

# Kaluza Analysis Tree Plots offer a fast and comprehensive overview of multi-color flow cytometry data

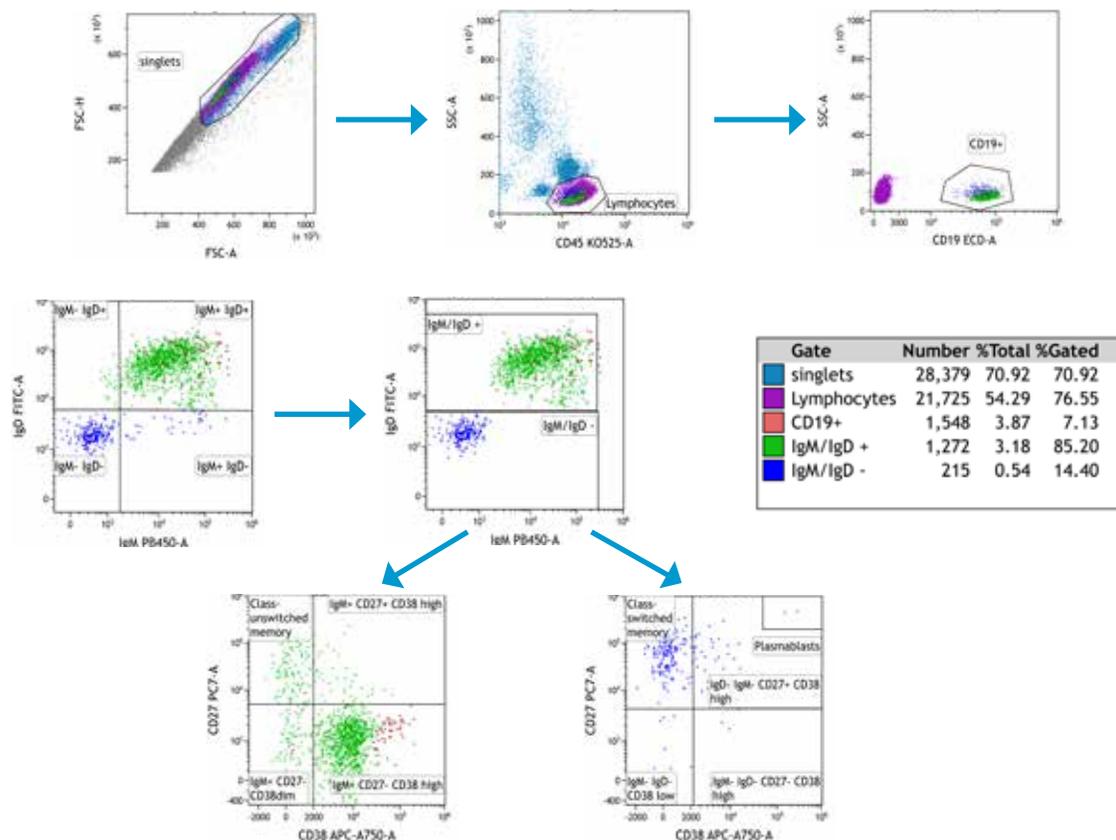
This white paper introduces the Tree Plot as a supervised approach for comprehensive deconvolution of flow cytometry data, bridging the gap between fully manual analysis and unsupervised computational analysis approaches. Step-by-step instructions for the generation of Tree Plots using Kaluza Analysis Flow Cytometry software are provided.

## Manual Flow Cytometry Data Analysis

Flow cytometry data is traditionally analyzed by sequential gating. The expression of markers is visualized one by one or two at a time and populations are identified subjectively by positioning so called gates on the displayed distributions in the histograms.

All events contained within this parent gate are then further assessed for other markers and more child gates are defined, building a gating hierarchy that sequentially excludes more and more events from the analysis, until only the population of interest remains (Figure 1) (Lugli, Roederer, & Cossarizza, 2010).

Boolean gating allows the user to combine previously defined gates using logical operators such as AND, NOT and OR to determine whether the events in the gate, or a combination of gates, are to be included or excluded.



