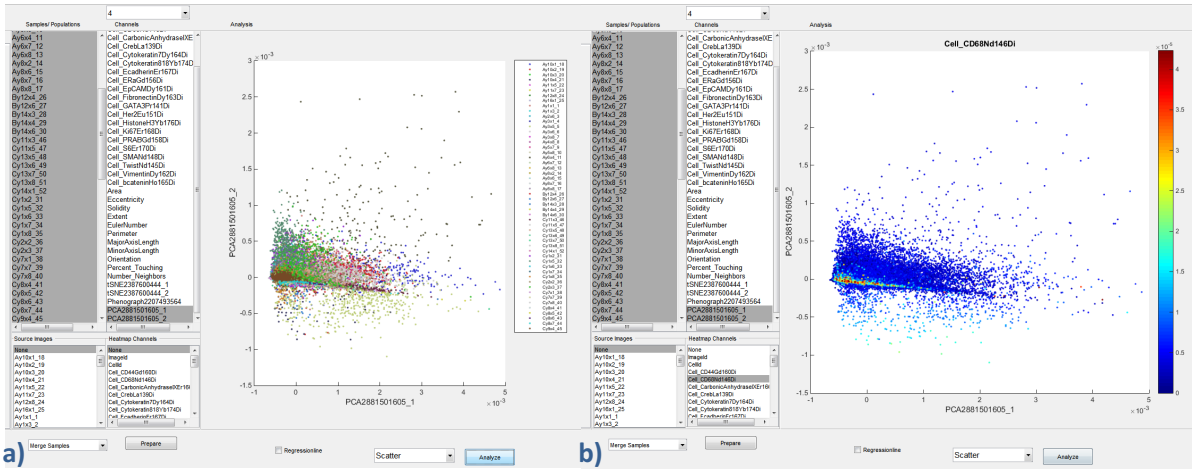


Figure 20

9. A k-means clustering can be run on any two-dimensional plot by selecting the “Run k-means” option. You will be asked to choose an amount of clusters and a number of iterations for the algorithm to run. The resulting cluster assignments for each cell are saved as a new channel at the bottom of the list. Scatterplots can then be overlaid with the color code corresponding to the k-means clusters (Figure 21).

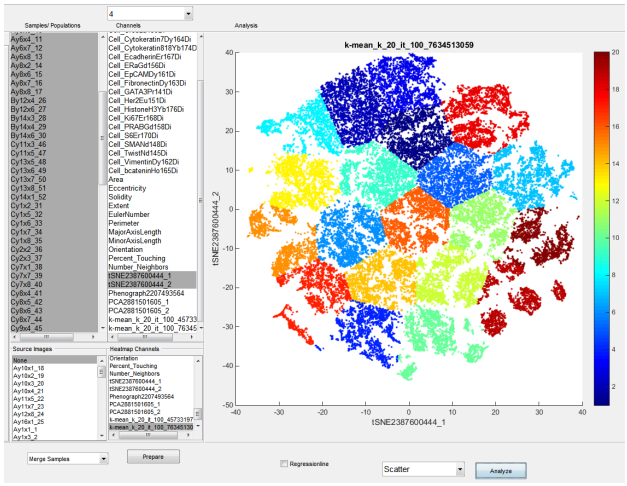


Figure 21

8. Prepare samples

Below the channel and gate selection boxes the “Sample Options” drop-down menu provides several options to prepare the loaded sample data or custom-made gates for further analysis. Click the “Prepare” button to apply the selected option (Figure 22).

1. “Import Neighbors” will create a new gate containing the neighbors of the cells in the currently selected gate.

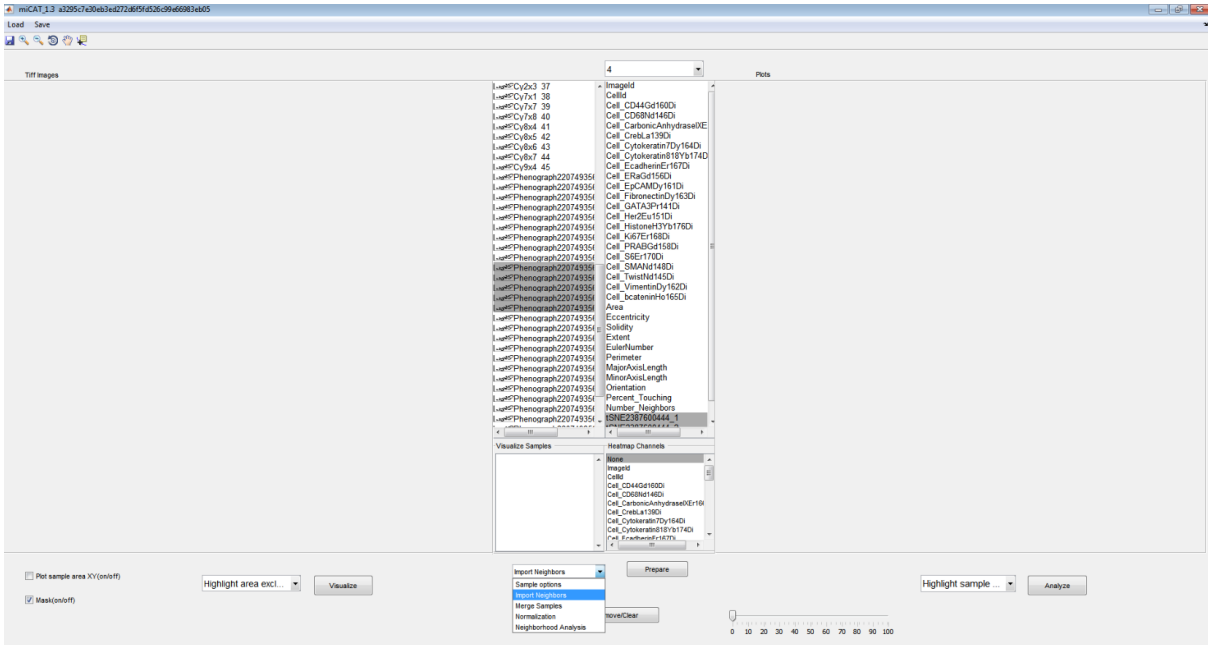


Figure 22

2. With “Merge Samples” you can pool multiple gates into one and give it a new name.
3. “Normalization” will take the Z-score of the selected marker and save it as a new channel at the bottom of the list.

9. Neighborhood analysis

The “Neighborhood Analysis” can be started from the “Prepare Samples” drop-down menu. Select the images (not the Phenograph clusters) to be considered from the gates box. This analysis can be run across all samples at once or across selected sample groups, and results can be compared.

