Network-based ML and graph theory algorithms for Precision oncology

Network-based analytics plays an increasingly important role in precision oncology.

Growing evidence in recent studies suggests that cancer can be better understood through mutated or dysregulated pathways or networks rather than individual mutations and that the efficacy of repositioned drugs can be inferred from disease modules in molecular networks.

This article reviews networkbased machine learning and graph theory algorithms for integrative analysis of personal genomic data and biomedical knowledge bases to identify tumor-specific molecular mechanisms, candidate targets and repositioned drugs for personalized treatment.

The review focuses on the algorithmic design and mathematical formulation of these methods to facilitate applications and implementations of network-based analysis in the practice of precision oncology.

We review the methods applied in three scenarios to integrate genomic data and network models in different analysis pipelines, and we examine three categories of network-based approaches for repositioning drugs in drug–disease–gene networks.

In addition, we perform a comprehensive subnetwork/pathway analysis of mutations in 31 cancer genome projects in the Cancer Genome Atlas and present a detailed case study on ovarian cancer.

Finally, we discuss interesting observations, potential pitfalls and future directions in network-based precision oncology

정밀 종양학(종양에 맞추는 항암제)을 위한 네트워크 기반 머신 러닝과 그래프 이론

(작성자 : 민경준 E-mail : teddyballll@naver.com)

네트워크 기반 분석은 정밀 종양학에서 점점 더 중요한 역할을 합니다.

최근 연구에서 증대하는 증거는 암이 개별 돌연변이보다 돌연변이 또는 조절 장애가 있는 경로나 네트워크를 통해 더 잘 이해될 수 있으며, 재조직 약물의 효능이 분자 네트워크의 질병 모듈에서 추론될 수 있음을 암시합니다.

이 문서에서는 개인의 유전적 데이터 및 생물 의학 지식베이스의 통합 분석을 위해, (네트워크 기반 기계 학습) 및 (그래프 이론 알고리즘)을 검토하여 (종양 특정 분자 메커니즘), (후보 표적(단백질) 및 맞춤 치료를 위한 신약재창출)를 살펴봅니다.

이 문서는 정밀 종양학의 실행에서 네트워크 기반 분석의 응용과 구현을 용이하게 하기위해, 알고리즘 설계 및 이러한 방법의 수학적 공식화에 중점을 둡니다.

우리는 세가지 시나리오에 적용된 방법을 검토하여 서로 다른 분석 파이프 라인에 게놈 데이터와 네트워크 모델을 통합합니다. 그리고 (약물-질병) 발생 네트워크에서 약물을 재배치하기 위해 네트워크 기반 접근법의 세가지 범주를 조사했습니다.

또한 Cancer Genome Atlas의 31개 암 유전체 프로젝트에서 돌연변이에 대한 포괄적인 (하위 네트워크/경로 분석)을 수행하고, 난소암에 대한 자세한 사례 연구를 제시합니다.

마지막으로 네트워크 기반 정밀 종양학에서 흥미로운 관찰, 잠재적 함정 및 향후 방향에 대해 논의합니다.

Introduction

The revolutionary large-scale genomic and sequencing technologies developed in the past two decades have enabled an understanding of cancer biology in individual tumors for personalized treatment.

Coordinated national and international efforts for cancer genome projects have been launched to characterize tens of thousands of individual tumors by somatic mutation, gene expression, copy number variation, DNA methylation, and various other types of genomic and epigenomic aberrations.

The large volume of accumulated cancer genomic data has facilitated the identification of precise oncogenes and tumor suppressors for the development of personalized therapeutic strategies.

One of the well-recognized new observations in these studies is that cancer is better characterized by frequently mutated or dysregulated pathways than driver mutations, which are often distinct in the tumors of the same type.

For example, studies have reported that only a few altered genes occur in more than 10% of the samples and that many other altered genes occur in less than 5% of the samples in the same tumor type.

Furthermore, certain cancer types, such as prostate cancer and pediatric cancers, are not driven by a few somatic mutations or copy number variations, and the mechanism might be better understood in the context of systems biology.

This important observation has led to a great effort to develop a collection of network -based computational methods to detect cancer pathways or subnetworks by integration of various genomic data, as shown in Fig.(1-a) and these methods can be classified into three categories depending on the scenario of applying the analysis pipeline.

Introduction

지난 20년 동안 개발된 획기적인 대규모 유전체 및 시퀀싱 기술(염기서열 해독기술)로 개인 종양의 암 생물학을 개인화된 치료법으로 이해할 수 있었습니다.

체세포 돌연변이, 유전자 발현, 복제 수 변이, DNA 메틸화 및 기타 여러 유형의 유전체 및 후성 유전학적 장애로 인해, 수만 개의 개별 종양을 특성화하기 위해서 암 유전체 프로젝트에 대한 국가적 및 국제적 노력이 협조되었습니다.

축적된 방대한 암 유전체 데이터는 개인화된 치료 전략 개발을 위해서 정확한 종양 유전자 및 종양 억제 인자의 확인을 용이하게 합니다.

이 연구들에서 잘 알려진 새로운 관찰들 중 하나는, 암은 유발 돌연변이보다 (빈번하게 특징지어지거나, 종종 동일한 유형의 종양에서 전혀 다르면서 조절되지 않는 경로로) 구성지어진다는 것이다.

예를 들어, 연구에 따르면 표본의 10% 이상에서만 약간의 변형된 유전자가 발생한다고 보고되었고, 많이 변형된 유전자는 동일한 종양 유형의 표본 중 5% 미만으로 발생했습니다.

또한 전립선암 및 소아암과 같은 특정 암 유형은 몇 가지 체세포 돌연변이나 복제 수 변이에 의해 유도되지 않으며, 메커니즘은 시스템 생물학의 맥락에서 더 잘 이해될 수 있습니다.

이 중요한 관찰은 <그림 (1-a)>에서와 같이 **다양한 유전체 데이터의 통합을 통해 암 경로 또는 서브 네트워크를 탐지하기위한 네트워크 기반 계산 방법 모음을 개발하기 위해 많은 노력을 기울였으며** 이러한 방법은 세 가지 카테고리 분석 파이프 라인을 적용하는 시나리오에 따라 다릅니다.

Network-based analysis has also attracted considerable attention in drug repositioning to reduce the cost of new drug development by using repositioned existing drugs on novel targets in drug–target networks for precision oncology.

네트워크 기반 분석은 정밀 종양학을 위한 (약물 – 표적(단백질)) 네트워크의 새로운 표적(단백질)에 재배치된 기존 약물을 사용함으로써, 신약 개발 비용을 줄이기 위해 신약재창출에 대한 상당한 관심을 보였다.

Based on the hypothesis that drugs tend to be more effective on target genes within or in the vicinity of a disease module in a molecular network, several network-based approaches have been used to explore networks of drugs, diseases and targets to reposition drugs for new targets, as listed in Fig.(1-b)

약물이 분자 네트워크의 질병 모듈 내부 또는 근처의 표적(단백질) 유전자에 더 효과적이라는 가설에 기초하여, <그림(1-b)>에 열거 된 바와 같이, 새로운 목표물에 대한 약물의 위치를 재조정하기위한 약물, 질병 및 표적(단백질)의 네트워크를 탐색하기 위해 몇몇 네트워크 기반 접근법이 사용되어 왔다.

In these methods, the drug–target relations can be inferred by various measures in the network, combining drug–drug, drug–target, drug–disease and disease–gene relations as shown in the drug–disease–target network in Fig.(1-d,e)

이 방법들에서, (약물 – 표적(단백질)) 관계는 <그림(1-d,e)>의 (약물 – 표적(단백질)) 네트워크에 나타난 바와 같이 (약물 – 약물), (약물 – 표적(단백질)), (약물 – 질병) 및 (질병 – 유전자) 관계를 결합하는 다양한 방법에 의해 추론 될 수 있다.

As summarized in Fig.(1-b) these methods can be classified into three categories based on the underlying computational formulation: methods using graph connectivity measures, link prediction methods and network-based classification methods.

<그림 (1-b)>에서 요약된 것처럼, 이 방법은 근본적인 계산 공식을 기반으로 세 가지 범주로 나눌 수 있습니다. 그래프 연결성 측정 방법, 링크 예측 방법 및 네트워크 기반 분류 방법을 사용하는 방법입니다.

The focus of this review article is to provide a comprehensive and unified survey of machine learning and graph theory algorithms for network analysis in precision oncology.

이 문서의 초점은 정밀 종양학에서 네트워크 분석을 위한 기계 학습 및 그래프 이론 알고리즘에 대한 포괄적이고 통합된 조사를 제공하는 것입니다.

We compare the methods by their distinctions in the methodology and mathematical formulations such that the methods can be better applied and improved appropriately for precision oncology.

우리는 이 방법을 정밀 종양학에 적절히 적용하고 향상시킬 수 있도록, 이 방법과 수학 공식에서의 두 방법을 비교합니다.

An overview of this article is given in Fig.(1)

이 기사의 개요는 <그림(1)>을 기반으로 합니다.

We not only review the resources of biomedical and molecular networks listed in Fig.(1-g) and the network-based methods listed in Fig.(1-a,b) but also present a comprehensive network-based pathway analysis of mutations in 31 cancer genome projects in the Cancer Genome Atlas (TCGA) list in Fig.(1-h) and a case study on ovarian cancer to show the promise of applying network-based analysis.

우리는 <그림 (1-g)>에 열거된 생물 의학 및 분자 네트워크 및 <그림 (1-a, b)>에 열거된 네트워크 기반 방법의 자원을 검토할 뿐만 아니라, <그림(1-h)>의 Cancer Genome Atlas (TCGA) 목록과 난소 암에 대한 사례 연구에서 31 가지 암 유전체 프로젝트의 돌연변이에 대한 포괄적인 네트워크 기반 경로 분석을 제시하여 네트워크 기반 분석 적용 가능성을 보여줍니다.

Fig 1

Overview of the methods for network-based precision oncology.

(a) The methods for integration of patient genomic data and molecular networks grouped under the three scenarios of data analysis pipelines.

(b) The methods for integration of drug–drug similarities, drug–target relations and target–target relations for drug repositioning, grouped under three algorithmic categories.

(c) Patient genomic profiles describe the genomic landscape of each patient sample.

(d) The patient genomic profiles are integrated with a molecular network, the human protein–protein interaction (PPI) network in the example.

(e) Drug and disease phenotypes are modeled in a network with connections to the target genes in the PPI network.

(f) An example of cancer subnetworks associated with recurrent ovarian cancer.

(36g) Resources of biomedical and molecular networks.

(h) List of the TCGA cancer studies

Fig 1

네트워크 기반 정밀 종양학을 위한 방법 개요.

(a) 환자 유전체 데이터와 분자 네트워크의 통합을 위한 방법은 데이터 분석 파이프 라인의 세 가지 시나리오에 따라 분류됩니다.

(b) (약물 – 약물) 유사성, (약물 – 표적(단백질)) 관계 및 약물 대체를 위한 (표적(단백질) – 표적(단백질)) 관계의 통합 방법은 3 가지 알고리즘 범주로 분류된다.

(c) 환자 유전체 프로필은 각 환자 표본의 유전체 풍경을 설명합니다.

(d) 환자 유전체 프로필은 이 예에서 분자 네트워크, 즉 인간 (단백질 – 단백질) 상호 작용 (PPI) 네트워크와 통합되어 있습니다.

(e) 약물 및 질병의 형질은 PPI 네트워크에서 표적(단백질) 유전자에 연결 되어있는 네트워크에서 모델링 됩니다.

(f) 재발 성 난소 암과 관련된 암 서브 네트워크의 예.

(g) 생물 의학 및 분자 네트워크의 자원.

(h) TCGA 암 연구 목록

BIOMEDICAL AND MOLECULAR NETWORKS

In the literature, various biological and biomedical network databases have been compiled to support network analysis.

Typically, the databases have been curated by the integration of high-throughput experimental screening results from studies in the literature and possibly computational predictions supervised by expert knowledge.

The networks represent the collections of molecules, phenotypes and drugs as nodes and their relations as edges in graphs.

In Table 1, we enumerate existing molecular networks, phenotype similarity networks or ontologies, and drug–target networks and the resources for obtaining these networks.

The properties of these networks, including their nodes, edges and graph structures, are also shown in Table

생체 분자 네트워크

문헌에서 네트워크 분석을 지원하기 위해 다양한 생물학 및 생물 의학 네트워크 데이터베이스가 컴파일 되었습니다.

일반적으로 데이터베이스는 문헌 연구 및 가능한 전문적인 지식에 의해 감독된 계산 예측과 같은 높은 처리량의 실험적 스크리닝 결과의 통합에 의해 설계되었습니다.

네트워크는 분자와 형질, 그리고 약물의 집합을 노드로 표현하고 그 관계를 그래프로 나타냅니다.

<표 1>에서 기존의 분자 네트워크, 표현형 유사성 네트워크 또는 종양학, 약물 대상 네트워크 및 이러한 네트워크를 얻기 위한 리소스를 나열합니다.

노드, 에지 및 그래프 구조를 포함하여 이러한 네트워크의 속성도 표에 표시됩니다.

1. Molecular networks:

Biological molecular networks describe relations among molecules, such as protein–protein interactions, gene co-expression, functional similarities, regulatory relations or biochemical reactions.

The new-generation high- throughput technologies have provided extensive content to construct such molecular networks.

Protein–protein interaction networks are available from several well-maintained databases.

Primarily, these networks include physical interactions determined by experiments and computationally derived interactions.

Proteome-wide protein–protein interactions capture the interplay among proteins based on the functional associates from co-membership of protein complexes and pathways.

A functional linkage network is a more comprehensive compilation of functional relations, physical interactions and co-expression in one network.