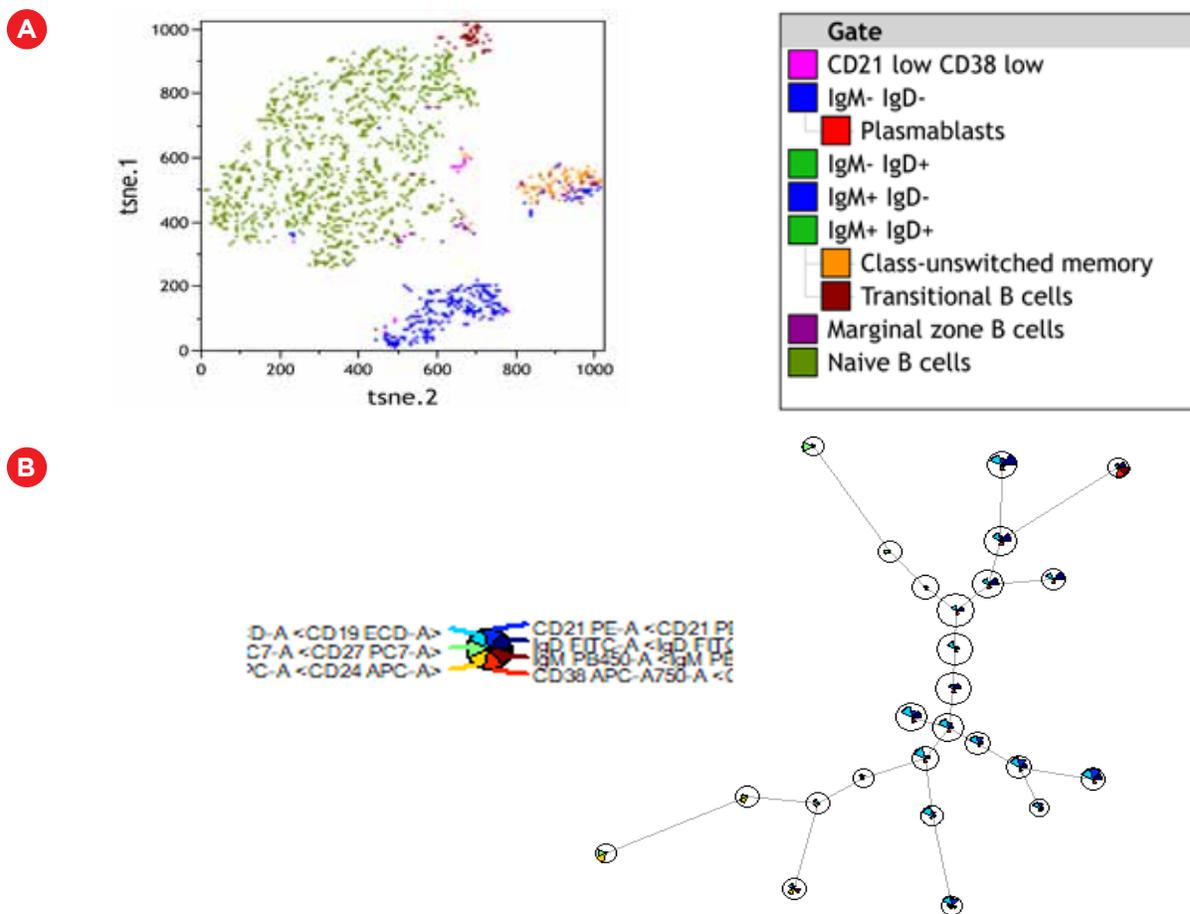
**Figure 1:** Sequential gating of selected B-cell subsets: Human peripheral blood was stained using the DuraClone IM B cells Tube according to IFU and acquired on a CytoFLEX LX. Data was analyzed in Kaluza Analysis. Cell populations are identified by sequential gating and Boolean gating.

Both approaches rely on user-defined gates and are highly subjective. Classical gating strategies have been shown to lack reproducibility and have high inter-operator variability (Maecker et al., 2005). In addition, it requires a lot of hands on time, as for a comprehensive review of all markers combinations in a 10 color panel, 45 histograms need to be evaluated (Saeys, Gassen, & Lambrecht, 2016).

