IN Carta

Image Analysis Software

Phenoglyphs User Guide

5087398 B
September 2021

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 combined with an intuitive interface to simplify workflows for advanced phenotypic classification and 3D image analysis.

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 whereby 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.
Multiparametric Classifier	Classification method that utilizes all measurable parameters or user-defined subset of parameters simultaneously to group structures of similar phenotypes.
Measure	A measurable characteristic of a segmented biological structure, cell, or organism.
Phenotype	Observable characteristics or traits of a segmented structure, cell, or organism.
Linked Target	Displays results of cell target measures as well as organelle target measures. • If there are multiple organelles within a given cell, those organelle measures are averaged. • If no cell targets are included in the analysis protocol, organelle data is aggregated and displayed per nucleus.
Supervised clustering	Classification method whereby a machine learning algorithm groups cells into pre-defined number of clusters based on examples you provide,
Unsupervised clustering	Classification method whereby a machine learning algorithm groups cells into clusters based on similarity. You may optionally indicate number of clusters.
Measurement	Set of object-level data extracted after image analysis.

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.

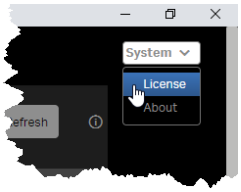
Our Support website—www.moleculardevices.com/service-support—describes the support options offered by Molecular Devices, including service plans and professional services. It also has a link to the Molecular Devices Knowledge Base, which contains documentation, technical notes, software upgrades, safety 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, please be prepared to provide the software version and the license ID.



- To display the software version, in the top right corner of the IN Carta window, click **System > About**.
- To display the license ID, in the top right corner of the IN Carta window, click **System > License**.

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 discussed 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 IN Carta Phenoglyphs workflow.

The information in this guide is valid for IN Carta software version 1.15 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

Application Overview

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.

High-Content Analysis

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.

IN Carta Phenoglyphs

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

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

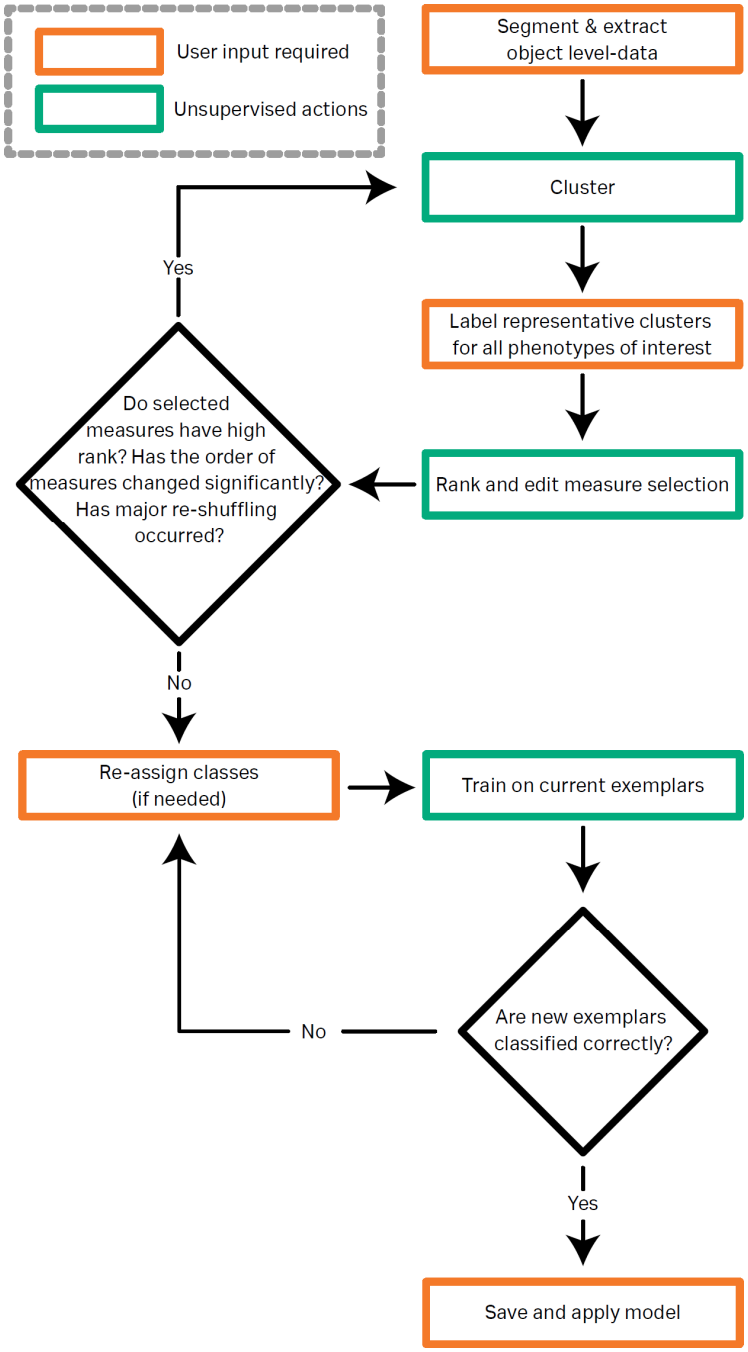
- Different types of cells with two or more visually distinct phenotypes
- Treated cells (chemical/RNAi/CRISPR)

