(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2011년09월23일
(11) 등록번호 10-1067352
(24) 등록일자 2011년09월19일

(51) Int. Cl.
G06F 19/00 (2011.01) *G01N 33/48* (2006.01)
(21) 출원번호 10-2009-0111749
(22) 출원일자 2009년11월19일
심사청구일자 2009년11월19일
(65) 공개번호 10-2011-0054926
(43) 공개일자 2011년05월25일
(56) 선행기술조사문헌
KR1020050060646 A
US20040024532 A1
KR1020030071225 A
KR1020080030142 A
전체 청구항 수 : 총 4 항

(73) 특허권자
한국생명공학연구원
대전 유성구 어은동 52번지
(72) 발명자
허철구
대전광역시 유성구 어은동 한빛아파트 133동 604호
최준경
대전광역시 유성구 신성동 121-11번지 201호
(뒷면에 계속)
(74) 대리인
최규환

심사관 : 이정숙

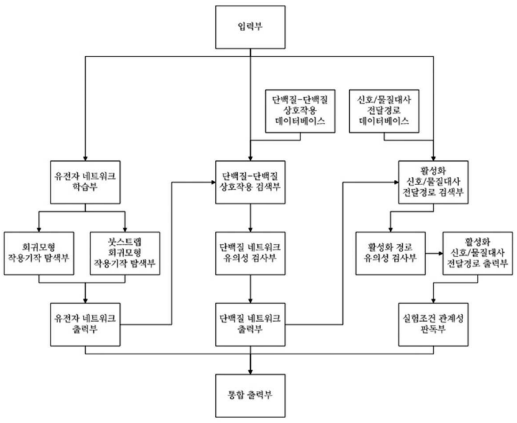
(54) 생물학적 네트워크 분석을 이용한 마이크로어레이 실험 자료의 작용기작, 실험/처리 조건 특이적 네트워크 생성 및 실험/처리 조건 관계성 해석을 위한 알고리즘을 포함한 시스템 및 방법과 상기 방법을 수행하기 위한 프로그램을 갖는 기록매체

(57) 요 약

본 발명은 마이크로어레이(microarray) 실험 자료를 이용하여 유전자 네트워크(gene network), 실험/처리 조건 특이적 단백질 네트워크(experiment/treatment condition-specific protein network), 실험/처리 조건 관계성 규명을 분석하기 위한 알고리즘을 포함한 시스템 및 방법과 상기 방법을 수행하기 위한 컴퓨터로 판독 가능한 프로그램을 기록한 기록매체에 관한 것이다.

본 발명은 어떤 질환 또는 질환 상태에 투여된 약물과 같이 마이크로어레이 실험 조건에서 반응하는 유전자를 여러 가지 네트워크로 분석하여 실험 조건과 관련성이 높은 유전자를 탐색하고, 탐색된 유전자의 단백질-단백질 상호작용 정보와 신호/물질대사 전달경로에 대한 정보를 활용하여 그 기능을 해석하는 것에 목적을 두고 있다.

대 표 도 - 도1



(72) 발명자

서승원

대전광역시 대덕구 와동 주공아파트 114동 406호

이승원

대전광역시 유성구 지족동 허브아파트 1114호

특허청구의 범위

청구항 1

단일 종에서 녹아웃(knockout), 약물(drug), RNAi, 과발현과 같은 다양한 조건의 처리를 수행한 마이크로어레이 실험 자료와 질환이나 질환 상태에서 투여된 약물과 같은 작용기작을 확인하기 위해 측정된 마이크로어레이 실험 자료를 입력하는 입력부;

입력된 마이크로어레이 실험 자료로부터 유전자 네트워크를 예측하기 위한 유전자 네트워크 학습부;

MNI(Mode-of-action by Network Identification) 기법을 이용하거나 붓스트랩핑 회귀 모형(Bootstrapping Regression Model)을 적용하여 실험 조건에서 작동하는 작용기작에 대한 유전자의 예측 순위 정보를 제공하는 작용기작을 탐색하는 알고리즘으로 구성된 탐색부;

학습한 유전자 네트워크와 탐색된 작용기작을 연동하여 유전자 네트워크를 출력하는 유전자 네트워크 출력부;

입력된 마이크로어레이 실험 자료에서 유의하게 발현된 유전자 또는 탐색된 작용기작을 이용하여 단백질-단백질 상호작용 정보를 검색하는 알고리즘으로 구성된 검색부;

검색된 단백질-단백질 상호작용 정보를 GNEA(Geneset Network Enrichment Analysis) 기법을 이용하여 마이크로어레이 실험 자료로부터 추출된 유의한 유전자 또는 탐색된 작용기작이 이루는 단백질 네트워크의 유의성을 검사하는 단백질 네트워크 유의성 검사부;

검사된 유의성에 따라 실험/조건 특이적인 단백질 네트워크를 출력하는 단백질 네트워크 출력부;

입력된 마이크로어레이 실험 자료에서 유의하게 발현된 유전자 또는 탐색된 실험/처리 조건 특이적인 단백질 네트워크에 포함된 유전자/단백질이 신호/물질대사 전달경로에서 존재하는지를 확인하는 활성화 신호/물질대사 전달경로를 검색하는 알고리즘으로 구성된 검색부;

검색된 활성화 신호/물질대사 전달경로(active signaling/metabolic pathway)에 포함되는 유전자의 목록을 특이값 분해(SVD; Singular Value Decomposition) 기법으로 압축하여 검색된 활성화 신호/물질대사 전달경로가 실제로 유의성을 가지고 있는지 검사하는 활성화 경로 유의성 검사부;

유의하게 밝혀진 활성화 신호/물질대사 전달경로를 표시하는 활성화 신호/물질대사 전달경로 출력부;

유의하게 발현된 유전자 또는 탐색된 실험/처리 조건 특이적인 단백질 네트워크에 포함된 유전자/단백질을 활성화 신호/물질대사 전달경로에 연동하여 실험/처리의 조건의 관계성을 해석하는 실험조건 관계성 판독부; 및

상기 유전자 네트워크 출력부, 단백질 네트워크 출력부, 활성화 신호/물질대사 전달 경로 출력부에 판독된 실험조건 관계성을 부여하여 생물학적 통합 네트워크를 출력하는 알고리즘으로 구성된 통합 출력부

를 포함하는 마이크로어레이 실험을 통하여 측정된 유전자의 발현값을 이용하여 유전자 네트워크, 실험/처리 조건 특이적 단백질 네트워크, 활성화 신호/물질대사 전달경로를 예측/분석하거나 실험/처리 조건의 관계성을 판독하여 생물학적 통합 네트워크로 출력하는 것을 특징으로 하는 시스템.

청구항 2

제1항에 있어서, 상기 탐색부의 알고리즘은 입력된 마이크로어레이 실험 자료로부터 유전자 네트워크 모형을 학습하고 작용기작을 예측하기 위하여 하기 식 4에 의해 붓스트랩핑 회귀 모형 방법을 이용하여 작용기작을 탐색하고 이를 유전자 네트워크로 출력하는 알고리즘인 것을 특징으로 하는 시스템:

$$\hat{\alpha}_{ij}^{*} = \frac{1}{B}\sum_{b=1}^{B}\hat{\alpha}_{ij}^{*b} \qquad \text{(식 4)}$$

(식 중, B는 붓스트랩핑 과정의 반복횟수를 나타내고, $\hat{\alpha}_{ij}^{*}$ 는 유전자 i에 대해 붓스트랩 표본으로 구한 회귀계수, $\hat{\alpha}_{ij}^{*}$ 는 주어진 관측치 쌍에서 복원랜덤추출한 붓스트랩 표본 $\boldsymbol{z}^{*} = (z_1^{*}, z_2^{*}, \cdots, z_h^{*})$ 을 적합시킨 회귀계수

를 나타낸다).

