Epigene Labs Announces a New Era in AI-Driven Omic Intelligence

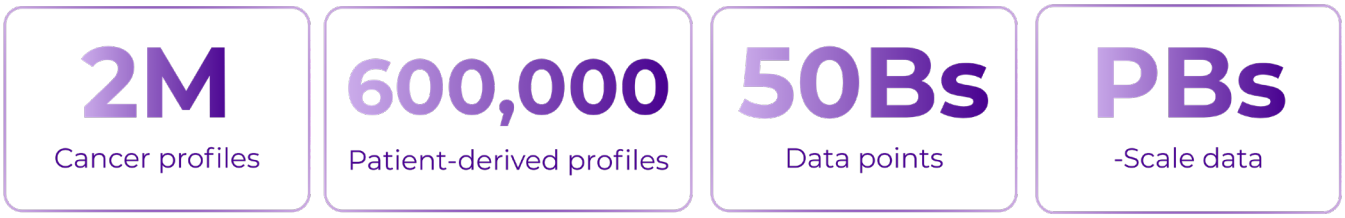
Dec 3, 2025

Seven years ago, Epigene Labs set out to build something ambitious: a new class of intelligence-augmenting technologies capable of **transforming fragmented biomedical data into the backbone of precision medicine research**. At the time, the field was defined by siloed datasets, inconsistent metadata, limited scalability, and heavy manual curation. Progress depended on researchers stitching together partial views of disease biology, each study adding a sliver of insight but rarely revealing the full picture.

From the beginning, our conviction was simple. If we could reconstruct disease biology at scale – by unifying genomics, transcriptomics, epigenomics, and clinical context across hundreds of thousands of patient samples – we could **give biopharma a radically more powerful foundation for discovering the drugs of tomorrow**.

Today, we are announcing the most significant milestone in our company's history. Over the past twelve months, we have expanded our curated cancer omic data collection by an order of magnitude. This 20x increase represents a dramatic acceleration after having crossed 100,000 curated profiles only last year. Epigene Labs now operates on more than **2 million cancer omic profiles**, including over **600,000 patient-derived profiles**, covering the full spectrum of molecular profiling technologies used in oncology research.

This achievement is not simply a matter of scale. It marks the moment Epigene Labs becomes a peer to the largest data-native players in the field while maintaining the scientific rigor, flexibility, and technological innovation that define our approach.



The Turning Point

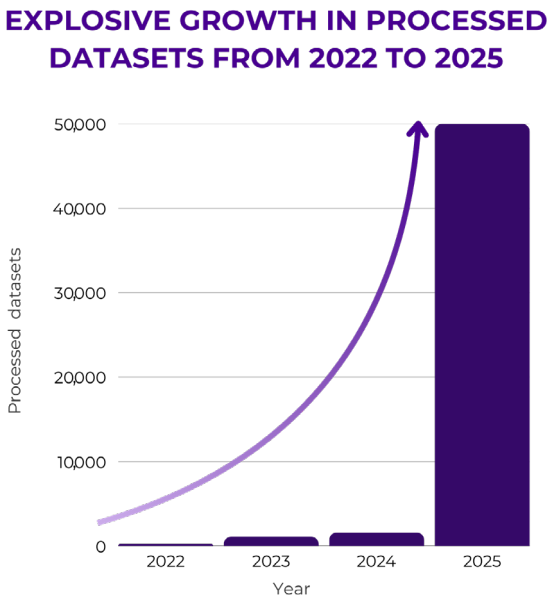
What makes this expansion transformative is not only the volume of data we now hold, but the maturity of the systems that produced it. The past year has been a story of overcoming scientific, organizational, and technological barriers that long constrained the field.

It took six years to reach 100,000 curated profiles. Reaching 2 million required redesigned workflows, new AI capabilities, and a fundamental rethinking of how biological expertise and machine intelligence can reinforce each other.