You will be asked to specify the amount of permutations and the significance cut-off for the P-values. Furthermore you can choose a “special cluster” on which to focus in addition to displaying the results for all Phenograph clusters. This will generate an additional individual output. “Extra information” can be anything you know about the selected group of samples, such as “Grade1Tumors” or “allImages”. This information will appear in the file name of the individual figure generated during the neighborhood analysis. Additionally, the number of pixels to expand from each cell when searching for neighboring cells can be set. If this field is left empty, the analysis will simply run across all pixel expansions from one to six and return results for each version. Finally, you can specify the percentage (0-1) cut-off for present interactions to be displayed in the heatmaps.

The heatmap displays cluster-neighborhood frequencies across all images present in at least the percentage of the images (Figure 23a, 10% (0.1 default)). For example, if you look at the square on the diagonal corresponding to cluster five on both axes, you are looking at how often cells of cluster five neighbor each other. The squares are displayed in different intensities of red or blue or remain white. Red means the cells of these clusters neighbor each other more frequently than they would in random permutations of the cell cluster labels for each image. Blue means that the cluster cells neighbor each other less frequently than in images with randomly permuted cell labels. The same

color code applies to the clustergrams (Figure 23b). Here you see the cluster “interactions” on the x-axis displayed as, for example, “5_5” for the frequencies of the cells of cluster five neighboring each other. On the y-axis each of the images are listed, so that cluster combinations that significantly deviate from random can be identified in each image. A separate clustergram is generated showing only the interactions with the “special cluster”.

The results of the neighborhood analysis are automatically saved to the “Custom Gates” folder.

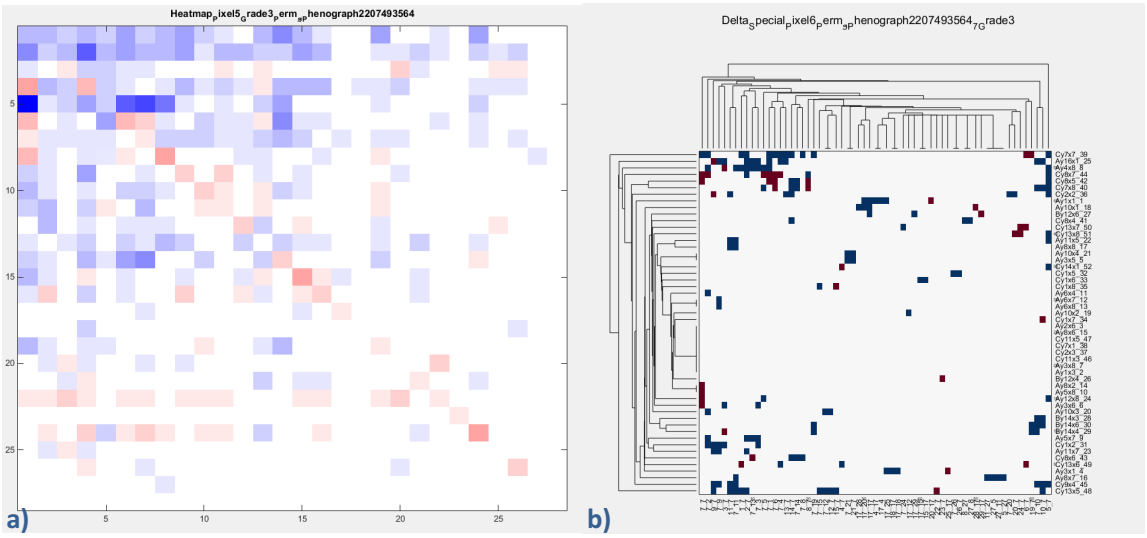


Figure 23

10. Save options

The “Save” button drop-down menu on the upper left yields additional saving options (Figure 24). Select “Save tiff figure” in order to save the currently visualized image on the left side of the interface. You will be prompted to specify in which folder and file format to save the image.

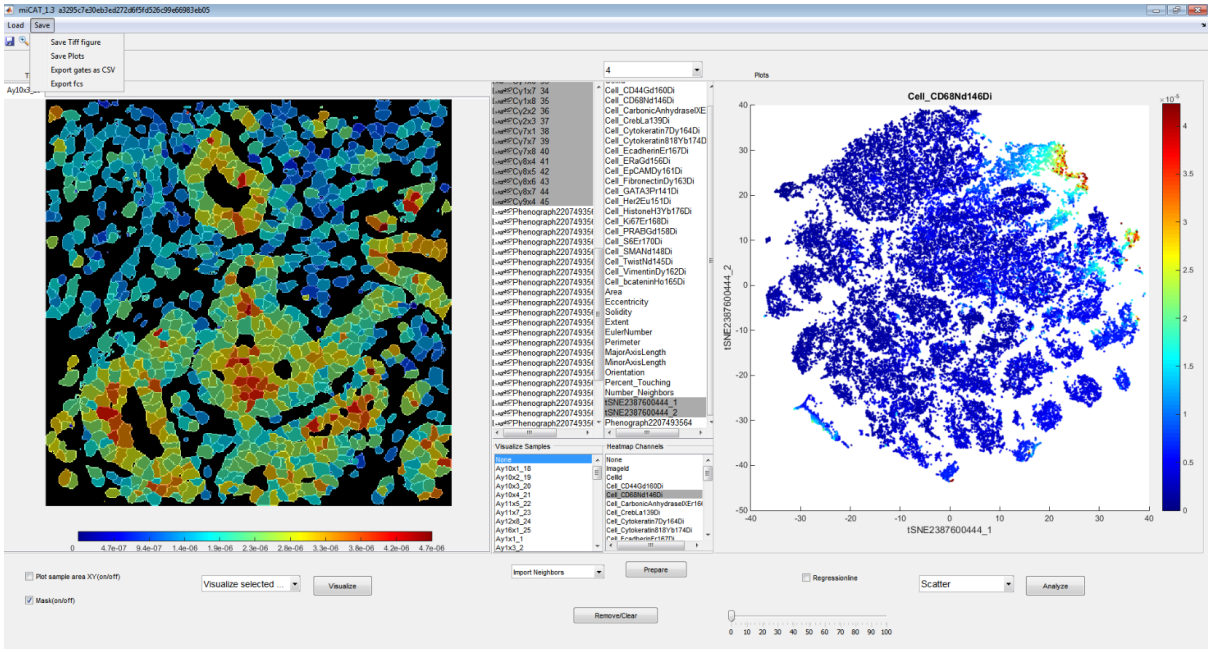


Figure 24

Similarly, the “Save plots” option will open the current plot displayed on the right half of the interface as a figure and prompt you to select a file folder in which to save it as an image (Figure 25).

