

생명정보학(bioinformatics)분야에서의 공학자들의 역할

이 동 권, 조 지 훈, 이 인 범
 포항공과대학교 화학공학과

1. 서론

최근 생명공학분야의 측정기기 및 실험장치들의 비약적인 발전으로 인해 생명공학분야의 발전속도가 상상할 수 없을 만큼 빨라지고 있다. 이러한 실험장치들에서 나온 데이터들을 분석하고 이를 생명공학분야에 응용하기 위해서는 과거보다 더 많은 논리적이고 수학적인 사고가 필요하게 되었다. 이로 인해 생명공학분야에서 학문의 통합화는 더욱 가속화되고 있으며 많은 수학자, 물리학자, 컴퓨터 공학자 그리고 시스템 엔지니어들이 이 새로운 분야를 발전시키기 위해 노력하고 있다. 이러한 노력들은 기존 의학분야에서 결과에 대응하는 식의 방법론법과는 달리 좀 더 효율적으로 환자를 치료할 수 있고 질병의 근본 원인을 제거할 수 있는 방법을 제공할 수 있을 것으로 예상된다. 예를 들어, 과거에는 당뇨병 환자들에게 주사기를 이용하여 일정량의 insulin을 주입하는 방식(open-loop control)에서 시스템 엔지니어의 노력으로 피드백 제어기(feedback controller)를 이용하여 환자의 상태를 확인하며 insulin을 주입하는 방식을 취하고 있다. 이는 glucose의 혈중 농도를 간접적으로 측정할 수 있는 혈당측정기기, 제어를 위한 insulin 펌프의 발달로 인해 가능할 수 있었다. 또한, 과거에는 사용할 수 없었던 고효율 어레이 기술, protein microarray 그리고 분광학적 기기 분석법을 이용하여 β -cell과 같이 당뇨병에 영향을 줄 것으로 예상되는 부분의 메커니즘을 연구하고 이를 통해 근본적인 문제점을 해결하고자 하는 방향으로 연구가 진행 중이고 이를 위해서는 더 많은 통계학적 공학적 기술이 필요할 것으로 예상된다.

앞에서 언급한 약물 주입에 있어서의 공정 제어기법 활용은 주로 제1형 당뇨병 환자의 insulin 주입 시스템을 이용한 혈당제어와 수술환자를 위한 마취제어 등이 있고 이외에도 인공신장제어와 같은 분야도 있다. 제1형 당뇨 환자의 제어에 관한 연구는 최근 Parker et al. [1]에 의해 많은 연구가 진행되고 있다. 1980년대까지 많은 연구자들이 노력으로 인해 당뇨병환자의 수학적 모델을 구할 수 있었고 이를 바탕으로 Lehmann과 Deutsch [2]는 음식물의 양이 인체에 미치는 영향을 수학적으로 표현한 외란 모델을 구성하였다. 이러한 수학적 모델을 이용하여 다수의 제어 알고리즘이 제안되었다. 그러나, insulin과 glucose사이 gain의 비선형성으로 인해 가장 널리 사용되는 PID 제어

기를 직접 사용하는 데는 커다란 어려움이 있고 이를 극복하기 위해 최적 제어 기법이나 모델 예측 제어기법을 적용하는 연구가 활발하게 진행 중이다 [1-3]. 마취제어 문제는 제1형 당뇨병 환자의 glucose 농도제어와 달리 다중 작업(multitask) 처리를 위한 제어 기법이 필요하고 이에 대한 연구는 Martin [4]과 Gentilini et al. [5]에 의해 진행되고 있다. 이외에도 고순도를 요구하는 제약 공장을 위한 공정 제어 기법에 관한 연구도 활발히 진행 중이다.

시스템 엔지니어나 공학자들이 개척하기 위해 노력하고 있는 또 하나는 생물학적 신호로부터 유용한 정보를 추출하기 위한 신호와 영상을 처리하는 signal processing과 관련된 분야이다. 이 분야는 microarray에서 나온 영상(image) 데이터나 2-D gel 실험을 통해 생성된 데이터를 intensity 데이터로 변환하는 분야에서도 사용되며 특히, 다양한 한의학이나 서양의학분야에서 이용되는 진단 장치에 의해 획득된 인체 신호들을 처리하는 분야에 많이 적용되리라 예상된다. 진단장치에 획득된 데이터를 다루는 분야를 biomedical signal processing이라 하며 주로 연구하고 있는 분야는 EEG (electroencephalographic)를 이용하여 간질환자가 발작하기 전에 이를 예측하는 분야와 ECG(electrocardiogram)를 사용하여 정확한 심장병 종류를 분류하거나 태아의 심장 이상을 감지하는 분야가 있다 [6-7]. 이 외에도 signal processing과 관련된 분야는 유방암환자의 단층 촬영 데이터를 이용하여 암인지 정상인지를 판단하거나 암의 subtype을 구분하는 분야도 있다.

