

# 질병 연구에 관한 생물정보학 동향

연세대학교 | 여윤구 · 박상현\*

## 1. 서론

생물학 연구는 기본적으로 실제 생물 내에서 일어나는 현상을 확인하고 그 원리를 분석하는 학문으로서, 정보를 처리하는 학문으로 생각할 수 있다. 생물학은 완전하게 동작하는 복잡한 계(system)를 분석하는 학문이지만, 그 내부 구조와 원리에 대해서는 거의 알려진 것이 없었다. 현대 분자생물학의 초기 상황을 다소 재미를 곁들여 설명한다면, 아무도 모르는 언어와 아무도 모르는 외국어 주석으로 작성된 프로그램 코드를 전산학을 전혀 몰랐던 사람이 분석하는 작업에 비유할 수 있을 것이다(심지어 띄어쓰기나 줄바꾸기도 전혀 되어 있지 않다). 또한 생명체는 같은 종일지라도 동시에 모두 다른 개체이기 때문에 각자 조금씩 다른 정보를 갖고 있으며, 생명체와 관계된 데이터 중 많은 것이 동적으로 변화하는 특성을 갖고 있다.

이러한 어려움으로 인해 생물학 연구에는 가급적 많은 양의 데이터를 확보하는 것이 매우 중요하다. 이러한 수요로 인해 근래 차세대 시퀀싱(Next generation sequencing) 기술 등 생물정보의 생산 기술이 폭발적으로 발전하였으며, 이에 따라 생물학자들이 처리해야 할 데이터의 양이 크게 증가하였다. 그러나 이렇게 엄청난 양의 데이터 생산량은 전통적인 생물학자들의 연구방식 만으로는 생산되는 데이터를 감당하기 어려운 것이었고, 생물학 도메인에서의 특수한 정보처리 기술인 생물정보학이 빠르게 발전하게 되었다.

데이터의 양적인 면 이외에도, 생물정보학의 연구성과가 누적되고 세포 내부 기작에 대한 이해가 깊어질수록 여러 데이터를 조합하여 연구에 이용할 필요성이 부각되었다. 예를 들어, 암과 관련된 유전자를 찾아내기 위해서는 암에 걸린 환자와 정상인의 DNA 서열을

비교함으로써 암에 걸린 환자에게만 다르게 나타나는 DNA 서열을 찾아낼 수 있다. 그러나 찾아낸 DNA 서열이 실제로 작동하는 것인지는 해당 DNA에 대응하는 mRNA(messenger RNA)나 단백질(Protein)의 활성화 정도를 분석해 보아야 확인할 수 있다. 이외에도 세포 내에는 핵산, RNA, 단백질 등의 다양한 구성 물질과 세포핵, 리보솜, 미토콘드리아 등의 여러 소기관이 있으며, 이들은 각각 생명 활동에서 중요한 역할을 수행한다. 서로 다른 세포 내 물질과 소기관의 동작을 측정하기 위한 데이터 역시 소기관과 물질의 종류만큼이나 다양하다.

이와 같은 다형(heterogeneous) 데이터를 종합적으로 활용하는 방법에 대한 필요성이 대두되었으며, 이에 대한 대안으로서 네트워크 이론이 활발하게 생물정보학 연구에 활용되고 있다. 최근 빅 데이터 연구에서도 주목받고 있는 네트워크는 정점(node)과 간선(edge)로 구성되는 데이터 구조로서, 정점 간의 유사도, 관련성, 선후 관계 등을 방향성(direction)과 가중치(weight)가 있는 간선으로 나타낼 수 있다. 개별 데이터를 간선 형태로 재해석함으로써 여러 종류의 다형 데이터를 하나의 개념적인 구조로 변환·통합할 수 있다는 점은 네트워크 이론이 가진 커다란 장점이다. 비록 데이터마다의 복잡한 변환 과정을 거쳐야 하지만, 이러한 장점으로 인하여 네트워크 이론은 생명체 전반을 통합적으로 이해하려는 시스템 바이올로지(Systems biology)의 주요 흐름 중 하나로 자리잡았다.

