William Blair

AI INSIGHTS

Equity Research
Healthcare | Life Science Tools and Diagnostics

March 30, 2026
Industry Report

Matt Larew +1 312 801 7795
mlarew@williamblair.com

Jacob Krahenbuhl +1 312 364 8171
jkrahenbuhl@williamblair.com

From Code to Clinic

How AI Is (and Isn't) Rewriting the Life of a Drug

Contents

Portfolio Manager Summary 3

Executive Summary 3

Ecosystem Overview 9

Public Company Coverage 11

How AI Will Impact the "Life of a Drug" 12

Why AI Is Relevant: Drug Discovery Is Expensive and Biology Is Programmable 12

Biological Foundation Models – In the Dry Lab, Data Is the Limiting Reagent 14

How AI Is Impacting Antibody Discovery – Smarter Funnels and Faster Cycles 20

Target Selection (Steps 1-2) 22

Step 1 – Target Identification 22

Step 2 – Target Validation 23

Step 2.5 – Antigen and Assay Preparation (Screening Preparation) 24

Lead Discovery (Steps 3-4) 25

Step 3 – Hit Identification, Antibody Design, and Primary Screening 25

Step 4 – Cloning and Expression 26

Lead Characterization (Steps 5-7) 27

Step 5 – Binding Characterization 27

Step 6 – Functional Assays 28

Step 7 – Developability Assessment 29

Optimization Cycle 30

Step 8 – Antibody Engineering and Optimization 30

Step 8.5 – Candidate Selection 31

IND-Enabling Development (Parallel Track of Steps 9-10) 31

Step 9 – Chemistry, Manufacturing, and Controls (CMC) 31

Step 10 – IND-Enabling Preclinical Studies 32

Appendix: Private Companies in AI 33

Portfolio Manager Summary

Here is the short version: AI is not replacing the physical work of drug discovery, but it is reshaping it. It is true that AI is increasingly being leveraged as a "dry lab" (i.e., computational) tool for hypothesis generation, but those hypotheses will always need to be physically tested in the "wet lab" (i.e., the real world) and the overwhelming majority of drug discovery spend relates to the work of proving what's true in the real world. Hence, the "wet lab obsolescence" narrative is overstated and misframed.

Separately, the most balanced assessment of AI in drug discovery today is that it is good at generating ideas. Not necessarily good ideas, but lots of them—quickly. Like all AI, the quality of outputs relies on the quality of inputs, and biology today lacks the requisite inputs (standardized, fit-for-purpose, experimental data) to generate desired outputs (novel drug ideas). Hence, for AI to reach its potential in biology, a significant amount of new experimental data needs to be generated, creating a durable tailwind for the tool stack sitting at those data bottlenecks.

On the **stocks**: the life science tools and services group has traded off categorically in recent weeks, partly on fears that AI will dramatically reduce wet lab utilization in drug discovery. However, on balance we view AI as a net tailwind for most companies we cover given where their products sit in the drug discovery workflow. Thus, the setup is asymmetric—alpha is available both if AI becomes a ubiquitous tool in drug discovery (underappreciated flow-through to physical lab work) and if it does not (dispelling the wet lab obsolescence thesis). Investors should be adding exposure as end-markets continue to heal and growth accelerates throughout 2026.

Executive Summary

In a 2021 report (Programming Life), we discussed how advances in AI were enabling a new model for drug discovery—one that leverages the inherent programmability of biology to automate and accelerate the design–build–test–learn (DBTL) cycle. As is often the case with new technology waves, excitement created hype that was more on par with promises made by the field than progress made on the field, which in turn created a bubble-popping cycle that led to persistent skepticism about the potential impact of AI in biology ever since. More recently, however, that paradigm has flipped on the heels of announcements by Anthropic about Claude for Life Sciences and broader awareness of progress made with in silico drug discovery tools (e.g., Microsoft CEO Satya Nadella's recent X post about Microsoft's GigaTIME). The S&P Life Sciences Tools & Services Index is down 19% year-to-date and 25% from January highs, in part related to fears that the recent inflection in news flow is indicative of an inflection point in utilization of the "dry lab" (i.e., computational work) at the expense of the "wet lab" (i.e., physical lab work) in future drug discovery efforts.

These concerns contemplate, perhaps with equal levels of imprecision, the potential for AI to 1) reduce discovery work by identifying targets and designing drugs and 2) reduce analytical work load by powering predictive models pertaining to efficacy and safety, with a potential offset from 3) more molecules entering the funnel due to the improving efficiency of drug discovery and development. There seems to be consensus (less wet lab demand) without conviction on the first two points, and hope (more shots on goal) without resolution on the second.

Our work builds on conversations with 20 companies and key opinion leaders in the space, a deep dive on biological models, and analysis of the drug discovery process and how cost distribution, AI disruption, and companies' products map to that process. Based on this, we see the narrative being reshaped in the coming years as the use-cases and limitations of AI in drug discovery are better understood and as evidence emerges that AI can serve as a tailwind to the broader ecosystem (e.g.,

powering the production and adaptation of biological models, shifting test volume to high-content and connected instruments, accelerating drug discovery, enhancing the efficiency of clinical trials, and scaling and automating manufacturing). In the coming quarters, we believe company results and communication will help investors better understand where AI poses real disruption risk (e.g., hypothesis creation) and where it does not (e.g., hypothesis validation) and how revenue exposure maps to those categories. We estimate that roughly 85% of preclinical discovery work for an antibody program carries low to modest residual AI disruption risk. This is not to be dismissive of the promise of AI in drug discovery but instead reflects the inherently physical nature of much of the drug discovery process.

Exhibit 1
Antibody Drug Discovery Workflow and Total Program Costs (per IND)

Antibody Drug Discovery	D-B-T-L	Current AI Adoption Level	Residual Disruption Risk	% pre-CMC	% total pre clin
1) Target Identification	L	HIGH	MODEST	11.0%	5.8%
2) Target Validation	T	MEDIUM	LOW	9.0%	4.7%
2.5) Antigen Design/Production	D-B	MEDIUM	MODEST	1.4%	0.7%
3) Hit ID, Ab Design, & Primary Screening	D-B-T-L	MEDIUM-HIGH	MODEST	13.8%	7.3%
4) Cloning & Expression	B	MIXED	LOW	2.1%	1.1%
5) Binding Characterization	T	LOW	LOW	6.9%	3.6%
6) Functional Assays	T	LOW	LOW	28.3%	14.9%
7) Developability Assessment	T	MEDIUM	MODEST	4.1%	2.2%
8) Antibody Engineering & Optimization	D-B-T-L	MIXED	MEDIUM	23.4%	12.4%
9) CMC	B-T	LOW	LOW		35.6%
10) IND-enabling Preclinical Studies	B-T	LOW	LOW		11.6%

Sources: Paul et al (Nature), PwC, Tufts CSDD, Contract Pharma 2010 mAb CMC, William Blair Equity Research

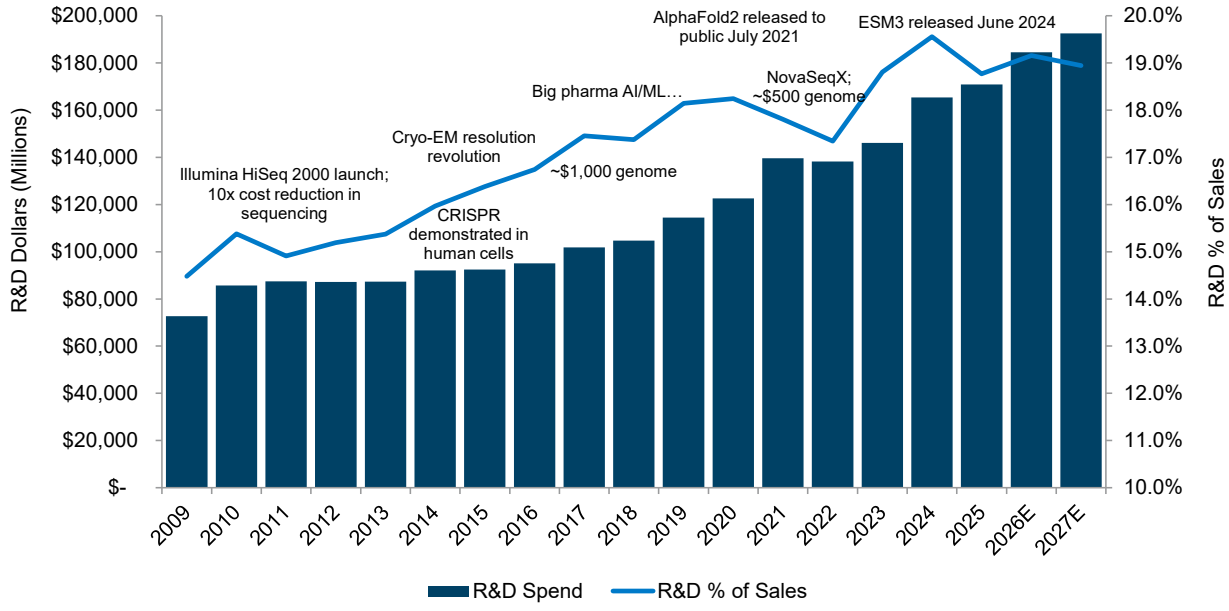
In other words, the "learning and hypothesizing" steps of drug discovery **are** benefiting from AI as a "tool in the tool kit," but the "validating and characterizing" steps are, and will remain, physical**.** This should mean lower time and energy/cost per shot on goal, and in turn more demand for tools to build, validate, and iterate those shots. Put differently, in the design–build–test–learn (DBTL) cycle of drug discovery, we see AI being leveraged increasingly for "D" and "L," and in turn creating even more "B" and "T" work to be done. We estimate over 80% of pre-IND work (and ~70% of pre-CMC work) fits in the "B" and "T" camps. Lastly, if successful, a quantifiable outcome will be that more drug candidates enter the clinic and are eventually approved, meaning more clinical trials and commercialized drugs, presumably yielding a Jevons paradox with more money spent on R&D if those dollars prove to have higher returns.

We also point out that for all the focus on the potential for AI to reduce wet lab discovery costs, the largest opportunity for savings per approved drug is in reducing clinical failure rates, particularly at Phase II. When comparing lowering the cost of preclinical steps by 20%, making them 20% faster, and reducing the failure rate by 20%, the results overwhelmingly show that the most value is generated by reducing clinical failure rate. Moreover, multiple models have found that the effect

is most pronounced at the Phase II clinical trial stage of drug development. Thus, we believe the biggest economic opportunity for AI in drug development is in helping improve the quality of candidates entering clinical trials rather than simply reducing the cost of each shot on goal.

As to the pushback that AI adoption will cause total pharma R&D spend to shrink, we point out that the list of technology innovations brought to bear in drug discovery over the past decade is lengthy, including those that directly reduce cost per insight (NGS is a great example, as the chart below details). However, pharma R&D—gross dollars and as a percentage of sales—continues to grow, including since AI/ML began adoption in 2017.

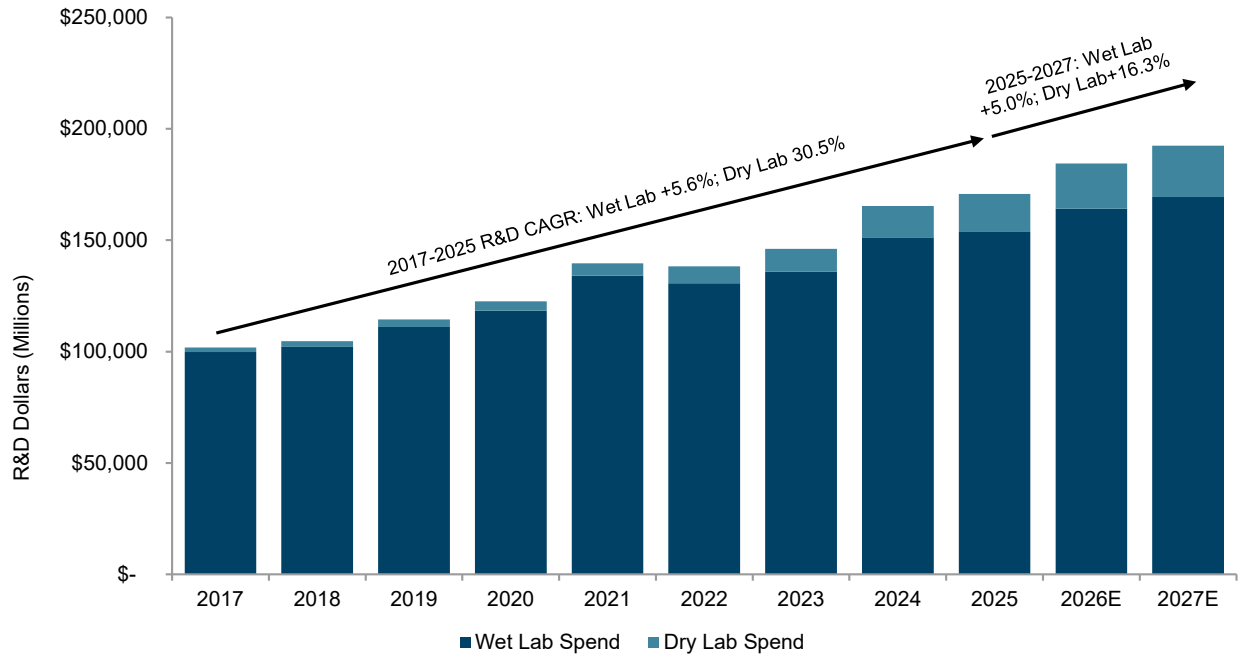
Exhibit 2
R&D Spend at Top 20 Pharma



Sources: Company websites, FactSet, William Blair Equity Research

The introduction of AlphaFold2 in 2021 made structure prediction more mainstream, and our industry conversations suggest that "compute" as a line item for research has grown from de minimis in 2017 to about 10% at big pharma today (which squares with a November 2025 Benchling survey pointing to about 14% of R&D budget spent on AI in total). Keep in mind that nearly 60% of pharma R&D spend is allocated to labor, which should be the category of spending most impacted by compressed timelines, while we estimate 10-15% of pharma R&D spend is tools-related spending (e.g., instruments, consumables, and reagents). Weighted-average life science tools revenue growth since 2017 (~7%) compares favorably to the prior 10-year period (~4%), although excluding the 2020-2021 COVID-related boom in growth, it has averaged ~3.5%. This includes declining revenue in 2023 and 2024 and low-single-digit growth in 2025 as end-markets suffered from destocking, pipeline rationalization, and regulatory and funding pressures.

Exhibit 3
Big Pharma Wet Lab vs. Dry Lab Spending



Sources: William Blair Research, FactSet

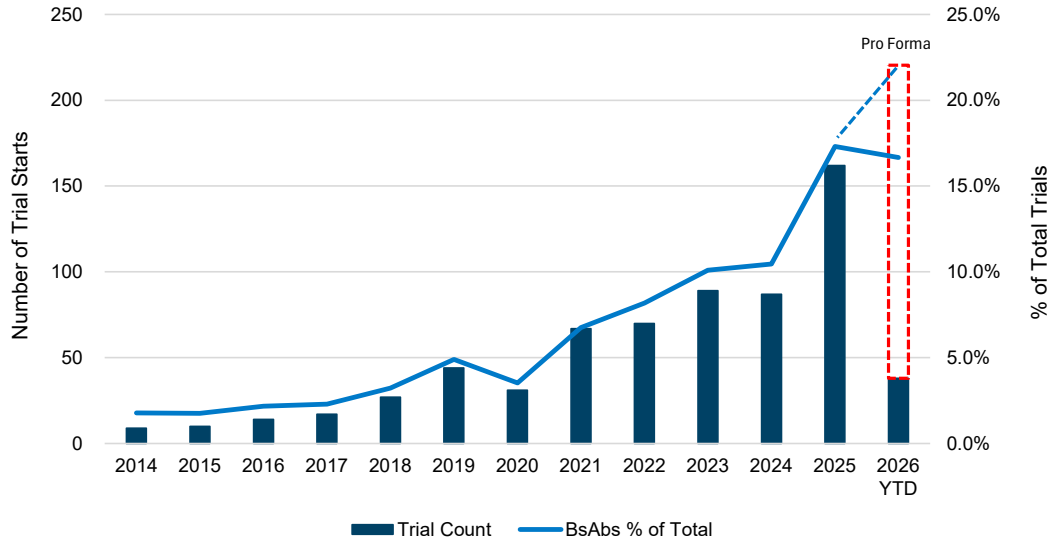
Further, if AI is leveraged successfully to improve drug discovery, it will in large part be powered by biological models to generate ideas (i.e., targets and drugs). For those models to "work," a robust (defined by size, diversity, quality, and context) collection of novel data must be generated, which will be a tailwind for high-resolution analysis (e.g., single cell, spatial). As we explain in detail later in the report, these models are quite heterogenous—built on different data inputs (wet versus dry lab) and with various layering requirements to be adapted into target or drug discovery engines. But in short, foundation models themselves are a starting point, and much like a company may adopt an LLM internally to be trained on in-house data and processes, so must a company layer in data to transform a foundation model into a disease/tissue/organ/cell-specific drug discovery platform.

It also should be said that while there is growing effervescence about the promise of AI in drug discovery, to date the most grounded assessment is that biological models are good at rapidly creating hypotheses. Not necessarily "good" or "better" hypotheses than traditional methods, just a lot more of them, faster. The most recent example of this is Boltz-2, an AI/ML model for protein-ligand cofolding that has become popular since its June 2025 release. A March 2026 independent evaluation found that while Boltz-2 can be useful as a fast first-pass triage tool, its affinity predictions are far less accurate/helpful (similar third-party assessments of AlphaFold3 have been published) All this to say: for AI to make a step-function change in drug discovery, high-quality experimental data is king, because the models ultimately run on (and are constrained by) the data that is generated and validated in the wet lab.

