

진화연산 기반 계층적 하이퍼네트워크 모델에 의한 암 특이적 microRNA-mRNA 상호작용 탐색

(Exploring Cancer-Specific microRNA-mRNA Interactions by Evolutionary Layered Hypernetwork Models)

김수진[†] 하정우^{**}
(Soo-Jin Kim) (Jung-Woo Ha)

장병탁^{***}
(Byoung-Tak Zhang)

요약 microRNA (miRNA)와 mRNA 조절 상호작용 탐색은 다양한 생물학적 현상에 있어 새로운 시야를 제공해 줄 수 있다. 최근 생물학적 프로세스에서 miRNA는 유전자 발현을 제어하고 세포를 기능적으로 조절하는 중요한 역할을 하는 요소로 밝혀졌다. 이에 복잡한 생물학 시스템에서 miRNA의 기능적 활동을 이해하기 위해서는 miRNA와 mRNA간 상호작용 분석은 필수적이다. 그러나 아직까지 복잡한 miRNA와 mRNA간 상호작용 관계를 추론하는

것은 어려운 문제이기 때문에 많은 연구자들이 실험적, 전산학적 접근 방법을 제안하며 활발한 연구를 진행하고 있다. 본 논문에서는 이종의 발현 데이터로부터 기능적으로 상호작용하는 miRNA-mRNA 조합을 탐색하기 위한 진화연산 기반의 새로운 하이퍼네트워크 모델을 제안한다. 이에 실험결과로 제안하는 방법을 인간 암 관련 miRNA와 mRNA 발현 데이터에 적용하여 암 특이적 miRNA-mRNA 상호작용 집합을 탐색하고 발견한 miRNA-mRNA 상호작용 관계가 생물학적으로 유의함을 제시한다.

키워드 : 기계학습, 진화연산, 생물정보학, miRNA-mRNA 상호작용

Abstract Exploring microRNA (miRNA) and mRNA regulatory interactions may give new insights into diverse biological phenomena. Recently, miRNAs have been discovered as important regulators that play a major role in various cellular processes. Therefore, it is essential to identify functional interactions between miRNAs and mRNAs for understanding the context-dependent activities of miRNAs in complex biological systems. While elucidating complex miRNA-mRNA interactions has been studied with experimental and computational approaches, it is still difficult to infer miRNA-mRNA regulatory modules. Here we present a novel method, termed layered hypernetworks (LHNs), for identifying functional miRNA-mRNA interactions from heterogeneous expression data. In experiments, we apply the LHN model to miRNA and mRNA expression profiles on multiple cancers. The proposed method identifies cancer-specific miRNA-mRNA interactions. We show the biological significance of the discovered miRNA-mRNA interactions.

Key words : Machine Learning, Evolutionary Learning, Bioinformatics, miRNA-mRNA Interactions

- 이 논문은 지식경제부 한국산업기술평가관리원(KI002138, 10035348), 교육과학기술부 한국연구재단(No. 2010-0017734, No. 2010-0018950), 질병관리본부 학술연구용역사업(2009-S6-B-001), BK21-IT 사업 및 컴퓨터연구소(ICT)에 의해 지원되었음
- 이 논문은 2010 한국컴퓨터종합학술대회에서 '진화연산 기반 계층적 하이퍼네트워크 모델에 의한 암 특이적 microRNA-mRNA 상호작용 탐색'의 제목으로 발표된 논문을 확장한 것임

[†] 학생회원 : 서울대학교 생물정보학 협동과정
sjkim@bi.snu.ac.kr

^{**} 학생회원 : 서울대학교 컴퓨터공학부
jwha@bi.snu.ac.kr

^{***} 종신회원 : 서울대학교 컴퓨터공학부 교수
btzhang@bi.snu.ac.kr

논문접수 : 2010년 8월 9일
심사완료 : 2010년 8월 25일

Copyright©2010 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 컴퓨팅의 실제 및 레터 제16권 제10호(2010.10)

1. 서론

최근 microRNAs(miRNAs)는 세포분화, 증식, 성장, 사멸 등 다양한 세포적 프로세스에 있어 주요 역할을 하는 중요한 조절자로 밝혀졌다. 특히 여러 유형의 종양에 수반하여 주요 유전자의 발현을 제어하고 세포를 기능적으로 조절함으로써 암의 발생과 진행을 유도하는데 있어 결정적인 역할을 한다고 많은 연구를 통해 보고되고 있다. 따라서 암 발생 메커니즘을 포함한 복잡한 생물학 시스템에서 miRNA의 기능과 활동 기전을 이해하기 위해서는 miRNA와 mRNA간 기능적 상호작용을 찾아내는 것은 필수적이라 할 수 있다. 그러나 miRNA와 mRNA간 정확한 조절 메커니즘을 설명하는데 아직 많은 어려움이 존재한다. 이 분야에 있어 이전 연구들은 주로 유전체 수준에서 miRNA와 목표 유전자의 예측에 초점을 맞추어 많은 전산학적 방법들이 제안되었다. 다

양한 생물학적 프로세스에서는 유전자 조절 네트워크의 중요한 구성요소로서 miRNA와 mRNA는 상호작용적으로 함께 작용한다. 즉 miRNA와 mRNA 상호작용 탐색은 다른 생리학적 조건에서 miRNA와 mRNA의 조합적 효과를 찾아내는데 필수적인 첫 단계라 할 수 있다. 이런 쟁점에 초점을 맞추어 Yoon and DeMicheli은 weighted bipartite graph 모델을 이용하여 유전자 조절에 수반되는 miRNA와 목표 유전자의 그룹을 찾는 전산학적 방법을 제안하였다[1]. 그러나 오직 서열 정보만을 기반으로 예측하기 때문에 복잡한 miRNA-mRNA 조절 집합을 결정짓기에는 충분하지 않다. Huang et al.은 miRNA-mRNA 상호작용을 탐색하기 위해 베이지안 네트워크 매개변수 학습을 응용한 확률적 프래임워크를 제안하였다[2]. 또한 Joung et al.은 공진화적 학습 전략에 기반한 바이클러스터링 기법을 사용하여 miRNA-mRNA 모델을 발견하였다[3]. 이 두 방법은 miRNA-mRNA 조합을 찾기 위해 여러 정보를 통합하여 모델을 구축하였지만 miRNA와 mRNA 상호작용의 다양성과 복잡성 때문에 여전히 기능적인 miRNA-mRNA 집합을 추론하는 것은 어려운 문제로 남아 있다.

본 연구에서 우리는 암 관련 발현 데이터로부터 암 특이적 miRNA-mRNA 상호작용 조합을 탐색하기 위한 새로운 방법을 제안한다[4]. 제안하는 방법은 새롭게 수정된 하이퍼네트워크 모델로 계층적 하이퍼네트워크(layered hypernetworks: LHNs)라고 한다. 제안하는 모델은 기존 하이퍼네트워크에서 진보된 모델로 학습 과정에 진화 연산이 내포되어 있는 계층적 구조를 가지고 있다. 한 개체 이상으로 구성된 데이터가 주어졌을 때 계층적 하이퍼네트워크는 두개의 층으로 구성된다. 하부 층의 하이퍼네트워크는 데이터의 각 개체 특성을 반영하여 구축되고, 상부 층의 하이퍼네트워크는 하부 층 하이퍼네트워크의 학습 결과로부터 샘플링에 의해 주어진 개체간의 관계를 분석한다. 이런 특성은 miRNA와 mRNA 상호작용 집합 발견과 같은 복잡하고 이중적인 문제를 분석하는데 유용하다. 계층적 하이퍼네트워크 프래임워크에서 거대한 검색공간에서 최적의 고차원 조합 집합을 찾기 위해서 학습 과정은 진화 연산에 의해 수행된다.

