

Gene expression

## ***Pinnacle*: a fast, automatic and accurate method for detecting and quantifying protein spots in 2-dimensional gel electrophoresis data**

Jeffrey S. Morris<sup>1,\*</sup>, Brittan N. Clark<sup>2</sup> and Howard B. Gutstein<sup>2</sup>

<sup>1</sup>Department of Biostatistics and <sup>2</sup>Department of Anesthesiology and Molecular Genetics, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd Unit 447, Houston, TX 77030-4009, USA

Received on April 30, 2007; revised on August 22, 2007; accepted on November 25, 2007

Advance Access publication January 14, 2008

Associate Editor: David Rocke

### ABSTRACT

**Motivation:** One of the key limitations for proteomic studies using 2-dimensional gel electrophoresis (2DE) is the lack of rapid, robust and reproducible methods for detecting, matching and quantifying protein spots. The most commonly used approaches involve first detecting spots and drawing spot boundaries on individual gels, then matching spots across gels and finally quantifying each spot by calculating normalized spot volumes. This approach is time consuming, error-prone and frequently requires extensive manual editing, which can unintentionally introduce bias into the results.

**Results:** We introduce a new method for spot detection and quantification called *Pinnacle* that is automatic, quick, sensitive and specific and yields spot quantifications that are reliable and precise. This method incorporates a spot definition that is based on simple, straightforward criteria rather than complex arbitrary definitions, and results in no missing data. Using dilution series for validation, we demonstrate *Pinnacle* outperformed two well-established 2DE analysis packages, proving to be more accurate and yielding smaller coefficient of variations (CVs). More accurate quantifications may lead to increased power for detecting differentially expressed spots, an idea supported by the results of our group comparison experiment. Our fast, automatic analysis method makes it feasible to conduct very large 2DE-based proteomic studies that are adequately powered to find important protein expression differences.

**Availability:** Matlab code to implement *Pinnacle* is available from the authors upon request for non-commercial use.

**Contact:** jefmorris@mdanderson.org

**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

### 1 INTRODUCTION

Proteomics is capable of generating new hypotheses about the mechanisms underlying physiological changes. The perceived advantage of proteomics over gene-based global profiling approaches is that proteins are the most common effector molecules in cells. Changes in gene expression may not be reflected by changes in protein expression (Anderson and Seilhammer, 1997; Gygi *et al.*, 1999). However, the large

number of amino acids and post-translational modifications make the complexity inherent in analyzing proteomics data greater than for genomics data.

Several methods have been developed for separating proteins extracted from cells for identification and analysis of differential expression. One of the oldest yet still most widely used is 2-dimensional gel electrophoresis (2DE, Klose, 1975; O'Farrell, 1975). In this method, proteins are first separated in one direction by their isoelectric points, and then in a perpendicular direction by molecular weight. As 2DE-based proteomic studies have become larger and more complex, one of the major challenges has been to develop efficient and effective methods for detecting, matching and quantifying spots on large numbers of gel images. These steps extract the rich information contained in the gels, so are crucial to perform accurately if one is to make valid discoveries.

In current practice, the most commonly used spot detection and quantification approach involves three steps. First, a spot detection method is applied to each individual gel to find all protein spots and draw their boundaries. Second, spots detected on individual gels are matched to a master list of spots on a chosen *reference gel*, requiring specification of vertical and horizontal tolerances since spots on different gels are rarely perfectly aligned with one another. Third, 'volumes' are computed for each spot on each gel by summing all pixel values within the defined spot regions.

Unfortunately, methods based on this approach lack robustness. Errors are frequent and especially problematic for studies involving large numbers of gels. The errors consist of three main types, spot detection, spot boundary estimation and spot matching errors. Detection errors include merging two spots into one, splitting a single spot into two, not detecting a spot and mistaking artifacts for spots. Also, automatically detected spot boundaries can be inaccurate, increasing the variability of spot volume calculations. Matching errors occur when spots on different gels are matched together but do not correspond to the same protein. In our experience, these errors are pervasive and can obscure the discovery of differential protein expression. Almeida *et al.* (2005) list mismatched spots as one of the major sources of variability in 2DE, and Cutler *et al.* (2003) identify the subjective nature of the editing

\*To whom correspondence should be addressed.

required to correct these errors as a major problem. Extensive hand editing is needed to correct these various errors and can be very time-consuming, taking 1–4 h per gel (Cutler *et al.*, 2003). Taken together, these factors limit throughput and bring the objectivity and reproducibility of results into question. Also, one must decide what to do about missing values caused by spots that are matched across some, but not all gels. A number of *ad hoc* strategies have been employed, but all have their weaknesses and result in biased quantifications.

In this article, we introduce a new method for spot detection and quantification for 2DE analysis, which we call *Pinnacle*. This method takes a different fundamental approach than the most commonly used methods, using a mean gel for spot detection and using pinnacles instead of volumes for spot quantification. As a result of these differences, *Pinnacle* is much simpler and quicker than existing alternatives, and it results in no missing data, more sensitive and specific spot detection, and as we demonstrate in validation studies, spot quantifications that are more accurate and precise. In Section 2, we describe and motivate the *Pinnacle* algorithm. In Section 3, we describe the validation and group comparison studies, providing details of the data sets used, the implementation details for the competing methods, and the statistical measures used or evaluation. Section 4 contains the results of the validation and group comparison studies, and Sections 5 and 6 contain a discussion of the benefits of using *Pinnacle* for spot detection and quantification and final conclusions.