IN Carta Phenoglyphs Workflow

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

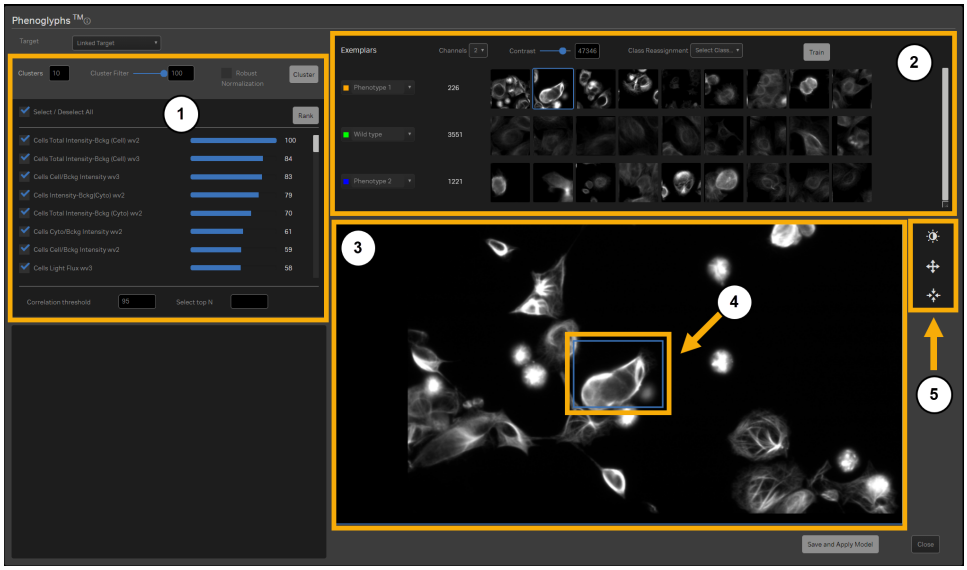
Steps	Phase
Clustering	Characteristic phenotypes are identified by the unsupervised classifier algorithm.
Data labeling	The cell clusters are manually labeled and visually reviewed based on the presented exemplars of phenotypes.
Ranking	Measures are ranked to assist you in selecting the most relevant measures for optimizing the model. Rank scores (scaled 1 to 100) show the significance of each measure.
Training a model	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 data set.

The following flow diagram shows the workflow when using IN Carta Phenoglyphs.



IN Carta Phenoglyphs Overview Screen

The following illustration shows the IN Carta Phenoglyphs overview screen.



Part	Description
1	Cluster settings
2	Exemplar viewport panel
3	Viewport panel
4	Selected exemplar
5	Image control

Chapter 3: Segmentation and Analysis

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

For details regarding segmentation of the image data, see the *IN Carta Protocol Design App Guide*.

Additional Recommendations

Note: Before phenotypic classification, it is unknown which cellular features are most representative.

