Sequencing to Synthesis: Precision Antibody Discovery Leveraging Machine Learning to Prioritize Leads

Ilaria DeVito[1], Crystal Richardson[1], Anthony Moore[1] , Zain Alvi[1], Pranay Vishwanath[1], Brian Sereni[1], Suhani Nagpal[2], Amy Migliori[2] , M. Frank Erasmus[3], Laura Spector[3], Michael Karney[1], Sumit Kumar[1], Laure Turner[1]
[1]GENEWIZ by Azenta Life Sciences, South Plainfield, NJ 07080; [2]OpenEye, Cadence Molecular Sciences, Santa Fe, NM 87508; [3] Specifica LLC, a Q[2] Solutions Company, Santa Fe, NM 87501

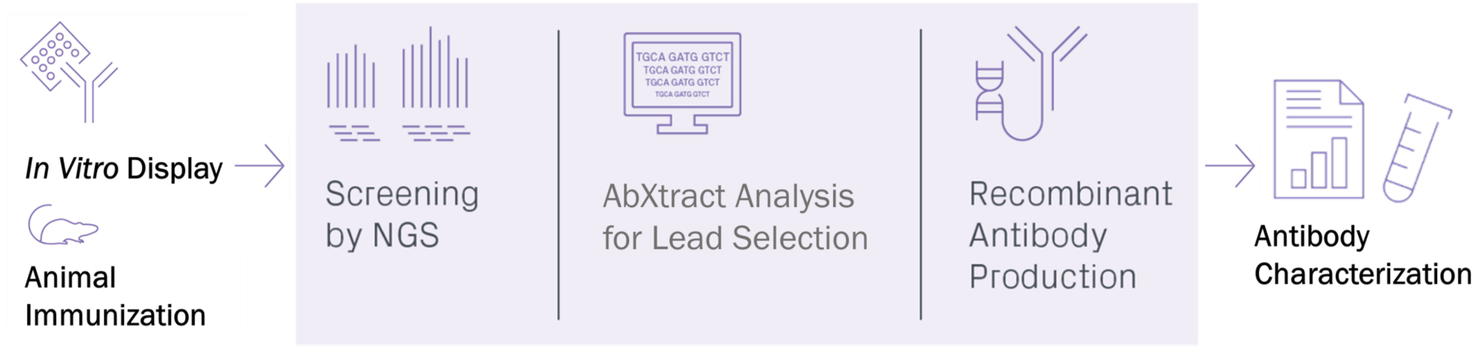
Abstract

Current industry practice for antibody (Ab) generation and validation can take months, involving extensive wet lab labor and resource usage. Even after an elaborate Ab campaign, the candidates often still have developability issues that can result in late-stage failure. This inefficient process is laborious and costly, thus there is a need for new ways to discover Ab lead candidates with precision and efficiency.

Here we present an innovative end-to-end Ab discovery solution combining the strengths of *next-generation sequencing (NGS)* and *computational analysis* that allows the rapid identification and expression of antibody sequences with a broad affinity range for the target and the most promising biophysical profiles for commercialization. The novel bioinformatic platform, AbXtract™, powered by OpenEye and developed by Specifica, enables end-to-end high-throughput screening of antibody library to generate 5 – 50x more antibody leads than traditional low-throughput approaches.

GENEWIZ's workflow increases success of antibody screening by identifying a greater number of clonotype leads, unlocking rare clones, and prioritizing leads with decreased developability liabilities.

Schematic Representation of Innovative GENEWIZ Precision *In-Silico* Antibody Screening



Methods

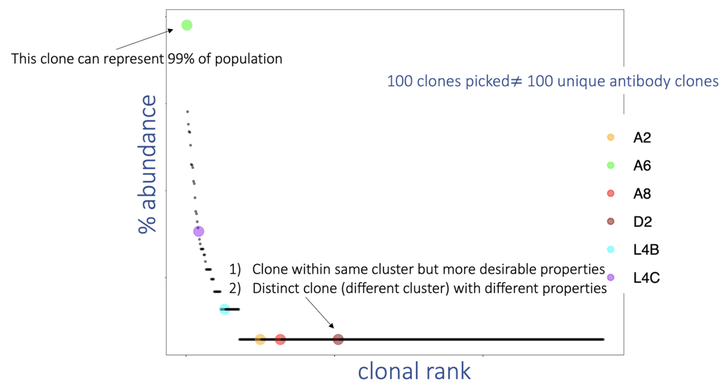


Figure 1. Antibody clones identified by NGS & computational analysis. Antibody discovery approaches that only rely upon random colony picking are often insufficient in their exploration of the underlying diversity. NGS & computational processing can guide lead cluster & candidate prioritization to 1) overcome situations where diversity is dominated by a few clones, 2) to use NGS statistics to rank order leads within and across clusters.

