Sequencing Hundreds of Thousands of Samples to Gain Insights into Novel Variants, Biology, and New Therapies

The Regeneron Genetics Center is using the HiSeq® 2500 System to perform exome sequencing of Geisinger Health System MyCode samples and other patient-consented data sets to identify relevant genetic associations.

Introduction

Important insights and discoveries arise when large patient data sets are generated and analyzed using the power of genomic sequencing methods. In January 2014, Regeneron Pharmaceuticals and Geisinger Health System announced a partnership focused on using genomic analysis to improve drug development and patient care capabilities. Under the partnership, Regeneron will sequence DNA samples from at least 250,000 patients in Geisinger’s MyCode program and pair this information with their Electronic Health Records to fuel novel gene discovery and genomic medicine implementation.

January 2014 was also when the Regeneron Genetics Center (RGC) was founded to mine other patient-consented clinical data sets for medically relevant genetic associations.

Researchers at RGC are using the HiSeq 2500 System and Infinium® HumanOmniExpressExome BeadChip array to analyze patient data. With the power of large-scale exome sequencing, the RGC hopes to identify genes that cause or influence a range of human diseases, some of which might present new targets for drug development. Regeneron has a world-class mouse genetics program and uses a genetics-driven drug discovery and development platform that has been instrumental in the production of several successful drugs, such as Arcalyst (rilonacept) for a rare genetic disease and Praluent (alirocumab) for hypercholesterolemia.

Q: What prompted the creation of the RGC?

Aris Baras (AB): Regeneron has been focused on mouse and human genetics since our beginning nearly 30 years ago. We realized that recent technological advances have opened up a new frontier in this space and heightened the impact it could have on developing new medicines for patients in need.

We have a history of utilizing and building new technologies to tackle challenges in drug development - for instance with our proprietary VelociSuite antibody platforms – and we took a similar approach in building the RGC.

One of the remaining bottlenecks in drug development is identifying the best new targets to explore and developing therapeutics against them. Human genetics is a key resource for identifying new targets, or validating existing targets and drug development programs. A significant portion of our pipeline includes targets and programs that are derived from human genetic discovery. We built the RGC to make this process stronger and more efficient.

Q: What value does high-throughput sequencing add to your goals at the RGC?

AB: To date, the things that have been most useful for Regeneron’s development efforts come from looking at the real variation of genes that can impact the phenotypes of interest. To do that, we have to perform large-scale sequencing of tens of thousands, if not hundreds of thousands, of individuals to identify these large-effect mutations that are exceedingly rare in the population. We need high-throughput technologies to detect rare, important, and functional mutations and correlate them with diseases so that we can gain insights into novel variants, new biology, and new therapies.

Q: Why did Geisinger and Regeneron partner on the MyCode Community Health Initiative?

AB: We wanted to build a large-scale and comprehensive database of genomic sequence information combined with rigorous and longitudinal health records. Approximately 4 million people who live in central and northeast Pennsylvania seek care through the Geisinger Health System. So Geisinger has a strong phenotypic database and a trusting relationship with its customer community. Many of its patients were willing to participate in the MyCode Community Health Initiative. What’s unique is that there are decades of multigenerational health records available within the Geisinger Health System. Most family members stay in the Pennsylvania area. Among the patients we’ve sequenced, there’s an average of more than 10 years of health records.

Q: Why did Geisinger and Regeneron partner on the MyCode Community Health Initiative?

AB: We wanted to build a large-scale and comprehensive database of genomic sequence information combined with rigorous and longitudinal health records. Approximately 4 million people who live in central and northeast Pennsylvania seek care through the Geisinger Health System. So Geisinger has a strong phenotypic database and a trusting relationship with its customer community. Many of its patients were willing to participate in the MyCode Community Health Initiative. What’s unique is that there are decades of multigenerational health records available within the Geisinger Health System. Most family members stay in the Pennsylvania area. Among the patients we’ve sequenced, there’s an average of more than 10 years of health records.

For Research Use Only. Not for use in diagnostic procedures.
Geisinger was also a pioneer in adopting electronic health records and now has been a leader in the use of those for analytical and research purposes. So, with the partnership, we have put together a huge database of all those individuals who have been sequenced, along with their health records.

Q: What do Geisinger and Regeneron hope to gain, both individually and together?
AB: For Regeneron, this partnership can help guide our drug development efforts by studying human genetics. Geisinger has a long history of being involved in genetics research, so this partnership builds on its existing research programs. Geisinger is also forward-thinking in using genomics and clinical genetics to serve its patients and run its health system better. Geisinger just celebrated its centennial this year, and I think that its legacy might involve these innovative genomics programs.

“We need high-throughput technologies to detect rare, important, and functional mutations and correlate them with diseases so that we can gain insights into novel variants, biology, and new therapies.”

Q: What is the value of this partnership to patients who participate?
AB: Geisinger has trained its physicians and genetic counselors to return some of these CLIA*-confirmed genetics findings, so for some patients, there might be an immediate gain of important health information. Beyond that, I think many people participate because they are contributing to long-term, meaningful research that will one day hopefully lead to better treatment and medicines for everyone. Because patients in Geisinger’s system place so much trust in the health system, we have the opportunity to reach out to these individuals and conduct further clinical research around interesting diseases, or individuals who have interesting genotypes that we want to learn more about. These are key steps in the drug development process.

