You May be Unique, but Your Adaptive Immune System is Not

ImmunoSEQ technology identifies commonalities in human adaptive immune systems, opening up new ways for diagnosing and treating disease.

Introduction

The human adaptive immune system rallies T and B lymphocytes to combat the multitude of pathogens we face in a lifetime. It generates a diverse repertoire of T- and B-cell antigen receptors to recognize and remember these pathogens, with effector cells leading an immediate immune response and memory cells poised to mount a stronger attack the next time the pathogen is encountered. Finding the cells that respond to specific pathogens and identifying their receptors could open the door to new diagnostics and therapeutics. Yet, until recently, scientists were unsure exactly how many receptors there were and which ones were involved in specific disease responses.

We all possess a varied repertoire of T cell receptors (TCRs) that are involved in cell-mediated immune responses, i.e. activating macrophages, natural killer cells, and cytotoxic T-lymphocytes. Most of their diversity is contained within the third complementarity-determining region (CDR3) of the TCR α and β chains. In the 1990s, researchers sequenced the genomic structure of these genes using capillary-based sequencing, describing hundreds of gene segments that can rearrange into mature TCR proteins. Extrapolating from the combinatoric possibilities using these germline gene segments gave a rough estimate of approximately 1×10^{11} possible TCR β chain CDR3, with experimental studies suggesting that 1×10^{6} of these are actually made per individual. Because the number of possible TCR sequences is much larger than the subset found in an individual, the belief has been that each person's TCR repertoire is effectively unique.

A more accurate measurement of the TCR repertoire was needed, one that would require significantly deeper TCR sequencing than could be obtained using capillary-based sequencing. To accomplish this, researchers turned to next-generation sequencers such as Illumina's Genome Analyzer. The trick was to sequence only the highly variable segment of the TCR β chain CDR3 regions carried in millions of T cells and develop computational algorithms to make sense of the huge amount of the data generated. A team of researchers from the Fred Hutchinson Cancer Research Center (FHCRC) led by Harlan Robins, Ph.D. and Christopher Carlson, Ph.D. solved the problem, developing a novel set of TCR sequencing assays and computational tools to sequence rearranged TCR genes at high throughput.

Christopher Carlson, Ph.D., co-founder of Adaptive TCR and an Assistant Member, Fred Hutchinson Cancer Research Center.

Known as ImmunoSEQ, the method was recently used to sequence more than five million TCRs simultaneously from seven healthy adults and the results were surprising—any two people may share tens of thousands of TCR β chains encoding exactly the same amino acid sequence. While striking a blow to the concept that each of us realizes a unique TCR β repertoire, the discovery bodes well for clinical advancements. Armed with an understanding of which TCRs respond to which pathogens or antigens, researchers could develop tools to diagnose and treat a broad range of ailments, from infections to organ transplants and autoimmune diseases.

With the ImmunoSEQ technologies as its foundation, Drs. Robins and Carlson co-founded Adaptive TCR to offer TCR sequencing and analysis services to researchers. We spoke with Dr. Carlson about how ImmunoSEQ was developed, what it has uncovered so far, and how it might be used to enhance diagnostic and treatment paradigms.

Q. What makes the adaptive immune system so special and so complex?

CC: In the adaptive immune system, the antigenic specificity of T cells is determined by the amino acid sequence in the hypervariable CDR3 region of the α and β chains of TCR. This small piece of germ line genome works like a card deck. We do not inherit T-cell receptors or antibodies from our parents. Instead, we inherit this card deck that can be shuffled to generate trillions of different antibodies and T-cell receptors. The number of combinatorial possibilities was so huge that most people thought you would hardly every find any two TCR sequences in common between individuals. ImmunoSEQ has helped us demonstrate that this is not true.

Q. What is ImmunoSEQ and how was it developed?

CC: ImmunoSEQ leverages the sequencing power of the Genome Analyzer, and our proprietary PCR assays and bioinformatics software to reveal details about the immune system that would otherwise be all but invisible. Most researchers use the Genome Analyzer to sequence millions of different genome segments. However, we wanted to sequence the same segment—the "card deck," or hypervariable

CDR3 region of T-cells. We first developed assays for key TCR genes that use a novel multiplex primer set to amplify this region of each T cell prior to sequencing. Then we created a bioinformatics software suite for managing and analyzing the large amount of data the assay generates. We believe it will replace spectratyping as the methodology of choice for assessing the clonal composition of T cell populations.

