

## NOVEL TECHNOLOGIES AND APPROACHES FOR OPTIMIZING ANTIBODY DISCOVERY

By Dr Catarina Carrao

*This piece represents the views of the author and not necessarily the views of Informa Connect Life Sciences or the Antibody Therapeutics and Engineering conference series.*

### INTRODUCTION

Successful antibody discovery has been dependent on the proper identification of a target and the mechanisms of action associated to the development of a specific disease. But, this conventional wisdom of the antibody-development roadmap is being challenged, by how long it takes from target research to the initiation of a clinical program. As the search for validated targets becomes more challenging, unconventional strategies, such as phenotypic screening, are becoming more significant.

### Phenotypic Screening: The Function-FIRST Antibody Discovery Platform (F.I.R.S.T™)

In a “function-first” approach of new antibody discovery, or phenotypic screening, the first step is targeting a specific cell type over others; for example, a malignant cell over its normal equivalent<sup>1</sup>. By using primary cells from patients, the phenotypic discovery of antibodies allows the development of personalized drugs. Antibodies against targets are then screened in clinically and mechanistically relevant, high-throughput, functional *in vitro* assays<sup>1</sup>. Finally, antibody targets are deconvoluted, and the most biologically and therapeutically promising antibodies are tested in state-of-the-art *in vivo* experimental models<sup>2</sup>. This strategy allows the identification of antibodies that mediate their mechanism of action through the biology of

the target receptor; and, also, through effector mechanisms (e.g., Fc-receptor binding), and a further engagement of immune effector cells.

Recently, Anne Ljungars and colleagues<sup>3</sup> from the department of Immunotechnology in Lund university (Sweden), have developed a platform, the function-FIRST antibody discovery platform (F.I.R.S.T™) for phenotypic discovery of therapeutic antibody-target combinations. By using Chronic Lymphocytic Leukaemia (CLL) primary cells throughout the discovery process (panning, screening, and functional studies), they discovered antibodies with improved cytotoxicity compared to the standard of care (*Rituximab*, a CD20-specific monoclonal antibody)<sup>3</sup>. The isolated antibodies were found to target six different receptors on CLL cells (e.g., CD32, CD200, and HLA-DR), which appeared to be the most potent targets for antibody-based cytotoxicity treatment<sup>3</sup>. One of the antibodies discovered by this study is currently being tested in phase I/II clinical trials on CD32b+ B-cell lymphomas and leukemias, showing that phenotypically discovered targets are translatable to the clinic (NCT02933320, NCT03571568).

### The Challenges of Next Generation Sequencing

The arrival of high-throughput sequencing brought an understanding to antibody repertoires and their dynamics<sup>4</sup>. There are an increasing number of

computational tools that aid in the immune repertoire characterization, like genome editing<sup>5</sup> and mammalian surface display<sup>6</sup>. This not only enables the screening of full-length antibodies for antigen binding, but also allows for the integration of immunoglobulin genes to predict biophysical properties optimal for downstream therapeutic development<sup>7</sup>.

Next generation sequencing (NGS) has allowed a massive increase in capacity, a relative low-cost and a short time analysis of thousands of antibody targets. One of the main challenges in the use of NGS for antibody selection relates to the size of the encoded genes. The smallest antibody fragments (variable domains) range between 300-400 bp in length, while the commonly used scFv and Fab antibody fragment formats range from 700/800 bp to >1,500 bp, respectively<sup>8</sup>. There are many NGS platforms that can sequence up to 300–400 bp with reasonable throughput, like the Illumina Miseq and Hiseq<sup>9</sup>, or Ion Torrent PMG<sup>10</sup>. Using paired-end reads can also generate long sequences, which is particularly useful for scFv formats, enabling the sequencing of multiple Complementarity-Determining Regions (CDRs) of VH and VL domains<sup>8</sup>. Some studies have focused on the center of the antigen binding site, which is a major determinant of antigen binding; as such, focusing on the relatively short VH CDR3 repertoire only<sup>11</sup>.