A transcriptional regulatory network models the molecular interactions between transcript factors/microRNA and target genes to regulate transcript expression.

A transcriptional regulatory network is a directed graph in which the edges connect a regulator to its targets.

A cellular metabolic network can be constructed by the co-membership of biochemical reactions among metabolites and enzymes.

Several graph structures can be used to represent metabolic pathways, e.g., labeled directed graphs, unions of bipartite graphs (per reaction) and hypergraphs, depending on the level of detail of metabolic reactions to be modeled with the graph.

1. 분자 네트워크 :

생물학적 분자 네트워크는 (단백질 – 단백질) 상호 작용, 유전자 동시 발현, 기능적 유사성, 규제 관계 또는 생화학 반응과 같은 분자 간의 관계를 설명합니다.

차세대 고효율 기술은 이러한 분자 네트워크를 구축할 수 있는 광범위한 컨텐츠를 제공합니다.

(단백질 – 단백질) 상호 작용 네트워크는 잘 관리된 여러 데이터베이스로부터 이용 가능하다.

주로 이러한 네트워크에는 실험 및 계산으로 유도된 상호 작용에 의해 결정된 물리적 상호 작용이 포함됩니다.

모든 형질의 (단백질 – 단백질) 상호 작용은 단백질 복합체와 경로의 공동 구성원으로부터의 기능적 결합체에 기초한 단백질 간의 상호 작용을 포착한다

기능적 연계 네트워크는 기능적 관계와 물리적 상호 작용 및 하나의 네트워크에서의 공동 표현을, 보다 포괄적으로 편집한 것입니다.

복사 조절 네트워크는 (복사 인자/마이크로 RNA)와 표적(단백질) 유전자 사이의 분자 상호 작용을 모델화하여 복사 발현을 조절합니다.

복사 조절 네트워크는 에지가 조절기를 표적(단백질)과 연결시키는 방향성 그래프이다.

세포 대사 네트워크는 대사 산물과 효소 간의 생화학 반응의 공동 구성원에 의해 구축될 수 있습니다.

그래프로 모델링 되는 대사 반응의 세부 수준에 따라 여러 개의 그래프 구조를 사용하여 대사 경로를 나타낼 수 있습니다

(예 : 방향성 그래프, 2 회성 그래프의 조합, 하이퍼 그래프).

Fig 2

Three scenarios for the integration of genomic data with molecular networks.

(a) Model-based integration formulates one unified learning framework regularized by a graph Laplacian.

The output of the model is network modules enriched by the selected genomic features and a prediction of treatment outcome/cancer phenotype.

(b) Preprocessing integration consists of the following two steps:

The first step detects subnetworks that differentiate the contrasted patient groups by the genomic features; in the second step, the subnetwork features are then fed into a standard learning model to generate predictions.

(c) Post-analysis integration of oncogenic alterations in the network also consists of two steps.

The oncogenic alterations are first detected across the patient profiles, and then the altered genes/loci are mapped to the network as seed genes for the module analysis.

For each scenario, the objectives of the approach, the inputs and outputs of the network- based analysis models/methods, and the advantages/limitations of each approach are also provided

Fig 2

유전체 데이터를 분자 네트워크와 통합하기위한 세 가지 시나리오.

(a) 모델 기반 통합은 Laplacian 그래프에 의해 정형화된 하나의 통합 학습 틀을 공식화한다.

모델의 아웃풋은 (선택된 유전체 특징과 (치료 결과/암 표현형)의 예측에 의해, 강화된 네트워크 모듈)입니다.

(b) 전처리 통합은 다음의 두 단계로 구성됩니다 :

첫 번째 단계는 유전체 기능에 의해 대조된 환자 그룹을 구별하는 서브 네트워크를 감지합니다. 두 번째 단계에서는 서브 네트워크 기능을 표준 학습 모델에 제공하여 예측을 생성합니다.

(c) 네트워크에서 발암성 변화의 (사후 분석 통합)은 두 단계로 구성된다.

발암성 변화는 환자 프로파일 전체에서 처음으로 검출되며, (변경된 유전자 / 유전자좌)는 모듈 분석을 위한 종자 유전자로서 네트워크에 매핑된다.

각 시나리오에 대해 접근법의 목적, 네트워크 기반 분석 모델 / 방법의 입력 및 출력, 각 접근법의 장점 / 한계가 제공됩니다

2. Phenotype similarity networks and ontologies:

Phenotypes, particularly disease phenotypes, are of special interest for cancer studies.

The analysis of diseases in the context of other related diseases can offer insight into their genotypic drivers.

Online Mendelian Inheritance in Man (OMIM) is a comprehensive compendium of human genes, genetic phenotypes and documentation of their phenotype–gene associations.

Phenotype similarity networks can be constructed based on the genetic resemblance or the synopsis of the diseases and sometimes by mRNA expression.

Human Phenotype Ontology (HPO) is another more comprehensive organization of all human disease phenotypes in an ontology.

The ontology is a directed acyclic graph that can be used as a network structure for learning phenotype–gene associations.

2. 형질 유사성 네트워크 및 종양학:

형질, 특히 질병의 형질(생물의 형태와 성질(형질))은 암 연구에 특히 중요합니다.

다른 관련된 질병들의 맥락에서 **질병 분석은 그들의 유전적 유발자에 대한 정보를 제공할 수 있습니다.**

온라인 멘델 상속 (OMIM)은 인간 유전자, 유전자 형질과 그들의 형질 유전자 연관성에 대한 문서를 포괄적으로 요약한 것입니다.

형질 유사성 네트워크는 (유전적 유사성 또는 질병의 개요와 때때로 mRNA 발현)에 기초하여 구축될 수 있다.

Human Phenotype Ontology (HPO)는 종양학에서 모든 인간 질병 형질의 조직 중 가장 종합적인 조직입니다.

종양학은 (표현형(생물의 형태와 성질(형질)) - 유전자의 연관성)을 학습하기위한 네트워크 구조로 사용될 수 있는 방향성 있는 비 순환 그래프이다.

3. Drug–target and drug–drug networks:

Drug–target associations can be modeled by a bipartite network with connections between the drugs and their targets.

The drug–target pairs are typically derived from FDA-approved or experimental drugs and their human protein targets available from various drug databases.

Several different types of drug–drug similarity networks have been derived for drug repositioning.

Drug–drug relations can be inferred based on similarity of molecular basis, chemical substructure, and phenotypes, such as known drug-indication relations, co-membership in drug combinations, and co-morbidity of diseases.

NETWORK-BASED ANALYSIS OF PERSONAL GENOMIC PROFILES

The goal of applying network-based analysis to personal genomic profiles is to identify aberrant network modules that are both informative of cancer mechanisms and predictive of cancer phenotypes.

These methods can be classified into three categories based on the design of the analysis pipeline in different scenarios, as shown in Fig.(2)

In these scenarios, the detection of the network modules facilitates two other goals: predicting cancer phenotypes and detecting driver genes.

Depending on how the network information is processed in the pipeline, the inputs and the outputs to the predictive models or network analysis methods can differ.

Below, we describe the three categories of the methods listed in Fig. (1-a) and then discuss the advantages and limitations of each of the categories.

3. (약물 - 표적(단백질))과 (약물 – 약물) 네트워크 :

(약물 – 표적(단백질)) 연합은 약물과 표적(단백질) 사이를 연결하는 양자 간의 네트워크 (bipartite network)에 의해 모델링 될 수 있다.

(약물 - 표적(단백질)) 쌍은 일반적으로 FDA승인 약물 또는 실험 약물과 다양한 약물 데이터베이스에서 사용할 수 있는, 인간 단백질 표적(단백질)에서 파생된다.

신약재창출을를 위해 몇 가지 다른 유형의 (약물 – 약물) 유사성 네트워크가 유도되었습니다.

(약물 – 약물) 관계는 분자 기반의 유사성을 기반으로 추론될 수 있지만, 화학 물질 하부 구조나, 알려진 (약물 중독 관계), 약물들의 공동 구성요소, 그리고 여러 질병을 동시에 겪는 현상과 같은 형질을 포함합니다.

개인 유전체 프로필의 네트워크 기반 분석

개인 유전체 프로필에 네트워크 기반 분석을 적용하는 것의 목표는 암의 구조와 암의 형질을 예측하는 비정상적인 네트워크 모듈을 확인하는 것입니다.

이러한 방법은 <그림 2>에서와 같이 다양한 시나리오에서 분석 파이프 라인의 설계를 기반으로 세 가지 범주로 분류할 수 있습니다.

이러한 시나리오에서 **네트워크 모듈을 감지하면 암 표현형을 예측하고 암 유발 유전자를 탐지하는 두 가지 목표가 수월 해집니다.**

네트워크 정보가 파이프 라인에서 처리되는 방법에 따라 예측 모델 또는 네트워크 분석 방법에 대한 입력 및 출력이 다를 수 있습니다.

아래에서는 그림 (1-a)에 나열된 세 가지 범주의 방법을 설명하고 각 범주의 장점과 한계를 논의합니다.

Model-based integration of whole-genomic profiles and a network Model-based integration formulates a single unified machine learning framework to integrate genomic profiles with a network as illustrated in Fig. (2-a).

전체 유전체 프로필과 네트워크 기반 모델 통합은 그림 (2-a)에서 설명한대로 유전체 프로필을 네트워크와 통합하는 단일 통합 기계 학습 프레임 워크를 공식화합니다.

The core technique is to introduce a network-based regularization into machine learning models such that the coefficients learned on the feature variables form dense subnetworks.

핵심 기술은 특징 변수에서 학습된 계수가 조밀한 서브 네트워크를 형성하도록 기계 학습 모델에 네트워크 기반 정규화를 도입하는 것입니다.

The most commonly used network-based regularization is the graph Laplacian regularizer shown in Fig. (3-a).

가장 일반적으로 사용되는 네트워크 기반 정규화는 그림 (3-a)에 표시된 Laplacian 그래프 정규화기입니다.

The graph Laplacian was first introduced for spectral graph analysis and then used for semi-supervised learning in machine learning.

Laplacian 그래프는 스펙트럼 그래프 분석을 위해 처음 도입된 다음 기계 학습에서 준 지도 학습에 사용되었습니다.

The graph Laplacian regularization is a summation of smoothness terms on the variables to encourage similar coefficients on the genes or other genomic features that are connected in the network.

Laplacian 정규화 그래프(Laplacian regularization)는 변수들에 대한 평활도 (smoothness) 항의 합계로서, 네트워크에서 연결된 유전자 또는 다른 유전체 특징에 대한 유사한 계수를 권장한다.

Below, we describe the graph Laplacian regularized methods in different learning frameworks as shown in Fig.(3-b,e).

아래에서 우리는 그림 (3-b, e)와 같이 서로 다른 학습 프레임 워크에서 Laplacian 정규화 된 방법을 설명합니다.

To precisely describe the models, we also list all the necessary notations in Table 2 and the exact mathematical formulations of the methods in Supplementary Table S2.

모델을 정확하게 설명하기 위해 <표 2>에 필요한 모든 표기법과 <보충 표 S2>의 방법에 대한 정확한 수학 공식을 나열합니다.

Fig 3

Model-based integration of whole-genomic profiles and a molecular network.

(a) The patient genomic profiles X along with the clinical information: the survival time, two patient subgroups for classification and treatment response of each individual patient are shown.

The network S is typically integrated into the genomic profile analysis with a graph Laplacian regularization.

The formulas of the graph Laplacian and its regularization are shown below.

The graph Laplacian regularization can be rewritten as summation of pairwise smoothness terms that promote smoothness among the connected genomic features in the network.

(b) The network-based linear regression and Cox regression models are illustrated in the figure with the graph Laplacian regularization term added to the original cost functions.

(c) Network-based classification is illustrated by a network-based SVM to classify the samples.

(d) Network-based semi-supervised learning models classify samples and detect disease markers on a bipartite graph.

The edges between samples and genomic features are weighted by the genomic profiles, and semi-supervised learning is based on the bipartite graph Laplacian.

(e) Network-based factorization models factorize the genomic profile X into the product of two matrices, U and H, which cluster patient samples and learn the latent features in the genomic profiles

Fig 3

전체 유전체 프로필과 분자 네트워크의 모델 기반 통합.

(a) 임상 정보와 함께 환자 유전체 프로필 :

생존 시간, 각 환자의 분류 및 치료 반응에 대한 두 개의 환자 하위 그룹이 표시됩니다.

네트워크 S는 전형적으로 Laplacian 그래프 정규화와 함께 유전체 프로파일 분석에 통합된다.

그래프의 공식 Laplacian과 정규화는 아래와 같습니다.

Laplacian 정규화 그래프는 네트워크의 연결된 유전체 특징 간의 매끄러움을 촉진하는 쌍 방향 평탄 조건의 합계로 재작성될 수 있습니다.

(b) 네트워크 기반 선형 회귀 모델과 Cox 회귀 모델은 원래의 비용 함수에 Laplacian 정규화 용어가 추가된 그래프와 함께 그림으로 설명됩니다.

(c) 네트워크 기반 분류는 샘플을 분류하기 위해 네트워크 기반 SVM에 의해 설명된다.

(d) 네트워크 기반의 준 지도 학습 모델은 샘플을 분류하고 이분 그래프에서 질병 마커를 탐지합니다.

샘플과 유전체 특징 사이의 에지는 유전체 프로필에 의해 가중치가 적용되며 준 지도 학습은 이분 Laplacian 그래프를 기반으로 합니다.

(e) 네트워크 기반 인수 분해 모델은 유전체 프로파일 X를 환자 표본을 모으고 유전체 프로파일의 잠재 특징을 학습하는 두 개의 행렬 U와 H의 곱으로 분해한다

In Fig.(3-b), the widely used regression and survival models are extended to include the graph Laplacian constraint for the analysis of genomic data.