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

Settings	Recommended Settings	Additional Comments
Features	Activate all Spatial, Intensity, and Texture features for all channels and targets during initial analysis protocol setup.	IN Carta Phenoglyphs ranks features that are required to properly classify phenotypes.
Targets	One or more target types can be used for IN Carta Phenoglyphs classification. **Note:** Only one target can be used at a time when classifying results created in VoluMetrics module.	The 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 IN Carta Phenoglyphs model can be trained with data from one or multiple wells.

4

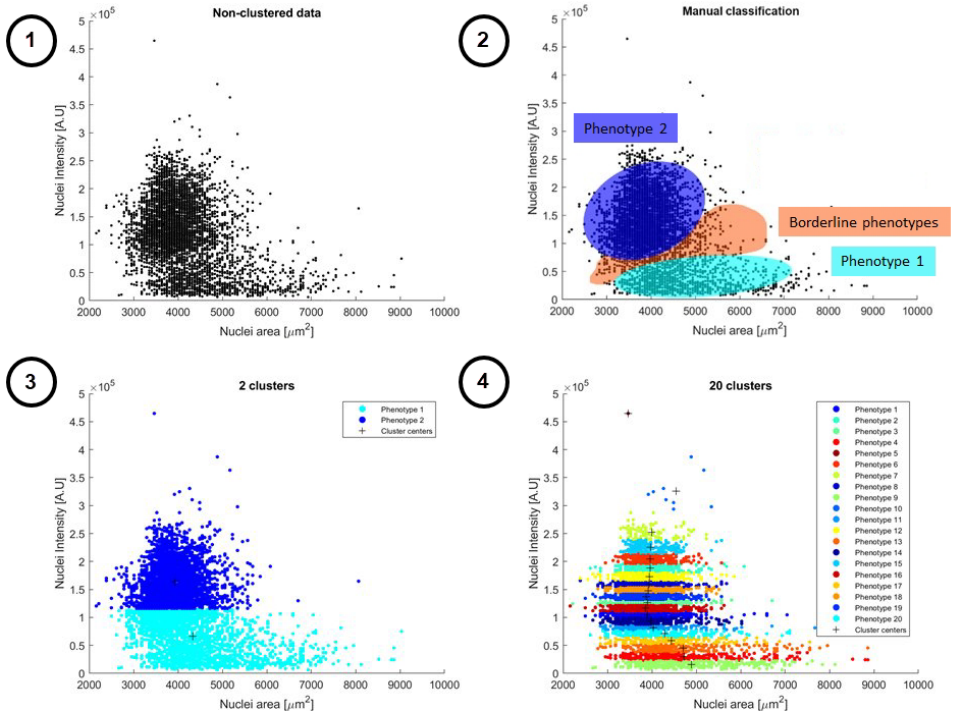
Cluster refers to a row of exemplars that represent a group of cells of a particular identified phenotype.

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 will be used for classification. If a data set has more than 5,000 cells, 5,000 cells will be 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 2 or 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
2	Only two clusters are available. All cells are placed either in Cluster 1 or Cluster 2. 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.
20	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 Viewports

The following table describes the exemplar viewport.

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.