생물학 연구에 관하여 생물정보학이 수행하는 역할은 크게 두 가지로 분류할 수 있다. 먼저 첫 번째 역할은 생물학 연구에 필요한 데이터 저장 기술이나 데이터 분석·처리 방법을 개발하는 것이다. DNA 서열의 정렬(alignment) 기술, 그래프 군집화 기법, 유전자 표현형 데이터베이스의 개발 등이 이 분야에 속한다. 두 번째 역할은 다형 데이터를 통합·분석함으로써 질병의 진단과 치료에 중요한 역할을 하는 표지(marker)

\* 종신회원

† 이 원고는 2013년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2012R1A2A1A01010775).

유전자를 찾아내거나, 생물체 전반에 대한 통합적인 이해를 제시하는 것이다. 본 고에서 주로 언급하는 생물정보학의 연구가 이 역할에 해당한다.

또한 본 고에서는 생물정보학의 여러 분야 중 특히 질병 연구와 관련된 분야를 다룰 것이다. 생물학 연구 분야 중에서도 질병과 관련된 분야는 의학, 약학 등과 밀접하게 연관되어 있어, 인력이나 재정 면에서 연구의 규모가 크며 연구 결과의 파급력이 큰 편이다.

본 고의 나머지 부분은 다음과 같이 구성될 것이다. 먼저 2장에서는 질병 연구와 관련된 생물정보학에서 사용하는 데이터의 종류와 특징에 대하여 간략하게 설명할 것이며, 3장에서는 개별 질병과 관련된 유전자, 단백질, 조작 관계(regulation relationship) 등을 탐색하는 생물정보학 연구에 대하여 설명할 것이다. 마지막으로 4장에서는 단일 질병에 대한 연구를 넘어 질병 간의 연관 관계를 탐색하는 연구에 대하여 논할 것이다.

## 2. 생물정보학에서 사용하는 데이터

생물정보학은 근본적으로 정보를 다루는 학문이므로 데이터의 종류와 특징에 대한 이해가 매우 중요하다. 본 장에서는 생물정보학에서 주로 사용하는 데이터에 대하여 간략하게 설명한다.

### • 모델 생물(model organisms)

연구 역량의 선택적 집중을 위하여 생물학자들은 세포 생물에서 포유류에 이르기까지 다양한 모델 생물을 결정하고, 연구 능력을 이에 집중한다. 원핵생물의 경우 대장균(*Escherichia coli*), 진핵생물 중 세포생물로는 효모(*Saccharomyces cerevisiae*)가 대표적이다. 이외에도 식물 중에는 애기장대(*Arabidopsis thaliana*)가 대표적인 모델 생물이며, 동물에서는 꼬마선충(*Caenorhabditis elegans*), 초파리(*Drosophila melanogaster*), 쥐(*Mus musculus*) 등이 대표적인 모델 생물에 해당한다.

*norhabditis elegans*), 초파리(*Drosophila melanogaster*), 쥐(*Mus musculus*) 등이 대표적인 모델 생물에 해당한다.

### • DNA 서열 데이터(DNA sequence)

DNA는 세포의 핵에 저장되어 있는 핵산으로서, 생명 현상 발현과 유전에 있어서 가장 기본이 되는 물질이다. DNA 서열은 A, T, G, C의 4가지 핵산을 문자로 간주함으로써 문자열 형태로 나타낼 수 있으며, 인간 DNA는 약 30억의 길이를 갖는다. DNA 서열 중에 유전자(gene)에 해당하는 일부 영역의 정보는 RNA로 합성된 다음 단백질을 합성하기 때문에 특히 중요하다. DNA와 질병 간의 관계를 연구하기 위해서는 GWAS (Genome-Wide Association Study)라고 불리는 방법론이 널리 이용된다[1]. GWAS는 환자와 정상인에 해당하는 여러 개체의 DNA 정보를 생성한 뒤, DNA 서열의 차이를 분석하면 질병과 관련된 DNA 서열에 대한 정보를 얻어내는 방법이다.

