Finding Improved Biologics with Twist's Machine Learning and Deep Learning Tools

Generally speaking, the more lottery tickets you buy, the better your chances of winning. This concept is true in biologics development, too: The more opportunities you get to design and test a candidate sequence, the better your chances of discovering one of therapeutic value. In reality, however, practical considerations significantly limit the number of opportunities researchers get. As such, the biologics discovery and development process requires difficult decisions about which candidates to synthesize, which to advance for further development, and which to leave behind. The stakes are high. Fortunately, Twist Biopharma has developed machine learning algorithms to help researchers see hidden patterns in their data and, in so doing, select the candidates that are most likely to succeed.

THE NUMBERS GAME OF BIOLOGICS DEVELOPMENT

In an ideal world, researchers would be able to synthesize and test every potential therapeutic candidate with equal rigor. However, practical limitations prevent this.

Consider a typical antibody which has approximately 60 CDR amino acids that are critical for target binding. To hone in on the most clinically valuable combination of amino acids, you may want to test every potential amino acid combination in each CDR. In this scenario, you would need to synthesize and test nearly 10^{80} unique sequences—that's more atoms than exist in the universe. Such a task is not physically possible. Most protein libraries max out at $10^{10–12}$ unique sequences in practice.

Researchers must therefore design their antibody libraries carefully in order to enrich therapeutically valuable sequences. After library panning has taken place, traditional sampling of candidate pools for screening is limited to highly abundant clones.

Next Generation Sequencing (NGS) has enabled researchers to dig deeper into their candidate pools providing vast numbers of therapeutic candidates to be mined in silico. Twist's high-throughput DNA synthesis platform is perfect to enable customers to make and test these antibodies.

After panning and screening, similar decisions will be needed to determine which candidates advance to more complex and costly preclinical *in vivo* and toxicology studies prior to development and manufacturing required for clinical trials.

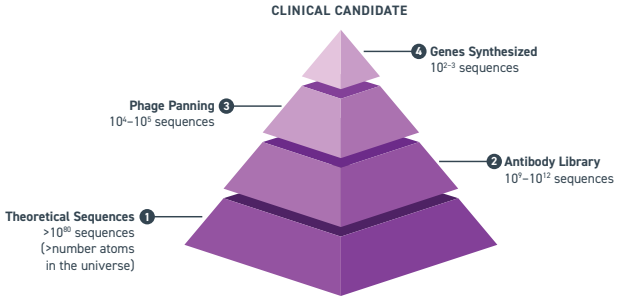

Figure 1: Illustration showing the narrowing capabilities of the biologics discovery process. Machine learning can help researchers maximize rational sampling of candidates at each stage.

THE VALUE OF MACHINE LEARNING

Machine Learning and Artificial Intelligence (AI) promise to help researchers analyze larger amounts of data and perform non-obvious pattern recognition. This allows for identification of more diverse candidate pools that are enriched for features that enhance therapeutic value. Such an *in silico* approach gives researchers the broadest possible pool of candidates, providing more opportunity for success.

TWIST MACHINE LEARNING IN BIOLOGICS DEVELOPMENT

Twist Biopharma has developed machine learning tools to support or help de-risk biologics discovery and development.

- AI assisted library design ensures that the initial antibody library samples a diversity of the sequencing space without wasting resources on high-liability, low-developability sequences.
- An unsupervised machine learning tool supports lead picking by analyzing NGS data from panning campaigns. This tool is designed to maximize sequence diversity going into downstream screening.
- New Machine Learning methods have been developed to assist with lead optimization, for example humanization to reduce immunogenicity.
- Machine Learning algorithms have also been trained to identify which candidates are likely to have superior biophysical properties such as reduced hydrophobicity.

These machine learning tools help Twist be the one-stop-shop for biologics development. Twist's industry leading DNA synthesis platform enables large-scale library synthesis which integrates with Twist Biopharma's phage display offering, IgG synthesis platform, and antibody optimization. Our platform enables researchers to design, build, test, and succeed all in the same place.