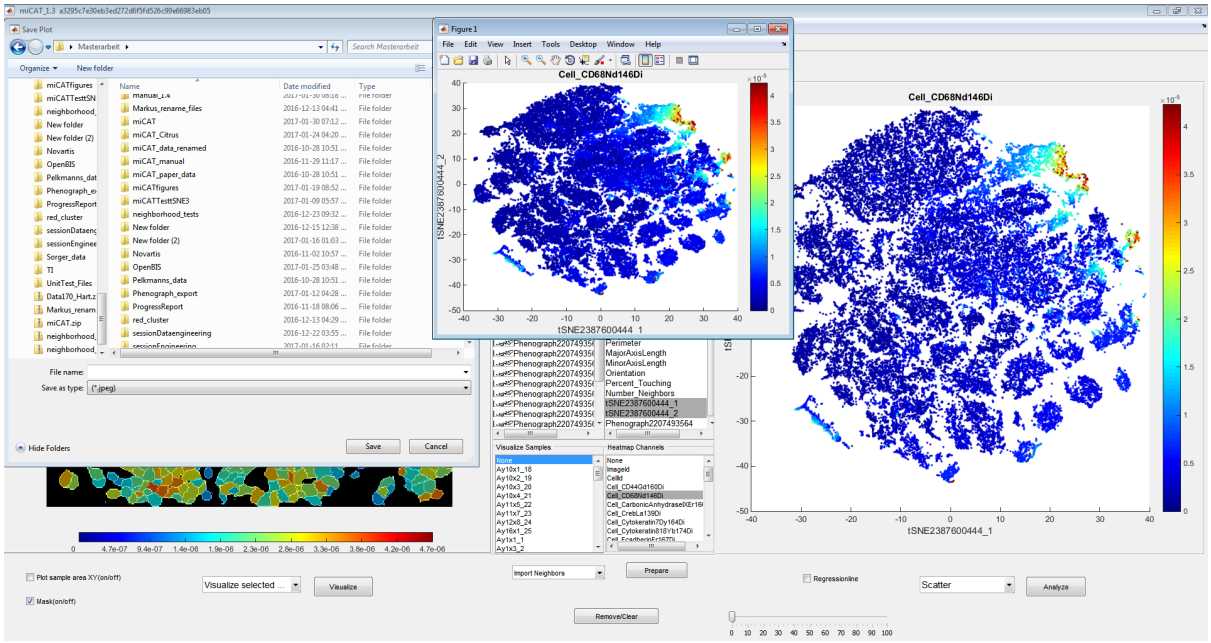
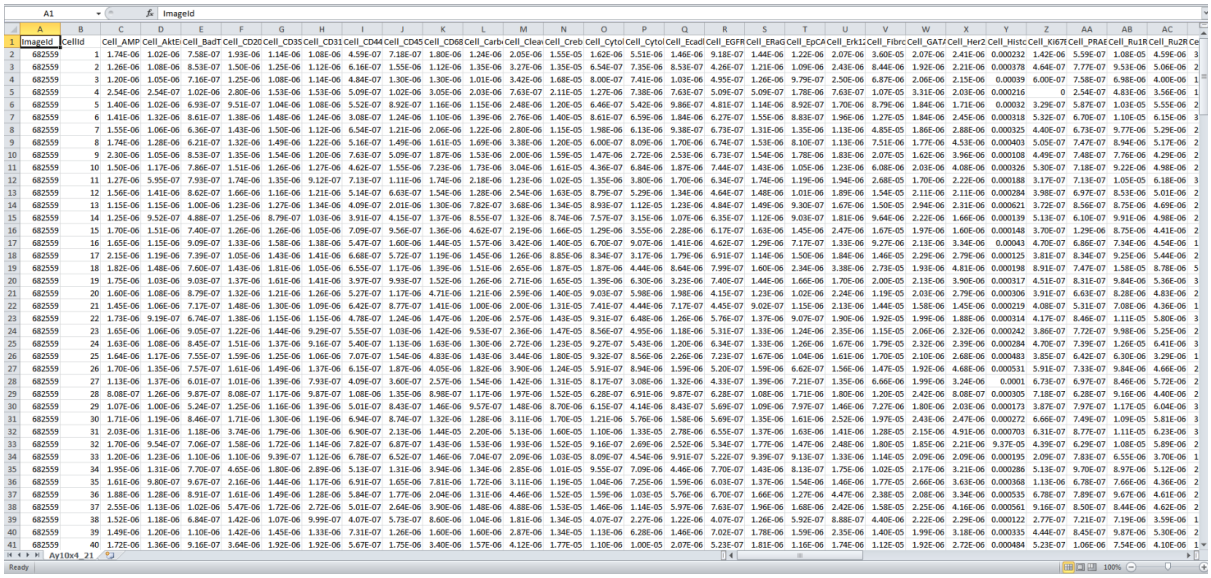


Figure 25

Click “Export gates as CSV” in order to save all of the information for the currently selected gates to your “Custom Gates” folder as CSV-files. These files contain the image and cell identifiers as the first two columns, followed by one column for each marker, each containing the measured intensities. After the marker columns the rest of the features (cell size, percent touching, number of neighbors, etc.) and their values are listed (Figure 26).