본 논문에서 주로 다루게 될 cDNA microarray나 oligonucleotide chip (DNA 칩)과 같은 고효율 어레이 기술(high throughput array technology)은 biotechnology 분야에서 방대한 데이터의 생산을 가능하게 했고 이로 인해 연구 방법의 패러다임이 변화하고 있는 실정이다. 기존 약성 종양의 분류는 주로 형태학·임상학적 또는 분자 수준에서의 몇 개의 매개 변수를 이용하여 분류해 왔다. 그러나 유전·발생학적으로 비슷한 형태를 가진 종양도 화학요법이나 방사선 치료에 대해 다른 효과를 보여주는 경우가 있기 때문에 위와 같은 방법이 신뢰도가 높다고 말하기는 힘들다. 또한 병리학적 또는 현미경을 이용하는 기존의 분석 방법은 시간이 많이 들고 실험자의 주관이 개입될 수도 있기 때문에 효율적이지 못하다. 이에 비하여, 유전자는 이러한 종양의 근본적인 특징들에 대한 정보를 내포하고 있기

때문에 유전자 수준에서의 DNA 칩 분석은 기존 방법에 비해 정확하고 효율적이라고 할 수 있다 [8-9]. 또한 이전까지의 연구에서 한 연구자가 유전자 하나의 기능과 역할을 밝히는 데에는 많은 시간과 노력을 필요로 했으나, DNA 칩은 한번의 실험으로 수 천 혹은 수 만의 유전자 발현양을 동시에 한번의 실험으로 데이터를 획득할 수 있다는 장점을 가지고 있다. 특히, 생명 현상과 암 발생 및 분화와 관련된 메커니즘은 매우 복잡하여 이에 대한 이해나 분석이 어려운 경향이 있는데 다변량 통계적 분석법 및 데이터 마이닝 기법을 접목하여 이를 해석할 경우 많은 시간 단축과 정확한 결론을 제공하므로 이를 활용한 연구가 활발히 진행되고 있다. 공정 데이터들과는 달리 샘플에 비해 상대적으로 많은 변수(유전자)가 있어서 기존의 데이터 분석법을 그대로 사용하기에는 어려움이 있고 특히, 샘플 자체의 편차, 실험상의 오차, 실험 장치에서 발생하는 외란이 섞여 있으므로 데이터를 규격화하고 보정 작업이 고급 데이터 분석 전에 선행되어야 한다.

앞에서 언급한 내용들 중 생명공학이나 생물학전공자가 아닌 공학자, 특히 시스템 엔지니어와 화학공학 엔지니어들이 연구하고 있는 생명공학관련 분야들을 정리해 보면 그림 1과 같다. 이 중 본 논문에서는 주로 microarray 기법을 이용한 질병 분석 방법에 대해 설명하고 다른 분야로의 유기적 결합방법과 확장에 대해 소개하도록 하겠다.

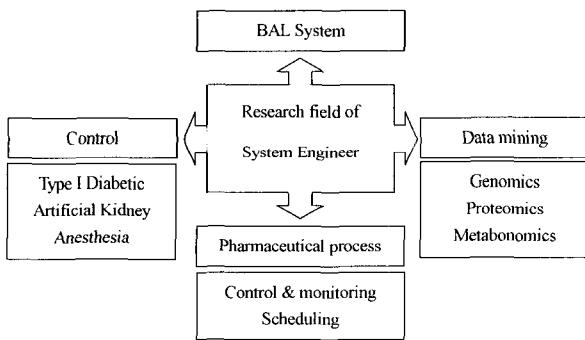
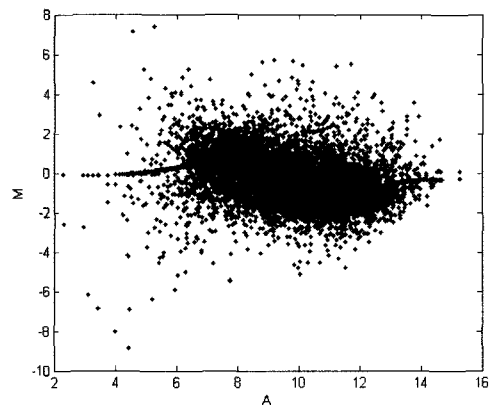


그림 1. Research field of System engineer.

