

Bioinformatics Integration Support Contract (BISC), Phase II
**IMPORT FLOW CYTOMETRY ANALYSIS: FLOCK,
RAS, CROSS SAMPLE COMPARISON USER GUIDE**



IMPORT

BIOINFORMATICS FOR THE FUTURE OF IMMUNOLOGY

Version 2.11

Contents

1.0 Introduction	1
2.0 FCSTrans: ImmPort .fcs file conversion	1
3.0 Conversion of .fcs Files to Tab Delimited Text Files	1
3.1 Generating a Marker Information File (.info).....	8
3.2 Editing Marker information of uploaded data.....	8
4.0 Uploading flow cytometry files into ImmPort.....	9
4.1 Accessing Data Management module.....	9
4.2 Uploading Flow Cytometry Files to the Data Management module	9
4.2.1 Single File Upload.....	9
4.2.2 Multiple File Upload	10
4.3 Selecting, Editing and Analyzing Uploaded files.....	12
4.4 Create Data Sets in the Data Management Module	13
5.0 Data Analysis with FLOCK	14
5.1 FLOCK v1 and v2.....	14
5.2 Accessing data files for FLOCK analysis	14
5.3 Providing descriptive analysis information.....	15
5.4 Marker Selection.....	16
5.5 FLOCK Parameter Selection	16
5.5.1 Using Auto-detect parameter setting	16
5.5.2 Using the Customize Parameter setting	16
5.5.3 Viewing FLOCK analysis results.....	17
5.6 FLOCK Analysis Detail.....	18
5.7 FLOCK Results visualization	18
5.7.1 FLOCK Population ID.....	20
5.7.2 Summary Table of all Analysis Populations	20
5.7.3 Centroid Table.....	21
5.7.4 Mean Fluorescence Intensity Table (MFI).....	21
6.0 Result Adjustment System (RAS).....	22
7.0 Cross Sample Comparison	22
7.1.1 Population Percentage table.....	24
7.1.2 Marker by Population visualization	25
7.1.3 Downloading Cross Sample results.....	26
8.0 FLOCK at sourceforge.....	26
9.0 Future Work.....	26

Appendices

APPENDIX A Glossary..... 28

BISC Documentation Style Guide Version History

Version	Date	Description
2.4	11/24/2008	2.4 release of the ImmPort Flow Cytometry using FLOCK User Guide
2.5	04/30/2009	2.5 release of the ImmPort Flow Cytometry using FLOCK and RAS User Guide
2.9	08/11/2010	2.9 release of the ImmPort Flow Cytometry using FLOCK, RAS and Cross Sample

1.0 INTRODUCTION

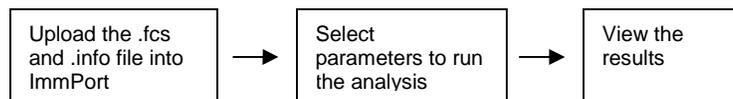
Recent advances in flow cytometry (FCM) instrumentation combined with the availability of new staining reagents have resulted in the ability to assess cell samples based on the quantification of many individual molecular characteristics simultaneously. These advances in experiment methodologies have led to a challenge in the interpretation of the resulting data. Historically, investigators have analyzed FCM data manually by utilizing gating strategies and visualization of two-dimensional dot plots to identify specific cell populations within samples. However, this approach is somewhat subjective and laborious, and does not scale well as the number of parameters increases. These observations have led to the realization that computational approaches to assist in the interpretation of high-dimensional FCM data are sorely needed. To address this need, UTSW has developed a system for automated population discovery in multidimensional FCM data – FLOCK (FLOW Clustering without K), which is designed to specifically take into account the unique feature of FCM data and produce an objective segregation of cell populations.

Flow Cytometry Analysis Workflow



Flow Cytometry data analysis within ImmPort employs the FLOCK algorithm and supports management of data files, creation of data sets, population centroid adjustment for results refinement and cross sample comparison.

2.0 FCSTRANS: IMMPORT .FCS FILE CONVERSION

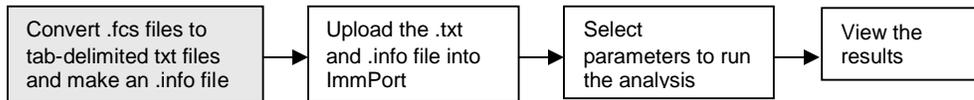


FCS (.fcs) files generated by instruments used in flow cytometry (FCM) experiments are in binary format and cannot be directly processed by independent analysis and visualization software. Reading and transforming a binary .fcs file requires understanding FCS file formats including FCS2.0 and FCS3.0. Unlike FCS2.0 which usually stores log-transformed data, FCS3.0 files keep the original raw instrument outputs in a linear mode. Analysis and visualization software transforms the linear-mode data before delivering the data to biologists.

Most existing software used for FCS file conversion, including flowCore of Bioconductor, FCSExtract, FCS2CSV, and LData, do not always provide satisfactory transformation of FCS3.0 files when compared with FlowJo MacOS version (Tree Star Inc.). FCSTrans, the ImmPort automated flow cytometry data converter, generates a data matrix output from FCS2.0 and FCS3.0 binary files which can be used by immunologists and bioinformaticians to perform lab-specific analysis and customized visualization of their FCM data. All uploaded .fcs files will be converted by FCSTrans to a data format suitable for FLOCK analysis without further action on the part of the user. FCSTrans is written in R and will be made an open source Bioconductor package. Details about FCSTrans can be found in the ImmPort Flow Cytometry Help menu

3.0 CONVERSION OF .FCS FILES TO TAB DELIMITED TEXT FILES

The FLOCK Beta release also accepts flow cytometry files which have been converted from their native .fcs format to a tab delimited text format, however this is not a required step as FCSTrans will automatically convert uploaded .fcs files. There are many flow cytometry analysis tools that can accomplish this task. The workflow for file format conversion using Tree Star FlowJo™ and FCSExtract are highlighted below as examples. Please note that only the Mac OS version of FlowJo™ supports this operation at this time.



The input to this stage is an .fcs file. The output of this module is a tab delimited .txt file. Each .txt file corresponds to one .fcs file. The text file will contain columns for the fluorescence values for each parameter in the flow cytometry assay, (e.g. FSC, SSC, fluorochrome 1, etc). The following screenshots demonstrate the conversion of the .fcs files using two versions of FlowJo™.

Figure 1: Upload a .fcs format flow cytometry file into FlowJo™ using standard procedures

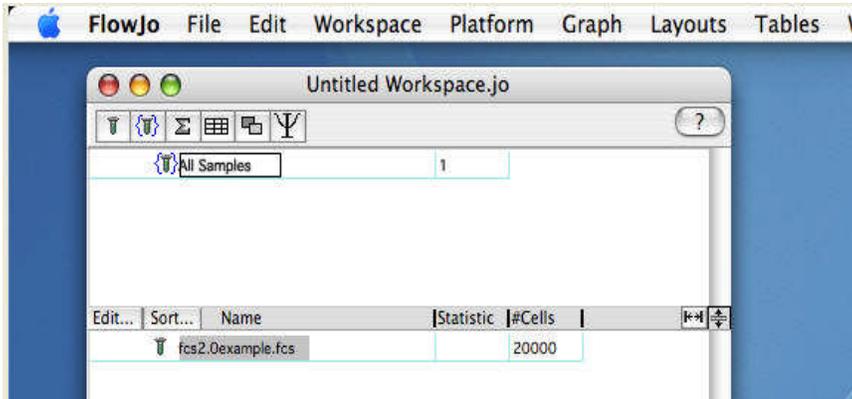


Figure 2: Select the 'Export' option from the 'Workspace' menu option

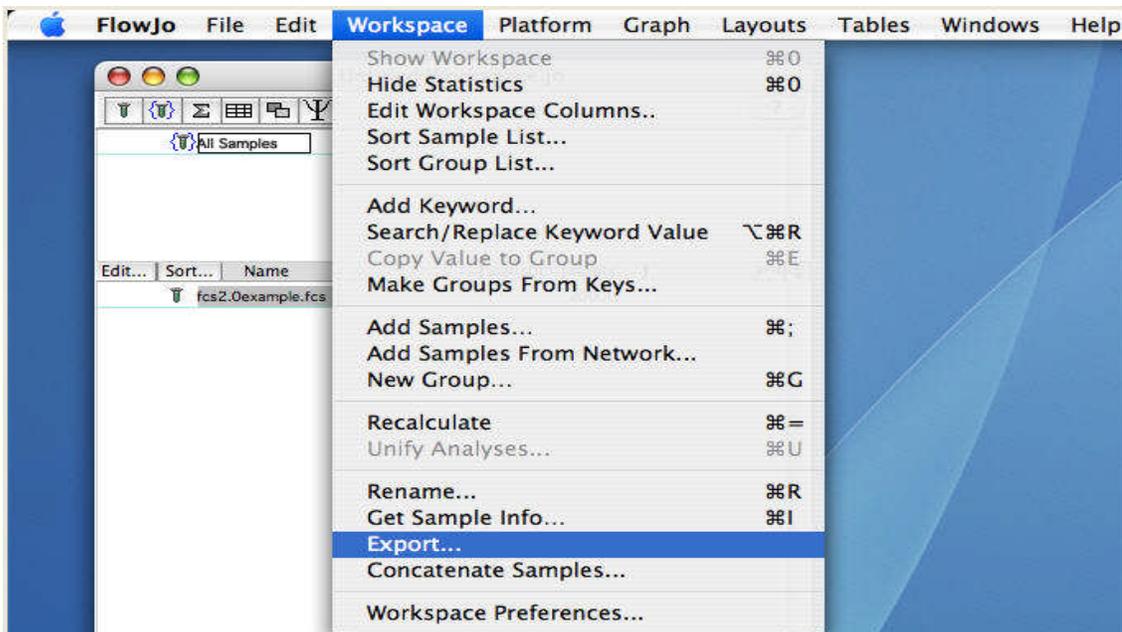


Figure 3: The 'Export' dialog box displays the default settings

All parameters are selected by default for export.