## 2 METHODS

Here we describe in detail the *Pinnacle* method introduced in this article. This method assumes that gels have been scanned without pixel saturation and have been suitably aligned using appropriate image registration software. In our analyses here, we used the TT900 program (Nonlinear Dynamics), although any effective image registration program could be used. For optimal performance of this method, any remaining misalignment should be less than the minimum distance between the pinnacles of two adjacent protein spots. We have found no difficulty aligning the gels in this study, or other gel sets we have analyzed.

Working on the aligned gels, the *Pinnacle* method consists of the following steps:

- (1) Compute the average gel.
- (2) Denoise the average gel using wavelet shrinkage.
- (3) Detect *pinnacles* on the wavelet denoised average gel.
- (4) Combine any pinnacles within a specified proximity.
- (5) Quantify each spot for each gel by taking the maximum intensity within a specified neighborhood of the pinnacle in the average gel.
- (6) Apply background correction filters and normalize the spot quantifications.

We next discuss each of these steps in more detail.

A key novel feature of this approach is that the average gel is used for pinnacle detection. We construct the average gel by averaging the intensities pixel-by-pixel across all gels in the experiment. Note that this ‘average gel’ differs from the composite gels constructed by PDQuest, Progenesis and other commercial software that are representations of the spots detected on all of the gels rather than simple

pixel-wise averages. It is unnecessary to do any background correction before computing the average gel.

In step 2, we apply wavelet-based denoising filters to denoise the average gel. Over the past ten years, wavelet de-noising has become a standard method for removing white noise from signals and images. On these gels, wavelet de-noising ‘smoothes out’ small irregularities in the average gel that are consistent with white noise while retaining the larger signals produced by true protein spots. Removal of these irregularities reduces the number of false positive spots detected.

To denoise, we used the undecimated discrete wavelet transform (UDWT), as implemented in version 2.4 of the Rice Wavelet Toolbox (RWT), which is freely available from their web site (<http://www.dsp.rice.edu/software/rwt.shtml>). The wavelet de-noising consists of the following three steps. First, given a particular choice of wavelet basis, wavelet coefficients are computed for the average gel. These coefficients represent a frequency-location decomposition in both dimensions of the image. The advantage of using the UDWT over the more computationally efficient and commonly used dyadic wavelet transform (DDWT) is that the results are translation-invariant, meaning that the de-noising is the same even if you shift or crop the image in either dimension, which results in more effective de-noising. We have found the results to be minimally sensitive to choice of wavelet basis; by default we use the Daubechies wavelet with four vanishing moments.

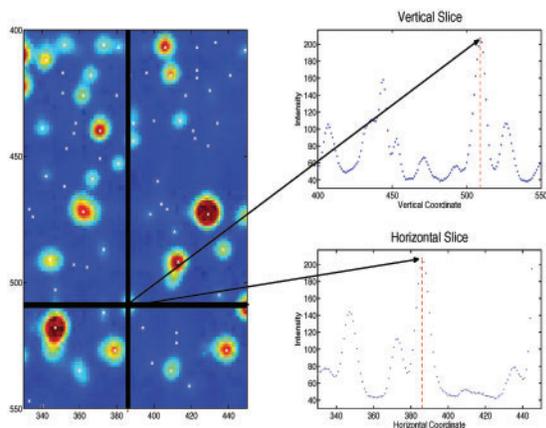
Second, hard thresholding is applied to the wavelet coefficients. By hard thresholding, we set all coefficients below a threshold  $\phi = \delta\sigma$  to 0, while leaving all coefficients  $\geq \phi$  unaffected. The parameter  $\sigma$  represents a robust estimator of the SD, following Donoho and Johnstone (1994) by using the median absolute deviation for the highest frequency wavelet coefficients divided by 0.6745, and  $\delta$  is a threshold parameter specified by the user, with larger choices of this parameter result in more de-noising. In the context of MALDI-MS, values of  $\delta$  between 5 and 20 were found to work well (Coombes *et al.*, 2005). For 2D gels, we have found that the background white noise is not as strong as MALDI-MS, so smaller values work better. Our default value is  $\delta = 2$ .

Third, the denoised signal is reconstructed by applying the inverse UDWT to the threshold wavelet coefficients. The thresholding works because white noise is equally distributed among all wavelet coefficients, while the signal is focused on a small number of coefficients. Thus, the thresholding zeroes out the large number of wavelet coefficients of small magnitude corresponding mostly to noise, while leaving the small number of coefficients of large magnitude corresponding to signal.

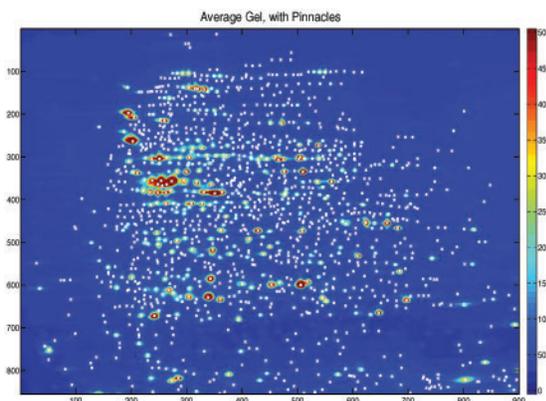
After de-noising, we next perform spot detection on the wavelet denoised average gel by detecting all *pinnacles*. We determine that a pixel location contains a pinnacle if it is a local maximum in both the horizontal and vertical directions on the gel (Fig. 1), and if its intensity was greater than some threshold, by default the 75th percentile on the gel. This leaves us with a list of pixel coordinates marking the pinnacles in the average gel that index the ‘spots’ of interest in the given gel set. Figure 2 shows the 1403 detected pinnacles on the average gel from one of our studies.

