VUGENE

WHITE PAPER

The Biology Interpretation Gap

Why Multi-Omics Data Generation Has Outpaced Mechanistic Decision-Making in Drug Development, and our approach to catch up

Juozas Gordevičius, PhD. Co-Founder and CSO at VUGENE
Jean Pineault, MSc. Global Head, Life Sciences Practice, HumanFactor Enterprises LLC

ABSTRACT

The capacity to generate multi-omics biological data has been and remains far ahead of the capacity to interpret it. That gap has existed for over a decade, but its consequences have become structurally acute: sequencing now costs under $300 per genome, proteomics runs at industrial throughput, and data volumes compound faster than interpretation infrastructure has evolved. The industry's response has been linear: more headcount applying the same approach at increasing cost. We characterize the nature of the interpretation gap, survey why existing approaches cannot close it, define the technical architecture needed to do so, and illustrate the operational consequences when that architecture is deployed.

SECTION 1

We first characterize how the interpretation bottleneck emerged and why multi-omics accelerated it; we then review why existing solution categories cannot resolve it, specify the architectural requirements for closing the gap, describe how one implementation (VUGENE) meets those requirements in practice, and illustrate operational and regulatory consequences across real-world engagements.

A Widening Gap: Why the Bottleneck in Drug Development Has Migrated from Data to Interpretation

The challenge of extracting biological meaning from high-dimensional molecular data is not new. Sequencing technologies have been generating genomic data at scale for over a decade. What has changed is the pace and breadth of the problem.

The cost of sequencing a human genome has declined from approximately $100 million in 2001 to under $300 today. Multiplexed mass spectrometry platforms now profile thousands of proteins per sample in hours. Single-cell RNA sequencing resolves transcriptional states across hundreds of thousands of cells per experiment. The volume and dimensionality of biological data now generated in a single drug program would have taken years of instrument time a decade ago.

The industry's response to this expansion has been largely linear: more scientists performing the same analytical work at greater cost, not additional data-generation capacity. The result is that the interpretation gap has widened structurally. The rate-limiting step in understanding drug biology is no longer data generation. It has been and remains data interpretation[1,2]. In regulatory and translational terms, this means that generating phenotypic and molecular readouts is routine, but converting those readouts into auditable, mechanism-of-action evidence suitable for INDs and phase-transition decisions is still the durable bottleneck.

Why multi-omics was adopted

Multi-omics profiling was adopted because single-omics approaches could not answer the questions that drug development programs most urgently need answered. Three specific needs drove adoption, each of which remains unresolved under existing interpretation infrastructure.

The first is target discovery and validation. A disease-relevant molecular target visible only in the proteome, or in the methylation landscape of enhancer regions, is invisible to transcriptomics alone. Integrating multiple molecular layers simultaneously allows researchers to identify novel targets with cross-layer evidence, and to validate existing ones against orthogonal biological signals. A target supported by transcriptomic, proteomic, and epigenomic evidence simultaneously is a materially stronger candidate than one supported by a single layer.

The second is mechanistic characterization of drug action. Understanding what an administered molecule does to a biological system at molecular resolution, which pathways it engages, which it does not, what off-target interactions it produces, and why, requires the cross-layer picture that no single omics modality can provide. This is the Mechanism of Action (MoA) evidence that regulatory bodies require and that development decisions depend on.

The third is patient stratification and biomarker development. The majority of Phase II efficacy failures do not occur because the drug has no effect. They occur because it was tested in the wrong patient population, without a mechanism-based enrollment criterion or a pharmacodynamic readout capable of detecting early signal. Identifying which molecular subpopulations carry the disease mechanism, and which biological signatures can serve as trial endpoints, requires the same cross-layer integration that MoA characterization demands.

All three purposes converge on the same operational requirement: interpretation must keep pace with data generation. A 7-week interpretation cycle versus a 7-month one is not a convenience difference. It determines how many molecule iteration cycles a program can complete before a development decision is forced. Extended interpretation cycles also burn capital and miss development milestones. The generation technology has delivered on its promise. The interpretation infrastructure has not kept pace[14]. For a typical early oncology or immunology program, this difference can mean three to four design–test–interpret loops before

first-in-human entry, versus one to two cycles under a 7-month lag, with direct consequences for target selection, dose finding, and biomarker strategy.

The cost of the gap

The experimental work required to generate multi-omics data demands considerable expertise and resources: study design, sample sourcing, quality control, and platform execution each carry their own complexity. What has shifted is where the rate-limiting step sits. Once data leaves the instrument, converting it into a mechanistic biological decision is, in the field's current state, the harder problem[3].

In drug development, this matters in precise economic terms. A molecule that produces the intended pharmacological effect but lacks a characterized mechanism cannot be efficiently advanced through preclinical development, cannot be positioned with confidence in a clinical protocol, and cannot be defended before a regulatory body. The primary driver of Phase II failure is inadequate clinical efficacy, which has now fallen to a 28% Phase II success rate and an all-time low likelihood of approval of 6.7% from Phase I entry. While these statistics reflect multiple contributing factors, early mechanistic blind spots materially increase the risk of mis-enrollment, suboptimal dose selection, and endpoint mis-specification, thereby amplifying the observed efficacy deficit. A significant proportion of these efficacy failures trace to insufficient mechanistic characterization earlier in the program: failure to identify the patient population carrying the relevant molecular signature, failure to characterize off-target engagement, and failure to build the biomarker strategy that would have detected early efficacy signals[4,13].

