High-throughput Sequencing Technologies are Revolutionising Antibody Discovery

Therapeutics based on antibodies (Abs) are at the forefront of revolutionary treatments for cancer, autoimmunity and many other diseases. The application of high-throughput sequencing (HTS) in monoclonal Ab (mAb) screening and selection is rapidly transforming the discovery process. By tightly integrating sequences with lab measurements, mAb candidates can be analysed in greater numbers and greater detail than ever before. Methods such as phylogenetic analysis allow researchers to better understand the differences between Abs based on their sequences and to rationally select the most promising candidates. Advanced modelling strategies are in their infancy but hold great potential for avoidance of developmental liabilities. In this article, we discuss the impact of HTS and immuno-informatics on the accuracy and speed of Ab discovery workflows, and how these technologies are revolutionising the Ab development field.

Antibody Therapeutics on the Rise

mAbs have emerged as a major class of therapeutics for immunological infectious diseases and cancer. mAb development has proven to yield effective drug molecules, with currently over 500 clinical studies ongoing. The success rate for these studies between Phase I and approval is 17–25%[1], which is double the general average (9.6% between 2006 and 2015[2]). Overall, the number of available therapeutic mAbs for clinical use has steadily increased since the first approval in 1986 (Figure. 1), and had already managed to seize almost half of the Top 20 U.S. therapeutic biologicals sales back in 2007[3].

The success of mAbs on the therapeutic market has driven scientific and technological discovery. Recently, two Nobel prizes were awarded to research directly involving Ab discovery; one covers expansion of the discovery toolbox with phage display[4]; the other highlights immune checkpoints, its most prominent therapeutic application[5]. Our improved understanding of the immune system and the use of novel technologies allow better Abs to be developed in a shorter amount of time. In this article, we will describe how HTS technologies are revolutionising the Ab discovery workflows and enable researchers to efficiently harness Ab characteristics for the development of sensitive and specific clinical impactful therapeutics.

Why Antibodies?

Diversity – Abs are incredibly diverse molecules. While most cellular receptors are hard-coded in the genome, Ab chains are generated by a semi-random recombination of gene segments and the addition of non-templated nucleotides, creating a huge variety of Ab. Mammal Abs are often formed by a dimer of two recombined Ab chains encoded in separate chromosomes, which together form the antigen-binding site. This creates truly unique binding sites, but also means that the full receptor cannot be sequenced together without creative molecular biology steps to link different mRNA products. The whole process has the capability to generate 10^{11} Abs with distinct specificities. This diversity contains Abs for an extremely wide range of targets. Finding a candidate among this large diversity thus becomes the objective and can be performed by screening for functionality (binding) and Ab sequence analysis.

Specificity – Abs can accumulate mutations that increase their binding affinity. Upon antigen binding, the B cell will become activated and will undergo clonal expansion to form a group of cells, called a clonotype. These cells can acquire mutations through somatic hypermutation (SHM), which further improves the affinity for the target (affinity maturation), leading to a class of extremely selective binders for any given antigen.

Versatility – Natural Ab molecules possess various relevant effector functions such as engaging the complement system

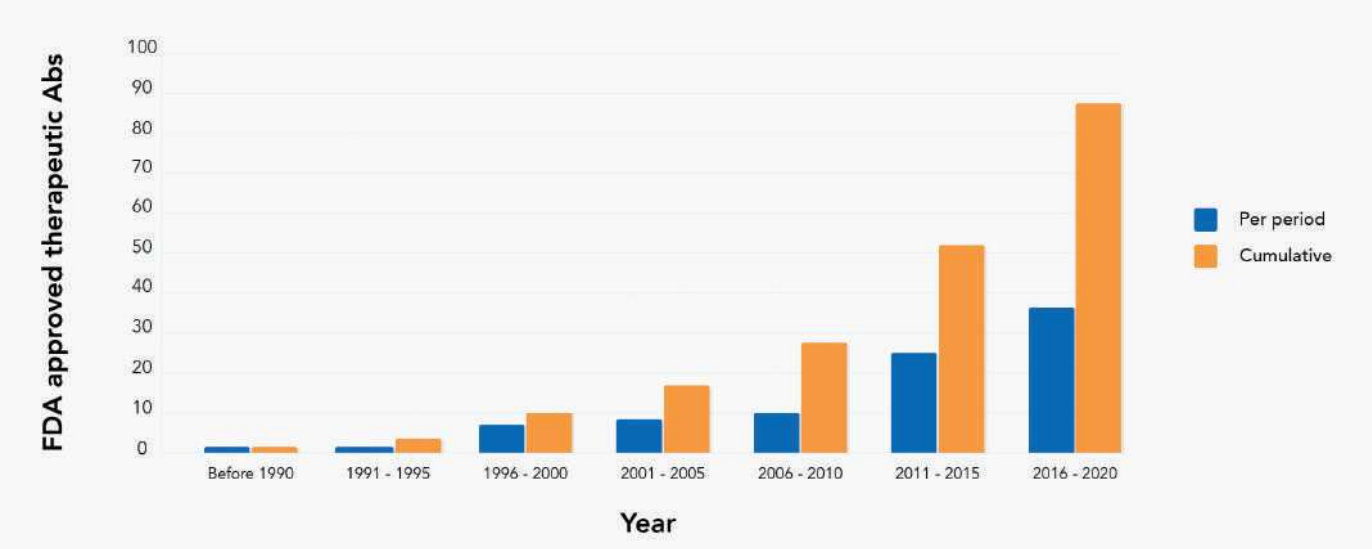

Figure 1. FDA-approved therapeutical mAbs per year, grouped by the number of approvals per time period (blue bars) and cumulative including that period (orange bars).

(ADCC) or Fc receptors, which takes their applications beyond mere target blocking. These effector functions can be further modified and refined by genetic engineering. Abs from different animal species have different properties or functionality advantages (e.g., nanobodies, IgNAR), and through bioengineering, Abs with multiple specificities (bispecific Ab) or linked to drugs (Ab-drug conjugates) can be constructed.

All-in-all, natural Abs are selected by the immune system to optimally bind to any antigen with strong affinity and high specificity and possess versatile effector functions. Our ability to select and engineer Abs with the right characteristics makes the B cell repertoire a perfect pool to fish for high-potential candidates to construct improved therapeutics.

The Different Roads to Rome

A variety of different Ab discovery platforms are in use today, but all include a selection process from thousands or even millions of distinct antibodies, to the one that is best for the job. Ab discovery workflows still rely heavily on *in vitro* (biochemical) screenings of Ab candidates[6]. Unfortunately, although considerable improvements have been made in the throughput of these techniques, the rate of discovery is limited by the biochemical or binding phase of the workflow. Moreover, due to inevitable technical limitations, good candidates can be lost along the way and flawed candidates may only be dropped at a late stage. Sequencing information captured throughout the process has shown to mitigate some of these issues and is therefore playing an increasingly important role in all discovery workflows[6].

Sequence Data is Key

There are several sequencing technologies available to elucidate adaptive immune receptors (Table 1). Sanger sequencing is often standardly performed in Ab discovery workflows for candidate validation purposes: the experiment start-up costs are low, and the read length is sufficient for almost any therapeutic construct (up to 1000 bp). Unfortunately, the throughput of this technology is very low compared to the natural Ab diversity. HTS technologies provide an unparalleled level of depth that, for the first time, allows for the interrogation of millions or receptor sequences. This broad candidate space, which can now be captured and interrogated, is revolutionising the Ab discovery workflows and paves the way for *in silico* prediction and prioritisation, reducing the amount of necessary time-consuming binding selection. Each discovery platform faces its own technical hurdles and leverages HTS data in a unique way to improve the discovery process (Figure 2).

Hybridoma – By fusing plasma cells with myeloma cells, immortalised B cell lines (hybridomas) can be created to stably produce mAbs[7]. This workflow is labour-intensive and typically only hundreds of distinct receptor variants can be tested for binding, leaving a large part of the candidate space unexplored[8,9]. HTS is increasingly applied to expand the candidate output (so-called 'hit expansion'). Millions of receptor sequences can be identified through HTS and clustered together with the small set of validated binders (typically Sanger sequences). This enlarged and diversified candidate scope not only increases the chance of success but can even improve the quality of the final selected candidate, leveraging naturally occurring affinity maturation.

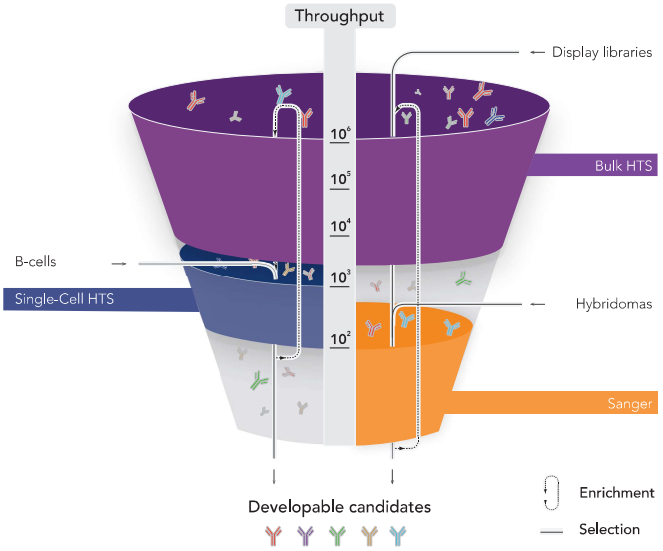

Figure 2. Schematic overview of the Ab discovery 'funnelling' process, starting from a throughput (y-axis) of millions, and progressively selecting candidates to thousands (single B-cell) or hundreds (hybridoma) of different Ab variants. After the validation of selected candidates, large-scale NGS data can be mined to expand the candidate pool (dotted flows upward), either to enrich the set of potential candidates, or to find Abs with better characteristics or without known sequence liabilities.

Display – The discovery of phage display technologies[10] revolutionised the way in which we can functionally screen candidates. The display of Ab fragments in phages and selection with panning rounds allows phenotypically screening up to 10 million distinct receptor variants in a single experiment. Display workflows can make use of HTS data for quality control (e.g., to ensure library diversity) and to analyse clone enrichment profiles throughout panning rounds. Analysis of the fold increase in clone abundance, rather than solely cloning abundant candidates from final panning output, expands the candidate pool with more diverse binding profiles.

TECHNOLOGY	CHAINS	THROUGHPUT	DEPTH	READ LENGTH*	RELATIVE COST**
Sanger	Single chain	Low/medium	Shallow (hundreds)	1000 bp	Low
Bulk	Single chain	High	Deep (millions)	600 bp	Medium
Single-cell	Paired chains	Low	Average (thousands)	250 bp	High

Table 1. Overview of sequencing technologies to characterise the Ab repertoire. A comparison of the three commonly applied sequencing technologies based on features that guide the decision process, including sequencing depth, chain information, read length, and cost.
** Most common commercially available length for Rep-Seq. ** Cost per sample, not base pair.*

Single-B cell – Single-B cell workflows are based on the direct amplification of VH- and VL-region-encoding genes from single human B cells and their subsequent expression in cell culture systems[11]. The identification of paired heavy and light chains is frequently done with HTS technologies, sometimes accompanied by transcriptome analysis, and typically yields thousands of distinct Ab variants[12]. As for hybridoma construction, HTS can be applied on the original (unsorted) B cell pool to enlarge the candidate scope and to estimate the clone abundance more accurately, indicative of antigen binding.

A Use Case: Finding Better Antibody Candidates in High-throughput Sequencing Data

A major bottleneck in some Ab discovery platforms is the expression and experimental testing of candidates. Even with cutting-edge isolation, expression and characterisation methods, the process is lengthy and yields less than 500 candidates, which is the tip of the iceberg considering the extremely diverse Ab repertoire typically generated against any antigen. Paired to this, affinity-based sorting methods are not perfect and can fail to select the best Abs. HTS offers the possibility to expand the candidate list by finding clonal relatives of the characterised Abs in a larger repertoire pool. For example, Phad *et al.* applied HTS of Abs to identify clonal relatives of single B cell candidates in different B cell compartments after immunisation[13]. These cells had undergone affinity maturation and had accumulated lots of mutations, which strongly correlated with neutralisation potency. This demonstrates how the discovery process can be improved with HTS by specifically identifying alternative Ab sequences with better characteristics, reducing time and cost and improving the Ab properties. This concept of 'target enrichment' is applied in all of the different discovery workflows, including hybridoma and phage display.

A Use Case: Expanding and Diversifying the Antibody Candidates in Phage Display

Phage display panning is a fast way to phenotypically select Ab candidates. Unfortunately, the calibration of the panning process to obtain a constant number of reliable candidates can be quite complex and may give uneven results. Panning is usually performed in sequential rounds, with amplifications in between to select for the stronger binders in a stepwise manner. Without sequencing candidates at intermediate steps, the panning process is a black box, the results of which can only be measured at the end. By using HTS in between the panning steps, the selected clones can be tracked and their expansion can be measured, which correlates with binding strength[14]. Moreover, the information obtained from HTS allows for a highly stringent final panning round to select the best candidate; if the final panning output is meagre, this information can be used to revisit previous panning rounds in a targeted manner.

Analytical Discovery Challenges

The analysis of Ab sequences has proven to be highly effective at improving selection and production of mAbs. However, the correct annotation of immune receptors, the interpretation of large-scale repertoires and the integration of different datatypes can be challenging.

Correctly Identifying Ab Sequences

The processing of Ab sequencing data requires specific techniques that are related to the peculiarities of Ab sequences. In a blood sample, the large diversity of Abs means that, on average, each one is represented by a single sequence in the data. This sequence contains segments from two to three highly polymorphic genes, and can, additionally, have a high proportion (>30%) of mutations. Sequencing data processing needs to take these factors into account to obtain the correct Ab sequences.

Interpreting Large-scale Ab Repertoires

Ab repertoires provide a wealth of information such as gene usage, sequence diversity and mutational load. The analysis and visualisation of these large datasets with millions of sequences for Ab development purposes (i.e., the identification of Abs with desired characteristics) has been shown to be challenging. A very common step in discovery workflows is the clustering of sequences in clonotypes or groups that (presumably) share properties such as binding affinity[15]. Given the volume of HTS data, this easily turns into an analysis bottleneck. Subsequently, the mutations in the clusters often are analysed to trace the sequence history from VDJ recombination to a set of Abs ('phylogenetic analysis' or 'lineage tracing'). Among others, this allows researchers to select Ab variants with similar binding characteristics, but different developability traits. The technical expertise required to perform phylogenetic analysis is substantial, hampering its use in day-to-day discovery workflows.

Leveraging Sequence Data for Ab Engineering

mAb development is partly about the selection of Ab candidates, but also involves Ab engineering to obtain or enhance desired properties. Traditionally, mAb optimisation is done by custom, experience, or (if all else fails) trial-and-error. Fortunately, the field is moving towards a more efficient strategy that is based on risk mitigation and rational design. Based solely on sequences, computational techniques can pinpoint likely origins of undesired properties (so-called sequence liabilities) and suggest engineering strategies for mitigation[16]. A great example is the use of public sequence databases to statistically assess the likelihood of liabilities associated with specific amino acid residues at certain positions in the Ab sequences. Using such models to analyse novel Ab candidates can prevent stability and developability issues. Moreover, promising work on sequence-based predictions of specific properties (e.g., 'humanness' of a sequence), has led to Ab sequences being altered 'rationally' in order to gain the most optimal sequences. This field is still moving rapidly and comes with some challenging tasks such as database mining and structural modelling.

Data Integration to Create a Complete Picture

The final challenge of Ab development is to bring all this information together such that a high-throughput analysis approach is possible. Linking existing and new data (e.g., Sanger and HTS) with new algorithms and methods will enable new insights to be gained. As new measurements and assay data (affinity, specificity for target, etc.) are added to quantify various aspects of Ab suitability, data analysis solutions should grow with the data to enable new insights.

The Right Toolset for the Job

Bioinformatics software solutions aim to reduce the time spent on data processing and simplify the access to valuable analysis methods and algorithms. This allows researchers without extensive bioinformatics expertise to independently

perform their discovery analysis and keeps them focused on what matters the most – the science. Academic researchers have created scripts and pipelines for Ab analysis, but these frequently have down-sides such as command-line interfaces, limited scope and/or usability, and sometimes sub-standard maintenance and validation approaches. Several commercial solutions exist for HTS data management and analysis, but these solutions are often generic and fail to take into consideration specific pitfalls related to Ab sequences. ENPICOM and others have developed data analysis solutions specifically tailored to the peculiarities of Ab sequence data. These comprehensive platforms sometimes incorporate (academic) point-solutions, in this way combining the best of both worlds: cutting-edge scientific methods and end-to-end robust solutions with a user-friendly interface. The integration of high-throughput screening and sequencing has proven very powerful for the rapid discovery of novel therapeutic Abs. Companies that leverage the full potential of HTS technologies and innovative immuno-informatic solutions will be ahead in this fast-growing field of therapeutic applications.

