

암 관련 단백질과 상호작용하는 microRNA에 가중치를 부여함으로써 유용한 정보 도출 microRNA of interaction cancer related protein

박 별 나, 김 학 용
충북대학교 생화학과

Park, Byeol Na and Kim Hak Yong
Department of Biochemistry, Chungbuk
National Univ.

요약

선행연구에서 우리는 암과 관련된 단백질-단백질 상호작용 네트워크와 단백질-질병 네트워크를 통해서 핵심 단백질 60개를 추출했다. 이 단백질들을 조절하여 암을 제어하기 위한 방법으로 miRNA(microRNA)를 이용하기 위해 단백질과 상호작용하는 miRNA와 miRNA 서열정보를 추출하였다. 한 단백질과 상호작용하는 miRNA의 수가 많았기 때문에 각각의 miRNA에 대해 우선순위를 주어 서 가중치를 부여했는데, 기준으로는 miRNA 서열길이, 수소결합 수 등으로 잡아주었다. 이 방법을 사용함으로써 밝혀지지 않은 단백질과 miRNA의 상호작용 서열을 찾는데 이용가능 할 것이다.

I. 서론

miRNA는 21-25개의 뉴클레오타이드(nucleotide, nt)로 이루어진 단일 염기쌍으로 생물의 유전자 발현을 제어하는 조절물질이다. 따라서 세포가 증식하거나 분화 또는 자연사 등의 생명활동에도 영향을 미칠 수 있는데 이는 암과도 밀접하게 연관되어 있다는 것을 의미한다. 암이 생성되는 과정에서 miRNA는 암을 유발시킬 수 있는 종양유전자(oncogene)으로 작용할 수도 있고, 암을 억제 할 수 있는 종양억제유전자(tumor suppressor)로 작용할 수도 있다[1].

현재 miRNA에 대한 연구가 굉장히 활발하게 진행되고 있음에 따라 그에 대한 방대한 양의 정보와 데이터베이스들이 웹상에 나와 있다. 이들 miRNA와 단백질의 상호작용 정보도 그 중 하나인데, 양이 너무 많다 보니, 정작 유용한 정보를 골라내기가 쉽지 않다. 따라서 각각의 단백질에 작용하는 수많은 miRNA에 여러 개의 기준을 가지고 가중치를 주어 우선순위를 정해준다면 웹상에서 단순히 가져온 정보들을 한 번 더 필터링 해줌으로써 유용한 정보를 좀 더 쉽게 얻을 수 있을 것이다. 우리가 선행연구에서 추출해 낸 암 관련 단백질들에 대한 miRNA를 이용한다면 암의 증식이나 분화, 자연사에 대한 조절하기 위한 재료로 유용할 것이다.

II. 본론

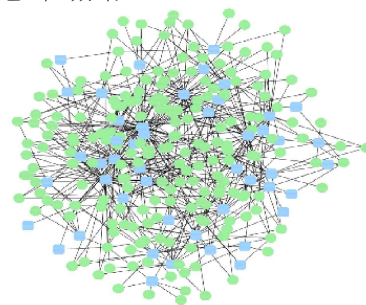
1. Dataset

암 관련 단백질 60개는 선행연구에서 추출하였고[2], 60개 단백질 중 48개의 단백질에 대해서만 상호작용하는 miRNA와 miRNA 서열 정보를 추출할 수 있었다. miRNA정보는 miRTarBase(<http://mirtarbase.mbc.nctu.edu.tw/>), TargetScan(<http://www.targetscan.org/>),

Mir2diseasebase(<http://www.mir2disease.org/>), TransmiR(<http://202.38.126.151/hmdd/mirna/tf/>)의 네 개의 웹사이트를 이용하였고 202개의 miRNA를 추출하였다.

2. 연구방법 및 연구결과

202개의 miRNA와 상호작용을 하는 48개의 단백질간의 관계를 보기 위하여 단백질-miRNA 상호작용 네트워크를 구축하였다 (그림 1). 파란색 네모난 노드는 단백질을 나타내고, 초록색 둥근 노드는 miRNA를 나타내며, 이들 사이의 연결 선수는 534개이다. 이 네트워크를 보면 한 개의 단백질이 여러 개의 miRNA와 상호작용을 하는 것을 알 수 있다.



▶▶ 그림 1. 단백질-miRNA 상호작용 네트워크

따라서 우리는 각각의 단백질이 가지고 있는 miRNA 우선순위를 정해주기 위해서 miRNA 서열정보에 가중치를 부여해 주었다. 가중치 부여 요인으로 miRNA 서열의 길이, miRNA 서열내의 수소결합 수, 하나의 miRNA 서열 내에서 C와 G 염기가 얼마나 잘 분산되어 있는지에 대한 값을 사용했다(표 1, 표 2). 먼저 miRNA 서열의 길이를 조사해 본 결과 거의 20-25개 사이로 염기 한 두 개 차이가 났다. 따라서 miRNA의 서열 길이가 단백

질과 상호작용을 하는데 미치는 영향을 크게 다르지 않을 것으로 간주하고 서열의 길이가 20일 때 '0'을 부여하고, 염기가 하나씩 늘어갈 때 마다 0.1씩 증가시켰다.

두 번째로 miRNA 서열 안에서의 수소결합 수의 경우, 2개의 수소결합을 가지는 염기A와 U(이하 퓨린계)에 비해 3개의 수소결합을 가지는 염기 C와 G(이하 피리미딘계)가 더 많을수록 miRNA와 단백질의 상호작용이 더 잘 일어날 것이다. 따라서 수소결합의 수를 식 (1)을 이용하여 계산해주었다.

$$\text{수소결합수} = (A + U) \times 2 + (C + G) \times 3 \quad (1)$$

대부분의 miRNA 서열내의 수소결합의 수는 50-60개 이었다. (0.05를 부여한 이유?) 각각의 수소결합 수에 0.05를 부여했다(표 1, 표 2).