Further, as models improve, we believe they will be applied to more targets, more modalities (bispecifics, nanobodies, ADCs, Fc fusions, minibodies), and more difficult problems (undruggable targets, novel epitopes), which heightens the demand for the development and characterization

work that public company exposure is weighted to. For reference, the number of bispecific antibody-focused clinical trials has grown from roughly 9 total trial starts in 2014 (2% of all combined therapeutic antibody trials) to over 150 in 2025 (>15% of trials).

Exhibit 4
Total BsAbs Clinical Trial Count

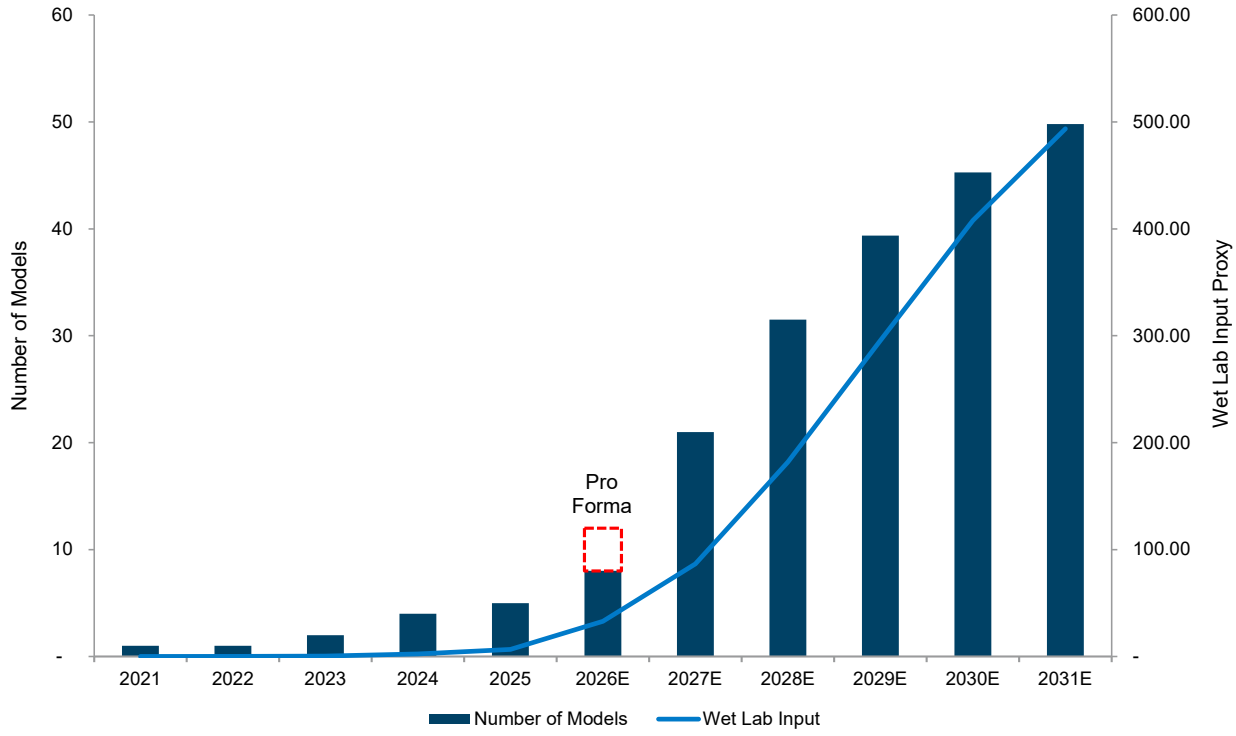


Sources: clinicaltrials.gov

Like all AI models, the usefulness of models is directly predicated on the quality and utility of the data they are trained on, something the field is grappling with as excitement and awareness broaden into the mainstream. For example, in Benchling's 2026 Biotech AI Report, more than half of failed pilots related to data challenges (availability and integrity) and ~75% of respondents called out data quality, standardization, and accuracy as a top three priority for AI-driven R&D.

On durability of demand, consumer-facing products that investors may be familiar with (e.g., Claude, ChatGPT) are built on natural language models that are continuously improved on by training more sophisticated algorithms on larger, more robust datasets. The same thing is occurring in life sciences, with the number of biological foundation models and size of training data growing dramatically in recent years (the exhibit below focuses exclusively on foundational models built on wet lab inputs). New entrants raise the bar for training parameters (e.g., Xaira's X-Cell Ultra model built on 25.6 million perturbed cells) and individual models are iterated over time. A good example is the evolution of Recursion's cell imaging dataset, which over four years increased from approximately 125,000 parameters to 2.2 million.

Exhibit 5
AI Foundation Model Wet Lab Inputs



Sources: William Blair Equity Research, Company websites

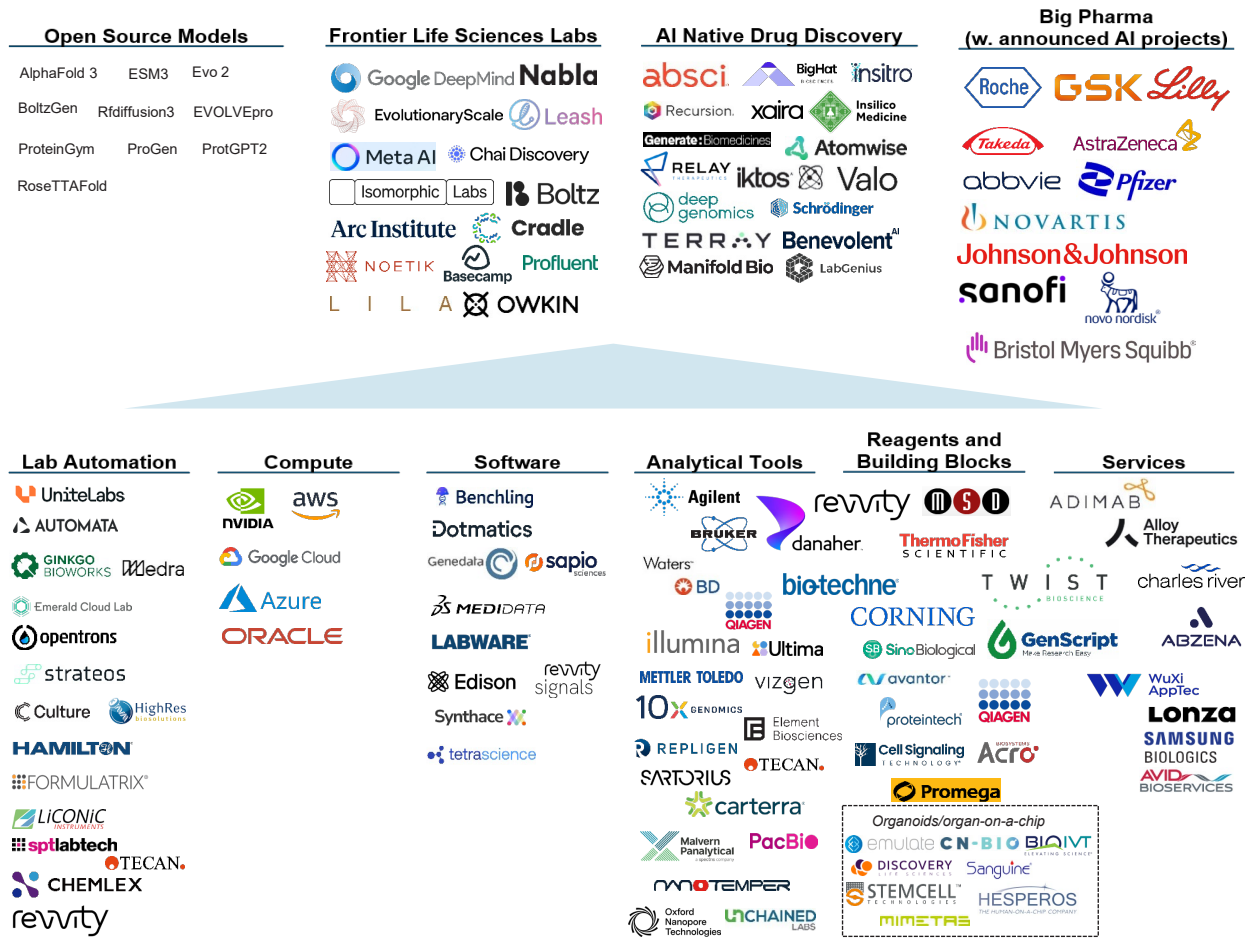
A reminder that 100% of the roughly 3 billion base pairs genome has been mapped, and the vast majority of the hundreds of thousands of transcripts in the transcriptome have been mapped; however, the proteome is far less understood (perhaps less than 20% of proteoform structures have been mapped) and potential for AI to create de novo proteins further widens the universe of possible therapeutics. Put differently, by Arc Institute in its recent BioReason-Pro preprint: over 250 million protein sequences are known, but fewer than 0.1% have confirmed functions. This knowledge gap is driving the scale of models, as defined by number of parameters per model (in some cases wet lab inputs), higher—and that will continue to increase in the future.

To reiterate our thesis: we believe the utility of AI in drug discovery is at an inflection point, and AI will be increasingly levered as a tool to understand human biology, identify druggable targets, repurpose existing drugs, and discover and/or design new drugs. This inflection point will lead to a significant increase in wet lab data points (primarily omics/sequencing and imaging) required to adapt existing and forthcoming foundation models to drug discovery engines. And ultimately, we expect AI to be successful in improving the efficiency of drug discovery (defined as time and cost expended per approved drug); however, the economic exposure for the life science tools ecosystem is not heavily weighted to hypothesis creation (where AI is making an impact) but rather to hypothesis validation (which occurs in the physical world). While demand-destructive categories exist (e.g., target ID), other demand may shift, both between steps (e.g., from lead ID to lead optimization) and between technologies (e.g., from animal to non-animal models). Overall, we believe the overwhelming majority of existing wallet share faces low or modest *residual* disruption risk.

Ecosystem Overview

The adoption of the dry lab in a historically wet-lab–dominated drug discovery field has altered the landscape. However, at its core, the ecosystem still consists of 1) companies generating novel biological insights and turning them into therapeutics, and 2) companies providing tools and services that enable those discoveries.

Exhibit 6
AI Drug Discovery Ecosystem



Sources: Company websites, William Blair Equity Research

In recent years, collaboration between the layers of the ecosystem has resulted in a number of partnerships, which we highlight below.

Exhibit 7
Pharma, Biotech, and AI Partnerships

Date	Partnership		Announcement Headline	Funding Amount ($)
	Pharma	Biotech / AI		
Mar-26	Eli Lilly	InSilico Medicine	*"Insilico Medicine Announces Global R&D Collaboration with Lilly"*	$2.75B
Mar-26	Roche	NVIDIA	*"Roche launches NVIDIA AI factory to accelerate the development of new therapeutics and diagnostics solutions"; The collaboration between the two companies began in 2023*	N/A
Feb-26	Takeda	Iambic Therapeutics	*"Iambic Announces Collaboration with Takeda to Advance AI-Driven Design of Small Molecules"*	$1.7B
Jan-26	Sanofi	CytoReason	*"CytoReason extends its collaboration with Sanofi to advance AI-driven drug discovery"*	$16M
Jan-26	Eli Lilly	NVIDIA	*"NVIDIA and Lilly Announce Co-Innovation AI Lab to Reinvent Drug Discovery In the Age of AI"*	$1B
Jan-26	Thermo Fisher	NVIDIA	*"Thermo Fisher Scientific Announces Strategic Collaboration With NVIDIA Leveraging AI to Advance Scientific Instrumentation and Accelerate Laboratory Performance"*	N/A
Jan-26	GSK	Noetik	*"GSK Licenses Noetik's AI Foundation Models in Anchor Partnership to Transform Cancer Therapeutic Research and Development"*	$50M
Jan-26	Bayer AG	Cradle	*"Bayer and Cradle enter collaboration to enhance AI-enabled antibody discovery and optimization"*	N/A
Jan-26	Bristol Myers Squibb	Microsoft	*"Bristol Myers Squibb Announces Collaboration with Microsoft to Advance AI-Driven Early Detection of Lung Cancer"*	N/A
Jan-26	Eli Lilly	Nimbus Therapeutics	*"Nimbus Therapeutics Announces Research Collaboration and License Agreement with Lilly for Novel Oral Obesity Treatment"*	$1.3B
Nov-25	Eli Lilly	InSilico Medicine	*"InSilico and Lilly enter a research and licensing collaboration to advance AI-driven drug discovery"*	$100M
Oct-25	Takeda	Nabla Bio	*"Nabla Bio Signs Second Takeda Collaboration to Advance AI-Driven Design of Protein Therapeutics"*	$1B
Aug-25	Absci	Almirall	*"Almirall and Absci Expand AI Drug Creation Collaboration Adding a Second Dermatology Target"*	$650M
Jun-25	AstraZeneca	CSPC Pharmaceutical Group	*"AstraZeneca enters strategic collaboration with CSPC Pharmaceuticals focused on AI-enabled research"*	$3.6B
Jun-25	Agenus	Noetik	*"Agenus and Noetik Enter Collaboration to Develop AI-Enabled Predictive Biomarkers for BOT/BAL Using Foundation Models of Virtual Cell Biology"*	N/A
Jan-25	TME Pharma	aimed analytics	*"aimed analytics & TME Pharma: AI-driven drug discovery"*	N/A
Jan-25	Pfizer	PostEra	*"PostEra announces expansion to $610M in their AI drug discovery collaboration with Pfizer"*	$610M
Sep-24	Novartis	Generate Biomedicines	*"Generate:Biomedicines Announces Multi-Target Collaboration with Novartis to Discover and Develop Protein Therapeutics with Generative AI"*	$1B
May-24	Sanofi	Formation Bio x OpenAI	*"Sanofi, Formation Bio and OpenAI announce first-in-class AI collaboration"*	N/A
Jan-24	Eli Lilly	Isomorphic Labs	*"ISOMORPHIC LABS ANNOUNCES STRATEGIC MULTI-TARGET RESEARCH COLLABORATION WITH LILLY"*	$1.7B
Jan-24	Novartis	Isomorphic Labs	*"ISOMORPHIC LABS ANNOUNCES STRATEGIC MULTI-TARGET RESEARCH COLLABORATION WITH NOVARTIS"*	$1.2B
Dec-23	AstraZeneca	Absci	*"Absci Announces Collaboration with AstraZeneca to Advance AI-Driven Oncology Candidate"*	N/A
Dec-23	Abbvie	BigHat Biosciences	*"AbbVie and BigHat Biosciences Announce Research Collaboration to Leverage Artificial Intelligence and Machine Learning to Discover Next-Generation Therapeutic Antibodies"*	$30M
Nov-23	Roche	NVIDIA	*"Genentech and NVIDIA Enter Into Strategic AI Research Collaboration to Accelerate Drug Discovery and Development"*	N/A
Nov-23	Amgen	PostEra	*"PostEra Announces a Research Collaboration with Amgen to Discover Small Molecule Therapeutics using Artificial Intelligence"*	N/A
Oct-23	Sanofi	BioMap	*"BioMap Establishes a Strategic Collaboration with Sanofi to Co-Develop AI Modules to Accelerate Drug Discovery for Biotherapeutics"*	>$1B
Jul-23	BioNTech	InstaDeep	*"BioNTech Completes Acquisition of InstaDeep"*	>$500M
Oct-22	Eli Lilly	Nimbus Therapeutics	*"Nimbus Therapeutics Announces Research Collaboration and License Agreement with Lilly for Small Molecule Activators of AMPK"*	$496M
Jan-22	Sanofi	Exscientia	*"Exscientia and Sanofi establish strategic research collaboration to develop AI-driven pipeline of precision-engineered medicines"*	$5.2B
Dec-21	Pfizer	PostEra	*"PostEra establishes $260M strategic AI Lab partnership with Pfizer"*	$260M
Dec-20	Pfizer	PostEra	*"PostEra enters a multi-year strategic partnership with Pfizer Inc. to advance machine learning for Drug Discovery"*	N/A
Oct-20	Bristol Myers Squibb	Insitro	*"insitro Announces Five-Year Discovery Collaboration with Bristol Myers Squibb to Discover and Develop Novel Treatments for Amyotrophic Lateral Sclerosis and Frontotemporal Dementia"*	$2.1B
Apr-19	AstraZeneca	BenevolentAI	*"AstraZeneca starts artificial intelligence collaboration to accelerate drug discovery"*	N/A

Note: This is a non-exhaustive list of major partnership announcements
Sources: Company websites, William Blair Equity Research

Public Company Coverage

For large-cap tools, exposure to preclinical drug discovery varies, with Bio-Techne (~60% of revenue; 25% early discovery, 35% translational) and Revvity (~50%) at the high end, followed by Thermo (~10%-15%), Agilent (10%), Waters (~7%), Mettler-Toledo (we estimate <5%, company has shared ~20% of sales are tied to R&D, but most of that is late-stage development and scale up), and Danaher (<5%). There is a barbell effect with small- and midcap names in the space: bioprocessing (Repligen) has minimal exposure and pharma packaging (Aptar, West, Stevanato) has no preclinical exposure, while Twist (~45%) has higher exposure and 10x is essentially 100% exposed.

Exposure does not necessarily map to AI disruption risk, however, at both a sector and company level. First, from a discovery workflow standpoint, roughly half of preclinical spend is in CMC and IND-enabling clinical studies. Second, in our view, the step with the most residual risk of disruption (lead engineering/optimization) represents about 10% of preclinical spend (20% of pre-CMC spend), while roughly 70% of spend occurs in steps we believe have low residual disruption risk.

Viewed through these lenses, pharma packaging (e.g., West, Stevanato, Aptar) carries no AI disruption risk and should benefit if the AI maximalist case (more drugs approved at a faster pace) plays out. Repligen is a pure play on bioprocessing, though does very modest exposure related to process development (we estimate about 10% of revenue), though we do not see much AI disruption risk to that step. Among diversified companies, Danaher and Sartorius have heavy bioprocessing exposure (and Danaher a large diagnostics franchise), making them low-downside plays, while Waters and Agilent are weighted downstream in development to QC work that has low disruption risk, in our view. Bio-Techne and Revvity have much higher exposure than peers, though as we explain below, that exposure is largely tied to validation and characterization steps, not hypothesis-creation steps that AI is disrupting. Twist, 10x, and Illumina are tied to discovery work but are positioned as winners given they help create the biological insights needed to adapt models to drug discovery engines.

As to specific winners and losers from our public company coverage list, we view clear winners as Twist Bioscience (increased demand for synthetic DNA) and 10x Genomics. Twist benefits from both the rise in the number of AI-driven antibody discovery campaigns (order size as much as 10 times higher than traditional campaigns) and the growing demand for more biological data that is needed to adequately and effectively train models. Here, Twist provides oligo pools needed to build the deep mutational scanning (DMS) libraries that are in turn the foundation for some biological foundation models. For example, frontier AI bio lab Cradle Bio has specifically called out Twist's role in its wet lab platform. Twist recognized $25 million of AI orders in fiscal 2025 (off a base of zero in the prior-year period) and we see that growing 50%-plus moving forward.

As to losers, while we discuss this in detail in the report, we see more risk at a specific technology level than at a company level. For example, binding affinity is an essential first test for antibody candidates, traditionally tested via ELISA (bind/no bind) followed by SPR (binding kinetics). However, Carterra's high-throughput SPR technology is potentially creating an opportunity to move straight to SPR because its high-throughput system matches the larger scale of binders emerging from AI lead generation programs. Another example is the growing momentum for non-animal models (NAMs) at the industry and regulatory level, which may mean that organoids/organ-on-a-chip (readout is high-content screening like Revvity's Opera Phenix) and predictive models might allow reduced reliance on in vivo aspects of the discovery process.

A private company appendix is included at the conclusion of this report, covering frontier life sciences labs, AI-native drug discovery companies, and software and hardware tools and services focused on delivering the scale and automation demanded in an AI drug discovery world.

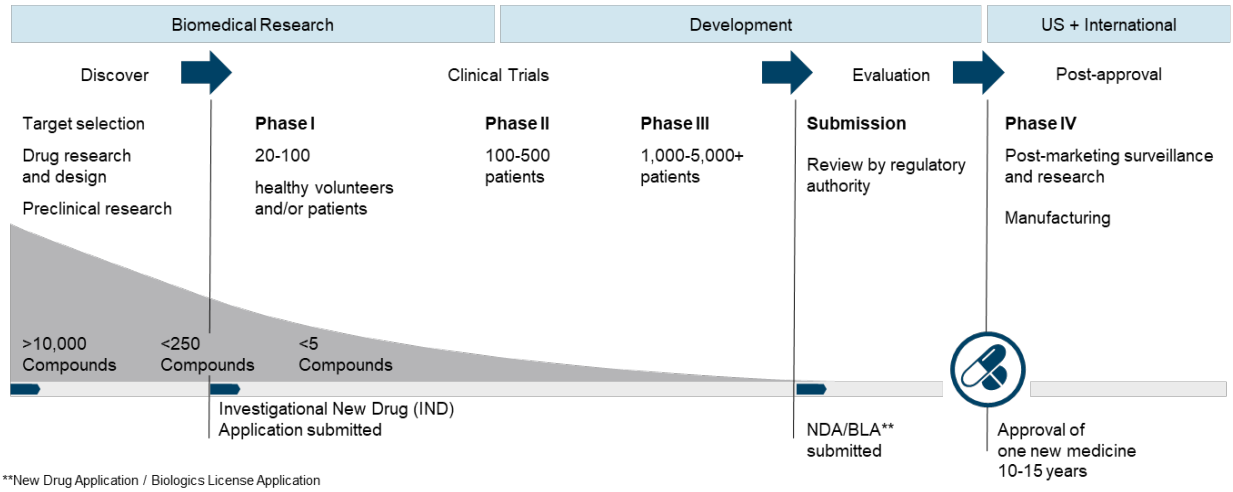
How AI Will Impact the "Life of a Drug"

Why AI Is Relevant: Drug Discovery Is Expensive and Biology Is Programmable

Drug Discovery Is Expensive

Investors are likely familiar with the Eroom's law problem of drug discovery. Despite major advances in science and technology, which have brought new tools and methods that enable scientists and drug researchers to work with ever-increasing efficiency, the cost of developing a new drug has roughly doubled every nine years since 1950. The effect has been termed Eroom's law, in homage to Moore's law, which refers to the observation that computing power has tended to double about every two years since the advent of computing. Metrics that are widely cited point to a 10-year-plus time horizon to get a new therapy on market, at a cost of more than $2 billion and with a 10% success rate.

Exhibit 8
The Path to a New Medicine



**New Drug Application / Biologics License Application

Sources: weforum.org, William Blair Equity Research

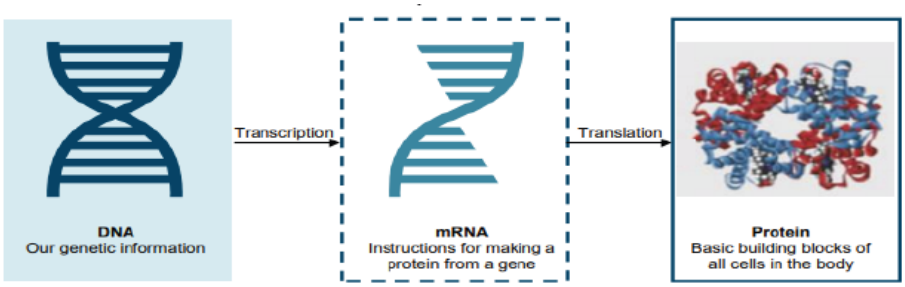
Biology Is Programmable

Cells do not act solely as stores of information but also as processors of information. In the same way that computers can be programmed with binary code, cells from every living creature on Earth can be programmed with the code of life, DNA. Unlike the binary code used for computers, which is made up of zeros and ones, DNA is made up of four nucleotides—adenine (A), thymine (T), cytosine (C), and guanine (G).

So, if biology is programmable, then how does the program work? While the specifics are elegant and complex, the basics are straightforward. Distinct regions within DNA strands called genes encode for the production of specific proteins that carry out nearly every task in an organism. The two-step process by which a cell's DNA is converted into functional proteins is referred to as the *central dogma* of molecular biology. First, DNA is transcribed into single-stranded messenger RNA (mRNA) in the nucleus of a cell during transcription. This mRNA is then transferred out of the nucleus to the cytosol, where it is translated by cellular organelles called ribosomes, which join amino acids together to form a functional protein. Moderna refers to mRNA as the "software of life" as it converts information stored as code (DNA) into applications (functional proteins). This process is illustrated below.

Exhibit 9

Schematic of DNA Transcription and mRNA Translation Events



Sources: TBIO company presentation, William Blair Equity Research

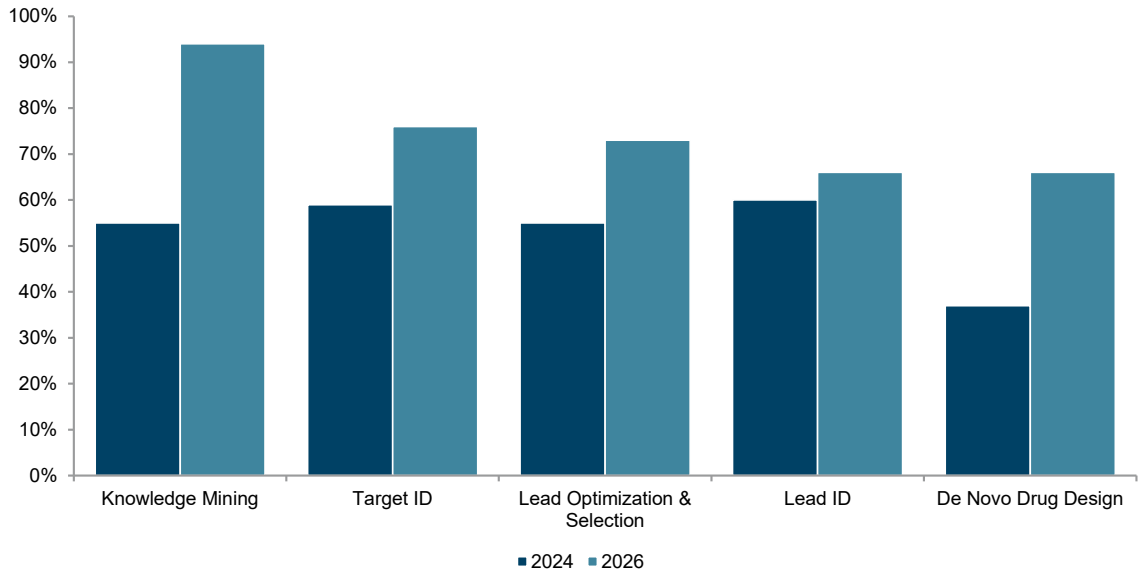
Importantly, the "language" of DNA, RNA and proteins is universal across biology. This, combined with technical improvements in the fields of sequencing, imaging, and multi-omics (together creating large multimodal data sets), makes biology well-suited to benefit from AI. These developments became more mainstream with AlphaFold (which won the 2024 Nobel Prize in chemistry), which demonstrated the ability to predict protein structures, and that effort has been built on in recent years as both the number of models and parameters per model have grown rapidly (as shown above).

While high-profile examples of AI-focused drug development tend to focus on computer-generated compounds, AI and ML are, at their core, techniques for a computer to make decisions based on data and information, and these techniques have the potential to be positioned at any, or even all, steps of the drug development process—from the understanding of biology, to drug development and lead optimization, through to the design and execution of the clinical trial process. Therefore, while AI- and ML-assisted drug development is a single concept, it represents a near limitless number of distinct and differentiated approaches for improving each step in the drug development process. By reducing the length of drug development, reducing costs associated with drug development, and increasing the probability of success of a program, these techniques aim to bring value to the process and to increase industry throughput, delivering more and higher-quality products to patients who need them the most. While AI- and ML- assisted drug development relies heavily on computational techniques, we highlight that it is distinct from more traditional computational or simulation techniques, which are focused on directly modeling or programming a system and have been used for several years in drug development, such as simulation models.

Pharma R&D programs have been gradually adopting AI/ML tools over the past decade, but the pace of adoption has increased in recent years as the cost of compute has come down and a number of foundation models have been released openly (primarily around protein folding prediction). This shows up in Benchling's annual "State of Tech in Biopharma" survey: in the 2023 edition, AI/ML was at ~59% adoption, trailed R&D data platforms and robotics/automation for adoption rates, and was third in investment priority (less than 40% of R&D respondents labeled it as "important or very important" for next-12-months' investment, last of the survey options). Only a quarter of respondents at large companies said they use AI/ML regularly, which was double the rate at smaller companies. In 2024, adoption moved to 67% at big pharma (34% "use regularly) and 23% at small biopharma (9% use regularly) and AI became the second-highest investment priority.

The 2026 version of Benchling's report narrowed its focus to about 100 biopharma companies actively using AI in R&D. We assume most/all of the top 20 pharma falls into that camp. These more nuanced survey results speak to the pace at which AI is being adopted for relevant workload. Bristol Myers's SVP IT R&D sums up the adoption curve well: "At BMS, AI is already supporting nearly every facet of our work."

Exhibit 10
AI Adoption by Use-Case



Sources: William Blair Equity Research, Benchling

Half of biopharma organizations regularly using AI have already seen "faster science," while roughly one-third have recognized better outcomes (e.g., better hit rates) and one-quarter have seen lower costs (e.g., fewer failed experiments).

Biological Foundation Models – In the Dry Lab, Data Is the Limiting Reagent

Investor debate regarding AI in drug discovery primarily centers on the extent to which AI can reduce utilization of certain categories of wet lab work by using predictive algorithms to narrow what needs to be experimentally tested. But there is a distinct way in which AI can *increase* wet lab demand: the deliberate generation of large, standardized biological datasets to build biological foundation models and the adaptation of those models to useful drug discovery engines. In this use-case, rather than being displaced, wet lab work is being done at scale to power the models.

A *biological foundation model* is a large-scale AI model pretrained on large biological datasets to learn reusable representations of biology that can be adapted to various downstream tasks. Models are purpose-built to address specific goals—for example, understanding disease biology (what is happening), identifying targets (what should be modulated), understanding proteins (what a protein looks like or does), or designing drugs.

From the outset it is important to reiterate that biology is a compelling target for "foundation model" approaches because core biological code (DNA → RNA → protein) is shared across organisms and because evolution generates immense natural variation that can, in theory, be learned as statistical structure. In comparison to natural language models where "tokens" more or less represent words, biological models operate on different building blocks. In protein models, basic units are amino acids; for virtual cell models, they are measurements of gene expression or other molecular features of a cell.

However, the analogy to natural language has limits: biological data are not just symbols that can be rearranged, they are experimental measurements that were produced under specific laboratory conditions. As a result, model performance is often limited less by computing power and more by the underlying data: whether experiments span enough biological diversity, are run with consistent protocols, produce reliable labels or phenotypes, and preserve meaningful biological context.

This creates a fundamental difference from language models. Large language models can be trained in a fully self-supervised way (i.e., predicting the next word provides its own training signal). In biology, however, the most valuable questions ("Does this protein fold?", "Does this genetic perturbation change cell state?", "Does this compound kill the cell?") generally require an experimental outcome generated in the lab. That experimental dependency makes data generation slower, more expensive, and a core bottleneck in scaling biological models.

Protein sequence data can be learned in a largely self-supervised way (e.g., models like ESM are trained by masking and predicting amino acids, similar to how language models learn from text), but sequence alone does not capture biological function. Determining whether a protein folds correctly, performs a specific function, or affects cells in a meaningful way typically requires direct wet-lab measurement. As a result, unlike internet-scale text data, much of the most valuable biological training signal must be generated through physical experiments like sequencing runs, perturbation-response datasets (e.g., how cells change when a gene is knocked down or a drug is applied), structural measurements, and functional assays.

Depending on the application, biological models may draw on many different data types, including DNA sequences, RNA and gene expression data, protein sequences and structures, chemical structures, cellular response profiles, and biological images such as microscopy or histology. These data categories differ fundamentally in how much data already exists, how expensive new data is to generate, and how much biological insight each experiment provides. Understanding these differences is critical for assessing which models can scale primarily on existing data and which depend on continued investment in new wet-lab experiments.

Protein language models (PLMs, such as ESM), *structure prediction models* (such as AlphaFold), and newer *generative protein design models* (such as RFdiffusion and Chai-2) are trained primarily on large, existing public datasets and therefore do not require significant new wet-lab work to scale. While these approaches differ in objective (sequence modeling, structure prediction, or de novo protein design), they all learn predominantly from accumulated public sequence and structure data (in repositories like UniProt and the Protein Data Bank) rather than newly generated experimental labels.

In contrast, models that aim to *predict protein fitness* or *biological function* require direct experimental input. For example, the ProteinGym benchmark aggregates deep mutational scanning datasets, which depend on extensive wet-lab workflows including gene synthesis, protein expression or display, functional selection, and next-generation sequencing. Similarly, cell-state and perturbation-response models (e.g., Chan Zuckerberg Initiative's virtual cell effort) require large-scale generation of new experimental data, including single-cell sequencing, CRISPR perturbation libraries, imaging, and other high-content assays.

A related but distinct category includes large *foundational datasets* or *response maps,* such as Connectivity Map (CMap), LINCS, and JUMP-CP. These projects are not models themselves but standardized experimental resources designed to systematically measure how cells respond to genetic or chemical perturbations, and they often serve as training data for downstream models.

While protein language and structure models have progressed rapidly given they do not require new experimental data, most draw from the same core public databases (e.g., UniProt, UniRef, PDB, GenBank). These datasets were assembled over decades from individual experiments conducted for diverse biological purposes, not explicitly designed as machine-learning training sets with standardized protocols, materials, and controls.

This creates two structural limitations, characterized well in a white paper from Basecamp Research in June 2025. First, the data often lacks the diversity, consistency, labeling quality, and experimental context required for next-generation biological models, leading to bias and noise. Second, because many groups rely on the same underlying public resources, incremental gains from scaling models on these datasets have diminished since the early 2020s. In effect, much of the field has encountered a "data wall," where further progress is increasingly constrained by access to new, well-curated, experimentally grounded biological data rather than by modeling techniques alone. Importantly, this "data wall" is not just about dataset size; it is also about whether the data are paired, contextual, and experimentally consistent – a problem being tackled by the field through the introduction of new instruments (e.g., 10x Genomics, Bruker/NanoString, Element Biosciences) that enable multi-omic analysis of the same cell.

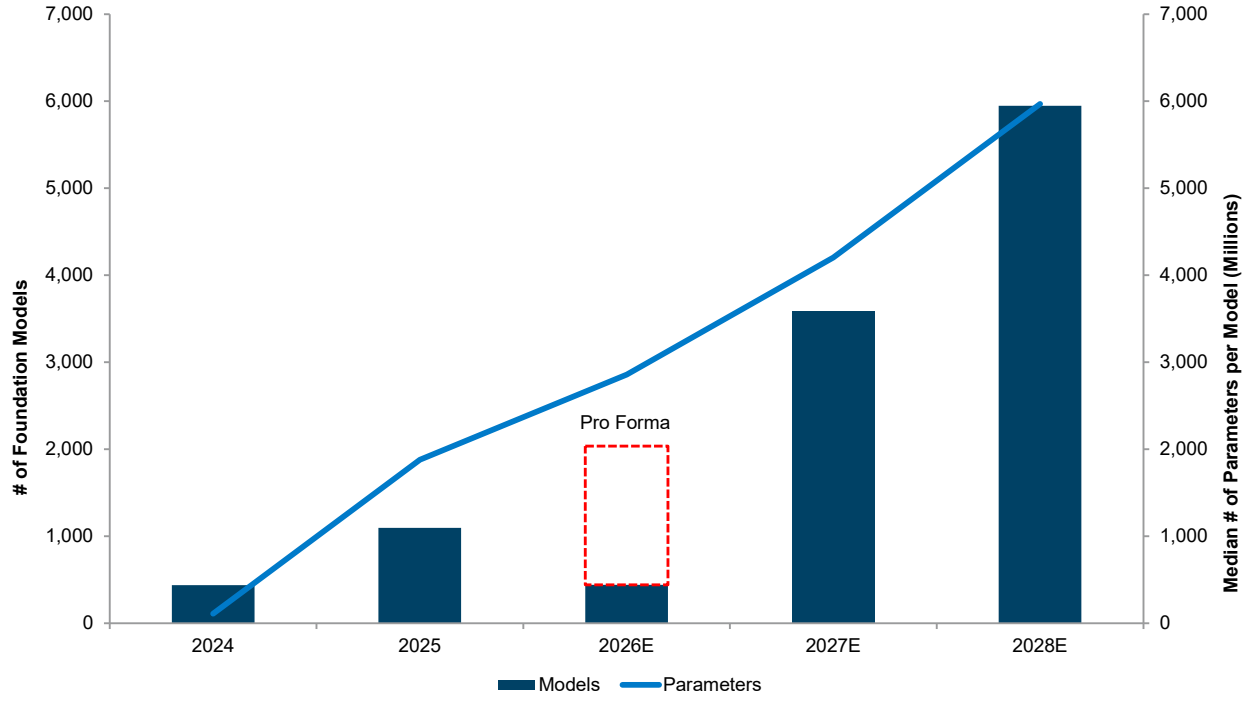
As a result, a visible wave of investment is forming around industrial-scale biological data generation with the explicit goal of training next-generation models rather than repurposing legacy datasets. While still early, these efforts are accelerating as groups attempt to break through the "data wall" limiting current biological foundation models.

In parallel, there has been a rapid increase in investment around *cellstate* and *perturbation response modeling,* where new data must be generated de novo. For example, the Chan Zuckerberg Initiative's Billion Cells Project (announced February 2025) is a multiyear effort to produce a single-cell dataset of 1 billion cells designed explicitly to support next-generation biological models, in partnership with technology providers such as 10x Genomics and Ultima Genomics. Similarly, Illumina and Broad Clinical Labs announced a collaboration focused on rapidly scaling single-cell workflows, with the long-term ambition of enabling multibillion-cell atlas scale datasets. Arc Institute is building a Virtual Cell Atlas with at least 300 million cells, utilizing 10x for single-cell and Ultima Genomics for NGS.

Commercial efforts are emerging as well: Xaira recently introduced X Cell, its first virtual cell model, trained on X-Atlas/Pisces, a genome-wide CRISPRi Perturb-seq dataset (using 10x for single-cell capture and Illumina for NGS) comprising 25.6 million perturbed single-cell transcriptomes across seven biologically diverse cellular contexts—one of the largest functional genomics datasets disclosed to date. Owkin's MOSAIC project is using 10x both for single-cell and spatial to build a massive pan-cancer spatial transcriptomics dataset to enable AI-based therapeutic prediction modeling and identification of putative drug targets. Most recently, Bioptimus announced STELA—in partnership with 10x (Xenium) and Broad, Bioptimus intends to profile up to 100,000 patient specimens (per Bioptimus, a ~20-fold increase in scale over existing spatial biology atlases) to scale its M-Optimus model.

Taken together, these examples illustrate an ongoing inflection in wet-lab data demand, as both the number of large-scale biological models and the volume of experimentally generated training data increase. As shown in the chart below (our analysis), both the number of published models and average wet lab inputs per model are rising sharply, underscoring that progress in many areas of biological AI is now constrained less by algorithms and more by the capacity to generate high-quality experimental data at scale.

Exhibit 11
Number of Biologic Foundation Models Published and Median Number of Parameters per Model



Sources: William Blair Equity Research, BioRxiv, Epoch.ai

Where the Wet Lab Fits In

There are two distinct sources of wet lab demand tied to "foundation-model" biology. The first is the wet lab required to create (or expand) the training data for certain model categories. The second is the wet lab required to turn a general model into a practical drug discovery engine, where companies run closed-loop DBTL in a specific disease and assay context.

First, wet lab demand for models themselves (i.e., training data)

Most foundation-model efforts in biology fall into two broad use-cases: protein-centric models (used heavily in hit/lead design and optimization, steps 3-8 in the antibody discovery workflow later in this report) and cell-state/perturbation-response models (used heavily in target discovery/validation, steps 1-2 in the workflow below). Within these, wet-lab requirements vary widely.

- *Protein language models (PLMs)* are large neural networks trained on protein sequences using self-supervised objectives (masked-token prediction and/or next-token prediction). The ESM family is a leading example, and later generations extend beyond sequence alone. These models are used primarily to help researchers decide *what* to make (i.e., propose sequences or guide which variants to explore; steps 1-3 below). In general, they can be trained largely on existing public sequence databases, so incremental wet-lab spending to train them is limited; the main **winners** here are compute and data infrastructure (e.g., Nvidia and the broader GPU/cloud ecosystem)

- *Structure prediction models* infer 3D structures from sequences (some recent models also evaluate interactions, like protein-protein or protein-ligand). Most commonly known are AlphaFold and Boltz. These models are leveraged primarily in steps 1-3 by injecting structural context into predictions—they help researchers predict the characteristics of what they make.

Like PLMs, these are primarily trained on existing structural databases and derived resources, so incremental wet-lab demand to train them is limited and the main **winners** here are compute and data infrastructure (e.g., Nvidia and the broader GPU/cloud ecosystem).

- *Generative protein design models* (e.g., RFdiffusion, Chai-2) sit between "PLMs" and "structure" in practical workflow terms: they propose new proteins (new sequences/structures) under design constraints. They typically scale on existing sequence/structure data, so they do not inherently require major new wet-lab data to train. The main **winners** here are compute and data infrastructure (e.g., Nvidia and the broader GPU/cloud ecosystem), though they do drive downstream wet-lab validation because they increase the number of design ideas for testing.

- *Protein fitness/function prediction models* measure how amino acid substitutions affect physical protein characteristics like binding/kinetics, stability, activity, and expression. The most well-known reference set is ProteinGym. These models are relevant for steps 3-8 as they help researchers predict what hypothetical sequences/structures will actually do in the physical world. They rely on drug mutational screening (DMS) assays, which involve building large custom variant libraries, selecting/screening for function, and sequencing. The wet lab is unavoidable here, though **winners** vary by step—for variant libraries (biggest force multiplier), primarily Twist (speed, cost, scale) but also Danaher (IDT); for functional screening/selection (less meaningful), Bio-Techne (target protein for binding selection), Carterra, Danaher (Beckman Coulter), and Waters (BD Bio) for screening and sorting; for NGS readout, primarily Illumina but also Ultima Genomics and Element Biosciences; and for NGS reagents, Illumina, New England Biolabs, and others.

- *Cell-state and perturbation-response models* are trained on how cells respond to perturbations (e.g., genetic knockdown/knockout, drug treatment), most commonly by imaging phenomics (high-content screening) or single-cell omics (primarily perturb-seq). These models are leveraged in steps 1 and 2—they help researchers understand the consequences of a target being perturbed. Examples include GEARS (predicting transcriptional outcomes of perturbations), scGPT/Geneformer-style foundation models trained on very large single-cell datasets, scGPT-spatial (trained on spatial transcriptomics data), RxRx3 (trained on imaging-based phenomics) and large public foundational response maps/datasets like CMap/LINCS and Jump-C. **Winners** differ by category of data. For imaging phenomics, cell painting is the standard approach: Revvity (largest exposure) and Danaher (Molecular Devices) stand out. Consumables (kits and plates) are low-cost/not needle-moving. For perturb-seq, 10x is most exposed. For spatial transcriptomics, 10x, Bruker, and Vizgen. Illumina, Ultima Genomics and Element Biosciences (readout) and Danaher/IDT and Twist (CRISPR libraries) also benefit. *Sequencing demand is significant, so this is likely the single largest driver of wet lab demand related to foundational models. Sequencing economics in these workflows are increasingly on a "per-cell" basis, which should favor systems that can deliver scale, simplicity, and value.*

In sum: on foundational models themselves, we view 10x, Twist, and Illumina as having the highest single-company exposure. Danaher, Revvity, Thermo, Techne, and Waters also participate, as do private companies like Carterra, Element Biosciences, New England Biolabs, and Ultima Genomics.

Second, on the transition from model to drug discovery engine

Biological foundation models themselves are starting points—capability blueprints that must be fed with more directed experimental data to have utility in drug discovery. Importantly, this data must be obtained with 1) context (e.g., cell type, disease state) and 2) labels (e.g., outcomes like binding, potency, toxicity), and be collected in an unbiased, auditable manner from a closed-loop system.

This transition is highlighted by the EVOLVEpro framework: "zero-shot" use of pretrained models often fails to reliably improve protein activity, while iterative cycles with small amounts of experimental data can drive meaningful functional gain. Commercial platforms are increasingly built around this idea as well. Absci explicitly describes its approach as wet lab plus AI in an iterative loop, supported by a large internal wet lab designed to generate proprietary training data and validate designs. Recursion describes an analogous industrialized model centered on automated experimentation at a very large scale (including public statements of ~2.2 million experiments per week), with model development tightly linked to that experimental throughput.

To turn a biological "model" into a "drug discovery engine," three things are necessary: context, labels (for example, a measured cell response to a perturbation), and a closed-loop system (with wet lab work as the "test" in a hypothesize-test-learn loop). The scale and durability of wet lab demand is dependent on how that transition is brought about:

Integration Into Existing Antibody Discovery Workflows

- ***PLM -> protein engineering engine:*** PLMs are increasingly used as active learning tools within directed evolution workflows, particularly for antibody optimization against a defined target. Functionally, this represents an evolution of late-stage lead optimization (analogous to step 8 in the workflow), which in practice is not a one-time gate but a repeated loop with re-expression (step 4) and confirmation of improved characteristics (steps 5-7). In a typical implementation, an initial antibody sequence is used to prompt a sequence model to propose a small set (10-15) of variants. These variants are synthesized, expressed, and tested for binding or other relevant properties. Experimental results are then fed back into the model, which proposes the next round of variants. This cycle is repeated until performance plateaus and an optimized antibody emerges. From a wet-lab demand perspective, volumes per iteration are modest, but iteration frequency typically increases. Demand is concentrated in gene synthesis (Twist, IDT), protein expression reagents (Thermo Fisher), and small amounts of target antigen (Bio-Techne). Depending on the readout, assay winners could be Danaher/Cytiva, Carterra (if SPR), Techne (if ELISA), Waters/BD, or Revvity (if flow). When adopted, this tool likely yields fewer variants per cycle, but more cycles—ultimately, DNA synthesis demand may be similar, with higher assay utilization.

An Adaptation Layer From Model to Engine

- ***Structure prediction tuning*** involves companies adapting general structure or complex prediction models by layering proprietary structure, binding, or biophysical data on top of pretrained systems. This activity primarily supports early hit finding and lead design and is largely a dry lab exercise. Incremental wet-lab demand is episodic rather than continuous and arises mainly when proprietary structural data must be generated, typically via Cryo EM or biophysical binding assays such as SPR.

- ***Virtual cell model adaptation*** involves retraining or finetuning general cell-state or perturbation-response models on new, disease- and context-specific experimental data. The goal is to convert a broad foundation into a decision-grade engine, pushing the models upstream into steps 1-2 (target ID and validation). Training data generation typically includes pooled CRISPR perturbations, large-scale single-cell RNA sequencing, and increasingly spatial transcriptomics or high-content phenomics. As a result, wet-lab demand spans CRISPR libraries and gene synthesis (e.g., Twist, Danaher/IDT), single-cell capture (10x), sequencing (Illumina, Ultima Genomics, Element Biosciences), spatial (10x, Bruker, Vizgen), and imaging (Revvity, Danaher/Molecular Devices). Because economics scale per cell and per perturbation, this category produces the largest and most durable wet-lab demand associated with foundation-model biology.

From our diligence across the space (conversations with big pharma, tools vendors, and model developers), we expect both foundation models themselves and their adaptation layers to grow meaningfully in the coming years. As with natural language processing, progress is increasingly constrained not by algorithms but by access to high-quality, standardized, fit-for-purpose experimental data. For tools companies positioned at those data bottlenecks, we view the resulting demand as structural rather than transient.

How AI Is Impacting Antibody Discovery – Smarter Funnels and Faster Cycles

Think of the drug discovery process in three phases: 1) mapping a disease pathway and identifying a control point that can change the downstream biology by being turned off and on or dialed up and down, 2) designing and generating a molecule to effect that change, 3) providing evidence it is likely to be safe to use in humans and can be reproducibly manufactured. In the first two phases, identification is followed by validation, so in essence most of the steps we discuss in the sections below can be lumped into four buckets: target ID, target validation, hit ID, and hit validation.

Within each phase there are multiple distinct steps doing some version of designing, building, testing, and learning (the DBTL cycle). In the sections below, we walk through the discovery workflow and identify which steps are most likely to be disrupted by AI (learning and designing) and which require the wet lab (building and testing). For simplicity, our focus here is on large molecules (e.g., antibodies); while distinct workflows exist for small molecules (and new modalities like cell and gene therapies), the process of finding and validating a target, discovering and designing a drug for that target, and generating evidence that it works are similar throughout.

Regardless of modality, we view the biggest impact of AI as identifying disease targets, predicting/generating molecules to affect those targets, and more efficiently eliminating candidates that are unlikely to have safety and efficacy profiles that translate to in-human biology. Where AI has more limited utility is in the physical testing and manufacturing process development steps, areas that today represent roughly three-fourths of preclinical spend per campaign.

To start an in-human Phase I clinical trial, an antibody program must:

1) pick and validate a target,

2) discover/generate binders to that target and pick leads

3) prove they bind, work in functional biology, and can behave like a drug (developability),

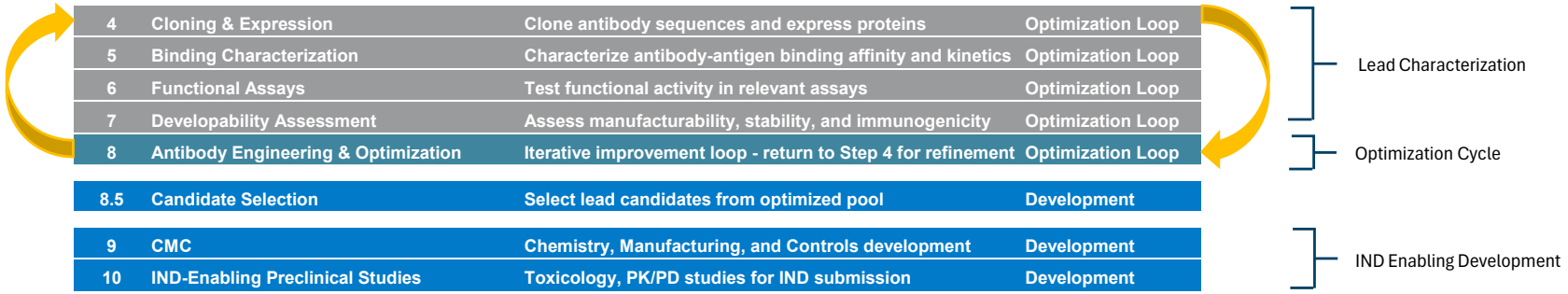
4) optimize and filter candidates until a clear lead emerges for nomination

5) run two parallel IND-enabling tracks—chemistry, manufacturing, and controls (CMC) (how you make and control the drug, including GMP Phase I supply) and nonclinical safety (to support the IND).

Exhibit 12

Antibody Discovery Workflow

Step	Name	Description	Phase
1	Target ID	Identification and selection of therapeutic target	Discovery
2	Target Validation	Validation of target relevance and druggability	Discovery
2.5	Antigen Assay & Preparation	Prepare antigens and develop screening assays	Discovery
3	Hit ID, Antibody Design & Primary Screening	Generate antibody hits and conduct initial screening	Discovery
4	Cloning & Expression	Clone antibody sequences and express proteins	Optimization Loop
5	Binding Characterization	Characterize antibody-antigen binding affinity and kinetics	Optimization Loop
6	Functional Assays	Test functional activity in relevant assays	Optimization Loop
7	Developability Assessment	Assess manufacturability, stability, and immunogenicity	Optimization Loop
8	Antibody Engineering & Optimization	Iterative improvement loop - return to Step 4 for refinement	Optimization Loop
8.5	Candidate Selection	Select lead candidates from optimized pool	Development
9	CMC	Chemistry, Manufacturing, and Controls development	Development
10	IND-Enabling Preclinical Studies	Toxicology, PK/PD studies for IND submission	Development



Sources: William Blair Equity Research

We go through these steps in more detail below, describing what occurs at each step, what tools are used to accomplish those tasks, public company participation in each step, and how AI is disrupting or may disrupt that work in the future. Note that the cost ranges provided are our estimates. There are no established benchmarks; however, our estimates are informed by a variety of published studies on the topic.

Target Selection (Steps 1-2)

Step 1 – Target Identification

What It Is
Researchers identify and prioritize a shortlist of disease-relevant targets (e.g., proteins, genes, and RNA) that are: a) plausibly disease-driving based on human evidence (e.g., genetic association, expression, and clinical observations), b) druggable by an antibody (i.e., with extracellular or cell-surface accessibility, favorable epitope geometry), and c) associated with an acceptable predicted safety profile (i.e., minimal expression in critical healthy tissues and a low likelihood of on- or off-target toxicity).