ImageId	CellId	Cell_AMP	Cell_AktE	Cell_BadT	Cell_CD20	Cell_CD35	Cell_CD31	Cell_CD44	Cell_CD45	Cell_CD68	Cell_Carb	Cell_Clea	Cell_Creb	Cell_Cyto	Cell_Cyto	Cell_Ecad	Cell_EGFR	Cell_ERaG	Cell_EpCA	Cell_Erk1	Cell_Fibrc	Cell_GAT4	Cell_Her2	Cell_Histc	Cell_KI67I	Cell_PRAE	Cell_Ru1R	Cell_Ru2R	Ce
682559	1	1.74E-06	1.02E-06	7.58E-07	1.93E-06	1.14E-06	1.08E-06	4.59E-07	7.18E-07	1.80E-06	1.24E-06	2.05E-06	1.55E-05	1.62E-06	5.51E-06	1.46E-06	9.18E-07	1.44E-06	1.22E-06	2.07E-06	3.60E-05	2.07E-06	2.41E-06	0.000232	1.42E-06	5.59E-07	1.08E-05	4.59E-06	3
682559	2	1.26E-06	1.08E-06	8.53E-07	1.50E-06	1.25E-06	1.12E-06	6.16E-07	1.55E-06	1.12E-06	1.35E-06	3.27E-06	1.35E-05	6.54E-07	7.35E-06	8.53E-07	4.26E-07	1.21E-06	1.09E-06	2.43E-06	8.44E-06	1.92E-06	2.21E-06	0.000378	4.64E-07	7.77E-07	9.53E-06	5.06E-06	2
682559	3	1.20E-06	1.05E-06	7.16E-07	1.25E-06	1.08E-06	1.14E-06	4.84E-07	1.30E-06	1.30E-06	1.01E-06	3.42E-06	1.68E-05	8.00E-07	7.41E-06	1.03E-06	4.95E-07	1.26E-06	9.79E-07	2.50E-06	6.87E-06	2.06E-06	2.15E-06	0.00039	6.00E-07	7.58E-07	6.98E-06	4.00E-06	1
682559	4	2.54E-06	2.54E-07	1.02E-06	2.80E-06	1.53E-06	1.53E-06	5.09E-07	1.02E-06	3.05E-06	2.03E-06	7.63E-07	2.11E-05	1.27E-06	7.38E-06	7.63E-07	5.09E-07	5.09E-07	1.78E-06	7.63E-07	1.07E-05	3.31E-06	2.03E-06	0.000216	0	2.54E-07	4.83E-06	3.56E-06	1
682559	5	1.40E-06	1.02E-06	6.93E-07	9.51E-07	1.04E-06	1.08E-06	5.52E-07	8.92E-07	1.16E-06	1.15E-06	2.48E-06	1.20E-05	6.46E-07	5.42E-06	9.86E-07	4.81E-07	1.14E-06	8.92E-07	1.70E-06	8.79E-06	1.84E-06	1.71E-06	0.00032	3.29E-07	5.87E-07	1.03E-05	5.55E-06	2
682559	6	1.41E-06	1.32E-06	8.61E-07	1.38E-06	1.48E-06	1.24E-06	3.08E-07	1.24E-06	1.10E-06	1.39E-06	2.76E-06	1.40E-05	8.61E-07	6.59E-06	1.84E-06	6.27E-07	1.55E-06	8.83E-07	1.96E-06	1.27E-05	1.84E-06	2.45E-06	0.000318	5.32E-07	6.70E-07	1.10E-05	6.15E-06	3
682559	7	1.55E-06	1.06E-06	6.36E-07	1.43E-06	1.50E-06	1.12E-06	6.54E-07	1.21E-06	2.06E-06	1.22E-06	2.80E-06	1.15E-05	1.98E-06	6.13E-06	9.38E-07	6.73E-07	1.31E-06	1.35E-06	1.13E-06	4.85E-05	1.86E-06	2.88E-06	0.000325	4.40E-07	6.73E-07	9.77E-06	5.29E-06	2
682559	8	1.74E-06	1.28E-06	6.21E-07	1.32E-06	1.49E-06	1.22E-06	5.16E-07	1.49E-06	1.61E-05	1.69E-06	3.38E-06	1.20E-05	6.00E-07	8.09E-06	1.70E-06	6.74E-07	1.53E-06	8.10E-07	1.13E-06	7.51E-06	1.77E-06	4.53E-06	0.000403	5.05E-07	7.47E-07	8.94E-06	5.17E-06	2
682559	9	2.30E-06	1.05E-06	8.53E-07	1.35E-06	1.54E-06	1.20E-06	7.63E-07	5.09E-07	1.87E-06	1.53E-06	2.00E-06	1.59E-05	1.47E-06	2.72E-06	2.53E-06	6.73E-07	1.54E-06	1.78E-06	1.83E-06	2.07E-05	1.62E-06	3.96E-06	0.000108	4.49E-07	7.48E-07	7.76E-06	4.29E-06	2
682559	10	1.50E-06	1.17E-06	7.86E-07	1.51E-06	1.26E-06	1.27E-06	4.62E-07	1.55E-06	7.23E-06	1.73E-06	3.04E-06	1.61E-05	4.36E-07	6.84E-06	1.87E-06	7.44E-07	1.43E-06	1.05E-06	1.23E-06	6.08E-06	2.03E-06	4.08E-06	0.000326	5.30E-07	7.18E-07	9.22E-06	4.98E-06	2
682559	11	1.27E-06	5.95E-07	7.93E-07	1.74E-06	1.35E-06	9.12E-07	7.13E-07	1.11E-06	1.74E-06	2.18E-06	1.23E-06	1.02E-05	1.35E-06	3.80E-06	1.70E-06	6.34E-07	1.74E-06	1.19E-06	1.94E-06	2.68E-05	1.70E-06	2.22E-06	0.000188	3.17E-07	7.13E-07	1.05E-05	6.18E-06	3
682559	12	1.56E-06	1.41E-06	8.62E-07	1.66E-06	1.16E-06	1.21E-06	5.14E-07	6.63E-07	1.54E-06	1.28E-06	2.54E-06	1.63E-05	8.79E-07	5.29E-06	1.34E-06	4.64E-07	1.48E-06	1.01E-06	1.89E-06	1.54E-05	2.11E-06	2.11E-06	0.000284	3.98E-07	6.97E-07	8.53E-06	5.01E-06	2
682559	13	1.15E-06	1.15E-06	1.00E-06	1.23E-06	1.27E-06	1.34E-06	4.09E-07	2.01E-06	1.30E-06	7.82E-07	3.68E-06	1.34E-05	8.93E-07	1.12E-05	1.23E-06	4.84E-07	1.49E-06	9.30E-07	1.67E-06	1.50E-05	2.94E-06	2.31E-06	0.000621	3.72E-07	8.56E-07	8.75E-06	4.69E-06	2
682559	14	1.25E-06	9.52E-07	4.88E-07	1.25E-06	8.79E-07	1.03E-06	3.91E-07	4.15E-07	1.37E-06	8.55E-07	1.32E-06	8.74E-06	7.57E-07	3.15E-06	1.07E-06	6.35E-07	1.12E-06	9.03E-07	1.81E-06	9.64E-06	2.22E-06	1.66E-06	0.000139	5.13E-07	6.10E-07	9.91E-06	4.98E-06	2
682559	15	1.70E-06	1.51E-06	7.40E-07	1.26E-06	1.26E-06	1.05E-06	7.09E-07	9.56E-07	1.36E-06	4.62E-07	2.19E-06	1.66E-05	1.29E-06	3.55E-06	2.28E-06	6.17E-07	1.63E-06	1.45E-06	2.47E-06	1.67E-05	1.97E-06	1.60E-06	0.000148	3.70E-07	1.29E-06	8.75E-06	4.41E-06	2
682559	16	1.65E-06	1.15E-06	9.09E-07	1.33E-06	1.58E-06	1.38E-06	5.47E-07	1.60E-06	1.44E-05	1.57E-06	3.42E-06	1.40E-05	6.70E-07	9.07E-06	1.41E-06	4.62E-07	1.29E-06	7.17E-07	1.33E-06	9.27E-06	2.13E-06	3.34E-06	0.00043	4.70E-07	6.86E-07	7.34E-06	4.54E-06	1
682559	17	2.15E-06	1.19E-06	7.39E-07	1.05E-06	1.43E-06	1.41E-06	6.68E-07	5.72E-07	1.19E-06	1.45E-06	1.26E-06	8.85E-06	8.34E-07	3.17E-06	1.79E-06	6.91E-07	1.14E-06	1.50E-06	1.84E-06	1.46E-05	2.29E-06	2.79E-06	0.000125	3.81E-07	8.34E-07	9.25E-06	5.44E-06	2
682559	18	1.82E-06	1.48E-06	7.60E-07	1.43E-06	1.81E-06	1.05E-06	6.55E-07	1.17E-06	1.39E-06	1.51E-06	2.65E-06	1.87E-05	1.87E-06	4.44E-06	8.64E-06	7.99E-07	1.60E-06	2.34E-06	3.38E-06	2.73E-05	1.93E-06	4.81E-06	0.000198	8.91E-07	7.47E-07	1.58E-05	8.78E-06	5
682559	19	1.75E-06	1.03E-06	9.03E-07	1.37E-06	1.61E-06	1.41E-06	3.97E-07	9.93E-07	1.52E-06	1.26E-06	2.71E-06	1.65E-05	1.39E-06	6.30E-06	3.23E-06	7.40E-07	1.44E-06	1.66E-06	1.70E-06	2.00E-05	2.13E-06	3.90E-06	0.000317	4.51E-07	8.31E-07	9.84E-06	5.36E-06	3
