Gene Expression

Automated image alignment for 2-D gel electrophoresis in a high-throughput proteomics pipeline

Andrew W. Dowsey¹, Michael J. Dunn² and Guang-Zhong Yang¹,*
¹Institute of Biomedical Engineering, Imperial College London, United Kingdom.
²UCD Conway Institute of Biomolecular & Biomedical Research, University College Dublin, Ireland.

Associate Editor: Prof. David Rocke

ABSTRACT

Motivation: The quest for high-throughput proteomics has revealed a number of challenges in recent years. Whilst substantial improvements in automated protein separation with liquid chromatography and mass spectrometry (LC/MS), aka ‘shotgun’ proteomics, have been achieved, large-scale open initiatives such as the HUPO Brain Proteome Project have shown that maximal proteome coverage is only possible when LC/MS is complemented by 2-D gel electrophoresis (2-DE) studies. Moreover, both separation methods require automated alignment and differential analysis to relieve the bioinformatics bottleneck and so make high-throughput protein biomarker discovery a reality. The purpose of this paper is to describe a fully automatic image alignment framework for the integration of 2-DE into a high-throughput differential expression proteomics pipeline.

Results: The proposed method is based on robust automated image normalisation (RAIN) to circumvent the drawbacks of traditional approaches. These use symbolic representation at the very early stages of the analysis, which introduces persistent errors due to inaccuracies in modelling and alignment. In RAIN, a 3rd order volume-invariant B-spline model is incorporated into a multi-resolution schema to correct for geometric and expression inhomogeneity at multiple scales. The normalised images can then be compared directly in the image domain for quantitative differential analysis. Through evaluation against an existing state-of-the-art method on real and synthetically warped 2-D gels, the proposed analysis framework demonstrates substantial improvements in matching accuracy and differential sensitivity. High-throughput analysis is established through an accelerated GPGPU implementation.

Availability: Supplementary material, software and images used in the validation are available at http://www.proteomegrid.org/rain/.

Contact: g.z.yang@imperial.ac.uk

1 INTRODUCTION

Proteomics is playing a major role in elucidating the functional role of many novel genes and their products and in understanding their involvement in biologically relevant phenotypes, both in normal cellular processes and disease. Differential (discovery) proteomics has become an important tool in the development of protein biomarkers for earlier and more accurate screening and diagnostic tests for the detection and treatment of disease, as well as screening all human proteins to ascertain their structures and functions - the major biology driven challenges in proteomics today. It has become evident in recent years that these large-scale challenges are too great for the resources of a single laboratory, so open international collaborations are being championed by the Human Proteome Organisation (HUPO - http://www.hupo.org/).

Tissue samples are complex mixtures of up to thousands of different proteins, so adequate identification by mass spectrometry (MS) is reliant on their pre-separation, of which 2-D gel electrophoresis (2-DE) has been the de-facto standard (Görg, et al., 2004). However, limited automation and reproducibility have led to success in combining multi-dimensional liquid chromatography (MDLC) with MS ‘shotgun’ approaches (Wolters, et al., 2001) in which a tryptic digest is separated by one or more dimensions of LC and then automatically subjected to MS/MS. These two separation methods were employed by 9 participating laboratories in the HUPO Brain Proteome Project’s pilot studies, with the result that 57% of proteins were only detected by 2-DE, and 29% only by LC (Reidegeld, et al., 2006). They conclude that both techniques are vital to ensure maximal proteome coverage in future analyses. It is therefore an important research goal that 2-DE data can be automatically aligned, modelled and differentially quantified in a high-throughput pipeline, and integrated with shotgun proteomics data at a much earlier stage in the workflow.

In 2-DE, the 1st dimension separation is isoelectric focusing, during which proteins migrate in a pH gradient until each protein reaches its isoelectric point (pI) where its net charge is zero. In the 2nd dimension, the proteins are separated orthogonally according to their molecular mass by electrophoresis in the presence of sodium dodecyl sulphate. The proteins are then visualised by pre-labelling or post-staining so that the gels can be digitised for analysis. However, several experimental factors hinder the separation power (as illustrated in Supplementary Fig. 1). Important changes in protein expression may therefore be obscured by comparing only two gels, so in practice experiments are replicated multiple times. Between sample comparison is further facilitated by the DIGE (Difference In Gel Electrophoresis) protocol (Görg, et al., 2004), where two cyanine dyes fluorescently label different sample mixtures prior to simultaneous separation on the same gel. Two gel...
images are then acquired using different emission filters. Nevertheless, this approach is still dependent on the accurate alignment and comparison of large sets of 2-D gel images in order to generate robust and sensitive data on differential protein expression.