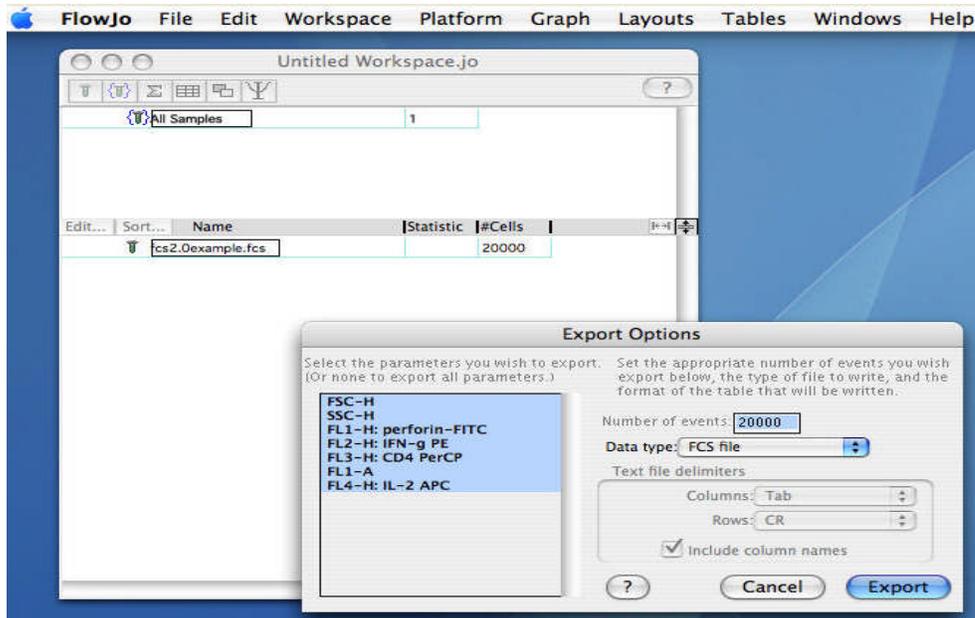


Figure 4: De-select any parameters that are not to be included in the text file

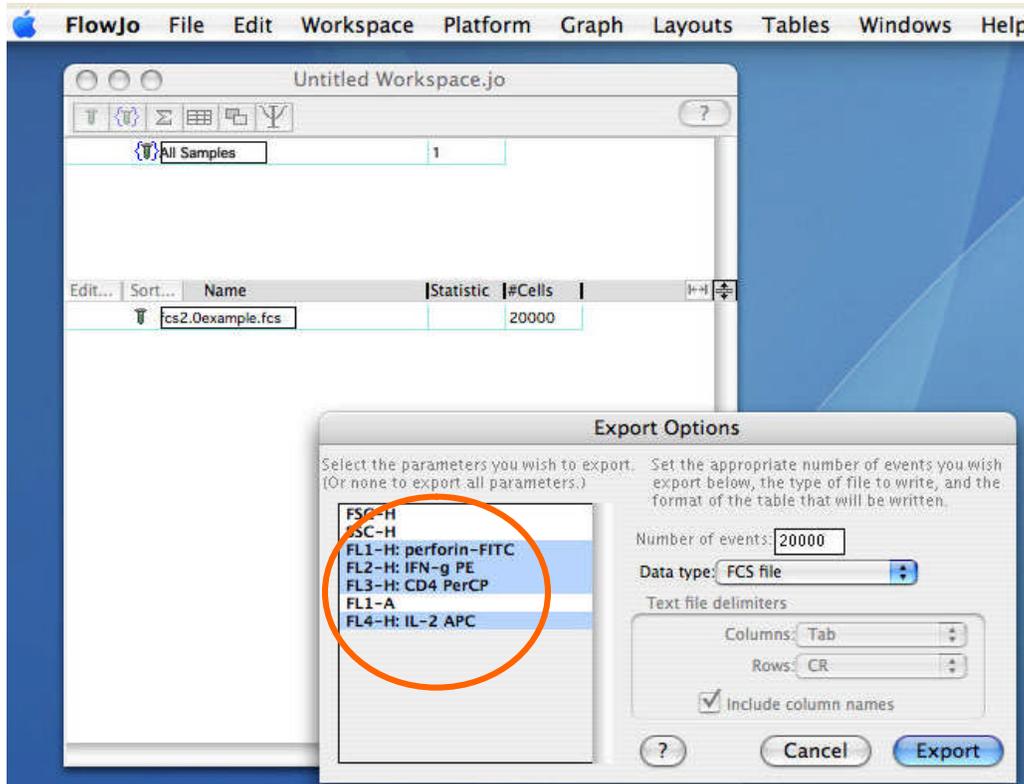


Figure 5: Change the 'Data Type' parameter to 'Channel numbers' from 'FCS file'

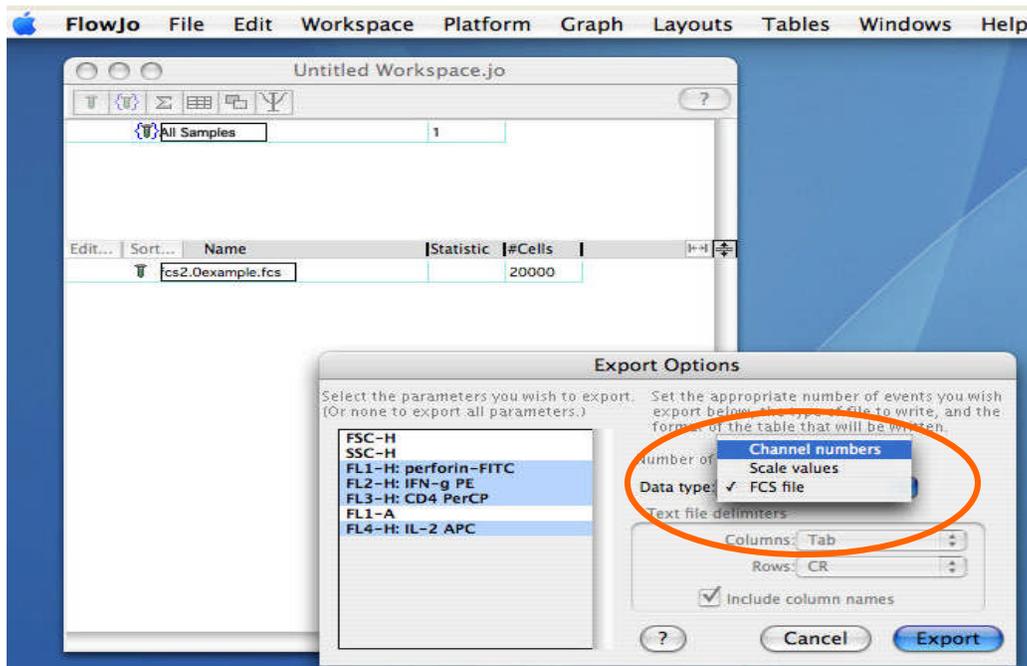
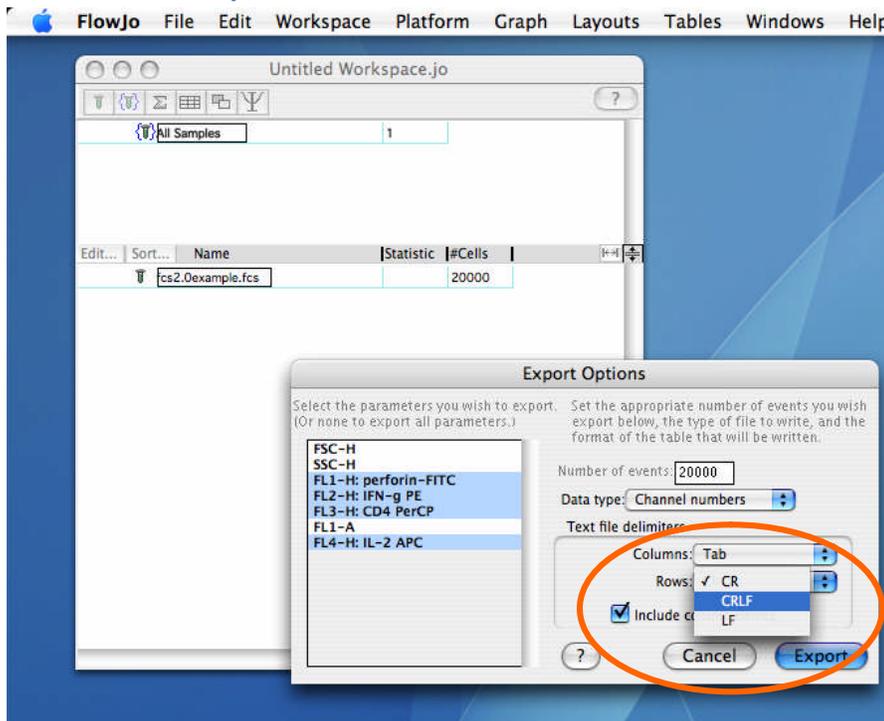


Figure 6: Select 'CRLF' option for Text file delimiter

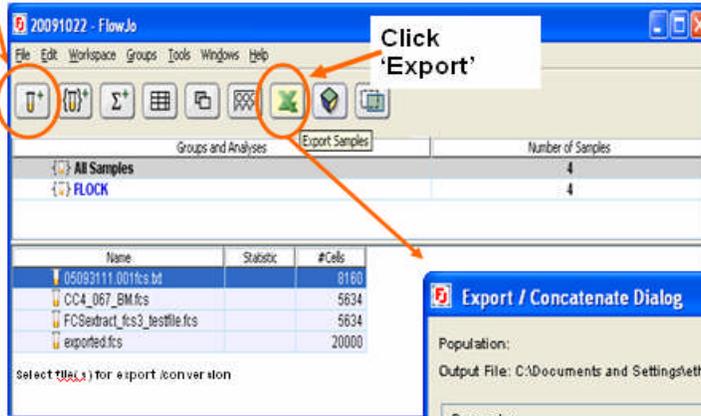


Clicking "Export" will allow you to save the .txt file to the folder you specify.

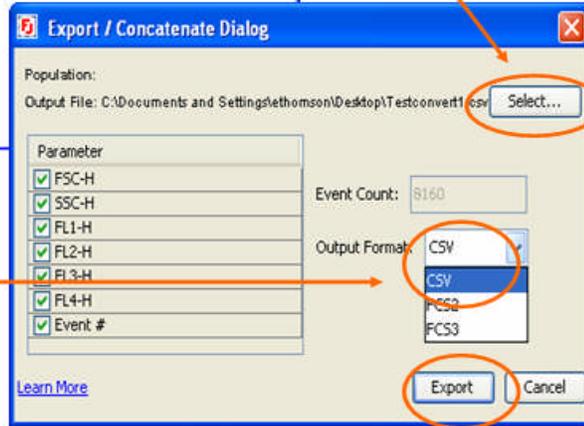
FlowJo™ v8.6.6 for Macintosh:

Figure 7: Select samples to be converted, select CSV as output format

Add Samples



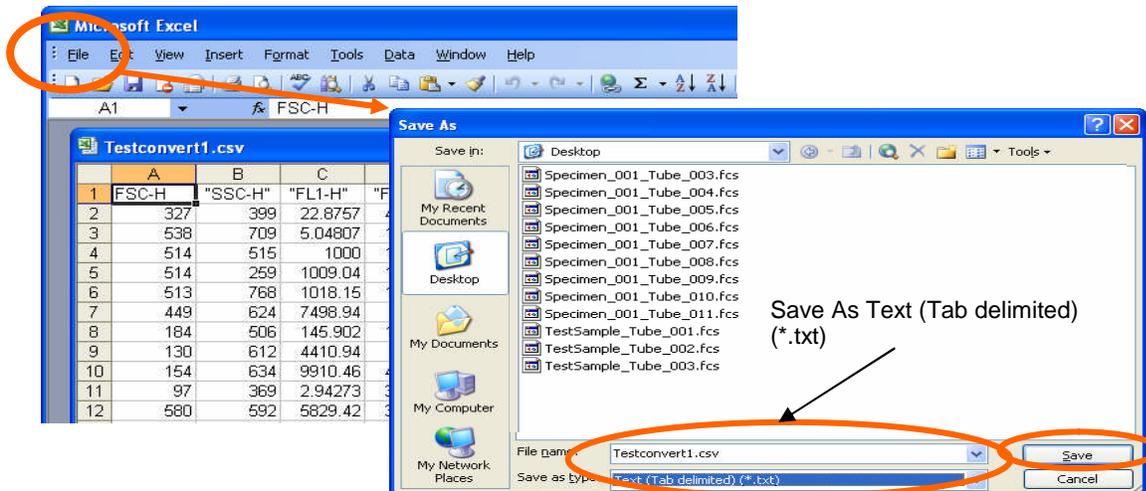
Click Select—choose where the CSV file should be stored



Output format:
 ▪ FLOCK requires .txt files.
 To convert to .txt, initially export as CSV.
 ▪ To convert CSV to .txt, go to the next slide

Figure 8: Save CSV file as a .txt file

- Locate CSV file based upon the location you chose for storing.
- Select to 'Save As' a Text (Tab delimited) (*.txt) file, click Save
- You will upload the new .txt converted file into ImmPort through the Data Management upload process



FCSExtract FCS file conversion to .txt format

Figure 9: Download FCSExtract, Run software

- Go to: <http://research.stowers-institute.org/efg/ScientificSoftware/Utility/FCSExtract/index.htm>
- FCSExtract applies to FCS v2.0 format
- Download FCS Extract (a)
- Open the .zip file (b)
- Double click the FCS Extract icon, and click 'Run' (c)

