Over the last ten years, a great shift in research prerogatives has taken place. Armed with knowledge of the roughly 22,000 human genes and other interspersed genetic elements that make up the human genome, scientists have become focused on identifying how this genetic material influences human disease. To this end, research has become preoccupied with genetic variants—that is, individual variations in the genetic code—for the roles they may play in disease, and a body of compelling data has been collected in this area. The use of methodologies like genome-wide association (GWA) studies has certainly led to progress in defining the possible genetic basis of a variety of conditions. But with this progress comes the realization that uncovering common genetic variants may only be scratching the surface.

Hundreds of GWA studies have been published over the last ten years examining the contributions of common variants to a variety of diseases, but according to David Goldstein, most of the work in understanding genomic control is yet to be done. Goldstein—who is director of the Center for Human Genome Variation at Duke University and a molecular genetics and microbiology professor—is now looking beyond the common genetic variants, and instead focusing on what is rare. He is not the only one, either. “Although there are differences in view about just how successful a GWA study is or isn’t,” he says, “it is now clear and accepted by all that, at best, it has revealed only a minority of the genomic control over disease.”

Shared by fewer than 3% of people in the population, individual rare variants, or combinations thereof, are a promising new direction and may answer the genomic control question that researchers like Goldstein are searching for. But their low frequency in the human population makes identifying these rare variations a challenge.

“The problems that still need to be overcome for finding rare variants have to do with resolution of microarrays and their inability to find all of the rare variants, either directly or indirectly,” says Stephen Scherer, director of the Center for Applied Genomics at Toronto’s Hospital for Sick Children and professor of molecular genetics at the University of Toronto. Microarrays—which have been the key tool for conducting GWA studies up to now—can only detect variation that has been designed onto the array. Therefore, rare single nucleotide polymorphisms (SNPs) or copy number variations that might be involved in a particular disease would be missed.

Tools of the Trade

“GWA studies have been the primary workhorse in the search for variation, but at its heart, GWA is designed specifically for common variation,” says Goldstein. “The whole paradigm works by identifying the polymorphisms in the human genome, and until recently it was the only technique that was comprehensive for variation.”

The microarray platforms used in GWA look at gene expression or structural changes within the genome. Tiling arrays, which are microarrays composed of short probes that can span an entire genome, can evaluate known and unknown genes, and are manufactured by companies including Affymetrix, NimbleGen, and Agilent. On average, those arrays based on photolithographic manufacturing can hold up to six million discrete features and each contain millions of copies of a single probe. Tiling arrays based on mechanical spotting or printing can hold an average of 400,000 features.

To accurately analyze any variant, researchers need large sample sizes to obtain statistical significance; to this effect, microarrays are designed specifically to analyze the target variant in high numbers of people at a low cost. But because microarrays must be programmed in advance to search for a specific known variant, these arrays are not a discovery tool. This has hampered their usefulness when it comes to rare variant identification.

The Next Generation of Variants

Now, when it comes to finding disease-associated rare variants, many scientists are pegging their hopes on a technology that has exploded in recent years: next-generation sequencing. The technology provides a means to analyze greater numbers of whole genomes in order to search for rare variants, but it is bogged down with voluminous data and technical limitations. Furthermore, sample sizes—though expanding—are still restrictive. In the face of these drawbacks, researchers are currently trying to determine how to better GWA studies by integrating the advantages of next-generation sequencing with microarrays.

In August, Robert Hegele—director of the London Regional Genomics Center and professor of medicine and biochemistry at the University of Western Ontario—published a GWA study on hypertriglyceridemia (1), a lipid disorder characterized by high levels of blood fat. Merging technology, his team used next-generation sequencing to find rare variants that would have been missed using microarrays. After analyzing the genomes of 500 people, the researchers found that the majority of variation was not caused by common variants (directly matched to the phenotype). The researchers identified four genes with most relevance as determined in the initial study, and resequenced these genes in the same patients using next-generation methods. What they found was that these genes containing common variants contained rare variants as well.

Next-generation sequencing systems are now being used to sequence large numbers...
of human genomes, and researchers hope that these efforts will lead to identifying more rare variants within the population. The 1000 Genomes Project, launched in 2008, plans to create an improved map of the human genetic variation by sequencing the genomes of 1200 individuals from around the globe. The project has already released preliminary data. And dropping costs means better access to next-generation sequencing tools, which will enable even more data on rare variants to emerge.

Data: Not Enough, or Too Much?

But even with increasing sequencing capability, Brett Abrahams, assistant professor of genetics and neuroscience at Albert Einstein College of Medicine, says that there are still issues with the technology’s scope. “Because rare variants are often observed in only one patient,” he says, “it becomes a challenge to be able to conclude that the variant plays a role in disease, without having any context to effectively compare it to, as one would with a common variant.”

The inherent problem with rare variants is that they are, in fact, rare. “Depending on the population, a typical GWA study is unable to look at rare variation effectively because the sample size isn’t large enough,” says Abrahams. “So if you see something that looks deleterious like a truncating mutation, and you never see it again in another affected individual, on a statistical level it is impossible to draw any inferences.”

According to Hegele, to adequately study rare variation would require as many as 2000–3000 participants. “What we consider to be a big GWA study using microarray technology is very small for studying rare variants,” says Hegele. “We need to have multiple people with the rare variant involved in a study to feel confident about drawing conclusions.”

This will require more efficient and cost-effective genome analysis. “I suspect there will need to be a third-generation-type breakthrough where the DNA sequence and structure is being assayed directly from the chromosome strand,” says Scherer. “I’m not complaining about the latest technologies—they are great—but there will need to be another leap forward in technology.” Such technology would lower the cost and time of sequencing to the point where studies of 2000 or more participants becomes finan-
cially feasible. But even if such capability is developed and a new library of rare variants constructed, the research community still must consider the greater question: how these variants impact disease.

Clinical Relevance

“The issue isn’t how we can identify these rare variants, but what can we conclude once we have determined them to be present,” says Abrahams. “The problem is compounded even further because some rare variants may actually exert effects through interaction with other variants.”

Writing recently in Cell, Mary-Claire King, a University of Washington professor of genome science, noted that there are indeed current strategies to identify rare variants in individuals (2). But she cautioned that researchers should be careful before claiming that any variant is associated with a specific disease. “Biological relevance,” she warns, “must be established before a mutation can be causally linked to a disorder.”

According to Hengele, the major research focus in the future will be to understand rare variants’ function in disease pathogenesis in order to predict responses to treatments. “If you know the pathway that is being affected, and trace it back to the genomic level, that might help researchers and physicians come to the right treatment for specific patients faster.”

The disease architecture may be such that the genetic underpinnings vary based on individuals, says Abrahams; then researchers will have no choice but to look at each case as its own entity. “But what people are hoping is that despite the actual disease contributors being different among individuals, the pathways that are impacted will be shared, allowing for therapeutics to be broadly effective.”

Goldstein foresees third-generation sequencing greatly influencing the hunt for rare variants. “In less than five years, we will have our genomes sequenced and interpreted,” he says. “We’ll be able to sequence large sections of genomes quickly, so most of the genes that carry variants will be identified. What happens with that information is the only question that remains.”

References


Written by Erin Podolak. [ ]