

AI 신약 개발 선도기업 리커전의 성공 전략

김수민

sumin.kim@lgb.co.kr

AI 기술이 신약 개발에도 변화의 바람을 불러일으키고 있다. 많은 시간과 비용, 낮은 성공 확률을 가진 기존 신약 개발의 문제를 해결하는 데 AI의 적용이 활발하다. 많은 AI 신약 개발 기업 중 리커전(Recursion Pharmaceuticals)은 5개 이상의 후보 물질을 임상에 진입시키며 앞서 달리고 있다.

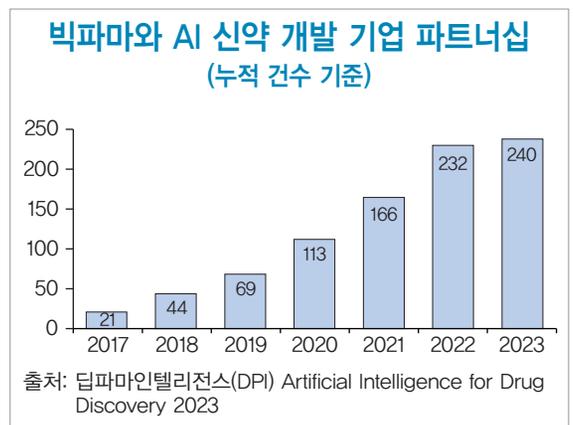
AI 기반의 드라이랩(Dry Lab)과 웨트랩(Wet Lab)을 통합 구축하여 신약 개발 과정의 효율성을 극대화한다는 점이 리커전의 핵심 경쟁력이다. 이를 위해 리커전은 ▲설립 초기부터 Wet Lab의 규모 확장과 자동화에 적극 투자하여 AI를 통해 세운 가설을 빠르게 검증할 수 있었으며, ▲이미지 해석에 특화된 AI 모델 '페놈베타'를 개발, 정상 혹은 약물을 처리한 세포의 미세한 변화를 분석하여 후보물질 발굴의 효율성을 높였다. 그리고 ▲외부 파트너십과 Wet Lab으로부터 광범위한 고품질 데이터 라이브러리를 확보하고, 자체 슈퍼컴퓨터를 통해 AI 모델 학습과 데이터 품질 제고의 선순환 구조를 구축할 수 있었다.

리커전 사례는 AI 신약 개발에서 AI 기술과 제약 산업 전문성의 결합을 통해 약물 파이프라인을 강화하는 것이 가치 제고의 핵심임을 보여준다. AI가 신약 개발에 필수적 수단의 하나로 자리매김함에 따라 제약 기업들은 데이터 확보나 자동화 분석 위한 파트너십 확대와 함께 전문 인력 확보와 육성 등 경쟁력 강화를 서둘러야 할 때이다.

2023년 11월 엔비디아는 AI 신약 개발 기업 리커전제약(Recursion Pharmaceuticals, 이하 리커전)에 5,000만 달러 규모의 투자를 발표하며 세간의 주목을 받았다. 향후 두 기업은 리커전 AI 모델을 탑재한 엔비디아의 신약 개발 플랫폼 ‘바이오니모(BioNeMo)’를 다른 기업이 활용할 수 있도록 서비스를 제공함으로써 신약 개발 파운드리 사업까지 협업의 범위를 확장할 예정이다. 한편 구글은 2024년 1월 자회사 아이소모픽 랩(Isomorphic Laboratories)을 통해 일라이 릴리(Eli Lilly), 노바티스(Novartis) 등 2개 기업과 30억 달러 규모의 AI 기반 신약 공동개발 계약을 체결했다. 이러한 움직임들은 AI가 본격적으로 신약 개발에 활용되면서, 제약 산업의 새로운 생태계가 만들어지는 모습으로 해석할 수 있다.

신약 개발에 소요되는 기간은 평균 10~15년, 비용은 1~3조 원에 달한다. 인체의 생명활동이 복잡해 신약 개발 과정에서 다양한 변수가 발생하고 실패 확률 또한 높기 때문이다. 하나의 후보물질이 상업화로 이어질 확률은 백만분의 1 이하다. 게다가 신약 개발의 타겟이 더욱 정밀해지고, 효능과 안전성에 대한 심사 기준이 점점 엄격해짐에 따라 개발 기간은 길어지며 비용 또한 늘어나고 있다. 이에 따라 R&D 투자 수익률은 계속해서 감소하고 있다. 딜로이트에 따르면, 제약 업계의 평균 R&D IRR은 2014년 7.2%에서 2022년 1.2%까지 떨어진 것으로 나타났다.

이러한 가운데, 최근 AI 신약 개발 기업들은 전 임상 단계까지 소요되는 R&D 기간과 비용을 25~50% 감축하는 등 실질적인 효과를 거두고 있다. 이미 다수의 글로벌 빅파마(Big Pharma)들은 AI 신약 개발 기업, 빅테크 기업과 적극적인 파트너십을 추진하고 있다. 예를 들면, 화이자(Pfizer)는 IBM과, 노바티스(Novartis)는 MS와, 로슈(Roche)는 리커전과 협력 관계를 구축하고 있다. 또한 엔비디아·IBM·구글 등 빅테크 기업들은 신약 개발용 AI 모델을 개발하며 제약 산업에 뛰어 들고 있다.

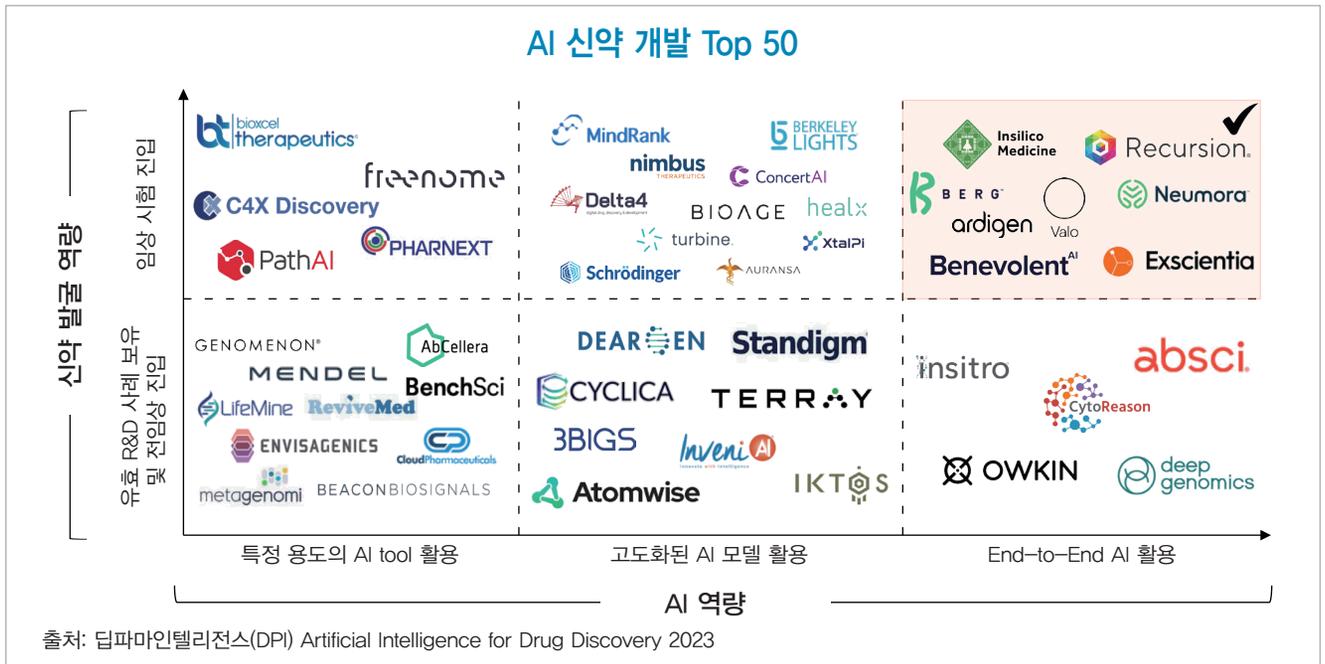


AI 신약 개발 선도하는 리커전

AI가 신약 개발의 중요한 도구로 자리매김하면서 AI 신약 개발 선도 기업들은 점차 AI의 활용 범위를 확장하며 신약 개발의 혁신을 이끌고 있다. 과거 AI 신약 개발 기업들은 R&D

AI 신약 개발 선도기업 리커전의 성공 전략

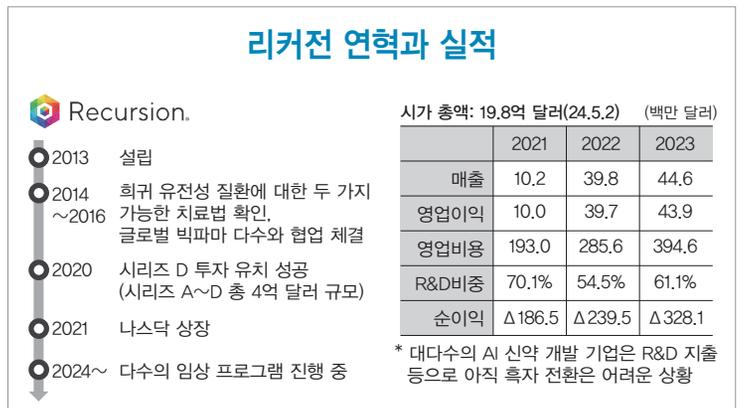
효율성이 가장 낮은, 후보물질 발굴 서비스에 초점을 맞췄다. 이들이 1세대라면, 최근 2세대 AI 신약 개발 기업들은 자체 신약 개발이나 파운드리 사업까지 영역을 확장하고 있다. 이들의 특징은 약물 디자인부터 후보물질 발굴까지 가능한 엔드투엔드(End-to-End) AI 모델과 첨단 자동화 실험실을 갖추고 있다는 점이다. 그러나 수백 개의 AI 신약 개발 기업 중, 엔드투엔드 AI 모델을 구축함과 동시에 AI로 발굴한 후보물질을 임상 1상 이상 단계에 진입시킨 기업은 현재 10곳 미만이다.



2013년 설립된 리커전은 현재 임상 2상에 4개, 2/3상에 1개 외에도 다수의 파이프라인을 보유하고 있으며 AI 신약 개발 기업 중 독보적인 성과를 보이고 있다.

그렇다면 AI 신약 개발 분야에서 리커전의 핵심 경쟁력은 과연 무엇일까?

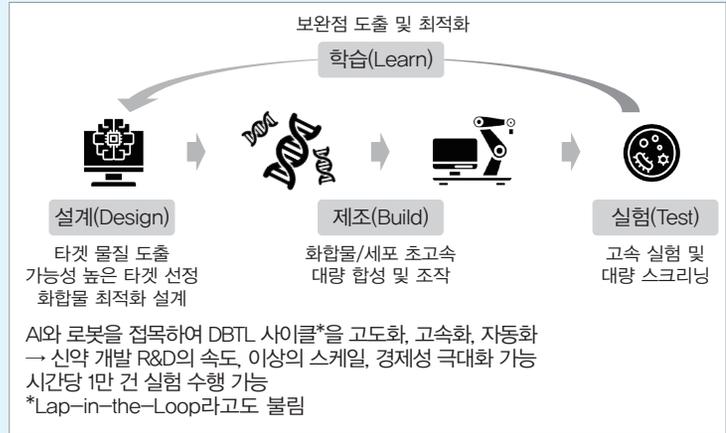
가장 주목할 점은 드라이랩(이하 Dry Lab)과 웨트랩(이하 Wet Lab)을 함께 구축하고 지속적으로 고도화했다는 것이다. 약물을 설계하고 합성하여 평가하고, 여기서 만들어진 데이터로 다시 약물 설계에 반영하는 과정을



1 임상2/3상은 임상 2상과 3상을 합쳐서 진행하는 것으로, 임상 1상을 통해 안전성이 충분하다고 판단된 이후 진행할 수 있다.

DBTL 사이클* (Design - Build - Test - Learn)

신약 개발 R&D 프로세스는 설계(Design), 합성(Build), 실험(Test), 학습(Learn) 단계를 거쳐 진행된다. 이러한 프로세스의 속도와 경제성을 극대화하려면 전산 역량을 갖춘 Dry Lab과 물리적 실험실인 Wet Lab 두 축이 함께 맞물려 움직여야 한다. 이를 통해 AI 기반으로 설계된 약물의 효능을 즉시 검증할 수 있고, 직접 실험·평가함으로써 균일하고 정형화한 양질의 데이터를 만들고, AI 모델의 예측 가능성을 높이는 선순환 구조를 확보할 수 있기 때문이다.