682559	20	1.60E-06	1.08E-06	8.79E-07	1.32E-06	1.21E-06	1.26E-06	5.27E-07	1.17E-06	4.71E-06	1.21E-06	2.59E-06	1.40E-05	9.03E-07	5.98E-06	1.98E-06	4.15E-07	1.23E-06	1.02E-06	2.24E-06	1.19E-05	2.03E-06	2.79E-06	0.000306	3.91E-07	6.63E-07	8.28E-06	4.83E-06	2
682559	21	1.45E-06	1.06E-06	7.17E-07	1.48E-06	1.30E-06	1.09E-06	6.42E-07	8.77E-07	1.41E-06	1.00E-06	2.00E-06	1.31E-05	7.41E-07	4.44E-06	7.17E-07	4.45E-07	9.02E-07	1.15E-06	2.13E-06	1.44E-05	1.58E-06	1.45E-06	0.000219	4.08E-07	5.31E-07	7.08E-06	4.36E-06	1
682559	22	1.73E-06	9.19E-07	6.74E-07	1.38E-06	1.15E-06	1.15E-06	4.78E-07	1.24E-06	1.47E-06	1.20E-06	2.57E-06	1.43E-05	9.31E-07	6.48E-06	1.26E-06	5.76E-07	1.37E-06	9.07E-07	1.90E-06	1.92E-05	1.99E-06	1.88E-06	0.000314	4.17E-07	8.46E-07	1.11E-05	5.80E-06	3
682559	23	1.65E-06	1.06E-06	9.05E-07	1.22E-06	1.44E-06	9.29E-07	5.55E-07	1.03E-06	1.42E-06	9.53E-07	2.36E-06	1.47E-05	8.56E-07	4.95E-06	1.18E-06	5.31E-07	1.33E-06	1.24E-06	2.35E-06	1.15E-05	2.06E-06	2.32E-06	0.000242	3.86E-07	7.72E-07	9.98E-06	5.25E-06	2
682559	24	1.63E-06	1.08E-06	8.45E-07	1.51E-06	1.37E-06	9.16E-07	5.40E-07	1.13E-06	1.63E-06	1.30E-06	2.72E-06	1.23E-05	9.27E-07	5.43E-06	1.20E-06	6.34E-07	1.33E-06	1.26E-06	1.67E-06	1.79E-05	2.32E-06	2.39E-06	0.000284	4.70E-07	7.39E-07	1.26E-05	6.41E-06	3
682559	25	1.64E-06	1.17E-06	7.55E-07	1.59E-06	1.25E-06	1.06E-06	7.07E-07	1.54E-06	4.83E-06	1.43E-06	3.44E-06	1.80E-05	9.32E-07	8.56E-06	2.26E-06	7.23E-07	1.67E-06	1.04E-06	1.61E-06	1.70E-05	2.10E-06	2.68E-06	0.000483	3.85E-07	6.42E-07	6.30E-06	3.29E-06	1
682559	26	1.70E-06	1.35E-06	7.57E-07	1.61E-06	1.49E-06	1.37E-06	6.15E-07	1.87E-06	4.05E-06	1.82E-06	3.90E-06	1.24E-05	5.91E-07	8.94E-06	1.59E-06	5.20E-07	1.59E-06	6.62E-07	1.56E-06	1.47E-05	1.92E-06	4.68E-06	0.000531	5.91E-07	7.33E-07	9.84E-06	4.66E-06	2
682559	27	1.13E-06	1.37E-06	6.01E-07	1.01E-06	1.39E-06	7.93E-07	4.09E-07	3.60E-07	2.57E-06	1.54E-06	1.42E-06	1.31E-05	8.17E-07	3.08E-06	1.32E-06	4.33E-07	1.39E-06	7.21E-07	1.35E-06	6.66E-06	1.99E-06	3.24E-06	0.0001	6.73E-07	6.97E-07	8.46E-06	5.72E-06	2
682559	28	8.08E-07	1.26E-06	9.87E-07	8.08E-07	1.17E-06	9.87E-07	1.08E-06	1.35E-06	8.98E-07	1.17E-06	1.97E-06	1.52E-05	6.28E-07	6.91E-06	9.87E-07	6.28E-07	1.08E-06	1.71E-06	1.80E-06	1.20E-05	2.42E-06	8.08E-07	0.000305	7.18E-07	6.28E-07	9.16E-06	4.40E-06	2
682559	29	1.07E-06	1.00E-06	5.24E-07	1.25E-06	1.16E-06	1.39E-06	5.01E-07	8.43E-07	1.46E-06	9.57E-07	1.48E-06	8.70E-06	6.15E-07	4.14E-06	8.43E-07	5.69E-07	1.09E-06	7.97E-07	1.46E-06	7.27E-06	1.80E-06	2.03E-06	0.000173	3.87E-07	7.97E-07	1.17E-05	6.04E-06	3
682559	30	1.71E-06	1.19E-06	8.46E-07	1.71E-06	1.30E-06	1.19E-06	6.94E-07	8.74E-07	1.32E-06	1.28E-06	3.11E-06	1.70E-05	1.21E-06	5.76E-06	1.58E-06	5.69E-07	1.35E-06	1.61E-06	2.52E-06	1.97E-05	2.43E-06	2.47E-06	0.000272	6.66E-07	7.49E-07	1.09E-05	5.81E-06	3
682559	31	2.03E-06	1.31E-06	1.18E-06	3.74E-06	1.79E-06	1.30E-06	6.90E-07	2.13E-06	1.44E-05	2.20E-06	5.13E-06	1.60E-05	1.10E-06	1.33E-05	2.78E-06	6.55E-07	1.37E-06	1.63E-06	1.41E-06	1.28E-05	2.15E-06	4.91E-06	0.000703	6.31E-07	8.77E-07	1.11E-05	6.23E-06	3
682559	32	1.70E-06	9.54E-07	7.06E-07	1.58E-06	1.72E-06	1.14E-06	7.82E-07	6.87E-07	1.43E-06	1.53E-06	1.93E-06	1.52E-05	9.16E-07	2.69E-06	2.52E-06	5.34E-07	1.77E-06	1.47E-06	2.48E-06	1.80E-05	1.85E-06	2.21E-06	9.37E-05	4.39E-07	6.29E-07	1.08E-05	5.89E-06	2
682559	33	1.20E-06	1.23E-06	1.10E-06	1.10E-06	9.39E-07	1.12E-06	6.78E-07	6.52E-07	1.46E-06	7.04E-07	2.09E-06	1.03E-05	8.09E-07	4.54E-06	9.91E-07	5.22E-07	9.39E-07	9.13E-07	1.33E-06	1.14E-05	2.09E-06	2.09E-06	0.000195	2.09E-07	7.83E-07	6.55E-06	3.70E-06	1
682559	34	1.95E-06	1.31E-06	7.70E-07	4.65E-06	1.80E-06	2.89E-06	5.13E-07	1.31E-06	3.94E-06	1.34E-06	2.85E-06	1.01E-05	9.55E-07	7.09E-06	4.46E-06	7.70E-07	1.43E-06	8.13E-07	1.75E-06	1.02E-05	2.17E-06	3.21E-06	0.000286	5.13E-07	9.70E-07	8.97E-06	5.12E-06	2
682559	35	1.61E-06	9.80E-07	9.67E-07	2.16E-06	1.44E-06	1.17E-06	6.91E-07	1.65E-06	7.81E-06	1.72E-06	3.11E-06	1.19E-05	1.04E-06	7.25E-06	1.59E-06	6.03E-07	1.37E-06	1.54E-06	1.46E-06	1.77E-05	2.66E-06	3.63E-06	0.000368	1.13E-06	6.78E-07	7.66E-06	4.36E-06	2
682559	36	1.88E-06	1.28E-06	8.91E-07	1.61E-06	1.49E-06	1.28E-06	5.84E-07	1.77E-06	2.04E-06	1.31E-06	4.46E-06	1.52E-05	1.59E-06	1.03E-05	5.76E-06	6.70E-07	1.66E-06	1.27E-06	4.47E-06	2.38E-05	2.08E-06	3.34E-06	0.000535	6.78E-07	7.89E-07	5.67E-06	4.61E-06	2
682559	37	2.55E-06	1.13E-06	1.02E-06	5.47E-06	1.72E-06	2.72E-06	5.01E-07	2.64E-06	3.90E-06	1.48E-06	4.88E-06	1.53E-05	1.46E-06	1.14E-05	5.97E-06	7.63E-07	1.96E-06	1.68E-06	2.42E-06	1.58E-05	2.25E-06	4.16E-06	0.000561	9.16E-07	8.50E-07	8.44E-06	4.62E-06	2
682559	38	1.52E-06	1.18E-06	6.84E-07	1.42E-06	1.07E-06	9.99E-07	4.07E-07	5.73E-07	8.60E-06	1.04E-06	1.81E-06	1.34E-05	4.07E-07	2.27E-06	1.22E-06	4.07E-07	1.26E-06	5.92E-07	8.88E-07	4.40E-06	2.22E-06	2.29E-06	0.000122	2.77E-07	7.21E-07	7.19E-06	3.59E-06	1
682559	39	1.49E-06	1.20E-06	1.10E-06	1.42E-06	1.45E-06	1.33E-06	7.31E-07	1.26E-06	1.60E-06	1.60E-06	2.87E-06	1.34E-05	1.13E-06	6.28E-06	1.46E-06	7.02E-07	1.78E-06	1.59E-06	2.35E-06	1.40E-05	1.99E-06	3.18E-06	0.000335	4.44E-07	8.45E-07	9.87E-06	5.30E-06	2
682559	40	1.72E-06	1.36E-06	9.16E-07	3.64E-06	1.92E-06	1.92E-06	5.67E-07	1.75E-06	3.40E-06	1.57E-06	4.12E-06	1.77E-05	1.10E-06	1.00E-05	2.07E-06	5.23E-07	1.81E-06	1.16E-06	1.74E-06	1.12E-05	1.92E-06	2.72E-06	0.000484	5.23E-07	1.06E-06	7.54E-06	4.10E-06	1

