Old Herborn University Seminar Monograph 31:
Evolutionary biology of the virome, and impacts in human health and disease.
Editors: Peter J. Heidt, Pearay L. Ogra, Mark S. Riddle and Volker Rusch.
Old Herborn University Foundation, Herborn, Germany: 89-96 (2017).

READING THE ANTIBODY REPERTOIRE AGAINST VIRUSES, MICROBES, AND THE SELF

URI LASERSON

Department of Genetics and Genomic Sciences and Precision Immunology Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

SUMMARY

Accessing the information in the antibody repertoire promises to provide an unprecedented window into the exposure history and disease susceptibility of an individual. However, the non-heritability and massive diversity of the receptors are major hurdles to interpreting the repertoire. We have developed the phage immunoprecipitation sequencing assay (PhIP-seq) that inexpensively generates high-dimensional functional profiles of a complex mixture of antibodies of unknown specificity. The assay leverages the recent advances in next-generation DNA sequencing along with synthesis of large pools of DNA oligonucleotides. We have successfully applied PhIP-seq to characterize auto-antigen profiles in patients with autoimmune diseases along with a global survey of antiviral immune responses. We are further developing the assay to apply to host-microbiome interactions, among other application areas.

INTRODUCTION

The repertoire of antibodies present in a single drop of blood should reveal an enormous amount of information about an individual, such as their history of infection, travel, vaccines, allergies, and even their susceptibility to future diseases (*Lerner* et al., 1991). However, multiple hurdles have obstructed our ability to glean such insights from the repertoire: proteins are more difficult to work with than nucleic acids, the non-germline nature of the repertoire complicates population studies (since most antibodies are rare antibodies), and the enormous sequence diversity has historically exceeded the capabilities of any measurement technology. The introduction of next-generation DNA sequencing technology (NGS) a decade ago (*Shendure* et al., 2017) has reinvigorated excitement in studying antibody and T cell receptor (TCR) repertoires (*Georgiou* et al., 2014). The use of NGS to study the immune repertoire has resulted in many insights about the core function of immunity and its role in virtually every disease, including infections (*Parameswaran* et al., 2013), cancer (*Li* et al., 2016; *Liu* and *Mardis*, 2017), and autoimmunity (*Dekosky* et al., 2015).

A holy grail in the field of adaptive immunity is the creation of a mapping between antibody/TCR sequences and their cognate epitopes, such that an antigen/epitope can be determined purely from inspecting the antibody sequence and vice versa. But despite the massive increase in antibody sequence data (*Breden* et al., 2017), most studies typically report either aggregate statistics (e.g., repertoire diversity met-

rics, germline gene usage) or compare them across experimental conditions to discover biomarkers (*Yaari* and *Kleinstein*, 2015). Alternatively, some studies have characterized antibody repertoires in the context of a single antigen and require low-throughput sorting of antigen-specific cells. The ability to predict immune receptor function from its sequence has only been demonstrated in very narrow situations and primarily for TCRs (*Emerson* et al., 2017; *Glanville* et al., 2017).

Here we describe phage immunoprecipitation sequencing (PhIP-seq) (*Larman* et al., 2011), an NGS-based assay that can determine the specificity of a repertoire of antibodies against hundreds of thousands of antigens simultaneously. The assay makes use of phage display libraries containing a set of antigens, each of which will be measured for binding to some antibody of interest. Because the universe of possible antigens grows exponentially with peptide length, such peptide libraries have typically been constructed as random oligomers with lengths <20 amino acids. However, libraries of random peptides "waste" clones on sequences that are not observed in nature. Furthermore, the use of short peptides precludes their ability to exhibit natural structure. To circumvent these limitations, we have leveraged improvements in the ability to synthesize large, complex pools of DNA oligonucleotides (*Kosuri* and *Church*, 2014). Similarly to pooled-screen experiments, the ability to synthesize large pools of DNA oligos to-order allows us to "synthesize" many experiments and execute them in parallel (*Gasperini* et al., 2016).

As described below, PhIP-seq has been used successfully to generate high-dimensional profiles of antibody immunity in the context of autoimmunity and viral infection. In the following sections, we review some of the results of these efforts, along with preliminary methods for expanding the scope of PhIP-seq to bacterial antigens.

