The Next Chapter for RefinedScience

Scaling the biological data layer for drug development and diagnostics in the AI era

Drug and diagnostic developers rely on data to guide every major decision, and the industry has generated an enormous volume of biological and clinical data. However, much of that data fails to capture the biology at a resolution sufficient to know what is actually driving human disease. RefinedScience is solving this problem by using primary human tissue, longitudinal clinical data, single-cell analysis, and functional validation to put the best possible disease data in the hands of translational leads and decision makers.

Too often, industry leaders make critical and expensive decisions using data from cell lines that poorly represent disease, animal models that fail to translate, public datasets with little or no clinical annotation, and bulk sequencing methods that average away cellular heterogeneity.

Hundreds of companies are crowding into the drug and diagnostic industry, assuring investors they can use AI to address these pain points and help translational leads make better decisions. Venture capital funding for AI-driven drug discovery and development reached approximately $11 billion across 348 rounds in 2025.

Despite this influx of new players and these large investments, outcomes data show AI alone is still unable to solve most of today's biggest development challenges. The impact of AI is mostly being seen in the initial drug-discovery phase, while evidence of improved later-stage clinical success remains limited. Advanced algorithms are generating more targets and designing new molecules faster, but they are doing so without the biological ground truth that translates to real patients and better health outcomes. In an analysis of Phase II trials, AI-discovered drugs had a success rate of roughly 40%, comparable to historical industry averages.

Even in the initial discovery phase of development, activity remains highly concentrated around already-familiar targets and already-tested biology. This is not because the scientific community is running out of promising ideas. There is a vast supply of biologically interesting and under-explored targets; however, too few organizations are currently willing or able to risk the funding and resources necessary to explore even the most promising of these hypotheses.

We built RefinedScience to remove these kinds of bottlenecks, address critical gaps in biological and clinical data, and give developers a better way to explore their most promising but under-supported ideas.

If we want to develop more effective therapies for human disease and better diagnostics that can detect disease at its earliest stages, we have to study human biology at a more direct and granular level—at the level of individuals and single cells.

If we want to develop AI capabilities that can accurately predict disease outcomes and treatment responses, we must improve the underlying models by incorporating patient-specific longitudinal data and single-cell resolution.

And if we want to help developers test their most promising but untested or under-explored targets, we have to provide them with reliable and accurate ways to de-risk their strongest hypotheses.

Pharma
Data show 60-70% failure rates in Phase II trials.

AI Labs
AI cannot infer biology that was never measured.

Diagnostic Developers
Biomarkers need biology that reflects real disease.

Figure 1. Most downstream players in biology are building on the same incomplete datasets: drug programs are still built from population averages rather than patient-specific biology; AI models are still constrained by lack of deep biological information and longitudinal outcome data; and diagnostic tools are still developed using low-resolution biological data.

During the last eight years, we've assembled the right people, tools, and capabilities to reveal the underlying biology that conventional approaches often miss. What began as a disease-specific effort in acute myeloid leukemia (AML) has evolved into a platform for building proprietary, multimodal data suites from real human tissue linked to clinical context. Looking ahead, RefinedScience's next chapter is about expanding into more disease areas, data sources, and tissue sources; advancing our AI-driven translational capabilities; and strengthening relationships with new and existing partners to maximize the real-world impact of our services and our platform.

How our model is different and why it matters

Drug- and diagnostic-development efforts often fail because the field does not have access to the right information, right context, or right amount of detail about the underlying biology driving the disease at the level of the individual. Consider a few examples:

- A target may look promising in a model but not be expressed in the right cellular state in actual human tissue.
- A biomarker may appear useful until the disease is examined at single-cell resolution.
- A treatment may affect the dominant cell population while missing the rare cells that drive relapse, resistance, or nonresponse.
- A drug may engage the intended biology and still fail if the trial enrolls the wrong patients.

These are areas where conventional approaches fall short. When a tumor or diseased tissue is sequenced in bulk, the result is an average signal—but human diseases are not averages. They are mixtures of cell types, cell states, tissue neighborhoods, clinical histories, treatment exposures, and molecular programs.

> "Billions are spent advancing drugs without understanding disease biology at the level of detail needed to make better development decisions."
>
> – Brett Atwood, RefinedScience CEO

In diseases such as pancreatic ductal adenocarcinoma (PDAC), where stroma can dominate the tissue landscape, bulk methods can be misleading. In rheumatoid arthritis (RA), where responder heterogeneity remains one of the largest unsolved problems in autoimmune drug development, the biology of response resides in specific tissue states and patient subgroups rather than in the general RA diagnosis.

This is why RefinedScience has focused on human tissue, single-cell resolution, clinical annotation, and functional validation with the goal of understanding which patients are most likely to respond, why others do not, where the actionable biology resides, and how clinical trials and diagnostics can be designed with higher probabilities of success.

With this in mind, we built the cornerstones of our approach around the following four components:

1. **Real human tissue.** Primary specimens and biobank access allow us to study disease within the actual biological context in which the disease occurs.
2. **Single-cell multimodal analysis.** Rather than averaging tissue into one bulk signal, we resolve disease into its component cell types, states, neighborhoods, and molecular programs.
3. **Longitudinal clinical annotation.** Biological data become more valuable when connected to diagnosis, treatment history, response, outcomes, pathology, molecular testing, and longitudinal clinical context.
4. **Wet-lab validation.** Computational insight should be tested in relevant human systems. Our wet-lab and functional-testing capabilities are designed to help validate targets, biomarkers, and therapeutic hypotheses before they become larger clinical bets.

The combined impact of these four components allows us to build datasets that reveal biology other approaches may miss, and then use those datasets to (a) create new intellectual property and drug-development strategies and (b) provide discovery services, translational validation, and clinical trial support to our partners.

How our capabilities and strategy align with current market needs

Our next chapter is about scaling our capabilities, and several key elements have converged to create a favorable growth opportunity for us. We can place these key elements into four broad categories:

1. **Pharma's needs have become more urgent.** The clinical-failure problem remains substantial, particularly at the transition from Phase II to Phase III. Many programs fail because the underlying biology, patient-selection strategy, or translational rationale is not strong enough before the trial begins. This is an industry problem that AI alone cannot solve without the right biological data at the right resolution.
2. **The overall market for proprietary biological datasets has matured.** Developers, AI companies, and research partners need to obtain as much validated, high-quality data as possible. Biopharma licensing partnerships reached $82.7 billion in announced value during Q1 2026, underscoring the industry's continued appetite for external innovation and datasets that can provide a competitive edge.
    - **AI companies** need high-quality datasets to train more useful models, which depend heavily on the quality and biological relevance of their training data.
    - **Pharma companies** need them to design better programs and trials with higher success rates.
    - **Investors** need them to de-risk potential investment opportunities and evaluate whether asset-development strategies are grounded in differentiated biology.
    - **Academic and clinical partners** need disease resources to generate new scientific insights, support publication and collaboration, and ensure therapies reach the right patients.
    - **Regulators** increasingly expect stronger biological rationales, biomarker strategies, and evidence-generation plans.
3. **Our own capabilities have matured.** We've built our core data infrastructure, strengthened privacy and security, expanded our laboratory capacity, and built the organizational infrastructure needed to support finance, legal, data science, bioinformatics, lab operations, and partner-facing execution.
4. **We have expanded into multiple disease areas.** AML remains foundational, but we have expanded into pancreatic cancer and myelodysplastic syndromes (MDS) and are currently adding new high-value indications where primary tissue availability, unmet need, and commercial relevance intersect. Our proprietary clinical data and unique biobank access via our connections with the University of Colorado Anschutz and UCHealth give us the potential to scale almost any indication.

Current and upcoming disease areas

Launched	In development	
Acute myeloid leukemia	Ovarian cancer	**Data suites for most new indications can be built and insight-ready in 3 to 6 months.**
Pancreatic cancer	Glioblastoma (GBM)	
Myelodysplastic syndromes (MDS)	Diffuse large B-cell lymphoma (DLBCL)	

Table 1. As of Q4 2026, we have full capabilities in three disease areas with three additional areas in development. Our recently expanded lab and research capacity allows us to build new data suites for partners who want insights based on indications not currently listed.

As stated above, we are meeting these market needs by expanding into more disease areas, data sources, and tissue sources; advancing our AI-driven translational capabilities; and developing new relationships while strengthening existing relationships with our partners.

These new and emerging opportunities present an exciting challenge for us

RefinedScience has transformed from an idea into a platform, building the scientific, technical, and commercial foundation needed to scale. Along the way, we have watched the industry evolve, creating opportunities that did not exist when we started this journey eight years ago.

Now, we are looking beyond simply generating more data or discovering more drug candidates. Our view of the future is one of abundance, with biology-derived data enabling more accurate diagnostics and the development of new, highly effective treatments and cures for disease—all developed more quickly and efficiently.

We are actively seeking partners in multiple areas, from pharmaceutical and biotech companies that would benefit from the insights generated from our data suites to diagnostics companies seeking to understand new markers for disease. If you would like to learn more or discuss a potential collaboration, please contact us at partnerships@refinedscience.com.

www.RefinedScience.com