1.1 Bioinformatics in 2-DE

A review of the bioinformatics for 2-DE is given in Dowsey et al. (2003). Briefly, the traditional image processing pipeline for 2-DE begins with preprocessing of the gel images to reduce noise and subtract out background. A two-step approach then resolves each protein spot individually. The gel is first segmented using the watershed transform before each segment is fitted with a parametric spot model to separate co-migrated spots. Two or more gels are compared for differential expression by point-pattern matching their spot lists annotated with metrics such as spot volume. The final spot correspondence list is then subject to statistical testing with significant manual validation using e.g. Student’s t-test and ANOVA, in order to find outliers that represent up or down regulated (differentially expressed) proteins. With this conventional approach, symbolic representation of the spots is determined at a very early stage of the pipeline. Once such a description is reached, intrinsic errors dependant on the spot modelling and matching persist throughout subsequent processing. Recent studies on the accuracy of spot detection algorithms have shown marked difference in performance (Wheelock and Buckpitt, 2005), highlighting major obstacles in realising the full potential of differential 2-DE.

If one focuses on the raw pixel values, features such as spot shape, streaks, smears, spot tails and background are available which are otherwise lost in the spot detection phase. This image registration approach is an established field in medical imaging (Maintz and Viergever, 1998). The problem is a quadruple, where the source image \( I_s \) is brought into alignment with the reference image \( I_r \), constrained by \( M \), the space of allowable transformations (mappings) on \( I_s \) and guided by optimisation of a similarity function which is at its maximum when the alignment is optimal:

\[
\arg \max_{m \in M} \text{sim}(I_s \circ m, I_r)
\]  

Z3 (Smilansky, 2001) was the first hybrid 2-D gel alignment algorithm to incorporate pixel information, by comparing rectangular image regions each containing a few detected spots. Veeseer et al. (2001) then presented MIR, a direct algorithm with no recourse to spot detection. They register the control points of a piecewise bi-Linear mapping with the BFGS quasi-Newton optimiser and closed form derivatives of the normalised cross-correlation (NCC) similarity measure. Convergence is robust since gels have smooth image gradients and therefore the NCC with respect to the control points is smooth and continuous. To avoid convergence to local optima, a multi-resolution approach is used. In a comparative study between MIR and Z3, MIR scored better 29 out of 30 times under expert visual quantification. Subsequently, the concept was extended with a current leakage model (Gustafsson, et al., 2002), which compared favourably with PDQuest (Bio-Rad Laboratories). In recent years, multi-resolution with complex wavelets has been proposed (Woodward, et al., 2004), as has a realistic deformation model based on tensor-product B-splines (Sorzano, et al., 2008), though no comparative validation was performed in either case.

Driven by the need for more sensitive differential expression analysis, recent studies in gene microarray research have sought to model and correct systematic bias between samples. Two threads of data normalisation can be identified: (a) calibration of nonlinearity between measured intensity and true expression, particularly between different cyanine dyes and spatially within each slide (bias-fields); (b) variance stabilisation to account for the predominantly multiplicative increase in variance with expression abundance, and so allow principled differential expression analysis. For proteomics data, the same requirements for cyanine dye calibration (Karp, et al., 2004), bias-field correction (Gustafsson, et al., 2004) and variance stabilisation (Kreil, et al., 2004) have been acknowledged, and microarray normalisation methods successfully applied in some cases. The issue of spatial inhomogeneity has also been reported in the areas of magnetic resonance imaging (Velthuizen, et al., 1998) and microscopy (Likar, et al., 2000). In each case, the bias-field is assumed to be smoothly varying and therefore separable from: (a) the underlying signal which is locally uniform; (b) differential expression which causes discontinuous changes in intensity. In 2-D gel protein patterns, the assumption of a locally uniform signal is violated, but bias-field estimation can still be performed relative to a second gel, since their difference should be locally uniform. We have previously demonstrated that a thin plate spline model can be used (Dowsey, et al., 2003), whilst Kultima et al. (2006) have similarly applied 2-D LOWESS (local polynomial regression fitting). The disadvantage of these techniques is that they use pre-matched spot list data and so require extensive manual verification, which is incompatible with an automated high-throughput pipeline. They must also cope with the substantial variance and missing values induced by conventional techniques (Wheelock and Buckpitt, 2005). The approach also precludes the use of data normalisation in the gel alignment phase.