CASE STUDY*: VHH HSHUFFLE HYPERIMMUNE

Twist aimed to isolate multiple VHH nanobody leads against a soluble target protein. First, a VHH library was designed with diversity included in CDR1, CDR2 and CDR3. In particular, the VHH's CDR3 loop included more than two million possible CDR sequences. After three rounds of panning, a candidate pool was sequenced using the newest generation NGS technology.

NGS sequencing from each round of biopanning enabled the use of both unsupervised and deep learning approaches to identify candidates for screening. >500 candidates were selected and included >100 CDR3 sequences indicating a diverse pool of candidates. Expression of clones with Twist High-Throughput Antibody production platform enabled downstream testing of VHH binding affinity by SPR.

Several high affinity binders were identified using NGS + Unsupervised and Deep Learning workflows. Importantly, some of the highest binding affinities were observed among the rare clones that would be missed during typical screening workflows.

In short, Twist's machine learning tools enabled maximum extraction of hits from discovery campaigns and ultimately identification of candidates with higher likelihood of success.

**Results from case studies are not predictive of results in other cases. Results may vary in other applications.*

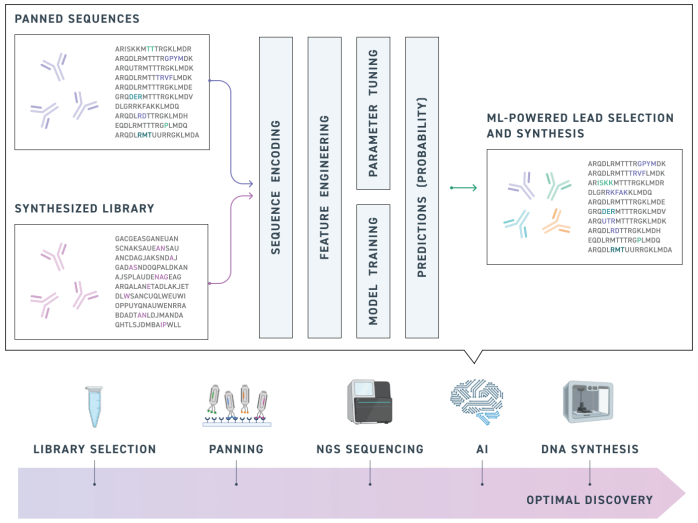

Figure 2: Workflow for antibody lead selection using machine learning.

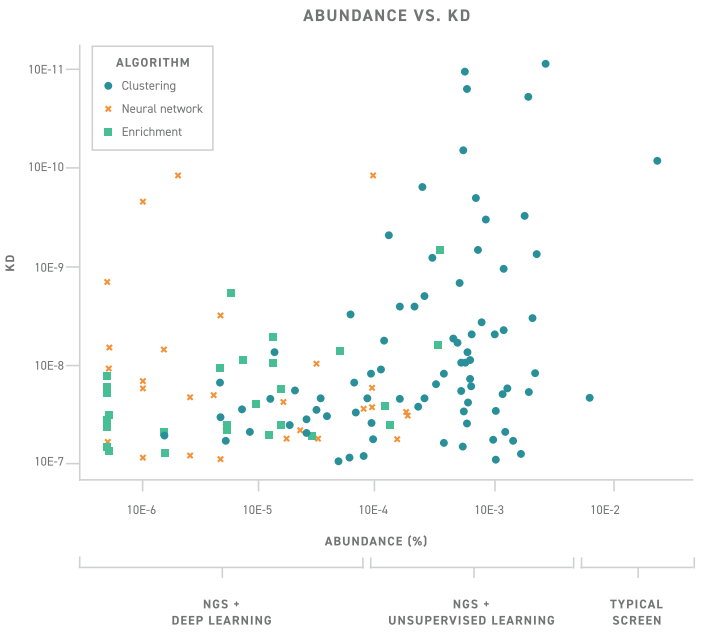

Figure 3: Plot showing the binding affinity (KD, y-axis) and abundance (x-axis) of candidates as well as the methods used to identify those candidates.

TALK TO US ABOUT YOUR NEXT PROJECT.

Contact the Twist Bioscience team at **support@twistbioscience.com** or visit **twistbioscience.com**