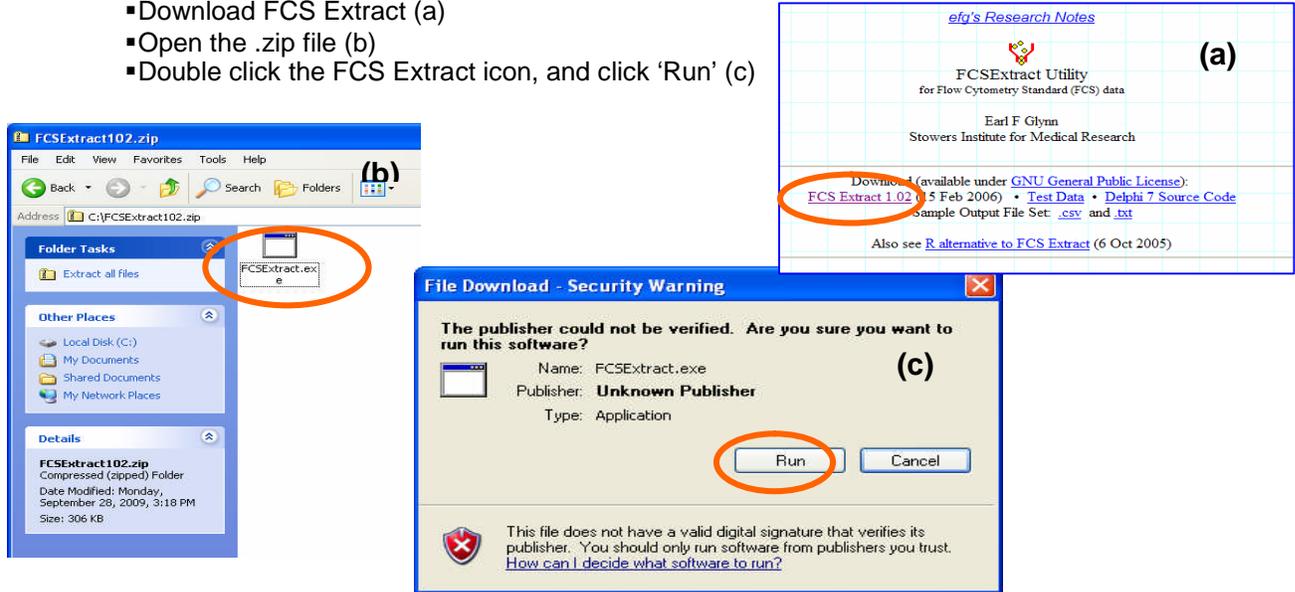


Figure 10: Select FCS file to be converted

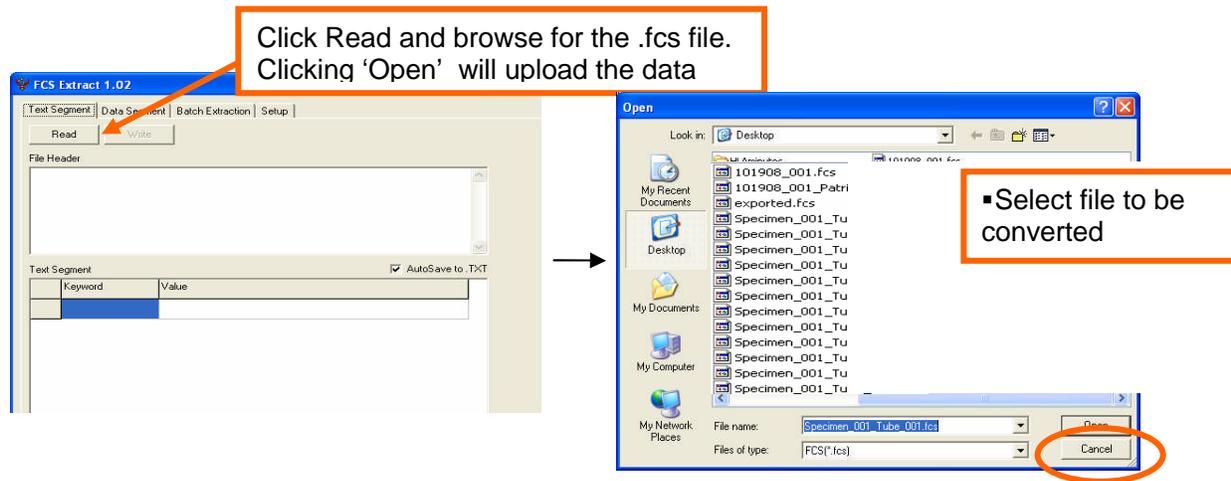


Figure 11: Convert FCS to CSV

FCS Extract 1.02

Text Segment | Data Segment | Batch Extraction | Setup

Read | **Write** | C:\Documents and Settings\ethomson\Desktop\Specimen_001_Tube_009.fcs

File Header

Version: FCS2.0
TextFirst: 256
TextLast: 1629
DataFirst: 1630
DataLast: 201629
AnalysisFirst: 0
AnalysisLast: 0

Text Segment

AutoSave to .TXT

Keyword	Value
\$FIL	Specimen_001_Tube_009.fcs
\$SYS	Windows XP 5.1
\$TOT	10000
\$PAR	10
\$MODE	L
\$BYTEORD	4,3,2,1
\$DATATYPE	I
\$NEXTDATA	0
CREATOR	BD FACSDiva Software Version 5.0.1
TUBE NAME	Tube_009
\$SRC	Specimen_001
EXPERIMENT NAME	48hr 101806_002
GUID	6ca19385-fa04-4c25-9979-d04724de61fc
\$DATE	19-OCT-2006

10,000 rows

Save As

Save in: Desktop

File name: Specimen_001_Tube_009.csv

Save as type: CSV (Comma delimited) (*.csv)

Save | Cancel

From pop-up, select where the .csv file should be stored. Click 'Save'. .csv file is now stored at the

- Click “Write”
- Select where the .csv file should be stored (Desktop, C: drive...)
- Un-check the AutoSave to .TXT box. If left checked, FCS Extract provides .txt output of the .fcs file header information only, and does not include the necessary fluorescence values (.csv format)

Figure 12: Convert CSV file to .txt file

Microsoft Excel

File | Edit | View | Insert | Format | Tools | Data | Window | Help

Specimen_001_Tube_009.csv

	A	B	C	D	E	F	G	H	I	J	K
1	FD	F1	F2	F3	F4	F5	F6	F7	F8	F9	
2		124	81	583	0	188	95	236	589	266	0
3		828	649	402	422	315	289	495	640	526	0
4		24	57	0	174	0	0	0			
5		420	185	675	218	279	104	481			
6		154	125	795	341	216	177	252			
7		234	120	674	0	347	84	367			
8		1023	493	313	368	291	267	536			
9		1023	760	462	497	393	357	646			
10		1023	448	326	480	251	315	533			
11		417	302	915	0	394	307	293			
12		1023	410	452	428	292	278	524			
13		26	30	447	183	84	106	200			
14		14	22	397	0	0	0	0			
15		244	258	811	63	233	334	556	453	465	0
16		1023	953	471	465	420	367	632	657	621	0
17		955	267	368	301	341	239	506	618	491	0

Save As

File name: Specimen_001_Tube_009.txt

Save as type: Text (Tab delimited) (*.txt)

Save | Cancel

- Find the .csv converted .fcs file (based on where it was saved in Slide 3). The file will be in an Excel format.
- Open the file
- Click on File, 'Save As' and select 'Text (Tab delimited(*.txt) from the Save as Type dropdown menu
- File will be save as .txt to the specified location and can be uploaded into ImmPort

If there are questions or comments, please contact the helpdesk at helpdesk@import.org

3.1 GENERATING A MARKER INFORMATION FILE (.INFO)

The Data Management module provides the user with two methods for annotating the uploaded .fcs or .txt files: uploading an .info file or editing the uploaded FCS file to provide more informative names for the markers and fluorochromes used in the flow cytometry assay.

The .info file is a two (2) column tab delimited text file. The first column contains the names of the parameters (markers) from the .fcs file and can be found in the header of the .fcs or .txt file. The second column contains names the user chooses to provide for the parameters (markers). In some cases, the parameter names are fluorochrome channel identifiers (e.g. FL1) and these may be renamed to indicate the fluorochrome and marker used to stain the cells (e.g. APC-anti-CD3).

Example .info file

Note that the columns are tab delimited.

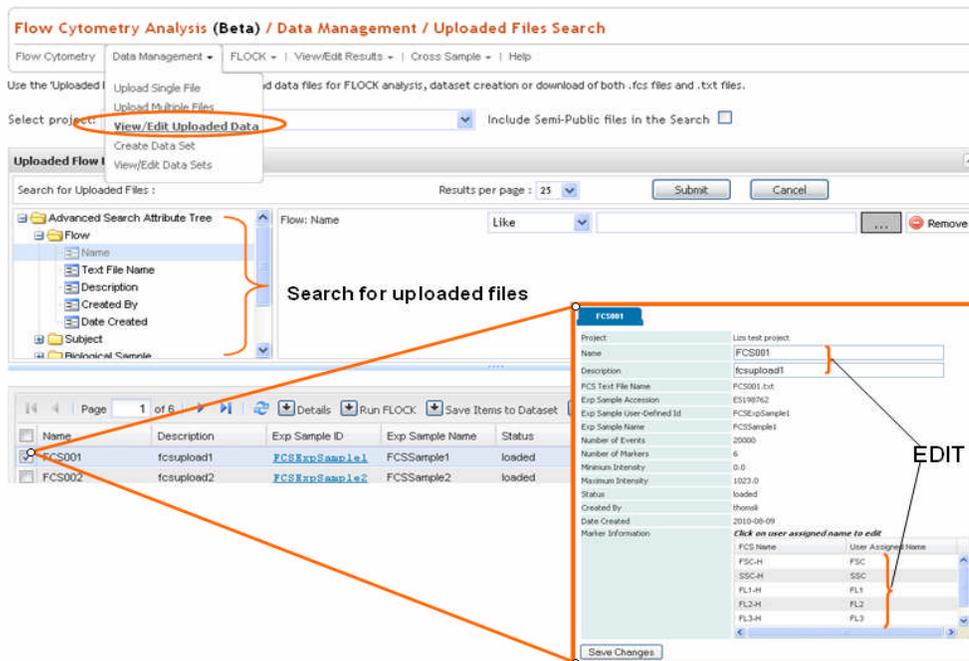
FSC	Forward Scatter
SSC	Side Scatter
FL1	APC-anti-CD3
FL2	Cy7-anti-CD4
FL3	Alex 350-CD8
FL4	PE-CD25

3.2 EDITING MARKER INFORMATION OF UPLOADED DATA

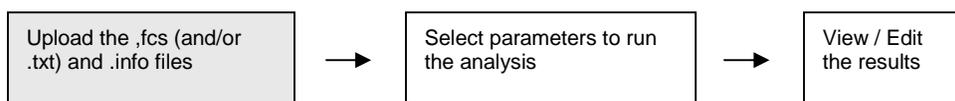
Uploaded data, not associated with a marker .info file at upload, will use marker information from the header of the FCS file as a default. Editing is possible through the file detail screen.

Figure 13: Editing uploaded .txt file

- From 'Data Management' in tool bar
- Click 'View/Edit Uploaded Data'
- Select file to be edited, click 'View Details'
- Edit from Detail screen, click 'Save'



4.0 UPLOADING FLOW CYTOMETRY FILES INTO IMMPORT



4.1 ACCESSING DATA MANAGEMENT MODULE

The Data Management module is accessed after login (<http://www.immport.org>) from the Tools section of the menu bar. Figure 14 captures the Tools dropdown menu with the Data Management ‘Upload Data’ highlighted. Uploaded files are stored and may be reused.