### • 전사 관련 데이터(transcriptomic data)

DNA 서열에서 인트론(intron) 부분이 따로 복제되어 mRNA(messenger RNA)를 형성하는데, 이 현상을 전사(transcription)이라고 한다. 생성된 mRNA가 리보솜으로 이동하여 단백질을 합성한다. 따라서, 연구하고자 하는 대상 내의 mRNA를 측정하면 그 상황에서 활성화된 유전자에 대한 정보를 얻을 수 있다. 세포 내의 RNA 활성화 정도를 측정하기 위해서 사용되는 대표적인 기법은 마이크로어레이(microarray)와 RNA-sequencing이 있다. RNA의 측정에 사용되는 cDNA(coding DNA) 마이크로어레이는 분자생물학 초기에서부터 현재까지 꾸준히 사용되는 기술로서, mRNA를 생성하는 cDNA의 발현량을 형광 물질의 색깔과 같이 상대적 강도(intensity)로 나타내는 기술이다. RNA-sequencing

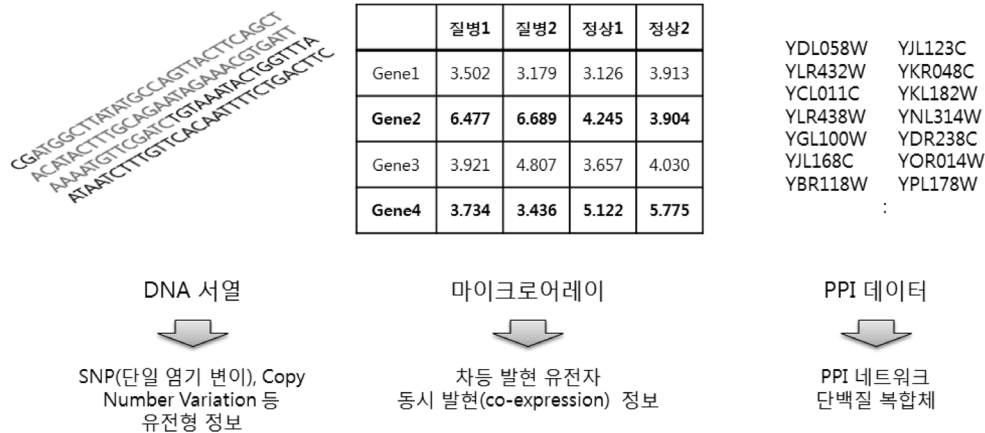


그림 1 대표적인 생물정보학 데이터의 예제

은 DNA 서열을 해독하는 기술을 RNA에 적용한 뒤, 해독된 RNA 서열의 양을 근거로 RNA 활성화 정도를 측정하는 기술이다.

• 단백질 상호작용 데이터(Protein-Protein Interaction data, PPI data)

단백질은 생명 활동의 실제 기능을 수행하는 기초 단위이다. 단백질은 독자적으로 기능하는 경우도 많지만, 두 개 이상의 단백질이 결합하여 기능을 수행하거나 한 단백질이 다른 단백질을 변형시키는 경우도 많으며, 여러 개의 단백질이 복합체(complex)를 구성하기도 한다. 이러한 단백질 상호작용 데이터는 주로 단백질과 단백질 간에 상호작용의 유무를 나타내는 바이너리 데이터로 표현된다.

• 유전자 조절과 관련된 데이터(gene regulatory information)

유전자가 생성한 단백질은 대사 물질을 생산하거나 물리적 기능을 수행하는 것 이외에도 다른 유전자의 활성화 정도를 조절하기도 한다. 복잡한 생명 활동의 원리를 보다 정확하게 이해하기 위해서는 유전자 조절에 대한 이해가 필요하다.

유전자 조절과 관련된 정보는 특정 유전자를 무력화(knock out)시킨 다음 그에 따른 변화를 관찰하는 생물학적 실험으로 밝혀지기도 하며, 데이터를 이용한 연구로는 마이크로어레이에서 동시 발현(co-expression) 하는 유전자를 기반으로 추론하기도 한다. 최근에는 chip-sequencing 기술이 새로이 개발되었는데, 이 기술을 이용하면 특정 단백질이 결합하는 DNA 서열상의 위치를 알아낼 수 있다. 이 DNA 서열상의 위치들은 해당 단백질이 전사인자(transcription factor)로 기능할 수 있는 후보 위치가 된다. 단백질이 전사인자로 기능한다는 것은, 그 단백질을 생성하는 유전자가 단백질이 전사인자로 작용하는 유전자를 조절할 수 있다는 것을 의미한다.

또, DNA-RNA-단백질의 구조 이외에 최근 발견된 조절 관계로 후성 조절(epigenetic modification)과 관련된 데이터가 있다. 이것은 DNA 이외의 원인으로 유전자의 발현이나 세포의 활동이 변동하는 것을 총칭한다. miRNA(micro RNA)의 작용은 대표적인 후성 조절 과정 중 하나로서, RNA 중 어떤 것이 단백질을 생성하는 대신 다른 mRNA에 결합하여 단백질을 생성하는 것을 억제하는 조절 작용이다. miRNA 하나가 여러 개의 RNA에 결합하는 경우가 많으며, 최근의 연구[2]에서는 인간의 단백질 생성 유전자 중 60% 이상이 miRNA에 의해 억제될 수 있는 것으로 알려졌다. 현재까

지 알려진 포유류의 miRNA 중 절반 이상이 다른 유전자의 인트론(intron) 사이에 존재하며, 그 유전자가 전사될 때 함께 전사되는 것으로 알려져 있다.

• 기타 데이터

이외에도 생물정보학에서 활용되는 데이터를 개략적으로 살펴보면 다음과 같다. 먼저 세포 내에 존재하는 각종 생화학 물질과 효소의 반응과 양을 나타내는 생화학 물질 데이터(biochemical data)가 있다. 특히 세포의 대사 활동(metabolism)에 관련된 연구의 경우 이러한 생화학 물질에 대한 데이터를 함께 연구하는 경우가 많다.

또한 이미 알려진 생물학적·의학적 지식이나 구조화된 정보 역시 생물정보학에서 널리 이용된다. 예를 들어 실험적인 방법으로 검증된 질병 관련 유전자를 기계 학습(machine learning) 방법의 학습 데이터로 활용할 수 있다. OMIM(Online Mendelian Inheritance in Man)[3]은 인간의 유전자와 질병 사이의 관계에 대한 정보를 모은 데이터베이스이며, KEGG(KEGG: Kyoto Encyclopedia of Genes and Genomes)[4]는 질병과 관련된 유전자, 생화학적 경로(pathway) 등에 대한 정보를 담고 있다. GO(Gene Ontology)[5]를 통해서는 유전자의 기능별, 종류별 분류에 대한 정보를 얻을 수 있다.

### 3. 개별 질병을 연구하는 생물정보학 방법

질병과 관련된 생물학 연구의 주된 흐름은 여러 가지의 근거에 기반하여 질병과 관련되었을 것으로 의심이 되는 후보 유전자나 물질을 결정하고, 후보에 대하여 조작을 가한 실험군과 대조군을 비교 분석하여 후보 물질의 실제 관련 여부를 검증하는 것이다. 초기 연구에서는 이러한 방법으로 여러 가지 질병 관련 유전자와 신약이 개발되었다. 그러나 연구가 지속됨에 따라 쉽게 연구가 가능한 후보 물질과 유전자의 풀(pool)이 점차 소진되었고, 신약 연구에 드는 비용과 시간이 급격히 악화되고 있다[6]. 특히 신약 및 치료법의 개발은 임상 실험까지 거쳐야 하기 때문에, 잘못된 후보에 대한 실험을 수행할 경우 최소 몇 년에서 길게는 10년 이상의 긴 시간을 허비하게 된다. 이러한 문제점으로 인해 많은 양의 정보를 분석하여 최대한 정확한 후보군을 제안해 주는 생물정보학 방법론의 필요성이 커지고 있다.

