

# FlowCAP-2: Participant Guide

May 22, 2011

## Introduction

This document provides detailed instructions for participation in the FlowCAP-2 Challenge. For more information please visit: <http://flowcap.flowsite.org/> or join the Google Group and mailing list at <http://groups.google.com/group/flowcap>.

In FlowCAP1, we evaluated the algorithms against cell population identification methods. However, this strategy heavily relies on the quality of the gold standard produced by humans. In FlowCAP2, we have avoided using manual gates when possible. In challenges that manual gates are still used as gold standards, the gates will be extensively reviewed by experts before and after the evaluation of the algorithms. There will be 2 groups of challenges:

1-Classification: In challenges 1, 2, and 3a, the goal is to find cell populations that correlate with a known clinical outcome.

2-Cell Population Identification: In challenges 3b, 4, 5, 6, and 7, the goal is replicate the gold standards produced by human experts.

## Challenge 1: HIV-Exposed-Uninfected versus Un-exposed

The goal of this challenge is to find cell populations that can be used to discriminate between HEU (n = 20) and UE (n = 24) infants. Blood samples were taken at 6 months after birth and were left unstimulated (for control) or stimulated with 6 Toll-like receptor molecules. In addition to raw FCS files, half of the subject labels are provided for training purposes. Your algorithm must use this data to label the rest of the samples. These labels will be used to evaluate your algorithm's performance.

- Raw FCS files, compensation matrixes, and transformed/compensated CSV files available below:
  - FCS: <http://www.terryfoxlab.ca/flowsite/FlowCAP2/download/HEUvsUE/FCS.zip>
  - CSV: <http://www.terryfoxlab.ca/flowsite/FlowCAP2/download/HEUvsUE/CSV.zip>
  - Compensation: <http://www.terryfoxlab.ca/flowsite/FlowCAP2/download/HEUvsUE/Compensation.zip>
- The meta-data consists of 7 fcs file names for 44 subjects and the subject labels for half of them
  - Meta-data: <http://flowcap.flowsite.org/download/FlowCAP2/HEUvsUE/HEUvsUE.csv.zip>
- Your task is to complete the missing labels in the meta-data. Rename the file to Algorithm-Dataset.csv (e.g., Kmeans-HEUvsUE.csv).

## Challenge 2: Acute Myeloid Leukaemia

The goal of this challenge is to find cell populations that can be used to discriminate between AML positive (n = 43) and healthy donor (n = 316) patients. Peripheral blood or bone marrow aspirate samples were collected over a 1 year period using 8 tubes (tube #1 is an isotype control and #8 is unstained) with different marker combinations. In addition to raw FCS files, half of the subject labels are provided for training purposes. Your algorithm must

use this data to label the rest of the samples. These labels will be used to evaluate your algorithm's performance.

- Raw FCS files and transformed/compensated CSV files are available below:
  - FCS: <http://www.terryfoxlab.ca/flowsite/FlowCAP2/download/AML/FCS.zip>–CS
  - V: <http://www.terryfoxlab.ca/flowsite/FlowCAP2/download/AML/CSV.zip>
- The meta-data consists of 8 FCS file names for each of the 789 subjects and the subject labels for half of them:
  - Meta-data: <http://www.terryfoxlab.ca/flowsite/FlowCAP2/download/AML/AMLTraining.csv.zip>
- Your task is to complete the missing labels in the meta-data. Rename the file to Algorithm-Dataset.csv (e.g., Kmeans-AML.csv).

Note that this challenge is also part of DREAM6 (<http://www.the-dream-project.org/>).

### Challenge 3A: Identification of Antigen Stimulation Group of Intracellular Cytokine Staining of Post-HIV Vaccine Antigen Stimulated T-cells.

The goal of this challenge is to correctly label the antigen stimulation group of post-HIV vaccine T-cells. The data set contains samples from 48 individuals (column "pub-id" in the metadata). Each individual received an experimental HIV vaccine. Samples were collected approximately 10 months later and T-cells challenged with two antigens (ENV-1-PTEG and GAG-1-PTEG, column "antigen" in the meta- data). The response of CD4+ and CD8+ T-cells was measured by flow cytometry for each of these groups. The cells were found to respond differently to the two antigen stimulations. This is essentially a classification challenge.

For training purposes we provide data from 24 individuals within each group. The antigen stimulation label is provided (column "antigen" in the metadata). The testing data (n=24) does not have an antigen stimulation group label. Participants must correctly identify the antigen stimulation group of the test data.