본 논문에서는 제안한 계층적 하이퍼네트워크 모델을 인간 암 관련 miRNA와 mRNA 발현 데이터[5,6]에 적용하여 높은 분류 정확도로 암 특이적 miRNA-mRNA 상호작용 집합을 탐색한다. 이에 실험 결과로 본 제안하는 모델은 지지벡터머신(support vector machines)에 경쟁력 있는 성능을 보이고, 결정트리(decision trees), 베이지안 네트워크(Bayesian networks), 나이브 베이즈(naïve Bayes), 랜덤 포레스트(random forests)에 비해서는 우월한 분류 성능을 보여주었다. 또한 발견된

miRNA-mRNA 조합들은 암에서 기능적으로 상호작용하는 집합으로 문헌 검색 등으로 생물학적 유의성을 검증하였다.

2. 하이퍼네트워크 모델

하이퍼네트워크는 다수 인자들 간의 고차원 연관 관계를 학습 할 수 있는 확률적 그래프 모델이다[7]. 이는 기존의 일반 그래프와는 달리 2개 이상의 정점을 동시에 연결 가능한 하이퍼에지(hyperedge)로 모델로서 하이퍼네트워크에서 정점은 데이터를 구성하는 인자(attribute)를 의미하며 하이퍼에지는 가중치가 할당된 인자들간의 고차원적 조합을 의미한다. 수식적으로 하이퍼네트워크 H 는 $H = (V, E, W)$ 로 각각 V, E, W 는 정점, 하이퍼에지, 가중치들의 집합으로 정의된다. k -차수(cardinality)의 하이퍼에지를 k -하이퍼에지라 하며 모든 하이퍼에지가 k -차수 하이퍼에지들로만 구성된 하이퍼네트워크를 k -하이퍼네트워크라 한다. 하이퍼네트워크는 이론적으로 데이터 $D = \{x^{(n)}\}_{n=1}^N$ 를 저장하기 위한 확률적 메모리 모델로 간주될 수 있기 때문에 k -하이퍼네트워크의 에너지는 다음과 같은 식으로 정의될 수 있다.

$$\varepsilon(x^{(n)}; W) = - \sum_{i=1}^{|E|} w_i^{(k)} I(x^{(n)}, E_i), \quad (1)$$

위의 식에서 $w_i^{(k)}$ 는 i 번째 k -차수 하이퍼에지 E_i 의 가중치이고, $x^{(n)}$ 는 하이퍼네트워크에 저장된 n 번째 데이터 패턴이며, $I(x^{(n)}, E_i)$ 는 하이퍼에지 E_i 에서 $x^{(n)}$ 의 요소 조합을 나타낸다. 이 때, 하이퍼네트워크로부터 데이터가 생성될 확률은 Gibbs 분포 가정 하에 다음과 같이 주어진다.

$$P(x^{(n)} | W) = \frac{1}{Z(W)} \exp\{-\varepsilon(x^{(n)}; W)\}, \quad (2)$$

$Z(W)$ 는 부분 함수로 $Z(W)$ 는 아래와 같은 식으로 정의된다.

$$Z(W) = \sum_{x^{(n)} \in D} \exp\left\{ \sum_{i=1}^{|E|} w_i^{(k)} I(x^{(n)}, E_i) \right\}, \quad (3)$$

즉, 하이퍼네트워크는 학습을 통해 랜덤 변수 조합의 확률 분포를 추정하는 모델이며, 모델 학습은 진화기법을 기반으로 한 구조학습과 파라미터 학습에 의해 진행된다.

