



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2011-0035652
(43) 공개일자 2011년04월06일

(51) Int. Cl.

G01N 33/68 (2006.01) G06F 3/00 (2006.01)

(21) 출원번호 10-2009-0093460

(22) 출원일자 2009년09월30일

심사청구일자 2009년09월30일

(71) 출원인

이화여자대학교 산학협력단

서울 서대문구 대현동 11-1 이화여자대학교내

(72) 발명자

조미경

서울특별시 성동구 금호동4가 1517 16통 8반

박현석

서울특별시 용산구 청과동3가 이화맨션 201호

(74) 대리인

특허법인우인

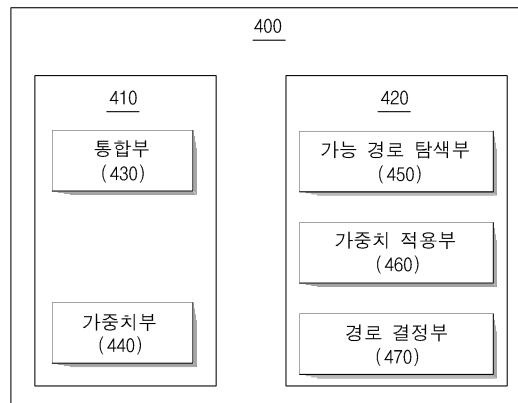
전체 청구항 수 : 총 12 항

(54) 단백질 신호 전달의 경로 탐색 방법 및 그 시스템

(57) 요약

단백질 신호 전달의 경로 탐색 방법 및 그 시스템이 개시된다. 본 단백질 신호 전달의 경로 탐색 방법은, 시작 단백질, 종료 단백질, 경로길이가 입력되면, DFS(Depth First Search) 방식으로 단백질의 신호 전달 가능경로를 탐색하는 단계, 단백질 위치에 대한 가중치 및 단백질 상호작용에 대한 가중치 중 적어도 하나를 이용하여 평가 함수를 산출하고, 상기 산출된 평가함수를 이용하여 상기 신호 전달 가능 경로를 평가하는 단계 및 상기 평가 함수의 결과를 내림차순으로 정렬하여 가장 큰 값을 신호전달 경로로 결정하는 단계를 포함한다. 그리하여 단백질 상호작용 정보와 단백질 위치정보를 기반으로 용이하게 신호전달 경로를 추출할 수 있다.

대표도 - 도4



특허청구의 범위

청구항 1

시작 단백질, 종료 단백질, 경로길이가 입력되면, DFS(Depth First Search) 방식으로 단백질의 신호 전달 가능 경로를 탐색하는 단계;

단백질 위치에 대한 가중치 및 단백질 상호작용에 대한 가중치 중 적어도 하나를 이용하여 평가 함수를 산출하고, 상기 산출된 평가함수를 이용하여 상기 신호 전달 가능 경로를 평가하는 단계; 및

상기 평가 함수의 결과를 내림차순으로 정렬하여 가장 큰 값을 신호전달 경로로 결정하는 단계;를 포함하는 단백질 신호 전달의 경로 탐색 방법.

청구항 2

제 1항에 있어서,

서로 다른 형태의 단백질 코드를 특정 단백질 코드로 단일화하는 단계;

단백질의 위치 분포도와 단백질의 위치간 상호작용 분포도를 이용하여 상기 단백질 위치에 대한 가중치와 상기 단백질 상호작용에 대한 가중치를 산출하는 단계;를 더 포함하는 것을 특징으로 하는 단백질 신호 전달의 경로 탐색 방법.

청구항 3

제 1항에 있어서,

상기 단백질 위치에 대한 가중치는,

$\text{Log}((\text{상기 단백질 위치에 속하는 단백질의 누적 경우의 수} / \text{모든 위치의 단백질 누적 경우의 수의 총합} * 100) * 100)$ 인 것을 특징으로 하는 단백질 신호 전달의 경로 탐색 방법.

청구항 4

제 1항에 있어서,

상기 단백질 상호 작용에 대한 가중치가, 단백질 a와 단백질 b의 상호작용에 대한 가중치라고 할 때, 상기 단백질 a와 단백질 b의 상호작용에 대한 가중치는 $\text{Log}((\text{단백질 a 위치들과 단백질 b 위치들간 모든 상호작용 경우의 수} / \text{모든 위치간 상호작용 경우의 수 총합} * 100) * 100)$ 인 것을 특징으로 하는 단백질 신호 전달의 경로 탐색 방법.

청구항 5

제 1항에 있어서,

상기 평가 함수는,

상기 단백질 상호작용에 대한 가중치 중 최고값인 것을 특징으로 하는 단백질 신호 전달의 경로 탐색 방법.

청구항 6

제 1항에 있어서,

상기 평가 함수는

상기 단백질 상호작용에 대한 가중치, 상기 단백질 상호 작용에 대응되는 각 단백질의 가중치의 곱 중 최고값인 것을 특징으로 하는 단백질 신호 전달의 경로 탐색 방법.

청구항 7

시작 단백질, 종료 단백질, 경로길이가 입력되면, DFS(Depth First Search) 방식으로 단백질의 신호 전달 가능

경로를 탐색하는 탐색부;

단백질 위치에 대한 가중치 및 단백질 상호작용에 대한 가중치 중 적어도 하나를 이용하여 평가 함수를 산출하고, 상기 산출된 평가함수를 이용하여 상기 신호 전달 가능 경로를 평가하는 가중치 적용부; 및

상기 평가 함수의 결과를 내림차순으로 정렬하여 가장 큰 값을 신호전달 경로로 결정하는 경로 결정부;를 포함하는 단백질 신호 전달의 경로 탐색 시스템.

청구항 8

제 7항에 있어서,

서로 다른 형태의 단백질 코드를 특정 단백질 코드로 단일화하는 통합부;

단백질의 위치 분포도와 단백질의 위치간 상호작용 분포도를 이용하여 상기 단백질 위치에 대한 가중치와 상기 단백질 상호작용에 대한 가중치를 산출하는 가중치부;를 더 포함하는 것을 특징으로 하는 단백질 신호 전달의 경로 탐색 시스템.

청구항 9

제 7항에 있어서,

상기 단백질 위치에 대한 가중치는,

$\text{Log}((\text{상기 단백질 위치에 속하는 단백질의 누적 경우의 수} / \text{모든 위치의 단백질 누적 경우의 수의 총합} * 100) * 100)$ 인 것을 특징으로 하는 단백질 신호 전달의 경로 탐색 시스템.

청구항 10

제 7항에 있어서,

상기 단백질 상호 작용에 대한 가중치가 단백질 a와 단백질 b의 상호작용에 대한 가중치라고 할 때, 상기 단백질 a와 단백질 b의 상호작용에 대한 가중치는 $\text{Log}((\text{단백질 a 위치들과 단백질 b 위치들간 모든 상호작용 경우의 수} / \text{모든 위치간 상호작용 경우의 수 총합} * 100) * 100)$ 인 것을 특징으로 하는 단백질 신호 전달의 경로 탐색 시스템.

청구항 11

제 7항에 있어서,

상기 평가 함수는,

상기 단백질 상호작용에 대한 가중치 중 최고값인 것을 특징으로 하는 단백질 신호 전달의 경로 탐색 시스템.

청구항 12

제 7항에 있어서,

상기 평가 함수는

상기 단백질 상호작용에 대한 가중치, 상기 단백질 상호 작용에 대응되는 각 단백질의 가중치의 곱 중 최고값인 것을 특징으로 하는 단백질 신호 전달의 경로 탐색 시스템.

명 세 서

발명의 상세한 설명

기술 분야

[0001] 단백질 신호 전달의 경로 탐색 방법 및 그 시스템의 예 관한 것으로, 보다 상세하게는 단백질의 위치 정보 및 단백질의 상호작용 정보를 이용하여 단백질 신호 전달의 경로를 탐색하는 방법 및 그 시스템에 관한 것이다.

배 경 기 술

- [0002] 1980년부터 시작된 DNA 염기서열결정방법 (Sequencing Method)의 개발로 유전체 정보에 대한 활발한 연구가 진행되고 있다. 그 결과 유전체 정보는 1977년 바이러스(Bacteriophage)(5.4 Kb)의 전체 염기서열을 해독된 이후, 1983년에는 람다(Lambda)(48.5Kb), 1997에는 인플루엔자(Haemophilus Influenzae)(1.8Mb)와 효모(Yeast)(12Mb), 1998에는 선충류인 C.elegans (98Mb)의 유전체 정보가 해독 되었다. 더욱이 2001년 인간 유전체 사업을 통해 인간의 유전자 지도(~3,000Mb)가 완성된 이후 많은 유전체 정보에 대한 연구가 진행되고 있다.
- [0003] 현재 효모, C.elegans, 초파리(Drosophila), 쥐(Mouse), 집쥐(Rat) 등 100종이상의 유전체 정보가 축적되고 있으며, 이 같은 정보는 데이터베이스에 계속 축적되고 있다.
- [0004] 최근에는 유전체 정보뿐만 아니라 단백질 상호작용에 대한 정보 또한 빠른 속도로 축적되고 있으며 이를 통해 대사(Metabolism) 및 신호전달(Signal Transduction), 그리고 세포주기(Cell Cycle)에 대한 정보를 수집할 수 있게되었다. KEGG를 비롯하여 EcoCyc, SPAD 등 다양한 종류의 세포 내 경로정보에 대한 데이터베이스가 그것이다. 특히 SPIDer는 대중적인 프로그램으로 효모의 유전체를 기반으로 단백질 상호작용을 데이터베이스로 정리하였다. SPIDer는 다양한 조건검색이 제공하며, 그래프를 이용하여 단백질 상호작용을 시각화 해주어 연구자들은 편리하게 광역적인 단백질 상호작용을 분석 할 수 있게 되었다.
- [0005] 최근 유전체 정보뿐만 아니라 단백질 상호작용에 대한 관심이 높아진 이유는 단백질 상호작용 경로를 구현하는 것이 생물학적 현상을 이해하는 것에 매우 중요하기 때문이다. 생명체를 구성하는 최소 단위인 세포의 생명활동은 세포 내의 다양한 종류의 단백질이 적재적소적소에서 정확하게 제 기능을 발휘함으로써 가능한 것이다.
- [0006] 특히 세포 안에서는 한 종류의 단백질이 단독적으로 작용하기보다는 기능적으로 연관되어 있는 단백질 상호작용이 정확하게 일어남으로써 대사작용 및 신호전달이 체계적으로 이루어진다.
- [0007] 그러나 현재 생물학적 측면에서는 외부자극에 대한 세포 내부의 복잡한 신호전달 경로가 정확히 어떻게 존재하는지 알아내기가 힘들다는 것이다.

발명의 내용

해결 하고자하는 과제

- [0008] 본 발명은 단백질 상호작용 정보와 단백질 위치정보를 기반으로 신호 전달 경로를 추출하는 시스템 및 방법을 제공하는 것이 목적이다.

과제 해결수단

- [0009] 상기 목적을 달성하기 위한 본 발명에 따른, 단백질 신호 전달의 경로 탐색 방법은, 시작 단백질, 종료 단백질, 경로길이가 입력되면, DFS(Depth First Search) 방식으로 단백질의 신호 전달 가능경로를 탐색하는 단계; 단백질 위치에 대한 가중치 및 단백질 상호작용에 대한 가중치 중 적어도 하나를 이용하여 평가 함수를 산출하고, 상기 산출된 평가함수를 이용하여 상기 신호 전달 가능 경로를 평가하는 단계; 및 상기 평가 함수의 결과를 내림차순으로 정렬하여 가장 큰 값을 신호전달 경로로 결정하는 단계;를 포함한다.
- [0010] 그리고, 서로 다른 형태의 단백질 코드를 특정 단백질 코드로 단일화하는 단계; 및 단백질의 위치 분포도와 단백질의 위치간 상호작용 분포도를 이용하여 상기 단백질 위치에 대한 가중치와 상기 단백질 상호작용에 대한 가중치를 산출하는 단계;를 더 포함하는 것이 바람직하다.
- [0011] 또한, 상기 단백질 위치에 대한 가중치는, $\text{Log}((\text{상기 단백질 위치에 속하는 단백질의 누적 경우의 수} / \text{모든 위치의 단백질 누적 경우의 수의 총합} * 100) * 100)$ 인 것이 바람직하다.
- [0012] 그리고, 상기 단백질 상호 작용에 대한 가중치가, 단백질 a와 단백질 b의 상호작용에 대한 가중치라고 할 때, 상기 단백질 a와 단백질 b의 상호작용에 대한 가중치는 $\text{Log}((\text{단백질 a 위치들과 단백질 b 위치들간 모든 상호 작용 경우의 수} / \text{모든 위치간 상호작용 경우의 수 총합} * 100) * 100)$ 인 것이 바람직하다.
- [0013] 또한, 상기 평가 함수는, 상기 단백질 상호작용에 대한 가중치인 것이 바람직하다.
- [0014] 그리고, 상기 평가 함수는 상기 단백질 상호작용에 대한 가중치, 상기 단백질 상호 작용에 대응되는 각 단백질의 가중치의 곱인 것이 바람직하다.
- [0015] 한편, 상기 목적을 달성하기 위한 본 발명에 따른, 단백질 신호 전달의 경로 탐색 시스템은, 시작 단백질, 종료 단백질, 경로길이가 입력되면, DFS(Depth First Search) 방식으로 단백질의 신호 전달 가능경로를 탐색하는 탐

색부; 단백질 위치에 대한 가중치 및 단백질 상호작용에 대한 가중치 중 적어도 하나를 이용하여 평가 함수를 산출하고, 상기 산출된 평가함수를 이용하여 상기 신호 전달 가능 경로를 평가하는 가중치 적용부; 및 상기 평가 함수의 결과를 내림차순으로 정렬하여 가장 큰 값을 신호전달 경로로 결정하는 경로 결정부;를 포함한다.

[0016] 그리고, 서로 다른 형태의 단백질 코드를 특정 단백질 코드로 단일화하는 통합부; 및 단백질의 위치 분포도와 단백질의 위치간 상호작용 분포도를 이용하여 상기 단백질 위치에 대한 가중치와 상기 단백질 상호작용에 대한 가중치를 산출하는 가중치부;를 더 포함하는 것이 바람직하다.

[0017] 또한, 상기 단백질 위치에 대한 가중치는, $\text{Log}((\text{상기 단백질 위치에 속하는 단백질의 누적 경우의 수} / \text{모든 위치의 단백질 누적 경우의 수의 총합} * 100) * 100)$ 인 것이 바람직하다.

[0018] 그리고, 상기 단백질 상호 작용에 대한 가중치가 단백질 a와 단백질 b의 상호작용에 대한 가중치라고 할 때, 상기 단백질 a와 단백질 b의 상호작용에 대한 가중치는 $\text{Log}((\text{단백질 a 위치들과 단백질 b 위치들간 모든 상호작용 경우의 수} / \text{모든 위치간 상호작용 경우의 수 총합} * 100) * 100)$ 인 것이 바람직하다.

[0019] 또한, 상기 평가 함수는, 상기 단백질 상호작용에 대한 가중치인 것이 바람직하다.

[0020] 그리고, 상기 평가 함수는 상기 단백질 상호작용에 대한 가중치, 상기 단백질 상호 작용에 대응되는 각 단백질의 가중치의 곱인 것이 바람직하다.

효 과

[0021] 본 발명에 의하면, 단백질 상호작용 정보와 단백질 위치정보를 기반으로 용이하게 신호전달 경로를 추출할 수 있다.

[0022] 나아가 신호전달 경로가 추출됨으로써 질병과 유전자간의 연관성 파악이 용이해질 수 있다. 또한, 신호전달 과정을 분석함으로써 신약 물질의 대상 단백질을 결정하는데 유용하게 이용될 수 있을 뿐만 아니라, 신약에 대한 정확한 대상 단백질을 선별하는 것은 신약개발과정시 불필요한 예산낭비를 줄여줄 수 있다.

발명의 실시를 위한 구체적인 내용

[0023] 이하에서는 도면을 참조하여 본 발명을 보다 상세하게 설명한다.

[0024] 세포를 구성하는 여러 물질(Molecules) 중 단백질은 매우 중요하다. 단백질은 생명체를 구성하는 성분으로, 효소 및 항체를 형성하여 생명현상을 유지시키기 때문이다. 단백질들은 각자의 기능을 독립적으로 하는 것이 아니라, 각각의 기능을 분담하고 협력하여 하나의 생물학적 기능을 수행한다.

[0025] 예를 들어, 호르몬이 세포에 영향을 주기 위해서는, 먼저 세포 표면에 존재하는 수용체에 호르몬이 결합하여야 한다. 이 결합은 Receptor Dimerization 같은 수용체의 구조적인 변화를 수반하기도 하며, 일련의 하위 단백질들의 Signal Transduction을 유발한다. 이는 핵 내의 다양한 유전자를 단백질로 발현시켜 세포에 특정한 영향을 주게 된다.

[0026] 모든 세포는 주위 환경의 자극을 감지하고, 그에 반응한다. 단순한 생물체인 박테리아도 포도당이나 아미노산이 있는 방향으로 이동하며, 많은 단세포 생물들은 다른 생명체가 분비하는 물질에 반응한다. 다세포 생물의 경우, 한 세포는 주위의 세포들로부터 받게 되는 다양한 신호들을 해석하고, 그에 따라 자신의 반응을 조절하고 다시 신호를 보내야만 한다.

[0027] 이러한 세포 간 신호전달이 이루어지면서, 세포의 대사, 이동, 증식, 생존 그리고 분화와 같은 세포의 활동이 조절된다. 세포 내 신호전달 과정은 외부 신호전달 물질을 세포 표면의 수용체 단백질이 인식하고 이를 세포 내에 전달함으로써 이루어진다.

[0028] 신호전달 경로는 세포 내 신호전달 과정을 말한다. 호르몬이나 환경변화 같은 외부의 자극이 올 경우, 세포는 이를 감지하여 세포 내로 전달하여 이에 대한 반응을 한다. 이 때 세포의 외부에서 오는 신호를 세포 내로 전달하는 과정을 신호전달 경로라고 하며, 대표적인 세포 내 신호전달 경로는 MAP Kinase Cascade (MAPKinase 신호전달 경로)등이 있다.

[0029] 신호전달 경로의 일반적인 특징은 첫째, 세포 막에 존재하는 수용체에 신호전달 물질이 결합을 하면서 복잡한 다단계 반응이 시작된다. 신호전달 물질로는 호르몬이 가장 대표적이고, 그 밖에 신경전달 물질이나 성장인자 심지어 포도당과 지방산도 신호전달 물질 역할을 할 수 있다.