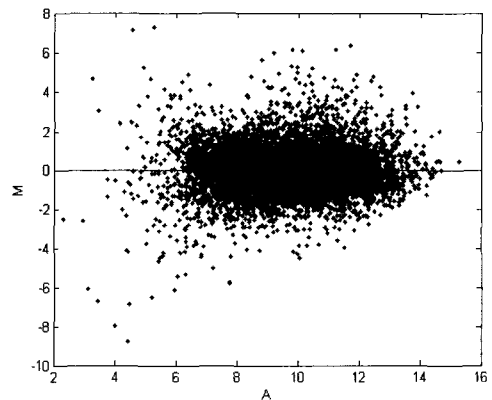
2. Microarray 관련 연구 분야

Microarray 데이터 분석과 관련된 분야는 크게 두 가지 분야로 나눌 수 있다. 첫 번째 분야는 데이터의 정규화(normalization), faulty probe 감지, 실험과정 중 missing value의 처리 등을 다루는 low-level 데이터 분석과 판별 분류나 군집화를 주목적으로 하는 high-level 데이터 분석으로 나눌 수 있다. microarray 데이터의 정규화는 데이터에 포함되어 있는 외란이나 실험상의 오차, 편차 등을 보정해 주는 과정이라고 할 수 있다. 정규화 되지 않은 데이터를 사용하게 될 경우 왜곡된 분석 결과를 도출하는 오류

를 범할 수 있으므로 high-level 데이터 분석을 수행하기 전에 적절한 방법을 통해서 반드시 정규화를 거쳐야 한다. cDNA chip의 정규화를 위해서는 주로 locally weighted regression (LOESS) 방법을 사용하며 oligonucleotide chip에 대해서는 다양한 연구결과가 발표되어 있으나 최근 발표된 연구 중에서 quantile normalization 방법이 효과적인 것으로 알려져 있다. 그림 2는 cDNA microarray 데이터를 LOESS를 사용하여 정규화한 결과를 보여주고 있다. 이와 같이 데이터의 정규화를 수행한 후 목적에 맞는 high-level 데이터 분석을 진행하게 된다.



(a) 정규화하기 전 데이터



(b) 정규화한 후 데이터

그림 2. comparison work of between before normalization and after one.

2.1. 유전자 선택과 판별분류

High-level 데이터 분석을 이용하여 주로 연구하고 있는 분야는 악성 종양에 관한 연구 분야라 할 수 있다. 악성 종양은 많은 subtype을 가지고 있고 각각의 subtype은 다른 처방 계획이 요구된다. 따라서 정확한 악성종양 subtype의 분류는 성공적인 치료를 위해 중요하다 [10-11]. 만약 각 환자 고유의 상태가 분명하게 정의되었다면 지도 학습법 (supervised learning)을 통해서 classifier를 설

계할 수 있다. 그러나 만약 이러한 고유 상태를 알 수 없는 경우 데이터 분석의 목적은 데이터의 similarity measure나 correlation을 이용하여 cluster, 즉 군집을 조사하는 것이 된다. 전자에 관한 연구는 고급화된 비선형 알고리즘을 이용하여 classifier를 구성하는 쪽에서 classifier 설계 위한 전단계인 유전자의 선택(gene selection)으로 연구방향이 옮겨가고 있다. 유전자 발현 데이터들은 샘플(환자)의 숫자에 비해 극단적으로 많은 변수(유전자)들을 가지고 있어서 보편적인 데이터 마이닝 기술을 이러한 데이터에 직접적으로 적용할 수 없다. 이러한 이유로 유전자 발현 데이터의 분석은 curse of dimensionality나 over-fitting과 같은 문제를 피하기 위해 각 타입의 질병과 상관관계가 높은 유전자를 추출해 내는 유전자 선택과정(gene selection)이 반드시 필요하다. 초기의 연구에서는 각각의 유전자를 하나의 변수로 보고 단변수 관점에서의 통계적 검정, 즉 T-test나 ranksum test와 같은 방법을 이용해서 두 그룹 사이에서 통계적으로 차이를 보이는 유전자들을 선택하는 방법을 취했다. 그러나 유전자 발현 데이터가 외란에 의한 영향이 많이 포함되어 있고, 소위 outlier라고 하는 극단적인 오류가 포함되는 경우가 많기 때문에 이러한 통계적 검정은 왜곡된 결과를 도출할 수 있는 가능성을 가지고 있다. 또한, multiple classification 문제에 직접 적용하기 힘들고 multiple testing의 문제를 해결하기 위해서는 시간이 많이 필요한 보정작업이 필요하다는 단점을 가지고 있다 [10].

이러한 통계적 접근 방법 외에도 유전자 선택(gene selection)에 대해 많은 연구들이 이루어졌다. Golub et al. [11]은 유전자와 악성 종양 subtype 사이의 관계를 측정하기 위한 척도로서 signal-to-noise ratio를 사용하였다. Guyon et al. [12]은 support vector machine에 recursive feature extraction을 사용하는 방법을 제안했다. 또한 Cho et al. [13]은 class들 사이에서 차이가 나타날 뿐 아니라 동일 class내에서는 발현양의 차이가 크지 않은 유전자 선택법을 제안했다. 그림 3은 Cho et al. [13]의 방법을 small round blue cell tumors (SRBCT) 데이터에 적용한 결과를 heat map을 이용하여 구현한 것이다.

