mAbs 3:1, 17-20; January/February 2011; © 2011 Landes Bioscience

Sequencing antibody repertoires

The next generation

Nicolas Fischer
NovImmune SA; Geneva, Switzerland

Key words: next generation sequencing, phage display, hybridoma, antibody discovery, in vitro selection, immunization

Submitted: 11/01/10

Accepted: 11/09/10

DOI: 10.4161/mabs.3.1.14169

Correspondence to: Nicolas Fischer; Email: nfischer@novimmune.com

©2011 Landes Bioscience. Do not distribute.

Genomic studies have been revolutionized by the use of next generation sequencing (NGS), which delivers huge amounts of sequence information in a short span of time. The number of applications for NGS is rapidly expanding and significantly transforming many areas of life sciences. The field of antibody research and discovery is no exception. Several recent studies have harnessed the power of NGS for analyzing natural or synthetic immunoglobulin repertoires with unprecedented resolution and exploring alternative paths for antibody discovery. Thus, appreciating and then exploiting these advances is essential for staying at the edge of antibody innovation.

Next Generation Sequencing

In the past decade several sequencing technologies that are alternative to the Sanger method have been developed. Although these next generation sequencing (NGS) platforms rely on different principles and differ in the number and length of sequencing reads that they deliver during a run, they have dramatically increased sequencing outputs and lowered costs.[1,2] Depending on the technology, the sequencing of a human genome can now be completed in weeks for less than $100,000. The availability of NGS platforms is thus completely changing the landscape in many areas of basic, applied and medical research. Genomic studies first benefited from these advances, as whole genome sequencing and re-sequencing became possible even for individual research groups.[1,3] However, as is often observed when major technological advances become available, the scope of their use rapidly spreads as scientists explore applications within their own field of interest. In particular, quantitative applications, such as gene expression (RNA-seq), chromatin immunoprecipitation sequencing (ChIP-seq), microRNA profiling, as well as epigenetic and chromatin structure studies (DNAse-seq), have been used to exploit the power of NGS.[4-8] A wide variety of key biological processes ranging from protein-DNA interactions to temporal and spatial gene expression profiles can now be studied at nucleotide resolution, often revealing additional and sometimes unsuspected layers of complexity.[9,10]

Therefore, it is not surprising that several groups have also applied NGS to antibody research. From recent reports and studies carried out in our laboratories, it appears that NGS will not only contribute to a better understanding of the humoral arm of the immune system, but can also significantly improve and streamline antibody discovery.

Characterization of Natural Antibody Repertoires

DNA sequencing was essential for understanding antibody structure, function and diversification by the immune system. Antibody variable and hypervariable regions were initially identified by sequence comparisons.[11] Similarly, sequencing greatly contributed to our understanding of how the immunoglobulin repertoire is generated through recombination of germline gene segments

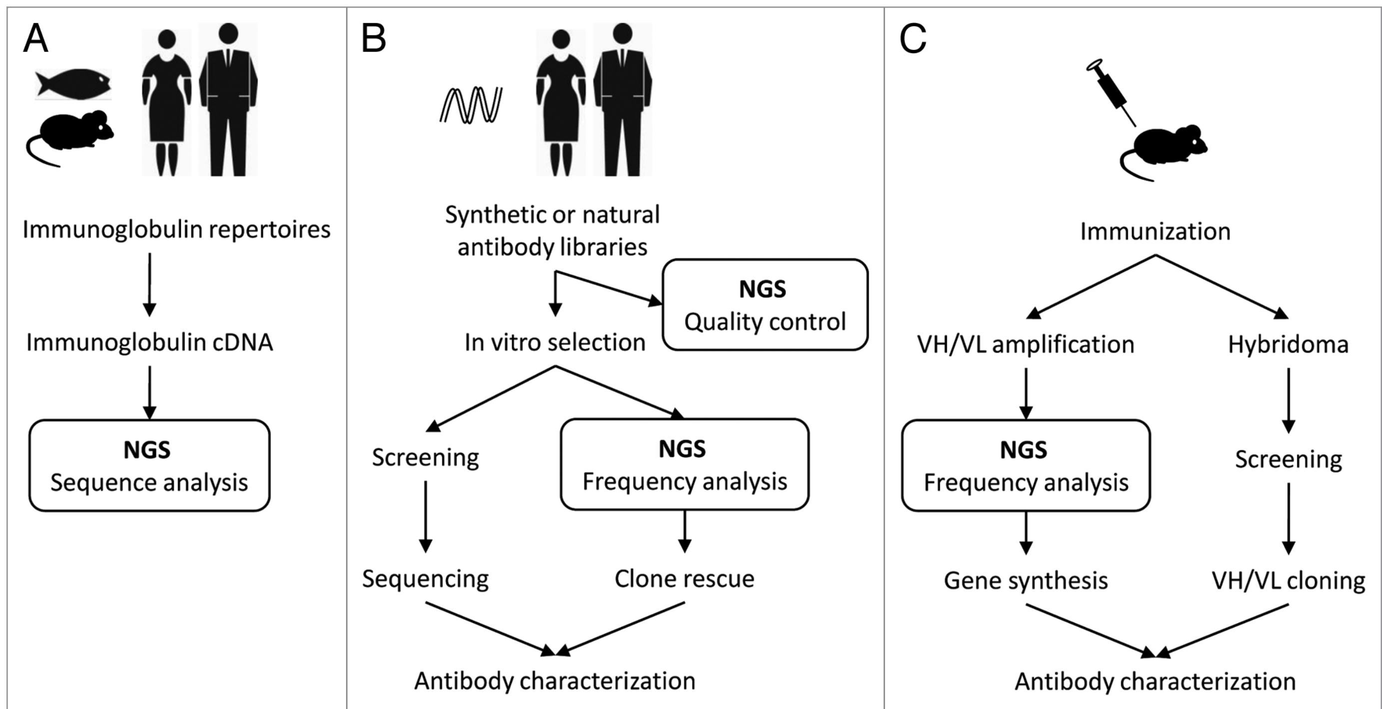

Figure 1. Different applications of next generation sequencing to antibody research and discovery. Deep sequencing of natural immunoglobulin repertoires (A). Quality control of natural and synthetic antibody libraries (B). Alternative approaches for antibody discovery, bypassing screening and applicable to both in vitro selection and immunization strategies (B and C).

Do not distribute.

and how additional mechanisms further introduce diversity to fine-tune the interaction with an antigen. While the human germline immunoglobulin genes have been identified, the limited output of Sanger sequencing provides only a glimpse into the expressed and constantly evolving nature of the antibody repertoire. NGS is ideally suited to extensively characterize this complex and plastic system (**Fig. 1A**). Although the potential size of the human repertoire (10^{11}) cannot be covered by current NGS platforms, the limited B cell number present in certain species, e.g., zebrafish, enables coverage and comparison of gene usage between individuals.[12-14]

Analyzing natural antibody sequences in which diversity is spread across all complementary determining regions (CDRs) requires long sequencing reads. Currently, the 454 system is the only NGS platform able to provide sequence of 250–400 base pairs (bp), but generates at least 10-fold less reads per run compared to other systems.[1] Another limitation of whole repertoire sequencing is that the original VH and VL pairing found in the lymphocyte cannot be determined, although it is an important contribution to diversity. Nevertheless, the sequencing of over a million immunoglobulin variable genes from several hundred individuals allowed for the analysis of germline usage, CDR lengths and amino acid composition with unprecedented depth.[15]

Quality Control of Antibody Libraries

Various approaches have been described for the generation of antibody libraries and their use for in vitro identification of antibodies against numerous targets.[16] Diversification strategies have been mainly based on the amplification and random association of naturally rearranged heavy and light chain variable genes or the introduction of synthetic diversity into the CDRs of selected antibody frameworks.[17,18] A crucial characteristic of a library is its size and diversity as they correlate with the number and affinity of target specific antibodies that can be isolated.[19] Diversity is limited by the number of transformants obtained at the bacterial or yeast transformation stage or by the number of molecules present in a defined volume for non-cell based in vitro selection systems. As such, claiming the largest library size has been a source of competition in the past. However, the actual functionality of the libraries, which is in fact the most important parameter, could only be superficially evaluated by sequencing. At most, only several hundred clones were routinely assessed, which represented a minute fraction of these libraries that usually had a size between 10^8 to 10^{10} purported variants.

