

질병유발 유전자 네트워크 규명을 위한 이질적 바이오 빅데이터의 시스템적 통합분석

LG전자기술원 | 최형석

1. 서론

고령화 시대를 넘어 100세 시대로 접어든 지금, 암과 뇌신경 및 만성질환과 같은 복합질환(complex disease)의 극복은 100세 건강수명을 유지하기 위한 필요조건이 되었고, 이를 해결하기 위해 BT(Bio Technology), HT(Health Technology), IT(Information Technology) 및 NT(Nano Technology)는 연구 및 산업 현장에서 실질적인 융합을 만들어 내고 있으며, 중심에 생물정보학(Bioinformatics)과 시스템생물학(Systems Biology)이 있다.

2000년대 초, 인간게놈프로젝트(Human Genome Project)를 통해 인간의 참조유전체(reference genome)가 최초로 해독된 이후 마이크로어레이(microarray) 기술을 이용해서 인간 유전체 연구가 전 세계적으로 본격적

으로 시작되어 지금까지 약 100만 건(약 15TB)에 달하는 데이터가 축적되었으며, 현재는 차세대유전체해독(Next Generation Sequencing; NGS) 기술의 개발에 의해 마이크로어레이보다 30,000배 이상의 정보량을 가진 NGS 데이터가 더 다양하고 이질적인 형태로 생산되고 있으며(유전적(genetic), 후성유전적(epigenetic) 현상뿐만 아니라 환경과의 상호작용까지 포함되는), 이를 이용하여 암과 같은 복합질환의 현상을 연구하여 질병의 조기진단 및 맞춤형 치료 방법을 개발하기 위한 다양한 시스템적이고 통합적인 바이오데이터의 분석연구가 이루어지고 있다.

기존의 연구는 유전자발현양상(gene expression profile) 데이터나 단일염기변이(Single Nucleotide Poly-

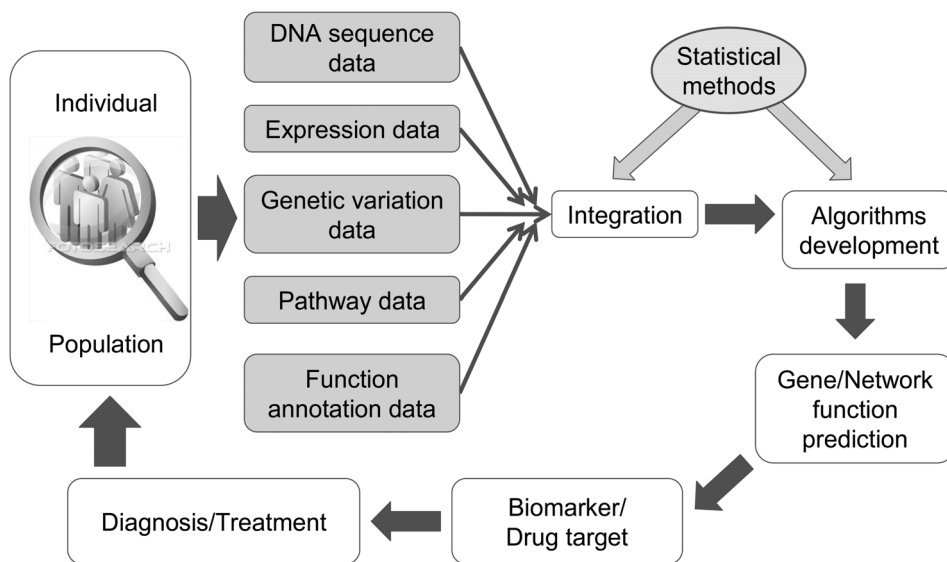


그림 1 이질적인 바이오데이터의 시스템적 통합분석의 개요

† 본 연구는 산업통상자원부 및 한국산업기술평가관리원의 산업융합원천기술개발사업의 일환으로 수행하였음(10040174, 차세대 생명정보 통합분석을 이용한 다중 바이오마커 개발).

morphism; SNP) 데이터를 이용하여 사례대조연구(case-control study)나 연관성연구(association study)를 통하여 유전자와 질병과의 관계를 유추하였다. 그러나 단순한 비교만으로 복잡계(complex system)인 생명현상을 설명하기 어려웠으며, 유전적 요소뿐만 아니라 마이크로RNA(microRNA)나 메틸화DNA(methylated DNA)와 같은 후성유전적인 요소들과 의미없는 요소로 여겨졌던 정크DNA(junk DNA)까지도 생명현상의 조절에 관여한다는 연구결과가 속속 제시되면서 분석에 사용될 데이터의 복잡성이 더욱 가중되었고, 이를 해결하기 위한 방법론의 개발이 필요하게 되었다.

본 고에서는 그림 1에서 보는 바와 같이 이질적인(heterogeneous) 바이오 빅데이터(bio big-data)의 종류 및 특징을 소개하고, 이를 시스템적으로 통합분석하여 특정 환경(질병상황 및 약물반응 등) 하에서 일어나는 다양한 유전적 요소들 간의 상호작용을 통해 특정 환경에서 작동하는 유전자조절네트워크를 예측하고 상황에 맞게 재구성하는 방법을 소개하고, 마지막으로 통합분석을 통해 질병의 원인 메커니즘을 추론하고 질병의 진단 및 치료에 적용할 수 있는 바이오마커를 발굴하는 방법을 소개하고자 한다[1].

2. 바이오 빅데이터

인간게놈프로젝트와 함께 생물정보학이 탄생하기 전까지는 생물학에서 다루는 정보의 양은 실험에 사용된 문서와 이미지 자료를 합쳐 수십 KB를 넘기 힘들었을 것이다. 그러나 30억 염기서열에 달하는 인간의 유전 정보가 해독된 후 마이크로어레이를 활용하여 본격적으로 유전체(Genomics) 연구가 시작되면서 정보량은 기본적으로 수 MB로 증가하였고, NGS 시대로 접어들면서 수백 GB로 증가하게 되었다.

데이터의 양의 증가뿐만 아니라 데이터의 종류 또한 더욱 이질적이고 다양하게 되었는데, 유전자 염기서열데이터(sequence data)와 발현데이터(expression data), SNP데이터를 포함한 다양한 유전적변이데이터(genetic variation data)를 비롯해서 단백질 및 대사산물 데이터, 후성유전체 데이터 등 다양하고 서로 독립적이지만 생물학적 기능 관점에서 서로 밀접하게 관련되어 있는, 다시 말해 시스템적으로 유기적으로 통합이 가능한 다양한 데이터가 생산되고 있다.

2.1 유전체 데이터와 후성유전체 데이터

유전체 정보의 시작이 되는 DNA 염기서열데이터는 미국의 GenBank, 유럽의 EMBL, 일본의 DDBJ에서

제공되고 있다[2]. 특히 미국의 NCBI(National Center for Biotechnology Information)와 유럽의 EBI(European Bioinformatics Institute)는 인간뿐만이 아닌 다양한 생물종의 DNA 서열데이터를 제공하고 있으며, 서열정보뿐만 아니라 유전자 발현정보, 단백질의 서열정보 및 구조정보와 상호작용 정보, 화학물질 정보와 유전질병에 대한 정보 등 다양한 정보를 제공하고 있다. 또한 ENCODE(Encyclopedia of DNA Elements) 프로젝트로 인해 유전자 이외의 기능적 요소(유전자 발현을 조절하는 DNA와 RNA의 서열 및 염색체의 구조 등)들에 대한 후성유전체 데이터가 다양하게 제공되고 있다.

미국의 UCSC Genome Browser와 유럽의 Ensembl Genome Browser를 통해서 유전체와 후성유전체 데이터에 쉽게 접근가능하며, 제공되는 Tool을 이용하여 필요한 데이터를 편리하게 조작하고 이용할 수 있다[3, 4].

2.2 마이크로어레이 데이터와 차세대유전체 데이터

• 마이크로어레이 데이터

DNA 마이크로어레이는 1995년 스탠포드 대학의 Brown 교수와 Davis 교수에 의해 개발된 후로 유전체학 연구에서 대표적으로 쓰이는 기술로써 한 생명체에서 발현되는 총 유전자를 한 번에 측정할 수 있는 기술이다.

마이크로어레이에는 대표적으로 두 종류가 있는데, 위에서 기술한 바와 같이 유전자의 발현량을 측정하는 마이크로어레이를 유전자발현(GE)칩(Gene Expression Chip)이라 하고, 유전체 DB로부터 얻어진 유전자의 서열데이터와 유전자칩으로부터 얻어지는 유전자의 발현데이터를 통합하여 유전체의 기능분석(functional genomics analysis)을 다양하게 수행할 수 있게 되었다. 다른 형태의 마이크로어레이는 유전자의 발현량 대신에 단일염기변이의 유무를 검출하기 위한 SNP칩으로써, 이를 이용하여 유전자 중의 특정 염기서열의 변이가 어떠한 표현형의 차이(개인간의 차이, 질병에 대한 개인차 등)를 유발하는지를 연구할 수 있다[5].

전세계의 마이크로어레이 데이터는 미국의 GEO(Gene Expression Omnibus)와 유럽의 ArrayExpress를 통해서 저장되고 관리되고 있으며, 원데이터와 실험정보 및 제조사별로 다양한 마이크로어레이 플랫폼에 대한 주석정보가 제공되고 있다[6,7].