3. 계층적 하이퍼네트워크 모델

계층적 하이퍼네트워크 모델은 기존 하이퍼네트워크를 발전시킨 모델로 두 개의 층으로 구성된 계층적 구조의 하이퍼네트워크 모델이다. 하부 층은 하나 이상의 모달리티로 구성되어 있는 데이터의 특성을 반영하는

모달리티 의존적인 하이퍼네트워크로 구축되고, 상부 층은 각 모달리티 특성을 반영하고 있는 하부 층의 하이퍼네트워크 결과로부터 모달리티간의 관계를 분석하고 통합한다. 기존의 하이퍼네트워크 모델과 마찬가지로 수식적으로 상부 층의 하이퍼네트워크는 파라미터로 가중치 벡터가 주어질 때 에너지 함수가 정의된다. 이에 두 개의 모달리티를 가지고 있는 주어진 데이터 셋이 D일 때, $D = \{x^{(n)}\}_{n=1}^N = \{(m^1, m^2)^{(n)}\}_{n=1}^N$, 하부 층의 k-하이퍼네트워크로부터 생성된 상부 층의 하이퍼네트워크 에너지 함수 $\varepsilon(x^{(n)}; W)$ 는 다음과 같이 정의된다.

$$\varepsilon(x^{(n)}; W) = - \sum_{i=1}^{|E|} W_i^{(k)} I \{ (m^1, m^2)^{(n)}, E_i \}, \quad (4)$$

m^1 과 m^2 은 데이터 샘플 x를 구성하고 있는 각 모달리티 변수의 벡터이다. 식 (2)와 같이 두 모달리티의 n 번째 데이터 생성의 확률 $P(x^{(n)}|W)$ 는 아래와 같다.

$$P(x^{(n)}|W) = \frac{1}{Z(W)} \exp[-\varepsilon \{ (m^1, m^2)^{(n)}; W \}] \quad (5)$$

4. 계층적 하이퍼네트워크의 진화적 학습 알고리즘

4.1 하부층 진화적 하이퍼네트워크의 학습

하부 층에서의 하이퍼네트워크는 데이터를 구성하는 모달리티 각각의 인자들의 조합 공간을 확률적으로 표현함으로 학습은 매우 넓은 문제공간을 검색하는 과정이라 할 수 있다. 이를 현실적으로 가능하게 하기 위해 랜덤 샘플링 기반의 진화적 학습방법을 이용하여 하이퍼에지의 적합도, 즉 하이퍼에지의 가중치를 학습한다. 이는 예제 데이터 셋에서 하이퍼에지를 구성하는 인자 값을 비교하여 정확하게 예측된 예제의 개수와 정확하게 예측되지 않은 예제의 개수를 기반으로 업데이트 된다. 가중치 w 는 두 요소의 선형 합으로 다음과 같이 계산된다.

$$w = \alpha \times (\#_c - (\#_w)^2) + (1 - \alpha) \times (\#_c + \#_w) \quad (6)$$

$\#_c$ 와 $\#_w$ 는 정확하게 예측된 훈련예제와 부정확하게 예측된 훈련예제의 개수이다. 식 (6)에서 첫 번째 항은 정확도 요소로 정확하게 예측된 예제가 많고 부정확하게 예측된 예제보다 적을 때 점수가 높아진다. 두 번째 항은 데이터에서 하이퍼에지의 커버리지를 나타내는 것으로 더 많은 훈련 예제에 일치하면 점수가 높아진다. 또 하이퍼네트워크의 다양성을 향상시키기 위한 전략으로 가중치에 기반하여 하이퍼에지를 교체하는 정책을 도입한다. 이전 연구[8]에서는 특정 값으로 교체되는 하이퍼에지의 비율을 고정했었는데 본 연구에서는 하부 층의 하이퍼네트워크를 진화시키기 위한 하이퍼에지 교체 비율을 유연하게 조정할 수 있는 함수 $R(t)$ 를 아래와 같이 정의하였다.

$$R(t) = (M - m) \times \exp\left(\frac{1-t}{c}\right) + m, \quad (7)$$

M 과 m 은 교체 비율의 최대값과 최소값을 나타내고 0과 1사이의 범위를 가지며, t 는 진화 과정에서의 세대 수이며, c 는 임의적 상수이다. c 를 조정함으로써 수렴 속도를 조절할 수 있다.