Epigene Labs Announces a New Era in AI-Driven Omic Intelligence

Dec 3, 2025

The result is a system that is not just bigger, but exponentially more productive. Over the same period in which we increased data volume twentyfold, our operational productivity increased fiftyfold. We **processed more datasets this year than in the previous six years combined** – an acceleration made possible by our human-in-the-loop AI architecture. Each dataset we ingest varies in structure, quality, and depth. To scale across this variability, we built a factory-style orchestration process: one that automates what should be automated, isolates expert intervention where it adds irreplaceable value, and dynamically selects the right model and method for each task.



While public discourse often frames AI as monolithic, our approach is deliberately pragmatic. We use classical machine learning when it outperforms large models on speed and stability, and we use advanced LLMs when they enable flexible, synthetic knowledge-driven curation.

The result is a curation engine capable of industrializing one of the most complex problems in data engineering for science: **transforming the world's fragmented omic data into coherent, richly annotated disease atlases.**

Reconstructing the Biology of Cancer at Scale

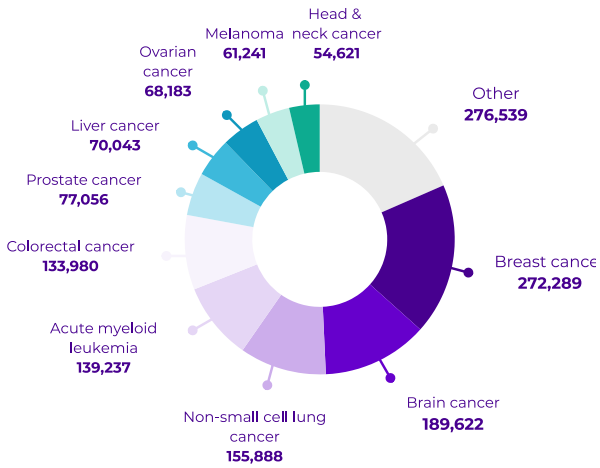
With this release, we are unveiling our first comprehensive mapping of the global cancer omics data landscape, starting with the predominant data repository in the field, NCBI's Gene Expression Omnibus (GEO). This map provides a high-resolution view of 21 and growing data elements, including information on:

- **Omics** and **sequencing technology**
- **Indication**: primary site, biopsy site, histological type, stage, grade, tumor type
- **Sample origin**: donor type, sample type
- **Treatment**: prior treatment, therapy type
- **Clinical outcome**: overall survival time, overall survival status, progression-free survival time, progression-free survival status
- **Demographics**: sex, age, biological origin
- **Geographical origin**

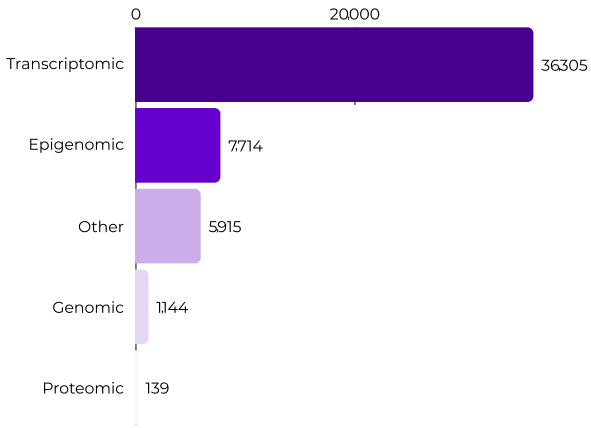
Epigene Labs Announces a New Era in AI-Driven Omic Intelligence

Dec 3, 2025

DISTRIBUTION OF CANCER TYPES



DISTRIBUTION OF THE OMICS



But the mapping is only the starting point. What truly differentiates our platform is the disease atlases built on top of this collection of curated datasets.

