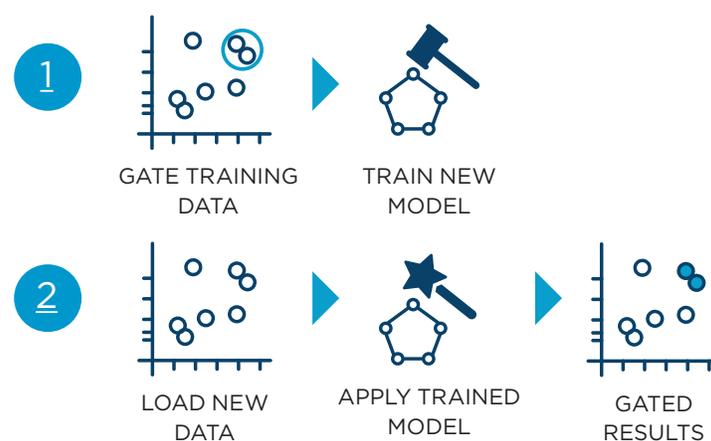




How to establish and evaluate machine learning-assisted automatic gating to improve reproducibility and reduce time spent on your flow cytometry data analysis

Objective

- Understand how to perform machine learning-assisted automatic gating to identify populations/subsets present in flow cytometry data
- Train a computational flow cytometry gating model on the Cytobank platform
- Learn how to assess the quality of an automatic gating model
- Apply a trained model on new data to infer the population affiliation of every event
- Recognize how pre-processing steps can affect the quality of machine learning algorithm analysis results



Manual gating has long been identified as a major source of variability for flow cytometric analysis (Maecker et al. 2005). Centralized data analysis as well as standardized gating strategies and stringent standard operating procedures have been suggested as means to increase the reproducibility but are time-consuming (Macchia et al. 2020). Machine learning-assisted analysis of cytometry data has proven advantageous, however, there is a lack of methods that allow users to automate their own user defined gating strategy for their specific marker panel in a hypothesis-driven setting (Hu et al. 2021). Here, we present an automatic gating approach that allows you to train a model on your manual gating strategy on a small number of samples without requiring coding skills. Automatic gating can reduce inter-operator variability of manual population identification and streamline the data analysis workflow. **(Figure 1).**

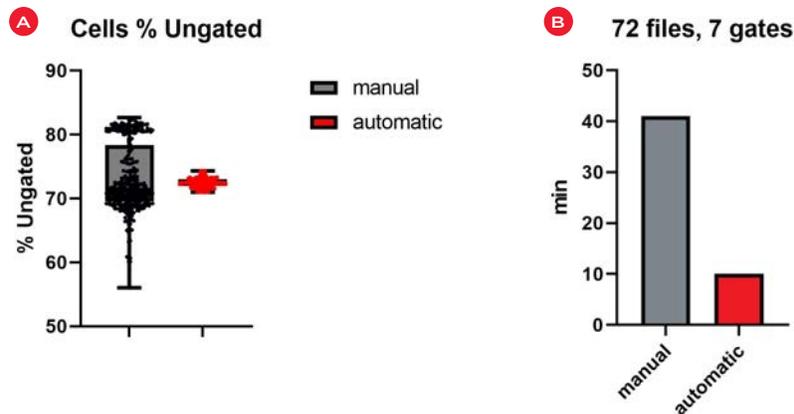


Figure 1. A) 96 technical repeats were analyzed by four independent operators (mean 73.06%, CV 7.3%). An automatic gating model was trained on a subset of files (n=24) analyzed by operator 1 and subsequently applied to the remaining 72 files by all 4 operators (mean 72.47%, CV 1.02%). **B)** Time required for training and application of automatic model is 75% shorter compared to manual gating of the full dataset for population identification in 72 files with 7 gates.

The Cytobank platform is a cloud-based analysis software with integrated analytical algorithms, as well as a structured and secure content management system for flow cytometry and other single-cell data (Kotecha et al. 2010). Cytometry data analysis includes two different approaches, the discovery of populations or expression patterns that have not been previously described and the hypothesis-driven detection of known cellular subsets. The algorithms and tools integrated in the Cytobank platform cover both scenarios and leverage the scalable compute and collaborative power of the cloud. This way, large analyses can be done in parallel, and the cloud-based storage provides the capability to automatically archive and easily share these data securely and safely.

In this application note, we illustrate how machine learning-assisted automatic gating algorithms can help to faithfully replicate manual gating to analyze flow cytometry data in a hypothesis-driven manner using example data from a standardization experiment consisting of control material acquired at laboratories with a dry reagent panel for basic immunophenotyping.

Automating your manual gating strategy on the Cytobank platform

Machine learning algorithms are increasingly used for the analysis of cytometry data. The most frequently used solutions in this context assist researchers in the unbiased discovery of structures in their high-dimensional data. In many cases, these algorithms are unsupervised, meaning that they do not require manually annotated data. However, they may not be fit-for-purpose in a hypothesis-driven setting, when scientists need to control the criteria based on which a population is defined.

Supervised machine learning approaches, which automatically assign cells or events to a defined population, learn how to do so from a so-called training set of annotated, manually gated data. In contrast, unsupervised machine learning algorithms infer structures from the information contained in the transformed multi-dimensional cytometry data.

The Cytobank automatic gating functionality allows users to define a set of populations to be identified and the criteria for how they should be gated. It includes two computational steps (**Figure 2**). In a first training step, the results of manual gating are used to train a model. The model “learns” how to replicate the population assignments users would typically make. In a second so-called inference step, the model is applied to new data and categorizes the events into the different populations without manual gate adjustments.

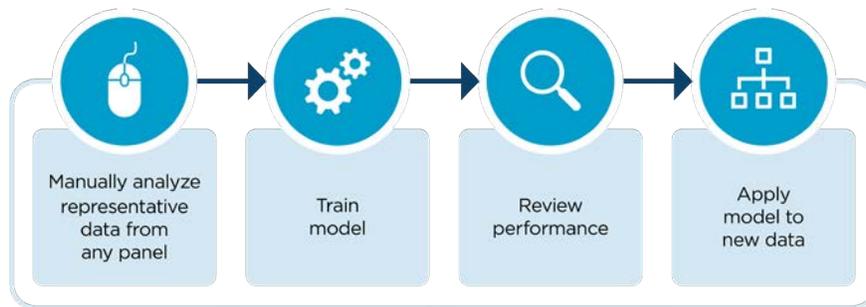


Figure 2. Workflow of Cytobank Automatic Gating.

At least 10 experimental files that have been manually gated are needed to train a model. Once you have established your gating strategy and made the necessary tailoring by file and/or population, apply the gates. You can initiate model training by going to the **Advanced Analysis > Automatic gating: Select Train new model > Create**. If you have previously trained a model or another Cytobank user has shared an experiment containing a trained model with you, you can start an inference run to assign the cells in new files to the populations the model has been trained on.

Model training with the Cytobank automatic gating functionality

During model training you can establish a model that is specific for your panel and gating strategy and that can subsequently be used for the automatic analysis of new data acquired using the same marker panel. A general key performance indicator (KPI) for the model is generated to help you compare models and you can drill into the performance at a population level.

Model training, validation and testing

For the training of supervised machine learning models, data is typically split into training, validation and test sets (Ripley 1996).

The training set is used to identify which settings or hyperparameter values lead to a model that best fits the training data.

The validation set is a part of the data that is excluded from model training and used to give an estimate of the model performance while the model's hyperparameters are being tuned.

The test set is used to assess the performance of a fully trained model. As the files in the test set were not used to train the model or to evaluate the performance in the context of hyperparameter tuning, it can be used to understand how accurately the model can predict populations in new data to understand the generalization error. On the Cytobank platform, the test set is used to determine the KPI of the model.

Files selected for training setup will automatically be split into training, validation and test set at a ratio of 60% of files for training, 20% for validation, and 20% for testing (**Figure 3**).

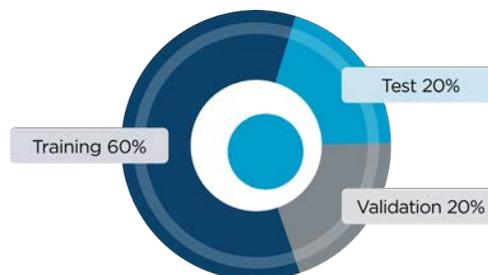


Figure 3. Distribution of files selected for model training.

For the training of a model that performs well on new data it is important that any variability present in the new data is also reflected in the training data. This includes technical variability such as batch effects from different reagent lots or varying instrument settings, and biological variability such as individual differences in marker expression levels. Also, the distribution between training, validation and test set should be comparable. The Cytobank automatic gating model training includes an automated step to support this. The files selected for training setup are clustered into a user-defined number of optimal clusters. Files for training, testing and validation are then sampled equally from the resulting clusters.

Tip for success

- Use as many samples for training as feasible.
- Perform gating as accurately as possible on your training data.
- Choose training data to reflect as much of the expected variability from sources like reagents, instrument settings or biological differences as possible.

Validation of supervised machine learning models

The average number of files in cytometry experiments is often considered low for training of machine learning models, with hundreds of files instead of thousands. While k-fold cross-validation is often positioned as the optimal choice for model validation, it tends to be unstable for small sample numbers. Therefore, the sample clustering approach was implemented as part of the Cytobank automatic gating functionality.

Population selection

The Cytobank automatic gating functionality allows you to select the populations to be included in the automatic gating model. Note that any populations higher in the hierarchy than the selected ones will automatically be included, as they are required for the step-wise assignment of cells to populations. All channels used in the population hierarchy to define the selected populations will be taken into account when training the model.

By default, all populations are considered equally important. When setting up model training you can assign a higher weight to one or several populations. By increasing the weight of populations of interest the performance on that particular population has more impact on the overall performance KPI calculation. When training and comparing several models by increasing the magnification, increasing the weight for populations you care most about will help you select a model that performs well on these populations.

Event sampling

The training of machine learning models can be compute-intensive and time consuming. To reduce the time required for large training datasets, the Cytobank automatic gating functionality supports up to 4 million events for model training and allows you to apply a downsampling method.

The event sampling offers 3 options:

- **Equal:** the same number of events is used from every sample.
- **Proportional:** Each population/file combination present will be sampled according to its relative abundance as a function of total available events.
- Use all events.

The downsampling for automatic gating model training optimizes maximum population diversity across all training set samples. It does not randomly select events, but sample events from all manually gated populations to ensure that events from all populations are included for training. This reduces the risk of losing low-frequency populations (**Figure 4**).

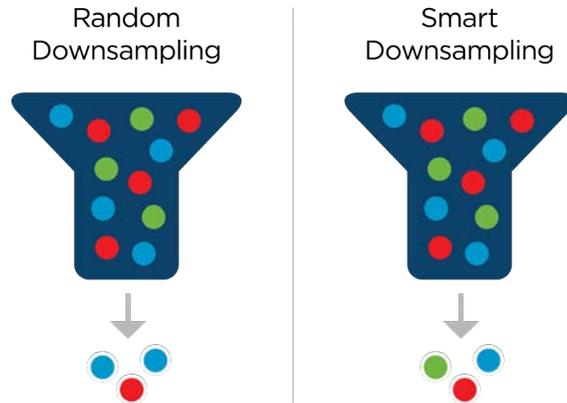


Figure 4. Smart downsampling ensures representation of rare populations.

Training magnification

Cytobank automatic gating model training explores a defined search space for the best model. By increasing the magnification, users can determine how many different models are being trained using different parameters on the same training data. The model with the highest KPI will be returned to the user. By increasing the weight of a population in a training run with a magnification greater than 1, you may be able to influence the model selection to return a model performing better on your population of interest. Of note, increasing the magnification also causes a proportional increase of the runtime.

Setting a seed

Many decisions in training machine learning models involve a random component. These include the sampling of training data as well as the parameters used for model training. To ensure repeatability of training results, a user-defined seed can be set instead of starting with a random seed. This allows the same decisions for sample and parameter selection to be made across training runs. In general, it is advisable to start with a random seed.

Tip for success

The Cytobank platform saves the seed value for training runs on the setup page after the run is completed to document the information and keep it available to users.

Create experiment with blind test files and their inferred populations

The Create experiment with blind test files and their inferred populations option will make it easier to visually evaluate model performance.

Data scaling and compensation

The automatic gating model learns from patterns in the compensated and scaled data. Like with all cytometry data analysis, it is important to make the necessary adjustments to account for fluorescence spillover in flow cytometry data and to scale data appropriately. The quality of compensation and scaling should be comparable between the dataset used for model training and any new dataset the model is used to conduct inference.

Please make any desired adjustments on compensation and scaling on the setup page before selecting **Run Autogate training analysis**.

Training duration

The time required for training depends on the number of files and events to be included and on the number of populations. It will also increase linearly with the magnification (**Figure 5**).



Figure 5. Training time depends on the number of populations included, magnification 1.

Evaluation of model performance

Once model training is complete, the training setup page is updated with information about the algorithm runtime and the KPI of the model (**Figure 6**).

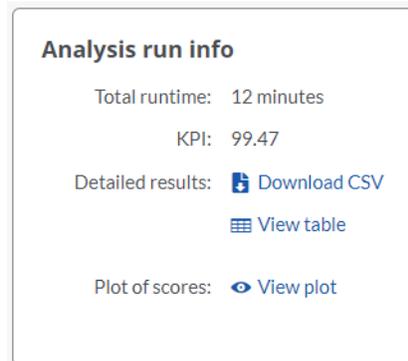


Figure 6. Cytobank Automatic gating training analysis run info.

The KPI for an autogating model is the weighted average F1 score across all predicted populations. The F1 score is the weighted average (harmonic mean) of precision and recall. Precision measures the percentage of events predicted to be part of a population that are actually true positive events and by this takes into account any false positive events. Recall is a sensitivity measure and indicates the percentage of true positive events that are accurately predicted. Therefore it takes into account any false negative events.

To determine the KPI of a model, for every event of the test set and for every population that is part of the model training, the Cytobank platform determines if the user classified the event as belonging to a particular population and if the autogating model classified the event as belonging to this population or not.

As an output of model training, you receive a table that shows the following information for every population for the test dataset (**Figure 7**).

Population	Actual	Predicted	True positives	Precision	Recall	F1 score
Singlets	231600	231608.0	231539.0	99.97	99.97	99.97
CD45+	207150	207152.0	207053.0	99.95	99.95	99.95
Cells	207345	207356.0	207254.0	99.95	99.96	99.95
Lymph	74682	74688.0	74617.0	99.9	99.91	99.91
CD4+	34655	34656.0	34622.0	99.9	99.9	99.9
CD45+ CD3+	58349	58356.0	58294.0	99.89	99.91	99.9
CD14+	21049	21048.0	21022.0	99.88	99.87	99.87
CD8+	19901	19907.0	19869.0	99.81	99.84	99.82
CD19+	7680	7679.0	7662.0	99.78	99.77	99.77
CD3-CD19-	8444	8452.0	8422.0	99.65	99.74	99.69
CD56+ CD16+ NK cells	6420	6435.0	6404.0	99.52	99.75	99.63
CD56bright NK cells	552	554.0	541.0	97.65	98.01	97.83

Figure 7. Table of detailed Cytobank Automatic gating training analysis run results.

Actual: the number of events included in this population based on manual gating.

Predicted: the number of events included in this population based on automatic gating.

True positives: number of events that have been defined manually as belonging to this population and that have been automatically defined as belonging to this population, overlap between actual and predicted.

Precision: number of true positive events divided by the number of all predicted events in an automatically defined population in %; high precision indicates a low number of false positives or type I errors.

Recall: number of true positive events divided by the number of actual events in a manually defined population in %; high recall indicates a low number of false negatives or type II errors.

F1 score: harmonic mean of precision and recall for a population, it balances precision and recall and is calculated as

$$2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}$$

As a rule of thumb, F1 score above 90% are considered *very good*, those between 80% and 90% as *good*, those between 50% and 80% as *OK*, and scores below 50% as *not good*.

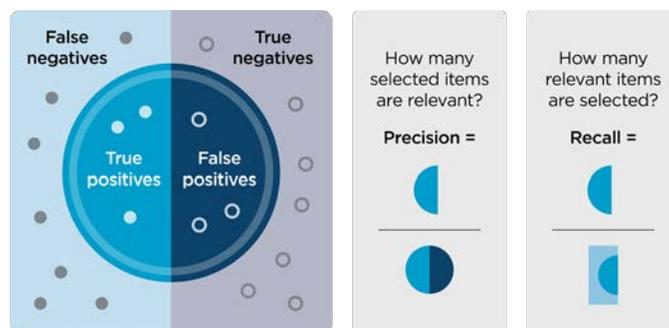


Figure 8. Visual summary of Precision and Recall.

In addition to the tabular output, you can also review box plots of the prediction scores by population showing the distribution across the files in the test set (**Figure 9**).

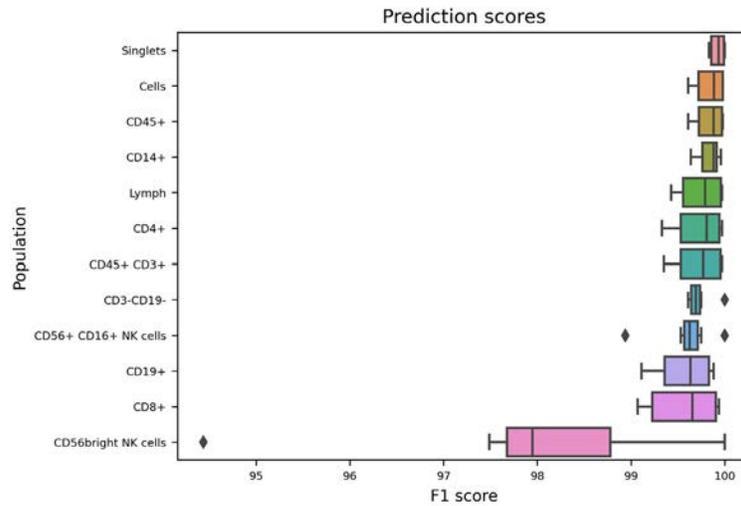


Figure 9. Plot of scores from analysis run info.

If you decided to automatically create a child experiment containing the blind test files by checking the relevant checkbox, you can access the created experiment from the autogate analysis training setup page.

The automatically created child experiments contain the subset of FCS files that were assigned to the blind test set. For every predicted population, the files now contain one additional parameter following the naming convention of auto_gate_Population name (**Figure 10**).

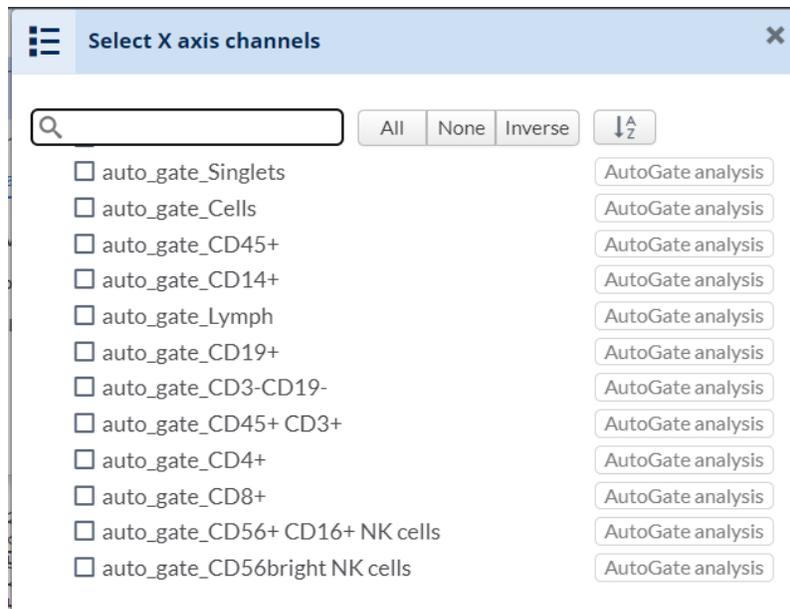


Figure 10. Channel list in illustration layout showing channels generated by automatic gating.

Populations are identified by cluster gates on the added channels following the naming convention of Population name_auto_cluster_gate (Figure 11).

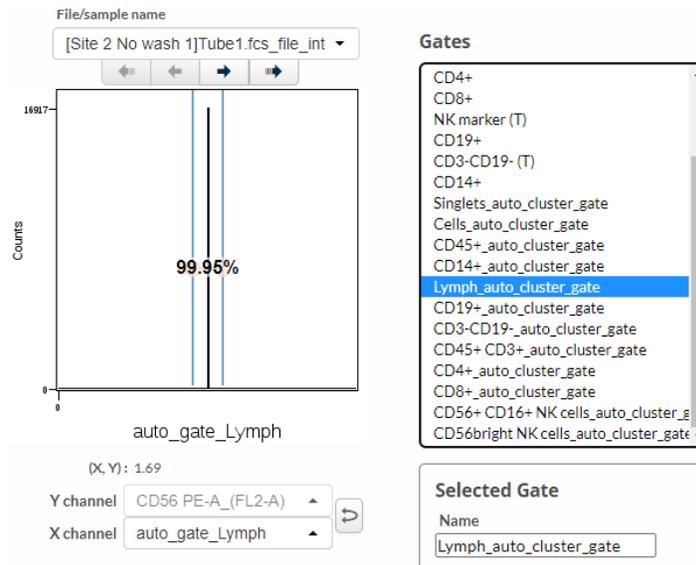


Figure 11. Cluster gates for automatically gated populations visualized in the Gating editor.

The population editor will show the automatically predicted population as Population name [auto] (Figure 12).

Population \ Gate	Singlets	Cells	CD45+	Lymph	CD45+ CD3+	CD4+	CD8+	NK marker (UR)	NK marker (UL)	NK marker (LL)	NK marker (LR)	CD19+	CD3-CD19-	CD14+	Singlets_auto_cluster_gate	Cells_auto_cluster_gate	CD
Lymph	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>										
Lymph [auto]	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	

Figure 12. Population manager showing manually defined and automatically gated populations.

To evaluate the performance of the autogating model within the Cytobank platform using summary charts, comparing the event count between the manually gated population and the automatic population as subgroups may be useful. A default illustration named Autogating boxplot of event counts comparing all manual and automatic populations is generated automatically (Figure 13).

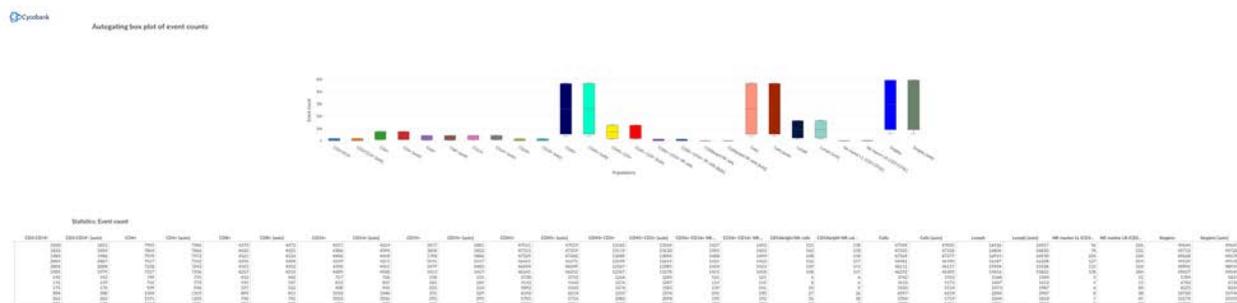


Figure 13. Example of default illustration generated for autogate model training – Autogating box plot of event counts.

To evaluate the performance for a specific population, clone the default illustration and deselect all additional populations, or generate a new summary chart.

In a boxplot comparing manual and automatic results for one population, toggling **Show dots** and **Show lines** to **on** visualizes additional information about any bias of the model. The statistics of this Illustration can be exported as a .csv file and used for the visualization and calculation of the correlation between manual and automatic results in a spreadsheet program (**Figure 14**).

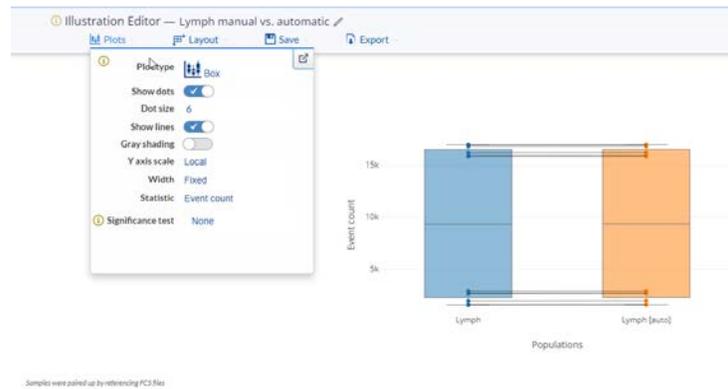


Figure 14. Illustration Editor Plots menu selection to compare automatic and manual population event counts.

To visualize the results on a per event basis, overlay dot plots can be used. Use populations in the overlaid dimension and select the automatically identified population of interest and the relevant parent population. If you select the same channels on the x- and y-axis that have been used for creating the manual gate, you can toggle on **Show gates** for a direct comparison between manual and automatic identification. Use plot size and **Dot size** to improve the resolution (**Figure 15**).