개별 질병을 연구하는 생물정보학의 방법론 중 널리 사용되는 방법론은 질병 데이터와 정상 데이터에서 서

로 다른 패턴을 보이는 대상을 탐색하는 방법이다. 이 방법에 입각한 연구는 질병 연구의 초기에서부터 마이크로어레이 데이터를 이용하여 주로 수행되었다[7]. 질병과 정상인으로부터 마이크로어레이를 생산한 뒤, 특징 선택(feature selection)이나 기계 학습 방법으로 이용하여 차별 발현되는(differentially expressed) 유전자를 찾아내는 방법이다. 이 방법으로 기존에 알려지지 않았던 후보 유전자를 찾아내는 등 여러 성과가 있었으나, 상대적인 강도로 발현량을 나타내는 마이크로어레이 자체의 한계점과 mRNA 발현의 불안정성으로 인해 한계점에 부딪혔다. 또한 miRNA와 같은 후성 조절 기작이 발견되면서, mRNA 발현량 분석만으로 유전자의 활성화를 모두 설명할 수 없게 되었다. 이에 따라 miRNA, PPI 등의 데이터를 추가로 이용하여 차별 발현 유전자를 찾아내는 연구가 활성화되었다. 2003년에는 Olga와 동료들이 효모의 다형 데이터를 통합하는 일반적인 베이지안 추론 방법을 제안하였으며[8], 2008년에는 꼬마선충의 mRNA 발현 데이터, PPI 데이터, DNA 시퀀스, 관련된 문헌 정보의 5가지 데이터를 하나의 네트워크 형태로 결합하는 연구가 제안되었다[9]. Folger와 그의 동료들은 대사 작용 네트워크와 암환자의 RNA 발현 데이터를 통합하여 새로운 표적 물질을 예측하였다[10].

질병 연구에서 널리 이용되는 또 다른 방법론은 ‘guilt by association (GBA)’[11] 개념에 입각한 연구 방법이다. GBA는 비슷한 기능을 하는 유전자들은 비슷한 유전적·물리적 특성을 나타내거나 서로 간에 상호 작용을 갖는다는 개념으로 정리할 수 있다. 이에 근거하여, 질병과 관련된 이미 알려진 유전자나 물질을 시작점으

로 삼고 이들과 유사한 특징을 갖는 새로운 대상을 탐색하는 방법이다[12,13].

역으로 GBA는 네트워크의 품질을 평가하는데 이용되기도 하는데, 데이터를 통합하여 구축한 네트워크 상에서 가까이 근접해 있는 유전자나 질병이 얼마나 비슷한 특성을 갖는가를 알아보는 것이다. 이 방법에 대한 의문점 또한 제시되고 있으나[14], 아직까지 GBA 방법은 네트워크의 평가에 널리 사용되고 있다.

#### 4. 질병 간의 연관 관계를 추론하는 방법

단일 질병에 대한 연구와 함께, 질병과 질병 간의 관계에 대한 연구 또한 활발하게 진행되고 있다. 질병과 질병 간의 관계는 어떤 질병과 다른 질병이 공통의 원인을 갖고 있어서 함께 발병하거나, 한 질병의 발현이 다른 질병의 원인이 되는 경우 등을 고려할 수 있다. 2007년 Goh와 동료들은 이러한 질병 간의 연관 관계에 대한 연구 범위를 인간에서 발생하는 유전 질환 전체로 확대하여 최초의 질병 네트워크를 제안하였다[15]. Goh가 구축한 질병 네트워크는 OMIM에서 얻은 유전자-질병관계를 이용하여 구축한 양분 네트워크(bipartite network) 형태이다(그림 2 참조). 구축된 양분 네트워크에서 어떤 질병이 다른 질병과 공통된 관련 유전자를 갖고 있으면 두 질병 사이에 간선을 생성하고, 두 유전자가 공통된 질병에 관련되어 있으면 유전자 사이에 간선을 생성하는 방식으로 질병 유전자 간의 네트워크와 질병 간의 네트워크를 동시에 생성하였다. 이렇게 구축된 질병 네트워크는 질병 분류에 따라 군집을 형성하는 경향을 보였으며, PPI, mRNA 발현 데