Figure 14: Accessing Data Management module

Flow Cytometry Analysis (Beta) / Data Management / Overview

Flow Cytometry | Data Management | FLOCK | View/Edit Results | Cross Sample | Help

The Data Management module provides the following options:

- Upload Single File**: Upload a single .fcs or .txt file for future analysis.
- Upload Multiple Files**: Upload multiple .fcs or .txt files for future analysis.
- View/Edit Uploaded Data**: View and edit uploaded data, adding sample information, renaming markers.
- Create Data Set**: Create a new data set, which can be used either in batch FLOCK runs or Cross Sample Comparison.
- Put multiple files into a Data Set**: Add multiple files to a data set, which can be used either in batch FLOCK runs or Cross Sample Comparison.
- View and edit the created dataset**: View and edit the created dataset, e.g., adding/removing files, adding information for files in the dataset.

The uploaded data file can be in either .fcs or .txt format. The filename can't contain the following characters: " \ and / . ImmPort automatically converts .fcs files to .txt files when the upload includes only .fcs files. The .txt files can be created using third party tools including Tree Star FlowJo™ on MacOS. For conversion details, please see the [help section](#).

Channels, such as forward scatter width (FSC-W), can distort FLOCK analysis result if included together with forward scatter height (FSC-H) in the .fcs data file. You can exclude these channels using Tree Star FlowJo™ on MacOS or you can modify and upload the .txt data file available after conversion on ImmPort. For details, please see the [help section](#).

The ImmPort FCS conversion component has been tested with a variety of BD (Becton, Dickinson and Company) flow cytometry instrument data files. The conversion results are highly consistent with those from Tree Star FlowJo™ channel output for FCS 2.0 and 3.0 files. Ongoing improvements will cover data files, both pre- and post-gated, from alternative flow cytometers. If you have data files from another vendor, please contact the [help desk](#) so that the ImmPort team can assist you with your data requirements.

Flow Cytometry Analysis Workflow

```

    graph LR
      A[Upload Data and Create Dataset] --> B[Identify Populations Automatically Using FLOCK]
      B --> C[Adjust FLOCK Results and Generate Centroid Files]
      C --> D[Compare Populations Across Different Samples]
  
```

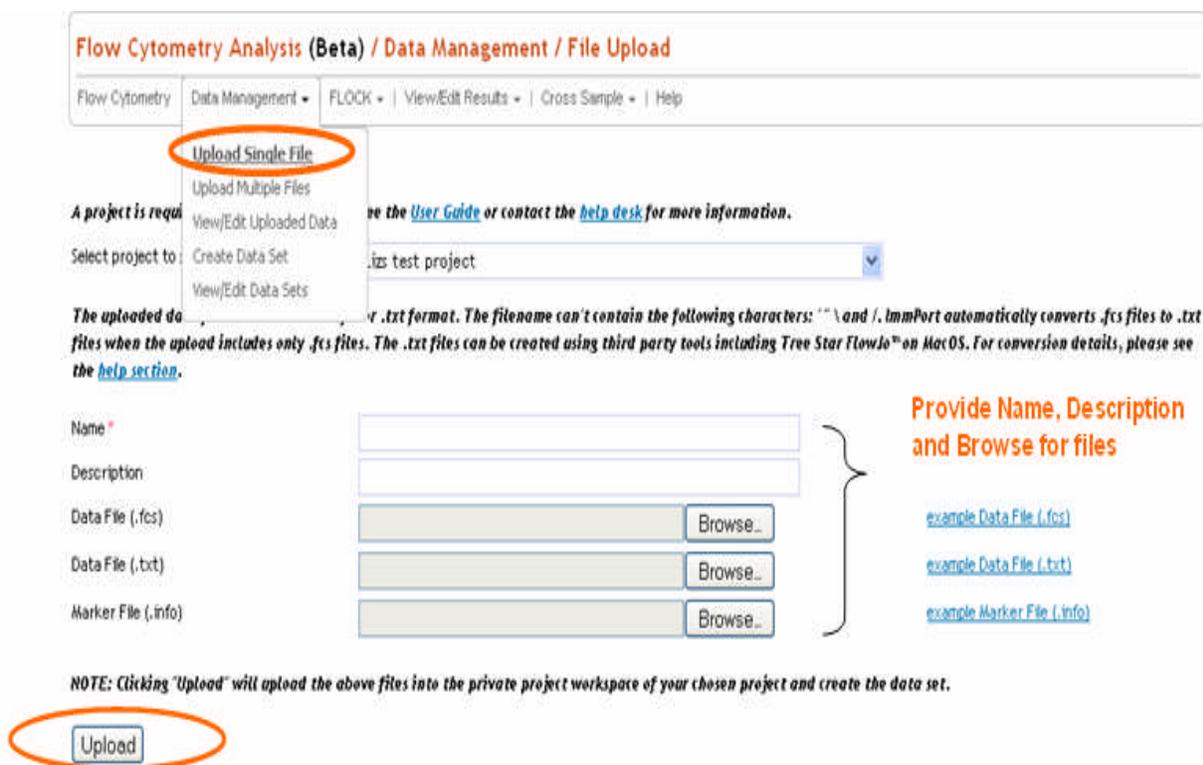
4.2 UPLOADING FLOW CYTOMETRY FILES TO THE DATA MANAGEMENT MODULE

Flow cytometry data files can be uploaded individually, or as multiple files in a .zip package, into ImmPort.

4.2.1 Single File Upload

Single file upload employs a standard ‘Browse’ option to select the flow cytometry (.fcs and/or .txt) and .info file from your computer. You will be prompted to choose a private workspace, from a dropdown menu, in which to store your files. Examples of .fcs, .txt and .info files can be found on the upload screen and utilized to test the upload process.

Figure 15: Flow Cytometry single file data upload in ImmPort



As Figure 15 shows, the addition of file information such as name and description is enabled; descriptive file-specific information will be helpful in discriminating files in the future, particularly in Cross Sample comparisons.

4.2.2 Multiple File Upload

To **upload multiple files** in a single upload process the creation of a .zip data package is required. The .zip data package consists of:

1. .fcs and/or .txt flow cytometry files
2. .info file if parameter (marker) names are to be changed (for more information see section 3.1)
3. flowTextFiles template. The flowTextFiles template can be found as a link on the Multiple File Upload screen (see Figure 16). Download and complete the template as an .xls spreadsheet (see Figure 17), convert to a .txt file prior to including in the .zip data package. Conversion of the flowTextFiles.xls to flowTextFiles.txt is necessary for the system to accurately parse the flow cytometry files and for those files to appear in the selected project. **Note: do not change the name of the flowTextFiles.txt.**

To access Multiple File Upload go to the Data Management drop down menu and select Upload Multiple Files. Figure 16 highlights the location of the flowTextFiles.xls download and the Browse function for .zip package selection and annotation.

Figure 16: Flow Cytometry multiple file data upload in ImmPort

Flow Cytometry Analysis (Beta) / Data Management / Package Upload

Flow Cytometry | Data Management - | FLOCK - | View/Edit Results - | Cross Sample - | Help

Upload Single File
Upload Multiple Files
 View/Edit Uploaded Data

Upload of Multiple Flow Cytometry Result Files

Complete Spreadsheet Data Multiple Files → Create Data Set View/Edit Data Sets → Info File (including the Detected) → Assemble and Submit a .zip Folder with Result Files, Spreadsheet and Info File → Review Data Submission

The ImmPort Flow Cytometry multiple file upload process supports uploading of FCS files (.fcs) and tab delimited text files (.txt) converted from FCS files. Uploaded .fcs files are automatically converted to .txt via FCSTrans, the ImmPort flow cytometry data converter. All uploaded files will be stored in a private project workspace. For more information please see the [User Guide](#) or conversion tutorials found in the menu bar under [Help](#).

Batch Upload:

1. Complete the **flowTextFiles.xls** template. **Required template**
 - This template provides the opportunity to rename data files, allows for the addition of descriptive information about each individual file being uploaded and enables marker name changes via the inclusion of a marker information file.
 - The .fcs or fcs .txt data file names may NOT contain the following characters: "*" \ and /.
 - The fcs: data files should be compensated before sending to ImmPort.
2. Save the Excel flowTextFiles.xls template as a tab delimited text file (flowTextFiles.txt).
Please do not rename the flowTextFiles.txt file.
3. ZIP the completed template and the data files.
 - Create the ZIP file by selecting files from within a folder versus selecting the folder itself.
 - Add marker (channel) information files at this time if they are to be included (optional). [example Marker File \(.info\)](#)
 - Please do not include spaces in the .zip file name.**
 - You may find it helpful to review an [example multiple fcs files upload package](#).
4. Select the project to which the data will be saved
5. Use the Browse feature to select the .zip file and click Submit. Time to complete the processing of submitted data is dependent on the size of the submission queue.
6. To review the uploaded data go to [Data Management, View / Edit Uploaded Data](#)

Research Project Title: TESTING: Bioinformatics Integration Support Project

Notes:

Upload Online

Browse...

Upload file information and Browse

Figure 17: flowTextFiles.xls template

flowTextFiles.xls								
	A	B	C	D	E	F	G	H
1	List of FLOW text files for batch upload	Version 2.8.						
2	Please do not delete or edit this column	1	2	3	4	5	6	
3	Column Name	FCS Binary File Name*	FCS Text File Name*	Preferred Display Name	Description	Marker File	Experiment Sample User-Defined ID	Experiment Sample ImmPort Accession
4		FCS01tst.fcs		FCS01	fcsupload1	FCSBatch.in	FCSExpSample01tst	
5		FCS02tst.fcs		FCS02	fcsupload2	FCSBatch.in	FCSExpSample02tst	
6		FCS03tst.fcs		FCS03	fcsupload3	FCSBatch.in	FCSExpSample03tst	
7								

Complete the flowTextFiles.xls template as follows:

1. Fill in **FCS Files Names** for each .fcs file being uploaded—one file per row. The files names entered into the spreadsheet must match **exactly** those of the .fcs files. Uploaded FCS files will be automatically converted using the ImmPort conversion process. Please see the ImmPort Conversion tutorial found under 'Help' in the Flow Cytometry Analysis menu.
2. Fill in **FCS Text File Name** if uploading .txt converted FCS files instead of, or in addition to, the .fcs files. Note: If both FCS and .txt converted FCS files are uploaded ImmPort will not convert the .fcs file.
3. **Preferred Display Name** enables editing of FCS text file names to user-preferred names (optional)
4. **Description** of the file being uploaded (optional)
5. **Marker file** is used to edit the marker names within the flow cytometry file—see section 3.1 of the user guide for more details. (optional)
6. **Experiment Sample User-defined ID** and **Experiment Sample ImmPort Accession**. (optional)
The ImmPort accession is assigned by ImmPort when an experiment sample is submitted. The **User-defined ID** is provided by the user during data upload via the Experiment Sample submission template. Adding Experiment Sample information allows ImmPort to link uploaded flow cytometry data to Experiment Samples within a Project.
7. Convert the flowTextFiles .xls spreadsheet to a Tab-delimited format and save.

To create the .zip data package:

1. Convert the completed flowTextFiles.xls spreadsheet to text format (.txt).
2. Select individual flow cytometry files (.fcs and/or .txt), .info files if warranted and the flowTextFiles.txt template
3. Zip all as individual files—no folders are permitted in the zip archive. If folders are included in the .zip archive the system will return an error and no files will be uploaded.

Once the .zip archive is complete, click Browse from the Multiple File Upload screen and select the .zip archive to be uploaded. Click Submit on the Multiple File Upload screen. For additional assistance contact the helpdesk at helpdesk@import.org or review the **Upload Multiple File** tutorial found under Help in the ImmPort flow cytometry menu.

