Collaborative Genotyping Effort Helps Identify Heterogeneity in the Diverse Hispanic Population

LABioMed/Harbor-UCLA Researchers Drs. Taylor and Rotter team up with Illumina FastTrack Services to identify disease-associated SNP markers prevalent in the Hispanic population.

Hispanic ancestry is a complex admixture of diverse origins that can be traced to Europe, Africa, the Americas, and the Caribbean. While Hispanics represent the second largest population in the United States, they tend to be underrepresented in research studies, including many high-profile genome-wide association studies (GWAS). GWAS is used to identify disease-associated genes, which can differ among populations, and enable personalized therapeutic strategies. To address this disparity, the National Heart, Lung, and Blood Institute (NHLBI) initiated the Hispanic Genotyping Array project to identify disease-associated genetic variants in The Hispanic Community Health Study/Study of Latinos (HCHS/SOL). Awarded to Illumina in 2012, the project is being overseen by the FastTrack Services group, a team of in-house expert scientists that perform client-specified genotyping and sequencing services.

Over the past year, Illumina FastTrack Services collaborated on the NHLBI contract with Kent Taylor, Ph.D., and Jerome I. Rotter, M.D., at the Los Angeles BioMedical Research Institute (LABioMed) at the Harbor-UCLA campus of the UCLA School of Medicine to identify a high-quality data set of single nucleotide polymorphisms (SNPs) that delineate the various groups comprising the Hispanic population. Dr. Taylor, an expert in bioinformatics and Native American ancestry, and Dr. Rotter, an expert in Hispanic genetic epidemiology studies, selected the SNPs and Illumina processed the arrays. This effort resulted in the Illumina Hispanic Array BeadChip, which will greatly accelerate gene findings for grave public health problems observed primarily in the U.S. Hispanic population. The array is a custom version of the Infinium® HumanOmni2.5 BeadChip, which includes common and rare SNP content from the 1000 Genomes Project. iCommunity caught up with Drs. Taylor and Rotter to hear about the drivers for this study and the results obtained so far.

Q: What are the specific goals of the Hispanic SNP data set study?

Jerome Rotter (JR): The goal is to identify SNPs that distinguish the heterogeneity among U.S. Hispanics that is not captured on existing arrays. These include SNPs that distinguish Hispanics of Caribbean origin (Puerto Rico, the Dominican Republic, and Cuba) from those descending from populations in the continental Americas (Mexico, Central America, and South America). To capture the range of Hispanic origins throughout the U.S., the study included samples from 14,000–16,000 Latinos in four cities—Miami, New York, Chicago, and San Diego.

> "Illumina FastTrack Services processed all 14,000 samples in about 7 weeks, which was incredible."

Kent Taylor, Ph.D. (left) is the Director of the Laboratory for Molecular Genomics, Bioinformatics, and High-Throughput Genotyping and Professor of Pediatrics at LABioMed/Harbor-UCLA Medical Center. Jerome I. Rotter, M.D. (right) is the Director, Institute for Translational Genomics and Population Sciences at LABioMed/Harbor-UCLA Medical Center.

Q: Why is disease research in the Hispanic population difficult?

Kent Taylor (KT): Accommodating the large heterogeneity is difficult because Hispanics are not a single group. There are Mexican-Americans, Central Americans, and South Americans. They have mixtures of European genes from when the Spaniards came to the Americas. They also have differing Native American genes from groups like the Aztec, Maya, Inca, as well as the Taino Indians in the Caribbean. The slave trade brought Africans to the Caribbean, Brazil, and Honduras among other countries. As a result, there is large heterogeneity within the Hispanic gene pool and it's very hard to sort out.

JR: The Hispanic population is the second largest group in the U.S. after the European-origin population, and it's the fastest growing. Genetic tools for Hispanics are not as well developed as they are for other populations, such as European-origin Americans, Chinese, or African Americans. Hispanics weren't even part of the original HapMap project, the endeavor to describe gene variants affecting health, disease, and response to drugs and environmental factors. When the NHLBI committed to this major epidemiologic study, they realized the importance of developing markers that reveal the heterogeneous Hispanic population, especially the Native American component.

Q: How were the SNPs selected for the array?

KT: We selected SNPs that could capture the diversity in the 1000 Genomes Project and differentiate Hispanics from African and European populations. This was based on a calculation known as Informativeness for Ancestry Assignment, a general measure for determining the amount of information that multiallelic markers provide about individual ancestry. We analyzed how much Native American SNP content was in the 1000 Genomes subjects and used that as a trait to find SNPs by normal GWAS methods. We added a number of SNPs, including SNPs from European or African American association studies for cardiovascular disease, mitochondrial SNPs, and SNPs associated with known major histocompatability complex diversity.

> "The data quality is excellent and that's due to the high-quality DNA used in the project and the excellent data quality from Illumina."

Q: How did you perform array analysis and what was the level of quality obtained?

KT: FastTrack Services processed all 14,000 samples on the Illumina HiScan™ system in about 7 weeks, which was incredible. We were beginning to receive data at the two-month mark.

JR: The data quality has been excellent and that's due to the high-quality DNA used in the project and the excellent data quality delivered by the Illumina array. Processing of the arrays has gone very well. Out of the 14,000 samples, we had a failure rate of only about 30. That's quite low.

Q: What role has sequencing played in identifying SNPs in Hispanics?

KT: Sequencing using the HiSeq® 2500 system enabled us to find many SNPs, with variations in Hispanic populations from Puerto Rico, Columbia, and Los Angeles-based Mexican-Americans. However, we found that the sequencing data was incomplete and doesn't capture Native American ancestry found in the Hispanic population. We need to sequence more Native American subjects.

Q: Has the data provided a clearer picture of the variations in Hispanic populations?

KT: Of the 14,000 samples run so far, we are pleasantly surprised that the selected SNPs allow us to differentiate between many of the Hispanic populations. We can easily see three distinct Caribbean signatures from Cuba, the Dominican Republic, and Puerto Rico, and three distinct groups from the Americas with signatures of Mexican-Americans, Central Americans from Guatemala and Honduras, and South Americans. We can differentiate between groups with SNPs from Mexico versus Central America populations. Even with our limited Native American SNP set, we can distinguish between Hispanics with Mayan and Aztec ancestry. We can also see a signature from the Inca population and from Native Americans found on the other side of the Andes. That's remarkable because existing standard arrays don't allow you to see all that.

Q: What other project-types do you believe are best suited for Illumina FastTrack Services?

JR: Illumina FastTrack Services is ideal for groups that don't have a high-throughput genotyping lab. If you're going to work under stringent contractual arrangements, the FastTrack Service makes more sense than an ordinary genotyping lab. The FastTrack laboratory has a certain service ethos, which is particularly appropriate for large-scale contracts.

Q: What are the next steps in your research?

KT: I'm performing an analysis to calculate just how much Hispanic heterogeneity is missed using standard genotyping arrays, and how much more information this new array offers researchers. Our colleagues will be able to use this data to rationalize their selection of this array for their studies.

We're also in the process of preparing a grant to do more sequencing of Native American subjects. We need to sequence samples from the various parts of the Americas so that we have more SNP data to inform the development of the next generation of Hispanic arrays.

Learn more about the HumanOmni2.5 BeadChip and Illumina FastTrack Services:

www.illumina.com/applications/genotyping/human-genotyping-arrays/omni-arrays.ilmn

www.illumina.com/services/microarray-services.ilmn

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

FOR RESEARCH USE ONLY

Pub. No. 370-2014-001 Current as of 14 January 2014

illumina®