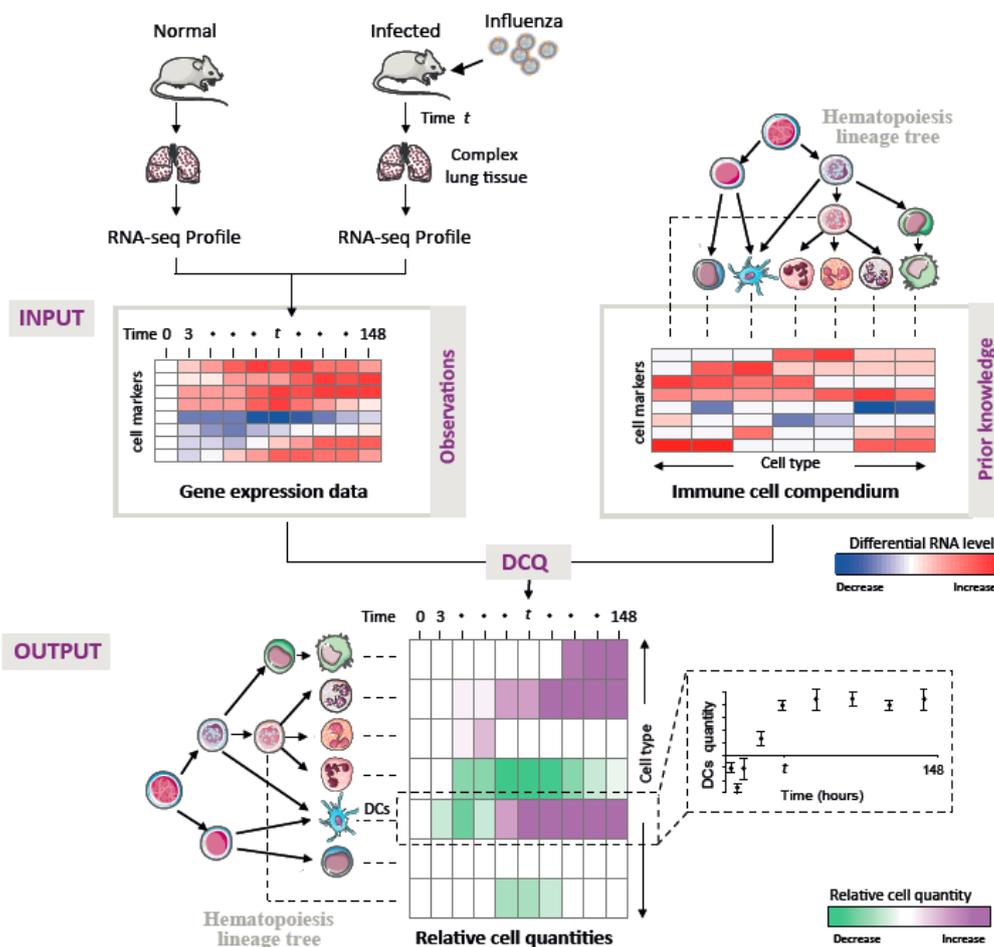


# DCQ application

## Quick start manual



### Digital cell quantification identifies global immune cell dynamics and the temporal roles of dendritic cell subtypes in Influenza infection

Zeev Altboum, Yael Steerman, Eyal David, Zohar Barnett Itzhaki, Liran Valadarsky, Hadas Keren-Shaul, Tal Meningher, Ella Mendelson, Michal Mandelboim, Irit Gat-Viks and Ido Amit

# 1. Quick start

The following manual describes how to use DCQ application while applying default parameters.

DCQ application is operational on different browsers including FireFox version 22 (22.0 and higher), Chrome version 28 (28.0.1500.72 and higher), Safari version 6 (6.0.2 and higher) and Explorer version 10 (10.0.9200 and higher).

## **General Overview:**

DCQ application takes the following inputs:

1. An immune cell compendium, consisting of transcriptional profiles of isolated immune cell subsets, taken from various tissues, stimulations and time points.
2. A collection of immune cell surface markers designed to discriminate between the immune cell types included in the compendium.
3. RNA-seq or microarray differential gene expression profiles of a test sample compared to a reference sample.

Here, inputs (1) and (2) are provided as part of the application.

Input (3), the differential gene expression file, should be uploaded as a data file by the user.

For the complete list of cells that appear in the compendium, please download [this file](#). In the compendium file shown are the general immune cell populations (column 1), the particular name for each cell type (column 2). The particular markers that have been exploited for the isolation of the cell type (by FACS) are detailed in column 3. Column 4 provides information on whether this cell type is stimulated or is a known effector or memory immune cell type and the tissue from which a cell type was isolated is detailed in column 5.

The output is the predictions of relative cell quantity for each immune cell type that appears in the compendium, for each of the test samples.

## **Running DCQ.**

**Step1:** Preparing the RNA-seq profiles of a test sample compared to a reference sample

There are several requirements for the RNA-seq / microarray file:

- The file should be tab-delimited file
- Each row represents an expression of a gene, and each column represents a test sample.
- The first row is the name of each test sample and the first column includes the genes' symbol.
  - The names of each test sample and each gene should be unique.
  - The names of the columns should be without spaces.
- Any gene with missing expression values should be excluded.

Proposed pre-processing of the data:

- Apply a log2 transformation
- Normalize each entry (test and reference) by the gene's average across all reference samples (test entry – average(reference sample))
- Calculate the standard deviation across all test samples and divide each entry with this standard deviation.

An example data file is provided [here](#).

**Step 2:** Choosing parameters

- a. Number of repeats – Using 1 repeat will generate only one output model. Using many repeats, DCQ calculates an ensemble of models, and outputs the average and standard deviation for each predicted relative cell quantity.
  
- b. Lambda min – A parameter of the L1 and L2 regularization. It is generally recommended to leave the default value. For more information about this parameter, see the [glmnet](#) package.

**Step 3:** Run the analysis and looking at the output files.

If you ran the analysis with only 1 repeat, the output will be a single file containing the relative quantity prediction for each cell type on every test sample.

If you ran the analysis with multiple repeats, the output will contain two file:

1. ###\_outMean.txt - The average relative quantities for each cell type in every test sample.
2. ###\_outSD.txt - The standard deviations over all repeats for each cell types in each test sample.

Each of the output files is a tab-delimited matrix, where columns are samples and rows are the various cell types. It is recommended to use [GENE\\_E](#) software in order to visualize the results.

\*\* If one or more of the markers used in the analysis are missing from the data, an additional output file by the name “missing\_markers.txt” will be provided. This file contains the names of the markers that were missing.