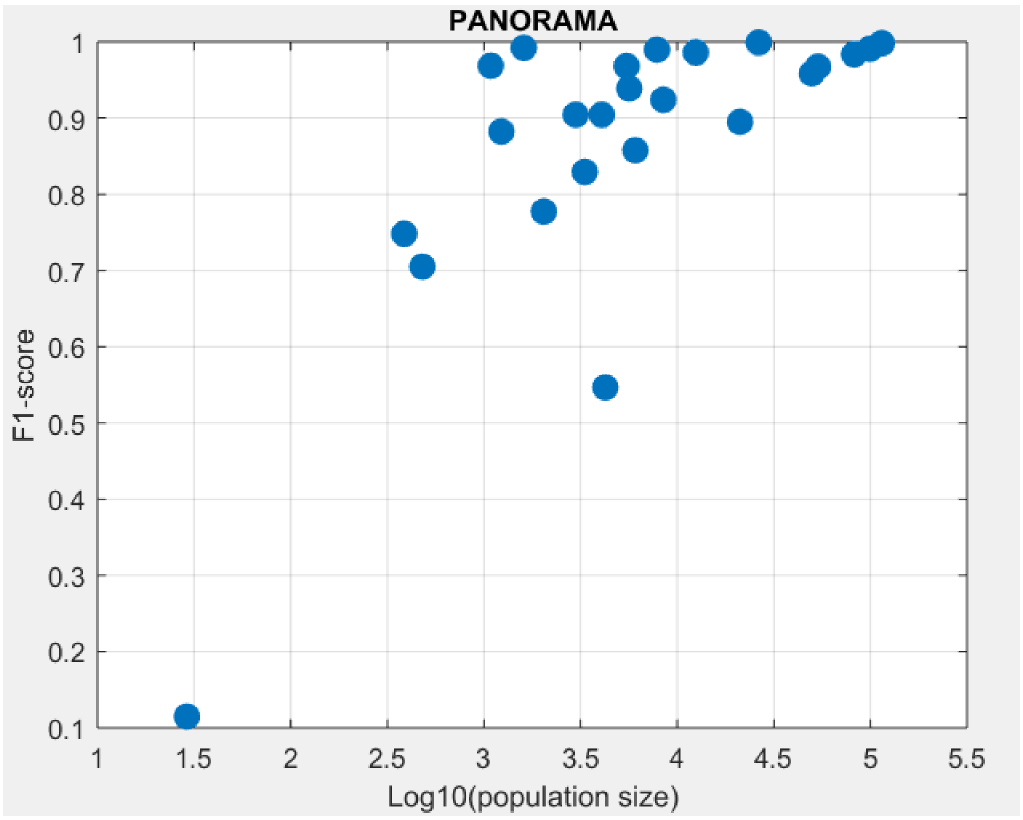
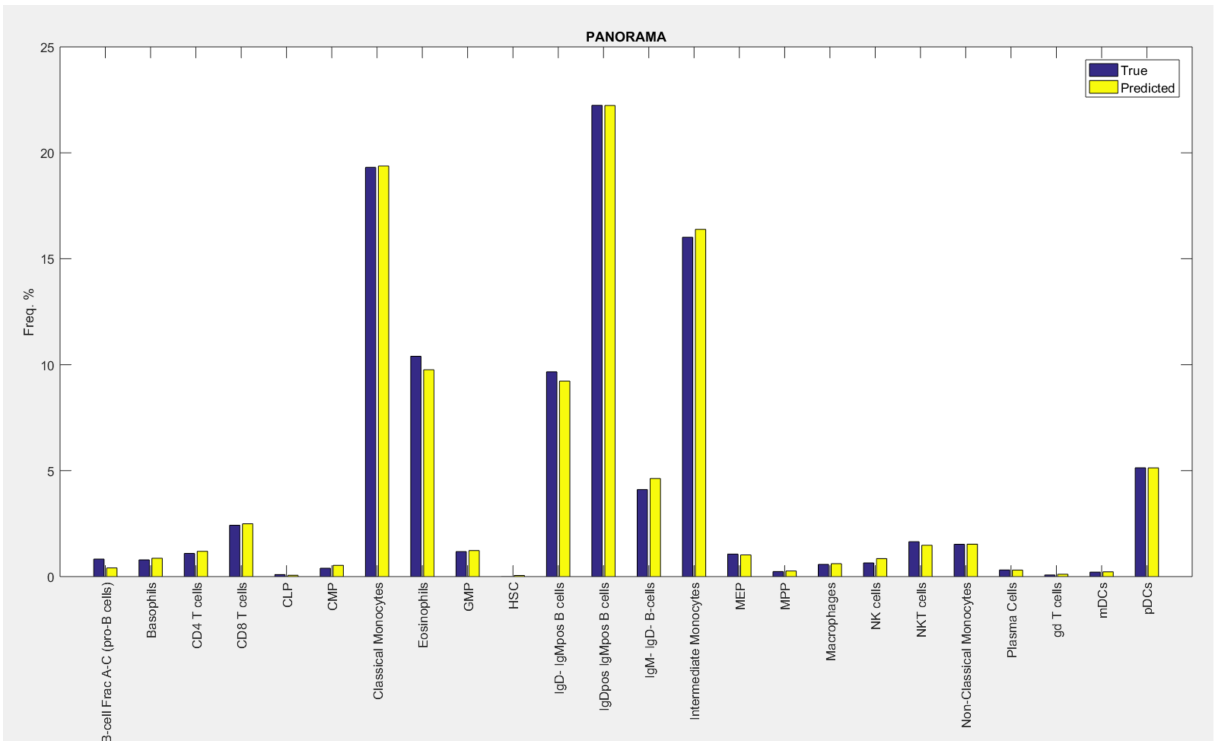
PANORAMA Dataset Classification

```
%% Read the Data and Preprocess

>> VarNames = {'Ter119';'CD45.2';'Ly6G';'IgD';'CD11c';'F480'; ...
   'CD3';'NKp46';'CD23';'CD34';'CD115';'CD19';'120g8';'CD8'; ...
   'Ly6C';'CD4';'CD11b';'CD27';'CD16_32';'SiglecF';'Foxp3'; ...
   'B220';'CD5';'FceR1a';'TCRgd';'CCR7';'Sca1';'CD49b';'cKit';...
   'CD150';'CD25';'TCRb';'CD43';'CD64';'CD138';'CD103';'IgM'; ...
   'CD44';'MHCII'};

>> SamplesData=struct('Data',[],'Labels',{});
>> H=dir(fullfile('Samples\', '*.csv'));
>> SamplesFiles = cellstr(char(H(1:end).name));

>> H=dir(fullfile('Labels\', '*.csv'));
>> LabelsFiles = cellstr(char(H(1:end).name));
>> clear H

>> for i=1:length(SamplesFiles)
>>     SamplesData(i).Data = csvread(['Samples\' SamplesFiles{i}]);
>>     SamplesData(i).Labels = table2cell(readtable(['Labels\'...
       LabelsFiles{i}],'ReadVariableNames',0,'Delimiter',','));
>> end
>> clear i SamplesFiles LabelsFiles

>> Labels = [];
>> for i=1:length(SamplesData)
>>     Labels = [Labels; SamplesData(i).Labels];
>> end
>> clear i

% Data is already arcsinh(5) transformed

%% run LDA Classifier with 5-fold cross-validation on samples

>> CVO = cvpartition(1:1:10,'k',5);
>> Accuracy = zeros(length(SamplesData),1);
>> training_time = zeros(CVO.NumTestSets,1);
>> testing_time = zeros(length(SamplesData),1);
>> CellTypes = unique(Labels);
>> ConfusionMat = zeros(length(CellTypes));
>> for i = 1:CVO.NumTestSets
>>     trIdx = find(CVO.training(i));
>>     teIdx = find(CVO.test(i));

>>     DataTrain=[];
>>     LabelsTrain=[];
>>     for j=1:length(trIdx)
>>         DataTrain = [DataTrain; SamplesData(trIdx(j)).Data];
>>         LabelsTrain = [LabelsTrain;SamplesData(trIdx(j)).Labels];
>>     end
>>     clear j
>>     tic
>>     classificationLDA = fitcdiscr(...
