

# 타겟 유전자 기반 miRNA 네트워크를 통한 질병 관련 miRNA 추출 방법

하지 환\*, 김현진\*, 박상현\*\*

## An Extraction Method of Disease Rrelated miRNA Through the Target Gene Based miRNA Network

Ji-Hwan Ha\*, Hyun-Jin Kim\*, and Sang-Hyun Park\*\*

---

이 논문은 2012년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임.  
(2012R1A2A1A01010775).

---

### 요 약

microRNA(miRNA)는 19-22개의 뉴클레오티드로 이루어지고 있으며 단백질로 번역되지 않는 것으로 잘 알려져 있다. 또한, miRNA는 유전자의 발현을 조절하는 기능을 가지는 코딩하지 않는(small-non-coding) RNA로서 세포 분화와 질병 발생(pathogenesis)에 중요한 역할을 한다는 것이 입증 되었고, 수많은 질병 관련 miRNA 계산 방법이 제안되어 왔다. 본 논문에서는 타겟 유전자(target gene)를 기반으로 한 miRNA 네트워크를 구축하고, miRNA 마이크로어레이 데이터를 활용하여 효과적인 질병 관련 miRNA 추출 방법을 제시한다. 본 논문에서 새로이 제안하는 알고리즘을 적용하여 실험을 수행한 결과, 기존 연구의 방법보다 더 좋은 성능을 보이는 것을 확인하였다.

### Abstract

In general, it is well known that microRNA(miRNA) is composed of 19-22 nucleotide which can not be translated into the protein. miRNA has been proved to be playing a vital role in cell differentiation and pathogenesis, and then numerous computational methods of disease related miRNA has been proposed. In this paper, we constructed miRNA network based on their target gene and suggest effective method of extracting disease related miRNA using miRNA micro array data. From the experiment results obtained from the new algorithm proposed in this paper, it can be seen that our proposed method shows more efficient performance compared to the results from the conventional method.

### Keywords

microRNA, network, microarray data, target gene, disease

---

\* 연세대학교 컴퓨터과학과  
\*\* 연세대학교 컴퓨터과학과 교수(교신저자)  
· 접수 일: 2014년 10월 01일  
· 수정완료일: 2014년 11월 12일  
· 게재확정일: 2014년 11월 15일

· Received: Oct. 01, 2014, Revised: Nov. 12, 2014, Accepted: Nov. 15, 2014  
· Corresponding Author: Sang-Hyun Park  
Dept of Computer Science, Yonsei University, Sinchon-dong, Seodaemun-gu  
Seoul, 120-749, Korea  
Tel.: +82 2 2123-5714, Email: sanghyun@cs.yonsei.ac.kr

## I. 서 론

miRNA(Micro RNA)는 19-22개의 뉴클레오티드로 이루어진 small non-coding RNA로서[1], DNA 유전 정보로부터 단백질이 합성되는 과정에서 mRNA(Messenger RNA)의 3' UTR(Untranslated Region)에 상보적인 결합을 통해 유전자 발현을 억제하거나 mRNA의 degradation을 일으키는 것으로 알려져 있다[2]. 또한, 최근 연구 결과에 따르면 miRNA는 세포 증식(Proliferation)과 사멸(Death) 등 생명 현상에 중요한 역할을 하는 것으로 밝혀졌다. 이러한 miRNA의 정상적인 발현 값(Normal Expression Value)과 비교를 했을 때, 비정상적인 발현 값의 증가와 감소는 질병 유발과 진단을 예측하는데 중요한 잣대가 될 것이다.