> **Data generation is a commodity. Interpretation is a scarce resource. The bottleneck has migrated.**

Why multi-omic integration is the hard problem

Each omics dataset is difficult to integrate with the others. A transcriptomics experiment produces tens of thousands of features. A proteomics experiment from the same biological system produces a partially overlapping, partially contradictory feature set measured by a different technology, subject to different noise structures, and operating on a different dynamic range. DNA methylation at CpG sites in promoter and enhancer regions reflects longer-term transcriptional regulation. Chromatin accessibility across large genomic regions governs transcription factor binding and reflects more dynamic upstream regulatory states. Neither has a simple or direct relationship to the transcriptional changes measured downstream. Metabolomics reflects the functional consequences of pathway activity across all of the above[3,14].

Each omics layer requires specialized tools and domain expertise to process correctly. Integration across layers requires an additional and distinct competency beyond any single-layer specialization. The combination of per-layer processing and cross-layer integration produces a complexity that scales faster than headcount can address: not because skilled analysts are not capable, but because the coordination overhead, the non-standardization of judgment calls across individuals, and the absence of shared institutional memory make the manual approach structurally inefficient at program scale. Adding more people doing the same work at greater cost cannot resolve a problem that is architectural in nature.

SECTION 2

Existing Approaches Cannot Close the Gap Fast Enough

Four categories of tools and services currently address parts of the multi-omics workflow. Each has a structural ceiling that prevents it from closing the interpretation gap.

CRO bioinformatics services

Contract research organizations and specialist bioinformatics consultancies deliver high-quality analysis performed by experienced scientists. The output is typically a project-specific report: differential expression results, pathway enrichment summaries, visualization of selected findings. As a result, biologically critical conclusions remain embedded in static documents and email threads rather than in a standardized, queryable interpretive layer that can compound across projects The scientist receiving that report is still responsible for the mechanistic interpretation. The question of what the data means for the drug program is not answered; it is returned to the client in a more organized form.

The structural limitation is not analyst skill. It is architecture. Each engagement starts from that analyst's individual baseline. Interpretive judgment calls are not standardized across projects: choices around multiple testing correction, gene set selection, minimum gene thresholds, background gene set definition, and species-specific pathway annotation each affect the biological conclusion and are made differently by different analysts. This analyst-dependence is tolerable at single-project scale but becomes a structural risk at portfolio scale, where reproducibility of mechanistic claims and consistency of biomarker narratives are prerequisites for regulatory defensibility. Two analysts applying the same declared methods to the same dataset will frequently surface different findings[15].

Workflow and pipeline automation tools

A second category automates the computational steps of bioinformatics analysis: read alignment, quality control, normalization, differential expression, and pathway enrichment. These platforms reduce the data engineering burden and accelerate the production of standard analytical outputs. They do not perform mechanistic interpretation but provide an expert to do so. A platform that automates the production of a volcano plot and a GSEA table has not answered the biological question. In practice, the scientific bottleneck therefore shifts from data wrangling to prioritization and narrative synthesis, then deciding which of the many statistically significant signals truly matter for the specific therapeutic hypothesis and development decision at hand. It has automated the generation of the inputs the scientist still needs to interpret.

The practical consequence is visible in day-to-day research operations: a single-cell sequencing experiment produces a UMAP plot. The UMAP shows cluster structure. The scientist must then determine which clusters are biologically meaningful, which markers define each population, how those populations relate to the biological question under investigation, and what the transcriptional differences between them signify mechanistically. The tool produced the visualization. Interpretation means knowing where to focus: filtering out obvious findings, known biology, likely

technical artifacts, and routine biological variability to identify the few signals that are meaningful in the context of the specific research question. That cognitive step has not been a subject of automation. It remains in the realm of the scientist. The bottleneck has shifted upstream, it has not been eliminated[14,15].

No-code visualization platforms

A related category targets the bench scientist who lacks bioinformatics or coding expertise. These platforms provide graphical interfaces for running standard single-omics analyses: RNA-seq differential expression, basic pathway enrichment, volcano plots, and heatmaps. They are valuable tools for exploratory analysis, but they do not deliver mechanistic conclusions.

The specific mechanism of failure is more precise than it might appear. Such a platform may provide a tool to compute and visualize a PCA of a dataset. But the scientist must still understand whether the data should be scaled, filtered, imputed, and subsetted beforehand. Each of those decisions, and the parameters chosen, produces a different result. The right choice depends on prior experience with that specific data type. A scientist without that experience will produce results, but not necessarily the right results. This gap between having a visualization and knowing how to configure it correctly for robust biological inference underscores why encoding parameter choices into a standardized, biologically informed workflow is as important as the algorithms themselves. An interpretation platform that encodes those parameter choices within a standardized, biologically-informed workflow eliminates this source of analyst-dependent variability at the point where it is most consequential.