REFERENCES

1. Kaplon H, Reichert JM. Antibodies to watch in 2019. MAbs [Internet]. 2019 Feb 17 [cited 2020 Dec 27];11(2):219–38. Available from: /pmc/articles/PMC6380461/?report=abstract
2. BIO, Biomedtracker, Amplion. Clinical Development Success Rates 2006-2015 - Pharma intelligence. 2016.
3. Scolnik PA. mAbs: a business perspective. MAbs [Internet]. 2009/03/21. 2009;1(2):179–84. Available from: https://pubmed.ncbi.nlm.nih.gov/20061824
4. The Nobel Prize in Chemistry 2018 - NobelPrize.org [Internet]. [cited 2020 Dec 27]. Available from: https://www.nobelprize.org/prizes/chemistry/2018/summary/
5. The Nobel Prize in Physiology or Medicine 2018 - NobelPrize.org [Internet]. [cited 2020 Aug 12]. Available from: https://www.nobelprize.org/prizes/medicine/2018/summary/
6. Bailly M, Mieczkowski C, Juan V, Metwally E, Tomazela D, Baker J, et al. Predicting Antibody Developability Profiles Through Early Stage Discovery Screening. MAbs [Internet]. 2020 Jan 1 [cited 2021 Jan 11];12(1):1743053. Available from: https://www.tandfonline.com/doi/full/10.1080/19420862.2020.1743053
7. Köhler G, Milstein C. Continuous cultures of fused cells secreting antibody of predefined specificity. Nature [Internet]. 1975 [cited 2020 Dec 27];256(5517):495–7. Available from: https://pubmed.ncbi.nlm.nih.gov/1172191/
8. Mackenzie M, Cambrosio A, Keating P. The commercial application of a scientific discovery: The case of the hybridoma technique. Res Policy. 1988 Jun 1;17(3):155–70.
9. Juy D, Legrain P, Cazenave PA, Buttin G. A new rapid rosette-forming cell micromethod for the detection of antibody-synthesizing hybridomas. J Immunol Methods. 1979 Nov 15;30(3):269–75.
10. McCafferty J, Griffiths AD, Winter G, Chiswell DJ. Phage antibodies: filamentous phage displaying antibody variable domains. Nature [Internet]. 1990 [cited 2020 Dec 27];348(6301):552–4. Available from: https://pubmed.ncbi.nlm.nih.gov/2247164/
11. Tiller T. Single B cell antibody technologies. Vol. 28, New Biotechnology. Elsevier; 2011. p. 453–7.
12. Mcdaniel JR, DeKosky BJ, Tanno H, Ellington AD, Georgiou G. Ultra-high-throughput sequencing of the immune receptor repertoire from millions of lymphocytes. Nat Protoc. 2016 Mar 1;11(3):429–42.
13. Phad GE, Pushparaj P, Tran K, Dubrovskaya V, Àdori M, Martinez-Murillo P, et al. Extensive dissemination and intraclonal maturation of HIV Env vaccine-induced B cell responses. J Exp Med. 2020;217(2).
14. Yang W, Yoon A, Lee S, Kim S, Han J, Chung J. Next-generation sequencing enables the discovery of more diverse positive clones from a phage-displayed antibody library. Exp Mol Med [Internet]. 2017 Mar 3 [cited 2021 Jan 11];49(3):308. Available from: www.nature.com/emm
15. Galson JD, Trück J, Clutterbuck EA, Fowler A, Cerundolo V, Pollard AJ, et al. B-cell repertoire dynamics after sequential hepatitis B vaccination and evidence for cross-reactive B-cell activation. Genome Med [Internet]. 2016 Jun 16 [cited 2021 Jan 14];8(1):68. Available from: http://genomemedicine.biomedcentral.com/articles/10.1186/s13073-016-0322-z
16. Seeliger D, Schulz P, Litzenburger T, Spitz J, Hoerer S, Blech M, et al. Boosting antibody developability through rational sequence optimization. MAbs [Internet]. 2015 Jan 1 [cited 2021 Jan 11];7(3):505–15. Available from: http://www.tandfonline.com/doi/full/10.1080/19420862.2015.1017695

Néstor Vázquez Bernat

As an Application Specialist at ENPICOM, Néstor focuses on analyzing customer requirements, and project setup, management, and execution. During his PhD in Immunology, he isolated and expressed monoclonal antibodies after vaccinations and developed high-throughput sequencing library preparation protocols for B cell repertoires in humans, non-human primates, and other animal models.

Email: n.vazquez@enpicom.com

Josine oude Lohuis

As a Product Manager at ENPICOM, Josine is leading the development of the new antibody discovery module (ADM). This includes conducting strategic field exploration studies, defining new specifications, and driving software development by leveraging her diverse bioinformatics background and deep knowledge of immune repertoire sequencing technologies.

Email: j.oudelohuis@enpicom.com

Henk-Jan van den Ham

As a Senior Scientist at ENPICOM, Henk-Jan is part of the research team focusing on the latest developments in immunology & bioinformatics, with a focus on adaptive immune repertoire analysis. With a PhD in theoretical Immunology & bioinformatics and post-doc experience in virology, Henk-Jan leads academic & industry projects, collaborations and grant applications at the interface of immunology, bioinformatics and software engineering.

Email: h.j.vandenham@enpicom.com