

★ 2026년 6월 2일(화) 조간 (온라인 6.1.(월), 낮 12:00 이후) 보도해 주시기 바랍니다.

보도자료	 한국연구재단 National Research Foundation of Korea
청렴한 연구지원 신뢰받는 NRF	<대외홍보실> 김윤숙 팀장, 고흥숙 연구원 ☎ 042-869-6119

<자료문의> 중앙대학교 생명과학과 윤성일 교수(02-820-5163)

질병 유전자, 시가 콕 집어낸다

- 병인 찾고 신약 개발 돕는 '설명 가능한 AI 프레임워크' 개발 -

- 대규모 유전자 데이터를 스스로 학습해 질병의 신호 경로와 핵심 유전자를 정밀 예측하는 인공지능 분석 체계가 구축됐다.
- 한국연구재단(이사장 홍원화)은 중앙대학교(윤성일 교수, 전민승 박사, 이민석 석사과정), 성균관대학교(양시영 교수, 남지호 박사), 한양대학교(조찬미 교수) 공동연구팀이 대규모 전사체 데이터를 기반으로 질병 연관 경로를 정밀하게 예측하고, 각 경로의 핵심 유전자를 근거와 함께 제시하는 인공지능 분석 기술 '세인트GSE(SaintGSE)'를 개발했다고 발표했다.
- 이번 연구 성과는 과학기술정보통신부와 한국연구재단이 추진하는 기초연구실 지원사업과 해양수산부 어류 유전자교정 기술 개발 등의 지원을 받아 수행됐으며, 국제학술지 '골관절염 및 연골(Osteoarthritis and Cartilage, JCR 상위 1.7%)'에 4월 16일 온라인 게재되었다.
- 차세대 시퀀싱 기술의 발전으로 방대한 전사체(유전자 발현) 데이터가 축적되고 있으나, 이를 질병의 분자 기전과 신호전달경로 변화로 연결해 질병의 정확한 원인을 해석하는 데에는 여전히 한계가 존재했다.
- 특히 기존 유전자 집합 분석법은 사전에 정의된 유전자 목록과 통계적 유의성에 크게 의존하기 때문에, 질병 특이적 발현 패턴과 핵

심 유전자 기여도를 정밀하게 설명하는 데 어려움이 있었다.

- 따라서 방대한 데이터 속에서 질병 관련 경로를 스스로 찾아내고, 그 판단 근거를 유전자 수준에서 정량적으로 증명할 수 있는 설명 가능한 인공지능 분석 기술이 절실한 상황이었다.
- 연구팀은 오토인코더*와 트랜스포머** 구조를 결합한 인공지능 모델 SaintGSE를 구축하여, 복잡한 전사체 데이터 패턴으로부터 질병 관련 신호전달경로의 활성화 여부를 정밀하게 예측하는 데 성공했다. 특히 설명가능한 AI(XAI) 기법을 도입해 인공지능이 질병 경로를 예측할 때 어떤 유전자를 결정적인 근거로 삼았는지 그 기여도를 수치화하여 제시할 수 있도록 만들었다.

* 오토인코더 : 데이터의 효율적인 표현 방법을 스스로 학습하는 비지도 학습 기반의 인공지능경망 구조

** 트랜스포머 : 데이터(문장, 유전자 서열 등) 내의 구성 요소들이 서로 어떤 관계를 맺고 있는지 맥락을 한눈에 파악하는 인공지능 모델

- 실제 연구팀이 이 모델을 골관절염 및 천연물 처리 전사체 데이터에 적용한 결과, 질병이 발생하는 분자 기전을 정확히 해석해 냈을 뿐만 아니라 치료 후보물질이 몸속에서 어떻게 작용하는지 그 기전까지 효과적으로 탐색할 수 있음을 증명했다.
- 이는 기존의 단순 통계 분석을 넘어 AI가 질병 신호를 직접 예측하고 원인 유전자까지 규명하는 새로운 분석 체계를 완성한 것으로, 향후 환자 맞춤형 정밀의학 및 맞춤형 신약 개발의 효율성을 획기적으로 높일 수 있는 만능 플랫폼의 기반을 마련했다는 평가를 받는다.
- 윤성일 교수는 “이번 연구는 인공지능을 활용해 복잡한 유전자 데이터 속에서 질병을 유발하는 핵심 경로와 원인 유전자를 설명 가능한 방식으로 도출해 낸 성과다”라며, “앞으로 다양한 신약 후보 물질의 작용 원리를 예측하는 정밀의학 분야의 핵심 기술로 널리 활용될 것으로 기대된다”고 밝혔다.

주요내용 설명

<작성 : 중앙대 전민승 박사, 성균관대 남지호 박사,
성균관대 양시영 교수, 중앙대 윤성일 교수>

논문명	SaintGSE: Transformer-based efficient and explainable gene set enrichment analysis
저널명	Osteoarthritis and Cartilage
키워드	Self-Attention and Intersample Attention Transformer (SAINT; 자기), eXplainable Artificial Intelligence (XAI; 설명가능 인공지능), Signaling pathway (신호전달경로), Osteoarthritis (퇴행성 관절염), Natural product (천연물)
DOI	10.1016/j.joca.2026.04.009
저자	윤성일 교수(교신저자/중앙대학교), 양시영 교수(교신저자/성균관대학교), 전민승 박사(제1저자/중앙대학교), 남지호 박사(제1저자/성균관대학교), 이민석 석사과정(공동저자/중앙대학교), 조찬미 교수(공동저자/한양대학교)

1. 연구의 필요성

- 차세대 시퀀싱 기술의 비약적인 발달에 따라 대규모 유전자 발현 데이터가 빠르게 축적되고 있으나, 이를 통합적으로 활용하여 질병의 생물학적 기전과 신호전달경로 수준에서 정밀하게 해석하는 데에는 여전히 한계가 있음
- 기존의 ORA(over-representation analysis) 및 GSEA(gene set enrichment analysis)와 같은 유전자 집합 분석법은 사전에 정의된 유전자 집합의 통계적 풍부도에 주로 의존하므로, 복잡한 전사체 패턴과 질병 특이적 분자 작동 원리를 충분히 반영하기 어려움
- 특히 골관절염과 같은 복합 질환은 노화, 염증, 연골 분해 등 다양한 생물학적 기전이 동시에 작용하므로, 질병과 관련된 신호전달경로를 정밀하게 예측하고 핵심 유전자를 함께 규명하는 분석 기술이 필요함
- 그러나 기존 분석법만으로는 어떤 신호전달경로가 실제 질병 상태와 밀접하게 연관되는지, 또 어떤 유전자가 해당 경로의 활성화에 핵심적으로

기여하는지를 체계적으로 제시하는 데 한계가 있음

- 설명가능 인공지능은 이러한 한계를 보완할 수 있는 유망한 접근법으로 주목받고 있으나, 많은 AI 모델은 예측 결과를 생물학적으로 해석하기 어렵거나 계산 비용이 커 실제 생명정보 분석에 적용하는 데 제약이 있음
- 이에 연구팀은 인공지능 기반 분석 모델 SaintGSE를 개발하여, 질병 관련 유전자 신호를 더 정밀하게 읽고 계산비용을 최적화한 설명가능 인공지능의 적용을 통해 핵심 유전자까지 함께 제시할 수 있는 새로운 분석 기술의 가능성을 확인하고자 함

2. 연구내용

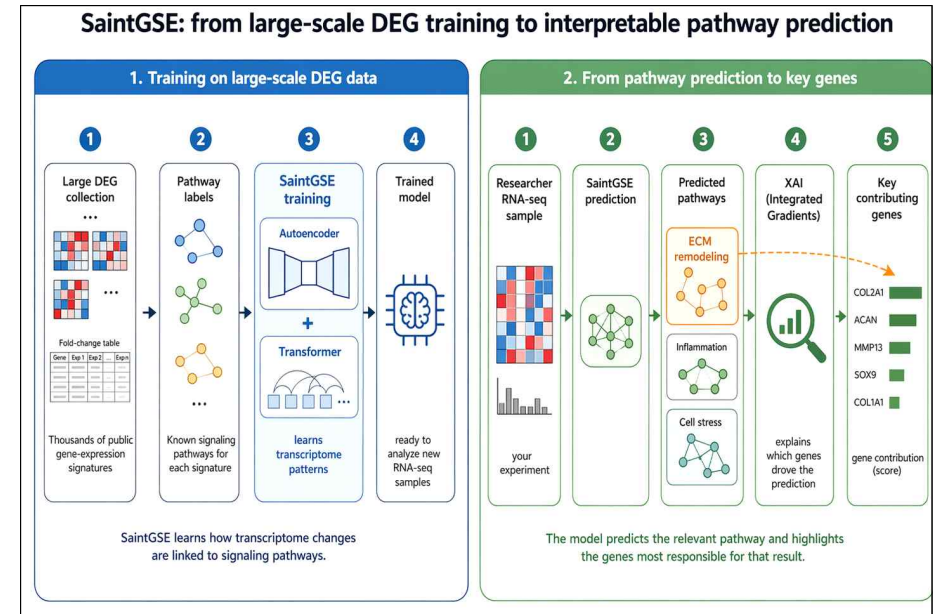
- 기존 유전자 집합 분석법의 한계를 보완하고 질병과 관련된 유전자 신호를 보다 정밀하게 해석하기 위해, 연구팀은 대규모 유전자 발현 데이터를 기반으로 신호전달경로의 활성 여부를 예측할 수 있는 인공지능 기반 분석 프레임워크 SaintGSE를 개발함.
- SaintGSE의 성능은 동일한 입력 조건과 평가 조건에서 기존 모델들과 비교 검증함. 특히 최근 표 형식(tabular) 데이터 분석에서 높은 성능으로 주목받는 딥러닝 모델인 FT-Transformer, TabNet, NODE와, 실제 예측 문제에서 강력한 기준 모델로 널리 활용되는 XGBoost, LightGBM, CatBoost와 비교한 결과, SaintGSE는 신호전달경로 예측 과제에서 전반적으로 우수한 예측 성능을 보였음. 이는 SaintGSE가 최신 표형 데이터용 딥러닝 모델 및 대표적인 기계학습 모델과 비교해도 경쟁력 있는 분석 성능을 갖추고 있음을 의미함.
- 특히 SaintGSE는 오토인코더와 자기 주목 및 샘플 간 주목 기반 트랜스포머 구조를 결합하여, 대규모 유전자 발현 정보로부터 신호전달경로의 활성 여부를 예측하고, 설명가능 인공지능 기법인 Integrated Gradients를 적용해 각 경로 예측에 기여한 핵심 유전자를 함께 제시할 수 있도록 설계함. 이를 통해 기존 통계 기반 분석으로는 파악하기 어려웠던 경로별 핵심 유전자를 보다 체계적으로 해석할 수 있음을 확인함.

- 연구진은 골관절염 전사체 데이터와 천연물 처리 데이터를 SaintGSE에 적용하여, 샘플별로 골관절염 치료와 관련될 수 있는 후보 유전자를 탐색함. 그 결과 SPIDR과 SH3BP4가 주요 후보 유전자로 예측되었으며, 이들은 이전 연구에서도 골관절염 관련 타겟 후보로 보고된 바 있음을 확인함. 또한 SaintGSE는 세포외기질(ECM) 관련 신호전달경로의 변화를 포착하여, 해당 천연물이 골관절염 증상 완화에 기여할 가능성을 제시함.
- 또한 mouse 및 human chondrocyte 실험과 mouse 비임상실험에서 SaintGSE가 예측한 경로 변화 양상이 실제 분자실험 결과와 부합함을 확인함으로써, 모델 예측의 생물학적 타당성을 검증함.

3. 연구성과/기대효과

- 본 연구는 대규모 유전자 발현 데이터로부터 질병 관련 신호전달경로를 예측하고, 해당 경로에 기여하는 핵심 유전자를 함께 제시할 수 있는 설명가능 인공지능 기반 분석 체계를 구축함으로써, 기존 유전자 집합 분석의 한계를 보완하고 질병 기전 해석의 새로운 가능성을 제시함.
- 특히 골관절염과 천연물 처리 데이터에서 예측된 세포외기질(ECM) 관련 신호전달경로 변화가 실제 mouse 및 human chondrocyte 실험, mouse 비임상실험 결과와 부합함을 확인함으로써, SaintGSE가 질병 관련 분자기전 해석과 후보 작용 기전 탐색에 활용될 수 있음을 제시함.
- SaintGSE는 향후 질병의 핵심 유전자와 신호전달경로를 보다 정밀하게 찾아내는 데 활용될 수 있으며, 바이오마커 발굴, 신규 치료표적 탐색, 약물 반응 예측 등 다양한 바이오 연구로 확장될 수 있을 것으로 기대됨.

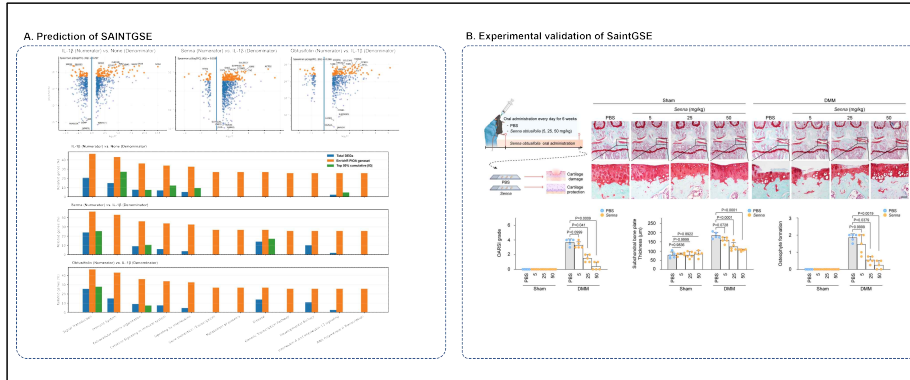
그림 설명



(그림1) SaintGSE의 학습 및 예측 프레임워크

SaintGSE는 대규모 DEG 데이터로부터 유전자 발현 변화와 신호전달경로의 연관성을 학습한 뒤, 연구자의 RNA-seq 샘플에 적용해 질병 관련 신호전달경로를 예측함. 또한 설명가능 인공지능 기법을 활용해 예측 결과를 유도한 핵심 유전자를 함께 제시함으로써, 질병 기전 해석과 후보 작용 기전 탐색이 가능하게 함.

그림설명 및 그림제공 : 중앙대학교 전민승 박사, 성균관대학교 남지호 박사, 성균관대학교 양시영 교수, 중앙대학교 윤성일 교수



(그림2) SaintGSE 예측 결과와 동물실험을 통한 천연물의 골관절염 완화 가능성 검증

A. SaintGSE 분석을 통해 천연물 처리 후 세포외기질(ECM) 관련 신호전달경로 변화와 핵심 기여 유전자를 예측하고(위 그림), 이를 기존 경로 분석 결과와 비교하여 제시함(아래 그림).

B. 골관절염 동물 모델에서 천연물 투여 후 연골 손상과 병리학적 지표가 완화됨을 확인하여, SaintGSE 예측 결과의 생물학적 타당성을 검증함.

그림설명 및 그림제공: 중앙대학교 전민승 박사, 성균관대학교 남지호 박사, 성균관대학교 양시영 교수, 중앙대학교 윤성일 교수

연구 이야기

<작성 : 중앙대 전민승 박사, 성균관대 남지호 박사, 성균관대 양시영 교수, 중앙대 윤성일 교수>

□ 연구를 시작한 계기나 배경은?

차세대 시퀀싱 기술의 발달로 다양한 질병과 생물학적 조건에서 대규모 유전자 발현 데이터가 빠르게 축적되고 있다. 그러나 이러한 데이터를 질병의 신호전달경로와 분자 기전 수준에서 해석하는 데에는 여전히 한계가 있었다.

기존 유전자 집합 분석법은 생명과학 연구에서 널리 활용되어 왔지만, 주로 사전에 정의된 유전자 집합과 통계적 유의성에 기반하기 때문에 복잡한 전사체 패턴과 질병 특이적 유전자 기여도를 충분히 설명하기 어려웠다. 연구팀은 이러한 한계를 보완하고자, 대규모 전사체 데이터를 직접 학습해 질병 관련 경로를 예측하고 그 근거가 되는 핵심 유전자를 설명할 수 있는 AI 기반 분석 기술을 개발하고자 하였다.

□ 연구 전개 과정에 대한 소개

연구팀은 먼저 다양한 공개 유전자 발현 데이터를 수집하고, 각 데이터가 신호전달경로와 어떻게 연결될 수 있는지 학습 가능한 형태로 정리하였다. 이후 고차원 유전자 발현 데이터를 효과적으로 압축하기 위해 오토인코더를 적용하고, 전사체 패턴을 학습하기 위해 자기주의 및 샘플 간 주의 기반 트랜스포머 구조를 결합하였다.

이를 통해 개발된 SaintGSE는 전사체 데이터로부터 질병 관련 신호전달경로의 활성 여부를 예측하도록 설계되었다. 또한 Integrated Gradients 기반 설명가능 인공지능 분석을 적용하여, 특정 경로 예측에 어떤 유전자가 기여했는지를 정량적으로 제시할 수 있도록 하였다. 연구팀은 골관절염 및 천연물 처리 전사체 데이터에 SaintGSE를 적용하여, 질병 기전 해석과 후보 작용 기전 탐색에 활용 가능함을 확인하였다.

□ 연구하면서 어려웠던 점이나 장애요소는 무엇인지? 어떻게 극복(해결)하였는지?

가장 큰 어려움은 대규모 전사체 데이터의 복잡성과 이질성이었다. 공개 데이터는 실험 조건, 질병 종류, 분석 플랫폼, 샘플 품질 등이 서로 달라 하나의 모델이 일관되게 학습하기 어려운 구조를 가지고 있었다. 연구팀은 데이터 전처리와 표준화 과정을 체계화하고, 다양한 데이터 조건에서도 적용 가능한 학습 구조를 구축함으로써 이러한 문제를 보완하고자 하였다.

또 다른 어려움은 AI 모델의 예측 결과를 생물학적으로 해석하는 문제였다. 단순히 예측 정확도가 높은 모델을 만드는 것만으로는 실제 생명과학 연구에 활용되기 어렵기 때문에, 모델이 어떤 유전자를 근거로 특정 경로를 예측했는지를 함께 제시하는 것이 중요했다. 연구팀은 설명가능 인공지능 기법을 적용하여 경로별 핵심 유전자 기여도를 산출함으로써, 예측 결과의 해석

가능성을 높이고자 하였다.

□ 이번 성과, 무엇이 다른가?

이번 연구의 차별성은 기존의 통계 기반 유전자 집합 분석을 넘어, 대규모 전사체 데이터를 학습한 AI 모델이 질병 관련 신호전달경로를 직접 예측하고 그 근거가 되는 핵심 유전자까지 제시했다는 점이다.

기존 분석법이 주로 유전자 목록의 통계적 풍부도를 평가하는 데 초점을 맞췄다면, SaintGSE는 전체 전사체 패턴을 기반으로 경로 수준의 변화를 예측하고, 각 예측에 기여한 유전자를 정량적으로 제시할 수 있다. 이를 통해 샘플 특이적 질병 기전 해석뿐 아니라 바이오마커 후보 탐색, 치료표적 발굴, 약물 또는 천연물의 작용 기전 분석 등에 활용될 수 있는 가능성을 제시하였다.

□ 실용화된다면 어떻게 활용될 수 있나? 실용화를 위한 과제는?

SaintGSE가 실용화된다면, 질병 전사체 데이터를 기반으로 어떤 신호전달경로가 변화했는지 예측하고, 그 변화에 관여하는 핵심 유전자를 탐색하는 분석 도구로 활용될 수 있다. 특히 복합 질환의 분자 기전 해석, 바이오마커 발굴, 신규 치료표적 탐색, 약물 반응 분석, 천연물 또는 후보물질의 작용 기전 예측 등에 응용될 수 있다.

실용화를 위해서는 다양한 질환과 조직, 임상 데이터에서 모델의 일반화 가능성을 추가로 검증할 필요가 있다. 또한 현재 모델은 경로 활성 여부를 이진 분류 형태로 예측하므로, 향후 경로 활성의 강도나 단계적 변화를 반영할 수 있는 모델로 고도화할 필요가 있다. 대규모 데이터 학습에 필요한 시간과 계산 비용을 줄이기 위한 모델 경량화와 학습 효율 개선도 중요한 후속 과제이다.

□ 꼭 이루고 싶은 목표나 후속 연구계획은?

후속 연구에서는 SaintGSE를 다양한 질환과 생물학적 조건으로 확장하여, 보다 범용적인 전사체 기반 경로 해석 플랫폼으로 발전시키고자 한다. 또한 단일 전사체 데이터뿐 아니라 단일세포 전사체, 공간전사체, 단백질체 등 다양한 오믹스 데이터와 연계하여 질병 기전을 보다 정밀하게 해석할 수 있는 방향으로 연구를 확장할 계획이다.

궁극적으로는 SaintGSE가 연구자들이 대규모 유전자 발현 데이터를 보다 쉽게 해석하고, 질병의 핵심 경로와 치료표적 후보를 발굴하는 데 도움을 주는 분석 기술로 활용되기를 기대한다.

□ 기타 특별한 에피소드가 있었다면?

이번 연구는 생명과학적 질문과 인공지능 모델링을 연결하는 과정에서 많은 시행착오를 거치며 진행되었다. 특히 단순히 성능이 높은 AI 모델을 개발하는 것에 그치지 않고, 생명과학 연구자가 실제로 해석하고 활용할 수 있는 결과를 제시하는 것이 중요하다는 점을 연구 과정에서 지속적으로 확인하였다.

연구팀은 모델의 예측 결과를 경로 수준에서 해석하고, 다시 유전자 수준의 근거로 연결하는 과정을 반복적으로 검토하였다. 이러한 과정은 시간이 오래 걸렸지만, SaintGSE가 단순한 예측 모델이 아니라 질병 기전 해석을 돕는 설명 가능한 분석 프레임워크로 발전하는 데 중요한 기반이 되었다.