If any pinnacles are found within a given  $2k_1 + 1 \times 2k_1 + 1$  square surrounding another pinnacle, then in step 4 these pinnacles are combined by keeping only the one with the highest intensity. This step removes spurious double peaks, and accommodates imperfect alignment, as described in the next step. In our experience, it is rare to see two protein spots with pinnacles  $\leq 5$  units from each other, given the resolution of our scanner, which yields a  $1024 \times 1024$  image of the gel, so by default we use  $k_1 = 2$ . Thus, we have found that we do not lose spots by this step.

In step 5, we quantify each spot for each individual gel by taking the maximum intensity within the  $2k_2 + 1 \times 2k_2 + 1$  square formed by taking the corresponding pinnacle location in the average gel and extending out  $\pm k_2$  units in the horizontal and vertical directions on the



**Fig. 1.** Demonstration of pinnacles on an average gel: the left plot contains a zoomed-in region of a denoised average gel, with detected pinnacles marked as white 'x's, with the focus on the spot detected with pinnacle at (509, 386). The units of the  $x$  and  $y$  axes are pixel distance from the origin (upper left corner of the image). The upper right plot contains a plot of the vertical slice at 386, and the lower right plot contains a plot of the horizontal slice at 509. This location was flagged as a pinnacle because it was a local maximum in both the horizontal and vertical slices, and had an intensity of  $\geq 120.86$ , the 75th percentile on the average gel.



**Fig. 2.** Average gel with detected pinnacles, Nishihara and Champion dilution series: The average gel was created by taking the pixel-wise average over the 28 gels in that series. 'Hotter' colors indicate regions of higher intensity, while 'cooler' colors indicate lower intensities. Intensities above 500 were censored to improve contrast. The units of the  $x$  and  $y$  axes are pixel distance from the origin (upper left corner of the image). White 'x's' mark the 1403 pinnacles detected using Pinnacle, which represent local maxima in both the  $x$ - and  $y$ -directions with intensities  $\geq 120.86$ , the 75th percentile intensity on the average gel.

individual gel. The width  $k_2$  should be at least as small as the proximity  $k_1$  in step 5; by default we use  $k_2 = k_1$ . This tolerance enabled us to find the maximum pinnacle intensity for the corresponding spot for each individual gel even when the alignment was not perfect. The accuracy of the alignment only needed to be within  $\pm k_2$  pixels in both the horizontal and vertical directions.

In the final step, we perform background correction and normalization on the quantifications. If the background appears relatively

uniform, we have found subtracting global minimum intensity for the gel works sufficiently well. Whenever the background appears to be spatially varying, we use a windowed minimum to estimate the background. Since we are using pinnacle intensities rather than spot volumes for quantifications, the background only needs to be estimated for the pixel locations containing pinnacles, so its calculation proceeds very quickly. Our default window is  $\pm 100$  pixels in the horizontal and vertical directions. One must ensure that the window is large enough to extend beyond each spot region to avoid attenuation of the quantified pinnacle intensities.

To normalize, we divide each pinnacle intensity on a given gel by the mean pinnacle intensity for that gel. We also note that it is possible to apply a wavelet-based de-noising to the individual gels before quantification. While conceptually appealing, we have found this to make little difference in practice, so by default we do not denoise the individual gels.

Given  $N$  individual gels and  $p$  spots, after this step we are left with an  $N \times p$  matrix of protein expression levels with no missing values. In profiling or group comparison studies, this matrix would be analyzed to find which of the  $p$  spots appear to be associated with factors of interest, and worthy of future study.

### 3 VALIDATION STUDIES

We compared the performance of Pinnacle with current versions of the commercial software packages Progenesis and PDQuest in detecting, matching and quantifying protein spots using two dilution series, and we compared their performance in differential expression using a group comparison study. The first dilution series was created by Nishihara and Champion (2002), and we prepared the second dilution series and the group comparison data in house. For the dilution series, the percentage of spots correctly matched across gels by the automatic algorithms was summarized. Reliability of spot quantifications was assessed by measuring the strength of linear association ( $R^2$ ) between the spot quantifications and the protein loads in the dilution series for each detected spot. Given the nature of the dilution series, methods yielding more accurate protein quantifications should result in  $R^2$  closer to 1. We assessed precision using the coefficient of variation (%CV) of spots within the different dilution groups. For the group comparison, we summarized the number and proportion of spots with differential expression  $P$ -values and local false discovery rates below pre-specified thresholds. All comparisons were based on results generated solely by the three algorithms, without any subsequent editing. In the remainder of this section, we provide detailed descriptions of the data sets, the competing algorithms, and the statistical measures used to compare the methods.

