knowing01

Missing the insights for the Omics data

Data is key to the discovery of new treatments in disease research. Automation, dropping costs and novel approaches lead to an increasing amount of data being generated every day. At the same time, data quality is steadily rising with the arrival of new measurement technologies. However, generating more and better data alone is not the answer.

To arrive at critical insights, we must first ask the right questions and carefully select the appropriate data and methodology.

This white paper examines today's data processing challenges in the biotech and pharma sectors, compares current solutions to knowing01's innovative elastic approach, and describes how knowing01's software uniquely addresses these challenges.

From single Omics to Multiomics challenges

The biotech and pharma industries are the **innovative risk takers** addressing today's **unmet medical needs.** They have a competitive advantage based on deep expertise in their – often under-studied – **focus area** with its **unique data scope.** Yet they face significant challenges today. The **return on investment for new treatments has reached an all-time low[1].** Approved drugs target only a small fraction of known proteins. Clinical trials sometimes have **unclear outcomes that are difficult to understand and interpret,** as some patients respond positively to treatment and others do not.

More and more, people believe that having more data could be the answer to our problems. With the steady growth in technology, we're now able to keep track of different kinds of tiny particles found in nature, called biomolecules. If the methods we're using don't give us the answers we need in one area, like genetics, researchers can now easily look into other related areas. This comprehensive exploration of related areas is known as multiomics data, and it's becoming a powerful tool for finding solutions.

Multiomics data encompasses multiple biological levels, such as genomics, proteomics, transcriptomics, epigenomics, metabolomics, etc. The study of these Omics data, also known as Multiomics analysis, offers an integrated perspective that contributes to a more comprehensive understanding of biological systems. It helps to discover new relationships between biological entities, identify relevant biomarkers or improve our knowledge of disease mechanisms.

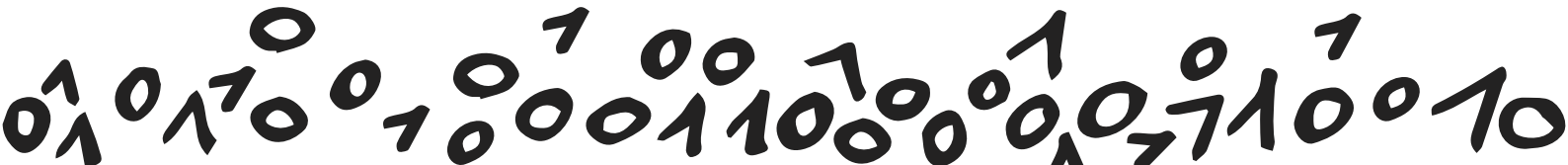
However, without the right tools and expertise, the expanding volume and complexity of data makes it impossible to obtain an integrated **Multiomics view.**

Is artificial intelligence the solution?

Artificial Intelligence (AI) and Machine Learning (ML) have emerged as powerful technologies that have and are revolutionizing various biotech fields such as the prominent examples: protein structure prediction (e.g., Alphafold), (microscopy) image analysis (e.g., for radiology and object recognition) and text writing (e.g., the large language model (LLM) wave with OpenAI's prominent ChatGPT).

However, these neural network-based AI solutions are of limited use when it comes to understanding complex disease mechanisms and the intricacies of biology, as they rely heavily on huge amounts of labeled data with known structures. Oversimplified, a neural network is learning from millions of examples and can figure out how to solve a highly complex problem. This is in stark contrast to the complex and multi-layered structure of Multiomics data.

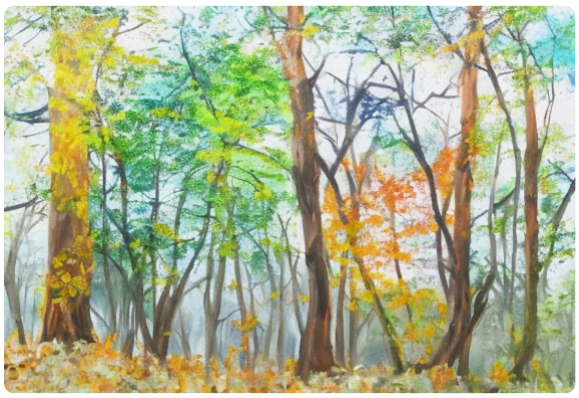
[1] Buntz, B. (2023, February 14). Pharma R&D investment returns sink to 13-year low. Retrieved from Drug Discovery & Development.