청구항 3

제1항에 있어서, 상기 단백질-단백질 상호작용 정보를 검색하는 알고리즘은 입력된 마이크로어레이 실험 자료의 유의한 유전자와 학습된 유전자 네트워크로부터 탐색된 작용기작으로 구성된 유전자 또는 단백질의 목록을 또는 작용기작만을 실험/처리 조건 특이적 단백질 네트워크로 검색하거나 유의성을 검사하는 알고리즘인 것을 특징으로 하는 시스템.

청구항 4

제1항에 있어서, 상기 활성화 신호/물질대사 전달경로를 검색하는 알고리즘은 입력된 마이크로어레이 실험 자료의 유의한 유전자와 학습된 유전자 네트워크로부터 탐색된 작용기작, 유의성이 검사된 실험/처리 조건 특이적 단백질 네트워크에 포함되는 유전자/단백질 목록 중에서 두가지 이상의 요소로 구성된 유전자 또는 단백질의 목록을 이용하여 활성화 신호/물질대사 전달경로에 검색하거나 실험/처리 조건의 관계성을 판독하는 알고리즘인 것을 특징으로 하는 시스템.

청구항 5

삭제

청구항 6

삭제

청구항 7

삭제

청구항 8

삭제

명 세 서

발명의 상세한 설명

기 술 분 야

[0001] 본 발명은 마이크로어레이(microarray) 실험 자료를 이용하여 유전자 네트워크(gene network), 실험/처리 조건 특이적 단백질 네트워크(experiment/treatment condition-specific protein network), 실험/처리 조건 관계성 규명을 위한 생물학적 통합 네트워크를 분석하기 위한 알고리즘을 포함한 시스템 및 방법과 상기 방법을 수행하기 위한 컴퓨터로 판독 가능한 프로그램을 기록한 기록매체에 관한 것이다.

배 경 기 술

[0002] 질환 상태의 세포에서 발현되는 유전자는 정상 상태의 세포에서 발현되는 유전자와 발현되는 양상이 틀려지는 것은 자명한 사실이다. 그러나 유전자가 기능을 발휘하는 것은 단일 유전자의 변화로 일어나는 것보다 여러 유전자가 상호작용을 하며 이루어지는 경우가 더 많고 질환 상태의 타켓 유전자와 상호작용하는 관련 유전자가 동시에 발현하기 때문에, 이러한 유전자의 발현이 어떻게 변화되는지를 측정하는 것은 상당히 어려운 일이다.

[0003] 1995년에 대량의 유전자가 발현하는 형태를 탐색하는 마이크로어레이 기법이 개발되었다(Schena et al. 1995. Science 270: 467-470). 이러한 마이크로어레이 기술이 점점 발달하면서 인간(human), 마우스(mouse)와 같은 고등생물의 모든 유전자가 특정 실험 조건에서 어떻게 발현이 변화되는지 한 번의 실험으로 밝히는 것이 가능하게 되었다.

[0004] 유전자는 여러 유전자가 상호작용하여 기능을 발휘하는 경우가 많은데, 이는 단백질(protein), 신호 전달(signaling pathway)이나 신진 대사(metabolism)에도 동일하게 해당된다. 이렇게 생명체의 여러 요소가 상호 복합적으로 작동하여 만들어지는 것을 생물학적 네트워크라고 칭하며, 생물학적 네트워크는 크게 유전자 네트워크

(gene network), 단백질-단백질 상호작용(protein-protein interaction), 신호/물질대사 전달경로(signaling/metabolic pathway)로 구분할 수 있다.

[0005] 마이크로어레이 기술을 이용하여 유전자 상호간에 어떻게 작동을 하는지를 학습하여 밝히는 유전자 네트워크 기법은 Friedman (2000. J. Comput. Biol. 7(3-4):601-20)에 의해 소개되었다. 이 기법은 시간의 변화나 실험의 변화에 따른 유전자의 발현 외형(expression profile)을 조사하여 상호간의 관련성이 높은 유전자를 밝히고, 두 유전자를 연결함으로써 유전자간의 연결성을 가진 네트워크를 생성할 수 있다. 이러한 네트워크를 이용하여 유전자간의 관계성을 탐색하고 유전자의 기능을 해석할 수 있다.

[0006] 단백질(protein)은 생명체내에서 유전자가 기능을 발휘하기 위하여 작동하는 분자로서 유전자와 마찬가지로 단백질이 단독으로 기능을 하는 경우보다, 두 개 이상의 폴리펩티드(polypeptide)가 아미노산(amino acid)간의 인력에 의해 결합하여 단백질 복합체(protein complex)를 이룬 형태처럼 (예, 헤모글로빈(hemoglobin): α글로빈(α-globin)과 β글로빈(β-globin) 한 쌍씩 4개의 소단위체로 구성) 여러 단백질이 복합적으로 작용하여 기능을 발휘하는 경우가 더욱 많다. 이러한 단백질간의 연결성에 대한 정보를 단백질-단백질 상호작용이라고 하며, 이 단백질-단백질 상호작용은 생명체가 생존하기 위한 여러 가지 기능을 발휘하는 단위로서 유전자의 기능을 파악하거나 해석하기 위한 필수적인 정보이다.

[0007] 신호/물질대사 전달경로는 세포의 대사, 이동, 증식, 생존, 분화 또는 시신경의 움직임과 같이 특정 기능을 수행하는 유전자와 단백질의 집합체이며, 상기의 유전자 네트워크, 단백질-단백질 상호작용 또는 유전자-단백질 상호작용을 포함하고 있다. 이러한 신호/물질대사 전달경로에는 유전자의 발현을 조절하는 전사 조절 인자(transcription factor) 단백질을 포함한 유전자와 단백질의 결합 네트워크 모형이다.

[0008] 질환 또는 질환 상태에 투여된 약물에 반응하는 유전자 발현 변화를 확인하기 위하여 기존에는 발현비율(fold-change), t-test, SAM(Significance Analysis of Microarrays) 등의 방법을 이용하여 정상(normal) 상태의 샘플(sample)을 이용한 마이크로어레이 실험과 질환 상태의 샘플 또는 질환 상태에서 약물이 투여된 상태의 샘플을 서로 비교하여 마이크로어레이 실험 사이의 유전자 발현 변화량을 계산하는 기법을 사용하였다. 그러나 이러한 방법은 실험을 수행할 때 마다 생기는 유전자의 발현 값 오차에 의해 잘못된 결과가 발생할 수 있으며, 질환 상태와 정상 상태의 유전자 발현 변화비가 높게 나와도 질환에 직접적인 관련이 있는 유전자인지 간접적으로 영향을 받는 유전자인지, 아무런 영향을 받지 않는 유전자인지 확인할 수가 없다. 그러나 만약 그러한 유전자들을 네트워크의 형태로 그릴 수 있다면, 조사된 질환 또는 약물 타깃과 같은 실험 처리에서 유전자가 다른 유전자에 주는 영향을 확인할 수 있다.

발명의 내용

해결 하고자하는 과제

[0009] 본 발명은 상기와 같은 요구에 의해 도출된 것으로서, 질환 또는 질환 상태에 투여된 약물과 같은 실험 처리에서 마이크로어레이 실험 조건에 반응하는 유전자의 작용기작을 탐색하고, 유전자 네트워크 또는 단백질 네트워크와 같은 생물학적 네트워크를 생성하여 발현 유전자의 인과관계를 확인하고, 실험/처리간의 관계성을 판독함으로써 발현 유전자의 기능을 분석하고자 한다.

