

기능성유전자 네트워크를 이용한 생물학 연구



이인석
연세대학교 생명공학과

E-mail: insuklee@yonsei.ac.kr

서론

최근 생물학 연구에서 가장 화두가 되는 것 중의 하나가 시스템 생물학 (Systems biology)이라 하겠다. 이는 복잡한 생물체를 단순화된 실험적 재현의 틀 안에서만 이해할 수 없다는 다수의 생물학자들의 결론, 그동안 축적되어온 여러 오믹스(Omics) 기술의 발달이 가져온 막대한 양의 실험생물학적 데이터, 그리고 고도로 발달된 컴퓨팅 능력과 정보과학이 가져온 생물정보학의 출현이 한 곳에 만나 또 하나의 새로운 학문의 줄기를 이루게 된 것이다. 시스템 생물학이 기본적으로 주장하는 것은 “전체 기능계(whole functional system)는 그것을 이루는 구성인자(component)들 사이의 상호관계 (relationship)를 통해 이해할 수 있고, 또한 역으로 각각의 구성인자의 기능은 전체 기능계 안에서 다른 구성인자와의 관계성에 의해 이해할 수 있다”는 것이다.

이는 주위에 흔히 수행되는 사회학적 관찰에서도 쉽게 이용되는 접근 방법이다. 예를 들면, 필자가 속한 여러 사회조직의 특성을 이해하기 위해서는 필자를 포함한 모든 조직구성원의 판별(identification)과 그들 간의 유기적 관계 관점에서의 구성에 대한 이해가 필수적으로 요구된다. 이러한 구성원간의 기능 관계도는 전체 조직의 이해뿐만 아니라 각 구

성원의 특성을 이해하는데도 강력한 도구가 된다. 예를 들면 필자는 필자가 속한 다양한 조직에서 각각 다른 상대 구성원들과의 다양한 관계도에 의해 필자의 다양한 사회적 기능을 관찰할 수 있다. 필자는 학교에서는 학생들과의 관계 안에서 선생으로서의 역할을 담당하게 되고 가정에서는 가족들과의 관계 안에서 가장으로서의 역할을 인식하게 된다. 이러한 사회학적 접근법 개념은 유전자들의 사회(Gene society)라 할 수 있는 세포계(cellular system)에도 동일하게 적용될 수 있다. 동일한 유전자도 그것이 어떤 유전자군(gene set)과 상호작용 하느냐에 의해 매우 이질적인 기능을 나타낼 수 있다. 유전자들의 다기능성(pleiotropy)은 이러한 사회학적 맥락에서 가장 적절하게 설명될 수 있다. 그러므로 시스템 생물학적 접근법의 핵심요소는 주어진 세포 혹은 개체 (organism)를 다양한 배경(context)에서 다각적으로 해석하는 것이다. 이들 배경에는 시간적(temporal), 공간적 (spacial), 환경적(envIRONMENTAL), 그리고 유전적(genetic) 요소들이 기여하며 이들 요소들에 의해 유전자들의 상호관계는 매우 동적으로 변화하게 된다. 그러므로 생명체의 기능, 즉 생리 현상들을 이해하기 위해서는 유전자들 간의 상호 기능관계의 정보가 매우 절실하다.

이러한 계 구성요소들의 상호관계를 가장 우수하게 모델링 할 수 있는 기술적 기반이 바로 네트워크이다. 이러한 이

유로 네트워크는 현대 시스템 생물학에서 가장 핵심적인 요소라고 하겠다. 시스템 생물학 분야 중에서도 특히 네트워크의 개발과 그 생물학적 응용에 집중하는 연구분야를 네트워크생물학(network biology)이라 칭한다. 이미 미국과 유럽의 생물학자들은 네트워크생물학이라는 주제 하에 매년 정기적으로 국제학회를 개최하고 있다. 이들 학회의 중점토의 내용은 첫째, 새로운 생물학적 네트워크 모델링의 방법론과 둘째, 네트워크를 이용한 실질적 생물학적 응용방법론 두가지로 크게 나눌 수 있다.

본 논단에서는 여러 가지 생물학적 네트워크 중에서 특히 기능성유전자네트워크(functional gene network)를 중심으로 여러 가지 유전자 네트워크의 개발법과 그것을 이용한 실제 생물학적 연구의 기법들에 대하여 살펴보고, 현재 네트워크생물학 기술의 한계와 향후 동향에 대하여 살펴보고자 한다.

유전자 네트워크의 종류와 그 모델링 기법

네트워크는 어떠한 경우에도 주어진 계(system)를 점(node)과 선(edge)으로 모델링한다. 여기서 점은 모델링에 사용되는 구성요소(component), 그리고 선은 구성요소들 간의 관계(relationship)를 나타낸다. 그러므로 모든 네트워크는 이들 점과 선의 속성(attribute)에 의하여 정의되어 질 수 있다. 이와 같이 유전자 네트워크들도 이상의 기준에 의하여 다음과 같이 분류될 수 있다.

1. 단백질 상호작용 네트워크 (protein-protein interaction network)

단백질은 생물체내의 여러 기능성 분자들 중에서도 가장 중요한 역할들을 담당하고 있고 대부분의 세포대사경로(cellular metabolic pathway) 화학반응들이 단백질들에 의하여 수행되어지고 있으므로 이들을 대상으로 하는 네트워크는 매우 유용하다. 단백질상호작용 네트워크(protein-protein interaction network, 이하 PPIN)에서 각 점은 단백질 그리고 두 점 사이의 선은 단백질들 간의 상호작용 특

히 물리적 접촉(physical contact)을 나타낸다. 현재 단백질들 간의 물리적 접촉을 감지하는 실험적 기술들은 비교적 잘 개발되어있다. 가장 보편화된 두 기법은 yeast two hybrid(이하 Y2H)와 affinity purification followed by mass spectrometry analysis(이하 APMS)이다. 이 두 기법들은 상호보완적이므로 두 방법들의 적절한 조합은 비교적 우수한 PPIN을 제공할 수 있다. 예를 들면 Y2H는 false positive가 빈번하지만 대사경로 단백질간의 일시적인 접촉까지 감지할 수 있다. 반면에 APMS는 비교적 안정적인 단백질체(protein complex)의 구성 단백질들 간의 관계에 우수한 감도(sensitivity)를 보인다. 현재까지 시스템 생물학에서 가장 많이 연구된 효모(yeast)의 경우, 이상의 두 기법들에 의한 실험 자료들을 모두 통합하면 50% 이상의 단백질체간의 상호관계들을 모델링 할 수 있다.

2. 전사조절 네트워크 (transcriptional regulatory network)

대부분의 PPIN은 두 단백질들 간의 관계를 기능적으로 대등한 위치에서 해석한다. 그러므로 PPIN에서는 관계에 어떠한 방향성(directionality)도 존재하지 않는다. 하지만 실제 생리학적 현상은 많은 경우 조절인자(regulator)와 표적인자(target)의 관계로 설명하는 것이 적절한 경우가 많다. 예를 들면 특정 대사경로에 참여하는 단백질군의 발현은 그들에게 공통적으로 특이성(specificity)을 갖는 전사조절인자(transcriptional factor, 이하 TF)에 의해 결정되는 경우가 많다. 이들 전사조절인자들과 그들의 표적유전자(target gene)들 간의 종속관계(master-slave relationship)를 방향성을 가진 선(directional edge)들로 나타낸 것이 전사조절 네트워크(transcriptional regulatory network, 이하 TRN)이다. 전사조절관계는 주로 두 가지 방법을 이용해 지도화 되어왔다. 첫째, Chromatin immuno precipitation (CHIP) 기술을 이용한 직접적인 접근법이다. 이 방법에서는 특정 전사인자를 주어진 세포의 특정 환경에서 전사조절을 위해 결합한 염색체 DNA에 다양한 방법을 통해 고정한 다음 이 전사인자가 어떤 표적 유전자의 상류염기서열