Different NGS platforms have recently been used to characterize the content and quality of scFv libraries in a much more precise and effective manner. The 454 platform provides the required read lengths for libraries containing natural or synthetic diversity in all CDRs and also information on VH-VL pairing.[15,20] The Illumina/Solexa platform delivers several millions of shorter reads (up to 100 bp) and has been used to characterize a scFv library in which the diversification was restricted to the CDR3.[21] Despite the fact that 10^5–10^7 sequences still represent a fraction of large libraries, NGS provides unprecedented amounts of information on V-gene family usage, CDR length

distribution, amino acid composition and ultimately, the quality of a library beyond its theoretical size and content (**Fig. 1B**).

Antibody Sequence Evolution after Immunization or Selection

A fascinating feature of the immune system is that following exposure to a foreign antigen, it elegantly shapes the immunoglobulin repertoire for optimal binding to the pathogen. NGS can provide a better understanding of this process by extensively sampling antibody sequences at any given time. The immunoglobulin repertoire from mice after immunization has already been analyzed using NGS and revealed the enrichment of antigen specific V-genes and the polarization of the repertoire.[22] Using this type of approach, it is possible to compare naïve and immune repertoires against different targets and obtain a deeper understanding of the humoral immune system, which could aid in designing vaccination strategies.[12]

Sequence evolution during in vitro antibody selection, a process that aims to mimic the natural immune response, can be easily scrutinized by NGS as the repertoire to be sampled is smaller. Indeed, during the first round of in vitro selection against a target, the number of sequences drops drastically (to 10^5–10^6 selected variants) and, therefore, virtually all the sequences can be covered by current sequencing platforms. For the first time, this technology provides a comprehensive view of antibody sequences, enabling their relative enrichment between selection rounds to be visualized.[21]

Will Antibody Screening Become Obsolete?

Whether monoclonal antibodies are generated by in vivo or in vitro approaches their identification relies mainly on screening of hybridoma supernatants or bacterially expressed antibody fragments. This pivotal screening step can be quite intensive, e.g., involving robotic liquid handling systems, in order to test the largest possible number of clones and find candidates with the desired characteristics (**Fig. 1B and C**). Therefore, the ultimate goal of any screening campaign would be to test all possible candidates available after a fusion or after several rounds of in vitro selection. The extensive sequence information that is now available through NGS provides a means to bypass screening as described in two recent studies.[21,22] In the first, heavy and light chains were identified and paired based on their frequencies in the repertoire of immunized mice.[22] After gene synthesis of the selected VH and VL pairs, a panel of target specific antibodies was successfully generated thus effectively bypassing classical screening (**Fig. 1C**). An additional benefit of this approach was that the bone marrow cell repertoire that cannot be recovered by hybridoma fusion could be exploited. In a second study from our laboratory, phage display selections against two targets were performed and enrichment of CDR3 sequences between rounds of selection was followed. The high frequency sequences were identified and the corresponding scFv were recovered by a simple PCR step and subsequently shown to be target specific.[21] Compared to classical screening of the same selection rounds, this in silico identification approach provided additional high affinity candidates that were altogether missed by the classical approach.

Omitting the screening step can be highly beneficial. Upfront screening invariably leads to the repeated identification of the same enriched candidates. In contrast, sequence based identification limits working with this redundancy, thus allowing more effective and elaborate testing. Furthermore, when target protein is precious or available in limited quantities, this approach minimizes reagent consumption. Taken together, implementing a strategy that includes NGS renders the antibody identification process streamlined and more productive.

Concluding Remarks

As the current platforms become more widely applied, NGS will continue to revolutionize antibody research and discovery. New sequencing technologies with further increased throughput are in the pipeline and will again change the landscape of fundamental and applied research.[1] However, the challenges and limitations linked to NGS should not be underestimated. For instance, current NGS platforms lead to a higher error rate compared to the Sanger method, which has to be taken into consideration during analysis and interpretation of the data.[23] In addition, handling, transfer, storage and bioinformatic analysis of large quantities of information are not only technically challenging, but also have an impact on the overall sequencing costs.[24,25] Nevertheless, as applied to antibodies, NGS may lead to a deeper understanding of the immune system and improve the ability to select suitable therapeutic candidates, in turn, providing a better process to generate tomorrow's drugs.