과제 해결수단

[0010] 상기 과제를 해결하기 위해, 본 발명은 마이크로어레이(microarray) 실험 자료를 이용하여 유전자 네트워크(gene network), 실험/처리 조건 특이적 단백질 네트워크(experiment/treatment condition-specific protein network), 실험/처리 조건 관계성 규명을 위한 생물학적 통합 네트워크를 분석하기 위한 알고리즘을 포함한 시스템 및 방법을 제공한다.

[0011] 또한, 본 발명은 상기 방법을 수행하기 위한 컴퓨터로 판독 가능한 프로그램을 기록한 기록매체를 제공한다.

효 과

[0012] 마이크로어레이 자료로부터 생물학적 통합 네트워크를 이용하여 질환이나 질환 상태에 투여된 약물과 같은 실험/처리의 타깃 유전자를 밝히는 것은 기존의 방법으로는 알지 못하는 유전자/단백질을 파악하고, 유전자의 연관성을 확인할 수 있으며, 약물 개발의 사전 자료, 약물에 반응하는 유전자 후보군을 찾고 유전자/단백질의 기능을 손쉽게 파악할 수 있도록 할 수 있다.

발명의 실시를 위한 구체적인 내용

[0013] 본 발명의 목적을 달성하기 위하여, 본 발명은

[0014] 단일 종에서 녹아웃(knockout), 약물(drug), RNAi, 과발현과 같은 다양한 조건의 처리를 수행한 마이크로어레이 실험 자료와 질환이나 질환 상태에서 투여된 약물과 같은 작용기작을 확인하기 위해 측정된 마이크로어레이 실험 자료를 입력하는 입력부;

[0015] 입력된 마이크로어레이 실험 자료로부터 유전자 네트워크를 예측하기 위한 유전자 네트워크 학습부;

[0016] MNI(Mode-of-action by Network Identification) 기법을 이용하거나 붓스트랩핑 회귀 모형(Bootstrapping Regression Model)을 적용하여 실험 조건에서 작동하는 작용기작에 대한 유전자의 예측 순위 정보를 제공하는 작용기작을 탐색하는 알고리즘으로 구성된 탐색부;

[0017] 학습한 유전자 네트워크와 탐색된 작용기작을 연동하여 유전자 네트워크를 출력하는 유전자 네트워크 출력부;

[0018] 입력된 마이크로어레이 실험 자료에서 유의하게 발현된 유전자 또는 탐색된 작용기작을 이용하여 단백질-단백질 상호작용 정보를 검색하는 알고리즘으로 구성된 검색부;

[0019] 검색된 단백질-단백질 상호작용 정보를 GNEA(Geneset Network Enrichment Analysis) 기법을 이용하여 마이크로어레이 실험 자료로부터 추출된 유의한 유전자 또는 탐색된 작용기작이 이루는 단백질 네트워크의 유의성을 검사하는 단백질 네트워크 유의성 검사부;

[0020] 검사된 유의성에 따라 실험/조건 특이적인 단백질 네트워크를 출력하는 단백질 네트워크 출력부;

[0021] 입력된 마이크로어레이 실험 자료에서 유의하게 발현된 유전자 또는 탐색된 실험/처리 조건 특이적인 단백질 네트워크에 포함된 유전자/단백질이 신호/물질대사 전달경로에서 존재하는지를 확인하는 활성화 신호/물질대사 전달경로를 검색하는 알고리즘으로 구성된 검색부;

[0022] 검색된 활성화 신호/물질대사 전달경로(active signaling/metabolic pathway)에 포함되는 유전자의 목록을 특이값 분해(SVD; Singular Value Decomposition) 기법으로 압축하여 검색된 활성화 신호/물질대사 전달경로가 실제로 유의성을 가지고 있는지 검사하는 활성화 경로 유의성 검사부;

[0023] 유의하게 밝혀진 활성화 신호/물질대사 전달경로를 표시하는 활성화 신호/물질대사 전달경로 출력부;

[0024] 유의하게 발현된 유전자 또는 탐색된 실험/처리 조건 특이적인 단백질 네트워크에 포함된 유전자/단백질을 활성화 신호/물질대사 전달경로에 연동하여 실험/처리의 조건의 관계성을 해석하는 실험조건 관계성 판독부; 및

[0025] 상기 유전자 네트워크 출력부, 단백질 네트워크 출력부, 활성화 신호/물질대사 전달 경로 출력부에 판독된 실험조건 관계성을 부여하여 생물학적 통합 네트워크를 출력하는 알고리즘으로 구성된 통합 출력부

[0026] 를 포함하는 마이크로어레이 실험을 통하여 측정된 유전자의 발현값을 이용하여 유전자 네트워크, 실험/처리 조건 특이적 단백질 네트워크, 활성화 신호/물질대사 전달경로를 예측/분석하거나 실험/처리 조건의 관계성을 판독하여 생물학적 통합 네트워크로 출력하는 것을 특징으로 하는 시스템을 제공한다.

[0027] 본 발명의 일 구현예에 따른 시스템에 있어서, 상기 탐색부의 알고리즘은 입력된 마이크로어레이 실험 자료로부터 유전자 네트워크 모형을 학습하고 작용기작을 예측하기 위하여 하기 식 4에 의해 붓스트랩핑 회귀 모형 방법을 이용하여 작용기작을 탐색하고 이를 유전자 네트워크로 출력하는 알고리즘일 수 있다:

[0028]
$$\hat{\alpha}_{ij}^{*} = \frac{1}{B}\sum_{b=1}^{B}\hat{\alpha}_{ij}^{*b} \qquad \text{(식 4)}$$

[0029] (식 중, B는 붓스트랩핑 과정의 반복횟수를 나타내고, $\hat{\alpha}_{ij}^{*}$ 는 유전자 i에 대해 붓스트랩 표본으로 구한 회귀계수, $\hat{\alpha}_{ij}^{*}$ 는 주어진 관측치 쌍에서 복원랜덤추출한 붓스트랩 표본 $\boldsymbol{z}^{*} = (z_1^{*}, z_2^{*}, \cdots, z_h^{*})$ 을 적합시킨 회귀계수를 나타낸다).

[0030] 본 발명의 일 구현예에 따른 시스템에 있어서, 상기 단백질-단백질 상호작용 정보를 검색하는 알고리즘은 입력

된 마이크로어레이 실험 자료의 유의한 유전자와 학습된 유전자 네트워크로부터 탐색된 작용기작으로 구성된 유전자 또는 단백질의 목록을 또는 작용기작만을 실험/처리 조건 특이적 단백질 네트워크로 검색하거나 유의성을 검사하는 알고리즘일 수 있다.

[0031] 본 발명의 일 구현예에 따른 시스템에 있어서, 상기 활성화 신호/물질대사 전달경로를 검색하는 알고리즘은 입력된 마이크로어레이 실험 자료의 유의한 유전자와 학습된 유전자 네트워크로부터 탐색된 작용기작, 유의성이 검사된 실험/처리 조건 특이적 단백질 네트워크에 포함되는 유전자/단백질 목록 중에서 두가지 이상의 요소로 구성된 유전자 또는 단백질의 목록을 이용하여 활성화 신호/물질대사 전달경로에 검색하거나 실험/처리 조건의 관계성을 판독하는 알고리즘일 수 있다.