Figure 15. Visualization of autogating populations in a dot plot overlay.

F1 score vs. R²

Automatic gating of cells or events in a cytometry assay is a classification problem. Classification describes the act of identifying to which class or category or population a cell or event belongs. Widely used performance metrics for machine learning classification models include precision, recall and the F1 score. The F1 score takes into account false positive and false negative classification of events per population.

In contrast, R² is typically used as a performance metric for regression testing. It describes how much variance of a dependent variable can be inferred from the amount of variance of an independent variable. It explains the degree to which the input variable, such as the manually gated event count of a certain population in a number of samples, describes the variation of the automatically gated or predicted event count of this population. When evaluating semi-quantitative cytometry assays, new analytical approaches are often compared to established methods. This can be done by measuring the correlation and calculating the bias.

Therefore, the F1 score is an appropriate metric to describe the performance of an autogating model that is used to classify cells or events into population categories. To further characterize a cytometry assay, the determination of correlation between two analytical approaches, e.g., manual vs. automatic, can be valuable, as this is an established concept that many researchers in the field are familiar with.

Tip for success

If your initial model training shows poor performance for a population you are especially interested in, consider repeating model training with a random seed. If the population continues to show poor performance, review the quality of your manual gating.

Applying an automatic gating model on new data

To use a trained model to predict the populations in new FCS files, open the experiment that contains the new FCS files. It does not matter whether this is the same experiment you used to train the model or if it is a separate experiment. Please ensure that data is appropriately pre-processed with respect to scale adjustments, compensation and panel assignment/channel names.

Go to the **Advanced Analysis > Automatic gating: Select Run inference using model > Create**.

On the setup page, select the FCS Files you want to include.

As soon as at least one FCS file has been selected, you will be able to select an automatic gating model to apply from the drop-down menu in the Settings section.

You may choose to clone the gates from the current experiment to the child experiment containing the automatically assigned populations. This can be helpful to compare manual and automatic results while you are evaluating the automated analysis approach.

Select the green **Run Autogate** inference analysis button at the top of the setup page to start population prediction. You will receive an e-mail notification once the child experiment has been created. Runtime will depend on the number of events and number of populations included (**Figure 16**).

Once inference is complete you can access the results by clicking the link in the notification e-mail or via the experiment navigation bar and the **Experiment Details** page.

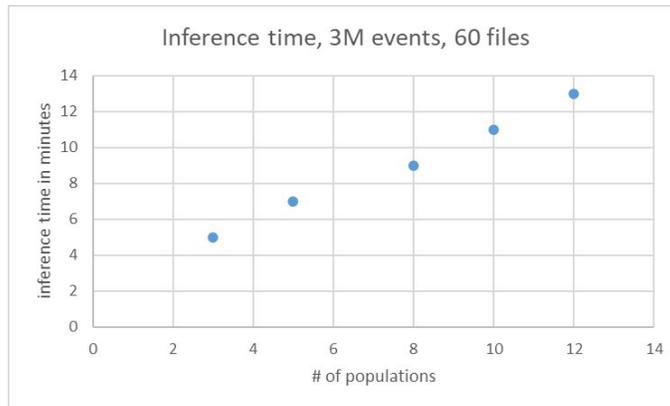


Figure 16. Inference analysis run time per number of populations.

The automatically created child experiment contains the subset of FCS files that were included in the inference analysis. For every predicted population these files now contain one additional parameter following the naming convention of auto_gate_Population name.

Populations are identified using cluster gates on the added channels following the naming convention of Population name_auto_cluster_gate. The population editor now shows the automatically predicted populations as Population name [auto] (**Figure 17**) and the populations are available in the Illustration editor.

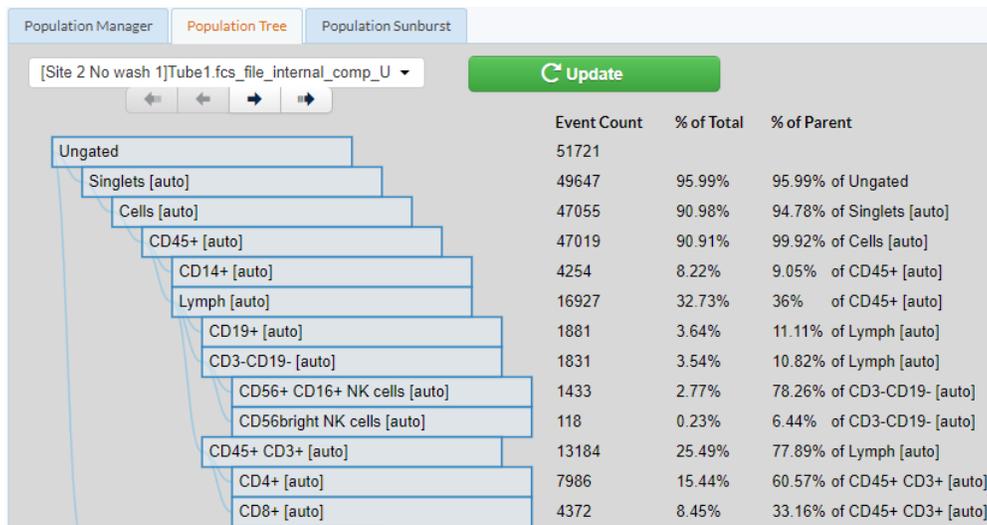


Figure 17. Automatic populations in Population Tree.

To visualize the autogating results at a single-cell level, dot plot overlays as shown in Figure 15 can be useful. Automatic populations can be used just like manually defined populations.

Figure 18 shows the results of two samples analyzed automatically:

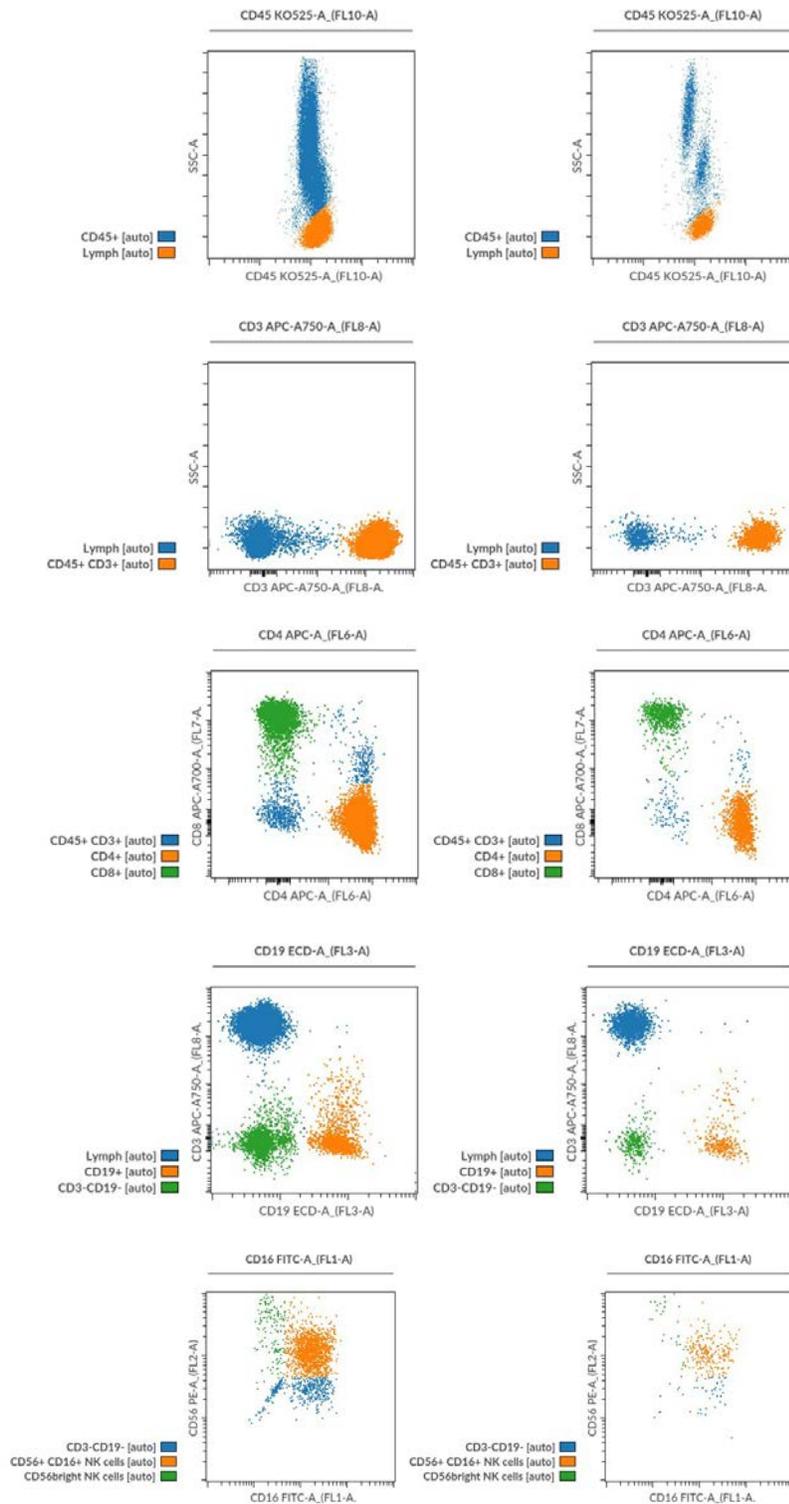


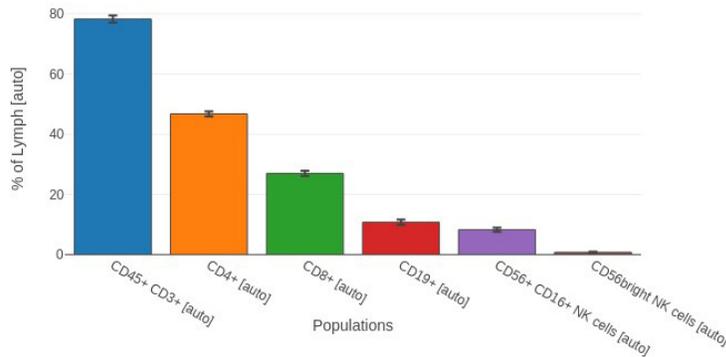
Figure 18. Commercially available control cells, a liquid preparation of stabilized human erythrocytes and leukocytes to verify the accuracy/reproducibility of the steps involved in immunophenotyping were stained using a DURAClone IM Phenotyping Basic Tube kit,* an expert-designed dry antibody panel for clinical research studies, which eliminates error-prone antibody pipetting and extensive reagent inventory management. (Beckman Coulter PN. B53309) A model was trained on 60 files and results visualized for two files from the test set.