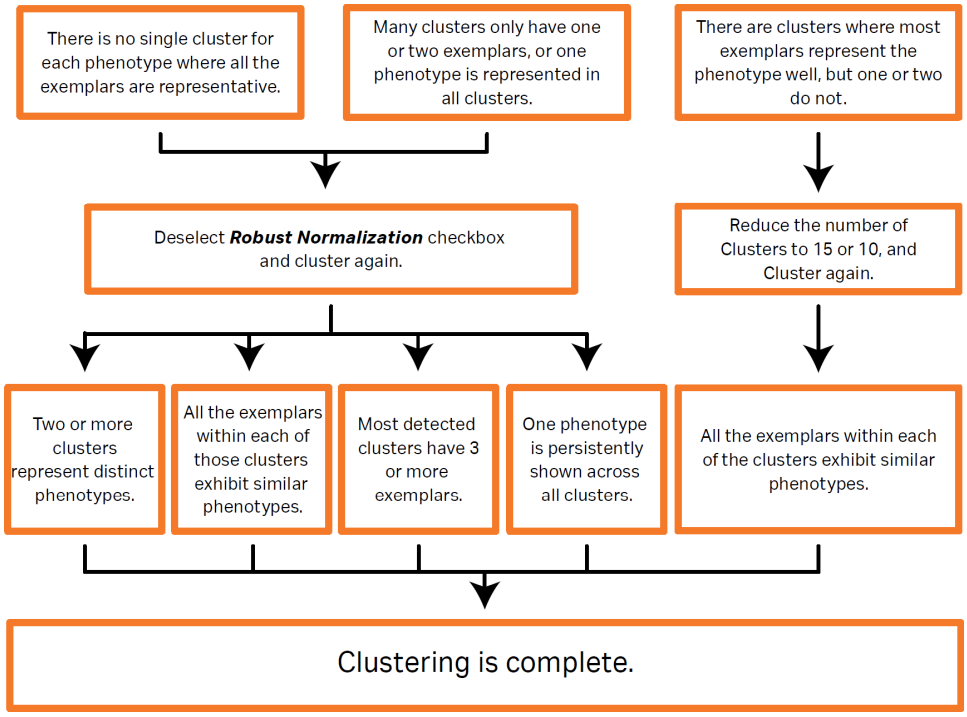
Cluster Parameters

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 14 for details. **Tip:** After clustering, similar phenotypes can be combined during the data labeling step. See Data Labeling on page 19 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

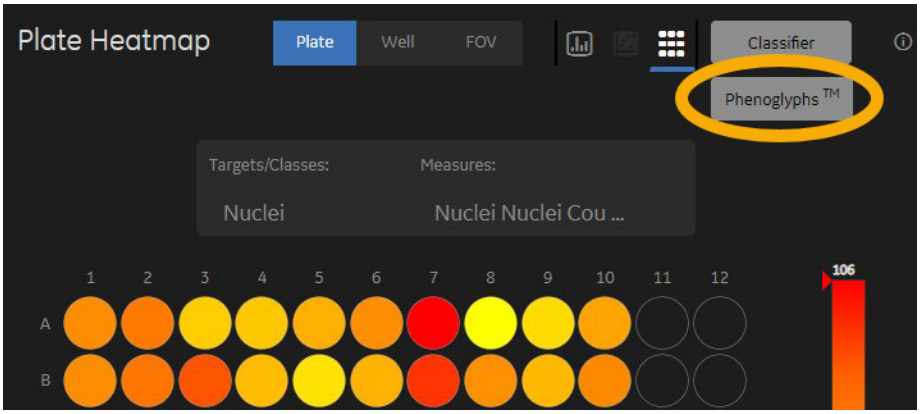
The following flow diagram describes the clustering results and the recommended actions to improve clustering.



Creating an Analysis Protocol

Do the following to create an analysis protocol in any IN Carta application and run analysis in **Interactive** or **Batch** mode. This is performed after segmentation of the data set and from the **Analysis Dashboard** | **Plate Heatmap**.

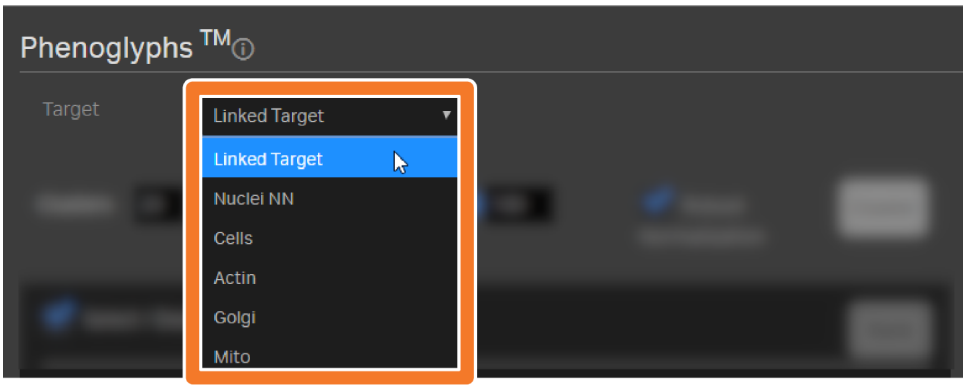
1. Click **Phenoglyphs** to start the classification workflow.