[0032] 본 발명은 또한,

[0033] a) 단일 종에서 녹아웃(knockout), 약물(drug), RNAi, 과발현과 같은 다양한 조건의 처리를 수행한 마이크로어레이 실험 자료와 질환이나 질환 상태에서 투여된 약물과 같은 작용기작을 확인하기 위해 측정된 마이크로어레이 실험 자료를 입력하는 단계;

[0034] b) 입력된 마이크로어레이 실험 자료를 이용하여 일반 회귀 모형 또는 붓스트랩핑 회귀 모형을 이용하여 유전자 네트워크를 생성하고 작용기작을 탐색하는 단계;

[0035] c) 탐색된 작용기작을 이용하여 실험/처리 조건 특이적인 단백질 네트워크를 예측하는 단계;

[0036] d) 탐색된 작용기작과 실험/처리 조건 특이적인 단백질 네트워크에 포함되는 유전자/단백질의 목록을 이용하여 활성화 신호/물질대사 전달경로를 검색하는 단계; 및

[0037] e) 검색된 활성화 신호/물질대사 전달경로를 이용하여 실험/처리 조건의 관계성을 판독하는 단계;

[0038] 를 포함하는 출력된 유전자 네트워크, 실험/처리 조건 특이적 단백질 네트워크, 활성화 신호/물질대사 전달경로 정보와 판독된 실험/처리 조건의 관계성을 이용하여 생물학적 통합 네트워크를 구축하는 방법을 제공한다.

[0039] 본 발명의 일 구현예에 따른 방법에서, 상기 c) 단계는 녹아웃(knockout), 약물(drug), RNAi, 과발현과 같은 단일 종의 다양한 조건의 처리를 수행한 마이크로어레이 실험 자료와 질환이나 질환 상태에서 투여된 약물과 같은 작용기작을 확인하기 위해 측정된 마이크로어레이 실험 자료를 이용하여 확인된 유전자 네트워크와 작용기작 정보를 이용하여 단백질 네트워크를 도출할 수 있다.

[0040] 본 발명의 일 구현예에 따른 방법에서, 상기 d) 및 e) 단계는 녹아웃(knockout), 약물(drug), RNAi, 과발현과 같은 단일 종의 다양한 조건의 처리를 수행한 마이크로어레이 실험 자료와 질환이나 질환 상태에서 투여된 약물과 같은 작용기작을 확인하기 위해 측정된 마이크로어레이 실험 자료를 이용하여 확인된 유전자 네트워크와 작용기작 정보 또는 이를 이용하여 도출한 단백질 네트워크의 정보를 이용하여 활성화 신호/물질대사 전달경로를 도출하거나 실험/처리 조건의 관계성을 판독할 수 있다.

[0041] 본 발명은 또한, 본 발명의 생물학적 통합 네트워크를 구축하는 방법을 수행하기 위한 컴퓨터로 판독 가능한 프로그램을 기록한 기록매체를 제공한다. 구체적으로, 마이크로어레이(microarray) 실험 자료를 이용하여 유전자 네트워크(gene network), 실험/처리 조건 특이적 단백질 네트워크(experiment/treatment condition-specific protein network), 실험/처리 조건 관계성 규명을 위한 생물학적 통합 네트워크를 분석하기 위한 방법을 수행하기 위한 컴퓨터로 판독 가능한 프로그램을 기록한 기록매체를 제공한다.

[0042] 컴퓨터로 판독할 수 있는 기록매체란 컴퓨터에 의해 직접 판독되고 엑세스될 수 있는 임의의 기록매체를 말한다. 이러한 기록매체로서는 플로피 디스크, 하드 디스크, 자기 테이프 등의 자기기록매체, CD-ROM, CD-R, CD, RW, DVD-ROM, DVD-RAM, DVD-RW 등의 광학기록매체, RAM이나 ROM 등의 전기 기록매체 및 이들 범주의 혼합물(예: MO 등의 자기/광학기록매체)을 들 수 있지만, 이들에 한정되는 것이 아니다.

[0043] 상기한 기록매체에 기록 또는 입력시키기 위한 기기 또는 기록매체 중의 정보를 판독하기 위한 기기 또는 장치의 선택은 기록매체의 종류와 엑세스 방법에 근거한다. 또한 여러 가지 데이터 프로세서 프로그램, 소프트웨어, 컴퍼레이터 및 포맷이 본 발명의 방법을 수행하기 위한 프로그램을 당해 매체에 기록시키기 위해 사용된다. 당해 정보는 예를 들면, 시판하는 소프트웨어로 포맷된 바이너리 파일(binary file), 텍스트 파일 또는 ASCII 파일의 형태로 나타낼 수 있다.

[0044] 첨부된 도면을 참조하여 본 발명의 바람직한 실시예를 보다 상세히 설명하기로 한다.

[0045] 도 1은 마이크로어레이(microarray) 실험 자료를 이용하여 유전자 네트워크(gene network), 실험/처리 조건 특이적 단백질 네트워크(experiment/treatment condition-specific protein network), 실험/처리 조건 관계성 규명을 위한 생물학적 통합 네트워크를 분석하기 위한 시스템의 개략도를 나타낸다.

[0046] 본 발명의 시스템은 앞서 기술한 입력부; 데이터베이스; 학습부; 탐색부; 검색부; 검사부; 출력부; 판독부를 포함한다.

[0047] 상기 입력부는 질환 또는 약물과 같은 처리를 통한 마이크로어레이 실험 자료를 입력하는 기능을 수행한다. 도 4는 입력부 화면을 나타낸다. 입력 양식에 필수요소인 마이크로어레이 실험 자료의 유전자 발현값을 입력한다.

[0048] 상기 데이터베이스에서 단백질-단백질 상호작용 데이터베이스는 단백질-단백질 상호작용 정보의 실험 정보를 포함하거나 텍스트마이닝(Textmining) 방법을 통한 예측 정보를 포함하고 있다. 신호/물질대사 전달경로 데이터베이스는 문헌정보를 통한 curate 방식의 정보를 포함하고 있다.

[0049] 상기 학습부는 입력된 마이크로어레이 실험 자료로부터 유전자 네트워크를 예측하기 위한 일반 회귀 모형 또는 붓스트랩핑 회귀 모형을 이용한 유전자 네트워크의 모형을 학습하는 기능을 한다.

[0050] 상기 검색부는 상기 구축된 데이터베이스를 검색하는 기능을 한다.

[0051] 상기 탐색부는 학습된 유전자 네트워크로부터 작용기작을 탐색하는 기능을 한다.

[0052] 도 5는 학습부, 탐색부, 검색부에 사용되는 모수를 설정하는 것을 나타낸다. 설정되는 모수에 따라 학습부, 탐색부, 검색부는 앞서 기술한 방법론에 따라 유전자 네트워크, 작용기작, 실험/처리 조건 특이적 단백질 네트워크, 활성화 신호/물질대사 전달 경로를 출력하거나, 실험/처리 조건의 관계성을 판독한다.

[0053] 상기 검사부는 검색부를 통하여 검색된 자료가 통계적으로 유의성이 있는지 검사한다. 검사의 수준은 p값(p-value)에 의존적으로 결정되며, 경험적인 유의수준으로 70%를 사용하거나 p값으로 0.05를 사용한다.

[0054] 상기 판독부는 출력되는 활성화 신호/물질대사 전달경로를 이용하여 실험/처리 조건의 관계성을 판독하는 기능을 한다. 판독은 앞서 출력된 신호/물질대사 전달 경로들 간의 관계성에 기인하며, 이는 도 6에서 표시하는 것과 같이 나타낼 수 있다.

[0055] 통합 출력부는 상기의 모든 출력부에서 나타내는 결과를 하나로 통합하는 기능을 한다. 이는 도 7에서 표시하는 것과 같이 나타낼 수 있다.

