

# MIAPE: Gel Informatics

Version 1.0, July 2009.

Christine Hoogland<sup>1</sup>, Martin O'Gorman<sup>2</sup>, Philippe Bogard<sup>2</sup>, Frank Gibson<sup>3</sup>, Matthias Berth<sup>4</sup>, Simon J Cockell<sup>5</sup>, Andreas Ekefjård<sup>6</sup>, Ola Forsstrom-Olsson<sup>6</sup>, Anna Kapferer<sup>6</sup>, Mattias Nilsson<sup>6</sup>, Salvador Martínez-Bartolomé<sup>7</sup>, Juan Pablo Albar<sup>7</sup>, Sira Echevarría-Zomeño<sup>8</sup>, Montserrat Martínez-Gomariz<sup>9</sup>, Johann Joets<sup>10</sup>, Pierre-Alain Binz<sup>11</sup>, Chris F Taylor<sup>12</sup>, Andrew Dowsey<sup>13</sup>, Andrew R Jones<sup>14,†</sup>

1. *Swiss Institute of Bioinformatics, Proteome Informatics Group, Genève, Switzerland*
2. *Nonlinear Dynamics, Cuthbert House, All Saints, Newcastle upon Tyne, NE1 2ET, UK*
3. *School of Computing Science, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK*
4. *Decodon, GmbH W.-Rathenau-Str, 49a, 17489 Greifswald, Germany*
5. *Bioinformatics Support Unit, Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK*
6. *Ludesi, Lund, Sweden*
7. *ProteoRed, National Center for Biotechnology-CSIC, 28049, Cantoblanco, Madrid, Spain*
8. *Agricultural and Plant Biochemistry and Proteomics Research Group, Department of Biochemistry and Molecular Biology, University of Córdoba, Córdoba, Spain*
9. *ProteoRed, Proteomic Facility, Universidad Complutense de Madrid-Parque Científico de Madrid (UCM-PCM), Madrid, Spain*
10. *UMR 0320/UMR 8120 Génétique Végétale, Gif-sur-Yvette, France*
11. *Swiss Institute of Bioinformatics, Rue Michel-Servet 1, CH-1211 Geneva 4, Switzerland and GeneBio SA, Av. de Champel 25, Geneva, Switzerland*
12. *EMBL Outstation, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK*
13. *Inst. of Biomed. Eng., Imperial Coll. London, London, UK*
14. *Department of Pre-clinical Veterinary Science, Faculty of Veterinary Science, University of Liverpool, Liverpool, L69 7ZJ, UK*

† Corresponding author, email: [andrew.jones@liv.ac.uk](mailto:andrew.jones@liv.ac.uk)

**This module identifies the minimum information required to report an informatics analysis performed with gel electrophoresis images, in a manner compliant with the aims as laid out in the 'MIAPE Principles' document.**

## Introduction

Gel electrophoresis is a reliable, well-characterised separation technique for proteins and peptides (inter alia), underpinning a wide variety of specific protocols. The performance of gel electrophoresis (whether one or two dimensional, denaturing or 'native', simple or multiplexed, etc.) is typically followed by gel

image capture, minimally to provide an (annotated) electronic record of the result. Such images (in particular, complementary images from multiplexed gels) are increasingly subject to informatics analyses such as spot detection, spot quantitation, warping and scaling of whole images to enable their comparison or combination, as well as comparative analysis between images (including measures of reproducibility, clustering, classification and differential analysis).

This document is a single module, as part of the MIAPE (Minimum Information about a Proteomics Experiment) family of reporting guidelines, produced by the Proteomics Standards Initiative ([www.psidev.info](http://www.psidev.info)). MIAPE

is structured around a parent document that lays out the principles of a reporting guideline [1]. In brief, a MIAPE module represents the minimum information that should be reported about a dataset to allow a reader to interpret and critically evaluate the conclusions reached, and to support their experimental corroboration. In practice a MIAPE module comprises a checklist of information that should be provided (for example about the protocols employed) when a data set is submitted to a public repository. The MIAPE modules specify neither the format that data should be transferred in, nor the structure of the repository. However, the PSI is not developing the modules in isolation; several compatible data exchange standards are now well established and supported by public databases.