Unlike single-study atlases published in the literature, Epigene's atlases unify data across cohorts, institutions, and technologies to reconstruct a broad and deep view of disease biology. These atlases routinely reach scales up to ten times larger than current references, offering models of cancer that capture heterogeneity, rare patient populations, microenvironmental variation, and cross-disease patterns that would otherwise remain hidden.

This depth enables two kinds of exploration that are especially valuable for biopharma:

1. **Mechanistic exploration of disease biology**, including pathways, targets, and tumor microenvironment signatures.
2. **Biomarker exploration powered by clinical and outcomes metadata**, enabling insights tied to survival, response, and treatment history.

These capabilities transform what is possible in early research. They allow partners to validate hypotheses across dozens of datasets (studies) rather than one or two, and to explore how biological signals behave not only within individual cancer types but across "related" and "unrelated" diseases.

How We Got Here

The path to this breakthrough has been marked equally by technical innovation and organizational change. Early in our history, we operated with a sequential workflow: biologists created reference ontologies, manually curated data, and handed processed datasets to machine learning engineers. Engineers built curation models based on the training data. Biologists validated the models.

Epigene Labs Announces a New Era in AI-Driven Omic Intelligence

Dec 3, 2025

But the separation created misalignment. Errors emerged not because the models were poorly designed but because they lacked the subtle biological context that drives correct ontology and labeling choices.

We learned, sometimes painfully, that the only way to achieve consistent quality at scale was to **fully integrate expertise.** Today, machine learning engineers sit side by side with biologists. They observe how experts curate data, understand the bottlenecks, and build tools around real workflows. This has reshaped not only our productivity but also our culture. **AI is not a replacement for biological expertise; it is a force multiplier that amplifies it.**

Another critical learning emerged from our early use of deep learning and natural language processing (NLP) models. We experimented with cutting-edge architectures, including BERT-like models, but quickly confronted a fundamental constraint: these approaches require vast volumes of labeled data to reach their potential. At the time, our pool of curated omic data was simply too small to train deep learning models. So we shifted to simpler machine learning systems – not because they were less modern, but because they matched the scale of our training data.

This pragmatism paid off. As our curated data grew, we gradually introduced more advanced methods, culminating in the current hybrid approach in which LLMs, classical ML, and engineered heuristics coexist, each selected for its strengths. This layered design ensures scalability without compromising quality, speed, or interpretability.

A Framework Built for the Future

Our AI-powered curation framework follows a well-defined structure:

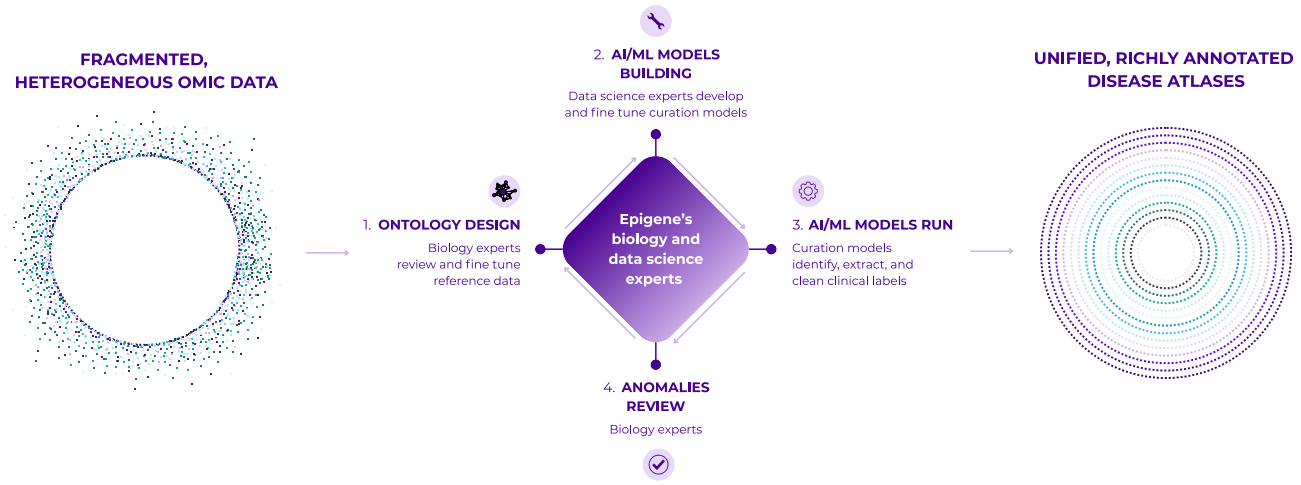
1. **Ontology co-design** between biologists and machine learning engineers
2. **Manual curation of a minimal critical mass of data**
3. **Model development and fine-tuning** tailored to each task
4. **Automated execution at scale** across dozens of thousands of samples
5. **Expert validation, feedback loops, and continuous improvement**

Quality control is embedded in every step. Rather than relying on a single performance metric, we use a suite of indicators tracking consistency, completeness, semantic accuracy, and biological plausibility. Many of these benchmarks are designed specifically for biopharma use cases, and we will soon publish scientific articles detailing the metrics we use.

This framework is now robust enough to generalize across therapeutic areas. Because patient demographics, sample descriptors, anatomical information, and many biological attributes are common across diseases, our system can rapidly extend into immunology, neurology, and other fields. The hardest part – building the framework itself – is done.

Epigene Labs Announces a New Era in AI-Driven Omic Intelligence

Dec 3, 2025



Building the Infrastructure for Scalable Biopharma Collaboration

This breakthrough enables a new level of partnership with biopharma organizations. The scale and consistency of our data allow us to develop and maintain private data marketplaces tailored to individual pharma needs. These marketplaces can integrate:

- **Public omic data** curated by Epigene
- **Commercial datasets** acquired or by our partners
- **Internal data** generated in clinical trials or research consortia

Delivery happens through multiple layers of access:

- Streamlined exploration through the company's user-friendly web app
- A forthcoming notebook environment for advanced computational users
- Secure multi-cloud architectures and cloud-to-cloud data bridges

This architecture provides a unified backbone on which partners can run discovery programs, validate hypotheses, or build internal bioinformatics pipelines and AI/ML models.

Beyond large-scale collaborations, this milestone also lays the foundation for a new product line: **off-the-shelf curated datasets** for smaller biotechs and academic groups. These datasets will be ready to use, affordably priced, and continuously refreshed as new public data becomes available.

Expanding the Boundaries of Biological Insight

One of the most powerful scientific outcomes of this expansion is the ability to analyze biology across diseases.

Epigene Labs Announces a New Era in AI-Driven Omic Intelligence

Dec 3, 2025

Because our atlases unify data from diverse studies, we can now examine relationships between cancer indications – and soon across therapeutic areas – compare immune signatures across tissues, and explore mechanistic similarities that could guide drug repurposing.

These cross-disease comparisons, long theorized but rarely feasible, will now be practical at scale. They open new avenues for understanding tumor evasion, immune activation, chronic inflammation, and therapeutic response.

Why This Matters for the Future of Precision Medicine

Reaching 2 million curated profiles is not just a numerical achievement. It represents a shift in what biopharma teams can attempt, test, and validate.

It means:

- Faster identification of **novel biomarkers** and **therapeutic targets**
- Higher-confidence **insights** validated across **multiple datasets**
- Improved decision-making for **go/no-go** choices
- New opportunities for **cross-disease research**

Most importantly, it means that Epigene Labs now offers a data and AI infrastructure capable of supporting the next decade of precision medicine innovation.

We now operate with unprecedented clarity on the global landscape of cancer omic data. We know its boundaries, its blind spots, its opportunities, and its unexplored regions. This visibility guides our next steps as much as it supports our partners.

Author: Akpéli Nordor