(upstream DNA sequence)에 결합하는지 염기서열판독(DNA sequencing)이나 마이크로어레이(microarray) 기술을 이용해 알아보는 방법이다. 전자의 경우는 CHIP-sequence, 그리고 후자는 CHIP-chip 기술로 명명되기도 한다. 둘째, 전사인자의 knockout(이하 TFKO)에 의한 유전자 발현의 변화를 이용하여 이들 전사인자에 의해 조절되는 표적유전자들을 발굴하는 방법이다. 이 방법이 CHIP 기술을 이용한 방법과 크게 다른 점은 CHIP 기술이 TF와 표적유전자 사이의 오직 직접적인 관계(direct relationship)만을 단백질과 DNA의 물리적 결합의 탐지를 통하여 알아내는 반면 TFKO 방법은 기능적 상관관계의 유전학적 탐지를 통한 간접적인 관계(indirect relationship)까지 제공한다는 것이다.

3. 기능성유전자 네트워크 (functional gene network)

위의 PPIN이나 TRN은 모두 특정 관계성(즉 PPIN은 단백질 간의 물리적 접촉관계, 그리고 TRN은 전사조절인자와 표적유전자 간의 조절관계)에 집중된 네트워크이다. 하지만 유전자들 간의 특정 세부 관계성에 집중된 네트워크는 몇 가지 단점이 있다. 첫째, 시스템 생물학에서 궁극적으로 추구하는 진정한 전체계에 대한 모델(holistic model)의 구축은 불가능하다는 것이다. 이는 세포내 모든 유전자들이 하나의 특정한 관계성에 의해 연관되어 있지는 않을 것이 당연하기 때문이다. 그러므로 진정한 전체계의 모델 구축을 위해서는 유전자들 사이의 매우 이질적(heterogeneous)인 관계들을 모두 고려하는 통합적인 네트워크의 구축이 필요하다. 하지만 단일 네트워크 내에 이질적인 계구성인자(heterogeneous system component)들과 이질적인 관계성(heterogeneous relationship)들을 모두 포함시키는 것은 기술적으로 매우 어렵다. 둘째, 위에서 언급한 이질성(heterogeneity)으로 인하여 다양한 종류의 생물학적 데이터를 통합하는 것 또한 기술적으로 매우 어려움이 따른다는 것이다. 사실 자료통합(data integration)은 자료분석(data mining)을 통한 시스템 연구의 핵심요소들 중 하나이다. 시스템을 분석하는 기술들은 모두 각각의 개별적인 특이성(specificity)과 탐지력

(sensitivity)이 있으므로 어느 단일 분석기술도 완전한 모델링을 성취할 수 없는 반면 이들의 상호 보완성을 잘 이용하면 단일 분석법에 비교하여 정확도(accuracy)와 completeness(완성도) 면에서 훨씬 향상된 모델의 구축이 가능하다.

위에서 언급한 PPIN과 TRN의 단점들을 보완하기 위해 개발된 새로운 네트워크 모델링기반(modeling platform)이 기능성유전자 네트워크(functional gene network)이다. 기능성유전자 네트워크는 세 가지 주요 장점을 지니도록 고안되었다. 첫째, 특정한 종류의 상호 관계성에 집중하기 보다는 일반적인 관계성인 유전자 간의 기능관련성(functional association)에 근거한 모델링을 한다. 유전자 간의 기능관련성은 그 유전자들의 산물인 단백질들 간의 물리적 접촉이나 전사인자와 그 표적유전자들의 조절관계 등을 모두 포함하는 포괄적인 정의의 관계성으로서 관계성 정의의 계급(hierarchy)에서 상위위치를 차지한다. 즉 다양한 종류의 하위수준 관계성들은 상위수준의 관계성으로 쉽게 재해석 되어 용이하게 통합될 수 있다. 결과적으로 손쉬운 자료의 통합을 가능케 함으로서 모델의 정확도와 완성도를 증대시킬 수 있다. 둘째, 통일된 확률론적 점수법(unified probabilistic scoring)을 이용하여 이질적인 근본데이터(raw data)들을 평가(benchmarking) 함으로서 이들을 표준화(standardization)하고 이 표준화는 첫 번째 특징과 함께 자료통합에 매우 유용하게 이용된다. 셋째, 현재 알려진 생물학적 지식들은 절대 구축 되어질 네트워크 모델에 포함시키지 않고 오로지 지속적으로 공급되는 대량생산 데이터들을 훈련(training) 시키는 목적으로만 이용한다. 그러므로 구축되는 네트워크는 현재 우리가 보고있는 세포계의 범위보다 훨씬 더 확장된 범위의 세포계에 대한 예측을 제공할 수 있다. 이로써 우리는 네트워크의 세포계에 대한 예측범위를 효과적으로 향상시킬 수 있다.

