

Fixing Proteomics Initiative: A Global Quest for Reproducibility

K. Bala¹, A. Borthwick², W. Dracup², P. Lavery², Martin O' Gorman² and H. Voshol³
¹Bio-Rad Laboratories, 6000 James Watson Dr., Hercules, CA 94547
²Nonlinear Dynamics, Keel House, Garth House, Newcastle upon Tyne, NE1 2JE
³Novartis Institutes for BioMedical Research, Basel, Switzerland



BIO-RAD

Introduction

Fixing Proteomics Campaign (HUPO Phase II Study) is a noncommercial, vendor-independent campaign dedicated to solving the experimental challenges that prevent proteomics from delivering on its potential. Encouraging reproducibility of results has emerged as the key factor toward achieving this end. The organization's website www.fixingproteomics.org provides background on the importance of and tips for applying quantitative proteomics techniques in a reproducible manner.

Here we focus on two-dimensional electrophoresis (2-DE), arguably the most widely used method for comparative protein profiling. In spite of some important technology developments, 2-DE remains a rather complex method and obtaining reproducible 2D gels with well-resolved spots continues to be a challenge to many researchers, especially to those starting in the field.

The goal of this project is to facilitate the introduction to and use of 2-DE by providing reference protocols, images and samples. This poster describes the first experiences with those reference materials, which were distributed to around 20 labs worldwide with the request to produce at least three gels and submit the images.

Materials & Methods

Sample: HeLa Cell Lysate (Prepared by CiBiotech)

↓
Suspend in Lysis Buffer

↓
1st Dimension IEF Analysis – 24cm IPG Strip, pH 4-7

↓
2nd Dimension – 12% TG/SDS Gels (20x25cm, 1.5mm Thick)

↓
Stain and Image

↓
Image alignment and analysis to compare similarity with a set of "Gold Standard" images visualized using Principal Component Analysis (PCA)

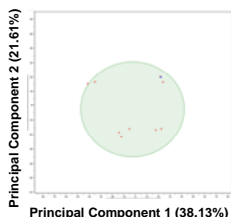


Fig.1. Principal Component Analysis (PCA). Spot expression measures from aligned gel images were used to determine the principle axes of variation and principle component 1 was plotted against principle component 2. Transforming and plotting the data in principle component space allows us to separate the gel images according to this variation. We can graph transformed gel image data on a biplot like the one here and compare one principle axis of variation against another. The biplot contains a lot of information and can be helpful in interpreting relationships between experimental groups and identify outlier gels, i.e. gels that have different properties to other gels in the same groups.

The reference sample consisting of a HeLa cell lysate in a standard 2-DE lysis buffer was provided to the labs listed in Table 1. Details on the sample preparation, gel running, detection and imaging can be obtained by visiting www.fixingproteomics.org. Submitted gel images were compared automatically to the "Gold Standard" reference gel image (Fig. 2 A and B) using a "Gel QC feature" being developed within Progenesis SameSpots software (Nonlinear Dynamics, Newcastle, UK). No spot editing was performed. Subsequently, normalized intensities of matched spots were subjected to Principal Component Analysis (PCA) to determine whether a certain image fell within a 95% confidence boundary of a predefined reference set, consisting of 7 gel images for intra-lab comparisons and 10 gel images for inter-lab comparison.

Example - Matching of Submitted Images

Figure 2 below shows Gold Standard gel image (A) compared to a gel image from a participatory lab (B). Study participant followed the protocol and generated the image that is compared to the Gold Standard. Highlighted region in the Red Box is for visual reference.

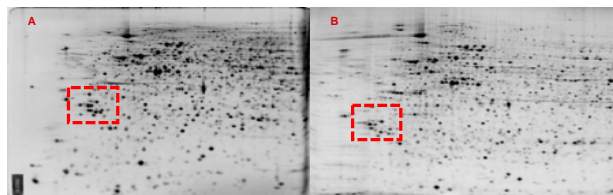


Fig.2. Comparison between Gel A (Gold Standard) to Gel B (Participant).

Comparison to the Reference Set

Figure 3 shows the comparison of four test gel images to a reference set of 7 gel images of the same sample from different users all within the same lab using PCA. All images fall within or on the 95% confidence boundary and show good measure of similarity.

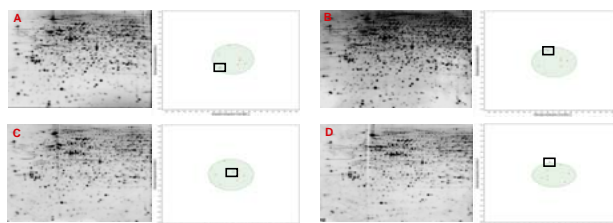


Fig.3. PCA biplots of 4 gel images (A,B,C,D) represented by the blue dots (within the black box) compared to a reference set made up of 7 gel images represented by the red dots. The elliptical green shape marks the 95% confidence boundary in the PCA space.

