



GIST(광주과학기술원) 보도자료

<http://www.gist.ac.kr>

| | | |
|---------|-------------------------|------------------------------|
| 보도 일시 | 배포 즉시 보도 부탁드립니다. | |
| 보도자료 담당 | 대외협력팀 김미연 팀장 | 062-715-2020 / 010-5302-3620 |
| | 대외협력팀 이나영 행정원 | 062-715-2024 / 010-2008-2809 |
| 자료 문의 | 전기전자컴퓨터공학부 남호정 교수 | 062-715-2641 / 010-2570-3580 |

약물-표적 단백질 상호작용 판별 인공지능 모델 개발

- 약물-표적 단백질 상호작용 예측.. 기존 실험 방법 대비 초기 약물 후보 발견까지 10~20배 성능 향상
- GIST 남호정 교수 연구팀, 생물정보학 분야 상위 저널 <PLOS Computational Biology>에 논문 게재

□ GIST(지스트, 총장 김기선) 전기전자컴퓨터공학부 남호정 교수 연구팀이 약물 개발의 초기 단계에서 수행되는 약물-표적 단백질 상호작용* 판별을 인공지능 모델로 수행하는 방법론을 제시하였다.

*약물-표적 단백질 상호작용: 표적 단백질이 주어진 약물과 실제 결합하는지 여부 (표적 단백질 상호작용: 화합물에 의해 생물학적 기작이 조절되는 단백질)

○ 연구팀은 기존에 실험적으로 수행되던 약물-표적 단백질 상호작용 판별을 합성곱 신경망(Convolutional neural network) 인공지능 모델을 이용하여 예측했다.

□ 약물 개발의 초기 단계에서 질병에 관여한다고 생각되는 표적 단백질이 선정된 후, 표적 단백질과 상호작용하는 화합물을 선별하게 된다. 실험적으로 수행되는 약물-표적 단백질 상호작용의 판별은 많은 시간과 비용을 소모한다. 또한, 실험적인 약물-표적 단백질 상호작용의 판별은 주어진 화합물 라이브러리*에서 무작위 화합물에 대하여 수행되기 때문에 실제 상호작용하는 화합물이 선별되는 비율은 매우 낮은 편이다.

*화합물 라이브러리: 약물개발 중 초기 화합물 선별과정에서 사용되는 화합물 세트

○ 약물과의 상호작용에서 단백질의 모든 서열이 관여하지는 않는다. 그러나 기존에 단백질을 표현하는 컴퓨터 모델들은 단백질의 모든 서열과 그에 기반한 물리-화학적 속성을 이용하였다. 따라서 본 연구에서는 합성곱 신경망

(Convolutional Neural Network, CNN)* 을 사용하여 약물과 상호작용하는 지역적인 서열 패턴을 추출하고, 이를 약물-표적 단백질 상호작용의 예측에 이용하였다.

*합성곱 신경망: 인간의 신경망을 본 뜬 딥러닝 기술 중 하나로 이미지 인식과 음성 인식 등 다양한 곳에 사용됨

- 24,000개 이상의 약물-표적 단백질 데이터를 시험한 결과 약 80%의 정확도를 보여주었다. 또한, 합성곱 신경망의 결과를 해석하여 추출된 지역 서열 패턴이 실제로 약물-표적 단백질 상호작용에 관련하는 부분임을 확인할 수 있었다.

- 현재 남호정 교수와 과학기술정보통신부 ‘인공지능 기반 신약개발 플랫폼 구축’ 과제를 공동으로 수행하고 있는 GIST 생명과학부 김용철 교수와 GIST 화학과 안진희 교수는 해당 인공지능 모델을 통하여 제시된 화합물에 대하여 실제 스크리닝 실험을 수행하였으며, 그 결과 키나아제(kinase), 이온 통로(ion channel), G-단백질 결합 수용체(G-protein coupled receptor, GPCR) 등의 단백질 군에서의 hit*의 비율이 10~20배의 상승하는 효과를 확인하였다.

*hit: 대단위 스크리닝 기법 중 초기 활성 물질

- 남호정 교수는 “이번 연구는 약물 개발의 초기 단계에서 수행되는 약물-표적 상호작용 판별을 인공지능을 이용하여 수행한 것”이라며, “기존의 무작위 스크리닝방식이 아닌 인공지능을 통해 예측된 후보군에 대하여 실험을 수행하면 약물 개발의 시간적 비용적 효율성을 크게 제고할 수 있을 것이다”라고 말했다.

- 현재 개발된 인공지능 모델은 한국화학연구원 연구진(김동욱 책임)과 공동 개발 중인 신약개발 플랫폼에 공식 탑재될 예정이며, 특정 단백질에 대한 상호작용 판별을 화합물은행 화합물에 적용하여 올해 말 서비스를 시작할 계획이다.