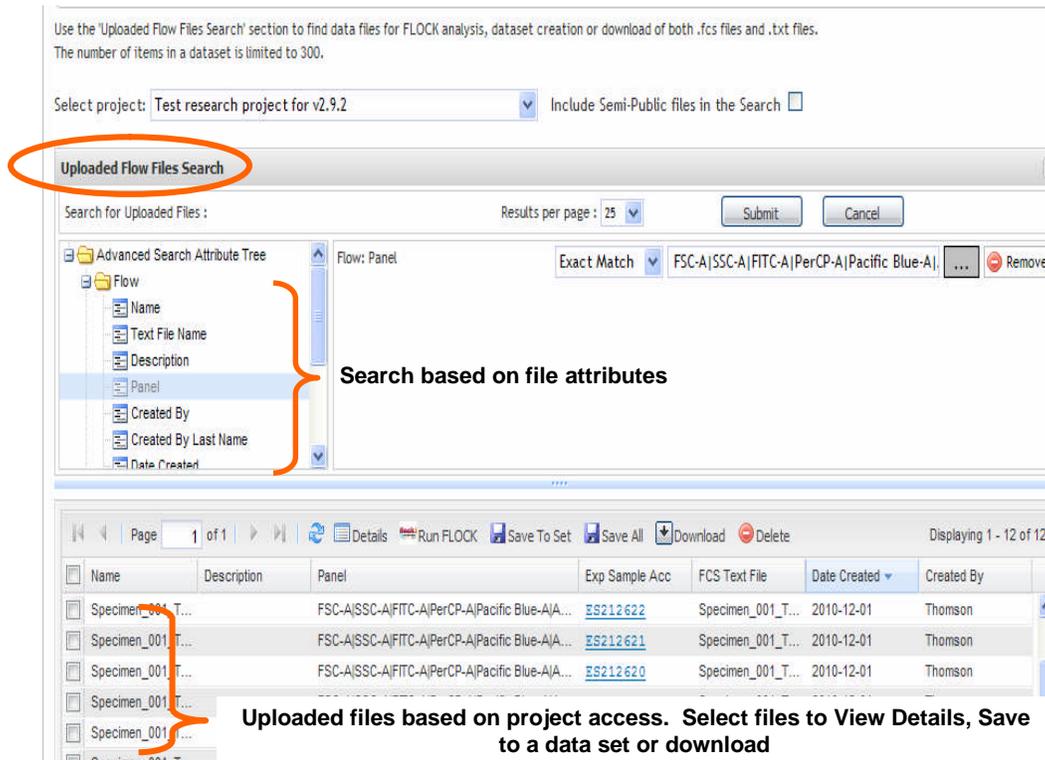
Making a list of many files from a desktop:

Listing FCS files in the flowTextFiles.xls spreadsheet can be time consuming if many files are involved. Please see the Making a List tutorial found under Help in the Flow Cytometry Tool menu.

4.3 SELECTING, EDITING AND ANALYZING UPLOADED FILES

Previously uploaded flow cytometry files may be selected for viewing, editing, creating data sets and FLOCK analysis through the Data Management module. Accessing the Data Management module, either from the Tools menu or from 'Data Management' dropdown within Flow Cytometry Analysis menu bar, will provide a link to 'View /Edit Uploaded Data'. The system will display the Uploaded Files Search screen which includes the Uploaded Flow Files Search section and a table displaying all uploaded data based on the project selected. The project selection drop down menu supports viewing of files based on project access.

Figure 18: Uploaded Files Search screen

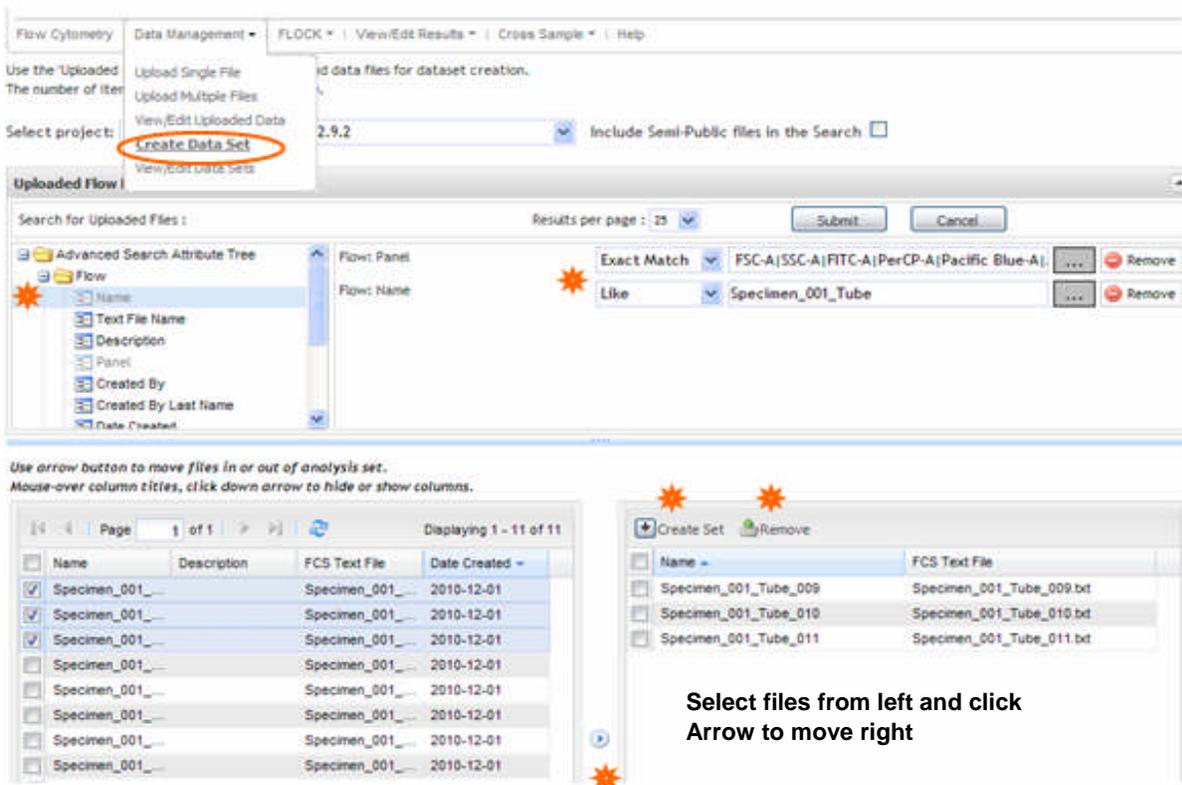


The Uploaded Files Search screen supports detailed viewing of individual or multiple files (tabbed display) from which it is possible to edit file information including file name, description and marker information. The Uploaded Files Search screen facilitates the creation of data sets for batch FLOCK analysis or Cross Sample analysis, running FLOCK for selected files (single or multi-select), downloading the converted .txt output for uploaded .fcs files or deleting files.

4.4 CREATE DATA SETS IN THE DATA MANAGEMENT MODULE

Grouping data into sets facilitates multiple-file FLOCK analysis and Cross Sample comparison. Figure 19 illustrates data set creation where the user selects files from the uploaded data table on the left and moves those files to the data set table on the right. Uploaded files displayed in the table (below, left) are based on the project chosen and the search criteria used in the Uploaded Files Search section. If no search is used all uploaded files for a chosen project will be displayed. Providing a name and project designation is required to save the data set. Providing a succinct description will aid in retrieval of the data set for future analyses. Once files have been selected, click Save. To view data sets select View/Edit Data Sets from the Data Management module drop down menu.

Figure 19: Data Set creation



5.0 DATA ANALYSIS WITH FLOCK

5.1 FLOCK v1 AND v2

ImmPort has two versions of the FLOCK algorithm; FLOCK v1 and FLOCK v2. The initial version of FLOCK, v1, supports the automatic identification of cell populations but requires defining appropriated bin and density threshold values for optimal results. Defining threshold values may be challenging—FLOCK v2 was developed to simplify the analysis process, reduce the need for defining thresholds and generates results similar to those achieved by manual gating without the time-consuming effort. Independent assessment of FLOCK v2 with other relevant methods will be released at FlowCAP:

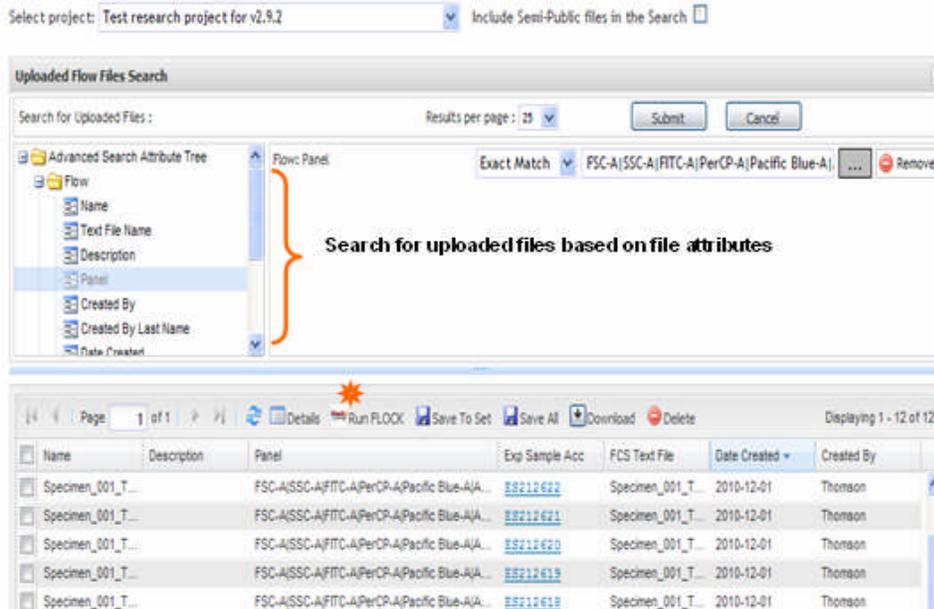
<http://flowcap.flowsite.org>. FLOCK v1 remains available to support existing analyses.

5.2 ACCESSING DATA FILES FOR FLOCK ANALYSIS

To access uploaded files for FLOCK analysis, click “FLOCK” from the Flow Cytometry menu, Data Management from the menu bar and select to analyze individual files or previously created data sets. Selection of more than one individual uploaded file for FLOCK analysis (batching) is enabled; however multi-selection is not supported for the analysis of data sets. Figure 20 shows accessing individual flow cytometry files for FLOCK analysis. The files displayed in the table below the search section are based on the project selected and search criteria. If no search is defined the system will display all files accessible within the selected project.

Figure 20: File selection

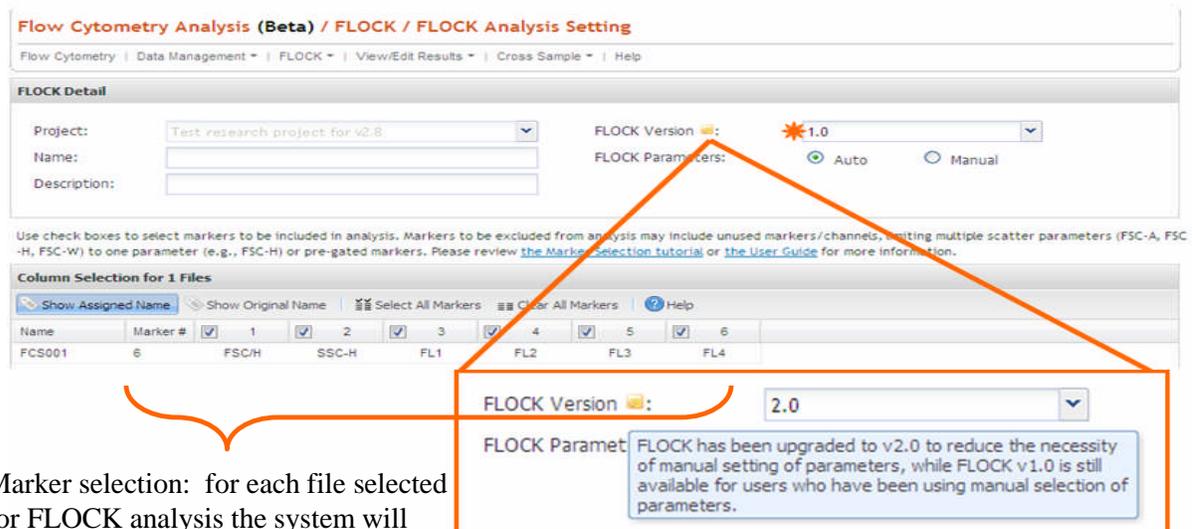
Use the 'Uploaded Flow Files Search' section to find data files for FLOCK analysis, dataset creation or download of both .fcs files and .txt files. The number of items in a dataset is limited to 300.