Figure 26

The “Export fcs” option will save all gates as fcs files, excluding the neighbor columns. This format is suitable for use in, for example, a Citrus analysis. The exported fcs files are stored in the “Custom Gates” folder.

miCAT – For power users

The miCAT code was written in Matlab version 2014b. For optimal performance be sure to have this version installed when running miCAT from the source.

1. Modular build of miCAT

miCAT is built modularly, which yields the advantage that adding new features is easy and does not require changes in any of the existing structure. In general, features in miCAT must include only two basic scripts: one callback from the GUI and one script executing the function. The main functions are not linked to the GUI and can be run independently.

2. Data retrieval

All data stored for the current session and necessary to perform any function can be retrieved from the GUI handles or included manually without the GUI. Throughout a session, the data are kept in an fcs format structure. There is one main matrix containing a column for each channel and a row for each individual cell of each image. This matrix is continuously updated during the session and will therefore also contain the custom gates and channels. The corresponding channel names for each image are saved in a cell array. All individual tiff files and corresponding masks are stored in a multidimensional matrix structure.

Supplementary Note 2 - Results

Supplementary Figure 5

The provided simple synthetic data set demonstrates the validity of the neighborhood module in detecting cell neighborhoods deviating from randomness (Supplementary Figure 5a-d). These artificial images were constructed in a chessboard pattern to simplify the visualization of the validation. As described in the methods section, the absolute cell number or the size of the image does not have a direct effect on the neighborhood analysis – only the frequencies of cell types present, their relative quantities, and connectivity are important (see equation 2, Methods). Therefore, the underlying synthetic data sets were designed to imitate the ratios and relative frequencies of cell types present in the experimental data set. The validation covered average and extreme cases (Supplementary Fig. 5e,f).