기능성유전자 네트워크 구축 현황

현재까지 이러한 기능성유전자 네트워크의 구축에 대한 노력은 단세포 진핵생물인 효모에서 가장 활발하게 진행되



Welcome to the www.YeastNet.org server

Networks: [v.2](#) Applications: [Phenotype Prediction](#)

Latest network versions

• **YeastNet v. 2**

YeastNet v. 2 is a probabilistic functional gene network of yeast genes, constructed from ~1.8 million experimental observations from DNA microarrays, physical protein interactions, genetic interactions, literature, and comparative genomics methods. In total, YeastNet v.2 covers 102,803 linkages among 5,453 yeast proteins (95% of the validated proteome).

YeastNet was constructed using a modified Bayesian integration of diverse data types, with each data type weighted according to how well it links genes that are known to share functions. Each interaction in YeastNet has an associated log-likelihood score (LLS) that measures the probability of an interaction representing a true functional linkage between two genes.

그림 1A. Yeastnet web site

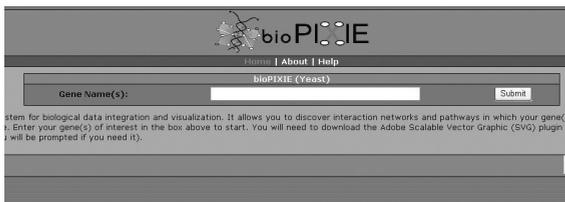


그림 1B. bioPIXIE web site

어져 왔다. 이들 중 정확도와 완성도가 가장 뛰어나고 실제 실험적 검증이 많이 진행된 두 모델은 Yeastnet (1, 2)과 bioPIXIE (3)이다(그림 1). 두 모델 모두 매우 발달된 자료통합(data integration) 기법을 이용하여 다수의 이질적인 자료들의 상호 보완적인 예측력을 효과적으로 이용함으로써 모델의 완성도를 높였다. 두 모델들 모두 Y2H와 APMS 기술에 의해 발굴된 단백질 상호작용, 마이크로어레이에 의한 유전자의 co-expression pattern, 유전자의 genome context로부터 유추된 기능 관련성 등의 많은 개별적 자료들을 포함하고 있다. 하지만 두 모델은 자료통합의 기법에 있어서는 약간의 차이를 보이고 있다. bioPIXIE는 베이즈넷(Bayes Net)을 이용해 전체적인 자료들 간의 상관관계(correlation)를 고려하는 자료통합 기법을 이용하는 반면 Yeastnet은 개별적 유전자 상관관계에 대해 가장 기여도가 높은 자료들로부터 순차적으로 높은 가중치(weight)를

적용해 자료를 통합하는 Weighted Sum 기법을 이용한다. 이 두 모델들은 비교적 수월하게 최신화(update)할 수 있으므로 지속적으로 제공되는 새로운 대량실험 자료들을 편입(incorporate)하여 현재에도 계속적으로 개선되고 있다.

최근에는 기능성유전자 네트워크의 구축이 다세포 동식물 체에도 가능해지고 있다. 그 중 선충(*C. elegans*)의 기능성 유전자 네트워크인 Wormnet은 (그림 2) 80% 이상의 약 2만개의 선충 유전자를 포함하고 있고 이전에 genome-scale RNA interference를 이용해 발굴된 40여개의 다양한 유전자저해(gene perturbation)에 의한 형질(phenotype)들에 대해 우수한 예측력을 나타냄이 보고되었다 (4). 이러한 결과는 기능성유전자 네트워크가 생물학이 궁극적으로 추구하고 있는 인체와 작물들의 예측모델의 구축이 가능하여 미래 의학과 농학 분야에 기여할 가능성이 많을 것을 시사하고 있다.



WormNet v. 1 is a probabilistic functional gene network of protein-encoding genes of *C. elegans*, constructed using a modified Bayesian integration of many different data types from several different organisms, with each data type weighted according to how well it links genes that are known to function together in *C. elegans*. Each interaction in WormNet has an associated log-likelihood score (LLS) that measures the probability of an interaction representing a true functional linkage between two genes.

To search WormNet please go to the basic search page and enter a gene, or list of genes, in the box provided, using Wormbase public or sequence gene names (e.g. *dys-1* or *F15D3.1*). The search will return all of the genes that are directly connected to an input gene (or set of genes), ranked according to their log-likelihood scores. The evidence for each interaction is indicated using the codes listed at the bottom of this page. WormNet is built on Wormbase release WS140 gene identifiers.

Current network statistics			
version	# genes	# linkages	coverage of WB140 protein-coding loci
v.1	16,113	384,700	82%

그림 2. Wormnet web site

기능성유전자 네트워크의 생물학적 응용기법

앞에서도 언급한 바와 같이 네트워크생물학의 두 가지 큰 과제는 네트워크의 완성도를 높이는 것 외에 구축된 네트워크를 실제 생물학 연구를 수행하는데 이용할 기법들을 개발하는 것이다. 현재에도 수많은 연구가 이 분야에 집중되고 있으나 그 중 몇 가지만을 소개해 보고자 한다.

1. 유전자의 신기능 예측

지난 한 세기 이상 동안 진행된 유전학 연구에도 불구하고 심지어 단순한 단세포 생물체인 효모에서조차 그 기능이 실험적으로 전혀 검증되지 않은 유전자들은 천여개에 이르고 있다. 유전자들의 다기능성(pleiotropy)을 고려한다면 전체 유전자 기능에서 우리가 현재 알고 있는 부분은 매우 한정적이라 하겠다. 이러한 유전자들의 미지의 기능을 효과적으로 예측할 수 있다면 전통적인 무작위적 탐색(random screen) 방법에 비교해 훨씬 효율적으로 유전자의 기능들을 발굴할 수 있을 것이다. 이러한 유전자 기능예측에 기능성유전자 네트워크가 매우 유용함이 여러 연구 결과들을 통하여 입증되었다.

여러 가지 기법 중 가장 많이 사용되는 것이 Guilt-By-Association (GBA) 기법이다. 즉 네트워크에서 서로 연결된 유전자들은 모두 기능적으로 관련성이 높을 것 이라는 가정 하에 특정 유전자에 연결된 모든 유전자들 중 기능적으로 이미 알려진 유전자들을 이용해 그 특정 유전자의 새로운 기능을 예측하는 것이다.

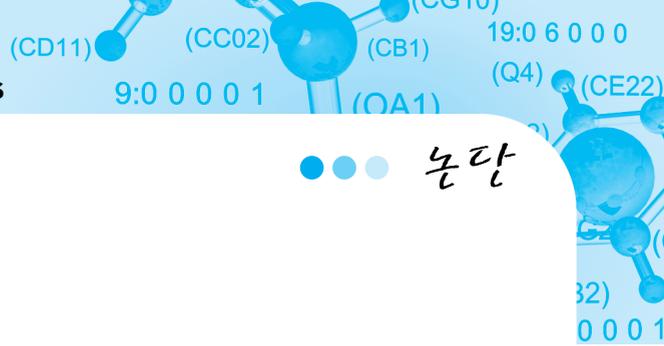
또 다른 접근법은 네트워크의 sub-graph 조직을 이용하는 것이다. 네트워크에서 연결성(connectivity)을 이용해 유전자들의 클러스터(cluster)들을 정의하고 각 클러스터 내에서 높은 빈도를 보이는 알려진 유전자들의 기능을 이용해 동일 클러스터의 기능이 알려지지 않은 구성유전자들의 기능을 예측하는 것이다. 이 접근법은 얼핏 GBA와 유사하나 GBA는 supervised learning만 가능한 반면 클러스터링 기법은 unsupervised learning도 가능 하다. 즉 정의된 클러스터 내에 기능적으로 알려진 유전자가 전혀 없더라도 우리는 주어진 클러스터 자체가 어떤 미지의 기능단위체(functional module)를 나타낸다고 가설을 세워볼 수도 있다. 이 기능단위체는 어쩌면 우리가 이제까지 전혀 발견하지 못했던 새로운 생물학적 기능단위체일 수도 있기 때문이다.