그림 (3-b)에서, 널리 사용되는 회귀 및 생존 모델은 유전체 데이터의 분석을 위한 Laplacian 그래프 제약을 포함하도록 확장된다.

The paper proposed a network constrained linear regression procedure that combines a graph Laplacian constraint with the L1-norm sparse linear regression to capture the relations among the regression coefficients.

이 논문에서는 Laplacian 그래프 제약과 L1- 표준 희박 선형 회귀를 결합하여 회귀 계수 간의 관계를 포착하는 네트워크 제한 선형 회귀 절차를 제안했습니다.

This network-based linear regression is equivalent to a standard LASSO optimization problem.

이 네트워크 기반 선형 회귀는 표준 LASSO 최적화 문제와 동일합니다.

The paper proposed a network-based Cox proportional hazards model (Net-Cox) for survival analysis.

이 신문은 생존 분석을 위해 네트워크 기반 Cox 비례 위험 모델 (Net-Cox)을 제안했습니다.

In Cox regression, the objective is to learn the regression coefficients β and the baseline hazard function h0(t) such that the instantaneous risk of an event at time t for a patient xi can be estimated by h tð Þ¼ jxi h0ðtÞexp xT i β .

Cox 회귀 분석에서, 목표는 환자 xi에 대한 시간 t에서의 사건의 순간 위험이 다음과 같이 추정 될 수 있도록 회귀 계수 β와 기본 위험 함수 h0 (t)를 학습하는 것입니다.

Similarly, the graph Laplacian constraint is introduced on the regression coefficients β.

마찬가지로 Laplacian 그래프 구속 조건이 회귀 계수 β에 도입됩니다.

By alternating between maximization with respect to β and h0(t), a local optimum can be found.

β에 대한 최대화와 h0(t)를 번갈아 가며 국부 최적이 발견 될 수 있다.

As shown in Fig.(3-c), the graph Laplacian constraint can also be introduced into linear classification models such as logistic regression and support vector machines (SVMs).

그림 (3-c)에서 볼 수 있듯이 Laplacian 그래프 구속 조건은 로지스틱 회귀 및 SVM (Support Vector Machine)과 같은 선형 분류 모델에도 도입될 수 있습니다.

Given the binary response vector y = (y1, ..., yn) T with yi ∈ {1, 0}, a Bernoulli likelihood function minus both the L1-norm and the graph Laplacian constraints is maximized to learn the linear coefficients.

yi ∈ {1, 0} 인 이진 응답 벡터 y = (y1, ..., yn) T가 주어지면, 선형 계수를 배우기 위해 베르누이 가능성 함수와 L1-표준 및 Laplacian 그래프 제약을 모두 극대화한다.

In the model, pð Þ¼ xi exp β0þxT i ð Þβ 1þexp β0þxT i ð Þβ is the probability that the ith sample is in class 1.

이 모형에서 i 번째 표본이 1등급에 있을 확률은 $p(\boldsymbol{x}_i) = \frac{\exp(\beta_0+\boldsymbol{x}_i^T\beta)}{1+\exp(\beta_0+\boldsymbol{x}_i^T\beta)}$ 이다.

The elastic-net procedure can be applied to maximize the regularized cost function.

elastic-net 절차는 정규화 된 비용 함수를 최대화하기 위해 적용될 수 있습니다.

The paper proposed a network-based SVM. Given the +1/−1 binary response vector y, the network-constrained SVM can be formulated as the addition of the hinge loss Pn i¼1 1 yi β0 þ xT i β

이 논문은 네트워크 기반의 SVM을 제안했다. + 1 / -1 바이너리 응답 벡터 y가 주어지면, 네트워크 제약 SVM은 경첩 손실과

$$\sum_{i=1}^{n} \left[1 - y_i\left(\beta_0 + \boldsymbol{x}_i^T\beta\right)\right]_+$$

þ and the graph Laplacian constraint, where the subscript "+" denotes the positive part, i.e., z+ = max{z, 0}.

Laplacian 그래프 제약 조건의 추가로 공식화될 수 있습니다. 여기서 첨자 "+"는 양수 부분을 나타내며, 즉 z + = max { z, 0}.

Semi-supervised learning methods can more conveniently explore the structures among both the genomic features and the patient samples by learning with the graph Laplacians, as shown in Fig.(3-d).

준 지도 학습 방법은 그림 (3-d)와 같이 Laplacian 그래프로 학습함으로써 유전체 특징과 환자 샘플 사이의 구조를 더 편리하게 탐색할 수 있습니다.

In the bipartite graph formulation introduced in the paper, gene expression data are represented as a bipartite graph with weighted edges between patient samples and genomic features.

논문에 소개된 이분 그래프 작성에서 유전자 발현 데이터는 환자 샘플과 유전체 특징 사이에 가중치가 있는 이분 그래프로 표시됩니다.

The bipartite graph captures the co-expression among the genes and the samples as bi-clusters in the graph such that both the sample clusters and feature modules are explored.

이분 그래프는 군집 표본과 기능 모듈 모두를 탐색할 수 있도록 그래프에서 이중 군집으로 유전자와 샘플 간의 동시 표현을 담아냅니다.

In the hypergraph formulation introduced in the papers, the gene expression data are represented as weighted hyperedges on the patient nodes, and a graph Laplacian on the hypergraph can be introduced for semi-supervised learning on the patient samples.

논문에 소개된 하이퍼 그래프 공식에서 유전자 발현 데이터는 환자 노드에서 가중치가 있는 하이퍼 에지로 표시되고 하이퍼 그래프의 Laplacian 그래프는 환자 샘플에서 준 지도 학습을 위해 도입 될 수 있습니다.

An additional graph Laplacian of a protein–protein interaction (PPI) network is then introduced to incorporate network information among the genomic features.

유전체 기능 중 네트워크 정보를 통합하기 위해 (단백질 - 단백질) 상호 작용 (PPI) 네트워크의 추가 그래프인 Laplacian이 소개됩니다.

It is also possible to regularize non-negative matrix factorization (NMF) models with a graph Laplacian, as shown in Fig. (3-e).

그림 (3-e)에서 볼 수 있듯이 비 음수 행렬 인수 분해 (NMF) 모델을 Laplacian 그래프로 정규화 하는 것도 가능하다.

NMF aims to find two non-negative matrices Um × k and Hn × k whose product can accurately approximate the data matrix X with X ≈ UHT.

NMF는 두 개의 음수가 아닌 행렬 Um × k와 Hn × k를 찾고, 그 행렬의 곱은 X ≈ UHT로 데이터 행렬 X에 정확하게 근사 할 수 있습니다.

Combining the geometrically-based constraint with the original NMF leads to the graph-regularized NMF, where Tr(·) denotes the trace of a matrix.

기하학적 기반 구속 조건을 원본 NMF와 결합하면 그래프 정규화된 NMF가 생성되며 여기서 Tr (·)는 행렬의 흔적을 나타냅니다.

Preprocessing integration to detect network-based features

The preprocessing integration methods comprise two steps, as illustrated in Fig.(2-b).

First, the genomic profiles and the network are processed together to generate network-based features; second, standard learning models are applied with the network-based features for predictions.

In this scenario, the integration of network and genomic data occurs before applying a learning model.

The paper first proposed a graph algorithm to detect discriminative subnetworks for classification of patient samples.

Highly discriminative genes are used as seed genes in a greedy search in a PPI network to find discriminative subnetworks, and then gene expression in each subnetwork is normalized as one feature value for classification with standard logistic regression.

A similar approach was later proposed for application with features of discriminative pathways instead of subnetworks.

In this approach, the gene expression in a pathway is normalized as one feature for the collection of pathways from a molecular signature database.

The paper used disease-specific subnetworks as features, where a set of known disease genes are first mapped into the PPI network and then the subnetworks of the disease genes are identified as disease module features.

The paper proposed implementing label propagation on the mutation data of each patient on a PPI network to generate network-smoothed features for classification of the patients.

The paper proposed to find a small subnetwork to connect all differentially expressed genes in a PPI network and then use the genes in the subnetwork as features to classify patient samples.

네트워크 기반 기능을 탐지하기위한 전처리 통합

전처리 통합 방법은 그림 (2-b)와 같이 두 단계로 구성됩니다.

첫째, 유전체 프로파일과 네트워크가 함께 처리되어 네트워크 기반 기능을 생성합니다. 둘째, 표준 학습 모델이 예측을 위한 네트워크 기반 기능과 함께 적용됩니다.

이 시나리오에서는 학습 모델을 적용하기 전에 네트워크 및 유전체 데이터의 통합이 발생합니다.

이 논문에서는 먼저 환자 표본 분류를 위한 차별적인 하위 네트워크를 탐지하는 그래프 알고리즘을 제안했다.

매우 차별적인 유전자는 PPI 네트워크에서 그리디(탐욕) 검색에서 종자 유전자로 사용되어 차별적인 하위 네트워크를 찾고, 각 하위 망의 유전자 발현은 표준 논리주의 회귀 분석을 위한 분류를 위한 하나의 특성 값으로 표준화된다.

유사한 접근법이 하위 네트워크 대신 차별적 경로의 특징을 가진 애플리케이션에 대해 나중에 제안되었다.

이 접근법에서, **경로의 유전자 표현은 분자 서명 데이터베이스로부터의 경로 수집을 위한 하나의 특징으로 표준화된다.**

이 논문은 질병 특이적인 서브 네트워크를 특징으로 사용했으며 알려진 질병 유전자 세트를 먼저 PPI 네트워크에 매핑 한 다음 질병 유전자의 서브 네트워크를 질병 모듈 기능으로 확인했습니다.

이 논문은 PPI 네트워크에서 각 환자의 돌연변이 데이터에 라벨 전파를 구현하여 환자 분류를 위한 네트워크 상의 특징을 생성할 것을 제안했다.

이 논문은 PPI 네트워크에서 차별적으로 발현된 모든 유전자를 연결하기 위해 작은 서브 네트워크를 찾은 다음 환자 샘플을 분류하는 기능으로 서브 네트워크의 유전자를 사용하는 방법을 제안했습니다.

This setting is the Steiner tree problem in graph theory, and a heuristic algorithm coupled with randomization was designed to combine multiple suboptimal Steiner trees to find an optimum solution with a higher probability.

이 설정은 그래프 이론의 Steiner 트리 문제이며 무작위화와 결합된 발견적 알고리즘은 여러 차선의 Steiner 트리를 결합하여 더 높은 확률로 최적 솔루션을 찾도록 설계되었습니다.

This category of algorithms is a very useful generalization of the earlier gene-set-based methods since the network structures suggest dynamic modules among the genes rather than a fixed set.

이 알고리즘 범주는 네트워크 구조가 고정된 집합보다는 유전자들 사이에 역동적인 모듈을 제안하기 때문에 초기 유전자 집합 기반 방법의 매우 유용한 일반화이다.

These modules can be data-specific and disease-specific for improved results.

이 모듈은 향상된 결과를 위해 데이터 별 그리고 특정 질병일 가능성이 있습니다.

Thus, the data-driven subnetwork discovery introduced by these methods is a key improvement over previous studies.

따라서 이러한 방법으로 도입된 데이터 기반 서브 네트워크 검색은 이전 연구보다 중요한 개선 사항입니다.

Post-analysis of oncogenic alterations in networks

네트워크에서 발암성 변화의 사후 분석

The post-analysis integration methods also consist of two steps, as illustrated in Fig. (2-c).

분석 후 통합 방법은 또한 그림 (2-c)와 같이 두 단계로 구성된다.

First, the genomic profiles are analyzed to generate a list of oncogenic alterations; second, the detected alterations are analyzed in the network.

먼저, 유전체 프로파일을 분석하여 발암성 변화 목록을 생성합니다. 둘째, 감지된 변경 사항이 네트워크에서 분석됩니다.

In this post-analysis integration, the network information is integrated in the analysis after the oncogenic alterations are first detected by standard statistical methods. The purpose of these methods is to assess how cancer-driving alterations disrupt a normal cellular system by examining the influences on network components.

이 사후 분석 통합에서 네트워크 정보는 표준 통계 방법으로 발암 유전자 변형이 처음 발견된 후 분석에 통합됩니다. 이 방법의 목적은 암을 유발하는 변화가 네트워크 구성 요소에 미치는 영향을 조사하여 정상적인 세포 시스템을 어떻게 파괴하는지 평가하는 것입니다.

The circuit flow algorithm first identifies differentially expressed genes and then the genomic aberrations by mutations and copy number variations (CNVs) associated with the differential gene expression.

회로 흐름 알고리즘은 미분 유전자 표현과 관련된 돌연변이 및 카피 수 변화 (CNVs)에 의해 차별적으로 표현된 유전자를 확인한 다음 유전체 수차를 확인합니다.

Next, a current flow algorithm is applied to find causal paths from the causal genes (altered genes) to the target genes (differentially expressed genes) in a PPI network.

다음으로, PPI 네트워크에서 인과적인 유전자 (변형된 유전자)에서 표적(단백질) 유전자 (차별적으로 발현된 유전자)로의 인과 경로를 찾기 위해 전류 흐름 알고리즘이 적용된다.

Finally, the causal genes are selected by a set-covering algorithm to explain all the differentially expressed target genes.

마지막으로, 원인 유전자는 차별적으로 발현된 모든 표적(단백질) 유전자를 설명하기 위해 (세트 – 커버링) 알고리즘에 의해 선택된다.

HotNet first maps gene alterations in a gene network and then employs a diffusion kernel to build an influence graph with the edges weighted by the influence between each pair of genes.

HotNet은 먼저 유전자 네트워크에서 유전자 변형을 매핑한 다음 확산 커널을 사용하여 각 쌍의 유전자 사이의 영향에 의해 가중치가 적용된 영향 그래프를 작성합니다.