These guidelines cover the origin and nature of source images derived from two-dimensional gel electrophoresis, algorithms (with parameters) used to process those images, any relationships between (sets of) images or spots known to exist prior to the experiment or established by analysis and the stable location at which the various (meta)data were deposited. They do not cover the running of a gel or any prior steps, nor do they cover the actual image capture process, which is covered in the MIAPE Gel Electrophoresis document. They are not explicitly designed to cover analysis of images derived from one-dimensional gel electrophoresis, although most of the guidelines could be applied in this context. Items falling outside the scope of this module may be captured in complementary modules (for instance the generation of the electrophoresis experiment), which can be obtained from the MIAPE website (<http://psidev.info/miape/>). Note that subsequent versions of this document may have altered scope, as will almost certainly be the case for all the MIAPE modules.

The following section details the reporting requirements of gel informatics:

1. General features
  - 1.1 Responsible person or institutional role
  - 1.2 Electrophoresis type
  - 1.3 Electrophoresis context
  - 1.4 Image(s)
  - 1.5 Image analysis software

2. Gel analysis design
  - 2.1 Type
  - 2.2 Replicates
  - 2.3 Groups (not applicable on blind design)
  - 2.4 Standards

3. Image preparation (if performed)  
*(edits performed on the image(s) using image software)*

- 3.1 Software
- 3.2 Preparation Steps
  - 3.2.1 Type
  - 3.2.2 Parameters

4. Image processing (if performed)  
*(Any processing performed by the bioinformatics software).*

- 4.1 Input image(s)
- 4.2 Software
- 4.3 Processing steps
  - 4.3.1 Algorithm name/version
  - 4.3.2 Parameters

5. Data extraction process  
*For the following possible processes, assemble the sections that are most appropriate for the workflow performed.*

- 5.1 Input image(s)
- 5.2 Feature detection (either automatic or manual) (if performed)
  - 5.2.1 Algorithm name/version
  - 5.2.2 Parameters
  - 5.2.3 Feature editing (yes/no, manual/automatic)
- 5.3 Matching (if performed)
  - 5.3.1 Algorithm name/version
  - 5.3.2 Parameters
  - 5.3.3 Reference image used
  - 5.3.4 Landmarks
  - 5.3.5 Match editing
- 5.4 Feature quantitation (if performed)
  - 5.4.1 Type
  - 5.4.2 Quantitation
  - 5.4.3 Background subtraction (if applicable)
  - 5.4.4 Normalization (if applicable)

6. Data analysis  
*Analysis protocol used, repeat the steps 6.1 – 6.6 as necessary*

- 6.1 Analysis intent
- 6.2 Input data
- 6.3 Data transformation
- 6.4 Software
- 6.5 Type
- 6.6 Parameters

- 7. Data reporting
  - 7.1 List of image features
  - 7.2 List of matches
  - 7.3 Analysis results

## Summary

The MIAPE: Gel Informatics minimum reporting requirements for the analysis and interrelation of gel electrophoresis data specify that a significant degree of detail be captured on the origin and context of gel images, their analysis and the resultant data set. However, providing the information required by this document should enable the effective interpretation and assessment of gel informatics data and potentially, support experimental corroboration. Much of the information required herein may already be stored in an electronic format, or may be exportable from analysis software; we anticipate further automation of this process.

