

게놈연구로 우리 무엇을 얻을까?

Sakaki Yoshiyuki

동경대학 의과학 연구소

게놈 연구는 염기서열 결정에서 기능해석 쪽으로 크게 발전하고 있으며, 생명을 체계적으로 탐구하는 연구 프로젝트가 광범위하게 진행되고 있다. 이러한 새로운 게놈 연구는 질병의 극복, 식량문제, 환경문제를 해결하는데 튼튼한 기반이 됨과 동시에, 생명이란 무엇인가, 인간이란 무엇인가라는 철학적인 물음에 새로운 해답을 제시할 것이다.

■ 서론

DNA 이중나선구조가 발견된 후, 게놈 DNA 서열 해독은 생명과학의 중심과제였다. 1970년대 유전자 재조합 기술과 DNA 염기서열 결정기술의 발전은 생명과학분야를 비약적으로 발전시켰으며, 유전자에 대한 지식을 한층 증대시켰다. 그러나 이것은 아직 생물이라는 블랙박스 내에서 각종 생체 반응에 관한 유전자를 하나하나 발견하는 것으로, 생명의 모든 것을 이해할 수는 없었다. 그러나 1980년대 PCR법 등 DNA 해석기술의 진보로 생명의 설계도인 게놈의 전체 해석을 완성하여 생명 전체상을 체계적으로 이해할 수 있는 길이 열렸다. 현재 다수의 미생물이나 몇몇의 작은 동물의 전체 유전정보(게놈 DNA 서열)의 해독을 완료하였으며, 인간 게놈 구조 해석도 완성단계에 이르러, 생명과학 세기라 불리는 21세기에 걸맞는 기반이 갖추어져 가고있다. 이런 게놈 연구의 발전을 배경으로 유전자 의료, 바이오 산업 등 게놈 정보를 실생활에 적용하고자 하는 움직임이 활발하며, 이는 게놈 연구를 급속하게 발전시킨 추진력이며 활력의 근원이라 할 수 있다. 그러나 그 이면에는 생명과학, 생물과학으로서 그 방법론이나 사고 방식을 뿌리부터 바꿀 정도로 큰 변화가 일어나고 있는 것도 깊이 인식해야한다. 본 고에서는 생물과학론적 관점에서 우리가 연구해야 할 주요한 과제로 ① 생명의 체계적·통합적 이해, ② 인간의 보다 깊은 이해, ③ 게놈 정보과학의 전개, ④ 게놈 연구와 사회와의 관계의 새로운 움직임에 대해 논하고자 한다.

I. 시스템으로 생물의 체계적·통합적 이해

게놈의 전체 유전정보가 바로 생명의 전체상을 밝힐 수는 없지만 생명현상을 밝혀내기 위한 접근 방법에 결정적인 변화를 가져온다는 것을 강조하고자 한다. 즉 지금까지 유전자 연구가 블랙박스 내에서 미지 유전자를 탐색하는 귀납적 접근방식이라고 한다면, 게놈 정보에서 접근하는 것은 논리적, 연역적으로 생명현상과 유전자의 관계를 규명하는 것이다.

이러한 접근방식은 전체 게놈 서열이 결정된 빵 효모, 선충, 초파리 계를 중심으로 시작되고 있다. 본 고에서는 최근 일본에서 전개되고 있는 최첨단 연구를 중심으로 소개하고자 한다.

효모의 6000개 전체 유전자를 바탕으로, 모든 단백질간의 상호작용을 해명하고자 하는 Ito Takashi는 게놈 서열 정보를 바탕으로 생명현상의 다이내믹한 움직임을 이해하기 위한 연구를 진행중이다.

또 초파리의 게놈정보를 신경연구로 생명과학의 최첨단 테마에 응용하려고 하는 Itou Kei의 연구와, post-sequence 해석에서 가장 앞서있는 선충을 재료로 연구하는 Iino Yuichi는, 게놈 정보가 생물학의 방법론에 어떤 영향을 미치는지에 대해 많은 정보를 제공할 것이다. 그러나 제 1선의 연구자가 아무리 좋은 접근방법으로 해석하여도 생명의 본질을 이해하기에는, 아직 해결해

야 할 과제가 많음을 인식해야 한다.

게놈의 과학적 접근방법을 지금까지 축적된 지식을 바탕으로 새로운 전개를 하고자 하는 움직임이 활발하며, Satou Noriyuki의 우렁쟁이(Ascidians) 연구가 대표적이다. 발생학의 분야에서 독보적인 위치를 차지하고 있는 우렁쟁이에 대한 연구를 게놈적 관점에서 융합시키고자하는 연구도 한참 진행중이다. 또한 이러한 새로운 움직임은 식물에서도 볼 수 있다. 지금까지 게놈 연구는 미생물이나 동물에 치우쳐져 있었지만, 애기장대(*Arabidopsis thaliana*)의 전체 게놈 해석이 완성되어 Tabata Satoshi를 중심으로한 식물 과학 게놈 연구도 많이 진행중이다.

II. 인간의 보다 깊은 이해

인간 게놈 연구는 질병 극복을 목표로 하고 있으며, 이는 현재 게놈 연구를 이끈 원동력이다. 지금까지 1000종 이상의 질병을 유발하는 원인 유전자를 동정하여, 질병 원인을 규명중이다. 그러나 지금까지 방법론은 생활 습관병 등 우리가 일상적으로 접하는 다인자 요인에 의한 질병규명 방법으로, 명확한 원인을 밝히기에는 미흡하였으나 전체 게놈 해석의 완료로 다인자질병에 대한 접근이 용이해졌다. Tamiya Ken의 이론은 새로운 접근방법의 이점을 쉽게 설명하고 있다. 그러나 인간 게놈 해석은 질병 극복 뿐 아니라 인간을 보다 깊이 이해하기 위한 수단으로 사용되어야한다. '인간이란 무엇인가?' 라는 질문은 우리가 예전부터 가져왔던 영원한 테마이지만, 인간 게놈 해석은 이 문제에 대한 새로운 해답을 제시할 것이다. Saitou Naruya은 게놈 진화학의 관점에서 인간 게놈 연구의 새로운 방향을 제시하고 있다.

III. 게놈 정보과학의 전개

게놈과학에서 항상 염두해 두어야하는 것은 게놈해석에서 나오는 방대한 데이터를 처리하여 그 중 새로운 규칙이나 사실을 발견해야한다. 게놈 과학은 기본적으로 정보를 취급하는 과학이고, 그 궁극의 목표는 산술적으로 생물 프로세스를 시뮬레이션하는 것일지도 모른다. Yada Tetsushi와 Matuo Yo, Tomita Masaru는 이 분야에서 일본에서 누구보다 앞선 연구자로, 정통 생물학과 새로운 정보과학을 접목한 연구자이다.

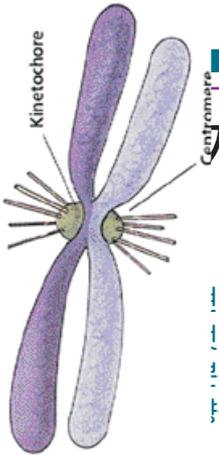
IV. 게놈과 사회

이러한 게놈에 대한 깊은 이해를 바탕으로, 유전자 치료나 바이오 산업 등 게놈정보를 실생활에 적용하려는 움직임이 활발하다. 유전자 진단, 복제기술, 유전자 재조합 작물 등의 기술력은 빠른 진보를 거듭하고 있지만, 특히, 개인 유전정보의 보호, 환경 부합 등 이에 대응하는 사회적 인식과 장비는 뒤쳐져 있는 상황이다.

아직까지 해결되지 않은 문제가 남아 있지만, 다가오는 21세기의 고령화 사회, 인구 100억 시대, 거기에서 발생하는 인구·식량·의료문제, 지역환경문제를 해결하기 위해서는 게놈·유전자 기술의 이용은 필요불가결한 요소이다. 그런 의미에서 다양한 문제에 대한 사회의 합의점 형성이 중요하다.

■ 결론

본 기획에서는 게놈 연구의 향후 전개에 대하여 다양한 측면에서 살펴보았다. 한정된 지면으로 각 저자들의 생각을 깊이있게 다루지는 못했으나, 게놈을 통한 새로운 생명과학의 숨결을 느낄 수 있었으면 한다.



게놈에서 Interactome으로

Ito Takashi

가나자와대학 암 연구소

!대 생물학의 기본 접근법은 구성요소간의 관계를 명확히 하여 상위기능의 메카니즘을 설명하는 것이다. 따라서 지금까지 구성요소 전체를 동정하는데 주력했던 게놈과학이 상호작용의 조직적 인식, 즉 Interactome쪽으로 이동되고 있다. 이러한 분석시스템을 통한 생명의 이해가 어디까지 영감을 미칠 것인가에 관심이 집중되고 있다.

■ 서론 : 게놈과학의 특성과 현황

게놈과학이란 무엇인가? 「게놈과학이란 가설에 근거하지 않은 측정과 계산에 의한 생물학이다. 이는 그 게놈과학의 특성을 가장 잘 설명한 말이다¹⁾. 무작위로 측정된 결과를 정보과학적으로 분석하여 가설로 이끄는 것으로, 많은 적든 어떤 가설 또는 거기에 가까운 것을 제시하여 결정하는 일반적인 생물학적 연구와 대조적인 것이다. 게놈과학은 먼저 전체를 분석하여 모든 구성 성분을 명확히 하는 환원론적 방법을 기본으로 한다. 그리고 그 결과를 근거로 합성론적 접근방법으로 시스템과 생명의 이해를 목표로 한다. 즉, 전반은 실험이, 후반은 정보과학적 분석이 중요한 역할을 한다. 이러한 관점에서 “게놈 프로젝트”라는 이름의 게놈 과학 제 1기를 살펴보자. 이미 알려진 바와 같이 게놈 프로젝트에서 shotgun sequencing 방법으로 무작위 해석하여 얻은 방대한 부분서열정보를 계산상으로 연결하여 모든 구조를 결정하였으며, 그 정보를 과학적으로 분석한 hypothetical protein이라는 “가설”이 제시되었다.

