



제어시스템공학 관점에서 바라 본 Bioinformatics 분야의 현황과 전망

박진현*, 이인범

*(주)피앤아니컨설팅, 포항공과대학교 화학공학과

1. 서론

HGP(Human Genome Project)가 완료된 이후 DNA microarray와 같은 자동화된 실험법이 상용화 되어감에 따라 과거에는 이용하기 힘들었던 무수히 많은 유전자 데이터가 나오고 있지만 이러한 데이터를 가지고 실제 질병을 진단하거나 처방하고 또한 신약의 개발에 이를 수 있는 후보 물질을 탐색하는 것 등의 실제적인 부분으로 적용을 한다는 것은 그리 쉽게 접근할 수 있는 부분은 아니다. 그 이유는 우선 생물학적인 Genomics, Proteomics, Genetics등과 같은 분자 생물학적인 지식을 갖추어야 하고 이와 더불어 공정제어 및 시스템 공학의 지식을 기반으로 하여 앞에서 언급한 생물학적인 지식과 접목을 시켜나가야 하기 때문이다.

현재 Bioinformatics라는 학문을 전문적으로 전공한 사람이 거의 전무한 상황에서 누가 과연 Bioinformatics를 가장 효율적으로 연구할 수 있는가에 대하여 다음과 같은 두 가지 경우에 대하여 논란이 있다. 먼저 분자생물학적인 지식을 습득하고 그 연후에 시스템 공학적인 지식(공정확인, 데이터 분석 알고리즘, 최적화, Data Mining 등)을 습득하여 Bioinformatics를 하는 경우와 시스템 공학적인 지식을 기반으로 생물학적 지식을 접목시켜 나가는 것 중 어느 것이 더 효율적인가 하는 논란이지만 필자는 감히 후자가 더 바람직하리라고 생각한다. 그 이유는 제어시스템공학을 전공하게 되면 복잡하고 비선형성이 강한 공정을 분석하고 모델링하는 체계적이고 합리적인 지식을 얻을 수 있기 때문에 생명현상도 복잡한 하나의 시스템으로 생각한다면 충분히 분석하여 알지 못했던 많은 현상을 밝혀나갈 수 있기 때문이다. 그렇다고 해서 제어시스템공학을 전공한 사람이 모든 것을 이룰 수 있다는 것은 아니고 실험을 하는 생물학자와 반드시 collaboration을 이루어서 연구가 이루어 질 때 그 파급효과는 가히 상상을 초월한다고 할 수 있다. 실제 제어시스템공학을 전공한 사람들이 할 수 있는 분야를 생명 공학적인 분야와 연관이 있는 부분을 보면 다음과 같다. 현재 많이 연구되고 있는 질병을 진단하거나 원인을 규명하는 부분은 화학 공장에 대한 Controller Fault Detection과 아주 유사하며 신약

후보 물질을 탐색하는 부분은 최적화의 기법이 반드시 적용되는 실험계획법 및 다변량 통계분석과 연관된다고 할 수 있다. 이외에도 Metabolic Network나 Genetic Network를 규명하는 부분도 제어시스템 공학 지식을 접목한다면 체계적이고 합리적인 분석을 통해 생명현상을 조명해 나갈 수 있으리라고 생각된다. 위에서 언급한 여러 부분들을 효과적이고 체계적으로 통합한다면 제어시스템공학을 전공한 사람들이 나아갈 수 있는 새로운 Paradigm이 될 수 있으리라고 할 수 있다.

2. 생물 시스템 모니터링 기술

생물체는 항상성 기작을 구성하는 요소들, 즉 유전자 혹은 단백질로 대표되는 각 전사체들의 발현정도를 나뉠대로의 고유한 규칙에 의해 변화시킴으로써 끊임없이 외부의 환경변화에 대응하고 있다. 즉 생물체는 그 자체가 하나의 아주 복잡한 시스템이므로 결국 제어시스템공학의 관점에서 바라본다면 생물체 역시 외부의 환경으로부터 자신을 지키기 위해 스스로 구성 요소들을 변화시킬 수 있는 능력을 지닌 하나의 고차원적인 시스템으로 볼 수 있다. 그러므로 생물시스템이 변화하는 메커니즘을 알기 위해서는 단지 몇몇 구성요소의 변화를 모니터링하는 것보다는 전체 시스템의 보다 많은 구성 요소들을 동시에 그리고 정확히 모니터링 할 수 있는 기술과 더불어 모니터링된 시스템을 이해하고 모델링, 분석 및 제어할 수 있는 기술들이 필요하다.

하지만 이전까지 대부분의 유전자 연구는 northern blot 기법과 같은 단일 유전자 및 분석기법이 주로 이용되었고 연구자가 알기를 원하는 단지 몇몇의 유전자 발현간의 상관관계를 분석하여 생물시스템에서의 유전자와 단백질 사이의 관계를 규명하려는 노력들이 주를 이루었다. 그러나 생물시스템의 부분적인 모니터링에 의존해온 기존의 연구방법만으로는 그 복잡한 양상들을 이해하기에는 한계가 있어 보였고 결국 1980년대 이후부터 생물 연구자들에게도 생물을 하나의 시스템적인 관점으로 바라보고자 하는 움직임이 생겨나게 되었다. 이러한 요구에 따라 동시에 많은 수의 유전자의 발현

분석을 수행하는 방향으로 생물연구방법이 전환하였고 결국 최근 수천에서 수 만개의 유전자 혹은 단백질의 발현을 모니터링 할 수 있는 기술인 DNA microarray 혹은 protein microarray 등과 같은 기술로 구체화되고 있다. 전체염기서열의 해독이 끝난 *S. cerevisiae*의 경우 전체 유전자 발현정보를 담을 수 있는 DNA 칩이 개발되어 단 한 번의 실험으로 한 생물 시스템 전체를 모니터링 할 수 있게 되었다. 또한 일련의 microarray technology와 관련된 분석기들이 상업화 되고 표준화 되어가면서 데이터의 신뢰도 및 실험의 질적 수준이 급속히 높아지고 있으므로 점차 생물연구는 이전에는 상상할 수도 없었던 높은 연구효율을 갖추어 가게 될 것으로 보인다.

이러한 관점에서 볼 때 제어시스템공학 분야에서 사용되어지고 있는 모니터링, 최적화, 생산일정계획, 모델링, 제어 등의 기술들을 접목한다면 생물시스템을 이해하고 분석 및 문제점을 해결할 수 있는 실마리를 제공해 주리라고 생각한다.

3. Microarray 데이터 분석 기술

DNA Chip적는 방법

Normalization

Chip Data 분석기술

4. Metabolic Network Analysis

Microarray technology의 발전은 생물 연구영역에서의 성과라기보다는 실험의 자동화를 위한 기계공학기술의 도입 및 microarray의 제작을 위한 광학기술의 접목 그리고 분리 및 정제를 위한 화학 분야등에서의 연구 성과들에 힘입은 바가 크다. 물론 앞서의 관련 기술들이 microarray technology의 효율을 점차 극대화시키는 있는 원동력이 되고 있지만 실제 생물 연구 영역에서의 주된 관심은 각 연구 성과들이 얼마나 잘 활용되어 실제 생물시스템의 현상을 잘 해석할 수 있느냐 문제일 것이다.

이러한 목표에 이르기 위해서는 많은 양의 신뢰도 높은 실험 데이터를 만들어 내는 것이 주된 연구 방법인 이용하는 생물학의 독특한 학문적 특성 때문에 DNA microarray(DNA chip) 혹은 2DE-Gel electrophoresis, protein microarray 등의 실험을 통해 얻어진 생물 데이터들에 대한 quality의 검증은 역시 매우 중요한 문

제 중 하나이다. 하지만 대부분의 자연현상들은 관찰 (monitoring)하고자 하는 노력이 시작되는 순간 데이터 속에는 관찰자의 인위적인 노력에 기인한 error가 생겨나게 되고 이는 실제 현상을 이해를 어렵게 하는 제한 요소가 된다. 생물시스템의 관찰에서도 마찬가지로 이러한 제한요소가 생성되어 in vivo와 in vitro라는 경계 속에서 연구자들의 실제 in vivo에 가까운 데이터를 얻고자 하는 샘플링에서의 노력은 생물연구분야에서도 통계적인 데이터 분석을 도입하게 하는 주된 문제가 되어왔다. 이렇게 생물시스템에 적용되기 시작한 통계적 데이터 분석은 관찰과정에서의 무수히 많이 생겨날 수 있는 인위적인 요소들을 확인하고 제거하는 과정을 반복해 실제로 매우 미세한 생물시스템을 최대한 실제에 가까운 곳에 위치시킨 후 구축된 통계모델 속에서 개별 구성 인자들의 요동(perturbation)에 대한 모니터링이 가능할 수 있도록 도와주는 가이드 역할로서 우선 적용되어 오고 있다.

결국 microarray 실험에서도 측정된 gene들의 발현 정도에 영향을 줄 수 있는 여러 종류의 systematic variation이 존재한다[10]. 따라서, 보다 신빙성 있게 발현량의 차이를 알아내고 정확한 패턴의 찾고자 하는 과정의 진단계로서 normalization 과정이 필수적이다. microarray chip에 나타나는 systematic variation의 원인은 실험 전과정에서의 실험적 uncertainty로부터 chip의 image 처리과정에 이르기까지 매우 다양하다. 따라서, 이러한 원인들 각각으로부터 발생하는 variation을 모두 찾아내어 제거하는 것은 불가능하다. 제한적으로나마 다른 종류의 fluctuation을 각기 다른 종류의 실험을 통해 얻어진 데이터를 이용하여 선택적으로 제거할 수 있다[11]. 이를 위해서는 미리 제거하고자 하는 fluctuation만을 이끌어내기 위한 계획적인 실험이 필요하다. 하나의 chip안에서의 normalization 뿐만 아니라, 두개 혹은 그 이상의 chip들간의 normalization도 고려해야 한다. 일반적으로 slide 내에서의 normalization은 두 dye를 통해 일정하게 발현되는 것으로 알려진 gene들 혹은 통계적 방법을 통해 거의 발현의 차이를 보이지 않는 gene들을 선택한 후, 이러한 gene들을 바탕으로 두 dye간의 선형 회귀식을 유도하여 이 회귀선에서부터의 거리를 측정하여 발현차이로 보고 있다. 이를 위해 기존의 log-log 그림이 아닌 이를 45도 회전시킨 형태의 MA plot를 이용하기도 하는데, 이 plot에서는 평균 intensity의 값이 x값으로 두 dye의 intensity의 차이가 y축에 나타나기 때문에 두 dye 간의 평균 intensity에 따른 발현량의 차이를 확인하는데 보다 쉬운 이점이 있다. 위에서 언급한 바와 같이 slide 전체의 gene을 하나의 회귀선을 통해 normalization하는 방법 이외에, dye의 bias가 gene의 intensity에 어느 정도 의존하는 경향을 고려한 방법

