

Exploring HLA Diversity in Brazil

Drawing upon the great genetic diversity of the people of Brazil, Erick C. Castelli, PhD, is developing a database for the exploration of human HLA variation.

Introduction

The sixth chromosome of the human genome is home to the human leukocyte antigen (HLA) system, a unique group of genes essential for human health and immunity. A major role of HLA is enabling the immune system to differentiate between healthy and diseased cells. The immune system can detect certain threats, such as the presence in a cell of a viral protein or the product of a mutated gene. When this happens, HLA brings evidence of the threat, in the form of a small piece of protein, to the surface of the cell. Cells bearing such markers are then rapidly recognized and destroyed by other components of the immune system. Among other wonders, the immune system makes a special exception in the case of pregnancy, sparing the fetus from annihilation by the mother's immune system. The basis of this "immune tolerance" mechanism appears to be associated with HLA-G, one of several genes in the HLA system.

As a biologist and a computer programmer, Erick C. Castelli, PhD, has been working with HLA for more than 10 years. His graduate research, carried out in the lab of Eduardo Donadi, MD, PhD, at Ribeirão Preto University of São Paulo, was focused on the relationship between *HLA-G* gene polymorphism and cancer susceptibility.

Now, as a member of the Department of Pathology at São Paulo State University (UNESP – Botucatu) in Brazil, Dr. Castelli is taking advantage of the country's vast population diversity to establish a database of human HLA gene variation. This database is becoming an essential resource for Dr. Castelli and other biologists who wish to understand how HLA genes are regulated and the associations between HLA region polymorphisms and human phenotypes.

iCommunity spoke with Dr. Castelli to learn how nextgeneration sequencing (NGS) is contributing to the advancement of knowledge about this complex and essential region of the human genome.

Q: What sparked your interest in conducting HLA research? Erick Castelli (EC): I began by studying the *HLA-G* gene in Dr. Donadi's laboratory. He's an expert in the field and has been conducting HLA research for more than 30 years. The *HLA-G*



Erick C. Castelli, PhD is a Principal Investigator in the Department of Pathology at the São Paulo State University (UNESP – Botucatu) School of Medicine.

gene produces an immune tolerance protein that is active at the fetal/maternal interface. We were trying to understand the genetics of *HLA-G*. It is quite different from the other HLA genes because it is highly conserved. Our main focus was on the population genetics around this gene. As we learned more about its structure, we were able to start looking at how the expression of the gene is managed.

Now that I am here in Botucatu, my research is expanding to include the other genes in the HLA region, and the associations between sequence variants and disease states. It is helpful that I am situated here in Brazil, which is a very good place to do HLA research.

Q: Why is Brazil a good place for HLA research?

EC: We have a very mixed population here in Brazil. In fact, we have a saying among scientists that "Brazil is a good variation repository." This country has been influenced by immigration from Europe, Africa, and Asia. The genetic background of the Brazilian population is highly diverse as a consequence of five centuries of interethnic crosses of people from three continents, including Portuguese, Africans, and autochthonous Amerindians. Later, other immigrants from Europe (mainly Italy and the Netherlands) and Asia (mainly Japan and China) came to Brazil in the past century.

I saw that variability when I started working on *HLA-G* eight years ago. Even though *HLA-G* itself is very highly conserved, we found divergent sequences when evaluating its regulatory regions. Any time I look at HLA in the Brazilian population, I

can find something new. We have genetic variation here that you do not find elsewhere, but most of the frequent variations found worldwide are also found in Brazil.

Q: What is the focus of your current research studies?

EC: We're conducting two different research studies. The first study is an evaluation of HLA genetics from a population point of view. For this study, we are using normal, healthy individuals. No one has performed deep sequencing of the HLA region on such a mixed population, and the available databases describing the polymorphisms in the HLA genes usually does not consider their regulatory regions. So, we are determining the genetic structure of this region and creating a database that can be used for future comparative studies. We're looking at the sequence of the entire region, not just the exons, because as we saw with *HLA-G* that there is a lot of interesting variation outside the exons.