PHAGE IMMUNOPRECIPITATION SEQUENCING

Given a monoclonal antibody or a complex mixture of antibodies (e.g., serum) of unknown specificity, the PhIP-seq assay can determine the specificity of the antibodies from a large set of candidate epitopes that have been selected in advance. Specifically, the "query antibody" is used to immunoprecipitate a phage display library expressing many antigens of interest. The phage clones that are pulled down by the antibody are quantified using next-generation DNA sequencing. Therefore, in a single ~$30 assay, small amounts of antibody (nanogram to microgram quantities) can be assayed against hundreds of thousands of candidate epitopes in parallel. In principle, the diversity of the epitope library is only limited by what can be cloned into the phage display library.

For example, the set of human protein sequences can be tiled with a set of ~400,000 unique peptide 36-mers. The DNA oligos that code for the 36-mers can be synthesized by a commercial DNA vendor (e.g., Twist Bioscience or Agilent Technologies) and cloned into a commercially available bacteriophage T7 protein display system (e.g., Novagen T7Select) to yield a phage library containing every possible human protein 36-mer. The phage library can then be used to identify unknown auto-antigens using serum samples from patients with autoimmune diseases.

It is estimated that a majority of

antibody epitopes are "conformational" rather than linear (*Sela* et al., 1967; *Barlow* et al., 1986). A clear limitation of this method is that the epitopes are limited to the length of DNA oligonucleotides that can be synthesized en masse. As of this writing, synthesis of large oligo pools is generally limited to 200-300 bp sequences, in which some of the sequence must be used for common adaptor sequence for cloning (allowing ~56 aa peptides). Even if we could synthesize or assemble significantly larger pieces of DNA, we would eventually reach the capacity of the phage display system to display larger proteins without interfering with the assembly of infective phage particles. Furthermore, because the peptides are processed in *E. coli*, we cannot include any mammalian post-translational modifications when applicable. These limitations ultimately lead to a loss of sensitivity and make PhIP-seq somewhat unsuitable for diagnostic applications.

It is our experience that robust immune responses generally include at least some responses to linear epitopes. Therefore, PhIP-seq is highly suitable for hypothesis generation experiments, similarly to genome wide association studies (GWAS). Instead of testing SNPs for association with a phenotype, we can use PhIP-seq to test whether particular antibody specificities are associated with a phenotype. Despite its limitations, PhIP-seq provides an inexpensive means to generate a very high-dimensional characterization of the antibody specificity repertoire.

DISCOVERY OF AUTO-ANTIGENS

PhIP-seq was first demonstrated by building a library of ~400,000 36-mer peptides tiling >24,000 human open reading frames with 7 aa overlaps (*Larman* et al., 2011). Our initial test searched for auto-antigens in several cancer patients with paraneoplastic neurological disorder (PND), a cancer-induced autoimmune syndrome. After successfully detecting known and novel auto-antigens in these patients, we undertook a larger survey of ~300 subjects with multiple sclerosis (MS), type 1 diabetes (T1D), or rheumatoid arthritis (RA), along with healthy controls (*Larman* et al., 2013A). Analysis of the repertoire of autoantibodies in healthy controls revealed that most individuals exhibit antibodies against a variety of self-epitopes. However, the vast majority of these self-specificities were "private" in the sense that they were observed in only a single individual, likely a result of a cross-reactive antibody that is not causing any pathology. Interestingly, some auto-antigen specificities were observed in >20 healthy control subjects, such at MAGEE1, ACVR2B, and TTN. But given the healthy status of these individuals, the most likely explanation is that they represent common cross-reactivities. Using the samples from T1D patients, we analysed the sensitivity of the PhIP-seq assay by comparing peptide enrichments to RIA assays for known auto-antibodies against insulin, ZnT8A, GAD2, and PTPRN. PhIP-seq with the 36-mer library was only able to recover reactivities against GAD2 and PTPRN in some patients. Analysis of RA patients did not find any new auto-antigens; however, we were able to cluster a subset of patients into one of two specificity profiles, which was not correlated with their seropositivity status. Finally, by performing a motif analysis on the positive peptide enrichments of

the MS cohort, we were able to recapitulate a previously discovered reactivity to the BRRF2 protein. In a separate PhIP-seq study, the auto-antigen for inclusion body myositis was discovered to be cytosolic 5'-nucleotidase 1A (*Larman* et al., 2013B).