Legacy Approach
This method includes profiling gene and protein expression in diseased versus healthy tissues using 1) microarrays (historically the gold standard, though now in decline), RNA-seq, or single-cell RNA-seq (with qPCR for validation); 2) proteomics (e.g., mass spectrometry) to confirm protein presence, abundance, and PTMs; 3) spatial omics to localize target expression to specific cell types and tissue microenvironments; 4) genetic association studies, including genome-wide association studies (GWAS) and rare-variant databases; and 5) manual literature review. Note this step has largely shifted to a two-phase approach with dry lab first, followed by selective use of wet lab to fill gaps (e.g., single-cell RNA seq to confirm cell-type-specific expression in a disease tissue, spatial proteomics to confirm location of target, and mass spec to confirm protein quantification*). Wallet share estimate: ~11% of discovery (pre-CMC) spending, ~6% of total preclinical spending, including heavy dry lab spend.*

How AI Is Being Used
Using AI compresses the timeline to target by automating manual literature and evidence review, improving target nomination and prioritization, and predicting protein structure and in-human toxicology. In the future, we see manual literature review and broad exploratory screening continuing to move out of favor, with a hybrid approach (in silico selection followed by focused wet lab to confirm cell-type-specific expression via scRNA-seq, spatial omics for tissue localization, MS-based chemical proteomics for druggability/accessibility) becoming industry standard. Significant resources will be deployed to build and adapt foundational biological models to understand disease biology and identify targets.

Exhibit 13
Step 1 - Target Identification

Current Market Leaders	AI Adoption		Winners	Losers
	Current	Residual Risk		
ThermoFisher SCIENTIFIC 10X GENOMICS BRUKER Waters BD illumina	**VERY HIGH** ML models built on knowledge graphs help prioritize targets with high druggability scores, favorable assessments from predictive toxicology and structure (e.g., epitope accessibility)	**LOW** Per Benchling, big pharma adoption in mid-to-late innings (60%-80%); biotech far earlier and will broaden to the point where industry standard is in silico target ID	Model generation/adaptation: * Single-cell (TXG, QGEN/Parse) * Spatial (TXG, BRKR, Vizgen) * NGS (ILMN, Ultima, Element) * Mass spec (TMO, BRKR) * Variant libraries (TWST, DHR) * Organoids: TECH, STEMCELL * Software: Benchling, UniteLabs, RVTY (Signals), DHR (IDBS), Siemens (Dotmatics) * Models: Isomorphic Labs, Cradle EvolutionaryScale, Noetik, Owkin, Terray, Boltz, AlphaFold	Manual knowledge acquisition: * Service providers focused on manual literature review, raw data bioinformatics * Microarray

Sources: Company websites, William Blair Equity Research

Step 2 – Target Validation

What It Is
Target validation confirms that the target drives the disease (directionally and magnitude) by demonstrating that target perturbation (e.g., genetic knockdown, knockout, and CRISPR; pharmacologic and antibody blockade; and chemical probes) produces the expected phenotypic change in relevant human systems.

Legacy Approach
The legacy method starts with loss-of-function or gain-of-function studies via siRNA/shRNA knockdown and CRISPR (KO, CRISPRi/a), followed by orthogonal expression profiling via tissue readouts like immunohistochemistry (IHC)/immunofluorescence (IF) , Western blot and capillary electrophoresis (protein expression and signaling confirmation), ELISAs (protein quantification and cytokine profiling), and flow cytometry (multi-parameter phenotyping). Next steps are functional assays relevant to mechanism and molecule type (e.g., calcium flux for GPCRs, ADCC/CDC reporter assays if IO target), proof-of-mechanism studies via animal models or in vitro human-like systems, and chemical probes to mimic pharmacologic perturbation. *Wallet share estimate: ~9% of discovery (pre-CMC), ~5% of total preclinical.*

How AI Is Being Used
In this step, AI is being used to prioritize validation experiments (thus reducing experiments on low-probability targets), automate analysis of multimodal readouts, and improve throughput. AI can help prioritize targets and predict druggability, but validation will still require wet lab work. In other words, researchers must demonstrate that perturbing a target does indeed produce an expected phenotypic outcome. In short, we expect stable demand for perturbation tools (e.g., CRISPR reagents, siRNA/antisense reagents) and validation tools (e.g., flow cytometry, Western blot and ELISA along with related reagents like antibodies), increasing demand for animal model alternatives (organoids, organ-on-a-chip) and complementary analytical techniques like high-content screening (Revvity, Danaher/Molecular Devices), 3D spatial (e.g. Vizgen's MERSCOPE Ultra), and lower demand (over time) for in vivo animal studies.

Exhibit 14
Step 2 - Target Validation

Current Market Leaders	AI Adoption		Winners	Losers
	Current	Residual Risk		
	**MEDIUM** More efficient filtering from Step 1 to reduce wasted experiments on low-probability targets	**LOW** Lab-in-a-loop / autonomous labs will be leveraged to speed up the validation cycle, but we believe orthogonal wet lab data will always be required in target validation.	* Perturbation: TMO, DHR (IDT), RVTY * Validation: WAT, DHR, RVTY, SRT, CTKB * Organoids: TECH, STEMCELL, EmulateBio * Data infrastructure: Benchling, Sapio Sciences, RVTY (Signals)	* In vivo CROs: CRL, LH (Covance), Inotiv, Jackson Labs *--> Expect impact over 5-10 years, not at once*

Sources: Company websites, William Blair Equity Research

Step 2.5 – Antigen and Assay Preparation (Screening Preparation)

What It Is
Produce screenable antigens and build, optimize, and validate the assay stack (primary binding, counter screens, and orthogonal confirmation) to power hit identification at step 3.

Legacy Approach
The legacy method uses antigens expressed in CHO or HEK cells, ELISA (bind/no-bind), followed by SPR/BLI (binding kinetics), and limited counter-screening. *Wallet share estimate: ~1%-2% of discovery (pre-CMC), ~1% of total preclinical.*

How AI Is Being Used
AI models are used to 1) predict characteristics of antigen constructs (e.g., optimal domain boundaries, epitope exposure, stability), reducing the number of variants that need to be tested; 2) optimize assay conditions to compress design and optimization timeline; and 3) recommend appropriate counter-screens based on likely artifacts (a false signal that makes an antibody look like a hit even though it is not binding in the right way, or at all). In the future, the antigen pool per program continues to shrink, reducing assay volume needed for validation. Bind/no-bind assays lose share due to AI influence on antigen design and growing adoption of kinetics as the first measurement.

Exhibit 15

Step 2.5 - Antigen and Assay Preparation

Current Market Leaders	AI Adoption		Winners	Losers
	Current	Residual Risk		
biotechne, ThermoFisher SCIENTIFIC, revvity, MSD, Agilent, danaher, carterra, GenScript, Sino Biological	**MEDIUM** Antigen construction is a physical process, but AI is actively being used in antigen and assay design, reducing trial-and-error	**MEDIUM** Less demand per program for antigen construction and low-information screens, with analytical share shift to higher-information assays.	* Antigen production: TECH (antigens), TMO (expression systems) * Assays: DHR (Cytiva), RVTY (AlphaLISA and HTRF), Carterra	* Share shift away from ELISA (BIO, TECH, RVTY)

Sources: Company websites, William Blair Equity Research

Lead Discovery (Steps 3-4)

Step 3 – Hit Identification, Antibody Design, and Primary Screening

What It Is

Starting from a validated target (from step 2), generate a list of candidate antibodies and identify "hits" (i.e., sequences that bind the target with minimum specificity in primary screens) that will later be down-selected.

Legacy Approach

Historically, the standard approach used in vivo immunization campaigns (often mice), followed by B-cell isolation, cloning, and sequencing. The other approach is in vitro, using phage/yeast display libraries, followed by panning or FACS enrichment, then conducting primary binding screens (e.g., ELISA for antibody-antigen binding, flow cytometry to detect binding on cell surfaces, SPR/BLI to assess kinetics and affinity, sequencing). Typically, simple mechanism-relevant reporter assays are used so binders with no functional effect are not advanced. There has been some adoption of single-cell screening (e.g., functional screening with Bruker's Beacon platform acquired from Berkley Lights, 10x Genomics' Chromium V[D]J immune repertoire sequencing). *Wallet share estimate: ~14% of discovery (pre-CMC), ~7% of total preclinical.*

How AI Is Being Used

This is one of the points of maximal disruption—this step is shifting from "search and sort" to "design and print"; though it is fair to say that AI has already become widely adopted here. AI proposes thousands to millions of antibody sequences (these could be generative or motif-guided or structure-guided variants), then algorithmically ranks a subset of sequences that are synthesized, expressed, and screened (e.g., binding, specificity, basic functional assessment). Thus, demand for gene synthesis, target protein, and initial binding screens will increase as AI generates more (from thousands to millions) of optimized variants (e.g., CDR tweaks for affinity, stability, or bispecific formats) that need to be synthesized and empirically tested. Assay results are fed back into AI models, creating new predictions, and more demand. In some cases, models work in tandem with sequencing immune repertoires from immunization campaigns.

Exhibit 16

Step 3 - Hit ID, Antibody Design and Primary Screening

Current Market Leaders	AI Adoption		Winners	Losers
	Current	Residual Risk		
GenScript ADIMAB TWIST BRUKER 10X GENOMICS	**MEDIUM-HIGH** AI design/rank + wet lab validation is the dominant approach for pharma, and collabs with AI natives (e.g., AZN/ABSI, ABBV/BigHat) are commonplace. AI natives (e.g., Xaira, Nabla) are making progress on de novo design, and both ABSI and recent IPO GENB have candidates in trials.	**MIXED** Demand is primarily shifted, not destroyed, as AI impacts antibody design. AI will reduce brute-force screening, manual hit list review, and some in vivo campaigns. But AI is a tailwind to targeted gene synthesis, primary binding screening assay volumes, and recombinant target proteins.	* Gene synthesis: TWST, DHR (IDT), GenScript * Target protein: TECH * Binding screens: TECH, DHR, Carterra RVTY, WAT, TMO * Repertoire profiling: TXG, ILMN Lab data management: Benchling, RVTY (Signals)	* In vivo immunization CROs, hybridoma reagents and consumables like PEG (TMO, MilliporeSigma) * In vitro: colony-style pickers like Molecular Devices (DHR)

Sources: Company websites, William Blair Equity Research

Step 4 – Cloning and Expression

What It Is

Selected hits (variable region sequences) are converted into physical antibodies by 1) obtaining DNA (fragments or full genes), 2) cloning into expression vectors (typically full-length IgG), 3) expressing in mammalian cells (typically CHO or HEK293), and 4) purifying the antibody for downstream testing. Importantly, physical production of purified antibody is necessary to test its properties.

Legacy Approach

VH/VL variable-region sequences are obtained as synthetic DNA (either as fragments requiring assembly/cloning into mammalian expression vectors or as full-length genes pre-cloned into expression-ready vectors). These sequences are then amplified in bacterial cultures to produce enough plasmid DNA for transfection. That plasmid DNA is delivered into mammalian cells (using reagents or electroporation), and the cells read the gene and manufacture the antibody in large volumes. Antibodies are harvested (centrifugation or filtration), and then purified (Protein A affinity capture), and concentrated. Lastly, QC is performed for concentration (using UV absorbance), purity (typically SDS-PAGE for assembly and HPLC for aggregation), and initial binding (often BLI). ***This step reads like a methods section for a reason—it is "wet lab heavy" and will remain so, given the need to produce purified antibody.*** *Wallet share estimate: ~2% of discovery (pre-CMC), ~1% of total preclinical.*

How AI Is Being Used

AI is primarily used in the transition from step 3 to step 4, by triaging hits based on predicted characteristics, including: expression yield, folding, solubility, aggregation propensity, developability, etc. In the future, AI triaging will become more broadly adopted, and "make-lists" will improve predictability and eliminate a higher percentage of candidates prior to entering the physical make/test steps (that begin with step 4). However, the crucial caveat is that AI campaigns typically generate thousands of candidates versus traditional campaigns generating tens of candidates, so the number of candidates physically made is likely to be stable or growing.

Exhibit 17

Step 4 - Cloning and Expression

Current Market Leaders	AI Adoption		Winners	Losers
	Current	Residual Risk		
	**MIXED** Large pharma and AI-native drug discovery companies leverage AI to create more targeted make-lists, though AI is less adopted in smaller biotech and academia.	**LOW** While AI adoption will continue to broaden, it will not eliminate the need to make (cloning, transfection, and purification) and QC candidate antibodies.	* Gene synthesis: TWST (meaningful advantage due to speed and scale), DHR (IDT) * Vector assembly/cloning and plasmid prep: QGEN, TMO, TWST (service) * Expression: TMO, Corning * Purification: DHR, RGEN, TMO Initial QC: SRT, A, WAT	* DNA synthesis companies with less-scalable and lower throughput offerings.

Sources: Company websites, William Blair Equity Research

Lead Characterization (Steps 5-7)

Step 5 – Binding Characterization

What It Is

Starting with purified antibodies from step 4, quantitative, decision-grade datasets for each candidate are generated: confirm binding to target antigen, measure affinity/kinetics, rank off-rates, assess specificity, and map epitope diversity.

Legacy Approach

ELISA is used for first-pass, semi-quantitative yes-or-no binding confirmation of initial candidates. Top candidates run on SPR or BLI (higher throughput) for kinetics, followed by epitope binning and specificity counter-screening. Recombinant target antigen is required for each test as the binding partner to candidate antibodies. Carterra's high-throughput SPR (HT-SPR) is gaining traction as an alternative to ELISA and SPR/BLI steps. Meso Scale Discovery's (MSD's) electrochemiluminescence (ECL) platform can be used for complex binding formats. *Wallet share estimate: ~7% of discovery (pre-CMC), ~3%-4% of total preclinical.*

How AI Is Being Used

Like step 4 (and subsequent steps), AI use is more operational than transformational. It is primarily to improve and expedite analysis of vast quantities of data generated in binding experiments. Predictive affinity models exist and will improve, though these complement (do not replace) empirical wet lab measurement. In the future, high-throughput binding technologies may replace first-pass ELISA and work well with the potentially higher-volume pool of candidates entering this step from AI-driven discovery programs.

Exhibit 18
Step 5 - Binding Characterization

Current Market Leaders	AI Adoption		Winners	Losers
	Current	Residual Risk		
bio-techne ThermoFisher SCIENTIFIC The Jackson Laboratory BIO-RAD revvity MSD danaher carterra	**LOW** Used as a tool to speed analysis and optimize/triage experiments rather than a replacement for experiments.	**LOW** Physical binding data will always be required to advance antibody programs (and from a regulatory perspective), but it is possible that predictive affinity models improve to the point where fewer candidates enter physical testing.	* Reagents: TECH * Instruments: Carterra, DHR, RVTY(AlphaLISA, HTRF) ** MSD, WAT, A winners for complex binding formats (e.g., bispecifics)	* ELISA reagents/kits (multiple companies) and RVTY plate readers benefit but face displacement risk from high-throughput binding instruments.

Sources: Company websites, William Blair Equity Research

Step 6 – Functional Assays

What It Is
Functional assays are used to test candidate antibodies for functional activity—essentially, does binding create the intended biological effect (e.g., receptor blockade, pathway modulation, cell death, immune system engagement) without causing obvious undesirable effects (e.g., nonspecific activation, cytokine storm risk signal, off-pathway activity)?

Legacy Approach
Cell-based functional assays are performed to answer a series of questions about whether the antibody produces the intended phenotype, including: receptor blockade/neutralization assays with Western blot or phospho-flow (does the antibody block downstream signaling?), agonism assays (does the antibody activate the target pathway and induce downstream signaling?), flow cytometry (does the antibody bind?), cytokine release assays (is the immune system engaged?), reporter gene assays (does the antibody trigger the intended signaling pathway?), and viability assays (does the antibody stimulate/inhibit cell division and what fraction of cells did the antibody destroy?). Assay use and volume vary depending on the mechanism of action. *Wallet share estimate: ~28% of discovery (pre-CMC), ~15% of total preclinical.*

How AI Is Being Used
AI is being used primarily to augment instrument capability, particularly for high-content imaging, and optimizing experiment flow. Like binding characterization, while AI can predict functional activity from sequences, regulators will always require physical evidence that an antibody performs its intended function (e.g., kills a cancer cell, blocks a receptor). Higher demand will be created if more antibody candidates are created.

Exhibit 19

Step 6 - Functional Assays

Current Market Leaders	AI Adoption		Winners	Losers
	Current	Residual Risk		
Waters, danaher, ThermoFisher SCIENTIFIC, BIO-RAD, revvity, biotechne, Promega, MSD, Agilent, SARTORIUS	**LOW** High adoption in new instrumentation (e.g., Revvity Opera Phenix OptIQ with Phenologic.AI, Molecular Devices' ImageXpress HCS.ai, Thermo' Attune CytPix), with growing adoption in flow cytometry data analysis.	**LOW** Physical evidence of what an antibody does in cells will always be required at this stage. Possible tailwinds from mix (shift from simple, single-endpoint testing to high-content, multiplexed testing) and greater candidate volume.	* Reagents: TECH, BIO, TMO, RVTY, Promega * Instruments: WAT (BD), RVTY, DHR, TMO, SRT, A	* Manual, singleplex ELISA * Single-color flow cytometry * Legacy plate readers like Revvity EnVision * Endpoint-only viability assays (as kinetic testing provided by Agilent's xCELLigence and Sartorius' Incucyte take more share).

Sources: Company websites, William Blair Equity Research

Step 7 – Developability Assessment

What It Is

This is a multi-parameter assessment of whether an antibody can be manufactured, formulated, and stored as a drug product. In other words, can a candidate be made into a drug?