Generative and predictive biology platforms

A fourth category uses machine learning to model biological systems and predict the response of those systems to perturbations. These platforms are hypothesis-driven and forward-looking: given a perturbation, what is the predicted downstream effect? This is a distinct and complementary function. It is not a mechanistic interpretation of empirical multi-omics data. The quality of any such predictive model is bounded by the breadth and quality of the data it is trained on and by the number of omics layers it can draw on. Curated, high-quality proprietary multi-omics inputs materially improve performance. Recursion's own leadership has described the preclinical-to-clinical translation gap as one of the biggest unsolved problems in the industry, and has iterated its platform architecture repeatedly, adding transcriptomics and proteome-wide target interaction modeling, to address the distance between phenomics-scale outputs and clinical-grade mechanistic evidence. That iterative architecture evolution is an acknowledgment that phenomics alone does not close the gap[5].

Current generative models in this space also carry reproducibility and reliability limitations that constrain their use in regulatory contexts. Model outputs that cannot be independently reproduced from the same input data, or that vary materially across runs, are not suitable as primary evidence in IND filings or regulatory submissions without substantial additional validation. Until training-data provenance, run-to-run stability, and error-characterization are routinely documented at a level comparable to established statistical workflows, these models will remain better suited to hypothesis generation than to serving as standalone mechanistic evidence.

The structural gap that emerges from this survey: few commercially deployed platforms combine cross-modal integration, explainability, reproducibility, and decision-ready mechanistic interpretation within a single workflow. Achieving all of these properties simultaneously, at production scale, and across diverse biological systems, remains an open challenge in the field.

SECTION 3

The Architecture Required

Closing the interpretation gap requires a platform architecture that addresses each of the structural failures identified above. Five requirements in Table 1 follow directly from the analysis. What happens in the absence of each is predictable and observable in current practice as summarized in Table 2.

Table 1 | Requirements for closing the interpretation gap.

REQUIREMENT	CONSEQUENCE OF ABSENCE
Simultaneous cross-modal integration	Sequential single-omics analysis cannot surface cross-layer mechanisms. A transcriptomic signal without concurrent proteomic and metabolomic context cannot be confidently attributed to a specific biological process. The mechanistic picture is constructed post hoc from partial views, introducing interpretive error at each step.
Automated preprocessing	Batch correction, cell-type deconvolution, ambient RNA removal, doublet detection, and normalization across modalities each require analytical judgment calls. When performed manually, expert time is consumed before any biological question is addressed. When inconsistently applied, downstream interpretation is unreliable. Codifying this translation step into a repeatable workflow ensures that similar data types and questions yield comparable mechanistic narratives, rather than diverging with each new analyst or consulting engagement.
Automated postprocessing and biological narrative generation	Statistical output is not biological interpretation. Differential expression tables, GSEA results, and pathway enrichment scores require translation into mechanistic statements. This translation must be systematic, not analyst-dependent. It starts from knowing the research question that leads to the study and is typically in the realm of consulting services or, in larger organizations, internal expert teams.
Explainability at every decision node	A mechanistic conclusion that cannot be traced back to the data that supports it cannot be audited, reproduced, or defended. Every inference must carry its evidentiary basis. At the extreme, any figure, data representation, or statistical test should be accompanied by a computing environment and machine code that translates raw input data to that particular outcome.

Cross-study compounding knowledge	Per-project resets eliminate the analytical advantage from exposure to a large and diverse body of prior studies. A platform starting from generic baselines cannot surface cross-program signals, cannot recognize unusual patterns based on prior experience, and delivers no more interpretive value on study 500 than on study 1.
Days-scale turnaround	A program awaiting multi-omics interpretation for four to eight weeks cannot run the molecule iteration cycles that early development decisions require. For hit-to-lead, lead optimization, and early biomarker feasibility studies, this cadence typically means days to a few weeks per cycle, not months, if mechanistic evidence is to inform rather than merely document decisions already taken. Interpretation must operate at the cadence of R&D decisions, not constrain it.

Table 2 | Data analysis approaches and their bottlenecks.

CATEGORY	WHAT IT DOES WELL	STRUCTURAL LIMITATION FOR INTERPRETATION
CRO bioinformatics services	High-quality bespoke analysis; expert QC	Knowledge trapped in reports; non-standardized judgment, no cross-study memory
Workflow / pipeline automation	Robust data engineering, reproducible pipelines	Stops at statistical outputs; no mechanistic narrative or prioritization
No-code visualization platforms	Accessibility for non-coders; quick plots	Parameter-choice sensitivity; no guidance toward biologically appropriate configurations
Generative / predictive platforms	Hypothesis generation; perturbation modeling	Not tied to empirical multi-omics; reproducibility/traceability concerns in regulatory context

SECTION 4

How VUGENE addresses these requirements

VUGENE is a mechanistic interpretation platform designed around the requirements described above. Its architecture integrates multiple omics modalities in a single analytical workflow: genomics, epigenomics (DNA methylation and chromatin accessibility), transcriptomics (bulk, single-cell, and spatial), proteomics, metabolomics, and lipidomics. Integration is simultaneous rather than sequential; cross-layer associations are computed across all modalities in the same analytical pass rather than synthesized post hoc from individual modality reports.

Preprocessing is automated and standardized: quality control, normalization, batch correction, cell-type deconvolution, ambient RNA removal, doublet detection, etc. are applied through a set of fixed omics specific pipelines whose parameter choices are learned from accumulated experience across 6,500+ completed studies. These pipelines evolve under explicit governance: parameter updates are driven by

benchmark datasets, performance metrics, and expert review, and are locked per workflow version, ensuring that each study can be reproduced against a clearly documented configuration. The pipeline produces consistent, reproducible results regardless of which analyst initiates a study. Postprocessing translates statistical outputs into structured mechanistic statements grounded in the specific research question driving the study. Concretely, this can mean moving from a set of cross-layer associations and enriched pathways to statements such as: 'Compound X inhibits pathway Y in disease-relevant cell type Z, with off-target activation of pathway W in hepatocytes,' directly answerable to the sponsor's hypothesis and decision needs. Every inference is traceable to its data source. The reasoning chain from raw input to biological conclusion is auditable by design.

The cross-study library built from 6,500+ engagements is a structural property of the platform. Pattern recognition calibrated against that study population produces interpretive depth that cannot be derived from a smaller or less diverse body of prior work. Typical turnaround is measured in days, not months, enabling the iteration cadence that active development programs require.

What is fundamentally different about VUGENE

Traditional multi-omics workflows are sequential and analyst-dependent. A transcriptomics experiment is processed and interpreted by one specialist. A proteomics experiment is processed and interpreted by another. The results are then manually compared and a scientist attempts to synthesize them into a mechanistic account. Each handoff introduces parameter choices that are not standardized, each modality is interpreted in isolation from the others, and the synthesis step depends on the individual judgment of whoever performs it. The output is not reproducible in the formal sense: a different analyst, or the same analyst at a different time, will make different choices and reach different conclusions from the same data. By contrast, an integrated, parameter-locked workflow produces a consistent interpretive baseline; scientific experts then build program-specific strategy and regulatory narratives on top of that baseline rather than recomputing it ad hoc for each project.

VUGENE's core innovation is the replacement of this sequential, analyst-dependent process with a single integrated analytical pass. All omics layers are processed simultaneously, using a standardized preprocessing pipeline whose parameter choices are fixed and documented. Cross-layer associations are computed computationally rather than assembled by manual comparison. Every step from raw data to biological conclusion is traceable, reproducible, and documented in a form that can be reviewed independently.

This is distinct from visualization tools, which display data but leave interpretation to the scientist. It is distinct from CRO bioinformatics services, which produce high-quality single-project analyses but do not standardize the interpretive process or compound knowledge across engagements. It is distinct from predictive AI platforms, which model what might happen under a given perturbation but do not interpret what actually happened in an empirical experiment. VUGENE operates on real experimental data and produces a mechanistic account of what the biology did, in a form that is auditable, reproducible, and suitable for regulatory and scientific scrutiny.

The platform does not eliminate the need for scientific expertise. It relocates where that expertise is applied: from data processing and report synthesis to higher-order biological interpretation, study design, and the translation of findings into program decisions. In practical terms, biology and clinical teams spend more time deciding which mechanisms to pursue, which biomarkers to advance into trials, and how to position MoA narratives for regulators and investors, and less time re-running pipelines or reconciling conflicting analyst outputs.

SECTION 5

Decision-Ready Biology in Practice

The following scenarios illustrate what changes operationally when the interpretation gap is closed. Each describes a class of engagement that occurs regularly in drug discovery and development contexts.

Pharma: systemic disease characterization and therapeutic optimization

A research team is investigating a rare disease in a small, heterogeneous patient cohort. The study objective is to define the molecular consequences of the disease-causing mutation at the molecular level and to track response to a targeted treatment across multiple patients. The multi-layer dataset spans transcriptomics, proteomics, metabolomics, and lipidomics from the same samples, generating more than 30,000 molecular features across four omics layers (Fig. 1).

Under a conventional workflow, each omics layer would be analyzed independently over several months, requiring specialist expertise at every stage. In small cohorts, analyzing individual molecular features in isolation allows biological signals to be obscured by noise, masking coordinated pathway dynamics, leaving the principal drivers of disease unresolved, and limiting detection of subtle treatment-associated molecular effects critical for therapeutic optimization.

> **Each omics layer was processed using standardized, automated pipelines tailored to its specific modality. Quality control, signal normalization, and missing-value imputation were applied to eliminate batch effects and technical artifacts. To account for inter-patient variability and repeated post-dose measurements within small cohort constraints, linear modeling frameworks were applied to isolate true disease- and treatment-associated effects over time. Recognizing that individual feature signals in small, noisy cohorts face severe statistical power limits, multi-layer data were aggregated at the functional biological level using Gene Set Enrichment Analysis (GSEA) and structured knowledge frameworks. This pathway-resolved approach transformed 30,000+ raw features into a clear cellular roadmap, mapping downstream pathology, quantifying temporal treatment rescue, and identifying residual molecular abnormalities to guide future therapeutic complementation.**

While multi-omic integration was carried out, feature-level cross-layer signals in small, noisy cohorts face severe statistical power limits. VUGENE overcame this limitation of feature noise by aggregating data at the pathway level within a structured biological knowledge framework. Crucially, while the causative genetic defect was known, the systems-level molecular consequences of the mutation remained unresolved. VUGENE mapped these core biological drivers for the first time, resolving pathology into distinct operational states, such as active cellular "Defense Modes" (immune/stress alerts) versus "Pause Modes" (growth/energy suppression).

