

GIST, 유전자 맞춤형 항암제 생성 AI 세계 최초 개발

개인별 유전자형 분석, 최적의 치료제 설계

- 전기전자컴퓨터공학과 남호정 교수 연구팀, 암 유전자 기반 맞춤형 항암제 후보물질 생성 AI 모델 'G2D-Diff' 개발... 150만 개 화학 구조와 120만 개 약물 반응 데이터 학습
- 기존 모델 대비 월등히 높은 정확도의 생성 능력(오차율 1%, 적합성 35~44% 개선) 보여 난치성 암 정밀치료 가능성 제시... 국제학술지 《Nature Communications》 게재



▲ (왼쪽부터) GIST 전기전자컴퓨터공학과 남호정 교수, 김현호 박사(현 국가독성과학연구소 선임연구원)

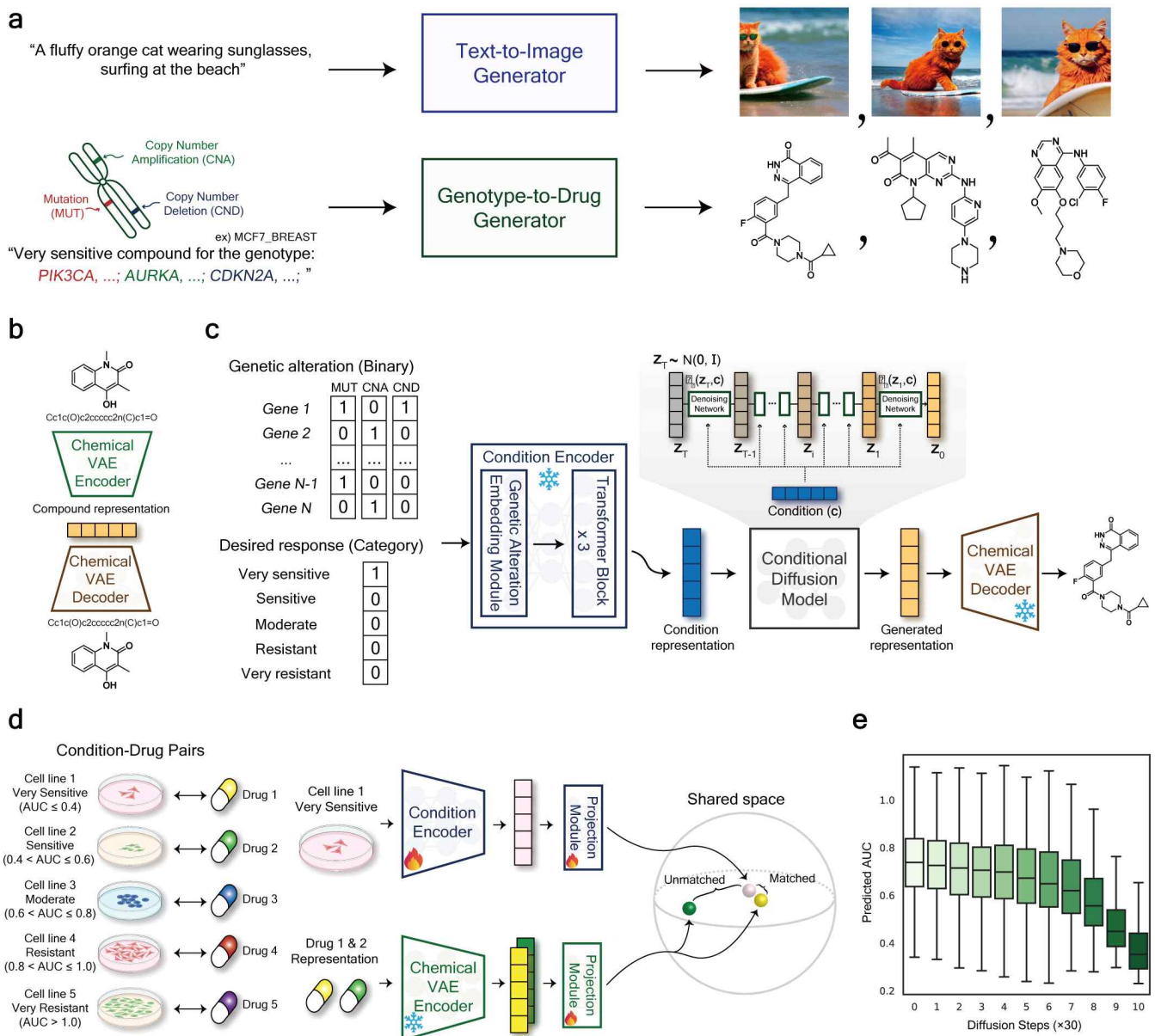
암은 동일한 유형의 질환이라도 환자마다 유전자형*이 달라 치료 효과에 큰 차이를 보인다. 특히 '삼중-음성 유방암(TNBC)*'과 같은 난치성 암은 표적이 명확하지 않아 기존 치료법으로는 충분한 효과를 기대하기 어렵다.

