Querying the Whole Genome to Elucidate the Role of HLA Genetics in Autoimmune Disease

Researchers combine human genetics with computational biology to identify novel pathogenic risk loci.

Introduction

When he was training as a rheumatologist, Dr. Soumya Raychaudhuri was treating many patients struggling with immune-mediated conditions. Fascinated and frustrated by the disease complexity, Dr. Raychaudhuri began developing an interest in studying immune diseases. As a postdoctoral fellow, he conducted rheumatoid arthritis (RA) research, applying large-scale single nucleotide polymorphism (SNP) genotyping to his earliest genetic studies.

Now an Associate Professor of Medicine at Harvard Medical School, Dr. Raychaudhuri and his team are focused on understanding genetic susceptibility for complex, immune-mediated diseases such as RA, type 1 diabetes (T1D), and tuberculosis (TB). Genetic susceptibility to autoimmune diseases, such as RA, is associated with variants in human leucocyte antigen (HLA) genes in the major histocompatibility complex (MHC) region on chromosome 6. The encoded HLA protein molecules form a peptide-binding pocket that presents antigenic peptides to CD4+ T-cells. Amino acid variations in exons encoding the HLA antigen-binding domains are thought to increase susceptibility to specific diseases.[1] Dr. Raychaudhuri integrates computational approaches, high-density array-based genotyping, and next-generation sequencing (NGS) to infer the precise variation in HLA that fully explains risk. His goal is to identify genetic variations that increase a person's likelihood of developing an immune-mediated disease.

"To understand the genetic factors that make you susceptible to immune-mediated diseases, you need to query the whole genome using NGS or high-throughput genotyping," said Dr. Raychaudhuri. Through this approach we've identified common factors across diseases such as RA, T1D, and even TB." iCommunity talked with Dr. Raychaudhuri to learn more about how he's using NGS to identify novel pathogenic risk loci.

Q: What is your approach to studying complex diseases?
Soumya Raychaudhuri (SR): In traditional Mendelian genetic diseases, a mutation in one causal gene increases disease risk. For complex diseases, it's a spectrum of alleles rather than just one. For example, we've collaborated in RA research where we defined over 100 alleles that play some role in contributing to RA risk. We are trying to understand how these alleles work together to cause disease. This has led us to investigate how these alleles influence gene expression, which has developed our interest in transcriptomics. We think that many of these alleles influence enhancers or promoters, which has steered us toward epigenetic assays, ChIP-Seq, and most recently ATAC-seq (assay for transposase-accessible chromatin sequencing) to define parts of the genome playing a role in active gene regulation. Ultimately it's a systems biology approach we employ to link this data together to determine how many alleles influence disease risk.

Q: Why are you focusing your research on RA, T1D, and TB?
SR: The more we dug into it, the more we began to understand that there was genetic overlap between these diseases. The genetic factors that make someone susceptible to RA are closely related to pathways that are relevant to CD4+ T-cells. We became interested in T1D because it has tremendous genetic overlap with RA. We studied TB because of the importance of the disease and the number of people that are affected by it, and because CD4+ T-cell biology plays an important role in TB susceptibility. There is a functional link between these 3 diseases, and they are all important in terms of their worldwide morbidity and mortality.

Q: What role does the MHC play in immune-mediated disease?
SR: In RA and T1D, the MHC plays an important role in susceptibility. In humans, the MHC complex harbors over 200 genes. HLA genes have many possible variations and are involved in recognizing foreign antigens, as well as ensuring a match for transplanted tissue. In most instances, HLA alleles encode polymorphic amino acid residues that sit in the binding group of HLA receptors. They likely influence the efficiency with which antigens are found, because specific pathogenic antigens trigger immune-mediated diseases.