이렇게 설정된 “가설”이라는 거대한 앞에 우리가 안타까워하는 것은, 어떤 의미에서 생명에 대한 지식이 가지는 결함의 크기, 즉 반대로 생명의 깊이였을지도 모른다. 예를 들면 대장균이나 출아효모 등 분자유전학의 초기부터 가장 많이 연구된 생물조차도 게놈 중 유전자의 약 반수 밖에 기능이 밝혀지지 않았다. 유전자 네트워크를 명확히 하고자 구성요소를 전부 꺼내 열거했지만, 그 중 반이 무엇을 하고 있는지 전혀 예측조차 하지 못하고 있다. 즉 게놈이라는 글자는 읽었지만 그 의미가 파악되지 않는 상태로 이것을 어떻게 해결할 것인가가 게놈과학이 해결해야 할 과제이다²⁾.

1. 상호작용에 의한 게놈해석

현대 생물학은 생물 기능에서 출발해 그것을 담당하는 분자·유전자의 방향으로 진행해 왔다. 생화학은 생명현상(일부분)을 시험관 내에서 재현할 수 있는 무세포계를 개발한 후, 각 구성요소를 동정하고 정제하여 재구성하여, 분자기능을 해석하는 것이 하나의 규범처럼 진행되어 왔다. 유전학은 표현형을 지표로

변이체를 screening하여, 대응하는 유전자를 각각 동정한 후, 상호간의 유전적 관계를 규명했다. 그러나 게놈과학의 결과, 그 흐름이 완전히 뒤바뀐다. 즉, 분자·유전자가 이미 전부 명확히 밝혀져 있으므로, 이전 연구와는 반대로 이와 연결되는 생물기능을 찾아야 한다.

다시말해 역 유전학적 접근 방법이라고 할 수 있다. 각 유전자를 파괴하여 그 표현형을 찾는 것이나 원래의 기능을 추정할 수 없는 유전자이므로 어떤 표현형이 어디에 나타날 것인가를 예상하기 어렵다. 변이체를 만든 후 이미 밝혀진 표현형의 screening을 통한 다양한 프로젝트가 진행되고 있다. 그러나 유전학적 연구가 많이 진행된 생물에서 게놈 해석 결과, 처음으로 발견한 유전자들은 전통적 screening법을 이용해서는 유전학적으로 기능을 추정하기 힘든 경우가 많으므로, 원리적으로 이 방법은 큰 성과를 기대할 수 없다. 따라서 표현형 해석 프로젝트에서는 기존에 시도하지 않았던 특정 세포기능에 대한 특이적 저해제에 대한 감수성 시험 등 여러 가지 새로운 방법의 연구가 진행중이다.

분자 수준에서 생명현상을 보면 생체구성 분자간 물리적 상호작용의 총체라 할 수 있다. 어떤 분자도 단독으로는 생물학적 기능을 발휘하지 못하며, 다른 분자와 상호작용을 통하여 그 기능을 나타낼 수 있으며, 기능추정이 불가능한 유전자 산물도 반드시 다른 생체구성분자와 상호 작용할 것이다. 기능추정이 불가능한 유전자 산물 X의 상호작용 파트너가 이미 그 기능이 알려진 분자 A라고 가정하자. A 유전자는 그것이 관여하는 분자 시스템에 관한 지식과 그 생물학적 의의를 알고 있으므로 그 정보는 말할 것도 없이 X 기능추정에 대하여 많은 정보를 제공한다. 그러나 기존분자 A에서 새로운 상호작용분자 X의 발견은 지금까지 알려지지 않았던 정보를 발견하는 것으로, A를 포함한 분자시스템에 많은 영향을 미친다(그림 1). 따라서 이러한 상호작용정보의 전체적 수집은 기능추정이 불가능한 유전자 산물의 기능해명과 동시에 세포내 분자 네트워크에서 기능이 알려지지 않은 미지의 새로운 유전자산물을 통하여, 그 네트워크상에서 각각의 구성요소(신규유전자)의 의미를 해석할 수 있으며 네트워크 자체를 이해하는 데도 크게 기여한다.