#### 3.1 Description of data sets

**3.1.1 Nishihara and Champion dilution series** Nishihara and Champion (2002) prepared a dilution series experiment using a sample of *E. coli* with seven different 2D gel protein loads spanning a 100-fold range (0.5, 7.5, 10, 15, 30, 40 and 50  $\mu\text{g}$ ). Four gels were run at each protein load. Details of the conduct of the 2DE are described in Champion *et al.* (2001), and the details of the staining and image capture procedures are described in Nishihara and Champion (2002). The images were provided to us courtesy of Dr Kathleen Champion-Francissen, and were

used to compare Pinnacle with PDQuest and Progenesis. Nishihara and Champion previously used this series to evaluate the performance of several 2D analysis packages by analyzing 20 corresponding spots from all the gels. By only investigating 20 spots, however, they did not gain an accurate picture of the methods' performance in detecting, matching and quantifying spots across the entire gel. Therefore, we evaluated analysis methods using all spots detected in this dilution series. Our Progenesis and PDQuest results for the 20 selected spots were comparable to the results previously obtained by Nishihara and Champion (2002) (data not shown).

**3.1.2 SH-SY5Y neuroblastoma cell dilution series** SH-SY5Y cells were grown to 60–70% confluence and then harvested. Cells were then resuspended using the ProteoPrep™ (Sigma) total extraction kit and the suspension ultrasonicated on ice for 15 s. bursts at 70% amplitude for a total time of 1 min. After sonication, the suspension was centrifuged at 15000g for 30 min. at 15°C. The samples were then reduced for 1 h at RT by adding tributylphosphine to a final concentration of 5 mM and alkylated in the dark for 1.5 h at RT by adding iodoacetamide to a final concentration of 15 mM. 11 cm IPG strips (Bio-Rad) were then rehydrated in 100 µl of sample buffer for 2 h at RT. Protein samples were then applied to the strips in 150 µl of buffer and IPGs were then focused for 100 kVh. Three replicate gels were run for each of six different protein loads (5, 10, 25, 50, 100, 150 µg). Voltage was increased from 0 to 3000 V over 5 h (slow ramp), 3–10 000 V over 3 h (linear ramp), followed by additional hours at 10 000 V. IPGs were then equilibrated in SDS-equilibration buffer containing 3 M urea, 2.5% (w/v) SDS, 50 mM Tris/acetate buffer (pH 7.0) and 0.01% (w/v) bromophenol blue as a tracking dye for 20 min. The equilibrated strips were then placed on 8–16% polyacrylamide gels (Bio-Rad) and proteins separated by size. Run conditions were 50 mA/gel until the bromophenol blue reached the end of the gel. Proteins were visualized using SYPRO ruby stain (Bio-Rad). Gels were fixed for 30 min. in a solution containing 10% methanol and 7% acetic acid. After fixation, gels were stained in 50 ml of SYPRO ruby overnight in the dark. The gels were next de-stained in 10% methanol and 7% acetic acid for 2 h, and then imaged using a Kodak Image Station 2000R. Gel images were subsequently cropped to exclude edge artifacts and streaks. The same cropped image area was used for all analytical protocols.

**3.1.3 Morphine group comparison data set** After institutional IACUC approval was obtained, 6 adult male Sprague-Dawley rats were implanted with either morphine 75 mg slow release pellets (National Institute on Drug Abuse) or placebo pellets subcutaneously under isoflurane anesthesia. Tolerance development was monitored daily by tail flick latency (Xu *et al.*, 2006). After 5 days, animals were sacrificed and spinal cords harvested. The substantia gelatinosa region was then dissected using the transillumination method as previously described (Cuello *et al.*, 1983). Proteins were extracted from this region and 2D gels run as previously described (Moulédous *et al.*, 2005).

## 3.2 Implementation details for competing methods

All gels used in these studies were processed and analyzed using three different methods: Progenesis PG240 version 2006 (Nonlinear Dynamics Ltd., Newcastle-upon-Tyne, UK), PDQuest Version 8.0 (Bio-Rad Laboratories, Hercules, CA, USA), and the Pinnacle method described in this article. Pinnacle, as described in Section 2, was applied to images that were first aligned using the TT900 software program (Nonlinear Dynamics), and involved average gel computation, pinnacle detection and pinnacle-based quantification using computer code written in MATLAB (version R2006a, The MathWorks, Inc.) using Windows XP-based PCs, with default settings used. All procedures were performed in our laboratory. Specific analysis steps are detailed below. Both Progenesis and PDQuest are designed to be run on unaligned gels. In order to ensure that any differences between Pinnacle and these methods are not due solely to the alignment, we also applied these methods to the gel images after they were aligned using TT900.

**3.2.1 Progenesis** Gels were processed using the Analysis Wizard, which is a stepwise approach for selecting pre-processing options. Gels were grouped by protein load, and the same gel was selected as the top reference gel for both the aligned and unaligned image sets. Background subtraction was performed using the Progenesis Background method. Combined warping and matching was selected for the unaligned images, and property-based matching was selected for the aligned images, as recommended by the manufacturer. Normalization was not done, since this would eliminate the linearity of quantifications with protein load that we use to evaluate reliability. The minimum spot area was set to one, and the split factor set at nine. These settings produced a similar number of spots to that reported by Nishihara and Champion using an earlier version of this program (Nishihara and Champion, 2002). No manual editing of the data was performed. The data were simply exported to Excel and spots present in 3 of 4 replicates in the Nishihara and Champion series, or 3 of 3 replicates in the SH-SY5Y dilution series determined. Spot volumes of zero were used for spots present on other gels with no match on the current gel.