Legacy Approach

Developability is phase-gated because full characterization of molecules is expensive. ***Phase I*** is high-throughput (100s-1000s of candidates) and includes SEC for aggregation, icIEF (capillary isoeletric focusing)/cSDS for charge profiling, and differential scanning fluorimetry (DSF) for thermal stability (at what temperature will the antibody denature?) and colloidal stability (will antibody molecules aggregate over time?) screening. ***Phase II*** (multi-parameter characterization of 10s-100s of candidates) assesses charge variant analysis (icIEF), size/purity (CE-SDS and SEC/SEC-UPLC), hydrophobicity (hydrophobic interaction chromatography), thermal transitions (differential scanning calorimetry), stability, and polyspecificity assays for off-target binding risk. ***Phase III*** (full assessment of 5-10 candidates) includes forced degradation and stress (e.g., thermal, oxidative, freeze-thaw) studies, formulation screening (e.g., buffer, excipients, pH), viscosity measurement (for subcutaneous feasibility), and subvisible particle analysis. *Wallet share estimate: ~4% of discovery (pre-CMC), ~2% of total preclinical.*

How AI Is Being Used

Phase I is most disrupted, with models flagging high-risk sequences as likely failures, so fewer candidates enter Phase II. Beyond that, AI primarily is used to augment instrument capabilities and better organize and assess the complex array of data this step generates. AI is making progress on in silico developability prediction (e.g., flagging aggregation-prone sequences), which could compress phases by eliminating candidates before physical screening, but ultimately candidates need to be physically assessed for developability before entering IND-enabling studies. Instrument platforms that are automated and high-throughput continue to gain share.

Exhibit 20
Step 7 - Developability Assessment

Current Market Leaders	AI Adoption		Winners	Losers
	Current	Residual Risk		
Waters, Agilent, biotechne, MANTEMPER, Malvern Panalytical, UNCHAINED LABS, danaher	**MEDIUM** Primarily in candidate triage/pre-screening.	**MIXED** Anticipate broader adoption and higher model efficacy leads to fewer candidates flowing through this step, pressuring reagents but not reducing the need for the instrumentation.	* TECH * WAT * A * Private: Malvern Panalytical, NanoTemper	* BIO (manual gel-based protein QC tools)

Sources: Company websites, William Blair Equity Research

Optimization Cycle

Step 8 – Antibody Engineering and Optimization

What It Is
This design–build–test–learn iterative cycle transforms "lead antibodies" into "drug candidates." Antibodies are engineered and redesigned based on the binding, functional, and developability data from steps 5-7. This step includes humanization (reducing immunogenicity of non-human-derived antibodies), affinity/kinetics optimization (often off-rate engineering), Fc/format engineering (effector function tuning, half-life extension, multispecific architecture), expression control, and liability removal (fixing sequence/PTM/developability problems). ***Importantly, this is not a one-time gated step, but rather a loop with steps 4 (reexpression) and 5-7 (confirmation of improved characteristics).***

Legacy Approach
Tools vary by task, and include humanization (CDRs are transplanted from the animal antibody to human framework regions), affinity maturation (gene synthesis of variant libraries, display and selection [phage/yeast + FACS sorting], and kinetic characterization with SPR/BLI), Fc/format engineering, and developability fixes. *Wallet share estimate: ~23% of discovery (pre-CMC), ~12% of total pre-clinical.*

How AI Is Being Used
AI is used upstream in the design process (e.g., de novo or targeted redesign) and to guide library diversity, humanization strategy, and affinity maturation. In the future, existing AI humanization tools will likely improve, helping eliminate low-quality candidates. Hence, volumes will likely lower, particularly for affinity maturation, but physical confirmation will always be required; counterintuitively perhaps, given less budget spent on low-probability candidates and quicker turns of the crank, AI may actually increase the number of optimization cycles per candidate. An important addendum is that complex biologics (e.g., bispecifics) increase the optimization task (at minimum, more cycles) and are likely to continue to grow as a percentage of the pipeline.

Exhibit 21

Step 8 - Antibody Engineering and Optimization

Current Market Leaders	AI Adoption		Winners	Losers
	Current	Residual Risk		
TWIST, danaher, Waters, BD, HAMILTON, TECAN, ADIMAB, WuXi AppTec, ABZENA, GenScript	**MIXED** By definition, high in AI drug discovery companies (e.g., Absci, Xaira), but lower adoption with CROs and legacy pharma.	**MEDIUM** AI reduces variants per cycle, though it's unclear if AI will reduce the number of optimization cycles needed, or counterintuitively increase the number of cycles given timeline compression and reduced budget waste on likely-to-fail candidates.	* Precision variant libraries: TWST * Binding: Carterra, DHR, SRT	* Random/degenerate mutagenesis libraries * CROs focused on manual optimization services/compete on library size * Endpoint-only binding, for example ELISA used with microplate readers from BIO, A, DHR, Tecan

Sources: Company websites, William Blair Equity Research

Step 8.5 – Candidate Selection

What It Is

A clinical candidate is formally nominated (and usually 1-2 backups). This decision relies on data generated in prior steps, but often includes minimal (i.e., confirmatory) wet lab testing itself.

IND-Enabling Development (Parallel Track of Steps 9-10)

Step 9 – Chemistry, Manufacturing, and Controls (CMC)

What It Is

In this step the nominated antibody is turned into a reproducible product, by establishing: 1) a scalable drug substance (DS) manufacturing process; 2) a drug product formulation and container/closure strategy; 3) analytical methods and specs to prove identity, purity, and potency; and 4) stability programs to support clinical use and the IND CMC section. ***Essentially, proving that you can make the exact same drug every time.***

Legacy Approach

Here, the physical production of the drug becomes the work product (rather than knowledge-acquisition and analysis about potential drug candidates). Wet lab work includes cell line development (typically a stable CHO cell line), upstream process development (optimizing bioreactor conditions) and downstream purification (e.g., Protein A capture and polishing chromatography protocols), formulation development (buffer, excipient, pH and concentration optimization), fill/finish (aseptic filling into vials/syringes/pens), analytical method development and qualification (identity, purity, potency, stability-indicating methods), and formal stability programs (real-time and accelerated conditions). This is an instrument-heavy step more thought of as QC and process development than "discovery." *Wallet share estimate: ~36% of total preclinical.*

How AI Is Being Used

None of these wet lab/manufacturing steps can be replaced by AI, but AI is being leveraged to optimize manufacturing and formulation/storage parameters. If AI is successful in compressing discovery timelines and increasing the number of nominations for the clinic, step 9 (and 10) will see a volume tailwind as more candidates enter CMC.

Exhibit 22

Step 9 – Chemistry, Manufacturing and Controls (CMC)

Current Market Leaders	AI Adoption		Winners	Losers
	Current	Residual Risk		
Lonza, ThermoFisher Scientific, Samsung Biologics, Bruker, WuXi AppTec, Fujifilm Diosynth Biotechnologies, Waters, Boehringer Ingelheim BioXcellence, Agilent, Biotechne, Revvity, Danaher, Promega, Sartorius, Biotechne, Revvity, Carterra, Mettler Toledo, Unchained Labs	**LOW** AI being leveraged in process development and design of analytical QC protocol, but CMC is where the physical production of the drug becomes the work product.	**LOW** The product is now physical (the drug) rather than digital (information about potential drugs), reducing disintermediation risk.	* Current market leaders will benefit if AI is successful in creating more drug candidates	* NA

Sources: Company websites, William Blair Equity Research

Step 10 – IND-Enabling Preclinical Studies

What It Is

A variety of studies are conducted to verify the antibody's mechanism of action and efficacy and evaluate potential toxicity risks prior to human studies and to estimate dosing for human studies. Along with CMC data, the IND package must include comprehensive regulatory dataset containing: GLP-compliant safety pharmacology, pharmacodynamics (PD), pharmacokinetics and toxicokinetics (PK/TK), immunogenicity, and toxicology studies.

Legacy Approach

Safety pharmacology focuses on evaluating potential adverse effects on vital organ systems (e.g., cardiovascular, respiratory, CNS), with assays specific to each. PD focuses on binding characterization, functional assays, and in vivo efficacy. PK/TK characterizes ADME (absorption, distribution, metabolism, excretion) profile and aids toxicology findings. Immunogenicity combines in silico prediction and in vitro testing (e.g., ADA testing, neutralizing antibody [Nab] assays). Toxicology historically has occurred in in vivo models with staggered time cutoffs and supporting histopathology. ***Notable FDA NAM policy announcement in April 2025 and $150 million NAM investment announced in March 2026 could shift more volume toward NAMs (e.g., organoid, organ-on-a-chip), though this will be gradual.*** *Wallet share estimate: ~12% of total preclinical.*

How AI Is Being Used

Some adoption in study design (e.g., dose selection, sampling schedules), predictive toxicology, and in silico ADME/PK. Demand for step 10 work is at worst neutral as AI adoption grows upstream and will feel a tailwind if AI gets more candidates into IND-enabling steps (9-10) faster. There is growing momentum for NAMs adoption, and predictive toxicology models (e.g., Tox21, ToxCast, DeepTox, ADMET Predicotr) are improving—both of which would displace or reduce demand for animal models, but this is a multiyear (at best) transition.

Exhibit 23

Step 10 – IND-Enabling Preclinical Studies

Current Market Leaders	AI Adoption		Winners	Losers
	Current	Residual Risk		
	**LOW** Study design (e.g., dose selection, sampling schedules), predictive toxicology, and in silico ADME/PK.	**LOW** Shift to NAMs (if it occurs) would be more in parallel with AI adoption rather than because of it. Animal models seem likely to persist, but number of animals and duration of study required may both compress over time..	* NAMs ecosystem, including companies helping produce NAMs (Bio-Techne, STEMCELL Technologies, Emulate Bio, BioIVT, Sanguine Biosciences, Discovery Life Sciences, CN Bio, Hesperos, MIMETAS) and developing technologies to evaluate them (Revvity, Sartorius); predictive toxicology (e.g., Axiom)	* CROs weighted to animal tox studies

Sources: Company websites, William Blair Equity Research

In aggregate, we see risk to about 10% of the preclinical budget, weighted heavily to design and learn steps as described prior. Importantly, this is tools-related spending (e.g., instruments, consumables, and reagents) that we estimate is 10-15% of total R&D, with considerable savings generated in other categories (e.g., labor and overhead) by compressed timelines. The overarching distinction is ***hypothesis creation versus hypothesis validation***. Work done to identify druggable targets and identify or generate drugs is being augmented by AI, and there are aspects of that workflow that will see reduced wet lab demand in the future. However, validation work will always require physical construction of targets and hits and related wet lab testing to characterize those hits. As of early 2026, no purely de novo AI-designed antibody has entered clinical trials without subsequent lab-based engineering and physical characterization.

Appendix: Private Companies in AI

On the following page is a non-exhaustive list of private companies participating in the AI drug discovery ecosystem.

Exhibit 24

Private Company Lanscape: Frontier Life Sciences Labs

Company	Overview	Total Funding ($Ms)	Latest Round ($Ms)	Key Investors	Year Founded
Arc Institute	Arc Institute is a Palo Alto-based independent nonprofit biomedical research organization cofounded by Patrick Collison (Stripe), Patrick Hsu, and Silvana Konermann. It provides scientists with no-strings-attached, multiyear funding to pursue high-risk, high-reward research into complex diseases including cancer, neurodegeneration, and immune dysfunction. Its technology centers leverage AI, genomics, and multiomics, producing landmark tools such as the Evo biological foundation model and bridge RNA genome editing.	$122M	N/A	N/A	2021
Basecamp Research	Basecamp Research is a London-based AI company that explores biodiversity—collecting genomic data from extreme environments across 28 countries—to build the world's most diverse biological knowledge graph and train its EDEN family of generative AI models. The platform, built on over 10 billion novel protein sequences and a 5-billion-relationship knowledge graph, powers protein and genome design for pharma, food, and industrial biotech.	$84M	$60M (Oct 2024)	Singular, Hummingbird, S32	2019
Boltz PBC	Boltz is an open-science AI research and product company cofounded by MIT researchers who developed the widely adopted Boltz series of biomolecular structure prediction models. The company is a public benefit corporation (PBC) and has secured a multiyear collaboration with Pfizer to build exclusive models for target selection, structure prediction, small-molecule affinity, and biologics design.	$28M	$28M (Jan 2026)	Zetta, a16z, Amplify	2024
Chai Discovery	Chai Discovery is a San Francisco-based AI biologics company founded in 2024 that develops de novo antibody design models. Its flagship Chai-2 model generates full-length, drug-like antibodies computationally, eliminating the need for traditional experimental screening.	$231M	$130M (Dec 2025)	General Catalyst, OpenAI	2024
Cradle Bio	Cradle is a Zurich-based AI protein engineering platform company that helps scientists design and optimize proteins using generative AI and wet-lab feedback loops. Its platform allows biotech and pharma companies to iteratively improve protein therapeutics, enzymes, and antibodies through AI-guided design, reducing experimental cycles.	$103M	$73M (Nov 2024)	IVP, Index Ventures	2021
EvolutionaryScale	EvolutionaryScale is a New York-based AI research company spun out of Meta that develops large language models for protein biology. Its ESM3 model—trained on evolutionary sequence, structure, and function data—can generate novel proteins, including a fluorescent protein that would require ~500 million years of natural evolution.	$182M	$142M (Jun 2024)	Lux Capital, Amazon, NVIDIA	2023
Isomorphic Labs	Isomorphic Labs is a London-based AI-first drug design company spun out of Google DeepMind. Its unified AI drug design engine, built on AlphaFold 3 and other next-generation models, covers multiple therapeutic modalities and has active partnered programs with Eli Lilly, Novartis, and J&J.	$602M	$579M (Mar 2025)	Alphabet (Google), Thrive	2021
Kardigan	Kardigan is a San Francisco-based heart health company that uses AI and its proprietary Cardiac Intelligence R&D platform to match patients to targeted cardiovascular medicines. Founded by the team behind MyoKardia (acquired by BMS for $13B), Kardigan acquired Prolaio in 2025 to add cardiovascular clinical data analytics. Its late-stage pipeline addresses dilated cardiomyopathy, acute severe hypertension, and calcific aortic valve stenosis.	$554M	$254M (Oct 2025)	Sequoia Heritage, ARCH	2023
Leash Bio	Leash Bio is an early-stage AI drug discovery start-up that raised $9.4M in April 2024, backed by SpringTide, Mitsui, and MFV. The company applies machine learning to accelerate early-stage drug discovery and development programs.	$9.4M	$9.3M (Apr 2024)	SpringTide, Mitsui, MFV	2021
Lila Sciences	Lila Sciences is a Cambridge, MA-based AI company incubated by Flagship Pioneering, focused on building a general-purpose AI agent that can autonomously design and validate novel therapeutics. The company deploys robotic labs and generative AI to close the loop between computation and experiment, targeting cancer, obesity, and immune diseases.	$551M	$115M (Oct 2025)	NVentures (NVIDIA), Flagship	2023
Nabla Bio	Nabla Bio is a Boston-based AI antibody design company that uses biologically informed machine learning to computationally design antibody drugs for intractable diseases. Founded as a Y Combinator company, Nabla's JAM platform generates drug-like antibodies directly from computational design without traditional experimental screening.	$37M	$26M (May 2024)	Radical Ventures, Khosla	2020
Noetik	Noetik is a San Francisco-based AI-native biotechnology company developing precision cancer immunotherapies using its OCTO and Perturb-map platforms. By training foundation models on one of the largest proprietary multimodal oncology datasets—integrating spatial biology, genomics, and clinical outcomes—the company predicts which therapies will work in which patient subpopulations.	$55M	$40M (Aug 2024)	Polaris, Zetta, DCVC	2022
Owkin	Owkin is a Franco-American techbio company using federated and multimodal AI trained on patient data from academic institutions to discover drug targets, optimize trials, and deploy digital pathology diagnostics. Its platform integrates spatial omics, single-cell sequencing, and pathology imaging to map complex biology at molecular, tissue, and disease scales.	$337M	$33M (Mar 2026)	Alpha Intel, Sanofi, BMS	2016
Pathos	Pathos AI is an oncology-focused AI drug development company cofounded by Tempus cofounder Eric Lefkofsky in 2022. It combines millions of real-world patient records with proprietary AI to identify the right patients, design adaptive trials, and run lean, biomarker-driven development programs.	$467M	$365M (May 2025)	NEA, Lightbank	2020
Profluent Bio	Profluent is an Emeryville, CA-based AI-first protein design company that develops generative models, including the ProGen3 series, to design novel functional proteins for biomedicine and agriculture. The company is focused on designing compact CRISPR-Cas gene editing systems programmed in human language, and has commercial partnerships with Revvity, Corteva, and Ensoma.	$150M	$106M (Nov 2025)	Bezos Expeditions, Altimeter	2022
Soley Therapeutics	Soley Therapeutics is a UCSF-founded biotech that combines AI with a proprietary cell stress-sensing platform to discover first-in-class medicines for complex diseases. Founded by a cardiologist and cancer biologist, the company's biology-first approach images cellular responses to identify novel drug targets across cardiovascular and oncology indications.	$220M	$200M (Jan 2026)	Surveyor, Breyer, Leone	2018

Sources: Company websites, PitchBook Data, Inc., Crunchbase, William Blair Equity Research

Exhibit 24 (continued)