Building on this systemic baseline, VUGENE evaluated the molecular impact of intervention. Although Bachmann–Bupp syndrome was an established ODC1 gain-of-function disorder and DFMO the target therapeutic rationale, VUGENE provided the first comprehensive, multi-layer molecular assessment of treatment response. By evaluating patient profiles over time, the platform tracked temporal response, quantified downstream pathway normalization, and uncovered residual molecular abnormalities to guide future therapeutic complementation.

The Corewell Health collaboration demonstrated this outcome: translating 30,000+ molecular features across biological layers into an actionable cellular roadmap – moving from data to decision in seven weeks across a study that traditionally requires six to twelve months of sequential analysis[6].

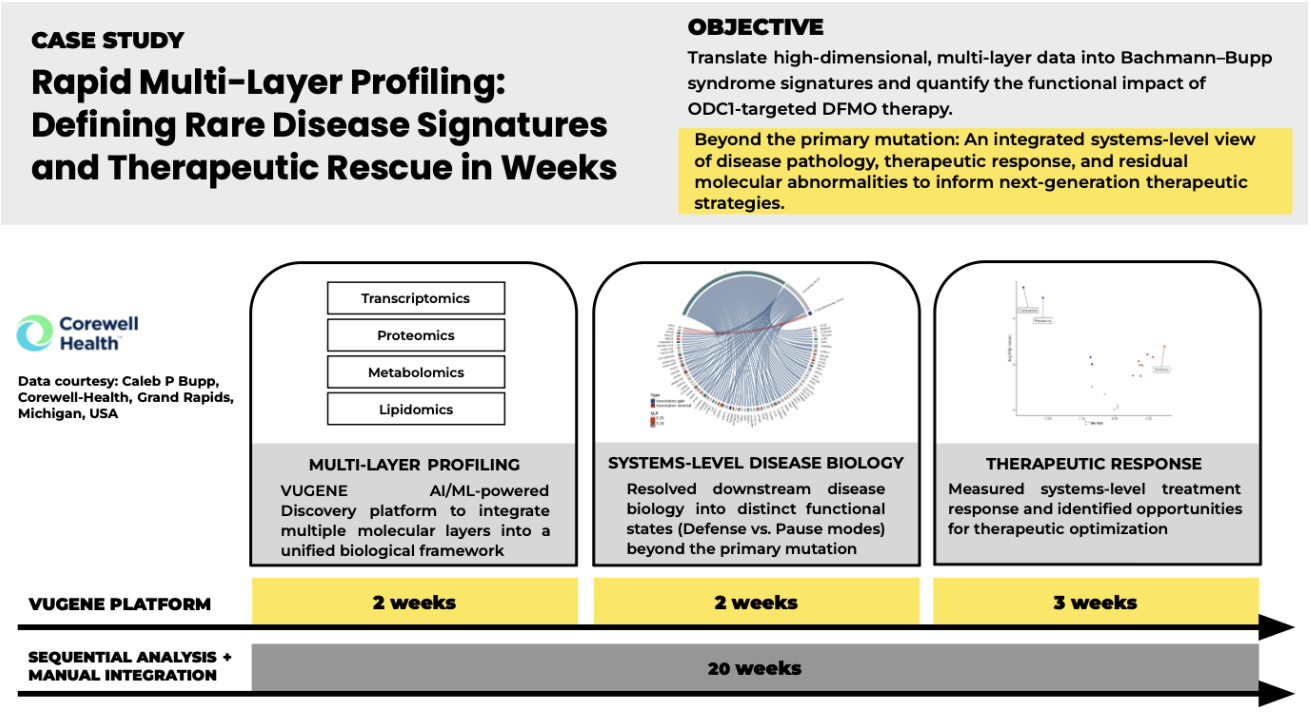

Fig. 1 | Case study, Dr. Caleb P. Bupp, Corewell Health.

CRO: embedded interpretation as a service differentiator

A contract research organization generates high volumes of multi-omics data on behalf of pharmaceutical clients. Its core competency is data generation: study design, sample preparation, sequencing and/or mass spectrometry, and delivery of raw and processed data files. Without an interpretation layer, it is a commodity vendor. Its data output is indistinguishable from any other facility of comparable technical quality.

When VUGENE's mechanistic interpretation layer is embedded within the CRO's service delivery, each data deliverable includes a biological decision, not only a dataset. For sponsor organizations, this reduces re-analysis cycles, accelerates biomarker triage, and enables more confident prioritization of interventions, turning each CRO engagement into a direct input into development strategy rather than just a data-generation event. The output is transformed from a commodity to a decision-enabling service. Headcount does not increase. The service offering is structurally differentiated.

The Clock Foundation engagement illustrates this at production scale (Fig. 2). The Clock Foundation develops and deploys epigenetic aging clocks for use in preclinical and clinical longevity studies, including PhenoAge, GrimAge, and multi-species clocks spanning diverse mammalian systems. Their operational requirement was not a single analytical project; it was a production-scale interpretation infrastructure capable of processing the diverse epigenetic aging study designs rapidly, in parallel, and reproducibly, while also supporting the development of novel epigenetic age predictors. VUGENE built an automated epigenetic data pipeline with simultaneous ML/AI and statistical testing, capable of processing multiple independent datasets in parallel. The Clock Foundation's scientific team shifted its focus from data processing to higher-order questions: which interventions modify biological age, in which populations, through which epigenetic mechanisms, and whether novel clock architectures can be trained on the accumulating dataset. The interpretation infrastructure enabled that shift without requiring additional analytical headcount to scale with volume. This engagement received the Impact Award from the Baltic-American Freedom Foundation and the U.S. Embassy in Lithuania, recognizing the translational significance of applying automated epigenetic interpretation at clinical study scale[16,7].

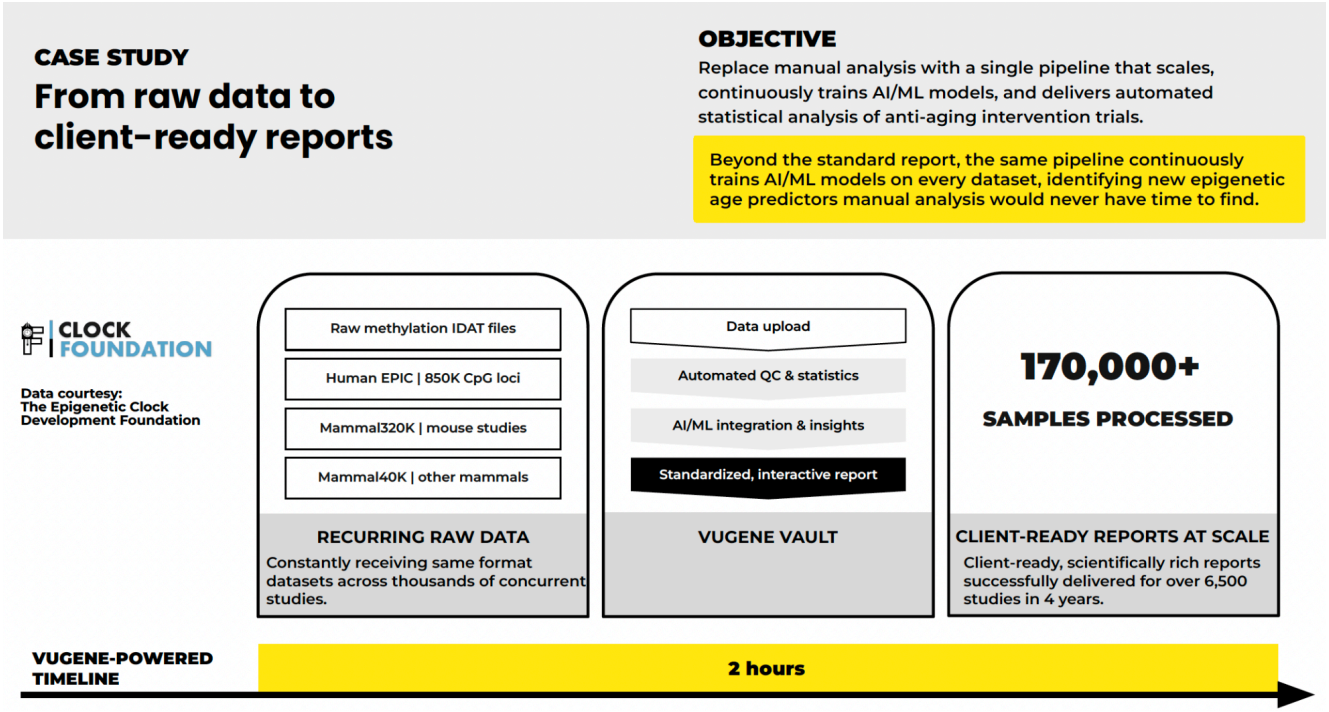

Fig. 2 | Case study, The Epigenetic Clock Development Foundation.

Regulatory: IND-supportive mechanistic evidence

The FDA's requirement for MoA evidence is not a single submission event. It is a continuous thread that runs from pre-IND through NDA or BLA and into the approved label. At the pre-IND stage, submissions should include evidence of preclinical activity and mechanism of action establishing a rationale for use in the intended patient population. At the IND, the MoA section must be supported by mechanistic evidence demonstrating a plausible biological rationale for the therapeutic hypothesis and characterizing the expected pharmacological effects and off-target liabilities. At Phase II, dose selection rationale, patient stratification strategy, and biomarker endpoint design are each expected to be mechanistically grounded. At NDA or BLA, the clinical pharmacology section must include a fully characterized MoA for each approved indication, free of speculative claims. At approval, FDA also determines the established pharmacologic class used in the Highlights of Prescribing Information, a designation derived from the mechanism of action, physiologic effect, or chemical structure of the active moiety. A MoA narrative constructed superficially at IND that cannot be substantiated at approval is a regulatory liability that accumulates quietly across the development program[8,9,10,11].

Mechanistic evidence produced by a manual, analyst-dependent process cannot reliably satisfy this requirement at any stage of the regulatory thread. The interpretive judgment calls embedded in the analysis are not standardized. The reasoning chain is not traceable from data to conclusion. Documentation describes what the analyst did; it does not make the output reproducible.

When mechanistic interpretation is produced by a standardized, automated, and auditable workflow, the evidentiary chain is intact by design. Maintaining internal validation dossiers that cover workflow performance, reproducibility metrics, and version-control history then allows sponsors to reference the interpretive infrastructure itself in regulatory interactions, aligning with emerging expectations around AI and complex analytics in drug development. The MoA narrative enters the IND not as a reconstruction assembled under deadline, but as a structured, auditable product of the interpretive workflow. That narrative can be extended and substantiated at each subsequent regulatory interaction without methodological inconsistency, because the same workflow, parameters, and biological knowledge base are applied consistently throughout the program. The FDA's 2025 draft guidance on artificial intelligence in drug development articulates principles that apply to any analytical methodology contributing to regulatory submissions: reproducibility of results from the same inputs, explainability of the reasoning chain from data to conclusion, and traceability of each analytical decision. These principles are not specific to any platform or architecture. They describe the evidentiary standard that mechanistic evidence must meet to be defensible in a regulatory context. Interpretation workflows designed around these principles from the start of preclinical development are better positioned to support submissions across the program lifecycle[12].

From a regulatory perspective, the outputs of a mechanistic interpretation platform of this type are best understood as structured, auditable supportive mechanistic evidence that complements, rather than replaces, primary clinical efficacy and safety data.

SECTION 6

The Compounding Library Effect

One property distinguishes a platform from a service: whether the value it delivers compounds over time or resets with each engagement.

Interpretive intelligence at scale

In a human expert-based model, completed projects do not improve the system's analytical capability. Each engagement draws on the knowledge and judgment of the assigned analyst. That knowledge is not transferred to the platform, cannot be queried across prior studies, and does not inform the interpretation of the next project. When the analyst leaves, the institutional knowledge leaves. The analytical output of 6,500 projects exists in 6,500 separate documents. It does not accumulate into a queryable, trainable body of biological intelligence.

VUGENE's platform has processed 6,500+ cross-modal studies across diverse biological systems, drug classes, species, and experimental designs. Models trained on that body of biological experience recognize patterns in cross-layer signals that a first encounter would miss. For example, the platform can more readily recognize recurrent cross-layer mechanistic signatures, such as characteristic immune activation motifs or shared pathway engagement patterns, across otherwise unrelated programs, strengthening mechanistic characterization earlier in a program. The interpretive accuracy and depth that results from exposure to a large and diverse study population is a structural property of the platform, not a feature of individual analysts. No comparably deployed cross-modal library exists in a commercially validated platform. A new entrant with capital can replicate the architecture. The learning curve encoded in 6,500 studies across multiple omics layers cannot be compressed with funding[3].

The Clock Foundation engagement, described in Section 5, illustrates this dynamic at scale. The 3,000+ studies completed under that single engagement have produced models calibrated to species-specific methylation dynamics, intervention-associated clock responses, and multi-species aging signatures that could not be derived from a smaller or less specialized study population. That calibration is an asset of the engagement, not a generic platform feature.

Client model calibration and continuity

Beyond the population-level library, VUGENE's models calibrate over time to the specific biological systems, assay designs, experimental protocols, and study history of each client. A client whose programs have been interpreted by the platform over multiple years has models tuned to their cell lines, their animal models, their preferred sequencing platforms, and the molecular signatures characteristic of their therapeutic area.

A platform switch under these conditions is not merely a vendor change. It requires reverting analytical models to generic baselines and losing the per-client calibration accumulated over the course of the relationship. For a program with an active IND, a methodology change introduces analytical discontinuity that must be revalidated: the mechanistic claims made under the prior methodology are no longer

reproducible by the new system without rederiving the full analytical path. In practice, this means reconciling prior MoA narratives, biomarker claims, and mechanistic safety arguments across platforms to maintain a coherent, auditable record and this work can be substantial if interpretive architectures differ materially. For active development programs, this creates continuity considerations that extend beyond vendor preference into the integrity of the analytical record.

> **Replicating the architecture may be achievable. Replicating years of accumulated biological interpretation experience is substantially more difficult. This is especially relevant for programs traversing from preclinical through late-stage development, where continuity of interpretive methodology simplifies integrated summaries, supports consistent MoA messaging, and reduces the risk of methodology-related questions at approval.**

CLOSING

The Gap Is Closing. The Question Is Who Closes It.

Regulatory expectations around mechanistic evidence are becoming more explicit. Mechanistic evidence is expected to be structured, auditable, and reproducible from the underlying data, regardless of the analytical method used to generate it. Programs that build interpretation infrastructure meeting these criteria early in development are better positioned as the evidentiary bar rises[12].

The biological data generation infrastructure of the pharmaceutical industry has already crossed the threshold where interpretation, not data, is the rate-limiting step. The programs that will advance most efficiently through development will be those in which mechanistic understanding keeps pace with the data being generated from them. That requires an interpretation architecture capable of operating at the scale, speed, and analytical depth the data demands.

The interpretation gap is closing. The question is who closes it, and when.

Acknowledgements

The authors thank Sanjeev Thohan, Ph.D. (https://www.linkedin.com/in/sanjeevthohan/) and Grant Blouse, Ph.D. (https://www.linkedin.com/in/grant-blouse-b065a73/) for their review of this paper. Both brought extensive drug development experience to a detailed scientific and editorial review of the draft.

References

1. National Human Genome Research Institute (NHGRI). DNA Sequencing Costs: Data. Tracks cost per genome from 2001 to present. https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data
2. Wetterstrand KA. The Cost of Sequencing a Human Genome. NHGRI. https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost
3. Kant S et al. Integrative Multi-Omics and Artificial Intelligence: A New Paradigm for Systems Biology. OMICS: A Journal of Integrative Biology, December 2025. Data heterogeneity, absence of standardized integration frameworks, preprocessing complexity as persistent obstacles. https://journals.sagepub.com/doi/10.1177/15578100251392371
4. Kuo YM et al. Consideration of the Root Causes in Candidate Attrition During Oncology Drug Development. Clinical Pharmacology in Drug Development, August 2024. Efficacy deficit is the leading cause of Phase II attrition. https://accp1.onlinelibrary.wiley.com/doi/full/10.1002/cpdd.1464
5. Lin F. Recursion's Fast-Track Road to Therapeutics Using AI-Based Maps of Biology. GEN Edge, October 2024. Interview with Recursion leadership (N. Khan, C. Gibson) on translation gap; iterative platform architecture. https://www.genengnews.com/topics/artificial-intelligence/recursions-fast-track-road-to-therapeutics-using-ai-based-maps-of-biology/
6. VUGENE. Case Study: Corewell Health. 30,000+ molecular features, 4 omics layers, Bachmann-Bupp Syndrome, MoA confirmed in 7 weeks. Data courtesy Caleb P. Bupp, Corewell Health, Grand Rapids, Michigan, USA. https://vugene.com
7. VUGENE. Learn How Automated Analysis and AI Modelling Helps Longevity Research. Clock Foundation partnership; 3,000+ studies; Impact Award, Baltic-American Freedom Foundation and U.S. Embassy in Lithuania. https://vugene.com/learn-how-automated-analysis-and-ai-modelling-helps-longevity-research/
8. FDA. Investigational New Drug (IND) Application. https://www.fda.gov/drugs/types-applications/investigational-new-drug-ind-application
9. FDA. Pre-IND Submission Requirements, Division of Anti-Infectives. MoA evidence required to establish rationale for clinical use. https://www.fda.gov/about-fda/cder-offices-and-divisions/division-anti-infectives-dai-information-pre-ind-submissions
10. FDA. Clinical Pharmacology Section of Labeling for Human Prescription Drug and Biological Products. MoA requirements for NDA/BLA; prohibits speculative claims. https://www.fda.gov/media/74346/download
11. FDA. MAPP 7400.13 Rev. 2. Determining the Established Pharmacologic Class (EPC) for use in the Highlights of Prescribing Information. EPC is determined from mechanism of action, physiologic effect, or chemical structure. https://www.fda.gov/media/86437/download
12. FDA. Considerations for the Use of Artificial Intelligence to Support Regulatory Decision Making for Drug and Biological Products. Draft guidance, 2025. Reproducibility, explainability, auditability standards. https://www.fda.gov/about-fda/center-drug-evaluation-and-research-cder/artificial-intelligence-drug-development
13. Norstella / Citeline. Why Clinical Development Success Rates Are Falling. May 2024. Phase II success rate 28%; LOA from Phase I at all-time low of 6.7%. https://www.norstella.com/insight/why-are-clinical-development-success-rates-falling/
14. Armstrong R. Advancing Drug Discovery Through Multiomics. Drug Discovery World, October 2025. Data integration as the foremost challenge; infrastructure bottleneck; absence of centralized framework. https://www.ddw-online.com/advancing-drug-discovery-through-multiomics-37448-202510/
15. Catalano M et al. Navigating Cancer Complexity: Integrative Multi-Omics Methodologies for Clinical Insights. Clinical Medicine Insights: Oncology, October 2025. Absence of standardized protocols; manual analyst dependency; reproducibility inconsistency. https://journals.sagepub.com/doi/10.1177/11795549251384582
16. eBioMedicine / The Lancet. Epigenetic Clocks: Advancing Biological Age Measures Towards Meaningful Clinical Use. February 2026. Epigenetic clocks in clinical trials; multi-omics integration as an active frontier. https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(26)00056-3/fulltext