Q. What hurdles did you have to overcome in developing the multiplex PCR design?

CC: First, it is a fairly high multiplex PCR primer. Just for T-cell receptor β there are 45 forward primers and 13 reverse primers. We had to structure those primers so they would all amplify at the same efficiency. We also did not want to sequence the primers, just the amplified region, so we created a 13-plex sequencing primer that worked like a charm, enabling us to gather data on any molecule in our library.

Q. Why did you choose the Genome Analyzer?

CC: We recognized that the Genome Analyzer's phenomenal capacity to simultaneously sequence millions of molecules could be valuable in deciphering the adaptive immune system. It can generate a 60 nucleotide sequence length that works perfectly with our technology. Our assays are easy to bar code, so the Genome Analyzer's ability to multiplex samples has been an advantage. We recently increased from 12-plex to 96-plex in the same lane of the flow cell. At this point, we're collecting data at about 5 to 10× redundancy.

The Genome Analyzer has contributed greatly to the success of the ImmunoSEQ assays and software. In a single experiment, we can now directly determine the rearranged TCR β CDR3 sequences carried in millions of αβ T cells. The use of T cell genomic DNA, as opposed to cDNA, as a template for sequencing also permits us to estimate the relative frequency of each CDR3 sequence in the population of T cells, enabling us to perform global comparisons of the actual repertoires in two or more individuals. No one has ever been able to achieve this before.

Q. How do you analyze the data generated by the Genome Analyzer?

CC: That has been a real challenge and something we recognized would be critical to solve. Our co-founder, Harlan Robins, Ph.D. is a computational biologist and has developed some very sophisticated algorithms to interpret the data, and provide tools and applications for researchers to analyze, visualize, and interpret their data in a meaningful way.

The ability to process and analyze even more data will be important as we move from the Genome Analyzer, with its single flow cell to the HiSeq™ 2000 System, with its dual flow cell technology. We are in the midst of determining how we can leverage that amount of real estate, solve the computational challenges, and provide valuable data back to our customers. The interpretation of data has turned out to be a large part of our business.

Q. What surprised you when you first began using ImmunoSEQ to analyze TCR repertoire?

CC: We first used ImmunoSEQ to answer a very fundamental question: How many TCR β receptors do we have? From the peripheral blood T cells of two adults, we determined the CDR3 sequence in millions of rearranged TCR β genes. We were slightly surprised by the scale of the repertoire within an individual, finding that the total TCR β receptor diversity is around 3×10^6, at least four-fold higher than previous estimates[1]. While this is large, it comprises a negligible fraction of the estimated 5×10^{11} theoretically possible TCR3 CDR3 sequences.

Q. How did you use ImmunoSEQ to perform the TCR repertoire overlap study?

CC: We determined the overlap and effective size of the TCR repertoire by sequencing the CDR3 regions from millions of rearranged TCR β chain genes in CD8 T cells of seven healthy adults and performing pair-wise comparisons between individuals[2]. More than 5 million TCR β CDR3 sequence reads were generated from ~ 1 million template genomes for fourteen libraries: a naïve CD8 library and a memory CD8 library from each of seven individuals.

> "We recognized that the Genome Analyzer's phenomenal capacity to simultaneously sequence millions of molecules could be valuable in deciphering the adaptive immune system."

Our multiplex PCR primer targets all possible combinations of the noncontiguous (V_β), diversity (D_β), and joining (J_β) gene segments of the β chain locus. As expected, much of the predicted diversity in these sequences is generated at the V_β-D_β and D_β-J_β junctions of the TCR receptor. We used a model of VDJ recombination to determine whether the CD8 TCRβ CDR3 repertoire in each individual is randomly sampled from all possibilities, and found that a signficant fraction of the sequences were drawn from a defined subset of less than .11% of the estimated 5×10^{11} possible sequences. This finding greatly increased the chances that there would be some TCR receptor commonality between individuals.