The complete data set consists of 240 FCS files. The data has been compensated, transformed and partially gated (gated for singlets, live cells and lymphocytes). Note that the data set contains positive and negative controls (sebctrl, negctrl) which are provided for challenge 3B. These samples are not part of this challenge, and do not need to have an antigen group label assigned to them to complete the challenge. Only the metadata rows where the antigen code is missing need to be labeled correctly.

### Challenge 3B: Identification of Responders and Non-Responders in Intracellular Cytokine Staining of Post-HIV Vaccine Antigen Stimulated T-cells

The data set is identical to challenge 3A. However, the goal is to use automated methods to identify responders and non-responders to the antigen stimulations, as defined by manual analysis. Important to this challenge are additional positive controls for each sample (sebctrl), in which positive staining cells can be detected for each cytokine following stimulation with an antigen that is known to produce a cytokine response. Also provided is a matched negative control (negctrl), where the sample was not challenged with antigen. This negative control is used as a patient-matched baseline.

The goal of this challenge is to identify each sample as either a responder or non-responder to the antigen stimulation. To facilitate this, we define precisely the criteria for labeling responders and non-responders in the original study. Response to an antigen is measured separately in the CD4+ and CD8+ T-cell subsets, and quantified via intracellular cytokine staining of T-cells (TNF-  $\alpha$ , IL4, IFN- $\gamma$ , and IL2, identified in the FCS files). In the original study, a T-cell subset is determined to respond to an antigen if the IFN-gamma or IL2 cytokines demonstrate a significant change in the number of positively staining cells relative to the patient matched negative control sample (negctrl). These responders and non-responders are coded 0/1 in the "response.cd4+" and "response.cd8+" columns of the metadata, respectively. The significance of the change is established via a one-sided Fisher's exact test for each cytokine.

The complete data set consists of 240 FCS files. The data has been compensated, transformed and partially gated (gated for singlets, live cells and lymphocytes). We present an example of the manual gating scheme in Figure 1.

For training purposes we provide data from 24 individuals. The training set metadata contains information on responder/non-responder status for CD8+ (column "response.cd8+") and CD4+ (column "response.cd4+") T-cell subsets relative to the IFN- $\gamma$  or IL2 cytokine positivity for the training data.

The remaining samples are provided with no response code (testing data, n=24). It will be up to the

participants to identify the responder / non-responder status of the test samples. Note that the negative controls and positive controls do not need a responder classification to complete this challenge. There are two negative controls per sample. You may choose to use any (or both) of them.

- Transformed/compensated FCS and CSV files are available below:

- FCS: <http://www.terryfoxlab.ca/flowsite/FlowCAP2/download/HVTN/FCS.zip>

- CSV: <http://www.terryfoxlab.ca/flowsite/FlowCAP2/download/HVTN/CSV.zip>

- The meta-data consists of labels for half of the samples:

- Meta-data: <http://www.terryfoxlab.ca/flowsite/FlowCAP2/download/HVTN/HVTNTraining.csv.zip>

- Your task is to complete the missing labels in the meta-data. Rename the file to Algorithm-Dataset.csv (e.g., Kmeans-HVTN.csv).

## Challenge 4: Multiple Sclerosis Treatment

The goals of this challenge are to i) recapitulate the results of expert manual gating using automated algorithms and ii) to find cell populations that correlate with treatment arms between healthy controls, multiple sclerosis patients treated with Copaxone and multiple sclerosis patients treated with interferon beta. Peripheral blood mononuclear cells were isolated from blood specimens taken before (t0) and 4 hr, 12 hr, 24 hr and 72 hr after the initiation of treatment. Samples were stained with an 8-marker T cell reagent panel. Data was collected on a viable lymphocyte gate based on FSC/SSC, with 100,000 gated events collected/sample.

- Raw FCS files are available below:

- FCS: <http://www.terryfoxlab.ca/flowsite/FlowCAP2/download/MS/FCS.zip>

the meta-data consists of a file describing the experiment samples used for each results:

- Meta-data: <http://www.terryfoxlab.ca/flowsite/FlowCAP2/download/MS/MS.xlsx.zip>

- Your task is to i) provide files with population memberships labels for every cell event in every sample file, and ii) to identify any cell population that correlates with treatment arm and provide population proportion statistics across all datasets.