Excluding events can introduce operator-specific bias. Data analysis in this case is solely driven by prior knowledge and assumptions made by the person defining the gating strategy. Populations not previously described in the literature or aberrant marker expression patterns may not be readily recognized (Irish, 2014).

## Computational Flow Cytometry Data Analysis

More recently, computational approaches such as dimensionality reduction and clustering techniques have become popular. Algorithms such as tSNE and FlowSOM (Fig. 2, example of SNE and FlowSOM) are unbiased and allow the discovery of previously unrecognized similarities between cell populations (Amir et al., 2013; Saeys et al., 2016). Clustering algorithms have been shown to match the results of manual gating and provide greater reproducibility (Aghaeepour et al., 2016). They may still be time consuming to run, but generally do not require the same amount of hands on time as manual analysis. Additionally, the commercially available solutions that do not require programming skills to run do not support the inclusion of expert knowledge that might aid in checking the plausibility of the results (Reiter et al., 2016).



**Figure 2:** Computational analysis of DuraClone IM B cells panel using the Kaluza R Console plugin. Scripts provided with the R Console plugin example files were adapted to run on CD19+ gated events. A) A tSNE analysis was performed including all fluorescence channels except CD45. Sequential gating was performed in parallel to allow for color backgating on the tSNE clusters. B) Star plots of CD19+ events after FlowSOM analysis.

## Tree Plot Visualization of Flow Cytometry Data

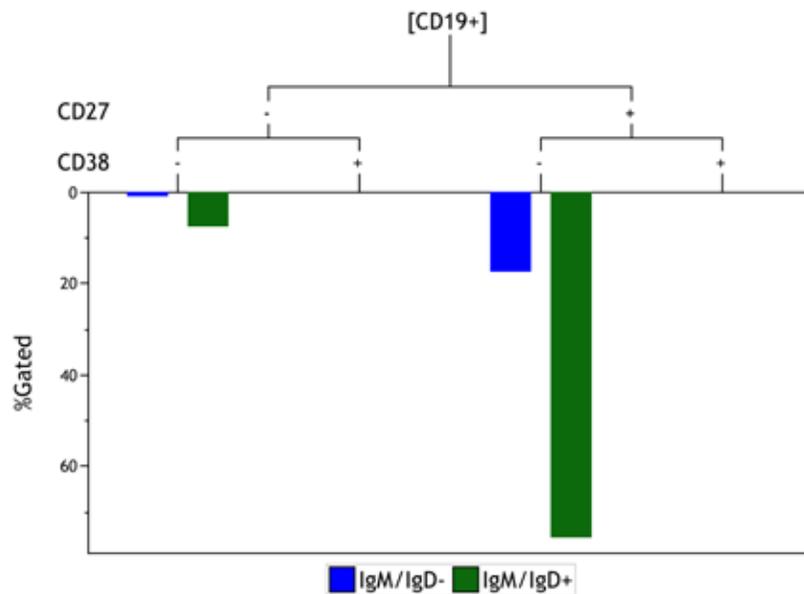
The Tree Plot offered by Kaluza Analysis software bridges the gap between analysis guided by expert knowledge, comprehensiveness and speed. It allows the assessment of physical characteristics of all events included in an analysis in a supervised manner. Unexpected marker combinations will become readily visible. Tree Plots provide a useful data comparison tool, as one Tree Plot can condense data from up to 28 bivariate plots, thus reducing the time required for data interpretation once the analysis strategy has been set up.

The Tree Plot classifies events into different classes/populations based on gating information. Events are assigned to a class depending on whether they are contained within or excluded from a given gate. The tree plot does not have any axes and thus no axis information. The maximum number of gates that can be used for classification on a tree plot is limited to eight. Since the number of classes is 2 to the power of the number of gates, with eight gates, a single tree plot can display 256 populations.

Tree Plots are especially effective when cell populations are characterized by markers showing a bi-modal or binary expression pattern.