According to our CEO and founder, Dr. Nikola Müller: "For Multiomics data, we **need AI solutions** that are able to comprehend and identify the hidden complex disease mechanisms. We need to **innovate the way we approach the complex, fragmented data of today.** I believe this can be best achieved with augmentation solutions rather than black box predictions."

So, in general, neural network-based AI is already powerful in many ways, but its rigid data requirements make it an unsuitable solution for understanding complex biology. Novel solutions that work with unlabeled and heterogeneous data are the best way forward in this application area.

We at knowing01 offer a flexible and adaptable approach, our elastic biodata model, which embraces a complementary AI-based knowledge graph technology.



Missing the forest for the trees

When research and development (R&D) teams search for novel targets or biomarkers, they often restrict their analysis to the available data set and manually sift through the information to identify and prioritize key findings. This highly manual and error-prone pattern-finding process is limited by the human mind's ability to comprehend and process vast amounts of data. We challenge this approach to identifying novel targets and biomarkers and call for more advanced computational methods that analyze data sets at scale and identify patterns and relationships that are overlooked today.

With existing solutions, disease researchers easily **miss the forest for the trees** when it comes to biomolecular data. They may overlook important contexts and connections in their data while focusing too narrowly on individual data points. While they are skilled experts at measuring features of a tree (focus area), learning everything about a single tree to derive a tree profile (disease pattern or therapy options), they easily miss the (often unknown) necessary context of the forest.

Without the context of the forest – its elements: soil, climate, birds, insects etc. – a single tree can be **easily misunderstood.** However, the process of incorporating existing observations of other elements in the forest and knowledge on how they are linked is highly unstructured, disorganized, and manual today.

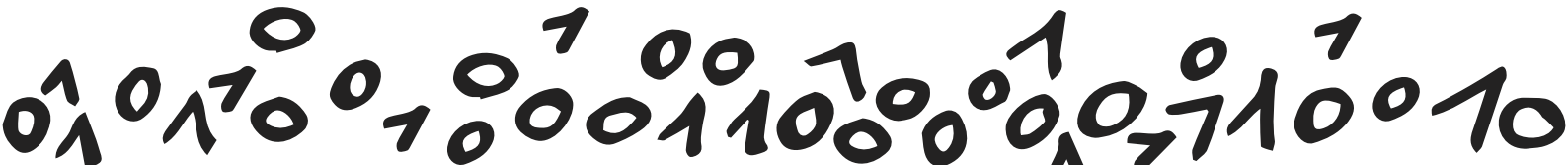
More specifically, research teams may **struggle to contextualize** their Multiomics data. The challenge is to resolve the many-to-many data links that require advanced bioinformatics and graph mining expertise. Available tools are either experimental at best or force the data into rigid output layers, resulting in a very narrow field of application. To truly unlock the value of all data, whether focus area specific or contextual, a novel approach is required that is elastic to the (data) shape of each tree and at the same time general enough to consider the entire forest.

So, what does this mean for the biotech and pharma industries? Failing to identify patterns in their data early on means continuing with **incomplete information and an increased risk of failure at later stages, after making considerable investments that cannot be recovered.**

knowing01's deep tech solution

At knowing01, we have developed an **elastic biodata model** that links Multiomics data in a flexible way, setting it apart from rigid and non-customizable software solutions. It retains the original uniqueness of the data with flexible adapters that render different Omics data layers comparable without transforming them.

The software elastically aligns novel and established data and is designed to rapidly unlock the encapsula-



ted data potential. In this way, our technology **innovates the process of understanding data by taking its context into account** – breaking free from the constraints of current rigid solutions.

Our generalized and flexible deep tech approach allows us to **address the contextualizing challenge with a single software,** solving the composition of a biotech or pharmaceutical company's unique data scope. Adding new context through additional data layers comes at minimal marginal cost, irrespective of the variety of the Multiomics data layer.

More specifically, knowing01's solution drives innovation in biotechnology and pharma through three benefits:

1. **Unlock unresolved data** to establish context for unique focus areas (the features of a tree).
2. **Ensuring maximum flexibility** of Multiomics data input and output (the elements of the forest).
3. **Freeing up resources** through automation and scalable algorithms that reduce manual work and accelerate the process from data to insights.

Application: Prioritize marker using Multiomics disease context

In the process of identifying relevant **disease markers,** R&D teams may face the challenge that a (Multi)omics analysis yields too many gene hits. A few genes or proteins should be selected for validation, e.g., by knock-out experiments, but which ones are most disease-relevant?