Then, a combinatorial problem is formulated to find the subnetworks of genes altered in a significant number of patients.

그런 다음 상당한 수의 환자에서 변형된 유전자의 하위 네트워크를 찾기 위해 조합 문제가 공식화됩니다.

Similarly, TieDIE and HotNet2, an extension of HotNet, apply network diffusion to analyze multiple types of genomic alterations, and NetPathID applies network diffusion to analyze CNVs in 16 types of cancers.

마찬가지로 HotNet의 확장 인 TieDIE와 HotNet2는 여러 유형의 유전체 변형을 분석하기 위해 네트워크 확산을 적용하고 NetPathID는 16가지 유형의 암에서 CNV를 분석하기 위해 네트워크 확산을 적용합니다.

PARADIGM is a probabilistic graphical model framework used to model the gene transcription, translation and post-translational events.

PARADIGM은 유전자 전사, 번역 및 번역 후 사건을 모델링하는 데 사용되는 확률적 그래픽 모델 프레임 워크입니다.

Each gene is modeled by a factor graph of DNA copy numbers, gene expression, protein levels and protein activities.

각 유전자는 DNA 복사 수, 유전자 발현, 단백질 수준 및 단백질 활동의 요소 그래프로 모델링 됩니다.

The factor graphs of genes are connected based on their regulatory relations in a pathway.

유전자의 요인 그래프는 경로의 규제 관계를 기반으로 연결됩니다.

The genomic and proteomic data are analyzed in the graphical models for the inference of pathway activities in each patient to derive integrated pathway activity (IPA) scores.

유전체 및 단백질 데이터는 그래픽 모델에서 분석되어 각 환자의 경로 활동 추정을 통해 통합 경로 활동(IPA)점수를 도출합니다.

The significantly altered genes/pathways can be identified using the IPA scores.

IPA 점수를 사용하여 유의하게 변형된 (유전자/경로)를 확인할 수 있습니다.

The mutual exclusivity module (MEMo) method is another widely used method in the TCGA project.

상호 배타성 모듈(MEMo) 방법은 TCGA 프로젝트에서 널리 사용되는 또 다른 방법입니다.

MEMo first builds a matrix representation of genes that are significantly altered by mutations or CNVs.

MEMo는 먼저 돌연변이 또는 CNV에 의해 크게 변형된 유전자의 매트릭스 표현을 만듭니다.

Then, the altered genes are connected by their proximal in the HPRD PPI network.

그런 다음 변경된 유전자는 HPRD PPI 네트워크에서 근위부에 의해 연결됩니다.

Finally, the cliques (a subgraph with all the gene pairs connected) are identified to analyze the mutual exclusivity in the patient data.

마지막으로, 파벌(모든 유전자 쌍이 연결된 하위 그래프)은 환자 데이터의 상호 배타성을 분석하기 위해 식별됩니다.

Signaling pathway impact analysis (SPIA) and mixed integer programming (MILP) are two examples of earlier pathway-based methods for genomic data analysis.

신호 경로 영향 분석 (SPIA) 및 혼합 정수 프로그래밍 (MILP)은 유전체 데이터 분석을 위한 초기 경로 기반 방법의 두 가지 예입니다.

SPIA applies an iterative algorithm similar to a random walk to measure the pathway perturbations in the regulatory network such that the impact of differentially expressed genes on a pathway can be evaluated.

SPIA는 무작위 걷기와 유사한 반복 알고리즘을 적용하여 차별적으로 발현된 유전자가 경로에 미치는 영향을 평가할 수 있도록 규제 네트워크의 경로 섭동을 측정합니다.

Fig 4

Methods for network-based drug repositioning.

(a) Graph connectivity measures consider the local structures of the networks to predict drug–target interactions.

This example shows the shortest path from each target node to the query drug (red node) in the graph.

(b) Link prediction models predict the relations between drugs and targets based on the global structures of the known interactions in the networks with matrix completion or random-walk approaches.

The known and predicted drug–target interactions are green and red, respectively, in the drug–target relation matrix.

(c) Network-based classification methods first extract the network topological features for all the targets in the networks.

For each drug, a classifier can be trained with the known targets of the drug as positive samples and the others as negative samples.

The learned classifiers can then be used to predict the new targets in the test set for each drug.

(d) The advantages and disadvantages of the methods in each category are compared

Fig 4

네트워크 기반 신약재창출을를 위한 방법.

(a) 그래프 연결성 측정은 (약물 – 표적(단백질)) 상호 작용을 예측하기 위해 네트워크의 로컬 구조를 고려한다.

이 예제는 **그래프에서 각 대상 노드에서 미확인 약물 (적색 노드) 까지의 최단 경로를 보여줍니다.**

(b) 링크 예측 모델은 매트릭스 완성 또는 랜덤 워크 접근법을 사용하여, 네트워크에서 알려진 상호 작용의 글로벌 구조를 기반으로 약물과 목표 간의 관계를 예측합니다.

알려지고 예측된 (약물 – 표적(단백질)) 상호 작용은 (약물 – 표적(단백질)) 관계 매트릭스에서 각각 녹색(알려진)과 적색(예측된)입니다.

(c) 네트워크 기반 분류 방법은 먼저 네트워크의 모든 대상에 대한 네트워크 위상 특징을 추출한다.

각 약물에 대해 분류기는 **알려진 표적(단백질)을 양성 표본으로, 나머지는 음성 표본으로 훈련할 수 있습니다.**

그런 다음 학습된 분류기를 사용하여 각 약물에 대한 테스트 세트의 새로운 목표를 예측할 수 있습니다.

(d) 각 카테고리의 방법의 장단점을 비교한다.

MILP is an optimization model to predict flux activity states of genes based on gene expression and a metabolic network.

MILP는 유전자 발현과 신진대사 네트워크를 바탕으로 유전자의 유동 활동 상태를 예측하기 위한 최적화 모델이다.

Comparison of the methods

방법의 비교

Network-based analysis of genomic data is based on the assumptions that cancer-driven aberrations often target different genes in the same pathway or subnetwork in the molecular network

유전체 데이터의 네트워크 기반 분석은 흔히 암에 의해 유도된 이상이 분자 네트워크의 동일한 경로 또는 하위 네트워크에서 다른 유전자를 대상으로 한다는 가정에 기초한다.

and that such systematic behavior can be observed as a coordinated change of genes' functions in pathways or network modules.

그리고 그러한 체계적인 행동이 경로나 네트워크 모듈에서 유전자 기능의 조정된 변화로서 관찰될 수 있도록 한다.

Network-based analysis is an effective approach because it has been observed that mutated genes in a cancer pathway can either co-occur in the same patients or be mutually exclusive among the patients,

네트워크 기반 분석은 암 경로의 돌연변이 유전자가 동일한 환자에서 동시에 발생하거나 환자 간에 상호 배제될 수 있다는 사실이 관찰되었기 때문에 효과적인 접근법이다.

and the systematic behavior is a more detectable and interpretable signal for the assessment of functional impacts of the aberrations.

그리고 체계적인 거동은 수차의 기능상의 영향을 평가하기 위해보다 탐지 가능하고 해석 가능한 신호이다.

It has also been shown that feature selection smoothed by graph Laplacian regularization based on the gene co-expression network is highly robust and generates more reproducible feature selections across independent datasets.

또한 유전자 동시 발현 네트워크를 기반으로 한 Laplacian 그래프 (Laplacian) 정규화에 의해 매끄럽게 처리된 특징 선택은 매우 견고하며 독립적인 데이터 세트 전반에 걸쳐보다 재현 가능한 특징 선택을 생성하는 것으로 나타났다.

Thus, the network-based approach is both well motivated and validated.

따라서 네트워크 기반 접근법은 동기가 부여되고 유효성이 확인됩니다.

The three categories of methods have different relative advantages and disadvantages.

세 가지 범주의 방법에는 서로 다른 상대적 장점과 단점이 있습니다.

Model-based integration methods are a fully supervised approach for both outcome prediction and subnetwork detection.

<모델 기반 통합 방법>은 결과 예측 및 하위 네트워크 탐지 모두에 대해 완전히 감독된 방식입니다.

The subnetworks are jointly discovered to contrast the control/case groups in the study based on a global optimization strategy, and thus these methods typically perform better in outcome prediction.

서브 네트워크는 글로벌 최적화 전략을 기반으로 한 연구에서 대조군 / 사례 그룹을 대조하기 위해 공동으로 발견되며, **따라서 이 방법은 일반적으로 결과 예측이 더 쉽습니다.**

In addition, the models can be tuned by a few clearly defined parameters, making it possible to train the models with cross-validation in contrast to the two-step methods in the other categories.

또한 모델은 몇 가지 명확하게 정의된 매개 변수로 조정할 수 있으므로, 다른 범주의 2단계 방법과 달리 **교차 유효성 검사를 통해 모델을 학습할 수 있습니다.**

The disadvantage is the need for more sophisticated optimization techniques, which are often less scalable.

단점은 확장성이 떨어지지만, 더욱 정교한 최적화 기술이 필요하다는 것입니다.

The preprocessing integration methods are more flexible in detecting customizable subnetwork features such that the detected features clearly reflect the hypothesized network-based characteristics.

<전처리 통합 방법>은 사용자 정의 가능한 서브 네트워크 기능을 보다 유연하게 탐지하여 탐지된 기능이 가정된 네트워크 기반 특성을 명확하게 반영하도록 합니다.

For example, the size and density of discriminative subnetworks can be precisely specified.

예를 들어, 식별 가능한 서브 네트워크의 크기와 밀도를 정확하게 지정할 수 있습니다.

However, it is not possible to guarantee that the detected subnetwork features are optimal features for prediction with the standard learning model in the second step.

그러나 두 번째 단계에서 표준 학습 모델을 사용하여 검색된 하위 네트워크 기능이 예측을 위한 최적의 기능임을 보장할 수 없습니다.

The post-analysis integration methods focus on associating mutations or other DNA aberrations with differential expression or certain other molecular phenotypes in the network context.

<분석 후 통합 방법>은 네트워크 상황에서 돌연변이 또는 기타 DNA 이상과 미분 표현 또는 특정 다른 분자 표현형을 연관시키는 데 중점을 둡니다.

Thus, these methods are highly informative regarding cancer mechanisms in the network.

따라서 이러한 방법은 네트워크에서 암 메커니즘과 관련하여 매우 유익합니다.

In model-based integration, Graph LASSO is another choice of graph-based regularization other than the graph Laplacian regularizer.

모델 기반 통합에서 Graph LASSO는 Laplacian 그래프 정규화기가 아닌 그래프 기반 정규화의 또 다른 선택입니다.

Graph LASSO imposes a LASSO loss on each pair of connected variables in the network rather than a squared error as with the graph Laplacian regularizer.

그래프 LASSO는 Laplacian 정규 표현식 그래프와 마찬가지로 제곱 오차가 아니라 네트워크의 연결된 변수 쌍마다 LASSO 손실을 부과합니다.

The LASSO loss terms force the coefficients of the connected pairs to be identical such that the inconsistent pairs are "sparse."

LASSO 손실 조항은 연결된 쌍의 계수가 동일하도록 하여 일치하지 않는 쌍이 "희소 (sparse)"되도록 합니다.

In practice, the assumption can be too strong in networks with overlapping clusters.

실제로, 군집들이 겹치는 네트워크에서는 가정이 너무 강할 수 있습니다.

In addition, optimization of Graph LASSO-constrained models is generally challenging, while the graph Laplacian regularizer is a quadratic constraint that is relatively straightforward to optimize.

또한 **Graph LASSO 구속 모델의 최적화는 일반적으로 어려운 일인 반면, Laplacian 그래프 정규 표현은 최적화하기가 비교적 간단한 이차 구속 조건입니다.**

Thus, Graph LASSO is a less common choice for network-based integration methods.

따라서 그래프 LASSO는 네트워크 기반 통합 방법에 대한 덜 일반적인 선택입니다.

NETWORK-BASED METHODS FOR DRUG REPOSITIONING

신약재창출을를 위한 네트워크 기반 방법

Network-based algorithms have also been developed for drug repurposing by exploring drug–drug similarities, drug–target relations and gene-gene relations.

(약물 – 약물) 유사성, (약물 – 표적(단백질)) 관계 및 (유전자 – 유전자) 관계를 탐구하여 약물 재사용을 위한 네트워크 기반 알고리즘도 개발되었습니다.

These methods can be largely classified into three categories, i.e., graph connectivity measures, link prediction models and network-based classification methods, as illustrated in Fig.(4).

이러한 방법들은 크게 그림4에서 볼 수 있듯이 그래프 연결성 측정, 링크 예측 모델, 네트워크 기반 분류법의 세 가지 범주로 분류할 수 있다.

The methods reviewed under each category are also listed in Fig.(1-b).

각 카테고리에서 검토된 방법은 그림(1-b)에도 나와 있습니다.

Below, we describe and compare the methods in the three categories.

아래에서는 세 가지 범주의 방법을 설명하고 비교합니다.

Graph connectivity measures

그래프 연결 측정 값

The methods in this category are based on measuring the connectivity among the nodes in the graph, such as neighboring relations, the number of shared neighbors and shortest paths, to derive drug–drug, drug–target or drug–disease relations, as illustrated in Fig.(4-a).

이 범주의 메소드는 인접 관계, 공유된 이웃의 수와 같은 그래프에서 노드 간의 연결성을 측정하는 것에 기반합니다

그리고 그림(4-a)와 같이 (약물 – 약물), (약물 – 표적(단백질)) 또는 (약물과 질병)의 관계를 도출하기위한 최단 경로를 제공한다.

Several early studies showed that drugs sharing similar chemical structures, transcriptional responses following treatment and text mining analysis often share the same target, where the implication is that the drug–drug network based on the similarities can be used to reposition a drug for the targets of similar drugs.

몇몇 **초기 연구들은 유사한 화학 구조를 공유하는 약물, 치료와 텍스트 마이닝 분석에 따른 초월 반응이 종종 같은 표적(단백질)을 공유한다는 것을 보여주었다.**

여기서 암시하는 바는 유사한 유사성에 기초한 (약물 – 약물) 네트워크가 유사한 약물의 대상을 위해 약물의 위치를 변경하는 데 사용될 수 있다는 것이다.

The paper derived drug–drug similarities based on mining the side-effect description from medical symptoms in the Unified Medical Language System ontology.

이 논문은 Unified Medical Language System 종양학의 의학적 증상에서 부작용에 대한 설명을 토대로 (약물 – 약물) 유사점을 도출했습니다.

The paper developed a method to predict similarities in terms of drug effect by comparing gene expression profiles following drug treatment across multiple cell lines and dosages.

이 논문은 여러 세포주와 용량에서 약물 치료 후 유전자 발현 양상을 비교함으로써 약물 효과면에서 유사성을 예측하는 방법을 개발했다.

Both studies validated the correlation between drug–drug similarity and the likelihood of two drugs sharing a common protein target.

두 연구 모두 (약물 – 약물) 유사성과 일반적인 단백질 표적(단백질)을 공유하는 두 약물의 가능성 사이의 상관 관계를 입증했다.

Based on the observations, the paper proposed a recommendation technique for predicting drug–target relations based on the drug–drug similarity matrix W computed based on the structural similarity of the drugs and sequence similarity of their targets and the known drug–target matrix A.

이 논문은 약물의 구조적 유사성과 표적(단백질)의 서열 유사성 및 알려진 (약물 – 목표) 매트릭스 A에 기초하여 계산된 (약물 – 약물) 유사성 행렬 W에 기초하여 (약물 – 표적(단백질)) 관계를 예측하는 권고 기법을 제안했다.

By a simple multiplication (R = WA), the scores in matrix R can be used to derive a ranking of the candidate targets against each drug.

간단한 곱셈 (R = WA)에 의해, 매트릭스 R의 스코어는 각 약물에 대한 후보 표적(단백질)의 랭킹을 유도하는데 사용될 수 있다.

The paper performed a large-scale analysis of ~7000 genomic expression profiles in the Gene Expression Omnibus with human disease and drug annotations to create a disease–drug network consisting of drug–drug, drug–disease and disease–disease relations.

이 논문은 인간의 질병 및 약물 주석이 있는 Gene Expression Omnibus에서 ~ 7000개의 유전체 표현 프로필을 대규모 분석하여 (약물 – 약물), (약물 – 질병) 및 (질병 – 질병) 관계로 구성된 (질병 – 약물) 네트워크를 창출했습니다.

The study shows that the derived disease–disease relations are highly consistent with the definition in the Medical Subject Headings disease classification tree and that the drug–disease relations can be used to generate hypothesized drug repositioning and side effects.

이 연구는 파생된 (질병 – 질병) 관계가 의학 주제 표제 질병 분류 트리의 정의와 매우 일치하고 (약물 – 질병) 관계가 가정된 약물 재 위치 및 부작용을 생성하는 데 사용될 수 있음을 보여줍니다.

The paper further generalized the inference to drug–disease proximity in the network by the hypothesis that an effective drug for a disease must target proteins within or in the immediate vicinity of the corresponding disease module in the molecular interaction network.

이 논문은 질병에 대한 효과적인 약물이 분자 상호 작용 네트워크에서 해당 질병 모듈의 바로 근처 또는 내부에 있는 단백질을 표적(단백질)해야 한다는 가설에 의해 네트워크에서 (약물 – 질병) 근접성에 대한 추론을 더 일반화했다.

They applied a shortest-path-based measure coupled with a randomization normalization technique to derive the drug–disease proximity scores for the inference.

그들은 추론을 위한 (약물 – 질병) 근접 점수를 도출하기 위해 무작위 표준화 기술과 결합된 최단 경로 기반 측정을 적용했다.

A recent work in the paper performed a correlation analysis of disease modules and drug targets in the functional linkage network.

이 논문의 최근 연구는 기능적 연계 네트워크에서 질병 모듈과 약물 표적(단백질)의 상관 분석을 수행했습니다.

The differentially expressed disease genes and the drug–target genes are first overlapped in the functional linkage network, and a mutual predictability score is then computed based on the neighboring relations among the genes to evaluate the repositioning of the drug for the disease.

차별적으로 발현된 질병 유전자와 약물 표적(단백질) 유전자를 먼저 기능적 연계 네트워크에서 중첩시키고, 이 유전자에 대한 인접한 관계를 토대로 상호예측성 점수를 계산하여 질병에 대한 약물의 재배치를 평가한다.

Link prediction models

예측 모델 연결

Link prediction models predict the relations between drugs and targets based on the global structures of the known interactions in the networks with matrix completion or random-walk approaches, as illustrated in Fig.(4-b).

링크 예측 모델은 그림 (4-b)에서 설명한 바와 같이 행렬 완성 또는 무작위 도보 접근법을 사용하여 네트워크에서 알려진 상호 작용의 전역 구조를 기반으로 약물과 대상 간의 관계를 예측합니다.

The paper predicted drug–target relations for drug repositioning based on a network of three types of relations: drug–drug structural similarity, target–target sequence similarity and drug–target relations from DrugBank.

이 논문은 (약물 – 약물) 구조적 유사성, (표적(단백질) – 표적(단백질)) 서열 유사성 및 DrugBank에서의 (약물 – 표적(단백질)) 관계의 3 가지 유형의 네트워크를 바탕으로 약물 위치 조정을 위한 (약물 – 표적(단백질)) 관계를 예측했다.

It was shown that exploring the network topology outperforms simple inference rules by graph connectivity measures such as similar drugs sharing the same target or similar targets sharing the same drug.

네트워크 위상 탐색은 동일한 표적(단백질) 또는 동일한 약물을 공유하는 유사한 표적(단백질)을 공유하는 등의 그래프 연결 조치로 단순한 추론 규칙을 능가하는 것으로 나타났다..

The paper applied an information-flow approach on a heterogeneous network of drug–drug, disease–disease and target–target similarities along with the known disease–drug and drug–target relations.

이 논문은 알려진 (질병 – 약물) 및 (약물 – 표적(단백질)) 관계와 함께 (약물 – 약물), (질병 – 질병) 및 (표적(단백질) – 표적(단백질)) 유사성의 이질적인 네트워크에서 (정보 – 흐름) 접근법을 적용했다.

The algorithm iteratively updates the disease–drug and drug–target relations and converges to stationary scores for the prediction of their relations.

이 알고리즘은 (질병 – 약물) 및 (약물 – 표적(단백질)) 관계를 반복적으로 업데이트하고 관계의 예측을 위해 정지 된 점수로 수렴합니다.

The paper introduced a bipartite graph-learning method based on kernel regression to learn a co-mapping of drugs in chemical space and targets (proteins) in genomic space into a common pharmacological space.	**이 논문은 화학적 공간에서 약물의 공동 매핑과 유전체 공간의 표적(단백질)을 일반적인 약리학적 공간으로 배울 수 있는 커널 회귀 분석에 기초한 이분 그래프 학습 방법을 소개했다.**
In the pharmacological space, the correlation between compound-protein pairs can be conveniently calculated to predict their interactions for drug repositioning.	**약리학적 공간에서, (화합물 – 단백질) 쌍간의 상관 관계는 신약재창출에 대한 상호 작용을 예측하기 위해 편리하게 계산 될 수 있다.**
The paper proposed a collaborative matrix factorization method to factorize known drug–target relations to predict new relations constrained by the drug–drug similarity network and the target–target similarity network.	**논문은 (약물 – 약물) 유사성 네트워크와 (표적(단백질) – 표적(단백질)) 유사성 네트워크에 의해 제약된 새로운 관계를 예측하기 위해 알려진 (약물 – 표적(단백질)) 관계를 인수분해하기위한 협업 행렬 인수 분해 방법을 제안했다.**
The paper proposed a manifold regularization semi-supervised learning method in which two classifiers in drug space and target space are learned and then combined to give a final score for drug–target interaction prediction.	**이 논문에서는 약물 공간과 표적(단백질) 공간의 두 분류자를 학습한 다음 결합하여 (약물 – 표적(단백질)) 상호 작용 예측을 위한 최종 점수를 부여하는 다각화된 정규화 준 지도 학습 방법을 제안했습니다.**
The paper applied several random-walk methods on a heterogeneous network of drug–drug similarities, target–target similarities and drug–target relations such that the global structure among all the networks can be used to improve the prediction of new drug–target pairs.	**이 논문은 (약물 – 약물) 유사성, (표적(단백질) – 표적(단백질)) 유사성과 (약물 – 표적(단백질)) 관계의 잡다한 네트워크에서, 여러 가지 랜덤 워크 방법을 적용하여 모든 네트워크 중 글로벌 구조가 (신약 – 표적(단백질)) 쌍의 예측을 향상시키는 데 사용될 수 있도록 했다.**

Network-based classification methods

네트워크 기반 분류 방법

Network-based drug repositioning can also be reformulated as a classification problem such that standard classification methods can be applied to predict the new targets of each drug, as illustrated in Fig.(4-c).

네트워크 기반 신약재창출은 또한 그림 (4-c)와 같이 표준 분류 방법을 적용하여 각 약물의 새로운 표적(단백질)을 예측할 수 있도록 분류 문제로 재구성될 수 있다.

These methods first extract the network topological features for all the targets in the networks.

이러한 방법은 먼저 네트워크의 모든 대상에 대한 네트워크 위상 기능을 추출합니다.

For each drug, a classifier can be trained with the known targets of the drug as positive samples and the others as negative samples.

각 약물에 대해, 분류기는 약물의 알려진 표적(단백질)을 양성 표본으로, 다른 표적(단백질)을 음성 표본으로 훈련시킬 수 있습니다.

The learned classifiers can then be used to predict the new targets in the test set for each drug.

그런 다음 학습된 분류기를 사용하여 각 약물에 대한 테스트 세트의 새로운 목표를 예측할 수 있습니다.

The paper proposed mapping disease-specific differentially expressed genes into a PPI network and using network topological features to detect new drug targets based on the known targets from the drug–target database by logistic regression.

이 논문은 질병 특이적으로 차별적으로 발현된 유전자를 PPI 네트워크로 매핑하고 네트워크 회귀 분석을 통해 약물 표적(단백질) 데이터베이스에서 알려진 표적(단백질)을 기반으로 신약 표적(단백질)을 검출하기 위해 네트워크 위상 특성을 사용하는 방법을 제안했다.

The paper also applied a supervised bipartite model to predict the probability of each drug–target interaction based on the known drug targets as labels and the target–target interactions as features, where the bipartite model was augmented with additional training samples from the neighboring drug–target relations.

또한 이 논문은 알려진 약물 표적(단백질)을 기반으로 한 약물 표적(단백질)과 (표적(단백질)-표적(단백질)) 상호 작용을 특징으로 하여 각 (약물-대상) 교호 작용의 확률을 예측하기 위해 슈퍼 분류 모형을 적용했다. 여기서, 두 부분 모델은 인접한 (약물-표적(단백질)) 관계의 추가 교육 샘플로 보강되었습니다.

The paper constructed a drug–drug kernel matrix based on chemical structure similarities and a target–target kernel matrix based on sequence similarities.

이 논문은 화학적 유사성에 기초한 (약물 – 약물) 커널 매트릭스와 서열 유사성에 기반한 (표적(단백질) – 표적(단백질)) 커널 매트릭스를 구성했다.

For each drug, using the known targets as the positive training samples, an SVM classifier is built with the target–target kernel matrix to classify the candidate genes for new targets.

각 약물에 대해, 알려진 표적(단백질)을 양성 훈련 표본으로 사용하여 (표적(단백질) – 표적(단백질)) 핵 매트릭스와 함께 SVM 분류기를 구축하여 새로운 표적(단백질)에 대한 후보 유전자를 분류합니다.

In addition, for each target and using the known drugs as the positive training samples, an SVM classifier is built with the drug–drug kernel matrix to classify the drugs for new repositioned drugs.

또한 각 표적(단백질)에 대해 알려진 약물을 양성 훈련 표본으로 사용하여 새로운 재조합 약물에 대한 약물 분류를 위해 SVM 분류기를 (약물 – 약물) 커널 매트릭스와 함께 구축합니다.

The paper adopted a similar approach with two additional advanced kernel methods, applying diffusion-types of kernels to integrate both the drug–drug kernel matrix and the target–target kernel matrix to predict the new targets of a drug or the new repositioned drugs for a target.

이 논문은 확산형 커널을 적용하여 (약물 – 약물) 커널 매트릭스 및 (표적(단백질) – 표적(단백질)) 커널 매트릭스를 이용하여 새로운 표적(단백질) 또는 표적(단백질)을 위한 새로운 위치 변경 약물을 예측할 수 있다.

Comparison of the methods

방법의 비교

The three categories of methods have different relative advantages and disadvantages, as shown in Fig.(4-d).

세 가지 범주의 방법은 그림 (4-d)와 같이 서로 다른 상대적 장점과 단점을 가지고 있다.

Graph connectivity measures are straightforward to implement based on standard graph algorithms, and the prediction results are easy to interpret with the edges and the paths in the graph.

그래프 연결 측정은 표준 그래프 알고리즘을 기반으로 구현하기가 쉽고, 예측 결과는 그래프의 가장자리와 경로로 해석하기 쉽습니다.

However, the prediction performance is typically worse since only relatively local information of the networks is considered by the graph algorithms.

그러나, 네트워크의 상대적으로 로컬 정보만 그래프 알고리즘에 의해 고려되기 때문에 예측 성능은 일반적으로 더 나쁘다.

