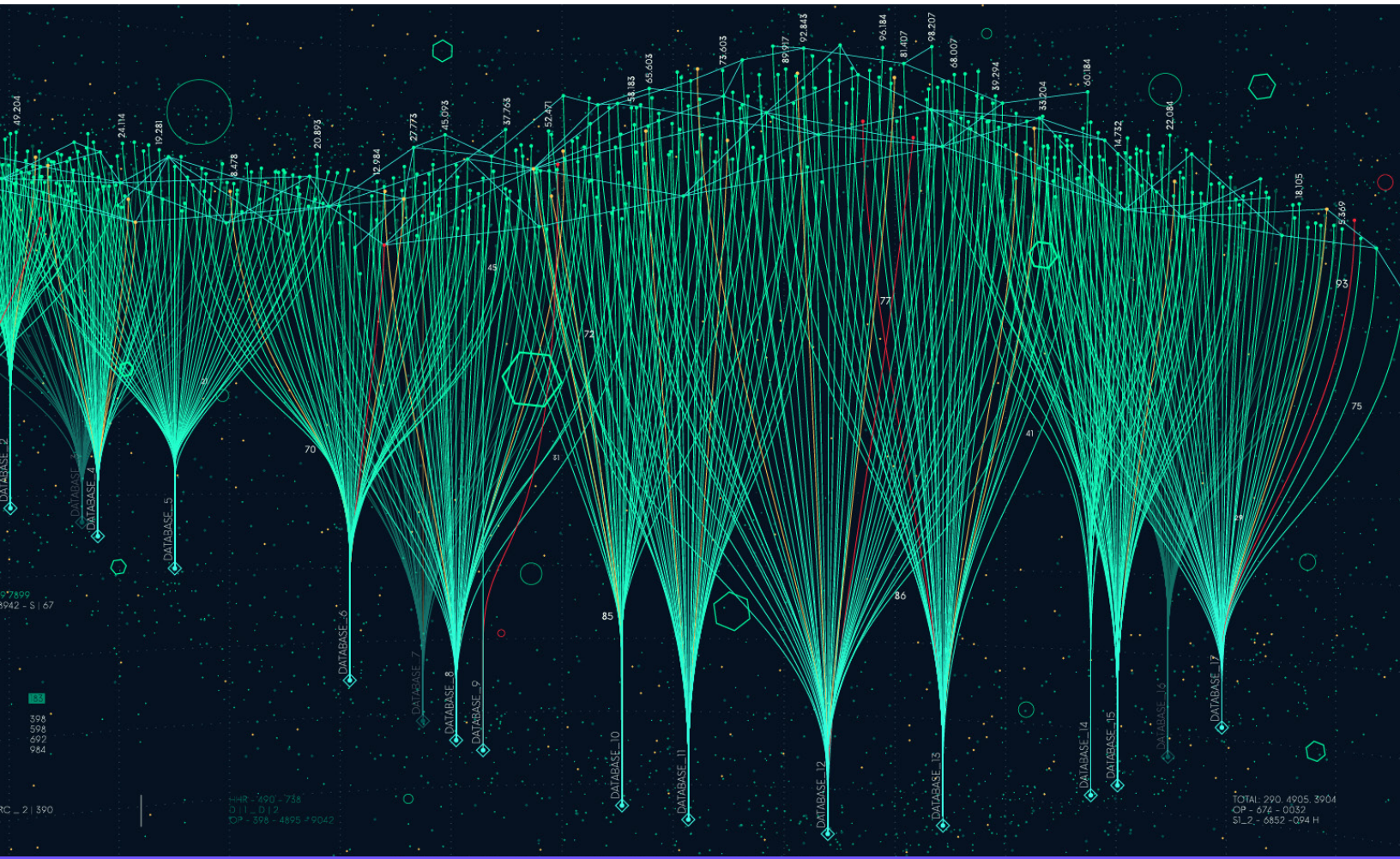
Benchling

CASE STUDY

Teaming up with a trailblazer to develop next-gen drug discovery



ANAGENEX

GOAL

Make drug discovery faster and more efficient by laying a solid data foundation to fuse large scale experimental data collection and machine learning

Anagenex is a drug discovery company combining DNA encoded library (DEL) technology with machine learning (ML) to achieve their vision of finding a drug candidate for every disease cheaply and quickly. Their goal is to radically accelerate the traditionally time- and labor-intensive portion of drug discovery with a fast and powerful approach based on combinatorial chemistry, machine learning and rapid, large scale iteration. They've relied on Benchling from day one, capturing every intermediate data point to ensure that their computational systems get a complete picture of lab processes.

COMPANY PROFILE

Number of employees: 1-10

Industry: Small Molecule, Biopharma

Location: Bay Area, CA

KEY RESULTS

100%

of scientists agree that Benchling has helped connect bench and ML work together

50%

decrease in time spent on data entry and cleanup

200%

estimated increase in speed due to starting with a robust data infrastructure

"I don't think anybody's put the pieces together quite the way we are, bringing both the lab and compute under the same roof and designing both types of experiments to take advantage of the other's unique capabilities or pitfalls from the very beginning. But it also means we have to build a special kind of company. Lab and compute have to be true partners, trusting one another, and making each other stronger. We focus on what software can do to serve lab needs and make the research process more efficient, more productive and more pleasant."

Nicolas Tilmans, CEO

CHALLENGES

Edge Use Cases

Treating DNA sequencing readouts as an intermediary step to understanding the molecule of interest is an indirect use case, creating an additional layer of complexity.

Batch Effect Identification

Each experiment can introduce batch effects that mask the biologically relevant effects and confuse the machine learning process.

Scalable Solution

Anagenex needed a robust solution that could start tracking from the very first experiments and scale quickly as the company grows.

THE STORY

Today's typical small-molecule development processes start with high throughput experimentation but quickly collapse to a low throughput iterative process. After identifying a target, scientists will sample collections of up to a few million compounds to find possible compounds that might interfere with the target of interest. If they can pinpoint a promising compound, a medicinal chemist will synthesize a few dozen variations of that molecule, testing those in more sophisticated assays. After several iterations of synthesizing dozens of variants at a time, hopefully they will have identified a compound that performs well enough to administer to an animal and eventually a human. This process has benefitted from automation; however, it often remains relatively slow, taking many years between compound discovery and clinically enabling studies.

Nicolas Tilmans, CEO, founded Anagenex to radically increase the throughput and reliability of drug discovery. Drawing from his background as both a biochemist and data scientist, he's using a technique combining DNA-encoded libraries (DELs) and machine learning (ML) to reliably identify far more promising molecules for a given target, streamlining the process of downstream compound optimization. These two technologies rely on large scale datasets, and combining them in novel ways requires a highly flexible and scalable data infrastructure. Nicolas chose Benchling to make sure Anagenex would have a digital lab solution that could keep up with its team's fast pace and cutting edge science.

Establishing the correct relational schema in Benchling tames the complexity of life science workflows, even with atypical use cases

DELs work by constructing potential binders in a stepwise process. Chemical building blocks and accompanying DNA sequence tags are added in alternating steps to create a the potential binder, a single compound. In relatively few steps, a massive collection of hundreds of millions of compounds can be constructed within a single tube. Each compound in the mixture is associated with a unique DNA sequence. The compound mixture is affinity selected against a target, retaining only tight binders to the target, which are then identified using next-gen sequencing (NGS). NGS reads out the DNA tags added to each step, which serves as a breadcrumb trail to identify the building blocks that created the successful tight binder.

This approach to handling DNA sequences - a pathway to deconvoluting the molecule of interest rather than being the endpoint of the experiment - is unusual. The NGS data needs to be combined with the original experimental data from library generation to decode a useful dataset. This departure from the usual focus on sequencing as an end goal might be too complex for legacy data management systems to handle. However, Nicolas was not worried - his team could create a custom data schema in their Benchling Registry that mapped the parent-child relationships between DELs, the DNA tags identifying the building blocks of each DEL, and the pools each DEL was added to. The team never had to worry about identifying the tight binder - even when hundreds of millions of possible binders were mixed with the target of interest - because they had very clear records with which to decode DNA sequencing readouts.

Standardized data fields let the team collect data for more parameters more completely, helping machine learning find biologically-relevant insights rather than batch effects

Anagenex uses ML to help direct the next round of DELs to be built by identifying binding trends in previous pools. However, machine learning is agnostic to whether it identifies a biologically relevant trend, or simply an experimental artifact, such as whether a molecule was located in the middle or the sides of a plate. Because Anagenex's science is so bleeding edge, the team is constantly adding to their knowledge of what matters for DEL screening in a machine learning context, and it's essential to keep thorough experimental records to track down any confounding variables.

The team equips their ML models with as much data about experimental information as possible to tease out what trends are batch effects and what are trends to act on. Time of runs, reagents used, readout method, etc., may all be important. Anagenex takes advantage of templatized registration

in Benchling to lower friction to data entry, keeping scientist compliance up and data records clean and complete. Furthermore, standardized data entry speeds up their ML workflows despite high volumes of data because no extra data cleaning is required.

Using a data-first software like Benchling matches Anagenex's computation-first business model and can scale with Anagenex as they grow

Benchling is a solution that can keep up as Anagenex grows. All data is collected, stored, and accessible, no matter if it's from a simple feasibility experiment or a highly complex experiment in a multi-step workflow. Nicolas recognizes that data is Anagenex's most precious asset because it is their solution to finding higher-quality drug candidates faster than their competitors.

In the highly competitive field of drug discovery, Anagenex needs to set itself apart from its peers by demonstrating a considerable increase in quality. To do so, Nicolas built Anagenex with software engineering and compute capabilities as a core part of the company. Like Benchling, Anagenex invests in its software solutions because they see them as a cornerstone of their business.

Anagenex is just hitting its stride. Their technology holds the promise of a highly effective drug discovery pipeline that could bring about a sizable impact in numerous disease areas. The combination of new chemical approaches with new computational tools could become the status quo of drug discovery, with Anagenex leading the transition to this new generation.

"We currently have four people working in the lab, and we're investing a fairly large amount of capital into Benchling because we believe that every last data point must be collected. We're asking data questions no one has asked because no one really could before. But that also means we must track all of our steps very carefully to catch minor variations or data drift and make sure the machine learning models don't lie to us."

Nicolas Tilmans, CEO