2. 유전자의 기능상실이 개체 형질에 가져올 결과 예측

네트워크를 이용해 유전자의 기능뿐 아니라 유전자의 기능상실이 초래할 형질변환의 결과도 예측하는 것이 가능할 수 있다. 왜냐하면 많은 생물체 수준의 형질들은 실제 단일 생물과정경로(biological process pathway)나 혹은 서로 밀접하게 연결된 다수의 생물과정경로들에 의해 조절되기 때문이다. 동일한 생물과정경로에 속한 유전자들은 기능성유전자 네트워크에서 서로 밀접하게 연결되어 있기 쉽다. 하지만 모든 형질 유전자들이 네트워크를 이용해 예측할 수 있는 것은 아니다. 이는 네트워크의 일반적인 예측력이 저조한 경우와 본래 주어진 형질을 조절하는 유전자들이 기능적 관련성이 적은 다수의 생물과정경로들에 의해 조절되는 경우들이 이유가 된다. 두 경우 모두 주어진 네트워크가 주어진 형질에 대해 예측력이 없으므로 네트워크를 이용한 유전자의 기능상실 형질예측에서 배제 되어야 한다. 효모의 경우에 이전에 많은 genome scale reverse genetics 접근을 이용해 발굴된 100 가지의 다른 형질에 관련된 유전자들을 이용해 테스트 해본 결과 기능성유전자 네트워크가 매우 우수한 예측력을 보임이 입증 되었다 (5). 또한 효모의 elongated cell morphology를 이용한 실험적인 검증을 통하여서도 기능성유전자 네트워크를 이용한 유전자 기능상실 형질의 예측이 가능함이 입증되었다.

3. 생존필수 유전자(essential gene)의 예측

네트워크는 그래프 모델(graph model)중의 하나이므로 그래프 이론(graph theory)을 이용한 생물학적 가설의 유출이 가능하다. 그 중 하나가 각각의 유전자의 네트워크에서의 centrality를 이용해 그 유전자의 생물체에 대한 생존필수성(essentiality)을 예측하는 것이다. 생존필수 유전자의 발굴은 특히 병원성 미생물에서 매우 중요하다. 왜냐하면 병원성 미생물의 생존필수 유전자는 그 병원균의 저해에 이용될 신약개발에 중요한 표적유전자를 제공할 수 있기 때문이다. 이러한 예측에서 우리는 “다른 유전자들과 기능관련성이 많은 유전자는 그 유전자의 기능상실이 많은 타 유전자들에게 유해한 영향을 미쳐 종국에는 전체 생물체의 생존에 위협을 줄



수 있다”는 가정에 기반을 둔다. 네트워크에서 타 유전자와 관련이 많은 유전자는 허브(hub)로 정의되며 hubness(혹은 centrality)와 생존필수성은 관련성(correlation)이 높은 것으로 다수의 네트워크 모델들을 통하여 보고되었다 (6).