4.2 상부층 진화적 하이퍼네트워크의 학습

상부 층의 하이퍼네트워크는 모달리티 간의 관계를 표현한다. 상부 층 하이퍼네트워크의 하이퍼에지는 하부 층의 각 하이퍼네트워크의 결과로 생산된 하이퍼에지 통합에 의해 생성되므로 각 모달리티 간의 상호작용 관계를 분석할 수 있다. 상부 층 하이퍼네트워크에서도 역시 하부 층 하이퍼네트워크로부터 산출된 하이퍼에지의 모든 조합을 학습하기에는 불가능하기 때문에 랜덤 선택에 기반하여 하이퍼에지들을 학습하여 통합한다.

4.3 하이퍼에지 분석을 위한 유의성 점수 측정

모달리티간 유의한 관계를 찾기 위해 우리는 본 연구에서 메트릭을 정의한다. 제안하는 메트릭은 유의성 점수 S 라 하며 이는 각 모달리티의 인자가 하이퍼에지에 동시에 존재하는 빈도에 기반하여 계산한다. 즉, 모든 하이퍼에지에서 두 인자의 각 쌍이 동시에 나타나는 빈도수를 기반으로 유의성 점수 s_{ij} 는 다음과 같다.

$$s_{ij} = \sum_{n=1}^N w_{ij}, w_{ij} = \begin{cases} 1 & ((i, j) \in E_n) \\ 0 & (otherwise) \end{cases}, \quad (8)$$

위와 같이 유의성 점수는 i 번째 인자와 j 번째 인자가 정점으로 동시에 나타나는 하이퍼에지의 수로 정의된다. 따라서 높은 유의성 점수를 가진 인자의 쌍은 유의미하게 상호작용하는 모듈로 간주될 수 있다.

5. 실험결과

5.1 데이터 전처리 및 실험 파라미터 설정

본 연구에서는 인간 암 관련 miRNA와 mRNA의 발현 프로파일 데이터를 사용하여 암 특이적 miRNA-mRNA 상호작용 조합을 탐색하였다[5,6]. 실험 데이터는 21개의 정상 조직과 68개의 암 조직으로부터 151개 miRNA와 10,262개 mRNA의 발현 정도를 측정된 것이다. 실험을 위해 각 샘플의 유전자 평균값을 기반으로 발현 수준을 표준화시키고, 그 표준화된 발현 값을 각 유전자의 샘플 평균값을 기준으로 miRNA와 mRNA의

표 1 계층적 하이퍼네트워크 실험 파라미터 설정

파라미터	값
차수 (#miRNA, #mRNA)	(3, 3)
교체비율 (최소값, 최대값)	(0.1, 0.9)
#샘플링비율	80
#통합비율	10

발현 수준을 0 또는 1로 이산화하였다. 또한 생물학적으로 더 유의성 있는 결과를 얻기 위해 전체 mRNA 목록에서 [9]에 따라 mRNA를 추출하여 실험을 수행하였다. 실험에 사용된 파라미터는 표 1과 같이 설정하였다.

5.2 분류 성능

표 2는 제안하는 알고리즘과 다른 기계학습 방법들과의 분류 성능을 비교한 결과이다. 다른 기계학습 방법으로는 WEKA에서 제공되는 지지벡터머신(support vector machines), J48 결정트리, 베이저안 네트워크(Bayesian networks), 나이브 베이즈(naïve Bayes), 랜덤 포레스트(random forests)를 사용하였으며 10-fold-cross validation을 이용하여 10번 반복 수행, 분류 정확도의 평균을 구하였다. 그 결과 계층적 하이퍼네트워크는 95.30% 분류 정확도로 결정트리, 나이브 베이즈, 랜덤 포레스트, 베이저안 네트워크보다 높은 성능을 보였고 지지벡터머신에도 비할만한 성능을 보여주었다. 그러나 계층적 하이퍼네트워크 모델은 분류뿐만 아니라 직접적으로 해석 가능한 결과를 제공할 수 있으므로 다수 인자간 상호작용 관계 분석이 필수적인 복잡한 생물학 문제를 해결하는데 있어 유용하다 할 수 있다.

