

LECTURE-7

Gel-based Proteomic Data Analysis

HANDOUT

PREAMBLE

The aim of gel-based proteomics experiments is identification of proteins spots, which are differentially expressed or unique under a given condition. After the gel is obtained using 2DE electrophoresis or DIGE, there may be some variations observed. These variations might be due to various reasons such as making of the gel, running of the gel, manual handling difference, scanning the gel, and differences and bias due to human individuality in analysis of the gel images. These differences may be biological or technical. Biological variations may be due to the different types of experimental sets conducted and technical variations may be due to the issues of running the gel or other experimental artifacts. Variations due to any kind may result in erroneous results. For identification of statistically significant spots, analysis of the obtained gels is very important. The gels are scanned and images are analyzed using image-processing software.

OUTLINE OF THE LECTURE

1. Artifacts in gels
2. Scanning of gels
3. Methods for analysis of a 2D gel
4. Analysis of DIGE gel

1. ARTIFACTS IN 2D GELS

There are many types of artifacts that may occur in gels, and results into bad gels, which can't be used for analysis. Some of the commonly occurring artifacts in 2D gels are as follows:

Poor sample preparation

At times, even though the gel run is smooth, artifacts may appear in gels. The cause for this might be poor sample preparation, which includes:

- Poor solubilization of protein in buffer
- Impurities due to running of crude protein sample
- Improper washing during protein extraction steps
- Presence of salts in protein sample which interferes with IEF
- Highly abundant proteins – like IgG, in samples can mask the lower abundant proteins

Horizontal Streaking

Horizontal streaking can occur as a result of different types of erroneous procedures.

Some of the reasons for streaking on gels are as follows:

- Poor sample preparation
- Presence of interfering agents such as carbohydrates or nucleic acid
- Presence of salts
- Due to incomplete resolution of the spots on gel

Chemical impurities

Many artifacts arise in gels while preparation of the gel. This may occur due to poor quality of the chemicals used for preparation of the gels.

- Impurities in acrylamide, bis-acrylamide
- Usage of old TEMED – as very less amount of TEMED is required per gel, thus the same stock may be used for a long time
- Presence of impurities in salts like Tris
- Impurities in urea – carbamylation train in gels

Improper gel running

At times, if there is no discrepancy with the protein sample preparation, or with the chemicals used to prepare the gels, issues may arise due to improper running of the gels. For any good 2D gel, proper IEF and second dimension is required. Some artifacts on gels may arise due to:

- Improper Isoelectric Focusing – it is important to have the correct program for focusing of the sample – this may lead to under-focusing or over-focusing, which may reflect on the quality of the final gel.
- Incorrect equilibration step – either with DTT or with IAA, which results in artifacts on the final gel.
- Improper gel running in second dimension.

2. Scanning of the Gel

For further analysis of the gels, it is necessary to capture an image of the gel spots. For capturing the gel image, the following steps are taken:

- The gel is laid carefully on the gel image scanner.
- Since the size of the gels is large, care has to be taken to handle the gel carefully while transferring it from the destaining solution to the scanner platform. The gel may break during this process, which results in a black colored line across the image, thus reducing the gel image quality. This discrepancy may affect the spots detected, and affect the gel image analysis.
- Another precaution that needs to be taken before capturing the image is that the gel has to be laid on the platform such that there are no air bubbles between the gel and the surface. This leads to poor image quality. This issue can be resolved by rolling a wet glass rod (previously dipped in destaining solution), and removing any air bubbles.
- The gel is then scanned using software. Image is previewed, and the colors are adjusted. The image is in gray scale, and can be adjusted for optimum intensity of all the spots.

3. Methods for analysis of a 2D gel

After capturing the image on the scanner, image analysis is done for interpretation of the results of the experiment. The ideal method is to use software for automated analysis but also validate the results manually. Thus, the results shown by the software must be verified by the user to confirm that spots are real and not artifacts. There are many software available to analyze 2D gels. The table-1 provides a list of few software:

Table-1. Software used for gel image analysis of traditional 2D gels.

2-DE software	Website
Image Master 2D Platinum	http://gelifesciences.com
PDQuest	http://discover.biorad.com
Delta 2-D	http://www.decodon.com
Dymension	http://www.syngene.com
Ludesi 2-D gel image analysis	http://www.ludesi.com
Progenesis	http://www.nonlinear.com

Illustration: 2D Gel Analysis

- *Following 2-D gel electrophoresis, the gels are scanned using a suitable scanner and the images obtained are analyzed by means of various available software. These allow interpretation of data present in the form of spots on the gel images.*
- *It is possible to load either a single or multiple gel images simultaneously. This is done by means of the 'load' option in the file menu. The saved gel images must be chosen, which are then displayed on the software. Several tools are available for analysis of the gels. It is possible to crop the gels by selecting a specific region that is to be studied and then selecting the 'crop gels' function. Cropping gels helps in selection of regions with high spot density or in trimming of regions with high background staining but no spots.*
- *Specific selected regions of the gel can be zoomed for viewing the spots more closely and for comparison of spots between two gels. This is particularly useful for gels having large number of spots. Overlaying of images is a particularly useful tool for comparison of two gels. The gels are*

overlaid such that they appear merged and spots that coincide will overlap with each other. This is extremely helpful while comparing clinical samples of control and treatment, providing clear indication of the proteins that are differentially expressed.

