



즉시 / 2022. 5. 11.(수)

문의: 김찬우 학사  
(현 워싱턴대, +1-206-532-6151, [chanwkim@cs.washington.edu](mailto:chanwkim@cs.washington.edu)) / 제1저자  
이한빈 학사과정  
(서울대 의대, [hanbin973@snu.ac.kr](mailto:hanbin973@snu.ac.kr) / 제1저자)  
연구책임자 한 범 교수  
(서울대 의대, 02-3668-7618, [buhm.han@snu.ac.kr](mailto:buhm.han@snu.ac.kr)) / 교신저자  
연구책임자 정기훈 교수  
(서울대 의대, 02-740-8207, [keehoon.jung@snu.ac.kr](mailto:keehoon.jung@snu.ac.kr)) / 교신저자

## 유전자의 바다를 향해하며 신약 개발 표적을 찾아내는 인공지능 알고리즘 ‘마르코폴로’

서울대 창의선도 신진 교수들과 학부생들,

단일세포 RNA 시퀀싱 분석의 신개념 AI (인공지능) 알고리즘 개발 성공  
연구결과는 국제학술지 『뉴클레익 애시드 리서치』 (IF=16.971) 최신호에 발표

서울대학교 신진 교수들 중 특히 우수한 연구로 영예의 ‘창의선도 신진연구자’로 발탁되었던 의과대학 한 범 교수, 정기훈 교수 연구팀은 AI (인공지능)을 적용하여 질병의 발생과 치료에 중요한 유전자를 발굴하는 ‘마르코폴로’ 알고리즘을 개발해 내었다. 항암 면역치료, 줄기세포 연구 등 주요 의생명 연구분야에서 최근 급부상 중인 단일세포 RNA 시퀀싱 분석법의 획기적 발전의 전기를 마련하여, 다양한 질환의 병인 기전을 밝히고 고부가가치를 창출하는 신약 개발 타겟 선정에 핵심적인 역할을 할 것으로 기대된다.

질환의 발병 기전을 연구하고 이를 통해 성공적인 신약 개발을 하려면, 질병의 치료 타겟이 될 수 있는 세포 종류와 “마커 유전자”를 발굴해야 하는데, 최근에는 단일세포 RNA 시퀀싱 기술이 개발되어 널리 활용되고 있다. 이와 같은 단일세포 전사체 분석은 세포 하나하나의 유전자 발현량을 개별적으로 측정할 수 있기 때문에, 희귀 세포 등 특정 세포군을 동정해내고, 그 특정 세포 종류에서만 발현되는 핵심 마커 유전자를 발굴 가능케 한다.

하지만 기존 단일세포 전사체 분석의 가장 큰 문제는, 각각의 세포 종류를 클러스터링하며 정의하는 과정에 주관적인 요소가 크게 작용하여, 세포 종류의 정확한 구분에 종종

오류를 범한다는 것이다. 즉, 실험에서 도출된 세포의 정보가 기본적으로 분류되지 않은 상태로 연구자에게 주어지기 때문에, 연구자들은 세포를 주관적 기준에 따라 혹은 기존 학계의 잘못된 보고에 따라 분류해주는 기존의 방법론을 우선 적용하게 되는데, 이 방법론의 낮은 정확도로 인하여 세포군들을 모호하게 정의하고 분류를 명확하게 하지 못하게 된다. 이는, 실제 생명현상과는 동떨어지는 그릇된 결과 분석을 초래하며, 궁극적으로 효율적인 신약개발 타겟의 발굴은 불가능하게 된다.

이러한 문제를 해결하기 위해 서울대 의대 한범/정기훈 교수 연구팀은, 세포 분류를 수행하지 않고 신약개발 타겟 마커 유전자를 효율적으로 발굴할 수 있는 AI (인공지능) 알고리즘 “마르코폴로”를 개발했다. 이 알고리즘은 단일세포 전사체 데이터를 통합적으로 분석하여 마커 유전자를 찾아내기 때문에, 기존의 선제적 및 인위적 세포 분류를 필요로 하지 않는다.

즉, 마커 유전자를 찾기 위해 인간의 수동적 세포 분류가 선행되어야 한다는 기존의 고정관념에서 탈피하고, 발상의 전환을 통해 인공지능으로 하여금 중요 유전자를 찾도록 한 것이다. 이렇게 함으로써 세포 분류를 할 때 발생할 수 있는 오류를 근본적으로 피할 수 있게 되었다. 따라서, 알고리즘이 어떠한 편견(bias)도 없이 데이터를 전체적으로 조망하며 유전자를 찾아내기 때문에, 드넓은 미지의 바다를 향해하며 신세계 아시아의 새로운 땅을 찾아내고 동방견문록을 편찬하게 된 “마르코 폴로”의 이름이 차용 되었다.

이례로 연구팀은 마르코폴로를 인간의 간 데이터에 적용하였고, 기존 방법론에서는 찾을 수 없었던 중요 마커 유전자(GNLY 유전자)를 찾아내었고, 또한 인간 줄기세포 데이터에 적용하여, 역시 기존 방법론에서는 동정할 수 없었던 마커 유전자(NODAL 유전자)를 발굴하는 데 성공하였다. 한범 교수와 정기훈 교수는, “단일세포 RNA 시퀀싱 분석법 기반 중요 마커 유전자의 발굴은 암과 면역질환 등의 발병 메커니즘을 심도 있게 이해하고, 이를 기반으로 효율적 신약개발을 성공적으로 진행하는 데에 매우 중요한 역할을 할 것” 이라고 말했다.

이번 연구는 한국연구재단의 중견연구자지원사업과 우수신진연구지원사업의 지원을 받아 진행되었으며, 인용지수 16.971의 『뉴클레익 애시드 리서치 (Nucleic Acids Research)』에 최근 게재되었다.

[붙임] 1. 연구결과 2. 그림설명

## 연구 결과

### MarcoPolo: a method to discover differentially expressed genes in single-cell RNA-seq data without depending on prior clustering

Chanwoo Kim<sup>1</sup>, Hanbin Lee<sup>1</sup>, Juhee Jeong, Keehoon Jung\* and Buhm Han\*

( *Nucleic Acids Research*, 2022

Online Article <https://doi.org/10.1093/nar/gkac216> )

<sup>1</sup> 공동 제1저자

\* 공동 교신저자

## 그림 설명

그림 1. 유전자의 바다를 향해하여 마커 유전자를 찾아내는 마르코폴로의 개념도.



(일러스트 상업적 사용 가능, 신문기사에 사용 가능)

그림 2. 마르코폴로 알고리즘의 작동 원리 도식도

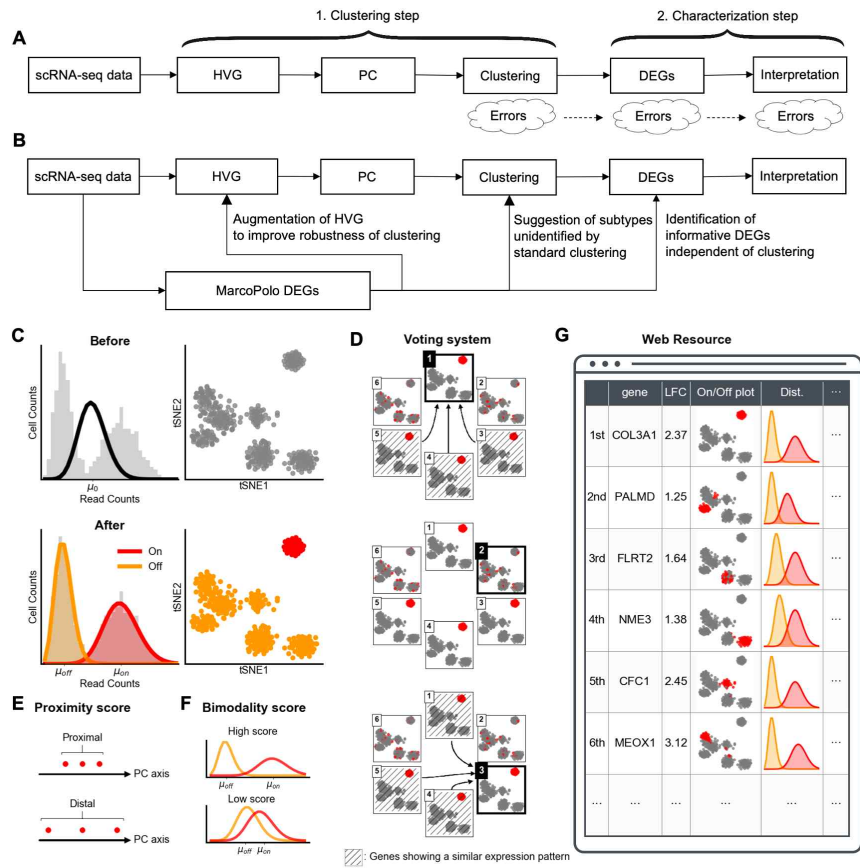


그림 3. 마르코폴로가 찾아낸 마커 유전자 (NODAL과 GNL1)