표 2 다른 기계학습 방법들과 성능 비교 결과

분류기법	정확도(표준편차)
지지벡터머신	97.30 (0.58)
계층적하이퍼네트워크	95.39 (0.98)
베이저안네트워크(#부모=3)	92.92 (1.59)
랜덤포레스트(#트리=50)	90.90 (0.98)
나이브베이즈	89.21 (0.79)
J48결정트리	84.41 (2.70)

5.3 암 특이적 miRNA-mRNA 상호작용 탐색

표 3은 계층적 하이퍼네트워크 모델에 의해 탐색된 높은 점수의 miRNA-mRNA 상호작용 조합을 보여주고 있다. 제안한 매트릭(유의성 점수 S)을 적용하여 통합된 하이퍼에지에서 동시에 나타나는 인자의 빈도수를 기반으로 miRNA와 mRNA의 유의한 관계를 추출하였다. 상호작용 집합을 구성하고 있는 각 miRNA와 mRNA 간 유의성 점수를 계산하고 평균을 구하였다. 또, 표 3에서 보여주고 있는 miRNA와 mRNA는 이전 연구들

을 통해 암과 관련되어 있는 miRNA와 mRNA로 다수 밝혀졌다. hsa-miR-21은 유방암, 난소암, 대장암, 간암 등 다양한 인간 암에서 높게 발현되어 PDCD4, PTEN, TPM1과 같은 암 억제 유전자의 발현을 억제하여 종양 유발을 유도한다[10]. 또 hsa-miR-29a는 폐암에서 과잉 발현되어 암 억제 유전자의 발현을 억제하고 암세포의 분화, 전이에 있어서도 기능적인 조절 역할을 한다[11]. 또한, hsa-miR-154 역시 만성 림프성 백혈병에서 유의하게 높은 수준으로 발현되고[12], hsa-miR-184은 신경교종 진행의 억제 기능을 한다고 알려져 있다[13]. 더욱이 표 3에서 보여주는 mRNA의 대부분이 다양한 암 이전에 기능적으로 연관되어 있음이 여러 연구를 통해 보고되었으며, 또한 같은 집합 내에 있는 mRNA은 생물학적 시스템에서 비슷한 생물학적 기능을 공유하고 있음을 보여준다.

5.4 발견된 miRNA-mRNA 상호작용 조합의 생물학적 유의성

우리는 제안한 모델에 의해 탐색된 miRNA-mRNA 상호작용 조합을 검증하기 위해 유전자 온톨로지(Gene Ontology)를 사용하여 같은 집합 내에 있는 mRNA의 기능적 응집성을 분석하였다. 찾아진 miRNA-mRNA 상호작용 조합이 잠재적인 기능을 내포하고 있다면 조합을 구성하고 있는 mRNA들은 생물학적으로 유의미한 기능적 관련성이 반영되어 있을 것이다. GOstat[14]을 이용하여 집합 I에 대해 통계적으로 유의한 GO term을 검증하였다. q -value<0.01에 해당하는 유전자 온톨로지 분석 결과, 전체적으로 집합 I을 구성하고 있는 mRNA는 세포 자살, 죽음, 분화, 발달 등 세포 주기 메커니즘 관련한 특정 기능의 카테고리에 속해 있었다. 이는 암 진행이나 발달에 매우 밀접한 관련이 있는 기능으로 제시하는 결과는 탐색된 상호작용 조합이 세포적 프로세스에서 암 특이적인 기능적 역할을 내포하고 있음을 암시한다 할 수 있다.