효과적으로 구축했다(DBTL 사이클*). 리커전이 이를 위해 추진했던 주요 전략은 크게 세 가지로 볼 수 있다.

1) 설립 초기부터 첨단 자동화 Wet Lab 구축

많은 AI 신약 개발 기업들은 AI 모델의 성능을 향상시키는 데 주력하고 있다. 그러나 약물을 합성·평가하는 Wet Lab을 자체 보유하지 않은 이상, AI가 세운 가설을 검증하고 해석하는 데 수 개월 이상의 리드타임이 추가적으로 소요된다.

리커전은 다른 AI 신약 개발 기업들과 달리 Wet Lab의 중요성을 일찌감치 파악하고, 설립 초기부터 Wet Lab 구축과 확장에 적극적이었다. 우선 물리적인 실험실을 구성하는 화합물(Compound) 투여기, 플레이트(Plate) 세척기, 배양 스테이션, 현미경 카메라 등에 로봇을 도입했다. 그 결과, 각종 발현 정보를 자동으로 수집할 수 있을 뿐만 아니라, 소규모 팀으로도 실험을 관리하고 작업 시간까지 단축할 수 있었다.

또한 동식물 등 생물체에서 직접 약물 효과를 검증하는 인비보(In vivo) 실험을 위해 온도 측정, 행동 관찰용 장비를 장착한 '디지털 케이지(Cage)'를 만들었다. 2023년에는 약물이 투여된 후 체내 이동을 관찰하기 위한 흡수(Absorption), 분포(Distribution), 대사(Metabolism), 배설(Excretion) 실험까지 내부에서 가능해지며, 전임상 연구에 필요한 모든 실험을 수행할 수 있는 규모의 Wet Lab을 확보했다.

한편, 매주 수백만 건의 실험을 수행하는 과정에서 오류를 최소화하고 균일한 데이터

품질을 보장하기 위해 실험 설계·조율에서부터 결과값 저장까지 모든 프로세스에 독자적 소프트웨어 플랫폼을 연동했다. 덕분에 정확한 반복 실험, 빠른 속도 및 다량의 실험을 수행할 수 있다. 현재 리커전은 주당 220만 건을 수행하며 생성된 실험 데이터를 바탕으로 AI 모델의 학습 능력과 예측 정밀도를 높이고 있다.

2) 강력한 이미지 해석 AI 모델, ‘페놈베타(Phenom-beta)’ 개발

기존에 약물을 발견하던 방식은 질병의 발생 원인에 대한 규명부터 시작한다. 따라서 질병에 대한 이해를 위해 몇 년의 시간을 보낸 후 유효한 화합물의 여부를 탐색하므로, 초반에 많은 시간이 걸린다. 그러나 리커전은 질병 자체를 이해하기보다는 질병의 결과로 드러나는 이미지를 관찰함으로써 유력한 후보물질을 찾아내는 데 집중했다.

리커전은 설립 초반에는 ‘셀페인팅(Cell Painting)’이라는 이미지 기반 프로파일링 소프트웨어를 활용하였다. 셀페인팅은 세포를 여러 색깔의 형광 염료로 염색하여 크기, 모양, 질감, 강도 등의 특징을 측정하는 기법이다. 즉, 세포의 표면적인 특징을 질병, 약물 또는 유전적 이상 현상의 영향에 대한 결과로 활용할 수 있다. 유전자가 발현되어 드러나는 세포의 특징적인 외형, 단백질 변형에 따른 세포의 변화 이미지 등을 살펴봄으로써 세포를 정상화할 수 있는 화합물이 무엇인지 찾아내는 것이다.

이후 리커전은 수백만 개의 이미지 데이터, 유전자 데이터 학습을 통해 난이도가 높은 복잡한 분석이 가능하도록 AI 모델을 업그레이드하였고, 2024년 1월 그 결과물인 ‘페놈베타’를 공개하였다. 페놈베타 모델을 활용하면 인간의 눈으로 감지할 수 없는, 미묘하지만 유의미한 이미지적 특징을 찾아내고 해석하기 쉬운 형태로 바꾸어 분석할 수 있다.

기존의 약물 발견 방식은 타겟 가설을 검증하기 위해 특정 생물학적 신호만을 탐색하기 때문에 분석 범위가 제한적이다. 반면, 페놈베타 모델은 세포의 모든 반응과 변화를 포착할 수 있어 보다 확률이 높은 약물 후보 물질 도출에 상대적으로 유용하다. 특히, 실험 과정에서 현미경으로 관찰되는 생물학적 반응은 이미지 형태로 캡처되기 때문에, 이를 해석하는 과정에서 페놈베타 모델의 강점이 발휘된다.

또한 이미지 기반의 오믹스(Omics)² 데이터는 기존 분석 방식 대비 달러당 데이터 밀도가 2~4배 높아 비용 대비 훨씬 많은 데이터를 생성할 수 있어 보다 빠르게 AI를 학습시킬 수 있었다.

2 오믹스는 유전체(Genome, DNA 대상), 전사체(Transcriptome, RNA 대상), 대사체(Metabolome) 등 분석을 총칭한다.

3) 대규모 고품질 데이터 라이브러리 확보

이미지 기반 AI는 타겟 세포와 화합물 간 결합을 관찰하는 단계에서 힘을 발휘한다. 그런데 어떤 타겟을 대상으로 연구를 진행할 지 결정하기 위해서는 수십억 개의 데이터를 통해 학습한 머신러닝 모델 구축이 선행되어야 한다. 이를 위해 양적으로 풍부한 데이터를 확보할 필요가 있다.

그러나 AI 신약 개발 기업들은 고품질 데이터로 구성된 대규모 데이터 라이브러리 확보에 어려움을 겪고 있다. 고품질 데이터는 AI 모델의 예측가능성을 높이기 위해 구조화·정형화해야 하며, 특정 분석을 위해 필요한 정보를 포괄적으로 담고 있어야 한다. 그러나 일반 공개된 데이터의 대부분은 결합이 있거나 편향되어 있어 AI 모델 학습에 유의미한 수준의 신뢰성을 확보하기 어렵다.

AI 신약 개발 분야의 권위자인 블라디미르 마카로브(Vladimir Makarov) 박사는 2022년, “특정 질병과 관련된 27개 공개 데이터 세트 중 단 4개만을 사용할 수 있었으며, 다른 사례에서 참조된 1,200개의 샘플 중 300개만을 AI 모델에 활용할 수 있었다”며, 공개 데이터의 저품질 이슈를 제기했다.

고품질 데이터는 특정 연구 목적에 맞게 설계된 실험을 직접 수행하는 과정에서 얻을 수 있으나, 주로 빅파마나 CRO(Contract Research Organization, 위탁 임상개발 기관) 등이 독점하고 있다. 특히, 실험의 시행착오 과정에서 생기는 부정적인 결과를 담은 데이터는 AI 학습에 있어서 필수적이거나, 공개되는 경우는 드물다. 따라서 고품질 데이터를 확보하려면 데이터 특화 파트너십을 확장하고 동시에 자체 실험실을 통해 데이터를 직접 생성할 수 있어야 한다.

리커전은 화합물·유전자 라이브러리 제공 전문 업체인 에나민(Enamine), 템퍼스(Tempus) 등과 협력하여 고품질의 데이터를 확보했다. 에나민은 세계 최대 규모의 화합물 라이브러리 제공 업체인데, 리커전은 에나민 라이브러리를 통해 360억 개의 화합물에 대한 단백질 표적을 예측했다. 템퍼스는 암환자 데이터 제공에 특화한 업체이며, 리커전은 템퍼스 데이터에 대한 라이선스에 기반하여 10만 명이 넘는 암환자의 DNA·RNA 데이터를 확보할 수 있었다.

여기에서 또 하나 주목할 사실은 다량의 고품질 데이터를 바탕으로 학습·추론하고 복잡한 분석을 수행하기 위해서는 일반 성능의 컴퓨터로는 역부족이라는 점이다. 리커전은 슈

퍼컴퓨터 ‘바이오하이브-1(Biohive-1)’을 통해 고품질 데이터를 충분히 학습할 수 있었다. 바이오하이브-1은 2023년 모든 산업 분야를 통틀어 세계에서 가장 강력한 슈퍼컴퓨터 상위 125위에 랭크되었다. 또한 엔비디아와 협력하여, Top 50위권 진입을 목표로, 바이오하이브-1의 용량을 4배로 늘릴 예정이다.

이외에도 리커전은 특정 연구 목적에 맞게 설계한 실험을 자체 Wet Lab에서 수행하여 즉각적으로 가설을 검증하고 데이터를 생성할 수 있다. 이 때 생성된 데이터는 AI 모델에 바로 적용할 수 있는 형태로 변형되어 리커전의 전용 데이터 세트에 저장된다. 이러한 노력 덕분에 리커전은 2017년 0.5페타바이트(Petabyte, 1015바이트, 이하 PB)에 불과했던 데이터 라이브러리 규모를 2023년 50PB까지 확장하고 지속적으로 데이터의 품질을 높일 수 있었다.

AI 신약 개발에서 가시적 성과를 얻기 위해서는

AI를 활용한 신약 개발은 R&D에 걸리는 시간과 비용을 단축해줄 뿐만 아니라, 기존 제품 대비 우수한 효능과 안전성을 지닌 약물을 개발할 수 있다는 점에서, 신약 발굴을 위한 필수적인 방법론으로 평가받고 있다. 리커전 사례를 통해 알 수 있듯이 AI 신약 개발에서 두각을 나타내기 위해서는 제약 산업에 대한 전문성을 바탕으로 AI 역량을 접목시키는 것이 중요하다. 신약 개발의 근본적인 메커니즘을 이해해야만 기존 신약 개발 방식의 문제와 발생 원인을 파악할 수 있으며, 이를 통해 신약 개발 과정에서 AI를 우선 도입해야 하는 지점을 알아낼 수 있다. 또한 AI 모델을 유의미하게 학습시키기 위해 구체적으로 어떤 데이터가 필요한지 판단할 수 있다. 이러한 이해를 바탕으로 AI를 활용할 때 비로소 AI의 효과가 강력하게 발휘될 수 있다.

뿐만 아니라, 질병 종류 혹은 R&D 프로세스에 특화된 AI 모델이 각각 존재하기 때문에, 구체적인 영역에 대한 전략을 수립하고 해당 영역의 AI 모델 확보와 고도화에 집중할 필요가 있다. 예를 들면, 약물 후보 발굴 외에도 AI를 적용하여 임상 시험도 효율화 할 수 있다. 임상 시험에서 중요한 단계는 대상 환자 모집인데, AI는 의료 기록, 환자 정보 등 다양한 데이터를 참고하여 시험에 적합한 환자군을 설계하는 데 유용하다. 이와 같은 노력들이 병행된다면 신약 개발 성공 확률을 높이는 데 일조할 수 있을 것이다.

한편, 리커전과 같은 기업들이 증가하며 제약 산업에서 ‘AI 신약개발 기업’이라는 새로운 사업모델 유형이 만들어지면서 동시 경쟁하는 체제가 심화되고 있다. 제약 업계에서는 신

약 파이프라인이 기업의 가치와 직결되기 때문에 AI를 활용한 파이프라인 구축 경쟁이 활발한 한편, 데이터의 소유권 이슈 등 협업의 문제도 드러나고 있다. 때문에 장기적인 관점에서 리커전처럼 자동화 Wet Lab을 구축하고, 궁극적으로는 독자적인 AI 신약 개발 플랫폼을 확보하는 기업이 더욱 주목을 받을 것이다.

이를 위해서는 최적의 인적 역량을 확보하는 것이 무엇보다 중요한 첫걸음이 될 것이다. 신약 개발을 이해하는 AI 인력과 AI를 이해하는 신약 개발 인력 양쪽 모두가 필요한데, 아직은 부족한 상황이다. 그렇기에 AI 신약 개발을 직접하거나, 이를 활용하려는 기업 모두 관련 전문 인력 확보에 더욱 적극적으로 나서야 할 시점이다. LG경영연구원