이 제시되기도 하였다[12]. 즉, 평균 intensity의 크기에 따라 적절하게 부분 회귀식을 사용하여 전체를 수 개의 회귀식의 합으로 나타내어(이는 전체적으로 볼 때 MA plot에서 하나의 부드러운 곡선의 형태로 나타난다)이 값들을 중심으로 벗어난 정도의 크기를 발현량의 차이로 보게 된다. 이러한 단일 slide normalization 기법은 자연스럽게 두개 혹은 그 이상의 slide에 기반한 multiple slide normalization으로 확장 가능하다.

또한 많은 array 실험을 반복해서 수행하다 보면 실험 과정 중에서 missing value가 생길 수 있는데 물론 이러한 문제에서는 해당실험을 다시 수행하는 것이 가장 좋은 방법이라 할 수 있다. 그렇지만 Troyanskaya et al.[13]은 비용과 시간적인 효율을 높이기 위한 차선책으로써 missing value 문제도 통계적인 기법을 이용하여 보다 잘 해결할 수 있는 방법을 제안 하기도 하였다.

이외에도 microarray를 이용하는 실험의 유효성을 검증하기 위한 다양한 통계적 모델들이 normalization을 위해 적용되고 있고 Lee 등[14]과 같은 연구 결과들은 microarray를 이용한 실험의 유효수준을 높이는데 훌륭한 가이드가 되어 high throughput array technology들의 질적인 향상을 견인해 나가고 있다.

5. 생물을 바라보는 Macroscope와 같은 역할로서 응용되는 생물정보학

이전까지의 생물학 연구가 하나의 유전자 혹은 단백질의 세포 내에서의 작용을 살펴보기 위해 현미경(microscope)을 사용해 왔다. 다시 말하자면 보다 미세한 것을 실제로 관찰하기 위해서 연구자들은 과거에 더 고 배율의 현미경을 이용하기 위한 투자를 마다하지 않았다.

하지만 미세한 현상을 관찰하는 것 만큼이나 거대한 현상을 관찰하는 것 역시 쉽지 않은 일이다. 예를 들면 작게는 수 백에서 수 만에 이르는 한 생물 시스템의 전체 유전자의 발현변화를 수십 종류의 세포에 적용하거나 수 만 가지의 화합물들을 처리한 경우와 같이 세포 내의 미세한 요동을 관찰하는 연구를 수행한다면 우선 그 방대한 데이터의 양만 생각해보아도 컴퓨터를 이용한 전산기술과 이를 분석해 낼 수 있는 분석기술이 반드시 필요함을 알 수 있다.

실제로 앞서 언급한 바에 부합되는 초대형 실험이 NCI에 의해서 수행되었다. NCI 60 data set 라고 널리 알려진 이 대형의 실험에서는 70,000 종 이상의 화합물들을 60개의 cancer cell-line에 적용하여 동시에 9703개의 유전자 발현 정보와 각 세포의 activity pattern들을 측정하였다[15]. 또한 이 실험을 통해 나온 결과들은 일

련의 normalization 과정을 거친 후에 data base에 축적되었고 이 후 다양한 알고리즘들을 적용하여 cluster analysis가 수행되었다. 그 분석의 한 결과로서 각 cell line들은 expression pattern의 correlation distance (distance=1-pearson correlation coefficient)가 구해지고 이를 통한 cell-line 들의 차이가 효율적으로 cluster tree 상에 도식화 되었다. 또한 118개의 약제에 대한 각각의 cell activity pattern에 따라서 average linkage hierarchical clustering이 앞서의 correlation distance에 따라 cluster tree로 도식화되었고 아울러 activity pattern 들과 gene expression pattern을 고려한 correlation coefficient에 의한 Euclidian distance에 따른 cluster tree 역시 구성되었다. 이러한 일련의 과정을 거쳐 실험 결과는 AT-matrix clustered image map(CIM) [그림 1]이라는 단 하나의 이미지 속에 drug activity와 gene expression profile간의 연관성이 효율적으로 통합적으로 표현되었다[16]. 결국 CIM을 통해서 연구자들은 각각의 화합물들이 각각의 유전자 발현에 미치는 영향을 한눈으로 파악할 수 있게 됨으로써 약으로 개발될 가능성이 있는 물질들에 대한 전반적이고 대략적인 정보를 획득할 수 있게 되는 것이다. 이러한 연구 결과는 high throughput array technology가 생물정보학과 결합되어 얼마나 유용한 정보들로 변모해 갈 수 있는지를 보여주는 좋은 예시가 된다.

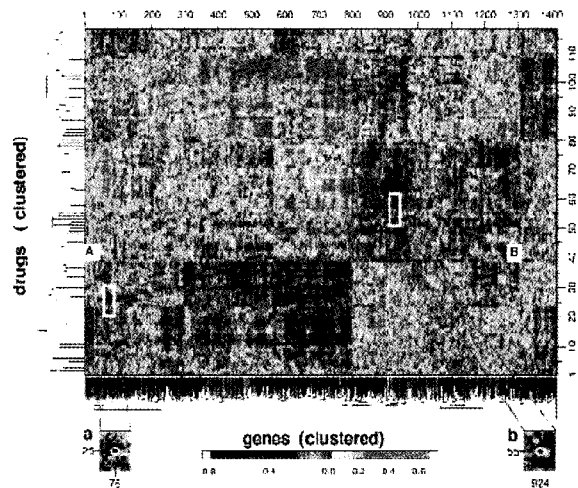


그림 1. 60 cell-line의 1376 gene들에 대한 118종류 compound들의 activity pattern들을 한눈에 파악할 수 있게 해주는 CIM(clustered image map). Red는 양(+)의 피어슨 상관계수를 나타내고 Blue는 음(-)의 상관관계를 나타낸다.

또한 S. cerevisiae에 300 종류의 세포처리, 즉 돌연변이를 일으키거나 화학 처리를 한 이 후 전체 유전자 발현 양상들을 모아놓은 compendium이라는 S. cerevisiae에 대해 다양한 정보들의 추출이 가능한 참고적인 데이

터 베이스가 구축되었다[17]. 이 데이터 베이스는 염기 서열의 분석이 끝난 몇 종류의 생물 시스템에 적용될 경우 보다 특정 유전자 또는 생물 내 효소들의 발현을 조절한 결과들간의 상관관계를 통해 그 세포내의 기능을 유추할 수 있게 해주는 좋은 참고자료가 될 것이다. 그러나 이렇게 많은 비용을 사용하여 구축된 데이터 베이스들은 어떤 목적으로 정보를 분석해 내느냐에 따라서 그 정보의 종류가 달라질 수 있게 되므로 수많은 정보들이 숨어있는 어레이 데이터의 높은 활용을 이끌어 내기 위해서는 각각의 목적에 맞는 적절한 분석방법의 개발이 수반되어야 할 것이다.

최근에는 gene expression profiling과 serum starvation 및 chemical inhibitor를 이용하는 cell synchronization기법의 융합으로 세포 내 주기에 따른 gene expression pattern을 통해 명확하게 세포내의 유전적 프로그램을 이해하려는 노력들이 점차적으로 증가하고 있다[18,19].

세포 내 유전적 현상 이해를 위해 gene expression pattern을 분석하기 위한 데이터분석 도구로는 크게 세 가지 도구를 주로 활용하게 된다. 그러므로 생물정보학에 유용한 데이터 분석 도구는 어떤 것들이 있고, 그러한 분석 툴을 이용하여 어떤 정보들을 추출할 수 있는지를 좀더 자세히 살펴보자.

가장 널리 활용되는 DNA 칩 데이터 분석 툴은 clustering, SVD(singular value decomposition) 방법이다. 세포주기 분석에는 위 두 가지 방법과 더불어 주기분석을 위한 phasing-세포주기, G1/S/G2/M로 genes을 재배열하는 방법이 유용한 데이터 분석법이다. Spellman 등[19]은 Yeast의 세포주기를 좌우하는 mRNA를 조절하는 주요 genes을 Fourier 변환과 genes사이에 상관관계에 기반한 clustering 방법을 이용하여 800개의 genes을 선택하였다. 일반적으로 사용되는 DNA 칩의 gene library가 5~6천여개 정도된다고 보면 800개의 주요 gene을 후보로 선정했다는 것은 세포내 대사를 조절하기 위한 인자를 찾는 데 경제적으로 매우 큰 효과를 준다고 할 수 있다. 좀더 구체적으로 살펴보면, 먼저 세포 주기를 이산 푸리에 변환(Discrete Fourier Transform)을 통해 계산하고, 계산된 주기를 이용하여 genes을 주기성을 가장 잘 설명하는 순으로 정렬한다. 주기성을 나타내는 인덱스로 정렬된 genes 중 의미 있는 genes을 선택하기 위해 적절한 Cut-off 값을 이용한다. Cut-off 값을 결정하기 위해 보편적으로 사용되는 방법은 mRNA 수준에서 세포 주기를 조절하는 gene 이라고 문헌 및 DB에서 알려진 것들을 최대한 포함하면서 관련 없는 gene을 지목할 가능성을 최소화 시키는(false-positive assessments) 값으로 설정하는 것이다. 선정된 genes에 대한 신뢰도는 세포 주기를 조절하는 몇 번의 실험을 통해 검증된다.

Alter 등[20]은 복잡하고 무수히 많은 array 및 gene을 SVD를 이용하여 eigenexpression (eigenarray 및 eigengene) 공간으로 변환함으로써 [그림 2-A] gene expression의 동특성을 분석하는 방법론을 제시하였다. 몇 개의-일반적으로 5개 내- 의미있는 eigengenes과 이에 상응하는 eigenarrays가 칩 내의 주요한 expression 정보를 내포하고 있다. 그림 2-B(a)는 390 세포주기시간을 갖는 세포 칩 데이터를 SVD를 통해 분해한 것이며, 두개의 eigenexpression으로 왼쪽의 칩 데이터가 설명할 수 있는 대부분의 정보들을 포함하고 있음을 알 수 있다. 두개의 eigenexpression 공간에서 세포 내에 일어나는 상태나 생물학적 표현형을 분석, 분류할 수 있다[그림 2-B(b)].