The test dataset contained three different “phenotypic” clusters of cells, constructed by PhenoGraph: green represents cells of phenotype #1, blue represents cells of phenotype #2, and red cells are part of phenotype #3. We did not add further phenotypes or structures, since image complexity does not have a direct effect on the neighbor analysis of an individual cell phenotype and cell connectivity is kept constant for the permutation test (see equation 2, methods).

All of the following examples used a pixel-expansion of 4 to define neighbors, 99 permutations were run, and a significance cut-off of 0.05 for the p-value was used. The hierarchically clustered heatmaps display the cell cluster interactions on the x-axis and the four test images on the y-axis. No interactions involving phenotype #2 were observed. Those columns were automatically cut out of the visualization since their interaction frequencies did not significantly deviate from randomness. Thus, the focus here is on phenotypes #1 and #3.

The alternating pattern of cells of phenotype #1 and #3 in image 1 prevents cells of the same cluster from being in each other’s neighborhood (Supplementary Fig. 5a,f). This example represents common case in the experimental data where two phenotypes are equally distributed

across the image. Therefore, the interactions #1↔#1 and #3↔#3 are significantly less frequent than in random permutations control which would almost always have some interaction between cells of the same phenotype. As expected, miCAT neighbor analysis identifies both #1↔#1 and #3↔#3 as avoiding each other (Supplementary Fig. 5a). Similar cell distributions are for example observed in samples Ay10x3 and Ay1x3 which both have two equally sized abundant cell types (Supplementary Fig. 5g).

A positive control for interaction is displayed as the bulk of phenotype #3 cells in image 2 (Supplementary Fig. 5b,f). This enrichment in interaction between cells of phenotype #3 with themselves (#3→#3) is displayed red in the corresponding row of the hierarchically clustered heatmap. This interaction occurs significantly more frequently in the actual test image than in a matched randomized control image because random permutations are likely to distribute cells of phenotype #3 among those of phenotype #1. This also leads to significantly fewer phenotype #1 cells neighboring phenotype #3 cells in the permuted images compared to our real image, hence the blue #1→#3 interaction in the hierarchically clustered heatmap (Supplementary Fig. 5a). This example represents the case of one phenotype forming a cluster that is measured as an interaction with itself and the avoidance of other phenotypes. The presence of a rare cell type next to an abundant cell type is a common scenario in the experimental data, with sample Ay5x7 and Ay5x8 being representative examples (Supplementary Fig. 5h)

Image 3 (Supplementary Fig. 5c) does not reach the significance cut-off for any neighbor interactions even though two of the cells of phenotype #3 are neighbors (Supplementary Fig. 5c). This example represents a single rare interaction, and the result visualizes the effect of the significance cut-off as expected. A simple example can be seen in image Ay7x1, an extreme case with few cell types (Supplementary Fig. 5i).

Even though a single interaction is not significantly different than chance, two rare phenotypes forming repeated interactions do measure as significant. In image 4, cells of phenotype #1 always neighbor a cell of phenotype #3 and vice versa (Supplementary Fig. 5d).

Here, the hierarchically clustered heatmap of significant interactions in image 4 shows that the interactions between cells of phenotype #1 and cells of phenotype #3 occur significantly more often (in both directions) than in a randomly shuffled image. The presence of small clusters of cells is a common occurrence, particular in Grade 1 tumor samples, and sample By14x4 and Cy13x5 are representative examples of this scenario (Supplementary Fig. 5j).

These rare interactions visualized in image 3 and 4 are regularly present in the experimental data as well (Supplementary Fig. 5b,i,j). This highlights the validity of the analysis over a wide range of cell ratios, relative frequencies from prevalent to rare cell types and interactions, and represents the full complexity of cell-cell interactions found within the experimental data.

Supplementary Figure 6

To quantitatively assess how segmentation impacts our single-cell analysis and neighbor identification, we had five independent users, two of them inexperienced in tissue segmentation (users 4 and 5), segment three identical images using variations of our standard tissue segmentation pipeline (Material and Methods). After segmentation, we compared the masks and derived single-cell data. The five masks varied slightly, ranging from over segmentation (white, user 4) to under segmentation (purple, user 5) masks (Supplementary Fig. 6a). To quantitatively assess the segmentation quality, we used the segmentation score described previously by Schüffler et al. (Supplementary Fig. 6b). This score is composed of four intuitive segmentation constraints: (1) mask should overlap with membrane signal; (2) mask should not overlap with nuclei signal; (3) segmented cells should contain maximal one nucleus; (4) mask should approximate the expected number of cells based on cell radius. No individual user or segmentation strategy provided the best scoring for all images.

We then studied the effect of variation in user segmentation on several levels. First, we compared single-cell data from all test subjects in 2D scatter plot distributions. Single-cell

distributions and correlations between co-expressed markers (Supplementary Fig. 6c), and spatially resolved features (Supplementary Fig. 6d) were highly similar. Second, PhenoGraph analysis of the segmented single cells by the five users yielded similar cell type identifications from the same images (Supplementary Fig. 6e) and similar cell numbers (Supplementary Fig. 6f). Only the results from the inexperienced user 5 differed from those of the others. Small variations in segmentation did result in some individual cells on the edge of a high-dimension single-cell cluster being assigned to a different, but similar, cell phenotype during the PhenoGraph analysis. This could result in small variations in the neighbor interaction analysis for rare phenotypes. Third, we compared the neighborhood analysis among users. Using our standard settings (Material and Methods) no individual user scored consistently better than any other (Supplementary Fig. 6a) and comparable neighbor cell interactions were identified and clustering separated the individual images (Supplementary Fig. 6g). Some differences were seen between users, with the biggest difference seen for user 5 compared to all other users (which given the points above comes without surprise). The main influence on the comparability of the neighbor analysis was found to be rare cell types that could be caused by mis-segmentation. Thus, to ensure that rare or variable neighbor interactions do not result in false positives throughout our analysis we only analyze interactions that are present in greater than 10% of our images. Another factor that influences the neighbor analysis is the distance parameter that defines if cells are neighbors. Varying the expansion of the neighborhood in our dataset made the neighbor analysis of different users more similar to each other without increasing false positive interactions (Supplementary Fig. 6g). We conclude that our segmentation is of high quality and our neighbor analysis is robust to changes in variations in segmentation using the segmentation methods applied here.

Supplementary Figure 7

Analysis of an IMC dataset of healthy mammary ducts and alveoli from six donors validated our ability to identify relevant cell interactions that were known a priori. Specifically,

interactions between basal/myoepithelial and luminal cell types in the outer and inner layers of mammary ducts are known features of both alveoli and ductal structures of the mammary gland.

