

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

김수진^{01,2} 하정우³ 장병탁^{1,2,3}

서울대학교 생물정보학 협동과정¹

바이오지능기술연구센터(CBIT)²

서울대학교 컴퓨터공학부³

sjkim@bi.snu.ac.kr, jwha@bi.snu.ac.kr, btzhang@bi.snu.ac.kr

Exploring cancer-specific microRNA-mRNA interactions by Evolutionary Layered Hypernetwork Models

Soo-Jin Kim^{01,2} Jung-Woo Ha³ Byoung-Tak Zhang^{1,2,3}

Interdisciplinary Program in Bioinformatics, Seoul National University¹

Center for Biointelligence Technology²

School of Computer Science & Engineering, Seoul National University³

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 하는 것으로 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. 이런 측면에 초점을 맞추어 많은 연구[1,2,3]들이 진행되었지만 miRNA와 mRNA 상호작용의 다양성과 복잡성 때문에 기능적인 miRNA-mRNA 집합을 추론하는 것은 여전히 어려운 문제로 남아 있다.

따라서 본 연구에서는 이종의 발현 데이터로부터 기능적으로 상호작용하는 miRNA-mRNA 조합을 탐색하기 위한 진화연산 기반의 새로운 하이퍼네트워크 모델, 계층적 하이퍼네트워크(layered hypernetworks: LHNs)를 제안한다[4]. 이는 기존의 하이퍼네트워크 모델을 발전시킨 모델로 이종적 데이터 특성을 반영할 수 있는 모달리티 의존적 하이퍼네트워크와 각 모달리티 간 관계를 분석, 조합 할 수 있는 통합 하이퍼네트워크, 두 개의 계층으로 구성된 계층적 하이퍼네트워크로 miRNA와 mRNA와 같이 이종적 요소간의 상호작용을 탐색하는데 있어 적합한 모델이라 할 수 있다. 계층적 하이퍼네트워크 모델에서 하부 층은 데이터 특성을 반영하는 하나 이상의 모달리티 의존적인 하이퍼네트워크로 구성되며, 이에 하이퍼네트워크 개수는 주어진 데이터의 모달리티 수와 동일하고 각 하이퍼네트워크는 각 모달리티를 대변하는 인자들로 구성된 데이터로부터 샘플링을 통하여 구축된다. 상부 층은 하부 층과 달리 오직 한 개의 하이퍼네트워크만 존재하며 이는 각 모달리티 하이퍼네트워크로부터 선택된 하이퍼에지 통합에 의해 생성되고 여러 모달리티 사이의 관계를 분석 할 수 있게 한다. 두 개의 모달리티를 가지고 있는 주어진 데이터 셋이 D 일 때, $D = \{x^{(n)}\}_{n=1}^N = \{(\mathbf{m}^1, \mathbf{m}^2)^{(n)}\}_{n=1}^N$, 하부 층의 k -하이퍼네트워크로부터 생성된 상부 층의 하이퍼네트워크 에너지 함수 $\varepsilon(x^{(n)}; W)$ 는 다음과 같이 정의된다.

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

\mathbf{m}^1 과 \mathbf{m}^2 은 데이터 샘플 x 를 구성하고 있는 각 모달리티 변수의 벡터이며, 두 모달리티의 n 번째 데이터 생성의 확률 $P(x^{(n)}|W)$ 는 아래와 같으며 $Z(W)$ 는 부분 함수이다.

$$P(x^{(n)}|W) = \frac{1}{Z(W)} \exp[-\varepsilon\{(\mathbf{m}^1, \mathbf{m}^2)^{(n)}; W\}], \text{ where } Z(W) = \sum_{x^{(m)} \in D} \exp\{-\varepsilon(x^{(m)}; W)\}$$

표 1 다른 기계학습방법들과의 성능 비교 결과

분류기법	SVMs	LHNs	Bayesian Networks	Random Forests	Naïve Bayes	J48
정확도 (표준편차)	97.30% (0.58)	95.39% (0.98)	92.92% (1,59)	90.90% (0.98)	89.21% (0.79)	84.41% (2.70)

또 기존의 하이퍼네트워크 모델과 마찬가지로 계층적 하이퍼네트워크 프레임워크의 학습전략 역시 거대한 문제공간에서 최적의 고차원 조합 집합을 찾기 위해 랜덤 샘플링 기반의 진화적 학습방법을 이용하였다.