Another challenge of the use of NGS is paying attention to sequencing errors; because DNA amplification inevitably results in context-dependent polymerase errors ( $10^{-5}$ – $10^{-6}$  per base), and errors will certainly be present in large NGS datasets that include billions of bases<sup>8</sup>. Not only that, but NGS technologies themselves can be vulnerable to the introduction of errors, such as cluster misamplification and base misincorporation<sup>4</sup>. To help identify PCR and sequencing errors, unique molecular identifiers (UMIs; stretches of 8–10 degenerate DNA bases) can be added to primers during the first two cycles of PCR amplification, meaning that reads that share the same UMI have a high probability of being derived from the same original template<sup>12</sup>. Such reads can then be grouped and used for error correction.

Following confirmation of the quality of the NGS read-data, the large sample size of NGS method logically requests the use of bioinformatics tools. The selection of a tool will be dependent on the number of NGS reads being analyzed. Programs such as IMPRe, IgBLAST, IMGT/High V-QUEST, and ImmundiveRsity, are used to identify V<sub>H</sub> and V<sub>L</sub> germlines as well as V<sub>H</sub> and V<sub>L</sub> CDRs<sup>8</sup>. Post-processing of the output of the bioinformatics tools is then needed to

understand the clone structure within the dataset, and to pair V<sub>H</sub> and V<sub>L</sub> domain sequences. The clone structures can be determined by applying sequence clustering tools, such as CD-hit<sup>13</sup> or UCLUST<sup>14</sup> to CDR3s alone. This can be done at either the amino acid or nucleotide sequence level (or, even to the full-length sequence), in order to group closely related sequences into a “clonal” groupset<sup>8</sup>. Finally, the scripts can be used to examine and review the diversity and other characteristics of the library. Fortunately, nowadays it is actually possible to outsource the sequencing and/or analysis of antibody libraries to commercial suppliers, with companies offering a range of options from multiplexing/sequencing to complete analysis from purified DNA or phages (e.g., CD Genomics, MCLAB).

## ‘Yin-Yang’ Panning: “Finding the Needle in the Haystack” of Antibodies

Once the library has been established, then the development of computational methods to aid antibody engineering relies on the successful exploration and exploitation of these new data sources. A field that currently benefits from a steady stream of new data from NGS is the B-cell Receptors (BCRs), that can be used as proxy for antibodies<sup>15</sup>. Recently, Leonard D. Goldstein and co-workers<sup>16</sup> from Genentech in San Francisco (USA), described the use of high-throughput single-cell B-cell receptor sequencing (scBCR-seq) to obtain accurately paired full-length variable regions in a massively parallel fashion. Besides of identifying additional candidate lineages, not recoverable as hybridomas; they found that 99% ( $n = 92/93$ ) of the clones were antigen-reactive, establishing scBCR-seq as a powerful tool for antibody discovery<sup>16</sup>.

The isolation of thousands of antibodies that might contain specificity for all therapeutically relevant cell surface receptors makes biopanning a crucial step to further identify specific antibodies. New strategies are being developed to facilitate the biopanning procedure. For example, panning in solution using biotinylated antigens seems to offer advantages in comparison to using immobilized ones<sup>17</sup>. Chia Chiu Lim and colleagues<sup>18</sup> from the Institute for Research in Molecular Medicine in Malaysia, recently presented a new process coined ‘Yin-Yang’ panning by using saturation of non-binders with blocking agents to promote positive binder enrichment with

antigen preparations devoid of purification, allowing the use of crude antigens for further selection. Also, using whole cells for panning appears to be a good way to present membrane embedded target antigens in their natural environment, and conformation, to antibody-bearing phages<sup>19</sup>.

Since, in the panning process, with its consecutive selection rounds, there is a tendency to mainly identify the most abundant antibodies, independent of functionality; there's also a need to facilitate discovery of low-abundant antibodies and, when appropriate, additional specificities. With this in mind, deep mining strategies have recently been used to uncover rare antibodies that bind to new targets and epitopes<sup>20</sup>. For example, a protein depletion step can be used as a deep mining strategy to reduce the fraction of antibodies binding to previously discovered hits and remove low therapeutic interest targets<sup>21</sup>. NGS can also be used as a deep mining tool for analysis of low frequency clones<sup>20</sup>, and provide an effective way to retrieve antigen-specific scFv clones that are difficult to identify by the conventional colony screening method<sup>22</sup>.

## Bioinformatics Tools: An Essential for Antibody Optimization

The high-throughput screening of antibodies needs to be done with clinically predictive *in vitro* and *in vivo* assays. Despite the range of *in vitro* measures to characterize physicochemical properties of bio-therapeutics, "*in vitro-in vivo*" correlations (IVIVC) have yet to be established<sup>23</sup>. As such, implementing tools during early phases of drug discovery to identify antibodies at risk of poor pharmacokinetics, for example, will eventually reduce the time needed by improving the lead antibody selection process. Simple *in silico* tools can be used to evaluate the presence or absence of certain molecular features, such as net charge and hydrophobicity of critical parts of antibodies, which can then be followed by more complex 3D modeling assessment of hot spot analysis, isoelectric point, or charge distribution<sup>23</sup>. Also, *in vitro* high throughput tools addressing non-specific binding (e.g., Baculovirus-ELISA, heparin binding, or extracellular matrix binding assay), binding to FcRn (human FcRn liquid chromatography) and poly-specificity (e.g., protein chips) can be used in a pharmacokinetic de-risk strategy<sup>24</sup>.

Good solubility is also crucial to avoid *in vivo* antibody

aggregation, which can lead to loss of activity and further degradation of the bio-therapeutic. Aggregation prone regions (APRs) around the CDRs of therapeutic antibodies can be predicted by algorithms<sup>15</sup>. Recently, Puneet Rawat and co-workers<sup>25</sup> from the Protein Bioinformatics Lab in the Indian Institute of Technology, collected experimental data on aggregation kinetics and used machine learning to identify protein aggregation rate enhancer and mitigatory mutations. This new algorithm enables *in-silico* optimization of antibodies with an eye on improved biophysical features<sup>25</sup>.

The identification of hydrophobic regions is also an important step in aggregation prediction that ideally requires a crystal structure of the antibody or a reliable homology model<sup>15</sup>. Tushar Jain and colleagues<sup>26</sup> from Adimab in Palo Alto (USA), developed a surface accessible area predictor that can be applied to an antibody sequence, to further create a propensity score that could be correlated with aggregation risk<sup>26</sup>. The low computational cost and high accuracy of this approach, allows real-time assessment of the hydrophobic character of an antibody. This not only supports the antibody ranking during the discovery process, but also a rational engineering approach to reduce hydrophobic liabilities<sup>26</sup>.

## The Use of Machine Learning Techniques for Target Deconvolution

One of the difficulties associated with phenotypic drug-discovery, in comparison with target-based, is that it requires downstream identification of targets engaged by the newly identified antibodies – called target deconvolution. This is an essential step to understand the compound mechanism of action, and for using the identified hits as tools for further dissecting a given biological process<sup>27</sup>. Since, this is a difficult and labour-intensive step, a cell binding heat map analysis could be used to group antibodies with the same specificity together, simplifying the target deconvolution process<sup>3</sup>. This allows a reduction in the number of analysed antibodies that must be investigated in low throughput, high-cost techniques, such as immunoprecipitation-mass spectrometry (IP-MS).

The evaluation of an antibody fragment-induced activation of major histocompatibility complex class II (MHCII)

molecules is also an important step in bio-therapeutics discovery<sup>15</sup>.