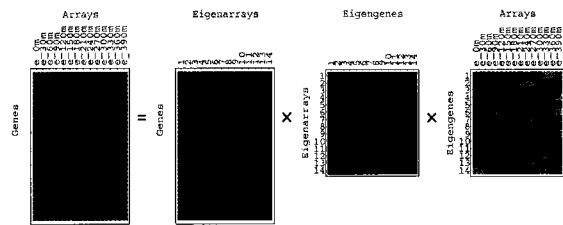


그림 2-A. 빨간색은 overexpression, 검정색은 no expression, 녹색은 underexpression

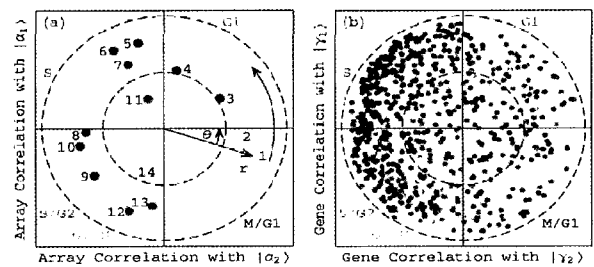


그림 2-B. (a) M/G1-노란색, G1-녹색, S-파란색, S/G2-빨간색, G2/M-주황색으로 세포내 주기를 14 array로 분류한 결과 (b) 두개의 eigenexpression에 의해 모델링된 Gene 상관관계 (Spellman에 의해 선택된 800 genes)

위의 방법론을 활용하면 세포 내에 일어나는 현상과 이와 관련된 주요 gene을 일반적인 세포 주기내에서 일어나는 현상 및 gene(모델)과 비교, 어떤 현상이 세포 내에 일어났는지, 기존 세포주기에 영향을 주는 주요 gene이 어떻게 조절되었는지를 신속히 분석, 파악할 수 있다.

결국 아주 복잡해보이고 많은 데이터들이 존재하는 생물 실험데이터에 생물정보학이 효율적으로 적용되어 실험진체를 한 눈에 파악할 수 있는 분석방법을 제공했다는 것이 앞서의 연구들 속에서 생물정보학적인 측면에서의 의미일 것이다. 결국 체계적이고 적절한 분석이 생물연구에 적용되어 앞서 얘기되었던 단순히 array 관



련실험의 질적수준을 측정 및 보정하는 역할을 넘어서 효율적인 데이터 마이닝 및 시각화를 통해 연구자에게 생물시스템의 전체정보를 읽어내는 것을 가능하게 하는 microscope 같은 역할을 역할로서 이는 복잡하기 그지 없는 생명현상을 연구해 나가는데 있어서 좋은 실마리를 제공해주고 있다.

6. 새로운 진단기법의 개발에 이용되는 생물정보학

미세한 분자수준에서의 생명현상을 규명하는 것이 생물학자들의 주된 관심이라면 의학관련 연구자들의 주된 관심은 보다 정확하게 인간의 분자생물학적인 현상들을 이해한 후 전체 인체시스템에서의 질병을 근본원인에 따라 정확하게 분류하여 진단해내는 것에 있다고 할 수 있다. 또한 치료는 정확한 진단에 근거해 실제 질병의 원인이라 생각되는 세포수준 혹은 조직수준의 문제점을 발견하고 적절한 문제해결 방안을 찾아나가는 과정이라고 할 수 있다. 최근에는 보다 정확한 진단 및 적절한 치료방법의 개발을 위해서 최근 각 질병들의 양상을 분자적인 수준에서 유전자 발현양상에 따라 재분류 함으로써 더욱 정확하게 진단기준을 제시할 수 있는 가능성에 대한 많은 연구들이 microarray를 통해서 진행되어 왔다.

Golub 등[21]은 급성 골수성 백혈병(AML)과 급성 림프 모세포형 백혈병(ALL) 환자의 골수 조직을 채취한 후 oligonucleotide chip (affymetrix chip : 현재 gene chip)을 통해 각각의 샘플에 대한 6817개의 gene expression profile을 제작하고 이렇게 구성된 데이터들을 neighborhood analysis 중의 하나인 Self Organizing Maps(SOMs)을 통한 클러스터링을 적용하여 class predictor를 구축하여 DNA 칩을 이용하여 보다 질병진단을 시도했다. 여기에서 각 질병의 predictor로서 적합한 유전자들을 클래스를 구분에 있어서 강력한 유전자들을 선별하였는데 결과적으로 선택된 유전자들은 ALL을 또 다시 T-ALL과 B-ALL로도 구분하여 줄 수 있음을 보여주어 상당히 고무적인 결과로 평가되고 있고 차후에 현재에는 분류하기에 모호해 보이는 다양한 임상적인 양상들을 구분 지을 수 있는 predictor들을 구성할 수 있을 가능성을 제시하였다.

Cho 등[22]은 백혈병 환자의 DNA 칩 데이터를 이용하여 백혈병 subtype을 구분할 수 있는 최적의 유전자(5~8개)를 선택하는 방법을 제시하였다. 이 방법은 기존의 사용되어지는 유전자를 선택하는 기준을 보다 체계적이고 통계이론에 근거를 둔 최적화된 접근 방법을 시도하였다.