In response to many unanswered questions surrounding automated 2-D gel alignment, we describe the RAIN (Robust Automated Image Normalisation) algorithm. The purpose of RAIN is to remove systematic geometric deformation and expression bias in the image domain for objective and fully automated alignment of 2-DE datasets. By performing image-based alignment as the first step of the pipeline, differential analysis and spot modelling can subsequently be performed simultaneously on all gels by fusion of the aligned images, thus leveraging strength across all images in the set (Morris, et al., 2008; Sorzano, et al., 2008).

2 METHODS

The main steps of the RAIN algorithm can be summarised as follows:

1. A smooth tensor-product B-spline surface is used to warp the source gel until all features are brought into correspondence with those in the reference gel. Volume-invariance ensures that expression is not amplified under dilatation, nor attenuated under contraction.

2. Simultaneously, additive and multiplicative bias fields (modelled as smooth B-spline surfaces) are fitted between the images to account for systematic regional changes in background and staining. In this way, the alignment is guided by the underlying expression pattern but not influenced by expression bias and background variation.

3. To assess image correspondence, the sum of squared residuals (SSR) is calculated - a pure pixel-based similarity measure. Since SSR as-
susses data with homogenous variance, an approximate arctanh transformation is applied to the raw pixel values before alignment.

4. Gradient descent is used to find the maximum of the similarity measure with respect to the parameters of the warp and bias-fields. A multi-resolution approach is necessary to reach a global optimum. First, coarse deformations are corrected on sub-sampled images. Then, progressively finer deformations are eliminated on progressively more and more detailed images.

5. For performance, RAIN is implemented using GPGPU (general purpose computation on graphics cards – http://www.gpugpu.org/).

The explanation and motivation for these components is detailed in the next five subsections. This section then concludes by presenting systematic tests to fix application-specific parameters and for algorithm validation.

2.1 Smooth volume-invariant B-spline warping

B-splines are the shortest possible polynomial splines with global continuity, and have been chosen for their efficiency in optimising geometrical transformations for image registration. They are smoothly connected piecewise polynomials where an additional smoothness constraint imposes continuity of the spline and its derivatives up to order (n-1), so that there is effectively only one degree of freedom per segment. A recursive definition beneficial for numerical computation is defined as (Cohen, et al., 1980):

\[
\beta(t) = \sum_{i=0}^{n+1} P_i B_{i,k}(t) \quad \text{where} \quad t_{\text{min}} \leq t < t_{\text{max}}, 2 \leq k \leq n + 1 \quad (2)
\]

where position along the curve is given by t between the start \( t_{\text{min}} \) and end \( t_{\text{max}} \), \( P_i \) are the position vectors of the \( n+1 \) vertices, and \( B_{i,k} \) are the normalised B-spline basis functions (see Supplementary Fig. 2a)). B-splines are linearly separable and therefore have a simple tensor-product extension to 2-D, as illustrated in Supplementary Fig. 2(b).

Standard warping techniques correct for optical distortion and therefore introduce protein over-expression after dilation and under-expression during contraction. Deformation in 2-DE is a physical manifestation, and so to preserve volume-invariance the intensity of each transformed pixel must be normalised, which can be achieved by considering the set of first order partial derivatives of the mapping with respect to its input coordinates. In RAIN, the B-spline warp is third order to ensure C2 continuity in the intensity normalisation as well as to provide a realistically smooth deformation for spot modelling. This ensures accurate quantification on warped gels.

2.2 Bias-field correction of intensity inhomogeneity

Image registration relies on a known deterministic mapping between the intensity values of the reference and source images. A global shift-invariant linear systematic bias can be accounted for by normalised cross-correlation, assuming data with homogenous variance, an approximate arcsinh transformation is applied to the raw pixel values before alignment. For performance, RAIN is implemented using GPGPU (general purpose computation on graphics cards – http://www.gpugpu.org/).