• NGS 데이터

2000년대 중반에 들어서면서 모세관 방식의 1세대 Sanger 시퀀서(sequencer; 염기서열해독기)보다 운용비용과 소요시간이 획기적으로 줄어든 2세대 NGS 시퀀서가 개발되면서 인간 유전체 한 건을 해독하는데 3조의

표 1 시퀀싱 기술의 발전과 특징 비교

	1세대	2세대	3세대	3.5세대	4세대
Key Concept	▪ capillary와 laser를 이용한 자동화	▪ slide glass 또는 bead 위에 DNA 조각을 집적화	▪ DNA 증폭이 필요없이 single DNA , 광학장비는 필요	▪ 광학장비 필요없는 반도체 시퀀서, 증폭은 필요	▪ 효소, 시약, 광학장비가 필요없는 완전한 전기전자장비
장단점	▪ Sanger sequencing (termination 방식)을 이용한 비효율성 ▪ 초고비용	▪ High-throughput ▪ HPC ¹⁾ 필요 ▪ 고비용	▪ DNA 증폭 불필요 ▪ 중합효소 불안정 ▪ HPC 불필요 ▪ 중비용	▪ 광학장비 불필요 ▪ DNA 증폭 필요 ▪ HPC 불필요 ▪ 중비용	▪ 고가의 화학물질 (효소, 시약) 불필요 ▪ 광학장비 불필요 ▪ 저가, 높은 정확도, 고속, 소형화 ▪ 저비용
필요기술	유전자조작 기술	O	X	X	X
	PCR반응(증폭)	O	O	X	X
	DNA합성반응	O	O	O	O
	광학	O	O	O	X
	컴퓨터공학	X	O	O	O
	전기전자기술	X	X	X	O
	Nano기술	X	X	X	X

1) High Performance Computing

비용으로 13년에 걸렸던 것을 불과 수천만원의 비용으로 수개월 안에 해독할 수 있게 되었고, 곧 수백만원에 수일 수준으로 떨어질 전망이다. 그러나 이러한 효율성 증대의 측면보다 더 큰 NGS 기술의 가치는 아날로그 데이터에서 디지털 데이터로의 전환이라는 측면인데, 마이크로어레이 기술은 특정 염기서열을 형광으로 염색시켜 놓고 이로부터 방출되는 형광의 양(연속형) 가지고 특정 염기서열을 측정하는 방식임에 비해 NGS 기술은 직접 특정 염기서열의 개수(이산형)를 세는 방식이다. 즉, 유전자 발현량을 디지털로 정량화함으로써 상대적인 비가 아닌 절대적인 측정값을 얻을 수 있게 되었다. 하지만 아직 NGS 기술로 얻어진 데이터에 대한 효율적인 분석방법의 개발이 필요한 상황이다.

NGS 데이터는 어느 분야보다도 빅데이터(big data)이다. 인간 한명에서 얻어지는 전장유전체(whole genome) 데이터는 250GB(30X coverage 기준)에 달하는데, 현재 중국의 BGI(Beijing Genome Institute)에서 매일 생산되는 데이터량이 6TB임을 생각하면 어떤 빅데이터보다도 빅(big)하다는 것을 알 수 있다. 더 나아가 NGS 적용분야의 확장성으로 인해 유전체, 엑솜(exome), 전사체(transcriptome), 후성유전체 등 다양한 타입의 데이터 생산이 가능해졌으며, 이로써 각 생물학적 정보흐름(DNA→RNA→Protein)의 단계에서 얻어지는 빅데이터를 시스템적으로 통합하여 전체 생명현상을 시스템적으로 분석할 수 있는 기반이 마련되었다.

현재는 NNGS(Next Next Generation Sequencing)라

불리는 3, 4세대 기술에 의해서 NGS 기술에서는 필요했던 염기서열의 PCR(Polymerase Chain Reaction; 중합효소연쇄반응) 단계를 없애으로써 한 가닥의 염기서열을 바로 실시간으로 해독할 수 있는 더욱 효율적인 기술이 개발되고 있으며, DNA 분자를 직접 다루기 위해서 나노기술과 바이오기술이 융합되고 있다. 시퀀싱 기술의 발전과 특징을 표 1에 정리하였다[8].

2.3 다양한 주석(Annotation) 데이터와 2차 통합 데이터

인간의 표준게놈의 염기서열데이터와 함께 유전체 및 후성유전체에 대한 모든 주석정보(유전자의 염색체상의 위치, 기능, 발현조절부위 등)가 UCSC Genome Browser 및 Ensembl Genome Browser를 통해 통합되어 제공되고 있고, 이를 통해 원하는 정보를 효과적으로 얻을 수 있다. 또한 인간 이외의 다양한(거의 모든) 생물종에 대한 염기서열 및 주석정보를 제공함으로써 각각의 종에 대한 기능유전체학(Functional Genomics) 뿐만 아니라 진화적인 관점에서 종들간의 유전자의 기능을 연구하는 비교유전체학(Comparative Genomics) 연구도 가능하도록 데이터를 제공하고 있다.

또한 특정 연구분야의 특이적인 정보와 기존의 다양한 유전체 주석정보를 통합하여 생산된 2차 유전체 데이터베이스가 다양한 연구분야에서 생산되고 있다. 예를 들면, 암 유전체 연구를 위한 통합DB인 OncoPrint, Gene Set 분석을 위한 DAVID 등이 있고, 이 외에도 다양한 2차 통합DB들이 존재한다[9,10].

3. 이질적 데이터의 통합분석을 통한 생명현상의 시스템적 분석

위에서 기술한 바와 같이 BT기술의 발전으로 생체 내에 존재하는 다양한 분자들(DNA, RNA, 단백질, 대사물질 등)로부터 다양한 연구목적에 맞게 다양하고 이질적인 바이오데이터의 생산이 가능해 졌다. 또한 바이오데이터가 대용량(high-throughput)으로 생산이 가능해짐으로써 데이터량에 있어서도 폭발적으로 증가하게 되었다. 이렇게 다양한 바이오 빅데이터에 의해 기존의 단편적인 소규모의 데이터 분석에서 효과적이었던 관심있는 한 부분에만 집중해서 전체를 조각으로 분할하여 보는 것(reductionism)이 아닌 각 부분들이 이루는 전체 네트워크를 통째로 시스템적으로 연구하는 것(holism)이 가능하게 되었고, 이질적인 데이터를 어떻게 통합하여 효과적으로 분석할 것인가에 대해 다양한 방법론들이 개발되고 있다.

바이오데이터는 데이터의 성질에 따라 다음과 같이 세 종류로 분류할 수 있다. A(아데닌), T(티민), G(구아닌), C(시토신), 네 종류 문자의 나열로 이루어진 염기서열데이터, 유전자의 실제 발현값으로써 실수값으로 이루어진 유전자 발현데이터, 유전자의 기능에 대한 정보가 문자열로 이루어진 주석데이터로 나눌 수 있다.

이러한 이질적인 바이오 빅데이터를 시스템적으로 통합하여 생체내의 분자들 사이의 상호관계를 예측하여 생명현상의 기본을 이루고 있는 유전자조절네트워크(Gene Regulatory Network; GRN)를 예측하고 재구성(reconstruction)하는 방법들에 대해서 살펴보고자 한다.

4. 유전자 조절 네트워크 추론 방법

유전자조절네트워크를 추론하는 방법에 대해서 다음과 같이 크게 세 가지로 나눌 수 있다. 첫 번째는 유전자 발현데이터만을 이용하는 방법이고, 두 번째는 유전자 서열데이터만을 이용하는 방법이다. 마지막으로 위의 두 가지 형태의 데이터 및 그 외 다른 바이오데이터를 통합하여 분석하는 방법을 소개하고자 한다[11].

4.1 유전자 발현 데이터 이용(Expression-based approach)

대표적으로 불린네트워크(Boolean Network) 방법, 베이시안네트워크(Bayesian Network) 방법, 미분방정식모델(Differential Equation Model) 방법이 있다. 이 방법들은 측정된 유전자의 발현량 데이터를 이용하여 이들 사이의 인과관계를 특정 함수관계를 이용하여 해석하려는 시도였다. 각 방법은 불린함수(Boolean function),

최대우도함수(Maximum Likelihood), 미분함수를 이용하여 결정적(deterministic)으로 예측하였고, 이들 방법들을 robust하게 확장하기 위해서 결정함수를 확률함수로 확장하여 확률적(stochastic)으로 예측하는 방법인 확률불린네트워크(Probabilistic Boolean Network) 방법, 동적베이시안네트워크(Dynamic Bayesian Network) 방법, 확률운동모델(Stochastic Kinetics Model) 방법으로 확장되었다. 또한 선형함수를 이용한 선형모델(Linear Model), 비선형함수를 이용한 인공신경망(Artificial Neural Net), 매개변수를 변화시켜 네트워크를 재구성하는 매개변수섭동모델(Parameter Perturbation Model) 등의 방법들이 개발되었다[11].

4.2 유전자 염기서열 데이터 이용(Sequence-based approach)

이 방법은 유전자의 발현량 사이의 상관관계를 이용하는 대신에 유전자에 존재하는 염기서열의 특정패턴(sequence motif)을 이용하여 조절관계를 예측하는 방법이다. 유전자 중에는 다른 유전자의 발현을 조절하는 전사인자(Transcription Factor; TF)를 만드는 유전자가 있고, 각각의 유전자는 자신을 조절하는 전사인자가 자신을 인식해서 결합할 수 있도록 특정 염기서열을 가지고 있다. 이것을 전사인자결합모티프(Transcription Factor Binding Motif; TFBM)라고 하는데, 이 특정 염기서열을 예측함으로써 유전자조절네트워크를 예측한다.