- The spots on the gels can be displayed as three-dimensional graph. Either the entire gel can be chosen or a particular region can be selected for this representation. The peaks obtained in the graphical representation are directly related to the spot intensity. Every spot on the gel can be detected by selecting the 'detect spots' option. Parameters such as 'smoothness', 'saliency' and 'min area' must be suitably adjusted for maximum clarity. Once this is done, each spot will either be encircled or marked with a cross, depending on the settings, along with the spot numbers.
- The software facilitates interpretation of the gel images by matching two different gel images obtained. The matching spots are marked and any variations in the spot position are indicated by blue lines in one of the gels. This provides an understanding about the reproducibility across gels. Information regarding the various physical parameters of each spot can be obtained via the spot table. This table provides the spot number, intensity, area and volume of spots as well as the saliency of spots. These parameters help in judging the quality of a gel.
- In addition to physical parameters, several statistical parameters can also be computed for each spot on the gel such as central tendency, mean, median, dispersion, coefficient of variation, standard deviation etc. Scatter plots and histograms can also be plotted for clear data analysis. These provide information regarding intra and inter-gel variations. It is also possible to specifically compare a particular selected spot across gels. When the gel is run with molecular weight markers, the weight of unknown proteins can be estimated from these. It is also possible to estimate the pI of the proteins. These parameters, in addition to other physical & statistical parameters, can be obtained for each spot.
- Software for DIGE analysis varies with respect to certain features compared to normal 2-DE analysis. It can compare three gels simultaneously, of which one is typically the pooled internal standard containing all spots. Any changes implemented in one gel such as cropping, spot selection etc. will be implemented across all three gels in DIGE. Other features and tools for DIGE analysis are same as those used for 2-DE analysis. Physical and statistical parameters of all spots on the gels can be determined through their corresponding reports. Three-dimensional graphical representation provides information regarding intensity of spots on the gel.

There are specific protocols to be followed depending on the software being used. The protocol followed for image analysis by using the Image Master 2D Platinum (IMP7) software is as follows:

STAGE 1:

1. The scanned images which are saved in the .mel or .tiff format are first imported to the image pool of the IMP7 image analysis software from the source folder.
2. It is important to keep in mind that all the images, which are to be analyzed are of the same dimensions and the contrast & brightness conditions are also same. This should be done prior to importing the images.
3. Within the IMP7 software, a new project is created with the required label.
4. A new workspace is then created in the project and all the images (control as well as treated; for all the reps under study) are considered in the workspace for analysis.
5. Now a match-set, which comprises of all the gel images (control as well as treated; for all the reps under study), is created.
6. This is followed by manual detection of spots; wherein the spot boundaries are individually marked in each of the gel images. The contrast and zooming features can be adjusted for better resolution of spot boundaries.
7. The spots can also be detected automatically by the software.

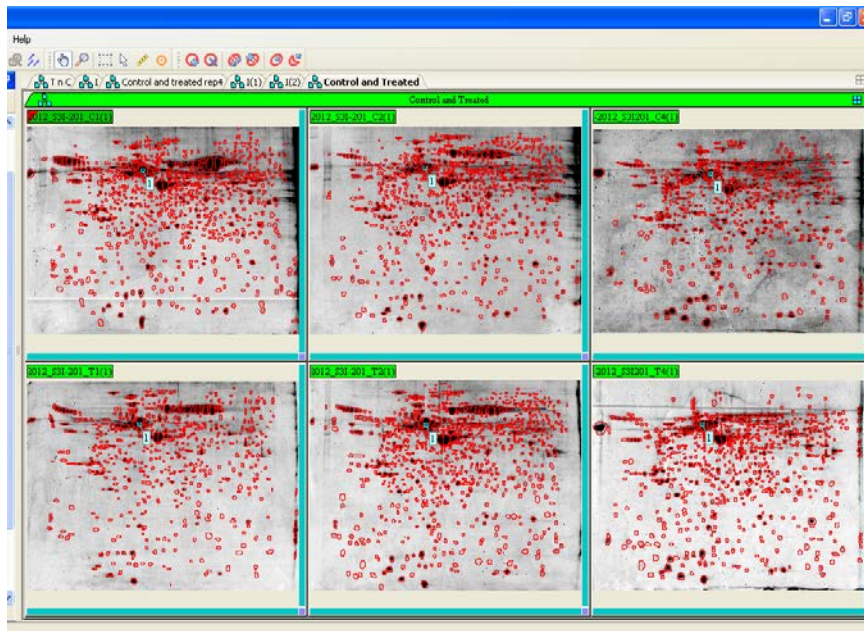


Fig 1. Representation of the software interface for setting landmarks

STAGE 2:

1. Once all the spots are manually detected in each of the gels, the next step of matching gels can be carried out. Ensure that the spots that are detected are approximately equal in number in each of the gels.
2. Create Match classes and then include gel images in the respective classes of 'Control' and 'Treated'.
3. Display images belonging to one of the two classes. For instance, take up 'Control' class. It has the gel images with spots that were manually detected. The first gel is taken as the 'Reference gel', when the individual gels are being matched amongst themselves.
4. Alternatively, all the gels can be considered in a single match class and then be compared. The first gel will be taken as 'Reference gel', by default. This can be

changed by simply dragging the required gel in the first position in the order of the created class.

5. Define landmarks (i.e. spots that are prominent and common to each of the gels that are being compared). Ensure that the entire gel area is covered when you are defining landmarks. Optimally 4-5 landmarks are sufficient for analysis.
6. The gels are now ready to be matched within the said class.
7. Click on 'Match gels' icon on the toolbar. It will now display the matched gels and the number of total matches that are identified. The Match count includes the spots, which are present in all the gels under consideration.

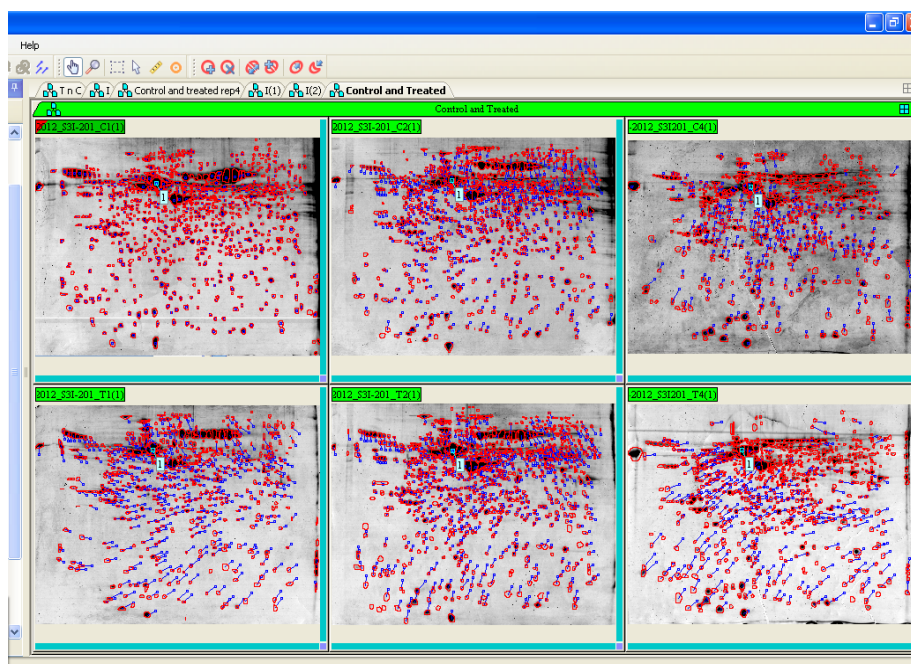


Fig 2. Matching of gels

8. From the Match Analysis Table, the ANOVA values for each of the matches can be obtained. The matches with ANOVA value <0.05 are considered to be significant. These spots are then checked if they are matched properly or not.

The software may miss out pre-marked spots and hence there might be loss of significant spots. In that case manually checking of the 3D views of the spots becomes essential.

9. Subsequently, if the spots are missed out in any of the gels, it could be compared and marked. Again, statistical analysis has to be performed.
10. Then calculate the ratio, which is avg. of treated intensities/ avg. of control intensities. Correspondingly, note down the trends of spots i.e. whether down-regulated or up-regulated.

DIGE

Illustration: 2D-DIGE Gel Scanning

- *After running 2D-DIGE, the scanner uses the property of dye used for labeling the sample to produce a protein profile image. The dye when exposed to different wavelength, due to excitation and emission property of the dyes the images can be captured.*
- *When the instrument is ON, it will go for automatic calibration to check all the connections. Take out the cassette for scanning, cassette provides the platform for gels to scan. A hollow cassette is used to scan large format gels in the glass plates. For small gels, cassette with glass platform is used. The cassette needs to be placed inside the scanner. Proper fit is necessary for better scanning. User needs to set the parameters for scanning and specify the scanning area depending on the size of the gel. For gel format: predefined gel format or user can define the gel format in user select.*
- *Selection of the Channel check boxes must be performed in sequential numerical order. The excitation and emission wavelength helps to produce optimum results.*
- *Cy2: Ex/Em:480/520 for 0.35exposure.*
- *Cy3: Ex/Em:540/595 for 0.90 exposure.*
- *Cy5: Ex/Em:635/680 for 0.70 exposure.*
- *User can set the exposure time depending upon sample quantity used.*
- *Once all the settings are done, user can go for scanning. Before scanning the software asks to save the image. Provide the location under the folder and name the gel properly.*
- *During scanning the image get displayed row by row for each channel. In case if image shows some red spots in the protein region, stop the scan and re-scan with less exposure time.The display gives the scanning report.*
- *Cropping and image editing depends on user requirements and just for appearance purpose user can make the changes. Cropping helps to remove unwanted area from analysis.*
- *In case the image gel plate is kept on “-” side, while scanning the gels can be kept either within the glass plate or it can be performed after removing it from the plates. Software provides option for user to rotate the image and later on save option. Now the gel image is ready to be uploaded and analyzed by software tools.*

Illustration: 2D-DIGE Gel Analysis

After running 2D-DIGE the protein profile obtained need to be analyzed for the protein spot of interest, showing significant fold change and appearing in most of the samples. As demonstrated in this animation,

NPTEL WEB COURSE – ADVANCED CLINICAL PROTEOMICS

the Decyder software helps to minimize system variation and enables to identify significant spots of biological importance.

4. DIGE Analysis using DeCyder

The traditional 2D gels are stained using dyes like CBB; however, the 2D-DIGE utilizes labeling of proteins using the fluorescent Cyanine dyes. There are 3 types of Cy dye – Cy3, Cy5 and Cy2, which are used for labeling the control, test and internal control samples. For this, control, test and internal control samples are run in the same gel. Images of the gel are captured using Scanner and 2D-DIGE gel analysis can be performed using the DeCyder software.

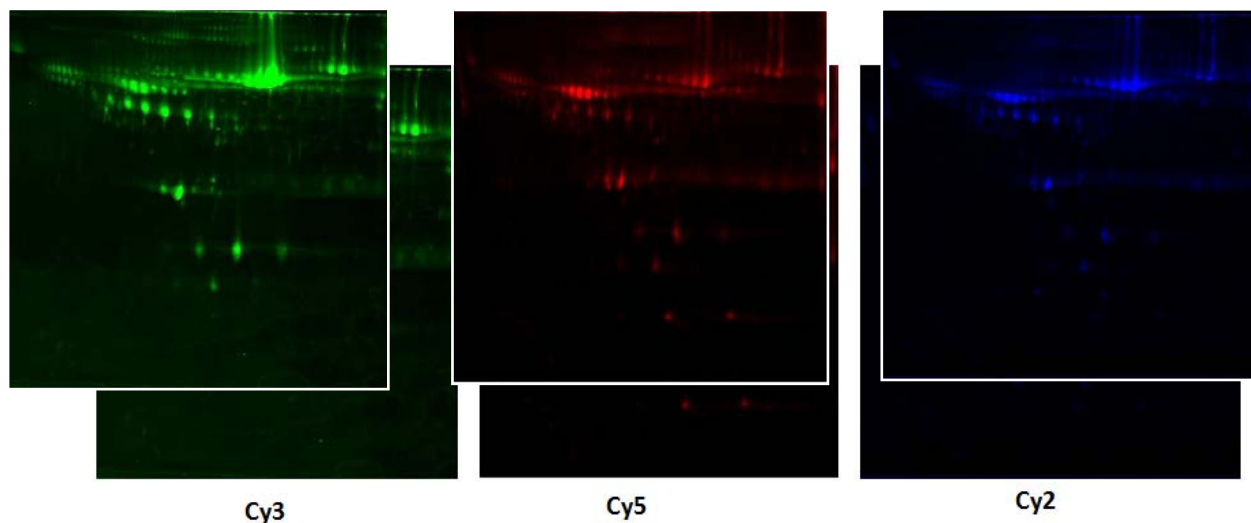


Fig. 3. Scanned images for DIGE of Cy3, Cy5 and Cy 2 respectively.