**3.2.2 PDQuest** Gels were processed using the Spot Detection Wizard. Gels were grouped by protein load, default background subtraction and default match settings were applied, and the same master gel was selected for both aligned and unaligned images. This was the same gel used as the top reference gel in Progenesis. We used the 'Give Manual Guidance' and 'Test Settings' features of the Advanced Spot Detection Wizard. We also used the speckle filter and the vertical and horizontal streak filter in the Advanced Controls, as recommended by the manufacturer. Using these settings we obtained a similar number of spots to that reported by Nishihara and Champion using an earlier version of this program (Nishihara and Champion, 2002). Again, we did not normalize spot volumes. No manual editing of the data was performed. The data were exported to Excel and spots present in 3 of 4 replicates in the Nishihara and Champion series, or 3 of 3 replicates in the SH-SY5Y dilution series determined.

Spot volumes of zero were used for spots present on other gels with no match on the current gel.

### 3.3 Statistical criteria used for validation

For each dilution series and method, we summarized the total number of detected spots. For Progenesis and PDQuest results, we computed the number of these that were ‘unmatched’, meaning that they were present on only one gel and not matched to any spot on any other gel. Pinnacle had no unmatched spots since by definition it yielded quantifications for every pinnacle on each gel. We removed all unmatched spots in Progenesis and PDQuest from consideration in the quantitative summaries. We also removed any spots that were not present in at least 3 out of 4 replicate gels for at least one of the protein load groups in the Nishihara and Champion study or 3 out of 3 replicates in the SH-SY5Y cell dilution series. This criterion was used by Nishihara and Champion (2002), and is commonly used by other investigators.

We used the results of the dilution series experiments to assess the matching percentage, reliability and precision of the different methods’ quantifications. The matching percentage for Pinnacle applied to aligned gels was 100%. For PDQuest and Progenesis, we estimated the matching percentage by randomly selecting 10% of the total number of spots that met the above criteria, and then checking by hand the number of times the automatic algorithms correctly matched the corresponding spot on all individual gels for which it was detected to the spot on the reference gel. Note that this measure only deals with matching errors, not detection errors, since gels for which a given spot was not detected at all did not count as a mismatch in terms of the match percentage. Also, incorrect spot splitting (e.g. matching a spot in one gel to the same spot and an adjacent one which were detected as one spot in another gel) was not considered a mismatch in this analysis.

The reliability of quantification for each spot was assessed by computing the coefficient of determination ( $R^2$ ) from a simple linear regression (implemented in Matlab, Mathworks, Inc.) of the mean spot quantification across replicates for each protein load group versus the true protein load. If the correlation ( $R$ ) was negative, then we set  $R^2 = 0$ . The idea driving this analysis was that if the gel ran properly and the quantification method used was robust, then the ratio of quantifications for a given spot for any two gels should be proportional to the ratios of the protein loads on those gels. We computed this measure for all detected spots, not just a select set, so we would get a realistic assessment of the performance of each method across the entire gel. We summarized the  $R^2$  across all spots within a gel by the mean, five-number summary (5th percentile, Q05, 25th percentile, Q25, the median, Q50, the 75th percentile, Q75 and the 95th percentile, Q95), and by counting the number of ‘reliable spots.’ Spots were considered reliable if  $R^2 > 0.90$ , which roughly corresponds to a correlation of at least 0.95 between the group mean spot quantifications and the protein load. The number of ‘reliable spots’ gave us a sense of the number of spots that were well quantified by a given method.

We assessed the precision of the quantifications by computing the coefficient of variation (%CV) for each spot detected in the entire gel set across the gels within each protein load group.

In the main text, we present the results from the 30  $\mu\text{g}$  protein load group for the Nishihara and Champion dilution series (as they did in their paper), and in the 50  $\mu\text{g}$  group for the SH-SY5Y dilution series; other results are available in Supplementary Tables. We summarized the %CV across all spots by the mean and 5 number summary (Q05, Q25, Q50, Q75, Q95), and we counted the number of detected spots with %CV < 20. Note that it was not possible to compute CVs for spots with group mean quantifications of zero, so those spots were left out of this analysis.

For the group comparison data, for each method we performed two-sample  $t$ -tests with unequal variance assumptions for each detected spot, and summarized the number and proportion of  $P$ -values  $\leq 0.001$ , 0.005, 0.01 and 0.05. We also summarized the number of spots with  $q$ -values < 0.10. A  $q$ -value is a measure of local false discovery rate, and estimates the probability that a given spot is a false positive if called significant (Storey, 2003).

## 4 RESULTS

For the Nishihara and Champion study (NH), Pinnacle detected 1403 spots (Fig. 2), which by definition were found and quantified for all gels. PDQuest detected 2692 spots of which 745 were ‘unmatched spots’ found on only one gel. An additional 571 spots were detected on more than one gel, but not included in the analyses because they were not found on 3 out of 4 gels for at least one group, the same exclusion criterion used by Nishihara and Champion (2002). The match percentage of the 1376 spots found on 3/4 gels in at least one group was 60%. Progenesis detected 1986 unique spots, of which 990 were unmatched and 121 not found on 3/4 gels in at least one group. The match percentage of the 875 spots found on 3/4 gels for at least one group was 84%. If we restricted attention only to those spots that had no missing values on any gel, as we did for Pinnacle, we would have been left with only 377 and 271 spots for PDQuest and Progenesis, respectively. These summaries are shown in Supplementary Table 1.