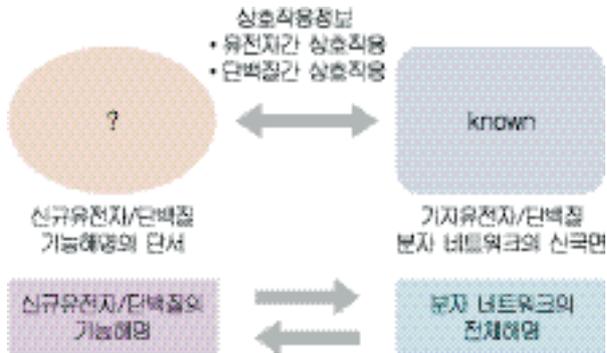


그림 1 상호작용정보에 의한 기능해석

기존 유전자산물과 신규 유전자산물의 상호작용은 양자의 기능해석에서 중요한 정보를 제공한다. 분자 네트워크 상에서 신규 유전자산물은 그 네트워크로 해석할 수 있으며, 신규 유전자산물의 기능해명과 분자 네트워크의 전체 해석은 결과적으로 동일하다.

이런 관점에서 보면, 게놈 과학도 최종적으로는 생체분자간의 상호작용을 전체적으로 해석하는 방향, 말하자면 Interactome 해석이다. 상호작용 해석 또는 요소간의 관계는 그야말로 현대생물학연구의 기본으로, 기존의 이론과 지식과의 대조적인 선입견 없이(즉, 가설에 기초하지 않고) 철저하게 상호작용 데이터 수집으로, 그 정보과학적 해석으로 가설을 확립할때 Interactome 해석의 존재 의의가 있다. Interactome 해석은 어떤 의미에서 게놈, transcriptome, proteome, metabolome 전체를 토대로 기능의 상호작용을 해석하므로 functional genomics의 큰 부분을 차지하고 있다고 해도 좋을 것이다. 그러나 현재 실질적인 상호작용해석은 유전자발현 profile와 단백질간 상호작용해석으로 이에 대해 조금 더 자세히 살펴보자.

II. 발현 profile에 의한 상호작용해석

Microarray/oligochip에 의한 transcriptome해석은, 게놈 기능해석 분야에서 가장 발전한 분야이다^{3), 4)}. 다양한 상황에서 변하는 유전자군이 공통된 제어기구 하에서 동일한 생물학적 목적을 공유하고 있다고 생각하는 해석은 매우 단순한 생각이다. Transcriptome 해석은 발현패턴의 유사도를 지표로 유전자를 그룹화 하는 것(clustering)이 주류이다. 각 cluster는 기지 유전자와 함께 신규 유전자도 포함되어 있어, 기지 유전자의 정보에서 미지 유전자의 기능을 추정 한다. 이러한 “생리적인” 유전자 발현 profile 해석과 동시에 보다 직접적으로 특정 유전자를 변화시켜 발생한 transcriptome의 변화로 유전자간의 제어관계를 알아내기 위한 연구가 진행중이다⁹⁾. 이 경우 단순하게 생각하면 전사인자와 그 표적유전자군의 관계이지만 현실적으로는 여러 가지 어려운 점이 있다.

그림 2에 나타난 가상 네트워크를 예로 생각해보자. 유전자는 직접 다른 유전자의 발현을 유도할 수 있는 전사유전자를 코딩하는 것(색으로 표시)과 그렇지 않은 것(검정색)으로 크게 나눌 수 있다. 전사인자는 크게 세 가지로 분류할 수 있다. 발현이 유도되면 곧 기능이 발현되는 것, 즉 활성화를 필요로 하지 않은 것(청), 항상 존재하면서 활성화를 그 기능발현에 필요로 하는 것(오렌지)과 두 가지 혼합형(녹색)으로 나눌 수 있다. 전사인자 이외의 유전자는 간접적으로 세포 내 상황에 변화를