Steps for using the Decydersoftware for analysis of gels are as follows:

Image Loading

The first step in DIGE analysis is loading the relevant images in the software. For this, the Image Loader option is selected. The .tiff files and .GEL files obtained from Typhoon scanner are used for acquiring the images into the software. All 3 images – Cy2, Cy3,

Cy5 – for one gel are selected and loaded. This image loading can be done for as many biological replicates as needed for the analysis.

Image Editing

All the imported gel images are selected and the 'Edit Gel Images' tab is selected. This view shows the Cy2 images i.e. Standards of all the gels. The Cy2 images can be cropped and rotated as required in this window. The changes made to Cy2 gels are extrapolated to the Cy3 and Cy5 gels as well i.e. other gels are cropped using the same dimensions. All the changes are saved, and the window is closed.

Image Import

The New Project icon at the right corner of the screen is clicked, and new file is created and named. The cropped images are imported in this folder.

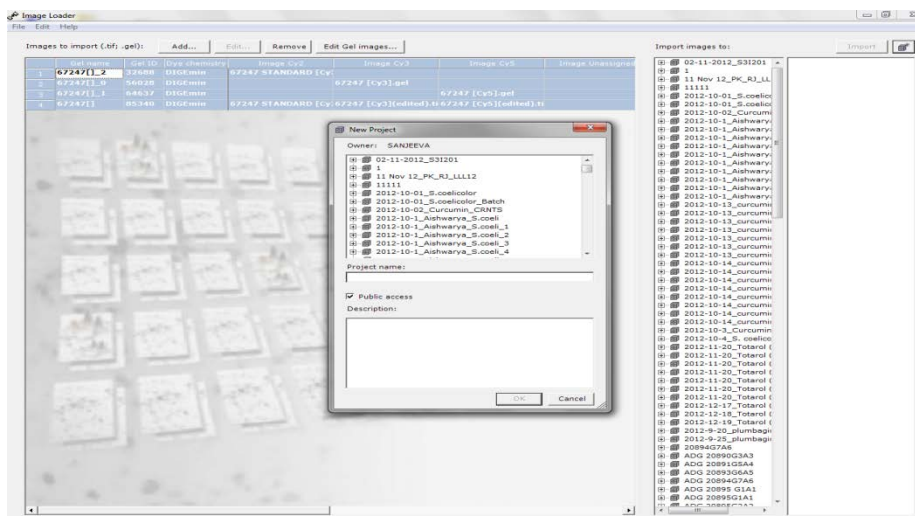


Fig 4. Representation of Image Loader in DeCyder

Image Analysis

Differential In-Gel Analysis (DIA)

For analysis of the Cy2, Cy3 and Cy5 images of one set the DIA is used. For this, a new DIA Workspace is created, and relevant images are imported into the workspace. The estimated numbers of spots are added, and the images are processed. After processing, the DIA gives 3D views of all the spots, and the Max Slope and Max Volume for all. The Max Slope for a real protein spot is between 1 to 1.5. Exclusion Filters can be set up to exclude certain areas because there is no distinction between real protein spots and dust particles.

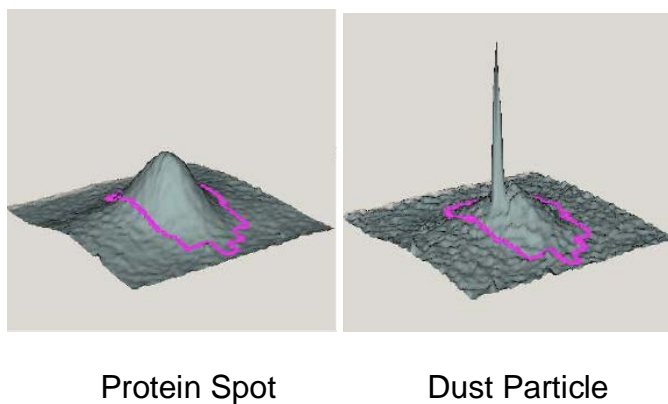


Fig. 5. Image of protein spot and dust particle in DIGE using DeCyder

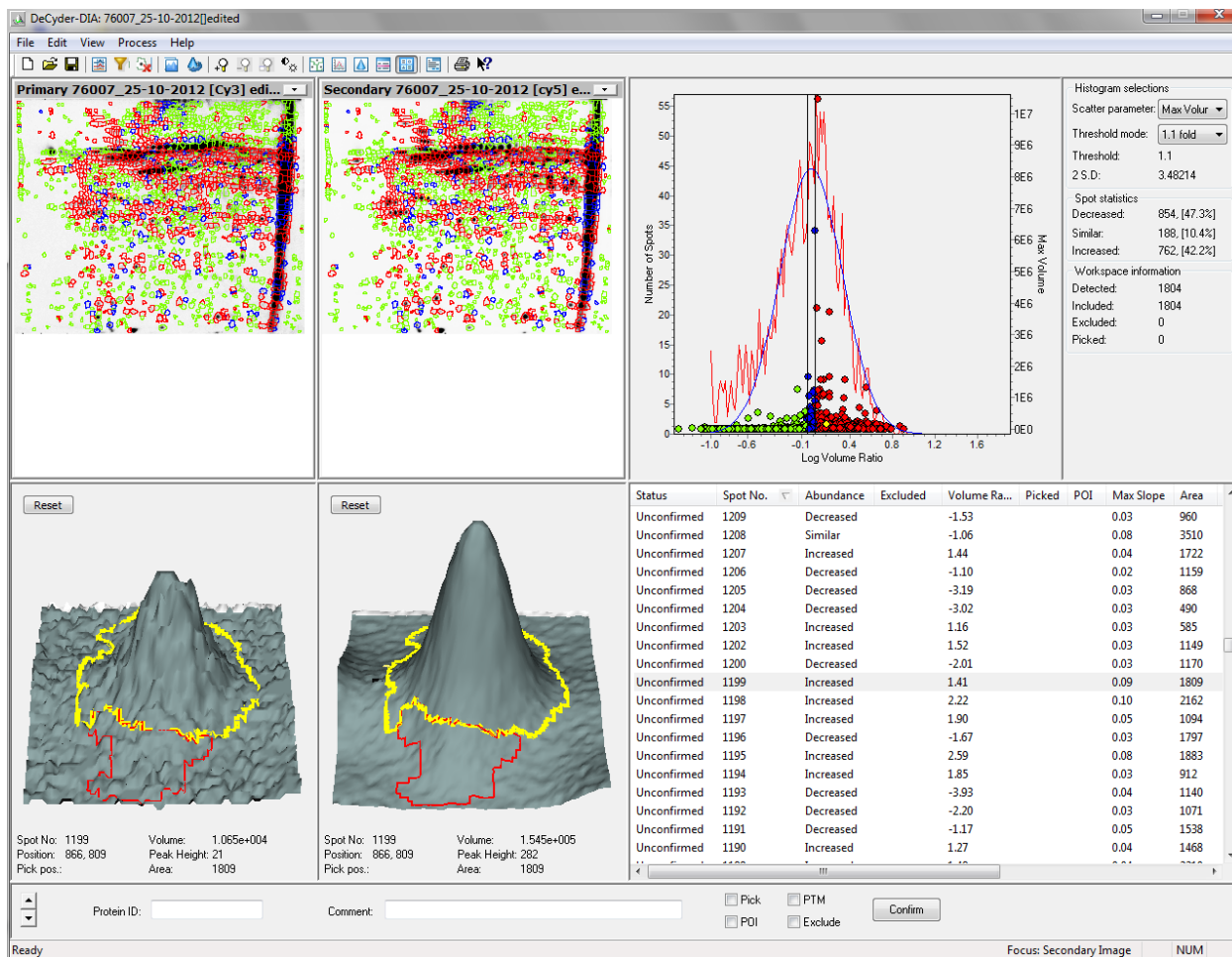


Fig 6. Representation of DIA in DeCyder

Biological Variation Analysis (BVA)

The BVA is used for analyzing the Cy2, Cy3 and Cy5 gels of more than one biological replicate, e.g. to compare different patients samples etc. The Batch Processor also helps in comparison of a large number of gels simultaneously. For this, all the gels to be analysed are selected. The DIA and BVA batch-list are created, and run. After this process, one of the Cy2 gel is selected as the Master gel, and other gels are matched against this gel. A new BVA Workspace is created. Two groups are made – control and test, and the gels are sorted appropriately.

BVA has 4 modes:

- Spot Map Mode

Only the standard gels are shown in this mode.

- Match Mode

All the gels can be matched in this mode. It provides an option for adding new matches, creating new spots, merging spots and splitting the spots, which are demarcated by the software. Then, a match-set can be created which incorporates all the changes. This process helps in accurate matching of the gels; however, verification should be done manually. This tool is especially helpful for marking those spots, which are actually preset but have not been detected by the software.

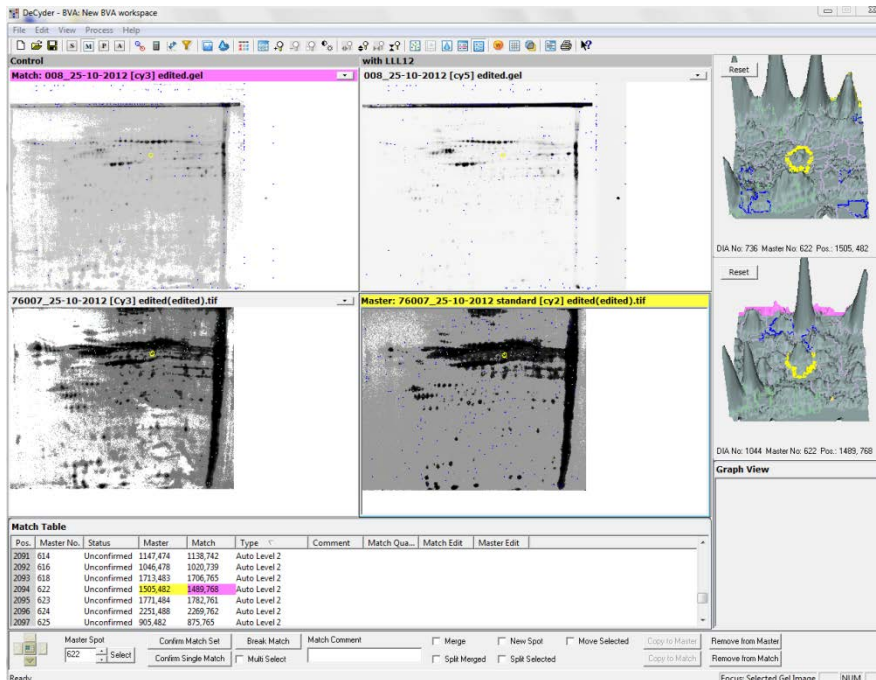


Fig 7. Representation of the Match Mode in BVA Workspace in DeCyder

- Protein Mode

The protein mode shows the statistical values for all the gels. It also shows the number of gels in which that spot has appeared. The values of the selected protein parameters: Student's T-test, One-way ANOVA, Two-way ANOVA and the fold change are displayed for each and every spot, along with the spot number on the Master gel.

- Appearance Mode

Appearance Mode gives the comparison of peak heights etc. for every component.

Pos.	Master ...	Status	Protein...	Appearan...	T-test	Av. Ratio	1-ANO...
1	576	Unconfirmed	9 (9)	0.0020	-2.47	0.0020	
2	1153	Unconfirmed	9 (9)	0.0023	1.42	0.0023	
3	1169	Unconfirmed	9 (9)	0.0098	1.87	0.0098	
4	536	Unconfirmed	9 (9)	0.011	3.73	0.011	
5	424	Unconfirmed	9 (9)	0.012	-1.42	0.012	
6	813	Unconfirmed	9 (9)	0.018	4.01	0.018	
7	642	Unconfirmed	9 (9)	0.026	2.39	0.026	
8	486	Unconfirmed	9 (9)	0.027	-2.08	0.027	
9	859	Unconfirmed	9 (9)	0.027	1.93	0.027	
10	1038	Unconfirmed	9 (9)	0.027	-1.59	0.027	
11	887	Unconfirmed	9 (9)	0.033	1.97	0.033	
12	1293	Unconfirmed	9 (9)	0.033	-2.31	0.033	
13	1266	Unconfirmed	9 (9)	0.039	-2.38	0.039	
14	566	Unconfirmed	9 (9)	0.041	1.55	0.041	
15	515	Unconfirmed	9 (9)	0.045	-2.12	0.045	
16	1488	Unconfirmed	9 (9)	0.052	1.55	0.052	
17	326	Unconfirmed	9 (9)	0.053	-1.92	0.053	
18	488	Unconfirmed	9 (9)	0.062	-2.38	0.062	
19	659	Unconfirmed	9 (9)	0.063	1.75	0.063	
20	647	Unconfirmed	9 (9)	0.064	4.10	0.064	

Fig. 8. Table seen in Protein Mode

- **3D View**

The 3D view for each spot can be seen, and compared with one another. This tool is especially helpful to compare the spots chosen by the software, and to detect spots where improper matching has been done. This mode is also useful to detect the authenticity of a spot – whether it is a protein spot or a dust particle.

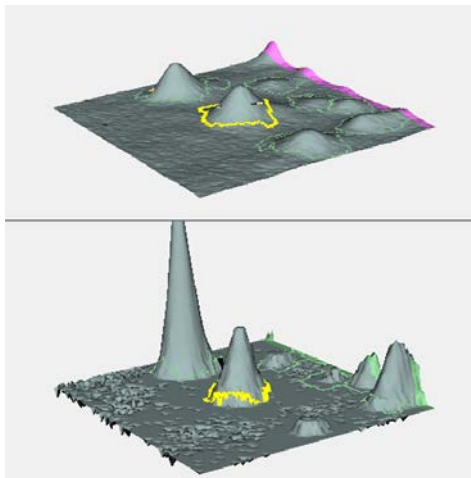


Fig. 9. Representation of 3D view of a spot

- Graph View

The Graph View shows the relative abundance of the protein in each gel. This is done for individual spots. The graph view plots the values of the control samples, and their average, as well as the experimental group and the average. The graph view also enables us to know whether a particular spot is up-regulated or down-regulated.

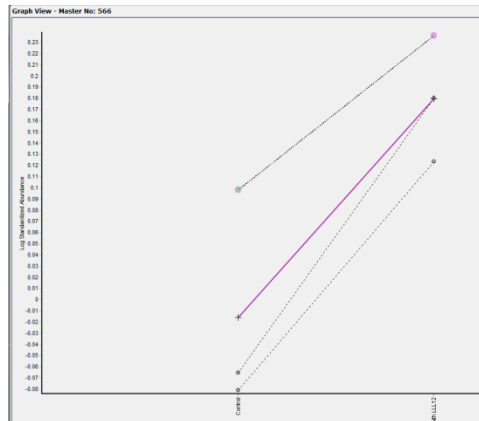


Fig. 10. Graph View

Extended Data Analysis (EDA)

The EDA can be used to compare the data obtained from the BVA analysis, which includes all the individual DIA analysis. Thus, data analysis can be done on a larger scale using EDA. EDA can be used to compare different datasets. DIA can be used to find out if any groups or clusters are present in the data. We can also find out whether there are proteins, which follow a pattern, or behave in a certain way. This may help to identify protein spots, which behave similarly as other spots across gels, and may give information for the trends of those proteins.

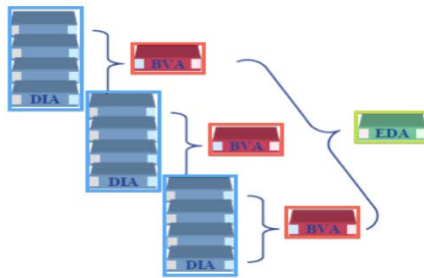


Fig. 11. Representation for EDA (Figure modified from - DIGE applications and Data Analysis, GE Manual)

Extended Data Analysis may be applied to 4 types of data :

(a) Principal components analysis:

The Principal Component Analysis is used to find out whether there are outliers in the data. It is also used to find whether protein expression is uniform in multiple sample from same experimental group.

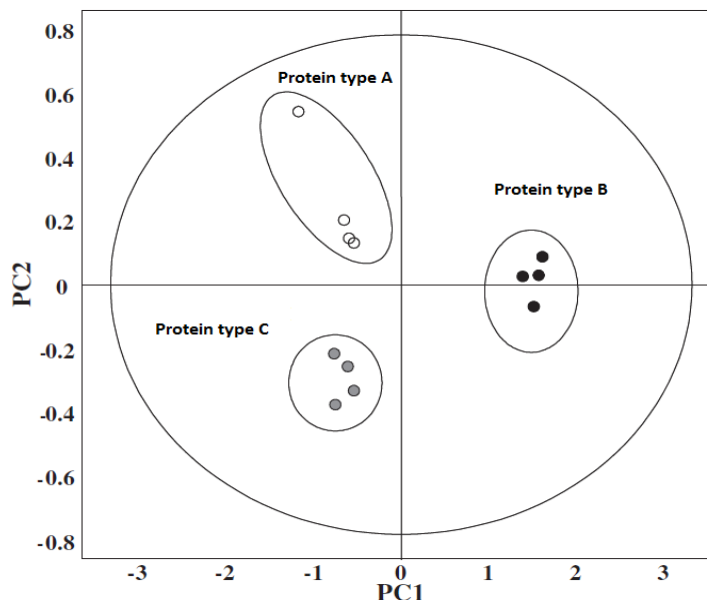


Fig. 12. PCA analysis on the basis of protein clusters

(b) Differential expression analysis:

This is done using statistical tests like Student's T-test or One-way ANOVA derived from the DIA data.