A Tree Plot includes (Fig. 3)

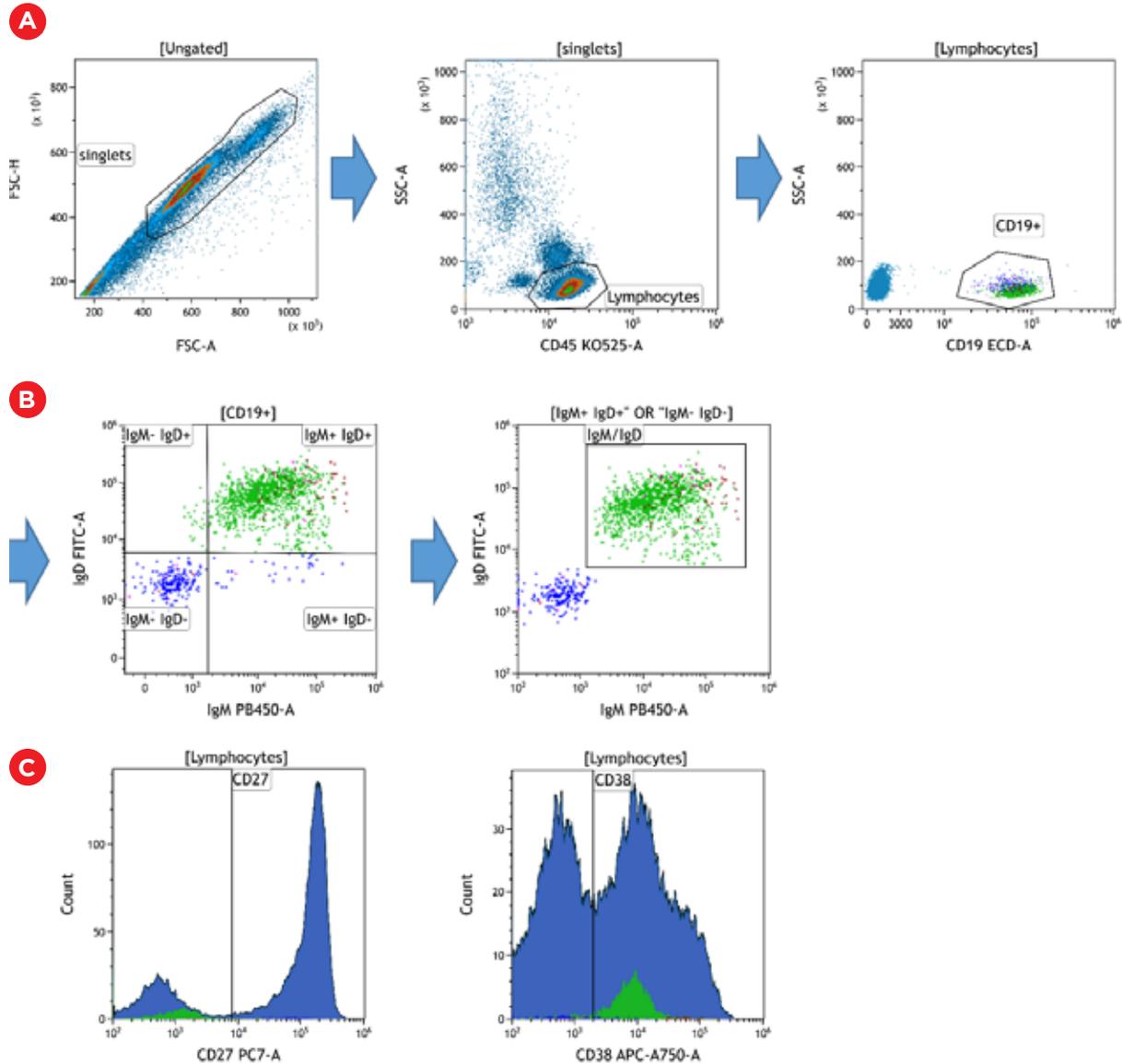
- Branches, which are used to categorize cell populations based on whether they have a negative or positive result for a specified phenotypic data type. Branches are located at the top of the plot.
- Bars, which allow for the comparison of possible negative/positive branch combination between two cell populations. Bars are the central focus of the Tree Plot, as they are the pictorial representation of this phenotypic classification system. Bars can be viewed as either Count or % Gated.