The top half of Table 1 contains reliability results for the NH study. Pinnacle yielded more reliable spot quantifications over this dilution series (mean  $R^2 = 0.924$ ) than either PDQuest (0.835) or Progenesis (0.883). Pinnacle found many more reliable spots (defined as  $R^2 > 0.90$ ) than PDQuest or Progenesis (1203 versus 847 or 666, respectively). Table 2 shows that Pinnacle also generated more consistent quantifications within the 30  $\mu\text{g}$  protein load group. Pinnacle generated a lower CV (mean 18.4) than either PDQuest (54.7) or Progenesis (40.3), and found far more spots with CV < 20% (983 versus 498 and 304, respectively). The results were similar for the other protein loads (Supplementary Tables 2–4).

To determine whether Pinnacle performed better than the other methods only because the gels were pre-aligned, we also ran PDQuest and Progenesis on the set of aligned gels. In general, we found that the alignment tended to slightly improve the reliability, but not to the levels of Pinnacle (Table 1). Alignment had inconsistent effects on match percentage, and decreased measurement precision for both PDQuest and Progenesis (Table 2).

**Table 1.** Reliability of quantifications for detected spots: summary of  $R^2$  measuring linearity of quantification method across protein loads within dilution series for all spots automatically detected by Pinnacle (Pinn) and for spots meeting the selection criteria below for PDQuest (PDQ) and Progenesis (Prog)

Study	Method	Match (%)	Number of spots <sup>a</sup>	Reliable spots ( $R^2 > 0.90$ )	$R^2$					
					Mean	Q05	Q25	Median	Q75	Q95
N+C	Pinn	100	1403	1203	0.924	0.613	0.951	0.978	0.990	0.995
	PDQ	60	1376	847	0.835	0.206	0.790	0.941	0.984	0.995
	Prog	84	875	666	0.883	0.253	0.909	0.973	0.990	0.996
	PDQ-a	71	1342	869	0.850	0.322	0.829	0.944	0.982	0.994
	Prog-a	80	887	713	0.894	0.446	0.919	0.963	0.981	0.990
SH-SY5Y	Pinn	100	1162	603	0.887	0.735	0.843	0.905	0.960	0.987
	PDQ	45	1297	406	0.735	0.141	0.637	0.820	0.923	0.982
	Prog	30	979	295	0.662	0.011	0.448	0.805	0.919	0.981
	PDQ-a	64	1103	391	0.753	0.137	0.669	0.841	0.931	0.980
	Prog-a	43	1092	384	0.698	0.014	0.570	0.845	0.932	0.982

<sup>a</sup>For Progenesis and PDQuest, the number of spots corresponds to the number of spots used in the quantitative analysis, which are those present in at least 3/4 replicate gels within at least 1 of the 7 protein load groups for the Nishihara and Champion study, and at least 3/3 replicate gels within at least 1 of the 6 protein load groups for the SH-SY5Y study.

Software settings are detailed in the Sections 2 and 3. No manual editing of the datasets was performed. PDQuest and Progenesis were run on the gel sets without alignment, since that is typically how these programs are used. We also ran these programs on the aligned gels (PDQ-a and Prog-a) to verify that the superior performance of Pinnacle was not solely due to image alignment. The summaries include match percentage, number of spots, number of reliable spots, and the mean, median, 5th, 25th, 75th and 95th percentile for the distribution of  $R^2$  across spots. A spot was considered reliable if  $R^2 > 0.90$ .

**Table 2.** Precision of quantifications for detected spots

Study	Method	Number of Spots <sup>a</sup>	Number of spots with %CV < 20	%CV					
				Mean	Q05	Q25	Median	Q75	Q95
N+C	Pinn	1403	983	18.4	7.6	12.7	16.5	21.2	31.6
	PDQ	1326	498	54.7	6.5	14.4	27.5	77.5	200.0
	Prog	821	304	40.3	11.1	16.9	23.6	41.3	121.8
	PDQ-a	1309	415	55.7	9.5	17.7	29.8	76.1	200.0
	Prog-a	852	144	47.4	14.4	22.6	30.5	50.3	145.5
SH-SY5Y	Pinn	1162	856	15.7	3.0	7.4	12.4	20.7	40.7
	PDQ	1166	267	64.4	8.0	21.4	45.5	91.1	173.2
	Prog	787	188	53.2	8.1	20.6	36.9	86.6	173.2
	PDQ-a	1001	272	58.8	7.0	19.0	38.0	88.1	173.2
	Prog-a	931	182	59.9	8.2	24.2	45.5	88.9	173.2

<sup>a</sup>The number of spots in this table corresponds to the number of spots from Table 1 for which there was at least one non-missing value for the 30 µg group in the Nishihara and Champion Study, and the 50 µg group in the SH-SY5Y study, since it is not possible to compute CVs for spots with group means of zero.

Summary of coefficient of variation (%CV) for the 30 µg protein load for the Nishihara and Champion study, and the 50 µg load for the SH-SY5Y cell study, for all spots automatically detected by Pinnacle (Pinn) and for spots meeting the selection criteria below for PDQuest (PDQ) and Progenesis (Prog). Software settings are detailed in the Methods. No manual editing of the datasets was performed. PDQuest and Progenesis were run on the gel sets before alignment, since that is typically how these programs are used. We also ran these programs on the aligned gels (PDQ-a and Prog-a) to verify that the superior performance of Pinnacle was not solely due to image alignment. The summaries include the number of spots, plus the mean, median, 5th, 25th, 75th and 95th percentile for the distribution of %CV across spots, and the number of spots with %CV < 20. The %CV results for the other protein load groups are presented in supplementary tables.