Link prediction models retrieve the global structures of the networks to predict drug–target interactions for better prediction performance.

링크 예측 모델은 네트워크의 글로벌 구조를 검색하여 더 나은 예측 성능을 위해 (약물-대상) 상호 작용을 예측합니다.

The disadvantages are the lack of a satisfactory interpretation of the predictions and that the implementation of the models often relies on advanced optimization algorithms.

단점은 예측에 대한 만족스러운 해석이 부족하고 모델 구현이 고급 최적화 알고리즘에 의존한다는 것입니다.

When sophisticated optimization is required, the scalability can be poor.

정교한 최적화가 필요한 경우 확장성이 떨어질 수 있습니다.

Network-based classification methods are more accurate for repositioning drugs with many known targets as the training samples but are not applicable to drugs with few or no known targets.

네트워크 기반 분류 방법은 알려진 대상이 많은 의약품을 훈련 샘플로 신약재창출할 때 더 정확하지만 알려진 대상이 거의 없는 의약품에는 적용되지 않는다.

The prediction results can be interpreted by the network topological features extracted from the networks, depending on the feature extraction strategy.

예측 결과는 네트워크에서 추출된 네트워크 위상 특징에 의해 해석될 수 있다.

Another important aspect of the comparison is whether a method can generate de novo predictions for drugs with no known targets or gene targets with no known drugs.

비교의 또 다른 중요한 측면은 알려진 표적(단백질)이 없는 약물에 대한 새로운 예측이나, 알려진 약물이 없는 유전자 표적(단백질)을 예측할 수 있는지 여부입니다.

Graph connectivity measures are often more biased towards highly connected nodes in the graph such that new drugs or less-studied genes typically receive low rankings.

그래프 연결성 측정는 종종 그래프에서 고도로 연결된 노드 쪽으로 편향되어 새로운 약물이나 덜 연구된 유전자가 일반적으로 낮은 순위를 받습니다.

Thus, de novo predictions are rarely made by graph connectivity measures.

따라서 그래프 연결성 측정에 의해 새로운 예측은 거의 발생하지 않습니다.

With no positive training pairs available, the network-based classification methods simply abandon the de novo cases.

사용할 수 있는 긍정적인 트레이닝 쌍이 없기 때문에 네트워크 기반 분류 방법은 그냥 새로운 경우를 포기합니다.

Link prediction models are often the most capable of making de novo predictions because global topological structures are generally less biased after proper normalization and control by randomization.

예측 모델 연결은 종종 글로벌 위상 구조가 적절한 정규화 및 무작위 화에 의한 제어 후에 편향되기 때문에 일반적으로 새로운 예측을 할 수 있습니다.

Fig. 5

Network-based analysis of highly mutated pathways of 31 cancer types in TCGA data.

The highly mutated pathways detected by (a) network-based analysis and (b) standard enrichment analysis.

The pathways of interest in the discussion are highlighted in blue, and the pathways only enriched by network-based analysis are highlighted in red

Fig. 6

Network-based analysis of patient mutation data in TCGA ovarian cancer.

The significantly mutated pathways in each patient detected by (a) network analysis and (b) the analysis of the original mutation data without the network.

(c) The survival plot of the three groups detected by the network-based pathway analysis of the TCGA ovarian cancer patients.

Derived by standard log-rank test, the p-values for comparing group 2 vs. group 3 and group 1 + group 2 vs. group 3 are both significant.

(d) The survival plot of the groups detected by the analysis of the original mutation data of the TCGA ovarian cancer patients

Fig. 5

TCGA 데이터에서 31 가지 암 유형의 고도로 변이 된 경로의 네트워크 기반 분석.

고도로 돌연변이 된 경로는 (a) 네트워크 기반 분석 및 (b) 표준 농축 분석에 의해 검출됩니다.

토론에서 관심있는 경로는 파란색으로 강조 표시되어 있으며, 네트워크 기반 분석으로만 강화된 경로는 빨간색으로 강조 표시되어 있습니다

Fig. 6

TCGA 난소암에서 환자 돌연변이 데이터에 대한 네트워크 기반 분석.

(a)네트워크 기반 분석 및

(b)네트워크가 없는 원래 돌연변이 데이터의 분석에 의해 탐지된 각 환자의 현저하게 변형된 경로.

(c) TCGA 난소 암 환자의 네트워크 기반 경로 분석에 의해 검출된 세 그룹의 생존 플롯.

표준 (로그 – 랭크) 테스트에 의해 유도된 (그룹 2 대 그룹 3) 및 (그룹 1 + 그룹 2 대 그룹 3)의 p 값은 모두 중요하다.

(d) TCGA 난소 암 환자의 원래 돌연변이 데이터 분석에 의해 검출된 그룹의 생존 플롯

NETWORK-BASED ANALYSIS OF TCGA MUTATION DATA AND A CASE STUDY ON OVARIAN CANCER

TCGA 돌연변이 데이터의 네트워크 기반 분석 및 난소암에 대한 사례 연구

To better discuss the network-based methods, we performed a network-based analysis of the mutated genes in the cancer genome projects in TCGA78–101 and summarized the enriched KEGG pathways in Fig.(5).

네트워크 기반 방법을 더 잘 논의하기 위해 TCGA78-101의 암 유전체 프로젝트에서, 돌연변이 된 유전자의 네트워크 기반 분석을 수행하고, 그림 (5)에서 강화된 KEGG 경로를 요약했습니다.

For the analysis, the mutation frequencies among the patients in the TCGA provisional studies

분석을 위해, TCGA임시 연구에서 환자들 사이의 돌연변이 빈도가 암 유전체 사이트인 cBioPortal에서 다운로드 되었습니다.

In the network-based analysis, label propagation (λ = 0.5), as described in Table S2 in the Supplementary Information was applied to the HPRD PPI network in each cancer study to capture the highly mutated subnetworks.

네트워크 기반 분석에서 보충 정보의 표 S2에 설명된 대로 표지 전파 (λ = 0.5)를 각 암 연구의 HPRD PPI 네트워크에 적용하여 고도로 변이된 서브 네트워크를 포착했습니다.

The initialization was the gene mutation frequency among the patients in each cancer study for label propagation.

초기화는 표지 전파에 대한 각 암 연구에서 환자들 사이의 유전자 돌연변이 빈도였다.

The summation of the stationary scores of the genes in a KEGG pathway is compared with the scores of 10,000 random gene sets of the same size to derive p-values.

KEGG 경로에서 유전자의 고정 점수의 합계를 같은 크기의 10,000개의 무작위 유전자 세트의 점수와 비교하여 p 값을 유도합니다.

In the analysis without the network, the highly mutated genes in each cancer type are overlapped with KEGG pathways with enrichment analysis to derive p-values by hypergeometric test.

네트워크가 없는 분석에서, 각 암 유형의 고도로 변이 된 유전자는 KEGG 경로와 중첩 분석되어 고지 학적 시험에 의해 p 값을 유도한다.

This network-based analysis clearly detects more significantly mutated pathways than the analysis without using the network, as shown in Fig.(5-a,b), respectively.

이 네트워크 기반 분석은 그림(5-a,b)와 같이 네트워크를 사용하지 않고 분석한 것보다 훨씬 더 돌연변이 된 경로를 분명히 감지합니다.

Interestingly, the network-based analysis in Fig.(5-a) indicates that the AMPK signaling pathway is affected in breast cancer (BRCA) and uterine corpus endometrial cancer (UCEC).

흥미롭게도 그림(5-a)의 네트워크 기반 분석은 AMPK 신호 전달 경로가 유방암(BRCA)과 자궁 체부 자궁 내막 암(UCEC)에서 영향을 받는다는 것을 나타냅니다.

Prior studies demonstrated that BRCA patients receiving metformin, a pharmacological activator of AMPK, showed complete pathologic response, implicating the role of AMPK in BRCA.

이전 연구에서 AMPK의 약리학적 활성제인 메트포르민을 투여 받은 BRCA 환자는 BRCA에서 AMPK의 역할을 암시하는 완전한 병리학적 반응을 보였다.

Similarly, the loss of the AMPK activator LKB1 promotes endometrial cancer progression and metastasis, implicating the AMPK pathway in endometrial cancer, and metformin inhibits endometrial cancer cell proliferation.

마찬가지로, AMPK 활성화기 LKB1의 손실은 자궁 내막 암의 AMPK 경로를 암시하는 자궁 내막 암의 진행 및 전이를 촉진하고, 메트포르민은 자궁 내막 암 세포의 증식을 억제한다.

The HIF-1 pathway has been predicted to be affected in renal clear cell carcinoma (KIRC), BRCA, endometrial cancer (UCEC), glioblastoma multiforme (GBM), cervical cancer (CESC), and lung cancer (LUAD), and these results are consistent with prior studies implicating the VHL/HIF-1 pathway in these cancers.

HIF-1경로는 신경암(KIRC), BRCA, 자궁 내막 암(UCEC), 유방암(GBM), 자궁 경부 암 등이 영향을 받을 것으로 예측되었다.

이러한 결과는 이러한 암에 VHL.HIF-1경로를 적용한 이전 연구와 일치합니다.

The Hippo pathway has been predicted to be affected in colorectal cancer, renal papillary carcinomas, stomach cancer, and liver cancer, and these results are consistent with recent cancer genomic studies.

Hippo 경로는 결장 직장암, 신장 유두암, 위암 및 간암에서 영향을 받을 것으로 예측되었으며, 이러한 결과는 최근의 암 유전체 연구와 일치합니다.

Finally, the PI3K-Akt pathway has been identified as one of the most frequently affected pathways in several cancer types, and several components of this pathway were reported to be mutated or amplified in various cancer types.

마지막으로, PI3K-Akt 경로는 여러 가지 암 유형에서 가장 빈번하게 영향을 받는 경로 중 하나로 확인되었으며, 이 경로의 여러 구성 요소는 다양한 암 유형에서 돌연변이 되거나 증폭되는 것으로 보고되었습니다.

Collectively, these results suggest that network analysis can identify clinically relevant pathways that are altered in different cancer types.

종합적으로, 이러한 결과는 네트워크 분석이 다양한 암 유형에서 변경되는 임상적으로 관련된 경로를 식별할 수 있음을 시사한다.

In the case study on the ovarian cancer patients shown in Fig.(6), the mutation data of the 316 TCGA ovarian cancer patients were downloaded from the Xena Public Data Hubs.

그림(6)의 난소 암 환자에 대한 사례 연구에서 316 TCGA 난소암 환자의 돌연변이 데이터가 Xena Public Data Hubs에서 다운로드 되었습니다.

Similar to the study in the paper, label propagation ($\lambda = 0.1$) was applied on the same HPRD PPI network in each patient to detect the patientspecific highly mutated subnetworks.

논문에서의 연구와 마찬가지로, 각 환자의 동일한 HPRD PPI 네트워크 상에 라벨 전파 ($\lambda = 0.1$)를 적용하여 고도로 변이된 환자 네트워크를 탐지했다.

The initialization was 1 for the mutated genes and 0 for the other genes and then normalized to sum to 1.

초기화는 돌연변이를 일으킨 유전자에 대해서는 1이었고, 다른 유전자에 대해서는 0이었고, 그리고 나서 1로 요약되도록 정규화 되었습니다.

Similarly, the summation of the stationary scores of the genes in a KEGG pathway was compared with the scores of 10,000 random gene sets of the same size to derive the p-value.

유사하게, KEGG 경로에서 정지된 유전자의 합계를, p 값을 도출하기 위해 동일한 크기의 10,000개의 무작위 유전자 세트의 점수와 비교하였다.

In the analysis without the network, the mutated genes in each patient are overlapped with KEGG pathways with enrichment analysis to derive p-values by hypergeometric test.

네트워크가 없는 분석에서, 각 환자의 돌연변이 된 유전자는 KEGG 경로와 중복 분석되어 초 고밀도 테스트로 p 값을 얻는다.

Hierarchical clustering was applied to cluster the patients into three groups using the –log10 (p-values) as features.

계층적 클러스터링을 적용하여 -log10(p 값)을 피처로 사용하여 세 그룹으로 환자를 군집화 했습니다.

The network-based analysis informs a clustering of the patients by a significant relevance to survival (Fig.(6-c)).

네트워크 기반 분석은 환자의 군집화를 생존에 중요한 관련성으로 알려줍니다(그림 (6-c)).

Notably, three subgroups of tumor samples can be identified from the network-based analysis shown in Fig.(6-c), compared to four subgroups in the mutation-based analysis without the network in Fig.(6-d).

특히, 종양 샘플의 세 하위 그룹은 그림(6-d)에서 네트워크가 없는 돌연변이 기반 분석의 네 하위 그룹과 비교하여, 그림(6-c)에 표시된 네트워크 기반 분석에서 확인할 수 있습니다.

Although subgroups identified by mutation-based analysis without the network show no significant association with disease-free survival, two of the subgroups detected by the network-based analysis (Subgroup 1 and Subgroup 3) show significant association with disease-free survival relative to Subgroup 2.

네트워크가 없는 돌연변이 기반 분석으로 확인된 하위 집단이 질병이 없는 생존과 유의한 관련성을 보이지는 않지만, 네트워크 기반 분석 (하위 그룹 1과 하위 그룹 3)에 의해 탐지 된 하위 그룹 중 2 개는 하위 그룹 2와 관련하여 무병 생존과 유의 한 연관성을 보입니다.

Interestingly, Subgroup 1 has the highest copy number alterations, whereas Subgroup 3 has the highest number of pathway alterations.

흥미롭게도, (서브 그룹 1)은 가장 높은 카피 번호 변경을 갖는 반면, (서브 그룹 3)은 가장 많은 경로 변경 번호를 갖는다.