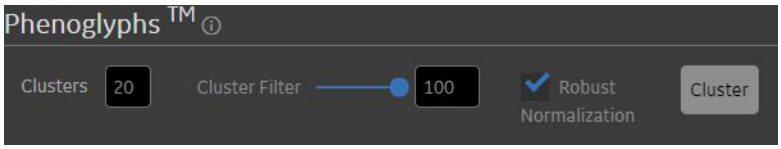
2. Select the intended target from the **Target** drop-down list.

Note:

- The default setting is to use **Linked Target**.
- If desired, it is possible to select a specific target to use during classification.

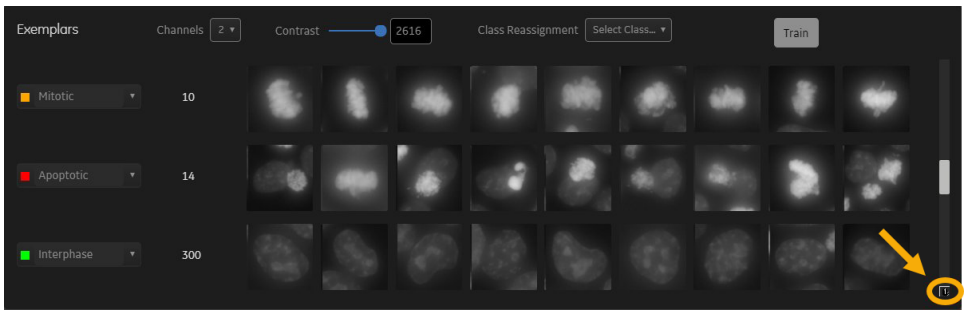


3. Adjust settings, if needed.



4. Click **Cluster**.
   When clustering is complete, each row in the exemplar pane represents a phenotypic cluster identified by the algorithm. In each row, exemplars are displayed from left to right according to how well they represent the phenotype, with the left-most exemplars most strongly representing the phenotype.
5. Check the clustering quality by viewing the exemplars.

> **Tip:** Clicking and dragging the box highlighted in the bottom right corner allows you to resize the exemplar viewport.



6. Click on an exemplar to view the cell within a field of view viewport for more context.
7. Check clustering quality based on the guidelines in Clustering Result Guidelines on page 16.

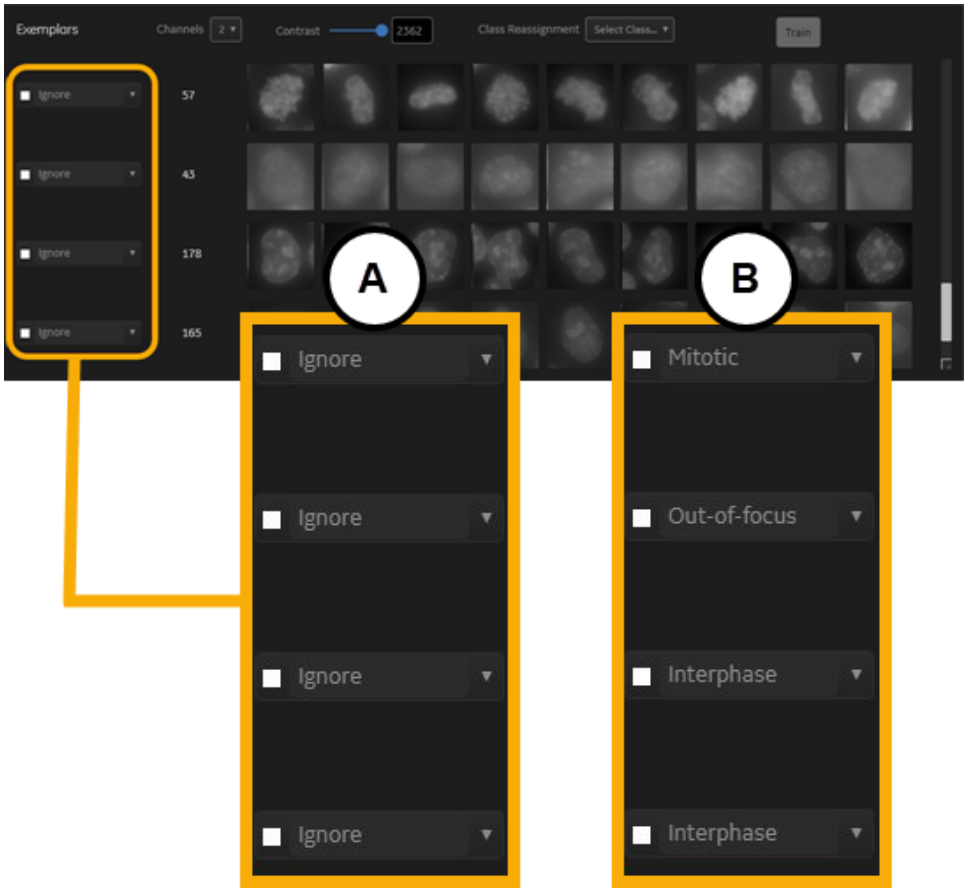
The goal of the first round of data labeling is to identify and rank features that are important for the supervised learning algorithm.

