HANDOUT

LECTURE-16

GEL-BASED ELECTROPHORESIS

TWO-DIMENSIONAL DIFFERENCE GEL ELECTROPHORESIS (2D-DIGE)

Slide 1:
The previous lectures have covered 2-DE workflow and how to perform 2-DE step by step, including, protein extraction, isoelectric focusing, equilibration, second dimension separation by using SDS-PAGE and finally the different types of staining methods available and how to scan and analyse stained gel images. The different advanced methods to creatively overcome limitations of gel-based approaches were also discussed. In the same theme, today, an other advanced method of gel-based proteomics, Difference in Gel Electrophoresis or DIGE will be covered.

Slide 2:
The previous lectures focused on gel-based proteomics, 2-DE workflow and the new methods used for the proteomics studies.

Slide 3:
In today's lecture, firstly, the different challenges which are associated with 2-DE are discussed, followed with a method to overcome those challenges, namely, 2-D flourescence difference gel electrophoresis (DIGE).

Slide 4:
First, see an image obtained when the all the steps preceding 2-DE, as well as the 2DE process itself are performed perfectly: In the end, good protein separation on the gel is obtained. This is a representative 2-DE gel image following seperation of the components of a human serum sample.

Slide 5:
Shown next is another good gel run of a bacterial sample, *B. subtilis*. But often this is not the case. Many times there are different types of issues which could give rise to very poor seperations. Issues could be arise from the sample preparation, the reagents involved, IEF, or from the staining step and different types of staining methods being used.

Slide 6:
There are various challenges associated with 2-DE.

Slide 7:
Disadvatages of 2-DE:

- Gel artefacts are major limitations as are gel to gel variations. These can arise from variations in electrophoretic run or different types of technical and biological variations.
- The image analysis is a very challenging process, with defination of the spot boundaries and intensity calculations introducing variations in sample analysis, with the added variation introduced by user bias.

Slide 8:
A few problematic gels are now shown, illustrating how even small mistakes or issues inherent to biological samples can give rise to very bad 2-DE gels. The example shows you sample preparation issues. In this gel, the TCA-acetone precipitation method was followed, but the washing was not performed well, so the TCA remained in the pellet and produced the pattern shown in this gel.

Slide 9:
This type of streaking and uneven gel pattern is obtained when solublization of the plant protein extract sample is performed without the step of precipitation.

Slide 10:
Samples with very high protein concentartions, such as crude serum. will show interference in the focusing step because of the presence of salt or too high amounts of

protein or other interfering components, resulting in such bad gels as shown in the example.

Slide 11:

Chemical impurities or the chemical ingradients in the sample can also give rise to bad gels. In this example,impurities in urea result in the carbamylation trail as shown in the slide.

Slide 12:

As very small amounts of TEMED are used during 1-DE, a single bottle of TEMED would continue to be used for a very long perriod. The pattern seen in the example is due to TEMED whose quality has declined during storage.

Slide 13:

If the TRIS used is not of the purity grade required, such artefacts as shown in the slide develop.

Slide 14:

Streaking in a gel can arise from various causes. It can arise from primarily from sample preparation issues, including the presence of interfering components, or nucleic acid or carbohydrate contamination, which cause streaking during isoelectric focussing.

Slide 15:

So as shown in this gel, the presence of salt contributes tovertical streaking as shown in the gel.

Slide 16:

Vertical streaking can arise from the use of a narrow pH gradient strip such as 4-7 pH gradient strip because all the proteins with a pI beyond 7 will stack together in that region.

Slide 17:

Abundant proteins: These form one of the major interfereing components in a sample prep. In a serum sample, albumin would consititute such an abundanat protein. In a plant sample, an example of an abundant protein is RuBisCO. These abundant proteins

mask many smaller proteins and create problems in the IEF step of 2DE, as shown in the example.There are various ways to overcome these limitations, mainly by the removal of abundant proteins as discussed in the previous lecture.

Slide 18:

Incomplete equilibration (only with DTT) can result in the pattern shown

.Slide 19:

Contrary to expectation, increasing the equilibration time does not improve gel quality, on the contrary actually, as seen in the gel incubated with DTT for 20 mins followed by IAA for 30 mins.

Slide 20:

The IEF process is a crucial one. Problems arising during focussing can be prevented by monitoring the run using suitable software programs. or by adjusting the voltage which may be insufficient, which can cause 'underfocussing' as shown in the image.

Slide 21:

A too high volatge can cause over-focusing, as shown in the gel. Optimizing the focusing protocol for an appropriate duration and voltage is very crucial.

