How to Predict Discovery Potential and False Positives in 2-Dimensional Electrophoresis Image Analyses

Ola Forsström-Olsson[1], Johan Ljunggren[1], Mattias Nilsson[1], Haris Alisic[1], Anna Kapferer[1], Andreas Hallberg[1], Andreas Ekefjärd[1], Steven T Elliott[2], Jennifer Van Eyk[2]

[1]Ludesi, Lund, Sweden; [2]Johns Hopkins University, Baltimore, Maryland, USA

Introduction

The desired outcome of most 2D gel image analysis projects is to find all the true positives with a minimal false positives rate.

In this study, we measured the correctness of different aspects of the image analysis in order to investigate if there is a way to predict the discovery potential and the false positives rate of the 2D gel image analysis results.

We had access to 30 different 2D gel electrophoresis projects comprising close to 500 analyzed gel images in total. All projects had been previously analyzed by their respective proteomics labs using conventional image analysis software such as PDQuest, DeCyder, Progenesis, Progenesis with SameSpots, and Image Master. All projects were also analyzed by the Ludesi Image Analysis Center.

A novel method of evaluating the correctness of different aspects of the analyses was devised. In addition, we also calculated the mean coefficient of variation (mean CV) of the spot volumes in all the analyses.

Methods

30 analyses, comprising a total of 474 gel images analyzed. Analyses using: The Ludesi Analysis Center, Progenesis, PDQuest, Decyder Progenesis, Progenesis SameSpots, and Image Master.

Evaluation of spots

A method was devised to count spots that are:

Correct	Belonging to a single, real protein spot with a proper border containing the entire spot but nothing else such as other spots or artifacts
False	Not belonging to a protein spot at all
Misshaped	Belonging to a real protein spot but with a faulty border, leading to a faulty measurement of volume
Missing	A real protein spot that has not been detected

Total Spot Correctness = Correct / (Correct + False + Misshaped + Missing)

For each of the 30 analysis 200 spots were evaluated and rated based on these criteria.

Evaluation of pair matches

Evaluate the matching, or non-matching, between a randomly selected spot in a randomly selected gel and the equivalent region in a different, randomly selected, gel (whether this region contains a spot matched to the first spot or not).

Correct
- 1st spot correctly matched to 2nd spot
- *Or* 1st spot correctly not matched to 2nd gel (meaning that no protein spot corresponding to the first spot can be seen in the image).

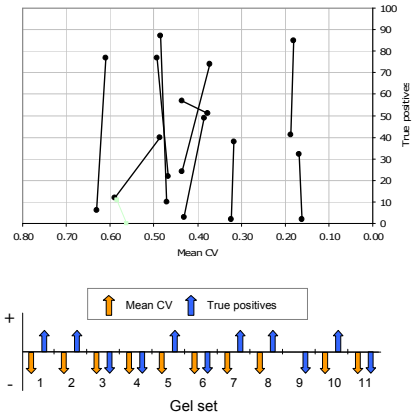
Incorrect
- 1st spot matched to the wrong spot in 2nd gel
- *Or* 1st spot incorrectly unmatched to 2nd gel (because there exists a spot in the 2nd gel that it should have been matched to).

Pair matching correctness = Correct / (Correct + Incorrect)

For each of the 30 analysis 200 spots were evaluated and rated based on these criteria..

Does the Mean CV relate to the rate of true positives and false positives?

A lower Mean-CV does not mean more true positives



Evaluation of true/false positives

True: The same protein spot has been matched in all gels where it exists. All the matched spots are correctly detected and segmented.

False: One or more errors in matching and/or spot detection/segmentation has occurred.

Evaluation of Mean CV's correlation to true/false positives

For each project, the Mean CV and the number of true positives were plotted from the two analyses (left on top)

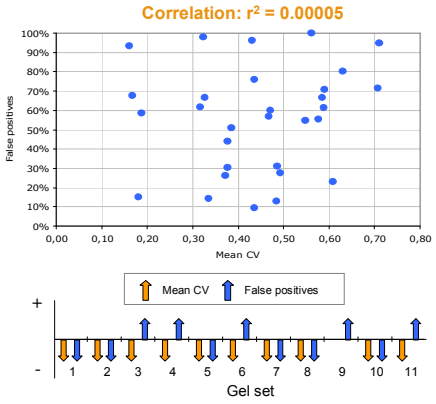
For each of the 30 analyses, the false positive rate and the Mean CV were plotted (right on top)

For each project, a trend plot was made, showing how a decrease in Mean CV affects the number of true positives (left, below) and the false positive rate (right, below).

The results clearly show that the Mean CV does not correlate very well with neither the number of true positives found in the projects, nor with the false positives rate in the projects.

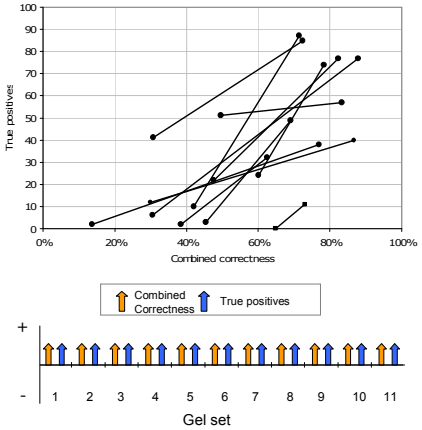
No The r^2 correlation coefficient for Mean CV's correlation to false positive rate was only **0.00005.**

A lower Mean CV does not mean a lower false positives rate



A more reliable measurement for true positives and false positive rate: "Combined Correctness"

Higher Combined Correctness always means more true positives



Evaluation of Combined Corectness' correlation to true resp false positives

Combined correctness = [Spot correctness] x [Pair matching correctness]

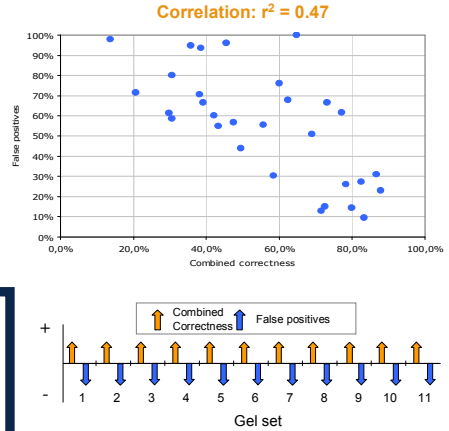
For each project, the Combined Correctness and the number of true positives were plotted from the two analyses (left on top)

For each of the 30 analyses, the false positive rate and the Combined Correctnes were plotted (right on top)

For each project, a trend plot was made, showing how an increase in Combined Correctness affects the number of true positives (left, below) and the false positive rate (right, below).

The r^2 correlation coefficient for Combined Correctness' correlation to false positives rate was **0.47,** considerably higher than for the Mean-CV.

Higher Combined Correctness always means a lower false positive rate



Conclusions

Mean CV does not seem to be reliably correlated to neither the number of true positives nor to the false positives rate in a 2D gel image analysis project.

The **Combined Correctness** is however highly correlated to the discovery potential of 2D gel image analysis:

- When the Combined Correctness is improved, the number of true positives goes up.
- When the Combined Correctness is improved, the false discovery rate goes down.