IN Carta

Image Analysis Software

Phenoglyphs User Guide

5087398 C
June 2023

MOLECULAR DEVICES

Product manufactured by Molecular Devices, LLC.
3860 N. First Street, San Jose, California, 95134, United States of America.
Molecular Devices, LLC is ISO 9001 registered.

Contents

Chapter 1: IN Carta Image Analysis Software

The IN Carta® Image Analysis Software provides powerful analytics for advanced phenotypic classification and 3D image analysis. It delivers robust, quantitative results from complex biological images and datasets using advanced AI technology.

Faster Data

- Intuitive design makes complex analysis accessible with minimal training.
- Shorten analysis time with true parallel processing.

Reliable Data

- Sophisticated algorithms generate reliable data with minimal user input.
- Improved segmentation algorithms represent cellular structures more accurately.

Results That Matter

- See real results quickly—from populations to single cells—using integrated data visualization tools.
- User-friendly interface guides you through your discoveries with continual updates that grow with your needs.

Terminology

The following table defines the terms and abbreviations used in this guide.

Term	Definition
Cluster	Category in which cells of a particular phenotype are grouped.
Decision Tree Classifier	Classification method in which populations are split into two or more distinct categories by applying a measure threshold. Thresholds are applied serially to further classify sub-populations.
Exemplar	Image of a biological structure serving as a typical example of a phenotype within a cluster.
F1 Score	A measure of model quality equal to the harmonic mean of precision and recall.
Measure	A measurable characteristic of a segmented biological structure, cell, or organism.
Measurement	Set of object-level data extracted after image analysis.
Multiparametric Classifier	Classification method that utilizes all measurable parameters or user-defined subset of parameters simultaneously to group structures of similar phenotypes.
Phenotype	Observable characteristics or traits of a segmented structure, cell, or organism.
Supervised clustering	Classification method in which a machine learning algorithm groups cells into predefined number of clusters based on examples you provide,
Unsupervised clustering	Classification method in which a machine learning algorithm groups cells into clusters based on similarity. You can indicate number of clusters.

Obtaining Support

Molecular Devices is a leading worldwide manufacturer and distributor of analytical instrumentation, software, and reagents. We are committed to the quality of our products and to fully supporting our customers with the highest level of technical service.

Our Support website—www.moleculardevices.com/service-support—describes the support options offered by Molecular Devices, including service plans and professional services. It also has a link to the Molecular Devices Knowledge Base, which contains documentation, technical notes, software upgrades, safety data sheets, and other resources. If you still need assistance, you can submit a request to Molecular Devices Technical Support.

Technical Support

To contact Molecular Devices Technical Support, submit a support request through the Molecular Devices Knowledge Base at support.moleculardevices.com.

You can also submit a support request by phone. For regional support contact information, go to www.moleculardevices.com/contact.

To expedite support, be prepared to provide the software version and your activation ID. To display this information, at the bottom left of the IN Carta window, click  **Settings**.

Documentation

Review the product documentation on the Molecular Devices Knowledge Base at support.moleculardevices.com. In addition, online Help is available within the IN Carta software.

Additional Resources

Web-based microscopy courses:

- www.microscopyu.com
- www.ibiology.org/ibioeducation/taking-courses/ibiology-microscopy-short-course.html

The *Molecular Probes Handbook* offers advice on fluorescent probes and can help you determine if there are better stains available for your analysis:

- www.thermofisher.com/us/en/home/references/molecular-probes-the-handbook.html

The *Assay Guidance Manual* details state-of-the-art approaches to high-content screening (HCS) and discusses challenges specific to HCS. It serves as a good introduction for new HCS practitioners.

- www.ncbi.nlm.nih.gov/books/NBK100913

About This Guide

This guide is intended for the scientist using the IN Carta software. It provides an overview of IN Carta Phenoglyphs™ and describes the Phenoglyphs workflow.

The information in this guide is valid for IN Carta software version 2.1 and is subject to change without notice. We recommend that you review the guide on the Molecular Devices Knowledge Base at support.moleculardevices.com for the most up-to-date information.

Chapter 2: IN Carta Phenoglyphs

2

Phenotypic screens are used to determine whether a treatment (for example, chemical compound) or condition (for example, genotype) causes alterations in the phenotype (for example, morphological changes) on a cellular or organismal level.

IN Carta Phenoglyphs is a multiparametric classifier which utilizes multivariate cell-level measurements to categorize segmented features into different phenotypic classes.

Phenoglyphs allows you to classify the following as a percentile portion in a discrete class of cells:

- Different types of cells with two or more visually distinct phenotypes
- Treated cells (chemical/RNAi/CRISPR)

High-Content Analysis (HCA) is a phenotypic screening method for active compounds, antibodies, or genes that affect a particular biological process. HCA aims to extract diverse quantitative information from thousands of data points within an experiment.

One of the key steps when analyzing imaging data from HCA screens is classifying the phenotype of cells in an automated fashion. Once each organelle has been assigned to a certain phenotype, researchers can determine the effects of the experimental conditions at a wider level.

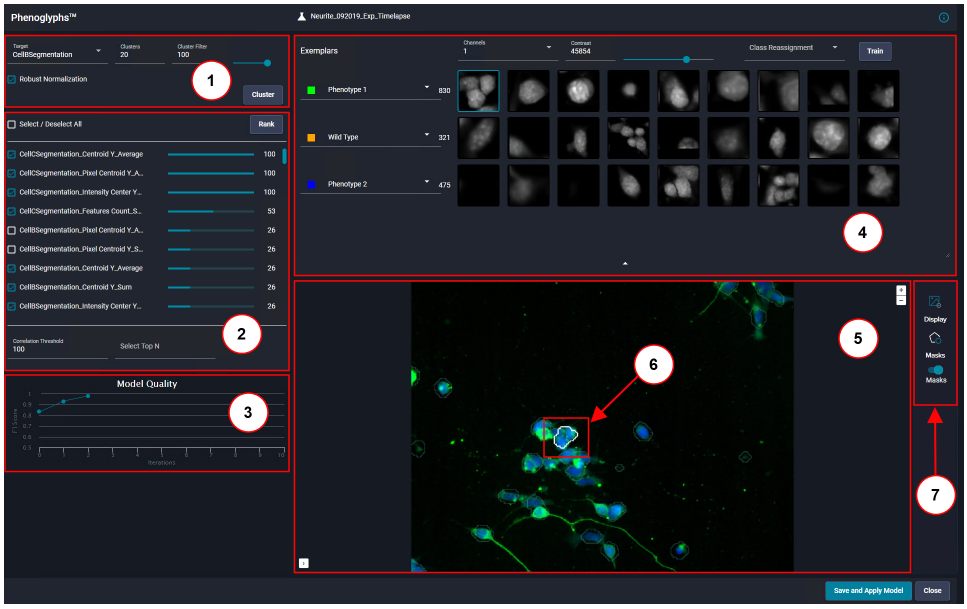
Note:
An additional license is required to use Phenoglyphs. Some functions are available to all users, but the following functions are not available without a valid license:

- Saving and applying a model.
- Running a protocol with a Phenoglyphs classifier.

If you are running the IN Carta software with a network license, you may be able to reserve a license to access Phenoglyphs. See Settings for details.

Phenoglyphs Window

The following illustration shows the Phenoglyphs window.



Part	Description
1	Cluster Pane
2	Rank Pane
3	Model Quality Pane
4	Exemplar Pane
5	Image Viewer Pane
6	Selected Exemplar
7	Image Viewer Tools

Phenoglyphs Workflow

The Phenoglyphs workflow is applied after segmentation. Analysis can be divided into the following steps.

Step	Phase
Clustering	Characteristic phenotypes are identified by the unsupervised classifier algorithm.
Labeling	The cell clusters are manually labeled and visually reviewed based on the presented exemplars of phenotypes.
Ranking	Measures are ranked to assist you in selecting the most relevant measures for optimizing the model. Rank scores (scaled 1 to 100) show the significance of each measure.
Training	A supervised machine learning algorithm is used to generate a training model based on the labeled clusters and selected measures.
Saving and Applying a Model	The model is saved with the protocol and applied to the current dataset.

The following diagram shows the Phenoglyphs workflow:

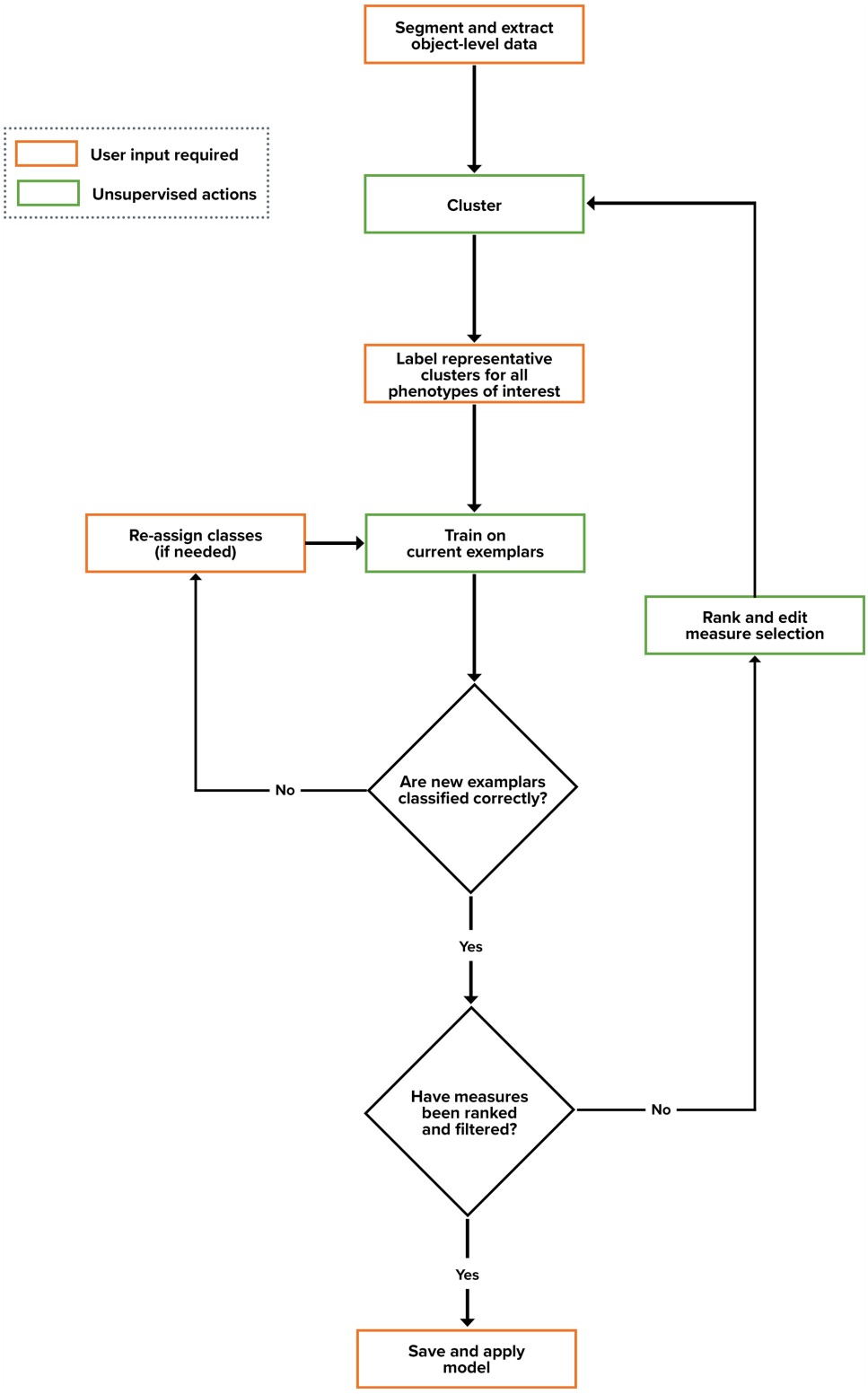
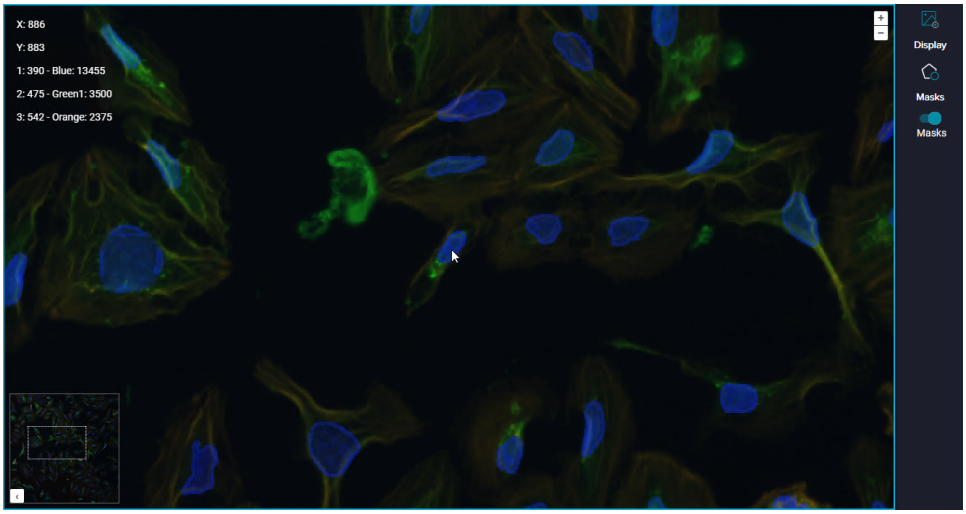