Computational methods can be used to identify potential MHC I and MHC II binding T-cell epitopes, as well as conformational B-cell epitopes<sup>28</sup>. The prediction of T-cell epitopes by machine learning approaches is tackled mostly by neural networks-based methods, that evaluate the binding affinity of a given short peptide towards MHC-I or II<sup>29</sup>. Recently, Masato Ogishi and Hiroshi Yotsuyanagi<sup>30</sup> from the Institute of Medical Sciences Research Hospital in the University of Tokyo (Japan), proposed a supervised machine learning framework that generates probabilistic estimates of immunogenicity, termed “immunogenicity scores”, based on computed sequence-based simulation. This method mimics the thermodynamic interactions between MHC complexes and T-cell receptor clone-types, which enables an efficient prioritization of epitope candidates highly prone to T-cell recognition in multiple individuals<sup>30</sup>.

## Nanopens and Nanobodies, the Future is Here

Accelerated development of antibody tool reagents is also an essential requirement for the successful progress of therapeutic antibodies in today’s fast-paced and competitive drug development marketplace. Aaron Winters and colleagues<sup>31</sup> from Amgen in California (USA), have recently reported a nanofluidic optoelectronic single B lymphocyte antibody screening technique (NanOBlast) applied to the generation of anti-idiotypic reagent antibodies.

This platform is capable of manipulating living cells within microfluidic chip nanopens using structured light, capable of quantifying free therapeutic IgG molecules directly from human patient serum, which facilitates important bio-therapeutics development decision-making<sup>31</sup>. New approaches to develop antibody-based therapeutics are also focused on different molecular formats, like the H-chain only antibody, or nanobody, that are naturally occurring in camelids and sharks<sup>15</sup>. The first therapeutic approval in 2018 of *Caplacizumab*, has raised the interest in these type of molecules. *Caplacizumab* is a humanized anti-von Willebrand factor, bivalent variable-domain-only immunoglobulin fragment. It inhibits the interaction between von Willebrand factor multimers and platelets, for the

treatment of thrombotic thrombocytopenic purpura (TTP), an immune-mediated deficiency disease<sup>32</sup>. The results of the phase III clinical trial have just come out, showing promising results when added to the standard treatment of affected patients.

## Conclusion

The annual number of antibody-based therapeutics to receive approval in either the European Union or the United States has reached double digits for the first time in 2017<sup>31</sup>. The specific binding characteristics of antibodies, together with their compliance to protein engineering make these molecules powerful research tools, and the perfect vehicle for a more straightforward road to the discovery of new medical treatments.

## ABOUT THE AUTHOR

Dr Catarina Carrao gained a PhD in Biochemistry from Northeastern Ohio Medical and Pharmacy University and an M.Phil in Biochemistry at the University of Beira Interior. She has worked as a researcher at Max F. Perutz Laboratories, Yale Cardiovascular Center at Yale University School of Medicine, and the Center Cardiovascular Research (CCR) at Charité Medical University.