한편, miRNA의 기능과 이와 관련된 질병정보는 매우 제한적이다. 따라서 컴퓨터 알고리즘을 이용하여 이를 밝히기 위한 많은 연구가 이루어져 왔다. Jiang et al[3]은 타겟 유전자(Target Gene)를 기반으로 한 miRNA 네트워크를 이용하여 표현형이 비슷한 질병들은 비슷한 기능을 가지는 miRNA들과 관련 있을 것이라는 가정 하에 miRNA와 질병 간의 관계를 밝혀낸 바 있으며, Gene Ontology(GO)를 사용한 miRNA와 질병 간의 예측방법으로 심장혈관계 질병 관련 miRNA도 처음으로 밝혀낸 바 있다. Chen et al[4]은 miRNA 네트워크에 랜덤워크(Random Walk) 알고리즘을 적용하여 miRNA와 질병간의 관계를 효과적으로 밝혀냈다. 하지만 이와 같은 방법들의 한계점은 이미 알려진 miRNA 데이터를 정보를 사용한다는 것이다. 이미 알려진 정보 데이터를 사용한다는 것은 정확한 의미에서 miRNA 추출방법 이라기 보다 이미 알려진 질병간의 통계 데이터에 한정되기 때문에, 예측 정도가 떨어질 수가 있으며, 예기치 못한 새로운 질병 관련 miRNA를 새로이 발굴 예측하기가 쉽지 않을 것으로 생각된다.

본 논문에서는 miRNA가 조절하는 타겟 유전자를 통해 miRNA의 기능을 유추하고, miRNA 마이크로 어레이 데이터를 사용하여 미리 알려져 있지 않다는 가정 하에, 특정 질병을 예로 들어 miRNA간의 질병 관련성의 유추 정도를 높이고, 질병 관련성을 새로이 밝혀내고자 한다. 우선 타겟 유전자를 기

반으로 한 miRNA 네트워크를 구축하고, miRNA 마이크로 어레이 데이터를 활용하여 효과적인 새로운 질병 관련 miRNA의 추출 방법을 제시한다.

본 논문의 구성은 다음과 같다. 2장에서는 본 논문에서 제안하는 miRNA의 추출방법 및 데이터 구축에 관한 내용으로, miRNA 네트워크 구축방법과 데이터 통합 스코어 방법을 소개한다. 3장에서는 실험방법 및 분석결과로, 특정 질병 miRNA 추출실험 방법과 그 결과를 기술하여 본 논문에서 주장하는 방법의 타당성을 설명한다. 4장에서는 결론 및 향후 연구 방향을 제시한다.

## II. 메소드 및 데이터 출처

### 2.1 miRNA 네트워크 생성

실제로 현존하는 miRNA와 관련된 정보, 데이터는 많이 부족한 현실이다. 따라서 miRNA의 기능을 유추하는 방법 중 하나로, miRNA가 조절하는 타겟 유전자의 기능을 통해 추론하는 방법이 자주 사용된다[3].

본 연구에서는 miRNA에 관련된 정보를 얻기 위해, 먼저 4개의 miRNA 관련 데이터베이스(mirTarbase, miRecords, miR2Disease, hmdd)에서 miRNA와 miRNA가 조절하는 타겟 유전자를 다운로드 받았다[5]-[8]. 표 1에는 논문에서 새로이 구축하고 추론방법에 사용된 각 miRNA 데이터베이스 출처와 각각이 포함하고 있는 miRNA와 타겟 유전자의 수를 나타내고 있다.

각 데이터베이스는 서로 중복된 miRNA 리스트를 가지고 있다. 따라서 데이터를 통합하는 과정에서 각 데이터베이스 별로 중복된 것을 제거하고 총 1125개의 miRNA와 40544개의 타겟 유전자를 확보할 수 있었다.

표 1 miRNA 데이터베이스

Table 1. miRNA database

데이터 출처	miRNA 수	Target gene 수
miTarbase	580	37,484
miRecords	281	1,740
miR2Disease	176	602
hmdd	263	1,060

확보 구축한 miRNA 리스트를 토대로 각 miRNA가 조절하는 유전자를 기반으로 하여 miRNA 네트워크를 생성하였다. miRNA를 노드(Node)로, 두 miRNA 사이에 공통된 타겟 유전자가 있으면 연결선(edge)을 추가하여 네트워크를 생성 구축하였다.

그림 1은 두 miRNA 노드 사이에 연결선을 생성하는 과정을 나타내고 있다. 본 연구에서 생성 구축한 네트워크에서는 miRNA가 조절하는 유전자의 수

는 최소 1개에서 최대 2546개로 다양한 범위를 갖는다. 두 miRNA 노드 사이에 공통된 타겟 유전자가 존재하면 에지를 연결하였다. 그리고 연결선에 공통된 타겟 유전자가 많을수록 이웃한 두 miRNA 노드는 유사한 기능을 할 것이라는 가정하였다. 또한, 그림 1 과정을 바탕으로 전체적인 miRNA 네트워크(그림 2) 기반을 만들게 된다.