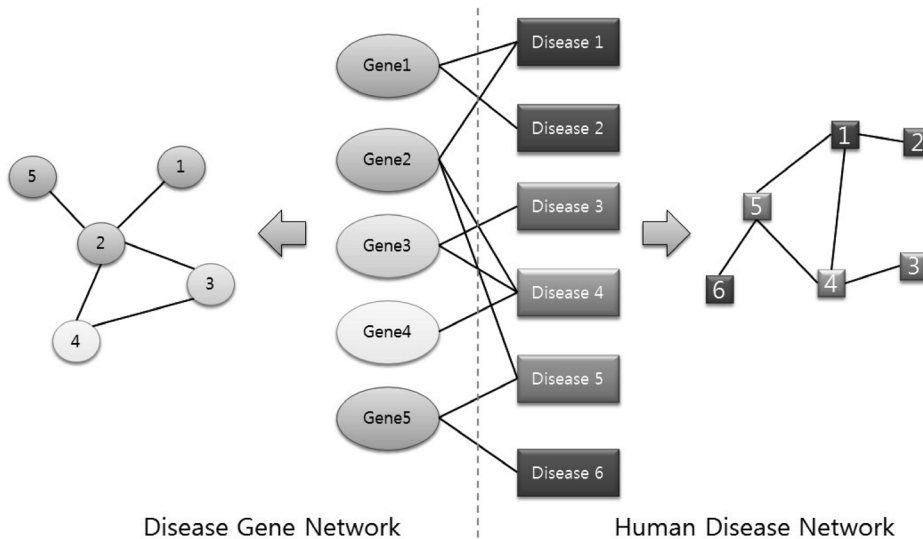


그림 2 Goh의 방법에 따라 구축된 유전자 네트워크와 질병 네트워크의 예

이더로 검증하였을 때 높은 지지도를 나타냈다. Goh에 의해 처음 제안된 질병 네트워크는 네트워크 기반의 질병 분석과 타겟 물질의 탐색 뿐만 아니라, 질병간 유사도를 기반으로 동시이환(comorbidity) 질병을 추측하고 분류되지 않은 질병을 분류하는 등 생물정보학 전반에 걸쳐 활용 가능성이 높다.

이후 이를 더욱 개량한 질병 네트워크에 대한 연구가 활발하게 진행되었다. Jiang과 그 동료들은 OMIM과 의학 용어의 데이터베이스인 MeSH를 이용하여 표현형(phenotype)의 네트워크를 구축하고, 유전자를 공유하는 질병들 뿐만 아니라 표현형이 비슷한 질병도 기능적으로 유사함을 밝혔다[16]. Lee와 동료들은 질병과 관련된 신진대사의 데이터를 이용하여 신진대사 질병 네트워크(metabolic disease network)를 구축하였다[17]. 이들의 연구 결과에 따르면, 유전자를 중심으로 한 접근 방식은 신진 대사 관련 질병의 특징을 정확하게 나타내지 못했다. 1,300만 건의 미국 의료 보험 청구 기록을 기반으로 분석한 결과, 이들이 구축한 질병 네트워크에서 많은 연결을 갖고 있는 질병들 사이에서 높은 동시 이환 현상이 관찰되었다. Satoru와 동료들은 인간 두뇌와 관련된 유전자와 질병 데이터에 텍스트 마이닝 기법으로 얻은 질병과 발병 위치 데이터를 결합하여 삼분 네트워크(tripartite network)를 구축하였다[18]. 이들의 연구 결과에 따르면 발병 위치에 따라 다양한 질병 네트워크가 형성되었으며, 소규모의 발병 위치가 많은 유전자 및 질병과 연관되어 있는 핵심 위치에 해당했다. 이 핵심 발병 위치는 별도의 GWAS 연구를 통해 밝혀진 핵심 영역을 포함하고 있었다.

## 5. 결론

지금까지 생물정보학의 기본적인 정의와 데이터, 질병과 관련된 연구 방법론의 동향에 대하여 간략하게 살펴보았다. 실제 생물정보학에서 사용할 수 있는 데이터와 방법론은 본 고에서 언급한 것보다 훨씬 방대하며, 특히 데이터의 생산 기술의 발전 속도가 매우 빠르다.

그러나 결국 모든 데이터는 생명 활동을 설명하는 여러 가지 방법 중 하나에 해당한다. 데이터가 점점 복잡해지고 방대해질수록 생물정보학자의 역할이 중요해질 것이다. 데이터의 상세한 분석만큼 생명 활동 전반에 대한 이해를 바탕으로 시스템적인 관점에서의 접근 또한 중요할 것이다.