IMC enabled identification of complex cell types with different combinations of luminal (CK19, CK8/18, CK7) and basal (CK5, CK14, SMA) markers (Supplementary Fig. 7b). Multiple luminal (Phenograph clusters #3 and #6) and basal (Phenograph clusters #4, #7 and #13) cell types were identified, and SMA^+ endothelial cells (Phenograph clusters 12 and 14) were distinguished from SMA^+ basal cells (Supplementary Fig. 7a,c). Specific luminal and basal cell types were associated with different epithelial structures, but more images are needed to validate this observation and identify alveoli or ductal specific cell phenotypes.

As visualized across all images, and in sample Ay1x5 alone (Supplementary Fig. 7a, fifth row), high-dimension analysis revealed differences in the cytokeratin expression of the epithelial cells from different ducts or alveoli even in neighboring structures of the same breast. Even within the unforeseen complexity of these normal tissues, neighborhood analysis identified significant interactions that correspond to known interactions within mammary ductal structures. Epithelial cell types (Phenograph clusters #3, #4, #6, and #7) had similar interaction patterns (Supplementary Fig. 7g). Within these interactions, as expected, all basal-luminal epithelial interactions were highly significant, and basal cells were more commonly in the neighborhood of a luminal cell than vice versa (e.g., row 6, column 7 > row 7, column 6: basal around luminal > luminal around basal). Interaction patterns also helped distinguish endothelial vessels, which had their own interaction profile (Supplementary Fig. 7g, row 12 and 14) and identified stromal cell populations that were associated with mammary epithelial structures (Phenograph clusters #1, #2, and #8) and those that avoided interactions (Phenograph clusters #5, #9, and #10).

Supplementary Note 3 – Discussion

Segmentation

Segmentation of tissues is a challenging task for the entire image cytometry field and all tissue analysis approaches. In 5-μm thick tissue sections, a small percentage of cells will overlap, only a portion of a cell may be present within the analyzed section, and small differences in segmentation between adjacent cells can result in a small spillover of signal between neighboring cells. The ability to visualize a variety of cell markers and membrane stains helps to improve the accuracy of segmentation masks. Different cell types are segmented using different markers and a combination of markers helps to define cell boundaries. This is especially true in heterogeneous tumor samples where different adhesion markers (E-cadherin) or cell markers (cytokeratins) are present in different samples. In addition to improving segmentation, the increased information provided by high-dimension imaging can also better identify any poorly segmented cells so that they can be excluded from further analysis.

Phenotype Identification

Many single-cell technologies depend on automated high-dimension clustering techniques for the analysis and understanding of cell types. Manual gating is no longer adequate for systems level studies as 2D gating is not easily scalable, is biased toward expected results, and is one of the most variable steps in cytometry studies[1]. High-dimension clustering, such as Phenograph, is able to identify associations between markers that may be missed using conventional gating and identifies rare cell populations[2]. Differences in cell segmentation can impact marker quantification and cell phenotype determination for individual cells, but we find that multiple segmentations by different users provide a similar distribution of single-cell measurements.

Neighbor Interactions

Only recently have high-dimension imaging and other spatially resolved techniques been developed that are able to identify many cell types within a single measurement, and neighbor analysis is dependent on the ongoing development of tools for the identification and analysis of these single cells. A change in the phenotype identified also results in the identification of a different neighbor interaction, and differences in segmentation can change the measured connectivity of a tissue thereby altering the distribution of potential interactions. The use of a permutation test to measure the significance of a neighbor interaction provides internal controls for all images thereby controlling for differences in image processing and analysis between different experiments, as well as the inherent variability of different tissues. In addition, setting the distance parameter that determines a neighbor interaction can alter or correct for differences in segmentation strategies. Overall, we observed that relevant neighboring interactions are consistently detected across different segmentations. A focus on interactions that are detected across multiple samples results in identification of the significant neighboring interactions in a sample set.

Bibliography

1. Mair, F. *et al.* The end of gating? An introduction to automated analysis of high dimensional cytometry data. *Eur J Immunol* **46,** 34–43 (2016).

2. Levine, J. H. *et al.* Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell* **162,** 184–197 (2015).

Supplementary Note 4 – Antibody Validation

The majority of antibodies used in this study have previously been validated for imaging mass cytometry. Additional example validations of antibodies used in this study provided by the research community are below.

Smooth Muscle Actin (SMA): clone 1A4, Abcam

This antibody has been validated on multiple tissues including breast tissue and basal cancer subtype by the vendor and has been used in many published studies.
http://www.abcam.com/alpha-smooth-muscle-actin-antibody-1a4-ab7817-references.html

EpCAM: clone 9C4, Biolegend

This antibody is validated for immunohistochemistry on paraffin-embedded sections and quality tested for immunofluorescence by Biolegend. Validation staining has also been completed on many tissues include breast cancer by The Human Protein Atlas:
http://www.proteinatlas.org/ENSG00000119888-EPCAM/antibody.

Fibronectin: clone 10/Fibronectin, Becton Dickinson

This antibody has been tested in multiple human cancer tissue types as well as immunofluorescence of *in vitro* cell lines and western blot of human breast tissues. It was used in:
Wagner D, Bonenfant N, Parsons C, Sokocevic D, Brooks E, Borg Z, *et al*. Comparative decellularization and recellularization of normal versus emphysematous human lungs. Biomaterials. 2014;35:3281-97.

Cytokeratin 5: clone EP1601Y, Abcam

This antibody is guaranteed for immunohistochemistry applications by the vendor and has tested on multiple tissues including the basal cancer subtype of breast cancer.

Cytokeratin 19: clone Troma-III, Developmental Studies Hybridoma Bank

This antibody that has been tested on a variety of tissues and was used in a number of published studies. It was used to detect specific cellular subpopulation in primary human breast culture:

Zubeldia-Plazaola A, Ametller E, Mancino M, Prats de Puig M, López-Plana A, Guzman F, *et al*. Comparison of methods for the isolation of human breast epithelial and myoepithelial cells. Frontiers in Cell and Developmental Biology. 2015;3:32.

Cytokeratin 14: polyclonal, Thermo Fischer

This antibody is validated for immunohistochemistry application, and amongst many other uses was shown to identify the basal cell population in human breast primary culture and mouse mammary tissues:

Hines WC, Yaswen P, Bissell MJ. Modelling breast cancer requires identification and correction of a critical cell lineage-dependent transduction bias. Nature Communications. 2015;6:null

Meyer DS, Brinkhaus H, Müller U, Müller M, Cardiff RD, Bentires-Alj M. Luminal expression of PIK3CA mutant H1047R in the mammary gland induces heterogeneous tumors. Cancer Research. 2011;71:4344.

Cytokeratin: clone AE3, Millipore

Performance of this antibody is guaranteed by the vendor in immunohistochemistry against type II keratins. It has been extensively used in combination with other antibodies to detect epithelial or carcinoma tissue.