- [0030] 둘째, 신호전달 물질이 수용체에 결합한 후 이 신호는 여러 가지 방식으로 다단계 증폭된다. 일반적으로 G-단백질-cAMP, Ras-MAPK, JAK-STAT 등의 경로가 있다. 이런 일련의 다단계 반응은 주로 Kinase에 의한 인산화로 조절되며, 대부분의 단백질은 인산화가 되면 활성을 갖게 되고 인산기가 떨어져 나가면 활성을 잃게 된다.
- [0031] 셋째, 많은 경우 부신호전달자(Second Messenger)의 도움으로 신호전달이 일어난다. 칼슘이온이 대표적인 경우이고, 지방관련 물질들도 있다.
- [0032] 넷째, 이런 신호전달 경로는 생체 반응을 조절 한다. 포도당의 저장 및 합성 같은 대사반응의 방향을 조절하거나, 유전자의 발현의 억제 또는 발현을 조절하여 여러 생체 반응에 영향을 준다.
- [0033] 한편 세포는 신호전달경로 이외에 대사경로(Metabolic Pathway)도 함께 가지고 있다. 신호전달경로는 외부의 자극을 받아서 세포내부의 수용체가 이를 받아서 이에 대한 반응을 보이는 일련의 단백질 상호작용이 일어나는 현상을 말한다. 그러나 대사경로란 세포 내에 필요한 화학물질(대사산물)을 생성하는 과정을 말한다. 세포는 외부의 변화에 반응하고 대처하기 위해, 신호를 감지하는 신호전달 경로와 더불어, 이에 필요한 산물을 얻기 위한 대사경로를 동시에 가지고 있다.
- [0034] 다음은 단백질 상호작용의 활용에 대해 설명한다.
- [0035] 단백질 상호작용 경로를 구현하는 것은 생물학적 현상을 이해하는 데에 매우중요하다. 세포의 생명활동은 세포 내의 다양한 종류의 단백질이 특정 상황에서 정확하게 기능을 발휘함으로써 가능한 것이다. 특히 세포 안에서는 한 종류의 단백질이 단독적으로 작용하기 보다는 기능적으로 연관되어 있는 단백질 상호작용이 일어나므로써 대사작용 및 신호전달이 체계적으로 이루어진다.
- [0036] 단백질 상호작용의 경로를 밝혀내는 것은 특정 생물학적 작용에 대한 기능적필연성에 대한 이해를 향상시킬 수 있으며, 아직 그 특성이 규명되지 않은 단백질의 기능을 예상해볼 수 있다. 더불어 역으로 특정 단백질 상호작용 단계에서 그 존재는 밝혀졌으나 생물학적 작용이 밝혀지지 않은 단백질의 기능을 파악하는 데에 중요하다.
- [0037] 단백질 상호작용은 단백질 신호전달 경로를 알아내는 데 중요한 정보로 활용되고 있다. 외부 자극이 세포에 가해지면 세포 내의 단백질들은 복잡한 단백질 상호작용을 통해 그 신호를 전달하여 유전자 발현과 같은 현상으로 연결한다.
- [0038] 세포 내 다양한 생화학적 작용을 일으키는 단백질은 세포 내에서 단독적으로 작용할 수도 있지만, 일반적으로는 다른 단백질 상호작용이 이루어지는 경우가 대다수이다. 따라서 특정 단백질과 이와 기능적으로 연관된 다른 단백질 상호작용을 이해하는 것은 특정 생화학적작용에 있어서 필수적인 부분이라 할 수 있다.
- [0039] 이와 관련하여 세포 내 단백질(Cellular Protein)의 Subcellular Localization 정보도 고려하여야 한다. 세포 안의 물질들은 Cytoskeleton(microtubule, actin filament, intermediate filament 등) 구조의 도움을 통해 세포 내에서 제 기능을 수행하는 데에 가장 적합한 곳에 위치하게 된다. 세포 내 다양한 세포 소기관 및 여러 단백질 등이 단순히 혼합되어 있는 상태로 보일지라도 실제로는 특정 생화학적 반응의 효율을 가장 높일 수 있는 최적 상태로 맞춰져 있는 것이다.
- [0040] 단백질 상호작용도 단백질간의 물리적인 거리가 대단히 짧은 상태에서 일어나는 것이므로 이러한 반응을 이해하려면 그 각각의 세포 내 단백질 위치정보를 이해하는 것이 바람직하다.
- [0041] 세포 내의 단백질들은 특정 위치에 존재하고 있다. 이는 단백질 상호작용이 어느 곳에서나 일어나는 것이 아니고, 특정한 위치에서만 일어날 수 있다는 것이 분자생물학 실험들을 통해 증명되고 있기 때문이다.
- [0042] 단백질 상호작용은 세포막으로 둘러 쌓여 있는 세포 소구역 안에 동시에 존재하여 가까운 거리에 존재하는 단백질 간에 일어난다.
- [0043] 도 1은 세포내 단백질 위치 정보를 나타낸 도면이다. 도 1과 같은 단백질 위치 정보에 대한 자료는 신호전달 경로를 연구하기 위한 단백질 위치정보에 활용 될 수 있다.
- [0044] 이하에서는 세포 내 단백질 위치정보를 이용하여 MAP Kinase 신호전달 경로를 추출하고 예측하는 방법에 대해 설명한다.
- [0045] 단백질 상호작용은 단백질 간 물리적으로 상당히 근접한 상태에서 일어난다. 예를 들어 막으로 구성된 위치에 존재하는 단백질은 그 위치안에서 우선적으로 단백질 상호작용이 일어난다. 따라서 본 실시예에서는 세포 내 단백질 위치가 알려진 Budding Yeast 의 단백질 위치정보를 이용한다.

- [0046] 도 2는 단백질 위치정보의 예를 도시한 도면이다. 도 2에 도시된 바와 같이, YAL001C 단백질은 핵(Nucleus)와 세포질(Cytoplasm)에 존재함을 알 수 있다.
- [0047] 한편, 단백질은 한곳에 머무르는 것이 아니라 이동하며 상호작용을 하기도 하기 때문에 하나의 단백질은 여러 위치정보를 가질 수 있다.
- [0048] 본 발명에서는 단백질 5,495 개를 가지고 다중 위치정보를 고려하여 다중 위치정보 각각을 21 개의 세포 내 단백질 위치에 분포 시킨 후 얻어진 7,786 개의 단백질 위치정보 데이터를 이용한다.
- [0049] 도 3은 세포 내 단백질 다중위치 분포도를 도시한 도면이다. 도 3에 도시된 바와 같이, 단백질은 하나 이상의 위치 정보를 갖고 있음을 확인할 수 있다.
- [0050] 도 4는 본 발명의 일 실시예에 따른 단백질의 위치 정보를 기반으로 단백질의 신호 경로를 추출하는 시스템에 대한 개략적인 블록도이고, 도 5는 본 발명의 일 실시예에 따른 단백질의 위치 정보를 기반으로 단백질의 신호 경로를 추출하는 방법에 대한 흐름도이다.
- [0051] 본 단백질의 신호 경로 추출 시스템을 LocSPF(Local Signaling Pathway Finder)라 명명할 수도 있고, 시스템은 크게 전처리부(410)와 신호전달 경로 탐색부(420)로 구분할 수 있다.
- [0052] 전처리부(410)는 단백질 코드를 단일화시키는 통합부(430)와 단백질 위치의 분포도 및 단백질 위치간 상호작용 분포도를 이용하여 단백질 위치에 대한 가중치와 단백질 상호작용에 대한 가중치를 구성요소로 하는 확률값 테이블을 작성하는 제1 가중치 산출부(440)로 구성된다.
- [0053] 전처리부(410)가 수행된 후 신호전달 경로 탐색부(420)는 단백질 상호작용의 신호전달 모든 가능 경로를 탐색하는 가능 경로 탐색부(450), 가중치 적용부(460) 및 경로 결정부(470)를 포함한다.
- [0054] 구체적으로, 가능 경로 탐색부(450)는 시작 단백질, 목표 단백질, 경로길이를 입력되면, DFS(Depth First Search) 방식으로 모든 가능경로를 탐색한다. 그리고, 가중치 적용부(460)는 모든 가능경로에 대해 단백질 위치에 대한 가중치와 단백질 상호작용에 대한 가중치를 이용하여 각 경로마다 평균값을 산출한다. 마지막으로, 경로 결정부(470)는 가중치 적용부(460)에서 얻은 각 경로마다의 평균값을 내림차순으로 정렬한 후 가장 큰 값을 신호전달 경로로 결정한다.
- [0055] 이하에서는 시스템의 각 구성요소를 보다 상세히 설명한다.
- [0056] 통합부(430)는 서로 다른 형태로 제공되는 단백질 코드를 특정 단백질 코드로 단일화한다(S510).
- [0057] 도 6은 본 발명의 일 실시예에 따른 데이터베이스 통합 방식을 도식적으로 도시한 도면이다.
- [0058] 본 실시예에서는 단백질의 신호 전달에 대한 설명의 편의를 도모하기 위해 단백질 중 효모(*Saccharomyces cerevisiae*) 단백질을 대상으로 신호 전달을 파악해 본다. 그리고, 상기한 효모 단백질 위치간 상호작용 정보와 효모 단백질 위치정보는 DIP 데이터 베이스(<http://dip.doe-mbi.ucla.edu/dip>)와 MIPS 데이터베이스(<http://www.helmholtz-muenchen.de/en/mips/>)에서 제공받는다.
- [0059] 이때 각각의 데이터베이스에서 제공하는 단백질 코드는 서로 다른 형태로 제공된다. 즉, DIP에서 사용되는 코드는 P 로 시작되는 코드이며 MIPS에서 사용되는 코드는 Y 로 시작되는 코드 형태이다. 코드 단일화를 위해 NCBI 데이터 베이스(<http://www.ncbi.nlm.nih.gov>)에서 제공되는 정보를 이용하여 MIPS 코드 형식으로 변환한다.
- [0060] 이와 같이 획득된 변환된 단백질 위치간 상호작용 정보는 약 46,121개이다.
- [0061] 도 7은 본 발명의 일 실시예에 따른 단백질 상호작용을 인접행렬로 표시한 도면이다. 도 7에 도시된 바와 같이, 인접행렬에서 두 단백질 사이에 단백질 상호작용이 있으면 '1' 로 표현하고 단백질 상호작용이 없으면 빈칸으로 낸다. 그리하여 행 A6 셀에 데이터 YAL008W 와 열 E1 셀에 데이터 YAL007C 와는 단백질 상호작용이 있음을 알 수 있다.
- [0062] 가중치부(440)는 단백질의 위치 분포도와 단백질의 위치간 상호작용 분포도를 이용하여 단백질 위치에 대한 가중치와 단백질 상호작용에 대한 가중치를 산출한다(S520).
- [0063] 예를 들어, 단백질 a 와 단백질 b 가 있다고 가정할 때, 단백질의 위치 정보와 단백질 상호작용 정보를 이용하여 단백질 위치에 대한 가중치와 단백질 상호작용에 대한 가중치를 구한다.
- [0064] 가중치부(440)는 단백질 a의 위치에 대한 가중치를 단백질 위치 정보를 이용해서 $\text{Log}(\text{단백질 a 위치에 속하는})$

단백질의 누적 경우의 수 / 모든 위치의 단백질 누적 경우의 수의 총합 * 100) * 100)로 구한다.

[0065] 구체적으로, 가중치부(440)는 하기 수학적 식 1과 같은 단백질 a의 위치 정보의 확률값을 산출한다.

수학적 식 1

$$\left(\frac{|\{i | i \in \text{proteins}, a \in \text{단백질 } i \text{의 위치}\}|}{\sum_{j \in \text{proteins}} |\text{단백질 } j \text{의 위치}|} \right) * 100$$

[0066]

여기서 ‘*’ 기호는 곱하기 기호이다.

[0067]

[0068] 그리고, 상기한 확률값이 작기 때문에 크기 보정을 위해 다시 100을 곱하고 로그를 취함으로써, 가중치부(440)는 하기 수학적 식 2와 같은 단백질 a의 위치에 대한 가중치를 산출한다.

수학적 식 2

$$\text{EvL}_a = \log \left(\left(\frac{|\{i | i \in \text{proteins}, a \in \text{단백질 } i \text{의 위치}\}|}{\sum_{j \in \text{proteins}} |\text{단백질 } j \text{의 위치}|} \right) * 100 \right) * 100$$

[0069]

[0070] 또한, 가중치부(440)는 단백질 a-단백질 b의 단백질의 상호작용에 대한 가중치를 단백질 상호작용의 정보와 각 단백질의 위치정보를 이용해서 $\log((\text{단백질 } a \text{ 위치들과 단백질 } b \text{ 위치들간 모든 상호작용 경우의 수} / \text{모든 위치간 상호작용 경우의 수 총합} * 100) * 100)$ 으로 구한다.

[0071] 구체적으로, 가중치부(440)는 하기 수학적 식 3과 같은 단백질 a-단백질 b의 단백질의 상호작용에 대한 가중치를 산출한다.

수학적 식 3

$$\begin{aligned} & \text{EvL}_{(a,b)} \\ &= \log \left(\left(\frac{|\{(i,j) | (i,j) \in \text{proteins}, a \in \text{단백질 } i \text{의 위치}, b \in \text{단백질 } j \text{의 위치}\}|}{\sum_{(i,j) \in \text{proteins}} |\text{단백질 } i \text{의 위치} \times \text{단백질 } j \text{의 위치}|} \right) \right. \\ & \quad \left. * 100 \right) * 100 \end{aligned}$$

[0072]

[0073] 가능 경로 탐색부(450)는 시작 단백질, 종료 단백질 및 경로 길이가 입력되면, 시작 단백질, 종료 단백질 및 경로 길이를 기준으로 Depth First Search 방식을 통해 신호전달 가능 경로를 산출한다(540).

[0074] 단백질 신호전달 경로는 세포 내 세포막에 위치하는 단백질로부터 출발하여 핵에 존재하는 단백질까지 도달하는 과정이다. 본 실시예에서는 효모 단백질 MAPK의 "High Osmolarity" 기능을 대표 모델로 하여 신호전달 경로를 예측한다.

[0075] 예를 들어, 시작 단백질(예:YIL147C)과 종료단백질(예:YMR037C) 그리고 경로길이(예:8)를 입력할 수 있다.

[0076] 본 발명에서는 통합부(410)에서 이미 추출해 놓은 단백질 상호작용 정보 46,121 개를 사용하여, 입력 받은 시작 단백질로부터 종료 단백질까지 경로길이를 제한하여 모든 신호전달 가능 경로를 찾을 수 있다. 이때, DFS(Depth First Search) 방식을 기반으로 시작 단백질로부터의 경로길이를 비교하여 경로길이 범위안에서 종료 단백질에 도달할 때까지 신호전달 경로를 탐색한다.

[0077] 도 8은 DFS 방식을 이용하여 탐색하는 예를 도식적으로 도시한 도면이고, 도 9는 본 발명의 일 실시예에 따른 단백질 상호작용의 모든 경로를 찾는 알고리즘에 관한 도면이며, 도 10은 도 9의 알고리즘의 결과를 도시한 도면이다.

[0078] 도면에 도시된 바와 같이, 시작 단백질(예:YIL147C)과 종료 단백질(예:YMR037C)이 입력되면, 도 9의 알고리즘에

의해 도 10의 결과가 도출된다.

- [0079] 가중치 적용부(460)는 단백질 위치에 대한 가중치 및 단백질 상호작용에 대한 가중치 중 적어도 하나를 이용하여 평가 함수를 산출하고, 산출된 평가 함수로 신호 전달 가능 경로를 평가한다(S550).
- [0080] 막으로 구성된 위치에 존재하는 단백질은 그 위치 안에 존재하는 단백질들끼리 우선적으로 단백질 상호작용이 일어날 수 있다. 따라서 단백질 위치정보를 신호전달 경로 찾기의 가중치로 반영하는 것이 바람직하다.
- [0081] 가중치 적용부(460)에서 사용되는 평가함수는 하기 수학식 4로부터 산출된다.

수학식 4

$$\prod_{i,j \in P \text{ \& Interaction Path}} (\text{Protein i와 Protein j의 Interaction Criterion 확률})$$

- [0083]
- [0084] 그리고, 도 11은 본 발명의 일 실시예에 따른 수학식 4의 평가 함수를 적용한 알고리즘을 도시한 도면이다.
- [0085] 구체적으로, 단백질 상호작용 신호전달 경로 찾기를 평가하기 위해서는 패스를 이루는 단백질 앞뒤 순서 간 모든 평가 값을 곱한다.
- [0086] 예를 들어 YLR332W-YOL109W-YBR069C-YLR372W-YMR058W-YDL029W-YHR030C-YPL089C를 평가해 본다. YLR332W-YOL109W, YOL109W-YBR069C, YBR069C-YLR372W, YLR372W-YMR058W, YMR058W-YDL029W, YDL029W-YHR030C, YHR030C-YPL089C 단백질 간 모든 단백질 상호작용 평가 값을 구하고 그 값들을 곱한다.
- [0087] 단백질 상호작용의 값을 평가하기 위해 단백질 위치정보와 단백질 상호작용 위치에 대한 상호작용 확률 값을 포함하는 이차원 배열 정보를 사용한다. 이때 단백질 위치에 대한 가중치는 각 단백질 위치의 확률 값 중 가장 큰 값을 갖고 있는 단백질의 위치 값이다.
- [0088] 단백질 상호작용의 확률 값은 배열(21위치X21위치)을 이용하여 두 단백질간 상호작용에서 갖는 모든 단백질 위치 상호작용 경우를 반영한 확률 값을 사용하며 단백질 위치의 상호작용에 대한 가중치는 모든 위치에 대한 상호작용의 확률값 중 가장 큰 값을 가지는 위치 값이다.
- [0089] 예를 들어 단백질 위치정보가 YMR088C: vacuolar membrane, 세포질 YLL040C: mitochondrion, endosome, 세포질 이면, YMR088C와 YLL040C의 모든 단백질 위치의 상호작용 경우의 수는 위치 개수 2*3한 값인 6개가 발생한다. 도 12는 본 발명의 일 실시예에 따른 단백질 위치의 상호작용에 대한 모든 경우의 수를 도시한 도면이다.
- [0090] 이하에서는 단백질 위치정보와 단백질 위치의 상호작용 값을 이용하여 2가지 형태의 평가 함수를 수학식으로 표현한다.
- [0091] 첫째, 일 형태의 평가함수는 단백질 I-단백질 J 위치에 대한 상호작용 가중치 중 최고값으로 하기 수학식 5와 같다.

수학식 5

$$\max_{(a,b) \in \text{Protein i의 위치} \times \text{Protein j의 위치}} (\text{Local a와 Local b의 Interaction 확률})$$

- [0092]
- [0093] 둘째, 다른 형태의 평가함수는 단백질 I 와 단백질 J 의 단백질 상호작용 위치에 대한 상호작용 가중치 중 최고값으로 하기 수학식 6과 같다.

수학식 6

$$\max_{(a,b) \in \text{Protein i의 위치} \times \text{Protein j의 위치}} (\text{Local a와 Local b의 Interaction 확률} \times$$

$$\text{Local a의 위치 가중치} \times \text{Local b의 위치 가중치})$$

[0094]

[0095]

즉, 단백질 I 와 단백질 J 의 단백질 상호작용 위치에 대한 상호작용 가중치를 평가하기 위해서 단백질 I 와 단백질 J 의 단백질 위치정보를 이용한다.

[0096]

도 13은 본 발명의 일 실시예에 따른 단백질 상호작용 위치에 대한 상호작용의 확률 값을 도시한 도면이다. 도면에 도시된 바와 같이, 단백질 I 의 위치가 없거나 단백질 J 의 위치가 없을 경우 1.0 의 값으로 평가한다. 그리고, 단백질 I 와 단백질 J 모두 하나 이상의 위치를 가질 경우 그 위치 간 쌍의 모든 경우의 수 중에서 평가함수를 통해 산출한 값 중 가장 큰 값을 선택하고, 선택된 단백질 상호작용 값들을 곱함으로써 한 개의 단백질 상호작용 신호전달 경로 찾기 단계의 최종값으로 한다.

[0097]

도 14는 A 위치 가중치 * B 위치 가중치 * A-B 위치 상호작용 가중치를 도시한 표이다.

[0098]

경로 결정부(470)는 가중치 적용부(460)에서 산출된 모든 단백질 상호작용 경로에 대해 단백질 위치에 대한 가중치와 단백질 상호작용 위치에 대한 가중치의 결과 중 어느 하나를 동순위를 허용하면서 내림차순으로 정렬하여 가장 큰 값을 신호전달 경로로 결정한다(S560).

[0099]

도 15는 본 발명의 일 실시예에 따른 경로길이 9를 제외한 전체 신호전달 경로에 대한 수행 결과표이다. 또한, 도 16은 본 발명에 따른 신호 전달 경로를 탐색하는데 소요된 시간을 도시한 표이다.

[0100]

이하에서는 단백질 정보와 단백질 상호작용 관계를 표현한 인접행렬을 바탕으로 시작 단백질에서 종료 단백질까지 그 경로길이가 6 인 신호전달 경로를 탐색한다.

[0101]

도 17은 본 발명의 일 실시예에 따른 경로길이가 6일 때의 결과표이다. 도 17에서 알 수 있듯이 KEGG 신호전달 경로 3 개의 경로에 대해 본 발명의 제1 방법(LocSPF_방법1)과 본 발명의 제2 방법(LocSPF_방법2)은 KEGG 신호전달 경로와 일치한다.