Slide 22:

There are various sources of variations which can be introduced from the biological as well as technical issues,. Often, the biological issues cannot be fixed or controlled—for example, inter-individual variation in clinical studies, even in an age and gender matched cohort.However, the technical variations can be tackled during the sample preparation, IEF and equilibration steps. So all of these steps may result in large variations which may not be fully resolved. The running of the control and test samples on different gels in 2-DE hence has a potential for questionable data.

Slide 23:

This limitations have been tackled by the introduction of two-dimensional difference gel electrophoresis or 2D-DIGE technology.

Slide 24:
The 2D-DIGE method was basically developed to overcome the gel-to-gel variability which was observed in 2-DE since allowed the multiplexing of the samples. This method was reported by Unlu *et al.* in 1997 and became very user-friendly and is now in use by thousands of laboratories worldwide. The DIGE method involves protein labelling using flourescent tags and separation techniques This method allows linear detection over a the wide range of the protein concentrations and is a very sensitive method with a very large dynamic range of the protein detection.

Slide 25:
DIGE:Basic principles

- Importantly, this method incorporates commonly used protein labelling chemistry to tag the proteins.
- N-hydroxysuccinimidyl or NHS ester derivatives of cyanine dyes Cy2, Cy3 and Cy5 are used in this method. The Cy3 and Cy5 dyes are used to label the controls and treated samples. One should always perform dye swapping so that there is no dye bias. If in one sample you are adding Cy3 in control and Cy5 in treatment then in other sample the reverse should be done. However, Cy2 is consistantly used for labelling the internal standard. Also, the size and charge of the Cy3 and Cy5 are matched so that the labeled samples can co-migrate within the gel.

Slide 26:
The next section discusses labelling methods in DIGE technology.

Slide 27:
When previously discussing the 2-DE workflow in the different types of staining methods currently available, such as Coomassie, Silver staining and SYPRO Ruby were discussed. In that context, Cyanine dyes, that are used in DIGE technology were briefly covered As mentioned, 3 different Cy dyes Cy3, Cy5 and Cy2 are used. The excitation and emission wavelength of these three are detailed. The Cy3, Cy5 excitation wavelengths are at 540 nm, 620 and 480 nm, respectively. Similarly the emission

wavelengths of Cy3, Cy5 and Cy2 are at 590 nm, 680 nm and 530 nm, respectively. These dyes are photostable, they are not sensitive to pH changes and are spectrally distinct. By using these dyes the DIGE method can be used.

Slide 28:

In DIGE there are two types of labelling method choices available. The most commonly used is lysine labelling or cysteine labelling. The lysine labelling method is also known as minimum labelling method wherein the charge and the size of dyes are matched. Multiple labels are not used and it is known as minimum labelling because it is labelling only 3% of all the lysine residues in the proteins. I will give you the reason why that is we are aiming for very small amount of proteome to be labeled and that will be covered in the next slide.

The saturation labelling is done bytagging the cysteine residues. This method is charge neutral and size matched and multiple labels are used, it is very sensitive and, used only for Cy3 and Cy5 tagging of proteins. Saturation labeling is not very widely used as its counterpart, the minimal labeling method.

Slide 29:

The minimal labelling procedure that targets lysine residues is now discussed. The amino acid lysine is selected as because as shown in the graph , which depicts the distribution of amino acids across the proteins and their percentage distribution, lysine is present at a frequency 5% so statistically almost every protein has atleast one lysine residue. There is hence a good probabilty that all proteins in the sample will have lysine residues and hence will be labeled. However if this is not the case, the alternate method of saturation labelling of cysteines can be employed.

Slide 30:

Minimal labelling or lysine labelling process

- Cy dyes react with the primary amino group of the target proteins and the N-terminal alpha-amino and lysine epsilon-amino groups by the process of nucleophilic substitution. So labelling of all the available lysine residues would

create a very hydrophobic environment. That was observed in the original paper by Unlu *et al.* To avoid this, the minimal labelling procedure was developed wherein only 1-2% of the total lysine residues are labeled so that overall it is not very hydrophobic.

Slide 31:
This slide explains the the labelling chemistry. So as you can see in the structure here the stucture of CyDye Flour comprises of a positive charged CyDye Flourophore, a linker and NHS ester reactive group. So at pH 8.5 the NHS ester couples with amino group of lysine in the protein and it forms a covalent bond.