여기에는 MEME(Multiple EM for Motif Elicitation) 방법, AlignACE 방법, Gibbs Motif Sampler 방법과 같은 ab initio 탐색방법과 Position Weight Matrices (PWMs) 및 기존에 정의되어 있는 모티프를 이용하여 모티프를 찾아내는 방법이 있다[12].

4.3 이질적인 데이터의 시스템적 통합분석(Integration-based approach)

전혀 다른 성질의 데이터인 발현데이터와 서열데이터를 통합하여 유전자조절네트워크를 보다 정확하게 분석하는 방법이 개발되고 있다. 그림 2에서 보는 바와 같이, 서열데이터를 이용하여 특정 전사인자가 인식하는 결합모티프를 추론하고, 발현데이터를 이용하여 특정 시간에 차등발현되는 유전자(Differentially Expressed Gene; DEG)를 예측하여 두 종류의 데이터를 통합하고, 통계모델인 로그선형모형(log-linear model)을 개발하여 최종적으로 특정 시간에 특정 유전자들과 그것을 조절하는 전사인자들의 집합을 추론하는 방법이 개발되었다[12]. 이와 같이 발현데이터와 서열데이

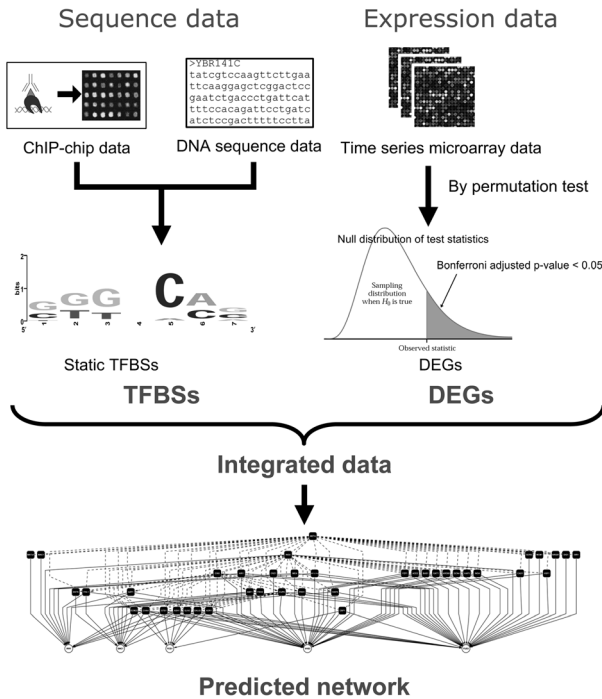


그림 2 서열데이터와 발현데이터의 통합분석

터를 통합하여 유전자조절관계를 예측하는 새로운 통계모형을 만드는 방법이 개발되고 있다.

5. 시스템적 통합분석을 이용한 질병 핵심인자의 발굴

그림 3과 같이 이질적인 데이터를 시스템적으로 통합하여 분석하는 방법을 통해 생물학적 메커니즘을 이해하고 생물시스템에서의 핵심인자를 찾아낼 수 있다. 이것을 위해서 기존에 이용되었던 서열데이터, 발현데

이터, 주석데이터에 새로운 데이터가 추가되어 새로운 알고리즘이 우선 개발되어야 한다. 최근에는 NGS 데이터와 같은 빅데이터뿐만 아니라 바이오데이터가 아닌 외부환경 데이터까지 통합하여 생명현상을 시스템적으로 해석하고자 하고 있으며, 실험검증을 통해 확인된 핵심인자는 특정 질병의 예측과 치료를 위한 바이오마커로 이용되어 진단 및 신약개발에 이용된다.

5.1 통합네트워크 추론

그림 4는 질병의 바이오마커를 개발하기 위한 시스템적 통합분석 알고리즘의 절차이다. 우선 특정 질병상태에서 생산된 데이터 및 일반적인 다양한 바이오데이터를 수집해서 가공하고 통합한다. 특정 질병상태에서 실험데이터를 직접 생산할 수도 있고, 기존의 데이터베이스를 통해서 얻을 수도 있다. 또한 마이크로어레이 실험을 통해서 발현데이터만 얻을 수도 있고, NGS 시퀀싱을 이용해서 서열데이터만 또는 서열데이터와 발현데이터를 동시에 얻을 수도 있다. 그림 5는 시스템적 통합분석 알고리즘의 일례로써, 발현데이터 및 서열데이터와 함께 다양한 주석정보로써 생체신호전달(Pathway) 정보 및 유전자 기능에 대한 정보, 전사인자나 마이크로RNA 정보, SNP나 CNV(Copy Number Variation) 정보 등을 통합하여 이를 시스템적으로 분석할 수 있는 알고리즘을 보여준다[13]. 시스템적 통합분석을 위해서 통계 알고리즘이나 기계학습 알고리즘이 사용되는데, 다양한 데이터를 사용함으로써 생기는 복잡성을 해결하고 다차원 데이터를 효과적으로 분석하기 위해서 Lasso나 Ridge Regression과 같이 데이터의 차원을 축소하거나, PCA(Principle Component Analysis), SVM

