

보도자료



보도일시	즉시 보도
	2023. 8. 11.(금)
문의	담당자: 최재용 연구원(02-740-8246)
	연구단장/연구책임자 의과대학 김종일 교수(02-740-8251) / 교신저자
	연구단/연구진 최재용, 김주현, 손호영 연구원(02-740-8246) / 공동 제1저자

동아시아인 임퓨테이션을 위한 레퍼런스 패널 및 GWAS 연구 발표

- 동아시아인 레퍼런스 패널 공개 서비스 시작-

■ 요약

연구성과/ 기대효과	<p>김종일 교수(서울대학교 의과대학 유전체의학연구소, 생화학교실), 임선화 교수(강원대학교 생화학교실), 박진호 교수(서울대학교 의과대학 가정의학교실) 등이 이끄는 연구진이 <u>2023년 동아시아인을 위한 레퍼런스 패널 (Northeast Asian Reference Database, NARD2)</u>을 제작하고, 이를 사용하는 전장유전체연관분석(GWAS) 연구를 통해 동아시아인 특이적인 신규 변이를 확인하였다는 연구결과를 발표했다.</p> <p>연구진이 제작한 신규 레퍼런스 패널은 14,393명의 전장유전체 시퀀싱(WGS) 데이터를 포함하고 있으며, 이중 11,000명이 동아시아인 유래 데이터로 현존하는 레퍼런스 패널 중에서 가장 많은 동아시아인을 포함하고 있다.</p> <p>해당 레퍼런스 패널을 72,298명의 한국인에 적용하여 시행한 GWAS 연구에서 키, 체질량지수(BMI), 혈압, 당뇨 관련 형질에 대해 39개의 신규 변이 및 다양한 유전자에 대한 연관성을 밝혔다. 추가로, 유전체 정밀 맵핑과 후성유전체 정보를 통해 논코딩 영역에 존재하는 변이에 대한 생물학적 조절 기작의 근거도 제시하였다.</p> <p>NARD2 레퍼런스 패널은 gmi.snu.ac.kr/imputation을 통해 누구나 사용할 수 있으며, 유전체 연구를 진행하는 동아시아인에게 중요한 자료가 될 것으로 기대한다.</p>
---------------	---

■ 본문

서울대학교 의과대학 유전체의학연구소 김종일 교수 (제1저자 최재용, 손호영 연구교수, 김주현 박사과정 대학원생), 강원대학교 임선화 교수, 분당서울대병원 서정선 교수 (제1저자 김성재, 유승근 박사), 서울대학교병원 가정의학과 박진호 교수, 국립암센터 김현진 교수 공동연구팀이 2023년 동아시아인을 위한 레퍼런스 패널 (Northeast Asian Reference Database, NARD2) 을 제작하고, 이를 사용한 전장유전체연관분석(GWAS, 용어설명 참조) 연구를 통해 동아시아인 특이적인 신규 변이를 확인한 연구 결과를 발표했다.

레퍼런스 패널은 개인의 유전자형 일부를 통하여, 다른 유전자형을 임퓨테이션(용어설명 참조) 할 수 있는 참조 자료로, 집단 규모의 유전체 연구를 하는데 필수적이다. 이전까지 발표된 레퍼런스 패널은 동아시아인이 상당히 적게 포함되어 이를 그대로 동아시아인에 적용하면 유추할 수 있는 유전형의 정확도가 떨어진다.

본 연구진은 총 14,393명의 전장유전체 데이터를 사용하여 NARD2 레퍼런스 패널을 제작하였고, 이중 동아시아인이 11,000명으로 현존하는 레퍼런스 패널 중에서 가장 많은 동아시아인을 포함하고 있다. 대규모의 다른 레퍼런스 패널과의 비교를 통해 NARD2가 한국인에서 임퓨테이션의 정확도가 가장 높다는 사실을 확인하였다.

GWAS는 레퍼런스 패널을 통해 확인한 다수의 유전자형과 특정 질병간 분석을 통해 연관성이 높은 변이를 찾아내는 연구 방법이다. 본 연구진은 질병관리본부로부터 분양받은 한국인 72,298 명의 유전체 데이터에 NARD2를 적용하여 임퓨테이션을 수행하고 이를 바탕으로, 키, 체질량지수 (BMI), 혈압, 당뇨 등 8가지 형질에 대해서 39개의 신규 변이를 포함하는 347개의 유전변이를 발견하였다. 신규 변이 39개 중에서 6개는 상당히 희귀한 변이로, 개발한 NARD2 패널이 한국인 희귀 유전변이 발굴에 충분히 활용 가능함을 확인하였음.

GWAS를 통해서 발견되는 변이 중 대부분은 논코딩 영역(용어설명 참조)에 위치하여 단백질에 영향을 주지 않기 때문에, 형질에 영향을 주는 생물학적 기전을 확인하기 어렵다. 이를 극복하기 위해, 본 연구진은 유전체 정밀 맵핑을 통해 인과성이 높은 변이를 탐색하고, 후성유전체(용어설명 참조) 정보를 다수 통합하여 유전자 발현에 직접적인 영향을 줄 수 있는 변이를 확인하였다. 이를 통해 GWAS에서 연관성을 확인한 변이가 전사인자의 결합에 영향을 주어 유전자의 발현을 조절하는 근거를 제시하였다.

본 연구에서 구축한 NARD2 레퍼런스 패널을 누구나 사용할 수 있도록 하는 임퓨테이션 서버를 gmi.snu.ac.kr/imputation를 통해 제공하고 있으며, 이를 통하여 동아시아인에 대한 유전체 연구에 활발히 사용될 것으로 기대한다.

이번 연구는 의생명과학 저널인 'Science Advances' (IF=13.6) 최신포에 온라인 출간되었다. (논문명: A whole-genome reference panel of 14,393 individuals for East Asian populations accelerates discovery of rare functional variants).

이번 연구는 한국연구재단을 통해 지원된 과학기술분야 기초연구사업 과제 중, 교육부의 재원으로 수행된 대학중점연구소, 기초과학연구역량강화 사업과, 과학기술정보통신부의 재원으로 수행된 중견연구사업의 지원으로 이뤄졌다.

□ 용어설명

전장유전체연관분석(GWAS) : 인간 유전체 전체 영역에 대하여 다수의 유전체 변이와 특정 형질의 연관성을 통계적으로 분석하는 기법. 각종 질병이나 특성의 유전자를 찾기 위하여 가장 흔하게 사용되는 방법이지만, 대규모의 샘플과 유전형 분석이 필요하다는 이유 때문에 대형 연구과제를 통해서만 수행할 수 있다는 단점이 있다.

임퓨테이션 : 이미 전장유전체서열을 알고 있는 레퍼런스 샘플이 많이 있을 경우, 이를 미리 학습시키면, 새로운 유전체서열 전체를 알고 있지 않고 일부만 가지고 있어도 나머지 서열에 대한 추정이 가능한데 이렇게 추정하는 방법을 임퓨테이션(imputation)이라고 한다. 마이크로어레이 등의 실험 기법으로 유전체 서열의 일부만 확보한 다음, 나머지 다른 서열을 임퓨테이션 할 경우 유전체서열 분석 비용을 절약할 수 있다.

논코딩 영역 : 인간 유전체서열 중에는 단백질을 구성하는 아미노산 서열을 결정하는 부위(코딩 영역)이 있고, 거기에 포함되지 않는 부위가 있는데 이를 논코딩 영역이라고 한다. 코딩 영역은 전체 서열의 1%를 조금 넘는 영역을 차지하고 있고 대부분은 논코딩 영역이다. 인간의 다양한 질병이나 형질에 논코딩 영역이 중요한 역할을 할 것이라고 생각되지만 이를 밝히는 방법이 제한적이어서 많은 연구가 진행되지 않고 있다.

후성유전체 : 유전체 서열의 변화가 없어도, 특정한 세포나 개체가 가지고 있는 형질이 그 다음 세대까지 전달되는 경우가 있는데 이를 후성유전(epigenetic)이라고 한다. 이는 대부분 유전체의 구조나 메틸화, 히스톤 단백질의 변화 등을 통해서 유전자 발현에 장기적인 영향을 주는 방식으로 이루어진다. 이러한 유전체의 특성을 후성유전체(epigenome)라고 한다.