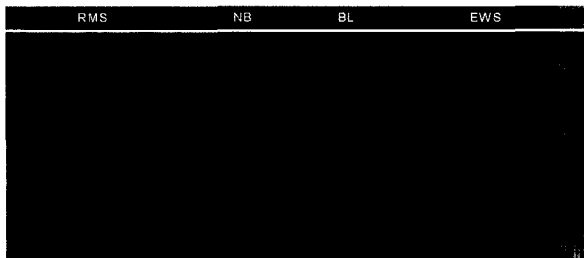


그림 3. Heat map of selected genes.

Classifier를 구현하는데 있어서도 많은 연구가 이루어져왔다. Golub et al. [11]은 각각의 유전자의 weight를

이용해서 단변수 classifier의 집합체를 이루는 weighted voting이라는 알고리즘을 제시하였다. 이후에는 고차원 데이터를 다룰 수 있는 다변량 통계적 방법을 이용하는 방법들이 발표되었는데, Nguyen and Rocke [14-15]가 partial least squares (PLS)를 이용하여 암의 subtype을 구분 지을 수 있는 classifier를 제안한 것이 대표적이라 할 수 있다. 그 외에도 확률론에 기반하여 Bayesian classifier를 제안하는 등의 연구가 이루어져 왔다. 최근에는 데이터의 차원에 무관하면서도 비선형적인 거동을 설명하는데 효과적인 kernel machine을 사용한 연구가 발표되고 있다. Guyon et al. [12]은 support vector machine을 이용한 classifier를 제안하였으며, Cho et al. [13]은 kernel machine의 장점을 살리면서 multiple classification에도 쉽게 적용할 수 있는 kernel Fisher discriminant analysis를 이용하였다.

2.2. 군집분석

Eisen et al. [16]은 효모의 발아 과정과 인간 섬유 세포의 혈청 반응을 microarray 실험을 통해 데이터화하였고 군집화 분석을 수행하였다. 이 연구는 군집화 분석이 유전자 발현 데이터의 해석에 유용하게 사용될 수 있다는 가능성을 보여주었다. 그 후, Alon et al. [17]은 대장암 환자의 유전자 발현 패턴을 oligonucleotide microarray를 이용해서 얻어 내고, 군집화 분석을 수행함으로써 정상인과 대장암 환자를 구분하였으며, 대장암에 특이하게 발현되는 유전자들을 찾아내는 등 암의 진단 목적으로 군집화 분석이 사용될 수 있음을 보였다. 이외에도 Alizadeh et al. [18]은 Non-Hodgkin's lymphoma의 아형인 diffuse large B-cell lymphoma(DLBCL) 환자의 임상적 치료효율이 다른 원인을 찾기 위해 microarray 실험을 통해 얻은 유전자 발현 프로파일과 계층적인 군집화 방법을 사용하였다. 물론 이들의 연구 전에도 DLBCL의 subgroup을 구분하기 위한 노력이 행해졌고 그 결과 많은 수의 단일 biomarkers들을 찾았으나 subgroup 사이의 차이가 무엇인지에 대한 명확한 설명을 하기에는 부족한 부분이 있었다. 그러나, Ailzadeh et al. [18]의 연구를 통해 germinal centre B-like DLBCL와 activated B-like DLBCL 사이에 발현양이 다른 수백 개의 유전자를 찾을 수 있었고 이러한 유전자들은 각 그룹과 B-cell 분화와 활성화 단계를 구별하는 것과 관계가 있다는 것을 밝혔다. Rosenwald et al. [19]는 DLBCL에 또다른 그룹이 하나 더 존재할 수 있다는 연구 결과를 발표하였다.

군집화 경향을 분석하기 위해서는 주성분 분석(principal component analysis)와 같은 투영 방법들을 사용할 수도 있으나 샘플의 숫자에 비해 변수들이 상대적으로 많아서 직접 사용하기는 힘들다. 하나의 유전자에 대해 전체 샘플 평균에 비해 변이성이 작은 유전자를 제거해 나가는 방식의

gene elimination을 군집화 분석과 주성분과 함께 쓰는 방식이 효과가 있을 것으로 기대된다. 또한 군집화 분석을 한 후에도 통계적 수학적 안정성을 보장하기 위해 유효성 검증을 함께 실시하는 것이 좋을 것으로 기대된다.

2.3. 암의 조기 진단과 신약 개발에의 적용

많은 경우에 있어서 악성 종양은 악성 종양 세포들이 다른 신체 장기로 전이되었거나 조직으로 침투될 때까지 진단되거나 치료되지 못하고 있다. 그리고 일단 폐암, 유방암, 대장암, 구강암과 악성 종양들이 원래의 조직으로부터 다른 조직으로 퍼져나가면 현재의 존재하는 일반적인 치료법으로 치료 성공이 매우 제한적이다. 따라서 많은 연구자들은 초기에 이를 진단할 수 있고 관측 가능한 biomarker들을 찾기 위해 노력하고 있다. 유전자의 돌연변이, 유전자의 전사나 유전정보의 번역에서의 변질과 그들의 protein 생산물에서의 변질 등은 질병에 대한 특정 biomarker로서의 역할을 할 수 있다. 특히, microarray와 같은 기술의 발달은 악성 종양 cell line에서 일어날 수 있는 유전자 발현양의 전체적인 변화량이나 특정 변화량을 빠르게 걸러내는 작업을 가능케 했다. 최근에는 조기 악성 종양 진단의 효율을 높이기 위해 단백질 수준에서 실험과 분석을 통해 biomarker들을 찾기 위한 연구들이 더 많이 수행되고 있다.

유전자 발현 분석 기술은 biomarker, drug target 규명 그리고 실제 작용 메커니즘의 규명과 같은 신약 발굴과 개발의 다양한 분야에서 유용하게 사용될 수 있다. Target 규명에서의 도움은 많이 알려진 바와 같이 microarray는 정상 조직과 비정상 조직의 유전자 발현들을 비교하여 차이가 나는 유전자를 찾아낼 수 있고 특히, 갓가지 질병에서 up-regulated 되었거나 down-regulated된 유전자를 찾는 데 사용될 수도 있다 [20-21]. 다양한 종양으로부터의 조직이나 cell line과 연관된 연구논문들도 있다. microarray를 이용하여 다수의 lead compounds들에 대해 효율과 독성에 대한 연구도 수행되었다. 특히, Scherf et al. [22]은 NCI (National Cancer Institute)에서 보유하고 있는 60개의 암 조직 cell line과 110여 종의 약물의 반응성 관계를 유전자 발현 데이터를 통해서 도출해냄으로써 drug resistance에 대한 연구를 분자 수준에서 수행하였다. 이러한 연구를 통하여 암 조직이 특정 약물에 얼마나 민감하게 반응하는지를 알아낼 수 있고, 차후에 효과적인 항암제의 개발에도 크게 기여할 수 있다는 가능성을 보여주었다.

3. 새로운 도전

3.1. Metabonomics와 Global cellular analysis

Nicholson et al. [23]은 질병 과정, 독성 반응 그리고 유전자 축진 등에서의 복잡한 물질 대사의 중요성을 분석하

기 위해 데이터에 기반한 분석적인 접근이 필요하다고 주장했다. 왜냐하면 유기체 내부에서 생체 고분자의 조직과 제어는 상호 의존적인 여러 반응에 의해서 이루어지며 환경적인 요소와 외부의 스트레스에 영향을 받는 등 복잡한 양상을 보이는데, 이러한 생물학적 활동이 다변수 데이터 형태로 표현될 수 있기 때문이다. Mass spectroscopy, gas chromatography/mass spectroscopy 그리고 NMR과 같은 분광학적 기기에 의해 생체 물질의 물질 대사의 데이터를 수집한 후, 통계적 분석법이나 패턴 인식 (pattern recognition) 방법들을 이용하여 이를 분석하면 중요한 정보를 도출해 낼 수 있을 것이라고 Nicholson et al. [23]은 밝히고 있다. 이러한 분야를 metabonomics라 하는데, metabonomics는 향후 기형아 검사, 신약 개발, 약제 독성 및 내성 분석과 같은 분야에서 유용하게 사용될 수 있을 것으로 보인다 [24].

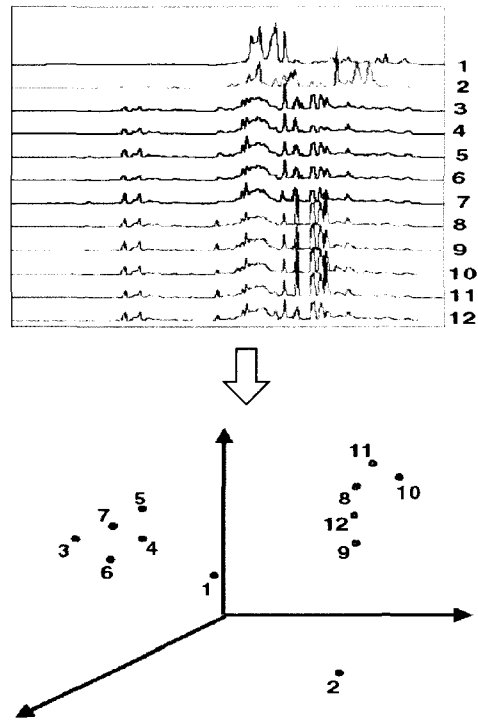


그림 4. NMR spectroscopy and data analysis.

Bochner [25]는 microarray, protein chip 그리고 phenotype microarray를 이용하여 유전자형질(genotype)과 외적 형질(phenotype) 사이의 관계 평가에 대한 연구를 수행하였다. 이 연구에서 phenotype microarray의 기본적인 사용 목적은 수천 개의 외적 형질들을 테스트하는 것이다. 이러한 정보들은 유전체학(genomics)와 단백질체학(proteomics) 연구들을 보완해 줄 수 있을 것으로 기대된다. 특히, Bochner [25]와 같은 연구뿐만 아니라 다양한 실험들을 유기적으로 연결하여 상호 연관성을 규명해 나가는 작업들은 생명 공학분야의 더 빠른 발전에 필수 불가결한 요소라 할 수 있다.

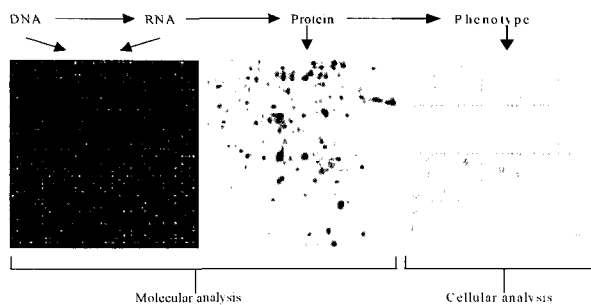


그림 5. Global cellular analysis.

3.2. 생체의 네트워크 분석

생명정보학은, 유전자나 단백질의 서열을 이용해서 유전체 구조나 기능부위를 해석하는 서열 분석과 유전체의 발현 패턴을 이용해서 profiling이나 다형성을 분석하는 기능 유전체학 등 여러 세부분야로 분류할 수 있다. 본 논문에서는 생명정보학의 주요 분야 중 하나인 고효율 어레이 기술 (high throughput array technology)과 접목된 데이터 마이닝 기법 (data mining approach)이나 다변량 통계 분석 기법 (multivariable statistical approach)에 대해 살펴보았다. 그러나 이러한 생명정보학을 이용한 연구들이 실질적으로 신약 개발이나 암과 같은 난치병이나 유전병의 진단치료 등에 공헌한 바는 기대했던 바와는 달리 상대적으로 미약했다. 이러한 근본적인 문제들을 극복하기 위해 기능 유전체학의 마지막 단계라고 할 수 있는 유전체 발현 조절 네트워크 분석에 대한 연구가 수행되고 있다.

생체의 네트워크 분석 (network analysis and reverse engineering)은 기본적으로 생체 데이터를 수집하고, 데이터와 가장 잘 부합하는 네트워크 모델을 만들어 나가는 과정이라고 할 수 있다. 이를 위해서는 기본적으로 모델링과 데이터 마이닝 기법을 필요로 하게 된다. 네트워크 분석은 크게 두 가지 세부분야로 나눌 수 있는데, 첫 번째는 데이터를 기반으로 한 기법 (data-driven approach)로서 high-throughput 기술을 이용하여 많은 양의 데이터를 수집한 후, 데이터 마이닝 기법을 이용하여 네트워크 구조를 모델링 하는 방법이고, 두 번째는 모델을 기반으로 한 기법 (model-based approach)으로서 생체내의 네트워크를 물질 수지와 같은 미분 방정식 형태로 모델링하고 적합한 모델을 구현해 나가는 방법이다 [26]. 첫 번째 접근 방법을 위해서는 네트워크 구성에 참여할 것으로 추측되는 대상 유전자 (단백질, 혹은 대사물질)를 선정해내는 일이 선행되어야 한다. 이후에는 hidden Markov model이나 Bayesian network을 이용하여 네트워크의 구조를 모델링하고, 실제 데이터와 적합도를 조사하게 된다. 두 번째 접근 방법은 deterministic modeling이라고 불리는 방법으로서 물질 수지 등을 이용, 상대적 유전자 발현양이나 대사 물질의 플럭스(flux)를 모사하여 네트워크 구조를 규명하는 방법이다. 데이터를 기반으로 한 연구 방식은 모델을 기반으로 한 접근 방식보다는

상대적으로 쉽고 간편하나 drug target이나 biomarker로 개발하기 위해 수많은 시간을 투자해야 하고 실패했을 경우 그 이유를 설명하기가 어렵다는 특징을 가지고 있다. 이와 반대로 모델을 기반으로 한 연구방식은 이에 대한 인과관계 설명이 상대적으로 쉽지만 이를 연구하기 위해서는 생체 시스템의 조절 메커니즘을 완벽히 이해해야 할 뿐만 아니라 수많은 자료의 해석에 많은 시간이 소요되고 모델링을 위한 수학적이나 공학적 방법론이 필요하여 전자의 방식보다는 어렵다고 할 수 있다.

4. 토론 및 결론

분자 생물학의 발전과 인간 유전체의 해석으로 인해 많은 양의 생명정보와 관련된 데이터가 생산되기 시작하면서 이러한 데이터들로부터 유용한 정보를 추출해 내기 위한 많은 시도가 이루어지고 있다. 과거의 의학 분야에서 주로 사용 하였던 실험분석 방법은 방대한 양의 데이터를 감당해내기 어렵기 때문에 데이터의 효율적인 해석을 위하여 컴퓨터 공학, 수학, 통계학 등과의 접목이 필요하게 되면서 생명정보학, 즉 bioinformatics라는 학문이 대두되었다. 이러한 생명정보의 공학적 분석은 생물체의 생리학적, 병리학적인 메커니즘을 밝혀내고 이해하는데 시간적으로나 경제적으로 큰 기여를 할 것으로 보인다. 본 논문에서 언급한 바와 같이 오랫동안 밝혀지지 않았던, 그리고 실험적으로는 도저히 구현하기 어려운 규모의 여러 가지 현상들이 DNA칩 데이터를 이용한 연구들에 의해 관찰되고 규명되었으며, 특히 암의 메커니즘이나 drug target을 찾아내는 일에 효과적으로 이용되고 있다. 앞에서 다루었던 DNA칩 데이터의 분석 외에도 proteomics나 metabonomics, toxicogenomics 등 여러 분야에서 공학자들의 역할을 필요로 하고 있음은 두말할 나위가 없다.

한가지 간과하지 말아야 할 것은 공학자들의 역할이 중요해지는 만큼, 기존의 생명과학이나 의학 전공자들과의 협력 또한 반드시 필요하다는 점이다. 예를 들어, 데이터의 분석을 통해서 drug target을 찾았다면 그 이후에는 반드시 임상적인 실험을 통해서 drug target candidate의 타당성과 안정성, 그리고 실용성을 입증해야만 비로소 의미 있는 하나의 연구가 완결되는 것이기 때문이다. 실제로 해외 유명 대학의 생명정보학 연구팀은 공과대학과 의학대학, 그리고 수학과나 통계학과를 포함한 자연과학대학의 연구진을 아우르는 조직을 갖추고 있다. 따라서 공학자들은 post-genome시대라는 새로운 도전을 맞이하면서 자신의 역량을 키울 뿐만 아니라 어떤 연구분야의 전공자와도 긴밀하게 공조할 수 있는 open mind를 갖춰야 할 것이다. 생명정보학이라는 광범위한 분야에서 공학자들의 활발한 연구활동을 통해 인간 질병에 관한 연구는 보다 수학적이고

논리적인 방법에 의해 수행될 것이며, 질병 치료를 위한 메커니즘의 규명, 치료법이나 신약의 개발이 지금보다도 훨씬 효율적으로 이루어질 것으로 기대한다.