Q: How are HLA genes involved in T1D?
SR: Our group has invested a significant amount of time trying to identify the individual alleles within particular HLA genes that drive the risk of T1D. It has long been known that T1D is associated with

Soumya Raychaudhuri, MD, PhD, is Associate Professor of Medicine, Harvard Medical School Divisions of Genetics & Rheumatology; Department of Medicine, Brigham and Women's Hospital; Member of Partners Center for Personalized Genomics; Professor in Genetics, University of Manchester; and Associate Member, Broad Institute.

the MHC and its HLA genes. In 1987, John Todd and his group identified the 57th amino acid site in HLA-DQß1, a peptide-presenting protein, and the important role it played in T1D.[2] It's become apparent that position alone doesn't explain all the risks attributed to HLA genes in T1D. We wanted to identify the specific driving alleles. We took advantage of recent published data from Dr. Suna Onengut, collaborators, and the Type 1 Diabetes Genetics Consortium, where they typed tens of thousands of cases and controls using the Immunochip platform.[3] The MHC was extensively queried and we were able to apply that statistically to localize the precise association. By analyzing these samples, we determined that half the risk of T1D comes from position 57 in HLA-DQß1, and most of the other half comes from positions 13 and 71 in HLA-DRß1. The DRß1 alleles and the DQß1 alleles interact with each other to confer differential risk of T1D. Not only does this study give us a sense of how to approach risk prediction using HLA and T1D, but it also identifies specific residues to pursue in functional follow-up studies.[4]

Q: What is the mechanism in which variation in DRß1 and DQß1 increases T1D risk?
SR: There's probably multiple pathogenic peptides in T1D that interact with HLA molecules, DRß1 and DQß1, in different ways. Depending on which version of DRß1 or DQß1 a person has, their system is either more or less effective in interacting with the different pathogenic antigens. This impacts who gets T1D and who doesn't. We hope that this study will inspire people to look into what the specific antigens are and how they're interacting. There's a tremendous amount of biochemistry that needs to be pursued.

> "We determined that half the risk of T1D comes from position 57 in HLA-DQß1, and most of the other half comes from positions 13 and 71 in HLA-DRß1."

Q: What are some approaches to assess loci functions and their impact on disease?
SR: It's important to have some type of physiological response to measure. We'll stimulate T-cells and look at the response, which varies from person to person. We'll try to understand how the disease alleles alter that physiological response. There are many ways to stimulate T-cells and many subtypes of different T-cells. While it is important to understand the genetics of any individual physiologic response, it is equally important to identify what the right response is and what the right cell type is for a given disease.

Q: How has the Immunochip enabled your studies?
SR: Immunochip is widely used in T1D, RA, psoriasis, ankylosing spondylitis, and a range of other diseases and is a good technology to perform gene-based discovery. It facilitates low-cost genotyping of a huge number of samples. The ImmunoChip densely types SNPs across the MHC, providing a high degree of accuracy. Dense SNP typing allows us to use HLA imputation to infer HLA alleles within the MHC. The advantage of HLA imputation is that we can take existing data sets produced by the Immunochip and accurately infer what the HLA alleles should be. The HLA alleles themselves are challenging to sequence. They are always a blind spot. Most clinical labs today still use PCR-based protocols to do HLA typing. We're hungry for a technology to sequence the HLA gene effectively.

> "Immunochip is widely used in T1D, RA, psoriasis, ankylosing spondylitis, and a range of other diseases and is a good technology to perform gene-based discovery."

Q: How were you involved in developing the Immunochip?
SR: Illumina collaborated with the autoimmune disease community to select the set of SNPs for the Immunochip. We were involved in picking RA susceptibility SNPs. It was a fantastic collaboration with Illumina. We all benefitted from the input of others across different immunological diseases, resulting in a product that we could all use. What's been fascinating is just how successful it's been in its broad application to different immune diseases. A wide range of immune diseases have been queried with the exact same platform. We can learn quite a lot by continuing to mine those data sets without ever having to recollect DNA. That's powerful.

Q: What other Illumina BeadChips are you using?
SR: We used the HumanExome BeadChip to study age-related macular degeneration (AMD), identifying several rare coding variants that are associated with AMD susceptibility. The initial discovery was done by targeted sequencing of various genes. We validated the variants found to be associated, and identified susceptibility-related variants in C9 and C3.[5] Those variants are on the HumanExome BeadChip. It is a useful technology.