CHARACTERIZING GLOBAL VIRAL IMMUNITY

Following the demonstration and use of the human PhIP-seq library, the "VirScan" PhIP-seq library was constructed containing 93,904 56-mer peptides (28 aa overlap) derived from 206 viral species that infect humans (e.g., HIV, HCV, EBV, influenza) (*Xu* et al., 2015). Sera from 569 individuals of various ages primarily from the United States but also including sera from Thailand, South Africa, and Peru were characterized with VirScan to generate a global survey of immunological memory/repertoire against viral infection. A subset of samples were tested with ELISA against HIV1, HCV, HSV1, and HSV2, and compared to the VirScan results. When aggregating VirScan epitopes to the virus level, VirScan exhibited >90% sensitivity for all viruses with 100% specificity for HIV1, HSV1, and HSV2. The VirScan-computed fraction of individuals with immunity to CMV (48.5%) and EBV (87.1%) recapitulated known prevalences of the diseases. In contrast, the computed prevalences of Influenza A (53.4%), Poliovirus (33.7%), and VZV (24.3%) were all lower than expected, possibly reflecting a narrowing of the immune response due to absence of the antigens, or the low specificity of PhIP-seq due to the limitation to linear epitopes. Most interestingly, the population analysis of VirScan provided a dramatic display of immunodominance across a large number of viral proteins. When individuals displayed antibodies against a 56-mer derived from a viral protein, they generally all showed specificity for the same 56-mer tile. However, some cases showed that the immunodominant epitopes varied with geolocale. Because the epitope mapping resolution of VirScan is only 28 amino acids, several 56-mer tiles were chosen for alanine scanning mutagenesis in order to pinpoint the location of the epitope. In most cases, individuals specific for a particular 56-mer tile also showed that the sera was specific for the same ~10-mer epitope. Overall, VirScan has been shown to provide an unprecedentedly high-dimensional view of global immunity against the human virome.

BUILDING LIBRARIES FOR THE MICROBIOME

The influence of the microbiome on human health and disease has been receiving increasing attention (*Cho* and *Blaser*, 2012). This interaction is largely mediated through the mucosal immune system (*McCoy* et al., 2017). Indeed, the majority of antibody synthesized by the human body is IgA that is secreted into the gut lumen to interact with the commensal flora and protect against pathogens (*Fagarassan* and *Honjo*, 2003). However, the specific mechanisms and antigens that drive this interaction are currently unknown (*Kubinak* and *Round*, 2016). Significant progress has been made with the development of the IgA-seq assay, in which gut microbes are sorted into

IgA-bound and IgA-free fractions, followed by 16S rRNA sequencing to determine which taxa are targeted by secretory IgA (sIgA) (*Palm* et al., 2014; *Kau* et al., 2015). While this technique has generated useful hypotheses in the pathogenesis of gut inflammation and other disorders, the specific antigens involved cannot be determined from the relatively low-resolution 16S data.

We have been developing library construction methods for bacterial antigens in order to characterize the sIgA repertoire at antigen-resolution. While we have designed synthetic libraries for particular bacterial species (*Staphylococcus aureus* and *Streptococcus pyogenes*), the microbiome as a whole is too genetically diverse to practically synthesize as DNA oligos using available technology. As an alternative, we have developed a protocol for cloning shotgun metagenomic libraries into the T7 phage display system. The two main advantages of this approach are that we are not limited by current knowledge of the genetic contents of the microbiome and we can also clone larger fragments than can be synthesized as oligonucleotides. However, because the genomic DNA of the microorganisms is randomly sheared, we expect only one sixth of the clones to be in the correct reading frame. While out-of-frame clones may show elevated levels of background noise, analysis of the sequences should distinguish which clones represent true ORF enrichments. Alternatively, a recent method has been developed based on padlock probes in order to capture large DNA fragments in the correct reading frame (*Tosi* et al., 2017). We hope that these libraries will have many applications, such as understanding the role of the mucosal immune system in conferring protection against infection, or understanding the host immune response in inflammatory bowel disease.

ALTERNATIVE TECHNOLOGIES

A number of assays similar to PhIP-seq have been developed that have different trade-offs. For ease of library construction and non-biasing of sequences, some have used random peptides in display systems (*Pantazes* et al., 2016). However, these libraries tend to "waste" many clones on epitopes that are not observed in nature. Others have used alternative protein display technologies to phage display (*Boder* and *Wittrup*, 1997; *Spatola* et al., 2013; *Zhu* et al., 2013). Most intriguingly, a system for characterizing TCR specificities has been developed by expressing large libraries of peptide-MHC complexes using yeast display (*Birnbaum* et al., 2014). However, the system is limited to MHC-II, and each MHC allele must be individually optimized and engineered to fold correctly. To overcome the size limitation of phage display, the Elledge group has also developed PLATO, a system analogous to PhIP-seq that uses ribosome display to express full ORFs (*Zhu* et al., 2013; *Larman* et al., 2014). In contrast to direct measurement of specificities, the recently published GLIPH technique can perform semi-supervised clustering of TCRs in order to determine their specificities (*Glanville* et al., 2017). Finally, protein microarrays are a well-characterized alternative to protein display-based methods (*Templin* et al., 2002). While they can sometimes accommodate full ORFs, they suffer some solid-state kinetics, generally require larger amounts of input material, and are relatively expensive and lower-throughput than NGS-based assays.

DISCUSSION

The simultaneous improvement of next-generation DNA sequencing technology and large-scale DNA synthesis technology has enabled the development of highly multiplexed assays at low costs. PhIP-seq provides the ability to assay antibody specificities with unprecedented breadth and will provide new insights into the landscape of host immune responses to their environments. One of the primary bottlenecks is the availability of libraries containing antigens of interest. While we described the use of libraries containing auto-antigens and viral antigens, we are actively developing methods for cloning bacterial libraries. Furthermore, we have designed libraries for allergens and toxins, among other classes of antigens of interest.

By generating a large number of PhIP-seq profiles, our data sets could be used to train B cell epitope predictors (e.g., BepiPred; *Larsen* et al., 2006). Most intriguingly, coupling the PhIP-seq assay to single-cell RNA-seq methods (*Stubbington* et al., 2017) opens the possibility of generating training data sets that could allow us to achieve the holy grail of predicting antibody specificity from primary sequence.

ACKNOWLEDGEMENTS

I would like to thank Ben Larman, Meimei Shan, and Tim O'Donnell for helpful discussions.

LITERATURE

Barlow, D.J., Edwards, M.S., and Thornton, J.M.: Continuous and discontinuous protein antigenic determinants. Nature 322, 747-748 (1986).

Birnbaum, M.E., Mendoza, J.L., Sethi, D.K., Dong, S., Glanville, J., Dobbins, J., Ozkan, E., Davis, M.M., Wucherpfennig, K.W., Garcia, K.C.: Deconstructing the peptide-MHC specificity of T cell recognition. Cell 157, 1073-1087 (2014).

Boder, E.T. and Wittrup, K.D.: Yeast surface display for screening combinatorial polypeptide libraries. Nat. Biotechnol. 15, 553-557 (1997).

Breden, F., Luning Prak, E.T., Peters, B., Rubelt, F., Schramm, C.A., Busse, C.E., Vander Heiden, J.A., Christley, S., Bukhari, S.A.C., Thorogood, A., Matsen, F.A. IV, Wine, Y., Laserson, U., Klatzmann, D., Douek, D.C., Lefranc, M.P., Collins, A.M., Bubela, T., Kleinstein, S.H., Watson, C.T., Cowell, L.G., Scott, J.K., and Thomas B. Kepler, T.B.: Reproducibility and Reuse of Adaptive Immune Receptor Repertoire Data. Front. Immunol. 8, 1418 (2017).

Cho, I. and Blaser, M.J.: The human microbiome: at the interface of health and disease. Nat. Rev. Genet. 13, 260-270 (2012).

Dekosky, B.J., Kojima, T., Rodin, A., Charab, W., Ippolito, G.C., Ellington, A.D., and Georgiou, G.: In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. Nat. Med. 21, 86-91 (2015).

Emerson, R.O., DeWitt, W.S., Vignali, M., Gravley, J., Hu, J.K., Osborne, E.J., Desmarais, C., Klinger, M., Carlson, C.S., Hansen, J.A., Rieder, M., and Robins, H.S.: Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell

repertoire. Nat. Genet. 49, 659-665 (2017).
Fagarasan, S.S. and Honjo, T.T.: Intestinal IgA synthesis: regulation of front-line body defences. Nat. Rev. Immunol. 3, 63-72 (2003).
Gasperini, M., Starita, L., and Shendure, J.: The power of multiplexed functional analysis of genetic variants. Nat. Protoc. 11, 1782-1787 (2016).
Georgiou, G., Ippolito, G.C., Beausang, J., Busse, C.E., Wardemann, H., and Quake, S.R.: The promise and challenge of high-throughput sequencing of the antibody repertoire. Nat. Biotechnol. 32, 158-168 (2014).
Glanville, J., Huang, H., Nau, A., Hatton, O., Wagar, L.E., Rubelt, F., Ji, X., Han, A., Krams, S.M., Pettus, C., Haas, N., Arlehamn, C.S.L., Sette, A., Boyd, S.D., Scriba, T.J., Martinez, O.M., and Davis, M.M.: Identifying specificity groups in the T cell receptor repertoire. Nature 547, 94-98 (2017).
Kau, A.L., Planer, J.D., Liu, J., Rao, S., Yatsunenko, T., Trehan, I., Manary, M.J., Liu, T.C., Stappenbeck, T.S., Maleta, K.M., Ashorn, P., Dewey, K.G., Houpt, E.R., Hsieh, C.S., and Gordon, J.I.: Functional characterization of IgA-targeted bacterial taxa from undernourished Malawian children that produce diet-dependent enteropathy. Sci. Transl. Med. 7, 276ra24 (2015).
Kosuri, S. and Church, G.M.: Large-scale de novo DNA synthesis: technologies and applications. Nat. Methods 11, 499-507 (2014).
Kubinak, J.L. and Round, J.L.: Do antibodies select a healthy microbiota? Nat. Rev. Immunol. 16, 767-774 (2016).
Larman, H.B., Zhao, Z., Laserson, U., Li, M.Z., Ciccia, A., Gakidis, M.A.M., Church, G.M., Kesari, S., Leproust, E.M., Solimini, N.L., and Elledge, S.J.: Autoantigen discovery with a synthetic human peptidome. Nat. Biotechnol. 29, 535-541 (2011).
Larman, H.B., Laserson, U., Querol, L., Verhaeghen, K., Solimini, N.L., Xu, G.J., Klarenbeek, PL., Church, G.M., Hafler, D.A., Plenge, R.M., Nigrovic, P.A., De Jager, P.L., Weets, I., Martens, G.A., O'Connor, K.C., and Elledge, S.J.: PhIP-Seq characterization of autoantibodies from patients with multiple sclerosis, type 1 diabetes and rheumatoid arthritis. J. Autoimmun. 43, 1-9 (2013A).
Larman, H.B., Salajegheh, M., Nazareno, R., Lam, T., Sauld, J., Steen, H., Kong, S.W., Pinkus, J.L., Amato, A.A., Elledge, S.J., and Greenberg, S.A.: Cytosolic 5′-nucleotidase 1A autoimmunity in sporadic inclusion body myositis. Ann. Neurol. 73, 408-418 (2013B).
Larman, H.B., Liang, A.C., Elledge, S.J., and Zhu, J.: Discovery of protein interactions using parallel analysis of translated ORFs (PLATO). Nat Protoc. 9, 90-103 (2014).
Larsen, J.E.P., Lund, O., and Nielsen, M.: Improved method for predicting linear B-cell epitopes. Immunome Res. 2, 2 (2006).
Lee, J.Y., Artiles, K.L., Zompi, S., Vargas, M.J., Simen, B.B., Hanczaruk, B., McGowan, K.R., Tariq, M.A., Pourmand, N., Koller, D., Balmaseda, A., Boyd, S.D., Harris, E., and Fire, A.Z.: Convergent antibody signatures in human dengue. Cell Host Microbe 13, 691-700 (2013).
Lerner, R.A., Barbas, C.F. 3rd., Kang, A.S., and Burton, D.R.: On the use of combinatorial antibody libraries to clone the “fossil record” of an individual’s immune response. Proc. Natl. Acad. Sci. USA 88, 9705-9706 (1991).
Li, B., Li, T., Pignon, J.C., Wang, B., Wang, J., Shukla, S.A., Dou, R., Chen, Q., Hodi, F.S., Choueiri, T.K., Wu, C., Hacohen, N., Signoretti, S., Liu, J.S., and Liu, X.S.: Landscape of tumor-infiltrating T cell repertoire of human cancers. Nat. Genet. 48, 725-732 (2016).
Liu, X.S. and Mardis, E.R.: Applications of Immunogenomics to Cancer. Cell 168, 600-612 (2017).
McCoy, K.D., Ronchi, F., and Geuking, M.B.: Host-microbiota interactions and adaptive immunity. Immunol. Rev. 279, 63-69 (2017).
Palm, N.W., De Zoete, M.R., Cullen, T.W.,

