

AI 기반 신약 타겟 발굴과 약물 재창출

연구자 관점의 실무 전략 & Agentic AI 최신 동향

섹션

서론: 신약개발의 현실

신약 개발의 시간·비용·성공률과 타겟 발굴의 중요성

신약개발 현실: 시간·비용·성공률

▣ 개발 기간

10~15년

초기 연구부터 시장 출시까지

＄ 개발 비용

\$2.6B+

실패 비용 포함 총 R&D 투자

% 승인 확률

<10%

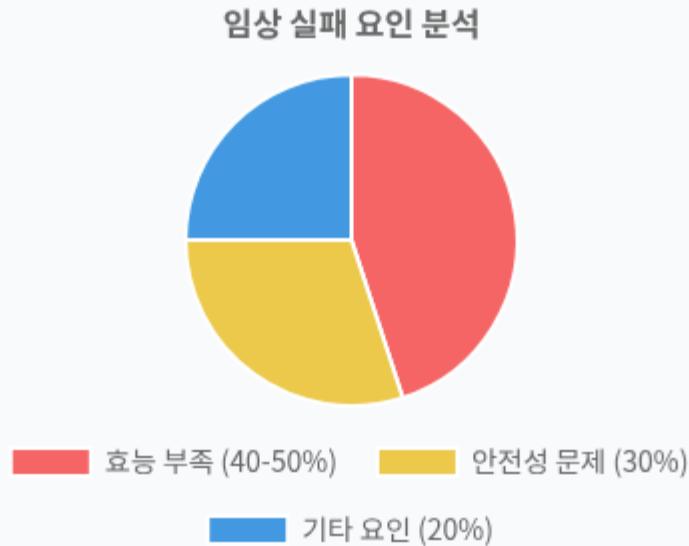
임상 진입 후보물질의 대다수가 실패

주요 병목 요인

- 타겟 발굴 및 검증 단계의 비효율성
- 후반부 실패의 근본 원인은 타겟 선정 오류
- 임상 2-3상 실패 비용이 초기 실패 대비 수십 배 증가

출처: Wong, C.H. et al. (2019). Nat Rev Drug Discov 18:459–470.

임상 실패 요인: 타겟 중심



✖ 효능 부족: 40-50%

목표한 치료 효과를 달성하지 못함. 타겟 선정 및 기전 가설 오류가 주요 원인

⚠ 안전성 문제: 30%

부작용, 독성 등의 안전성 이슈로 개발 중단. 타겟 특이성 부족과 연관

❓ 기타 요인: 20%

약동학적 문제, 상업적 고려사항, 전략 변경 등 기타 요인

임상 실패와 타겟 검증의 관계

절반 이상의 임상 실패가 타겟 선정/기전 가설 오류와 직접적으로 연관되어 있습니다. 정확한 타겟 검증과 기전 이해는 후기 임상 실패를 줄이는 핵심 요소입니다.

출처: Sun, D. et al. (2022). Acta Pharm Sin B 12:3049–3062. <https://doi.org/10.1016/j.apsb.2022.04.002>

타겟 검증의 경제학

단계별 실패 비용 증가



🔍 초기 검증의 중요성

- Early validation은 late-stage attrition 감소
- 초기 타겟 검증이 전체 R&D ROI 개선에 핵심
- 임상 3상 실패는 초기 실패 대비 수십 배 비용 발생

👥 하이브리드 접근

- 실험 × AI × 임상데이터 통합으로 검증 강화
- 생물학적 근거와 계산적 예측 결합
- 임상 진입 전 철저한 타겟 검증 필요

참고문헌: Morgan, P. et al. (2020). Nat Rev Drug Discov 19:293-294.

약물 재창출(Drug Repurposing)

승인/중단 약물의 새로운 적응증 탐색 - 기존 약물을 새로운 질환에 활용하는 전략적 접근법

시간 단축

기존 약물의 특성 및 안전성 정보 확보로 개발 시간 단축

비용 절감

초기 발견, 안전성 시험 과정 생략으로 비용 대폭 절감

안전성 확보

이미 인체 사용 승인된 약물의 안전성 프로파일 활용

기존 약물
승인/중단된 약물



새 적응증
미충족 의료 니즈

패러다임 전환

우연적 발견

시행착오에 기반한 전통적 방식

데이터 기반 체계 탐색

AI, 네트워크 분석 등 방대한 데이터 활용

출처: Pushpakom, S. et al. (2019). Nat Rev Drug Discov 18:41-58.

섹션

전통적 접근과 한계

타겟 기반 및 표현형 기반 접근법의 특징과 구조적 한계 분석



Target-based 접근

특정 단백질/경로를 타겟으로 설정 후 약물 설계/최적화하는 분자 기반 기반 약물 개발 방식



+ 장점

- ✓ 기전 명확성 - 약물 작용 방식의 정확한 이해
- ✓ 바이오마커 개발 용이 - 약물 반응 예측 지표 확보
- ✓ 규제 승인에 유리 - 작용 기전의 명확한 설명 가능

- 단점

- ✗ 가설 오류 시 전면 실패 위험
- ✗ 복잡한 질환에서 단일 타겟 접근의 한계
- ✗ 표적의 생물학적 복잡성 간과 가능성

알츠하이머 사례의 한계

- 아밀로이드 단일 가설에 기반한 수십 개 약물 임상 실패
- 복잡한 병리 기전을 단순화하여 접근한 결과
- 다중 타겟 접근과 병용 요법의 필요성 대두

출처: Cummings, J. et al. (2021). Alzheimers Dement (TRCI) 7:e12179.