*DURAClone IM Phenotyping Basic Tube is for Research Use Only, not for use in diagnostic procedures.

Summary plots and the associated statistical table are useful to generate summary visualizations (Figure 19).



Lymphocyte subsets



Statistics: Percent of population for Varying by Subgroups of Lymph [auto]

CD45+ CD3+ [auto]	CD4+ [auto]	CD8+ [auto]	CD19+ [auto]	CD56+ CD16+ NK...	CD56bright NK cel...
77.887	47.179	25.829	11.112	8.466	0.697
78.049	46.794	26.3	10.839	8.346	0.821
77.342	47.094	25.54	10.667	8.854	0.874
77.857	46.471	26.53	9.545	8.835	0.845
77.756	45.473	27.323	9.304	8.934	0.885
78.233	46.366	26.893	8.956	8.962	0.929
81.124	47.664	27.904	9.785	7.639	0.379
79.789	47.923	27.092	11.407	7.13	0.372
79.769	47.106	26.774	11.223	7.096	0.453
78.294	45.373	27.898	11.318	7.912	0.894
78.932	45.335	27.916	11.099	7.223	0.677
80.755	46.701	28.519	9.238	8.211	0.33

Figure 19. Visualization of autogating population percentage in a bar graph.

Troubleshooting your automatic gating

No populations are available for model training

The Cytobank platform stores two principal versions of your gates. One version is what you see at any given time in the Gating Editor. These are termed Scratch Gates. The second version of your gates are what the rest of your experiment is working off of at any given time. These are termed Experiment Gates. These gates are used for statistics, illustrations, algorithms such as Automatic gating and everything that is not in the Gating Editor.

When actions are taken in the gating interface, differences accumulate between the Scratch Gates and the Experiment Gates. These differences could be small, as in the case of simply nudging a gate, or they can be large, as in the case of a long session of gating. The apply button sets the Experiment Gates to the state of the Scratch Gates.

If your populations are not available in the setup page for an Automatic gating model training, return to the Gating Editor and select **Apply** to update the Experiment Gates to the current state of the Scratch Gates.

A single population shows an F1 score of 0 for the trained model

The Automatic gating algorithm introduces randomness at several stages of model training. In some cases, a model will fail to recognize individual populations. When this happens, select Copy settings and re-start the training analysis with a random seed.

Individual populations show high variability in the plot of scores for the trained model

A high level of variability is often observed for populations that are challenging to gate manually. In this situation, consider reviewing and iterating your manual gate positions to improve the consistency of the training data.

A single or a few populations show a worse F1 score than the rest for the trained model

Populations with modulated expression patterns are often difficult to gate manually. This may lead to lower F1 scores. Consider reviewing and iterating your manual gate positions to improve the consistency of the training data.

As previously discussed, the automatic gating algorithm introduces randomness at several stages of model training. In some cases, a model will show lower performance for individual populations. When this happens, select Copy settings and re-start the training analysis with a random seed. You can also increase the weight of the affected populations to skew the model training toward favoring these populations.

When running an inference analysis, some populations fail to be predicted

Any model learns the structure of the data during training. The structure of the data is influenced by the parameter scales and compensation settings. It is not necessary to use the exact same compensation matrix and scale settings. Rather, the scaled and compensated data used for training should look comparable to the data for inference analysis. In most cases when the same panel is used and instrument settings are consistent it is appropriate to import scales from a previous experiment.

If a model fails to detect populations completely, ensure you have applied appropriate scale and compensation settings

When using a model trained on quadrant gates, events may be assigned to more than one quadrant population. The Cytobank platform treats every quadrant as an independent rectangular gate. An Automatic gating model may therefore predict an event to fall into more than one population. This is expected behavior.

For Research Use Only. Not for use in diagnostic procedures.

This protocol is for demonstration only, and is not validated by Beckman Coulter Life Sciences. Beckman Coulter Life Sciences makes no warranties of any kind whatsoever express or implied, with respect to this protocol, including but not limited to warranties of fitness for a particular purpose or merchantability or that the protocol is non-infringing. All warranties are expressly disclaimed. Your use of the method is solely at your own risk, without recourse to Beckman Coulter Life Sciences. Not intended or validated for use in the diagnosis of disease or other conditions.

© 2023 Beckman Coulter, Inc. All rights reserved. Beckman Coulter, the stylized logo, and the Beckman Coulter product and service marks used herein are trademarks or registered trademarks of Beckman Coulter, Inc. in the United States and other countries. All other trademarks are the property of their respective owners.

For Beckman Coulter's worldwide office locations and phone numbers, please visit Contact Us at

[beckman.com](https://www.beckman.com) 2022-GBL-EN-100375-v2