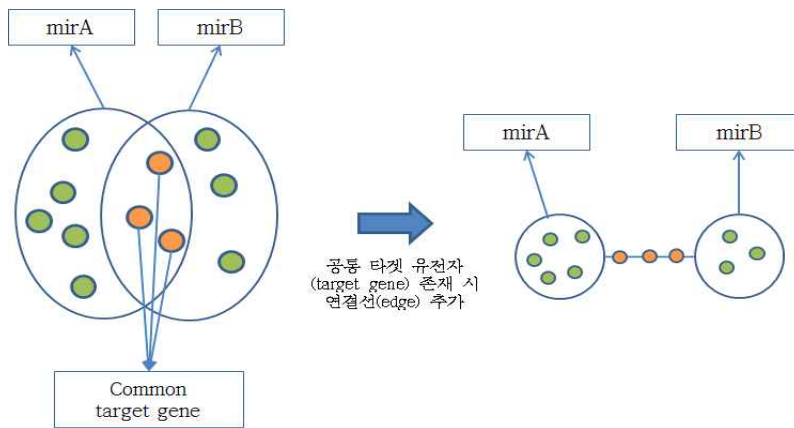


그림 1. miRNA 노드 간 연결선 생성 과정  
Fig. 1. Process of creating edge between two miRNA nodes

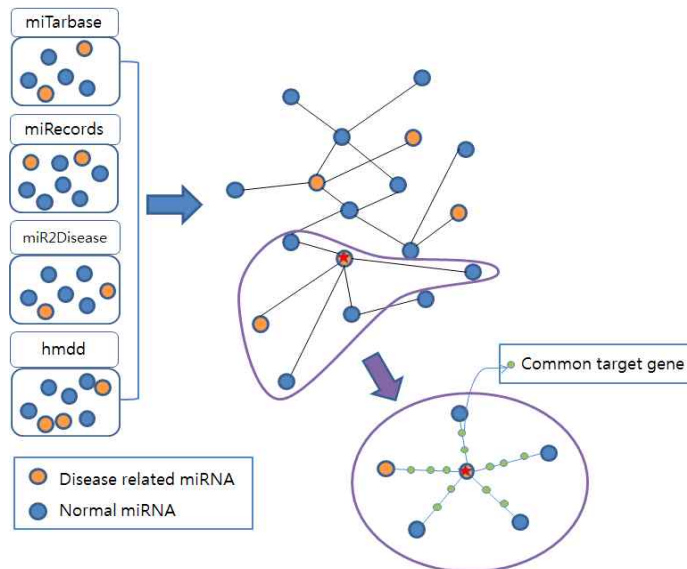


그림 2. 타겟 유전자를 기반으로 한 miRNA 네트워크  
Fig. 2. miRNA network based on target gene

## 2.2 스코어 방법

miRNA 네트워크를 이용하여 질병 관련 miRNA를 스코어 내는 방법은, 크게 miRNA가 조절하는 타겟 유전자를 통해 miRNA의 질병 관련 유무를 판단하는 방법과, 특정 질병에서 나타나는 miRNA 이상 발현 값(Expression Value)을 이용하는 것으로 나누어진다. 이 정보들을 통해 네트워크상에서 존재하는 각 miRNA 노드의 질병 관련성을 점수화한다.

### 2.2.1 타겟 유전자를 이용한 스코어방법

기존의 논문에서는 miRNA의 기능을 추론하는 방법 중 하나로 miRNA가 조절하는 타겟 유전자를 통해 추론하는 방법을 사용하고 있다[9]. 본 논문에서는 miRNA 노드의 주변 연결선에 포함 되어 있는 타겟 유전자를 토대로 질병 관련 유무를 검사하고, 질병과 관련이 있는 타겟 유전자의 개수를 노드의 스코어에 합산하는 방법을 사용하였다.