* **유전자형(Genotype)**: 다양한 유전자들의 돌연변이 정보와 복제수 변이 정보를 종합한 형태.

* **삼중-음성 유방암(Triple-Negative Breast Cancer, TNBC)**: 에스트로겐 수용체, 프로게스테론 수용체, 그리고 HER2 단백질이 모두 음성인 유방암을 뜻한다. 이 세 가지 주요 수용체가 발현되지 않아 기존의 호르몬 치료제나 표적 치료제가 효과를 발휘하지 못하는 난치성 암종으로 분류된다. 때문에 TNBC는 치료가 매우 어려워 예후가 좋지 않은 경우가 많으며, 새로운 맞춤형 치료법 개발이 절실히 요구되고 있다.

광주과학기술원(GIST, 총장 임기철)은 전기전자컴퓨터공학과 남호정 교수 연구팀이 **암 환자의 유전자형을 분석해 개인 맞춤형 항암제 후보물질을 제안하는 생성형 인공지능(AI) 모델을 세계 최초로 개발했다고 밝혔다.**

연구팀이 개발한 AI 모델은 **암세포마다 다르게 나타나는 유전자형 정보와 약물 반응 데이터를 학습하여, 환자 개인에게 최적화된 새로운 항암제 후보물질을 생성할 수 있다.** 이로써 환자 맞춤형 정밀의료는 물론, 기존의 치료법이 잘 듣지 않는 난치성 암에 대한 새로운 해법을 제시할 수 있게 됐다.



▲ **G2D-Diff 모델 개요.** a) 텍스트-이미지 생성기 작동 방식을 모티브로 한 G2D-Diff의 작동 방식 컨셉. b-c) 모델의 구조 d) 모델의 주요 부분인 Condition encoder를 학습하는 방식. e) 모델 작동 예시-Diffusion step이 진행될수록 조건 적합성이 증가하고 있음.

지금까지의 생성형 AI 기반 항암제 개발 연구는 몇 가지 한계를 안고 있었다. 첫째, 암과 같은 복합 질환에서는 **치료 표적이 불분명한 경우가 많아 생성된 약물의 효과가 제한적**이었다. 둘째, 임상 현장에서 확보하기 어려운 **특수한 데이터에 의존하는 경우가 많아 실제 활용 가능성이 낮았다.**

연구팀은 이러한 한계를 극복하기 위해 **약 150만 개의 화학 구조와 120만 건의 약물 반응 데이터를 학습한 생성형 AI 모델 'G2D-Diff'를 개발했다.** 이 모델은 실제 임상에서 확보할 수 있는 유전자 정보(돌연변이 및 복제수 변이)와 목표로 하는 약물 반응 수준을 입력하면, 이에 최적화된 항암제 후보물질을 자동으로 설계한다.

G2D-Diff는 텍스트-이미지를 생성하는 AI 모델과 유사한 방식으로 작동한다. 예를 들어, '특정 암 유전자형에 매우 민감한* 약물'이라는 조건을 입력하면, 그 조건에 맞는 항암제 분자 구조를 생성하는 것이다.

이 모델은 ▲분자구조를 수치로 표현하는 '화학적 변분 오토인코더(Chemical VAE)' ▲입력 조건(유전자형과 약물 반응 목표 등)을 수치화하는 '조건 인코더(Condition encoder)' ▲조건에 맞는 새로운 분자 구조를 생성하는 '조건부 확산 모델(Conditional Diffusion Model)'로 구성되어 있다.

* **민감함**: 약물이 암 세포 사멸에 대해 민감하게 작용하여 사멸을 촉진한다는 의미.

G2D-Diff는 기존 생성형 AI 모델들에 비해 모든 성능 지표에서 압도적인 성능을 보였다. 특히 최고 성능을 가진 모델로 알려진 IBM의 'PaccMannRL*'과 비교했을 때, 다양성(Diversity), 실현 가능성(Feasibility), 조건 적합성(Condition Fitness) 등에서 모두 우수한 성능을 나타냈다.

특히, 생성된 화합물들이 입력된 유전자형 조건에 얼마나 잘 부합하는지 평가하는 '조건 적합성(Condition Fitness)' 항목 중 약물 반응성 예측에서는 기존 모델이 약 51%의 평균 오차율*을 보인 반면, G2D-Diff는 약 1% 수준의 평균 오차율을 기록했다.

또한 생성된 분자 구조는 약물 유사성(QED) 및 합성 접근성(SAS) 측면에서 실제 약물군들과의 분포 유사성*이 기존 모델보다 평균 35~44% 더 높게 나타나, 실제 신약으로 개발될 가능성도 높다는 점을 입증했다.

* **PaccMannRL** : 유전자 발현 정보와 화학 구조 데이터를 통합해 항암제의 효능을 예측하고, 강화학습 기반으로 새로운 약물 후보를 생성하는 생성형 AI 모델이다. 주의 메커니즘(attention mechanism)을 통해 암세포에 중요한 유전자나 분자 구조에 집중하며, 원하는 약물 반응을 보이는 화합물을 설계할 수 있도록 최적화되어 있다. 이를 통해 실험 이전 단계에서 효율적인 신약 후보물질 탐색이 가능하다.

* **평균오차율 (%)**:
$$\frac{|\text{생성물 예측민감도 평균} - \text{대조군 예측민감도 평균}|}{\text{대조군 예측민감도 평균}} \times 100$$

* **분포 유사성**: 실제 약물군 분포와의 Jensen-Shannon 발산(분포 차)을 평가지표로 사용함. 이 척도는 두 분포 사이의 거리를 측정하는 척도로 흔히 사용됨.

연구팀은 G2D-Diff 모델을 난치성 암의 대표 사례인 **삼중-음성 유방암에 적용해 실제 활용 가능성을 검증했다.**

환자들의 유전자 변이 정보를 입력해 생성한 후보물질들은 암세포 증식을 억제하는 핵심 단백질인 PI3K, HDAC, CDK* 등을 정확히 표적으로 삼았다.

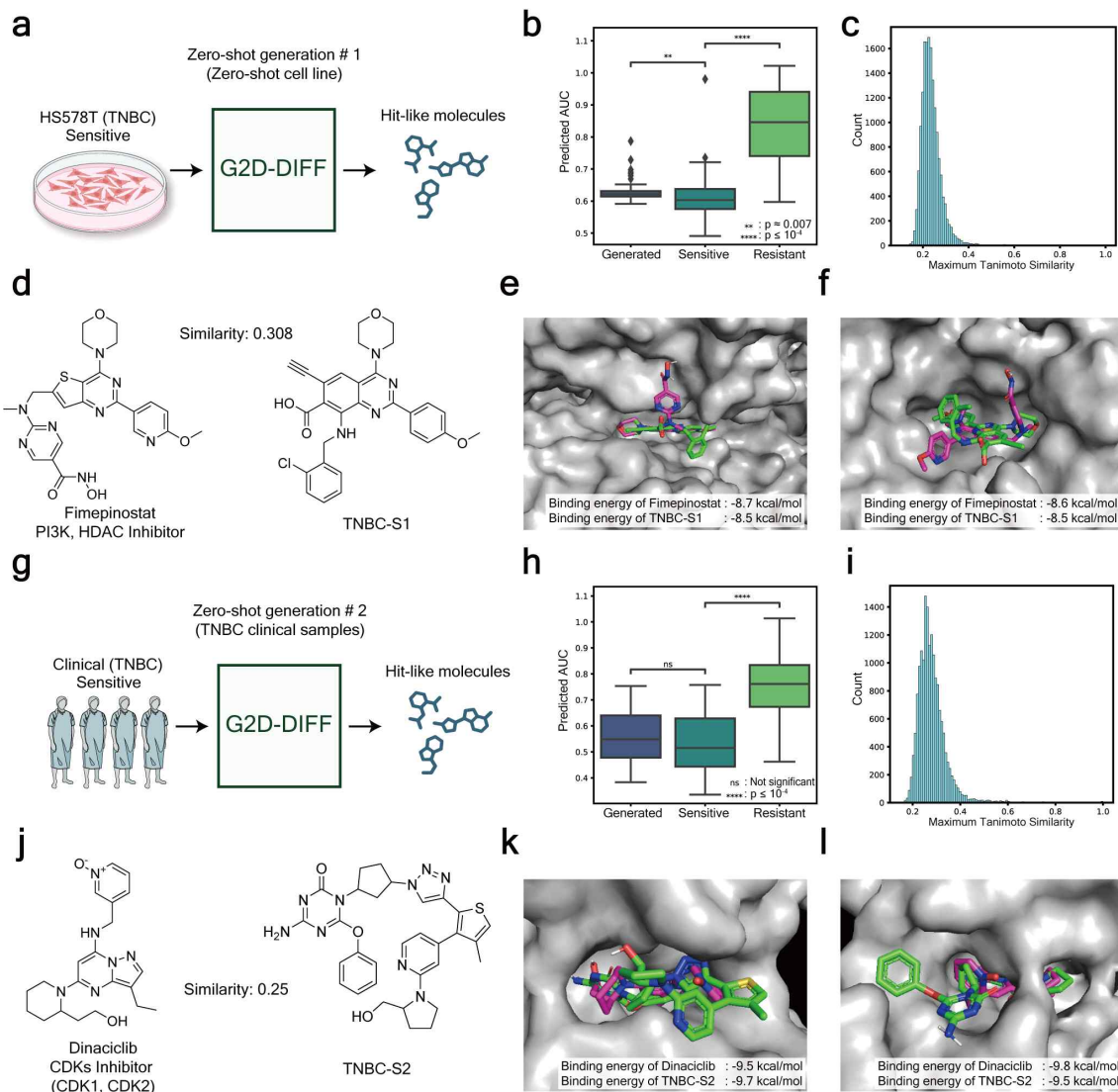
또한, 이들 화합물의 화학 구조가 기존 치료제와 전혀 다르면서도 동일한 치료 효과를 낼 수 있다는 것이다.

컴퓨터 도킹 시뮬레이션*을 통해 이 화합물들이 실제로 암세포의 표적 단백질에 결합할 수 있음도 확인했다.

이는 AI가 단순히 기존 약물을 모방하는 데 그치지 않고, 환자 개개인의 유전적 특성을 분석해 그에 최적화된 완전히 새로운 치료제를 설계할 수 있음을 보여주는 획기적인 성과로 볼 수 있다.

* **PI3K, HDAC, CDK** : 암세포의 성장, 분화, 생존 및 세포주기 조절에 관여하는 핵심 단백질로, 항암제 개발의 주요 표적으로 주목받고 있다. PI3K는 세포 내 생존 신호 전달 경로를 조절하고, HDAC는 유전자 발현을 조절하는 후성유전적 효소이며, CDK는 세포주기 진행을 조절하는 효소로, 이들의 비정상적인 활성화는 종양 발생 및 진행과 밀접하게 관련되어 있다.

* **도킹 시뮬레이션(Docking simulation)**: 컴퓨터를 이용해 약물 분자가 표적 단백질에 어떻게 결합하는지를 예측하는 방법이다. 이 시뮬레이션을 통해 약물과 단백질 간의 상호작용을 분석함으로써, 약물이 목표로 하는 단백질에 효과적으로 결합할 수 있는지를 사전에 평가할 수 있다.



▲ **G2D-Diff의 활용가능성 검증 연구 내용.** a) 삼중-음성 유방암 세포주 HS578T 에 대한 생성 실험 개요. b-c) 생성된 분자들의 조건 적합성 및 높은 구조적 다양성을 확인한 결과. d) 생성한 분자와 기존 약물과의 구조적 유사도 분석. e-f) 타겟 단백질과 도킹 시뮬레이션 결과 g) 삼중-음성 유방암 환자군에 대한 생성 실험 개요. h-i) 생성된 분자들의 조건 적합성 및 높은 구조적 다양성을 확인한 결과. j) 생성한 분자와 기존 약물과의 구조적 유사도 분석. k-l) 타겟 단백질과 도킹 시뮬레이션 결과.

G2D-Diff의 또 다른 강점은 '해석 가능성'이다. 이 모델은 어텐션 메커니즘*을 활용해 특정 유전형을 가진 암종에서 어떤 유전자나 생물학적 경로가 약물 설계에 중요한지 파악할 수 있다.

이를 통해 단순히 새로운 분자를 생성하는 데 그치지 않고, 해당 분자가 왜 효과적인지를 유전자 및 생물학적 경로 수준에서 설명함으로써 치료제의 타당성을 뒷받침하는 과학적 근거까지 함께 제시할 수 있다.

이는 기존의 '블랙박스'*형 AI 모델과 달리, 연구자들이 왜 특정 분자가 생성되었는지 이해할 수 있도록 돕는 중요한 기술적 진전이다.

* 어텐션 매커니즘(Attention mechanism): AI가 여러 정보 중 중요한 부분에 집중하는 기법

* 블랙박스(Black box): 내부 작동 원리를 알 수 없는 시스템

남호정 교수는 "이번 연구는 개인 맞춤형 의학의 새로운 가능성을 연 것으로, AI 기술이 난치성 암 환자들에게 새로운 희망을 제공할 수 있을 것으로 기대한다"고 말했다.

김현호 박사(제1저자)는 "G2D-Diff는 신약 개발 과정에서 가장 어려운 초기 후보물질 탐색 단계의 효율성을 비약적으로 향상시킬 수 있어, 항암제 개발 기간을 크게 단축할 수 있을 것"이라고 강조했다.

GIST 전기전자컴퓨터공학과 남호정 교수가 지도하고 김현호 박사(현 국가독성과학연구소 선임연구원)와 배봉성·박민수·신예원 석박통합과정생·Trey Ideker 교수(University of California, San Diego)가 함께 수행한 이번 연구는 ▲과학기술정보통신부·한국연구재단 중견연구자지원사업 ▲보건복지부·과학기술정보통신부 연합학습 기반 신약개발 가속화 프로젝트(K-MELLODDY) ▲국가독성과학연구소 기본사업 ▲미국 국립보건원(NIH) Bridge2AI 프로그램의 지원을 받았다. 연구 결과는 국제학술지《네이처 커뮤니케이션즈(Nature Communications)》에 2025년 7월 1일 온라인 게재됐다.

논문의 주요 정보

1. 논문명, 저자정보

- 저널명 : Nature Communications (IF 14.7, 2023년 기준)
- 논문명 : A genotype-to-drug diffusion model for generation of tailored anti-cancer small molecules

- 저자 정보 : 김현호(제1저자, GIST 전기전자컴퓨터공학과, 현
국가독성과학연구소), 배봉성(제2저자, GIST 전기전자컴퓨터공학과),
박민수(제2저자, GIST 전기전자컴퓨터공학과), 신예원(제2저자, GIST
AI융합학과),
Trey Ideker(공동교신저자, University of California, San Diego,
Department of Medicine), 남호정(교신저자, GIST
전기전자컴퓨터공학과, AI융합학과)