**Figure 3:** Tree Plot display of IgM/IgD- CD19+ cells (blue bars) and IgM/IgD+ CD19+ cells (green bars) comparing co-expression patterns of CD27 and CD38 as branches of the plot. Data was generated using DuraClone IM B cells kit.

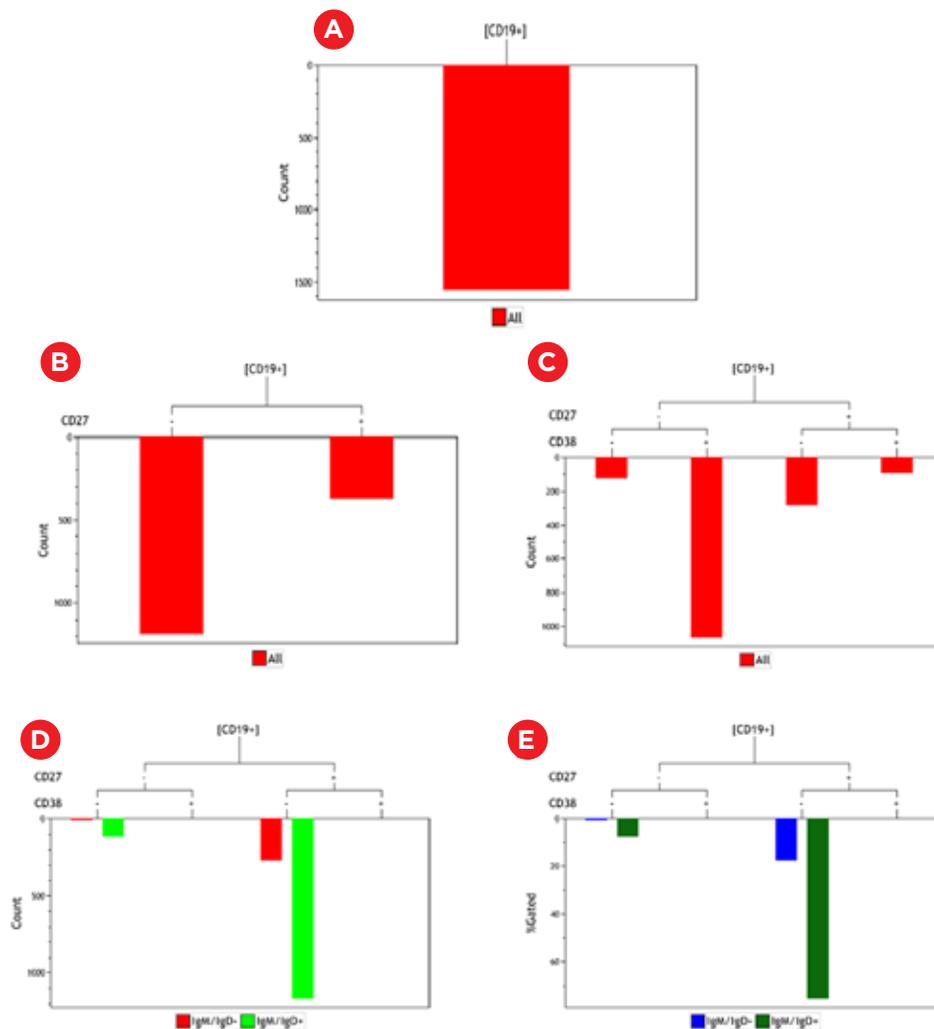
Tree Plot creation, step by step: (Fig 4: T-cell tube)

1. Create plots displaying every phenotype you wish to include in your Tree Plot as Branch or Bar (Figure 4).
2. On each plot, create a gate that includes the events that are positive for the respective phenotype. Kaluza will automatically label bars and branches that include the phenotype with + and those that exclude the phenotype with -. Therefore it is best to label the gate with the population name only, omitting + or - for CD markers (Figure 4).