Note: It is not necessary to label all rows of exemplars.

Labeling Data

To label data:

1. Visually identify clusters where 100% of the exemplars represent a clear phenotype.
2. Double-click the class label in the drop-down lists, as in A, replacing **Ignore** with an appropriate name for the class, as in B.



Note: Once a class label is created, it can be re-used for any other **Cluster** with the same phenotype.

3. Choose **Ignore** for clusters that do not clearly show a consistent phenotype.
4. Create a separate class for debris, out-of-focus cells, and cropped cells at the border.

   Tip: Cells at the border can also be removed at the segmentation step using the **Exclude cells touching the edges** check box in the **Filters** section of analysis settings.

5. Once classes of interest are created, click **Cluster**.

   Note: At least two labeled classes are needed to proceed with clustering.

   Clustering is repeated based on the selected subset of measures from the chosen representative clusters.

Once clustering is complete, continue to the measure selection step in the next section.

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 1 to 100) show the significance of each measure. The higher the ranking score, the more important the feature is to separate the labeled classes.

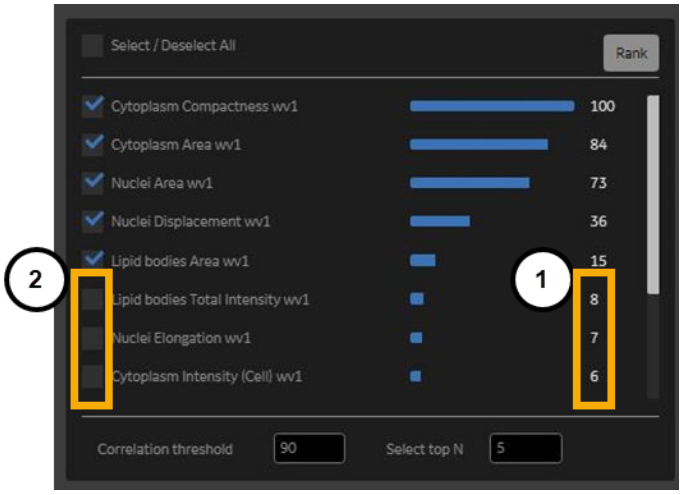
By default, all features are included when separating the classes, even low-ranking ones. Removing low ranking and highly redundant features helps to increase the speed of classification and prevent the model being excessively specific for a limited number of objects in the data set (model overfitting).

Measure Selection

Consider the following recommendation about measure selection:

- There is no specific number of measures that must be present for training.
- Limit the number of selected measures to 4 * L, where is L the number of labels/phenotypes.
- Consider “leveling off” of Rank values.

The following illustration is an example of the Rank values in measure selection.