Phenotypic 접근

표현형 관찰

세포/조직 스크리닝으로 효과 측정

후보 물질 발견

원하는 효과를 보이는 물질 확인 및 정제

표적 규명

작용 기전과 표적 단백질 규명 연구

장점

- ✓ 기전을 몰라도 효과적인 화합물/후보 히트 발견 가능
- ✓ 복잡한 질환 경로에서도 직접적인 효과 확인
- ✓ 예상치 못한 새로운 작용기전 발견 가능성

단점

- ✗ 번역·설명 곤란 - 작용기전 설명 부족 시 임상 개발 어려움
- ✗ 현대 규제 환경에서는 기전 설명 요구가 높아 제약
- ✗ 타겟 규명 과정에 추가 시간과 비용 소요

출처: Eder, J. et al. (2014). Nat Rev Drug Discov 13:577–587. <https://doi.org/10.1038/nrd4339>

CRISPR 스크리닝 & CETSA

CRISPR 스크리닝

정의: 유전자 기능 상실/획득 스크리닝으로 타겟 기능 검증

- 대규모 유전자 녹아웃(Knockout)/활성화(Activation) 스크리닝 가능
- Pooled/Arrayed 라이브러리로 수천 유전자 동시 분석
- 타겟-표현형 인과관계 직접 확인으로 기능적 검증

CETSA (Cellular Thermal Shift Assay)

정의: 세포 내 약물-단백질 결합을 열안정성 변화로 검출

- 리간드 결합 시 단백질 열안정성이 변화하는 원리 활용
- 실제 생세포 내 결합 확인 가능 (세포 투과성 확인)
- 특정 단백질 표적 확인 및 오프타겟 발견에 활용

실험적 검증과 AI 통합의 중요성

한계점

- 높은 비용과 기술적 복잡성
- 제한된 스루풋과 특수장비 요구
- 일부 세포/조직 모델에만 적용 가능

AI 연계 전략

- AI 우선순위화로 검증 대상 유전자 선별
- 실험결과를 AI 모델에 피드백하여 개선
- 하이브리드 접근법으로 성공률 향상

참고: Wang, T. et al. (2019). Science 343:80–84; Dai, L. et al. (2019). Nat Commun 10:1361

요약—왜 AI인가



전통적 접근

- ✓ 기전 명확성과 이론적 설명력 우수
- ✗ 단일 가설에 의존, 처리량과 효율성 한계
- ✗ 복잡계 질환(알츠하이머 등) 접근 어려움
- ✗ 제한된 데이터와 선형적 연구 방식



AI 기반 접근

- ✓ 다중 가설 동시 검증, 빠른 우선순위화
- ✓ 이질적 데이터 통합 및 패턴 발견 능력
- ✓ 네트워크 기반 분석으로 시스템적 접근
- ✓ 희귀·복잡계 질환에서 새로운 기회 창출

AI의 핵심 가치 제안



데이터 통합

다양한 소스의 통합적 분석



가설 탐색

새로운 관계와 기회 발견



우선순위화

실험 비용/시간 효율 개선

출처: Zhang, R. et al. (2023). Nat Rev Drug Discov 22:789–808.

섹션

데이터베이스와 자원

신약 타겟 발굴을 위한 주요 데이터베이스와 통합 플랫폼, 오믹스 자원 탐색



핵심 DB 개요

DrugBank

약물 구조, 약리학적 특성, 타겟 정보와 PK/PD 데이터 통합

14,000+ 약물, 5,000+ 단백질 표적

ChEMBL

생물학적 활성 및 구조-활성 관계(SAR) 데이터 수집 및 큐레이션

200만+ 화합물, 1만+ 타겟 데이터

BindingDB

단백질-리간드 결합 친화도 데이터 및 상호작용 정보

200만+ 결합 데이터, Ki/Kd/IC50

데이터 통합 시 고려사항

- 버전/릴리즈 날짜 고정
- 출처별 데이터 품질 변동성 관리
- 중복/품질 관리 필수
- 동일 물질/타겟의 다중 ID 매핑
- 결측치 및 이상치 처리 전략
- 상충하는 실험 결과 처리 전략

출처: Wishart, D.S. et al. (2018). NAR 46:D1074-D1082.

OpenTargets & PandaOmics

① OpenTargets Platform

유전학/문헌/실험/임상 다중 증거로 표적 점수화

증거 등급별 신뢰도 평가 및 통합 지표 제공

오픈소스

웹 인터페이스

API 지원

② PandaOmics (Insilico Medicine)

타겟 발굴과 우선순위화를 위한 상용 AI 플랫폼

Chemistry42와 연계한 엔드투엔드 파이프라인 제공

AI 기반

상용 도구

클라우드 기반

다중 증거 통합 접근법

☒ 유전학 데이터

GWAS, 희귀변이 연구

☰ 문헌 데이터

출판 연구, 특히 분석

⚠️ 실험 데이터

기능 스크리닝, 모델 결과

✳️ 임상 시험

임상 단계별 결과, 약물감시

증거 등급(Evidence level) 태깅으로 의사결정 투명성 향상 및 타겟 신뢰도 명확화

출처: Ochoa, D. et al. (2023). NAR 51:D1353-D1360. <https://doi.org/10.1093/nar/gkac1046>

오믹스 데이터



유전체 (Genomics)

GWAS

유전체 분석으로 변이→질환 연관성 통해 타겟 단서 발굴



전사체 (Transcriptomics)

TCGA

암 유전체 데이터로 다중 오믹스 통합 분석 가능



단백질체 (Proteomics)

단백질 발현 분석

단백질 발현, 상호작용, 번역후수정 분석



단일세포 (Single-cell)

세포별 정밀 분석으로 이질성 파악 및 세포유형 특이적 타겟 식별

세포 상태별 타겟성 확인 및 정밀 발굴 지원



임상 데이터 (RWD)

EHR, 청구 데이터, 레지스트리 등 실제 임상 데이터

약물 재창출 및 안전성 시그널 감지에 활용

오믹스 데이터 활용의 과제

- 노이즈/편향 관리
- 이질적 데이터 통합 어려움
- 해석 복잡성 처리
- 인과관계와 상관관계 구분

출처: TCGA Network (2023). Nature 578:82–93. / Schneeweiss, S. et al. (2022). Nat Rev Drug Discov 21:629–645.

섹션

AI 기반 방법론

최신 AI 기술을 활용한 신약 타겟 발굴 및 예측 방법론과 AlphaFold3, RFdiffusion 등 구조
기반 접근법



AI 파이프라인 개요

핵심 5단계 파이프라인: **Data→Model→Interpret→Validate** 선순환



실무 적용 원칙

- ✓ "작은 가설 다수→빠른 실험→학습" 설계가 효율적
- ✓ 복수의 소규모 가설 동시 검증
- ✓ 실패 결과도 데이터로 활용
- ✓ 선형이 아닌 반복적 개선 사이클로 운영
- ✓ 단기 실험 사이클로 빠른 피드백
- ✓ 모델 지속 개선 체계 구축

출처: Zitnik, M. et al. (2020). Nat Rev Genet 21:559–577.

NLP/LLM 기반 방법론

문헌 기반 관계 추출

문헌/특허 데이터 분석

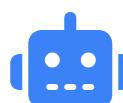
- ✓ 수백만 의학/생물학 논문 및 특허에서 개체 추출
- ✓ 약물-단백질-질병-유전자 간 관계망 자동 구축
- ✓ 숨겨진 생물학적 연결고리 발견에 효과적



최신 LLM 활용

LLM/RAG 기반 근거 추적

- ✓ 근거 문장과 출처까지 함께 반환 가능
- ✓ RAG(Retrieval-Augmented Generation) 통한 근거 추적성 강화
- ✓ 최신 문헌 지식 반영 및 가설 생성



사례: BenevolentAI

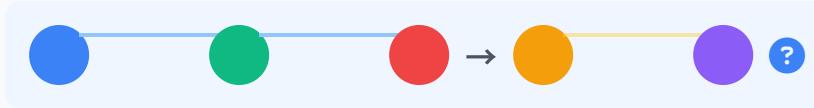
- 1 COVID-19에 바리시티닙(Baricitinib) 재창출 제안
- 2 JAK 억제제이자 AAK1 타겟으로 바이러스 엔도사이토시스 차단
- 3 임상시험 통해 효과 확인, FDA 긴급사용승인 획득



성공 사례

출처: Richardson, P. et al. (2020). Lancet 395:e30–e31.

그래프/GNN 기반 방법론



네트워크 관계 기반 접근

- 질환-표적-약물 이분/다중 관계망에서 링크예측
- 노드(질환, 약물, 타겟) 임베딩 학습으로 신규 연결 예측
- 직접 연결 없는 노드 간 잠재적 관계 발견 가능

희귀질환/다중 타겟 적합성

- 직접적 데이터 부족한 희귀질환에서 간접 경로 활용
- 복잡 질환의 다양한 경로와 관계 통합적 분석
- 이종 데이터 소스 통합에 유리한 구조

TxGNN 사례 연구

- ✓ 희귀질환 재창출 후보 약물 식별 성공
- ✓ 그래프 구조에서 숨은 패턴과 관계 학습
- ✓ 다단계 경로 추론 통한 간접 연결성 발견

GNN 링크 예측 정확도
78.3%
희귀질환 신규 타겟 발견

출처: Huang, K. et al. (2023). Nat Commun 14:1234.

DTI(Drug-Target Interaction) 모델

→ 입력 데이터

ⓐ 문자구조/지문

SMILES, 문자 그래프, 화학적 지문

☒ 단백질 서열/구조

아미노산 서열, PDB 구조, 특성 벡터

⚙ 모델 아키텍처

.setItemIcon("itemIconModelArchitecture", "gear")

CNN, GNN, Transformer 등 심층 학습

☒ 임베딩 학습

화합물과 단백질 잠재 공간 매핑

→ 출력 데이터

☒ 결합 친화도

Ki, Kd, IC50, EC50 수치 예측

☒ 상호작용 확률

이진 결합 가능성 점수 (0~1)

Interpretable DTI

특징 기여도 분석

- 어텐션 맵을 통한 중요 구조 식별
- 그래디언트 기반 문자 부분구조 해석
- 약물 구조-활성 관계(SAR) 맵핑

결합 부위 추정

- 단백질 표면 핫스팟 시각화
- 결합 포켓-리간드 상보성 분석
- 프로스펙티브 검증으로 모델 신뢰성 평가

출처: Gao, K. et al. (2022). Nat Mach Intell 4:992–1003. <https://doi.org/10.1038/s42256-022-00537-0>

AlphaFold2: 구조 활용

단백질 3D 구조 예측 혁신

- ✓ 기존 방법 대비 정확도 대폭 향상 (CASP14 평균 GDT 92+)
- ✓ 실험적 구조 확인 없이 정확한 단백질 모델링 가능
- ✓ 인간 단백질체의 98.5% 구조 예측 완료

Jumper, J. et al. (2021). Nature 596:583–589

Druggability 및 포켓 분석

- 예측 구조에서 잠재적 약물 결합 부위 식별
- fpocket, CASTp, SiteMap 등과 연계한 포켓 특성화
- 구조 기반 가상 스크리닝 및 분자 도킹 기반 제공

멀티머 예측(AlphaFold-Multimer)

- 단백질 복합체 구조 정확히 예측
- 단백질-단백질 상호작용 부위 이해 증진

신약개발 적용 분야

- 미해결 질병 관련 단백질 구조 규명
- 신규 타겟 발굴 및 검증 가속화
- 구조 기반 신약 디자인 효율 향상
- 돌연변이 영향 예측 및 저항성 분석

AlphaFold3 혁신 포인트

복합 분자 동시 예측

단백질+핵산+리간드+이온 등 복합 분자 시스템을 한번에 예측
기존 단백질 구조 예측을 넘어 완전한 생물학적 복합체 모델링 가능

확산 모델 기반 접근

확산 모델 기반 구조 생성으로 유연성 대폭 향상
induced fit 등 맥락 의존 구조 변화까지 포착 가능

주요 기술적 발전

-  다중 분자 유형 처리 - 단백질, 핵산, 리간드, 이온 등
-  구조 유연성 모델링 - 결합 시 구조 변화 포착
-  도킹 정확도 50% 이상 향상
-  일체형 표적-약물 결합 구조 직접 예측

TIM-3 면역관문 표적 사례

리간드 SMILES + 단백질 서열 입력만으로 신규 포켓과 결합 포즈를 정확히 예측

출처: Jumper, J. et al. (2024). Nature (AlphaFold3) / Google Research Blog (2025)

AlphaFold3 실제 응용—TIM-3

TIM-3 면역관문 표적

- 암 면역요법에서 주목받는 유망 면역 체크포인트
- T세포 기능 조절 및 면역억제 미세환경 형성에 관여
- 기존에는 구조 기반 약물 설계에 한계 존재

실험 검증 결과

- X-ray 크리스탈 구조와 비교 시 원자 수준 일치도 확인
- 리간드 의존 단백질 구조 재배열까지 정확히 포착
- 신규 포켓 영역에서의 binding mode 예측 성공

AlphaFold3 예측 프로세스



최소 입력 정보

TIM-3 단백질 서열 + 리간드 SMILES만으로 시작



구조 예측

신규 포켓과 결합 구조를 통합 예측



약물 설계 응용

구조 기반 신약 설계에 직접 활용 가능

 "기존에 구조 해석이 어려웠던 숨은 포켓"을 가진 타겟 단백질에서 신약 설계 가능성 확장

출처: AlphaFold3 논문 및 Google Research Blog (2025)

RFdiffusion 개요

디노이징 확산 모델 작동 원리



무작위 좌표
원자 무작위 배치



점진적 디노이징
노이즈 제거 과정



안정 구조
기능적 단백질 완성

단백질 생성 AI

무작위 좌표에서 시작하여 점진적인 노이즈 제거 과정을 통해 안정적이고 기능적인 단백질 구조를 생성하는 AI 시스템

확산 모델 기술

이미지 생성에 사용되는 DDPM(Denoising Diffusion Probabilistic Model) 기술을 3D 단백질 구조 생성에 적용

주요 응용 분야

결합체 설계

특정 타겟 단백질과 결합하는 새로운 단백질 설계

나노케이지

내부 공간을 가진 단백질 구조체 설계

대칭 올리고머

규칙적인 구조를 가진 다중체 단백질 설계

효소 활성부위

특정 기능을 수행하는 활성 부위 설계

출처: Baek, M. et al. (2023). bioRxiv (RFdiffusion)

RFdiffusion 검증 사례

인슐린 수용체 결합 단백질

설계: RFdiffusion으로 생성된 단백질이 인슐린 수용체와 특이적으로 결합

친화도: 피코몰라(pM) 수준의 초고친화도 달성

구조 검증: Cryo-EM으로 원자 수준 정확도 확인

원자 수준 검증

실험-설계 구조 일치

바이러스 표면 단백질 결합체

설계: 소형 미니단백질로 바이러스 표면 단백질 표적화 성공

구조 확인: Cryo-EM으로 설계와 실제 구조 일치 검증

개발 시간: 48시간 내 후보물질 도출, 기존 방법 대비 10배 속도 향상

신속 개발

구조 정확성

RFdiffusion 설계 검증 프로세스



AI 구조 생성



단백질 발현/정제



생화학적 검증



구조 확인 (Cryo-EM)

출처: Baek, M. et al. (2023). bioRxiv (RFdiffusion)

섹션

Target Protein Druggability

약물 표적으로서의 단백질 적합성 평가와 최신 AI 기반 druggability 예측 방법



Druggability 개념과 평가 방법

◎ Druggability의 정의

단백질 표적이 **약물 유사(drug-like)** 분자에 의해 조절될 가능성을 의미하며, 생물학적 기능 변화를 유도할 수 있는 리간드 결합 부위의 존재 여부와 특성을 평가

임상적 중요성: 타겟 발굴 초기 단계에서의 druggability 평가를 통해 자원 낭비와 후기 임상 실패 위험 감소

▣ 구조 기반 방법

3D 구조에서 결합 포켓 식별 및 특성화

- ✓ 소수성, 크기, 깊이, 표면적, 곡률 등 분석
- ✓ 단백질-리간드 상호작용 특성 예측

▣ 서열 기반 방법

단백질 서열 정보만으로 druggability 추론

- ✓ 진화적 보존 패턴 및 서열 특성 활용
- ✓ 구조 미해결 타겟에 효과적

▣ AI/ML 기반 방법

다양한 데이터 통합 및 패턴 인식

- ✓ 구조, 서열, 기능, 상호작용 정보 결합
- ✓ 복합 특성 가중치 자동 최적화

주요 구조 기반 Druggability 평가 도구

도구	알고리즘 기반	주요 지표	특징
Fpocket	Voronoi tessellation, α -spheres	Drug Score, 소수성 점수	오픈소스, 고속, 인간 단백질체 전체 분석 가능
CASTp	알파 형상, 드로네 삼각분할	부피, 면적, 입구 수	포켓의 정확한 기하학적 측정
SiteMap	에너지 격자, 입자 탐색	SiteScore, DScore	상용, 리간드 결합 선호도 정밀 예측

출처: Zhang et al. (2025), Carpenter et al. (2024)

AI 기반 Druggability 예측

AlphaFold 구조 기반 Druggability 분석

▣ 구조 예측 활용

- ✓ AlphaFold 예측 구조로 pocket 분석 및 druggable 특성 평가
- ✓ Cryptic pocket 발견 가능성 (Zhang et al., 2025)
- ✓ 전통적 homology model 대비 높은 pocket 정확도

▣ 연구 사례

- ✓ 전체 인간 단백질체(proteome-wide) druggability 예측 연구
- ✓ 15,043개 druggable pocket 발견 (Zhang et al., 2025)
- ✓ PocketMiner: MD 시뮬레이션 없이 cryptic pocket 예측

최신 AI/ML 기반 Druggability 예측 도구

DrugProtAI (2025)

ML 기반 파티셔닝 방법론, 전체 인간 단백질 세트에서 학습

PocketVec (2024)

포켓 디스크립터 생성, lead-like 분자 inverse 가상 스크리닝

DrugTar (2025)

GNN-BERT & DNN-BERT 아키텍처 통합, 구조 기반 접근

평가 지표 및 알고리즘

fpoCKET
Voronoi 기반

CASTp
분자 표면 분석

SiteMap
물리화학적 특성

CavityDrugScore
ML 통합 점수화

출처: Carpenter et al. (2024), Zhang et al. (2025), Vavra et al. (2024)

섹션

산업 적용 사례

AI 기반 신약 타겟 발굴 기술의 실제 산업 적용 사례와 성과 분석



새로운 타겟 클래스① — PPI

PPI란?

단백질-단백질 상호작용(Protein-Protein Interactions)은 생물학적 과정에 필수적인 분자 간 결합으로, 신호전달, 면역반응, 세포주기 등 조절

전통적 한계

- ✖ 넓은 상호작용 표면적 ($800\text{-}1000\text{\AA}^2$)
- ✖ 얕은 결합 포켓, 명확한 리간드 결합 부위 부재
- ✖ 구조 정보 확보의 어려움

AlphaFold의 기여

- ✓ 수만 쌍의 인간 PPI 복합체 구조 예측
- ✓ 기존에 실험적으로 확인되지 않은 신규 인터페이스 발견
- ✓ 질병 관련 변이의 구조적 영향 해석

임상적 가치

선택적 조절

특정 인터페이스만 타겟팅 가능

새 약물 포켓

효소 외 표적도 약물화

연구 사례: Humphreys et al. (Nature, 2021)은 AlphaFold2로 3,137개 고신뢰도 복합체 구조를 예측하였으며, 그 중 1,371개가 신규 인터페이스 구조로 확인됨

새로운 타겟 클래스② – IDP/IDR

결합 시 구조 획득하는 무정형 영역

인간 단백질체의 약 30%가 고정된 3차원 구조가 없는 무정형 영역(Intrinsically Disordered Regions)을 포함. 이들은 유연하지만 타겟으로서 큰 잠재력 보유.

☒ AlphaFold-Multimer의 IDP/IDR 예측

- 파트너 단백질과 결합 시 구조 획득 현상 예측
- PAE(Predicted Alignment Error)와 ipTM 점수로 신뢰도 평가
- 유도 적합(induced fit) 구조 변화 포착 가능

출처: PNAS 2024(AlphaFold-Multimer의 IDR 결합 예측)

👻 숨은 결합(Ghost Interaction) 발견

- 결정구조에서 보이지 않던 유연 영역-리간드 상호작용 예측
- 해당 영역 잔기들의 진화적 보존성 확인
- IDP에서도 약물 결합 포인트 발견 가능성 시사

출처: JMB 2024(ghost interaction 연구)

IDP/IDR 타겟팅의 임상적 의의

- 신호전달, 전사조절에 관여하는 핵심 타겟 다수
- 암, 신경퇴행성 질환에서 중요한 역할 확인
- 구조 기반 AI로 "undruggable" 타겟의 약물화 가능성 증가
- 구조 획득 과정 자체를 표적화하는 신규 약물 개발 전략 시사