[0056] 학습(training) 자료를 생성하기 위하여, 질환이나 질환 상태에 투여된 약물의 작용기작을 파악하고자 하는 생명체 종과 동일한 종의 다양한 조건으로 처리를 한 마이크로어레이 실험 자료를 획득하여 동일한 유전자에 대해 처리된 다양한 조건에 대한 반복된 마이크로어레이 실험의 발현 값을 평균 내어 도 2와 같이 가로로 나열한 형태로 N x M 형태의 자료 행렬을 생성한다. 이때 유전자의 개수를 N, 독립된 마이크로어레이 실험의 수를 M으로 하여 구분한다. 또한 반복된 마이크로어레이 실험에서 표준편차를 계산하여 동일한 형식의 자료 행렬을 생성한다. 단, 평균 발현 값에 대한 자료 행렬과 표준편차에 대한 자료 행렬의 형식은 동일해야 한다.

[0057] 검사(test) 자료를 생성하기 위하여, 질환이나 약물에서의 작용기작을 확인하고자 하는 마이크로어레이 실험의 반복된 실험을 평균 내어 N x M 형태의 자료 행렬을 생성한다. 또한 표준편차를 계산하여 동일한 형태의 자료 행렬을 생성한다. N은 학습 자료와 동일하게 유전자의 수를, M은 마이크로어레이 실험의 수를 나타낸다. 학습 자료와 검사 자료의 유전자의 순서는 동일해야 한다(Xing et al. 2006. Nature Protocol 1(6):2551-4).

[0058] 상기의 자료 행렬을 이용하여 질환이나 약물에서의 작용기작에 대한 영향력을 계산하기 위하여 회귀 모형을 사용한다. 유전자 i에 영향을 주는 요소가 유전자 j라는 벡터라고 가정할 경우의 영향력 함수 f_i 회귀 모형은 식 1과 같다.

$$f_i(y_1,\ldots,y_N,u_i) = u_i\prod_j y_j^{n_{ij}} - d_i y_i \qquad \text{(식 1)}$$

[0059]

[0060] 위의 회귀 모형 식 1의 각 기호는 다음과 같다. 측정된 마이크로어레이 상의 유전자의 수를 N이라 하고, 유전자 i에 집중되는 영향력을 y_i, 유전자 i에 합성률(synthesis rate)에 따른 네트워크상의 외부 영향력을 u_i, 유전자 i에서 유전자 j가 미치는 영향력을 표시한 모수를 n_{ij}, 유전자 i의 발현 감소율(degradation rate)을 d_i라고 할

경우, 전체 유전자 네트워크의 모형을 계산할 수 있다(di Bernardo et al. 2005. Nature Biotechnology 23(3):377-83).

[0061] 위의 식을 마이크로어레이 실험으로부터 획득한 자료 행렬에 적용하고 모수를 예측하기 위하여 steady-state 상태로 가정하면 $f_i(y_1,\ldots,y_N,u_i)$가 0이 된다. 그러면 위의 식 1을 다음의 식 2로 변형할 수 있다.

[0062]
$$\sum_j a_{ij}x_j = -p_i \qquad \text{(식 2)}$$

[0063] 식 2의 각 기호는 다음과 같다.

$$a_{ij} = \begin{cases} n_{ij} & ,j \neq i \\ n_{ij}-1 & ,j = i \end{cases}$$

$$x_j = \log_{10}\left(\frac{y_j}{y_{jb}}\right)$$

$$p_i = \log_{10}\left(\frac{u_j}{u_{jb}}\right)$$

[0064]

[0065] 위의 식 2에서 x_j를 마이크로어레이 실험 관측 값과 동일하게 맞추기 위하여 x_j와 p_i에 로그 변환(logarithm formation)을 수행한다.

[0066] 위의 식 2에서 y_{jb}와 u_{jb}는 입력되는 마이크로어레이 실험 자료가 항상 비교에 의한 값으로 나오기 때문(실험군과 대조군의 비교)에 붙는 분모(denominator)이고, 마이크로어레이 실험 자료에서 대조군(control)의 값이다. 식 2를 이용하기 위해서는 a_{ij}와 p_i의 값을 예측하여야 하는데 마이크로어레이 실험 자료에서 소수의 유전자만 특정 실험 조건에서 반응하는 것이라고 가정하면 p_i가 대부분 0이 됨으로, p_i를 0이라고 가정하고 a_{ij}를 예측한다. a_{ij}가 예측이 되면, 예측한 a_{ij}값과 마이크로어레이 실험 자료 행렬을 사용하여 p_i를 예측한다. 그럼 예측된 p_i를 이용하여 a_{ij}를 다시 예측하고, 다시 p_i를 예측한다. 이 과정을 반복적으로 수행하여 최적의 a_{ij}와 p_i를 찾는다. 유전자 네트워크 학습부는 이러한 과정을 거쳐서 유전자 네트워크를 생성하게 된다. 학습을 위하여 입력된 단일 종에서 다양한 조건을 처리한 마이크로어레이 실험 자료가 용량이 많을 경우 학습의 수행 속도를 위하여 특이값 분해를 수행하여야 한다. 이때 특이값 분해를 통하여 포함되는 고유값(singular value)의 수는 총 실험의 경험적으로 유의한 수준인 70% 선을 유지하도록 해야 한다.

[0067] 위의 식 1과 식 2를 이용하여 작용기작을 탐색하기 위하여 추출된 유전자의 목록을 외부 영향력의 내림차순으로 정렬한다. 내림차순은 z-score나 modified z-score를 기준으로 정렬하는데, z-score는 다음과 같이 계산한다.

[0068]
$$z-score = z_i = \frac{|p_i|}{\sigma_{p_i}}$$
$$modified\ z-score = z_i^m = z-score + \frac{x_i}{\sigma_{x_i}} \qquad \text{(식 3)}$$

[0069] z-score나 modified z-score에 의해 내림차순으로 정렬함으로써 출력되는 유전자는 질환이나 약물과 같은 실험/처리 상태의 마이크로어레이 실험 자료에서 영향력이 큰 작용기작부터 순서대로 나타나는 것이다. 정렬한 z-score나 modified z-score에 의해서 영향력을 많이 발휘하는 유전자를 제외하고 나머지를 목록에서 삭제한다. 이는 실험/처리 조건에만 발현되는 작용기작을 찾기 위함이다. 이러한 결과를 식 2에 다시 적용하여 반복적으로 최적의 결과를 획득할 때까지 반복적으로 수행한다. 반복횟수는 자료에 의존적임으로 최적의 결과가 나올 때까지 사전에 알고 있는 약물 타깃을 이용하여 학습 자료를 시험해서 미리 반복횟수를 결정하고 수행해야 한다.

[0070] 작용기작을 탐색하는 부분은 상기의 일반 회귀 모형을 사용하지 않고 붓스트랩핑 회귀 모형(Bootstrapping Regression Model)을 사용할 수 있다. 붓스트랩핑 회귀 모형을 사용하는 경우에 네트워크의 모형을 생성하는 과정에서 일반 회귀 모형이 아닌 붓스트랩핑 회귀 모형을 사용해야 한다.

[0071] 단계 1; 주어진 관측치 쌍을 $\boldsymbol{z} = (z_1, z_2, \cdots, z_h)$ 으로부터 복원랜덤추출한 확률표본을 붓스트랩 표본 $\boldsymbol{z}^* = (z_1^*, z_2^*, \cdots, z_h^*)$ 이라 할 때, $\boldsymbol{z} = (z_1, z_2, \cdots, z_h)$ 으로부터 복원추출로 하나의 붓스트랩표본 $\boldsymbol{z}^* = (z_1^*, z_2^*, \cdots, z_h^*)$ 을 구성한다.

