miRNA 와 mRNA 통합 분석을 위한 웹 기반 시스템 개발

김다연 [1], 고윤희 [1]
[1] 한국외국어대학교 바이오메디컬공학과
dayeonkim@hufs.ac.kr, younko@hufs.ac.kr

Development of web-based system for miRNA and mRNA integrated analysis

Da-Yeon Kim[1], Younhee Ko[1]
[1]Division of Biomedical Engineering, Hankuk University of Foreign Studies

요 약

기존의 질병 관련 연구들은 대부분 유의미하게 변화되는 유전자들을 찾아내고(Differentially Expressed Genes, DEGs), 이들이 연관된 생물학적 패스웨이(biological pathway)를 찾아내는 방향으로 이루어졌다. 더불어 miRNA(microRNA)가 많은 mRNA 의 발현을 조절하며, 실제 면역, 대사 및 세포 사멸을 포함한 여러 필수 생리학적 및 질병에 매우 중요한 역할을 한다고 밝혀지며, 바이오 마커로써의 miRNA 를 찾아내고자 하는 연구가 활발히 진행되기 시작하였다. 하지만 mRNA 나 miRNA 의 독립적인 연구만으로는 명확한 질병과의 연관성이나 기능을 이해하기에는 어려움이 있다. 따라서 본 연구에서는 질병 상태에서 유의미하게 변화되는 miRNA 와 이러한 miRNA 에 의해 조절되는 mRNA 를 함께 고려하여 분석함으로써, 실제 질병의 발병 원인이 되는 생물학적 패스웨이나 메커니즘을 밝히고자 하였다. 또한, miRNA 와 mRNA 의 연관성을 찾기 위해, PPI(protein-protein interaction) 네트워크에 기반을 둔 RWR(Random Walk with Restart Algorithm)를 적용하여, 직접적 연관성뿐 아니라, 유전자 간의 숨겨진 간접적인 패스웨이를 고려하여 분석하기 위한 웹 기반 시스템을 개발하였다. 이 시스템은 mRNA-miRNA 를 함께 고려한 통합 분석을 통해 숨겨진 질병의 메커니즘을 이해하고 치료 방법을 찾아내는 데 크게 공헌할 것이다.

1. 서론

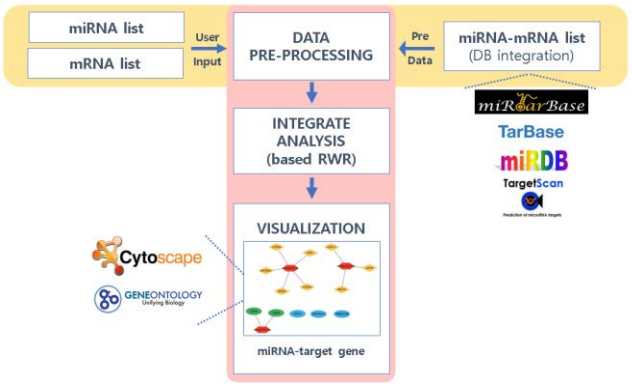
miRNA(microRNA)는 18~26 개의 뉴클레오타이드를 가진 작은 비암호화 RNA 로서 표적 messenger RNA(mRNA)와 상호작용하여 발현을 억제한다[1]. 이들은 면역, 대사 및 세포 사멸 같은 인간의 필수 생물학적 과정에서 중요한 역할을 하므로 다양한 질병의 바이오 마커 혹은 치료 표적의 가능성에 관한 연구에서 중요하게 연구되고 있다[2]. 지난 몇 년간 이뤄진 miRNA 연구는 대부분 질병 특이적 miRNA 식별과 miRNA 의 표적유전자를 찾아내는 것에 초점이 맞춰 연구되어 왔다[3]. 더 나아가 NGS 기반의 실험적 기술의 개발과 다양한 알고리즘의 향상으로 많은 miRNA-표적유전자(target gene) 상호작용 데이터가 축적되었으며, 이를 기반으로 miRNA 와 질병 간의 인과관계를 밝히는 연구가 활발히 진행되기 시작하였다. 또한, 기존의 질병 관련 연구들은 질병으로 유의미하게 변화되는 유전자 (Differentially Expressed Genes)들을 찾고, 이들이 연관된 생물학적 패스웨이(Biological pathway)를 찾음으로써, 질병의 원인이나 치료 방법을 찾고자 하였다. 이러한 miRNA 와 mRNA 는 서로 생물학적으로 상호작용을 하며, 실제 질병의 발병에 간접적으로 영향을 미치고 있다. 최근에는 mRNA 와 miRNA 의 발현 패턴을 함께 분석함으로써 질병의 원인을 찾고자 하는 연구들이 진행되기 시작하였다.