Still, we were really surprised by the results of our pair-wise comparisons. When we compared the naïve repertoires of any two individuals, standard models would have predicted approximately 10 to 50 overlapping sequences that were amino acid identical. We found 10,000—in any comparison, independent of ethnicity or Class I HLA match. The overlap between the memory samples was smaller, because the composition of each person's memory repertoire is determined by his or her cumulative history of antigenic exposures. Nonetheless, the mean overlap between the memory repertoires of any two individuals exceeded 1,000 sequences.

Q. What could the discovery of common TCR repertoires mean for the development of diagnostics and therapeutics?

CC: Such a significant overlap between two unrelated individuals suggests that the phenomenon of public T cell responses—shared CDR3

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACG
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTT
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACG
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGACG
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA

sequences associated with an immune response to specific foreign or self antigens—may be more prevalent than previously thought. Its' potential implications in the diagnosis of disease could be profound, since these sequences could serve as surrogate markers for infections, cancer, and autoimmune diseases. Our customers are already using ImmunoSEQ to sequence the adaptive immune systems across populations and identify commonalities as they relate to certain diseases.

There could also be potential diagnostic opportunities if these sequences could be detected before the onset of disease, or used to determine the maturity of an immune system in transplant patients.

Q. You have mentioned a number of different therapeutic and diagnostic applications where ImmunoSEQ could provide value. Starting with transplantation, what information could it provide that would benefit monitoring pre- or post-transplant?

CC: ImmunoSEQ could help us define a "normal" healthy immune system by determining the number of T-cell clones necessary for a patient's immune system to effectively fight infection. In treating leukemia and lymphoma, the therapeutic strategies include bone marrow or cord blood transplants to restart the body's blood cell factory. In a bone marrow transplant, the host's immune system is weakened, with the new immune system destroying any residual from the original. With cord blood, you are transplanting a truly immature immune system. In both cases, it takes that immune system awhile to grow up to the point where it possesses millions of different T-cell clones. What we do not yet know is when during that process is a person ready for the real world? ImmunoSEQ could be used to determine when a patient's immune system is making enough T-cell clones, so that prophylactic antivirals and antifungals would no longer be necessary.

Q. Could ImmunoSEQ also benefit vaccine development by providing similar information?

CC: One of the challenges in vaccine development is that very large studies are required to prove that you are protecting people against infection. Since they cannot actively inoculate someone to prove they are immune, vaccine developers have to make educated guesses as to which formulation of the vaccine to carry forward into a large trial. ImmunoSEQ could assess each formulation, generating a profile of the number of different clones and the enthusiasm with which they expand within a modest number of individuals. Such a profile could positively impact the speed of development and effectiveness of new vaccines, if the profile predicts protective activity.

Q. Autoimmune diseases are often difficult to diagnose, resulting in advanced disease progression before treatment is begun. How could ImmunoSEQ be used to detect these diseases at an earlier stage?

CC: Autoimmune diseases occur when there is an overactive immune response against self-antigens. In this case your genetics are not deterministic, with many autoimmune diseases having a concordance rate in identical twins of only 30 to 50%. What is fascinating is that a given human leukocyte antigen (HLA) type may predispose a person to one disease, but a number of non-HLA loci may be necessary for disease onset. Genome-wide association studies (GWAS) have identified a number of genes that play a role in numerous autoimmune diseases. For example, the PTPN22 gene encodes a protein expressed in lymphoid tissues and has been associated with a range of autoimmune disorders, including Type 1 diabetes, rheumatoid arthritis, systemic lupus erythematosus, and Grave's disease[3]. It may be that the disease is initiated in predisposed people only after the wrong T cell receptor is made. We are already collaborating with researchers led by Jerry Nepom at the Benaroya Institute, concerning diabetes and multiple sclerosis (MS) to look at these diseases in more depth using ImmunoSEQ and see what the adaptive immune system can tell us.

Many different autoimmune diseases, such as lupus, are rather cryptic in their presentation, making it difficult to diagnose whether a patient is suffering from active lupus or another autoimmune disease such as MS. By profiling active and memory T cells, ImmunoSEQ could help provide a more definitive diagnosis. All this is speculative, but it is exhilarating to know we now possess the tools to actually test it.