4. 질병요인(Disease causal) 유전자와 그들의 조절인자(modulator) 예측

인간의 질병들 중, 단일 유전자에 기인하는 질병(monogenic disease)은 전체 질병의 10% 정도만 차지하는 것으로 추정되고 있다. 그러므로 대부분의 인간 질병은 다수의 유전자의 복합적인 작용에 기인하는 질병(polygenic disease)으로 또는 복잡질병(complex disease)으로 일컬어 지기도 한다. 그러므로 이들 질병의 치료법을 개발하기 위해선 각 질병을 일으키는 원인이 되는 유전자(causal gene)들과 그들의 조절 유전자(modulator)들을 모두 밝히고 그들 간의 복잡한 상관관계를 이해하는 것이 선결되어야 할 숙제이다. 예를 들어 암의 경우 현재까지 밝혀진 관련 유전자의 수는 수백개에 이른다. 질병도 하나의 표현형질이므로 GBA 접근법을 이용하면 주어진 질병에 관련된 새로운 유전자들을 예측할 수 있다.

네트워크를 이용한 유전학 연구

가능성 유전학의 생물학적 응용에서 필자가 마지막으로 소개하고 싶은 것은 네트워크를 이용한 새로운 그리고 효율적인 유전학 연구의 기법이다. 유전학의 궁극적인 목표중의 하나는 생물체의 표현형질과 유전자들 간의 관계성을 확립하는 것이다. 즉 어느 유전자가 어느 표현형질들을 조절하는지 또는 어떤 표현형질은 어떤 유전자들에 의하여 조절되는지 알고자 하는 것이다. 이를 위하여 전통적으로 사용되어 왔던 접근법은 forward genetics였다. 이 접근법은 phenotype-driven approach로서 주로 주어진 생명체의 유전체에 대해 무작위적으로 변이를 유도한 후 관심형질을 나타내는 돌연변이에서 변형된 유전자를 발굴하는 방법이다. 이 접근법은 오랫동안 유전학 연구를 주도해 왔으나 시

간과 노력을 많이 소모하고 변이에 의한 형질효과가 강하지 않은 유전자들은 분리하기가 쉽지 않다는 단점이 있다.

이후 기능유전체학(functional genomics)의 발달로 효모를 비롯한 몇몇 모델 생물체들에겐 genome scale of mutant library가 가용하게 되었다. 새로운 유전학적 기법인 reverse genetics에서는 이러한 mutant library를 이용해 각 유전자의 변이가 일으킬 수 있는 가능한 형질변이를 직접적으로 탐색함으로써 유전학적 탐색(genetic screen)의 식별력(sensitivity)을 매우 증가시킬 수 있었다. 하지만 이러한 예민한 방법들도 유전체의 모든 유전자들을 대상으로 하는 대량 탐색에선 고속(high-throughput)의 분석법을 사용하기 때문에 관련된 유전자를 분리하지 못하고 간과하게 되는 경우가 많이 있다. 그래서 많은 경우 reverse genetics 접근법은 가능성이 높은 소수의 후보유전자들을 우선 선별하여 검증하는 후보접근법(candidate approach)을 많이 사용한다.

이렇게 주어진 형질에 관련된 가능성이 많은 유전자 후보들을 선별하는데 기능성유전자 네트워크를 매우 유용한 예측모델로 이용할 수 있다. 예를 들면 이미 유전자 기능상실에 의한 변이형질을 알고 있는 유전자들에게 네트워크상에서 연결된 새로운 유전자들은 역시 기능상실에 의하여 동일한 변이형질을 나타내는 유전자일 확률이 매우 높다. 그러므로 새로운 형질 유전자를 탐색하기 위하여 유전체의 유전자를 모두 테스트 할 필요 없이 이들 확률이 높은 유전자들에 집중된 테스트를 하면 시간과 노력 그리고 경비를 획기적으로 줄일 수가 있을 것이다. 이와 같은 유전학 연구 기법은 “네트워크기반 집중탐색법(network-guided focused screen)”으로 명명 되어졌다 (5). 그리고 최근에는 같은 연구 접근법이 동·식물에서도 적용될 수 있음을 선충(Caenorhabditis elegans)의 유전자 네트워크를 이용한 연구에서 보고되었다 (4).

결론 및 향후전망

본 논단에서 필자는 네트워크생물학을 정의하고 그것이

시스템생물학과 어떠한 관련성을 갖는지 살펴보았다. 또한 생물학 연구에 유용하게 사용될 수 있는 네트워크에는 어떠한 종류들이 있는지 그리고 그 중 기능성유전자 네트워크는 다른 네트워크 모델들과 어떤 차이점들이 있는지도 살펴보았다. 마지막 부분에서는 기능성유전자 네트워크를 사용하여 어떻게 생물학 연구를 수행할 수 있을지에 대해서도 살펴보았다.

기능성유전자 네트워크를 이용한 예측의 유용성은 여러 실험적 검증을 통하여 이미 증명되었다 (1-5). 이 예측모델은 우수한 자료통합력과 해석의 수월성이라는 장점이 있지만 또한 단점도 존재한다. 그 중 가장 큰 단점은 유전자들 간의 상호 관계의 존재를 모델링 하지만 그 관계성의 종류(즉 두 유전자가 단백질간의 물리적 접촉에 의해 연결되어 있는지, 혹은 전사 조절인자와 그것의 표적 유전자 사이의 종속적인 관계인지 등등)를 모델링 하지는 않는다는 점이다. 그러므로 이들 관계의 보다 자세한 메커니즘을 알기 위해서는 개별적인 유전자 상호작용에 대한 연구가 별도로 수행 되어져야 한다. 하지만 이러한 여분노력의 요구는 이전의 탐색과정에서 획득되어지는 대량의 시간적 물질적 이익을 감안 한다면 전체 연구 과정에서 고려 할 때 충분히 상쇄 될 수 있는 부분이라 생각된다.

인간의 질병, 특히 복잡하고 흔한 암이나 당뇨 같은 병을 유발하는 유전적 인자들에 대한 지식은 이들 질병들을 제어하기에는 아직도 매우 부족한 실정이다. 그러므로 우리에게 더 많은 질병관련 유전자들과 그들 사이의 상호 기능관계들을 규명하는 것은 매우 절실하다. 이러한 노력의 하나로 지금 미국을 비롯한 선진국들은 Genome-wide association (GWA) study에 천문화적인 연구비를 투자하고 있다. 하지만 현재의 기술력은 수많은 후보 유전자들 중에서 질병유전자를 감지하기에 여전히 많은 통계학적 어려움을 겪고 있다. 기능성유전자 네트워크를 이용한 접근법은 앞으로 이런 경우에도 후보 유전자의 수를 효과적으로 줄여 줌으로써 신규 질병후보유전자 선별에 강력한 도움을 줄 것으로 예상된다.

【참고문헌】

1. Lee, I., et al., *A probabilistic functional network of yeast genes*. Science, 2004. **306**(5701):p.1555-8.
2. Lee, I., Z. Li, and E.M. Marcotte, *An Improved, Bias-Reduced Probabilistic Functional Gene Network of Baker's Yeast, Saccharomyces cerevisiae*. PLoS ONE, 2007. **2**(10):p.e988.
3. Myers, C.L., et al., *Discovery of biological networks from diverse functional genomic data*. Genome Biol, 2005. **6**(13):p.R114.
4. Lee, I., et al., *A single gene network accurately predicts phenotypic effects of gene perturbation in Caenorhabditis elegans*. Nat Genet, 2008. **40**(2):p.181-8.
5. McGary, K.L., I. Lee, and E.M. Marcotte, *Broad network-based predict ability of S.cerevisiae gene loss-of-function phenotypes*. Genome Biol, 2007. **8**(12):p.R258.
6. Jeong, H., et al., *Lethality and centrality in protein networks*. Nature, 2001. **411**(6833):p.41-2.

저 | 자 | 약 | 력

이인석

1988~1990	B.D., Hanyang University (Biology)
1993~1996	M.S., Western Illinois University (Biology)
1996~2002	Ph.D., Univ. of Texas at Austin (Molecular genetics & Microbiology)
2003~2006	Postdoc., Univ. of Texas at Austin (Center for Systems and Synthetic Biology)
2006~2008	Research Associate, Univ. of Texas at Austin (Center for Systems and Synthetic Biology)
2008~현재	Assistant Professor, Yonsei University (Dep. of Biotechnology)