그림 2는 스코어 방법의 예를 나타내는 개념도이며, 그림 2에서 검은색은 질병 관련 miRNA, 흰색은 정상 miRNA를 나타낸다. 별표로 표기된 miRNA의 질병 관련 유무를 판단하고 싶다고 하였을 때, 본 논문에서는 별표 노드의 주변 연결선에 포함되어 있는 타겟 유전자의 정보를 이용하였다.

### 2.2.2 miRNA 마이크로어레이를 이용한 스코어방법

miRNA 네트워크에 마이크로어레이 데이터를 적용하기 위해 TCGA 데이터베이스에서 534개의 miRNA 발현값을 가지고 있는 마이크로어레이 데이터를 다운로드 받았다. TCGA(The Cancer Genome Atlas)는[10] 미국 국립암연구소(NCI)와 국립인간유전체연구소(National Human Genome Research Institute)에 의해 관리되는 바이오인포매틱스 관련 데이터베이스이다.

miRNA의 이상 발현은 암유전자 발현에 영향을 미쳐, 암 발병(Pathogenesis)에 영향을 준다고 밝혀져 있다. 따라서 miRNA의 차등 발현은 암을 진단 할 수 있는 하나의 지표가 될 것이다[11]. 이를 근거로 본 논문에서는, 마이크로어레이 데이터의 전처리 과

정을 살펴보면 특정 질병이 주어졌을 때, 각 miRNA 샘플 발현값을 비교하여 0으로부터 가장 먼 값을 선택하고 이 값을 토대로 각 노드에 적용하여 스코어를 합산하는 방법을 두 번째 방식으로 사용하였다.

### 2.2.3 스코어 합산 및 순위화 과정

본 논문에서의 스코어 방법은, 우선 2.2.1절의 첫 번째 스코어 방법을 통하여 miRNA가 조절하는 타겟 유전자의 기능으로부터 miRNA의 질병 관련성을 점수화 한 후, 2.2.2절의 두 번째 스코어 방법을 통하여 특정 질병에서 이상 발현되는 miRNA 발현 값을 가지고 각 노드의 점수를 부여하는 방식도 함께 수행하였다. 이 두 가지 방법에 의한 점수를 합산하여 적용하는 가중치 합산 방식을 수행함으로써 보다 나은 실험 결과를 얻을 수 있었다.

위에서 제시한 두 가지의 스코어 방법을 통해 합산된 점수를 토대로 점수가 높은 순서대로 miRNA를 1위부터 1125위까지 순위화 하였다. 그리고 상위 랭킹에 포함된 miRNA일수록 질병과 관련이 높을 것이라 가정하고, 상위 n개의 miRNA를 뽑아 얼마나 많은 질병 관련 miRNA를 포함하고 있는지 확인해보았다.

## 2.3 데이터 출처

본 논문에서 사용된 데이터의 출처는 다음과 같다. 앞서 기술한 바와 같이 miRNA는 miRNA가 조절하는 타겟 유전자를 통해서 그 기능을 유추한다. 이에 miRNA와 miRNA가 조절하는 타겟 유전자의 정보를 2.1절의 표 1에 나타낸 바와 같이 각각의 데이터베이스로부터 다운로드 받았다. 이 표 1의 데이터를 바탕으로 miRNA 네트워크를 생성하였다. 생성된 네트워크를 바탕으로 각각의 연결선에 포함되어 있는 타겟 유전자와 특정 질병간의 관련 유무를 검사하였는데 이는 표 2의 CTD 데이터베이스에서 다운로드 받은 데이터를 이용하였다[12].

유방암에서 나타되는 miRNA 발현값은 TCGA 데이터베이스에서 다운로드 받은 miRNA 마이크로어레이 데이터를 이용하였으며, 마지막으로 특정 질병이 주어졌을 때, 스코어 방법으로 순위화한

miRNA에서 상위 n개를 뽑았을 때, miRNA와 질병 간의 관련 유무는 표 2에서의 hmdd 데이터베이스에서 다운로드 받은 데이터를 사용하였다.