[0102]

결과를 분석해 본 결과 경로 1에 대한 본 발명의 제1 방법은 가중치 최종 값이 상위 3%에서 KEGG에서 제공하는 신호전달 경로와 일치하는 결과를 얻으며 나머지 모두에서는 상위 1%에서 KEGG에서 제공하는 신호전달 경로와 일치하는 결과를 얻어 본 발명이 유효함을 알 수 있다.

[0103]

상기한 본 발명의 제1 방법과 본 발명의 제2 방법은 서로 다른 평가 함수를 적용한 것으로, 본 발명의 제1 방법은 단백질 A 단백질 B 의 단백질 상호작용 위치에 대한 상호작용 가중치를 평가함수로 적용하였으며, 본 발명의 제2 방법은 단백질 A 위치의 가중치, 단백질 B 위치의 가중치, 단백질 A 단백질 B 상호작용 위치에 대한 상호작용 가중치의 곱을 평가 함수로 적용한 것이다.

[0104]

“High Osmolarity” 기능에 대해 경로1의 본 발명의 제1 방법의 결과를 가지고 가중치 결과값을 기준으로 최고 값 순으로 나열하였을 때, 가중치 점수의 상위 3%이내의 수준에서 KEGG에서 제공하는 신호전달 경로와 일치함은 도 17를 통해 알 수 있다. 또한 경로1의 제2 방법, 경로2와 경로3의 제1 방법과 제2 방법을 적용한 결과의 가중치를 큰 값순으로 나열하여 KEGG에서 제공하는 신호전달 경로를 비교해 본 결과 Top스코어에서 찾을 수 있다. 결국 목표 경로 찾기에서 찾고자 하는 경로를 제1 방법과 제2 방법에 의해서 가중치가 최고 값을 갖는 그룹에서 목표 경로를 찾고 있음을 보여 준다.

[0105]

본 발명의 제1 방법 및 제2 방법에 의해 최고의 가중치 값을 갖는 신호전달 경로가 KEGG의 신호 전달 경로와 일치한다는 것은 본 발명의 제1 방법 및 제2 방법이 유용함을 증명한다.

[0106]

하지만, 동일한 가중치값을 갖는 상당수의 신호전달경로가 존재 할 시에는 좀더 체계적인 방법이 요구되게 된다.

[0107]

도 18은 LocSPF 시스템을 이용하여 구한 Top 스코어의 신호전달 경로의 수이다. 도 18에서 알 수 있는 것처럼 경로길이가 6 인 경우는 중복된 데이터의 수가 많지 않지만 경로의 길이가 7 이상되면 중복데이터의 수가 급격히 늘어난다.

[0108]

도 19는 본 발명의 제1 방법에 따른 경로 길이 5를 포함하는 경로 길이 6의 결과표이고, 도 20은 본 발명의 제2 방법에 따른 경로 길이 5를 포함하는 경로 길이 6의 결과표이다.

- [0109] 도 19 및 20에 도시된 바와 같이, "Starvation" 기능에 대해서 본 발명의 제1 방법, 본 발명의 제2 방법의 경우 모두 KEGG에서 제공하는 신호전달 경로 길이6에 대해 경로 길이5가 일치함을 알 수 있다. 이는 신호전달 경로에서 미지의 단백질이 있다는 것을 예측할 수 있다.
- [0110] 한편, KEGG에서 제공되는 코드와 실제 실험에서 사용된 MIPS 데이터베이스 제공코드가 표현이 서로 상이하다. 그리하여 코드 대조표가 필요한데, 도 21은 MIPS와 KEGG 단백질의 코드 대조표의 일 예를 도시한 도면이다.
- [0111] 한편, 아직 밝혀 지지 않은 미지의 단백질이 무수히 많이 존재 할 때, 본 발명을 통해서 앞으로 밝혀질 신호전달 경로의 단백질까지 예상하여 신호전달 경로를 추출할 수 있다.
- [0112] KEGG에서 제공하는 신호전달 경로 길이가 7일 경우, 본 발명에서는 그 경로 길이를 8로 하여 미지의 단백질을 실제 KEGG의 신호전달 경로에 삽입한다. 그 결과가 KEGG의 신호전달 경로와 동일한지 확인한다.
- [0113] 도 22는 본 발명에 따른 경로 길이가 7일 때의 결과표이고, 도 23은 본 발명에 따른 경로 길이가 8일 때의 결과표이다.
- [0114] 도 22 및 도 23에 도시된 바와 같이, KEGG MAPK의 "High Osmolarity" 기능에서 제공하는 단백질 신호전달 경로 길이를 7에서 8로 변경한다. 이는 아직 밝혀지지 않은 미지의 단백질이 경로 중간에 삽입 되어 있음을 전제하는 것이다. 그 방법과 과정은 앞의 경로길이 6일 때와 동일하다.
- [0115] "High Osmolarity" 기능에서 제공하는 신호전달 경로에 미지의 단백질이 포함되어 있음을 전제로 하여 실험을 한 결과 KEGG에서 제공하는 신호전달 경로와 일치한다는 점에서 본 발명의 신호 전달 경로 탐색 시스템 및 방법은 유용하다는 것을 알 수 있다.
- [0116] 본 발명은 단백질의 여러 위치 중 단백질 위치 가중치가 최고인 값을 선택하게 된다. 따라서 단백질 위치정보가 여러 가지일 경우, 여러 단백질 위치의 후보군 중에 위치 값이 최고인 값을 가진 단백질이 대상으로 선택된다. 이와 같은 방식으로 단백질의 신호 경로를 컴퓨터를 통해 탐색함으로써, 보다 용이하게 단백질의 신호 경로를 탐색할 수 있고, 예측할 수 있다.
- [0117] 또한, 이상에서는 본 발명의 바람직한 실시예에 대하여 도시하고 설명하였지만, 본 발명은 상술한 특정의 실시예에 한정되지 아니하며, 청구범위에서 청구하는 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 기술분야에서 통상의 지식을 가진자에 의해 다양한 변형실시가 가능한 것은 물론이고, 이러한 변형실시들은 본 발명의 기술적 사상이나 전망으로부터 개별적으로 이해되어져서는 안될 것이다.

도면의 간단한 설명

- [0118] 도 1은 세포내 단백질 위치 정보를 나타낸 도면,
- [0119] 도 2는 단백질 위치정보의 예를 도시한 도면,
- [0120] 도 3은 세포 내 단백질 다중위치 분포도를 도시한 도면,
- [0121] 도 4는 본 발명의 일 실시예에 따른 단백질의 위치 정보를 기반으로 단백질의 신호 경로를 추출하는 시스템에 대한 개략적인 블록도,
- [0122] 도 5는 본 발명의 일 실시예에 따른 단백질의 위치 정보를 기반으로 단백질의 신호 경로를 추출하는 방법에 대한 흐름도,
- [0123] 도 6은 본 발명의 일 실시예에 따른 데이터베이스 통합 방식을 도식적으로 도시한 도면,
- [0124] 도 7은 본 발명의 일 실시예에 따른 단백질 상호작용을 인접행렬로 표시한 도면,
- [0125] 도 8은 DFS 방식을 이용하여 탐색하는 예를 도식적으로 도시한 도면,
- [0126] 도 9는 본 발명의 일 실시예에 따른 단백질 상호작용의 모든 경로를 찾는 알고리즘에 관한 도면,
- [0127] 도 10은 도 9의 알고리즘의 결과를 도시한 도면,
- [0128] 도 11은 본 발명의 일 실시예에 따른 수학적 4의 평가 함수를 적용한 알고리즘을 도시한 도면,
- [0129] 도 12는 본 발명의 일 실시예에 따른 단백질 위치의 상호작용에 대한 모든 경우의 수를 도시한 도면,
- [0130] 도 13은 본 발명의 일 실시예에 따른 단백질 상호작용 위치에 대한 상호작용의 확률 값을 도시한 도면,

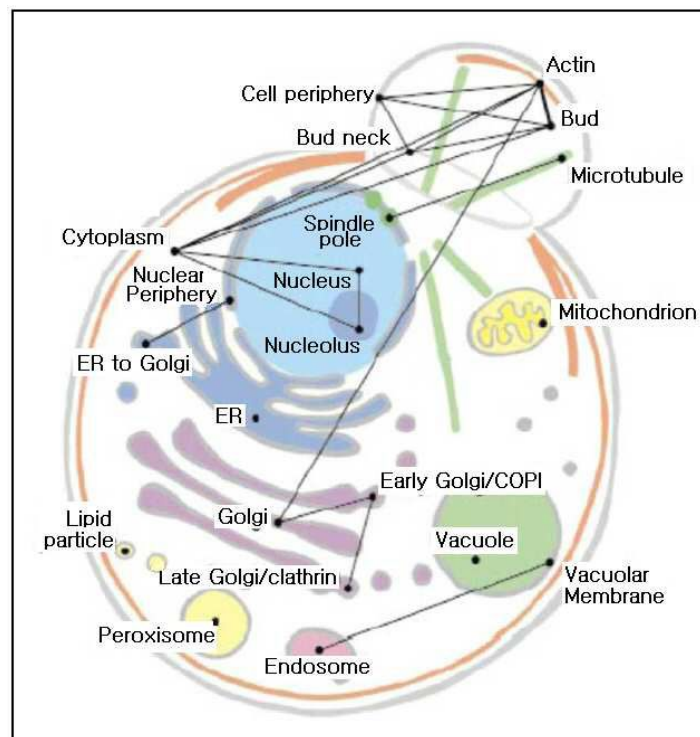
- [0131] 도 14는 A 위치 가중치 * B 위치 가중치 * A-B 위치 상호작용 가중치를 도시한 표,
- [0132] 도 15는 본 발명의 일 실시예에 따른 경로길이 9를 제외한 전체 신호전달 경로에 대한 수행 결과표,
- [0133] 도 16은 본 발명에 따른 신호 전달 경로를 탐색하는데 소요된 시간을 도시한 표,
- [0134] 도 17은 본 발명의 일 실시예에 따른 경로길이가 6일 때의 결과표,
- [0135] 도 18은 LocSPF 시스템을 이용하여 구한 Top 스코어의 신호전달 경로의 수,
- [0136] 도 19는 본 발명의 제1 방법에 따른 경로 길이 5를 포함하는 경로 길이 6의 결과표,
- [0137] 도 20은 본 발명의 제2 방법에 따른 경로 길이 5를 포함하는 경로 길이 6의 결과표,
- [0138] 도 21은 MIPS와 KEGG 단백질의 코드 대조표의 일 예를 도시한 도면,
- [0139] 도 22는 본 발명에 따른 경로 길이가 7일 때의 결과표,
- [0140] 도 23은 본 발명에 따른 경로 길이가 8일 때의 결과표이다.