>>         DataTrain, ...
>>         LabelsTrain);
>>     training_time(i)=toc;          %in seconds

>>     for j=1:length(teIdx)
>>         tic
>>         Predictor = predict(classificationLDA, ...
           SamplesData(teIdx(j)).Data);
>>         testing_time(teIdx(j))=toc;           %in seconds
>>         Accuracy(teIdx(j)) = nnz(strcmp(Predictor, ...
           SamplesData(teIdx(j)).Labels)) ...
           /size(SamplesData(teIdx(j)).Labels,1);
>>         ConfusionMat = ConfusionMat + ...
           confusionmat(SamplesData(teIdx(j)).Labels, ...
           Predictor,'order',CellTypes);
>>     end
>>     clear j
>> end
>> Total_time = sum(training_time)+sum(testing_time);
>> training_time = mean(training_time);
>> testing_time = mean(testing_time);
>> cvAcc = mean(Accuracy)*100;
>> cvSTD = std(Accuracy)*100;
>> disp(['LDA Accuracy = ' num2str(cvAcc) ' ' char(177) ...
   ' ' num2str(cvSTD) ' %'])

LDA Accuracy = 97.1205 ± 0.33418 %

>> clear i Predictor classificationLDA trIdx teIdx CVO Accuracy
   DataTrain LabelsTrain


%% Performance evaluation

% F1 measure
>> Precision = diag(ConfusionMat)./sum(ConfusionMat,1)';
>> Recall = diag(ConfusionMat)./sum(ConfusionMat,2);
>> Fmeasure = 2 * (Precision.*Recall)./(Precision+Recall);
>> MedianFmeasure = median(Fmeasure);
>> Subset_size = sum(ConfusionMat,2);
>> WeightedFmeasure = (Subset_size./sum(Subset_size))'*Fmeasure;

>> disp(['Median F1-score = ' num2str(MedianFmeasure)])

Median F1-score = 0.93149

>> figure,scatter(log10(Subset_size),Fmeasure,100,'filled')
>> title('PANORAMA')
>> xlabel('Log10(population size)'),ylabel('F1-score')
>> box on, grid on
```



```
%% Population Frequency

>> True_Freq = sum(ConfusionMat,2)./sum(sum(ConfusionMat));
>> Predicted_Freq = sum(ConfusionMat,1)'./sum(sum(ConfusionMat));
>> Max_Freq_diff = max(abs(True_Freq-Predicted_Freq))*100;

>> disp(['delta_f = ' num2str(Max_Freq_diff)])

delta_f = 0.63688

>> figure,bar([True_Freq*100 Predicted_Freq*100])
>> xticks(1:24)
>> xticklabels(CellTypes)
>> xtickangle(90)
>> set(gca,'FontSize',10)
>> set(gca,'XLim',[0 25])
>> legend({'True','Predicted'},'FontSize',10)
>> legend show
>> ylabel('Freq. %'),title('PANORAMA')
```



PANORAMA
Freq. %
25
20
15
10
5
0
True
Predicted
B-cell Frac A-C (pro-B cells)
Basophils
CD4 T cells
CD8 T cells
CLP
CMP
Classical Monocytes
Eosinophils
GMP
HSC
IgD- IgMpos B cells
IgDpos IgMpos B cells
IgM- IgD- B-cells
Intermediate Monocytes
MEP
MPP
Macrophages
NK cells
NKT cells
Non-Classical Monocytes
Plasma Cells
gd T cells
mDCs
pDCs