표 2. 질병 관련 miRNA 데이터 및 miRNA 마이크로어레이 출처 유방암 관련 유전자 데이터  
Table 2. Sources of disease related miRNA data and miRNA microarray data (breast cancer related)

데이터 출처	질병 관련 유전자 수 (Breast cancer)	miRNA 마이크로어레이	miRNA 질병 관련 정보
CTD	5072	-	-
TCGA	-	534	-
hmdd	-	-	533

### III. 실험 결과 및 분석

#### 3.1 실험 결과

본 논문에서 제시하는 방법의 효율성을 살펴보기 위해 유방암 관련 miRNA를 추출해보았다. 유방암은 전 세계적으로 여성에게 가장 많이 발병하는 암의 한 종류로서 암중에서도 가장 다양한 종류를 가지는 암 중 하나이다. 유방암은 다른 암에 비해 치료 방법이 많은 편이며 그에 따른 치료 효과도 좋은 편이다. 유방암 환자의 5~10%가 유전적 원인인 것으로 알려져 있으며 BRCA1과 BRCA2와 같은 유전자의 돌연변이가 유방암을 발생하는 원인으로 알려져 있다. 그리고 아직까지 수많은 관련 연구도 활발하게 수행되어서 유방암과 관련된 유전자도 다른 질병 관련 유전자 보다 훨씬 많이 밝혀져 있다. 이에 본 논문에서는 유방암 관련 유전자를 샘플로 채택하여 miRNA와 유방암 사이의 관련성을 유추해보았다.

총 1125개의 miRNA 중 유방암 관련 miRNA가 192개가 존재한다고 할 때, 상위 n개를 선택했을 때 유방암 관련 miRNA를 포함한 수와 정확성을 다음 표 3에 나타내었다. 표 3에서와 같이 유방암 관련 miRNA를 추출하였을 때 최소 50%에서 최대 60% 사이의 정확도를 보임을 알 수 있었다. 정확도 계산은 식 (1)로 수행하였다.

$$\text{정확도 (Accuracy)} = \frac{\text{질병 관련 miRNA 수}}{\text{상위 n개 miRNA 수}} \quad (1)$$

표 3. 유방암 관련 miRNA 수와 정확도(상위 n개 선택시)  
Table 3. Number of breast related miRNA and its accuracy (when selecting top n)

상위 n개 선택시	유방암 관련 miRNA 수	정확도
10	5	50%
20	11	55%
30	17	57%
40	24	60%
50	28	56%
60	32	53%
70	39	56%
80	46	58%
90	52	58%
100	58	58%

#### 3.2 비교 실험 분석

본 논문에서 제시하는 질병 관련 miRNA 추출법의 성능을 평가하기 위해 Jiang et al[3]의 질병 관련 miRNA 추출 알고리즘과 비교 실험을 수행하였다. Jiang et al의 알고리즘 방식으로 유방암 관련 miRNA 추출 실험을 수행하였을 때에는, 상위 100개 선택시 30개의 유방암 관련 miRNA를 추출하여 30%의 정확도를 보였다. 반면, 본 논문에서 제시하는 알고리즘으로는 37% ~ 42%의 정확도로 더 나은 결과를 보임을 알 수 있었다.

한편, 3.1절에서의 실험에서는 본 논문 표 1에 나와 있는 데이터를 사용하였지만, 정확한 비교 실험을 위해 Jiang et al의 논문에서 사용된 데이터를 사용한 실험을 함께 수행하였다. 그 결과 3.1절의 실험결과로 얻어진 정확도(50% ~ 60%)보다 다소 조금 떨어진 정확도(37% ~ 42%)를 보였지만, 이 경우도 기존 알고리즘의 정확도 30%와 비교했을 때, 우수한 성능임을 알 수 있다.

그림 3은 각 논문이 제안하는 서로 다른 miRNA 추출방법을 통해 얻어진 상위 100개의 miRNA에 포함된 유방암 관련 miRNA수를 나타내고 있다. 기존 알고리즘[3]에서는 상위 100개의 miRNA에서 30개의 유방암 관련 miRNA를 추출한 반면, 본 논문에서 제안한 알고리즘을 통해서 37개의 유방암 관련 miRNA를 추출하여 보다 나은 성능을 나타냈다는 것을 그림 3을 통해 입증하고 있다.

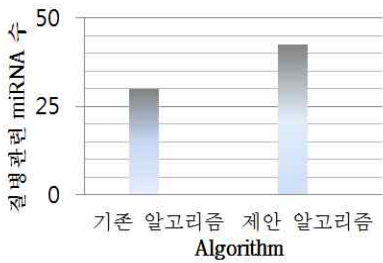


그림 3. 알고리즘 성능 평가 결과 (기존 알고리즘 대 제안 알고리즘)

Fig. 3. Performance test results of experimental algorithm (conventional algorithm vs. proposed algorithm)

표 4. Precision, Recall 비교

Table 4. Comparisons of precision and recall

알고리즘	Recall	Precision
기존 알고리즘	0.25	0.3
제안 알고리즘	0.3	0.37

마지막으로, 본 연구에서는 알고리즘의 성능을 추가적으로 평가하기 위해 기존 알고리즘과 본 논문에서 제안하는 알고리즘의 Precision과 Recall 값을 각각 구하여 비교해 보았다. 표 4에 나타난 바와 같이, 비교 논문에서는 0.25의 Recall값과 0.3의 Precision 값을 갖는 반면, 본 논문의 알고리즘에서는 0.3과 0.37의 Recall, Precision 값으로 더 나은 결과를 얻을 수 있었다. 이처럼 본 논문에서 제안하는 질병 관련 miRNA 추출 방법이 Recall, Precision에서 모두 다 더 좋은 결과를 보임을 확인할 수 있었다.

본 논문에서는 타겟 유전자로부터 miRNA의 기능을 유추하고 miRNA 마이크로어레이 데이터로부터 각 miRNA의 질병 관련성 유무를 점수화 하였다.

본 논문이 제시하는 방법으로 더 좋은 결과를 보인 차이점을 분석해보면, Jiang et al의 논문에서는 표현형(Phenotype)이 유사한 질병간에는 서로 연관이 높다는 가정하에 특정 질병 뿐만 아니라, 표현형이 비슷한 다른 질병의 정보까지도 miRNA의 질병 관련성 점수화시에 사용된다는 차이가 있었다. 이는 특정 질병에 관련해서 정확한 예측을 할 때, 정확성을 떨어뜨릴 수 있다고 생각된다. 반면, 본 논문에서는 특정 질병에만 관련있다고 밝혀진 타겟 유전자의 정보를 이용하여 miRNA의 질병관련 유무를

점수화하였다. 또한, miRNA 마이크로 어레이 데이터의 이상 발현값을 네트워크의 각 노드에 새로이 적용하는 방식을 채택함과 동시에, 질병관련성 유무의 점수화 방법을 달리 수행함으로써 보다 높은 정확도의 성능을 얻을 수 있었다.

#### IV. 결 론

본 논문에서는 타겟 유전자를 기반으로 한 miRNA 네트워크를 구축하고, miRNA 마이크로어레이 데이터를 활용하여 효과적인 새로운 질병 관련 miRNA 추출 방법 수행하였다. 특히, 본 논문에서 새로이 제안하는 알고리즘을 적용하여 질병 관련 miRNA 추출실험을 수행한 결과, 그 추출 성능이 기존의 방식보다 7~12% 높은 성능 향상을 얻을 수 있었다. 또한, 본 논문에서 제안하는 알고리즘의 성능을 평가한 결과, Recall값과 Precision값이 각각 0.3과 0.37로 기존의 알고리즘의 것(0.25/0.3)과 비교하여 더 나은 결과를 얻을 수 있었다.

금후 연구에서는 GO term, RNA-seq 데이터를 사용, miRNA 네트워크를 포괄적으로 접근하여 보다 더 효율적으로 질병 관련 miRNA를 밝혀내고자 한다. 아울러, 본 논문에서는 특정의 유방암 관련 miRNA 데이터를 예로 이용하였으나, 금후는 각종 질병 관련의 복합적인 질병 관련 miRNA 혼합 데이터로부터, 각각의 특정 질병 miRNA를 높은 성능으로 정확하게 각각 추출하는 연구를 수행하여 본 논문에서 제안 한 알고리즘의 타당성을 확인할 필요가 있다고 사료된다.