Image Viewer

Use the **Image Viewer** pane to view the image from the experiment.



Click and drag to pan around the image.

Double-click the **Image Viewer** pane to center the image on the spot that you click.

Use the mouse wheel to zoom in and out. Alternatively, click + **Zoom In** and − **Zoom Out**. Zoom is reset and image is recentered when a new well or FOV is selected.

Click › **Overview Map** to open a thumbnail view of the image. Use the thumbnail to navigate around a zoomed image.

Hover over the image to show x/y coordinates of the current mouse pointer location along with the pixel intensity for each wavelength.

Click at the bottom of the **Exemplar** pane to expand the **Image Viewer** pane. Click to restore the default view.

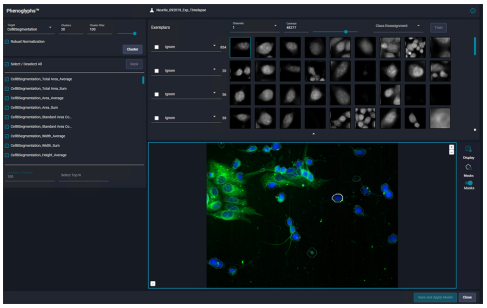


Image Viewer: Default Pane

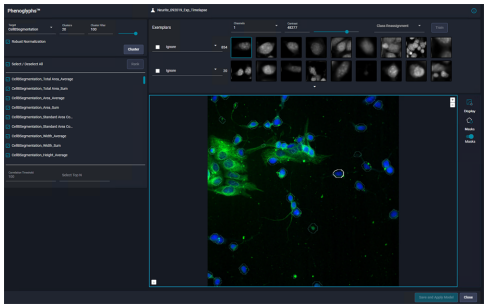
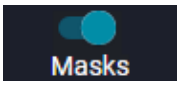


Image Viewer: Expanded Pane

The following controls are available:

Display: Opens the **Image Display** dialog, where you can set contrast and brightness. See Image Display on page 13 for details.

Masks: Opens the **Mask Properties** dialog, where you can toggle visibility of the mask for the selected target on/off. You can also change mask type, color, and opacity. See Mask Properties on page 14 for details.



Masks: Toggle visibility of the target mask on/off. (Shown in the on position.) See Mask Properties on page 14 for details on adjusting the visibility of each mask.

Image Display

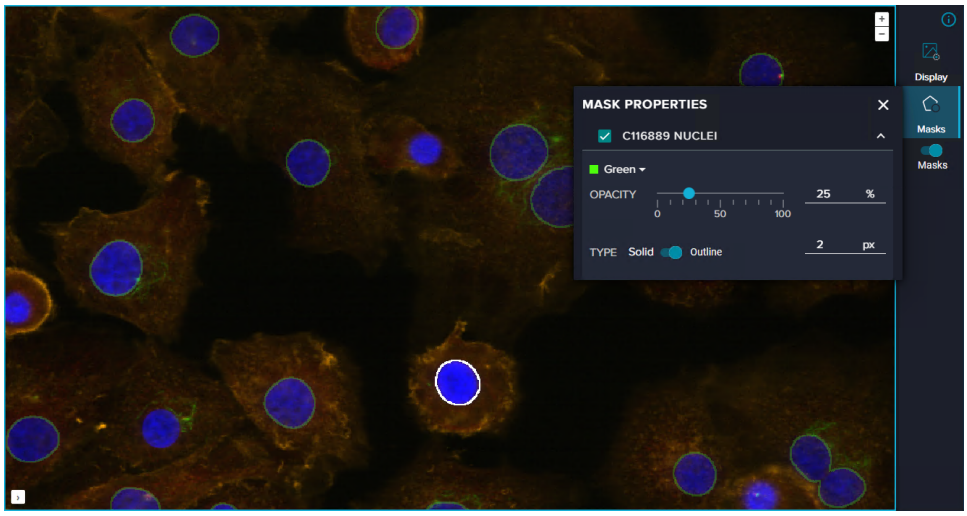
In the **Image Viewer** pane, click  **Display** to display the **Image Display** dialog.

Use the controls to set the contrast and brightness. The display color for the wavelength is based on the acquisition channel wavelength; you can change this color, if needed. A blended image is shown by default. Deselect one or more wavelengths to hide channels.

Mask Properties

In the **Image Viewer** pane, click  **Masks** to display the **Mask Properties** dialog.

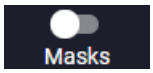
Use the controls to change mask type, color, and opacity. You can set the default mask color for targets is provided in the Protocol Editor Table; the default color for classes is defined during classification.

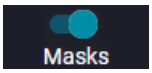


To change the mask appearance for a Target/Class in the Image Viewer, then adjust the **Type**, **Color**, **Opacity**, and **Pixel Size**.

The selections in the **Mask Properties** dialog persist when you select a new examplar.

In addition, the following controls are available:



Masks Off or 

Masks On: Toggle to control mask visibility on/off.

Chapter 3: Segmentation and Analysis

Phenoglyphs classification can only be performed on segmented and analyzed data.

See the *IN Carta Protocol Design App Guide* for details on segmentation of the image data,

Additional Recommendations

The following table summarizes the recommendations to consider for segmentation prior to Phenoglyphs classification.

Note: Before Phenoglyphs classification, it is unknown which target features are most representative.

Settings	Recommended Settings	Additional Comments
Measures	Activate all Spatial, Intensity, and Texture measures for all channels and targets during initial analysis protocol setup.	Phenoglyphs ranks measures that are required to properly classify phenotypes.
Targets	One target type can be used for Phenoglyphs classification. **Note:** Parent targets created in the Flexi-Protocol application have aggregated child target data.	Targets must include all phenotypes to be classified.
Wells	The analyzed wells must include all phenotypes wanted in the classification. **Note:** Include negative and positive phenotypic examples.	The Phenoglyphs model can be trained with data from one or multiple wells.

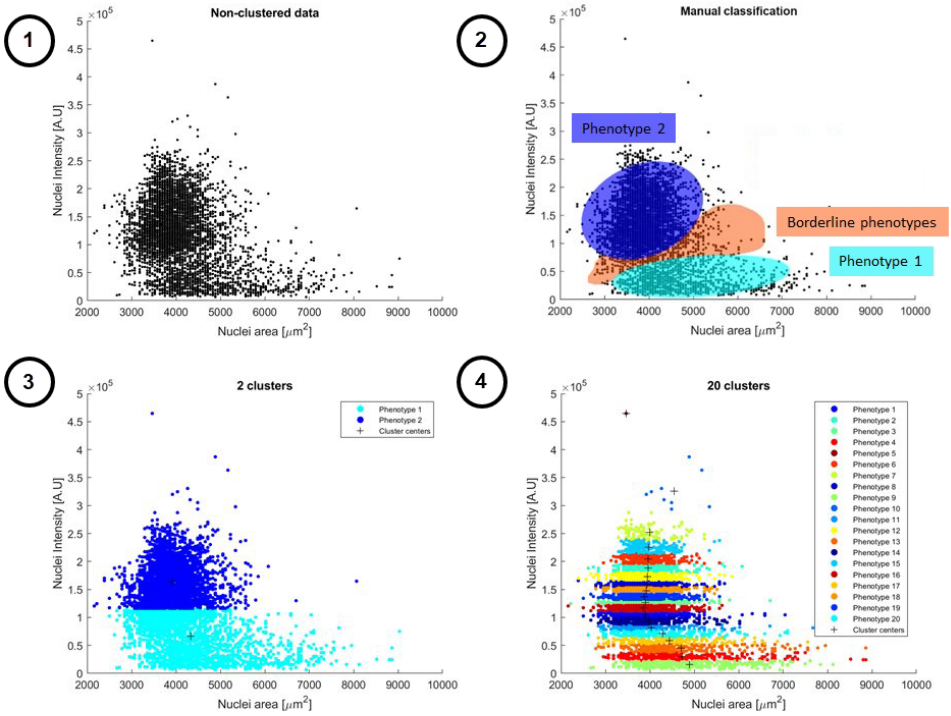
Cluster refers to a row of exemplars that represent a group of cells of a particular identified phenotype.

For clustering, up to 5,000 random cells from across a plate are grouped using the unsupervised **k-means++** algorithm.

Note: Up to 5,000 cells are used for classification. If a dataset has more than 5,000 cells, 5,000 cells are picked at random for classification.

k-means++ attempts to group multi-dimensional data (morphological, spatial, intensity, and texture) obtained for each cell into a specified number of clusters. If desired, a subset of wells can be selected before analysis to limit clustering to phenotypes found in those wells.

The scatter plots below show cells arranged by nuclear area and nuclear intensity. Graphs 3 and 4 show unsupervised **k-means++** classification into the specified 3 or 50 clusters. The default setting is 20 clusters.



Number of Clusters

A higher number of clusters reduces heterogeneity in the cluster and enables visualization of borderline phenotypes. Following clustering, similar phenotypes can be combined during the data labeling step.

The following table describes the classification result based on the number of selected clusters.