(c) Pattern analysis:

Pattern analysis enables clustering of the data to find patterns in the expression profiles in EDA data. For this, no prior information about the variables used is required. Items with similar expression profiles get clustered like as proteins, spot maps, and experimental groups. The results from Pattern Analysis are represented as a Hierarchical Cluster displayed as a heat map with dendrogram, showing different classes in the data set.

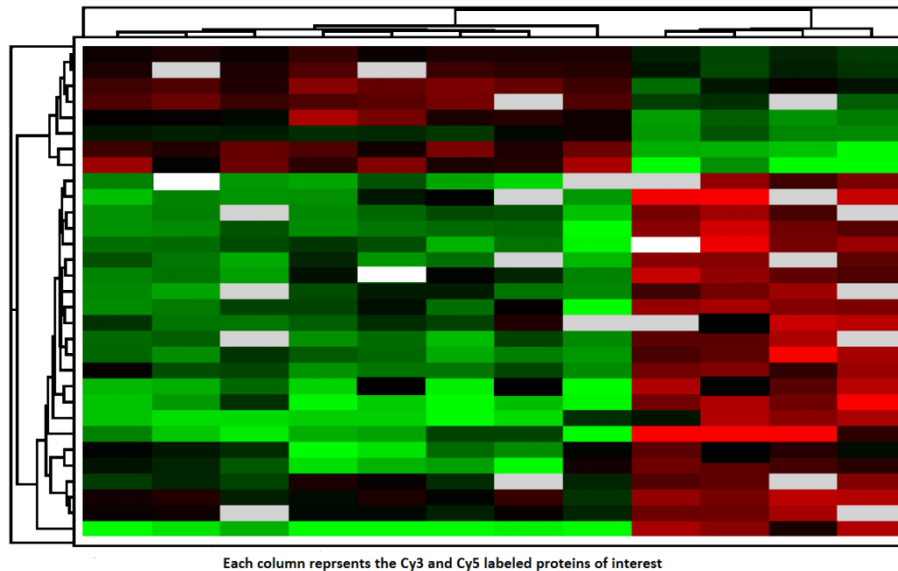


Fig. 13. Pattern Analysis in EDA for Heirarchical Clustering

(d) Discriminant analysis

The Discriminant Analysis is used for classifying the unknown data into known classes. This helps in identification of the proteins that may be indicators of a particular disease i.e. this analysis may help in identification of biomarkers for diseases.

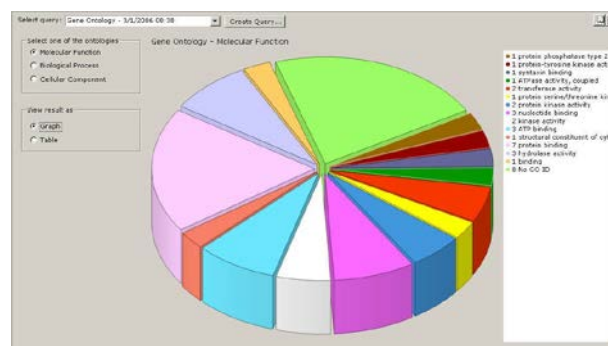
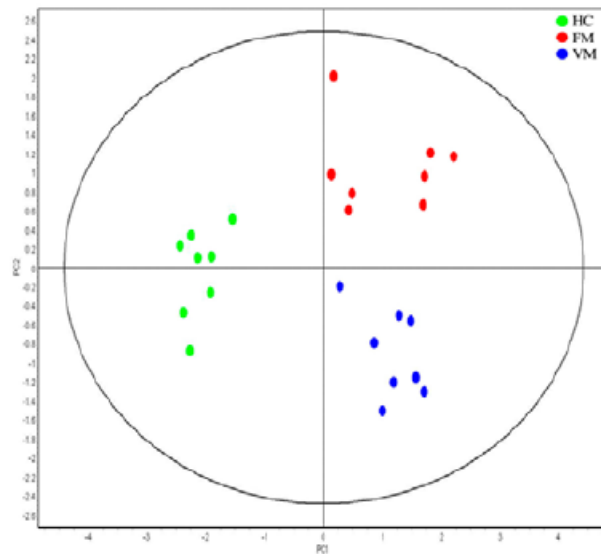