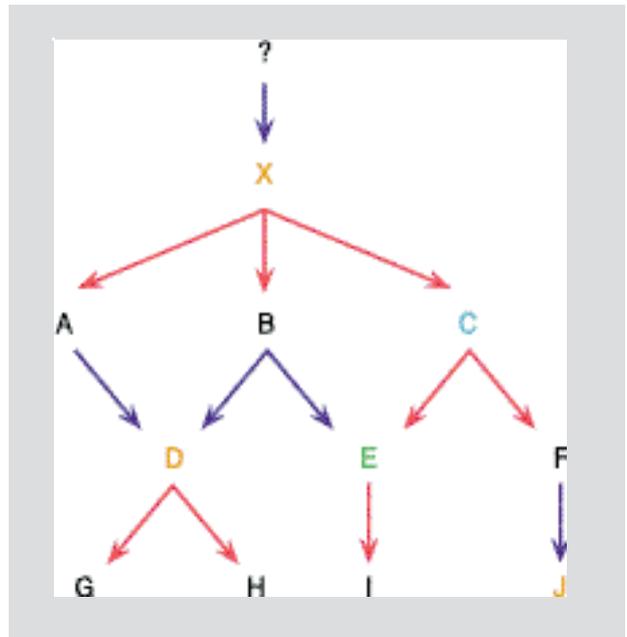


그림 2 가상적 유전자 발현제어 네트워크

유전자 발현제어 네트워크의 관점에서 보면 유전자에는 전사인자를 코딩하는 것(색)과, 그렇지 않은 것(흑)의 2종류가 있다. 전자는 기능 발현의 활성화를 필요로 하는 것(오렌지)과 발현 유도를 필요로 하는 것(청), 두 가지 다를 필요로 하는 것(녹색)으로 나눌 수 있다. 전사인자는 직접 다른 유전자의 발현을 유도하는 것(빨간 화살표)과 세포내의 상황을 변화시켜 전사인자를 활성화하는 것(보라색 화살표)으로 다른 유전자의 발현을 유도 할 수 없다. X는 신규유전자로 기능주장이 불가능한 전사인자라고 하면, 그 활성화 시그널을 미리 예측할 수는 없다. 그래서 X를 항시적 활성화로 하류유전자 후보를 동정하여, 그 안에 포함되어있는 기지유전자의 기능을 토대로 X의 생물학적 역할과 활성화 자극을 탐색한다.

주는 것으로 특정 전사인자에 의한 활성화 시그널이 필요하다. 그림 2의 X가 완전한 신규 전사인자로 기능발현에 활성화를 필요로 하는 형(오렌지)일 경우를 생각해보자. 완전한 신규 전사인자이므로 미리 그 활성화 자극을 알 수 없다. 물론 X를 파괴해도 자극이 없으면 기능을 나타내지 않으므로 어떤 일도 일어나지 않는다. 반대로 단순히 과다 발현하면, 기능할 수 없는 단백질이 만들어져 하류유전자에 영향을 미치지 않으므로 표적을 알 수가 없다. 이러한 경우 어떤 실험을 해야하는가? 가장 유용한 해결책으로 전사인자가 항상 활성화한 상태에서 세포 내에도 실험을 하고 있다. 이미 몇 개의 기지 전사인자를 모델로, DNA 결합도메인을 남긴 채 다른 도메인을 VP16과 같은 강한 전사활성화 도메인으로 치환하여 발현시키면 활성화 시그널이 없어도 표적유전자의 발현을 유도할 수 있음을 확인할 수 있다. 예를 들면 효모의 GAL4를 chimera화 하여, 본래의 활성화 자극인 galactose가 없어도 그 하류 표적유전자군의 발현을 유도할 수 있다(chimera 유전자를 디자인하기 위해서는 각 전사인자 핵미리의 대표 멤버의 고차구조에 근거한 기능구조해석을 충분히 해야하며, 또 어떤 상대와 heterodimer가 형성되어서 전사인자로 작용을 하는지와 단백질간 상호작용정보가 필요하다). 이렇게 발현 profile을 해석하면(feedback 등 까다로운 것은 고려하지 않았다.), 결과적으로 A, B, C, E, F, G, H, I는 발현량이 증가한 유전자로 분류할 수 있으나, 이들 중 X의 일차표적을 특정 짓기는 어렵다. 상류부분의 공통서열을 검색하는 것도 좋은 방법이지만, 이 경우는 A, B, C와 G, H에서 모두 공통서열을

찾을 수 있어, 어떤 것이 X의 직접표적인지 알 수 없다. 결국 제어관계를 유도하는데는, A, B, C등 발현변동을 나타내는 유전자의 변이체를 만들어 같은 분석을 반복해야 한다(전체를 해석하지 않으므로 부분을 모르는 것은 게놈과학에서 자주 발생하는 경우이다).

또 하나 문제는 변이체 자체가 갖는 한계이다. 변이체는 어떤 의미에서 변이로 세포에 대혼란이 있을 후 안정된 세포로서 기능을 발휘할 수 있는 상태가 된 것이다. 따라서 많은 부차적인 적응이 일어나므로 변이의 직접효과를 보기 어렵다. 이러한 문제를 해결하기 위하여 위의 항시적 활성형 chimera 전사인자 자체의 발현을 유도적 promotor로 조절하여 발현 profile의 변동을 해석한다. 그 결과 일차표적(A, B, C) 발현의 상승으로 이차적인 것이 유도되는 것을 알 수 있었다. 단순한 변이체를 총망라한 해석 정보에서 세밀한 정보를 추가하여, 확실하게 네트워크를 해명할 수 있다. 네트워크 해석을 위하여 정교한 변이체를 조직적으로 제작하는 것은 유용한 것 같다.