5.3 PROVIDING DESCRIPTIVE ANALYSIS INFORMATION

FLOCK analysis requires providing a name for the analysis task, choosing the private workspace where to save the FLOCK analysis results, selection of markers to be included in the analysis and defining FLOCK analysis parameters (auto or manual). Task descriptions are optional but can provide valuable information about the analysis for future reference. Figure 21 shows the FLOCK Analysis Setting screen. For details about **Marker Selection** see section 5.3

Figure 21: Define analysis task



Marker selection: for each file selected for FLOCK analysis the system will display a single marker per column. Check boxes allow for selection.

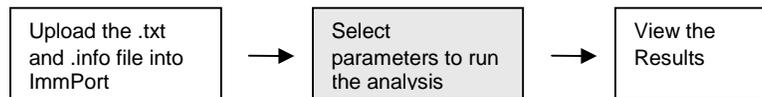
5.4 MARKER SELECTION

Marker selection is a pre-processing step in the ImmPort flow cytometry analysis pipeline which supports the selection of pertinent markers/fluorescence channels by data analysts prior to FLOCK analysis. Inclusion of non-pertinent channels distorts the real data distribution resulting in suboptimal analysis results. Scenarios where analyses would benefit from marker selection include:

- FCS files containing unused channels—this situation arises when an FCS file records signal from channels not stained in an experiment. One example of such a situation would be an experiment where stains were used for FL1, FL2 and FL4 but the FCS file contains signal from FL3. Another situation would be the inclusion of FSC-A, FSC-H and FSC-W while only FSC-A was intended for measurement.
- FCS files containing pre-gated markers—if a manually gated FCS file is uploaded to ImmPort there may be no reason to further partition the events based on the gated markers.

Selection of markers prior to FLOCK analysis is performed on uploaded files by the data analyzer—in the case of uploaded .fcs files, ImmPort automatically converts these to files to .txt prior to display for marker selection. Marker selection does not affect the original uploaded files as the ImmPort system extracts the user-specified data subset for analysis and visualization. In Cross Sample comparisons, data columns are automatically selected to be consistent with those of the centroid file being applied (see Figure 32 for more details).

5.5 FLOCK PARAMETER SELECTION



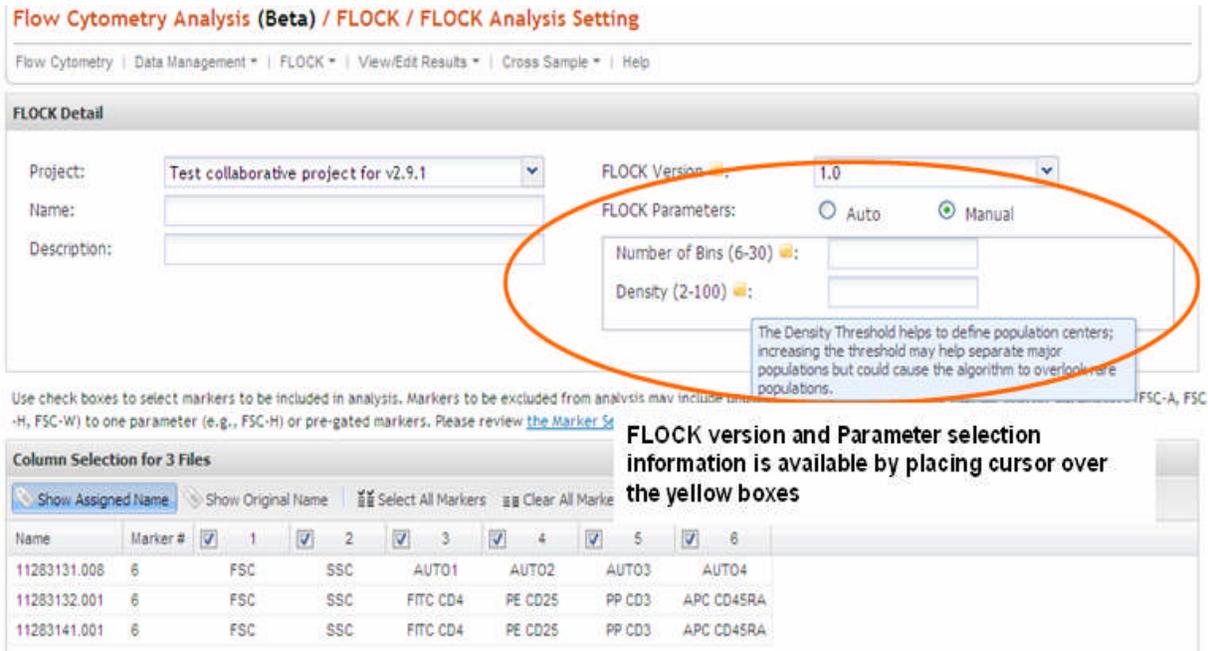
5.5.1 Using Auto-detect parameter setting

The default parameter setting for FLOCK is auto-detect. Auto-detect allows FLOCK to determine the best parameters for the analysis.

5.5.2 Using the Customize Parameter setting

Customizing the FLOCK parameter settings includes defining the number of bins and density with which FLOCK will analyze the data. Figure 22 is a screenshot of FLOCK parameter selection.

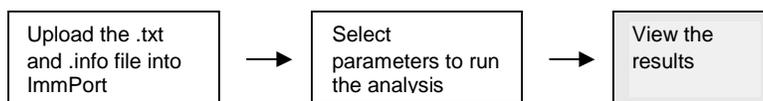
Figure 22: Parameter selection



5.5.2.1 Number of bins is an integer specifying how many equal-sized regions the data will be partitioned into on each axis. Increasing the number of bins increases the sensitivity to detect rare populations but may also result in single populations being divided.

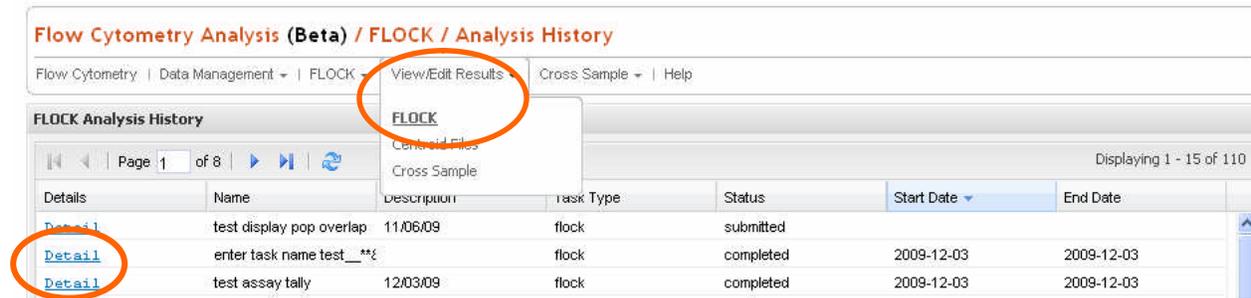
5.5.2.2 Density Threshold is the cut-off value to separate the dense regions from background. It is a floating number that helps define population centers; increasing the threshold may help separate major populations but could cause the algorithm to overlook rare populations.

5.5.3 Viewing FLOCK analysis results



After parameter selection is completed and “Run FLOCK” is clicked, you will be notified of the Task ID. To access the FLOCK results, click “View/Edit Results” from the menu bar and then click “FLOCK”. A table of all the submitted FLOCK analyses will display from which details of the analyses may be viewed by clicking “Detail”. (Figure 23)

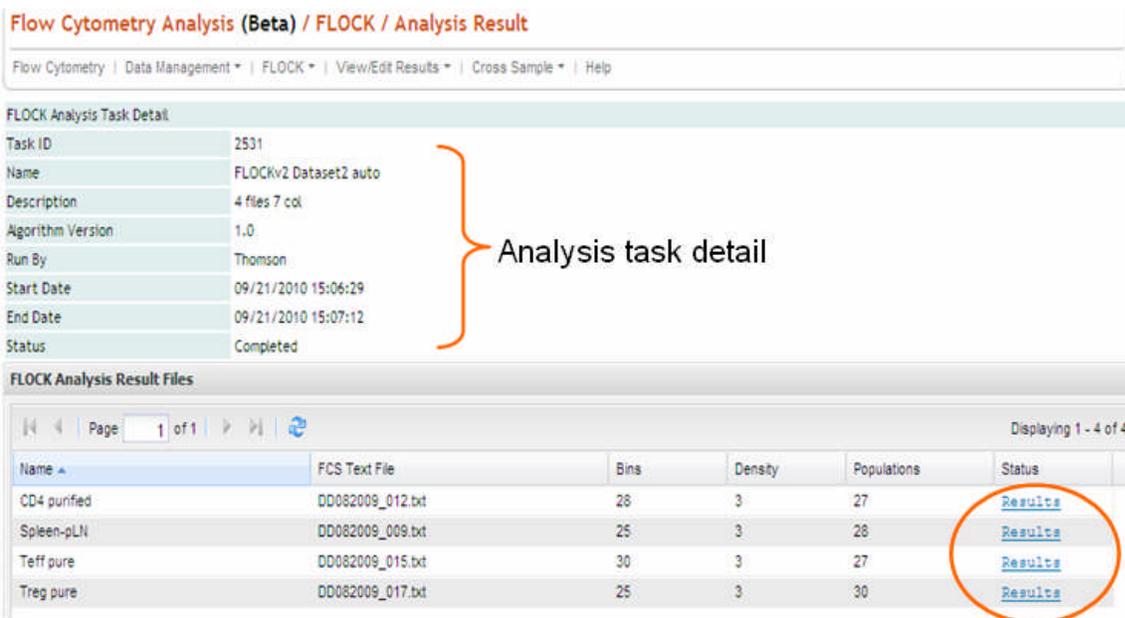
Figure 23: Show the analysis results from the FLOCK results table



5.6 FLOCK ANALYSIS DETAIL

Clicking “Detail” (Figure 23) will display the analysis task detail and provide a “Results” link to visualize FLOCK output.

Figure 24: FLOCK analysis task detail



5.7 FLOCK RESULTS VISUALIZATION

The “Results” link from the Analysis Task Detail screen (Figure 24) will display, in a separate screen, the thumbnail overview where each population is assigned a color (Figure 25)—clicking any thumbnail will link to the FLOCK/RAS Result Overview (Figure 26). The screen also displays the Population Selection table (discussed in a later section) which allows the user to select populations to view in the overview matrix.