또, 그림 1은 상호작용 집합 I에서 실험적인 증거를 바탕으로 생물학적으로 더 의미 있는 miRNA-mRNA 상호작용 조합의 부분 집합을 추출한 결과이다. 위에서 언급한 바와 같이 hsa-miR-21과 hsa-miR-29a는 세포 성장과 발달에 많은 영향을 주는 조절인자로 인간 악성

표 3 계층적 하이퍼네트워크에 의해 탐색된 높은 유의성 점수의 miRNA-mRNA 상호작용 집합

No.	miRNAs	mRNAs (평균점수)
I	hsa-miR-21, has-miR-29a	SFERS2 (959), NFIB (956), CRYAB (917), ALDH1A1 (916), CDKN1B (901), SART1 (864.5), HSPA1A (861.5), TGFBR2 (851.5), PPAP2B (837), PPP2CB (820), ARMCX1 (815.5), FLNB (814), HSP90B1 (797.5), GADD45G (795.5), MYLK (766), CTNBN1 (758.5), EPS8 (755), HNRPK (748.5), ZAK (745), FGFR1 (740.5), RRAGA (732.5), MT1X (730.5), CYP51A1 (728), BIRC4 (726), ZNF133 (713), GNA11 (712.5), YWHAE (712), ALDH2 (708.5), BNIP3L (708.5), TPM1 (704.5)
II	hsa-miR-154, hsa-miR-184	NFIB (824.5), SFERS2 (821.5), CRYAB (809), CDKN1B (792), ALDH1A1 (788), PPAP2B (772), HSPA1A (767), SART1 (756), PPP2CB (752), TGFBR2(737.5), FLNB (707), EPS8 (705.5)

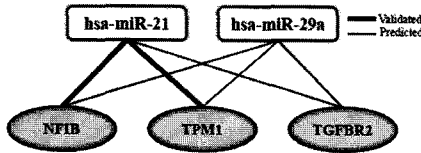


그림 1 실험적 증거가 뒷받침된 생물학적으로 유의한 miRNA-mRNA 상호작용 조합의 서브 집합 (validated: 실험적으로 검증된 목표유전자, predicted: 전산학적으로 예측된 목표유전자)

증양을 포함하여 다양한 기능적 장애를 이끄는 데 주요 역할을 한다. 또, 그림 1을 구성하고 있는 유전자 NFIB [15], TPM1[16], TGFBR2[17]는 암과 매우 밀접한 관련이 있음이 밝혀져 있다. 특히 NFIB와 TPM1은 실험적으로 hsa-miR-21의 목표 유전자임이 검증되었으며, 이는 서로 밀착하여 발암 현상을 유도하는데 중대한 영향을 준다고 보고되어 있다[15,16]. 결과적으로 제안한 방법이 생물학적으로 유의하게 명백히 암과 관련되어 있는 암 특이적 miRNA-mRNA 상호작용 조합을 발견하는데 유용한 모델임을 추론할 수 있다.

6. 결론

본 논문에서는 이종의 발현 데이터로부터 암 특이적 miRNA-mRNA 상호작용 조합을 효과적으로 탐색할 수 있는 새로운 방법을 제안하였다. 제안한 계층적 하이퍼네트워크는 하나 이상의 모달리티 사이의 관계를 분석, 조합 할 수 있는 구조로 되어있어 miRNA와 mRNA와 같이 이종적 요소 간의 생물학적으로 타당한 상호작용 조합을 탐색하는데 적합한 모델이다. 또한 높은 분류 성능과 더불어 다양한 리소스 사이의 연관관계 이해를 위한 해석 가능한 결과를 제공함으로써 복잡한 생물학 문제를 해결하는데 유용하다. 실험적으로 제안한 모델을 인간 암 관련 miRNA와 mRNA 발현 데이터에 적용하여 높은 분류 정확도로 암 특이적 miRNA-mRNA 상호작용 조합을 탐색하였다. 또 발견한 miRNA-mRNA 상호작용 조합에 대해 유전자 온톨로지 분석과 문헌 검색을 수행하여 생물학적으로도 유의함을 제시하였다.