따라서 본 연구에서는 miRNA 와 miRNA 의 표적유전자 그리고 유의미하게 변화되는 mRNA 와 간접적으로 연결된 생물학적 패스웨이 정보들을 통합하여 분석할 뿐만 아니라 시각 자료를 제공하는 웹 기반 시스템을 개발하였다.

2. 웹 시스템 구현 과정

RWR 기반 miRNA 기능 분석 웹 기반 시스템은 Linux 운영 체제 내 Apache 웹 서버에서 실행된다. 그림 1 은 서버의 동작 흐름이다. 사용자로부터 2 개의 데이터를 받으면 서버는 (1) 데이터 전처리, (2) RWR 알고리즘 실행, (3) 시각화 단계를 수행한다. miRNA-표

적유전자 상호관계 네트워크를 시각화하고 target GO term 순위를 제공함으로써 mRNA 와 miRNA 의 기능 분석에 필요한 자료를 제공한다.



(그림 1) 웹 시스템 동작 흐름도

2.1 데이터 수집 및 선별

mRNA 와 miRNA 통합 분석을 위해 miRNA 표적유전자, 유전자 간의 상호작용 정보, 생물학적 패스웨이에 관한 Gene Ontology(GO) 용어 정보를 수집하여 시스템 데이터베이스를 구성하였다.

miRNA 의 표적유전자는 잘 알려진 4 개의 데이터베이스(TargetScan[4], miRDB[5], DIANA Tarbase[6], miRTarBase[7])로부터 다운로드받아 사용하였다. 이와 같은 miRNA 의 표적유전자 데이터베이스는 데이터 생성 방식에 따라 크게 두 유형으로 나뉘는데, 실험을 통해 얻는 방법과 프로그램으로 예측하는 방법이다. 예측된 데이터만 사용했던 기존의 방법[8-9]과 달리 제안된 시스템에서는 두 유형의 데이터베이스 모두 사용했다. TargetScan 와 miRDB 데이터베이스 각자의 프로그램으로 miRNA 표적유전자를 예측한 결과가 있으며, 두 결과가 서로 상이하기 때문에 교집합으로 존재하는 miRNA-표적유전자 데이터만 선별했다. DIANA Tarbase 와 miRTarBase 데이터베이스는 실험적으로 검증된 miRNA 표적유전자 목록으로써, 합집합으로 존재하는 모든 데이터를 사용했다. 이렇게 선별된 데이터를 모두 통합하여, 본 연구에서는 miRNA 2,668 개, mRNA 18,902 개, 상호작용 800,723 개로 구성된 miRNA 표적유전자 데이터 세트를 사용하였다.

유전자 간 상호작용 정보는 STRING[10] 데이터베이스로부터 다운로드 받았으며 총 11,938,498 개의 PPI(Protein-protein interaction) 정보가 본 연구를 위해 이용되었다. NCBI 에서 제공하는 유전자 아이디를 이용해 protein ID(string ID)를 gene id(entrez ID)로 매핑하여 유전자 18,144 개에 대한 유전자 간의 상호작용 11,428,553 개를 확보할 수 있었다.

또한, AmiGO 데이터베이스로부터 유전자들의 기능에 대한 정보인 GO 정보 중 생물학적 패스웨이에 관련된 용어와 그것에 관련된 유전자 정보를 활용하여 표적유전자들의 분석에 사용하였다. 5 개 이하의 유전자와 매핑되는 GO 용어는 제거해 총 12,415 개 유전자에 대한 가지 GO 18,434 용어를 수집하였다.

2.2 방법

본 시스템은 해당 질병이나, 실험 상황에 따라 유의미하게 변화되었다고 알려진 miRNA 리스트와 mRNA 리스트를 입력으로 하여, 다음의 과정을 수행한다.

2.2.1 전처리 과정

RWR 알고리즘을 실행하는 데 필요한 데이터 세트를 생성하는 단계이다. 시스템 데이터베이스의 miRNA 표적유전자 데이터세트로부터 사용자 정의 miRNA 에 대한 표적유전자 세트를 생성한 뒤 사용자 정의 mRNA 와 통합하여 하나의 시드(seed) 유전자 세트를 생성한다. 이후 시드 유전자 세트와 STRING 데이터베이스 내 모든 유전자 세트를 통합하여 전체 유전자 세트를 생성한다. 시드 유전자 세트와 전체 유전자 세트는 RWR 모델의 입력데이터가 된다.

2.2.2 RWR 알고리즘 실행

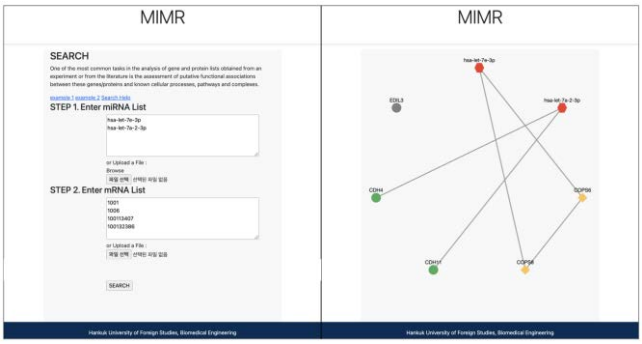
랜덤 워크는 노드 간의 상호작용을 엣지로 표현한 네트워크상에서 관심 있는 노드를 기점으로 무작위로 주변 엣지를 선택해 이웃 노드로 움직이는 것을 말한다. RWR 알고리즘은 랜덤 워크를 반복적으로 수행해 각 노드를 방문할 확률을 계산한다[11]. 특정 노드의 방문 확률을 1 에서 빼면 관심 있는 노드로부터 해당 노드가 떨어진 거리 점수를 계산할 수 있다. 이런 방식은 관심 있는 노드와 직접 상호작용하는 노드뿐만 아니라 간접적으로 연결되는 노드 또한 고려할 수 있게 된다. 제안한 시스템은 RWR 알고리즘을 이용해 PPI 네트워크로부터 입력받은 시드 유전자 세트와 인간의 전체 단백질 코딩 유전자 간의 거리를 점수화한다. 이때, 거리 점수를 스레시홀딩(thresholding)하여 시드 유전자와 연관성이 높은 유전자들을 뽑아낸다. 결국, 특정 환경에서 유의미하게 변화된 miRNA 표적유전자와 mRNA 그리고 이들과 직·간접적으로 연결된 유전자 세트를 모두 더한 확장된 시드 유전자 세트를 얻을 수 있게 된다.

그다음 해당 질병과 연관성 높은 생물학적 패스웨이를 찾기 위해 GO 분석을 한다. GO 분석은 관심 있는 유전자가 특정 패스웨이에 얼마나 있는지 수치화하여 해당 질환과 연관성이 높은 패스웨이를 찾아낸다. 제안한 시스템에서는 RWR 알고리즘으로 계산한

거리 점수로부터 xd-score 를 계산해 확장된 시드 유전자 세트에 대한 패스웨이의 유의성을 평가했다.

2.2.3 시각화

시각화 단계에서는 확장된 시드 유전자 세트와 GO 분석 결과를 바탕으로 mRNA-miRNA 통합 분석한 결과를 시각화하여 나타낸다. xd-score 을 이용한 생물학적 패스웨이 순위표뿐만 아니라 miRNA 와 확장된 시드 유전자 간 모든 상호작용을 시각화한 네트워크를 제공한다. 그림 2 는 웹 시스템의 초기 입력 화면과 cytoscape.js 를 이용해 구현한 네트워크 시각화 결과를 보여준다. 네트워크의 노드는 5 가지(사용자 정의 miRNA, 사용자 정의 mRNA, miRNA 의 표적 유전자, 사용자 정의 mRNA 이며 miRNA 의 표적인 유전자, 간접적으로 연결된 유전자)로 보다 다양한 정보를 제공한다.