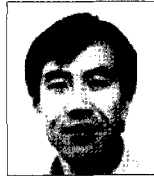
참고문헌

1. Parker, R. S., F. J. Doyle III, J. H. Ward and N. A. Peppas, "Robust H Glucose Control in Diabetes Using a Physiological Model," *AICHE J.*, vol. 46, 2000.
2. Lehmann, E. D., and T. Deutsch, "A Physiological Model of Glucose-Insulin Interaction in Type 1 Diabetes Mellitus," *J. Biomed. Eng.*, vol. 14, 1992.
3. Ollerton, R. L., "Application of Optimal Control Theory to Diabetes Mellitus," *Int. J. Control*, vol. 50, 1989.
4. Martin J. F., Fuzzy control in anesthesia. *J. Clin. Monitoring*, vol. 10, 1994.
5. Gentilini, A., W. F. Christian, A. H. Glattfeder, M. Morari, T. J. Sieber, R. Wymann, F. W. Schnider, and A. M. Zbinden, "Multitasked Closed-loop Control in Anesthesia," *IEEE Eng. in Med. & Biol.*, vol. 20, 2001.
6. Litt, B. and J. Echaz, "Prediction of epileptic seizures," *Lancet Neurol.*, vol. 1, 2002.
7. Tarvainen, M. P., P. O. Rantaho, and P. A. Karjalainen, "An advanced detrending method with application to HRV analysis," *IEEE Trans Biomed Eng.*, vol. 49, 2001.
8. Friend, S. H. and R. B. Stoughton, "The magic of microarray," *Scientific American*, February, 2002.
9. Hanahan, D. and R. A. Weinberg, "The Hallmarks of Cancer," *Cell*, vol. 100, 2000.
10. Dudoit, S., Y. H. Yang, T. P. Speed, and M. J. Callow, "Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments," *Stat. Sinica.*, vol. 12, 2002.
11. Golub, T. R., D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander, "Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring," *Science*, vol. 286, 1999.
12. Guyon, I., J. Weston, S. Barnhill, and V. Vapnik., "Gene selection for cancer classification using support vector machine," *Mach. Learn.*, vol. 46, 2002.
13. Cho, J.-H. Lee, D. Park, J. H. and Lee, I.-B., "New gene selection method for classification of cancer subtypes considering within-class variation," *FEBS Lett.*, vol. 551, 2003.
14. Nguyen, D. V. and Rocke, D. M., "Tumor classification by partial least squares using microarray gene expression data," *Bioinformatics*, vol. 18, 2002.
15. Nguyen, D. V. and Rocke, D. M., "Multi-class cancer classification via partial least squares with gene expression profiles," *Bioinformatics*, vol. 18, 2002.
16. Eisen, M. B., Spellman, P. T., Brown, P. O. and Botstein, D., "Cluster analysis and display of genome-wide expression patterns," *Proc. Natl. Acad. Sci.*, vol. 95, 1998.
17. Alon, U., Barkai, N., Notterman, D. A., Gish, K., Barra, Y., Mack, D. and Levine, A. J., "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays," *Proc. Natl. Acad. Sci.*, vol. 96, 1999.
18. Alizadeh, A. A., M. B. Eisen, R. E. Davis, C. Ma, I. S. Lossos, A. Rosenwald, J. C. Boldrick, H. Sabet, T. Tran, X. Yu, J. I. Powell, L. Yang, G. E. Marti, T. Moore, Jr., J. Hudson, L. Lu, D. B. Lewis, R. Tibshirani, G. Sherlock, W. C. Chan, T. C. Greiner, D. D. Weisenburger, J. O. Armitage, R. Warnke, R. Levy, W. Wilson, M. R. Grever, J. C. Byrd, D. Botstein, P. O. Brown, and L. M. Staudt, "Distinct Types of Diffuse Large B-cell Lymphoma Identified by Gene Expression Profiling," *Nature*, vol. 403, 2000.
19. Rosenwald, A., Wright, G., Chan, W. C., Connors, J. M., Campo, E., Fisher, R. I., Gascoyne, R. D., Muller-Hermelink, H. K., Smeland, E. B. and Staudt, L. M. "The use of molecular profiling to predict survival after chemotherapy for diffuse large B-cell lymphoma," *New. Engl. J. Med.*, vol. 246, 2000.
20. Khan, J., Simon, R., Bittner, M., Chen, Y., Leighton, S. B., Pohida, T., Smith, P. D., Jiang, Y., Gooden, G. C., Trent, J. M. and

Meltzer, P. S. "Gene expression profiling of alveolar rhabdomyosarcoma," *Cancer Res.*, vol. 58, 1998.

21. Perou, C. M., Jeffrey, S. S., van de Rijn, M., 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," *Proc. Natl. Acad. Sci.*, vol. 96, 1999.
22. Scherf, U., Ross, D. T., Waltham, M., Smith, L. H., Lee, J. K., Tanabe, L., Kohn, K. W., 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," *Nat. Genet.*, vol. 24, 2000.
23. Nicholson, J. K. J. Connelly, J. C. Lindon and E. Holmes, "Metabonomics: a platform for studying drug toxicity and gene function," *Nat. Rev. Drug. Discov.*, vol. 1, 2002.
24. [Http://www.acc.umu.se/~tnkjtg/Chemometrics/Editorial](http://www.acc.umu.se/~tnkjtg/Chemometrics/Editorial)
25. Bochner, B. R. "New technologies to assess genotype-phenotype relationships," *Nat. Rev. Genet.*, vol. 4, 2003.
26. Bolouri, H. and Davidson, E. H. "Modeling transcriptional regulatory networks," *BioEssays*, vol. 24, 2002.

..... 저자약력



《이 동 권》

- 1995년 영남대학교 화학공학과 졸업.
- 1997년~2000년 공정산업의 지능자동화 연구센터 연구원.
- 2004년 포항공과대학교 화학공학과 (공학박사).



《조 지 훈》

- 1999년 연세대학교 화학공학과 졸업.
- 2001년~현재 포항공과대학교 화학공학과 박사과정.



《이 인 범》

- 1977년 연세대학교 화학공학과 졸업.
- 1987년 Purdue University 화학공학과 (공학박사).
- 1988년~현재 포항공과대학교 화학공학과 교수.
- 1998년~현재 공정산업의 지능자동화 연구센터 소장.
- 2000년~현재 포항공과대학교 환경연구소 소장.
- 2002년~2003년 제어자동화시스템공학회 교육이사.
- 2004년~현재 한국공학한림원 정회원.