Q: How many samples have you sequenced to date and how many do you hope to sequence by the 5-year mark?
AB: We’ve sequenced 60,000 Geisinger samples so far and hope to sequence 250,000 within the next few years.

Q: Have you identified any new variants as a result of sequencing these samples?
AB: We’re making many discoveries. Researchers from the RGC and Geisinger recently published a paper showing that inactivating mutations in the ANGPTL4 gene are associated with a significantly reduced risk of coronary artery disease (CAD) in humans.5 We hope this is the first of many notable peer-reviewed publications that will arise from the collaboration.

Q: What is ANGPTL4 and what conditions are associated with it?
AB: ANGPTL4 encodes for a protein that inhibits lipoprotein lipase (LPL), an enzyme involved in the breakdown of triglycerides, a form of fat derived from foods. Studies have found that activation of LPL leads to the reduction of circulating triglycerides, increased levels of which are thought to be an independent risk factor for ischemic cardiovascular disease. It was hypothesized that genetic mutations inactivating ANGPTL4 would lead to activation of LPL, low levels of circulating triglycerides and reduced risk of cardiovascular disease. We were looking to validate this hypothesis through genetic analyses and functional modeling.

“HiSeq 2500 Systems are incredibly reliable and they produce high-quality data every time we run them.”

Q: What specific variants of ANGPTL4 did you pursue further in mouse studies?
AB: We found that individuals with 1 or 2 copies of the p.E40K mutation, which was previously known to inactivate ANGPTL4, had about a 19% lower risk for CAD. Patients with one of 13 other ANGPTL4 loss of function mutations that we newly identified had an almost 45% reduction in CAD risk.

Using our proprietary VelociGene technology, we were able to develop animal models that supported the human genetic findings. The Velocimmune platform was used to create a human monoclonal antibody inhibitor of ANGPTL4 that reduced triglyceride levels in mice and nonhuman primates.

Q: Why did you choose HiSeq 2500 Systems for your studies?
John Overton (JO): We chose them based on our extensive experience with it and the consistency of the data. HiSeq 2500 Systems are incredibly reliable and they produce high-quality data every time we run them. We’re able to sequence thousands of samples simultaneously. We run the HiSeq 2500 system in high-output mode and aim to sequence to a depth of 20× coverage across at least 85% of the targeted bases in our exome design.

Q: In what instances are you performing HumanOmnixExome BeadChip analysis?
JO: We complement the exome sequencing we’ve done with our Geisinger samples using the HumanOmnixExome genotype kit. We do that with all our population-based studies. Performing BeadChip analysis and exome sequencing gives us the coding and noncoding variation down to a low allele frequency.

Q: In addition to exome sequencing, are you also performing RNA-Seq?
JO: We’ve done very little RNA-Seq so far. However, we are starting to ramp up those studies with human samples in collaboration with Geisinger.

*Clinical Laboratory Improvement Amendments

For Research Use Only. Not for use in diagnostic procedures.
Q: What are the next milestones in the Regeneron/Geisinger partnership?
AB: There are some numerical milestones for the partnership. For example, we are aiming to sequence 250,000 patients in the next few years, which will help power the discovery and understanding of even more genes implicated in disease.

We’re also at the stage where results and impactful discoveries matter more to us than anything else. In addition to moving our pipelines forward, we want to advance science and benefit our patients through research. We hope these projects will enable development of important new therapeutics in areas of great unmet need, benefiting Geisinger patients and people worldwide.

We’re also thinking about deeper phenotyping initiatives. Geisinger has a tremendous number of clinical records and phenotype data available. We’re considering new initiatives including molecular phenotyping, where we can add RNA, proteomic, and metabolomics data to strengthen our ability to make gene-disease associations.

“Performing BeadChip analysis and exome sequencing gives us the coding and noncoding variation down to a low allele frequency.”

Q: How will the results of this partnership lay the groundwork for new options in the diagnosis, prevention, and treatment of various medical conditions?
AB: The foundation and groundwork is already there. For Geisinger, they already have the CLIA infrastructure to take validated findings and genetic results and return those to patients. That program is really ramping up. They’ll also take the CLIA-validated pathogenic variants and associate them with clinical phenotypes before contributing those findings in public databases like ClinVar. That data will be a major contribution to the field, thanks to the scale of the program that Geisinger has.

On our end, Regeneron has already been successful with a couple of programs, like our PCSK9 inhibitor, going from genetic discovery all the way to development and commercialization of therapies, but we expect the RGC will make this process more efficient and potentially shorter. We’re operating at such scale that we are attempting to work on every target in our pipeline, as well as to identify new targets.

We’ve made great advances in understanding our existing targets and advancing our programs by providing new human genetics evidence around disease associations. Overall, we’re excited to be contributing to Regeneron’s powerful R&D engine, all in the goal of bringing important new medicines to patients, faster and repeatedly.

References

Learn more about the Illumina system mentioned in this article: