

Clustering of 2D-Gel Images

Won Hur

Kangwon National University, School of Biotechnology and Bioengineering

TEL: +82-33-250-6276, FAX: +82-33-243-6350

Abstract

Alignment of 2D-gel images of biological samples can visualize the difference of expression profiles and also inform us candidates of protein spots to be further analyzed. However, comparison of two proteome images between case and control does not always successfully identify differentially expressed proteins due to sample-to-sample variation. Because of poor reproducibility of 2D-gel electrophoresis, sample-by-sample variations and inconsistent electrophoresis conditions, multiple number of 2D-gel image must be processed to align each other to visualize the difference of expression profiles and to deduce the protein spots differentially expressed with reliability.

Alignment of multiple 2D-Gel images and their clustering were carried out by applying various algorithms and statistical methods. In order to align multiple images, multiresolution-multilevel algorithm was found out to be suitable for fast alignment and for distorted images. Clustering of 12 different images implementing a k-means algorithm gives a phylogenetic tree of distance map of the proteomes. Microsoft Visual C++ was used to implement the algorithms in this work.

서론

현재 프로테옴 연구는 프로테옴 자체의 분석과 프로테옴 비교분석으로 주로 나눌 수 있다. 일반적으로 프로테옴 분석은 주로 생명체의 개체나 혹은 일부 조직에 대한 프로테옴 map 을 2DGel로 만들고 이를 TOF 질량분석 장치와 이미 알고 있는 단백질 서열정보의 도움을 받아 2차원 map상의 단백질을 하나하나 규명해 나가는 작업을 말한다. 특히 전사 프로필과 실제 단백질의 발현 프로필이 매우 다르며 동시에 복잡한 것으로 밝혀져 프로테옴 분석의 중요성이 더욱 강조되고 있다[1,2]. 프로테옴 연구의 다른 한 주류는 프로테옴 비교분석 연구이다. 이것은

비교유전체학과 마찬가지로 정상적인 세포와 질환 혹은 결합 상태에 있는 세포의 프로테옴 2DGel 이미지를 비교하여 그 원인 단백질 혹은 질환이나 결합 상태와 관련이 있는 단백질을 찾아내는 것이다. 개체간 변이 및 실험상의 오차를 극복하기 위해서는 다수의 시료로부터 프로테옴 2DGel 이미지 얻어도 이들을 서로 비교 분석하는 것이 필요하다[3-5]. 그 이유는 시료의 양을 정확하게 조정하고 동일한 조건에서 electofocusing 과 electrophoresis를 수행하고 발색과 탈색을 표준화시켜도 여전히 매 실험마다 동일한 시료의 2DGel 이미지도 상당한 변화를 보인다. 아울러 2D-gel 이미지를 유사한 것끼리 서로 모음으로서 단백질 스팟의 변이의 정도 및 발현 정도의 분포에 따른 클러스터링이 가능하다.

현재까지의 연구 내용은 고해상도의 재현성이 높은 2D-Gel 프로테옴 이미지를 구성하고 이를 분석하는 소프트웨어를 개발하는 것이다. 프로테옴 이미지 데이터 베이스에서 단백질 정보를 서로 비교 분석 혹은 검색할 수 있는 시스템은 아직 많이 연구된 바가 없다. 전술한 바와 같이 프로테옴의 다양성으로 인하여 적은 숫자의 프로테옴 2D젤 이미지로서는 대조구와 시험구의 이미지가 수개 이하의 단백질 스팟만이 명확하게 차이날 경우를 제외하고는 의미 있는 결과를 얻기 어려운 경우가 많다. 이와 같은 한계를 극복하기 위하여 본 연구에서는 다수의 프로테옴 2D Gel 이미지 분석 방법을 연구 개발하였다.