The last rows of Tables 1 and 2 contain the results from the dilution series created from SH-SY5Y neuroblastoma cell extracts. Pinnacle detected 1013 spots, while PDQuest identified 1297 spots that were found on 3/3 gels in at least one group, with a match percentage of 45%. Progenesis detected 979 spots on 3/3 gels in at least one group with a match

percentage of 30%. Pinnacle again yielded more reliable spot quantifications over this dilution series (mean  $R^2 = 0.887$ ) than either PDQuest (0.735) or Progenesis (0.662), and found many more reliable spots (603) than either PDQuest (406) or Progenesis (295). Again, Pinnacle generated more consistent measurements (mean CV in 50 µg load group 15.7) than either

**Table 3.** Comparison of methods for morphine group comparison dataset

	Method	Number of spots	$q\text{-val} < 0.10$	$P\text{-values}$			
				$<0.001$	$<0.005$	$<0.01$	$<0.05$
Morphine group comparison	Pinn	1687	308	6 (0.4%)	40 (2.4%)	80 (4.7%)	330 (19.5%)
	PDQ	992	2	3 (0.3%)	11 (1.1%)	22 (2.2%)	94 (9.5%)
	Prog	864	0	1 (0.1%)	9 (1.0%)	17 (2.0%)	84 (9.7%)
	PDQ-a	930	2	2 (0.2%)	13 (1.4%)	24 (2.6%)	89 (9.6%)
	Prog-a	972	32	8 (0.8%)	18 (2.1%)	31 (3.5%)	126 (14.4%)

Summary of number of detected spots, and number (%) of spots with  $P$ -values less than 0.001, 0.005, 0.01, and 0.05 for morphine group comparison data set, and with  $q$ -values less than 0.10, for the different preprocessing methods.

PDQuest (64.4) or Progenesis (53.2), and found far more spots with  $CV < 20\%$  (856 versus 267 and 188, respectively). Again, we found that alignment had inconsistent effects on the performance of PDQuest and Progenesis. Reliability and match percentage improved for both methods, but was still far inferior to Pinnacle. Precision improved for PDQuest but worsened for Progenesis.

Table 3 summarizes the results of the group comparison study. Using Pinnacle tended to result in more spots with small  $P$ -values. It found a greater number and proportion of spots with  $P$ -values  $\leq 0.001$ , 0.005, 0.01 and 0.05 than Progenesis without alignment or PDQuest with or without alignment. Progenesis with alignment found similar numbers and proportions of spots with  $P$ -values  $\leq 0.001$ , but considerably fewer spots with  $P$ -values  $< 0.005$ ,  $< 0.01$  or  $< 0.05$  than Pinnacle. After adjusting for multiplicities, Pinnacle found considerably more spots with  $q$ -values  $< 0.10$  than the other methods.

## 5 DISCUSSION

We have described and validated a new method for detecting and quantifying protein spots on 2DE gel sets. Designed for aligned gel images, it is automatic, fast and yields reliable results without any need for hand editing. Our results demonstrated that quantifications using Pinnacle are more reliable and precise than two currently popular analysis methods, yielding many more reliable spots, and having no missing data issue. It runs very quickly, taking just 56.6, 32.9 and 29.7 seconds, respectively, for the data sets considered in this article. Pinnacle is considerably simpler than methods like Progenesis and PDQuest, and this simplicity is the key not just to its speed but also its superior performance. There are several factors contributing to this effect.

First, image alignment is generally easier and more accurate than one-at-a-time spot matching across a gel series done after spot detection. Image alignment software efficiently uses information from nearby spots on the gel to guide the process. As shown by our validation studies, however, the improvement from Pinnacle is not solely from using the aligned gels images.

Second, as demonstrated in other contexts (Morris *et al.*, 2005), spot detection using the average gel is not just quicker, but should result in greater sensitivity and specificity compared

with spot detection on individual gels. This is because features corresponding to true protein spots will tend to be present on many gels and thus will be reinforced in the average gel, while artifacts and noise will tend to average out. The central limit theorem suggests that the noise level in the average gel will be less than the noise level in an individual gel by a factor of  $\sqrt{N}$ , and thus, it becomes easier to see the protein signal. Roughly speaking, the arithmetic average gel should have greater sensitivity for peak detection than individual gels for any proteins present in at least  $1/\sqrt{N}$  of the gels. By this principle, we should be able to more reliably detect fainter spots, thus improving the realized dynamic range of the 2D gel analysis. This also suggests that sensitivity, specificity and spot detection should improve, not deteriorate, as more gels are included. This is in marked contrast to standard methods that detect spots on individual gels, since in that setting, the occurrence of missing spots and matching errors tend to increase with larger gel sets.

Third, accurate pinnacle detection is aided by the wavelet denoising that adaptively removes noise without severely attenuating the true protein spots. In recent years, wavelet denoising has become a standard tool in nearly every area of signal processing, so is a natural tool to use in denoising 2DE images.

Fourth, the use of pinnacles instead of spot boundaries to define and quantify spots greatly reduces computational complexity, and decreases the variability of spot quantifications. Provided the gel images are not saturated, protein spots typically appear as mountain-like structures with well-defined pinnacles. It is quicker and easier to detect these pinnacles than to detect spots using more complex algorithms. Also, unlike spot boundaries, pinnacles are consistent and well defined even when spots overlap. The reduced variability comes from the fact it is not necessary to detect spot boundaries, a difficult and error-prone exercise.