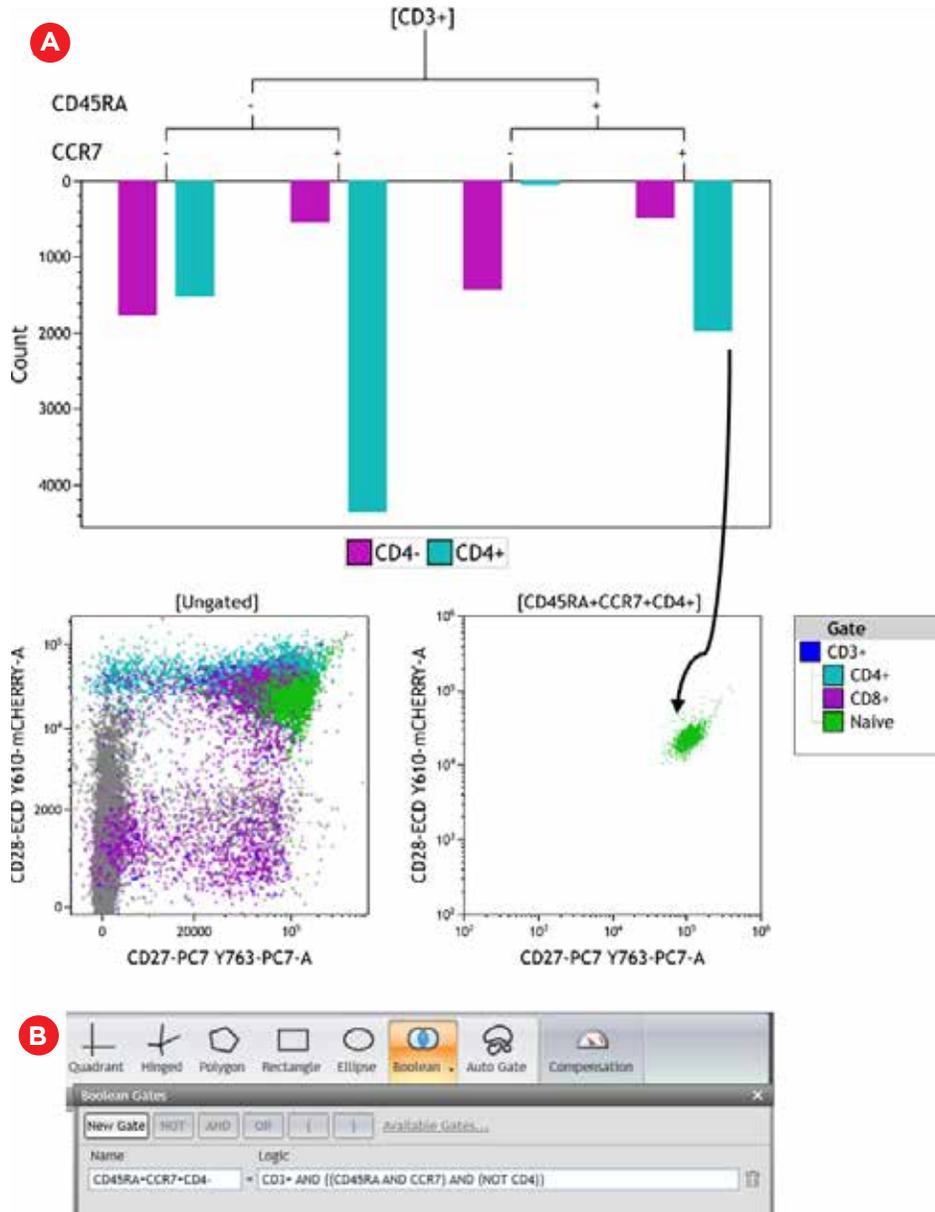
**Figure 4:** Identification of populations and phenotypes with the DuraClone IM B cells panel. A) Sequential gating to identify CD19+ lymphocytes. The CD19+ population will be used as input gate. B) CD19+ lymphocytes are further divided based on their IgM and IgD expression levels. A Boolean gate was created to allow exclusion of IgM- IgD+ and IgM+ IgD- cells from the Tree Plot. The IgM/IgD gate will be used to define the bars of the Tree Plot. C) Histogram plots were created to visualize CD38 and CD27 expression. To facilitate identification of the positivity cut-off Lymphocytes were chosen as input gate. The “Divider” gate type was used to mark the cut-off between positive and negative events. The positive side of the Divider was labelled with the name of the marker displayed.