Number of Selected Clusters	Result
3	Only three clusters are available. All cells are placed either in Cluster 1, Cluster 2, or Cluster 3. Obvious representatives of each cluster have a high probability to be placed correctly, while less obvious representatives can be misclassified. You must manually re-assign classes.
50	More clusters are available, including: • A few large clusters representing each phenotype with obvious representatives. • Several small clusters containing representatives of borderline phenotypes. • A few clusters containing unexpected cells (for example, debris, out-of-focus cells, or cropped cells at the border).

Exemplar Pane

The following table describes the Exemplar pane.

Parameters	Function	Application
Channels	Use the Channels drop-down list to view the exemplar image in a different channel.	Switching between channels is useful when verifying phenotypes. It is applicable for cases when a phenotype is clearly visible in one channel, but not the other.
Contrast	Adjust Contrast using the slider to ensure the optimal display settings.	Adjusting the contrast may be required to improve visualization of exemplars in individual clusters. Contrast set in Exemplar pane is independent of contrast setting in the Image Display dialog of the Image Viewer.

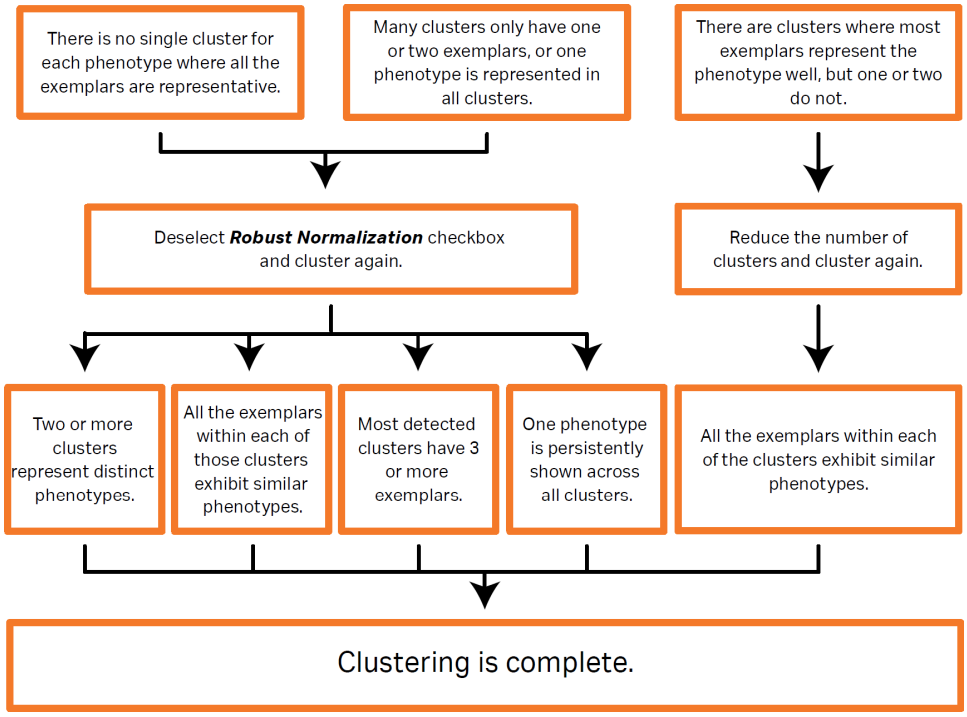
Cluster Parameters

The following table describes cluster parameters.

Clustering Parameters	Function	Application
Cluster	Number of clusters controls expected number of phenotypes.	A higher number of clusters reduces heterogeneity in the cluster and enables visualization of borderline phenotypes. See Number of Clusters on page 18 for details. **Tip:** After clustering, similar phenotypes can be combined during the data labeling step. See Labeling on page 23 for details. Set the field value to 20, even if the expected number of phenotypes is less than 5.
Cluster Filter	The Cluster Filter limits the range of exemplar objects shown for review. Every detected cluster will have a center defined by a minimum intra-class variance.	By default, any of the exemplars from a cluster can be shown. Reducing the value will exclude the objects that are further away from the cluster center. Only prominent representatives from each class are displayed. Set Cluster Filter value at maximum (100) to visualize the entire range of exemplars for a given cluster.
Robust Normalization	Robust Normalization specifies what kind of data is used by the normalization algorithm.	When activated, a robust method is used for scaling the data to an appropriate range. The robust algorithm is more complex and considers outliers while normalizing the data. If not activated, a simplified method for data normalization is used. Use Robust Normalization for the initial clustering. If obvious phenotype(s) are not present in any of the groups of exemplars, clustering can be repeated without Robust Normalization.

Clustering Result Guidelines

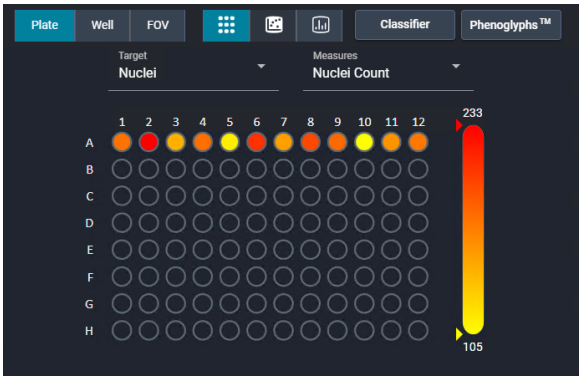
The following diagram describes the recommended actions to improve clustering.