> "Our customers are already using ImmunoSEQ to sequence the adaptive immune system across populations and identify commonalities as they relate to certain diseases."

Q. Could ImmunoSEQ be used for monitoring disease state in individuals with a relapsing, remitting course of an autoimmune disease?

CC: For many autoimmune diseases, there are episodes where patients are actively suffering (flares), followed by intermittent periods of remission. By profiling an individual and following them over time, we might be able to identify the clones involved in their disease state, particularly the ones that pop up during a flare. We could use ImmunoSEQ to monitor people and detect flares presymptomatically, enabling earlier intervention and hopefully better long-term prognosis.

We feel the technology could be used as a diagnostic, enabling improved autoimmune therapy selection. Treatments are getting more sophisticated, but their success is hit or miss. For anti-TNF therapy for Crohn's disease, there are responders and non-responders. Our technology could be used to identify if a person will respond to therapy.

Q. Could the identification of these autoimmune disease–associated clones be turned into targets for therapeutics?

CC: It would be a home run if we found a target that was useful in treating the disease. The complication is that the longer your disease has been going on, the more clones may be involved. At that point, interfering with any one of them probably would not be enough to tamp down the disease. It is possible that as we learn more about it, we might be able to identify a specific clone that is a good therapeutic target.

Q. What about other applications?

CC: Every time we present information about our technology, people come up to us with ideas on how it could be used, sometimes things we had never considered. For example, there is a lot of research going on into which cells are part of an antigen response. Right now, researchers stimulate a population of cells with a given pathogen or antigen, followed by a laborious process to separate out the cells that are antigen responsive. We could bypass all that by using ImmunoSEQ to create a pre-stimulation and post-stimulation profile of a population of cells and identify the exact sequences of the responding molecules. Those sequences could be plugged into an expression vector and presto, you have isolated them without performing any cell sorting and simplified a process that now takes months into a few weeks.

Q. Are there ImmunoSEQ assays for any other types of immune system molecules?

CC: In the human genome there are four T cell receptors genes: alpha, beta, gamma, and delta. We focused our initial efforts on beta, because of its diversity and the range of clinical applications it could impact. We have now expanded the ImmunoSEQ assay to other gene loci, including the TCR γ chain and IGH, encoding the heavy chain in antibodies. The addition of these two gene loci will deepen our ability to observe the adaptive immune system in action.

It has been very exciting to demonstrate that the assumptions the immunology field has operated under for the last 20 years were flawed. More importantly, we feel the ImmunoSEQ technology will help us to gain a deeper understanding of the adaptive immune system, enabling us to better test hypotheses and develop and measure the effectiveness of therapeutics and diagnostics to improve patient care.

References

1. Robins HS, Campregher PV, Srivastava SK, Wacher A, Turtle CJ, et al. Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells, Blood 114: 4099-4107.
2. Robins HS, Srivastava SK, Campregher PV, Turtle CJ, Andriesen J, et al. Overlap and effective size of the human CD8+ T cell receptor repertoire, Sci. Transl. Med. 2:47ra64.
3. Criswell LA, Pfeiffer KA, Lum RF, Gonzales B, Novitzke J, et al. Analysis of families in the Multiple Autoimmune Disease Genetics Consortium (MADGC) Collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes, Am. J. Hum. Genet. 76:561-571.

Illumina, Inc. • 9885 Towne Centre Drive, San Diego, CA 92121 USA • 1.800.809.4566 toll-free • 1.858.202.4566 tel • techsupport@illumina.com • illumina.com

FOR RESEARCH USE ONLY

© 2011 Illumina, Inc. All rights reserved.
Illumina, illumina*Dx*, Solexa, Making Sense Out of Life, Oligator, Sentrix, GoldenGate, GoldenGate Indexing, DASL, BeadArray, Array of Arrays, Infinium, BeadXpress, VeraCode, IntelliHyb, iSelect, CSPro, GenomeStudio, Genetic Energy, HiSeq, HiScan, TruSeq, Eco, and MiSeq are registered trademarks or trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners.
Pub. No. 070-2011-006 Current as of 02 March 2011

illumina®

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTT
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGAGC
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA