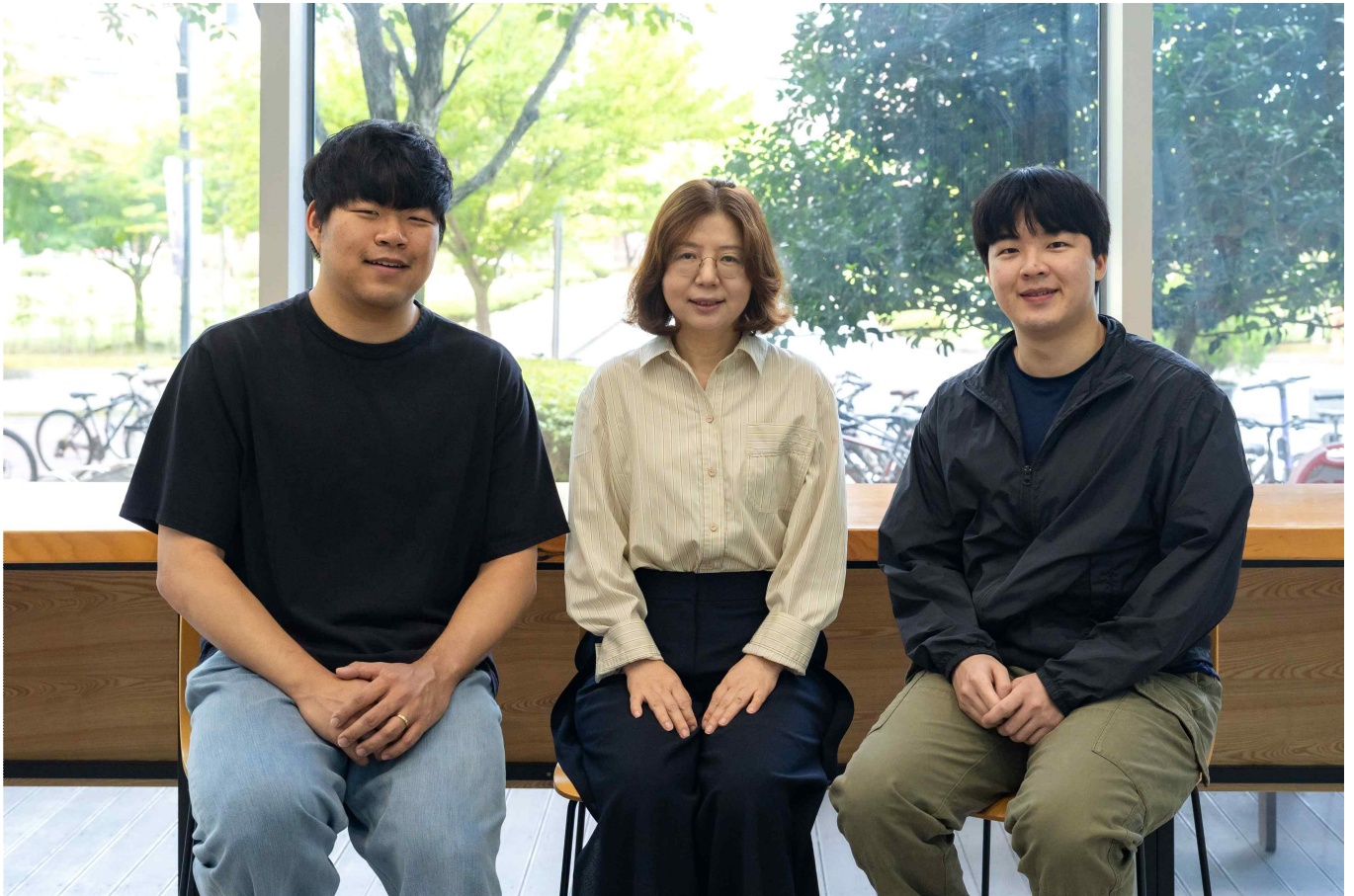


“조직 사진에선 안 보이는 차이, 유전자 정보와 함께 찾는다”

GIST, 유전자 발현 예측하고, 조직 영역 구분하는 AI 개발

- AI학과 이현주 교수팀, 약 98만 개 조직 이미지-유전자 정보 쌍 학습한 AI 프레임워크 ‘ST-ConMa’ 개발... 두 정보의 특징과 서로의 관계까지 함께 학습
- 유전자 발현 예측 12개 평가 지표 중 10개에서 최고 성능... 뇌 조직 영역 구분에서도 기존 모델 대비 정확도 향상
- 국제학술지《Briefings in Bioinformatics》 게재



▲ (왼쪽부터) GIST AI학과 최동하 박사, 이현주 교수, 염태림 석박사통합과정생

광주과학기술원(GIST·지스트, 총장 임기철)은 **AI학과 이현주 교수 연구팀**이 사람의 장기와 신체 부위를 이루는 조직의 이미지와 유전자 활동 정보를 함께 활용해, 겉으로 보이는 형태만으로는 알기 어려운 조직 내 영역별 차이를 분석하는 인공지능(AI) 프레임워크 ‘에스티 콘마(ST-ConMa)’를 개발했다고 밝혔다.

우리 몸의 조직은 같은 장기 안에서도 위치에 따라 세포의 종류와 활발하게 작동하는 유전자가 다르다. ‘조직의 모습을 담은 이미지’와 어떤 유전자가 어느 위치에서 얼마나 활발하게 작동하는지를 나타내는 ‘유전자 발현’ 정보를 함께 살펴보면, 위치마다 나타나는 이러한 차이를 더 세밀하게 파악할 수 있다.

ST-ConMa는 조직의 형태와 유전자 활동 사이의 관계를 학습해, 조직 이미지에서 유전자 발현을 예측하고 서로 다른 특성을 가진 영역을 구분하는 등 다양한 공간

전사체* 분석에 활용할 수 있다.

* **공간전사체(Spatial Transcriptomics)**: 조직의 어느 위치에서 어떤 유전자가 얼마나 활발하게 발현되는지를 조직 이미지와 함께 분석하는 기술이다.

최근에는 이러한 공간전사체 데이터를 AI로 분석하는 연구가 활발하지만, **기존 모델** 상당수는 유전자 발현 예측이나 조직 영역 구분 등 **특정 분석에 맞춰 개발**돼 다른 분석에 폭넓게 활용하기 어려웠다.

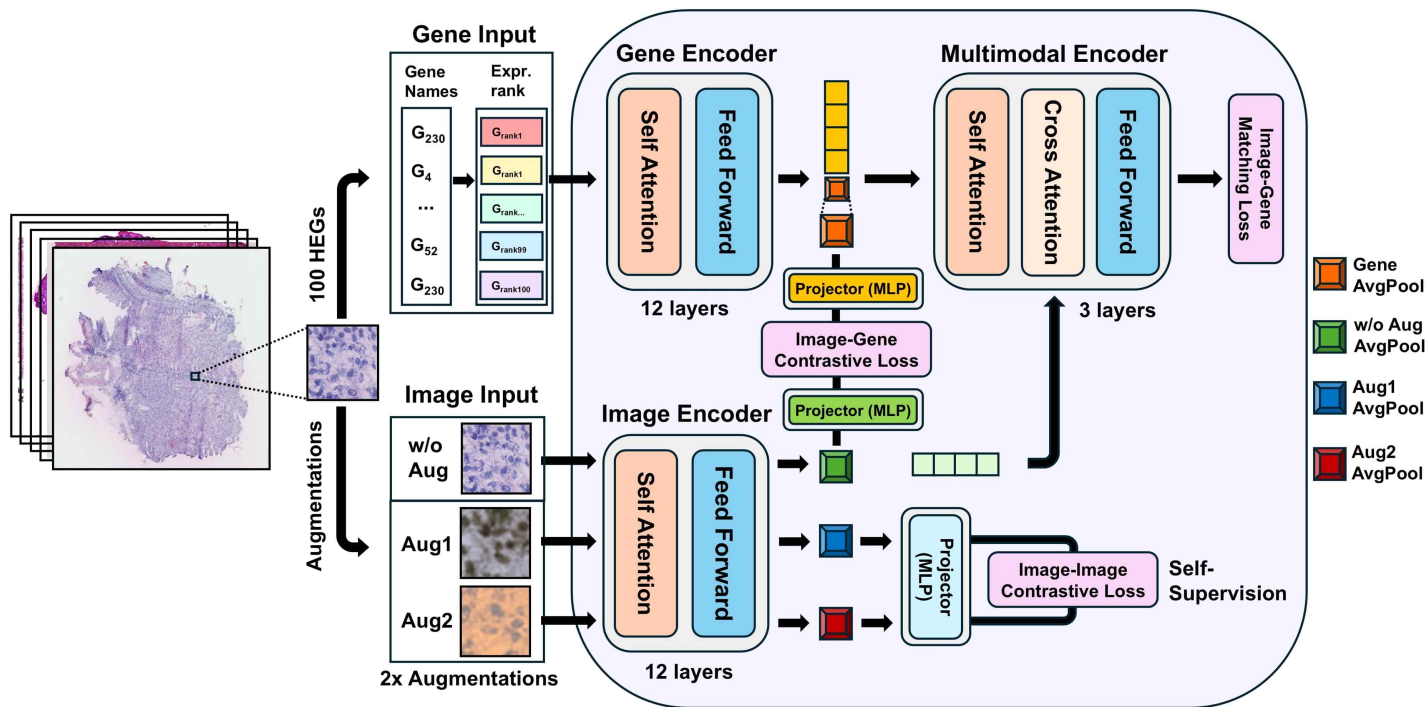
조직 이미지와 유전자 정보를 함께 학습하는 AI도 개발됐지만, **기존 방식은 두 정보를 각각 학습한 뒤 서로 얼마나 잘 맞는지를 비교하는 데 주로 초점을 맞췄다.**

이 때문에 조직의 특정한 형태와 유전자 활동이 어떻게 연결되는지와 같은 **두 정보 사이의 세밀한 관계를 함께 학습하는 데 한계가 있었다.**

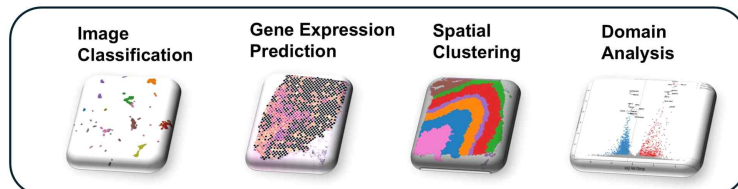
연구팀은 이러한 한계를 해결하기 위해 **조직 이미지와 유전자 정보가 서로 정보를 주고받으면서 두 정보 사이의 관계까지 학습하도록 ST-ConMa를 설계했다.**

사람의 여러 조직에서 얻은 534개 시료의 약 98만1천 개 ‘**조직 이미지-유전자 정보 쌍**’을 미리 학습시키고, 이미지와 유전자 정보를 각각 학습하는 동시에 두 정보가 서로 관련된 경우와 그렇지 않은 경우를 구별하도록 했다.

a Pre-training



b Downstream Tasks

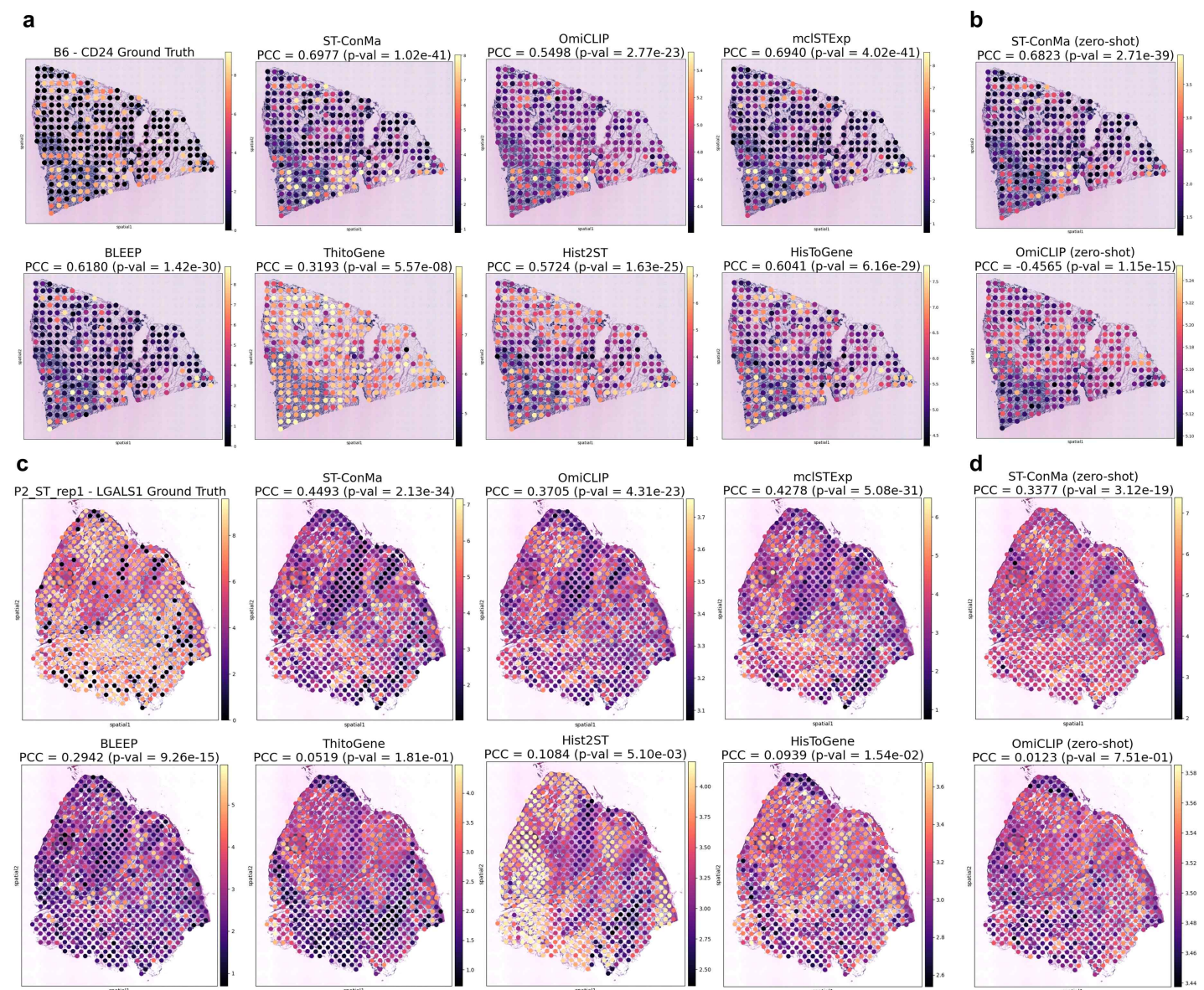


▲ 조직 이미지와 유전자 정보를 함께 학습하는 ‘ST-ConMa’의 작동 원리. 사람의

조직에서 얻은 이미지와 해당 위치의 유전자 발현 정보를 한 쌍으로 학습하고, 각 정보의 특징과 두 정보 사이의 관계를 함께 파악한다. 학습된 ST-ConMa는 조직 이미지에서 유전자 발현을 예측하거나 서로 다른 특성을 가진 조직 영역을 구분하는 등 다양한 공간전사체 분석에 활용할 수 있다.

연구팀은 먼저 두 정보의 관계까지 학습한 것이 실제로 조직 이미지에서 유전자 발현을 예측하는 데 도움이 되는지 확인했다.

유방암과 피부에서 발생하는 암의 한 종류인 피부 편평세포암, 정상 간 조직 데이터를 대상으로 조직 이미지에서 유전자 발현을 예측하는 기존 AI 모델 6종과 성능을 비교했다.



▲ 유방암과 피부 편평세포암 조직에서 유전자 발현을 예측한 결과. 실제 조직에서 측정한 유전자 발현의 위치별 분포와 ST-ConMa를 비롯한 여러 AI 모델의 예측 결과를 비교한 그림이다. 왼쪽은 HER2 양성 유방암의 CD24, 오른쪽은 피부 편평세포암의 LGALS1 유전자 발현 결과를 보여준다.

그 결과, 유전자 발현 예측 성능을 평가한 총 12개 지표 가운데 10개에서 ST-ConMa가 가장 높은 성능을 보였다.

특히 유방암과 피부 편평세포암에서는 한 환자의 데이터를 학습에서 제외한 뒤 해당 환자의 유전자 발현을 예측하는 방식으로 평가했다. 이를 통해 **학습에 포함되지 않은 환자의 조직에서도 유전자 발현을 예측할 수 있음을 확인했다.**

이어 연구팀은 유전자 발현 예측뿐 아니라 **조직 내부의 서로 다른 영역을 구분하는 데도 ST-ConMa를 활용할 수 있는지 검증했다.**

사람의 뇌 앞쪽 영역인 전전두엽 조직을 분석한 결과, ST-ConMa는 실제 **조직 구조와의 일치도를 평가하는 지표에서 기존 비교 모델보다 12.9% 높은 성능을 보였다.** 특히 기존 모델에서는 주변 층과 경계가 뚜렷하지 않았던 **대뇌피질의 4층을 더 명확하게 구분했다.**

유방암 조직에서도 ST-ConMa를 이용해 조직 내부의 서로 다른 영역을 구분할 수 있었다. 구분된 영역의 유전자 특징을 추가로 분석한 결과, **면역 반응이 활발한 영역 안에서도 서로 다른 유전자 특징을 보이는 영역을 다시 구분하는 등 종양 미세환경을 더욱 세밀하게 나눌 수 있었다.**

이현주 교수는 “공간전사체에서는 조직의 모습과 그 안에서 일어나는 유전자 활동을 함께 이해하는 것이 중요한데, 이번 연구는 두 정보의 관계까지 학습해 여러 분석에 활용할 수 있는 기반을 마련했다”며 “앞으로 **다양한 조직과 질환 데이터를 추가로 학습해 복잡한 조직 미세환경을 분석하는 데 활용할 수 있도록 연구를 이어가겠다**”고 말했다.

GIST AI학과 이현주 교수가 지도하고 염태림 석박사통합과정생이 제1저자로 참여한 이번 연구는 과학기술정보통신부·한국연구재단 중견연구자지원사업, 과학기술정보통신부·정보통신기획평가원 AI 최고급 신진연구자 지원사업의 지원을 받았다.

연구 결과는 **생명정보학 분야 국제학술지 《Briefings in Bioinformatics》에 2026년 9월 28일 온라인으로 게재됐다.**

논문 정보

○ 논문명, 저자 정보

- 저널명 : Briefings in Bioinformatics

(JCR 10% 이내, IF 7.3, 2025년 기준)

- 논문명 : ST-ConMa: a multimodal foundation framework for spatial transcriptomics via image-gene contrastive and matching

learning

- 저자 정보 : 염태림(제1저자, AI학과 석박통합과정생), 최동하(공저자, AI학과 박사), 이현주(교신저자, AI학과 교수)