3. Select the Tree Plot icon from the Plots & Tables ribbon tab to add a new Tree Plot.
4. Choose the input gate (if needed) to filter your data by selecting the [Ungated] hyperlink located at the top of the plot and choosing the gate from the pop-up menu.
5. Use the <Choose Branches> hyperlink to choose the Branches of the tree; Branches can be any gate within the Data Set. Each Branch added to the tree further classifies gated events based on whether they are positive or negative for the phenotypic characteristic defined in the Branch(es) of greater precedence.  
A legend for the Bars, including the colors and the definition of the positive/negative phenotypic data classification specifically associated with each bar, is, by default, located at the bottom of the plot.
6. Select the appropriate Y-axis data type for viewing the bars. The default measurement type is Count. If you wish to change the measurement type to % Gated, select the <Count> hyperlink, and from the pop-up list, choose % Gated.



**Figure 5:** Tree Plot Setup for the DuraClone IM B cells panel. A) A new Tree Plot was created and CD19+ chosen as input gate. B) CD27 was chosen as the first branch. Divider gates are found in the Quadrant category of the gate list. C) CD38 was chosen as the second branch. The Tree Plot now contains bars representing all 4 for possible combinations of CD27 and CD38 expression. D) IgM/IgD was added as bars. E) The y-axis scale was switched to %Gated and different colors assigned to the bars by accessing the Tree Plot Radial menu and selecting the Coloring option.

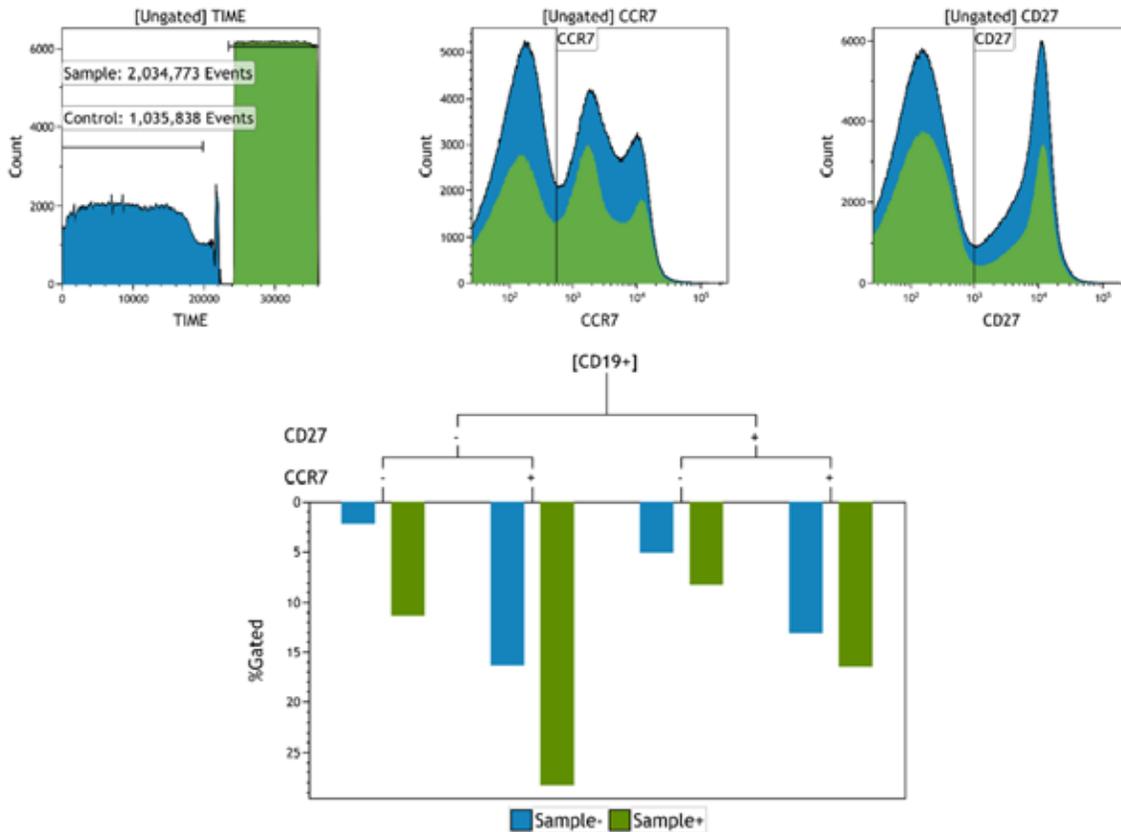
- For further analysis of any population visualized as a bar in a Tree Plot, these events can automatically be extracted by the generation of a Boolean gate. You may use Bars from Tree Plots as an input gate for other plots, including other Tree Plots. To gate a plot using a Bar, press the (Alt) key and select and drag the Bar onto the appropriate plot and release your mouse button to complete the process (Figure 6).