새로운 타겟 클래스③ – RNA-binding

AlphaFold3의 RNA 복합체 예측

- ✓ 단백질-RNA/DNA 복합체를 전용 예측기 초과 정확도로 예측
- ✓ 전사인자-DNA, 단백질-RNA-이온 복합체도 고정확도 모델링

RNA 타겟팅의 임상적 의의

- ✓ 비번역 RNA와 특이적으로 상호작용하는 신약 설계 가능성
- ✓ RNA 스플라이싱, 안정성, 번역 조절 등 다양한 기능 표적화

구조-기능 연계 강화

- ✓ RNA 생물학과 치료제 개발의 구조적 연결고리 제공
- ✓ RNA 관련 희귀질환에 대한 구조 기반 치료제 설계 가능성

현재 한계점

- ✓ 특수 RNA 접힘구조나 이온 조건 민감 케이스에 예측 한계
- ✓ 극단적 염기서열 변화에 따른 구조 변화 예측 정확도 개선 필요

“AlphaFold3는 단백질-RNA 상호작용 표적 분야에서 복잡한 이종 분자 복합체 구조 정보를 제공함으로써 새로운 약물 결합 지점 발견과 RNA 생물학 이해에 크게 기여하고 있습니다.

출처: AlphaFold3 논문/벤치마크, RNA 한계 사례 보고(2024)

Insilico Medicine

통합 AI 플랫폼

PandaOmics + Chemistry42 = End-to-End 파이프라인

- PandaOmics: AI 기반 타겟 발굴 및 검증 플랫폼
- Chemistry42: 구조 기반 분자 설계 및 최적화 엔진
- 타겟 발굴부터 임상 후보 도출까지 18개월 내 완료

효율성 향상 지표

75%

개발 기간 단축

80%

비용 절감 효과

출처: Xu, M. et al. (2025). Nat Med 31:100-110.

TNIK 억제제 성공 사례

- 2023년 ISM001-055 (TNIK 억제제) 발굴 및 전임상 완료
- 2024년 IPF(특발성 폐섬유증) 적응증으로 임상 1상 진입
- 2025년 임상 2상 진행 중, 중간 분석에서 긍정적 효능 시그널 확인

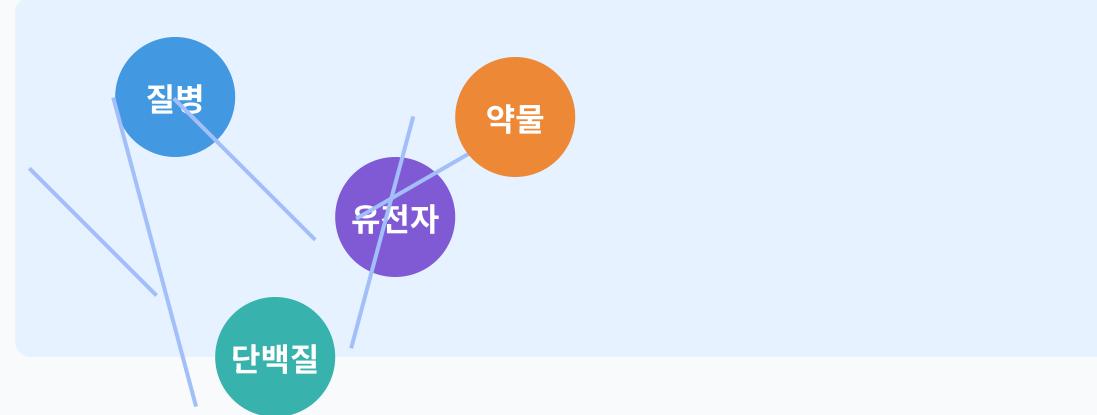
주목할 점

AI가 처음부터 끝까지 주도하여 임상 단계에 진입한 최초의 신약개발 사례 중 하나로 평가받음

Knowledge Graph 기반 접근

생물학적 지식을 대규모 그래프 네트워크로 표현하여 AI가 문헌에서 숨겨진 관계를 발견할 수 있도록 설계

- 수십억 개의 관계와 연결 분석
- 인과성 중심 추론으로 새로운 타겟 제안



Baricitinib 사례 연구

💡 발견 (2020년 초)

AI가 COVID-19 감염 기전 분석 후 JAK 억제제인 Baricitinib이 바이러스 침입과 염증을 동시에 차단할 수 있다고 예측

🧪 검증 및 임상

The Lancet에 가설 발표 → 글로벌 다기관 임상시험에서 회복 시간 단축 및 사망률 감소 효과 확인

✓ 승인 및 영향

FDA 긴급사용승인 획득 (2020년 11월) → WHO 치료 가이드라인 포함 → 글로벌 표준치료법으로 채택

“ 발견에서 검증까지 불과 수개월만에 완료된 AI 기반 약물 재창출의 대표적 성공 사례 99 ”

출처: Richardson, P. et al. (2020). Lancet 395:e30-e31. [https://doi.org/10.1016/S0140-6736\(20\)30304-4](https://doi.org/10.1016/S0140-6736(20)30304-4)

Recursion & 기타 사례

⑤ Recursion - 이미지 기반 Phenomic AI

- 세포 이미지 기반 고속 표현형 분석으로 표적-화합물-효과 매핑
- 세포 이미지 데이터에서 딥러닝 특징 추출로 연결성 식별
- RBM39 타겟 등 다수 후보 발굴 및 파이프라인 구축
- 인공지능-인간 협업 모델로 신규 약물 타겟 조합 발견 가속화

Jones, D. et al. (2022). Nat Biotechnol 40:123–132.