그런데 다양한 조건에서 자극에 대한 응답성으로 clustering한 경우나, 위와 같이 목적유전자를 탐색하는 방법에서도, 하나의 군 유전자에서 이루어진 집합을 제시한다. 그로부터 공통의 생물학적 기능을 발견하는 것은 불 분명한 상류 시그널을 발견하는 계기가 된다. 예를 들면 GAL4가 완전히 새로운 전사인자라고 하자. 그 서열 자체를 보고 galactose를 떠올리는 사람은 없지만, 앞서 말한 바와 같이 GAL4 chimera 전사인자의 표적 유전자중에 galactose 대사계 유전자군에서 galactose를 연상하는 것은 지극히 간단하고 자연스러운 것이다. Transcriptome은 해석 결과에서 얻은 유전자군을 대상으로 하는 게놈과학에서는 적어도 어느 정도 까지 인간의 머리에 의지하지 않고 작업을 수행하는 것이 필요하고 기존의 문헌정보를 얼마나 잘 이용하는가가 가장 중요한 포인트이다.

발현 profiling에 대해서 한가지 더 설명하자. 유전자의 기능을 나타내는 것이 단백질인 것을 감안하면 mRNA가 아니라 단백질의 발현을 측정해야한다. 현 시점에서는 기술적으로 힘들므로 차선책으로 취급이 용이한 mRNA를 측정한다. 머지않아 항체(또는 그 대체물)를 spot한 chip이 만들어지면, 단백질 수준에서 발현 profiling을 측정 할 수 있을 것이다. Proteome을 연구하는 사람은 mRNA 수준과 단백질 수준의 상관관계가 상당히 낮다고 강조하는데, proteome레벨에서의 발현 profiling은 transcriptome 레벨보다는 확실히 유효할 것이다. 또한 proteome 레벨에서는 단백질 양뿐 아니라 post-transcriptional modification의 질적 변화도 포함한 profiling이 가능해야 한다. 이는 기본적으로 유전자 발현량의 변화가 없는 부분으로, transcriptome 해석으로 얻을 수 없었던 고도 정보를 얻을 수 있다.

III. 단백질 상호작용의 전체적 분석

단백질의 다양한 기능은 다른 생체구성분자와의 상호작용을 매개로 발현되며, 이는 구조단백질, 대사경로 효소, 시그널 전달분자, 전사인자 등 모든 단백질에 해당된다. 특히 상호작용 단백질 간에서 커다란 역할을 하여, 일반적인 연구 뿐 아니라, 게놈과학에서도 중요하다. 이 이론의 기본개념은 상호작용하고 있는 단백질은 하나의 군으로서 공통된 기능을 갖고 있다는 "guilt by association"이다.

단백질 간의 상호작용 해석은 크게 두 종류로 나눌 수 있다.

하나는 효모 two-hybrid 법(Y2H), phage display법, RNA-peptide 법 등 핵산을 이용하여 그 결과를 얻는 방법이고 또 하나는 면역침강법 등으로 affinity capture 복합체의 구성 성분을 질량 분석법으로 동정하는 방법이다. 전자는 게놈이 코드할 수 있는 단백질을 모두 발현시켜 그 상호작용 가능성을 binary mode를 기본으로 해석하는 것으로 bottom up 또는 게놈적인 접근 방법이라고 한다. 후자는 실제의 proteome에 가까운 형태로 해석하는 것으로, top down 또는 proteome 적인 접근방법이라고 한다. 현재 실질적으로 조직적인 해석을 한 경우는 출아효모를 대상으로 two-hybrid 해석을 한 필자의 그룹과 CuraGen사가 있다^{5), 6)}. 이 연구의 흐름을 간단히 소개하겠다(그림 3).

먼저 효모 6000 ORF(open reading frame)을 각각 Y2H의 bait와 prey로 cloning하여, 각각을 접합형이 다른 주에 도입하여 clone을 제작한다. 다음에 각각을 pooling하여 pool 간에 접합을 실시하여 Y2H를 선택한다. 선택한 clone을 tag sequence를 이용하여 상호작용정보를 얻는다. 이러한 해석으로 저자는 4500 정도, CuraGen사는 700에 가까운 상호작용을 검출하였다. 그러나 두 결과에서 공통된 부분은 190 정도에 불과하며, 이는 screening의 깊이, construct의 차이 등 여러 가지 요인을 들 수 있으며, 단독 Y2H 방법의 한계로 두 개의 데이터를 상호적으로 통합해서 생각해야 한다.