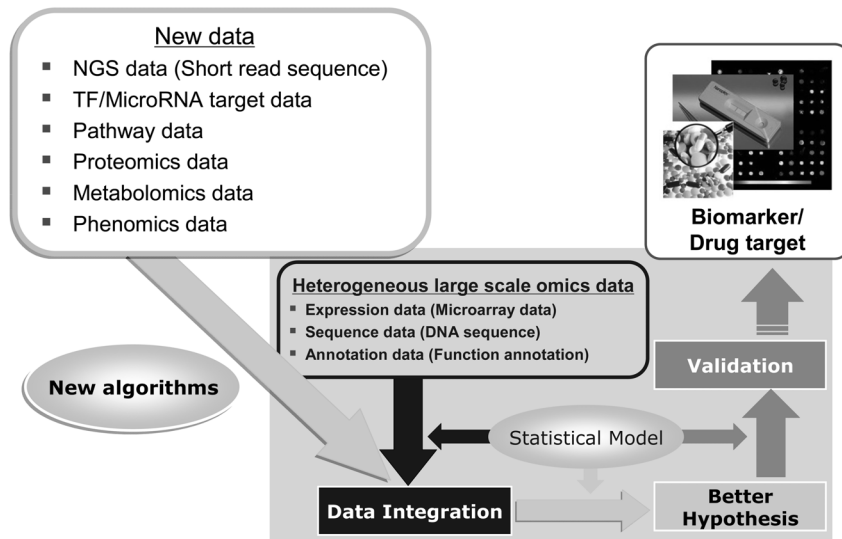


그림 3 이질적인 바이오 빅데이터의 통합분석

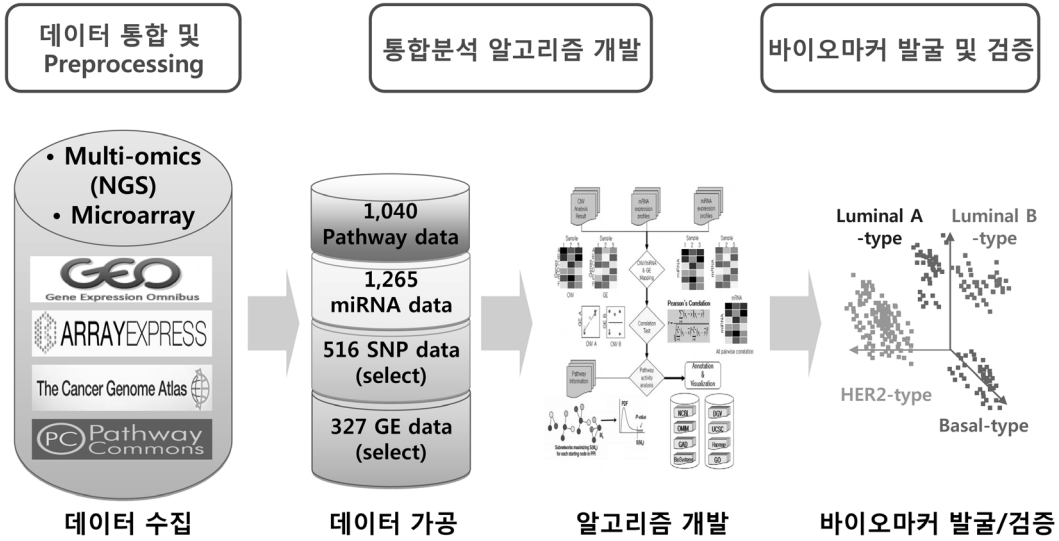


그림 4 바이오마커 발굴을 위한 통합분석 절차

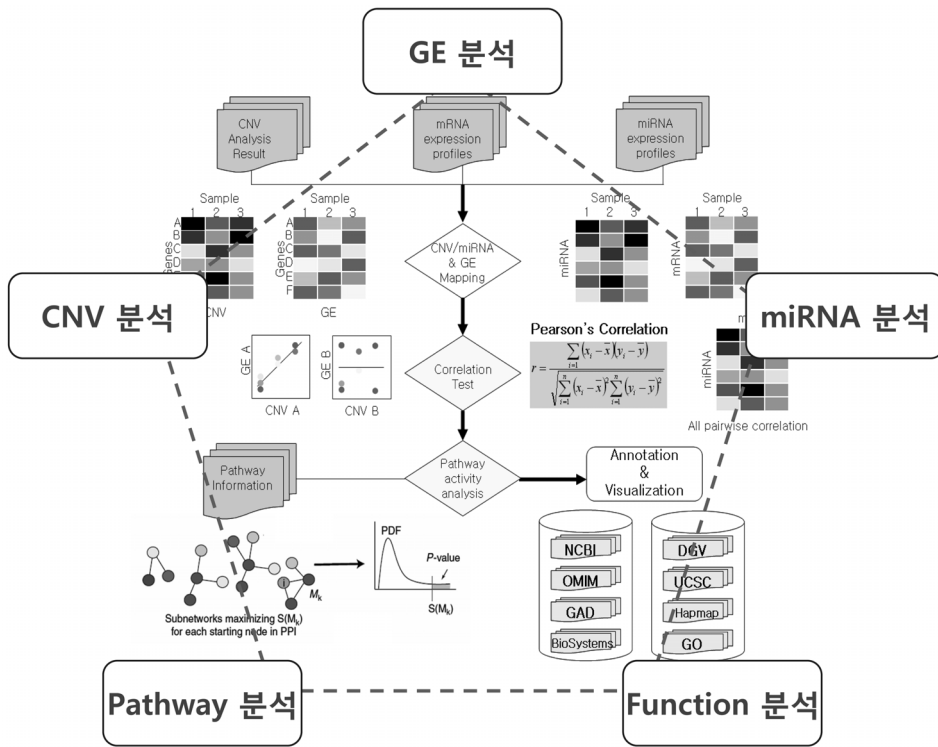