- GIST 전기전자컴퓨터공학부 남호정 교수(교신저자)가 주도하고, 이인구 연구원(공동 1저자)과 금종수 연구원(공동 1저자)이 참여한 이번 연구는 과학기술정보통신부(MSIP)의 바이오·의료기술개발사업과 전통천연물기반유전자-동의보감사업(Bio-Synergy Research Center)의 지원을 받아 수행되었으며, 연구성과는 생명정보학의 저명 학술지 PLoS Computational Biology(IF 3.95)에 2019년 6월 14일(금) 온라인 게재되었다. <끝>

논문의 주요 내용

1. 논문명, 저자정보

- 저널명 : PLoS Computational Biology
- 논문명 : DeepConv-DTI: Prediction of drug-target interactions via deep learning with convolution on protein sequence
- 저자 정보 : 이인구 연구원(공동 1저자, GIST), 김종수 연구원(공동 1저자), 남호정 교수(교신저자, GIST)

용어 설명

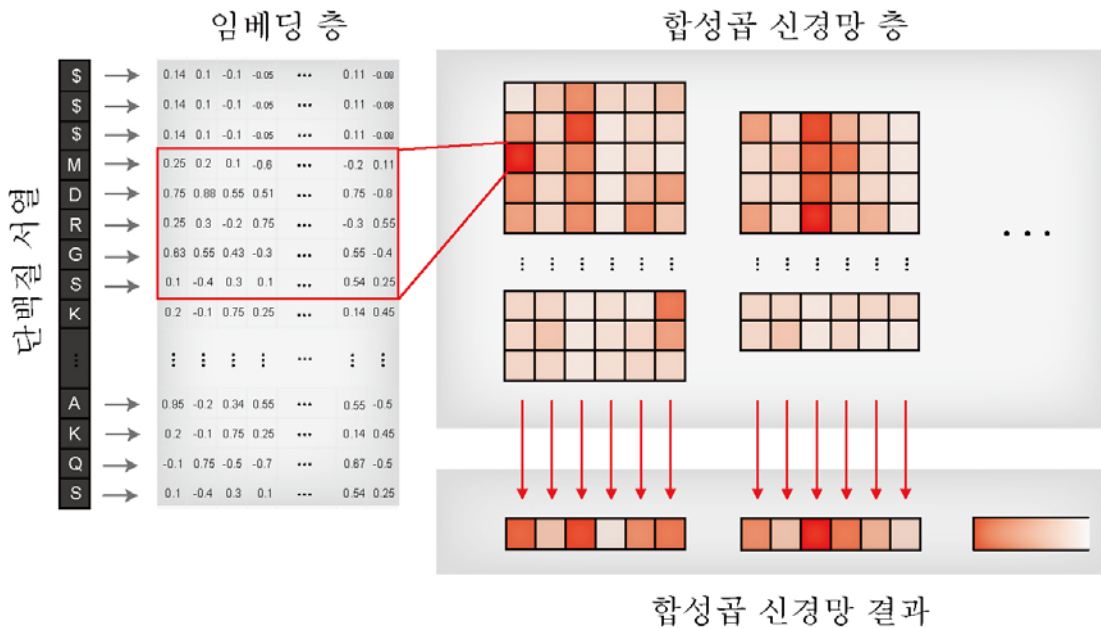
1. 약물-표적 단백질 상호작용 (Drug-target interaction)

- 약물이 실제로 인체 내에서 작용하기 위해서는 특정 단백질과 결합하여, 해당 단백질의 형태를 변화시키고, 최종적으로 그 단백질이 수행하는 생물학적 기작을 변화시킬 수 있어야 한다. 따라서, 어떤 화합물이 단백질과 결합하는지의 여부(약물-표적 단백질 상호작용)를 판별하는 것이 약물 개발에 있어 기초적이고 중요한 단계라고 할 수 있다.
- 그러나, 대규모 화합물 라이브러리에 대한 실험적인 검증은 비용과 시간이 많이 들며, 이를 해결하기 위하여 인공지능을 통한 약물-표적 단백질 상호작용 판별 연구가 진행되고 있다.

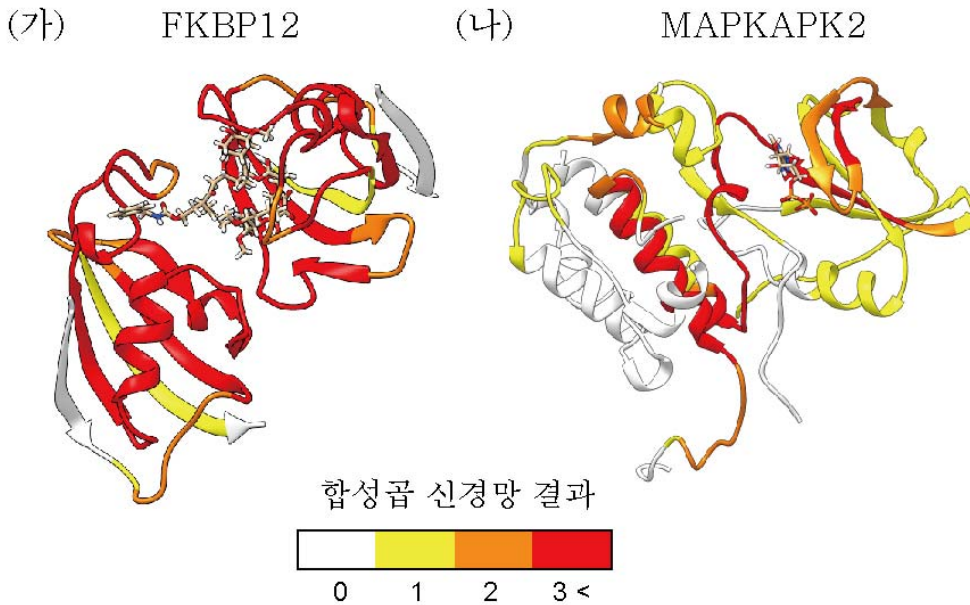
2. 합성곱 신경망 (Convolutional neural network, CNN)

- 합성곱 신경망은 인간 시각 인지를 모사한 신경망의 일종으로 지역적인 패턴을 파악하는데 사용되고 있으며, 이미지와 컴퓨터 비전의 경우에는 객체 검출부터 분류문제등 다양한 분야에서 활용되고 있다.
- 본 연구에는 합성곱 신경망을 이미지가 아닌 단백질 시퀀스에 적용함으로써 단백질의 지역 서열 패턴을 파악하고자 하였다.

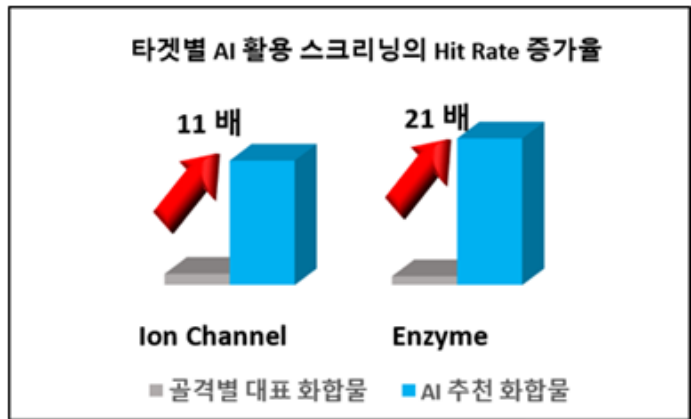
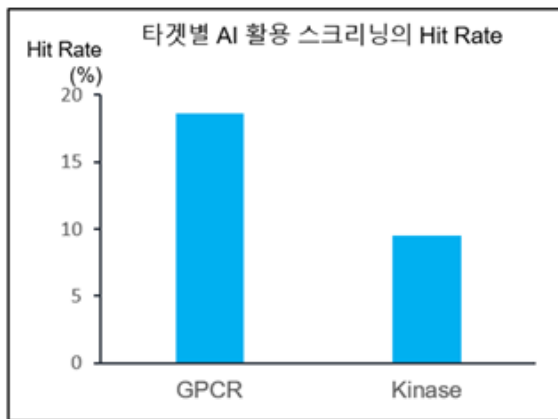
그림 설명



[그림 1] 합성곱 신경망을 이용한 단백질의 지역 서열 패턴 추출. 단백질 서열의 각각 아미노산은 임베딩 층에서 벡터로 치환되며, 치환된 단백질 서열상에서 합성곱 신경망을 통하여 약물과 상호작용할 때 중요한 지역 서열 패턴을 찾아내게 된다.



[그림 2] 단백질 FKBP12과 MAPKAPK2 상에서의 합성곱 신경망 결과를 해석한 결과, 화합물의 결합 지역이 그 외 지역보다 강조됨을 확인할 수 있었다.



[그림 3] GIST 생명과학부 김용철 교수 연구실과 GIST 화학과 안진희 교수 연구팀의 스크리닝 결과 AI 추천 화합물이 GPCR, Kinase 군에서 높은 hit 비율을 보여주었으며, (좌) Ion Channel, Enzyme 군에서는 골격별 대표 화합물 라이브러리에 비해 10~20배의 높은 hit 비율을 보여주었다. (우)