With knowing01's solution, biotech and pharma teams can apply a **context-aware data comparison** approach and **prioritize the interesting disease-relevant gene candidates.** This can be done with data from any species or Omics level while leaving the meaningful integration up to us. In addition, only genes that share common regulation in the existing data are selected, thus obtaining a smaller set of key markers. Through a flexible ranking scheme, key markers can be prioritized based on their upregulation in the disease and overlapping data sets. With these selected genes, the teams can now return to the lab and proceed with validation.

For most research fields, especially novel ones, data from various Omics layers must be combined to consider all available data. This results in complex, network-like relationships between data points. Standard analysis methods force them into rigid structures and output formats, which can cause significant information loss. When the data from an organism highlights "mutation", "protein modification" or "gene", why should the output layer be determined by the analysis tool instead of the disease biology and the research team?

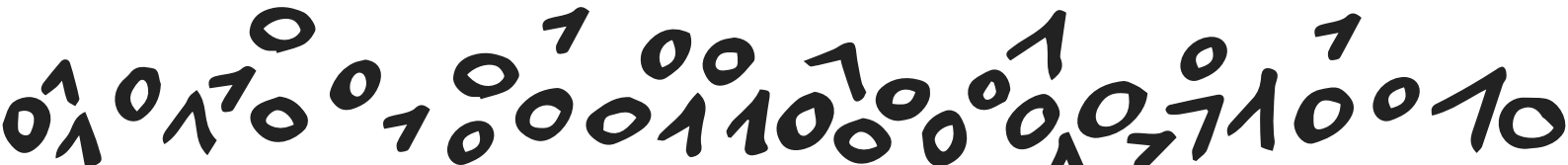
With the knowing01 solution, markers can be examined and set in context without data loss thanks to the flexibility of our underlying elastic biodata model. We automatically resolve the combinatorics of data links, thus drastically reducing the time data scientists spend on repetitive analysis and data comparison tasks, ultimately contributing to higher quality results.

Application: Augmented target identification process

Identification of a target gene is one of the main challenges in drug development. Yet (bio)pharmaceutical companies often approach this by manually reviewing public resources gene by gene. Such a process lacks structure and can be extremely time-consuming.

With knowing01's deep tech solution, a research team can easily perform a structured search to find reported evidence for a gene. Thereby, the genes of interest are run through the search function, which pulls relevant entries from the research team's data as well as from defined library databases, e.g., publications from PubMed. Results are reported in a transparent and reproducible manner, with gene inclusions and exclusions well documented.

knowing01 software can add significantly to the automation of gene target identification. Our AI-based knowledge graph technology digests unique input and output layers of Omics and Omics-related data and accommodates the changes in molecule names



and definitions over time, e.g., the complementary gene definitions of US or EU gene reference databases. The knowing01 software reduces manual errors and eliminates the need for costly algorithm and proprietary software development to an extent other solutions cannot provide.

Conclusion

Traditional data processing approaches fall short in addressing todays challenges that biotech and pharma industries are facing in understanding complex disease mechanisms and identifying novel targets and biomarkers. Multiomics data analysis offers new ways to generate comprehensive insights, but it requires sophisticated computational methods. Recent innovations in AI and ML are severely limited when it comes to unraveling yet unknown biology, since they rely on labeled data with known structures. At knowing01, we have developed an elastic biodata model that links all Multiomics data without a priori data transformations to provide a more comprehensive understanding of biological systems. Our technology helps to unlock and bring together the value of millions of data points and existing knowledge to fully capture the “tree in the forest” or biomolecular composition.

Do the challenges and use cases described apply to you? Do you have a lot of data but little insight? Talk to us, we at knowing01 are open for new collaborations.

knowing01

Introducing the groundbreaking spinoff from a renowned German research center, knowing01! Our team, led by a trailblazing female founder and expert in Multiomics data integration, has tackled the bioinformatic challenges that have plagued the industries for years. With our innovative software solution, Biotech’s can finally gain control over their data, allowing for accurate insights and meaningful knowledge expansion. We’ve tackled the complex curse of dimensionality and turned it into a powerful solution for your success.

Our dedication is to sustainability and scalability ensure that our solution will grow with your company, providing the support you need to make a **real impact in the industry**. Join the revolution with knowing01!

We’d love to hear from you

support@knowing01.com • Phone +49 89 74678081