Barry, N.A., Stefanowski, J., Hao, L., Degnan, P.H., Hu, J., Peter, I., Zhang, W., Ruggiero, E., Cho, J.H., Goodman, A.L., and Flavell, R.A.: Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. Cell 158, 1000-1010 (2014).

Pantazes, R.J., Reifert, J., Bozekowski, J., Ibsen, K.N., Murray, J.A., and Daugherty, P.S.: Identification of disease-specific motifs in the antibody specificity repertoire via next-generation sequencing. Sci. Rep. 6, 30312 (2016).

Sela, M., Schechter, B., Schechter, I., and Borek, F.: Antibodies to Sequential and Conformational Determinants. Cold Spring Harbor Symp. Quant. Biol. 32, 537-545 (1967).

Shendure, J., Balasubramanian, S., Church, G.M., Gilbert, W., Rogers, J., Schloss, J.A., and Waterston, R.H.: DNA sequencing at 40: past, present and future. Nature 550, 345-353 (2017).

Spatola, B.N., Murray, .JA., Kagnoff, M., Kaukinen, K., and Daugherty, P.S.: Antibody repertoire profiling using bacterial display identifies reactivity signatures of celiac disease. Anal. Chem. 85, 1215-1222 (2013).

Stubbington, M.J.T., Rozenblatt-Rosen, O., Regev, A., and Teichmann, S.A.: Single-cell transcriptomics to explore the immune system in health and disease. Science 358, 58-63 (2017).

Templin, M.F., Stoll, D., Schrenk, M., Traub, P.C., Vöhringer, C.F., and Joos, T.O.: Protein microarray technology. Drug Discov. Today 7, 815-822 (2002).

Tosi, L., Sridhara, V., Yang, Y., Guan, D., Shpilker, P., Segata, N., Larman, H.B., and Parekkadan, B.: Long-adapter single-strand oligonucleotide probes for the massively multiplexed cloning of kilobase genome regions. Nat. Biomed. Eng. 1, 0092 (2017).

Xu, G.J., Kula, T., Xu, Q., Li, M.Z., Vernon, S.D., Ndung'u, T., Ruxrungtham, K., Sanchez, J., Brander, C., Chung, R.T., O'Connor, K.C., Walker, B., Larman, H.B., and Elledge, S.J.: Viral immunology. Comprehensive serological profiling of human populations using a synthetic human virome. Science 348, aaa0698 (2015).

Yaari, G. and Kleinstein, S.H.: Practical guidelines for B-cell receptor repertoire sequencing analysis. Genome Med. 7, 121 (2015).

Zhu, J., Larman, H.B., Gao, G., Somwar, R., Zhang, Z., Laserson, U., Ciccia, A., Pavlova, N., Church, G., Zhang, W., Kesari, S., and Elledge, S.J.: Protein interaction discovery using parallel analysis of translated ORFs (PLATO). Nat. Biotechnol. 31, 331-334 (2013).