이러한 해석 결과로 얻어진 방대한 상호작용 데이터에서 무엇을 얻을 수 있을까? 「X는 Y에 결합한다, Y는 Z에 결합한다」라고 하는 이중관계에 대한 정보가 축적되면 이를 통합하여 보다 복잡한 상호작용 네트워크를 구성해야 한다. 예를 들면 그림 4에 나타난 네트워크는 저자가 해석한 것으로 세포내 소포체 운반의 막융합은 기지 단백질과 기능이 알려지지 않은 단백질에 의해 형성된다. 이러한 중층적인 상호작용은 기능이 알려지지 않은 단백질군도 소포체 운반에 관여할 가능성을 강하게 시사한다. 또 유전적 상호작용 정보 면에서 보면 다른 cluster와의 연관성을 알

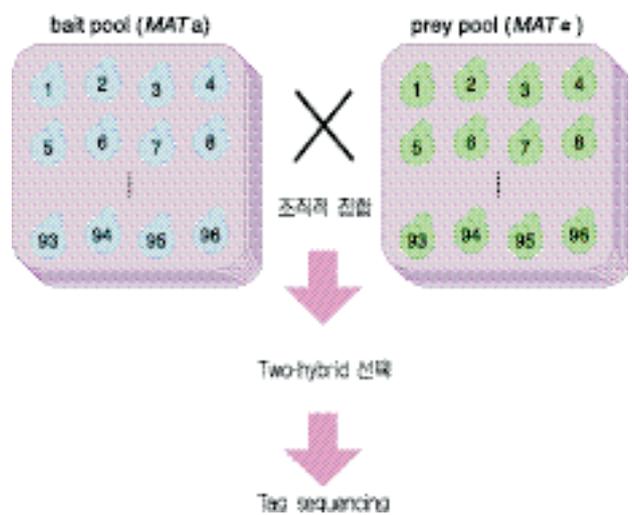


그림 3 전체적 효모 two-hybrid 해석의 예

Bait 및 prey로 게놈 중의 각 유전자를 발현하는 clone을 접합형이 다른 주를 이용하여 제작한다. 이를 일정한 수만큼 pooling하여 pool 간에 접합을 실시한다. 형성된 2배체 세포내에는 각 한 종류의 bait와 prey가 공존한다. 두 개의 상호작용을 two-hybrid로 선택하여, bait와 prey의 tag sequence (interaction sequence tags; IST)로 동정한다.

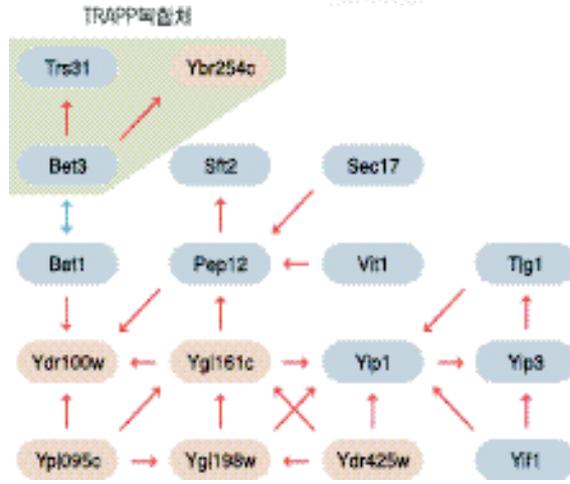


그림 4 효모 two-hybrid 상호작용 네트워크의 예

효모 two-hybrid(Y2H) 프로젝트에서 얻은 데이터로 구축한 네트워크. 세포 내 소포체 운반에 관한 단백질(핑크)과 함께 cluster를 이루고 있다. 빨간 화살표는 Y2H 상호작용으로, bait에서 prey로 방향성을 보여주고 있다. 파란 화살표는 유전적 상호작용을 의미하고, 중앙부의 네트워크와 좌상부의 TRAPP 복합체 구성분자군과 연결되어 있다.

수 있으며, 같은 세포내 물질운송계 TRAPP 복합체의 기지구성인자도 포함되어 있다. 또한 이 중 기능이 알려지지 않은 Ybr254c는 TRAPP복합체의 20 kDa 인자임을 알 수 있었다. 상호작용 네트워크가 그 기능성을 예측할 수 있게 하였으며, 이러한 상호작용정보의 축적은 신규단백질의 기능을 추정할 수 있는 열쇠이며 기존 분자 네트워크의 지식을 확대하는 계기가 된다.

방대한 상호작용 데이터에서 네트워크 모델 구축을 위해서는 계산기의 힘이 절대적으로 필요하다. 그러나 상호작용 이항정보를 기계적인 단순 연결작업만 해도 몇 천개의 단백질에서 형성되는 거대한 cluster가 형성된다. 이 거대한 cluster는 개개의 상호작용에 대한 시간적·공간적 정보로 결여되어 있고, 복수의 복합체에 포함되는 공통인자가 이를 매개로 본래와 다른 별개의 복합체를 연결하기 때문이다. 또한 Y2H는 생물학적인 근본적인 평가가 들어 있지 않은 것도 문제이다. Y2H 데이터에서 나온 네트워크는 어디까지나 가설이며, 각각의 상호작용에 대한 신뢰성도 생각해야 한다. 그 중요도는 실험의 정도(재현성)나 두 단백질에 대한 지식에 따라 결정되며, 양자의 상호작용이 생물학적으로 작용하는지를 판단해야 한다. 여기에서도 발현 profiling과 같이 clustering된 유전자에 관한 지식을 기본으로 공통사항을 찾는 작업이 요구되어, 기존의 방대한 문헌정보를 잘 이용해야 한다.

