

# RegStatGel: Proteomic software for identifying differentially expressed proteins based on 2D gel images

Feng Li<sup>1,2\*</sup>, Françoise Seillier-Moiseiwitsch<sup>3</sup>

<sup>1</sup>Department of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, Maryland, USA; <sup>2</sup>Division of Biometrics II, Office of Biostatistics, Center for Drug Evaluation and Research, Food and Drug Administration, 10903 New Hampshire Ave, Silver Spring, MD, 20993; <sup>3</sup>Infectious Disease Clinical Research Program, Department of Preventive Medicine and Biometrics, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Road, Bethesda, MD, 20814; Feng Li - Email: feng.li@fda.hhs.gov; Phone: 3017964974; \*Corresponding author

Received June 10, 2011; Accepted June 13, 2011; Published August 02, 2011

## Abstract:

Image analysis of two-dimensional gel electrophoresis is a key step in proteomic workflow for identifying proteins that change under different experimental conditions. Since there are usually large amount of proteins and variations shown in the gel images, the use of software for analysis of 2D gel images is inevitable. We developed open-source software with graphical user interface for differential analysis of 2D gel images. The user-friendly software, RegStatGel, contains fully automated as well as interactive procedures. It was developed and has been tested under Matlab 7.01.

**Keywords:** 2D gel, 2D PAGE, proteomic software, image analysis, differential analysis, statistical analysis

**Availability:** The software is freely available at <http://www.mediafire.com/FengLi/2DGelsoftware> for academic use.

## Background:

Two-dimensional gel electrophoresis (2D gel) remains an important tool in proteomics for protein separation and quantification. The discovery of interesting proteins relies on the analysis of 2D gel images [1]. However, there are large amount of variations contained in the gel images, which should be appropriately accounted for using statistical methods. Unlike analysis of microarray images, there are limited published research and freely available software packages on statistical differential analysis of 2D gel images. The main challenges are the discrimination between actual protein spots and noise, the quantification of protein spots thereafter, and spot matching for individual comparison [2, 3]. Although there are commercial software packages for 2D gel image analysis (e.g. PDQuest, Dymension), considerable human intervention is needed for spot identification and matching. Moreover, the comparison of the quantitative features is either based on simple t-test or relies on external statistical tools for analysis. We developed open-source software with graphical user interface, RegStatGel, which is fast, fully automated and robust, with an emphasis on identifying differentially expressed proteins instead of striving for accurate quantification. Moreover, the RegStatGel software incorporates more advanced statistical tools for identifying differentially expressed proteins. It implements a novel analysis procedure as described elsewhere [4].

## Methodology:

The software, RegStatGel, bypasses the spot-matching bottleneck by detecting spots using the mean gel image and comparing protein features within watershed boundaries. RegStatGel uses the watershed algorithm to

automatically define the region surrounding a spot. The analysis procedure is described in detail [4]. It contains fully automated as well as interactive analysis procedures. Snapshots of the main interface of RegStatGel and of the sub-items of selected menus are shown in **Figure 1**. RegStatGel is implemented in Matlab 7.01. There are three modes of operation within RegStatGel: Fully automated mode: the user loads the images and then clicks on the menu to choose the fully automated analysis under the 'Gel Analysis' menu. Interactive automated mode: the user will be automatically prompted to choose the parameters for some procedures or skipping some procedures for exploratory purpose. The interactive mode can be accessed via the submenu under 'Gel Analysis'. Exploratory mode: the user can explore the protein regions sequentially by clicking the 'next' or 'previous' buttons. The user can also use the slider to choose the size of the image section that needs closer inspection. The fully automated analysis procedure implemented in RegStatGel contains the following key steps: (1) smoothing and rescaling of gel images; (2) construction of the master watershed map using the mean image; (3) segmentation of watershed regions; (4) quantification of watershed regions; (5) separation of protein regions into independent sets of correlated proteins; (6) statistical analysis using one way or two-way ANOVA or MANOVA; (7) selection of interesting proteins regions while controlling the false discovery rate at selected level.

## Software Input:

RegStatGel can load gel images with various formats. At first run, the user will be asked some information about the experimental design such as how many experimental groups and how many images within each group. The software

then loads images by groups. The user can load files saved during a previous run. The software features default settings for parameters. The user can easily edit the default parameters in self-evident menus.

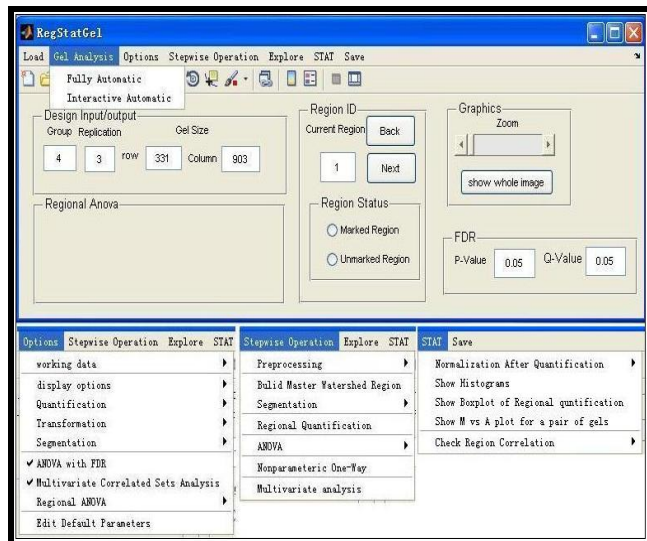


Figure 1: Snapshots of RegStatGel.

## Software Output:

During the automated mode, the software will display the status of each key step. At the end of the automated analysis, RegStatGel shows an image with differentially expressed proteins highlighted. The software provides menus to save results from different steps, such as the quantification of the proteins and the list of proteins detected. Under the exploration mode, the user can choose to investigate the detected spots individually by looking at their 3D shape.

## Caveat & Future Development:

It should be noted that RegStatGel focuses on post-alignment analysis so there is no image alignment function. The software is not sensitive to misalignment as long as it is slight. The next version will have additional options for statistical analysis.

## Acknowledgement:

Françoise Seillier-Moiseiwitsch was supported in part by NIH grant 5R01GM075298. This article reflects the views of Feng Li and should not be construed to represent FDA's view or policies.

## References:

- [1] O'Farrell PH. *J Biol Chem.* 1975 **250**: 4007 [PMID: 236308]
- [2] Kaczmarek K *et al. Proteomics.* 2004 **4**: 2377 [PMID: 15274133]
- [3] Li F & Seillier-Moiseiwitsch F. *Methods Enzymol.* 2011 **487**: 595 [PMID: 21187240]
- [4] <http://dx.doi.org/10.1016/j.ccsda.2011.05.013>

Edited by P Kanguane

Citation: Li & Seillier-Moiseiwitsch. *Bioinformatics* 6(10): 389-390 (2011)

**License statement:** This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.