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. 또한 생물학적으로 더 유의미한 결과를 얻기 위해 전체 mRNA 목록에서 [7]에 따라 mRNA를 추출하여 실험을 수행하였다. 이에 실험 결과로 10-fold-cross validation을 이용하여 10번 반복 수행한 분류 정확도의 평균을 보여준다. 표 1과 같이 본 제안하는 모델은 SVMs(support vector machines)에 경쟁력 있는 성능을 보이고, 결정트리(decision trees), 베이저안 네트워크(Bayesian networks), 나이브 베이즈(naïve Bayes), 랜덤 포레스트(random forests)에 비해서는 우월한 분류 성능을 보여주었다. 또한 제안한 모델에 의해 발견된 miRNA-mRNA 조합들은 암에서 기능적으로 상호작용하는 집합으로 유전자 온톨로지(Gene Ontology) 분석, 문헌 검색 등으로 생물학적 유의성을 검증하였다. 그림 1은 실험적인 증거를 바탕으로 생물학적으로 더 의미 있는 miRNA-mRNA 상호작용 조합의 부분집합을 추출한 결과를 보여준다. 이와 같은 여러 분석 결과를 통해 제안한 방법이 높은 분류 정확도로 생물학적으로 유의하게 명백히 암과 관련되어 있는 암 특이적 miRNA-mRNA 상호작용 조합을 발견하는데 유용한 모델임을 확인 할 수 있다.

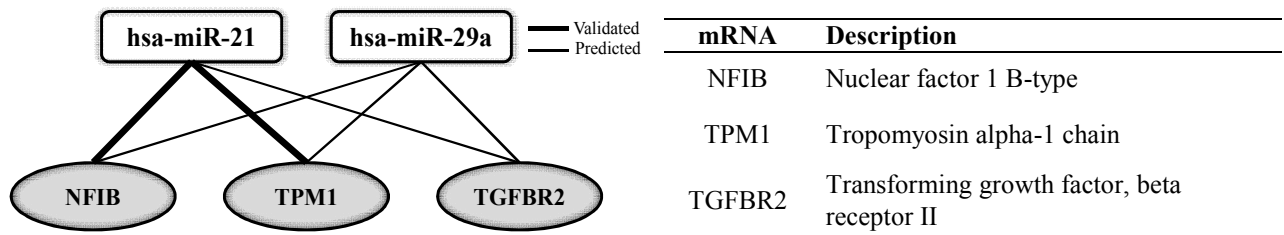


그림 1 실험적 증거가 뒷받침된 생물학적으로 유의한 miRNA-mRNA 상호작용 조합의 부분집합 (validated: 실험적으로 검증된 목표유전자, predicted: 전산학적으로 예측된 목표유전자)

감사의 글

이 논문은 지식경제부 및 한국산업기술평가관리원의 IT 산업원천기술개발사업(KI002138, 차세대 맞춤형 서비스를 위한 기계학습 기반 멀티모달 복합 정보 추출 및 추천기술 개발, MARS), 교육인적자원부 학술진흥재단(KRF-2008-314-D00377), 교육과학기술부 한국연구재단(No. 2010-0017734) 및 컴퓨터연구소(ICT)에 의해 지원되었음.

참고문헌

[1] Yoon, S. and Micheli, G. De, Prediction of regulatory modules comprising microRNAs and target genes, *Bioinformatics*, 21, pp. ii93-ii100, 2005.

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

[3] Joung, J. -G. et al., Discovery of microRNA-mRNA modules via population-based probabilistic learning, *Bioinformatics*, 23(9), pp. 1141-1147, 2007.

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

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

[6] Ramaswamy, S. et al., Multiclass cancer diagnosis using tumor gene expression signature, *PNAS*, 98(26), pp.15149-15154, 2001.

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