현재는 단백질 간 상호작용을 담은 카타로그는 이제 시작한 단계로 「어떤 단백질과 어떤 단백질이 결합할 수 있다」라는 단순한 이항관계정보만이 축적되어 있다. Y2H에 의한 방법도 원리적으로도 한계가 있고, 향후 그 외의 방법으로 카타로그화 할 경우 「언제, 어디서, 어느 정도의 상호작용이 일어날 것인가」에 중심을 둔 시간적·공간적 해상도와 정량성을 갖춘 “상호작용 profiling”이 되어 할 것이다. 이러한 고차 상호작용 정보를 어떻게 수집할 것인가, 그에 따라 어떤 방법론을 개발할 것인가,

시간적·공간적 요소를 어떻게 잘 정리하여 알기 쉬운 방법으로 표현해 갈 것인가가 큰 과제이다. 또한 이러한 profiling과 동시에 “상호작용 파괴”에 의한 각각 상호작용의 생물학적 의미 부여도 중요하다. 이는 상호작용 도메인의 단독 과다 발현으로 해당 상호작용의 경합 저해로 상호작용의 의미를 부여하고, screening 할 수 있어, 상호작용의 구조기반 즉, 상호작용 도메인의 조직적 동정에 중요한 열쇠가 될 것이다. 현재 상호작용 도메인의 조직적 동정을 목표로 정보과학 및 실험의 양면에서 접근하고 있다.

■ 결론 : 상호작용분석에서 네트워크 생명기능으로

현대 생물학은 각각의 수준에서 구성요소간의 상호작용정보를 모아 그들의 상관관계를 밝혀 기능의 구조를 설명하기 위한 것이다. 예를 들면 구조 생물학 분야에서 구성 원자간의 상호작용을 측정해 단백질 분자라고 하는 시스템의 전체상을 명확히 하고, 이를 바탕으로 그 분자의 기능발현을 규명하는 것이다. 발생생물학에서 세포간의 상호작용에 의한 조직, 기관 형성을 설명하고, 생태학에서는 개체간의 상호작용을 기본으로 집단을 이해하는 것이다. 그러므로 세포수준에서 분자간의 상호작용 네트워크가 중시되는 것은 지극히 자연스러운 것이다. 그러나 특정한 방향성을 가지고 해석한 후, 명확하게 하는 것이 아니라, 전체적인 상호작용을 측정해 전반적인 이해를 높이는 것이 일반적인 생물학과 게놈과학의 커다란 차이이며, 양자를 서로 보완해 전체의 이해를 넓혀나가야 할 것이다.

그러나 여기서 시스템 전체를 망라해 인간이 이해할 수 있도록 정리하는 문제가 있다. 데이터가 계산기에는 저장된다 하더라도 보통 사람의 머리에는 전부 들어 갈 수 없다. 또한 그 정보를 명확하게 이해시키기 위하여 그래프로 표현하는 것은 좋은 방법이라고 생각되나 그래프도 복잡하게 되면, 인간의 머리로는 이해할 수 없게 되므로, 이해하기 쉬운 형태로 표현할 필요가 있으며, 이는 향후 가장 중요한 문제가 될 것이다.

또한 유전자와 생물의 기능을 연결할 때 가장 근본적 불안은 인간이 과연 어느 정도까지 생물기능을 인식하여 정리할 수 있는가 이다. 유전자 일람표를 만든 후, 향후 Interactome해석을 통해 기능적으로 가까운 것을 모아 전체를 정리하는 작업(clustering)으로 발전해야 할 것이다. 그들의 대응관계에 따라 생물기능의 표면방식을 보면, 원래 정의나 용어가 생물학의 각 분야나 대상이 같은 생물임에도 불구하고 통일되어 있지 않으므로, 그것을 먼저 정리하는 생물기능 사서(ontology) 작업이 한창 진행 중이다. 그러나 원래 생물기능의 어느 정도가 생물학에서 인식되어 표현되고 있을까? 발견되지 않거나 명확하지 않은 생물기능이 있는 것은 아닐까? 기능을 알지 못하는 분자 네트워크에 연결해야 할 생물기능은 기지 리스트 상에 있는 것만으로 충분한 것인가? 에 주목해야 할 것이다. 또 반대로 분자 네트워크에서 지금까지 누구도 깨닫지 못했던 새로운 생물기능을 발견하는 것도 간과해서는 안된다.

그러나 그러기 위해서는 생물 그 자체를 허심탄회하게 관찰하고, 이미 밝혀진 현상을 배제하고 같은 현상을 주의 깊게 관찰하는 것이 중요하다.