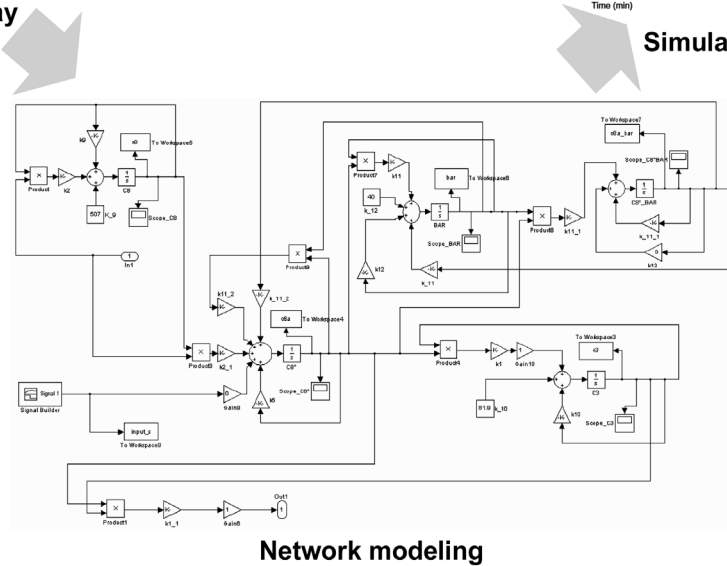
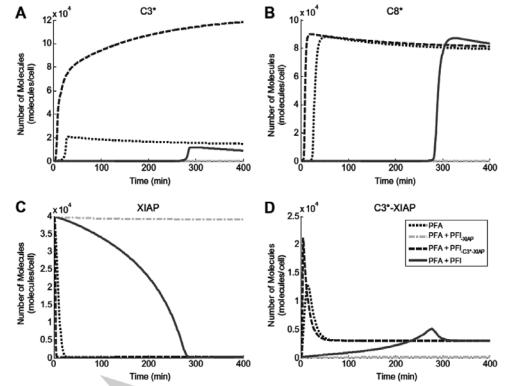
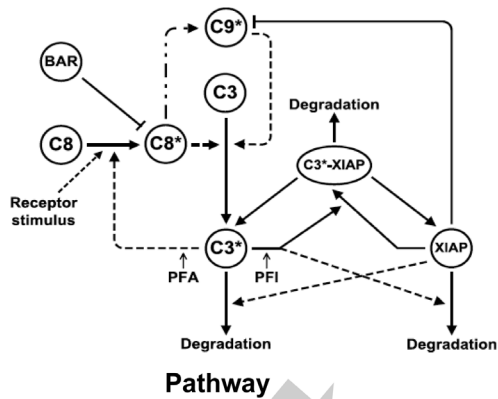
그림 5 이질적인 바이오 빅데이터의 시스템적 통합분석 알고리즘의 예

(Support Vector Machine)과 같이 차원축을 변형하는 방법들이 적용되고 있다[14].