Alizadeh 등[23]은 육안 및 몇 가지 단순한 면역조직

생화학적인 진단 표지자들로는 잘 구분 할 수 없어 모호한 기준선에 서 있는 질병들을 각각의 유전자 발현정도를 고려하여 유사성에 따라서 그룹화하거나 구분하기 위한 시도로서 Diffuse Large B-Cell Type Lymphoma를 그 대상으로 하여 환자의 유전체 발현양상에 따라 진단기준이 새로이 구성 될 수 있음을 보여주었다.

이 연구에서는 현재까지도 명확한 진단기준을 제시하지 못하고 있는 Non-Hodgkins leukemia의 한 아형(subtype)인 DLBCL을 연구 대상으로 하였는데 Gloub 등[18]과 마찬가지로 DNA microarray를 통해 각 환자의 gene expression profile을 얻어내고 그 데이터들을 hierarchical clustering를 통해 새로운 진단 기준인 Germinal Center B-like DLBCL과 Activated B-like DLBCL로 진단적인 가치가 있게 분류될 수 있음을 보여주었다. 또한 유전적 발현양상에 따른 환자의 진단기준이 Kaplan-Meier plot을 통해 환자 생존률에 대해서 비교 분석해 볼 때 기존의 DLBCL 진단 기준인 international prognostic index(IPI)를 통한 환자 진단 기준보다 환자의 예후 판정에 있어 더 나은 판별력을 보여줄 수 있다는 것을 확인함으로써 당 기술이 질병진단기준의 확립에 있어서도 좋은 도구가 될 수 있음을 제시하였다. 이는 현재까지의 의학적인 모든 진단 기준들이 다시금 질병에 따른 유전자의 미세변화를 통해 새로이 분류될 수 있음을 보여주는 한 예로서 현재까지 몇몇 면역조직화학적 진단 표지자들을 기반으로 의사의 경험적으로 얻어져 다소 주관성이 개입되어 있던 질병분류의 한계를 극복하여 한 차원 높은 차세대 진단법의 개발가능성을 확인했다는 점에서 그 의미를 찾을 수 있다.

Perou 등[24]은 기존의 조직화학적 진단기준 특히 종양의 악성과 양성을 구분하는 기준이기도 하고 치료법 선택에 있어서도 최대 관심사인 세포의 증식성의 정도를 분석하기 위해서 DNA microarray를 이용한 gene expression profile을 사용하였다. 이는 기존의 세포의 증식성 정도를 판단하는 기준이었던 Ki-67 단백질에 대한 항체동정 방법이나 IFN-regulation의 정도를 판단하는 STAT1 염색법 등의 면역조직화학적 진단 기준들을 gene expression pattern을 이용하여 다시 얻어낼 수 있음을 보여주었다. 이를 위해 Perou et al.은 실제 유방암 환자로부터 얻어진 조직들과 실험실에서 각각 처리된 세포주로부터 얻어진 gene expression profiling 데이터에 바이오인포매틱스가 적절히 이용되어 총 1247개의 gene cluster diagram으로부터 proliferation associated cluster와 IFN-regulated cluster가 추출했고, 각 선별된 cluster들이 기존의 주관성이 들어가기 쉬운 병리현상에도 객관적 기준을 제시할 수 있는 가능성을 보여주었다.

Choi 등[25]은 cDNA microarray를 통해서 환자의 원

발성 간암을 조직학적인 아형인 pseudoglandular type 와 solid type으로 구분할 수 있는 classifier를 Support Vector Machine 기법을 이용하여 찾아냄으로써 다변량 통계분석기법을 통해 다양한 목적으로 질병을 분류할 수 있음을 보여주었다.

최근까지는 연구결과가 특정 몇몇의 질환에 적용되고 있지만 기존 의학에서 경험적으로 얻어진 결과이거나 상당부분 주관적인 판단기준이 적용되고 있는 진단영역에서 microarray와 다변량통계기법을 이용한 분석기법들은 세포의 형태학, 조직학, 병리학 등으로부터 제공되는 다양한 결과들을 체계적으로 담아낼 수 있는 가능성을 보여주고 있기에 환자의 개별 특성에 따른 질병을 이해함에 있어서도 보다 객관적이고 명확한 시각을 제시해 줄 것으로 보인다[26].

7. 결론

포스트 지노믹스 시대가 시작되면서 DNA microarray 및 protein microarray를 이용한 실험들이 현재에도 다양한 장소에서 다양한 목적으로 수행되고 있다. 현재의 고효율 어레이 기술과 생물정보학의 상호이해를 바탕으로 한 적절한 융합은 세포의 분자수준에서의 미세 현상들을 해석해 나감으로써 생물내의 생리학 및 병리학적 인 메커니즘을 이전보다 효율적이고 체계적으로 이해하는 데 있어서 훌륭한 역할을 해내고 있는 것으로 보이며 현재의 기술수준 만으로도 세포내의 모니터링 도구로서 각 유전자 및 단백질의 기능을 규명하는 것과 의료적 이용을 위한 진단 및 신약 발굴을 위한 도구의 하나로 이용될 가능성은 매우 높다고 할 수 있다.

물론 현재까지의 microarray를 이용한 연구 결과들은 array technology로서 얻어지는 데이터 속의 수많은 실험상의 인위적 요소와 실험장치의 한계성들이 내재되어 있기에 아직까지는 극 미세 수준에서의 생명 현상들에 대한 정확한 규명이 전통적인 생물연구방식을 이용한 결과들에 비해 크게 나아진 점이 없어 보일 수도 있다. 또한 개별적인 유전자 및 단백질들의 기능 및 특성들에 대해 현재까지 알려진 정보들이 너무 적기 때문에 그 자체만으로는 실제 적용가치가 많지않게 보일지도 모른다. 그러나 기술적인 문제점들은 점차 개선되어 보다 미세한 생물현상을 정확히 관찰 할 수 있는 관련 기술들이 하나 둘씩 개발되고 있기에 더욱 신뢰도 있는 데이터를 추출해 낼 수 있을 것으로 기대되고 있으며 앞으로 각 유전자 및 단백질들의 기능이 하나 둘씩 밝혀질 것으로 보여지므로 더욱 다양한 생물관련 정보들이 생물정보학자들에게 제공될 것이다.

따라서 고효율 어레이 바이오테크와 접목된 생물정

보학은 머지않아 많은 정보들을 속에 숨어있는 생명현상의 비밀을 풀어내어 체계적이면서도 명확하게 생명현상을 풀어나갈 수 있는 실마리를 우리에게 제공해 줄 수 있을 것으로 기대된다.

참고문헌

1. Thomas, P.S., "Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose", *Proceedings of the National Academy of Sciences*, Vol. 77, No. 9, pp. 5201-5205 (1980).
2. Joseph L. D., Vishwanath R. I., Brown, P. O., "Exploring the metabolic and genetic control of gene expression on a genomic scale", *Science*, Vol. 278, pp. 680-686 (1997).
3. Augenlicht, L. H., Wahrman, M. Z., Halsey, H., Anderson, L., Taylor, J., and Lipkin, M., "Expression of cloned sequences in biopsies of human colonic tissue and in colonic carcinoma cells induced to differentiate in vitro", *Cancer Research*, Vol. 47, No. 22, pp. 6017-6021 (1987).
4. Duggan, D. J., Bittner, M., Chen, Y., Paul, and Trent, J. M., "Expression profiling using cDNA microarrays", *Nature Genetics Supplement*, Vol. 21, pp. 10-14 (1999).
5. Schadt, E. E., Li, C., Su, C., and Wong, W. H., "Analyzing high-density oligonucleotide gene expression array data", *Journal of Cellular Biochemistry*, Vol. 80, pp. 192-202 (2000).
6. Blohm, E. H. and Guiseppe-Elie, A., "New developments in microarray technology", *Current Opinion in Biotechnology*, Vol. 12, pp. 41-47 (2001)
7. Rabilloud, T., "Detecting proteins separated by 2-D gel electrophoresis", *Analytical Chemistry*, Vol. 72, 48A-55A (2000).
8. Patterson, S. D. and Aebersold, R., "Mass spectrometric approaches for the identification of gel-separated proteins", *Electrophoresis*, Vol. 16, pp. 1791-1814 (1995).
9. Haab, B. B., Dunham, M. J., and Brown, P. O., "Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions", *Genome Biology*, Vol. 2, No. 2, research0004.1-0004.13 (2001).
10. Teseng, G. C., Oh, M., Rohlin, L., Liao, J. C., and Wong, W. H., "Issues in cDNA microarray