The objective of our image registration approach is to maximise the similarity measure with respect to the B-spline parameters of the warp and bias-fields \( C = \{X, Y, U, V\} \). The choice of optimisation technique for this purpose depends on the presence of local optima in the vicinity of the global optimum. Given that the disparity between 2-D gets scales from global changes to small imperfections, significant local optima are expected. Robust global optimisation algorithms such as simulated annealing can be used, but for image registration with this range of misalignment, local gradient descent coupled with a multi-resolution approach is similarly robust and less computationally demanding.

With a multi-resolution approach, coarse deformations are initially accounted for by using heavily sub-sampled images, before finer and finer deformations are eliminated on progressively more detailed images. Choosing the optimal image resolution for the order of deformation leads to significant improvement in computational efficiency because, (a) no unnecessary
sarily detail is processed for each warp, and (b) the B-Spline transformation has local support, so (i) the computation for each coefficient is limited to its area of influence, and (ii) areas of the gel free from finer distortions can be ignored whilst the rest are optimised. However, if the optimiser does not converge close enough to the optimum, there will be too much residual deformation to correct in the next level. In this framework, the Haar basis (Strang, 1993) was chosen for its low computational complexity.

In this approach, the B-spline transformation is also required to conform to a multi-resolution approach. As shown in Supplementary Fig. 2(c), the optimal B-spline surface found at each level is propagated as the starting estimate for the next level by re-parametrisation with the Osler algorithm (Cohen, et al., 1980). This maintains the topography of the transformation whilst doubling the number of constituent patches along each dimension. By treating our nonlinear least squares problem as a general unconstrained minimisation, quasi-Newtonian algorithms with super-linear convergence can be employed. The optimiser chosen for RAIN is the limited-memory BFGS method (Nocedal, 1980). The method requires first order partial derivatives to be supplied, whilst the Hessian is approximated by the sum of a diagonal matrix and a fixed number of rank-one matrices. B-spline locality can be easily exploited e.g. for a third order B-spline, only the 16 patches around each vertex (less near the borders) are taken into consideration when calculating its partial derivative. To apply the BFGS method, the closed-form partial derivatives of Equation (4) were calculated as:

$$
\frac{\partial \text{SSR}}{\partial i_j} = \sum_{p \in L} \left[ \frac{t_i \ln(t_i + \alpha) - \ln(t_i + \alpha)}{i_j + \alpha} \right] \frac{\partial t_i}{\partial i_j}
$$

(5)

For X, Y, U, V \in C:

$$
\frac{\partial I}{\partial X} = e^{\lambda} \ln(\lambda \beta) \nabla \lambda t \left( \frac{\partial \ln(t)}{\partial \lambda} \right) \nabla \lambda \beta
$$

$$
\frac{\partial I}{\partial Y} = e^{\lambda} \ln(\lambda \beta) \nabla \lambda t \left( \frac{\partial \ln(t)}{\partial \lambda} \right) \nabla \lambda \beta
$$

$$
\frac{\partial I}{\partial U} = e^{\lambda} \ln(\lambda \beta) \nabla \lambda t \left( \frac{\partial \ln(t)}{\partial \lambda} \right) \nabla \lambda \beta
$$

$$
\frac{\partial I}{\partial V} = \frac{\partial \beta}{\partial \lambda}
$$

where \( \nabla J_i(x, y) \) is the gradient image in direction \( \alpha \), and:

$$
\frac{\partial \beta}{\partial z} \left( x, y \right) = \frac{\partial \beta}{\partial \alpha} x \otimes y
$$

$$
\frac{\partial \beta}{\partial z} \left( x, y \right) = \frac{\partial \beta}{\partial \alpha} x \otimes y
$$

where Z = X, Y, U or V, \( \otimes \) and \( \otimes \) are the 4×4 B-spline surface control points and \( \otimes \) is the Kronecker product.

2.5 Computational details

Pseudocode for the RAIN algorithm is given in Supplementary Fig. 3. Briefly, multi-resolution pyramids are first constructed for the reference and source images. RAIN then commences with an initial registration to account for positional and contrast/brightness variation in gel scanning. The shifted-log SSR similarity measure is then optimised on the coarsest multi-resolution level with a reduced transformation of the warp \{X, Y\} and bias-fields \{U, V\} until neither lead to an improvement in similarity. This approach was utilised because with simultaneous optimisation, the deformation field frequency diverged due to volatility in variable perturbation whilst the Hessian stabilised during the first few iterations. In other words, a simultaneous poor estimation of the warp and an opposing (and therefore nullifying) bias-field estimate would result in an erroneous increase in similarity.

For a substantial speedup, the B-spline transformation and similarity measure were implemented using GPGPU technology on Nvidia (http://www.nvidia.com/) consumer graphics card hardware with the Cg programming language, OpenGL and Nvidia-specific extensions. A schematic diagram is presented in Supplementary Fig. 4. In brief, once the resolution of the B-spline patches is chosen, the array of sampling coordinates does not change, nor does the Basis function. The vector-matrix is therefore pre-multiplied and stored as a texture. The B-spline transformation is linearly separable, so the y dimension is implemented as a separate pass. The x dimension is then calculated from the results of the y dimension, and the warp is performed. The residuals and the partial derivatives are then calculated and summed, as in Supplementary Fig. 5. However, since graphics cards have no accumulator registers to combine results in parallel, the summation is performed by a 4 stage reduction, which is a multi-pass accumulation of an array of data by a local operator working up an image pyramid (see Supplementary Fig. 6). The final SSR value and derivatives are transferred back to the main processor for limited-memory BFGS optimisation using the VXL library (http://vxl.sourceforge.net/).

2.6 Parameter tuning and validation

Three application-specific parameters require tuning through comprehensive testing on large 2-D gel sets. These are: the linearity in the shift-log transformation, \( \alpha \); the resolution of the lowest multi-resolution level (single B-spline patch resolution); and the total number of multi-resolution levels.

To determine the optimal values of \( \alpha \) and the B-spline patch resolution, several experiments were conducted to investigate the likelihood of local optima and curvature of the similarity measure. \( \beta_x, \beta_y, \beta_\alpha \) and \( \beta_\beta \) were set to single B-spline patches (4×4 degrees of freedom each). A ‘ground truth’ alignment was attained by a conservative multi-resolution BFGS optimisation with resolution levels from 4×4 to 2048×2048 and no shifted-log, which was verified manually. The experiment was then repeated with image resolutions of 4×4, 8×8, 16×16 and 32×32, and log-shifts from \( \alpha = 10^\beta \) to total linearity. Currently, an automated solution for \( \alpha \) was not sought, since there is evidence it does not vary significantly for a particular 2-DE protocol, and therefore can be determined once per workflow.

To determine the number of multi-resolution levels required, the point at which the modelling became unconstrained and unrealistic was observed. At a particular multi-resolution level, the transformation deforms individual spots and not spot patterns, so no further improvement in spot alignment is seen. However, so that the warp will only deform in directions with sufficient evidence, regularisation should be added at this level and the algorithm then terminate. Since the registration did not require regularisation to converge to an accurate spot alignment, in this study we simply threshold the derivatives of the similarity measure at the final level.

Once the final multi-resolution level was determined, RAIN’s operating bounds were investigated through back-registration of synthetically warped gels. The hierarchical B-spline coefficients of the geometric warp and bias-fields were perturbed by a normally distributed random process, from a single B-spline patch to 32×32 patches, thus simulating deformation at all scales. The normal distribution was set to zero mean, whilst the standard deviation was varied. A set of warps was therefore created for each input gel, each one slightly more severe than the last. For real-world validation of the complete algorithm, an expert was given the task of quantifying spot mismatches between gels processed by RAIN and MIR double-blind. In line with the MRI vs Z3 comparison (Veeseer, et al., 2001), the threshold criterion used was that 75% of the spot shape must overlap. The ‘score’
devised by Veesar et al. was also computed, which takes into account the spatial distribution of mismatches (see Supplementary Table 1). Both intra- and inter-sample gel sets were examined, to assess each algorithm’s robustness to differential expression. Large sets of 2-D gels were provided through proteomics research carried out at Imperial College London and University College Dublin (UCD). The 4 sets under study were: (a) UCD Silver Stain: A large gel study of human brain proteins to differentiate between two mental disorders and controls; (b) UCD DIGE: Brain proteins of 30 female mice at three growth stages; (c) Imperial: Imperial™ stained human osteosarcoma cell proteins after exposure to hormone PTH for 1, 4, 8, 16 and 32 hours; (d) Veesar et al.: 2 silver stained experiments published (http://www.doc.ic.ac.uk/vip/2d-gel) for evaluation of future gel alignment algorithms. Complementing the results presented here, an extensive investigation of RAIN on HUPO Brain Proteome Project datasets for 9 participating laboratories is presented in Dowsey et al. (2006).

3 RESULTS

3.1 Parameter tuning

Parameter tuning was conducted on a representative gel from each of the four validation sets. In Fig. 1 and Supplementary Fig. 7, representative coefficients of X and Y in the minimisation of ‘ssr’ are presented for a Veesar et al. silver stained gel. These illustrate: the most local minima near the global minimum; the broadest valley; and the greatest change between resolutions. Bias-field graphs are not shown since they are significantly more stable. From these experiments, it can be concluded that: (a) The most favourable values for variance stabilisation were \( \alpha = 0.1 \) for silver and Imperial™ stained gels (large additive element) and \( \alpha = 0.0005 \) for DIGE (small linear element in line with their increased dynamic range); (b) As shown in Supplementary Fig. 8, if the resolution is too low (4×4), the minima are too far from the real minimum. At high resolution (32×32), local minima nearer the global minimum confuse the optimiser; (c) Similar overall alignment accuracy was reached with resolutions of 8×8 and 16×16, but the computed minimum with 16×16 was less approximate and so led to a smoother transformation, and was therefore deemed the optimal resolution.

![Fig. 1](image1.png)

Fig. 1. The objective function for row 3, column 2 of \( X \) and row 3, column 4 of \( Y \) of a single B-spline patch (\( X \) & \( Y \) are 4×4 matrices; see Supplementary Fig. 2 for a visual representation). Offset 0 represents the global optimum. (left) 16×16 resolution with no log-shift and log-shift of 0.001, 0.01, 0.1. (right) resolutions of 4×4, 8×8, 16×16 and 32×32 with no log-shift.

Table 1. Cumulative execution time in seconds for 0 to 8 levels of RAIN.

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>Sdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>20.9</td>
<td>1.95</td>
</tr>
<tr>
<td>1</td>
<td>43.1</td>
<td>1.57</td>
</tr>
<tr>
<td>2</td>
<td>68.8</td>
<td>1.41</td>
</tr>
<tr>
<td>3</td>
<td>98.1</td>
<td>4.01</td>
</tr>
<tr>
<td>4</td>
<td>162</td>
<td>8.03</td>
</tr>
<tr>
<td>5</td>
<td>230</td>
<td>16.2</td>
</tr>
<tr>
<td>6</td>
<td>392</td>
<td>33.1</td>
</tr>
<tr>
<td>7</td>
<td>671</td>
<td>33.1</td>
</tr>
<tr>
<td>8</td>
<td>1306</td>
<td>80.2</td>
</tr>
</tbody>
</table>

Level 0 is translation. Levels 1 to 8 optimise from 1 B-spline patch to 128×128 patches. Trials run on a 2Ghz Pentium M with Nvidia 6800Ultra. Multi-resolution performance was then determined, from 0 levels (translation only) to 8. Table 1 illustrates that each level increases run time by ~65%. At multi-resolution level 8, unconstrained warping and bias-fields at the scale of individual spots led to mismatches in regions of co-migrated differential expression. Since there are no discontinuities to constrain the bias-fields, the differential expression is propagated to co-migrated spots. This in turn causes the warping to deform towards these spots, eventually leading to a false positive match for the unexpressed protein. For this reason, and since no additional spot matches were observed at level 7, a multi-resolution pyramid of 6 levels was chosen for this study. Fig. 2 shows the result of multiplicative and additive bias-field correction during alignment for these 6 multi-resolution levels.

3.2 Validation

Subjective validation was performed between technical replicates (intra-set assessment) and between control and sample gels (inter-set assessment). Intra-set assessment was performed between 36 pairs of UCD Silver, UCD DIGE and Imperial. Technical replicates are not present in Veesar et al.. In Table 2 and Supplementary Table 2, mean and standard deviation is shown for spot mismatches per gel and score respectively, excluding outlier results such as total failures and problem regions which would skew the distribution. The results show RAIN matches substantially more spots without introducing a greater number of false-positives. In the DIGE experiment, MIR failed to converge at all in 6 gel pairs due to the significantly higher background bias in DIGE.