## REFERENCES

1. Frendeus B. Function-first antibody discovery: Embracing the unpredictable biology of antibodies. *Oncoimmunology*. 2013;2:e25047.
2. Veitonmaki N, Hansson M, Zhan F, Sundberg A, Lofstedt T, Ljungars A, Li ZC, Martinsson-Niskanen T, Zeng M, Yang Y, Danielsson L, Kovacek M, Lundqvist A, Martensson L, Teige I, Tricot G and Frendeus B. A human ICAM-1 antibody isolated by a function-first approach has potent macrophage-dependent antimyeloma activity in vivo. *Cancer Cell*. 2013;23:502-15.
3. Ljungars A, Martensson L, Mattsson J, Kovacek M, Sundberg A, Tornberg UC, Jansson B, Persson N, Emruli VK, Ek S, Jerkeman M, Hansson M, Juliusson G, Ohlin M, Frendeus B, Teige I and Mattsson M. A platform for phenotypic discovery of therapeutic antibodies and targets applied on Chronic Lymphocytic Leukemia. *NPJ Precis Oncol*. 2018;2:18.
4. Lopez-Santibanez-Jacome L, Avendano-Vazquez SE and Flores-Jasso CF. The Pipeline Repertoire for Ig-Seq Analysis. *Front Immunol*. 2019;10:899.
5. Bradford J and Perrin D. A benchmark of computational CRISPR-Cas9 guide design methods. *PLoS Comput Biol*. 2019;15:e1007274.
6. Parola C, Neumeier D, Friedensohn S, Csepregi L, Di Tacchio M, Mason DM and Reddy ST. Antibody discovery and engineering by enhanced CRISPR-Cas9 integration of variable gene cassette libraries in mammalian cells. *MAbs*. 2019;11:1367-1380.
7. Pogson M, Parola C, Kelton WJ, Heuberger P and Reddy ST. Immunogenomic engineering of a plug-and-(dis)play hybridoma platform. *Nat Commun*. 2016;7:12535.
8. Rouet R, Jackson KJL, Langley DB and Christ D. Next-Generation Sequencing of Antibody Display Repertoires. *Front Immunol*. 2018;9:118-118.
9. Guo Y, Chen K, Kwong PD, Shapiro L and Sheng Z. cAb-Rep: A Database of Curated Antibody Repertoires for Exploring Antibody Diversity and Predicting Antibody Prevalence. *Front Immunol*. 2019;10:2365.
10. Barreto K, Maruthachalam BV, Hill W, Hogan D, Sutherland AR, Kusalik A, Fonge H, DeCoteau JF and Geyer C R. Next-generation sequencing-guided identification and reconstruction of antibody CDR combinations from phage selection outputs. *Nucleic Acids Research*. 2019;47:e50-e50.
11. Rettig TA, Ward C, Bye BA, Pecaut MJ and Chapes SK. Characterization of the naive murine antibody repertoire using unamplified high-throughput sequencing. *PLoS One*. 2018;13:e0190982-e0190982.
12. Smith T, Heger A and Sudbery I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res*. 2017;27:491-499.
13. Huang Y, Niu B, Gao Y, Fu L and Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*. 2010;26:680-682.
14. Jespersen MC, Mahajan S, Peters B, Nielsen M and Marcatili P. Antibody Specific B-Cell Epitope Predictions: Leveraging Information From Antibody-Antigen Protein Complexes. *Front Immunol*. 2019;10:298-298.
15. Norman RA, Ambrosetti F, Bonvin AMJJ, Colwell LJ, Kelm S, Kumar S and Krawczyk K. Computational approaches to therapeutic antibody design: established methods and emerging trends. *Briefings in Bioinformatics*. 2019.
16. Goldstein LD, Chen YJ, Wu J, Chaudhuri S, Hsiao YC, Schneider K, Hoi KH, Lin Z, Guerrero S, Jaiswal BS, Stinson J, Antony A, Pahuja KB, Seshasayee D, Modrusan Z, Hotzel I and Seshagiri S. Massively parallel single-cell B-cell receptor sequencing enables rapid discovery of diverse antigen-reactive antibodies. *Commun Biol*. 2019;2:304.
17. Wenzel EV, Roth KDR, Russo G, Fuhner V, Helmsing S, Frenzel A and Hust M. Antibody Phage Display: Antibody Selection in Solution Using Biotinylated Antigens. *Methods Mol Biol*. 2020;2070:143-155.
18. Lim CC, Woo PCY and Lim TS. Development of a Phage Display Panning Strategy Utilizing Crude Antigens: Isolation of MERS-CoV Nucleoprotein human antibodies. *Sci Rep*. 2019;9:6088.
19. Stark Y, Venet S and Schmid A. Whole Cell Panning with Phage Display. *Methods Mol Biol*. 2017;1575:67-91.