재료 및 방법

Multiresolution and multialignment 알고리즘을 이용한 이미지 alignment와 클러스터링을 소프트웨어적으로 구현하기 위하여 Visual C++(Microsoft, USA)를 사용하였다. 2D-Gel 이미지는 인터넷에서 일련의 시료에 대한 다수의 이미지를 제공하는 Flickr 사이트에서 다운로드받아 사용하였다. 클러스터링을 위해서는 12개의 Fetal Alcohol Syndrome의 case 와 control의 프로테옴 이미지를 사용하였다.

결과 및 고찰

다수의 이미지를 정렬시키기 위해서 먼저 기준이 되는 이미지를 결정한 후 여러 이미지를 순차적으로 정렬시키는 방법을 사용하였다. 2D-Gel 이미지로부터 단백질 스팟을 먼저 찾아 내지 않고 원시 이미지를 사용하여 여러 단계의 해상도에서 보조 그리드 좌표를 활용하여 순차적으로 이미지를 정렬시키는 multiresolution-multilevel algorithm을 활용하여 Visual C++로 메모리가 허용하는 한 다수의 이미지를 정렬시키는 소프트웨어를 개발하였다. 500 x 500 pixel의 이미지를 정렬시키

는데 약 30초 소요되었다. Figure 1.은 Alignment 전후의 소프트웨어의 스크린 샷이다.

기준 이미지에 대하여 다른 이미지들은 2x2, 4x4, 8x8 및 16x16으로 이미지를 분할시켜 정렬을 하고 그리드 정보만을 기록하여 보관하여 원시 이미지 정보를 가능한 변화시키지 않도록 하였다. 이후 2DGel 이미지를 Alignment를 통하여 하나의 기준 2DGel 이미지에 정렬된 다른 2차원 이미지 상관계수를 구하여 Phylogenetic tree를 생성시켜 직관적으로 프로테옴 이미지의 클러스터링을 할 수 있는 소프트웨어를 구현하였다. 여기에 12개의 Fetal Alcohol Syndrome(FAS)의 case와 control의 프로테옴 이미지를 전술한 방법으로 정렬시킨 후 클러스터링을 시도하였다(Figure 2). 그러나 FAS의 프로테옴 이미지는 클러스터링 결과는 case와 control로 구분되지 못하고 나이 및 성별로 클러스터링 되는 경향을 보였다. FAS의 경우 2DGel 이미지의 전체의 경우 case의 경우 발현되는 차이보다 성별 혹은 나이에 따른 발현의 차이가 더 뚜렷하기 때문인 것으로 판단된다. 따라서 이것을 해결하기 위해서는 Principal Component Analysis 방법을 통하여 case의 경우 발현량의 차이를 보이는 단백질 스팟을 추출하는 방법을 시도해 볼 가치가 있다.

References

1. Jensen O. N., Larsen M. R., Roepstorff P. (1998), Mass spectrometric identification and microcharacterization of proteins from electrophoretic gels: Strategies and applications, *Proteins, Supplement 2* , 74-89.
2. Lopez M. F. (2000), Better approaches to finding the needle in a haystack: Optimizing Proteome analysis through automation, *Electrophoresis* **21**, 1082-1093.
3. Harry J. L., Wilkins M. R., Herbert B. R., Packer N. H., Gooley A. A., Williams K. L. (2000), Proteomics: Capacity versus utility, *Electrophoresis* **21**, 1071-1081.
4. Jungblut P. R., Schaible U. E., Mollenkopf H. J., Zimny-Arndt U., Raupach B. M. (1999), Comparative proteome analysis of Mycobacterium tuberculosis and Mycobacterium bovis BCG strains: towards functional genomics of microbial pathogens, *Molecular Microbiology* **33**, 1103-1117..
5. Haynes P.A., Yates J.R. (2000) Proteome profiling-pitfalls and progress, *Yeast* **17**, 81-87.