Q: How are you using NGS in your research?
SR: We are sequencing many genes and use different Illumina sequencing systems for different studies. We work through the Broad Institute or Partners Sequencing Center to access the NextSeq® 500 and HiSeq® Systems, in rapid and high-output modes respectively. If we need longer read lengths, we use the NextSeq System. If we have many samples and flow cells to fill, we'll use the HiSeq System. Illumina sequencing systems are widely used and well supported. We've been happy with the results from both sequencing platforms.

Recently, we've been performing transcriptome studies using SMART-seq2 and Illumina TruSeq® Targeted RNA Expression Kits on the NextSeq and HiSeq Systems. The SMART-Seq2 enables us to create libraries out of low RNA quantities, down to the single-cell level. That's been an important feature for us.

Q: How important is the sequencing speed of the NextSeq and HiSeq Systems?
SR: Sequencing on the NextSeq and HiSeq Systems is fast. The sequencing speed is rarely a rate-limiting step for us. In fact, we spend more time analyzing the data than actually sequencing.

Q: What software tools are you using to integrate various genomic data?
SR: We develop our own software tools to perform data analysis. Over the last year, our focus has been integrating genetic data with epigenetic data. We've written a series of methods, including GoShifter and epiGWAS, to help geneticists take their genetic data, put it in the context of epigenetic data, and prioritize individual candidate alleles that might be driving susceptibility. Epigenetic markers identifying gene regulatory sites and alleles that play a role in gene regulation are most likely to be causal in a locus. We also have undertaken efforts to look at transcriptional data in the context of genetic data.

> "Sequencing on the NextSeq 500 and HiSeq Systems is fast... In fact, we spend more time analyzing the data than actually sequencing."

Q: How do you see genomics impacting patient care for autoimmune diseases?
SR: I think it's clear that genomics will play a huge role in solving Mendelian and familial diseases in the next decade. What role genomics plays in complex diseases is an open research question. We and many others are interested in finding out the best ways to understand the genomics underlying complex diseases. It's an exciting area.

Q: What are the next steps in your research?
SR: In the last 5 years, there's been a tremendous amount of discovery in terms of identifying alleles and the role they play in immune-mediated diseases. We are trying to localize those associations to find what the causal alleles are and what they do. In HLA, we've had some success in RA and T1D, but we'd like to extend that success outside the HLA, to other loci and figure out what the driving alleles are across the genome. We want to understand how those driver alleles are changing the function of specific cell types and specific physiologic responses.

References:

1. Santamaria P, Yang Y. *Autoimmune Disease: Genetics.* John Wiley & Sons Ltd.; 2005.
2. Todd JA, Bell JI, and McDevitt HO. HLA-DQβ gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature*. 1987; 329: 599–604.
3. Onengut-Gumuscu S, Chen WM, Burren O, et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat. Genet.* 2015; 47: 381–386 (2015).
4. Hu X, Deutsch AJ , Lenz TL, et al. Additive and interaction effects at three amino acid position in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. *Nat. Genet*. 2015; 47:898–905.
5. Seddon JM, Yu Y, Miller EC, et al. Rare variants in CFI, C3 and C9 are associated with high risk of advanced age-related macular degeneration. *Nat. Genet.* 2013; 45: 1366–1370.

Learn more about the Illumina products and systems mentioned in this article:

HiSeq System, www.illumina.com/systems/hiseq_2500_1500.html

NextSeq 500 System, www.illumina.com/systems/nextseq-sequencer.html

TruSeq Targeted RNA Expression Kits, www.illumina.com/products/truseq-targeted-rna-expression-kits.html

Infinium HumanExome BeadChip Kit, www.illumina.com/products/infinium_humanexome_beadchip_kit.html

Infinium ImmunoArray-24 v2 BeadChip Kit, www.illumina.com/products/infinium-human-immunoarray-24-beadchip.html

Illumina • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

For Research Use Only. Not for use in diagnostic procedures.

 Pub. No. 1070-2015-004 Current as of 04 November 2015

illumina®