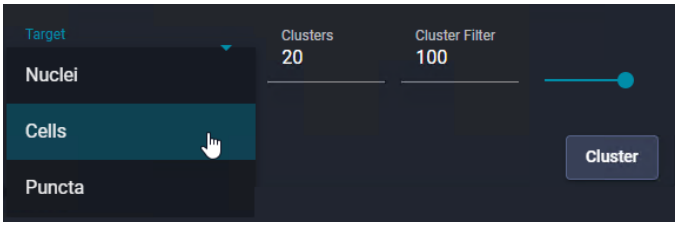
Creating a Machine Learning Model

After segmentation is complete, create a machine learning model that you can use for training.

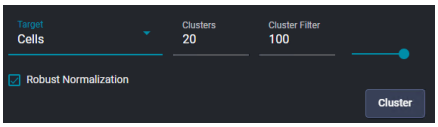
1. On the **Interactive: Data** page, in the Chart Dashboard, click **Phenoglyphs** to start the classification workflow.



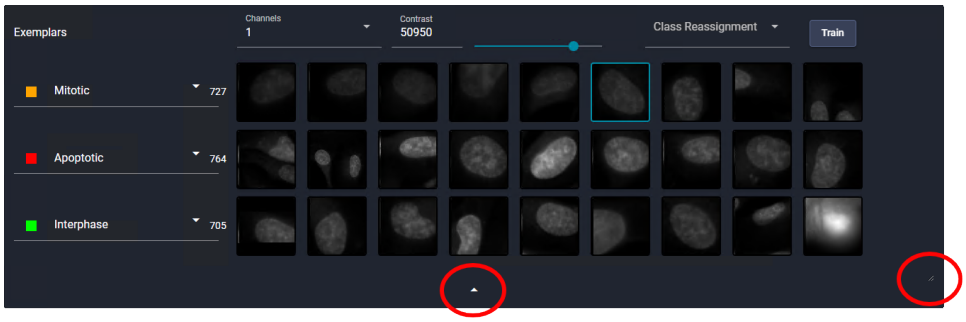
2. Select the target from the **Target** drop-down.



3. As needed, adjust the settings.



4. Click **Cluster**.



Each row in the Exemplar pane represents a phenotypic cluster identified by the algorithm. Exemplars display from left to right according to how well they represent the phenotype, with the left-most exemplars most strongly representing the phenotype.

Tip: Click and drag the box at the bottom right corner of the Exemplar pane to resize it. Click the triangle icon at the bottom of the pane to expand or restore the Image Viewer.

5. Check the clustering quality by viewing the exemplars.
6. Click on an exemplar to view the cell within the Image Viewer for more context.
7. Check clustering quality as described in Clustering Result Guidelines on page 20.

After clustering is complete, continue to the labeling step. See Labeling on page 23 for details.

Chapter 5: Labeling

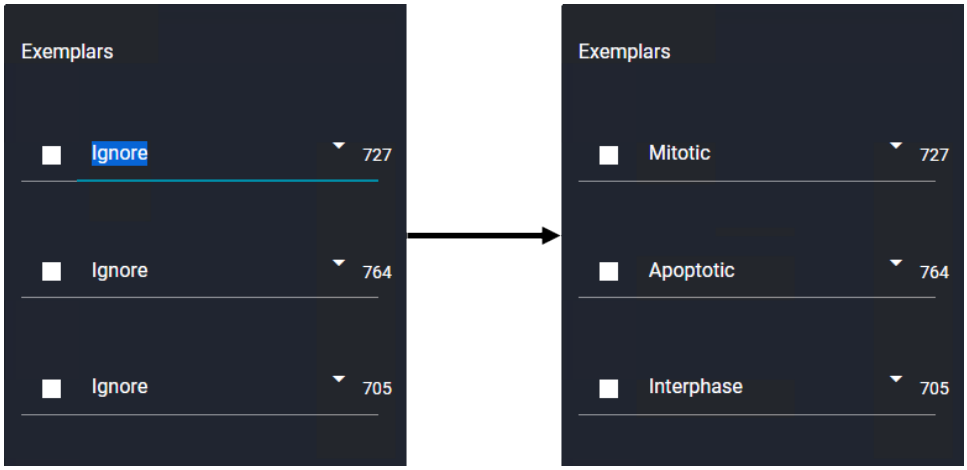
5

The goal of labeling is to identify and label features that are important for the supervised learning algorithm.

Note: You do not need to label all rows of exemplars.

To label data:

1. Visually identify clusters where 100% of the exemplars represent a clear phenotype.
2. Double-click **Ignore** in each class label, and replace it with an appropriate label for the class.



Note:
- Labels are case-sensitive.
- Once you create a class label, you can re-use it for other clusters with the same phenotype.

3. Use the **Ignore** label for clusters that do not clearly show a consistent phenotype.
4. As needed, create separate classes for debris, out-of-focus cells, and overlapping cells.
5. Once you create classes of interest, click **Cluster**.

Note: You must label at least two classes to proceed with clustering.

6. Repeat clustering based on the selected subset of measures from the chosen representative clusters.

After labeling is complete, continue to the ranking step. See Ranking on page 25 for details.

Chapter 6: Ranking

Selecting a subset of cellular features used for classification improves the performance of an algorithm. Preliminary ranking of the measures helps identify the most and the least relevant features.

Rank scores (scaled from 1 to 100) show the significance of each measure. The higher the ranking score, the more important the measure to separate the labeled classes.

By default, all measures are included when separating the classes, even low-ranking ones. Removing low ranking and highly redundant measures helps to increase the speed of classification and prevent the model being excessively specific for a limited number of objects in the dataset (known as model overfitting).

Selecting Measures

Consider the following recommendations about selecting measures:

- There is no specific number of measures that must be present for training.
- Limit the number of selected measures to 4 times the number of labels/phenotypes.
- Consider how Rank values “level off”.

The following illustration is an example of the Rank values.

The last four Rank values level off reaching a plateau. For this reason, you can clear the check boxes, removing them from training.