## REFERENCES (CONT.)

20. Ljungars A, Svensson C, Carlsson A, Birgersson E, Tornberg UC, Frendeus B, Ohlin M and Mattsson M. Deep Mining of Complex Antibody Phage Pools Generated by Cell Panning Enables Discovery of Rare Antibodies Binding New Targets and Epitopes. *Front Pharmacol.* 2019;10:847.
21. Williams GS, Mistry B, Guillard S, Ulrichsen JC, Sandercock AM, Wang J, Gonzalez-Munoz A, Parmentier J, Black C, Soden J, Freeth J, Jovanovic J, Leyland R, Al-Lamki RS, Leishman AJ, Rust SJ, Stewart R, Jermutus L, Bradley JR, Bedian V, Valge-Archer V, Minter R and Wilkinson RW. Phenotypic screening reveals TNFR2 as a promising target for cancer immunotherapy. *Oncotarget.* 2016;7:68278-68291.
22. Yang W, Yoon A, Lee S, Kim S, Han J and Chung J. Next-generation sequencing enables the discovery of more diverse positive clones from a phage-displayed antibody library. *Experimental & Molecular Medicine.* 2017;49:e308-e308.
23. Avery LB, Wade J, Wang M, Tam A, King A, Piche-Nicholas N, Kavosi MS, Penn S, Cirelli D, Kurz JC, Zhang M, Cunningham O, Jones R, Fennell BJ, McDonnell B, Sakorafas P, Apgar J, Finlay WJ, Lin L, Bloom L and O'Hara DM. Establishing in vitro in vivo correlations to screen monoclonal antibodies for physicochemical properties related to favorable human pharmacokinetics. *MAbs.* 2018;10:244-255.
24. Dostalek M, Prueksaritanont T and Kelley RF. Pharmacokinetic de-risking tools for selection of monoclonal antibody lead candidates. *MAbs.* 2017;9:756-766.
25. Rawat P, Kumar S and Michael Gromiha M. An in-silico method for identifying aggregation rate enhancer and mitigator mutations in proteins. *Int J Biol Macromol.* 2018;118:1157-1167.
26. Jain T, Boland T, Lilov A, Burnina I, Brown M, Xu Y and Vasquez M. Prediction of delayed retention of antibodies in hydrophobic interaction chromatography from sequence using machine learning. *Bioinformatics.* 2017;33:3758-3766.
27. Lee J and Bogyo M. Target deconvolution techniques in modern phenotypic profiling. *Curr Opin Chem Biol.* 2013;17:118-26.
28. Soria-Guerra RE, Nieto-Gomez R, Govea-Alonso DO and Rosales-Mendoza S. An overview of bioinformatics tools for epitope prediction: implications on vaccine development. *J Biomed Inform.* 2015;53:405-14.
29. Sidhom J-W, Pardoll D and Baras A. AI-MHC: an allele-integrated deep learning framework for improving Class I & Class II HLA-binding predictions. *bioRxiv.* 2018:318881.
30. Ogishi M and Yotsuyanagi H. Quantitative Prediction of the Landscape of T Cell Epitope Immunogenicity in Sequence Space. *Front Immunol.* 2019;10:827.
31. Winters A, McFadden K, Bergen J, Landas J, Berry KA, Gonzalez A, Salimi-Moosavi H, Murawsky CM, Tagari P and King CT. Rapid single B cell antibody discovery using nanopens and structured light. *mAbs.* 2019;11:1025-1035.
32. Scully M, Cataland SR, Peyvandi F, Coppo P, Knöbl P, Kremer Hovinga JA, Metjian A, de la Rubia J, Pavenski K, Callewaert F, Biswas D, De Winter H and Zeldin RK. Caplacizumab Treatment for Acquired Thrombotic Thrombocytopenic Purpura. *New England Journal of Medicine.* 2019;380:335-346.