Inter-assessment was performed between the three sets of UCD Silver (24 pairs), the five time-points of Imperial (24 pairs), and the 10 pairs in the Veesar et al. easy and medium groups. Inter-assessment was not applicable to DIGE gels since the matching is always performed against the pooled sample. Both MIR and RAIN’s matching performance decreased under differential
expression, but RAIN is more robust as the images become less and less similar: the Imperial gels exhibit large disparities, especially at the latter time-points. MIR proved unsatisfactory on these gels, diverging 3 times and mismatching many spots in the other gels. This notably occurred in heavily smeared regions, whereas RAIN’s additive bias-field equalised the disparity. The UCD Silver gels showed less dissimilarity but differences between the schizophrenia and control gels were still significant. Two of the medium difficulty Veeser et al. gels exhibit problem regions due to issues in the 1st dimension IPG strip. Ignoring these areas, RAIN only跨越式 4 problem regions. See Fig. 3(a) for an example.

Table 2. Subjective expert validation of mismatched spots (<75% overlap) after alignment of the four gel sets with MIR and RAIN.

<table>
<thead>
<tr>
<th></th>
<th>UCD Silver</th>
<th>Imperial</th>
<th>UCD DIGE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MIR</td>
<td>RAIN</td>
<td>MIR</td>
</tr>
<tr>
<td>Mean</td>
<td>15.50</td>
<td>1.17</td>
<td>20.12</td>
</tr>
<tr>
<td>Stdev</td>
<td>2.40</td>
<td>1.36</td>
<td>12.10</td>
</tr>
<tr>
<td>Inter-sample</td>
<td>UCD Silver</td>
<td>Imperial</td>
<td>Veeser et al.</td>
</tr>
<tr>
<td></td>
<td>MIR</td>
<td>RAIN</td>
<td>MIR</td>
</tr>
<tr>
<td>Mean</td>
<td>17.67</td>
<td>6.33</td>
<td>25.77</td>
</tr>
<tr>
<td>Stdev</td>
<td>10.97</td>
<td>6.84</td>
<td>13.98</td>
</tr>
</tbody>
</table>

\(^1\)Excluding 6 total failures (divergence of the warp).
\(^2\)Excluding 4 problem regions. See Fig. 3(a) for an example.

mismatched 2 spots over the 10 pairs and MIR did an acceptable job, mismatching a mean 4.8 spots per gel. Fig. 3 shows two of the gel pairs picked for their difficulty: (a) shows a 1st dimension protein migration problem that causes a void in one of the gels. RAIN is able to match more spots around the problem region, as well as in more sparse regions; (b) illustrates a match between Imperial control osteoblast gels, subject to saline (control) and hormone PTH (sample) for 32 hours. A large pattern of proteins is missing from the sample gel, which RAIN’s additive bias field recreates in the later stages of the algorithm so nearby matching is not affected.

The results of the synthetic warping investigation are shown in Fig. 4(a). One representative gel was chosen from the UCD DIGE, UCD Silver and Imperial sets, and two from the Veeser et al. set (one ‘easy’ and one ‘medium’). Five random processes were generated and, for each one applied to each gel, warps were output by varying the standard deviation from 0.100 to 0.350 at 0.025 intervals. In total, 5 × 3 × 11 = 275 pairs were matched with RAIN and analysed for spot mismatches and the Warping Index (Thevenaz and Unser, 2000), which computes the mean distance error in the vector field resulting from the composition of the synthetic transformation with the back-registered transformation. We found good agreement between the Warping Index and spot mismatches. RAIN was generally able to reconstruct the original images well (spot mismatches < 10 and Warping Index < 0.5% of image size) when the warp had a standard deviation of \(\sigma \leq 0.250\). Fig. 4(b) illustrates an example with \(\sigma = 0.225\). Until \(\sigma = 0.325\), major regions of the gel were matched, but at \(\sigma = 0.350\) RAIN generally diverged completely (a series of increasing synthetic warps is given in Supplementary Fig. 9). In the four gel sets used for validation, gels exhibiting misalignment similar to a synthetic warp with \(\sigma \geq 0.250\) were not evident, and can in general be considered poor quality.

Fig. 3. Results of subjective validation of MIR (left) and RAIN (right) with mismatched spots boxed. (a) Veeser et al. set: RAIN’s additive bias-field actually recreates the missing protein pattern (see insert with the full bias-fields applied).