Controlling the Number of Measures

There are three methods to control the number of measures used for classification, which can reduce redundancy:

- Manually
- Correlation Threshold
- Select Top N

Method	Description
Manually	Manually select and deselect measures.
Correlation Threshold	Set a value to remove redundant features (that is, two features have similar trends). Redundant features have high correlation (greater than 90). Removing highly correlated features prevents overfitting of a model during training. For example, setting a threshold of 90 would: • Refine which pairs of features are correlated by over 90%. • Retain only the higher-ranking feature within any of the pairs.
Select Top N	Specify the number of the most important measurements that contribute to class separation.

Ranking and Editing Measures

To rank and edit measures:

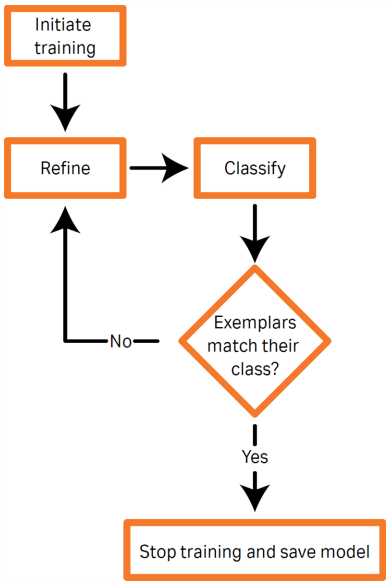
1. Click **Rank** to sort the features based on their importance.
2. Examine the Rank values.

> **Tip:** A typical distribution includes a few measures that score very highly (greater than 70) and then a sharp drop off. If there is no sharp drop in ranking, set a correlation threshold and manually select a subset of measures to include. The general rule for number of measures in this case is **Number of classes + 2**. For example, if you have defined three classes, then five measures would be sufficient.

3. Clear the check boxes for redundant measures.
4. Repeat the Clustering steps (see Creating a Machine Learning Model on page 21) and Labeling steps (see Labeling on page 23).

After ranking, continue to the training step. See Training on page 29 for details.

The following diagram shows the steps to train a model.



Consider the following when training a model:

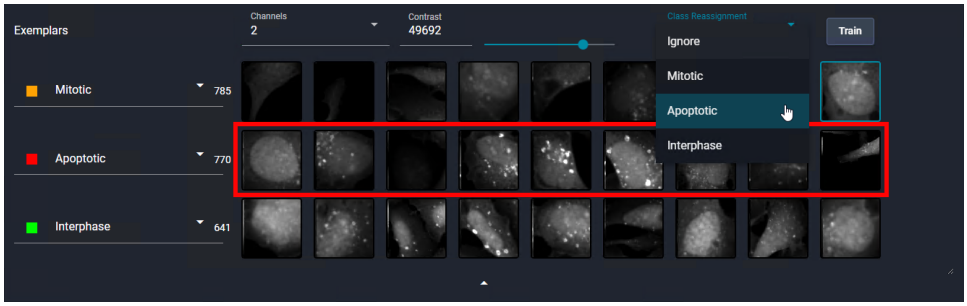
- If you are unsure about which phenotype to classify the exemplar, use the **Ignore** label.
- If a specific exemplar is not very clearly representative of the phenotype, use the **Ignore** label. When given ideal examples of the phenotype, the algorithm is better suited than the human eye to assign cells close to the boundary between classes.

  Note: Ignoring an object does not remove it from the analysis; it just does not include it in the training set.

- As needed, create separate classes for debris, out-of-focus cells, and overlapping cells.

Exemplar Position

The following table clarifies the position of the exemplars.



Position of Exemplars	Clarification
Left	Exemplars are more representative of the class.
Right	Exemplars are not as prominent in the class according to the model trained so far.

Tip: Take note of the exemplars on the left side of the rows. They must clearly represent the chosen phenotype. The consistent presence of the wrong exemplars on the left side of the rows makes it necessary to repeat the Phenoglyphs workflow starting at the clustering step. See Clustering on page 17 for details.

When to Stop Training

Stop training and proceed to saving the model when you observe one or more of the following:

- The F1 score has not improved much in the last few training iterations. See Model Quality on this page for details.
- All exemplars match their class after two or three training iterations.
- Few cells on the right side of an exemplar row must be re-assigned as **Ignore**.
- Exemplars on the right are continually ignored and cannot be placed into a class.

If classes are not improving (that is, only two or three exemplars from a class match another existing class) after 10-15 training iterations, restart the Phenoglyphs workflow with different clustering parameters.

Model Quality

Use the graph in the **Model Quality** pane to review the F1 score as training progresses. The F1 score is a measure of model quality equal to the harmonic mean of precision and recall.

An F1 score that does not improve over iterations can be an indicator to stop training.

An F1 score of 1.0 would represent perfect quality model that has been trained on consistent set of exemplars and will generate highly accurate results. Any score greater than 0.5 indicates predictive value. F1 scores less than 0.5 are not shown on the graph.

Hovering over each point in the graph will show the F1 score as a tooltip.

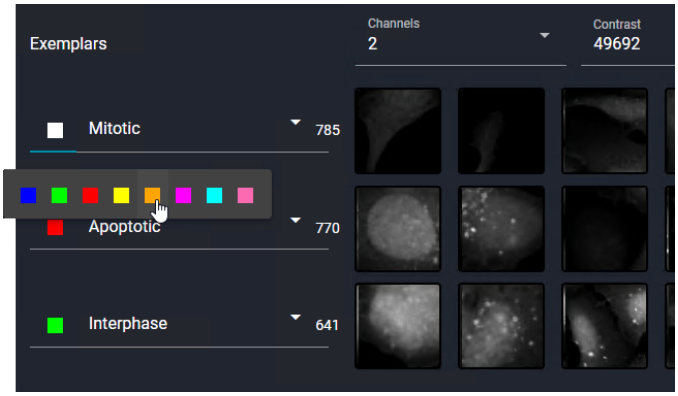
The software begins capturing F1 score data after the second training iteration (which is Iteration 0 in the graph).