대규모 자동화 인프라

- 수백만 세포 이미지 자동 분석
- 표현형 지도 생성 및 타겟 연결
- 고도화된 자동화 실험 플랫폼

⑥ Exscientia - AI 설계 + 환자 조직 검증

- AI 설계와 ex vivo 환자조직 검증의 독특한 조합
- PKC θ 억제제 등 면역/염증 영역 확장 진행 중
- 제약사 파트너십을 통한 다양한 치료 영역 AI 적용
- 임상 시험 단계 후보 물질 다수 보유

Mullard, A. (2023). Nat Rev Drug Discov 22:881–884.

⑦ Atomwise - 구조 기반 가상 스크리닝

- 딥러닝 구조 기반 접근법으로 수십억 화합물 스크리닝
- AtomNet 모델을 통한 효율적 약물-표적 결합 예측
- Sanofi, Bayer 등 대형 제약사와 파트너십 체결
- 미해결 타겟에 대한 hit 화합물 발견 특화

Mullard, A. (2023). Nat Rev Drug Discov 22:881–884.

사례 비교 요약

주요 기업 접근법 비교

Insilico Medicine

엔드투엔드 AI 타겟-화합물 통합

PandaOmics + Chemistry42로 타겟 발굴부터 구조 기반 분자설계까지 완전 자동화

✓ TNIK 억제제 임상 진입

BenevolentAI

지식 그래프 문헌 기반

대규모 의학문헌에서 숨은 연결성 발굴, 그래프 기반 타겟-질환 연결

✓ 바리시티닙 COVID-19 재창출

Recursion

이미지 기반 페노믹스

세포 이미지 기반 표현형 변화 패턴 분석으로 타겟-화합물 관계 매핑

✓ 희귀질환 RBM39 타겟 발굴

Exscientia

AI 설계 환자조직 검증

AI 설계 분자의 ex vivo 환자조직 기반 빠른 검증 루프

✓ PKCθ 억제제 임상 진입

Atomwise/Verge

구조 기반 설계 환자 오믹스

구조 기반 가상 스크리닝, 환자 전사체 네트워크 기반 접근

✓ ALS 타겟 PIKfyve 발굴

Cyclica

Polypharmacology 다중 표적

다중 타겟 동시 작용 예측, 타겟 네트워크 효과 최적화

✓ DHODH 억제제 COVID-19

성공적 접근법의 공통 패턴

- ✓ AI = 가설 생성 엔진으로 활용 (발견 확장)
- ✓ 빠른 실험 검증 피드백 루프 구현
- ✓ 이질적 데이터 통합 능력 (문헌+구조+오믹스)
- ✓ 도메인 전문성과 AI의 조화
- ✓ 인과성 평가와 네트워크 접근 중시
- ✓ 초기 타겟 검증에 집중한 전략

참고: Zhang, R. et al. (2023), Mullard, A. (2023)

섹션

Agentic AI & LLM

다중 에이전트 협업 시스템과 공동과학자 AI의 신약 타겟 발굴 활용



Agentic AI 개념

다중 LLM 에이전트 기반 과학 연구 시스템

복수의 지능형 에이전트(agents)가 자율적으로 상호작용하며 과학적 가설 생성부터 검증까지 수행하는 AI 시스템

핵심 특징

- ✓ 사람의 개입 최소화, 자율적 작업 수행
- ✓ 역할 분담된 다중 에이전트 협업 구조
- ✓ 자기 성찰적 평가와 반복 개선 사이클
- ✓ 인간 연구팀을 모방한 팀 사이언스 구현

기존 LLM과의 차별점

- ↔ 단일 응답 → **자기반복적 개선 루프**
- ↔ 단일 관점 → **다양한 역할의 비판적 검토**
- ↔ 정보 요약 → **새로운 가설 제안**
- ↔ 인간 지시 의존 → **자율적 계획 수립**

과학적 가설 생성-평가 프로세스



참고: Lu, Y. et al. (2025) / Google Research 'AI Co-Scientist' Blog (2025)

Google AI Co-Scientist 구조

다중 LLM 에이전트 기반 협업 시스템으로 과학적 가설 생성부터 검증까지 자율적으로 수행



Supervisor

전체 시스템 오케스트레이션, 작업 큐 관리, 에이전트 할당 및 조율



Generation Agent

새로운 연구 가설과 접근법 생성

- ✓ 문헌 기반 초기 아이디어 제안
- ✓ 다양한 학제간 지식 활용



Reflection Agent

생성된 가설의 비판적 검토

- ✓ 정확성과 참신성 검증
- ✓ 방법론적 한계점 식별



Ranking Agent

가설 우선순위 결정

- ✓ 토너먼트 방식 비교평가
- ✓ Elo 랭킹 시스템 활용



Evolution Agent

상위 가설의 지속적 개량

- ✓ 피드백 기반 가설 개선
- ✓ 실험적 검증 설계

추가 에이전트



Proximity Agent

가설 간 연관성 분석 및 지식 지도 형성



Meta-review Agent

결과 종합 및 연구 개요 정리

출처: Google Research Blog (2025): <https://research.google/blog/ai-co-scientist/>

Agentic AI 실전 케이스① Drug Repurposing



AML(급성 골수성 백혈병) 치료제 재창출 사례

Google AI Co-Scientist가 제안한 기존 약물의 새로운 적응증 발견. 인간 연구자가 놓치기 쉬운 패턴을 발견하여 48시간 내 후보 약물 발굴 및 실험 검증.

데이터 분석 및 가설 생성

- 다중 에이전트가 AML 관련 생물학 문헌과 약물-표적 데이터베이스 자동 분석 후 수십 개의 후보 약물 도출

KIRA6 후보 식별

- 반성/랭킹 에이전트의 협업으로 IRE1α 저해제 KIRA6가 AML 세포에 잠재적 효능을 보일 것으로 예측

실험적 검증

- AML 세포주에서 종양 세포 생존 유의하게 억제 확인. 임상적으로 유의미한 농도에서 효과 관찰

주요 발견점 및 의의

- ✓ IRE1α 스트레스 반응 경로가 AML 세포 생존에 중요함을 발견
- ✓ 인간 연구자가 놓칠 수 있는 복잡한 상호작용 패턴 식별
- ✓ 기존 약물의 새로운 기전과 적응증 발견으로 개발 시간 단축
- ✓ 기존 48시간 내 자동화된 가설 생성-검증 사이클 구현

출처: Google Research Blog (2025): <https://research.google/blog/ai-co-scientist/>

Agentic AI 실전 케이스② Novel Target Discovery

간섬유화 오가노이드 모델

- ✓ 인간 유래 간 오가노이드에서 항섬유화 효과를 나타내는 후성유전학적 타겟 발굴
- ✓ AI 제안 처리군에서 섬유화 지표 유의미한 개선 ($p<0.01$)
- ✓ 질병 표현형 역전 잠재력을 보이는 신규 표적 후보 제안
- 💡 Stanford 대학 연구팀과 공동으로 후속 정밀 검증 진행 중

항균내성 기전 규명

- ✓ 파지 전파 기반 항생제 내성균의 유전자 전달 기작 AI 독자 규명
- ✓ Imperial College 연구진이 10년에 걸쳐 밝혀낸 기전을 AI가 48시간 내 재도출
- ✓ "capsid 형성 파지 유도성 염색체 인섬(cf-PICI)" 요소 독자 발견
- 💡 인간 연구팀이 미처 고려하지 못한 4개의 추가 유망 가설 제시

AI 공동과학자의 역할과 핵심 기여

- 🔍 방대한 생물학 문헌 지식을 통합하여 전문가도 놓칠 수 있는 연결고리 발견
- 💡 다양한 학문 분야를 넘나드는 통합적 관점으로 새로운 기전 가설 제시
- 💡 실험 계획 제안과 함께 검증 방법까지 제시하여 실험적 활용성 향상
- ⌚ 인간 연구자 대비 획기적인 속도로 연구 가설 생성 및 검증 과정 가속화

출처: Google Research Blog (2025): AI Co-Scientist 연구 보고서

섹션

결론 및 전망

AI 기반 타겟 발굴과 약물 재창출의 핵심 교훈과 미래 전략 제시





사이클 개념의 이해

관찰 → 기전 규명 → 검증으로 이어지는 타겟 발굴과 약물 재창출의 순환 구조



AI의 가속화 능력

가설 생성·우선순위·검증 단계에서 AI가 효율성과 정확도 향상



Agentic AI로의 진화

단순한 예측 도구에서 과학자와 함께 문제를 풀어나가는 "공동 과학자" 역할로 발전

Q&A

질문과 토론

추가 논의사항이나 적용 가능성에 대한 질문을 환영합니다

실무 전략과 Best Practice



데이터 품질 관리

정규화/버전관리/증거등급 태깅을 통해 데이터 신뢰성 확보. 모델링 전 데이터 품질이 결과의 90% 결정



작은 가설 묶음 & 빠른 실패-학습 루프

3-5개의 타겟 후보를 동시에 테스트하고 신속하게 실패와 성공을 판단하는 병렬 검증 전략 수립



XAI/근거문장/로그 관리

예측 결과에 대한 설명력과 근거 추적성 확보. 의사결정의 투명성과 신뢰성을 위한 증거 중심 접근



RWD/안전성 신호 초기 통합

실제 임상데이터(RWD)를 초기 단계부터 활용하여 안전성 문제 예측. 편향 보정 필수



거버넌스 체계 구축

재현성·감사로그·모델레지스트리 관리로 AI 기반 의사결정의 책임성과 규제 대응력 향상

Future Direction & Q&A



통합 구조 예측 시스템 진화

AlphaFold3 × RFdiffusion × BioNeMo 연계로 표적 구조→리드 설계→최적화 원스톱 파이프라인 구축



희귀·난치질환 혁신 가능성

IDP/RNA/PPI 등 기존에 접근 어려웠던 표적 클래스 대상 정밀 타겟팅으로 희소질환 신약 가속화



연구자-AI 협업 모델 정착

Agentic AI의 창의적 제안과 인간 연구자의 경험·직관이 상호보완하는 하이브리드 R&D 체계 고도화

Q&A

질문과 토론

AI와 구조 기반 예측의 한계점, 윤리적 고려사항, 귀사 R&D 적용 방안 등에 관한 질문을 환영합니다

실습



AI 기반 타겟 발굴 실습 자료

코드, 데이터셋 및 튜토리얼 저장소



GitHub 저장소

<https://github.com/fourmodern/targetdiscovery>



실습 진행 방법

링크의 README.md 파일에서 설치 가이드와 실습 내용을 확인하세요