Experiment-2:

Differential protein expression analysis

Objective: To study the difference between the protein expression profiles of normal and cancer tissue cells.

Gel-based high throughput comparison between the protein expression profiles of normal and diseased tissues can be done in either of the two ways, viz., 2-D gel electrophoresis or 2D-DIGE. Since both techniques are widely used in modern day proteomics, each of them is described in this experiment. To again complete knowledge of the experiment, it is important to understand each of the following sections. Hence, we recommend that the content of each section be read in the given order.

SECTION I: Differential protein expression analysis using 2-D gel electrophoresis

- Protein isolation: Total protein content of the cell is extracted and the sample is readied for performing a 2-D electrophoresis.
- Rehydration of IPG strips: The labeled sample is loaded onto the IPG strips.
- IEF: This is the first dimension wherein sample proteins are separated on the IPG strips on the basis of their isoelectric point.
- Equilibration of the strips: The IPG strip are saturated with the SDS buffer system required for the second-dimension separation.
- SDS-PAGE: This is the second dimension wherein the proteins are further separated based on their molecular weight.
Staining and Scanning of the gel: This provides the image needed to carry out analysis.

Software analysis: This helps to identify the differentially expressed protein spots.

SECTION II: **Differential protein expression analysis using 2D-DIGE**

Protein isolation: Total protein content of the cell is extracted

Sample preparation: Samples from the two different tissues are labeled with fluorescent dyes and are readied for performing 2D-DIGE.

Rehydration of IPG strips: The labeled sample is loaded onto the IPG strips.

IEF: This is the first dimension wherein sample proteins are separated on the IPG strips on the basis of their isoelectric point.

Equilibration of the strips: The IPG strip are saturated with the SDS buffer system required for the second-dimension separation.

SDS-PAGE: This is the second dimension wherein the proteins are further separated based on their molecular weight.

Staining and Scanning of the gel: This provides the image needed to carry out analysis.

Software analysis: This helps to identify the differentially expressed protein spots.

After the details of the technique are understood, the reader is encouraged to go through the stimulations, protocols and manuals to get better insight of the process.

---

**Theory:**
Proteomics is defined as the simultaneous analysis of all the complex proteins expressed at a given time-point in a particular cell type, tissue extract or plasma/serum blood sample. It tries to unravel the mystery behind many unresolved complications that cannot be explained by genomic information alone. Since proteins are the ultimate effector molecules in any cell, a study of its entire protein complement could provide important insights into various biological functions.

Amongst a number of high-throughput proteomic techniques used to resolve the complex protein mixtures, Two Dimensional (2-D) gel electrophoresis is the most preferred method. With the help of this technique, protein separation is carried out based on two different properties of proteins namely, isoelectric point and molecular weight. It is a simple yet powerful tool to study the high-throughput protein expression profiles as a function of biological variations resulting from difference in sample conditions, like disease state or drug treatment. Visualization of highly resolved proteins on 2-D gels followed by mass spectrometry (MS) identification is a routinely used practice for comparative proteome analysis. An advanced version of this technique called Two Dimensional Difference Gel Electrophoresis (2D-DIGE) gives a completely new dimension to protein analysis. The main advantage of this technique is that it enables the researcher to separate up to three different samples on a single gel, eliminating gel-to-gel variations. This significantly increases the accuracy and reproducibility of the results.

Both, the 2-D gel electrophoresis and the 2D-DIGE, can be used to study the differences in the protein expression profile of a cell or organism under varying conditions. Details of the 2-D gel electrophoresis and 2D-DIGE are explained in the following sections.

**SECTION I : Differential protein expression analysis using 2-D gel electrophoresis**
1. **Protein isolation:**

Cancer is caused due to the accumulation of gene aberrations that are transcribed and translated with many modifications, which finally result in transformation of a normal cell to a tumor cell. Though genome and transcriptome studies have increased our understanding of cancer biology, the discord between mRNA and protein expression level indicates the need for proteomic studies. Aberrant posttranslational modifications of proteins are known to regulate cancer phenotypes. Therefore, they provide a rich source for biomarker identification and subsequent development. The identification of such biomarkers that are present in specific cancer tissues can lead to early detection and successful treatment of multiple cancer types. Modern proteomics, by its ability to detect dynamic changes in protein expression, localization and modification, has become a powerful tool to map differential protein expression profiles and deliver the functional information that will promote insights in the biology of cancerous cells.

With this aspect in mind, normal and tumor tissue samples are procured from various approved clinical sources. Samples from tumors at different stages of growth are also collected to investigate the progression of cancer in that tissue. After surgical resection, these samples are snap frozen using liquid nitrogen and immediately stored at -80°C. For isolation of proteins, the tissue is washed thoroughly with PBS buffer and then sonicated to disrupt the cells. Cell debris is removed by centrifugation. Proteins are then precipitated from the supernatant, using the Trizol method for protein isolation, and quantified by the Bradford method as described earlier. The reader is advised to look up the details of these methods and the following subsections in the relevant sections of Experiment 1.
2. **Rehydration of the strips:**

IPG strips provide an immobilized pH gradient for separation of proteins during the first dimension of 2-D gel electrophoresis, i.e. IEF. They are rehydrated with sample protein solution along with the rehydration buffer before carrying out IEF. The volume of the rehydration solution depends on the length of the IPG strip being used. This rehydration solution is added to a well in the reswell tray and the IPG strip is placed on it such that the gel side is in contact with the solution. The well is then filled with a cover fluid which provides prefect condition for rehydration and avoids strip drying and sample precipitation. The rehydration process is carried out overnight for 10-20hrs.

3. **First dimension-IEF:**

In isoelectric focusing, proteins are separated on the basis of their isoelectric point (pI). In the presence of a pH gradient and under the influence of an electric field, a protein tends to move to a position in the gradient where it attains its pI. IPG strips provide such an immobilized pH gradient. The process is carried out under constant voltage conditions in a stepped manner, initially applying a low voltage to avoid protein aggregation and protein precipitation, followed by maximum voltage at which proteins get resolved properly from each other.

4. **Equilibration of the IPG strips:**

The equilibration step for focused IPG strip treatment is carried out before subjecting the strip to the second dimension of 2-D gel electrophoresis. The equilibration step is carried out in two steps for 15min each, initially by adding DTT and later by adding IAA to the equilibration solution.
5. **Second dimension - SDS-PAGE:**

The second dimension of 2-D gel electrophoresis is Sodium Dodecyl Sulphate - Polyacrylamide gel electrophoresis (SDS-PAGE). In this step, the proteins are separated under the applied electric field on the basis of their molecular mass/weight. The SDS masks the charge present on the proteins, giving roughly a constant net negative charge per unit mass to all the protein-SDS complexes. This ensures that the electrophoresis carried out in the SDS-PAGE gel is based only on the molecular weight of the protein.

6. **Staining and Scanning of 2-D electrophoresis gels:**

After the second dimension of 2-D electrophoresis is done, the separation patterns of the protein samples are visualized by staining the gels. This is done by exposing the gels to specific dyes which bind to proteins embed in the gels and help visualization of maximum number of protein spots. Coomassie brilliant blue, a commonly used dye, interacts with the proteins embedded in the gel by non-covalent forces like electrostatic and Van der Walls interactions. Dye that is not bound to the protein diffuses out of the gel during destaining step. The proteins then appear as blue spots or bands on a clear background. Gels are then subjected to a corresponding destaining step before they are scanned using a gel documentation instrument. A representative image is shown in Fig.1. Such images of the stained gels can then be used for comparison of the protein expression profile of normal and tumor tissue across different gels with the help of commercially available software.
**Fig. 1: 2-D electrophoresis gel image:** Image of a typical 2-D electrophoresis gel showing normal and tumor tissue proteins separated on a 4-7 pH range on the X axis and molecular weight on the Y axis.

7. **Software analysis:**

Scanned images of 2-D electrophoresis gels show thousands of spots, each spot representing a single protein or a group of protein isoforms, having a particular pI and molecular weight. With the help of commercially available software, each spot is defined by an outline which is automatically or manually drawn around it. The software then digitizes the image file into pixels. The sum of the intensities of all the pixels present within the defined region of a spot is recorded and co-related with the quantity of proteins present in each spot. Such an analysis gives a comprehensive output in statistical terms, which are easy to interpret and can be extended to a
wider biological scenario. One such software is ImageMaster 2D Platinum, a high-throughput 2-D imaging software for almost parameter-free spot detection. Although it contains several tools which need to be explored in detail, the basic steps for image analysis are as follows:

- A master folder is first created. Images of gels to be analyzed are imported, opened in the software and labeled appropriately.
- The cropping tool is used to select the region on the gel having maximum spot density and exclude the regions without spots from analysis.
- The spot picking tool selects the spots on the gel using certain user-defined criteria. The software then records the pixel intensities of the spots.
- A 3-D graphical representation of the spots can be seen using the 3-D view tool.
- Spot parameters such as volume, intensity, possible pI, molecular weight etc. can be obtained through the software.

The resulting data can be compiled together to understand the global expression profile of the particular sample. A schematic representation of the entire process of differential protein expression analysis by 2-D gel electrophoresis is represented in Fig. 2.
Fig. 2: 2-D gel electrophoresis and analysis: A schematic representation of all the processes involved in global expression analysis, starting from preparation of protein sample to analysis of 2-D gels is shown.
SECTION II: **Differential protein expression analysis using 2D-DIGE**

1. **Protein isolation:**

   Normal and tumor tissue samples, which are procured as described in the above section, can also be used to study the differential protein expression patterns using 2D-DIGE. Tissue samples are disrupted by sonication and the proteins are extracted from the whole cell lysates by the Trizol method for protein isolation. Once a pure preparation of protein, devoid of any impurities, is obtained, the exact concentration of proteins in the solution is determined by Bradford method.

   The differences in the protein expression profile of the normal and tumor tissue samples are then determined using 2D-DIGE. Though both 2-D gel electrophoresis and 2D-DIGE are routinely used in high throughput proteomic assays, 2D-DIGE has a distinct advantage. The characteristic features of 2D-DIGE are its high sensitivity and wide linear dynamic range in quantitative estimations. This helps to generate data of high accuracy and biological significance.

2. **Sample preparation:**

   The only difference between the two techniques, 2-D gel electrophoresis and 2D-DIGE, is that in 2D-DIGE, the samples are labeled with fluorescent dyes, usually Cy dyes, before rehydrating the IPG strips and carrying out the two dimensional electrophoresis.

   i. **Why is Cy dye used:** CyDye labeling provides significant added advantage over other staining techniques like silver and coomassie staining. Fluorescent detection offers a broad range of sensitivity for protein detection, with simple and robust, staining protocol and accurate quantitative reproducibility. CyDye DIGE Fluor minimal dyes consist of three dyes (Cy2, Cy3 and Cy5), having distinctly different excitation and emission spectra, that makes them perfect for multi-color detection (Fig. 3). The dyes are both pH insensitive and mass- and charge-matched, which ensures that identical proteins labeled with different dyes migrate to the same position during 2D-DIGE run. This particular property helps to resolve three different samples on a single gel. These fluorescent dyes can detect as little 25 pg of protein, and give a
linear response to protein concentration of up to five orders of magnitude. As against this, sensitivity of silver stain is down to 100 pg of protein while that of Coomassie blue is about 15 ng of protein. Altogether Cy dyes are size and charge-matched, pH insensitive, spectrally resolvable and highly sensitive with photostable property.

**Fig. 3: Chemistry of Cy dyes:** A schematic representation of chemical structure and the absorbance maxima of each dye is shown.

ii. **Cy dye labeling:** The CyDye DIGE Fluor minimal dyes contain N-hydroxysuccinimidyl ester (NHS) reactive group which specifically forms a covalent bond with the epsilon amino group of lysine residues by an amide linkage. The charge carried by dye is +1 at neutral or acidic pH, which replaces the intrinsic +1 charge of lysine, making sure the pI of protein is not altered. The concentration of the dye used during labeling is adjusted such that only 1-2% of lysine residues get labeled in each protein. The result is that each labeled protein carries only one dye molecule and can be visualized as a single protein spot on the gel. Using the CyDye DIGE Fluor minimal dyes, up to three samples can be resolved on a single 2D-DIGE gel. The three samples include two samples taken from different conditions and an internal standard, which is pool of these two samples (Fig 4). Dye labeling is done prior to running 2D-DIGE.
Fig 4: Chemistry of CyDye labeling: The NHS ester linker of the dye reacts with the lysine group of protein. The charge present on the lysine group is replaced by the dye, making it charge and mass-matched reaction. The reaction is optimum at pH 8.5.

Care should be taken to exclude primary amines and DTT from the sample before labeling. The protein concentration should be between 1-10mg/ml and pH of labeling reaction must be around 8.0 to 9.0 for efficient labeling. The reconstitution of CyDye and Labeling reaction must be carried out in dark room.

iii. Importance of internal standard: The major concern for 2-D electrophoresis is system variation, which arise due to different gel run conditions encountered during electrophoresis, user-to-user usage variation or in some case very low performance of scanning /analysis software. These variations can be checked by incorporating an internal standard within each gel run. The internal standard is a pool of two samples taken from different conditions and acts as reference gel image during the analysis (Fig. 5). The co-detection algorithm in the DeCyder 2-D software uses this reference gel image for spot detection, quantitation, positional matching and differential protein abundance analysis. This limits the experimental variations and makes sure accurate intra- and inter-gel matching is carried out.
**Fig. 5: Schematic representation of the 2D-DIGE experiment:** Protein samples labeled with different dyes are mixed together to run on a single IPG strip to carry out 1D-IEF followed by 2D SDS-PAGE. Later the gel obtained is scanned in a scanner to get three gel images from the same gel, one each of Cy2, Cy3 and Cy5.

3. **Rehydration of the strips:**

IPG strips provide an immobilized pH gradient for separation of proteins during the first dimension of 2-D gel electrophoresis, i.e. IEF. They are rehydrated with sample protein solution along with the rehydration buffer before carrying out IEF. The volume of the rehydration solution depends on the length of the IPG strip being used. This rehydration solution is added to a well in the reswell tray and the IPG strip is placed on it such that the gel side is in contact with the solution. The well is then filled with a cover fluid which provides perfect condition for rehydration and avoids strip drying and sample precipitation. The rehydration process is carried out overnight for 10-20hrs.

4. **First dimension-IEF:**

In isoelectric focusing, proteins are separated on the basis of their isoelectric point (pI). In the presence of a pH gradient and under the influence of an electric field, a protein tends to move to a position in the gradient where it attains its pI. IPG strips provide such an immobilized pH
gradient. The process is carried out under constant voltage conditions in a stepped manner, initially applying a low voltage to avoid protein aggregation and protein precipitation, followed by maximum voltage at which proteins get resolved properly from each other.

5. **Equilibration of the IPG strips:**
The equilibration step for focused IPG strip treatment is carried out before subjecting the strip to the second dimension of 2-D gel electrophoresis. The equilibration step is carried out in two steps for 15min each, initially by adding DTT and later by adding IAA to the equilibration solution.

6. **Second dimension - SDS-PAGE:**
The second dimension of 2-D gel electrophoresis is Sodium Dodecyl Sulphate - Polyacrylamide gel electrophoresis (SDS-PAGE). In this step, the proteins are separated under the applied electric field on the basis of their molecular mass/weight. The SDS masks the charge present on the proteins, giving roughly a constant net negative charge per unit mass to all the protein-SDS complexes. This ensures that the electrophoresis carried out in the SDS-PAGE gel is based only on the molecular weight of the protein.

7. **Scanning of 2D-DIGE gels:**
Scanning of the 2D-DIGE gels is carried out by using Typhoon, a highly sensitive variable-mode imager which meets the specific requirements of the 2D-DIGE experiment (Fig. 6). The difference in excitation and emission signals of the CyDye DIGE Fluor dyes helps the imager to detect the signals optimally with exceptional signal to noise ratio. The combination of CyDye DIGE Fluor dyes and the Typhoon imager provides a sensitivity down to 125pg of protein and enables quantitative and reproducible data production. The imager platform provides 3 set of the images and the file outputs are saved in a format compatible with DeCyder 2-D Differential Analysis Software.
**Fig. 6: Schematic representation of the Imager:** Protein samples labeled with different dyes are mixed together to run on a single gel. Three different images of the same gel are captured and used for differential expression analysis.

**8. Software Analysis:**

DeCyder 2-D Differential Analysis Software exploits the properties of CyDye DIGE Fluor dyes. It is a fully automated analysis platform which uses the three gel images obtained from a single gel to perform detection, quantitation, positional matching and analysis of the images to identify differentially expressed protein spots. The DeCyder 2-D software comprises of four modules, viz., Image loader, DIA, BVA and EDA (Fig. 7).

i. **Image loader:** The scanned images from the Typhoon imager are imported on to the analysis platform to proceed with analysis of the gels. This module also provides an option for cropping the gel image to select and focus on the gel area of interest and exclude the regions of less importance.

ii. **Differential In-gel Analysis (DIA):** This module uses a co-detection algorithm. It forms a cumulative image by overlaying the three images obtained from a single gel, and then co-detects
and produces identical spot boundaries for each gel image. This is followed by background subtraction, gel artifact removal, in-gel normalization, and quantitation of spot protein abundance for each image. The abundance values are expressed as a ratio, which helps for a better comparison of each protein spot between the sample gel images and with the in-gel internal standard reference image. Hence, with the help of DIA, all protein spots on a single gel can be detected and for each spot, changes in the expression levels between the two samples can be determined.

iii. Biological Variation Analysis (BVA): This module performs gel to gel matching of spots across multiple gels, allowing quantitative comparison of protein expression. It works on the original scanned image files and DIA files that have undergone spot detection. All resulting image files are then compared and matched to a single master image and common protein spots across all the gels are identified. A number of statistical analysis tools can be used in BVA module to highlight the protein spots which show significant change in expression under varying conditions.

iv. Extended Data Analysis (EDA): The module works on the BVA input data files. In addition to all the functions that can be carried out in the BVA module, the EDA can perform additional analyses. One of these analyses is the Principal Component Analysis (PCA), which gives an overview of all the data and detects the outliers. Another analysis performed by EDA is the Pattern analysis that identifies the pattern of protein expression. Discriminant analysis in EDA is useful for locating biological markers and assigning samples to known classes depending on their expression profiles. Interpretation in EDA is used to find out the functions, processes and pathways that are related to the identified protein by integrating the biological information with in-house or public databases in-house database.
Fig. 7: **Schematic representation of DeCyder 2D Analysis**: Initially the scanned image files are subjected to spot detection in DIA. All resulting image files are then compared and matched to a single master image and common protein spots across all the gels are identified in BVA.