(a) 초기 입력 화면 (b) 네트워크 시각화 결과

(그림 2) 제안한 웹 시스템

3. 결론

본 논문은 mRNA-miRNA 를 통합 분석이 가능한 웹 기반 시스템을 제안했다. 이 시스템은 기존 웹 시스템과 달리 예측된 miRNA 의 표적유전자 정보뿐만 아니라 실험적으로 증명된 miRNA 의 표적유전자 정보 모두 통합하여 사용했다. 따라서 단순히 알고리즘으로 예측한 결과물보다 신뢰성 높으며 실험적으로 증명된 데이터만 쓰는 것보다 많은 miRNA 표적을 고려할 수 있게 되었다. 더불어 PPI 네트워크에 기반을 둔 RWR 알고리즘을 통해 miRNA 와 직·간접적으로 상호작용하는 유전자를 식별했다. 이는 기존의 웹 서비스에서 지원하지 않았던 방법으로, 해당 질환에서 중요하게 작용하는 miRNA 와 mRNA 의 숨겨진 분자적 작용 기전을 찾아내는 데 이바지할 것이다.

따라서, 제안한 웹 시스템은 입력된 miRNA 정보와 mRNA 정보로부터 숨겨진 생물학적 패스웨이를 제공하며 연구자들이 쉽게 질병의 바이오 마커 식별 및 치료 표적 탐색을 수행하도록 기여할 수 있을 것이다.

참고문헌

[1] ZHAO, Y., & SRIVASTAVA, D. “A developmental view of microRNA function.” Trends in Biochemical Sciences. Elsevier BV. Volume 32. Issue 4. Pages 189-197. 2007

[2] Chen, Y., Gao, D.-Y., & Huang, L. “In vivo delivery of miRNAs for cancer therapy: Challenges and strategies.” Advanced Drug Delivery Reviews. Elsevier BV. Volume 81. Pages 128-141.2015

[3] Chen, L., Heikkinen, L., Wang, C., Yang, Y., Sun, H., & Wong, G. “Trends in the development of miRNA bioinformatics tools. Briefings in Bioinformatics.” Oxford University Press (OUP). Volume 20. Issue 5. Pages 1836–1852. 2019

[4] Agarwal, V., Bell, G. W., Nam, J.-W., & Bartel, D. P. “Predicting effective microRNA target sites in mammalian mRNAs.” eLife. eLife Sciences Publications, Ltd. 2015

[5] Chen, Y., & Wang, X. “miRDB: an online database for prediction of functional microRNA targets.” Nucleic Acids Research. Oxford University Press (OUP). 2019

[6] Vlachos, I. S., Paraskevopoulou, M. D., Karagkouni, D., Georgakilas, G., Vergoulis, T., Kanellos, I., … Hatzigeorgiou, A. G. “DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions.” Nucleic Acids Research. Oxford University Press (OUP). 2014

[7] Huang, H.-Y., Lin, Y.-C.-D., Li, J., Huang, K.-Y., Shrestha, S., Hong, H.-C., … Huang, H.-D. “miRTarBase 2020: updates to the experimentally validated microRNA–target interaction database.” Nucleic Acids Research. Oxford University Press (OUP). 2019

[8] Nam, S., Li, M., Choi, K., Balch, C., Kim, S., & Nephew, K. P. “MicroRNA and mRNA integrated analysis (MMIA): a web tool for examining biological functions of microRNA expression.” Nucleic Acids Research. Oxford University Press (OUP). 37(Web Server issue): W356–W362, 2009

[9] Nam, S., Kim, B., Shin, S., & Lee, S. “miRGator: an integrated system for functional annotation of microRNAs.” Nucleic Acids Research. Oxford University Press (OUP). Volume 36. Issue suppl_1. Pages D159–D164. 2008

[10] Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., … Mering, C. von. “STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets.” Nucleic Acids Research. Oxford University Press (OUP). 2018

[11] Glaab, E., Baudot, A., Krasnogor, N., Schneider, R., & Valencia, A. “EnrichNet: network-based gene set enrichment analysis.” Bioinformatics. Oxford University Press (OUP). 28(18): i451–i457. 2012