## 참고문헌

- [ 1 ] Robert J. Klein et al, "Complement Factor H Polymorphism in Age-Related Macular Degeneration," *Science*, Vol. 308, No. 5720, pp. 385-389, 2005
- [ 2 ] Robin C. Friedman et al, "Most mammalian mRNAs are conserved targets of microRNAs," *Genome Research*, Vol. 19, No. 1, pp. 92-105, 2009
- [ 3 ] McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), Online Mendelian Inheritance in Man, OMIM®. (<http://omim.org/>)
- [ 4 ] Minoru Kanehisa, Susumu Goto, "KEGG: Kyoto Encyclopedia of Genes and Genomes," *Nucleic Acids Research*, Vol. 28, No. 1, pp. 27-30, 2000
- [ 5 ] Michael Ashburner et al, "Gene Ontology: tool for the unification of biology," *nature genetics*, 25, pp. 25-29, 2000
- [ 6 ] Fabio Pammolli, Laura Magazzini, Massimo Riccaboni, "The productivity crisis in pharmaceutical R&D," *Nature Reviews Drug Discovery*, Vol. 10, pp. 428-438, 2011
- [ 7 ] Joshua M. Stuart, Eran Segal, Daphne Koller, Stuart K. Kim, "A Gene-Coexpression Network for Global Discovery of Conserved Genetic Modules," *Science*, Vol. 302, No. 5643, pp. 249-255, 2003
- [ 8 ] Olga G. Troyanskaya, Kara Dolinski, Art B. Owen, Russ B. Altman, and David Botstein, "A Bayesian framework for combining heterogeneous data sources for gene function prediction (in *Saccharomyces cerevisiae*)," *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 100, No. 14, pp. 8348-8353, 2003
- [ 9 ] Insuk Lee et al, "A single gene network accurately predicts phenotypic effects of gene perturbation in *Caenorhabditis elegans*," *nature genetics*, No. 40, pp. 181-188, 2008
- [ 10 ] O. Folger, L. Jerby, C. Frezza, E. Gottlieb, E. Ruppim and T. Shlomi, "Predicting selective drug targets in cancer through metabolic networks," *Molecular Systems Biology*, 7:501, 2011
- [ 11 ] Chiang AP, Butte AJ. "Systematic evaluation of drug-disease relationships to identify leads for novel drug uses," *Clinical Pharmacology & Therapeutics*, 86, pp. 507-510, 2009
- [ 12 ] Timothy R. Hughes et al, "Functional Discovery via a Compendium of Expression Profiles," *Cell*, Vol. 102, Issue 1, pp. 109-126, 2000

- [13] Miguel Angel Pujana et al, "Network modeling links breast cancer susceptibility and centrosome dysfunction," *nature genetics*, 39, pp. 1338-1349, 2007
- [14] Jesse Gillis, Paul Pavlidis, " "Guilt by Association" Is the Exception Rather Than the Rule in Gene Networks," *PLoS Computational Biology*, 8(3): e1002444, 2011
- [15] Goh et al, "The human disease network," *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 104, No. 21, pp. 8685-8690, 2007
- [16] Xingpeng Jiang et al, "Modularity in the genetic disease-phenotype network," *FEBS Letters*, Vol. 582, Issue 17, pp. 2549-2554, 2008
- [17] D.-S Lee et al, "The implications of human metabolic network topology for disease comorbidity," *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 105, No. 29, pp. 9880-9885, 2008
- [18] Satoru Hayasaka, Christina E. Hugenschmidt, Paul J. Laurienti, "A Network of Genes, Genetic Disorders, and Brain Areas," *PLoS one*, Vol. 6, Issue 6, e20907, 2011

## 약력



### 여윤구

2009 연세대학교 컴퓨터과학과 (학사)  
 2011 연세대학교 컴퓨터과학과 (석사)  
 2011~현재 연세대학교 컴퓨터과학과 박사과정  
 관심분야 : Bioinformatics, Data mining  
 E-mail : yyk@cs.yonsei.ac.kr



### 박상현

1989 서울대학교 컴퓨터공학과 (학사)  
 1991 서울대학교 컴퓨터공학과 (석사)  
 2001 UCLA 컴퓨터과학과 (공학박사)  
 1991~1996 대우통신 연구원  
 2001~2002 IBM T. J. Watson Research Center  
 Post-Doctoral Fellow  
 2002~2003 포항공과대학교 컴퓨터공학과 조교수  
 2003~2006 연세대학교 컴퓨터과학과 조교수  
 2006~2011 연세대학교 컴퓨터과학과 부교수  
 2011 현재 연세대학교 컴퓨터과학과 교수  
 관심분야 : Data mining, Bioinformatics, Index structure for Flash  
 memory, SSD  
 E-mail : sanghyun@cs.yonsei.ac.kr