표 1. CCND1에 대한 miRNA 가중치

miRNA	length	H-H		최종 합산
miR-503	23	59	1.6900	11.54
miR-17	23	57	1.2833	11.03
miR-302c	23	56	1.0000	10.70
miR-302a	23	55	0.7714	10.42
miR-421	23	56	0.6923	10.39
miR-20a	23	55	0.7071	10.36
miR-19a	23	54	0.6400	10.24
miR-155	23	55	0.5143	10.16
miR-193b	22	57	1.4444	8.69
miR-34a	22	55	1.4000	8.55
let-7b	22	54	1.1667	8.27
miR-16-1*	22	54	1.0833	8.18
miR-449a	22	54	1.0000	8.10
miR-15b	22	53	0.8308	7.88
miR-15a	22	53	0.7615	7.81
miR-16	22	54	0.5833	7.68
miR-16-1	22	54	0.5833	7.68
miR-424	22	52	0.6286	7.63
miR-590	22	52	0.5714	7.57
miR-340	22	50	0.3750	7.28

표 2. PTEN에 대한 miRNA 가중치

miRNAs	length	H-H		최종 합산
miR-17	23	57	1.2833	11.03
miR-221	23	57	1.1917	10.94
miR-217	23	56	1.0769	10.78
miR-18a	23	56	1	10.70
miR-19b	23	55	0.7714	10.42
miR-20a	23	55	0.7071	10.36
miR-19a	23	54	0.64	10.24
miR-214	22	57	2.3111	9.56
miR-25	22	55	1.5	8.65
miR-216a	22	54	1.1667	8.27
miR-494	22	54	0.9167	8.02
miR-141	22	53	0.9	7.95
miR-21	22	52	0.8	7.80
miR-26a	22	53	0.6923	7.74
miR-222	21	53	1.32	6.07

마지막 인자인 miRNA의 서열 내에서 피리미딘계의 분산정도이다. 퓨린계가 피리미딘계로 바뀌는 순간 또는 피리미딘계가 퓨린계로 바뀌는 순간에 0.1을 부여하는 방식으로 가중치를 부여했다. 예를 들어 'AGGCUGCGGGAACGUUC UGCAG'라는 miRNA 서열이 있다고 가정한다. 첫 번째 A다음에 G가 나왔으니 0.1 점, 세 번째는 똑같은 G가 나왔으니 0 점, 네 번째는 G와 같은 피리미딘계인 C가 나왔으니 0 점, 다섯 번째는 퓨린계의 U가 와서 다시 0.1 점을 주는 식으로 계산한 뒤 그 값들을 다 더하고 [피리미딘계 염기 수] / [퓨린계 염기 수]로 나뉘 주는 방식이다(표 1, 표 2). 이때 최대값을 넘지 않아야 하는데 최대값은 miRNA 서열 안에서 퓨린계의 수가 피리미딘계의 수보다 적을 경우에는 식(2)을 이용하였고, 클 경우에는 식(3)을 이용하였다. 예외적으로 서열 내에서 피리미딘계가 통틀어 한 개 일 경우나 퓨린계의 수가 피리미딘계의 수와 같을 경우에는 식 (4)를 이용하여 최대값을 구했다.

$$0.2 \times x \times \frac{[\text{서열전체길이}] - x}{x} \quad (2)$$

$$0.2 \times (y-1) \times \frac{[\text{서열전체길이}] - x}{x} \quad (3)$$

$$[0.2 \times (y-1) + 0.1] \times \frac{[\text{서열전체길이}] - x}{x} \quad (4)$$

x = 퓨린계 수 / y = 피리미딘계 수

위에서 설명한 세 가지 요인들을 다 더한 값이 각각의 miRNA가 가지는 가중치이다(표 1, 표 2의 5번째 열).

본 논문에는 대표적으로 CCND1과 PETN 단백질들과 상호작용하는 miRNA를 나타내었다. CCND1은 cell cycle 단계 중 G1기에서 S기로 넘어가는 시기에 cyclin D1-CDK4라고 불리는 단백질의 조절 구성요소인데, cyclin dependent kinsase(CDK)관련 단백질이 cell cycle에 중요한 역할을 하기 때문에 CCND1은 cell cycle 단계에서 핵심적이라고 볼 수 있다. PETN은 단백질이나 지질에 phosphate를 떨어지게 하는 protein phosphatase, lipid phosphatase의 하나 종류로, 가장 중요한 기능은 tumor suppressor이다. 따라서 이들 단백질과 상호작용하는 miRNA에 가중치를 준다면, 더 높은 가중치 점수를 가진 miRNA의 작용이 활발하게 이루어 질 수 있을 거라고 본다.

III. 결론

본 논문에서는 암관련 단백질과 상호작용하는 miRNA에 대해 가중치를 부여함으로써 우선순위에 따라 miRNA를 사용하여 암을 조절하는데 유용한 miRNA 정보를 제공하였다. 이는 단백질과 miRNA간의 복잡한 상호작용들 중에서 좀 더 가능성 있는 후보 miRNA를 도출하는데 적합하다. 본 연구에서 가중치를 주는데 사용한 3가지의 기준은 비교적 보편화된 방법이라고 생각될지도 모른다. 따라서 향후 연구에서는 p-value 개념을 도입하여 신뢰도를 높은 값으로 개선할 것이다.

■ 참고 문헌 ■

- [1] microRNAs as oncogenes and tumor suppressor
[2] 단백질 60개 추출한 논문(윤경언)