Private Company Lanscape: AI-Native Drug Discovery

Company	Overview	Total Funding ($Ms)	Latest Round ($Ms)	Key Investors	Year Founded
Atomwise	Atomwise is a San Francisco-based AI drug discovery company that uses its AtomNet deep learning platform to analyze protein structures and predict drug-target interactions for small molecule drug discovery. The company has conducted thousands of virtual drug screens for academic and pharma partners, with collaborations with Sanofi, Bayer, Lilly, and Pfizer. It was among the first to apply convolutional neural networks to structure-based drug design.	$246M	$123M (Dec 2019)	Sanofi, Bayer, Lilly, Pfizer	2012
BigHat Biosciences	BigHat Biosciences is a South San Francisco-based AI antibody design company using its Milliner platform to integrate machine learning with automated, high-throughput wet labs to design safer and more effective antibody therapeutics. The platform synthesizes and characterizes hundreds of antibodies per week, generating proprietary data to train increasingly accurate ML models.	$148M	$75M (Jul 2022)	S32, a16z, Amgen, BMS	2019
Deep Genomics	Deep Genomics is a Toronto-based AI genetic medicines company that uses an AI platform to analyze RNA biology and identify therapeutic targets for genetic conditions. The company has built a large-scale RNA atlas and proprietary AI system to predict how genetic variants affect splicing, enabling discovery of oligonucleotide therapies for rare diseases.	$241M	$180M (Jul 2021)	SoftBank, CPPIB, Fidelity	2014
Earendil Labs	Earendil Labs is an AI drug discovery company that applies AI and computational methods to accelerate early-stage drug discovery programs.	$912M	$787M (March 2026)	Dimension Capital, DST Global, Sanofi	2024
Formation Bio	Formation Bio is a New York-based AI-native pharmaceutical company that acquires or licenses clinical-stage drug assets and uses technology to accelerate and reduce the cost of clinical development. Its AI platform optimizes trial design, patient recruitment, and data collection across the development lifecycle.	$619M	$372M (Jun 2024)	a16z, Sanofi, Sequoia	2013
Iktos	Iktos is a Paris-based AI drug discovery company that combines generative AI with robotic chemical synthesis through its Makya end-to-end platform, significantly accelerating small-molecule drug discovery timelines.	$21M	$16.4M (Mar 2023)	M Ventures, Debiopharm	2016
Insilico Medicine	Insilico Medicine is a Cambridge, MA-based clinical-stage AI biotech applying an end-to-end generative AI stack—target identification, molecule design, and clinical forecasting—to drug discovery. Its Pharma.AI platform has generated multiple IND-stage programs across fibrosis, oncology, and CNS disease. Its lead IPF candidate rentosertib showed positive Phase IIa results, marking one of the first end-to-end AI-discovered drugs in human trials.	$842M	$123M (Mar 2025)	Warburg Pincus, Qiming	2014
Insitro	Insitro is a San Francisco-based AI drug discovery company founded by Daphne Koller that integrates high-throughput cellular biology with machine learning to discover and develop therapeutics. Its platform combines in vitro cellular data with human clinical data to build predictive disease models across metabolism, oncology, and neuroscience.	$643M	$400M (Feb 2022)	a16z, GV, ARCH, Foresite, B Capital	2018
Kailera Therapeutics	Kailera Therapeutics is a clinical-stage biotech developing next-generation GLP-1-based therapies for obesity and metabolic disease. The company in-licensed a portfolio of drug candidates from Jiangsu Hengrui Pharmaceuticals, including lead injectable dual GLP-1/GIP agonist KAI-9531. Its lead program demonstrated ~22% mean weight loss in Phase II and is advancing into global Phase III trials.	$1,000M	$1.0B (Oct 2025)	Atlas, Bain Capital, RTW	2024
LabGenius Therapeutics	LabGenius (now LabGenius Therapeutics) is a London-based ML-driven antibody discovery company that uses its EVA platform to co-optimize multiple antibody properties simultaneously via closed-loop robotic automation, synthetic biology, and machine learning. It focuses on discovering next-generation therapeutic antibodies for solid tumors.	$93M	$45M (May 2024)	Atomico, Lux, Obvious	2012
Manifold Bio	Manifold Bio is a Boston-based AI drug discovery company that combines protein barcoding and in vivo multiplexed measurement with computational protein design to build tissue-targeted biologics. Its platform allows rapid, parallel in vivo screening of protein variants, providing a unique dataset for training predictive models of biological behavior.	$77M	Series A ext. (Nov 2024)	Playground Global, Section 32, Triatomic	2019
Terray Therapeutics	Terray Therapeutics is a Monrovia, CA-based AI-driven small-molecule drug discovery company whose tNova platform integrates ultra-high-throughput experimentation with generative AI, having measured over 5 billion target-ligand interactions.	$226M	$120M (Oct 2024)	Bedford Ridge, NVentures (NVIDIA), Maverick	2018
Valo Health	Valo Health is a Lexington, MA-based AI drug discovery company using its Opal Computational Platform to integrate large-scale human data, machine learning, and tissue biology for drug discovery. Its pipeline spans cardiometabolic, renal, oncology, and neurodegenerative diseases, with a landmark $2.76B partnership with Novo Nordisk signed in 2025.	$595M	$450M (Jan 2021)	Koch, HBM, Atinum	2019
Xaira Therapeutics	Xaira Therapeutics is a San Francisco-based AI drug discovery company cofounded by David Baker and launched in 2024 with $1B-plus in backing. It integrates protein design AI (RFdiffusion, RFantibody), functional genomics, and proteomics to connect biological targets to engineered molecules.	$1,300M	$800M (Aug 2025)	ARCH, Foresite, Sequoia	2023

Sources: Company websites, PitchBook Data, Inc., Crunchbase, William Blair Equity Research

Exhibit 24 (continued)

Private Company Landscape: Lab Automation

Company	Overview	Total Funding ($Ms)	Latest Round ($Ms)	Key Investors	Year Founded
Automata	Automata is a London-based lab robotics company that provides AI-ready, software-defined robotic automation platforms for life science labs. Its LINQ system integrates hardware and software to automate sample processing workflows, and its cloud-enabled orchestration layer coordinates multi-instrument protocols.	$144M	$45M Series C (2026)	Dimension, A.P. Moller, Danaher, Octopus	2015
ChemLex	ChemLex is a lab automation and informatics company focused on chemical data management and synthesis automation workflows for drug discovery laboratories. The company develops software and automation solutions to streamline compound library management, reaction tracking, and synthetic chemistry workflows for pharma and biotech clients.	$71M	N/A	N/A	2022
Culture Biosciences	Culture Biosciences is a San Francisco-based bioprocess development company that provides cloud-connected bioreactors and a remote real-time monitoring platform to scale up and optimize fermentation and cell culture processes for biopharma. Its platform allows scientists to run parallel bioreactor experiments with AI-powered data analytics integrated with Google Cloud.	$100M	$80M Series B (2021)	Northpond Ventures, GV, Lux Capital	2016
Emerald Cloud Labs	Emerald Cloud Lab is a San Francisco-based cloud laboratory platform that enables researchers to design and run biological experiments entirely through software, with physical execution performed by fully automated robotic lab systems. Scientists submit protocols remotely and receive results without setting foot in a lab, enabling 24/7 experiment throughput.	$152M	Series B (2019)	Bezos Expeditions, Andreessen Horowitz	2010
Formulatrix	Formulatrix is a Bedford, MA-based life science instrument company that designs automated systems for protein crystallization, liquid handling, and laboratory informatics used in structural biology and drug discovery research. Its Rock Maker crystallization platform and MANTIS liquid handler are widely used in pharma and academic structural biology labs worldwide.	N/A	N/A	Private	2002
HighRes Biosolutions	HighRes Biosolutions is a Beverly, MA-based laboratory automation integration specialist that designs and builds large-scale automated screening systems for pharma and biotech, combining BioRobot robotic plate handlers with its Cellario laboratory execution software. The company is a major integrator for high-throughput drug discovery automation, orchestrating complex multi-instrument workflows for clients including top-10 pharma companies.	N/A	N/A	Private	2001
LiCONiC	LiCONIC is a Schaan, Liechtenstein-based provider of automated incubation and storage systems—including CO2 incubators, cryostorage units, and compound management solutions—used in pharmaceutical drug discovery and cell biology research. Its StoreX automated incubators and CryoStore systems are integrated with robotic handling platforms in high-throughput screening labs globally.	$16M	N/A	Private	1990
Medra	Medra is a San Francisco-based lab automation company whose Continuous Science Platform uses AI-driven general-purpose robotics to automate up to 70% of existing lab instruments without custom integrations, using vision and software controls. The platform generates structured experimental datasets to create a self-improving feedback loop between physical automation and AI reasoning.	$63M	N/A	N/A	2021
Opentrons Labworks	Opentrons is a New York-based lab automation company that democratizes access to liquid handling robotics with its OT-2 and Flex open-source pipetting robots and the Opentrons Protocol Library. The company's mission is to make automated, reproducible biology accessible to any lab at a fraction of the cost of traditional automation.	$329M	$200M (Sep 2021)	SoftBank, Khosla, DCVC, Y Combinator	2013
Strateos	Strateos is a Menlo Park, CA-based cloud laboratory automation company that offers automation-as-a-service and on-premises robotic lab platforms for synthetic biology, medicinal chemistry, and drug discovery. Its LodeStar platform allows remote control of automated lab infrastructure, and the company has deployed fully robotic cloud labs for Eli Lilly and other pharma partners.	$90M	$56M Series B (2021)	DCVC, Lux Capital, Andreessen Horowitz	2012
UniteLabs	UniteLabs is a Munich-based lab automation OS company that builds standardized cloud connectivity for laboratory instruments across manufacturers, enabling AI-driven biotech research workflows. Its platform includes 40-plus bidirectional connectors, a workflow engine, and a Python SDK that allows researchers to automate and orchestrate multi-instrument protocols.	$4M	€2.77M Pre-Seed (Feb 2025)	NAP (Cavalry Ventures), Acurio, LANA Ventures	2017

Sources: Company websites, PitchBook Data, Inc., Crunchbase, William Blair Equity Research

Exhibit 24 (continued)

Private Company Landscape: Software

Company	Overview	Total Funding ($Ms)	Latest Round ($Ms)	Key Investors	Year Founded
Benchling	Benchling is a San Francisco-based cloud R&D platform for life sciences, providing an integrated suite of tools including an electronic lab notebook (ELN), LIMS, molecular biology suite, bioregistry, and workflow management for biotech and pharma R&D teams.	$450M	$100M Series F (Oct 2021)	Altimeter, Tiger Global, Benchmark, Thrive	2012
Medidata (Dassault)	Medidata (a Dassault Systèmes company) is a New York-based clinical trial technology company providing cloud-based software for clinical data management, electronic data capture, risk-based monitoring, and trial analytics. Acquired by Dassault Systèmes for $5.8B in 2019, Medidata's Rave platform is used by hundreds of pharma companies and CROs to run clinical trials globally.	Acquired ($5.8B)	Acquired by Dassault Systèmes (2019)	Dassault Systèmes	1999
Dotmatics	Dotmatics is a Boston-based scientific informatics company providing R&D data management, ELN, chemistry, and biology software to pharma, biotech, and academic research teams globally.	N/A	N/A	Insight Partners	2005
Edison Scientific	Edison is a life sciences R&D informatics platform that provides AI-powered scientific data management, workflow orchestration, and analytics tools to help biopharma companies accelerate drug discovery and development. The platform integrates with existing lab systems and instruments to create a connected, data-driven R&D environment.	$70M	N/A	N/A	2017
LabWare	LabWare is a Wilmington, DE-based privately held laboratory informatics company that has provided LIMS and ELN software to regulated industries—including pharma, biotech, food and beverage, and environmental labs—since 1987. Its LabWare LIMS is one of the most widely deployed in the industry, supporting GxP compliance, data integrity, and full audit trails across lab operations.	N/A	N/A	Private	1987
Sapio Sciences	Sapio Sciences is a Durham, NC-based scientific data management company that provides a unified LIMS, ELN, and AI-powered automation platform tailored for pharmaceutical, genomics, and clinical research labs. Its Jarvis data management solution and highly configurable platform serve data-intensive R&D environments requiring complex workflows, large-scale data capture, and regulatory compliance.	N/A	N/A	Private	2004
Synthace	Synthace is a London-based digital experiment platform company that enables life science researchers to design, automate, and analyze experiments in a no-code environment, integrating with liquid handlers from SPT Labtech, Tecan, and Hamilton. Its platform captures structured experiment data and metadata, accelerating assay development and bioprocess optimization workflows.	$72M	$35M Series C (2021)	Horizons Ventures, Sofinnova Partners	2011
TetraScience	TetraScience is a Boston-based scientific data platform company that provides a cloud-native data pipeline infrastructure for life sciences R&D, enabling automated ingestion, harmonization, and AI-readiness of data from lab instruments, ELNs, and LIMS. Its Tetra Scientific Data Cloud supports biopharma companies across research, development, and manufacturing quality control, with integrations for 200-plus instrument types.	$92M	$42M Series B (2022)	Insight Partners, Lux Capital, Tiger Global	2014

Sources: Company websites, PitchBook Data, Inc., Crunchbase, William Blair Equity Research

Exhibit 24 (continued)

Private Company Landscape: Analytical Tools

Company	Overview	Total Funding ($Ms)	Latest Round ($Ms)	Key Investors	Year Founded
Carterra	Carterra is a Salt Lake City-based provider of high-throughput surface plasmon resonance (HT-SPR) technology for antibody discovery and characterization. Its LSA instrument delivers up to 100x the throughput of conventional SPR platforms while requiring only 1% of the sample, enabling screening of thousands of antibody candidates in a single run. Carterra is widely used by leading pharma and biotech companies for epitope binning, kinetics, and lead selection.	$80M	$60M Series C (2021)	Northpond Ventures, Novo Holdings	2005
Element Biosciences	Element Biosciences is a San Diego-based next-generation sequencing company that has developed the AVITI sequencing system, a benchtop platform offering high-accuracy, high-throughput sequencing at competitive costs. The company's AVITI chemistry achieves very low error rates and flexibility across genome, exome, and transcriptome applications, positioning it as a competitive alternative to Illumina.	$400M	$200M Series C (2021)	Data Collective (DCVC), Foresite, Illumina Ventures	2017
Malvern Panalytical	Malvern Panalytical is a Malvern, U.K.-based analytical instrument company providing materials characterization solutions—including dynamic light scattering, rheometry, X-ray diffraction, and particle sizing—used across pharma, biotech, materials science, and semiconductor industries. The company is a division of Spectris plc and is widely used for protein formulation characterization in biopharmaceutical development.	N/A	N/A	Spectris plc	2017
NanoTemper Technologies	NanoTemper Technologies is a Munich-based life science instrument company best known for its Monolith (MST) and Prometheus (nanoDSF) instruments for label-free protein interaction and stability characterization. Its MicroScale Thermophoresis technology measures biomolecular binding affinities in solution with minimal sample consumption, making it essential in fragment screening and early drug discovery. NanoTemper has raised over $100M and serves pharma, biotech, and academic labs globally.	N/A	N/A	Permira, INKEF Capital	2008
Ultima Genomics	Ultima Genomics is a Fremont, CA-based next-generation sequencing company that has developed a novel circular wafer-based sequencing platform targeting $1/genome sequencing costs. Its approach uses a modified flow cell design and proprietary biochemistry to dramatically reduce per-base sequencing costs, with the aim of making whole-genome sequencing practical for large-scale population and clinical genomics.	$600M	$400M Series B (2022)	Andreessen Horowitz, General Atlantic, EMED	2016
Vizgen	Vizgen is a Cambridge, MA-based spatial multi-omics company that pioneered single-cell spatial genomics with its MERSCOPE platform, built on its proprietary MERFISH technology developed at Harvard. Vizgen's tools are used across oncology, immunology, and neuroscience research.	$184M	$48M (Jan 2026)	ARCH Venture Partners, M Ventures, Northpond Ventures	2019

Sources: Company websites, PitchBook Data, Inc., Crunchbase, William Blair Equity Research

Exhibit 24 (continued)
Private Company Landscape: Reagents and Building Blocks

Company	Overview	Total Funding ($Ms)	Latest Round ($Ms)	Key Investors	Year Founded
ACROBiosystems	ACROBiosystems is a Newark, DE-based life science reagent company that specializes in recombinant proteins, antibodies, and bioassay tools for targeted drug discovery, including checkpoint inhibitors, CAR-T cell therapy, and ADC development. The company's catalog covers over 4,000 proteins and critical reagents for immuno-oncology and bispecific antibody research.	N/A	N/A	N/A	2010
BioIVT	BioIVT is a Westbury, NY-based provider of high-quality biological specimens and value-added services for pharmaceutical and biotech research. The company specializes in control and disease-state biospecimens including human and animal tissues, cell products, blood, and other biofluids, serving drug discovery, biomarker, toxicology, and ADME research.	$20M	N/A	N/A	1981
Cell Signaling Technology	Cell Signaling Technology is a Danvers, MA-based life science reagent company specializing in antibodies, ELISA kits, and research tools for cell biology, signal transduction, and cancer biology research.	N/A	N/A	N/A	1999
CN Bio	CN Bio is a Cambridge, UK-based organ-on-a-chip company developing its PhysioMimix microphysiological systems for human-relevant preclinical drug testing. The company offers single- and multi-organ models (liver, lung, intestine) to generate efficacy, ADME, and safety data as alternatives to animal models.	$56M	$21M Series B (April 2024)	Bayland Capital, CN Innovations Holdings, CITIC Securities, Innovate UK	2008
Discovery Life Sciences	Discovery Life Sciences is a Huntsville, AL-based biospecimen and biomarker specialist providing highly characterized human biospecimens, cellular starting materials, and multi-omic analytical services (genomics, proteomics, flow cytometry, molecular pathology).	$208M	N/A	Water Street Healthcare Partners, Ares Capital Corporation	2004
Emulate	Emulate is a Boston-based organ-on-a-chip company that develops human organ-on-a-chip technology to create physiologically relevant in vitro models for drug discovery, safety testing, and personalized medicine. Its Organ-Chip platform, developed from Harvard's Wyss Institute, mimics the microarchitecture and function of human organs including the gut, liver, kidney, and lung.	$275M	$82M Series D (2020)	Northpond Ventures, Perceptive Advisors, GlaxoSmithKline	2014
Hesperos	Hesperos is an Orlando, FL-based organ-on-a-chip company offering its Human-on-a-Chip pumpless, serum-free microphysiological platform for multi-organ drug efficacy and toxicity testing without animal models.	$8M	N/A	N/A	2015
Mimetas	MIMETAS is a Leiden, Netherlands-based organ-on-a-chip company developing its OrganoPlate microfluidic platform for 3D tissue culture, drug compound testing, and disease modeling. Its plates support kidney, liver, gut, brain, and cancer tissue models for toxicology, ADME, and drug efficacy screening.	$32M	N/A	N/A	2013
Promega	Promega is a Madison, WI-based life science reagent and instrument company providing over 4,000 products for genomics, protein analysis, cell biology, and drug discovery research. The company's portfolio spans bioluminescent assays (NanoLuc), cell viability reagents, PCR kits, and purification products used by researchers worldwide.	N/A	N/A	N/A	1978
Proteintech	Proteintech is a Chicago-based antibody and reagent company producing over 30,000 antibodies, ELISA kits, and recombinant proteins validated for research and drug discovery applications. The company is known for its strong citation record and commitment to antibody validation, serving life sciences researchers worldwide.	N/A	N/A	N/A	2002
Sanguine Biosciences	Sanguine is a San Diego, CA-based biospecimen services company that collects and provides disease-state and healthy human biospecimens—including blood, PBMCs, leukopaks, and biopsies—directly from donors for pharmaceutical and biotech researchers. Its direct-to-donor model integrates electronic medical records with specimens to support precision medicine, autoimmune, oncology, and cell & gene therapy research.	$28M	N/A	N/A	2010
STEMCELL Technologies	STEMCELL Technologies is a Vancouver, BC-based cell culture and stem cell research tools company providing over 2,000 specialized media, separation products, and tools for hematopoietic, neural, and pluripotent stem cell research. The company is a leading supplier of serum-free culture systems and EasySep magnetic separation reagents used in academic and pharma research worldwide.	N/A	N/A	Private	1993