* 도면의 주요 부분에 대한 부호의 설명 *

[0142]	410: 전처리부	420: 신호전달 경로 탐색부
[0143]	430: 통합부	440: 가중치부
[0144]	450: 가능 경로 탐색부	460: 가중치 적용부
[0145]	470: 경로 결정부	

도면

도면1



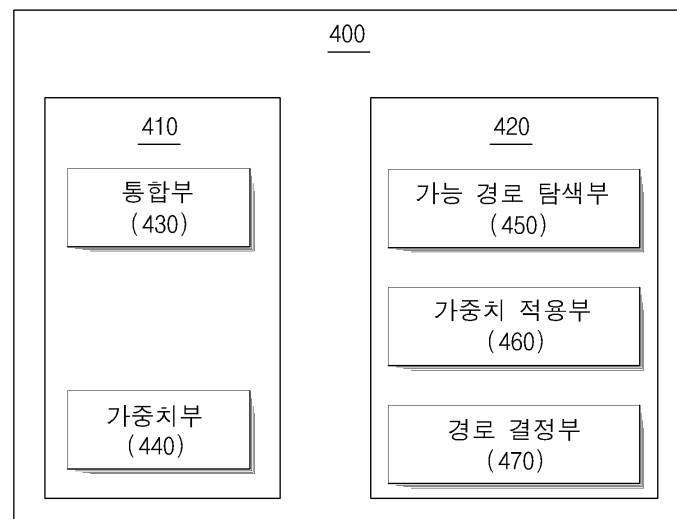
도면2

단백질 이름	위치 1	위치 2	위치 3	위치 4	위치 5
YMR089C	미토콘드리온	세포질			
YIL101C	핵	세포질			
YER139C	세포질				
YPL240C	세포질				
YDL060W	핵	세포질			
YDR024W	세포질	핵			

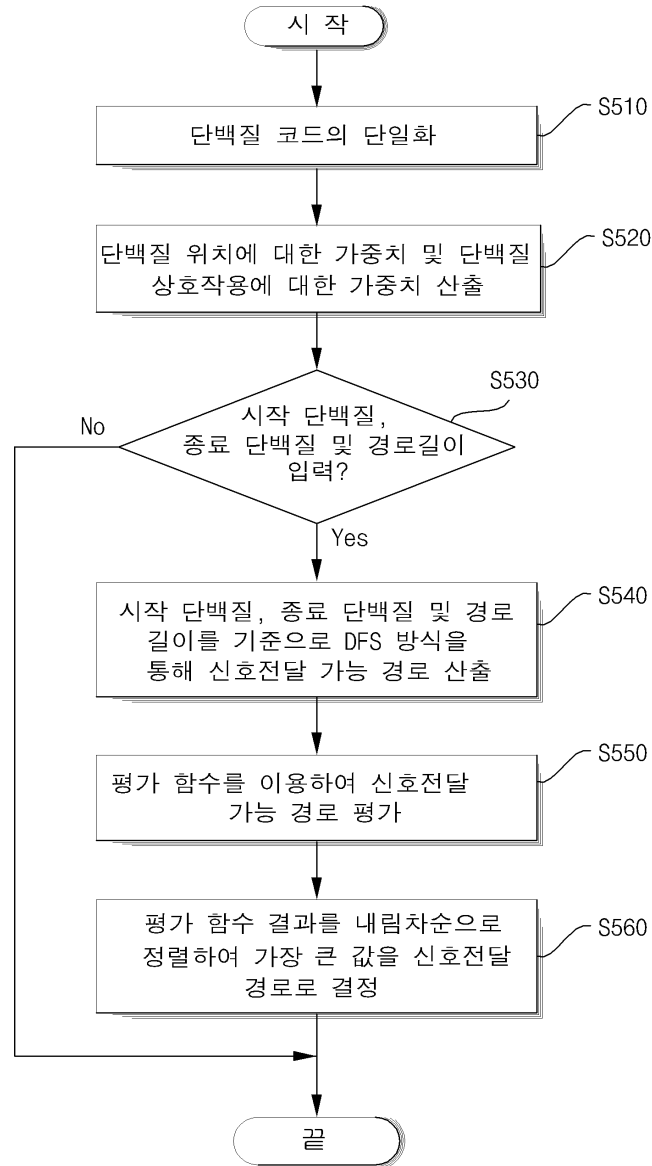
도면3

단백질 위치	단백질 개수	단백질 위치	단백질 개수
Nucleus	2,001	Golgi	180
Cell Periphery	209	Microtubule	117
Bud	88	Bud Neck	99
Lipid Particle	26	Endosome	57
Early Golgi	55	Nuclear Periphery	62
ER to Golgi	6	ER	553
Cytoplasm	2,666	Vacuolar Membrane	95
Nucleolus	207	Spindle Pole	81
Actin	32	Vacuole	220
Peroxisome	47	Mitochondrion	939
Late Golgi	46	7,786	

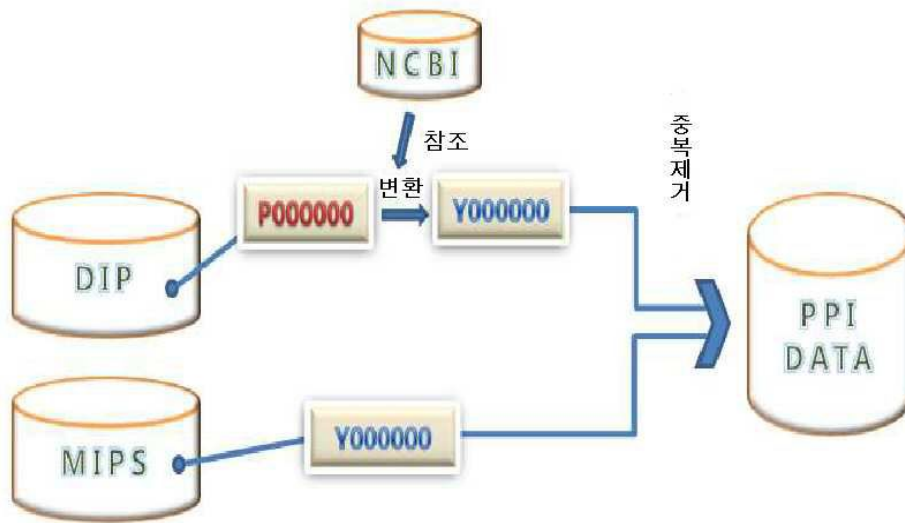
도면4



도면5



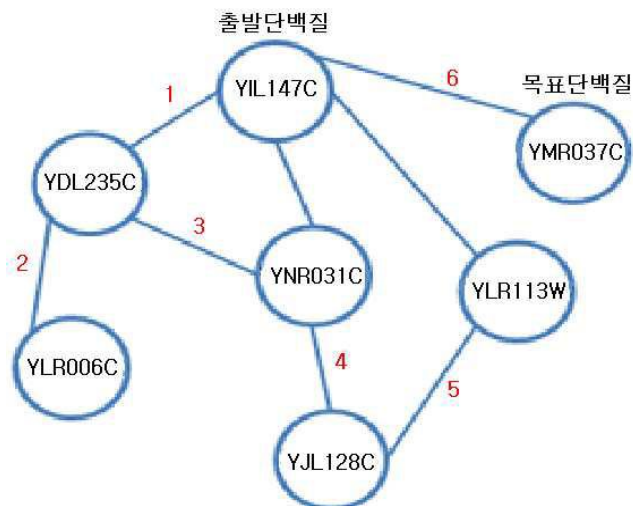
도면6



도면7

	A	B	C	D	E	F	G	H	I
1	구분	YAL001C	YAL002W	YAL005C	YAL007C	YAL008W	YAL009W	YAL010C	YAL011W
2	YAL001C			1	1				
3	YAL002W	1		1	1				
4	YAL005C	1	1		1	1	1	1	1
5	YAL007C		1	1		1		1	1
6	YAL008W		1	1	1			1	1
7	YAL009W	1	1	1	1	1		1	1
8	YAL010C	1	1		1		1		1
9	YAL011W	1	1	1	1	1	1	1	
10	YAL012W	1			1	1	1		
11	YAL014C	1			1	1	1		
12	YAL015C	1		1	1	1	1		

도면8



도면9

```

PPI Interaction Path Search
Class PathFind{
    InputDataReader reader //reading node & edge info. from input
    file)
    Input start-protein, target-protein and node maximum length
    Save this input info. into the matrix.
    Dfs (start-protein, target-protein, node maximum length)
}

```

```

//by Depth-First Search, find the possible path.

1: Dfs(start-protein, target-protein, node maximum length)
2: {
3:     push (start-protein index, link_info);
4:     for all e∈ existing adjacent nodes to the source-protein do
5:         for (current stack depth<node-maximum length) do
6:             find adjacent vertex and push the value.
7:             if (finding the target protein) break;
8:             end if
9:         end for
10:        pop the discovered possible path.
11:    end for
12: }

```

도면10

	A	B	C	D	E	F	G	H
1	start : YLR332W, target : YPL089C, maximum length : 8							
2	YLR332W	YOL109W	YBR069C	YPR19W	YHR026W	YJR091C	YDR167W	YPL089C
3	YLR332W	YOL109W	YBR069C	YLR372W	YBR054W	YJR091C	YDR167W	YPL089C
4	YLR332W	YOL109W	YBR069C	YLR372W	YJR010C	YJR091C	YDR167W	YPL089C
5	YLR332W	YOL109W	YBR069C	YLR372W	YML123C	YJR091C	YDR167W	YPL089C
6	YLR332W	YOL109W	YBR069C	YLR372W	YHR026W	YJR091C	YDR167W	YPL089C
7	YLR332W	YOL109W	YBR069C	YLR372W	YMR058W	YDL029W	YHR030C	YPL089C
8	...							
9	(중략)							