*These guidelines will evolve. To contribute, or to track the process to remain 'MIAPE-compliant', browse to the website at <http://psidev.info/miape>*

## Appendix I. The MIAPE Gel Informatics glossary of required-terms

<i>Classification</i>	<i>Definition</i>
<b>1. General features</b>	
1.1 Responsible person or role (or institutional role if more appropriate); provide name, affiliation and stable contact information	The (stable) primary contact person for this data set; this could be the experimenter, lab head, line manager etc. Where responsibility rests with an institutional role (e.g. one of a number of duty officers) rather than a person, give the official name of the role rather than any one person. In all cases give affiliation and stable contact information.
1.2 Electrophoresis type	A description of the gel electrophoresis experiment type, such as a two-dimensional gel, a two-dimensional differential electrophoresis gel.
1.3 Electrophoresis context	The project or experiment for which the present analysis was initiated. A reference to the relevant instance of the MIAPE Gel Electrophoresis document(s) [2] should be given here if available.
1.4 Image(s)	Indicate image names and identifiers. Describe the type of input image(s) (for example visible light scan, synthetic image or fluorescent scans). Information related to the image acquisition is covered in the related MIAPE Gel Electrophoresis document (Section 7). If not, provide details as per Section 7 of MIAPE Gel Electrophoresis.
1.5 Image analysis software	Image analysis software information. Software name and version number, vendor (or if not available a literature reference or a URI).
<b>2. Gel analysis design - The design with respect to number and type of gels, such as directed or blind, to capture the overall design of the experiment</b>	
2.1 Type	State the analysis design with respect to the assignment of images to groups. For example, directed (when groups are defined) or blind (groups are unknown).
2.2 Replicates	Type (for example biological or technical) and number of replicates.

2.3 Groups (not applicable on blind design)	Collection of images based on user-defined classification (for example control or disease set, time course, doses, stimuli etc). State the images within each group (with references to the images documented in Section 1.5).
2.4 Standards	List standards used in the following categories: internal standards - an additional sample run in the same physical gel for a specific purpose, for example, an internal standard in a DIGE experiment; external standards - any additional sample run in different gels for a specific purpose (for example characteristic alignment); or within sample standards - a protein or protein mixture added in a defined amount for purposes such as normalization or alignment.
<b>3. Image preparation</b> – Any edits performed on the image(s) using image software prior to specific bioinformatics analyses or processing. Each preparation step should be listed under section 3.2 with a relative order.	
3.1 Software	Software used for the manipulation, if different from the analysis software defined in Section 1. Software name and version number, vendor (or if not available, provide a literature reference or a URI).
<b>3.2 Preparation steps</b> - The editing steps (usually untraceable) performed on the image.	
3.2.1 Type	Type and description of editing performed, for example rotation, scaling, resizing or resolution change, inverting gray level, format change, defining an inclusion region (regular or irregular cropping).
3.2.2 Parameters	Any parameters used in the editing step.
<b>4. Image processing</b> – Any processing performed by the bioinformatics software to prepare the images for data extraction. List the image processing steps performed with parameters and their relative order.	
4.1 Input image(s)	The image(s) on which processing is performed, given by reference to images specified in Section 1.
4.2 Software	Software used for the processing, if different from the analysis software defined in Section 1. Software name and version number, vendor (or if not available, provide a literature reference or a URI).
<b>4.3 Processing steps</b> - Information about the analysis processing steps, for example denoising, background subtraction (only when used as an image processing step), image alignment and defining an irregular region of interest (basic image cropping is reported under Image	