Figure 25: FLOCK analysis results detail—thumbnail matrix with Population Selection table

Result Adjustment System(RAS) - Task Name: FLOCK_analysis 1 File Name: FCS001.txt

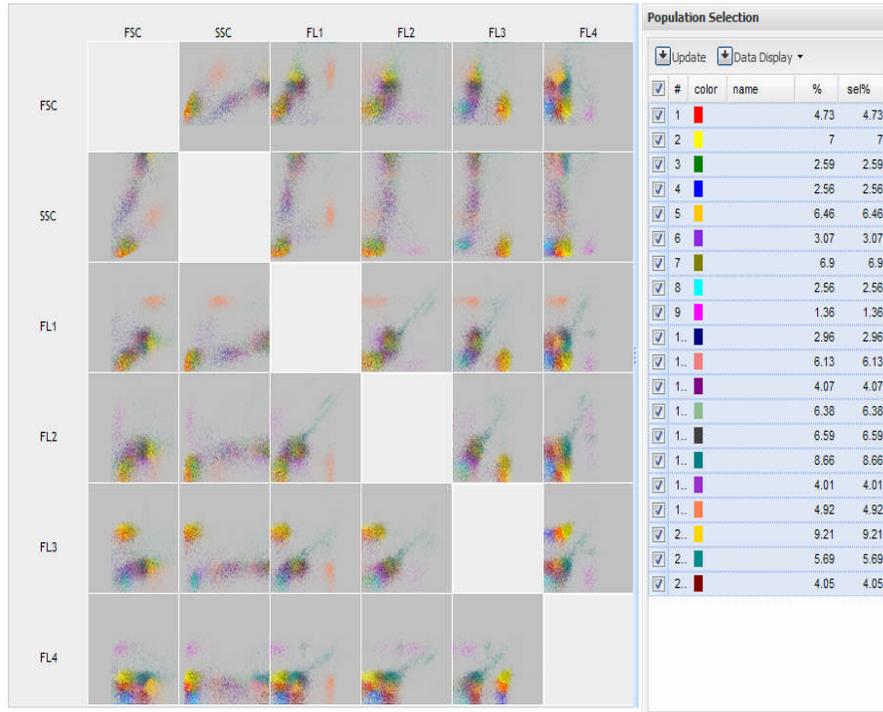


Figure 26: Small Image Matrix of FLOCK Results



Click on a thumb nail image in Figure 26 to view a larger image and adjust populations generated by FLOCK. Please see Section 6 for result adjustment details.

5.7.1 FLOCK Population ID

FLOCK assigns a population ID to each event in the flow cytometry result file. The FLOCK results display enables you to view different representations of the FLOCK results. Each Population ID is assigned a color for display in two dimensional matrices. For example, population ID = 1 corresponds to color RED and population ID=2 corresponds to color yellow.

Available population information can be obtained from the Population Selection Table where the dropdown Data Display menu provides the option to view Centroid, Score or Mean Fluorescent Intensity values.

5.7.2 Population Selection Table: Summary Table of all Analysis Populations

Each population group is likely to represent distinct proportions of the total cell population and these proportions can be very useful in immunology research and disease diagnosis. Proportion of population can also be used to generate standard deviation (or variation) of population proportions across different samples. (Figure 27) Use the Data Display dropdown menu to view Centroid, Score or Mean Fluorescent Intensity values. To view a subset of populations in the matrix overview use the selection boxes and click 'Update'—the 2-D display matrix will display only the selected populations across all marker pairs.

Figure 27: Population Selection Table

#	color	sel%	mfi FSC	mfi SSC	mfi FL1	mfi FL2	
1	Red		391	111	94	48	
2	Yellow		397	119	100	70	
3	Green		392	180	110	89	
4	Blue	2.56	419	130	118	137	
5	Orange	6.46	100	437	140	124	183
6	Purple	3.07	485	232	233	223	
7	Olive	6.9	457	164	149	194	
8	Cyan	2.56	405	175	115	144	
9	Magenta	1.36	377	152	83	492	
1..	Dark Blue	2.96	402	452	207	287	
1..	Red	6.13	565	914	263	337	
1..	Purple	4.07	497	639	197	299	
1..	Green	6.38	535	683	294	391	
1..	Black	6.59	614	1009	359	341	
1..	Teal	8.66	596	984	246	321	
1..	Purple	4.01	602	951	311	194	
1..	Orange	4.92	732	445	651	135	
2..	Yellow	9.21	707	1011	284	315	
2..	Teal	5.69	800	1000	398	444	
2..	Red	4.05	625	992	295	298	

The Score value is a number indicating the degree this population expresses a marker. The index code is:

- 1 implies negative expression
- 2 implies low expression

- 3 implies positive expression
- 4 implies highly positive.

The user may fill the column “Name” with whatever he/she wants to name the population, e.g., CD4+T Helper cells.

5.7.3 Centroid Table

The values in this table (Figure 28) are the coordinates of the displayed centroids that can be adjusted by the user through the graphical interface to modify the populations generated by FLOCK

Figure 28: Centroid Table

Flow Cytometry Analysis (Beta) / RAS / Result Overview

Overview ▾ Summary Tables ▾ Download ▾ | Centroid ▾ | Help ▾ |

Result Ad Results) - Task Name: Run FLOCK Analysis File Name: FLOCK_Data_Fil

Population Id	Centroid	Mean Fluorescent Intensity (MFI)	CD21	AA4.1	CD19	CD24	CD23
1			40.0	42.0	18.0	75.0	38.0
2	82.0	100.0	24.0	37.0	244.0	365.0	335.0
3	234.0	170.0	63.0	180.0	388.0	421.0	339.0
4	229.0	175.0	280.0	127.0	54.0	227.0	163.0
5	56.0	45.0	13.0	32.0	147.0	267.0	161.0
6	366.0	278.0	271.0	297.0	539.0	510.0	453.0
7	417.0	335.0	312.0	444.0	609.0	580.0	214.0
8	398.0	369.0	311.0	353.0	593.0	561.0	574.0
9	560.0	45.0	29.0	24.0	48.0	77.0	124.0

5.7.4 Mean Fluorescence Intensity Table (MFI)

The value at row X column Y is the mean of the fluorescence intensity on parameter/marker Y of all events of the population X. (Figure 29)

Figure 29: Mean Fluorescence Intensity Table (MFI)

Flow Cytometry Analysis (Beta) / RAS / Result Overview

Overview ▾ Summary Tables ▾ Download ▾ | Centroid ▾ | Help ▾ |

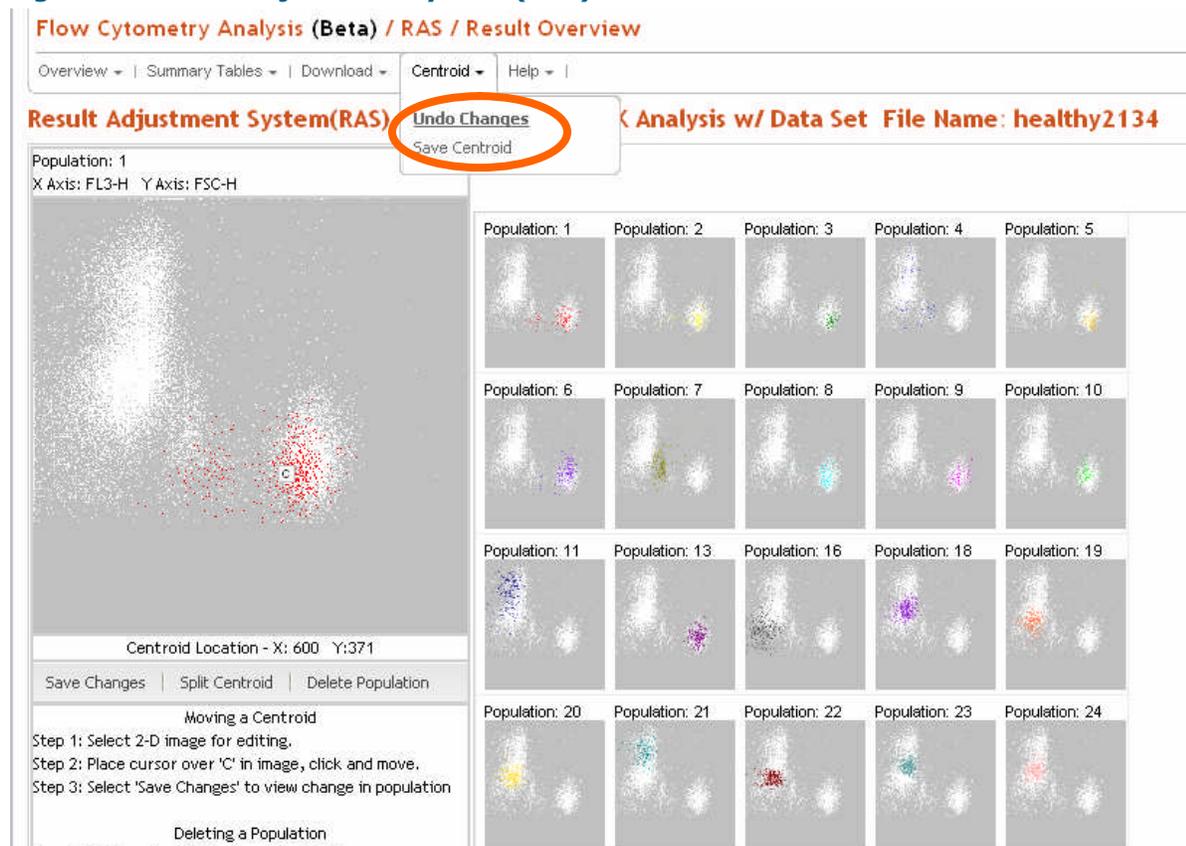
Result Ad Results Task Name: Run FLOCK Analysis File Name: FLOCK_Data_File

Population Id	Centroid	Mean Fluorescent Intensity (MFI)	CD21	AA4.1	CD19	CD24	CD23
1			40.0	42.0	18.0	75.0	38.0
2	82.0	100.0	24.0	37.0	244.0	365.0	335.0
3	234.0	170.0	63.0	180.0	388.0	421.0	339.0
4	229.0	175.0	280.0	127.0	54.0	227.0	163.0
5	56.0	45.0	13.0	32.0	147.0	267.0	161.0
6	366.0	278.0	271.0	297.0	539.0	510.0	453.0
7	417.0	335.0	312.0	444.0	609.0	580.0	214.0
8	398.0	369.0	311.0	353.0	593.0	561.0	574.0
9	560.0	45.0	29.0	24.0	48.0	77.0	124.0

6.0 RESULT ADJUSTMENT SYSTEM (RAS)

The Result Adjustment System (RAS) exists outside of FLOCK and is intended for use after FLOCK analysis has been completed. The user selects the analysis results from the FLOCK Analysis History; clicking on any thumbnail of the result overview will provide the user with the opportunity to adjust the populations (one population at a time).

Figure 30: Result Adjustment System (RAS)



As Figure 30 shows, the user can select, from the right side of the display, the population he/she wants to modify and move or split the population centroid or delete the population (i.e. reassign its events to other populations) through the left panel. The 2-D displays and summary tables will automatically refresh to reflect the changes made by the user. The user is also allowed to return to the original FLOCK result by discarding all manual changes (click Undo Centroid Adjustment in the navigation bar above the 2-D displays)

Saving centroids, by selecting Save Centroid from the menu bar (see figure 28) allows the user to retain centroid adjustments. Saved centroid files can be used within the Cross Sample comparison module to map populations across multiple flow samples. The current ImmPort release supports the use of RAS and saving centroid with FLOCK 1 output—FLOCK 2 output will be supported in a future release.

7.0 CROSS SAMPLE COMPARISON

Cross Sample Comparison automatically maps populations of cells across multiple flow samples and computes the summary statistics for downstream analysis. Currently Cross Sample Comparison is available for FLOCK v1 output only—FLOCKv2 output Cross Sample Comparison will be available in a subsequent release. To perform cross sample analysis the user is required to define a data set which

contains the samples to be compared using the Data Management module. See section **4.4 Create Data Sets in the Data Management Module** for details.

A centroid file is required for comparison and is generated by running FLOCK on one or more files of the data set or on a file external to the data set to which the data set is being compared. See section **6.0 Result Adjustment System (RAS)** for details on adjusting and saving population centroids. Note: the number of markers (columns in each file) must be consistent within the data set and the saved centroid file in order to use Cross Sample comparison.