[0072] 단계 2; 유전자 i에 대해 붓스트랩 표본으로 구한 회귀계수를 $\hat{\alpha}_{ij}^*$, j=1,...N 라 하면, 주어진 관측치 쌍에서 복원랜덤추출한 붓스트랩 표본 $\boldsymbol{z}^* = (z_1^*, z_2^*, \cdots, z_h^*)$ 을 식 2에 적합시키고, 회귀계수 $\hat{\alpha}_{ij}^*$ 를 계산한다. 단계 1과 단계 2의 과정을 독립적으로 B회 반복한다. 유전자 i에 대해 B회의 붓스트랩 반복으로 계산한 회귀계수 $\hat{\alpha}_{ij}^{*b}$, b=1,...,B 의 평균을 계산하여 하기의 식 4와 같이 추정하고자 하는 모형계수 $\hat{\alpha}_{ij}^*$ 를 구한다.

$$\hat{a}_{ij}^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\alpha}_{ij}^{*b} \qquad \text{(식 4)}$$

[0073]

[0074] 실험/처리 조건 특이적 단백질 네트워크를 검색하기 위하여 마이크로어레이 실험 자료로부터 유의한 유전자를 탐색하거나 상기의 작용기작 분석으로부터 추출된 유전자 순위 목록을 단백질-단백질 상호작용 데이터베이스에 맵핑(mapping)하여 단백질-단백질 상호작용 정보를 검색한다.

[0075] 검색된 단백질-단백질 상호작용 정보를 이용하여 Liu 등 (2007, PLoS Genet. 3(6):e96)에서 제안한 GNEA 방법을 응용하여 높은 값의 단백질-단백질 상호작용 네트워크(HSN; High-Scoring protein-protein interaction Network)를 추출한다. 맵핑을 하는 마이크로어레이 실험 자료는 질환이나 약물의 작용기작을 파악하기 위해 도 3과 같이 동일한 형태의 검사 자료를 사용하여 실험/처리 조건 특이적(experiment/treatment condition-specific)인 단백질-단백질 상호작용 정보가 될 수 있도록 해야 한다. 또한 각 유전자가 실험/처리 조건에서 변화된 순위에 의해 맵핑을 수행한다.

[0076] 실험/처리 조건 특이적 단백질-단백질 상호작용 정보가 올바르게 추출된 것인지를 확인하기 위해 필요한 유전자 목록을 마이크로어레이 실험 조건과 관련성이 있는 신호/물질대사 전달경로(signaling/metabolic pathway) 정보로부터 획득한다.

[0077] 상기에서 마이크로어레이 실험 정보와 단백질-단백질 상호작용 정보를 맵핑하여 추출된 단백질-단백질 상호작용 네트워크 정보를 피셔의 정확도 검증(Fisher's Exact Test)을 이용하여 유의성을 검증한다. 피셔의 정확도 검증은 다음의 방법을 사용하여 수행한다.

	HSN 내부 유전자	HSN 외부 유전자
추출된 유전자	a	b
추출되지 않은 유전자	c	d

$$p = \frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{n}{a+c}} = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!} \qquad \text{(식 5)}$$

[0078]

[0079] 질환에서 작용 기작으로 작동하는 것으로 예측된 유전자가 높은 값의 단백질-단백질 상호작용 네트워크의 내부에 포함되면 a;

[0080] 질환에서 작용 기작으로 작동하는 것으로 예측된 유전자가 높은 값의 단백질-단백질 상호작용 네트워크의 외부에 포함되면 b;

[0081] 질환에서 작용 기작으로 작동하지 않는 것으로 예측된 유전자가 높은 값의 단백질-단백질 상호작용 네트워크의 내부에 포함되면 c;

[0082] 질환에서 작용 기작으로 작동하지 않는 것으로 예측된 유전자가 높은 값의 단백질-단백질 상호작용 네트워크의 외부에 포함되면 d;

[0083] 이며 a, b, c, d는 유전자의 개수이다.

[0084] 위의 식 5를 이용하여 추출된 조건 특이적 단백질-단백질 상호작용 정보가 유의하게 추출된 것인지를 확인한다. 유의하게 추출된 조건 특이적 단백질-단백질 상호작용 정보를 높은 값의 단백질-단백질 상호작용 네트워크(HSN; High-Scoring protein-protein interaction Network)라고 한다.

[0085] 실험/처리에 대한 조건 관계성 판명을 위하여 신호/물질대사 전달경로 데이터베이스로부터 유의한 유전자 또는 실험/처리 조건 특이적 네트워크에 포함된 유전자/단백질을 신호/물질대사 전달경로 정보에 맵핑을 한다. 신호/물질대사 전달경로의 연관성을 측정하기 위해 Hu 등(2008, Pac. Symp. Biocomput. 255-66) 방법론을 이용하여 검사(test) 자료의 조건에 관련성이 있는 질환이나 약물 투여 처리된 자료를 학습 데이터로부터 추출하고, 신호/물질대사 전달경로(metabolic/signaling pathway) 정보의 순으로 도 2와 같은 자료 행렬을 다시 생성한다. 일례로 검사 자료가 2형 당뇨(type 2 diabetes)에 해당할 경우, 2형 당뇨와 관련이 있는 인슐린, 비만 등과 같은 관련된 실험의 마이크로어레이 실험 자료를 추출한다. 또한 추출한 마이크로어레이 실험과 관련된 신호/물질대사 전달경로에 대한 유전자의 리스트를 추출한다. 마이크로어레이 데이터를 신호/물질대사 전달경로에서 추출된 유전자 리스트로 여과하여, 도 2의 왼쪽의 경우에는 신호/물질대사 전달경로에서 추출된 유전자 리스트에 포함되는 발현이 유의한 유전자만을 추출하고, 도 2의 오른쪽의 경우에는 신호/물질대사 전달경로에서 추출된 유전자 목록에 포함되는 질환이나 약물의 반응 타깃과 같은 상기의 분석을 통해 획득한 유전자의 목록만을 추출한다.

[0086] 추출한 자료 행렬 A을 이용하여 특이값 분해를 수행하여 eigenarray 행렬 U, 특이값(singular value)으로 구성된 행렬 $\sum$와 eigengene 행렬 V^T로 분해한다.

[0087] $$A = U\Sigma V^T \qquad (식\ 6)$$

[0088] 이때 $\sum$ 행렬에서 k'개의 특이값을 결정해야 하는데, 이 경우도 경험적으로 사용하고 있는 (70/Z)%을 선택한다. Z는 추출된 행렬 A의 실험 수를 뜻한다.

[0089] 상기에 추출된 마이크로어레이 실험에서 신호/물질대사 전달경로의 활성도(pathway activity level)를 결정한다(Tomfohr et al. 2005. BMC Bioinformatics 6:225). 이것은 신호/물질대사 전달경로에 포함되는 유전자가 유의하게 나타난 실험을 결정하는 것이 가능하다. k'개의 유의한 유전자가 각 실험에서 신호/물질대사 전달경로의 활성도(I_j)를 결정하기 위하여 다음의 식 7을 사용한다.

[0090] $$l_j = \sum_{i=0}^{k'} VT_{ij}^2, \qquad VT = V^T \qquad (식\ 7)$$

[0091] j는 마이크로어레이 실험을 나타내며;

[0092] VT는 특이값 분해로부터 계산된 eigengene 행렬 V^T이고;

[0093] k'은 $\sum$ 행렬에서 결정된 특이값의 수이다.

[0094] 특이값 분해에 의해 생성된 신호/물질대사 전달경로의 활성도를 계산하면 마이크로어레이 실험 자료로부터 유의하게 나타난 신호/물질대사 전달경로를 파악한다. 이것을 위하여 신호/물질대사 전달경로의 활성도를 이용하여 SAM(Significance Analysis of Microarrays)를 이용한다(Tusher et al. 2001. Proc Natl. Acad. Sci. USA 98(9):5116-21).