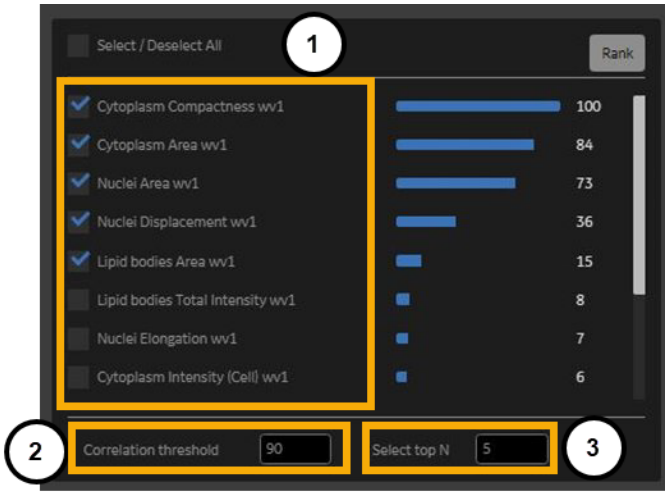


The last three measured Rank values (1) level off reaching a plateau.

The plateaued exemplars are deselected (2) and removed from training.

Controlling the Number of Features

There are three ways to control the number of features used for classification, which can reduce redundancy. Options to control the number of features are shown in the table below.



Part	Method	Description
1	Manually	Manual selection/deselection of features, one-by-one.
2	Correlation Threshold	Removes redundant features (if two features have similar trends, one of them may be omitted from the list of measurements used to generate a prediction model). Redundant features have high correlation (>90). Removing highly correlated features prevents overfitting of a model during Train process For example, setting a threshold to 90 would: • Refine which pairs of features are correlated by over 90%. • Retain only the higher-ranking feature within any of the pairs.
3	Select Top N	Selects a specified number of the most important measurements that contribute to class separation.

Ranking and Editing Measures

To rank and edit measures:

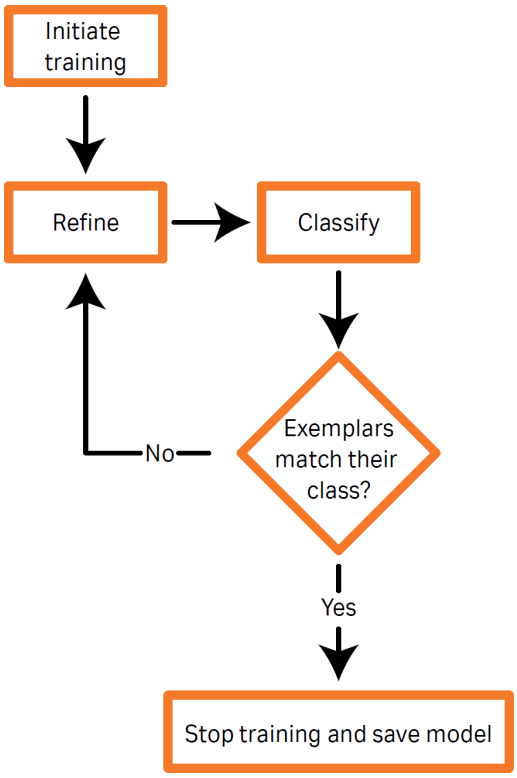
1. Click **Rank** to sort the features based on their importance.
2. Examine the Rank scores.

> **Tip:** A typical distribution has a handful of measures that score very highly (>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 empirical 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. Deselect redundant measures.
4. Repeat the Cluster Data steps (see Creating an Analysis Protocol on page 17) and Label Data steps (see Labeling Data on page 19).

After ranking and editing measures, proceed to training the model in the next section.

The following flow diagram shows the steps to train a model.