These results are analogous to the spectrum of somatic alterations described by ref. 112

이러한 결과는 참고문헌 112에서 설명한 체세포 변경 범위와 유사하다.

Although those authors placed ovarian cancer in class C, defined by extensive copy number alterations, the spectrum of somatic alterations can be further described as subgroups with higher copy number changes, mixed, and higher mutations within ovarian cancer.

이 저자들은 광범위한 사본 번호 변경으로 정의된 C 등급의 난소 암을 배치했지만, 체세포 변경의 스펙트럼은 난소 암에서 더 높은 복제 번호 변경, 혼합 및 높은 돌연변이를 갖는 하위 그룹으로 더 자세히 기술할 수 있습니다.

This case study shows that via network analysis, several subtypes of ovarian cancer can be grouped together for further assessment of clinical values, such as occurrence, relapse and treatment resistance.

이 사례 연구는 네트워크 분석을 통해 난소 암의 여러 하위 유형을 그룹화하여 발생, 재발 및 치료 저항성과 같은 임상적 가치를 더 평가할 수 있음을 보여줍니다.

This information may also be valuable for the design or assessment of treatment strategies.

이 정보는 치료 전략의 설계 또는 평가에 유용할 수도 있습니다.

Collectively, the network analysis unveils important cancer pathways and their correlation to subtypes of cancers that would not be identifiable by original mutation data analysis.

종합적으로, 네트워크 분석은 원래 돌연변이 데이터 분석으로 식별될 수 없는 암의 하위 유형에 대한 중요한 암 경로와 그 상관 관계를 밝혀 낸다.

DISCUSSION

Precision oncology tailors cancer treatment and repositions drugs based on personal genomic information.

There are several promising aspects of the application of network-based analysis in precision oncology.

With a network to capture the molecular organization in the cellular system, genomic data analysis is both more accurate and descriptive.

The smoothness constraint introduced into the model-based integration methods is helpful in eliminating false positives and false negatives in high-dimensional genomic data.

The network analysis identifies molecular targets in the context of pathways or interaction partners in a subnetwork that are interpretable for molecular mechanisms.

For example, in the case study in Fig.(6-a), the mutation information of each individual patient is propagated on the PPI network to detect the patient-specific subnetwork and improve the quality of the patient clustering by a significant relevance to survival.

As a consequence, network-based analysis often reports consistent marker genes across different studies of the same cancer or more comparable results in pan-cancer analysis.

Collectively, it is evident that network-based methods employ molecular and biomedical networks to extract useful personal genomic information, and build better predictive models for target identification, phenotype prediction and drug repositioning.

DISCUSSION

정밀 종양학은 암 치료 및 맞춤 유전체 정보를 기반으로 의약품을 신약재창출합니다.

정밀 종양학에서 네트워크 기반 분석의 적용에 대한 몇 가지 유망한 측면이 있습니다.

세포 시스템에서 분자 조직을 잡기 위한 네트워크로, 유전체 데이터 분석은 더 정확하고 설명적이다.

모델 기반 통합 방법에 도입된 부드러움 제한 조건은 고차원 유전체 데이터의 잘못된 긍정(false positive) 및 잘못된 부정(false negative)을 제거하는데 유용합니다.

네트워크 분석은 분자 메커니즘에 대해 해석할 수 있는 서브 네트워크의 경로 또는 상호 작용 파트너의 맥락에서 분자 목표를 식별합니다.

예를 들어, 그림(6-a)의 사례 연구에서 각 개별 환자의 돌연변이 정보는 PPI 네트워크에 전파되어 환자 특정 서브 네트워크를 탐지하고, 생존에 중요한 관련성에 의해 환자 군집화의 품질을 향상시킵니다 .

결과적으로, 네트워크 기반의 분석은 동일한 암에 대한 여러 연구 또는 전립선 암 분석에서의 비교 가능한 결과에 대해 일관된 표시 유전자를 보고하는 경우가 있습니다.

종합적으로, 네트워크 기반 방법은 유용한 개인 유전체 정보를 추출하기 위해 분자 및 생물 의학 네트워크를 사용하며, 표현형 예측, 신약재창출을를 위한 더 나은 예측 모델을 구축할 수 있습니다.

Conceptually, network-based analysis also adopts mutation patterns that are mutually exclusive or co-occurring.

특징적으로, 네트워크 기반 분석은 상호 배타적이거나 동시에 발생하는 돌연변이 패턴을 채택합니다.

Mutually exclusively mutated genes are often located on the same pathway, and network analysis propagates the mutually exclusive signals to identify the pathway by a significant signal.

상호 배타적으로 변이된 유전자는 종종 동일한 경로 상에 위치하며, 네트워크 분석은 상호 배타적인 신호를 전파하여 중요한 신호에 의해 경로를 식별한다.

Cooccurring mutated genes in a pathway/dense network module also mutually strengthen the mutation signals.

경로/고밀도 네트워크 모듈에서 돌연변이 유전자를 공동 발견하면 돌연변이 신호가 상호 강화됩니다.

The results in Fig.(6) clearly support that the mutation patterns are accurately captured in the case study on ovarian cancer by label propagation.

그림(6)의 결과는 라벨 확산에 의해 난소암에 대한 사례 연구에서 돌연변이 패턴이 정확하게 포착된다는 것을 명확히 뒷받침한다.

In drug repositioning, both molecular networks and drug–drug or phenotype similarity networks play important roles.

신약재창출에서 분자 네트워크와 (약물 – 약물) 또는 표현형 유사 네트워크 모두 중요한 역할을 합니다.

It has been repeatedly observed that genes associated with the same (or similar) diseases tend to lie in a dense module in the PPI network.

동일한 (또는 유사한) 질병과 관련된 유전자가 PPI 네트워크의 고밀도 모듈에 있는 경향이 반복적으로 관찰되었습니다.

This observation has motivated effective network-based methods to predict new disease genes.

이 관찰은 새로운 질병 유전자를 예측하기위한 효과적인 네트워크 기반 방법을 유도했습니다.

The analysis of gene modules in the PPI of similar diseases has also suggested associations between diseases and gene functions or pathways.

유사한 질병의 PPI에서 유전자 모듈의 분석은 또한 질병과 유전자 기능 또는 경로 사이의 연관을 제안했다.

When drug targets and disease genes are analyzed together in the PPI network, their proximities are useful for drug repositioning.

약물 표적(단백질) 및 질환 유전자가 PPI 네트워크에서 함께 분석될 때, 이들의 근접성은 신약재창출에 유용하다.

The methods compared in Figs.(2, 4) have different relative advantages and disadvantages.

그림 (2, 4)에서 비교된 방법은 서로 다른 상대적 장점과 단점을 가지고 있다.

The considerations involve a variety of key properties, including the performance of the methods, the interpretation of the results, the difficulty of implementation, the scalability to genome-wide analysis, and the characteristics of the training data.

고려 사항은 방법의 수행, 결과의 해석, 구현의 어려움, 유전체 전체 분석에 대한 확장성 및 훈련 데이터의 특성을 포함하는 다양한 주요 특성을 포함한다.

The appropriate choice of a network-based method for a particular analysis can be customized based on the information gained from these comparisons.

특정 분석을 위한 네트워크 기반 방법의 적절한 선택은 이러한 비교에서 얻은 정보를 기반으로 사용자 정의할 수 있습니다.

For example, drugs with more known targets can be repositioned by the network-based classification models, while drugs with no known targets in the candidates can be repositioned by the link prediction methods.

예를 들어, 알려진 대상에 대한 약물은 네트워크 기반 분류 모델에 의해 재배치할 수 있지만, 후보자에 알려진 표적(단백질)이 없는 약제는 링크 예측 방법으로 위치를 변경할 수 있습니다.

Depending on whether the analysis must be highly scalable to a huge network, simple graph connectivity measures or link prediction methods can be used.

분석을 대규모 네트워크로 확장해야 하는지 여부에 따라 간단한 그래프 연결 방법 또는 링크 예측 방법을 사용할 수 있습니다.

In the application of network-based analysis, there are also several practical issues and limitations.

네트워크 기반 분석의 적용에는 몇 가지 실질적인 문제와 한계가 있습니다.

1. Molecular networks often contain biased information.

1. 분자 네트워크는 종종 편향된 정보를 포함합니다.

Well studied genes tend to have more connections in the PPI network, and they are also targets of more drugs and are associated with more disease phenotypes.

잘 연구된 유전자들은 PPI네트워크에서 더 많은 연관성을 가지는 경향이 있고 더 많은 약물의 표적(단백질)이 되며 더 많은 질병 표현형과 연관되어 있다.

Typically, it is important to exercise normalizations and repeat the experiments on randomized networks to assess the statistical significance of the results.

일반적으로 정규화를 실행하고 무작위 네트워크에서 실험을 반복하여 결과의 통계적 중요성을 평가하는 것이 중요합니다.

The biases also prevent the prediction of de novo disease genes or target genes if the gene has no association with known diseases or is not a target of any drug.

편견은 또한 유전자가 알려진 질병과 관련이 없거나, 어떤 약물의 표적(단백질)이 아닌 경우, 새로운 질병 유전자 또는 표적(단백질) 유전자의 예측을 예방합니다.

2. The empirical results of network-based methods rely on tuning parameters.

2. 네트워크 기반 방법의 경험적 결과는 튜닝 매개 변수에 의존합니다.

The parameters often balance how much belief is imposed on the network topologies.

매개 변수는 종종 네트워크 위상에 적용되는 신뢰도의 균형을 조정합니다.

When excessive weights are assigned to the network topology, there will be an "over-smoothing" effect such that nearly uniform scores are expected among the genes in even large and sparse neighborhoods.

과도한 가중치가 네트워크 위상에 할당되면 크고 좁은 지역에서도 거의 균일한 점수가 예상되는 " over-smoothing"효과가 발생할 것입니다.

Thus, a proper procedure for determining the appropriate (optimal) parameters is critical, for example, by applying cross-validation and wet-lab validation.

따라서 적절한 (최적의) 매개 변수를 결정하기위한 적절한 절차가 중요합니다. 예를 들어 상호 유효성 확인 및 습식 연구소 유효성 검사를 적용하는 것이 중요합니다.

3. Commonly, a molecular network describes a general relation, such as protein–protein physical interaction or functional linkage.

3. 일반적으로, 분자 네트워크는 (단백질 – 단백질) 물리적 상호 작용 또는 기능적 결합과 같은 일반적인 관계를 기술한다.

In some cases, the relations can be either positive or negative, e.g., gene co expression.

일부 경우, 관계는 양성 또는 음성 일 수 있는데,

예를 들어, 유전자 공동 발현 (gene co expression)가 있다.

A practical approach is to apply a signed graph Laplacian.

실용적인 접근법은 서명 된 Laplacian 그래프를 적용하는 것입니다.

The models applied with a signed graph Laplacian can be solved in a manner similar to those with the normal graph Laplacian by the same algorithms.

서명 된 Laplacian 그래프로 적용된 모델은 동일한 알고리즘으로 일반 Laplacian 그래프와 비슷한 방식으로 해결할 수 있습니다.

Finally, this article targets the scope of precision oncology, including steps for understanding cancer mechanisms, finding targets and repositioning drugs, while previous survey studies have focused on detecting cancer-driven aberrations and understanding of the aberrations in molecular networks/pathways.

마지막으로 이 기사에서는 암 종양을 이해하고 대상을 찾고 약물을 재배치시키는 단계를 포함하여, 정밀 종양학의 범위를 대상으로 합니다.

이전의 조사 연구는 암에 의한 수차를 발견하고 분자 네트워크 / 경로의 수차를 이해하는 데 초점을 맞추었습니다.

This article also surveys several categories of algorithms, including model-based integration and preprocessing integration with machine learning methods, while previous reviews primarily surveyed the methods in one of the three categories, namely, post-analysis integration of oncogenic alterations in networks.

또한 이 기사에서는 모델 기반 통합 및 기계 학습 방법과의 사전 처리 통합을 비롯하여 알고리즘의 여러 범주를 조사합니다.

이전의 리뷰는 세 가지 카테고리 중 하나의 방법, 즉 네트워크에서 발암성 변화의 사후 분석 통합을 주로 조사했다.

Thus, this article offers a different scope and a more comprehensive survey of computational methods.

따라서 이 기사에서는 계산 방법에 대해 다른 범위와 포괄적인 설문 조사를 제공합니다.

FUTURE DIRECTIONS

Several challenges remain in the application of network-based analysis in precision oncology.

These challenges concern the data quality, deployment for research or clinical use, and scalability of network analysis.

To precisely model the molecular interactions and drug–target relations, networks of better quality are required.

It is known that most molecular networks and drug–target databases are incomplete and biased towards well-studied proteins/genes.

Thus, continuing effort on the improvement of the networks with additional experimental data is important.

In addition, network modeling with higher resolution is also crucial to model complex molecular functions at higher precisions.

For example, proteins are present in the isoforms of genes, and thus isoform–isoform interactions are the true interactions to model in a network; mutations or other structure variations of a protein can also change the protein–protein binding or drug–protein docking in a specific tumor.

Furthermore, even within each tumor, heterogeneous cell populations exist, and the drug targets and molecular interactions could be different for each cell population if measured by single-cell RNA sequencing.

To partially address this issue, several computational methods for quality control of PPI screening have been proposed to reduce the number of falsepositive and false-negative PPIs due to spurious errors and systematic biases from the high-throughput techniques.

Currently, it is still impossible to construct these more accurate networks at a large scale due to the limitation of the current highthroughput experimental methods for measurement of molecular interactions or drug screening.

향후 방향

정밀 종양학에서 네트워크 기반 분석의 적용에는 몇 가지 과제가 남아 있습니다.

이러한 과제는 데이터 품질, 연구 또는 임상 사용을 위한 배포 및 네트워크 분석의 확장성에 관련됩니다.