본 연구에서는 계층적 하이퍼네트워크 학습 과정에 있어 랜덤 선택 기반으로 진행하였지만 향후 연구에서는 계층적 하이퍼네트워크 프레임워크에 상관관계 정보와 같은 사전 지식(prior knowledge)을 도입하여 생물학적으로 더 의미 있는 결과를 얻을 수 있도록 개선할 것이다.

참고 문헌

[1] S. Yoon and G.D. Micheli, "Prediction of regulatory modules comprising microRNAs and target

genes," *Bioinformatics*, 21, pp.ii93-ii100, 2005.

[2] J.C. Huang *et al.*, "Detecting microRNA targets by linking sequence microRNA and gene expression data," In *Proceedings of RECOMB*, pp.114-129, 2006.

[3] J.-G. Jeong *et al.*, "Discovery of microRNA-mRNA modules via population based probabilistic learning," *Bioinformatics*, vol.23, no.9, pp.1141-1147, 2007.

[4] S.-J. Kim *et al.*, "Evolutionary layered hypernetworks for identifying microRNA-mRNA regulatory modules," *IEEE World Congress Computational Intelligence (WCCI 2010)*, pp.2299-2306, 2010.

[5] J. Lu *et al.*, "MicroRNA expression profiles classify human cancers," *Nature*, 435, 834-838, 2005.

[6] S. Ramaswamy *et al.*, "Multiclass cancer diagnosis using tumor gene expression signature," *PNAS*, vol.98, no.26, pp.15149-15154, 2001.

[7] B.-T. Zhang, "Hypernetworks: A molecular evolutionary architecture for cognitive learning and memory," *IEEE Computational Intelligence Magazine*, vol.3, no.3, pp.49-63, 2008.

[8] J.-W. Ha *et al.*, "Evolutionary hypernetwork models for aptamer-based cardiovascular disease diagnosis," *The Genetic and Evolutionary Computation Conference (GECCO 2007)*, pp.2709-2716, 2007.

[9] P.A. Futreal *et al.*, "A census of human cancer genes," *Nature Reviews Cancer*, 4, pp.177-183, 2004.

[10] S. Zhu *et al.*, "MicroRNA-21 targets tumor suppressor genes in invasion and metastasis," *Cell Research*, 18, pp.350-359, 2008.

[11] M. Fabbri *et al.*, "MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B," *PNAS*, vol.104, no.40, pp.15805-15810, 2007.

[12] G.A. Calin *et al.*, "MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias," *PNAS*, vol.101, no.32, pp.11755-11760, 2004.

[13] M. Malzkorn *et al.*, "Identification and functional characterization of microRNAs involved in the malignant progression of gliomas," *Brain Pathology*, 2009.

[14] T. Beissbarth and T.P. Speed, "GOstat: Find statistically overrepresented Gene Ontologies within a group of genes," *Bioinformatics*, vol.20, no.9, pp.1464-1465, 2004.

[15] S. Fujita *et al.*, "miR-21 gene expression triggered by AP-1 is sustained through a double-negative feedback mechanism," *Journal of Molecular Biology*, vol.378, no.3, pp.492-504, 2008.

[16] S. Zhu *et al.*, "MicroRNA-21 targets the tumor suppressor gene tropomyosin 1 (TPM1)," *Journal of Biological Chemistry*, vol.282, no.19, pp.14328-14336, 2007.

[17] R.L. Elliott and G.C. Blobe, "Role of transforming growth factor beta in human cancer," *Journal of Clinical Oncology*, vol.23, no.9, pp.2078-2093, 2005.