- analysis : Quality filtering, channel normalization, models of variations and assessment of gene effects”, *Nucleic Acids Research*, Vol. 29, No. 12, pp. 2549-2557 (2001).
11. J. Schuchhardt, D. Beule, A. Malik, E. Wolski, H. Eickhoff, H. Lehrach, and H. Herzfel, “Normalization strategies for cDNA microarrays”, *Nucleic Acid Research*, Vol. 28, No. 10, e47 (2000).
 12. Yang, Y. H., S. Dudoit, P. Luu, and Speed, T. P., “Normalization for cDNA microarray data”, *Technical Report* (2000).
 13. Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein D., and Alnan, R. B., “Missing value estimation methods for DNA microarrays”, *Bioinformatics*, Vol. 17, No. 6, pp. 520-525 (2001).
 14. Lee, M. T., Kuo, F. C., Whitmore, G. A., and Sklar, J., “Importance of replication in microarray gene expression studies : Statistical methods and evidence from repetitive cDNA hybridizations”, *Proceedings of the National Academy of Sciences*, Vol. 97, No. 18, pp. 9834-9839 (2000).
 15. Scherf, U., Ross, D. T., Waltham, M., Smith, L., H., Lee, J. K., Tanabe, L., Kohn, K., Reinhold, W. C., Myers, T. G., Andrews, D. T., Scudiero, D. A., Eisen, M. B., Sausville, E. A., Pommier, Y., Botstein, D., Brown, P. O., and Weinstein, J. N., “A gene expression Database for the molecular pharmacology of cancer”, *Nature Genetics*, Vol. 24, pp. 236-244 (2000).
 16. Weinstein, J. N., Myers, T. G., O’Connor, P. M., Friend, S. H., Fornace, A. J., Kohn, K. W., Fojo, T., Bates, S. E., Rubinstein, L. V., Anderson, N. L., Buolamwini, J. K., van Osdol, W. W., Monks, A. P., Scudiero, D. A., Sausville, E. A., Zaharevitz, D. W., Bunow, B., Viswanadhan, V. N., Johnson, G. S., Wittes, R. E., Paull, K. D., “An information-intensive approach to the molecular pharmacology of cancer.” *Science*, 275 : 343-349. (1997).
 17. Hughes, T. R., Marton, M. J., Jones, A. R., Roberts, C. J., Stoughton, R., Armour, C. D., Bennett, H. A., Coffey, E., Dai, H., He, Y. D., Kidd, M. J., King, A. M., Meyer, M. R., Slade, D., Lum, P., Stepaniants, S. B., Shoemaker, D. D., Gachotte, D., Chakraburtt, K., Simon, J., Bard M., and Friend, S. H., “Functional discovery via a compendium of expression profiles”, *Cell*, Vol. 102, pp. 109-126 (2000).
 18. Iyer, V. R., Eisen, M. B., Ross, D. T., Schuler, G., Moore, T., Lee, J. C. F., Trent, J. M., Staudt, L. M., Hudson, J., Boguski, M. S., Lashkari, D., Shalon, D., Botstein, D., Brown, P. O., “The transcriptional program in the response of human fibroblasts to serum”, *Science*, Vol. 283, pp. 83-87 (1999).
 19. Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D., and Futcher, B., “Comprehensive identification of cell cycle-regulated genes of the yeast *saccharomyces cerevisiae* by microarray hybridization”, *Molecular Biology of the Cell*, Vol. 9, pp. 3273-3297 (1998).
 20. Alter, O., Brown, P. O., and Botstein, D., “Singular value decomposition for genome-wide expression data processing and modeling”, *Proceedings of the National Academy of Sciences*, Vol. 97, pp. 10101-10106 (2000).
 21. Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S., “Molecular classification of cancer : class discovery and class prediction by gene Expression Monitoring”, *Science*, Vol. 286, No. 15, pp. 531-537 (1999).
 22. Cho, J., Lee, D., Park, J. H., Kim, K., and Lee, I., “Optimal approach for classification of acute leukemia subtypes based on gene expression data”, *Biotechnology Progress Submitted*.
 23. Alizadeh, A. A., Eisen, M. B., Davis R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., Powell, J. I., Yang, L., Marti, G. E., Moore, T., Hudson, J., Lu, L., Lewis, D. B., Tibshirani, R., Sherlock, G., Chan, W. C., Greiner, T. C., Weisenburger, D. D., Armitage, J. O., Warnke, R., Levy, R., Wilson, W., Grever, M. R., Byrd, J. C., Botstein, D., Brown, P. O., and Staudt, L. M., “Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling”, *Nature*, Vol. 403, pp. 503-511 (2000).
 24. Perou, C. M., Jeffrey, S. S., Rijn, M. V., Rees, C. A., Eisen, M. B., Ross, D. T., Pergamenschikov, A., Williams, C. F., Zhu, S. X., Lee, J. C. F., Lashkari, D., Shalon, D., Brown, P. O., and Botstein, D., “Distinctive gene expression patterns in human mammary epithelial cells and breast cancers”, *Proceedings of the National Academy of Sciences*,

Vol. 96, pp. 9212-9217 (1999).

25. Choi, S. W., Lee, D., Lee, I., Kim, M., Kim, J., and Park, J. H., "Classification of hepatocellular carcinoma based on the gene expression from cDNA microarrays using SVM and CDA", *AICHe*

2001 Annual Meeting, (2001).

26. Masters, J. R. W., Lakhani, S. R., "How diagnosis with microarrays can help cancer patients", *Nature*, Vol. 404, No. 27, pp. 921 (2000).

회원혜택과 입회절차

회원 혜택

- ① 논문 및 학회지를 인터넷에서 무료로 다운로드 받을 수 있습니다.
- ② 논문지 및 학회지, 학술행사에 투고된 논문을 인터넷상에서 영구보존 및 관리해드립니다.
- ③ 학회지, 논문지, 학술대회 논문을 투고할 수 있습니다.
- ④ 자신이 투고한 논문심사의 모든 과정을 인터넷을 통하여 신속하게 검색 할 수 있습니다.
- ⑤ 인터넷을 통하여 학회에서 제공하는 정보(회원인물검색, 학술행사, 해외유명사이트정보 등)를 제공받을 수 있습니다.
- ⑥ 본 학회에서 주관하는 학술대회에 대한 정보를 얻으실 수 있으며, 참가 시 할인혜택을 받으실 수 있습니다.
- ⑦ 회원 개개인의 연구업적을 개인별로 데이터베이스화하여 대외에 홍보할 수 있습니다.
- ⑧ 기타 본 학회 정관에 명시된 정회원 및 학생회원의 권리를 부여합니다

회원의 입회절차

- ① 입회원서 작성(원서에 직접작성 또는 홈페이지 <http://icase.or.kr> 내에서 작성)하신후,
- ② 입회비 및 연회비 납부 (정회원: 50,000원(입회비 10,000원 포함),
학생회원: 25,000원(입회비 5,000원 포함))
- ③ 회원등록확인(홈페이지내 회원검색 에서 가입여부 확인)
- ④ 위의 회원혜택부여 및 회원관리

회비납부 안내

홈페이지 <http://icase.or.kr>을 통한 납부

- 홈페이지내 우측상단의 전자결제서비스 에서 신용카드 전자결제 가능

온라인 납부처

- 국민은행(영업3부 지점) : 760-01-0038-589
- 외환은행(강남역 지점) : 057-22-03148-4
- 한미은행(역삼동 지점) : 102-53439-253
- 한빛은행(테헤란로 지점) : 169-189026-13-002
- 지로용지: 첨부지로용지 사용