분자 상호 작용과 (약물 – 표적(단백질)) 관계를 정확하게 모델링하려면 더 나은 품질의 네트워크가 필요합니다.

대부분의 분자 네트워크와 (약물 – 표적(단백질)) 데이터베이스는 불완전하며 잘 연구된 (단백질 / 유전자)에 편향 되어있는 것으로 알려져 있다.

따라서 추가적인 실험 데이터로 네트워크 개선에 대한 지속적인 노력이 중요합니다.

또한 고해상도의 네트워크 모델링은 더 높은 정밀도로 복잡한 분자 기능을 모델링하는 데에도 중요합니다.

예를 들어, 단백질은 유전자의 동형 단백질들에 존재하므로 (동형 단백질 – 동형 단백질) 상호 작용은 네트워크에서 모델링하는 진정한 상호 작용입니다.

돌연변이 또는 단백질의 다른 구조 변형은 또한 특정 종양에서 단백질 - 단백질 결합 또는 약물 - 단백질 도킹을 변화시킬 수 있다.

또한 각 종양 내에서도 이질적인 세포 집단이 존재하며,

단일 세포 RNA 염기 서열 분석으로 측정된 경우 약물 표적(단백질) 및 분자 상호 작용은 각 세포 집단마다 다를 수 있습니다.

이 문제를 부분적으로 다루기 위해, 거짓 오류 및 체계적 편향으로 인한 거짓 긍정 및 거짓 부정 PPI의 수를 줄이기 위해 PPI 스크리닝의 품질 관리를위한 몇 가지 계산 방법이 제안되었습니다. 그리고 이것은 높은 처리량의 기술에서 비롯되었습니다.

현재, 분자 상호 작용 또는 약물 스크리닝 측정을 위한 고차원 실험 방법의 한계로 인해, 이러한 더 정확한 네트워크를 대규모로 구축하는 것은 여전히 불가능합니다.

While many network-based methods have been developed to support precision oncology, the implementations of the methods are independent, with non-standardized tools that are never easily accessible as a useful collection to oncologists for research or clinical use.

정밀 종양학을 지원하기 위해 많은 네트워크 기반 방법이 개발되었지만, 방법의 구현은 독립적이며, 연구나 임상 사용을 위해 종양 전문의에게 유용한 콜렉션으로, 쉽게 접근할 수 없는 표준화되지 않은 도구를 사용합니다.

Thus, there is a strong need to develop a software platform that integrates standardized biomedical, biological network data, and analytic software components to support comprehensive network-based analysis of patient genomic data and drug repositioning for precision oncology.

따라서, (표준화된 바이오 메디컬, 생물학적 네트워크 데이터, 정밀 종양학을 위해 환자 유전체 데이터 및 약물 위치 변경에 대한 포괄적인 네트워크 기반 분석을 지원하는 분석 소프트웨어 구성 요소)를 통합하는 소프트웨어 플랫폼을 개발할 필요가 있다.

.

This platform should be based on a sophisticated system design to meet oncologists' requirements and support customization of the analysis pipeline.

이 플랫폼은 종양 전문의의 요구 사항을 충족시키고 분석 파이프라인의 사용자 정의를 지원하는 정교한 시스템 설계를 기반으로 해야 합니다.

The concept of part of such a platform was proposed in the paper5 as an integrative network-based infrastructure to identify new druggable targets and repositionable drugs through the targeting of significantly mutated genes identified in human cancer genomes.

이러한 플랫폼의 일부 개념은 인간 암 유전체에서 확인된 돌연변이된 유전자를 표적(단백질)으로 하여 새로운 위험한 표적(단백질) 및 재배치 가능한 약물을 확인하기위한 통합 네트워크 기반 인프라로서 문서5에서 제안되었다.

In the future, the existing tools can be reimplemented as apps on a platform such as Cytoscape or another software environment similar to GALAXY for NGS data analysis to facilitate the development and deployment of the software system for precision oncology.

앞으로 Cytoscape와 같은 플랫폼이나 NGS 데이터 분석을 위한 GALAXY와 유사한 다른 소프트웨어 환경에서, 기존 도구를 다시 구현하여 정밀 종양학을 위한 소프트웨어 시스템의 개발 및 배포를 용이하게 할 수 있습니다.

Finally, scalability is always an issue in network-based analysis since it is common to model millions of genomic features, hundreds of thousands of drugs and tens of thousands of phenotypes in a very large network.

마지막으로, 네트워크 기반 분석에서는 확장성이 항상 중요한 문제입니다. 대용량 네트워크에서 수백만 개의 유전체 기능, 수십만 개의 약물 및 수천 가지 표현형을 모델링하는 것이 일반적이기 때문에 네트워크 기반 분석에서는 항상 문제가 됩니다.

For example, in an isoform–isoform interaction network, hundreds of thousands of nodes are contained in a single graph that cannot be loaded onto a computer with less than 100 GB of memory.

예를 들어, (동형 단백질 – 동형 단백질) 상호 작용 네트워크에서 100GB 미만의 메모리로 컴퓨터에 로드 할 수 없는 수십만 개의 노드가 하나의 그래프에 포함됩니다.

Such big-data analysis will require more scalable algorithms and efficient computing platforms. For example, the standard label propagation can be applied to low-rank approximations of big graphs, enabling work with networks of millions of nodes.

이러한 대용량 데이터 분석에는 보다 확장성이 뛰어난 알고리즘과 효율적인 컴퓨팅 플랫폼이 필요합니다. 예를 들어 표준 레이블 전파는 큰 그래프의 낮은 순위 근사치에 적용되어 수백만 노드의 네트워크로 작업할 수 있습니다.

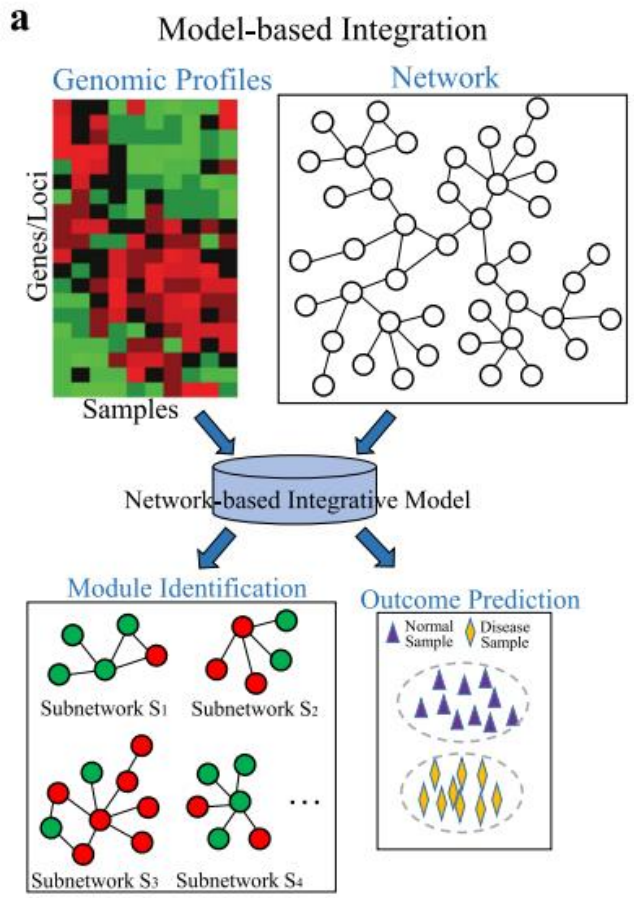
Parallel implementations of the network-analysis methods, especially the optimization algorithms in those model-based approaches, are also necessary.

네트워크 분석 방법의 병렬 구현, 특히 이러한 모델 기반 접근 방식의 최적화 알고리즘도 필요합니다.

시퀀싱 기술 : 게놈 염기서열 해독 기술인 시퀀싱(sequencing)기술

➔ http://www.bioin.or.kr/fileDown.do?seq=27626

FIG 2-A



Objectives: subnetwork module detection and cancer phenotype prediction
Inputs: genomic profiles, a molecular network and phenotype labels
Outputs: cancer subnetwork modules and phenotype predictions
Advantages: one unified framework, global optimization strategy and better outcome prediction
Limitation: more difficult optimization and lower scalability

목표 : 서브네트워크 모듈의 탐지와 암에 대한 표현형(암의 형태와 특징에 대한 예측) 예측

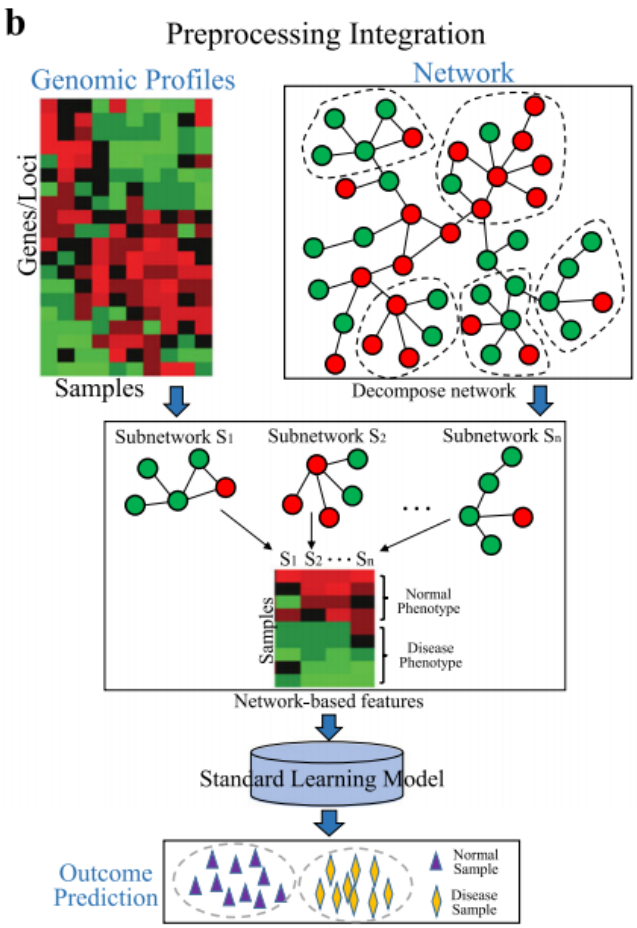
인풋 : 유전체 프로필, 분자 네트워크, 표현형 라벨

아웃풋 : 암 서브네트워크 모듈과 표현형 예측

이점 : 하나로 결합된 프레임워크, 세계적으로 최적화된 전략,
더 나은 예측 결과

한계점 : 최적화가 더 어렵고, 확장성이 떨어진다.

FIG 2-B



Objectives: detection of network-based feature and cancer phenotype prediction
Inputs: network-based features and phenotype labels
Outputs: phenotype predictions
Advantages: flexible and customizable subnetwork features
Limitation: not optimal network-based features for the prediction

목표 : 네트워크 기반 구성물의 탐지와 암에 대한 표현형 예측

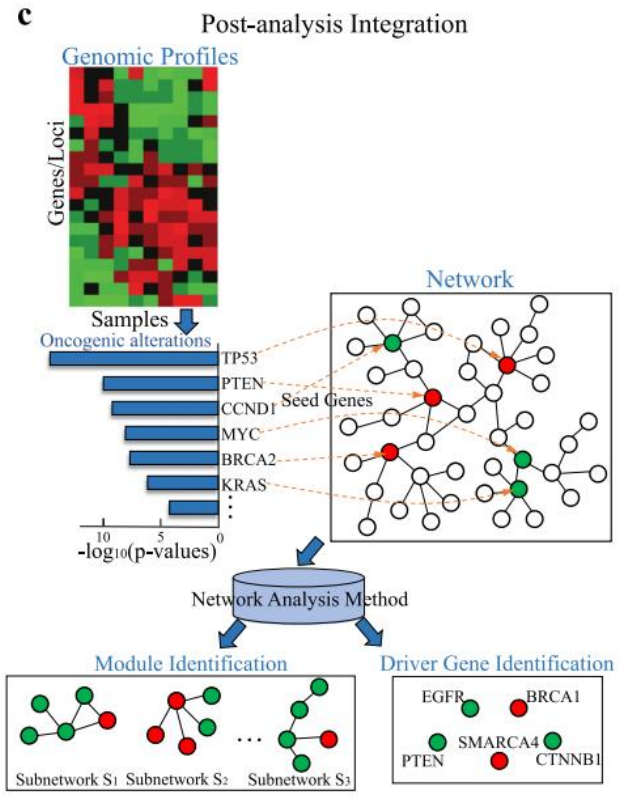
인풋 : 네트워크 기반 구성물(유전체 프로필 + 부패 네트워크)과 표현형(유전자와 환경의 영향에 의해 형성된 생물의 형질) 라벨

아웃풋 : 표현형 예측

이점 : 융통성있고 맞춤형이 가능한 서브네트워크 구성물

한계점 : 예측을 위한 최선의 네트워크 기반 구성물이 아니다.

FIG 2-C



Objectives: detection of subnetwork modules and cancer driver genes
Input: oncogenic alterations detected from genomic profiles and a molecular network
Output: cancer driver genes and cancer subnetwork modules
Advantage: highly informative of cancer mechanisms in the network
Limitation: relying on accurate detection of oncogenic alterations

목표 : 서브네트워크 모듈의 탐지와 암 유발 유전자의 탐지

인풋 : 유전체 프로필과 분자 네트워크에서 발견된 발암화(발암화 및 기타 분자들의 위치가 네트워크에 표시됨)

아웃풋 : 암 유발 유전자와 암 서브네트워크 모듈

이점 : 네트워크에서의 암 메커니즘에 대해 매우 유익한 정보

한계점 : 발암화에 대한 정확한 탐지에 의존한다.