Fig. 4. (a) Graph of RAIN on a synthetically warped gel, where the hierarchical B-spline coefficients were perturbed by a normally distributed random process with standard deviation from 0.100 to 0.350 of B-spline patch length. For the mean of all 25 tests, the Warping Index is shown in blue and the number of mismatched spots is shown in red. (b) Example synthetic warp of standard deviation 0.225 (green) overlaid onto the original gel (magenta), before (top) and after (bottom) RAIN registration.

4 DISCUSSION AND CONCLUSIONS

We have shown that RAIN is a substantial improvement over a previous state-of-art algorithm MIR, due to incorporation of realistically smooth volume-invariant warping, bias-field correction, shifted-log variance stabilisation, DIGE gel compatibility and a
GPGPU implementation. RAIN handles both artefacts and intensity inhomogeneities, and remains robust as deformation severity increases. MIR alignment takes ~3 seconds, so it can be performed interactively. It is fast because the transformation is simple and the number of optimisable parameters is low. Whilst RAIN is relatively slow (~7 minutes), the higher sensitivity suits it to pipeline processing, where gels are processed in batch and computational complexity is alleviated with the aid of the proTurbo distributed cluster-image-computing framework (Dowsey, et al., 2004).

A question is left as to whether the aligned image output by RAIN could be left bias-corrected. In RAIN, an implicit Euclidian constraint is defined, so that regulated expression without local support from unregulated expression (e.g. in sparse gel areas) will invariably be normalised out. Whilst this behaviour is suppressing all inhomogeneities for image alignment, to ensure correct quantification of differential expression the bias-field would need to be more constrained in sparse areas to account for the consequent reduction in support. One example of this technique is LOWESS (Kultima, et al., 2006), where the n nearest neighbour proteins are used to smooth the data, regardless of their proximity.

We believe that a holistic ‘statistical expression analysis’ (SEA) approach to differential expression analysis, where group-wise analysis is performed simultaneously on whole sets of gels, is required to unlock the information trapped in 2-DE and LC/MS data. Small insignificant expression changes over two datasets could become significant when reinforced by the same consistent changes in the others. With SEA, the richness in information and its associated uncertainties are fully captured in a model independent manner. Our ProteomeGRID (http://www.proteomegrid.org/) project is an initiative to interface RAIN and proTurbo with SEA and a proteomics repository. This will realise our main goal of providing a centralised and automated bioinformatics resource for proteomics experiments. Gels and LC/MS data annotated with standardised metadata (http://www.psidev.info/) will be submitted via a web services interface for direct comparison with RAIN. The normalised data will then be automatically mined for differential expression with SEA, drawing on the statistical norms for that tissue type, protocol and laboratory from the integrated proteomics database. The statistical norms will also be updated with the results, whilst the differential proteins will be identified by integrated and automated online MS informatics.

To conclude, it is clear that a solution for the full automation of the bioinformatics pipeline will be required to realise large-scale proteomics for drug discovery and proteome mapping. In this paper, we have presented RAIN as one component in this system. Further challenges include: integration of LC/MS data into the RAIN/SEA pipeline; and automatic image-based variance stabilisation and constrained bias-field correction, which will pave the way for sensitive and robust SEA-based differential analysis.

FUNDING
Engineering and Physical Sciences Research Council UK (EP/E03988X/1 to A.W.D., GR/T06735/01 to G.-Z.Y.); Science Foundation Ireland (04/RPI/B499 to M.J.D.)

ACKNOWLEDGEMENTS
This work was supported by the Engineering and Physical Sciences Research Council UK (GR/T06735/01 and EP/E03988X/1). Research in M.J.D.’s laboratory is funded by Science Foundation Ireland (04/RPI/B499). We thank members of M.J.D.’s and David Cotter's (Royal College of Surgeons in Ireland) research groups (Paul Boersema, Jane English, Melanie Foecking, Kyla Pennington) and Judit Nagy and Nikolay Zhukovsky (Imperial College London) for providing the gel images for this study.

REFERENCES
Dowsey, A.W., et al. (2006) Examination of 2-DE in the Human Proteome Organisation Brain Proteome Project pilot studies with the new RAIN gel matching technique, Proteomics, 6, 5030-5047.