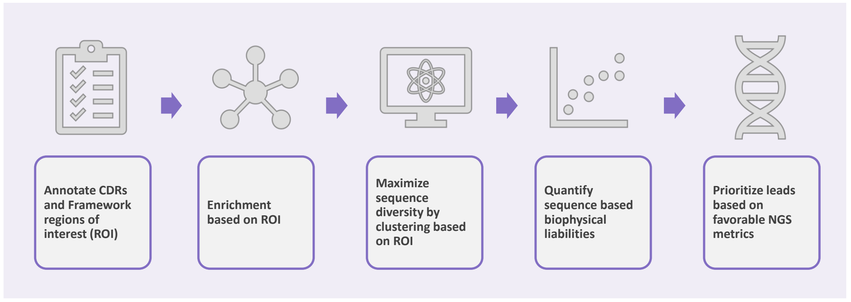


Figure 2. AbXtract Workflow- The bioinformatics pipeline to allow for prioritization of optimal leads starts with quality filtering of input NGS data followed by annotating the sequences to identify regions of interest (ROI) and extract features. Relative abundance and enrichment based on ROI can be calculated, the module then identifies abundant and rare clusters with unsupervised ML based on density-based clustering and quantifies sequence-based biophysical liabilities. The platform prioritizes the leads based on favorable NGS metrics to provide a gene synthesis-ready output.

Results

AbXtract identified ≥1000 antibodies (Ab) candidates selected against SARS-CoV-2 Spike trimer protein, its monomer S1 and the receptor binding domain (RBD). Utilizing PacBio NGS followed by *in silico* AbXtract analysis and recombinant Ab production, 200 Ab were produced as IgG and recognized receptor binding domain (RBD). The RBD binding affinities of 143 Abs ranged from 34 pM to 1µM, with 30 Ab better than 100 pM, as measured by surface plasmon resonance (Figure 2). The Ab selected with the traditional colony screening method (pink dots) had overall worse affinity when compared to clones identified by AbXtract (blue dots).

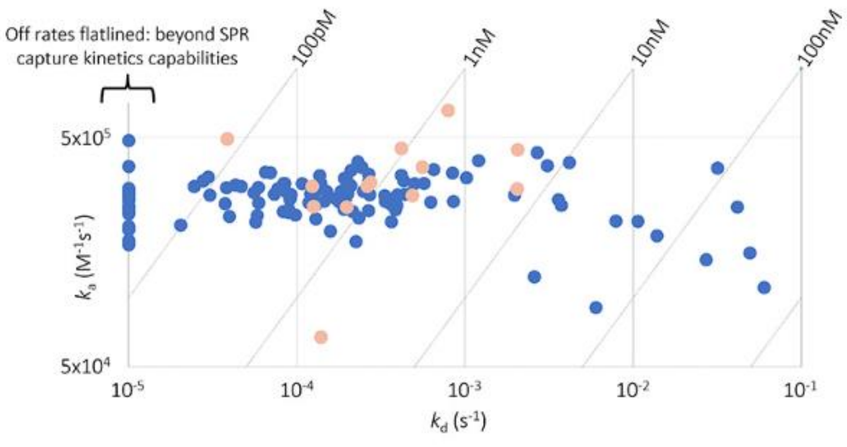


Figure 3. Clones and their affinities by traditional screening method versus NGS Isoaffinity plot of 13 picked clones (Pink Dots), or 143 clones identified by AbXtract synthesized, expressed, and purified (Blue Dots). Affinities are indicated by the diagonal lines.