It has long been assumed that spot volumes should correspond to true protein abundance, so we were initially surprised to find that the pinnacle-based method resulted in more reliable and precise quantifications than volume-based methods. However, as we demonstrate in Theorem 1 of the supplementary material, the pinnacle intensity should be strongly correlated with the spot volume when a given spot has a common shape across gels. Our empirical investigations

suggest that this assumption holds in practice for the vast majority spots on gels. Mahon and Dupree (2001) have similarly observed that pinnacle intensities are linearly related to spot volumes. Our studies indicate that our pinnacle-based method results in considerably smaller CVs than conventional spot volume-based analysis methods, which in profiling studies would result in greater statistical power for detecting differentially expressed proteins, as in our group comparison results. Also, Pinnacle's unambiguous spot definition on all gels results in no missing data, which is another factor in increasing quantitative precision.

We used dilution series experiments in this article because, unlike typical laboratory studies, these allow an investigation of both reliability and precision of quantifications. However, our method is intended for the standard laboratory setting, and we have applied it numerous times in that setting, and found it to perform very well. We have had no difficulty performing image alignment across gels from different individuals in laboratory studies of like tissue types. Also, note that the image alignment and use of the average gel do not require all gels to have the same proteins present. Since the Central Limit Theorem suggests the noise is reduced by a factor of  $\sqrt{N}$  in the average gel, sensitivity should be increased for spots present in at least  $1/\sqrt{N}$  of the gels. For example, in a study with 100 gels, use of the average gel should yield increased sensitivity for all protein spots present in at least 10% of the gels, which would likely include the proteins of interest in typical laboratory studies.

## 6 CONCLUSION

The lack of efficient, effective and reliable methods for 2D gel analysis has been a major factor limiting the contribution of 2DE to biomedical research. Currently, gel analysis is extremely time consuming and subjective, and it is difficult to conduct the larger studies required to have adequate statistical power for detecting proteins differentially expressed across experimental conditions. Ineffective pre-processing leads to reduced numbers of accurately detected and matched spots and unreliable, imprecise quantifications. This can cause investigators to miss potentially important discoveries that could be made from their data. The Pinnacle method is automatic, quick, robust, precise and without potential biases that could be introduced by manual editing. It tends to perform better, not worse, in larger studies, so is well-suited for the larger studies now being conducted. This simple, yet novel method has the potential to help maximize the impact of 2DE on biological research, and also has the potential to be applied to perform spot detection

and quantification in other settings where image data with spots are encountered, including DIGE and LC-MS.

## ACKNOWLEDGEMENTS

We thank Dr Kathy Champion-Francissen for the use of gel images from her 2002 study and Miguel Diaz for excellent technical assistance. This work was supported by grants from the National Cancer Institute (CA107304) to J.S.M. and from the National Institute on Drug Abuse (DA15146 and DA18310) and the National Institute on Alcohol Abuse and Alcoholism (AA13886) to H.B.G.

*Conflict of Interest:* none declared.

## REFERENCES

- Almeida, J.S. et al. (2005) Normalization and analysis of residual variation in two-dimensional gel electrophoresis for quantitative differential proteomics. *Proteomics*, **5**, 1242–1249.
- Anderson, L. and Seilhammer, J. (1997) A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis*, **18**, 533–537.
- Champion, K.M. et al. (2001) Similarity of the *Escherichia coli* proteome upon completion of different biopharmaceutical fermentation processes. *Proteomics*, **1**, 1133–1148.
- Coomes, K.R. et al. (2005) Improved peak detection and quantification of mass spectrometry data acquired from surface-enhanced laser desorption and ionization by denoising spectra with the undecimated discrete wavelet transform. *Proteomics*, **5**, 4107–4117.
- Cuello, A.C. and Carson, S. (2003) In Cuello, A.C. (ed.) *Brain Microdissection Techniques*. John Wiley and Sons, New York, pp. 37–116.
- Cutler, P. et al. (2003) A novel approach to spot detection for two-dimensional gel electrophoresis images using pixel value collection. *Proteomics*, **3**, 392–401.
- Donoho, D. and Johnstone, I.M. (1994) Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, **81**, 425–455.
- Gygi, S.P. et al. (1999) Correlation between protein and mRNA abundance in yeast. *Mol. Cell Biol.*, **19**, 1720–1730.
- Klose, J. (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues: a novel approach to testing for induced point mutations in mammals. *Humangenetik*, **26**, 231–243.
- Mahon, P. and Dupree, P. (2001) Quantitative and reproducible two-dimensional gel analysis using Phoretix 2D Full. *Electrophoresis*, **22**, 2075–2085.
- Morris, J.S. et al. (2005) Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum. *Bioinformatics*, **21**, 1764–1775.
- Nishihara, J.C. and Champion, K.M. (2002) Quantitative evaluation of proteins in one- and two-dimensional polyacrylamide gels using a fluorescent stain. *Electrophoresis*, **23**, 2203–2215.
- O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.*, **250**, 4007–4021.
- Storey, J.D. (2003) The positive false discovery rate: A Bayesian interpretation and the  $q$ -value. *Ann. Stat.*, **31**, 2013–2035.
- Xu, J.J. et al. (2006) Intermittent lumbar puncture in rats: a novel method for the experimental study of opioid tolerance. *Anesth. Analg.*, **103**, 714–720.