[0095] 유의하게 나타난 신호/물질대사 전달경로의 연결성을 확인하기 위하여, 각 처리 조건의 마이크로어레이 실험의 신호/물질대사 전달경로의 활성도를 Spearman's rank correlation을 이용하여 신호/물질대사 전달경로간의 관련성을 계산한다.

$$\rho = \frac{\sum_{i=1}^{n} R(x_i)R(y_i) - n\left(\frac{n+1}{2}\right)^2}{\left(\sum_{i=1}^{n} R(x_i)^2 - n\left(\frac{n+1}{2}\right)^2\right)^{0.5}\left(\sum_{i=1}^{n} R(y_i)^2 - n\left(\frac{n+1}{2}\right)^2\right)^{0.5}} \qquad \text{(식 8)}$$

[0096]

[0097] 식 8에서 $R(x_i)$와 $R(y_i)$는 Spearman's rank correlation을 계산하고자 하는 마이크로어레이 실험 결과의 신호/물질대사 전달경로의 활성도 값에 대한 오름차순 정렬된 값들이다. 스피어만 순위 관계성(Spearman's rank correlation)을 계산했을 때, 그 값이 긍정적(positive) 관계성이 0.6 이상이거나 부정적(negative) 관계성이 -0.6 이하의 값을 가진다면, 두 가지 마이크로어레이 실험에 처리된 조건은 관계성을 가지는 것으로 판단할 수 있고, 이것은 두 실험에 처리된 신호/물질대사 전달경로사이에 관계가 있다는 것이다. 관계성의 값이 0.6이라는 절대값은 상호 관계성을 조사하는 통계적인 방식의 경험적 근거이다. 경험적 근거는 correlation이 절대값 0.4 이상을 약한 관계성(weak relation), 절대값 0.7 이상을 강한 관계성(strong relation)이라고 한다.

[0098] 스피어만 순위 관계성이 절대값 0.6 이상인 경우, 입력 데이터인 마이크로어레이 실험 자료에서 유의하게 나타나는 신호/물질대사 전달경로이며, 신호/물질대사 전달경로가 서로 연결되어 있거나, 신호/물질대사 전달경로들이 하나 이상의 공통된 요소의 유전자를 포함할 경우, 스피어만 순위 관계성의 값으로 긍정적/부정적 관계성을 가지고 있다고 판단할 수 있다. 이를 이용하여 신호/물질대사 전달경로의 연결성을 설립하고 실험/처리 조건의 관계성을 판독한다.

[0099] 상기의 모든 과정에서 출력되는 유전자 네트워크, 실험/처리 조건 특이적 네트워크, 활성화 신호/물질대사 전달경로, 판독된 실험/처리 조건의 관계성을 통합하여 생물학적 통합 네트워크를 출력한다.

[0100] 이와 같이, 본 발명이 속하는 기술 분야의 당업자는 본 발명이 그 기술적 사상이나 필수적 특징을 변경하지 않고서 다른 구체적인 형태로 실시될 수 있다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적인 것이 아닌 것으로서 이해해야만 한다. 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 등가 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

도면의 간단한 설명

[0101] 도 1은 마이크로어레이(microarray) 실험 자료를 이용하여 유전자 네트워크(gene network), 실험/처리 조건 특이적 단백질 네트워크(experiment/treatment-specific protein network), 실험/처리 조건 관계성 규명을 위한 생물학적 통합 네트워크를 분석하기 위한 시스템의 개략도를 나타낸다.

[0102] 도 2는 입력부에서 처리하는 마이크로어레이 실험의 학습 자료의 형태를 나타낸다.

[0103] 도 3은 입력부에서 처리하는 마이크로어레이 실험의 검사 자료의 형태를 나타낸다.

[0104] 도 4는 입력부 화면을 나타낸다.

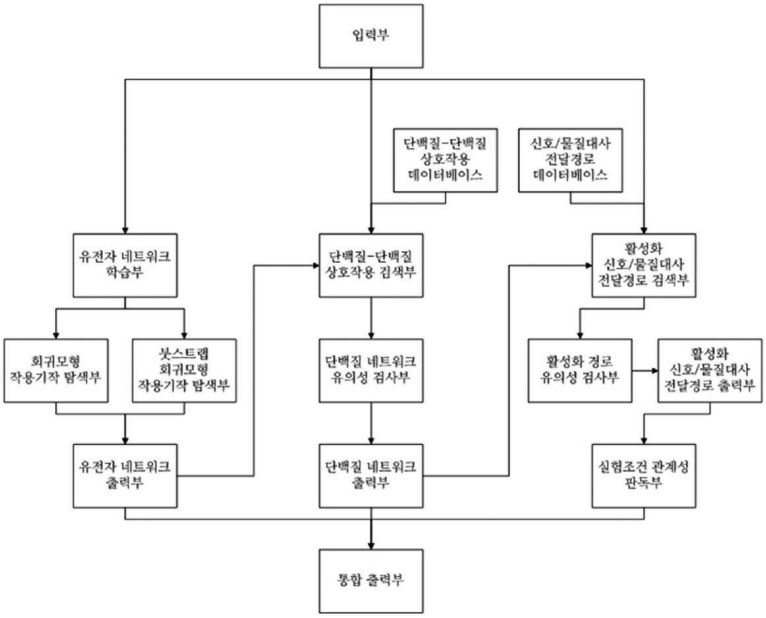
[0105] 도 5는 학습부, 탐색부, 검색부에 사용되는 모수의 형태를 나타낸다.

[0106] 도 6은 실험/처리 조건 관계성 판독부의 활성화 신호/물질대사 전달경로의 관계성을 이용하여 실험/처리 조건의 관계성을 판독된 형태를 나타낸다.

[0107] 도 7은 출력되는 모든 출력부의 결과를 하나로 통합하여 표현되는 생물학적 통합 네트워크의 형태를 나타낸다.

도면

도면1



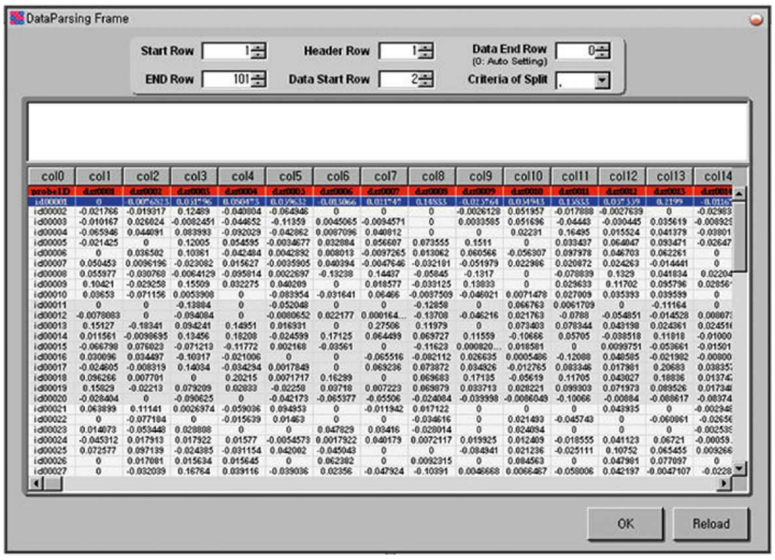
도면2

	$실험_1$	$실험_2$	…	$실험_M$
$유전자_1$				
$유전자_2$				
$유전자_3$				
$유전자_4$	반복된 실험으로부터 획득한 유전자			
…	평균/표준편차 발현 값			
…				
…				
…				
$유전자_N$				