#### References

- [1] David P Bartel, "MicroRNAs: genomics, biogenesis, mechanism and function", Cell, Vol. 116, No. 2, pp. 281-297, Jan. 2004.
- [2] V. Narry Kim, "MicroRNA biogenesis: coordinated cropping and dicing", Nature Reviews Molecular Cell Biololgy, Vol. 6, pp. 376-385, May 2005.
- [3] Jiang Q., Hao Y., Wang G., Juan L., Zhang T., Teng M., Liu Y., and Wang Y., "Prioritization of



disease microRNAs through a human phenome-microRNAome network", BMC System Biology, 4 S2., May 2010.

- [4] Chen X, Liu MX, and Yan GY, "RWRMDA: predicting novel human microRNA-disease associations", Mol Biosyst, Vol. 8, pp. 2792-2798, July 2012.
- [5] Hsu SD, Lin FM, Wu WY, Liang C, Huang WC, and Chan WL, et al., "MiRTarBase: a database curates experimentally validated microRNA-target interactions", Nucleic Acids Res, Vol. 39, pp. 163-169, Nov. 2010.
- [6] Xiao, F., Zuo, Z., Cai, G., Kang, S., Gao, X., and Li, T., "miRecords: an integrated resource for microRNA-target interactions", Nucleic Acids Research, Vol. 37, pp. 105-110, Nov. 2008.
- [7] Jiang Q. et al., "miR2Disease: a manually curated database for microRNA deregulation in human disease", Nucleic Acids Research, Vol. 37, pp. 98-104, Sep. 2008.
- [8] Li, Y. et al., "HMDD v2.0: a database for experimentally supported human microRNA and disease associations", Nucleic Acids Research, Vol. 42, pp. 1070-1074, Nov. 2013.
- [9] Xia Li, Qianghu Wang, Yan Zheng, Sali Lv, Shangwei Ning, Jie Sun, and Teng Huang, "Prioritizing human cancer microRNAs based on genes' functional consistency between microRNA and cancer", Nucleic Acids Res, Vol. 39, pp. 153, Oct. 2011.
- [10] <http://cancergenome.nih.gov>, The Cancer Genome Atlas.
- [11] Boyeon Seo and Youngmi Yoon, "Prognosis Prediction of Prostate Cancer by Integrating Microarray Data and Clinical Factors", JKIIIT, Vol. 10, No. 5, pp. 139-150, May 2012.
- [12] <http://ctdbase.org>, The Comparative Toxicogenomics Databases.

## 저자소개

하 지 환 (Ji-Hwan Ha)



2013년 8월 : 부산대학교  
바이오정보전자과 졸업(학사)  
2013년 9월 ~ 현재 : 연세대학교  
컴퓨터과학과 석사과정  
관심분야 : 바이오인포매틱스,  
데이터마이닝, 데이터베이스

김 현 진 (Hyun-Jin Kim)



2010년 : 연세대학교 컴퓨터과학과  
졸업(학사)  
2010년 ~ 현재 : 연세대학교  
컴퓨터과학과 통합과정  
관심분야 : 바이오인포매틱스,  
데이터마이닝, 텍스트 마이닝,  
그래프마이닝, 데이터베이스

박 상 현 (Sang-Hyun Park)



1989년 서울대학교 컴퓨터공학과  
졸업(학사)  
1991년 : 서울대학교 대학원  
컴퓨터공학과(공학석사)  
2001년 UCLA 대학원  
컴퓨터과학과(공학박사)  
1991년 ~ 1996년 : 대우통신 연구원

2001년 ~ 2002년 : IBM T. J. Watson Research Center  
Post-Doctoral Fellow

2002년 ~ 2003년 : 포항공과대학교 컴퓨터공학과 조교수

2003년 ~ 2006년 : 연세대학교 컴퓨터과학과 조교수

2006년 ~ 2011년 : 연세대학교 컴퓨터과학과 부교수

2011년 ~ 현재 : 연세대학교 컴퓨터과학과 교수

관심분야 : 데이터베이스, 데이터마이닝, 바이오인포매틱스,  
적응적 저장장치 시스템, 플래쉬메모리 인덱스, SSD.