Access Cross Sample comparison from the Flow Cytometry menu bar (Figure 31)

Figure 31: Cross Sample comparison

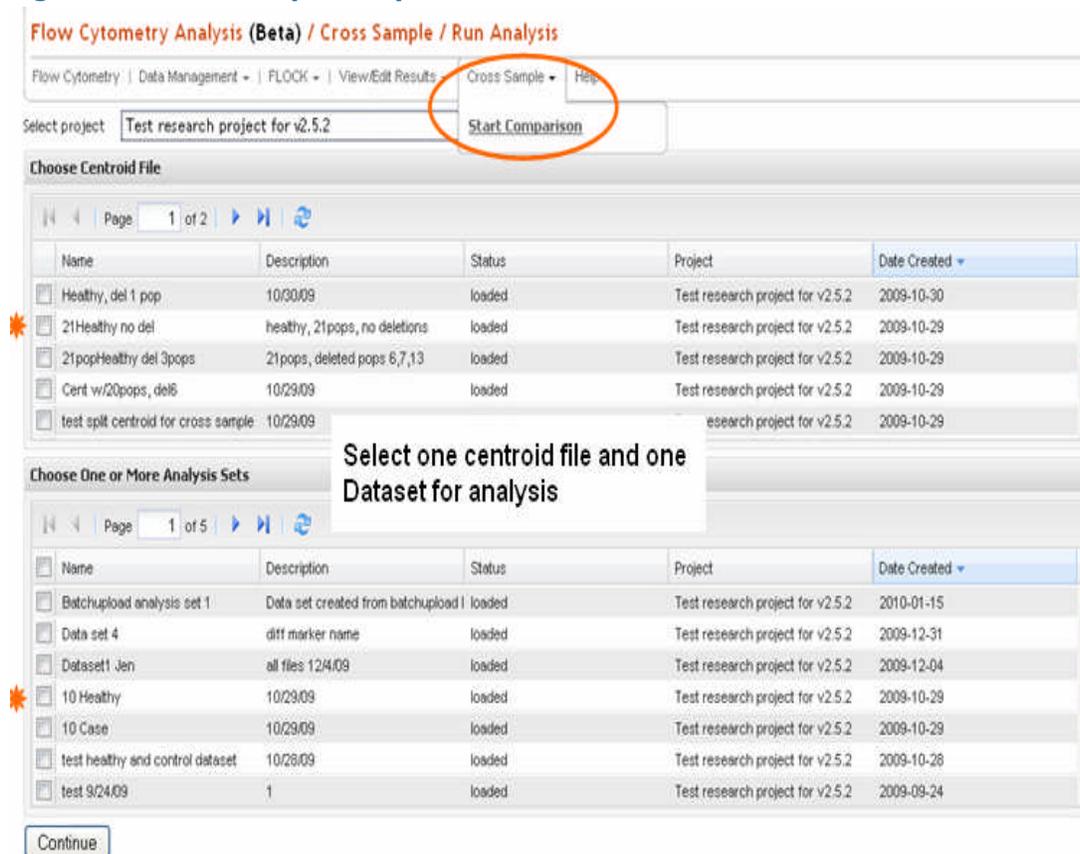


Figure 32: Cross Sample comparison

Marker display of all files in dataset compared to centroid selected for Cross Sample comparison. Green color indicates marker name identical, red indicates marker name is not identical.

Flow Cytometry Analysis (Beta) / Cross Sample / Cross Sample Analysis Setting

Flow Cytometry | Data Management | FLOCK | View/Edit Results | Cross Sample | Help

Cross Sample

Project: Test research project for v2.8
 Name: Cross Sample Analysis 1
 Description: 3 Files 5 columns

Marker Selection

Show Assigned Name Show Original Name Help

Type	Name	Marker #	1	2	3	4	5	6	
Centroid	save cent no cha	6	FSC-H	SSC-H	FL1-H	FL2-H	FL3-H	FL4-H	
Flow File	FCSfile003	6	FSC-H	SSC-H	FL1-H	FL2-H	FL3-H	FL4-H	
Flow File	206B_Olym	6	FSC-H	SSC-H	FL1-H	FL2-H	FL3-H	FL2-A	FL4-H
Flow File	Sample 002	6	FSC-H	SSC-H	FL1-H	FL2-H	FL3-H	FL4-H	FL4-H Time

Not included in analysis

- Markers displayed are compared to the centroid which appears in row one of the table
- Green indicates identical marker names
- Red indicates a marker name mismatch between the files in the data set and the centroid file.
- Note only those files which are included in the centroid are included in the Cross Sample comparison

Run Cross Sample

Figure 33: Cross Sample comparison-viewing results

Flow Cytometry Analysis (Beta) / Cross Sample / Analysis Detail

Flow Cytometry | Data Management | FLOCK | View/Edit Results | Cross Sample | Help

Cross Sample Analysis

Task ID: 973
 Name: Compare dataset 4
 Description: 2 files: diff marker names
 Start Time: 12/31/2009 12:33:13
 End Time: 12/31/2009 12:33:29
 Status: [Analysis Results](#)

Centroid Information [Detail](#)

Name: Test centroid save_del pop1
 Description: 12/16/09

Flow Analysis Sets

Name	Description	Status
Data set 4	diff marker name	loaded

Cross Sample Analysis Result Files

Individual file results

Name	Description	FCS Text File	Status
Test IE upload2	12/16/09	02054004_002.txt	Results
test .info files	Jen3infofile	02054004_001fewrows.txt	Results

7.1.1 Population Percentage table

The Population Percentage table displays the proportion of each population within a sample based on the centroid applied in the cross sample comparison. (Figure 34) Individual samples appear in rows and sample name links will take the user to a 2-D thumbnail display where the user can continue with centroid refining.

Figure 34: Cross Sample Population Percentage table

Flow Cytometry Analysis (Beta) / Cross Sample / Analysis Result

Percentage Table | Marker by Population | Download ▾

Cross Sample Analysis - Task Name: test cross 11/02/09

File	Pop 1	Pop 2	Pop 3	Pop 4	Pop 5	Pop 8	Pop 9	Pop 10	Pop 11	Pop 12	Pop 14	Pop 16	Pop 17	Pop 18	Pop 19
healthy2054	2.69	3.95	4.27	1.48	3.41	2.33	4.8	1.16	7.39	2.54	12.45	7.64	10.79	12.47	3.8
healthy2094	4.32	7.58	5.54	3.9	5.92	2.03	2.7	2.44	2.16	3.83	4.61	22.5	4.89	7.32	5.2
healthy2124	3.75	4.82	6.88	1.88	5.47	2.54	3.69	0.63	5.86	1.96	9.28	19.61	8.66	7.32	5.3
healthy2174	2.84	4.56	8.45	1.4	2.49	1.65	3.98	0.29	7.18	3.52	14.49	3.33	8.95	20.36	4.3
healthy2194	2.39	8.08	7.55	2.99	12.85	4.27	2.28	2.21	4.29	5.81	8.22	4.39	11.82	9	3.9
healthy2064	3.93	7.81	3.8	2.37	4.59	2.73	2.6	0.98	4.67	3.5	8.5	14.7	9	11.2	4
healthy2114	1.75	5.85	5.76	1.3	7.83	3.31	2.41	1.83	8.26	1.56	11.74	3.08	11.74	10.56	2.3

7.1.2 Marker by Population visualization

The Marker by Population visualization is accessed through the Analysis Result menu bar (see Figure 35). Populations of interest are selected from a drop down menu as are the markers of interest. Visualization includes 2-D images for each marker pair and population selected for all samples in the comparison.

Figure 35: Cross Sample Marker by Population

Flow Cytometry Analysis (Beta) / Cross Sample / Analysis Result

Percentage Table | **Marker by Population** | Download ▾

Cross Sample Analysis - Task Name: test cross 11/02/09

Population: 5 | Marker 1: FSC-H | Marker 2: SSC-H |

Population: 5 Marker 1: FSC-H Marker 2: SSC-H

7.1.3 Downloading Cross Sample results

Within the menu bar for Cross Sample comparison the user will find Download options (Figure 36). Downloads include analysis statistics which can be opened via Excel (Figure 37) and includes the sample name, number of events for each population, population proportion, mean, standard deviation and coefficient of variance for each marker. Additionally, all analysis results may be downloaded via the **All Results** option.

Figure 36: Download options

Flow Cytometry Analysis (Beta) / Cross Sample / Analysis Result

Percentage Table | Marker by Population | Download ▾

Cross Sample Analysis - Task Statistics 11/02/09

All Results

File	Pop 1	Pop 2	Pop 3	Pop 4	Pop 5	Pop 6	Pop 7	Pop 8	Pop 9	Pop 10	Pop 11	Pop 12
healthy2054	2.69	3.95	4.27	1.48	3.41	2.33	4.8	1.18	7.39	:		
healthy2094	4.32	7.58	5.54	3.9	5.92	2.03	2.7	2.44	2.16	:		
healthy2124	3.75	4.82	6.88	1.88	5.47	2.54	3.69	0.63	5.86	:		
healthy2174	2.84	4.56	8.45	1.4	2.49	1.65	3.98	0.29	7.18	:		
healthy2194	2.39	8.08	7.55	2.99	12.85	4.27	2.28	2.21	4.29	:		

Figure 37: Statistics download

_home_jboss_flockResults_830_statistics.all-1.txt [Read-Only]

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1						FSC-H			SSC-H			FL1-H			FL2-H
2	Population	File Name	File Id	Events	Percentage	Mean	STD	Coefficient	Mean	STD	Coefficient	Mean	STD	Coefficient	Mean
3	1	healthy2054	71	538	2.69	393	49	0.12	107	34	0.32	465	29	0.06	285
4	2	healthy2094	71	789	3.95	436	57	0.13	135	46	0.34	468	38	0.08	354
5	3	healthy2124	71	853	4.27	418	73	0.17	166	49	0.3	70	62	0.89	183
6	4	healthy2174	71	296	1.48	405	76	0.19	247	151	0.61	384	110	0.29	532
7	5	healthy2194	71	682	3.41	407	56	0.14	156	51	0.33	71	48	0.68	125
8	6	healthy2054	71	465	2.33	381	50	0.13	112	40	0.35	458	36	0.08	91

8.0 FLOCK AT SOURCEFORGE

FLOCK source code is available at <http://importflock.sourceforge.net/>

9.0 FUTURE WORK

Implementing FLOCK in ImmPort aims to allow ImmPort users to study and compare cell phenotypes including expression levels of cell-surface markers, the proportion values of cell populations, and their changes across different samples and different disease developmental stages. Future implementation of FLOCK in ImmPort will include better support for high-dimensional flow cytometry data. Results of FLOCK analysis will be parsed and archived into ImmPort FCM result database to allow query and management. Also results of FLOCK could be easily encoded in cell ontology terms since different cell types are defined based on their marker expressions which can be output by FLOCK in a standardized way. Therefore cell phenotype information can be easily reused in other integrated analysis ImmPort has

and will provide. Also the analysis is done in an automated way to free ImmPort users from laborious and subjective gating and comparison efforts.

Future development cycles will extend the UI of FLOCK to facilitate a more easy-to-use environment. Other useful visualizations such as heat maps can be provided to allow an intuitive comparison between subject groups. The FLOCK algorithm will be improved and re-designed based on feedback from ImmPort users. As FLOCK becomes more widely adopted by biomedical researchers it will continue to offer significant advantages to the ImmPort research community.

If you have any questions or comments on how to use FLOCK for Flow Cytometry analysis, please contact ImmPort help desk at helpdesk@immport.org.

APPENDIX A GLOSSARY

Glossary	
Term	Acronym
Biological Sample	BS
Data Management	DM
Experiment Sample	ES
Identifier	ID
Query Interface	QI
Research Projects	RP
Search Research Data	SRD
University of Texas Southwestern Medical Center	UTSW