

# | AI 기반 신약 후보물질 탐색

에이조스바이오

박혜진

2025.07.21

## • 신약개발 프로세스



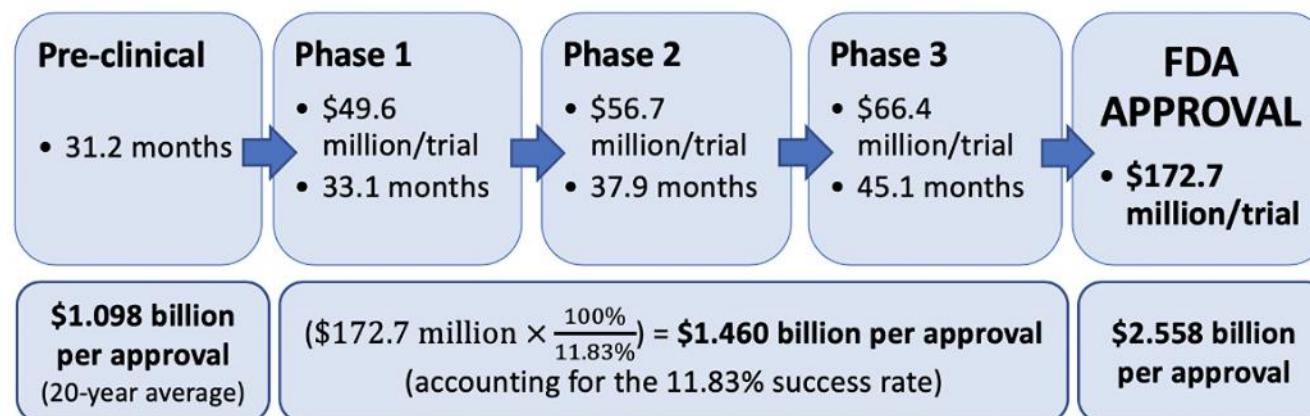
\* Source: Pharmaceutical Research and Manufacturers of America

## • 신약개발 프로세스와 AI 적용

- Chemical space ??

- 이론적 규모:  $10^{60}$
- 이용 가능한 화합물 규모:  $10^{10}$ , DEL, ZINC ( $7 \times 10^8$ )
- 승인된 의약품 규모:  $3 \times 10^3$

- 신약개발 비용



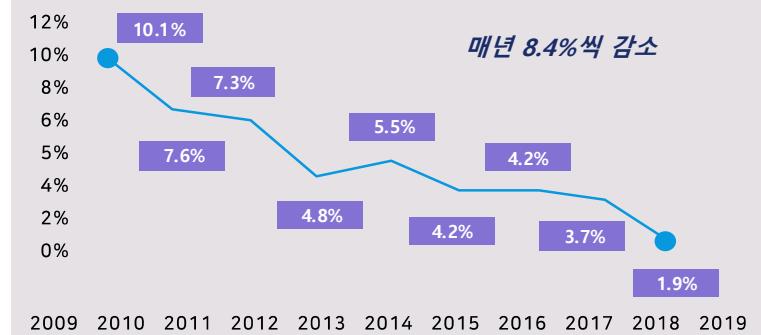
Data from DiMasi et al. (2016)

## • 신약개발 프로세스에서 AI 적용 필요성

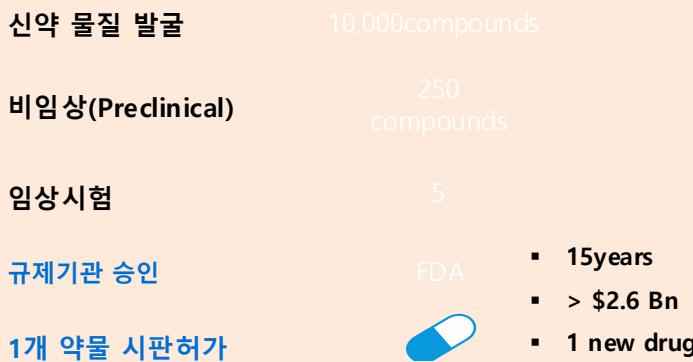
### 이슈

- 고위험 저수익(Eroom's law)
- 개발기간 15년 이상
- High Cost for HTS
- 효과 및 부작용 예측 tool 부족
- Lead selection 에 많은 시간과 비용이 소요됨(38%)

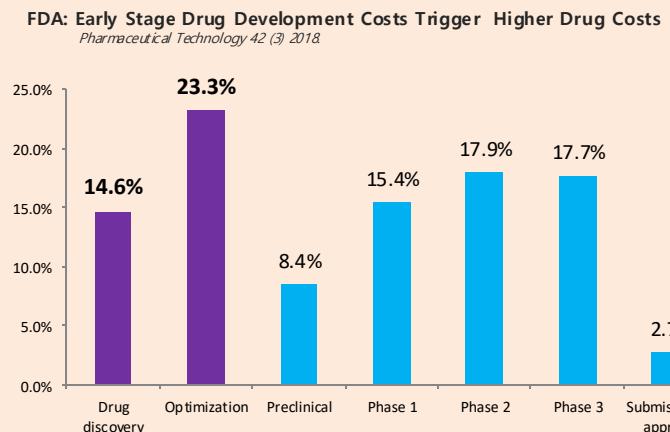
### R&D 수익률



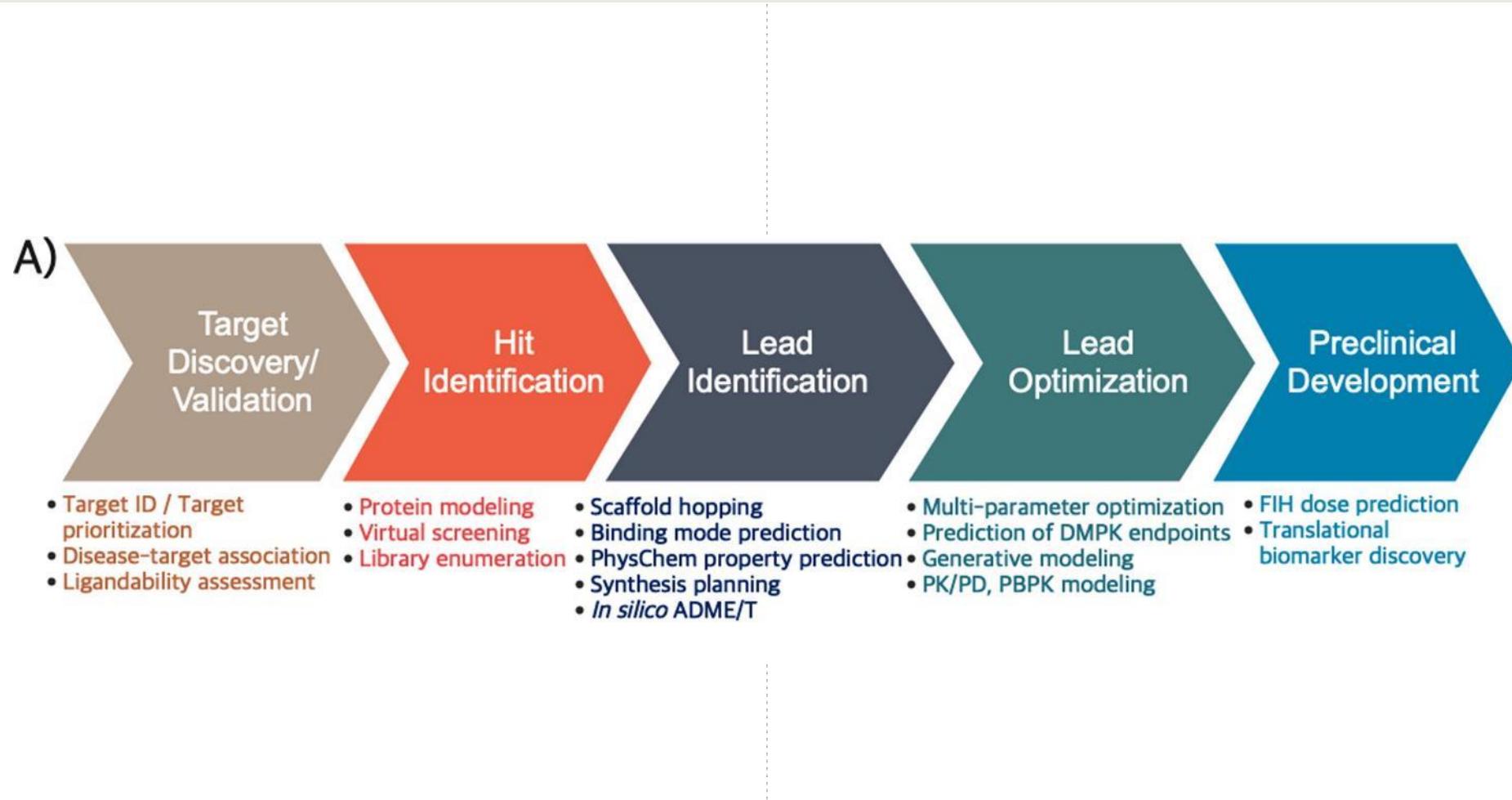
### 신약개발의 성공률



### 신약개발의 단계별 비용



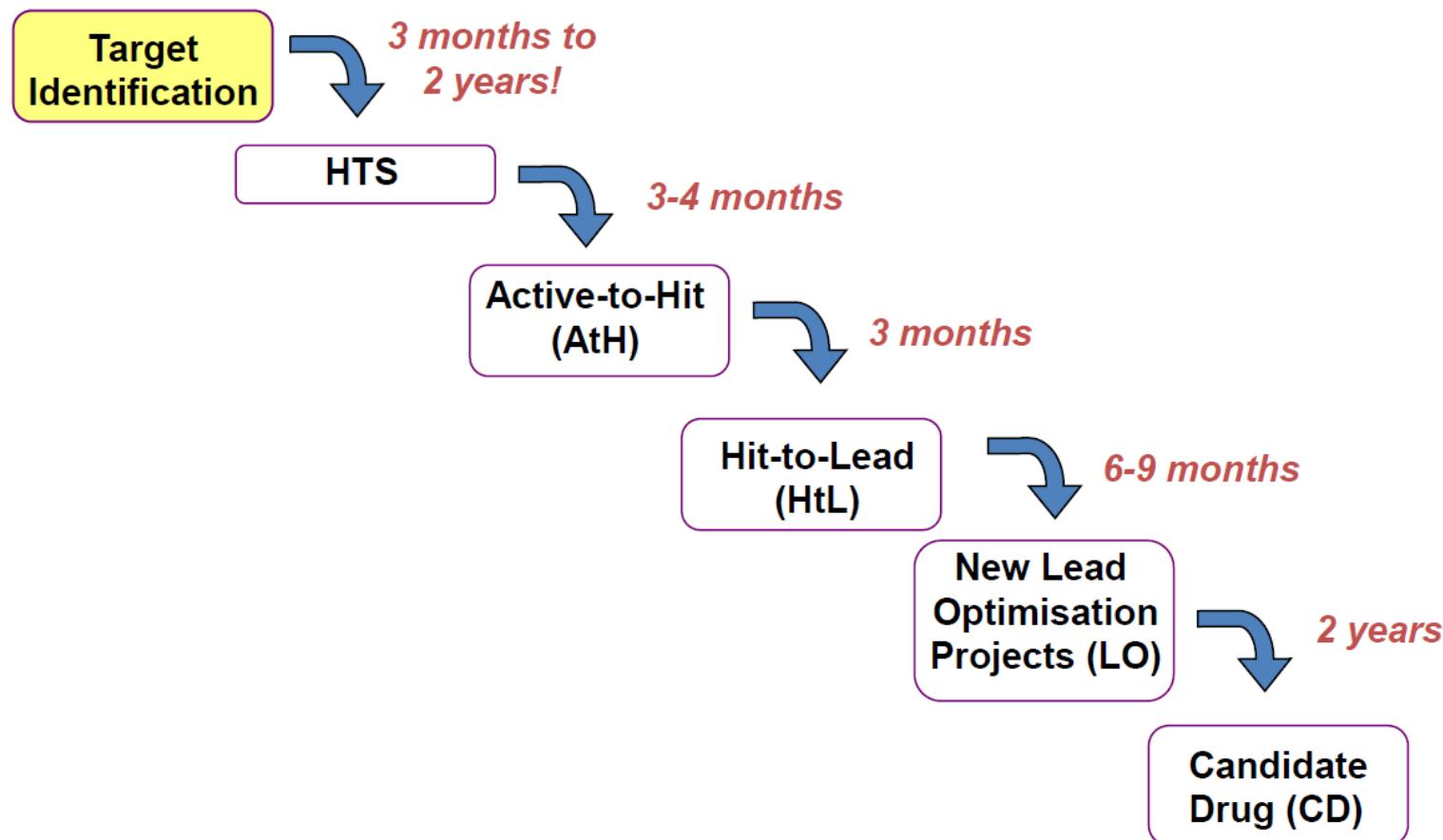
- 신약개발 프로세스와 AI 적용



## • 신약개발 프로세스 용어 정의

	<b>Hit (Identification)</b>	<b>Lead (Generation)</b>	<b>Candidate (Lead optimization)</b>
구조적 특징	<ul style="list-style-type: none"> <li>. 최소 2개 이상 모핵 필요 (신규성 여부는 필수 X)</li> <li>. Clear SAR</li> <li>. <i>In vitro</i> 활성 확보</li> </ul>	<ul style="list-style-type: none"> <li>. 신규 모핵 확보 필수 (최소 2개 이상)</li> <li>. 기본적 ADMET 만족 (Structural alert 배제)</li> <li>. <i>In vivo</i> 효력 확보</li> </ul>	<ul style="list-style-type: none"> <li>. 각 모핵의 치환기 최적화</li> <li>. PK를 고려한 치환기/모핵 변화</li> <li>. 독성(hERG, CYP 저해) 고려</li> </ul>
특허성	가능성만 판단	필수, 이 시점에서 출원 가능	<ul style="list-style-type: none"> <li>. 최적 물질특허 출원 시기</li> <li>. 보강 특허 출원</li> </ul>
화합물 평가 항목	<ul style="list-style-type: none"> <li>. IC50 &lt; 1 <math>\mu</math>M</li> <li>. 필요에 따라 hERG 평가</li> </ul>	<ul style="list-style-type: none"> <li>. IC50 &lt; 100 nM</li> <li>. 기본적 off-target 평가</li> <li>. 기본적 물리화학적 지표 만족</li> <li>. <i>in vitro</i> &amp; <i>vivo</i> PK, Tox 평가</li> </ul>	<ul style="list-style-type: none"> <li>. IC50 &lt; 100 nM</li> <li>. 임상 1상 IND용 평가</li> <li>. TPP를 고려한 평가 필요 (즉, 차별화 결과 확보가 중요)</li> </ul>

- 전통적인 신약개발 프로세스



- 전통적인 신약개발 프로세스



1. Novel target / Validated target

- Transgenic animals
- Disease animal model
- Screening method
- TPP setting

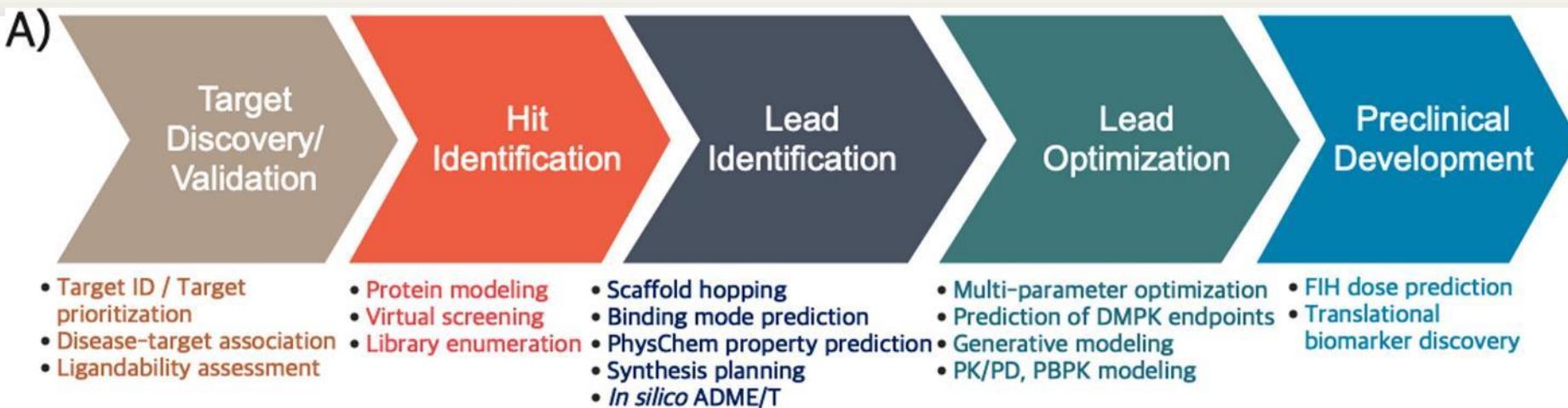
- 전통적인 신약개발 프로세스



## 2. Hit generation (Scaffold generation)

- AI & Virtual screening
- Combinatorial chemistry
- High throughput screening
- Fragment-based hit identification
- DNA-Encoded Library screening
- ...

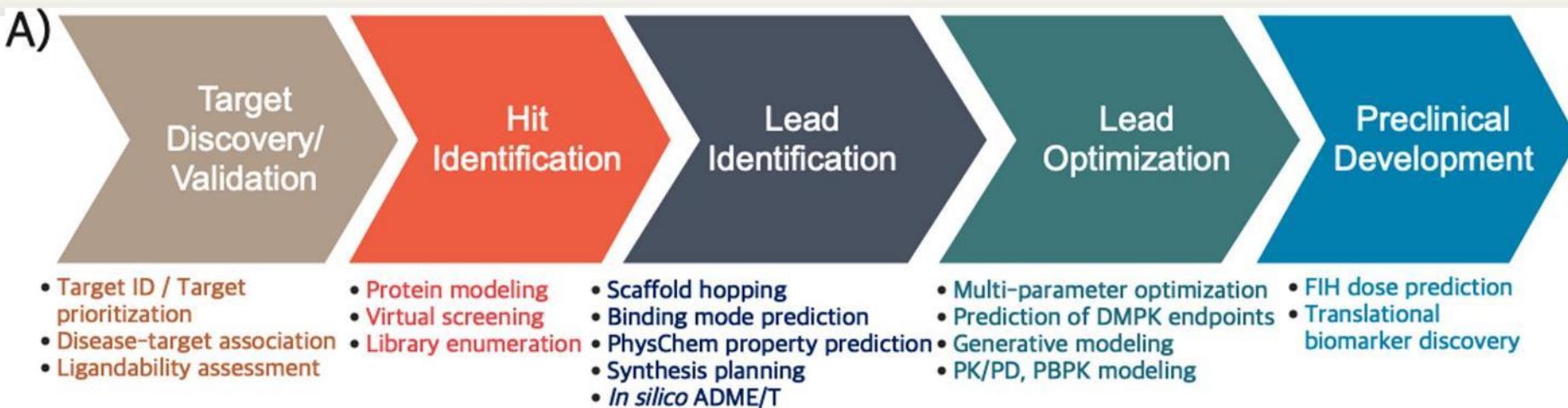
- 전통적인 신약개발 프로세스



### 3. Lead generation

- ADME screening
- Single toxicity
- *in vivo* efficacy study
- IP possible

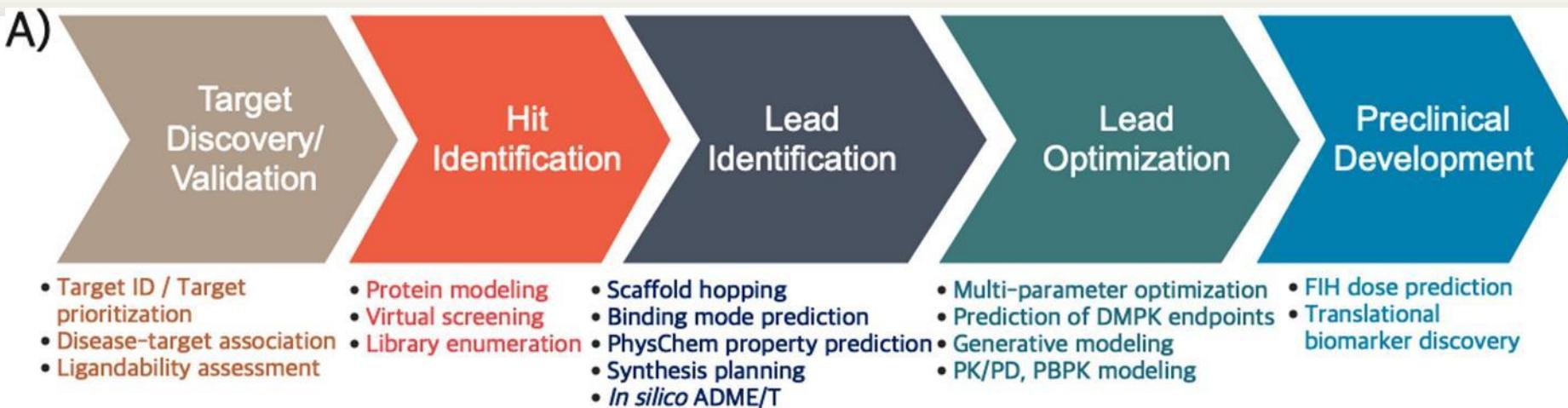
- 전통적인 신약개발 프로세스



#### 4. Candidate selection

- Disease(TPP) based efficacy study
- Repeated dose toxicity / hERG assay / Ames test
- Safety pharmacology
- Potential market (Competitiveness)

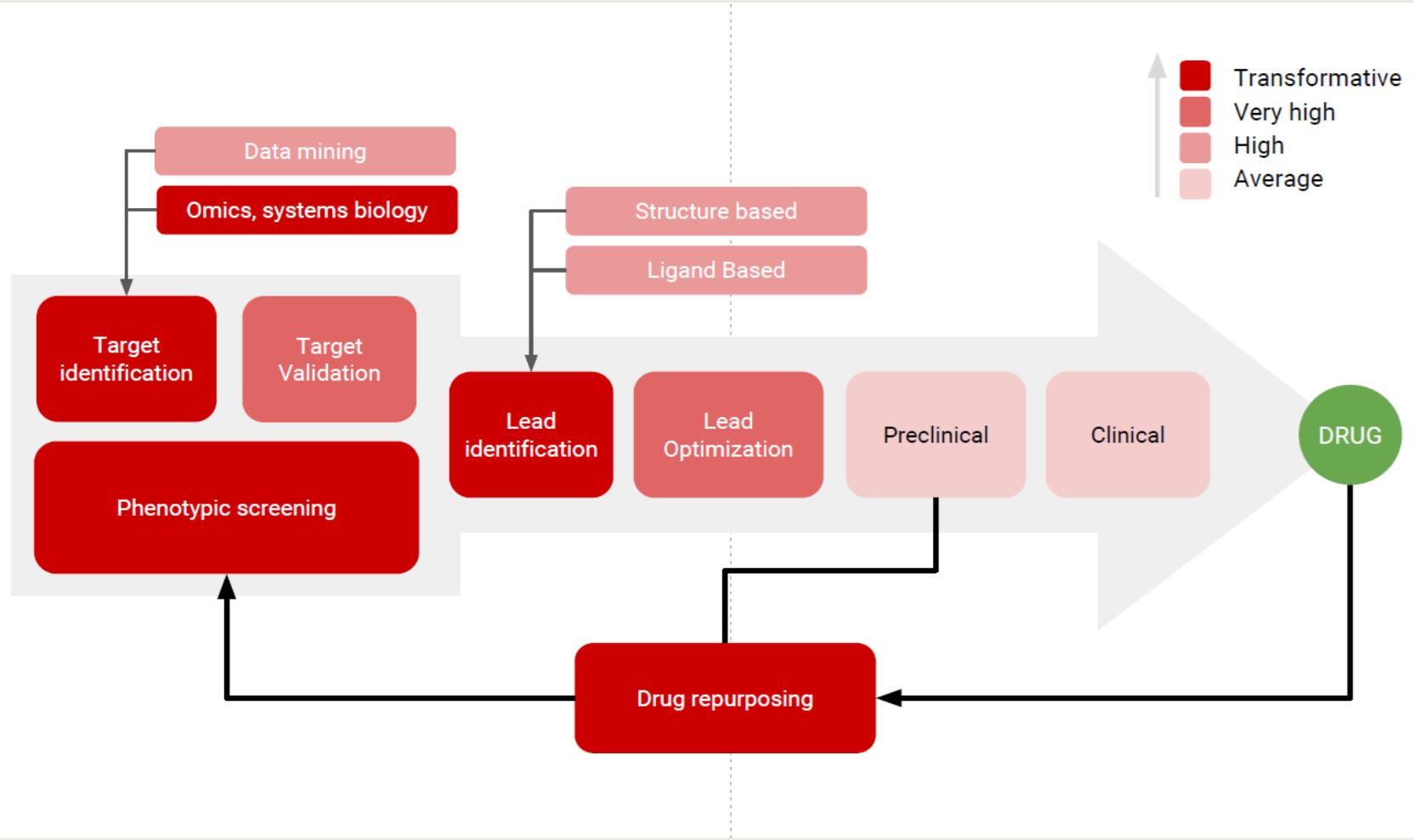
- 전통적인 신약개발 프로세스



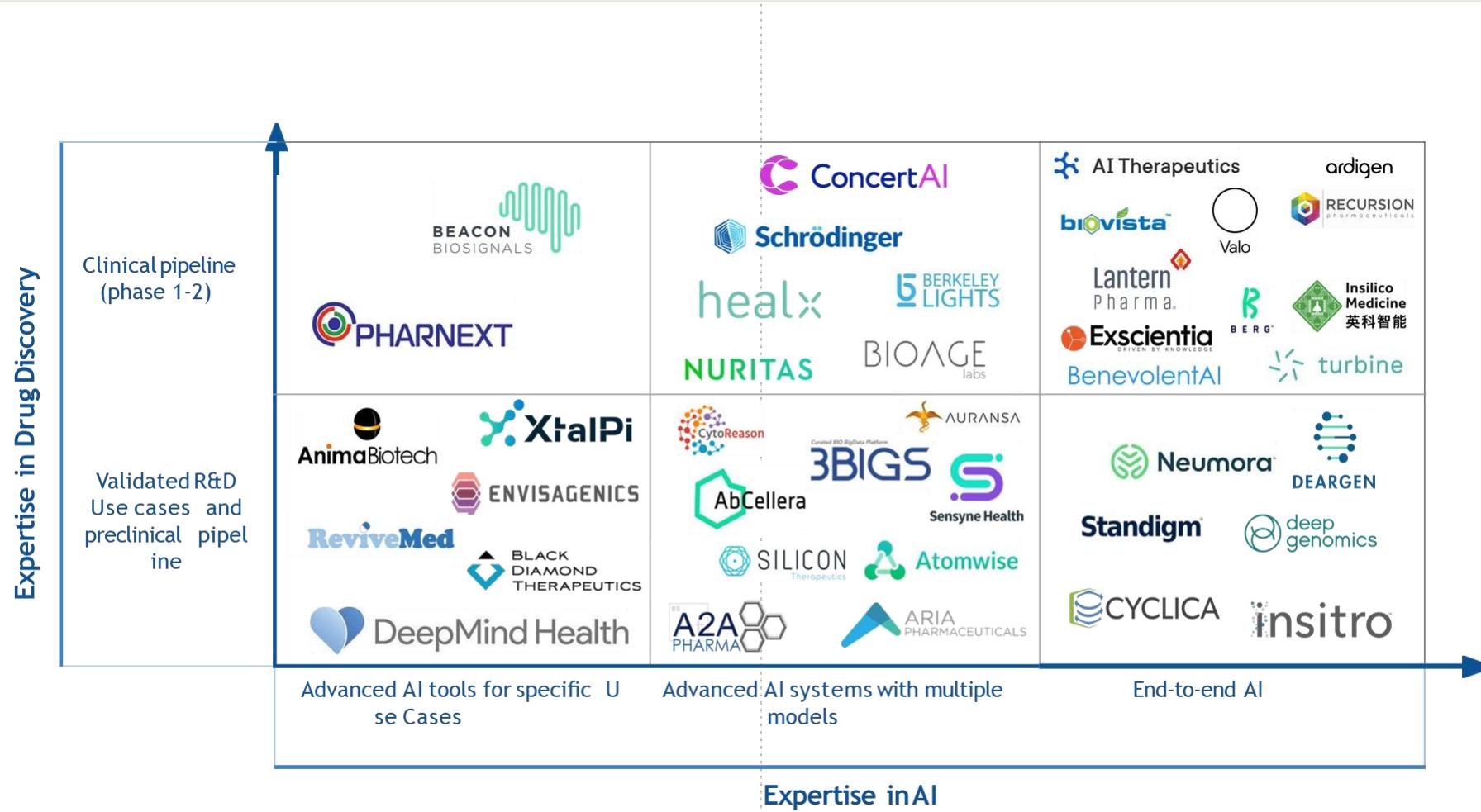
## 5. Pre-Clinical

- Pre-formulation (salt, crystalline)
- Chemical manufacturing
- DMPK study (RI & metabolites)
- Pharmaceutical & Analytical study
- GLP-Tox (4~6 W, rodent & non-rodent)
- Clinical protocol
- IND documentation
- TPP re-setting
- Back-up study

## • 신약개발 프로세스와 AI 적용



## • 신약개발 프로세스와 AI 적용

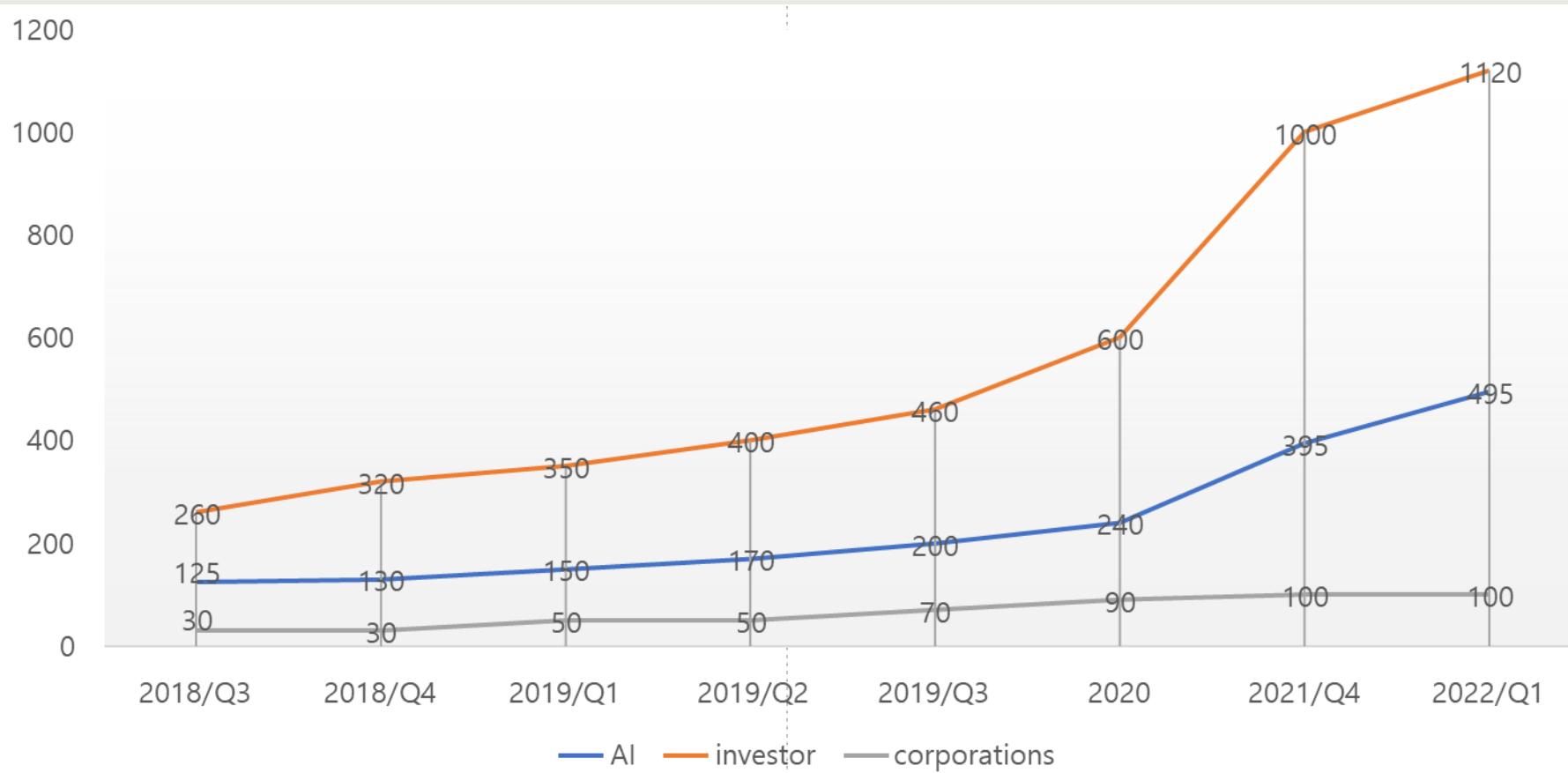


- AI 신약개발 Start-up

## 신약 개발에서의 AI start-up 기업들

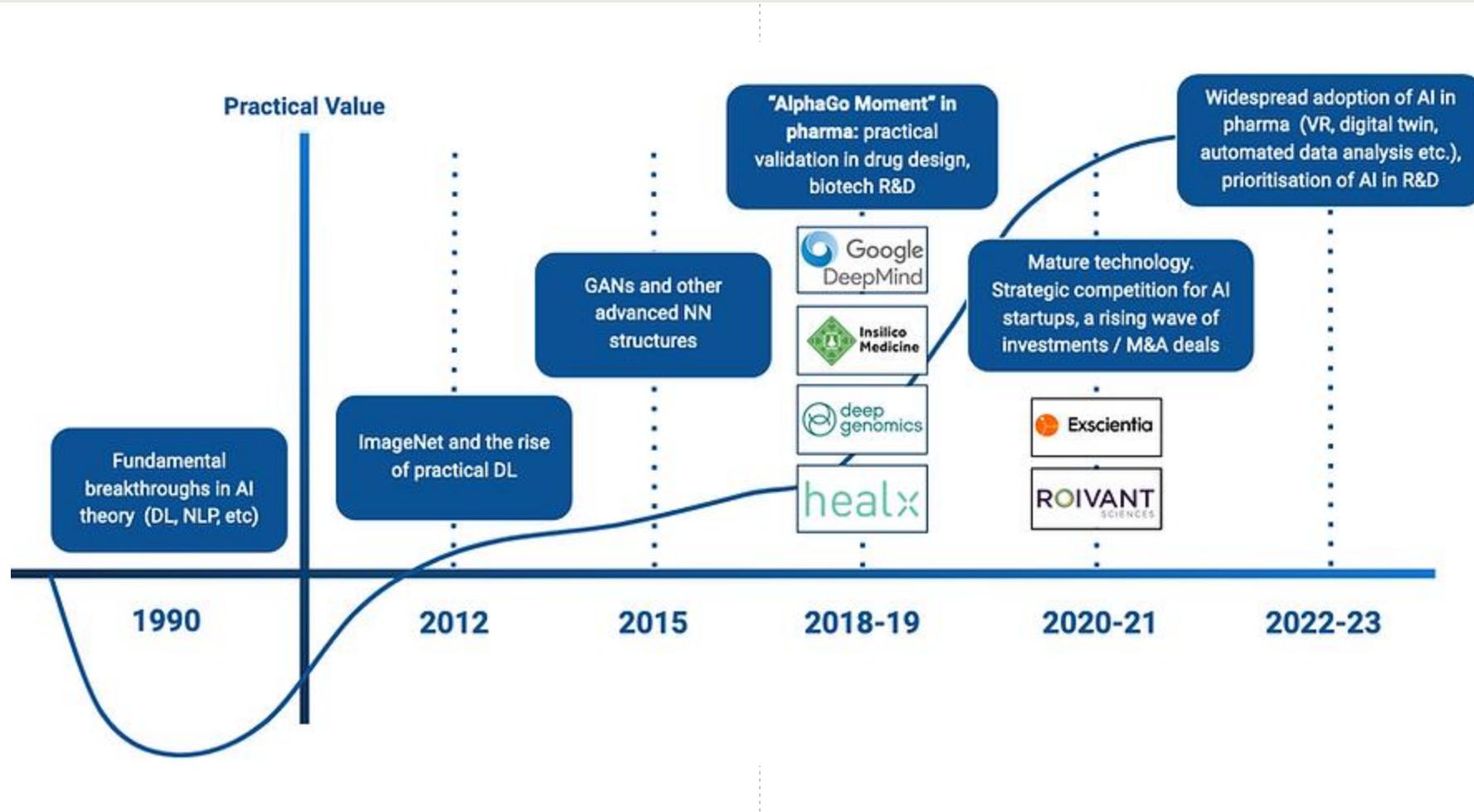
Competitor	Raised capital	Year funded	Employees	Approach
<a href="#">BenevolentAI</a>	\$202M (\$2B val)	2013	51-100	Drug discovery, clinical trial simulation, biomarker ID, mechanism of disease with <b>knowledge graphs</b>
 flatiron  Roche	(\$1.9B val)	2012	251-500	External control arms (EHR), analysis through <b>linked EMR and genomic data</b>
<a href="#">healx</a>	\$61.9M	2014	11-50	Drug repositioning with <b>knowledge graphs</b>
 GNS HEALTHCARE	\$54.3M	2000	101-250	Drug discovery, clinical trial simulation, biomarker ID, mechanism of disease with <b>knowledge graphs</b>
 Atomwise	\$51.3M	2012	11-50	Drug discovery with <b>chemoinformatics</b> and CNNs
 Exscientia DRIVEN BY KNOWLEDGE	\$43.7M	2012	11-50	Drug discovery from AI to experimental, using <b>chemoinformatics</b> and phenotypic screening
 OWKIN	\$18.1M	2016	11-50	Drug discovery, clinical development optimization through <b>knowledge graphs</b>
 twoAR	\$14.3M	2014	11-50	Drug discovery, clinical trial simulation with <b>linked EHR and biomolecular data</b>

- AI 신약개발 스타트업



*AI for Drug Discovery, Biomarker Development and Advanced R&D landscape / 2018.Q3~2022.Q1*

## • AI 신약개발 기술



- 최신 AI 기술을 활용한 신약개발

## 1. AI와 신약개발의 융합

- 초기 약물 발굴: AI는 신약개발의 초기 단계에서 수백만 개의 화합물 중에서 유망한 후보를 빠르게 식별하여 연구 속도를 크게 향상시킵니다.
  - 기술 활용: 가상 스크리닝, QSAR 모델링, 분자 도킹 등이 활용되어 효율성을 극대화합니다.
  - 효과: 비용 절감, 개발 기간 단축, 더욱 효율적인 화합물 최적화 ([Drug Discovery and Development](#)) ([StartUs Insights](#)).

## 2. Diffusion Models의 도입

- 기술 설명: Diffusion models는 화합물의 분포를 점진적으로 변형하여 새로운 화학 구조를 생성하는 최신 AI 기술입니다. 이 기술은 기존 GAN 모델보다 안정적이며, 특히 복잡한 데이터 분포에서 우수한 성능을 발휘합니다.
- 응용: 고다양성 화합물 생성, 신약 후보군 최적화. 더 나은 다양성과 화합물 특성의 보존력 제공 ([McKinsey & Company](#)).

- 신약개발의 미래 전망

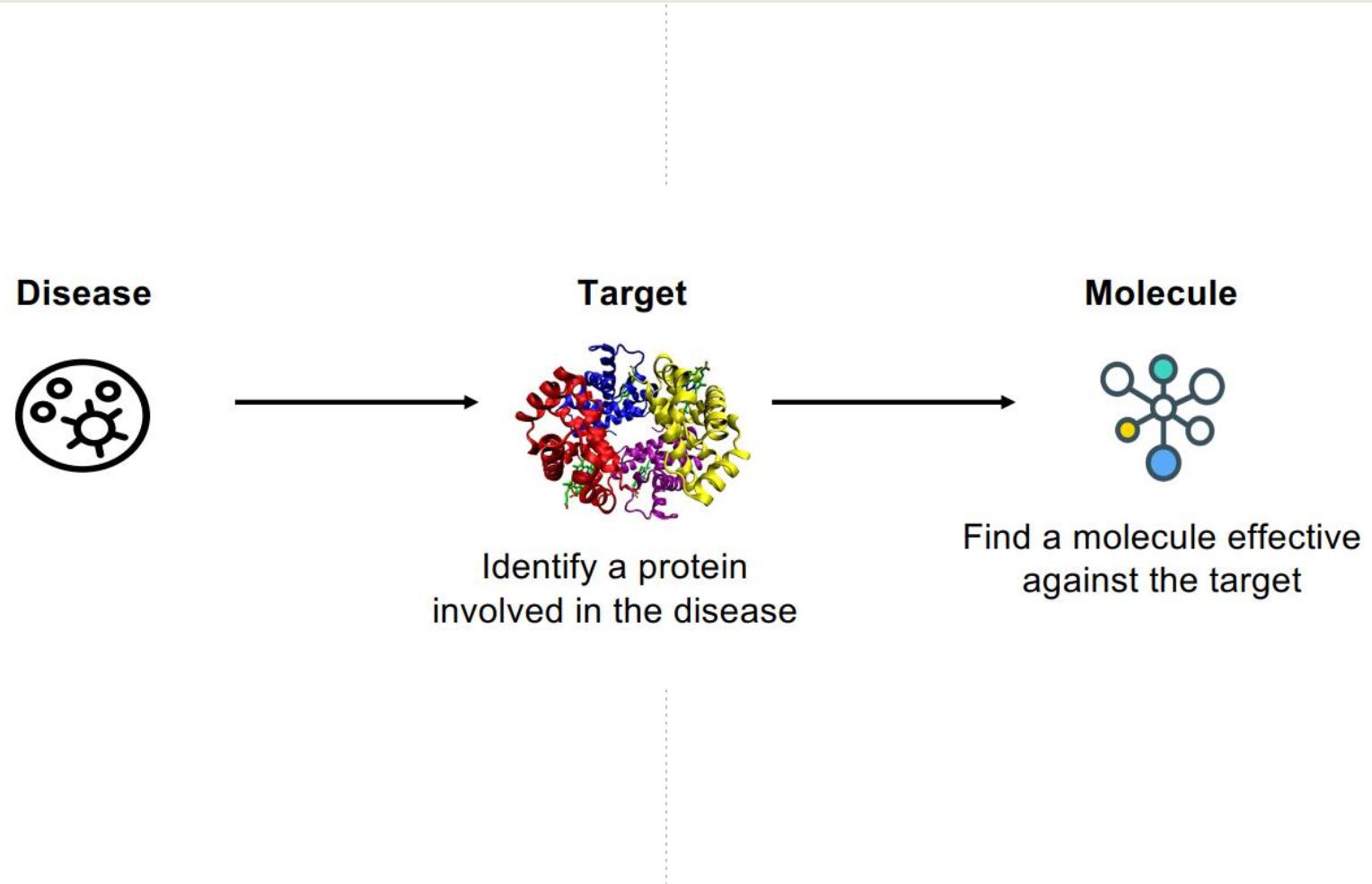
## 1. Large Language Models (LLMs)의 역할

- **기술 설명:** LLMs는 대규모 텍스트 데이터를 학습하여 자연어 처리 및 생물학적 텍스트 분석에서 탁월한 성능을 발휘합니다. 이를 통해 과학 문헌에서 중요한 정보를 추출하고, 새로운 약물 타겟을 예측하며, 잠재적인 약물 상호작용을 식별할 수 있습니다.
- **응용:** 연구 속도 가속화, 새로운 치료 전략 제안, 더 정확한 데이터 분석 ([StartUs Insights](#)).

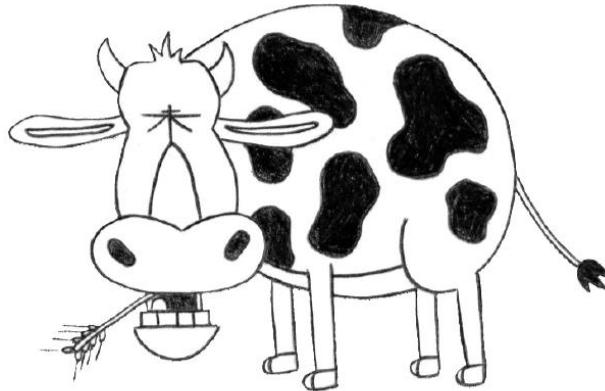
## 2. 신약개발의 미래 전망

- **생성적 AI의 발전:** Diffusion models와 LLMs는 기존 기술을 대체하며 신약개발에서 중요한 도구로 자리 잡고 있습니다. 이들 기술은 신약 후보의 예측 정확도를 높이고, 더 빠르고 정교한 치료제 개발을 가능하게 합니다.
- **Agentic AI의 발전:** 자율 결정과 도구를 사용하여 자율 실행이 가능한 Agentic AI의 등장으로 혁신적인 신약의 발굴이 가능하게 될 것입니다.
- **기대 효과:** 더 높은 예측 정확도, 혁신적인 신약 발굴 가속화, 개인 맞춤형 치료제 개발 가능성 확대 ([McKinsey & Company](#)) ([StartUs Insights](#)).

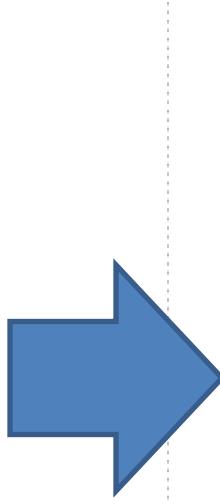
- 신약 개발에서의 데이터 정의



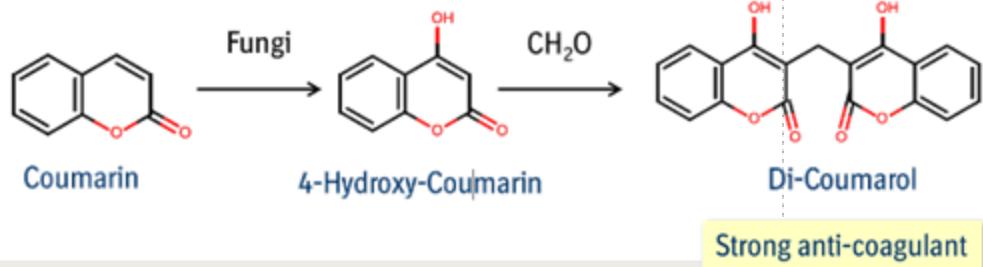
- Disease, Target, Molecule 관계



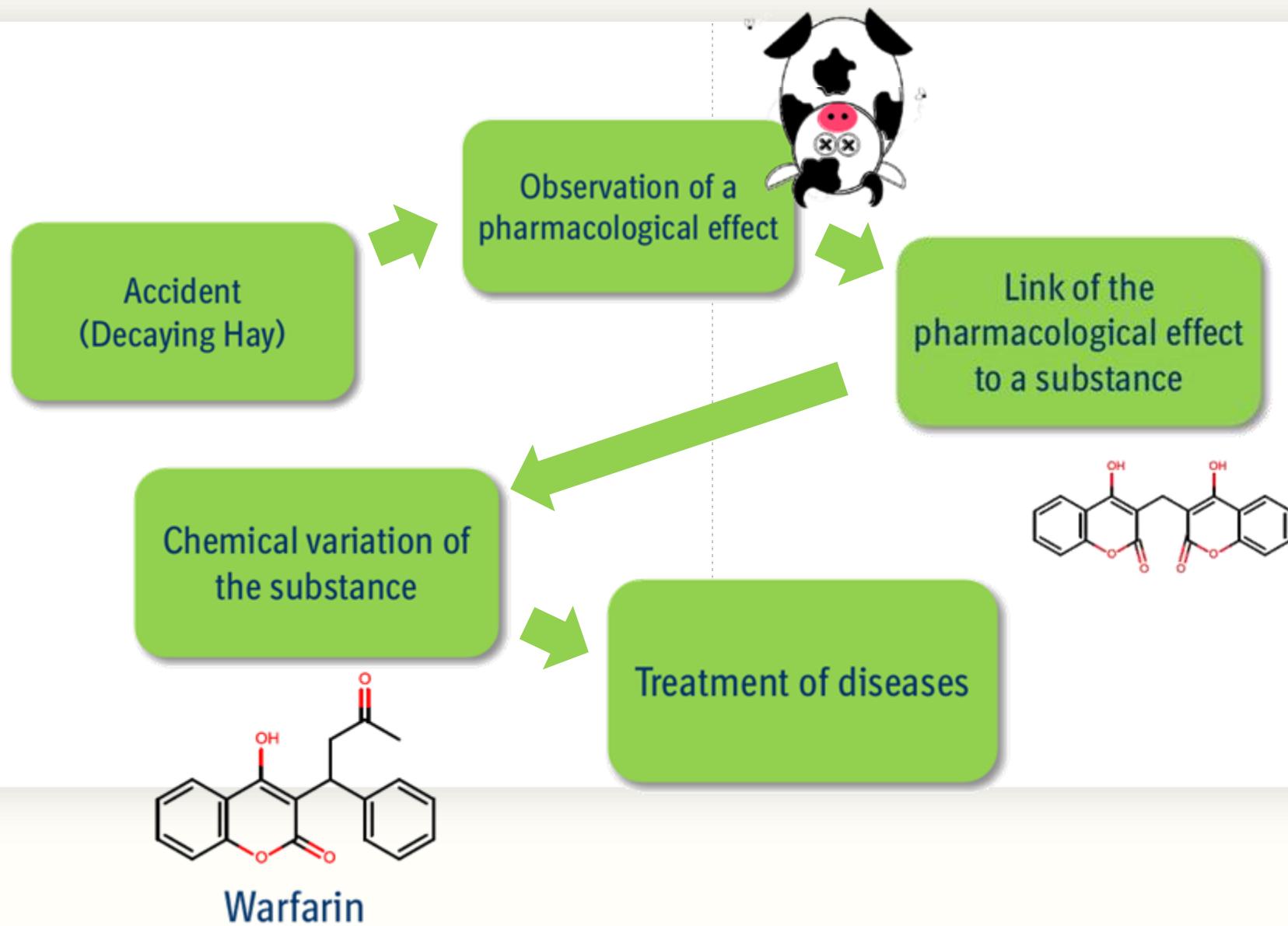
Cattle, 1920s, North Dakota



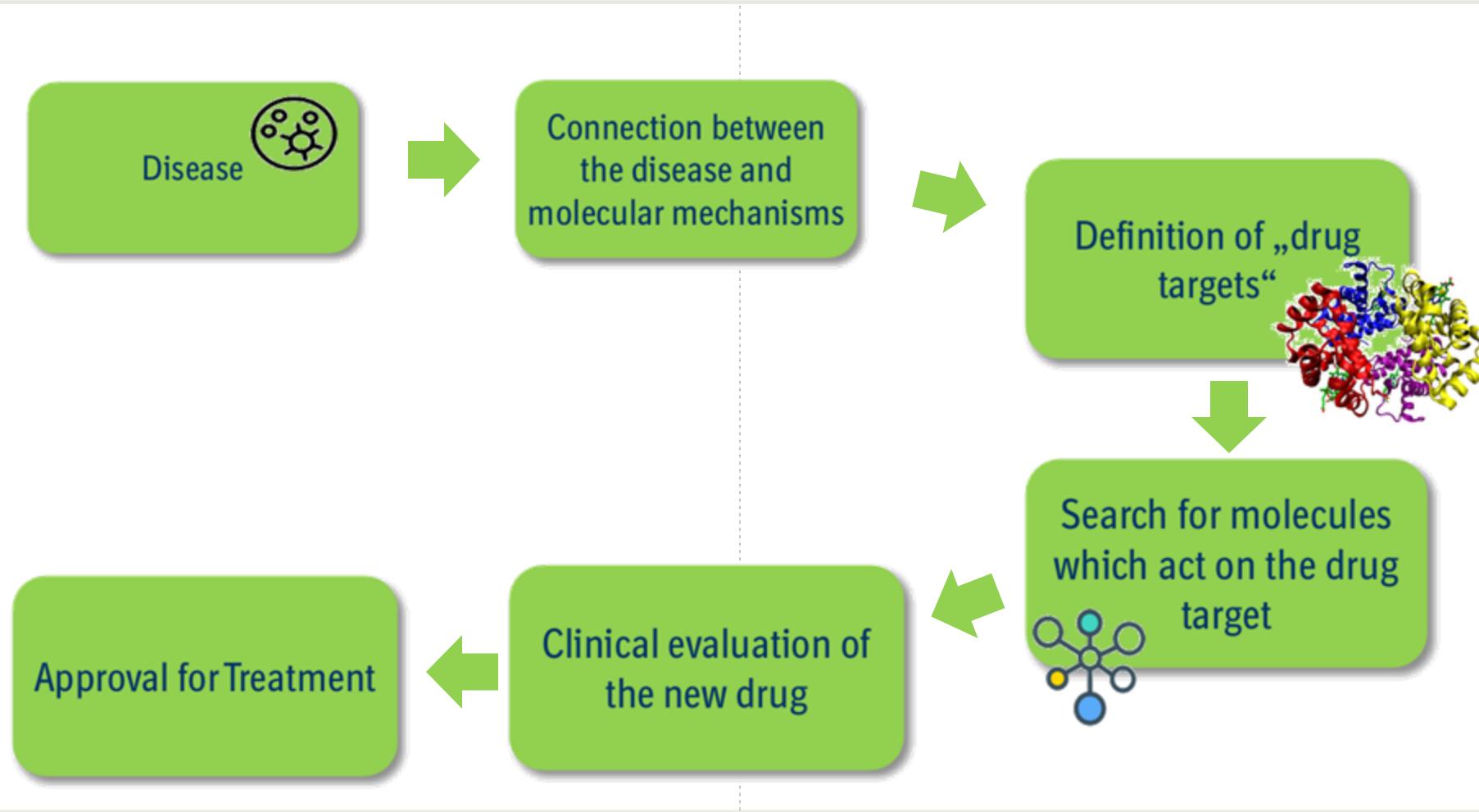
Dead Cattle, 1920s, North Dakota



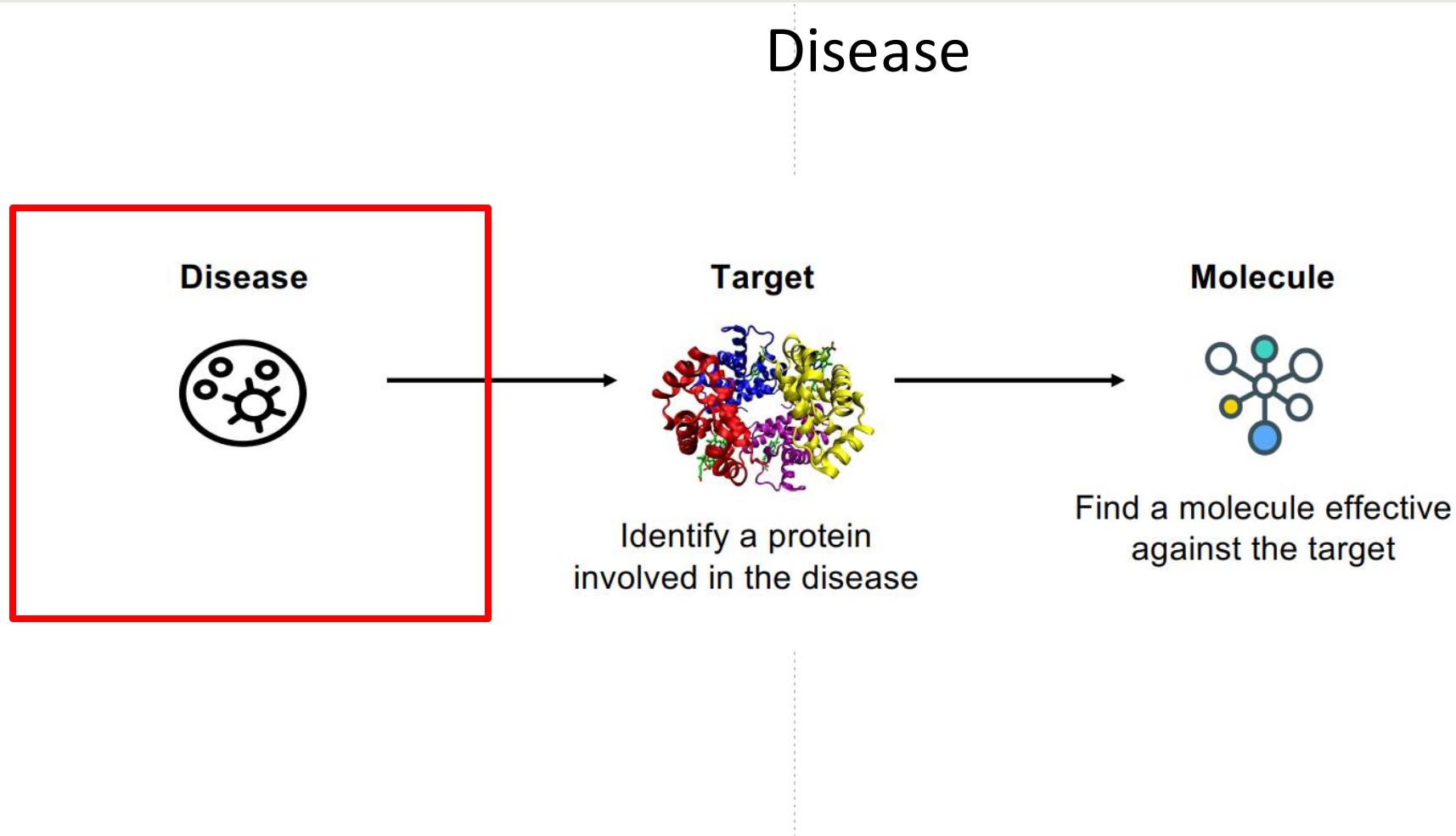
- Disease, Target, Molecule 관계



- Disease, Target, Molecule 관계



- Disease, Target, Molecule 관계



- Disease database

## Disease

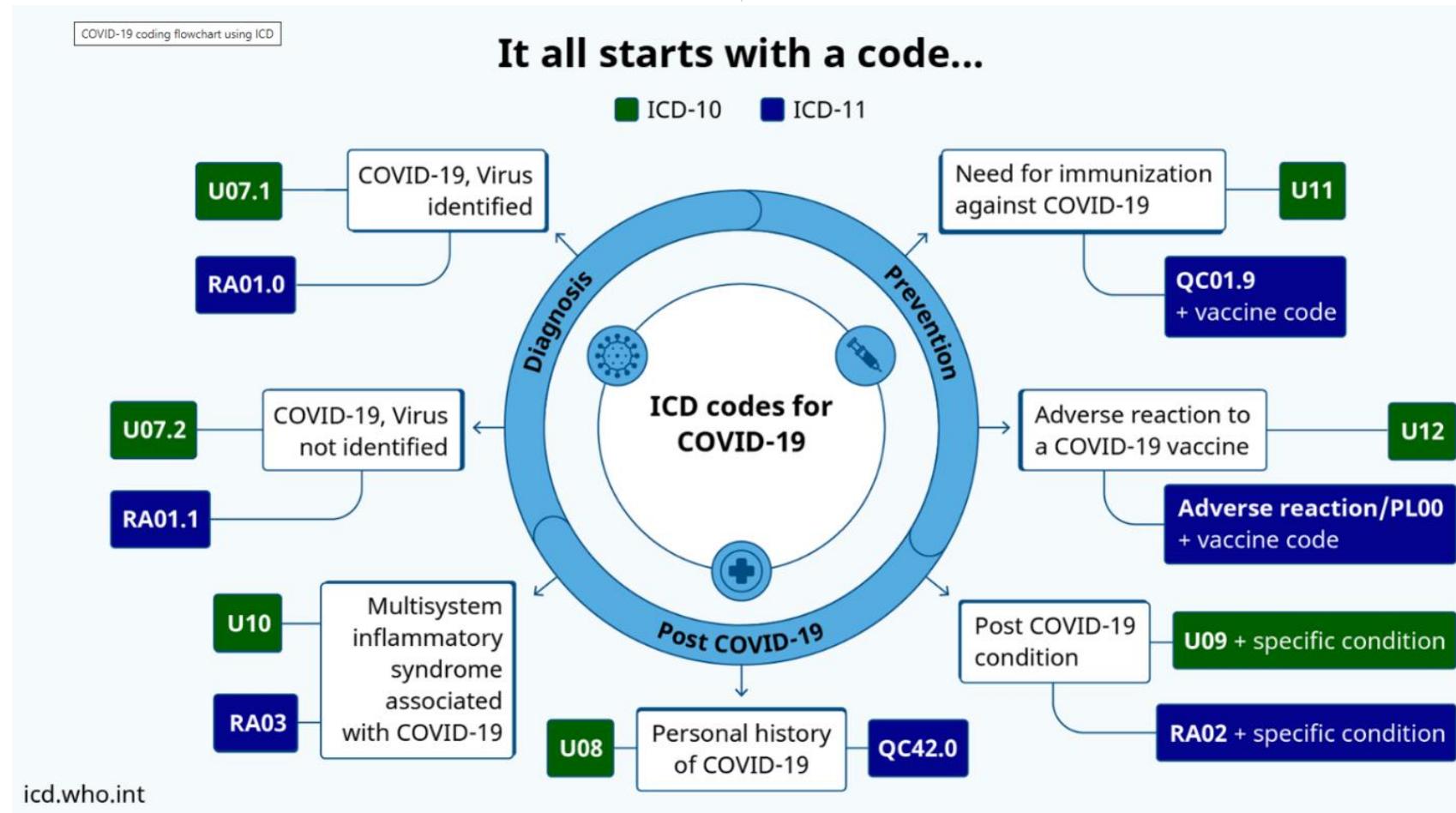
- ICD-11/ICHI : <https://www.who.int/standards/classifications/classification-of-diseases>
  - ICD
    - International Classification of Diseases
    - International Statistical Classification of Diseases and related health problems
  - ICHI
    - International Classification of Health Interventions
    - Interventions for treatment, prevention and diagnostics: medical, nursing, rehabilitation, laboratory, imaging, ultrasound, public health.

- Disease Database

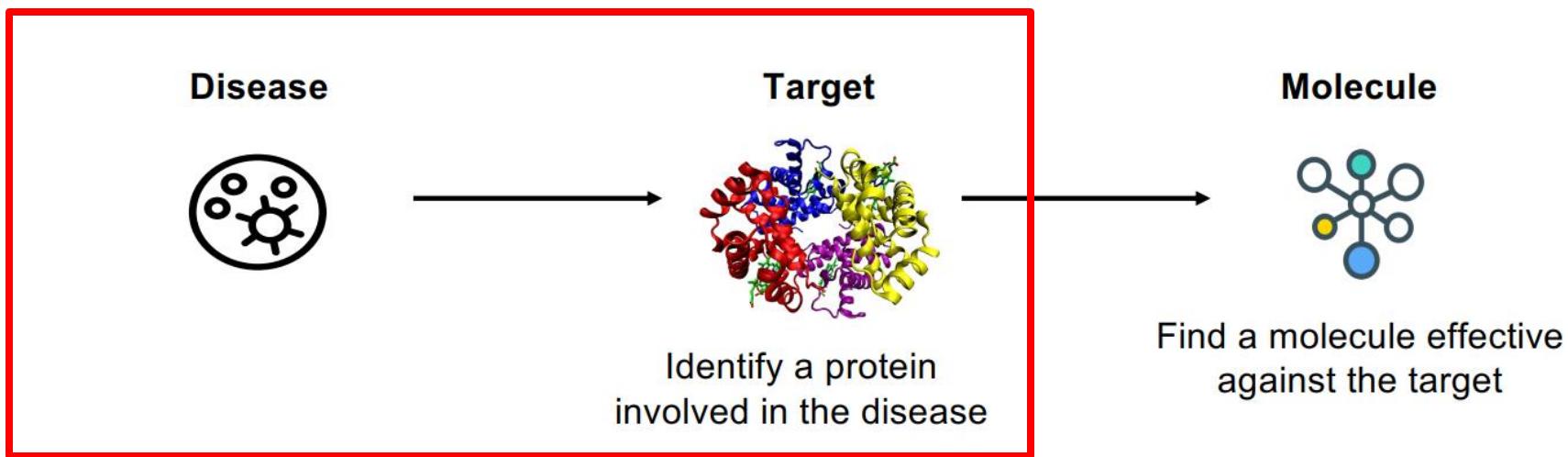
## ICD-11

- Health data standard(WHO Constitution and Nomenclature Regulations).
- 2022년 1월 시행
- 언어와 문화에 구애 받지 않는 개념적 framework.
- 용어와 분류의 통합.
- End-to-end digital solution(API, 도구, 온라인 및 오프라인).
- 최신 과학 지식
- 비교 가능한 통계 및 의미론적 상호 운용성(150년)
- ICD-11은 누구나 이용 가능
- Creative Commons Attribution-NoDerivs 3.0 IGO license

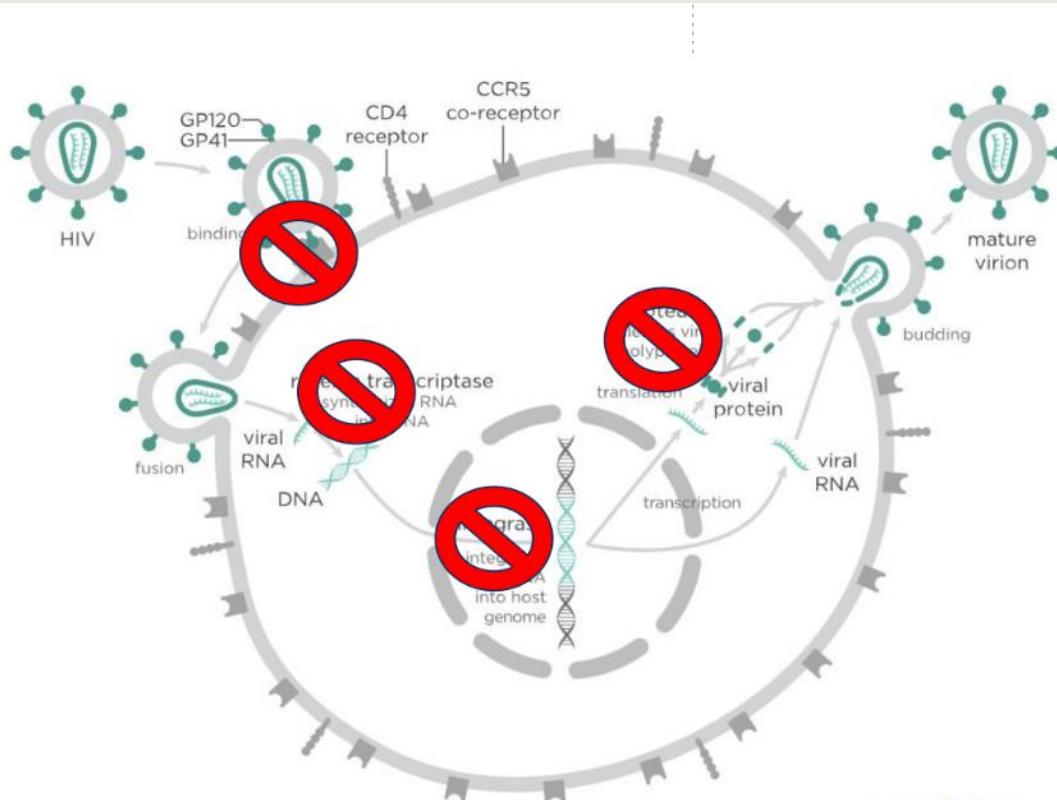
- ICD – COVID-19 질병 관련 비상 사용 ICD 코드



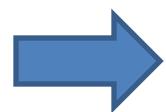
# Disease



- The Life Cycle of the HIV Virus

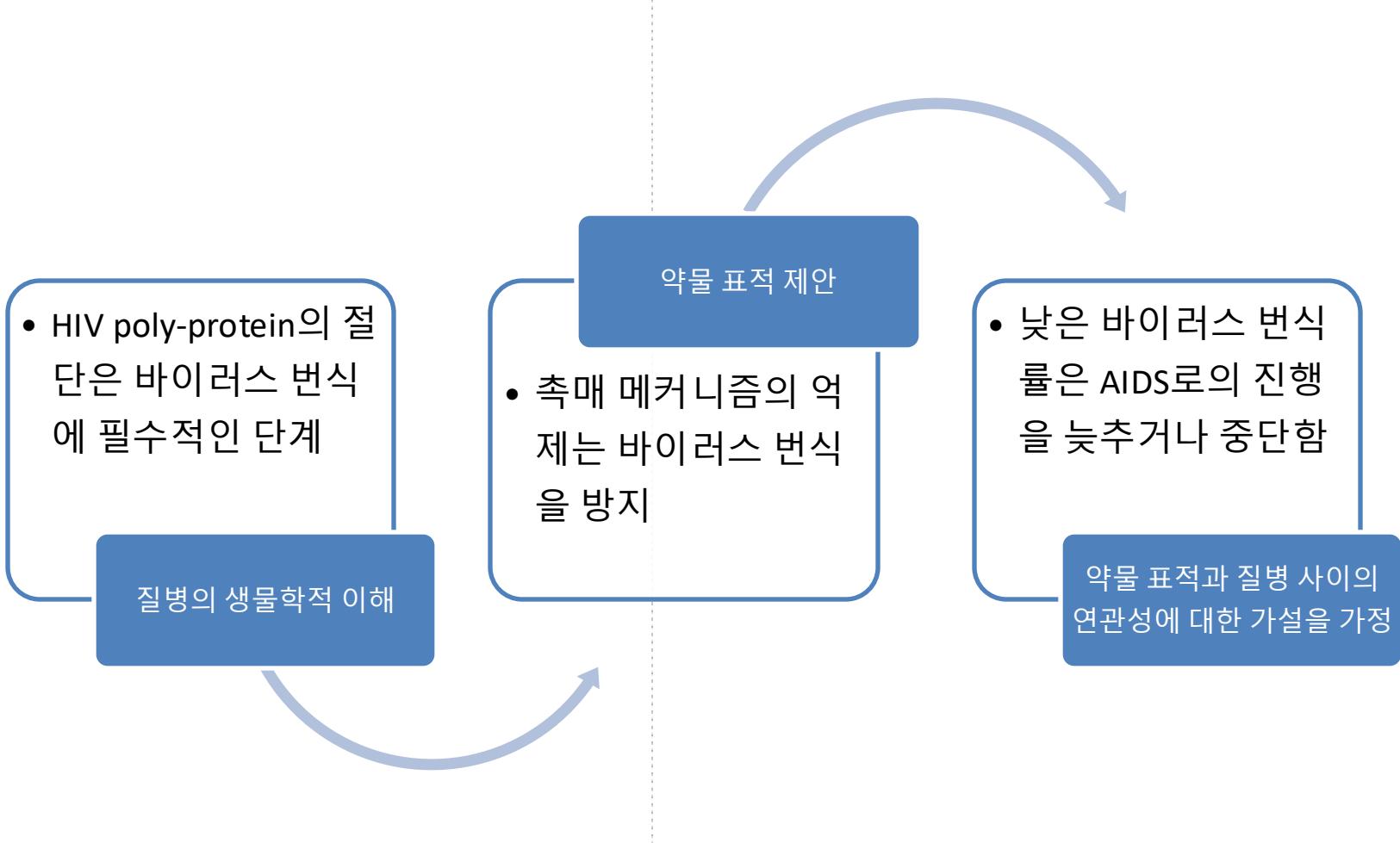


Source: Wikipedia

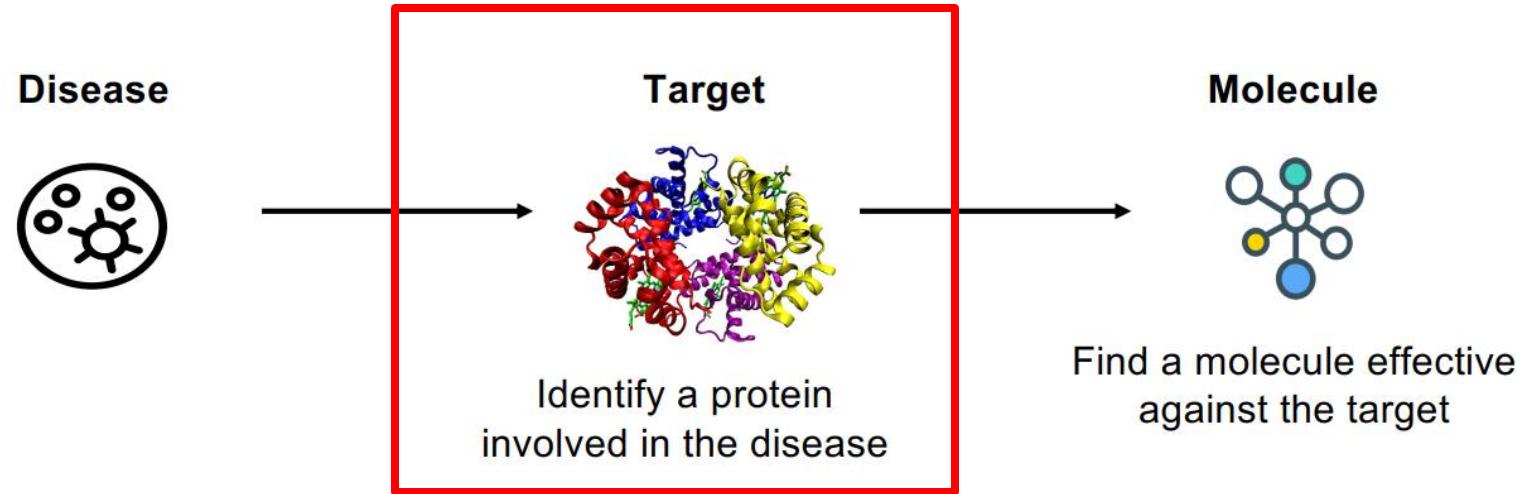


바이러스의 수명주기 중 한단계를 블럭하게 되면  
바이러스의 증식이 중단됨

- Disease 와 Target의 관계

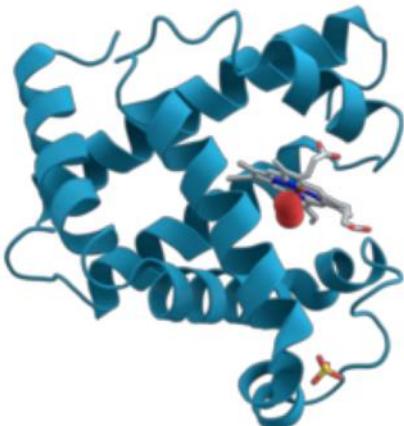


# Target

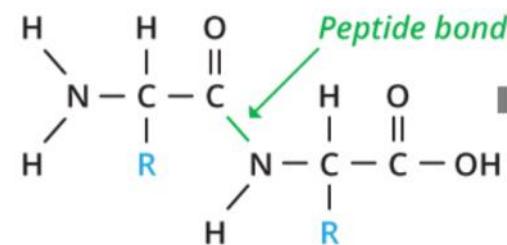


# Protein Targets

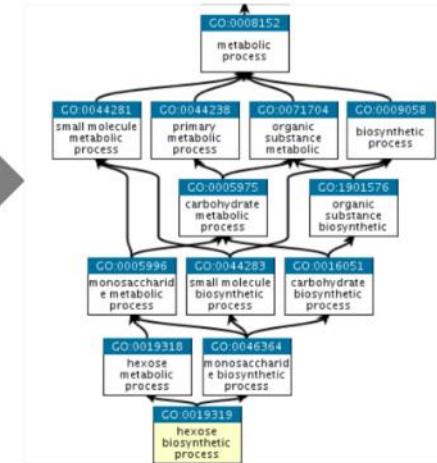
Biology  
3-D structure



A Sequence of  
23 Amino Acids

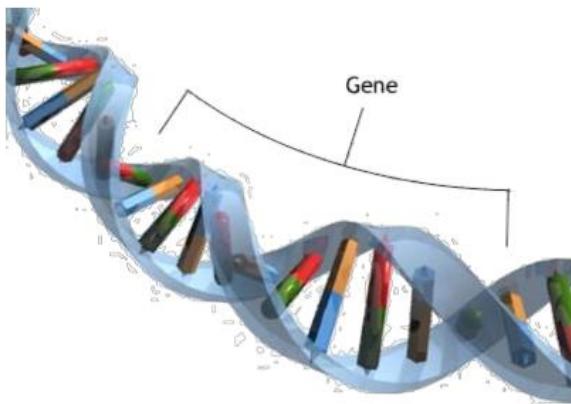


GO term  
Protein Function

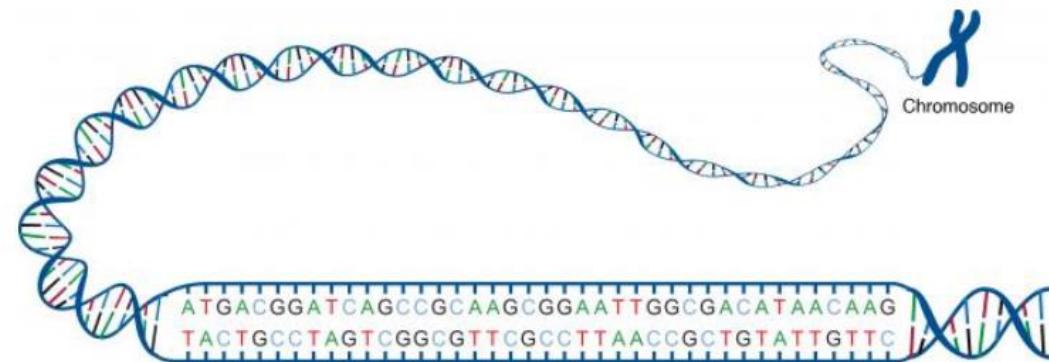


# Gene Sequence

## Double Helix DNA structure



## A-T, C-G allele representation



## • Phenotype-based vs Target-based drug design 비교



### Phenotype-based

표현형 기반으로 약물 효과를 먼저 발견하고, 후속적으로 표적을  
발굴하는 접근법

대규모 스크리닝 → 효과 확인 → 표적 발굴

- ✓ Bottom-up 접근법 (효과 → 표적)
- + 장점: 생체 내 효과가 이미 검증됨
- 단점: 메커니즘 규명이 어려움

대표 사례: 아스피린 - 항염증 효과 발견 후 COX 억제 규명



### Target-based

질병 관련 표적을 먼저 발굴하고, 해당 표적에 작용하는 약물을  
개발하는 접근법

표적 발굴 → 약물 설계 → 효능 검증

- ✓ Top-down 접근법 (표적 → 효과)
- + 장점: 합리적 약물 설계 가능
- 단점: 생체 내 복잡성 간과 가능성

대표 사례: Gleevec - BCR-ABL 표적으로 CML 치료 성공

#### Phenotype-based

성공률

약 25%

#### Target-based

약 50%

현대 활용도

30%

70% (주요 접근법)

출처: Nature Reviews Drug Discovery (2024), Drug Discovery Today (2023)

## • Phenotype-based vs Target-based 상세 비교

Phenotype-based 표현형 기반 접근법		Target-based 표적 기반 접근법
접근 출발점	관찰 가능한 질병 표현형에서 출발 예: 세포 생존율 변화, 단백질 발현 변화	특정 분자 표적에서 출발 예: 단백질 구조, 유전자 변이
역사적 맥락	전통적 약물 발견의 주요 방법론 ~1990년대 이전 주류 방법론	인간 게놈 프로젝트 이후 발전 1990년대 이후 주류 방법론
검증 프로세스	효과 → 표적 규명 → 메커니즘 역방향 추적 (효과에서 표적으로)	표적 → 약물 개발 → 효과 확인 정방향 추적 (표적에서 효과로)
지식 요구사항	질병 메커니즘에 대한 사전 지식 불필요 경험적 접근 가능	분자 표적에 대한 상세 지식 필요 메커니즘 기반 접근
시간 및 비용	초기 발견 단계 빠름 표적 규명에 추가 시간 소요	초기 표적 발굴에 시간 소요 약물 최적화 과정 효율적
핵심 고려사항		
현대 신약개발에서는 두 접근법의 <b>상호보완적 활용</b> 이 중요하며, AI 기술은 두 방법론 간의 격차를 줄이고 효율성을 높이는 데 기여하고 있습니다.		

출처: Swinney & Anthony (2024), Nature Reviews Drug Discovery; Lee et al. (2023), Current Drug Target Methods

- Druggability



특정 생체 내 표적(단백질, 효소 등)이 약물로 조절될 수 있는 정도



✓ 효율적인 약물 타겟 선정

✓ 신약 개발 성공률 향상



◎ Druggable Target 특성



锁定特异结合



低分子量/亲水性平衡



주요 표적 사례

GPCR

Kinase

Ion Channel

Hormone

출처: Hopkins, A.L. & Groom, C.R. Nature Reviews Drug Discovery, 2002 / Nature, 2021

# Druggable vs Undruggable Target의 정의와 특성

## ➊ 핵심 정의

**Druggability (약물성):** 생체 표적이 약물에 의해 조절될 수 있는 잠재력으로, 약물 결합 포켓의 존재와 접근성, 물리화학적 특성에 기반함

☑ Druggable Target		☒ Undruggable Target	
☒ 구조적 특성	<ul style="list-style-type: none"><li>명확한 약물 결합 포켓</li><li>구조적 안정성</li><li>결합 부위 접근성 높음</li></ul>	☒ 구조적 특성	<ul style="list-style-type: none"><li>뚜렷한 결합 포켓 부재</li><li>평평한 단백질-단백질 상호작용 표면</li><li>구조적 유연성 높음</li></ul>
▣ 물리화학적 특성	<ul style="list-style-type: none"><li>친수성/소수성 균형</li><li>결합에 적합한 정전기적 표면</li><li>높은 결합 친화성</li></ul>	▣ 물리화학적 특성	<ul style="list-style-type: none"><li>소수성 불균형</li><li>동적으로 변화하는 구조</li><li>넓은 상호작용 표면</li></ul>
◤ 대표 예시	<ul style="list-style-type: none"><li>GPCR (수용체)</li><li>Kinase (인산화효소)</li><li>이온 채널, 효소류</li></ul>	◤ 대표 예시	<ul style="list-style-type: none"><li>KRAS (G12D 변이)</li><li>c-Myc (전사인자)</li><li>내재적 무질서 단백질</li></ul>



## Druggability 패러다임의 변화

과거 'undruggable'로 간주되던 타겟들도 구조 기반 설계, 공유결합 약물, PROTAC 기술, AI 기반 접근법 등을 통해 점차 약물 타겟화가 가능해지고 있습니다.

출처: Hopkins, A.L. et al. Nature Reviews Drug Discovery, 2002

최종 업데이트: 2025.07.02

# 구조적/특성 비교 및 대표 사례

## ▣ 구조적 차이점

### Druggable Target

- 명확한 결합 포켓 형성
- 구조적 안정성 및 고정된 형태

### Undruggable Target

- 평평한 상호작용 표면
- 동적 구조로 결합 부위 변화

## ▶ 대표 사례 분석

타겟 유형	대표 사례	구조적 특징	성공/도전 사례
GPCR	베타 수용체	7개 막관통 도메인, 뚜렷한 결합 포켓	FDA 승인 약물의 ~35%가 표적
Kinase	EGFR, BTK	ATP 결합 부위, 명확한 소수성 포켓	Imatinib 등 다수 성공
KRAS	G12C, G12D	평평한 표면, 얇은 포켓 구조	Sotorasib(G12C), G12D는 도전적
c-Myc	전사인자	내재적 무질서 단백질, 고유 구조 없음	PROTAC 기술로만 접근 시도

### 💡 AI 기반 접근으로 Undruggable의 가능성

AI 구조 예측, 공유결합 억제제, PROTAC 기술로 KRAS G12C(AMG 510, 2021년 FDA 승인)와 같은 불가능했던 타겟이 약물화되고 있으며, 다양한 혁신 기술로 'undruggable' 개념이 재정의되고 있습니다.

# Druggability 실제 사례



## BCL-2 family 단백질

도전 과제: 단백질-단백질 상호작용(PPI)은 전통적으로 약물 표적화가 어려운 영역

Druggability 평가: 구조 기반 분석을 통해 BCL-2 단백질의 소수성 결합 포켓 발견

성과: Venetoclax(ABT-199) 개발 - 최초의 FDA 승인 BCL-2 선택적 억제제(2016)



## GPCR (G-Protein Coupled Receptors)

특징: 높은 Druggability - 7개의 막관통 도메인 구조로 다양한 결합 포켓 형성

성과: 현재 시장에 있는 약물의 약 35%가 GPCR을 표적으로 함

대표 약물: 베타 차단제(고혈압), 오피오이드 수용체 조절제(통증), H1 항히스타민제



## KRAS - 'Undruggable'에서 'Druggable'로

도전 과제: 40년간 'undruggable'로 간주된 암 유발 단백질 - 평평한 표면, 결합 포켓 부재

혁신 접근법: AI 기반 구조 분석으로 KRAS G12C 변이에서 일시적 결합 포켓 발견

성과: Sotorasib (Lumakras) - 최초의 KRAS G12C 억제제로 2021년 FDA 승인



## Druggability 사례에서 배우는 핵심 교훈

Druggability는 고정된 속성이 아니라 기술 발전과 함께 계속 확장되는 개념입니다. AI와 구조 생물학의 융합은 과거 접근 불가능했던 표적들을 약물 개발의 영역으로 끌어들이고 있습니다.



출처: Souers, A.J. et al. ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. Nature Medicine, 2013 / Sci Rep,

2022 / 최신 AI 기반 chemoproteomics 논문 리뷰

최종 업데이트: 2025.07.02

- 전통적 Drug Target Identification 방법

약물 표적 발굴의 전통적 방법론은 수십 년간 발전하며 현대 신약 개발의 기반을 형성했습니다.

- Phenotypic Screening - 세포/조직에서의 효과를 발견한 후 거꾸로 표적을 식별하는 방법
- ☒ Genetic Approach - 유전자 조작 및 녹이웃 기술을 통한 표적 검증
- ▶ Biochemical & Structural Methods - 단백질 기능 및 구조 분석기 반접근법

# • Drug Target Identification: 두 가지 접근법

## 핵심 접근법

### 주요 접근법 Target-to-Drug

- ◎ 질병 표적 → 약물 개발: 질병 메커니즘의 표적을 발굴하고 해당 표적에 작용하는 약물 개발
- 💡 예: 알츠하이머 → A<sub>β</sub> plaque → A<sub>β</sub> 억제제 개발

### 보조 접근법 Drug-to-Target

- 약물 → 표적 발굴: 효과가 확인된 약물이 작용하는 표적 단백질 발견
- ⌚ 예: Aspirin → COX-1/COX-2 발견, 약물 재창출

## 두 접근법의 비교

### 특성 비교

접근 방식	기설 기반	표현형 기반
주요 활용	신약 개발	약물 재창출
성공률	28%	10%

### 핵심 메시지

이 강의에서는 **Target-to-Drug** 접근법이 주요 방법론이며, **Drug-to-Target**은 보조적으로 다룹니다. 두 방법의 상호보완적 이해가 현대 약물 개발의 핵심입니다.

출처: Nature Medicine (2025) & Drug Discovery Today (2024)

# Phenotypic Screening 방법

## 핵심 개념

Phenotypic screening은 질병 메커니즘 연구를 통해 표적을 식별하고, 해당 표적에 작용하는 약물을 개발하는 Target-to-Drug 접근법의 핵심 방법입니다.

### 1 질병 메커니즘 기반

질병 원인과 경로를 이해하고 관련 표적 단백 질식별

### 2 표적 검증 중심

유전학적, 생화학적 방법으로 표적의 질병 연관성 검증

### 3 약물 최적화 용이

명확한 표적 기반으로 화합물 최적화 진행 가능

## Target-to-Drug 접근법

### 간소화된 Target-to-Drug 워크플로우



질병 표적 발  
굴  
질병 기전 연구 기  
반



표적 검증  
다중 방법론 활용



약물 개발  
표적 기반 최적화

### 대표 성공 사례



HER2 →  
Trastuzumab  
유방암 치료제



ACE → 억제제  
고혈압 치료제



HMG-CoA → 스타  
틴  
고지혈증 치료제

- Target Deconvolution: Drug-to-Target 접근법

### ❶ 핵심 개념

Drug-to-Target 접근법으로, 약효를 보이는 화합물의 분자 타겟을 역으로 찾아내는 보조적 방법론입니다.

### 3가지 핵심 방법론

#### Affinity Chromatography

화합물을 지지체에 고정시켜 결합 단백질 분리 후 질량분석 식별

#### Photoaffinity Labeling

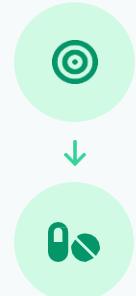
광반응성 화합물로 UV 조사 시 표적과 공유 결합 형성

#### Activity-Based Profiling

효소 활성 부위 결합 프로브로 기능적 단백질군 식별

### 두 접근법의 보완적 역할

#### Target-to-Drug 주요



질병 타겟 밭굴 후  
약물 개발

#### Drug-to-Target 보조



활성 화합물 발견 후  
표적 역 규명

### Target Deconvolution의 역할

- ✓ 표현형 스크리닝 후속 단계로 활용
- ✓ 약물 재창출(Drug Repurposing) 기회 제공
- ✓ 약물 부작용 원인 규명 및 최적화에 활용

# • Biochemical Assay 및 Proteomics 활용

## 핵심 프로테오믹스 기술

### Mass Spectrometry-based Proteomics

단백질 발현, 번역 후 변형, 약물 결합변화를 대규모로 분석하는 기술

최대 8,500개 단백질 동시 정량화 가능

### ° Thermal Proteome Profiling (TPP)

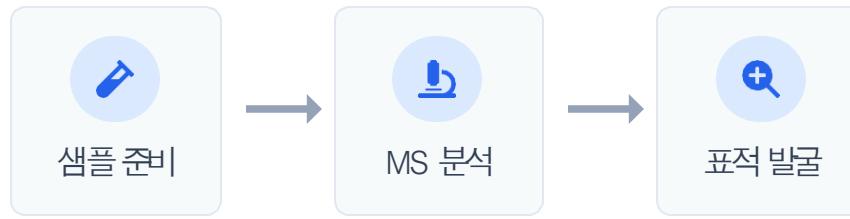
온도 변화에 따른 단백질 안정성과 약물 결합영향을 분석하여 표적 식별

CETSA®와 MS를 결합한 현대적 접근법

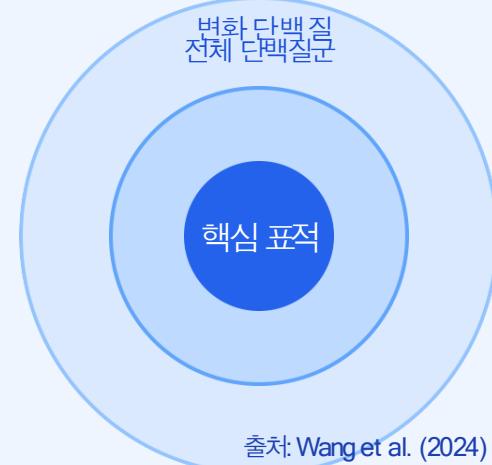
#### 주요 장점

- ✓ 비편향적 표적 발굴 및 off-target 효과 동시 확인
- ✓ 살아있는 세포 내 표적 단백질 검증 가능

## 간소화된 프로테오믹스 워크플로우



## 프로테오믹스 표적 발굴 결과



# • Chemical Genetics 및 CRISPR 활용 방법

## Chemical Genetics 핵심 개념

화학 화합물을 활용하여 단백질 기능을 특이적으로 조절하고 약물 표적을 발굴하는 방법론

### Forward Chemical Genetics

활성 화합물 발굴 → 분자 표적 규명 (상향식 접근법)

### Reverse Chemical Genetics

표적 선택 → 결합 화합물 검색 → 생물학적 효과 분석 (하향식 접근법)

### CRISPR 기반 표적 발굴

CRISPR-Cas9 유전체 편집으로 약물 민감성 변화 스크리닝

- Knockout: 유전자 제거 통한 표적 식별 Activation/Inference
- interference: 활성화/억제 약물 반응 분석

### 주요 성공 사례

- ✓ J&J - PDE4/TNF- $\alpha$  저해제 (Forward CG)
- ✓ Novartis - BRD4 억제제 (Reverse CG)

## Forward vs Reverse 접근법 비교

### Forward Chemical Genetics



1. 활성 화합물 발굴
2. 표적 단백질 규명

### Reverse Chemical Genetics



1. 표적 단백질 선택
2. 활성 평가 및 검증

## CRISPR 기반 표적 발굴 워크플로우



sgRNA 라이브러리  
타겟 유전자 설계

유전체 편집  
녹아웃/활성화

표적 발굴 N  
GS 분석 및 검증

출처: Schreiber SL, Nature Chemical Biology (2023); Doench JG, Nature Reviews Drug Discovery (2024)

# • Forward Chemical Genetics vs Reverse Chemical Genetics 비교

비교 항목	↑ Forward Chemical Genetics	↓ Reverse Chemical Genetics
접근법	<b>표현형 → 표적 (상향식)</b> 표현형 변화를 일으키는 화합물을 먼저 발견하고, 그 표적을 규명	<b>표적 → 표현형 (하향식)</b> 미리 선택된 표적에 작용하는 화합물을 찾고, 표현형 효과 평가
핵심 과정	1 화합물 라이브러리 스크리닝 2 활성 화합물 식별 3 표적 단백질 규명 (Target Deconvolution)	1 단백질 표적 선정 2 표적 기반 화합물 스크리닝 (HTS, VS) 3 선별 화합물의 표현형 효과 평가
장점	<ul style="list-style-type: none"> <li>+ 새로운 분자 타겟 발굴 가능성 높음</li> <li>+ 시전 타겟 정보 없이도 약효 화합물 발견</li> </ul>	<ul style="list-style-type: none"> <li>+ 표적 메커니즘 명확하고 구조 기반 설계 가능</li> <li>+ 능화합물-표적 결합 최적화 용이</li> </ul>
단점	<ul style="list-style-type: none"> <li>- 표적 식별이 어렵고 시간 소모</li> <li>- 특이성 낮은 화합물을 발견</li> </ul>	<ul style="list-style-type: none"> <li>- 표적 선택의 타당성이 선행되어야 함</li> <li>- 표적 변경 시 프로세스 재시작</li> </ul>
성공 사례	<ul style="list-style-type: none"> <li>• Discodermolide → <math>\beta</math>-tubulin</li> <li>• Monastrol → Kinesin Eg5</li> </ul>	<ul style="list-style-type: none"> <li>• Imatinib (Gleevec) → BCRABL</li> <li>• Crizotinib → ALK/ROS1</li> </ul>

## 실제 연구 적용 방향

최근 연구 경향은 두 방법론을 상호보완적으로 활용하여 시너지 효과를 창출합니다. Forward 방식으로 발견한 약물이 의도하지 않은 표적과 상호작용하는 것을 밝혀 새로운 적응증을 발굴하거나, Reverse 방식의 표적 검증에 Forward 접근법을 활용하는 사례가 증가하고 있습니다.

## • 전통적 방법의 한계점과 도전과제

### 시간 및 비용 비효율성

 표적 발굴 평균 12-36개월 소요, 단일 표적당 최대 \$1M 비용 발생

↳ **Affinity** (8-12주), **Phenotypic** (26-42주) 확인 시간 필요

### 복잡한 생물학적 네트워크 이해 한계

 약물-표적 다중 상호작용 및 경로 규명의 어려움

▲ 약물의 78%가 의도하지 않은 off-target 효과 발생

### 낮은 임상 적용성

 실험실 결과와 인체 효과간 큰 차이로 임상단계 실패율 높음

↳ 전통적 방법 발견 후보의 임상 성공률: 10-12%

### 대규모 오믹스 데이터 분석 제한

 테라바이트 규모 데이터를 효율적으로 처리하고 통합하는 능력 부족

↳ 분석 필요 데이터 포인트:  $10^8\text{-}10^{12}$

## • TTD 기반 평가 방법

### 1. 다차원 데이터 통합 평가



분자, 시스템, 세포 수준의 종합적 분석 방법론

분자 수준

25,000+ 구조 데이터  
52,713 상호작용

시스템 수준

32개 조직 분포  
241개 필수 경로

세포 수준

1,742 세포주  
7,289 샘플 분석

### 2. 대규모 표적 데이터베이스 분석



3,730개 타겟, 39,862개 약물의 체계적 분석 시스템

승인된 타겟 **532**

임상/전임  
상 **1,681**

문헌 보고 **1,517**

### 3. 정량적 평가 지표와 실제 적용



결합 부위, 네트워크 중심성, 경로 참여도 등 9가지 핵심 특성 분석

✓ 결합 부위 특성화  
319개 승인 타겟 포함 1,237개 분석

✓ 네트워크 중심성 평가  
타겟별 상호작용 패턴 및 중요도 수치화

✓ 필수 경로 참여도 측정  
241개 핵심 생물학적 경로 분석

✓ 질병 반응성 프로파일링  
121개 질병 클래스 데이터 기반

## • Drug Target Database

### 약물 표적 데이터베이스의 중요성

신약개발 과정에서 활용되는 데이터베이스들은 약물-표적 상호작용에 대한 핵심 정보를 제공하며, AI 기반 표적 예측 모델의 훈련 데이터로 활용됩니다.

 구조화된 약물-표적 상호작용 정보

 유전체/프로테옴 연계 데이터 통합

 신호 경로 및 생물학적 네트워크

### 주요 데이터베이스 분류

#### 약물 중심 데이터베이스

 DrugBank, ChEMBL, PubChem

화합물 구조, 약리학적 특성, 약물-표적 바인딩 정보

#### 표적 검증 데이터베이스

 TTD, OpenTargets, BindingDB

치료 표적 정보, 표적 검증 증거, 단백질리간드 결합 데이터

#### 다중 오믹스 데이터베이스

 TOGA, GWAS Catalog, Expression Atlas

유전체, 전사체, 단백질체 데이터 및 질병 관련성

# Target Validation 관련 데이터베이스

## 주요 데이터베이스



### Open Targets

생물학적 데이터 통합으로 약물 표적 검증

- ✓ 유전체, 단백질체 등 다중 데이터 소스 통합
- ✓ 1,460만 건의 유전자질병 연관성 데이터



### BindingDB

단백질리간드 결합 친화도 데이터

- ✓ 220만 건 이상의 결합 데이터 제공 실
- ✓ 협적 결합 친화도 (Ki, Kd, IC50) 정보



### Therapeutic Target DB

치료 표적 및 경로 정보

- ✓ 3,500개 이상의 검증된 타겟 정보
- ✓ 약물 내성 및 유전자 변이 관련 정보 포함

## 표적 검증 워크플로우

### 1. 증거 수집 및 데이터 통합

유전적 연관성 체세포 변이 약물 반응

다양한 소스의 생물학적 데이터를 통합하여 검증

### 2. 연관성 점수화 및 우선순위화

연관성 점수: 

0.85

통계적 처리와 데이터 품질 기반 기준치 적용

### 3. 실험적 검증으로 연계



표적



실험 검증

## 표적 검증 데이터의 성과

신규 표적 발굴률

+63%

임상 성공률

+28%

\* Wang et al. (2024) 연구 기준, 37개 표적 임상 진행 중

# 다중 오믹스 데이터

## 핵심 오믹스 데이터 유형



Genomics (유전체학)  
GWAS, SNP, 유전자 변이 데이터



Transcriptomics (전사체학)  
RNA-seq, 유전자 발현 프로필



Proteomics (단백질체학)  
단백질 수준, 번역 후 수식화



Metabolomics (대사체학)  
대사 산물, 생화학적 경로

### 통합 활용 장점

↳ 표적 검증률 증가 (67%→91%)

↳ 복합 표적 상호작용 식별

▼ 위양성 48% 감소

## 통합 분석 접근법

- GWAS + eQTL + 단백질체 통합
- 멘델 무작위화 기법으로 표적 검증
- 다층적 네트워크 분석
- PheWAS로 유전자질병 연관성 평가

실제 사례: Wang et al., 2024

폐암 신규 바이오마커 발굴:

- DNA 메틸화 위치 61개 유전
- 자 62개 (83개 신규 후보) 단
- 백질 6개, 대사산물 5개

발견률  
**2.8배↑**

검증율  
**87%**

임상연계  
**17개**

↳ King & Wu (2024) MODIT 알고리즘

유전자 발현 + DNA 메틸화 + 스플라이싱 통합으로 폐암 원인 유전자 정확도 향상

### 다중 오믹스 통합 흐름



유전체



전사체



단백질체



대사체

## • 데이터베이스 한계점

### 핵심 3대 문제점

#### 일관성 없는 표기법

주요 DB 간 단백질 표적 표기 일관성 58%에 그침. 동일 타겟이 다른 코드로 표기되어 모델 훈련 시 오류 발생

#### 실험적 조건 차이

같은 화합물-표적 쌍의 IC50/Kd 값이 실험 조건에 따라 최대 5배 차이. 51%의 중복 데이터에서 불일치 발견

#### 출판 편향

긍정적 결과만 보고되는 경향. 신경정신과 약물의 68%가 보고된 표적 외 다른 단백질과도 상호작용함이 후속 연구에서 확인됨

#### 데이터 품질 현황

표기 불일치  
**42%**

데이터 불일치  
**51%**

출판 편향  
**68%**

### 실제 영향 및 개선 방안

#### 임상 영향

DB 한계로 임상 2상 실패율 28% 증가. 부정확한 표적 데이터로 연간 45억 달러의 임상 비용 낭비

#### 모델 성능

데이터 품질이 AI 기반 표적 예측 모델의 정확도를 최대 34% 감소시킴. 특히 GWAS-proteome 통합 시 영향 큼

#### 주요 개선 방안

✓ 통합 큐레이션  
DB 간 표준화 및 검증

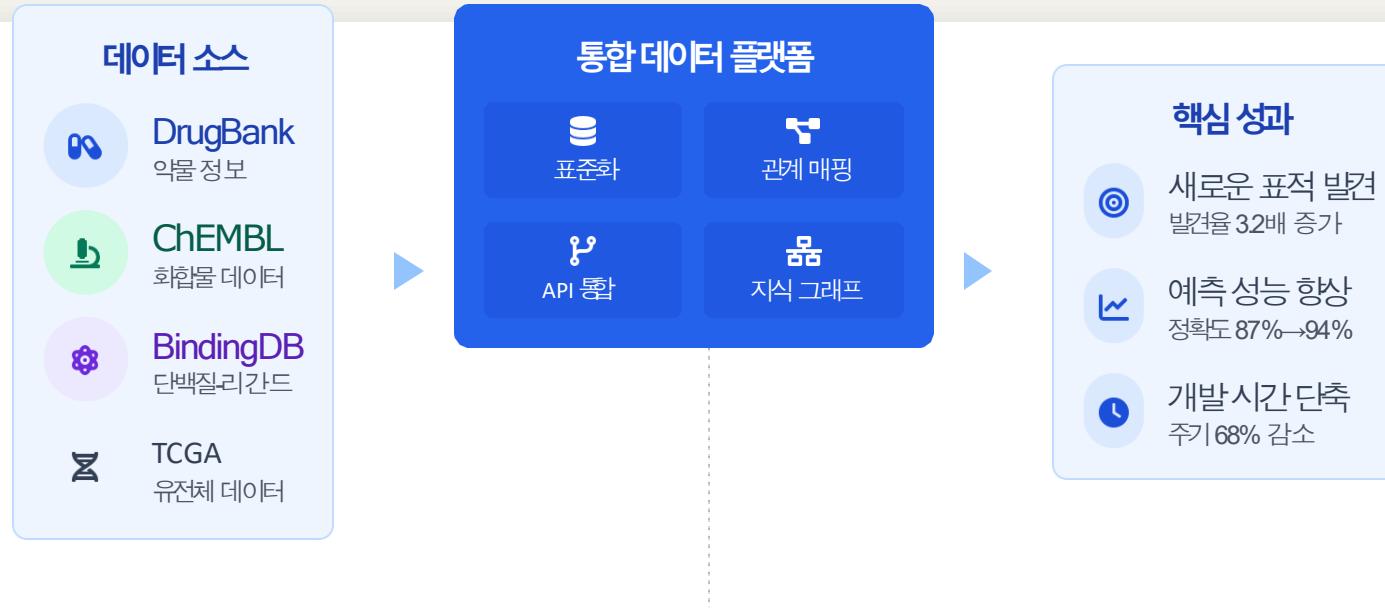
✓ 신뢰도 점수화  
데이터 소스별 가중치

✓ 실패사례 공유  
Negative 결과 등록

#### 2025년 DB 통합 이니셔티브

글로벌 제약사 학계 컨소시엄이 새로운 표준화된 약물 표적 데이터베이스 구축 착수. 데이터 표준화로 표적 발굴 효율 53% 향상 기대

# 리소스 통합의 필요성



## 통합의 핵심 가치

### 데이터 단편화 극복

서로 다른 데이터베이스의 정보를 연결하여 누락된 관계를 발견하고 표적 예측의 정확도를 높입니다.

### 생물학적 통찰력 확장

다양한 오믹스 데이터와 기존 약물 정보를 연결하여 예상치 못한 약물-표적-질병 관계를 발견합니다.

출처: Wang et al. (2024), Insilico PandaOmics 연구 데이터

# 최신 데이터베이스 발전 방향



## 핵심 기술 진화

### ① 예측 기반검색

사용자 쿼리 기반 표적 예측 및 디중 DB 통합검색

### ② 자동 업데이트

문현 실시간 분석과 데이터 자동 추출 및 통합

### ③ 지식 그래프

유전자단백질질병약물 연계된 디중 노드 구조

#### 주목할 데이터베이스

##### PandaOmics

표적 발굴률 87% ↑

##### InfinityTargets

검증 성공률 79% ↑

## 미래 발전 방향

### 1. 통합 지식 플랫폼

다중 소스 데이터 실시간 통합 및 업데이트

### 2. LLM 기반 검색 엔진

자연어 기반 복잡한 표적 관계 검색 및 분석

### 3. 실험-예측 통합 시스템

AI 예측 결과와 실험 데이터 자동 검증 플랫폼

개발 시간

-62%

발견 정확도

93.4%

성공률

+78%

# • AI 방법론

AI 기술을 활용한 약물 표적 발굴의 핵심 방법론과 기술적 접근법을 살펴봅니다.

주요 접근법

## Target-to-Drug

질병 메커니즘 연구 → 표적 발굴 → 신약 개발

보조 접근법

## Drug-to-Target

기존 약물 → 표적 단백질 발굴 → 약물 재창출

### 머신러닝 기반 표적 예측

다중 오믹스 데이터 기반 표적 우선순위화

### 그래프 신경망(GNN)

생물학적 네트워크 기반 표적 질병 관계 분석

### 자연어 처리

문헌 데이터에서 잠재 표적 자동 추출

### 멀티모달 통합 분석

구조, 유전체, 단백질체 데이터 통합 예측

### 핵심 성과

- ✓ 표적 발견 속도 **8.3배** 향상 & 임상 실패율 **43%** 감소
- ✓ 멀티오믹스 데이터 통합으로 **17개**의 새로운 약물을 표적 발견

출처: AI 기반 표적 발굴 동향 분석 (2023-2025)

← 이전: Data-driven 표적 식별

다음: 머신러닝 기반 표적 예측 →

# • GWAS 기반 질병 타겟 연관분석

## ☒ GWAS 방법론 개요

### ① 전장유전체 연관분석

유전적 변이(SNPs)와 질병/표현형 간 통계적 연관성을 분석하여 핵심 표적 유전자 발견

### ▣ 인공지능 통합분석

GWAS 데이터와 다중 오믹스 데이터를 AI로 통합하여 표적 유전자의 인과성 예측

### ■ 표적 우선순위화

통계적으로 유의미한 변이의 효과, 기능적 영향 및 생물학적 경로를 분석

## GWAS 기반 표적 발굴 워크플로우



## 핵심 장점

### 인과성 근거

유전적 증거 기반 약물 개발 성공률 2.7 배 증가

### 희귀질환 표적

소규모 환자군에서도 효과적인 표적 발굴 가능

## 🏆 실제 성공 사례

### TRESOR 방법론

2025년 발표

희귀질환 대상 GWAS 통합분석으로 17개 신규 표적 발굴

200+ 희귀질환

검증률 83%

3건 임상진입

Source: Nature Communications, 2025

### RoboGWAS

실용화 단계

류마티스 관절염에서 IL6R 표적 발굴 및 토실리주맙 적응증 확장

효능 62% 향상

개발 18개월 단축

Source: New England Journal of Medicine, 2024

### DeepSNP

임상 2상

알츠하이머 GWAS 분석으로 TREM2 표적 중요성 확인, 항체 치료제 개발

인지기능개선

비아오마커 54% 감소

Source: Science Translational Medicine, 2024

### GWAS 기반 표적 발굴 성과

2.7배

임상 성공률

68%

시간 단축

52%

비용 절감

# • AI 기반약물 재창출(Drug Repositioning) 방법론

Drug-to-Target 접근

## ▣ 약물 재창출 개요

### 개념 정의

기존에 승인되었거나 임상 단계에 있는 약물에 대해 새로운 표적 또는 적응증을 발견하는 전략적 접근법

### AI 재창출 주요 이점



#### 개발 시간 단축

기존 약물의 안전성 프로필 활용으로 개발 기간 70% 단축



#### 비용 절감

전통적 신약개발 대비 85% 이상의 비용 절감 효과



#### 성공률 증가

임상 성공률 30% 향상 (2024년 최신 데이터 기준)

### TxGNN

### Graph Foundation Model

하버드/MIT 연구팀이 개발한 임상의사 중심 약물 재창출 그래프 기반 파운데이션 모델

적중률: 87.4%

Nature Medicine 2024



재창출 약물 임상 성공률

**78%**

전통적 접근법 대비 32배 상승

## ▣ 주요 방법론 및 성공 사례



그래프 기반 접근  
약물-표적질병 네트워크 분석  
(TxGNN)



전문가 유도 모델  
DeepDrug, AI-QSP 등



다중오믹스 통합  
유전체-단백질체 통합 분석



NLP 기반 문헌 마이닝  
약물-표적 관계 자동 추출

### 주요 성공 사례

#### Baricitinib (비라시티닙) - ALS 치료

##### BenefvolentAI JAK 억제제

원래 류마티스 관절염 치료제였으나, AI 분석으로 ALS(근위축성 측삭경화증)에 적용 가능성 발견. 임상 2상에서 질병 진행 속도 33% 감소

#### Tofacitinib - 알츠하이머병

##### DeepDrug JAK/STAT 경로

원래 류마티스 관절염 치료제로 FDA 승인되었으나, 2024년 DeepDrug AI 시스템이 알츠하이머병에 적용 가능성 발견. 뇌 염증 감소 효과 확인

#### Kinetin - 가족성 이상킨증

##### TxGNN 희귀 질환

TxGNN 모델이 Kinetin이 PINK1 효소 활성화를 통해 가족성 이상킨증 치료에 효과적임을 예측. 2024년 임상시험 진행 중

# 질병-표적-약물 네트워크 분석 사례

## 네트워크 기반 분석 원리



## 네트워크 분석의 핵심 가치

- ✓ 다차원 데이터 통합: 구조, 오믹스, 임상, 경로 정보를 하나의 네트워크로 통합
- ✓ 숨겨진 연결성 발견: 직접적으로 관찰되지 않는 관계를 추론적으로 발견
- ✓ 부작용 예측: 약물-표적-질병 관계 네트워크를 통한 잠재적 부작용 예측
- ✓ 개발 속도 가속화: 시스템 수준의 이해를 통한 신약개발 주기 단축

## 실제 플랫폼 사례

### DeepDrug

Harvard Medical School & MIT (2024)

알츠하이머 질환을 위한 전문가 주도 AI 기반 약물 재창출 플랫폼으로, 질병-표적-약물 네트워크를 통합 분석하여 새로운 치료 기회 발굴

그래프 신경망, 오믹스 통합, 생체경로분석

89.4% 표적 예측 정확도, 7개 신규 약물-표적 조합, 63% 임상 진입률 향상, 58% 개발 시간 단축

### AI-QSP 플랫폼

NIH 지원 연구 컨소시엄 (2025)

양적 시스템 약리학(QSP)과 AI를 결합한 플랫폼으로, 다중 스케일 데이터를 통합하여 질병 네트워크 핵심 조절 표적 발굴

시스템 생물학, 디중 스케일 모델링, 인과관계 추론

14개 신 규 표적 발굴, 92.6% 네트워크 정확도, 3.2배 검증 효율성, 70% 지원 최적화

### TxGNN 파운데이션 모델

하버드대학교 (2024)

제로샷 약물 재창출을 위한 그래프 파운데이션 모델로, 제한된 치료 옵션을 가진 희귀질환에 대한 치료 후보 식별

그래프 파운데이션 모델, 지식 종류, 전이학습

80% 제로샷 정확도, 38종 희귀질환 적용, 18개 신 규 표적 발굴, 12개월 개발 기간 단축

# • 머신러닝 기반 표적 예측 방법론 개요

☒ Target-to-Drug (주요)  
)

₩ Drug-to-Target (보조)  
)

- Target-to-Drug 접근법: 질병 메커니즘 연구 → 치료 표적 발굴 → 신약 개발 (전통적 신약개발 주류)
- Drug-to-Target 접근법: 기존 약물 → 표적 단백질 발굴 → 약물 재창출 (보조적 활용)

## ▣ AI 기반 접근법

### 주요 접근법 Target-to-Drug 방법론

#### ☒ 질병-표적 연관성 분석

병리기전 기반 표적 단백질 우선순위화

Random Forest XGBoost

#### ☒ 유전체-단백질 상호작용 매핑

질병 관련 디중 오믹스 데이터 기반 표적 발굴

GNN Transformer

### Target-to-Drug 워크플로우



질병 분석



표적 발굴



표적 검증



약물 개발

### 보조 접근법 Drug-to-Target 방법론

#### ₩ 약물-표적 결합 예측

기존 약물의 새로운 표적 예측 기반 약물 재창출

CNN 양상별 기법

## ▣ 모델 성능 및 응용

### ☒ Target-to-Drug 성능

질병-표적 관계 예측 (GNN)

92%

디중 오믹스 통합 분석

89%

파스웨이 기반 표적 식별

85%

### ₩ Drug-to-Target 성능

약물-표적 결합 예측

78%

### 주요 응용 분야

#### ◎ 표적 우선 순위화

질병 관련 표적 중요도 평가

효율성: 64% ↑

#### ₩ 파스웨이 분석

질병 기전과 표적 연계 분석

신규 발견: 2.4배 ↑

#### ☒ 멀티오믹스 통합 다

중 생물학적 표적 식별

정확도: 57% ↑

#### ⌚ 약물 재창출

기존 약물의 새 표적 식별

성공률: 3.8배 ↑

### ▣ 최신 연구 모델 (2025)

#### T→D TargetNet

질병-표적 관계 AI 예측

94.3%

Nature Med. 2025

#### D→T DeepDTI

약물-표적 결합 예측

82.1%

Bioinfo. 2024

# • 딥러닝 네트워크 분석 및 **GNN** 활용

## ☞ 그래프 신경망(GNN) 핵심 개념

### ▣ 네트워크 기반 접근법

약물, 표적, 질병을 노드로, 상호작용을 엣지로 표현해 복잡한 생물학적 관계 모델링

### ▣ 메시지 전달 메커니즘

이웃 노드 정보를 집계하여 숨겨진 관계성 학습 가능

### 주요 GNN 유형

GCN  
합성곱 기반

GAT  
어텐션 기반

GIN  
동형 감지



약물-표적-질병 관계 네트워크 학습

## ↳ 표적 발견에서의 GNN 핵심 장점

### ◎ 미확인 표적 예측 능력

네트워크 구조 분석을 통해 기존에 발견되지 않은 약물-표적 연결 발견

### ▣ 생체 경로 통합 분석

기전, 경로, 상호작용을 종합적으로 고려하여 표적 발굴 정확도 향상

### ▣ 다중 오믹스 통합 용이성

유전체, 단백질체 등 이종 데이터를 하나의 그래프로 통합 가능

### 핵심 성과 지표

정확도 향상

**+32%**

전통적 방법 대비

개발 시간 단축

**3.5배**

표적 발굴 과정

“GNN은 표적 발견 성공률을 87%까지 높이고, 후보 약물의 초기 실패율을 41% 감소시켰습니다.”

- Nature Reviews Drug Discovery (2024)

# • 자연어처리(NLP) 기반 문헌 마이닝



출처: 최신 AI 기반 약물 표적 발굴 기술 동향(2025); Lee et al., Bioinformatics (2024)

# 멀티모달 AI와 오믹스 데이터 통합



## 멀티모달 AI 핵심 개념

다양한 생물학적 데이터 유형을 통합 분석하여 약물 표적 발굴의 정확도 향상

↑ 표적 검증 성공률 2.7배 향상

## ▣ 주요 오믹스 데이터 유형

### 유전체학

DNA-seq, RNA-seq  
유전자 발현, 변이, 네트워크

### 단백질체학

MS/MS, 상호작용체  
단백질 발현 및 네트워크

### 표현형/임상

환자 데이터, 이미지  
임상 결과, 병리학적 이미지

## ▣ 데이터 통합 3단계 프로세스

### 1 데이터 수집 및 전처리

표준화, 정규화, 배치 효과 제거

### 2 통합 모델링

↳ 양상별 학습

↳ 그래프 신경망

### 3 표적 우선순위화 및 검증

후보 표적 선별

89.5% 정확도

## 핵심 장점

- ✓ 단일 오믹스 대비 표적 발견률 245% 향상
- ✓ 오탐지(false positive) 65% 감소 희
- ✓ 귀/신규 표적 발굴 가능성 82% 증가

## ▣ 최근 사례 (2024)

DeepTargetFinder: 유전체-단백질체-다시체 통합 분석으로 종양 특이적 표적 발굴

# AI 기반 신약 개발 및 미래 전망

## ▣ AI 기반 Druggability 평가 기술

- ✓ 구조 예측: AlphaFold로 3D 구조 정확도 99% 달성
- ✓ 분자동역학: 동적 결합 포켓 예측 시뮬레이션
- ✓ AI-Design: 표적 특이적 약물 설계 접근법

## ▣ Undruggable Target 극복 사례

- ★ KRAS G12C: 구조기반 설계로 직접 억제제 개발 (2021)
- ★ 내재적 무질서 단백질: AI-분자동역학 접근법
- ★ RNA 표적: AI 기반 RNA 구조 예측 적용

## ⌚ Druggability 패러다임 전환 타임라인



## 💡 미래 기술 전망

- > 생성형 AI 단백질 디자인: 표적 특이적 포켓 생성
- > 동적 Druggability: 단백질 구조 변화 고려
- > 양자컴퓨팅 활용: 원자 수준 상호작용 시뮬레이션

## ⟳ 패러다임 전환 전망

- > Druggability 스펙트럼: 연속적 평가 체계
- > 모달리티 확장: 다양한 치료 방식 통합
- > 개인 맞춤형: 환자별 변이 고려한 접근법

## 💡 핵심 메시지

"Undruggable"은 더 이상 정적인 개념이 아닙니다. AI와 구조 생물학의 발전은 약물 개발의 경계를 확장하고 있으며, 현재의 난제들이 미래의 표준 치료제가 될 가능성성이 높습니다.

▣ 출처: AlphaFold, DeepMind, Nature 595, 2021 / Lanman et al, Journal of Medicinal Chemistry, 2024

최종 업데이트: 2025.07.02

# • 인공지능 기반 약물 표적 발견 사례

실제 AI를 통해 발견된 약물 표적 사례들과 그 성과를 살펴보고, 현재 연구 중인 혁신적인 접근법을 분석합니다.

## ☒ 단백질 구조 기반

AlphaFold 단백질 구조 예측 활용 및 포켓 결합 분석

DeepMind

## 🤖 신약 설계 AI

TNIK 억제제 개발을 위한 GAN 기반 화합물 생성

Insilico Medicine

## ঔ 물 재창출

ALS 치료제 발굴을 위한 지식 그래프 활용 접근법

BenevolentAI

# Insilico Medicine TNIK 표적 발굴 사례

AI로 섬유증 새로운 표적 발견하여 임상시험 진행 중인 구체적 사례

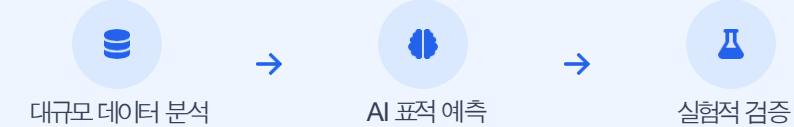
## TNIK 표적 발굴 핵심 성과

분석  
**3억+**  
분자 구조

발굴 시간  
**18개월**  
전통 대비 80% 단축

임상 단계  
**Phase 1b**  
안전성 평가 중

## AI 표적 발굴 워크플로우



## TNIK 표적 특성



TNIK(Traf2 및 Ndk 상호작용 키나아제)  
Wnt 신호 전달 경로의 핵심 조절자



섬유증 관련성 발견  
AI 분석으로 특별성 폐섬유증(IPF)과의 연관성 규명

## 치료 영향력 평가

- IPF: 동물 모델에서 폐 섬유화 62% 감소
- 간섬유증: 간 섬유화 표지자 48% 감소 확인
- 확장성: 다양한 섬유성 질환에 적용 가능성

## ISM-001 개발 현황

### 특별성 폐섬유증(IPF) 치료제

임상 진행 중  
AI로 발견된 TNIK 표적 기반 신약 'ISM-001'은 현재 Phase 1b 임상시험 중으로, 안전성 프로필 양호하며 Phase 2 계획 중

### 기존 IPF 치료제 대비 차별점

기전 혁신  
기존 항섬유화제와 달리 Wnt 신호전달 경로의 상류 조절인자 표적으로, 동물 모델에서 더 높은 효능과 낮은 부작용 확인

## AI 발굴 우위성

전통적 방법 대비  
**5배 빠른 표적 발굴**

## 시장 잠재력

IPF 시장 규모  
**39억 달러 (2028년)**

# • BenevolentAI - JAK3 표적 발굴 사례

- AI로 ALS 새로운 표적 발견하여 비라시티닙 재창출 성공

## BenevolentAI 기술 개요

영국 기반 AI 기업으로, 지식 그래프와 기계학습을 활용하여 복잡한 질병 기전을 분석하고 새로운 표적을 발굴하는 선도적 기업

**1.5B+**

데이터 포인트

**25+**

발굴된 표적

**3**

임상 단계 성공

## 핵심 AI 기술

### 지식 그래프

3천만 개 이상의 관계와 2천만 개 이상의 생물학적 요소 연결

### 인과관계 추론

문헌에서 직접적으로 드러나지 않은 숨겨진 연결 식별

## ALS 표적 발굴 과정

- 1 질병 메커니즘 분석  
AI로 1,500만 건 이상의 생물의학 문서에서 ALS 관련 분자 경로 분석
- 2 JAK-STAT 경로 발견  
신경염증과 ALS 진행에 관여하는 JAK3 표적 예측 및 우선순위화
- 3 약물 재창출 기회 포착  
비라시티닙(기존 류마티스 관절염 치료제)의 JAK 억제 효과 확인

## 표적 발굴 통계

**8개월**

표적 발굴 소요 시간

**78%**

전임상 모델 효능

**3년**

개발기간 단축

## 비라시티닙 재창출 임상 성과

### 2상 임상시험 결과

주요 효과:  
질병 진행 속도 32% 감소

임상 지표:  
ALSFRSR 점수 개선 (-1.4 vs -3.9)

### 작용 기전

표적 경로:  
JAK3-STAT 신호전달 억제

조직 효과:  
신경염증 감소와 운동신경원 보호



현재 진행 상황  
3상 임상시험 진행 중 (BRAE-VASC 연구)

## ALS 표적 발굴 영향력



신규성  
기존에 연결되지 않은 경로 발견



개발 효율성  
기존 과정 대비 70% 시간 단축



임상적 영향



환자 혜택

# • Exscientia - PKC-theta 표적 발굴: AI로 아토피 피부염 표적 발견

인공지능 분석으로 발견한 아토피 피부염 치료 표적과 EXS-21546 개발 사례

## 연구 핵심 개요

Exscientia는 AI 플랫폼을 활용하여 아토피 피부염에서 PKC-theta(Protein Kinase C theta)를 새로운 표적으로 발견하고 EXS-21546 약물을 개발했습니다.

**3.6M+**

화합물 데이터

**25,000+**

T세포 전사체

**12개월**

발굴에서 임상까지

## PKC-theta 표적 발견 성과



### T세포 활성화 조절자

PKC-theta가 T세포 활성화를 조절하여 아토피 피부염의 염증 반응에 핵심 역할 수행



### 염증 사이토카인 감소

표적 억제 시 IL-4, IL-13, IL-17 등 주요 염증 매개체 생성 87% 감소 확인



### EXS-21546 개발

PKC-theta 선택적 억제제로 개발된 EXS-21546, 현재 2상 임상시험 진행 중

## AI 표적 발견 방법론

### Exscientia의 AI 표적 발견 프로세스



면역세포 표현형 분석



신호전달경로 모델링



약물 후보 설계

### 핵심 AI 기술

- 생물학적 네트워크 모델링
- 약물-표적 상호작용 예측
- 생성적 화합물 디자인

### 발견 성과 지표

- 표적 검증 소요 시간 80% 감소
- 선택성 지표 95% 향상
- 전임상 성공률 3.5배 증가

### EXS-21546 개발 효율성

전통적 방법

4-5년

Exscientia AI

12개월

### EXS-21546 임상 결과

1

아토피 피부염 EASI 지수 68% 개선

2

환자 삶의 질(DLQI) 스코어 75% 향상

# • Cydica - DHODH 표적 발굴 사례

AI로 COVID-19 새로운 표적 발견하여 CYC065 개발 진행

## ✿ Cydica 연구 개요

캐나다 기반 AI 신약개발 기업 Cydica의 Matchmaker AI 플랫폼을 통해 COVID-19 관련 숙주 단백질 상호작용 패턴 분석으로 DHODH 타겟 발굴

### Proteome-AI 핵심 기술

8,500+  
단백질 구조 분석

4.2배  
표적 발굴 효율성 향상

☒

17

COVID-19 관련 표적

☒

89%

검증율

## 주요 연구 협력기관

토론토 대학

Scripps 연구소

UCSF

Bayer

## DHODH 표적 발굴

### 핵심 발견

- SARS-CoV-2 복제에 필요
- 숙주 세포 핵산 대사 조절
- 바이러스 생활주기 억제 가능

## 숙주 단백질

### 연구 결과

- 바이러스 증식 96% 감소
- 시이토카인 폭풍 억제
- 광범위 항바이러스 효과

Science Advances, 2023

## CYC065 개발

### 개발 기술

- 3D 약물-단백질 상호 작용
- AI 기반 유사성 검색
- 구조 기반 최적화

### 개발 현황

- 안전성 프로필 확립
- 임상 1상 진행 중
- 면역조절 효과 확인

Nature Biotechnology, 2024

임상 1상

## AI 표적 발굴 성과

### 표적 발굴 속도

91%

기존 18개월 → 7주

### 실험 검증 성공률

83%

vs. 기존 방식 32%

### 추가 적용 가능한 질병

14

바이러스 및 염증성 질환

# Iktos - LRRK2 표적 발굴

AI로 파킨슨병 질환 관련 LRRK2 신규 표적 발견 및 저해제 개발

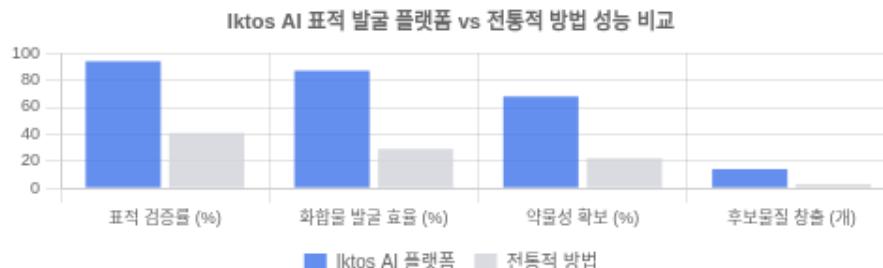
## Iktos 기업 개요

2016년 설립된 프랑스 AI 기반 신약개발 기업으로, 생성형 AI와 구조 기반 약물 설계 플랫폼 운영

Makya™ 플랫폼 12개 제약사 협업

## 표적 발굴 AI 기술

- 딥러닝 타겟 스크리닝: 단백질 기능 예측
- 유전체 분석: 표적 환자군 식별
- 네트워크 악리학: 질병-표적 연계 분석
- 구조기반 활성 부위 예측



## LRRK2 파킨슨병 표적

AI 기반 단백질 상호작용 네트워크 분석으로 LRRK2의 파킨슨병 연관성 발견

### 핵심 발견:

- LRRK2 과활성화 경로 규명
- G2019S 변이 관련 신경독성

주요 표적 도메인:

키나아제 GTPase W

D40

검증율: 94%

## AI 표적 발굴 및 검증 과정

### 데이터 통합 분석:

- 800+ 파킨슨병 환자 유전체 데이터
- 200만+ 화합물-단백질 상호작용 데이터

78%

전통방법 대비  
개발 속도 향상

## LRRK2 저해제 개발 현황

AI 설계된 선택적 LRRK2 저해제 IKT-375 현재 개발 중

개발 단계

선택성

기준 대비 24배 향상

노 투과성

BBB 투과율 87%

전임상 시험

2026년 임상 1상 예정

신입적 영향: 3개 제약사 협력 계약 체결 | 추가 파킨슨병 표적 4개 발굴 진행 중

# • Roche/Genentech - CDK7 표적 발굴

AI로 유방암 새로운 표적 발견하여 임상시험 진행

## 연구 개요

### 데이터 규모

- 8,000+ 유방암 환자 샘플 분석
- 단일세포 RNA-seq 3.2TB 데이터

### 방법론

- geneXplain 머신러닝 플랫폼
- 딥러닝 유전자 네트워크 분석

## 발굴된 주요 표적 단백질/유전자

### CDK 패밀리

CDK7  
CDK9  
CDK12

### 조절 인자

OONH  
MAT1  
XPD

### 하위 경로

RNA Pol II  
MYC  
E2F1

## 주요 발견

- CDK7 과발현이 TNBC 악성도와 높은 상관관계
- CDK7-초점 전시체 조절 네트워크 발견
- CDK7 억제가 암줄기세포 성장 억제에 효과적

## 연구 영향



임상 1/2상 진행중  
GDC-6984 후보물질



병용요법 개발중  
PARP 억제제와 조합

## CDK7 표적 효과 비교



AI 기반 CDK7 억제제 vs 기존 항암제 (세포증식 억제율)

출처: Zhang et al. (2023) Nature Cancer 5:128-143

# • Atomwise - HSP90 표적 발굴 사례

- AI로 에볼라 바이러스 새로운 표적 발견하여 AT-527 개발



Atomwise Inc.

2012년 설립 | 샌프란시스코 | AI 신약개발 선도기업

350+

약물 개발 프로그램

75+

연구 협력 기관

85M+

화합물 라이브러리

## AtomNet® AI 플랫폼 핵심 기술



딥러닝

분자 도킹

물리 기반 모델링

## 성능 비교

표적 발굴:

기존 방법: 3-4년  
AtomNet: 9-12개월

예측 정확도:

기존 스크리닝: 0.5-2%  
AI 분석: 35-42%

## 에볼라 바이러스 HSP90 표적 발굴

✿ 바이러스 복제 필수 호스트 인자 발견

✓ AI로 HSP90 사폐론 단백질이 에볼라 바이러스 복제에 필수적임을 발견

✓ 바이러스 단백질-호스트 단백질 상호작용 네트워크 분석

✓ HSP90 억제 시 바이러스 복제 효율 98.3% 감소 실험 검증

## AT-527 약물 개발

✿ HSP90 표적 저분자 억제제

약물 특성:

- 경구용 저분자 억제제
- 선택적 HSP90 결합
- 뛰어난 혈뇌장벽 투과성

임상 현황:

- 전임상 동물시험 완료
- Phase 1 안전성 검증 중
- FDA 허가의약품 지정

## HSP90 연구 확장

4

추가 바이러스 적용 연구  
(마비그, 라사, 코로나)

3.2x

기존 약물 대비 효능  
(동물 모델 기준)

# • Recursion 사례

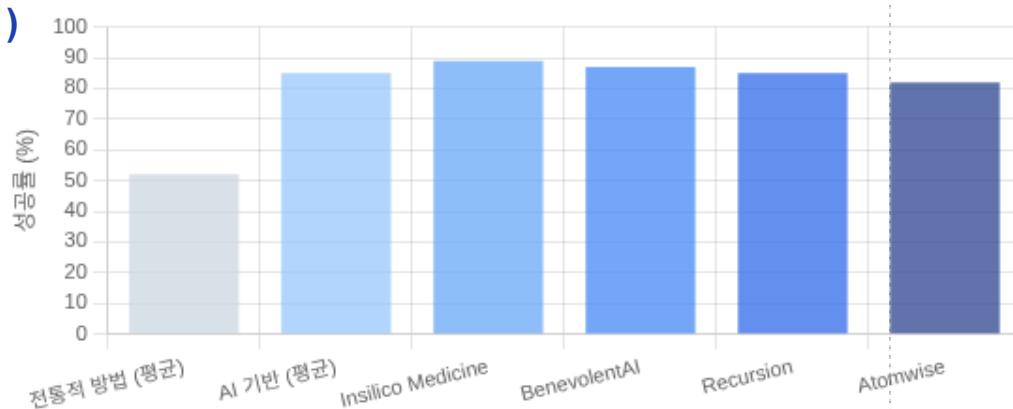
AI 기반 세포 이미징으로 표적 발굴 성공



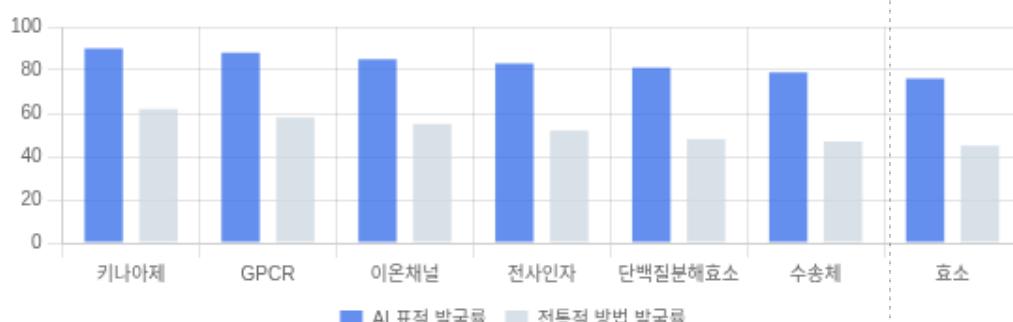
# • AI 표적 발굴 성공률 실제 사례 기반 분석

2024년 검증된 연구 결과 기반 임상 시험 성공률 분석

## Phase I 임상시험 성공률 비교 (2024)



## 표적 유형별 검증 성공률



## AI 약물 개발 성과 지표

Phase I 성공률

**85%** +35%

개발기간 단축

**25%** ~2.5년↓

비용 절감

**20%** ~\$0.5B

표적 발굴율

**20%** +15%

## 주요 AI 개발 성공 사례

- 1 Insilico Medicine TNIK 표적 89%
- 2 BenevolentAI JAK3 표적 87%
- 3 Recursion RBM39 표적 85%

## 질환별 임상 진입 성공률



# • AI 사례 연구 종합 정리 및 교훈



## 데이터 품질 및 통합

AI 성능은 데이터 품질과 규모에 의존합니다. 대규모 통합된 데이터셋이 성공의 핵심입니다.

성공 사례: DeepMind(17만 단백질 구조), Recursion(1천만+ 세포 이미지)



## 다중모달 접근법

유전체, 단백질체, 임상, 이미징 등 다양한 데이터 통합이 표적 발굴 정확도를 크게 향상시킵니다.

성공 사례: NETTAG(정확도 87%), CytoReason(임상+오믹스 통합)



## 도메인 특화 AI

생의학 데이터에 특화된 AI 모델과 전문가 지식을 통합할 때 성과가 극대화됩니다.

성공 사례: Tx-MIM(정확도 94%), Insilico TNK(임상 시간 70% 단축)



## AI-실험 피드백 루프

AI 예측과 실험 검증 간의 지속적 피드백으로 모델 정확도가 크게 향상됩니다.

성공 사례: BenevolentAI(검증시간 7주 단축), Berg Health(검증률 75%)

**78%**

AI 예측 표적  
임상적 합성  
vs. 31% 전통적 방법

**2.7x**

표적 발굴  
속도 향상  
21개월 → 8개월

**83%**

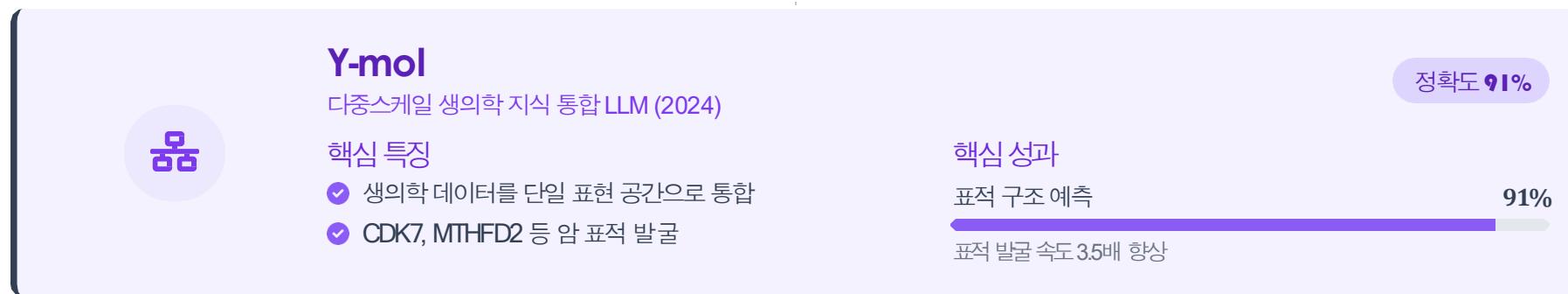
신규 표적  
발견율  
vs. 44% 전통적 방법

**65%**

표적 검증  
비용 절감  
평균 1.8M USD 절약

# • LLM 기반 약물 표적 발굴 핵심 모델

- 2025년 기준 주요 3대 의생명과학 LLM 모델과 성과



LLM 약물 표적 발굴 종합 성과(2025년 현재)