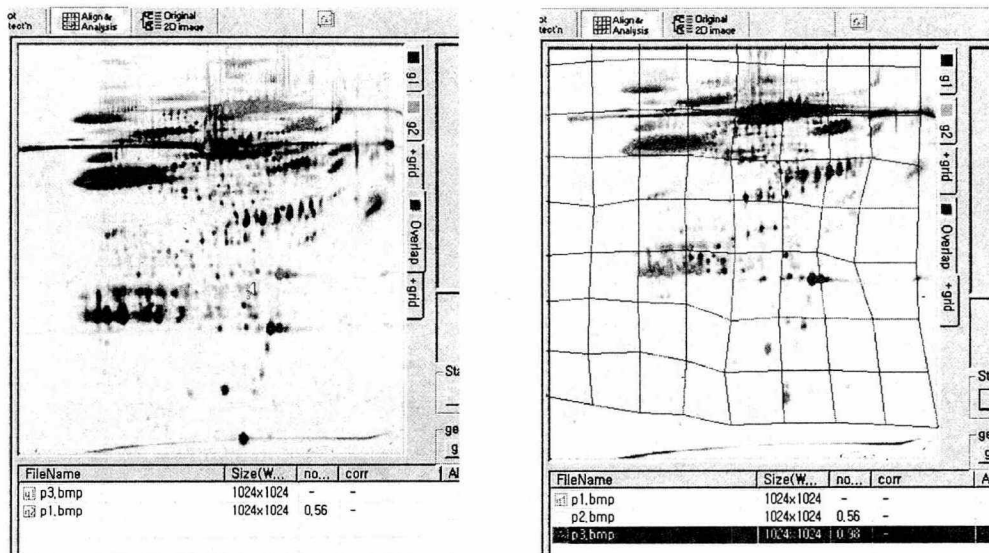


Figure 1. Screen shot of alignment of multiple 2D-gel images using multiresolution-multilevel algorithm.

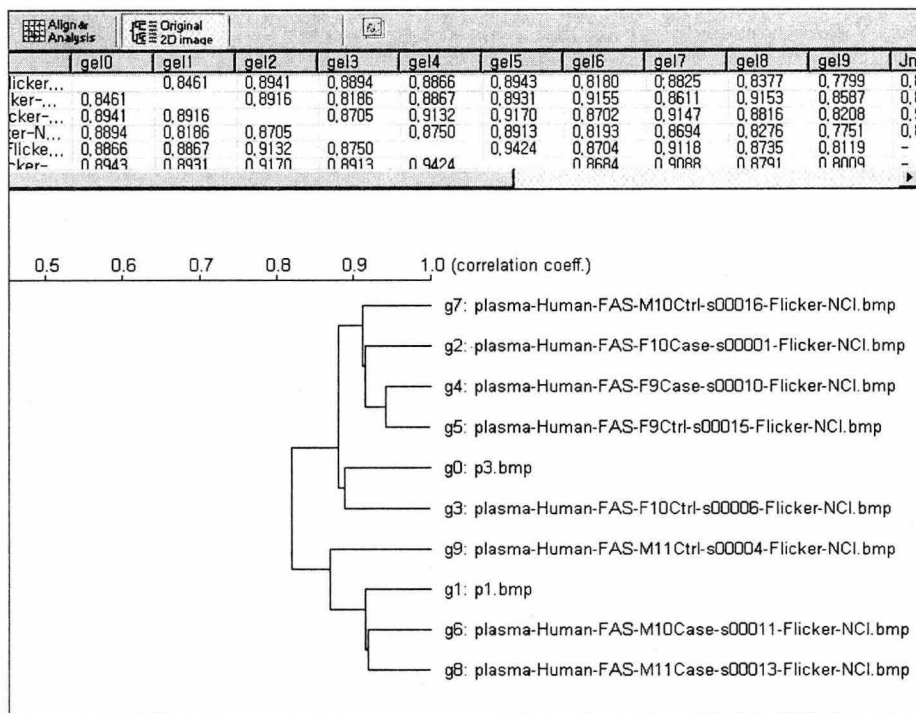


Figure 2. Clustering of 2D gel proteome images of cases and controls of fetal alcohol syndrome