Sources: Company websites, PitchBook Data, Inc., Crunchbase, William Blair Equity Research

Exhibit 24 (continued)

Private Company Landscape: Services

Company	Overview	Total Funding ($Ms)	Latest Round ($Ms)	Key Investors	Year Founded
Abzena	Abzena is an antibody engineering and discovery services company that provides proprietary platforms for therapeutic antibody optimization, including Fc engineering, bispecific antibody design, and affinity maturation. The company partners with biotech and pharma clients to advance antibody drug candidates from discovery through IND-enabling studies.	N/A	N/A	N/A	2004
Adimab	Adimab is a Lebanon, NH-based antibody discovery services company that provides fully human antibody discovery using its proprietary yeast-display platform to pharma and biotech clients. The company has generated hundreds of antibody leads for partners including Amgen, AstraZeneca, and Novartis under partnership models that allow clients to retain IP ownership. Adimab is highly profitable and has maintained a unique fee-for-service plus equity model.	$140M	N/A	Private	2007
Alloy Therapeutics	Alloy Therapeutics is a Waltham, MA-based drug discovery services and platform company that provides integrated discovery capabilities—including transgenic animal models, antibody engineering, and AI drug design—to enable pharma and biotech partners to develop next-generation therapeutics.	$130M	$190M Series C (Jun 2022)	General Atlantic, OrbiMed, NVentures (NVIDIA)	2017

Sources: Company websites, PitchBook Data, Inc., Crunchbase, William Blair Equity Research

The prices of the common stock of other public companies mentioned in this report follow:

10x Genomics, Inc. (Outperform)	$19.48
Absci Corporation	$2.78
Agilent Technologies, Inc.	$110.24
AptarGroup, Inc. (Market Perform)	$121.99
Avantor, Inc. (Market Perform)	$7.41
Bio-Rad Laboratories, Inc.	$265.40
Bio-Techne Corporation (Outperform)	$50.95
Bristol Myers Squibb Company (Market Perform)	$58.54
Bruker Corporation	$33.72
Charles River Laboratories International, Inc. (Outperform)	$159.16
Danaher Corporation (Outperform)	$181.52
Eurofins Scientific SE	€62.08
FUJIFILM Holdings, Inc.	¥9.17
Genscript Biotech Corporation	HKD 11.10
Illumina, Inc.	$117.67
Lonza Group AG (Outperform)	CHF 491.80
Merck KGaA	€24.04
Mettler-Toledo International Inc.	$1232.04
Microsoft Corporation (Outperform)	$356.77
Novartis AG Sponsored ADR	$148.18
Nvidia Corporation (Outperform)	$167.52
Recursion Pharmaceuticals, Inc.	$2.94
Repligen Corporation (Outperform)	$110.12
Revvity, Inc.	$83.90
Samsung Biologics Co., Ltd.	₩1606000
Sartorius AG	€206.80
Shimadzu Corporation	¥12.26
Stevanato Group S.p.A. (Outperform)	$13.43
Thermo Fisher Scientific, Inc. (Outperform)	$473.36
Twist Bioscience Corporation (Outperform)	$44.53
Waters Corporation (Outperform)	$289.16
West Pharmaceutical Services, Inc. (Outperform)	$243.35
WuXi Biologics, Inc.	HKD 8.10

IMPORTANT DISCLOSURES

William Blair or an affiliate beneficially own or control (either directly or through its managed accounts) 1% or more of the equity securities of Bio-Techne Corporation and Twist Bioscience Corporation as of the end of the month ending not more than 40 days from the date herein.

William Blair or an affiliate is a market maker in the security of 10x Genomics, Inc., 908 Devices Inc., Alpha Teknova, Inc., AptarGroup, Inc., Avantor, Inc., Bio-Techne Corporation, Danaher Corporation, Ginkgo Bioworks Holdings, Inc., Maravai LifeSciences Holdings, Inc., MaxCyte, Inc., Repligen Corporation, Stevanato Group S.p.A., Thermo Fisher Scientific Inc., Twist Bioscience Corporation, Waters Corporation, West Pharmaceutical Services, Inc. and Option Care Health, Inc.

William Blair or an affiliate expects to receive or intends to seek compensation for investment banking services from 10x Genomics, Inc., 908 Devices Inc., Alpha Teknova, Inc., AptarGroup, Inc., Avantor, Inc., Bio-Techne Corporation, Danaher Corporation, Ginkgo Bioworks Holdings, Inc., Maravai LifeSciences Holdings, Inc., MaxCyte, Inc., Repligen Corporation, Stevanato Group S.p.A., Thermo Fisher Scientific Inc., Twist Bioscience Corporation, Waters Corporation, West Pharmaceutical Services, Inc. and Option Care Health, Inc. or an affiliate within the next three months.

Officers and employees of William Blair or its affiliates (other than research analysts) may have a financial interest in the securities of 10x Genomics, Inc., 908 Devices Inc., Alpha Teknova, Inc., AptarGroup, Inc., Avantor, Inc., Bio-Techne Corporation, Danaher Corporation, Ginkgo Bioworks Holdings, Inc., Maravai LifeSciences Holdings, Inc., MaxCyte, Inc., Repligen Corporation, Stevanato Group S.p.A., Thermo Fisher Scientific Inc., Twist Bioscience Corporation, Waters Corporation, West Pharmaceutical Services, Inc. and Option Care Health, Inc.

This report is available in electronic form to registered users via R*Docs™ at https://williamblairlibrary.bluematrix.com or www.williamblair.com.

Please contact us at +1 312 236 1600 or consult https://www.williamblair.com/equity-research/coverage for all disclosures.

Matt Larew attests that 1) all of the views expressed in this research report accurately reflect his/her personal views about any and all of the securities and companies covered by this report, and 2) no part of his/her compensation was, is, or will be related, directly or indirectly, to the specific recommendations or views expressed by him/her in this report. We seek to update our research as appropriate. Other than certain periodical industry reports, the majority of reports are published at irregular intervals as deemed appropriate by the research analyst.

DOW JONES: 45216.10
S&P 500: 6343.72
NASDAQ: 20794.60

Additional information is available upon request.

Current Rating Distribution (as of March 31, 2026):

Coverage Universe	Percent	Inv. Banking Relationships *	Percent
Outperform (Buy)	73	Outperform (Buy)	13
Market Perform (Hold)	26	Market Perform (Hold)	2
Underperform (Sell)	1	Underperform (Sell)	0

*Percentage of companies in each rating category that are investment banking clients, defined as companies for which William Blair has received compensation for investment banking services within the past 12 months.

The compensation of the research analyst is based on a variety of factors, including performance of his or her stock recommendations; contributions to all of the firm's departments, including asset management, corporate finance, institutional sales, and retail brokerage; firm profitability; and competitive factors.

OTHER IMPORTANT DISCLOSURES

Stock ratings and valuation methodologies: William Blair & Company, L.L.C. uses a three-point system to rate stocks. Individual ratings reflect the expected performance of the stock relative to the broader market (generally the S&P 500, unless otherwise indicated) over the next 12 months. The assessment of expected performance is a function of near-, intermediate-, and long-term company fundamentals, industry outlook, confidence in earnings estimates, valuation (and our valuation methodology), and other factors. Outperform (O) - stock expected to outperform the broader market over the next 12 months; Market Perform (M) - stock expected to perform approximately in line with the broader market over the next 12 months; Underperform (U) - stock expected to underperform the broader market over the next 12 months; not rated (NR) - the stock is not currently rated. The valuation methodologies include (but are not limited to) price-to-earnings multiple (P/E), relative P/E (compared with the relevant market), P/E-to-growth-rate (PEG) ratio, market capitalization/revenue multiple, enterprise value/EBITDA ratio, discounted cash flow, and others. Stock ratings and valuation methodologies should not be used or relied upon as investment advice. Past performance is not necessarily a guide to future performance.

The ratings and valuation methodologies reflect the opinion of the individual analyst and are subject to change at any time.

Our salespeople, traders, and other professionals may provide oral or written market commentary, short-term trade ideas, or trading strategies-to our clients, prospective clients, and our trading desks-that are contrary to opinions expressed in this research report. Certain outstanding research reports may contain discussions or investment opinions relating to securities, financial instruments and/or issuers that are no longer current. Investing in securities involves risks. This report does not contain all the material information necessary for an investment decision. Always refer to the most recent report on a company or issuer. Our asset management and trading desks may make investment decisions that are inconsistent with recommendations or views expressed in this report. We will from time to time have long or short positions in, act as principal in, and buy or sell the securities referred to in this report. Our research is disseminated primarily electronically, and in some instances in printed form. Research is simultaneously available to all clients. This research report is for our clients only. No part of this material may be copied or duplicated in any form by any means or redistributed without the prior written consent of William Blair & Company, L.L.C.

This is not in any sense an offer or solicitation for the purchase or sale of a security or financial instrument.

The factual statements herein have been taken from sources we believe to be reliable, but such statements are made without any representation as to accuracy or completeness or otherwise, except with respect to any disclosures relative to William Blair or its research analysts. Opinions expressed are our own unless otherwise stated and are subject to change without notice. Prices shown are approximate.

This report or any portion hereof may not be copied, reprinted, sold, or redistributed or disclosed by the recipient to any third party, by content scraping or extraction, automated processing, or any other form or means, without the prior written consent of William Blair. Any unauthorized use is prohibited.

If the recipient received this research report pursuant to terms of service for, or a contract with William Blair for, the provision of research services for a separate fee, and in connection with the delivery of such research services we may be deemed to be acting as an investment adviser, then such investment adviser status relates, if at all, only to the recipient with whom we have contracted directly and does not extend beyond the delivery of this report (unless otherwise agreed specifically in writing). If such recipient uses these research services in connection with the sale or purchase of a security referred to herein, William Blair may act as principal for our own account or as riskless principal or agent for another party. William Blair is and continues to act solely as a broker-dealer in connection with the execution of any transactions, including transactions in any securities referred to herein.

For important disclosures, please visit our website at williamblair.com.

This material is distributed in the United Kingdom and the European Economic Area (EEA) by William Blair International, Ltd., authorised and regulated by the Financial Conduct Authority (FCA). William Blair International, Limited is a limited liability company registered in England and Wales with company number 03619027. This material is only directed and issued to persons regarded as Professional investors or equivalent in their home jurisdiction, or persons falling within articles 19 (5), 38, 47, and 49 of the Financial Services and Markets Act of 2000 (Financial Promotion) Order 2005 (all such persons being referred to as "relevant persons"). This document must not be acted on or relied on by persons who are not "relevant persons."

This report is being furnished in Brazil on a confidential basis and is addressed to the addressee personally, and for its sole benefit. This does not constitute an offer or solicitation for the purchase or sale of a security by any means that would constitute a public offering in Brazil under the regulations of the Brazilian Securities and Exchange Commission (*Comissão de Valores Mobiliários*) or an unauthorized distribution under Brazilian laws and regulations. The securities are authorized for trading on non-Brazilian securities markets, and this report and all the information herein is intended solely for professional investors (as defined by the applicable Brazilian regulation) who may only acquire these securities through a non-Brazilian account, with settlement outside Brazil in a non-Brazilian currency.

William Blair & Company, L.L.C. licenses and applies the SASB Materiality Map® and SICSTM in our work.

William Blair

Equity Research Directory

John Kreger, Partner Director of Research +1 312 364 8612
Kyle Harris, CFA, Partner Operations Manager +1 312 364 8230

Scott Hansen, Partner Associate Director of Research +1 212 245 6526

CONSUMER

Sharon Zackfia, CFA, Partner +1 312 364 5386
Group Head–Consumer
Lifestyle and Leisure Brands, Restaurants, Automotive/E-commerce

Jon Andersen, CFA, Partner +1 312 364 8697
Consumer Products

Phillip Blee, CPA +1 312 801 7874
Home and Outdoor, Automotive Parts and Services, Discount and Convenience

Dylan Carden +1 312 801 7857
E-commerce, Specialty Retail

ECONOMICS

Richard de Chazal, CFA +44 20 7868 4489

ENERGY AND POWER TECHNOLOGIES

Jed Dorsheimer +1 617 235 7555
Group Head–Energy and Power Technologies
Generation, Efficiency, Storage

Neal Dingmann +1 312 801 7835
Oil and Gas, Rare Earth

Tim Mulrooney, Partner +1 312 364 8123
Energy and Environmental Services

FINANCIAL SERVICES AND TECHNOLOGY

Adam Klauber, CFA, Partner +1 312 364 8232
Group Head–Financial Services and Technology
Financial Analytic Service Providers, Insurance Brokers, Property & Casualty Insurance

Andrew W. Jeffrey, CFA +1 415 796 6896
Fintech

Cristopher Kennedy, CFA +1 312 364 8596
Fintech, Specialty Finance

Jeff Schmitt +1 312 364 8106
Wealthtech, Wealth Management, Capital Markets Technology

GLOBAL SERVICES

Tim Mulrooney, Partner +1 312 364 8123
Group Head–Global Services
Commercial and Residential Services

Andrew Nicholas, CPA +1 312 364 8689
Consulting, HR Technology, Information Services

Trevor Romeo, CFA +1 312 801 7854
Staffing, Waste and Recycling

HEALTHCARE

Biotechnology

Matt Phipps, Ph.D., Partner +1 312 364 8602
Group Head–Biotechnology

Sami Corwin, Ph.D. +1 312 801 7783

Lachlan Hanbury-Brown +1 312 364 8125

Andy T. Hsieh, Ph.D., Partner +1 312 364 5051

Myles R. Minter, Ph.D., Partner +1 617 235 7534

Healthcare Technology and Services

Ryan S. Daniels, CFA, Partner +1 312 364 8418
Group Head–Healthcare Technology and Services
Healthcare Technology, Healthcare Services

Steven Lichtman +1 646 561 9373
Medical Technology

Brandon Vazquez, CFA +1 212 237 2776
Dental, Animal Health, Medical Technology

Life Sciences

Matt Larew, Partner +1 312 801 7795
Life Science Tools, Bioprocessing, Healthcare Delivery

Andrew F. Brackmann, CFA +1 312 364 8776
Diagnostics

Max Smock, CFA +1 312 364 8336
Pharmaceutical Outsourcing and Services

INDUSTRIALS

Brian Drab, CFA, Partner +1 312 364 8280
Co-Group Head–Industrials
Advanced Manufacturing, Industrial Technology

Ryan Merkel, CFA , Partner +1 312 364 8603
Co-Group Head–Industrials
Building Products, Specialty Distribution

Louie DiPalma, CFA, Partner +1 312 364 5437
Aerospace and Defense, Smart Infrastructure

Ross Sparenblek +1 312 364 8361
Diversified Industrials, Robotics, and Automation

TECHNOLOGY, MEDIA, AND COMMUNICATIONS

Jason Ader, CFA, Partner +1 617 235 7519
Co-Group Head–Technology, Media, and Communications
Infrastructure Software

Arjun Bhatia, Partner +1 312 364 5696
Co-Group Head–Technology, Media, and Communications
Software

Dylan Becker, CFA +1 312 364 8938
Software

Louie DiPalma, CFA, Partner +1 312 364 5437
Government Technology

Jonathan Ho, Partner +1 312 364 8276
Cybersecurity, Security Technology

Patrick McIlwee, CFA +1 312 364 8655
Software

Sebastien Naji +1 212 245 6508
Infrastructure Software, Semiconductor and Infrastructure Systems

Maggie Nolan, CPA, Partner +1 312 364 5090
IT Services

Ralph Schackart III, CFA, Partner +1 312 364 8753
Internet and Digital Media

Stephen Sheldon, CFA, CPA, Partner +1 312 364 5167
Vertical Technology – Real Estate, Education, Restaurant/Hospitality

EDITORIAL AND SUPERVISORY ANALYSTS

Steve Goldsmith, Head Editor and SA +1 312 364 8540
Katie Anderson, Editor and SA +44 20 7868 4451
Audrey Majors, Editor and SA +1 312 364 8992
Beth Pekol Porto, Editor and SA +1 312 364 8924
Lisa Zurcher, Editor and SA +44 20 7868 4549