## Challenge 5: Omalizumab and Rush Immunotherapy

The goal of this challenge is to find cell populations that correlate with treatment arms between placebo, omalizumab alone, rush immunotherapy alone, and the combination of omalizumab and rush immunotherapy (see Casale TB, et al., J Allergy Clin Immunol 2006;117:134-40 and Casale TB, et al., J Allergy Clin Immunol 2007;120:688-95 for experiment design details and outcome results). Peripheral blood mononuclear cells were isolated from blood specimens taken before and after treatment at various timepoints. Samples were stained with 7 different 4-marker reagent panels.

Although you are welcomed to analyze the entire dataset, the first part of the challenge - to compare the automated results with expert manual gating - will only be performed with the data from Visit 5 (all four arms). The CD56-CD23-DUMP-CD11c and FI-1-FI-2-FL-3-FL-4 folders and tubes can be ignored.

- The dataset is available below:

- <https://www.immport.org/immportWeb/display.do?content=FlowCap>

- Your task is to identify any cell population that correlates with treatment arm and provide population proportion statistics across all datasets.

## Challenge 6: Stimulation of Influenza-specific cytokine-expressing T cells

The goal of this challenge is to identify rare populations (e.g. 25 cells per million) that are induced by influenza antigen stimulation of PBMC. Human PBMC were incubated in triplicate with pools of influenza peptides (G1 to G6), DMSO control medium, tetanus peptides and Staph Enterotoxin B (SEB). After 10 hours, cells were stained

with a panel of antibodies specific for surface antigens and cytokines, then fixed and analyzed.

Your task is to identify the rare cytokine-secreting activated T cell populations that are induced by influenza and tetanus antigen stimulation. The specific example is to identify at least one population of activated CD4 memory T cells expressing IFN $\gamma$  and TNF $\alpha$  that is more abundant in some antigen-stimulated samples than in the negative control (DMSO) samples. These cells are CD3 $^+$  CD14 $^-$  CD4 $^+$  CD8 $^-$  CD45RA $^-$  CD69 $^+$  Live/dead $^-$  CXCR5 $^-$  IFN $\gamma^+$  TNF $^+$ . As a secondary goal, identify additional antigen-induced populations producing other cytokine combinations. Both high yield and high signal:noise are important. Results should be submitted as output files with the identified populations tagged with an additional parameter.

In addition to raw .fcs files, the meta-data consists of the file names and antigen stimulation conditions for the samples.

– FCS: <http://www.terryfoxlab.ca/flowsite/FlowCAP2/download/RarePopulations/FCS.zip>

– Meta-data: <http://www.terryfoxlab.ca/flowsite/FlowCAP2/download/RarePopulations/RarePopulations.xlsx.zip>

## Challenge 7: Dilution of activated cells in control populations.

The goal of this challenge is to determine the sensitivity with which your algorithm can identify SEB activated T cell populations. Human PBMC were incubated with control medium, or SEB. After 10 hours, cells were stained with a panel of antibodies specific for surface antigens and cytokines, fixed and analyzed. Stimulated and unstimulated cell populations were then combined electronically in different ratios.

Your task is to identify the activated T cell populations that are induced by SEB stimulation, analyzing each sample independently. The two specific examples are:

1) Identify at least one population of activated CD4 memory T cells expressing IFN $\gamma$ , IL-2 and TNF $\alpha$  that is present in the SEB-stimulated (S) sample but not in the negative control (D) sample. These cells are CD3 $^+$  CD14 $^-$  CD4 $^+$  CD8 $^-$  CD45RA $^-$  CD69 $^+$  Live/dead $^-$  CXCR5 $^-$  IFN $\gamma^+$  TNF $^+$  IL-2 $^+$ .

2) Identify at least one population of activated CD4 memory T cells expressing IL-17 and TNF $\alpha$  that is present only in the SEB-stimulated (S) sample. These cells are CD3 $^+$  CD14 $^-$  CD4 $^+$  CD8 $^-$  CD45RA $^-$  CD69 $^+$  Live/dead $^-$  CXCR5 $^-$  IL-17 $^+$  TNF $^+$ .

As a secondary goal, identify additional antigen-induced populations producing other cytokine combinations. The sensitivity of detection of each population by your algorithm will be evaluated from the dilution series. Results should be submitted as output files with the identified populations tagged with an additional parameter. In addition to raw .fcs files, the meta-data consists of the file names and ratios of control and SEB-stimulated cells in the samples.