The second study is focused on the relationship between genetic structure in the HLA region and various disease states. Specifically, we are looking at the genetic variation in the HLA region between several tumor types and autoimmune diseases (such as diabetes), and normal individuals, and characterizing the correlations between HLA polymorphisms and such diseases.

"We have a saying among scientists that 'Brazil is a good variation repository."

Q: What's unique about your approach to studying the HLA region?

EC: First, we are sequencing the entirety of every gene — all the promoter segments and the 3' untranslated regions, in addition to the coding regions. Second, we're performing deep sequencing on many individuals. And third, we are creating new tools that are allowing us to complete this work. These include our own software analysis program and strategies to develop a good sequence map of the HLA region.

Q: What are some of the challenges you've encountered in your HLA research?

EC: We are starting from the beginning, in the sense that there is no established methodology to evaluate all HLA gene segments and regions. Much of the genetic analysis of HLA has been focused on answering specific clinical questions, mainly detecting the presence of known sequences to make a patient/donor comparison. We need to look at much more of the data so we can ask many different questions. Also, working with HLA is not the same as working with another genome region. The HLA genes are very similar to each other. You cannot mix several genes at the same time, unless specific bioinformatics approaches are developed to correctly separate data from each gene.

Today our main problem is the presence of repeated sequences within the region and the fact that the HLA genes are quite similar. Sometimes, for a short nucleotide segment (~ 40 nucleotides), it is not possible to define the origin of this segment. We have written new software and created data flow pipelines using open-source tools to help us address this challenge, but we are still developing strategies to work with the data.

"We are starting from the beginning, in the sense that there is no established methodology to evaluate all HLA gene segments and regions."

Q: How long have you been working with the MiSeq System?

EC: We have three MiSeq systems here in the University campus, and I have been working with the MiSeq System for the past two years. We are performing sequencing with NGS systems to determine HLA variability.

Using Sanger sequencing, we were limited to working with just one gene at a time. It took me a whole year to evaluate just one or two genes. We moved to NGS because it enables us to work with several genes at the same time. With the MiSeq system, I can work with several thousand nucleotides in one experiment and process all of my genes at the same time.

Q: What have you discovered in your research?

EC: We are finding a lot of polymorphism in HLA regions outside of the ones that have been investigated in clinical studies. Over the entire region, we find a great amount of variability that has not been seen before because the Brazilian population is so much more diverse than other countries. This is not really surprising to us. What's exciting is that it's all new information. We're creating a database covering the genetic diversity of the HLA region in Brazil. It will be a comprehensive HLA database since it also considers the regulatory regions and introns.

Q: Were you surprised by the extent of polymorphism in the HLA regions?

EC: I was expecting some extent of variation. We weren't sure how much there would be. It was very exciting for us to discover such a high degree of variability in the HLA region in the segments that are usually not evaluated.

"With the MiSeq System, I can work with several thousand nucleotides in one experiment and process all of my genes at the same time."

Q: Is there any potential clinical relevance in the polymorphisms you are finding in the HLA region? EC: A large part of our curiosity about HLA genes and their surrounding sequences is that we believe that most of these polymorphisms have some effect on gene expression. After we have established our initial database and methodology, we want to investigate that more closely. We want to establish specific associations between polymorphisms and the expression levels of HLA genes.

We already know that some of HLA genes, like *HLA-G*, are associated with immune tolerance, and several groups in the world are trying to figure out how they work. There are several situations where being able to control gene expression in this area would be useful. For instance, it would be beneficial if we could treat autoimmune disorders by enhancing immune tolerance by increasing *HLA-G* expression.

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

FOR RESEARCH USE ONLY.

© 2015 Illumina, Inc. All rights reserved. Illumina, MiSeq, and the pumpkin orange color are trademarks of Illumina, Inc. and/or its affiliate(s) in the U.S. and/or other countries. Pub. No. 070-2014-005 Current as of 06 January 2015

illumina