Consider the following when training a model:

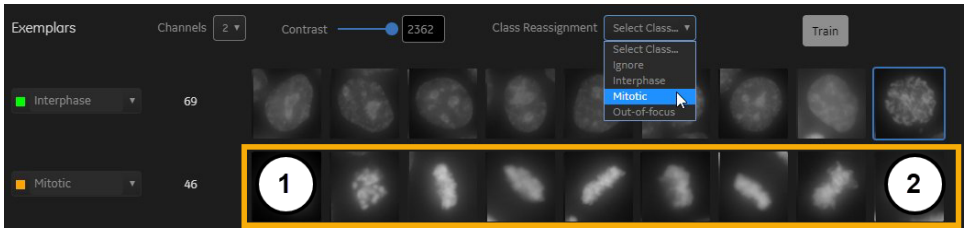
- If you are unsure about which phenotype to classify the exemplar, select **Ignore**.
- If a specific exemplar is not very clearly representative of the phenotype, label it **Ignore**. When given ideal examples of the phenotype, the algorithm is better suited than the human eye to assign cells close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

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

- **Ignore** any out-of-focus cells, overlapping cells, or cells that extend past the image boundary.

Exemplar Position

The following table clarifies the position of the exemplars.



Part	Position of Exemplars	Clarification
1	Left	Exemplars are more representative of the class.
2	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 requires that the process be redone starting at the cluster and training steps.

When to Stop Training

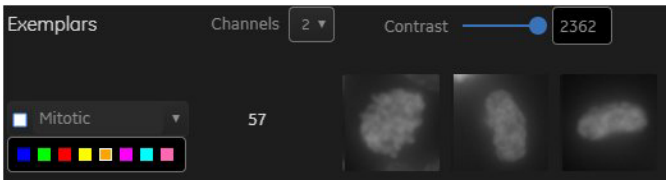
The following table helps you determine when to stop training.

Observation	Recommended Action
All exemplars match the class they are in after two or three iterations of training.	Stop training and proceed to saving the model.
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.	
Classes are not improving (that is, two or three exemplars from a class match another existing class) after 10-15 training cycles.	Restart the process with different clustering parameters, including: • Measure Selection • Normalization Type • Cluster Number

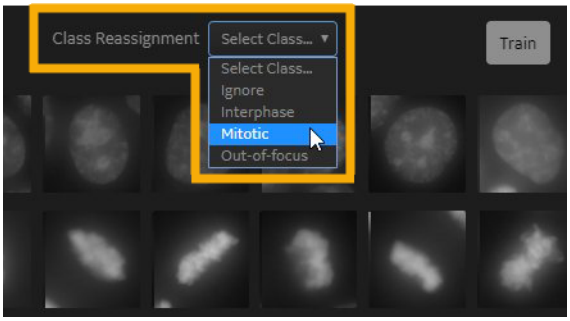
Training a Model

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 list to reassign it to another class or **Ignore** it.



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 26 for details.

Chapter 8: Saving and Applying a Model

Saving and applying the model applies the IN Carta Phenoglyphs classification model to the entire data set and saves it as a part of the current protocol for use with other data sets.

Class Information

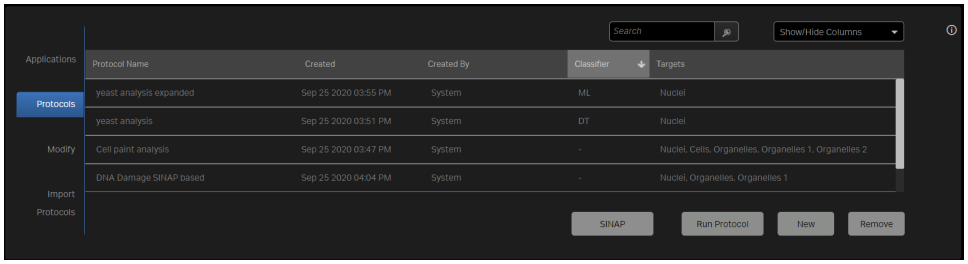
The following illustration shows class information after saving a model.

Part	View	Description
1	Data Table	Views single objects.
2	Plate Heatmap	Visualizes the percentage of cells in a class or plots a measure as a function of class.
3	Field of View	Viewport using the class-specific mask colors.

Saving and Applying a Model

Do the following to save and apply the model:

1. When training is complete, click **Save Model**.
2. Protocols with IN Carta Phenoglyphs-based classification can be executed in **Interactive** or **Batch** mode (see the respective info-panel for details). Protocols with an IN Carta Phenoglyphs classifier will display **ML** in the **Classifier** column.



3. View class information once processing is complete.

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 B

MOLECULAR DEVICES