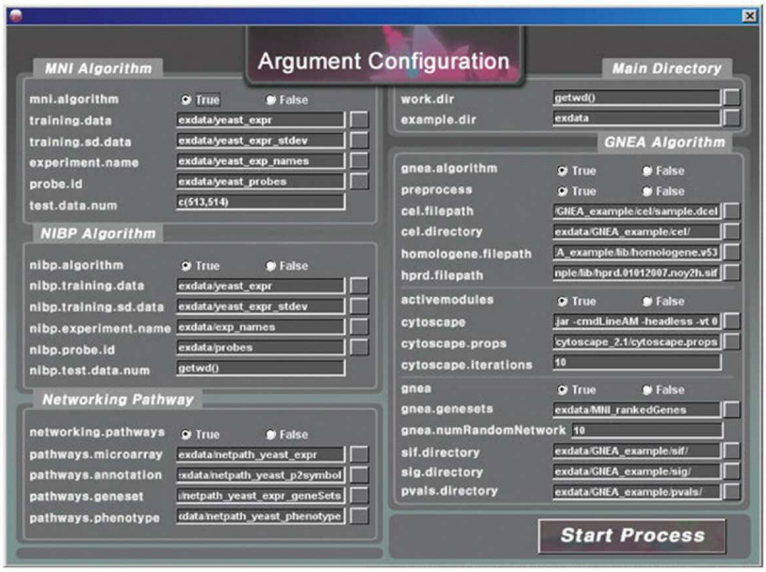
도면3

	당뇨$_1$	인슐린$_2$	…	비만$_Z$
유전자$_1$				
유전자$_2$				
유전자$_3$				
유전자$_4$	마이크로어레이 실험으로부터			
…	획득한 유전자 발현 값			
…				
…				
…				
유전자$_K$				

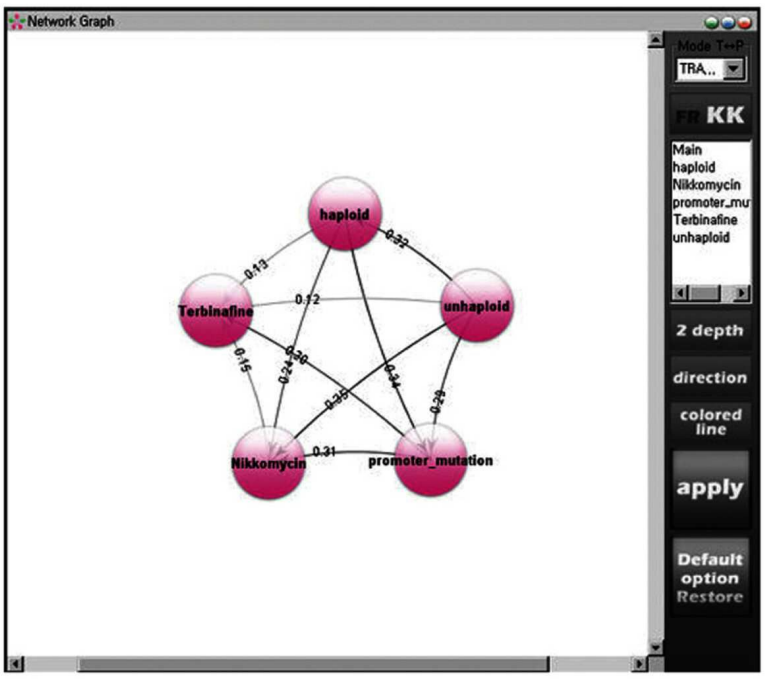
도면4



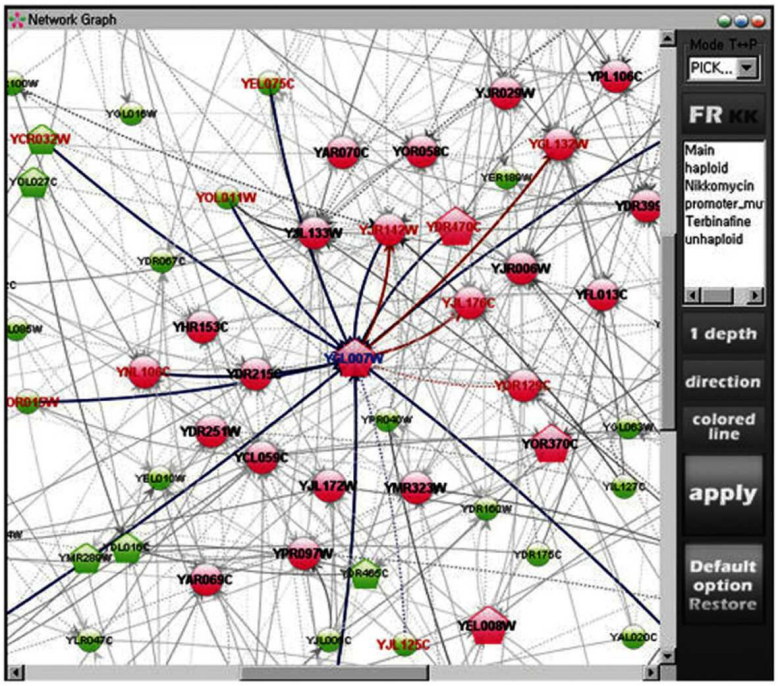
도면5


Argument Configuration
MNI Algorithm
mni.algorithm True False
training.data exdata/yeast_expr
training.sd.data exdata/yeast_expr_stdev
experiment.name exdata/yeast_exp_names
probe.id exdata/yeast_probes
test.data.num c(513,514)
NIBP Algorithm
nibp.algorithm True False
nibp.training.data exdata/yeast_expr
nibp.training.sd.data exdata/yeast_expr_stdev
nibp.experiment.name exdata/exp_names
nibp.probe.id exdata/probes
nibp.test.data.num getwd()
Networking Pathway
networking.pathways True False
pathways.microarray exdata/netpath_yeast_expr
pathways.annotation rxdata/netpath_yeast_p2symbol
pathways.geneset inetpath_yeast_expr_geneSets
pathways.phenotype rdata/netpath_yeast_phenotype
Main Directory
work.dir getwd()
example.dir exdata
GNEA Algorithm
gnea.algorithm True False
preprocess True False
cel.filepath GNEA_example/cel/sample.dcel
cel.directory exdata/GNEA_example/cel/
homologene.filepath A_example/lib/homologene.v53
hprd.filepath nple/lib/hprd.01012007.noy2h.sif
activemodules True False
cytoscape jar -cmdLineAM -headless -vt 0
cytoscape.props cytoscape_2.1/cytoscape.props
cytoscape.iterations 10
gnea True False
gnea.genesets exdata/MIII_rankedGenes
gnea.numRandomNetwork 10
sif.directory exdata/GNEA_example/sif/
sig.directory exdata/GNEA_example/sig/
pvals.directory exdata/GNEA_example/pvals/
Start Process

도면6


Network Graph
haploid
Terbinafine
unhaploid
Nikkomycin
promoter_mutation
0.13
0.32
0.12
0.15
0.24
0.20
0.35
0.34
0.29
0.31
Mode T↔P
TRA..
FR KK
Main
haploid
Nikkomycin
promoter_mu
Terbinafine
unhaploid
2 depth
direction
colored line
apply
Default option Restore

도면7


Network Graph
Mode T↔P
PICK...
FR KK
Main
haploid
Nikkomycin
promoter_mu
Terbinafine
unhaploid
1 depth
direction
colored line
apply
Default option Restore
YEL075C
YJR029W
YPL106C
YCR032W
YAR070C
YOR058C
YGL137W
YOL011W
YJL133W
YJR142W
YDR479C
YJR006W
YFL013C
YJL177C
YHR153C
YGL007W
YDR215C
YDR127C
YDR251W
YOR370C
YCL059C
YJL172W
YMR323W
YPR097W
YAR069C
YEL008W
YJL125C