5.2 네트워크의 동역학 분석

이질적인 바이오데이터를 통합하여 유전자조절네트워크를 추론하고 네트워크의 핵심인자를 발굴하기 위해서 동역학분석을 이용한다. 실제 생명시스템을 모사하기 위해서 그림 6과 같이 조절네트워크로부터 생체 분자 사이의 반응을 컴퓨터시뮬레이션을 통하여 동역

학 특성을 예측하여 생체네트워크에서 주요한 핵심인자를 도출할 수 있고, 그것의 조절메커니즘을 발견할 수 있다. 그림 6은 세포가 특정 상황에 처했을 때 자살하는 메커니즘인 세포사멸(Apoptosis)을 유발하는 네트워크를 추론하고, 각 구성요소 사이의 조절반응식을 구하여 컴퓨터시뮬레이션을 통해서 조절네트워크의 동역학을 분석함으로써 특정 핵심인자가 어떻게 세포사멸 시스템을 조절하는지를 밝힌 예를 보여주고 있다[15].



Network modeling

그림 6 생체조절네트워크의 동역학 시뮬레이션 분석의 예

6. 결론

BT기술의 발전과 바이오데이터 생산기술의 발전으로 다양한 바이오데이터가 생산됨으로써 이질적인 성질의 데이터가 쏟아지고 있다. 따라서 기존에는 단편적인 데이터로부터 제한된 결과를 얻을 수 밖에 없었지만 현재는 가능한 다양한 데이터를 통합함으로써 기존에 얻을 수 없었던 결과를 창발적(emergent)으로 얻을 수 있게 되었다. 그러나 바이오데이터와 같이 특히 이질적인 데이터를 통합하기 위해서는 이를 시스템적으로 통합하기 위한 알고리즘의 개발이 필수적이다. 복잡한 다차원 데이터를 효과적으로 분석하기 위해서 다양한 알고리즘이 개발되고 있으며, 시스템이론을 적용하여 생체분자의 동역학을 예측할 수 있다. 이러한 시스템적 통합분석 알고리즘을 통해 생체내의 조절네트워크를 예측하고 동역학을 분석함으로써 앞으로는 기존의 생물학 실험으로는 발굴하기가 어려웠던 암과 같은 복합 질병의 진단 및 치료를 위한 바이오마커 개발이 가능해질 것으로 보인다.

참고문헌

- [1] Werner, T., "Next generation sequencing in functional genomics", Brief Bioinform., Vol. 11. No. 5. pp.499-511, 2010.
- [2] International Nucleotide Sequence Database Collaboration, <http://www.ncbi.nlm.nih.gov/genbank/collab>
- [3] UCSC Genome Browser, <http://genome.ucsc.edu>
- [4] Ensembl Genome Browser, <http://www.ensembl.org>
- [5] Heller, M. J., "DNA microarray technology: devices, systems, and applications", Annu Rev Biomed Eng., Vol. 4, pp.129-53, 2002.
- [6] Gene Expression Omnibus, <http://www.ncbi.nlm.nih.gov/geo>
- [7] ArrayExpress, <http://www.ebi.ac.uk/arrayexpress>
- [8] Metzker, M. L., "Sequencing technologies-the next generation", Nature Rev Genet., Vol. 11. pp.31-46, 2010.
- [9] Oncomine, <http://www.oncomine.org>
- [10] DAVID, <http://david.abcc.ncifcrf.gov>
- [11] Cho, K.-H., Choo, S.-M., Jung, S. H., Kim, J.-R., Choi,

- H.-S., Kim, J., “Reverse Engineering of Gene Regulatory Networks”, IET Syst Biol., Vol. 1, No. 3, pp.149-163, 2007.
- [12] Choi, H.-S., Kim, Y. C., Cho, K.-H., Park, T., “Integrative analysis of time course microarray data and DNA sequence data via log-linear models for identifying dynamic transcriptional regulatory networks”, Int J Data Min Bioinform., Vol. 7, No. 1, pp.38-57, 2013.
- [13] Eo, H.-S., Heo, J. Y., Choi, Y., Hwang, Y., Choi, H.-S., “A pathway-based classification of breast cancer integrating data on differentially expressed genes, copy number variations and microRNA target genes”, Mol Cells, Vol. 34, Issue 4, pp.393-8, 2012.
- [14] Davidson, E. H., The Regulatory Genome: Gene Regulatory Networks In Development And Evolution, 1st Ed., p.304, Elsevier, 2006.
- [15] Choi, H.-S., Han, S., Yokota, H., Cho, K.-H., “Coupled positive feedbacks provoke slow induction plus fast switching in apoptosis”, FEBS Lett., Vol. 581, Issue 14, pp.2684-2690, 2007.

약 력



최형석

2008 서울대학교 생물정보학(박사)
 2008~2008 국립암센터 박사후연구원
 2009~2010 이화시스템생물학연구소 박사후연구원
 2010~현재 LG전자기술원 책임연구원
 관심분야: 바이오마커, 유전자 네트워크, 차세대 유전체(NGS)

E-mail : hyungseok.choi@lge.com