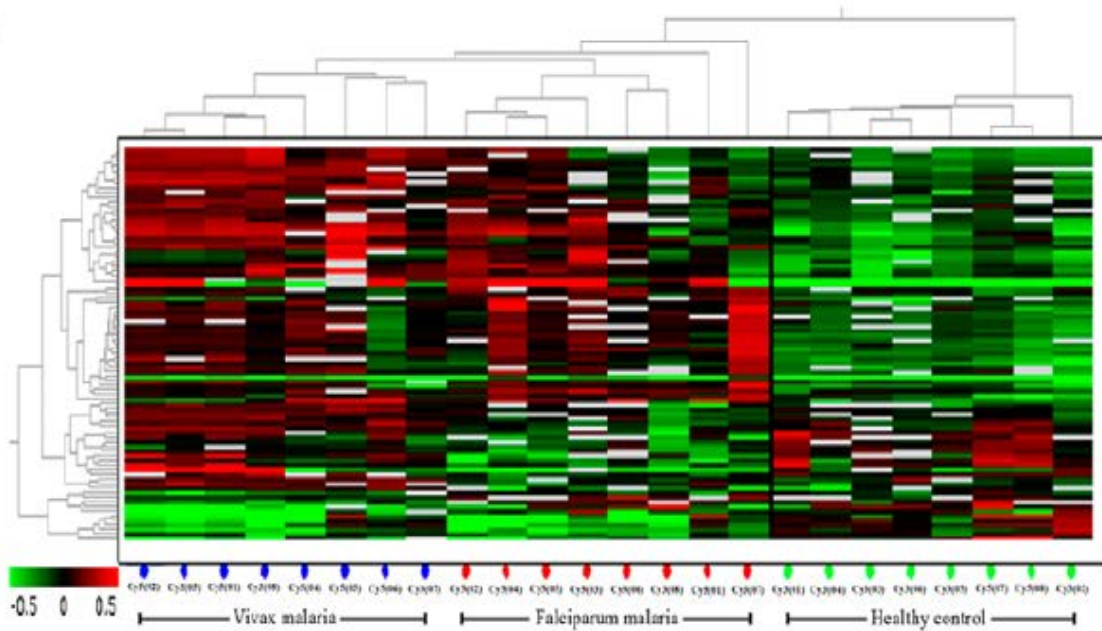
Fig. 14. EDA – Discriminant Analysis

Representation of data using EDA

A



B



C

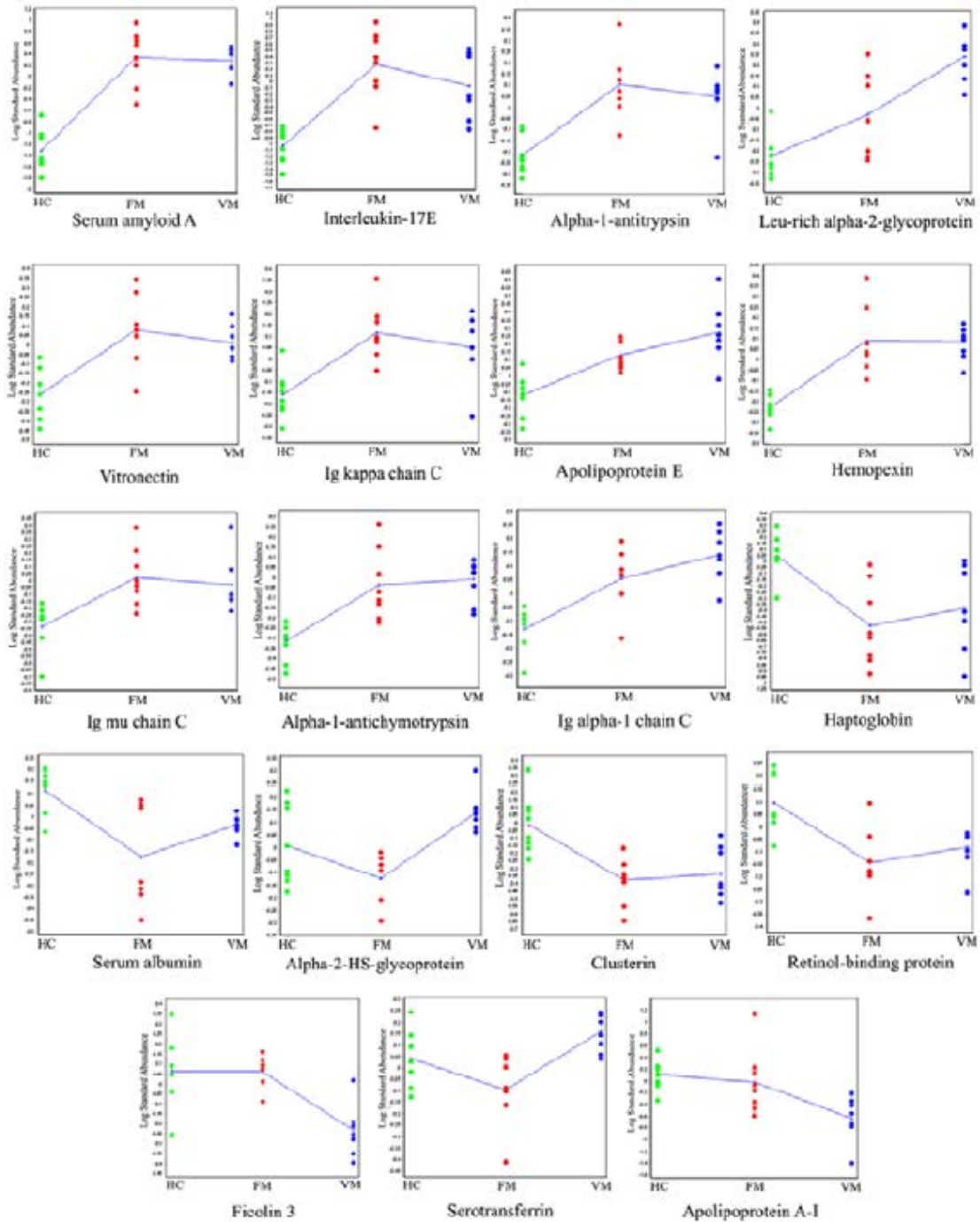


Fig. 15. Data analyzed from EDA. The EDA analysis was conducted for control (healthy subject), against two tests – two different types of infectious diseases. (A) PCA analysis - clustering of different phenotypical classes. The PC1 component separates the control

group from the other set, while the PC2 clusters the diseased groups separately.(B) Dendrogram showing the separation of different experimental groups after the hierarchical cluster analysis (Red - represents up-regulated; Green - represents down-regulated; Black - represents spots with no significant change in expression level). (C) Differentially expressed proteins in malaria patients represented as standardized log abundance of spot intensity.

References:

- Tieqiaozhanget.al., Time-resolved excited state dynamics of cyanine dye, chemical physics letters: 2008, 236-240
- Rita Marouga., Stephen David., Edward Hawkins., et al. Analytical and Bioanalytical Chemistry. 2005, 382, 669-678.
- Chen JH., Chang YW., Yao CW., et al., PNAS. 2004, 7,17039-17044.
- Maldonado AM.,Echevarría-Zomeño S., Jean-Baptiste S., et al., Proteomics. 2008, 71,461-472.
- GE Healthcare handbook: 2D-Electrophoresis, principles and methods.
- https://www.gelifesciences.com/gehcls_images/GELS/Related%20Content/Files/1314742967685/litdoc375-635_typhoon_userguide_20110831023220.pdf
- <http://www.med.unc.edu/sysprot/files/DeCyder%20Manual%20from%20GE.pdf>