**Figure 6:** Boolean gate creation from Tree Plot. Human peripheral blood was stained using the DuraClone IM T cells Tube according to IFU and acquired on a CytoFLEX LX. Data was analyzed in Kaluza Analysis. A) The CD45RA+ CCR7+ CD4+ bar (aqua), was selected while holding down the ALT key and dragged onto a CD27 vs. CD28 dot plot. B) A Boolean gate representing the chosen bar was automatically created.

Tip: As you hover your mouse over a Bar, the names of the gates, including the positive or negative classification, displays in the Branches that are associated to that Bar.

Tree Plots may also be used to compare two samples by merging two data files and separating them by gating as shown in Figure 7.



**Figure 7:** Tree Plot comparing a control sample (blue) and a test sample (green) as bars. CD19+ events were used as input population and CD27 and CCR7 as branches. Data was generated by Liechti et al. Liechti, T., Günthard, H. F. and Trkola, A. (2018), OMIP-047: High-Dimensional phenotypic characterization of B cells. Cytometry, 93: 592-596. doi:10.1002/cyto.a.23488.

## Summary

While assumption driven manual data analysis may not be ideal for data exploration and discovery of novel populations, for defined research questions and analysis of known cellular subsets it is still relevant today. With the patented Tree Plot Kaluza offers a tool for fast and comprehensive deconvolution of multi-color data, rapid generation of statistical results and comparison between samples. Tree Plots are especially valuable for the analysis of markers that show a bimodal expression pattern.

## References:

1. Aghaeepour, N., Chattopadhyay, P., Chikina, M., Dhaene, T., Van Gassen, S., Kursu, M., ... Brinkman, R. R. (2016). A benchmark for evaluation of algorithms for identification of cellular correlates of clinical outcomes: A Benchmark for Identification of Cellular Correlates of Clinical Outcomes. *Cytometry Part A*, 89(1), 16-21. <https://doi.org/10.1002/cyto.a.22732>.
2. Amir, E. D., Davis, K. L., Tadmor, M. D., Simonds, E. F., Levine, J. H., Bendall, S. C., ... Pe'er, D. (2013). ViSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nature Biotechnology*, 31(6), 545-552. <https://doi.org/10.1038/nbt.2594>.
3. Irish, J. M. (2014). Beyond the age of cellular discovery. *Nature Immunology*, 15(12), 1095-1097. <https://doi.org/10.1038/ni.3034>.
4. Lugli, E., Roederer, M., & Cossarizza, A. (2010). Data analysis in flow cytometry: The future just started. *Cytometry Part A*, 77A(7), 705-713. <https://doi.org/10.1002/cyto.a.20901>.
5. Maecker, H. T., Rinfret, A., D'Souza, P., Darden, J., Roig, E., Landry, C., ... Sekaly, R. P. (2005). [No title found]. *BMC Immunology*, 6(1), 13. <https://doi.org/10.1186/1471-2172-6-13>.
6. Reiter, M., Rota, P., Kleber, F., Diem, M., Groeneveld-Krentz, S., & Dworzak, M. (2016). Clustering of cell populations in flow cytometry data using a combination of Gaussian mixtures. *Pattern Recognition*, 60, 1029-1040. <https://doi.org/10.1016/j.patcog.2016.04.004>.
7. Saeys, Y., Gassen, S. V., & Lambrecht, B. N. (2016). Computational flow cytometry: Helping to make sense of high-dimensional immunology data. *Nature Reviews Immunology*, 16(7), 449-462. <https://doi.org/10.1038/nri.2016.56>.