Preparation).	
4.3.1 Algorithm name/version	The name and version of the algorithm as part of the specified software (if available) or default algorithm of software.
4.3.2 Parameters	Parameters used in applying the algorithm (for example smooth, min area, saliency, etc.). If image alignment is performed as a processing step, the reference image and landmarks used should be recorded here - landmarks should be provided in a suitable format (e.g. boundary chain, image map, regular shape etc) as for section 7.2.
<b>5. Data extraction process</b> - State the name of the data extraction process, including the steps and order. For the following possible processes, describe whenever used.	
5.1 Input image(s)	The input image(s) for data extraction. If processing has been performed, the output image of processing should be provided using a Uniform Resource Identifier (URI) or a Digital Object Identifier (DOI), otherwise a reference to the image specified in Section 1 can be provided. The input image set can include auxillary/ synthetic images where they are essential to the result (such as fusion images) and they are available for export from the software
<b>5.2 Feature detection</b> - Describe the process by which the features have been assigned on the image(s). In terms of algorithm name, version and feature editing (automatic or manual)	
5.2.1 Algorithm name/version	The name and version of the algorithm as part of the specified software – (if available) or default algorithm of software.
5.2.2 Parameters	Parameters used in applying the algorithm (for example smooth, min area, saliency, etc.).
5.2.3 Feature editing	Specify if feature editing has been performed and if so, whether feature editing followed a manual or user-guided automatic process.
<b>5.3 Matching</b> - Describe the process by which the features have been matched on the image(s). In terms of algorithm name, version, parameters and match editing	
5.3.1 Algorithm name/version	The name and version of the algorithm as part of the specified software (if available).

5.3.2 Parameters	Parameters or arguments used in applying the algorithm (if appropriate).
5.3.3 Reference image used	Name and identifier of the gel image used as a reference for the matching.
5.3.4 Landmarks	Describe if landmarks have been manually and/or automatically assigned.
5.3.5 Match editing	Specify if manual and/or automatic match editing has been performed by a user-guided process.
<b>5.4 Feature quantitation</b> – Describe the process by which the features detected have been quantified. In terms of measurement used, quantitation method, background subtraction (if used), normalization method.	
5.4.1 Type	Describe the measurement used (for example spot volume -normalized or not, peak volume, peak height).
5.4.2 Quantitation	Define the method used for quantitation including data sources and/or parameters.
5.4.3 Background subtraction	State whether background subtraction has been performed. Define the method used for background subtraction including data sources and/or parameters (for example lowest on boundary).
5.4.4 Normalization	State whether normalization has been performed on the quantified values. Define the method used for normalization (for example percentage of total feature volume, ratiometric) that produced the normalized quantities given in the features, including data sources and/or parameters.
<b>6 Data analysis</b> – analysis protocol used, repeat the steps as necessary, that created the data reported in Section 7.	
6.1 Analysis intent	Describe the analysis intent (for example “finding the features with a statistical significant differential expression”).
6.2 Input data	Source data used as input to the method (for example intensity, area, volume normalized or not, etc.), provide a description of any selection or filtering procedures used on the outputs of Section 5.

6.3 Data transformation	Describe any transformations performed on the data (with parameters where appropriate) prior to the analysis e.g. log transformation, median centering.
6.4 Software	Software used for the data analysis, if different from the software defined in Section 1.6. Software name and version number, vendor (or if not available, provide a literature reference or a URI).
6.5 Type	State the data analysis method used (such as student t-test, heuristic clustering, PCA, determination of the false-discovery rate).
6.6 Parameters	Parameters used in applying the method (blind or not, number of classes, method for handling missing values and/or outliers, etc.) if applicable
<b>7 Data reporting</b>	
7.1 List of image features	List of all the quantified values and location of image features. This may be provided by a reference, such as a URI or in a file. The outlines of features should be provided in a suitable format (e.g. boundary chain, image map, regular shape etc) if spot outline detection has been performed by a non-proprietary algorithm. If proteins have been identified, annotations to features should be provided with references out to external resources, for example compliant with MIAPE Mass Spectrometry Informatics.
7.2 List of matches	List of all the matched feature relationships if feature matching has been performed. This may be provided by a reference, such as URI or in a file.
7.3 Analysis results	List the results of each Data Analysis recorded in Section 6, such as the set of features that displays significant differential expression. This may be provided by a reference, such as URI or in a file.



## References

- [1] Taylor CF, et al. The minimum information about a proteomics experiment (MIAPE). Nat Biotechnol. 2007 Aug;25(8):887-93. PMID: 17687369
- [2] Gibson F, et al. MIAPE Gel Electrophoresis. <http://psidev.info/miape/ge>