– FCS: <http://www.terryfoxlab.ca/flowsite/FlowCAP2/download/Dilutions/FCS.zip>

– Meta-data: <http://www.terryfoxlab.ca/flowsite/FlowCAP2/download/Dilutions/Dilutions.xlsx.zip>

## How to Participate

For each challenge, submit your algorithm name, the requested results, and the software/documentations required for reproducing your work to [flowcap@flowsite.org](mailto:flowcap@flowsite.org). By default, one submission per research group is allowed. Exceptions can be made in cases that there are significant differences between the two different sets of results. In these cases, contact the organizing committee before the deadline.

## Timeline

The deadline for submission of results is 31.AUG.2011. The FlowCAP2 summit will be held 22-23.SEP.2011 in Bethesda.

## Agreement

Publishing the datasets provided by FlowCAP is prohibited until the project publishes the results. The datasets and results of the FlowCAP project will be publicly available for any use after publication. The software and results submitted to FlowCAP cannot be withdrawn but can be de-identified upon submitter's request.

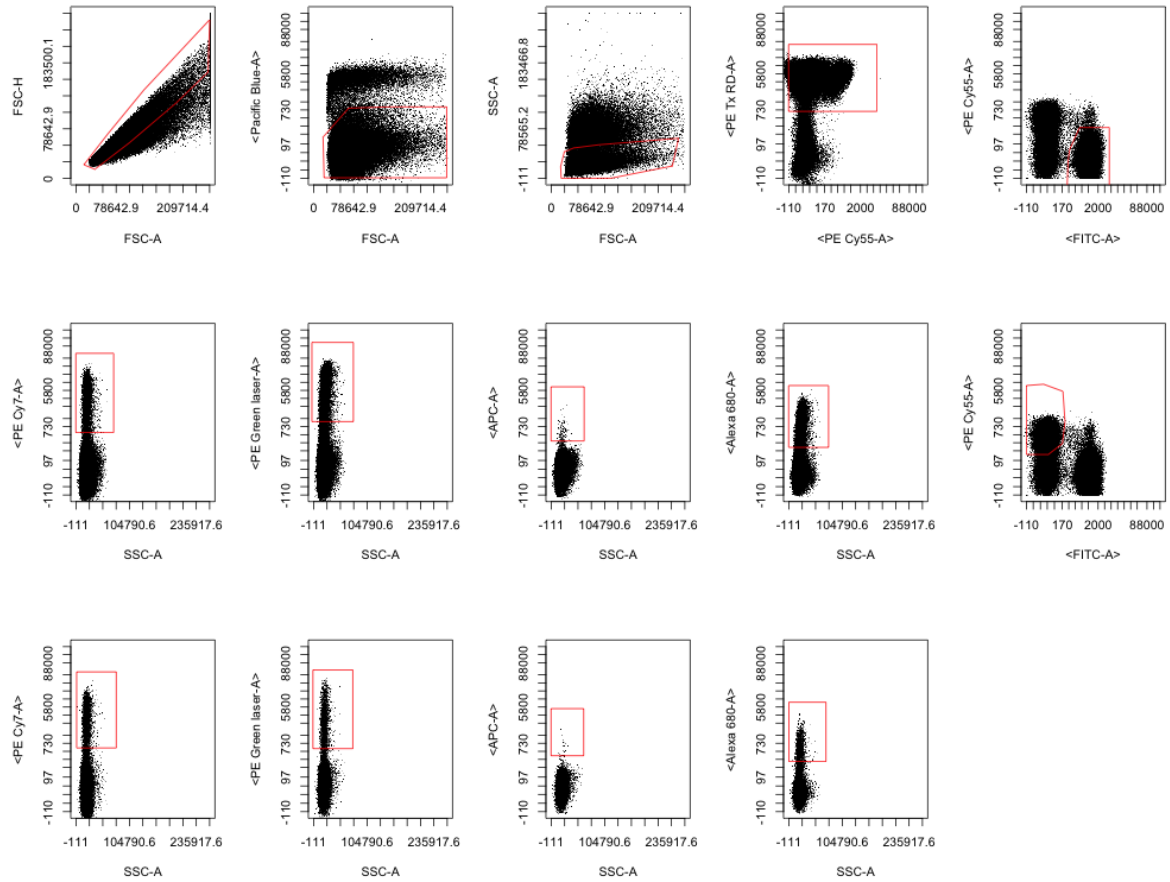


Figure 1: Manual Gating Scheme for the HVTN data. Example of an SEBCTRL sample.