Acknowledgements

Marie Kosco-Vilbois and Ulla Ravn for stimulating discussions.

©2011 Landes Bioscience. Do not distribute.

References

1. Metzker ML. Sequencing technologies—the next generation. Nat Rev Genet 2010; 11:31-46.
2. Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol 2008; 26:1135-45.
3. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, et al. The complete genome of an individual by massively parallel DNA sequencing. Nature 2008; 452:872-6.
4. Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu AL, et al. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res 2008; 18:610-21.
5. Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, et al. Dynamic regulation of nucleosome positioning in the human genome. Cell 2008; 132:887-98.
6. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009; 10:57-63.
7. Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, et al. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature 2008; 453:1239-43.
8. Wold B, Myers RM. Sequence census methods for functional genomics. Nat Methods 2008; 5:19-21.
9. Kahvejian A, Quackenbush J, Thompson JF. What would you do if you could sequence everything? Nat Biotechnol 2008; 26:1125-33.
10. Sorek R, Cossart P. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. Nat Rev Genet 2010; 11:9-16.
11. Wu TT, Kabat EA. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. J Exp Med 1970; 132:211-50.
12. Dimitrov DS. Therapeutic antibodies, vaccines and antibodyomes. MAbs 2010; 2:347-56.
13. Perelson AS, Oster GF. Theoretical studies of clonal selection: minimal antibody repertoire size and reliability of self-non-self discrimination. J Theor Biol 1979; 81:645-70.
14. Weinstein JA, Jiang N, White RA III, Fisher DS, Quake SR. High-throughput sequencing of the zebrafish antibody repertoire. Science 2009; 324:807-10.

15. Glanville J, Zhai W, Berka J, Telman D, Huerta G, Mehta GR, et al. Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. Proc Natl Acad Sci USA 2009; 106:20216-21.
16. Hoogenboom HR. Selecting and screening recombinant antibody libraries. Nat Biotechnol 2005; 23:1105-16.
17. Lloyd C, Lowe D, Edwards B, Welsh F, Dilks T, Hardman C, et al. Modelling the human immune response: performance of a 10^{11} library human antibody repertoire against a broad panel of therapeutically relevant antigens. Protein Eng Des Sel 2009; 22:159-68.
18. Rothe C, Urlinger S, Lohning C, Prassler J, Stark Y, Jager U, et al. The human combinatorial antibody library HuCAL GOLD combines diversification of all six CDRs according to the natural immune system with a novel display method for efficient selection of high-affinity antibodies. J Mol Biol 2008; 376:1182-200.
19. Griffiths AD, Williams SC, Hartley O, Tomlinson IM, Waterhouse P, Crosby WL, et al. Isolation of high affinity human antibodies directly from large synthetic repertoires. EMBO J 1994; 13:3245-60.
20. Ge X, Mazor Y, Hunicke-Smith SP, Ellington AD, Georgiou G. Rapid construction and characterization of synthetic antibody libraries without DNA amplification. Biotechnol Bioeng 2010; 106:347-57.
21. Ravn U, Gueneau F, Baerlocher L, Osteras M, Desmurs M, Malinge P, et al. By-passing in vitro screening—next generation sequencing technologies applied to antibody display and in silico candidate selection. Nucleic Acids Res 2010; 38:e193.
22. Reddy ST, Ge X, Miklos AE, Hughes RA, Kang SH, Hoi KH, et al. Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. Nat Biotechnol 2010; 28:965-9.
23. Kircher M, Kelso J. High-throughput DNA sequencing—concepts and limitations. Bioessays 2010; 32:524-36.
24. Batley J, Edwards D. Genome sequence data: management, storage and visualization. Biotechniques 2009; 46:333-4.
25. Richter BG, Sexton DP. Managing and analyzing next-generation sequence data. PLoS Comput Biol 2009; 5:e1000369.

©2011 Landes Bioscience.
Do not distribute.