Slide 32:
The amino group of lysine residues is positively charged at neutral or acidic pH. The Dye is also positively charged so that the net pI will be unchanged. Therefore Cy3 and Cy5 dyes will not alter mass or charge overall and they are very well spectrally resolved. This is the reason this method is prefered.labeled

Slide 33:
Sample Preparation:

- Reconstitute the dye in dimethylformamide (DMF).
- The lysis buffer in which your protein sample is dissolved consists of 30 mM Tris, 7 M urea, 2 M thiourea, 4% CHAPS, pH 8.5. It is necessary to sure that there is no IPG buffer or carrier ampholyte or reductant added in this lysis buffer.
- Minimal amounts of flours added: Since you want to perform minimal labelling, only 400 pmoles of Cy dyes is added to to 50 µg of protein the reaction should be performed on ice for 30 min.For staining with coomasie or the Silver stains, a large amount of protein is required. But as these flourescent dyes are very sensitive, much lower amounts of protein can be readily detected.
- Quenching: Importantly, the reaction should be quenched with excess of primary amine, free lysine; 1 µL of 10 mM lysine, which will bind up all the non-conjugated dye molecules.

Slide 34:
The Cy2 dye is used to label the internal standard.

Slide 35:
A simple example is presented here, where you have 4 healthy controls and 4 patients suffering from a disease. So to represent all the proteins present in each of these samples in healthy individuals and diseased patients first you need to make an internal pool which is the representative of all of these 8 samples. So you can take 25 µL from each these 8 samples and make to 200 µL of one pooled sample which is your internal pooled standard. This pooled sample is then labeled with the Cy2 dye so that it can be used as the internal standard.

Slide 36:
This internal standard has many advantages and is reponsible for the actually one of major success of the DIGE procedure in overcoming the limitations of 2-DE.

- Every protein is present in the internal pool population and thus will be present on each gel run in the experiment.
- This reduces gel-to-gel variations.
- Gel-to-gel matching and comparison is much easier and more efficient.
- The internal standard is creating a reference point for every protein on each gel in the experiment. Later on when a software performs data analysis, the use of the internal standard produces into accurate spot statistics and more meaningful biological can be interpreted from this data analysis.

Slide 37:
The following is the step-by-step DIGE procedure.

Slide 38:

The first step following protein extraction is internal standard preparation.

Slide 39:

Next, the control sample, treatment sample and internal standard should be labeled with Cy dyes. Internal standard is always labeled with Cy2 dye but the control and treatment samples should be dye swapped.

Slide 40:

Once the labelling is done then the samples should be mixed and all the three labeled samples should be combined in one single e tube.

Slide 41:

The next step is rehydration which has already been covered.

Slide 42:

Next, the rehydrated strip can be focused in an isoelectric focusing instrument. It is a good idea to avoid any light exposure during rehydration, dye reconstitution process as well as during the IEF run.

Slide 43:

Once IEF is finished, the equilibration is done in 2 steps followed by SDS-PAGE. So the overall procedure is the exactly the same as 2-DE. The only variation here is that all your control and treatment samples have been mixed together and now are separated on a single IPG strip. Then, this strip is placed on a SDS-PAGE gel and proteins from control and sample treatments are are separated on a single gel.

Slide 44:

Now this one gel can be scanned using flourescence scanner such as the DIGE imager as shown -- and Cy3, Cy2 and Cy5 patterns can be obtained.

Slide 45:

The slide gives you an overview of the whole DIGE process where one needs to mix Cy2, Cy3 and Cy5 labeled samples in one tube and resolve this on one gel on SDS-

PAGE and scan that gel with different excitation and emission wavelengths to obtain 3 gel images derived from the same gel.

Slide 46:
A few representative gels obtained from the DIGE experiment, scanned for Cy3, Cy5 and Cy2 are shown.

Slide 47:
The software also generates the combined image obtained from superimposing all 3 dye image. s The overlapped image which shows you the abundance of certain proteins. In the overlapped image the red and green are indicative of over- and under-expression, respectively and the white or yellow colour representing that proteins for which there is no fold change in expression between control and treatment samples.

Slide 48:
Although the DIGE image analysis is a software driven process, it is much more automated compared to 2-DE analysis but it requires the use of a lot of tools for data interpretation.

Slide 49:
From the gel,you have obtained 3 images and those can be further analysed by using 'DIA' of the DeCyder software.This software takes these 3 images into account andcreates a single 'image pool' by the process of DIA. At the next level, multiple DIA from different gels can be matched in the 'BVA'. At the next level, BVAs generated can then further be used for obtaining statistical information about these protein changes, whether those are statistically significant or not.

Slide 50:
is shown aa representative gel here where 2 spots are highlighted from diseased and healthy controls. So one can look for the abundance of those proteins in the 3D views and shown on the right to determine if the proteins are up- or down regulated in healthy vs diseased vs other treatments.

Slide 51:
Now These 3D views across all the samples in both groups are examined. So for example, group of 6 patients and 6 healthy controls is shown. The protein of interest is showing up-regulation in all the 6 patients compared to the healthy individuals. So by looking at the data carefully patient-wiseand examining reproducibility, one can determine if the protein has some biological relevance and is suitable for further charectarization.

Slide 52:
The next level of statistical analysis past the BVA is the EDA or extended data analysis. Now the EDA module has different types of tools in the software such as principle component analysis (PCA). The principle component is the linear combination of optimally-weighted observed variables. PCA is used to test whether protein expression is uniform across all the multiple samples derived from the same experimental group, be it test or control. So if your results are consistant across the whole population, the same protein pattern is observed in all the controls and in all the patient samples In such a scenario,spot map from the same experimental group will be located in the same region.

Slide 53:
To clarify this point see the image. Herein, one control(red dots) and two treatment groups (blue and yellow dots) representing the same protein in different groups are depicted., If the dots from a single group cluster togetehr,it means it is highly significant. , as seen in the example.

Slide 54:
Using this software further information about different types of discriminants and classes can be obtained. In other words, the discriminant analysis is used to classify uniform data for the known classes. One can initially take all the proteins which are significant and try to group those based on the known classes.

Slide 55:
Hierarchical clustering is another method giving useful information. The figure is showing upregulation (in red) or downregulation (in green) for a number of proteins (depicted by the dendogram on the left). This method of depiction thus shows the fold change for individual members of all 3 groups (the blue which is the internal control), the red which is the disease or test group (in the middle) and for healthy controls in green (on the right) So one can tell at a glance that a majority of proteins behave similarly in all the subjects of the control group, but in the test group. fall into 2 subpopulations wherein some genes are upregulated in some patient samples from the disease or treatment group, while they are downregulated in other samples from the same group. In other words, the test group shows more variation.

Slide 56:
Next shown is an animation of how to perform a DIGE experiment, which should help recall concepts explained so far.

Animation 1

Difference in gel electrophoresis:
The pooled internal standard for DIGE is prepared by mixing equal amounts of all samples that are being run in the experiment. This prevents problems of gel-to-gel variations.

Dye labeling
Each protein sample as well as the internal standard is labeled with a differently fluorescing cyanine dye which allows all protein samples to be simultaneously run on a single gel. The dye binds covalently to the epsilon-amino group of lysine residues in proteins.

Mixing and 2DE
The labeled protein samples are mixed and run on a single 2-DE gel. Separation takes place on the basis of isoelectric points of the proteins in first dimension and based on

molecular weight of the proteins in the second dimension with the smaller proteins migrating further along the gel.

Gel viewing

The gel containing all the protein samples can be viewed by illuminating it alternately with excitation wavelengths corresponding to the various Cyanine dyes.

Superimposed 2D-DIGE image

Now the superimposed DIGE gel depicting all protein spots of multiple samples can be viewed. Information on molecular weight and pI of proteins can be obtained from the gel. But most importantly by looking at the gel, if performing abundance proteomics comparing the expression of two different protein samples, then a pattern can be immediately observed on the gel by a visual examination of red and the green spots. If there is no change it will appear as white owing to the superimposition of one on the other.

Slide 57:

After understanding the concepts so far, which would be a better method to separate proteins in the serum samples of 250 patients in a clinical trial? If your answer is 2-DE, then that is not correct. You need some method which minimize variations across the large population studied, in other words, DIGE.

Animation 2

Slide 58:

This animation elaborated on the the statement.

2-DE, although a very useful technique, may not be the best option in this case for analyzing serum proteins from a large number of patients as it would involve running several individual gels which would be a time consuming process. Also variations across the gels would make comparison of results a problem.

Whereas DIGE is an extremely valuable tool for analysis of a large number of samples simultaneously without having to encounter the problem of gel-to-gel variation. The control and test samples can be differentially labeled using the cyanine dyes and run on a single gel.

Slide 59:
To summarize the disadvantages of 2-DE and advantages of DIGE

- In 2-DE there are gel-to-gel variations which can be overcome by DIGE by separating the proteins on one gel.
- In 2-DE there are large sample variations whereas in DIGE these variations can be reduced by making an internal standard which provides reference points for all the proteins.
- In the 2-DE data analysis there is user-bias introduced during definition of spot boundaries and statistical analysis. The co-detection method involved by the software analysis in DIGE eliminates lot of these issues,

Slide 60:
So in summary, today's lecture covered the different challenges which are associated with 2-DE followed by the DIGE workflow, with a comparison of 2DE vs DIGE. So overall from this lecture, the take home message is that DIGE is a superior method to separate protein mixtures that can provide very quantitative information in gel-based proteomics.