Reclustering (that is, clicking **Cluster**) clears the current graph.

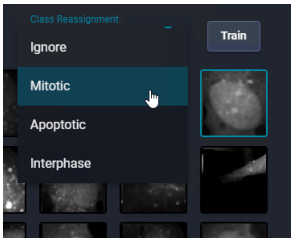
Training a Model

Do the following to manually refine the labeled classes:

1. Review the labeled classes and assign colors to each class.



2. Visually examine exemplars in each class. If there are exemplars that do not clearly represent the defined phenotype, click on the exemplar and use the **Class Reassignment** drop-down to reassign it to another class or use the **Ignore** label.



3. Click **Train**.
   The model is refined, which results in the following:
   - Any like classes will be combined.
   - All classes labeled **Ignore** will be removed.
   - New exemplars will be presented for review.
   - Exemplars are ordered according to their significance for a class based on current classification model.
4. Repeat step 3 to refine the model further, if needed.
5. Repeat the classification and refining steps (steps 2 to 4) until exemplars in all classes clearly represent the defined phenotype. Depending on the complexity of the data, as few as three or as many as ten training iterations may be required.

> **Tip:** More training is not detrimental to the model, provided only clear representatives of each phenotype are included in each iteration. See When to Stop Training on page 31 for details.

Chapter 8: Saving and Applying a Model

Save and apply the model to apply the Phenoglyphs classification model to the entire dataset and save it as a part of the current protocol for use with other datasets.

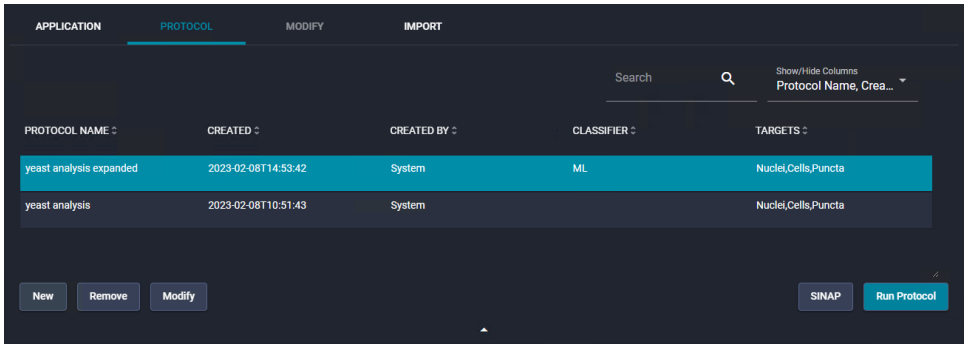
When training is complete, in the Phenoglyphs window, click **Save and Apply Model**.

Protocols with a Phenoglyphs classifier display **ML** (indicating machine learning) in the **Classifier** column of the Protocol Editor Table. You can run these protocols from the **Interactive: Analyze** page or the **Batch** page.

Note:
An additional license is required to use Phenoglyphs. Some functions are available to all users, but the following functions are not available without a valid license:

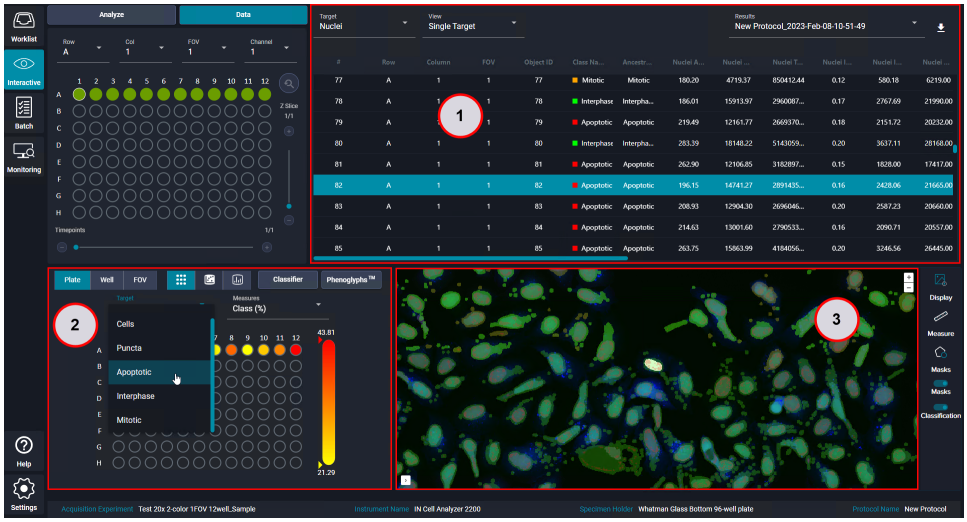
- Saving and applying a model.
- Running a protocol with a Phenoglyphs classifier.

If you are running the IN Carta software with a network license, you may be able to reserve a license to access Phenoglyphs. See Settings for details.



Class Information

The following illustration shows class information on the **Interactive: Data** page after saving a model.



Part	View	Description
1	Results Analysis Table	Lists results; each row is a single object.
2	Chart Dashboard	Indicates the percentage of cells in a class or plots a measure as a function of class.
3	Image Viewer	Uses class-specific mask colors.

Visit us on Spectranet.

Contact Us

Phone: +1-800-635-5577
Web: moleculardevices.com
Email: info@moldev.com

Visit our website for a current listing of worldwide distributors.

5087398 C

MOLECULAR DEVICES