도면11

Evaluation	
Class	Evaluation{
	Reading paths, criterion formed by two dimensional array
	from input file
	Input maximum length
	Evaluate (node maximum length)
	}
	//Evaluate the score of the path.
1:	Evaluate(node maximum length)
2:	{
3:	
4:	for all path ∈ paths do
5:	for (protein1, protein2) ∈ path until maximum
	length do
6:	Multiply score with value of
	EvaluationEach(protein1, protein2)
7:	&
8:	Save into score variable
9:	end for
10:	end for
11:	}
	//Evaluate the score between two proteins
1:	EvaluateEach(protein1, protein2)
2:	{
3:	
4:	if protein1.locations = no location or protein2.locations =
	no location
5:	return 1.0f
6:	else if protein1.locations ≠ no location
	and protein2.locations ≠ no location
7:	choose max value from criterion
	each protein1.locations and protein2.locations
8:	return max value
9:	}

도면12

	A	B	C	D	E	F	G	H	I	J
1	Y56A3A2-YGL178W									
2										
3	YAL001C-YBR123C									
4	nucleus(1)-nucleus(1) mitochondrion(20)-nucleus(1) cytoplasm(5)-nucleus(1)									
5	YAL002W-YKR026C									
6	endosome(13)-cytoplasm(5)									
7	YAL002W-YLR148W									
8	endosome(13)-golgi(10) endosome(13)-vacuole(19) endosome(13)-vacuolar membrane(16) endosome(13)-									
9	YAL002W-YLR396C									
10	endosome(13)-vacuole(19) endosome(13)-vacuolar membrane(16) endosome(13)-cytoplasm(5)									
11	YAL002W-YMR231W									
12	endosome(13)-vacuole(19) endosome(13)-vacuolar membrane(16) endosome(13)-endosome(13)									
13	YAL002W-YPL045W									
14	endosome(13)-vacuole(19) endosome(13)-vacuolar membrane(16)									
15	YAL003W-YKL081W									
16	cytoplasm(5)-mitochondrion(20) cytoplasm(5)-cytoplasm(5)									
17	YAL003W-YML064C									
18	cytoplasm(5)-er(15) cytoplasm(5)-spindle pole(18)									

도면13

	golgi	microtubul	bud neck	endosome	nuclear periphery	er	vacuolar membran e	cell periphery	spindle pole	vacuole	mitochondrion
bud	0.0401	0.1126	0.1168	0.0035	0.0084	0.1259	0.0141	0.1140	0.0232	0.0408	0.0957
nucleus	0.4981	0.7161	0.3306	0.0957	0.4601	1.3542	0.1716	0.4101	0.5121	0.4284	1.8178
lipid particle	0.0077	0.0021	0.0021	0.0007	0.0028	0.0127	0.0000	0.0021	0.0014	0.0035	0.0056
early golgi	0.1435	0.0183	0.0091	0.0190	0.0063	0.1597	0.0218	0.0134	0.0070	0.0443	0.0478
er to golgi	0.0091	0.0014	0.0007	0.0000	0.0028	0.0141	0.0007	0.0014	0.0000	0.0007	0.0021
cytoplasm	0.8688	1.1185	0.6423	0.1611	0.4017	2.0134	0.3095	0.7471	0.5762	0.6725	2.0345
nucleolus	0.0345	0.0324	0.0183	0.0056	0.0218	0.0689	0.0113	0.0148	0.0176	0.0239	0.1646
actin	0.0380	0.1674	0.0457	0.0091	0.0028	0.0661	0.0056	0.0457	0.0077	0.0225	0.0739
peroxisome	0.0063	0.0035	0.0063	0.0021	0.0098	0.0302	0.0035	0.0049	0.0021	0.0070	0.0232
late golgi	0.1112	0.0317	0.0162	0.0091	0.0056	0.0689	0.0162	0.0246	0.0056	0.0310	0.0577
golgi	0.4854	0.0844	0.0514	0.0654	0.0246	0.4896	0.0696	0.0809	0.0246	0.1569	0.1949
microtubule	0.0844	0.5375	0.1850	0.0098	0.0197	0.1316	0.0120	0.1541	0.2026	0.0478	0.1899
bud neck	0.0514	0.1850	0.1787	0.0056	0.0127	0.0907	0.0120	0.1505	0.0394	0.0464	0.1273
endosome	0.0654	0.0098	0.0056	0.0422	0.0042	0.0626	0.0246	0.0120	0.0070	0.0457	0.0260
nuclear periphery	0.0246	0.0197	0.0127	0.0042	0.1548	0.1104	0.0063	0.0239	0.0204	0.0218	0.1013
er	0.4896	0.1316	0.0907	0.0626	0.1104	1.9571	0.1632	0.2884	0.0556	0.4242	0.5656
vacuolar membrane	0.0696	0.0120	0.0120	0.0246	0.0063	0.1632	0.0816	0.0253	0.0070	0.1112	0.0760
cell periphery	0.0809	0.1541	0.1505	0.0120	0.0239	0.2884	0.0253	0.1815	0.0274	0.0893	0.1864
spindle pole	0.0246	0.2026	0.0394	0.0070	0.0204	0.0556	0.0070	0.0274	0.2490	0.0204	0.1189
vacuole	0.1569	0.0478	0.0464	0.0457	0.0218	0.4242	0.1112	0.0893	0.0204	0.1928	0.1773
mitochondrion	0.1949	0.1899	0.1273	0.0260	0.1013	0.5656	0.0760	0.1864	0.1189	0.1773	0.9356

도면14

	bud	nucleus	lipid particle	early golgi	er to golgi	cytoplasm	nucleolus	actin	peroxisome	late golgi	golgi
bud	763	43,356	6	27	1	114,705	231	77	23	71	635
nucleus	43,356	38,690,790	1,208	12,465	160	44,778,920	485,743	10,811	6,219	10,360	179,393
lipid particle	6	1,208	0	2	0	2,584	23	2	1	2	36
early golgi	27	12,465	2	260	0	33,009	80	17	2	37	1,421
er to golgi	1	160	0	0	0	473	2	0	0	0	10
cytoplasm	114,705	44,778,920	2,584	33,009	473	70,130,553	403,366	29,468	16,219	25,623	416,920
nucleolus	231	485,743	23	80	2	403,366	33,218	51	14	161	1,284
actin	77	10,811	2	17	0	29,468	51	134	1	22	219
peroxisome	23	6,219	1	2	0	16,219	14	1	174	3	54
late golgi	71	10,360	2	37	0	25,623	161	22	3	98	920
golgi	635	179,393	36	1,421	10	416,920	1,284	219	54	920	15,727
microtubule	1,159	167,662	6	118	1	348,896	784	627	19	170	1,778
bud neck	1,017	65,499	5	50	0	169,519	375	145	29	74	915
endosome	18	10,912	1	60	0	24,481	66	17	6	24	671
nuclear periphery	46	57,078	5	22	1	66,396	280	6	29	16	275
er	6,128	1,498,498	182	4,857	47	2,968,302	7,892	1,170	786	1,754	48,737
vacuolar membrane	118	32,630	0	114	0	78,395	221	17	16	71	1,191
cell periphery	2,096	171,520	11	154	2	416,279	639	306	48	237	3,043
spindle pole	165	83,007	3	31	0	124,418	295	20	8	21	359
vacuole	790	188,599	20	536	1	394,452	1,089	158	73	313	6,212
mitochondrion	7,906	3,415,532	137	2,471	12	5,093,041	31,997	2,220	1,025	2,492	32,936

	microtubule	bud neck	endosome	nuclear periphery	er	vacuolar membrane	cell periphery	spindle pole	vacuole	mitochondrion
bud	1,159	1,017	18	46	6,128	118	2,096	165	790	7,906
nucleus	167,662	65,499	10,912	57,078	1,498,498	32,630	171,520	83,007	188,599	3,415,532
lipid particle	6	5	1	5	182	0	11	3	20	137
early golgi	118	50	60	22	4,857	114	154	31	536	2,471
er to golgi	1	0	0	1	47	0	2	0	1	12
cytoplasm	348,896	169,519	24,481	66,396	2,968,302	78,395	416,279	124,418	394,452	5,093,041
nucleolus	784	375	66	280	7,892	221	639	295	1,089	31,997
actin	627	145	17	6	1,170	17	306	20	158	2,220
peroxisome	19	29	6	29	786	16	48	8	73	1,025
late golgi	170	74	24	16	1,754	71	237	21	313	2,492
golgi	1,778	915	671	275	48,737	1,191	3,043	359	6,212	32,936
microtubule	7,357	2,143	66	143	8,511	133	3,767	1,920	1,231	20,867
bud neck	2,143	1,751	32	78	4,968	112	3,115	316	1,011	11,837
endosome	66	32	137	15	1,974	133	142	32	573	1,393
nuclear periphery	143	78	15	595	3,787	37	310	102	297	5,898
er	8,511	4,968	1,974	3,787	598,495	8,574	33,336	2,489	51,608	293,697
vacuolar membrane	133	112	133	37	8,574	736	503	54	2,323	6,777
cell periphery	3,767	3,115	142	310	33,336	503	7,928	464	4,108	36,586
spindle pole	1,920	316	32	102	2,489	54	464	1,634	364	9,043
vacuole	1,231	1,011	573	297	51,608	2,323	4,108	364	9,329	36,622
mitochondrion	20,867	11,837	1,393	5,898	293,697	6,777	36,586	9,043	36,622	824,966

도면15

KEGG_단백질수	실험_단백질수	찾은_단백질수	방법1_Pathway수	방법1(%)	방법2_Pathway수	방법2(%)
6	6	6	3	75%	3	75%
		5	1	25%	1	25%
		합계	4	100%	4	100%
7	8	7	3	33%	3	33%
		5	3	33%	4	44%
		4	2	22%	1	11%
		2	1	11%	1	11%
		합계	9	100%	9	100%

도면16

시작단백질	종료단백질	KEGG-Depth	본 System	Program_Start	Program_End	OutputFile_Size	소요시간
YER118C	YMR037C	6	6	2008-11-11 5:42 PM	2008-11-11 6:33 PM	2.12k	0시간 51분
YER118C	YKL062W	6	6	2008-11-11 5:45 PM	2008-11-11 6:43 PM	10.6k	0시간 58분
YER118C	YMR043W	6	6	2008-11-11 5:47 PM	2008-11-11 6:38 PM	34.3k	0시간 51분
YER118C	YGR040W	6	6	2008-11-11 5:49 PM	2008-11-11 6:41 PM	175k	0시간 52분
YLR332W	YPL089C	7	8	2008-11-04 4:23 AM	2008-11-04 4:38 PM	639k	12시간 15분
YLR332W	YER111C	7	8	2008-11-04 4:31 AM	2008-11-04 6:26 PM	1.15m	13시간 55분
YLR332W	YLR182W	7	8	2008-11-04 4:27 AM	2008-11-04 4:39 PM	2.13m	12시간 12분
YIL147C	YMR037C	7	8	2008-11-04 4:15 AM	2008-11-06 10:06 PM	215k	17시간 51분
YIL147C	YKL062W	7	8	2008-11-04 4:16 AM	2008-11-05 9:55 AM	506k	5시간 39분
YIL147C	YMR043W	7	8	2008-11-04 4:21 AM	2008-11-05 10:54 AM	2.32m	6시간 33분

도면17

Function	경로	구분	Pathway						방법1_찾은노드수	방법2_찾은노드수
			1	2	3	4	5	6		
High osmolarity	경로1	MAPK	YER118C	YHL007C	YLR362W	YJL128C	YLR113W	YMR043W	6	6
		LocSPF_방법1	YIL147C	YDL235C	YLR006C	YNR031C	YJL128C	YLR113W	6	6
		LocSPF_방법2	YIL147C	YDL235C	YLR006C	YNR031C	YJL128C	YLR113W	6	6
	경로2	MAPK	YER118C	YHL007C	YLR362W	YJL128C	YLR113W	YMR037C	6	6
		LocSPF_방법1	YIL147C	YDL235C	YLR006C	YNR031C	YJL128C	YLR113W	6	6
		LocSPF_방법2	YIL147C	YDL235C	YLR006C	YNR031C	YJL128C	YLR113W	6	6
	경로3	MAPK	YER118C	YHL007C	YLR362W	YJL128C	YLR113W	YKL062W	6	6
		LocSPF_방법1	YIL147C	YDL235C	YLR006C	YNR031C	YJL128C	YLR113W	6	6
		LocSPF_방법2	YIL147C	YDL235C	YLR006C	YNR031C	YJL128C	YLR113W	6	6

도면18

Function	KEGG_ 경로길이	LocSPF_ 경로길이	경로	구분	Start 단백질	Target 단백질	Top스코어 증분데이터_수
High Osmolarity	7	8	경로1	LocSPF_방법1	YIL147C	YKL062W	285
				LocSPF_방법2	YIL147C	YKL062W	285
			경로2	LocSPF_방법1	YIL147C	YMR037C	181
				LocSPF_방법2	YIL147C	YMR037C	181
			경로3	LocSPF_방법1	YIL147C	YMR043W	8
				LocSPF_방법2	YIL147C	YMR043W	1218
	6	6	경로1	LocSPF_방법1	YER118C	YMR043W	9
				LocSPF_방법2	YER118C	YMR043W	27
			경로2	LocSPF_방법1	YER118C	YMR037C	14
				LocSPF_방법2	YER118C	YMR037C	15
			경로3	LocSPF_방법1	YER118C	YKL062W	26
				LocSPF_방법2	YER118C	YKL062W	26

도면19

VER118C YLR229C YHL007C YLR362W YDL159W YGR040W								
	A	B	C	D	E	F	G	H
1	방법1	node수	art :VER118C	target:YGR040W		maximum length :6		
2	530415.3	5	VER118C	YER114C	YHL007C	YLR362W	YDL159W	YGR040W
3	530415.3	5	VER118C	YHL007C	YLR362W	YLR313C	YDL159W	YGR040W
4	530415.3	5	VER118C	YHL007C	YLR362W	YLL021W	YDL159W	YGR040W
5	530415.3	5	VER118C	YHL007C	YLR362W	YDR103W	YDL159W	YGR040W
6	530415.3	5	VER118C	YHL007C	YLR362W	YDL159W	YDR103W	YGR040W
7	530415.3	5	VER118C	YHL007C	YLR362W	YBL016W	YDL159W	YGR040W

도면20

VER118C YLR229C YHL007C YLR362W YDL159W YGR040W								
	A	B	C	D	E	F	G	H
1	방법2	node수	art :VER118C	target:YGR040W		maximum length :6		
2	15576.61	5	VER118C	YER114C	YHL007C	YLR362W	YDL159W	YGR040W
3	15576.61	5	VER118C	YHL007C	YLR362W	YLR313C	YDL159W	YGR040W
4	15576.61	5	VER118C	YHL007C	YLR362W	YLL021W	YDL159W	YGR040W
5	15576.61	5	VER118C	YHL007C	YLR362W	YDR103W	YDL159W	YGR040W
6	15576.61	5	VER118C	YHL007C	YLR362W	YDL159W	YDR103W	YGR040W
7	15576.61	5	VER118C	YHL007C	YLR362W	YBL016W	YDL159W	YGR040W

도면21

KEGG	MIPS	KEGG	MIPS	KEGG	MIPS
BCK1	YJL095W	MLP1	YKR095W	SSK1	YLR006C
CDC42	YLR229C	MSN2	YMR037C	SSK2	YNR031C
FAR1	YJL157C	MSN4	YKL062W	STE11	YLR362W
FUS3	YBL016W	PBS2	YJL128C	STE12	YHR084W
GPA1	YHR005C	PKC1	YBL105C	STE18	YJR086W
HOG1	YLR113W	RAS2	YNL098C	STE2	YFL026W
KSS1	YGR040W	RHO1	YPR165W	STE20	YHL007C
MCM1	YMR043W	RLM1	YPL089C	STE3	YFL026W
MID2	YLR332W	SHO1	YER118C	STE4	YOR212W
MKK1	YOR231W	SLN1	YIL147C	STE7	YDL159W
MKK2	YPL140C	SLT2	YHR030C	YPD1	YDL235C

도면22

Function	경로	구분	Pathway							방법1_찾은노드수	방법2_찾은노드수
			1	2	3	4	5	6	7		
High osmolarity	경로1	MAPK	YIL147C	YDL235C	YLR006C	YNR031C	YJL128C	YLR113W	YKL062W	7	7
		LocSPF_방법1	YIL147C	YDL235C	YLR006C	YNR031C	YJL128C	YLR113W	YKL062W	7	7
		LocSPF_방법2	YIL147C	YDL235C	YLR006C	YNR031C	YJL128C	YLR113W	YKL062W	7	7
	경로2	MAPK	YIL147C	YDL235C	YLR006C	YNR031C	YJL128C	YLR113W	YMR037C	7	7
		LocSPF_방법1	YIL147C	YDL235C	YLR006C	YNR031C	YJL128C	YLR113W	YMR037C	7	7
		LocSPF_방법2	YIL147C	YDL235C	YLR006C	YNR031C	YJL128C	YLR113W	YMR037C	7	7
	경로3	MAPK	YIL147C	YDL235C	YLR006C	YNR031C	YJL128C	YLR113W	YMR043W	7	7
		LocSPF_방법1	YIL147C	YDL235C	YLR006C	YNR031C	YJL128C	YLR113W	YMR043W	7	7
		LocSPF_방법2	YIL147C	YDL235C	YLR006C	YNR031C	YJL128C	YLR113W	YMR043W	7	7

도면23

Function	경로	구분	Pathway								방법1_찾은노드수	방법2_찾은노드수
			1	2	3	4	5	6	7	8		
High Osmolarity	경로1	MAPK	YIL147C	YDL235C	YLR006C	YNR031C	YJL128C	YLR113W	YKL062W		7	7
		LocSPF_방법1_1	YIL147C	YDL235C	YLR006C	YNR031C	YJL128C	YLR362W	YLR113W	YKL062W	7	7
		LocSPF_방법2_1	YIL147C	YDL235C	YLR006C	YNR031C	YJL128C	YLR362W	YLR113W	YKL062W	7	7
		LocSPF_방법2_2	YIL147C	YDL235C	YLR006C	YDL013W	YNR031C	YJL128C	YLR113W	YKL062W	7	7
	경로2	MAPK	YIL147C	YDL235C	YLR006C	YNR031C	YJL128C	YLR113W	YMR037C		7	7
		LocSPF_방법1_1	YIL147C	YDL235C	YLR006C	YNR031C	YJL128C	YLR362W	YLR113W	YMR037C	7	7
		LocSPF_방법1_2	YIL147C	YDL235C	YLR006C	YDL013W	YNR031C	YJL128C	YLR113W	YMR037C	7	7
		LocSPF_방법2_1	YIL147C	YDL235C	YLR006C	YNR031C	YJL128C	YLR362W	YLR113W	YMR037C	7	7
		LocSPF_방법2_2	YIL147C	YDL235C	YLR006C	YDL013W	YNR031C	YJL128C	YLR113W	YMR037C	7	7
	경로3	MAPK	YIL147C	YDL235C	YLR006C	YNR031C	YJL128C	YLR113W	YMR043W		7	7
		LocSPF_방법1_1	YIL147C	YDL235C	YLR006C	YNR031C	YJL128C	YLR362W	YLR113W	YMR043W	7	7
		LocSPF_방법2_1	YIL147C	YDL235C	YLR006C	YNR031C	YJL128C	YLR362W	YLR113W	YMR043W	7	7
		LocSPF_방법2_2	YIL147C	YDL235C	YLR006C	YDL013W	YNR031C	YJL128C	YLR113W	YMR043W	7	7