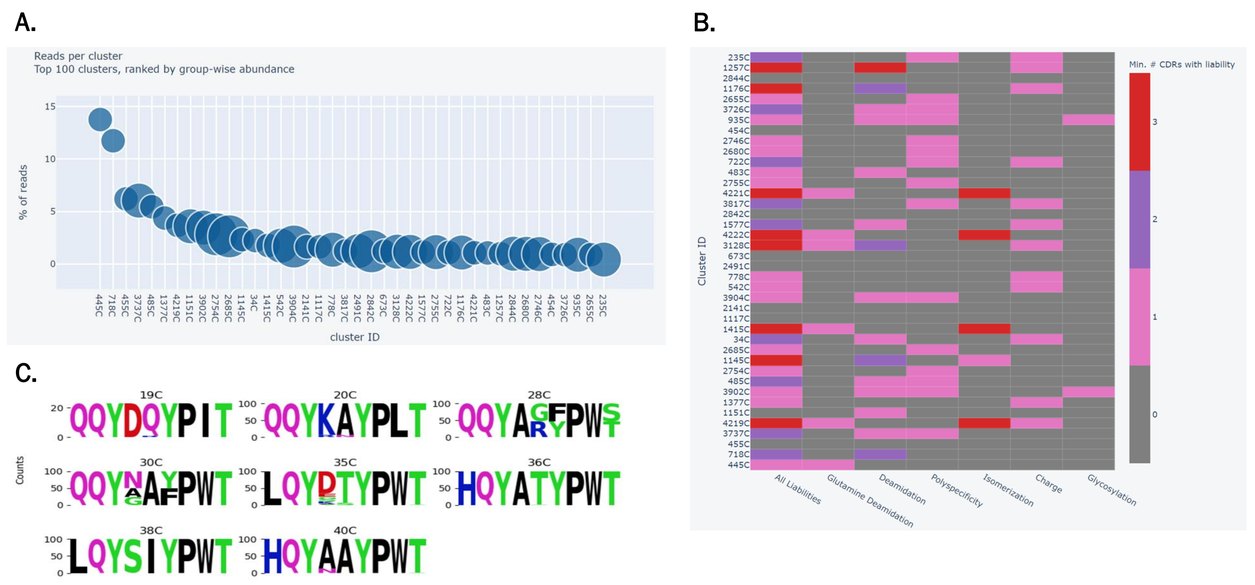


Figure 4. AbXtract Report. (A) Percentage of NGS Reads versus Cluster. The figure highlights the reads per cluster for the top 100 clusters that have been ranked by group-wise abundance. A group refers to one sample of a multiplexed project. The size of blue circle indicates the unique clones within each cluster. (B) Cluster IDs versus CDR Liabilities. This chart shows the minimum number of CDRs with sequence liabilities across each clustered sequences. The top 100 clusters are ranked by group-wise abundance. Furthermore, the clusters with less liabilities are predicted to provide a better candidate for antibody development and production. (C) Consensus sequences for the LCDR3 by cluster for the top 100 full-length sequences. LCDR3 confers binding affinity, and these sequences can be used for design of antibody candidates for gene synthesis and production.

Advantages

The workflow we presented compared with traditional colony screening methods:

- **Uncover more leads**: Increase the number of leads 5X to 10X compared to random colony picking.
- **Greater library diversity:** NGS, with its deep level of sequencing information, allows to more comprehensively explore the complete pool of antibodies. Even rare clones are identified that are typically missed by low-throughput methods.
- **Decrease sequencing liability in downstream development:** NGS combined with computation processing can help reduce sequence liabilities to improve efficiency of downstream antibody production.
- **Faster antibody discovery/ screening:** Identify top leads with high target affinity and beneficial development properties with less rounds of screening.
- **Convenience of one partner:** Partner with GENEWIZ to allow your antibody discovery engineers and scientists to focus on research and development of subsequent pipeline for antibodies therapies.

Contacts & Additional Info

To learn more or to request a project consultation and a quote please reach out to technical team members:
Dr. Anthony Moore, Ph.D. anthony.moore@azenta.com or
Dr. Crystal Richardson, Ph.D. at crystal.richardson@azenta.com

Scan the QR Code for More Info

References

1. Erasmus, MF et al. Insights into Next Generation Sequencing Guided Antibody Selection Strategies. To Be Submitted.
2. Ferrara, F, Erasmus, M.F., et al. A pandemic-enabled comparison of antibody discovery platforms: a naïve library directly delivering antibodies as potent as immune sources. Nature Comm. 13:462 (2022).
3. Teixeira, A., Erasmus, M.F et al . Drug-like antibodies with high affinity, diversity and developability straight from next-generation antibody libraries. MAbs vol 13, 1 (2021).

GENEWIZ® From Azenta Life Sciences

OpenEye SCIENTIFIC — CADENCE MOLECULAR SCIENCES

SPECIFICA a Q² Solutions Company