Figure 4 shows the comparison of 4 gel images to a reference set of 10 gel images from global labs running the same sample using PCA. Biplot of the first two principle components includes an elliptical green shape which marks the 95% confidence boundary in this PCA space. Three gel images (A,B,C) fall within this boundary, while the fourth gel image (D), with clear distortions at the anodic end in the high molecular weight region is outside the boundary.

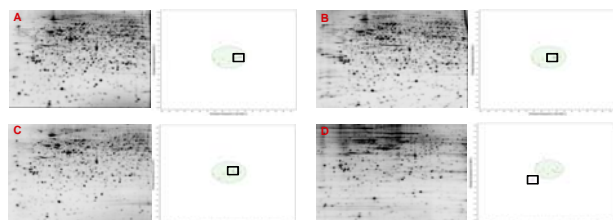


Fig.4. PCA biplots of 4 gel images (A,B,C,D) represented by the blue dots (within the black box) compared to a reference set made up of gel 10 images represented by the red dots. The elliptical green shape marks the 95% confidence boundary in this PCA space.

We found that, while PCA is a useful tool to reduce the dimensionality of the dataset, our implementation is biased towards gels that are, geographically similar to the reference gel. Therefore sometimes gels that are visually of good quality are called *Outliers*, because of differences in the spot pattern which cannot be reconciled by automatic alignment alone.

Reference Materials Lead to Quality Improvements

Initially several participants had difficulties reaching an optimal result with the reference sample (Fig. 5. Gel 1). In several cases, the observed inconsistencies immediately led to suggestions towards improvement (Fig. 5. Gel 2) - Peer to Peer Education.

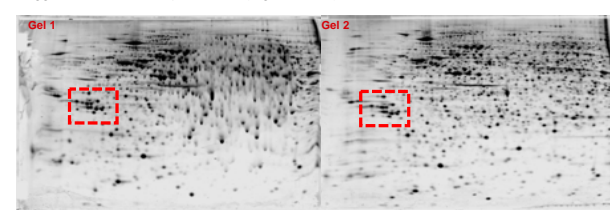


Fig.5. User G Attempt 1 Gel 1 vs User G Attempt 1 Gel 2: illustrates the value of consultative approach leading to the improvement in the quality of the image in Gel 2.

Table 1. Enrolled Participants - Fixing Proteomics Initiative

Participant	Lab	Country
Ben Herbert	University of Technology, Sydney	Australia
Alamgir Khan/Mark Baker	Australian Proteomics Facility	Australia
Toni Posch	Bio-Rad Laboratories, Munich	Germany
Ravi Srideshmukh/Poonam Gautam	Center for Cellular and Molecular Biology	India
Mike Dunn	University College, Dublin	Ireland
Julie Polden/Aisling Robinson	University College, Dublin	Ireland
Alexander Archakov/A.Melnik/S.Moshkovski	Russian Academy of Medical Sciences	Russia
Maxy Chung	National University of Singapore	Singapore
Jose Bermudez	Universidad de Santiago de Compostela	Spain
Francisco Canals	Vall d'Hebron University, Barcelona	Spain
Peter James	Lund University	Sweden
Glen Kemp/Kaveh Emani	NEPAF, Newcastle Upon Tyne	UK
Jun Wheeler/Iolanda Vendrell	NIBSC, Hertfordshire	UK
Aran Paulus/Katrina Academia	Bio-Rad Laboratories, Hercules	USA
Amrita Cheema/Alex Kirilyuk	Georgetown University	USA
Wayne Chadwick	National Institute on Aging, Baltimore	USA
Aldrin Gomes	University of California, Davis	USA
Melissa Sondaj	University of California, Los Angeles	USA
Philip Andrews/Mary Hurley	University of Michigan	USA
Jim Malone/Petra Gilmore	Washington University, St. Louis	USA

Results

- Intra-lab Reproducibility – Highly reproducible for gels run within the same lab
- Inter-lab Reproducibility – 10/17 labs (~60%) were able to generate gel images which fall within the 95% confidence bound as determined by PCA.

Conclusion and Lessons Learned

- This project illustrates that a complex proteomics method such as 2-DE can benefit greatly from the availability of reference materials. With the immediate feedback on the quality of their gels, users can gain confidence in being able to generate reproducible data with their precious samples. Overall, ~60% of the participating labs were able to produce reproducible data even under low stringency.
- Use of same starting material will be helpful
 - Availability of reference images (intra and inter-lab) could reduce bias
 - Even the best labs will see user-dependent variation
 - Minimize variation arising from product quality
 - Iterative approach will help reduce issues - streaking, poor spot resolution and artefacts.

Next Steps

- Recommending standard protocol to validate performance
- Distribution of test standards & kits
- Creation of central image repository
- Global education to produce high quality and highly reproducible data
- May be expand study under stringent conditions

Acknowledgement

The authors gratefully acknowledge the contributions of all participants in this global initiative and thank Alain Miller (CiBiotech) for a continuous supply of high-quality reference samples.